

## Data-driven Communication Research: Paradigmatic Shift and Methodological Innovation

Shaojing Sun<sup>1</sup>, Ying Ding<sup>2</sup>, Giuseppe A. Veltri<sup>3</sup>, Hongzhong Zhang<sup>4</sup>, Li Tang<sup>5</sup>, Yuxiang Zhao<sup>6</sup>, Yujia Zhai<sup>7</sup>

<sup>1</sup>Fudan University, China

<sup>2</sup>Indiana University, USA

<sup>3</sup>University of Leicester, UK

<sup>4</sup>Beijing Normal University, China

<sup>5</sup>Fudan University, China

<sup>6</sup>Nanjing Technological University, China

<sup>7</sup>Nankai University, China

### Abstract

Over the past decade, the growing amount of accessible data, including but not limited to academic publications, media contents, and social tweets, has brought both opportunities and challenges to social and behavioral science researchers. Advances of data analytics, on the other hand, open a new window for researchers to look at human communication processes. Traditionally, empirical communication research is more built on deductive reasoning, finite sampling, and parametric statistics. Now, data-driven research affords a panoramic view and innovative analysis of large-scale data. As such, this SIE is aimed to achieve the following goals: 1) to elucidate the linkage between data science and communication studies; 2) to introduce methods and tools which can inform communication and new technology studies; 3) to illustrate how data-driven research can help address substantive communication questions.

**Keywords:** communication science, data science, data-driven discovery

**Copyright:** Copyright is held by the authors.

**Contact:** shaojingsun@gmail.com

### 1 Purpose and Intended Audience

The advances of new technologies and abundance of public data have rendered a high demand for innovative communication research. Over the past decades, researchers have mainly relied on traditional methods (e.g., finite sampling, manual coding) to study communication phenomena. Such an approach, however, may suffer from the limitations of samples and human coding errors. The development of data science has offered new opportunities to examine communication data (e.g., scholarly publications and media contents), by improving the rigor and objectivity of the research process.

The SIE is attempted to draw researchers from communication, data science, and information science, to discuss potential collaborations. More specifically, the discussion is intended to advance data-driven communication research, addressing substantive questions which may seem to be impossible to answer with traditional methods. To that end, the session will explore existing resources for conducting data-driven communication research. Furthermore, the session will demonstrate some case applications by integrating communication studies and data science.

The iConference is an ideal avenue for gathering researchers with common interest to explore data-driven

communication research. The proposed session draws participants from the communication, information, and data science to discuss both conceptual and methodological issues pertaining to the discovery of communication knowledge. Building on the recent work of the panelists, the session will focus on the following: a) to explain the difference between data-driven and theory-driven communication research, b) to explore substantive communication domains which may benefit most from the integration of the aforementioned disciplines, and c) to demonstrate the value of applying innovative data analytics to study communication.

## 2 Proposed Activities

Using a roundtable discussion format, the session will proceed as the following:

- Introduction to data-driven communication research (10 minutes)
- Seven talks followed by Q & A (20 minutes each)
- Title: Extracting knowledge from large-scale communication data (Shaojing Sun)

With the rapid growth of communication data and scholarships, it is promising to extract knowledge from them and offer guidance for future scholarships. Researchers have relied on conventional methods such as content analysis and meta-analysis to extract knowledge from existing data. We argue that developing data analytics renders new ways to scrutinize data and generate new scholarship. We discuss the commonalities and differences between conventional and cutting-edge methods for extracting knowledge. Furthermore, drawing cases from communication research, we demonstrate how data science can inform the innovation of communication theories and methodologies.

Shaojing Sun is a professor of the School of Journalism at Fudan University. He has published about fifty peer-reviewed journal articles on communication, psychology, public health, among others. His research employs a variety of quantitative methods including but not limited to survey, experiments, meta-analysis, text mining, etc.

- Title: Automatic text mining of media data (Giuseppe Veltri)

Abstract: compared to traditional content analysis, text mining has its unique advantage in analyzing large-scale media data. Automatic text analysis is a powerful tool, though, with its limitations. Awareness of these shortcomings is necessary in the process of interpretation, and particular attention should be paid to the distinction between implicit and explicit meanings of the texts. Semantic analysis refers to how meaning units denote one another within a system of relations (Morris, 1955). Therefore, any semantic analysis focuses on the identification of relationships between concepts in terms of their explicit meanings and does not go beyond the explicit meaning of a text (Krippendorff, 2003). We explicate the pros and cons of employ automatic text mining to analyze media data and address communication issues.

Bio: Giuseppe is a professor of communication at the University of Leicester, UK. He has done extensive work on text mining and social network analysis of communication data. He has held workshops annually on analyzing communication data at LSE, University of East Anglia, and other universities. His work has been published in journals including Social Network Analysis, Nature Materials, Big Data & society, among others.

- Title: Diffusion of Innovations: A revisit from the information science perspective (Ying Ding)

Abstract: Innovations refer to new ideas that have value and impact. Innovations are connected with creative, breakthrough, and original. They are the essential driven forces to advance science, technology, health, and society. How to diffuse innovations is critical to enable massive and sustainable adoption of innovations, and improvement

through re-invention. Diffusion of Innovations is the classic masterpiece in the field of communication in 1960s. It is time to revisit it as its four key elements have been completely changed because of big data and social media. This talk will showcase several ideas on how to enable data-driven analysis to better understand diffusion of innovations.

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.

- Title: Investigating communication effects with big data (Hongzhong Zhang)

Communication effects are to large extent related to structure and nature of messages. The data-driven approach has offered a new way to investigate communication effects, on top of conventional psychological experiments or survey studies. Moreover, data-driven approach can help identify meaningful communication issues and help customize communication messages before proceeding to experimental studies. Here, we look at accessible communication data, and discuss how a data-driven approach may inform communication theories building as well as testing.

Bio: Hongzhong Zhang is professor and associate dean of the School of Journalism and Communication at Beijing Normal University. He has published work on public opinion, media effects, and journalism by using various empirical research methods. In recent years, he has spearheaded a series of workshops on big data and communication research in China. His work has facilitated the collaboration of researchers and professionals in communication and big data.

- Title: Big data and data-driven research: reflections from China's most influential researchers (Li Tang)
- Title: Understanding young people's we-intention to contribute in Danmaku websites: Motivational, social, and subculture influence (Yuxiang Zhao)
- Title: Measuring the diffusion of an Innovation (Yujia Zhai)

Abstract: Innovations transform our research traditions and become the driven force to advance individual, group, and social creativity. Meantime, interdisciplinary research is increasingly being promoted as a route to advance the complex challenges that we face as a society. Using topic modeling and knowledge extraction methods, we will discuss the sequence and progress of the growth of innovation, explore the patterns of applying or improving the innovative ideas and reveal the content of knowledge flows between different disciplines.

Bio: Yujia Zhai is a Ph.D. student in information Science at Nankai University. His research interest is in understanding innovation diffusion and using machine learning/text mining and natural language processing (NLP) methods for solving social problems and assisting decision making process. His former research projects involved

extracting knowledge entities and emerging topics from unstructured and structured dataset ranging from scientific articles to social networks, news articles, social media and general web pages.

### 3 Relevance to the Conference/Significance to the Field

Communication research, by its nature, is closely related to information science and data science. Focusing on human-generated data in its various formats (e.g., publications, media contents, social media posts), communication studies are aimed to reveal the linkage between information, communication, individuals, organizations, and society as a whole. this session can help advance the dialogue between researchers from communication, informatics, and information science. A fundamental driving force is the rapid progress of digital communication. this session will draw researchers from related disciplines together to tackle interdisciplinary or transdisciplinary questions pertaining to both substantive domains and analytical methodologies.

### 4 Indicate the length of your event

This SIE will span 180 minutes.

### 5 References

Sudhahar, S., Veltri, G. A., & Cristianini, N. (2015). Automated analysis of the US presidential elections using Big Data and network analysis. *Big Data & Society*, 2, 1-28.