Dynamic Evolution of Data Journalism Research in China over the Past Decade: A CiteSpace-Based Data Mining Analysis (Postprint)

Authors: Hu Yanyan

Date: 2023-10-08T00:00:00+00:00

Abstract

Objective: Data journalism, as a cross-disciplinary and multi-disciplinary news reporting paradigm, has garnered widespread attention in recent years. Therefore, it is essential to systematically examine its dynamic evolutionary trajectory and development trends, thereby providing a foundational basis for exploring cutting-edge issues in China' s data journalism field.

Method: This study examines data journalism-related literature indexed in the China Academic Journals Network Publishing Database (CNKI), employing the bibliometric software CiteSpace to generate knowledge maps for visual analysis.

Results: The analysis aims to excavate and examine the developmental patterns, research themes, and evolution of hotspots in domestic data journalism research over the past decade since 2013.

Conclusion: The findings reveal that following a short-term explosive growth period from 2013 to 2016, research output entered a declining phase beginning in 2017, with scholarship gradually becoming more rational. The knowledge maps demonstrate that domestic research hotspots over the past decade have primarily concentrated on the conceptualization of data journalism, production workflows, value impacts, and talent cultivation, gradually shifting from early-stage investigations of fundamental concepts and visualization practices toward research on the narrative turn in data journalism, talent development, and the application of emerging technologies.

Full Text

Preamble

Dynamic Evolution of China' s Data Journalism Research Field in the Past Decade: A Data Mining Analysis Based on CiteSpace

(Xinhua Newspaper Media Group, Nanjing, Jiangsu 210005)

Abstract

[Objective] As an interdisciplinary and multi-domain news reporting approach, data journalism has attracted widespread attention in recent years. It is therefore necessary to systematically examine its dynamic evolution path and development trends to provide a foundation for exploring frontier issues in China' s data journalism field. **[Method]** This study examines data journalism-related literature indexed in the China Academic Journals Network Publishing Database (CNKI), employing the bibliometric software CiteSpace to generate knowledge maps for visual analysis. **[Results]** The analysis aims to excavate and examine the development trends, research themes, and hotspot changes in domestic data journalism research over the past decade since 2013. **Conclusion** The study reveals that after a short-term explosive growth phase from 2013 to 2016, related research entered a cooling-off period beginning in 2017, with scholarship gradually becoming more rational. The knowledge maps indicate that domestic research hotspots over the past decade have primarily focused on the concept of data journalism, production processes, value and impact, and talent cultivation. Moreover, the research has gradually shifted from early studies on basic concepts and visualization practices to investigations of narrative regression, talent development, and the application of new technologies in data journalism.

Keywords: data journalism; knowledge mapping; CiteSpace; narrative regression; talent cultivation
Classification: G240
Document Code: A
Article ID: 1671-0134(2023)04-053-05
DOI: 10.19483/j.cnki.11-4653/n.2023.04.011
Citation Format: Hu Yanyan. Dynamic Evolution of China' s Data Journalism Research Field in the Past Decade: A Data Mining Analysis Based on CiteSpace [J]. China Media Technology, 2023(04): 53-57.

Introduction

Data journalism has emerged against the backdrop of rapid technological developments in cloud computing, big data, artificial intelligence, and mobile internet. As a novel interdisciplinary and multi-domain news reporting approach, it has attracted extensive attention since its inception. In 2009, *The Guardian* became the first media organization worldwide to establish a digital news department. The inaugural "International Data Journalism" roundtable conference held in Amsterdam in 2010 provided a detailed elaboration of data journalism from a production process perspective. The 48-hour workshop at the MozFest conference in London in 2011 brought together numerous data journalism advocates and experts to compile the first data journalism production guide, *The Data Journalism Handbook*. The Global Data Journalism Awards were established in

2012, and the Chinese translation of *The Data Journalism Handbook* was published in 2013. In March 2013, Xinhua News Agency established its data news department and launched a dedicated data news column, becoming one of the earliest domestic media outlets to practice data journalism. In October 2013, Caixin Media established the Caixin Data Visualization Laboratory to create its "Digital Talk" channel. The first China Data Journalism Competition was held in 2015. Additionally, initiatives such as *The Paper*' s "Meishu Shuo" and NetEase' s "Data Reading" column have continued to explore and innovate in content production, narrative models, and knowledge generation over the past decade.

Given the current state of widespread attention to data journalism in both academia and industry and its significant contemporary relevance, this study adopts the internationally recognized scientific research evaluation method of bibliometrics. Using literature from CNKI's core journals in the data journalism field as the research sample and employing the information visualization software CiteSpace, this paper excavates, analyzes, organizes, and maps knowledge graphs to gain in-depth understanding of the research status, dynamic evolution paths, and development trends in China' s data journalism field, thereby providing a foundation for exploring frontier issues.

1. Research Design

1.1 Research Methods

CiteSpace is a visualization software developed by Professor Chen Chaomei to assist in bibliometric analysis [1]. It can display important information changes in the development process of a given research field through visualization, presenting key literature, research hotspots, and development trends. This study employs CiteSpace' s basic functions, including word frequency co-occurrence analysis, cluster analysis, and research frontier detection, to reveal research hotspots and development trends in the data journalism field, aiming to achieve a comprehensive understanding of the dynamic evolution of China' s data journalism research over the past decade.

1.2 Data Sources

The data for this study were sourced from CNKI using advanced search. The search terms were "data journalism" or "data-driven journalism" in the subject field, with publication years limited to 2013-2022. Journal sources were restricted to "Peking University Core," "CSSCI," and "CSCD" categories, with the discipline set to "Journalism and Media." After screening out invalid literature, 751 journal articles were obtained. The bibliographic data were exported, converted, saved, and then subjected to bibliometric analysis.

2.1 Analysis of Publication Volume

Examining the annual publication volume of data journalism research in CNKI core journals over the past decade, we generated the publication volume chart shown in [Figure 1: see original paper]. As illustrated in the figure, the number of published papers on "data journalism" research entered a rapid growth phase beginning in 2013. It was precisely in 2013 that the Chinese translation of *The Data Journalism Handbook*—originating from the 48-hour workshop at the 2011 London Mozilla Festival—was introduced to China, marking the popularization of data journalism in the country. From 2013 to 2016, annual publications surged from 25 to a peak of 149, demonstrating a short-term explosive growth trend. During this stage, domestic scholars' research enthusiasm for data journalism grew increasingly intense, research content deepened continuously, and involved themes expanded progressively. However, starting in 2017, domestic data journalism research entered a cooling-off period, with core journal publications declining from 149 in 2016 to 131 in 2017 and continuing to decrease. This indicates that research has gradually become more rational.

2.2 Analysis of Publishing Authors

Analyzing authors' publication volumes can reflect the distribution of academic resources in a given field. Based on CiteSpace statistical analysis results, we mapped major data journalism authors and their publication volumes from 2013 to 2022 (). The study found that high-volume authors are primarily scholars from journalism and communication schools at major universities.

Regarding author collaboration, this study used CiteSpace to generate a knowledge map of data journalism authors from 2013 to 2022 ([Figure 2: see original paper]), where connecting lines between nodes represent collaborative relationships and larger nodes indicate higher publication volumes. Analysis of [Figure 2: see original paper] reveals that author collaboration networks in domestic data journalism research are relatively loose, with scholars tending to produce work independently. The field has not yet formed significant group effects. However, this more open and free research model facilitates the accumulation of diverse research outcomes and enables multi-dimensional innovation.

3. Analysis of Research Hotspots and Frontiers in China' s Data Journalism

3.1 Keyword Co-occurrence Analysis

Keywords provide a high-level summary of a paper' s main content and core themes. When a keyword appears frequently in literature within a research field, it often reflects hotspot research topics. A co-occurrence analysis map uses circles labeled with keywords as nodes, connecting lines between nodes indicate co-occurrence relationships, and circle radius represents keyword frequency [2]. After importing the 751 bibliographic records into CiteSpace, we obtained the

keyword co-occurrence map shown in [Figure 3: see original paper]. The map comprises 402 keyword nodes and 475 connecting lines. Further examination of [Figure 4: see original paper] reveals that data journalism research hotspots over the past decade have concentrated on data journalism concepts, big data, visualization, media convergence, news production and communication, data literacy, and talent cultivation.

3.2 Research Knowledge Clustering

To further identify relationships among keywords and summarize research themes in the data journalism field, this study conducted cluster analysis on closely related keywords from the literature. The resulting clusters represent current hotspot research areas. Cluster credibility is evaluated using Modularity Q and Mean Silhouette values. The cluster analysis results are shown in [Figure 4: see original paper], with a modularity of 0.8609 and a mean silhouette coefficient of 0.9727, indicating good clustering effects and a reasonable cluster map. A total of 17 automatic clusters were identified in domestic data journalism research from 2013 to 2022. To clarify the knowledge system and development 脉络 of data journalism research, this paper combines map analysis results with close literature reading to categorize the 17 clusters into four major research hotspots.

The first category is research on basic concepts and innovative development of data journalism, primarily including clusters "#0 data journalism" and "#3 big data." The year 2013 marked China' s big data 元年, and simultaneously, the Chinese translation of *The Data Journalism Handbook* was officially introduced, expanding the fundamental concepts of journalism in China. Some domestic scholars summarized and proposed various characteristic elements of data journalism while reviewing international research. Fang Jie and Yan Dong' s 2013 paper discussed the connotation and characteristics of data journalism and clarified related concepts [3]. Ren Ruijuan proposed that data journalism elements include data, hyperlinks, interactivity, contextualization, visualization, and semantics [4].

The second category is research on data journalism production processes, mainly including clusters "#2 visualization," "#6 *The Guardian*," "#7 precision journalism," "#8 traditional media," "#9 news narrative," "#10 data science," "#11 media data," "#12 data mining," "#15 news production," and "#16 openness." Related papers focus on data sources, data mining and analysis, narrative models, data presentation, and the application of new technologies in data journalism. First, massive data sources are the prerequisite for data journalism, enhancing the accuracy and persuasiveness of news reports by connecting multiple information streams through data. Liu Yinghua et al. argue that open data is a type of data that can be freely used, reused, and redistributed by anyone [5]. Bi Qiuling notes that open data is primarily provided by government agencies, universities and research institutions, and non-governmental organizations, characterized by openness, accessibility, completeness, and timeliness

[6]. Second, technology drives data forward, with data modeling, data mining, NLP semantic analysis, knowledge graphs, and recommendation algorithms being key technologies applied in data journalism [7]. These new technologies also demonstrate development trends in data journalism. Finally, data journalism innovates news narrative models through visualization pathways. Xu Xiangdong argues that in data journalism reporting, the "narrator" is partially materialized as "data," and narrative structures evolve from two-dimensional to three-dimensional forms. Visualization presentation collates, organizes, and summarizes cumbersome and massive data information, forming a relatively smooth visual narrative chain that enhances explanatory functions while strengthening user interactive experiences [8].

The third category is research on the value and impact of data journalism, primarily including clusters "#4 news communication" and "#5 news value." In the era of intelligent media, data journalism serves as a vibrant and innovative news format that will become a valuable supplement to traditional news genres such as messages, features, and in-depth reports. Ren Ruijuan mentions in *Prediction and Discovery: Theory and Practice of Data Journalism* that a crucial value of data journalism in the big data era is its predictive and discovery functions [9]. Through structured and knowledge-based data processing, meaningful data connections are explored, and based on real-world data analysis, the present and past are examined to make probabilistic predictions about future developments, truly achieving a journalistic transformation from "post-event summary" to "pre-event planning." Data journalism provides new opportunities for mainstream media and journalists to reconstruct their identities as professional content producers [10].

The fourth category is research on talent cultivation and data literacy in data journalism, mainly represented by clusters "#1 data literacy" and "#13 technology empowerment." Related papers focus on data journalism talent training models and data literacy among journalism professionals. Xu Xiangdong' s article "A Comparison and Reflection on Data Journalism Talent Training Models in China and the United States"explores suitable data journalism talent training models for China' s journalism education and media industry practices by analyzing models and characteristics from both countries [11]. Fang Jie and Gao Lu' s study "Data Journalism: A Field Urgently Needing Professional Standards—A Quantitative Study Based on Five Domestic Data Journalism Columns" found through content analysis of five domestic data journalism columns over one year that the domestic data journalism field urgently needs to establish professional standards and that practitioners' data literacy requires significant improvement [12].

3.4 Research Frontier Analysis

Burst keywords are those whose frequency suddenly shows significant growth within a certain period. Through burst term analysis, we can understand research frontiers, shifts in research focus, and the latest hotspot dynamics, help-

ing to predict subsequent development trends and explore new potential frontier research questions. Using CiteSpace's "Burstness" function, this study generated a burst keyword map for China' s data journalism research from 2013 to 2022, shown in [Figure 6: see original paper], where "Year" represents the keyword' s emergence time, and "Begin" and "End" represent the start and end times of its hotspot status.

The earliest burst keywords include "big data," "social media," "information graphics," "news editing," and "database," with related research focusing on data journalism concepts, impacts, and associated big data and visualization practices. In recent years, particularly since 2018, new burst terms such as "artificial intelligence," "narrative," "data science," and "technology empowerment" have emerged, indicating that research on the application and development of artificial intelligence and data science in data journalism, studies on news narrative characteristics, and talent cultivation issues have increasingly attracted academic attention.

Conclusion

Based on CiteSpace visualization software and data journalism-related literature from CNKI, this study summarizes the development trends, research themes, and hotspot changes in domestic data journalism research since 2013, reaching the following conclusions.

From 2013 to 2016, data journalism research experienced explosive growth, with domestic scholars' research enthusiasm increasing daily. Beginning in 2017, core journal publications declined, and domestic data journalism research entered a cooling-off period, gradually becoming more rational.

High-volume authors in domestic data journalism research are primarily scholars from journalism and communication schools at major universities, with core scholars represented by Yu Guoming, Fang Jie, Chen Changfeng, and Xu Xiangdong. Author collaboration networks remain relatively loose, with scholars preferring independent work and no significant group effects yet formed. However, this open research model facilitates the accumulation of diverse and innovative research outcomes.

Based on automatic clustering results and close literature reading, this paper categorizes domestic data journalism research hotspots over the past decade into four areas: basic concepts and innovative development, production processes, value and impact, and talent cultivation/data literacy.

Combining keyword timeline views and burst term analysis, research on the application of new technologies such as artificial intelligence and data science in data journalism, studies on narrative characteristics, and talent cultivation have increasingly attracted academic attention. Current research hotspots are shifting from early-stage basic concepts and big data visualization practices to new technology development, narrative regression, and talent cultivation.

References

[1] Chen Yue, Chen Chaomei, Liu Zeyuan, et al. The Methodological Functions of CiteSpace Mapping Knowledge Domains [J]. Studies in Science of Science, 2015(2): 242-253.
[2] Ren Yanli, Wang Tong. Knowledge Mapping Analysis of Domestic Information Literacy Education Research: Based on CSSCI Papers (2008-2018) [J]. Heilongjiang Researches on Higher Education, 2018(9):
[3] Fang Jie, Yan Dong. "Data Journalism" from a Global Perspective: Concepts and Practices [J]. International Press, 2013(6): 73-83.
[4] Ren Ruijuan. A Feasibility Study on Constructing a Constructive News Reporting Framework Based on Data Journalism [J]. Contemporary Communication, 2020(4): 45-50.
[5] Liu Yinghua. Development Trends of Open Data and Its Application in Data Journalism [J]. Youth Journalist, 2019(15): 82-83.
[6] Bi Qiuling. Application of Open Data in Data Journalism [J]. Hubei Social Sciences, 2016(7): 190-194.
[7] Wang Lin. Data Journalism and Data Value Protection [J]. China Media Technology, 2022(11): 53-54.
[8] Xu Xiangdong. Turn, Deconstruction, and Reconstruction: A Study on Data Journalism Visualization Narrative [J]. International Press, 2019(11): 142-155.
[9] Ren Ruijuan. Prediction and Discovery: Theory and Practice of Data Journalism [M]. Beijing: Science Press, 2019: 78-79.
[10] Shi Lifeng. Analysis of Visualized Data Journalism in Domestic Online Media [J]. Media, 2015(20): 48-49.
[11] Xu Xiangdong. A Comparison and Reflection on Data Journalism Talent Training Models in China and the United States [J]. International Press, 2016(10): 100-109.
[12] Fang Jie, Gao Lu. Data Journalism: A Field Urgently Needing Professional Standards—A Quantitative Study Based on Five Domestic Data Journalism Columns [J]. International Press, 2015(12): 105-124.

Author Bio: Hu Yanyan (1986-), female, from Suzhou, Anhui, is an assistant engineer whose research focuses on data journalism.
(Responsible Editor: Zhang Xiaojing)

Note: Figure translations are in progress. See original paper for figures.