Postprint: Analysis of Diffusion Characteristics and Themes in China's Government Big Data Policies

Authors: Duan Yaoqing, Shang Ting, meticulous

Date: 2023-04-01T16:15:56+00:00

Abstract

[Purpose/Significance] This study investigates the patterns and mechanisms of government big data policy diffusion, providing theoretical support for policy formulation and enhancing implementation effectiveness.

[Method/Process] Employing social network analysis and thematic analysis, this study examines 213 government big data-related policies, constructs a reference relationship network among policies, and analyzes the temporal, spatial, and thematic characteristics of policy diffusion.

[Results/Conclusion] In the temporal dimension, China's government big data policy diffusion exhibits typical S-curve diffusion characteristics; in the spatial dimension, it demonstrates geographically uneven distribution and hierarchical gradient features; in the thematic dimension, it shows characteristics of integrating inheritance and innovation, where the core value orientation of referenced policies is inherited during the diffusion process.

Full Text

Preamble

Characteristics and Thematic Analysis of Government Big Data Policy Diffusion in China

Duan Yaoqing[1],[2], Shang Ting[1],[2], Zhou Mi[1],[2]

[1]School of Information Management, Central China Normal University, Wuhan 430079
[2]Hubei Research Center of Data Governance and Intelligent Decision-making, Wuhan 430079

Abstract: [Purpose/Significance] This study explores the patterns and mechanisms of government big data policy diffusion to provide theoretical support for policy formulation and enhance implementation effectiveness. [Method/Process] Using social network analysis and thematic analysis, we analyzed 213 government big data-related policies, constructed a reference relationship network among policies, and examined the temporal, spatial, and thematic characteristics of policy diffusion based on this network. [Results/Conclusions] Temporally, China's government big data policy diffusion exhibits a typical S-curve pattern. Spatially, it shows geographically unbalanced distribution characteristics and hierarchical shading features. Thematically, it demonstrates a fusion of inheritance and innovation, with the core value orientation of referenced policies being preserved during the diffusion process.

Keywords: policy diffusion; government big data; social network analysis; thematic analysis
Classification Number: G250.2
DOI: 10.13266/j.issn.0252-3116.2020.13.017

Government big data refers to massive datasets generated or acquired through collection, processing, and transformation by administrative organs at all levels, as well as public institutions and social organizations authorized by laws and regulations to perform administrative functions in the process of fulfilling their duties. Compared with other types of big data, government big data has relatively higher value density and greater development potential, making it a strategic priority for national development. Governments at all levels have actively introduced policies related to government big data, with increasing adoption and mutual referencing among policies. These policies not only promote the development of government big data but also provide guidance, standardization, and safeguards for its implementation. Policy diffusion theory offers an explanatory framework for the inter-regional spread of policies, and studying the diffusion of government big data policies helps to accurately grasp the underlying mechanisms, thereby improving both policy formulation and implementation efficiency.

Policy diffusion theory generally derives from sociology, information science, and communication studies, typically referring to the process by which a policy program spreads to another region and is adopted and referenced by the local government [**?**]. Scholars both domestically and internationally have conducted in-depth research on policy diffusion, covering concepts, characteristics, influencing factors, outcome variables, and diffusion patterns and paths [**?**], yielding numerous results in fields such as education policy, social policy, sports industry policy, and informatization policy. Studies on diffusion characteristics and patterns represent a classic research direction in policy diffusion. Policy diffusion exhibits distinct features across multiple dimensions including time, space, and actors, and analyzing these processes reveals diffusion patterns.

Wang Puqu et al., combining Chinese public policy diffusion practice, classified policy diffusion patterns into four types [**?**]: top-down hierarchical

diffusion, bottom-up policy adoption and promotion, inter-regional and inter-departmental diffusion, and diffusion from policy-advanced regions to policy-follower regions. Ding Wenyao et al., based on policy diffusion and policy transfer theories, studied the diffusion characteristics and patterns of local big data policies in China from temporal, spatial, and content dimensions, finding that the temporal diffusion curve of big data policies follows an S-shape and conforms to the top-down hierarchical diffusion model [**?**]. Liu Hongbo and Lin Bin also found that the diffusion of China's artificial intelligence policies is primarily directive and top-down hierarchical [**?**]. Foreign scholars have conducted extensive research in this area. V. Gray proposed the national interaction model, assuming a nationwide communication network among local governments that enables free interaction between policies [**?**]. M. H. Go proposed the regional diffusion model, suggesting that government policy formulation may be influenced by policies in neighboring administrative regions [**?**]. Walker proposed the leader-follower model, arguing that certain local governments are leaders in policy formulation while others learn and draw lessons from them rather than competing under pressure [**?**, **?**]. F. S. Berry's vertical influence model posits that local governments reference national policies rather than those of other local governments [**?**]. In practice, mixed diffusion patterns are common, with vertical and horizontal diffusion occurring simultaneously to ensure accurate and rapid policy spread within the social system [**?**]. Whether the characteristics and patterns of these policy diffusion models are similar to or innovative compared with China's government big data policy diffusion requires further analysis and discussion.

Research methods for policy diffusion have also diversified. Early research focused on measuring diffusion and adoption rates, representing an early quantitative approach. Pei Lei et al. first introduced policy text analysis into policy diffusion research, conducting thematic tracking of China's informatization policy diffusion from regional and temporal dimensions [**?**]. Zhang Jian et al. studied the diffusion of China's scientific and technological achievement transformation policies through bibliometric and keyword temporal analysis, examining the process and characteristics from four dimensions: intensity, breadth, speed, and direction [**?**]. Li Jian et al. used event history analysis to examine the diffusion mechanisms of provincial government purchasing services in China [**?**].

Although scholars have conducted in-depth research on policy diffusion, analysis of government big data policy diffusion remains relatively limited. The content diffusion relationships, diffusion extent, and whether and to what degree content innovation occurs in government big data policies remain unclear. Based on social network analysis and thematic analysis, this study analyzes China's government big data policies to uncover underlying patterns, identify characteristics in policy content formulation, and explore innovation points, thereby providing reference and inspiration for future research and a scientific basis for policy formulation.

2 Data Sources and Research Methods

Policy texts were sourced from the official portal of the Central People's Government of China ("China Government Network") and local government portals. To avoid omissions, we supplemented our search with the domestic legal information retrieval system "Beida Fabao." Using "government big data," "government affairs," and "big data" as keywords on these websites, with a cutoff date of December 31, 2018, we conducted manual screening to obtain a final sample of 213 valid policy documents. Since government big data development is closely related to technological and industry application developments, our dataset includes both specialized government big data policies and related policies.

3 Research Process

We analyzed the diffusion of 213 government big data policies from temporal, spatial, and thematic dimensions to reveal their main characteristics. First, we coded the policies and identified reference relationships through content analysis, followed by basic statistical analysis. We described the temporal diffusion characteristics by analyzing the cumulative distribution of policy reference diffusion. Second, based on content analysis and statistical results, we mapped the regional distribution of reference frequencies and the administrative-level reference network to reveal spatial and hierarchical diffusion characteristics. Finally, we conducted thematic analysis of the core diffusion sub-networks in the administrative-level reference network to examine content inheritance and innovation during diffusion.

3.1 Policy Text Coding and Reference Relationship Identification

Coding policy texts involves assigning classification identifiers. We coded the 213 collected policies using a three-tier system: central, provincial, and municipal/county. Central policies issued by the Party Central Committee, State Council, and other central party and government departments were coded as A (A1, A2, …, An). Provincial, autonomous region, and municipal government policies were coded as B (B1, B2, …, Bn). Municipal and county government policies were coded as C (C1, C2, …, Cn).

We analyzed policy texts to identify reference relationships. When a reference relationship exists, the referencing policy typically lists the referenced policy title in book quotation marks. Reference relationships 主要包括 central agencies referencing other central agency policies, local governments referencing central agency policies, municipal/county agencies referencing provincial policies, and horizontal referencing among local governments at the same level. A reference network with policies as nodes can visually demonstrate diffusion across different levels. We then introduced social network analysis and visualization techniques to construct the government big data policy diffusion reference network for analyzing hierarchical diffusion.

From the 213 policy documents, we extracted 187 pairs of policy reference relationships. For example, the Ministry of Industry and Information Technology's "Big Data Industry Development Plan (2016-2020)" states: "To thoroughly implement the spirit of the Fifth Plenary Session of the 18th CPC Central Committee, implement the national big data strategy, and fulfill the State Council's 'Outline for Promoting Big Data Development' in compiling this plan…" This constitutes a reference relationship, with the "Big Data Industry Development Plan (2016-2020)" as the referencing policy and the "Outline for Promoting Big Data Development" as the referenced policy.

3.2 Temporal Characteristics of Policy Reference Diffusion

Based on the extracted policy reference relationships, we conducted a time-series statistical analysis of reference frequencies, with results shown in Table 1 . The data reveal high reference frequencies between 2015 and 2017, indicating rapid diffusion of government big data policies during this period.

To further explore the changing trend of policy diffusion cumulative volume, we imported the cumulative reference data by year into MATLAB for curve fitting, using 2013 as the starting point. The scatter plot of data in Table 1 exhibited S-curve characteristics, so we adopted a typical S-curve model. The resulting functional relationship between cumulative diffusion volume and diffusion time (Formula 1) achieved a coefficient of determination R^2 of 0.9776, indicating good fit. The fitted curve is shown in Figure 1 [Figure 1: see original paper].

$$0.0041 + 0.15e^{-t} \quad \text{(Formula 1)}$$

Where t (in years) represents diffusion time, and Y represents the cumulative reference frequency at time t. Analysis of Figure 1 reveals that government big data policy diffusion is currently in the early stage of slowing down. The temporal characteristics of government big data policy diffusion are not significantly different from those of general public policies, though the overall diffusion period is relatively short, closely related to the high priority placed on government big data development at the national level.

3.3 Spatial Characteristics of Policy Reference Diffusion

We statistically analyzed reference frequencies by province to understand the overall diffusion status across regions, with visualization results shown in Figure 2 [Figure 2: see original paper]. Guizhou, Shanxi, and Shanghai exhibit high reference frequencies and deep diffusion. Guizhou holds significant advantages in government big data development, leading nationally. In contrast, Northwest and Northeast China show low reference frequencies and shallow diffusion, requiring strengthened guidance and active learning from better-performing regions.

To further understand the diffusion process across administrative levels, we used nodes to represent policy documents, with node size indicating reference frequency and arrows showing reference relationships (diffusion direction) from referenced to referencing policies. Using Gephi's ForceAtlas algorithm for network layout, we visualized the reference network shown in Figure 3 [Figure 3: see original paper]. The largest node is central policy A3, the "Outline for Promoting Big Data Development," indicating the highest reference frequency. Central policies carry substantial weight in the network, and analysis of the A3-centered sub-network reveals high reference frequencies from both other central policies and local policies at all levels, demonstrating horizontal diffusion within the same administrative level and "central-to-local" diffusion patterns.

Government big data policy diffusion follows administrative-level patterns. This vertical diffusion includes not only sequential diffusion by administrative level but also cross-level diffusion from central to municipal/county governments. Horizontal diffusion phenomena also exist among municipal governments and within central and provincial policies. Sequential and cross-level diffusion constitute the main forms, with administrative directives playing a strong facilitating role—after higher-level policies are issued, lower-level governments actively learn and formulate local policies based on the spirit of higher-level policies. Small-scale horizontal diffusion represents learning and borrowing among governments, though this characteristic is not prominent in the overall diffusion process and warrants further strengthening and guidance.

3.4 Thematic Diffusion in Core Reference Networks

To further understand thematic changes in government big data policy diffusion, we analyzed the diffusion status of core nodes in the reference network. We identified 10 core diffusion relationships, with source policies A3, A14, A18, A19, A31, A32, A33, A35, B63, B157, and diffusion policy sets coded as a3, a14, a18, a19, a31, a32, a33, a35, b63, b157. We used the "Natural Language Processing and Information Retrieval Sharing Platform" (NLPIR) developed by the Chinese Academy of Sciences to extract keywords from each dataset, with results shown in Table 3 .

We selected three commonly used indicators for thematic tracking analysis: inheritance ratio, diffusion ratio, and innovation ratio [**?**]. If a referenced policy has N themes and a referencing policy has M themes, with n shared themes, we define: - **Inheritance ratio (Hscope)**: proportion of themes inherited from referenced policy = n/N - **Diffusion ratio (Dscope)**: proportion of themes diffused in referencing policy = n/M - **Innovation ratio (Iscope)**: proportion of innovative themes in referencing policy = 1 - Dscope

These indicators and calculation methods are detailed in Table 4 . Calculation results are shown in Table 5 .

Analysis of Table 5 reveals that, compared with high diffusion and innovation ratios, several diffusion sub-networks exhibit low diffusion ratios but high inno-

vation ratios. This indicates that local policies have not fully comprehended and implemented central policy spirit, and the balance between local adaptation and adherence to central guidance needs optimization. Additionally, the inheritance ratio for provincial/municipal policy diffusion is lower than that for central-to-local diffusion, suggesting that top-down directive diffusion is relatively more effective than horizontal policy learning.

(1) Thematic Inheritance Analysis. Thematic inheritance reflects the shared core value orientation between referenced and referencing policies. Analysis of inherited keywords in Table 3 reveals six inherited themes: 1) **Promoting government big data openness and sharing**, reflected in keywords such as "data," "open," and "share," aiming to promote inter-departmental data sharing and public data resource openness; 2) **Emphasizing government big data platform construction**, reflected in "platform," "system," and "architecture," focusing on integrating various government information platforms; 3) **Emphasizing government big data application**, reflected in "application," "market," and "service," promoting big data in healthcare, social security, education, and transportation; 4) **Improving government big data quality**, reflected in "quality," "catalog," and "standard," establishing standard systems; 5) **Ensuring government big data security**, reflected in "security" and "safety," strengthening safeguard systems; and 6) **Promoting big data technology development**, reflected in "infrastructure" and "technology," advancing core technology research. These six themes cover crucial aspects of government big data development, with data services and applications as goals, and platform construction, standards, quality, and security as foundations.

(2) Thematic Innovation Analysis. Thematic innovation reflects optimizations and improvements made by referencing policies during implementation, demonstrating local policy formulation characteristics. Analysis of innovative keywords in Table 3 reveals six innovative themes: 1) **Actively cultivating big data industry demonstration zones**, reflected in "industrial park," "leading enterprise," and "pilot," establishing national demonstration zones; 2) **Government data asset rights confirmation**, reflected in "data asset," promoting asset-based management; 3) **Government big data marketization**, reflected in "marketization," "service," and "financial products," enabling joint development and utilization; 4) **Government big data talent cultivation**, reflected in "talent" and "personnel training," strengthening professional development; 5) **Government big data technology integration**, reflected in "cloud computing" and "innovation," deepening integration with IoT and AI; and 6) **Government big data application innovation**, reflected in "smart community" and "decision-making," improving intelligent application levels. These six innovative themes involve more detailed content related to policy implementation, reflecting local characteristics. Data asset management and marketization are pathways for value-added services, while talent provides intellectual support. Industry demonstration zones summarize experiences and drive comprehensive development.

4 Conclusions and Outlook

Based on social network analysis and thematic analysis of 213 government big data policies, this study examined diffusion across temporal, spatial, hierarchical, and thematic dimensions.

Temporally, China's government big data policy diffusion exhibits a typical S-curve pattern: slow initial diffusion, rapid mid-term diffusion, and gradually slowing later diffusion. The process is currently entering the later slowing stage. The temporal characteristics are similar to general public policies, though the overall diffusion period is relatively short, reflecting high national priority.

Spatially, geographic distribution is unbalanced, with diffusion depth showing agglomeration effects and neighboring regions influencing each other. Guizhou, Inner Mongolia, Shanxi, and Shanghai show deep diffusion, while Northwest and Northeast China show shallow diffusion. Hierarchically, diffusion demonstrates shading characteristics, including not only sequential administrative-level diffusion but also cross-level diffusion from central to municipal/county governments. Small-scale "municipal-municipal" and intra-provincial diffusion reflects learning effects among neighboring regions and same-level governments.

Thematically, government big data policy diffusion combines inheritance and innovation. Inherited themes are macro-level, covering openness/sharing, platform construction, application, quality, security, and technology. Innovative themes are more detailed, covering industrial parks, data assets, marketization, talent cultivation, technology integration, and application innovation. Higher-level policies emphasize overall framework design and macro guidance, while local policies balance adherence to central spirit with local adaptation.

These conclusions offer guidance for policy formulation, implementation, and forecasting. Based on our findings, we propose: 1) strengthening inter-regional policy learning and exchange to address uneven distribution and limited horizontal diffusion; 2) actively promoting policy competition and incentive mechanisms alongside administrative directives to foster multi-dimensional diffusion; and 3) strengthening governance and support systems, including infrastructure, data quality, security, and processing technology for governance, and funding, data rights, talent, and management for support.

This study describes the spatiotemporal and thematic diffusion characteristics of China's government big data policies but has limitations. First, policy theme identification could be enhanced, potentially through LDA topic models. Second, thematic diffusion analysis could be optimized, possibly using word vector matching instead of thematic indicators. Finally, policy analysis could consider additional dimensions such as institutional distribution, policy structure, and policy type impacts on diffusion.

References:

[?] Ma Haiqun, Zhang Bin. Analysis of the diffusion of safety intelligence laws, regulations, and policies in China [?]. Journal of Intelligence, 2019, 38(7): 28-34.
[?] Zhou Yingnan, Huang Sai, Song Xiaoman. Policy diffusion research review and future prospects [?]. East China Economic Management, 2019, 33(5): 150-157.
[?] Wang Puqu, Lai Xianjin. Analysis of patterns and mechanisms of public policy diffusion in China [?]. Peking University Journal (Philosophy and Social Sciences Edition), 2013, 50(6): 14-23.
[?] Ding Wenyao, Zhang Zili, Yu Guoxian, et al. Diffusion patterns and transfer characteristics of local big data policies in China [?]. Big Data, 2019, 5(3): 76-95.
[?] Liu Hongbo, Lin Bin. Research on mechanisms and paths of artificial intelligence policy diffusion—A typological analysis perspective [?]. Chinese Public Administration, 2019(4): 38-45.
[?] Gray V. Innovation in the states: A diffusion study [?]. The American Political Science Review, 1969, 63(3): 880-899.
[?] Go M H. Building a safe state hybrid diffusion of building code adoption in American states [?]. American Review of Public Administration, 2016, 46(6): 713-733.
[?] Walker J L. The diffusion of innovations among the American states [?]. The American Political Science Review, 1973, 67(4): 1174-1185.
[?] Boushey G. Punctuated equilibrium theory and the diffusion of innovations [?]. The Policy Studies Journal, 2012, 40(1): 127-149.
[?] Berry F S. Sizing up state policy innovation research [?]. Policy Studies Journal, 1994, 22(3): 442-456.
[?] Pei Lei, Zhang Qiping, Li Xiangju, et al. Thematic tracking research in China's informatization policy diffusion [?]. Library and Information, 2016(6): 63-71.
[?] Zhang Jian, Huang Cui, Ye Xuanting, et al. A bibliometric study of public policy diffusion in China—Taking scientific and technological achievement transformation policy as an example [?]. China Soft Science, 2016(2): 145-155.
[?] Li Jian, Zhang Wenting. Research on the diffusion of government purchasing services—An event history analysis based on data from 31 provinces [?]. China Soft Science, 2019(5): 60-67.
[?] Aronson J. A pragmatic view of thematic analysis [?]. The Qualitative Report, 1995, 2(1): 1-3.
[?] Braun V, Clarke V. Using thematic analysis in psychology [?]. Qualitative Research in Psychology, 2006, 3(2): 77-101.
[?] Shi Qian, Pei Lei, Li Xiangju, et al. Interactive diffusion between informatization policy theory and practice—Taking Jiangsu and Zhejiang informatization policy samples as examples [?]. Journal of the China Society for Scientific and Technical Information, 2016, 35(10): 1081-1089.