



Bibliographic records reveal trends in Chinese history scholarship: Evidence from Library of Congress metadata

Kai Li, Wenyi Shang, and Yingying Han

DOI: <https://doi.org/10.47989/ir31iConf64266>

Abstract

Introduction. Books are key research outputs in the humanities, but often poorly indexed in scholarly databases. This paper presents a preliminary analysis of metadata in the Library of Congress (LC) to analyse scholarship on Chinese history.

Method. We extracted metadata of all 18,564 books on Chinese history from LC published in three major regions, Greater China, the U.S., and Europe, and mapped chronological information of each book onto historical periods.

Analysis. We conducted quantitative analysis on the information to understand the general publication trends of books on Chinese history and their historical period representation, both the overall patterns and the similarities and differences within the three regions.

Results. Two major surges appear in Greater China and U.S. publications, with Greater China leading since the 1980s and the U.S. surpassing Europe by the 1960s. Later periods attract more exclusive attention, while earlier periods are often embedded in multi-dynastic narratives in Western scholarship, though U.S. research has recently diversified.

Conclusion. This study pioneers the use of bibliographic records to analyse Chinese historiography, offering a complement to scholarship focused on canonical works and highlighting the value of library metadata for quantitative humanities research. We share our dataset and code at <https://anonymous.4open.science/r/Code-for-Analysis-of-Chinese-History-Collection-in-the-Library-of-Congress-Catalog6BC2>.

Introduction

There is a growing body of scholarship on collections as data that treats library, archive, and museum holdings as datasets and uses computational methods to identify patterns, trends, and insights (Padilla et al., 2019; Burrows & Verhoeven, 2014; Mordell, 2019). Books are key research outputs in the social sciences and humanities (Maleki, 2022; White & Zuccala, 2018), yet they are not comprehensively indexed in scholarly databases as suggested by empirical evidence (Delgado-Quirós & Ortega, 2025; Mongeon & Paul-Hus, 2016). As a result, there is a clear gap between the importance of books for understanding social sciences and humanities scholarship and the inability of quantitative researchers to systematically analyse them on a large scale using existing research infrastructure. With the increasing availability of large-scale library catalog data from major research libraries, however, it is now possible to revisit the history of humanities scholarship in ways analogous to how quantitative studies of science have illuminated the history of scientific scholarship (Li et al., 2025).

This short paper draws on the complete set of Library of Congress (LC) books as of 2019 (Library of Congress Cataloging Distribution Service, 2019), one of the largest research-oriented library collections in the world. By using this data source, we aim to offer a preliminary attempt to revisit the historiography on China after the Second World War (1945). Previous surveys of the field, such as Zhang et al. (2010) and Zhang et al. (2013), have relied almost entirely on a small body of canonical texts. By shifting the focus from this narrow foundation to tens of thousands of volumes, new possibilities emerge for extending and enriching scholarly discourse. This perspective parallels Franco Moretti's concept of *distant reading*, which advocates the use of quantitative methods to investigate *'the great unread'* alongside canonical literary works (Moretti, 2000, 2013). In this study, we apply that approach to nonfiction on Chinese history, tracing the broader trajectory of scholarship and mapping the contours of post-1945 historiography on China, thereby enabling an examination of its evolution on a scale unattainable through traditional approaches limited to a small corpus of canonical works.

Specifically, we focus on books published after 1945 in three regions that account for the overwhelming majority of LC holdings on Chinese history: the Greater China region ('*Greater China*' hereafter), which we define as including mainland China, Taiwan, Hong Kong, and Macau (Lo, 2016), the U.S., and Europe. From these, we constructed three separate corpora, one for each region, to enable comparative analysis. This preliminary study is guided by the following research questions (RQs). For each question, we examine not only the overall patterns but also the similarities and differences across the three corpora.

RQ1: What are the general publication trends in books on Chinese history published after 1945 in the LC collection, in terms of publication numbers, their proportion relative to the total collection, and language distribution?

RQ2: How does the LC collection represent different historical periods in books on Chinese history published after 1945, and how have these patterns changed over time?

Methods

In this short paper, we used the open-access version of the MARC Distribution Services (MDS) supplied by the LC published in 2019 (Library of Congress Cataloging Distribution Service, 2019). This dataset contains more than 11 million book records from the LC catalog. It has been used in previous research and has been proven to be useful for empirical research (Li et al., 2025).

We used the Library of Congress Subject Heading (LCSH) terms assigned by professional librarians in the bibliographic record to decide if a book is about Chinese history and the specific periods of Chinese history a book covers. In this research, we used the 651 field (Subject Added Entry - Geographic Name) to select our sample (Library of Congress, 2025). In this field, subfield a (\$a)

refers to a main geographical term to describe the aboutness of a resource and subfield x (\$x) represents any additional general topical term. Given this structure, we used the query 651\$a == 'China' & 651\$x == 'History,' as it can capture those books that are directly about Chinese history.

After excluding 1,118 works of fiction identified through the 33rd digit of the MARC 008 field, we obtained 20,786 unique books, representing the complete set of books on Chinese history held in the LC.

From this complete set, we selected books published after the Second World War (i.e., publication year 1946 or later). The filtered dataset consists of 19,819 books (95.35% of the total), with their annual distribution shown in Figure 1. We then identified books published in Greater China, the U.S., or Europe based on the publication-place information recorded in digits 15–17 of the MARC

008 field, which uses the controlled vocabulary of the MARC Code List for Countries (<https://www.loc.gov/marc/countries/>) to encode publication locations. Of the 19,819 post-1945 books, the vast majority (18,564 books, 93.67% of the total) are published in these three regions. These 18,564 books form the basis of our analysis and are divided into three regional corpora for our comparative analysis.

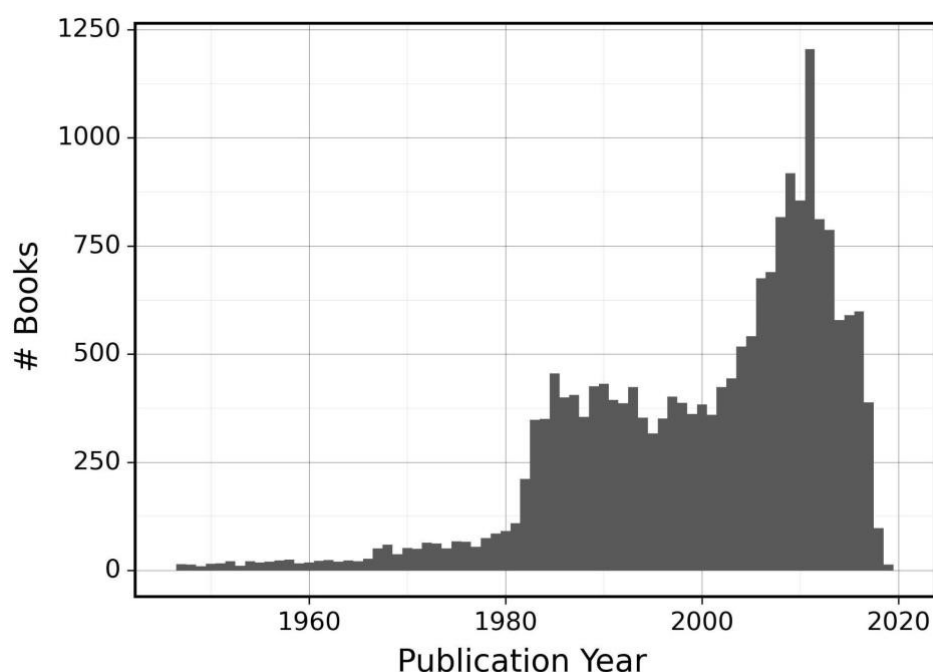


Figure 1. Distribution of books on Chinese history published from 1946 to 2019 (n = 19,819).

Finally, we extracted temporal subject information to examine the historical periods represented in the books. From the MARC 650 and 651 fields, we collected all chronological subdivisions (subfield z) related to Chinese history and manually mapped them onto dynasties and modern periods to assess how different periods of Chinese history are covered in the collection. Furthermore, we mapped all individual dynasties and modern periods onto long periods in order to generate a broader view of Chinese history. These long periods follow the periodisation adopted by the *Journal of Chinese History*, one of the most prominent academic journals in the field, namely:

Pre-Tang (–618 C.E.), Tang-Song-Yuan (618–1368 C.E.), Ming-Qing (1368–1912 C.E.), and the 20th Century (1912 C.E.–). Details of the mapping process are available in

Appendix_Table_1_Period_Mapping.csv in our anonymous GitHub repository (<https://anonymous.4open.science/r/Code-for-Analysis-of-Chinese-History-Collection-in-the-Library-of-Congress-Catalog-6BC2>), which also includes an additional dataset (described below in Section 3.2) and the code used in this paper.

Results and discussion

General publication trends

We first analysed the proportion of books on Chinese history relative to the total number of books published each year after 1945, both in the entire LC catalog and within each of the three regional corpora. As shown in Figure 2, the representation of Chinese history in the entire LC catalog, in Greater China, and in the U.S. exhibits two major surges in the early 1980s and early 21st century. These trends correspond to China's 'reform and opening up' beginning in 1978 and its accession to the World Trade Organisation (WTO) in 2001. We tentatively interpret these patterns as follows, while acknowledging that further work—such as close examination of individual books—is needed for a more robust explanation: the political and social reforms of the late 1970s increased academic openness within mainland China, encouraging new historical writing that moved beyond the strictures of Marxist historiography. At the same time, the end of China's isolation from the outside world fostered greater international scholarly engagement with Chinese history. Similarly, WTO membership marked another turning point: domestically, rapid economic growth expanded resources for historical research within China, while internationally, China's rising global influence attracted greater scholarly attention to its past.

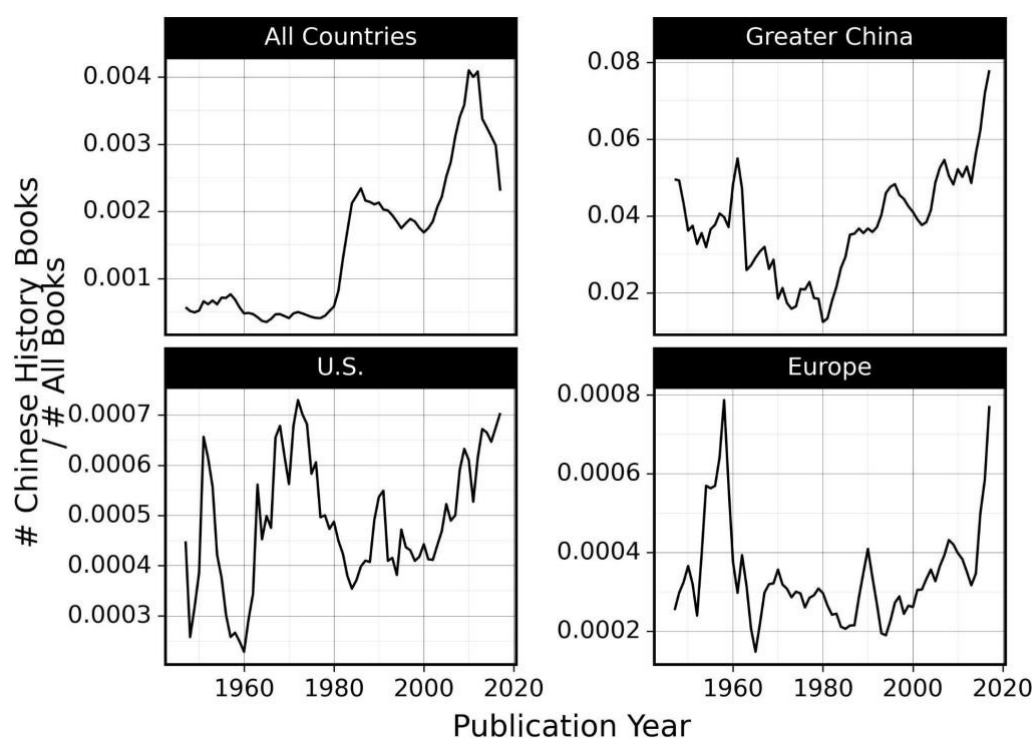


Figure 2. Proportion of books on Chinese history relative to the total number of books published each year, in the entire LC catalog and in the three regional corpora.

Notably, the rapid growth of the 1980s is not evident in the European corpus, suggesting that what Perry (2010) characterises as Western scholars' turn 'up to the archives and down to the villages' (p. 265) in response to China's opening was driven primarily by U.S. researchers. In addition, while the first panel shows a decline in the proportion of Chinese history books within the LC catalog in the

late 2010s, this trend is absent in the three regional corpora, all of which continue to rise. This indicates that the decline does not reflect waning interest in Chinese history but results from the growing share of publications from other regions, where Chinese history is rarely a focus, within the entire LC catalog.

To further compare publication trends across the three regions, we visualised the number of books published over time in each corpus. We plotted three-year moving averages to reduce short-term fluctuations, and the y-axis is displayed on a logarithmic scale to ensure readability. Figure 3 shows that publications from Greater China lagged behind those from the U.S. and Europe during the Cultural Revolution (1966–1976), but quickly surged ahead after the onset of ‘reform and opening up’ in 1978, eventually surpassing both and maintaining a substantial lead. In addition, U.S. publications have outnumbered European ones since the early 1960s and have held a modest advantage ever since. These results suggest that the U.S. overtook Europe in Chinese studies earlier than the 1970s, which Xu (2010) identifies as the turning point (p. 4).

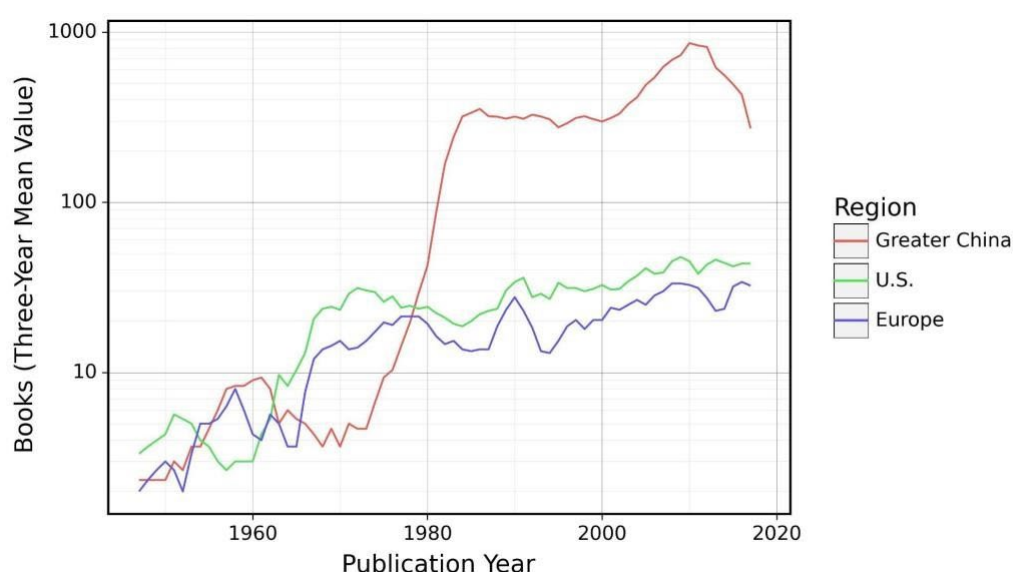


Figure 3. Number of books published each year in the three regional corpora.

Table 1 illustrates language use across the three regional corpora. The results reveal a strong correlation between place of publication and language: the overwhelming majority of books published in Greater China are in Chinese, while most of those published in the U.S. are in English. By contrast, books published in Europe display greater linguistic diversity, with more than half written in local languages such as Russian, French, and German.

Region	Total books	% Chinese	% English	% Others
Greater China	15,624	98.44%	1.24%	0.32%
US	1,750	4.74%	94.63%	0.63%
Europe	1,190	0.08%	43.36%	56.55%

Table 1. Distribution of book languages in the three regional corpora.

Historical period representation

To further explore patterns of historiography on China and address RQ2, we analysed the scholarly attention given to different dynasties and modern periods using two measures derived from

temporal information in the book metadata: (1) *Full Count*, in which each book about a period is counted as 1, and (2) *Fractional Count*, in which a book covering n periods contributes $1/n$ to each period. For example, a book tagged with the chronological subdivision '581–907' would add 1 to the *Full Count* of both the Sui dynasty (581–618 C.E.) and the Tang dynasty (618–907 C.E.), but only

0.5 to the *Fractional Count* of each of the two dynasties. Building on these measures, we introduced an indicator, the *Dedication Index*, defined as the ratio of *Fractional Count* to *Full Count*, to assess the extent to which scholarship on a given period focuses exclusively on that period. A *Dedication Index* of 1 indicates that every book associated with a period is devoted solely to it, rather than treating the period as part of broader, multi-period studies. Detailed results for the *Full Count*, *Fractional Count*, and *Dedication Index* for each dynasty and modern period are available in our anonymous GitHub repository (Appendix_Table_2_Dynasty_Representation.csv).

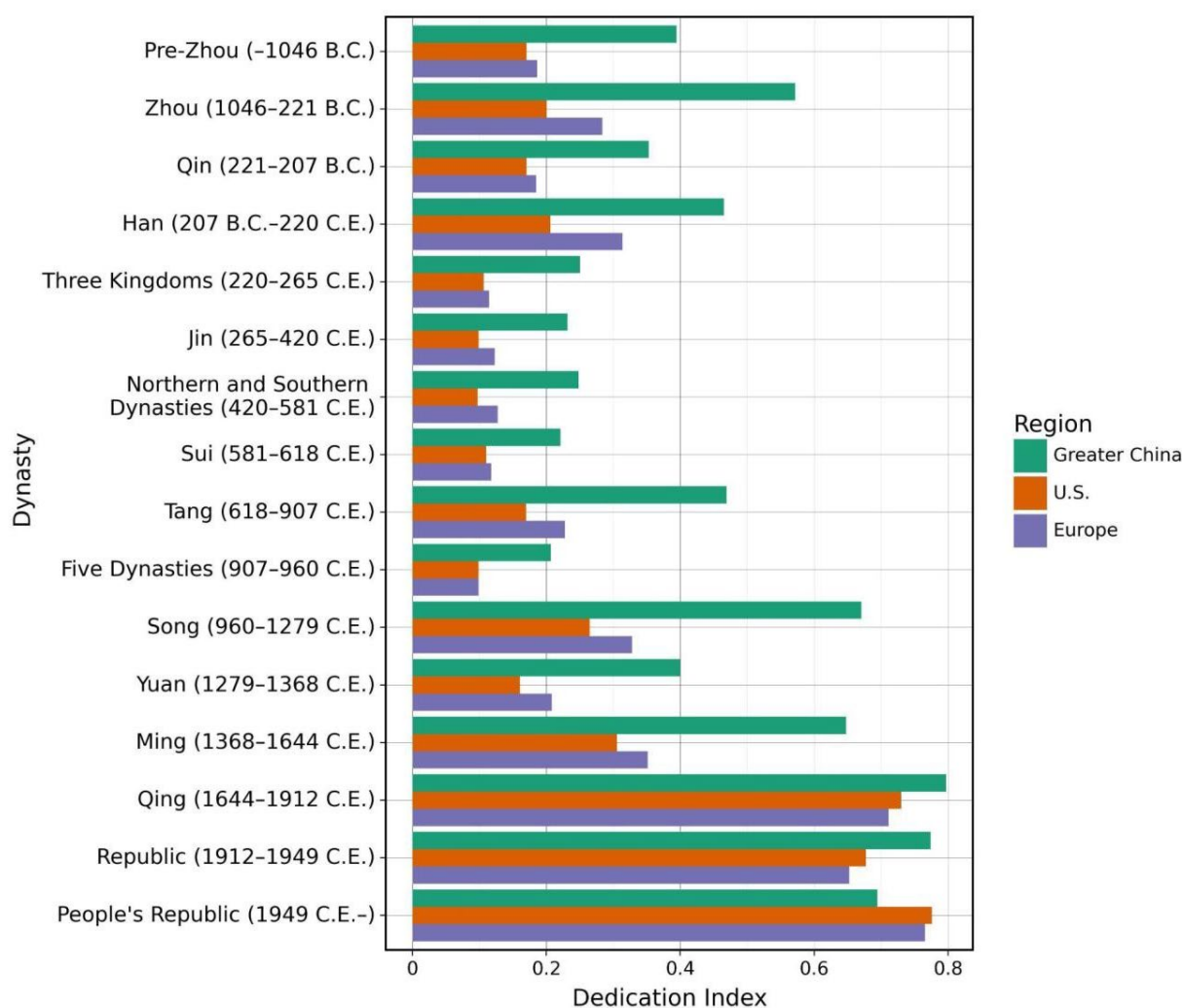


Figure 4. Dedication index by dynasty and modern period in the three regional corpora.

Figure 4 reveals both temporal and regional patterns. Temporally, the most recent dynasties and modern periods (Qing, Republic, and People's Republic) are more likely to receive exclusive scholarly attention, with *Dedication Index* values above 0.65 across corpora. By contrast, earlier dynasties are much less likely to be the sole focus of a book, especially during the 'early medieval period' (Three Kingdoms, Jin, Northern and Southern Dynasties, and Sui), which Chennault and Pearce (2013) describe as 'one of the most complex and least familiar periods of Chinese history' (p. 11), with the *Dedication Index* falls to 0.25 or lower in every corpus.

Regionally, with the sole exception of the People's Republic period, Greater China shows the highest *Dedication Index* for each dynasty and modern period, with especially pronounced differences for earlier dynasties. Between the U.S. and Europe, the latter consistently shows slightly higher *Dedication Index* values for most dynasties and modern periods, except the three most recent ones. These findings suggest that the historiography of China diverges significantly across regions for earlier periods: in Greater China, each dynasty is more often treated as a distinct domain, whereas in the West, earlier dynasties are more typically discussed within extended narratives spanning multiple dynasties—a pattern more evident in the U.S. than in Europe.

Finally, we examined changes in the share of different long periods, as defined in the last paragraph of Section 2. Because this analysis further subdivides the books published in each regional corpus each year by historical periods represented, we excluded the years 1946–1978 to avoid rapid fluctuations caused by limited sample sizes. These years, prior to the onset of China's '*reform and opening up*' in 1978, account for only 4.9% of the three regional corpora (901 of 18,564 books).

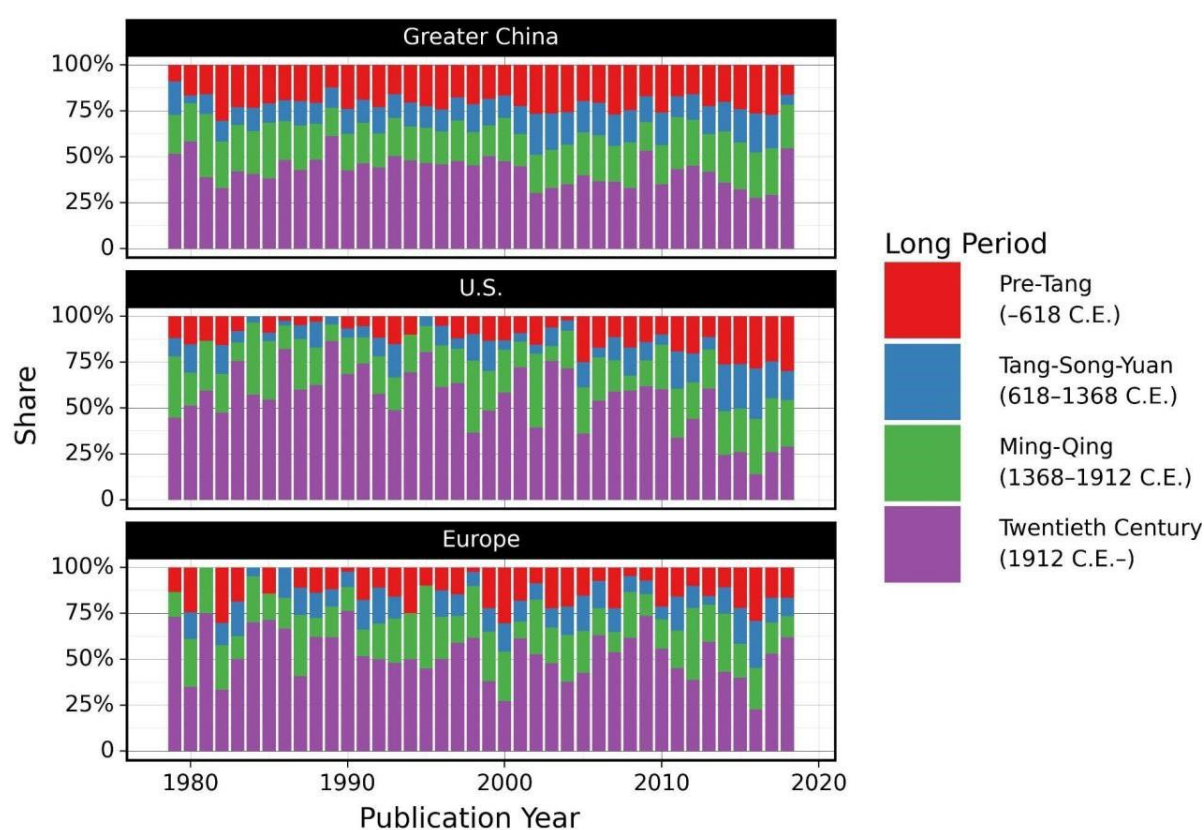


Figure 5: Share of long periods of Chinese history in the three regional corpora, 1979–2019.

Overall, Figure 5 shows that the distribution of long periods in the Greater China corpus remained relatively stable, and although there are larger fluctuations in the European corpus, no consistent pattern of increasing or decreasing share emerges for any long period. In contrast, a distinct trend appears in the U.S. corpus: in the last two decades of the twentieth century, books on Chinese history were dominated by works on the two later long periods, Ming-Qing and the Twentieth Century, which together averaged 83% of all books between 1979 and 2000. In the twenty-first century, however, the share of the two earlier periods, Pre-Tang and Tang-Song-Yuan, increased substantially, reaching an average of 50% between 2014 and 2018. This shift is not surprising, given that later periods historically attracted far more attention in U.S. scholarship on Chinese history.

For example, Zhang et al. (2013) devote four chapters to the Qing dynasty, while all pre-Qing dynasties together receive only five (pp. 1–173). Chennault and Pearce (2013) date the emergence of early medieval China as a distinct domain of research in the U.S. only to the late 1970s (p. 11), while Birge (2013) identifies the 21st century as a period of ‘revival of Yuan studies in North America’ (p. 65).

Previous surveys of studies of Tang and Song history, however, present different narratives. De Weerdt (2013) traces a ‘significant transition’ in studies of Song history to 1970, with the establishment of the *Song Studies Newsletter* (p. 23). Lu (2010) characterises 1970–1990 as the ‘mature period’ of studies of Tang history (p. 87), followed by the 1990s as a period of ‘relative silence’ (p. 94). Our empirical results suggest that although foundational works in these fields appeared as early as the 1970s, the number of publications grew significantly only in the 21st century. This highlights how examining large-scale publication patterns through quantitative analysis of all works in a field can provide a complementary perspective on its vitality beyond what canonical works alone reveal.

Conclusion, implication, and future work

Our results reveal that, in terms of general publication trends, there were two major surges in the early 1980s and early 21st century in the proportion of books on Chinese history relative to the total collection. Among the three regional corpora, Greater China has substantially led in publication numbers since the 1980s, while the U.S. has outnumbered Europe since the 1960s. Regarding historical period representation, later periods are more likely to have books devoted exclusively to them, whereas earlier periods are typically discussed within broader narratives spanning multiple dynasties in Western scholarship. Moreover, U.S. research has recently shifted to a more diversified distribution in which earlier periods received growing attention. These results confirm many broad patterns noted by existing scholarship, but they also challenge arguments based on a limited number of canonical works, such as dating the U.S. surpassing Europe in Chinese history scholarship to the 1960s rather than the 1970s and identifying the growth in Tang and Song publications as a 21st-century phenomenon, despite foundational works appearing earlier.

This is one of the first articles to use bibliographic records to analyse the historiography of Chinese history. Our preliminary results confirm and advocate for the value of library metadata in supporting quantitative analysis of humanities scholarship. We further call on the LIS community to place greater emphasis on the potential of metadata beyond its traditional role of providing access to library collections, and to explore how it can be incorporated into large-scale quantitative analysis or the development of AI applications in the future.

Our future work will proceed along several directions. First, we will expand our work by conducting large-scale comparative analyses between the LC book collections and those of other major libraries, such as the British Library and Yale University Library, to draw broader findings. Another comparable corpus is the journal articles on this topic, where we are interested in understanding how academic attention is paid to different periods and topics in these two fundamental publication venues for historians. All these comparisons will provide more contextual understanding of the evolution of the Chinese history scholarship. We will also employ large language model (LLM) applications to extract more granular subject information from book abstracts, enabling us to examine the thematic content of books beyond the often-limited subject terms assigned by librarians.

Through advancing our scholarship using different approaches, we seek to engage more deeply with historians who take traditional approaches to studying Chinese history, examining whether our results align with or diverge from established historiography. We aim to demonstrate the unique values of library metadata to understand the scholarship systematically and comprehensively in social science and humanities. We believe this preliminary study, together with

its future development, will foster more meaningful conversations among humanists, library and information scientists, and digital humanities researchers, contributing to the development of a more diverse and robust data infrastructure.

About the authors

Kai Li is an Assistant Professor in School of Information Sciences, University of Tennessee, Knoxville, USA. He received his Ph.D. from the Department of Information Science at Drexel University, and his research interest are scholarly communication, quantitative science studies. He can be contacted at kli16@utk.edu.

Wenyi Shang is an Assistant Professor in the iSchool, University of Missouri, USA. He received his Ph.D. from the University of Illinois Urbana-Champaign, and his research interests are in the digital humanities. He can be contacted at wenyishang@missouri.edu.

Yingying Han is an Assistant Professor in the School of Information Sciences at the University of Tennessee, Knoxville. She received her Ph.D. from the University of Illinois Urbana-Champaign, and her research interests include digital preservation of cultural heritage, critical archival studies, and community engagement. She can be contacted at yhan25@utk.edu.

References

- Birge, B. (2013). Yuan studies in North America: Historical overview, contributions, and current trends. In H. Zhang, Z. Xue, S. Jiang, & G. L. Lugar (Eds.), *A scholarly review of Chinese studies in North America* (pp. 54–78). Association for Asian Studies, Inc.
- Burrows, T., & Verhoeven, D. (2014, June 2). Digital scholarship in the humanities and creative arts: The HuNI virtual laboratory. *EDUCAUSE Review*.
<https://er.educause.edu/articles/2014/6/digital-scholarship-in-the-humanities-and-creative-arts-the-huni-virtual-laboratory>
- Chennault, C. L., & Pearce, S. (2013). History and society of early medieval China. In H. Zhang, Z. Xue, S. Jiang, & G. L. Lugar (Eds.), *A scholarly review of Chinese studies in North America* (pp. 11–22). Association for Asian Studies, Inc.
- De Weerdt, H. (2013). Song studies. In H. Zhang, Z. Xue, S. Jiang, & G. L. Lugar (Eds.), *A scholarly review of Chinese studies in North America* (pp. 23–53). Association for Asian Studies, Inc.
- Delgado-Quirós, L., & Ortega, J. L. (2025). Research entity information and coverage in eight free access scholarly databases. *Online Information Review*.
- Library of Congress. (2025). MARC 21 Format for Bibliographic Data. Library of Congress.
<https://www.loc.gov/marc/bibliographic/>
- Library of Congress Cataloging Distribution Service. (2019). Books all: MDS Connect dataset, 24 MARC Distribution Services (data set) books all. Library of Congress. Cataloging Distribution Service. <https://www.loc.gov/item/2020445551>
- Li, K., Dobreski, B., & Busch, M. (2025). How are library cataloguing metadata used differently over time: a large-scale quantitative analysis of the Library of Congress catalog. *Journal of Documentation*, 81(2), 385–402.
- Lo, W. Y. W. (2016). The concept of greater China in higher education: adoptions, dynamics, and implications. *Comparative Education*, 52(1), 26–43.

- Lu, Y (2010). Xifang Tang shi yanjiu gaiguan [An Overview of Western Scholarship on Tang History]. In H. Zhang, Z. Xue, & S. Jiang (Eds.), *Beimei Zhongguoxue—Yanjiu gaishu yu wenxian ziyuan* [Chinese studies in North America—Research and resources] (pp. 83–110). Zhonghua shuju.
- Maleki, A. (2022). OCLC library holdings: Assessing availability of academic books in libraries in print and electronic compared to citations and altmetrics. *Scientometrics*, 127(2), 991–1020.
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106, 213–228. <https://doi.org/10.1007/s11192-0151765-5>
- Mordell, D. (2019). Critical Questions for Archives as (Big) Data. *Archivaria* 87, 140–161. <https://muse.jhu.edu/article/724731>.
- Moretti, F. (2000). Conjectures on world literature. *New Left Review*, 1, 54–68.
- Moretti, F. (2013). *Distant reading*. Verso.
- Padilla, T., Allen, L., Frost, H., Potvin, S., Russey Roke, E., & Varner, S. (2019). Always already computational: Collections as data—Final report. Zenodo. <https://doi.org/10.5281/zenodo.3152935>
- Perry, E. J. (2013). Studying Chinese politics: Farewell to revolution? In H. Zhang, Z. Xue, S. Jiang, & G. L. Lugar (Eds.), *A scholarly review of Chinese studies in North America* (pp. 264–291). Association for Asian Studies, Inc.
- White, H. D., & Zuccala, A. A. (2018). Libcitations, WorldCat, cultural impact, and fame. *Journal of the Association for Information Science and Technology*, 69(12), 1502–1512.
- Xu, Z. (2010). Xu [Introduction]. In H. Zhang, Z. Xue, & S. Jiang (Eds.), *Beimei Zhongguoxue—Yanjiu gaishu yu wenxian ziyuan* [Chinese studies in North America: A scholarly review and bibliographic resources] (pp. 1–7). Zhonghua shuju.
- Zhang, H., Xue, Z., & Jiang, S. (Eds.). (2010). *Beimei Zhongguoxue—Yanjiu gaishu yu wenxian ziyuan* [Chinese studies in North America—Research and resources]. Zhonghua shuju.
- Zhang, H., Xue, Z., Jiang, S., & Lugar, G. L. (Eds.). (2013). *A scholarly review of Chinese studies in North America*. Association for Asian Studies, Inc.

© [CC-BY-NC 4.0](#) The Author(s). For more information, see our [Open Access Policy](#).