# ReOrienting Digital History: New frontiers in Chinese Studies

### Cécile Armand [orcid](https://orcid.org/0000-0002-9107-6443) Aix-Marseille University

[cc-by-nc-nd](https://creativecommons.org/licenses/by-nc-nd/4.0/) ©. Published by De Gruyter in cooperation with the University of Luxembourg Centre for Contemporary and Digital History. This is an Open Access article distributed under the terms of the [Creative Commons Attribution License CC-BY-NC-ND](https://creativecommons.org/licenses/by-nc-nd/4.0/)

from IPython.display import Image, display  
display(Image("./media/liuri\_wordcloud.png"))



China, Digital humanities, Computational turn, Area studies

This article serves as the introduction to the special issue on Chinese digital history in the *Journal of Digital History* (JDH). After presenting the papers gathered in this special issue, it scrutinizes the significance of area studies and advocates a paradigm shift towards computational methodologies in both Chinese studies and the broader humanities disciplines.

## Introduction

This journal special issue is the outcome of the [“New Foundations in Chinese Digital History” Conference](https://summi.enpchina.eu/members/) held in June 2022 under the auspices of the “Elites, Networks, and Power in modern China” [(ENP-China)](https://www.enpchina.eu/) project in Aix-en-Provence. Drawing connections between papers gathered in a special issue is often a challenge. Indeed, each article constitutes a unique research driven by specific interests and involving specific expertise. In this particular issue, the breath of topics, historical periods, sources, and methods covered by the papers makes the editor’s task particularly difficult. What do they actually have in common, besides sharing the same origin and having something to do with Chinese history and digital methods? How do their approaches differ from the digital scholarship produced outside the field of Chinese studies? The question is not an artificial one. It is a unique opportunity to explore the meanings and implications of digital history within and beyond the China field.

The five papers that will progressively be published and constitute this special issue illustrate the multifarious ways digital methods can be applied in the field of Chinese studies. Paul Vierthaler takes the famous late Ming novel *Jinpingmei* (Plum in the Golden Vase) as a test case for developing a generalizable approach for automatic detection of intertextuality and disambiguation of sources in Chinese literary history. He is not only able to map the uneven distribution of intertextuality across the novel, but also to computationally assess the influence of earlier works like *Shuihu Zhuan* and the likely influence of *Jinpingmei* on contemporary works such as *Cinüe*. Subsequently, Elisabeth Kaske and Florian Kessler utilise Monte Carlo simulation and network analysis to reveal the “hidden rules” of the Qing mandarinate based on a popular game of the time, shedding light on both the construction of the Qing social reality, as well as the game designers’ perceptions of it. Their findings highlight the significance of hereditary privileges, which challenges the commonly held view of a meritocratic system, and that of office purchases, which introduced an element of strategy but do not diminish the importance of hereditary privileges and examination degrees. Taking the Chinese newspaper *Shenbao* as his core source, Christian Henriot combines natural language processing, network analysis, and topic modelling to determine who was newsworthy in the early Chinese press, and in which context eminent and less eminent people appeared in the Shanghai newspaper. He found that a small segment of literati emerged amidst a sea of nobodies, which mirrors the concerns and expectations of *Shenbao*’s intended readers and reveals the process of news making in the early phase of Chinese journalism. Cécile Armand also employs topic modelling - a statistical method aimed at identifying latent themes in text corpora based on the most frequent word co-occurrences - to map the development of a transnational public sphere in China during the Republican period. She demonstrates that despite growing tensions during the interwar years, major newspapers, both Chinese and foreign, and multinational organisations like the Rotary Club consistently emphasised non-violent modes of action, through meetings, entertainment venues, lectures, and philanthropy, in sharp contrast with the war of resistance advocated by nationalist organisations and political publications during the same period. Finally, Marylin Levine and Heidi Yu Huang utilise a combination of quantitative analysis, hierarchical clustering, network analysis, and “mini-biographical portraits” to explore a group of Chinese students in Europe. Their “enriched prosopography” approach illuminates individual and group dynamics and helps to understand why, in the context of growing politicization after 1927, some became revolutionaries, while others chose to prioritise their academic careers and became political intellectuals.

Despite their diversity, all articles share three key characteristics: first, a taste for experimentation and a remarkable ingenuity in devising solutions to overcome the common set of challenges they face; second, a critical attitude towards the use of digital tools and methods; and third, a strong interdisciplinary engagement. These three common traits derive from the challenges of working with computational tools and historical data. One of the main difficulties is the large volume of data and the wide scale of observation, which cannot be handled through the “traditional” methods of close reading and micro-analysis. To address this challenge, the authors imagine various reduction methods such as classification and filtering (Vierthaler, Armand), hierarchical clustering (Levine and Huang), pruning techniques (Henriot, Levine), and contextualised case studies. For example, Vierthaler analyses the specific textual connections between *Jinpingmei* and one of its most influential sources (*Shuihu Zhuan*). Kaske and Kessler examine more specifically the figure of the prefect, while Henriot and Levine focuses on particular ego-networks and clusters. The authors continuously shift their lenses to navigate from individuals to groups (Henriot, Levine and Huang), from words to documents and corpora (Vierthaler, Henriot, Armand). To overcome the issue of article separation in massive press corpora, Armand applies topic modelling to selected segments of texts using concordance. Another challenge is the complexity of historical data, which cannot always be easily bent to computational tools. For example, Kaske and Kessler use simulation and probabilities to handle the complex system of bureaucratic appointments. To tackle the variety of social life as reported in contemporary newspapers, Henriot combines the network analysis of actors with the topic modelling of article contents. Missing information is another big issue. For example, the uncertain date of publication of *Jinpingmei*, incomplete game rules, truncated names in *Shenbao*, and gaps in archival documentation raise significant challenges. One of the strategies to fill the holes consists in formulating rational hypotheses informed by the authors’ intimate knowledge of the context and previous scholarship.

The recognition of these challenges shapes the authors’ modesty and drives their commendable efforts to make historical research more transparent. Their contribution to “digital hermeneutics” () encompasses its two key dimensions, namely source and tool criticism. The authors critically confront different source materials and algorithmic models to unveil their potential biases. For example, Vierthaler experiments with different models for text classification and develops proofs of concept to test their robustness before applying them to unknown cases. His philosophy is to “use the simplest methods possible that provide reliable results”. For this particular study, he chose to rely on the well-established “bag-of-words” representation of text (n-gram), instead of the trendy, yet less stable transformer-based models. Henriot and Armand compare different topic models, while Kaske and Kessler contrast different versions of the game and consider various possible scenarios for reaching the highest position with the highest return. This critical and experimental state of mind also helps the authors to anticipate the risk that the selected methods may “fail” or may not provide the expected level of accuracy, which is remarkably explained in Vierthaler’s paper. While he acknowledges the limitations of his “bag-of-word” approach, he is also able to draw important lessons from them. Trying to understand why and where the model fails leads to new discoveries about the original texts and to identifying the mentions of persons’ names as a major source of noise.

This “thinkering” attitude - a neologism coined from the ingenuous merging of “thinking” and “tinkering”, which describes “the action of playful experimentation with technological and digital tools for the interpretation and presentation of history” () - demands strong interdisciplinary training or collaboration, particularly with computer sciences. In both cases, the authors manifest a deep commitment with the alter disciplines, which goes far beyond the passive reliance on “engineer servants”, as it is often the case among self-claimed digital humanists.

Despite their computational twist and turns, the authors never lose sight of their scholarly questions and never depart from their disciplinary grounding. On the contrary, the strength of their analyses lies in their genuine engagement with the existing scholarship and their intimate knowledge of the intellectual debates in their respective fields. This knowledge not only helps them to formulate sound research hypotheses; it also provides a basis to strengthen, nuance, or revise previous findings, and to open new avenues for future investigation. For example, Vierthaler is able to make new hypotheses regarding the likely author and date of publication of *Jinpingmei*. Kaske and Kessler demonstrate the understudied importance of the prefect position as a springboard toward high officialdom. Henriot convincingly argues that a computational approach to the historical press challenges the previous, binary conceptualisation of newspapers as either texts or sources, while Armand provides insights into the circulation of information and journalistic practices among Chinese and foreign periodicals. Her corpus-based approach also contributes to substantiate the elusive notion of the “public sphere” and to precisely map its discursive practices. Levine and Huang’s findings from network analysis highlights Cai Yuanpei’s ambivalent relationship with the French model of education and suggests that Wang Jingwei may have been a key bridging figure between Japan and Vichy France during WWII.

The *Journal of Digital History* (JDH) offers an ideal venue for reshaping the ways of writing Chinese history in the digital age. The authors skilfully appropriate the unique, multi-layered structure of the journal to deploy their experimental process and their critical perspective at the hermeneutic level, while using the narrative layer to expose their scholarly interpretations and engage with previous literature. The data and code are publicly accessible on Github, which significantly increases transparency and makes a huge step toward the ideal of making experiments reproducible and results more easily verifiable. Ultimately, the three common threads that knit the articles together significantly expand the boundaries of Chinese studies and reshape the very notion of the “digital”.

## Unlocking Area Studies

What makes Chinese studies different from other fields in which digital methods have been applied? To what extent is it relevant to single out *Chinese* digital history? For sure, some of its features require specific training and expertise: the Chinese language, or more accurately, *languages* in the plural, given its multiple writing styles and accelerating changes since the late nineteenth century, (); the primary source materials available to China historians, although the digitisation of original sources and their conversion into data tend to smoothen material differences; and expert knowledge of the Chinese historical context. But after all, most of these requirements apply to other domains outside the China field as well. The distance between Chinese and non-Chinese digital history, ultimately, is no greater than the distance among the papers in this issue. It essentially depends on the data historians utilise and the research questions they address, which in turn, determines the tools and the methods they adopt. To some extent, the “digital” is bringing Chinese studies closer to other fields of research which also make use of digital data and methods. The greatest promise of the digital age is perhaps to relieve “area studies” from the scholarly isolation they have been enduring for decades (). The digital may appear as an opportunity to reconnect the so-called area studies with more “central” areas of scholarship through a common language - the language of data - and a shared set of tools and methodologies.

For now, however, Chinese studies and digital humanities (DH) alike are still the subjects of specific publications or special journal issues in Europe (see for instance, the special issue [“Beyond Guanxi”](https://historicalnetworkresearch.org/journal-of-historical-network-research-vol-5-no-1-2021-beyond-guanxi-chinese-historical-networks/) () in the *Journal of Historical Network Research*, or the more recent [special issue on East Asian Studies](https://link.springer.com/journal/42803/volumes-and-issues/4-1) in the *International Journal of Digital Humanities*) (). This *JDH* special issue aims to join this wider effort to bridge the “old Europe” with the growing DH community in Asia. During the last two decades, the field of digital humanities has experienced a formidable expansion in the Chinese-speaking world (). DH conferences, research centres, and training programs have multiplied in China, Taiwan, and Hong Kong. While Anglo-Saxon countries have been at the forefront of digital humanities over the last forty years, these new China-born initiatives will be setting the tone in the next decade. They are already paving the way towards what may become the future of digital humanities – the new *computational* paradigm driven by artificial intelligence.

## Beyond the Digital: The Computational Turn

At first sight, the computational turn may sound like a return to the origins of Digital Humanities, which grew out of the forgotten “humanities computing” in the early 1960s (). However, the computational turn marks the beginning of a radically new era in humanities research. With the rapidly expanding field of artificial intelligence, deep learning, and large language models, humanities researchers can now envision “qualitative” research and deep reading at a wider scale, instead of merely counting words or mapping large trends. The new computational paradigm not only makes irrelevant most of the criticism raised against “cliometrics” and macro-analysis, but it radically challenges the binary framework that has emerged with the digital humanities, exemplified by the series of oppositions between quantitative/qualitative, macro/micro, close/distant reading. Levine and Huang’s “integrative approach”, which proposes to understand macro-historical changes through the lenses of individual choices and actions, is an excellent illustration of the potential of this multi-dimensional perspective.

The five articles in this special issue are all contributing a leap forward in this direction. Their authors share what might be called a “computational way of thinking”, which encompasses three key dimensions: the capacity to translate historical questions into machine-actionable tasks; the aptitude to interpret computational outputs from a humanistic and critical perspective; and the ability to weave their discoveries with previous findings into a consistent, compelling narrative that speaks to a wide community of readers beyond their specific field of expertise. These are the three vital skills that scholars will have to acquire in the near future if they want to maintain their relevance in the age of algorithms and artificial intelligence.

## Bibliography