



## Introduction

---

THOMAS J. MAZANEC, JEFFREY R. THARSEN, and JING CHEN

The goal of this special issue is to highlight some exemplary ways in which the digital humanities are being applied to the study of classical Chinese literature. By *digital humanities* we mean methods of humanistic inquiry assisted by digital sources and tools. The articles in this issue cover a wide range of such sources and methods, but rather than focus on theory or methodology, they provide concrete case studies that offer new insights driven by digital tools and databases. These articles do not just promise to open up new avenues of inquiry but represent tangible efforts to make good on that promise. They put forth bold conclusions about the history of traditional Chinese literary culture, showing how these technologies can help support and extend the traditional concerns of philology and literary studies: to reexamine classical literary texts within the contexts of their production, reception, and circulation.

The digital humanities, simply put, are the humanities aided by computers. Nearly all literary scholars now access research digitally, scour source texts in massive online corpora, collaborate via e-mail and cloud-based word processors, and use online technologies to distribute their work. Digital tools are already a pervasive part of nearly all forms of scholarship.

Of course, some humanistic studies are more firmly rooted in the digital than others. We generally apply the term *digital humanities* to works that use one or more computing technologies as a methodological cornerstone. These often involve some amount of data modeling and quantitative analysis. Though often couched in the language of innovation, such approaches do not mark a sharp break with earlier methods. The history of systematically sorting, counting, and analyzing literary texts is long. In sinophone academia, it begins at least as

far back as the large-scale *kaozheng* 考證 (evidential studies) scholarship of the Qing dynasty (1644–1911); can be traced through early twentieth-century reformers' obsession with numbers, charts, and graphs; and continues to the present.<sup>1</sup> In anglophone academia, digital approaches to literary texts, such as stylometry, distant reading, bibliography, and literary sociology and geography, can claim a similarly venerable pedigree.<sup>2</sup>

The digital humanities, like their predecessors, cobble together an amalgam of methods to advance new claims about humanistic topics. Like other humanists, practitioners of the digital humanities tend to be self-critical about their methods: we understand that the very framing of a question shapes the answer, that data modeling is itself an act of interpretation, and that computer-assisted analysis becomes meaningful only with human intervention. The digital humanities do not disrupt previous generations of humanistic inquiry; they support and extend them.

Digital technologies also allow scholars to more clearly present the process of their research. Like the older technology of the footnote, online repositories let one check the sources underlying a claim, trace its steps, and reexamine its conclusions if necessary.<sup>3</sup> To this end, we have created a data repository for each of the articles in this issue, where readers may download original data sets, technical appendices, and related documentation (see [www.chinese-poetryforum.org/?page\\_id=1512](http://www.chinese-poetryforum.org/?page_id=1512)). In this way, we attempt to model openness in our scholarship—a practice that is already ascendant in some circles, and one that we hope may become ubiquitous.

Our first article comes from Donald Sturgeon, the founder and curator of one of the best-known textual databases in the field, the Chinese Text Project ([ctext.org](http://ctext.org)). In his article, Sturgeon describes how to find patterns of “text reuse” throughout early Chinese texts. His method is designed to help identify highly similar word usage and direct borrowings between discrete passages and to discover more amorphous forms of intertextual relationships. Sturgeon's examples, primarily drawn from the *Mozi* 墨子, include both detection of specific, small-scale textual parallels and calculation of overall lexical similarity between chapters—a method for algorithmic identification of parallels that can be applied to any text or corpora. In the second half of the article, he provides a detailed comparison between several text reuse metrics (cosine similarity and term frequency–inverse document frequencyweighting) and his “n-gram overlap” method, with the end goal of being able to identify and interrogate highly similar lexical elements even when they occur within radically different rhetorical structures. Computational techniques, Sturgeon concludes, are efficient when it comes to performing large-scale data-driven tasks; these practices are useful for identifying potential correlations, and then it is up to the experts to

interpret their significance and causes. His new methodology and digital tool kit for identifying text reuse will prove invaluable to scholars seeking to make sense of early Chinese textual relationships on a large scale.

The next article, by Evan Nicoll-Johnson, analyzes text reuse in one of its traditional forms: annotation. Specifically, he focuses on shared bibliographic notes in two important texts from the fifth and sixth centuries, the *Sanguozhi* 三國志 (Record of the Three Kingdoms) and *Shishuo xinyu* 世說新語 (New Account of Tales of the World). By creating a citation network from these notes, Nicoll-Johnson shows the extent to which these two very different texts emerged from a shared bibliographic environment. The early medieval period saw a surge in access to books, and historiographers drew extensively from this new resource to write their histories. By representing these texts' notes as a network, Nicoll-Johnson not only sheds light on the multipolar relations between many texts but also provides a new visual metaphor for the production and circulation of such texts. In this way, early medieval texts should be understood not as discrete units but as composites of dozens or hundreds of shared passages. This kind of argument, about the blurry boundaries between original texts and compendia, had been made in the predigital era, but only recent technology allows it to be so powerfully articulated and deeply felt in a visual form.

The next four articles offer various takes on the celebrated poetry of the Tang dynasty (618–907). The boldest, methodologically speaking, is Mariana Zorkina's article on "poems on things" (*yongwu shi* 詠物詩). Zorkina uses distributional semantics and neural networks to describe common correlations between words, lexical strings, and whole poems in this popular verse genre. Her work is valuable because it describes with precision the baseline of Tang poetic discourse. If individual style, as some claim, is deviation from a norm,<sup>4</sup> then the norms limned by Zorkina and her algorithm are crucial for understanding the signature achievements of Li Bai 李白 (701–62), Du Fu 杜甫 (712–70), and dozens of other beloved poets. Additionally, Zorkina's article points to deeper, unexpected congruencies. There is a strong correlation, for example, between the usage of *tiger* (*hu* 虎) and *happiness* (*xi* 喜) in poems on things. This is not because tigers bring joy but, rather, because both terms partake of shared linguistic modes governing poetic expression. Zorkina's research thus demonstrates how computers can highlight previously unseen patterns that close reading can then explicate.

Chao-lin Liu (with Mazanec and Tharsen) also proposes ways of understanding the macroscopic patterns of Tang poetry. The main purpose of Liu et al.'s article is to introduce readers to the possible applications of a textual algorithm Liu developed, FindCommon, to the study of classical Chinese poetry.

His tool is not just an improved version of the classic concordance or word search—the basic functions one might expect from such an algorithm. It goes further by offering a window on word co-occurrences, individual poets' styles, quotation and allusion, the social relations between poets, and diachronic changes in the poetic tradition. Liu's program even goes so far as to highlight limitations in the textual sources and their digital analysis, such as multiple authorial attributions to a single poem in *Quan Tang shi* 全唐詩 (Complete Poems of the Tang Dynasty)—his algorithm points to gaps in the archive that require further philological investigation to resolve.

Thomas J. Mazanec's article takes a deep dive into one of Liu's concerns, the relations between poets as asserted in their works. In search of a new take on literary history that does not privilege a few emblematic texts, his article attempts to reconstruct the late Tang poetic world as a vast network of imagined literary relations. Combining social-network analysis with close readings, the article concludes that mobility became increasingly important to a poet's place in the network as the Tang collapsed in the late ninth century. This, in turn, calls attention to the centrality of figures normally marginalized in Tang literary history, such as Jia Dao 賈島 (779–843) and Buddhist poet-monks. By using the framework of the network, the “dynamic literary history” Mazanec proposes sees poets not as static icons but as actors who move between genres, modes, styles, cliques, and locations.

Location is the main topic of Wang Zhaopeng and Qiao Junjun's article (translated by Mazanec), which looks at the geographic distribution of the Tang poetic world. Using a meticulously curated database of poets' hometowns and the places they traveled, Wang and Qiao reach several important conclusions about the geography of Tang poetry. First, northern cities produced more poets until the late Tang (835–907), when they began to be outnumbered by southerners. Second, no matter where poets were from, most poetry was written in the south, meaning that much of it was written by northern poets while traveling far from home. Third, the two capitals of the Tang, Chang'an 長安 and Luoyang 洛陽, held by far the greatest appeal for poets; nevertheless, poetry was produced everywhere, even in undeveloped backwaters and remote provinces. By focusing on geography, Wang and Qiao highlight the conditional nature of Tang poetry: the vast majority was produced in response to specific circumstances, on specific occasions, in specific places. Tang poems are ephemera as much as they are monuments, and that very ephemerality is one source of their power.

Ephemera become monuments through canonization, and canonization is precisely the subject of Timothy Clifford's article. With his contribution, we jump ahead six centuries to see how later anthologists made sense of the classical literary tradition. Clifford's network analysis of the contents of sixteenth-

and seventeenth-century “ancient-style prose” (*guwen* 古文) anthologies argues that such collections represent successive attempts to overturn the canon of model examination essays. He offers a new take on Ming literary history by demonstrating that the debate over prose was organized not around an imitative-expressionist dichotomy but around proposals of new canons entirely: Qin-Han dynasty prose, transdynastic prose, and *xiaopin* 小品 (informal essays). The last of these, which more strongly emphasized the contributions of women, was purely an invention of seventeenth-century printers and had no precedent in earlier eras. Like many articles in this special issue, Clifford uses digital methods to provide a broad framework for his analysis and then digs deep into his source material (here, prefaces to the anthologies) to develop that framework into a strong thesis. In so doing, he provides a new model for the ways that Ming anthologists collected, sorted, and debated over the increasingly unwieldy classical literary tradition.

Our final article, by Huang Yi-long and Bingyu Zheng, also grapples with the enormity of the classical Chinese textual tradition. It introduces to an English-language audience a method Huang developed over several decades called “electronic textual research” (*e-kaoku*, abbreviated ETR) that uses digital tools to advance traditional philological inquiry. Huang and Zheng demonstrate the usefulness of this method to the study of eighteenth-century literature by uncovering obscure literary allusions used in poems and by determining identities and relationships of people in the social circle of Cao Xueqin 曹雪芹 (c. 1715–63), author of the celebrated novel *Dream of the Red Chamber* (*Honglou meng* 紅樓夢). In their painstakingly researched examples, Huang and Zheng show that digital tools allow for greater depth as well as greater breadth of analysis. ETR is proof that digital literary studies is not limited to distant reading; it enhances close reading, too.

Huang and Zheng’s article, which demonstrates the continuity of digital technology with philological methods, nicely summarizes the general contributions of this special issue. Rather than focus on developing new tools for their own sake, these articles emphasize the way new tools and methods can support and extend long-standing practices of humanistic inquiry, such as the interpretation of classical literature in the contexts of its production, circulation, and reception. In this way, we envision a future in which the digital humanities have been normalized. We predict that digital sources and methods will no longer constitute their own field; instead, they will become part of a methodological tool kit familiar to any humanist. New generations of sinologists, who will learn programming languages alongside Japanese and French, will employ word vectors and network graphs as approaches to Chinese literature in conjunction with close reading and manuscript analysis. There will not be a divide between

“digital” and other scholars. Articles will primarily be judged by the content of argument, not the medium of their analysis. The future of the digital humanities, in short, is their own erasure.



THOMAS J. MAZANEC 余泰明  
University of California, Santa Barbara  
mazanec@ucsb.edu

JEFFREY R. THARSEN 康森傑  
University of Chicago  
tharsen@uchicago.edu

JING CHEN 陳婧  
University of Illinois, Urbana-Champaign  
chen125@illinois.edu

### Acknowledgments

The articles collected in this special issue were first presented February 9–10, 2018, at the University of California, Santa Barbara, for the conference “Patterns and Networks in Classical Chinese Literature: Notes from the Digital Frontier.” We would like to thank the University of California, Santa Barbara, for hosting us and for the generous support of its Interdisciplinary Humanities Center, College of Letters and Science, Humanities and Fine Arts, Center for Taiwan Studies, East Asia Center, Center for Information Technology and Society, and Departments of East Asian Languages and Cultural Studies, Comparative Literature, Linguistics, and History, as well as from the Forum on Chinese Poetic Culture.

As guest editors of this special issue, we express our gratitude to the anonymous readers for every article in this issue, who have provided constructive and detailed suggestions and comments. We also thank Xiaohui Zhang for providing editorial assistance for this special issue.

Finally, we would like to thank the general editors, Zong-qi Cai and Yuan Xingpei, for their tireless efforts and for giving us the opportunity to put this collection together.

### Notes

1. On Qing-dynasty *kaozheng*, see Elman, *From Philosophy to Philology*, esp. 177–98 for *kaozheng* scholars’ use of data. On data-driven literary studies in the early twentieth century, see Detwyler, “Aesthetics of Information,” 39–106. In recent years, historians of China (not literary scholars) have been at the forefront of data-driven humanities research. See, e.g., Hilde De Weerd’s monograph *Information, Territory, and Networks*, as well as the many symposia she has convened and tools she has helped develop.
2. For distant reading, see Underwood, “Genealogy of Distant Reading”; for stylometry, see Mosteller and Wallace, *Inference and Disputed Authorship*; and for digital prosopography and geography, see Bradley and Short, “Texts into Databases.”
3. On the long evolution of the modern, documentarian footnote, see Grafton, *Footnote*.
4. See, e.g., Barthes, “Style and Its Image.”

**References**

- Barthes, Roland. "Style and Its Image." In *Literary Style: A Symposium*, edited by Seymour Chapman, 3–10. London: Oxford University Press, 1971.
- Bradley, John, and Harold Short. "Texts into Databases: The Evolving Field of New-Style Propopography." *Literary and Linguistic Computing* 20, suppl. (2005): 3–24.
- De Weerd, Hilde. *Information, Territory, and Networks: The Crisis and Maintenance of Empire in Song China*. Cambridge, MA: Harvard University Asia Center, 2016.
- Detwyler, Anatoly. "The Aesthetics of Information in Modern Chinese Literary Culture, 1919–1949." PhD diss., Columbia University, 2015.
- Elman, Benjamin A. *From Philosophy to Philology: Intellectual and Social Aspects of Change in Late Imperial China*. Cambridge, MA: Council on East Asian Studies, Harvard University, 1984.
- Grafton, Anthony. *The Footnote: A Curious History*. Cambridge, MA: Harvard University Press, 1997.
- Mosteller, Frederick, and David L. Wallace. *Inference and Disputed Authorship: The Federalist*. Reading, MA: Addison-Wesley, 1964.
- Underwood, Ted. "A Genealogy of Distant Reading." *Digital Humanities Quarterly* 11, no. 2 (2017). [www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html](http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html).