

# Exploring Lives of China Ming-Qing Female Poets: Collaboration of the “Ming-Qing Women’s Writings Digital Archive and Database” and the “China Biographical Database”

HU Jia-jia

Beijing Normal University, Beijing, China

Based on the “Ming-Qing Women’s Writings (MQWW), a digital archive and database” project and the “China Biographical Database (CBDB)” project, which have been collaborating since 2008, this paper uses different digital methods to analyse the Ming-Qing women poets’ lives, such as the themes of their poems, their geographical distribution, and their social networks. The aim of this paper is to show researchers various prospects for the integration of digital humanities projects.

*Keywords:* China Ming-Qing Female poets, database, MQWW, CBDB, GIS, social network

## Introduction

Recent decades have witnessed strong interest in digital resources and computer tools among scholars, researchers, and students in North America, Europe, Asia, and elsewhere in the world. Numerous digital archives and databases, as well as various software applications, are being developed to aid research in the fields of history, literature, linguistics, philology, anthropology, archaeology, and others. These digital projects are built under the guidance of different methodologies and aim at various objectives. However, more and more usages exceeding their original designs are being explored through the collaboration and integration of these digital projects to obtain unexpected perceptions and views of humanities research. This paper is an attempt to gain a fresh view of Chinese female poets who lived during the Ming and Qing Dynasties (1368-1911) by exploring two digital projects in conjunction: the “Ming-Qing Women’s Writings (MQWW)” project (McGill Library Digital Initiatives, 2017) and the “China Biographical Database (CBDB)” project (Harvard University, Academia Sinica, and Peking University, 2018).

The MQWW designed and implemented by the McGill Library Digital Initiatives team is dedicated to the digitization of collections of writings by women held in the libraries participating in the project. These collections were mainly published during the Ming (1368-1644) and Qing (1644-1911) dynasties, and some were published in the Republican period (1911-1949).

The CBDB originates with the work of Robert M. Hartwell (1932-1996) (Peter, 1999). The development of the CBDB now is a joint project of Fairbank Centre for Chinese studies at Harvard University, Institute of History and Philology of Academia Sinica, and Centre for Research on Ancient Chinese History at Peking University.

Both the MQWW and CBDB projects allow free downloading of their data in Microsoft Access format and encourage innovative analysis of their data. The two on-going projects keeps updating yearly. As of the writing of this paper, the latest Microsoft Access version of the MQWW was updated in December 2016 and the CBDB was updated in April 2017.

The CBDB project team has been collaborating with their counterparts at the MQWW project since 2008 (Lik, 2016). The first and basic step for the CBDB to incorporate biographical data from the MQWW into its database is disambiguation, i.e., ensuring that each character in the MQWW has only one unique identification (ID) in the CBDB database. Considering there may be hundreds of people sharing the same name across the thousands of years of Chinese historical records, the disambiguation work needs significant careful textual research work. Fortunately, the data structure of the CBDB helps the disambiguation process be more efficient and effective. To assist with historical social network analysis, CBDB records not only personal information but also types of relationships between persons, such as kinship, friendship, partnership, marital ties, and alumni. There are over 900 pairs of family and social relations coded in the CBDB, which are classified into 10 general categories (Peter, 2015). As a result, each person can be identified by his/her position in the social network. Moreover, as the amount of data increases, the relationships become more numerous and disambiguation becomes easier. Therefore, each person appearing in the MQWW online database has a permanent system ID in the CBDB, which enables a user of either database to consult the other one

However, each database has its own focus on personal information. The MQWW lists details of individual collections authored by the person, as well as remarks and biographies on the person recorded in its database, and every list item has a link to the corresponding digitalized text material, whereas the CBDB elaborates upon each person's kinship relations as well as social relations. Based on the two digital projects, this paper conducts a set of analyses on Ming-Qing women poets from three aspects: writings, geographic distribution, and social network (Charles, 1998).

### **Favourite Poem Themes of Ming-Qing Female Poets**

The main focus of the MQWW project is on the valuable corpus of texts for research on Chinese women's literature, history, and culture of the Ming and Qing periods (Grace, 2010). It now contains 269 collections that can be searched using many access points, including author name, title of collection, and poetic form. Most works in those collections are poems (including traditional Chinese poetry genres: *Shi* 詩, *Ci* 詞, *Qu* 曲). Figure 1 shows a simple example of corpus analysis on the MQWW. It is a statistical analysis of 34,675 poem titles in the form of a word cloud to give an intuitive view of the favourite themes of female writers during the Ming and Qing periods.

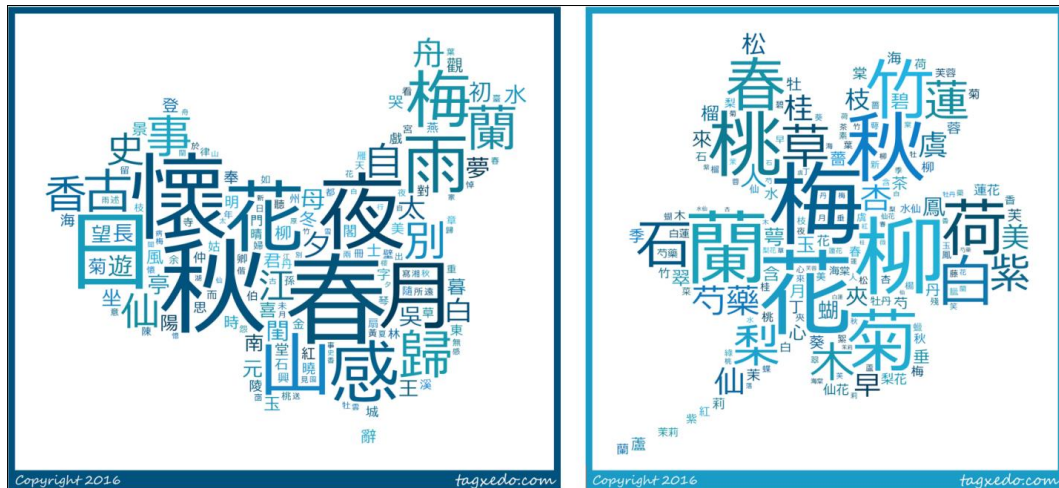


Figure 1. Two word clouds of 34,675 poem titles from the MQWW.

The left part of Figure 1 shows that the most frequently used words in 34,675 poem titles from the MQWW are “秋 (autumn),” “春 (spring),” “夜 (night),” “花 (flower),” “月 (moon),” “雨 (rain),” “懷 (miss),” and “感 (feel).” The right part shows a deep dive into female poets’ favourite flora. In addition to “梅 (plum blossoms),” “蘭 (orchid),” “竹 (bamboo),” and “菊 (chrysanthemum),” which are regarded by Chinese literati as “花中四君子 (four Junzi in flowers).” The “柳 (willow),” “荷 (lotus),” “桃 (peach blossoms),” and “梨 (pear blossoms)” also frequently appeared in titles by female poets. What is more interesting, “白 (white)” and “紫 (purple)” are the colours most preferred by female poets. These two word clouds were generated at tagxedo.com. A stop word list is used to skin Chinese function words, including numerals, measure words, prepositions, and conjunctions, such as “二 (two),” “首 (piece),” and “和 (and).”

As many critics have noted, this simple analysis is only a description, not an explanation. In the author’s defence, digital methods can provide some intriguing views that extend beyond paper research and offer information in a faster and perhaps a more beautiful way. So, why are white and purple more frequently used than other colours in female-authored poems on flowers? Are these colours also loved by male poets? What kind of spirit do they represent in the cultural values of educated women during the later Chinese imperial period? To answer these questions may require more queries, statistics, and comparisons by computers as well as the creative mental work of experts.

### Geographic Distribution of Ming-Qing Female Poets

For the sake of seeing patterns of time and/or space in large amounts of data, the CBDB has elaborated upon the temporal and spatial information regarding people when incorporating biographical data from the MQWW into its data structure. In the CBDB, almost every address from different historical periods corresponds to a pair of latitude and longitude data for use in a GIS, which is the result of years of hard work by the China Historical GIS (CHGIS) project (see <http://www.fas.harvard.edu/~chgis/>). The MQWW database contains information on over 5,000 poets and other writers (including men and women). However, only 2,512 female poets who are believed to have lived in the Ming and Qing periods have native place information that can be transformed into geographic coordinates according to the CHGIS. Figure 2 shows a GIS analysis map of the 2,512 female poets. The data are concentrated in east-central China, especially the Yangtze River Delta and



its surrounding area, as Figure 3 shows. However, this is only a static distribution, which can be transformed to a dynamic evolution process by adding temporal information.

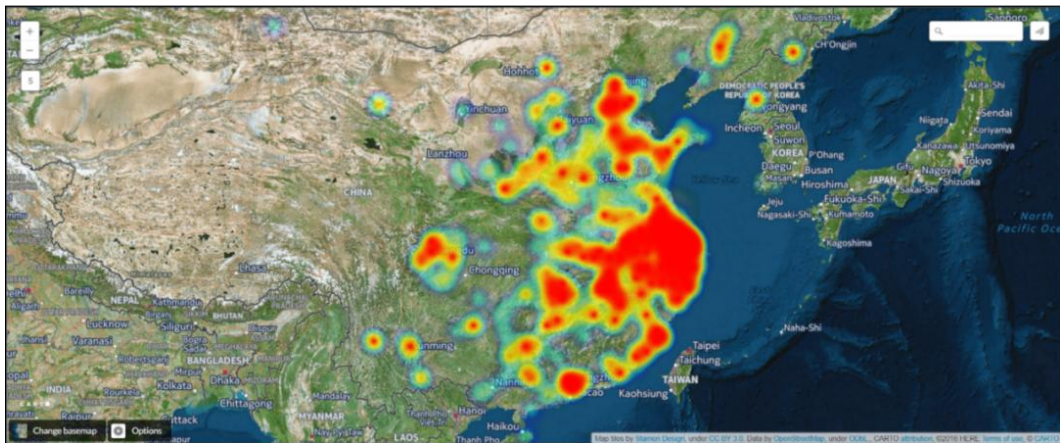


Figure 2. Geographical distribution of 2,512 female poets during the Ming and Qing periods (a “heat map” layer in CartoDB).

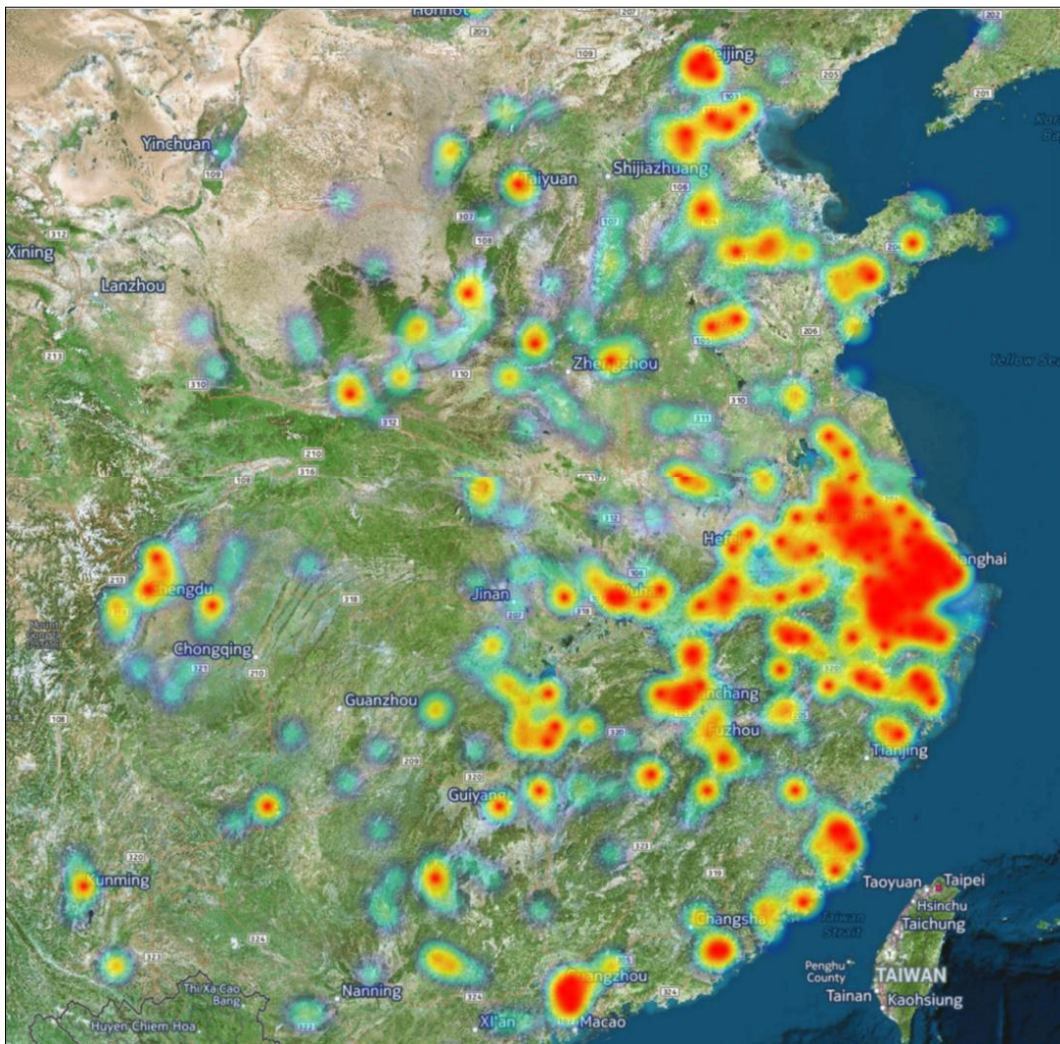


Figure 3. Scaled-up map of Figure 2 focused on east-central China.



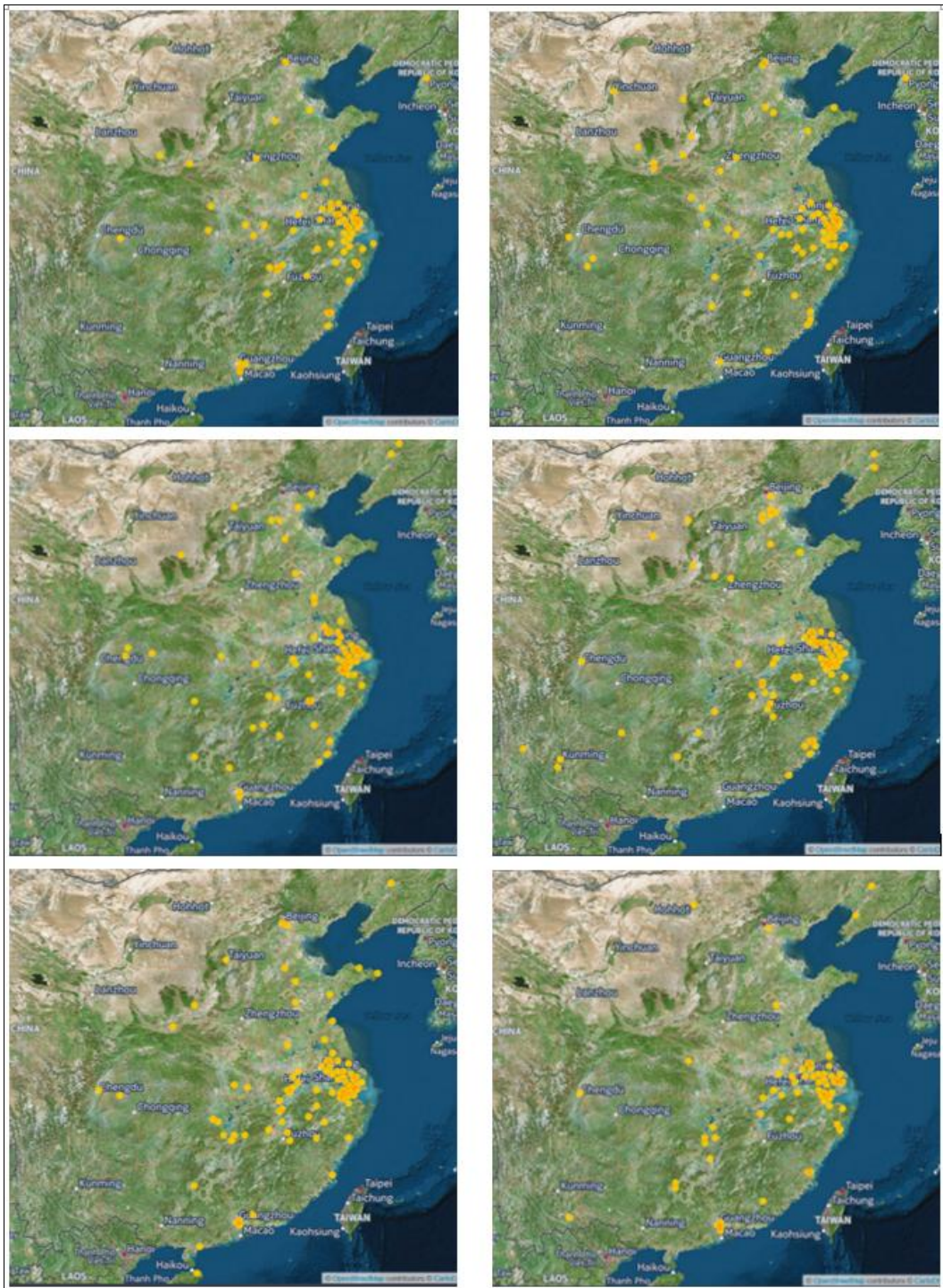


Figure 4. Change in geographical distribution of 1,178 female poets during the Ming and Qing periods (a “torque” layer in CartoDB). The change order is from left to right and from top to bottom.

In addition to the “birth year” and “death year” for each person, the CBDB introduced an artificial attribute called “index year” for computational purposes. The “index year” is a single year value calculated by complex rules to locate a person as a point in time (Michael, 2015). Figure 4 shows how the geographical distribution of 1,178 female poets changes according to their “index year.” It is noted here that some of the

2,512 female poets in Figure 2 do not have precise “index year” values in the CBDB, and are thus removed in Figure 4. To illustrate the dynamic process more clearly on paper, the change is divided into six phrases corresponding to different reign periods: (1) before late Ming (including 206 women); (2) late Ming to early Qing (including 179 women); (3) Qing Kangxi 康熙 to Yongzheng 雍正 (including 176 women); (4) Qing Qianlong 乾隆 (including 199 women); (5) Qing Jiaqing 嘉慶 to Daoguang 道光 (including 242 women); and (6) Qing Xianfeng 咸豐 and after (including 176 women). This division considers both the length of time and the number of women whose “index year” are located in the corresponding period, so as to make different phases more comparable.

### Family Relations among Ming-Qing Female Poets

In addition to listing all direct relations of a person, the CBDB also provides powerful tools to explore social network-defined kinship ties and other social relations. Its “Query Social Networks” tool allows the user to select the particular forms of social association to investigate and to set the range of years to consider.

Figure 5 shows the search results of exploring the social network of Gan-Lirou (甘立嫫) (a female poet in the MQWW). There is a list of pairs of people who are connected by types of association (including kinship and non-kinship type), which can be output to three different social network analysis (SNA) programs: UCInet, Pajek, and Gephi.

The screenshot shows the 'Looking at Social Networks' application window. The main area displays a table of social network relationships for Gan-Lirou. The table has columns for Name, 姓名, Linked to, 社會關係人姓, Kin/Nc, and Link. The data is as follows:

Name	姓名	Linked to	社會關係人姓	Kin/Nc	Link
Gan Lirou	甘立嫫	Gan He	甘禾	K	F
Gan Lirou	甘立嫫	Gan Rulai	甘汝來	K	FF
Gan Lirou	甘立嫫	Xu Yueiv	徐曰呂	K	H
Gan Lirou	甘立嫫	Gan Yue'e	甘月娥	K	Z+
Gan Lirou	甘立嫫	Gan Ligong	甘立功	K	B+
Gan Lirou	甘立嫫	Gan Liyan	甘立言	K	B+
Gan Lirou	甘立嫫	Gan Liyou	甘立猷	K	B-
Gan Lirou	甘立嫫	Xu Binian	徐必念	K	S
Gan Lirou	甘立嫫	Xu Xintian	徐心田	N	Postface of book written by
Gan Lirou	甘立嫫	Xu Xintian	徐心田	K	S2
Gan Lirou	甘立嫫	Xu Weilun	徐維倫	K	HF
Gan Lirou	甘立嫫	Song Rong	宋蓉	N	Preface of book by
Gan Lirou	甘立嫫	Wang Ruohong	王若閏	N	Preface of book by
Gan Lirou	甘立嫫	Liu Binshi	劉彬士	N	Epitaph written by
Gan Lirou	甘立嫫	Ru Fen	茹焚	N	Portrait eulogy written by
Gan Lirou	甘立嫫	Gu Gao	顧皋	N	Epitaph written by
Gan He	甘禾	Gan Rulai	甘汝來	K	F
Gan He	甘禾	Gan Yue'e	甘月娥	K	D
Gan He	甘禾	Gan Ligong	甘立功	K	S
Xu Yueiv	徐曰呂	Xu Xintian	徐心田	K	S2

The interface also includes a sidebar with search filters (e.g., 'Gan Lirou', '甘立嫫', 'From: 1700', 'To: 1900') and a bottom section with export options (Save to UCInet, Save to Pajek, Save to Gephi, etc.).

Figure 5. Search result of Gan-Lirou's social network relationships.

Figure 6 visualizes the social network of Figure 5 using Gephi. Red node means female, blue node means male. Pink edge means kinship relation, green edge means non-kinship relation. It shows not only the female poet Gan-Lirou's (甘立嫫) family and friends but also her family's friends as well as her friends' families and friends' friends.



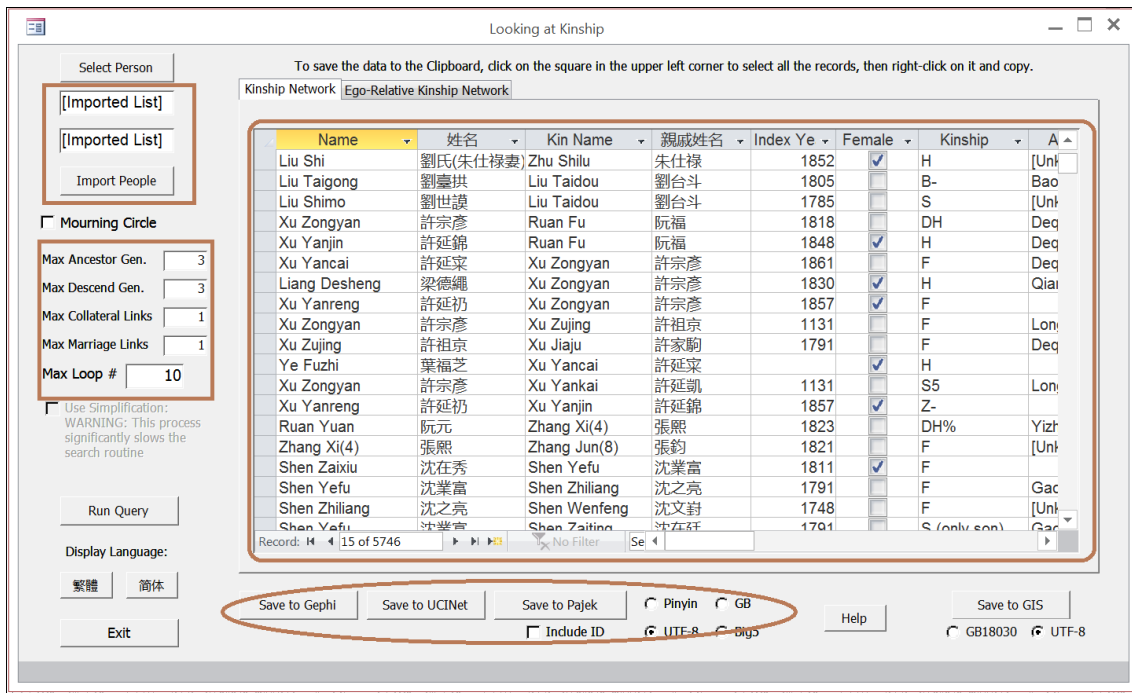


Figure 7. Exploring the kinship network of 2,512 female poets in the MQWW.

The search result lists all the kinship relations meeting the set criteria of imported people. In this case, it shows 5,746 family members of 2,512 MQWW female authors (including themselves) who have records in the CBDB. The result can be output to SNA programmes for further analysis. Using component analysis in Gephi, the 5,746 family members of the 2,512 MQWW female authors can be divided into 1,721 connected components. A component of an undirected social network is a subgraph that is disconnected from other nodes but in which any two nodes are connected to each other by paths. In this case, a component means an alliance of several families through blood or marriage.

Figure 8 shows such a component, composed of 24 female poets in the MQWW and their family members in the CBDB. There are 90 nodes and 161 edges in total in this component. The 90 nodes are divided into three kinds: male members (blue nodes), female poets (red nodes), and other female members (carmine nodes). In this paper, a female poet means she must have at least one poem collected in the MQWW. Some women who have been mentioned but have no works in the MQWW are not regarded as poets for the purposes of this paper. The size of a node is proportional to its degree. The degree of a node here is the number of edges connected to it. In this case, Wang-Duanshu (王端淑) has the most direct connections with other family members. Additionally, there are three colours of edges, which represent three kinds of relations between family members: orange means relations between peers (like brothers, sisters, cousins, or husband and wife), green means relations between parents and children (like mother and daughter, father and son-in-law, uncle and nephew), and purple means relations between grandparents and grandchildren.



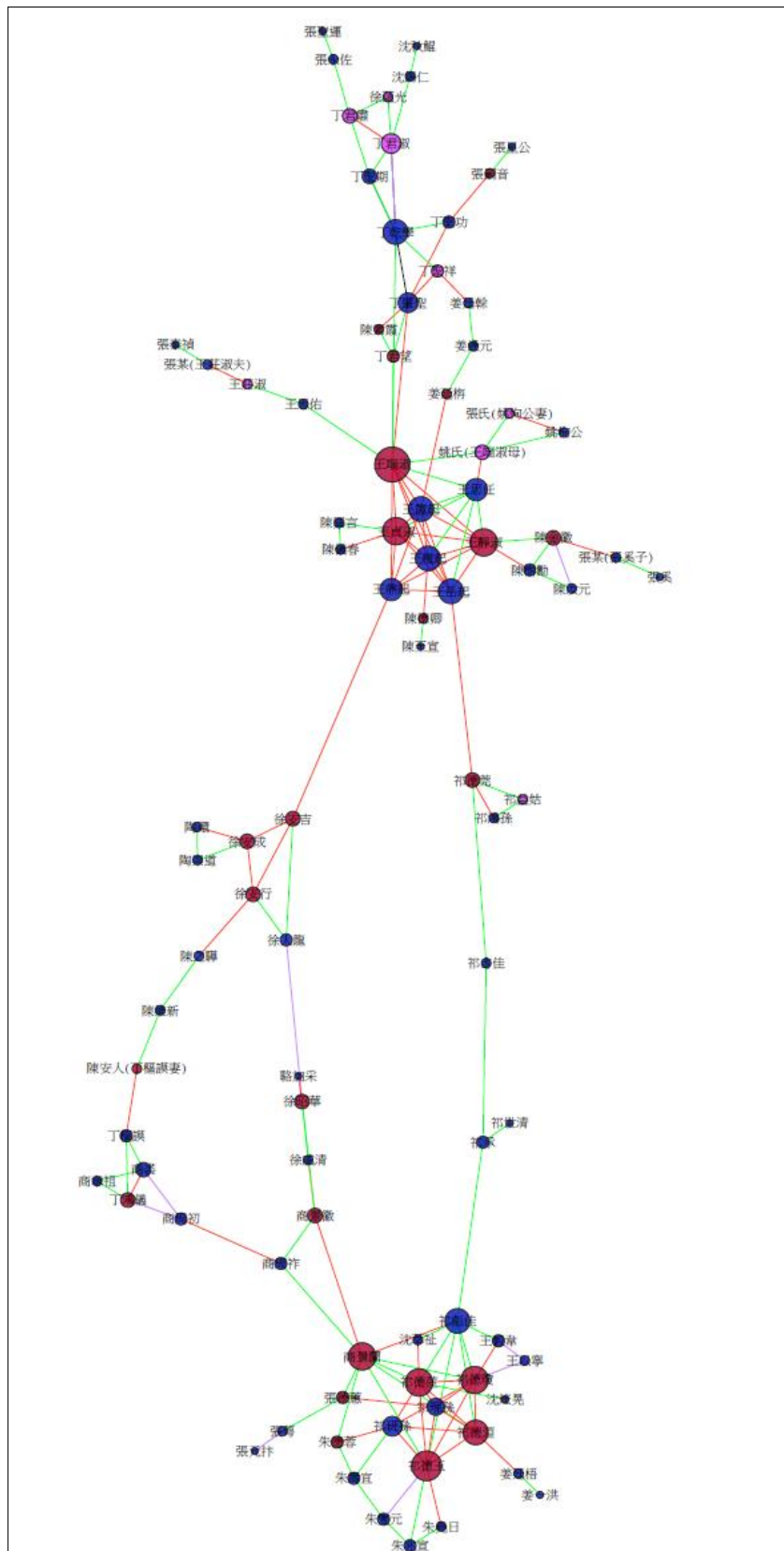


Figure 8. An alliance of families including 24 female poets.

This example can be viewed as a local projection of the multi-generation marital circle in Chinese literati families. It is interesting to ask whether the families who valued women’s education were more likely to marry or if the marriages of literati families encouraged literary communication among their female members.

### Literary Relationships Between Female Poets

In addition to family relationships, there are also literary relationships between poets based on the exchange of poems or correspondence. These literary relationships can be searched in the CBDB with its “Query Social Networks” tool by selecting types of relations as “writings” (Michael, 2015). The CBDB could search all direct relations described under the “writings” type between any two female authors provided in the imported list of people, as Figure 9 shows. What is more, the result can be output with peoples’ geographic information for GIS analysis.

Figure 10 shows a map of Ming-Qing female poets from different regions connected by their direct literary exchange. This map is generated in QGIS. It seems to indicate that the Jiangnan 江南 (south of the Yangtze River) area was a centre of women literary communications during the Ming and Qing dynasties.

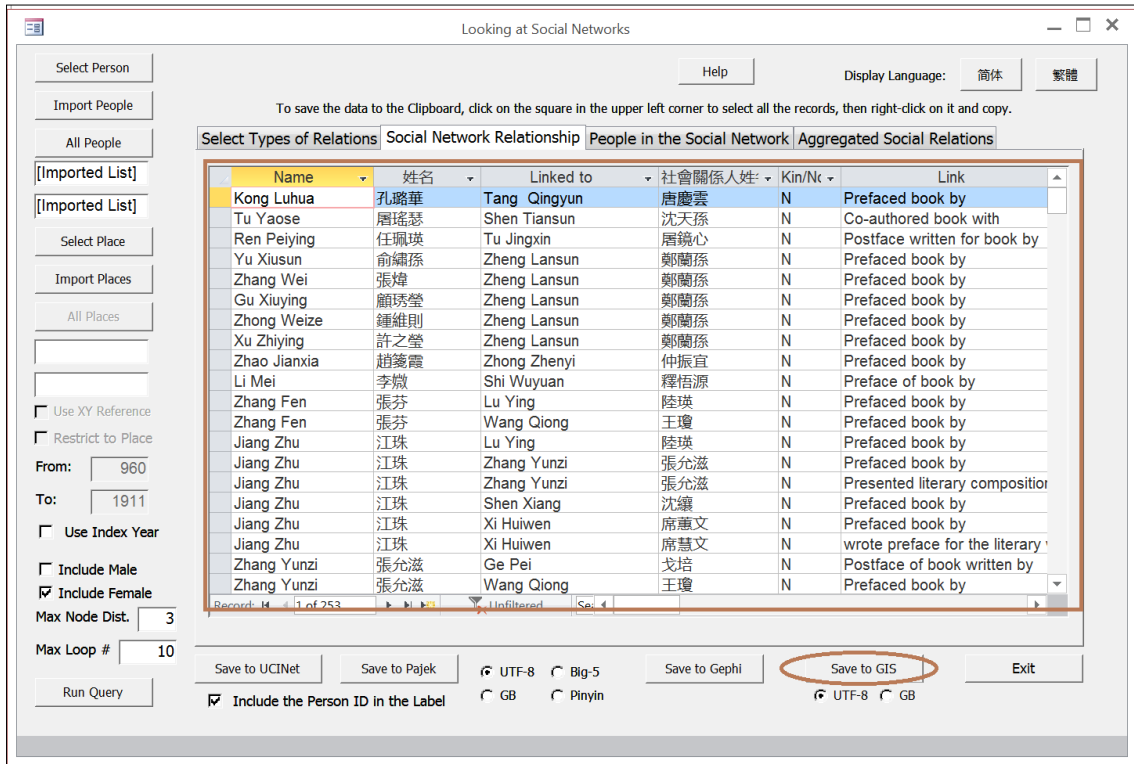


Figure 9. Exploring direct literary relationships of 2,512 female poets in the MQWW.

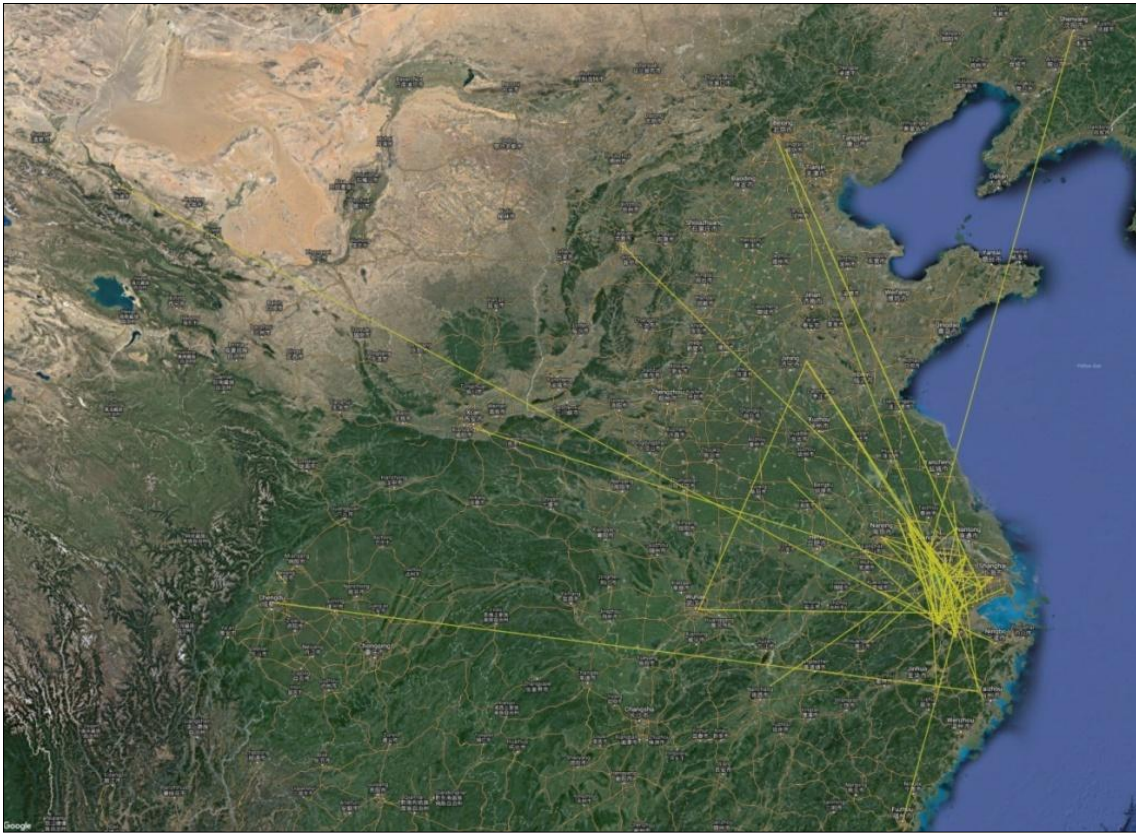


Figure 10. Literary relationships among female authors of different regions.

## Conclusion

Although this paper conducted a set of analyses on Ming-Qing women writers, its purpose is not to enrich related research. This paper is an example to show researchers what can be expected by integrating different digital humanities projects and methods. With the increasing number of digital projects, collaboration is becoming more and more important because “1+1 > 2”. Sometimes, interesting questions beyond the expectations of both users and designers can be asked based on computer descriptions. In addition, of course, to understand and explain those descriptions requires more knowledge and deeper insight. Meanwhile, it is urgent for researchers to prepare themselves to adapt to technological advances and changing paradigms in digital times.

## References

- Fong, G. S., & Widmer, E. (2010). *The inner quarters and beyond: Women writers from Ming through Qing*. Leiden Boston: Brill.
- Harvard University, Academia Sinica, and Peking University, China Biographical Database. (2018). Retrieved from <https://projects.iq.harvard.edu/cbdb>
- Lik, H. T. (2016). *The digital humanities as an emerging field in China*. Retrieved from <https://blogs.nottingham.ac.uk/chinapolicyinstitute/2016/06/13/the-digital-humanities-as-an-emerging-field-in-china>
- McGill Library Digital Initiatives. (2017). *Ming Qing women's writings*. Retrieved from <http://digital.library.mcgill.ca/mingqing>
- Michael, A. F. (2015). *The China biographical database user's guide*. Retrieved from [http://projects.iq.harvard.edu/files/cbdb/files/cbdb\\_users\\_guide.pdf](http://projects.iq.harvard.edu/files/cbdb/files/cbdb_users_guide.pdf)



- Peter, K. B. (1999). *The late Robert M. Hartwell Chinese historical studies, Ltd. Software*. Retrieved from <http://pnclink.org/annual/annual1999/1999pdf/bol.pdf>
- Peter, K. B., Michael, A. F., Chen, S., Chen, S. P., & Wang, H. S. (2015). *General introduction to the CBDB project*. Retrieved from <http://projects.iq.harvard.edu/cbdb/presentations-and-papers>
- Wetherell, C. (1998). Historical social network analysis. *International Review of Social History*, 43, Supplement.