



图书馆论坛

Library Tribune

ISSN 1002-1167, CN 44-1306/G2

## 《图书馆论坛》网络首发论文

题目：明代书信数字化研究的新尝试——以哈佛大学 CSA 明代书信计划为例  
作者：陈士银  
收稿日期：2022-11-01  
网络首发日期：2023-02-02  
引用格式：陈士银. 明代书信数字化研究的新尝试——以哈佛大学 CSA 明代书信计划为例[J/OL]. 图书馆论坛.  
<https://kns.cnki.net/kcms/detail//44.1306.G2.20230201.1346.001.html>



**网络首发：**在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认：**纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

\*本文系国家社科基金后期资助项目“明初礼制变革研究”（项目编号：20FZSB007）的阶段性成果。

## 明代书信数字化研究的新尝试 ——以哈佛大学 CSA 明代书信计划为例\*

陈士银

**摘要** 基于社会关系网络数据众包平台 CSA (Crowdsourcing System for Association Data) 以及中国历代人物传记资料库 CBDB (China Biographical Database) 的明代书信计划 (Ming Letters Project) 是向学界提供通过书信作者、收信人、书信名称及其所属文集进行检索的数据平台。自 2020 年成立以来, 该计划已经收录 1,650 位明代人物, 超过 52,399 封书信信息, 另有约 50,000 封书信信息待收录。通过整理书信往来人物, 建立其与 CBDB 数据库的关联, 学者可以批量利用这些书信, 展开探索相关人物的身份信息, 包括姓名、别名、籍贯、入仕、官名、著述、亲属关系与社交网络等, 丰富对明代历史人物及其社会关系的研究。

**关键词** 明代书信; 数字化; 明代书信计划; 中国历代人物传记资料库

## A New Attempt of Digitalization Study about Ming Letters: Based on Harvard University CSA Ming Letters Project

CHEN Shiyin

**Abstract:** Based on CSA (Crowdsourcing System for Association Data) and CBDB (China Biographical Database), Ming Letters Project is a newly-established data platform providing service from retrieval items, like the author, addressee, title of letters and books. This project has involved 52,399 letters from 1,650 Ming people, since 2020, with another 50,000 letters remained to be involved. Researchers from various academic fields can expand further study about the identity information of people in Ming dynasty, for instance, name, surname, imperial examination, official rank, book, relative and social network, with the relevance between MLP and CBDB.

**Keywords:** Ming letters; digitalization; MLP; CBDB

### 0 前言

近年数字人文作为新型研究方法, 通过学科交叉的视角、可视化的手段, 被广泛运用到文学、历史学、哲学、考古学、社会学、音乐、艺术等领域。数字化研究可以大规模、批量化、形象化地分析研究对象<sup>[1][2]</sup>, 促进对研究对象涵盖范围、影响规模以及社交网络的认识, 并从中抽绎新的结论, 获得新的启示。以明代书信为例, 学界目前的研究集中在收藏、展览、鉴赏以及文学、史学等方面。举其大者, 即有石守谦、杨儒宾编纂出版的《明代名贤尺牍集》<sup>[3]</sup>, 金桂台通过明代文学书信对性灵论的探讨<sup>[4]</sup>, 史小军、梁娟对书信体文论发展<sup>[5]</sup>、明代文人书信体文论的研究等<sup>[6]</sup>。迄今为止, 尚未看到国内外学者借助数字化的视角介入明代书信研究。依托传统的历史文献研究方法, 学者往往通过研究个体或者某类人的书信, 分析其中的互动往来, 厘清其中的人物关系。这种研究方法的优势和不足一目了然: 优势在于对个案的深入研究, 不足在于无法兼顾书信以及通信人物的巨大体量。以《明代名贤尺牍集》为例, 该书虽然有利于书法鉴赏, 裨益文史研究<sup>[7]</sup>, 但仅收录 216 位明人, 252 通尺牍。根据哈佛大学与我馆台湾汉学研究中心的合作, 至少有将近 10 万封书信需要研究, 是《明代名贤尺牍集》收录书信数量的 400 倍。如果采用传统的团队合作以及研究方法, 将 10 万封书信分批研究, 也会耗费大量的时间和精力, 劳而少功。不宁唯是, 这种传统的研究成果也无法向读者提供检索、交互功能, 大大降低国内外学界互相交流、学习的可能。有鉴于此, 笔者通过哈佛大学现有的 CBDB 项目, 将明代书信涉及的通信人物作为突破点, 引入数字化研究的思路, 换一种角度研究历史问题。通过厘清通信双方,

既可以增益 CBDB 数据库中明代历史人物、著作等方面的数据,又可以搭建明代书信检索平台,方便学界在此基础上,进一步开展对明代书信的研究。

## 1 为什么要启动哈佛大学 CSA 明代书信计划?

哈佛大学 CSA 明代书信计划的启动既非偶然,也非平地起高楼,而是依托 CBDB 的优势项目。2006 年哈佛大学费正清中国研究中心与北京大学中国古代史研究所、我国台湾“中央”研究院历史语言研究所基于美国汉学家郝若贝(Robert Hartwell, 1932—1996)的数据与程序,建立中国历代人物传记资料库项目 CBDB (China Biographical Database project, CBDB)。通过搜集和整理中国历代人物,该数据库旨在系统性地收录中国历史上的传记数据,并将其公诸学术之用。截至 2022 年 8 月,CBDB 已经收录 521,442 名中国历史人物<sup>[8]</sup>。目前该数据库的负责人为哈佛大学东亚语言和文明系教授包弼德(Peter K. Bol)。包弼德教授主要研究方向为中国中古时期的思想文化,著有《历史上的理学》<sup>[9]</sup>《斯文:唐宋思想的转型》<sup>[10]</sup>等。在他带领下,CBDB 数据库从唐宋时期开始,工作范围陆续延展到明清时期。故此,大量搜集明代人物便成为 CBDB 面临的一个重要方向。在这种背景下,作为 CBDB 的子项目,CSA 明代书信计划应运而生。

就哈佛大学 CBDB 的项目进展而言,唐宋时期的历史人物资料相对丰富,而明清时期的历史人物资料相对薄弱,亟待发掘。那么,该如何推进明清时期历史人物的数字化研究?以明代为例,近年来,海内外学界陆续出版了人物、著作资料索引等方面的著作,有关机构也建立了一些数据资料库,尤其在文集、方志、谱牒、书画等方面均有成果可以借鉴。由于已经出版的著作,尤其是研究性质的作品,涉及版权问题,所以哈佛大学 CBDB 团队着手磋商处理,而磋商过程则需要一个周期。至于依托古籍资料的各类数据库资源,则各有特色,其中包含的人物传记资料需要分类整理。整理的前提同样需要多方协商合作,才能获得基础的批量化数据,而非从零开始,逐条增加。

就明代书信本身的应用研究而言,海内外学界迄今尚无类似的数据资料处理平台,而合作方我国台湾汉学研究中心慨允分享已有资料。故此,2020 年通过与我国台湾汉学研究中心的合作,哈佛大学 CBDB 团队获得大量明代书信资料,随之对涉及的历史人物展开分析。为了更好地利用这些资料,CBDB 团队创立用于识别社会关系网络数据的众包平台 CSA (Crowdsourcing System for Association Data)。在此基础上,CBDB 项目启动 CSA 明代书信计划(Ming Letters Project, MLP),利用自身数据库的优势,并组织海内外高校和研究机构文学、历史学、哲学、计算机、社交网络等方面的专家学者,激活数量庞大的明代书信,最大限度地挖掘、利用其中的身份信息,建立明代历史人物(特别是明代中后期人物)的社会关系网络。截至 2022 年 6 月,该计划已经收录 1,650 位明代人物,超过 52,399 封书信信息<sup>[11]</sup>,另有约 50,000 封书信信息待收录。

## 2 数字化研究与历史考证相结合: CSA 明代书信计划的工作方法

### 2.1 辨别书信,完成编码

CSA 明代书信计划的思路即是借助数字计算,将明代书信和历史人物数字化。书信的辨识与编码是 MLP 的基础工作,这不仅直接决定后期 Web 平台操作的工作体量,也是 MLP 子项目接入 CBDB 的重要保障。面对数量众多的明人文集及其包含的各类文章,MLP 首先要从这些文章中辨识并剥离出书信部分。辨别工作并不复杂,因为很多明人文集本身就按照体例分卷,而且关于书信的文章一般以“与某某书”“答某某书”“致某某书”等方式存在,辨识度较高。项目组成员试图利用按类统计的筛选方式加快剥离速度,但是明人书信的不少标题又出现随机性,比如有的标题仅仅出现一个人名,或者一个别称,甚至连续多封信件,从第二封信件开始只用“又”字标题。故此,辨别书信的工作依然需要人工完成。对于

不同版本的文集，MLP 优先采用卷数更大的版本作为操作版本，同时保留其他版本的数据信息。

剥离工作相对复杂，因为 MLP 既要保留这些书信的基本信息，比如文集的作者、书信名称、所在卷数以及文集名称，同时要对这些书信进行编码，方便后期接入 MLP 的 Web 工作平台。作为 CBDB 的子项目，MLP 在对明代书信进行编码时，既要考虑 MLP 的编码，也要兼顾 CBDB 已经存在的各种编码（如图 1）。不然，在后期操作的过程中，可能存在冲突，诱发数据紊乱。一旦完成有效的编码，后面的数据操作便可通过系统运行。这里的编码不仅包括书信作者、通信对象，还包括所属文集等信息。有的文集收录在 CBDB 系统之中，便可以自动匹配原来的著作编码，有的文集则需要利用系统重新编制。这些文集信息既可以保证数据平台在史料运用过程中的真实性，也方便后期数据库使用者的查证，以及根据文集的数据分布，展开进一步研究。



图 1 CBDB 代码表（局部）

## 2.2 考订收信人，撰写札记

数字人文研究的手段在于数字计算，基础则离不开人文研究，特别是考证功夫。编码之后，将书信包含的身份信息接入 MLP 的操作平台。MLP 并不处理书信的正文内容，而是从确定收信人的身份入手。这就需要来自文学、历史学、哲学等学科领域的专家学者发挥所长，根据极其有限的书信标题等信息确定收信人的身份。

一般而言，名声越大的作者，如王守仁、王世贞、申时行、张居正，通信对象的身份越容易锁定。对绝大多数声名有限的作者，确定通信对象则相对困难。另外，古人通信并不像今天收发信件、Email 那样，载明通信对象，而是经常称呼对方的字、号、别字、别号、他称、官职、官职别称等。即便是名人，一旦使用生僻的别名，尚且难以确定其人。至于一般历史人物的别名，就更难以匹配到具体的个人。

如何根据代称确定历史人物是团队开展研究遇到的一大难点。至于官职，明人极少直

接称呼对方官职的正式名称,而是用他称、俗称代替。比如,常见的六部长官用《周官》名称代替,用冢宰、司徒、宗伯、司马、司寇、司空代替吏部、户部、礼部、兵部、刑部、工部的尚书。如果是侍郎,则加“少”字,如少司空代指工部侍郎。至于地方督抚,则用督府代指总督,方伯代指巡抚,总戎代指总兵。至于地方官员,明人用大尹代指左、右布政使,另用大参、少参代指参政、参议。同样的“知府”一职,又会出现“太守”“郡守”“府尊”“太府”等代称。另外,明人还习惯用清戎代称同知,宪副代称按察副使,佥宪代称按察使司佥事,兵宪代称都指挥使司佥事等。通过对类似身份信息的整理、比对,学者可以借助 MLP 勾勒出明代历史人物书信中的“代称世界”,继续推进这一方面的研究。

各种名儒、名宦(特别是中央公卿、地方布政使以及总督、巡抚等官职),相对容易确定。知府以下的官员,史料记载有限,这就需要项目组学者发挥所长,借助出版的工具书以及建立的各种数据库(如美国哈佛大学 CBDB 数据库、北京爱如生公司开发的中国基本古籍库、方志库、别集库、谱牒库)穷尽所能,找出此人。如果不能找出,暂时阙疑,等待日后数据库的完善。在发现新的线索之后,再返回阙疑之处,补足身份信息。

另外,对通信作者提到的“众弟子”“族人”等信件,项目组即便保留数据,也无法从中提取有用信息,姑予搁置。换言之,本阶段的 MLP,乃至 CBDB,均以人名为工作的发端。如果某历史人物没有留下具体的名字或者别称,只是以“众人”“诸人”的笼统面目出现,则不在工作范围之内。

考订收信人是一项漫长且繁重的工作。为了平衡考订具体的通信人与推进工作进度,MLP 鼓励项目组成员根据个人研究的领域,采用“先易后难”的原则,批量确定同一作者、不同通信人的信息,提高效率。考订收信人信息的工作分为两组:一组初步确定收信人,二组对初步确定的收信人信息进行复核。考订过程中,为了提高工作的准确度,在保留具体史料出处(刻本、卷数等)的前提下,项目组成员会额外记录收信人的信息,与 CBDB 数据库进行比对,撰写考证札记,方便日后补入 CBDB 数据库之中。

### 2.3 补正人物,文字数据化

项目组成员在确定收信人的工作过程中,会处理大量新旧人物。旧人物是指已经收录在 CBDB 数据库中拥有人物编码及其相关信息的人物,新人物则是未收录其中的人物。对于新人物,需要从头做起,建立全新的身份档案。此外,如果需要补充旧人物的信息,也会将 MLP 的最新数据导入 CBDB。

随着工作推进,和 CBDB 一样,MLP 处理人物的数量逐渐上升,从几百人到几千人,进而几万人,难免会遇到很多重名的人物。对于这些重名的人物,MLP 会根据朝代、字号、籍贯等信息加以甄别。如果重名的两人或者多人系同一人物,则予以技术上的消歧。根据编码生成的时间顺序,以较早出现的编码替代稍后产生的编码,从而提高数据的精准度。根据考证札记的文字信息,MLP 生成线上操作的多个编辑页面,将这些可以充实 CBDB 的文字信息一一切割、分类,后期批量导入 CBDB 数据库。自成立以来,MLP 着重搜集和处理的人物信息如下:

姓、名、性别、字、号、别名、谥号、民族、传记地址、户籍地址、出生年份、死亡年份、入仕(中举人、中进士及其年份)、任职(先后任官及年份)、亲属关系(夫妻、兄弟、姐妹、父母、曾祖父母、舅甥、从兄弟等)、社会交往(师生、朋友、同年、同僚、上司/下属、推荐/被推荐、赠送诗文、作序/传/墓志铭、仇敌/被害等)、著作、文献出处等。

随着更多资料补入,这些信息分类也会相应调整。不难想见,如果能将明代 10 万封书信所包含的上述人物信息全部或者大部提取完毕,将大力推动学界对明代历史人物的研究。

除明人文集包含的书信之外,还有诸多散件,也具有重要的学术价值。对于这类书信,该如何处理?作为 CBDB 项目的组成部分,哈佛大学 CSA 明代书信计划同样不会设置项

目的完成周期,比如三到五年,或者外加一轮滚动期,而是持续运行。就像 CBDB 项目一样,从启动到引起海内外学界的广泛关注,该项目已经走过 16 个春秋。在资助方和合作方的共同努力下,该项目还会继续运行下去,不存在限期结项的问题。故此,明代书信计划在现阶段任务执行的过程中,也会注意对类似书信的搜集,比如王阳明、王世贞等人的散见书信。同时,对这些散见书信加以利用的前提在于补辑、缀合、辨伪等基础工作的完成,而类似工作并非专注于人物传记资料的 CSA 明代书信计划所擅长。鉴于项目团队有限的人力、物力,期待学界在这方面贡献出更多成果,也期待尽快将这方面书信资料的人物信息补入 CSA 明代书信计划之中,共同推动对明代书信资料的整合和利用。

### 3 明代书信数字化研究可以为学界带来什么?

#### 3.1 搭建数据平台,方便检索明代书信

持续发掘新的资料来源以及使用新的技术手段是促进学术向前发展的不竭动力,而哈佛大学 CSA 明代书信计划最突出的贡献就是提供一个全新的数据检索平台。该平台尽可能竭泽而渔地收录明代所有历史人物的书信,方便学者检索利用。基于不同的学科背景(比如文学、历史学、哲学、社会学、博物馆学、图书馆学等),不同的学者对明代书信都会有相应的研究视角与方法。就 CBDB 而言,哈佛大学 CSA 明代书信计划的特色正是基于数字化的手段,结合已有的数据优势,以数据库的形式参与并推动学界对明代书信的研究。自上线以来,MLP 处在持续更新之中,以确定这些书信的通信人为突破口。随着数据库的日益完善,学者通过繁体字检索书信作者、收信人、书信名称及其所属文集等信息,就可以看出相关人物的通信概况(如图 2 所示)<sup>[12]</sup>。根据这些概况,回到 CBDB 数据库中,又可以更大程度地发现此人的身份信息(如图 3 所示)。



图 2 CSA 明代书信计划数据库检索“王世贞”书信界面(局部)

王世貞

ADVANCED SEARCH

100 Ming Letters 20200311

2

Submit Proposals

Discard

100 Ming Letters 20200311

100 Ming Letters 20200311

Page 11

Page

| 行號   | 作品  | 編號 | 作者姓名 | 作者生年 | 人物ID  | 作品標題 | 關係類別 | 關係人姓名 | 關係人生年 | 關係人ID | 關係人生年 | 關係人生年 | 年 | 月 | 日 | 文體                 | 品級    | 非   | CSDB編號 | CSDB編號 | 備註 |
|------|-----|----|------|------|-------|------|------|-------|-------|-------|-------|-------|---|---|---|--------------------|-------|-----|--------|--------|----|
| 9895 | 王世貞 | 1  | 1526 | 1590 | 34717 | 李攀龍  | 朋友   | 李攀龍   | 1571  | 1514  | 1514  | 1570  |   |   |   | 貴州四庫第一卷七十四卷續編二百零七卷 | 57974 | 117 | 117001 | 3611   |    |
| 9896 | 王世貞 | 2  | 1526 | 1590 | 34717 | 徐子檢  | 朋友   | 徐子檢   | 1590  | 1517  | 1517  | 1578  |   |   |   | 貴州四庫第一卷七十四卷續編二百零七卷 | 57974 | 118 | 118001 | 3611   |    |
| 9897 | 王世貞 | 3  | 1526 | 1590 | 34717 | 董道玉  | 朋友   | 董道玉   | 1573  | 1525  | 1525  | 1593  |   |   |   | 貴州四庫第一卷七十四卷續編二百零七卷 | 57974 | 118 | 118002 | 3611   |    |
| 9898 | 王世貞 | 4  | 1526 | 1590 | 34717 | 莊敬王  | 朋友   | 莊敬王   | 1573  | 1525  | 1525  | 1593  |   |   |   | 貴州四庫第一卷七十四卷續編二百零七卷 | 57974 | 119 | 119001 | 3611   |    |
| 9899 | 王世貞 | 5  | 1526 | 1590 | 34717 | 李子德  | 朋友   | 李子德   | 14723 | 1525  | 1525  | 1560  |   |   |   | 貴州四庫第一卷七十四卷續編二百零七卷 | 57974 | 119 | 119002 | 3611   |    |
| 9900 | 王世貞 | 6  | 1526 | 1590 | 34717 | 李德甫  | 朋友   | 李德甫   | 12840 | 1514  | 1514  | 1583  |   |   |   | 貴州四庫第一卷七十四卷續編二百零七卷 | 57974 | 120 | 120001 | 3611   |    |

图 3 CBDB 检索“王世贞”页面(局部), 收录社会关系 318 条

配合 CBDB 的使用, 哈佛大学 CSA 明代书信计划提供的又不仅是一个数据库。与传统百科式、传记式的人物记载不同, MLP 侧重通过历史人物的姓名、别名、户籍地址、官名、入仕、著述、亲属以及社交机构等各种身份信息、社会关系, 还原此人的历史世界。学者可以按图索骥, 继续探讨和该历史人物有关的人物与事件。

### 3.2 检索资料的再处理, 通过可视化手段让人物研究更形象

截至 2021 年 12 月, CBDB 收录明代历史人物 215,755 个。伴随 MLP 的推进与补充, CBDB 收录的明代人物从数量上和信息丰富程度上还会大规模增加。明代历史人物的身份信息暴露越多, 就越有可能建立彼此之间的互动网络。

根据 MLP 项目的不完全收录, 截至 2022 年 5 月, 申时行发出的信件最多, 达到 1,803 封书信, 张居正次之, 为 822 封。鉴于申、张二人都担任过内阁首辅, 可以通过书信的角度分析顶层官员与中高层官员的往来, 探究明代内阁首辅参与国家治理的过程。另就 CBDB 的关系数量而言, 宋濂的资料最充分, 达到 700 条 (如图 4 所示), 但是和宋代朱熹 2,580 条社会关系量相比差距尚远。所以, MLP 和 CBDB 一直处在持续更新的状态 (如王守仁、湛若水、徐阶、叶向高等人的资料还在扩充之中), 尽最大可能一步步完善明代历史人物的数据资料。

| 人物 ID  | 人物姓名 | 发出信件数量 | 人物 ID  | 人物姓名 | 关系数量 |
|--------|------|--------|--------|------|------|
| 123973 | 申時行  | 1803   | 10097  | 宋濂   | 700  |
| 69011  | 張居正  | 822    | 34717  | 王世貞  | 318  |
| 131392 | 黃承玄  | 740    | 127987 | 李維楨  | 317  |
| 65500  | 熊廷弼  | 667    | 33860  | 呂柟   | 303  |
| 34758  | 魏大中  | 629    | 28410  | 王逢   | 238  |
| 30240  | 馮夢禎  | 591    | 28691  | 李東陽  | 235  |
| 56535  | 姚希孟  | 554    | 12087  | 楊維禎  | 223  |
| 123637 | 馮琦   | 551    | 28022  | 危素   | 191  |
| 125593 | 尹伸   | 511    | 30374  | 王守仁  | 184  |
| 30243  | 馮時可  | 490    | 10726  | 王禕   | 178  |
| 34919  | 屠隆   | 481    | 131191 | 湛若水  | 177  |
| 34717  | 王世貞  | 476    | 34649  | 崔銑   | 175  |
| 35222  | 袁宏道  | 457    | 27852  | 戴良   | 170  |
| 124068 | 沈演   | 457    | 34493  | 王直   | 167  |
| 65870  | 湯顯祖  | 436    | 34532  | 章懋   | 166  |
| 30593  | 徐學謨  | 433    | 29602  | 程敏政  | 161  |
| 439144 | 張璠   | 416    | 66729  | 楊寓   | 159  |
| 135569 | 方應祥  | 415    | 127530 | 汪道昆  | 151  |
| 132172 | 趙南星  | 414    | 34579  | 王鏊   | 149  |
| 130250 | 梅鼎祚  | 410    | 68547  | 徐階   | 142  |
| 126634 | 王庭譔  | 391    | 33465  | 吳寬   | 141  |
| 128411 | 何東序  | 379    | 28825  | 張邦奇  | 139  |
| 131951 | 葉向高  | 376    | 34742  | 顧憲成  | 136  |
| 68522  | 俞大猷  | 373    | 126359 | 于慎行  | 131  |
| 134653 | 朱長春  | 367    | 28347  | 楊榮   | 130  |
| 125581 | 李材   | 361    | 128983 | 姜寶   | 129  |
| 54133  | 黃汝亨  | 351    | 68133  | 費宏   | 126  |
| 34747  | 鄭元標  | 349    | 131951 | 葉向高  | 126  |
| 35112  | 吳國倫  | 348    | 24554  | 林俊   | 124  |
| 134652 | 何三畏  | 348    | 130797 | 陳文燭  | 121  |

图 4 MLP 发出信件数量以及 CBDB 关系数量排名前 30 的明代历史人物

在 MLP 数据平台完成明代书信人物的检索之后, 还可以通过 MLP 本身的算法和 Gephi 等可视化软件, 对检索资料再处理, 通过图像呈现明代历史人物的书信往来情况。以王世贞为例, 截至 2022 年 5 月, 王世贞 (人物 ID: 34717) 发出的书信有 476 封, CBDB 关系量 318 条 (如图 4 所示)。根据通信人物, 可以大致看出王世贞的书信圈 (如图 5 所示)。根据社会关系, 可以看出他大概的社会关系圈 (如图 6 所示)。根据这些图像资料, 得到的不再是单一或者多个历史人物的信息, 而是不断网格化、形象化的海量数据信息。这些数据和图像互相补充, 会让王世贞的通信网络、社会关系更加丰满。

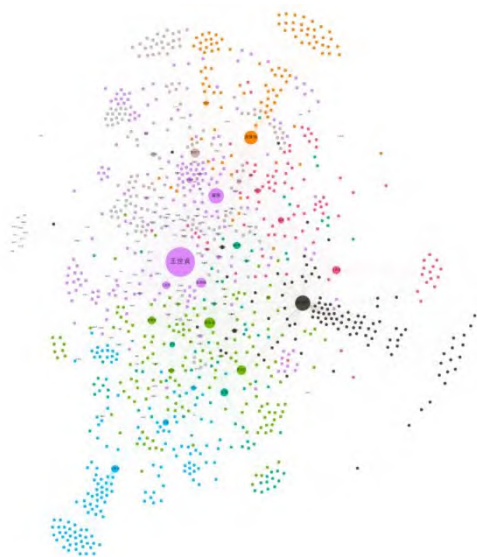


图 5 据 MLP 现有资料整理王世贞的通信网络

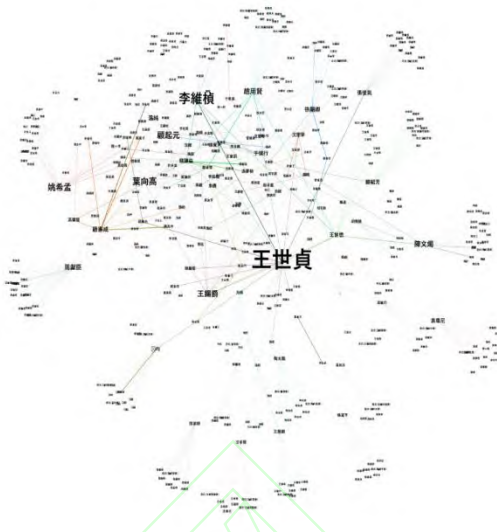


图 6 据 CBDB 现有资料整理王世贞的社会关系

### 3.3 结合 CBDB，促进对明代人物社会关系的研究

哈佛大学 CSA 明代书信计划的数据平台与 CBDB 数据库结合后，让学界对明代书信的研究增益。仅仅根据 MLP 平台，学者得到的关于书信标题、文集以及收发人的数据信息。但是，将 MLP 平台与 CBDB 数据配合使用，明代历史人物的身份信息就会更加翔实。

如上所言，MLP 不仅处理书信收发人的身份信息，还与 CBDB 进行匹配。对常见的历史人物，学者可以利用社会关系的角度综合分析历史人物的迁徙信息、家庭关系、仕宦变迁、师友交游、著作关系等。史籍中出现的历史人物不再是单一的人，而是可以通过数字化手段迅速确定身份信息及其社会关系的对象。随着书信资料的不断补充，越来越多的明代历史人物及其社会关系会浮出水面。

MLP 平台与 CBDB 的结合使用相当于提供学者一份便捷的明代历史人物的数据档案。这份数据档案不仅提供历史人物的基本身份信息，还提供以书信为主的社交数据网络。通过书信圈，进而探索历史人物的籍贯圈、入仕圈、亲属圈、师友圈、著作圈等，让明代历史人物在可视化的社会关系网格中“活起来”，助力海内外学者的学术研究。

总之，在 CSA 明代书信计划的开展过程中坚持一个原则，即始终把数字化研究和人文研究相结合，用数字化研究服务人文研究。具体而言，要保证数据库收录的历史人物、书信、著作等相关信息要有明确的史料来源和必要的考证功夫，增强数字人文研究的可信度。项目研究的重点在于对明代历史人物的书信进行竭泽而渔式地收录和整理，弄清通信人物（尤其是较少被关注的小人物）的社会身份，建立他们的关系网络。本项目研究的难点有二：一是小人物的考证。由于资料有限，仅仅通过书信的标题和内容，很难一一确定通信人的姓名，这就需要不断积累经验，提高对历史人物的识别度；二是数据库程序的完善。由于本项目在启动和运行过程中，并无同类的系统可供借鉴，故而从海量书信的抓取、分类，到 CSA 平台的建设、调试，再到人物、书信的编码和导入，以及对通信人物网络的后期构图等，都需要不断更新和完善程序。

## 4 特色、不足及其发展前景

作为哈佛大学 CBDB 项目的组成部分，MLP 计划的特色体现在多个方面：（1）明代海量书信资料的首次数字化。到目前为止，哈佛大学 CSA 明代书信计划是海内外唯一一个大规模处理明代书信的群体传记学数字研究计划。该计划旨在不断收录明代历史人物的书信，

并从数字化的角度予以整理分析；(2) 以人物传记为中心，基于通信人物往来的数据可视化。通过 MLP 整理分析的结果，与 CBDB 相匹配，有助于形成可视化的结果，比如直观地展现明代历史人物的社交网络等；(3) 明代书信计划的持续开放性。该计划欢迎海内外学者提供更多的明代书信资料，帮助平台确定更多收信人的身份信息，发挥交互研究的价值。与此同时，MLP 计划不会设定完成项目的时间上限，而是持续搜集新的书信并予以数字化处理。至于该计划的平台数据（目前 MLP 仅支持繁体中文搜索，CBDB 则提供中英文两种信息），也会免费开放给从事明代中国研究的海内外学者，为学界研究明代历史人物以及社会关系提供数据支持和信息检索。

明代书信的数字化研究与明代书信研究之间是互相促进的关系，对大量明代书信的数字化研究会激发学界研究明代书信的兴趣，并不会也不可能取代明代书信研究。MLP 在提供便利的同时，也存在不足：(1) 该计划侧重书信往来的双方而非书信内容本身，重视文字资料以及数据分析，无法兼顾书信的版本价值、艺术价值、美学价值等；(2) 该计划在提供明代通信人物资料的基础上，学者如果想要进一步展开研究，比如通过限定时段、地域等信息进行纵向、横向比较，或者进行可视化制图，则需要熟练 CBDB、MLP、Gehpi、GIS 等软件系统的切换与操作。换言之，如果使用者不能综合利用以上软件系统，恐怕难以发挥明代书信计划检索结果的最大价值。这就要求使用者不仅具有良好的专业人文知识，同时要有一定水平的数字人文技能。

自项目启动两年多来，笔者一直在思考如何充分、便捷、高效地发掘明代书信的价值。CSA 明代书信计划依托已经运行多年相对成熟的 CBDB 项目，从确定通信双方入手，最大程度地挖掘历史人物的身份信息，并建立历史人物的通信网格，进而完善历史人物的社交网络，并以数据库的方式将相关信息向国际学界实时分享。作为一种全新的尝试，这是海内外学界第一次对明代 10 万封书信规模的大型数据研究。随着更多资料的补入，明代书信的数字化研究拥有广阔的前景，既会丰富学界对明代历史人物及其社会关系的认识，也会为其他历史时期（如唐、宋、元、清、民国）、其他国家和地区的书信研究提供新的研究视角、思路与方法。

最后，在 MLP 项目工作过程中，切感结合数字人文与专业知识的必要性。简言之，既要懂计算机技术的人掌握专业知识，又要让掌握专业知识的人懂计算机技术。值得欣喜的是，近年海内外高校涌现出不少数字人文方面的期刊、团队和专业增设。有理由相信，有了更多“数字+人文”的人才培养和团队建设，未来会有更多“数字人文+”的可预期效果。

## 致谢

本文完成于笔者在哈佛大学访学期间，得到哈佛大学 CBDB 项目经理王宏魁先生的鼎力支持，谨致谢忱！

## 参考文献

- [1] 李敬敏, 骆遥.《坤輿万国全图》地图数字复原与讨论[J].测绘科学, 2021(4):121-134.
- [2] 参修萌, 李雪飞.明代官服的结构研究与数字化复原[J].丝绸, 2021(12):110-116.
- [3] 石守谦, 杨儒宾.明代名贤尺牘集[M].台北:何创时书法艺术文教基金会, 2013.
- [4] 金桂台.明代文学书信之性灵献酬论[J].中国文学研究辑刊, 2008(1):151-173.
- [5] 史小军, 梁娟.古代书信体文论发展阶段初探[J].学术论坛, 2005(7):159-163.
- [6] 史小军, 梁娟.明代文人书信体文论探究[J].广西民族学院学报(哲学社会科学版), 2006(2):164-168.
- [7] 王林军.尺牘书疏, 千里面目——《明代名贤尺牘集》编辑手记[J].出版广角, 2020(3):94-95.
- [8] 中国历代人物传记数据库 (CBDB).简介[EB/OL].[2022-9-1].<https://projects.iq.harvard.edu/chinese/cbdb>
- [9] 包弼德.历史上的理学[M].王昌伟译.杭州:浙江大学出版社, 2010.
- [10] 包弼德.斯文:唐宋思想的转型[M].刘宁译.南京:江苏人民出版社, 2017.
- [11] 中国历代人物传记数据库 (CBDB).CSA 明代书信计划[EB/OL].[2022-07-

- 01].<https://projects.iq.harvard.edu/chinesecbdb/csa-%E6%98%8E%E4%BB%A3%E6%9B%B8%E4%BF%A1%E8%A8%88%E5%8A%83>
- [12] 中国历代人物传记数据库 (CBDB). 明代书信计划数据检索 [EB/OL].[2022-07-31].<https://csa.cbdb.fas.harvard.edu>

**作者简介** 陈士银，博士，扬州大学社会发展学院历史系讲师，哈佛大学访问学者，chenshiyin007@126.com。  
**收稿日期** 2022-11-01

(责任编辑：刘木子)

