Article

Peer Review:

Copyright:

Open Access:

Digital Preservation:

ARTICLE

How the Digital is Changing Research and Teaching on Asia

Peter Bol
Harvard University, US
pkbol@fas.harvard.edu

Tools built for scholarship in a digital environment allow us to take more information into account, to share it more broadly, and to analyze it with greater precision. But they also require mastering new skills and making conceptual leaps. The importance of the digital for Asian studies applies generally to the liberal arts as a whole. Digital advances offer us all opportunities for reaching a broader audience through online teaching.

Today, the place of the liberal arts in higher education is under attack. On the one hand, politicians question the value of the liberal arts in preparing young people for the workforce. On the other hand, students make their own choices. Increasingly burdened by debt to meet the high cost of college and dropping out in large numbers before graduation, students at even the most elite universities—if the decline in humanities enrollment is any indication—are becoming more and more vocational. This vocationalism is evident on EdX and Coursera, the two major open online learning platforms, based on the current assumption that people are more inclined to pay money to earn money than they are to pursue learning that develops the less tangible skills of being a thoughtful and reflective person. In the *Analects*, Confucius says, "Those who learned in antiquity did so for themselves; those who learn today do so for others." Learning for oneself is the heart of the humanities and is integral to the liberal arts. The liberal arts are apparently in trouble, yet I do not despair. Advances in digital technology offer ever-increasing methods for expanding student engagement and reaching would-be learners around the world.

The Asian humanities face even more challenges than do the liberal arts generally. In the United States, advocates of "the Humanities" are not always inclined to go beyond the bounds of Mediterranean civilization. I may point out that China also has a tradition of art, literature, philosophy, music, and historiography, but I cannot argue that those traditions have been important to the cumulative cultural traditions of Mediterranean civilization until very recent times (from the perspective of several millennia). The reverse is true as well, of course, but at least we can point out that South Asian, Southeast Asian, and East Asian cultures had interesting and complex histories of mutual influence well before they began to master the cultural knowledge and practices of the West.

Globalization is not unidirectional, but how do we prepare our students in the U.S. to be culturally global, given that almost half of the world's population resides in Asia? Asking our students to become conversant with both Greek and Chinese thinkers; Judeo-Christian, Buddhist, Hindu, and Confucian texts; Michelangelo and Guo Xi; and so on, is asking a lot. We want our students to know more, but most of us are not in the business of stuffing students full of knowledge of the Asian greats

just for the sake of knowing more. And how do we deal with the multiple registers of culture, with works of popular culture as well as the sophisticated products of high culture? There is too much to do, and all the while we also want to encourage our students to "learn for themselves."

On top of all of this there is a second challenge that should concern us just as greatly: the decline in students taking foreign language courses. Compared to higher education in Asia and Europe, the percentage of U.S. students in foreign language courses was low to begin with. In 2013 there were about 1.6 million students in foreign language courses, out of nearly 21 million.[1] Although foreign language enrollments had increased prior to 2009, there has been a decline of nearly 7% since then.[2] And yet it turns out that one in five U.S. residents speaks a foreign language at home.[3] From my perspective, a serious engagement with Asia, past and present, requires learning the language. While we may agree with all the reasons the American Council on the Teaching of Foreign Languages gives for learning foreign languages, college students themselves apparently do not.

It turns out, however, that Americans do want to study foreign languages—just not in school. Consider the case of Duolingo, www.duolingo.com, which began in 2009 at Carnegie Mellon University and now has 150 million registered users around the globe, over half of whom are studying a language other than English.[4] It helps, of course, that Duolingo is free to use. The case of Duolingo serves two purposes in this essay: it shows that there are many people who want to learn a foreign language, and that they are willing, even eager, to do so in a digital environment, provided they receive the tools with which they can learn for themselves.

---

[1] National Center for Education Statistics, http://nces.ed.gov/fastfacts/display.asp?id=98.

[2] Colleen Flaherty, "Not a Small World after All," *Inside Higher Ed*, February 11, 2015. https://www.insidehighered.com/news/2015/02/11/mla-report-shows-declines-enrollment-most-foreign-languages. The most recent MLA report on enrollments dates from 2013. https://www.mla.org/Resources/Research/Surveys-Reports-and-Other-Documents/Teaching-Enrollments-and-Programs/Enrollments-in-Languages-Other-Than-English-in-United-States-Institutions-of-Higher-Education.

[3] The Center for Immigration Studies, based on the Census Bureau's 2013 American Community Survey http://cis.org/record-one-in-five-us-residents-speaks-language-other-than-english-at-home.

[4] The Wikipedia entry on Duolingo is more informative than the company website, https://en.wikipedia.org/wiki/Duolingo.

This brings me to the heart of my topic: research and learning in a digital environment. The case I want to make is simple. We have a growing number of computational tools for use in research and teaching in Asian studies, an increase that has outpaced our proficiency, and we have ever more opportunities to teach Asian studies online. We are not using these tools yet to their full capacity; we should be.

Why Tools?

How do academic fields advance? In the first half of the twentieth century there were few positions in higher education devoted to Asian studies. Today, we specialize by country and period. Specialization may distance us from the novice learner but it has also led to the advancement of knowledge and a proliferation of positions. Fields also advance through changes in shared conceptualization of the questions disciplines ask through "paradigm shifts." No less important is a third way in which knowledge advances: the development of new tools. One need only think of the telescope and microscope to realize that the ability to see what was previously invisible to the naked eye changed the state of knowledge about the world. The tools developed during the last few decades for research allow us to take more information into account and to analyze and visualize it in ways that had until now been beyond the power of an individual.

Here I want to introduce four tools I have used in my own field of Chinese history: relational databases, text-mining, geographic information systems (GIS), and social network analysis. The employment of these tools requires data, and data has to be collected and made accessible. I use the China Historical GIS (CHGIS) and the China Biographical Database (CBDB), two international projects I have personally participated in, as examples of data resources for using these tools.

The Relational Database

The China Biographical Database, or CBDB, is a collaborative project between the Fairbank Center for Chinese Studies at Harvard University, the Institute of History and Philology of Academia Sinica 中央研究院歷史語言研究所, and the Center for Research on Ancient Chinese History at Peking University 北京大學中國古代史研

究中心. As of May 2017 it had data on about 380,000 men and women, mainly from the seventh to the early twentieth century. The database can be downloaded as a standalone or consulted online.[5] **Figure 1** depicts its spatial coverage and **Table 1** gives its temporal distribution by dynasty.
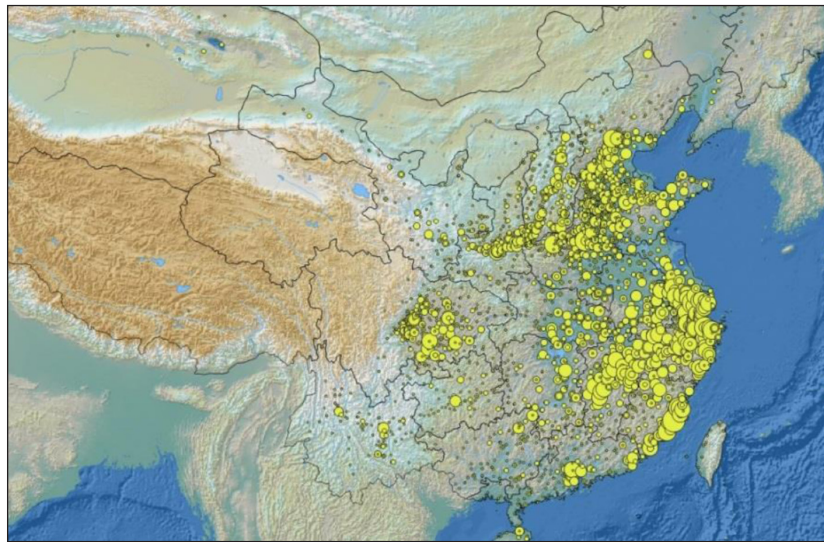
Figure 1: Spatial distribution of persons in CBDB.

Table 1: Temporal distribution of persons in CBDB by dynasty.

Dynasty	Persons
Tang 唐	52751
Five Dynasties 五代十國	2900
Song 宋	45199
Liao 遼	325
Jin 金	300
Yuan 元	19915
Ming 明	163522
Qing 清	62226
Republic 民國	3222
Other	20366

---

[5] See the CBDB website, https://projects.iq.harvard.edu/cbdb, for the history of the project and current work and to download instructional materials and the entire database.

This table of rows and columns is a common and well-understood way of organizing data. Here is part of a CBDB table with basic biographical information:

A relational database differs in that it relates many tables to each other. To illustrate, suppose we had information on the offices and social associations of a person but only one table to store the data. We might produce something like this.[6]

Name 姓名	Dates 日期	Offices 任官	Associations 社會關係
Sima Guang 司馬光	1019–1086	(1) 1059 度支勾院 Budget Auditor; (2) 1085 門下侍郎 Executive of the Chancellery; (3) 1086 左僕射兼門下侍郎 Left Executive, Dept of Ministries [....]	(1) Yuanyou coalition member (元佑黨); (2) An Dun 安惇 Desires opposed by; (3) Chao Buzhi 晁補之 Sacrificial prayer written by; (4) Chen Jian 陳薦 Sacrificial prayer written for; (5) Chen Min 陳敏 Honored by; (6) Cheng Yi 程頤 Recommended; (7) Ding Du 丁度 Sacrificial prayer written for; (8) Fan Chunli 範純禮 Patron of; [....]

This formulation organizes data but does not facilitate asking who else served as a budget auditor, for instance. To do that the data could be reorganized as follows:

---

6 This and the following illustrations are by Michael A. Fuller, designer of the CBDB database.

Name 人物		Dates 日期	
Sima Guang 司馬光		1019-1086	

Person 人物	Posting Date 任命日期	Office Title 官名
Sima Guang 司馬光	1059	度支勾院 Budget Auditor
Sima Guang 司馬光	1085	門下侍郎 Executive of the Chancellery
Sima Guang 司馬光	1086	左僕射兼門下侍郎 Left Executive, Dept of Ministries

Person 人物	Association Type 社會關係	Associate 社會關係人
Sima Guang 司馬光	Yuanyou member (元佑黨)	(not applicable)
Sima Guang 司馬光	Desires opposed by	An Dun 安惇
Sima Guang 司馬光	Sacrificial prayer written by	Chao Buzhi 晁補之
Sima Guang 司馬光	Patron of	Fan Chunli 範純禮
Sima Guang 司馬光	Sacrificial prayer written for	Ding Du 丁度

By creating three tables (person dates, offices, and associations), I can now sort my information by any column, but having to retype names, offices, etc. is also terribly inefficient. Moreover, it makes it easy to introduce errors. Consider, too, that I have more kinds of data I would like to keep track of: where did a person live, serve, die, etc.; how did he enter office; who was her father, mother, brother, sister, husband, etc.; what did he write; and so on.

The solution to this problem begins with a distinction between code tables and data tables. Let us create one table listing all possible office titles, another with all possible associations, and another with all people, giving each an ID number. We would only have to enter a name or title once (and correct a mistake once). We then would create data tables in which we related person to office posting and person to associate and association type.

People 人物

ID	Name	Dates
1	Sima Guang 司馬光	1019-1086
2	An Dun 安惇	…
3	Chao Buzhi 晁補之	
4	Chen Jian 陳荐	
5	Chen Min 陳敏	
6	Cheng Yi 程頤	
7	Ding Du 丁度	
8	Fan Chunli 範純禮	

Office Titles 官名

ID	Office Name
1	度支勾院 Budget Auditor
2	門下侍郎 Executive of the Chancellery
3	左僕射兼門下侍郎 Left Executive, Dept of Ministries

Associations 社会关系

ID	Association Type
1	Yuanyou coalition member (元佑党)
2	Desires opposed by
3	Sacrificial prayer written by
4	Sacrificial prayer written for
5	Honored by
6	Recommended
7	Patron of

Postings Data 任官资料

Person ID	Office ID	Posting Date
1	1	1059
1	2	1085
1	3	1086

Associations Data 社会关系资料

Assoc Type ID	Person ID	Assoc ID
1	1	-1
2	1	2
3	1	3
4	1	4
5	1	5
6	1	6
4	1	7
7	1	8

This type of document is supremely efficient and the software for a relational database allows one to relate these tables together on the basis of IDs.
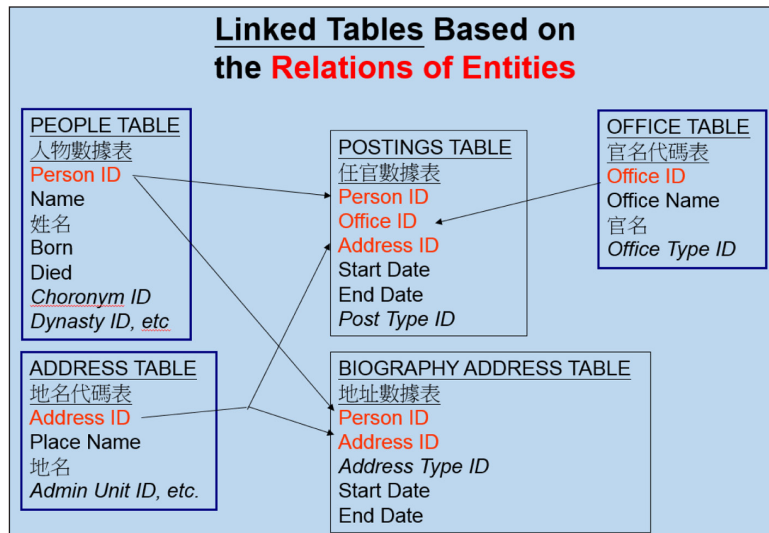

The advantages of a relational database are:

- Data is linked to each other: good for identifying connections and patterns.
- We can search with any entity—not just person's name.
- Information about entities appears just once in the database. Errors in information need to be corrected just once.
- Complex queries with multiple conditions are easily enabled.

This last point, complex querying, is really the heart of the matter. We want to derive new information about large numbers of people from the data we have collected. Suppose, for example, I wanted to know where all the people who were labeled as members of the Yuanyou coalition, a political faction, came from in order to see if they were associated with some places more than others. In my relational database that simple query results in a list of 217 people, including their dates and places of origin. These points could then be mapped, as below:

Or we could go even further and inquire into the degree of intermarriage among those faction members for whom we have kinship data. The figure below illustrates this for one section of the network, with faction members circled in red.
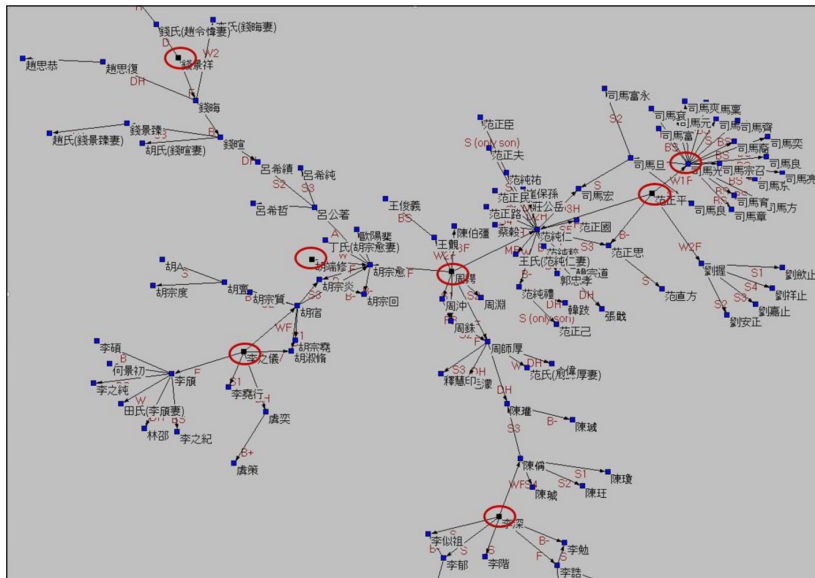
S = son, D = daughter, F = Father, M = mother, B = brother, H = husband, W = wife. The query included only F, M, B, S of the Yuanyou members.

Text-Mining

There are many ways of getting information by mining digital text corpora, large and small.[7] Here, I want to illustrate two we have used to collect data for CBDB. Consider the following passage from the biography of Lü Zuqian in the *History of the Song Dynasty*. I have color-coded different kinds of information.

We already possess the names of people, kinship relations, office titles, place names, types of social association, and modes of entry into government service. The goal is to identify these factoids in a digital text, to tag them with the appropriate code numbers, and to extract them into data tables that can be appended to the relation database with a minimum of manual work.

To accomplish this we wrote a program in Python, a programming language, in which we included "regular expressions."[8] In essence, a regular expression is a representation of a pattern of characters that are to be searched for in the target text. The power of a regular expression is that it can include large numbers of

---

7   Two books by Franco Moretti illustrate some of the possibilities: *Graphs, Maps, Trees: Abstract Models for a Literary History* (London; New York: Verso, 2005) and *Distant Reading* (London: Verso, 2013).

8   For an introduction to regular expressions see https://en.wikipedia.org/wiki/Regular_expression.

variations. Suppose we want to find all references to a year of death. We know that the typical way in which a Chinese historical text expresses this is Reign Period Name+number+the word "Year 年"+the word "Died卒". To capture this we write a regular expression that includes, in order, *all* the reign period names in China's history, plus all numbers up to 60, plus the words Year and Died. We can also add a command to the program to convert the date to the western calendar. At CBDB, we built a "regex machine" that users can download to use on their own without having to write code.[9]

This kind of targeted text-mining, which aims to identify, code, and extract data depends on having extensive code tables. In addition to the reign period names, CBDB has over 21,000 place names (with the administrative units they belong to and their begin and end dates) and over 30,000 office titles. These codes can be used by other online programs via application programming interfaces (APIs). A particularly successful example of this is MARKUS.[10] To quote the website:

> With MARKUS you can upload a file in classical Chinese (and perhaps in the future other languages) and tag personal names, place names, temporal references, and bureaucratic offices automatically. You can also upload your own list of key terms for automated tagging. You can then read a document while checking a range of reference works at the same time, or compare passages in which the same names or keywords appear. Or, you can extract the information you have tagged and use it for further analysis in our visualization platform and other tools.

The user uploads a text (.txt) file, selects from the automated markup options (or does it manually), and runs the program. Here, with the biography of Lü Zuqian used above, are two steps in the process. First is the selection of options.

---

[9] https://projects.iq.harvard.edu/cbdb/cbdbregexmachine.

[10] http://dh.chinese-empires.eu/beta/.
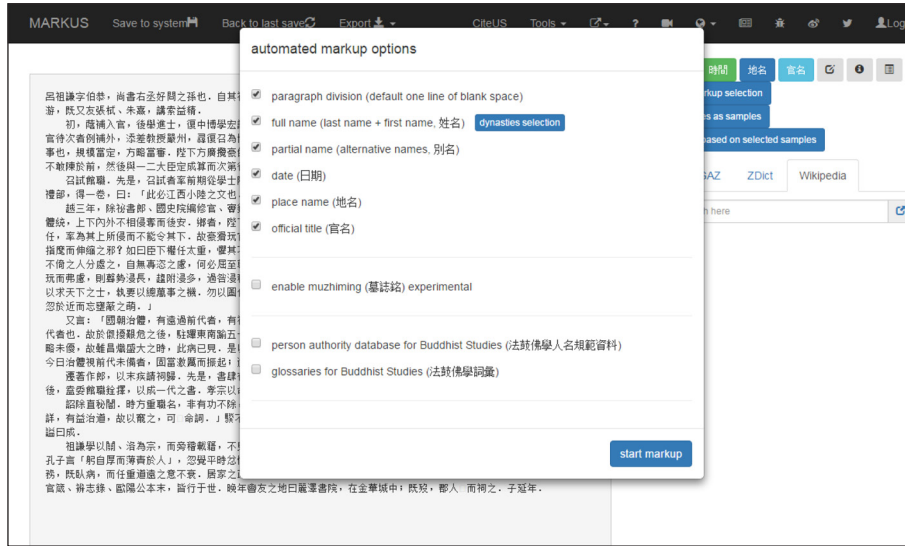
The second figure shows the marked-up text (which the user can correct) with a report on what kinds of data were discovered and choices for downloading the results.
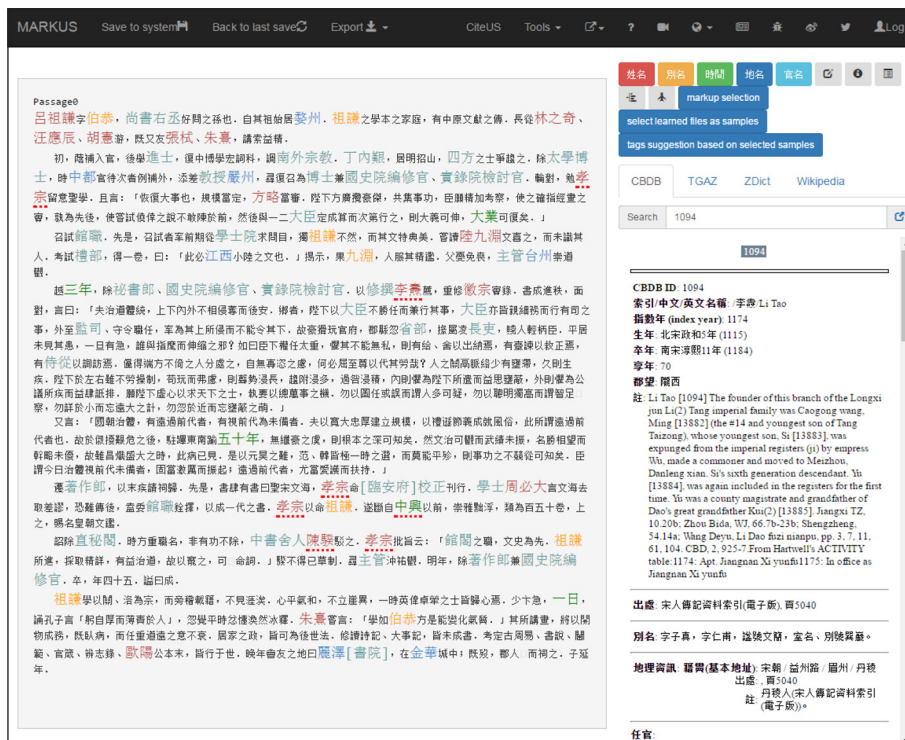
Geographic information systems (GIS)

GIS is a software program that draws on a database with geographic information to create maps. The concept is simple even if the software is not. In contrast to a printed map, which may include many kinds of information on a single printed surface, a GIS enables one to treat different kinds of information as different layers. For example, I could call up a digital elevation model, in which every pixel has a color value to indicate elevation and combine it with a dataset of lines representing rivers.
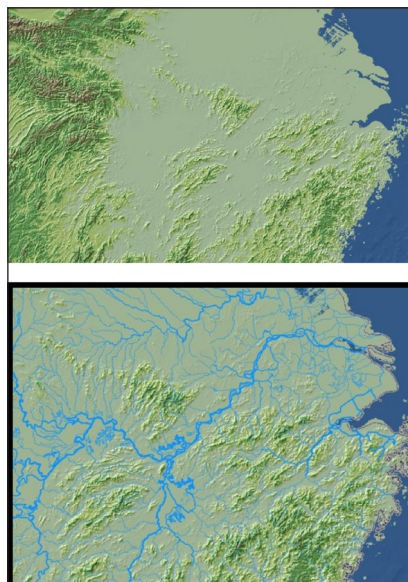
To this we can add points representing prefectural and county seats.
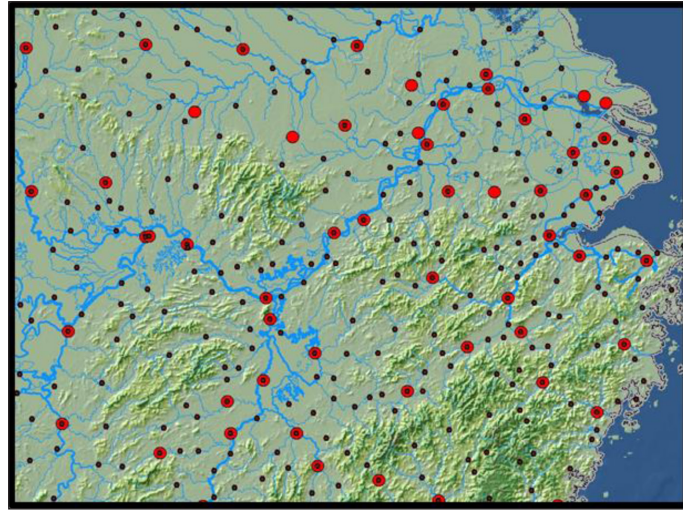
All the datasets in these examples are freely available from the China Historical GIS project, a collaboration between the Harvard Fairbank Center and Fudan University.[11] As the website explains:

> The main objective of the CHGIS project is to create a flexible tool, in the form of a documented database of places and administrative units, which can be used to investigate any sort of geographically specific data related to China. The unique ID numbers for each of the CHGIS temporal instance records can be used as geocodes in relational databases, or to mark up texts, enabling users to import their own datasets into the CHGIS platform. Users will be able to associate their own data with CHGIS records, and then use the CHGIS database to sort, query, and display their data for different historical periods and at different levels of aggregation.
>
> The CHGIS project is designed to provide a GIS platform for scholarly and scientific research. The advantage of creating the CHGIS, rather than print-ing paper maps, is that the contents can be modified, improved, or anno-tated by the user for their own purposes.

---

[11] http://www.fas.harvard.edu/~chgis/.

Having downloaded the CHGIS datasets, the user can search the database for administrative units and capitals for any given time in Chinese history, can create customized digital maps for particular times and places, or can join their own datasets for spatial analysis, thematic mapping, or other specialized statistical modeling according to their own interests. Also included in the datasets are layers for historical coastlines, major rivers, and generalized elevations.

CHGIS extends from 221 BCE to 1911 CE (Fudan is currently working on 1911–2015), allowing the user to query the data for any year. Here are the prefectural borders and county seats in Ming China in 1400 and the native places of the followers of the twelfth century philosopher Zhu Xi.

To use CHGIS datasets one must install free Quantum GIS software, and then of course learn how to use it. CHGIS also provides an online gazetteer, allowing users to search for and locate historical place names.[12]

People who have developed GIS data can make it available through CHGIS. They can also upload it to WorldMap.harvard.edu and share it with a class or with the world. As of May 2017 there were 20,000 registered users (among the 1,500,000

---

[12] http://www.fas.harvard.edu/~chgis/search/.

unique visitors) who had added 29,000 data layers, in addition to 70,000 remote service layers. Users have created 7,000 map collections. Readers may want to try the modern China map http://worldmap.harvard.edu/chinamap, the map of the 2011 Tohoku disasters in Japan http://worldmap.harvard.edu/japanmap, or any of the historical series for China, ChinaXmap 1.0 to 11.0, beginning with http://worldmap. harvard.edu/maps/7063. Free registration allows one to create one's own map using existing data layers or upload data layers of one's own making.

Social Network Analysis

"In the short run actors make relations, but in the long run relations make actors."[13] This statement seems to me to capture a basic assumption of those who see social networks as the key to understanding social processes. Despite the limited survival of historical sources, knowing even part of the story of who people were related to, who they taught, or who they sponsored is useful. CBDB can export queries to both of the two free social network analysis and visualization programs: Pajek and Gephi.

The challenge to using these tools is discovering the interpersonal relations in the sources. In the Chinese case we particularly benefit from two kinds of sources. Literary collections provide a wealth of information on kinship and literary exchanges with other literati. Local gazetteers, which cover all of China by the end of the Ming dynasty, contain biographies (which may mention kin), examination and school lists (which may tell us when people were in the same class), lists of local officials (which may tell us who served together), etc.

Two things that can be gleaned from network analysis are particularly valuable. The first is the measurement of comparative centrality, although this is distorted by the partial survival of literary collections. In the first example, I have asked who Lü Zuqian 呂祖謙 wrote letters to and in turn to whom his addressees wrote. It is evident that there are no letters from the vast majority of addressees. But at the same

---

[13] From the marketing blurb for John F. Padgett and Walter W. Powell, eds. *The Emergence of Organizations and Markets* (Princeton: Princeton University Press, 2012). http://press.princeton.edu/titles/9909. html. For historical network analysis generally see http://historicalnetworkresearch.org/.

time we see that we have a much smaller number of major figures who are in contact with each other and have separate cohorts of their own followers.
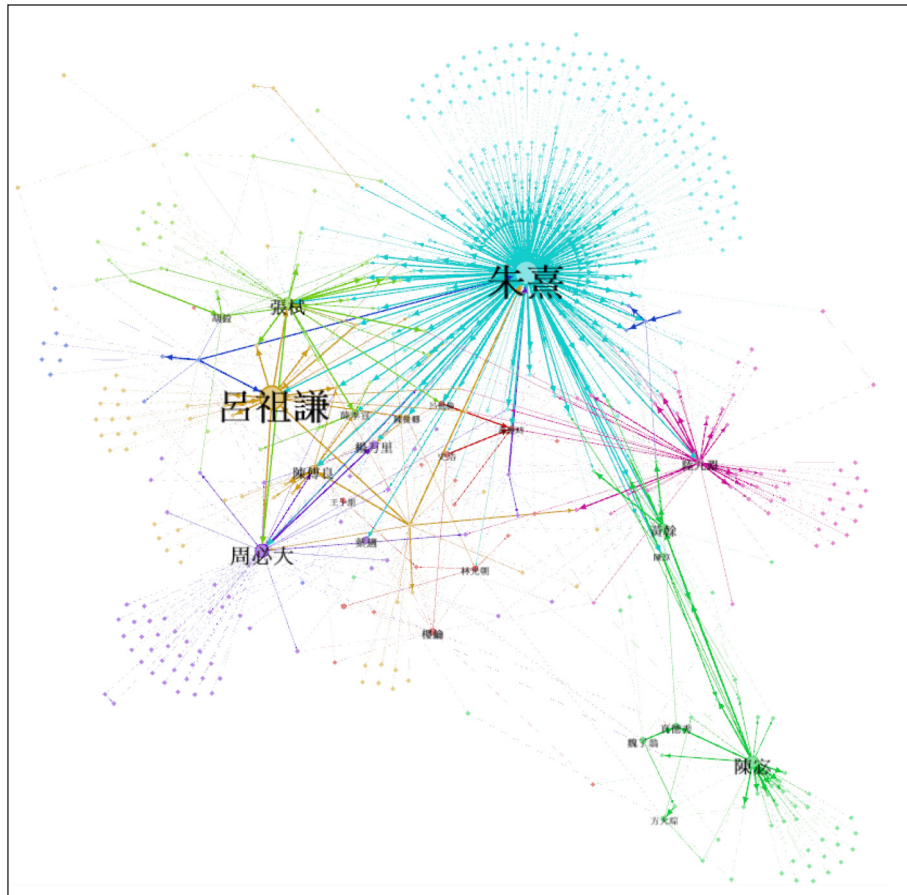
Letters to and from addressees of Lü Zuqian's letters were 597 persons. Node size (and corresponding label size) shows centrality of the person. Color shows communities.

The second thing we can learn is the role of certain individuals in mediating between others. This brokering, termed "betweenness centrality," is illustrated in the figure below. At either end of the network are two contemporaries who served at court at the same time, the literary intellectual Su Shi on the left and the moral philosopher Cheng Yi on the right. There is no evidence for a direct link through writing between the two. When we look for paths by which they were connected,

we see that in all but one case the path between them involved two or three others. The one exception is the figure at the middle of the graph, Sima Guang, who was connected to more people in the graph than any other. In fact, Sima served as chief councilor when Su and Cheng were at court.
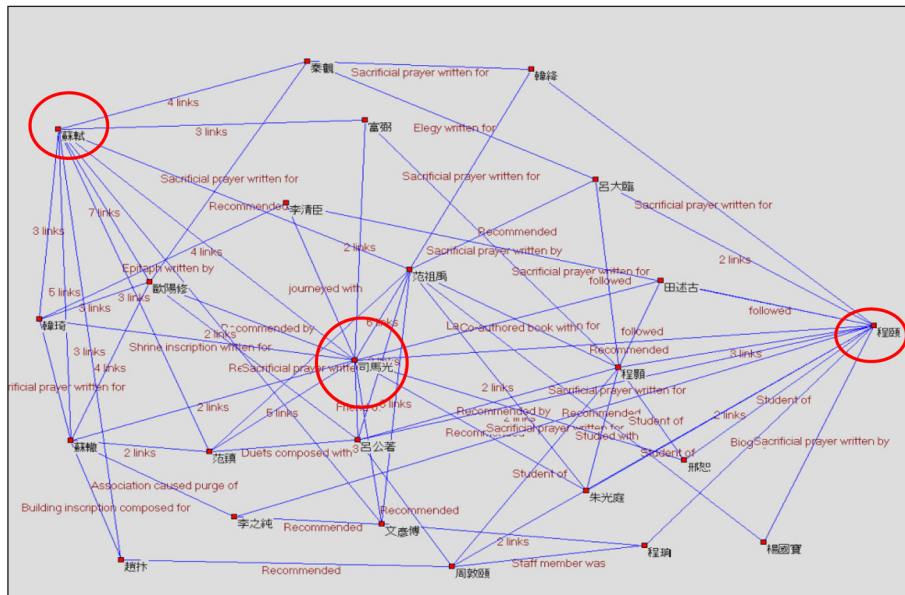
In this discussion some things are evident. First, the software tools are all widely available but require time and energy in learning how to use them. Second, the use of these tools also requires data in a digital form. Third, most obviously in the case of relational databases and to a lesser extent with GIS, the tools can also function as repositories for organizing and preserving data and preparing data for analysis. Fourth, in contradistinction to the tradition of the individual scholar working alone, developing and applying these tools involves collaboration between technical knowledge and content expertise. This last point, the necessity of collaboration, applies also to teaching in a digital environment.

Teaching Asian Studies Online

Every year about one third of the 21 million students in higher education in the U.S. take at least one course online for academic credit. Most of these courses are small, involve contact between the instructor and the students, offer
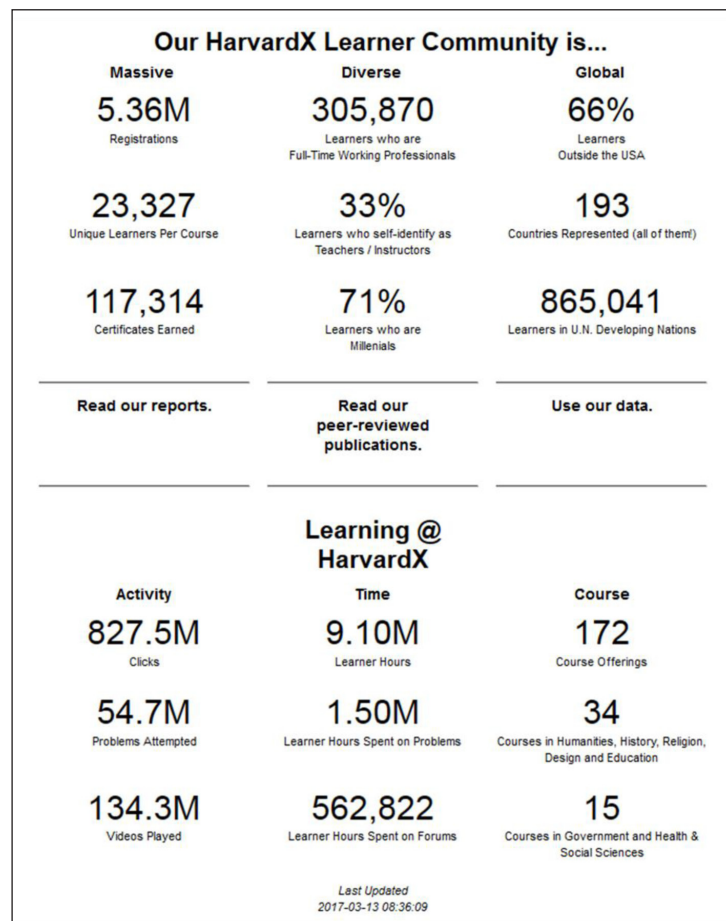
academic credit, and charge tuition. They may include videos of lectures, written assignments, online discussions, tests, and so on. Online learning is a great boon to those who, for a variety of reasons, cannot attend classes in person. Online courses for credit and online degrees are good for students and an important source of revenue for institutions. So we already know that people are prepared to pay to learn online. The advent of "massive open online courses" (MOOCs) did not change this. MOOCs are different from traditional online courses because they are built to scale, they are free, and they do not offer academic credit from the originating institution. Increasingly, however, they offer various forms of certification for a fee much smaller than normal tuition. MOOCs are important because they have shown that the audience for demanding structured learning—for academic courses, in short—is many times larger than that which higher education has been serving.

The design of MOOCs differs from typical online courses built around videos of lectures and slides. Ideally, a MOOC is designed for learning outcomes. Rather than simply putting a lecture online, we first ask what learning outcome is desired and then design the course to accomplish that end. The focus is on the learner rather than the instructor. Knowing that attention diminishes after a few minutes, courses are designed to actively engage the learner; a straightforward lecture rarely lasts more than five to eight minutes before something—an assessment, a simulation, a dynamic map—requires the learner to respond. Well-organized discussion forums are a must in liberal arts courses where thinking *with* the stuff one has learned is as important as remembering the stuff itself.

The two major platforms for open online courses are edX, created by Harvard and MIT as a not-for-profit, and the larger Coursera. Both carry a variety of successful courses in Asian studies. Harvard, for example, has produced Visualizing Japan together with MIT and the University of Tokyo, as well as the ChinaX history series of ten mini-courses, the ChinaX Book Club on five modern novelists, a forthcoming ChinaX Humanities, and a ChinaX Contemporary China. All of these can be found on Harvard's Online Learning Portal. The open format of these offerings goes against the trend in open online learning, which increasingly offers courses that learners will
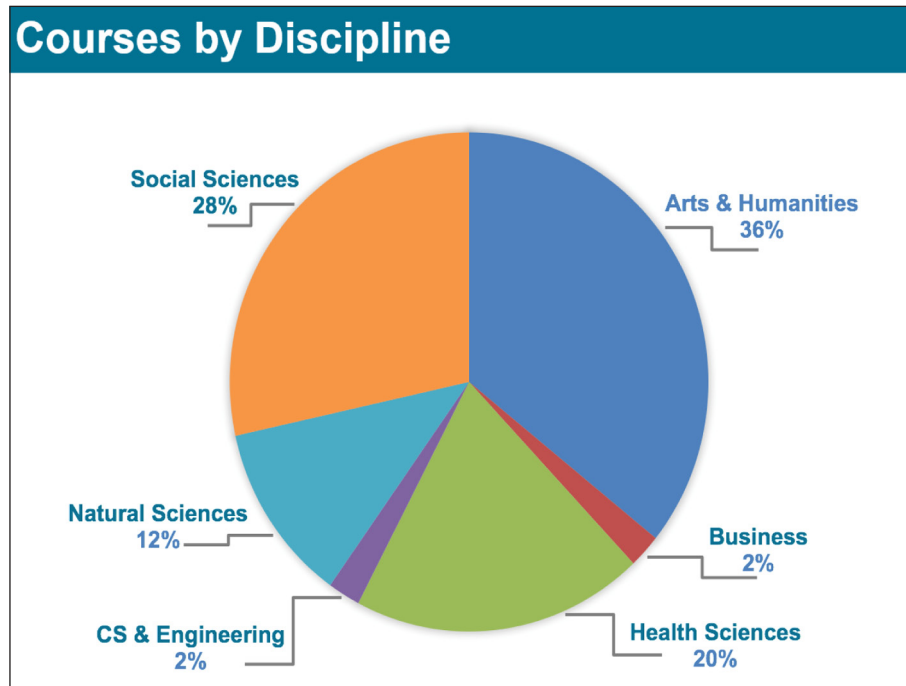
buy certificates for in order to advance their careers. Courses in Asian Studies—like liberal arts courses generally—appeal instead to those who seek lifelong learning.

We have learned several things from the over five million course registrations for HarvardX courses, some of which came as a surprise. The most important, in my view, is that contrary to the general trend, about 80% of learners say they are motivated by a desire for life-long learning, with only about 20% primarily interested in career advancement and about 70% already graduated from college. Statistics on HarvardX, https://harvardx.harvard.edu/insights, tell us more.

Our HarvardX Learner Community is...

Massive	Diverse	Global
5.36M	**305,870**	**66%**
Registrations	Learners who are Full-Time Working Professionals	Learners Outside the USA
23,327	**33%**	**193**
Unique Learners Per Course	Learners who self-identify as Teachers / Instructors	Countries Represented (all of them!)
117,314	**71%**	**865,041**
Certificates Earned	Learners who are Millenials	Learners in U.N. Developing Nations
Read our reports.	Read our peer-reviewed publications.	Use our data.

Learning @ HarvardX

Activity	Time	Course
827.5M	**9.10M**	**172**
Clicks	Learner Hours	Course Offerings
54.7M	**1.50M**	**34**
Problems Attempted	Learner Hours Spent on Problems	Courses in Humanities, History, Religion, Design and Education
134.3M	**562,822**	**15**
Videos Played	Learner Hours Spent on Forums	Courses in Government and Health & Social Sciences

Last Updated 2017-03-13 08:36:09

To our surprise, over two-thirds of the learners are between the ages of 22 and 31, and two-thirds are from outside the U.S. Most important in my view are the 35%

who are teachers. I suspect that our success in reaching teachers, many of whom take courses outside of their area of expertise, has much to do with the fact that Harvard MOOCs are predominantly focused on the liberal arts and sciences.

Courses by Discipline

Social Sciences
28%

Arts & Humanities
36%

Natural Sciences
12%

Business
2%

CS & Engineering
2%

Health Sciences
20%

It seems to me that, as we have seen in the case of foreign languages, there may be a decline in college enrollment but a simultaneous growth in interest for the liberal arts outside of college. After all, the audience between the ages of 22 and 82 is many times larger than the traditional college-age audience.

The greatest obstacle to taking our Asian studies courses online for a global audience is the cost of creating a course and then mounting it on a platform that attracts a worldwide audience, such as edX. It is already possible for faculty from several schools to join together to create a course and take advantage of the technologies that that make the digital world creative and interactive, but this has to be a labor of love, for there is, unfortunately, not likely to be revenue to recompense our efforts.

If for the moment schools with greater resources are more able to create MOOCs, what we can ask is that they make those learning opportunities easily accessible to their colleagues so that they can use them in their own courses. We know that colleagues at other schools have used parts of HarvardX courses in their own teaching, but this can be cumbersome, requiring learners to access the course on edX.org, and one cannot be certain that the course will always be available. For this problem we have developed a solution at Harvard which we hope to make available in a year or so to all schools. DART, "Digital Assets for Reuse in Teaching," allows one to search all HarvardX courses, find the particular piece of a course one is interested in, and drag and drop it into one's own course website. Thus, for example, one could search for the "Chinese dynasty song"—a mnemonic for remembering the major Chinese dynasties—or for an interactive map showing changes in the borders of Chinese dynasties over the last 3500 years. I have made them available here and welcome their use by my Asian Studies colleagues around the world.

Competing Interests

The author has no competing interests to declare.