DOI: 10.3785/j.issn.1008-942X.CN33-6000/C.2016.02.021

中国古典文学研究的几种可视化途径

——以汤显祖研究为例

徐永明
(浙江大学 人文学院,浙江 杭州 310028)

[摘 要] 当代西方教学和科研中常用的可视化工具主要有ArcGIS、QGIS、CHGIS、CartoDB、Worldmap、CBDB、GEPHI等数据库和软件。其中"中国历史地理信息系统"和"中国历代人物传记数据库"是由哈佛大学包弼德教授主持的项目,与中国的文史研究直接相关。利用上述数据库和软件,以汤显祖为例,可展示汤显祖行迹和活动地点、社会关系地理分布、社会关系点线表达等多种可视化效果图,给人以直观明了、耳目一新之感。这些数据库和可视化呈现方法对中国古典文学的研究和教学将起到极好的辅助作用。

[关键词] 中国古典文学;地理信息;汤显祖;可视化;哈佛大学;CBDB;CHGIS

历经几千年的中国古代文学,无论是原典作品还是研究成果,在"大数据"时代的数据库专家眼里,都是可以通过程序处理的大数据,这些大数据可按照人们的需要建成大大小小的各种类型的数据库,并可将其中的一些数据进行可视化展示。笔者并非计算机专业的数据库专家,但因在西方高校访学,常见到国外学者和研究生利用相关软件和数据库将研究对象进行可视化呈现,颇有直观明了、耳目一新之感。经过了解和学习,笔者认为这些数据库和可视化呈现方式若运用到中国古典文学的研究和教学中,则不啻为一种良好的辅助手段。故不揣浅陋,以明代戏曲家汤显祖研究为例,将相关的数据库和软件及操作步骤做一介绍演示,希望对读者有所帮助。

一、利用ArcGIS、QGIS 、CHGIS、CartoDB、Worldmap等地理信息系统软件和网站将作家的行迹和活动地点可视化

ArcGIS是由美国Esri公司开发的功能强大的分析性制图软件,可广泛用于一切与地理和空间有关的研究。从20世纪80年代初开发的第一代的ARC/INFO 1.0,到现在最新的ArcGIS 10.3版本,已有三十多年的历史。根据功能和产品类型的不同,ArcGIS不同款项的售价由几千美元到数万美元不等。美国哈佛大学购买了ArcGIS的使用权,在校师生可在自己的电脑上安装和使用该软件。但是,我国高校和科研机构很少有大规模集体购买供师生使用的,因此,该软件在中

[收稿日期] 2016-02-02
[在线优先出版日期] 2016-04-29
[本刊网址·在线杂志] http://www.zjujournals.com/soc
[网络连续型出版物号] CN33-6000/C
[基金项目] 教育部人文社会科学研究一般项目(10YJA751096)

[作者简介] 徐永明(http://orcid.org/0000-0002-3929-4101),男,浙江大学人文学院教授,博士生导师,文学博士,主要从事元明清文学研究。

国的使用受到了很大限制。

QGIS 是“Quantum GIS”的简称，是由 QGIS 发展团队开发的开源性地理信息系统软件，使用者可以免费到其网站（http://www.qgis.org）下载最新版本的 QGIS 软件。QGIS 项目始于 2002 年 5 月，发展至今已有 13 年的历史。与 ArcGIS 一样，QGIS 也是与地理和空间有关的分析性制图软件。

CHGIS 是“中国历史地理信息系统”（China Historical Geographic Information System）的简称，是由哈佛大学东亚语言与文明系包弼德（Peter K. Bol）教授主持的项目，项目经理为贝明远（Lex Berman）。它是一个开源性质的中国地理信息系统网站，网址为：http://www.fas.harvard.edu/～chgis/。该项目与复旦大学史地所合作，将中国历史地名和历史地图矢量化，并且以关系型数据库的方式记录地名的层级及沿革信息，由此，凡涉及中国古代历史地名的，都可通过数字化的中国历史地理信息系统得以可视化的展示。网站提供了中国历史地名的经纬度，但可下载的矢量历史地图只有清代的，明代和明代以前的只能查到部分地名的经纬度，没有矢量化的行政区域图。

CartoDB 是一个云上的地理空间数据库，使用者可以将已获得的经纬度数据批量导入 CartoDB 网站，从而快速创建基于地图的可视化效果，创建的地图可以在网上存储或公开发布，这也是一个开源性的网站。

Worldmap 是哈佛大学地理分析中心（the center for geographic analysis）开发的一个全球地理信息研究成果发布和共享平台。其中中国部分，包括了人口统计、宗教、交通、城市研究、少数民族和语言、能源、环境、教育、气候、公共健康、经济、历史等诸多领域的地理信息和地图。譬如，与文学有关的，有宋元明清的科举考试分布图、明清驿站路线图等。

下面以汤显祖为例，利用 QGIS 在地图上展示汤显祖行迹和活动地点。制作出来后的效果如图 1 所示（由于地图出版规定，此处不展示底图，只给出相对位置图）。那么，这张图是如何制作出来的呢？其步骤和方法如下：



图 1　汤显祖行迹和活动地点图

（1）安装 QGIS 软件。

（2）查出汤显祖行迹和活动地点（根据徐朔方先生撰写的《汤显祖年谱》）[1]。

（3）查出汤显祖行迹和活动地点的经纬度。这一步要利用 CHGIS，即“中国历史地理信息系统”网站，读者可以直接到该网站上查出历史地名的经纬度，也可以利用包弼德 CBDB 项目团队成员王宏甦先生开发的搜索界面去查找，海外的搜索界面网址为 http://oopus.info/chgis/name，国内的搜索界面网址为 http://oopus.info/chgis/cn。还可以用项目经理贝明远开发的搜索界面查找，其搜索界面网址为 http://maps.cga.harvard.edu/tgaz/。获得经纬度数据后，拷入 excel 表中，字段分别合命名为：name，X，Y。这里要注意的是，由于汤显祖是明朝人，故要查的地名一定是明朝行政区域下的地名，因为有些地名在不同历史时期的地理位置是有变化的。笔者查得的汤显祖行迹和活动地点的经纬度如表 1 所示。

（4）将 excel 表存为 CSV 格式文件，并上传到 QGIS 系统中。注意，上传入口在打开的 QGIS 左侧一大逗号处。点击确定后，在 filter 栏里输入 Xian 1980，双击下方的 Xian 1980。

（5）到 CHGIS 网站下载 v4_citas90_cnty_pgn_utf_stats 文件包。其下载路径为：DATA—China Historica GIS—Version 4 Datasets（with descriptions）—CITAS-1990-Counties（polygons）—Data Archive—1990 CITAS Counties(With Stats，UTF-8)—Dataset。

表1 汤显祖行迹和活动地点经纬度

name	X	Y	name	X	Y	name	X	Y	name	X	Y
临川	116.35	27.985	南昌	115.9	28.675	宣城	118.74	30.947	保昌	114.3	25.119
南京	118.77	32.053	黄州	114.87	30.447	杭州	120.17	30.294	南海	113.26	23.135
绍兴	120.58	30.005	吉安	114.97	27.103	赣州	114.93	25.847	恩平	112.31	22.192
梅岭	114.34	25.322	广州	113.26	23.135	东莞	113.75	23.047	肇庆	112.45	23.057
香山	113.37	22.526	澳门	113.55	22.2	长沙	112.98	28.198	温州	120.65	28.018
阳江	111.96	21.845	琼州	110.36	20.008	徐闻	110.16	20.33			
遂昌	119.26	28.588	滕县	117.16	35.085	丽水	119.91	28.449			
扬州	119.44	32.391	北京	116.37	39.931	通州	120.85	32.01			

(6)将下载的 v4_citas90_cnty_pgn_utf_stats 文件包解压,然后回到 GGIS 界面,点击左侧的图标,上传刚解压的 v4_citas90_cnty_pgn_utf_stats 文件夹中后缀为.shp 文件。将 CSVs 拖至.shp文件之上,且置于上方。

(7)点击 CSV 文件的属性,在 labels 状态下勾选 label this layer with,选择下拉的 name,然后在下方设置颜色和字体大小。

(8)在 QGIS 菜单上方的地图链接中导入 Google 或 Bing 地图。路径为:plugins—manage and install plugins—open layers —Web-openlays plugin—googlemap—googlephysics。

如果底图使用卫星地图,则其可视化呈现将又会是另一种效果,图略。

除了 QGIS 外,制图者还可以免费利用 CartoDB 网站制作作家的行迹和活动地点图。其步骤和方法如下。

(1)在 https://cartodb.com/上注册。

(2)登录后,点击右侧的红灯,选择 your dashboard,然后选择 new map。

(3)点击 connect dataset,上传带有 name、X、Y 三个字段的 excel 数据表。

(4)在 dataview 中点击 the_geom GEO,选择经纬度 XY 栏,然后就可以 MapView(预览)了。右边选项框可以设置参数。

(5)将制作好的地图在网上保存或发布,也可以另存到本地电脑。

二、利用 CBDB 及上述地理信息系统软件将作家的社会关系地理分布可视化

CBDB 是“China Biographical Database Project”的简称,中文名称为“中国历代人物传记数据库”(网址为:http://isites.harvard.edu/icb/icb.do? keyword=k16229)。该项目也是由哈佛大学包弼德教授主持,合作单位有北京大学中国古代史研究中心和我国台湾地区“中央研究院”历史语言研究所。“中国历代人物传记数据库”是目前世界上最大的中国历史人物传记资料分析数据库,迄今上线的中国历代人物已有 36 万人之多,此外,中国地方志等其他数据源中近 50 万人的数据还在源源不断地添加中。该数据库不仅能查找人物的生卒年、字号别名、籍贯、科举仕进等最基本的传记资料信息,而且还可以查找人物的亲属关系、社会关系等,其籍贯等历史地名均有经纬度的数据。该数据库目前也是免费开放的,使用者可以在线查询或将数据库(access)下载到本地电脑查询。

譬如,我们要了解汤显祖的亲属关系和社会关系,即可通过 CBDB 的亲属关系和社会关系查

找功能获得相关数据。图 2 即为 CBDB 的线下查询界面。

一个人的社会关系网络包括了各种类型的社会类别，如其中的“学术”关系，就包括了师生关系、学术交往、主题相近、学术成员、学术襄助、文学艺术交往、学术攻讦等；“政治”关系，则包括了官场平等关系、官场下属关系、官场上司关系、官场奥援、荐举保任、政治对抗等关系。这些关系是计算机按事先设定的关系的关键词从海量的文本里抓取出来的，故可能有人的眼力所不及的有价值的数据，但也有的数据并不能反映一个人实际的社会交往。譬如，A 的集子流传到 B 地，B 地的 C 看到了 A 的集子，有可能会在文章里发表对 A 的集子的阅读感受，于是 A 和 C 的关系自然被计算机捕捉到了。当然，A 和 C 存在一定关系，但 A 和 C 在实际生活中并没有交往。所以，CBDB 里搜索出来的社会关系不全是实际的社会关系，这需要使用者对搜索结果进行鉴别。最好的办法就是结合作家的年谱，筛选出较亲密、较重要且有实际交往的社会关系人员。图 3 是 CBDB 的社会关系网络查询界面。



图 2　CBDB 线下查询界面



图 3　CBDB 线下社会关系网络查询界面

表 2 就是笔者结合 CBDB 查询和徐朔方先生的《汤显祖年谱》制作的汤显祖社会关系经纬度表。其中经度 X 和纬度 Y 的数据有的是 CBDB 自动生成的，有的是根据“中国历史地理信息系统”查寻补入的。

表2 汤显祖社会关系经纬度

NameChn	AddrChn	X	Y	NameChn	AddrChn	X	Y
陈于陛	南充	106.080 7	30.798 99	袁应祺	兴化	119.835 3	32.934 97
戴洵	奉化	121.406 9	29.651 66	张居正	荆州	0	0
冯梦祯	秀水	120.753 2	30.767 47	张四维	蒲州	115.919 9	38.682 15
顾宪成	无锡	120.297 7	31.574 61	张位	新建	115.897 7	28.674 9
顾允成	无锡	120.297 7	31.574 61	赵南星	高邑	114.611 5	37.604 76
胡桂芳	金溪	116.776 3	27.910 08	赵用贤	江阴	120.266 1	31.908 77
胡应麟	兰溪	119.478 9	29.204 45	朱长春	乌程	120.099 3	30.864 96
姜士昌	丹阳	119.569 9	31.995 8	邹元标	吉水	115.132 2	27.214 37
李维桢	京山	113.116 9	31.024 82	臧懋循	长兴	119.901 4	31.013 89
李贽	晋江	118.589 9	24.909 64	李化龙	长垣	114.682 7	35.198 36
刘应秋	吉水	115.132 2	27.214 37	张凤翼	长洲	120.618 6	31.312 71
龙宗武	泰和	114.894 9	26.792 1	张献翼	长洲	120.618 6	31.312 71
罗汝芳	南城	116.627 4	27.559 72	彭兴祖	长洲	120.618 6	31.312 71
梅鼎祚	宣城	118.742 5	30.946 94	钱希言	常熟	120.733 8	31.646 58
欧大任	顺德	113.253 9	22.847 86	顾大章	常熟	120.733 8	31.646 58
沈懋学	宣城	118.742 5	30.946 94	钱谦益	常熟	120.733 8	31.646 58
申时行	长洲	120.618 6	31.312 71	刘芳誉	陈留	114.524 5	34.673 2
沈思孝	嘉兴	120.753 2	30.767 47	姜士昌	丹阳	119.569 9	31.995 8
谭纶	宜黄	116.210 2	27.546 39	钟宗望	东莞	113.749 8	23.046 62
汤凤祖	临川	116.351 3	27.984 78	叶向高	福清	119.381 4	25.727 92
汤会祖	临川	116.351 3	27.984 78	陈邦瞻	高安	115.372 3	28.425 6
汤良祖	临川	116.351 3	27.984 78	刘天虞	高陵	109.080 5	34.533 33
汤儒祖	临川	116.351 3	27.984 78	袁宏道	公安	112.226 5	30.057 53
汤尚贤	临川	116.351 3	27.984 78	袁宗道	公安	112.226 5	30.057 53
汤显祖	临川	116.351 3	27.984 78	袁中道	公安	112.226 5	30.057 53
屠隆	鄞县	121.542 7	29.866 32	马犹龙	固始	115.671 1	32.183 54
王弘诲	定安	110.318 1	19.702 47	王一鸣	黄冈	114.865 5	30.446 99
王汝训	聊城	115.987 5	36.446 72	石昆玉	黄梅	115.934 9	30.079 2
王世懋	太仓州	121.098 6	31.451	乐石帆	嘉兴	120.753 2	30.767 47
汪镗	鄞县	121.542 7	29.866 32	岳元声	嘉兴	120.753 2	30.767 47
谢杰	长乐	119.518 8	25.959 84	何晓	江山	118.615 9	28.735 31
虞淳熙	杭州右卫	120.168 6	30.294 13	李至清	江阴	120.266 1	31.908 77
余有丁	鄞县	121.542 7	29.866 32	谢廷谅	金溪	116.776 3	27.910 08

续表 2

NameChn	AddrChn	X	Y	NameChn	AddrChn	X	Y
高应芳	金溪	116.776 3	27.910 08	张汝霖	山阴	120.578 3	30.004 52
郭惟贤	晋江	118.589 9	24.909 64	陆梦龙	山阴	120.578 3	30.004 52
张大复	昆山	120.948 2	31.386 11	叶干	遂昌	119.263 5	28.587 89
曾如春	临川	116.351 3	27.984 78	李三才	通州	120.854 6	32.010 47
周宗镐	临川	116.351 3	27.984 78	达观禅师	吴江	120.637 8	31.167 07
帅机	临川	116.351 3	27.984 78	高攀龙	无锡	120.297 7	31.574 61
徐良傅	临川	116.351 3	27.984 78	邹迪光	无锡	120.297 7	31.574 61
丘兆麟	临川	116.351 3	27.984 78	张师绎	武进	119.952 3	31.782 78
汤维岳	临川	116.351 3	27.984 78	汪应蛟	婺源	117.844 6	29.244 73
周弘祖	麻城	115.031	31.180 92	丁此吕	新建	115.897 7	28.674 9
朱尔玉	南丰	116.529 9	27.214 31	汤宾尹	宣城	118.742 5	30.946 94
祝世禄	鄱阳	116.663 8	28.994 17	张岳	余姚	121.152 8	30.049 07
袁世振	蕲州	115.345 2	30.062 5	吕胤昌	余姚	121.152 8	30.049 07
黄汝亨	仁和	120.168 6	30.294 13	孙如法	余姚	121.152 8	30.049 07
卓发之	仁和	120.168 6	30.294 13	董裕	乐安	115.831 8	27.426 29
王思任	山阴	120.578 3	30.004 52				

有了带经纬度的数据后，就可以利用 ArcGIS、QGIS 及 CartoDB 等软件或网站制作人物社会关系的地理分布图了。其制作方法与人物的行迹和活动地点图制作方法类似，这里不再罗列操作步骤，用 ArcGIS 制作出来的效果如图 4 所示（由于地图出版规定，此处不展示底图，只给出相对位置图）。

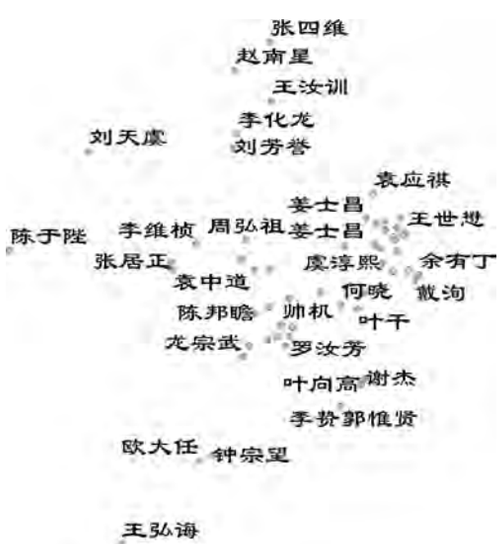


图 4　汤显祖社会关系地理分布图

三、利用CBDB、GEPHI等数据库和软件将人物的社会关系以点线的方式可视化

有了从CBDB获得的社会关系数据,经过编辑加工后,就可以利用GEPHI将人物的社会关系可视化。Gephi是一款用于各种图表和网络的可视化探测软件,是开源和免费的(https://gephi.org/)。文史工作者可以用来分析人物的社会关系,将结果以点线的方式可视化呈现。不过,该软件需要JAVA 1.7语言的工作环境,电脑里需事先安装JAVA 控件。

用GEPHI来展示人物的社会关系需要两张表,一张是节点表(nodes),一张是边表(edges)。节点表包含ID(序号)和Label(即人物姓名)两个字段,边表则包含Source(源)和Target(目标)两个字段,边表主要显示人物的对应关系,是一对多的关系。在表中,主要是用ID来表示对应关系。以汤显祖为例,其节点表和边表分别如表3、表4所示。

表3 Nodes表

ID	Label	ID	Label	ID	Label	ID	Label	ID	Label	ID	Label	ID	Label	ID	Label
1	曹学佺	17	顾宪成	33	李贽	49	钱谦益	65	汤儒祖	81	王世贞	97	袁宏道	113	张岳
2	曾如春	18	顾允成	34	梁辰鱼	50	钱希言	66	汤尚贤	82	王思任	98	袁世振	114	赵南星
3	陈邦瞻	19	郭惟贤	35	刘芳誉	51	丘兆麟	67	汤维岳	83	王锡爵	99	袁应祺	115	赵氏(屠隆母)
4	陈思进	20	何晓	36	刘天虞	52	申时行	68	汤显祖	84	王一鸣	100	袁中道	116	赵用贤
5	陈文烛	21	贺灿然	37	刘应箕	53	沈懋学	69	唐长孺	85	王稚登	101	袁宗道	117	钟宗望
6	陈于陛	22	胡桂芳	38	刘应秋	54	沈明臣	70	屠本畯	86	谢杰	102	岳元声	118	周弘祖
7	陈与郊	23	胡应麟	39	龙膺	55	沈思孝	71	屠隆	87	谢廷谅	103	臧懋循	119	周履靖
8	达观禅师	24	黄汝亨	40	龙宗武	56	石昆玉	72	万世德	88	谢肇淛	104	张大复	120	周宗镐
9	戴洵	25	江盈科	41	陆梦龙	57	帅机	73	汪道昆	89	徐良傅	105	张凤翼	121	朱尔玉
11	董裕	27	乐石帆	43	吕胤昌	59	孙如法	74	汪镗	90	徐学谟	106	张佳胤	122	朱长春
10	丁此吕	26	姜士昌	42	罗汝芳	58	宋世恩	75	汪应蛟	91	许自昌	107	张居正	123	祝世禄
12	冯梦祯	28	李化龙	44	马犹龙	60	谭纶	76	王百谷	92	杨氏(屠隆妻)	108	张汝霖	124	卓发之
13	傅光宅	29	李三才	45	茅维	61	汤宾尹	77	王衡	93	叶干	109	张师绎	125	邹迪光
14	高攀龙	30	李惟寅	46	梅鼎祚	62	汤凤祖	78	王弘诲	94	叶向高	110	张四维	126	邹元标
15	高应芳	31	李维桢	47	欧大任	63	汤会祖	79	王汝训	95	余有丁	111	张位		
16	顾大章	32	李至清	48	彭兴祖	64	汤良祖	80	王世懋	96	虞淳熙	112	张献翼		

表4 Edges表

Source	Target	Source	Target	Source	Target	Source	Target	Source	Target	Source	Target
68	6	68	9	68	12	68	17	68	18	68	22
68	23	68	26	68	31	68	33	68	38	68	40
68	42	68	46	68	47	68	53	68	52	68	55

续表 4

Source	Target	Source	Target	Source	Target	Source	Target	Source	Target	Source	Target
68	60	68	62	68	63	68	64	68	65	68	66
68	71	68	78	68	79	68	80	68	74	68	86
68	96	68	95	68	99	68	107	68	110	68	111
68	114	68	116	68	122	68	126	68	103	68	28
68	105	68	112	68	48	68	50	68	16	68	49
68	35	68	117	68	94	68	3	68	36	68	97
68	101	68	100	68	44	68	84	68	56	68	27
68	102	68	20	68	32	68	87	68	15	68	19
68	104	68	2	68	120	68	57	68	89	68	51
68	67	68	118	68	121	68	123	68	98	68	24
68	124	68	82	68	108	68	41	68	93	68	29
68	8	68	14	68	125	68	109	68	75	68	10
68	61	68	113	68	43	68	59	68	11	71	12
71	90	71	81	71	126	71	53	71	83	71	106
71	73	71	26	1	46	71	115	71	92	71	54
71	37	71	105	71	85	71	70	71	30	71	13
71	34	71	5	71	23	71	4	71	68	71	39
71	58	71	69	71	96	71	125	71	72	71	77
71	25	71	10	71	119	71	50	71	45	71	88
71	7	71	76	71	91	71	1	71	21		

这两个表导入到 GEPHI 里后，就会产生汤显祖由点线关联的社会关系图，其效果如图 5：

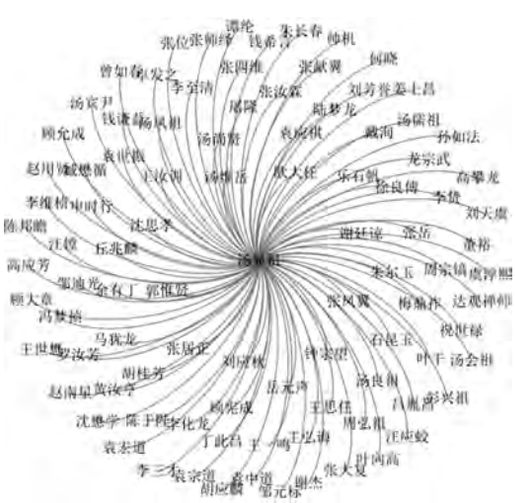


图 5　汤显祖社会关系点线表达图

GEPHI 不仅能产生一个人物的点线社会关系图,而且还可以产生两个到多个人物群落的点线关系图。图 6 是汤显祖和明代另一个戏曲家屠隆的人物群落的点线关系图。图 7 是汤显祖、屠隆和汪道昆三人的社会关系群网络。

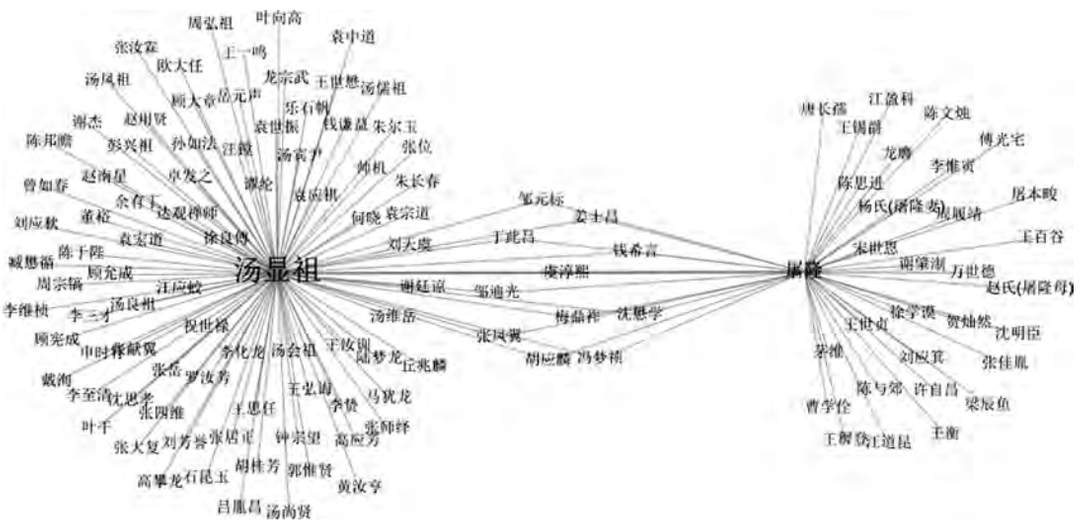


图 6　汤显祖与屠隆社会关系网络点线表达图

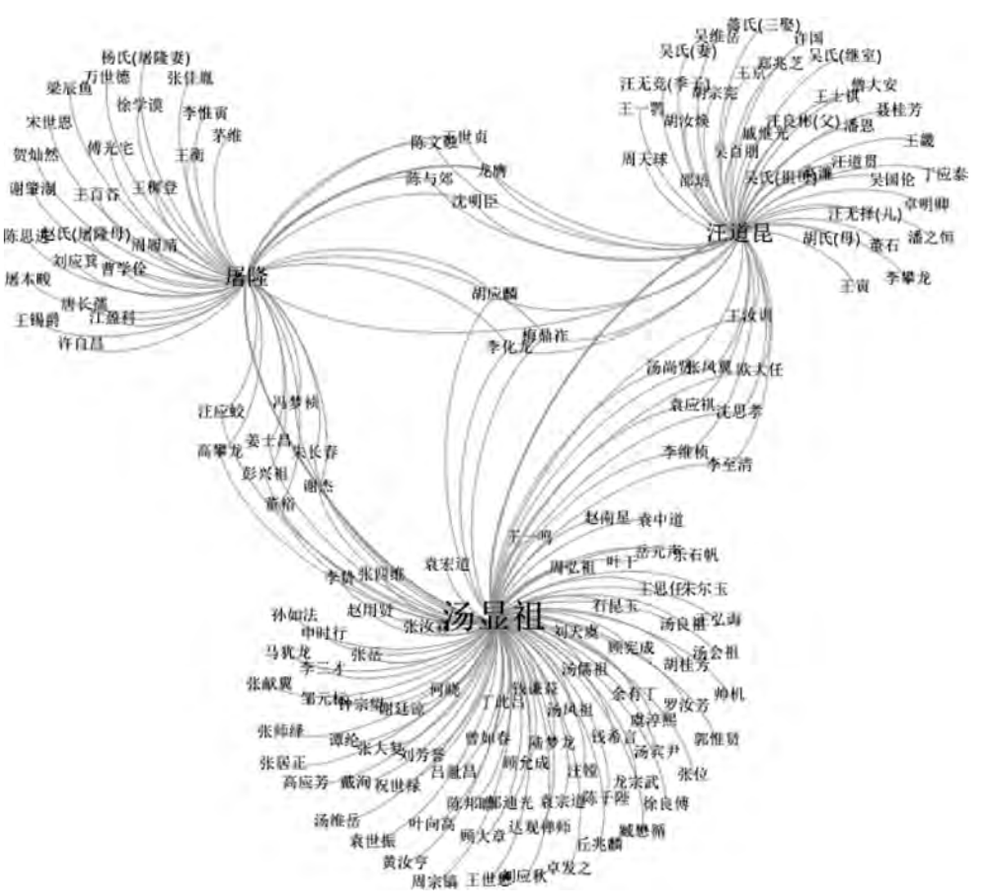


图 7　汤显祖、屠隆和汪道昆三人的社会关系网络图

通过点线的方式展示作家的社会关系网络，则作家自身的社会关系网、彼此间共同的相识者就一目了然了。

用点线表示数据间彼此关系的软件还有 UCINET、Nodexl、Pajek 等，因篇幅所限，这里不再介绍。

四、结　语

通过上面与可视化有关的数据库和软件的介绍，可以知道文学研究的可视化一是需要数据库的支撑，二则需要较好的软件。文史数据库的建设需要有前瞻的眼光，需要有精通计算机的专业人才以及长期不断的资金投入。哈佛大学包弼德教授建立的“中国历史地理信息系统”和“中国历代人物传记数据库”，经过十多年的建设，现在功能越来越强大，应用前景也越来越广泛，由于是开源的数据库，我们乐观其做大做强。譬如，关于明以前的矢量化的中国历史地图，我们就希望能早日出现，这样，如果制作某一朝代的作家活动地点图，有当朝的地图作为底图，就更显得真实可靠。另一方面，我们希望国内的学术界在文史数据库建设方面也要有所作为，呼吁有关部门加大中国文史数据库建设的资金投入，不要等到哪一天来开发祖宗留下的“大数据”时，发现有价值的数据库都已被打上了异邦的标签。中国古代文学作品涉及大量的人名、地名、物品、器皿、服饰、动植物等可以可视化的对象，如何将这些事物可视化地呈现，值得我们去研究。在软件方面，上述软件都是西方人开发的，在使用的时候，我们也发现会受到许多限制。譬如，就字体来说，上述软件可供选择的字体就非常有限。就 QGIS 链接的地图来说，可选择的当代地图只有必应和谷歌地图，而没有百度地图。因此，我们也希望中国的软件开发商能够开发出适合中国人使用的可视化软件。

[参　考　文　献]

[1] 徐朔方：《汤显祖年谱》，见《晚明曲家年谱》，杭州：浙江古籍出版社，1993 年。[Xu Shuofang, *The Chronicle of Tang Xianzu*, in *The Chronicles of Dramatists in Late Ming Dynasty*, Hangzhou: Zhejiang Classics Publishing House, 1993.]

（本文英文全文请参见《浙江大学学报(人文社会科学版网络版)》，http://www.zjujournals.com/soc/CN/article/downloadArticleFile.do? attachType=PDF&id=11330）

Some Visualization Approaches to the Study of Classical Chinese Literature: A Case Study on Tang Xianzu

Xu Yongming

(*School of Humanities, Zhejiang University, Hangzhou* 310028, *China*)

Abstract: The most commonly-used Western visualization database and software in the study of Chinese culture and history include ArcGIS, QGIS, CHGIS, CartoDB, Worldmap, CBDB, GEPHI, etc, among which ArcGIS and QGIS are cartographical systems working for the analysis of spatial and geographic information. The high cost of the ArcGIS system prohibits its widespread use in China. On the contrary, QGIS is an open-source system suitable for the researchers of culture and history all over the world. CartoDB is a cloud-computing database, a

GIS platform that provides geographical information for display in a web browser.

Chaired by Professor Peter K. Bol of the Department of East Asian Languages and Civilization in Harvard University, CHGIS (China Historical Geographical Information System) and CBDB (China Biographical Database) are two database projects directly related to the study of Chinese culture and history. Working in collaboration with Fudan University, CHGIS has established a database of populated places and historical administrative units in pre-modern China. In the form of a digitalized database of places and administrative units, any sort of geographically specific data related to China can be easily displayed at CHGIS. Working closely with the History Department of Peking University and "Academia Sinica" in Taiwan, CBDB is currently the largest database with biographical information about historical figures in China. It not only displays data on individuals, but also provides data on kin relations, social association, official career, as well as ranks and positions a person held, etc. Launched by the Center for Geographic Analysis at Harvard University, Worldmap is an online platform for visualizing and sharing spatial data around the globe. GEPHI is an interactive visualization and exploration software for all kinds of networks, complex systems, and dynamic and hierarchical graphs. Researchers of culture and history can use this platform to explore an individual's social network, represented by innumerous nodes and edges in the graph visualization.

Taking advantage of the software and database above, one can have access to the geographical references in graph data. For example, locations of a writer's life activities, social network, as well as different types of social associations, all of which can be seen in a visualized spatial distribution and refreshes user's experience. Taking Tang Xianzu as an example, this essay demonstrates how the itinerary, social network and social association of this great dramatist of Ming Dynasty are illustrated in graph visualization.

The approach of visualization in literary study would need support not only from the database but also from suitable software. The study of pre-modern Chinese literature requires data on personal names, place names, objects, vessels, clothing, animals, plants, etc., all of which can be visualized by the visualization software and database mentioned above. However, how to take advantage of the software and database in the study of pre-modern Chinese literature should be further explored. We also hope software developers in China and elsewhere develop more visualization software suitable for the researchers in this field in the future.

Key words: classical Chinese literature; geographic information; Tang Xianzu; visualization; Harvard University; CBDB; CHGIS

