

中央委员共事情况可视化

数据分析和确定选题

本次作业数据是第十九届中央委员的统计数据，包括委员基本信息（姓名、性别、民族、出生地、出生年月、毕业院校、专业背景、现任职务）和履历。每人的履历包括多条记录，其中每条记录包括时间和部分描述性记录，描述性记录包括任职所在地和对应职位。

根据现有数据，我认为比较有意思的探索点如下：

- 1) 由任职时间和任职所在地，可以判断委员之间是否在同一地区任职过，由此判断共事情况。
- 2) 由任职时间和对应职位，可以得出在某一级别的时间跨度，进一步可以比较委员之间升迁快慢。

确定选题：由于时间有限，本次作业选择委员共事情况作为课题进行探究。

数据预处理

- 1) 提取任职所在地：首先，我整理了全国省市县三级行政单位名称，见 location.xlsx，从履历字符串中提取行政区是一个字符串匹配的过程，即：从履历中提取包含的行政区名称。说明：对于履历信息中存在工作地名称不属于标准三级行政区划的情况，如“东北”和“清华大学”，作忽略处理。
- 2) 添加任职时间：为对应工作地添加任职时间，统一精确到年份
- 3) 挖掘委员之间共事次数：对任意两个委员，判断是否在同一地点工作过，对任职过的同一地点，判断是否在同一时间段。注：由于行政单位最高是省级，因此这里的同一地点不包括中央级别。

可视化

设计宗旨：将各委员之间共事情况呈现出来。其实，共事情况相当于一个网络图，结点代表委员，边代表两个人之间的联系。每个人可能与多人共事过，另外，两人之间可能有多次共事，网络图需要恰当的反映这些信息。

设计过程：将一个人共事过的委员的数量用结点的颜色反映，颜色越深，代表与更多的人共事过；将两人之间共事的次数用边的粗细或深浅表示，边的颜色越深，代表两人之间共事次数越多。

可视化结果：

如图 Network.html，各个结点代表委员，颜色从浅黄色到深蓝色表示与其他委员共事次数依次增多，鼠标停留在结点上方可以显示该委员的名字、共事次数、共事过的委员。共事次数较多的委员结点分布在图的中间，次数较少的分布在图的四周。边的深浅表示两人之间共事次数，由于本数据集两人之间共事次数大多为 1 次，所以大部分边的颜色一致。

另外，在图的右上方有交互操作悬浮按钮，包括**放大、缩小、漫游、调整尺度、选择感兴趣区域显示、标签悬浮方式**等，可以适应不同人的显示需求，有较大的探索空间。

讨论

本作品分析：

- **好处：**以网络图的方式清晰显示了各委员之间的共事情况；在人数较多、结点分布紧密的情况下，将共事情况以悬浮窗形式展现出来，有效地解决了文字与结点重叠问题。
- **难点：**本作品在前期数据处理阶段耗时较多，一是整理省市县行政单位名称，二是由于委员履历信息中的任职时间格式有些混乱，尝试了较多匹配方式才将时间提取了出来。另外，可视化方面，如何保证繁多的结点不重叠，也付出了较多尝试。

待探究之处：

- 在表示委员与哪些委员共事过时候，现有策略是采用悬浮窗方式将共事委员名字显示，此处可以添加更多的交互设计：如选定一个委员时，与之相连的委员能够高亮。
- 在数据预处理阶段，各省市县单位的经纬度信息已经整理了出来，可以考虑将个人任职经历以地图可视化出来。

（代码、数据和可视化结果见 GitHub <https://github.com/lamyang/Data-Visualization-Course/tree/master/Project%20%20Co-working%20information%20of%20central%20committees> ）