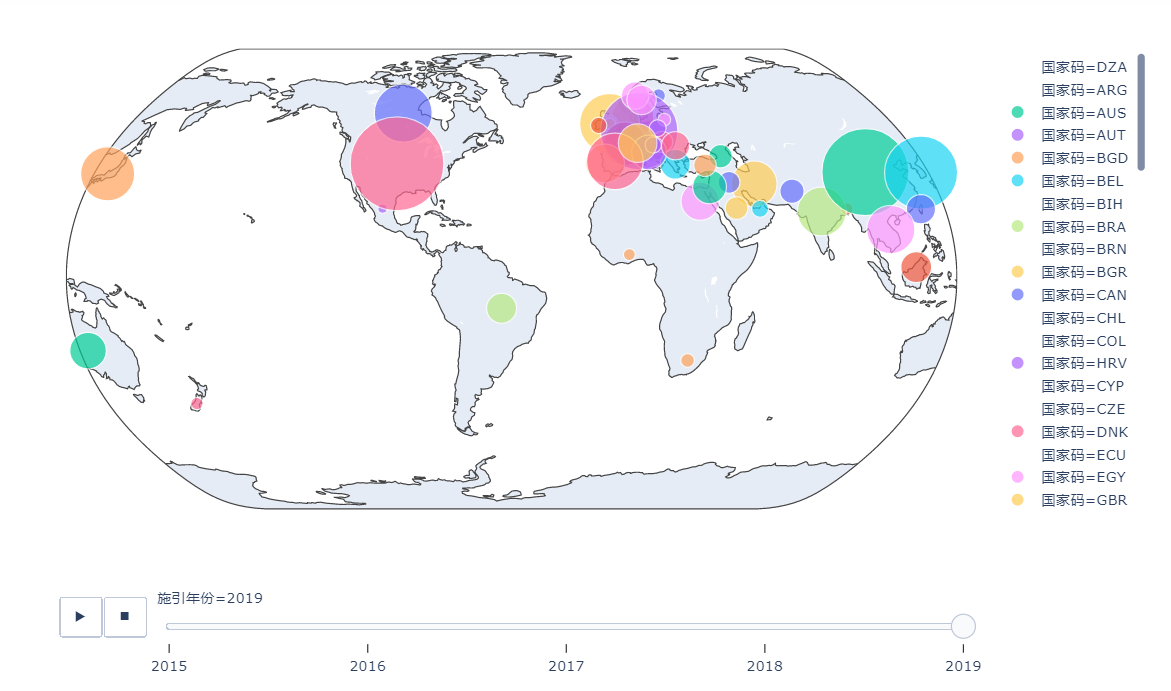
**上期遗留问题**



已解决“动态地图引文信息图只显示少量数据点”的问题：plotly\_express在添加时间维度时，数据点只采集第一年的个数；如果后续年份有新加入数据点，则不会显示。因此，通过在第一年的数据中添加引用数为0的国家，即可规避该缺陷。

[tips]单个封装的函数，其功能不应过多；实现复杂功能，应当分块编写多个函数再调用。否则，极易出错，且排查困难。

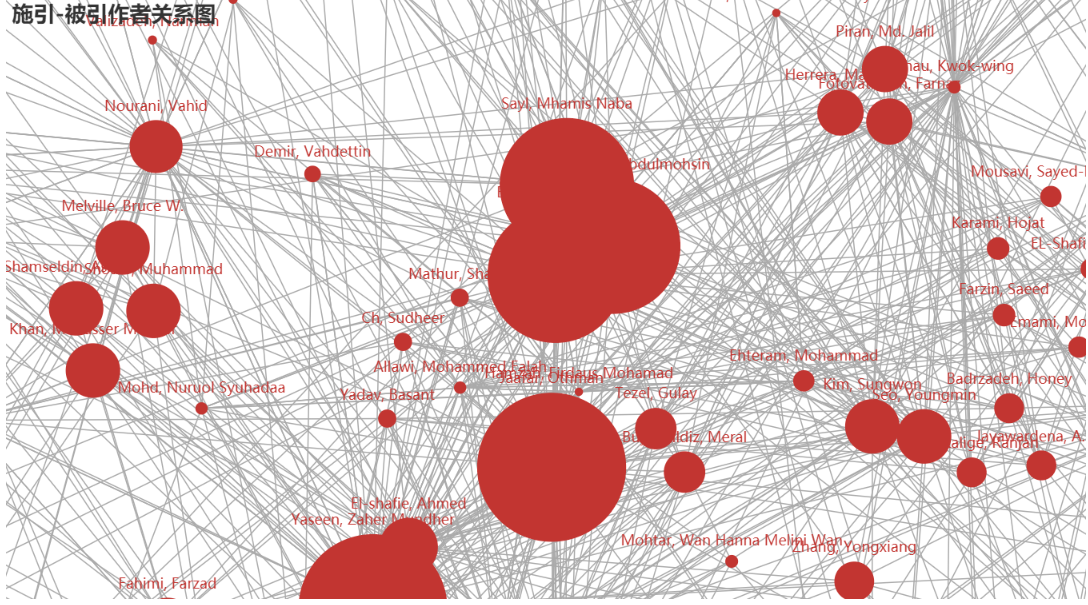
**关于pyecharts**

pyecharts是一个用于生成 Echarts 图表的类库。实际上就是 Echarts 和 Python 的对接。Echarts 是百度开源的一个数据可视化 JS 库，主要用于数据可视化。

pyecharts我们可以作图的类型分别为：Bar（柱状图/条形图），Bar3D（3D 柱状图），Boxplot（箱形图），EffectScatter（带有涟漪特效动画的散点图），Funnel（漏斗图），Gauge（仪表盘），Geo（地理坐标系），Graph（关系图），HeatMap（热力图），Kline（K线图），Line（折线/面积图），Line3D（3D 折线图），Liquid（水球图），Map（地图），Parallel（平行坐标系）Pie（饼图），Polar（极坐标系），Radar（雷达图），Sankey（桑基图），Scatter（散点图），Scatter3D（3D 散点图），ThemeRiver（主题河流图），WordCloud（词云图）。

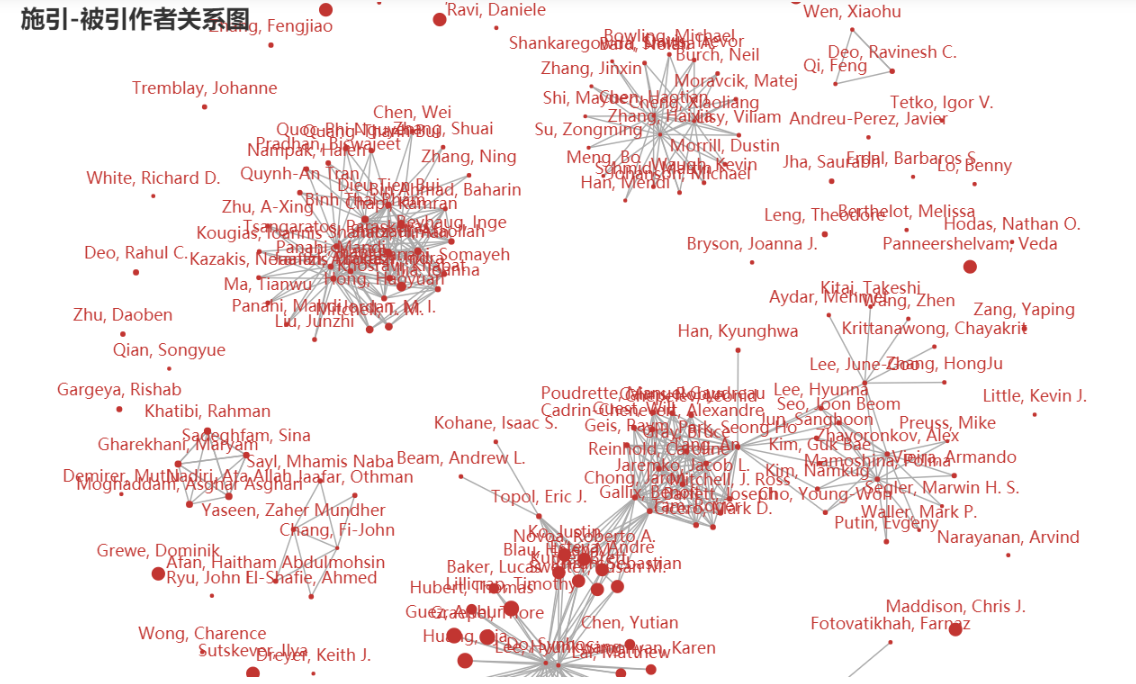
对于引文数据的可视化，可以使用pyecharts中的关系图Graph、地理坐标系图Geo、热力图HeatMap、词云图WorldCloud来绘制。在可视化的过程中，最繁琐、最费时的步骤是数据的结构化与数据清洗。pyecharts的内置函数使用的数据源一般是特定格式的文件，因此需要将引文网络中的每个节点编译为对应格式的数据。完成了数据处理工作，整个数据分析工作也就完成了大半。

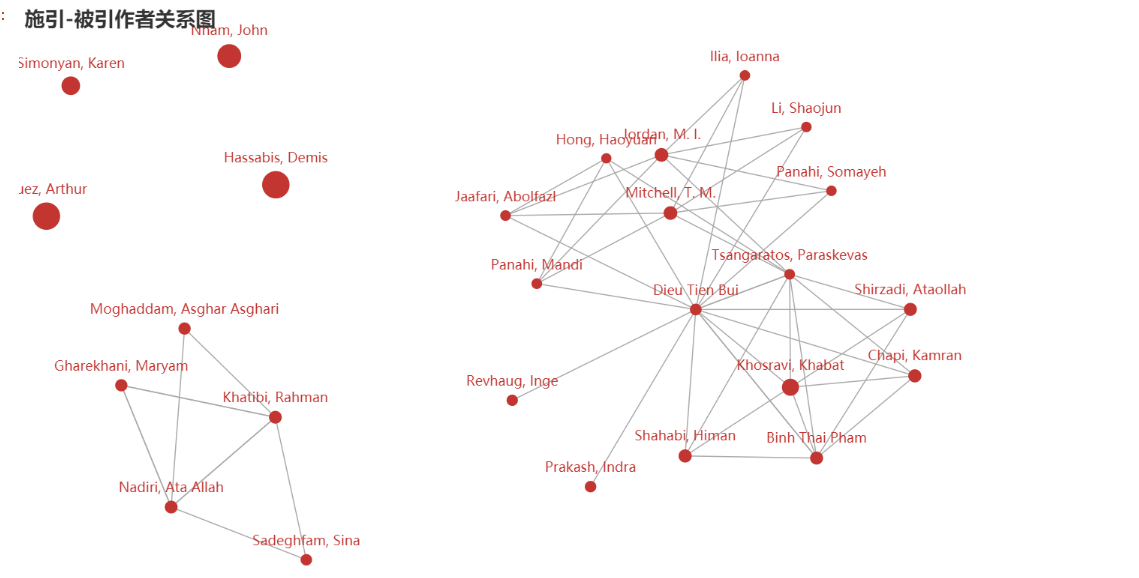
**可视化尝试①：引文网络图（无categories参数）**



如果涉及大量的作者，那么网络图的连线会过多，杂乱的线将使得可视化效果较差。在这种情况下，可以不再显示引用连线，通过聚类算法直接得到一些类内差异小、类间差异大的学者群体。

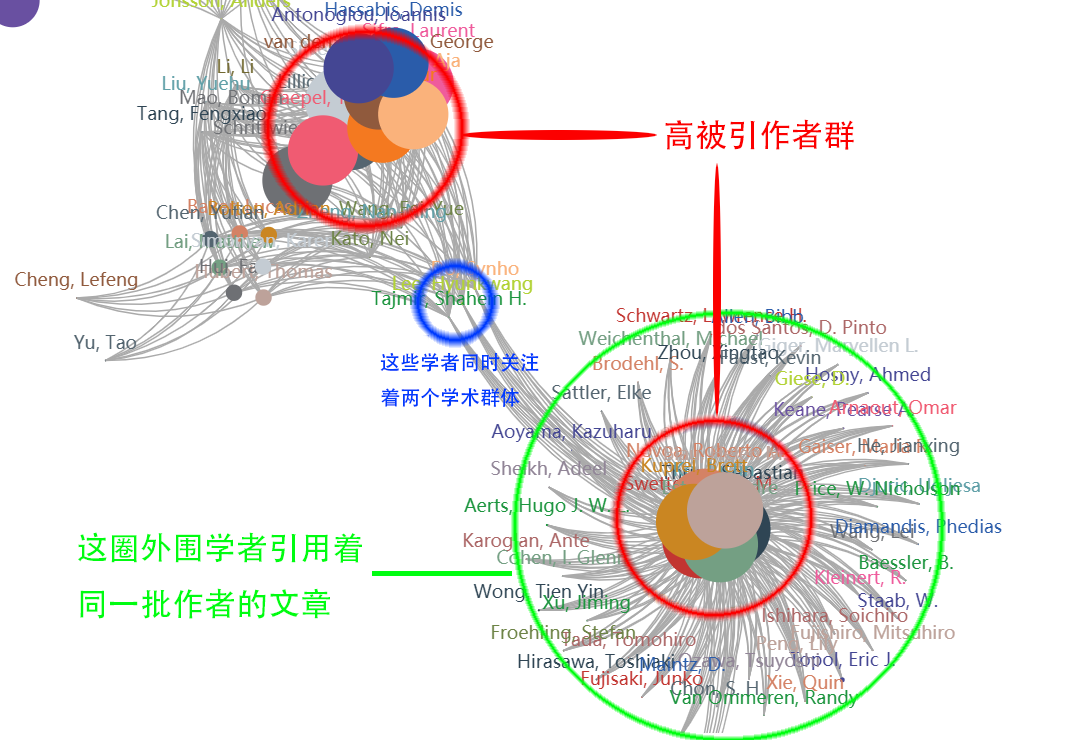
**可视化尝试②：引文网络图（无categories参数）**

下面尝试减少作者的数量，再绘制网络图：  




由于测试用数据源的问题，关系图出现了一些孤立的点。如果使用实际的数据集、确保各个作者间有确实的联系，那么此问题就可以规避了。

**可视化尝试③：引文网络图（有categories参数）**



通过将categories参数传入Graph类图，可以看到引文网络初步形成了肉眼可见的密集群体。但是网络图仍然比较杂乱，而且大量的学者名字占据了太多图表位置，使得可视化表现力很一般。因此，在接下来为使引文网络的效果更好，需要调整categories等参数。