**2020年中国排名前一百大学知识图谱**

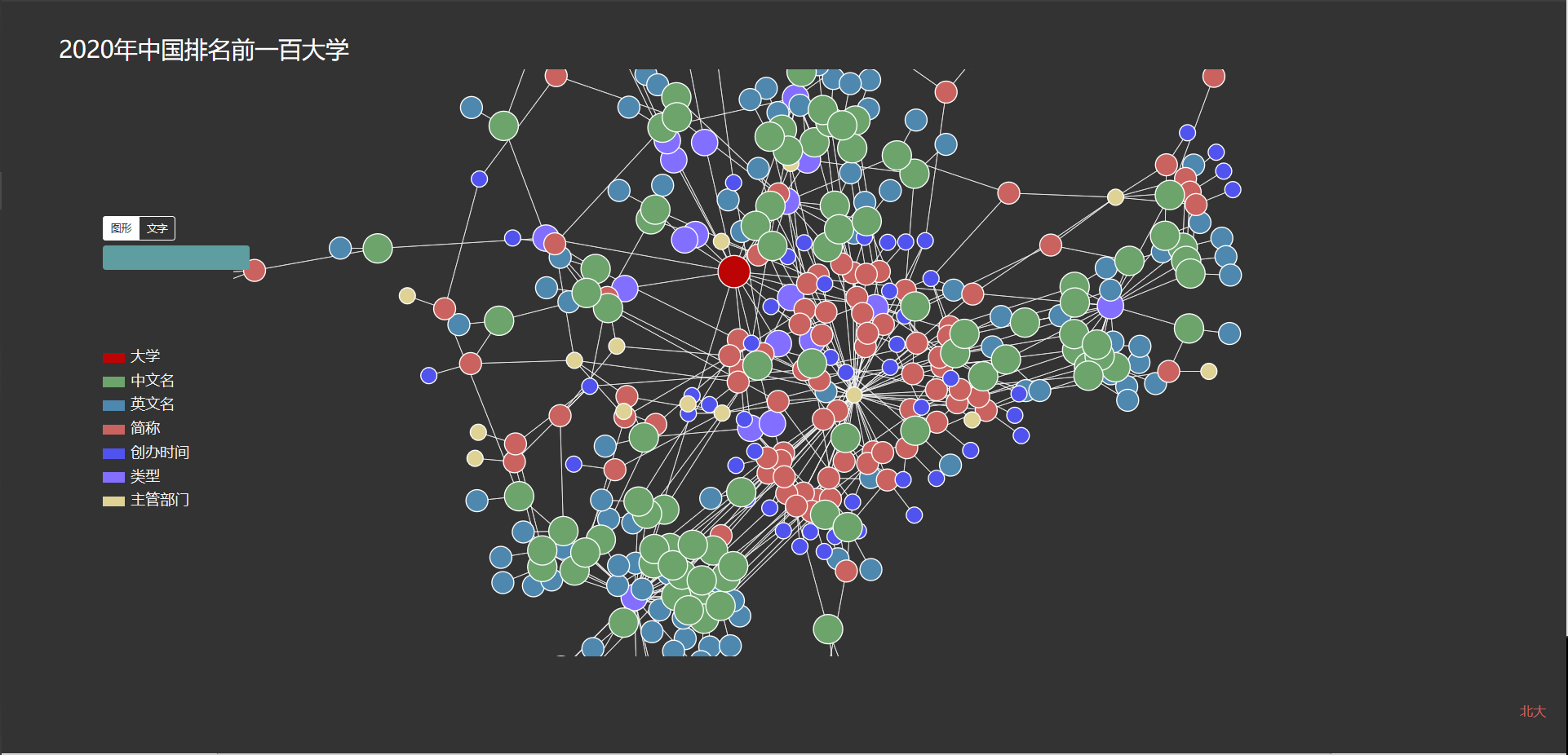
本项目是根据2020年校友会发布的中国前一百大学排行榜来获取中国百强大学的相关信息。提取的主要信息包括有大学的中文英文名称，大学的缩写，大学所属的类型，大学的主管部门和创办时间等等。

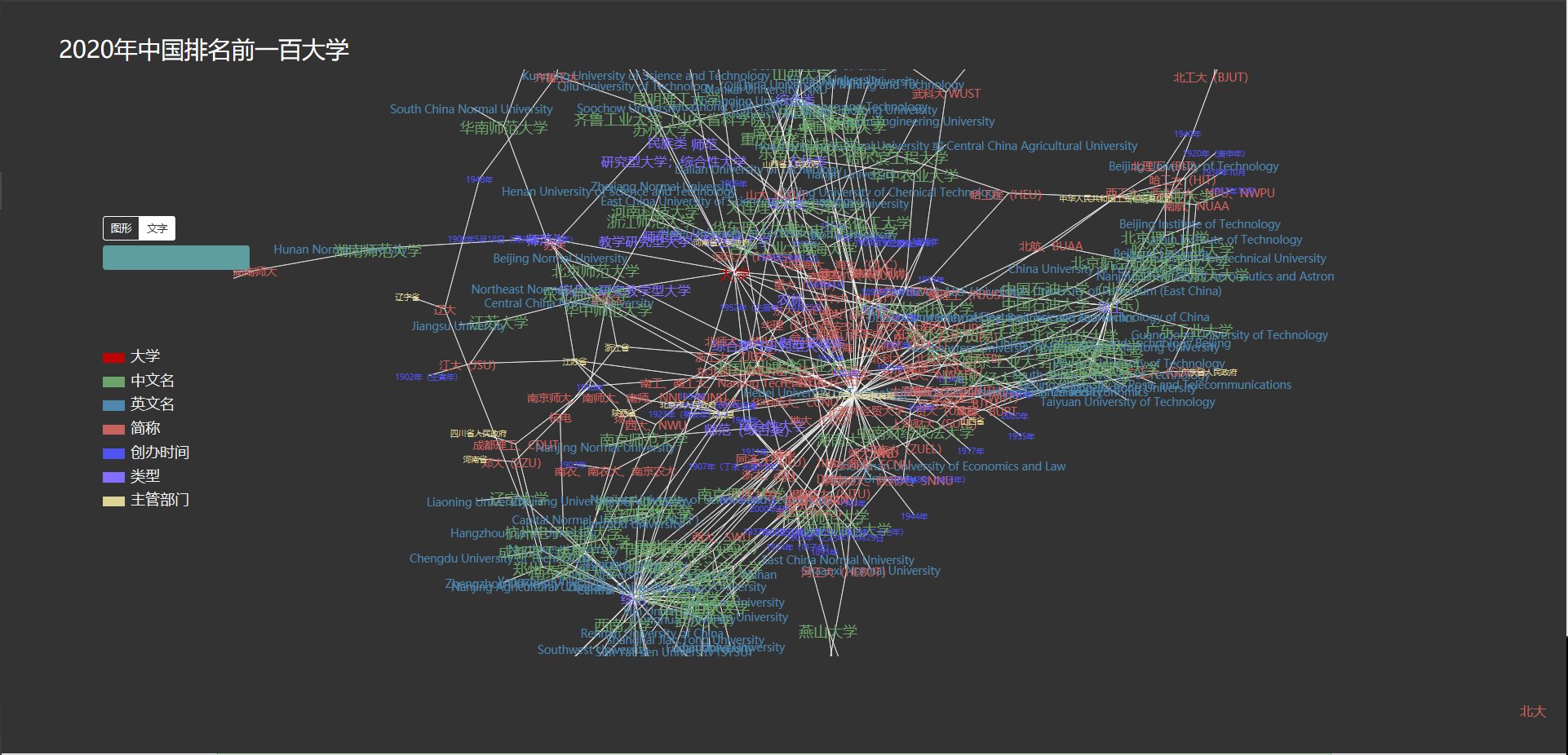
数据的来源主要是来自于百度百科的这一百所大学的相关词条，从网络上获取到100所大学的名称之后通过爬虫技术爬取百度百科资料，从而获得相关信息，主要是在“基本信息”这一项目板块中的信息。



在数据的爬取过程中，不可避免地会出现有部分学校有部分特征缺失，在这边我们使用“无信息”标签进行替代。最后获取到的数据中有综合、师范、理工、农林、财经、政法、民族等7个大类的大学。最后将所有爬取到的数据汇总，存入json和csv文件中。之后将所提取的特征分别取出，存入到不同的txt文件中，为之后建立node和link做准备。

接下来使用D3.js的模板进行数据的可视化。可视化首先就是需要建立node和每个node之间的link，将每个特征的每一个值都提取出来将其作为node，并且对于每个有联系的node，存在关系的实体，我们使用link将其链接起来。在可视化的过程中对于不同类型的node，使用不同的颜色进行标注，link就是简单的使用线段来进行node之间的连接。同时既有图形显示，也可以选择文字信息展示，最后得到的结果如下图所示：





与此同时，此项目还能完成节点的详细信息以及其关联性息的查找，例如如果需要查询“浙江大学”的相关信息，可以在搜索栏中输入浙江大学，同时也能够将鼠标移动到相应的node上，这时候会自动出现相关联的信息以及本身的信息。

