Neo4j

思路：

构建单篇文献的知识基因，每篇文章为节点，其属性有该篇文献的知识基因和发表时间，节点是Subject，Action，Object，Year四个类型。

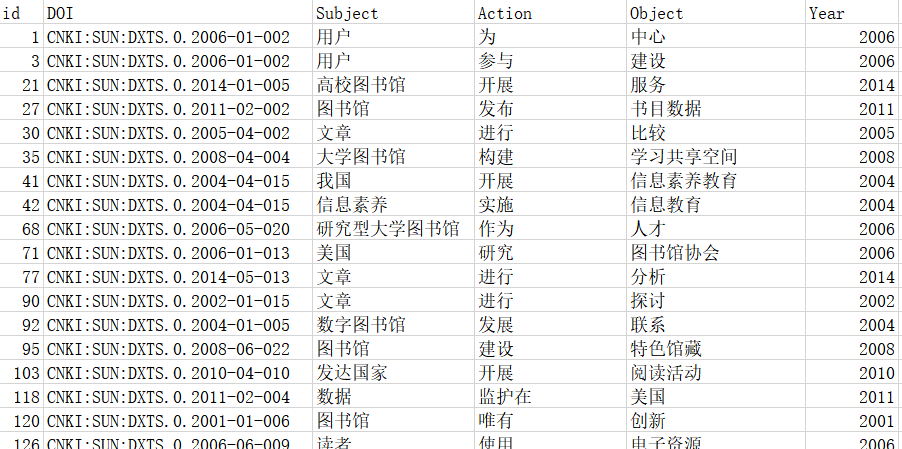
用Neo4j对单篇文献的知识基因进行语义网络组织，期望更清晰地掌握知识基因的横向传播情况，了解知识基因随时间变化状态。

一篇文章可能有多个知识基因，一个知识基因可能对应多篇文章，同一年下可能有多篇文章。

基于上述思路，进行知识基因网络组织，构建知识基因图谱。

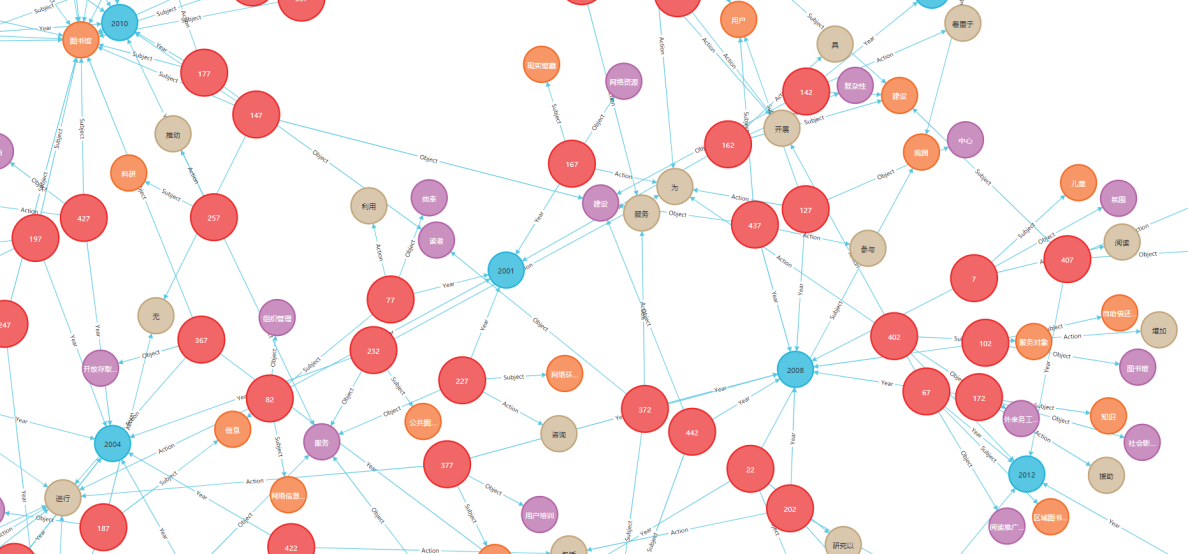
配置：Neo4j 4.1 + Python3.6

输入数据如下，该数据是所有单片文献知识基因和对应的年份，共计17350条：

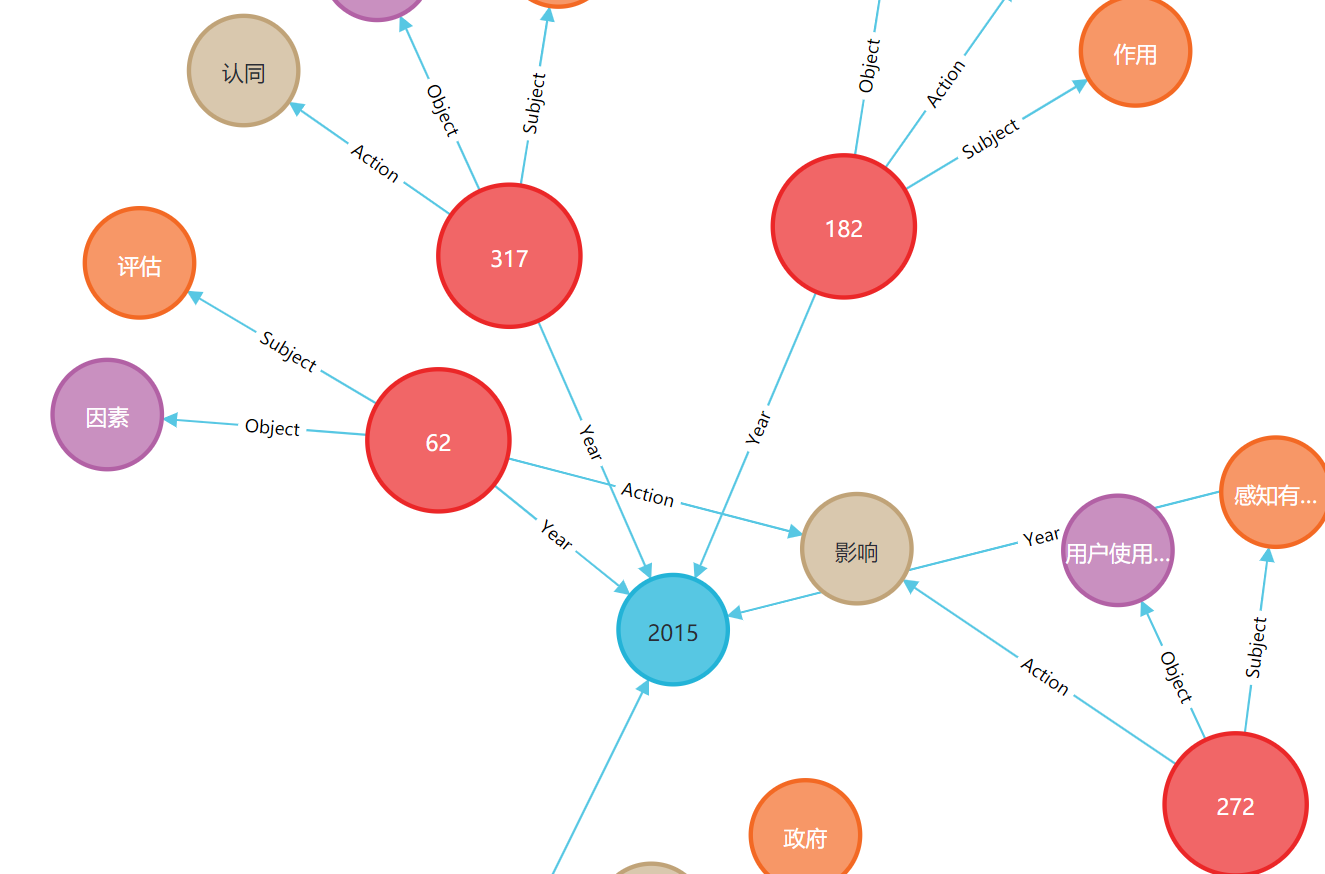


文章用唯一编号DOI表示，每个DOI有S、A、O、Year四个属性，S、A、O、Year分别作为节点建立，关系用Subject、Action、Object、Year表示。

每输入一条数据，分别为其创建DOI，Subject，Action，Object和Year的节点和彼此的关系，遍历所有数据后删除重复节点。最后建立有效节点5027个，关系共计10436条，部分网络如下所示，红色表示文章，用ID代替：



如下图所示，在2015年发表的文章有62号、317号、182号、272号等等，其中62号文章的知识基因之一是“评估-影响-因素”，“影响”一词也是构成272号文章的知识基因的组成部分之一，该文知识基因是“感知有用性-影响-用户使用意愿”。



当构建知识基因图谱后，可以清晰明了地发掘知识基因组成部分的相互关联情况.......

加入时间因素后，能够反映某些时间段的发文数量和知识基因流动情况。

加入单篇文献后，能够观察某个知识基因属于哪些文献，某篇文献拥有哪些知识基因等等。。

上述基因图谱将知识的组织从简单的线性关系转化成为语义网络，相比之下更加直观清晰，通过Neo4j语句可以查询更多语义关系......

限于篇幅和文章重点，没有加入更多因素，如作者分布、被引次数分布、被引文献分布、参考文献分布等等，在后续的研究中，计划结合文献的更多特征属性，进一步探究知识基因图谱的语义组织关系。