第7次上机课任务

颠覆性指一项工作对前人工作的颠覆程度，通常用颠覆性指数（Disruption Index）来衡量。极端情况，当一篇极具颠覆性(D=1)的研究G（称为焦点文献）发表后，后续的研究只引用G而不再引用G的参考文献。颠覆性指数的计算公式请见式 (1) ，其中Ni表示只引用焦点文献的引文数量，Nj表示既引用焦点文献又引用焦点文献参考文献的引文数量，Nk表示只引用焦点文献参考文献的引用数量。D值大于0表示论文更具颠覆性，D值小于0表示论文更具延续性。关于D值的详细信息请参见论文*Large teams develop and small teams disrupt science and technology*

尽管D指数可以很好的评估研究的颠覆性，但是D指数目前还存在一下缺陷

（1）D值具有时间依赖性，一篇论文的D值可能随发表时间改变，例如发表初期一项研究的D值小于0，随着时间的推移，D值大于0，即在发表后的不同时间评价同一项研究的颠覆性可能得到不同的结果；

（2）D值计算难度较大，不仅需要识别目标文献的施引文献，而且需要识别目标文献参考文献的施引文献；

（3）D值基于引文的指标，与引文具有很强的相关性；

（4）D值受参考文献数量的影响，当目标文献引用的参考文献的数量很少时，D值会很大，而目标文献引用的参考文献的数量很多时，D值会很小；此外，D值与目标文献的参考文献的发表时间有关，如果目标文献参考文献中经典文献比例很高，那么式1中的Nk 会很大，会影响目标文献的D

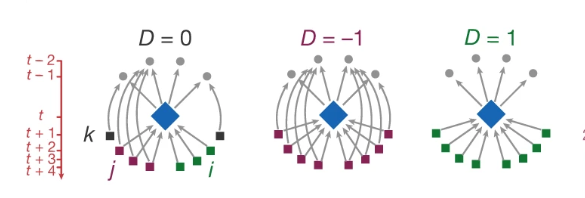


图1 D值计算示意图

(1)

数据：citation relationship.csv中提供了APS收录的19种期刊的引用关系，其中cited id表示被引文献编号，citing id表示施引文献编号，另一个文件中提供了19种期刊发表的所有论文的题录数据。

实验要求：请针对上述提及的关于D指数存在的问题或者可能存在的其他问题，对D指数进行改进。请指出改进后的颠覆性指数解决了当前的哪些问题，并评估改进后的D指数和原始D指数之间的一致性。

Id 表示每篇论文的编号

Doi 表示论文的DOI

Title 表示论文标题

Journal 表示论文发表的期刊名称

Issue表示期刊的期

Volume表示期刊的卷

Papestart表示论文起始页码

Date 表示论文发表的日期

Numpages 表示论文页数

ArticleType 表示文献类型

Authors 表示论文的作者

Year表示论文发表的年份

CiteN表示论文的被引次数

RefN表示论文的参考文献数量