# 区块链领域研究热点及其变迁趋势

## ——基于WOS2016-2020年的文献分析

高 伦，朱思成

南京大学信息管理学院

**摘要：**本研究通过CiteSpace可视化软件对2016-2020年间WOS核心数据库收录的5115篇区块链领域研究文章从摘要、关键词、作者、引用关系等角度绘制知识图谱并进行数据挖掘，分析发现区块链研究的研究热点、变迁趋势等。

**关键词：**区块链；研究热点；变迁趋势；WOS

1. **引言**

区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链是一种防篡改、共享的数字化账本，用于记录公有或私有对等网络中的交易，并由集体维护交易数据的分布式数据库。

2008年，中本聪第一次提出了区块链的概念，在随后的几年中，区块链成为了电子货币比特币的核心组成部分。近年来，世界对比特币的态度起起落落，但作为比特币底层技术之一的区块链技术日益受到重视。区块链在金融、保险、物联网、公共服务、数字版权等诸多领域都得到了广泛的应用。2019年10月24日，在中央政治局第十八次集体学习时，习近平总书记强调，“把区块链作为核心技术自主创新的重要突破口”“加快推动区块链技术和产业创新发展”。

在此背景下， 本文基于web of science数据库，对2016-2020年间国际区块链研究领域相关文献进行计量及对比分析，以期更好地把握该领域的研究热点、变迁趋势及规律。

1. **数据来源与研究工具**

**2.1 数据来源**

从开题目的出发，为了获取区块链领域的最新研究进展，本研究将文献检索范围限定在2016-2020年间的Web of Science核心数据库。检索规则为：TS=("blockchain" OR "block chain" OR "block-chain") AND PY=(2016-2020)。文献类型仅选取研究类文献Article，最终得到文献2294篇。

获取方法：Github上的wos\_crawler工具是一个wed of science核心集合爬虫（链接：https://github.com/tomleung1996/wos\_crawler）。该工具支持爬取任何合法高级检索式的检索结果，并将结果保存为Plain text格式，最后将爬取结果解析导入数据库，包含文献的基本信息（标题、摘要、关键词、被引量等）以及作者机构，分类，参考文献等信息。笔者通过使用此工具，获取了5538篇文献，再通过citespace的文献数据处理功能，分类出文献类型为Article的2294篇文献。

**2.2 研究工具**

**2.2.1 CiteSpace**

CiteSpace是一种多元、分时、动态的指数图谱绘制工具，因其可视化的突出特点，受到了国内学者们的高度关注并得到了广泛应用，截止2020年4月，仅CNKI就有近3800篇文献应用了CiteSpace计量工具。

CiteSpace的分析原理主要分以下几类。

其一，共被引分析（Co-citation analysis）。论文引用其他论文可以看作知识重组产生新知识的过程；而论文被引用可以看作这个过程的延续。学科领域就是在这样的引用和被引用中不断产生新的知识和突破，渐渐就会形成引文网络。1970年，由美国情报学家Small和苏联情报学家Marshakova几乎同时提出了文献共被引的概念，它是指两篇及以上的文献同时被其他文献所引用，则这两篇文献构成共被引关系；同时引用这两篇文献的文献篇数被用来衡量共被引关系的强度，强度越大，表示这两篇文献的相似程度越大，关系越密切。实现共被引分析需要先从文献信息中归纳得到引证矩阵，在此基础上通过矩阵原理生成共被引矩阵，再通过图论的原理将共被引矩阵进行网络化、可视化。

其二，共词分析（Co-word analysis）。专业的学者，可以很好地提炼出最能代表其作品的关键词来反映其作品内容，同时，学者标注其作品关键词时也会受到领域内其他重要成果的影响而学习并使用相似的关键词来标注。词频是指某词语在文献中出现的次数，针对词频的文献分析可以通过从文献中提取词频较高的词语并结合关键词来研究领域热点和发展动向。共词分析是在词频分析的基础上进行更高层次的分析，其基本原理是对一组词两两统计其在同一组文献中出现的次数，以此衡量其关联程度。其基本步骤与共被引分析类似，先从文献中得到关键词信息，形成矩阵，在此矩阵基础上通过矩阵数学原理生成共词矩阵，最后可视化，形成共词网络。关联程度高的多个词语的结点，便会形成更大的聚类点，便于研究人员对领域热点以及趋势把握分析。

其三，聚类分析。俗语说“物以类聚，人以群分”，自古以来，人类就知道依靠经验和知识来进行分类，聚类分析依靠更科学的数学算法将分类的过程定量化。聚类分析又称群分析，是以相似性为基础的分析方法。许多聚类算法都依靠欧几里得或者曼哈顿距离来度量点与点之间的相似性。在一个聚类中的点之间比不在同一聚类中的点之间具有更高的相似性。CiteSpace的聚类算法的依据是谱聚类算法，这是一种基于图论的算法，会将样本看作结点，样本的相似度看作带权的边，如此一来，聚类的问题就会演变为图的分割问题，其对类似共引网络这种关注链接关系甚于结点本身的聚类要求有着天然的优势，这是因为一般的K-Means等聚类算法都是建立在凸球星样本空间上的，一旦样本控件非凸，算法结果就会只关注部分而失去对全局的把控。图论的方法很好地解决了这一问题，使得谱聚类能够识别任意形状的样本空间并收敛于全局最优解。

其四，高频与高中介中心性。高频，代表文献拥有较高的被引量，说明其是某个甚至多个领域的重要知识基础。中介中心性，指的是一个结点担任其他两个结点之间最短路的桥梁的次数。拥有高中介中心性的文献代表其与多篇文献形成共被引关系，其与多篇文献均有关联，是领域内的联系枢纽，也是领域内的重要文献。若一篇文献同时具有高频性和高中介中心性，则可以代表领域内该时期的研究热点。

本文基于CiteSpace知识图谱，对最终检索得到2294篇文献进行文献计量分析，包括共被引分析，共词分析等，并绘制知识图谱，分析区块链领域的研究热点及趋势，以了解近五年国际区块链技术研究的最新进展，掌握其发展的方向和趋势演变特征。

2.2.2 HistCite

HistCite是一款功能强大的引文分析工具，可以快速绘制出某个研究领域发展的脉络，快速锁定领域的重要文献。

LCS（Local Citation Score）是指某一文献被本地数据集中的文献所引用的次数。导入HistCite的文献都与设置的检索词有关，因此可以认为这些文献都属于本文的研究方向，若一篇文献被本地数据集中的文献引用次数较高，则可以说明它在这个领域内的重要程度较高；若文献的LCS值在本地数据集中数一数二，则意味着此文献很可能是该领域的开山之作，因为之后其他文献都与其有关。

LCR（Local Cited References）是指本地被引参考文献数，也即某一篇文献的参考文献在本地数据集中的数量。LCR值的大小可以作为衡量一篇文献是否为关注该领域的重要文献，因为一篇文献引用本地数据集中的文献数量越多就越能说明其十分关注本领域的研究动向，从该文献很可能找到领域发展的新方向。

在Histcite中导入区块链领域的本地数据集，得到文献的LCS值可以作为衡量文献在领域内的重要程度甚至开创程度；得到文献的LCR值可以作为发现文献在领域内发展的新方向的依据。本问通过该工具对2294篇文献进行引文网络分析，绘制出文献之间的引用关系网络，计算具有较高被引量的文献的LCS和LCR，寻找区块链领域近五年重要程度较高的文献和发展方向更新颖的文献，并据此分析区块链领域的研究热点和变迁趋势。

2.2.3 文献计量在线分析平台

文献计量在线分析平台（<https://bibliometric.com/>），是一个集文献总量，合作关系，影响分析，关键词，引用关系等功能为一体的在线数据分析平台。该平台获得过2013年中国科学院国家科学图书馆“科研教育开放信息创新应用大赛”三等奖，其分析结果直观，图形美观大方。

该平台通过web端服务，以图形化的方式，协助研究人员对引文数据进行文献计量分析。简单，易上手，是其对比其他文献计量分析工具最大的特点，用户只需上传指定格式的数据集即可一键获得分析结果，从大体上把握领域的研究动向，避免了参数调节等复杂工序。

本文通过导入平台指定的utf-8格式的数据，分析了多个国家近五年在区块链领域发文的数量情况、国家间的合作关系、关键词的占比情况以及数量变化趋势，以求分析不同国家地域在区块链领域研究的热点问题和发展趋势。

1. **研究过程与结果分析**

**3.1 年度发文量**

论文的发表数量是衡量某领域学术发展水平和科研成果的重要指标。图1是2016-2020年各年份文献发表数量，从图中可以看出：在2016年，研究者对于区块链的研究较少，全年仅有120篇。而从2017年开始，区块链技术在实际应用领域取得了广泛的成功，文献发表数呈现迅猛增长的态势，在2019年达到一个极大值2289篇。鉴于本文撰写时间（2020年4月），2020年份发文量仅有392篇。

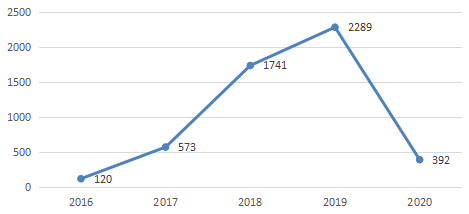
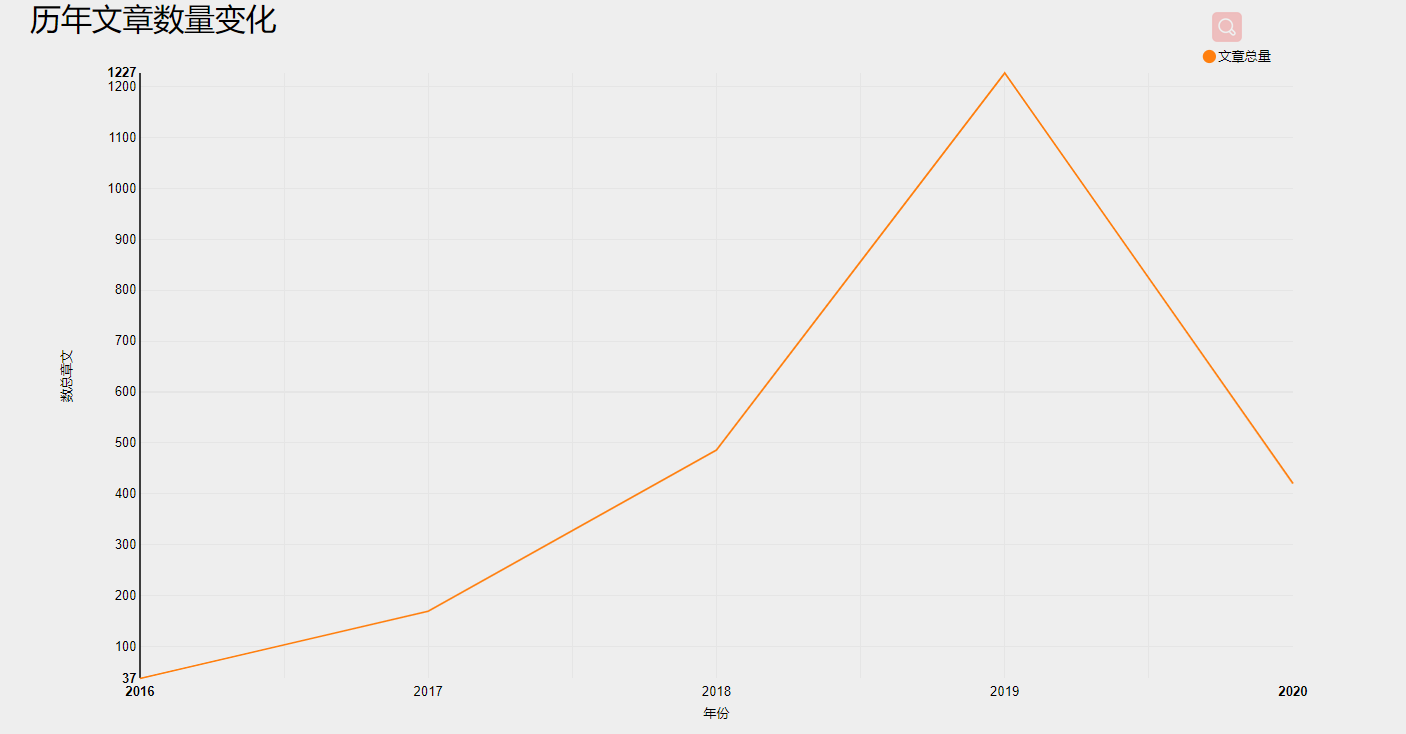


图 1 区块链研究各年份发文量



**3.2 研究热点**

研究热点是指在某一特定时间段内，被许多研究者所一致关注并研究探讨的话题，从知识理论的角度看，中心性和高频词的关键词可以作为代表。中心性作为衡量节点权力的大小，反映了该点在网络中的重要性。关键词的共现频次越高，点中心性越高，说明节点在该领域越重要。

**3.2.1 关键词词频**

如表1所示，区块链研究文献中出现频次较高的关键词有“”

表1

CiteSpace的关键词聚类功能可以明确某研究领域的热点和发展趋势。在知识图谱中，关键词节点用圆形来表示，节点大小与对应主题出现的频次正相关。节点年轮颜色及厚度表示出现时段，色环越后说明改颜色对应年份出现的频次越高。将数据导入CiteSpace中，（这一段写具体处理过程），得到区块链领域的研究热点，如图2所示。

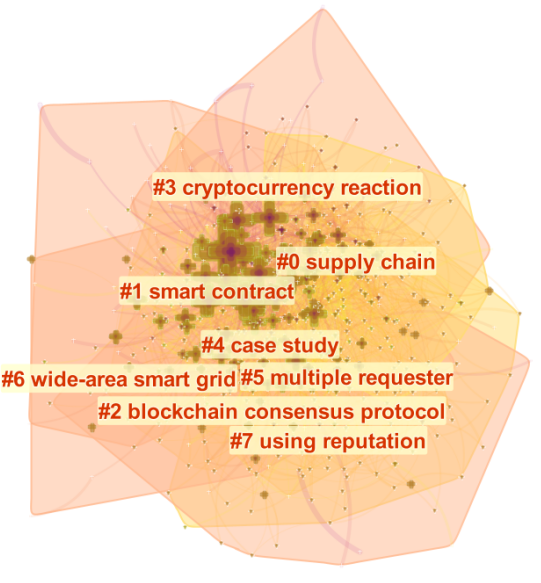


图2

**3.2.2 摘要词频**

**3.3 研究趋势**

**3.3.1 关键词**

突现词指在较短时间内出现较多或使用频率较高的词，根据突现词的次品变化可以判断研究领域的前言与趋势。根据CiteSpace的相关功能，得到区块链领域突现关键词及对应的凸显率和被引热度曲线，如图6所示。

“bitcoin”、”digital currency”、”fintech”、”ontology”、”econmics”、”trust”、”finance”、”innovation”是区块链领域的研究热点。其中”bitcoin”的凸显率远超过第二的”digital currency”，主要体现在2016-2018年，”digital currency”和”trust”体现在2016-2017年，”fintech”、”ontology”、”econmics”、”finance”和”innovation”都体现在2017-2018年。这在一定程度上说明，目前区块链研究前沿主要体现在这些领域。

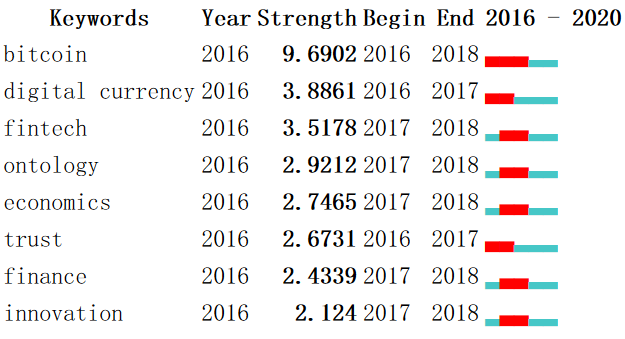


图 6 区块链关键词凸显率

为了更加清晰地了解每个时间段的研究热点，观察发展脉络及研究热点的变化，本研究在聚类图基础上，按时间片段统计了区块链前沿关键词时序图普，如图7所示。

**3.3.2 摘要**

1. **结论与思考**

**参考文献：**

[1] 许振宇,吴金萍,霍玉蓉.区块链国内外研究热点及趋势分析[J].图书馆,2019(04):92-99.

[2] 程豪,张峥.基于CiteSpace分析的区块链技术可视化研究[J].物流科技,2019,42(02):7-11+22.

[3] Chen C. CiteSpace Ⅱ: Detecting and visualizing emerging trends and transient patterns in scientific literature[J].Journal of the American Society for Information Sci-ence and Technology,2006,57(3):359-377.

[4] 王娟,陈世超,王林丽,杨现民.基于CiteSpace的教育大数据研究热点与趋势分析[J].现代教育技术,2016,26(02):5-13.

[5] 陈超美（2016）CiteSpace 的分析原理. 科学知识图谱：前沿与实践. 北京，高等教育出版社.

[6]邱均平,马瑞敏,李晔君.关于共被引分析方法的再认识和再思考[J].情报学报,2008,27(1):69-74. DOI:10.3969/j.issn.1000-0135.2008.01.011.