

区块链和比特币相关主题的知识结构分析: 共被引和耦合聚类分析视角

李牧南^{1,2}

摘要 区块链与比特币作为当前一种重要的新兴技术和数字货币,随着相关应用的拓展,一些理论和应用问题开始得到更多的关注.近年来,区块链和比特币的相关概念已经逐渐扩散到其他的学术领域.但是,对区块链和比特币相关主题的知识结构(Intellectual structure),包括知识基础(Intellectual base)和研究前沿(Research front)的探讨较少,量化研究基本为空白.本文通过对 Web of Science 数据库收录的相关文献进行计量分析,并结合科学图谱分析工具,以较为直观的方式,部分呈现了区块链与比特币相关主题的知识结构,其中包括以共被引聚类分析呈现的知识基础和以耦合聚类分析和关键词共现分析呈现的研究前沿.对于推进区块链和比特币相关主题的知识扩散,以及对未来区块链和比特币的相关学术研究具有一定的参考价值.

关键词 区块链, 比特币, 知识结构, 知识基础, 研究前沿, 共被引聚类, 耦合聚类, 共词分析

引用格式 李牧南. 区块链和比特币相关主题的知识结构分析: 共被引和耦合聚类分析视角. 自动化学报, 2017, 43(9): 1509–1519

DOI 10.16383/j.aas.2017.c160648

Analyzing Intellectual Structure of Related Topics to Blockchain and Bitcoin: From Co-citation Clustering and Bibliographic Coupling Perspectives

LI Mu-Nan^{1,2}

Abstract As an emerging technology and an important digital currency, blockchain and bitcoin have attracted more and more attention from diversified researchers for the past three years. Actually, the conceptual framework of blockchain and bitcoin has increasingly diffused into other research and application areas. Therefore, to analyze the intellectual structure of the related topics to bitcoin and blockchain sounds significant and valuable. However, the relevant research literature based on the bibliometrics keeps the blank. In this paper, co-citation clustering and bibliographic coupling are utilized to analyze and present the intellectual structure. The related result of analysis could be valuable for the next relevant exploration.

Key words Blockchain, bitcoin, intellectual structure, intellectual base, research front, co-citation clustering, bibliographic coupling, co-word analysis

Citation Li Mu-Nan. Analyzing intellectual structure of related topics to blockchain and bitcoin: from co-citation clustering and bibliographic coupling perspectives. *Acta Automatica Sinica*, 2017, 43(9): 1509–1519

2009 年 1 月, 中本聪 (Satoshi Nakamoto) 发布了全球第一款基于区块链 (Blockchain) 技术的数

字货币 — 比特币 (bitcoin)^[1]. 作为一种分布式数据库技术^[2], 区块链在网络加密、数字签名、数字防伪等互联网信息安全领域的优异性能得到了愈发广泛的关注. 作为数字或虚拟货币的比特币一度引起了政府机构的高度关注和部分风险投资者的追捧. 2016 年 6 月, 比特币的全球市值超过了 100 亿美元, 大约占全球所有数字货币 (Digital currencies) 市场的 80%, 但依然与传统货币的价值规模相去甚远.

世界各国对待比特币等数字货币的态度并不统一, 存在一定争议. 但是近年来, 随着互联网金融、众筹、物联网等新兴概念和应用的兴起, 区块链作为比特币实现的基础技术, 逐渐引起了理论和产业界的高度关注. 因此, 针对区块链技术的研究在近年开始从传统的计算机和信息科学领域向经济、金融等

收稿日期 2016-09-12 录用日期 2017-02-03
Manuscript received September 12, 2016; accepted February 3, 2017

国家自然科学基金 (71673088), 广东省软科学研究基金 (2017A070706003) 资助

Supported by National Natural Science Foundation of China (71673088) and Soft Science Research Funds of Guangdong Province (2017A070706003)

本文责任编辑 袁勇
Recommended by Associate Editor YUAN Yong

1. 华南理工大学工商管理学院 广州 510641 2. 广东省创新方法与决策管理系统重点实验室 广州 510641

1. School of Business Administration, South China University of Technology, Guangzhou 510641 2. Guangdong Key Laboratory of Innovation Methods and Decision Management Systems, Guangzhou 510641

领域扩散. 作为一种计算机科学的新兴技术, 区块链技术在很多重要的研究领域都处于起步阶段, 因此有必要对其知识结构进行深入分析, 从而发现其知识基础和潜在的研究前沿, 为后续相关研究提供有价值的参考.

1 文献综述

区块链技术的最大贡献是在没有第三方机构的情况下, 通过分布式节点的验证和共识机制解决了去中心化系统的双重支付问题^[3]. 由于区块链技术具有典型的去中心化、透明性、公平性和公开性特质, 因此可以推广到能源互联网中; 这是区块链技术走向新兴互联网应用的一个重要方向^[4]. 此外, 研究人员还尝试将区块链技术应用到采样机器人的信息传感过程中^[5]. 区块链技术极大解决了交易的信息不对称问题, 提高了交易效率和降低了交易成本. 从货币的基本意义和作用角度看, 数字货币是传统货币在互联网金融风险 and 交易成本递减双重驱动下的必然趋势^[6]. 与此同时, 区块链不仅仅是一种解决交易信任机制的技术, 更是一种解决争端和达成共识的重要手段, 甚至可以称之为区块链思维^[7].

从专业术语的潜层语义分析 (Latent semantic analysis, LSA) 角度看^[8-9], 比特币和区块链无疑是两个语义高度相关的同义词; 尽管比特币可能更多出现于货币、金融和经济领域的研究文献, 而区块链则更多在计算机科学领域出现. 假如把区块链和比特币当成一个整体, 可以认为比特币本质上是区块链技术基础之上的一个重要应用, 而区块链技术本身则具有更广阔的应用前景. 当然, 区块链也存在一定的负面影响. 搭建在区块链之上的加密货币, 例如比特币就曾经被一些不法使用者用来购买一些违禁药品和毒品, 较早支持比特币在线支付交易的境外著名电子商务网站 Silk Road 就因为涉嫌非法交易毒品和其他违禁药品被美国联邦调查局查封^[10]. 但是, 由此而将负面影响全部归结为区块链技术和比特币或许还需要进一步探讨.

无论是加密货币、数字货币或者虚拟货币, 都是互联网经济发展的一个自然衍生物, 并且对传统中心化的记账和信用鉴权体系、交易和支付系统、甚至金融系统都存在一定的颠覆式创新效应. 因此, 以比特币为代表的数字货币给传统的金融和货币监管体系带来了新的挑战和新的激励^[11]. 尽管 2009 年 1 月中本聪才挖出了第一个比特币, 但是 2012 年以来, 比特币的数量增长较快, 而且随着投资价值的逐步涌现, 比特币开始吸引一些投机性资金的介入, 并且由于其严格的交易匿名性和网络信息加密机制, 可能也会在一段时期内产生类似金融泡沫的风险^[12].

根据 2013 年夏天修订的牛津英语大词典, Bitcoin 和 bitcoin 因为首字母的大小写不同而存在显著性差异, 大写字母开头的 Bitcoin 一般代表一种加密货币的网络协议, 而 bitcoin 则是加密货币的一个单位, 即一个比特币^[13]. 因此, 比特币既代表典型技术层面的网络协议, 又代表经济和金融领域的数字货币, 具有典型的语义二元性特征. 尽管比特币代表一种数字货币或者加密货币协议, 但是目前区块链本身和工作量证明机制也存在一定的局限性, 例如比特币目前的开采 (发行) 难度呈现出指数级增长, 即当前单个人和单台普通电脑成功挖出比特币的概率几乎为零, 而且随着参与机器数量和计算性能的增加, 最终可能导致开采成本超过比特币本身的价值, 这也是当前比特币和其他替代数字货币面临的挑战^[14]. 此外, 在并行交易量和并行计算吞吐能力方面, 比特币等数字货币尚未具备全球范围内大规模应用的条件. 但是, 区块链作为一种重要网络共识系统的原型, 开始得到更多相关研究者的关注, 并且其开采 (发行) 难度控制机制也得到了进一步的探讨^[15].

2 研究设计与数据采集

2.1 研究设计

针对一个新兴研究主题知识结构 (Intellectual structure) 的分析具有积极意义, 对于该新兴领域或者主题的知识扩散, 促进学科交叉和更多研究者参与相关领域的讨论和研究具有一定的参考价值. 一般而言, 传统文献计量方法中的共被引分析 (Co-citation), 共词分析 (Co-word) 和耦合聚类分析 (Bibliographic coupling) 等都可以用于对某个研究主题的知识结构分析^[16-17]. 严格来说, 知识结构本身也是一个模糊概念, 其内涵和外延依然存在一定的争议, 但是可以认为某个研究主题的知识结构存在多个不同的维度, 因此通过不同的分析视角 (维度) 都可以部分揭示该主题知识结构, 对于探索该研究主题的研究者而言都具有一定的借鉴和参考意义^[18].

聚类分析是文献计量分析中常见的方法, 主要是由于聚类分析可以采取无监督学习模式, 通过设定一定的距离、相似度公式和聚类迭代阈值, 就可以对文献数据进行分析. 一般而言, 共被引聚类分析采用的距离公式有余弦 (Cosine) 和杰卡德 (Jaccard) 距离等, 如式 (1) 和式 (2) 所示.

$$\text{Cosin}(X, Y) = \frac{\sum_{i=1}^n (X_i \times Y_i)}{\sqrt{\sum_{i=1}^n (X_i)^2 + \sum_{i=1}^n (Y_i)^2}} \quad (1)$$

$$\text{Jaccard}(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} \quad (2)$$

如果说式 (1) 和 (2) 代表了通用的相似度计算公式, 如何构建两个待比较的文本向量则是文献聚类分析的核心, 经典的算法有 TF-IDF (Term frequency-inverse document frequency) 及其各类改进和加权算法, 其中 TF 表示某个具体词或者词组在文献中出现的频率, 如式 (3) 所示, 而 IDF 则可以过滤一些通用词, 如式 (4) 所示。

$$tf_{i,j} = \frac{n_{i,j}}{\sum_{k=1}^j n_{k,i}} \quad (3)$$

$$idf_i = \log \frac{|D|}{1 + |\{j : t_i \subseteq d_j\}|} \quad (4)$$

在式 (3) 中, $n_{i,j}$ 表示第 i 个词 (Term) 在第 j 个文档 (Document) 中出现的频次, $\sum_{k=1}^j n_{k,i}$ 代表第 j 个文档中所有词出现的频次累计。式 (4) 中, $|D|$ 代表当前语料库中所有文档数量, $|\{j : t_i \subseteq d_j\}|$ 表示包含第 i 个词的文档数量, 因为 $|\{j : t_i \subseteq d_j\}|$ 理论上可能为零, 所以作加 1 处理。在实际处理过程中, 还需要考虑进行权重, 例如根据应删除 (Stopwords) 列表对这些词的权重赋零等处理。考虑到共被引网络的节点数和连边数都较多, 导致网络规模较大, 常规的网络拓扑和聚类算法的时间复杂度较高, 因此还可以考虑采用 BH (Barnes-hut) 多核聚类方法, BH 方法的时间复杂度只有 $O(n \log n)^{[19-20]}$ 。

考虑到当前的文献计量工具 (例如 CiteSpacet 和 HistCite 等软件) 均支持较为完善的文献聚类算法和科学图谱功能^[21], 因此采用文献计量工具 CiteSpace 进行相关的辅助分析。然而, 相关文献的检索和数据采集依然需要进行系统化的考虑和进行多次文献检索实验, 以及相关的文献摘要阅读和筛选。

2.2 数据采集

文献检索和数据采集是科学计量的重要基础性工作。理论上说, 随着学科交叉和会聚 (Convergence) 趋势的出现, 很难有一种基于逻辑运算的组合检索式可以找到相关主题的全部文献^[22-23]。此外, 基于 LDA (Latent dirichlet allocation) 为代表的主题 (Topic) 和关键词 (Key words) 的概率分布模型也让传统的信息检索面临一定的挑战^[24]。但是, 传统的组合检索式依然是信息检索领域较为高效的手段和工具之一。一般而言, 基于主题关键词的组合检索式需要征询领域专家的意见, 以及一些定量化的分析手段, 例如采取简单检索方式产生一批样本文献, 然后通过关键词的共现网络分析提取出同义词集合, 同时还要考虑剔除一些通用关键词, 最大程

度保留特征词。基于 Web of Science 数据库, 检索时间为 2016 年 8 月 31 日, 得到如下检索实验结果, 如表 1 所示。

表 1 的检索式主要参考了维基百科, 2013 版的牛津英语大词典以及文献 [1, 3] 的部分表述内容。从表 1 的信息检索实验来看, 当前相关文献中, 会议论文比重较大; 考虑到期刊论文出版时间存在滞后, 因此可以预见, 未来 2~3 年相关研究论文会出现较大增长。在对比了表 1 中的 5 个检索结果, 笔者决定采用第 5 个文献集作为实证研究的数据基础。通过对第 5 个文献集 221 篇文献采取人工摘要阅读的方式进行数据过滤和噪声检测, 最终发现其中有一篇关于 Proof of Work 的论文实际上是研究职业健康问题, 与区块链和比特币的主题显著不相符, 其他 220 篇文献全部符合主题, 这也从一个侧面反映了检索实验的准确性。当然, 也不能完全排除由于作者个人阅读偏好和知识背景所带来的判断偏差。

当然, 对于表 1 中列出的 5 种组合检索式还存在一定的优化空间, 而且随着区块链和比特币相关主题的发展, 也一定会涌现出新的术语和关键词, 但是通过摘要阅读还是基本可以保障所检索文献的相关性, 当然也可能存在个人判断偏好产生的误差。

3 文献计量分析

3.1 相关文献的描述性统计分析

针对 220 篇与区块链和比特币高度相关的文献, 可以发现当前主要的研究群体集中在美国和英国 (英格兰为主) 等欧美发达国家, 如表 2 所示, 文献的类型分布则如表 3 所示。

在表 2 中可以看到, 美国和英国、德国的发文占比远远高于其他国家, 显示在这三个地区研究区块链和比特币的活动相对更为活跃, 瑞士、印度、中国和西班牙占据了第二梯队, 但是与前三名国家的数量差距显著。

在表 3 中可以发现, 在这 220 篇相关文献中占比最大的是计算机科学相关领域, 其次是工程、经济学、通信和金融领域。因此, 当前有关区块链的研究更多还是基于技术和工程视角, 而作为区块链的一个应用产品——比特币则得到了部分经济和金融领域研究人员的关注。此外, 区块链在商业管理领域有 4 篇, 运作管理和自动控制系统各只有 3 篇相关文献; 显然, 当前区块链与运作管理、自动化系统的结合还处于起步阶段, 大部分交叉研究领域基本处于空白。

此外, 从发文的时间序列看, 2013 年是个拐点, 区块链和比特币相关文献自 2013 年以来呈现出突破式增长, 如图 1 所示, 而且 2014 和 2015 两年的文献占比接近总文献数的 70%, 达到了 153 篇。

表 1 不同组合检索式实验结果一览表

Table 1 The different experiments of literature retrieving

序号	组合检索式	文献数量	文献类型分布
1	TS = (“blockchain*”) 索引 = SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, CCR-EXPANDED, IC 时间跨度 = 1986 ~ 2016	30	会议论文: 21 期刊论文: 2
2	TS = (“blockchain*”) NOT TS = (“bitcoin*”) 索引 = SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, CCR-EXPANDED, IC 时间跨度 = 1986 ~ 2016	4	会议论文: 2 期刊论文: 2
3	TS = (“blockchain*” OR “bitcoin*” OR “proof of work” OR “Ethereum*” OR “crypto-currenc*” OR “digital currenc*”) 索引 = SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, CCR-EXPANDED, IC 时间跨度 = 1986 ~ 2016	264	会议论文: 158 期刊论文: 70
4	TS = (“bitcoin*” OR “proof of work” OR “Ethereum*” OR “crypto-currenc*” OR “digital currenc*”) NOT TS = “blockchain*” 索引 = SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, CCR-EXPANDED, IC 时间跨度 = 1986 ~ 2016	234	会议论文: 137 期刊论文: 68
5	TS = (“blockchain*” OR “bitcoin*” OR “proof of work” OR “Ethereum*” OR “crypto-currenc*” OR “digital currenc*”) 精炼依据: 文献类型: (PROCEEDINGS PAPER OR ARTICLE) 索引 = SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, CCR-EXPANDED, IC 时间跨度 = 1986 ~ 2016	221	会议论文: 158 期刊论文: 70

表 2 文献占比最高的前 5 个国家 (地区)
Table 2 Top 5 countries or territories of the literature proportion

国家	论文数量	占比 (%)
美国	63	28.64
英国	23	10.45
德国	21	9.55
瑞士	10	4.55
印度	9	4.09
中国	9	4.09
西班牙	9	4.09

从图 1 可以看出, 自从中本聪在 2008 年 11 月提出了基于区块链技术的比特币协议, 并且于 2009 年 1 月发行第一枚比特币开始, 大约经历了 3 年的静默期. 在 2013 年迎来了学术界的重视, 并且在 2014 和 2015 两年, 相关出版物出现了爆发式增长, 尽管会议论文占比相对较高, 但是可以预见未来几年相关期刊论文也会出现较大幅度增长.

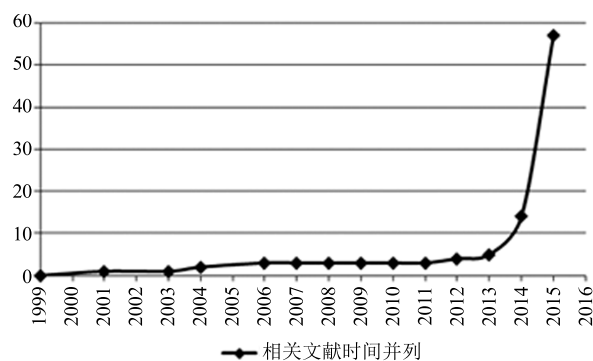


图 1 区块链和比特币相关主题文献的时间序列分析 (1999~2015)

Fig. 1 The time series of the related literature to blockchain and bitcoin (1999~2015)

3.2 相关文献的共被引聚类分析

共被引聚类分析主要是针对主要相关文献的引文网络特征进行分析, 通过引入聚类手段可以部分呈现共被引文献的主题, 进一步阅读这些关键文献, 可以大致推断出该主题文献的主要知识基础 (Intel-

表 3 文献占比最高的前 10 个科目 (Category of WOS)
Table 3 Top 10 categories of WOS ordered by the proportion of literature

WOS 的科目	论文数量	占比 (%)	学科方向
Computer science theory methods	113	51.36	计算机科学
Computer science information systems	93	42.27	计算机科学
Computer science interdisciplinary applications	39	17.73	计算机科学
Engineering electrical electronic	29	13.18	工程 (电气)
Computer science software engineering	28	12.73	工程 (软件)
Computer science hardware architecture	25	11.36	计算机科学
Economics	15	6.82	经济学
Computer science artificial intelligence	12	5.46	计算机科学
Telecommunications	12	5.46	通信
Business finance	11	5.00	金融
Multidisciplinary sciences	9	4.09	综合

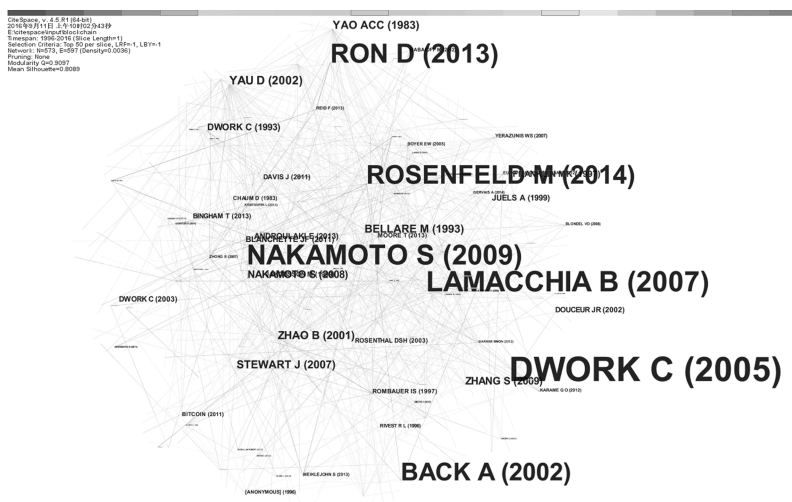


图 2 基于 CiteSpace 的共被引网络图 (中心节点度排序)
Fig. 2 Co-citation networking based on CiteSpace

lectual base)^[21, 25]. 通过借助较为流行的文献计量工具 CiteSpace, 并选择余弦相似度函数和 TF-IDF、极大似然估计等文本处理方法, 得到的共被引网络如图 2 所示, 聚类输出的主要簇如表 4 所示.

图 2 中采取了基于中心度的排序方法, 主要是为了突出在共被引网络中一些共被引频次较低, 但是可能对网络结构和知识流动较为重要的文献节点. 共被引聚类分析在迭代次数 30 000 次的情况下, 总共得到了 34 个簇, 表 4 列出了文献规模最大的 10 个共被引簇, 共被引频次前 10 的论文如表 5 所示.

从表 4 可以看出, 当前与区块链和比特币较为相关的知识基础包括有限理性、比特币协议、信息加密、并发的安全计算、存储复杂性、交易隐匿性、工作量证明机制、拒绝服务攻击和对等网络技术. 其中轮廓系数 (Silhouette coefficient) 是 K -mean

和 K -medoids 等聚类方法常见的评价指标, 其主要工作原理如下:

给定一个数据集 DS , 任意一个数据元素 i 应该与簇 (Cluster) 内其他成员的平均距离 $a(i)$ 保持足够小, 而与其他簇内成员的最短平均距离 $b(i)$ 则尽可能大, 如果所有的元素都满足这个要求, 则聚类之间的边界清晰, 聚类效果良好, 而轮廓系数 $s(i)$ 就是用来衡量这种效果的参数, 显然 $s(i)$ 是 $a(i)$ 和 $b(i)$ 的函数, 具体计算为^[26]

$$S(i) = \begin{cases} 1 - \frac{a(i)}{b(i)}, & \text{若 } a(i) < b(i) \\ 0, & \text{若 } a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1, & \text{若 } a(i) > b(i) \end{cases} \quad (5)$$

表 4 共引网络聚类的文献规模前 10 的簇
Table 4 Top 10 clusters ordered by size in co-citation clustering

ID	文献数量	轮廓系数	主题 (Log-likelihood ration, LLR)	主题 (TF-IDF 加权)
0	67	0.823	mining pool; bitcoin ecosystem; bitcoin protocol	rational zero; bitcoins
1	41	0.962	client puzzle; dos attack; computational power	denial; cryptography
2	37	1.000	set-up assumption; unified framework; concurrent secure computation-both	a unified framework for concurrent security: universal composability from stand-alone non-malleability
3	34	0.999	transaction prefixe; human behaviour offer; empirical inquiry	the case for public work
4	32	0.976	storage complexity; constructing peer-to-peer system; storage-enforcing commitment scheme	difficulty notions; storage complexity
5	28	0.965	effective method; rejection notice; compromising party	spam-combating system
7	24	0.998	hidden internet service; silk road; online media	silk road
6	24	0.982	potential attack; several month; advanced bot	proof-of-work
9	23	0.957	transaction prefixe; human behaviour offer; empirical inquiry	fairnet - how to counter free riding in peer-to-peer data structures
8	23	0.937	bread pudding protocol; security protocol; untrusted computational device	application; work
10	22	1.000	business process; security intent; business process expert	secure business process modelling; SOA system
11	21	1.000	electronic payment system; broad classes; encryption method	multi-analyses electronic payment system

根据式 (5), 轮廓系数的取值范围为 $[-1, 1]$, 并且越靠近 1, 表明该簇的轮廓更加清晰, 聚类效果显著. 此外, 图 2 采取了基于中心节点度的排序方式, 主要是为了突出在图 2 中具有更加显著意义的研究者. 作者共被引网络属于一种典型的社会网络, 在传统的社会网络分析中, 一般认为具有中心性更高的节点往往对于整个社会网络的连通性和知识流动性具有更加重要的作用, 而 PageRank 排序则可以部分反映该作者在网络中的影响力^[27-29].

3.3 文献耦合聚类分析

针对知识结构中研究前沿文献的研究依然是一个存在争议的问题, 因为相对知识基础, 研究前沿

(Research front) 是一个更加动态的概念. 例如, 有部分学者认为文献耦合网络中的中心节点度较高文献属于较为前沿的文献^[30]. 但是, 一些新兴研究, 尤其是一些交叉研究可能并不在网络的主干节点上, 因此基于中心节点度的判断似乎也不够全面. 有研究者基于关联规则挖掘, 结合主题关键词和突现词之间的潜在耦合关系, 提出了一种新的研究前沿探测方法^[31].

在图 3 中, 基于极大似然估计-LLR 的主题聚类分析可以看到一些新兴研究领域, 例如基于区块链的碳排放交易策略以及可再生能源交易平台等, 聚类输出的主要簇如表 6 所示. 图 3 中节点中心度前 10 的关键文献如表 7 所示.

表 5 图 2 中共被引频次 (Frequency of co-citation) 最高的 10 篇文献
Table 5 Top 10 articles with high frequency of co-citation in Fig. 2

序号	论文	频次
1	Nakamoto S. Bitcoin: a peer-to-peer electronic cash system [Online], available: https://bitcoin.org/bitcoin.pdf , 2008.	62
2	Ron D. Shamir A. Quantitative Analysis of the Full Bitcoin Transaction Graph. <i>LNCS</i> , 2013, vol.7859(3-4): 6-24.	24
3	Reid F, Harrigan M. An Analysis of Anonymity in the Bitcoin System. <i>IEEE Third International Conference on Privacy</i> , 2013, vol.1, 1318-1326.	23
4	Barber S, Boyen X, Shi E, et al. Bitter to Better How to Make Bitcoin a Better Currency, <i>LNCS</i> , 2012, 7397: 399-414.	23
5	Meiklejohn S, Pomarole M, Jordan G, et al. A fistful of bitcoins: characterizing payments among men with no names. <i>Proceedings of the ACM Conference on Internet Measurement Conference</i> , 2013, vol.1, 127-140.	22
6	Ober M, Katzenbeisser S, Hamacher K. Structure and Anonymity of the Bitcoin Transaction Graph. <i>Future Internet</i> , 2013, 5(2): 237-250.	17
7	Androulaki E, Karame G, Roeschlin M, et al. Evaluating User Privacy in Bitcoin. <i>LNCS</i> , 2013, 7859: 34-51.	16
8	Moore T, Christin N. Beware the Middleman: Empirical Analysis of Bitcoin-Exchange Risk. <i>LNCS</i> , 2013, 7859: 25-33.	14
9	Miers I, Garman C, Green M, Rubin AD. Zerocoin: Anonymous Distributed E-Cash from Bitcoin. <i>Security and Privacy</i> , 2013: 397-411.	10
10	Christin N. Traveling the silk road: a measurement analysis of a large anonymous online marketplace. <i>International Conference on World Wide Web</i> , 2012, 2(3): 293-304.	10

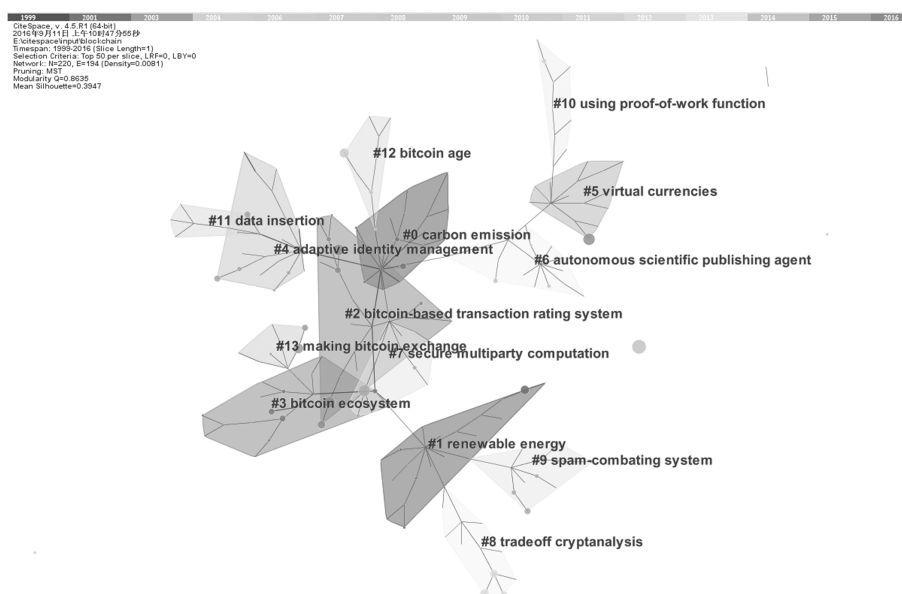


图 3 基于 CiteSpace 的相关文献耦合聚类分析

Fig. 3 Bibliographic coupling analysis based on CiteSpace

表 6 相关文献的耦合聚类分析结果表 (文献规模排名前 5)
Table 6 Top 5 clusters of bibliographic coupling analysis

ID	文献数量	轮廓系数	主题 (Log-likelihood ration, LLR)	主题 (TF-IDF 加权)
0	21	0.984	carbon emission; cryptocurrency system; using p2p network traffic	trading infrastructure model bitcoins
1	20	0.991	renewable energy; smart grid; difficulty control	empirical analysis bitcoin network
2	18	0.97	bitcoin-based transaction rating system; peer-to-peer affine commitment; algorithmic trading	weird trick
3	17	0.935	bitcoin ecosystem; conjoint analysis; future online banking service	empirical analysis bitcoins
4	17	0.972	adaptive identity management; certified bitcoin; making puzzle	cloudlet-based multi-lateral resource exchange framework bitcoin
5	15	0.979	virtual currencies; new phenomenon; financial market	denial

表 7 图 3 中文献节点中心性 (Centrality of node) 最高的 10 篇文献
Table 7 Top 10 articles with high centrality in Fig. 3

序号	论文	中心性
1	Donet JAD, Pérez-Solà C, Herrera-Joancomar J. The Bitcoin P2P Network. <i>The Workshop on Bitcoin Research</i> , 2014, 8438 : 87–102.	1.12
2	Wang LQ, Liu Y. Exploring Miner Evolution in Bitcoin Network. <i>LNCS</i> , 2015, 8995 : 290–302.	0.87
3	Ron D, Shamir A. How Did Dread Pirate Roberts Acquire and Protect his Bitcoin Wealth? <i>LNCS</i> , 2014, 8438 : 3–15.	0.8
4	Dev JA. Bitcoin mining acceleration and performance quantification. <i>IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)</i> , 2014, vol.1, 1–6.	0.6
5	Bamert T, Decker C, Elsen L, et al. Have a Snack, Pay with Bitcoins. <i>IEEE Thirteenth International Conference on Peer-to-peer Computing</i> , 2013, 19 (17): 1–5.	0.51
6	Meiklejohn S, Orlandi C. Privacy-Enhancing Overlays in Bitcoin. <i>LNCS</i> , 2015, 8976 : 127–141.	0.49
7	Herrera-Joancomar J. Research and Challenges on Bitcoin Anonymity. <i>LNCS</i> , 2015, 9229 : 163–180.	0.36
8	Novak M, Skalak P. Virtual currencies. <i>International Scientific Conference on Knowledge for Market Use — Women in Business in the Past and Present</i> , 2015, vol.1, 880–887.	0.34
9	Vasek, Marie, Moore, Tyler. There's No Free Lunch, Even Using Bitcoin: Tracking the Popularity and Profits of Virtual Currency Scams. <i>LNCS</i> , 2015, 8975 : 44–61.	0.33
10	Sapuric S, Kokkinaki A. Bitcoin Is Volatile! Isn't that Right? <i>Lecture Notes in Business Information Processing</i> , 2014, 183 : 255–265.	0.19

表 6 的分析结果是传统文献耦合分析基础上的 *K-mean* 聚类结果. 从轮廓系数看, 聚类的效果较好, 而且规模前 5 的 6 个聚类的文献数量达到了总数的 49%, 具有一定的代表性. 聚类结果也部分反映出基于区块链和比特币的原始框架已经开始产生一些跨学科交叉研究, 例如基于区块链的碳排放交易机制和可再生能源交易机制等. 这些新兴交叉研究领域的出现也部分显示出区块链架构和比特币协议存在较为广阔的应用前景. 而且在碳排放交易和可再生能源市场, 与区块链和比特币相关的研究也只是刚刚起步, 但也指出了一个新的研究方向.

3.4 主题词共现分析

主题词共现分析可以用来对语义相似度进行一定程度的检验^[18, 32]. 引入主题词共现分析的原因有两个: 1) 可以呈现区块链和比特币高度相关的主题词. 2) 可以部分验证本文所采用的检索式是否科学合理, 可以认为是对本文信息检索式的稳健性检验. 相关文献关键词的共现分析 (*Co-occurrence analysis*) 如图 4 所示.

从图 4 可以看出, 与区块链和比特币语义高度相关的共现词包括加密货币、工作量证明机制、网络安全和虚拟货币等. 其中, 关键词共现频次和节点中心性排名前 10 的关键词如表 8 所示.

一般而言, 共词分析中的共现 (*Co-occurrence*) 频次可以部分反映出该术语在网络中的影响力 (*Im-*

pact), 中心性指标可以反映出该术语在网络知识共享和流动中的中介作用^[33-34]. 从表 8 可以看出, 高频关键词包括加密货币、工作量证明机制、数字货币、隐匿性、虚拟货币、加密理论和网络攻击等; 节点中心性较高的关键词包括电子现金、金融投资、非法药品交易 (*Silk road*) 和洗钱等. 此外, 关键词的共现分析也可以从一个侧面显示本文所选择的组合检索式, 以及基于语义相似度的信息检索方式基本合理.

4 研究结论与局限性

区块链和比特币无疑是当前较为热点的研究主题之一. 但是, 通过本文的文献计量发现, 在当前的学术研究活动中, 区块链依然是比特币的一个衍生或者关联概念. 如果将比特币排除在检索式之外, 在 *WOS* 中只能找到 4 篇相关文献, 即使在 *SCOPUS* 数据库中也只能检索不到 20 篇相关文献.

通过关键词的共现分析可以看出, 区块链与比特币的文献语义存在高度相关性. 当前针对区块链和比特币相关主题的研究文献主要集中在近 3 年, 而且会议论文占比为 72%, 期刊论文比例相对偏低. 通过对 *WOS* 数据库中 1986 年以来的 220 篇相关文献进行文献计量分析, 可以发现区块链与比特币相关主题研究主要还是集中在计算机科学领域, 而金融和经济领域占比相对较低, 运作管理、自动控制系统、能源与可持续发展等新兴领域目前还处于起

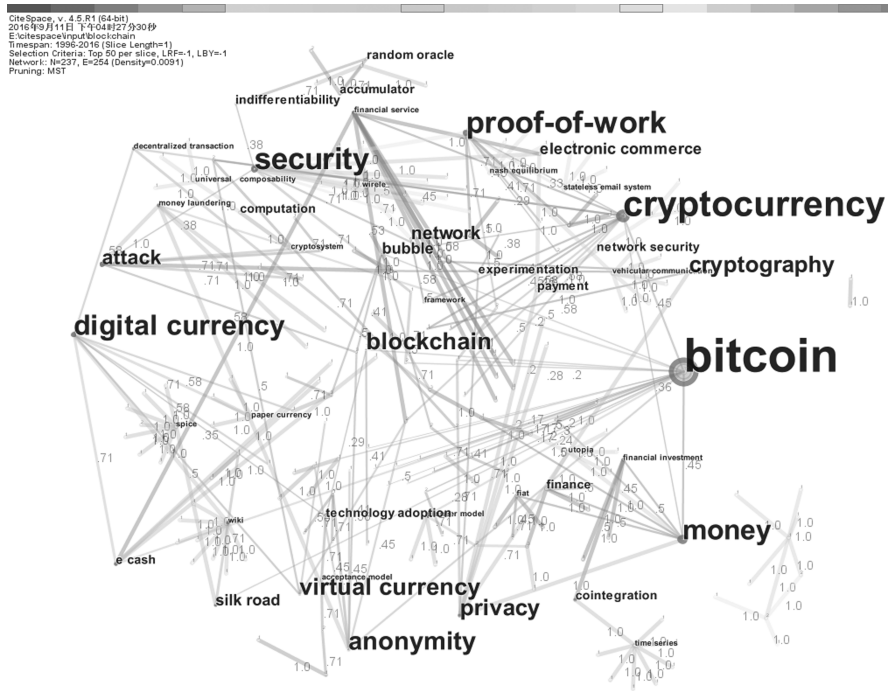


图 4 基于 CiteSpace 的相关文献关键词共现分析 (语义相似度分析)
 Fig. 4 Co-occurrence analysis of the relevant literature based on CiteSpace

表 8 共现频次和节点中心性排名前 10 的关键词
Table 8 Top 10 keywords ordered by frequency of co-occurrence and centrality

序号	共现频次	节点中心性	关键词	序号	节点中心性	共现频次	关键词
1	75	0.62	bitcoin	1	0.62	75	bitcoin
2	23	0.5	cryptocurrency	2	0.5	23	cryptocurrency
3	14	0.38	proof-of-work	3	0.41	11	money
4	13	0.3	security	4	0.38	14	proof-of-work
5	11	0.41	money	5	0.3	13	security
6	10	0.22	digital; currency	6	0.22	10	digital currency
7	8	0.1	anonymity	7	0.21	5	attack
8	7	0.06	blockchain	8	0.16	2	e cash
8	7	0.02	virtual; currency	9	0.15	1	financial; investment
9	6	0.12	privacy	10	0.14	3	silk road
9	6	0.03	cryptography	10	0.14	1	money laundering
10	5	0.21	attack				

步阶段.

通过共被引聚类分析可以发现: 信息加密、并发的安全计算、存储复杂性、交易隐匿性、工作量证明机制、拒绝服务攻击和对等网络等是区块链与比特币相关主题的重要知识基础. 而在基于文献耦合聚类分析中, 可以发现当前区块链的概念已经开始扩散到其他学科领域, 并且呈现出较快的文献增长趋势. 从表 6 的聚类分析结果可以看出, 针对区块链和比特币应用拓展的研究已经成为一个主流方向, 其中区块链和比特币相关概念在碳排放交易和可再生能源市场领域的交叉研究已经形成了一定的局部热点效应, 并且是当前较为新兴的两个主题. 这恰恰反映出区块链技术和比特币的概念具有较强的学科渗透性, 甚至可以扩散到可再生能源和碳排放交易领域^[35-36].

本文也存在一定的局限性: 1) 国内外目前针对区块链和比特币的科技文献计量分析基本为空白, 如何选择合适的主题词簇 (Term clump) 本身就可能存在一定争议; 2) 本文选择的数据集来源于 Web of Science, 也可能需要进一步与来自 SCOPUS 以及谷歌学术的文献数据进行对比分析; 3) 在分析知识网络中的作者影响力方面, 还可以引入节点的 Betweenness 和 PageRank^[27] 等指标进一步深入分析.

总体而言, 本文借助文献计量分析的主流方法探讨了区块链和比特币相关主题的知识结构问题, 分别基于共被引聚类分析部分呈现了相关主题的知识基础, 基于耦合聚类分析和关键词共现分析部分捕捉了当前区块链和比特币的研究前沿. 但是, 无论是知识结构, 还是知识基础和前沿都是一个动态演化的文献计量概念, 都需要在未来一段时间内进行进一步的深入探索.

References

- Nakamoto S. Bitcoin: a peer-to-peer electronic cash system [Online], available: <https://bitcoin.org/bitcoin.pdf>, January 3, 2009
- Blockchain (database) [Online], available: <https://en.wikipedia.org/wiki/Blockchain>, June 1, 2017
- Yuan Yong, Wang Fei-Yue. Blockchain: the state of the art and future trends. *Acta Automatica Sinica*, 2016, **42**(4): 481-494
(袁勇, 王飞跃. 区块链技术发展现状与展望. 自动化学报, 2016, **42**(4): 481-494)
- Zhang Ning, Wang Yi, Kang Chong-Qing, Chen Jiang-Nan, He Da-Wei. Blockchain technique in the energy internet: preliminary research framework and typical applications. *Proceedings of the CSEE*, 2016, **36**(15): 4011-4022
(张宁, 王毅, 康重庆, 程将南, 贺大伟. 能源互联网中的区块链技术: 研究框架与典型应用初探. 中国电机工程学报, 2016, **36**(15): 4011-4022)
- Zhao He, Li Xiao-Feng, Zhan Li-Kui, Wu Zhong-Cheng. Data integrity protection method for microorganism sampling robots based on blockchain technology. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 2015, **43**(S1): 216-219
(赵赫, 李晓风, 占礼葵, 吴仲城. 基于区块链技术的采样机器人数据保护方法. 华中科技大学学报 (自然科学版), 2015, **43**(S1): 216-219)
- Dai Jin-Ping, Li Yan. Will money finally disappear: the future of digital cash. *Nankai Journal (Philosophy, Literature and Social Science Edition)*, 2016, (4): 141-149
(戴金平, 黎艳. 货币会消亡吗? — 兼论数字货币的未来. 南开学报 (哲学社会科学版), 2016, (4): 141-149)
- Swan M. Blockchain thinking: the brain as a decentralized autonomous corporation. *IEEE Technology and Society Magazine*, 2015, **34**(4): 41-52
- Evangelopoulos N, Zhang X N, Prybutok V R. Latent semantic analysis: five methodological recommendations. *European Journal of Information Systems*, 2012, **21**(1): 70-86
- Müller O, Schmiedel T, Gorbacheva E, vom Brocke J. Towards a typology of business process management professionals: identifying patterns of competences through latent

- semantic analysis. *Enterprise Information Systems*, 2016, **10**(1): 50–80
- 10 Van Hout M C, Bingham T. “Surfing the silk road”: a study of users’ experiences. *International Journal of Drug Policy*, 2013, **24**(6): 524–529
- 11 Chu J, Nadarajah S, Chan S. Statistical analysis of the exchange rate of bitcoin. *PLoS One*, 2015, **10**(7): e0133678
- 12 Cheung A, Roca E, Su J J. Crypto-currency bubbles: an application of the Phillips-Shi-Yu (2013) methodology on Mt. Gox bitcoin prices. *Applied Economics*, 2015, **47**(23): 2348–2358
- 13 Böhme R, Christin N, Edelman B, Moore T. Bitcoin: economics, technology, and governance. *Journal of Economic Perspectives*, 2015, **29**(2): 213–238
- 14 Van Alstyne M. Why Bitcoin has value. *Communications of the ACM*, 2014, **57**(5): 30–32
- 15 Kraft D. Difficulty control for blockchain-based consensus systems. *Peer-to-peer Networking and Applications*, 2016, **9**(2): 397–413
- 16 Yan B N, Lee T S, Lee T P. Mapping the intellectual structure of the internet of things (IoT) field (2000–2014): a co-word analysis. *Scientometrics*, 2015, **105**(2): 1285–1300
- 17 Liu Z G, Yin Y M, Liu W D, Dunford M. Visualizing the intellectual structure and evolution of innovation systems research: a bibliometric analysis. *Scientometrics*, 2015, **103**(1): 135–158
- 18 Li M N, Chu Y Q. Explore the research front of a specific research theme based on a novel technique of enhanced co-word analysis. *Journal of Information Science*, 2016, DOI: 10.1177/0165551516661914
- 19 Gan Z C, Xu Z L. Efficient implementation of the Barnes-Hut octree algorithm for Monte Carlo simulations of charged systems. *Science China Mathematics*, 2014, **57**(7): 1331–1340
- 20 Zhang J C, Behzad B, Snir M. Design of a multithreaded barnes-hut algorithm for multicore clusters. *IEEE Transactions on Parallel and Distributed Systems*, 2015, **26**(7): 1861–1873
- 21 Chen C M. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 2006, **57**(3): 359–377
- 22 Zhang Y, Zhou X, Porter A L. How to combine term clumping and technology roadmapping for newly emerging science and technology competitive intelligence: “problem and solution” pattern based semantic TRIZ tool and case study. *Scientometrics*, 2014, **101**(2): 1375–1389
- 23 Li M N. A novel three-dimension perspective to explore technology evolution. *Scientometrics*, 2015, **105**(3): 1679–1697
- 24 Al-Salemi B, Ab Aziz M J, Noah S A. LDA-AdaBoost.MH: accelerated adaboost.MH based on latent Dirichlet allocation for text categorization. *Journal of Information Science*, 2015, **41**(1): 27–40
- 25 Hammarfelt B. Interdisciplinarity and the intellectual base of literature studies: citation analysis of highly cited monographs. *Scientometrics*, 2011, **86**(3): 705–725
- 26 Rousseeuw P J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 1987, **20**: 53–65
- 27 Zhang Q P, Feng Z, Li X, Zheng X L, Zhang L. 25 years of collaborations in IEEE intelligent systems. *IEEE Intelligent Systems*, 2010, **25**(6): 67–75
- 28 Borgatti S P. Centrality and network flow. *Social Networks*, 2005, **27**(1): 55–71
- 29 de Amorim R C, Hennig C. Recovering the number of clusters in data sets with noise features using feature rescaling factors. *Information Sciences*, 2015, **324**: 126–145
- 30 Jarneving B. Bibliographic coupling and its application to research-front and other core documents. *Journal of Informetrics*, 2007, **1**(4): 287–307
- 31 Li Mu-Nan. Probing the research frontiers of competitive intelligence based on the association rules mining. *Journal of Intelligence*, 2016, **35**(3): 54–60
(李牧南. 基于关联规则挖掘竞争情报研究前沿分析. 情报杂志, 2016, **35**(3): 54–60)
- 32 Hu J M, Zhang Y. Research patterns and trends of Recommendation System in China using co-word analysis. *Information Processing and Management*, 2015, **51**(4): 329–339
- 33 Kostoff R N, Schaller R R. Science and technology roadmaps. *IEEE Transactions on Engineering Management*, 2001, **48**(2): 132–143
- 34 Xie P. Study of international anticancer research trends via co-word and document co-citation visualization analysis. *Scientometrics*, 2015, **105**(1): 611–622
- 35 Mihaylov M, Jurado S, Avellana N, Van Moffaert K, de Abril I M, Nowe A. NRGcoin: virtual currency for trading of renewable energy in smart grids. In: Proceedings of the 11th International Conference on European Energy Market. Krakow, Poland: IEEE, 2014. 1–6
- 36 Al Kawasmi E, Arnautovic E, Svetinovic D. Bitcoin-based decentralized carbon emissions trading infrastructure model. *Systems Engineering*, 2015, **18**(2): 115–130



李牧南 华南理工大学工商管理学院副教授, 广东省创新方法与决策管理系统重点实验室副主任, 美国佐治亚理工学院访问学者. 2006 年获得华南理工大学计算机应用技术专业博士学位. 2008 年从华南理工大学管理科学与工程博士后流动站出站. 主要研究方向为文本挖掘与大数据, 新兴技术管理, 信息系统.

E-mail: limn@scut.edu.cn

(**LI Mu-Nan** Associate professor at the School of Business Administration, South China University of Technology, deputy director of the Key Laboratory on Innovation Methods and Decision Management System of Guangdong, and the visiting scholar at Georgia Institute of Technology, USA. He received his Ph. D. degree from South China University of Technology in computer science in 2006, and experienced two-year postdoctoral career in management science. His research interest covers text mining and big data, management of emerging technology, and information systems.)