DOI:10.13530/j.cnki.jlis.173005

关联驱动的领域知识群落生长*

滕广青

摘　要　知识的发展演化一直是图书情报学界重点关注的主题。随着复杂网络理论的复兴,以网络思维探索知识发展过程中结构关系的演化成为学术界的共识。本文以知识间关联关系为基础,对社会化标注模式下 Folksonomy 知识组织模式中领域知识群落的生长展开研究。基于关联频度提取层次网络,利用 k—丛和派系识别知识网络中的松散型与紧密型领域知识群落,从频度、关联、数量、规模、时序多个维度进行交叉复现分析。研究结果表明,领域知识群落的基本生长路径为"关联关系→松散群落→紧密群落";知识间关联关系的增长是领域知识群落生长过程中数量繁衍和规模扩容的保障;关联关系频度和数量的积累是领域知识群落生长过程中核心凝聚的过滤器。知识群落生长模式与规律的揭示,有助于从知识间的互促互扰关系方面拓展领域知识组织视野并把握知识发展脉络。但本研究在维度的划分粒度方面还有待进一步加强。图 3。表 4。参考文献 36。

关键词　知识网络　知识关联　知识群落　Folksonomy

分类号　G254

Correlation-Driven Domain Knowledge Community Growth

TENG Guangqing

ABSTRACT

The evolution of domain knowledge has always been the focus of the library and information academia. Exploring the evolution of structural relationships in the process of knowledge development with network thinking has become the consensus of academia. The evolution process of the domain knowledge communities can be tracked and analyzed in view of the changes of the relationships among knowledge units to reveal the growth pattern and the law of the domain knowledge communities. This study makes an extraction of 2 075 related articles in specific knowledge domains in folksonomy knowledge organization mode with 203 tags and the correlations number 2 953 in total. The time span is from 2005 to 2015 and the timer shaft is divided into 11 units of time-window. The original domain knowledge networks are constructed based on the correlations between tags. Using the frequency of correlation as the threshold, the knowledge networks at level are extracted based on scale-free and fractal theory. The k-plexes and cliques in the

* 本文系国家自然科学基金项目"基于网络结构演化的 Folksonomy 模式中社群知识组织与知识涌现研究"(编号:71473035)和教育部人文社会科学研究规划基金项目"基于后结构主义网络分析的 Folksonomy 模式中社群知识非线性自组织研究"(编号:14YJA870010)的研究成果之一。(This article is an outcome of the project "Study on Community Knowledge Organization and Knowledge Emergence in Folksonomy Based on Network Structure Evolution" (No.71473035) supported by National Natural Science Foundation of China and the project "Study on Community Knowledge Nonlinearity Self-organization in Folksonomy Based on Post-structuralism Network Analysis" (No.14YJA870010) supported by Humanities and Social Science Foundation, Ministry of Education of the People's Republic of China.)

通信作者:滕广青,Email:tengguangqing@163.com,ORCID:0000-0002-1053-0959(Correspondence should be addressed to TENG Guangqing, Email:tengguangqing@163.com, ORCID:0000-0002-1053-0959)

knowledge networks at level are calculated separately, and the loose knowledge communities and close-knit knowledge communities are identified by k-plexes and cliques. On this basis, the time series analysis on tags, correlations, k-plexes, cliques in the knowledge networks is executed from the quantitative dimension. The long-term trends in the evolution of domain knowledge communities are identified. The number of tags, the number of correlations, the number of k-plexes, the average scale of k-plexes, the number of cliques, and the average scale of the cliques in the original knowledge networks and the knowledge networks at level in the key time windows are cross-discovered. The influence of knowledge correlation on the growth of domain knowledge community is thus revealed. The results show that the basic growth path of the domain knowledge community is 'correlation→loose knowledge community→close-knit knowledge community'. Knowledge correlation achieves the connection between the knowledge nodes. When the knowledge correlations are rich to a certain extent, they structure the loose knowledge communities identified by k-plexes. With the further enrichment of knowledge correlations, the close-knit knowledge communities identified by cliques emerge. The increase of the correlations between knowledge units is the guarantee of the quantity multiplication and scale expansion in the growth of domain knowledge communities. Both quantitative evolution analysis and cross-discovery show that in the case of the same scale of the knowledge network, as long as the numbers of correlations continue to increase, the correlations will still lead to more loose knowledge communities and close-knit knowledge communities. At the same time, the new correlations have the opportunities to bridge the old k-plexes or cliques into larger k-plex or clique, resulting in a decrease in the number of knowledge communities and an increase of scale. The frequency of correlation and the accumulation of numbers are the filter of core cohesion during the growth of domain knowledge communities. The knowledge network at level filters to get the significant correlations by threshold. Some once-prominent knowledge nodes and correlations fade out of the knowledge community, and the number and scale of the communities are reduced and condensed. In this article, a knowledge network at level extraction method based on scale and fractal, which extends the identification way of knowledge and correlation in knowledge network analysis, is proposed to provide a new way for knowledge networks to process large-scale data. The correlation-driven domain knowledge community growth patterns in this study can capture the most critical factors in the evolution of knowledge communities in the process of knowledge evolution. Although there is no elasticity in time granularity and dimensions of cross-discovery, the revelation of correlation-driven domain knowledge community growth patterns helps to grasp the evolution venation of knowledge community, which has a positive effect on revealing the law of domain knowledge development. 3 figs. 4 tabs. 36 refs.

KEY WORDS

Knowledge network. Knowledge correlation. Knowledge community. Folksonomy.

0 引言

自从 Popper[1]157-192 在其经典著作《客观的知识——一个进化论的研究》中以进化论的视角洞察和辨析客观知识的发展进程以来，知识的生长与演化就成为图书情报学界一个被长期关注的主题。然而在知识世界内部，任何知识都不是孤立于其他知识而独自成长发展的，即使是那些看上去似乎是一枝独秀的知识也必然有关联知识与之伴生。因此，知识的生长演化过程在本质上是一定范围内知识群落的繁荣与

变迁过程。正因为如此,在领域知识发展的研究中,以结构关系见长的网络思维起到至关重要的作用。特别是随着复杂网络分析理论与方法的复兴,知识发展问题的相关研究也从早期经典的引文知识网络逐渐发展出关键词知识网络、图书知识网络、专利知识网络、标签知识网络等多种形态。尤其是在当今开放网络大众参与的网络现实中,建立在社群集体认知基础上的 Folksonomy 知识组织模式,在舍弃专家精英的同时,也充分地显露出 Web2.0 时代网民群体的智慧[2]。基于标签知识网络的领域知识群落生长研究也自然成为学术界关注的热点问题。

有鉴于此,本研究从知识之间关联关系的角度出发,采用层次网络的 k—丛和派系相结合的网络分析方法,分别对标签知识网络中的松散和紧密型领域知识群落进行识别和挖掘。考虑到知识演化发展过程的多维性以及现实中网络的动态性[3],研究中依循时间序列分析的逻辑线索,将多维相关数据沿时间轴的绕动态势渐次推进。通过多维数据的交叉复现,尝试从知识关联的视角对领域知识群落生长过程中的模式与规律进行探索和揭示。

1 研究综述

回溯图书情报学的发展历史,以网络思维对知识相关问题展开研究,要追溯到两位图书情报学的巨匠,他们于二十世纪五六十年代分别在《科学》(*Science*)杂志上发表了两篇经典之作。一篇是 Garfield[4] 基于引文关系从网络结构的视角诠释了科学知识之间的继承与发扬;另一篇是 Price[5] 通过对引文网络的研究,发现并提出了知识传承中的马太效应。这两篇文章今天已经成为图书情报领域科学计量学的理论基石。进入 20 世纪 70 年代,Popper[1]258-283 的“三个世界”理论从科学哲学的层面进一步以知识客观性、自在性的角度夯实了知识发展问题研究的理论基础,并率先以进化论的视角对知识的发展进行阐述。其后,Belkin 和 Robertson[6] 又在 Popper 的客观知识世界(世界 3)的基础上,将知识本身与知识结构关联起来,即知识的本质就是一种结构,也据此开创了情报学的属性结构学派。至此,人们开始尝试从结构关系的视角探索知识的发展与传承。显然,在结构关系的阐释与揭示方面,网络思维具有得天独厚的优势。然而其后的研究中,尽管在这一领域也陆续有成果面世,但仍然主要局限于引文网络,虽然奠定了基础却并未成为真正的学科热点。

图书情报学界真正以网络思维审视和洞察知识发展相关问题的热潮涌起于 20 世纪末。其启动的导火索是美国圣母大学的 Barabási 等人[7] 发表在《科学》(*Science*)杂志上的 *Emergence of Scaling in Random Networks* 一文,该文以真实数据为基础,对 WWW 网络、引文网络、演员合作网络、电力网络等诸多现实网络自组织(Self-organize)状况下的节点度的幂律(Power-law)分布以及网络的无标度(Scale-free)状态进行阐述。同一时期,Watts[8] 和 Albert[9] 等学者先后在《自然》(*Nature*)杂志上发表关于小世界(Small World)网络的群体动力学以及 WWW 网络直径问题的科研文章。这些成果不但标志着国际学术界网络结构主义[10] 的复兴,更是将网络科学(Network Science)[11] 推至学术界的显学。至此,知识的网络结构及其自组织理论再次诠释了 Popper 提出的客观知识世界的观念,同时也从网络结构关系的层面令学术界不得不重新认识当初 Popper 提出的知识间逻辑关系的重要性。

其后的研究中,基于引文关系的知识网络依旧在知识发展脉络研究中占有重要地位。Johnson[12] 等人通过对引文网络的中心性研究,发现网络中的直接关联能够促进学术成果的广泛引用和发展传承,并且识别出引用数量与网络紧密中心度之间的正比关系。Polites[13] 及其合作者则在中心性基础上将凝聚子群、等价结构等更多的网络分析指标用于测度引文知识网络。他们的研究不但拓展了引文知识网络的分

析手段，还基于知识网络的视角为学术期刊关系及影响力分析探索出更客观多维的技术路径。近年来引文知识网络的研究中，更是呈现出中心性、层级性、群聚性等多个层面齐头并进的局面[14]，以引文知识网络揭示领域知识发展的相关研究越来越走向成熟。在经典引文知识网络研究发展的同时，基于其他类型知识网络的知识发展研究也呈现出欣欣向荣的景象。关键词知识网络[15]、图书知识网络[16]、专利知识网络[17]等纷纷被应用到领域知识结构关系及其发展的研究工作中。近年来，随着基于网民群体智慧的 Folksonomy 知识组织模式的兴起，从数字图书馆到各类资源分享网站，大量的网络知识资源采用 Folksonomy 知识组织模式进行架构。学术界认为，这种开放网络环境下由群体认知驱动的社群知识组织模式既能够反映 Popper 的“世界 3”的知识客观性，同时还彰显了“世界 2”（主观精神世界）对“世界 3”的能动作用[18]，由此也催生了基于标签同现关系的标签知识网络的研究。Chojnacki[19] 以“用户—资源—标签”三元组为基础，构建 Folksonomy 知识组织模式的 3—模知识网络。通过对网络的聚类系数的分析，发现此类知识网络的聚类系数与随机网络不同，具有显著的高位性。Weng 和 Menczer[20] 构建了标签共生网络，基于标签同现关系提取主题聚类集群，并对不同主题聚类中标签的熵进行测度。他们的研究表明，早期多样性的标签意味着未来受欢迎程度高，而低多样性则有助于个体积累社群中的社会影响力。

国内学术界基于复杂网络分析的视角对领域知识发展问题展开研究兴起于 21 世纪初，相关成果才刚刚积累。相关研究工作也基本遵循着以传统引文知识网络为起点，到知识网络类型逐渐丰富，再到开放网络环境下的标签知识网络日渐兴起的发展过程。马费成[21] 从经典的引文知识网络出发，秉承 Popper 的知识进化论视角，基于知识网络的发展演化过程建立了知识网络的时序演化模型。由于研究中使用原始的引文知识网络，因此该引文知识网络是规模单调递增的有向网络。该研究表明，单调增长的引文知识网络中，点度择优机制在网络全局上保障了经典科学知识的传承性，而时间优先机制则在局部范围内促进了最新成果的吸纳和延展。刘向等[22] 基于 ISI 中特定领域学术期刊论文引用关系建立了基于复杂网络的知识演化模型。该研究揭示了领域知识演化过程中知识引用行为的马太效应背后的时间抑制性，这种时间抑制性能在一定程度上抑制引文知识网络点度择优机制所导致的从众影响。在此基础上进一步对引文知识网络中的理论来源、领域内聚等问题进行分析[23]。邱均平等[24] 基于 Web of Science 的引文数据库对“Knowledge Network”领域知识分别建立地区知识网络、引文知识网络、关键词知识网络，使知识网络的类型得以扩展，并基于上述知识网络对领域知识的发展现状及研究热点进行分析。在 Folksonomy 知识组织模式下的社群知识研究方面，马费成等[25] 基于 CiteULike 标签同现关系构建标签知识网络，并基于标签知识网络对社群知识展开了一致性分析、中心性分析、核心—边缘分析。研究表明，标签知识网络同时具备“小世界”和“无标度”属性。此外，本课题组也在前期的研究工作中基于标签知识网络，采用层次派系分析方法对紧密型领域知识群落的时间序列演化[26] 进行了分析。

综上所述，关于领域知识生长进化的研究主题在图书情报学界历久弥新，以网络思维对知识的发展与传承演化进行研究已经成为学术界的共识。而网络思维关注的不是事物本身，而是事物之间的关系[27]292-293。尤其是针对知识网络结构关系的时间序列分析，更有助于揭示知识发展演化过程中交叉、衍生、融合等现象背后的规律。特别是在当今开放网络全民参与的背景下，网络思维无疑扩展了研究者的视野，同时也增加了研究工作的难度。正像 Newman[28]28-31 在《网络科学引论》中所指出的，时间确定性网络（随时间变化的网络）在社会中称为纵向（Longitudinal）研究，是一个在未来值得关

注的领域。因此,本研究在基于标签同现关系构建的原始知识网络的基础上,采用层次 k—丛和层次派系分析方法,沿时间轴对标签知识网络中具有显著意义的知识群落的生长过程进行跟踪与分析,以期从中识别和挖掘领域知识群落生长的模式与规律。

2 理论框架

知识群落(Knowledge Community)[29]不是某一单一孤立的知识点,而是基于知识之间关联关系形成的具有互促互扰关系、具有生命力的知识群簇。图书情报学界早期对知识群簇的界定主要遵循知识组织体系的树形结构[10],界定的依据主要来自于领域专家主观认知的他组织方式。因此,无论是《杜威十进制分类法》(*DDC*)还是《中国图书馆分类法》,都是以树形结构中的特定分支或更细小的枝杈为知识群簇。这种知识群簇划分方式的机械性和割裂性在今天逐渐显露,其中大量的知识关联被忽略或隐匿,其自组织的生长演化也无从谈起。随着人类科学技术的不断发展以及网络思维的复兴,知识之间的交叉、融合等现象日渐凸显,以网络节点和连线呈现知识及其关联的方式被学术界普遍接受。其中多数的研究成果主要以知识节点或者知识节点与知识关联作为知识群簇界定的重要依据,由于静态研究不涉及时间序列的生长性,因此相关术语也使用知识群簇而非具有生命意义的知识群落。同时,考虑到课题组的前期研究成果已经证实,在知识网络规模(节点数量)不变的情况下,知识关联(连线数量)的变化仍然对知识群落发展演化具有重大影响[26]。因此,本研究重点对关联驱动的知识群落生长展开研究。

知识及其关联关系在网络拓扑结构中表现为节点与节点间的连线。在 Barabási 等人[7]发现并验证了现实社会中的诸多大型复杂网络属于具有幂律(Power-law)分布特征的无标度(Scale-free)网络之后,以引文、关键词、标签等元素之间的关联关系构建领域知识网络已成为图书情报界的共识。本研究基于知识关联数量与关联频度识别知识网络中具有显著意义的关联关系,并据此提取层次网络。进而采用 Seidman 和 Foster[30]的 k—丛(k-plex)理论,并结合 Luce 和 Perry[31]的派系(Clique)理论,在层次网络的基础上对领域知识群落的生长模式进行分析。

k—丛和派系的概念都源自于复杂网络,主要用于识别和提取复杂网络中不同关联紧密程度的凝聚子群。k—丛是非完备的网络凝聚子群,子群中的每一个节点都与该子群中至少 n-k 个(n 为子群规模)节点关联。派系则是具有完备性的网络凝聚子群,子群中的每一个节点都与该子群中其他所有节点(n-1 个)关联。一般情况下,k—丛是比派系松散的网络凝聚子群(只有当 k—丛中的 k=1 时,此时的 1—丛等于派系)。以 n=4、k=2 为例,成员为 4 的情况下,二者的区别参见图 1。

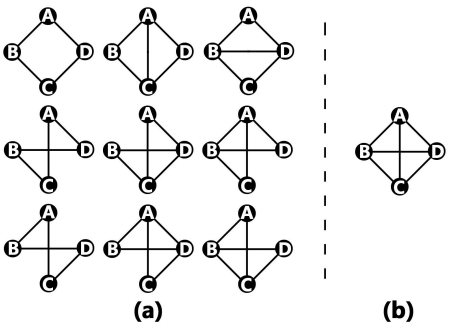

图 1 复杂网络中的 k—丛与派系

图 1 中,(a)为 n=4、k=2 条件下 4 成员 k—丛的所有形态,(b)为成员为 4 的派系。由图 1 可知,(a)中 9 种 k—丛形态中 A、B、C、D 四个节点之间的关联紧密程度显然都低于(b)中的派系。同时(a)中每一行的第一种状态在生长出一条新的关联关系后都可以演变为该行后两种状态之一,而每一行的后两种状态也可以在一条特定的关联关系退化后演变为该行的第一种状态。而且从拓扑结构上讲,每一种 k—丛状态

都是派系的一种非完备状态，即若要构成派系则首先要达成一种 k—丛。从这一点上来看，k—丛比派系具有更好的鲁棒性和宽松性，因而也更适合用于识别现实网络中的非完备子群[28]123-126。在课题组前期的研究中，已经证明层次派系能够识别和提取知识网络中的紧密型领域知识群落。由于 k—丛可以用于识别知识网络中更常见的较松散（次紧密）的领域知识群落，因此将派系和 k—丛结合使用，有助于基于群落中知识的关联程度揭示知识群落的生长模式。

本研究将 k—丛与派系和时间序列分析相结合，在各个时间窗口，重点对具有显著意义的知识网络中的 k—丛和派系分别识别和跟踪的同时，采用大数据交叉复现的思想对知识关联的变化与知识群落的生长进行交叉复现分析。数据不仅代表事实，还隐藏着发展规律[32]。因此研究中将完全采用数据说明，以多个维度的数据演变为基础，通过知识关联的频度与数量、松散型知识群落的数量与规模，以及紧密型知识群落的数量与规模等多个维度的交叉复现，对知识群落在时间轴上的生长模式进行识别与揭示。

3 研究方法

3.1 数据获取

本研究以卡塞尔大学（University of Kassel）知识与数据工程（KDE）团队、维尔茨堡大学（University of Würzburg）数据挖掘与信息检索（DMIR）团队和德国 L3S 研究中心共同管理的 Bibsonomy 文献出版共享系统为基础数据源。考虑到历史久远的知识领域数据资料不够完整，近期热点领域时效性过于突出且周期较短，而曾经的热点领域有些已降温并逐步沉寂，因此最具有代表性的知识领域是既不过热也不过冷且具有中长发展周期的知识领域。研究中选择“Database”（数据库）领域作为研究对象，以概念“Database”进行检索，采用自主研发的爬虫工具对检索结果进行抓取，能够获得文献题目、用户标签、标注时间等相关数据，检索时间为 2016 年 8 月 12 日。相关数据的统计结果如表 1 所示。

表 1 文献、标签及关联数量的时间序列分布

年份	文献数量（累计值）	标签数量（累计值）	关联关系（累计值）
2005	1	3	3
2006	88	38	155
2007	273	71	423
2008	539	85	608
2009	726	91	704
2010	942	101	791
2011	1 226	114	872
2012	1 622	156	1 908
2013	1 741	160	2 128
2014	1 891	165	2 359
2015	2 075	203	2 953

表 1 中，时间跨度为 2005—2015 年。以每一年为一个时间刻度，表中的“行”数据为该时间窗口下的文献数量、标签数量、关联关系数量，其中关联关系由标签在同一文献资源中的同现关系决定。上述三种数据（文献、标签、关联）的发生具有连续性，一经产生就不会消失，全部时间窗口累计共获得该领域相关文献2 075篇，涉及相关标签 203 个，关联关系（标签同现关系）2 953对。

鉴于知识节点及其关联关系的不可消失性，所获得的原始知识网络的演化过程必然是数量与规模不断增加和扩大的过程。这种演化过程虽然可以一定程度上识别知识的成长与繁荣状况，却难以捕捉知识生长过程中的迟滞与衰退现象。因此，研究中还将在原始知识网络的基础上进一步提取具有显著意义的层次知识网络。

3.2 研究流程与方法

3.2.1 层次知识网络提取

现实社会中的知识网络属于大规模无标度

复杂网络。传统复杂网络理论中对大规模复杂网络内具有显著意义的k—丛的提取往往采用Wasserman和Faust[33]提出的c层次k—丛(k-plex at level c)方法,而对具有显著意义的派系的提取则使用Doreian[34]的c层次派系(Clique at Level c)方法。本研究采用先提取层次网络,在层次网络的基础上再提取k—丛和派系的方法。以终态网络为基础,对网络节点的度分布和关联频度分布进行统计,双对数坐标系下的相关结果如图2所示。

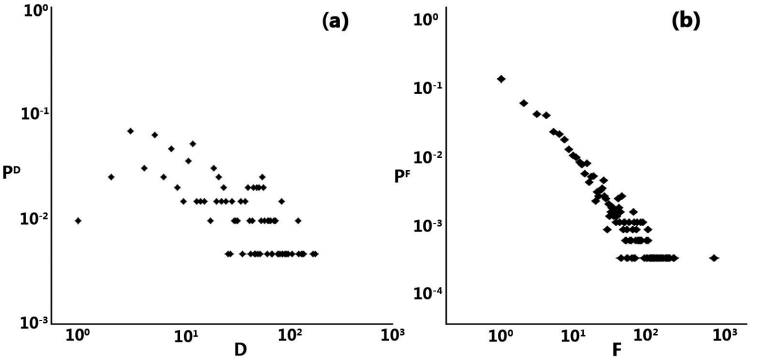

图2 知识节点度分布与知识关联频度分布

图2(a)为知识网络的节点度分布情况。可以发现,多数的知识节点(P^D值较高的节点)具有较小的度值(D值较低),而少量的知识节点(P^D值较低的节点)具有较高的度值(D值较高),即该知识网络中节点的度分布总体上趋近于幂律分布。这一结果与Barabási和Albert发表于《科学》(*Science*)杂志上的B—A模型[7]相吻合。

与Barabási和Albert的研究不同的是,本研究进一步分析了知识网络中的知识关联频度分布(参见图2(b))。从图2(b)中可以发现,知识网络的关联频度分布能够更好地符合幂律分布的特征。知识网络中大量的关联关系(P^F值较高的关联)具有很低的频次(F值很低),只有少量的关联关系(P^F值较低的关联)具有很高的频次(F值很高)。鉴于幂律分布的无标度性,正如波特兰州立大学的Mitchell[27]333-334教授所指出的,幂律分布就是分形(Fractal)。分形具有极大的自相似性(Self-similarity)[35],分形理论与幂律分布理论结合为基于关联频度提取层次网络奠定了基础。研究中设定关联关系平均频度为阈值,即将大于等于频度均值(算术平均数)的知识关联视为具有显著意义的知识关联提取层次网络。各个时间序列的原始网络平均关联频度,以及依据频度均值提取的层次网络的关联标签数和关联关系数如表2所示。

表2 层次网络提取的参数与结果

年份	频度均值	层次网络	
		关联标签数	关联关系数
2005	1	3	3
2006	1.832 3	21	50
2007	1.839 2	41	123
2008	2.498 4	41	131
2009	3.039 8	48	135
2010	3.464 0	54	177
2011	4.022 9	52	177
2012	6.354 8	72	405
2013	6.348 7	77	438
2014	7.358 2	88	509
2015	8.422 6	95	620

表中的频度均值保留了4位小数,如果均值不是整数(整数均值直接作为实际阈值),则实

际使用时基于“大于等于”的原则实为取整后加1，即实际阈值=Int(频度均值)+1。由此可知，在时间序列上2005—2006、2007—2008、2008—2009、2010—2011、2011—2012、2013—2014、2014—2015等窗口跃迁时阈值发生了跳跃。

3.2.2 知识群落提取

为了更好地跟踪分析知识网络中特定领域知识群落的生长发展状况，在具有显著性意义的层次网络(见表2)的基础上，依据Seidman和Foster[30]的k—丛理论提取松散型知识群落，同时采用Luce和Perry[31]的派系理论提取紧密型知识群落。此时获得的k—丛和派系在结果上与Wasserman[33]的c层次k—丛和Doreian[34]的c层次派系是一致的(c为阈值)，但是在计算的复杂度上却得到了极大的简化。设子群的最小成员数量为3，且k=2，分别基于已获得的层次网络计算提取11个时间窗口下两种不同类型的具有显著意义的领域知识群落，相关统计指标如表3所示。

表3 显著领域知识群落

年份	k—丛数量	派系数量
2005	1	1
2006	112	12
2007	556	31
2008	838	34
2009	784	30
2010	956	36
2011	971	34
2012	1 980	68
2013	2 098	75
2014	2 694	85
2015	2 981	93

此时，由于时间序列上的层次网络本身就是基于知识关联频度提取的具有显著意义的知识网络，因此能够很大程度上代表领域知识之间的关系结构。同时在此基础上得到的松散知识群落(k—丛)和紧密知识群落(派系)也同样满足显著性要求，具有很高的代表性。至此，可以结合各个时间窗口下两种不同类型的显著领域知识群落的具体特征，对基于标签同现关系构建的知识网络的领域知识群落演化进程展开细致的分析，揭示其演化生长的模式与规律。

4 研究结果

4.1 知识群落数量演化分析

从表2和表3中的数据可以发现，在同一时刻的时间截面上，知识间关联关系数量大于同时期的知识节点(标签)数量，k—丛数量更是远大于派系数量。由此得出的初步结论是，一个知识领域中，知识关联的数量比知识节点的数量更为丰富；松散知识群落的数量比紧密知识群落的数量更为丰富。为了更好地观察和跟踪显著性领域知识群落在时间轴上的变化趋势，将表2中的层次网络数据与表3中的数据进行归一化处理，以消除量纲的影响。由处理后的数据共同构造的层次网络及其满足显著性条件的知识群落演化趋势如图3所示。

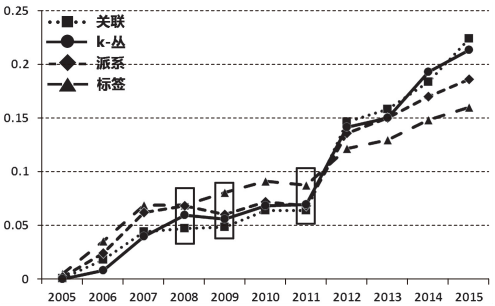

图3 层次网络及其显著知识群落演化趋势

从图3中可以发现，层次网络中的标签数量、关联关系数量、k—丛、派系四项指标从长期发展趋势来看都呈现出增长趋势。也就是说，随着时间的延展，知识网络中的知识节点与知识关联关系保持持续增长的趋势，而无论是相对松散的领域知识群落还是紧密型领域知识群

落的数量也随着知识节点和关联关系的增长在长期态势上处于增长趋势。

此处需要说明的是,在层次网络中,由于关联频度均值的变化,前一时间窗口中具有显著性的关联关系可能在下一时间窗口中失去显著地位,从而消失。如,若 t1 时间窗口中原始网络的关联频度均值为 2,如果有两对关联关系的频度 r1 = 2、r2 = 2,则这两对关联关系及其所对应的标签都被提取在 t1 时间窗口的层次网络中。如果随着标签同现频度的增加,t2 时间窗口中原始网络的关联频度均值跳跃为 3,若此时 r1 = 2、r2 = 3,则关联关系 r1 会在 t2 时刻的层次网络中消失,其所对应的标签如果没有与其他标签具备符合阈值条件的关联频度也会消失,由此导致 k—丛和派系也会受此影响。

结合表 2 和表 3 的数据对图 3 中曲线的细节观察可以发现,相对于整个时间区间内各项指标普遍增长的总体趋势而言,也有一些时间窗口中标签、关联、k—丛、派系的数量并非都保持了严格的单调性,因此在整个时间序列上,2008、2009、2011 这 3 个时间窗口尤其值得关注。

2008 时间窗口中,标签数量保持不变,关联关系数量、k—丛数量、派系数量都继续增加。从这一现象上看,对于知识群落而言,尽管知识节点和关联关系都是构成知识群落的关键因素,但后者比前者更重要[36]。关于这一点,课题组在前期专门针对紧密型领域知识群落的研究中也得到了同样的结论。

2009 时间窗口中,标签数量、关联关系数量都增加,而 k—丛数量、派系数量则在减少。按照 k—丛与派系的网络原理,这一现象似乎有些不可思议,毕竟网络凝聚子群的建立基础是网络节点与关联关系。对于这一现象的一种解释是,由于幂律分布条件下无标度网络增加的关联关系一定程度上被新增的节点稀释,另一方面群落数量的减少还可能伴随群落规模的扩大。关于后者将在下文交叉复现分析中做出更详细的阐述。

2011 时间窗口中,标签数量、派系数量减少,关联关系数量保持不变,k—丛数量增加。这一时间窗口表现出的现象与 2008、2009 时间窗口的情况又有不同。其中,标签数量增加与关联关系数量不变导致 k—丛数量增加还相对容易理解。但此时 k—丛数量增加的同时派系数量减少却似乎又有些令人费解,显然这一现象再次关系到群落规模等其他维度上的发展变化,需要深入的交叉复现分析。

至此,知识群落数量演化分析表明,无论是松散型的领域知识群落还是紧密型的领域知识群落,其数量在时间轴上的总体趋势是随着知识节点数量与关联关系数量的不断增长而增长的。这与人类科学技术发展过程中知识的不断增加、细化及交叉是相辅相成的。而且结合前文图 1 所表明的群落原理可知,在领域知识凝聚成群落的过程中,总是先借助一部分关联关系形成较松散的领域知识群落,再借助后续生长出的关联关系进一步构成紧密型的领域知识群落,这也是同一个时间窗口中 k—丛数量远大于派系数量(见表 3)的原因。当然,此过程中也存在个别时间窗口中四项指标的数量变化方向不协调的现象,对此本研究将进一步采用交叉复现的思想予以分析。

4.2 领域知识群落交叉复现分析

领域知识群落生长的总体模式相对鲜明,提炼与识别也相对简明清晰。知识群落的数量演化分析虽然在长期发展趋势上对群落生长的总体脉络做出了趋势研判,但是若要探究其中的规则细节则有必要进行多维度交叉复现分析。对 2008、2009、2011 这 3 个特殊时间窗口的多维度数据相对于上一个时间窗口的变化情况进行整合,原始知识网络与层次知识网络的标签数量、关联关系数量、k—丛数量、k—丛平均规模、派系数量、派系平均规模变化情况整合后的结果如表 4 所示。

表 4　知识群落变化的多维度整合

年份	原始知识网络						层次知识网络					
	标签数量	关联数量	k—丛数量	平均规模	派系数量	平均规模	标签数量	关联数量	k—丛数量	平均规模	派系数量	平均规模
2008	↗	↗	↗	↗	↗	↗	→	↗	↗	↘	↗	↘
2009	↗	↗	↗	↗	↗	↗	↗	↗	↘	↗	↘	↗
2011	↗	↗	↗	↘	↗	↗	↘	→	↗	↘	↘	↘

（→：不变，↗：增加，↘：减少）

表 4 中的知识群落变化情况集合了频度、数量、规模等多组维度的交叉复现，其增减变化情况都是针对前一时间窗口而言。从表 4 左右两部分知识网络的具体变化情况看，在具有显著性的层次知识网络的相关指标变化情况较为离散的同时，原始知识网络的相关指标变化情况总体上较为一致，基本上都是处于增长状态。原始知识网络中，标签数量增加、关联关系数量增加的同时，k—丛数量和派系数量也相应增加。结合表 1 的数据，说明人类知识发展过程中知识间关联的生长速度要比知识节点的生长速度更快。这种快速生长的关联关系使得知识节点之间建立连接，构成 k—丛，进而成为派系，从而完成由点/线到松散群落再到紧密群落的生长过程。结合表 4 中原始知识网络的“平均规模”指标看，在知识节点与知识关联不断构造出新的 k—丛和派系的同时，k—丛和派系的规模也在不断壮大。即在领域知识群落数量不断增加的同时群落的规模也在不断扩容。其中，2011 时间窗口 k—丛平均规模比上一时间窗口下降，表明即使在知识节点与知识关联只能生长无法消失（知识及其关联一经产生便不会消失，且原始知识网络没有经过筛选提取）的背景下，新产生的较小规模的 k—丛如果数量较多还可能会拉低平均规模这一指标。

然而表 4 右半部的层次知识网络是基于关联频度提取的具有显著性意义的知识网络，正像从一个生物种群中提取出那些在体力、智力等方面更优秀的个体一样，幂律分布与分形理论已经为这种提取工作提供了最好的理论支持。结合表 2 中的“频度阈值”指标可以发现，2008、2009、2011 三个时间窗口相比各自的上一时间窗口阈值都有增加，就像生物种群的整体体力、智力水平的提高促使选拔优秀者的标准也在提高一样。

2008 时间窗口中，由于层次知识网络的提取阈值从上一时间窗口的“2”提高到“3”（见表 2），因此其间必然发生知识节点与知识关联的新陈代谢。代谢的结果是标签数量不变关联关系数量增加，显然相对于上一时间窗口而言，知识网络规模不变的情况下（节点数量不变）连线数量增加了。由此导致的结果是 k—丛数量增加，但 k—丛的平均规模减小（见表 4）。即新生长出规模较小的 k—丛，并且拉低了平均规模。同理，派系数量增加且规模减小，同样说明新生长出规模较小的派系，并且拉低了派系的平均规模。这一现象反映出领域知识群落的生长过程中数量繁衍的基本规律。

2009 时间窗口中，层次知识网络的提取阈值再次由“3”跳跃到“4”（见表 2），标签数量和关联关系数量都有所增加。然而这一时间窗口层次知识网络中标签与关联关系的增长并没有得到与同一时刻原始知识网络一样的 k—丛与派系数量增长的结果，而且也不同于上一时间窗口层次知识网络的 k—丛与派系的增长。与规模维度的 k—丛平均规模和派系平均规模交叉对比可以发现，k—丛与派系数量减少的同时其平均规模却扩容了（见表 4）。显然，新生长的

标签和关联关系,将原本不同的 k—丛合并成规模更大的 k—丛,将不同的派系联合成体积更大的派系。由此,标签与关联关系数量的增长也能够帮助不同的 k—丛和派系发育成规模更大的群落,揭示了领域知识群落生长过程中规模扩张的基本规律。

2011 时间窗口中,层次知识网络的提取阈值又一次发生跳跃,由"4"到"5"(见表 2),只是这一次阈值跳跃的结果导致层次知识网络标签数量减少而关联关系数量保持不变。正如前文所述,频度均值的提高会造成部分知识节点由于关联频度不达标而从知识网络中陨落消失。这一时间窗口的 k—丛数量和平均规模变化情况与 2008 时间窗口一致,是松散型知识群落数量繁衍的表现,此处不再赘述。而派系数量及其平均规模同时减少的现象(见表 4)则与以往不同。松散知识群落数量的增加没有导致紧密知识群落数量的相应增加,说明在领域知识群落的生长过程中,松散知识群落是构成紧密知识群落的必要条件而不是充分条件。派系数量减少的同时其规模也在缩小,说明一些原本比较显著的知识节点和关联关系由于显著程度不够又淡出了紧密知识群落,反映出领域知识群落生长过程中核心知识群凝聚与迭代筛选的基本规律。

5 结论与讨论

本研究基于标签同现的关联关系构建领域知识网络,并依据幂律分布和分形理论在原始知识网络的基础上提取了具有显著意义的层次知识网络。采用 k—丛和派系的方法识别不同知识网络中的松散知识群落和紧密知识群落,沿着领域知识发展演化的时间序列,综合关联、数量、规模、显著性等多个维度,对领域知识群落的生长进行交叉复现分析。综合时序分析与交叉复现的结果,可以得出以下结论。

(1)领域知识群落的基本生长路径为"关联关系→松散群落→紧密群落"。传统知识管理理论认为知识点是知识体系中最基本的元素,同时也是以往人们最为关注的焦点。然而在网络思维的理念下,没有知识关联的知识节点只能是一个个孤立的节点,无法体现知识之间互促互扰的伴生关系。从分析中的数据可以发现,知识关联关系实现了知识节点之间的连接,并且随着领域知识的发展演化(时间序列)迅速增加(见表 1、表 2)。当知识关联关系(节点之间的连接)丰富到一定程度,就形成了 k—丛所识别的松散型领域知识群落;随着知识关联关系进一步丰富,在松散型知识群落的基础上进而涌现出派系识别的紧密型领域知识群落(见表 3)。这一点从 k—丛和派系的形成原理(见图 1)中也可以得到验证。从网络科学的视角理解,领域知识群落的生长路径遵循着由"点"到"线"再到"网"的发展过程,其中的"网"态又包含从松散到紧密的过程。

(2)知识间关联关系的增长是领域知识群落生长过程中数量繁衍和规模扩容的保障。研究中的数量分析部分已经表明,在知识发展的总体趋势上,领域中的知识节点、关联关系数量、松散知识群落、紧密知识群落在长期趋势上不断地积累与增长。而且,采用层次知识网络分析还可以发现,无论是数量分析还是交叉复现分析,都显示出即使在知识网络规模不变(节点数量不变)的情况下,只要关联关系数量继续增加,就依然会衍生出更多的松散型领域知识群落和紧密型领域知识群落(如表 4 中 2008 时间窗口)。同时,从知识群落的规模维度的交叉复现看,正是因为新生长出的领域知识群落体积较小,才从总体上拉低了知识群落(k—丛和派系)的平均规模。另一方面,知识间关联关系的增长也同样能够促进领域知识群落的规模扩容。表 4 中 2009 时间窗口的 k—丛和派系变化情况就很好地证明了这一点。新增加的关联关系有机会将原本不同的 k—丛或派系桥接成更大的 k—丛或派系,从而形成知识群落数量减少而规模扩容的生长现象。

(3)关联关系频度和数量的积累是领域知

识群落生长过程中核心凝聚的过滤器。以往领域知识核心的识别一般都是基于知识节点的度值(Degree),但在时间序列的动态演化中,度值只升不降。高度值知识节点的众多知识关联中可能仅有少量关联具有显著性(高关联频度),基于度值识别的领域知识核心会出现质量良莠不齐的现象。研究中发现,原始知识网络中度值不为0的条件下,其关联频度最大为688,最小为1,这种巨大的频度差距一定程度上制约了知识演化模式的识别与揭示。基于幂律分布和分形理论的层次知识网络,通过阈值过滤筛选出具有显著意义的关联关系。频度维度与数量维度的交叉,使得不具备显著性的关联关系淘汰出层次知识网络,进而过滤出具有显著意义的领域知识群落。在此基础上结合规模维度的相关变化情况,能够揭示出领域知识核心凝聚的基本模式。表4中2011时间窗中的群落变化情况就是一个典型的代表。领域知识发展演化过程中,网络总体关联频度的积累使得过滤阈值发生跳跃(见表2)。部分曾经显著的知识节点和关联关系淡出紧密型知识群落,群落数量和规模由此减小而得到凝聚。显然,在领域知识核心凝聚过程中不是简单的叠加,其中还包括知识的更新和迭代。

尽管研究中沿时间序列的线索,以关联、频度、数量、规模等多个维度交叉复现分析的方法,对知识关联驱动下的领域知识群落生长演化进程进行跟踪与分析,但是研究中也存在不完善之处。以年份为时间刻度虽然一定程度上能够反映出领域知识群落生长演化的模式与规律,但是对于其中特定现象细节的呈现与展示却受到一定制约。后续的研究将在丰富数据数量与维度的同时细化时间粒度,随着时间刻度的细化使领域知识的演化更加清晰,更深度地分析与揭示领域知识群落生长过程中的模式与规律。

参考文献

[1] Popper K. 客观的知识——一个进化论的研究[M]. 舒炜光,卓如飞,梁咏新,等,译. 杭州:中国美术学院出版社,2003.(Popper K. Objective knowledge:an evolutionary approach[M]. Shu Weiguang,Zhuo Rufei,Liang Yongxin,et al,trans. Hangzhou:The China Academy of Art Press,2003.)

[2] Surowiecki J.群体的智慧:如何做出最聪明的决策[M]. 王宝泉,译. 北京:中信出版社,2010:1-25.(Surowiecki J. The wisdom of crowds[M]. Wang Baoquan,trans. Beijing:China CITIC Press,2010:1-25.)

[3] Barabási A L,Frangos J. Linked:the new science of networks science of networks[M]. Cambridge:Perseus Publishing,2002:219-226.

[4] Garfield E. Citation indexes for science:a new dimension in documentation through association of ideas[J]. Science,1955,122(3159):108-111.

[5] Price D J de S. Networks of scientific papers[J]. Science,1965,149(3683):510-515.

[6] Belkin J N,Robertson E S. Information Science and the phenomenon of information[J]. Journal of the American Society for Information Science,1976,27(4):197-204.

[7] Barabási A L,Albert R. Emergence of scaling in random networks[J]. Science,1999,286(5439):509-512.

[8] Watts D J,Strogatz S H. Collective dynamics of 'small world' networks[J]. Nature,1998,393(6684):440-442.

[9] Albert R,Jeong H,Barabási A L. Internet:diameter of the world-wide web[J]. Nature,1999,401(6749):130-131.

[10] 滕广青,贺德方,彭洁,等. 结构与秩序:知识组织领域中结构主义思想的演进[J]. 情报理论与实践,2015,38(4):6-10.(Teng Guangqing,He Defang,Peng Jie,et al. Structure and order:the evolution of structuralism in knowledge organization [J]. Information Studies:Theory & Application,2015,38(4):6-10.)

[11] Lewis T G.网络科学:原理与应用[M]. 陈向阳,巨修练,等,译. 北京:机械工业出版社,2011,16-17.(Lewis T G. Network science:theory and applications[M]. Chen Xiangyang,Ju Xiulian,et al,trans. Beijing:China Machine Press,2011,16-17.)

[12] Johnson B,Oppenheim C. How socially connected are citers to those that they cite?[J]. Journal of Documentation,2007,63(5):609-637.

[13] Polites G L,Watson R T. Using social network analysis to analyze relationships among IS journals[J]. Journal of the Association for Information Systems,2009,10(8):595-636.

[14] Giannakis M. The intellectual structure of the supply chain management discipline:a citation and social network analysis[J]. Journal of Enterprise Information Management,2012,25(2):136-169.

[15] Seo J,Lee J Y,Ahn S,et al. Keyword hierarchy construction using co-word analysis[J]. International Information Institute(Tokyo). Information,2016,19(6B):2397-2402.

[16] Coscia M,Giannotti F,Pensa R. Social network analysis as knowledge discovery process:a case study on digital bibliography[C]//ASONAM '09 Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining,Washington DC:IEEE Computer Society,2009:279-283.

[17] Krafft J,Quatraro F,Saviotti P P. The knowledge base evolution in biotechnology:a social network analysis[J]. Economics of Innovation and New Technology,2011,20(5):445-475.

[18] Kilduff M,Tsai W. 社会网络与组织[M]. 王凤彬,朱超威,等,译. 北京:中国人民大学出版社,2007:128-147.(Kilduff M,Tsai W. Social network and organizations[M]. Wang Fengbin,Zhu Chaowei,et al,trans. Beijing:China Renmin University Press,2007:128-147.)

[19] Chojnacki S,Klopotek M. Random graph generative model for folksonomy network structure approximation[J]. Procedia Computer Science,2012,1(1):1683-1688.

[20] Weng L,Menczer F. Topicality and impact in social media:diverse messages,focused messengers[J]. PloS one,2015,10(2):e0118410.

[21] 马费成,刘向. 科学知识网络的演化模型[J]. 系统工程理论与实践,2013,33(2):437-443.(Ma Feicheng,Liu Xiang. Evolvement model for scientific knowledge networks[J]. Systems Engineering-Theory & Practice,2013,33(2):437-443.)

[22] 刘向,马费成. 科学知识网络的演化与动力——基于科学引证网络的分析[J]. 管理科学学报,2012,15(1):87-94.(Liu Xiang,Ma Feicheng. Evolution and dynamics of scientific knowledge network:based on the study of scientific citation network[J]. Journal of Management Sciences in China,2012,15(1):87-94.)

[23] 刘向,马费成,王晓光. 知识网络的结构及过程模型[J]. 系统工程理论与实践,2013,33(7):1836-1844.(Liu Xiang,Ma Feicheng,Wang Xiaoguang. Formation and process model of knowledge networks[J]. Systems En-

gineering-Theory & Practice,2013,33(7):1836-1844.)

[24] 邱均平,吕红. 基于知识图谱的知识网络研究可视化分析[J]. 情报科学,2013,31(12):3-8.(Qiu Junping, Lü Hong. Visualization analysis of research on knowledge network based on mapping knowledge domains[J]. Information Science,2013,31(12):3-8.)

[25] Ma F,Li Y. Utilising social network analysis to study the characteristics and functions of the co-occurrence network of online tags[J]. Online Information Review,2014,38(2):232-247.

[26] 滕广青. Folksonomy 模式中紧密型领域知识群落动态演化研究[J]. 中国图书馆学报,2016,42(4):51-63.(Teng Guangqing. Dynamic evolution of the close-knit domain knowledge communities in folksonomy[J]. Journal of Library Science in China,2016,42(4):51-63.)

[27] Mitchell M. 复杂[M]. 唐璐,译. 长沙:湖南科学技术出版社,2011.(Mitchell M. Complexity:a guided tour[M]. Tang Lu,trans. Changsha:Hunan Science & Technology Press,2011.)

[28] Newman M E J. 网络科学引论[M]. 郭世泽,陈哲,译. 北京:电子工业出版社,2014.(Newman M E J. Network:an introduction[M]. Guo Shize,Chen Zhe,trans. Beijing:Publishing House of Electronics Industry,2014.)

[29] 滕广青,杨明秋,田依林,等.Folksonomy 模式中的知识群落及其核心知识分析[J]. 图书情报工作,2015,59(22):124-129.(Teng Guangqing,Yang Mingqiu,Tian Yilin,et al. Analysis on knowledge communities and core knowledge in folksonomy[J]. Library and Information Service,2015,59(22):124-129.)

[30] Seidman S B,Foster B L. A graph-theoretic generalization of the clique concept[J]. Journal of Mathematical Sociology,1978,6(1):139-154.

[31] Luce R,Perry A. A method of matrix analysis of group structure[J]. Psychometrika,1949,14(2):95-116.

[32] 涂子沛. 数据之巅:大数据革命,历史、现实与未来[M]. 北京:中信出版社,2014:81-82.(Tu Zipei. Big data:history,reality and future[M]. Beijing:China CTTIC Press,2014:81-82.)

[33] Wasserman S,Faust K. Social network analysis:methods and applications[M]. Cambridge:Cambridge University Press,1994:277-282.

[34] Doreian P. A note on the detection of cliques in valued graphs[J]. Sociometry,1969,32(2):237-242.

[35] Lesmoir-Gordon N,Rood W,Edney R. 分形学[M]. 杨晓晨,译. 北京:当代中国出版社,2014:23-24.(Lesmoir-Gordon N,Rood W,Edney R. Introducing fractals:a graphic guide[M]. Yang Xiaochen,trans. Beijing:Contemporary China Publishing House,2014:23-24.)

[36] Gleick J. 信息简史[M]. 高博,译. 北京:人民邮电出版社,2013:409-421.(Gleick J. The information:a history,a theory,a flood[M]. Gao Bo,trans. Beijing:Posts & Telecom Press,2013:409-421.)

滕广青 东北师范大学计算机科学与信息技术学院教授。吉林 长春 130117。

(收稿日期:2016-12-26;修回日期:2017-02-16)