

Folksonomy 模式中紧密型领域知识群落动态演化研究^{*}

滕广青

摘 要 具有高度感知易用性的 Folksonomy 知识组织模式,在实践应用中却呈现离散纷杂的外在表象。对 Folksonomy 知识组织模式中领域知识群落及其结构关系的研究,对于洞悉领域知识发展过程中的衍生、交叉、融合等现象具有重要意义。本研究以复杂网络分析中的层次派系分析技术为主要研究方法,基于标签共同标注关系构建领域知识网络。从标签间关联关系出发,对特定领域知识网络中紧密型知识群落的发展过程进行时间序列的动态跟踪与分析。研究结果表明:标签数量在领域成熟期趋于饱和,但标签间的连线数量却持续增长,紧密型领域知识群落规模逐渐扩大;紧密型领域知识群落数量总体递增的同时也存在波动,这种波动与知识群落自身的扩张、衰减、派生、融合的演化过程有关;随着领域知识的发展,紧密型知识群落之间的交叠密度呈上升趋势,并基于交叠关系形成了一个更大的知识群落,成为领域知识网络的部分骨架结构,进而能够展示出多个主要发展方向。对领域知识群落演化规律的揭示,有助于把握领域知识演进的发展脉络,并揭示 Folksonomy 知识组织模式中领域知识的发展模式与规律。图 4。表 4。参考文献 32。

关键词 知识组织 Folksonomy 领域知识 知识群落 复杂网络

分类号 G254

Dynamic Evolution of the Close-knit Domain Knowledge Communities in Folksonomy

TENG Guangqing

ABSTRACT

For its high level of perceived ease of use, Folksonomy presents a discrete and perplexing appearance in its practical application. Studies about domain knowledge communities and their structural relationships within Folksonomy (knowledge organization mode) make up an important research topic in library and information academia and are crucial for a thorough understanding of numerous phenomena, such as deviation, crossover and fusion, in the development of domain knowledge.

^{*} 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 MOE (Ministry of Education in China) Project of Humanities and Social Sciences.)

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)

This study makes an extraction of 2 231 related articles in specific knowledge domains with 647 tags and the inter-tag-relationships number is 11 749 in total. The tagging-time span is from 2006 to 2015 and the timer shaft is divided into 10 units of time-window. The study adopts as the major research approach the clique analysis technique which is commonly used for complex network analysis, and domain knowledge networks are constructed via co-tagging relations of the tags. The cliques at level are constructed via dynamic thresholds based on the association relationships between tags and the close-knit domain knowledge communities are identified by the frequency and completeness of the association relationships. The dynamic tracking and analysis of the close-knit knowledge communities in specific domain knowledge are executed with the time series as clues. Finally, the skeleton structure of domain knowledge is extracted in perspective of overlapping cliques.

The present study concludes that, in the development of domain knowledge based on Folksonomy, tag population in domain mature stage tends to become saturated while the association relationships between the tags (lines) continue to yield a sustainable increase and the scale of close-knit domain knowledge communities presents a tendency of fluctuating expansion. Two factors account for this fluctuating expansion. Firstly, there are remarkable dynamic changes in the development of domain knowledge; some significant relationships between tags in the initial developmental stage become no longer significant in the follow-up stage due to the rapid development of other relationships, and some insignificant relationships in the initial stage become increasingly significant in the follow-up period due to the rapid development of their own. Secondly, expansion, attenuation, derivation and fusion of domain knowledge communities co-exist in the domain knowledge development process, and this multi-mode-coexisting evolutionary process also leads to fluctuations in the formation of close-knit domain knowledge communities. With the development of domain knowledge, overlap density among close-knit domain knowledge communities is on the rise and gradually forms a larger knowledge community based on overlapping relationships which become part of the skeleton structure of domain knowledge network and comes to display a number of major multi-dimensional development directions. The idea and approach of clique at level and overlapping cliques employed in this study expand the application of complex network analysis technique in the field of library and information science, promote the crossing of disciplinary knowledge, and provide a new perspective for the study of knowledge organization.

This study proposes close-knit domain knowledge communities (CDKC) and is helpful for the grasp of the core content and factors of domain knowledge from complicated tag sets. The tracking and analysis of close-knit domain knowledge communities in Folksonomy via time series are dynamic in nature and capable of further development, and the transformation of static analyses into dynamic ones helps to penetratingly reveal the mode and pattern of domain knowledge development. It is virtually impossible for the present study to exhaust the previous literature and studies in this area, but the revelation of the evolutionary patterns and rules concerning domain knowledge contribute to the grasp of the developing venation of domain knowledge evolution, and also facilitate the studies aiming to reveal the development patterns and rules of domain knowledge in Folksonomy. 4 figs. 4 tabs. 32 refs.

KEY WORDS

Knowledge organization. Folksonomy. Domain knowledge. Knowledge community. Complex network.

0 引言

图书情报领域的任何一种分类法,本质上都是对领域知识的组织和架构。无论是早期《法国大百科全书》所代表的“线性”结构,还是盛行全球的《杜威十进制分类法》(DDC)所遵从的“树形”结构,都蕴藏着对构建基于知识关联关系的知识组织结构的共识。Folksonomy 知识组织模式自 21 世纪初诞生以来,迅速风靡互联网。从知识的客观性来看,Folksonomy 无疑是迄今为止距离人类感知最近的知识组织模式。然而,也恰恰是这种源自于 Web2.0 环境的近距离感知,使得其外在形式显示出离散与无序的特征。人们在享受其高度的感知易用性所带来的便利的同时,也对其离散表象下知识的关联与结构展开了研究和探索,并取得了丰硕的成果。特别是复杂网络理论引入图书情报学领域之后,各类领域知识网络中的小世界、无标度等特征逐渐被揭示出来,基于网络结构属性探索 Folksonomy 散乱外表背后领域知识之间的关联、交叉、渗透、融聚等问题,逐渐显露出其得天独厚的优势。

然而,对表象背后真谛的揭示往往需要动态的分析视角,好在 20 世纪 70 年代 Popper^[1]就将知识的演进与发展同进化论结合起来,不仅奠定了知识组织研究领域动态分析的理论基础,更把领域知识的生长、衰老现象赋予了自组织、自适应的生命特征。最新的研究成果表明,复杂网络理论应用于知识组织相关问题研究,有助于识别、分析、研判领域知识生长模式的复杂性与多维性。

有鉴于此,本研究从复杂网络的视角出发,采用层次派系的分析方法,对 Folksonomy 知识组织模式下紧密型领域知识群落的发展过程进行基于时间序列的动态演化分析,尝试揭示其散乱表象背后领域知识群落的衍生、交叉、融聚等现象的潜在模式,以期对 Folksonomy 模式中领域知识群落及其发展演化规律做出有益的探索。

1 研究综述

Folksonomy 知识组织模式自诞生以来就以高度的感知易用性席卷了众多类型的网站,同时,其用户标签的离散性和平层性也成为学术界研究和突破的焦点。其中,离散表象下标签间潜在的关联关系无疑成为近年来 Folksonomy 知识组织模式相关研究的主要课题,并且已经积累了较为丰富的研究成果。

自从 2004 年 Vander Wal^[2]正式提出“Folksonomy”这一概念以来,以 Folksonomy 模式组织知识资源,给网络知识组织实践带来了巨大的便利,同时也引起了国际学术界的普遍关注。特别是在关于 Folksonomy 模式中知识关联关系的研究方面,取得了较为显著的成果。德国卡塞尔大学的 Schmitz^[3]等人的研究表明,标签不仅仅代表简单的知识点,而且其背后存在潜在的知识关联,这种关联关系将会影响 Folksonomy 模式下的知识组织。然而,这一时期关于 Folksonomy 模式中知识关联问题的研究主要集中在语义关联层面。Laniado^[4]及其合作者借助 NetWord 中概念的静态层级结构在 Folksonomy 中扩充标签间的语义关系,为基于 Folksonomy 的用户导航创造了更多的可能路径。Solskinnsbakk^[5]通过为标签创造语义表示,采用无监督的机器学习方法,将潜在的标记素材与关联规则挖掘结果进行整合后,构建 Folksonomy 的语义结构框架,使得标签的语义表示成为该结构不可分割的组成部分。Zhitomirsky - Geffet^[6]通过分析多用户在同一资源上的协作标注行为,提出了基于多用户不同视角的标签(概念)关联方法和总体框架,并借助标签间的关联关系构建了多维度的动态领域本体。

与此同时,国内学者在 Folksonomy 模式中语义层面的标签知识关联研究也取得了相应的进展。北京大学信息管理系的周鑫^[7]等通过对用户标签概念外延的界定,挖掘用户标签间的关联关系。邵杨芳^[8]等从解决标签的平面性引

起的标签间缺乏语义关联和无法揭示资源的深层语义及资源间语义关系的问题出发,提出基于受控词表的优化方案,以此提高 Folksonomy 模式下知识组织和检索的效率。此外,滕广青^[9]等基于形式概念分析的方法,通过标签聚类分析和关联规则挖掘,对标签间的语义关联进行揭示。陈丽霞^[10]等在对等网环境下分析用户标签的语义等级关系。这些研究成果对 Folksonomy 知识组织研究中语义层面的标签间知识关联关系的剖析起到了积极的作用。

随着复杂网络分析(Complex Network Analysis, CNA)^[11]理论引入知识组织研究领域,“网络结构决定网络功能”^[12]的思想为 Folksonomy 模式下的知识组织研究提供了新的思想与方法。一些研究者尝试采用复杂网络分析的方法,对 Folksonomy 知识组织模式中知识的关联与结构关系展开研究。特别是基于标签的领域知识网络的小世界和无标度属性被揭示出来^[13],直接证明了复杂网络理论在标签关联研究方面的有效性。波兰国家科学院的 Chojnacki^[14]认为,Folksonomy 模式中用户、资源和标签既然是相互连接的三元组,研究中就应该以更直观的 3—模网络来呈现。与普通随机网络不同的是,“用户—资源—标签”3—模网络中聚类系数随着时间的推移从高位缓慢下降,揭示了其中的系统动力学和聚类系数的高位性。Mas^[15]利用“社群驱动”展示了 Folksonomy 模式下结构网络中知识概念间的关联关系。日本学者柳(H. Yanagimoto)^[16]同样基于标签构建了领域知识网络,并采用贝叶斯定理对标签间关联关系强度进行测评,实验结果表明该方法能够更好地判断关联关系强度,并且可以删除其中的虚假关系。

国内学术界虽然在复杂网络应用于 Folksonomy 知识组织模式的研究方面起步较晚,但也取得了一定进展。通过对标签网络中语义信息的测度,已经证明标签之间的连接具有较强的语义相关性,并且随着标签间距离的增加相关性减弱^[17]。此外,李静^[18]基于网络分析验证

了标签网络具有小世界效应和无标度特性的同时,采用新的聚类算法挖掘标签网络中的社团结构,有助于挖掘标签网络中隐含的语义关联。吴江^[19]基于网络凝聚子群对用户标签的网状分类结构进行分析。滕广青^[20]通过对标签网络中心性的考察,对标签网络的语义紧密性进行分析。苏晓萍^[21]在标签网络中使用随机游走的方法实现了标签的个性化推荐。

综上所述,以标签作为知识的表征对 Folksonomy 知识组织模式展开研究已经成为学术界的共识。学术界关于标签间关联关系的研究成果在不断丰富的同时,研究视角也从朴素关联转向网状关联以及基于聚类的子群内部关系。从相关研究工作的发展趋势来看,网络思维的引入,有助于深层次识别和揭示领域知识聚类、交叉的模式与规律。而且,Uddin^[22]等人的研究结果已经表明,在建立标签之间的关联关系时,不仅要考虑标签、资源、用户三者之间的关系,更要考虑三者之间的动态关联。因此,本研究采用复杂网络理论中的层次派系分析方法,对 Folksonomy 知识组织模式中基于标签的领域知识网络进行时间序列上的动态跟踪与分析,以期从中发现和提炼紧密型领域知识群落的演进路径和演化规律。

2 理论框架

Folksonomy 知识组织模式诞生于 Web2.0 的时代背景下,与传统知识组织模式的本质区别在于其开放性的用户标注行为。这种开放性的标注行为体现了社群用户对领域知识的群体认知,从超越领域专家的层面对 Price^[23]所提出的“无形学院”(Invisible College)进行了诠释。本研究借鉴 Crane^[24]采用网络思维研究无形学院的思想,基于标签间的邻接关系构建领域知识网络。领域知识网络为无向网络,通过网络结构关系能够反映领域知识之间的多维关联。

特定的知识领域,无论其涵盖范围的大小,总会有一些知识点之间存在相互依赖、相互促

进的关联关系,并形成知识群落^[25]。如果知识群落中,一个知识点能以较高的显著性(高频性)对其他所有(完备性)知识点起到基础或支撑作用,并促进这些知识点发展,或者产生需求进而拉动其发展,则在领域知识发展层面表现出知识点之间的紧密性。这种基于紧密关联关系构成的领域知识群落称为紧密型领域知识群落,它往往对于领域乃至学科的知识发展起到至关重要的影响。在复杂网络中可以用卢斯—佩里(Luce-Perry)^[26]的派系(Clique)概念对紧密型领域知识群落加以揭示。所谓派系,简单说就是复杂网络中的最大完备子图。由于复杂网络分析的理论与技术在我国图书情报学界的应用正处于发展阶段,因此,目前Clique的汉译并不统一,比较典型的译法包括派系^[27]、集群^[28]、团伙^[29]等。本研究中采用派系分析方法识别紧密型领域知识群落需要满足两个条件:一是高频关联,偶尔少量的关联不能代表领域知识的紧密核心关系;二是完备关联,基于部分知识节点之间关联关系的知识群落也不足够紧密。非派系(非紧密型领域知识群落)与派系(紧密型领域知识群落)的区别参见图1。

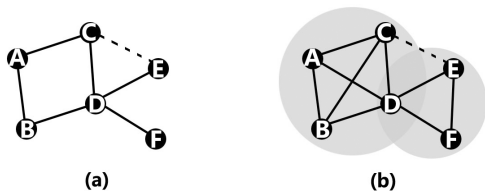


图1 复杂网络中的非派系与派系

图1中,6个知识节点都具有一定程度的关联关系,其中节点C与节点E之间的关联关系频次极低,远低于其他的关联关系,视为不具有显著性,研究中通过动态阈值筛选。即如果节点C与节点E之间仅有少量关联(虚线表示),而其他由实线表示的关联有百余次,则节点C与节点E之间的关联在领域中不具有高频显著性,被筛选。进一步讲,图1(a)中节点A与节点C、节点C与节点D、节点D与节点B、节点B与节点A之间都具有达到阈值标准的高频关联关

系,但节点A与节点D、节点C与节点B之间没有达到阈值标准的高频关联关系,则4个知识节点之间没有形成完全的相互支撑关系,其紧密程度不够,未形成派系,因此也未构成紧密型领域知识群落。反之,图1(b)则形成 $\{A, B, C, D\}$ 和 $\{D, E, F\}$ 两个派系,分别构成两个紧密型领域知识群落。

本研究采用时间序列分析的指导思想,对不同时间窗口下领域知识网络中的派系加以识别和跟踪。通过时间轴的延展呈现领域知识网络中紧密型知识群落的生长和衰老过程,并对此类知识群落的演化模式与规律进行揭示。与此同时,由于知识点之间关联频次的不同,知识群落内部的紧密程度也存在差异,越是领域中基于高频关联形成的知识群落,越具有典型性和代表性。因此,本研究在Doreian^[30]层次派系的基础上,采用动态阈值的方法构建不同时间窗口下的层次派系,并基于层次派系对紧密型领域知识群落的动态演化展开研究。

3 研究方法

3.1 研究数据

本研究以德国卡塞尔大学知识与数据工程小组架设的Bibsonomy网站资源为基础数据来源,该网站采用Folksonomy知识组织模式架构。以“network”为检索标签,共获得该领域相关文献2 231篇,这些文献被用户标注的各类标签总计647个,涵盖时间区间为2006—2015年。以年份为时间刻度,2006—2015年共10个时间窗口中的相关基础指标统计数据如表1所示。

3.2 研究流程与方法

3.2.1 矩阵转换与网络构建

根据2006—2015年不同的时间窗口,首先构建文献与标签的隶属关系矩阵,即将文献与其对应的标签建立隶属关系。在此基础上,基于不同标签相对于同一文献的共同标注关系,

表1 领域文献及标签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)
2006	134	139
2007	248	248
2008	735	443
2009	936	518
2010	1 296	576
2011	1 480	609
2012	1 739	642
2013	1 902	647
2014	2 166	647
2015	2 231	647

* 检索日期为 2015-09-18

将隶属关系矩阵转换为多值邻接矩阵,即当标签 A 与标签 B 共同标注同一文献时,则两个标签之间构成邻接关系,在邻接矩阵中计为 1(没有共同标注同一文献计为 0,表示不构成邻接关系)。进一步,如果标签 A 与标签 B 共同标注了两篇文献,则计为 2。以此类推,数值越大,表明两个标签之间的关系越密切。由此,获得 2006—2015 年时间序列下 10 个时间窗口中的标签多值邻接矩阵。

分别根据 10 个时间窗口中的多值邻接矩阵构建相对应的领域知识网络,即以标签为领域知识网络中的节点,如果两个标签之间具有邻接关系(标注同一篇文献),在网络中就在这两个标签之间建立一条连线。由于不同的标签对(一对基于共同标注关系的关联标签)标注同一篇文献的频次不同,因此领域知识网络中连线的频值也存在差异,一对标签共同标注同一文献的频次越高,则该对标签之间连线的频值越大。由此可以获得基于用户标签的 2006—2015 年各个时间窗口中的领域知识网络。

3.2.2 紧密型领域知识群落提取

本研究采用复杂网络理论中的卢斯—佩里

(Luce-Perry)^[26]派系对领域知识网络中的紧密型领域知识群落加以揭示。为了更好地跟踪分析特定领域知识群落中知识之间相互促进、相互依赖的伴生演化关系,采用 Doreian^[30]层次派系的研究视角,重点关注具有较高关系强度的紧密型领域知识群落。考虑到本研究中的时间序列因素,采用动态阈值的方法,选取各个时间窗口中领域知识网络各个节点所对应连线频值的均值为临界值,提取 C 层次派系,即设定只有超过连线频值均值的关联关系才视为有效高频关系,所提取的层次派系中任何一对节点之间的关系强度都不小于 C(均值),并且在层次派系外的任何一个节点到该派系中所有节点的关系强度都小于 C。

计算 2006—2015 年 10 个时间窗口中各个领域知识网络中节点(标签)间连线频值的均值,并以该值为临界值分别计算提取 10 个时间窗口中相应的紧密型领域知识群落,相关统计指标见表 2。

表2 时间序列紧密型领域知识群落

年份	临界值	群落数量
2006	1.9391	42
2007	1.9317	95
2008	2.0453	141
2009	1.9976	338
2010	2.2066	299
2011	2.3130	347
2012	2.5266	484
2013	2.6508	569
2014	2.9952	842
2015	3.0599	555

至此,可以结合各个时间窗口中紧密型领域知识群落的具体特征,对 Folksonomy 知识组织模式中紧密型领域知识群落的生长、衰老现象展开分析,揭示其发展演化的模式与规律。

4 研究结果

4.1 紧密型领域知识群落的总体趋势

10 个领域知识网络中的连线代表领域知识

(标签)之间的关联关系,连线数量的多少反映了不同时间窗口中领域知识之间关联关系的丰富程度。对上述 10 个领域知识网络中的连线数量及相关指标进行统计,所得结果见表 3。

表 3 时间序列领域知识网络中的连线数量

年份	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
连线数量	640	1 362	4 348	6 139	7 504	8 197	10 051	10 568	11 524	11 749

将表 3 中领域知识网络的连线数量与表 1 中文献数量、标签数量和表 2 中的知识群落数

量多个指标相结合,获得领域知识群落时间序列指标对比情况(见图 2)。

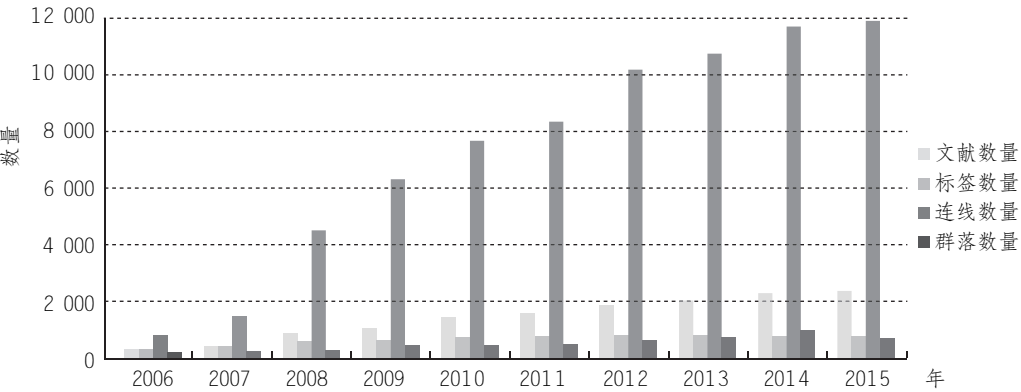


图 2 紧密型领域知识群落时间序列相关指标对比

从图 2 中可以发现,在 2006—2015 年时间区间内,文献数量逐年递增,但是后半段增速趋缓。相应的标签数量在前 1/3 段增势比较明显,中间 1/3 段增速趋缓,后 1/3 段则基本上保持平稳。值得注意的是,在整个时间周期内,领域知识网络的连线数量在保持递增的同时,还表现出与标签数量之间的巨大差异。从时间序列角度可以看出,当后 1/3 段领域知识点发展接近饱和状态时(2013、2014、2015 年标签数量均为 647,没有增加),相应年份中领域知识网络的连线数量却依然持续增加。这一现象非常符合《信息简史》的作者 Gleick^[31]的观点:知识的连通性比知识点本身更为重要。毕竟知识的发展与创新不但包括新知识点的产生,更包括新的知识关联关系的发现。这一现象同时也说明,

尽管这一时期领域内并没有新的知识点产生,但领域知识仍然在发生着演化与变迁。与此同时,在层次派系与动态阈值的筛选下,紧密型领域知识群落数量呈波动式递增的趋势,进一步反映出领域知识中知识群落生长与衰老并存的演化模式。关于紧密型领域知识群落在时间序列上的生长与衰老演化问题,将选择具体的子领域对紧密型领域知识群落的演化过程进行更深入的分析 and 说明。

4.2 紧密型领域知识群落的动态演化分析

以初始时刻即 2006 时间窗口中领域知识网络中的“mining”(挖掘)为起点知识点,并删减与领域知识无关的个别标签,在“network”(网络)限定的知识领域中围绕“mining”及其所对应的知识

群落展开紧密型领域知识群落的深入分析。在2006时间窗口中,采用C层次派系方法获得的包含“mining”的紧密型领域知识群落有且只有1个,命名为群落A,该知识群落同时包含“network”“mining”“pagerank”(网页级别)、“personalization”(个性化)4个知识点。4个知识点中的任意1个知识点与其他3个知识点之间的关系强度(连线频值)都大于均值,属于高频关联,是2006时间窗口中领域知识网络中具有重要地位的知识群落,其相互关系见图3。

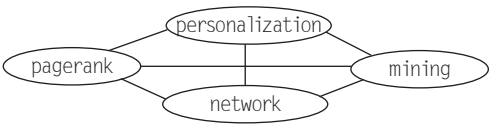


图3 初始时刻紧密型领域知识群落

随着时间的推移,“mining”所涉及的知识群落在领域知识发展演化的同时,其内部知识点之间的伴生关系也在不断发生变化。2006—2015年10个时间窗口中,“network”知识领域里与知识点“mining”相关的紧密型领域知识群落见表4。

表4 时间序列下的紧密型领域知识群落

年份	群落	知识点
2006	A	network ,mining ,pagerank ,personalization
2007	A	network ,mining ,pagerank ,personalization
	B	network ,mining ,social ,sna
2008	A	network ,mining ,pagerank ,personalization
	C	network ,mining ,social
2009	A	network ,mining ,pagerank ,personalization
	B	network ,mining ,social ,sna
	D	network ,mining ,social ,data
2010	A	network ,mining ,pagerank ,personalization
	D	network ,mining ,social ,data
2011	A	network ,mining ,pagerank ,personalization
	D	network ,mining ,social ,data
2012	A	network ,mining ,pagerank ,personalization
	E	network ,mining ,analysis ,pagerank
	F	network ,mining ,social ,sna ,analysis
	G	network ,mining ,social ,data ,analysis
2013	A	network ,mining ,pagerank ,personalization
	E	network ,mining ,analysis ,pagerank
	F	network ,mining ,social ,sna ,analysis
	G	network ,mining ,social ,data ,analysis
2014	A	network ,mining ,pagerank ,personalization
	E	network ,mining ,analysis ,pagerank
	H	network ,mining ,social ,sna ,data ,analysis ,prediction
	I	network ,mining ,social ,sna ,data ,analysis ,user
	J	network ,mining ,social ,data ,analysis ,interaction ,prediction
	K	network ,mining ,social ,data ,analysis ,interaction ,user ,web
2015	J	network ,mining ,social ,data ,analysis ,interaction ,prediction
	L	network ,mining ,social ,data ,analysis ,interaction ,web
	M	network ,mining ,social ,data ,analysis ,link ,prediction
	N	network ,mining ,social ,sna ,data ,analysis ,link ,prediction

从表 4 中可以看出,2006 时间窗口中知识群落 A 以网页级别及个性化为主要方向(鉴于知识网络本身与挖掘起点受所选领域和起点知识点限定,故对“网络”与“挖掘”方向不再赘述),该知识群落中“network”“mining”“pagerank”“personalization”4 个知识点相互支持,相互依赖。

2007 时间窗口,保留了知识群落 A 并且产生了 1 个新的领域知识群落,命名为群落 B。相对于 2006 时间窗口的知识群落而言,新生的知识群落 B 由“network”“mining”“social”(社会)、“sna”(社会网络分析)4 个知识点组成,该知识群落以社会科学领域中的社会网络分析为主要方向。此时,新老知识群落(群落 B 与群落 A)相互交叠,以“network”和“mining”为基础,派生出两个存在较大差异的研究方向,反映这一时期新的主要方向的派生。

接下来的 2008 时间窗口仍然为两个领域知识群落,其中,知识群落 A 仍保持不变,但是刚刚在前一时间窗口(2007)新生的知识群落 B 中的“sna”却淡出了群落 B,剩余知识点组成了新的知识群落,命名为群落 C。由知识群落 B 衰减为群落 C,这其中不排除动态阈值变化的影响因素,但可以发现,截至这一时刻,“sna”在知识群落 B 中并不稳定,即相对于成长中的整个领域知识网络而言,“sna”与知识群落 B 中其他 3 个知识点之间的关系强度尚不足够显著。

当时间轴推进到 2009 时间窗口时,“mining”所对应的知识群落进一步发展为 3 个。与前一时间窗口相比,知识群落 A 保持不变,知识群落 C 恢复到知识群落 B 状态(“sna”被重新吸纳),同时派生出由“network”“mining”“social”“data”(数据)构成的新知识群落,命名为群落 D。显然,知识群落 D 是以社会科学领域数据处理为主要方向的。

在 2010、2011 时间窗口,知识群落 A 继续保持,“sna”再次淡出,从而导致知识群落 B 解体(由于知识群落 D 的存在,剩余部分不再构成最大完备子图)。这两个时间窗口下都只保留了

知识群落 A 和知识群落 D,再一次反映出“sna”在知识群落中地位波动的同时,也开始展露出该知识领域对“data”的日渐关注。

2012、2013 时间窗口,领域知识群落发展相对稳定。这两个时间窗口中都包含 4 个领域知识群落,除了保留知识群落 A 外,还新生成了由“network”“mining”“analysis”(分析)、“pagerank”组成的知识群落,命名为群落 E;由“network”“mining”“social”“sna”“analysis”组成的知识群落,命名为群落 F;由“network”“mining”“social”“data”“analysis”组成的知识群落,命名为群落 G。从群落内的知识点构成不难发现,3 个新生的知识群落是在早期知识群落的基础上派生与扩张后形成的(如群落 F 由群落 B 扩张形成,群落 G 由群落 D 扩张形成),并且明显以数据分析为主要方向。

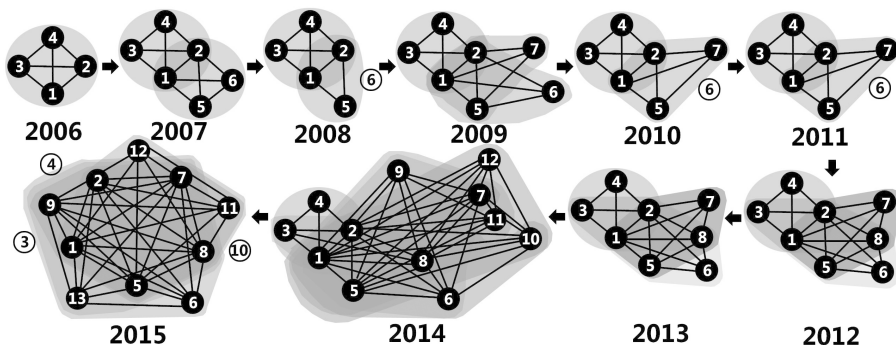
2014 时间窗口下,领域知识群落的发展更为繁荣。这一时期共形成了 6 个领域知识群落,除包含前一时间窗口下的群落 A、群落 E 外,还新生成了群落 H、群落 I、群落 J、群落 K。知识群落 H 是在群落 F 与群落 G 融合的基础上与“prediction”共同构成的;知识群落 I 是在群落 F 与群落 G 融合的基础上与“user”(用户)共同构成的;知识群落 J 是在群落 G 的基础上与“interaction”(互动)和“prediction”共同构成的;知识群落 K 是在群落 G 的基础上与“interaction”“user”“web”共同构成的。从这一时期的紧密型领域知识群落发展状态可以发现,紧密型领域知识群落开始由数据分析向预测研究转移,在领域知识群落的演化过程中,多个知识群落融合与单个知识群落扩展现象交织在一起。

2015 时间窗口下,初始时刻的知识群落 A 消失,“pagerank”和“personalization”淡出,说明网页级别与个性化已经不再与“network”和“mining”保持高度紧密关系,或者说原本紧密的关系随着领域知识的发展,在这一时间窗口中已经不再显著或突出。前一时间窗口(2014)中产生的知识群落 K 中原有的“user”淡出,缩减为新的知识群落,命名为群落 L。知识群落 J 继续

保留,另一方面知识群落 J 中原有的“interaction”淡出的同时,“link”(链接)被吸纳进来,形成新的知识群落,命名为群落 M。知识群落 H 也吸纳了“link”扩展生成新的知识群落,命名为群落 N。从此时紧密型领域知识群落的构成来看,这一时间窗口中衰减与扩张并存,群落 J、群落 M 与群落 N 继续引领着数据分析预

测的发展方向。

纵观整个时间序列上领域知识的演化过程,基于层次派系识别挖掘出的紧密型领域知识群落在不断生长、扩张、缩减、融合。2006—2015 年 10 个时间窗口中,“network”知识领域里知识点“mining”相关知识群落的演化序列如图 4 所示。



(注:1-network,2-mining,3-pagerank,4-personalization,5-social,6-sna,7-data,8-analysis,9-prediction,10-user,11-interaction,12-web,13-link)

图 4 时间序列下的紧密型领域知识群落演化

从图 4 中可以发现,随着时间的推移,紧密型领域知识群落的交叠现象越来越突出。如果按照 Alba^[32]关于交叠派系的观点,图 4 中知识群落交叠的密度比单个知识群落更有意义,交叠的结果最终可以被看作是一个规模更庞大的知识群落。由于在复杂网络中的派系被视为网络的骨架,交叠派系被视为网络的骨架结构^[28],因此,图 4 中时间序列后期呈现出的交叠的紧密型领域知识群落,自然构成了领域知识网络的部分骨架结构。针对“network”领域中“mining”所对应的知识群落而言,这一结果揭示出,从时间轴的初始时刻开始,领域知识群落的发展沿着“网页级别与个性化→社会网络分析→数据处理分析→数据分析与预测”的路径演进。

5 结论与讨论

本研究采用复杂网络理论中的派系思想,

基于用户标签的共同标注关系构建特定领域知识网络。以年份作为时间刻度,基于动态阈值,采用层次派系的方法对 2006—2015 年区间内 10 个时间窗口中的紧密型领域知识群落展开时间序列分析。通过对特定领域知识群落的生长、衰减、扩张、融合进行跟踪,对领域知识群落动态演化发展的模式与规律进行揭示与分析。综合上述对特定领域知识群落的演化发展过程进行的时间序列动态分析,初步可以得出如下结论。

(1) Folksonomy 知识组织模式下的领域知识发展过程中,在文献数量总体持续增加的同时,越接近领域成熟时期(时间序列的后期),标签(知识点)数量越趋近于饱和(见表 1),但标签间的连线数量(知识点之间的关联)却以远远大于标签数量的优势持续增长(见图 2),并且紧密型领域知识群落的规模也在总体上呈波动递增趋势(见表 4)。可见,知识之间的关联关系对于领域知识的发展起着至关重要的作用。由此

揭示出,科学研究与知识发展的过程中,知识之间逻辑关系的揭示与明晰无疑比知识点本身具有更重要的价值。这一结果验证了 Popper^[1]关于知识间逻辑关系重要性的观点。

(2)领域知识发展过程中,基于紧密关系的领域知识群落数量总体递增的同时也存在一定的波动。这种时间序列上的波动主要由两方面的因素造成:一方面是受到层次派系的动态阈值的影响,毕竟随着领域知识网络中标签(知识点)数量的增长与饱和,以及标签间连线(知识间关联关系)数量的持续增长,临界值在保持增长趋势的同时也会随之产生相应的起伏;另一方面是紧密型领域知识群落自身扩张、衰减、派生、融合的演化过程的表现。通过具体时间窗口(如 2015)衰减与扩张并存的现象可以发现,后者的作用更大,也恰恰是这种知识发展过程中的生长与衰老,彰显着领域乃至学科热门方向的转化与迁移。

(3)随着领域知识的发展,相关的紧密互惠型领域知识群落内部涵盖的知识点数量虽偶有起伏,但总体呈递增趋势,知识群落之间存在较

大程度的交叠现象,而且交叠密度呈上升趋势。因此,无论是按照 Alba^[32]的标准,还是 Scott^[27]的标准,Folksonomy 知识组织模式中相互交叠的派系在时间轴的后期,都会满足合并成一个规模更大圈子的标准。因此,尽管复杂网络理论中派系的概念较为严苛,但相互交叠的紧密型领域知识群落也会基于交叠关系构成一个更大的、次紧密的知识群落,并成为领域知识网络的部分骨架结构,从这一更大知识群落的视野出发能够展示出多个主要发展方向。

本研究从领域知识间关联关系的角度,采用复杂网络中层次派系的方法,将 Folksonomy 知识组织模式中领域知识群落的相关研究,由静态分析推进到更为复杂的动态演化研究。但是研究工作也还存在待完善之处,所选择的数据源毕竟难以穷尽学术界领域知识的所有文献和全部领域相关人员对该领域知识的认知。未来的研究将进一步采集更加全面的数据资料,综合地对领域知识群落发展演化过程中的模式与规律进行深入探索和揭示。

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,2003:258-283. (Popper K. Objective knowledge: an evolutionary approach[M]. Shu Weiguang, Zhuo Rufe, Liang Yongxin, et al, trans. Hangzhou: The China Academy of Art Press, 2003: 258-283.)
- [2] Vander Wal T. Folksonomy[EB/OL]. [2015-12-01]. <http://www.vanderwal.net/folksonomy.html>.
- [3] Schmitz C, Hotho A, Jäschke R, et al. Mining association rules in folksonomies[C]//Batagelj V, Bock H, Ferligoj A, et al. Data science and classification. Berlin: Springer-Verlag, 2006, VI: 261-270.
- [4] Laniado D, Eynard D, Colombetti M. Using wordnet to turn a folksonomy into a hierarchy of concepts[DB/OL]. [2015-12-27]. <http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-314/51.pdf>.
- [5] Solskinnsbakk G, Gulla J A. A hybrid approach to constructing tag hierarchies[C]//Meersman R, Dillon T, Herrero P. On the move to meaningful Internet systems, OTM 2010. Berlin: Springer, 2010: 975-982.
- [6] Zhitomirsky-Geffet M, Bar-Ilan J, Miller Y, et al. A generic framework for collaborative multi-perspective ontology acquisition[J]. Online Information Review, 2010, 34(1): 145-159.
- [7] 周鑫, 王军. 基于概念外延的 Folksonomy 语义关系挖掘方法[J]. 现代图书情报技术, 2008, 24(10): 6-10. (Zhou Xin, Wang Jun. Semantic relations mining in folksonomy based on extensions of concepts[J]. New Technology of Library and Information Service, 2008, 24(10): 6-10.)

- [8] 邵杨芳,贾君枝,贺培风. 基于受控词表的 Folksonomy 优化系统分析与设计[J]. 情报科学,2014,32(2): 112-117. (Tai Yangfang, Jia Junzhi, He Peifeng. Analysis and design of the controlled vocabulary based optimization system of Folksonomy [J]. Information Science, 2014, 32(2): 112-117.)
- [9] 滕广青,毕强,高娅. 基于概念格的 Folksonomy 知识组织研究:关联标签的结构特征分析[J]. 现代图书情报技术,2012(6): 22-28. (Teng Guangqing, Bi Qiang, Gao Ya. A study on knowledge organization of Folksonomy based on concept lattice: analysis on structural characteristics of related tags [J]. New Technology of Library and Information Service, 2012(6): 22-28.)
- [10] 陈丽霞,窦永香,秦春秀. 利用社会化标签实现 P2P 语义推荐[J]. 图书情报工作,2011,55(22): 110-113. (Chen Lixia, Dou Yongxiang, Qin Chunxiu. Semantic recommender based on social-tagging in p2p networks [J]. Library and Information Service, 2011, 55(22): 110-113.)
- [11] Lewis T G. 网络科学:原理与应用[M]. 陈向阳,巨修炼,等,译. 北京:机械工业出版社,2011: 87-140. (Lewis T G. Network science: theory and applications [M]. Chen Xiangyang, Ju Xiulian, et al, trans. Beijing: China Machine Press, 2011: 87-140.)
- [12] Freeman L C. Centrality in social networks conceptual clarification[J]. Social Networks, 1979, 1(3): 215-239.
- [13] Shen Kaikai, Wu Lide. Folksonomy as a complex network [DB/OL]. [2015-12-28]. <http://www.claymountain.com/svn/gradu/L%C3%A4hdeartikkelit/Folksonomy/ShWu05.pdf>.
- [14] Chojnacki S, Klopotek M. Random graph generative model for folksonomy network structure approximation[J]. Procedia Computer Science, 2012, 1(1): 1683-1688.
- [15] Mas M D. Intelligent interface architectures for folksonomy driven structure network [C] // Barolli L, Xhafa F, Vitable S, et al. 2012 Sixth International Conference on Complex, Intelligent, and Software Intensive Systems. Washington, DC: IEEE Computer Society CPS, 2012: 519-525.
- [16] Yanagimoto H, Yoshioka M. Relationship strength estimation for social media using folksonomy and network analysis [C] // 2012 IEEE International Conference on Fuzzy Systems. Piscataway: IEEE, 2012: 1-8.
- [17] Wu C, Zhou B. Tags are related: measurement of semantic relatedness based on folksonomy network [J]. Computing & Informatics, 2011, 30(1): 165-188.
- [18] 李静. Folksonomy 的网络性质研究与应用[J]. 情报科学, 2009, 27(10): 1483-1491. (Li Jing. Study and application of network properties of folksonomy [J]. Information Science, 2009, 27(10): 1483-1491.)
- [19] 吴江. 自由分类标签类聚成网状分类结构研究与实现[J]. 图书情报知识, 2011(1): 75-81. (Wu Jiang. Research and implementation of folksonomy tags clustering into network catalogue map [J]. Document, Information & Knowledge, 2011(1): 75-81.)
- [20] 滕广青,毕达天,任晶,等. Folksonomy 中用户标签的语义紧密性研究[J]. 现代图书情报技术, 2013(12): 48-54. (Teng Guangqing, Bi Datian, Ren Jing, et al. Study on semantic closeness of user tags in folksonomy [J]. New Technology of Library and Information Service, 2013(12): 48-54.)
- [21] 苏晓萍,楼俊钢. 结合超图投影和随机游走的个性化推荐方法[J]. 情报学报, 2012, 31(8): 830-836. (Su Xiaoping, Lou Jungang. An approach for personalized recommendation via hypergraph projection and random walk [J]. Journal of the China Society for Scientific and Technical Information, 2012, 31(8): 830-836.)
- [22] Uddin M N, Duong T H, Nguyen N T, et al. Semantic similarity measures for enhancing information revival in folksonomies [J]. Expert Systems with Applications, 2013, 40(5): 1645-1653.

- [23] Price D J S, Beaver D. Collaboration in an invisible college[J]. American Psychologist, 1966, 21(11): 1011-1018.
- [24] Crane D. Social structure in a group of scientists: a test of the “invisible college” hypothesis[J]. American Sociological Review, 1969, 34(3): 335-352.
- [25] 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.)
- [26] Luce R, Perry A. A method of matrix analysis of group structure[J]. Psychometrika, 1949, 14(2): 95-116.
- [27] Scott J. 社会网络分析方法[M]. 刘军, 译. 重庆: 重庆大学出版社, 2007: 95-100. (Scott J. Social network analysis: a handbook[M]. Liu Jun, trans. Chongqing: Chongqing University Press, 2007: 95-100.)
- [28] Nooy W D, Mrvar A, Batagelj V. 蜘蛛: 社会网络分析技术[M]. 林枫, 译. 北京: 世界图书出版公司北京公司, 2012: 98-104. (Nooy W D, Mrvar A, Batagelj V. Exploratory social network analysis with pajek[M]. Lin Feng, trans. Beijing: Beijing World Publishing Corporation, 2012: 98-104.)
- [29] 林聚任. 社会网络分析: 理论、方法与应用[M]. 北京: 北京师范大学出版社, 2010: 137-143. (Lin Juren. Social network analysis: theory, method and application[M]. Beijing: Beijing Normal University Press, 2010: 137-143.)
- [30] Doreian P. A note on the detection of cliques in valued graphs[J]. Sociometry, 1969, 32(2): 237-242.
- [31] 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.)
- [32] Alba R D. A graph-theoretic definition of a sociometric clique[J]. Journal of Mathematical Sociology, 1973, 3(1): 113-126.

滕广青 东北师范大学计算机科学与技术学院副教授。吉林 长春 130117。

(收稿日期: 2016-03-14; 修回日期: 2016-04-18)