

DOI:10.12154/j.qbzlgz.2019.04.003

领域知识群落的演变模式与知识传承*

安 宁 孙熊兰 滕广青 栾 宇 (东北师范大学信息科学与技术学院 长春 130117)

摘 要 文章采用社群发现算法对领域知识网络巨分支内部的群落结构进行识别与提取,通过知识群落组成率、输出率以及群落间传承强度的计算与分析,对知识群落的演变模式与知识传承特征进行时间序列的动态跟踪与分析。研究结果表明,领域知识网络在发展演变过程中更多地处于非连通状态;领域知识发展越成熟群落之间的知识交叉融合越频繁;交叉融合性的知识传承多发生在巨分支内部的知识群落之间。

关键词 知识网络 知识群落 群落演变 知识传承

The Evolution Model and Knowledge Heritage of Domain Knowledge Community

An Ning Sun Xionglan Teng Guangqing Luan Yu

(School of Information Science and Technology, Northeast Normal University, Changchun, 130117)

Abstract The present research uses community discovery algorithm to identify and extract the community structure inside the giant branches of domain knowledge networks. Through the calculation and analysis on inscape of knowledge communities, export of knowledge communities and lineage between communities, the time series dynamic tracking and analysis on evolution pattern and knowledge lineage characteristics of knowledge communities are carried out. The research results show that the domain knowledge networks are more non-connected in the process of evolution; the more mature the domain knowledge develops, the more frequently the knowledge cross-fusion between communities; the cross-fusion knowledge lineage mostly occurs between the knowledge communities within the giant branch.

Keywords knowledge network, knowledge community, community evolution, knowledge lineage

1 引言

自从Popper^[1]将人类知识的发展同进化论结合起来,知识的演化就成为图书情报学界长期研究的重点。用非形式化的语言来说,演化是指研究对象在时间轴上所展现的变化。对于领域知识来说,其内部的知识单元以及知识关联关系在时间轴上并不是一成不变的,而是具有一定的动态生长性,会随着科学研究发现以及人类认知的发展变化而变化。因此,识别领域知识的演化模式对于把握该领域知识发展脉络与趋势具有重要的意义。随着网络科学(Network Science)^[2]的复兴以及网络

分析方法的创新,基于复杂网络分析对领域知识及其内在关联关系进行研究的科学范式得到了学术界的广泛认可。许多真实网络并不是完全随机或者完全规则的,这些网络往往都具有很强的模块性或社群结构,在网络拓扑结构上表现出聚集现象。而在任何学科领域内部,知识单元之间同样也并不是完全孤立存在的个体,而是基于潜在的关联关系呈现出一定的群聚性特征。因此,对知识群落的研究有助于更好地理解知识的发展与演化,从中观层面洞察整个领域的知识结构与功能,掌握知识的衍生、交叉以及湮灭等发展过程。

有鉴于此,本文采用复杂网络分析中的社群发现

*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)的研究成果之一。

算法来识别领域知识内的知识群落,以多种群落演变指标沿时间序列对其进行动态跟踪与分析,以期对领域知识群落的演化模式与知识传承做出有益的探索。

2 相关研究综述

对网络中社群结构的研究已经有较长的历史了,网络中的社群划分与数学中的图论、计算机科学中的图分割思想以及社会学中的层次聚类密切相关。随着对网络性质的深入研究,逐渐将复杂网络中的社群发现推广应用到了生物^[3]、经济^[4]、社交^[5]等多个领域。Newman和Girvan^[6]通过对实验网络和真实网络(包括引文网络、合著网络等)进行研究发现,大多数复杂网络中都具有较为明显的社群结构。加之人们逐渐观察到基于知识关联关系而呈现出的知识群聚特性,从网络群落结构的视角研究知识的群聚性也逐渐得到了图书情报学术界的普遍认可。

在图书情报学界针对知识群落的相关研究中,Fudholi等^[7]研究发现,知识群落中有价值的知识经常会发生变化,并且与所在知识群落内部的知识之间密切相关,识别这些知识有助于更好地理解知识群落的主题发展。Wallace等^[8]利用复杂网络层次社群快速展开方法对共被引网络内的群聚性问题进行研究,结果表明社群发现方法与传统的共被引分析相比揭示了更多的知识领域的结构细节,在科研主题识别上具有天然的优势。Yi和Choi^[9]利用派系过滤算法识别关键词网络中的知识群落,分析网络结构和群落结构之间的相似性,研究表明关键词网络中的群落结构可以作为更高层次科学知识的替代元素。Dong等^[10]在作者合著关系网络中,利用快速贪婪算法识别合著网络中的社群结构,结合渗透理论探究具有社群结构的网络的稳定性。McCain^[11]利用社群发现算法对引文网络中的科研主题进行识别,研究表明在引文网络中存在明显的社群结构,不同的社群可以揭示领域研究的发展演变过程。随着对知识发展演进过程的更多关注,学术界将更多的注意力转向了基于群落结构的领域知识动态演化研究。Lancichinetti和Fortunato^[12]基于社群结构挖掘其内部潜在的知识关联关系,并且利用一致性聚类算法划分出引文网络中不同知识群落组织,识别科学话题的产生、生长、分裂、衰退以及消亡的演变过程。Liu等^[13]基于链路演化的方法识别作者合著网络的社群结构,探索时间序列合著网络中社群结构的动态演变。

Wang等^[14]使用社区发现算法对时间序列共词网络进行社群划分,研究表明共词网络社群的演化可以表现为产生、消亡、分裂、合并、扩张与收缩六种演化状态。

与此同时,国内很多学者对知识的群聚性也进行了探索,并取得了丰富的成果。刘自强等^[15]利用Fast Unfolding社群发现算法处理具有语义特征的关键词网络,对语义社群进行语义主题识别。朱梦娴等^[16]采用社群发现算法对关键词网络进行关键词聚类分析,同时利用Z-value对知识群落进行核心圈分析。李纲等^[17]利用Jaccard系数与余弦相似性系数对合著网络的分析表明,全网络范围的作者研究兴趣相似程度存在一定比例的差异性(互补性),社群内部作者平均研究兴趣相似性及互补性均高于网络整体水平。另一方面,国内学术界也逐渐开始关注知识群落的动态演化研究。白如江等^[18]根据词与词之间的关系以及群落重叠度进行知识主题分析,通过知识群落演化揭示不同知识领域的互相影响情况。程齐凯等^[19]的研究发现,不同的时间段内知识群落的数量、大小、密度、结构等属性变化,既包括社群自身内部节点、关系和结构的变化,也包括社区间关系和位置的变化。滕广青^[20]进一步对紧密型知识群落分析发现,紧密型知识群落的波动与群落自身的扩张、衰减、派生、融合的演化过程有关,群落交叠形成的骨架结构能够展示出多个发展方向。

综上所述,基于知识网络的社群结构来探索知识的生长演化特征已经得到学术界的普遍认可并广泛应用。然而,针对知识群落之间衍生、交叉、融合现象背后的知识传承问题却鲜有问津。有鉴于此,本研究采用社群发现算法对知识群落进行识别,通过知识群落的组成率、输出率和传承强度分析,尝试揭示领域知识群落的演变模式与知识传承特征。

3 理论框架

网络科学领域的著名学者Newman^[21]在对静态真实网络的结构进行研究时发现,网络中给定的两个节点之间可能不存在路径,即是一种非连通的网络。非连通网络中的节点被划分成不同的子群分支,网络中任意两点之间可能不存在任何路径。Newman又进一步指出,在许多真实网络中通常都会有一个最大分支(又称巨分支),该分支占据了网络中的绝大部分(一般都会大于网络的一半),而剩余部分则被划分为大量的小分支,这些小分支之间彼此也并不相连,如图1所示。

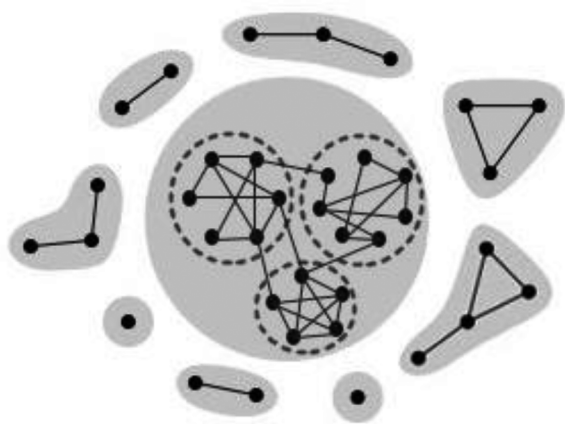


图1 具有分支和社群结构的网络

图1中阴影部分代表网络中不同的分支,图的中心位置为网络的巨分支,周边为小分支。对于巨分支来说,虽然其内部各节点彼此之间相互连通,但是在巨分支内部仍存在较为明显的社群结构,即每个社群内部的节点之间的连接相对紧密,各社群之间的连接相对来说比较稀疏。领域知识发展演进过程中,一方面受到外部人类知识发展变化的影响,另一方面则基于领域知识内部知识单元之间关联关系的兴衰更迭而变化。这些变化促使各知识节点聚集成不同的分支结构,因此可以说领域知识网络通常是由不同的分支群落所构成的。一般来说,领域知识网络中非连通的小分支群落的规模往往较小,通常代表了领域知识内部的非主流研究方向。而本研究的目的在于探索领域知识网络中知识群落的演变模式与知识传承特征,因此研究中将这非连通的、孤立的小分支群落统一视为一个知识群落,称之为“小分支群落”;而对于领域知识网络中的巨分支采用社群发现算法对其进行知识群落的划分。

目前识别复杂网络中的社群结构大致分为图划分和社群发现两类算法,其主要区别在于是否明确网络中社群的规模和数量。本研究的背景是一个动态演化的环境,事先并不确定各知识群落的具体情况,因此在研究中采用社群发现中的Louvain算法^[23]来划分与识别领域知识网络中巨分支内部的知识群落。为了判别知识群落被划分的是否合适,研究中采用Newman和Girvan^[6]提出的模块度指标来对群落划分质量进行衡量,用 Q 表示。模块度的基本思想就是把划分出群落后的网络与相应的零模型进行比较,以度量群落划分的质量。而所谓的零模型就是指与所研究的网络具有某些相同的性质,而在其他方面完全随机的随机网

络模型。目前在分析网络群落结构时,通常把与待研究的网络具有相同度序列的随机网络(一阶零模型)做比较对象^[23]。因此,常用的模块度 Q 定义为:

$$Q = \frac{1}{2M} \sum_{ij} (a_{ij} - \frac{k_i k_j}{2M}) \delta(C_i, C_j) \quad \text{公式(1)}$$

公式(1)中 M 是网络中的边数, $A=(a_{ij})$ 是网络的邻接矩阵, k_i 和 k_j 分别是网络中节点 i 和节点 j 的度值, C_i 和 C_j 分别表示节点 i 与节点 j 在网络中所属的群落;如果两个节点属于同一个群落, δ 取值为1,否则为0。事实上,一个网络的模块度就是该网络的群落内部边数与相应的零模型的群落内部边数之差占整个网络边数 M 的比例。有研究发现,当 Q 值在0.3~0.7时表明网络中出现了较强的群落结构^[24]。

研究中采用的Louvain算法就是基于模块度的社群发现算法,该算法不需事先确定群落信息,并且在研究大型网络上具有较好的效率和效果表现。该算法通过不断将节点进行凝聚,实现将整个网络分解为多个群落。Louvain算法通过不断聚集各节点来实现模块度值最优的结果,从而确定网络中群落的划分。研究中,基于上述的理论和思想,沿时间序列识别领域知识网络中巨分支内的知识群落,并分别测算知识群落演变过程中的组成率和输出率,进而计算知识群落的传承强度。根据计算结果,动态地分析与挖掘知识群落的衍生、交叉与湮灭等演变模式与规律,揭示知识传承的特征。

4 研究方法

4.1 研究数据

本研究采用以Folksonomy知识组织模式架构的Bibsonomy.org网站资源为基础数据源,以“information”为检索主题,共获得16 760篇文献,标签9 255个,时间跨度为1995~2017年。为了去除噪音数据和瞬变跳跃,更好地跟踪知识单元的生长与湮灭现象,在数据处理时以3年为时间刻度,1年为步长,平滑移动得到 t_1-t_2 共计21个时间窗口,即时间窗口 t_1 中包含了1995~1997年内的相关数据,时间窗口 t_2 中涵盖了1996~1998年内的相关数据,以此类推,得到21个时间窗口下的相关基础统计数据,如下页表1所示。

由表1中的标签以及关联数量的变化可知,在时间窗口下领域知识网络中标签(知识单元)及其共同标注关系(关联关系)存在着生长与湮灭现象。研究工作将表1中的标签作为网络的节点,标签之间的同标注关系为网

表1 时间序列知识网络节点与关联数量

| 时间窗口 | 标签数量 | 关联数量 | 时间窗口 | 标签数量 | 关联数量 |
|------|-------|--------|------|-------|--------|
| t1 | 70 | 283 | t12 | 4 659 | 60 092 |
| t2 | 87 | 346 | t13 | 4 911 | 51 646 |
| t3 | 126 | 543 | t14 | 4 362 | 34 791 |
| t4 | 126 | 526 | t15 | 4 031 | 33 494 |
| t5 | 133 | 565 | t16 | 3 462 | 29 797 |
| t6 | 93 | 348 | t17 | 2 829 | 21 759 |
| t7 | 92 | 351 | t18 | 2 202 | 15 178 |
| t8 | 166 | 510 | t19 | 2 023 | 13 027 |
| t9 | 546 | 2 615 | t20 | 2 144 | 16 575 |
| t10 | 1 848 | 21 679 | t21 | 1 912 | 11 750 |
| t11 | 3 526 | 48 022 | | | |

络连线,分别构建 t_1-t_{21} 时间窗口的领域知识网络。由于本研究的目的在于探索领域知识网络中知识群落的演变过程,因此所构建的领域知识网络为二值无向网络。

4.2 知识群落的识别

时间序列下的领域知识网络并非一定或一直处于

连通的状态,可能存在着孤立的小分支。在领域知识网络中,虽然孤立的小分支自身就代表着一定的知识群落,但是却不能代表该领域知识的主流方向。考虑到研究工作主要考察领域知识中知识群落的衍生、交叉、湮灭以及各群落间知识的传承情况,因此在研究中将这些小分支统一视为一个小分支群落(记作 C_x)来进行考察。其次针对领域知识网络中最大规模的分支(巨分支)采用 Louvain 算法^[22],对知识网络的巨分支结构进行群落划分与识别。最后基于时间序列动态探测领域知识群落的演变模式与知识传承特征。由于各时间窗口中知识网络拓扑结构的差异,导致不同时间窗口中所划分出的群落数目与规模也有所不同。对 t_1-t_{21} 时间窗口中各领域知识网络中知识群落的数量以及规模进行计算,结果如表2所示。其中第一行代表各时间窗口,第一列代表了群落编号。为了清晰地展示各时

表2 时间序列知识群落数量及规模

| | t ₁ | t ₂ | t ₃ | t ₄ | t ₅ | t ₆ | t ₇ | t ₈ | t ₉ | t ₁₀ | t ₁₁ | t ₁₂ | t ₁₃ | t ₁₄ | t ₁₅ | t ₁₆ | t ₁₇ | t ₁₈ | t ₁₉ | t ₂₀ | t ₂₁ |
|-----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| C _x | 4 | 4 | 0 | 0 | 0 | 11 | 7 | 27 | 56 | 115 | 169 | 177 | 216 | 183 | 216 | 196 | 197 | 242 | 233 | 199 | 201 |
| C ₀₁ | 16 | 19 | 41 | 32 | 40 | 21 | 20 | 37 | 83 | 502 | 795 | 1 173 | 801 | 910 | 627 | 746 | 455 | 585 | 346 | 549 | 303 |
| C ₀₂ | 14 | 18 | 25 | 25 | 32 | 18 | 18 | 24 | 65 | 400 | 663 | 841 | 691 | 570 | 565 | 564 | 454 | 283 | 231 | 208 | 270 |
| C ₀₃ | 11 | 12 | 15 | 15 | 17 | 12 | 15 | 23 | 57 | 194 | 400 | 826 | 637 | 554 | 560 | 546 | 417 | 166 | 218 | 202 | 185 |
| C ₀₄ | 9 | 9 | 13 | 14 | 15 | 12 | 10 | 18 | 48 | 114 | 390 | 501 | 621 | 511 | 423 | 313 | 336 | 114 | 174 | 146 | 168 |
| C ₀₅ | 7 | 7 | 12 | 13 | 15 | 7 | 8 | 17 | 48 | 99 | 334 | 238 | 558 | 425 | 321 | 214 | 120 | 104 | 136 | 143 | 158 |
| C ₀₆ | 5 | 7 | 8 | 11 | 8 | 6 | 8 | 15 | 43 | 92 | 301 | 215 | 332 | 334 | 220 | 191 | 118 | 91 | 108 | 122 | 97 |
| C ₀₇ | 4 | 6 | 7 | 10 | 6 | 6 | 6 | 5 | 36 | 66 | 220 | 134 | 302 | 206 | 195 | 169 | 110 | 79 | 87 | 110 | 85 |
| C ₀₈ | | 5 | 5 | 6 | | | | | 36 | 65 | 103 | 132 | 265 | 144 | 192 | 116 | 88 | 57 | 83 | 83 | 69 |
| C ₀₉ | | | | | | | | | 35 | 47 | 44 | 127 | 105 | 113 | 180 | 84 | 68 | 49 | 83 | 77 | 68 |
| C ₁₀ | | | | | | | | | 25 | 38 | 36 | 103 | 92 | 95 | 120 | 48 | 64 | 49 | 81 | 73 | 47 |
| C ₁₁ | | | | | | | | | 8 | 24 | 18 | 69 | 75 | 75 | 93 | 46 | 56 | 49 | 62 | 64 | 46 |
| C ₁₂ | | | | | | | | | 6 | 18 | 16 | 43 | 67 | 56 | 85 | 38 | 53 | 40 | 60 | 52 | 35 |
| C ₁₃ | | | | | | | | | | 12 | 11 | 17 | 60 | 33 | 61 | 34 | 52 | 39 | 31 | 34 | 32 |
| C ₁₄ | | | | | | | | | | 12 | 8 | 16 | 50 | 29 | 61 | 19 | 48 | 32 | 16 | 33 | 32 |
| C ₁₅ | | | | | | | | | | 12 | 8 | 11 | 16 | 25 | 27 | 16 | 29 | 31 | 14 | 22 | 31 |
| C ₁₆ | | | | | | | | | | 11 | 6 | 11 | 13 | 16 | 13 | 14 | 24 | 30 | 12 | 12 | 23 |
| C ₁₇ | | | | | | | | | | 7 | 4 | 8 | 10 | 15 | 13 | 14 | 21 | 28 | 11 | 6 | 22 |
| C ₁₈ | | | | | | | | | | 6 | | 5 | | 14 | 9 | 11 | 20 | 21 | 10 | 5 | 12 |
| C ₁₉ | | | | | | | | | | 6 | | 4 | | 12 | 7 | 11 | 11 | 13 | 9 | 4 | 11 |
| C ₂₀ | | | | | | | | | | 4 | | 4 | | 9 | 7 | 10 | 11 | 13 | 7 | | 10 |
| C ₂₁ | | | | | | | | | | 4 | | 4 | | 8 | 7 | 10 | 11 | 12 | 7 | | 7 |
| C ₂₂ | | | | | | | | | | | | | | 7 | 6 | 10 | 8 | 11 | 4 | | |
| C ₂₃ | | | | | | | | | | | | | | 5 | 5 | 8 | 7 | 10 | | | |
| C ₂₄ | | | | | | | | | | | | | | 5 | 5 | 7 | 7 | 7 | | | |
| C ₂₅ | | | | | | | | | | | | | | 4 | 5 | 6 | 7 | 7 | | | |
| C ₂₆ | | | | | | | | | | | | | | 4 | 4 | 5 | 6 | 6 | | | |
| C ₂₇ | | | | | | | | | | | | | | | 4 | 4 | 5 | 6 | | | |
| C ₂₈ | | | | | | | | | | | | | | | | 4 | 5 | 6 | | | |
| C ₂₉ | | | | | | | | | | | | | | | | 4 | 5 | 5 | | | |
| C ₃₀ | | | | | | | | | | | | | | | | 4 | 4 | 5 | | | |
| C ₃₁ | | | | | | | | | | | | | | | | | 3 | 4 | | | |
| C ₃₂ | | | | | | | | | | | | | | | | | 3 | 4 | | | |
| C ₃₃ | | | | | | | | | | | | | | | | | 3 | 4 | | | |
| C ₃₄ | | | | | | | | | | | | | | | | | 3 | | | | |

间窗口中的群落情况,表2中每一时间窗口下的知识群落根据规模由高到低进行排序(小分支群落单列),即并不意味着 t_1 时间窗口中群落 C_{01} 与 t_2-t_{21} 时间窗口中的群落 C_{01} 一定相同,或者一定存在遗传或继承关系。

表2中对各时间窗口中知识群落的数量以及规模进行了统计汇总,结合表1中的相关数据可以发现,在 t_3-t_5 时间窗口中领域知识网络处于连通状态(即不含小分支群落 C_n);而发展过程中的领域知识网络绝大多数时刻处于非连通的状态,即在领域知识网络中存在游离于巨分支之外的孤立分支,而这些小分支所涵盖的该领域内的知识单元不超过16%,进一步说明这些孤立的小分支并非是该领域内的主流方向与核心知识。根据表2中对知识群落的划分结果可以发现,虽然网络巨分支中的知识节点彼此之间相互连通,但是知识节点之间纷繁的关联关系使得巨分支内部仍存在明显的群落结构,且不同群落的规模存在一定差距,如 t_{12} 时间窗口内群落 C_{01} (规模为1 173)与群落 C_{21} (规模为4)之间的差距达到了1169。为了进一步考察所划分出的知识群落是否合理,根据公式(1)分别计算各个时间窗口下的领域知识网络的模块度(Q),如表3所示。

表3 时间序列领域知识网络模块度

| 时间窗口 | t_1 | t_2 | t_3 | t_4 | t_5 | t_6 | t_7 | t_8 | t_9 | t_{10} | t_{11} |
|------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 模块度 | 0.456 5 | 0.455 8 | 0.453 5 | 0.422 8 | 0.429 9 | 0.378 7 | 0.435 0 | 0.586 0 | 0.621 2 | 0.493 7 | 0.509 7 |
| 时间窗口 | t_{12} | t_{13} | t_{14} | t_{15} | t_{16} | t_{17} | t_{18} | t_{19} | t_{20} | t_{21} | |
| 模块度 | 0.484 5 | 0.482 9 | 0.489 9 | 0.487 2 | 0.519 5 | 0.525 8 | 0.639 8 | 0.680 3 | 0.695 7 | 0.640 8 | |

根据表3可知,各个时间窗口下的领域知识网络的模块度值均在0.3~0.7之间,说明研究中依据 Louvain 算法对巨分支进行划分得到的知识群落较为合理。

4.3 群落演变指标计算

结合表2中的数据可以发现,在时间序列早期知识群落数量较少且规模较小,随着领域知识的时序演进知识群落数量逐渐增多,并且有少量的大规模知识群落逐渐涌现出来;而在时间轴后面的区段中,知识群落数量以及最大群落的规模都有所减少。对于一个知识群落 β 来说,随着领域知识的发展,知识群落 β 内部所包含的知识单元并不是固定不变的。从知识群落形成来源的角度来看,知识群落 β 可能继承于上一时间窗口中的某个知识群落,也可能是由上一时间窗口中多个群落中的知识单元重组而形成的,甚至可能是由

新生知识所构成。而从知识群落发展的视角来说,在进入下一时间窗口后,其可能被继承或被融合,甚至可能出现湮灭现象。因此,研究中以知识群落的组成率 I_t 、输出率 E_t 来追踪领域知识群落的动态演变过程与模式。

$$I_t = \frac{\alpha_{t-1} \cap \beta_t}{\beta_t} \quad \text{公式(2)}$$

$$E_t = \frac{\beta_t \cap \gamma_{t+1}}{\beta_t} \quad \text{公式(3)}$$

在公式(2)和公式(3)中, α_{t-1} 表示 $t-1$ 时间窗口内的知识群落 $C_{01}, C_{02}, \dots, C_n$; β_t 表示 t 时间窗口内的知识群落 $C_{01}, C_{02}, \dots, C_n$; γ_{t+1} 表示 $t+1$ 时间窗口内的知识群落 $C_{01}, C_{02}, \dots, C_n$ 。然而,由于各个知识群落的规模、成员等方面存在差异,不能直接通过二者对各群落之间继承与遗传的知识传承情况做出具体测度。因此研究中采用传承强度 L_t 的指标,对群落之间知识传承的完整(或变异)程度进行衡量。

$$L_t = \frac{2 * I_t * E_t}{I_t + E_t} \quad \text{公式(4)}$$

公式(4)中,传承强度 L_t 是组成率 I_t 与输出率 E_t 的乘积比上组成率 I_t 与输出率 E_t 的算术平均数,反映两个相邻时间窗口之间知识群落演变中的知识传承性。研究中基于传承强度 L_t 可以结合群落的组成率 I_t 、输出率 E_t 两项指标,综合考察各个知识群落之间的变异程度与模式。至此,可以结合

各个时间窗口下的领域知识网络的具体特征,对领域知识群落的衍生、交叉、融合与湮灭等过程进行跟踪与分析,从而揭示知识群落演变的模式与知识传承特征。

5 研究结果

5.1 领域知识网络巨分支的统计特征

研究中为了探索领域知识网络中知识群落的演变模式,主要针对领域知识网络中的巨分支进行群落划分,因此研究中首先对领域知识网络中巨分支的基本测度指标进行统计,获得 t_1-t_{21} 时间窗口下领域知识网络的巨分支的节点数量、边数量、密度、聚类系数、特征路径长度、熵值以及谱半径等相关指标,如下页表4所示。

表4汇总了各时间窗口下知识网络的巨分支的基

表4 时间序列领域知识网络巨分支的统计特征

| 时间窗口 | 节点数量 | 边数量 | 密度 | 聚类系数 | 特征路径长度 | 直径 | 熵 | 谱半径 |
|----------|-------|--------|---------|---------|---------|----|---------|-----------|
| t_1 | 66 | 279 | 0.130 1 | 0.333 0 | 1.972 5 | 3 | 3.250 0 | 14.542 7 |
| t_2 | 83 | 342 | 0.100 5 | 0.257 3 | 1.984 1 | 3 | 3.296 6 | 15.717 1 |
| t_3 | 126 | 543 | 0.069 0 | 0.180 2 | 1.990 5 | 3 | 3.439 7 | 18.972 1 |
| t_4 | 126 | 526 | 0.066 8 | 0.163 5 | 2.005 0 | 3 | 3.278 4 | 18.854 2 |
| t_5 | 133 | 565 | 0.064 4 | 0.159 0 | 2.004 0 | 3 | 3.246 4 | 19.436 6 |
| t_6 | 82 | 327 | 0.098 5 | 0.205 2 | 1.901 5 | 2 | 2.815 1 | 15.700 7 |
| t_7 | 85 | 344 | 0.096 4 | 0.233 4 | 1.974 2 | 3 | 3.159 7 | 15.727 7 |
| t_8 | 139 | 482 | 0.050 3 | 0.291 2 | 3.088 3 | 8 | 3.366 4 | 14.916 0 |
| t_9 | 490 | 2 580 | 0.021 5 | 0.316 8 | 2.942 9 | 7 | 4.436 9 | 25.126 7 |
| t_{10} | 1 733 | 21 586 | 0.014 4 | 0.330 2 | 2.636 3 | 7 | 5.547 8 | 86.615 4 |
| t_{11} | 3 357 | 47 896 | 0.008 5 | 0.247 9 | 2.599 3 | 7 | 5.790 0 | 112.600 3 |
| t_{12} | 4 482 | 59 982 | 0.006 0 | 0.199 7 | 2.682 5 | 7 | 5.700 0 | 118.453 7 |
| t_{13} | 4 695 | 51 529 | 0.004 7 | 0.172 4 | 2.787 0 | 8 | 5.419 8 | 109.457 5 |
| t_{14} | 4 179 | 34 691 | 0.004 0 | 0.119 3 | 2.968 5 | 8 | 5.067 9 | 77.687 4 |
| t_{15} | 3 815 | 33 315 | 0.004 6 | 0.127 7 | 2.897 4 | 9 | 5.125 3 | 86.797 9 |
| t_{16} | 3 266 | 29 586 | 0.005 5 | 0.156 5 | 2.865 1 | 8 | 5.210 2 | 83.000 4 |
| t_{17} | 2 632 | 21 535 | 0.006 2 | 0.170 7 | 2.977 2 | 11 | 4.991 6 | 77.678 5 |
| t_{18} | 1 960 | 14 734 | 0.007 7 | 0.550 4 | 3.356 5 | 10 | 4.591 7 | 91.395 7 |
| t_{19} | 1 790 | 12 740 | 0.008 0 | 0.600 4 | 3.536 3 | 10 | 4.502 8 | 91.370 1 |
| t_{20} | 1 945 | 16 355 | 0.008 7 | 0.511 6 | 3.276 9 | 10 | 4.737 8 | 91.315 8 |
| t_{21} | 1 711 | 11 513 | 0.007 9 | 0.262 1 | 3.125 5 | 10 | 4.667 1 | 61.329 2 |

本指标,结合表1中领域知识网络中标签总量以及关联关系总量可以发现,领域知识在发展演进的过程中,会呈现内部知识单元连通的现象(如表1中 t_3 — t_5 时间窗口)。但是在更长的知识发展进程中,领域知识网络内部存在一个巨大的分支结构。该巨分支结构涵盖了该领域80%以上的知识节点(对照表1和表4)。事实上,表4中的巨分支还囊括了知识网络90%以上的知识节点之间的关联关系,侧面说明领域知识网络中的巨分支在一定程度上可以反映该领域内的主流研究方向与核心知识。结合表1和表4中相关数据的波动变化可以发现,在 t_8 时间窗口后,知识网络与巨分支的节点数量以及关联关系数量显著增加,而在 t_{13} 时间窗口之后其数量总体上呈现下降趋势,进一步说明在领域知识的发展进程中存在明显的知识单元的新生与湮灭现象。表4中知识网络巨分支的低密度则说明,尽管巨分支在整体上处于连通状态,但仍是一个较为稀疏的网络结构。

另一方面,表4中领域知识网络中巨分支结构的聚类系数明显高于同等规模的随机网络,说明知识网络巨分支结构中的知识节点存在着聚集现象。网络巨分支的特征路径长度平均为2.65跳距离,最大直径为11,表明巨分支内任意两个知识节点之间最多只需要

11跳距离就能够实现相互连接。因此,可以说明领域知识网络中的巨分支结构同样具有较高聚类系数、较短特征路径长度的小世界特性。表4中熵值的变化情况说明领域知识在发展过程中随机性与非随机性交互影响。谱半径是对网络拓扑结构进行描述的,因此可以根据其变化判断领域知识网络内部的拓扑结构是否发生变化,进而识别领域知识的发展演进路线。结合网络指标的统计特征,将领域知识网络巨分支分为如下几个阶段:Ⅰ阶段(t_1 — t_7),这一阶段各时间窗口下的巨分支规模较小,熵值和谱半径相对较低,巨分支的随机性较低,在时间轴上的变化程度小;Ⅱ阶段(t_7 — t_{12}),大量节点开始涌入,巨分支规模迅速扩张,熵值和谱半径升高,随机性增强,在时间轴上的变化程度较大;Ⅲ阶段(t_{12} — t_{21}),巨分支开始紧缩,熵值和谱半径有所降低,大量节点离开巨分支甚至产生湮灭现象。

5.2 群落演变与传承强度分析

在知识群落的动态演进过程中,组成率 I_i 与输出率 E_i 是衡量知识群落内各成员要素沿时间轴演化过程中传递知识的指标。其中组成率 I_i 可以反映知识群落的成员构成,可以看出其内部知识单元来自上一时间窗口中哪些知识群落,侧重于知识群落的来源情况。由于社群分析的数据量庞大,行文中无法全部展示,以知识群落从 t_{19} 时间窗口进入 t_{20} 时间窗口的情况为例,组成率 I_i 计算结果如表5所示。输出率 E_i 可以跟踪知识群落内的知识单元去往下一时间窗口中哪些知识群落之中,以知识群落从 t_{20} 时间窗口进入 t_{21} 时间窗口的情况为例,输出率 E_i 计算结果如表6所示。

表5和表6中, C_{0i} — C_n 表示所对应的时间窗口内的知识群落, C_x 表示对应时间窗口下的小分支群落。 C -out表示所对应时间窗口下的知识群落在相邻的下一时间窗口中发生湮灭的知识单元集合, C -in表示所对应时间窗口内的新生的知识单元集合。为了便于厘清知识群落演进的接续性,将表5中的行数据和表6中的列数据相结合进行分析。 t_{20} 时间窗口的知识群落 C_{01} 对应 t_{19} 时间窗口中的群落 C_{01} 的组成率值为1(参见表5),这意味着 t_{20} 时间窗口的知识群落 C_{01} 完全继承了 t_{19} 时间窗口中群落 C_{01} 内部的知识单元;在进入 t_{21} 时间窗口后,知识群落 C_{01} 内部85.25%的知识单元出现了湮灭现象(参见表6)。表5中 t_{20} 时间窗口的知识群落 C_{02} 则并非主要继承 t_{19} 时间窗口中的某一知识群落,而是汇

表5 t₂₀时间窗口知识群落组成率(节选)

| I _t | | t ₁₉ | | | | | | | | | | | | | | | | t ₂₀ |
|-----------------|----------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|----------------------------------|-----------------|----------------------------------|-----------------|-----------------|-----------------|----------------------------------|-----------------|----------------------------------|-----------------|----------------|-----------------|
| | | C ₀₁ | C ₀₂ | C ₀₃ | C ₀₄ | C ₀₅ | C ₀₆ -C ₀₇ | C ₀₈ | C ₀₉ -C ₁₀ | C ₁₁ | C ₁₂ | C ₁₃ | C ₁₄ -C ₁₇ | C ₁₈ | C ₁₉ -C ₂₁ | C ₂₂ | C _X | C-in |
| t ₁₉ | C-out | 0.031 3 | 0 | 0.091 5 | 0 | 0.060 3 | ... | 0.069 2 | ... | 0.024 6 | 0.053 6 | 0.013 4 | ... | 0.091 5 | ... | 0.183 | 0.187 5 | |
| t ₂₀ | C ₀₁ | 1 | 0 | 0 | 0 | 0 | ... | 0 | ... | 0 | 0 | 0 | ... | 0 | ... | 0 | 0 | 0 |
| | C ₀₂ | 0 | 0 | 0.014 9 | 0 | 0.064 4 | ... | 0.005 | ... | 0.044 6 | 0.084 2 | 0.014 9 | ... | 0.054 5 | ... | 0.311 9 | 0.044 6 | 0.361 4 |
| | C ₀₃ -C ₀₅ | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| | C ₀₆ | 0 | 0 | 0 | 0 | 0 | ... | 0.2 | ... | 0 | 0 | 0 | ... | 0 | ... | 0 | 0 | 0.8 |
| | C ₀₇ -C ₁₁ | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| | C ₁₂ | 0 | 0 | 0 | 0 | 0 | ... | 0 | ... | 0 | 0.954 5 | 0 | ... | 0 | ... | 0 | 0 | 0.045 5 |
| | C ₁₃ -C ₁₉ | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| | C _X | 0 | 0 | 0.030 2 | 0 | 0.025 1 | ... | 0.005 | ... | 0.005 | 0 | 0 | ... | 0.025 1 | ... | 0.040 2 | 0.517 6 | 0.266 3 |

聚了t₁₉时间窗口中多个群落(C₀₃、C₀₅、C₀₈、C₁₁、C₁₂、C₁₃、C₁₈、C₂₂、C_X)内的知识单元和部分t₂₀时间窗口的新生知识单元(C-in)形成的;表6的t₂₁时间窗口中该群落内的知识单元又分散输出到多个知识群落中,这说明该群落并未形成相对稳定的核心成员簇。表5中t₂₀时间窗口的知识群落C₀₆内部的知识单元有80%来自新生知识单元,20%来自t₁₉时间窗口的群落C₀₈;表6的t₂₁时间窗口中该群落完全融入群落C₀₇内,但是在限于篇幅原因未列示的t₂₁时间窗口的组成率计算结果中,这部分知识仅仅占据群落C₀₇知识单元的2%,表6的数据还显示,有多个群落(包括新生知识单元)都向t₂₁时间窗口的群落C₀₇输出知识单元。t₂₀时间窗口中的群落C₁₂继承了t₁₉时间窗口的群落C₁₂内95.45%的知识单元(参见表5);进入t₂₁时间窗口后有9.09%的知识单元发生了湮灭,剩余知识单元均匀地输出到知识群落C₁₂和C₁₈内(参见表6)。

从上述领域知识群落动态演化过程的分析中已经初步发现,知识群落组成率I_t与输出率E_t的结合分析,可以清晰地跟踪知识群落在时间序列上的演变过程,即跟踪群落内部知识构成的来源及其去向。研究中进一步将传承强度L_t引入分析过程,综合考察群落间知识的传承程度与模式。以t₂₀时间窗口内的知识群落进入t₂₁时间窗口各群落的情况为例,各个知识群落间的传承强度如表7所示。

表6 t₂₀时间窗口知识群落输出率(节选)

| E _t | | t ₂₀ | | | | | | | | | t ₂₁ |
|-----------------|----------------------------------|-----------------|-----------------|----------------------------------|-----------------|----------------------------------|-----------------|-----------------|----------------------------------|----------------|-----------------|
| | | C ₀₁ | C ₀₂ | C ₀₃ -C ₀₅ | C ₀₆ | C ₀₇ -C ₁₀ | C ₁₁ | C ₁₂ | C ₁₃ -C ₁₉ | C _X | C-in |
| t ₂₀ | C-out | 0.852 5 | 0.183 2 | ... | 0 | ... | 0 | 0.090 9 | ... | 0.311 6 | |
| t ₂₁ | C ₀₁ | 0.090 2 | 0 | ... | 0 | ... | 0 | 0 | ... | 0 | 0 |
| | C ₀₂ | 0 | 0.217 8 | ... | 0 | ... | 0 | 0 | ... | 0 | 0.003 2 |
| | C ₀₃ | 0 | 0 | ... | 0 | ... | 0 | 0 | ... | 0 | 0 |
| | C ₀₄ | 0.008 2 | 0.104 | ... | 0 | ... | 0 | 0 | ... | 0 | 0.022 5 |
| | C ₀₅ | 0.008 2 | 0 | ... | 0 | ... | 0 | 0 | ... | 0 | 0.003 2 |
| | C ₀₇ | 0.008 2 | 0.039 6 | ... | 1 | ... | 0 | 0 | ... | 0.005 | 0.038 6 |
| | C ₀₈ -C ₀₉ | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| | C ₁₀ | 0 | 0.029 7 | ... | 0 | ... | 0 | 0 | ... | 0 | 0.035 4 |
| | C ₁₁ | 0 | 0 | ... | 0 | ... | 0.931 5 | 0 | ... | 0 | 0.003 2 |
| | C ₁₂ | 0 | 0 | ... | 0 | ... | 0 | 0.454 5 | ... | 0 | 0 |
| | C ₁₃ | 0 | 0.019 8 | ... | 0 | ... | 0 | 0 | ... | 0.005 | 0.025 7 |
| | C ₁₄ -C ₁₆ | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| | C ₁₇ | 0.008 2 | 0.099 | ... | 0 | ... | 0 | 0 | ... | 0.005 | 0.151 1 |
| | C ₁₈ | 0.016 4 | 0.207 9 | ... | 0 | ... | 0.041 1 | 0.454 5 | ... | 0.025 1 | 0.196 1 |
| | C ₁₉ -C ₂₀ | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| C ₂₁ | 0.008 2 | 0.074 3 | ... | 0 | ... | 0.027 4 | 0 | ... | 0 | 0.263 7 | |
| C _X | 0 | 0.024 8 | ... | 0 | ... | 0 | 0 | ... | 0.648 2 | 0.151 1 | |

表7 t₂₀-t₂₁时间窗口间知识群落传承强度(节选)

| L _t | | t ₂₀ | | | | | t ₂₁ |
|-----------------|----------------------------------|----------------------------------|-----------------|-----------------|----------------------------------|----------------|-----------------|
| | | C ₀₁ -C ₁₀ | C ₁₁ | C ₁₂ | C ₁₃ -C ₁₉ | C _X | C-in |
| t ₂₀ | C-out | ... | N/A | 0.007 08 | ... | 0.167 116 | N/A |
| t ₂₁ | C ₀₁ -C ₁₀ | ... | ... | ... | ... | ... | ... |
| | C ₁₁ | ... | 0.957 746 | N/A | ... | N/A | 0.005 263 |
| | C ₁₂ | ... | N/A | 0.625 | ... | N/A | N/A |
| | C ₁₃ -C ₁₇ | ... | ... | ... | ... | ... | ... |
| | C ₁₈ | ... | 0.017 493 | 0.068 493 | ... | 0.021 322 | 0.209 983 |
| | C ₁₉ -C ₂₁ | ... | ... | ... | ... | ... | ... |
| | C _X | ... | N/A | N/A | ... | 0.645 | 0.183 594 |

表7中的数据显, t₂₀时间窗口的群落C₁₁演变成t₂₁时间窗口的群落C₁₁的传承强度达到了0.957 746,说明t₂₀时间窗口的群落C₁₁在演进过程中极大程度上被t₂₁时间窗口的群落C₁₁所继承。结合表6中的输出率来看, t₂₀时间窗口的群落C₁₁内部约93%的知识单元进入t₂₁时间窗口的群落C₁₁(即输出率约为93%)。并且,在

t_{21} 时间窗口的组成率(本文未列示)计算结果中,来自 t_{20} 时间窗口群落 C_{11} 的知识单元占据了 t_{21} 时间窗口的群落 C_{11} 内约99%的知识单元,进而综合表现出极高的传承强度。而表7中 t_{20} 时间窗口的群落 C_{12} 以0.625的传承强度进入 t_{21} 时间窗口的群落 C_{12} ,以0.068 493的传承强度进入群落 C_{18} ,以0.007 08的强度发生了湮灭现象。结合表6中 t_{20} 时间窗口的群落 C_{12} 的输出率可以发现,其仅有少量(约为9%)的知识单元在 t_{21} 时间窗口中发生了湮灭,而其余知识单元均匀地分别进入 t_{21} 时间窗口的群落 C_{12} 和群落 C_{18} (输出率分别约为45%)。 t_{20} 时间窗口的群落 C_{12} 虽对这两个群落的输出程度相同,但是传承强度存在很大差异。基于 t_{21} 时间窗口的组成率(本文未列示)计算结果的分析中发现,这一现象主要是由于 t_{21} 时间窗口的群落 C_{12} 完全继承于 t_{20} 时间窗口的群落 C_{12} ,而 t_{21} 时间窗口的群落 C_{18} 内仅有4%的知识单元继承于 t_{20} 时间窗口的群落 C_{12} 。上述结果说明,传承强度 L_i 可以更好地衡量领域知识发展演进过程中群落之间的知识传承情况。

结合上述基于谱半径、节点数量等指标信息对知识发展进程划分的3个阶段,分别考察Ⅰ、Ⅱ、Ⅲ阶段内各知识群落的组成率 I_i 、输出率 E_i 以及传承强度 L_i 的分布情况,如图2所示。

图2中,3个指标对应的散点图的密集中心处于低组成率、低输出率、低传承强度的区域。这一特征持续于领域知识发展演进的整个过程,特别是在Ⅲ阶段(t_{12} — t_{21})中该特征尤为明显。这一现象说明在领域知识发展演进的过程中,大多数知识群落之间往往会有大量的知识单元发生跨群落的交叉融合,对应的知识群落的

传承强度较小。即使在效果表现较弱的Ⅰ阶段,结合知识群落的组成率和输出率数据(未列示)发现, t_3 时间窗口群落 C_{03} 在演变成 t_4 时间窗口群落 C_{04} 的过程中,其内部仅有12%的节点输出给群落 C_{04} ,而这部分知识单元仅占群落 C_{04} 内部知识节点的20%,且 t_3 时间窗口的群落 C_{03} 演变成 t_4 时间窗口群落 C_{04} 的传承强度仅为0.151 5。类似的情况在领域知识演进的全过程中占有很高的概率。图2中的一个特殊现象是在B、C的邻近区域内,在Ⅰ阶段(t_1 — t_7)几乎没有高输出率、低组成率(或低输出率、高组成率)的演变现象,而在Ⅱ(t_7 — t_{12})、Ⅲ(t_{12} — t_{21})阶段中这两种现象逐渐涌现出来。这一结果表明,随着领域知识的发展演进,群落内的少量知识单元在下一时间窗口可能逐渐成为某群落内的核心知识簇(低输出率,高组成率);或发生知识群落内部大量知识节点进入下一时间窗口某群落中,但在新的群落中仅占有很小的比例(高输出率、低组成率)。

5.3 演化模式与传承特征分析

经过上述分析可以发现,通过组成率 I_i 与输出率 E_i 可以跟踪知识群落的演变来源与去向,而基于传承强度 L_i 则可以考察前后不同时刻知识群落之间的知识传承情况。因此,研究中利用小提琴图(Violinplot)对知识群落传承强度 L_i 在时间序列上的分布及其概率密度进行分析,如图3所示。

小提琴图(Violinplot)结合了箱线图与核密度图的特征展示数据的分布及密度。由图3可知, t_1 — t_2 时间窗口间的传承强度的中位数居于靠近中间的位置,密度图宽变化不大,表明群落传承强度的分布比较分散。在后续的时间序列中,密度图宽逐渐呈现出“上收敛—

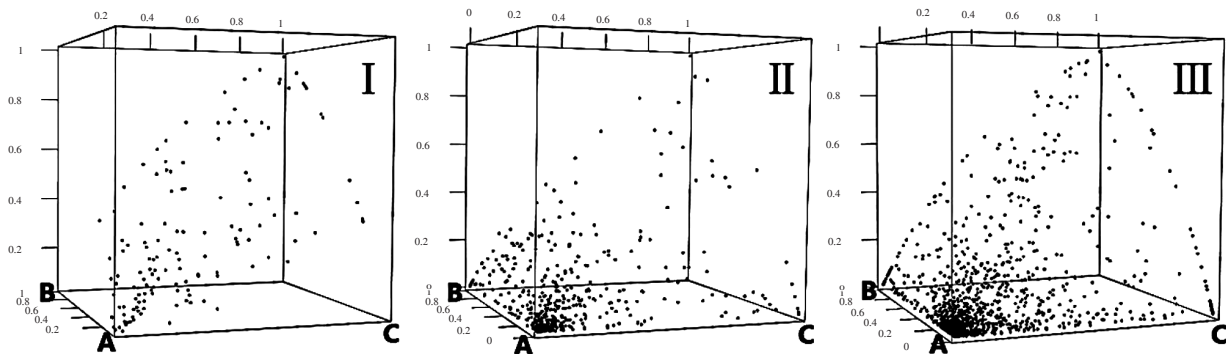


图2 知识群落组成率、输出率及传承强度分布

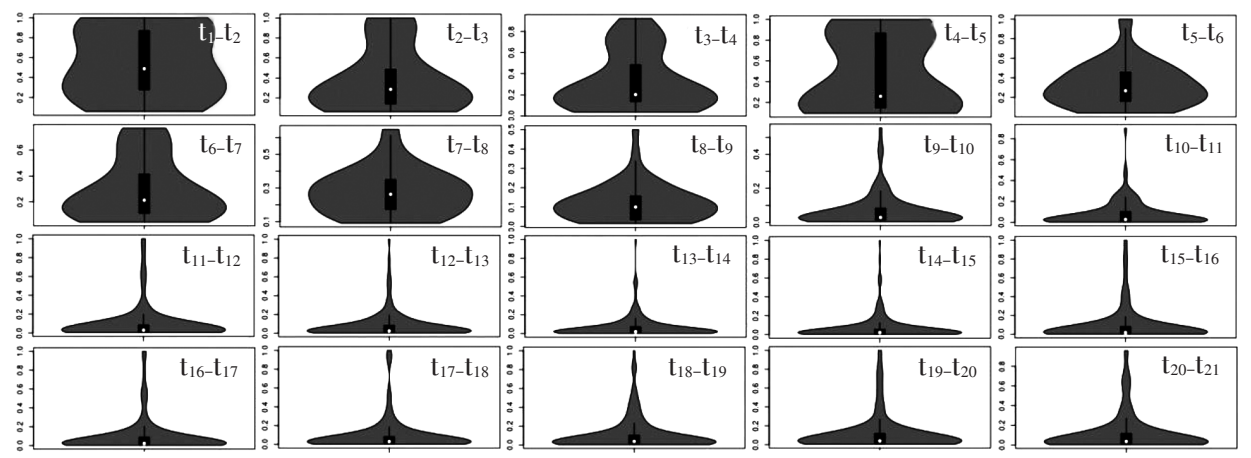


图3 知识群落传承强度的时序变化

下扩张”的态势。即使在表现较弱的 t_4 - t_5 时间窗口间,中位数也明显被拉近下四分位数,而且截止到 t_8 - t_9 阶段,各个跨时间窗口演变的传承强度小提琴图的上侧须(中间细线)较长,说明该区段仍然存在比较明显的离散值。进一步沿时间轴推移可以发现,密度图宽“上收敛-下扩张”的趋势愈发明显,进一步验证了图2中的数据分布特征。这说明在领域知识发展演进过程中,知识群落的演变特征逐渐集中并涌现出来,即在演进过程的中后期知识网络内各个群落并非会高概率地被下一时刻的某一知识群落独立完全继承,通常情况下存在传承强度较低(分散交叉)的群落演化模式。这同时说明该领域的知识进入了蓬勃发展阶段,跨群落的知识交叉与融合频繁发生。知识群落在时间序列上的演变过程中,可能进入下一时间窗口的巨分支内部的群落或小分支群落,甚至可能进入了湮灭的知识单元集合。因此有必要对知识群落演变强度 L_t 的相关指标进行统计计算(结果如表8所示),进而更深入地探索知识群落的演变模式与知识传承。

表8中,Time为时间窗口,Clu为知识群落,Pre-Tot表示 $t-1$ 间窗口的各个知识群落演变为 t 时间窗口该群落的传承强度总和,Pre-Sd表示 $t-1$ 时间窗口各个知识群落演变为 t 时间窗口该知识群落的传承强度的标准差,Pre-Nu表示 t 时间窗口该知识群落与 $t-1$ 时间窗口中有演变关联的知识群落的总数,Pre-BT表示 $t-1$ 时间窗口巨分支内部各个知识群落对于 t 时间窗口该知识群落的传承强度总和,Pre-BSd表示 $t-1$ 时间窗口巨分支内部各个知识群落对于 t 时间窗口该知识群落的传承强度的标准差,Pre-BNu表示 t 时间窗口该知识群落与 $t-1$ 时间窗口巨分支中有演变关联的知识群落的总数,Pre-In表示 t 时间窗口新生知识进入 t 时间窗口该知识群落的传承强度,Pre-X表示 $t-1$ 时间窗口的小分支群落演变为 t 时间窗口该群落的传承强度。Ne-*同上,是该知识群落向 $t+1$ 时间窗口演变的传承强度统计指标,其中Ne-Out表示该群落在 $t+1$ 时间窗口发生湮灭的传承强度。

根据表8中的相关数据可以发现, t_2 时间窗口中的知识群落 C_{01} 的Pre-Tot、Pre-Nu、Pre-BT、Pre-BNu值均

表8 知识群落传承强度指标统计测算(节选)

| Time | Clu | Pre-Tot | Pre-Sd | Pre-Nu | Pre-BT | Pre-BSd | Pre-BNu | Pre-In | Pre-X | Ne-Tot | Ne-Sd | Ne-Nu | Ne-BT | Ne-BSd | Ne-BNu | Ne-Out | Ne-X |
|----------|----------|---------|---------|--------|--------|---------|---------|---------|-------|---------|---------|-------|---------|---------|--------|---------|---------|
| ⋮ | | | | | | | | | | | | | | | | | |
| t_2 | C_{01} | 1 | N/A | 1 | 1 | N/A | 1 | N/A | N/A | 1 | N/A | 1 | 1 | N/A | 1 | N/A | N/A |
| | C_{02} | 0.9 | N/A | 1 | 0.9 | N/A | 1 | N/A | N/A | 1.024 5 | 0.428 9 | 2 | 0.941 2 | N/A | 1 | 0.083 3 | N/A |
| ⋮ | | | | | | | | | | | | | | | | | |
| t_{13} | C_{12} | 1.049 8 | 0.090 8 | 15 | 0.855 | 0.089 4 | 13 | 0.192 8 | 0.002 | 1.065 | 0.087 7 | 18 | 0.858 6 | 0.085 5 | 16 | 0.196 3 | 0.010 2 |
| ⋮ | | | | | | | | | | | | | | | | | |

为1,说明 t_2 时间窗口中的群落 C_{01} 主要是由 t_1 时间窗口巨分支内的一个群落演变而来,表明这部分知识群落的传承强度极高,即发生了完全继承的演变模式。群落 C_{01} 的Ne-Tot、Ne-Nu、Ne-BT、Ne-BNu值同样为1,说明在 t_3 时间窗口中该群落又整体进入了网络巨分支内部的一个知识群落之中,该群落内部的知识单元较为稳定成熟,并未发生分裂现象。 t_2 时间窗口的群落 C_{02} 也是由 t_1 时间窗口的网络巨分支内部的一个群落演变而来,但是其Pre-Tot与Pre-BT值为0.9却说明群落 C_{02} 并未获得 t_1 时间窗口关联群落的全部知识。在随后的演变过程中,其Ne-Sd值相对较高,说明该群落在进入 t_3 时间窗口的过程中传承强度分布较为离散,主要以0.941 2的强度进入巨分支内部的群落之中,其余则发生了湮灭现象。 t_3 时间窗口的群落 C_{12} 的Pre-Tot值为1.05,Pre-Sd值为0.090 8,说明群落 C_{12} 来自 t_2 时间窗口的传承强度虽高,但是与之相关联的知识群落共有15个,其中0.192 8的传承强度来源于新生知识单元集合,有0.002的强度是由 t_2 时间窗口中的小分支群落所提供,说明群落 C_{12} 主要是由 t_2 时间窗口中的众多群落分裂并重新融合而来。该群落在后继的演变中又相对均匀地分裂进入 t_4 时间窗口中包括小分支群落和湮灭知识单元集合在内的18个知识群落之中。

事实上,表8所代表的数据共涉及369个知识群落和2155次知识群落之间的演变。其中1457次群落演变发生在巨分支内部群落之间,占总演变次数的67.61%,平均传承强度仅为0.117 775。高演变次数与低传承强度意味着,领域知识跨群落的交叉融合主要发生在巨分支内部的知识群落之间。

6 结论与讨论

本文对社会化标注系统中所获取的数据进行平滑处理,构建不同时间窗口下的领域知识网络,基于社群发现算法对领域知识网络内的巨分支进行群落划分。通过对时间序列领域知识群落的来源与去向的动态跟踪,分析知识群落的演变模式与知识传承特征。综合上述对领域知识群落演化过程的分析结果,初步可以得出如下结论。

(1)领域知识网络在发展演变过程中更多地处于非连通状态。以往关于网络非连通状态的研究结论主要是针对静态网络得出的。文中针对知识网络的动态

分析结果表明,知识网络仅在较短的时段内处于连通状态,更多时刻则处于非连通的状态(参见表2),由庞大的巨分支结构和一些孤立的小分支结构所组成。由于巨分支涵盖了领域中绝大多数知识单元(>80%)和更丰富的知识关联关系(>90%),因此代表了该领域的主流方向和核心知识。从领域知识动态演化的视角诠释了网络分支结构的常态性。

(2)领域知识发展越成熟群落之间的知识交叉融合越频繁。研究结果显示,尽管低组成率、低输出率和低传承强度并存的现象贯穿领域知识发展全程,但是随着领域知识的发展演进,低组成率、低输出率和低传承强度并存的演变模式逐渐涌现为主流的群落演化模式。图2中分阶段的点密集区域和图3中密度图宽的时序变化,都说明了这一演化模式的渐进性。这意味着领域知识发展越成熟,演进过程中知识群落之间的知识交叉与融合越频繁。这一结论为领域知识发展成熟度的判识提供了一种新的思路。

(3)交叉融合性的知识传承多发生在巨分支内部的知识群落之间。研究中发现,领域知识发展演进过程中的群落演变主要存在于巨分支内部的知识群落之间(67.61%),但是平均传承强度只有0.117 775。说明了领域知识的交叉融合更多发生在巨分支内部的知识群落之间,低值的传承强度意味着群落演变过程中的交叉融合现象高于完整独立传承。这一方面证实了知识网络的巨分支在知识组织、知识生长等研究中的代表性,同时也为以往以网络巨分支代表整个知识网络的研究工作提供了相关支撑。

本研究采用社区发现算法识别并提取了领域知识网络中巨分支内部的知识群落,并在时间序列上对巨分支内部知识群落、小分支群落、湮灭以及新生知识单元集合进行动态跟踪与分析。研究中提出基于知识群落的组成率、输出率以及传承强度等指标来识别与分析领域知识演进过程中的群落演化模式与知识传承特征,不仅丰富了知识组织、知识传播的相关理论与方法,同时也有助于对于科研合作网络、信息传播网络中知识传播与共享进行分析与揭示。研究中也存在不足之处,由于数据量与计算量较大,无法全部展示和逐一说明,论述中难免有疏漏之处。未来的研究工作将进一步针对具体问题展开更深入细致的分析,以期更全面准确地对领域知识发展演进过程中的知识传承模式

与规律进行探索与揭示。

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.
- [2] 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: 4-5.
- [3] Ravasz E, Somera A L, Mongru D A, et al. Hierarchical organization of modularity in metabolic networks[J]. Science, 2002, 297(5586): 1551-1555.
- [4] Gui X, Li L, Cao J, et al. Dynamic communities in stock market [J]. Abstract and Applied Analysis, 2014, 2014(10): 1-9.
- [5] Ali A, Qadir J, Rasool R U, et al. Big data for development: applications and techniques[J]. Big Data Analytics, 2016, 1(1): 2.
- [6] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2): 026113.
- [7] Fudholi D H, Rahayu W, Pardede E. A data-driven dynamic ontology[J]. Journal of Information Science, 2015, 41(3): 383-398.
- [8] Wallace M L, Gingras Y, Duhon R. A new approach for detecting scientific specialties from raw cocitation networks[J]. Journal of the American Society for Information Science and Technology, 2009, 60(2): 240-246.
- [9] Yi S, Choi J. The organization of scientific knowledge: the structural characteristics of keyword networks[J]. Scientometrics, 2012, 90(3): 1015-1026.
- [10] Dong G, Fan J, Shekhtman L M, et al. Resilience of networks with community structure behaves as if under an external field [J]. Proceedings of the National Academy of Sciences, 2018, 115(27): 6911-6915.
- [11] McCain K W. Assessing an author's influence using time series historiographic mapping: the oeuvre of conrad hal waddington (1905-1975)[J]. Journal of the American Society for Information Science and Technology, 2008, 59(4): 510-525.
- [12] Lancichinetti A, Fortunato S. Consensus clustering in complex networks[J]. Scientific Reports, 2012, 2(13): 336.
- [13] Liu Q, Liu C, Wang J, et al. Evolutionary link community structure discovery in dynamic weighted networks[J]. Physica A: Statistical Mechanics and Its Applications, 2017, 466(C): 370-388.
- [14] Wang X, Cheng Q, Lu W. Analyzing evolution of research topics with NEViewer: a new method based on dynamic co-word networks[J]. Scientometrics, 2014, 101(2): 1253-1271.
- [15] 刘自强, 王效岳, 白如江. 语义分类的学科主题演化分析方法研究:以我国图书情报领域大数据研究为例[J]. 图书情报工作, 2016, 60(15): 76-85, 93.
- [16] 朱梦娴, 程齐凯, 陆伟. 基于社会网络的学科主题聚类研究[J]. 情报杂志, 2012, 31(11): 40-44.
- [17] 李纲, 李岚凤, 毛进, 等. 作者合著网络中研究兴趣相似性实证研究[J]. 图书情报工作, 2015, 59(2): 75-81.
- [18] 白如江, 冷伏海. k-clique社区知识创新演化方法研究[J]. 图书情报工作, 2013, 57(17): 86-94.
- [19] 程齐凯, 王晓光. 一种基于共词网络社区的科研主题演化分析框架[J]. 图书情报工作, 2013 (8): 91-96.
- [20] 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.
- [21] 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: 80-90, 152-160.
- [22] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008(10): P10008.
- [23] 汪小帆, 李翔, 陈关荣. 网络科学导论[M]. 北京: 高等教育出版社, 2012: 130-140.
- [24] Tang L, Liu H. 社会计算:社区发现和社会媒体挖掘[M]. 文益民, 闭应洲, 译. 北京:机械工业出版社, 2013: 20-30.

【作者简介】安 宁,女,1993年生,东北师范大学信息科学与技术学院硕士研究生。

孙熊兰,女,1996年生,东北师范大学信息科学与技术学院硕士研究生。

滕广青,男,1970年生,东北师范大学信息科学与技术学院教授,博士生导师(通讯作者)。

栾 宇,男,1988年生,东北师范大学信息科学与技术学院硕士研究生。

收稿日期:2019-02-14

