

基于频度演化的领域知识关联关系涌现^{*}

滕广青

摘 要 领域知识的生长演化问题一直是图书情报学界重点关注的主题。以网络科学思维探索知识生长过程中的关联关系涌现问题,能够对知识关联的生长模式与机制进行揭示。本研究提取知识关联关系累计 44 万余对,关联频度 87 万余次,共划分为 11 个时间窗口。在对时间序列领域知识网络结构属性初步判识的基础上,对关联频度分布进行时间序列分析。并从领域知识生长过程中的关联频度数量、关联频度占比、邻近窗口状态等方面,对知识关联关系及其频度进行跟踪与分析。研究结果表明,知识关联关系生长过程中,关联关系频度的分布符合幂律分布,且在领域知识发展的成熟期表现得更好。知识关联关系的生长过程具有频度层面的“富者更富”的属性,且主要遵循“择优强化”机制。尽管研究所使用的基于社会化标注系统的知识网络尚不足以囊括所有类型的知识网络,但是基于频度演化的知识关联关系涌现模式与机制,有助于促进知识网络、知识生长等领域的研究工作,对于社交网络、传播网络、交通网络等研究不无裨益。图 5。表 4。参考文献 30。

关键词 知识网络 知识关联 关联频度 关联涌现

分类号 G254

Emergence of Correlation in Domain Knowledge Based on Frequency Evolution

TENG Guangqing

ABSTRACT

The growth and evolution of domain knowledge have always been the focus of Library and Information Science. Exploring the emergence of correlations in the process of knowledge growth with network science thinking can reveal the growth patterns and mechanisms of knowledge correlations. The present research extracts a total of over 440 000 pairs of knowledge correlations, 870 000 times of correlation frequency, which are divided into 11 time windows. The small-world and scale-free properties of time series domain knowledge network topology are determined by the short average path length, the high clustering coefficient and the power-law distribution of degree. On this basis, the knowledge correlations and frequencies in the process of domain knowledge growth are tracked and analyzed from the aspects of the number of correlation

^{*} 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)

frequencies, the proportion of correlation frequencies, the status of neighboring windows, etc.

Results have shown that in the process of knowledge correlation growth, the frequency distribution of correlations is in accordance with power-law. Through the time series analysis of the frequency of knowledge correlations, it is found that the frequency distribution of knowledge correlations in the same network can better fit the power-law than the degree distribution of nodes, without the phenomenon of “top-heavy distribution” usually found in the degree distribution of nodes, and the power-law distribution of the frequency of knowledge correlations performs better in the latter part of the time series. This phenomenon shows that although the knowledge correlations determine the topology of knowledge networks, only a few correlations have extremely high frequency values, while most of them have only a small number of frequency values, and the process of frequency of knowledge correlations is that of frequency emergence. On the other hand, the growth process of knowledge correlations has the property of “the rich gets richer” in frequency, and mainly follows the mechanism of “preferential reinforcement”.

Statistics show that as the frequency of knowledge correlations grows and accumulates, the few frequency “rich” (high-frequency correlation) occupy more and more frequency “wealth” than most frequency “poor” (low-frequency correlation). The gap between the “rich” and the “poor” in frequency of knowledge correlations becomes more and more obvious in the second half of the timeline as the domain knowledge grows and develops. The status of the neighboring windows shows that most of the knowledge correlations in the frequency “extremely rich” status often have the “extremely rich” status in the previous time window. Correlations with more frequency “wealth” will attract additional frequencies with a higher probability in the process of growth and development of domain knowledge, reflecting the frequency growth mechanism of “preferential reinforcement”. The repeated superimposition of the micro-rules of “preferential reinforcement” on the time series have created the phenomena of “the rich get richer” at the frequency level in the emergence of correlation. The study has also found the “bursts reinforcement” phenomenon that the correlations in frequency “extremely poor” status suddenly jump into the “extremely rich” status. The main reason for this phenomenon is that academia has produced significant findings or inventions, which provide a possibility for identifying significant academic achievements based on the frequency of knowledge correlation.

Although the knowledge networks based on social tagging system used in this study are not comprehensive to cover all types of knowledge networks, the emergence patterns and mechanisms of knowledge correlation based on frequency evolution help to promote research in knowledge networks and knowledge growth, and they also benefit studies on social networks, communication networks, transport networks, etc. 5 figs. 4 tabs. 30 refs.

KEY WORDS

Knowledge network. Knowledge correlation. Correlation frequency. Correlation emergence.

0 引言

随着网络科学相关研究成果的不断发表与

传播,现实社会中诸多真实网络的生长性在学术界已经成为不争的事实。图书情报学领域内各类知识网络随着领域知识的不断积累与发展也同样表现出其生长发展的特征。以往考量网

络的生长性主要通过对网络节点和单频次连线规模的测度,但是网络关联关系频度的增加同样是网络生长不可忽视的因素。特别是在知识科学领域,没有哪个知识因子或知识单元能够一枝独秀,总是要在不同维度、不同程度地与其他知识形成关联关系。其中的关联程度能够反映知识之间关联关系的强弱、深浅。正是借助知识之间潜在的关联关系,图书情报学界才能对人类知识的发展、传承、热点、前沿等问题展开研究。作为知识网络研究重点内容的知识关联关系从网络科学的理论与技术中获得了强大助力的同时,学术界也注意到知识之间的关联关系并非是一成不变的。处于生长过程中的知识网络,不仅节点数量和关联关系的数量与日俱增,更重要的是知识关联频度也在潜移默化地发生着改变。由于日益强大的网络分析理论与技术为知识网络结构关系的呈现与揭示提供了有力的支撑,知识关联关系在时间序列上的涌现问题也成为图书情报学界关注的一个焦点。

有鉴于此,本研究从知识关联关系的频度出发,在对领域知识网络拓扑结构的时间序列特征进行识别的基础上,从关联频度分布、关联频度增长、频度占比变化、邻近窗口状态等方面沿时间轴进行跟踪与分析。通过对知识关联关系在网络全局范围的分析和局部关系频度的剖析,尝试从关联频度演化的视角对领域知识关联关系涌现的外显模式与内部机制进行探索和揭示。

1 研究综述

以网络思维对学科知识的结构与发展问题展开研究,是近年来图书情报学发展的重要方向之一。早期的启蒙阶段可以追溯到 20 世纪 50、60 年代 Garfield 和 Price 分别在《科学》(Science)杂志上发表的借助引文网络对以科学论文为载体的学科知识传承问题的研究成果^[1-2]。知识网络研究真正进入蓬勃发展时期是以 20 世

纪末网络科学(Network Science)的复兴为标志的^[3]。1998 年,Watts 和 Strogatz 在《自然》(Nature)杂志上发表的论文,从小世界网络动力学的视角阐明了网络总体上的重大变化源自于局部微小的变动^[4]。同年,Albert 和 Barabási 等人也在《自然》(Nature)杂志撰文,专门论述了庞大的 WWW 网络相对较短的直径问题^[5]。1999 年,Barabási 和 Albert 又在《科学》(Science)杂志上发表文章,基于网络节点度值的幂律(Power-Law)分布,确立了多种真实网络的无标度(Scale-Free)属性^[6]。这些真实网络中,既包括 WWW 网络等一般意义上的信息网络,也包括图书情报学中的引文网络。基于这些研究成果,关注信息传播与交流的信息网络在信息跟踪^[7]、信息分类^[8]等方面都取得了研究进展。

随着一系列网络科学经典之作的问世以及应用于图书情报学研究领域,图书情报学界关于知识网络的相关研究也取得了丰富的成果。这期间既有传统引文知识网络在网络科学理论支持下的重新诠释,也有关键词知识网络、标签知识网络、作者合作知识网络等领域知识网络的崛起。Johnson 及其合作者将拥有较高被引率的作者提取出来并基于多维尺度分析(MDS)生成体现引用关系的知识网络,重新对引文知识网络进行了解读^[9]。Ronda-Pupo 等人将网络中心度分析与核心—边缘分析相结合应用于关键词知识网络,分析了核心概念成为学科支柱的途径^[10]。Weng 和 Menczer 基于网络熵值的测算,对标签知识网络中的主题聚类进行了研究^[11]。Henneberg 等人将网络分析中的结构洞(Structural Holes)理论应用于作者合作知识网络的作者圈、学术派系分析,探究了合作知识网络背后的底层结构^[12]。随着研究工作的深入,单纯的静态网络分析已经不能满足研究工作的需要,研究者们纷纷从知识网络动态演化发展的角度展开研究。Toral 及其合作者将因素分析与知识网络分析相结合,基于知识网络指标特征,对科研合作关系在时间序列上的演化进行跟踪,将研究工作由静态分析发展到动态分

析^[13]。Krafft 等人基于专利文献中的知识内部结构生成技术节点及其相互关系构成的专利知识网络,采用网络密度和多种中心度指标的测算,挖掘知识库结构与时间维度演进的模式^[14]。

与此同时,国内学术界也涌现出一批基于网络科学思维的图书情报学研究成果。邱均平等学者借助网络分析的方法,从引文知识网络、关键词知识网络等方面对学科知识进行可视化分析^[15]。易明等采用网络密度分析、中心性分析、凝聚子群分析、结构等价分析的方法,对标签知识网络及其个性化信息服务进行研究^[16]。在知识网络动态研究方面,马费成等基于引用关系的发展演化过程建立了引文知识网络的时序演化模型^[17]。赵蓉英等基于文献的时空分布侦测知识网络中的词频变动,以此挖掘领域研究热点与前沿^[18]。滕广青利用 k-丛和派系识别知识群落,从时间序列知识网络演化过程中,以多维交叉复现的视角对知识群落的生长展开研究^[19]。吴江等人基于学科知识网络,探测网络的时序演变过程以及不同学科层次内知识流动路径^[20]。大量网络科学理论与方法的引入,极大地丰富了图书情报学界对知识网络分析的手段与途径。

综上所述,网络科学在知识结构特征与知识生长机理的解析方面的优势已经被图书情报学界普遍认可和广泛应用。研究工作从较早的静态分析逐渐转向动态研究,其中以动态视角针对知识网络的研究,更有助于揭示出知识生长发展现象背后的规律。因此本研究借助网络科学的思维,从知识关联频度的视角对领域知识关联关系的生长涌现沿时间序列进行跟踪与分析,以期从中识别和揭示知识关联关系涌现的模式与机制。

2 理论框架

无论是传统的分类法、叙词表,还是语义网环境下的领域本体,都不是知识概念的简单罗

列或机械堆积,其中最为重要的是知识概念之间的逻辑关系。Barabási 在《科学》(Science)杂志上对网络科学的十年回顾与展望中也明确指出,网络节点之间的连接关系才是至关重要的^[21]。因此,知识关联关系也理所当然地成为图书情报学界重要的研究课题。与一般意义上的信息网络不同,图书情报学中的知识网络专注于知识之间的逻辑关系,而一般意义上的信息网络则关注信息传播交流的渠道。在图书情报学界,知识网络(Knowledge Network),是呈现知识之间逻辑关系的网络,是基于知识关联关系架构起网络的骨架。从 Price 在其引文分析的论文题目中使用“Network”一词开始^[2],知识网络先后发展出引文知识网络、专利知识网络、关键词知识网络、标签知识网络等。目前这些类型的知识网络已经得到图书情报学界的普遍认可,并广泛应用于核心知识识别、知识耦合分析、知识发展预测等诸多研究领域。

知识关联关系,抽象地说是知识概念之间的逻辑关系。早期的知识关联关系主要以学科体系及其分支为逻辑框架,包括主题词表、叙词表等。主要依靠领域专家的主观判断,具有鲜明的他组织性。在知识自组织思想的影响下,引用关系、同现关系(关键词、标签等)等陆续成为更真实灵活的知识关联关系,表现出强大的自组织特性。另一方面,知识之间的关联关系也不可避免地存在强弱之分。强关联关系不但在显著性、权威性等方面优于弱关联关系,而且还反映知识关联的深度^[22]。在知识网络中,这种关联关系的强弱差异可以通过知识节点之间的关联频度加以体现。知识关联频度,是反映知识之间关联关系强弱程度的指标,是根据引用频次、同现频次等建立和测算的。图书情报学中,高频度的知识关联关系往往代表学术界的普遍认可和接受,低频度的知识关联关系则代表尚未被普遍认可或有待认可的知识关联关系。本研究正是基于对知识关联关系的这一认识,从关联频度的层面对知识关联关系涌现展开研究。

学术界以往对领域知识的识别与分析主要基于知识概念自身的频次,提取领域知识的高频关键词就是这一方法的具体体现。在网络科学理论与技术被引入图书情报学研究之后,对网络中知识节点的度值的测度便成为图书情报学界判识领域中重要知识的主要手段。对知识网络 hub 节点的提取与分析则是这一方法的具体应用。本研究对领域知识关联关系分析所采用的主要是关联关系频度这一指标,该指标与节点出现频次和节点度值皆不相同。节点出现频次针对的是单一节点,节点度值虽然涉及节点之间的关联关系,但却仍然是从单一节点的角度进行考查。关联关系频度则始终是针对两个节点之间关联频度(强弱)的考查。从节点出现频次、节点度值、关联频度三者相互比较的视角,对知识节点之间关联频度的解析如图 1 所示。

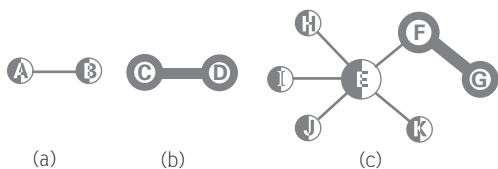


图 1 关联关系频度解析

图 1 中,节点的大小反映该节点所代表的知识概念出现的频次以及节点度值的大小,用灰色代表出现频次,用白色代表节点度值。出现频次越高则节点(灰色区域)越大,度值越高则节点(白色区域)越大。连线的粗细反映两个节点之间关联关系频度的大小,关联频度越高则连线越粗。图 1(a)反映两个知识节点之间关联频度为 1 的状态。节点 A 与节点 B 的出现频次都为 1(只出现一次),而且节点 A 与节点 B 的度值也都为 1(只有 1 个节点与之关联),因此(a)中两个节点出现频次一致,度值一致,故而节点大小一致,黑白两色区域大小一致。(b)反映两个知识节点之间关联频度为 5 的状态,节点间连线更粗。由于节点 C 与节点 D 之间的关联频度为 5,所以两个节点出现频次也都是 5。

代表度值的白色区域比代表出现频次的灰色区域要小,是因为两个节点此时的度值都为 1。(c)中节点 F 与节点 G 之间的关联频度为 5,节点间连线粗壮。节点 F 的灰色区域和白色区域比节点 G 的相应区域略大,因为节点 G 的出现频次为 5 且度值为 1,节点 F 的出现频次为 6 且度值为 2。节点 E 的出现频次为 5 且度值为 5,与其他相连节点的关联频度为 1。节点 H、I、J、K 则都是出现频次为 1 且度值为 1,与节点 E 的关联频度为 1。图 1 的关联频度解析说明,较高的关联频度在理论上并不能保证两个节点具有较高的度值,但是高关联频度却在保障关联关系高频显著性的同时,对两个节点较高的出现频次也做出了一定的保障。

另一方面,Newman 在 2010 年就已经指出,随时间序列变化的网络在未来是一个值得关注的领域^[23]。因此,知识网络在时间序列上的动态演化值得学术界深入研究。随着领域知识的不断积累和发展,总是有一部分在学科领域中处于重要地位的知识关联关系从众多的知识关联中脱颖而出。这些不同于其他知识关联的重要关联关系的产生与凸显,形成了领域知识关联关系的涌现。正如 Lewis 所指出的,涌现是一个动态的过程,经过若干时间步骤的微规则应用,最终导致网络的显著改变^[24]。因此,基于频度演化的领域知识关联关系涌现研究,在以关联频度为衡量标尺的基础上,还需要构建时变知识网络进而展开时间序列研究。研究工作以知识网络的结构属性为基础,但关注的重点并不局限于网络的拓扑结构,还包括知识关联关系在统计学上的显著性,这也是本研究与以往研究的显著不同之处。也正是基于这一思想,获得学术界普遍认可的高频度的知识关联关系与偶然产生的低频度的知识关联关系,在针对领域知识网络的分析过程中得以区分。

本研究将知识关联频度和时间序列分析相结合,在对各个时间窗口所对应的领域知识网络的拓扑结构进行初步识别的基础上,重点对知识网络中领域知识关联关系的频度分布与频

度生长进行跟踪和分析。基于关联频度的演化揭示领域知识生长过程中知识关联关系涌现的模式与机制。

3 研究方法

3.1 研究数据

本研究以 Bibsonomy 文献共享平台作为基础数据源,该平台采用社会化标注系统组织网站知识资源。研究中采用自主研发的网络爬虫工具,以“ontology”为标签组概念进行数据抓取。由此获得的数据为“ontology”主题领域相关数据,因此基于该数据的知识关联关系研究即为“ontology”主题领域的知识关联关系研究。研究中,共获得本体领域的相关文献 6 992 篇,相关标签 5 139 个,时间跨度为 2005—2015 年。Bibsonomy 平台系统的标签组概念代表一个由众多相关子标签组成的标签组群,因此以“ontology”作为标签组概念抓取的数据中也包含“owl”“protégé”“rdfs”等子标签及其与文献的标注关系。以自然年份为时间刻度,将时间区间划分为 $t_0 \sim t_{10}$ 共 11 个时间窗口,所获得的文献和标签在各个时间窗口中的时间序列分布如表 1 所示。

表 1 文献与标签数量的时间序列分布

时间	文献数量(累计值)	标签数量(累计值)
t_0	29	54
t_1	655	538
t_2	1 641	1 473
t_3	2 795	2 220
t_4	3 975	3 153
t_5	4 615	3 690
t_6	5 290	4 170
t_7	6 104	4 453
t_8	6 382	4 737
t_9	6 609	4 996
t_{10}	6 992	5 139

考虑到研究工作的焦点集中于领域知识发展过程中知识关联关系的涌现研究,当前时间窗口的状态都是基于上一个时间窗口的状态发展演变而来的。因此表 1 中的数据分别为该时间窗口下的文献数量和标签数量各自的累计值,即当前时间窗口的标签包含前一时间窗口的标签(当前时间窗口的文献也包含前一时间窗口的文献),或者说当前时间窗口的数据是在前一时间窗口数据的基础上发展而来的。尽管相对于当期发生值而言,基于累计值的测度会无形中增加分析过程中的计算量,但是采用累计值对领域知识发展进行跟踪与分析更符合领域知识演化变迁中知识生长进化的理论^[25]。

3.2 研究流程与方法

3.2.1 知识网络构建及其基础属性

在已获得的原始数据集的基础上,以标签为网络节点,以标签同现关系为网络连线,构建 $t_0 \sim t_{10}$ 时间序列领域知识网络。其中,如果两个标签被用于标注同一篇文献,则这种标签同现关系使得两个标签在矩阵中具备邻接关系,标签所反映的知识单元之间也相应地存在关联关系,知识网络中在代表两个标签的知识节点之间以一条连线描述这种知识关联关系,由此构建的知识网络为无向网络(同现关系不分方向)。由于表 1 中数据取值为累计值,因此每一个时间窗口的领域知识网络不但包含前一时间窗口领域知识网络的所有节点和连线,而且在前一时间窗口领域知识网络的基础上增加了新的知识节点和网络连线以及连线的频度。因此研究中构建的领域知识网络也是多值网络。结合表 1 中标签数量的相应数据可知,知识网络的规模随时间轴的延展持续递增,因此研究中构建的时间序列的领域知识网络首先是一个在规模上处于生长过程中的网络。

针对较大规模知识网络的分析工作的有效开展,需要对上述基于真实数据构建的时间序列领域知识网络的属性特征进行初步描述。首先,采用 Latapy 大规模网络聚类的计算方法^[26],

分别计算 $t_0 \sim t_{10}$ 全部时间窗口领域知识网络的聚类系数;其次,借助 Brandes^[27] 的算法分别计算 $t_0 \sim t_{10}$ 全部时间窗口领域知识网络的特征路径长度。计算结果显示,研究中构建的领域知识网络在 $t_0 \sim t_{10}$ 全部时间区间内具有稳定的高聚类系数($C \geq 0.83$)和较短的特征路径长度($L \leq 2.49$)。显然,这一结果符合 Watts 等人发表于《自然》(Nature)杂志的对网络中小世界效应的判定标准^[4]。因此,基于真实数据构建的领域知识网络不但在拓扑结构方面具有小世界效

应,而且在知识网络生长的时间序列过程中始终保持稳定的小世界网络的特征。

Barabási 和 Albert 于 1999 年在《科学》(Science)上发表论文提出,大多数真实网络拓扑结构的度序列分布符合幂律(power-law)分布^[6],此处进一步对 $t_0 \sim t_{10}$ 时间窗口的领域知识网络的度序列分布加以考查。以节点度值 D 为横轴,以度分布 P^D 为纵轴,采用双对数直角坐标系,得到时间序列领域知识网络度序列分布如图 2 所示。

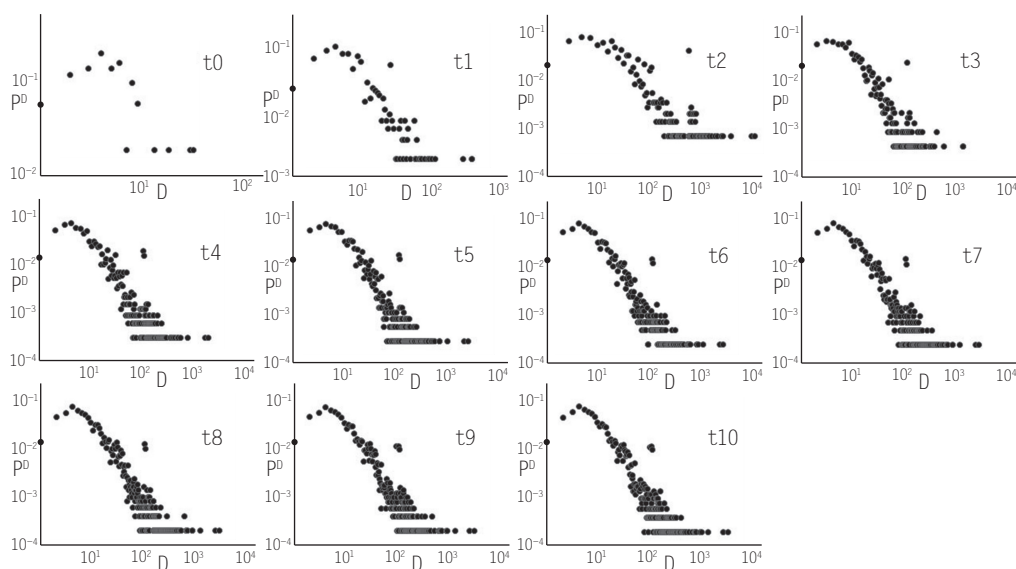


图 2 时间序列知识网络度序列分布

图 2 中,领域知识网络初态(t_0 时间窗口)的度序列分布显然没有较好地遵循幂律分布,其 R^2 值仅为 0.443 6。 $t_1 \sim t_5$ 时间窗口的 R^2 值依次为 0.694 8、0.707 2、0.728 9、0.747 6、0.751 4,幂律拟合程度均持续梯次提高。在这之后 $t_6 \sim t_{10}$ 时间窗口的 R^2 值保持在 0.765 之上,度序列分布较好地拟合幂律分布。尽管 Barabási 等人基于相对成熟的真实网络得出现实网络的无标度属性,但是从图 2 中度序列分布的时间序列来看,领域知识网络的初态并非无标度网络,而是随着领域知识的发展,知识网络的结构逐渐趋近于无标度网络。因此,从知识网络度序列

分布的角度来说,时间序列的领域知识网络是一个标度涌现网络。这一事实在确定领域知识网络符合 Barabási 等人所界定的生长性网络的同时,为后续基于频度的领域知识关联关系涌现研究,奠定了知识网络在拓扑结构方面的基础。此外,在图 2 中还可以发现,即使处于幂律拟合较好的时间轴后半段,度序列分布也存在一个 Newman 曾经指出的呈弯曲状的“头重分布”(Top-heavy Distribution)^[23]。正是基于“头重分布”,Newman 做出了真实网络在高度值区的尾部服从幂律分布的总结。综合这些情况,研究中使用的基于真实数据构建的时间序列领

域知识网络的属性特征是,在整个时间区间内具有较为恒定的小世界网络属性和随时间序列逐渐涌现的无标度网络属性。对于较短的特征路径长度、度序列幂律分布、较高的聚类系数三个特征在同一个网络出现的现象,Newman 指出,到目前为止还没有具有说服力的模型来解释这三个特征如何同时出现在一个网络中,这是网络科学领域有趣的开放性问题^[28]。

3.2.2 知识关联频度的提取

通过对所构建的领域知识网络的描述统计,从网络拓扑结构的视角确定了时间序列领域知识网络的属性特征。但是知识网络在拓扑结构方面的属性特征仅仅是知识关联分析的基础,特征路径、聚类系数、度序列分布主要是从知识节点的视角对知识关联进行的考查,关注的重点在于知识关联关系的结构属性。因此,研究中还需要专门针对知识关联及其频度做出具体的测度与分析。基于累计值构建的领域知识网络的成长过程中,不但节点数量增长、关联关系增长,关联的频度也在增长。即使在同一个知识网络中,知识关联关系的频度也存在较大的强弱差异。有的知识关联仅仅基于一次同现关系,此类知识关联往往尚未取得学术界的共识;有的知识关联则基于成百上千次同现关系,显然此类知识关联已经得到学术界的普遍认可。进一步在知识网络拓扑结构的基础上,专门对网络中每一对知识关联关系及其关联频度进行提取。相关数据的统计结果如表2所示。

表2中的关联关系数量显示,从知识网络初态(t0时间窗口)的162对关联关系,到知识网络终态(t10时间窗口)的71 467对关联关系,知识网络关联关系数量在时间序列上保持高速增长。相对于同期时间窗口的知识网络的节点数量(参见表1)而言,这种高速增长的趋势更加明显。另一方面,随着时间轴的延展,知识网络关联关系在数量上急剧增长的同时,关联频度的数量和最大值也在不断升高。同时,各个时间窗口知识网络关联频度的均值却相对平稳。如果说网络中知识节点的数量与关联关系数量在

表2 知识关联关系与关联频度

时间	关联关系数量	关联频度数量	关联频度最大值	关联频度均值
t0	162	181	7	1.117 3
t1	3 474	9 376	199	2.698 9
t2	16 395	28 535	287	1.740 5
t3	24 560	43 219	368	1.759 7
t4	40 077	67 021	475	1.672 3
t5	45 048	75 937	548	1.685 7
t6	51 435	88 668	569	1.723 9
t7	56 908	121 952	976	2.143 0
t8	62 222	132 510	1 055	2.129 6
t9	68 791	145 634	1 115	2.117 1
t10	71 467	162 811	1 245	2.278 1

时间序列上的增长符合网络规模层面的生长性,那么表2中关联频度数量在时间序列上的增长则说明了领域知识网络在关联关系数量与关联频度方面的生长性。

即使一对关联关系在不同时间窗口所连接的两个知识节点没有变化,关联频度在不同时间窗口之间的差异(频度增长)也会使得这一关联关系的强弱程度在不同时间窗口之间有所不同。表2中全时间区间内,领域知识网络的关联关系数量合计44万余对,关联频度数量合计87万余次。与Barabási等人从网络拓扑结构的视角进行的网络标度涌现研究不同,本研究的核心焦点是基于关联频度对知识关联涌现展开研究。因此,44万余对的关联关系与87万余次的关联频度如何分布、如何发展是本研究重点关注的对象。研究中,将分别从知识网络全局和知识关联个体两个层面对领域知识的关联关系展开时间序列的跟踪与分析,探索其中潜在的模式与机制。

4 研究结果

4.1 网络全局知识关联关系分析

研究中基于真实数据的时间序列领域知识

网络在节点数量、关联关系数量、关联频度三方面共同的生长性,使得领域知识网络中关联关系的生长不同于 Watts 等人基于概率 P 的网络重连,也不同于 Balabasi 等人以特定节点数量和单一频次关联关系模拟的网络生长。为了更好地从领域知识网络全局角度观察和跟踪知识关

联频度在时间序列上的变化情况,对时间序列领域知识网络的知识关联频度分布进行分析。以知识关联频度 F 为横轴,以关联频度分布 P^F 为纵轴,采用双对数坐标系,得到时间序列领域知识网络的关联频度分布如图 3 所示。

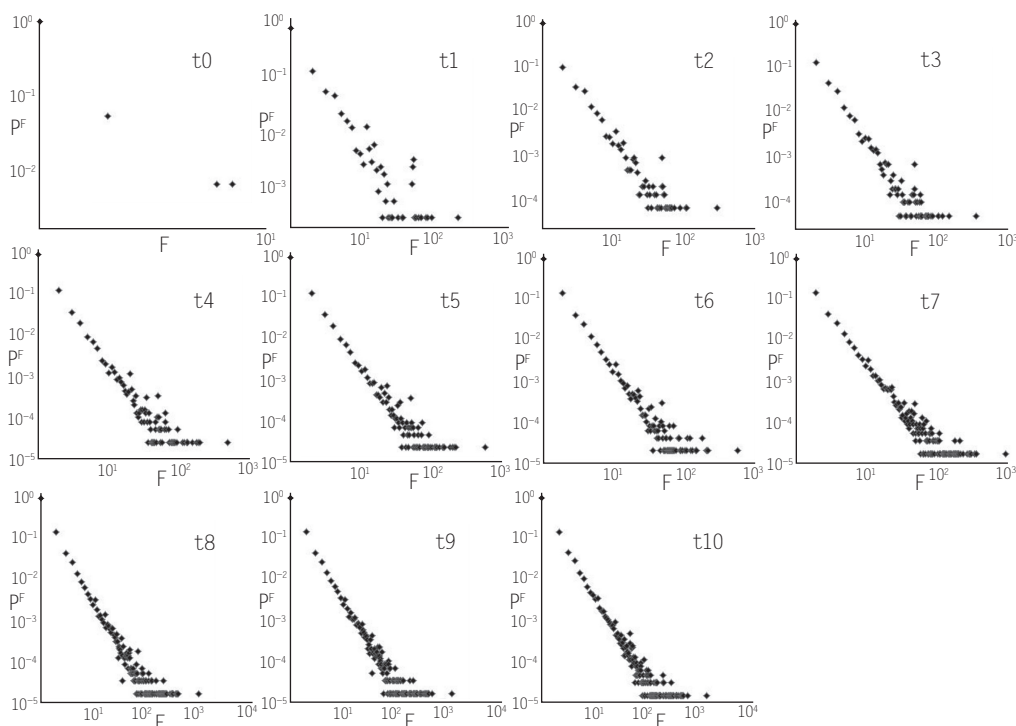


图 3 时间序列知识网络关联频度分布

图 3 中,知识网络的初态(t_0 时间窗口)的关联频度分布尽管在数值计算上已经遵循幂律分布,但由于网络初态数据过于稀疏,因此并不能很好地说明问题。在其后的时间窗口中,关联频度分布的幂律分布趋势逐渐明显。 t_1 时间窗口对应的 R^2 值为 0.776 6。从 t_2 时间窗口开始 R^2 值始终都保持在 0.851 8 之上。显然,与图 2 中的度序列分布相比,同一时刻领域知识网络的关联频度分布比知识网络的度序列分布更加符合幂律分布。经过计算得出,关联频度分布遵循如下关系公式:

$$P^F \sim F^{-\gamma_{frequency}} \quad (\text{公式 1})$$

公式 1 中, P^F 为关联频度分布, F 为知识关联频度, 幂指数 $\gamma_{frequency} = 2.05 \pm 0.46$ 。这一现象表明,在全时间区间内的领域知识网络中,少量的知识关联拥有很高的关联频度,大量的知识关联只具有很低的关联频度。即,领域知识当中只有少量的知识关联得到学术界的普遍认可。在关联频度分布的幂律拟合程度高于度序列分布的幂律拟合程度的同时,关联频度分布的幂律拟合程度也在时间序列上一定程度地存在初态网络拟合略好、终态网络拟合很好的局面。显然,知识关联频度的生长过程也是一个频度涌现的过程。

此外,从图3中还可以发现,领域知识网络的关联频度分布并没有出现度序列分布中 Newman 所说的真实网络中呈弯曲状的“头重分布”(参见图2)。从这个现象上来看,领域知识网络的关联频度分布并非像度序列分布一样只是在长尾部分符合幂律分布,而是在全部关联关系

的范围内更高程度地符合幂律分布。

进一步以时间序列为横轴,以关联关系数量和关联频度数量为纵轴,并且纵轴采用对数坐标轴构建直角坐标系。得到领域知识网络关联关系数量和关联频度数量在时间轴上的折线图,如图4所示。

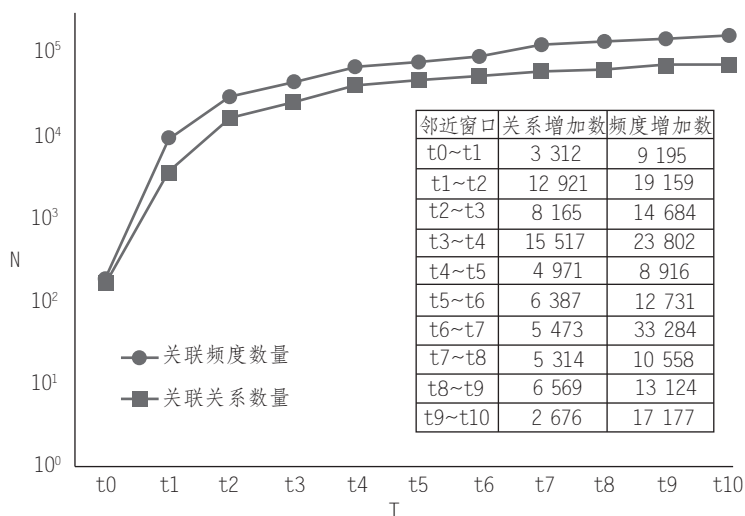


图4 关联关系数量与关联频度数量

尽管从此前表2中的数据已经可以发现,同一时间窗口中关联频度数量远大于关联关系数量,但是图4给出了更明确的展示。除去开始时刻二者相差无几外(频度数量略高于关系数量),后续的时间窗口中关联频度数量始终处于关联关系数量之上。并且图4中采用对数坐标轴的纵轴更加说明,二者在数量上的差距随时间轴的延展越来越大。另一方面,从图4中关联关系与关联频度的增加数可以发现,相邻时间窗口知识网络新增的关联频度并没有完全等价于新增的关联关系。这一发现初步说明,不但在原有网络中关联频度分布不均衡,可能在新增的关联关系中,关联频度分布也不均衡;甚至其中一部分新增的关联频度可能附加在一些原有的关联关系之上,加强了这些原有的知识关联关系。

至此,网络全局范围的知识关联分析表明,

领域知识网络中知识关联频度的分布符合幂律分布,领域中只是极少数的知识关联关系拥有很高的频度,大多数知识关联关系仅具有很低的频度。而且关联频度分布在幂律拟合方面比知识网络的度序列分布表现的更好,不但具有更高的 R^2 值,而且没有出现真实网络度序列分布中常见的“头重分布”现象。关联频度的幂律分布意味着知识关联关系的生长是超越了随机性演化的涌现过程。通过对全网络范围关联数量与关联频度增加情况的考查还发现,新增加的关联频度不但可能附加于新增的关联关系,还可能被用于强化原有的关联关系,并且可能是不均衡的。这一发现为领域知识发展过程中知识关联频度“富者更富”做了铺垫。当然,网络全局范围的考查尚不足以揭示更多的细节,对此本研究将进一步选取部分局部的知识关联关系从个体层面予以分析。

4.2 局部知识关联关系分析

知识关联关系的基本特征在网络全局层面上表现得较为明确,但在一些细节上尚不明晰。网络全局范围的考查虽然在时间序列上对生长过程中的知识网络关联关系做出了总体模式的研判,但是若要探究其中的规则细节则有必要从具体的个体关联关系着手进行细节分析。以往针对网络结构关系的研究工作往往提取其中

的高 Hub 节点给予特别关注,本研究则对高频度的知识关联关系给予重点分析。从频度分布的角度来讲,这部分研究就是专门针对频度“富者”是否“更富”的问题进行细节层面的跟踪分析。以关联频度为衡量标尺,提取每个时间窗口所对应的领域知识网络中关联频度排名前十位的关联关系。这些关联关系的频度值如表 3 所示。

表 3 频度排名前十位的关联关系频度值

排名 时间	1	2	3	4	5	6	7	8	9	10
t0	7	6	2	2	2	2	2	2	2	2
t1	199	86	76	71	62	61	57	53	51	50
t2	287	108	105	88	72	67	65	63	62	61
t3	368	151	122	108	107	88	84	79	78	76
t4	475	193	182	170	152	130	129	126	116	108
t5	548	212	200	191	178	146	142	139	124	114
t6	569	220	214	205	183	183	149	147	133	125
t7	976	374	372	370	349	314	305	298	297	295
t8	1 055	416	401	397	380	341	328	326	322	321
t9	1 115	474	422	420	401	385	375	362	354	346
t10	1 245	552	453	442	428	425	422	411	404	363

表 3 中,各个时间窗口排名第一的知识关联关系的频度值所对应的就是表 2 中的关联频度最大值。在排名前十位的关联关系内部,排名第一的关联关系的频度值与排名第二、排名第十的关联关系的频度值之间的差异,随着领域知识的生长越来越大。这充分显示出关联频度的“富者更富”属性。此外,将表 3 中的频度值与表 2 中的频度均值对比可以发现,不但排名第一的关联关系频度值与当期知识网络关联关系的频度均值有较大的差异,即使是排名第十的关联关系的频度值也与频度均值存在差异。而且随着领域知识网络的生长,这种差异同样逐渐拉大。从这个意义上讲,关联频度“富

者更富”的属性在知识关联关系的生长过程中是十分鲜明的。

另一方面,结合表 2 中各个时间窗口的关联关系数量,计算获得频度前十位的关联关系在当期总关联关系数量中的占比。同理,对表 3 中频度前十位的关联关系的频度计算合计值,结合表 2 中各个时间窗口的关联频度数量,计算频度前十位的关联关系的频度在当期总关联频度数量中的占比。以时间序列为横轴,以占比量为纵轴,纵轴采用对数坐标轴,由此获得频度前十位的关联数量与关联频度的比重在时间轴上的折线图如图 5 所示。

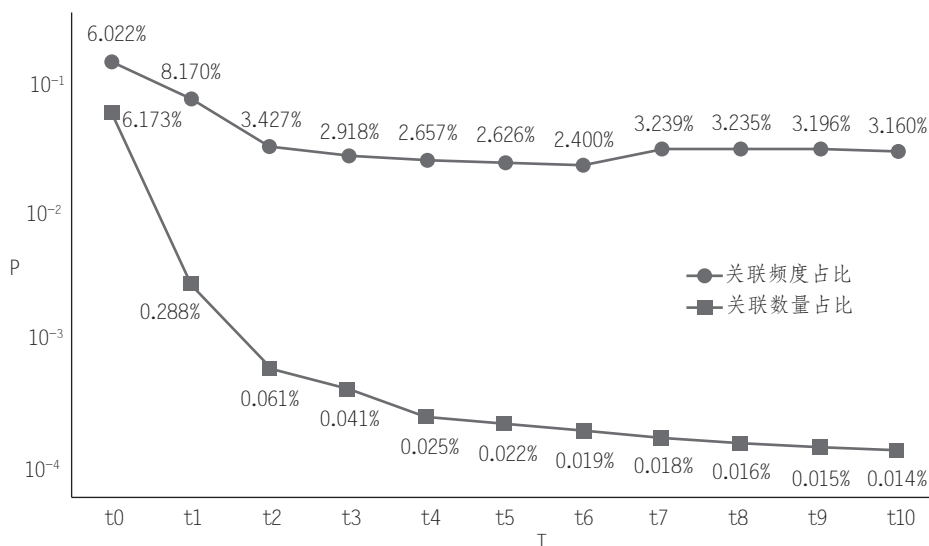


图5 频度排名前十位的关联数量与频度数量比重

图5中,频度前十位的关联关系在网络全部关联关系数量中的占比持续下降,这是领域知识网络关联关系数量随着时间轴延展持续增加的必然结果。另一方面,频度前十位的关联关系的频度在网络全部关联频度数量中的占比,除了开始阶段有所下降外,在更多的时刻却相对平稳,并始终高于同一时刻的关联关系数量占比。从图5中各个时间窗口的数据标识来看,频度前十位的关联关系,以较低的数量占比拥有了相对较高的频度占比。因此,频度前十位的关联关系在关联频度层面属于“富者”。与此同时,随着时间轴的延展,图5中数量占比与频度占比的相对差距越来越大。这种不断扩大的相对差距表明,关联频度层面的“富者”随着知识网络的生长变得相对“更富”。如,t3时间窗口中0.041%的关联关系拥有2.918%的关联频度,t10时间窗口中0.014%的关联关系更是占据了3.160%的关联频度。从相对平稳且较高的频度占比与不断增大的频度总值来看,越来越少的关联关系(图5中逐渐走低的数量占比)拥有更高的关联频度。至此可以判断出,领域知识网络关联关系的生长规律确实存在关联频度“富者更富”的属性。这一结果与

Barabási 和 Albert 在实验室条件下的标度涌现模拟仿真中得出的网络规模生长中节点度值的“富者更富”属性如出一辙。

为了对知识关联关系生长过程中关联频度“富者更富”属性背后的机制做出更细致准确的分析,进一步对各个时间窗口频度排名前十位的关联关系在邻近时间窗口的表现进行考查。将知识关联关系在前邻时间窗口中的状态分为“极富”“较富”“较贫”“极贫”。如果某关联关系的频度处于极少数的前十位的位置即是频度“极富”状态,如果关联频度小于排名第十位的关联关系频度但是大于频度均值(参见表2)则处于频度“较富”状态,如果关联频度小于等于频度均值则处于频度“较贫”状态。显然,由于此前的分析已经证实知识关联关系的频度分布符合幂律分布,因此处于频度“较贫”状态的关联关系数量要多于处于频度“较富”状态的关联关系数量,处于频度“较富”状态的关联关系数量也远多于处于频度“极富”状态的关联关系数量。“极贫”状态则是指在前邻时间窗口知识网络中尚不存在的关联关系。由此获得关联频度排名前十位的关联关系在前一个邻近时间窗口的状态如表4所示。

表 4 频度前十位关联关系在前邻窗口中的状态

排名 时间	1	2	3	4	5	6	7	8	9	10
t1	⊙	○	⊙	⊙	○	○	○	○	○	●
t2	⊙	⊙	⊙	⊙	○	⊙	⊙	⊙	⊙	⊙
t3	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
t4	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
t5	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
t6	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
t7	⊙	○	⊙	⊙	○	⊙	○	○	⊙	⊙
t8	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
t9	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙
t10	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙	⊙

* 注: ⊙极富, ⊙较富, ●较贫, ○极贫。

总体上看,表 4 中在 t1~t10 各个时间窗口中处于频度“极富”状态的、排名前十位的知识关联关系在前一时间窗口大多处于同样的频度“极富”状态。后一个时间窗口中的频度“极富”者往往也是前一个时间窗口中的频度“极富”者。从关联关系个体的层面再次解释了知识关联关系在频度方面的“富者更富”属性。

在领域知识生长的幼生期(时间序列前端),由于知识领域的内容与范围还处在不断充实和丰富阶段,加之此时数据量较小,因此 t1 时间窗口中除了少部分频度“极富”的关联关系来自于 t0 时间窗口中频度“极富”的关联关系外,大部分频度“极富”的关联关系在前一时间窗口(t0 时间窗口)中属于频度“较贫”甚至“极贫”的状态。随着领域知识的逐渐发展成熟,t2 时间窗口中更多的“极富”关联关系来自于前一时间窗口(t1 时间窗口)中频度“极富”或“较富”的关联关系。发展到 t3 时间窗口,“极富”状态的关联关系已经全部来自于前一时间窗口的频

度“极富”和“较富”的关联关系。t4 时间窗口“极富”状态的关联关系则更多地来自于前一时间窗口的“极富”关联关系。这一态势在 t5 时间窗口进一步延续。到了 t6 时间窗口甚至发展为所有频度“极富”的关联关系全部来自于前一时间窗口的频度“极富”关联关系。这一过程可以总结为,前一时间窗口中频度“极富”的知识关联关系有很高的概率成为后一时间窗口中频度“极富”的知识关联关系。通过更强劲地吸纳新增的关联频度,高频度的知识关联优先涌现,并且在领域知识度过幼生期之后表现得逐渐明显。如果说 Barabási 和 Albert 发现了网络规模生长中节点度值的“择优连接”(Preferential Connectivity)机制,那么表 3 则演绎出了知识关联关系生长过程中关联频度“择优强化”的内部机制。

值得注意的是,表 4 中 t7 时间窗口再次出现类似于 t1 和 t2 时间窗口的现象,即前一时间窗口频度“极贫”的关联关系成为后一时间窗口

频度“极富”的关联关系的现象。这一情况显然与上文知识关联关系生长过程中关联频度的“择优强化”机制不符。研究中,对这些知识关联的具体内容加以考查发现,这部分知识关联关系都是与人工智能(AI)相关的知识关联。事实上,自从 Hinton 在《科学》(Science)杂志上提出了深层网络训练的解决方案^[29]之后,直到 2012 年(t7 时间窗口)Hinton 团队以卷积神经网络 AlexNet 在 ImageNet 图像识别比赛中碾压群雄,才向世人证明了深度学习的潜力并真正引起了学术界的普遍关注。也是从这一时刻(t7 时间窗口)开始,人工智能研究进入了爆发期,进而使得本体研究领域涌现出高频度的人工智能相关的知识关联。由此可见,知识关联关系生长过程中,除了关联频度的“择优强化”之外,还存在一种关联频度的“爆发强化”机制。当然,“爆发强化”并非是一种常见现象,而是需要在学术界产生了重大影响的发明或发现才能引发。在之后的时间窗口(t8~t10)中,已经进入“极富”状态的、与人工智能(AI)相关的知识关联关系继续高概率地成为下一时间窗口中频度“极富”的关联关系。高频的知识关联优先涌现,知识关联关系的生长机制再次体现为关联频度“择优强化”机制。

5 结论与讨论

本研究基于社会化标注知识组织模式中的标签同现关系构建时间序列的多值领域知识网络。在对领域知识网络拓扑结构时间序列特征进行归纳的基础上,首先从网络全局的范围对知识关联频度分布展开时间序列分析,并针对高频度的局部知识关联关系从频度演化的角度进行具体分析。综合网络全局范围与高频度的局部关联关系分析的结果,可以得出以下结论:

(1)知识关联关系生长过程中,关联关系频度的分布符合幂律分布。尽管知识网络节点度

序列的幂律分布已经在学术界成为不争的事实,但是本研究通过动态跟踪发现知识网络节点的度值分布在时间序列后期表现得更好。然而,更重要的是通过对知识关联关系频度的时间序列分析发现,同一个网络的知识关联关系频度分布比节点度值分布能够更好地拟合幂律分布(拥有更高的 R^2 值),而且没有出现节点度值分布中常见的“头重分布”现象(参见图 3)。知识关联频度的幂律分布同样也是在时间序列的后期表现得更好。这一研究发现意味着,尽管知识之间关联关系决定着知识网络的拓扑结构,但是只有少数关联关系拥有极高的频度值,大多数关联关系仅具有少量的频度值。从关联频度的层面来讲,少数关联关系占据了极高的频度“财富”成为“富者”,频度分布并不均匀。据此可以在领域知识生长的多个时间窗口根据关联频度设定阈值,提取具有统计学显著意义的知识关联关系和等效的层次知识网络^[30]。

(2)知识关联关系的生长过程具有频度层面的“富者更富”的属性。以往关于知识网络的研究中,即使涉及网络的生长性问题,所关注的焦点也往往集中在节点数量以及节点之间关联关系数量的生长。本研究则通过时间序列分析发现了知识关联关系在关联频度层面的生长性(如表 2 中的关联频度数量)。特别是知识关联关系的生长过程中表现出鲜明的频度层面的“富者更富”属性。“富者更富”是一个相对的状态也是一个动态的状态。尽管知识关联关系的频度不断生长积累,但是少数频度“富者”与大多数频度“贫穷者”相比,却占据着越来越多的频度“财富”。表 3 的数据和图 5 的折线走势都表明,知识关联关系在频度层面的贫富差距,随着领域知识的生长发展,在时间轴的后半段表现得愈发明显。知识领域中重要的知识关联凭借其在频度方面的优势,在知识发展的过程中进一步高比率地占有频度“财富”,从而也越鲜明地从众多关联关系中涌现和凸显

出来。

(3) 知识关联关系的频度“富者更富”属性主要遵循“择优强化”机制。领域中重要的知识关联凭借其强劲的频度优势将其他关联关系远远地抛在后面,从而在众多的关联关系中脱颖而出。研究中发现大多数在频度层面处于“极富”状态的知识关联关系,往往在前一时间窗口已经处于频度“极富”状态。显然,拥有更多频度“财富”的关联关系在领域知识的生长发展过程中会以更高的概率吸引获得新增频度的附加。也就是说,新增加的关联频度并不是均匀地附加在每一对关联关系之上,而是优先选择那些曾经的“富者”进行“择优强化”。如果说 Barabási 等人提出的“择优连接”的关注对象是关联关系的数量,那么本研究发现的频度“择优强化”则关注的是关联关系的质量。也正是“择优强化”微规则在时间序列上的反复叠加,造就了关联关系涌现中频度层面的“富者更富”现象。当然,研究也发现处于频度“极贫”状态的关联关系瞬间一跃进入频度“极富”

状态的现象。尽管这种频度“爆发强化”的现象极为罕见,但是通过对关联细节的挖掘发现,引起“爆发强化”的原因主要是学术界产生了影响重大的发现或发明。这也为基于知识关联频度识别和判识重大学术成果提供了一种可能。

本研究从关联频度的层面对知识关联关系的涌现过程进行动态分析,其中知识关联关系生长过程中的频度幂律分布、“富者更富”“择优强化”等结论不但有助于促进知识网络、知识生长等领域的研究工作,对于同样涉及关联频度差异的社交网络、传播网络、交通网络等研究也有裨益。当然,研究工作中也存在尚不完善之处;研究使用的基于社会化标注系统构建的知识网络,虽然其开放模式在知识时效性方面更具有优势,但是尚不足以囊括所有类型的知识网络。未来的研究中将对更多类型的知识网络展开更加细致的分析,从而更深层次地接近领域知识生长发展的真相,揭示知识生长的模式与规律。

参考文献

- [1] Garfield E. Citation indexes for science: a new dimension in documentation through association of ideas[J]. Science, 1955, 122(3159): 108-111.
- [2] Price D J de S. Networks of scientific papers[J]. Science, 1965, 149(3683): 510-515.
- [3] Barabási A-L. Network science[M]. Cambridge: Cambridge University Press, 2016: 20-41.
- [4] Watts D J, Strogatz S H. Collective dynamics of ‘small world’ networks[J]. Nature, 1998, 393(6684): 440-442.
- [5] Albert R, Jeong H, Barabási A-L. Internet: diameter of the world-wide web[J]. Nature, 1999, 401(6749): 130-131.
- [6] Barabási A-L, Albert R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.
- [7] Moon I C, Kim Y-M, Lee H-J, et al. Temporal issue trend identifications in blogs[C]//Proceedings of the 2009 International Conference on Computational Science and Engineering, 2009, 4: 619-626.
- [8] Wang C, Song Y, Li H, et al. Text classification with heterogeneous information network kernels[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016: 2130-2136.
- [9] Johnson B, Oppenheim C. How socially connected are citers to those that they cite?[J]. Journal of Documenta-

- tion, 2007, 63(5): 609-637.
- [10] Ronda-Pupo G A, Guerras-Martin L A. Dynamics of the evolution of the strategy concept 1962-2008: a co-word analysis[J]. Strategic Management Journal, 2012, 33(2): 162-188.
- [11] Weng L, Menczer F. Topicality and impact in social media: diverse messages, focused messengers[J]. PloS One, 2015, 10(2): e0118410.
- [12] Henneberg S C, Swart J, Naudé P, et al. Mobilizing ideas in knowledge networks: a social network analysis of the human resource management community 1990-2005[J]. The Learning Organization, 2009, 16(6): 443-459.
- [13] Toral S L, Bessis N, Martínez-Torres M R. External collaboration patterns of research institutions using shared publications in the web of science[J]. Program: Electronic Library and Information Systems, 2013, 47(2): 170-187.
- [14] 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.
- [15] 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, Lv Hong. Visualization analysis of research on knowledge network based on mapping knowledge domains [J]. Information Science, 2013, 31(12): 3-8.)
- [16] 易明, 王学东, 邓卫华. 基于社会网络分析的社会化标签网络分析与个性化信息服务研究[J]. 中国图书馆学报, 2010, 36(2): 107-114. (Yi Ming, Wang Xuedong, Deng Weihua. A research on the tag network analysis based on social network analysis (SNA) and the personalized information service [J]. Journal of Library Science in China, 2010, 36(2): 107-114.)
- [17] 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.)
- [18] 赵蓉英, 王静. 社会网络分析(SNA)研究热点与前沿的可视化分析[J]. 图书情报知识, 2011(1): 88-94. (Zhao Rongying, Wang Jing. Research of international social network analysis in frontier domains in visualized information[J]. Document, Information & Knowledge, 2011(1): 88-94.)
- [19] 滕长青. 关联驱动领域知识群落生长[J]. 中国图书馆学报, 2017, 43(3): 58-71. (Teng Guangqing. Correlation-driven domain knowledge community growth[J]. Journal of Library Science in China, 2017, 43(3): 58-71.)
- [20] 吴江, 金妙, 陈君. 基金视角下的学科知识流动网络构建与分析[J]. 图书情报工作, 2016, 60(8): 79-85. (Wu Jiang, Jin Miao, Chen Jun. Construction and analysis of interdisciplinary knowledge flow network based on co-occurrences of funding application codes[J]. Library and Information Service, 2016, 60(8): 79-85.)
- [21] Barabási A-L. Scale-free networks: a decade and beyond[J]. Science, 2009, 325(5939): 412-413.
- [22] 李纲, 纪曼, 陈璟浩, 等. 基于社会网络的知识共享研究综述[J]. 信息资源管理学报, 2013, 3(2): 50-58. (Li Gang, Ji Man, Chen Jinghao, et al. A review of knowledge sharing based on social network[J]. Journal of Information Resources Management, 2013, 3(2): 50-58.)
- [23] 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: 28-31, 157-168.

- (Newman M E J. Network: an introduction[M]. Guo Shize, Chen Zhe, trans. Beijing: Publishing House of Electronics Industry, 2014: 28-31, 157-168.)
- [24] 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, 142-171. (Lewis T G. Network science: theory and applications[M]. Chen Xiangyang, Ju Xiulian, et al, trans. Beijing: China Machine Press, 2011, 142-171.)
- [25] 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.)
- [26] Latapy M. Main-memory triangle computations for very large (sparse (power-law)) graphs[J]. Theoretical Computer Science, 2008, 407 (1-3): 458-473.
- [27] Brandes U. A faster algorithm for betweenness centrality[J]. Journal of Mathematical Sociology, 2001, 25(2): 163-177.
- [28] Newman M E J. The structure and function of networks[J]. Computer Physics Communications, 2002, 147(1-2): 40-45.
- [29] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507.
- [30] 滕广青, 白淑春, 韩尚轩, 等. 基于无标度与分形理论的层次知识网络原理解析[J]. 图书情报工作, 2017, 61(14): 132-140. (Teng Guangqing, Bai Shuchun, Han Shangxuan, et al. Analysis on the principle of knowledge network at level based on scale-free and fractal theory[J]. Library and Information Service, 2017, 61(14): 132-140.)

滕广青 东北师范大学信息科学与技术学院信息管理系教授, 博士生导师。吉林 长春 130117。

(收稿日期: 2017-11-23; 修回日期: 2018-01-29)