

知识网络演化模型研究述评*

张 斌 李亚婷

摘 要 知识网络演化模型可用于探讨和发现知识发展脉络的关键特征,相关研究主要集中在统计物理、计算机、图书情报等领域。本文对网络模型经典理论进行回顾,从引证网络和合作网络两个方面对具有代表性的演化模型进行梳理,介绍网络演化模型评价的一些进展,最后提供了一个相对清晰的研究框架。研究发现:针对这方面的研究近年来有成为图书情报学领域研究热点的趋势;已有引证网络和合作网络的研究大多利用 BA 模型及其改进模型来研究演化机理;未来的研究应结合网络的局部结构特征和节点的外部属性信息来建立混合择优模型;众多知识网络演化模型在解释真实网络演化时,依然存在着不小差距,利用链路预测的原理来设计和评价演化模型有望成为一个新的研究方向。图 2。参考文献 79。

关键词 知识网络 演化模型 引证网络 合作网络 链路预测

分类号 G302

A Review of the Evolution Model of Scientific Knowledge Network

ZHANG Bin & LI Yating

ABSTRACT

The evolution model of knowledge network is an abstract expression of knowledge network's inherent interaction mode and process, a method to explore the key features of knowledge development and a bond to connect the external structure with internal interaction mechanism. Currently, researches on this issue are scattered in the fields of Statistical Physics, Computer Science, and Library & Information Science. It has become a research hotspot in Library & Information Science in recent years. Knowledge network is the product of knowledge exchange, and both classic citation network and co-authorship network are dominant representations for it. Many scholars studied citation network and co-authorship network to explore the interaction among scholars, collaboration evolution mechanism, knowledge dissemination and development from the perspective of network evolution.

Library & Information Science has long been concerned about the evolution model of knowledge network. Early Price Model laid the ideological foundation for BA scale-free network discovery. As knowledge

* 本文系国家自然科学基金面上项目“知识网络的形成机制及演化规律研究”(编号:71173249)和国家自然科学基金重点国际(地区)合作研究项目“大数据环境下的知识组织与服务创新研究”(编号:71420107026)的研究成果之一。(This article is an outcome of the general project “Research on Formation Mechanism and Evolution Laws of Knowledge Networks” (No. 71173249) and the major international (regional) joint research project “Research on Knowledge Organization and Service Innovation in the Big Data Environments” (No. 71420107026) supported by National Natural Science Foundation of China.)

通信作者:张斌, Email: zb0205@126.com, ORCID: 0000-0002-5591-7874 (Correspondence should be addressed to ZHANG Bin, Email: zb0205@126.com, ORCID: 0000-0002-5591-7874)

network has small-world structure and scale-free characteristic, growth mechanism and preferential attachment mechanism are essential to understanding its formation and evolution. Citation network can reflect the flow and the inheritance relation of knowledge clearly. Moreover, temporal and spatial factors and nodes heterogeneity can affect its preferential attachment mechanism, and function expression of new nodes and edges can help simulate its growth mechanism. By the appropriate modification, these factors can also be used for co-authorship network.

Co-authorship network reflects the collaboration relationship among knowledge creators. Based on co-authored papers, this kind of network has important significance in understanding the pattern of knowledge exchange and the establishment of interdisciplinary collaboration. Similarity-based algorithms can help examine the link prediction effect of co-authorship network and effectively identify the main influence factors on connection, which lays the foundation for the construction of evolution model. Most current researches still use BA Model and its improvements to study evolution mechanism. A future important research direction would be adding some external attributes of nodes to build hybrid preferential attachment model. As scientific collaboration contains knowledge exchange process, analysis ideas based on scientific research team evolution or network co-evolution can help identify the dynamic factors of the evolution and discover the scientific knowledge diffusion rules.

Many evolution models of knowledge network still have some bias when explaining the evolution of the real network, which leads to the future direction of evaluating network models. Compared to the traditional method which directly analyzes the features of the network structure, the method that uses link prediction to evaluate is more complicated, and applies more advanced mathematical knowledge, yet it is more convincing in the quantitative comparison level. In the future research, the evolution model and link prediction should be integrated into one research framework, such as: 1) When the network evolution mechanism is unclear, the main features that affect the network connection can be found through statistical analysis of the real world network and observation of link prediction; 2) Based on these features and the previous research findings, the evolution model of the specific network can be constructed, events implied rules can be revealed; 3) Use link prediction to design a quantitative evaluation method, which can evaluate advantages and disadvantages of the constructed network model and its evolution mechanism. 2 figs. 79 refs.

KEY WORDS

Knowledge network. Evolution model. Citation network. Co-authorship network. Link prediction.

在情报学理论研究中,知识网络是由科学知识节点和知识关联构成的一类网络,是知识交流网络化的产物^[1],是在社会关系基础上,交流、转移、共享知识资源特别是隐性知识后形成的网络结构。知识网络属于复杂网络的一种,可以进行网络抽象。人们在认识复杂网络形成机制时,总结出一些规律,并将其用网络模型进行刻画。小世界网络(WS模型)^[2]和无标度网络(BA模型)^[3]的发现,再现了真实网络的

小的平均距离、大的聚集系数、不均匀连接等特征,从而掀起了网络科学研究的新一轮高潮^[4],并且在知识网络的结构和演化问题上取得了很多研究成果^[5]。

网络演化主要考虑的是节点的出现和消失、节点间连边的改变这两个方面,其内在的演化机制决定了它所呈现出的纷繁复杂且有趣的特征。早期的研究关注模型描述,通过统计分析的手段对演化过程进行解释。之后,学者们

又从不同角度提出了许多过程模型,通过仿真的手段来揭示这些内在演化机制是如何作用的^[6]。演化过程反映知识网络的时空结构变迁,而演化模型是知识网络内在作用模式及作用过程的抽象表达,可用于探讨和发现知识发展脉络的关键特征,也是将外在结构与内在作用机理联系起来的纽带,有助于达到表里互相诠释的效果。

经典的引证网络和合作网络都是知识网络的一种显性表征。本文以检索式 TS=(“citation network” OR “co-authorship network” OR “collaboration network” AND “evolution model”)在 Web

of Science 中进行检索,文献类型为 Article,时间跨度为 1989 年至 2015 年,共得到 387 篇文章,主要集中在统计物理、计算机、图书情报等领域,近年来相关研究呈现快速上升趋势(见图 1)。在补充国内外一些成果和跟踪引证文献研究的基础上,本文对网络模型的类型、Price 模型和 BA 模型等经典理论模型进行回顾,从引证网络和合作网络两个方面对具有代表性的演化模型进行全面梳理和评述,介绍网络演化模型评价的进展,在总结当前研究现状的基础上,提出未来的研究重点。

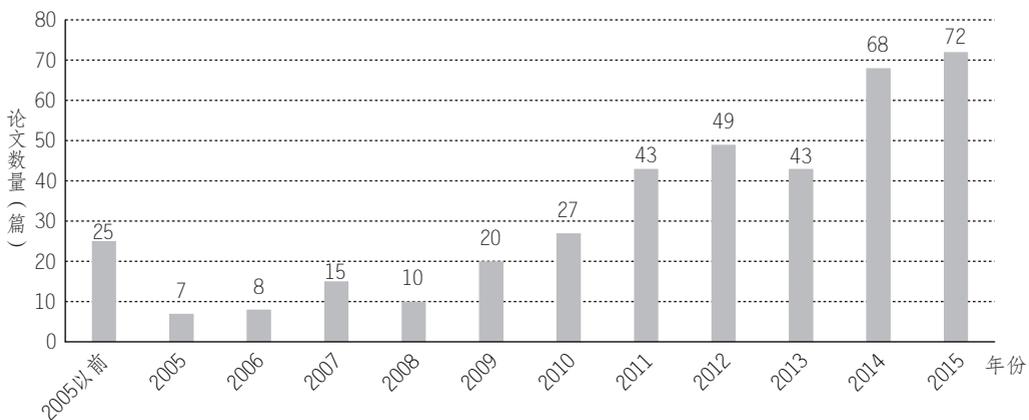


图 1 知识网络演化模型研究论文时间分布

1 理论回顾

统计物理、计算机、图书情报等领域从各自角度构建了不同类型的网络模型,这些模型大致可分为节点属性模型和网络演化模型两大类^[7]。在节点属性模型中,新连接的产生依赖于节点的外部属性信息,表现为同质性(Homophily)^[8-9],比如,术语空间模型^[10-11],每个节点通过各自属性信息来确定在社会或地理空间中的“位置”,以最(次)邻近原则进行择优,即“近水楼台先得月”。节点属性模型往往通过机器学习的方式来确定针对不同网络的各不相同的属性参数组合,以达到较好地预测新连接的

效果,很少考虑网络演化问题。

网络演化模型的基本假设是特殊的网络演化机制促使形成了特殊的网络结构特征,关注重点是网络演化的机制与过程。进一步地,网络演化模型还可以细分为动态模型(Dynamical Model)和增长模型(Growing Model)。动态模型通常起始于一个空的网络,重复对一组固定规模的节点添加和删除连边,直至网络结构特征在统计学意义上不再变化;而增长模型通常起始于一个小的种子网络,添加节点和连边,直至网络达到一个预期规模。增长模型是假设当网络规模足够大时,网络结构特征的相关统计结果会保持不变。与动态模型相比,增长模型一般不考虑删除连边的问题,不用模拟网络从无

到有的全部演化过程,而是模拟一个已存在网络的后续演化前景。对知识网络而言,不管是引证关系还是合作关系,基本上都是从出版物记录中抽取出来,连接一旦形成便保留下来,很少进行删除处理,新连接不断形成并加入到网络中来,同时也很难追溯网络的形成原点。显然,从知识网络的形成特点上看,增长模型更为适用一些。

针对知识网络的增长,Price^[12-13]研究科学文献之间的引用关系网络,发现入度分布服从幂律,并给出了最初的网络模型,称这种“贫者愈贫,富者愈富”的现象为“累积优势”,在社会学当中人们称之为“马太效应”。在 Price 模型中,增长机制指的是论文的数量是不断增长的,新发表的论文会引用以前发表的一些论文作为参考文献;积累优势机制指的是以前发表的论文被一篇新论文引用的概率与该论文已经被引用的次数成正比,这实际上就是择优。之后,Barabási 和 Albert^[3]使用了更为简洁的增长和择优机制来构建具有幂律度分布的 BA 模型。事实上,Price 模型和 BA 模型都是增长模型,而 BA 模型可以被视为 Price 模型无向化后的一个特例。它们可以在一定程度上揭示知识网络不均匀连接的形成机理,但是与真实知识网络仍然存在着很大不同,比如它们考虑到了知识节点的结构信息对于连边形成的影响,却忽略了本身所具有的一些外部属性对连边形成的贡献。知识网络演化模型的关键是增长和择优机制的有效实现,于是经典的演化模型需要改进和扩展。

2 引证网络的演化模型

知识网络中的引证关系能体现出知识流动,因此很多研究使用引证网络来探索知识交流的驱动因素。在真实引证网络中,不同论文的参考文献数量是有差异的,这与该论文所属研究领域、发表时间等因素相关。研究显示,论文篇均参考文献数量随着发表时间呈现出增

长趋势^[14],即出度随时间增长。同时,一篇论文在新发表时被引次数为 0,即入度为 0。Price 模型对上述情况进行了简化,假设每篇论文的参考文献数量为常数 m ,同时假设一篇已有论文被新论文引用的概率与该篇论文已经被引用的次数再加上一个正常数 a 成正比,从而使得任何一篇论文都有被引用的可能。这样,Price 模型的入度分布近似服从于幂指数 $\gamma = 2 + a/m$ 的幂律分布。通常, $a \leq m$, 于是 $\gamma \in (2, 3]$, Price 模型属于非均匀网络。

2.1 择优机制

BA 模型沿用了与 Price 模型类似的度择优机制 $\Pi(k_i) = k_i / \sum_j k_j$, 其中 k_i 是旧节点 i 的度数,新节点倾向于与具有较高度数的旧节点相连接。但更为一般的择优概率是采用 $\Pi(k_i) = g(k_i) / \sum_j g(k_j)$ 来表达, $g(k_i)$ 是预先设定的函数。Liu 等^[15]提出了 $g(k_i) = (1-p)k_i + p$ 的随机和择优混合模型,其中 p 是随机连接的概率。而 Krapivsky 等^[16]提出了 $g(k_i) = k_i^r$ 的非线性择优模型,其中 $r \geq 0$ 是非线性指数。Jeong 等^[17]对很多实际网络进行测量,发现:因特网和引证网络, $r \approx 1$,呈线性关系;而演员合作网络和科研合作网络, $r \approx 0.8$,呈亚线性关系;当 $r > 1$ 时,为超线性关系^[18]。对于有向网络, Dorogovtsev 等^[19]提出了 $g(k_i^{\text{in}}) = k_i^{\text{in}} + A$ 的原始吸引模型,其中 k_i^{in} 是旧节点 i 的入度, $A \geq 0$,称为原始吸引强度。显然,原始吸引模型与 Price 模型中的择优过程是一致的,同时它也有对应无向化的表达形式。

2.1.1 时间因素

在 BA 模型中,由于时间因素,越早加入的节点具有越高的度,也越容易通过度择优机制而被连接。对于庞大且复杂的知识系统来讲,新加入的节点不太可能获得网络中所有节点度数的完整信息。Fortunato 等^[20]提出了一个基于部分信息的秩次择优增长模型来替代 BA 模型,将节点进入的顺序规定为秩次 $R_i = i$,并以此定

义“年龄”为从旧到新,新节点按照年龄择优概率 $\Pi(i) = i^{-r} / \sum_j j^{-r}$ 进行选择,其中 $r \geq 0$ 。秩次择优增长模型没有改变度分布的性质,可以将它理解为对度择优和年龄择优进行了某种转换或映射。

现实中人们对知识的继承和发展的行为表现是:一方面会倾向于利用经典知识和文献,另一方面会追逐科学前沿,吸收最新的研究成果。Price^[12]很早就指出约有 50% 的引证与论文发表时间有关,人们偏好于引用经典文献和近期文献,而其他部分的引用则是随机的。刘向和马费成^[21]通过引入度择优和时间择优进行分析以反映科学知识的继承与更新过程,其中度择优保证对经典科学理论的继承,而时间择优促使对新近知识的吸收。他们认为度择优所体现出的马太效应的作用是全局性的,时间效应所体现出的后发优势的影响则是局部性的,时间效应会在一定程度上平抑度择优所导致的马太效应的负面影响。结合秩次择优增长模型来看,度择优和时间择优其实可以统一在时间维度上进行整体考虑。然而,Price 提到的随机引用又该如何去刻画和解释呢?有学者从空间因素上展开研究。

2.1.2 空间因素

知识创造过程一般基于有限知识领域,不仅受到时间因素制约,还会受到空间因素制约,表现为视阈受限。人们倾向接收“距离”自己较近的知识,这种“距离”表现在时间上,是对经典文献和近期文献的继承和吸收;而表现在空间上,是知识会以分门别类的形式出现,新加入的知识节点倾向模仿或复制已有知识节点的行为,跨领域的知识继承和创造相对较少。

人们在引用一篇文献时,有很大可能性会引用这篇文献的参考文献,而不是独立寻找流行文献。针对这种情况,Krapivsky 和 Redner 提出了平稳增长复制模型^[22]和加速增长复制模型^[23]。平稳增长复制模型假设新节点会随机选择一个旧节点进行复制,或者以 $1-r$ 的概率与旧节点相连接,或者以 r 的概率与旧节点指向的

节点相连接。值得注意的是,该模型的复制操作等价于度择优机制。而加速增长复制模型则假设新节点不仅与旧节点相连接,还会与旧节点指向的节点相连接。在该模型中,新节点的连接数会随着时间的推移而缓慢增加,网络的平均连接数按对数增长。他们统计了 *Physical Review* 杂志 110 年里发表的论文数据,发现每篇论文的引文数量随时间呈对数增长,很好地印证了上述结论。

模仿或复制行为体现了视阈受限,这说明一个独立的知识创造者的知识空间往往是特定的。Li 和 Chen^[24]提出的局域世界 (Local-world) 演化模型可以在一定程度上解释这种行为,在每一时间间隔,新增节点会从规模为 m_0 的网络中随机选择 M 个旧节点作为其局域世界,并在该局域世界中按照度择优方式建立 m 条连接。在 t 时刻, $m \leq M \leq m_0 + t$ 。当 $M = m$ 时,意味着新增节点与选择的局域世界中所有节点都建立连接,同时也表明择优机制失效,网络度分布服从指数分布;当 $M = m_0 + t$ 时,局域世界即整体网络,模型即为 BA 模型,网络度分布服从幂律分布;当 $m < M < m_0 + t$ 时,该模型的度分布会呈现出从指数分布到幂律分布的演化。基于知识空间受限的假设,局域世界的思想可以很自然地应用在知识网络中。

知识创造者虽然视阈受限,但也可以在若干个领域之间进行知识继承和创造,Price^[12]就指出参考文献中的外部连接约占 10%。马费成和刘向^[25]在度择优和时间择优的基础上,增加了交叉连接机制,使得新节点会随机选择另一个局域世界以度择优方式进行连接,这种跨领域交叉连接满足了学科知识交叉引用的要求,同时还能形成一定的聚集拓扑特征。该模型所构建的局域世界数量是一定的,意味着一般学科所包含的知识领域是稳定的,短期之内不会出现全新的知识领域。但随着网络规模逐渐扩大,整个网络的聚集性会由于静态局域世界数量的设定,呈现出下降趋势。于是,刘向等^[26]考虑到知识网络兼具小世界结构和不均匀连接的

特征,构造了一个特定结构形成的过程模型,他们引入核心知识节点和相关知识节点这两个概念,相关知识节点为核心知识节点的邻居节点。当一个新节点加入时,会从网络中以度择优方式选择 x 个核心知识节点进行知识领域定位,从这 x 个核心知识节点的邻居节点中选择 $m - x$ 个相关知识节点以表现知识领域受限。当 $x > 1$ 时,表明定位了多个知识领域,从而体现出跨领域交叉的特征。

2.1.3 节点的异质性

知识网络中节点的度及其增长速度与该节点的“距离”有关,这一现象体现在时空维度下的不同择优机制中,但更为深刻地看,这还与节点的内在属性相关,即节点本身具有异质性。以论文引证网络为例,尽管每个月都有大量的论文刊发出来,但不同期刊所刊载论文的质量差距却很大。有很多论文即使发表出来很长时间也很难获得引用,而一些高质量的论文却可以在较短时间内引起关注,并获得大量引用。一般认为,论文质量是影响引证网络形成的内在动力,而网络结构信息只是其外在表现。

Bianconi 和 Barabási^[27] 将节点的内在属性称为适应度 (Fitness), 择优概率与节点的度和适应度的乘积成正比。适应度模型假设每个节点加入网络时就有个固定的适应度,服从一定分布规律。尽管从形式上看,适应度模型与 BA 模型差别不大,但对于适应度模型的分析则要复杂得多,取决于适应度的分布情况。实际上,不同节点表现出的适应度是异质的,而且还会随着时间而衰减。以往基于择优的模型将它们分开看待,Medo 等^[28] 同时考虑这两种因素,并构建了相应的模型,依据不同的输入假设,度分布可以呈现出指数、对数正态或者幂律衰减,从而更好地模拟真实知识系统。

适应度模型及其改进模型可能会有以下特征:①先到者赢,若网络中节点的适应度相同且不为零,即节点是同质的,那就退化为 BA 模型,呈现出无标度特征,越早加入的节点具有越高的度,也越容易被连接;②适者更富,若网络中

节点的适应度不同,即节点是异质的,那不论节点的年龄大小,适应度较高的节点会越容易被连接,但适应度会随时间而衰减;③赢者通吃,随着时间的演化,新节点不断加入,适应度最高的那些节点会获得网络中一定比例的连接数,从而表现出“赢者通吃”现象。

2.2 增长机制

和 Price 模型一样,BA 模型也假设网络每次新加入一个节点,新增节点的连接数为常数 m 。为了和实际网络连接数加速增长的情形相符,后续研究采用幂函数^[29]、对数函数^[30] 等增长函数来替代常数 m ,得到的模型即为变 m 模型。但对于大多数真实网络,连接数的增加往往是开始比较快,之后逐渐变慢且往往会有一个上限值,这种情况可以用一个饱和函数 $m(i) = [M(1 - e^{-i})] + 1$ 来反映, M 为最多增加的连接数^[31]。BA 模型强调网络度分布应独立于演化时间或网络规模。而在初始网络确定、择优实现方法与 BA 模型一致的情况下,变 m 模型的网络度分布会依赖于演化时间或网络规模,即变 m 模型的网络度分布对特定时间都是服从幂律分布的,但却是时间相依的。事实上,变 m 模型加入了对实际网络演化历史的抽象,将新增节点的连接数表示为关于时间的增长函数,以此来表现这种与时间相依的非平稳特性。因此,可将 BA 模型称为平稳的无标度网络,而变 m 模型则称为非平稳的无标度网络。

知识网络中新增节点的连接情况可描述成人们对知识的继承和发展,而网络中节点的被连接情况随时间的变化可反映为知识的历时老化过程。Avramescu^[32] 认为文献老化与其本身的质量和数量密切相关,不同质量、不同种类文献的老化趋势可以描述为四种典型曲线,这些曲线可以从知识节点的历时被连接数的单调性和曲线拐点来粗略判断。马费成和刘向^[33] 考虑新增节点出度增长为线性函数和对数函数两种情况,选择度择优和时间择优两种方式来构建演化模型,研究结果表明:出度增长会使得度分

布更加均匀,且增长速度越快,度分布越平坦;当出度线性增长时,若度择优较强,知识节点的历史被连接数单调上升,不表现为老化现象,若度择优较弱,则是单调下降的,表现为老化现象;当出度对数增长时,一直表现为老化现象;时间效应使得新加入的节点获得了较多的利用,会自动平抑马太效应的负面影响。

上述模型对增长机制的考虑都是基于均匀的单节点增长,事实上还可以将这一限制放开,考虑非均匀的多节点增长模式。马费成和刘向^[34-35]分别讨论了在新增节点数量是时间的线性、对数、指数、阶跃、logistic等函数,择优机制为度择优时,网络演化模型的度分布和节点的历时被连接数的变化情况。他们认为,如果以节点的历时被连接数来表征知识利用效率,那么知识利用效率与其所属知识领域的发展趋势有关,随着扩张而增加,随着衰退而减小。在这些增长模式中,指数增长模式下度分布的幂指数相对较小,较为平坦,此时马太效应的影响相对较弱,还没有表现出“赢者通吃”的特征,更有利于知识传播和利用,适合于新生的知识领域;而logistic增长模式下度分布的幂指数为 $2 + 1/m$,可能代表了更为成熟的知识领域,logistic增长的前端可以看成是指数增长,中部是线性增长,尾端是对数增长,直至最后的零增长,相应地展现了事物产生、发展、成熟直至衰老的过程。

3 合作网络的演化模型

合作网络体现了知识创造主体间的合作关系,以合著论文为基本表现形式。学者们对合作网络的结构^[36]和演化^[37]以及合作行为模式^[38]进行统计和建模分析,总结为以下特征:①网络具有小世界结构,具有较高的聚集系数,会形成明显的合作小团体;②网络具有无标度特征,度分布服从幂律分布;③不管是领域内还是领域间,建立合作遵循度择优机制;④不同学科领域的合作行为有着明显差别,比如生物科学

家往往比数学家或物理学家拥有更多的合作者,但并不容易形成“拥有共同合作者的两个科学家也将成为合作者”这一现象。更广泛的统计结果发现:在合作规模上,科学与工程领域高于社会科学领域,社会科学领域高于艺术与人文领域,科学与工程以及社会科学领域团队合作增长趋势明显,而艺术与人文领域则没有显著变化^[39-40];在合作方式(单个作者、校内合作、校际合作)上,这些领域也存在着结构性的差异^[41]。

对不同学科合作网络的研究,主要围绕这样三个问题展开:①影响建立合作关系的主要特征有哪些?②能否利用这些特征来构建网络模型,进而分析合作网络的演化过程?③知识是如何通过科研合作行为得到传播、转移和扩散的?这些问题都涉及对网络演化机理的讨论。

3.1 局部结构特征

识别影响建立合作关系的主要特征,可以从利用节点之间的相似性进行链路预测入手。这一研究思路假设两个节点之间的相似性越大,它们之间存在连接的可能性也就越大。相似性表达的是一种接近程度,它有很多刻画的方法。其中,最为直接的就是利用节点的外部属性信息。两个人如果具有相同的专业领域、研究兴趣等,就说明他们很相似,在一定的条件下,就可能产生合作。应用节点的外部属性信息可以得到一些主要特征,如:基金支持、地理距离、社会距离、研究领域等,显著影响科研合作^[42],但在很多情况下获得这些信息并不容易。

与节点的外部属性信息相比较,已观察到的网络的局部结构特征或者节点的历史行为信息则更容易获取且更为可靠。合作网络通常具有高聚集性,这意味着科学家之间的合作一般遵循三元闭包(Triadic Closure)原则,会体现出传递性^[43]。也就是说,新连接的产生依赖于网络的局部结构特征。通过节点局部信息,如节点的共同邻居和度,可以计算出两个节点之间

的相似性,不妨称为结构相似性。Liben-Nowell和Kleinberg^[44]在对大型合作网络的实证研究中,发现仅考虑节点共同邻居的方法和Adamic/Adar指标就可以取得非常好的预测效果。因此,可以通过结构相似性指标来观察合作网络中的链路预测效果,从而分析影响连接建立的主要特征^[45]。

辨识网络局部结构特征可以增强对合作机制的理解。具有相似全局结构统计特征的网络,可能由于其生成机理的不同,反映出不同的局部结构。Milo等^[46]引入模体(Motif)来刻画网络中反复出现的相互作用的基本模式,它们在真实网络中出现的概率远高于相应的随机网络。使用模体分析合作网络的局部结构,有助于准确理解合作网络的演化机理^[47]。科学家会根据自身所具有的科研合作资本来选择合作伙伴,与网络分析结合起来,体现出关系嵌入和结构嵌入。关系嵌入行为可以理解为科学家倾向建立和保持长期合作关系;而结构嵌入行为可以理解为科学家倾向通过第三方共同伙伴与新合作者建立合作关系。关系嵌入行为和结构嵌入行为反映了一种组织惯例和成长路径,强化了对已有网络结构的依赖,也导致了网络模体和反模体形式的共同特征的出现;而从信息不对称和人力资本条件约束下科研合作全局演化的社会选择机制来看,这两种行为也可以视为科研合作成本与效率择优的结果。

3.2 网络模型构建

3.2.1 BA模型及其改进模型

与引证网络在时序上单向无回路的特征不同,合作网络在每一时间间隔,除了有新节点加入网络,旧节点也可以尝试建立新的连接,或者断边重连,或者对现有连接进行重复连接。在建立新的连接时,当前最流行的择优机制依然是度择优。但是Abbasi等^[48]认为,相比度中心

性和接近中心性,中介中心性是反映择优过程的一个更为显著的指标。随着网络不断增长,节点在择优选择时,会从局部的度中心性过渡到全局的中介中心性。一个现实的解释是博士生导师或者博士后合作导师会被一个科研新人作为优先合作对象。正是由于他们的中介作用,合作网络优先选择那些具有高的中介中心性的节点,朝着不同方向延展。而针对断边重连的情形,Albert和Barabási^[49]提出过一个扩展的BA模型,它会以某一概率随机重连网络中已有连边,具有幂律度分布,且幂指数可以通过对若干参数的选取而取值为(2,3)上的实数。

节点之间重复连接的频次可以很自然地转换为连边权重,这提供了表示连边权重演化的一种方式,同时也暗含着权重会在连接建立上起到重要作用。定义点强度为与节点连接的所有连边的权重之和,即 $s_i = \sum_j w_{ji}$ 。早期的网络模型将度择优扩展到点强度择优,即择优概率正比于点强度^[50];还认为点强度高的节点在吸引新增节点连接的同时,与其相连连边的权重会增加一个改变量 Δw ^[51]。这些模型再现了真实加权网络中节点度、强度、连边权重三者服从幂律分布的特征^[52]。为了反映现实存在的局域世界,Pan等^[53]提出一个加权的局域世界演化模型,将点强度择优驱动、连边权重逐渐加强以及局域世界等机制放在一起考虑,反映指数分布加权网络向幂律分布加权网络过渡的全过程。但这些模型依然存在一个很大的不足,即无法再现很多真实网络所具备的大聚集系数这一特征。

Li等^[54]构造了一个与经济物理学家合作网络性质一致的加权网络演化模型,利用局部结构信息去提升网络的聚集系数,将之命名为局部路径相关(Local-path-related)的择优机制,数学表达为:

$$P_{n \rightarrow i} = (1-p) \frac{k_i}{\sum_j k_j} + (p-\delta) \frac{s_i}{\sum_j s_j} + \delta \frac{l_{ni}}{\sum_{j \in \partial_n^d} l_{nj}}$$

其中,节点 n 连接到节点 i 的概率由三部分组成, k_i 为节点 i 的度数, s_i 为节点 i 的点强度, l_{ni} 为节点 n 和节点 i 的相似距离。 ∂_n^d 表示距离节点 n 的级数小于或等于 d 的所有近邻节点的集合,如 ∂_n^1 表示最近邻的集合, ∂_n^2 表示最近邻和次近邻的集合。对 p 和 δ 的不同取值可使择优机制退化到 BA 模型及其改进模型上去。该模型考虑了局部路径对演化过程的影响,当 $\delta \neq 0$ 时,能够明显提升网络的聚集系数,这是一个非常有意义的改进。

3.2.2 加入节点的外部属性信息

很多学者在构建演化模型时对融合网络结构和节点属性做过一些尝试。Rosvall 和 Sneppen^[55] 为了研究信息流在团队形成过程中的作用,使用交流和社交导航机制来对人们的兴趣和社交结构进行建模。一个现实的案例是博士生在经历多年的学习和研究工作后,他的科研技能和兴趣与导师非常相似。该博士生在确定毕业去向时,不可避免地会受到导师的影响。经验表明,他有很大可能去导师的科研伙伴那里,这样比较容易建立联系。也就是说,当该博士生为了更容易地获得他感兴趣的信息而扩展“朋友圈”时,会使用在网络中流动的信息,而这种信息流代表的是对整个社会系统的整合。进一步地,可以对兴趣搜索、朋友推荐等社交行为进行形式化表达,并据此来构建网络演化模型,通过仿真实验来观察生成网络的拓扑结构特征^[56-57]。

Sun 等^[58-59] 将影响建立合作关系的主要特征分为三类:基于邻居节点的、基于路径的以及基于外部属性的。基于邻居节点的预测指标选取了共同邻居 (CN) 和 Adamic-Adar, 基于路径的预测指标选取距离 (Distance) 和 Katz, 而基于外部属性的预测指标选取兴趣相似度 (Interest Similarity) 和择优连接 (PA)。兴趣相似度计算的是作者已发表论文题目的余弦相似度。他们针对物理学 (PRE)、计算机科学 (DBLP) 和图书情报学 (LIS) 这三个学科领域数据集,使用决策树算法对各项指标的预测效果进行分析,发现

基于路径的和基于外部属性的预测指标效果非常好。之后,他们设计了一个融合网络路径和节点属性的合作网络演化模型,称为基于事件和兴趣的模型 (Event-Interest Model, EIM)。与以往的 BA 模型及其改进模型不同,该模型在演化过程中,择优机制由两部分组成:第一,事件,认为连边的产生最初是由事件驱动的;事件主要是指作者之间的科学交流活动,包括正式的会议或者非正式的会面等,在这样的事件中,作者之间有机会认识彼此或者加强联系,从而为将来合作创造条件。第二,兴趣,认为研究兴趣会最终决定作者之间是否会真的进行合作。

对事件的模拟,结合随机游走和随机连接两种机制,从两个不同角度来反映建立合作关系的特征。随机游走过程采用转移概率矩阵来描述,每一个被访问到的节点都成为事件的参与者,从而反映出:①以前合作过的作者之间可能会再度合作;②拥有共同邻居的作者之间可能会进行合作;③有积极合作意愿的作者会更容易建立合作关系。而随机连接则反映了科研工作者选取合作者的随机性,使得他有一定概率接触到领域之外的人,从而产生跨领域合作。对兴趣的模拟,采用了术语空间模型,将节点以均匀概率投影到空间中,“位置”远近表示兴趣的相似程度。EIM 模型为增长模型,达到预期网络规模即可停止模拟,仿真实验结果生成了真实网络所具有的主要拓扑结构和分布特征。

尽管很多学者都有着相似的融合动机,但对于网络结构和节点属性的理解会因待解决问题不同而千差万别,所构建的网络模型的类型和方法也是不同的。这些加入节点的外部属性信息的演化模型颇为复杂,不仅考虑了随机性驱动机制,还考虑了确定性驱动机制。虽然仿真实验结果可以部分反映真实网络所具有的主要特征和性质,但是仍然很不全面。究其原因,是由于这些模型与自然界和人类社会中普遍存在的随机性和确定性机制还有一定差距,由多种机制混合择优而构建的网络演化模型还需要进一步探讨,这也是网络模型理论研究工作的

重要方向之一^[60]。

3.3 合作行为与知识交流

科研合作行为中包含着知识交流环节。关于合作行为与知识交流的传统研究思路是,通过对合作网络的时序演化和拓扑结构量变与质变等进行观察和统计分析,来探讨知识扩散的模式和机制^[61],其缺点是难以深入至动力学分析。知识传播动力学的研究难点在于构建包含知识因素的网络模型。这里总结两种基于知识因素的研究思路:第一,由科研团队演化而引起整个合作网络“自底向上”的涌现;第二,结合引证网络和合作网络,进行协同演化,研究知识的增长与扩散。

3.3.1 科研团队演化

合作网络的生成和演化,实际上是一个个科研团队在一定的合作机制下反复组合、解散的过程。从动态重组的角度来构建演化模型,要比从单一的择优机制建立连接,显得更符合实际情况。团队的组合和结构的变迁影响着团队内部知识共享、创造以及可能的跨团队知识交互,同时后者也会反作用于前者。该研究思路可以归纳为:科研团队演化是整个网络演化的基础,个体间的知识交流是整个网络中知识增长和扩散的基础,进而观察在不同的团队机制下的网络宏观统计特征。

Guimera等^[62]提出过一个很简洁的团队自组织模型,它由三个参数决定,即团队规模、新成果中的新成员部分、现有成员再次进行合作的倾向。该模型表明一个大型的连通社区结构的出现可以描述成一个相变过程。研究发现,对艺术和科学领域的团队来说,这一团队自组织机制决定了合作网络的结构和团队绩效。夏昊翔等^[63]在这一研究基础上,结合知识增长和扩散现象,建立了一个仿真模型。他们引入四个实体类(个体类、团队类、网络类、知识类)、两个活动类(知识传播类、协作类)和部分辅助类,实现了三个加入知识因素的网络模型,对基于团队自组织机制的建模框架进行了初步尝试。

团队自组织机制会使得知识与网络处于协同演化状态。Luo等^[64]假设在一个随机网络中,节点可以与其邻居之间交流知识,或者沿着连边移动到一个新的位置。其中,知识转移规则是两个已连接的节点,只有在它们的知识距离小于一个给定阈值的情况下,才能进行知识交流。在阈值范围内,知识距离越大,知识交流的效果越好;邻居调整规则是节点可以向着长程位置移动或者驻留在局部的群落之中。仿真结果表明,知识转移和邻居调整之间的双向影响使得网络结构和知识扩散在全局范围内协同演化,而且在知识转移过程中,可以很明显地观察到小世界结构的出现与消失。

基于局域世界的考虑,Yang等^[65]提出了一种非均匀合作网络的知识扩散模型。该模型引入择优扩散机制和知识吸收能力,知识吸收能力与节点的度数相关。在每一时间间隔,随机选择节点作为知识发送者,从该节点之前的合作者中,以正比于合作发文数量的择优概率选择知识接收者。采用平均知识存量、存量的方差和方差系数来衡量知识的增长、扩散和扩散效果。在进行三组对比实验后,发现网络规模和知识演化机制影响着知识扩散。之后,Wang等^[66]还将知识吸收过程作为一个重要影响因素考虑进来,对上述模型进行了优化和改进。此外,还可以将合作网络结合传染病模型进行知识传播动力学方面的探索^[67-68]。传染病模型适用对象最好是有向网络,由于合作网络在知识扩散方面更多地体现为知识共享,因此在实际研究工作中需要对传染病模型进行改造。当然,对于团队合作关系,可以通过调查来确定谁处于主导地位,谁处于次要地位,将无向网络改造为有向网络,从而明确知识的发送方和接收方,但这需要对团队演化动态进行更加细致的分析。关于这点尚未产生很好的研究成果,有待进一步探讨。

3.3.2 网络协同演化

引证网络是典型的知识扩散网络,从基础知识到研究前沿的扩散路径,可以反映出知识

范式转变和知识创新的主要进程。科学计量学家很早就开始利用引证网络进行科学知识的扩散研究,较为著名的如 Garfield 等^[69]利用引证时序图来表征知识扩散的历史网络,用以分析遗传学领域的知识扩散过程。正是由于引证网络的这种特殊作用,如果将合作网络与引证网络相结合,分析它们的协同演化,就有可能得出促进科学知识发展的新模式。

Börner 等^[70]提出过一个过程模型,命名为 TARL (Topics, Aging, and Recursive Linking) 模型,用来模拟作者合作网络和论文引证网络的同时生长状况,试图捕捉作者和论文在知识生产、存储、传播过程中所起的作用。研究假设信息扩散直接发生在作者合作过程中,间接发生在阅读他人的论文中。该模型对每个作者和每篇论文都赋予单一主题,并根据主题的特异性进行层次化组织。同时,综合考虑同质网络内的连接和异质网络间的连接,结合主题的层次化,来表现知识的增长与扩散。利用该模型生成网络统计特征,对发表在 *PNAS* 上长达 20 年的文献数据进行验证,得到的结论有:作者偏好引用最近发表的论文,作者倾向引用他们以前阅读过的论文中所引用的论文。网络分层思想可以拓展到“文章—作者—机构—国家”这样的结构上,称之为“瀑布”模型,利用路径分析来识别在各个实体层次上的重要节点。Liu 等^[71]认为这些路径及重要节点是观察由合作关系导致知识扩散的重要依据。参照之前 Börner 等的做法,将引证关系引入到这些实体层次之上或者之间,或许能得到一些新的结论。

当前对于合作网络的研究大多是基于论文合著关系的,或者进一步将合著论文频次转化为权重。当关注点在知识交流上时,仅仅考虑合著关系就显得比较单薄。Li 等^[72-73]所在研究小组收集过一个经济物理学家合作网络,将作者之间的合作、引用和致谢等三种关系综合在一起考虑。当然,从网络构造上讲,可以分别构造出三个不同网络来进行研究,也可以将这三种不同层次上的知识交流以连边权重的方式融

合在一起,从而得到相似权或者相异权。后者的做法实际上是将合作网络结构和知识扩散耦合在一起而产生一个加权网络。

将合作网络结合引证网络进行协同演化研究并不容易,究其原因是要找到一个合适的融合途径比较困难,它们毕竟代表两种有区别但又有联系的研究思路。对合作网络和引证网络按照某种方式进行分层组织,再对同质网络内和异质网络间的连接所构成的路径进行分析是非常有效的,而在一定的规则设定下,将分层网络整合为单一网络也不失为一种选择。

4 网络演化模型的评价

网络演化模型的评价是指所构建的模型在解释真实网络方面所能达到的程度。传统的研究思路是,根据演化模型生成模拟网络,将模拟网络主要结构特征的统计结果与真实网络进行对比,相关指标越接近真实网络,表明模型的解释程度越好。但问题是,刻画网络结构特征的统计量非常多,学者们在提出各自的演化机制时,关注角度和期望要解释的内容也不尽相同,对所比较指标的选取常常陷入学术争论中。构建演化模型的关键往往是确定两个节点之间产生连边的机制,这与链路预测的功能是相符的。因此,链路预测有望为网络演化模型提供一个简单而有效的比较平台。

利用链路预测进行评价的一种直观方法是,在给定网络情形下,将某个时点之后出现的连边作为测试集,将各种演化机制应用到当前网络结构上,利用 PR 曲线、ROC 曲线、AUC、排序分等评价方法来比较新生成的连边和测试集中的连边^[59]。需要说明的是,这一方法有时会受制于采集到的数据,因为这些数据本身可能存在缺失,从而对评价结果造成干扰。在随机去除网络中小部分的连边之后,还能保持一致性的结构特征,可以反映出网络中蕴含的规律。基于这一观点,Lü 等^[74]通过邻接矩阵的扰动设计了一个通用的结构一致性指标,且这一指标

不需要针对网络的任何先验知识。他们在大量真实网络中进行实验,结果表明该指标在评价链路预测的准确性和网络结构的鲁棒性上表现都很优秀。进一步地,该指标还能用于监测网络演化机制的突然变化。

也有学者尝试将链路预测应用在已提出的网络理论模型上。Ahn 和 Jung^[75]将六种链路预测指标(CN、AA、RA、Jaccard、PA、Simrank)应用到 WS 模型和 BA 模型中,试图解释预测精度和模型参数之间的相关性,发现网络的平均度与预测精度之间存在正相关性。王文强等^[76-77]提出利用仿真生成模拟网络,在知道演化情况的基础上,参考似然分析法来量化比较各种不同机制对网络的增长行为。不妨认为促使似然取最大值时的模型参数为最优,这样就可以对演化模型进行量化比较。同时,还可以对相似性

指标进行耦合,利用链路预测评价耦合算法的精确性,从而分析和推断网络演化机制^[78]。

5 总结与展望

通过上述回顾和梳理可知,图书情报学领域很早就关注到知识网络演化模型这一问题,早期的 Price 模型更是奠定了 BA 无标度网络发现的思想基础。知识网络具有小世界结构和不均匀连接的特征,增长和择优机制是理解其形成和演化的关键。已有研究多以引证网络和合作网络为具体对象,通过网络演化来探讨学者之间相互影响、合作的演化机理,以及知识的传播和发展脉络。本文总结出知识网络演化模型的研究框架(见图 2),并从研究方法和研究内容两个方面指出未来研究需要进一步深化之处。

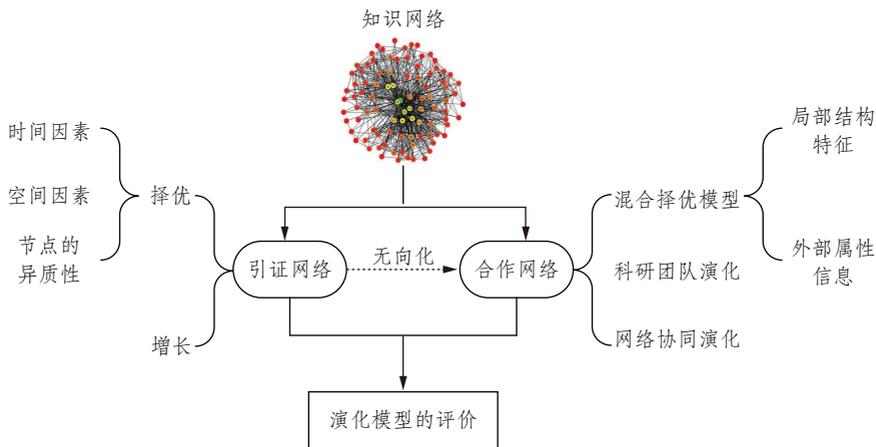


图 2 知识网络演化模型研究框架

在研究方法上,当前大多数研究都是利用 BA 模型及其改进模型来研究演化机理问题。引证网络能明确地体现出知识的流动和前后承接关系,时空因素、节点的异质性等会影响择优机制,新增节点和连边的函数表达可以模拟增长机制。这些驱动因素在经过合适改造后,对于合作网络也是适用的。合作网络体现了知识创造主体之间的合作关系,包含着知识交流环节,可以结合网络的局部结构特征和节点的外

部属性信息来建立混合择优模型。一种方法是从结构信息中抽象出很多属性,然后和外部信息各个属性放在一起构建模型;另一种方法是量化不同属性对连边形成的贡献,然后找到把不同的外部信息和结构信息融合起来的方法,提高模型解释真实网络的准确性。后者更有价值,也更具挑战性。

众多学者对真实网络的实证研究表明:真实网络既不是规则网络,也不是随机网络,而是

确定性和随机性混合而成的网络,表现出复杂性与多样性。知识网络演化模型为观察到的知识交流现象提供了解释。为此,未来的研究应对模型的适应性和有效性提出要求。

(1) 知识网络演化模型的适应性,指的是模型与研究对象相适合。随着信息技术发展和全球化影响,知识网络将人类知识在时空上构成了一个无限延展的关联系统,但依然受到时空、经济、政治等因素的严重影响,虽然科学界一直在努力削弱这些不利情况的影响。以科研合作为例,科学家之间的合作关系就算已经确定,合作终止或者不再保持活跃性的概率也不是可以忽略不计的。从这个意义上讲,合作网络中的连接都应该是过去的、短暂的,但从形式上却被永远记录下来,从而导致对任何给定时间范围的网络的相关描述都可能是不准确的。为了克服这一限制,一种实际的作法是滑动不同长度的时间窗口以得到实验数据^[79]。这种对数据的获取和处理方式,虽然可能会影响到一些网络的结构特征,比如凝聚性,但却更加接近于社会学家的观点。因此,研究对象与其网络抽象

是否适合,是研究工作中必须面对的一个现实问题。

(2) 知识网络演化模型的有效性,指的是模型所达到的解释程度。众多知识网络演化模型在解释真实网络演化时,依然存在着不小差距,因此对模型进行评价应当是今后着力的方向。与传统的直接比较网络结构特征的统计量的方法相比,利用链路预测的原理来设计评价方法要复杂得多,对数理知识的运用也更为高深,但后者在量化比较层次上更有说服力。未来应当将演化模型和链路预测集成到一个研究框架中,不妨遵循以下步骤:①在不明确网络的演化机制情况下,通过对真实网络的物理统计分析以及观察链路预测效果,找到影响该网络中连接建立的主要特征;②根据这些主要特征,再结合相关理论模型中的研究成果,能构造出针对特定网络对象的演化模型,从而揭示出其中蕴含的规律;③可以利用链路预测理论设计量化评价方法,来评价所构建的网络模型及其演化机制的优劣。

参考文献

- [1] Seufert A, Von Krogh G, Bach A. Towards knowledge networking[J]. *Journal of Knowledge Management*, 1999, 3(3): 180-190.
- [2] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks[J]. *Nature*, 1998, 393(6684): 440-442.
- [3] Barabási A L, Albert R. Emergence of scaling in random networks[J]. *Science*, 1999, 286(5439): 509-512.
- [4] Barabási A L. Scale-free networks: a decade and beyond[J]. *Science*, 2009, 325(5939): 412-413.
- [5] 刘向, 马费成, 陈潇俊, 等. 知识网络的结构与演化——概念与理论进展[J]. *情报科学*, 2011, 29(6): 801-809. (Liu Xiang, Ma Feicheng, Chen Xiaojun, et al. Structure and evolution of knowledge network: concept and research review[J]. *Information Science*, 2011, 29(6): 801-809.)
- [6] Dorogovtsev S N, Mendes J F F. Evolution of networks[J]. *Advances in Physics*, 2002, 51(4): 1079-1187.
- [7] Toivonen R, Kovanen L, Kivela M, et al. A comparative study of social network models: network evolution models and nodal attribute models[J]. *Social Networks*, 2009, 31(4): 240-254.
- [8] McPherson M, Smith-Lovin L, Cook J M. Birds of a feather: homophily in social networks[J]. *Annual Review of Sociology*, 2001, 27: 415-444.
- [9] Centola D, Gonzalez-Avella J C, Eguiluz V M, et al. Homophily, cultural drift, and the co-evolution of cultural groups[J]. *Journal of Conflict Resolution*, 2007, 51(6): 905-929.
- [10] Boguñá M, Pastor-Satorras R, Díaz-Guilera A, et al. Models of social networks based on social distance attachment

- [J]. *Physical Review E*, 2004, 70(5):056122.
- [11] Wong L H, Pattison P, Robins G. A spatial model for social networks[J]. *Physica A: Statistical Mechanics and its Applications*, 2006, 360(1):99-120.
- [12] de Solla Price D J. Networks of scientific papers[J]. *Science*, 1965, 149(3683):510-515.
- [13] de Solla Price D J. A general theory of bibliometric and other cumulative advantage processes[J]. *Journal of the American Society for Information Science*, 1976, 27(5):292-306.
- [14] Biglu M. The influence of references per paper in the SCI to impact factors and the Matthew effect[J]. *Scientometrics*, 2008, 74(3):453-470.
- [15] Liu Z H, Lai Y C, Ye N, et al. Connectivity distribution and attack tolerance of general networks with both preferential and random attachments[J]. *Physics Letters A*, 2002, 303(5):337-344.
- [16] Krapivsky P L, Redner S, Leyvraz F. Connectivity of growing random networks[J]. *Physical Review Letters*, 2000, 85(21):4629-4632.
- [17] Jeong H, Néda Z, Barabási A L. Measuring preferential attachment in evolving networks[J]. *EPL (Europhysics Letters)*, 2003, 61(4):567-572.
- [18] Golosovsky M, Solomon S. Stochastic dynamical model of a growing citation network based on a self-exciting point process[J]. *Physical Review Letters*, 2012, 109(9):098701.
- [19] Dorogovtsev S N, Mendes J F F, Samukhin A N. Structure of growing networks with preferential linking[J]. *Physical Review Letters*, 2000, 85(21):4633-4636.
- [20] Fortunato S, Flammini A, Menczer F. Scale-free network growth by ranking[J]. *Physical Review Letters*, 2006, 96(21):218701.
- [21] 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.)
- [22] Krapivsky P L, Redner S. Organization of growing random networks[J]. *Physical Review E*, 2001, 63(6):066123.
- [23] Krapivsky P L, Redner S. Network growth by copying[J]. *Physical Review E*, 2005, 71(3):036118.
- [24] Li X, Chen G. A local-world evolving network model[J]. *Physica A: Statistical Mechanics and its Applications*, 2003, 328(1):274-286.
- [25] 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.)
- [26] 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 Engineering—Theory & Practice*, 2013, 33(7):1836-1844.)
- [27] Bianconi G, Barabási A L. Bose-Einstein condensation in complex networks[J]. *Physical Review Letters*, 2001, 86(24):5632-5635.
- [28] Medo M, Cimini G, Gualdi S. Temporal effects in the growth of networks[J]. *Physical Review Letters*, 2011, 107(23):238701.
- [29] Dorogovtsev S N, Mendes J F F. Effect of the accelerating growth of communications networks on their structure[J]. *Physical Review E*, 2001, 63(2):025101.
- [30] Shi D H, Chen Q H, Liu L M. Markov chain-based numerical method for degree distributions of growing networks

- [J]. *Physical Review E*, 2005, 71(3): 036140.
- [31] Shi D H, Zhou H J, Liu L M. A discussion of Barabási-Albert's 1999 paper[J]. *Physics Procedia*, 2010, 3(5): 1767-1774.
- [32] Avramescu A. Actuality and obsolescence of scientific literature[J]. *Journal of the American Society for Information Science*, 1979, 30(5): 296-303.
- [33] 马费成, 刘向. 知识网络的演化(III): 连接机制[J]. *情报学报*, 2011, 30(10): 1015-1021. (Ma Feicheng, Liu Xiang. Evolution of knowledge networks (III): attachment mechanism[J]. *Journal of the China Society for Scientific and Technical Information*, 2011, 30(10): 1015-1021.)
- [34] 马费成, 刘向. 知识网络的演化(I): 增长与老化动态[J]. *情报学报*, 2011, 30(8): 787-795. (Ma Feicheng, Liu Xiang. Evolution of knowledge networks (I): growth and obsolescence[J]. *Journal of the China Society for Scientific and Technical Information*, 2011, 30(8): 787-795.)
- [35] 马费成, 刘向. 知识网络的演化(II): 增长老化与知识产生时点的关系[J]. *情报学报*, 2011, 30(9): 916-921. (Ma Feicheng, Liu Xiang. Evolution of knowledge networks (II): relationship between growth, obsolescence and obsolescence time-point of knowledge[J]. *Journal of the China Society for Scientific and Technical Information*, 2011, 30(9): 916-921.)
- [36] Newman M E J. The structure of scientific collaboration networks[J]. *Proceedings of the National Academy of Sciences*, 2001, 98(2): 404-409.
- [37] Barabási A L, Jeong H, Neda Z, et al. Evolution of the social network of scientific collaborations[J]. *Physica A: Statistical Mechanics and its Applications*, 2002, 311(3): 590-614.
- [38] Newman M E J. Coauthorship networks and patterns of scientific collaboration[J]. *Proceedings of the National Academy of Sciences*, 2004, 101(suppl 1): 5200-5205.
- [39] Wuchty S, Jones B F, Uzzi B. The increasing dominance of teams in production of knowledge[J]. *Science*, 2007, 316(5827): 1036-1039.
- [40] Ma F, Li Y, Chen B. Study of the collaboration in the field of the Chinese humanities and social sciences[J]. *Scientometrics*, 2014, 100(2): 439-458.
- [41] Jones B F, Wuchty S, Uzzi B. Multi-university research teams: shifting impact, geography, and stratification in science[J]. *Science*, 2008, 322(5905): 1259-1262.
- [42] 贾茜, 李亚婷, 张斌. 科学合作的形成及其影响因素[J]. *情报理论与实践*, 2014, 37(6): 40-45. (Jia Xi, Li Yating, Zhang Bin. Formation and influencing factors of scientific collaboration[J]. *Information Studies: Theory & Application*, 2014, 37(6): 40-45.)
- [43] Rapoport A. Spread of information through a population with socio-structural bias: I. assumption of transitivity[J]. *The Bulletin of Mathematical Biophysics*, 1953, 15(4): 523-533.
- [44] Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks[J]. *Journal of the American Society for Information Science and Technology*, 2007, 58(7): 1019-1031.
- [45] 张斌, 马费成. 科学知识网络中的链路预测研究述评[J]. *中国图书馆学报*, 2015, 41(3): 99-113. (Zhang Bin, Ma Feicheng. A review on link prediction of scientific knowledge network[J]. *Journal of Library Science in China*, 2015, 41(3): 99-113.)
- [46] Milo R, Shen-Orr S, Itzkovitz S, et al. Network motifs: simple building blocks of complex networks[J]. *Science*, 2002, 298(5594): 824-827.
- [47] 缪莉莉, 韩传峰, 刘亮, 等. 基于模体的科学家合作网络基元特征分析[J]. *科学学研究*, 2012, 30(10):

- 1468–1475. (Miao Lili, Han Chuanfeng, Liu Liang, et al. Using motif to characterize building block of scientific collaboration networks[J]. *Studies in Science of Science*, 2012, 30(10):1468–1475.)
- [48] Abbasi A, Hossain L, Leydesdorff L. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks[J]. *Journal of Informetrics*, 2012, 6(3):403–412.
- [49] Albert R, Barabási A L. Topology of evolving networks: local events and universality[J]. *Physical Review Letters*, 2000, 85(24):5234–5237.
- [50] Antal T, Kravivsky P L. Weight-driven growing networks[J]. *Physical Review E*, 2005, 71(2):026103.
- [51] Barrat A, Barthélemy M, Vespignani A. Weighted evolving networks: coupling topology and weight dynamics[J]. *Physical Review Letters*, 2004, 92(22):228701.
- [52] Barrat A, Barthélemy M, Pastor-Satorras R, et al. The architecture of complex weighted networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(11):3747–3752.
- [53] Pan Z, Li X, Wang X. Generalized local-world models for weighted networks[J]. *Physical Review E*, 2006, 73(5):056109.
- [54] Li M, Wu J, Wang D, et al. Evolving model of weighted networks inspired by scientific collaboration networks[J]. *Physica A: Statistical Mechanics and its Applications*, 2007, 375(1):355–364.
- [55] Rosvall M, Sneppen K. Reinforced communication and social navigation generate groups in model networks[J]. *Physical Review E*, 2009, 79(2):026111.
- [56] 张学义, 胡兴雨, 吴俊, 等. 基于兴趣的科研合作网络演化模型[J]. *计算机工程与应用*, 2010, 46(30):104–107, 111. (Zhang Xueyi, Hu Xingyu, Wu Jun, et al. Evolutionary model of interest-based scientific collaboration network[J]. *Computer Engineering and Applications*, 2010, 46(30):104–107, 111.)
- [57] Sun X, Dong J, Tang R, et al. Topological evolution of virtual social networks by modeling social activities[J]. *Physica A: Statistical Mechanics and its Applications*, 2015, 433:259–267.
- [58] Sun X, Lin H, Xu K. A social network model driven by events and interests[J]. *Expert Systems With Applications*, 2015, 42(9):4229–4238.
- [59] Sun X, Lin H, Xu K, et al. How we collaborate: characterizing, modeling and predicting scientific collaborations[J]. *Scientometrics*, 2015, 104(1):43–60.
- [60] 方锦清. 网络科学的理论模型探索及其进展[J]. *科技导报*, 2006, 24(12):67–72. (Fang Jinqing. Exploring theoretical model of network science and research progresses[J]. *Science & Technology Review*, 2006, 24(12):67–72.)
- [61] 陈柏彤, 张斌. 科学知识扩散研究框架[J]. *图书情报工作*, 2014, 58(15):48–57. (Chen Baitong, Zhang Bin. A research framework for scientific knowledge diffusion[J]. *Library and Information Service*, 2014, 58(15):48–57.)
- [62] Guimera R, Uzzi B, Spiro J, et al. Team assembly mechanisms determine collaboration network structure and team performance[J]. *Science*, 2005, 308(5722):697–702.
- [63] 夏昊翔, 王国秀, 宣照国, 等. 针对科研合作网络演化建模的基于 Agent 实验平台原型[J]. *情报学报*, 2010, 29(4):634–640. (Xia Haoxiang, Wang Guoxiu, Xuan Zhaoguo, et al. A prototype of agent-based modeling platform for studying the evolution of scientific collaboration network[J]. *Journal of the China Society for Scientific and Technical Information*, 2010, 29(4):634–640.)
- [64] Luo S, Du Y, Liu P, et al. A study on coevolutionary dynamics of knowledge diffusion and social network structure[J]. *Expert Systems with Applications*, 2015, 42(7):3619–3633.

- [65] Yang G Y, Hu Z L, Liu J G. Knowledge diffusion in the collaboration hypernetwork[J]. *Physica A: Statistical Mechanics and its Applications*, 2015, 419: 429–436.
- [66] Wang J P, Guo Q, Yang G Y, et al. Improved knowledge diffusion model based on the collaboration hypernetwork [J]. *Physica A: Statistical Mechanics and its Applications*, 2015, 428: 250–256.
- [67] Cowan R, Jonard N. Network structure and the diffusion of knowledge[J]. *Journal of Economic Dynamics and Control*, 2004, 28(8): 1557–1575.
- [68] 李盛庆, 蔡国永. 复杂网络领域科研合著网络演化及知识传播特点研究[J]. *现代图书情报技术*, 2013(5): 64–72. (Li Shengqing, Cai Guoyong. Study on network evolution and knowledge dissemination of scientific collaboration network in the field of complex networks[J]. *New Technology of Library and Information Service*, 2013(5): 64–72.)
- [69] Garfield E, Sher I H, Torpie R J. The use of citation data in writing the history of science[R]. Institute for Scientific Information Inc. Philadelphia PA, 1964.
- [70] Börner K, Maru J T, Goldstone R L. The simultaneous evolution of author and paper networks[J]. *Proceedings of the National Academy of Sciences*, 2004, 101(suppl 1): 5266–5273.
- [71] Liu Y, Rousseau R, Guns R. A layered framework to study collaboration as a form of knowledge sharing and diffusion[J]. *Journal of Informetrics*, 2013, 7(3): 651–664.
- [72] Li M, Fan Y, Chen J, et al. Weighted networks of scientific communication: the measurement and topological role of weight[J]. *Physica A: Statistical Mechanics and its Applications*, 2005, 350(2): 643–656.
- [73] Zhang P, Li M, Wu J, et al. The analysis and dissimilarity comparison of community structure[J]. *Physica A: Statistical Mechanics and its Applications*, 2006, 367: 577–585.
- [74] Lü L, Pan L, Zhou T, et al. Toward link predictability of complex networks[J]. *Proceedings of the National Academy of Sciences*, 2015, 112(8): 2325–2330.
- [75] Ahn M W, Jung W S. Accuracy test for link prediction in terms of similarity index: the case of WS and BA models [J]. *Physica A: Statistical Mechanics and its Applications*, 2015, 429: 177–183.
- [76] 王文强, 张千明. 链路预测的网络演化模型评价方法[J]. *电子科技大学学报*, 2011, 40(2): 174–179. (Wang Wenqiang, Zhang Qianming. New method of assessing network evolving models based on link prediction [J]. *Journal of University of Electronic Science and Technology of China*, 2011, 40(2): 174–179.)
- [77] Wang W Q, Zhang Q M, Zhou T. Evaluating network models: a likelihood analysis[J]. *Europhysics Letters*, 2012, 98(2): 28004.
- [78] 刘宏鲲, 吕琳媛, 周涛. 利用链路预测推断网络演化机制[J]. *中国科学: 物理学 力学 天文学*, 2011, 41(7): 816–823. (Liu Hongkun, Lü linyuan, Zhou Tao. Uncovering the network evolution mechanism by link prediction [J]. *SCIENTIA SINICA Physica, Mechanica & Astronomica*, 2011, 41(7): 816–823.)
- [79] Tomassini M, Luthi L. Empirical analysis of the evolution of a scientific collaboration network[J]. *Physica A: Statistical Mechanics and its Applications*, 2007, 385(2): 750–764.

张斌 武汉大学中国传统文化研究中心博士后、武汉大学国家文化发展研究院讲师。湖北 武汉 430072。

李亚婷 武汉大学信息资源研究中心博士研究生。湖北 武汉 430072。

(收稿日期: 2016-05-18; 修回日期: 2016-07-12)