# 科学知识网络的演化与动力①
## ——基于科学引证网络的分析

刘 向, 马费成
(武汉大学信息资源研究中心, 武汉 430072)

**摘要**: 研究科学知识演化的动力及规律是探究科学知识创造及发展过程的关键. 基于复杂网络的方法建立了科学知识的演化模型,揭示了知识演化的马太效应中潜隐的时间因素的作用,时间效应一定程度上平抑了度择优所导致的马太效应的负面影响. 模型通过引入度择优和时间优先连接以反映了科学知识的继承与更新过程,其中度择优机制保证对经典科学理论的继承,时间优先机制促使对新近知识的吸收. 数理分析与模拟结果表明度择优所体现的马太效应的作用是全局性的,时间效应所体现的后发优势的影响则是局部的.

**关键词**: 知识网络; 演化网络; 引证网络; 演化模型; 动力
**中图分类号**: G20; N94 **文献标识码**: A **文章编号**: 1007 - 9807(2012)01 - 0087 - 08

## 0 引 言

探究科学知识的创造及发展过程历来是信息与知识管理领域的重要课题. 引文网络(citation networks)[1],知识地图(knowledge map)[2-3],概念地图(concept map)[4],共现网络(con-occurrence networks)[5],知识超网络(knowledge supernetwork)[6]等分别在知识演化领域的不同方面做出了很多有益的探索,取得了良好的成效. 然而当前的研究对知识网络内在的演化机理的研究仍然相对较少,本文试图在这一方面做出一些探索.

研究科学知识演化的动力及规律是探讨知识演化问题的一个关键,而引文网络是研究知识演化问题的一个有效依托. 知识的发展与演化是一个抽象的过程,其结构和表现形式难以直接获得,只能通过引文网络或共现网络等进行可视化与直观化,从而间接的进行研究. SCI 创始人 Garfield 等[7]很早便意识到科学引文网络可以反映科学知识之间传承、发展的关系,并且尝试利用引文网络研究科学知识发展的历史、脉络和结构; Bernal, Price, Leak 和 Shryock 等学者也均表示对这一想法的认同, Garfield[8]针对几个领域的引文网络分析也证实了这一想法的有效性. 立足于统计物理学的角度, Price[1]对引文网络的连接度进行了细致分析,发现网络的连接度满足指数为 2.5 - 3.0 的幂率(power law)分布,而科学知识发展的累积优势(Cumulative Advantage)过程[9]是促成这种现象的根本原因, Price 将这一过程理解为马太效应(Matthew Effect)的作用. 马太效应在知识的时空分布上是普遍存在的, Simon[10], Bradford[11]等也分别发现文章的词频分布、期刊分布和作者分布的极限形式均为幂率函数,这一发现已成为情报学领域的核心规律之一. 幂率分布也称为无标度(scale-free)分布,对于这一分布形成过程的更明晰表述是 Barabási & Albert[12]做出的,他们通过引入增长(growth)和择优连接(preferential attachment)机制构建了无标度网络的演化模型(BA 模型),揭示了网络无标度特征形成

① 收稿日期: 2010 - 09 - 07; 修订日期: 2011 - 09 - 16.
基金项目: 国家自然科学基金重点资助项目(70833005).
作者简介: 马费成(1947—),男,教授,博士生导师,通讯作者. Email: fchma@whu.edu.cn

的内在机理. Newman[13] 比较了 Price 模型和 BA 模型,认为后者是前者的抽象和一般化,择优连接机制和优势积累过程本质意义上是一致的. 而实际的较大规模的统计分析却显示了一些差异,Redner[14] 通过统计 ISI( Institute for Scientific Information) 1981 年至 1997 年的 783 339 篇文献及 Physical Review D 上 20 年的 24 296 篇文献,发现引文网络的入度分布具有指数约为 3 的幂率尾( power-law tail) ,但是在较低连接度区间上则存在明显的对幂率分布的偏离,即前部有一段形似指数分布的下弯. 可见,除了马太效应之外,还有其他的因素影响科学知识网络演化的过程.

时间是影响引证的另一个因素. Price[1] 很早便指出在参考文献利用上,人们除了引用经典文献之外,也倾向于使用最新发表的文献,他以为约有 50%的引证与论文发表时间有关,30%的引证与近期发表的文献是强相关的,而这 30%之中约一半是对近 1 至 6 年发表的文章的引用,即使排除文献的指数增长影响,对近期文献的使用也会出现一个上升的趋势. 然而在 BA 模型中,通过度择优使得后续节点更倾向于连接老节点,越早加入的节点具有越高的度,对新加入的节点的连接则较少.

可见,时间对连接机制具有重要的影响作用. 本文对知识网络的分析也以引文网络为依据,尝试在科学知识网络的演化过程中考虑时间因素,构建科学知识网络的演化模型,分析网络的拓扑结构及演化特征,探讨知识演化的马太效应中潜隐的时间因素的影响,以及马太效应与时间效应之间的关系.

## 1 问题描述

在科学知识网络中,节点( vertex) 表示知识单元,由考察粒度的不同,可以取为图书、期刊论文、情报片段等,边( edge) 则代表知识单元之间的引证关系. 由于引证是时间先后次序上的单向连接,故科学知识网络是一个无环( acyclic) 的有向网络( directed networks) .

节点的出度( out-degree) 是指从该节点指向其他节点的边的数目,在引文网络中即为论文的参考文献数量; 节点的入度( in-degree) 指从其他节点指向该节点的边的数目,在引文网络中即为被引的次数. 引证过程能很好地体现群体性行为的特征,能揭示知识继承与发展的过程,基于 Redner 的统计结果的特征,本文重点讨论入度分布的动力学成因.

新知识产生的过程即表述为向科学知识网络中添加新节点的过程,新节点与已存节点之间的连接边则表示新知识对已存知识的继承、引用关系. 为了讨论的方便和模型的一般性,本文采用 Price 模型和 BA 模型中同样的方法,网络中每时间步增加一个节点.

对于新节点与哪些老节点建立连接边取决于连接的策略,本文中建立两种选择机制: 度优先连接机制和时间优先连接机制: 1) 度优先连接机制即新加入节点更倾向于连接高连接度的已存节点. 实践显示,科学文献被越多的资料引证,则此文献被读者发现的机率越大,从而再次被引证的机会也越大. 故本文认为在科学知识增长网络中,节点被连接的概率与此节点的入度成正比. 2) 时间因素反映在择优连接机制上,需要考虑现实中人们对知识的继承和发展行为特征: 人们偏向追逐研究前沿( research front) ,倾向于对最新研究成果的吸收和拓展. 在科学知识网络中即为新节点更易于连接最近加入的节点.

## 2 演化模型

### 2.1 假设条件

科学知识网络的构成及演化受诸多因素影响,为了模型研究重点的突出及讨论的一般性,作出如下假设:

1) 节点是同性质的,边也是同性质的;

2) 每时间步增加一个节点;

3) 节点的出度取定值;

4) 择优连接受节点入度和节点加入时间的影响;

条件 1) 保证科学知识网络的同质性、一致性,例如所有节点均代表期刊论文,边代表论文之

间的引证关系; 针对条件 2) ,在任意短的时间内,每步增加一个节点的假设是成立的; 条件 3) 取节点出度为确定值,即文献的参考文献数量平均值,Vazquez[15] 指出引文网络的出度服从指数分布,则平均值为入度分布的特征连接度; 条件 4) 忽略了其他影响条件,从科学知识网络的以往研究来看,上述两个条件具有很强的代表性,度优先连接概率取线性关系[16]; 因在信息空间中,人们对较近时间的信息反映敏感,而对于较远时间的信息则往往不加区别,故时间优先连接取为超线性关系.

### 2.2　模型构建

以增长和连接机制构建科学知识网络演化模型,将时间因素加入连接策略之中,模型的构造算法如下:

增长( growth) : 从一个具有 $m_0$ 个节点的网络开始,每次引入一个新的节点;

连接( attachment) : 新节点连接到 $m$ 个已存在的不同节点上( $m \leqslant m_0$) ,按如下概率选择节点进行连接操作:

1) 以概率 $p$ 连接新节点到一个已经存在的节点 $i$ 上,连接的概率 $P(i)$ 与节点 $i$ 的入度 $k_i$ 成正比( 度优先连接) ,即满足如下关系

$$P(i) = \frac{k_i + 1}{\sum_j (k_j + 1)}$$

2) 以概率 $q$ 连接新节点到一个已经存在的节点 $h$ 上,连接的概率 $P(h)$ 与节点 $h$ 的加入时间 $t_h$ 的 $\alpha$ 次方成正比( 时间优先连接) ,即满足如下关系

$$P(h) = \frac{t_h^{\alpha}}{\sum_l t_l^{\alpha}}$$

其中 $0 \leqslant p, q \leqslant 1; p + q = 1; \alpha > 1; i \cap h = \varnothing$. 度优先连接中采用 $k_i + 1$,以保证入度为 0 的节点的连接概率不为 0.

## 3　模型分析

节点 $i$ 在 $t$ 时间步的入度为 $k_i$,在时间步 $t$ 所有节点的入度和为 $\sum_i (k_i + 1) = t(m + 1) + m$,按照节点的入度选择优先连接的节点,则连接概率 $P(i) = \frac{k_i + 1}{\sum_j (k_j + 1)} = \frac{k_i + 1}{t(m + 1) + m}$. 每时间步只增加一个节点,不妨将时间步与节点号取等 $t_h = h$,由 $\sum_l t_l^{\alpha} = \sum_l l^{\alpha} \approx \int_1^t l^{\alpha} \mathrm{d}l \approx \frac{1}{\alpha + 1} t^{\alpha+1}$,故有 $P(h) = \frac{t_h^{\alpha}}{\sum_l t_l^{\alpha}} \approx \frac{(\alpha + 1) h^{\alpha}}{t^{\alpha+1}}$. 采用均场方法( mean-field approach) [17] 求解,上述模型的动力方程为

$$\frac{\partial k_i}{\partial t} = m\left(p \frac{k_i + 1}{\sum_j (k_j + 1)} + q \frac{t_i^{\alpha}}{\sum_l t_l^{\alpha}}\right) = m\left(p \frac{k_i + 1}{t(m + 1) + m_0} + q \frac{(\alpha + 1) i^{\alpha}}{t^{\alpha+1}}\right) \quad (1)$$

对于大 $t$ 有

$$\frac{\partial k_i}{\partial t} = mp \frac{k_i + 1}{t(m + 1) + m_0} \quad (2)$$

初始条件为 $k_i(i) = 0$,求解得 $i$ 节点在 $t$ 时刻的度

$$k_i = -1 + \left((t + \frac{m_0}{m + 1}) \Big/ (i + \frac{m_0}{m + 1})\right)^{\beta} \quad (3)$$

指数 $\beta = pm/(m + 1)$ ,认为 $i$ 为 $t$ 上的均匀分布,密度函数为 $\rho(i) = 1/(t + m_0)$ ,则有概率函数

$$P\{k_i(t) < k\} = P\left\{i > \left(t + \frac{m_0}{m + 1}\right)(k + 1)^{-\frac{m+1}{mp}} - \frac{m_0}{m + 1}\right\} = 1 - \frac{1}{t + m_0}\left(t + \frac{m_0}{m + 1}\right) \times (k + 1)^{-\frac{m+1}{mp}} + \frac{m_0}{(t + m_0)(m + 1)} \quad (4)$$

求解密度函数

$$p(k) = \frac{\partial P\{k_i(t) < k\}}{\partial k} = \frac{m + 1}{mp}(k + 1)^{-\gamma} \quad (5)$$

其中 $\gamma = 1 + (m + 1)/mp$. 解析结果显示,单纯入度驱动,即取 $p = 1$ 的增长网络是 $\gamma \approx 2$ 的无标度网络,而当 $p = (m + 1)/2m$ 时,$\gamma = 3$; 而当 $p \in [(m + 1)/2m,\ (m + 1)/1.5m]$ 时,$\gamma \in [2.5, 3.0]$.

上述结果是在 $t$ 为大值时的解析解,式( 2) 中忽略了无穷小 $(\alpha + 1) i^{\alpha}/t^{\alpha+1}$,下面单独对时间优先连接机制进行分析,即取 $q = 1, p = 0$,仅考虑

时间因素，则有

$$\frac{\partial k_i}{\partial t} = m\frac{(\alpha+1)\,i^{\alpha}}{t^{\alpha+1}} \tag{6}$$

解得

$$k_i = \frac{m(\alpha+1)}{\alpha}\left[1-\left(\frac{i}{t}\right)^{\alpha}\right] \tag{7}$$

固定 $t$，则 $k_i$ 随 $i$ 单调减少，当 $t\to\infty$ 时，有 $k_i = m(\alpha+1)/\alpha$. 同样求解

$$P\{k_i(t) < k\} = P\left\{i > t\left(1-\frac{\alpha k}{m(\alpha+1)}\right)^{1/\alpha}\right\} = 1-\left(1-\frac{\alpha k}{m(\alpha+1)}\right)^{1/\alpha} \tag{8}$$

有 $1-\alpha k/m(\alpha+1)\geqslant 0$，即 $k\in[0, m(\alpha+1)/\alpha]$，计算密度函数

$$p(k) = \frac{\partial P\{k_i(t) < k\}}{\partial k} = \frac{1}{m(\alpha+1)}\left(1-\frac{\alpha k}{m(\alpha+1)}\right)^{-\gamma} \tag{9}$$

其中 $\gamma = 1-1/\alpha$. 均场方法在低连接度区域会出现较大的偏差，上述密度函数仅作为参考，对于大致估计度区间起一个辅助作用.

由上可见，在大时间范围上，演化模型遵循幂率分布，且当度优先连接概率 $p=(m+1)/2m$ 时，形成指数 $\gamma=3$ 的无标度网络. 而对时间优先连接机制而言，连接度收敛于确定值，即 $k\to m(\alpha+1)/\alpha(t\to\infty)$，在整个区间上是近似均匀的，对整体连接度的影响作用有限.

分别对 $p=1$ 和 $q=1$ 的情况进行求解，如下图 1 所示. 其中，左上、左下图表示入度分布，右上、右下图表示最后加入的1 000个节点所连接节点在时间上的分布，其中 $m_0=10$，$m=9$，运算次数均为 $N=20\,000$. 可见，单纯入度驱动指数为 $\gamma\approx 2.1$ 的无标度网络，越早加入的节点入度越高，而时间优先网络则存在特征标度，偏向连接新近加入的节点.

度优先策略倾向于连接较早产生的节点，后加入节点的连接度较低，时间优先连接策略则对较晚产生的节点的选择较多，而相对度优先策略来说，最近时间连接对于低连接度区域产生较大影响，即对后加入节点影响较大，而对高连接度区域的影响则比较小.

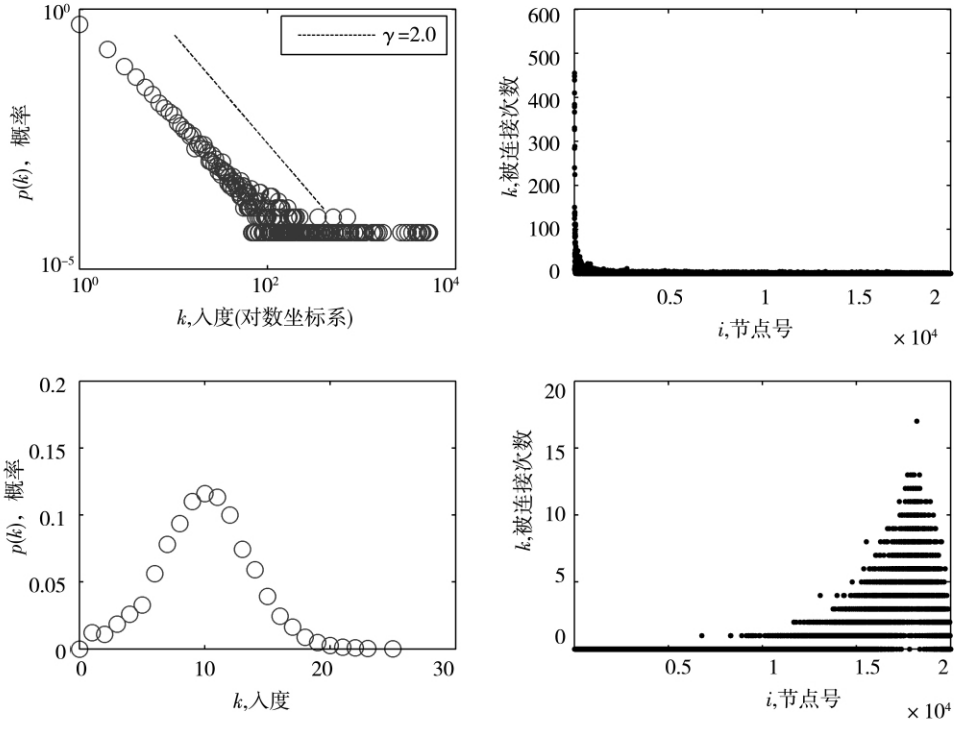


**图 1　左上、右上图是仅采用度优先连接($p=1, q=0$) 时的统计图，左下、右下图是仅采用时间优先连接($p=0, q=1, \alpha=10$) 时的统计图**

Fig. 1 Top left/right graph is the statistics chart of degree preferential attachment when $p=1, q=0$. bottom left/right is the statistics chart of time preferential attachment when $p=0$, $q=1, \alpha=10$

## 4　仿真实验

对演化模型采用编程实现,主要考察两个方面的问题: 其一是演化模型的度分布; 其二是考察加入了时间优先机制,其作用力如何?演化网络中是否仍然是先加入节点连接度越高; 最后是截取一定时间段节点,考察它们所连接的节点在整个时间区间上的分布. 具体过程与结果如下:

1) 实验数据

采集了 ISI 上 SCIE( 1998 年至 2010 年 12 月) 数据库中 CELL 及其系列子刊( 统计了 9 种) 的文献题录. 通过精炼子集方式仅包含 ARTICLE、PROCEEDINGS PAPER 两种文献类型 21 992 条数据,排除了 EDITORIAL MATERIAL、REVIEW 等类型. 这是由于前两种类型能好的体现知识的创新和继承关系,且文献在写作模式上具有很好的一致性; 而 REVIEW 等属于总结、概括性的文献,与前述几种具有较大的差别,所以本文中没有对其进行考虑,而仅选择了前两种. (1) 数据去重和一致性处理,为保持网络数据类型的一致性,对数据的参考文献进行过滤,排除图书、网址等非上述两种主要类别的其它文献类型的影响. 得到以下数据: 总记录数 19 698,总文献数373 697,平均参考文献数量 18.97,时间跨度 13 年( 1998 – 2010) ,被引文献时间跨度 343 年( 1667 – 2010) ,被引文献中 1990 年前发表数据占总体的 12.73%. (2) 子网络提取. 由于1998 年12 月以前的文献的参考文献我们不能获得,所以这一时间以前所产生的连接数据不能得到,而本文中所构建模型可以统计到所有节点的出 / 入度,为了与后续模型进行比较,排除 1998 年 12 月以前的节点,仅提取1998 年12 月至2010 年12 月所生成的节点的连接子图,共计节点数 17 340,边数 52 168,平均连接数 $< k > = 3$.

统计两个方面的情况: 一是实验数据的入度分布; 二是截取2010 年发表的1 831 篇文献,统计其所引用文献的年代分布. 如图2 所示: (1) 入度分布具有幂率尾,前端向下弯曲,此点与 Redner 的统计分布图[14] 在直观上是一致的; (2) 出度分布弯曲弧度明显, 为指数分布, 这一点和 Vazquez[15] 的结果是一致的,具有特征连接度可以确定. (3) 右图显示2010 年论文对近期文献的利用较多,快速上升,这与 Price[1] 的统计也是一致的.

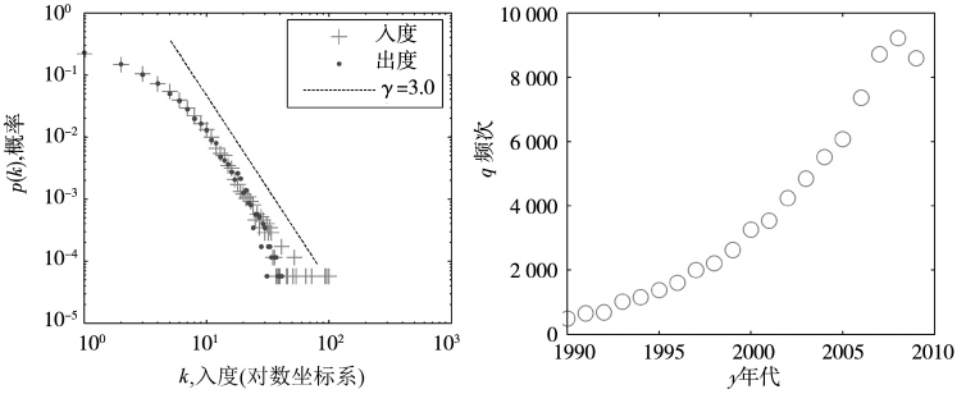


**图 2　左图是度分布图,包括入度分布和出度分布; 右图是 2010 年发表的论文的参考文献的年代分布,截取的是 1990 – 2010 段**

Fig. 2 The left graph shows the degree distribution of vertexes, including in-degree and out-degree; the right graph shows the publishing time of the references of the article in 2010

2) 仿真分析

分别取两组数据进行仿真. 第一组: 平均连接数为 $< k > = 18.97$,取整数 $m_1 = 19$ 进行实验,幂指数取 3,则度优先连接的概率应为 $p_1 = (m_1 + 1)/2m_1 = 0.526$,则 $q_1 = 0.474$,$\alpha = 2$ 进行测试. 图 3 是第一组数据的测试结果,运算次数为 $N = 50\,000$ 次,其中,中图和右图中 $i$ 既可表示时间步,又可表示节点号.

第二组: 在子图中平均连接数为 $< k > = 3$,取 $m_2 = 3$,故有 $p_2 = (m_2 + 1)/2m_2 = 0.667$ 和 $q_2 = 0.333$,$\alpha = 10$ 进行测试.

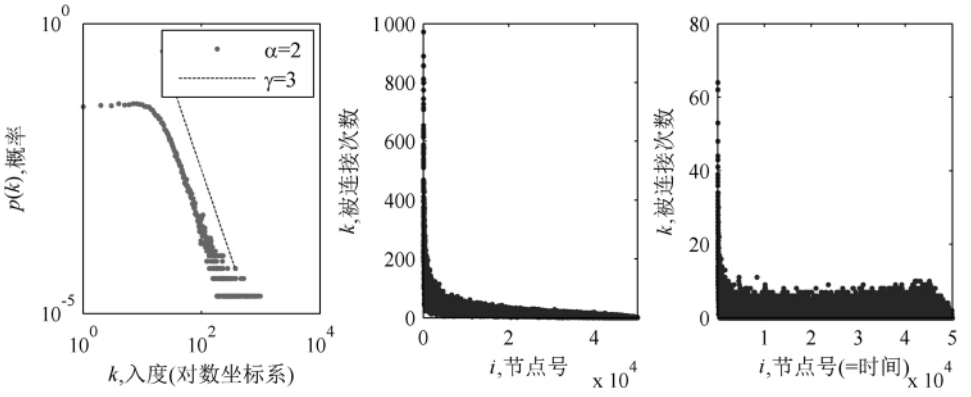


**图 3　左图是节点的度分布图,中图是节点(按加入时间排列)的被连接次数散点图,右图是最后加入的 5 000 个节点所连接节点在时间上的分布图**

Fig. 3 The left graph shows the degree distribution of vertexes, the centre shows the scatter diagram of attached times of total vertexes. The right shows the attached times of the vertexes attached by the vertexes that added the last 5 000 steps

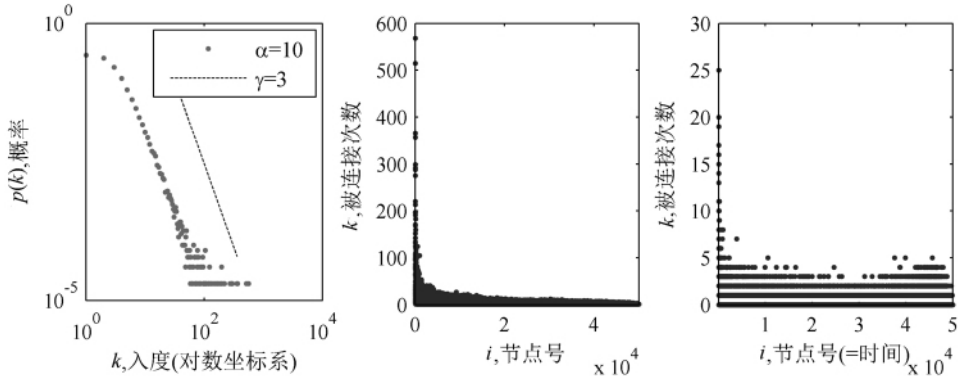


**图 4　左图是节点的度分布图,中图是节点(按加入时间排列)的被连接次数散点图,右图是最后加入的 5 000 个节点所连接节点在时间上的分布图**

Fig. 4 The left graph shows the degree distribution of vertexes, the centre shows the scatter diagram of attached times of total vertexes. The right shows the attached times of the vertexes attached by the vertexes that added the last 5 000 steps

如图3、4 所示,度分布有幂率尾,指数 $\gamma = 3$,与前文要求是一致的,模拟结果在分布曲线前端对幂函数产生了一定偏离,呈现下弯形态,且下弯幅度随着 $\alpha$ 值增大,度分布曲线前部下弯、时间分布后部翘尾的幅度均增大,可见,时间优先连接机制明显降低了低连接度节点的比例; 对于总体节点(如中图所示),显示先加入节点连接度越高,仍然保持幂率分布的图形特征.

对应图 2(中),观察图 3、4 后部形态: 仿真中截取了[45 000, 50 000]上生成的节点对前续节点的连接情况,对于最后加入的5 000 个节点所连接节点在时间上的分布后部上翘,图 3、4(右) 前部和后部连接较多,中部分布较为均匀. 图 3、4(左) 头部,图 3、4(右) 后部,弯曲的幅度均随 $\alpha$ 的增大而增大,这与时间优先连接机制的作用强度是相关的.

3) 实验结果

演化模型度分布模拟结果与实际统计结果在形态上基本相似,但也存在一些差异,主要是前端弯曲区段较小,而实际统计结果中前端区域则较长. 需要指出的是,演化模型中前端弯曲程度与参数 $\alpha$ 的取值有关,当取值较大时,下弯幅度越大,取值较小时则会平缓下弯.

关于时间分布上的尾部上翘,图 2(中) 是一段截取的时间区间,只有尾部形态,这与图 3(右) 的尾部是一致的. 而对于其它区段的特征,Price[1] 指出人们对经典文献和近期文献引用较多,而其他的部分的引用则是随机的,高连接度节点对应科学知识网络中的奠基性经典文献,而最后加入的节点则对应研究前沿. 这一说法与图 3、

4( 右) 在图形形态上基本是一致的,前部的高连接度节点受到持续关注,后部的新文献也引用较多,其它部分则是较为平缓的随机均匀分布.

图3、4( 中) 显示知识演化模型仍然具有幂率分布的图形特征,可见时间优先连接机制的影响是局部的、区段性的,它的加入没有改变曲线总体的度择优特征. 即有,优势积累效应的作用是全局性的,而时间优先连接机制的作用是局部的,时间优先只能促成短时的连接快速上升,对全局的幂率分布影响不大.

## 5　讨　论

度择优保证了对重要的高连接度节点的持续关注,而时间优先机制则显示了对新近节点的热衷,本文综合此两机制构建了择优和时间优先混合作用的科学知识网络演化模型,以下就文中存在的几点问题作一些讨论:

1) 科学知识网络连接机制的影响因素是多方面的,很明显的一个方面是随机连接,Barabási 和 Albert 很早便证明了随机连接机制作用形成的增长网络度分布为指数函数[11],即当 $\partial k_i/\partial t = m/(m_0 + t)$,得到 $p(k) \propto e^{-k/m}$. 倘若如前文的分析要保证指数为 3 的幂率尾,则需 $p = (m + 1)/2m > 0.5$,而根据文献[1],又 $q > 0.3$,故其它影响因素的作用程度大致在 0.1 左右,相对度择优和时间优先连接来说,作用效果较小.

2) 度择优导致了科学知识网络中马太效应现象的出现. 马太效应虽然为知识学习提供了方便,但是妨碍了知识的更新换代和新知识的脱颖而出,如何平抑马太效应的不利影响一直是情报学家关心的一个重要课题. 但是考虑时间因素的影响,人们在知识利用中不但倾向于利用经典知识,对新知识也赋予了很大关注,从而自发的克服了马太效应的影响. 且由于时间择优本身形成一个有特征标度的分布,见如图 1( 左下) ,对每一知识点的连接数是均匀的 $k \to m(\alpha + 1)/\alpha(t \to \infty)$,所以这也在一定程度上体现了所有知识在最初产生时在人们看来是平等的,而不仅仅是马太效应所体现的差别对待.

3) 本文的时间优先连接造成了一种后发优势,其形成的结果与现实的统计在直观上是一致的. 由模型分析可见,时间优先连接机制不会改变连接度的幂率分布整体特征,其作用是局部的. 当然要满足特定的幂率分布指数 $\gamma$,时间效应的作用概率 $q$ 就是一定的了. 改变 $p$ 和 $q$ 值,$\gamma$ 可以取到 [2.5,3.0] 区间内的任意数值.

4) 时间优先连接的作用强度. 对于 $\alpha$ 的取值问题,需结合实际的问题,例如电子、纳米等学科更重视最新文献,$\alpha$ 宜取较大值,而物理、化学、数学等学科则更重视经典文献资料,$\alpha$ 则可取较小值.

5) 出度对节点的连接度也是有影响的,比如涉及大量参考文献的综述类论文,被参考引证的次数通常也较多. 这一点在本文中没有作重点考虑,这是因为: 其一,科学文献的出度与度连接增长之间没有直接的联系,其影响程度弱于入度的直接驱动; 其次,本文演化模型中节点的出度取的是定值,考虑进去没有太大意义. 不过,出度的影响也可以作为后续的一个探讨主题.

## 6　结束语

本文着重探讨了知识演化过程中的时间因素影响,基于演化网络的方法与理论,构造了科学知识网络演化模型,分析了马太效应与时间效应之间的关系. 演化模型引入了度择优和时间优先两种连接机制,其中度优先连接机制保证了对重要知识的连接,而时间优先连接机制则促成对最新知识的接受和知识的更新,两种机制的结合形成了知识演化在研究基础与研究前沿之间的平衡. 模拟结果显示,度择优的作用是全局性的,而时间优先连接机制的作用则是局部的,它只能促成连接数的短时快速上升,不能改变全局的大趋势.

文中对于时间优先连接取为超线性关系,其实指数关系、对数关系也是可以考虑的,这是下一步研究的一个方向. 此外,本文构建的演化网络模型是一个一般化模型,对论文图书引文网络、专利网络等均具有一定的解释作用,然而其与实际网络仍然有一些不同,比如聚类系数、特征路径长度等均有实际网络存在一定差异,这与实际网络的层次结构有关,涉及更细致的拓扑建模,本文中没有作为重点,这也是下一步的研究方向之一.

**参 考 文 献:**

[1] Price D J. Networks of scientific papers [J]. Science, 1965, 149: 510 – 515.
[2] Brookes B C. The foundations of information science: Part I. Philosophical aspects [J]. Journal of Information Science, 1980, 2: 125 – 133.
[3] Borner K, Mane K K. Mapping topics and topic bursts in PNAS [J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(1): 5287 – 5290.
[4] 马费成, 郝金星. 概念地图在知识表示和知识评价中的应用(Ⅰ)——概念地图的基本内涵 [J]. 中国图书馆学报, 2006,32(3): 5 – 9.
Ma Feicheng, Hao Jinxing. Applications of concept maps in knowledge representation and knowledge evaluation (Ⅰ): Fundament of concept network [J]. The Journal of The Library Science in China, 2006,32(3): 5 – 9. (in Chinese)
[5] 王晓光. 科学知识网络的形成与演化(Ⅰ): 共词网络方法的提出 [J]. 情报学报, 2009, 28(4): 599 – 605.
Wang Xiaoguang. Formation and evolution of science knowledge network(Ⅰ): A new research method based on co-word network [J]. Journal of the China Society for Scientific and Technical Information, 2009, 28(4): 599 – 605. (in Chinese)
[6] 席运江,党延忠,廖开际. 组织知识系统的知识超网络模型及应用 [J]. 管理科学学报,2009, 12(3): 12 – 21.
Xi Yunjiang, Dang Yanzhong, Liao Kaiji. Knowledge supernetwork model and its application in organizational knowledge systems [J]. Journal of Management Sciences in China, 2009,12(3): 12 – 21. (in Chinese)
[7] Garfield E. Citation indexes for science [J]. Science, 1955, 122: 108 – 111.
[8] Garfield E. Citation Indexing: Its Theory and Application in Science, Technology, and Humanities [M]. Philadelphia: ISI Press, 1983: 69 – 123.
[9] Price D J. A general theory of bibliometric and other cumulative advantage processes [J]. J. Amer. Soc. Inform. Sci. 1976, 27: 292 – 306.
[10] Simon H A. On a class of skew distribution functions [J]. Biometrika, 1955, 42: 425 – 440.
[11] Bradford S C. Sources of information on specific subjects [J]. Engineering, 1934, 137: 85 – 86.
[12] Barabási A L, Albert R. Emergence of scaling in random networks [J]. Science, 1999, 286: 509 – 512.
[13] Newman M E. The structure and function of complex networks [J]. SIAM REVIEW, 2003, 45(2): 167 – 256.
[14] Redner S. How popular is your paper? An empirical study of the citation distribution [J]. Eur. Phys. J. B., 1998, 4: 131 – 134.
[15] Vazquez A. Statistics of citation networks [EB/OL]. Arxiv preprint cond-mat/010503, 2001.
[16] Jeong H, Nda Z, Barabási A L. Measuring preferential attachment in evolving networks [J]. Europhys. Lett., 2003, 61: 567 – 572.
[17] Barabási A L, Albert R, Jeong H. Mean-field theory for scale-free random networks [J]. Physica A, 1999, 272: 173 – 187.

# Evolution and dynamics of scientific knowledge network: Based on the study of scientific citation network

*LIU Xiang, MA Fei-cheng*
Center for Studies of Information Resources of Wuhan University, Wuhan 430072, China

**Abstract**: Research on the evolution and dynamics is the key point to explore the processes of the creation and development of scientific knowledge. By using the method of complex networks, we constructed a evolving model which revealed the effect of the time factor on the evolution of knowledge and the relationship between it and Matthew Effect, and discovered that the time effect decreases the adverseness of Matthew Effect. The model introduced both degree and time preferential attachments to describe the accepting and updating processes of scientific knowledge. The degree preferential attachment ensures the acceptance of classic theory, and the time preferential attachment encourages the absorption of recent knowledge. Simulation results showed that the effect of preferential attachment, which produced Matthew Effect, is overarching; and the effect of time attachment is partial.
**Key words**: knowledge networks; evolution networks; citation network; evolution model; dynamics