

# 基于科研合作网络演化的 个体网络生命过程分析\*

□ 王丁弘 赵奉英 吴斌 / 北京邮电大学智能通信软件与多媒体北京市重点实验室 北京 100876

**摘要:** 将复杂网络的研究方法引入到科研合作网络的研究中,为分析和评价科研人员提供了一个新途径。文章以合作网络为背景,微观深入地研究了科研合作网络中个体成员的网络生命特性,从全局网络和局部社团演化的角度定量分析个体网络生命过程对网络发展的影响。文章以大量的科技文献数据为实验数据集,以网络演化为线索跟踪个体成员的生命过程,同时考虑网络演化中的普通成员和核心成员对社团演化的不同影响,定量分析演化特性和个体网络生命过程。数据分析证实了科研团队要持续不断地发展,既要不断吸纳新成员为科研团队注入新活力,同时又要要有相对稳定的中坚力量维持着团队的科研方向。更进一步地,文章的研究方法可以扩展到对其他社会组织分析,追踪分析相关组织的发展趋势及关键人物对组织演化的影响。

**关键字:** 科研合作网,个人网络生命过程,个体活跃度

DOI: 10.3772/j.issn.1673-2286.2010.08.004

## 1 引言

近些年来,复杂网络越来越受到科研人员的关注,科研合作网作为一种典型的社会网络,最初对它的探究只是针对网络的宏观静态统计特征及网络拓扑结构。随着研究的深入,微观的社团演化特征以及对个体科研工作者的影响力的研究越来越备受关注。与此同时,将科研合作网引入各种学术搜索服务的应用也越来越多。近来,基于不同的角度和不同的学科构建科研合作网、论文引用网,结合网络自身的结构特征及科研合作中论文引用的各种统计特征的科研合作网络的研究,为分析科技文献数据提供了新的方法。

本文主要从科研合作网络的微观特性出发,在传统的科技文献数据分析的基础上,基于时间演化与微观分析的角度对科研合作网络、科研合作团队及个体科研工作者的成长过程进行量化分析。追踪科研合作团队演化路径,通过给定的社团演化特征量化方法分析了特定科研工作者在科研合作过程中的生命特性。文中引入了新的演化关系确立原则:前后社团之间的

普通成员和核心成员的双重重叠度来确定前后社团是否存在演化关系,即在两个连续的时间片内,社团成员存在一定的稳定性才认为社团间存在演化关系。这种方法与以往的只考虑了核心成员的演化路径分析相比更符合实际。同时文中分别基于全局网络生成和社团演化的角度分析了个体的网络生命过程特性。通过成员活跃时长及成员在网络演化过程中所处的地位变化,量化分析了个体成长与网络演化的相互影响。需要说明的是,本文中涉及的个体生命概念相对而言是个体在合作网中表现的生命现象,也可以说是个体的学术网络生命特征,合作网可以从一个侧面反映个体科研和学术活动情况,本文仅从这个角度引申出个体生命过程概念,会与个体实际科学研究过程存在一定的偏差,但从实验分析看对这一概念的探索性研究是有意义的,也具有一定的可行性。

本文着重关注网络社团演化中不同的个体成员在社团演化中的作用。基本结构如下:第2节是对相关课题的介绍及本文的研究特点;第3节定义了文中所用到的相关符号及公式说明;第4节介绍了个体成长过程实验分析及结果讨论;第5节是对全文的总结。

\* 本文得到国家自然科学基金项目(90924029)和国家“十一五”科技支撑计划项目(2006BAH03B05)资助。

## 2 相关工作

近年来,随着复杂网络研究的纵向研究的深入,越来越多的学者开始关注科研合作网络的复杂网络模型的分析研究,同时取得了较多的科研成果。着手于科研合作网的微观研究,文中将目前科研合作网中个体的研究分为三个层面进行阐述:一是个体对全局性的网络生成模型的影响,具体体现在个体活跃度的分析上。二是个体对局部的网络社团演化的影响。三是科研活动中的合作关系的预测。其中对于第一个层面研究具有代表性的是Palla等人在Barabasi的BA模型<sup>[1]</sup>基础上所作的改进。改进体现在分析研究科研合作网的生成因素时,优先链接不仅存在于新增节点与网络中已有节点,而且还存在网络中已有节点间的链接<sup>[2]</sup>。对于第二个层面研究的典型代表是康奈尔大学的Lars Backstrom等人对DBLP的科技文献数据的研究。Theodoros Lappas等人则根据现有的合作网络找到符合一定条件的专家<sup>[3]</sup>。针对第三个层面研究的代表是美国宾州州立大学Jian Huang等人提出的随机泊松分布-最优决策树模型<sup>[4]</sup>(Stochastic Poisson model with Optimization Tree),它根据当前科研合作者的邻居节点来预测任意两个节点之间在未来的时间里再次合作的可能性,即链接预测<sup>[5]</sup>,并在计算机领域中六个不同的数据集上进行了验证。

分析网络个体成员时,通常按其其在网络中的重要程度的不同,将他们分为普通成员和核心成员,通过对比来分析两种成员在网络演化中的影响力。清华大学的唐杰等人通过分析网络不同成员间的相互作用,提出了更为广泛性的个体研究模型:基于话题的个人影响力传播模型——TAP(Typical Affinity Propagation)<sup>[6]</sup>。此模型专门对科研合作网络中个人影响力进行量化。模型把个人影响力的量化分为三个步骤:首先抽象出网络图模型 $G=(V,E)$ 。其次利用ACT模型(Author-Conference-Topic模型)找出科研合作网络中所涉及的话题/主题,并根据作者间的论文发表关系,得出话题在科研工作者的分布。最后综合前两步得到科研合作网络图及话题分布函数,得出了基于某一特定话题的科研合作网中作者间相互影响的关系图。

本文采用了基于科研合作网的社团演化分析,引入了基于核心成员与普通成员双重重合度的度量方法来跟踪社团的演化,以成员活跃时长和成员变化率来考察个体成员的生命过程对社团演化的影响。分析实

验结果表明:社团核心成员的平均活跃时间要远远大于其新生成员的活跃时间,这也为核心成员为社团的发展带来了充分的时间,而新成员可能只是在一段特别短暂的时间内参与到科研活动中,而后就再也没有进行科研活动,那么这类成员对社团的发展所起到的作用自然不能与核心成员相提并论。本文基于个体生命特性的不同将节点分为普通成员、核心成员、新成员,进而分析不同个体对社团演化的影响,并着重对基于个体不同的生命阶段在全局网络和局部社团网络的表现定量研究。将个体的影响力分析贯穿到个体的整个生命过程,即出生阶段、成长壮大阶段、衰落阶段。实验基于DBLP数据集进行量化分析,展现了典型个体研究工作者整个生命周期及其主要特征。

## 3 符号及公式定义

### 3.1 符号定义

表1 符号定义

符号名称	含义
$CN^{(t)}$	t时刻所形成的科研合作网络
$A^{(t)}$	$CN^{(t)}$ 中节点的集合,即:t时刻科研合作网络中作者/科研工作者的集合
$L^{(t)}$	$CN^{(t)}$ 中边/合作关系的集合,即:t时刻科研合作网中科研合作关系的集合
$\alpha_i$	作者/科研工作者 $\alpha_i$
$Cen(\alpha_i)$	作者的中心性 $\alpha_i$
$l_{ij}$	$\{(\alpha_i, \alpha_j)   \alpha_i \in A, \alpha_j \in A \wedge \alpha_i, \alpha_j \text{ 至少合作写过一篇论文}\}$
$c_i^{(t)}$	t时刻的社团
$c_i^{(t)} \rightarrow c_i^{(t+1)}$	$c_i^{(t)}$ 是 $c_i^{(t+1)}$ 的前驱,或 $c_i^{(t+1)}$ 是 $c_i^{(t)}$ 的直接后继
$c_i^{(t)} \Rightarrow c_i^{(t+1)}$	$c_i^{(t)}$ 是 $c_i^{(t+1)}$ 的祖先,或 $c_i^{(t+1)}$ 是 $c_i^{(t)}$ 的后代
$ c_i^{(t)} $	社团 $c_i$ 在t时刻的规模
$Edge(c_i^{(t)})$	社团 $c_i$ 在t时刻边的集合
$Mem(c_i^{(t)})$	社团 $c_i$ 在t时刻时的成员集合
$Core(c_i^{(t)})$	社团 $c_i$ 在t时刻时的核心成员的集合

### 3.2 公式定义与说明

#### (1) 个体活跃度 (Author Activity):

在某一具体数据集中, 对于任意的作者  $\alpha_i$ , 即:  
 $\forall \alpha_i \in A$ , 定义  $\alpha_i$  发表第一篇论文的时间为其出生时间, 表示为  $\text{Birth}(\alpha_i)$ 。类似地,  $\text{Death}(\alpha_i)$  表示为作者  $\alpha_i$  的衰落时间, 即在当前的数据集中作者  $\alpha_i$  最后发表论文的时间。那么对于任意的时刻  $t$ , 且  $t \in [\text{Birth}(\alpha_i), \text{Death}(\alpha_i)]$ , 如果作者  $\alpha_i$  在  $t$  时刻发表了一篇论文  $p$ , 那么就认为论文合作事件发生了, 而论文的作者对于该论文  $p$  的贡献记作  $\text{Contr}(p, \alpha_i)$ , 此时用作者的活跃度  $\text{ActVal}(\alpha_i^{(t)})$  来表示相对于其最近的一个时间片  $t_j$  作者的活跃情况。这样就可以表示出作者  $\alpha_i$  在其整个生命周期内的活跃情况,  $\text{ActVal}(\alpha_i^{(t)})$  的具体计算方法为:

$$\text{ActVal}(\alpha_i^{(t)}) = \sum_j \text{Contr}(p_j, \alpha_i) \times \exp(-\lambda |t - t_j| \text{Contr}(p_j, \alpha_i)) \quad \text{公式 (3-1)}$$

公式说明:

1)  $t_j = \min\{ |t - t'|, t' < t \}$ , 即:  $t_j$  为在时间片  $t$  之前, 且离其最近的时间片。在此时间片作者  $\alpha_i$  有论文发表。

2)  $\text{Contr}(p_j, \alpha_i)$ : 是指在  $t_j$  时刻发表的论文  $p_j$ , 作者  $\alpha_i$  的贡献值。特别地, 如果论文  $p_j$  有  $n$  个合作者, 那么其中任意一个合作者对该论文的贡献值记作  $1/n$ 。

3)  $\lambda$ : 根据不同的科研合作网络而不同的数据常量。

很显然, 如果作者  $\alpha_i$  在一个时间片之后不再有论文发表, 则其活跃度是随着时间的增长而递减的, 直到其又有新的论文发表。

## (2) 社团成员稳定率 (Member Stability of Community)

$\forall c_i, t_0 = \text{Birth}(c_i), t_k = \text{Death}(c_i), \text{Mem}(C_i, t_j) = \{ \cup_{m \in \text{Mem}(c_i^k)} \text{Mem}(c_i^k) | c_i \Rightarrow c_i^k \}$ , 社团  $c_i$  的成员稳定率记作  $\text{MS}(c_i)$ :

$$\text{MS}(c_i) = \frac{|\text{Mem}(c_i^k) \cap (\text{Mem}(c_i^k) \cup \dots \cup \text{Mem}(c_i^k))|}{|\text{Mem}(c_i^k) \cup \dots \cup \text{Mem}(c_i^k)|} \quad \text{公式 (3-2)}$$

类似地,  $\text{Core}(C_i^k) \{ \cup_{m \in \text{Core}(c_i^k)} \text{Core}(c_i^k) | c_i \Rightarrow c_i^k \}$ , 社团  $c_i$  核心成员的稳定性可表示为  $\text{CS}(c_i)$ 。

$$\text{CS}(c_i) = \frac{|\text{Core}(c_i^k) \cap (\text{Core}(c_i^k) \cup \dots \cup \text{Core}(c_i^k))|}{|\text{Core}(c_i^k) \cup \dots \cup \text{Core}(c_i^k)|} \quad \text{公式 (3-3)}$$

## (3) 个体稳定性 (Activity Stability of Author):

根据社团的演化路径, 可以跟踪作者  $\alpha_i$  在其生命周期内所参与的所有社团, 从而可以得知其科研活动范围的变化情况, 用  $\text{AS}(\alpha_i)$  表示任意两个相邻时间内作者  $\alpha_i$  所参与社团情况的变化率。其计算公式为:

$$\text{AS}(\alpha_i) = \min\left(\frac{|C^{(0)} \cap C^{(t+1)}|}{|C^{(0)}|}, \frac{|C^{(0)} \cap C^{(t+1)}|}{|C^{(t+1)}|}\right) \quad \text{公式 (3-4)}$$

公式说明:

其中  $t \in [\text{Birth}(\alpha_i), \text{Death}(\alpha_i)]$ , 且  $C^{(0)}$  和  $C^{(t+1)}$  分别表示作者  $\alpha_i$  在  $t$  和  $t+1$  时刻所参与到的科研社团的集合,  $|C^{(0)} \cap C^{(t+1)}| = |c_i^{(0)}|$ , 其中  $c_i^{(0)} \in C^{(0)}$ , 且  $c_i^{(t+1)} \in C^{(t+1)}$  使得  $c_i^{(0)} \rightarrow c_i^{(t+1)}$  成立。

## (4) 作者核心率 (Core Ratio of Author):

通过跟踪科研工作者所参与到的社团, 不难发现, 在其科研活动初始时其可能不属于任何一个社团, 而随着其科研工作的深入与开展, 他们会逐渐加入到特定的社团中, 从而也就确立了其大致的科研方向, 而在此后的研究活动中可能成为社团的核心成员。为了表示科研工作者在其生命周期内与社团之间的关系, 用核心率来表示, 即: 在任意的时间片  $t$  内, 作者  $\alpha_i$  的核心率可表示为以他/她为核心的社团数量与其所参与的社团数量之比, 用  $\text{ACR}(\alpha_i^{(t)})$  来表示。

公式说明:

$$\text{ACR}(\alpha_i^{(t)}) = \frac{|C_c^{(t)}|}{|C_m^{(t)}|} \quad \text{公式 (3-5)}$$

其中,  $C_c^{(t)} = \{c_j^{(t)} | \alpha_i \in \text{Mem}(c_j^{(t)}) \wedge \alpha_i \in \text{Core}(c_j^{(t)})\}$   
 $C_m^{(t)} = \{c_j^{(t)} | \alpha_i \in \text{Mem}(c_j^{(t)}) \wedge \alpha_i \notin \text{Core}(c_j^{(t)})\}$

## 3.3 公共数据集

本文以公共数据集中的科技文献的作者信息为节点, 以合作关系为边分别构建三个数据集的科研合作网络。其中DBLP国际学术科研合作网<sup>[7]</sup>的数据是从DBLP官方网站获取的DBLP Computer Science Bibliography计算机科技文献。万方科研合作网的数据是万方数据库中所收录从1987年到2007年的关于医学领域的论文文献。美国康奈尔大学电子图书馆“密物质”物理领域合作者网络<sup>[8]</sup>是Newman等人从康奈尔大学“密物质”物理学科的电子文献中构造的科研和作者网络。

## 4 基于社团演化分析的个体成长过程分析

表2 科研合作网络数据集

数据集	时间跨度	节点数	最大边数	备注
DBLP科研合作网	1959-2008	516,113	484,713	公共数据集
万方科研合作网	1987-2007	108,383	65,794	公共数据集
康奈尔大学“密物质”理论科研合作网 (Cond-mat)	1993-2006	68,612	52,738	公共数据集

### 4.1 基于核心成员与普通成员分析的社团演化路径建立

对社团间演化关系的确立，本文采用了基于CommTracker<sup>[9]</sup>的改进算法。传统的CommTracker<sup>[9]</sup>算法的模型假设：社团的核心成员一旦出现，不可能立即消失，同时大部分普通社团成员在其科研工作初期不是核心成员。其事实依据是从普通成员到核心成员有一段成长过程，但在这个成长过程中也不会消失。基于以上两点假设，提出如下判定条件： $Core(c_i^{(t)}) \cap Mem(c_i^{(t+1)}) \neq \emptyset$ ， $Core(c_i^{(t+1)}) \cap Mem(c_i^{(t)}) \neq \emptyset$ 。t时刻社团与t+1时刻社团若满足这两个条件就可以说明社团 $c_i^{(t)}$ 与社团 $c_i^{(t+1)}$ 存在演化关系。这样就限定了t与t+1时刻两个社团间的演化关系。从以下实际情况考虑，可以得知这两个限定条件仍然很宽泛。假设中只要求前后两个时间片内的社团其核心成员存在着一定的重叠关系，但事实上核心成员很可能代表着多个不同的研究方向，甚至参与到不同的科研活动中，而不同的科研活动产生了不同的科研团队，而不仅仅只是一个科研团队。因此基于这种情况考虑，本文提出改进：认为前后时间片内两社团存在演化是在其社团普通成员也存在一定的稳定性的前提下成立的。

当确立了社团间的演化关系，对于任意时刻的社团 $c_i^{(t)}$ ，就可以从当前时刻t向前寻找其祖先社团 $\{C^{(t-k)}, C^{(t-k+1)}, \dots, C^{(t-1)}\}$ ，直至t-k时刻的任意一个社团 $c_i^{(t-k)}, c_i^{(t-k)} \in C^{(t-k)}$ ， $c_i^{(t-k)}$ 都是一个新生社团为止。同理，可以递归地寻找 $c_i^{(t)}$ 的后继社团 $\{C^{(t+1)}, C^{(t+2)}, \dots, C^{(t+h)}\}$ ，这样就会得到

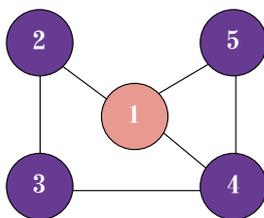


图1  $t_0$ 时刻社团

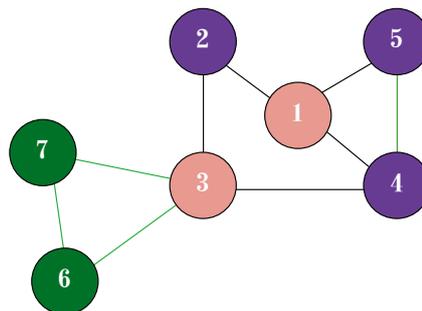
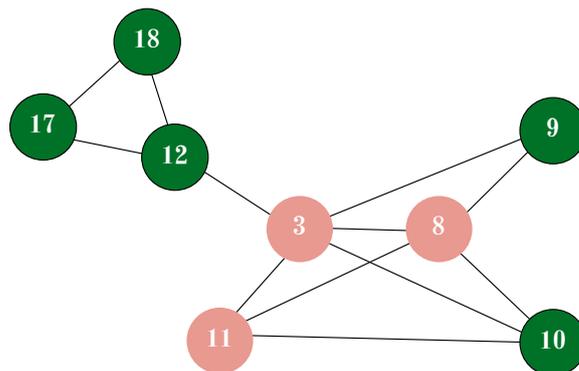
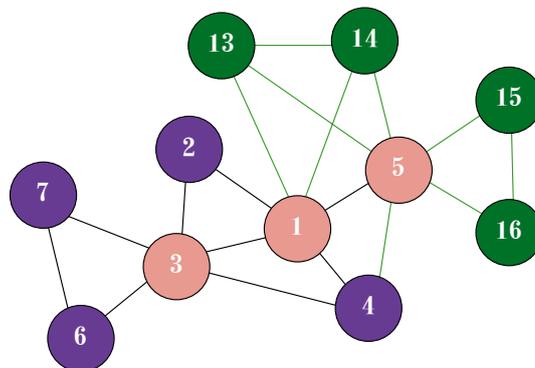


图2  $t_1$ 时刻社团



(a) 社团 $C_3^t$



(b) 社团 $C_4^t$

图3  $t_2$ 时刻社团

社团 $c_i^{(0)}$ 在整个演化过程中所出现的社团集合，记作： $EVOL(c_i^{(0)})=\{C^{(t-k)}, C^{(t-k+1)}, \dots, C^{(t-1)}, c_i^{(0)}, C^{(t+1)}, C^{(t+2)}, \dots, C^{(t+h)}\}$ 。本文将以图例形式演示社团演化路径的具体建立过程。

图1-图3分别显示了从 $t_0$ 时刻到 $t_3$ 时刻三个时间片内的四个不同的社团。显然 $Core(c_1^b) \cap Mem(c_2^b) = \{1\}$ ，且 $Mem(c_1^b) \cap Core(c_2^b) = \{3\}$ ，所以社团 $c_1^b$ 和 $c_2^b$ 存在演化关系，即： $c_1^b \rightarrow c_2^b$ 。同理社团 $c_2^b$ 和社团 $c_4^b$ 也存在演化关系。但此时，社团 $c_2^b$ 与社团 $c_3^b$ 并不存在演化关系，虽然社团 $c_2^b$ 和社团 $c_3^b$ 满足 $Core(c_2^b) \cap Mem(c_3^b) = \{3\}$ ，且 $Mem(c_2^b) \cap Core(c_3^b) = \{3\}$ 。其原因就在于节点3是社团 $c_2^b$ 和社团 $c_3^b$ 的唯一公共节点，但在社团 $c_3^b$ 中还存在着核心节点8，11，而这两个核心点并未在 $c_2^b$ 中出现。很明显这是节点3所参与的其他的科研社团。如果加入科技

的消失及节点19、20的加入，节点8失去了核心作用，而节点12逐渐成长为核心节点。以 $t_0$ 到 $t_3$ 四个时隙七个社团为例，展示了其整个演化过程，并得到两演化集合分别为： $\{c_1^b, c_2^b, c_4^b, c_7^b, c_6^b\}$ 和 $\{c_3^b, c_5^b\}$ 。

## 4.2 基于成员生命特性的社团演化路径特征的量化分析与统计结果分析

在上一小节通过对算法CommTracker<sup>[9]</sup>中的部分改进建立了社团演化路径。本节中进一步将不同特性的社团成员分为三类：普通成员、新生成员、核心成员，并对各成员的生命特性进行量化分析，同时给出社团演化路径特性的量化方法。本节采用了两个量化指标：成员活跃时长和成员变化率，同时展现了成员变化率的统计结果及分析。

### 4.2.1 成员活跃时长

成员活跃时间用来描述社团成员在其所参与到的社团活动中的持续时间，用符号 $ALS(\alpha_i, c_j^{(0)})$ 来表示。本文中采用两个量化指标来分析成员活跃时长，分别是相对于社团的成员出生时间 $Birth(\alpha_i, c_j^{(0)})$ 及衰退时间 $Death(\alpha_i, c_j^{(0)})$ 。 $Birth(\alpha_i, c_j^{(0)})$ 表示作者 $\alpha_i$ 成为社团 $c_j^{(0)}$ 的成员或第一次出现在社团的祖先社团中的时间。 $Death(\alpha_i, c_j^{(0)})$ 表示作者 $\alpha_i$ 最后一次出现时刻，且在此后再也没有在社团 $c_j^{(0)}$ 的后代社团中出现过。特别地，在 $t$ 时刻之前从未在社团 $c_j^{(0)}$ 的演化路径中出现过的成员的集合可表示为 $NewMem(c_j^{(0)}) = \{\alpha_i \mid \alpha_i \in Mem(c_j^{(0)}) \cap Birth(\alpha_i, c_j^{(0)}) = t\}$ 。然后通过求这三类不同的社团成员的平均活跃时长来考察不同的成员与社团的生命持续时间之间的关系。分别用 $ALS(Mem(c_j^t))$ 、 $ALS(NewMem(c_j^t))$ 、 $ALS(Core(c_j^t))$ 表示，结果发现 $ALS(NewMem(c_j^t)) < ALS(Mem(c_j^t)) < ALS(Core(c_j^t))$ 。

社团核心成员的平均活跃时间要远远大于其新生成员的活跃时间，这充分保证了核心成员对社团发展的积极影响。而新生成员可能只是在一段特别短暂的时间内参与到科研活动中，而后就再也没有进行科研活动，那么这类成员对社团的发展所起到的作用自然不能与核心成员相提并论。基于以上分析，更能体现本论文从微观角度对科研合作网络进行研究的现实意义。

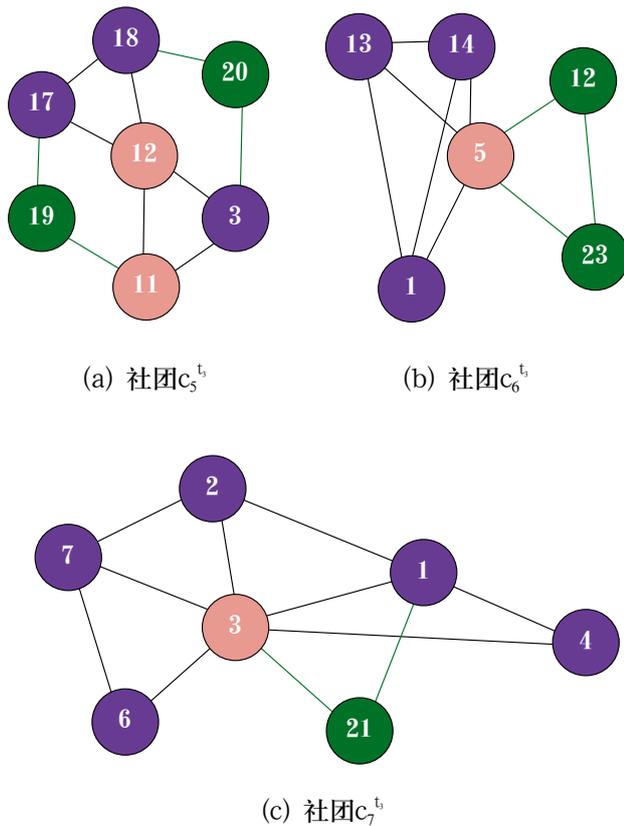


图4  $t_3$ 时刻社团

文献关键字、领域信息等额外信息，就可以更加直观地论证本文所提出的假设是有现实意义的。

接着加入 $t_3$ 时刻的社团，如下图4所示：

显然此时 $c_4^b$ 已经分裂为 $c_7^b$ 和 $c_6^b$ 两个社团，分别以节点5和3为核心成员，而节点1已经失去了核心的作用而成为普通成员。社团 $c_3^b$ 已经演化为 $c_5^b$ ，但随着节点9

### 4.2.2 成员变化率

成员变化率是影响社团成长的重要指标。新生成员的活跃时间远远比核心成员的活跃时间短，这与科研团队发展与成长过程相符合。因为科研团队本身是向科研工作者开放的，这样科研人员的进入与离开就是很正常的事情，同时这也是科研团队不断发展所必需的。但一个科研团队从形成确立到具有稳定的科研目标或任务也需要一定的时间，此过程中形成了科研带头人，把这类人称作社团的核心成员，并且正是科研带头人引领着整个团队的科研方向。虽然团队因新成员的加入及老成员的退出在不断更新着团队的成员组成，但科研团队的科研目标及任务不可能转瞬即逝，因此较大的成员变化率表明一个团队要发展就要不断地引进新成员。但科研团队要巩固加强其科研活动的地位就必须要保持核心成员的稳定性，只有这种有活力且有方向的科研团队才能发展得更好，生命持续时间才能更长。所以科研团队要保持开放性、同时加强其中间力量显得尤为重要。这也是研究科研团队的演化所得出的对科研团队的发展具有现实指导意义的地方。

### 4.2.3 成员变化率的统计分析

在上一小节的社团演化路径分析中，不难发现不同社团成员在社团科研活动中的活跃时间也是不相同的。现在根据建立起来的社团演化路径来分析社团普通成员及社团核心成员在整个社团成长的生命周期内

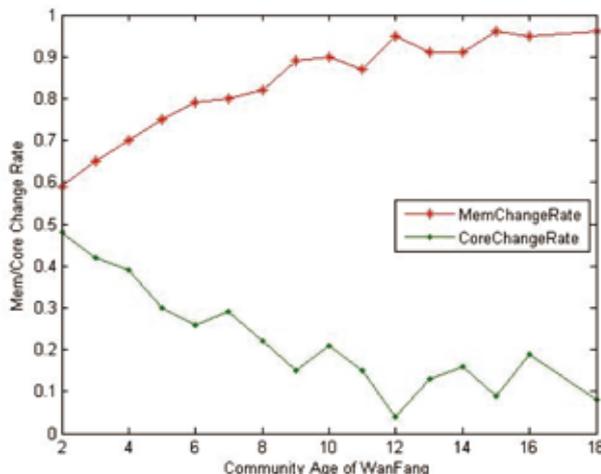


图6 社团成员变化率及核心成员变化率（万方）

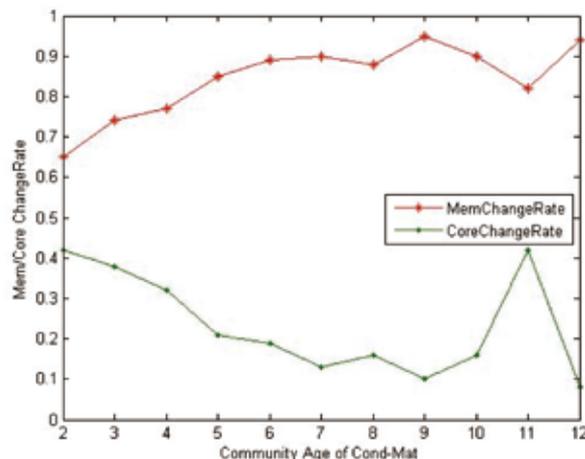


图7 社团成员变化率及核心成员变化率（Cond-Mat）

变化情况，继而考察其对社团成长的影响，并且在三个不同的数据集上得到了相似的计算结果，分别如图5、图6、图7所示。

从三个不同的数据集的统计结果，可以得出完全一致的结论：生命持续时间长的社团同时表现出较大的成员变化率和核心成员的稳定率。社团生命持续时间的长短与其不断更新社团成员的组成能力以及维持核心成员的相对稳定性的能力息息相关。对于不同的科研合作网络所表现出的共同的特性，给出如下的解释：科研合作网络中的科研团队，主要是在研究所或大学校园中的科研实验室中的科研活动中逐步形成的；在科研机构刚刚形成或创建时鲜为人知，因而会表现出较低的成员变化率。因为在最初的形成时期其吸纳新成员的能力有限，但又表现出比较高的核心成

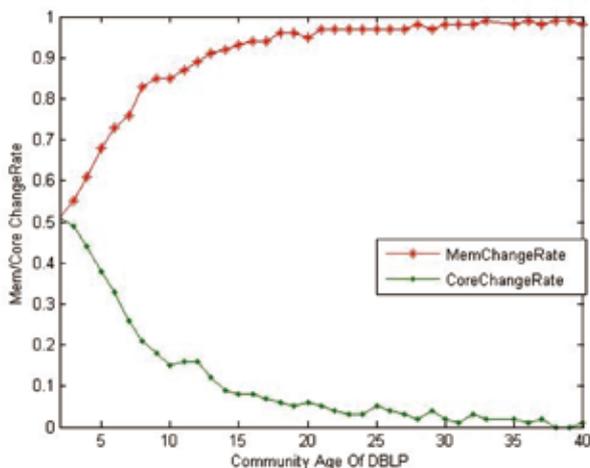


图5 社团成员变化率及核心成员变化率（DBLP）

员变化率，是因为在创建之初，科研团队对科研方向或目标的不确定性造成的。而随着科研团队的不断成长与发展，科研机构也为越来越多的学生、科研工作者所熟知，因此会有越来越多的人加入到科研团队中，同时由于学生的毕业与科研工作者对科研工作的重新选择，又会导致科研社团成员的流失，如此反反复复，不断有新成员的加入，又有老成员的离去，所以就表现出越来越高的成员变化率。这种现象也说明此类型的科研团队具有较强的生命力，从而为其新的发展带来新的生机。社团的核心成员则表现出极大的稳定性。这进一步说明，虽然在科研团队中不断有新成员的加入与老成员的离去，但科研机构仍然存在，科研机构的科研任务与科研方向仍然存在，核心成员的相对稳定性维持着其科研方向及目标的相对稳定性。这对一个社团的发展是相当重要的，因为任何组织如果没有了科研骨干则没有组织力和战斗力。

### 4.3 基于全局网络和社团演化的个体成长过程量化

本文以时间为窗口，以个体科研工作者为考察对象，在研究社团演化过程的基础上分析社团成长过程及其科研活动是否存在共性，如果存在又如何刻画其共性进行研究。文中采用全局的科研合作网络的视角和局部的社团演化两个不同的范围对科研工作者的活动进行量化与刻画。

#### 4.3.1 个体活跃度分析

对于个体活跃度的量化分析中，本文以时间为单

位，同时根据科研工作者间的合作关系及其科研成果的输出量（发表论文量）及他们对论文的贡献值为指标，符号定义为  $VctVal(\alpha_i^{(t)})$ 。本文中的这一量化方法有着传统统计方法所不能及的优点：

首先  $VctVal(\alpha_i^{(t)})$  同时考虑到了三个不同的因素：论文发表量、论文合作中的贡献值、时间。这样就从更深层而细微的角度来考察其科研情况，而不再只是单一的论文发表数量或其合作者的数量值进行统计。其次引入了时间衰减参数  $\lambda$ ，这可以起到两个作用：一是根据不同的科研合作网络作具体的调整，因此具有灵活性；二是随着时间的变化，在相当一段时间内，科研工作者依然存在，只是没有科研成果的发表，那么就可以看到其活跃度随着时间的增加而在不断地衰减，而不是没有显著变化，而当其又有成果发表后，会根据其成果的多少及其具体的贡献值再次计算其活跃值，从而可以完整地反映其整个生命周期内的活跃情况。

#### 4.3.2 基于社团演化的个人生命过程的量化分析

个体活跃度的量化结果只能反映科研工作者在整个科研合作网络中的科研情况，并不能具体地反映出其科研活动所处的范围，也就是他在科研活动中所起到的作用。本节中在社团演化的基础上将个体的生命周期划分为新生期、成长壮大期和衰减期，对个人的成长过程及他参与到的社团及其在社团中的地位进行刻画与研究。最后再结合全局的科研活跃度来分析个体的整个生命过程。具体的量化指标有科研工作者科研活动的稳定性以及在其参与到的社团中是否是核心成员，进而从整体上反映出其科研地位。

表3 节点3成长过程跟踪统计

量化参数	$t_0$	$t_1$	$t_2$	$t_3$
参与社团	$c_1^t$	$c_2^t$	$c_3^t$ $c_4^t$	$c_7^t$ $c_5^t$
核心社团	$\Phi$	$c_2^t$	$c_3^t$ $c_4^t$	$c_7^t$
社团演化路径	$c_1^t$	$c_1^t \rightarrow c_2^t$	1) $c_1^t \rightarrow c_2^t \rightarrow c_4^t$ 2) $c_3^t$	1) $c_1^t \rightarrow c_2^t \rightarrow c_4^t \rightarrow c_7^t$ 2) $c_3^t \rightarrow c_5^t$
核心率	0	1	1	0.5
变化率	0	0	0.5	0

仍然以图1到图4为例，对3号节点所参与社团情况以及其在社团中的核心率进行计算，从而得到以下观察数据：

从表3中的数据可以看出节点3在四个时隙上所参与的社团及所参与社团间的演化关系。很显然，节点3在 $t_0$ 时刻只是一个普通成员（有的节点甚至在初始时刻不是任何社团的成员），所以其核心率为零； $t_1$ 时刻成为社团 $c_2^b$ 的核心成员，但社团 $c_2^b$ 由社团 $c_1^b$ 演化而来，而不是一个新生的社团，所以此时节点3的变化率仍为零，表示其所参与的社团范围并没有新的变化； $t_2$ 时刻节点3参与到了一个新生的社团 $c_3^b$ ，所以此时其变化率为0.5。用同样的方法，可以继续对节点3的核心率及变化率做类似的分析。

可以得出：以社团演化为基础可以更加细致地跟踪其个体成员在科研活动中的活动情况，清晰地看到其所参与的社团以及在社团中是否是核心成员，从而来判断其最新的科研动态及当前所处的生命状态。

#### 4.4 量化分析实验结果及分析

通过分析科研工作者的活跃度、所参与的科研团队以及在科研团队中的地位来量化科研工作者在每个科研阶段中的成长情况，从而绘制出个体整个生命周期内科研活动的变化。

图8与图9分别展示了从DBLP数据集中获得的两位学者的成长过程图。前者为数据挖掘领域的专家，后者为社会网络分析领域的专家。图中分别用红色、绿色与蓝色曲线图表示科研工作者科研活动的活跃度、

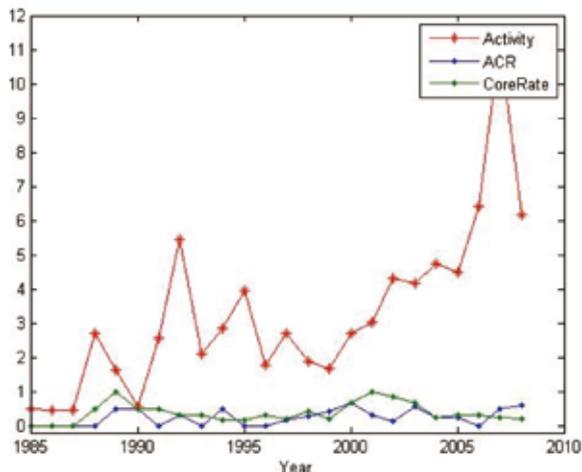


图8 数据挖掘领域 (Data Mining) 专家A个人成长过程图

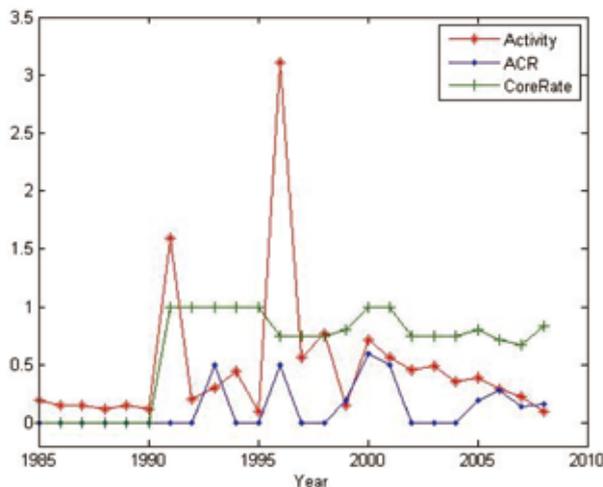


图9 网络科研领域 (Networking) 专家B个人成长过程图

所参与社团的变化率，以及在其所参与到的社团中的核心率。从图中不难发现，两个科研工作者的成长过程可大致分为三个阶段：新生期、成长壮大期、衰退期。

在图8中，作者A在1987年之前，图9中作者B在1990年之前，他们的活跃度都比较平稳且比较低，并且其核心率及所参与社团的变化率都为零。这说明科研工作者在刚步入科学研究时，科研成果比较少，很可能还没有形成自己的科研团队抑或没有参与到已有的科研团队的科研活动之中，所以本文将科研工作者还未成为团队的核心、活跃度较低的这段时期称为新生阶段。而在此阶段之后（图8中从1987年以后，图9中从1990年以后），可以看到作者A与作者B的活跃度有明显提高，核心率也在明显上升，而参与的社团也有明显的变化，这说明在新生阶段的积累后，他们开始慢慢参与到越来越多的科研活动中，并且成为科研团队的一员，并慢慢成为科研团队的核心。称这一阶段为成长壮大期。这一阶段的显著特征如前所述：活跃度明显升高、核心率升高、参与社团的变化较大。在此阶段之后，从图9中比较明显地看出，作者B的活跃度变得越来越小，但他/她仍然是科研团队的核心，并且还会参与到不同的科研活动中。最合理的解释是作者B凭借其以前的积累建立起了稳定的科研合作关系与科研地位，而又有越来越多的人接替他的工作继续进行科学研究，而这些新加入的成员不可能在短暂的时间内取代其核心成员的地位而成长为核心成员，而必须也要像B一样经历其新生阶段、成长壮大阶段才能

慢慢成长起来, 进而成为科研团队的核心。

## 5 结论

本文首先利用基于核心点与普通成员双重重合的方法建立了社团演化路径, 给出了社团的演化特性的量化方法, 提出了基于社团演化的个人成长过程的量化方法, 并对个人成长过程进行了阶段性划分, 即出生阶段、成长壮大阶段、衰落阶段。然后在DBLP数据集的实验分析的基础上, 展现了典型个体科研工作者整个生命周期的主要特征, 从时间演化的角度动态跟踪个体科研工作者的成长过程, 并且给出了成长过程

的量化指标, 通过分析个体科研工作者的成长过程, 提供了很好的模型。在进行大量的实验基础上, 本文总结了一个共同的规律: 关注科研团队中个体的生命过程(即新生期、核心期、衰落期), 有利于及时把握团队的发展动态, 为科研团队的演化提供指导。科研团队要可持续发展, 既要不断吸纳新成员, 为其注入新活力, 同时又要有相对稳定的中坚力量, 以维持团队的科研方向。本文虽然是基于大量的科技文献的研究, 但其研究方法及实验假设可以扩展到其他社会组织的分析中, 跟踪关心的团队中心个体来预测团队的演化趋势。

### 参考文献

- [1] BARABASI A L, ALBERT R. Emergence of scaling in random networks [J]. Science, 1999,286(5439):509-512.
- [2] PALLA G, BARABASI A L, VICSEK T. Quantifying social group evolution [J]. Nature, 2007,446(7136):664-667.
- [3] LAPPAS T, LIU K, TERZI E. Finding a team of experts in social networks: KDD 2009: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009 [C]. New York, 2009.
- [4] Huang J, Zhuang Z, Li J, etc. Collaboration over time: characterizing and modeling network evolution: WSDM 2008: Proceedings of the international conference on Web search and web data mining, 2008 [C]. California, 2009.
- [5] GETOOR L, DIEHL C P. Link Mining [J]. SIGKDD Explorations, 2005,7(2):3-12.
- [6] TANG J, SUN J, WANG C, et al. Social influence analysis in large-scale social networks: KDD'09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009 [C]. New York, 2009.
- [7] DBLP [EB/OL]. [2010-04-09]. <http://kd1.cs.umass.edu/data/dblp/dblp-info.html>.
- [8] Newman network data [EB/OL]. [2010-04-06]. <http://www-personal.umich.edu/mejn/netdata/>.
- [9] WANG Y, WU B, PEI X. CommTracker: A core-based algorithm of tracking community evolution [C]// ADMA, Chengdu, China, 2008.

### 作者简介

王丁弘, 硕士研究生, 研究方向为数据挖掘、复杂网络。通信地址: 北京邮电大学179信箱 100876。E-mail: wangdinghong@bupt.edu.cn  
吴斌, 副教授, 主要研究领域为数据挖掘、复杂网络及智能信息处理。通讯地址: 同上。E-mail: wubin@bupt.edu.cn

### Analyses of Academic Life-Cycle Based on Evolution of Scientific Collaboration Networks

Wang Dinghong, Zhao Fengying, Wu Bin / Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts & Telecommunications, Beijing, 100876

Abstract: Scientific collaboration network is one of the most typical complex networks. Traditional research on complex networks is focused on macroscopic properties of networks. In this paper, a microscopic analysis method which can be used to evaluate and analyze scientific researchers has been put forward. Based on the larger scale of datasets collected for co-authorship networks, we give in-depth analysis of single researcher. Furthermore, quantitative evolution analyses are demonstrated by tracking the common members. From the experimental results, we apparently conclude that introducing new members and stabilizing the core members play the same important role in the further development of a scientific research team. Meanwhile, the methods proposed in this paper can also be used on the analysis of other social networks such as terrorist organization, which can lead to finding a trend of organization and key individual's impact to evolution of organization.

Keywords: Scientific collaboration networks, The growing process of individual, Activity of individual

(收稿日期: 2010-05-31)