

1973—2019年中国气象观测领域的 科研合作特征和机制

侯威 杨晓武 张楠

(中国气象局综合观测司, 北京 100081)

摘要: 量化气象观测领域的科研合作有助于科学规划和管理观测研究团队。基于1973—2019年的17777篇气象观测文献, 利用复杂网络方法, 对中国气象观测领域的科学研究分析表明, 绝大部分气象观测科研人员以团队形式开展研究, 且以团队内部合作为主, 团队间的合作较少。气象观测科研团队的合作机制主要包括完全择优法则和随机-择优混合法则, 且随机-择优混合法则的作用突出; 团队的规模影响气象观测科研团队的合作程度, 团队规模越大, 合作程度越高。

关键词: 气象观测, 科研人员, 合作机制, 复杂网络方法

DOI: 10.3969/j.issn.2095-1973.2022.05.012

Characteristics and Mechanism of Scientific Research Cooperation in the Field of Chinese Meteorological Observation During 1973–2019

Hou Wei^{1,2}, Yang Xiaowu¹, Zhang Nan¹

(Department of Integrated Observations, China Meteorological Administration, Beijing 100081)

Abstract: To better understand and quantify the collaborations in metrological research areas, co-authorship is studied as an important entry point to analyze the development of scientific research in this field. 17,777 literatures related to meteorological observation from 1973 to 2019 are selected as the research object. By the use of the complex network method, a co-authorship network is constructed to reflect the collaboration of scientists in this field. The analysis results of the modular characteristics and the topological structure of the network show that most scientists in the field of meteorological observation collaborate as scientific research teams, but this kind of cooperative relationship focuses on internal team cooperation, with less cooperation between different teams. The cooperation mechanism mainly includes completely preferred rule and random-preferred hybrid rule, with the latter more pronounced. And the size of the research team is an influencing factor for the cooperation. The larger the team size, the higher the cooperation degree.

Keywords: meteorological observation, research team, cooperation mechanism, complex network

0 引言

随着现代科学的发展, 迫切需要理解、量化和预测科学研究及其产生的结果, 以科学地规划和管理研究团队。正是在这一背景下, 科学学应运而生。科学学是一门研究科学整体的学科或学科门类, 它以整个科学技术知识体系及其活动为研究对象, 探讨科学的社会性质、作用和发展规律以及科学的体系结构、规划、管理和科学政策等^[1-2], 包括研究各类科学合作的内在机理并建立相应的描述模型, 以理解科学研究的

发展和创新过程^[3]。因此, 科学学将整个科学系统视为一个复杂系统, 对这一复杂系统的内在结构、动力学与进化机制进行系统研究。作为描述和研究复杂系统的理论基础和有力工具, 复杂网络^[4-5]为科学学研究提供了重要的方法, 其中一个重要方面就是对科研人员之间相互交叉、相互渗透合作关系的复杂性特征的研究^[6-9]。

气象观测是指对地球大气圈的物理、化学特征及其变化进行长期、连续、系统的观察和测定, 并对获得的信息和数据进行整理和质量控制的过程^[10-11]。气象观测具有准确性、代表性与可比较性, 可为天气预警预报、气候预测预估及气象服务、科学研究提供高质量的观测数据^[12]。因此, 气象观测一直处于大气

收稿日期: 2021年11月5日; 修回日期: 2021年12月15日

第一作者: 侯威(1981—), Email: houwei@cma.gov.cn

资助信息: 国家自然科学基金项目(42005058, 41675092)

科学发展的前沿，是推动气象科学发展的基础和动力^[13-16]。气象观测是一门多领域交叉学科，涉及数学、物理、信息、材料、电子技术等多学科领域，随着各类先进装备制造和先进信息技术的不断发展和应用，全球气候变化背景下的气象观测科学研究也日趋专业化、异构化和复杂化，气象观测领域的科学技术和研究方向也在不断发展和增加^[17-19]。因此，气象观测领域科学和技术的发展需要不同学科背景、不同研究方向的科研人员密切合作。气象观测科学是一个复杂系统，需要利用复杂网络深入揭示气象观测领域科研合作的行为模式、机制与规则。对气象观测领域科研合作的发展与质量进行深入探究不仅有助于整合科研资源，提升科研人员的自主和协同创新能力，也是科学学的重要研究领域之一^[20-22]。但到目前为止，还没有针对包括气象观测在内的气象领域不同学科或者研究方向的定量研究。

本研究将论文合著作为分析该领域科学研究发展的切入点，基于复杂网络的相关理论构建1973—2019年的历年科研人员合作网络：采用共同作者网络^[23-24]，描述气象观测领域科研人员之间的合作关系，分析气象观测领域科研人员合作的内在特征和机制，以更好地促进科研合作与交流、提升科研产出与科研质量、推动气象观测科学技术的突破与创新^[25-27]。

1 资料与方法

1.1 文献数据

文献体现了科研人员之间合作关系。本研究的文献数据来源于中国现有的三个综合性中文期刊全文数据库：《中国学术期刊网络出版总库》《中文科技期刊数据库》和《万方数据》。这三个期刊数据库是中国连续出版的大规模集成化、多功能学术期刊全文数据库，内容覆盖自然科学、工程技术、农业、哲学、医学、人文社会科学等各个领域，收录出版国内各个学科领域专业性期刊、综合性期刊以及港台核心期刊8000余种，包括创刊至今出版的学术期刊4600余种。

(1) 论文检索。分别在三个期刊数据库中采用中图分类号P41，针对大气探测（气象观测）方向进行文献检索，检索时间为所有年份，期刊来源为所有期刊；将从三个期刊数据库中检索到的文献信息均保存为NoteExpress格式（包含作者、年份、标题、期刊、关键字、作者机构和作者地址等关键信息），并导入NoteExpress文献检索与管理软件。

(2) 文献清洗。基于NoteExpress文献检索与管理软件，采用“作者+年份+标题+期刊+作者机构”的联合关键词，对所有检索到的文献信息进行查重，将

重复的文献信息删除，最终得到18638篇文献，时间跨度从1936—2020年。剔除2020年收录尚不完整的文献（共432篇）以及部分缺少作者信息的历史文献（429篇），共得到17777篇文献（图1），时间跨度从1973—2019年。为防止同名同姓作者的影响，进一步提取17777篇文献的作者信息和作者机构信息，将同名同姓但不同机构的作者加以区分，区分方法是在姓名后加数字后缀，如张某1、张某2，以此类推。

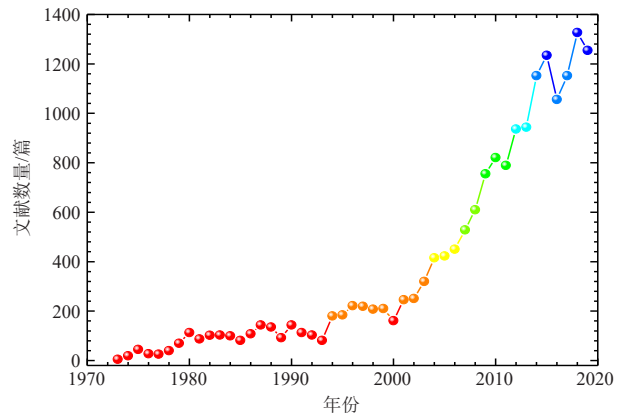


图1 1973—2019年历年文献数

Fig. 1 The amount of the literatures from 1973 to 2019

1.2 网络构建

本研究所要构建和研究的是科研合作网的一种：共同作者网络，这是一个加权无向网络。无向网络是指网络中两个节点间的连边如果仅有一个方向可通行则称为有向边，如果两个方向均可通行则称为无向边。若一个网络中所有的边都无向，则称此网络为无向网。加权网络是指节点间连边有权的图。可记网络为 $G=(V,E,W)$ ，其中 $V=\{v_1,v_2,\dots,v_n\}$ 为网络的节点集； $E=\{e_{ij}\}$ 为网络的边集，如节点 v_i 和 v_j 之间存在连边，则 $e_{ij}=1$ 否则 $e_{ij}=0$ ； n 为网络中节点的总数，也就是网络的规模； $W=\{w_{ij}\}$ 表示节点 v_i 和 v_j 之间所连接边的权重，本研究中的边权属于相似权 $w_{ij}\in[0,\infty)$ ， $w_{ij}=0$ 表示两点之间无连接， $w_{ij}=1$ 表示两点之间连接了一次，以此类推，边权越大表示两个节点的联系越紧密。

本研究将作者作为网络中的节点^[28-29]，整个网络的节点数之和是科学家的总数；如果两个作者共同署名发表一篇科研论文，那么这两个科学家之间就存在合作关系，代表两个作者的点之间就存在一条连边；在本研究中不区分作者的署名顺序，同一篇学术论文的多个作者的合作关系用全连接的方式表达，也就是所有作者两两之间均存在一条连边；若两个作者多次共同署名发表科研论文，则把合作次数，即合作

发表的学术论文数目作为代表两个作者的点之间边的权重。

1.3 网络特征量

为了刻画复杂网络全面性质，复杂网络的特征量和度量方法常被用于表示复杂网络的拓扑结构特性和动力学性质，包括：节点度、点权和边权，聚集系数、模块性等^[30]。

(1) 节点度、点权、单位权和差异性

节点度也称为连通度，节点 v_i 的度 $k(i) = \sum_{j=1, i \neq j}^n e_{ij}$ ，是指与节点 v_i 直接相连的其他节点的个数。在共同作者网络中，节点度就是与某个作者产生论文合作关系的其他作者的数量，描述了某个作者的合作范围。在加权网中，节点 v_i 的点权定义为 $s(i) = \sum_{j=1, i \neq j}^n w_{ij}$ ，即节点的点权等于该点所有边的权重之和，代表节点 v_i 与其他节点直接相连的次数，描述某个作者与其他作者的合作次数。点强度既考虑节点的近邻数，又考虑该节点和近邻之间的权重，是该节点局域信息的综合体现。点权分布 $P(S)$ 与度分布 $P(K)$ 的作用类似，主要是考察节点具有点权 s 的概率，这两个分布结合在一起，提供加权网络的基本统计信息。节点 v_i 的点权增量定义为 $s'(i) = \sum_{j=1, i \neq j}^n w_{ij} - \sum_{j=1, i \neq j}^n e_{ij}$ ，代表节点 v_i 与其他节点超过一次直接相连的次数，描述某个作者与其他作者的合作程度。

当边权与网络拓扑结构无关时，点权与度的函数关系为 $s(k) \cong \langle w \rangle k$ ，其中 $\langle w \rangle$ 为边权的平均值。当边权与网络拓扑结构相关时，点强度与度的函数关系一般为 $s(k) \cong Ak^\beta$ ， $\beta=1$ 但 $A \neq \langle w \rangle$ 或者 $\beta \neq 1$ 。在点权的基础上，还可以引入单位权 $u(i) = \frac{s(i)}{k(i)}$ ，表示节点连接的平均权重。但即使节点具有相同的度值和点强度，具有相同的单位权，也可能会情况迥异。例如，在单位权相同的情况下，可能是每条边的权重都接近于单位权 $u(i)$ 的数值，也可能是一条边或少数边上的权重处于优势。用 $y(i)$ 表示节点 v_i 所连接的边上权重分布的离散程度， $y(i) = \sum_{j \in n_i} \left[\frac{w_{ij}}{s(i)} \right]^2$ ，其中 n_i 是节点 i 的近邻集合。如果节点 v_i 的 $k(i)$ 条边的权重值相差不大，则 $y(i)$ 与度值 $k(i)$ 的倒数成正比，如果只有一条边的权重起主要作用，则 $y(i) \approx 1$ 。由定义可知 $y(i)$ 依赖于节点的度值 $k(i)$ ，所以通常更关心对所有度值相同的节点的 $y(i)$ 的平均值 $y(k)$ ， $y(k) \propto \frac{1}{k}$ 代表边权分布比较均匀，差异性

不大， $y(i) \approx 1$ 代表边权分布的差异性较大，独立于节点的度值 k 。

(2) 网络密度、聚集系数和模块

网络密度作为展现网络整体特征的重要参数，能够很好地反映网络的连通性。网络连通性是指如果网络中两个节点之间存在一条连通链路，称它们是连通的；如果网络中任何两个节点都是连通的，则称网络是连通的，即具有网络连通性。网络密度指的是网络中节点之间关系的紧密程度， $L = \frac{2m}{n(n-1)}$ ，其中 n 为网络中节点的数目， m 为网络中边的数目。在共同作者网络中，网络密度越大则意味着越多的科研人员之间存在合作关系，整个网络的连通性越好，科研人员之间的合作越广泛。

聚集系数是一个用来衡量一个复杂网络集团化程度的重要特征参数。加权网络中节点 v_i 的聚集系数^[31]

可以定义为 $c(i) = \frac{\sum_{jk} w_{ij} \cdot w_{jk} \cdot w_{ki}}{\max \left(w_{ij} \sum_{jk} w_{ij} \cdot w_{ki} \right)}$ ，描述网络中与

节点 v_i 直接相邻的节点间实际存在的边数目占最大可能存在的边数的比例，近邻之间联系越紧密，该节点的聚集系数越高。在计算出每个节点的聚集系数之后，就可以得到所有度值为 k 的节点的平均聚集系数 C_k 和整个网络的平均聚集系数 $C = \frac{1}{n} \sum_{i=1}^n c(i)$ ，其中 n 为网络中节点的总数，也就是网络的规模。对共同作者网络而言，一个集团或团队中的作者往往相互熟悉，网络的聚集系数越高，与某一个作者合作过的其他两个作者彼此也合作过的可能性越大。

现实世界中的许多网络是由模块结构组成的。模块结构有两个显著特征，一是模块内部的节点间高度连接，有着直接的相互作用；二是模块与模块之间只有少数甚至没有连接，模块与模块或模块与非模块之间有着清晰的边界。在复杂网络研究领域，模块也称作社区或群落。2002年，Girvan等^[32]给出了一种识别模块结构的新算法（GN算法），其基本划分算法如下：①计算复杂网络中每一条边的边介数^[33-34]；②找到介数最高的边并将它从网络中移除；③重复步骤①和②，直到每个节点就是一个模块为止。GN算法有很高的准确度，但其算法复杂度比其他算法的时间复杂度更高。因此，该方法适用于对中等规模的复杂网络进行模块划分。在共同作者网络中，不同的模块代表不同的科研团队；不同的模块之间的连边代表了不同的科研团队之间存在合作关系，这些模块成为连通模块。

2 网络整体特征分析

1973—2019年共同作者网络的聚集系数在增大(图2a),表明网络的集团化或者模块化程度在增加,网络密度在1993年之前总体是增大的,之后持续减小,但历年共同作者网络密度的最大值仅有0.011,表明网络的连通性不够好。1973—2019年共同作者网络的节点数持续增加(图2b),网络中模块的数量也是增加的,但在2003年之前没有出现连通模块,历年网

络连通模块数的最大值为17(2017年),即47年中最大规模团队合作涉及17个科研团队;历年连通模块数占有模块数的最高比例仅有2.3%(2017年),其余年份均不超过1%。随着网络规模的增大,网络中孤立点数量也是增加的(图2c),虽然从2012年开始增速加快,但历年孤立点的总数远小于非孤立点;历年孤立点数量占有节点数量的百分比也在持续下降,但从2012年开始整体保持平稳。

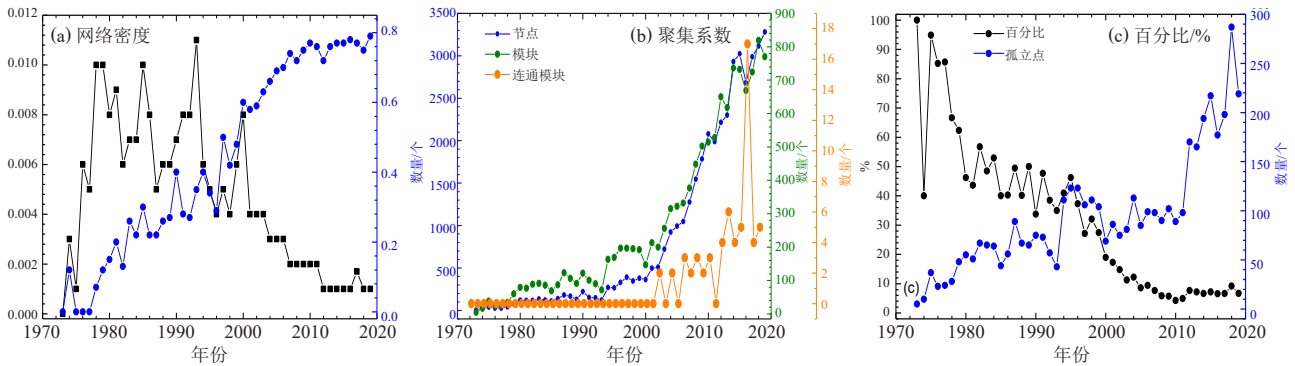


图2 1973—2019年共同作者网的整体特征

Fig. 2 The overall characteristics of the co-author network from 1973 to 2019

1973—2019年气象观测领域科研人员数量在持续增加(图2),共同作者网络的规模不断增大,绝大部分科研人员开始由“单打独斗”的研究形式逐渐转为以组成科研团队的形式开展研究,合作范围也在逐渐扩大,科研团队的数量也在明显增加,但随着网络规模的进一步持续增大,加入该领域科研人员及其所组成团队数量的增多,科研人员之间的合作范围并没有显著增强。各年的共同作者网由若干个相互独立的模块组成,同一模块中的任意两个节点之间至少存在一条连通路线,但绝大部分模块之间没有连线。这表明,虽然大部分科研人员选择相互合作而组成相当数量的科研团队,但这种合作关系大都以团队内部合作为主,团队之间合作相对较少。2019年共同作者网络的拓扑结构很好地展示了气象观测领域科学家合作共同作者网络的上述特征(图3),其他年份共同作者网络的拓扑结构与此类似(图略)。由此可见,研究对象是一个高度不连通网络,这一特征决定了很难分析通常意义下整个网络的统计特性(如计算网络平均路径长度、聚类系数等网络特征量),也很难说明其现实意义。所以,结合研究对象的实际特点,本研究进一步分析网络的模块化即科研团队的相关网络特征。

3 模块特征分析

在共同作者网络中存在节点之间具有相对较强的、直接的、紧密的连接模块,这些模块的规模一般相对较小并具有“高内聚、低耦合”的基础特征,

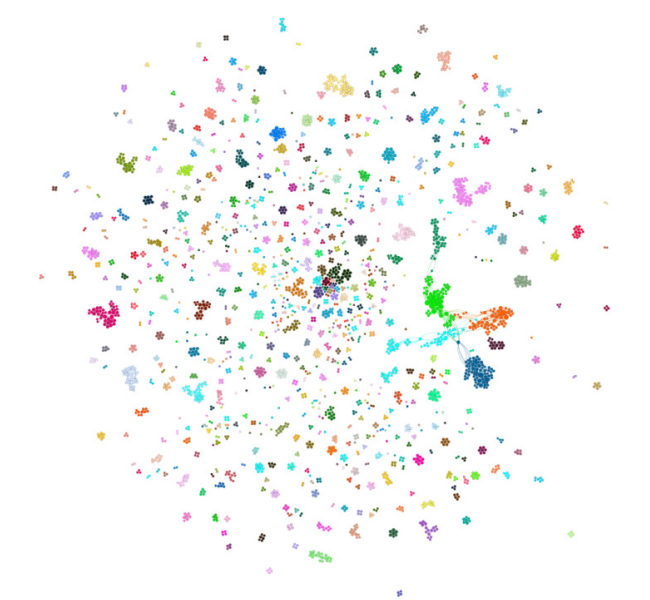


图3 2019年共同作者网络结构拓扑图

Fig. 3 The topological diagram of co-author network structure in 2019

是组成更大规模模块或连通模块的基础。如图4所示,可将共同作者网因合作而形成的所有模块分为两大类^[35]: A类模块为孤立点或全连接模块,即独撰论文的孤立作者和模块内的任意两个作者之间都有合作的全连接模块,代表内部合作型科研团队; B类模块为非全连接模块,即模块内的任意两个作者或任意两个A类模块之间的合作并不是完全确定的,从拓扑结构

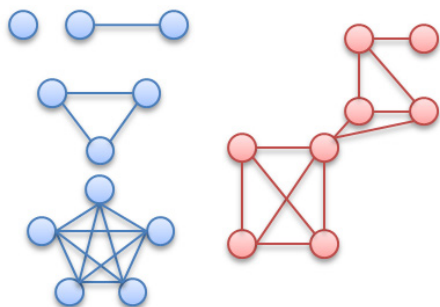


图4 两类模块类型示意图 (蓝色: A类模块; 红色: B类模块)
Fig. 4 Schematic diagram of two types of modules (blue: type A module; red: type B module)

上来看任意两个作者间总存在连通的合作途径(连通链路),代表开放合作型科研团队。B类模块是由一些A类模块中的活跃作者,与其他A类模块中的活跃作者之间建立起联系而形成的连通模块。A类模块和B类模块的区分十分有利于研究并建立共同作者演化的数学模型及仿真,深入揭示科学合作研究的形成和演变过程及其学科、地域、部门等现实特征。事实上,在更多研究领域、更大空间范围和更长时间尺度的背景下,不同A类模块之间的合作最终会形成具有十分复杂拓扑结构的科学家合作网,所以B类模块的形成和演变过程更能体现科研工作者之间合作关系随时间的演变过程^[36-37]。模块中的节点数、边权数分别代表了团队的人数规模、合作次数,节点的度、点权代表科研人员的合作范围及其与其他人员之间的合作次数,边的边权则代表了两个科研人员之间的合作次数。

3.1 模块特征量统计

1973—2019年,共存在11122个A类模块,模块总节点数为25675,总边权数为29547。平均每个模块中有2.3个节点、边权数为2.7;模块的最少节点数为1,最少边权数为0并且历年都有出现,最多节点数为14,最大边权数为91(2000年)。节点的度平均值为1.12,度最大值为13(2000年),平均点权为1.15,最大点权为21(2017年),度最小值和最小点权均为0,且历年都有出现,边权离散度的平均值为0.515、最大值为1(占有所有存在连边节点数的24.3%)、最小值为0.077。节点的单位权为1.03,最大边权为5(2015年和2017年),最小边权为1,且历年都有出现。

1973—2019年,共存在1351个B类模块,模块的总节点数为14793,总边权数为38105。平均每个模块有10.9个节点、边权数为28.2;模块的最少节点数和最少边权数均为4并且历年都有出现,最多节点数为113(2018年),最大边权数为563(2017年)。节点的度平均值为2.34,度最大值为36(2015年和2017年),平均

点权为2.58,最大点权为56(2017年),度最小值和最小点权均为1且从1980年开始历年都有出现,边权离散度的平均值为0.297、最大值为1(占有所有节点数的3.5%)、最小值为0.028。节点的单位权为1.1,最大边权为6(2010年),最小边权数为1且从1980年开始历年都有出现。

综上所述,1973—2019年,内部合作型团队的数量及其总人数规模要远多于开放合作型团队,但总合作次数却远小于开放合作型团队,开放合作型团队拥有更大、更灵活的发展空间。开放合作型团队已经形成了数百人规模的超大团队且其最小人数规模已超过绝大部分内部合作型团队。内部合作型团队人数规模、合作次数的平均值、最大值和最小值也都远小于开放合作型团队,同时单个开放合作型团队中人数规模、合作次数的值域范围都远大于内部合作型团队。开放合作型团队中科研人员的平均合作范围和合作次数高于内部合作型团队,最大合作范围和合作次数也远超内部合作型团队,两种类型团队中两个科研之间的最小合作范围无明显差异,并且开放合作型团队中人员之间的合作程度更为均衡,很少出现科研人员A与科研人员B之间的合作,在科研人员A全部合作关系中占主导地位的情况。两种类型团队中两个科研人员之间的最小合作次数相同,且平均合作次数和最大合作次数差异不大。

3.2 模块特征量分布

不同节点数(图5a)的A类模块数量呈漂移幂律分布 $P(N) \propto |N - \alpha|^\gamma$ (拟合置信度取0.95,下同),仅有1个节点的模块占比最大(39.8%),节点数为1~4的模块数量占比达91.6%。不同边权数内部合作型团队的数量整体分为上下两个相对独立的部分(图5c₁),上半部分满足指数分布 $P(N) \propto e^{-N/\gamma}$,下半部分满足幂律分布 $P(N) \propto N^{-\gamma}$;上下两部分随着边权数逐渐增加而趋向于一致。进一步分析发现,少数A类模块团队中的部分节点之间存在边权数大于1的边,图5c₂和5c₃就是对这一现象的描述,其中图5c₂代表了模块中各边的边权数均为1的情况,图5c₃代表了模块中存在边权数大于1的边的情况,只有2.7%的模块内部存在边权数大于1的边,表明少数内部合作型团队中的部分作者会合作发表多篇学术论文。

不同节点数开放合作型团队数量呈对数正态分布(图5b),即 $P(N) \propto \frac{1}{\sqrt{2\pi}\alpha N} e^{-\frac{(\ln(N/\mu))^2}{2\alpha^2}}$,模块数量随其包含节点数的增多而先增加后减少,节点数为6的模块数量最多(14.9%),其中节点数为4~7的模块数

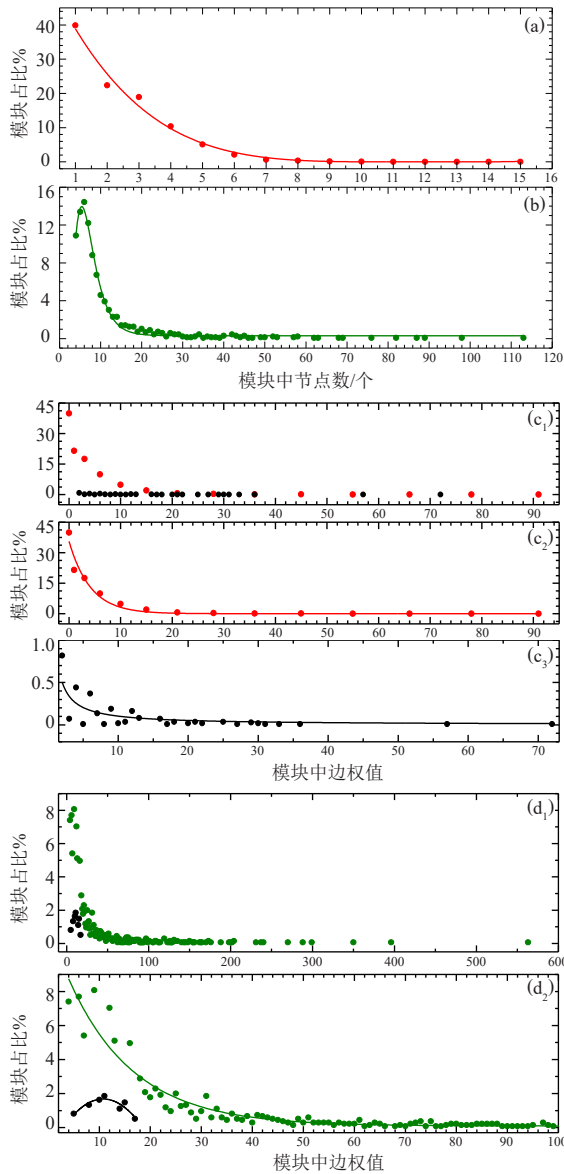


图5 不同节点数和不同边权数A类模块 (a, c) 和B类模块 (b, d) 数量的概率分布

Fig. 5 The probability distribution of type A modules (a, c) and type B modules (b, d) with different node numbers and edge weights

量占比50.9%且彼此差异不大;有49.0%的模块内部存在边权数大于1的边,表明接近半数开放合作型团队中作者之间存在多次合作。不同边权数开放合作型团队的数量整体同样分为上下两个相对独立的部分(图5d₁),其中上半部分满足指数分布 $P(N) \propto e^{-N}$,下半部分满足正态分布 $P(N) \propto e^{-((N-\mu)/\sigma)^2}$ (图5d₂)。下半部分对应边权数的值域范围相对较小($N \in [5, 17]$),并且数量呈正态分布的模块总体较少(占有B类模块数量的8.73%)。对于图5d₂中呈正态分布的模块,如图6a所示,其边权数越大,所包含节点数的值域范围也越

大,产生相同边权数的模块可以在较少节点数和较多节点数两种情况之间可以平滑过渡,边权数越大的模块可供选择的范围也越大;而呈指数分布的模块(图5d₁)则没有这一现象,能产生相同边权数模块的节点数的值域范围都非常狭窄(图6b)。B类模块中不同节点数和边权数的分布特征差异明显,说明B类模块中节点数和边权数之间的关系并不确定,开放合作型团队内的合作机制比内部合作型团队复杂。

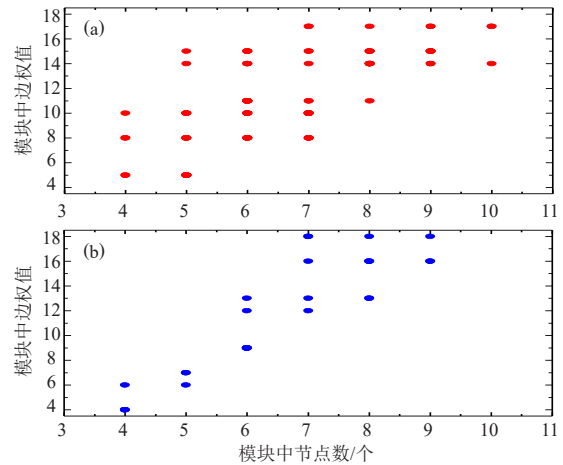


图6 图5d₂中呈正态分布的模块中节点数与边权数的对应关系
Fig. 6 The corresponding relationship between the number of nodes and the value of edge weights in the module with normal distribution in Fig. 5d₂

按本研究的定义,A类模块代表孤立点或全连接模块,模块具有均匀的网络拓扑结构。A类模块可分为两类,第一类是不存在边权大于1的模块(孤立点),模块中的节点之间拥有“平等”的度关系,节点的度值加1即为节点所在模块的节点数,并且模块中不同节点之间形成的连边不存在时间上的先后顺序,即团队中的科研人员之间只会合作一次,不同科研人员的合作时间是一样的;第二类是极少数存在部分节点具有边权大于1的边,其节点的度值和其所在模块的节点数无确定性关系,模块中不同节点之间形成的连边存在时间上的先后顺序。拟合置信度取0.95,A类模块的度(图7a)和边权(图7c)均服从正态分布;由于A类模块中具有边权大于1的边的节点数相对于总节点数而言非常稀少,因此依然可以认为模块中节点的度值与其所在模块的节点数成正比,内部合作型团队节点的度分布是由不同规模团队数量的分布直接导致的。A类模块的点权增量服从指数分布(图7e),只有少数团队中的部分科研人员会和团队中其他人员先后多次合作。

B类模块的度分布(图7b)服从对数正态分布。对数正态分布是某些因素起了突出作用但是未起到左右

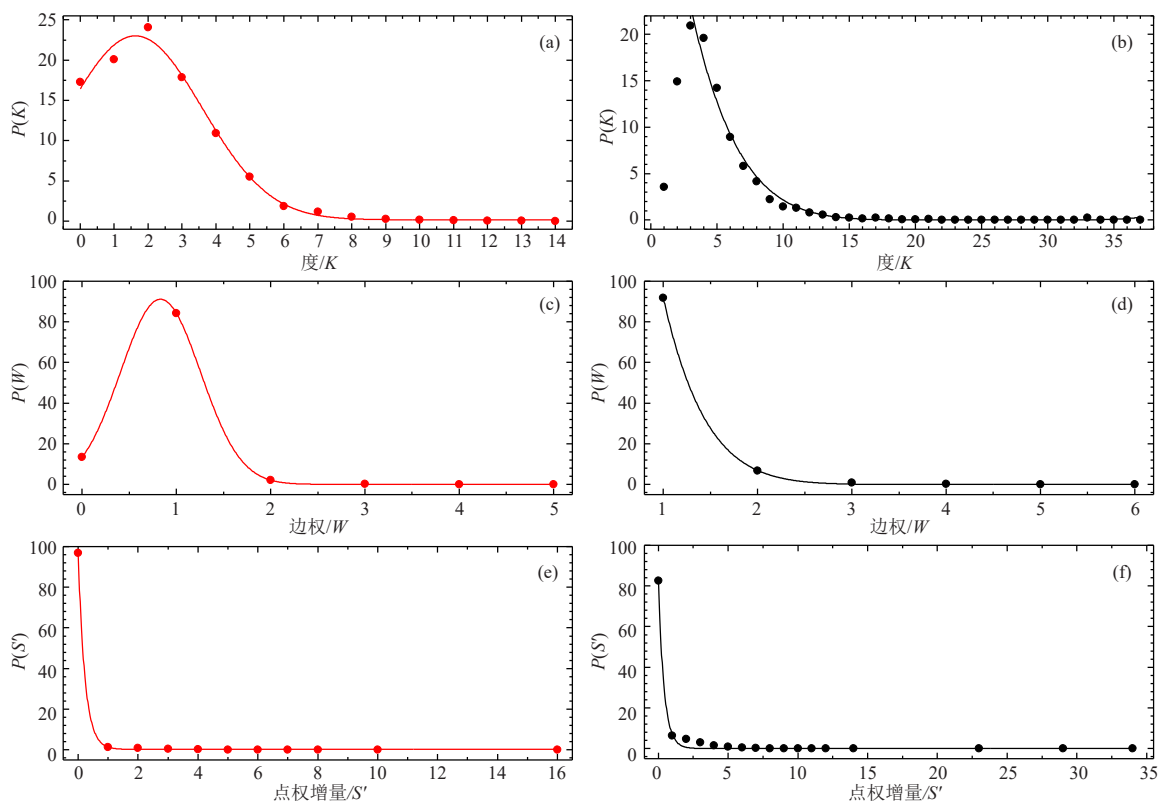


图7 A类模块(红色)和B类模块(黑色)的度分布(a, b)、边权分布(c, d)和点权增量分布(e, f)

Fig. 7 The distribution of degree (a, b), edge weight (c, d) and increment of node weight (e, f) of type A modules (red) and type B modules (black)

全局的结果^[38]。合作网络的演化动力学分为随机选择和优先选择,当节点以概率 p 随机选择与其他节点相连,而以概率 $1-p$ 选择与度值较高的节点相连,此时网络的度分布就会呈现漂移幂律分布^[39-43]。研究^[44]表明,如果一个模块中包含的节点数为 k (图5b),不管 N 为常量还是满足(对数)正态分布、幂律分布或指数分布,模块中节点的度分布在各种 p 参数下都遵从漂移幂律分布。按本研究的定义,一个B类模块是由多个在不同时间形成的A类模块所组成的,即B类模块中不同节点之间形成的连边存在时间上的先后顺序,也就是一个开放合作型团队中不同科研人员的加入时间是不一样的。图7b可以认为一种“去头幂律分布”^[45];其中 $k \in [1, 2]$ 对应于完全择优法则,也就是科研人员只和其他具有广泛合作关系的科研人员进行合作,并且几乎不与合作范围(度值)比自己低的科研人员合作;当 $k \in [3, 36]$ 满足漂移幂律分布,说明开放合作型团队满足随机-择优混合法则,即科研人员以概率 p 随机选择合作对象,以概率 $1-p$ 选择那些合作范围较大的科研人员作为合作对象。这说明开放合作型团队的形成具有两种机制,第一种机制是完全择优法则;第二种机制是随机-择优混合法则,第二种机制起了突出作用但是没有影响全局(占有节点数的81.6%)。开放

合作型团队的边权(图7d)也服从漂移幂律分布,同样满足上述机制。因此可以认为B类模块的网络拓扑结构具有自相似性,即所有B类模块与部分B类模块、或者单个B类模块与其局部之间(整体和部分)、规模较小的模块与规模较大的模块之间(部分与部分)呈现某种相似性,将开放合作型团队的规模视为尺度,则可以认为开放合作型团队所有尺度上都具有自相似的性质,开放合作型团队构成的合作网络具有无标度性质。点权增量(图7f)服从指数分布,点权增量和边权越大,满足要求的节点的数量就越少,说明少数科研人员与其他人员之间存在多次合作,合作程度较深,多数科研人员与其他人员之间只存在一次合作关系。

3.3 团队合作特征

A类模块和B类模块中节点的边权离散度均随节点的度成反比 $y(k) \propto \frac{1}{k}$,说明两种类型模块中节点的边权分布比较均匀,内部合作型团队中合作范围相同的科研人员的对外合作次数的差异不大,开放合作型团队中也存在同样的现象(图8)。图9给出了拥有边权大于1的边的节点数占相同度值总节点数的百分比,以分析具有不同合作范围的人员的多次合作情况。在A类模块中(图9a),随着节点度值的增大,其点权增

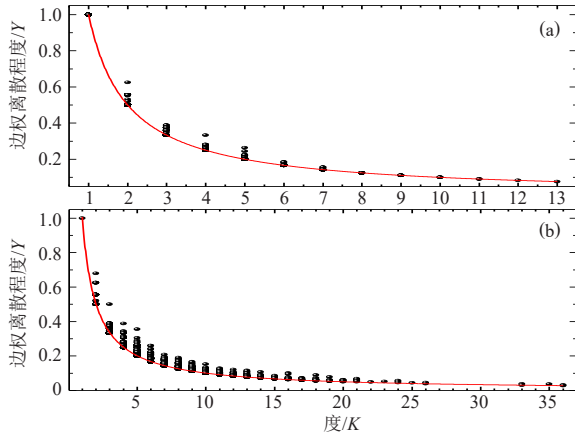


图8 A类模块(a)和B类模块(b)中节点边权离散程度随节点度的变化

Fig. 8 The variation of node edge weight dispersion with node degree and the probability distribution of node edge weight dispersion in type A modules (a) and type B modules (b)

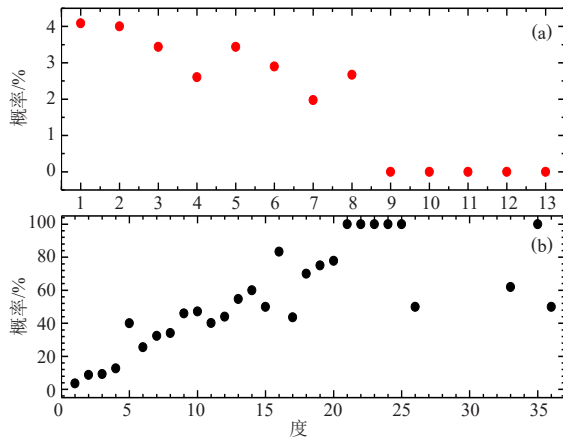


图9 A类模块(a)和B类模块(b)中不同度值节点至少存在一条边权大于1的边的出现概率

Fig. 9 The probability of multiple connections between nodes with different degrees and other nodes in type A modules (a) and type B modules (b)

量不为零的概率也随之减小,即度值越大的节点,越少概率具有边权大于1的边,即合作范围越大的人员,越多出现与每个合作人员只合作一次的情况。模块中的节点数越多,模块中两个节点之间出现多次连边的现象越少,也就是内部合作型团队的规模越大,团队中两个人之间就越少出现多次合作。图10给出了模块中边权与该边两端节点度的关系,即 w_{ij} 与 k_i, k_j 之间的关系,也就是某一权重的边连接了两个何种度值的节点,以分析具有不同合作范围的人员之间的合作次数。A类模块中较大的边权并不是出现在两个具有较大度值的节点之间(图10a),而是出现在两个度值偏小(不超过5)的节点之间,说明较小规模内部合作型团队中的人员之间倾向于多次合作。结合图5和图7,

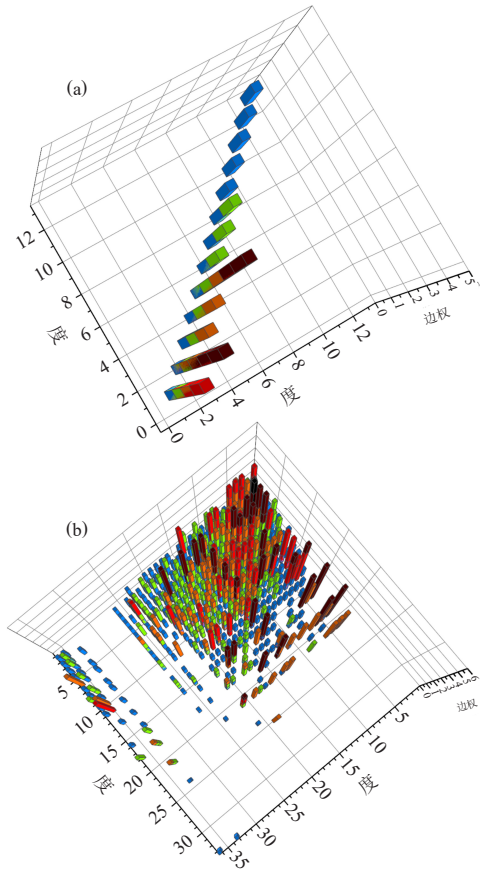


图10 A类模块(a)和B类模块(b)中不同度值节点间的边权
Fig. 10 The edge weights between nodes with different degree values in type A modules (a) and type B modules (b)

可以认为内部合作型团队中人员之间的合作机制主要取决于团队的规模,出现这一情况可能是由于团队的规模越小,团队中人员彼此之间越熟悉,即内部合作型团队中的人员倾向于优先和更熟悉的人员进行再次合作。

B类模块中(图9b),随着节点度值的增大,其点权增量不为零的概率也随之增大,即度值越大的节点,其存在边权大于1的边的可能性就越大,说明开放合作型团队中合作范围越大的人员,越少出现只与每个合作人员只合作一次的情况,这也证明开放合作型团队中择优机制的存在,合作范围大的人员会成为其他人员优先选择的再次合作对象。B类模块中较大的边权同样没有出现在两个度值很大的节点之间(图10b),而是频繁出现在度值相对较小(不超过15)的两个节点之间,尤其是当两个节点度值差异不大时,而且度值相对较小的各个节点之间都存在不同权重的连边;说明在开放合作型团队中,两个具有中等及以下合作范围的人员之间都会进行合作且倾向于多次合作,当二者的合作范围大致对等时更是如此。同

时,如图9b所示,对于度值大于25的节点,其存在边权大于1的边的概率整体上反而降低,说明当人员的合作范围很大时,其被优先选择作为再次合作对象的概率却在变小,和所有合作人员只合作一次的情况在增加,表明开放合作型团队中的“择优”选择也是有条件,这一点在图10b中也有体现,当一个节点的度值大于15时,其与其他度值的节点间较少出现较大边权,或者与其他度值的节点间连接变得稀疏;而当一个节点的度值大于25时,不仅与其他度值的节点间连接变得稀疏,连边也很少出现较大边权。

4 结论和讨论

4.1 结论

本研究基于《中国学术期刊网络出版总库》《中文科技期刊数据库》和《万方数据》三个期刊数据库中1973—2019年气象观测领域共17777篇文献的作者信息,采用复杂网络理论和方法,将作者作为网络中的节点、两个作者共同署名发表一篇科研论文代表两个节点之间就存在一条连边、由节点组成的模块代表科研团队,逐年构建了1973—2019年反映该领域科研人员之间合作的加权无向共同作者网络,并对这一网络的节点度、点权和边权、聚集系数、模块性等网络特征量进行了分析,其中模块中的节点数、边权数分别代表了团队的人数规模、合作次数,节点的度、点权代表了科研人员的合作范围及其与其他人员之间的合作次数,边的边权则代表两个科研人员之间的合作次数。结果表明:

1) 气象观测领域科研人员合作的共同作者网络是一个高度不连通网络:随着科研人员数量的持续增加,科研团队的数量也在明显增加;绝大部分科研人员以组成科研团队的形式开展研究,且以团队内部合作为主,团队间的合作较少;也就是单个科研人员更倾向于和其他固定的科研人员合作而不参与其他团队的研究,或者具有“交叉学科”研究能力的科研人员数量稀少。根据网络中模块“高内聚、低耦合”的基础特征,将科研团队分为内部合作型团队和开放合作型团队两种类型:对于内部合作型团队,仅有一人的团队数量最多,绝大部分团队包含1~4个人,仅有极少数内部合作型团队中的部分作者会合作发表多篇学术论文,说明有相当数量的科研人员不选择合作的方式,即使合作也仅限于少数人之间,合作多次则十分罕见;对于开放合作型团队,包含6人的团队数量最多,超过半数的团队由4~7人组成,接近半数开放合作型团队中的部分作者之间存在多次合作,说明这种类型团队中的部分科研人员具有较强的科研能力,能

够保持较好的持续性合作。

2) 内部合作型团队的数量及其总人数规模要远多于开放合作型团队,但总合作次数却远小于开放合作型团队,内部合作型团队人数规模、合作次数也都远小于开放合作型团队,开放合作型团队中科研人员合作范围和合作次数高于内部合作型团队,并且开放合作型团队中人员之间的合作程度更为均衡;整体而言,开放合作型团队的合作效率要高于内部合作型团队。总体上,两类团队中多数科研人员与其他人员之间只存在一次合作关系;内部合作型团队中合作范围越大的人员,越多出现与每个合作人员只合作一次的情况,较小规模内部合作型团队中的人员之间倾向于多次合作;而开放合作型团队中合作范围越大的人员,越少出现只与每个合作人员只合作一次的情况,两个具有中等及以下合作范围的人员之间都会进行合作且倾向于多次合作,尤其是当两个人员的合作范围大致对等时;说明内部合作型团队中核心骨干在科研人员间的“粘连”作用不如开放合作型团队。总体而言,开放合作型团队中人员的合作程度要优于内部合作型团队。

3) 不同人数的内部合作型团队的数量呈漂移幂律分布;当团队中所有作者都只有一次合作时,不同合作次数内部合作型团队的数量满足指数分布,否则不同合作次数内部合作型团队的数量满足幂律分布。不同人数的开放合作型团队的数量呈对数正态分布;当只能由值域范围非常狭窄的团队人数产生相同的合作次数时,不同合作次数开放合作型团队的数量满足指数分布,否则呈正态分布。

4) 开放合作型团队的合作具有两种机制,第一种机制是完全择优法则;第二种机制是随机一择优混合法则,第二种机制起了突出作用但是没有影响全局,而且开放合作型团队中的“择优”选择也是有条件,当人员的合作范围很大时,其被优先选择作为再次合作对象的概率却在变小;将开放合作型团队的规模视为尺度,可以认为开放合作型团队所有尺度上都具有自相似的性质,开放合作型团队构成的合作网络具有无标度性质;而内部合作型团队的合作机制主要取决于团队的规模。

4.2 讨论

考虑到复杂系统往往是随时间变化的动态过程,在得到网络结构复杂性特征的基础上,仍然需要采用复杂网络理论建立该领域共同作者网络的演化模型,分析该演化模型的拓扑统计特性,以及在该演化模型上合作机制的动力学行为。限于篇幅,本研究主要关

注气象领域科研人员合作共同作者网络的实证分析，但在基于网络特征量对气象观测领域共同作者网的拓扑性质进行研究过程中，一些拓扑性质的产生机制也同样值得关注，例如：

1) 不同人数的开放合作型团队的数量呈对数正态分布，对数正态分布^[38,46]常用于拟合人口数量增长、财富增长、生物组织生长等增长率独立随机且与当前规模无关的过程^[47]，而不同规模开放合作型团队数量的增长率与团队数量是相互独立的，这可能意味着不同规模开放合作型团队数量的变化可以用几何布朗运动进行解释^[47]。不同人数的内部合作型团队的数量呈指数分布，内部合作型团队中节点的度分布呈正态分布，后者可以由前者转换得到；研究表明，当网络中的节点数量足够多的时候，随机网络的度分布符合正态分布^[48]，而科研人员之间的合作不可能是随机行为，同时，一般认为如果一个事件受到许多微小又相互独立的随机因素的影响，但其中任何一个因素都不起决定性作用，则可认为该事件服从正态分布^[38]，所以除了相互之间的熟悉程度，构成内部合作型团队的科研人员之间的合作关系受其他何种因素的影响值得进一步研究。

$$2) \text{ 对于对数正态分布 } P(N) \propto \frac{1}{\sqrt{2\pi\alpha N}} e^{-\frac{(\ln(N/\mu))^2}{2\alpha^2}}$$

当 α 相对于 N 足够大时， $P(N)$ 趋向于幂律分布 $P(N) \propto N^{-\alpha}$ ，而当 α 相对于 N 较小时， $P(N)$ 则趋向于正态分布 $P(N) \propto e^{-((N-\mu)/\sigma)^2}$ ，因此对数正态分布是介于幂律分布和正态分布之间的一种分布；对于漂移幂律分布 $P(N) \propto |N-\alpha|^\alpha$ ，当 $\alpha \rightarrow 0$ 时， $P(N)$ 趋向于幂律分布，而当 $\alpha \rightarrow \infty$ 或者在 $\sum N_i = 1$ ，且 $\alpha \rightarrow 1$ 时， $P(N)$ 趋向于指数分布，因此漂移幂律分布是介于幂律分布和指数分布之间的一种分布， α 的大小则表示偏离程度；由此可见，上述5种分布可以归为“指数分布 \leftarrow 漂移幂律分布 \rightarrow 幂律分布 \leftarrow 对数正态分布 \rightarrow 正态分布”的数学过程，而且这5种概率分布在国内外不同学科、期刊或者跨领域的科研合作网络研究中都有出现。本研究针对中国气象观测科研人员合作共同作者网的研究中，不同网络特征量分别满足这5种可以连续变换的分布函数。内部合作型团队和开放合作型团队中网络特征量所展现出的丰富的统计特征，可能包含了两类团队的团队规模、单个科研人员的合作范围、不同科研人员之间的合作次数和合作程度等机制随时间的演变信息，因此这一现象值得用网络建模对两类团队的演变过程进行进一步研究。

3) 网络中模块的数量随时间是增加的，但在2003年之前没有出现过连通模块，历年网络连通模块数的最大值为17（2017年），而最大连通模块代表科研团队之间的合作，是科研合作最高级形态，对其形成和演变机制及其演变过程中内部拓扑性质随时间的变化进行研究，更有利于揭示气象观测领域科研合作潜在的共性特征和异质性变化。

参考文献

- [1] 吕达仁. 大气物理学二十世纪成就与世纪展望. 北京: 气象出版社, 2001.
- [2] 张大林. 大气科学的世纪进展与未来展望. 气象学报, 2005, 63(5): 812-824.
- [3] 孟智勇, 张福青, 罗德海, 等. 新中国成立70年来的中国大气科学研究: 天气篇. 中国科学(地球科学), 2019, 49: 1875-1918.
- [4] 王体健, 高太长, 张宏昇等. 新中国成立70年来的中国大气科学研究: 大气物理与大气环境篇. 中国科学(地球科学), 2019, 49: 1833-1874.
- [5] 张文建. 世界气象组织综合观测系统(WIGOS). 气象, 2010, 36(3): 1-8.
- [6] 张文建. WMO全球观测系统演进发展实施计划(EGOS-IP)执行摘要. 气象科技进展, 2015, 5(1): 72-74.
- [7] Stegehuis C, Litvak N, Waltman L. Predicting the long-term citation impact of recent publications. Journal of informetrics, 2015, 9(3): 642-657.
- [8] Clauset A, Newman M E J, Moore C. Finding community structure in very large networks. Physical Review E, 2004, 70(6 Pt 2): 066111.
- [9] Fortunato S, Bergstrom C T, Börner K, et al. Science of science. Science, 2018, 359(6379): eaao0185.
- [10] Bernal J D. 科学的社会功能. 北京: 商务印书馆, 1995.
- [11] 王续琨. 交叉科学结构论. 大连: 大连理工大学出版社, 2004.
- [12] 陈士俊. 科学学: 对象解析、学科属性与研究方法——关于科学学若干基本问题的思考. 科学学与科学技术管理, 2010, 31(5): 30-37.
- [13] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks. Nature, 1998, 393(6684): 440-442.
- [14] Barabási A L, Albert R. Emergence of scaling in random networks. Science, 1999, 286(5439): 509-512.
- [15] 汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用. 北京: 清华大学出版社, 2006.
- [16] Newman M E J. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences of the USA, 2001, 98(2): 404-409.
- [17] Newman M E J. Scientific collaboration networks: I. Network construction and fundamental results. Physical Review E, 2001, 64: 016131.
- [18] Newman M E J. Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality. Physical Review E, 2001, 64: 016132.
- [19] 秦大河, 孙鸿烈. 中国气象事业发展战略研究 现代气象业务卷. 北京: 气象出版社, 2004.
- [20] 宋连春, 李伟. 综合气象观测系统的发展. 气象, 2008, 34(3): 3-9.
- [21] World Meteorological Organization. World Weather Watch, 22nd Status Report on Implementation. WMO, 2005.
- [22] 黄荣辉. 大气科学发展的回顾与展望. 地球科学进展, 2001, 16(5): 643-657.
- [23] Zare-Farashbandi F, Geraei E, Siamaki S. Study of co-authorship network of papers in the Journal of Research in Medical Sciences using social network analysis. Journal of Research in Medical Sciences: the Official Journal of Isfahan University of Medical Sciences, 2014, 19(1): 41-46.
- [24] Barabási A L, Jeong H, Néda Z. Evolution of the social network of scientific collaborations. Physica A: Statistical Mechanics and its

(下转119页)

检查效果。

从上述案例可以发现, 经过不断开展基于PDCA循环的内部质检工作, 能够使业务工作持续改进, 进而实现空间天气业务的流程化、规范化。

3 总结

在国家卫星气象中心大力推进质量管理体系建设的背景下, 为落实中心的质量方针, 保障中心各项业务全面、有效运行, 空间天气室自2019年9月起定时开展空间天气室内部质量监督检查工作。通过这项工作的不断推进和完善, 空间天气室全体人员质量管理意识普遍增强, 体系与实际业务融合度不断提高; 业务人员基于风险思维对业务过程进行管理, 加强与地方台站沟通, 及时处理数据故障, 数据到达率大幅提高; 以用户需求为导向, 提升了用户对于数据产品质量和空间天气服务的满意度。在接下来的工作中, 将继续通过PDCA循环的内部监督检查来识别并改进业务中存在的问题, 不断提升空间天气监测、预报、服务水平。

深入阅读

- 郭建广, 张效信, 2011. 国际上的空间天气计划与活动. 气象科技进展, 1(4): 18-25.
- 陆梅, 2017. ISO9001:2015质量管理体系审核员培训教程. 北京: 中国标准出版社: 17-29.
- 王洁, 2018. ISO9001质量管理体系在质检工作中的应用. 中国标准化, (4): 167-169.
- 王亮, 姬翔. 2020. 拓路开局, 做观测质量守护者. 中国气象报, 2020-6-15(3).
- 王劲松, 吕建永, 2010. 空间天气. 北京: 气象出版社.
- 王劲松, 2011. 中国气象局的空間天气业务. 气象科技进展, 1(4): 6-12.
- 王延东, 2019. ISO9001标准在气象观测领域的探索与实践. 质量与认证, (9): 67-68
- Milgram L, Spector A, Treger M, 1999. Managing Smart. Boston: Gulf Professional Publishing.
- WMO, 2013. Guide to the Implementation of a Quality Management System for National Meteorological and Hydrological Services. Geneva: WMO.

(作者单位: 朱琳玲, 中国气象局空间天气重点开放实验室/国家卫星气象中心(国家空间天气监测预警中心)/许建民气象卫星创新中心/中国气象局综合观测司; 郭建广、敦金平、吕景天、于超、赵明现、叶茜, 中国气象局空间天气重点开放实验室/国家卫星气象中心(国家空间天气监测预警中心)/许建民气象卫星创新中心; 张建磊、李颖冲, 中国气象局综合观测司)

(上接94页)

- Applications, 2002 311(3): 590-614.
- [25] Onel S, Zeid A, Kamarthi S. The structure and analysis of nanotechnology co-author and citation networks. *Scientometrics*, 2011, 89(1): 119-138.
- [26] NF Naranjo-Estupiñán, Mora Q J, D Jaimés-Vega, et al. Co-authorship research networks in public health in Santander. *Biomédica Revista Del Instituto Nacional De Salud*, 2014, 34(2): 300-307.
- [27] Uddin S, Hossain L, Abbasi L, et al. Trend and efficiency analysis of co-authorship network. *Scientometrics*, 2012, 90(2): 687-699.
- [28] Goldstein M L, Morris S A, Yen G G. A Group-based Yule model for bipartite author-paper networks. *Physical Review E*, 2005, 71: 026108.
- [29] Ramasco J J, Dorogovtsev S N, R Pastorsatorras. Self-organization of collaboration networks. *Physical Review E*, 2004, 70: 036106.
- [30] 方锦清, 汪小帆, 郑志刚, 等. 一门崭新的交叉科学: 网络科学(上). *物理学进展*, 2007, 27(3): 239-343.
- [31] Holme P, Park S M, Kim B J, et al. Korean university life in a network perspective: Dynamics of a large affiliation network. *Physica A: Statistical Mechanics and Its Applications*, 2007, 373(36): 821-830.
- [32] Girvan M, Newman M E J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 2002, 99(12): 7821-7826.
- [33] Freeman L C. A set of measuring centrality based on betweenness. *Sociometry*, 1977, 40(1): 35-41.
- [34] Barthélemy M. *The European Physical Journal B*, 2004, 38: 163-168.
- [35] 刘杰, 陆君安. 一个小型科研合作复杂网络及其分析. *复杂系统与复杂性科学*, 2004, 1(3): 56-61.
- [36] 顾东蕾. 药物基因组学科研合作网络的统计学性质. *中华医学图书情报杂志*, 2008, 17(6): 7-10.
- [37] 张大伟, 薛惠锋, 寇晓东. 复杂网络领域科学合作状况的网络分析研究. *情报杂志*, 2008, 27(8): 143-145.
- [38] 盛骤, 谢式千, 潘承毅. *概率论与数理统计(第4版)*. 北京: 高等教育出版社, 2010.
- [39] Liu Z H, Lai Y C, Ye N, et al. Connectivity distribution and attack tolerance of general networks with both preferential and random attachments. *Physics Letters A*, 2002, 303: 337-344.
- [40] Li X, Chen G R. A local world evolving network model. *Physica A: Statistical Mechanics and its Applications*, 2003, 328(1-2): 274-286.
- [41] Zhang P P, Chen K, He Y, et al. Model and empirical study on some collaboration networks. *Physica A: Statistical Mechanics and Its Applications*, 2006, 360: 599-616.
- [42] Chang H, Su B B, Zhou Y P, et al. Assortativity and act degree distribution of some collaboration networks. *Physica A: Statistical Mechanics and Its Applications*, 2007, 383: 687-702.
- [43] Zhou T, Wang B H, Jin Y D, et al. Modelling collaboration networks based on nonlinear preferential attachment. *International Journal of Modern Physics C*, 2007, 18(2): 279-314.
- [44] 刘水晶, 钱霞, 冯雅超, 等. 4个合作-竞争系统的特殊合作性质. *科技导报*, 2008, 26(21): 79-84.
- [45] 张培培, 何阅, 周涛, 等. 一个描述合作网络顶点度分布的模型. *物理学报*, 2006, 55(1): 60-67.
- [46] Mitchell R L. Permanence of the log-normal distribution. *Journal of the Optical Society of America (1917-1983)*, 1968, 58(9): 1267-1272.
- [47] Reed W J, Jorgensen M. The double pareto-lognormal distribution—a new parametric model for size distributions. *Communications in Statistics-Theory and Methods*, 2004, 33(8): 1733-1753.
- [48] Albert R, Barabási A L. Statistical mechanics of complex networks. *Review of Modern Physics*, 2002, 74: 47-97.