

合作网络高阶聚集特性研究

杨忠明¹, 梁本来², 秦勇³, 卢庆武⁴

- (1. 广东科学技术职业学院 计算机工程技术学院, 广东 珠海 519090;
2. 中山职业技术学院 信息工程学院, 广东 中山 528404;
3. 东莞理工大学 计算机学院, 广东 东莞 523808;
4. 惠州学院 教育技术中心, 广东 惠州 516007)

摘要:情报学领域的研究表明,研究工作者形成的科研合作关系网络是一个非连通网络,但是网络统计参数呈现出无尺度网络特性和小世界网络特性,合作网络是一种描述科研合作的典型网络。提出高阶聚集系数的概念,以期刊作者的合作网络数据为计算案例,计算期刊作者的高阶聚集系数 $C(h)$,形成图表并分析其高阶聚集系数 $C(h)$ 在图中所表示的意义。通过采集国内计算机类一级学报的大量合作者样本数据并通过计算分析表明,合作者为20人左右时,合作团队拥有广泛的知识面和较高的合作水平,较易做出高水平的创新性成果,也意味着高水平的论文作者通常和学者20人左右直接或者间接保持合作关系。高阶聚集系数同样可用于语言网络等其他呈现出小世界效应的网络分析。

关键词:合作网络;小世界网络;高阶聚集系数;平均距离

中图分类号:TP391.1

文献标识码:A

文章编号:1673-629X(2017)02-0200-05

doi:10.3969/j.issn.1673-629X.2017.02.046

Study on Characteristics of High Order Clustering for Collaboration Network

YANG Zhong-ming¹, LIANG Ben-lai², QIN Yong³, LU Qing-wu⁴

- (1. College of Computer Engineering and Technique, Guangdong Polytechnic of Science and Technology, Zhuhai 519090, China;
2. College of Information Engineering, Zhongshan Polytechnic, Zhongshan 528404, China;
3. College of Computer Science, Dongguan University of Technology, Dongguan 523808, China;
4. Information Technology Center, Huizhou University, Huizhou 516007, China)

Abstract: The research made in the field of information science shows that scientific collaboration relationship network formed by the researchers is unconnected. While the characteristics of scale-free network and small-world network presented from the network statistical parameter shows that collaboration network is a typical network describing the scientific collaboration research. The concept of high order clustering coefficient is presented. With the use of network data of journal authors, it will calculate journal collaborators' high order clustering coefficient $C(h)$, and form a chart and analyze the significance of $C(h)$ in the chart. The analysis of a large amount of collaborator's sample data collected from the first-class domestic journal of computer shows that when the number of collaborators is around 20, the cooperation team has extensive knowledge and a high level of cooperation and it is easier for the team to make a high level of innovative achievement. That means a high-level author will usually keep a cooperative relationship with about 20 scholars directly or indirectly.

Key words: cooperation networks; small world network; high order clustering coefficient; average distance

1 概述

网络可以用来描述人与人之间的社会关系、计算机之间的网络联接、网页之间的超链接、科研文章之间

的引用关系等,而统计物理学是从微观到宏观的桥梁。研究网络中顶点与边的度值与权值等微观性质与网络的几何性质、效率与稳定性等宏观性质之间的关系正

收稿日期:2016-03-06

修回日期:2016-06-15

网络出版时间:2016-11-22

基金项目:国家自然科学基金资助项目(61170193);广东省工业高新技术领域科技计划项目(2013B010401036);广东省高等学校优秀青年教师培养计划项目(YQ2014187);广东省自然科学基金项目(S2013010015940);广东省教育厅科技创新项目(2013KJCX0178)

作者简介:杨忠明(1980-),男,硕士,副教授,CCF会员,研究方向为计算机网络、智能算法。

网络出版地址:<http://www.cnki.net/kcms/detail/61.1450.TP.20161122.1227.018.html>

是复杂网络研究的核心内容。因而,与图论的研究有所不同,复杂网络的研究更侧重于从各种实际网络的现象之上抽象出一般的网络几何量,并用这些一般性质指导更多实际网络的研究,进而通过讨论实际网络上的具体现象发展网络上模型的一般方法,最后讨论网络本身的形成机制。在模型研究、演化机制与结构稳定性方面的丰富研究经验是在复杂网络研究领域得到广泛应用的原因;而图论与社会网络提供的网络静态几何量及其分析方法是复杂网络研究的基础,并得到了充分发展。合作网络属于复杂网络中社会网络的一种,它的提出是复杂网络系统研究的一个新的方向^[1]。小世界网络中利用聚集度的计算来评价网络指标、反映网络性质的合作网络已成为评价网络指标的一种重要评价方法。根据权威文献表明,合作网络是一种优化评价网络指标的评价系统,自1993年到2010年,持续多年的研究更表明了合作网络的评价优化。科学合作的范围随着科学的进步而逐渐增长。论文合著是科学合作的表现之一,随着不同领域的合作者的介入,科学合作成果的质量也越来越好,合作网络成为更复杂的网络。

一个典型网络是由许多节点与连接两个节点间的一些边组成的,而复杂网络简单来说呈现高度复杂性的网络,具有结构复杂、网络进化、连接多样性、动力学复杂性、节点多样性和多种复杂性融合的特点。复杂网络是具有自组织、自相似、吸引子、小世界、无标度网络中部分或全部性质的网络。合作网络是描述合作关系的网络,它属于复杂网络中社会网络的一种。

两个现实的复杂网络模型被广泛应用:小世界模型和无尺度模型。文中重点讨论小世界网络,其特征是局部集群连接的长的线,平均距离之间的顶点的对数的增长与网络中 N 的大小有关。以日常语言看,它反映的是相互关系的数目可以很小但却能够连接世界的事实,且小世界模型社会网络就是小世界的一种。科研人员合作网络是描述科研人员合作关系的网络,通常把每个科研人员作为网络中的一个顶点(或称节点),如果两个科研人员之间共同发表过一篇论文,这两个顶点之间就连接一条边。科研人员之间形成了一个复杂网络,该网络的节点的复杂性、网络演化特性以及网络的合作行为等问题,所呈现出的密集度可以作为评价期刊的一个指标。研究基于合作网络的期刊高阶聚集特性分析,通过其看聚集度的计算发现目前研究热点,同时聚集度越大,合作越紧密,可作为评价期刊的一个重要指标。

Watts 和 Strogatz^[2]构造出一种介于规则网络和随机网络之间的网络(WS网络)。随后 Newman 和 Watts^[3]给出一种新的网络构造方法。在 NW 网络中,

原有的连边并不会被破坏,而是以一个很小的概率在原来的规则网络上添加新的连边,这样构造出的网络也同时具有大的簇系数和小的平均距离。Barabasi A^[4]对科学合作社会网络的演化进行研究,认为作者合作社会网络是复杂网络的代表。

Watts 和 Strogatz 的研究方法是通过其中一个规则网络转化为小世界网络,这种方法很少或基本没有改变顶点度的分布和量化参数。小世界网络中一个重要特征就是与随机网络相比有相当高的聚集系数,甚至比任意网络的都要高。聚集系数的定义如下。顶点 v 在图 G 中有相邻节点度 $k(v)$, $k(v)$ 的相邻节点可能被连接成 $k(v)(k(v)-1)/2$ 条边。实际存在的最大可能的边的分数数量就是图 G 的聚集系数 C_v ,所有节点 v 的平均聚集系数就是网络 G 的聚集度 C 。 C 的值接近于1意味着网络是高通的。

确定性小世界网络的生成方式种类繁多,肖文俊等^[5]通过 Cayley 图建立了一个确定性小世界网络模型,基于 Cayley 图的小世界网络具有结构简单和明显的适应性。网络 $G=(V,E)$ 作为图论的概念是指由一个点集 $V(G)$ 和一个边集 $E(G)$ 组成的一个图,且 $E(G)$ 中的每条边 e_i 有 $V(G)$ 的一点对 (u,v) 与之对应。记顶点数 $N=|V|$,边数 $L=|E|$ 。如果任意 (u,v) 与 (v,u) 对应同一条边,则称为无向网络,否则为有向网络;如果任意 $|e_i|=1$,则称为无权网络,否则为加权网络。从统计物理学的角度来看,网络是一个包含大量个体及个体之间相互作用的系统,是把某种现象或某类关系抽象为个体(顶点)及个体之间相互作用(边)而形成的用来描述这一现象或关系的图^[6]。

国内外学者在合作网络的构造、静态统计性质分析和演化模型等方面进行了研究。2006年 Cardillo 等分析了 Los Alamos 的2000–2005年间 cond-mat 数据库,发现该库的成长是通过合作网络图的连接增长体现的,主要特征体现在度-度相关性以及节点的聚集相关性^[7]。周涛等提出了基于二分图资源的推荐算法^[8]。李晓佳等^[9]在自建的合作网络上,实现了层次聚集法和介数聚集法,在 Newman 的 Q 函数基础上讨论了聚集过程中的最佳集团数。还有学者从不同角度对社团网络及其应用做了研究,如范超等^[10]利用合作网络及社会网络的结构知识对 CNM 算法作了改进,提高了社团发现的精度。

黄开木等^[11]以近30年国内“竞争情报”领域作者合著发表期刊论文为研究对象,分析表明该领域科学家的产量、最大合著群体的产量和所有合著群体的规模都满足幂律分布。通过对该合著网络最大连通子图的统计分析,发现该合著群体具有高聚类和小世界性。

文中采集国内计算机领域一级学报的大量合作者

样本数据,分析期刊作者的合作网络数据,计算其高阶聚集特性,分析期刊的作者合作情况,把高阶聚集系数的实验结果作为评价和衡量期刊作者合作的重要指标。通过采集样本期刊的合作者数据,计算合作者的高阶聚集系数,分析期刊合作者情况,考虑以高阶聚集系数作为评价期刊优劣的指标之一。

文中提出的合作网络高阶聚集特征还可用于其他呈现出小世界效应的网络分析。在复杂网络的研究中,语言网络作为一个新的研究方向^[12],大都符合 WS 小世界模型,大多数节点只需经过少量的边便可到达。在聚集系数上,与随机语言网络相比,真实语言网络的聚集系数较高。Minett 等基于词典资源,根据汉语词汇中的共字关系构建了汉字网络^[13],由汉字构建的网络表现出明显的高聚集系数和无尺度特征。刘海涛对中文、英文等 15 种语言的依存句法网络进行聚类研究^[14],得出这 15 种语言网络均是小世界网络,但聚集系数存在显著差异,通过选择网络 7 组特征,对 15 种语言进行聚类,发现英语和汉语网络整体上较接近。

2 合作网络高阶聚集特征描述

2.1 聚集系数及其含义

定义一个网络或图表 $G = (V, E)$, 设 V 有 N 个顶点或节点, E 有 M 个边或线段, 每条边定义为一对顶点(有序配对的有向图)。让一个顶点 v 有相邻节点 $k(v)$, $k(v)$ 的相邻节点可能被连接成 $k(v)(k(v) - 1)/2$ 条边。分数最大的边事实上存在于相邻的 v 和聚集系数 C_v 之间, 平均聚集系数对所有 $v \in V$ 是那个聚集系数 C 的网络 G 。一个网络 C 接近 1 可能包括聚集和有高度的连接, 或稀疏的连接在本地群集中。

2.2 高阶聚集系数的定义

文中延伸 Watts 和 Strogatz 的聚集系数, 定义 h 是相邻顶点 v 与顶点 G 的距离。假设 v 有邻节点 h 和 $k_h(v)$, $k_1(v)$ 与 $k(v)$ 的定义见 2.1。然后最多可以有 $k_h(v)(k_h(v) - 1)/2$ 条边连接 h 的邻节点 v 。这个分数 $C_v(h)$ 允许的边真实存在于 h 邻节点 v 与聚集系数 h 之间。假设当 $k_h(v) = 1$ 时 $C_v(h) = 1$, 这也涵盖了 $h = 0$ 时的特例。平均 $C_v(h)$ 对所有 $v \in G$, 考虑高阶聚集系数的概念, 聚集系数为 h 时为 $C(h)$ 。聚集系数为 1 时 $C(1)$ 是聚集系数定义在 2.1 的 C 。

顶点为 N 的网络有 M 条边, 可得公式 $P(0) = 1/N$, $P(1) = 2M/N^2 > 1/N$ 。在 $h = 1$ 以外, 不能精确地表达该值为 $P(h)$, 除了以几个固定网络为例。很多网络(小世界或反之), $P(h)$ 的值与 h 一起上升直到最大值, 然后下降至距离 h 与网络直径 D 越来越接近。

对于高阶聚集系数, 这一趋势开始于聚集度减少, 从 $C(0) = 1$ 到 $C(1) = C$, 然后是进一步减少。由于事

实上 h 在增加, q_h 的节点数目按一定距离 h 的节点在增加, 这些节点属于几个派系; 因此, 它们之间存在很多边是不可能的。在 h 向 D 靠拢的过程中, 会出现不同的效果。一个极端的例子是其中每个节点在网络中到距离 D 的唯一节点(这有一个截然相反的单节点), 导致 $C(D) = 1$ 。同样的情况, $C(D - 1)$ 可能广泛应用于鉴别存在多个直径和相同的相对的顶点。应注意前面的意思是在 h 向 D 靠拢的过程中 $C(h)$ 会变大, 并非 $C(h)$ 一定要大。

因为 $P(h)$ (上升然后下降) 和 $C(h)$ (下降然后上升) 的趋势是相反的, 所以 $P(h) C(h)$ 应有一个上限值。因为一个常数 c 经常接近 1 但很少超过 1, 得出:

$$P(h) C(h) \leq c \log N / N \quad (1)$$

在这特例中 $h = 1$, 方程(1)意味着 $P(1)C(1) \approx \log N / N$ 。从小世界网络中知道 $P(1) = 2M/N^2 \approx \log N / N$ 。对于该类网络符合 $C(1) = C$ 。

目前最有效的补充证据在方程(1), 利用这个确定性的小世界网络提出这个模型。事实上, 通过这个模型, 可以发现这种趋势表现在方程(1)。反观确定性模型, 这是基于 Cayley 的图, 提供了模型所需的支撑信息, 从而可得该模型的聚集系数:

$$C = \frac{at(at-1)}{(at+t-l)(at+t-l-1)} \quad (2)$$

模型中, $t = \log_2 N$ 和 $a = (2^t - 1)/t$ 是自由调整参数, 关系到互连密度, 从而影响了 C 的价值。注意那个非常广泛的网络($N, t \rightarrow \infty$), 当 a 是常数时 C 趋向 $a^2/(a+1)^2$ 。适当选择 a , 能获得不同的聚集系数, 同时保持一个小的顶点使 $at + t - 1 = (a+1) \log_2 N - 1$ 。

不像现实网络中计算 $C(h)$ 非常困难, 文中提出的确定性模型服从数学分析, 能用封闭形式表达一个高阶聚集系数。在这个确定性模型中, 对于数字 m 相邻的顶点之间, 在相邻顶点 h 的任何顶点给出表达式:

$$m = (2^t - 1)(2^{t-1} - 1) \binom{t-l}{h-1} \quad (3)$$

数字 $k_h(v)$ 的相邻顶点 h 的顶点 v 是有界的:

$$(2^t - 1) \binom{t-l}{h-1} \leq k_h(v) \leq (2^t - 1) \binom{t-l}{h-1} + \binom{t-l}{h} \quad (4)$$

鉴于高阶聚集系数 $C(h)$ 与 $m/(k_h(v))^2$ 成正比, 很容易得到高阶聚集系数为:

$$C(h) \approx 1 / \binom{t-l}{h-1} \quad (5)$$

因此, 定义聚集系数是基于相邻的节点, 高阶聚集系数定义了涉及到一个更广泛的相邻距离参数 h 。利用实际上更广泛的复杂网络的实验数据, 依据文中提

出的确定性小世界网络模型,将网络中有关 $C(h)$ 和 $P(h)$ 的距离分布,定义为有一定概率的随机选择的距离为 h 的一对顶点。

高阶聚集系数是聚集系数的高度概括,并受聚集系统相关公式所控制。通过利用期刊作者所构成的网络数据进行实验和分析证明,不等式 $P(h) C(h) \leq c \log N / N$ 在小世界网络中成立。这一结果明显表明该乘积 $P(h) C(h)$ 有一个上限值,该上限值小于 $P(h)$ 和 $C(h)$ 的最大值。

高阶聚集系数的计算可以应用在期刊合作者上。把期刊合作者的数据进行高阶聚集系数 $C(h)$ 的计算,可以得出一组关于顶点距离、阶数、高阶聚集系数 $C(h)$ 和 $C_d * P_d$ 等的的数据。通过实验把数据制作成图表进行分析。期刊合作者的高阶聚集系数能得出一组曲线图,而这组曲线图的高峰能表明这时合作者人数在期刊中的普遍性,以及此时的合作者人数能做出较好的科研成果。也可以根据曲线图的高峰段来判断合作者的作品情况,可作为期刊的评价方法之一。

3 实验

文中主要延伸 Watts 和 Strogatz 的聚集系数,计算期刊合作者的高阶聚集系数 $C(h)$,形成图表并分析高阶聚集系数 $C(h)$ 所表示的意义。

通过中国期刊网采集了《计算机研究与发展》(2000-2010)2 610 篇论文共 2 764 位作者数据、《计算

机学报》(2000-2010)2 109 篇论文共 1 647 位作者数据和《软件学报》(2000-2010)2 602 篇论文共 2 462 位作者数据。通过分类整理,分别把《计算机研究与发展》、《计算机学报》和《软件学报》三大期刊的作者分别输出,排列整理成文本数据,利用这些数据并通过延伸 Watts 和 Strogatz 的聚集系数,形成高阶聚集系数 $C(h)$ 以计算期刊合作者的高阶聚集系数。高阶聚集系数能反映出期刊作者在不同阶段的聚集度。通过高阶聚集系数的计算与分析,能从结果中判断出期刊作者的密集性。通过输出的结果把三大期刊三组数据分别进行计算并作图,再进行比较。

3.1 高阶聚集系数的计算方法与环境

2.3 节推导了高阶聚集系数的过程,利用此方法设计程序进行高阶聚集系数的计算。对实验数据首先计算最大连通子图,计算出顶点数与平均距离,继而计算各距离的点对数。根据平均距离数计算聚集度。

文中从中国期刊网将《计算机研究与发展》、《计算机学报》和《软件学报》(2000-2010)的作者信息进行数据采集,分别整理成 3 个单独存放作者信息的文档,再把论文是两位以上的作者进行分离和融合,形成每行为两位作者名称的格式。然后利用高阶聚集系数计算程序对 3 大期刊的作者数据进行计算。最后把数据信息分别整理成图表。

3.2 实验结果

计算结果如图 1 所示。

计算机研究与发展				计算机学报				软件学报			
$C(h)$	distance distribution (P)	Product C^*P	distance h	$C(h)$	distance distribution (P)	Product C^*P	distance h	$C(h)$	distance distribution (P)	Product C^*P	distance h
0.752	0.001 441 4	0.001085	1	0.74769	0.0024287	0.001815	1	0.75812	0.0016693	0.001265	1
0.244	0.005 484 1	0.001339	2	0.26009	0.008504	0.002211	2	0.21813	0.00746	0.001627	2
0.174	0.009 054 2	0.001584	3	0.18986	0.0134056	0.002545	3	0.15216	0.0126137	0.001919	3
0.109	0.014 813	0.001626	4	0.12905	0.0187424	0.002418	4	0.08047	0.0222483	0.001790	4
0.074	0.023 724 2	0.001760	5	0.09915	0.0262763	0.002605	5	0.04580	0.0360278	0.001650	5
0.046	0.034 998 3	0.001618	6	0.07985	0.0372038	0.002970	6	0.03379	0.051511	0.001741	6
0.030	0.050 070 9	0.001532	7	0.05340	0.0485437	0.002592	7	0.02378	0.0706178	0.001679	7
0.023	0.067 402 5	0.001553	8	0.04106	0.0557994	0.002291	8	0.01676	0.0913807	0.001532	8
0.017	0.084 241 9	0.001447	9	0.03354	0.0628618	0.002108	9	0.01239	0.1080545	0.001339	9
0.012	0.095 687 6	0.001223	10	0.02277	0.0734883	0.001673	10	0.01283	0.1171147	0.001503	10
0.009	0.100 241	0.000917	11	0.01821	0.0842705	0.001534	11	0.01014	0.1202001	0.001219	11
0.007	0.099 172 3	0.000792	12	0.01740	0.0866371	0.001507	12	0.00895	0.1104743	0.000988	12
0.007	0.091 241 1	0.000715	13	0.01671	0.0828053	0.001384	13	0.01293	0.0871836	0.001127	13
0.009	0.078 931 4	0.000737	14	0.01831	0.0786732	0.001440	14	0.02617	0.0636784	0.001666	14
0.012	0.066 451 6	0.000807	15	0.02163	0.068603	0.001484	15	0.05349	0.0414373	0.002216	15
0.017	0.054 188 8	0.000952	16	0.02950	0.0554711	0.001636	16	0.09528	0.0247464	0.002357	16
0.028	0.041 957 9	0.001185	17	0.04195	0.0424794	0.001782	17	0.10598	0.0148309	0.001571	17
0.053	0.029 489 9	0.001578	18	0.05868	0.0333299	0.001956	18	0.10949	0.0084097	0.000920	18
0.064	0.020 369 2	0.001314	19	0.08708	0.0259384	0.002258	19	0.09244	0.0047741	0.000441	19
0.093	0.013 224 9	0.001234	20	0.10907	0.0192898	0.002104	20	0.07019	0.0026658	0.000187	20
0.103	0.008 191	0.000846	21	0.12810	0.015562	0.001993	21	0.04917	0.0013986	0.000068	21
0.068	0.004 700 8	0.000319	22	0.13921	0.0132012	0.001837	22	0.02972	0.0007899	0.000023	22
0.044	0.002 552 1	0.000112	23	0.13590	0.0111311	0.001512	23	0.01334	0.0004334	0.000005	23
0.026	0.001 285 6	0.000034	24	0.13399	0.0081233	0.001088	24	0.00432	0.0002165	0.000000	24
0.022	0.000 690 9	0.000015	25	0.11054	0.0070801	0.000782	25	0.00155	0.0000515	0.000000	25
0.006	0.000 277 3	0.000001	26	0.10568	0.0059964	0.000633	26	0.00025	0.0000076	0	26
0.010	0.000 093 2	0.000001	27	0.10288	0.0044501	0.000457	27	0.00005	0.000004	0	27
0.006	0.000 021 2	0.000000	28	0.08411	0.0034372	0.000289	28				
0.000	0.000 001 6	0	29	0.05889	0.0023143	0.000136	29				
				0.05490	0.0013663	0.000075	30				
				0.04586	0.0010756	0.000049	31				
				0.02859	0.0008005	0.000022	32				
				0.01674	0.0003504	0.000005	33				
				0.00972	0.0002272	0.000002	34				
				0.00449	0.0001151	0.000000	35				
				0.00207	0.0000177	0	36				

图 1 高阶聚集系数计算结果

计算得出的图呈现驼峰分布型,即会出现 2 次高峰,在 3~5 的作者点集中均出现 1 次高峰,在后期点集中会出现 2 次高峰。此 2 次高峰将会说明作者合作群的合作关系是比较容易做出高水平创新研究成果的。

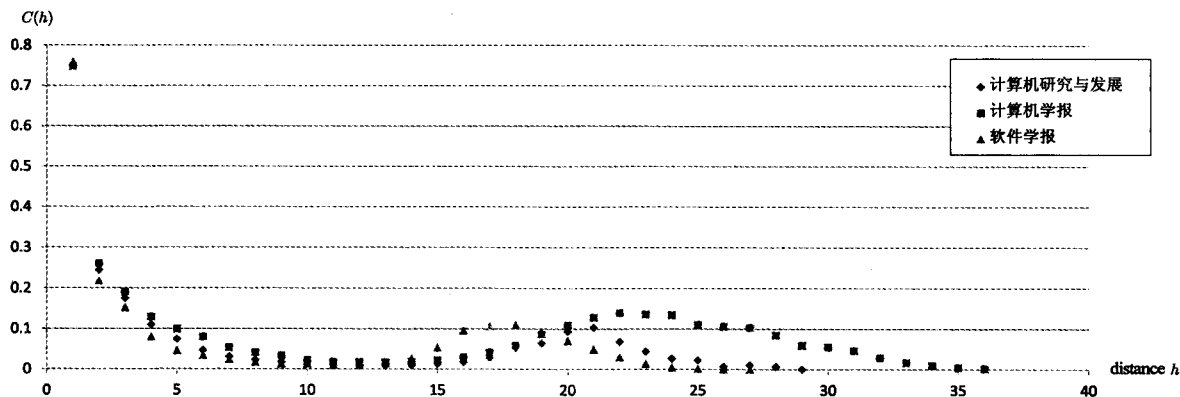


图 2 三大期刊高阶聚集系数分布情况

3.3 实验分析与讨论

从图 2 中可以看出,图中出现了两个高峰。在合作者为 2~3 人的平均距离实验结果中出现第一高峰,即 2~3 人的合作情况比较普遍。随着合作者数量增加高阶聚集系数减小,但是合作者 20 人左右时出现第二聚集系数高峰。此高峰表明合作者 20 人左右时,合作团队拥有广泛的知识面和较高的合作水平,较易做出高水平的创新性成果,也意味着高水平的论文作者通常和学者 20 人左右直接或者间接保持合作关系。可以看出,这三组数据的第二高峰都处于相似的阶段,证明了高水平的论文作者和学者 20 人左右直接或间接保持合作关系。

4 结束语

文者提出的高阶聚集系数是聚集系数的高度概括,并受聚集系数相关公式所控制。通过利用期刊作者所构成的网络数据进行了实验分析,表明高阶聚集系数提供了对小世界网络结构的新见解,为小世界网络的探索提供了一个新的渠道。

通过实验数据结果可以看出,高阶聚集系数越高,期刊合作者合作关系越紧密,越能做出有水平的研究成果。2~3 人的合作网络比较普遍,但并不意味着 2~3 人的合作者模式容易做出高水平成果。通过第二聚集系数高峰进行分析更具说服力,在合作者规模为 20 左右的合作关系比较容易做出高水平创新研究成果。文中所提出的高阶聚集系数还可应用于语言网络等其他呈现小世界效应的网络分析。

参考文献:

[1] 高霞,陈凯华.合作创新网络结构演化特征的复杂网络

三大期刊的合作者数据构成了合作网络,图 2 显示了三大期刊高阶聚集系数的分布情况:点集表示期刊合作者的高阶聚集系数。

分析[J].科研管理,2015,36(6):28-36.

- [2] Watts D J, Strogatz S H. Collective dynamics of small word networks[J]. Nature, 1998, 393(6684): 440-442.
- [3] Newman M E J, Watts D J. Scaling and percolation in the small-world network model[J]. Physical Review E, 1999, 60: 7332-7342.
- [4] Barabasi A, Jeong H, Neda Z, et al. Evolution of the social network of scientific collaborations[J]. Physical A, 2002, 311(3-4): 590-614.
- [5] Xiao Wenjun, Parhami B. Cayley graphs as models of deterministic small-world networks[J]. Information Processing Letters, 2006, 97(3): 115-117.
- [6] 吴金闪,狄增如.从统计物理学看复杂网络研究[J].物理学进展, 2004, 24(1): 18-46.
- [7] Cardillo A, Scellato S, Latora V. A topological analysis of scientific coauthorship networks[J]. Physica A Statistical Mechanics & Its Applications, 2006, 372(2): 333-339.
- [8] Zhou Tao, Red J, Medo M, et al. Bipartite network projection and personal recommendation[J]. Physical Review E Statistical Nonlinear & Soft Matter Physics, 2007, 76(2): 70-80.
- [9] 李晓佳,张鹏,狄增如,等.复杂网络中的社团结构[J].复杂系统与复杂性科学, 2008, 5(3): 19-42.
- [10] 范超,王厚峰.社交网络中的社团结构挖掘[J].中文信息学报, 2014, 28(1): 56-63.
- [11] 黄开木,樊振佳,卢胜军,等.我国竞争情报领域期刊论文合著网络研究[J].情报杂志, 2015, 34(2): 142-147.
- [12] 韩普,王东波,路高飞,等.语言网络研究进展[J].中文信息学报, 2014, 28(1): 9-18.
- [13] Minett J W, Peng G, Wang W S. The networks of syllables and characters in Chinese[J]. Journal of Quantitative Linguistics, 2008, 15(3): 243-255.
- [14] 刘海涛.语言复杂网络的聚类研究[J].科学通报, 2010, 55(27-28): 2667-2674.