

基于社会网络分析的合成生物学合著网络研究

刘 娅

(中国科学技术信息研究所, 北京 100038)

摘 要:以美国科学引文索引数据库扩展版数据库(SCI-EXPANDED)收录,发表于2006—2011年的2 119篇合成生物学相关的论文题录为对象,通过对著者所属国家/机构所构成的合著关系分析,在一定程度上揭示了合成生物学这一新兴领域中不同国家/机构在科研合作中所发挥的作用。研究发现:大部分论文发表数量较多的国家/机构在合著网络中处于较核心的地位,少数国家/机构具有非常明显的合作优势,而部分国家/机构的在合作网络中的地位与其论文数量规模并不相称。同时,该领域的合作中国家/机构层面均存在一些合作关系紧密的小团体。

关键词:合成生物学; SCI论文; 社会网络分析; 合著网络

中图分类号: C91-03; G353.1; Q81 **文献标识码:** A **DOI:** 10.3772/j.issn.1009-8623.2013.02.005

合成生物学(synthetic biology)是生物学研究的新兴领域,主要采用“自下而上”或“自上而下”的思路,通过人工的“设计”或“重设计”,构建具有全新功能或改善功能的生命系统,以服务于人类社会不断发展的需要^[1]。

2010年12月,合成生物学在《科学》杂志评出的十大科学突破中排第2位,在《自然》杂志盘点的2010年12件重大科学事件中排第4位。目前,很多国家正在此领域纷纷布局,加大对相关研究的支持力度^[2]。

为了全面把握合成生物学研究的发展现状,对该领域的研究网络进行分析非常必要。通过分析可以系统性地增加人们对参与主体之间关系“模式”或“结构”的洞见。

科研论文作为记录科学技术活动成果的一种重要载体,能够在一定程度上通过著者情况反映出科研主体之间的合作关系^[3-5]。有鉴于此,以社会网络分析相关理论为指导,开展合成生物学研究科研论文的合著分析,以揭示合成生物学研究网络的构成,丰富人们对于不同领域科学技术活动的知识交流与扩散的认识。

1 研究方法数据来源

1.1 研究方法

研究方法拟采用社会网络分析(Social Network Analysis, SNA)方法。SNA是社会学领域内根据数学方法、图论等发展起来的一套比较成熟的定量分析方法,用以分析不同社会单元所构成网络中社会关系的结构及其属性^[6-9]。

基于科研论文合著关系,构建一个群体内成员的相互关系网,并对其进行研究和讨论,所有分析均是对1-模网络的分析。考虑研究数据来源于科研论文,而目前大数据量前提下科研论文中作者重名的辨识问题国内外尚未有高效解决方案,因此,研究范围限定在国家和机构两级。

研究中,网络结构采用社群图法和矩阵代数法2种方法来表示。社群图法中,节点代表网络中的成员(论文所属国家/科研机构),节点间连线代表成员之间的合著关系。合著关系没有方向性,但有数量多少,因此社群图是赋值的无向图。在矩阵法中,用矩阵行和列代表网络成员(论文所属国家/科研机构),矩阵元素的值代表成员之间的关

作者简介:刘娅(1970—),女,硕士,研究员,主要研究方向为科技政策与管理。

收稿日期:2013-01-07

系。具体分析时,开展考虑关系强度的分析,采用赋值矩阵;开展不考虑关系强度的分析,采用二值化邻接矩阵。

研究分析的内容包括:

(1) 网络构成分析。计算网络成员构成、网络规模大小、网络密度等基本属性指标。

(2) 中心性分析。对网络中成员的点度中心度、中间中心度以及接近中心度的进行测度。点度中心度用以衡量成员是否处于网络的中心地位;中间中心度测度成员对网络中资源的控制能力;接近中心度测度网络中成员不受其它成员控制的程度。

(3) 凝聚子群分析。发现网络中具有紧密关系的小群体,分析网络中小群体的合作范围广度。

具体计算和可视化处理通过社会网络分析软件 UCINET 6.0 完成^[10]。

1.2 数据来源与处理

选取 Thomson Reuters 公司的科学引文索引数据库扩展版 (SCI-EXPANDED) 为数据来源。鉴于 SCI-EXPANDED 并没有设立直接对应合成生物学的学科类别,同时,目前合成生物研究自身尚未形成较为成熟的研究体系,因此,经过对合成生物学研究综述性文献的梳理,以及对当前研究活动的跟踪并参考专家意见,确定构建一个比较客观、但覆盖内容相对宽泛的检索策略——检索主题需体现该领域较为集中的主要研发活动;同时,由于当前合成生物学相关应用仍处于研究探索的较早期阶段,因此宜使用对该领域研究成果具有一定概括性的检索词来定位潜在应用研究,以覆盖较多的应用领域^[11-14]。多次测试后最终确定检索条件如下:入库时间为 2006—2011 年,发表语言为英语,文章类型为论文 (Article)。检索主题包含:“synthetic biology”或“gene circuit”或“gene circuits”或“genetic circuit”或“genetic circuits”或“genetic device”或“genetic devices”或“synthetic life”或“synthetic lives”或“synthetic tissue”或“synthetic tissues”或“synthetic cell”或“synthetic cells”或“synthetic genome”或“synthetic genomes”或“synthetic gene”或“synthetic genes”或“minimal genome”或“minimal genomes”或“essential gene”或“essential genes”;检索字段包括:题名、文摘、作者关键词和附加关键词。

检索时间为 2012 年 4 月 9 日—2012 年 4 月 11 日,共获得 SCI 论文题录数据 2 119 条。利用 Thomson Reuters 公司的 TDA 软件对下载的题录数据进行处理。首先,利用软件的自动清洗功能,对论文所属“国家”以及“机构”字段进行清洗;随后,采用人工方式对“机构”字段的清洗结果进行修正。在此基础上,根据合著关系分别建立国家和机构的赋值正方阵。

2 合成生物学研究合著网络

2.1 国家合著网络

2.1.1 网络构成

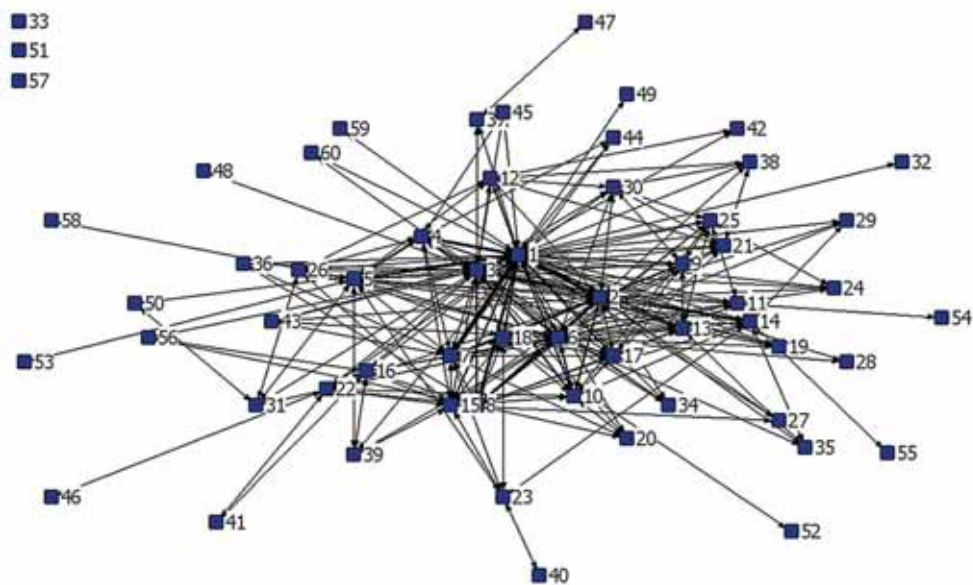
2006—2011 年产出的 2 119 篇 SCI 合成生物学论文由 60 个国家的科研人员参与撰写,这些论文的合著关系构成了反映合成生物学研究国家合著状况的无向网络,见图 1 所示。

由图 1 可见,2006—2011 年合成生物学的国家网络规模不算很大,网络成员 (节点) 共 60 个,其中有 3 个孤立点,即 33 号罗马尼亚、51 号智利和 57 号突尼斯;其他 57 个节点均存在连线关系 (连线数 454 条),即这 57 国家具有合著关系。由此可见,国家间合作研究在合成生物领域研究网络中是较为普遍的。

对 57 个有连线关系节点组成网络的二值邻接矩阵进行计算 (国家合著多值正方阵经过二值化处理,只反映是否存在合作关系而不反映合作强度,“1”表示具有合著关系,“0”表示没有合著关系,2.2.1 相同),得到网络整体的密度值为 0.518 8,各节点之间的平均捷径距离为 2.151,建立在距离基础上的凝聚力指数为 0.524,分裂指数为 0.476。网络中大多数国家之间的捷径距离 ≤ 2 (成员间捷径距离是 2 的情况出现 1 884 次,捷径距离是 1 的情况出现 454 次),因此,总体上两个国家之间平均只要通过一个国家就可以建立联系,网络整体的凝聚性较好 (见图 2 所示)。

2.1.2 中心性分析

57 个关联节点的二值邻接矩阵中心性计算结果见图 3 (数据均为标准化处理后的结果),Degree 为点度中心度,Closeness 为接近中心度 (UCINET 软件 Closeness Centrality 模块的 nCloseness 指标而非 Farness 指标),Betweenness



国家代码：
1. 美国；2. 英国；3. 德国；4. 中国；5. 日本；6. 法国；7. 加拿大；8. 西班牙；9. 意大利；10. 瑞士；11. 印度；
12. 韩国 13. 以色列；14. 荷兰；15. 澳大利亚；16. 巴西；17. 丹麦；18. 瑞典；19. 俄罗斯；20. 墨西哥；21. 奥地利；22. 阿根廷；
23. 伊朗；24. 爱尔兰；25. 比利时；26. 新加坡；27. 新西兰；28. 波兰；29. 葡萄牙；30. 捷克；31. 马来西亚；32. 芬兰；
33. 罗马尼亚；34. 匈牙利；35. 希腊；36. 挪威；37. 埃及；38. 斯洛伐克；39. 泰国；40. 土耳其；41. 斯洛维尼亚；42. 越南；
43. 南非；44. 巴基斯坦 45. 菲律宾；46. 哥伦比亚；47. 卡塔尔；48. 克罗地亚；49. 冰岛；50. 沙特阿拉伯；51. 智利；52. 立陶宛；
53. 孟加拉；54. 斯里兰卡；55. 印度尼西亚；56. 玻利维亚；57. 突尼斯；58. 保加利亚；59. 古巴；60. 阿拉伯联合酋长国

图 1 2006—2011 年 SCI 合成生物学论文发表的国家合著网络

Average distance = 2.151
Distance-based cohesion ("Compactness") = 0.524
(range 0 to 1; larger values indicate greater cohesiveness)
Distance-weighted fragmentation ("Breadth") = 0.476
Frequencies of Geodesic Distances

	Frequency	Proportion
1	454.000	0.142
2	1884.000	0.590
3	774.000	0.242
4	78.000	0.024
5	2.000	0.001

图 2 合成生物学论文国家合作网络中 57 个具有合著关系成员之间距离计算结果

为中间中心度，Eigenvector 为特征向量中心度。考虑数据规模，图 3 仅列出了国家代码在前 30 位的节点。

计算显示，1 号美国在 4 个指标上的表现均远远高于其他国家。其点度中心度为 73.214，这表明网络中美国与其他国家的直接联系最多（41 个国家），因此，具有很强的发展交往的能力；美国

的接近中心度达到 77.778，说明其受网络中其他成员控制的程度很低；再者，美国通过中间中心度表现出来对资源控制的能力也最强（33.728）；最后，特征向量中心度的数据表现也显示，在网络整体结构意义上美国是最核心的成员。上述不同评价指标所得结果的一致性说明，美国是当前世界合成生物学研究的合著网络中最重要、

Normalized Centrality Measures					
		1	2	3	4
		Degree	Closeness	Betweenness	Eigenvector
		-----	-----	-----	-----
1	1	73.214	77.778	33.728	49.249
2	2	46.429	62.921	6.109	40.025
3	3	62.500	71.795	20.668	46.470
4	4	23.214	52.830	0.884	24.547
5	5	33.929	58.333	9.852	30.128
6	6	50.000	64.368	7.752	41.611
7	7	32.143	57.143	3.158	31.197
8	8	26.786	54.902	1.788	26.659
9	9	21.429	52.830	0.314	25.517
10	10	21.429	53.333	0.611	23.667
11	11	16.071	51.852	3.747	17.899
12	12	16.071	50.909	0.845	16.044
13	13	21.429	52.336	0.680	24.389
14	14	21.429	53.846	5.099	21.434
15	15	25.000	54.369	2.399	24.788
16	16	17.857	51.852	2.340	16.714
17	17	21.429	53.333	0.295	26.171
18	18	26.786	55.446	4.971	26.546
19	19	8.929	47.458	0.000	13.767
20	20	8.929	47.863	0.015	12.095
21	21	17.857	50.909	0.274	20.286
22	22	17.857	51.376	5.375	18.620
23	23	10.714	44.094	3.626	9.193
24	24	8.929	48.696	0.033	12.407
25	25	17.857	50.000	0.318	19.530
26	26	14.286	50.450	0.201	16.535
27	27	5.357	45.902	0.000	7.784
28	28	3.571	45.161	0.000	6.200
29	29	7.143	47.059	0.000	10.394
30	30	16.071	49.123	0.167	18.326

图 3 合成生物学论文国家合著网络中 57 个具有合著关系成员的中心性计算结果（部分）

最关键的国家。此外，3 号德国在 4 项指标上的表现也呈现较好一致性，各指标值均较高，因此，也是国家合著网络中非常重要的一个成员。

其他成员国家在 4 项指标表现上，各自没有呈现出类似上述 2 个国家的一致性，而是各有千秋。例如，同是网络中较核心成员的 6 号法国和 2 号英国，法国的个体行动能力较强，对其他网络成员依赖较低，但其控制网络资源的能力并不算很高；英国虽然对其他网络成员依赖较低，但其个体行动能力和控制资源的能力比法国要低一些。5 号日本，有较强的控制资源能力，但其并未在交往能力和不受控制能力方面体现出较多优势。4 号中国，虽然 SCI 论文数量达到 154 篇，但在中心性的各项指标上均没有上乘表现，说明其未在网络中处于核心地位。与中国相比，18 号瑞典，尽管 SCI 论文数量只有 28 篇，但其各项指标均超过了中国，因此在网络中反而处于较核心的地位（特征向量中心

度 26.546）。

此外，计算结果显示，包括俄罗斯、新西兰在内的 28 个国家的中间中心度为 0，说明相当数量国家基本不具备掌控网络资源的能力。

2.1.3 凝聚子群分析

研究中，合成生物学论文国家合著网络的凝聚子群分析采用建立在互惠性基础上的凝聚子群分析，即派系分析（Cliques）。由于派系是最大的完备子图，因此，派系中任何两个成员都具有合著关系。由于国家之间合著论文数量不一，因此考虑合著强度来分析 57 个国家的无向多值关系网。UCINET6.0 不能直接分析多值网络中的派系，但可以通过间接方式计算。采用 C 层次派系，设定临界值 C 为 2（合著论文数）以表示合著强度。在此基础上，对国家合著多值正方阵进行二值化处理和派系分析（派系最小规模设定为 3）。

对合成生物学论文国家合著网络的凝聚子群分

析结果见图4所示,网络中存在20个最小规模为3的2层次派系。

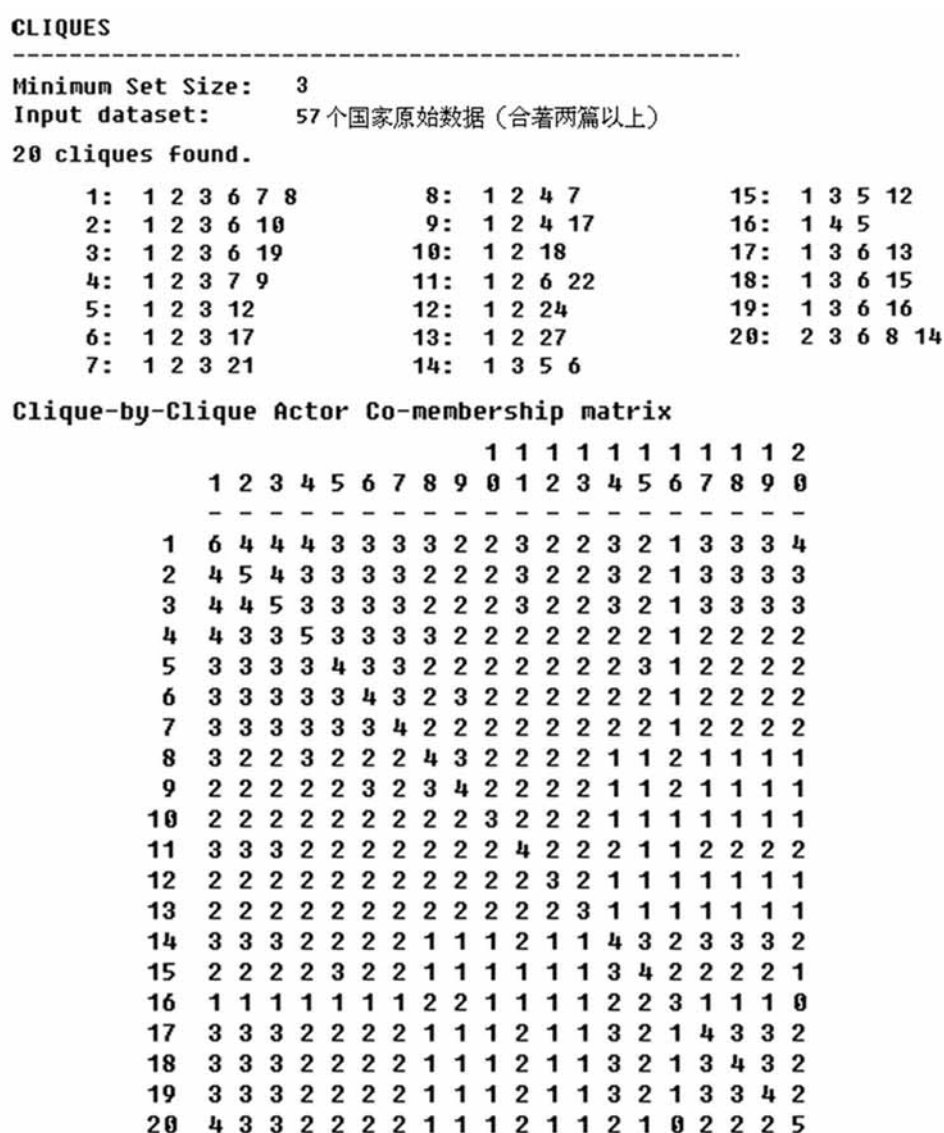


图4 合成生物学论文国家合著网络的2层次派系分析结果(部分)

图4显示的网络层次派系代表了合成生物学论文国家合作网络中具有较高凝聚力的小团体,其中任何一个派系中的任意两个成员的关系强度都大于2,即任何一个派系中两点所代表的国家间的合著论文均在3篇及以上。20个派系所包含的国家有19个,其余38个国家不属于任何最小规模为3的2层次派系。20个派系中,规模为4以上的派系16个,规模为5以上的派系5个,这表明国家合著网络中构成最小规模为3的2层次派系的国家的地域选择性不算太强。此外,分析20个派系两两共同拥有成员的情况,可以看到两个派系共

同拥有3个及以上成员的情况出现了51次,而未共同拥有成员的情况只有1次,这说明20个派系有较大的重叠性。

对各最小规模为3的2层次派系的成员构成进一步分析,美国出现在19个派系中,英国出现在14个派系中,德国出现在13个派系中,法国出现在9个派系中,这表明这些国家的合著圈较多,美国、德国和英国这3个国家是大多数派系的重要参与者。同时,还可以清楚看到美国和英国同时出现在13个派系中,美国和德国同时出现在了12个派系中,美国、德国和英国也同时出现在7个派系

中，说明这 3 个国家之间的联系非常紧密，它们之间无论是两方合作还是参与多方合作都表现活跃，知识共享程度较好。中国出现在派系 8、9 和 16 中，其合作圈内的国家主要是美国、英国、日本等几个国家。凝聚子群分析进一步显示出中国在合成生物学研究合作中的地域范围较窄。

2.2 机构合著网络

2006—2011 年共有 1 538 家机构发表了合成生物学研究 SCI 论文。为了使机构合著关系分析更具现实意义，以下分析对象选择为发表论文在 5 篇以上的 149 家机构。

2.2.1 网络构成

149 家机构 SCI 论文合著关系所构成的无向网

Average distance = 3.218
Distance-based cohesion ("Compactness") = 0.357
(range 0 to 1; larger values indicate greater cohesiveness)
Distance-weighted fragmentation ("Breadth") = 0.643

Frequencies of Geodesic Distances

	Freque	Proporti
	-----	-----
1	864.000	0.041
2	4368.000	0.206
3	7968.000	0.376
4	5690.000	0.269
5	1874.000	0.089
6	358.000	0.017
7	46.000	0.002
8	2.000	0.000

图 6 合成生物学论文机构合著网络中 146 个具有合著关系成员之间距离计算结果

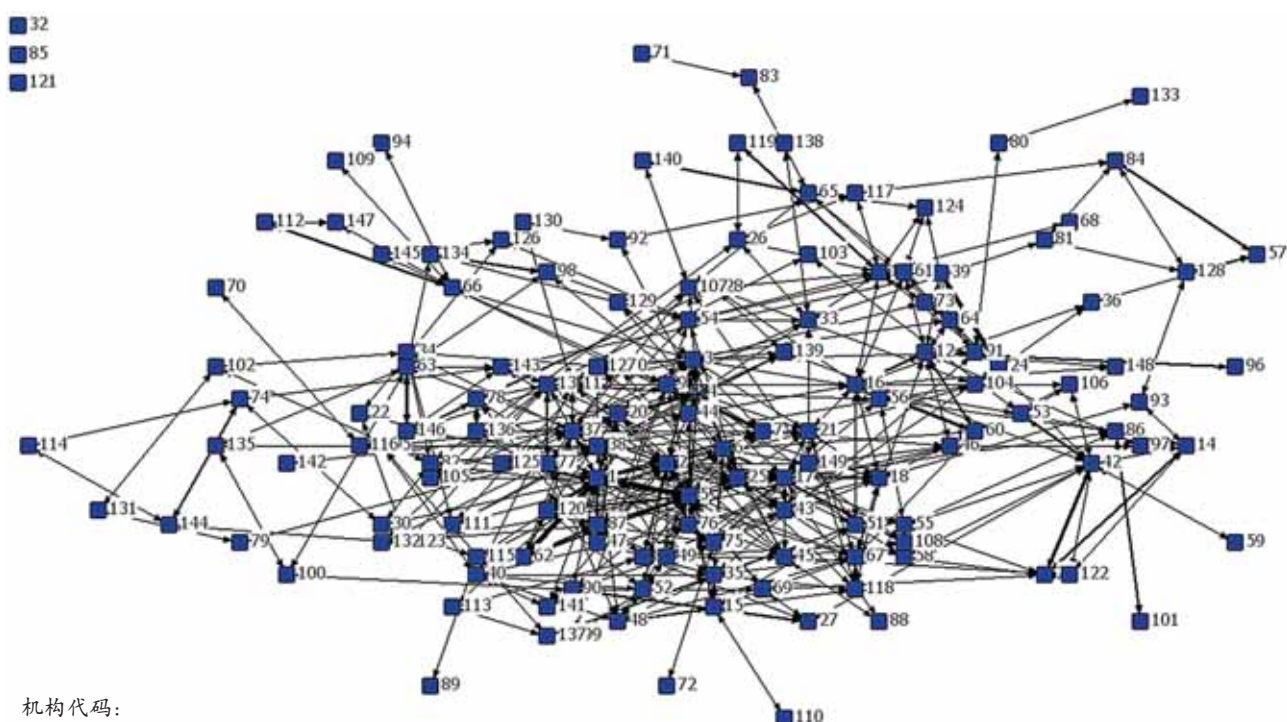
2.2.2 中心性分析

对 146 个关关节点的二值邻接矩阵进行中心性计算，其结果见图 7 所示。可知网络中，大致呈现论文发表量多的机构在网络中的中心性表现也较为突出的态势。其中，加州大学圣地亚哥分校具有最佳表现，尽管就论文发表量而言该校处于机构中的第 2 位（49 篇，哈佛大学为 64 篇），但其在网络中的交往能力、控制资源能力比哈佛大学更高，并且对网络中其他成员的依赖度比哈佛大学更低，因此该校在网络中处于最核心的地位（特征向量 50.878）。

就网络成员的交往能力而言，加州大学圣地亚哥分校、哈佛大学、麻省理工学院、多伦多大学、

络如图 5（见第 33 页）所示。除 32 号台湾“国立”清华大学、85 号弗吉尼亚理工大学和 121 号印度理工学院三家机构未其他机构形成任何合著论文以外，其余 146 家机构之间均在不同程度上拥有合著关系，连线数为 864 条。对 146 个节点组成网络的二值邻接矩阵进行计算，得到网络整体密度值为 0.040 8，各节点之间的平均捷径距离为 3.218，建立在距离基础上的凝聚力指数为 0.357，分裂指数为 0.643。网络中大多数国家之间的捷径距离 ≤ 3 （成员间捷径距离是 3 的情况出现 7 968 次，捷径距离是 2 的情况出现 4 368 次，捷径距离是 1 的情况出现 864 次），因此，两个机构之间平均通过两个机构才可以建立联系（见图 6），网络整体上不具有很强的连通性。

加州大学伯克利分校、北卡罗来纳大学、华盛顿大学、威斯康辛大学以及纽约大学比较活跃；在不受其他成员的控制与影响方面，加州大学圣地亚哥分校、多伦多大学、麻省理工学院、加州大学伯克利分校、哈佛大学、斯坦福大学、魏茨曼科技大学、北卡罗来纳大学表现比较突出；而在控制网络中其他成员交往的能力方面，加州大学圣地亚哥分校、多伦多大学、麻省理工学院、加州大学伯克利分校、哈佛大学、巴斯德研究所、帝国理工学院能力较强；特征向量计算结果显示，在整体意义上加州大学圣地亚哥分校、麻省理工学院、哈佛大学、多伦多大学、加州大学伯克利分校、北卡罗来纳大学、华盛顿大学、纽约大学、魏茨曼科技大学、波



机构代码:

1.哈佛大学; 2.加州大学圣地亚哥分校; 3.加州大学伯克利分校; 4.多伦多大学; 5.麻省理工学院; 6.华盛顿大学; 7.斯坦福大学; 8.加州大学旧金山分校; 9.加州理工学院; 10.杜克大学; 11.明尼苏达大学; 12.法国国家科学研究中心; 13.中国科学院; 14.东京大学; 15.伊利诺伊大学; 16.巴斯德研究所; 17.威斯康辛大学; 18.特拉维夫大学; 19.伦敦大学帝国理工学院; 20.波士顿大学; 21.康奈尔大学; 22.加州大学洛杉矶分校; 23.大阪大学; 24.布里斯托尔大学; 25.魏茨曼科技大学; 26.西班牙高等科学研究委员会; 27.瑞士联邦理工学院; 28.约翰·霍普金斯大学; 29.加拿大研究理事会; 30.密歇根大学; 31.宾夕法尼亚大学; 32.台湾国立清华大学; 33.爱丁堡大学; 34.圣保罗大学; 35.哥伦比亚大学; 36.曼彻斯特大学; 37.北卡罗来纳大学; 38.美国国家癌症研究所; 39.那不勒斯费德里克二世大学; 40.洛克菲勒大学; 41.国立首尔大学; 42.日本科学技术振兴; 43.加州大学戴维斯分校; 44.马萨诸塞大学; 45.牛津大学; 46.耶鲁大学; 47.普林斯顿大学; 48.俄罗斯科学院; 49.加州大学圣巴巴拉分校; 50.芝加哥大学; 51.哥本哈根大学; 52.韩国科学技术发展研究院; 53.京都大学; 54.麦克吉尔大学; 55.奈良科学技术研究所; 56.英属哥伦比亚大学; 57.剑桥大学; 58.庆应义塾大学; 59.弗莱堡大学; 60.巴黎第六大学; 61.巴黎第十一大学; 62.马萨诸塞州医院; 63.田纳西大学; 64.苏黎世大学; 65.斯塔菲研究所; 66.慕尼黑工业大学; 67.以色列理工大学; 68.伦敦大学学院; 69.卡罗林斯卡学院; 70.埃默里大学; 71.巴伦西亚大学; 72.亚利桑那大学; 73.英国医学研究理事会; 74.新加坡国立大学; 75.纽约大学; 76.艾伯特·爱因斯坦医学院; 77.纽约州立大学石溪分校; 78.德克萨斯农工大学; 79.文特尔研究所; 80.美国应用分子进化基金会; 81.诺丁汉大学; 82.匹兹堡大学; 83.巴伦西亚理工大学; 84.谢菲尔德大学; 85.弗吉尼亚理工大学; 86.名古屋大学; 87.霍华德·休斯医学研究所; 88.宾夕法尼亚州立大学; 89.天津大学; 90.巴塞尔大学; 91.佛罗里达大学; 92.格拉斯哥大学; 93.耶那大学; 94.康斯坦茨大学; 95.马里兰大学; 96.米兰大学; 97.罗马大学; 98.默克集团; 99.德克萨斯大学; 100.弗吉尼亚大学; 101.冈山大学; 102.布朗大学; 103.中山大学; 104.贝勒医学院; 105.卡尔加里大学; 106.巴黎高等师范学院; 107.哥德堡大学; 108.墨尔本大学; 109.莱顿大; 110.墨西哥国立自治大学; 111.内布拉斯加大学; 112.马普学会生物化学研究所; 113.德州大学奥斯丁分校; 114.筑波大学; 115.韦恩州立大学; 116.佐治亚理工学院; 117.纽卡斯尔大学; 118.希伯莱大学; 119.赫尔姆霍兹传染病研究中心; 120.布莱根妇女医院; 121.印度理学院; 122.日本理化学研究所; 123.上海交通大学; 124.瑞士生物信息学研究所; 125.凯斯西储大学; 126.阿尔伯塔大学; 127.亚利桑那大学; 128.伯明翰大学; 129.邓迪大学; 130.格罗宁根大学; 131.密苏里大学; 132.洛斯阿拉莫斯国家实验室; 133.南加利福尼亚大学; 134.加拿大默克·福斯特公司; 135.德克萨斯大学达拉斯分校; 136.密歇根州立大学; 137.犹他大学; 138.华威大学; 139.乌尔茨堡大学; 140.鹿培法布拉大学; 141.纽约西奈山医学院; 142.范德比尔特大学; 143.弗吉尼亚州立邦联大学; 144.南洋理工大学; 145.科罗拉多州立大学; 146.橡树岭国家实验室; 147.俄亥俄大学; 148.法国国家农科院; 149.克利夫兰普斯研究所

图5 合成生物学 SCI 论文数大于 5 篇的机构合著网络

Normalized Centrality Measures

		1	2	3	4
		Degree	Closeness	Betweenness	Eigenvector
1	1	17.931	42.398	8.569	41.522
2	2	21.379	45.171	14.446	50.878
3	3	11.034	42.522	10.321	28.779
4	4	15.862	44.753	12.989	31.345
5	5	17.931	44.207	10.835	42.730
6	6	10.345	39.945	2.685	27.792
7	7	7.586	40.616	1.455	24.905
8	8	7.586	36.802	1.834	19.219
9	9	7.586	37.958	4.102	14.894
10	10	6.897	36.802	4.297	9.485
11	11	6.207	38.770	2.281	17.092
12	12	7.586	34.118	3.996	3.016
13	13	7.586	39.189	4.673	10.585
14	14	3.448	26.460	0.401	0.950
15	15	4.138	34.038	1.653	5.755
16	16	8.276	38.667	6.858	6.276
17	17	10.345	39.402	5.321	21.533
18	18	7.586	37.760	3.673	16.284
19	19	8.276	35.024	6.039	2.638
20	20	6.897	38.978	0.521	25.380
21	21	8.276	39.402	3.376	18.133
22	22	2.069	32.222	0.724	4.061
23	23	4.138	28.101	0.457	2.590
24	24	2.759	31.590	0.313	2.070
25	25	9.655	40.390	2.253	27.238
26	26	3.448	31.317	1.060	1.364
27	27	2.759	29.897	0.797	3.521
28	28	4.138	32.731	1.486	2.212
29	29	2.759	28.827	1.768	0.834
30	30	2.759	33.257	1.322	5.741
31	31	6.207	35.714	2.254	13.766
32	32	5.517	36.616	4.972	6.452
33	34	3.448	32.294	0.932	4.255
34	35	7.586	35.109	1.911	7.565
35	36	2.069	29.058	1.239	0.792
36	37	11.034	40.278	4.594	29.437
37	38	8.276	37.760	4.715	10.745
38	39	2.759	31.116	0.965	1.236
39	40	4.828	32.880	0.622	6.215
40	41	2.759	31.183	0.275	3.192

图7 合成生物学机构合著网络中146个具有合著关系成员的中心性计算结果(部分)

士顿大学处于网络核心地位。

计算结果也显示,某些出色的专业研究机构在合著网络中并没有很好的表现。例如,79号美国文特尔研究所(J. Craig Venter Institute,简称JCVI),近年来,该研究所开展的多项研究工作在业内虽具有很高的影响力,但由于其合著论文较少,因此,在机构合著网络中并未处于核心地位。此外,中国大陆地区有4家论文产出在5篇以上的机构(即13号中国科学院、89号天津大学、103号中山大学以及123号上海交通大学)在网络中的表现一般。相比较而言,中国科学院在网络中处于中等地位,其交往能力和控制其他成员的能力较弱,但接近中心度比较高,原因在于其与网络中哈佛大

学、多伦多大学以及巴斯德研究所等一些核心成员都有关联,因此,其在网络中的“权力”得到一定提升。计算结果也显示,网络中24个研究机构的中间中心度为0,说明部分研究机构不具备掌控网络资源的能力。

2.2.3 凝聚子群分析

146家机构之间合成生物学合著论文达到1篇以上的不多,说明合著行为并不算很紧密。因此,对机构的凝聚子群进行分析,不考虑合著中的多值关系,而直接采用派系进行分析。

经过多次试验,选定派系最小规模为4,计算结果显示,机构合著网络中一共发现41个最小规模为4的派系(见图8)。

CLIQUEs

Minimum Set Size: 4
 Input dataset: 146 家机构原始数据 (二值化)
 41 cliques found.

1: 1 2 5 25 37 120	15: 2 4 5 20 37	29: 18 21 25 67 118
2: 1 2 5 20 37 120	16: 2 4 11 20 37	30: 23 42 55 58
3: 1 2 5 87 120	17: 2 6 8 43 75	31: 6 8 31 108
4: 1 2 3 5 7	18: 2 3 6 8	32: 4 5 38 44
5: 1 2 5 75 87	19: 2 8 25 43	33: 38 111 127 139
6: 1 2 3 5 20	20: 2 6 17 43 75 149	34: 17 55 56 86
7: 1 2 3 6 7	21: 2 17 55 58	35: 61 64 65 124
8: 1 2 6 75 87	22: 2 18 75 76	36: 1 5 25 37 62 120
9: 1 2 20 47	23: 2 5 49 52 120	37: 1 5 62 87 120
10: 1 2 75 76	24: 2 63 143 146	38: 1 5 75 90
11: 1 2 76 141	25: 2 6 75 87 149	39: 4 11 37 107
12: 2 3 4 5 7	26: 1 13 77 105	40: 1 3 5 7 125
13: 2 3 4 5 20	27: 14 23 42 122	41: 1 5 37 125
14: 2 3 4 11 20	28: 3 4 7 21	

Clique-by-Clique Actor Co-membership matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41		
1	6	5	4	3	3	3	2	2	2	2	2	2	2	1	3	2	1	1	2	1	1	1	3	1	1	1	0	0	1	0	0	1	0	0	0	5	3	2	1	2	3		
2	5	6	4	3	3	4	2	2	3	2	2	2	3	2	4	3	1	1	1	1	1	1	3	1	1	1	0	0	0	0	0	1	0	0	0	4	3	2	1	2	3		
3	4	4	5	3	4	3	2	2	2	2	2	2	1	2	1	1	1	1	1	1	1	1	3	1	2	1	0	0	0	0	0	1	0	0	0	3	4	2	0	2	2		
4	3	3	3	5	3	4	4	2	2	2	4	3	2	2	1	2	1	1	1	1	1	1	2	1	1	1	0	2	0	0	0	1	0	0	0	2	2	2	0	4	2		
5	3	3	4	3	5	3	2	4	2	3	2	2	2	1	2	1	2	1	1	2	1	2	2	1	3	1	0	0	0	0	0	1	0	0	0	2	3	3	0	2	2		
6	3	4	3	4	3	5	3	2	3	2	2	3	4	3	3	2	1	2	1	1	1	1	2	1	1	1	0	1	0	0	0	1	0	0	0	2	2	2	0	3	2		
7	2	2	2	4	2	3	5	3	2	2	2	3	2	2	1	1	2	3	1	2	1	1	1	1	1	2	1	0	2	0	0	1	0	0	0	0	1	1	1	0	3	1	
8	2	2	3	2	4	2	3	5	2	3	2	1	1	1	1	3	2	1	3	1	2	1	1	1	4	1	0	0	0	0	1	0	0	0	0	1	2	2	0	1	1		
9	2	3	2	2	2	3	2	2	4	2	2	1	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	
10	2	2	2	2	3	2	2	3	2	4	3	1	1	1	1	1	2	1	1	2	1	3	1	1	2	1	0	0	0	0	0	0	0	0	0	0	1	1	2	0	1	1	
11	2	2	2	2	2	2	2	2	2	3	4	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	
12	2	2	2	4	2	3	3	1	1	1	5	4	3	3	2	1	2	1	1	1	1	2	1	1	2	1	1	0	0	3	0	0	0	2	0	0	0	1	1	1	1	3	1
13	2	3	2	3	2	4	2	1	2	1	4	5	4	4	3	1	2	1	1	1	1	1	2	1	1	1	0	0	2	0	0	0	2	0	0	0	1	1	1	1	2	1	
14	1	2	1	2	1	3	2	1	2	1	1	3	4	5	3	4	1	2	1	1	1	1	1	1	1	1	0	0	2	0	0	0	1	0	0	0	0	0	0	2	1	0	
15	3	4	2	2	2	3	1	1	2	1	1	3	4	3	5	4	1	1	1	1	1	1	2	1	1	0	0	1	0	0	0	2	0	0	0	2	1	1	2	1	2		
16	2	3	1	1	1	2	1	1	2	1	1	2	3	4	4	5	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	1	0	0	0	1	0	0	3	0	1	
17	1	1	1	1	2	1	2	3	1	2	1	1	1	1	1	5	3	3	4	1	2	1	1	3	0	0	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0		
18	1	1	1	2	1	2	3	2	1	1	1	2	2	1	1	3	4	2	2	1	1	1	1	1	2	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	1	0	
19	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	2	4	2	1	1	1	1	1	1	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0		
20	1	1	1	1	2	1	2	3	1	2	1	1	1	1	1	4	2	2	6	2	2	1	1	4	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	
21	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	4	1	1	1	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	
22	1	1	1	1	2	1	2	1	3	2	1	1	1	1	1	2	1	1	1	2	1	1	4	1	1	2	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
23	3	3	3	2	2	2	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	1	0	0	0	2	2	1	0	1	1	
24	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
25	1	1	2	1	3	1	2	4	1	2	1	1	1	1	1	3	2	1	4	1	2	1	1	5	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	
26	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1		
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
28	0	0	0	2	0	1	2	0	0	0	0	3	2	2	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	
29	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
31	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
32	1	1	1	1	1	0	0	0	0	0	2	2	1	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
36	5	4	3	2	2	2	1	1	1	1	1	1	0	2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
37	3	3	4	2	3	2	1	2	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
38	2	2	2	2	3	2	1	2	1	2	1	1	1	0	1	0	0	1	0	0	1	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
39	1	1	0	0	0	0	0	0	0	0	0	1	1	2	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
40	2	2	2	4	2	3	3	1	1	1	1	3	2	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
41	3	3	2	2	2	2	1	1	1	1	1	1	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

图8 合成生物学机构合著网络的派系分析结果 (最小规模为4)

41 个派系中规模为 4 的有 19 个，规模为 5 的有 19 个，规模为 6 的有 3 个，包含的成员（机构）有 53 个，其余 93 个成员（机构）不属于任何最小规模为 4 的派系。分析 41 个派系两两共同拥有成员的情况，可以看到两个派系共同拥有 3 个及以上成员的情况仅出现了 78 次，而未共同拥有成员的情况有 352 次，这说明整体而言 41 个派系具有相对较强的独立性，它们之间的重叠性并不是很大。

对 41 个派系的成员构成进行分析可以发现一些机构的合著圈较多。例如，加州大学圣地亚哥分校出现在 25 个派系中，哈佛大学出现在 17 个派系中，麻省理工学院出现在 16 个派系中，加州大学伯克利分校出现在 9 个派系中，多伦多大学出现在 8 个派系中。此外，加州大学圣地亚哥分校和哈佛大学、哈佛大学和麻省理工学院均同时出现在了 11 个派系中，加州大学圣地亚哥分校和麻省理工学院同时出现在 10 个派系中，加州大学圣地亚哥分校和加州大学伯克利分校同时出现在 7 个派系中，加州大学圣地亚哥分校和华盛顿大学同时出现在 6 个派系中。中国的 4 家研究机构中，只有中国科学院进入一个规模为 4 的派系，其成员还包括美国的哈佛大学和纽约州立大学石溪分校，以及加拿大卡尔加里大学。

3 结论与探讨

利用合成生物学领域发表于 2006—2011 年的 SCI 论文，在国家层次和机构层次上构建了相应的合著关系网络，分别对其网络构成、网络中心性以及凝聚子群进行了分析，并形成如下结论：

（1）由于合成生物学属于近年来才发展起来的新兴研究领域，因此，与其他一些成熟学科相比，该领域过去 6 年中发表的 SCI 论文尽管增长率较高，但数量尚不多，目前年度规模未超过 500 篇。在此背景下，参与论文撰写的国家和科研机构数量相对而言比较有限，无论是国家合著网络还是机构合著网络的网络规模都不是很大。数据也显示，参与论文撰写的 60 个国家中的 57 个、1 538 家机构中的 1 227 家拥有合著关系，所以，尽管以论文形式反映的合成生物学研究规模不算很大，但科研合作关系在该领域还是比较广泛的。

（2）分析显示合成生物学国家合著网络整体连

通性较好，网络具有中等程度凝聚力，成员之间互动性较为紧密。可以判断，就合成生物学研究中整体而言，国家之间信息的交流与传播比较通畅。在成员的具体表现上，美国、德国的各种中心性测度均具有较好一致性，二者是国家合著网络中的最核心成员。尤其是美国，其在网络中的活跃性、控制其他成员能力和自主独立性远远超出了其他网络成员。其他成员则各有千秋，一些国家在合作活跃性方面展示出优势，另一些国家则在控制网络资源方面具有较强能力。凝聚子群分析表明，合成生物学国家合著网络中存在着一些关系紧密的小团体，而这些小团体具有较强重叠性。这是由于美国、德国、英国等科技强国较多同时出现在不同派系中而造成的。由此说明，部分国家之间在合成生物学研究中有非常良好的互动性。而大多数国家一如其在传统科研领域的表现一样，无论是在合作的活跃程度方面，还是在网络中的引领和促进作用方面，都没有在合成生物学这一新兴领域展示出较强的能力。

（3）合成生物学机构合著网络整体而言连通性不高，146 家 SCI 论文发表数量大于 5 篇的机构中，两个机构之间直接或通过一个机构建立起来的关联关系仅约占所有关联关系的 25%。该网络中的合著关系整体属于比较疏离的状态，机构之间的合作行为较为松散，由此判断信息和资源在机构网络中的流动性一般，合作网络对网络成员行为的影响可能并不大。分析同时也显示，机构网络中同样存在一些合作关系较为紧密的小团体，但小团体之间未共同拥有成员的情况占到 43%。由此可以判断，这些小团体内部尽管信息交流较频繁，但相互之间较强的独立性有可能阻碍信息或知识在更大范围内的有效共享。此外，机构在网络中是否处于核心地位与其论文发表能力具有一定关系，其中美国加州大学圣地亚哥分校表现突出。美国的科研机构在合著网络中具有较好的整体优势，其相当一部分科研机构的交往合作能力很强，同时掌控网络资源的能力也较强，因此，对网络的流动性和信息传播至关重要，是网络中重要成员。

（4）分析显示，尽管就论文发表量而言我国处于国家排名中的第 4 位（154 篇），但通过合著网络所展示的研究合作能力并不强。尤其是对网络资源的控制能力方面，我国还处于较弱的状况。此

外, 研究机构目前在合成生物学研究中的合作圈较小, 合作对象也较为有限。因此, 中国的研究机构在研究合作的拓展方面未来还需要较多努力, 向美国、德国等强势成员国或者优势研究机构靠拢, 通过争取合作机会以发展自己在研究网络中的能力, 同时也获取比较有价值的资源。

本研究基于论文合著关系对合成生物学研究当前的研究合作现状进行了一定揭示。应当看到, 研究所依据的合著关系是以国家/机构是否参与某篇论文的撰写来直接定义的, 而并未考虑作者所属国家或机构在论文中的署名排序。如果考虑署名排序, 一篇论文的不同作者所属的国家或机构应当具有不同权重, 由此国家/机构之间的关系以及在合著网络中发挥的作用也应当不同, 下一步研究可以以此为思路来继续推进。此外, 由于科研论文大数据量前提下作者重名的辨识问题目前尚无一以贯之的解决手段, 因此, 本研究并未对著者个人合作网络进行分析, 未来此方面工作还需要探索。■

参考文献:

- [1] Synthetic Biology community. Synthetic Biology [EB/OL]. [2012-02-17]. <http://syntheticbiology.org>.
- [2] 赵学明, 陈涛. 合成生物学: 进展与展望 [EB/OL]. [2011-07-01]. http://synbio.tju.edu.cn/res_pdf/shell_res_doc_3_1_6_1.pdf.
- [3] 李亮, 朱庆华. 社会网络分析方法在合著分析中的实证研究[J]. 情报科学, 2008, 26(4): 549-555.
- [4] 邱均平, 瞿辉. 我国科研机构合作网络知识扩散研究——以“生物多样性”研究为例[J]. 图书情报知识, 2011(6): 5-11.
- [5] Liu X, Bollen J, Nelson M L, et al. Co-authorship Networks in the Digital Library Research Community[J]. Information Processing & Management, 2005, 41(6): 1462-1480.
- [6] 刘军. 社会网络分析导论[M]. 北京: 社会科学文献出版社. 2004:
- [7] 刘军. 整体网分析讲义—UCINET 软件实用指南[M]. 上海: 上海人民出版社, 2009: 108.
- [8] 罗家德. 社会网分析讲义[M]. 北京: 社会科学文献出版社, 2010.
- [9] 约翰·斯科特. 社会网络分析法[M]. 重庆: 重庆大学出版社, 2011.
- [10] Borgatti S P, Everett M G, Freeman L C. Ucinet 5 for Windows: Software for Social Network Analysis [EB/OL]. Harvard, MA: Analytic Technologies (1999) [2011-09-15]. <http://wenku.baidu.com/view/75a10eeb172ded630b1cb662.html>.
- [11] 赵学明, 王庆昭. 合成生物学: 学科基础、研究进展与前景展望[J]. 前沿科学, 2007(3): 56-66.
- [12] 宋凯, 黄熙泰. 合成生物学导论[M]. 北京: 科学出版社, 2010.
- [13] 张春霆. 人造生命与合成生物学[N]. 科技导报, 2011, 29(01).
- [14] 张春霆. 合成生物学研究的进展[J]. 中国科学基金, 2009(2): 65-69.

A Social Network Analysis of Research Collaboration in Synthetic Biology

LIU Ya

(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: Based on the theory of social network analysis, this paper analyzes the pattern of research collaboration in global synthetic biology using co-authorship data from 2 119 articles published in 2006-2011 and collected by the SCI-EXPANDED database. The empirical study shows that most countries/organizations with a large number of publications are playing a key role in co-authorship networks, while a few countries'/organizations' status in the networks doesn't match to their performance in publication. In addition, the study reveals a few groups with close collaboration relationship both in country network and organization network.

Key words: synthetic biology; SCI articles; social network analysis; co-authorship network