

《社会网络分析》期末作业

人工智能主题领域文献的 合著网络、关键词网络及学科共现网络分析

周妍 徐山川 沈奕娜 宋小双

2014/1/3

摘要

本文利用社会网络分析技术对近五年的人工智能主题领域相关文献进行分析。在对文献进行基本描述统计的基础上，对人工智能主题领域文献的作者合著网络、机构合著网络、国家合著网络、学科共现网络以及关键词共现网络的度分布、平均路径长度、网络密度、聚类系数、介数分布、连通集团规模等指标进行分析，从而对当前人工智能领域的合作情况与研究现状进行梳理。

关键词： 社会网络分析 人工智能 文献计量 合著网络 共词网络 学科共现网络

目录

1	引言.....	1
2	数据和方法.....	1
2.1	数据准备.....	1
2.2	数据预处理.....	2
2.3	分析方法及工具.....	2
3	文献基本情况.....	2
3.1	中文文献基本情况.....	2
3.1.1	发文量.....	3
3.1.2	合作率.....	3
3.1.3	合作水平.....	3
3.2	英文文献基本情况.....	4
3.2.1	发文量.....	4
3.2.2	主要发文国家统计.....	5
3.2.3	主要发文机构统计.....	5
3.2.4	合作率.....	6
3.2.5	合作水平.....	6
3.2.6	学科分布统计.....	7
4	合著分析、学科分析与关键词分析.....	8
4.1	作者合著网络分析.....	8
4.1.1	国内文献作者合著网络.....	8
4.1.2	国外文献作者合著网络.....	10
4.1.3	国内外文献作者合著对比.....	13
4.2	机构合著网络分析.....	14
4.2.1	国内文献机构合著网络.....	14
4.2.2	国外文献机构合著网络.....	17
4.2.3	国内外文献机构合著对比.....	22
4.3	国家合著网络分析.....	22
4.3.1	度的分布.....	23
4.3.2	网络特征参数.....	25
4.3.3	介数及其分布.....	25
4.3.4	连通集团规模及其分布.....	26
4.3.5	核心—边缘结构分析.....	27
4.4	关键词共现网络分析.....	27
4.4.1	国内文献关键词共现网络.....	28
4.4.2	国外文献关键词共现网络.....	29
4.4.3	国内外文献关键词共现网络对比.....	30
4.5	学科共现网络分析.....	30
4.5.1	整体网络分析.....	30
4.5.2	学科中心性分析.....	31
4.5.3	学科聚类分析.....	32
5	总结与讨论.....	34

6	参考文献.....	34
7	小组分工.....	35

人工智能主题领域文献的 合著网络、关键词网络及学科共现网络分析

1301212641 周 妍

1101214041 徐山川

1301212634 沈奕娜

1301212635 宋小双

1 引言

被称为世界三大尖端科技之一的人工智能（Artificial Intelligence），是自 1956 年 Dartmouth 学会提出后发展起来的新型学科，具有涉及学科广、技术高端、使用范围广等特点。迄今为止，人工智能领域已经经历了近 60 年的发展，二十世纪七八十年代尤其是其发展的黄金年代。如今人工智能领域已经取得了众多研究成果，被广泛应用于企业管理、医学、矿业、技术研究、电子技术等多种领域^{1,2}。二十一世纪是信息化时代，作为现代信息技术的精髓，人工智能技术将成为新世纪科学技术的前沿和焦点。

目前，人工智能技术在发达国家发展很快。尤其是在美国，发展更为迅速。在人工智能技术领域十分活跃的 IBM 公司，在智能电脑方面有了新的突破，成功地生产了具有人脑千分之一智力的电脑，而且正在开发功能更为强大的超级电脑。据内部消息透露，预计该超级电脑研制成后，其智力水平将大致与人脑相当。除了 IBM 公司外，其他公司也在加紧这方面的研究，估计在未来几年内其成果将更为惊人。

在我国，很长一段时间内，专家们都把研制具有人的行为特征的类人性机器人作为奋斗目标。机器人的发展水平不仅与计算机科技水平相关，而且与一个国家工业的各方面发展水平密切相关。中国科技大学在国家基金的支持下，经过十年攻关和钻研，于 2000 年成功研制出我国第一台类人性机器人³。

可见，随着计算机技术的不断进步和发展，人工智能已经越来越受到社会各界的高度重视，该领域具有良好、广阔的发展前景⁴。通过研究人工智能主题领域的合著网络、学科网络以及关键词网络，可以掌握人工智能领域的作者、机构、国家之间的合作情况，该领域所涉及的学科、研究方向及其相互关系，从而对该领域的发展现状进行更好的梳理。

2 数据和方法

2.1 数据准备

本研究所采用的文献题录数据分为中英文两部分，检索时间点为 2013 年 12 月 20 日。

中文采用中国学术期刊网络出版总库中收录的主题至少包含“人工智能”，“智能代理”，“智能机器”，“机器学习”，“自然语言处理”，“模式识别”，“机器人学”，“专家系统”八个关键词之一，年份为 2009 年至今的文献 15297 篇。

英文采用 Web of Science 收录的主题至少包含“Artificial Intelligence”，“Intelligent Agent”，

¹ 韦雄奕，解晨. 人工智能发展问题探究与思考[J]. 计算机光盘软件与应用, 2013, 8: 091.

² 孙晔，吴飞扬. 人工智能的研究现状及发展趋势[J]. 价值工程, 2013, 32(28): 5-7.

³ 邢王秀. 人工智能发展现状与未来发展[J]. 中国-东盟博览, 2013,02:62.

⁴ 万雷. 论人工智能的现状与发展方向[J]. 科技风, 2012 (15): 54-54.

“Intelligent Machine”, “Machine Learn”, “Natural Language Processing”, “Pattern Recognition”, “Robotics”, “Expert System” 八个关键词之一(中文关键词对应的英文关键词), 年份为 2009 年至今的文献 29799 篇。

2.2 数据预处理

中文数据只选取分析所需要的“标题”, “关键词”, “作者”, “机构”, “年份”五个字段的内容并剔除包含空字段或者征文简章的记录, 最终保留 14646 篇。然后对机构字段尽量概化到大学级别。例如在机构中会存在很多学院级别的单位如“中国科学院软件所”, “中国科学院计算所”, 我们把这样的机构都概化为“中国科学院”。在作者重名方面, 由于 CNKI 所提供的题录中大部分文献没有建立作者与机构的一一对应关系, 因此无法依据作者所在的机构处理作者重名问题。

英文数据的处理和中文的数据类似, 选取“标题”, “关键词”, “作者”, “作者地址”, “年份”, “学科”六个字段的内容, 并从“作者地址”字段中提取出大学级别的机构名称以及所在的国家并剔除包含空字段的记录, 最终保留 23002 篇。由于英文文献提供了作者与文献的一一对应关系, 因此我们利用作者与机构的组合处理了作者重名问题。

2.3 分析方法及工具

在数据抓取方面, 中文数据采用 python 编写简单的程序对数据进行抓取, 英文数据采用 Web of Science 所提供的导出功能获取数据。之后采用 python 以及 excel 对数据进行清理和格式整理并生成社会网络分析工具所需要的.net 文件。接着利用 excel 对中英文文献进行基本的描述统计分析。然后在合著网络、学科网络与关键词网络分析部分采用 Pajek 和 Gephi 分别对作者、机构、国家、学科、关键词进行分析与整理并绘制网络图。

3 文献基本情况

本研究共获取有效中文文献题录 14646 条, 有效英文文献题录 23002 条。下面对文献的基本情况进行介绍, 包括中文文献的发文量、合作率与合作水平, 以及英文文献的发文量、主要发文国家、主要发文机构、合作率、合作水平以及学科分布。其中, 合作率与合作水平是度量合作程度的两个指标。合作率是指作者数大于等于 2 的论文占全部论文数的比例, 合作水平则用所有论文的平均作者合作度表示。一篇论文的作者数即为该论文的作者合作度。

3.1 中文文献基本情况

根据本研究的 14646 条有效中文文献题录, 对 2009—2013 年人工智能主题领域中文文献的发文量、合作率以及合作水平情况进行统计。统计结果如下表所示:

表 1 2009—2013 年人工智能主题领域中文文献发文量、合作率、合作水平

	2009	2010	2011	2012	2013	近五年
文献数量	3281	3204	3008	2973	2180	14646
合作率	76.23%	74.25%	73.27%	71.38%	67.75%	72.94%
合作水平	2.597	2.631	2.651	2.626	2.582	2.619

3.1.1 发文量

2009—2013 年, 人工智能主题领域中文文献的年度发文量呈逐年递减趋势, 自 2009 年的 3281 篇下降至 2013 年的 2180 篇 (尽管由于检索时间点的原因, 2013 年最后 1 个多月的题录未收入本研究中, 但是并不影响发文量年度递减的趋势)。

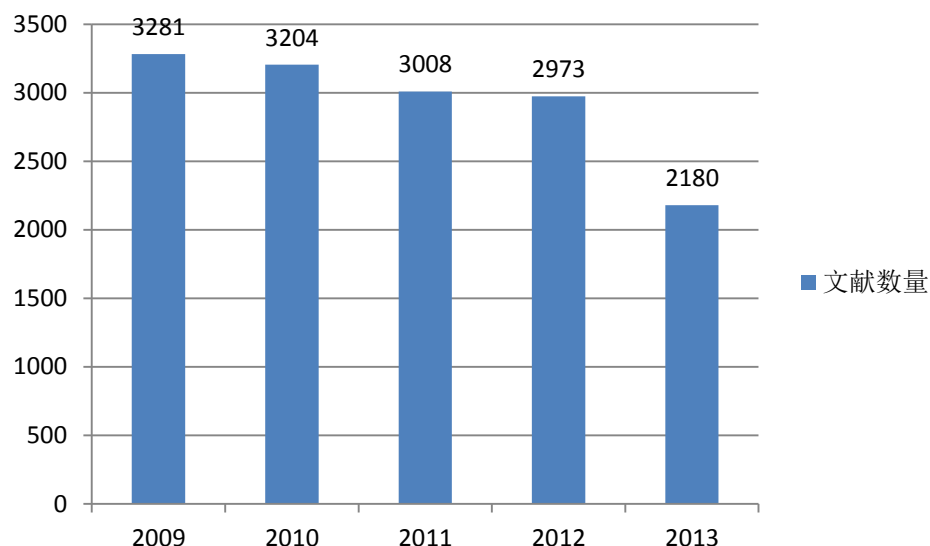


图 1 2009—2013 年人工智能主题领域中文文献年度发文量

3.1.2 合作率

2009—2013 年, 人工智能主题领域中文文献的年度合作率呈逐年递减趋势, 自 2009 年的 76.23% 下降至 2013 年的 67.75%。近五年的平均合作率为 72.94%。

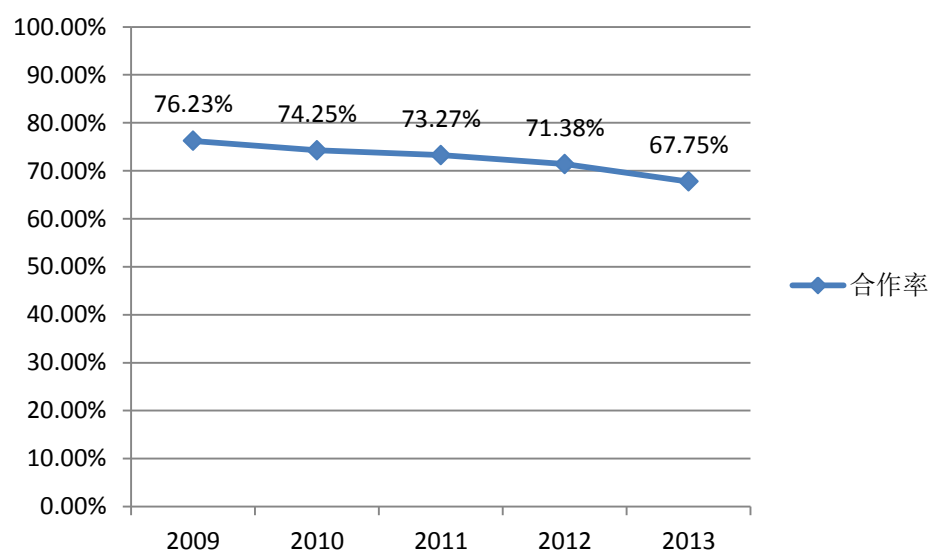


图 2 2009—2013 年人工智能主题领域中文文献年度合作率

3.1.3 合作水平

2009—2013 年, 人工智能主题领域中文文献的年度合作水平呈先上升后下降的趋势, 自 2009 年的 2.597 上升至 2011 年的 2.651, 此后又下降至 2013 年的 2.582。近五年的平均合作水平为 2.619, 即平均每篇文献由 2.62 个作者完成。

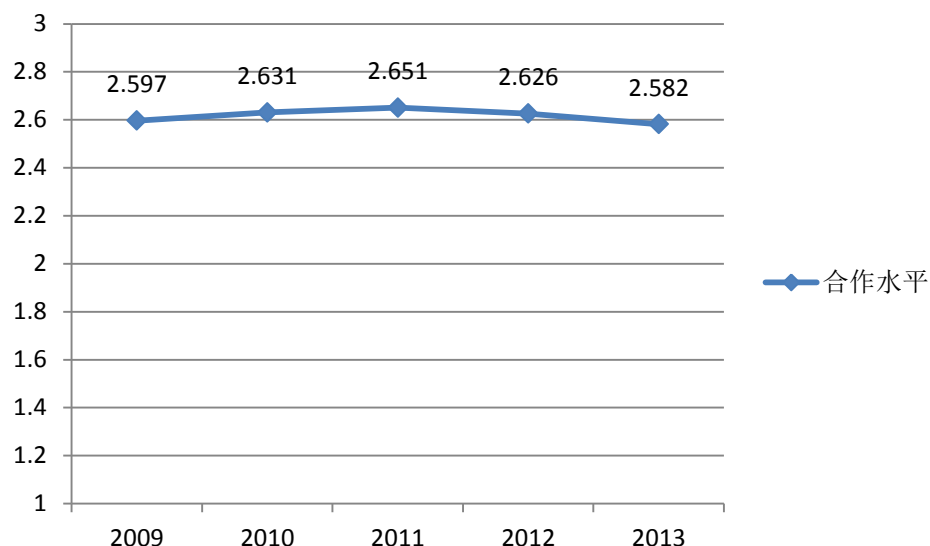


图 3 2009—2013 年人工智能主题领域中文文献年度合作水平

3.2 英文文献基本情况

根据本研究的 23002 条有效英文文献题录,对 2009—2013 年人工智能主题领域英文文献的发文量、合作率以及合作水平情况进行统计。统计结果如下表所示:

表 2 2009—2013 年人工智能主题领域中文文献发文量、合作率、合作水平

	2009	2010	2011	2012	2013	近五年
文献数量	4000	4081	4479	5391	5051	23002
合作率	88.28%	89.07%	91.18%	91.01%	92.73%	90.60%
合作水平	3.593	3.424	3.503	3.661	3.714	3.588

3.2.1 发文量

2009—2013 年,人工智能主题领域英文文献的年度发文量总体来看呈上升趋势,自 2009 年的 4000 篇上升至 2012 年的 5391 篇。由于检索时间点的原因,2013 年最后 1 个多月的题录未收入本研究中,因此 2013 年的发文量有所下降。

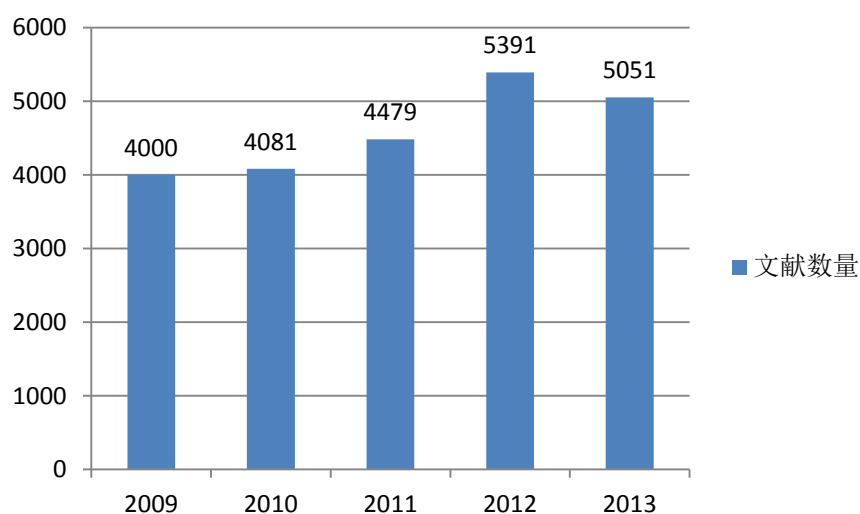


图 4 2009—2013 年人工智能主题领域英文文献年度发文量

3.2.2 主要发文国家统计

2009—2013 年，人工智能主题领域英文文献发文量较多的国家依次为美国、中国、英国、西班牙和德国。中美两国的发文量远高于其他国家和地区，该领域的研究成果来源主要集中在欧洲和北美，而亚洲，如韩国、印度、伊朗、土耳其、新加坡，也开始兴起有关人工智能领域的研究。

表 3 2009-2013 年人工智能主题领域英文文献主要发文国家统计

国家/地区	记录
USA	7661
PEOPLES R CHINA	4367
ENGLAND	2182
SPAIN	2056
GERMANY	1914
CANADA	1556
FRANCE	1445
TAIWAN	1356
ITALY	1347
JAPAN	1340
SOUTH KOREA	1065
AUSTRALIA	989
INDIA	808
IRAN	762
TURKEY	666
NETHERLANDS	613
SWITZERLAND	590
BRAZIL	559
SINGAPORE	517
POLAND	484

3.2.3 主要发文机构统计

2009—2013 年，人工智能主题英文文献发文量较多的机构依次为中国科学院、南洋理工大学、哈佛大学、麻省理工大学、上海交通大学以及卡耐基梅隆大学，尤其是中科院，发文量远高于其他机构，说明中国在人工智能领域的研究取得了一定的成就。在发文量排名前 200 的机构中，有 8 个美国机构、发文量为 1694 篇，5 个中国机构、发文量为 1303 篇，中国从事该领域研究的一流机构的研究成果并没有落后于美国太多。

表 4 2009-2013 年人工智能主题领域英文文献主要发文机构统计

机构名称	机构题录	地区	记录
中国科学院	CHINESE ACAD SCI	中国	483
南洋理工大学	NANYANG TECHNOL UNIV	新加坡	284
哈佛大学	HARVARD UNIV	美国	283
麻省理工大学	MIT	美国	262
上海交通大学	SHANGHAI JIAO TONG UNIV	中国	244
卡耐基梅隆大学	CARNEGIE MELLON UNIV	美国	240
东京大学	UNIV TOKYO	日本	210

多伦多大学	UNIV TORONTO	加拿大	208
香港科技大学	HONG KONG POLYTECH UNIV	中国	202
斯坦福大学	STANFORD UNIV	美国	198
加州大学伯克利分校	UNIV CALIF BERKELEY	美国	196
清华大学	TSINGHUA UNIV	中国	193
浙江大学	ZHEJIANG UNIV	中国	181
法国国家科学研究院	CNRS	法国	174
加州大学圣地亚哥分校	UNIV CALIF SAN DIEGO	美国	174
伊利诺伊大学	UNIV ILLINOIS	美国	174
伦敦大学帝国理工学院	UNIV LONDON IMPERIAL COLL SCI TECHNOL MED	英国	174
国立新加坡大学	NATL UNIV SINGAPORE	新加坡	170
马里兰大学	UNIV MARYLAND	美国	167
伦敦学院大学	UCL	英国	162

3.2.4 合作率

2009—2013 年,人工智能主题领域英文文献的年度合作率整体来看呈上升趋势,自 2009 年的 88.28%上升至 2013 年的 92.73%。近五年的平均合作率为 90.60%。

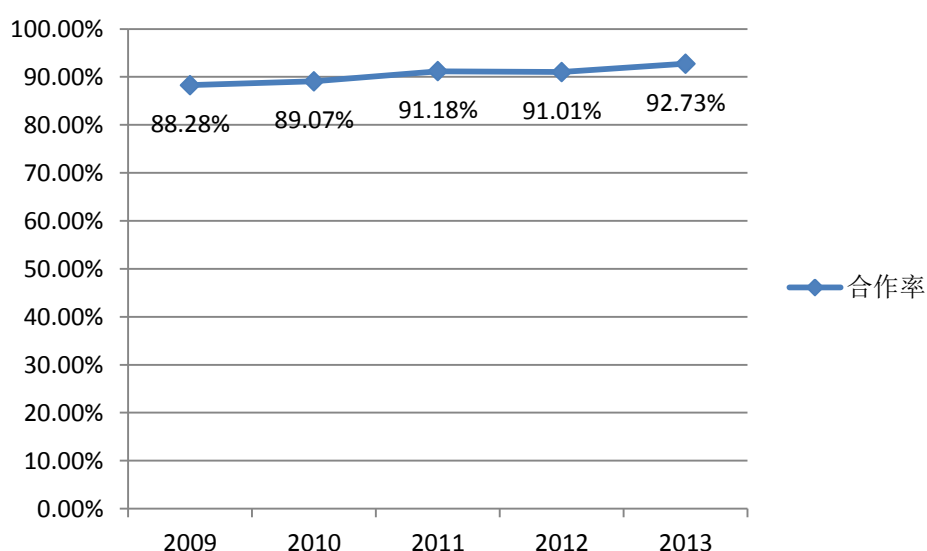


图 5 2009—2013 年人工智能主题领域英文文献年度合作率

3.2.5 合作水平

2009—2013 年,人工智能主题领域英文文献的年度合作水平呈先下降后上升的趋势,自 2009 年的 3.593 下降至 2010 年的 3.424,此后一直上升至 2013 年的 3.714。近五年的平均合作水平为 3.588,即平均每篇文献由 3.6 个作者完成。

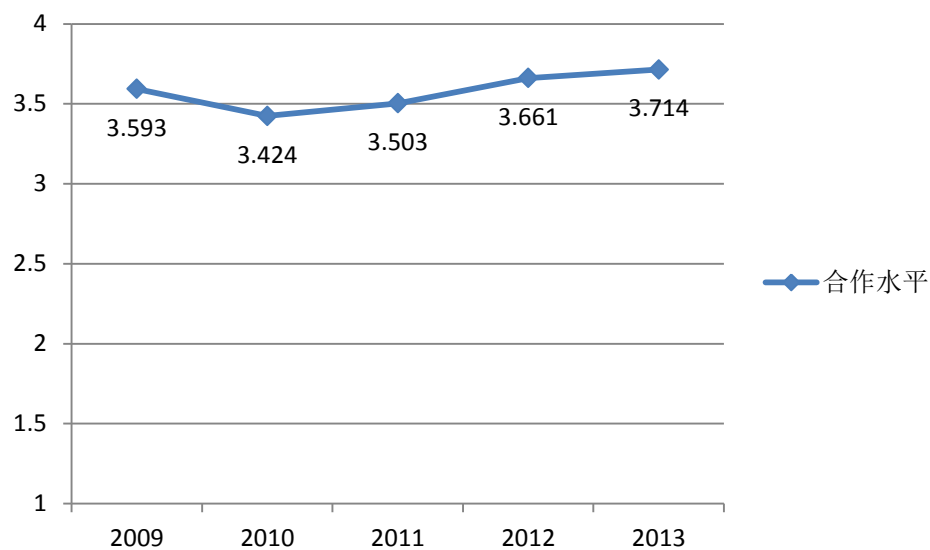


图 6 2009—2013 年人工智能主题领域中文文献年度合作水平

3.2.6 学科分布统计

2009—2013 年，人工智能主题英文文献主要分布的学科领域依次为计算机科学、电工与电子技术、机器人学、自动化系统，大多为工科领域的文献，此外还有一些神经科学、生物化学材料等理科领域的文献。此处的学科分类以 WOS 的分类为准。

表 5 2009-2013 年人工智能主题领域英文文献学科分布统计

Web of Science 类别	记录
COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE	7980
ENGINEERING ELECTRICAL ELECTRONIC	5666
COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS	2425
ROBOTICS	2352
AUTOMATION CONTROL SYSTEMS	2313
OPERATIONS RESEARCH MANAGEMENT SCIENCE	2199
COMPUTER SCIENCE INFORMATION SYSTEMS	2122
NEUROSCIENCES	1722
MULTIDISCIPLINARY SCIENCES	1486
COMPUTER SCIENCE SOFTWARE ENGINEERING	1349
BIOCHEMISTRY MOLECULAR BIOLOGY	1250
BIOCHEMICAL RESEARCH METHODS	1226
COMPUTER SCIENCE THEORY METHODS	1225
INSTRUMENTS INSTRUMENTATION	1119
ENGINEERING MULTIDISCIPLINARY	1116
MATHEMATICAL COMPUTATIONAL BIOLOGY	1008
ENGINEERING BIOMEDICAL	939
ENGINEERING MANUFACTURING	939
BIOTECHNOLOGY APPLIED MICROBIOLOGY	903
CHEMISTRY ANALYTICAL	882

4 合著分析、学科分析与关键词分析

4.1 作者合著网络分析

在本研究中,作者合著网络以作者为节点,节点的大小代表节点的度值。如果两个作者合著过一篇文章,则他们之间就存在一条无向边,边的权重为两个作者合著的次数。下面分别分析近五年来人工智能主题领域国内外文献的作者合著网络,并对国内外作者合著情况进行对比。

4.1.1 国内文献作者合著网络

根据 14646 条有效中文题录中的作者信息建立合著网络,删去网络中的孤立点(即没有与任何其他作者合著过的孤立作者节点),最终得到包含 24768 个节点、67349 条边的整体作者合著网络。下图是度值大于等于 6 的作者节点构成的作者合著网络图:

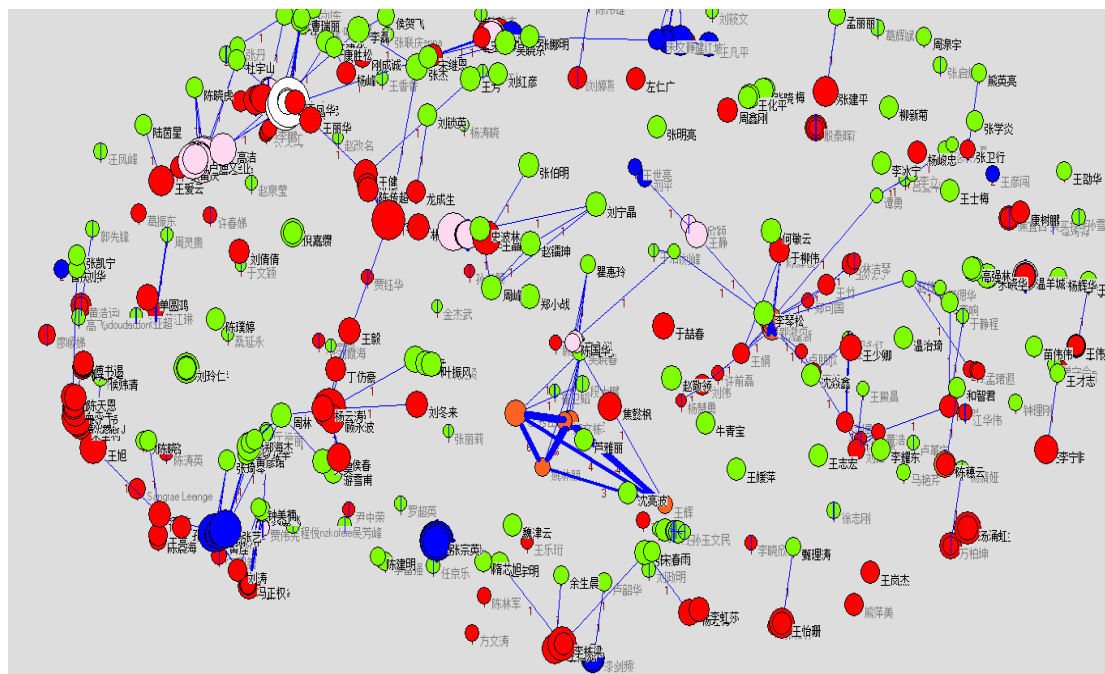


图 7 2009—2013 年国内文献作者合著网络(由度值大于等于 6 的作者节点构成)

4.1.1.1 度的分布

整体作者合著网络的平均度值为3.438,即平均每个作者与另外3.44个作者存在合著关系。整体网络的度分布基本符合幂律,呈现无标度网络的特征(如下图所示,对度值和作者数分别取了对数)。

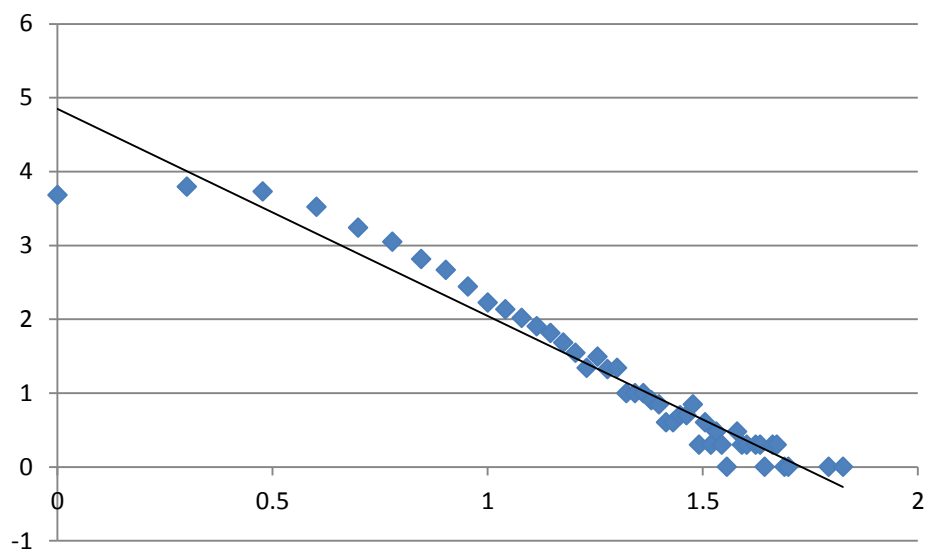


图 8 中文文献作者合著网络的度分布

其中，度值最大的前 20 个作者情况如下：

表 6 中文文献作者合著网络中度值最大的前 20 个作者

排序	度值	作者姓名
1	67	张涛
2	62	王伟
3	50	刘强
4	49	刘伟
5	47	王刚
6	47	王静
7	46	刘勇
8	46	李伟
9	44	王磊
10	43	张伟
11	43	颜贤忠
12	42	张杰
13	42	王健
14	40	刘杰
15	40	张琪
16	39	李强
17	39	李鹏
18	38	刘大有
19	38	张磊
20	38	李彦明

4.1.1.2 网络特征参数

整体作者合著网络中，除去 511341194 个作者对之间不存在合著关系，其余存在合著关系的作者对之间的平均距离为 10.00189，即整体作者合著网络的平均路径长度为 10，这意味着任意两个作者之间平均要通过 9 个人才能互相建立联系。与“六度空间”理论相比，人工智能主题领域中文文献作者合著网络的“小世界效应”较弱，平均网络距离较大。整个

网络中最大的节点对之间的距离为 26，存在于“乔闹生”和“葛荣”两个节点之间，也即网络直径为 26。

整体作者合著网络的网络密度为 0.00013881，说明该网络的整体连接度很低，处于很松散的状态。整体作者合著网络的聚类系数为 0.52461235。利用 pajek 生成一个同样包含 24768 个节点、67349 条边的随机网络，其聚类系数为 0.00011872。可见人工智能主题领域中文文献的整体作者合著网络的聚类系数还是比较高的，这说明该合著网络中存在一些内部彼此熟悉的团体，团体内人员容易形成合著关系。此外，较高的聚类系数反映出“小世界网络”的聚类特性在该合著网络中得到了佐证。

4.1.1.3 介数及其分布

介数最大的前 20 个作者依次为王伟、张涛、刘鹏、刘伟、王健、王娟、李强、刘强、王静、王勇、刘杰、李伟、李静、王磊、刘大有、刘勇、王涛、赵勇、陈伟、张磊。这 20 个作者的介数值之和占有所有作者介数值总和的 14.5%，且前 200 个作者的介数值之和已占总和的 52.97%。可见所有作者的介数值整体呈现前端集中的特征。

4.1.1.4 连通集团规模及其分布

中文文献作者合著网络的最大连通集团包含 10101 个节点，这是一个巨大的连通分支，占网络节点总数的 40.78%。第二和第三大连通集团分别包含 41 个节点和 40 个节点。节点数不小于 10 的连通集团共 126 个。节点数不小于 5 的连通集团共 780 个。

4.1.1.5 年度合著网络特征参数演变

分别统计 2009—2013 年每年的作者合著网络的平均路径长度、网络直径、网络密度、聚类系数，结果如下：

表 7 2009—2013 年年度中文文献作者合著网络的网络特征参数

	2009	2010	2011	2012	2013
平均路径长度	1.99936	1.99937	1.99932	1.99989	1.99985
网络直径	2	2	2	3	4
网络密度	0.000636	0.000633	0.000681	0.000694	0.000952
聚类系数	0.001972	0.001981	0.00223	0.002362	0.003166

可见，2009—2013 年每年的作者合著网络的平均路径长度都近似为 2，即对于年度作者合著网络而言，任意两个作者只需通过 1 个作者就能建立相互联系，具有明显的“小世界效应”。网络直径呈小幅上升趋势，由 2009 年的直径为 2 上升至 2013 年的直径为 4。网络密度略有上升趋势，但是始终低于 0.001，说明年度网络的整体连接度很低，处于很松散的状态。网络的聚类系数也呈上升趋势，由 2009 年的 0.00197 上升至 2013 年的 0.00317，高于一般的随机网络，说明该合著网络中存在一些内部彼此熟悉的团体。

4.1.2 国外文献作者合著网络

英文文献题录中存在一些合著作者非常多的论文，例如“Stand-alone cosmic muon reconstruction before installation of the CMS silicon strip tracker”一文合著作者为 522 人，合著机构有 47 家。这种论文会导致其中的每个作者即便发文量为 1，合著度数都至少为 521，从而影响整体合著网络的特性。因此，我们剔除了合著作者数最多的前 20 篇文献，根据剩余的 22982 条有效英文题录中的作者信息建立合著网络，删去网络中的孤立点，最终得到包含 64237 个节点、204677 条边的整体作者合著网络。下图是取每篇文献的前 6 位作者节点构成的作者合著网络图：

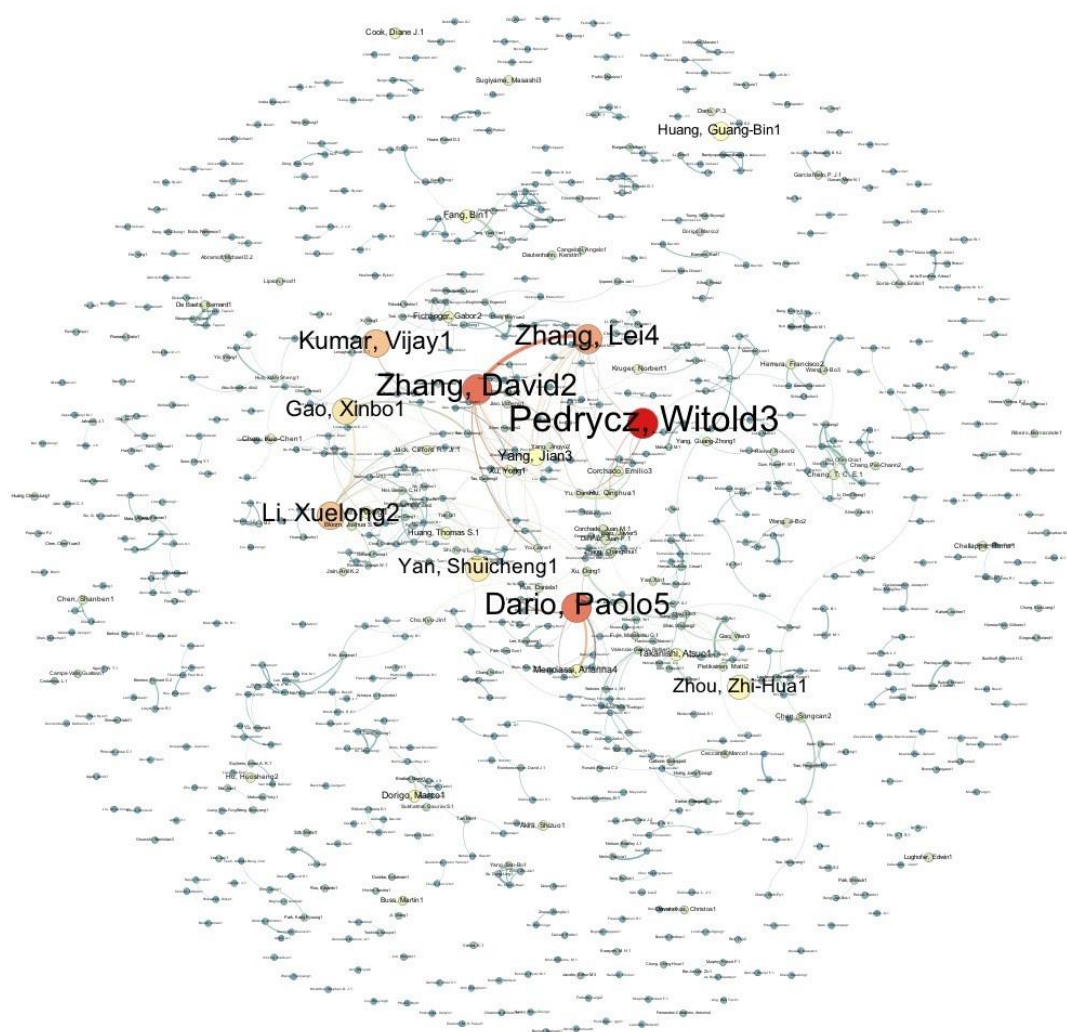


图 9 2009—2013 年国外文献作者合著网络（由每篇文献的前 6 位作者节点构成）

从图中可以明显看出几个度值较大的节点，即合作程度较高的几位作者。其中，Pedrycz Witold 是加拿大亚伯达大学教授，也是 IEEE 成员；Zhang David 是香港理工大学教授；Dario Paolo 是意大利圣安娜高等学校生物机器人学院教授，是生物医药工程教授；Zhang Lei 是香港理工大学计算机学院副教授；Kumar Vijay 是宾夕法尼亚大学教授；Gao Xinbo 是西安电子科技大学电子工程学院教授。

4.1.2.1 度的分布

整体作者合著网络的平均度值为 4.372，即平均每个作者与另外 4.37 个作者存在合著关系。整体网络的度分布基本符合幂律，呈现无标度网络的特征（如下图所示，对度值和作者数分别取了对数）。

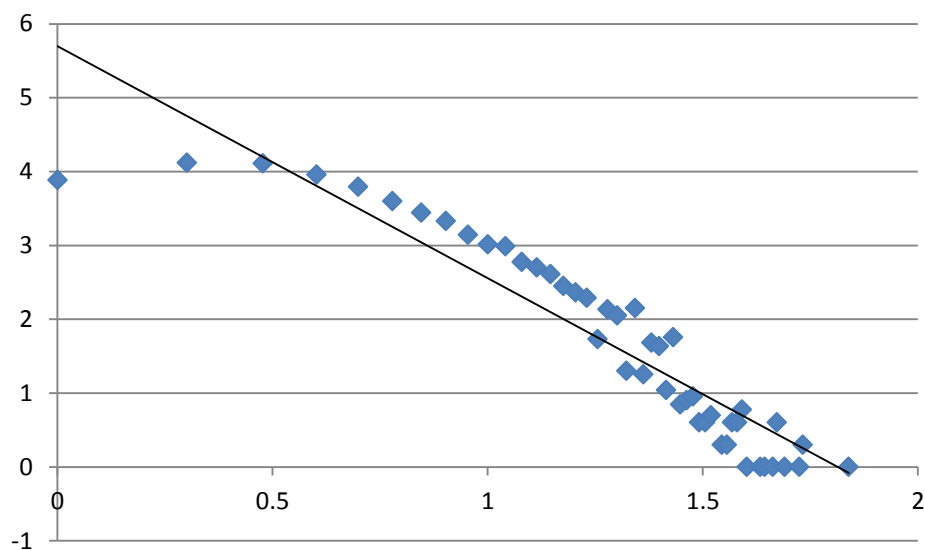


图 10 中文文献作者合著网络的度分布

其中，度值最大的前 20 个作者情况如下：

表 8 英文文献作者合著网络中度值最大的前 20 个作者

排序	度值	作者姓名
1	69	Pedrycz, Witold3
2	54	Akira, Shizuo1
3	54	Bloom, Joshua S.1
4	53	Dario, Paolo5
5	49	Pospisil, S.1
6	47	Dominik, M.1
7	47	Horne, K.1
8	47	Kulkarni, S. R.1
9	47	Tsapras, Y.2
10	46	Zhang, David2
11	44	Sukhatme, Gaurav S.1
12	43	Zhang, Lei4
13	40	Jack, Clifford R., Jr.1
14	39	Dautenhahn, Kerstin1
15	39	Fichtinger, Gabor2
16	39	Hughes, Stephen R.1
17	39	Takanishi, Atsuo1
18	39	Vykydal, Z.1
19	39	You, Jane1
20	38	Bischoff, Kenneth M.1

4.1.2.2 网络特征参数

整体作者合著网络中，除去 4119176358 个作者对之间不存在合著关系，其余存在合著关系的作者对之间的平均距离为 11.50031，即整体作者合著网络的平均路径长度为 11.5，这意味着任意两个作者之间平均要通过 10.5 个人才能互相建立联系。与“六度空间”理论相比，人工智能主题领域英文文献作者合著网络的“小世界效应”较弱，平均网络距离较大。

整个网络中最大的节点对之间的距离为 40, 存在于“Achiche, Sofiane³”和“Cardamone, Luigi¹”两个节点之间, 也即网络直径为 40。

网络密度与聚类系数

整体作者合著网络的网络密度为 0.00006807, 说明该网络的整体连接度很低, 处于很松散的状态。整体作者合著网络的聚类系数为 0.84690165, 相对较高, 而且与聚类系数的最大值 1 差距也不大, 说明该合著网络中存在一些内部彼此熟悉的团体, 团体内人员很容易形成合著关系, 即网络内存在比较普遍的小团体。此外, 较高的聚类系数反映出“小世界网络”的聚类特性在该合著网络中得到了佐证。

4.1.2.3 介数及其分布

介数最大的前 20 个作者依次为 Li, Xuelong²、Zhang, Baochang¹、Xu, Dong¹、Zhang, Lei⁴、Zhou, Huiyu²、Hu, Huosheng²、Wijesoma, Wijerupage Sardha¹、Adams, Martin¹、Patrikalakis, Nicholas M.¹、Leonard, John J.¹、Huang, Thomas S.¹、Zhong, Bineng¹、Tang, Xiaoou²、Liu, Jianzhuang²、Tsang, Ivor W.¹、Whitcomb, Louis L.¹、Han, Jiawei²、Eustice, Ryan M.¹、Yu, Philip S.¹、Cao, Longbing¹。这 20 个作者的介数值之和占有所有作者介数值总和的 39.95%, 且前 200 个作者的介数值之和已占总和的 91.06%。可见所有作者的介数值整体呈现明显的前端集中特征。

4.1.2.4 连通集团规模及其分布

英文文献作者合著网络的最大连通集团包含 2518 个节点, 占网络节点总数的 3.92%。第二和第三大连通集团分别包含 311 个节点和 279 个节点。节点数不小于 10 的连通集团共 975 个。节点数不小于 5 的连通集团共 4275 个。

4.1.2.5 年度合著网络特征参数演变

分别统计 2009—2013 年每年的作者合著网络的平均路径长度、网络直径、网络密度、聚类系数, 结果如下:

表 9 2009—2013 年年度英文文献作者合著网络的网络特征参数

	2009	2010	2011	2012	2013
平均路径长度	1.23287	1.38704	1.28215	1.45916	1.46247
网络直径	5	7	8	7	8
网络密度	0.000319	0.000315	0.000281	0.000234	0.000256
聚类系数	0.954851	0.957318	0.954993	0.950509	0.945957

可见, 2009—2013 年每年的作者合著网络的平均路径长度呈小幅上升趋势, 但是都低于 1.5, 即对于年度作者合著网络而言, 任意两个作者只需通过半个作者就能建立相互联系, 具有明显的“小世界效应”。五年来网络直径始终维持在 7 左右。网络密度呈现下降趋势, 且始终低于 0.001, 说明年度网络的整体连接度很低, 处于很松散的状态。网络的聚类系数始终保持在 0.95 左右, 与聚类系数的最大值 1 非常接近, 说明年度合著网络中存在很多内部彼此熟悉的团体, 团体内人员很容易形成合著关系, 即网络内存在普遍的小团体。综合考察年度作者合著网络的平均路径长度和聚类系数, 可以发现每年的作者合著网络呈现出很短的平均路径长度和很高的聚类系数, 即具有明显的“小世界网络”的特性。

4.1.3 国内外文献作者合著对比

对比国内外文献作者合著网络, 主要有以下几点发现:

1. 中英文文献作者合著网络的节点度分布均符合幂律, 呈现无标度网络的特性, 且英

文文献作者合著网络的平均度值高于中文文献作者合著网络。

2. 中英文文献作者合著网络的平均路径长度类似（均略大于 10），但英文文献作者合著网络的聚类系数明显高于中文文献作者合著网络。

3. 中英文文献作者合著网络的介数分布均呈现前端集中特征，且英文文献作者合著网络介数分布的前端集中特征更明显。

4. 英文文献作者合著网络的最大连通集团的规模明显低于中文文献作者合著网络的最大连通集团。

4.2 机构合著网络分析

在本研究中，机构合著网络以机构为节点，节点的大小代表节点的度值。如果两个机构合著过一篇文章，则它们之间就存在一条无向边，边的权重为两个机构合著的次数。下面分别分析近五年来人工智能主题领域国内外文献的机构合著网络，并对国内外机构合著情况进行对比。

4.2.1 国内文献机构合著网络

根据 14646 条有效中文题录中的机构信息建立合著网络，删去网络中的孤立点（即没有与任何其他机构合著过的孤立机构节点），最终得到包含 4091 个节点、5647 条边的整体机构合著网络，如下图所示：

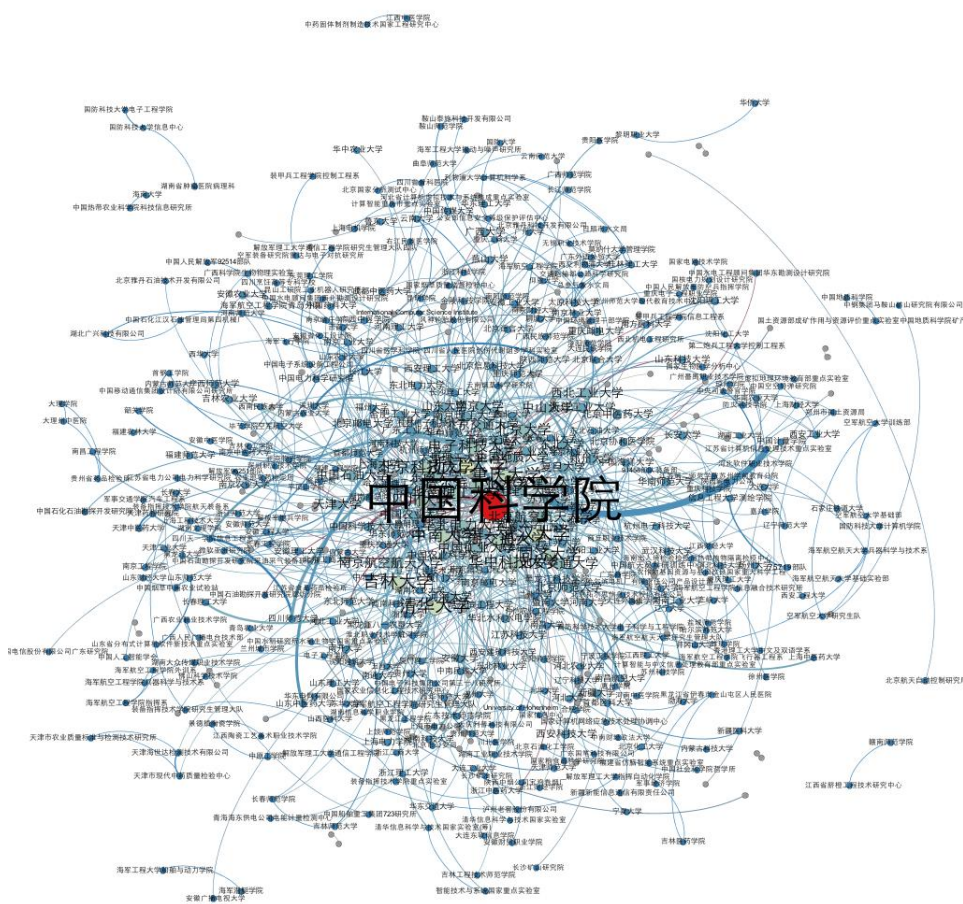


图 11 2009—2013 年国内文献机构合著网络

4.2.1.1 度的分布

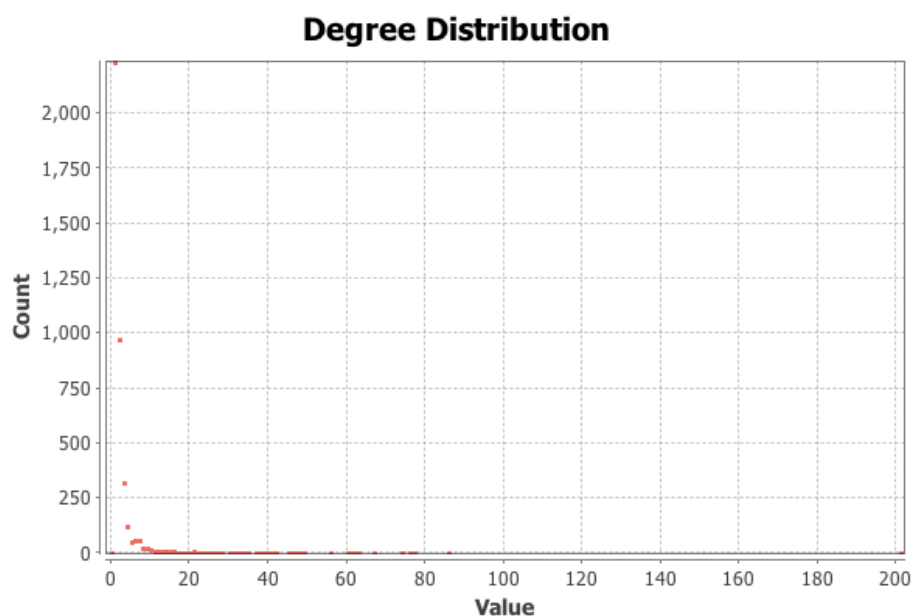


图 12 中文文献机构合著网络的度分布

度分布如上图所示，大体符合幂率分布。其中度数最大的 10 个节点如下图所示。中科院的度数远大于其他节点，排在之后的是浙江大学（86），武汉大学（77），清华大学（76），吉林大学（74）。

表 10 中文机构合著度数前十的机构

机构名称	度值
中国科学院	202
浙江大学	87
武汉大学	77
清华大学	76
吉林大学	74
重庆大学	72
中南大学	67
上海交通大学	63
北京科技大学	62
同济大学	62

4.2.1.2 网络特征参数

整体机构合著网络中，存在合著关系的机构对之间的平均距离为 4.438，即整体作者合著网络的平均路径长度约为 4.5，这意味着任意两个机构之间平均通过 3.5 个机构即可互相建立联系，具有明显的“小世界效应”。整个网络中最大的节点对之间的距离为 11，存在于“军械工程学院基础部”和“香港中文大学计算机科学与工程学系”两个节点之间，也即网络直径为 11。

整体机构合著网络的网络密度为 0.00067499，说明该网络的整体连接度很低，处于很松散的状态。整体机构合著网络的聚类系数为 0.05950900，高于一般的随机网络，说明该合著网络中存在一些内部彼此熟悉的团体。综合考察年度作者合著网络的平均路径长度和聚类系数，可以发现每年的作者合著网络呈现出较短的平均路径长度和较高的聚类系数，即具

有“小世界网络”的特性。

4.2.1.3 介数及其分布

如下表所示，介数大于 0.03 的节点有六个，与度数最大的节点基本吻合。

表 11 介数大于 0.03 的节点

机构名称	度值	相对介数
中国科学院	202	0.184
浙江大学	87	0.038
清华大学	76	0.035
武汉大学	77	0.033
吉林大学	74	0.031
北京科技大学	62	0.031

介数最大的前 20 个机构的介数值之和占所有机构介数值总和的 37.56%，且前 200 个机构的介数值之和已占总和的 86.81%。可见所有机构的介数值整体呈现前端集中的特征。

4.2.1.4 连通集团规模及其分布

图中连通子图共有 480 个，其中存在一个节点数为 2927 的极大连通子图，占据了整个图节点的 71.55%。这个连通子图涵盖了绝大部分的高校。其他连通子图最大的包含 21 个节点，主要是军队中的机构。

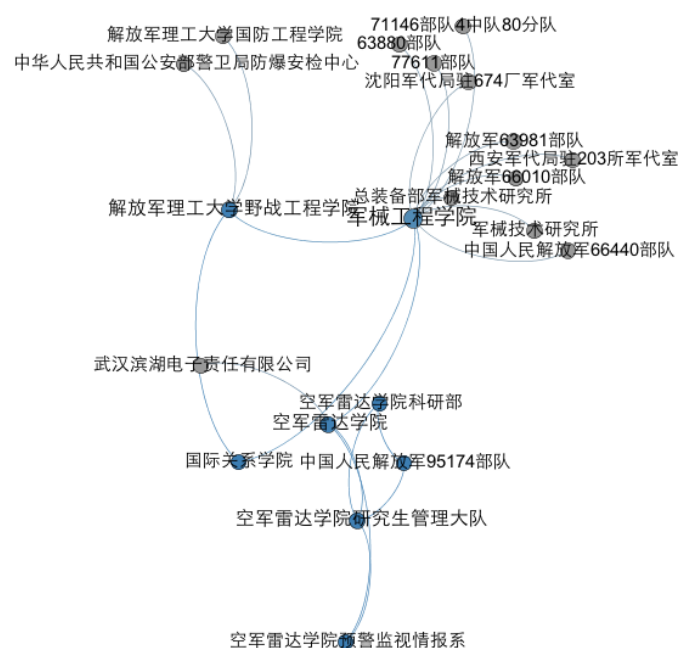


图 13 军队机构连通子图

4.2.1.5 年度合著网络特征参数演变

分别统计 2009—2013 年每年的机构合著网络的平均路径长度、网络直径、网络密度、聚类系数，结果如下：

表 12 2009—2013 年年度中文文献机构合著网络的网络特征参数

	2009	2010	2011	2012	2013
平均路径长度	5.35998	5.29669	6.06926	5.48015	5.33703
网络直径	12	14	15	13	13
网络密度	0.001869	0.001727	0.001785	0.001882	0.002211
聚类系数	0.131148	0.127704	0.165022	0.159494	0.179487

可见，2009—2013 年每年的机构合著网络的平均路径长度基本都保持在 5 左右，即对于年度机构合著网络而言，任意两个作者通过 4 个作者就能建立相互联系，具有明显的“小世界效应”。五年来网络直径始终维持在 13 左右。网络密度呈现先下降后上升的趋势，不过基本处于 0.002 左右，说明年度网络的整体连接度比较低，处于比较松散的状态。网络的聚类系数呈现波动上升趋势，从 2009 年的 0.13 增长至 2013 年的近 0.18，高于一般的随机网络，说明该合著网络中存在一些内部彼此熟悉的团体。综合考察年度作者合著网络的平均路径长度和聚类系数，可以发现每年的作者合著网络呈现出较短的平均路径长度和较高的聚类系数，即具有“小世界网络”的特性。

4.2.2 国外文献机构合著网络

与国外文献作者合著网络部分的处理类似，我们剔除了合著作者数最多的前 20 篇文献，根据剩余的 22982 条有效英文题录中的机构信息建立合著网络，删去网络中的孤立点，最终得到包含 8844 个节点、28322 条边的整体作者机构网络，如下图所示：

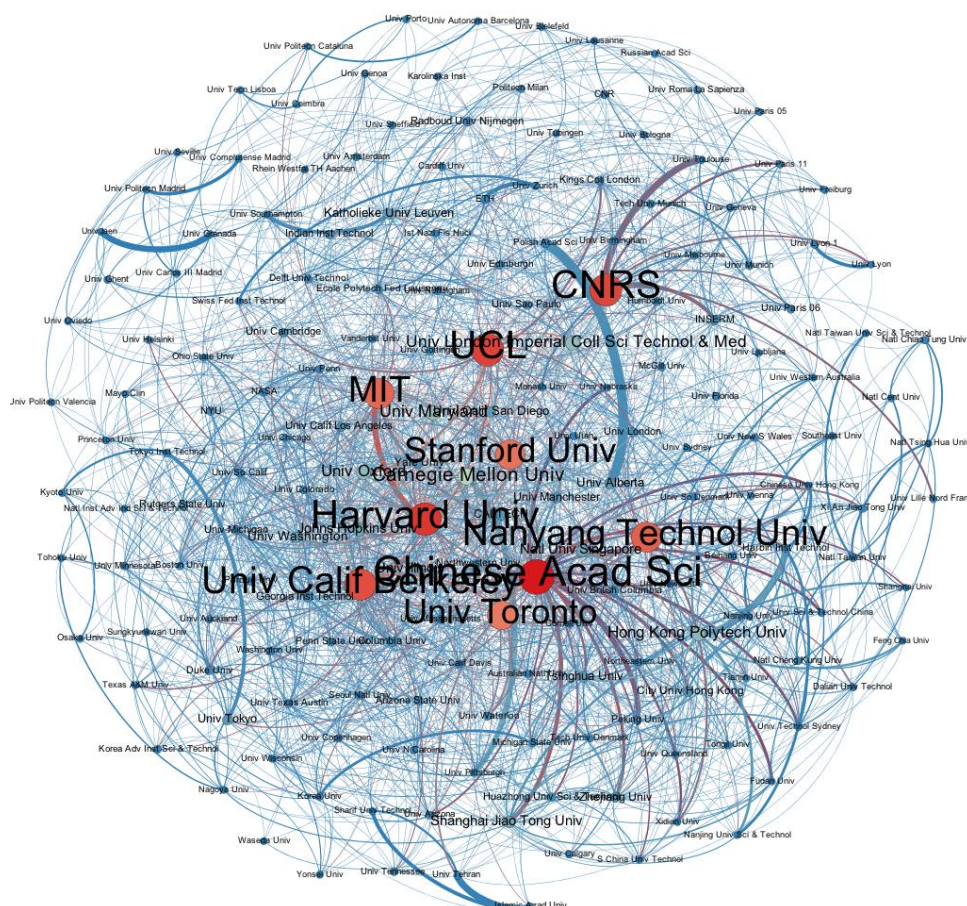


图 14 2009—2013 年国外文献机构合著网络

4.2.2.1 度的分布

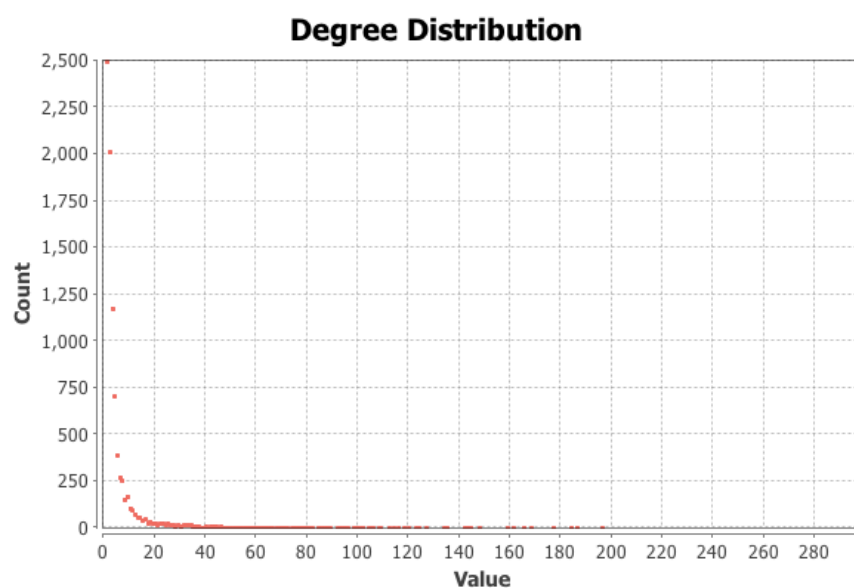


图 15 英文文献机构合著网络的度分布

如上图所示，机构合著的度分布大体符合幂率分布，是无标度网络。整个图的平均度数为 6.407。其中中国科学院的度最大为 298，第二为哈佛大学，第三为伦敦大学学院。在机构合作方面，中国科学院远超其他国家的机构。在带权度数前 20 的机构中，我国的中科院，香港理工，上海交通大学，清华大学，香港城市大学位列其中。可见我国的人工智能领域的机构合作活跃度是比较高的。

表 13 英文文献机构合著网络中度数前五的节点

机构名称	度数
Chinese Acad Sci	298
Harvard Univ	196
伦敦大学学院 UCL	186
法国国家科学研究中心 CNRS	184
Univ Calif Berkeley	177

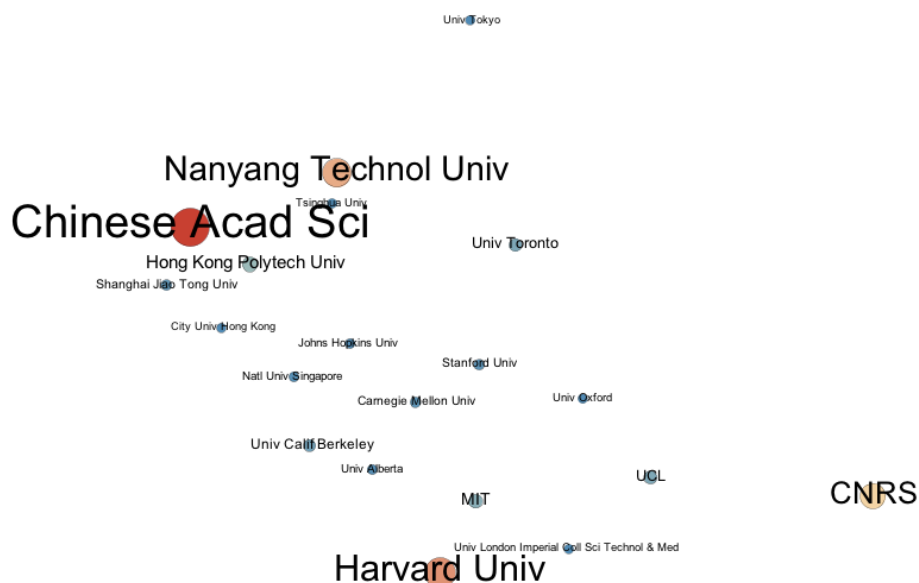


图 16 英文文献机构合著网络中带权度数前 20 的机构

查看中科院的节点，可见与中科院合作程度高的节点依然是国内的一些机构，例如上海交通大学，北京大学，清华大学，浙江大学，中国科技大学；相对于和它合作程度较高的国外机构，如南阳理工，内布拉斯加大学，东北大学，亚利桑那大学，东京大学，合作紧密程度要高一些。

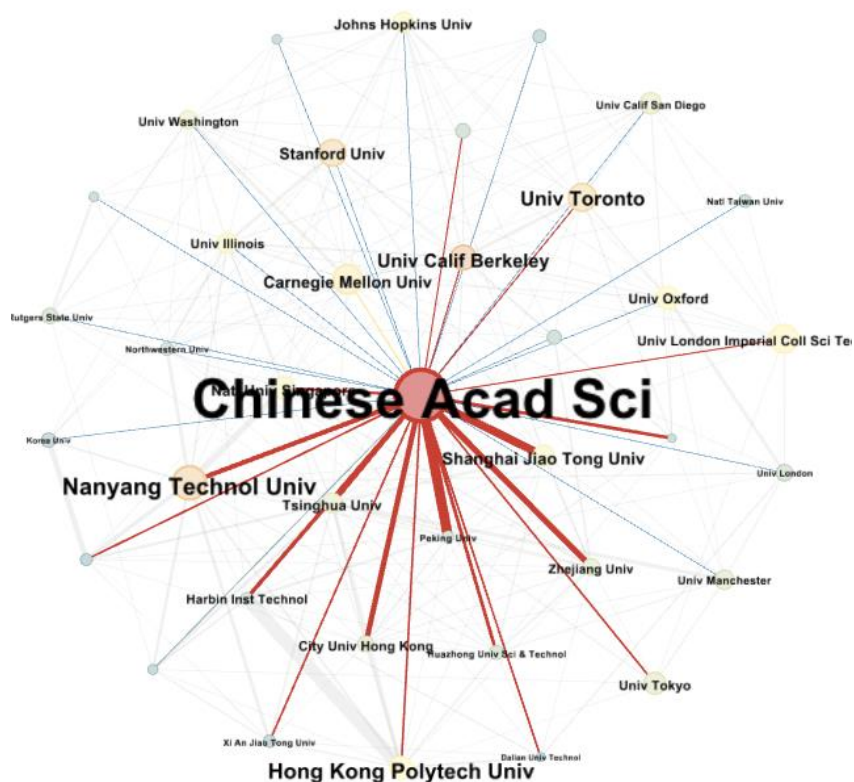


图 17 中科院与其他机构的合作情况示意

再看哈佛大学的节点，可能是由于地理位置的原因，它与 MIT 的合作程度非常高。除此之外，由于麻省总医院（Massachusetts Gen Hosp）是美国哈佛医学院的教学医院之一，

它们之间的合作程度也比较高。

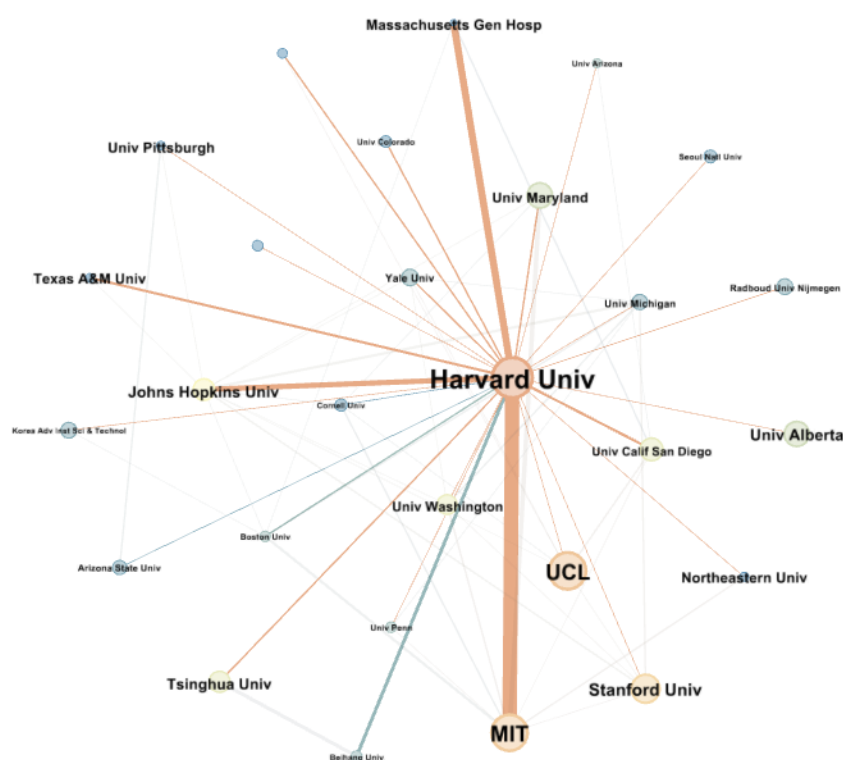


图 18 哈佛大学与其他机构的合作情况示意

4.2.2.2 网络特征参数

整体机构合著网络中，除去 13460042 个机构对之间不存在合著关系，其余存在合著关系的机构对之间的平均距离为 4.13713，即整体作者合著网络的平均路径长度约为 4，这意味着任意两个机构之间平均通过 3 个机构即可互相建立联系，具有明显的“小世界效应”。整个网络中最大的节点对之间的距离为 13，存在于“Akoren Cok Programli Lisesi”和“YIET”两个节点之间，也即网络直径为 13。

整体机构合著网络的网络密度为 0.00072453，说明该网络的整体连接度很低，处于很松散的状态。整体机构合著网络的聚类系数为 0.11879164，高于一般的随机网络，说明该合著网络中存在一些内部彼此熟悉的团体。综合考察年度作者合著网络的平均路径长度和聚类系数，可以发现每年的作者合著网络呈现出很短的平均路径长度和较高的聚类系数，即具有“小世界网络”的特性。

4.2.2.3 介数及其分布

如下表所示，介数大于 0.03 的节点有六个，与度数最大的节点基本吻合。

表 14 介数大于 0.03 的节点

机构名称	度值	相对介数
Chinese Acad Sci	298	0.054
Harvard Univ	196	0.035
CNRS	184	0.035
UCL	186	0.032
MIT	165	0.031

介数最大的前 20 个机构的介数值之和占所有机构介数值总和的 16.4%，且前 200 个机构的介数值之和已占总和的 58.72%。可见所有机构的介数值整体呈现前端集中的特征。

4.2.2.4 连通集团规模及其分布

图中连通子图共有 336 个，其中存在一个节点数为 8047 的极大连通子图，占据了整个图节点的 90.99%。其他连通子图最大的只有 7 个节点，多是几个不太知名的机构一起合著了一篇文章所构建的。例如下图的这个网络就是由 7 家机构只合著一篇文章而形成的。

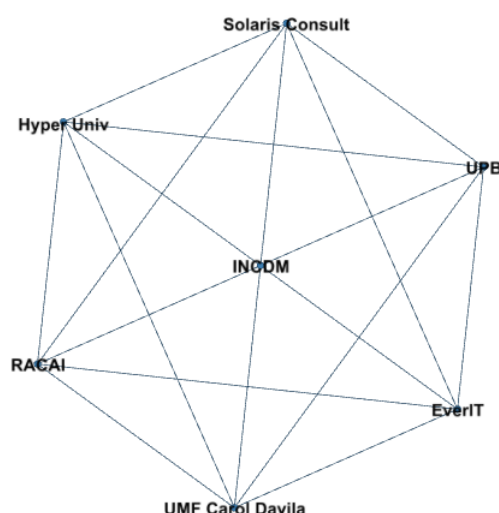


图 19 7 家机构合著一篇文章形成的小规模连通子图

4.2.2.5 年度合著网络特征参数演变

分别统计 2009—2013 年每年的机构合著网络的平均路径长度、网络直径、网络密度、聚类系数，结果如下：

表 15 2009—2013 年年度英文文献机构合著网络的网络特征参数

	2009	2010	2011	2012	2013
平均路径长度	5.28981	5.06841	5.09836	4.98174	4.73248
网络直径	12	14	13	14	10
网络密度	0.001578	0.001545	0.001434	0.001259	0.00139
聚类系数	0.265097	0.30019	0.295662	0.225356	0.228992

可见，2009—2013 年每年的机构合著网络的平均路径长度呈现小幅下降趋势，不过基本保持在 5 左右，即对于年度机构合著网络而言，任意两个作者通过 4 个作者就能建立相互联系，具有明显的“小世界效应”。五年来网络直径始终维持在 13 左右。网络密度呈现波动下降趋势，基本处于 0.0014 左右，说明年度网络的整体连接度比较低，处于比较松散的状态。网络的聚类系数呈现先上升后下降的趋势，不过始终高于 0.22，高于一般的随机网络，说明该合著网络中存在一些内部彼此熟悉的团体。综合考察年度作者合著网络的平均路径长度和聚类系数，可以发现每年的作者合著网络呈现出较短的平均路径长度和较高的聚类系数，即具有“小世界网络”的特性。

4.2.3 国内外文献机构合著对比

对比国内外文献机构合著网络，主要有以下几点发现：

1. 从中英文文献机构合著网络的平均聚类系数来看，英文文献机构合著网络的聚类系数明显高于中文文献机构合著网络；
2. 中文文献机构合著网络有明显的核心（中国科学院），而英文文献机构合著网络则具有多个大的节点，呈现出多极化的特点；
3. 英文文献机构合著网络的最大连通集团规模明显高于中文文献机构合著网络的最大连通集团。

4.3 国家合著网络分析

国家合著是指一篇文献由来自于不同国家的学者合作完成这样一种现象，它反映出某一科研领域的国际合作水平。在本研究中，国家合著网络以国家为节点，节点的大小代表节点的度值。如果两个国家合著过一篇文章，则它们之间就存在一条无向边，边的权重为两个国家合著的次数。根据 23002 条有效英文题录进行国家抽取，生成国家合著表，导入 Gephi 进行作图，以绘制国家合著网络。为精简网络，去掉度值小于 10 的节点，得到如下图所示的国家合著网络：

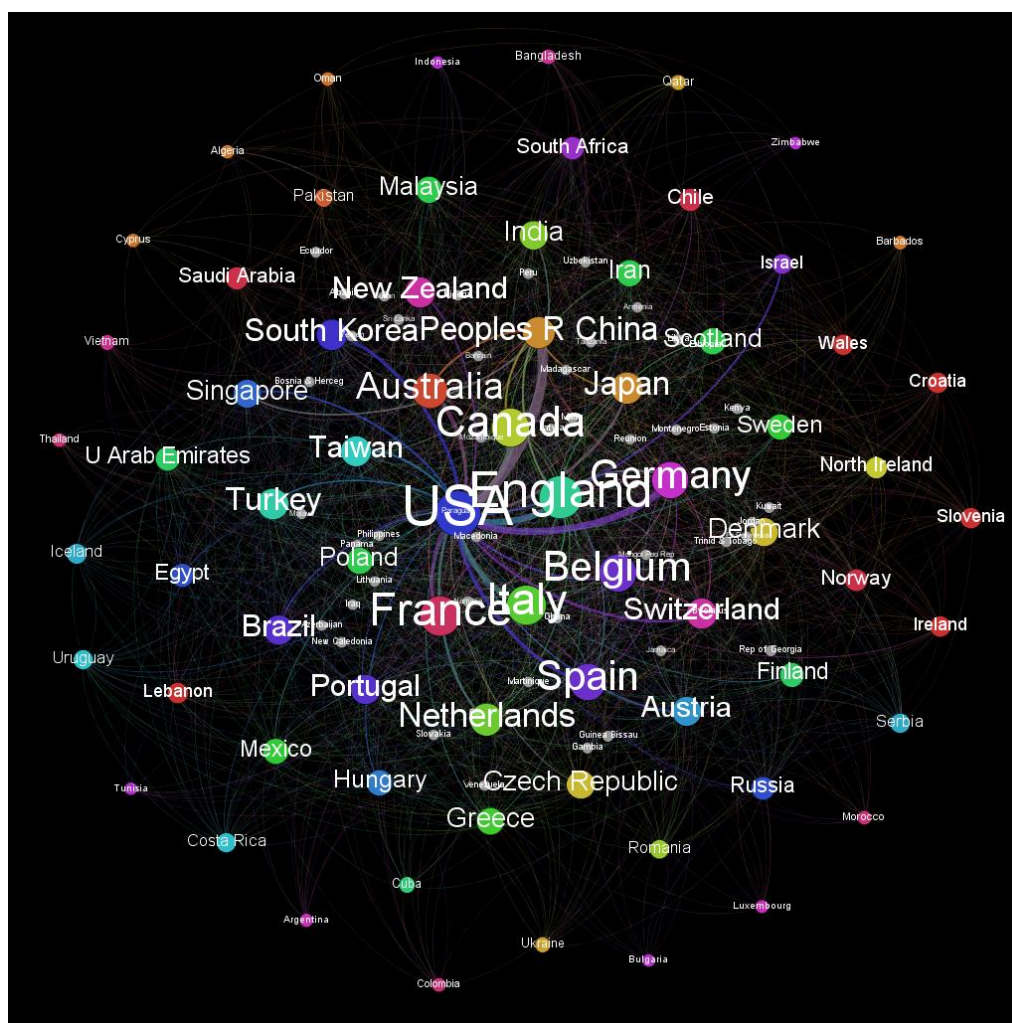


图 20 2009—2013 年国家合著网络（由度值大于等于 10 的节点构成）

目前在人工智能领域进行研究的国家有 118 个，从图中可以看出，美国在人工智能研究领域的国际合著现象最为广泛，和 84 个国家都存在合作，合作次数达到了 5362 次。其次是英格兰、法国、意大利、加拿大、比利时、西班牙、德国、澳大利亚、荷兰、日本、中国。从度值的大小和分布可以看出，各个国家在该领域的合作比较广泛，其中美国、英国、美国和中国、美国和中国之间的科研合作比较紧密。

4.3.1 度的分布

国家合著网络的度分布并不服从幂律，这可能是由于节点数目有限，网络规模较小而造成的。国家合著网络的平均度值为 22.847，在 118 个国家中，超过 44% 的国家度值超过平均数，有 88% 的节点度值超过 1，换言之，仅和一个国家合著过文章的国家数量很少，大部分国家合著范围都比较广。下图是由度值超过平均值的国家构成的合著网络：

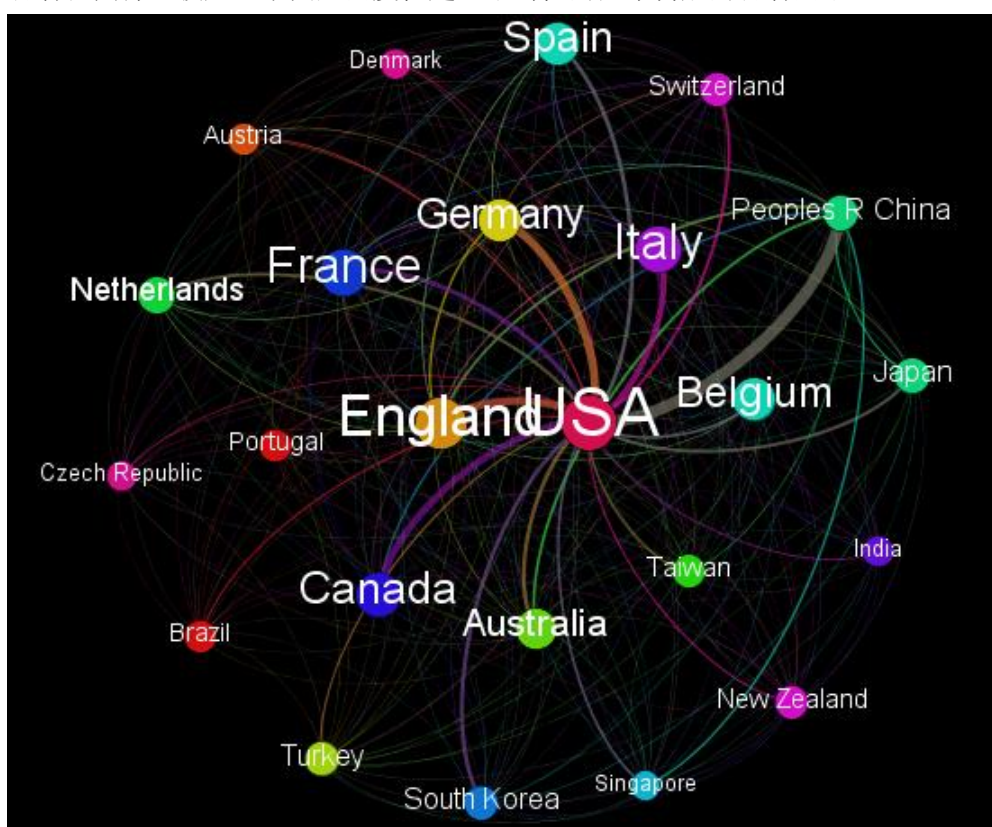


图 21 度值超过平均值的国家构成的合著网络
由度值低于平均值的国家构成的合著网络则如下图所示：

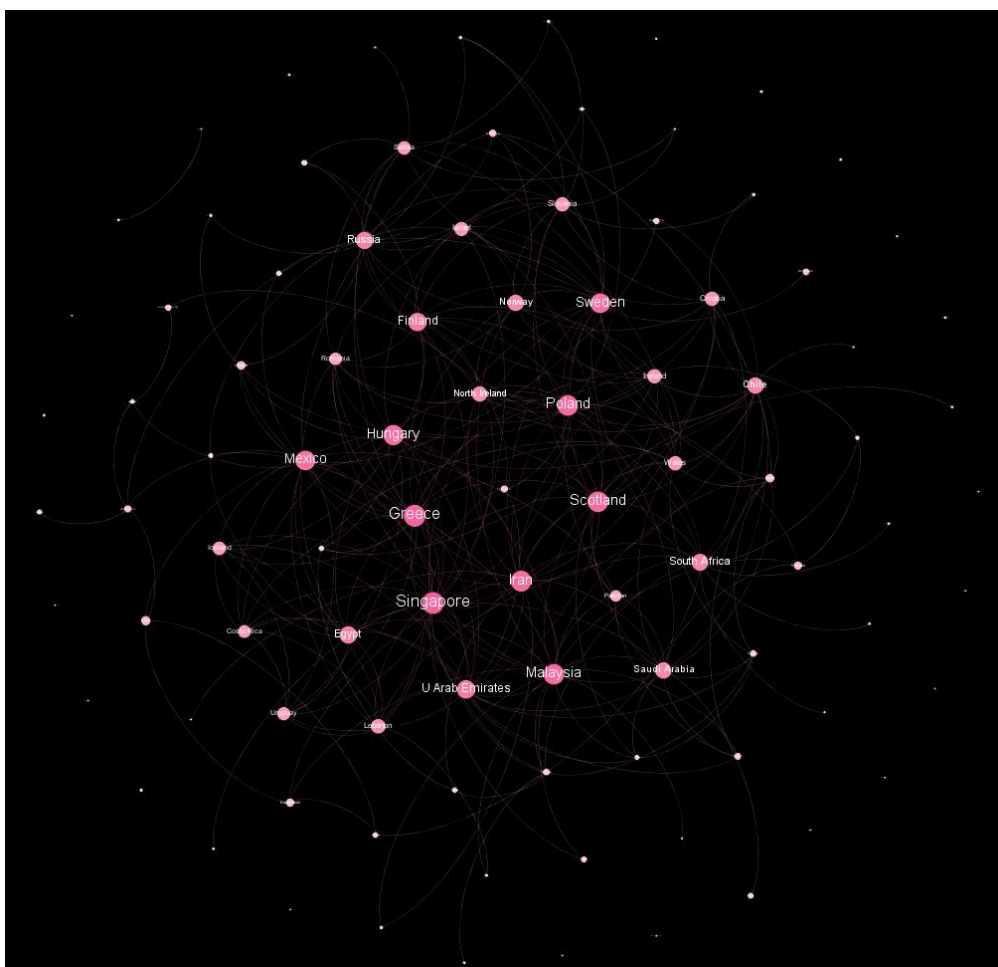


图 22 度值较小的国家构成的合著网络

整体国家合著网络中，加权度的平均值为 432，在 118 个国家中，只有超过 11% 的国家加权度值超过了平均数。合作发文量超过平均值的国家的加权度值如下图所示：

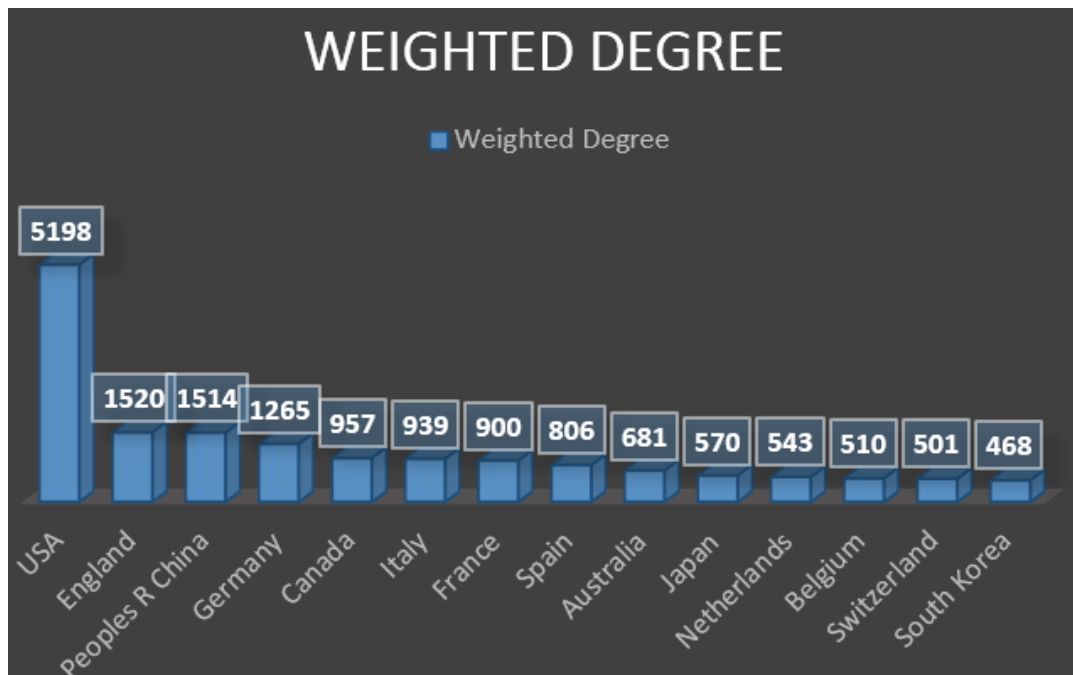


图 23 合作发文量超过平均值的国家的加权重度值

从以上三图中可以看出，度值排名靠前的国家在该领域的研究很紧密，国家之间交流广泛，而度值排名靠后的国家在该领域的研究比较松散，国家之间交流较少。从合作发文量来看，目前发文量较高的国家集中在美国、英国、中国、德国、加拿大、意大利等国家，一些国家例如中国的合作国家数只有 53 个，但是合作发文量则达到了 1514 篇，说明其与很多国家之间存在紧密的合作关系。

4.3.2 网络特征参数

表 16 国家合著网络的网络特征参数

	网络节点	平均路径长度	网络密度	聚类系数
国家合著网络	118 个合著国家	1.98	0.19	0.59

国家合著网络中的平均路径长度为 1.98，即任意两个国家之间平均只要通过 1 个国家就可以互相建立联系，与其他网络相比，这是一个较低的网络距离，具有“小世界效应”，因此是一个信息通畅、合作国家之间可以进行快捷交流的网络。

国家合著网络的密度为 0.19，与实际网络的最大密度值 0.5 相比，网络密度低，因此国家的合著网络整体连接比较稀疏。

与其他社会网络相比，国家合著网络的聚类系数相对较高，这进一步佐证了小世界效应在国家合著网络中的存在。另一方面也说明网络内存在一些内部彼此熟悉的团体，团体内国家之间容易形成合著关系。

总的来看，人工智能领域的国家合著网络具有较小的平均路径长度、较高的聚类系数、较低的网络密度。这三个特征参数表明，在实际网络中，整体的连接比较稀疏，同时网络中存在着很多紧密联系的“小团体”，虽然小团体内部联系紧密，但是团体之间联系松散。

4.3.3 介数及其分布

在国家合著网络中，介数的最小值为 0，整体网络的全部 118 个国家中，有 32% 的节点介数为 0，在剩余的国家中，又有 32% 的节点介数小于 10。另一方面，有 20 个国家的介数

大于 100，并且这 20 个国家的介数值占到了总体的 80%以上，拥有相当高的资源控制能力。所以，在国家合著网络中，资源控制能力的分布不平衡，较少的国家控制了大量的资源，而大部分国家则缺乏控制资源的能力。度和介数排名前 20 位的国家如下表所示：

表 17 国家合著网络中度和介数排名前 20 的国家

名次	国家（度）	国家（介数）
1	美国	美国
2	英国	英国
3	法国	法国
4	意大利	意大利
5	加拿大	西班牙
6	比利时	比利时
7	西班牙	德国
8	德国	韩国
9	澳大利亚	加拿大
10	荷兰	印度
11	日本	中国
12	中国	马来西亚
13	土耳其	巴西
14	韩国	葡萄牙
15	新西兰	新西兰
16	瑞士	伊朗
17	台湾	澳大利亚
18	巴西	新喀里多尼亚
19	葡萄牙	土耳其
20	奥地利	南非

从上表中可以看出，度值和介数之间有着较强的关联性，通常而言度值大的国家介数值也大，这表明少数国家在和大量国家合著的同时，还为其他国家提供了相互联系的机会。

4.3.4 连通集团规模及其分布

国家合著网络的最大连通集团包括了所有的 118 个顶点，也就是说国家合著网络总体是一个巨大的连通分支，任意两个国家都能通过若干个国家产生联系。

4.3.5 核心—边缘结构分析

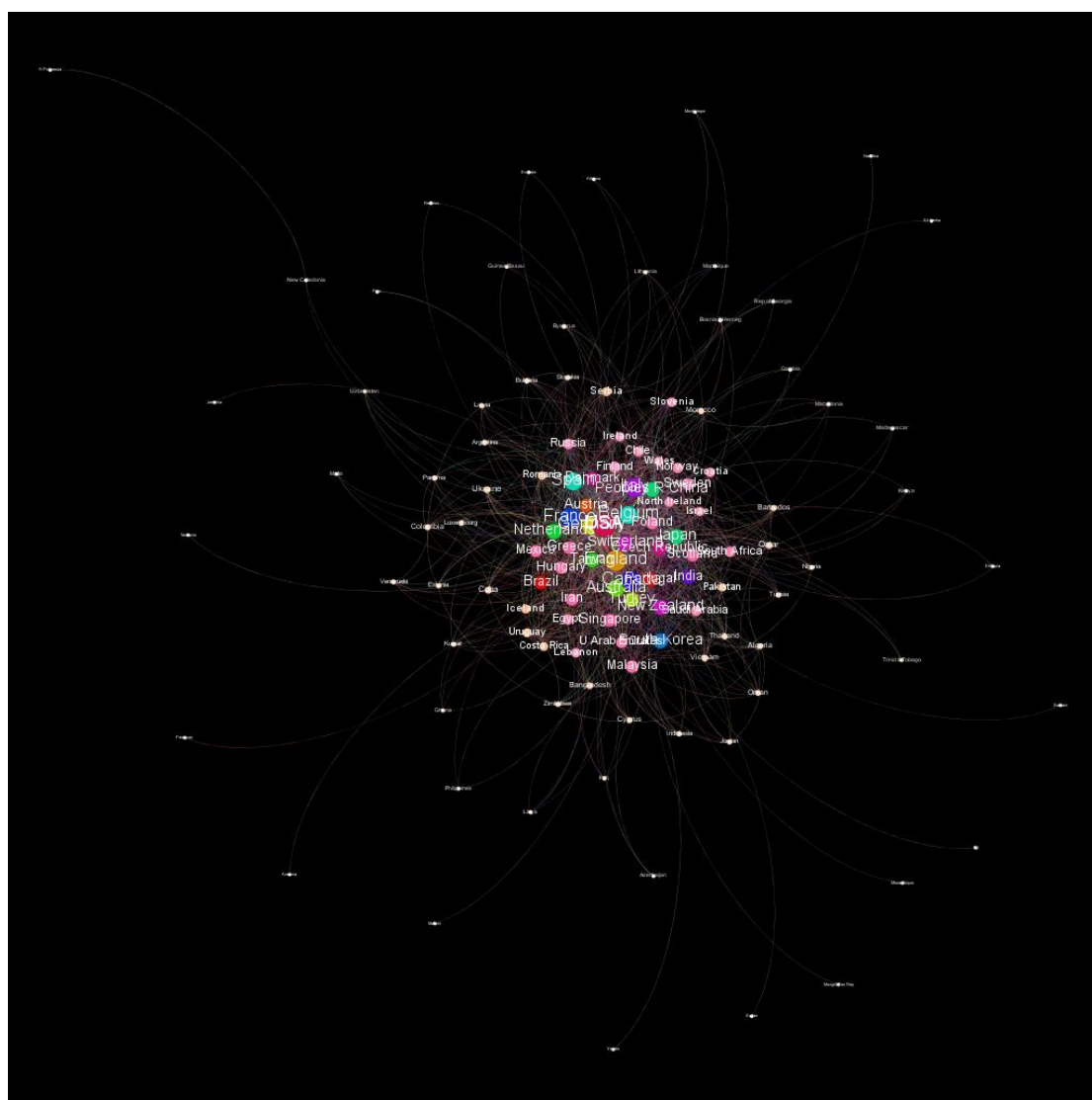


图 24 国家合著网络的核心—边缘结构

从根据接近中心度形成的国家合著网络核心—边缘结构来看，在人工智能领域，美国、英国、德国等发达国家占据着核心地位，这些国家在该领域的研究成果也一直处于领先地位，对其他国家的研究有引导作用。人工智能领域国家合著网络的核心—边缘结构包含几层含义：①各国间的合作网络只有一个核心，并不能分为几个相对独立的簇；②在国际合作中，可以分为处于合作核心的国家和处于边缘的国家两类；③处于合作核心国家间的合作强度明显高于处于合作边缘的国家间的合作强度；④处于合作边缘的国家有与处于合作核心国家进行合作的趋向。

4.4 关键词共现网络分析

在本研究中，关键词共现网络以关键词为节点，节点的大小代表节点的度值。如果两个关键词出现在同一篇文章内，则它们之间就存在一条无向边，边的权重为两个关键词共现的次数。下面分别分析近五年来人工智能主题领域国内外文献的关键词共现网络，并对国内外文献关键词共现情况进行对比。

4.4.1 国内文献关键词共现网络

根据 14646 条有效中文题录中的关键词信息建立共现网络，删去网络中的孤立点（即没有与任何其他关键词共现过的孤立关键词节点），最终得到包含 27063 个节点、101699 条边的整体关键词共现网络。

4.4.1.1 网络基本情况

2009—2013 年人工智能领域的国内文献关键词共现网络有 27063 个节点，101699 条边，是一个节点多、边多的大型复杂网络。网络基本参数如表所示。

表 18 国内关键词共现网络基本参数

平均度	平均加权重	网络直径	平均路径长度	图密度	平均聚类系数
7.516	8.696	9	3.00	0.000	0.871

该网络的平均度值为 7.516，平均路径长度为 3，聚类系数为 0.871，具有很小的平均路径长度和很大的聚类系数，存在明显的“小世界网络”特征；网络密度很小，接近为 0，说明在关键词共现网络中，所有节点的整体联系较为松散；聚类系数为 0.871，值很大，说明该网络中存在明显的学科群，这些学科群内部各学科联系较为紧密，而学科群之间联系比较松散。

4.4.1.2 中心性分析

统计国内文献关键词共现网络的节点度数、介数及接近中心性，发现所有关键词的介数和接近中心性都为零，说明在该关键词共现网络中，各个关键词的位置分布都较为平均，没有处于核心位置、具有重要连接作用的关键词，这也说明该网络较为松散。由于介数和接近中心度为零，我们考察中心性只参考度数，度数最高的 10 个关键词如表所示。

表 19 国内文献关键词共现网络度数最高的前 10 个关键词

关键词	度数
人工智能	3485
专家系统	3437
模式识别	3030
机器学习	1569
支持向量机	1382
故障诊断	1249
神经网络	1120
自然语言处理	821
数据挖掘	703
特征提取	588

度数最高的 10 个关键词中，许多关键词也是我们的检索词，自然它们成为该网络中较为核心的关键词。此外，“支持向量机”“神经网络”“数据挖掘”“特征提取”这 4 个与数据挖掘有关的关键词的度数排名也较靠前，说明数据挖掘在人工智能领域研究中具有比较重要的位置，国内对人工智能的研究主要集中在对人类知识的挖掘。

4.4.1.3 聚类分析

对国内文献的关键词进行聚类，得到 671 个类。其中有 5 个类位于网络的核心位置，占网络的 62.95%，其余 666 个类均低于 5%，共占 37.05%。这 5 个大类所包含的关键词数量很多，涉及的学科及主题分布较广，不易对关键词类进行概括和命名，在此不再罗列。

4.4.2 国外文献关键词共现网络

根据 23002 条有效英文题录中的关键词信息建立关键词共现网络，删去网络中的孤立点，最终得到包含 56372 个节点、244786 条边的整体关键词共现网络。

4.4.2.1 网络基本情况

2009—2013 年人工智能领域国外文献的关键词共现网络的基本特征参数如下表所示：

表 20 国外文献关键词共现网络基本特征参数

平均度	平均加权度	网络直径	平均路径长度	网络密度	平均聚类系数
8.685	9.392	10	6.203	0.000	0.873

该网络的平均路径长度为 6.203，聚类系数为 0.873，具有较小的平均路径长度和很大的聚类系数，即存在明显的“小世界网络”特征；网络密度很小，接近为 0，说明在关键词共现网络中，所有节点的整体联系较为松散；平均聚类系数为 0.873，值很大，说明该网络中存在明显的学科群，这些学科群内部各学科联系较为紧密，而学科群之间联系比较松散。

4.4.2.2 中心性分析

中心性主要从三个方面进行考察：度数、介数以及接近中心度。统计关键词的度数、介数及接近中心性，和国内文献关键词共现网络类似，所有关键词的介数和接近中心性都为零，说明在国外文献关键词共现网络中，各个关键词的位置分布都较为平均，没有处于核心位置、具有重要连接作用的关键词，这也从另一方面印证了该网络较为松散。同样，我们考察中心性只参考度数，对度数进行排序，国外文献关键词共现网络度数最高的 10 个关键词如表所示。

表 21 国外关键词共现网络度数最高的前 10 个关键词

关键词	度数
Machine Learning	4125
Pattern Recognition	2896
Support Vector Machines	2686
Feature Selection	1931
Classification	1851
Neural Networks	1849
Robotics	1766
Data Mining	1137
Artificial Intelligence	1118
Fuzzy Logic	876

国外文献关键词度数最高的 10 个关键词中，“Machine Learning”“Pattern Recognition”“Robotics”“Artificial Intelligence”4 个关键词是我们的检索词，自然成为网络中较为核心的

关键词。此外，“Support Vector Machines”“Classification”“Feature Selection”“Fuzzy Logic”“Neural Networks”“Data Mining”这 4 个与数据挖掘有关的关键词的度数排名也较靠前，说明数据挖掘在国外的人工智能领域研究中也具有比较重要的位置，国外对人工智能的研究也主要集中在对人类知识的挖掘。

4.4.3 国内外文献关键词共现网络对比

通过对比国内外文献关键词共现网络可知，在人工智能研究领域，国内外有许多共同点。首先，两个网络都比较稀疏，说明国内外人工智能领域研究涉及的主题非常多，遍及各个学科，因此整个网络较为分散；同时，两个网络聚类系数都很高，可以对网络进行聚类，而且聚类处理后发现整个网络是由少数大类控制的，占据网络的大部分资源，说明有些关键词的共现频次很高。其次，无论国内外，中心性较高的关键词都有人工智能、模式识别、机器学习、支持向量机、神经网络、数据挖掘、特征提取，尤其是“数据挖掘”“支持向量机”“特征提取”“模式识别”这些数据挖掘领域的关键词，虽然不是检索词，仍在网络中具有重要的地位，说明国内外人工智能领域当前的研究重点都是对人类知识、语言、思维模式的挖掘，以使机器能够更好地理解、模拟人的行为。

然而，当前国内外人工智能领域的研究仍存在一些差别。首先，国内关键词共现网络的平均路径为 3，而国外约为 6，国内网络的关键词之间的间隔较少，说明当前国内人工智能领域的研究所涉及范围不如国外广泛，平均每两个关键词之间通过 2 个关键词就可以建立联系。此外，国外度数比较靠前的关键词有机器人学、模糊逻辑等，而国内度数比较靠前的关键词有自然语言处理，说明国内研究仍主要集中在理论上的研究，而国外已经开展了有关实践，如通过机器人对人类智能进行仿生；此外，国外在理论上的研究也较为领先，模糊逻辑已经成为重要的关键词。

4.5 学科共现网络分析

由于国内文献题录的“学科类别”字段数据是缺失的，因此对学科的共现网络分析只针对国外文献。在本研究中，学科共现网络以学科为节点，节点的大小代表节点的度值。如果一篇文章同时归属于两个学科，则这两个学科之间就存在一条无向边，边的权重为两个学科共现的次数。根据 23002 条有效英文题录中的学科信息建立学科共现网络，删去网络中的孤立点，最终得到包含 170 个节点、760 条边的整体学科共现网络。

4.5.1 整体网络分析

2009—2013 年人工智能领域文献的学科(本研究以 WOS 类别作为划分学科类别的依据)共现网络如下图所示：

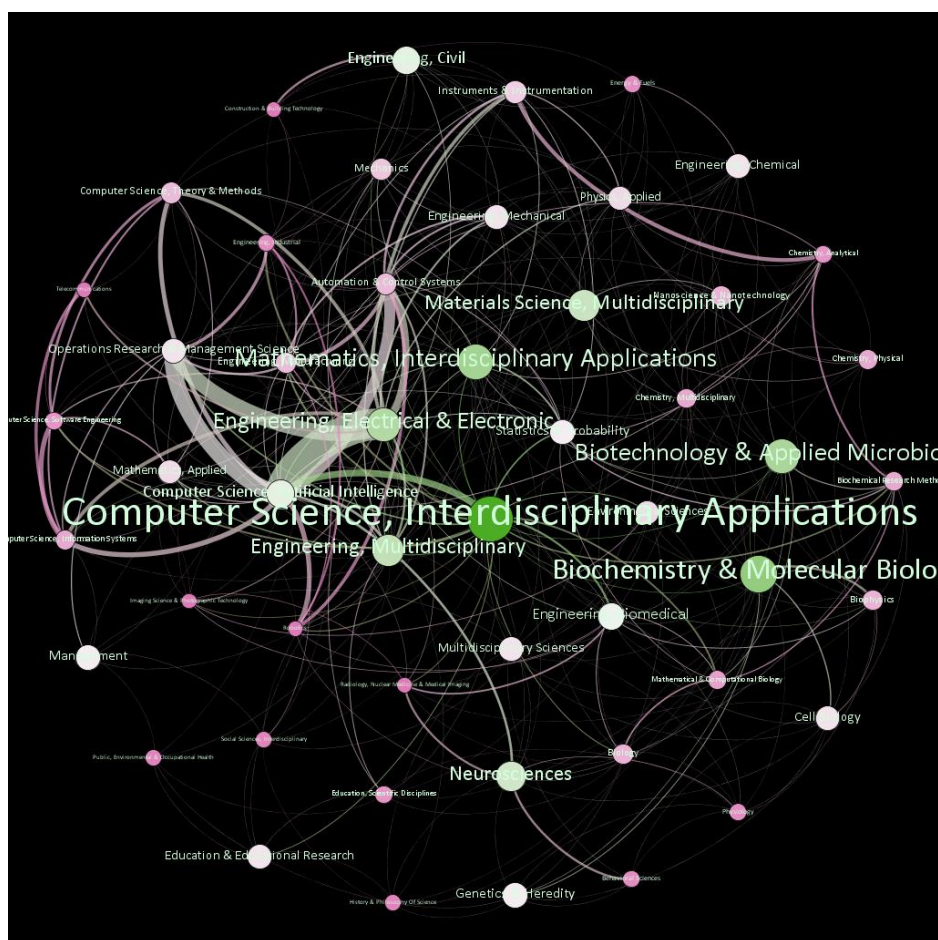


图 25 国外文献学科共现网络

在该网络图中，计算机科学、工程学、生物工程等学科在网络中位于较核心的位置，最核心的学科为“计算机科学：跨学科应用”，但是该节点与其周围节点的大小没有明显区别，说明网络整体分布较为分散。网络主要参数如表所示。

表 22 国外文献学科共现网络基本特征参数

平均度	平均加权重	网络直径	平均路径长度	网络密度	平均聚类系数
8.941	383.518	6	2.77	0.053	0.536

该网络的平均加权重很高，但是网络密度较低，说明存在一些学科和其他学科联系较为密切；网络密度低，但平均聚类系数很高，说明该网络存在一些大的学科群，学科群内部的学科联系较为紧密，但学科群之间联系松散。

4.5.2 学科中心性分析

统计学科共现网络中每个的学科的度数、介数和接近中心度，发现该网络中只有 8 个学科的介数大于 0，其他 162 个学科的介数约为 0，说明有 8 个学科处于该网络中较为核心的位置。同样，接近中心度只有 12 个学科的值显著大于 0。介数排名前 10 的学科如下表所示：

表 23 学科共现网介数最高的前 10 个学科

学科名称	介数
材料科学	0.37
物理化学	0.161
应用物理	0.096

纳米科学及纳米技术	0.063
工具学	0.059
应用化学	0.024
集成材料物理	0.006
光学	0.003

接近中心度排名前 10 的学科如下表所示:

表 24 学科共现网络接近中心度最高的前 10 个学科

学科名称	接近中心度
材料科学	0.812
物理化学	0.765
应用物理	0.765
纳米科学及纳米技术	0.722
工具学	0.650
集成材料物理	0.650
化学	0.619
分析化学	0.591
材料科学	0.591
电化学	0.520

综合介数和接近中心度可知,中心性高的学科是材料化学、物理化学、应用物理,大多为物理、化学、材料学领域的研究,说明这些学科在人工智能领域的研究中具有重要的连接作用,把计算机科学、工程学等学科与材料科学等学科相连。光学以后的学科的介数较低,接近为零,这可能由于关键词共现网络中的节点数目太多而引起的,但在如此庞大的网络中,物理、化学、材料科学等学科仍然具有重要的位置。

4.5.3 学科聚类分析

利用 gephi 的 modularity class 功能,对 2009—2013 年人工智能领域英文文献的学科进行聚类分析,聚类的对象是“WOS 类别”。一共得出 7 个类,如下表所示:

表 25 学科聚类分析

学科领域	子领域	学科类别
化学&生物类	生物	Biochemistry & Molecular Biology, Biodiversity Conversation, Biophysics, Cell Biology, Evolutionary Biology, Genetics& Heredity 等
	化学	Chemistry: Organic, Chemistry: Inorganic& Nuclear 等
计算机&工程类	计算机科学	Artificial Intelligence, Cybernetics, Hardware & Architecture, Information S Systems, Software Engineering,
	工程学	Electrical& Electronics, Telecommunications
	数学	Mathematics
工程&材料类	工程类	Engineering: Chemical, Engineering: Marine, Engineering: Petroleum, Imaging Science& Photographic Technology
	材料	Material Science: Characterization & Testing, Material Science: Textiles Polymer Science 等
工商管理&心理类	管理类	Business, Management, Public Administration
	心理社会类	Psychology, Sociology, Women's Studies, Industrial Relations

化学 & 物理 & 材料类	化学&物理	Chemistry: Analytical, Chemistry: Physical, Physics: Applied, Physics: Condensed Matter, Spectroscopy
	材料科学	Ceramics, Coatings& Films, Nanoscience & Nanotechnonology,
综合类	生物学	Biology, Biochemical Research Methods 等
	工程学	Automation& Control Systems, Engineering: Biomedical, Engineering: Manufacturing, Engineering: Mechanical 等
	物理	Physics: Fluids & Chemical, Physics: Particles & Fields, Physics: Multidisciplinary 等
	医学	Anatomy & Morphology, Nursing
	社会学	Social Issues
	其他	Theater, Music, Art, Astronomy& Astrophysics, History of Social Science 等
医学类	医学	Dentistry, Medicine, Nutrition & Dietetics, Surgery, Pediatrics, Veterinary Science 等
	心理学	Psychology: Development, Psychology: Experimental

170 个学科聚类后形成 7 大类，如上表所示，分别为化学&生物类、计算机&工程类、工程&材料类、工商管理&心理类、化学&物理&材料类、综合类和医学类。聚类后的学科共现网络图如下图所示：

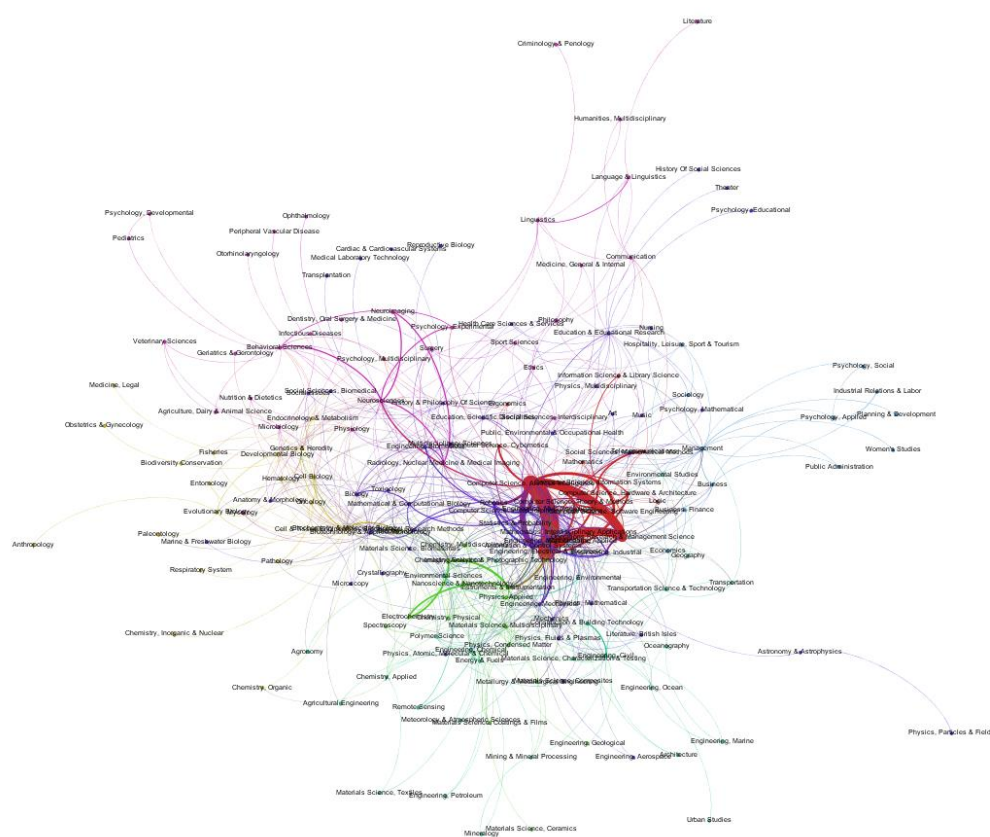


图 26 聚类后的学科共现网络图

第一类是化学&生物类（图中黄色部分），主要包括生物化学及分子生物学、生物多样性、细胞生物、生物物理等学科，这些学科主要是从生物角度研究进行仿生研究。第二类是计算机&工程类（图中红色部分），主要包括人工智能、信息系统、软件工程、数学等学科，这一类主要是从计算机技术的角度研究智能计算、计算机神经网络、机器学习等，是人工智能领域研究的核心部分，如图所示。第三类是工程&材料类（图中深绿色部分），主要包括化学工程、石油工程、海洋工程、图像处理技术、材料科学等，这一类主要研究如何寻找、制造、改进仿生材料，以更好地模拟自然生物。第四类是工商管理&心理类（图中蓝色部分），包括管理、经济、心理、社会学等学科，这一类主要研究人工智能相关研究的社会伦理问题，但这部分的研究并不多，在整个网络中也处于较为边缘的位置（位于网络的右上方）。第五类是化学&物理&材料类（图中蓝绿色部分），包括分析化学、物理化学、应用物理、纳米科学等，与第三类不同的是，这一类更注重对基础理论的研究，包括材料的分子结构以及相关材料技术原理等，研究层面更小，主要在分子层面进行研究，这一类研究分布较广，在网络的下方。第六类是综合类（图中紫色部分），包括生物学、工程学、物理学、医学、社会学、航空航天、戏剧、音乐、一书、社会学研究方法等多种学科，由于人工智能领域本就是一个综合学科，其应用的基础领域遍及各个学科，因此会有这么一批综合的大类，虽然涉及的学科最多，但在网络中分布却很集中。第七类是医学&心理系类（图中淡绿色部分），包括牙医学、要学、手术学、发展心理学和实验心理学等，这一类主要研究人工智能技术的应用，包括医学中临床手术的应用以及心理学治疗，这一类也较为集中，分布在网络中较为核心的位置。

5 总结与讨论

本研究利用社会网络分析技术对近五年的人工智能主题领域文献进行分析。在对文献进行基本描述统计的基础上，对人工智能主题领域文献的作者合著网络、机构合著网络、国家合著网络、学科共现网络以及关键词共现网络的度分布、平均路径长度、网络密度、聚类系数、介数分布、连通集团规模等指标进行分析，梳理了人工智能领域的合作情况与研究现状。

本研究的主要不足在于对作者重名、机构重名、关键词同义等问题的处理不够充分。在作者方面，目前对中文文献没有进行处理，对英文文献根据作者与机构之间的一一对应关系进行了处理，但是没有处理同一作者先后隶属于不同机构等问题。在机构方面，目前对部分机构进行了概化处理，但是对同一机构的不同称谓等问题处理得不够充分。此外，本研究并没有处理关键词同义的问题。更完善的数据清理是本研究今后的改进方向。

6 参考文献

- [1] 韦雄奕, 解晨. 人工智能发展问题探究与思考[J]. 计算机光盘软件与应用, 2013, 8: 091.
- [2] 孙晔, 吴飞扬. 人工智能的研究现状及发展趋势[J]. 价值工程, 2013, 32(28): 5-7.
- [3] 邢王秀. 人工智能发展现状与未来发展[J]. 中国-东盟博览, 2013, 02: 62.
- [4] 万雷. 论人工智能的现状与发展方向[J]. 科技风, 2012 (15): 54-54.
- [5] 张洋, 刘锦源. 基于 SNA 的我国竞争情报领域论文合著网络研究[J]. 图书情报知识, 2012, 2: 014.
- [6] 张继洋, 李宁. 科学合著网络研究进展分析[J]. 情报理论与实践, 2012, 35(004): 124-128.
- [7] 孟微, 庞景安. 我国情报学科研究合著网络研究及其特征参数分析[J]. 情报理论与实践, 2009, 8: 12-15.
- [8] 朱庆华, 范哲, 施文蔚. 信息系统领域中外合著网络研究[J]. 情报理论与实践, 2011, 11: 6-10.

- [9] 李凌云, 王海军, 王佳. 虚拟实验研究领域作者合著的社会网络分析[J]. 实验室研究与探索, 2011, 30(11): 56-60.
- [10] 左美云, 温晓薇, 华晓清. 知识管理领域学者间合著网络分析[J]. 信息资源管理学报, 2012, 2(4): 4-15.
- [11] Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks[C]//ICWSM. 2009.
- [12] 魏晓俊, 谭宗颖. 基于核心-边缘结构的国际科技合作网络分析——以纳米科技(1996-2004) 为例[J]. 图书情报工作, 2006, 50(12).
- [13] 孟微, 庞景安. 我国情报学科研合著网络研究及其特征参数分析[J]. 情报理论与实践, 2009, 8: 12-15.

7 小组分工

小组成员	负责部分
周妍	英文文献发文量、合作率、合作水平统计, 作者合著网络分析, 撰写报告大纲、报告的第 3 和第 5 部分以及 4.1 节
徐山川	中文题录数据抓取, 中英文题录数据清理, 机构合著网络分析, 撰写报告的第 2 部分以及 4.2 节
沈奕娜	中文文献发文量、合作率、合作水平统计, 国家合著网络分析, 撰写报告的第 1 部分以及 4.3 节
宋小双	英文题录数据抓取, 英文文献主要发文国家、机构和学科分布统计, 关键词和学科共现网络分析, 撰写报告的 4.4 节和 4.5 节