****

**本科生毕业设计（论文）外文翻译**

**外文原文题目：Feature analysis of multidisciplinary scientific collaboration patterns based on PNAS**

**中文译文题目：基于PNAS的多学科科学合作模式的特色分析**

**毕业设计（论文）题目：美国科学研究系统建模及合作模式挖掘**

**学 院 管理与经济学部**

**专 业 信息管理与信息系统**

**年 级 2019级**

**姓 名 蒋世华**

**学 号 3019209018**

**指导教师 王文俊 教授**

**摘 要**

合作模式的特色通常被看作不同学科之间的差异。同时，在孵化一些交叉学科的现代科学研究也出现了不同学科合作的明显特色。基于1999年～2013年的多学科杂志PNAS发表的52803篇论文数据分析了在生物、物理和社会科学间的合作特色。从这些数据中，我们发现合作模式的传递性和同配性与那些作者合作者和作者论文的相同分布一样，都满足广义泊松分布和幂律分布的混合。而且，我们发现有相当一部分作者从事交叉学科研究，而不仅仅是具有许多合作者或具有很多论文的作者。这个事实提供了一个了解多学科和交叉学科合作模式方面的一个途径。

**关键词：**合作模式；交叉学科；超图；复杂网络

1. **介绍**

自然和社会科学提供了很多方法分别去学习、预测和解释自然现象和社会(人类行为和精神状况)[[1]](#endnote-0)，这些专业化的科学知识形成了各种学科。同时，为了解决那些解决方案超出单个学科范围的问题，研究人员需要整合多个学科的数据、技术、概念和理论[[2]](#endnote-1)[[3]](#endnote-2)[[4]](#endnote-3)[[5]](#endnote-4)。学科之间的相互作用孕育了很多的跨学科，模糊了自然科学和社会科学的边界，并产生了许多重要的科学突破[[6]](#endnote-5)[[7]](#endnote-6)[[8]](#endnote-7)。

研究跨学科或学科间的合作模式有助于理解合作行为和知识融合方式的多样性。因为自然科学和社会科学主要依靠论文，多学科期刊的论文为这项研究提供了信息和可靠的平台[[9]](#endnote-8)[[10]](#endnote-9)[[11]](#endnote-10)[[12]](#endnote-11)。这里我们调查了1999-2013年发表在《美国国家科学院院刊》（PNAS）上的52803篇论文。数据集的内容涵盖三个科学类别：社会科学和自然科学中的两个主要子科学，即生物科学和物理科学。

合作关系可以被表示为合作网络。因此可以从网络的角度来研究合作模式。来自不同科学领域的合作网络出现了特定的相似性，例如合作者的部分传递性、合作者数量的同质性、每个作者的合作者分布的右偏分布[[13]](#endnote-12)[[14]](#endnote-13)[[15]](#endnote-14)[[16]](#endnote-15)[[17]](#endnote-16)[[18]](#endnote-17)[[19]](#endnote-18)。这些共性也出现在三个作者的合作网络中（分别来自PNAS的三个科学类别）。我们深入了解这些共性的规律和原因。我们发现每个作者的合作者分布和每个作者的论文分布遵循相同的分布类型：广义泊松分布和幂律的混合。我们通过作者吸引合作的能力的多样性为分布类型和这些共性提供了可能的解释。

之前的一系列工作讨论了科学[[20]](#endnote-19)[[21]](#endnote-20)[[22]](#endnote-21)、学科[[23]](#endnote-22)[[24]](#endnote-23)[[25]](#endnote-24)[[26]](#endnote-25)、大学[[27]](#endnote-26)、期刊[[28]](#endnote-27)[[29]](#endnote-28)和研究团队[[30]](#endnote-29)的跨学科定量指标。一些工作探讨了跨学科性和科学影响之间的相关性[[31]](#endnote-30)[[32]](#endnote-31)[[33]](#endnote-32)[[34]](#endnote-33)（例如，引文捕获能力[[35]](#endnote-34)[[36]](#endnote-35)[[37]](#endnote-36)）。基于这些参考文献的特定总体思路，我们通过交叉学科的论文共现，以及基于复现的一些指标计算如Rao-Sterling多样性[[38]](#endnote-37)，和介数中心性[[39]](#endnote-38)，研究PNAS的交叉学科活动。

我们进一步通过多学科研究了跨学科的合作模式，并发现相当一部分物理和社会科学的作者和论文参与了跨学科研究。从巨大组成部分的数据中提取的多学科合作者关系网络分别包含了超过88%的生物、超过80%的物理和超过71%的社会科学的作者。相当多的作者对巨大成分的形成做出了贡献。作者活动和生产力对巨大组成部分的贡献随着时间的推移而增加。案例研究所显示的高度跨学科性可能不能代表一般的合作模式，因为作者可以向多学科期刊提交比特定领域期刊更多的跨学科成果。

本论文的结构如下：在第2节中描述了数据处理过程；在第3节分析了相似性和相互作用；并在第4节中得出结论。

1. **数据集**

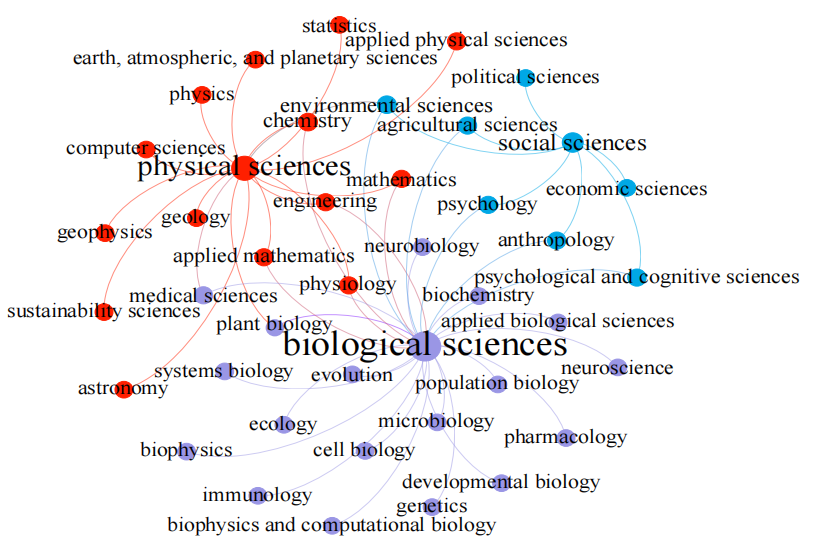
**2.1 使用该数据集的原因**

案例研究涉及两个概念，即多学科性（来自不同学科的研究人员在其学科内进行研究）和跨学科性（超越学科界限的研究）[[40]](#endnote-39)。多学科可以被看作是多学科交叉地出现在其中的学科的结合。一本涵盖自然科学和社会科学的多学科期刊可以用来分析科学类别之间的相互作用。这本期刊也可以用来比较多学科的合作模式并找到相似之处。PNAS发表了高质量的研究论文，并且提供这些论文的可靠的学科信息。该期刊还为分析全球合作模式提供了一个高质量的数据平台，因为其近一半的论文来自美国以外的作者。

包括SCIENCE、Nature和Nature Communications不提供论文的学科信息。《英国皇家学会学报》专注于物理科学和生命科学之间的交叉学科研究，但不涉及社会科学。我们的分析仅限于PNAS，这给我们的发现带来了局限性。例如，社会科学媒体不仅依赖论文，还依赖书籍xixii。因此，所获得的结果必须仔细解释为在所选期刊上发表论文的研究人员的模式。然而，由于PNAS的影响力和代表性，案例研究可能有助于理解多学科和跨学科合作模式的各个方面。

**2.2 学科信息**

数据集的大多数论文被分为三个一级学科（生物、物理和社会科学）和39个二级学科（表1）。跨学科论文被分为几个学科。数据包含43304篇生物学论文（包括3957篇生物物理学论文），占总数的82.01%。该数据还包含5987篇物理论文和1310篇社会论文。二级学科以上的跨学科论文2961篇，占总数的5.61%，交叉学科的重要差异并不意味更特别受到PNAS的偏好。



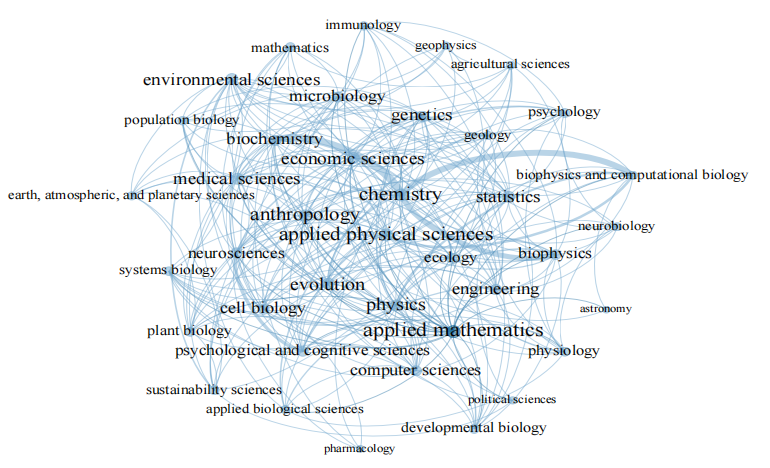
**图1 一级学科和二级学科的关系**这个网络基于PNAS 1999～2013年的多学科论文的信息构建，如果两个学科是同一篇论文的一级学科和二级学科它们就会相连。节点大小代表了节点的度。

事实上，参与自然科学（尤其是生物科学）的研究人员数量远远超过参与社会科学的研究人员[[41]](#endnote-40)。仅有1842篇论文被归类为一级学科。对于这些论文，它们的二级学科被认为是缺失的，但在我们以前的工作中，它们被认为与一级学科相同[[42]](#endnote-41)。因此表1中的数据与参考文献xlii中的数据不同。

基于论文的学科信息，我们构建了一个网络来表达一级学科和二级学科之间的关系（图1），其中如果两个学科是同一篇论文的一级学科和二级学科，它们之间相互连接。我们还可以构建一个网络来表达二级学科之间的相互作用（图2），其中每个节点都是一个学科，如果有一篇论文同时属于它们，则两个节点是连接的。这些网络可能会随着新发表论文的形成而发展。因此，使用最新数据，人们可能会有一个更全面的观点。

**2.3合作者网络**

识别真实的作者，即消除作者姓名的歧义，是一个重要、耗时但必要的合作网络分析。



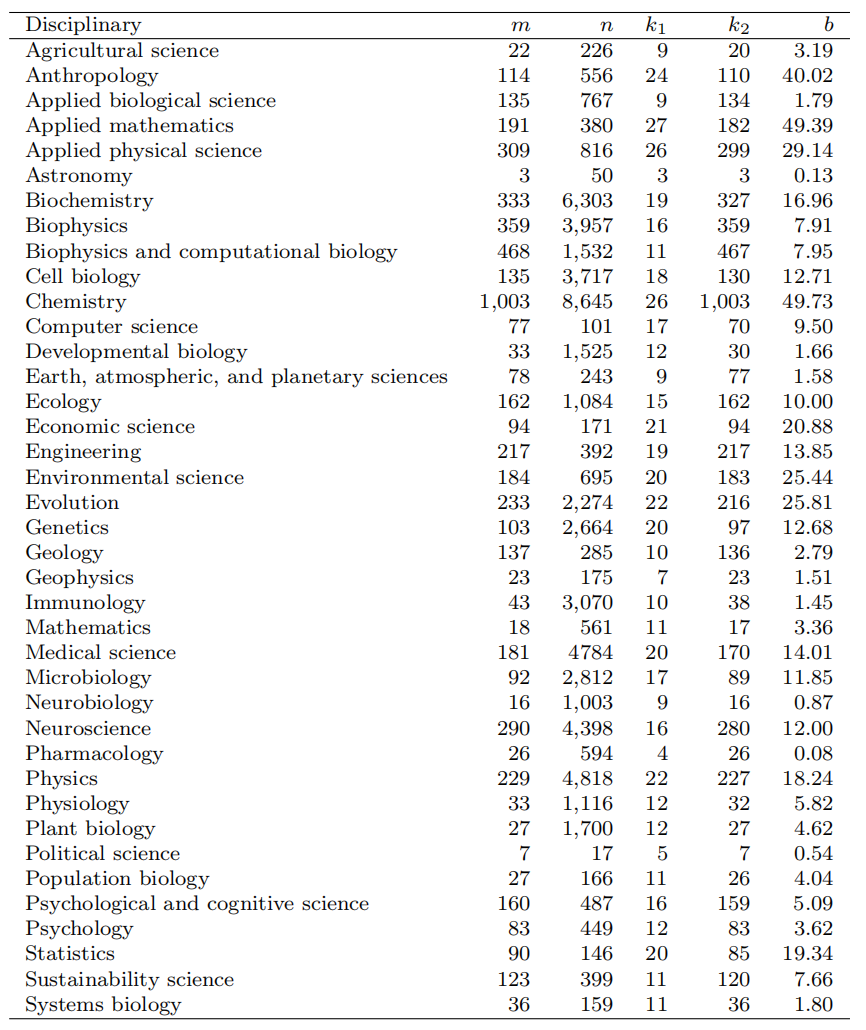
**图2 二级学科间的交互作用** 带有权重的网络基于PNAS 1999～2013年交叉学科的论文信息构建，边的宽度表示了边的权重：在两个连接的学科间的交叉学科论文数量

有几种方法使用论文中提供的名称信息（例如，基于初始值的方法[[43]](#endnote-42)）。基于初始值的方法的主要错误识别是由两个或多个不同的作者合并为一个引起的。因此，它减少了唯一作者的数量，并扩大了巨大组成成分的大小。获得更多的信息（如电子邮件地址）有助于减少合并错误，但也带来了收集信息的困难。

在PNAS 1999-2013中，93.1%的作者提供了全部的姓氏。因此，论文上提供的名字被直接用来识别作者。然而，使用姓和第一个给定名称的首字母将产生许多名称消歧的合并错误[[44]](#endnote-43)。这些作者在数据中的比例为2.9%，这部分作者以发表一篇以上论文为条件的比例为0.3%。同时，如果一些作者提供了完全相同的名字，即使使用全名也会产生合并错误。中文名字被发现是名字重复的原因xliv。我们计算了在100个主要中文姓氏中，名字小于6个字符和姓氏的比例。这些作者在数据中的比例为2.7%，而在这些作者中以发表一篇以上论文为条件的比例为1.1%。这四个比例的值表明，名字重复的影响是有限的。表2中列出了特定子数据的这些比例。

如果作者没有全程提供自己的名字，这里采用的方法将把一位作者分成两位或两位以上。这样会分割巨大组成部分的大小，这些指数被用作跨学科研究普遍性的证据。因此3.5、3.6小节的结论是有争议的。

**表1 在PNAS 1999～2013年二级学科的特定标签**



表中n为论文的数量，m为交叉学科的数量，k1为度，k2为度的权重，b为在图2的带权重的网络中计算的介数中心性

此外，所采用的方法造成的不准确并没有改变每个作者合作者的基本事实分布类型和每个作者的论文分布类型xliv。

**表2 分析网络的特定统计标签**



标签a和b分别表示了提供首字母和姓氏的比例、这些作者中以发表一篇以上论文为条件的比例。标签c和d分别表示了作者姓氏在100个主要的中国姓氏并且小于6个字符的比例、这些作者中以发表一篇以上论文为条件的比例。

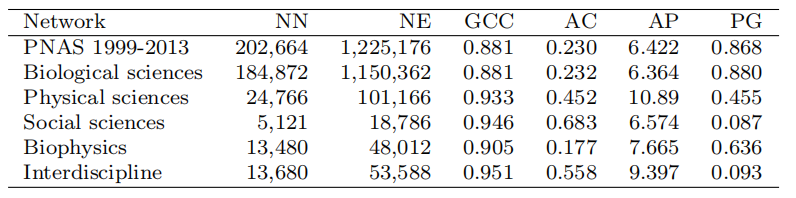
**3 数据分析**

**3.1 网络性质**

合作是一种多元关系，n∈Z+，因此它可以用超图来表示，超图是图的推广，其中一条边（称为超边）可以连接任意数量的节点。将作者表示为节点，将每篇论文的作者组（论文团队）表示为超边。然后，我们可以从超图中提取一个合作者关系网络作为一个简单图，其中在每个超边的每两个节点之间形成边，并且将多个边作为一个边处理。节点的“度”和“超度”分别用来表示合作者的数量和作者的论文数量。

数据显示，生物科学（6.624）和物理科学（5.254）的平均论文团队规模大于社会科学的平均团队规模（4.634）。这种规模关系符合研究团队规模为通常在自然科学中较大，在社会科学中较小xli。现在让我们考虑特定学科或科学类别中被考虑的论文的合作网络。所有这些网络都是高度集群的、具有度相似性的，其平均最短路径长度以节点数的对数表示（表3中的logNN≈AP）。这些性质并不意味着所有的网络都是小世界。社会科学网络是个例外，它甚至没有超过10%的作者。然而，这并不意味着社会科学的研究是孤立进行的。事实上，71.5%的社会科学作者属于由整个数据生成的作者网络的巨大组成部分。因此，分析作者在单个学科中的合作是有局限性的。因此，我们在所有学科的环境中进行分析。

**表3 分析网络的特定统计标签**



标签分别为节点的数量（NN），边的数量（NE），全局聚类系数（GCC），度相似性（AC），平均最短路径（AP），巨大组成部分的节点比例（PG）。前两个网络的平均最短路径大约计算了400,000对节点的样本。

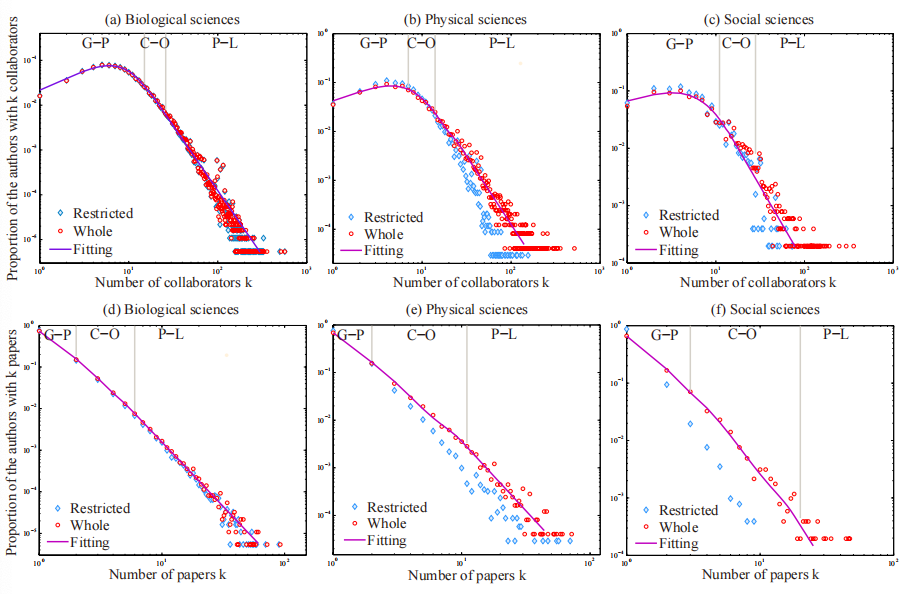
**3.2 度和超度**

在数据上统计度和超度（不限于单个科学类别），并观察三个作者集（分别来自三个科学类别）的度分布和超度分布。我们发现尽管合作级别因科学类别而异，但所有分布都出现了钩头、肥尾和它们之间的交叉，这可以被视为合作网络的一个共同特征（图3）。头部和尾部可以分别通过对数正态分布和幂律分布来拟合[[45]](#endnote-44)。

这些分布也可以作为一个整体，通过广义泊松分布和幂定律分布的混合来拟合。拟合参数如表4所示。我们进行了两个样本的Kolmogorov-Smirnov（KS）检验，以比较两个数据向量的分布：节点索引（即度，超度），从相应的拟合分布中提取的样本。零假设是两个数据向量来自相同的分布。每个拟合的p值表明，在5%的显著性水平上，检验不能拒绝零假设。注意到由于大量作者数量较少，χ2拟合度检验在这里不适用。

把作者当作样本，混合分布意味着这些样本来自不同的群体，即合作者和论文较少的作者的合作模式与合作者和论文较多的作者的合作模式不同。在参考文献[[46]](#endnote-45)中，对出现的混合型经验度分布给出了一种可能的解释（不含学科）。对于相同的一般思想，可以对超度分布采用类似的解释，如下所示。

研究人员是否合作发表论文可以被视为“是/否”的决定。因此，研究者的超度等同于希望与该研究人员合作的候选人在一系列决定中成功的次数。假设这些候选人的人数是n。假设每个候选人的合作概率为p。那么超度数将遵循二项式分布B(n，p)，并且泊松分布将满足np的期望值。由于作者吸引合作者的能力的多样性，不同作者的期望值np各不相同。



**图3 每个作者的合作者/论文分布** 页面展示了PNAS 1999～2013年的分布（红色部分），和每个学科分类的论文（蓝色部分）。拟合分布是广义泊松分布和幂律分布的混合，拟合参数在表4“G-P”、“C-O”、“P-L”分布表示了广义泊松分布、混合和幂律分布。

作者的决定可能具有依赖性。例如，与有出版经验的研究人员合作有助于发表论文。因此我们可以将超度视为遵循广义泊松分布的随机变量（允许事件的发生概率具有记忆性[[47]](#endnote-46)）。在经验数据中，大多数超度都围绕着它们的模式。因此，我们可以认为它们遵循一些广义泊松分布，在它们的模式周围有一个期望值，因此形成了超度分布的广义泊松部分。一些作者经历了一个复杂的论文过程，这使得超度分布向右倾斜并形成一个肥尾。

**3.3 合作网络的传递性**

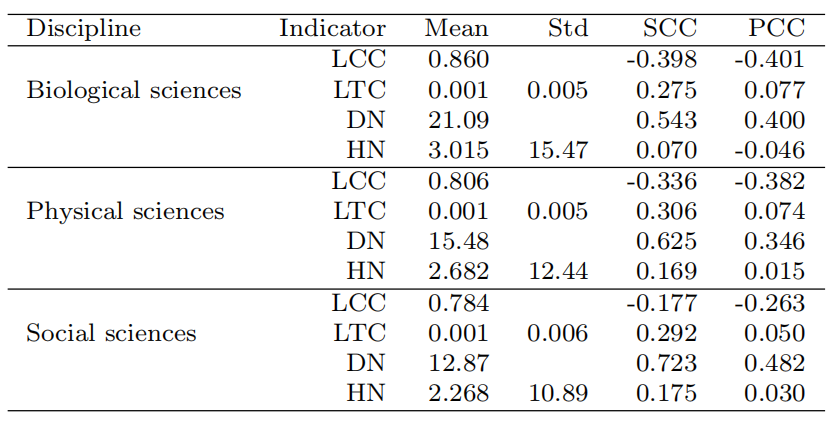
社会中的传递性是“我朋友的朋友也是我的朋友”，这是社会关系网络的典型特征。在学术界中，作者的合作者很可能彼此熟悉，因此也是合作者。例如，组织和机构背景可能会导致合作者关系的传递性，从而有助于出现合作者网络的集群结构。

网络的传递性可以用图论中的两个指标来量化，即全局聚类系数（在“三角形”中连成三元组的部分）和局部聚类系数（节点的两个邻居连接的可能性）。高传递性是合作网络的一个共同特征xv。

但是传递性在多大程度上是由于作者在学术社会中的活动造成的呢？这种活动可以通过合作者的数量，即度来部分地反映。因此，可以通过度和局部聚类系数之间的相关系数来描绘传递性。注意到相关系数表示两个变量或它们的秩之间的线性关系的程度。除非对于每个期望条件的Y给出X来表示E(Y|X)作为X中的线性或近似线性函数，不然变量X和Y的系数通常不能完全描述相关性。对于给定的度的局部聚类系数的条件期望值是用CC（k）表示度为k的节点的平均局部聚类系数。图4所示的CC（k）的近似线性趋势保证了表5中相关性分析的有效性。减少的趋势不能从度的信息中得出。一个节点的局部聚类系数的分母随着其度的二次增长，但分子不能根据度的信息计算得到。

CC（k）均值活动的下降趋势是否会抑制传递性？肯定的答案是违背常识的。在PNAS 1999-2013中，74.62%的作者在数据中只发表了一篇论文，99.9%的论文团队规模遵循广义泊松部分，即每个团队规模的平均论文约为6.028。广义泊松部分的边界由参考文献xlvi中概率密度函数的边界点检测算法检测（见附录）。因此大部分小度的作者的局部聚类系数接近于1（图4）。一些作者经历了一段时间的合作，他们的度是通过计算论文得到。对于这些作者来说，他们在不同论文中的合作者无法合作，这降低了他们的局部聚类系数。因此令人困惑的事情和常识并不矛盾，而是由于通过计算静态网络中的“三角形”来测量传递性这一动态性质的不足。

**表5 度与传递性/聚类系数指标的相关性**



图中指标为局部聚类系数（LCC），合作者的局部传递性（LTC），节点邻居的平均度（DN），节点邻居的平均超度（HN）。我们计算了这些作者的平均指标，这些指标和度的斯皮尔曼等级相关系数（SCC）和皮尔森积矩相关系数（PCC）。对于两个较小的皮尔森积矩相关系数（PCC），我们计算了它们的标准差。

为了设计一个更合理的测量传递性的指标，从最开始对合作的最初定义来考虑：一次研究合作中的两个合作者（他们还没有合作过）在未来合作的概率。这个概率可以通过合作的动态超图的时间信息来计算。基于作者的平均概率测量全局传递性，其值在每个科学类别中都很低（表5）。注意到PNAS 1999-2013中的计算是有限的，在其他期刊或其他时间段可能会发生变化。因此，这里传递性的值可能会被低估。度为k的作者的传递性概率（图4中的TC（k））的增加趋势意味着合作有助于传递性。这符合常识：一个拥有许多合作者的研究人员很可能会介绍他的合作者进行合作。

**3.4 合作关系的同质性**

合作是基于特定研究人员的共同特征，包括兴趣和地理位置。同质性现象出现在许多社交关系中，在网络科学中被称为同质混合xviii。每种学科的作者更倾向于与社会活动和生产力相似的合作者合作吗？作者的社会活动和生产力可以分别用度和超度这两个指标来量化，然后通过作者的指标和作者邻居的平均指标这两个变量之间的相关系数来描述指标的偏好。正相关意味着同质，负相关意味着不同质，零则无偏好。

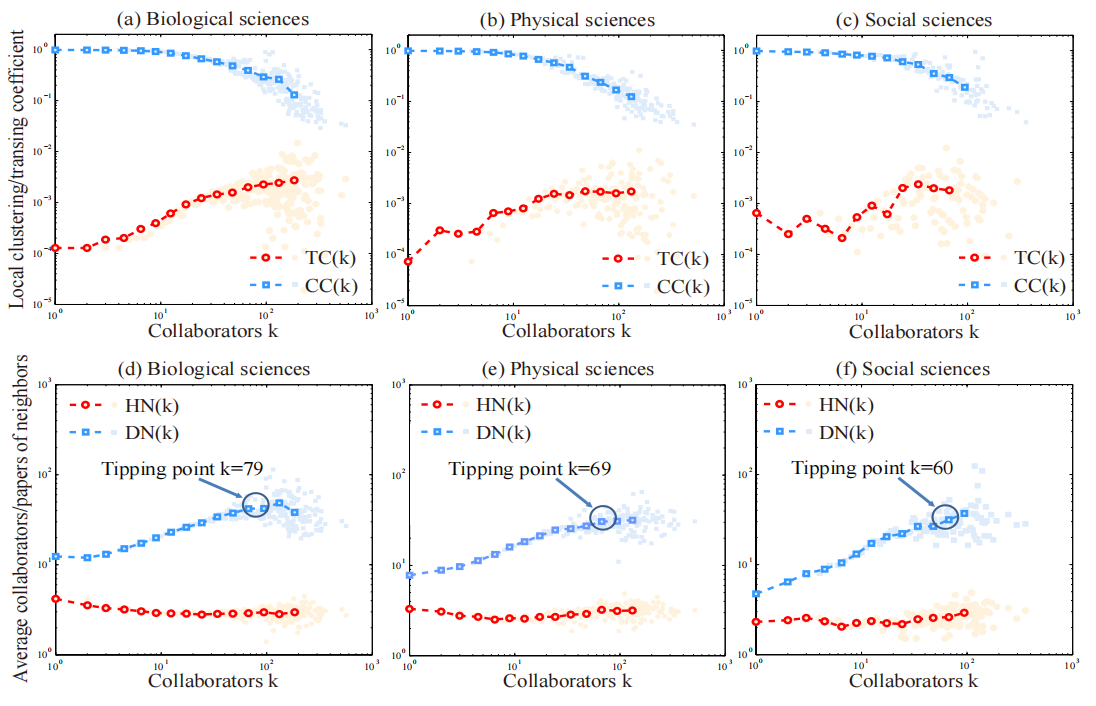
度的同质性是合作网络的一个特征xviii。这是否意味着善于交际的研究人员（与许多合作者）会更倾向于与其他善于交际的研究员合作，不善于社交的与不善于社交的合作呢？在之前的一项研究[[48]](#endnote-47)中，我们发现排名前5.99%的大多数善于社交的作者（根据度来衡量）与另一位此类作者合作的比例为99.5%。这一比例甚至可能被低估，因为这些作者可能在1999年之前或其他情况下合作。注意到名称消歧方法的拆分和合并错误在一定程度上影响了比例。即便如此，这一比例仍然很高。

然而，如果善于社交的研究人员只与善于社交的研究人员合作，那么就会有许多善于社交的研究人员，这违背了经验度分布。现在让我们分析作者的社会活动对度的同质性的影响。对于具有k度的作者，用DN（k）表示其邻居的平均度。每个经验数据集的DN（k）都存在趋势的变化：头部有明显的增加趋势，而尾部则没有（图4）。这意味着度同质性主要由小度的作者控制。

DN（k）的趋势的临界点由参考文献xlvi（附录中列出）中常用函数的边界点检测算法检测得出。该算法的输入是DN（k）、g(·) = log(·) 和h(x) =a1x3+a2x2+a3x+a4(x,ai∈R,i= 1,...,4)。使用这些输入是基于对DN（k）的观察。大多数作者的度都在他们的众数5左右，只有少数作者拥有大度。因此，作者的邻居很可能是小度的作者。因此，对于小度作者来说，这些作者与其邻居之间的度差距很小，而对于大度作者来说则很大，这导致了DN（k）的趋势变化。

在每个科学类别中，超度和邻居的平均超度之间的相关系数约为零（表5）。对于具有k超度的作者，用HN（k）表示其邻居的平均超度。这意味着选择合作者不受生产力因素的影响。事实上，研究团队的成员可能有各种各样的科学年龄（新加入者、现任者），因此有不同的超度。由于合作主要发生在一个研究团队中，一个作者的合作者可能具有不同的超度，这导致了HN（k）的稳定趋势。

基于HN（k）的平均值大于2，以及74.62%的作者在数据中仅有一篇论文，我们可以得出，在数据中很大一部分作者与至少一位在PNAS 1999-2003发表过第一篇论文的作者合作。这些作者在生物学、物理学、社会科学上的比例分别为79.22%、71.17%和65.12%。这一比例可能被高估了，因为其中一些作者可能在1999年之前在PNAS上发表论文。

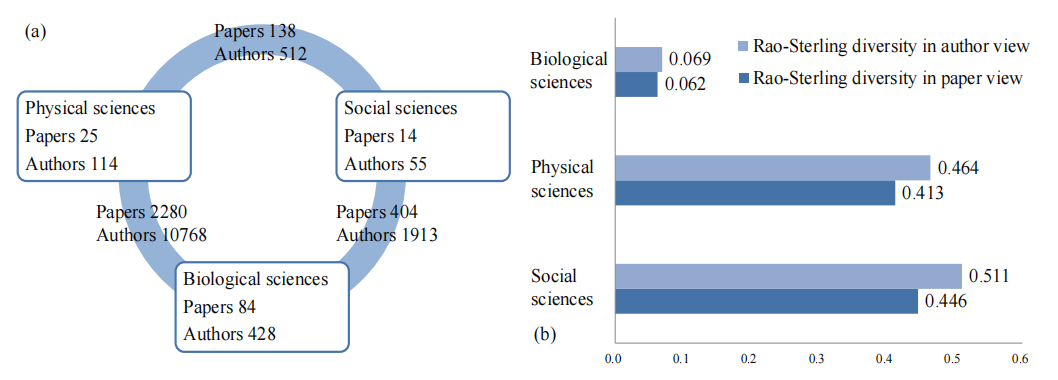


**图4 给定特定指标的条件期望值** 从度为1到最大值，我们计算了在k度的节点上的平均局部聚类系数（CC（k））、合作者的局部传递性（TC（k））、节点邻居的平均度（DN（k））、节点邻居的平均超度（HN（k））。数据按横坐标归类来提取藏在噪声中的趋势。

**3.5 在学科层面的跨学科性**

同类别比例衡量跨学科研究的活动。发表跨学科论文的社会、物理和生物科学作者分别占49.2%、46.0%和7.3%。常识表明，社会科学家单独从事科学研究。社会科学的比例表明常识在PNAS中并不成立。参考文献[[49]](#endnote-48)还表明，近几十年来，社会科学的跨学科性有所增加。

上述分析过程可以应用于二级学科，以获得更准确的结果。然而，有些学科只有少量的论文，例如只有17篇政治学论文。因此，对这些学科的分析失去了统计学意义。因此，我们从另一个角度分析了二级学科之间的相互作用，将它们可视化为图2中的网络。网络是连接的，也就是说没有任何学科是孤立的。该网络的度和介数中心性前三的节点是应用数学、化学和人类学（表1）。它意味着这些学科的理论、方法和问题直接或间接地使用或研究在许多学科中。对于每个一类学科，我们将其二类学科收缩为一个节点，并计算收缩节点的介数中心性。它们的介数中心性（生物科学47.51，物理科学163.81，社会科学161.72）支持上述的分析。

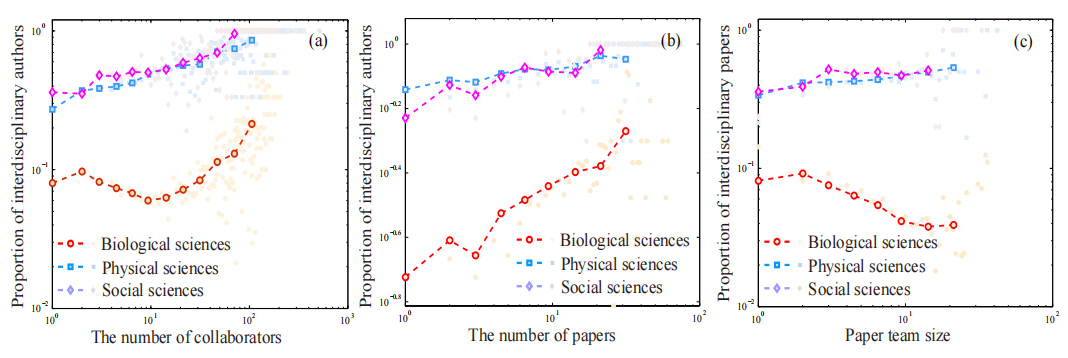


**图5 PNAS 1999～2013年在生物、物理、社会科学的交叉学科研究** 面板a中展示了这三个科学学科的论文和这些作者参与动态交互的部分，并且这些参与交互的部分不是单学科。面板b中展示了每个科学学科的作者/论文视角的Rao-Sterling多样性，其测量了交叉学科研究的学科多样性。

同类别比例仅描述跨学科活动。现在，让我们通过Rao-Sterling指数xxxviii∆=∑i,j(i≠j)dijα(pipj)β来衡量每个科学类别中跨学科研究的学科多样性，其中，pi和pj是论文/作者在i和j科学类别中的比例，dij是在i和j类别的不同贡献水平。学科信息用于将作者分为科学类别：如果他的一篇论文属于某个学科，则作者可以被分为该学科，从而被分为相应的科学。注意到如果一位作者的论文属于多个学科，那么他可以被分为几个科学类别。这里，我们让所有i、j满足α=β=dij= 1 ，因此计算的Rao-Sterling指标衡量了科学类别水平上跨学科研究的平衡加权变化。作者观点和论文观点的指标表明，社会科学和物理科学跨学科研究的学科多样性远高于生物科学（图5）。

**3.6 在作者层面的跨学科性**

我们分析了作者度/超度与进行跨学科研究的可能性之间的关系，以及论文团队规模与成为跨学科论文的概率之间的关系。图6显示，在每个科学类别中，跨学科研究不仅仅是由拥有大度或拥有大超度的作者进行。

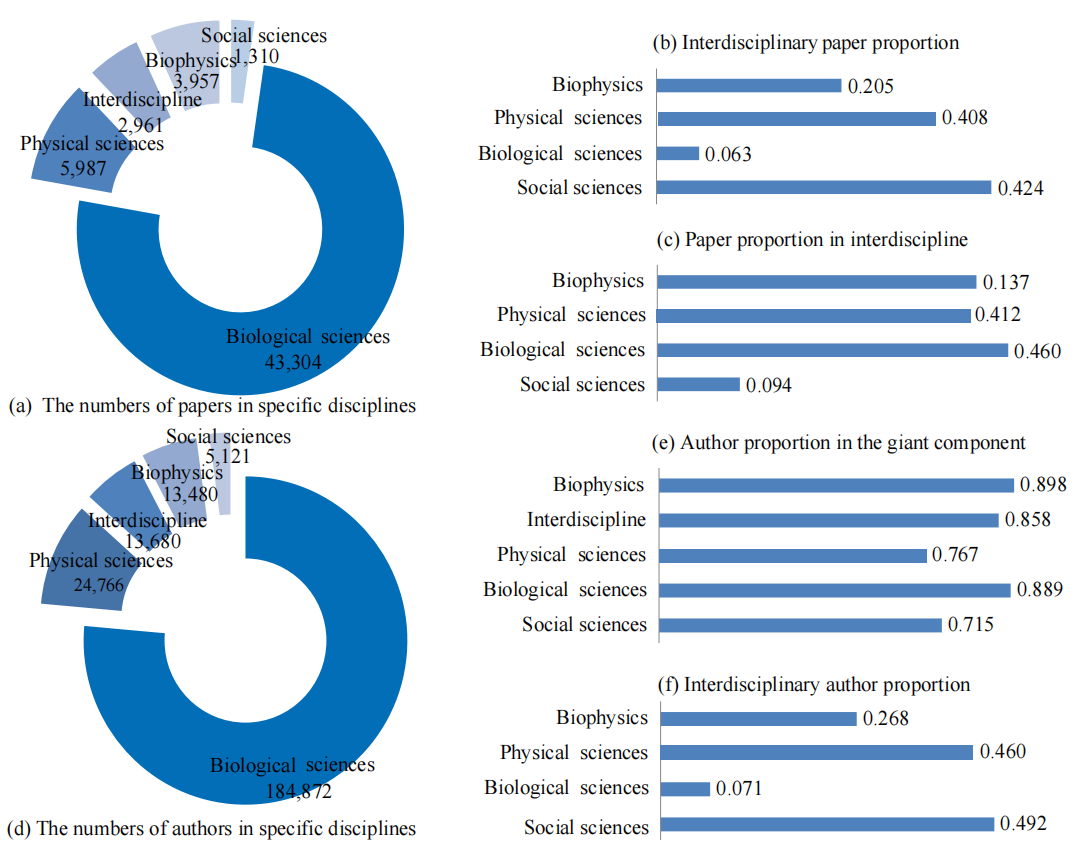


**图6 在作者和论文的特定指标以及他们的交叉学科性之间的关系** 面板a、b展示了作者的度/超度和进行交叉学科研究的可能性的关系。面板c展示了论文团队规模和成为交叉学科论文的可能性的关系。

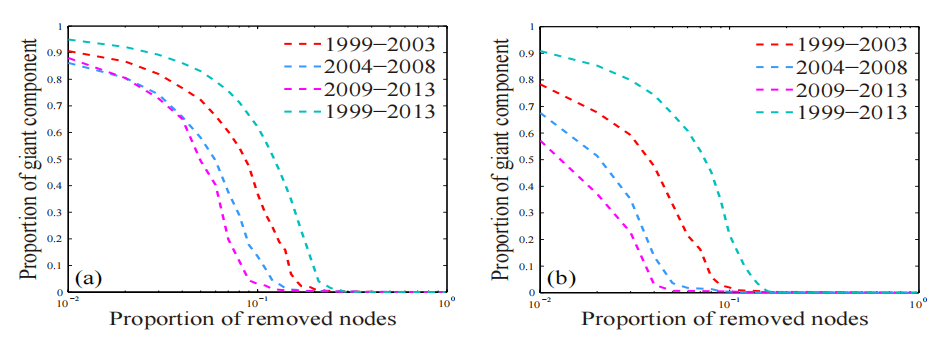
图6还表明，大度或超度作者很可能参与跨学科研究，团队规模大的论文很可能是跨学科的。这些现象似乎被预期是随机的。取几类的一组元素（合作者、论文），随机选择一个子集。那么，一个较大的子集更有可能包含来自多个类的元素。这种推理虽然合理，但并不正确，因为研究人员并没有随机选择主题和合作者。研究成本（时间和精力的投资）使研究人员倾向于在自己熟悉的领域内工作。此外，这个推理基于合作伙伴的选择范围仅限于经验数据，这在现实中是不成立的。

我们分析了合作网络PNAS 1999-2013的庞大组成部分，该网络包含超过86.8%的作者。社会、物理和生物科学的作者分别占71.5%、76.7%和88.9%（图7e）。注意到由初始基本方法引起的作者错误识别增加了巨大组成部分的大小xliv。因此我们通过论文中提供的作者姓名来识别作者（这可能会将一位作者分成两位），以获得保守的结果。

跨学科研究和多学科研究为包含每个科学类别的大多数作者的巨大组成部分做出了贡献。我们分析了巨大组成部分的作者比例与作者活动/生产力之间的关系。将作者从高度和超度分别移除到低度，并计算出巨大成分的比例。从删除作者的比例与巨大成分的比例之间的关系曲线可以发现，巨大成分的形成是由相当多的作者贡献的，例如按度排名的前10%的作者（图8）。考虑三个时期的关系，即1999-2003年、2004-2008年和2009-2013年。随着时间的推移，关系曲线向左移动，这意味着作者的活动和生产力在这个巨大组成部分的形成中发挥着越来越重要的作用。



**图7 特定学科的交叉程度** 对于每一个考虑到的学科i，我们分别用集合Ai、AiI、Pi、和PiI表示了它的作者、它参与交叉学科研究的作者、它的论文、它的交叉学科论文。用S表示PNAS 1999～2013合作网络的巨大组成部分。面板a中指标为｜Pi｜，面板b中为｜Pi∩Ii｜/｜Ii｜，面板c中为｜Pi∩PiI｜/｜PiI｜，面板d中为｜Ai｜，面板e中为｜Ai∩S｜/｜Ai｜，面板f中为｜Ai∩AiI｜/｜Ai｜。



**图8 巨大组成部分和度/超度的关系** 分别将节点从高度/超度移到低度/超度，对于不同的度和超度来说，移除节点的部分和巨大组成部分的关系曲线展示了大部分作者为巨大组成部分作出了贡献，在三个时间周期（1999-2003、2004-2008、2009-2013）的关系曲线左移趋向展示了作者活动和生产力对巨大组成部分的构成的贡献在上升。

1. **结论和展望**

我们对PNAS 1999-2013的案例研究验证了生物、物理和社会科学中合作模式的传递性和同质性。数据表明，这三个科学类别的度分布类型是相同的，它们是广义泊松分布和幂律的混合。这也适用于超度。我们通过作者的“是/否”决定以及他们吸引合作的不同能力，为这种分布类型的出现提供了解释。

数据显示，相当多的作者从事跨学科研究，合作网络PNAS 1999-2013的巨大组成部分包含了每个科学类别的大多数作者。我们采用网络视角分析了二级学科之间的相互作用，并通过度和介数中心性等网络指标量化了它们的跨学科性。我们发现特定的二级学科（如应用数学和人类学）在跨学科研究中发挥着重要作用。

由于PNAS的重要性及其论文的准确学科信息，该案例研究有助于理解多学科和跨学科的合作模式。数据的选择可能会影响我们关于跨学科研究结果的细节。我们的研究结果可能不能被解释为一般研究人员的模式。例如，我们无法期望通过分析特定领域的期刊来观察到高度的跨学科性。我们在完成案例研究时提出了一个问题：跨学科研究的基础依据是什么？虽然对这个问题的探究的讨论超出了本文的范围，但下面提供了一个简单的讨论。

在科学的发展过程中，学科有分裂的趋势：分为子学科和特定主题。尽管研究对象不同，但它们的研究范式是共同的，可以分为四类，即理论研究、实验、模拟和数据驱动[[50]](#endnote-49)。同时，许多科学问题过于复杂，无法通过单一学科的方法论来理解。将来自不同学科的理论和方法论观点整合在一起，为研究问题创造了统一的方法论，甚至为特定学科中的概念提供了工具[[51]](#endnote-50)，这推动了跨学科学科的形成[[52]](#endnote-51)。

系统科学作为一门典型的跨学科学科，研究系统从简单到复杂，从自然到社会科学。系统的各部分以及各部分之间的关系可以抽象为网络。网络（模型、算法等）研究的迅速发展孕育了一个新的学科，即网络科学。一些来自生物、物理和社会领域的研究人员在网络框架下调查了他们各自的问题[[53]](#endnote-52)，例如我们的案例研究。

为了跟进上述内容，人们会认为常见的研究范式和方法论，特别是那些整合为跨学科学科的研究范式，为科学类别之间的相互作用以及合作网络中巨大组成部分的形成提供了依据。分析论文内容有助于验证这些范式和方法的普遍性，这似乎很有希望。超过一半的PNAS 1999-2013论文包含“系统”和“控制”这两个主题词xlii。论文在一定程度上包含主题词的比例很高，这反映了主题的典型性。然而，这并不意味着一篇包含“系统”一词的论文和一篇应用系统科学研究成果的论文之间有很大的关系。因此，在语义层面上评估普遍性是一个有待进一步研究的课题。

参考文献

1. Weingart P (2012) A short history of knowledge formations. In R. Frodeman, J. Thompson Klein, & C. Mitcham (Eds.), The Oxford Handbook of Interdisciplinarity (pp. 3-14). Oxford, England: Oxford University Press. [↑](#endnote-ref-0)
2. National Academies (U.S.). Committee on Facilitating Interdisciplinary Research (2004).Facilitating interdisciplinary research. Washington: National Academy Press. Retrieved from http://www.nap.edu/books/0309094356/html/ [↑](#endnote-ref-1)
3. Hurd JM (1992) Interdisciplinary research in the sciences: Implications for library organizations. Coll Res Liber 53(4), 283-297. [↑](#endnote-ref-2)
4. Cooper G (2013) A disciplinary matter: Critical sociology, academic governance and interdisciplinarity. Sociology 47(1), 74-89. [↑](#endnote-ref-3)
5. Hadorn GH, Pohl C, Bammer G (2012) Solving problems through transdisciplinary research. In R. Frodeman, J. Thompson Klein, & C. Mitcham (Eds.), The Oxford Handbook of Interdisciplinarity (pp. 431-452). Oxford, England: Oxford University Press. [↑](#endnote-ref-4)
6. Siedlok F, Hibbert P (2014) The organization of interdisciplinary research: Modes, drivers and barriers. Int J Manage Rev 16(2), 194-210. [↑](#endnote-ref-5)
7. Liu Y, Rafols I, Rousseau R (2012) A framework for knowledge integration and diffffusion.J Doc 68(1), 31-44. [↑](#endnote-ref-6)
8. Gooch D, Vasalou A, Benton L (2017) Impact in interdisciplinary and cross-sector research: Opportunities and challenges. J Assoc Inf Sci Technol 68(2), 378-391. [↑](#endnote-ref-7)
9. Larivi´ere V, Gingras Y, Archambault E (2006) Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. Scientometrics 68(3): 519-533. [↑](#endnote-ref-8)
10. Moody J (2004) The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. Am Sociol Rev 69(2), 213-238. [↑](#endnote-ref-9)
11. Gl¨anzel W, Schoepflflin U (1999) A bibliometric study of reference literature in the sciences and social sciences. Inform Process Manag 35(1): 31-44. [↑](#endnote-ref-10)
12. Hicks D (1999) The diffiffifficulty of achieving full coverage of international social science literature and the bibliometric consequences. Scientometrics 44(2): 193-215. [↑](#endnote-ref-11)
13. Sarig¨ol E, Pfifitzner R, Scholtes I, Garas A, Schweitzer F (2014) Predicting scientifific success based on coauthorship networks. EPJ Data Science 2014:9. [↑](#endnote-ref-12)
14. Barab´asi AL, Jeong H, N´eda Z, Ravasz E, Schubert A, Vicsek T (2002) Evolution of the social network of scientifific collaborations. Physica A 311: 590-614. [↑](#endnote-ref-13)
15. Newman M (2001) The structure of scientifific collaboration networks. Proc Natl Acad Sci USA 98: 404-409. [↑](#endnote-ref-14)
16. Newman M (2004) Coauthorship networks and patterns of scientifific collaboration. Proc Natl Acad Sci USA 101: 5200-5205. [↑](#endnote-ref-15)
17. Xie Z, Ouyang ZZ, Li JP (2016) A geometric graph model for coauthorship networks. J Informetr 10: 299-311. [↑](#endnote-ref-16)
18. Newman M (2002) Assortative mixing in networks. Phys Rev Lett 89: 208701. [↑](#endnote-ref-17)
19. Tomasello MV, Vaccario G, Schweitzer F (2017) Data-driven modeling of collaboration networks: A cross-domain analysis. EPJ Data Science 6: 22. [↑](#endnote-ref-18)
20. Braun T, Schubert A (2003) A quantitative view on the coming of age of interdisciplinarity in the sciences, 1980-1999. Scientometrics 58(1), 183-189. [↑](#endnote-ref-19)
21. Levitt JM, Thelwall M, Oppenheim C (2011). Variations between subjects in the extent to which the social sciences have become more interdisciplinary. J Assoc Inf Sci Technol 62(6), 1118-1129. [↑](#endnote-ref-20)
22. Porter AL, Roessner JD, Cohenm AS, Perreault M (2006). Interdisciplinary research: Meaning, metrics and nurture. Res Eval 15(3), 187-195. [↑](#endnote-ref-21)
23. Porter AL, Rafols I (2009) Is science becoming more interdisciplinary? Measuring and mapping six research fifields over time. Scientometrics 81(3), 719-745. [↑](#endnote-ref-22)
24. Rafols I, Meyer M (2010) Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. Scientometrics 82(2), 263-287. [↑](#endnote-ref-23)
25. Abramo G, D’Angelo CA, Costa F (2012) Identifying interdisciplinarity through the disciplinary classifification of coauthors of scientifific publications. J Assoc Inf Sci Technol 63(11): 2206-2222. [↑](#endnote-ref-24)
26. Chen S, Arsenault C, Gingras Y, Larivi´ere V (2015) Exploring the interdisciplinary evolution of a discipline: The case of Biochemistry and Molecular Biology. Scientometrics 102(2), 1307-1323. [↑](#endnote-ref-25)
27. Bordons M, Zulueta MA, Romero F, Barrig´on S (1999) Measuring interdisciplinary collaboration within a university: The effffects of the multidisciplinary research programme.Scientometrics 46(3), 383-398. [↑](#endnote-ref-26)
28. Leydesdorffff L, Goldstone RL (2014) Interdisciplinarity at the journal and specialty level:The changing knowledge bases of the journal Cognitive Science. J Assoc Inf Sci Technol 65(1), 164-177. [↑](#endnote-ref-27)
29. Zhang L, Rousseau R, Gl¨anzel W (2015) Diversity of references as an indicator for interdisciplinarity of journals: Taking similarity between subject fifields into account. J Assoc Inf Sci Technol 67(5), 1257-1265. [↑](#endnote-ref-28)
30. Lungeanu A, Huang Y, Contractor NS (2014) Understanding the assembly of interdisciplinary teams and its impact on performance. J Informetr 8(1), 59-70. [↑](#endnote-ref-29)
31. Larivi´ere V, Gingras, Y (2010) On the relationship between interdisciplinarity and scientifific impact. J Assoc Inf Sci Technol 61(1), 126-131. [↑](#endnote-ref-30)
32. Larivi´ere V, Haustein S, B¨orner K (2015) Long-distance interdisciplinarity leads to higher scientifific impact. Plos One 10(3), e0122565. [↑](#endnote-ref-31)
33. Rinia EJ, van Leeuwen TN, van Raan AFJ (2002) Impact measures of interdisciplinary research in physics. Scientometrics 53(2), 241-248. [↑](#endnote-ref-32)
34. Wan J, Thijs B, Gl¨anzel W (2015) Interdisciplinarity and impact: Distinct effffects of variety, balance, and disparity. Plos One 10(5), e0127298. [↑](#endnote-ref-33)
35. Levitt JM, Thelwall M (2009) The most highly cited library and information science articles: interdisciplinarity, fifirst authors and citation patterns. Scientometrics 78(1), 45-67. [↑](#endnote-ref-34)
36. Levitt JM, Thelwall M (2008) Is multidisciplinary research more highly cited?: A macrolevel study. J Assoc Inf Sci Technol 59(12), 1973-1984. [↑](#endnote-ref-35)
37. Chen S, Arsenault C, Larivi´ere V (2015) Are top-cited papers more interdisciplinary?J Informetr 9(4): 1034-1046. [↑](#endnote-ref-36)
38. Stirling A (2007) A general framework for analyzing diversity in science, technology and society. J Roy Soc Interf 4(5), 707-719. [↑](#endnote-ref-37)
39. Leydesdorffff L (2007) Betweenness centrality as an indicator of the interdisciplinarity of scientifific journals. J Assoc Inf Sci Technol 58(9), 1303-1319. [↑](#endnote-ref-38)
40. Van den Besselaar P, Heimeriks G (2001, July). Disciplinary, multidisciplinary, interdisciplinary: Concepts and indicators. In ISSI (pp. 705-716). [↑](#endnote-ref-39)
41. Kagan J. The three cultures: Natural sciences, social sciences, and the humanities in the 21st century. Cambridge University Press, 2009. [↑](#endnote-ref-40)
42. Xie Z, Duan XJ, Ouyang ZZ, Zhang PY (2015) Quantitative analysis of the interdisciplinarity of applied mathematics. Plos One 10(9): e0137424. [↑](#endnote-ref-41)
43. Milojevi´c S. (2013). Accuracy of simple, initials-based methods for author name disambiguation. J Informetr 7(4): 767-773. [↑](#endnote-ref-42)
44. Kim J, Diesner J (2016) Distortive effffects of initial-based name disambiguation on measurements of large-scale coauthorship networks. J Assoc Inf Sci Technol 67(6):1446- 1461. [↑](#endnote-ref-43)
45. Milojevi´c S (2010) Modes of Collaboration in Modern Science: Beyond Power Laws and Preferential Attachment. J Assoc Inf Sci Technol 61(7): 1410-1423. [↑](#endnote-ref-44)
46. Xie Z, Ouyang ZZ, Li JP, Dong EM, Yi DY (2018) Modelling transition phenomena of scientifific coauthorship networks. J Assoc Inf Sci Technol 69(2): 305-317. [↑](#endnote-ref-45)
47. Consul PC, Jain GC (1973) A generalization of the Poisson distribution. Technometrics 15(4): 791-799. [↑](#endnote-ref-46)
48. Xie Z, Xie ZL, Li M, Li JP, Yi DY (2017) Modeling the coevolution between citations and coauthorship of scientifific papers. Scientometrics 112, 483-507. [↑](#endnote-ref-47)
49. Levitt JM, Thelwall M, Oppenheim C (2011) Variations between subjects in the extent to which the social sciences have become more interdisciplinary. J Assoc Inf Sci Technol 62(6), 1118-1129. [↑](#endnote-ref-48)
50. Hey T, Tansley S, Tolle KM (2009) The fourth paradigm: data-intensive scientifific discovery, Microsoft research, Redmond, Washington. [↑](#endnote-ref-49)
51. Haythornthwaite C (2006). Learning and knowledge networks in interdisciplinary collaborations. J Assoc Inf Sci Technol 57(8), 1079-1092. [↑](#endnote-ref-50)
52. Grauwin S, Beslon G, Eric Fleury, Franceschelli S, Robardet C, Rouquier JB, Jensen P (2012) Complex systems science: dreams of universality, interdisciplinarity reality. J Assoc Inf Sci Technol 63(7), 1327-1338. [↑](#endnote-ref-51)
53. Brier S (2013) Cybersemiotics: a new foundation for transdisciplinary theory of information, cognition, meaningful communication and the interaction between nature and culture. Integr Rev 9: 222-263. [↑](#endnote-ref-52)