

信息网络关键节点对之删除判定*

赵 星

摘 要 节点对,指网络中一条联系及其两端节点的组合。作为信息网络中底层的联系组件,节点对表征最基础的信道,但却难以被直接测度。本文尝试引入删除法,构建节点对的间接测度,并给出分别测量网络整体节点数量(信息存量)和联系权重(信息流量)损失的 L_N 和 L_S 指数,用于挖掘无向网络中的关键节点对。h 指数研究科研合作网络的实例显示,删除法挖掘出了该网络中具有组织作用、具有桥梁作用或具有相似结构对等性的各类关键节点对。基于删除法的信息网络间接测度方法,有望促成信息网络联系组件测度研究的新视域。图 7。表 2。参考文献 26。

关键词 信息网络 复杂网络 社会网络 节点对 联系测度

分类号 G250.2

Exploring the Vital Node Pairs in Information Networks by Deletion Algorithm

ZHAO Xing

ABSTRACT

The flow of information is one of the major values of information. In the evolving information world, nodes are generated by information content and carrier. Meanwhile, links derive from information behaviors and interactions. Then nodes and links construct networks. Such networks present the abstract picture of the real information world.

In the previous studies of networks measures, the power-law distribution of nodes in complex networks, the centrality measures in social networks, and h-type networks measures in Library Science all focus on nodes instead of links. However, studying the interactions and relations of information is an important task of information analysis. From the point of view of connection, the most basic components can be expressed as node pairs, a combination of a link and nodes at both ends. In information networks, a node pair usually represents a complete channel, which has significant mining and measuring values.

This paper selects node pairs, which are the most basic component of links, as research object and attempts to measure them indirectly. The study will use the “deletion algorithm”, which is common in computer science, to construct the measure method of key node pairs. The general idea is to delete the

* 本刊“青年学术论坛”特约稿(Special contribution for the Youth Academic Forum sponsored by this Journal)

本文系国家自然科学基金青年项目“h 型信息网络测度的机理与实证研究”(编号:71503083)的研究成果之一。(This article is an outcome of the youth project “The Mechanism and Empirical Study of h-Type Information Network Measures” (No. 71503083) supported by National Natural Science Foundation of China.)

通信作者:赵星,Email: xzhao@infor.ecnu.edu.cn, ORCID:0000-0001-9347-590X (Correspondence should be addressed to ZHAO Xing, Email: xzhao@infor.ecnu.edu.cn, ORCID:0000-0001-9347-590X)

targeted node pair in the information networks instead of measuring it directly, then indicators are designed to observe the changes of the whole networks. The degree of change reflects the importance of the node pair, which results in a certain degree of equivalence in the node pairs that have same networks variation.

Based on the deletion algorithm, the author proposes the L_N and L_S indexes respectively for measuring the numbers of nodes (information stock) and the loss of link's weight (information flow) in order to find key node pairs in the undirected networks. An example of scientific collaboration network related to the topic of h-index shows that the method has unearthed various types of key node pairs that have similar structural equivalence, playing an organizational role or performing connection functions in the networks.

This study suggests that the development of innovative measures is the basis of information networks research. In the meanwhile, the measures of link and its components in information networks are key to measuring the information networks. The indirect measures has provided new ideas for the investigation of links in information networks. The deletion algorithm is just one of the indirect measure methods. It is expected that the combination of direct and indirect measure methods will lead to the complete framework of information network analysis. 7 figs. 2 tabs. 26 refs.

KEY WORDS

Information networks. Complex networks. Social networks. Node pairs. Link measures.

0 引论

流动与传承,是信息的核心价值之一。在信息世界的不断演化中,信息内容与载体生成节点,信息行为及交互衍出联系,节点与联系织为网络。这类网络,抽象出了信息世界的现实图景。

本世纪以来,由社会学的社会网络和物理学的复杂网络共同推进的网络科学(Network Science)已成为分析网络的最佳范式^[1-2]。2010年,著名物理学家Newman出版了他的集大成之作*Networks: An Introduction*^[3]。该书根据系统的类别特征,将电力、交通等工程类网络归为技术网络,将生态网络等生命科学类网络划为生物网络,把人类社会关系网络归于社会网络,将引文网络^[4]、信息流动网络^[5]和网页链接网络^[6]等图情学研究内容纳入信息网络。Newman还指出,信息网络的特点,是网络中的节点和联系标识了信息(或知识)的存储和扩散^[3]。可见,直面图情学内容的信息网络已受到了主流学界的重视,构成了现实网络的重要大类。

节点与联系是网络的基本单元。在过去的

研究中,无论是复杂网络对于节点度幂律分布的普适发现^[7],还是社会网络对于中心度测量的孜孜以求^[8],亦或图书情报学在h型网络测度上的开拓创新^[9],就分析对象而言,多将目光凝固于网络中的节点,而非联系。万物可分,分则可解,这一还原论的惯性思维在网络分析中依然挥之不去。但事实上,网络分析与网络科学诞生之初,便带有强烈的系统论观点,其范式长于观测整体结构,而非限于节点自身。因此,关注“联系”这一网络中的重要组件或能构成新的研究视域^[10-11]。特别是,图情学科长期以来在信息相关性、信息共现^[12]等议题的探索发现:信息分析的要义之一是抽象与测量信息之间的关联。因此,信息网络分析的任务除了研究节点,还有测度联系及其组件。而从联系出发,最基本的组件可表述为节点对(Node Pairs),即一条联系和两端节点的组合。信息网络中,一组节点对通常表示一条完整信道。一组关键节点对即一条在网络中具有重要结构功能的信息渠道,具有重要的挖掘和测度价值。

然而,涉及联系的组件在测度复杂性上要远高于节点。首先,联系数据属于关系型数据,

与传统统计数据中实体变量(如个人的经济收入)的独立存在不同,每一条联系除了联系自身,还依存于两端的节点(如甲和乙相互传递了情报);即,完整的测量联系一定程度上需要转化为测量节点对。其次,网络分析的范式要求,对于组件的测度,更重要的还并非其自身属性,而是其在整体中的功能结构;即,测量节点对本质上要求测度其对整体网络的影响和作用。因此,直接测度包括节点对在内的联系组件具有一定难度,也容易陷于重局部而轻整体的局限。

本文尝试从间接的角度对联系组件进行测量,以节点对这一最基础的联系组件为研究对象,展开具体讨论。研究将移用计算机学科中朴素而常见的“删除法”,构建信息网络分析中的关键节点对测度方法。总体思路为:不直接测度信息网络中的节点对,而是先将该节点对从网络中删除,然后再设计简洁的测度指标,观测网络整体的变化;网络变化程度体现了该节点对的重要性,而导致网络变化幅度相同的节点对,则具有结构上的某种程度对等性。本文方法的基本理念是:失之交臂,方知珍惜。系统中存在的重要组件未必总能被直接观察到,而当其被阻断或失去时,其关键作用才会“突然”被普遍认知,信息网络中的关键通道,常是如此。

1 理论方法

1.1 节点对

信息网络中,最为简单的信息交流常由信源、信宿及其之间的信道三者共同构成。信息传播即信源通过信道将信息传递给信宿。已有的信息网络分析中,信源和信宿的角色类型通常是一致的,即信息传播者和接受者都归于节点。例如,学者合作网络本质上表述的是学者间相互交流拥有的知识和信息并创造新的科学知识的系统。两名学者节点之间的信息互动构成最基本的学术交流行为。社会网络分析认为,这种一条联系及两端节点的组合(节点对)是最简单的网络关系^[13]。在图情学领域,这也是能够抽象信息

交流的最为基础的网络组件。基于此,本文尝试规范节点对(Node Pairs)的表述,具体如下。

定义 1 网络中,一条联系及两端节点所共同构成的组件,可称为节点对,使用 NP_{ij} 表示, i 和 j 为联系两端的节点。

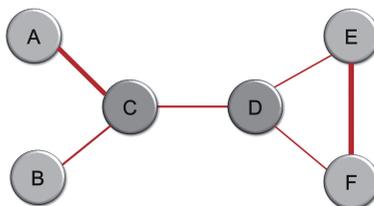


图 1 网络中的节点对示例

如图 1 所示,在该网络中,共有 6 条节点对,即 NP_{AC} 、 NP_{BC} 、 NP_{CD} 、 NP_{DE} 、 NP_{DF} 、 NP_{EF} 。从联系的角度出发,根据联系是否具有方向和权重,网络可分为无权无向网络、无权有向网络、带权无向网络和带权有向网络四种类型^[14]。本文主要基于无权无向网络和带权无向网络进行讨论,在无向网络的分析框架厘清之后,有向网络的应用可参照文献[15]的处理方式,将网络中的联系根据方向分别计算,本文不再赘述。

无向网络中,节点对有如下基本性质。

性质 1 若 $NP(i)$ 为涵盖节点 i 的节点对总数, $D(i)$ 为节点 i 的度, $WD(i)$ 为节点 i 的权重,则对于任一节点 i ,恒有:

$$NP(i) = D(i) \leqslant WD(i) \quad (1)$$

性质 1 说明,无向网络中,含有某个节点的节点对总数等于该节点的度。

性质 2 在网络中,若 NP 为网络节点对总数, NI 为网络中的联系总数, NI_s 表示网络中联系强度的总和,则恒有:

$$NP = NI \leqslant NI_s \quad (2)$$

性质 2 说明,无向网络中,节点对总数等于联系总数。

性质 3 在节点数量为 n 的网络中, NP 为网络节点对总数, $D(i)$ 为节点 i 的度,则在网络中,恒有:

$$NP = \frac{\sum_1^n D(i)}{2} \quad (3)$$

性质 4 在节点数量为 n 的网络中, NP 为网络节点对总数, 恒有:

$$0 \leq NP \leq \frac{n(n-1)}{2} \quad (4)$$

公式(4)中 $\frac{n(n-1)}{2}$ 为网络联系数量最大理论值。当网络为全连通网络时, $NP = \frac{n(n-1)}{2}$ 成立。

性质 1—4 共同表明, 节点对在量上与网络中总体节点和联系具有直接关系, 其数量与网络中节点和联系存在宏观或微观层面上的数理联系, 值得后续理论研究关注。同时也得出启示, 在测度上, 节点对的测量可考虑与节点和联系的指标进行关联。

1.2 删除法与节点对测度

1.2.1 节点对之删除测度

位置决定地位, 网络中所处的结构地位显著影响其社会资本, 这是社会学“社会资本”理论^[16-17]的基本观点。与此类似, 节点对在网络中的位置、地位和功能也显然会有不同。同网络中的节点一样, 节点对在网络中所处位置、结构等存在差别, 导致了网络中节点对的重要性也有所不同。如图 1 所示, 示例网络中节点对 NP_{CD} 连接了网络左右, 若将其删除, 整个网络将被分割为多个零散组件, 无法连通。可见, 假设节点对消失, 再观测网络整体的变化, 是一种测量节点对位置结构重要性的可尝试方法。

实际上, 在计算机等领域被广泛使用的“删除法”或类似视角早已在网络分析中被提及

或使用, Callaway 和 Newman 等人^[18]研究了节点损失对网络整体的影响, 发现虽然无标度网络比随机网络具有更强的抗毁性, 但在选择性打击下, 只需攻破 5% 的核心节点, 无标度网络就基本瘫痪。Nardelli 和 Proietti^[19]探索了删除节点后网络最短路径的变化, 以此评判节点的重要性。然而, 删除法至今没有成为重要节点挖掘的主流方法, 究其根由, 在于节点已有较多可以直接测度的定量方法。特别是伴随社会网络成为显学的中心性方法, 在理论和应用上都有相当的认可度^[3]。删除法作为一种间接测量节点重要性的方法, 在节点分析中并无太多优势。

但是, 关于信息网络关键节点对的挖掘这一问题, 删除法或更有潜力。首先, 如引论中所述, 节点对涉及关系型数据, 这类数据直接测量难度较大, 故以删除法为基础构建间接测度, 是值得考虑的方向。其次, 系统观点和整体研究是网络分析, 特别是联系分析的要义, 直接测度容易陷于细节而忽略网络整体, 间接的整体测度可望发掘出更高层次的信息。再者, 目前信息网络的测度方法中, h 型测度^[20-22]已构成一类具有学科特色的直接测度, 若能再辅以简洁有效的间接测度, 信息网络的测度方法体系将进一步完善。

基于以上讨论, 可以在节点对被删除后, 将最大联通网络的变化作为节点对重要性量化的参考依据(如图 2 所示)。前述节点对理论性质的讨论也表明, 节点对与网络中节点和联系的数量及权重密切相关。因此, 本文将给出两类测度指标: 一类指向节点对被删除后, 网络节点的损失程度, 一定程度表征信息网络的信息存

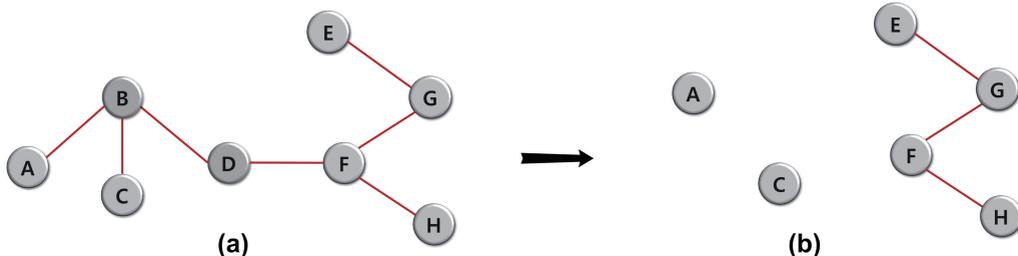


图 2 节点对 NP_{BD} 被删除后网络的变化示意

量损失;一类面向节点对被删除后,网络联系权重的损失程度,一定程度描述信息网络的信息流量损失。

1.2.2 信息节点损失测量: L_N

信息网络中,网络节点全连通是信息扩散和交互的物理基础。节点作为信息交流传播的参与者,一个节点所处的连通网络规模越大,就意味着该节点接收和扩散信息的潜在范围越广。依据节点对被删除后最大连通网络节点规模的变化,给出节点对基于删除法的节点(信息存量)损失测度指标 L_N ,定义如下:

定义 2 L_N 指某节点对从网络中被删除后,剩余网络中最大连通子网相比于原网络丢失的节点数占原网络节点数的比率。

由定义 2 可见,节点对的 L_N 值越大,表明该节点对的“消失”对网络造成的信息存量损失程度越大,此视角下,该节点对在网络中地位越重要。根据节点对测度指标 L_N 的定义,其计算公式为:

$$L_N(NP_{ij}) = \frac{(N - N_{max})}{N} \times 100\% \quad (5)$$

其中, N_{max} 为节点对被删除后,剩余最大连通子网的节点数量, N 为原网络节点总数。以图 2(a) 网络为例,在该 8 个节点和 7 条联系组成的网络中,共有 7 组节点对,分别为 NP_{AB} 、 NP_{BC} 、 NP_{BD} 、 NP_{DF} 、 NP_{EG} 、 NP_{FG} 、 NP_{FH} 。若要计算 $L_N(NP_{BD})$,先将节点对 NP_{BD} 从网络中删除,删除该节点对后得到的网络如图 2(b) 所示,该网络中最大连通网络由节点 E、F、G、H 组成,共有 4 个节点,相比原网络减少了 4 个节点,则节点对 NP_{BD} 的 L_N 值为 50%,即若节点对 NP_{BD} “消失”,网络整体将至少失去 50% 的信息节点。

节点对测度指标 L_N 及相关参数有如下性质:

性质 5 当 N 为网络节点总数, N_{lost} 表示新网络中最大连通子网相比原连通网络丢失的节点数,恒有:

$$2 \leq N_{lost} \leq N - 1 \quad (6)$$

性质 6 当 N 为网络节点总数,对于任意节

点对的 L_N ,恒有:

$$\frac{2}{N} \leq L_N \leq \frac{N - 1}{N} \quad (7)$$

性质 6 约定了 L_N 的理论取值范围。

1.2.3 信息联系损失测量: L_S

丰富的信息世界要求多维的测度视角^[23],完整反映节点对重要性还需要考察网络在联系方面的变化。信息联系表征信息行为,是信息传递的路径和渠道。信息网络中多是带权网络^[14],带权网络中的联系强度又常体现信息流量的强弱,具有重要分析价值。节点对从网络中被删除后,相对于原网络,新的最大连通子网中总联系强度缺失得越多,说明损失的信息流量或潜在信息通道越多。根据节点对被删除后最大连通网络联系强度的变化,给出节点对基于删除法的联系(信息流量)损失测度指标 L_S ,定义如下:

定义 3 L_S 指某节点对从网络中删除后,剩余网络中最大连通子网相比于原网络所丢失的联系权重占原网络所有联系权重的比率。

由定义 3 可见,节点对的 L_S 值越大,表明该节点对的“消失”对网络造成了越大规模的信息流量损失,此视角下,该节点对在网络中的地位越重要。根据定义 3,节点对 NS_{ij} 的 L_S 的计算公式为:

$$L_S(NS_{ij}) = \frac{(S - S_{max})}{S} \times 100\% \quad (8)$$

其中, S_{max} 为节点对在网络中被删除后,剩余的最大连通子网的联系强度之和, S 为原网络联系权重总和。 L_S 的计算类似于 L_N ,不同之处在于纳入计算的是联系强度,故不再列举算例。 L_S 有以下性质。

性质 7 当网络中联系最小强度为 $\min(S)$ 时, S 表示网络中所有联系强度之和,对于任一节点对的 L_S ,恒有:

$$0 < \frac{2 * \min(S)}{S} \leq L_S \leq 1 \quad (9)$$

连通网络中,若某节点对从网络中删除,则至少会失去两条联系(如线状网络的两端节点

对),此时,相比于原网络,总联系强度至少减少最小联系强度的两倍数,故有性质7。特殊的,在无权网络中,公式(9)变为 $0 < \frac{2}{S} \leq L_S \leq 1$ 。

无论对于 L_S 还是 L_N ,当有节点对的计算结果数值等同时,说明这些节点对在相关意义上具有结构对等性,即其结构功能在一定程度上等同甚至可相互替换。对于标准的星状网络和环状网络而言,其所有节点对都具有结构对等性,如图3所示。

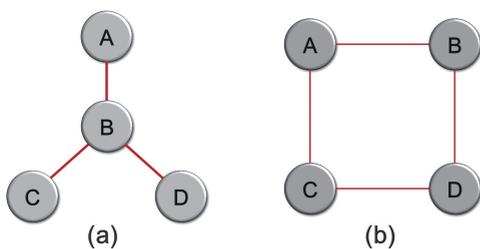


图3 所有节点对之间具有结构对等性的星状(a)与环状(b)网络

2 案例实证

在理论讨论之后,现给出现实案例用于演示本文方法的实现。科研合作网络是图情学领域研究最多的网络类型之一,同时也是较为典型的信息网络类型。学者节点通过各种方式进行联系交流与研究合作,形成知识和信息的交互升华与组合创新,并以共同署名的方式发表成果,已构成当今学术界最重要的创新方式之一^[24]。本文选用2005—2012年共计8年的h指数研究论文作为样本,构建作者之间的科研合作网络进行后续实证。这一选样主要基于以下理由:首先,h指数作为2005年提出的全新概念^[25],文献边界明确;其次,h指数研究是本世纪以来信息计量学乃至整个图情学最重要的热点主题,截至2018年7月31日,Hirsch提出h指数的原始文献已在谷歌学术中被引用了7 852次,仅用十余年时间就超越了所有文献计量学

经典文献,堪称学科史上最值得关注的案例之一;第三,本文作者自2007年起对h指数就有粗浅的跟踪研究,对此领域有一定了解,利于数值结果的内容解读。此外,选择2005年到2012年而不是到现今的数据,主要是因为h指数研究聚焦和成果密集涌现的时间是2005—2012年^[9]。之后,因h指数的热度,有大量文献仅是提及引用,并非专门的h指数研究,不宜纳入本文数据。

2.1 数据采集与处理

依据以上样本选择思路,利用Web of Science数据库,分两步进行数据采集与处理。

(1) 全网构建

以Hirsch提出h指数的原始文献*An index to quantify an individual's scientific research output*为始,检索了Web of Science核心集,2005—2012年间所有引用该文献的article类型的文献,共计874篇,从这些文献中提取和清洗出1 448名作者及作者间的合作关系,以作者为网络中的节点,作者之间的共署名关系作为联系,作者间合作频次为联系强度,作者的所有论文被引总频次为节点权重,构建出h指数研究主题下的科研合作网络。最终网络共有1 448个节点和2 146条联系,整体结构如图4所示。

(2) 最大连通子网抽取

全网联通是信息能顺利流动的基础,也是非涉密类信息网络努力演化和发展的方向。对于科研合作网络而言,大范畴的广泛合作更是科学界孜孜以求的目标。故本文方法设计为针对联通网络的计算,需抽取整体网络中的最大连通网络,结果如图5所示,该网络正好是整个h指数研究中最重要合作群体,包含了Rousseau、Egghe、Leydesdorff、Bornmann等欧洲最著名的文献计量学家,亦包含了叶鹰、梁立明和金碧辉等具有重要国际影响的我国文献计量学家。值得注意的是,在h指数研究初期,于理论模型和期刊h指数等方面做出重要贡献的Glanzel、Schubert并未出现在网络中,是因为统计期内他

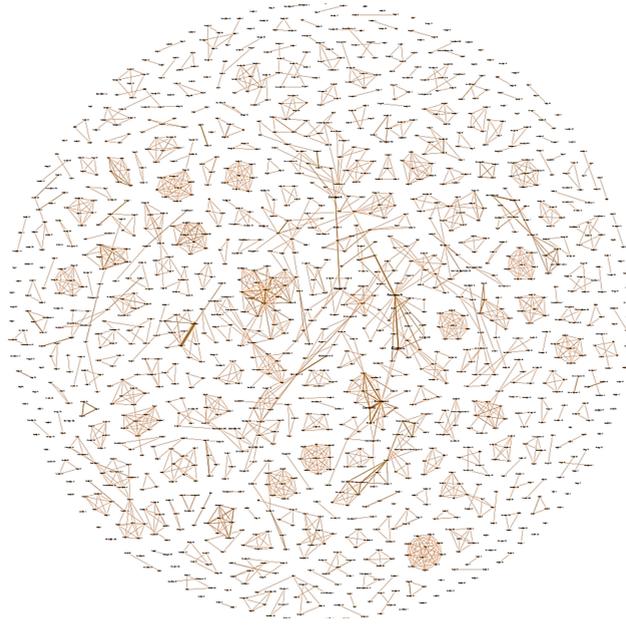


图 4 以 h 指数研究主题文献构建的作者合作网络

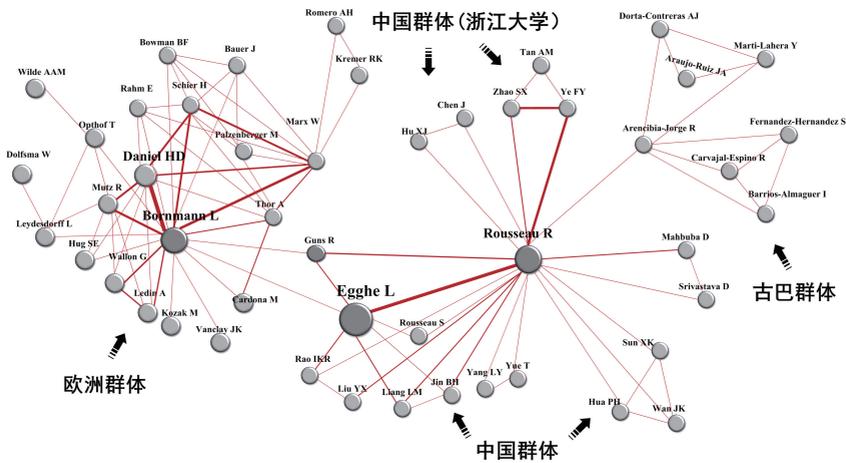


图 5 以“h 指数”论文施引文献为基础构建的科研合作网络中最大连通网络

们与这一最大连通网络中的学者并无合作关系。实际上,这两位著名的普莱斯奖得主同期领导了另一个合作群体。虽然几乎所有活跃的欧洲文献计量学家都参与了 h 指数的研究,但这一网络构建结果也折射出欧洲文献计量学界实际上有派别之分。当然,总体而言,某一主题的科研合

作网络虽并不一定整体连通,但其中的较大连通子网,常还是能表征出具有合作联系的重要群体。

2.2 结果与分析

2.2.1 关键节点对的 L_N

使用图 5 中的网络,依次交替删除所有节点

对,同时使用 L_N 和 L_S 观测网络整体变化。本文网络案例中,理论上 L_N 指标的最大值为 99.13%,此时要求该节点对从网络中删除,余下所有的节点都是彼此没有联系的孤立节点;最小值为 1.74%,要求该节点对从网络中剔除,余

下的所有节点都处在同一连通网络中,当然,这都是极端理论状态。经过计算,案例网络中 L_N 的实际最大值为 53.0612%,最小值为 4.0816%。其中, L_N 指标排名前 5 的 22 组节点对如表 1 所示。

表 1 L_N 指标排名前 5 节点对

排名	NP_{ij}			$L_N(NP_{ij})$	排名	NP_{ij}			$L_N(NP_{ij})$
1	Egghe L	-	Rousseau R	53.061%	5	Bornmann L	-	Thor A	44.898%
2	Egghe L	-	Guns R	48.980%	5	Rao I K R	-	Rousseau R	44.898%
3	Bornmann L	-	Egghe L	46.939%	5	Bornmann L	-	Vancley J K	44.898%
3	Bornmann L	-	Guns R	46.939%	5	Bornmann L	-	Ophhof T	44.898%
5	Bornmann L	-	Daniel H D	44.898%	5	Bornmann L	-	Palzenberger M	44.898%
5	Bornmann L	-	Mutz R	44.898%	5	Bornmann L	-	Rahm E	44.898%
5	Bornmann L	-	Marx W	44.898%	5	Bornmann L	-	Kozak M	44.898%
5	Bornmann L	-	Schier H	44.898%	5	Bornmann L	-	Hug S E	44.898%
5	Bornmann L	-	Wallon G	44.898%	5	Bornmann L	-	Cardona M	44.898%
5	Bornmann L	-	Ledin A	44.898%	5	Bornmann L	-	Bowman B F	44.898%
5	Bornmann L	-	Leydesdorff L	44.898%	5	Bauer J	-	Bornmann L	44.898%

由表 1 可见,从删除后网络节点损失的角度,网络中节点对 $NP_{Egghe L, Rousseau R}$ 最为重要。若将该节点对从网络中删除,过半的节点将被排出剩余的最大连通网络,网络节点将被分割为 10 个规模不等的互不连通的子群甚至成为孤立节点^①,网络整体已呈分崩离析的状态(如图 6 所示)。这一结果也与 h 指数研究的实际情况相符。Egghe 和 Rousseau 作为同一时代的比利时信息计量学家代表,都是数学出身,均为信息计量学数理派的倡导者和代表人物,长于理论模型建构和推导,对部分实证研究持保留态度,类似的范式 and 学缘使他们成为数十年的亲密战友,还共同创办了 *Journal of Informetrics (JOI)* 这一信息计量学国际顶级刊物。Egghe 和 Rousseau

紧密合作进行 h 指数研究期间,正是 h 指数大发展的时期,大量关于 h 指数的高被引论文涌现并刊载于他们主编的 *JOI*,直接促使该刊影响因子超过了 4^②,*JOI* 亦成为历史上第一本影响因子达到 4 的图情学国际刊物^③。节点对 $NP_{Egghe L, Rousseau R}$ 某种意义上也是整个网络的实际组织者,Egghe 连通多位欧亚顶尖计量学家,Rousseau 带动了欧洲地区之外的大量计量学研究团队,在这一节点对的组织和领导下,整个网络贡献了 h 指数的 h 核概念、Egghe-Rousseau 模型、幂律模型、g 指数、R 指数和 h 型信息网络测度等众多成体系的 h 指数研究成果。可见, L_N 较好地识别出了这一具有组织作用的关键节点对。

除了 Egghe 和 Rousseau 构成的节点对之外,

① 孤立节点 Rousseau S 是有“信息计量学之父”之称的 Rousseau R 的女儿,但实际上是经济学学者,与其他图情学和信息计量学交集甚少。

② 2014 年后,Rousseau R 和 Egghe L 相继淡出 *Journal of Informetrics* 编辑部,h 指数的热度也开始消退,该刊的影响因子近年已滑落到 2-3 之间。

③ 此处的图情学刊物指实际上属于图书情报范畴的传统领域期刊,不包括 SSCI 分类中被纳入图情学科的管理学(信息系统)、计算机科学等学科的刊物。

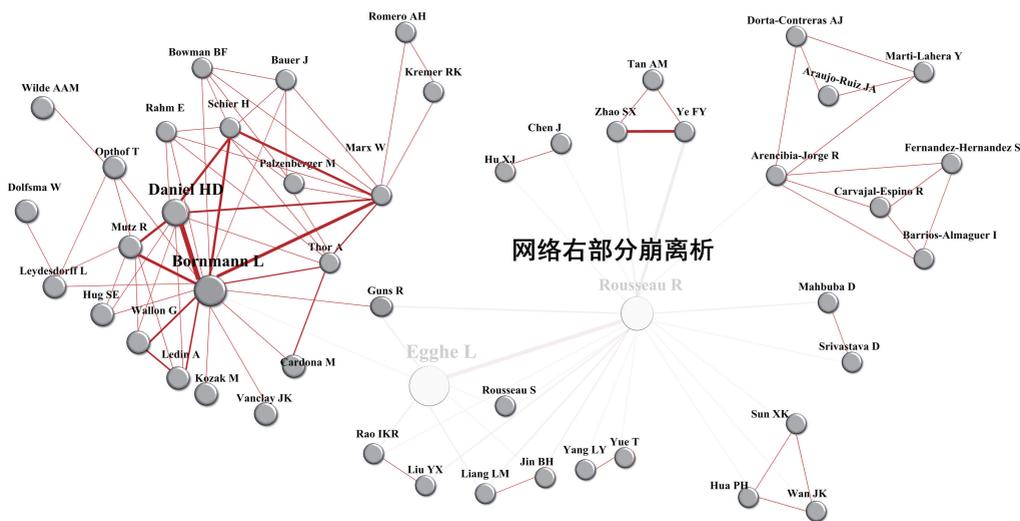


图6 节点对 $NP_{Egghe L, Rousseau R}$ 被删除后分崩离析的网络

节点对 $NP_{Egghe L, Guns R}$ 也具有重要的作用。有趣的是, Guns 作为 Rousseau 最得意的弟子之一, 其与 Egghe 的合作节点对 L_N 值超过了与其导师的合作节点对。由图 5 和图 6 可见, 出现这一结果的重要原因节点对 $NP_{Egghe L, Guns R}$ 在位置上具有特殊性, 若将该节点对删除, 欧洲学者群体和其他大洲学者的联系将完全断裂, 剩余的网络中, 左部子群虽有 Bornmann 被众星捧月, 右部子群有 Rousseau 领导中国和古巴学者进行合作, 但两个子群再无交集。删除节点对 $NP_{Rousseau R, Guns R}$ 的后果则没有这么严重, 部分中国学者与欧洲学者群体之间还有 Egghe 作为通道。可见, L_N 较好地识别出了这一具有桥梁作用的关键节点对。

表 1 中第三个有趣现象是 Bornmann 的反复出现。Bornmann 作为欧洲计量学家中的中生代“新星”, 可谓当今最活跃的信息计量学家之一, 近年每年发表国际论文超过 20 篇, 是为数不多的能在莱顿学派和 Leydesdorff 之间左右逢源的学者。他的活跃很大程度上带动了欧洲学术群体的合作创新, 形成了重要的子群, 表 1 中 22 组节点对中, 有 19 组与之相关, 并多具有相同的 L_N 。可见, 这些与 Bornmann 相关的节点对大多

具有结构对等性, 在网络中扮演类似的结构角色。具体分析可以发现, 这些节点对于整体网络的影响大多其实并不显著, 其节点对测度值较高主要是因为与 Bornmann 存在合作联系构成节点对。“重要的人的朋友亦重要”, Bornmann 的结果启示, 节点的重要性可以通过节点对传递给联系的另一端。

2.2.2 关键节点对的 L_S

将目光转移到联系权重上, 表 2 给出了表征删除节点对后信息流量损失大小的 L_S 指标。案例网络中, 理论上 L_S 指标的最大值为 100%, 即该节点对从网络中删除后, 余下网络中不存在任何一条联系, 所有的节点都是孤立节点。另一方面, L_S 指标理论最小值为 1.162 8%, 即该节点对从网络中剔除后, 网络中仅缺失两条权重为 1 (最低权重) 的联系, 且余下的所有节点都处在同一连通网络中。实际测算结果如下, 该网络 L_S 指标的实际最大值为 68.023 3%, 实际最小值为 1.744 2%。其中, L_S 值排名前 4 的节点对共计 20 条, 如表 2 所示。

由表 2 可见, 在本文案例中, 使用 L_N 和 L_S 的测度结果在总体上具有一致性, 表 1 和表 2 的节点对大多一致, 这说明, 案例中节点的损失与

表 2 L_S 指标排名前 4 节点对

排名	NP_{ij}		$L_S(NP_{ij})$	排名	NP_{ij}		$L_S(NP_{ij})$
1	Egghe L	-	Guns R	4	Bornmann L	-	Bowman B F
2	Rousseau R	-	Egghe L	4	Bauer J	-	Bornmann L
3	Bornmann L	-	Guns R	4	Bornmann L	-	Wallon G
4	Bornmann L	-	Marx W	4	Bornmann L	-	Ledin A
4	Bornmann L	-	Daniel H D	4	Bornmann L	-	Leydesdorff L
4	Bornmann L	-	Schier H	4	Bornmann L	-	Ophof T
4	Bornmann L	-	Mutz R	4	Bornmann L	-	Hug S E
4	Bornmann L	-	Thor A	4	Bornmann L	-	Cardona M
4	Bornmann L	-	Palzenberger M	4	Bornmann L	-	Vancley J K
4	Bornmann L	-	Rahm E	4	Bornmann L	-	Kozak M

联系权重的损失密切相关,尽管性质 1—7 的理论讨论也或多或少支持了这一结果,但节点对删除、节点损失和联系权重损失之间的实证关系仍需更多案例验证。特别是在不同的网络类型中,这三者的关系需要具体讨论。

当然, L_N 和 L_S 的结果在细节上仍有一些差异。首先,当主要的关键节点对被删除时,表 2 中联系权重损失比率普遍高于表 1 中节点数量的损失。其次,在两个指标的测评中,最重要的节点对位次被互换。在 L_S 的测量中,节点对 $NP_{Egghe L, Guns R}$ 被删除(如图 7 所示)会导致左右

两个网络的割裂,造成最多联系强度的损失。这也与前述 L_N 的相关讨论相互印证, $NP_{Egghe L, Guns R}$ 在实证中被证实具有关键桥组件的结构功能,沟通了两个拥有大量信息存量和信息流量的子群。虽然 $NP_{Egghe L, Rousseau R}$ 的删除也会导致网络被割裂,且与 $NP_{Egghe L, Guns R}$ 的 L_S 数值差异不大。但无论是信息计量学还是 h 指数研究,欧洲群体的主导地位依然明确,其信息存量和流量仍有优势。综上所述, L_N 和 L_S 的结合使用,可望配合挖掘出网络中占据最关键结构位置的节点对。

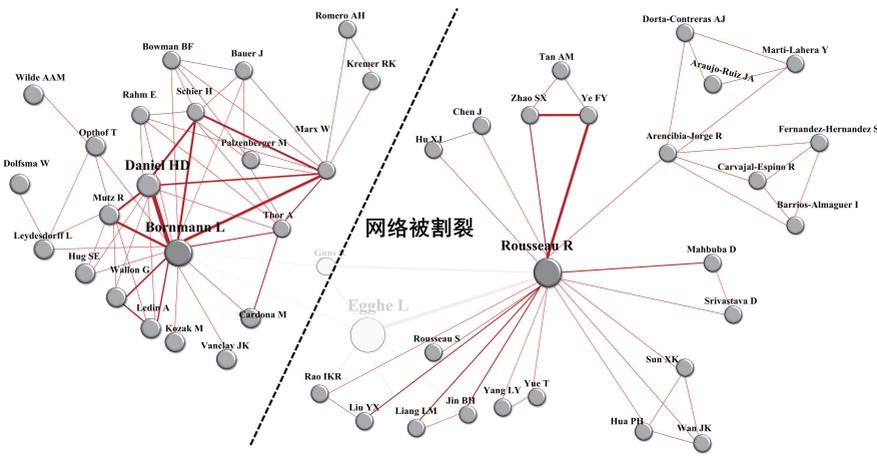


图 7 节点对 $NP_{Egghe L, Rousseau R}$ 删除后被割裂的网络

3 结语

近年来,网络科学和网络分析已进入各学科根据自身网络特点各自推进的阶段。作为以系统论为哲学观点,以关系型数据定量分析为范式特色的方法型研究议题,测度方法在信息网络研究中更具有基础性。

本文针对联系及其组件的测度,从节点对这一底层联系组件出发,移用计算机领域删除法导出节点对的间接测度方法,并给出分别测评节点数量(信息存量)损失和联系权重(信息流量)损失的 L_N 和 L_S 指数,用于挖掘和分析无向网络中的关键节点对。h指数研究主题论文构建的实证网络显示,本文方法可挖掘出该网络中的关键节点对,并与实际情况基本一致,识别出桥梁型的关键节点对, L_N 和 L_S 指数也得到了总体可相互印证但侧重略有不同的结果。本文方法还可用于考察节点对的结构对等性。

当然,本文仅是以删除法间接测度网络联系组件的探索性研究。全文仅讨论了无向网络,针对有向网络时,需要考虑将有向网络分解为两个固定方向的网络^[15]或发展其他创新方

法。在图情学科中,计算简洁性及物理意义明确性一直是指标是否具有应用潜力的关键要素,例如现今评价中最常用的h指数和影响因子等都是很简洁的测度。本文给出测评指标的 L_N 和 L_S 指数,在设计时尽可能考虑了计算和含义的简明性,但未必是最佳的测评视角和方法,下一步工作可在此基础上考虑更为综合和创新的测度指标,例如进一步将损失节点和联系的重要性纳入计算。删除节点对后若得到较大的 L_N 和 L_S 指数取值,说明网络中可能存在较显著的“结构洞”^[26],本文方法与“结构洞”理论的分析,也将是值得探索的后续方向。科研合作网络仅是信息网络的一种类型,更多的信息网络类型呼唤具有特色的新测度,热土有待开拓,盼图情学界能早日给出可影响整个网络科学的创新性测度方法。

致谢:2017年《中国图书馆学报》青年学术论坛召开期间,叶鹰先生、卓连营先生和刘炜先生对本研究提出了宝贵指导意见。之后的论文写作过程中,黄萃女士、刘晓钟先生、裴明明先生、楼雯女士和杨洁女士提供了重要建议和帮助,特致谢意。

参考文献

- [1] Vespignani A. Twenty years of network science[J]. Nature, 2018, 558(7711):528-529.
- [2] Borgatti S P, Mehra A, Brass D J, et al. Network analysis in the social sciences[J]. Science, 2009, 323(5916):892-895.
- [3] Newman M E J. Networks: an introduction[M]. Oxford: Oxford University Press, 2010.
- [4] 宋歌. 学术创新的扩散过程研究[J]. 中国图书馆学报, 2015, 41(1):62-75. (Song Ge. The diffusion process of academic innovation[J]. Journal of Library Science in China, 2015, 41(1):62-75.)
- [5] Zhou N, Zhan X X, Lin S, et al. Information diffusion on communication networks based on Big Data analysis[J]. Electronic Library, 2017, 35(4):745-757.
- [6] Figuerola C G, Berrocal J L A. Web link-based relationships among top European universities[J]. Journal of Information Science, 2013, 39(5):629-642.
- [7] Barabasi A L. Scale-free networks: a decade and beyond[J]. Science, 2009, 325(5939):412-413.
- [8] Borgatti S P. Centrality and network flow[J]. Social Networks, 2005, 27(1):55-71.
- [9] 赵星, 李盛庆, 叶鹰. H型指数与H型测度研究[M]. 北京: 科学出版社, 2018. (Zhao Xing, Li Shengqing, Ye Ying. Studies on the H-type indices and H-type measurement[M]. Beijing: Science Press, 2018.)

- [10] Zhao S X, Ye F Y. Power-law link strength distribution in paper cocitation networks[J]. Journal of the American Society for Information Science and Technology, 2013, 64(7):1480-1489.
- [11] Zhao S X, Zhang P L, Li J, et al. Abstracting the core subnet of weighted networks based on link strengths[J]. Journal of the Association for Information Science and Technology, 2014, 65(5):984-994.
- [12] 阮光册, 夏磊. 基于词共现关系的检索结果知识关联研究[J]. 情报学报, 2017, 36(12):1247-1254. (Ruan Guangce, Xia Lei. Knowledge connection of retrieval results based on co-word analysis[J]. Journal of the China Society for Scientific and Technical Information, 2017, 36(12):1247-1254.)
- [13] Chan K, Liebowitz J. The synergy of social network analysis and knowledge mapping: a case study[J]. International Journal of Management and Decision Making, 2006, 7(1):19-35.
- [14] 赵星. 带权信息网络之计量测度研究[D]. 杭州: 浙江大学, 2014. (Zhao Xing. Informetric measures in weighted information networks[D]. Hangzhou: Zhejiang University, 2014.)
- [15] Zhao S X, Ye F Y. Exploring the directed h-degree in directed weighted networks[J]. Journal of Informetrics, 2012, 6(4): 619-630.
- [16] Portes A. Social capital: its origins and applications in modern sociology[J]. Annual Review of Sociology, 1998, 24(1):1-24.
- [17] Shen J, Bian Y J. The causal effect of social capital on income: a new analytic strategy[J]. Social Networks, 2018, 54:82-90.
- [18] Callaway D S, Newman M E J, Strogatz S H, et al. Network robustness and fragility: percolation on random graphs[J]. Physical Review Letters, 2000, 85(25):5468-5471.
- [19] Nardelli E, Proietti G, Widmayer P. Finding the most vital node of a shortest path[J]. Theoretical Computer Science, 2003, 296(1):167-177.
- [20] Schubert A. A Hirsch-type index of co-author partnership ability[J]. Scientometrics, 2012, 91(1):303-308.
- [21] Shao B, Xiao F, Zhao S X. Exploring the h-type measure and its theoretical model in the context of e-commerce[J]. Current Science, 2016, 110(12):2276-2279.
- [22] Zhai L, Yan X, Zhang G. Bi-directional h-index: a new measure of node centrality in weighted and directed networks[J]. Journal of Informetrics, 2018, 12(1):299-314.
- [23] 薛霏, 鲁特·莱兹多夫, 叶鹰. 学术评价的多变量指标探讨[J]. 中国图书馆学报, 2017, 43(4):63-73. (Xue Fei, Leydesdorff L, Ye Ying. Probing multivariate indicators for academic evaluation[J]. Journal of Library Science in China, 2017, 43(4):63-73.)
- [24] Zhao S X, Tan A M, Yu S, et al. Analyzing the research funding in physics: the perspective of production and collaboration at institution level[J]. Physica A Statistical Mechanics & Its Applications, 2018, 508:662-674.
- [25] Hirsch J E. An index to quantify an individual's scientific research output[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(46):16569-16572.
- [26] 郭秋萍, 赵静, 郭祥. 基于结构洞的人际情报网络分析[J]. 情报理论与实践, 2016, 39(3):26-31. (Guo Qiuping, Zhao Jing, Guo Xiang. Analysis of human intelligence network based on structural hole[J]. Information Studies: Theory & Application, 2016, 39(3):26-31.)

赵星 华东师范大学经济与管理学部信息管理系、华东师范大学学术评价与促进研究中心研究员。上海 200241。

(收稿日期:2018-08-09)