

科研合作网络弹性研究与实证

■ 岳增慧 方曙

【摘要】 阐述科研合作网络弹性的概念、研究意义与应用,并以全球 100 所高校在图书情报学领域所组成的科研合作网络为例,选取网络最大簇规模和网络效率作为网络弹性测度指标,讨论节点度失效、介数失效以及随机失效策略下该科研合作网络的弹性。结果显示,科研合作网络对随机节点失效具有较强的鲁棒性,其网络容错能力较强;对选择性节点失效的网络抗攻击能力较弱;网络效率相对于网络最大簇规模更适合作为科研合作网络弹性的测度指标。

【关键词】 科研合作网络 弹性 节点失效 网络最大簇规模 网络效率

【分类号】 G350.7

1 引言

随着科学技术发展的趋全球化、科学研究的复杂化以及新的研究范式的出现,科研合作日益成为科学研究的主流方式。科研合作对于科研实体间传播与交流知识、提高科研水平和质量、实现科研实体科研能力与价值的提升、推动科学技术的突破与创新等方面具有十分重要的意义。

科研合作网络是指科研实体通过合作关系而形成的网络。一个好的科研合作网络对于科研合作诸多功能的发挥起着至关重要的作用。科研合作网络的非同质拓扑结构决定了网络在发生节点或边的故障时,网络的结构和性能会发生改变,从而导致在一定时期和一定范围内,科研合作网络的正常性能遭到破坏,功能无法有效地发挥,进而影响科研工作与科技发展的进程、效率与可持续性。

为了解决上述问题,保证科研合作网络高效、持续的功能输出,尽可能缩短“故障”恢复时滞,减轻网络结构的变化给网络性能带来的影响,促进科研工作的顺利完成以及科学技术的快速、持续进步与发展,需要根据科研合作网络的实际特性与功能需求,探索科研合作网络的弹性(又称抗毁性、恢复性等)。

目前,国内外专门针对科研合作网络弹性的研究还比较少。2008年,龚玉环等构建了一个基于引文网、合著网、机构合作网的三维网络模型,对如何进行知识创新进行了初步的探讨,并提出了科研合作网络的闭锁效应与抗毁性问题^[1]。2011年,Liu Xiaofan等

对科研合作网络中核心人物突然去世的情况进行了网络演进的攻击恢复力研究,认为从长期演进过程来看,科研合作网络的社会关系网络特性使其具有较强的恢复力^[2]。同年,M. Franceschet研究了计算机科学领域的合作网络弹性,认为该领域合作网络中的“明星合作者”仅可对其个体中心网络的性能产生影响,并不能控制整体网络的性能^[3]。2012年,吕丽利用现有复杂网络抗毁性指标“效能函数/网络效率”对一个加权科研合作复杂网络进行了网络抗毁性实证研究^[4]。

本文主要阐述科研合作网络弹性的概念、研究意义与应用,并对全球 100 所高校近年来在图书情报学领域所组成的科研合作网络弹性进行实证研究。

2 科研合作网络弹性的概念、研究意义及应用

2.1 概念

科研合作网络弹性是指科研合作网络中的科研实体(节点)或合作链路(边)在发生随机失效或受到故意攻击的条件下,科研合作网络维持其效能的能力。其中,科研实体(节点)或合作链路(边)的消退/中断的几种情况主要可以归纳为表 1。

2.2 研究意义

科研合作网络是一种典型的社会网络,同时也属于复杂网络的范畴,具有复杂网络所特有的小世界特性以及无标度特性。科研合作网络本质上的非同质拓扑结构决定了网络中每个节点及每条链路的重要程度

【作者简介】 岳增慧,中国科学院国家科学图书馆成都分馆博士研究生,E-mail: yzh66123@126.com; 方曙,中国科学院国家科学图书馆成都分馆研究员,博士生导师。

收稿日期: 2013-04-15 修回日期: 2013-05-20 本文起止页码: 86-89 95 本文责任编辑: 王传清

表 1 科研实体或合作链路的消退/中断类型

类型	自然消退/中断	非自然(强制)消退/中断
科研实体(节点)	由科研实体自身因素而导致的消退(如生理原因、退休、研究方向或需求与兴趣的转移等)	非自身因素或意愿而导致退出合作网络的现象(如上级或政府的宏观调控等)
合作链路(边)	合作最终目的已实现或自身不愿意/不需要继续合作而造成的中断	由于外部环境因素,如合作成本(距离、资金、精力等)上升而中止合作或者上级政策宏观调控要求中止合作;由于节点(科研实体)消退而引起合作关系中断

是不同的,一旦网络的某个关键节点或者链路发生故障,大量与之相连的节点会失去连接方向,网络的结构和性能会发生改变,从而导致在一定时期和一定范围内,科研合作网络的正常性能遭到破坏,功能无法有效地发挥。其影响大致表现在以下方面:

- 比较典型的情况就是当某位学术权威突然退出研究,短时间内该领域其他科研工作者可能会失去方向指引,进而使得相关研究在短期内出现衰退局面^[1]。

- 处于科研合作网络枢纽位置的科研实体或合作关系的消退,会使得网络中其他实体间合作距离变大,甚至不再连通,从而导致信息交流与知识传播的时延扩大、成本升高,甚至联系中断,对科研实体间知识的传播与交流、创新思想的凝聚、科技信息的共享都将产生不利影响。

- 在大科学时代,一项科研工作可能不仅涉及庞大的数据还要涉及不同的学科领域,需要多方配合协同才能顺利完成。合作网络中的不同科研实体在整个科研工作或领域中所起的作用也不尽相同,有些科研实体是本领域的核心与灵魂,对于任务的统筹、科技关键点的突破、难题的解决与攻克起到了举足轻重的作用。一旦其由于各种原因发生消退,将可能引起科研工作进程的减缓、科研生产率的降低,科研成果的质量也难以保证;同样,不同科研实体间的合作关系价值也有差异,具有高价值(高产出、高质量、高枢纽性等)的合作关系如果发生中断,也会在一定时期内对科研产出产生负面影响,甚至会阻滞领域科学的创新、突破与发展。

- 科研实体与其他实体进行合作,可以通过交流讨论,学习新的学术思想、技能与研究方法,并内化为自身的隐形知识,在交流与合作中提高个体的科研与创新能力。网络的不稳定变化会使实体间由合作而建立的这种关系缺乏稳定性和深入性,从而导致交流学习的不充分性以及较低的可持续性,进而影响科研实体科研能力与价值的提升。但是,过分稳固、封闭的合

作交流也不利于新知识、新思想的产生,我们在研究中必须权衡二者的相互关系。

- 上述几点的最终目的都是要促进科学的快速、持续发展与创新,因此,科研合作网络中的科研实体或合作链路发生变化,科研合作网络的效能也会随之改变,对于科研工作的顺利完成以及科学技术的快速进步与发展都将产生一定的影响。

从科研合作网络的长期演进来看,由于其特有的社会网络性质(社会性),关键节点或边消退后,其他科研实体会利用自身的社会能动性进行自组织调整,从而保证形成新的具有良好科研功能的合作网络^[2]。但是,从原始网络功能的破坏到实体自组织再到新网络的形成会有一些的时滞期,这一时滞的存在减缓了科研工作的进程乃至科学进步的脚步。为了保证科研合作网络高效、持续地实现功能输出,必须及时识别网络的运行状态,并采取相应的措施(如政府干预、宏观调控等)尽可能缩短恢复时滞、减轻网络结构的变化给网络性能带来的影响,推进科研进程,提高科研效率。

如何衡量一个科研合作网络在发生节点或边的故障时,维持其网络功能在可接受范围内的能力的大小;怎样构建面对网络“攻击/故障”时,仍具有高性能保持能力的科研合作网络,这些都涉及到了科研合作网络弹性问题。为了解决上述问题,需要根据科研合作网络的实际特性与功能需求,探索科研合作网络的弹性。

2.3 应用

科研合作网络弹性的研究,对于科技状态监测与政策制定具有一定的理论意义和应用价值。通过该研究,可以通过有效的测度手段得到并及时辨别现有网络的弹性大小,一旦发现其弹性接近或低于所需的风险阈值,可马上采取相应策略进行网络的优化与调整(自组织调整+政府政策引导、干预与调控等),以保证科研合作网络的正常运作和功能的稳定发挥,进而促进科技快速、持续进步与学科发展。具体应用包括:①为科研合作网络弹性状态监测提供科学、合理、有效的测度手段;②识别弹性功能最优的科研合作网络的构成要素,为相关部门统筹组建科研合作网络提供参考;③协助科技决策部门辨识特定科研合作网络状态,识别重要的或者潜在的合作伙伴,进而对合作网络采取优化或修复措施。

- 防御性策略:加强对科研合作功能发挥起决定性作用的科研实体与合作关系的维护;引导新的科研实体的加入;组建潜在高价值合作链路。

• 修复性策略: 一旦发生网络功能障碍, 科技部门采取相应的链路重组, 实施新人才发掘与培养或其他调控策略进行网络功能快速修复。

另外, 从功能弹性角度对关键科研实体与合作关系的评估与辨识, 可以为科技人才评价提供可借鉴的参考。

3 实证研究

3.1 研究对象选择

本文根据《美国新闻及世界报道》公布的 2012 - 2013 年世界大学排行榜^[5], 选取排名前 100 位的高校近几年来在图书情报学领域所组成的科研合作网络作为研究对象。该科研合作网络中的节点为各个高校(机构)边为高校间的科研合作关系。

3.2 数据收集

利用 Web of Science 中的 SCI - EXPANDED、SSCI 网络数据库进行检索, 得到 2000 - 2013 年“Information Science & Library Science”类目中类型为“Article”的文献共 35 364 篇。将文献数据导入 Thomson Data Analyzer(简称 TDA)软件, 经过数据清洗等步骤后, 最终生成上述 100 所高校在图书情报学领域的科研合作矩阵。

考虑到网络数据库的数据更新, 为了保持数据的一致性, 笔者在 2013 年 3 月 29 日集中采集了数据。

3.3 科研合作网络弹性测度指标

3.3.1 网络最大簇规模 网络最大簇规模是指节点失效后, 网络中最大连通子图的节点数目, 反映了科研合作网络整体连通性的变化。该值越高, 说明网络中节点间的连通程度越好, 越利于科研合作实体间信息的传播与交流。

3.3.2 网络效率 网络效率为网络中任意两个节点之间距离倒数之和的平均值, 反映了信息在动态网络上的扩散和控制能力。网络效率越高, 说明网络中节点的连接程度越紧密, 信息传播越快; 反之说明节点的连接越疏松, 信息传播越慢^[4]。该值亦反映了科研合作实体间合作效率的高低。

3.4 节点失效策略

按照失效对象的不同, 科研合作网络的失效策略(又称攻击策略)可分为两种: 节点失效和边失效。本例重点对节点失效情况进行模拟讨论。按照失效方式的不同, 又可划分为随机失效和选择性失效两类。

3.4.1 随机失效 随机地移除节点或边, 反映了网络的容错能力^[6]。

3.4.2 选择性失效 按照一定的规则, 有选择地移除节点或边, 反映了网络的抗攻击能力^[6]。本例中选取了以下两种选择性失效方式^[7]。

• ID(initial degree) 攻击, 即初始度攻击, 按照初始网络的节点度大小顺序移除节点。

• IB(initial betweenness) 攻击, 即初始介数攻击, 按照初始网络的节点介数大小顺序移除节点。

3.5 数据处理与结果分析

采集到的高校科研合作矩阵是对称矩阵, 其中第 x 行、第 y 列的数据代表 x 和 y 的合作频次。利用 Matlab 编程模拟不同节点失效策略下, 该高校科研合作网络的连通性(网络最大簇规模)和网络效率的性能变化。

3.5.1 基于网络最大簇规模的科研合作网络弹性分析 节点度失效、介数失效和随机失效三种不同策略下, 高校科研合作网络失效节点数与网络最大簇规模的关系变化如图 1 所示:

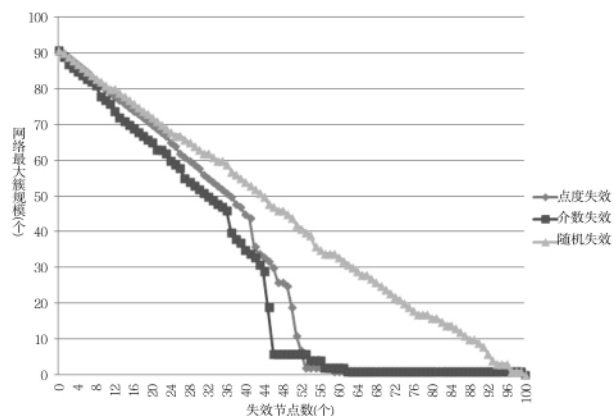


图 1 科研合作网络的网络最大簇规模变化

由图 1 可看出, 随机失效策略下, 网络最大簇规模变化最缓慢, 呈平缓下降趋势, 说明节点随机失效对科研合作网络连通性(网络最大簇规模)影响最小, 即科研合作网络对随机节点失效具有较强的鲁棒性, 其网络容错能力较大。点度失效策略下, 网络最大簇规模变化较为急促, 当移除网络中一半的节点时, 该科研合作网络中高校之间几乎不再连通, 说明节点度失效对科研合作网络连通性(网络最大簇规模)影响较大, 网络抗攻击能力较小。介数失效策略下, 网络最大簇规模变化最快, 当移除网络中 36% - 44% 的节点时, 网络最大连通子图中的节点数迅速下降, 说明介数点度失效对科研合作网络连通性(网络最大簇规模)影响最大, 网络抗攻击能力最弱。

3.5.2 基于网络效率的科研合作网络弹性分析 节

点点度失效、介数失效和随机失效三种不同策略下,高校科研合作网络失效节点数与网络效率的关系变化如图2所示:

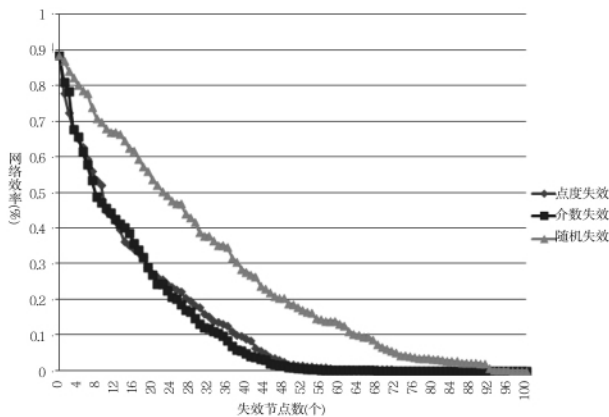


图2 科研合作网络的效率变化

由图2可看出,随机失效策略下,网络效率呈缓慢下降趋势,说明节点随机失效对科研合作网络效率影响最小,即科研合作网络对随机节点失效具有较强的鲁棒性,其网络容错能力较强。点度失效和介数失效策略下,网络效率变化都较为急促,当移除网络中8%的节点时,该科研合作网络的效率损失迅速高达50%;当移除网络中46%的节点时,其网络效率下降为0,说明这两种节点失效策略都会对科研合作网络的效率产生很大的影响,其网络抗攻击能力明显降低。

结合图1和图2可以看出,在相同的节点失效策略下,科研合作网络的效率比网络连通性(网络最大簇规模)的性能变化率更快、更加明显,说明网络效率相对于网络最大簇规模更能敏锐、准确地体现科研合作网络弹性性能,即网络效率更适合作为科研合作网络弹性的测度指标。

4 结论

本文通过对高校科研合作网络弹性进行实证研究,得到以下结论:①科研合作网络对随机节点失效具有较强的鲁棒性,其网络容错能力较强;②科研合作网络对选择性节点失效的网络抗攻击能力较弱;③网络效率相对于网络最大簇规模更适合作为科研合作网络弹性的测度指标。

在进一步的研究中需要注意以下几点:①以上关于科研合作网络弹性的结论都是建立在100所高校在图书情报学领域相互合作基础上的,其可推广性还需要进一步的讨论,因此需要扩展研究领域;扩大样本规模,避免样本统计随机性过大而给分析结果带来影响;

进行定性分析,补充和完善定量分析结果。②本文主要考察了在“一定时期和一定范围之内”的科研合作静态网络上的弹性变化,然而由科研合作网络的自组织修复能力而导致的网络动态变化是影响科研合作网络弹性的一个关键性因素,因此,需要深入探讨科研合作动态网络中的弹性问题,力求全面准确地反映科研合作网络在不同环境参数下的弹性性能。③现有的网络弹性测度指标与算法大都针对复杂网络这一具有共性的网络对象进行一般化、普适化的研究,但是,由于不同性质的网络(如科研合作网络、生物网络、交通网络、通信网络等)所具有的拓扑特性以及所承担的功能是不同的,在开展网络弹性研究时,针对不同的研究对象,所采用的方法和策略也应不同。因此,结合科研合作网络的自身特性,有针对性地设计出能准确反映科研合作网络需求的弹性测度指标十分必要。④在大力促进科研合作网络发展的过程中,需要建立网络弹性的风险预警机制。具体地说,就是根据不同情况,权衡科研合作网络的实际结构及功能需求,结合科研合作网络弹性测度指标,构建网络弹性的辨识参数(通常以科研合作网络弹性度量阈值或区间的形式表达)和相应的网络调整策略。加强对科研合作网络弹性具有重大影响的节点(科研实体)和链路(合作关系)的关注和维护,制定实际科研合作网络高弹性优化策略(防御策略)以及针对各类故障的最优修复策略,引导科研合作网络的健康发展,保证科研合作网络的正常运作及功能的稳定发挥。

参考文献:

- [1] 龚玉环,卜琳华. 科研合作复杂网络及其创新能力分析[J]. 科学管理研究, 2008(12): 30-32, 36.
- [2] Liu Xiaofan, Xu Xiaoke, Small M, et al. Attack resilience of the evolving scientific collaboration network [J]. PLoS ONE, 2011, 6(10): 1-7.
- [3] Franceschet M. Collaboration in computer science: A network science approach [J]. Journal of the American Society for Information Science and Technology, 2011, 62(10): 1992-2012.
- [4] 吕丽. 一个科研合作复杂网络模型的相关实证研究[D]. 广州:暨南大学, 2012.
- [5] 2012/2013年世界大学排名 U. S. News (1-100) [EB/OL]. [2013-03-29]. <http://www.university-list.net/rank1.htm>.
- [6] 吴俊,谭跃进. 复杂网络抗毁性测度研究[J]. 系统工程学报, 2005, 20(2): 128-131.
- [7] Holme P, Kim B J, Yoon C N, et al. Attack vulnerability of complex networks [J]. Physical Review E, 2002, 65(5): 1-14.

(下转第95页)

程,解释了实证研究中的主要步骤和内容。通过领域选择的理解和把握,选择了12个涉及基础、应用和技术领域且包括同域和异域领域作为实证领域。在数据处理和软件编写上,笔者对数据处理过程和基本原则进行了详细的描述。在学科的分类和数据元素的归属这个关键问题上,重点说明了将文章、作者和引文归属到一级学科的原因。本文后半段说明了各种数据表格的结果和获得机制,其中对引文类共现矩阵作了重点阐述,因为该矩阵是所有表格中的核心,也是进一步分析和研究的核心。本文初步获得了各领域经程序运算的数据表格和结果,为下一步进行指标计算、分析和研究学科交叉现象做好了充足的准备。

本文关注对跨学科现象的测度,文献计量学方法正好契合了这种测度要求,所以利用文献计量学研究跨学科现象具有一定的合理性和优势。本文是前期的数据集构建过程与分析,后续的研究策略是:在构建跨

学科测度指标的基础上,从验证其有效性、分析跨学科若干模式以及分析作者合作对跨学科研究的影响等观察角度,对多学科度、专业度、学科交叉度和合作度等计量指标进行计算、分析与验证,进而探索跨学科研究的本质和规律。

参考文献:

- [1] 杨良斌,金碧辉. 跨学科测度指标体系的构建研究[J]. 情报杂志, 2009, 28(7): 65-69.
- [2] 杨良斌,周秋菊,金碧辉. 基于文献计量的跨学科测度及实证研究[J]. 图书情报工作, 2009, 53(10): 87-90.
- [3] Bo J. A comparison of two bibliometric methods for mapping of the research front[J]. Scientometrics, 2005, 65(2): 245-263.
- [4] 陈强,廖开际,奚建清. 知识地图研究现状与展望[J]. 情报杂志, 2006, 25(5): 43-46.
- [5] [EB/OL]. [2013-01-16]. <http://portal.isiknowledge.com/portal.cgi?DestApp=ESI&Func=Frame>, 2009-04-04.

Data Set Construction and Data Preparation of Interdisciplinary Measure Indices

Yang Liangbin

Department of Information Technology, University of International Relations, Beijing 100091

[Abstract] Choosing twelve empirical fields, which include foundation, application and technology fields and comprises homology and difference between fields, this paper describes the construction process and basic principles of data set in detail. Then it expounds the reasons why articles, authors and citations belong to first-level discipline, and discusses the co-occurrence matrix of citation category. Finally, it gets the data forms and results of each field and well prepares for calculating indexes, analyzing and researching the phenomenon that three disciplines cross in the future.

[Keywords] bibliometrics interdisciplinary multi-disciplinarity interdisciplinarity measure indices

(上接第89页)

Resilience of Scientific Collaboration Network: An Empirical Research

Yue Zenghui^{1,2} Fang Shu¹

¹Chengdu Branch of the National Science Library of Chinese Academy of Sciences, Chengdu 610041

²University of Chinese Academy of Sciences, Beijing 100190

[Abstract] The paper expounds the concept, research significance, and application of scientific collaboration network resilience, and takes a scientific collaboration network composed of 100 selected universities in the field of LIS as an example. Size of giant component and network efficiency are further selected as the metrics of network resilience to explore the scientific collaboration network resilience according to the node removal strategies of degree centrality, betweenness centrality and randomized ones. Results from the case study suggest that the scientific collaboration network has good robustness and strong fault tolerant ability for the random failure, while the anti-attack capability of the network for the selective failure is poor, and network efficiency is a better measurement index for the scientific collaboration network resilience than size of giant component.

[Keywords] scientific collaboration network resilience node failure size of giant component network efficiency