

doi:10.3772/j.issn.2095-915x.2015.06.012

学术网络中科研人员影响力分析方法研究

刘京旋, 杜永萍, 杜晓燕, 付利华
(北京工业大学计算机学院 北京 100124)

摘要: 随着科研合作的多元化, 评判科研人员的影响力也将受到影响, 本文在由科研人员及其所发表的学术论文构成的网络中, 利用关联信息进行科研人员的影响力分析。在基于科研论文质量计算初始影响力的基础上, 融入科研人员之间的合作关系, 包括直接影响与间接影响, 设计算法实现综合的科研人员影响力计算。在真实实验数据集上进行影响力排序, 对不同研究领域的科研人员分析其合作网络以及对其影响力计算的作用, 进一步验证了方法的有效性。

关键字: 学术影响力, 学术网络, 排序

分类号: G2

Evaluation Method on the Influence of Researchers in the Academic Network

Liu JinXuan, Du Yongping, Du Xiaoyan, Fu LiHua

(Institute of Computer Science, Beijing University of Technology, Beijing 100124)

Abstract: With the development of cooperation in the research field, the evaluation to the researcher's influence will also be affected. This paper analyzes the influence of researcher by using the correlation information in the academic network, which is composed of researchers and their published academic papers. Based on the initial influence of academic paper quality, we integrate the cooperative relationship between researchers, including direct influence and indirect influence, and design the algorithm to implement the final influence evaluation of researchers. The experiment is implemented on the real data set to verify the effectiveness of the method. The cooperation network is constructed in different research field and the ranking of the influence is also demonstrated.

Key word: Academic influence, academic network, ranking

基金支持: 国家科技支撑计划子课题(2013BAH21B02-01), 北京市自然科学基金资助项目(4153058), ISTIC-THOMSON 科学计量学联合实验室开放基金。

作者简介: 刘京旋(1992-), 女, 硕士研究生, 研究方向自然语言处理。杜永萍(1977-), 女, 副教授, 研究方向信息检索, 自然语言处理。杜晓燕(1990-), 女, 硕士研究生, 研究方向自然语言处理。付利华(1976-), 女, 副教授, 研究方向: 智能信息处理。收稿时间: 2015年10月10日。

1 引言

随着互联网快速发展,从海量的科技文献数据中筛选出对用户有价值的文献是一项有意义的工作。如果在每篇科研论文的参考信息中有科研人员及其发表论文的影响力值,我们就能快速的筛选出质量相对比较高的论文,这将是一件非常有意义的工作。

学术网络中科研人员影响力的分析方法至关重要,有研究根据网络发表科技论文学术影响力评价指标权值,来进行科研人员的论文影响力的评测^{[1][2]}。在本文中,研究科研人员权威值主要根据科研人员所发表的学术论文的质量、数量,和与科研人员有过合作关系的科研工作者的权威值而决定。评判科研人员的学术论文的质量由论文引用次数、下载量、论文发表所在期刊的影响力等方面决定。

一些已有的研究通过科研合作网络中边的权重反映科研人员的科研产出能力,通过计算边权值来评定作者影响力^[3],在本文中建立作者合作网络图,重点研究节点与节点的关系,从而实现评定作者影响力。科研人员之间的“引用影响力+合作影响力”^[4]对于评价科研人员权威值是一个很好的依据。在本文中,科研人员合作关系对科研人员的权威值分为两级,直接合作关系决定直接影响,间接合作关系决定间接影响。

本论文第2节介绍基于科研论文质量的影响力计算;第3节将科研人员的合作关系融入计算模型;第4节是在数据集上的方法验证,分析了影响力计算方法的有效性。最后是结论。

2 基于科研论文质量的学术网络用户影响力计算

科研人员的工作通常体现在其发表的学术论文当中,因此论文的质量对其影响力评价起关键

作用。在传统出版环境下,纸质的学术出版物的影响力主要是通过引文索引和同行推荐评议来实现的^[5],前者具有局限性,后者具有个人主观性;且传播速度受传统出版模式影响。目前,互联网逐渐替代传统媒介方式,成为传播科研知识的主流媒体。因此,如何采用合适的方法评价科研人员所发表的学术论文质量成为重要的研究方向。

通常情况下,学术论文影响力可从论文网络下载次数、论文被引用次数、论文发表期刊的权威性这几个方面考虑。

被引频次和同行评审二者都是用于学术影响力评价的最具代表性的指标^[6],学术论文的被引用次数、下载次数越多,说明该论文受关注的程度越高,它对本学科及其相关领域的影响也就越大,其学术影响力越大,也在一定程度上体现了该论文的学术价值。因此,就论文引用次数和论文网络下载情况这两项可以明确的从统计数字上直观的反映论文在科学发展和文献交流中所起的作用,从历史的角度反映了论文的学术影响力。

期刊因子表明论文发表所在的期刊在行业中的权威性,权威性高的期刊上刊登的论文通常质量较高,因此,期刊因子成为评判论文影响力的二级指标。

论文影响力因素指标级别分类如表1所示。

表1 论文影响力因素指标级别分类

指标级别	指标
一级指标	论文转载量
	论文引用次数
	论文网络下载量
	论文获奖情况
二级指标	期刊因子
	论文基金资助

本文工作采用了一级指标中论文网络下载次数和论文引用次数这两个衡量标准和二级指标中期刊因子衡量标准。另外,基于H指数的科研人

员学术评价^[7]中讨论引用次数和下载次数超过数值 H 时论文的影响力对作者权威性起作用。因此,参数的设定也对评论作者影响力合理性起了重要作用。本文设定了三个参数来评定科研人员学术论文影响力,并根据科研人员所发表的论文的影响力来给科研人员设定一个初始影响力值。

2.1 数据集

我们采用的针对研究科研人员影响力计算的数据集包含 1015 条文献数据,如表 2 所示。每一条数据是以论文为节点,包含论文 id、标题、关键字、所有作者、发表所在期刊、发表年限、引用次数以及下载次数等论文相关信息。

表 2 影响力评价初步模型的数据类型样例

论文 Id	论文题目	关键字	作者	发表期刊	年份	引用次数	下载次数
1	基于汉明距离的文本相似度计算	文本分类;信息检索;汉明距离	张焕炯;王国胜;钟义信	计算机工程与应用	2001	70	604
2	一个基于向量空间模型的中文文本自动分类系统	文本分类;中文文本;向量空间模型;特征提取	朱华宇;孙正兴;张福炎	计算机工程	2001	121	380

2.2 作者影响力计算方法

对于科研人员而言,发表的论文质量高,在行业中认知度和权威性高。我们将科研人员所发表的论文质量作为研究科研人员影响力的基本指标。

根据论文引用次数 R 、论文下载量 S 、论文发表所在期刊的期刊因子 T 设定三个权重参数评价一篇论文的影响力,分别为 α 、 β 、 γ 。

论文影响力计算如公式(1)所示:

$$\text{Influence}(\text{paper}_i) = \alpha * R_i + \beta * S_i + \gamma * T_i \quad (1)$$

其中, $\alpha + \beta + \gamma = 1$ 。

根据论文影响力,计算作者初始影响力,如公式(2):

$$\text{InitialInflu}(\text{Author}_j) = \sum_{i \in I} \text{Influence}(\text{paper}_i) \quad (2)$$

其中, I 表示 Author_j 发表的所有论文的集合。

3 融入合作关系的学术网络用户影响力计算改进模型

在评价学术网络用户影响力基本模型的基础上,我们将科研人员合作关系融入影响力的计算,进一步客观评价。

将 2.1 节的数据集整理为以科研人员为节点的数据集,共有 2081 个科研人员节点,将合作过的科研人员或有间接合作的科研人员进行合作网络归类,并将合作网络编码。

数据样例如表 3 所示。

表 3 以科研人员为节点的评价模型数据样例

作者 id	作者姓名	作者影响力	作者发表论文 id	作者的合作伙伴 id
0	张焕炯	0	1、32	1、2、88、2
1	王国胜	0	1	0、2
2	钟义信	0	1、9、32	0、1、21、22、0、88

科研人员的合作伙伴对其影响分为两个级别，直接影响和间接影响，直接影响即为直接与其有合作关系的科研人员对他的影响，间接影响为与他的合作伙伴有合作关系的科研人员（与他本人无直接合作关系）对他的影响。

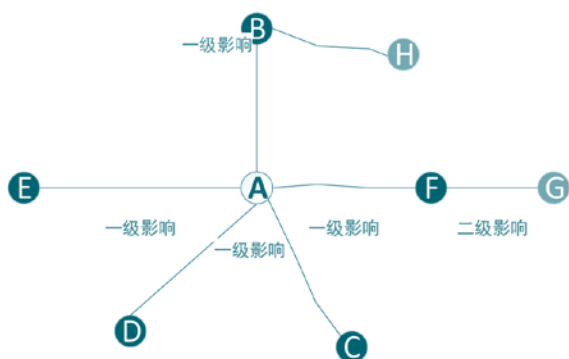


图1 科研人员影响力关系图

在图1中，对于A节点（科研人员A）而言，B、C、D、E、F都是跟其有直接合作关系的，对A属直接影响，影响参数设为 θ ；G与F有直接合作关系，与A处于间接合作状态；同样，H与A处于间接合作状态。G和H对A为间接影响，影响参数设 ϕ 。其中 $\theta > \phi$ ，并且 $\theta + \phi = 1$ 。

合作影响力计算如公式（3）。

$$\begin{aligned} \text{CoopInflu}(\text{Author}_j) &= \theta * \text{DirectInflu}(\text{Author}_j) + \phi * \text{IndirectInflu}(\text{Author}_j) \\ &= \theta * \sum_{m \in M} \text{DirectInflu}(\text{Author}_m) + \phi * \sum_{N \in M} \text{IndirectInflu}(\text{Author}_n) \end{aligned} \quad (3)$$

其中，M表示与 Author_j 有直接合作关系的科研人员的集合；N表示与 Author_j 有间接合作关系的科研人员的集合。

对A作者而言，其影响力计算表示为：

$$\begin{aligned} \text{CoopInflu}(\text{Author}_A) &= \theta * [\text{DirectInflu}(\text{Author}_B) + \text{DirectInflu}(\text{Author}_C) + \text{DirectInflu}(\text{Author}_D) \\ &+ \text{DirectInflu}(\text{Author}_E) + \text{DirectInflu}(\text{Author}_F)] \\ &+ \phi * [\text{IndirectInflu}(\text{Author}_G) + \text{IndirectInflu}(\text{Author}_H)] \end{aligned}$$

科研人员最终影响力由初始影响力与合作影响力构成，计算如公式（4）。

$$\text{Influ}(\text{Author}_j) = A * \text{InitialInflu}(\text{Author}_j) + B * \text{CoopInflu}(\text{Author}_j) \quad (4)$$

其中， $A+B=1$ 。

4 实验结果与分析

4.1 总体影响力排序

实验中用到的数据来自中国知网，年限跨度为2001年到2005年，涉及期刊共43种，包含1015篇论文数据。将数据集中的2081个科研人员节点数据进行影响力计算并进行排序，选取排名前10的科研人员进行数据分析。

图2为此数据集中排名前十的科研人员影响力值柱状图；图3为发表论文数量柱状图；图4为科研人员直接合作伙伴数量。

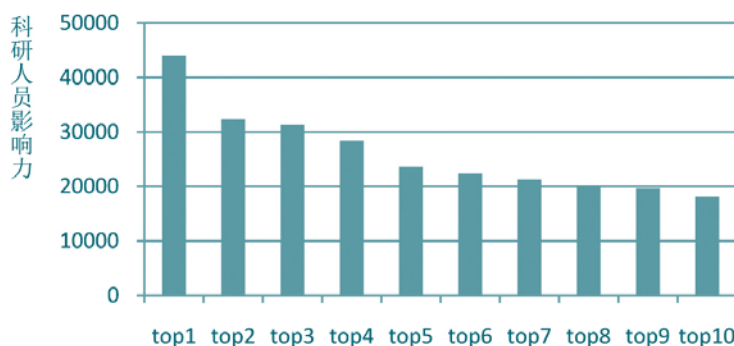


图2 Top10的科研人员影响力

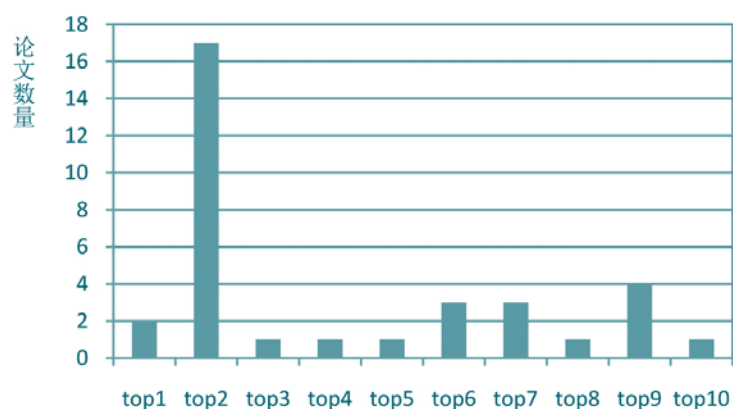


图 3 Top10 的科研人员发表论文数量

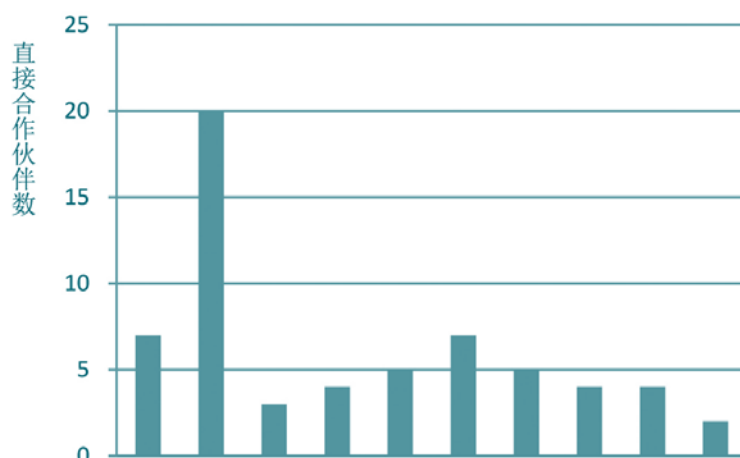


图 4 Top10 的科研人员的直接合作伙伴数量

由图 2- 图 4 分析可知，科研人员影响力受到不同因素的作用：

(1) 科研人员所发表的论文质量起着关键作用。如 Top 1 的科研人员的发表的论文数量和合作伙伴数量均低于 Top 2 的科研人员，但 Top 1 科研人员影响力最高是由于其发表论文的质量高，提升了影响力。

(2) 科研人员的所发表的论文数量对影响力排序作用较大。如 Top2 的科研人员发表的论文数量较多，他的影响力也得到提升。

(3) 科研人员的合作伙伴对其影响力计算也具有重要的作用，如 Top 6 科研人员与 Top 7 科研人员发表论文数量相当，但前者合作伙伴数量高于后者，使得影响力得到了进一步提升。

4.2 合作关系网络

科研人员之间具有直接或间接的合作关系，图 5 展示了在数据集上的合作关系网络，节点代表科研人员，节点之间连线代表科研人员之间具有合作关系。

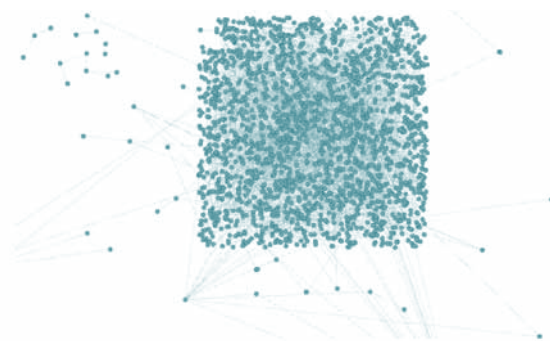


图 5 作者合作关系网络图

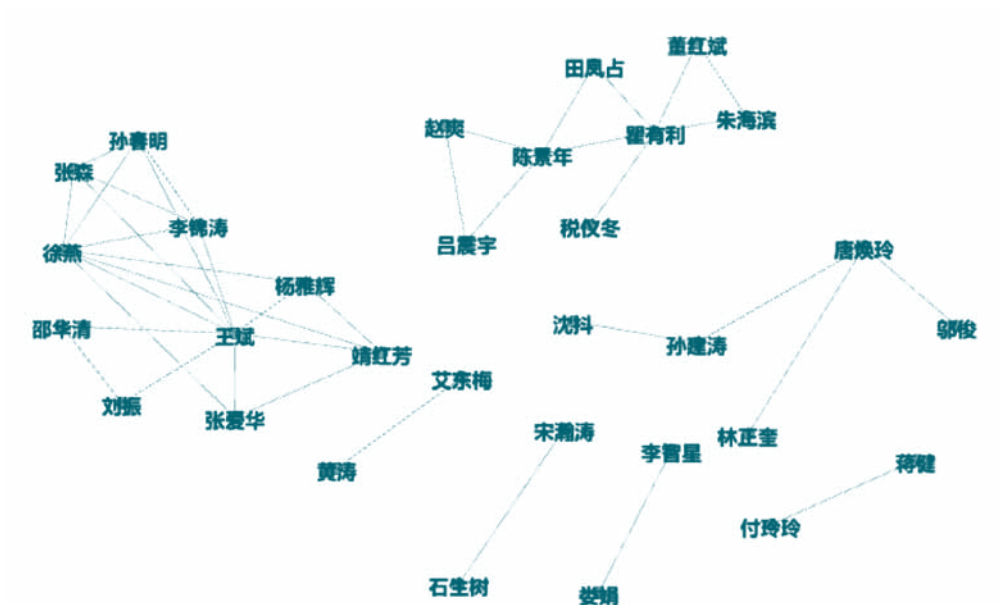


图6 核心作者合作关系网络图



图7 数据挖掘领域的作者合作关系网络

同时，选取影响力计算值较高的科研人员的合作网络进行展现，如图6所示。

不同研究领域的科研人员均构成各自的学术合作网络，图7所示为在数据挖掘方向上的合作网络。（其中标深色方块的节点为影响力较高的科研人员节点）

5 结束语

本文对科研人员影响力的评价进行了研究，论文被引次数、论文下载量、期刊因子等因素对评价论文质量起决定性作用，在对科研人员发表论文质量进行评价的基础上，引入合作关系改进

对影响力的计算。在真实数据集上,构建不同领域的学术合作网络,根据直接影响与间接影响,设计影响力计算模型,进行影响力排序,对实验结果进行分析,进一步验证了方法的有效性。

参考文献:

- [1] 李贺,周金娉,李春好.我国网络发表科技论文的学术影响力评价研究[J].情报理论与实践,2012(02):77-81.
- [2] 曹兴、周密、刘芳.网络科技论文学术影响力评价指标体系研究[J].科学决策,2010(9):19-30.
- [3] 肖连杰,吴江宁,宣照国.科研合作网中节点重

要性评价方法及实证研究[J].科学学与科学技术管理,2010(6):12-15.

[4] 杜建,张玢,唐小利.基于作者引用与合作关系的学术影响力测度研究进展[J].图书情报工作,2013(08):135-140.

[5] 袁曦临,常娥.网络发表学术论文的影响力评价研究[J].图书情报工作,2011(10):51-54.

[6] 沈小玲,徐勇,严卫中.基于网络文献计量的科技论文学术影响力综合评价研究[J].图书情报工作,2013(21):95-103.

[7] 君燕,洪伟,吴承祯等.基于H指数的科研人员学术评价及其相关性分析[J].福建农林大学学报(哲学社会科学版),2012(03):58-63.