利用 k-shell 分析合著网络中的作者传播影响力*

张金柱

（1.中国科学院国家科学图书馆，北京 100190；2.中国科学院研究生院，北京 100049）

[摘要]作者的传播影响力一般以度和介数来计算，而 k-shell 值表示作者在网络中所处的位置，能更好的表征影响力，结论颇具颠覆性。本文以情报学领域的 12 种期刊在 2000-2009 年间的 7389 位作者形成的合著网络为例，分别基于度和 k-shell，介数和 k-shell 对作者传播影响力进行比较分析。结果表明，k-shell 值较度、介数能更好的表征作者的传播影响力。这种方法希望可以推广到基于科技文献数据的其它网络中，如识别文献共被引网络、文献耦合网络中最具传播影响力的关键文献。

[关键词]传播影响力，合著网络，k-shell，度，介数

[分类号]G353

Influential Spreaders in Co-author Network Based on K-shell

Zhang Jinzhu

（1. National Science Library， Chinese Academy of Sciences， Beijing 100190; 2. Graduate University of the Chinese Academy of Sciences, Beijing, 100049）

[Abstract] Influential spreaders in co-author network are computed often by degree or betweenness centrality, but k-shell indicates the location in network and performs better. The paper’s data comes from 12 journals between 2000 and 2009 which contains 7,405 different authors, then computes the degree, betweenness centrality and k-shell and does comparative analysis. The results show that k-shell does better in identification of influential spreaders in ca-author network. This method could be also used in co-citation network and coupling network for identification of influential spreaders.

[Keywords] influential spreaders, co-author network, k-shell, degree, betweenness centrality

1 引言

在复杂网络的研究中，针对节点的传播影响力评估做了大量的工作，如科学家合作网络中的关键人物、通信网络的核心交换机、Web 中的权威页面等具有特定含义的重要节点[1,2]。因此，通过对节点的传播影响力评估来找出重要的“关键节点”将是一项非常有意义的工作，如发现科学家合作网络中的核心人物，重建基于重要节点的网络，阻止网络病毒的传播及扩散等等。

*本文系国家自然科学基金面上项目“科学结构特征及其演化动力学分析方法与应用研究”（项目编号：71173211）的研究成果之一。

合著网络作为复杂网络的一种表现形式，在合著网络中发现具有重大传播影响力的作者对学科建设、科研合作与评价、信息传播都起着重要作用，其中，节点表示作者，边表示相连接的两个作者合作发表过文章。作者的传播影响力以度（degree）和介数（betweenness centrality）来计算是最普遍的，度高的节点拥有更多的合作关系，而介数高的节点则能使更多的作者产生关联，有更多的最短路径通过[3]。k-shell 值则表示作者在网络中所处的位置，能更好的表征传播影响力，度、介数高的作者可能处于整个网络的边缘位置，而非中心位置，其传播影响力较低；反之，度、介数较低的节点也可能处于网络的中心位置，其传播影响力较高[4]。

本文拟以 2000 年到 2009 年的情报学领域的 12 种期刊数据为基础，对度、介数、k-shell 值三者进行综合分析，分别基于度和 k-shell，介数和 k-shell 来测度合著网络中的作者传播影响力，并对其中的典型案例进行分析，包括度大而 k-shell 值较小、介数大而 k-shell 值较小、k-shell 值最大时度和介数的表现情况，最终验证 k-shell 值较度、介数能更好的表示作者的传播影响力。这种方法希望可以推广到基于科技文献数据的其它网络中，如识别文献共被引网络、文献耦合网络中最具传播影响力的关键文献。

2 数据和方法

本文选择情报学领域的 12 种期刊作为数据源，跟踪情报学的发展和演化情况[5-7]。本文选取的时间区间为 2000 年到 2009 年，数据集涵盖了 117 个国家，2325 个研究机构，7,405 位不同作者撰写的 8,374 篇论文[8]。经过去重后的作者数为 7389，由此形成合著网络。去重主要思路是：以作者全称作为作者去重的标准，具有相同作者全称，但作者缩写不同的被认为是同一作者，由于数据中并不是每位作者均有全称，因此可能去重还存在遗漏。在进行替换时，以作者缩写较长的名称替换较短的名称。去重的详细信息如表 1 所示。

表 1 作者去重结果说明

被替换作者	新作者	替换次数	被替换作者	新作者	替换次数
Guan, J	Guan, JC	2	Ju, B	Ju, BY	4
Chen, H	Chen, HC	11	Yan, E	Yan, EJ	3
Wang, C	Wang, CN	1	Fry, J	Fry, F	6
Peritz, B	Peritz, BC	1	Wang, P	Wang, PL	1
Ou, S	Ou, SY	2	Dhawan, S	Dhawan, SM	7
Gupta, B	Gupta, BM	8	Zhang, W	Zhang, WD	2
Schlogl, C	Schloegl, C	2	Kim, S	Kim, SJ	8
Ke, W	Ke, WM	1	Aoe, J	Aoe, JI	10

复杂网络中一般使用疾病传播模型来模拟疾病在网络上的传播动力学，包括节点的传播范围、能力、速度[9-11]。一般认为，网络上传播能力强的为度、介数高的节点，因为度高的结点有更多的邻居节点，合作范围更广，容易造成更大的传播范围[1, 2, 12]；而介数高的节点表示通过其的最短路径数较多，某位作者控制网

络中其他作者之间交往的能力较强，与其他人的交流更广泛[13-16]。k-shell 考虑的则是节点在网络中所处的位置，度、介数高的节点未必处于此网络的中心位置，而可能处于网络的边缘位置。一般来说，越靠近中心位置的节点其传播影响力更大，而边缘位置的节点其传播影响力相对低[4, 17]。度考虑的是节点本身的局部性质，介数考虑的是节点本身的全局性质，而 k-shell 值不仅考虑了节点本身的特性，也充分考虑了该节点各阶邻居节点的特性，尤其是考虑了该节点和哪些节点相连，信息更加丰富。

Kitsak 等[4]基于 SIR（Susceptible Infected Recovered）模型和 SIS（Susceptible Infected Susceptible）模型对四种网络进行了建模分析，这四种网络包括：LiveJournal.com 网站上 340 万市民的友谊网络；伦敦大学计算机科学系的电子邮件联系网络；瑞典的医院住院病人的接触网络；由 imdb.com 标记的同一电影中有合作关系的演员网络。结果表明：对于单个传播源情形，度高或介数高的节点不一定是最具传播影响力的节点，而通过 k-shell 分解分析确定的网络核心节点（即 k-shell 值大的节点）才是最具传播影响力的节点。度或者介数高的节点不一定是最具传播影响力的节点的原因在于，如果它们位于整个网络的边缘位置，那么它们在传播中的作用就非常微弱。而某些度、介数值较低却位于网络核心位置的节点将对传播过程产生重大的影响。

k-shell 是图论里的一个经典的概念，网络的外壳和边缘的 k-shell 为 1，然后往内像剥洋葱一样进入网络的核心（k-shell 值大的区域）。如图 1（a）中的黄色节点拥有较高的度，却处于边缘位置。k-shell 的计算过程也较为简便：首先找出所有度为 1 的节点置于第一层并剔除这些节点，即 k-shell=1，在余下的节点中继续寻找度为 1 的节点置于第一层并剔除，直到没有度为 1 的节点，如图 1（b）所示，k-shell=2 或 3 的节点计算与此相同，分别如图 1（c）、图 1（d）所示。

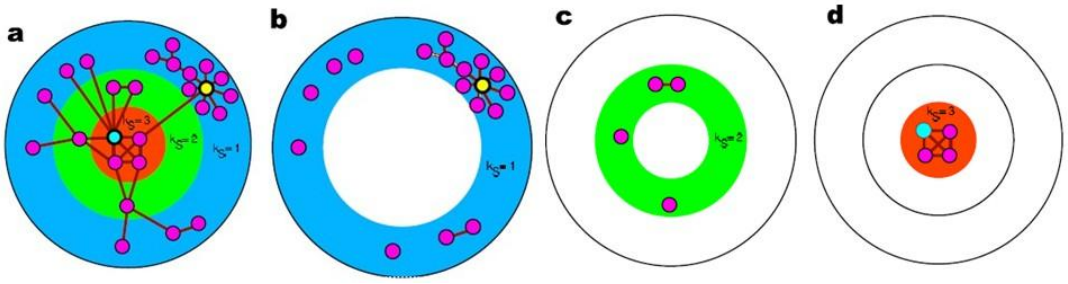

图 1 （a）中黄色节点度较大，但处于 k-sehll 网络的边缘位置；（b）k-shell 值为 1 的节点，处于最外层；（c）k-shell 值为 2 的节点；（d）k-shell 值为 3 的节点，处于最中心位置[4]。

3 结果分析

为了验证 k-shell 较度、介数能更好的表征作者的传播影响力，因此在合著网络中分别计算三种指标并进行比较分析，主要包括：基于度和 k-shell 的作者传播影响力比较分析、基于介数和 k-shell 的作者传播影响力比较分析。在比较分析中，分别对总体情况进行说明，然后选择其中的特例情况进行解释，如度高但 k-shell 值较小的作者其传播影响力分析、介数高但 k-shell 较小的作者其传播影响力分析、k-shell 值最大的作者其传播影响力分析。结果表明，传播影响力最强的作者

其合作者数量处于中等水平，合作者传播影响力较强，合作者均处于网络的较中心位置，作者与合作者共同形成了相应的团体，科研交流密切而广泛，可能是未来图情领域的新兴力量代表。而图情领域的专家间合作力度明显不够，造成他们的传播影响力较低。

3.1基于度和 k-shell 的作者传播影响力比较分析

3.1.1 总体情况分析

图 2 中横轴表示 k-shell 值，处于同一竖线的作者具有相同的 k-shell 值，纵轴的度表示作者的合作者数目。颜色较深的节点表示多个节点处于同一坐标，重合作者数较多，如节点 1，2 分别为 15 个和 21 个节点重合。

图 2 中处于左上方较为稀疏的节点为度较大，而 k-shell 值较小的作者，如 k-shell 值为 4 时，Oppenheim, C 的合作者数为 36，Jarvelin, K 的合作者数为 28。k-shell 值为 5 时，Rousseau, R（鲁索）的合作者数达到 41，Chen, HC 的合作数为 40，Marchionini, G Thelwall, M 的合作者数为 29，值得注意的是，这些节点的 k-shell 值和度值仅为 5 的作者（共 197 位）相同，在网络中处于第 5 层的较边缘位置，造成它们在整个网络上的传播能力较弱。Klingsporn, B 的合作者数为 26，Stock, Wg 的合作者数为 25，k-shell 值同为 22，为 k-shell 值最大且度也较大的节点，而 Marchionini, G 和 Kostoff, Rn 的合作数均为 27，而 k-shell 值仅分别为 8 和 9。节点 2 包含了 21 个度为 22 的节点，k-shell 值也是最大的。

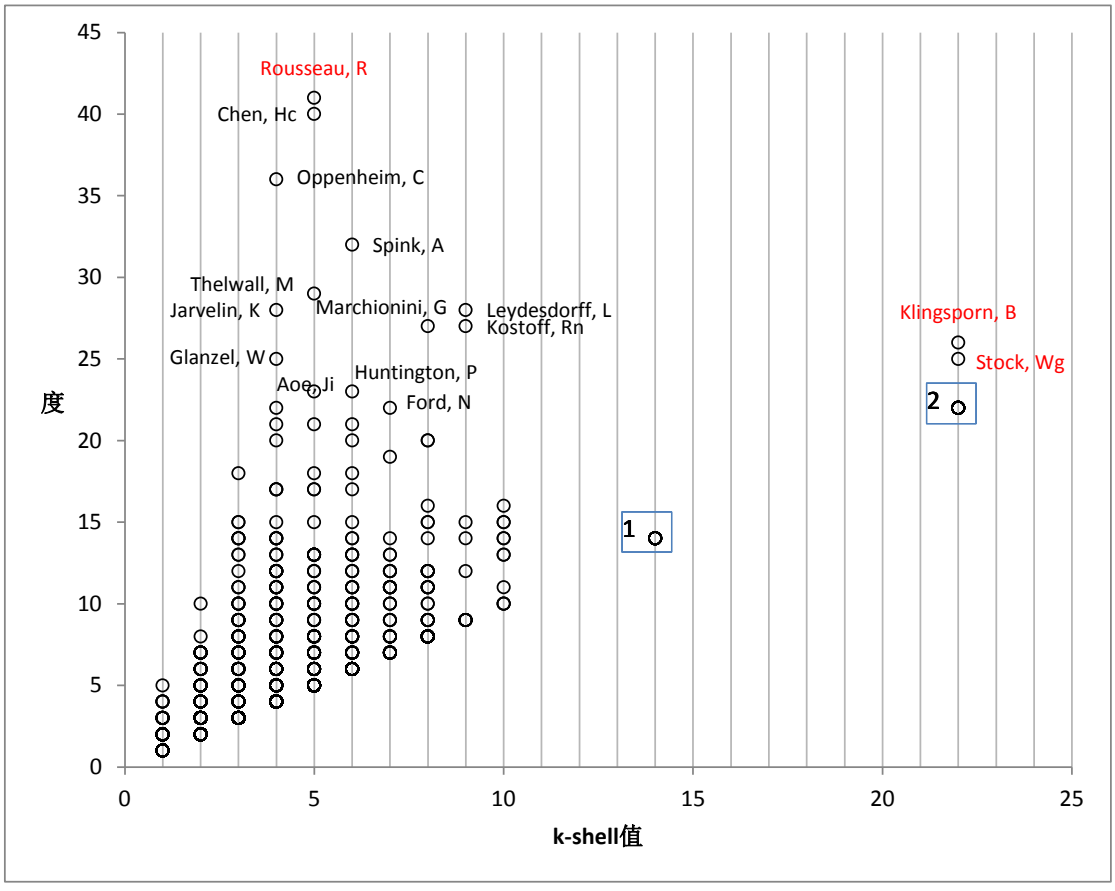

图 2 基于度和 k-shell 的作者影响力比较分析

3.1.2 度高但 k-shell 值较小的作者传播影响力分析

Rousseau, R 作为度高但 k-shell 值较小的典型代表，也为我们所熟知，应是传播影响力较高的作者，而分析结果却另人意外。选取 Rousseau, R 的 41 位合作者说明，如表 2 所示，其合作者的度普遍偏小，小于或等于 5 的占到 27/41，小于 10 的占到 34/41，使得 Rousseau, R 的 k-shell 值较小，处于网络的较边缘位置，使其传播影响力降低。而提高 Rousseau, R 的传播影响力的方式是其合作者的传播影响力得到提到，特别是小于或等于 5 的作者。因此，K-shell 值能更好的表征传播影响力，符合实际情况，Rousseau, R 的 k-shell 值为 5，处于网络的较边缘位置，其合作者的 k-shell 值也普遍偏小，其传播影响力较小。当然，Rousseau, R 作为图情领域的重要学者之一，此处测度的传播影响力仅在特定数据集上部分评估其工作和研究。

表 2 Rousseau, R 的合作者的度和 k-shell 值

合作者	度	k	合作者	度	k	合作者	度	k
Glanzel, W	25	4	Wang, Y	5	4	Shi, F	2	2
Kretschmer, H	17	4	Chen, L	5	2	Sun, Xx	2	2
Zitt, M	13	5	Zhang, L	5	4	Liu, Jw	2	2
Liang, Lm	12	4	Frandsen, Tf	5	3	Van Hecke, P	2	2
Rowlands, I	12	5	Ahlgren, P	4	2	Bogaert, J	2	2
Wu, Ys	11	7	Rao, Ikr	4	2	Liu, Yx	2	2
Egghe, L	11	3	Arencibia-jorge, R	4	3	Otte, E	1	1
Jin, Bh	9	3	Ramanana-rahary, S	4	3	Guns, R	1	1
Pan, Yt	9	7	Yang, Nh	3	3	Ye, Fy	1	1
Ma, Z	9	7	Li, L	3	2	Jin, B	1	1
Van Hooydonk, G	7	3	Rousseau, S	2	2	Ren, Sl	1	1
Zuccala, A	7	4	Asonuma, A	2	2	Smyers, M	1	1
Liu, Xm	6	3	Fang, Y	2	2	Hu, Xj	1	1
Jiang, Gh	6	3	Jarneving, B	2	2			

3.1.3 k-shell 值最大的作者传播影响力分析

Klingsporn, B 和 Stock, Wg 的合作者数分别为 26、25，这些合作者的度、介数、k-shell 值如表 3 所示，所有度为 22 的作者其 k-shell 值也为 22，均处于网络的最核心位置，每个节点的传播影响力、扩散能力都很强，也就造成了这两位作者的传播影响力较大。值得注意的是，这些合作者的度却处于中间偏上的位置，并不是度最大的合作者，并且相对集中，可能意味着这些作者已逐渐形成较好的交流环境，并形成了相应的团体，是未来的新兴研究力量。同时，度值高的作者需要加强合作，提高彼此的传播影响力。介数在此处几乎没有起到作用，这些合作者还没有起到使不同作者、不同研究领域关联起来的桥梁作用，而且对其传播影响力影响较小。

表 3 Klingsporn, B 和 Stock, Wg 的合作者的度、介数和 k-shell 值

Klingsporn, B 的合作者	度	介数	k	Stock, Wg 的合作者	度	介数	k
Kosinski, M	22	0	22	Schloegl, C	4	0	3
Kuntze, J	22	0	22	Stock, M	1	0	1
Lee, Jr	22	0	22	Schmidt, S	1	0	1
Osterhage, A	22	0	22	Werner, K	22	0	22
Probost, M	22	0	22	Altvater-mackensen, N	22	0	22
Risch, T	22	0	22	Balicki, G	22	0	22
Schmitt, T	22	0	22	Bestakowa, L	22	0	22
Sturm, A	22	0	22	Bocatius, B	22	0	22
Weller, K	22	0	22	Braun, J	22	0	22
Hornbostel, S	4	0	4	Brehmer, L	22	0	22
Von Ins, M	9	0.001	4	Brune, V	22	0	22
Bohmer, S	4	0	4	Eigemeier, K	22	0	22
Neufeld, J	4	0	4	Erdem, F	22	0	22
Stock, Wg	25	0.001	22	Fritscher, R	22	0	22
Werner, K	22	0	22	Jacobs, A	22	0	22
Altvater-mackensen, N	22	0	22	Klingsporn, B	26	0.001	22
Balicki, G	22	0	22	Kosinski, M	22	0	22
Bestakowa, L	22	0	22	Kuntze, J	22	0	22
Bocatius, B	22	0	22	Lee, Jr	22	0	22
Braun, J	22	0	22	Osterhage, A	22	0	22
Brehmer, L	22	0	22	Probost, M	22	0	22
Brune, V	22	0	22	Risch, T	22	0	22
Eigemeier, K	22	0	22	Schmitt, T	22	0	22
Erdem, F	22	0	22	Sturm, A	22	0	22
Fritscher, R	22	0	22	Weller, K	22	0	22
Jacobs, A	22	0	22				

3.2基于介数和 k-shell 的作者影响力比较分析

3.2.1 总体情况分析

图 3 中横轴表示 k-shell 值，处于同一竖线的节点具有相同的 k-shell 值，纵轴表示作者的介数大小。颜色较深的节点表示此处重合节点数较多，多个节点处于这一坐标。

图 3 中处于左上方较为稀疏的节点为介数较大，而 k-shell 值较小的作者，如 k-shell 值为 4 时，Glanzel, W 的介数为 0.96，Lam, W 的介数为 0.74；k-shell 值为 8 时，Kelly, D 的介数为 1，为最大值，Marchionini, G 的介数为 0.77，Tang, R 的介数为 0.53，Sun, Y 的介数为 0.45。Rousseau, R 的介数则处于中上水平，为 0.61。值得注意的是，k-shell 值为 8 而介数大于 0.45 的作者与介数为 0 的作者（共 46

位）的 k-shell 值相同，均处于网络中第 8 层的较边缘位置，造成其在整个网络上的扩散能力不强。Leydesdorff, L 的介数为 0.74，其 k-shell 值为 9，Stock, Wg 和 Klingsporn, B 的合作者数分别为 25、26，k-shell 值同为 22，为 k-shell 值最大的节点中度也较大的节点，介数却为 0。

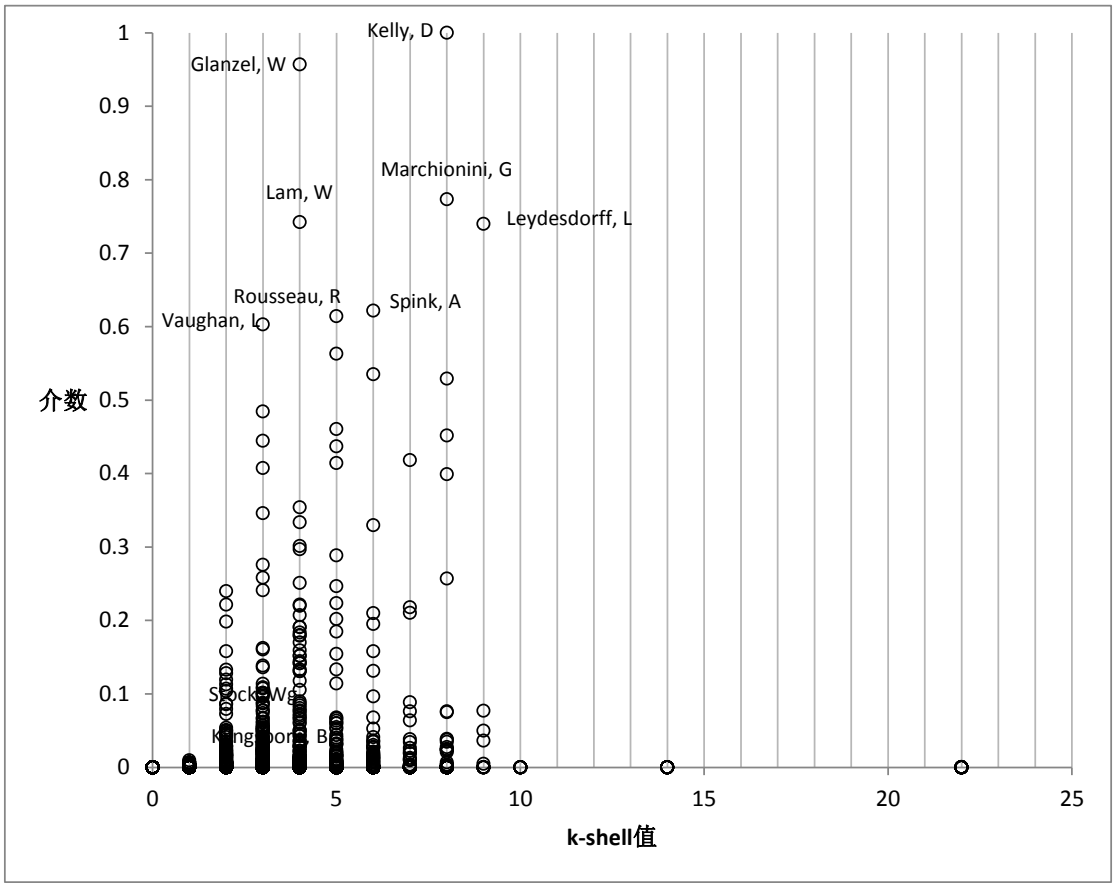

图 3 基于介数和 k-shell 的作者影响力比较分析

3.2.2 介数高但 k-shell 值较小的作者传播影响力分析

Kelly, D 作为介数最高但 k-shell 值较小，选取 Kelly, D 的 20 位合作者说明，如表 4 所示。Kelly, D 的合作者中，Sun, Y，Croft, Wb，Cool, C 和 Harper, Dj 的介数大于 0.1，占到 4/20，大于 0 小于 0.1 的占 8/20，而等于 0 的占 8/20。Kelly, D 的合作者介数均较低造成其传播影响力较弱。即使部分合作者有较高的介数，但 k-shell 值均较小，处于网络的边缘位置，不利于信息的传播，也造成 Kelly, D 的传播影响力较低。

表 4 Kelly, D 的合作者的介数和 k-shell 值

合著作者	介数	k	合著作者	介数	k	合著作者	介数	k
Sun, Y	0.452	8	Rittman, R	0.024	8	Sikora, C	0	6
Croft, Wb	0.437	5	Kantor, P	0.024	8	Park, Sy	0	6
Cool, C	0.210	6	Bai, B	0.024	8	Yuan, Xj	0	4
Harper, Dj	0.109	3	Perez-carballo, J	0.012	6	Murdock, V	0	4
Belkin, Nj	0.068	6	Lin, Sj	0.004	6	Landau, B	0	2
Strzalkowski, T	0.039	8	Small, S	0	8	Fu, X	0	1

Wacholder, N	0.035	8	Yamrom, B	0	8			

4 总结与展望

通过对情报学领域合著网络分析发现，k-shell 值较度、介数能更好的表征作者的传播影响力，这种方法也可以推广到基于科技文献数据的其它网络中，如识别文献共被引网络、文献耦合网络中最具传播影响力的关键文献。

Kitsak 等[4]证实高 k-shell 值的节点是最具传播影响力的单一传播源，当存在多个传播源的时候，传播的规模很大程度依赖于初始传播源之间的距离，此时，度高的节点往往比 k-shell 值大的节点更具传播效率。因为传播存在交叉感染现象，k-shell 大的节点往往在网络的核心，一般聚集在一起，交叉程度强；而度高的节点可以分散在网络的不同区域，交叉程度低。因此，在考虑多个传播源时，应该选择不同 k-shell 值、不直接相连的度值较高的节点作为初始传播源。在多个传播源的情况下计算合著网络中的作者传播影响力在本文中还没有考虑，需要在下一步工作中继续研究。

参考文献：

[1]. Albert R, Jeong H, Barabási A L. Error and attack tolerance of complex networks[J]. Nature, 2000, 406(6794): 378-382.
[2]. Cohen R, et al. Breakdown of the Internet under intentional attack[J]. Physical review letters, 2001, 86(16): 3682-3685.
[3]. Freeman L C. Centrality in social networks conceptual clarification[J]. Social networks, 1979, 1(3): 215-239.
[4]. Kitsak M, et al. Identification of influential spreaders in complex networks[J]. Nature Physics, 2010, 6(11): 888-893.
[5]. White H D, McCain K W. Visualizing a discipline: An author co-citation analysis of information science, 1972-1995[J]. Journal of the American Society for Information Science, 1998, 49(4): 327-355.
[6]. Zhao D Z, Strotmann A. Evolution of Research Activities and Intellectual Influences in Information Science 1996-2005: Introducing Author Bibliographic-Coupling Analysis[J]. Journal of the American Society for Information Science and Technology, 2008, 59(13): 2070-2086.
[7]. Chen C M, Ibekwe-SanJuan F, Hou J H. The Structure and Dynamics of Cocitation Clusters: A Multiple-Perspective Cocitation Analysis[J]. Journal of the American Society for Information Science and Technology, 2010, 61(7): 1386-1409.
[8]. 张金柱. 情报学的学科结构及其演化分析[J]. 情报资料工作, 2011(3): 34-37.
[9]. Newman M E J. The structure and function of complex networks[J]. SIAM Review, 2003, 45(2): 167-256.
[10]. Pastor-Satorras R, Vespignani A. Immunization of complex networks[J]. Physical Review E, 2002, 65(3): 036104.
[11]. Lloyd A L, May R M. How viruses spread among computers and people[J]. Science, 2001, 292(5520): 1316.
[12]. Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks[J]. Physical review letters, 2001, 86(14): 3200-3203.
[13]. Chen C. Searching for intellectual turning points: Progressive knowledge domain visualization[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(Suppl 1): 5303.
[14]. Chen C. Predictive effects of structural variation on citation counts[J]. Journal of the American Society for Information Science and Technology, 2011.
[15]. Freeman L. Centrality in social networks conceptual clarification[J]. Social networks, 1979, 1(3): 215-239.
[16]. Friedkin N E. Theoretical foundations for centrality measures[J]. American journal of sociology, 1991: 1478-1504.
[17]. Daley D J, Gani J, Gani J M. Epidemic modelling: an introduction[M]. NY: Cambridge University Press., 2001.

（Email：zhangjinzhu@mail.las.ac.cn）