

一种基于 K -Shell 的复杂网络重要节点发现算法

顾亦然,王 兵,孟繁荣

(南京邮电大学 自动化学院,江苏 南京 210023)

摘 要:复杂网络中的重要节点通常数量较少,但是对网络的影响却很大。为了能够有效地发现网络拓扑结构中的重要节点,文中基于 K -Shell 算法,在考虑节点自身重要度的基础上,考虑了邻居节点对自身节点的重要度贡献,提出 KSA(K -Shell-Affect)算法。该算法引入影响度概念,用节点自身的 K -Shell 值和与对其邻居节点的影响度来表征其对邻居节点的重要度贡献。对具有明显社团结构的 Zachary 网络进行仿真表明,该算法可行有效,克服了 K -Shell 划分结果的粗粒化,能够正确找到网络中的重要节点,具有一定的合理性,尤其在具有社团结构的网络中,能够十分有效地找到社团内部的核心节点。

关键词:重要节点; K -Shell; 重要度贡献; 影响度

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2015)09-0070-05

doi:10.3969/j.issn.1673-629X.2015.09.015

An Algorithm of Important Nodes Finding for Complex Network Based on K -Shell

GU Yi-ran, WANG Bing, MENG Fan-rong

(College of Automation, Nanjing University of Posts and Telecommunications,
Nanjing 210023, China)

Abstract: There are only a few important nodes in the complex network, which have a great impact on the complex network. In order to discover the important nodes in the complex network effectively, propose a novel algorithm-KSA(K -Shell-Affect) based on K -Shell. The algorithm considers the property of node itself, as well as the important contributions of the adjacent nodes by introducing the concept of affect. The important contribution to the adjacent nodes is characterized by K -Shell and affect to the adjacent nodes. The simulation on the Zachary network, which has a significant community structure, shows that the algorithm is feasible, effective and reasonable, and overcomes the coarse result of the K -Shell. Especially for a network of community structure, can find the core node of the community structure effectively.

Key words: important node; K -Shell; important contributions; affect

0 引 言

复杂网络往往有着大量节点,同时节点之间有着复杂的连接关系,它是复杂系统的高度抽象。复杂网络的研究已涉及因特网、电力与交通网络、生物网络、经济与金融网络以及社会网络等与生活息息相关的领域。复杂网络的小世界特性^[1]与无标度特性^[2],使得复杂网络的重要节点相比于网络中的其他节点能够更大程度上影响网络的功能与结构。在信息传播、病毒免疫、网络安全、电力枢纽等不同领域,对少数重要节点的保护或攻击,往往可以快速影响到网络中的大部

分节点状态^[3],如在一个无标度网络中,蓄意攻击网络中少量的重要节点会导致整个网络的瘫痪^[4-5];微博中具有影响力的少数名人所发的微博信息就能很快地传播到整个网络^[6]。因此,如何发现复杂网络中的重要节点已成为一项十分重要的工作。

寻找复杂网络中最重要的节点这一课题逐渐受到研究者的关注,衍生出许多早期的评估复杂网络中节点重要性的方法,包括度中心性、介数中心性、接近中心性、特征向量中心性等。由于这些指标存在一定的局限性,近年来,不断有新的重要节点发现算法被提

收稿日期:2014-09-20

修回日期:2014-12-24

网络出版时间:2015-06-23

基金项目:国家自然科学基金资助项目(61373136);教育部人文社科规划基金(12YJAZH120)

作者简介:顾亦然(1972-),女,博士,教授,CCF 会员,研究方向为复杂网络理论与应用、嵌入式系统、通信网络等;王 兵(1988-),男,硕士研究生,研究方向为复杂网络的建模、仿真与分析。

网络出版地址:

出。例如, Kitsak 等^[7]提出了用分解法将外围的节点逐层剥掉,处于内层的节点具有较大的值、较高的影响力,但是分解法是一种粗粒化的重要节点排序算法,没有很好的区分度。周漩等通过定义节点效率和节点重要度评价矩阵,提出了一种利用重要度评价矩阵来确定复杂网络关键节点的方法^[8]。李鹏翔等^[9]提出的节点删除的最短路径法,认为一个节点的破坏性可以用移除该节点后所引起的距离变化来表征。陈勇等^[10]提出的节点删除的生成树法认为,如果节点越重要,那么删除这个节点后,对应网络的生成树的数目就越少。文献^[11]提出的节点收缩法,将一个节点与它的邻接点收缩成一个新节点,认为节点越重要,那么收缩这个节点后,网络的凝聚度越大。文献^[12]基于节点的连接度和局部连通性定义了节点重要度函数,认为节点重要度与节点的度值成正比,与删除节点后该节点的邻居节点中依然连通的节点对数目成反比。文献^[13]提出了邻域与关键域的概念,在节点的邻域与关键域中,任意节点对之间的最短路径通过该节点则说明该节点越重要。

上述文献所提及的算法虽然在一定程度上能够有效衡量出复杂网络中的重要节点,但是它们都存在一定的缺陷:对于节点删除的生成树法,只能用在连通网络中,如果一个节点删除后,网络变得不再连通,那么这些节点的重要性认为是一致的,就难以判断了。对于节点删除的收缩法,由于每次收缩一个节点,都要计算一次网络的平均路径长度,时间复杂度比较高,不适用于计算大规模网络。对于重要度贡献矩阵算法,节点对邻居节点的重要度贡献是相同的,真实网络中,节点对邻居节点的重要度贡献其实不尽相同。基于以上考虑,文中为了有效地对复杂网络中的重要节点进行评估,提出了一种基于 K -Shell 的重要节点发现算法—KSA,基于节点对其邻居节点的影响度得到节点对邻居节点的重要度贡献。

1 复杂网络重要节点发现

复杂网络通常由大量的节点组成,节点之间通过连边进行相互作用,连通的节点对之间都存在着一定的重要度依赖关系,认为邻居节点之间的相互影响最主要、最直接。文献^[7]提出 K -Shell 分解法,它是一种粗粒化的划分结果,不能很好地区分同一 K -Shell 值的节点的重要性,而且没有考虑节点对之间的相互贡献。文献^[8]虽然考虑了节点对邻居节点的重要性贡献,但是认为节点对其邻居节点的重要性贡献是相同的,这存在一定的不合理性。例如 A 有 B, C 两个邻居节点,如果 A 和 B 之间的连接关系比 A 和 C 之间的连接关系更“亲密”,那么就认为 A 对 B 的影响度大于

对 C 的影响度。文中引入影响度概念,来表征相连节点对之间的影响。通常,在人际关系网络中,连接较为亲密的朋友的影响要大于那些连接关系一般的朋友的影响。

1.1 K -Shell 分解法

K -Shell 是图论里一个经典的概念,是一种粗粒化的节点重要性分类方法。不妨假设网络中不存在度为 0 的孤立节点。从度中心性的角度看,首先,将所有度值为 1 的节点及其连边去掉,这时网络中可能又会出现一些新的度值为 1 的节点,再把这些节点及其相连的边去掉,重复这种操作,直至网络中不再有度值为 1 的节点为止。这种操作形象上相当于剥去了最外面一层壳,这些被去掉的节点及它们之间的连边成为网络的 1-shell。剩下的新网络中每个节点的度值至少为 2,继续进行剥壳操作,即重复把网络中度值为 2 的节点及其相连的边去掉直至不再有度值为 2 的节点为止。这些被去除的节点及其之间的连边成为 2-shell。以此类推,直至网络中的每一个节点对应于唯一的 K -Shell 指标,并且 k_s -shell 中所包含的节点的度值必然满足 $k \geq k_s$ 。

图 1(a) 是一个可分解为三层壳的简单网络。图 1(b)-(d) 分别显示了 1-shell, 2-shell 和 3-shell 所包含的节点和边。

图中的两个黑色节点为该网络度值最大的两个节点,其度值均为 8。然而它们具有不同的 K -Shell 值,一个位于最里层, $k_s = 3$; 一个位于最外层, $k_s = 1$ 。

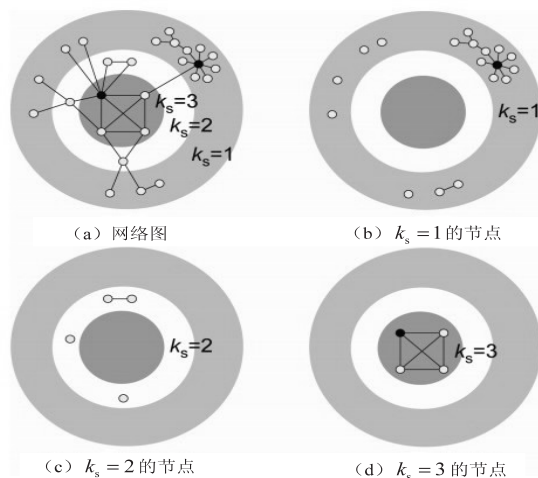


图 1 一个可分解为三层壳的简单网络

1.2 影响度

文中引入影响度概念,节点 i 与节点 j 的影响度定义如下:

$$w_{ij} = \frac{|d_i \cap d_j|}{|d_i \cup d_j|} \quad (1)$$

其中, d_i 和 d_j 分别为节点 i, j 的度;节点间的影响度 w_{ij} 就是节点 i, j 的共同邻接点占它们总邻接点的比

例, $w_{ij} \in [0, 1]$, 即拥有的共同邻接点数越多, 影响度越大。

现实生活中, i 对 j 的影响程度和 j 对 i 的影响程度是存在区别的, 如在一个真实社交网络中, i 和 j 是相互认识的, 但是 i 的社交圈很大, j 的社交圈很小, 那么 i 对 j 产生影响的概率就大于 j 对 i 产生影响的概率。

因此, 在一个网络中, 如果节点 i 和 j 存在边相连, 且节点 i 的度值远大于节点 j , 那么可以认为 j 对 i 的影响度大于 i 对 j 的影响度。

此外, 现实生活中, i 和 j 可能恰好是两个社交圈之间的桥梁, 彼此再没有公共的朋友。如图 2 所示的网络关系图中, 如果对于节点对 B 、 D 运用公式(1)得到 $w_{BD} = 0$, 这显然不合理。为此, 在网络中添加一个背景节点 S' , 使得节点与网络中的每个节点都存在连边, 这样, 任意两个相连的节点都至少存在一个公共节点, 避免了上述情况, 如图 3 所示。

这样, 会使得网络中任意相连节点对之间的共同节点数与总邻居节点数加 1, 简化得到将亲密度公式的分子分母都加 1。

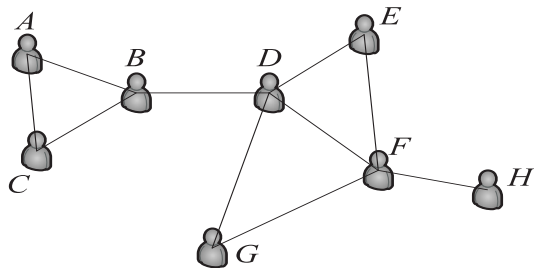


图 2 网络关系图

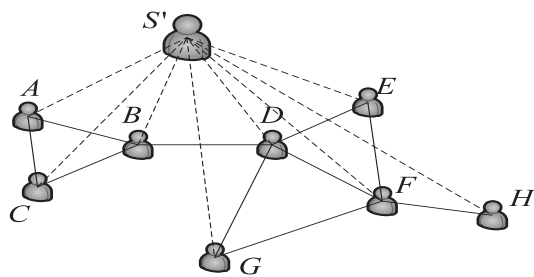


图 3 加入背景节点后的网络关系图

综上所述, 对公式(1)进行改进, 改进后的影响度: 如果 $d_i - d_j >$ 阈值 (通常取阈值为网络平均度值的 2 倍), 补偿 j 对 i 的影响度, 即:

$$w_{ij} = \frac{|d_i \cap d_j| + 1}{|d_i \cup d_j| + 1} + \alpha * (d_i - d_j) \quad (2)$$

否则, 影响度公式如下:

$$w_{ij} = \frac{|d_i \cap d_j| + 1}{|d_i \cup d_j| + 1} \quad (3)$$

1.3 算法模型

K -Shell 分解算法在时间复杂度为 $O(n)$ 的情况

下, 可以有效找到网络中核心节点, 但是得到的只是一种粗粒化划分结果, 层次性较差。

一个节点的重要度大小不仅与自身属性有关, 还依赖于网络中其他节点对其重要度的贡献, 尤其是邻居节点的贡献。基于此, 文中提出了一种算法模型。引入亲密度, 克服了节点对邻居节点的贡献是一样的不足, 使其更加合理。

由于网络的邻接矩阵反映了节点之间相连接的情况, 即节点与节点之间的连接关系。考虑到节点对邻接点的重要度贡献关系, 参照邻接矩阵的形式构建重要度贡献矩阵, 记为 C :

$$C = \begin{bmatrix} k_s(1) & \delta_{12}w_{12}k_s(2) & \cdots & \delta_{1n}w_{1n}k_s(n) \\ \delta_{21}w_{21}k_s(1) & k_s(2) & \cdots & \delta_{2n}w_{2n}k_s(n) \\ \vdots & \vdots & \ddots & \vdots \\ \delta_{n1}w_{n1}k_s(1) & \delta_{n2}w_{n2}k_s(2) & \cdots & k_s(n) \end{bmatrix} \quad (4)$$

式中, δ_{ij} 为网络邻接矩阵中的元素, 当节点 i 与节点 j 存在连边时 $\delta_{ij} = 1$, 否则等于 0; w_{ij} 为节点 j 对节点 i 的影响度, 由影响度定义公式知道, w_{ij} 与 w_{ji} 不一定相等; $k_s(i)$ 为节点 i 的 K -Shell 值。矩阵中, 位于 (i, j) 的元素表示节点 j 对节点 i 的重要度贡献值。

根据重要度贡献矩阵, 综合考虑节点自身的 K -Shell 值和相邻节点的重要度贡献值, 定义节点 i 的重要度 I_i 为:

$$I_i = k_s(i) * \sum_{j=1, j \neq i}^n \delta_{ij}w_{ij}k_s(j) \quad (5)$$

可以看出, 节点 i 的重要度 I_i 由两部分乘积构成, 一个因子是节点自身的 K -Shell 值, 另一个是其邻接点对节点 i 的重要度贡献值之和。

具体算法步骤如下:

Step1: 运用 K -Shell 分解算法, 得到网络中每个节点的 K -Shell 值, 得到的结果放在一个向量中。

Step2: 得到任意节点对之间的共同邻接点数与总邻接点数, 计算得到任意节点对之间的影响度。

Step3: 根据公式(4)得到影响度贡献矩阵。

Step4: 根据公式(5)计算每个节点的重要度 I 。

在算法中, K -Shell 分解算法的时间复杂度为 $O(n)$, 而计算节点对间的影响度的时间复杂度为 $O(n^2)$ 。所以本算法的时间复杂度取决于影响度的计算, 为 $O(n^2)$ 。

2 实验分析

Zachary 网络是社会网络分析领域中经典的数据集, 其来源于美国一所大学中的空手道俱乐部 34 名成员间的关系网, 1 号节点和 34 号节点分别是各团体的领袖, 由于空手道俱乐部数据真实且社区结构明显, 用

于许多算法的验证和分析。

Zachary 网络 K -Shell 划分结果如图 4 所示。

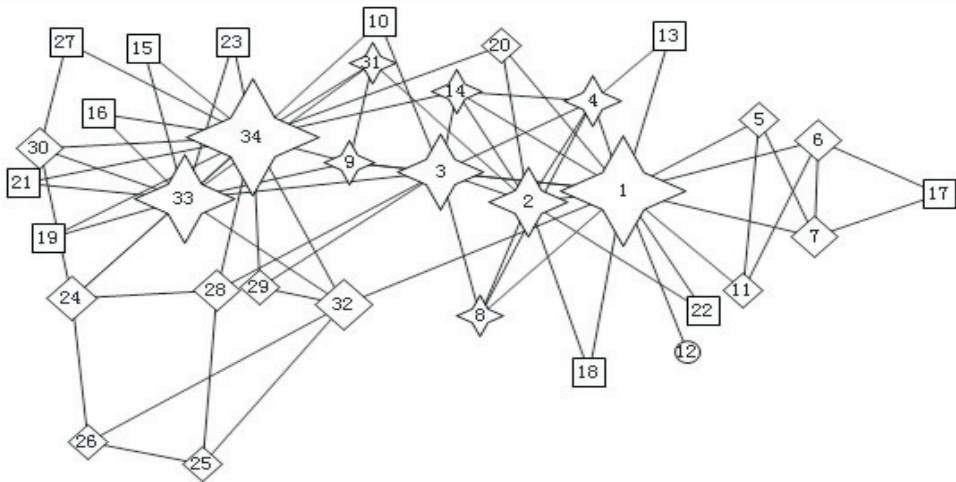


图 4 Zachary 网络 K -Shell 划分结果

图中,圆形代表 k_s 为 1 的节点,方形代表 k_s 为 2 的节点,菱形代表 k_s 为 3 的节点,星形代表 k_s 为 4 的节点。对此网络分别计算了 K -Shell、接近中心性、聚

类系数、环路系数、PageRank 算法、文中算法六种节点中心性参数指标值,得到的结果如图 5 所示。

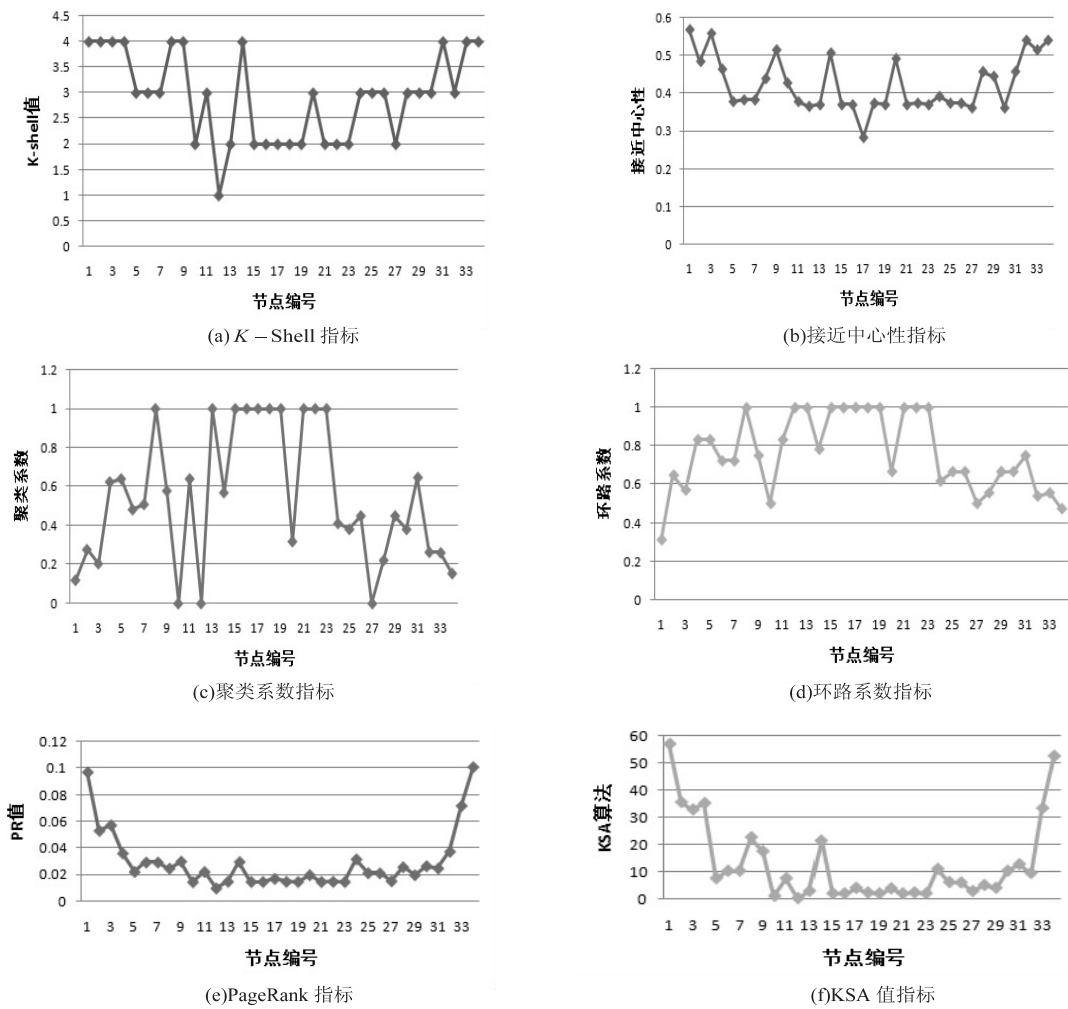


图 5 Zachary 网络中六种参数的计算结果对比

由图 4 和 5(a)可以直观发现,根据 K -Shell 算法,节点 1,2,3,4,8,9,14,31,33,34 的 k_s 值为 4,能够正确有效地找到网络中较为重要的节点,但是结果缺乏较高

的层次分辨率,即这些重要节点没有更好的区分度;由图(b)可以看出,完全基于节点间最短路径的接近中心度算法也能得到比较合理正确的结果,但是

具有较高的时间复杂度;由图(c)、(d)看出,聚类系数和环路系数对于发现网络中的核心节点表现不是很好;由图(e)、(f)可以直观看出,KSA 算法与在 Google 引擎上取得巨大成功的 PageRank 算法具有很大的相似性。

对节点的重要性分析发现:PageRank 算法认为节点 3 的重要性大于节点 2 和节点 4,而文中提出的算法认为节点 3 的重要性小于节点 2 和节点 4。空手道俱乐部网络存在两个社团,采用经典的 GN 算法^[14]进行社团划分时,节点 3 具有较大的歧义性,在多次划分结果中,有时属于左边的社团,有时属于右边的社团;而采用谱平分法划分社团时,节点 3 被错误划分在左边的社团中。因此,从社团核心节点的角度来看,节点 2 与节点 4 的重要性要大于节点 3,说明了文中算法的有效性。

3 结束语

发现网络中的重要节点具有十分重要的意义,例如可以提高信息的传播速度,免疫重要节点能够有效减缓甚至控制流行病的传播。文中基于 $K - \text{Shell}$ 算法,考虑了节点间的重要度贡献,引入影响度概念,使节点对其邻接点的重要度贡献值具有差异,这样更符合实际情况。根据节点间的重要度贡献构造了一个重要度贡献矩阵,由此提出了一个关键节点的发现算法。通过实验验证分析,该方法可行有效,能够有效发现网络中的重要节点,在具有社团结构的网络中,能够十分准确地发现社团内部的核心节点。

参考文献:

- [1] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks[J]. Nature,1998,393:440-442.

(上接第 69 页)

- 古石油化工,2009(14):28-30.
- [5] 张 军,李洪奇.常规测井资料质量自动验收方法研究与软件应用[J].天然气工业,2010,30(3):44-47.
- [6] 朱光明,李庆春,胡建平.数字信号分析与处理[M].西安:陕西人民教育出版社,2003.
- [7] 周 伟,桂 林,周 林,等. MATLAB 小波分析高级技术[M].西安:西安电子科技大学出版社,2006.
- [8] 雍世和,张超谟.测井数据处理与综合解释[M].东营:石油大学出版社,1996.
- [9] Keogh E, Pazzani M. An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback[C]//Proceedings of the 4th international conference on knowledge discovery and data mining. New York, NY:IEEE,1998:239-241.

- [2] Barabási A L, Albert R. Emergence of scaling in random networks[J]. Science,1999,286:509-512.
- [3] Albert R, Jeong H, Barabási A L. Error and attack tolerance of complex networks[J]. Nature,2000,406:378-382.
- [4] Callaway D S, Newman M E J, Strogatz S H, et al. Network robustness and fragility: percolation on random graphs[J]. Phys Rev Lett,2000,85(25):5468-5471.
- [5] Cohen R, Erez K, Ben-Avraham D, et al. Breakdown of the Internet under intentional attack[J]. Phys Rev Lett,2001,86(16):3682-3685.
- [6] Weng J, Lim E P, Jiang J, et al. Twitterank: finding topic-sensitive influential twitterers [C]//Proc of WSDM. New York:ACM Press,2010:261-270.
- [7] Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks[J]. Nat Phys,2010,6:888-893.
- [8] Zhou Xuan, Zhang Fengming, Li Kewu, et al. Finding vital node by node importance contribution matrix in complex networks[J]. Acta Physica Sinica,2012,61(5):050201.
- [9] 李鹏翔,任玉晴,席西民.网络节点(集)重要性的一种度量指标[J].系统工程,2004,22(4):13-20.
- [10] 陈 勇,胡爱群,胡 啸.通信网中节点重要性的评价方法[J].通信学报,2004,25(8):129-134.
- [11] 谭跃进,吴 俊,邓宏钟.复杂网络中节点重要度评估的节点收缩方法[J].系统工程理论与实践,2006,26(11):79-83.
- [12] 蒋丰景,陈玥琪.复杂网络中节点重要度的一个评估指标[J].西安工程大学学报,2014,28(1):140-142.
- [13] 张斌武,邹 森,王 勤.复杂网络上节点重要度的确定方法[J].兰州理工大学学报,2013,39(3):85-87.
- [14] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proc Natl Acad Sci,2001,99:7821-7826.

- [10] Degnan J J. Theory of the optimally coupled Q-switched laser [J]. IEEE Journal of Quantum Electronics,1989,25(2):214-220.
- [11] Domings P. MetaCost: a general method for making classifiers cost-sensitive[C]//Proceedings of the 5th ACM SIGKDD international conference on knowledge discovery and data mining. San Diego, CA:ACM,1999:155-164.
- [12] 刘 静,杨秀文.小波变换在孔隙度曲线去噪分析中的应用[J].河南石油,2002,16(2):4-5.
- [13] 张建业,潘 泉,张 鹏,等.基于斜率表示的时间序列相似性度量方法[J].模式识别与人工智能,2007,20(2):271-274.
- [14] Basiev T T, Karasik A Y, Osiko V V. Technologies of perforation of closely spaced micron holes with the help of neodymium[J]. Quantum Electronics,2009,39(4):385-387.

一种基于K-Shell的复杂网络重要节点发现算法

作者：[顾亦然](#)，[王兵](#)，[孟繁荣](#)，[GU Yi-ran](#)，[WANG Bing](#)，[MENG Fan-rong](#)
作者单位：[南京邮电大学 自动化学院](#)，[江苏 南京](#)，[210023](#)
刊名：[计算机技术与发展](#)[ISTIC](#)
英文刊名：[Computer Technology and Development](#)
年，卷(期)：[2015 \(9\)](#)

引用本文格式：[顾亦然](#). [王兵](#). [孟繁荣](#). [GU Yi-ran](#). [WANG Bing](#). [MENG Fan-rong](#) 一种基于K-Shell的复杂网络重要节点发现算法[期刊论文]-[计算机技术与发展](#) 2015 (9)