

基于局部核值的节点影响力计算方法研究

罗爱民¹ 杨圩生²

(国防科技大学信息系统工程重点实验室 长沙 410073)¹ (徐州工程兵学院 徐州 221000)²

摘要 识别复杂网络传播动力学中最重要的节点具有重要的应用价值。针对复杂网络影响力分析的问题,在核值影响力计算方法的基础上,提出了一种基于局部核值的影响力算法。以 Email 无标度网络数据为实验对象,来验证该方法的合理性和有效性。该方法为复杂网络影响力分析的应用研究提供了基础。

关键词 影响力,局部核值,复杂网络

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.8.007

Algorithm Research on Influence of Node Based on Local k-shell

LUO Ai-min¹ YANG Wei-sheng²

(Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Changsha 410073, China)¹

(College of Engineering Corps Xuzhou, Xuzhou 221000, China)²

Abstract It has important application value to identify the most important nodes in complex network transmission dynamics. Aiming at the problem of complex network influence analysis, a method of measuring influence based on local k-shell was provided in this paper, which considers the nuclear value of neighbor nodes. Finally, the method was verified to be rational and valid by using the data set of Email scale-free network as experimental object. It is foundation for the application research of influence of complex network.

Keywords Influence, Local k-shell, Complex network

1 概述

现实生活中的网络动力学现象时有发生,例如社会网络中的热点信息传播、交通网络中的汽车流堵塞、计算机网络中的邮件病毒传播、电力等基础设施网络中的级联故障等。复杂网络影响力相关研究多集中在信息、病毒传播影响力最大化问题^[1-3]。这些网络动力学现象往往不是整个网络节点共同造成的,而是网络中部分节点结构的损坏或功能的失效造成的全局性问题。由于网络节点“位置”的差异,不同节点在网络动力学中产生的作用有所不同,因此网络节点影响力分析成为了复杂网络动力学研究的重要内容。

复杂网络影响力是指在特定的网络动力学过程中,网络个体的动力学行为造成整个网络动力学现象的程度。目前,复杂网络影响力研究^[4-7]多是以网络中心性度量方法为基础的一种延伸,通常将网络拓扑结构特征参数作为衡量节点影响力大小的标准,如节点度中心性、介数中心性、接近中心性等。这些参数在一定程度上反映了节点在拓扑结构中的地位。而复杂网络影响力更多的是关注节点在局部网络甚至整个网络的动力学行为。因此,复杂网络影响力模型的建立不仅仅要考虑节点本身的性质,更要考虑节点局部网络结构甚至是整个网络结构的性质。很多学者^[8,9]认为核值(k-shell)影响力度量方法是一种最有效的节点影响力计算方法。然而

核值影响力算法虽然考虑了节点本身的性质,但忽略了其局部网络结构甚至整个网络结构对节点的影响,按照这种方法得到的网络节点的核值粒度较粗,这样很可能造成节点影响力大小排序分层不明显,甚至可能引起排序颠倒的问题。

本文针对复杂网络影响力分析的问题,在核值影响力计算方法的基础上,考虑邻居节点的核值,提出一种基于局部核值的影响力算法。该算法能够更准确地区分不同节点在网络中的“位置”,能够更好地分析和利用复杂网络动力学特征,为支持复杂网络的相关研究提供支持。

2 基于局部核值的影响力计算方法

2.1 节点核值计算方法

网络节点核值计算采用 k-壳分解法。k-壳分解法就是根据节点度的大小以及节点度所处的层次,由外而内的剥壳方法。

k-壳分解法的计算步骤如下。

(1)在网络模型中,将节点度为 1 的节点及其相关联的边从网络模型中去掉。这时网络中又出现了节点度为 1 的节点,继续把这些度为 1 的节点及其相关联的边去掉。重复以上操作,直至网络中没有节点度为 1 的节点。将所有被去掉的节点以及相关联的边称为网络的 1-壳。这些节点的核值为 1。

(2)在去掉了 1-壳后的新网络中,将节点度 $i=2$ 的节点

到稿日期:2015-06-20 返修日期:2015-10-23 本文受国家自然科学基金项目(71171197)资助。

罗爱民(1970—),女,博士,教授,主要研究方向为信息系统分析与设计,E-mail:amluo@nudt.edu.cn;杨圩生(1990—),男,助教,主要研究方向为信息系统分析与设计。

及其相关联的边去掉,直到网络中没有节点度小于或等于 i 的节点。在这一轮中所有被去掉的节点以及相关连的边称为网络的 2-壳。这些节点的核值为 2。

(3)对新得到的网络模型,令节点度为 $i(i=3,4,\cdots,k)$,重复前面的步骤,直至去掉网络中的最内层节点,得到网络的 k -壳。这时网络中节点的核值为 k 。

对于图 1 所示的简单网络模型,按照 k -壳分解法求得各节点核值在图 1 中,最里面的 $k=3$ 表示该圆内的节点 1,2,4,5 的核值为 3,最外圈的圆环 $k=1$,表示其中的节点核值为 1。

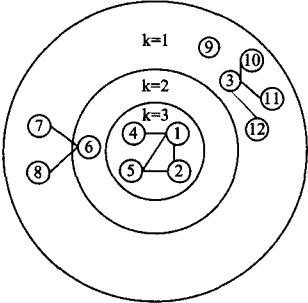


图 1 网络模型及各节点的核值

2.2 局部核值计算方法

对于图 1 所示的网络模型,按照核值计算结果,节点 1 和节点 2 的核值均为 3。如果以核值作为影响力的标准,那么节点 1 和节点 2 的影响力相同。而实际上,节点 1 和节点 2 的影响力是不同的,因为从网络的整个拓扑结构来看,由于节点 1 的邻居节点 3 的度较大,节点 3 对节点 1 的影响力是正相关的,因此从逻辑上来讲,节点 1 的影响力要大于节点 2 的影响力。

为更加精确地区分不同节点的影响力,可采用局部核值度量影响力。

节点 i 的局部核值 $v_i(k\text{-nshell})$ 为:

$$v_i(k\text{-nshell}) = \sum_{j \in N(i)} v_j(k\text{-shell})$$
 其中, $N(i)$ 为节点 i 的邻居节点集合, $v_j(k\text{-shell})$ 表示节点 j 的核值。

2.3 基于局部核值的节点影响力算法

基于局部核值($k\text{-nshell}$)的影响力度量方法的主要思想是以节点的邻居节点核值之和为依据,进而对节点影响力大小进行层次排序。

节点影响力的大小用局部核值来度量。基于局部核值计算节点影响力的具体算法实现如下。

算法 1 k-nshell

```

Input: 图 G
Output: 网络节点的局部核值 k-nshell
Step1 计算节点核值 k-shell
    k-shell=1;
    While (图 G 中有节点)
        for (图 G 中的每个节点  $v_i$ )
            if (节点  $v_i$  的度  $\text{degree} \leq k\text{-shell}$ )
                节点  $v_i$  的核值  $v_i(k\text{-shell}) = k\text{-shell}$ ;
                将节点  $v_i$  从图 G 中去除;
            end
        end
        k-shell=k-shell+1;
    end
Step2 计算节点局部核值 k-nshell
    for (图 G 中的每个节点  $v_i$ )
        节点  $v_i$  的局部核值  $v_i(k\text{-nshell}) = 0$ ;
        for (图 G 中的其它节点  $v_j$ )
            if (节点  $v_i$  和节点  $v_j$  相邻)
                 $v_i(k\text{-nshell}) = v_i(k\text{-nshell}) + v_j(k\text{-shell})$ ;
            end
        end
    end
Step3 return k-nshell;
```

分析该算法的时间复杂度:当网络模型的节点总数为 n 时,在 Step1 计算节点核值 $k\text{-shell}$ 中,有 while 和 for 两个循环语句,因此其时间复杂度为 $O(n^2)$;在 Step2 计算节点局部核值 $k\text{-nshell}$ 中,有两个 for 循环语句,因此其时间复杂度为 $O(n^2)$ 。那么,该基于局部核值的节点影响力算法的时间复杂度为 $O(n^2)$ 。

对于图 1 所示的网络模型,分别用核值法($k\text{-shell}$)、局部核值法($k\text{-nshell}$)对该网络模型进行节点影响力大小分析与排序,其结果如表 1 所列。

表 1 基于核值法和局部核值法的网络模型影响力分层排序

排序层次	核值 (k-shell)	对应节点	局部核值 (k-nshell)	对应节点
1	3	1,2,4,5	11	4,5
2	2	6	10	1
3	1	3,7,8,9,10,11,12	9	2
4			8	6
5			7	3
6			2	7,8
7			1	9,10,11,12

从表 1 可以看出,以局部核值来评价节点影响力,节点 1 和节点 2 处于不同分层,节点 1 的影响力大于节点 2 的影响力。局部核值法并没有改变基于核值法得到的不同层次之间节点的排序,而仅仅对同一层次内节点的影响力大小进行进一步分层排序。因此,相比核值法而言,局部核值法是一种更加细化的影响力度量方法。

3 算法性能分析

为了进一步说明基于局部核值计算影响力的有效性,以 Email 网络数据集为研究对象,对核值和局部核值计算方法进行对比分析。

该 Email 网络数据集来源于西班牙洛维拉·依维尔基大学(Univeristy Rovira i Virgili)师生之间的电子邮件通信关系,将该网络中的个体抽象为网络节点,将个体之间的通信关系抽象为无向边。经处理,网络中共有 1133 个节点和 5451 条边。可以证明该 Email 网络模型为无标度网络。

利用 Matlab 工具,分别求出 Email 网络中各节点的核值($k\text{-shell}$)、局部核值($k\text{-nshell}$),结果如图 2 和图 3 所示。

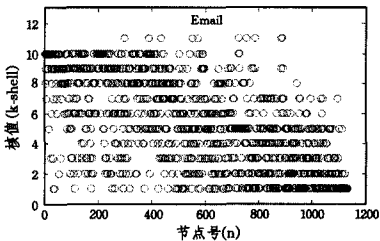


图 2 Email 网络节点核值($k\text{-shell}$)

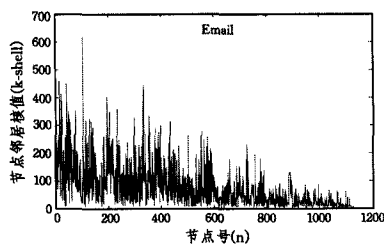


图3 Email网络节点局部核值(k-nshell)

对比分析图2和图3,可以发现利用标准的k-shell分解方法时节点的核值分层数仅为11。相对网络节点总数而言,节点的核值分层数较少,分层粒度较粗。而局部核值法可以区分较多的节点层次,因此利用局部核值法评估节点影响力能够更精确地反映节点之间的差别。

表2列出了通过核值(k-shell)法与局部核值(k-nshell)法得出的Email网络排名前10位的节点编号。可以看出,采用核值法与局部核值法得到的Top-10节点的差别很大。

表2 Email网络在核值法、局部核值法下的最具影响力Top-10节点

排序号	1	2	3	4	5	6	7	8	9	10
k-shell-节点号	299	389	434	552	571	726	756	788	885	886
k-nshell-节点号	105	16	42	333	23	196	41	332	233	76

为了说明基于局部核值的节点影响力计算方法的优劣,通过比较基于节点度的节点影响力计算方法、局部核值的计算方法和核值计算方法在Email网络的病毒传播动力学模型中的应用结果来进行进一步分析。

根据SIR模型的定义,SIR模型中3类节点分别为易感状态S(Susceptible)、已感状态I(Infected)和免疫状态R(Refractory)。设在一个节点总数为 N 的网络模型中, t 时刻网络中处在易感状态、已感状态、免疫状态的节点比例分别为 $s(t)$, $i(t)$, $r(t)$,且 $s(t)+i(t)+r(t)=1$ 。SIR模型的微分方程如下:

$$\begin{aligned} s'(t) &= -as(t)i(t) \\ i'(t) &= as(t)i(t) - bi(t) \\ r'(t) &= bi(t) \end{aligned}$$

其中, a 表示单位时间内的感染概率, b 表示单位时间内的免疫概率, $s'(t)$, $i'(t)$, $r'(t)$ 分别表示各种状态下的节点数在单位时间内的变化情况。

为了评价不同影响力度量方法在Email无标度网络模型中的动力学表现,设计如下实验:在相同的参数设置下,首先根据影响力算法确定网络中各个节点影响力大小的排序,其次将一定比例的最具影响力Top- k 节点作为初始感染节点,分析各种状态的节点随时间的变化情况。以传播过程中已感状态节点(I)的比例变化为评估指标。实验参数设置为:初始节点感染率 $p_0=0.01$,初始免疫状态节点比例为0;感染概率 $a=0.2$;免疫概率参数 $b=0.1$;最大时间步 $t_{max}=80$;独立实验次数 $n_{shiyen}=100$ 。

图4显示了基于节点度法、核值法和局部核值法排序的Email网络已感状态节点(I)的比例变化情况。从中可以看出,在初始的时间步内,基于节点度影响力方法和基于局部核值影响力方法的已感状态节点增长速率是最快的,两者差别很小;核值法的感染速度最慢。随着时间的推进,基于局部

核值影响力方法的已感状态节点增长速率快于基于节点度影响力方法的。当SIR动力学过程达到峰值状态时,局部核值法具有最大的峰值,即该方法下被感染的节点比例是最大的,这验证了局部核值法在节点影响力大小排序上的优越性。

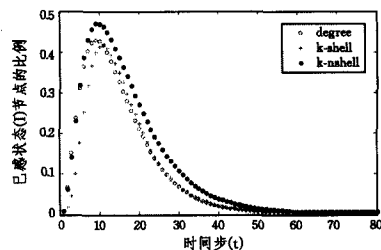


图4 3类影响力度量方法下的已感状态节点(I)的比例变化结果

结束语 本文提出了基于局部核值的节点影响力计算方法。该方法既体现了核值影响力度量方法挖掘网络“核心”的优点,也利用了节点度影响力度量方法选取网络聚集中心的优点。基于实验分析,可以认为对于无标度网络,基于局部核值的影响力度量方法在网络节点影响力排序上是最优的。针对其他复杂网络类型算法的有效性验证和改进是下一步的主要研究内容。

参考文献

- [1] Funk S, Gilad E, Watkins C, et al. The spread of awareness and its impact on epidemic outbreaks[J]. Proceedings of the National Academy of Sciences, 2009, 106(16): 6872-6877
- [2] Goltsev A V, Dorogovtsev S N, Oliveira J G, et al. Localization and spreading of diseases in complex networks [J]. Phys Rev Lett, 2012, 109(12): 128702
- [3] Song Yu-rong, Jiang Guo-ping, Xu Jia-gang. An epidemic spreading model in adaptive networks based on cellular automata[J]. Acta Physica Sinica, 2011, 60(12): 120509 (in Chinese)
宋玉蓉, 蒋国平, 徐加刚. 基于适应网络的元胞自动机病毒传播模型研究[J]. 物理学报, 2011, 60(12): 120509
- [4] Pinto P C, Thiran P, Vetterli M. Locating the source of diffusion in large-scale networks[J]. Physical Review Letters, 2012, 109(6): 068702
- [5] Chen Duan-bing, Lu Lin-yuan, Shang Ming-sheng, et al. Identifying influential nodes in complex networks[J]. Physica A: Statistical Mechanics and its Applications, 2012, 391(4): 1777-1787
- [6] Chen W, Wang Y, Yang S. Efficient Influence Maximization in Social Networks[C] // Proceedings of KDD09. Paris, France, 2009: 199-208
- [7] Narayanam R, Narahari Y. A Shapley Value-Based Approach to Discover Influential Nodes in Social Networks [J]. IEEE Transactions on Automation Science and Engineering, 2011, 8(1): 130-147
- [8] Ren Zhuo-ming, Liu Jian-guo, Shao Feng, et al. Analysis of the spreading influence of the nodes with minimum K-shell value in complex networks [J]. Acta Physica Sinica, 2013, 62(10): 108902
- [9] Daniel M, De P F. K-Shell Decomposition for Dynamic Complex Networks[C] // Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt'10). Avignon, France, 2010: 499-507