

复杂网络中最小 K-核节点的传播能力分析*

任卓明 刘建国[†] 邵凤 胡兆龙 郭强

(上海理工大学复杂系统科学研究中心, 上海 200093)

(2012年9月27日收到; 2013年1月16日收到修改稿)

K-核分解方法对于识别复杂网络传播动力学中重要节点具有重要的价值, 然而该方法无法对复杂网络中大量最小 K-核节点的传播能力进行准确度量. 本文主要考察最小 K-核节点的传播行为, 利用其邻居的 K-核信息, 提出一种度量这类节点传播能力的方法. 实证网络数据集的传播行为仿真结果表明, 该方法与度、介数等指标相比更能准确度量最小 K-核节点的传播能力.

关键词: 复杂网络, 传播能力, K-核分解, 最小 K-核节点

PACS: 89.75.Da, 05.10.-a

DOI: 10.7498/aps.62.108902

1 引言

准确度量复杂网络中节点的传播能力具有十分重要的现实和理论意义^[1,2]. 该工作对于预防网络攻击^[3]、阻碍计算机病毒在网络上的蔓延^[4]、防止传染病在人群中的流行^[5]、抑制流言在社会中的扩散^[6]和引导信息在社交网络中的传播^[7,8]等方面有着非常重要的作用. 节点重要性通常用中心性指标如度指标、紧密度指标^[9]以及介数指标^[10]等度量. 度指标只能刻画节点的局部信息, 紧密度与介数都因需要计算最短路径而导致算法的时间复杂性非常高. Kitsak 等^[11]认为网络传播动力学中最重要的节点并非传统的度最大或介数最大的 Hub 节点, 而是具有最大 K-核值的节点. Chen 等^[12]利用节点的多级邻居信息度量复杂网络中的最重要节点. 周漩等^[13]则提出重要度评价矩阵识别网络中的最重要节点. 虽然这些工作较常用的度指标、紧密度指标以及介数指标更能准确找到网络中最有影响力的节点, 却没有对 K-核很小节点的传播能力进行分析, 而这类节点往往在网络总节点数中占很大比例. 表 1 给出 PGP^[14] (pretty

good privacy network) 和 AS^[15] (autonomous system) 两个网络中这类节点所占比例, 从中可以看出网络中诸如 K-核、度为 1 和介数为 0 等这些节点所占的比例非常高, 而且这类节点 K-核、度或介数的值是完全相同的, 因此仅仅依靠自身的 K-核、度或介数信息不能很好地区分这类节点的传播能力. 我们利用其邻居集中的最大 K-核信息, 提出一种度量该类节点传播能力的方法. 数值试验显示, 利用邻居节点的 K-核信息可以准确刻画这部分节点的传播能力. 本文结构如下: 首先采用 K-核分解^[16]方法将网络划分为不同等级层次的社团, 筛选出最小 K-核节点集合; 然后利用该类节点邻居集中的最大 K-核值, 度量其传播能力; 最后通过大量真实网络的 SIR 传播仿真, 分别比较度指标、介数指标和新指标的表现. 结果表明该新指标更能够准确度量最小 K-核节点传播能力.

表 1 PGP 和 AS 网络中的部分节点所占比例

网 络	总节 点数	K-核为 1 的比例/%	度为 1 的 比例/%	介数为 0 的比例/%
PGP	10680	49.12	39.61	52.15
AS	6201	37.97	36.89	52.47

* 国家自然科学基金 (批准号: 71071098, 71171136, 91024026)、上海市青年科技启明星计划 (A 类)(批准号: 11QA1404500)、上海市教委科研创新项目 (批准号: 11ZZ135, 11YZ110)、教育部科学技术研究重点项目 (批准号: 211057)、上海市一流学科 (系统科学) 建设项目 (批准号: XTKX2012) 和上海市研究生创新基金 (批准号: 5411115004) 资助的课题.

[†] 通讯作者. E-mail: liujg004@ustc.edu.cn

2 理论基础与方法

2.1 节点的传播能力

采用 susceptible-infected-removed (SIR)^[17,18] 模型仿真网络最小 K -核节点的传播行为, SIR 模型将传染范围内的人群划分为 3 类: 1) S 类, 易感人群 (susceptible), 指未得病但缺乏免疫能力、与感病者接触后容易受到感染的人群; 2) I 类, 染病人群 (infected), 指染上传染病的人群; 3) R 类, 移除人群 (recovered/removed), 指已经治愈并获得了免疫能力或者已经死亡、不再对相应动力学行为产生任何影响的人群. 在 SIR 模型中, 染病人群为传染的源头, 以概率 β 把传染病传给易感人群, 经过 Tr 时间后, 以概率 δ 治愈或死亡. SIR 传染过程如图 1 描述.

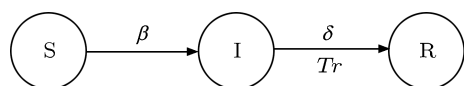


图 1 SIR 模型传播过程示意图

假设网络中节点 i 为传染的源头, 独立重复此节点 SIR 仿真 M 次, 则节点 i 传播能力可以表示为

$$\bar{X}_i = \frac{1}{M} \sum_{m=1}^M X_i(m), \quad (1)$$

$$\sigma(i) = \sqrt{\frac{\sum_{m=1}^M (X_i(m) - \bar{X}_i)^2}{M}}, \quad (2)$$

其中 $X_i(m)$ 为节点 i 在第 m 次仿真时的传播值, 即每次 SIR 仿真结束时, 网络中状态为 I 类与 R 类的节点个数, \bar{X}_i 为节点 i 在 M 次仿真中的平均传播值, $\sigma(i)$ 为标准差, 它表达了 $X_i(m)$ 与 \bar{X}_i 的偏离程度. 若 $\sigma(i) = 0$ 表示每次 $X_i(m)$ 取值相同. $\sigma(i)$ 较大, 说明在不同仿真实验中, 节点 i 的每次传播值差别较大, 即该节点传播能力发散. 本文用平均传播值 \bar{X}_i 和标准差 $\sigma(i)$ 度量节点的传播能力.

2.2 K-核分解

K -核分解^[16,19] 方法通过递归地移去网络中所有度值小于或等于 K 的节点, 它能描述网络结构特征, 揭示网络层次性质.

假设网络 $G = (V, E)$ 是由 $|V| = N$ 个节点和 $|E| = E$ 条边所组成的一个无向网络, 则 K -核的定义如下: 由集合推导出的子网络 $H = (C, E|C)$, 当且仅当对 C 中的任意节点 V , 其度值均大于 K , 具有这一性质的最大子网络的补集被称为 K -核, 简称 K_s . K_s 分解示意图如图 1, 该网络被划分为 3 层不同的核.

根据 K -核分解的定义可知, 最小 K -核节点为网络最外层 ($K_s = 1$) 的节点. 如图 2 所示, 最小 K -核节点即为编号为 1 至 12 的节点.

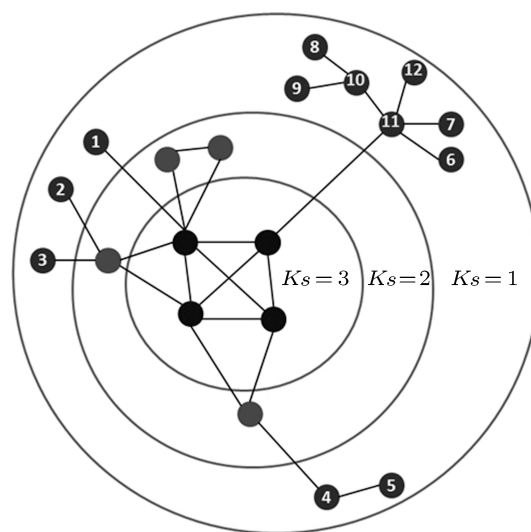


图 2 K -核分解示意图

2.2.1 最小 K -核节点的深度

由于最小 K -核节点的 K_s 值是相同的, 而且根据 K -核分解原理, 度为 1 以及大部分介数为 0 的节点属于最小 K -核节点, 因此仅仅依靠节点自身 K -核、度或介数信息不能很好区分这类节点的传播能力, 于是我们提出了基于最小 K -核节点邻居集合中最大 K_s 值的深度指标 $d(i)$, 表示为

$$d(i) = \max \{K_{s_j}\}, \quad j \in J(i), \quad (3)$$

其中 $J(i)$ 为节点 i 的邻居集合, K_{s_j} 为节点 j 的 K_s 值. 该指标可以判断最小 K -核节点与网络的其他层级节点的连接关系, 尤其可以用来判定一个节点是否为中心节点的邻居. 表 2 给出了图 2 中最小 K -核节点的深度, 如 $d(10) = 1$ 表示 10 号节点的邻居集合中的最大 K_s 值为 1, $d(11) = 3$ 表示 11 号节点的邻居集合中的最大 K_s 值为 3.

表 2 图 2 的网络中最小 K-核节点的深度值

节点标号	深度	节点编号	深度
1	3	7	1
2	2	8	1
3	2	9	1
4	2	10	1
5	1	11	3
6	1	12	1

3 数值仿真与结果分析

3.1 实验数据及相关参数

实验数据采用 PGP 和 AS 网络. PGP 信任网是一个每个节点都包含公钥和私钥的双向信任连接的无向网络, AS 网络为 CAIDA 2005 年的路由级拓扑测量数据获得的无向网络. 其基本统计特性如表 3 所示.

3.2 最小 K-核节点的传播能力分析

采用 SIR 模型分析网络最小 K-核节点传播性质. 在 SIR 模型中, 令 $\delta = 1, Tr = 2$, 即感染者 I 类经过 $Tr = 2$ 的时间后变成 R 类. 首先假设网络中的某个最小 K-核节点染病, 然后独立重复该节点 SIR 仿真 200 次. 文献 [8] 认为最小 K-核节点的平均传播值很小, 没有考虑最小 K-核节点传播值的发散性. 图 3 分别给出了感染率不同的两个网络中所有最小 K-核节点的传播值大小及标准差的分布范围. 由图可知, 感染率越小, 最小 K-核节点间的传染值的差距越小, 传染值的发散程度也越小. 但随着感染率的增大, 最小 K-核节点间的传染值差距明显增大, 而且传染能力的发散程度加剧. 这表明感染率不断增大时, 最小 K-核节点间传播能力的差异越明显.

表 3 PGP 和 AS 网络的基本参数

网络	节点数	边数	平均度	最大 K 核
PGP	10680	24316	3.92	31
AS	6201	12170	4.6	12

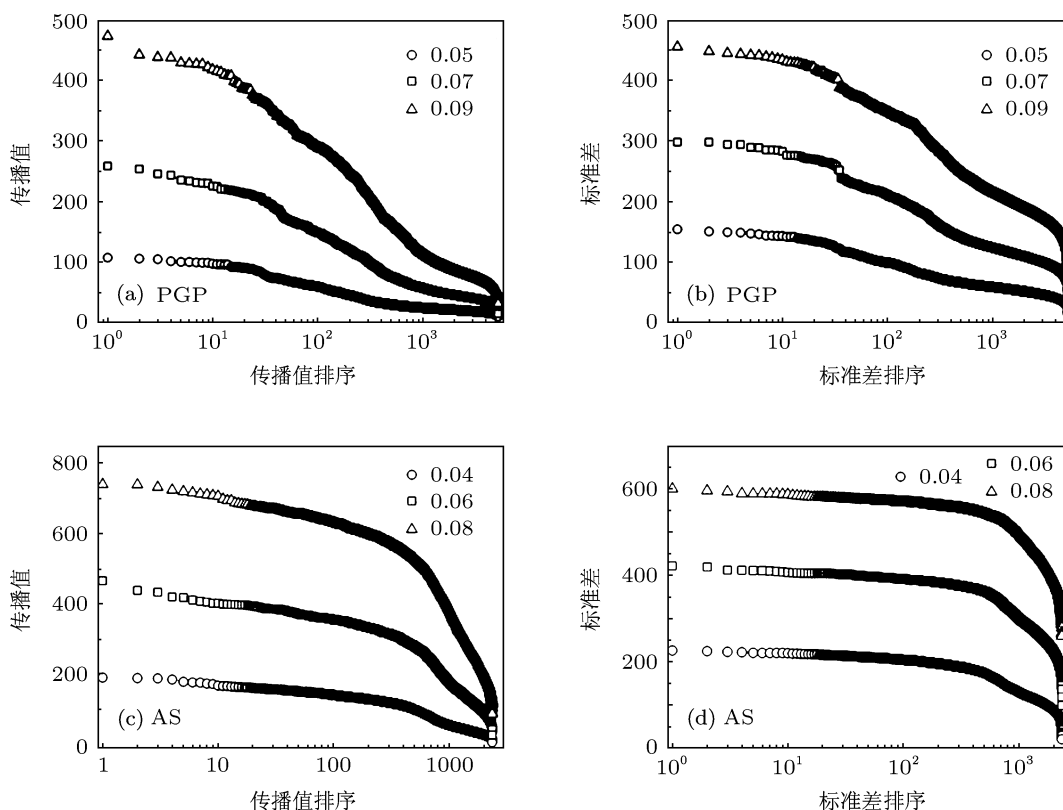


图 3 最小 K-核节点的传播能力 (a), (b) 分别为 PGP 网络最小 K-核节点的传播值和标准差排序; (c), (d) 为 AS 网络的相应结果

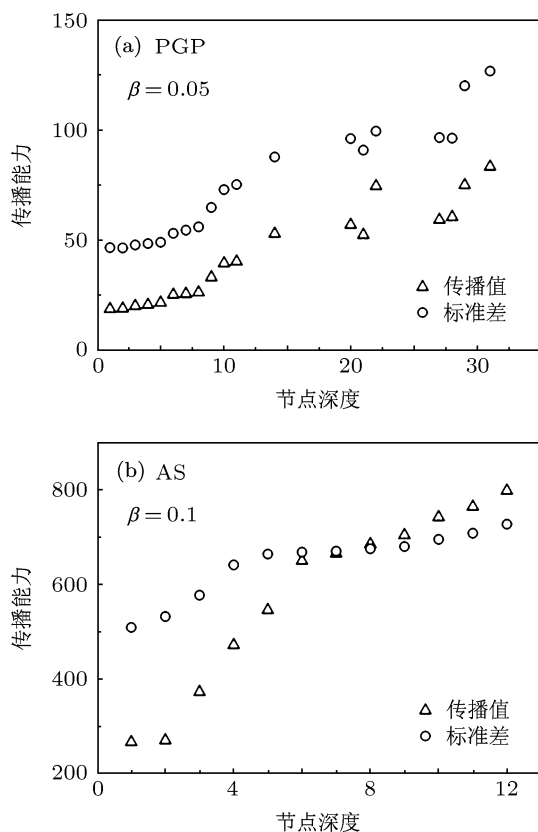


图4 不同感染率节点深度与传播能力的关系 (a) PGP 网络, 感染率为 0.05; (b) AS 网络, 感染率为 0.1

大量最小 K-核节点间传播能力是不一样的, 为

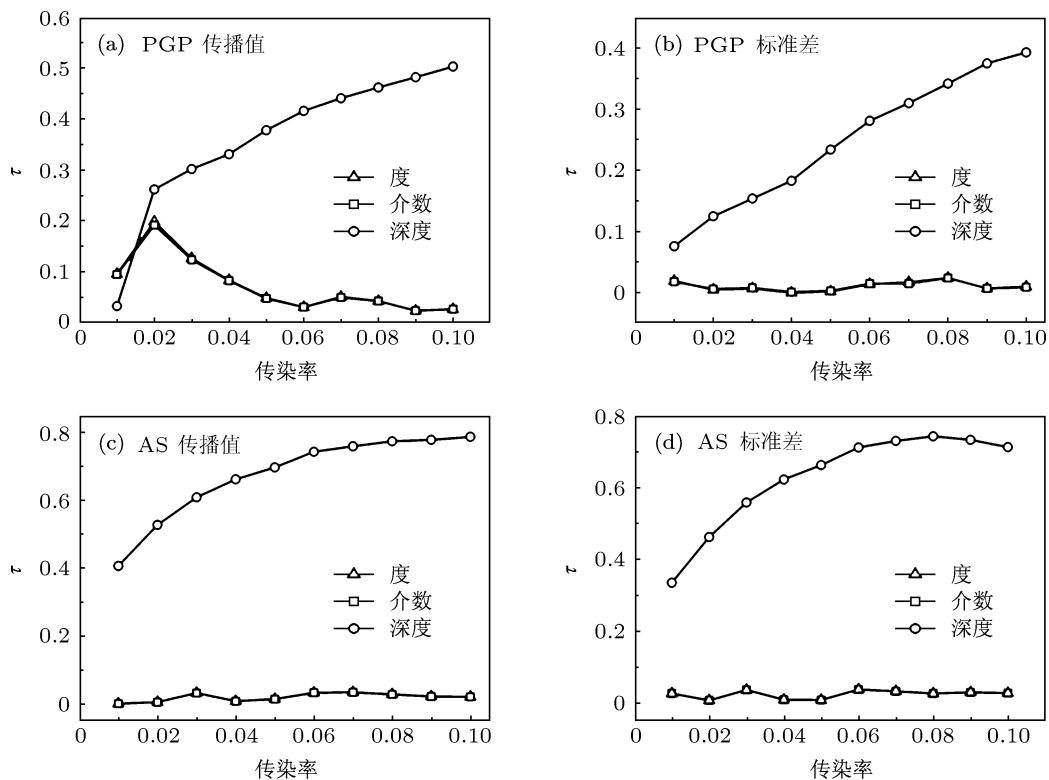


图5 不同感染率下, 节点深度 $d(i)$ 与节点传播能力的相关性 (a), (c) 为最小 K-核节点的度、介数和节点深度 $d(i)$ 指标分别在 PGP 和 AS 网络中与传播值的相关性; (b), (d) 为最小 K-核节点的度、介数和节点深度 $d(i)$ 指标分别在 PGP 和 AS 网络中与传播值的标准差的相关性

为了区分最小 K-核节点传播能力的差异性, 我们采用最小 K-核节点深度 $d(i)$ 度量其传播能力, $d(i)$ 越大, 其传播值越好, 但标准差越大. 图 4 给出了 PGP 和 AS 网络分别在感染率为 0.05 和 0.1 的情况下, 最小 K-核节点深度与传播能力之间的关系. 从中可以看出 $d(i)$ 越大, 传播值越大, 同时其传播值的标准差也越大, 也就是该节点传播能力发散程度越大.

3.3 深度与节点传播能力的相关性分析

我们采用 Kendall's Tau 系数^[20] 比较最小 K-核节点的度、介数、深度与节点真实传播能力之间的相关性. Kendall's Tau 系数是用来描述两个序列之间相关性的指标, 取值范围为 $[-1, 1]$, Kendall's Tau 系数越大说明两个序列一致性越高, 则该指标能越准确反映最小 K-核节点的实际传播能力, 其公式为

$$\tau = \frac{2}{N(N-1)} \sum_{i < j} \text{sgn}[(x_i - x_j)][(y_i - y_j)], \quad (4)$$

其中 N 为最小 K-核节点数, x_i 表示节点 i 的传播值或其标准差, y_i 表示节点 i 的度、介数或最小 K-核节点深度. $\text{sgn}(x)$ 为分段函数, 即 $x > 0$ 时, 函数值为 $+1$; $x < 0$ 时, 函数值为 -1 ; $x = 0$ 时, 函数值为 0 .

图 5 给出了在不同感染率下 PGP 和 AS 网络的最小 K -核节点度、介数、深度与实际传播能力的相关性. 从中可以发现节点的深度 $d(i)$ 比度和介数指标更能准确度量最小 K -核节点的传播能力, 而且随着传染率的增大, 最小 K -核节点深度的 τ 值越大, 这说明随着传染率的增大, 节点深度 $d(i)$ 更能准确预测节点的传播能力.

4 结论

真实网络中存在大量的 K -核值非常小的节点, 而传统的 K -核分解方法无法对这部分节点的传播能力进行度量. 节点的传播能力很大程度上取决于节点在网络中所处的位置, 因此 K -核很小的节点中也存在对网络的传播动力学产生巨大影响的节点. 度信息只能刻画局部信息, 介数等指标则需要全局信息, 并且计算量巨大, 不适用于大规模动态网络. 本文考虑最小 K -核节点的邻居信息, 利用邻居的

K -核信息提出了度量这部分节点传播能力的深度指标 $d(i)$. 对 PGP 和 AS 网络的实验结果表明, 最小 K -核节点的平均传播值和标准差随着感染概率的增大而增加, 该结果表明最小 K -核节点中存在部分对网络传播很重要的节点. 节点深度 $d(i)$ 指标与真实网络重要性的对比结果表明, 利用邻居信息的节点深度 $d(i)$ 比度和介数指标更能够准确地度量最小 K -核节点的传播能力.

本文利用邻居节点的 K -核信息对复杂网络中的最小 K -核节点的传播能力进行了分析, 该工作可以进一步扩展到考虑多级邻居的 K -核信息. 度量网络中节点的传播能力具有重要的现实意义和理论意义, 本文的工作有助于研究者认识网络中节点的位置与其传播能力之间的关系.

感谢上海理工大学复杂系统科学研究中心田鑫的交流与讨论.

-
- [1] Zhou T, Fu Z Q, Niu Y W, Wang D, Zeng Y, Wang B H, Zhou P L 2005 *Progress in Nature Science* **15** 513 (in Chinese) [周涛, 傅忠谦, 牛永伟, 王达, 曾燕, 汪秉宏, 周佩玲 2005 自然科学进展 **15** 513]
- [2] Li X, Liu Z H, Wang B H 2010 *Complex Systems and Complexity Science* **Z1** 34 (in Chinese) [李翔, 刘宗华, 汪秉宏 2010 复杂系统与复杂性科学 **Z1** 34]
- [3] Liu J G, Wang Z T, Dang Y Z 2006 *Mod. Phys. Lett. B* **20** 815
- [4] Balthrop J, Forrest S, Newman M E J, Williamson M M 2004 *Science* **304** 527
- [5] Keeling M J, Rohani P 2008 *Princeton Univ. Pr.* p10
- [6] Moreno Y, Nekovee M, Pacheco A F 2004 *Phys. Rev. E* **69** 066130
- [7] Zhang Y C, Liu Y, Zhang H F, Cheng H, Xiong F 2011 *Acta Phys. Sin.* **60** 050501 (in Chinese) [张彦超, 刘云, 张海峰, 程辉, 熊菲 2011 物理学报 **60** 050501]
- [8] Xiong X, Hu Y 2012 *Acta Phys. Sin.* **61** 150509 (in Chinese) [熊熙, 胡勇 2012 物理学报 **61** 150509]
- [9] Freeman L C 1977 *Sociometry* **40** 35
- [10] Goh K I, Oh E, Kahang B, Kim D 2003 *Phys. Rev. E* **67** 017101
- [11] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Nat. Phys.* **6** 888
- [12] Chen D B, Lü L Y, Shang M S, Zhang Y C, Zhou T 2011 *Physica A* **391** 1777
- [13] Zhou X, Zhang F M, Li K W, Hui X B, Wu H S 2012 *Acta Phys. Sin.* **61** 050201 (in Chinese) [周漩, 张凤鸣, 李克武, 惠晓滨, 吴虎胜 2012 物理学报 **61** 050201]
- [14] Boguna M, Pastor S R, Díaz G A, Arenas A 2004 *Phys. Rev. E* **70** 05612
- [15] Leskovec J, Kleinberg J, Faloutsos C 2005 *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD '05)* ACM, New York, August 21–24 2005 p177
- [16] Carmi S, Havlin S, Kirkpatrick S, Shavitt Y, Shir E 2007 *Proc. Natl. Acad. Sci. USA* **104** 11150
- [17] Pastor-Satorras R, Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200
- [18] Zhou T, Liu J G, Bai W J, Chen G R, Wang B H 2006 *Phys. Rev. E* **74** 056109
- [19] Zhang J, Zhao H, Yang B, Sun H L 2011 *J. Northeastern Univ. (Natural Science)* **32** 372 (in Chinese) [张君, 赵海, 杨波, 孙华丽 2011 东北大学学报(自然科学版) **32** 372]
- [20] Kendall M 1938 *Biometrika* **30** 8

Analysis of the spreading influence of the nodes with minimum K-shell value in complex networks*

Ren Zhuo-Ming Liu Jian-Guo[†] Shao Feng Hu Zhao-Long Guo Qiang

(Complex Systems Science Research Center, University of Shanghai for Science and Technology, Shanghai 200093, China)

(Received 27 September 2012; revised manuscript received 16 January 2013)

Abstract

The K-shell decomposition for identifying influential nodes plays an important role in analyzing the spreading performance in complex networks, which generates lots of nodes with the smallest K-shell value. The spreading influences of these nodes cannot be distinguished by the K-shell decomposition method, as well as the degree and betweenness indices. In this paper, by taking into account the K-shell information of the target node, we develop a new method to identify the node spreading ability with the minimum K-shell value. The experimental results for pretty good privacy and autonomous system networks show that the presented method could generate more accurate spreading list than the degree and betweenness indices.

Keywords: complex network, spreading performance, K-shell decomposition, minimum K-shell

PACS: 89.75.Da, 05.10.-a

DOI: 10.7498/aps.62.108902

* Project supported by the National Natural Science Foundation of China (Grant Nos. 71071098, 71171136, 91024026), Shanghai Rising-Star Program, China (Grant No. 11QA1404500), the Innovation Program of Shanghai Municipal Education Commission, China (Grant Nos. 11ZZ135, 11YZ110), the Key Project of Chinese Ministry of Education, China (Grant No. 211057), the Shanghai Leading Academic Discipline Project of China (Grant No. XTKX2012) and the Innovation Fund Project for Graduate Student of Shanghai, China (Grant No. 54-11-115-004).

[†] Corresponding author. E-mail: liujg004@ustc.edu.cn