

基于多阶邻居壳数的向量中心性度量方法*

王凯莉# 邬春学 艾均#† 苏湛

(上海理工大学光电信息与计算机工程学院, 上海 200093)

(2019年5月5日收到; 2019年7月17日收到修改稿)

K-壳分解法在度量复杂网络中节点的重要性方面具有重要的理论意义和应用价值. 但K-壳方法中, 存在大量壳值相等的节点, 从而无法精确地比较这些具有相同壳值节点的相对重要性. 因此, 本文基于网络中节点自身壳值与其多阶邻居的壳值, 设计利用向量的形式来表示节点在复杂网络中的相对重要性程度, 提出了多阶邻居壳数向量中心性方法, 并设计了该中心性向量比较方法. 通过在七个真实网络中进行消息传播与静态攻击实验, 发现基于多阶邻居壳数向量的中心性方法具有计算复杂度低, 能够有效发现具有高传播能力的节点, 在传播实验中具有优越的性能. 并在静态攻击实验过程中倾向于优先破坏网络中的传播核心结构. 多阶邻居壳数向量中心性方法在保留K-壳中心性信息的前提下, 极大提高了节点重要性的区别程度, 平衡了对节点在复杂网络中联通结构的重要性的度量和对传播结构重要性的度量, 因此具有重要理论意义与应用价值.

关键词: 中心性, 多阶邻居壳数向量中心性, 易感-感染传播模型, 静态攻击

PACS: 64.60.aq, 89.75.Hc, 89.75.-k

DOI: 10.7498/aps.68.20190662

1 引言

复杂网络理论可以用来研究多种类型的复杂系统, 这些系统中的主要元素被抽象成节点, 主要元素间的关系抽象成节点之间的连边. 近些年来, 科学界见证了复杂网络研究领域诞生了一系列的重要成果^[1,2]. 除了针对边进行预测应用在推荐系统等问题中之外^[3,4], 复杂网络模型里中心性研究也具有较高的应用价值^[5], 中心性是网络拓中度量节点重要程度的一个指标^[6], 对中心性的研究工作有着非常重要的理论意义, 除了搜索引擎对网页网络的排序外, 还有对恐怖分子和毒品贩卖等网络进行蓄意攻击^[7]、防止传染病在生物群体中进行传播^[8]、阻断谣言的传播^[9]以及引导有效信息在网络中迅速传播^[10]等等应用领域^[11]. 因此, 研究网络

中节点的重要性程度具有重要的理论意义和应用价值, 也是复杂网络研究中的热点问题之一^[12,13].

经典评估节点重要性的方法有度中心性 (Degree)^[14]、介数中心性 (BC)^[15]、接近中心性^[16]和特征向量中心性 (EC) 等. 度中心性仅考虑与该节点相连的邻居节点, 结果存在片面性, 而介数中心性和接近中心性都需要计算最短路径, 计算过程相对复杂. 于是在2010年 Kitsak 等^[17]提出了K-壳分解法 (K-shell), 他们认为将网络按层把节点移除, 越靠近壳心的节点重要性越高. K-壳分解法优点在于方法简单且计算复杂度低. 研究表明^[13], K-壳分解法的排序结果过于颗粒化, 使得很多节点处在同一壳层上, 节点的重要性区分度不大. 其次, K-壳分解法在网络分解时仅考虑剩余度的影响, 这相当于认为同一层的节点在外层的邻居数目对节点重要性并不重要, 这样得到的中心性是不合

* 国家自然科学基金青年科学基金 (批准号: 61803264) 资助的课题.

共同第一作者.

† 通信作者. E-mail: ajjun@outlook.com

理的.

国内外学者提出了基于 K-壳的改进方法. 罗仕龙等^[18]利用节点权重和权重分布重新定义边的扩散性, 提出适用于加权网络结构的基于冗余边过滤的 K-壳分解排序算法: filter-shell. Jiang 等^[19]受 K-壳分解方法去除外围节点后网络聚类系数会增大的启发, 选择 K-壳方法中壳值最高且相互连接的节点作为核心, 形成初始群来确定网络节点的重要性. Gong 和 Kang^[20]在 K-壳方法的基础上提出了一种新的社区网络识别方法, 该方法能够识别加速疾病传播的有影响力的传播者. 朱晓霞等^[21]为了对节点影响力给出具体排序, 在已有的各种最具影响力节点识别方法的基础上, 提出了一种基于社团结构和 K-shell 节点法的节点影响力识别方法. 其基本思想是利用某个节点处于不同社团的邻居节点的 ks 值判断节点影响力, 以识别 ks 值相同的节点的不同影响力. 李慧嘉^[22,23]等利用动态迭代技术, 提出了一种新型的社团探测技术, 能够准确而快速地识别网络中的社团结构. 首先引入一种动态系统, 可以使社团归属从随机状态逐步收敛到最优划分, 进一步利用严格的数学分析给出了社团归属在离散时间内收敛到最优的条件. Zhang 等^[24]根据度值和核值的差值来识别复杂网络中的重要节点. 在上述研究中讨论了对网络加权和引入社团等信息来评估节点的重要性, 但是尚未讨论目标节点的邻居节点对目标节点重要性排序的影响.

在本文的研究中, 为了提高网络中节点重要性的区分度, 同时考虑在相同条件下提高重要节点的影响力范围. 本文提出了多阶邻居壳数的向量中心性方法: multi-order K-shell vector(MKV). 通过对大量网络进行 SI 传播实验和静态攻击网络实验来验证 MKV 方法的有效性, 分别与度中性、K-壳中心性、介数中心性和特征向量中心性四种中心性方法进行比较. 实验结果表明, MKV 方法的计算复杂度低, 在对网络进行静态攻击时, MKV 方法能较大程度地破坏网络. 本文提出的中心性方法有效地提高了节点的重要程度的区分度, 并且通过传播实验分析发现该中心性方法具有较强的传播能力. MKV 方法能够发现具有较强传播能力的节点, 同时这些节点对网络造成的破坏也远胜于 K-壳方法. 通过对真实的网络进行实验, 对网络进行攻击时, MKV 方法能较大程度地破坏网络; 同时在网络中进行信息传播时, MKV 方法的传播能力相对

较强, MKV 方法可应用于实际网络中快速搜索具有传播能力强的节点.

2 基于多阶邻居壳数的向量中心性方法与现有经典方法

2.1 基于多阶邻居壳数的向量中心性方法

本文的算法是基于 K-壳分解法进行的, K-壳分解法是一个递归的过程, 具体的分解过程如下: 第一步, 找到网络 G 中所有度值为 1 的节点, 将度值为 1 的节点及其连边一并删除, 得到一个子网络 G1, 再选取子网络 G1 中所有度值为 1 的节点, 将该类节点及其连边删除, 直至子网络 G2 中不包含度值为 1 的子节点. 将这一步骤中删除节点的壳值 ks 定义为 1; 第二步, 删除子网络 G2 中度值为 2 的节点及其连边, 重复第一步的递归过程, 将这一步骤中删除节点的壳值 ks 定义为 2. 重复上述过程, 直到所有的点都分解到某个壳中^[12]. K-壳分解的示意图如图 1 所示.

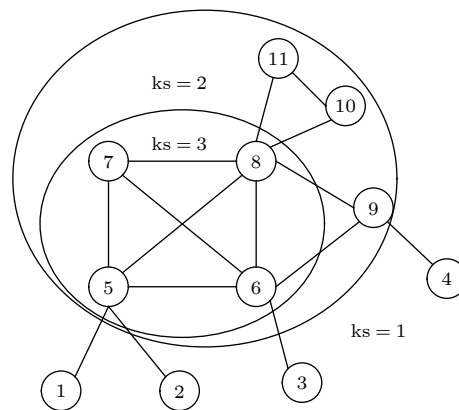


图 1 K-shell 分解法的示意图
Fig. 1. A diagram of the K-shell.

发掘二阶邻居的方法^[25]是: 1) 查找目标节点 v_i 的直接 (一阶) 邻居集合 N_i^1 ; 2) 查找 N_i^1 中包含的所有节点的邻居节点集 N_i^2 , 且满足 $N_i^1 \cap N_i^2 = \emptyset$. 以此类推, 得出节点 v_i 的多阶邻居.

设计该算法的思路是因为在 K-壳分解法中节点的区分度很低, 需要设计一种新的方法提高节点壳数所度量的区分度. 本文将节点自身的壳值与其多阶邻居的壳值构造一个向量, 把多阶邻居的壳数放入权值不同向量元素中, 通过一个向量而不是一个简单的数值来度量节点的重要性程度, 该向量代表这个节点在网络中能够波及到的其各阶邻居节

点的壳值的大小.

需要指出的是, 算法中多阶邻居的壳值, 指的是多阶低壳值邻居. 计算目标节点的多阶邻居过程中, 可能会出现部分壳值高于目标节点壳值的邻居, 在本文设计的算法中, 这些邻居并不统计.

对任意节点集合 V 中的节点 v_i , 该节点的多阶邻居节点集合为 N_i . 节点集合 N_i 中节点的数量为 n_i . 设对每个节点 v_i 都存在一个 m 维 (每个网络中节点最大的壳值是不同的, 此处的 m 是随着网络最大壳值的变化而变化的) 向量 \mathbf{x}_i , 可将其表示为 $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{im})^T$. 每一个向量 \mathbf{x}_i 中的元素 x_{ij} 可以通过如下的式子来计算 $C_j(M_j(v_k))$, $x_{ij} = \sum_{k=1}^{n_i} C_j(M_j(v_k))$, $j = 1, 2, \dots, m$, 其中 $M_j(v_k)$ 是一个给定的中心性指标, 在本文的情况下, 是节点 v_k 的 K-shell 中心性, $v_k \in N_i$. $C_j(\xi)$ 是一个如下给出的函数.

$$C_j(\xi) = \begin{cases} 0, & \text{if } \xi \neq j \\ 1, & \text{if } \xi = j \end{cases} \quad (1)$$

记 $C_d(v_i) = \sum_{j=0}^{n_i} x_{ij}$. 那么, x_i 就是节点 v_i 的邻居向量. MKV 算法的流程图如图 2 所示.

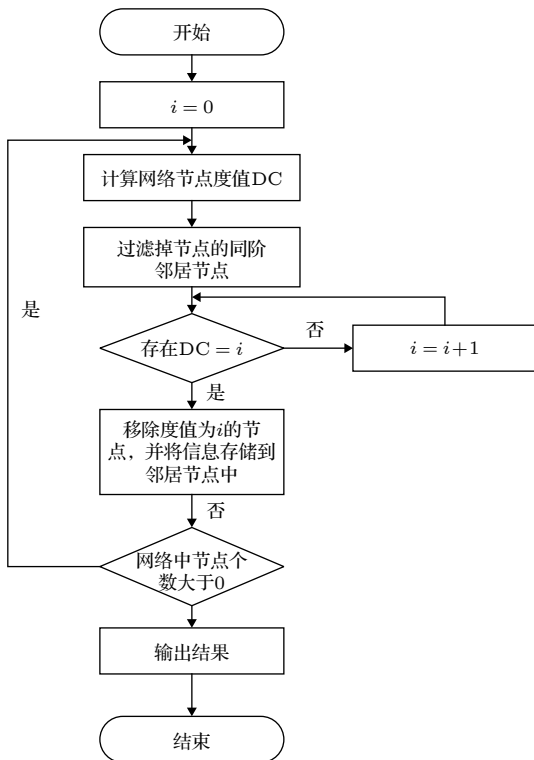


图 2 MKV 算法实现流程图

Fig. 2. MKV algorithm implementation flow chart.

通过流程图可以看出, 算法通过不断迭代, 将低壳数节点的信息记录在对应的高壳数邻居的向量中, 故此最终形成的向量, 带有该节点低壳数邻

居的信息, 而阶数在同一网络中固定不变, 在不同网络中则不尽相同.

基于多阶邻居壳数的向量中心性构造的基本思想是本文希望将一个简单的壳值按照某种方式展开成一个向量. 一方面, 该方法可以使得节点的重要性更具区分度, 另一方面, 也可以通过对每个节点的自身及其多阶邻居的综合考察来考虑该节点在网络中的重要程度. 这样, 对于向量中的每个元素, 用一个数字来表示目标节点有多少个多阶邻居节点具有某一壳层的壳值. 对于节点 v_i , 向量 x_i 中元素 x_{i1} 保存的是目标节点 v_i 多阶邻居中壳数为 1 的节点总数, x_{i2} 保存的是目标节点 v_i 多阶邻居中壳数为 2 的节点总数, 以此类推, x_{im} 表示节点 v_i 目标节点自身的壳数. 因此, 目标节点 v_i 的多阶邻居通过向量中的不同元素将其考虑成具有几个不同壳值等级, 这样可以直观地了解节点 v_i .

以图 1 中的节点 8 和节点 9 为例来解释向量的含义, 节点 8 的展开向量为 $x_8 = (1, 3, 1)^T$, 表示节点 8 的多阶邻居中壳值为 1 的节点个数有一个, 是节点 4, 壳值为 2 的节点数有 3 个, 分别是节点 9、节点 10 和节点 11, 节点 8 自身的壳数是 3; 节点 9 的展开向量为 $x_9 = (1, 1, 0)^T$, 该向量表示节点 9 的多阶邻居中壳值为 1 的节点存在一个, 是节点 4, 节点 9 自身的壳数为 2, 不存在壳数为 3 的多阶邻居.

多阶邻居的壳数向量能够清楚地表明该节点在整个网络中的潜在影响力, 不仅表明了节点的层次性, 还可以表明该节点在多阶邻居中的影响力. 在 K-shell 方法中节点壳值相同时, 节点的影响力往往是不同的, 但是本文的方法可以度量出上述节点的重要性.

2.2 多阶邻居壳数向量的比较方法

给定任意的两个向量 $\mathbf{x}_q \in X$ 与 $\mathbf{x}_p \in X$, 记 $\varepsilon = \mathbf{x}_q - \mathbf{x}_p = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m)^T$. 使

$$\varepsilon^* = \begin{cases} \varepsilon_m, \varepsilon_m \neq 0 \\ \varepsilon_i, \varepsilon_{i+1} = \dots = \varepsilon_m = 0, \varepsilon_i \neq 0. \\ \varepsilon_1, \varepsilon_2 = \dots = \varepsilon_m = 0 \end{cases}$$

一个给定的二元关系 \prec 如下: 如果 $\varepsilon^* < 0$, 那么 $\mathbf{x}_q \prec \mathbf{x}_p$, 对应地, \mathbf{x}_q 的迭代壳数的向量中心性小于 \mathbf{x}_p . 同理, 如果 $\varepsilon^* > 0$, 那么 $\mathbf{x}_q \succ \mathbf{x}_p$, 多阶邻居壳数向量 \mathbf{x}_q 大于 \mathbf{x}_p . 同理, 如果 $\varepsilon^* = 0$, 那么 $\mathbf{x}_q = \mathbf{x}_p$, 也就是说多阶邻居壳数向量 \mathbf{x}_q 等于 \mathbf{x}_p .

根据上述基于多阶邻居壳数向量的比较的定

义, 向量比较算法流程图如图 3 所示.

本文下面将进一步给出说明, 来证明本章的方法可以将目标网络中的所有节点进行有效的排序, 即, 向量集合 S 对于给定的二元关系 \prec 是严格的全序集 [26].

定理: 二元关系 \prec 使集合 S 是严格全序集, 等价于, 二元关系 \prec 满足如下条件:

- 1) $\forall x, y \in S, x \prec y, y \prec x, x = y$ 三个式子中, 仅有一个成立 (三分性);
- 2) $\forall x, y \in S$, 如果 $x \prec y$, 那么 $y \prec x$ 不成立 (不对称性);
- 3) $\forall x, y, z \in S$, 如果 $x \prec y$ 并且 $y \prec z$, 那么 $x \prec z$ 成立 (传递性).

证明如下:

首先, 对任意的 $x, y \in S$, 使得 $\xi = x - y$. 那么一定有 $\xi_* < 0, \xi_* > 0$ 或 $\xi_* = 0$ 三者当中一个关系成立. 因此, 基于二元关系 \prec 的定义, 二元关系 \prec 是三分的.

其次, 如果 $x \prec y$, 那么 $\xi_* < 0$. 所以, $\xi_* > 0$ 一定不成立, 也就等价于 $y \prec x$ 不成立.

最后, 对于任意给定的 $x, y, z \in S$, 使得 $\alpha = x - y, \beta = y - z$, 那么 $\xi = \alpha + \beta$, 于是 $x - z = (x - y) - (y - z) = \alpha + \beta = \xi$. 基于二元关系 \prec , 如果 $x \prec y$, 那么 $\alpha_f < 0$ 或者 $\alpha_f = 0$ 或者 $\alpha_m < 0$ 中的某一个成立. 类似地, $y \prec z$ 意味着 $\beta_f < 0$ 或者 $\beta_f = 0$ 或者 $\beta_m < 0$ 中的某一个成立. 因此, $\xi_* = \{\xi_f | \xi_f \neq 0\} \cup \{\xi_m | \xi_f = 0\} < 0$ 并且 $x \prec z$. 综上所述, 二元关系 \prec 具有传递性.

通过上面的证明, 可以看出多阶邻居向量和二

元关系 \prec 的定义保证向量集合 S 是严格全序的. 因此, 本文给出的定义也就能完成对所有节点重要性的排序.

2.3 相关常用中心性方法

2.3.1 度中心性 (Degree Centrality)

度中心性是指一个节点的度越大就意味着这个节点越重要. 一个包含 N 个节点的网络图中, 节点度的中心性值定义为

$$DC_i = \frac{k_i}{N - 1}, \quad (2)$$

其中 k_i 表示节点 i 的度值, $N - 1$ 是指节点最大可能的度值为 $N - 1$.

2.3.2 介数中心性 (Betweenness Centrality)

介数中心性是指经过某个节点的最短路径的数目来刻画节点重要性的指标. 将介数定义为

$$BC_i = \sum_{s \neq i \neq t} \frac{n_{st}^i}{g_{st}}, \quad (3)$$

g_{st} 是从节点 s 到节点 t 的最短路径数目, n_{st}^i 是从节点 s 到节点 t 的 g_{st} 条最短路径中经过节点 i 的最短路径条数.

2.3.3 特征向量中心性 (Eigenvector Centrality)

一个节点的重要性既取决于其邻居节点的数量 (即该节点的度), 也取决于其邻居节点的重要性, 记 x_x 为节点 i 的重要性度量值, 则:

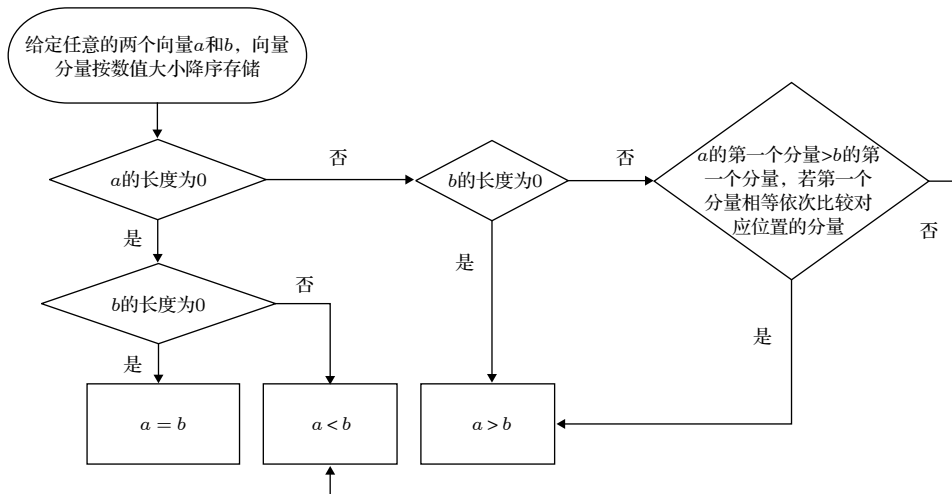


图 3 向量比较大小算法流程图

Fig. 3. Flow chart of vector comparison size algorithm.

$$x_x = c \sum_{j=1}^N a_{ij} x_j, \quad (4)$$

其中 c 为一个比例常数, 记 $\mathbf{x} = [x_1, x_2, x_3, \dots, x_n]^T$, 经过多次迭代到达稳态时, 可以写成如下的矩阵形式:

$$\mathbf{x} = cA\mathbf{x}. \quad (5)$$

2.4 基于多阶邻居壳数的向量中心性方法时间复杂度分析

假设图 G 中节点个数为 N , 边的数量为 M , 度中心性的计算复杂度在稠密矩阵中为 $O(N^2)$, 在稀疏矩阵中为 $O(M)$; 早期的介数中心性的时间复杂度为 $O(N^3)$ ^[27], Brandes^[28] 改进的介数中心性时间复杂度为 $O(NM)$; EC 算法的时间复杂度为 $O(N^2)$; K-shell 的时间复杂度是 $O(N)$.

本文提出的 MKV 方法要对网络中的每个节点的信息进行记录, 而对其中一个节点的邻居进行遍历的时候, 效率最差的情况就是 B 树的情形, 这种情形下时间复杂度是 $O(\log N)$, 网络中节点总数是 N , 所以 MKV 的时间复杂度是 $O(M \log N)$.

当变量 N 增大时, 时间复杂度的排序是 $O(1) < O(\log N) < O(N) < O(N \log N) < O(N^2) < O(N^3) < O(2^n)$. MKV 比 BC, EC 方法的时间复杂度小, 表明 MKV 方法的算法效率更高; 与 Degree, K-shell 方法相比时间复杂度相对较低, 但是其节点包含的信息要远高于这两种方法.

3 实验数据与分析

3.1 实验数据

为了证明本文设计的多阶邻居壳数的向量中心性方法的有效性和适用性, 本文选用了七个经典

的复杂网络模型: 由小说《悲惨世界》中人物关系构成的悲惨世界复杂网络 (简记为 LesMiserables)^[29], 美国西部电力网拓扑结构构成的电力网 (简记为 PowerGrid)^[30], Rovira I Virgili 大学 (塔拉戈纳) 成员之间电子邮件交换组成的网络 (简记为 Email)^[31], 2006 年以前在网络科学领域所有科学家的论文合作网络数据集 (Coauthorships in network science, 简记为 Coauthor)^[32], 生活在新西兰两个街区的海豚之间的网络 (Dolphin social network, 简记为 Dolphin)^[33], 大肠杆菌代谢网络 (简记为 C. Elegans)^[34] 以及著名的空手道俱乐部数据 (Zachary's karate club, 简记为 Club)^[35]. 上述这七个复杂网络来源于大量的复杂网络研究文献, 网络的具体信息如表 1 所列.

3.2 中心性的区分度实验

对于任意一种给定的中心性方法, 求得的节点中心性在数值上会存在一致的, 这就代表中心性的值是重复的. 在理论上, 具有相同中心性数值的节点的重要程度是没有办法区分的. 而在一个有 N 个节点的网络中, 中心性的重复度的定义为 $RR = r/N$, 其中 r 表示该中心性方法中重复的中心性值的个数, 而中心性的区分度则可以定义为 $DR = (N - r)/N$ ^[26]. 对本文用到的七个网络的中心性区分度分析计算得到结果如表 2 所列.

根据中心性区分度的定义, 可以得出一个结论: 中心性指标的区分度 DR 值越大, 那么这就意味着网络中节点重要程度的区分度就越高. 通过表 1 得出, MKV 方法的 DR 值要远远高于 K-shell 和度值中心性 Degree, 虽然本文方法的中心性指标的区分度 DR 比 BC 和 EC 小, 但是与 BC 的 DR 值已经很接近. 由此可以得出, MKV 方法明显地提高了节点重要性的区分度.

表 1 本文中用到的几个网络基本信息

Table 1. Several basic network information used in this paper.

网络	节点数量	边数量	平均度值	聚集系数	平均路径长度	同配系数
LesMiserables	77	254	6.597	0.736	2.641	-0.4756
PowerGrid	4941	6594	2.669	0.107	18.989	0.4616
Email	1134	6556	11.563	0.526	1.992	-0.0436
Coauthor	1589	2742	3.451	0.878	5.823	0.0035
Dolphin	62	159	5.129	0.303	3.357	-0.1652
C.Elegans	453	4596	9.007	0.657	2.664	-0.1085
Club	34	78	2.29	0.588	2.408	-0.2145

表 2 中心性的区分度在不同网络中的取值情况 (越大越好)
Table 2. Value of Centrality Distinction in Different Networks(the larger, the better).

DR	BC	EC	Degree	K-shell	MKV
C.Elegans	66.225%	88.962%	8.830%	2.208%	40.839%
Club	61.765%	79.412%	32.353%	11.765%	47.059%
Coauthors	9.880%	29.264%	1.447%	0.692%	9.880%
Dolphin	87.097%	96.774%	19.355%	6.452%	70.968%
Email	62.346%	94.533%	4.586%	0.970%	53.263%
Power	59.279%	86.217%	0.324%	0.101%	8.561%
LesMiserables	41.558%	67.532%	23.377%	10.390%	37.662%

3.3 基于多阶邻居壳数的向量中心性的复杂网络的静态攻击

本文通过蓄意对网络进行静态攻击并观察网络的形态, 以此来验证评估节点的重要性. 具体过程如下:

- 1) 计算每种中心性的度量值, 按照评估节点重要性算法的度量大小选取网络中重要性最高的 n 个节点;
- 2) 蓄意地逐个移除网络中选中的重要节点及其连边;
- 3) 计算并估计上述的攻击过程对网络造成的影响;
- 4) 将删除的节点数量从 0.1% 增加到 30%, 重复上述过程, 观察网络的状态, 并做出相关记录和分析.

在对复杂网络进行静态攻击的仿真实验中, 网络受到攻击后会破碎成零散的连接部件, 这些连接部件的数量表明了网络的破碎程度. 而网络中的最大连通子团可以表示剩余网络的网络连通性. 所以在评判网络的连通性时既要考虑子团的数量, 也需要考虑最大连通子团的节点数量. 美国电力网 power 在蓄意攻击后的最大连通子团以及子团数量的变化情况如图 4 和图 5 所示.

最大连通子团的节点数量越少表明网络的连通性越差, 在网络中移除的节点越重要. 在图 4 中, 在删除节点数量到 2% 的时候 MKV 方法中最大连通子团的节点数量变化趋势接近于除了 Degree 外的几种方法, 而删除节点大于 2% 时本文方法的最大子团数量迅速减小, 相较于 EC 和 K-shell 两种方法是有一定优势的, 当删除节点达到 15% 后, MKV 方法的结果要优于其他方法. 同样的在图 5 中, 子团数量在删除节点小于 7% 时, MKV 方法和 K-shell 方法结果很相近, 而删除节点高于 13%

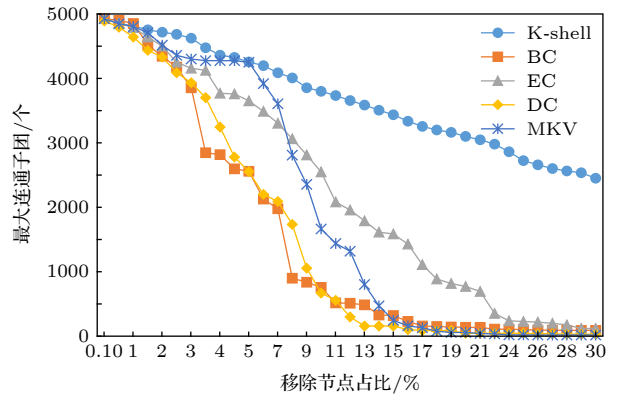


图 4 美国电力网 power 在静态攻击实验中最大连通子团变化情况

Fig. 4. Largest connected component during static attack experiment on power.

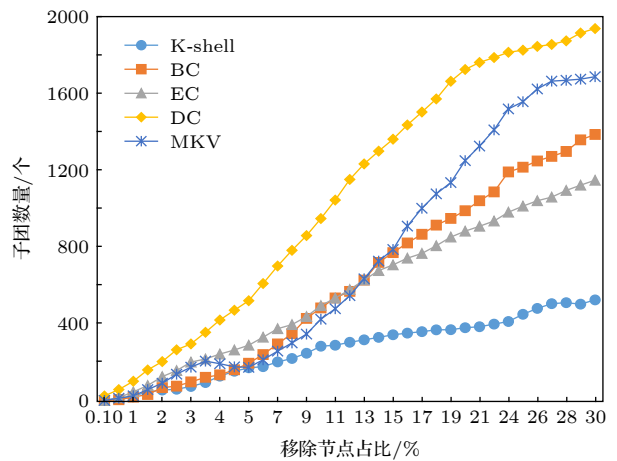


图 5 美国电力网 power 在蓄意攻击实验中子团数量变化情况

Fig. 5. Total component number during static attack experiment on power.

时, MKV 方法效果明显优于 EC, BC 和 K-shell 方法, 与 Degree 方法的结果趋于一致. Degree 方法在移除节点初期, 最大连通子团节点数量变化明显的原因是, 按 Degree 方法选择的重要节点移除时, 节点的邻居数量相对其他方法更多, 移除重要

节点及其连边后最大连通子团的节点数量会锐减,同时子团的数量会迅速上升.在 power 网络的实验中,攻击开始初期, MKV 方法的效果不及 EC, BC 和 Degree,但是后期的效果是具有优势的,可见,对网络的蓄意攻击, MKV 方法尽可能考虑的是长期收益,而非短期收益.

为了量化度量攻击实验与节点重要性之关系,本文对上述七个网络中选取移除节点的不同时间段的最大连通子团节点数量和子团数量进行了统计.选取 K-shell 方法为基准方法,根据公式 $rl = a/b$ 求出七个网络在这些时间段的最大连通子团数量的平均值和最大值,此处的 b 表示 K-shell 最大连通子团节点的数量, a 分别代表其他四种中心性方法对网络蓄意攻击同时刻的最大连通子团节点的数量,则 rl 表示其他方法攻击网络核心结构时最大连通子团比 K-shell 方法提升了多少.同理,子团数量根据公式 $rb = c/d$ 求出其他方法在子团数量上比 K-shell 方法优化多少,再统计其平均值和最小值,此处的 d 代表 K-shell 中心性

方法对网络蓄意攻击同时刻的连通子团的数量, c 表示其他方法得到的连通子团的数量,则 rb 表示其他方法的子团数量对比 K-shell 方法的提升.统计结果如图 6 和图 7 所示.

在图 6 中,以 K-shell 方法得到的结果为基准,其他方法分别与 K-shell 方法比较,图 6 中的值越小越好,在不同的网络中 MKV 方法要优于 K-shell 方法,与其他方法相比效果不明显,但是与度值相比, MKV 方法提高了节点重要性的区分度,与 BC, EC 方法相比,在时间复杂度上得到了有效的提升. power 网络不是一个社交网络,其平均路径长度较大且聚类系数较低,在这个网络中最大连通子团的效果相对是有提升的.因此对 power 这种类型的网络进行蓄意攻击时, MKV 方法更具优势.

在图 7 中,值越大表明网络受到蓄意攻击时得到的子团数量越多,网络越破碎,该度量节点重要性的方法就越具备优势.从图 7 中看到五种评估节点重要性的方法得到的子团数量的最终结果, MKV 方法得到的子团数量的结果要完全优于 K-

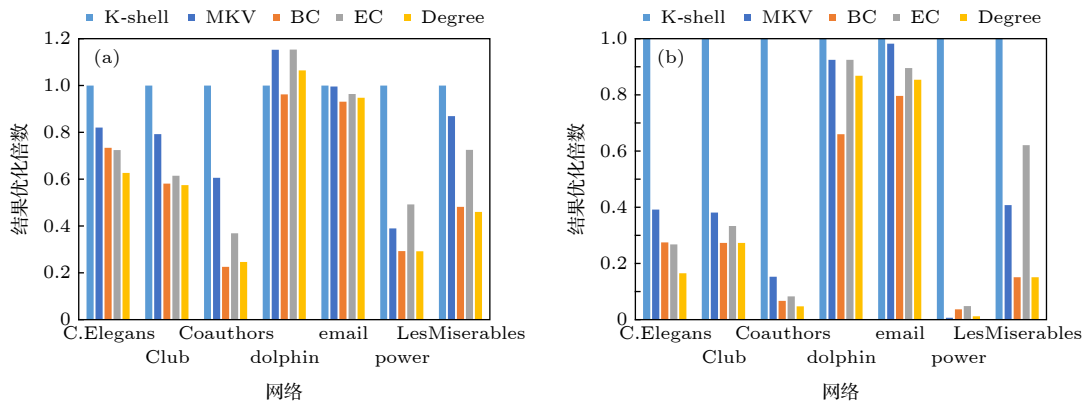


图 6 七个网络受到蓄意攻击中最大连通子团的统计情况(越小越好) (a) 最大连通子团平均值情况; (b) 最大连通子团最大值情况
Fig. 6. Largest connected component during static attack experiments on the seven networks(the smaller, the better).

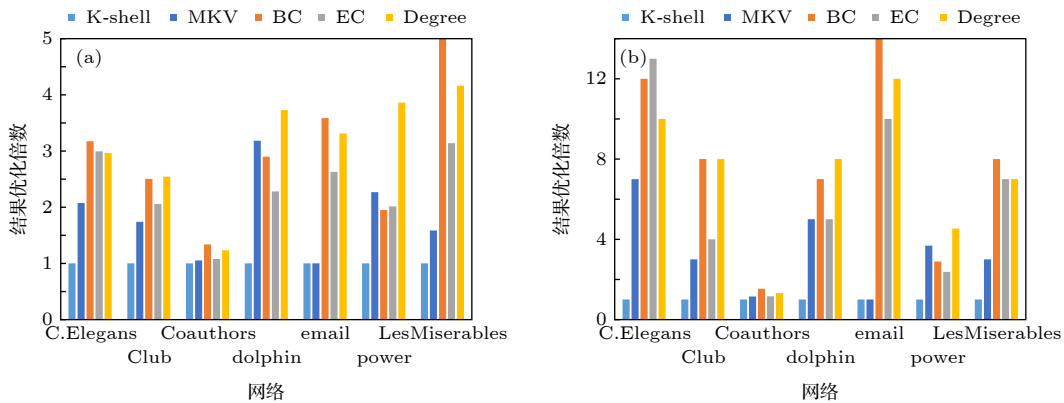


图 7 七个网络静态攻击重要节点的子团数量的统计情况(越大越好) (a) 子团数量平均值情况; (b) 子团数量最大值情况
Fig. 7. The number of components during static attack experiments on the seven networks (the larger, the better).

shell 方法. MKV 方法的效果虽不及其他几种方法,但是在时间复杂度上要优于 BC 和 EC,在节点重要性的区分度上要优于度值中心性. 在 power 网络和 dolphin 网络中,子团数量的效果占很大优势,原因是这两个网络的聚集系数较低且平均路径长度较长,在这类网络中, MKV 方法得到的结果占优势. 综合上述图片的内容,在对不同网络进行蓄意攻击时, MKV 方法明显优于 K-shell 方法. 对于低聚类系数跟高平均路径长度的网络, MKV 方法是占优势的,且效果相对明显. MKV 方法虽不是平均降低网络连通性的最佳选择,但是 MKV 方法计算复杂度相对较低且倾向于优先破坏网络中的传播核心结构.

3.4 基于多阶邻居壳数的向量中心性的复杂网络传播实验

本文采用 SI(susceptible-infected) 模型^[36]来仿真网络中重要节点的传播行为, SI 模型将传染范围内的人群划分为两类: 1) S 类: 易感人群, 这类人是指还未染病但是与感染病患者接触后以一定的概率 α 受到感染; 2) I 类: 染病人群, 这类人是指已经患病的人群. SI 传染过程如图 8 所示.

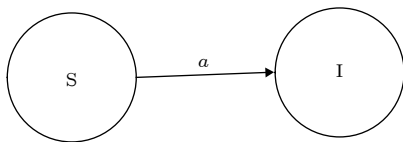


图 8 SI 传播模型图

Fig. 8. SI propagation model diagram.

计算每种中心性的度量值,将得出的度量值按数值降序排列,选取网络中每种中心性度量值最大的节点设置为感染节点 I ,该节点以感染比率 α 去感染易感节点,将该传染过程迭代 M 次,统计感染节点的数量.并用所求得感染节点数量来度量所选初始节点在网络中的传播能力.

以美国电力网 power 为例,选取复杂网络中按评估节点重要方法度量值最高的节点设置为感染节点 I ,感染节点以 $\alpha = 0.001$ 的概率去感染易感节点 S ,将这个过程迭代 1000 次,统计出迭代过程结束后的感染节点数量,选取部分迭代次数的感染节点的数量绘制统计图.美国电力网 power 的不同迭代次数感染节点统计图和大肠杆菌代谢网络 C.Elegans 不同迭代次数感染节点统计图如图 9

和图 10 所示.

在 SI 模型的实验中,传播感染的节点数量越多,证明该节点在网络中的重要程度越高.如图 9 中结果显示的那样,多阶邻居壳数的向量中心性 MKV 在美国电力 power 网络中传播的效果是第二优的,通过分析得到 EC 算法的时间复杂度为 $O(N^2)$,而 MKV 方法是时间复杂度为 $O(M \log N)$,在相同的条件下, MKV 的传播效果虽然比 EC 的传播效果差,但是 MKV 方法要比 EC 的计算代价低.而在图 10 大肠杆菌代谢网络 C.Elegans 的 SI 传播实验中,迭代壳数的向量中心性 MKV 方法要优于其他四种方法.因为本文的目的是验证哪个中心性参数在度量节点重要程度上有最好的结果,所以本文试图在这几种中心性传播过程中能感染最多的节点.为了量化这一目标,本文从上述七

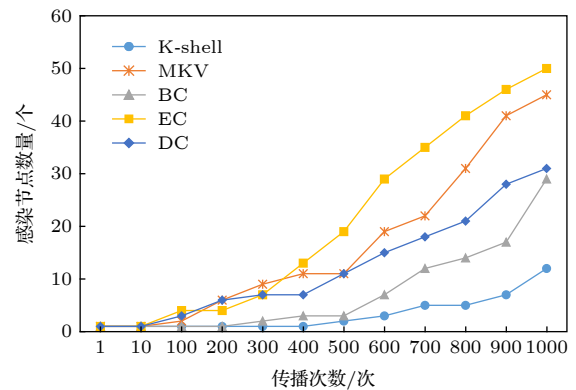


图 9 美国电力网 power 在 SI 传播模型下感染节点的变化情况

Fig. 9. Change of infected nodes in Power of American Electric Power network under SI propagation model.

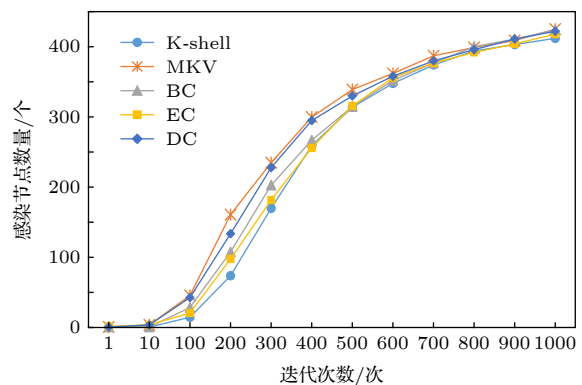


图 10 大肠杆菌代谢网络 C. Elegans 在 SI 传播模型下感染节点的变化情况

Fig. 10. Changes of infection nodes in C. Elegans network under SI propagation model.

个网络 SI 传播实验中选取不同迭代次数的感染节点数对其统计. $r_{si} = p/q$, q 表示 K-shell 方法感染节点数量, p 表示其余中心性方法感染节点数量,

r_{si} 表示其他方法感染的节点数量要比 K-shell 方法感染节点的数量高多少, 分别求出这些时间段的平均值和最大值. 统计结果如图 11 所示.

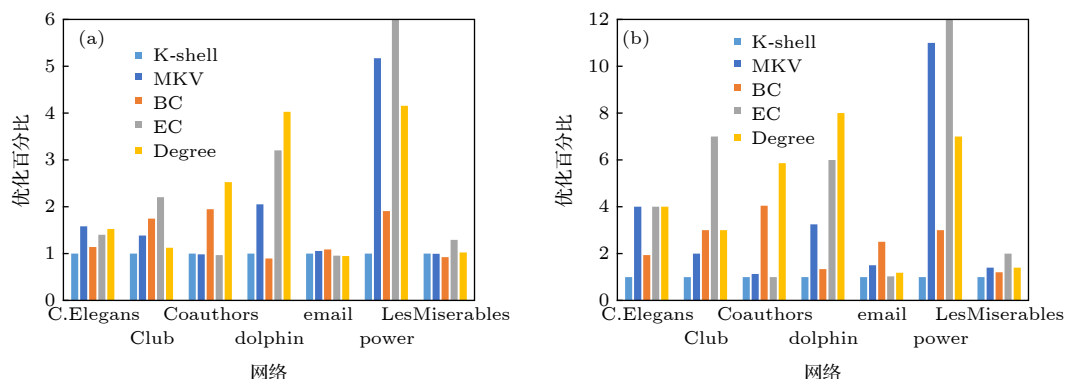


图 11 七个网络传播感染节点的统计情况 (越大越好) (a) SI 传播节点数量比较平均值; (b) SI 传播节点数量比较最大值
Fig. 11. The statistical result of infected nodes of seven network (the larger, the better).

对网络中进行 SI 传播实验时, 平均值和最大值的数值越大表明传播感染节点的效果越明显. 在图 11 中, 值越大表明感染的节点数量越多, 度量节点重要性的方法就越好. 在上述七个网络中, MKV 方法得到的感染节点的数量与 K-shell 方法相比领先的比率是明显的. 在 C.Elegans 这个网络中, MKV 方法的传播效果要完全优于其他方法, 在 power 网络中 MKV 方法的传播效果仅略低于 EC 方法传播的效果, 除此之外, 感染节点的数量要远高于其他方法的平均值和最大值, 在 Coauthors 网络和 LesMiserables 中 MKV 方法所占优势不太明显, 原因是这两个网络的平均度值相对较大且网络节点连接密集, 对该类型网络进行感染传播时, MKV 方法的优势不太明显. 因此在传播感染的实验中, MKV 在传播过程中在大多数网络中优于 K-shell 方法, 在半数网络中传播的效果要优于 BC 和 Degree. MKV 方法与度值相比, 提高了节点重要性的区分度, 而与 BC, EC 方法相比, 降低了时间复杂度, 所以 MKV 方法能够快速有效地发现具有强传播力的节点, 是网络传播信息的最佳选择. 对于少数平均度值高、网络连接密集的网络 MKV 方法的传播优势不太明显; 但是能够在大多数网络中发现具有强传播能力的节点, MKV 方法是信息传播的最佳选择.

4 结 论

本文提出了多阶邻居壳数的中心性方法,

MKV 方法将节点自身及多阶邻居节点的重要性考虑在内, 并通过对七个不同类型真实网络进行 SI 传播和蓄意攻击网络两种实验, 与其他四种常用中心性方法进行了比较.

结果表明: 本文设计的 MKV 方法基于 K 壳中心性, 还保留了 K 壳中心性对节点在网络中层次的信息, 同时展示出节点低阶壳数邻居的分布情况. 此外, MKV 方法具有高区别度、计算复杂度仅略高于度值, 但比度值提供了更多网络结构的信息; 在传播效率上, 仅次于特性向量中心性, 计算却更快捷; 在攻击实验过程中绝大多数情况下优于 K 壳中心性, 证明 MKV 方法在度量节点重要性程度上具有一定特点与部分优势.

参考文献

- [1] Barabasi A L, Albert R 1999 *Science* **286** 509
- [2] Newman M, Girvan M 2004 *Phys. Rev. E* **69** 423
- [3] Ai J, Su Z, Li Y, Wu C X 2019 *Physica A* **527** 121155
- [4] Ai J, Liu Y Y, Su Z, Zhang H, Zhao F Y 2019 *Europhys. Lett.* **126** 38003
- [5] Brin S, Page L 2012 *Computer Networks* **56** 3825
- [6] Newman M 2010 *Networks: An introduction* (Oxford: Oxford University Press) p327
- [7] Cohen R, Erez K, Ben-Avraham D, Havlin S 2001 *Phys. Rev. Lett.* **86** 3682
- [8] Keeling M J, Rohani P, Pourbohloul B 2008 *Clinical Infectious Diseases* **47** 864
- [9] Moreno Y, Nekovee M, Pacheco A F 2004 *Phys. Rev. E* **69** 066130
- [10] 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]
- [11] Li H, Gao G, Chen R 2019 *Int. J. Software Engineer.*

- Knowledge Engineer.* **29** 93
- [12] Liu J G, Ren Z M, Guo Q, Wang B H 2013 *Acta Phys. Sin.* **62** 178901 (in Chinese) [刘建国, 任卓明, 郭强, 汪秉宏 2013 物理学报 **62** 178901]
- [13] Ren X L, Lü L Y 2014 *Chin. Sci. Bull.* **59** 1175 (in Chinese) [任晓龙, 吕琳媛 2014 科学通报 **59** 1175]
- [14] Chen D, Lin Y L, Shang M S 2012 *Fuel and Energy Abstracts* **391** 1777
- [15] Freeman L C 1977 *Sociometry* **40** 35
- [16] Opsahl T, Agneessens F, Skvoretz J 2010 *Social Networks* **32** 245
- [17] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Nat. Phys.* **6** 888
- [18] Luo S L, Gong K, Tang C S, Zhou J 2017 *Acta Phys. Sin.* **66** 188902 (in Chinese) [罗仕龙, 龚凯, 唐朝生, 周靖 2017 物理学报 **66** 188902]
- [19] Jiang L C, Zhao X, Ge B, Xiao W, Ruan Y 2019 *Physica A* **516** 58
- [20] Gong K, Kang L 2018 *J. Syst. Sci. Inf.* **6** 366
- [21] Zhu X X, Zhao X, Liu M M 2017 *Application Research of Computers* **34** 2582 (in Chinese) [朱晓霞, 赵雪, 刘萌萌 2017 计算机应用研究 **34** 2582]
- [22] Li H J, Yan G, Liu Z D, Li G J, Zhang X S 2017 *Scientia Sinica Mathematica* **47** 16 (in Chinese) [李慧嘉, 严冠, 刘志东, 李桂君, 章祥荪 2017 中国科学: 数学 **47** 16]
- [23] Li H J, Li A H, Li H Y 2017 *Chinese Journal of Computers* **40** 15 (in Chinese) [李慧嘉, 李爱华, 李慧颖 2017 计算机学报 **40** 15]
- [24] Zhang J P, Xu H, Yang J, Lum L J 2018 *ICPCSEE* Zhengzhou, China, September 21–24, 2018 p28
- [25] Wang H, Zhang Z, Li L, Wang M 2016 *Acta Electronica Sinica* **44** 2330 (in Chinese) [王浩, 张赞, 李磊, 汪萌 2016 电子学报 **44** 2330]
- [26] Ai J, Zhao H, Carley K M, Su Z, Li H 2013 *Eur. Phys. J. B* **86** 163
- [27] Freeman L C 1979 *Social Networks* **1** 215
- [28] Brandes U 2001 *Math. Sociology* **25**
- [29] E.Knuth D 1993 *The Stanford GraphBase: A Platform for Combinatorial Computing* (Vol.1)
- [30] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [31] Guimera R, Danon L, Diaz-Guilera A, Giralt F, Arenas A 2003 *Phys. Rev. E* **68** 065103
- [32] Newman M 2006 *Phys. Rev. E* **74** 036104
- [33] Lusseau D, Schneider K, Boisseau O, Haase P, Slooten E, Dawson S 2003 *Behav. Ecol. Sociobiol.* **54** 396
- [34] Duch J, Arenas A 2005 *Phys. Rev. E* **72** 027104
- [35] Zachary W W 1977 *J. Anthropol. Res.* **33** 452
- [36] Wang X F, Li X, Chen G R 2012 *Network Science: an Introduction* (Beijing: Higher Education Press) pp306–307 (in Chinese) [汪小帆, 李翔, 陈关荣 2012 网络科学导论 (北京: 高等教育出版社) 第306—307页]

Complex network centrality method based on multi-order K-shell vector^{*}

Wang Kai-Li[#] Wu Chun-Xue Ai Jun^{#†} Su Zhan*(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)*

(Received 5 May 2019; revised manuscript received 17 July 2019)

Abstract

The K-shell has important theoretical significance and application value in measuring the importance of nodes in complex networks. However, in the K-shell method, most of nodes possess an identical K-shell value so that the relative importance of the identical K-shell nodes cannot be further compared with each other. Therefore, based on the K-shell value of nodes in the complex network and the K-shell values of multi-order neighbors in complex networks, in this paper we use the vectors to represent the relative importance of node in each of complex networks, which is named multi-order K-shell vector. Multi-order K-shell vector centrality defines a vector indicating the number of multi-order neighbors with different K-shells and groups them into elements of the vector. Each vector infers to not only the original K-shell of the given node but also the number of its multi-order neighbors and their K-shell values, which indicates the propagation capability of the given node. An approach to comparing two multi-order K-shell vectors is also presented, which is used to sort the vectors to evaluate the node importance. The method is explored by comparing several existing centrality methods. Through the experiments of SI propagation and static attack experiments in seven real-world networks, it is found that multi-order K-shell vector centrality provides low computational complexity, effectively evaluates nodes with high propagation capability, which confirms the improved performance in susceptible infected model propagation experiments. On the other hand, the static attack experiments show that the multi-order K-shell vector tends to preferentially select the core structure with powerful propagation capability in the network. The multi-order K-shell vector greatly improves the difference rate of node centrality under the premise of preserving the K-shell structure information, as well as balancing the importance measure of nodes in the complex network and the structure evaluation of propagation capability. The multi-order K-shell vector is not appropriate for all types of networks when considering the result of network attacking. For the networks with low clustering coefficients and high average path lengths, multi-order K-shell vector method is dominant and the effect is relatively obvious. By contrast, multi-order K-shell vector surpasses most of centrality approaches when spreading information is our priority. In a few networks, eigenvector centrality presents a slightly better performance with a larger computational complexity. The proposed centrality measure is therefore of great theoretical and practical importance.

Keywords: centrality, multi-order K-shell vector, susceptible infected model, static attack**PACS:** 64.60.aq, 89.75.Hc, 89.75.-k**DOI:** 10.7498/aps.68.20190662

^{*} Project supported by the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 61803264).

[#] These authors contributed equally.

[†] Corresponding author. E-mail: aijun@outlook.com