

加权网络中基于冗余边过滤的 k -核分解排序算法

罗仕龙 龚凯 唐朝生 周靖

A ranking approach based on k -shell decomposition method by filtering out redundant link in weighted networks

Luo Shi-Long Gong Kai Tang Chao-Sheng Zhou Jing

引用信息 Citation: *Acta Physica Sinica*, 66, 188902 (2017) DOI: 10.7498/aps.66.188902

在线阅读 View online: <http://dx.doi.org/10.7498/aps.66.188902>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2017/V66/I18>

您可能感兴趣的其他文章

Articles you may be interested in

度相关性对无向网络可控性的影响

[Effect of degree correlations on controllability of undirected networks](#)

物理学报.2017, 66(2): 028901 <http://dx.doi.org/10.7498/aps.66.028901>

面向成本-收益好的无标度耦合网络构建方法

[A toward cost-effective scale-free coupling network construction method](#)

物理学报.2016, 65(9): 098901 <http://dx.doi.org/10.7498/aps.65.098901>

有倾向性重连产生的反匹配网络

[Disassortative networks generated by directed rewiring](#)

物理学报.2015, 64(19): 198901 <http://dx.doi.org/10.7498/aps.64.198901>

基于度的正/负相关相依网络模型及其鲁棒性研究

[The model of interdependent network based on positive/negative correlation of the degree and its robustness study](#)

物理学报.2015, 64(4): 048902 <http://dx.doi.org/10.7498/aps.64.048902>

一种信息传播促进网络增长的网络演化模型

[A complex network evolution model for network growth promoted by information transmission](#)

物理学报.2014, 63(15): 158902 <http://dx.doi.org/10.7498/aps.63.158902>

加权网络中基于冗余边过滤的 k -核分解排序算法*

罗仕龙¹⁾ 龚凯^{1)†} 唐朝生²⁾ 周靖¹⁾

1)(西南财经大学经济信息工程学院, 成都 611130)

2)(河南理工大学计算机科学与技术学院, 焦作 454000)

(2016年12月16日收到; 2017年6月21日收到修改稿)

k -核分解排序法对于度量复杂网络上重要节点的传播影响力具有重要的理论意义和应用价值, 但其排序粗粒化的缺陷也不容忽视. 最新研究发现, 一些真实网络中存在局域连接稠密的特殊构型是导致上述问题的根本原因之一. 当前的解决方法是利用边两端节点的外部连边数度量边的扩散性, 采取过滤网络边来减少这种稠密结构给 k -核分解过程造成的干扰, 但这种方法并没有考虑现实网络上存在权重的普遍性. 本文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 -核分解排序算法: filter-core. 通过世界贸易网、线虫脑细胞网和科学家合著网等真实网络的SIR (susceptible-infected-recovered) 传播模型的仿真结果表明, 该算法相比其他加权 k -核分解法, 能够更准确地度量加权网络上具有重要传播影响力的核心节点及核心层.

关键词: 加权网络, k -核分解, 冗余边, 传播影响力

PACS: 89.75.-k, 89.75.Fb, 05.10.-a

DOI: 10.7498/aps.66.188902

1 引言

现实社会中的许多系统都可以很自然地抽象成复杂网络^[1], 例如航空网络^[2]、在线虚拟网络^[3]、社会网络^[4]等. 在这些复杂网络中, 鉴别重要节点的研究工作对于预防基础设施遭受攻击^[5]、阻止传染病在人群中爆发^[6-8]、抑制谣言信息在社会上扩散^[9]等方面有着非常重要的理论意义和指导作用. 一般而言, 重要节点是指具有重要传播影响力的少数节点^[10], 通过传播途径快速地影响网络上大多数节点. 因此, 如何有效地度量网络节点的传播影响力既是发现重要节点的主要方法, 也是复杂网络研究领域的热点问题^[11,12].

近年来, 复杂网络节点的核数研究引起了学者们的高度关注^[13-15]. 核数(core)是指节点在网络上的位置核心性^[16], 基本思想是通过递归过程移

除网络中当前度数小于或等于 k 值的剩余节点, 剩下最大核数的节点组成网络核心层. Kitsak等^[17]基于 k -核分解思想(k -shell decomposition), 提出用核数度量节点的传播影响力, 取得了相比度数^[18]和介数^[19]更准确的排序结果. k -核分解法优点在于方法简单且计算复杂度低, 但存在排序结果粗粒化等缺陷. 国内外学者对此提出了许多改良措施^[20-25], 这些研究多数都是基于无权网络结构. 有些学者^[26-28]利用权重信息定义加权度数以替换核分解过程中的当前剩余度, 提出了加权网络上的 k -核排序算法.

最新实证研究发现^[29], 一些网络上存在局域连接稠密的特殊构型是导致 k -核分解无效的根本原因之一. 对此, 文献^[30]提出过滤某些边来减少稠密结构对 k -核分解过程造成干扰的方法, 指出边是否为应该过滤的冗余边(redundant links)取决于边在传播过程中具备的扩散性(diffusion im-

* 国家自然科学基金(批准号: 61602331)、中央高校基本科研业务费专项资金(批准号: JBK170133, JBK160130, JBK150503)、四川省教育厅科研基金(批准号: J17ZB0434)和互联网金融创新及监管协同创新中心资助的课题.

† 通信作者. E-mail: gongkai1210@swufe.edu.cn

portance), 并利用边两端节点的外部连边数度量边的扩散性, 但这种方法并不适用于加权网络. 我们考虑到现实网络上边含有权重的普遍性, 利用节点的权重特征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 -核分解排序算法: filter-core. 仿真结果显示, 利用节点的权重特征过滤冗余边后求得的节点核数, 能够更准确地度量加权网络上重要节点的传播影响力. 本文结构如下: 首先提出节点权重和权重分布指数的计算方法, 求出节点传播系数, 根据两端节点的传播系数计算出边的扩散系数, 过滤该系数低于阈值的所有网络边; 接着, 在剩余网络上实施 k -核分解求出节点核数; 最后通过世界贸易网、线虫脑细胞网和科学家合著网等真实加权网络上的 SIR (susceptible-infected-recovered) 传播仿真, 分别与其他加权 k -核分解法 [26–28] 的排序准确性进行比较分析. 实验结果表明, 本文算法能够更准确地识别加权网络上具有重要传播影响力的核心节点及核心层.

2 算法理论

最新研究证实 [29], k -核分解法在许多真实网络上失效的根本原因之一是这些网络中存在着局域连接稠密的微观构型, 这些微观构型可以是稠密三角形, 或是其他能够形成稠密回路的特殊构型. 这些微观构型会导致 k -核分解过程把网络上局域区域连接稠密的一些节点误判为核心节点, 而这些节点影响力受限于连接稠密的局部区域, 并不具备重要影响力; 同时, 这些微观构型的稠密性也容易导致网络节点的邻居重叠程度过高, 造成多源传播覆盖区域出现太多重叠, 严重地制约多源传播影响力. 对此, 文献 [30] 提出一种过滤网络冗余边来减少稠密结构从而提高 k -核分解准确性的方法, 指出边是否应该过滤取决于该边在传播过程中具备的扩散性, 例如无权网络上边两端节点的外部连边数越多, 意味着传播途径越广, 这条边的扩散性越强. 受到上述思想的启发, 我们考虑到现实网络上边含有权重的普遍性 [31], 提出利用节点的权重特征重新定义边的扩散性. 一般地, 加权网络上节点权重等于边的权重和, 其大小意味着该节点对周围邻居影响程度的强弱, 同时, 边的权重分布特征又意味着该节点在传播过程中传播途径的多样性或局限

性 [32]. 因此, 我们认为具备高权重且权重呈均匀分布特征的节点间的边, 能够在传播过程中起到很重要的扩散作用. 简单地说, 加权网络上边的扩散性依赖于两端节点的权重, 也依赖于两端节点的权重分布特征. 本文用符号 I_{ij} 表示边的扩散系数, I_i 表示节点的传播系数, D_i 表示节点权重, H_i 表示节点权重分布指数. 这里, I_{ij} 通过计算边两端节点的传播系数 I_i 求出:

$$I_{ij} = \frac{I_i + I_j}{2}, \quad (1)$$

$$I_i = D_i \times \frac{1}{H_i}. \quad (2)$$

利用边的权重定义节点权重时, 最常见的做法是按照网络边的权重均值进行归一化处理 [26], 但这样处理后的权重是一个绝对值. 我们提出节点权重的计算公式如下:

$$D_i = \sum_j D_{ij} = \sum_j \frac{w_{ij}}{\sum_i w_{ij}}, \quad (3)$$

这里 w_{ij} 表示边 e_{ij} 的实际权重. 为了量化节点的权重分布特征, 参考文献 [33] 中的度分布异质指数, 本文提出权重分布指数 H_i 的计算式:

$$H_i = \frac{\sum_{j=1}^{k_i} \sum_{k=1}^{k_i} |D_{ij} - D_{ik}|}{2k_i D_i}. \quad (4)$$

在附录 A 中, 我们利用 Kendall [34] 相关系数 τ 进一步分析本文所提出的权重特征指标与节点传播影响力之间的相关性, 结果表明 $\tau(I_i) > \tau(D_i) > \tau(H_i)$.

利用边过滤后实施 k -核数分解的思想, filter-core 算法分成过滤和分解两个阶段. 首先, 在过滤阶段, 计算并求出原始网络上所有边的扩散系数 I_{ij} . 如果该系数小于阈值 θ , 则视为冗余边进行过滤, 直至剩余网络中所有边的系数 I_{ij} 都大于阈值. 这里要根据具体情况进行算法阈值 θ 的选取: 选值过大会严重破坏网络结构, 得到的核数不能代表节点在原始网络上的核心性; 选值太小又无法实施有效地过滤. 对此, 我们遵循文献 [26] 中提出的过滤后网络核心层规模趋稳作为选取阈值 θ 的依据. 接着, 在核数分解阶段, 移除剩余网络中当前度数等于 1 的所有节点, 被移除后的节点处在网络的边缘核层, 核数 $ks = 1$; 继续移除剩余度数等于 2 的网络节点, 核数 $ks = 2$; 直至所有节点都被赋予相应

的核数后,核数分解结束.此时,最大核数的节点称为网络核心节点,最大核数的核层称为网络核心层. filter-core 算法在过滤阶段的计算复杂度为 $O\left(2E + \sum_{i=1}^N k_i^2 + N + E + E\right)$, 分解阶段 [35] 的计算复杂度为 $O(N + E)$, 其中 N 表示网络节点规模数, E 表示网络边数. 经过数学推导可知, 算法的计算复杂度为 $O(N + E^{3/2})$.

为了验证 filter-core 算法与其他加权 k -核分解法的排序优劣, 本文选用三种适用于加权网络结构的 k -核分解法: weighted degree 方法 [26] 通过重新定义度数和权重的乘积关系作为 k -核分解过程中使用的剩余度数 (本文简称 m-core), s-core 算法 [27] 是利用节点的边权重和进行核数分解, weighted k -shell 方法 [28] 将节点度数与权重的线性相加定义成核数分解时的剩余度数 (本文简称 p-core).

3 实验数据和算法应用

3.1 实验数据

本文实验数据选用三个真实加权网络: 世界贸易网 (ITN) [36]、线虫脑细胞网 (CEL) [37] 和科学家合著网 (SCIE) [38], 来检验 filter-core 算法度量节点传播影响力的准确性. 其中, ITN 权重表示国家之间的贸易额, 权重分布接近广延指数特征; CEL 权

重表示秀丽隐杆线虫的细胞数量, 权重幂率系数等于 1.76 ± 0.08 ; SCIE 权重表示科学家之间合作发表的论文数量, 权重分布呈指数特征. 表 1 统计了网络结构特征量, $\langle k \rangle$ 表示平均度, l 表示节点间最短路径, C 表示聚类系数 [39]. 可以看出, ITN 表现出高聚类系数与短路径的小世界特性, 又具有度数分布服从无标度的性质 [40]. 相反, CEL 聚类系数偏低且路径短, 更接近随机网络结构. SCIE 具有高聚类系数和路径长的双重特征, 呈现规则网络的特性. 上述网络表征不同的结构特征, 能够更好地评估算法在真实加权网络上的有效性.

表 1 网络拓扑的结构特征量

Table 1. Structural properties of real weighted networks.

网络	N	E	$\langle k \rangle$	l	C
ITN	181	13163	145.4	1.17	0.89
CEL	297	2148	14.5	2.46	0.29
SCIE	379	914	4.82	6.04	0.74

3.2 算法实际应用

根据算法理论可知, 阈值 θ 选取是根据过滤结束后网络核心层规模是否趋稳为依据, 图 1 描述的是 filter-core 算法应用在具体网络上的阈值选取过程. 图中横坐标表示阈值 θ , 纵坐标表示核心层的

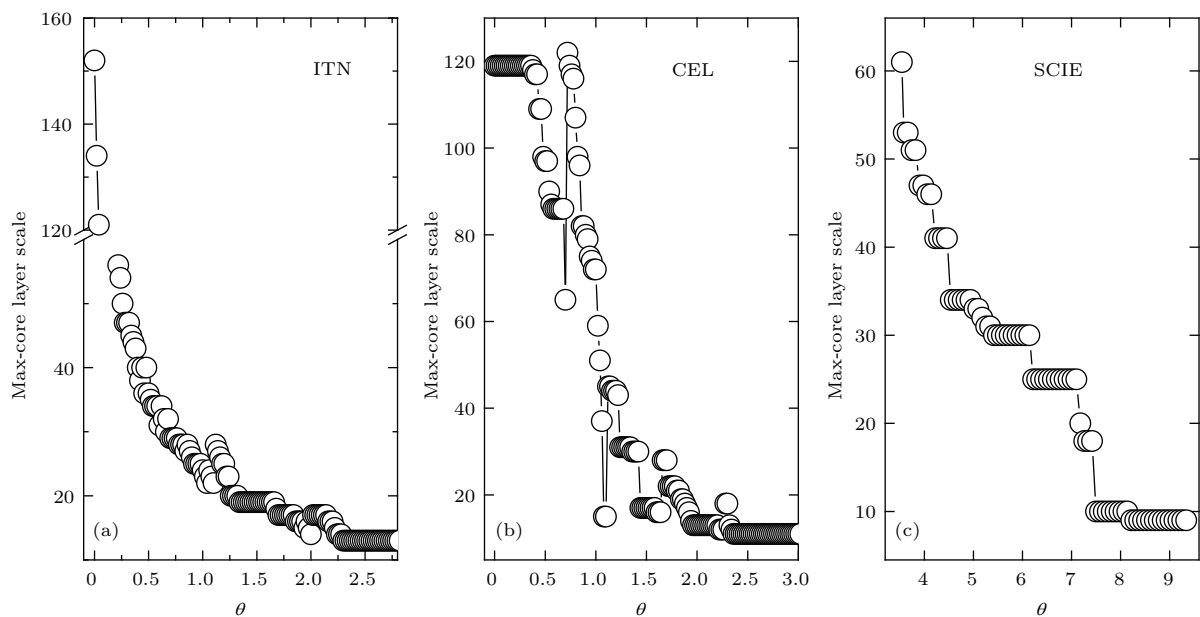


图 1 核心层的节点规模随阈值的变化 (a) ITN; (b) CEL; (c) SCIE

Fig. 1. Scale of the max-core for different redundant link thresholds θ in weighted networks: (a) ITN; (b) CEL; (c) SCIE.

节点规模数. 由图1可见, 随着阈值 θ 的不断增长, 初期臃肿的核心层规模逐渐缩减, 达到一定值时, 核心层规模基本趋于稳定. 例如ITN网络阈值 $\theta = 0$ 时, 核心层规模数等于152; 当阈值 $\theta > 2.2$ 时, 规模数在很长一段都稳定在13附近. 依据以上分析, 得出 $\theta(\text{ITN}) = 2.3$, $\theta(\text{CEL}) = 2.5$, $\theta(\text{SCIE}) = 8$.

为了进一步分析 filter-core 算法求出的核数能够代表节点在原始网络上的核心性, 图2描述了过滤阈值处的剩余网络的完整情况, 图中纵坐标表示极大连通系数 gcc. 极大连通系数是指过滤后最大连通网络的节点规模占原始网络节点规模的比例, 系数越大, 说明剩余网络的结构完整性越

高. 观察不难发现(圆圈符号), 三个网络在过滤阈值处都保持着较高的网络完整性($\text{gcc}(\text{ITN}) = 1$, $\text{gcc}(\text{CEL}) = 0.84$, $\text{gcc}(\text{SCIE}) = 0.54$). 相反, 如果连边扩散系数 I_{ij} 是根据两端节点权重进行计算(倒三角符号), 将会明显削弱剩余网络的结构完整性($\text{gcc}(\text{ITN}) = 1$, $\text{gcc}(\text{CEL}) = 0.74$, $\text{gcc}(\text{SCIE}) = 0.05$). 同样, 如果按照两端节点权重分布指数进行计算时(上三角符号), 剩余网络结构也会受到严重破坏. 经比较分析可知, filter-core 算法选取阈值过滤后的剩余网络结构保持着较高的完整性, 分解所得的核数能够从很大程度上代表节点在原始网络上的核心位置.

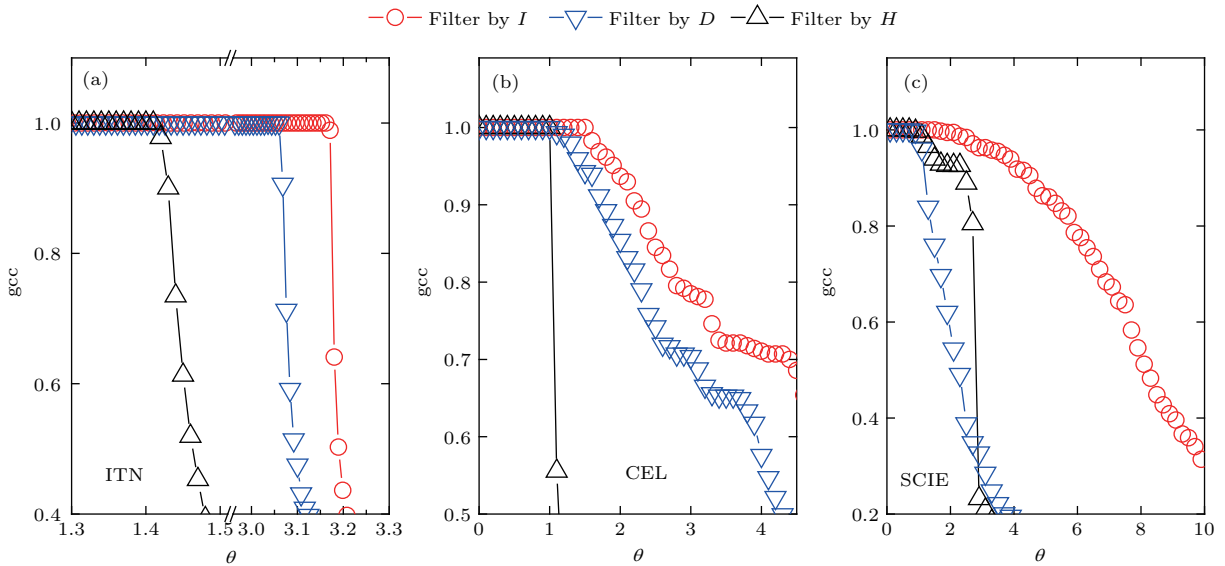


图2 (网刊彩色) 过滤阈值与剩余网络完整性 (a) ITN; (b) CEL; (c) SCIE

Fig. 2. (color online) Giant connected component gcc as the function of link filter threshold θ in weighted networks: (a) ITN; (b) CEL; (c) SCIE.

4 传播模型与指标

本文采用 SIR 模型 [26] 仿真加权网络上的传播过程. SIR 模型是把节点按照健康状态划分成三类: 易感态 S 是指当前未感染但存在感染可能性的节点, 感染态 I 是指当前患病且具有感染能力的节点, 康复态 R 是指已治愈且不再对传播行为产生任何影响的节点. 感染节点以感染概率 λ 将传染病传给邻居, 感染节点以康复率 β (取值等于 1) 进行治愈并免疫. 在加权网络的传播过程中, 节点间的感染概率并非恒等 [41]. 权重在传播路径上发挥着极其重要的作用, 感染概率与节点同其自身邻居间的权

重关系密切相关. 根据文献 [26], 感染概率 λ 的计算公式如下:

$$\lambda_{ij} = m \frac{w_{ij}}{\sum_i w_{ij}}, \quad (5)$$

参数 m 表征疾病感染性的强弱或者谣言迷惑性的强弱等. 传播模型用符号 M 表示节点的传播影响力:

$$M_i = \frac{1}{n^{\text{tol}}} \sum_{n=1} M_i(n), \quad (6)$$

其中 $M_i(n)$ 表示节点 v_i 在第 n 次传播结束后的感染节点数, M_i 表示节点 v_i 在 n^{tol} 次传播过程的感染节点平均数. 本文对每个网络进行 $n^{\text{tol}} = 10^5$ 次

的独立重复实验.

图3描述了节点核数 k_s 、节点传播影响力均值 $\langle M \rangle$ 和模型参数 m 三者间的变化关系, 图中点线表示核数 k_s . 观察可见, 高核数节点的传播影响力总是大于核数居中和核数偏低的节点. 也就是说, 当参数 m 固定时, 核数 k_s 越大, 节点传播影响力越强. 进一步分析发现, 当固定核数 k_s 时, 参数 m 与节点传播影响力 $\langle M \rangle$ 之间呈正相关. 参数 m 取值范围有严格限制, 取值太高会导致感染概率 λ 增大, 造成任意节点都会以很高概率感染网络其他节点, 无法有效地验证算法的优劣. 对此, 参考原模型的取值方法, 得出不同网络的参数值: $m(\text{ITN}) = 6$, $m(\text{CEL}) = 2$, $m(\text{SCIE}) = 3$.

基于 SIR 模型仿真, 本文利用传播误差率 ε 、传播发散性 σ 、传播时间 T 和感染规模 R 等指标比较分析 filter-core 与其他加权 k -核分解法的排序准确性. 传播误差率是指核数排序的结果与节点实际传

播影响力排序结果的符合度^[17]:

$$\varepsilon(p) = 1 - \frac{M_{\text{filter-core}}(p)}{M_{\text{eff}}(p)}, \quad (7)$$

其中 p 代表节点比例, $M_{\text{filter-core}}(p)$ 表示按照 k -核算法 (filter-core) 进行核数降序的节点传播影响力 M 的均值, $M_{\text{eff}}(p)$ 表示根据节点实际传播影响力降序的 M 均值. ε 越小, 说明排序结果的准确性越高. 传播发散性 σ 是指核心节点的传播影响力 M 与核心层里所有节点的平均传播影响力的离散程度, 即标准差:

$$\sigma = \sqrt{\frac{1}{|\psi|} \sum_i^{\psi} (M_i - \langle M \rangle)^2}, \quad (8)$$

其中 ψ 表示核心节点的集合, $|\psi|$ 表示核心节点的数量. σ 越大, 说明节点间的传播影响力的离散程度越高. 传播时间 T 和感染规模 R 是指核心层里的核心节点作为多源传播节点, 在传播时间内感染的网络节点数. 在规定传播时间 T 内, 感染规模 R 越大, 说明整个核心层的传播影响力越强.

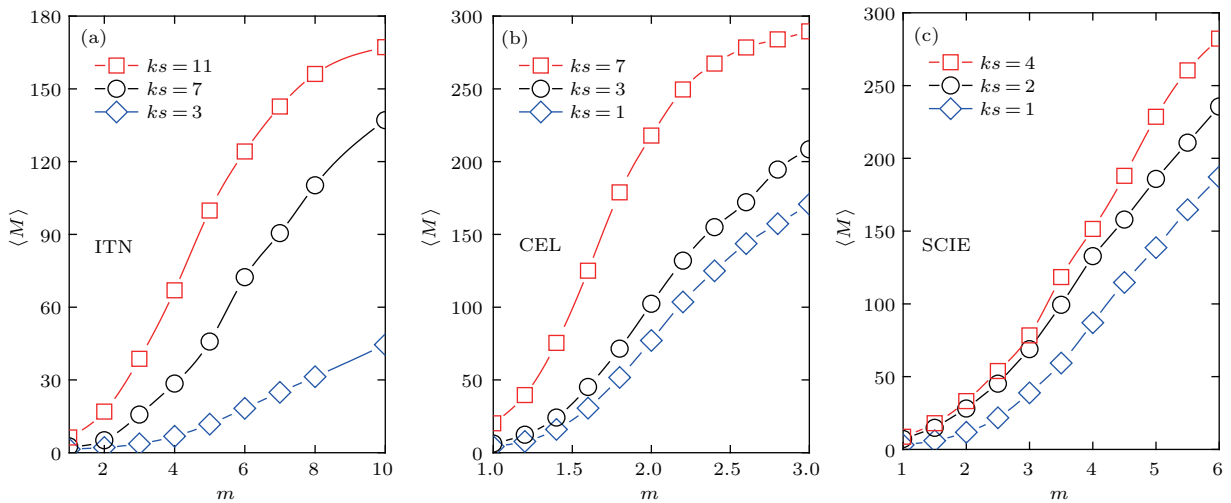


图3 (网刊彩色) 不同核数节点的传播影响力随 m 值的变化趋势 (a) ITN; (b) CEL; (c) SCIE

Fig. 3. (color online) Comparison of $\langle M \rangle$ vs. m for different cores in weighted networks: (a) ITN; (b) CEL; (c) SCIE.

5 结果分析

图4描述了核数分解排序后的节点比例 p 与传播误差率 ε 的变化, 横坐标表示按照核数降序的节点比例, 纵坐标表示传播误差率, 圆圈符号 (倒三角、方块、十字) 表示采用 filter-core 算法 (m-core, s-core, p-core) 分解得到的核数排序结果, 数值 ε 越

低, 说明排序准确性越高. 观察图4(a) 不难发现, filter-core 核数排序后的误差率最低; 其次是 s-core 和 p-core, 这两者排序的结果几乎重叠但误差率都不低; 误差率最高的是 m-core 分解后的核数排序. 进一步观察图4(b) 和图4(c) 可见, 我们提出的算法在 CEL 网络和 SCIE 网络上依然保持着较大优势, 能够很好地度量加权网络上具有重要传播影响力的少数节点.

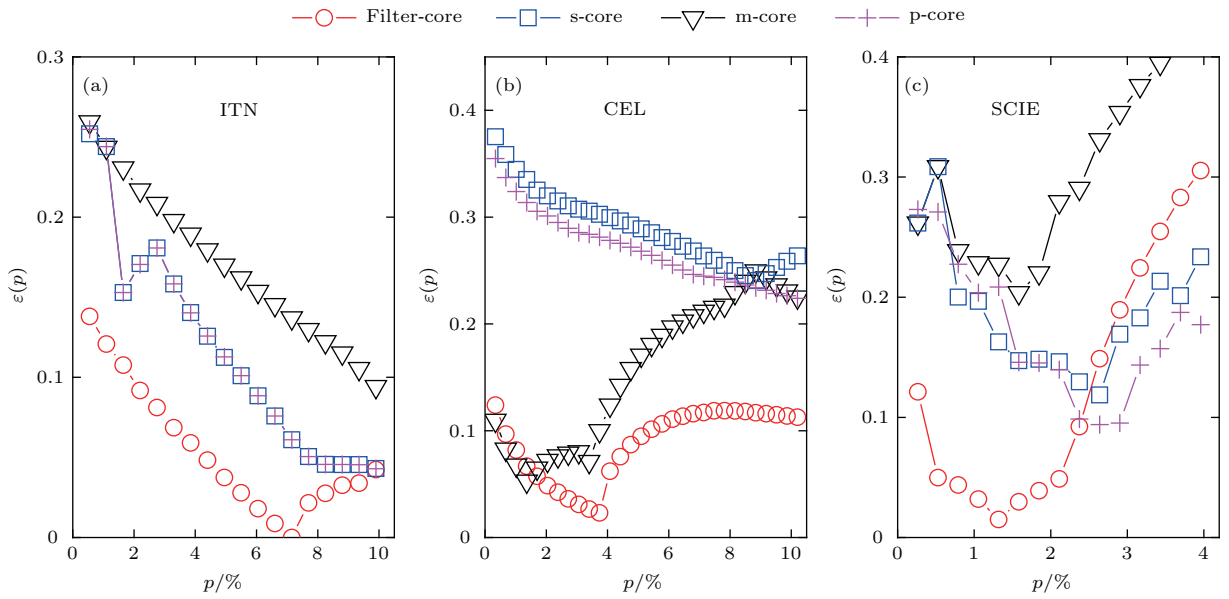


图4 (网刊彩色) 传播误差率随节点比例的变化 (a) ITN; (b) CEL; (c) SCIE

Fig. 4. (color online) The imprecision index ϵ as the function of node ratio p in weighted networks: (a) ITN; (b) CEL; (c) SCIE.

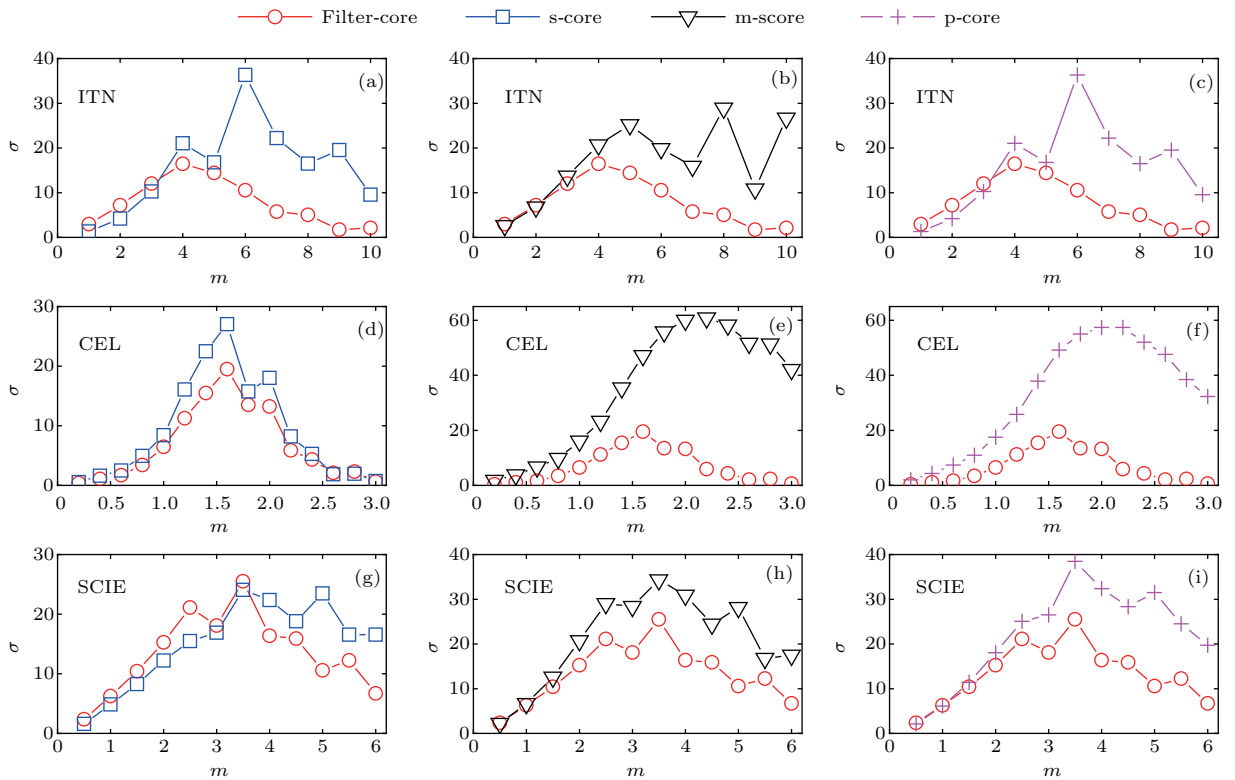


图5 (网刊彩色) 核心层节点的传播发散性随 m 值的变化 (a)—(c) ITN; (d)—(f) CEL; (g)—(i) SCIE

Fig. 5. (color online) Standard deviation σ of each node's spreading influence for max-core vs. m in weighted networks: (a)—(c) ITN; (d)—(f) CEL; (g)—(i) SCIE.

在 k -核分解法中, 最大核数的节点集构成了网络核心层, 而处在核心层里的所有节点应当具有近似相同的传播影响力^[30]. 对此, 进一步比较分析

filter-core 算法与其他加权 k -核分解法找到的核心层节点在传播发散性方面的优劣表现, 如图 5 所示. 整体而言, filter-core 算法在三个网络上的传播发

散值 σ 在大多数时候都是最低, 也就是说, 本文算法找到的网络核心层, 其内部核心节点之间的传播影响力表现得最接近. 观察图 5 (d)—(f) 可以发现, CEL 网络上的传播参数 m 取值偏小时, filter-core 算法相比其他算法在传播发散方面的优势并没有受到影响, 造成该现象的原因与网络权重和网络结构都有关: CEL 网络的权重分布具有幂率分布特征, 这种分布特征有利于本文算法在参数 m 较小时依然能够区分出具有重要扩散性的网络边, 有助于提高冗余边过滤的效果. 另外, k -核分解法不适用于 Barabási-Albert (BA) 无标度网络 [20], 根据表 1 结构特征量分析得知, CEL 网络结构更接近随机网络. 经上述分析可知, filter-core 算法更适用于权重分布具备幂率特征且结构不具备 BA 无标度的加权网络.

图 6 描述了 filter-core 算法与其他 k -核分解法找到的核心层节点作为多源传播节点时, 传播时间 T 与感染规模 R 之间的变化. 由图 6 可见, ITN 网络感染规模 $R \approx 80\%$ 时, filter-core 算法的传播时间 $T = 2$, 相反其他算法的传播时间 T 都大于 6; 在 CEL 网络中, 当传播时间 $T = 8$ 时, 相比 s-core 算法 (m-core, p-core) 感染规模 $R \approx 77\%$ (70%, 69%), filter-core 算法的感染规模更高 ($R \approx 80\%$); 在 SCIE 网络中, filter-core 算法在传播时间 $T = 10$ 的感染规模 $R \approx 61\%$, 明显强过 s-core 算法 (m-core, p-core) 在相同传播时间内的感染情况 (55%, 53%, 37%). 分析可知, filter-core 找到的核心层节点在多数传播时间内的感染规模都明显高于其他 k -核分解法. 上述结果进一步验证了我们算法能够更准确地度量加权网络上核心节点的传播影响力.

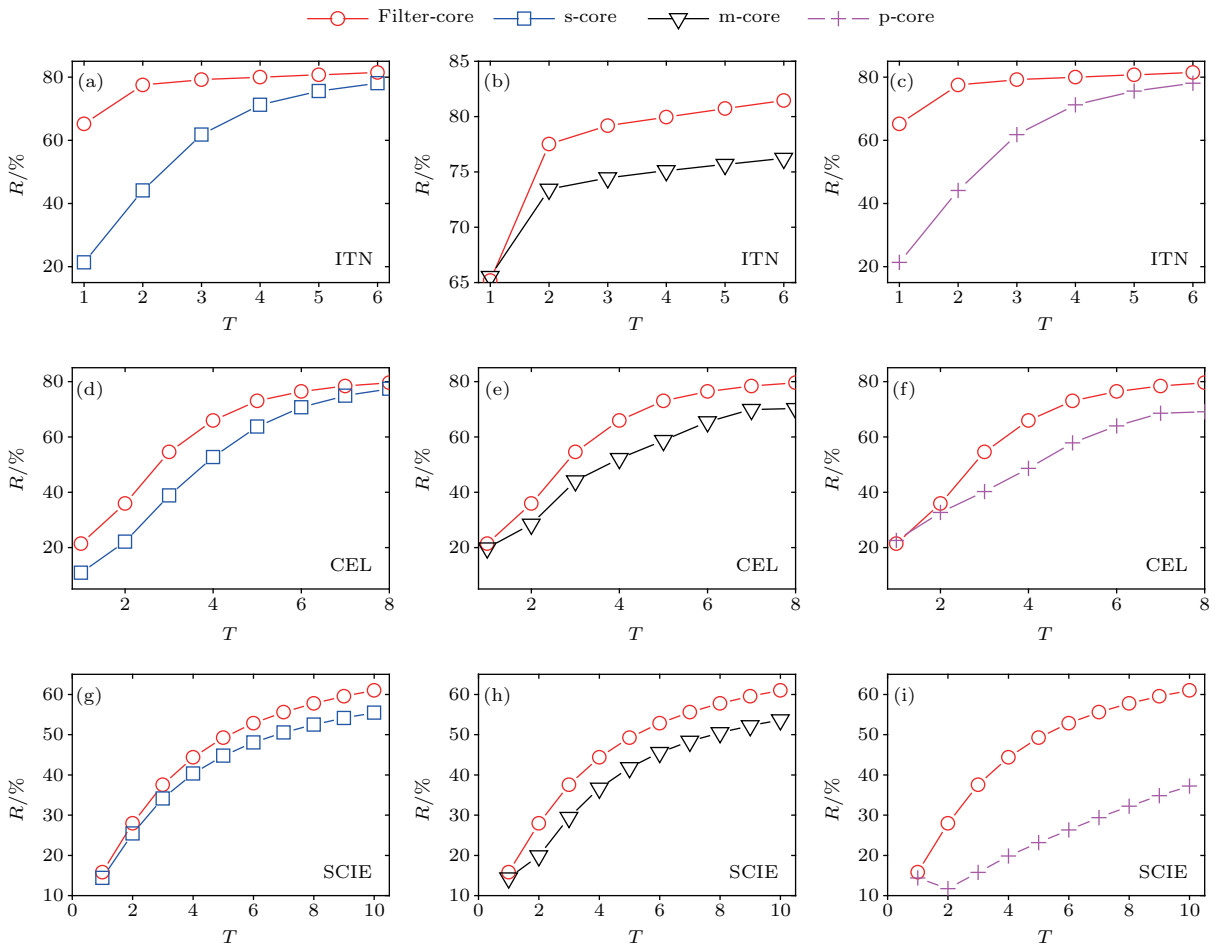


图 6 (网刊彩色) 核心层节点的传播时间与感染规模之间的变化 (a)—(c) ITN; (d)—(f) CEL; (g)—(i) SCIE
 Fig. 6. (color online) Spreading time T vs. number of recovered R profiles in weighted networks: (a)—(c) ITN; (d)—(f) CEL; (g)—(i) SCIE.

6 结论与展望

真实网络上存在局域连接稠密的特殊构型是导致 k -核分解结果粗粒化的本质原因之一. 本文考虑到现实网络上存在权重的普遍性, 利用节点权重和权重分布定义边的扩散性进而实施边过滤, 提出适用于加权网络结构的基于冗余边过滤的 k -核分解排序算法: filter-core. 通过对世界贸易网、线虫脑细胞网和科学家合著网等真实数据的 SIR 传播实验表明, 该算法的传播误差率相比其他加权 k -核分解法的值更低, 该结果表明我们算法得出的节点核数能够更准确地度量加权网络上重要节点的传播影响力; 利用传播发散性、传播时间和感染规模等指标进行比较分析, 结果进一步表明该算法能够更准确地识别加权网络上具有重要传播影响力的网络核心层.

本文利用权重特征和边过滤的思想对加权网络上节点的核数排序进行了研究, 该工作有助于人们进一步发现复杂网络上的重要节点以提高传播效率或阻止疾病传染及谣言传播等, 同时也为深入研究真实网络的核心层结构等后续工作提供了有

力的支持.

感谢审稿专家对文章提出的宝贵意见和建议, 这对本文研究工作的完善和提高起到了非常重要的作用.

附录 A 权重特征指标与节点传播影响力之间的相关性

这里采用 Kendall τ 相关系数, 基于真实加权网络世界贸易网 (ITN)、线虫脑细胞网 (CEL) 和科学家合著网 (SCIE), 分析节点的权重特征指标传播系数 I_i ((2) 式)、权重 D_i ((3) 式) 和权重分布指数 H_i ((4) 式) 与节点的传播影响力 M ((6) 式) 之间的相关性. 观察图 A1 可见, 综合考虑权重和权重分布的节点传播系数 I_i 与节点传播影响力 M 的相关系数 $\tau(\text{ITN}) = 0.861$, $\tau(\text{CEL}) = 0.744$, $\tau(\text{SCIE}) = 0.371$, 明显高于节点权重 D_i 与传播影响力 M 之间的相关系数 $\tau(\text{ITN}) = 0.738$, $\tau(\text{CEL}) = 0.648$, $\tau(\text{SCIE}) = 0.221$, 也明显高于权重分布指数的相关系数 $\tau(\text{ITN}) = 0.531$, $\tau(\text{CEL}) = 0.597$, $\tau(\text{SCIE}) = 0.103$. 结果表明, 加权网络上利用节点权重和权重分布能够更准确地反映出节点在传播中的重要性, 也进一步说明利用边两端节点的传播系数 I_i 可以更有效地度量边在传播过程中发挥的扩散作用.

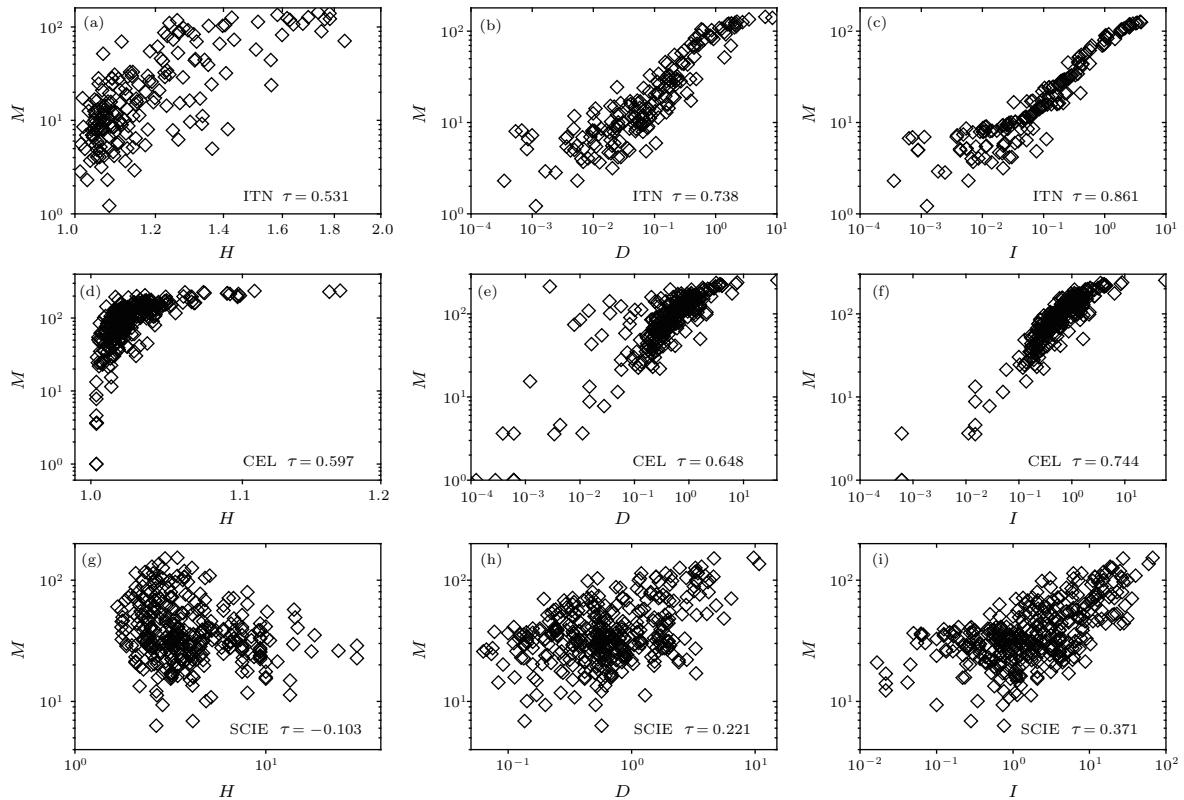


图 A1 节点的权重特征指标与传播影响力之间的相关性 (a)–(c) ITN; (d)–(f) CEL; (g)–(i) SCIE

Fig. A1. M vs. measuring index profiles in weighted networks: (a)–(c) ITN; (d)–(f) CEL; (g)–(i) SCIE.

参考文献

- [1] Wang X F, Li X, Chen G R 2012 *Network Science: an Introduction* (Beijing: Higher Education Press) pp3–18 (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3—18页]
- [2] Liu H K, Zhou T 2007 *Acta Phys. Sin.* **56** 106 (in Chinese) [刘宏鲲, 周涛 2007 物理学报 **56** 106]
- [3] Zhang Z K, Liu C, Zhan X X, Lu X, Zhang C X, Zhang Y C 2016 *Phys. Rep.* **651** 1
- [4] Liu C, Zhan X X, Zhang Z K, Sun G Q, Hui P M 2015 *New J. Phys.* **17** 113045
- [5] Cohen R, Erez K, Benavraham D, Havlin S 2001 *Phys. Rev. Lett.* **86** 3682
- [6] Liu J G, Lin J H, Guo Q, Zhou T 2016 *Sci. Rep.* **6** 21380
- [7] Keeling M J, Rohani P 2008 *Modeling Infectious Diseases in Humans and Animals* (Princeton: Princeton University Press) p366
- [8] Gong K, Tang M, Hui P M, Zhang H F, Younghae D, Lai Y C 2013 *Plos One* **8** e83489
- [9] Liu C, Zhang Z K 2014 *Commun. Nonlinear Sci.* **19** 896
- [10] Crucitti P, Latora V, Marchiori M, Rapisarda A 2004 *Physica A* **340** 388
- [11] 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]
- [12] 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]
- [13] Rombach M P, Porter M A, Fowler J H, Mucha P J 2014 *Siam J. Appl. Math.* **74** 167
- [14] Bassett D S, Wymbs N F, Rombach M P, Porter M A, Mucha P J, Grafton S T 2013 *Plos Comput. Biol.* **9** e1003171
- [15] Liu J G, Ren Z M, Guo Q, Chen D B 2014 *Plos One* **9** e104028
- [16] Seidman S B 1983 *Social Networks* **5** 269
- [17] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Nat. Phys.* **6** 888
- [18] Chen D, Lü L, Shang M S, Zhang Y C, Zhou T 2012 *Physica A* **391** 1777
- [19] Freeman L 1977 *Sociometry* **40** 35
- [20] Zeng A, Zhang C J 2013 *Phys. Lett. A* **377** 1031
- [21] Hou B, Yao Y, Liao D 2012 *Physica A* **391** 4012
- [22] Liu J G, Ren Z M, Guo Q 2013 *Physica A* **392** 4154
- [23] Hu Q C, Yin Y S, Ma P F, Gao Y, Zhang Y, Xing C X 2013 *Acta Phys. Sin.* **62** 140101 (in Chinese) [胡庆成, 尹龔燊, 马鹏斐, 高旻, 张勇, 邢春晓 2013 物理学报 **62** 140101]
- [24] Ren Z M, Liu J G, Shao F, Hu Z L, Guo Q 2013 *Acta Phys. Sin.* **62** 108902 (in Chinese) [任卓明, 刘建国, 邵凤, 胡兆龙, 郭强 2013 物理学报 **62** 108902]
- [25] Cao J X, Dong D, Xu S, Zheng X, Liu B, Luo J Z 2015 *Chin. J. Comput.* **38** 238 (in Chinese) [曹玖新, 董丹, 徐顺, 郑啸, 刘波, 罗军舟 2015 计算机学报 **38** 238]
- [26] Garas A, Schweitzer F, Havlin S 2012 *New J. Phys.* **14** 83030
- [27] Eidsaa M, Almaas E 2013 *Phys. Rev. E* **88** 062819
- [28] Wei B, Liu J, Wei D, Gao C, Deng Y 2015 *Physica A* **420** 277
- [29] Liu Y, Tang M, Zhou T, Do Y 2015 *Sci. Rep.* **5** 9602
- [30] Liu Y, Tang M, Zhou T, Do Y 2015 *Sci. Rep.* **5** 13172
- [31] Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A 2004 *Proc. Natl. Acad. Sci. USA* **101** 3747
- [32] Yan W, Zhou T, Wang J, Fu Z Q, Wang B H 2005 *Chin. Phys. Lett.* **22** 510
- [33] Hu H B, Wang X F 2008 *Physica A* **387** 3769
- [34] Kendall M G 1938 *Biometrika* **30** 81
- [35] Batagelj V, Zaversnik M 2003 arXiv:cs/0310049v1
- [36] Lee K M, Yang J S, Kim G, Lee J, Goh K I, Kim I M 2011 *Plos One* **6** e18443
- [37] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [38] Newman M E J 2006 *Phys. Rev. E* **74** 036104
- [39] Saramäki J, Kivela M, Onnela J P, Kaski K, Kertész J 2007 *Phys. Rev. E* **75** 027105
- [40] Li X, Jin Y Y, Chen G R 2003 *Physica A* **328** 287
- [41] Yan G, Fu Z Q, Chen G 2008 *Eur. Phys. J. B* **65** 591

A ranking approach based on k -shell decomposition method by filtering out redundant link in weighted networks*

Luo Shi-Long¹⁾ Gong Kai^{1)†} Tang Chao-Sheng²⁾ Zhou Jing¹⁾

1) (School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 611130, China)

2) (School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China)

(Received 16 December 2016; revised manuscript received 21 June 2017)

Abstract

The k -shell decomposition method of identifying the influential nodes which accelerate spread or hinder propagation, plays an important role in analyzing the spreading performance of complex network, but it is too coarse in terms of ranking granularity. Recent study shows that the accuracy of the k -shell decomposition method in determining node coreness is significantly affected by the mutually densely connected local structures. Existing approach tries to filter out the confusion of the classical k -shell decomposition method, caused by such densely local structures, through redefining a diffusion importance value which is the number of out-leaving links at/from the nodes connected by a link. This value is used to quantify the potential influence of a link in spreading process. However, the existing approach is not suitable for ubiquitously weighted networks. In this paper, we develop a new ranking approach (filter-core) to identify the node spreading influence in weighted network. Here, we concern that the redundant links, although they are less important in the spreading process, form mutually densely connected local structures, which lead to the classical k -shell decomposition method unable to accurately determine the coreness of node in network. By redefining a new diffusion importance value for each link based on the weights of its connected nodes and the weight distribution, we filter out the redundant links which have a relatively low diffusion importance in the spreading process. After filtering out all redundant links and applying the classical k -shell decomposition method to the residual network, we obtain a new coreness for each node, which is more accurate to indicate spreading influence of node in the original network. Our approach is applied to three real weighted networks, i.e., the international trading network, the neural network of *C. elegans*, and the coauthorship network of scientists. And the susceptible-infected-recovered epidemic spreading model is used to make a comparison of performance between our approach and other three k -shell methods (including the weighted degree decomposition method, the s -core decomposition method, and the weighted k -shell method) in terms of four quantitative indices, i.e., the imprecision function, the standard deviation of infected fraction of max coreness node, the spreading time, and the number of recovered nodes at the end of spreading process. The experimental results show that our proposed approach is more accurate to identify the influential spreaders than the weighted degree decomposition method, the s -core decomposition method, and the weighted k -shell method, and also helps to more accurately decompose the network core structure for an optimal analysis in weighted network.

Keywords: weighted networks, k -shell decomposition, redundant link, spreading influence

PACS: 89.75.-k, 89.75.Fb, 05.10.-a

DOI: 10.7498/aps.66.188902

* Project supported by the National Natural Science Foundation of China (Grant No. 61602331), the Fundamental Research Funds for the Central Universities of Ministry of Education of China (Grant Nos. JBK170133, JBK160130, JBK150503), the Scientific Research Foundation of the Education Department of Sichuan Province, China (Grant No. 17ZB0434), and the Collaborative Innovation Center for Electronic Finance and Financial Regulation.

† Corresponding author. E-mail: gongkai1210@swufe.edu.cn