

一种改进的基于信息传播率的复杂网络影响力评估算法

阮逸润 老松杨 王竣德 白亮 侯绿林

An improved evaluating method of node spreading influence in complex network based on information spreading probability

Ruan Yi-Run Lao Song-Yang Wang Jun-De Bai Liang Hou Lü-Lin

引用信息 Citation: *Acta Physica Sinica*, 66, 208901 (2017) DOI: 10.7498/aps.66.208901

在线阅读 View online: <http://dx.doi.org/10.7498/aps.66.208901>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2017/V66/I20>

---

您可能感兴趣的其他文章

Articles you may be interested in

多层单向耦合星形网络的特征值谱及同步能力分析

[Synchronizability and eigenvalues of multilayer star networks through unidirectionally coupling](#)

物理学报.2017, 66(18): 188901 <http://dx.doi.org/10.7498/aps.66.188901>

网络集聚性对节点中心性指标的准确性影响

[Effect of variable network clustering on the accuracy of node centrality](#)

物理学报.2016, 65(2): 028901 <http://dx.doi.org/10.7498/aps.65.028901>

两层星形网络的特征值谱及同步能力

[Synchronizability and eigenvalues of two-layer star networks](#)

物理学报.2016, 65(2): 028902 <http://dx.doi.org/10.7498/aps.65.028902>

基于联合矩阵分解的节点多属性网络社团检测

[Community detection based on joint matrix factorization in networks with node attributes](#)

物理学报.2015, 64(21): 218901 <http://dx.doi.org/10.7498/aps.64.218901>

基于时滞耦合映像格子的多耦合边耦合网络级联抗毁性研究

[Study on cascading invulnerability of multi-coupling-links coupled networks based on time-delay coupled map lattices model](#)

物理学报.2014, 63(7): 078901 <http://dx.doi.org/10.7498/aps.63.078901>

# 一种改进的基于信息传播率的复杂网络影响力评估算法\*

阮逸润<sup>1)†</sup> 老松杨<sup>1)</sup> 王竣德<sup>1)</sup> 白亮<sup>1)</sup> 侯绿林<sup>2)</sup>

1)(国防科技大学, 信息系统工程重点实验室, 长沙 410073)

2)(国防大学联合勤务学院, 北京 100858)

(2017年5月19日收到; 2017年7月4日收到修改稿)

评价网络中节点的信息传播影响力对于理解网络结构与网络功能具有重要意义. 目前, 许多基于最短路径的指标, 如接近中心性、介数中心性以及半局部 (SP) 指标等相继用于评价节点传播影响力. 最短路径表示节点间信息传播途径始终选择最优方式, 然而实际上网络间的信息传播过程更类似于随机游走, 信息的传播途径可以是节点间的任一可达路径, 在集聚系数高的网络中, 节点的局部高聚簇性有利于信息的有效扩散, 若只考虑信息按最优传播方式即最短路径传播, 则会低估节点信息传播的能力, 从而降低节点影响力的排序精度. 综合考虑节点与三步内邻居间的有效可达路径以及信息传播率, 提出了一种 SP 指标的改进算法, 即 ASP 算法. 在多个经典的实际网络和人工网络上利用 SIR 模型对传播过程进行仿真, 结果表明 ASP 指标与度指标、核数指标、接近中心性指标、介数中心性指标以及 SP 指标相比, 可以更精确地对节点传播影响力进行排序.

**关键词:** 复杂网络, 传播影响力, 信息传播率, 传播路径

**PACS:** 89.75.Fb, 89.75.Hc

**DOI:** 10.7498/aps.66.208901

## 1 引言

自然界中诸多的复杂系统都可以网络的形式存在, 我们的生活被各种各样的网络所包围<sup>[1-4]</sup>, 比如互联网、电力网络、社交网络和航空网络. 科学界真正开启网络研究的热潮是在网络的无标度特性<sup>[5]</sup>和小世界特性<sup>[6]</sup>被发现之后. 网络的无标度特性说明复杂系统内部存在严重的不均匀分布, 不同节点对网络结构和功能的影响大不相同; 而小世界特性表示网络中信息传递速度快, 大部分节点通过少数几步就可以到达其他节点. 当前, 越来越多的学者将研究的焦点放在网络节点个体的分析上, 其中, 如何准确识别网络节点的传播影响力是当前研究热点之一, 该工作对于控制谣言在社交网络上的传播<sup>[7,8]</sup>、传染病控制<sup>[9,10]</sup>、设计有效广告投放策

略进行病毒式营销等<sup>[11-14]</sup>具有非常重要的作用.

目前已提出许多经典的中心性指标用于对节点的传播影响力进行排序, 包括度中心性<sup>[15]</sup>、半局部度中心性<sup>[16]</sup>、特征向量中心性<sup>[17,18]</sup>、接近中心性<sup>[19]</sup>、介数中心性<sup>[20]</sup>、HITS<sup>[21]</sup>、PageRank<sup>[22,23]</sup>、LeaderRank<sup>[24]</sup>与H指数等<sup>[25]</sup>. 其中度排序方法最为简单直观, 但其精度有待进一步提高; 半局部中心性指标有限地扩大了源节点领域的覆盖范围, 在提高算法精度的同时兼顾了算法的时间复杂度. Kitsak等<sup>[26]</sup>提出了K核分解算法, 该算法通过逐步剥离网络外围度数小的节点, 可以较为准确地识别网络中最有影响力的内核节点, 但该方法对于网络整体节点的排序结果粒度较粗, 节点间的传播影响力区分度不够. 核数中心性指标<sup>[27]</sup>认为节点的影响力由其邻居在网络中的地位决定, 节点与网络

\* 国家自然科学基金 (批准号: 61302144, 61603408) 资助的课题.

† 通信作者. E-mail: ruanyirun@163.com

中K核值大的节点间存在的连接越多, 则其影响力越大. Liu等<sup>[28]</sup>考虑多阶邻居节点的中心性, 提出了一种邻域中心性指标用于评价节点影响力, 发现算法排序效果并不总是随着邻居阶数的增大而变好. 段杰明等<sup>[29]</sup>基于自规避随机游走思想, 提出一种综合考虑网络结构局域信息和标签扩散的影响力排序算法. Liu等<sup>[30]</sup>综合考虑传染率、康复率和有限的时间步三个因素用于评价网络节点影响力. 更多关于节点传播影响力排序方面的研究可以参见文献<sup>[31—33]</sup>.

最近, Bao等<sup>[34]</sup>指出节点的传播影响力由节点与邻域节点间的最短路径数、最短路径长度以及传播概率决定, 并基于此设计了一种半局部(SP)算法用于评价节点传播影响力. 最短路径表示节点间信息传播途径始终选择最优方式, 然而实际上网络中的消息、谣言或者资讯等在节点间进行传播时并不会遵循最短路径, 信息扩散的过程更类似于随机游走<sup>[35]</sup>. 在集聚系数高的网络中, 节点的局部高聚簇性有利于信息的有效扩散, 若只考虑信息按最优传播方式即最短路径传播, 则会低估节点信息传播的能力, 从而降低节点影响力的排序精度. 综合考虑节点与局域三步内邻居的有效可达路径及信息传播率, 本文提出了一种SP指标的改进算法, 即ASP算法. 在多个真实世界网络和人工网络中的实验表明, ASP算法与SP指标、核数中心性指标、半局部度中心性指标以及介数中心性指标相比, 更能准确评估节点的传播影响力.

## 2 相关研究

假设无向网络  $G = (V, E)$  包含  $|V| = N$  个节点和  $|E| = M$  条边, 其网络结构可用邻接矩阵  $A = (a_{ij})_{N \times N}$  表示, 当节点  $i$  与节点  $j$  之间存在连接时,  $a_{ij} = 1$ , 否则  $a_{ij} = 0$ .

### 2.1 度指标

度排序方法最为简单直观<sup>[15]</sup>,  $k_i$  表示节点  $i$  的邻居节点数:

$$k_i = \sum_{j \in A_i} a_{ij}, \quad (1)$$

式中  $A_i$  表示节点  $i$  一跳邻域内的邻居. 度指标反映了节点的直接影响力, 信息从度越大的节点发起, 网络中第一时间接收到该信息的节点数越多, 但度

指标忽略了邻居节点间的差异性, 导致其排序精度不高.

### 2.2 接近中心性

接近中心性(CC)反映了节点到达其他节点的难易程度<sup>[19]</sup>, 可表示为

$$C_{C\_i} = \frac{N}{\sum_{j=1}^N d_{ij}}, \quad (2)$$

式中  $d_{ij}$  为节点  $i$  到节点  $j$  的距离,  $N$  代表网络节点总数.

### 2.3 介数指标

介数指标(BC)以网络中经过某个节点的最短路径的数目来量化节点在网络中的地位<sup>[20]</sup>, 定义为

$$C_{B\_i} = \sum_{s \neq i \neq t \in V} \frac{n_{st}^i}{g_{st}}, \quad (3)$$

式中  $g_{st}$  表示从节点  $s$  到节点  $t$  的最短路径数,  $n_{st}^i$  表示其中经过节点  $i$  的最短路径数.

### 2.4 核数指标

Bae和Kim<sup>[27]</sup>同时考虑节点度与邻居节点的K核值, 提出了一种核数指标(Cnc)用于评价节点传播影响力. 该指标定义为

$$C_{nc\_i} = \sum_{j \in A_i} K(j), \quad (4)$$

式中  $K(j)$  表示节点  $j$  的K核值.

### 2.5 SP指标

Bao等<sup>[34]</sup>从信息传播的角度分析认为, 从网络中的某一节点  $i$  发起的信息要成功传递到另一个节点  $j$ , 其概率由节点  $i$  和  $j$  之间的最短路径数目、最短路径长度以及信息传播率决定, 他们将网络平均度  $\langle k \rangle$  的倒数近似为信息传播率, 设计了SP指标用于评价节点传播影响力, 表示为

$$f_{SP\_i} = \sum_{j \in \varphi_i} n_{ij} (1/\langle k \rangle)^{d_{ij}}, \quad (5)$$

式中  $\varphi_i$  表示与节点  $i$  距离小于等于网络半径的节点集合,  $n_{ij}$  表示节点  $i$  和节点  $j$  之间的最短路径数目,  $n_{ij} (1/\langle k \rangle)^{d_{ij}}$  近似表示节点  $i$  成功将信息传播至节点  $j$  的概率.

### 2.6 ASP 指标

实际上, 信息按照最短路径传播只是理想中的路由方式. 从某一个节点  $i$  发起的信息要最终传递至另一个节点  $j$ , 信息传递的路径理论上可以是  $i$  和  $j$  之间的任一可达路径, 因此在计算节点  $i$  将信息传播至节点  $j$  的可能性时, 若只考虑最短路径, 必然会降低算法的精度. 以图 1 为例, 由于只考虑最短路径, 节点  $i$  和  $k$  将信息传递至节点  $a$  的概率都为  $3(1/\langle k \rangle)^3$ . 依次计算信息从节点传递到其他邻域邻居的概率, 最终得到节点  $i$  和  $k$  的 SP 指标值相等. 然而直观上由于节点  $i$  的局部高聚簇性, 节点  $i$  将信息传递至  $a$  的过程中相比于节点  $k$  具有更多的路径选择, 因此可以推断节点  $i$  的传播影响力大于节点  $k$ . 疾病传播模型 (SIR 模型) 信息传播仿真实验也验证了我们的猜想.

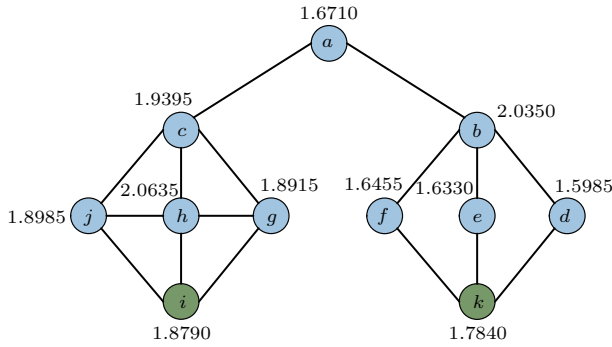


图 1 11 个节点组成的网络图, 节点感染概率取  $\beta = 0.2$  (传播阈值  $\beta_{th} = 0.3409$ ) 时在 SIR 模型上进行 2000 次独立仿真实验得到 11 个节点的信息传播影响力

Fig. 1. A simple network with 11 nodes. The spreading influence of each node is obtained by simulation on the SIR model over 2000 independent runs with  $\beta = 0.2$  ( $\beta_{th} = 0.3409$ ).

基于以上分析, 我们设计了 SP 指标的改进算法, 即 ASP 算法, 其表达式为

$$f_{ASP\_i} = \sum_{j \in \Gamma_i} f_{ASP\_ij} = \sum_{j \in \Gamma_i} \left[ \sum_l n_l (1/\langle k \rangle)^l \right], \quad 1 \leq l \leq 3, \quad (6)$$

式中  $f_{ASP\_ij}$  表示节点  $i$  将信息传播至节点  $j$  的成功率,  $\Gamma_i$  表示节点  $i$  三步内的邻居节点集合,  $l$  表示节点  $i$  到节点  $j$  的可达路径的长度,  $n_l$  表示节点  $i$  到节点  $j$  可达路径长度 (不包括回路) 为  $l$  的路径总数.

Fowler 和 Christakis<sup>[36]</sup> 认为节点的影响范围不仅限于直接相邻的节点, 还能间接影响与邻居节

点相邻的节点, 三阶以内都可能产生影响, 并提出了三阶影响力原则. 同时由于  $(1/\langle k \rangle)^l$  随着  $l$  的增大会快速衰减, 因此考虑所有的可达路径并不必要, ASP 算法只将节点与邻域节点间长度不大于 3 的可达路径纳入计算范围.

对于一步邻居节点,

$$f_{ASP\_ij} = \sum_l n_l (1/\langle k \rangle)^l = 1/\langle k \rangle + |n(i) \cap n(j)| \cdot (1/\langle k \rangle)^2 + [A^3(i, j) - |n(i)| - |n(j)| + 1] \cdot (1/\langle k \rangle)^3, \quad (7)$$

表 1 ASP 算法框架

Table 1. The framework of ASP algorithm.

Input: adjacent matrix $\mathbf{A} = (a_{ij})_{N \times N}$ of social network
$G = (V, E),  V  = N,  E  = M$
Output: ASP(V)
1 Input: $\mathbf{A} = (a_{ij})_{N \times N}$
2 Compute reachability matrix element $A^3(i, j)$
3 For $i = 1$ to $N$
4 For $j = 1$ to $k_1$ ( $k_1$ is the number of 1-step neighborhoods of node $i$ )
5 Compute the probability that the node $i$ transmits information to the node $j$
6 $f_{ASP\_ij} = (1/\langle k \rangle) +  n(i) \cap n(j)  \cdot (1/\langle k \rangle)^2 + [A^3(i, j) -  n(i)  -  n(j)  + 1] \cdot (1/\langle k \rangle)^3$
7 End for
8 For $l = 1$ to $k_2$ ( $k_2$ is the number of 2-step neighborhoods of node $i$ )
9 Compute the probability that the node $i$ transmits information to the node $l$
10 $f_{ASP\_il} =  n(i) \cap n(l)  \cdot (1/\langle k \rangle)^2 + A^3(i, l) \cdot (1/\langle k \rangle)^3$
11 End for
12 For $m = 1$ to $k_3$ ( $k_3$ is the number of 3-step neighborhoods of node $i$ )
13 Compute the probability that the node $i$ transmits information to the node $m$
14 $f_{ASP\_im} = A^3(i, m) \cdot (1/\langle k \rangle)^3$
15 End for
16 $f_{ASP\_i} = \text{sum}(f_{ASP\_ij}) + \text{sum}(f_{ASP\_il}) + \text{sum}(f_{ASP\_im})$
17 End for

式中  $\mathbf{A}$  代表网络邻接矩阵,  $A^3(i, j)$  表示节点  $i$  与节点  $j$  之间长度为 3 的所有可达路径,  $A^3(i, j) - |n(i) - n(j)| + 1$  为消除可达路径中的回路后剩下的路径数,  $|n(i) \cap n(j)|$  表示节点  $i$  与节点  $j$  的共同邻居数, 即长度为 2 的路径数.

对于二步邻居节点,

$$f_{ASP\_ij} = |n(i) \cap n(j)| \cdot (1/\langle k \rangle)^2 + A^3(i, j) \cdot (1/\langle k \rangle)^3. \quad (8)$$

对于节点  $i$  的二步邻居节点  $j$ , 节点  $i$  和节点  $j$  之间不存在长度为 3 且有回路的路径, 因此节点  $i$  到节点  $j$  的可达路径数为  $A^3(i, j)$ .

对于三步邻居节点,

$$f_{ASP\_ij} = A^3(i, j) \cdot (1/\langle k \rangle)^3. \quad (9)$$

以图 1 中节点  $i$  与节点  $c$  为例说明 ASP 计算过程. 节点  $i$  到节点  $c$  路径长度为 2 的路径有 3 条, 路径长度为 3 的路径有 4 条, 根据 ASP 指标, 可计算节点  $i$  成功传递信息至节点  $m$  的可能性为  $3(1/\langle k \rangle)^2 + 4(1/\langle k \rangle)^3$ , 同理, 可依次计算节点  $i$  传递信息至节点  $j, h, g$  与  $a$  的可能性, 由此得到节点  $i$  的影响力值.

为降低计算复杂度, SP 指标只考虑三步内的邻居节点. 与 SP 指标相比, ASP 指标多考虑了节点与一步邻居中长度为 2 和 3 的可达路径以及与二步邻居中长度为 3 的可达路径. 尽管如此, 通过表 1 中的算法框架依然可以看出, 改进指标几乎不增加计算的复杂度.

### 3 评价标准

为了衡量各种指标影响力排序结果的准确性, 采用 SIR 模型 [37,38] 对传播过程进行模拟. 在 SIR 模型中, 节点存在三种状态: 1) 易染状态 S, 个体易被感染, 但还未被感染; 2) 感染状态 I, 个体已经被感染; 3) 恢复状态 R, 个体被感染后以概率  $\lambda$  被治愈, 治愈后对该疾病免疫, 不再被感染. 不失一般性, 本文所有 SIR 模型实验都设  $\lambda = 1$ . 使用 SIR 模型验证所提算法评价节点影响力的准确性. 将网络中任意一个节点设为感染状态 I, 作为初始传播源, 其余节点设为易感状态 S, 在每个时间步中, 处于状态 I 的节点以传播率  $\beta$  尝试感染处于状态 S 的邻居节点, 并且自身以概率  $\lambda = 1$  恢复成状态 R, 重复此过程直至网络中没有处于状态 I 的节点. 定义节点的实际传播影响力  $\Phi(i) = \frac{1}{M} \sum_{m=1}^M \Phi'(i)$ , 其中  $M$  表

示实验重复次数,  $\Phi'(i)$  表示节点  $i$  作为传播源, 一次传播实验中网络中处于状态 R 的节点总数.

实验采用 Kendall tau 相关系数 [39,40] 衡量各指标排序结果的准确性, 其表达式为

$$\tau(R_1, R_2) = \frac{n_c - n_d}{\sqrt{(n_t - n_u)(n_t - n_v)}}, \quad (10)$$

式中  $R_1$  与  $R_2$  表示  $n$  个节点的两种不同排序序列,  $n_c$  和  $n_d$  分别表示两种排列的同序对和异序对的个数,  $n_t = n(n-1)/2$ ,  $n_u = \sum_{i=1}^s u_i(u_i-1)/2$ ,

$n_v = \sum_{i=1}^t v_i(v_i-1)/2$ .  $n_u$  与  $n_v$  分别针对序列  $R_1$  与  $R_2$  计算, 以  $n_u$  为例 ( $n_v$  的计算类推), 将  $R_1$  中的相同元素组成小集合,  $s$  表示序列  $R_1$  中拥有的小集合数,  $u_i$  表示第  $i$  个集合所包含的元素数. 由不同的评估指标得到节点的影响力值序列, 同时根据 SIR 模型传播过程得到各个节点影响节点数, 按照相同次序构成另一等级序列, 利用 Kendall tau 相关系数进行准确性计算,  $\tau$  值越高, 表明评估指标的结果与 SIR 模型仿真结果越接近, 评估结果越准确.

### 4 真实数据集实验结果

为了验证各指标评估节点传播影响力的效果, 实验选取 6 个真实数据集, 包括 Word, Netscience [41], Email [42], Yeast [43], Blog [44] 和 Router [45], 这些网络的拓扑结构统计特征如表 2 所列. 其中,  $N$  与  $M$  分别表示网络节点数与连边数,  $C$  表示网络集聚系数,  $D$  为网络直径,  $\beta$  表示传播概率,  $\beta_{th} = \langle k \rangle / \langle k^2 \rangle$  表示传播阈值, 其中  $\langle k \rangle$  表示节点平均度,  $\langle k^2 \rangle$  表示节点二阶平均度.

表 2 6 个真实网络的拓扑参数

Table 2. Topological parameters of six real networks.

网络名	$N$	$M$	$\beta$	$\beta_{th}$	$\langle k \rangle$	$C$	$D$
Word	112	425	0.08	0.073	7.59	0.173	5
Netscience	379	914	0.13	0.125	4.82	0.741	17
Email	1133	5451	0.06	0.054	9.62	0.570	8
Yeast	1458	1948	0.15	0.140	2.67	0.071	19
Blog	3982	6803	0.08	0.073	3.42	0.284	8
Router	5022	6258	0.08	0.073	2.49	0.033	15

在 6 个真实网络数据集和 3 个模拟数据集上, 比较提出的 ASP 指标与 SP 指标、度指标、核数指标、介数中心性指标以及接近中心性指标. 图 2 描述了不同网络中心性指标与实际影响力  $\Phi(i)$  之间

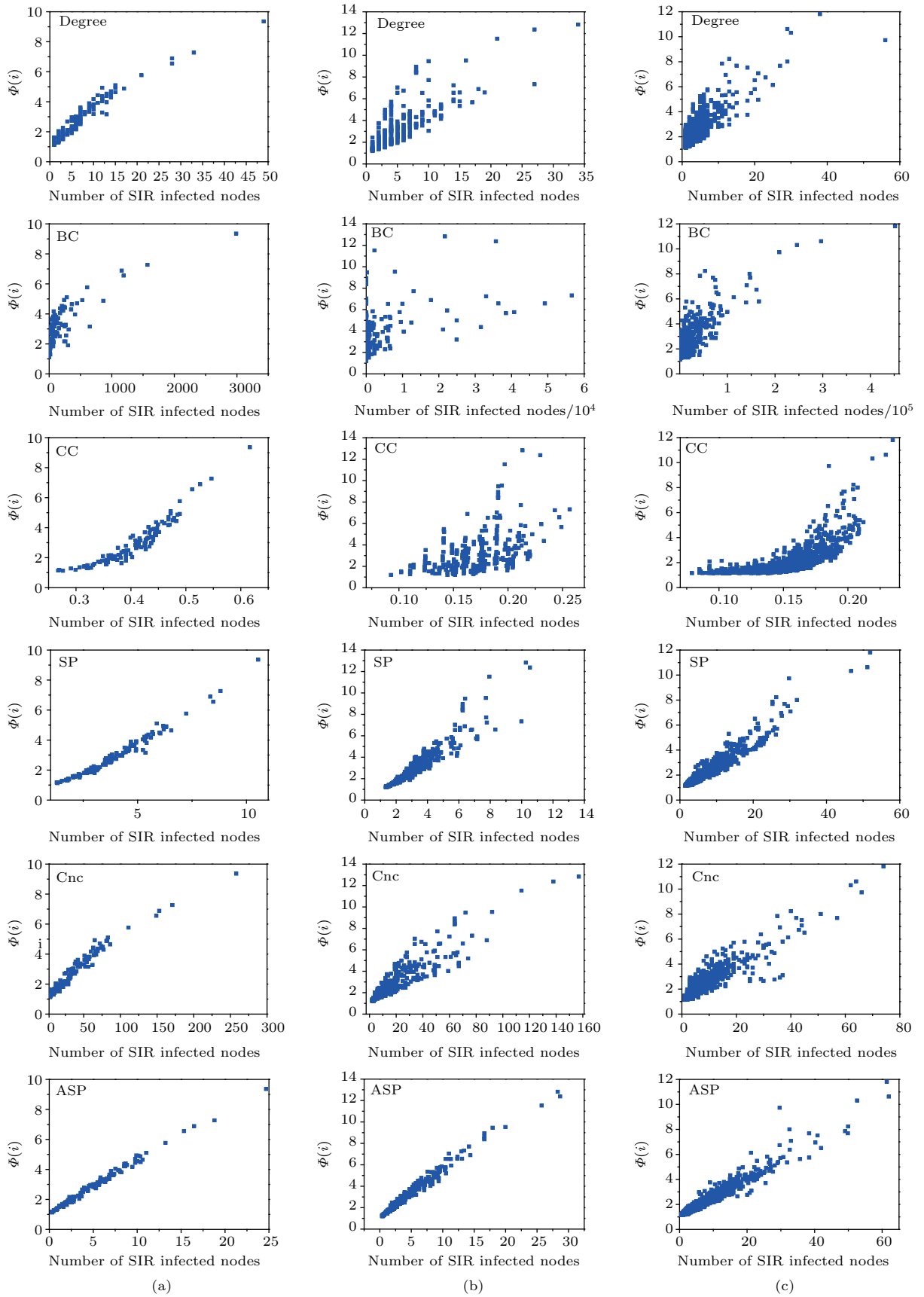


图2 六个真实网络中不同指标评估值与SIR模型感染节点数的相关性 (a) Word; (b) Netscience; (c) Yeast; (d) Email; (e) Blog; (f) Router

Fig. 2. Correlation between different ranking values and number of SIR model infected nodes in six real-world networks: (a) Word; (b) Netscience; (c) Yeast; (d) Email; (e) Blog; (f) Router.

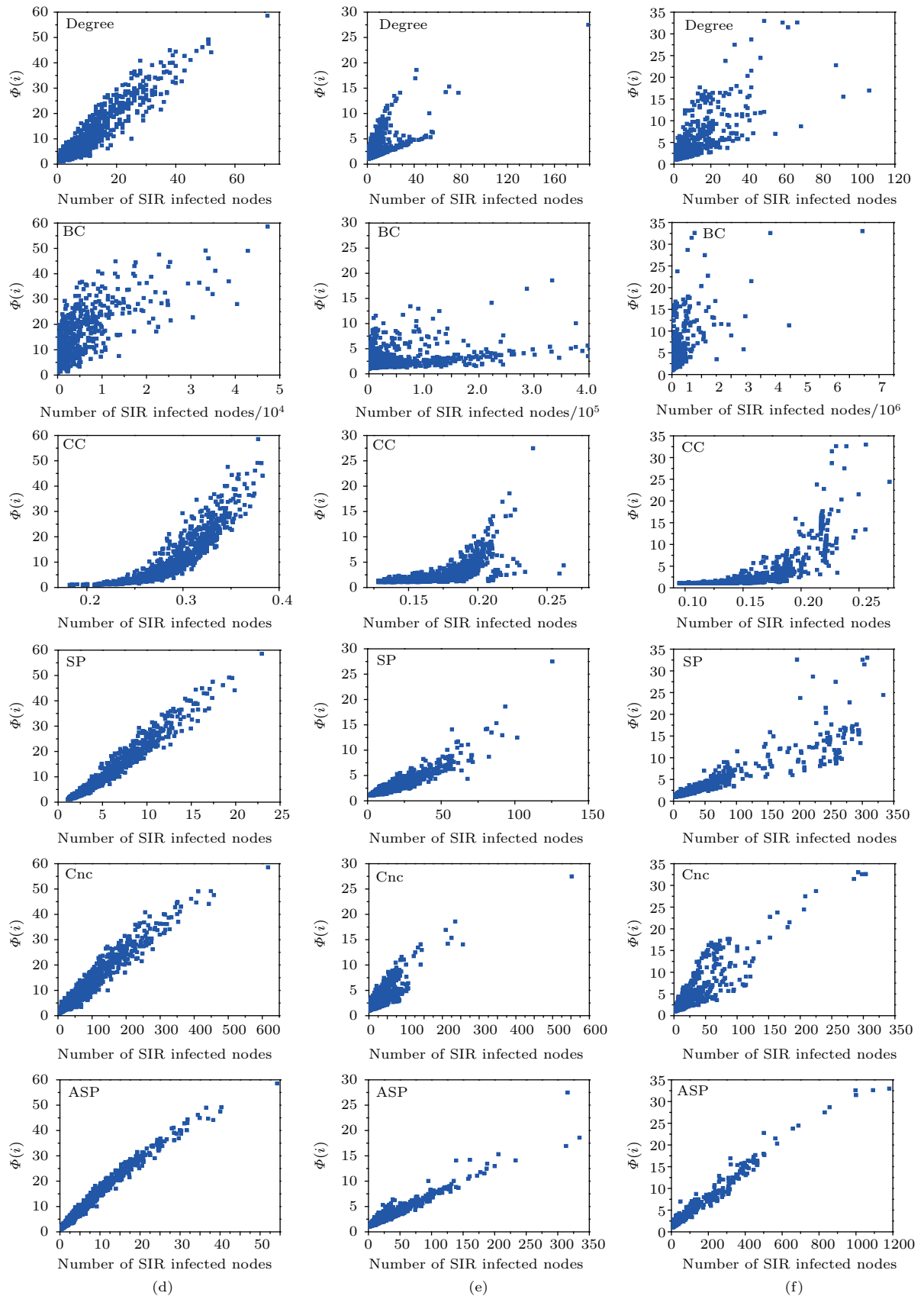


图2 六个真实网络中不同指标评估值与SIR模型感染节点数的相关性(续) (a) Word; (b) Netscience; (c) Yeast; (d) Email; (e) Blog; (f) Router

Fig. 2. Correlation between different ranking values and number of SIR model infected nodes in six real-world networks (continued): (a) Word; (b) Netscience; (c) Yeast; (d) Email; (e) Blog; (f) Router.

的相关性, 相关程度越高, 表明算法对节点传播影响力的测量越准确. 由于节点的影响力由最终被感染的网络节点数量决定, 因此为了正确评价节点的真实影响力, 感染概率  $\beta$  的值不宜选得过大或过小, 若  $\beta$  值过小, 信息传播容易局限于节点邻域. 相反, 若  $\beta$  值过大, 不论传染过程从哪个节点发起, 整个网络都很快被感染, 很难区分单个个体的影响力.

为保证传播能够进行, 实验设定感染概率  $\beta$  等于网络传播阈值  $\beta_{th}$ , SIR 实验独立运行 1000 次取平均结果.

从图 2 可以看出, 接近中心性和介数中心性指标与 SIR 影响节点数的相关性相对较弱, 接近中心性与 SIR 影响节点数总体呈正相关, 介数中心性的结果较为发散. 这是因为社会化网络的社区化使得

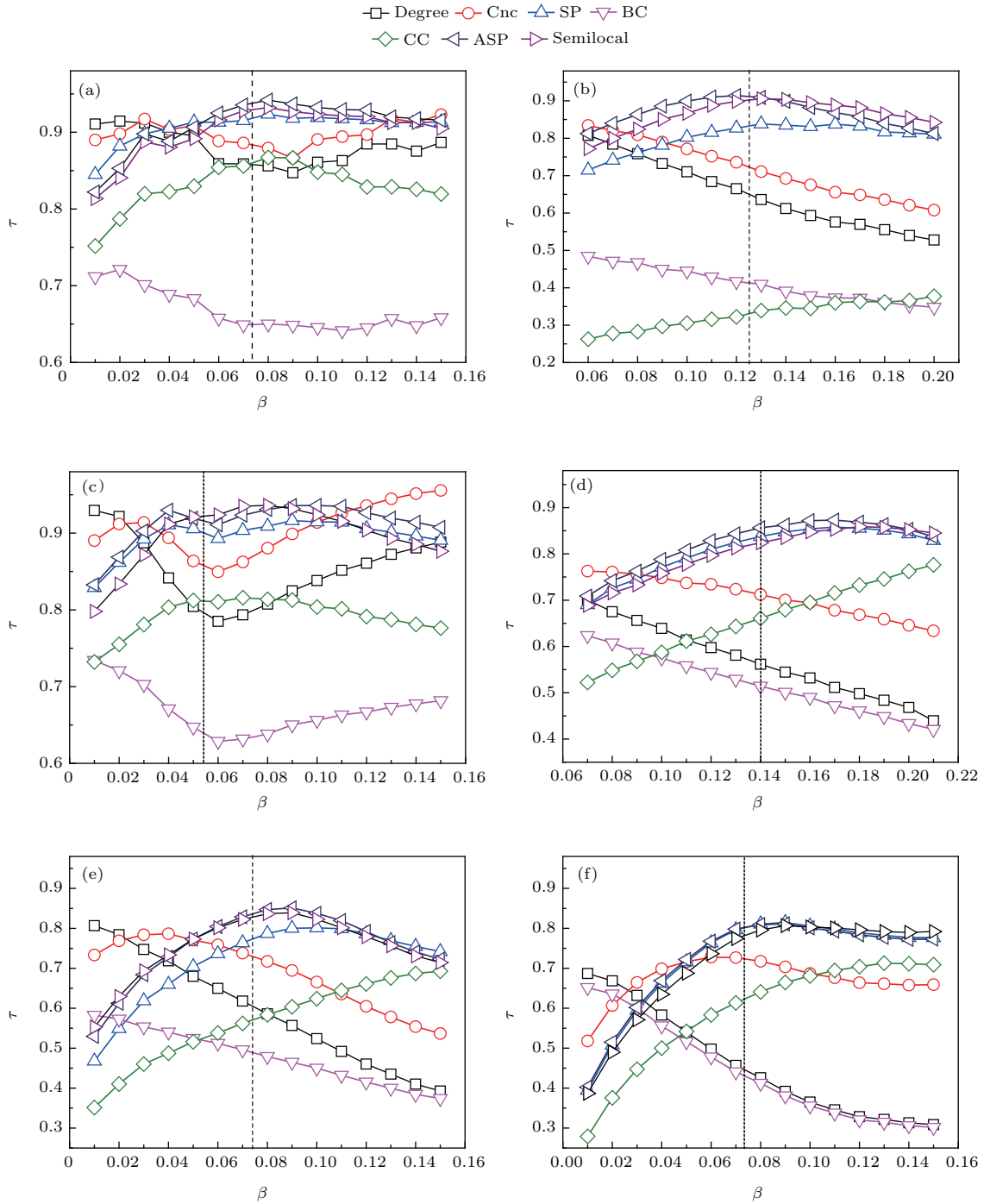


图 3 (网刊彩色) 不同指标评估准确性对比 (a) Word 数据集; (b) Netscience 数据集; (c) Email 数据集; (d) Yeast 数据集; (e) Blog 数据集; (f) Router 数据集

Fig. 3. (color online) Comparison of accuracy evaluation among various indices: (a) Word; (b) Netscience; (c) Email; (d) Yeast; (e) Blog; (f) Router.



绝大多数节点的介数很小, 通过介数进行影响力排序, 节点间区分度不大, 而实际上网络中介数相近的节点的传播能力存在较大差异. SP 指标、核数中心性和 ASP 指标的评估值与 SIR 影响节点数则呈现较强的正相关性, 其中 ASP 指标的相关性结果比 SP 指标好, 可见 ASP 指标在评价节点传播影响力时具有优势.

在相关性实验中, 信息传播率为网络传播阈值, 实验结果只体现特定传播率下的相关性情况. 为了更全面地评价各个指标在不同传播率下的排序准确性, 设置传播率区间为  $[|\beta_{th}| - 7\%, |\beta_{th}| + 7\%]$  (若  $\beta_{th} \leq 0.07$ , 传播率区间取为  $[0.01, 0.15]$ ), 并且将  $\tau$  值作为准确性度量值进行实验, 结果如图 3 所示. 从图中可以看出, 传播率较小时核数指标准确率普遍较高, 这是由于核数指标考虑了节点度与核数, 当传播率较小时, 从源节点发起的 SIR 传播过程容易局限于局部邻域, 此时节点度越大感染到的节点也越多, 核数指标正好适合这一情况. 当传播率在传播阈值附近时, 除了 Router 网络, ASP 指标的准确性比其他指标高, 这是因为传播率适中时, 节点局部高聚簇性能够使节点获得更多的将信息扩散出去的途径, ASP 指标充分考虑了这种因素. 当传播率更大时, 可以发现 ASP 指标的优势与 SP 指标相比在逐渐削弱, 这是因为传播率大到一定程度时, 信息可以轻易地扩散出去, 此时节点的局部聚簇性对信息的扩散作用并不明显. 在 Router 网络中, 由于网络结构较为稀疏, 节点间的冗余链接少, 因此 ASP 指标与 SP 指标的实验结果相差不多.

### 5 模拟数据集上的实验结果

除了真实数据集, 实验还使用了 Lancichinetti-Fortunato-Radicchi (LFR) [46] 数据模型生成的人工数据集. 通过设置不同的 LFR 参数, 可以生成不同拓扑特征的网络结构. 设置 LFR 参数如下:  $N = 2000$ ,  $c_{min} = 20$ ,  $c_{max} = 50$ ,  $k_{max} = 30$ ,  $\mu = 0.1$ , 其中  $N$  为网络节点数,  $c_{min}$  和  $c_{max}$  分别代表社区的最小和最大规模,  $k_{max}$  表示网络的最大度,  $\mu$  为混合参数. 调整平均度  $\langle k \rangle$  来调节网络的紧密程度, 分别生成  $\langle k \rangle = 5, 10, 15$  的三个网络数据集.

三个模拟数据集上的实验结果如图 4 所示, 实验结果证明随着网络稀疏性与信息传播率的变化,

ASP 指标与 SP 指标对节点影响力排序的相对准确性也发生变化. 在  $\langle k \rangle = 5$  与  $\langle k \rangle = 10$  的 LFR 数据集中, 传播率较小时, SP 指标略优于 ASP 指标, 这个结果与真实数据集上的原因类似, 都是因为传播率偏小时节点的真实影响力接近于度; 当传播率更大时, ASP 指标相比其他指标有明显的优势. 尤其在集聚程度高的网络中, 如图 4(c) 所示, ASP 指标在不同传播率下比 SP 指标更具有优势.

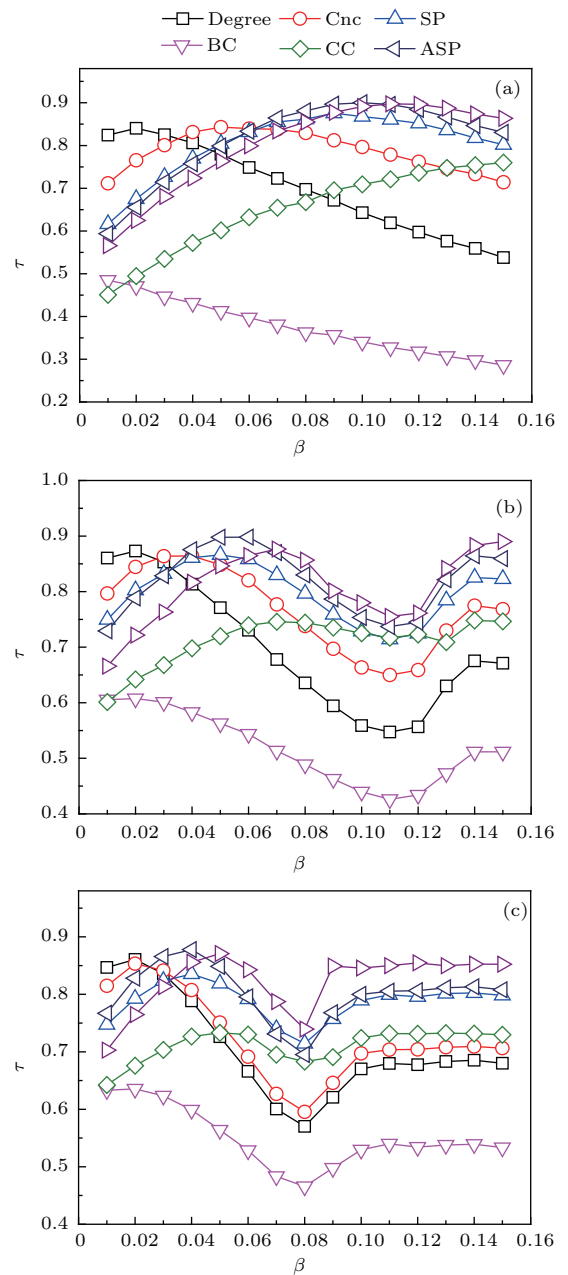


图 4 (网刊彩色) LFR 模拟数据集上各指标影响力排序准确性对比 (a)  $\langle k \rangle = 5$ ; (b)  $\langle k \rangle = 10$ ; (c)  $\langle k \rangle = 15$   
 Fig. 4. (color online) Comparison of accuracy evaluation among different centralities on three LFR datasets: (a)  $\langle k \rangle = 5$ ; (b)  $\langle k \rangle = 10$ ; (c)  $\langle k \rangle = 15$ .

## 6 考虑不同阶次内的邻居对算法排序结果的影响

最后比较了ASP算法考虑不同阶次内邻居时的排序效果, 考虑到 $(1/(k))^l$ 随着 $l$ 的增大会快速衰减, 因此对于4阶及4阶以上的邻居, 只将其最短路径纳入计算范围. 结果如图5所示, 算法效果并不总是随着阶次的提高而变好, 大多在3阶处取得最优. 可见考虑更高阶的邻居, 只会增大算法的计算复杂度, 对算法精度的提升帮助并不大, 因此ASP算法只将节点与邻域节点间三步内的邻居纳入计算范围.

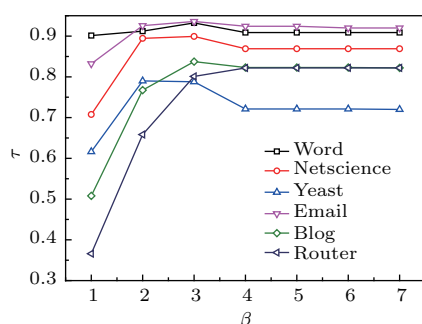


图5 (网刊彩色) 不同阶次内邻居对算法排序结果的影响 ( $S$ 表示算法中邻居节点的最高层级)

Fig. 5. (color online) The effect of neighbors in different orders on the results of algorithm sorting ( $S$  represents the highest order of neighbor nodes in the algorithm).

## 7 结论

准确度量复杂系统中节点的传播影响力, 对于控制流言在网络中的传播、预防网络攻击、设计有效的广告投放策略等具有现实意义. 综合考虑节点与三步内邻居间的有效可达路径以及信息传播率, 本文提出了一种SP指标的改进算法, ASP算法. 在同样只考虑三步内邻居的条件下, 与SP指标相比, ASP指标几乎不增加算法复杂度. 在多个真实数据集和人工数据集上的实验证明, 本文提出的指标可以在更广的信息传播率下取得更为准确的排序结果. 本文算法对于理解节点局部聚簇性对节点传播影响力的影响具有一定意义.

### 参考文献

[1] Dorogovtsev S N, Mendes J F F, Samukhin A N 2000 *Phys. Rev. Lett.* **85** 4633

[2] Lü L Y, Medo M, Yeung C H, Zhang Y C, Zhang Z K, Zhou T 2012 *Phys. Rep.* **59** 1

[3] Papadopoulos F, Kitsak M, Serrano M A, Boguna M, Krioukov D 2012 *Nature* **489** 537

[4] Tang J, Perra M A, Guasch T 2016 *Transport Res. C* **67** 357

[5] Barabási A L, Albert R 1999 *Science* **286** 509

[6] Watts D J, Strogatz S H 1998 *Nature* **393** 440

[7] Lü L Y, Chen D B, Zhou T 2011 *New J. Phys.* **13** 123005

[8] Medo M, Zhang Y C, Zhou T 2009 *Europhys. Lett.* **88** 38005

[9] Pastor-Satorras R, Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200

[10] Albert R, Barabási A L 2002 *Rev. Modern Phys.* **74** 47

[11] Castellano C, Fortunato S, Loreto V 2009 *Rev. Modern Phys.* **81** 591

[12] Yang J, Yao C, Ma W, Chen G 2010 *Physica A* **389** 859

[13] Morone F, Makse H A 2015 *Nature* **524** 65

[14] Zhang J X, Chen D B, Zhao Z D 2016 *Sci. Rep.* **6**

[15] Albert R, Jeong H, Barabási A L 1999 *Nature* **401** 130

[16] Chen D B, Lu L Y, Shang M S, Zhang Y C, Zhou T 2012 *Physica A* **391** 1777

[17] Stephenson K, Zelen M 1989 *Soc. Netw.* **1** 11

[18] Borgatti S P 2005 *Soc. Netw.* **27** 55

[19] Sabidussi G 1966 *Psychometrika* **31** 581

[20] Freeman L C 1977 *Sociometry* **40** 35

[21] Kleinberg J M 1999 *JACM* **46** 604

[22] Brin S, Page L 1998 *Comput. Networks. Isdn.* **30** 107

[23] Radicchi F, Fortunato S, Markines B, Vespignani A 2009 *Phys. Rev. E* **80** 056103

[24] Lü L Y, Zhang Y C, Yeung C H, Zhou T 2011 *PLoS ONE* **6** e21202

[25] Lü L Y, Zhou T, Zhang Q M, Stanley H E 2016 *Nat. Commun.* **7** 10168

[26] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Nat. Phys.* **6** 888

[27] Bae J, Kim S 2014 *Physica A* **395** 549

[28] Liu Y, Tang M, Zhou T, Do Y 2016 *Physica A* **452** 289

[29] Duan J M, Shang M S, Cai S M, Zhang Y X 2015 *Acta Phys. Sin.* **64** 200501 (in Chinese) [段杰明, 尚明生, 蔡世民, 张玉霞 2015 物理学报 **64** 200501]

[30] Liu J G, Lin J H, Guo Q, Zhou T 2016 *Sci. Rep.* **6** 21380

[31] 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]

[32] 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]

[33] Lü L Y, Chen D B, Ren X L, Zhang Q M, Zhang Y C, Zhou T 2016 *Phys. Rep.* **650** 1

[34] Bao Z K, Ma C, Xiang B B, Zhang H F 2017 *Physica A* **468** 391

[35] Newman M E J 2005 *Soc. Netw.* **27** 39

[36] Fowler J H, Christakis N A 2008 *Br. Med. J.* **337** a2338

[37] Newman M E J 2002 *Phys. Rev. E* **66** 016128

[38] Pastor-Satorras R, Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200

[39] Kendall M G 1945 *Biometrika* **33** 239

[40] Knight W R 1966 *J. Amer. Statist. Assoc.* **61** 436

[41] Newman M E J 2006 *Phys. Rev. E* **74** 036104

[42] Guimera R, Danon L, Diaz-Guilera A, Giralt F, Arenas A 2003 *Phys. Rev. E* **68** 065103

- [43] Jeong H, Mason S P, Barabasi A, Oltvai Z N 2001 *Nature* **1** 41
- [44] Xie N 2006 *M.S. Dissertation* (Bristol: University of Bristol)
- [45] Spring N, Mahajan R, Wetherall D 2002 *IEEEACM Trans. Netw.* **1** 2
- [46] Lancichinetti A, Fortunato S, Radicchi F 2008 *Phys. Rev. E* **78** 046110

# An improved evaluating method of node spreading influence in complex network based on information spreading probability\*

Ruan Yi-Run<sup>1)†</sup> Lao Song-Yang<sup>1)</sup> Wang Jun-De<sup>1)</sup> Bai Liang<sup>1)</sup> Hou Lü-Lin<sup>2)</sup>

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

2) (*Joint Service College, National Defense University, Beijing 100858, China*)

( Received 19 May 2017; revised manuscript received 4 July 2017 )

## Abstract

How to evaluate the node spreading ability and how to find influential nodes in complex networks are crucial to controlling diseases and rumors, accelerating or hindering information from diffusing, and designing effective advertising strategies for viral marketing, etc. At present, many indicators based on the shortest path, such as closeness centrality, betweenness centrality and the (SP) index have been proposed to evaluate node spreading influence. The shortest path indicates that the information transmission path between nodes always selects the optimal mode. However, information does not know the ideal route from one place to another. The message does not flow only along geodesic paths in most networks, and information transmission path may be any reachable path between nodes. In the network with high clustering coefficient, the local high clustering of the nodes is beneficial to the large-scale dissemination of information. If only the information is transmitted according to the optimal propagation mode, which is the shortest path propagation, the ability to disseminate the node information would be underestimated, and thus the sorting precision of node spreading influence is reduced. By taking into account the transmission rate and the reachable path between a node and its three-step inner neighbors, we design an improved method named ASP to generate ranking list to evaluate the node spreading ability. We make use of the susceptible-infected-recovered (SIR) spreading model with tunable transmission rate to check the effectiveness of the proposed method on six real-world networks and three artificial networks generated by the Lancichinetti-Fortunato-Radicchi (LFR) benchmark model. In the real data sets, the proposed algorithm can achieve a better result than other metrics in a wide range of transmission rate, especially in networks with high clustering coefficients. The experimental results of the three LFR benchmark datasets show that the relative accuracy of ranking result of the ASP index and the SP index changes with the sparseness of the network and the information transmission rate. When the information dissemination rate is small, SP index is slightly better than the ASP index. The reason for this result is that when the transmission rate is small, the node influence is close to the degree. However, when the transmission rate is greater, the accuracy of the ASP index is higher than those of other indicators. This work can shed light on how the local clustering exerts an influence on the node propagation.

**Keywords:** complex network, spreading influence, information spreading probability, transmission path

**PACS:** 89.75.Fb, 89.75.Hc

**DOI:** 10.7498/aps.66.208901

\* Project supported by the National Natural Science Foundation of China (Grant Nos. 61302144, 61603408).

† Corresponding author. E-mail: [ruanyirun@163.com](mailto:ruanyirun@163.com)