

一种有效的基于三角结构的复杂网络节点影响力度量模型

韩忠明 陈炎 李梦琪 刘雯 杨伟杰

An efficient node influence metric based on triangle in complex networks

Han Zhong-Ming Chen Yan Li Meng-Qi Liu Wen Yang Wei-Jie

引用信息 Citation: *Acta Physica Sinica*, 65, 168901 (2016) DOI: 10.7498/aps.65.168901

在线阅读 View online: <http://dx.doi.org/10.7498/aps.65.168901>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2016/V65/I16>

您可能感兴趣的其他文章

Articles you may be interested in

花簇分形无标度网络中节点影响力的区分度

Discriminability of node influence in flower fractal scale-free networks

物理学报.2015, 64(20): 208901 <http://dx.doi.org/10.7498/aps.64.208901>

复杂网络可控性研究现状综述

Recent progress in controllability of complex network

物理学报.2015, 64(18): 188901 <http://dx.doi.org/10.7498/aps.64.188901>

基于扩展度的复杂网络传播影响力评估算法

Evaluating influential spreaders in complex networks by extension of degree

物理学报.2015, 64(8): 088901 <http://dx.doi.org/10.7498/aps.64.088901>

面向结构洞的复杂网络关键节点排序

Ranking key nodes in complex networks by considering structural holes

物理学报.2015, 64(5): 058902 <http://dx.doi.org/10.7498/aps.64.058902>

度关联无标度网络上的有倾向随机行走

Biased random walks in the scale-free networks with the disassortative degree correlation

物理学报.2015, 64(2): 028901 <http://dx.doi.org/10.7498/aps.64.028901>

一种有效的基于三角结构的复杂网络节点影响力度量模型*

韩忠明^{1)2)†} 陈炎¹⁾ 李梦琪¹⁾ 刘雯¹⁾ 杨伟杰¹⁾

1) (北京工商大学计算机与信息工程学院, 北京 100048)

2) (食品安全大数据技术北京市重点实验室, 北京 100048)

(2016年5月10日收到; 2016年6月15日收到修改稿)

度量复杂网络中的节点影响力对理解网络的结构和功能起着至关重要的作用. 度、介数、紧密度等经典指标能够一定程度上度量节点影响力, k -shell 和 H-index 等指标也可以应用于评价节点影响力. 然而这些模型都存在着各自的局限性. 本文基于节点与邻居节点之间的三角结构提出了一种有效的节点影响力度量指标模型 (local triangle centrality, LTC), 该模型不仅考虑节点间的三角结构, 同时考虑了周边邻居节点的规模. 我们在多个真实复杂网络上进行了大量实验, 通过 SIR 模型进行节点影响力仿真实验, 证明 LTC 指标相比于其他指标能够更加准确地度量节点的传播影响力. 节点删除后网络鲁棒性的实验结果也表明 LTC 指标具有更好效果.

关键词: 复杂网络, 节点影响力, 三角结构, 关键节点

PACS: 89.75.Hc

DOI: 10.7498/aps.65.168901

1 引言

复杂网络的结构、功能以及两者间的联系一直以来都是网络科学的一个重要研究领域^[1-4], 识别复杂网络中高影响力节点对于复杂网络中的结构、传播以及同步等机理有着重要的理论意义^[5-7], 对理解网络中信息、疾病和谣言等传播和控制、新产品的市场推广营销等应用有着非常实用的价值^[8,9].

度量复杂网络节点影响力从网络科学的诞生起就获得了研究者的大量关注, 最直观的指标是度中心性 (degree centrality)^[10]. 度中心性能一定程度上度量节点的影响力, 然而度中心性并没有考虑节点在网络中所处的位置和邻居节点的影响力. 度值相同的两个节点, 若其中一个节点的邻居节点影响力都很大, 另一个都很小, 则这两个节点

的影响力显然差异很大. Chen 等^[11] 结合节点的度和其邻居节点的度提出了局部中心性 (local centrality), 该指标对节点影响力的度量效果较之度有了很大的提高. 闵磊等^[12] 认为度中心性只有在传播率较小的情况下才有较好的影响力度量效果, 通过对当前节点的邻居节点度的累加进一步扩展了节点度, 提出了扩展度 (exdegree) 指标. Fowler 和 Christakis^[13] 提出的三度影响力原则认为节点不仅可以影响与节点直接相邻的节点, 还能间接影响与邻居节点相邻的节点, 三度以内都可能产生影响. 基于节点度的指标忽略了节点间的结构和节点在网络中所处的位置.

介数中心性和紧密中心性是复杂网络中两种经典的中心性指标, 介数中心性 (betweenness centrality)^[14] 定义为网络中任意两个节点之间的最短路径经过该节点的次数, 介数中心性描述的是信息

* 国家自然科学基金 (批准号: 61170112)、教育部人文社会科学研究基金项目 (批准号: 13YJC860006) 和北京市教委科学研究面上项目 (批准号: KM201410011005) 资助的课题.

† 通信作者. E-mail: hanzm@th.btbu.edu.cn

在网络中传播时经过节点的繁忙程度,若移除介数大的节点,则可能会造成网络拥堵,不利于信息在该网络中的传播.紧密中心性(closeness centrality)^[15]则能够表达节点到达网络其他节点的快慢程度,节点的紧密度越大,则其与网络中其他所有节点平均路径长度越短.

复杂网络中节点的群聚现象^[16]引起了研究节点影响力的相关学者的广泛关注,并从节点的聚类系数与网络的社团结构两个角度去研究节点影响力,节点的聚类系数可以反映节点的邻居节点间联系的紧密程度.Chen等^[17]则从节点获取新邻居节点的能力角度验证了聚类系数对节点影响力的作用.许多真实网络存在着明显的群聚现象通过社团划分算法^[3]可以把网络中的节点划分到多个社团中,在多个社团之间起到桥梁作用的节点其影响力较大,研究者^[18]通过结合节点所连接的社团数量和传统中心性指标来挖掘网络中最有影响力的节点,此类算法的缺点是依赖于社团划分,算法复杂度较高,且在社团数量较少的网络中效果不明显.研究者^[19,20]在影响力度量模型中引入结构洞理论来挖掘连接多个群体的影响力大的节点.然而,此类算法对于一些存在联系紧密的多个核心节点的网络影响力度量效果并不好.Zhang等^[21]提出了一种VoteRank算法用来识别网络中多个最具影响力的节点,并将这些节点作为种子节点能使影响力在大范围内传播.

基于随机游走的影响力度量典型方法有PageRank^[22],HITS^[23]以及LeaderRank等^[6].针对PageRank存在排序结果不惟一的缺陷,Li等^[6]对此改进并提出了LeaderRank算法.在原有网络上增加一个对所有节点双向连接的节点,解决了排序不惟一的问题,并利用排序结果找到网络最具影响力的节点.Li等^[24]通过加权对LeaderRank算法进行了改进.HITS算法同时考虑节点的中心和权威性的排序方法.基于随机游走的影响力方法用邻居节点来刻画节点的影响力,在避免噪声的同时却忽略了节点自身的性质.Kitsak等^[25]对社交网络研究时发现高介数或者度值大的节点不一定是最有影响力的节点.并利用k核分解将节点在位置上从边缘层到核心层分为不同的核数层,认为核心节点(核数值较大的节点)才是影响力大的节点^[25,26].然而,由于网络中可能有局部联系很紧密的小团体,核数值较大的节点未必是网络的核心节

点,多个核数值相同的节点影响力也有可能差距很大.研究者^[27,28]针对此问题进行了改进,H指数是常见的评价学者或期刊影响力的指标,Lü等^[29]研究发现H指数在度量复杂网络中节点传播影响力有很好的效果,并且证明了度、H指数以及核数之间的关联,且H指数对节点影响力的度量比核数有了明显的提升.

综合分析节点影响力研究的相关成果,可以看出好的影响力度量指标必须同时考虑节点在网络中所处的位置和节点周边邻居节点的规模大小.三角结构(triangle)是复杂网络中的一种重要结构特征,但利用三角结构度量节点影响力研究却很少见,Kitsak等^[25]通过k-shell分解去寻找网络中的核心节点,认为网络中的核心区域节点之间通常连接更加紧密,网络边缘区域节点连接则相对稀疏.三个节点两两相互连接,形成一个稳定的三角结构.三角结构具有很高的稳定性.节点与其邻居节点间的三角结构数量多少可以衡量其与邻居节点之间的紧密程度,节点与其邻居节点形成的三角结构数量越多,则他们连接越紧密.然而只考虑节点与其邻居节点之间的结构并不能体现节点所处团体的规模大小,三度影响力等理论认为节点不仅可以影响其邻居节点,还能有效影响邻居节点的邻居节点,为了更加精确度量节点的影响力,还需要考虑邻居节点的度信息,这样能避免局部紧密小团体成为“伪核心”的问题.基于以上思想,本文结合节点与邻居节点之间的三角结构和邻居度提出了一种新颖的节点影响力度量指标(local triangle centrality, LTC),该指标同时考虑了节点与邻居节点的结构和数目特征.LTC具有简单、快速、有效的特性,能够用来在大规模网络上快速计算节点的影响力,在此基础上通过排序可以挖掘出复杂网络中最具影响力的节点.

2 LTC影响力度量模型

用无向图 $G = (VE)$ 表示复杂网络的拓扑结构, V 是图 G 中节点的集合 E 是边的集合 $|V| = n$, $|E| = m$,分别表示图 G 中有 n 个节点和 m 条边,节点 v 的度可表示为 $k(v) = \sum_{u \in V} e_{v,u}$,其中

$$e_{v,u} = \begin{cases} 1, & \text{节点}v\text{和节点}u\text{存在连边,} \\ 0, & \text{节点}v\text{和节点}u\text{不存在连边.} \end{cases}$$

NE_v 表示节点 v 的邻居节点集合, $NE_v = \{w \in V | wv = 1 \text{ 且 } w \in V\}$, 节点 v 和节点 u 之间形成的三角结构数量, 即节点 v 和节点 u 的共同邻居数量可表示为 $T(v, u)$, 其计算公式为

$$T(v, u) = |NE_v \cap NE_u|. \quad (1)$$

本文结合三角结构和节点的邻居节点特征提出节点的影响力度量指标 LTC, 节点 v 的 LTC 指标计算方法如下:

$$C_T(v) = \frac{\sum_{w \in NE_v} S(T(v, w))k(w)^\alpha}{\text{Max}_{v \in V} \left(\sum_{w \in NE_v} S(T(v, w))k(w)^\alpha \right)}, \quad (2)$$

其中, $S(x)$ 为一个响应函数, 设为 sigmod 函数 $S(x) = (1 + \exp(-x))^{-1}$, sigmod 函数可以将 $T(v, u)$ 的值映射到 0 和 1 之间, S 函数具有单调递增的性质, 即节点 v 与其邻居节点之间的三角结构越多, 节点 v 处于核心区域的可能性越大. α 取值为 1, 即节点 v 的邻居节点的度数越大节点 v 的影

响力越大. 分母是为了归一化. LTC 指标同样适用于加权网络, 将连边的权重作为 α 参数的取值可以对加权网络中的节点影响力进行有效度量.

为分析 LTC 指标和其他典型指标的差异, 我们选择空手道俱乐部 (Karata) 网络作为实例, 分别计算节点的 LTC 值、度中心性、介数中心性、紧密中心性、核数以及 H 指标, 然后对比分析, Karata 网络的结构如图 1 所示, 该网络中的节点通过社团划分算法可以分为 2 个社团.

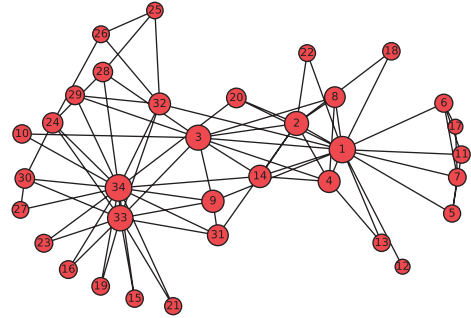


图 1 空手道俱乐部网络 Karata, 共 34 个节点, 78 条连边
Fig. 1. Karate Network with 34 nodes and 78 edges.

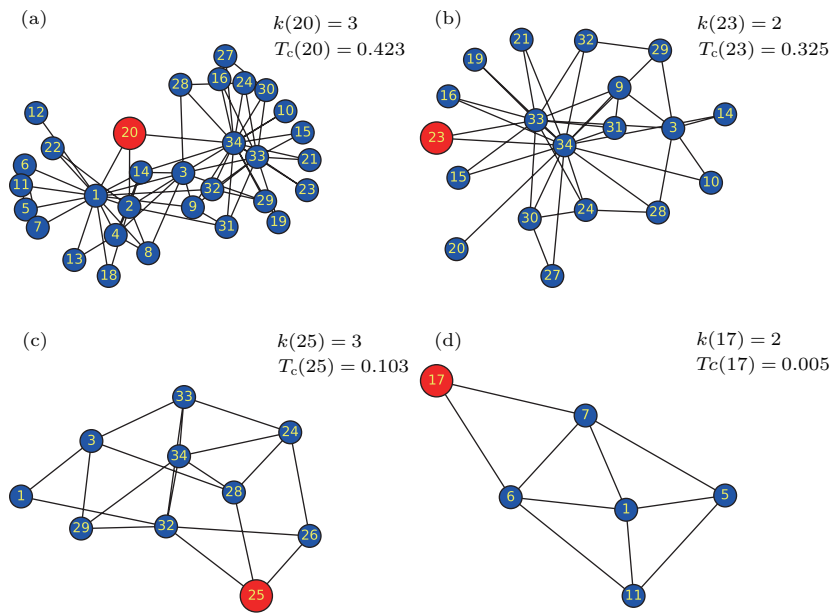


图 2 (网刊彩色) Karata 中的 4 个节点距离 2 以内的节点所构成子图
Fig. 2. (color online) Sub graphs in Karata network of four nodes with the distance less than two.

图 2 是图 1 的 Karate 网络中分别与节点 20, 23, 25, 和 17 距离 2 以内的节点所构成的子图. 对于 Karata 网络这类小规模网络而言, 度中心性可以得到节点 1、节点 2、节点 33 和节点 34 为核心节点. 但对于多个度值相同且较小的节点, 度中心性并不能够准确度量节点的影响力. 如图 2 (a)—

图 2 (d) 所示, 节点 20 和 25 的度都是 3, 但节点 20 与核心节点 1 和节点 2 之间相互连接形成了三角结构, 并与另一个社团的核心节点 34 连接, 在两个社团之间形成了桥接的作用, 其影响力显然比处于社团边缘的节点 25 要大. 节点 23 和 17 的度同样为 2, 但节点 23 与核心节点 33 和 34 连接成三角结构, 而

节点17与其邻居节点之间联系稀疏,且其邻居节点的度较小.显然节点23的影响力要远大于节点17的影响力,对于节点20,23,25和17,度中心性指标并不能区分影响力,LTC指标能够区分并准确地度量其影响力.

表1给出了Karata网络中所有节点各中心性指标对应的值, $F(t)$ 值为通过仿真实验(见3.1节SIR传播模型)获得的节点影响力.表1中对

表1 图1中节点的各项指标值及其传播影响力
Table 1. Different index values of nodes in Karata network.

v	Degree	Betweenness	Closeness	Coreness	H-index	LTC	$F(t)$
1	16	0.438	0.569	4	5	0.912	5.383
2	9	0.054	0.485	4	4	0.708	4.697
3	10	0.144	0.559	4	5	0.867	4.857
4	6	0.012	0.465	4	4	0.661	3.777
5	3	0.001	0.379	3	3	0.234	2.323
6	4	0.03	0.384	3	3	0.269	2.314
7	4	0.03	0.384	3	3	0.269	2.429
8	4	0	0.44	4	4	0.568	3.41
9	5	0.056	0.516	4	4	0.759	4.008
10	2	0.001	0.434	2	2	0.138	2.318
11	3	0.001	0.379	3	3	0.234	2.363
12	1	0	0.367	1	1	0.046	1.805
13	2	0	0.371	2	2	0.182	2.388
14	5	0.046	0.516	4	5	0.711	4.108
15	2	0	0.371	2	2	0.268	2.526
16	2	0	0.371	2	2	0.268	2.495
17	2	0	0.284	2	2	0.01	1.696
18	2	0	0.375	2	2	0.219	2.344
19	2	0	0.371	2	2	0.268	2.396
20	3	0.032	0.5	3	3	0.361	3.181
21	2	0	0.371	2	2	0.268	2.483
22	2	0	0.375	2	2	0.219	2.477
23	2	0	0.371	2	2	0.268	2.455
24	5	0.018	0.393	3	4	0.495	3.125
25	3	0.002	0.375	3	3	0.056	2.148
26	3	0.004	0.375	3	3	0.064	2.19
27	2	0	0.363	2	2	0.169	2.235
28	4	0.022	0.458	3	3	0.291	3.176
29	3	0.002	0.452	3	3	0.278	2.772
30	4	0.003	0.384	3	3	0.46	3.092
31	4	0.014	0.458	4	4	0.49	3.1
32	6	0.138	0.541	3	3	0.555	3.869
33	12	0.145	0.516	4	5	0.789	4.837
34	17	0.304	0.55	4	5	0.797	5.637

图2中的4个节点进行了加粗,通过其对应传播影响力比较,可以看出度、核数以及H指数等指标对于节点20这类在社团间起桥接作用的节点的影响力度量效果较差.从表1可以看出对于介数值为0的多个节点而言,介数无法度量这些节点的影响力.紧密中心性的度量效果在Karata网络上除了区分度不够高之外总体较好.度、核数和H值指标存在多个值相同其传播影响力差异却较大的问题.从图1中可以看出,Karata网络的核心是由多个连接紧密的节点构成,而网络的边缘则连接较为稀疏,LTC指标能体现节点的结构重要性并且解决了核数和H指数存在的多个值相同其传播影响力差异却较大问题.

表2给出了LTC指标的求解算法,算法依次遍历图G中的节点 v ,计算节点 v 与每个邻居节点之间的三角结构数量以及每个邻居节点的度值,然后根据(2)式对每个邻居节点对节点 v 的影响力贡献值进行累加从而得到节点 v 的影响力,遍历结束最终得到图G中所有节点的影响力.

表2 LTC节点影响力度量算法
Table 2. Algorithm for computing LTC index.

Input:	social network $G = (V, E)$
Output:	$T_C(V)$
1	Input: $G = (V, E)$
2	For node v in V do:
3	For u in node v 's neighboring nodes do:
4	Compute the number of triangle $T(v, u)$
5	According to formula (1)
6	Compute sigmoid ($T(v, u)$)
7	End For
8	Accumulation the contribution of node u to node v According to formula (2)
9	Get the influence of node v
10	End For
11	Get the influence of node V

下面分析算法的时间复杂度,对于给定的复杂网络G,网络中共拥有 N 个节点,在存储图结构时采用的边列表的形式,对于G中任一节点 v ,需要计算节点 v 和其每个邻居节点之间的三角结构数量以及邻居节点的度,因此LTC指标度量网络中所有节点的影响力时间复杂度为 $O(N\langle k^2 \rangle)$,从表3给出的各指标算法时间复杂度中可以看出介数中心性和紧密中心性的时间复杂度过高不适用于大规模网络.

表3 各影响力指标算法时间复杂度
Table 3. Time complexity of algorithms of different indexes.

指标	Degree	Betweenness	Closeness	Coreness	H-index	LTC
时间复杂度	$O(N)$	$O(N^3)$	$O(N^3)$	$O(N)$	$O(N\langle k^2 \rangle)$	$O(N\langle k^2 \rangle)$

3 实验设计与评价指标

3.1 实验设计

为了验证LTC指标模型的有效性和适用性,我们分别从节点的传播影响力和网络的鲁棒性两个方面对代表性节点影响力指标进行对比实验,实验对比方法包含度中心性(Degree)、介数中心性(Betweenness)、紧密中心性(Closeness)、核数(Coreness), H指数(H-index)和LTC指标(LTC).

在分析节点的传播影响力时,利用传播模型对节点的影响力进行仿真,然后在不同网络中分析各指标与仿真获得的节点传播能力的相关性,并将LTC指标与代表性指标在单个节点影响力的度量效果上进行对比.

3.2 传播模型

SIR模型^[30]被广泛应用于模拟和研究复杂网络中的信息、观念、病毒以及疾病等传播过程^[31],通过SIR模型进行仿真和模拟可以得到网络节点的传播能力.为了使仿真结果更加接近真实网络中的情况,我们按照文献^[29, 32]的方法对SIR模型以及感染率的阈值选取做了改进.

3.3 Kendall相关系数 τ

肯德尔(Kendall)相关系数可以用来衡量两种排序结果的一致性.假设两个关于 n 个节点的序列 X 和 Y , $X = (x_1, x_2, \dots, x_n)$, $Y = (y_1, y_2, \dots, y_n)$,将 X 和 Y 中的元素一一对应组成一个新的序列 $XY = ((x_1, y_1), \dots, (x_i, y_i), \dots, (x_n, y_n))$.取 XY 中的两个元素 $XY_i = (x_i, y_i)$ 和 $XY_j = (x_j, y_j)$,若 $x_i > x_j$ 且 $x_i > y_j$,或 $x_i < x_j$ 且 $x_i < y_j$,则认为这两个元素是一致的;若 $x_i > x_j$ 且 $x_i < y_j$,或 $x_i < x_j$ 且 $x_i > y_j$ 则认为这两个元素是不一致的;若 $x_i = x_j$ 且 $x_i = y_j$ 这两个元素既不是一致的也不是不一致的.肯德尔相关系数 τ 定义为

$$\tau(X, Y) = \frac{C - D}{(1/2)n(n - 1)},$$

其中 C 代表 XY 任意两个元素对中拥有一致性元素对的数量, D 代表 XY 中不一致的元素对的数量, τ 的取值范围在 -1 — 1 之间,当 τ 为 1 时认为两个序列完全正相关, τ 取 0 时不相关, τ 取 -1 时完全负相关.

3.4 数据集

为了客观评价模型的有效性和适用性,我们选取了7个公开的不同规模的复杂网络数据集. Karate网络为空手道俱乐部网络, Jazz网络是一个爵士音乐合作网络, Facebook网络为科学家合作网, Netscience网络为原始公开数据集网络中的最大联通子集, Email网络为洛维拉·依维尔基里大学成员之间的邮件交互网络, Blog网络为MSN博客空间中博主之间的交流网络, CA-HepPH网络为高能物理领域科学家合作网络.表4所示为实验采用的各真实复杂网络的基本特征,其中 n 为真实复杂网络的节点数量, m 为边的数量, $\langle k \rangle$ 为平均度, $\langle d \rangle$ 为平均距离, $\langle c \rangle$ 为节点的平均聚类系数, β_c 为SIR模型中节点的感染率.

表4 真实网络的基本特征
Table 4. The basic characteristics of the real networks.

Network	n	m	$\langle k \rangle$	$\langle d \rangle$	$\langle c \rangle$	β_c
Karate	34	78	4.588	2.408	0.571	0.147
Jazz	198	2742	27.697	2.235	0.617	0.026
Facebook	324	2218	13.691	3.054	0.466	0.048
Netscience	379	914	4.823	6.042	0.741	0.142
Email	1133	5451	9.622	3.606	0.22	0.056
Blog	3982	6803	3.417	6.227	0.284	0.078
CA-HepPH	12008	118521	19.74	5.212	0.611	0.007

4 实验结果与分析

4.1 节点影响力度量指标和真实传播影响力相关性分析

首先通过3.1节中的SIR模型对真实复杂网络中单个节点的传播影响力进行仿真实验,然后比较

节点的各指标值与节点的传播影响力相关性. 实验中在模拟节点的传播过程时, 取单个节点作为初始传播源, 经 t 时间后达到稳态 (即整个网络中不存在处于感染态的节点), 用 $F(t)$ 表示单个节点的实际传播影响力, $F(t)$ 表示网络中处于免疫态的节点的数量. 对于单个节点重复独立实验 1000 次, $F(t)$ 的最终取值为 1000 次实验的均值.

图 3 是不同指标在真实网络上与节点传播影响力 $F(t)$ 的关系, 纵坐标为 $F(t)$ 的值, 横坐标为归一化后的指标值. 其中核数、H 指和 LTC 指标未作归一化. 如图 3(a)—图 3(f) 所示, 在 Facebook 网络中节点的 LTC 值和 H 指标值相对于其他指标与节点的真实影响力的相关性更高, 度值、介数值和核数大的节点其真实影响力并不一定大. 如图 3(g)—图 3(i) 所示, 在 Netscience 网络中除了 LCT 指标外其他指标都出现了两个问题: 指标值大的节点真实影响力却很小, 和相同指标值的多

个节点真实影响力差异较大. 如图 3(m)—图 3(r) 所示, 在 Email 网络中只有核数指标效果较差, 其他指标效果表现较好. 在 Email 网络中核数最大的节点真实影响力并不是最大, 类似结果在图 3(v) Blog 网络中也明显体现. Email 网络的节点平均聚类系数较小, 故网络中的核心节点并不是连接最紧密的节点, 而 LTC 指标不仅考虑了节点邻居节点的紧密程度, 且考虑周边节点的规模大小, 因此适用于不同特性的复杂网络. 在 Email 网络中, 只有度和 LTC 指标能够找到真实影响力最大的节点.

从真实网络的实验结果上可以看出, 度中心性、紧密中心性、核数、H 指以及 LTC 指标和节点的传播影响力大致都呈现出正相关的趋势, 即指标值越大, 节点的真实传播能力越强. 而介数和节点的传播影响力之间的相关性并不强, 且介数值较大的节点其传播影响力并不一定大, 在多个网络中都存

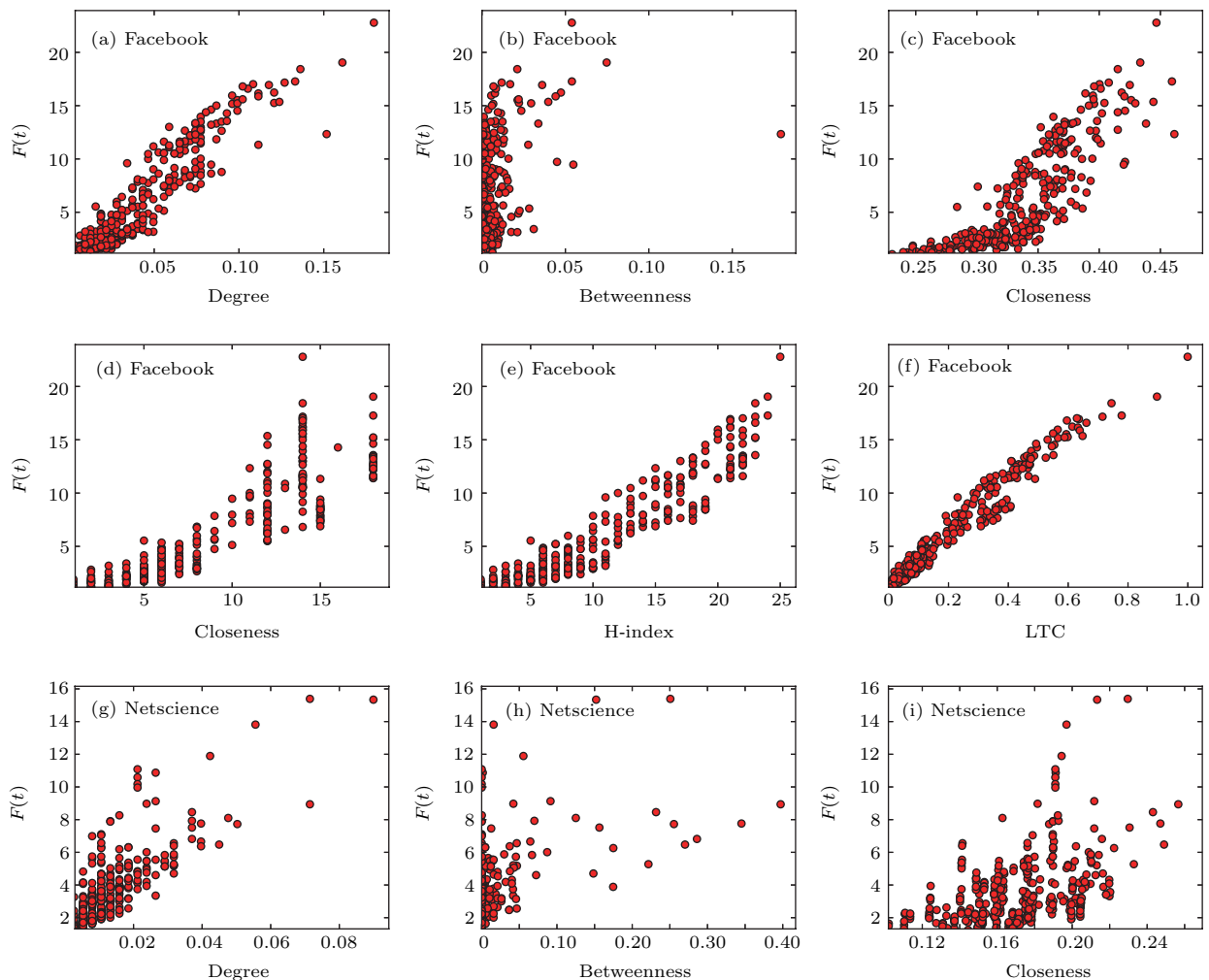


图 3 节点影响力与 6 种指标的相关性

Fig. 3. The correlation between the influence of nodes and the six indexes.

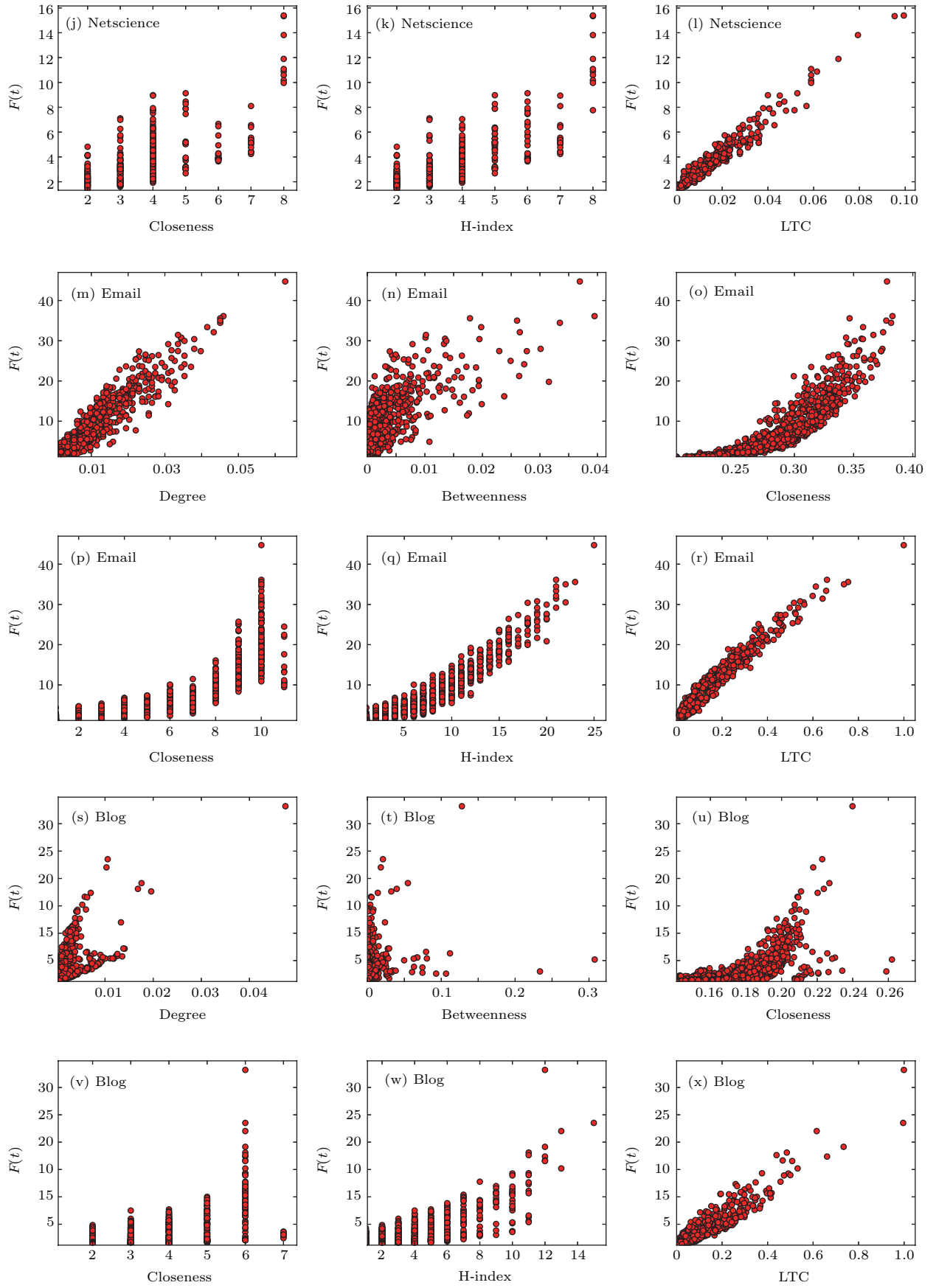


图3 节点影响力与6种指标的相关性(续)

Fig. 3. The correlation between the influence of nodes and the six indexes (Continued).

在类似问题. 同样度数和核数较大的节点的传播能力也不一定很强, 在 Facebook 和 Netscience 网络中存在多个度数较大传播影响力却很小的节点, 核数值较大节点影响力却很低的现象在 Email 和 Blog 网络中尤为明显, 网络中局部连接过于紧密的小团体的存在导致其核数较大但真实影响力却很小的现象在此处得到了很好的验证. 然而在 LTC 指标中则不存在这个问题, 因为其不仅考虑了节点与周边节点的联系紧密程度, 还考虑了周边邻居节点的规模大小. 此外, 观察 Netscience 和 Blog 网络上的结果, 度中心性、紧密中心性、核数和 H 指数这四个指标存在一个共同问题, 指标值相同的多个节点, 其传播影响力却拥有很大的差异, 紧密中心性和 H 指数指标在不同网络表现优劣不同, 在

Facebook 和 Email 网络中表现较好, 其他网络则较差. 在图 3 中 4 个网络上, LTC 值相同的多个节点其影响力差异很小, LTC 指标和单个节点真实影响力表现出良好的线性相关性, 比其他指标效果更好.

表 5 详细列出了度中心性、紧密中心性、核数、H 指以及 LTC 指等指标与节点的传播影响力 $F(t)$ 的肯德尔系数, 肯德尔系数越大说明节点的指标与真实网络中的传播影响力相关性越强, 也就表明指标的影响力度量效果越好. 从表中可以看出 LTC 指标在所有网络中拥有最佳或接近最佳效果. H 指数相对于核数有了明显的提高, 仅次于 LTC 介数的效果最差. 肯德尔系数的结果表明 LTC 在不同网络上的适用性最好.

表 5 在 6 种不同规模真实网络中, 节点影响力与各指标的肯德尔相关系数 τ
Table 5. Kendall correlation coefficient τ values in six real networks with different scale.

网络	Netscience	Facebook	Jazz	Email	Blog	CA-HepPH
Degree	0.611	0.789	0.824	0.787	0.595	0.719
Betweenness	0.389	0.365	0.469	0.63	0.58	0.403
Closeness	0.343	0.717	0.725	0.813	0.485	0.704
Coreness	0.563	0.769	0.797	0.81	0.576	0.667
H-index	0.609	0.814	0.858	0.818	0.589	0.566
LTC	0.831	0.881	0.912	0.866	0.767	0.689

4.2 影响力度量指标差异分析

为了进一步分析 LTC 与各指标影响力度量效果差异性, 将 LTC 与介数中心性和 H 指数分别在 4 个代表网络上的单个节点影响力进行了对比性分析. 图 4 所示是 LTC 与介数中心性的对比结果. 图 4 横坐标代表节点 LTC 指标值, 纵坐标为对比指标值, 颜色坐标代表节点通过 SIR 仿真实验获取的传播影响力大小.

如图 4(a), 图 4(b) 和图 4(d) 所示在 Facebook, Netscience 和 Blog 网络中介数值最大的节点的真实影响力并不是最大, 甚至比较小. 并且介数值相近的不同节点其影响力变化范围较大. 而在 4 个网络中, LTC 指标最大的节点影响力都是最大, 并且 LTC 值相近的节点影响力变化都很稳定. 在 Email

网络中, 节点平均聚类系数较小, 网络中节点的群聚现象不明显. 如图 4(c) 所示, 节点的介数和 LTC 指标表现出高度的一致性, 节点影响力度量效果都较好. 综合在 4 个网络中的表现, LTC 指标对节点影响力度量效果更好.

图 5 所示是 LTC 与 H 指数的对比结果. 如图 5(a) 和图 5(c) 所示, 在 Facebook 和 Email 网络中 LTC 和 H 指数的效果都很好. 如图 5(b) 所示, 在 Netscience 网络中, H 指数为 8 的多个节点影响力波动范围较大. 如图 5(d) 所示, 在 Blog 网络中 H 指数值最大的节点影响力并不是最大, 而 LTC 度量出高影响力的个体节点处于网络的核心位置, 所以与传播影响力具有较高的一致性, 波动很小. 因此, LTC 指标对节点影响力的度量效果比 H 指数更加精确.

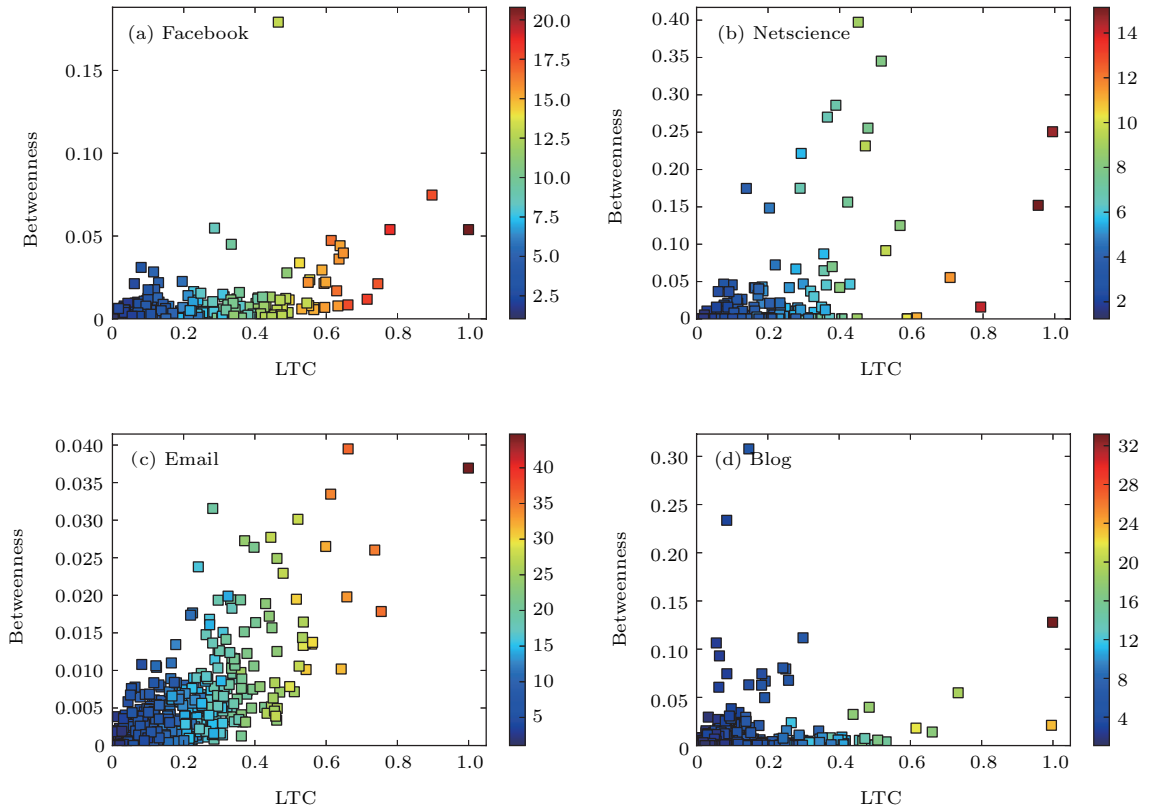


图4 (网刊彩色) LTC和介数在4个代表网络中的影响力度量效果对比

Fig. 4. (color online) Effect comparison of LTC and Betweenness in 4 representative networks.

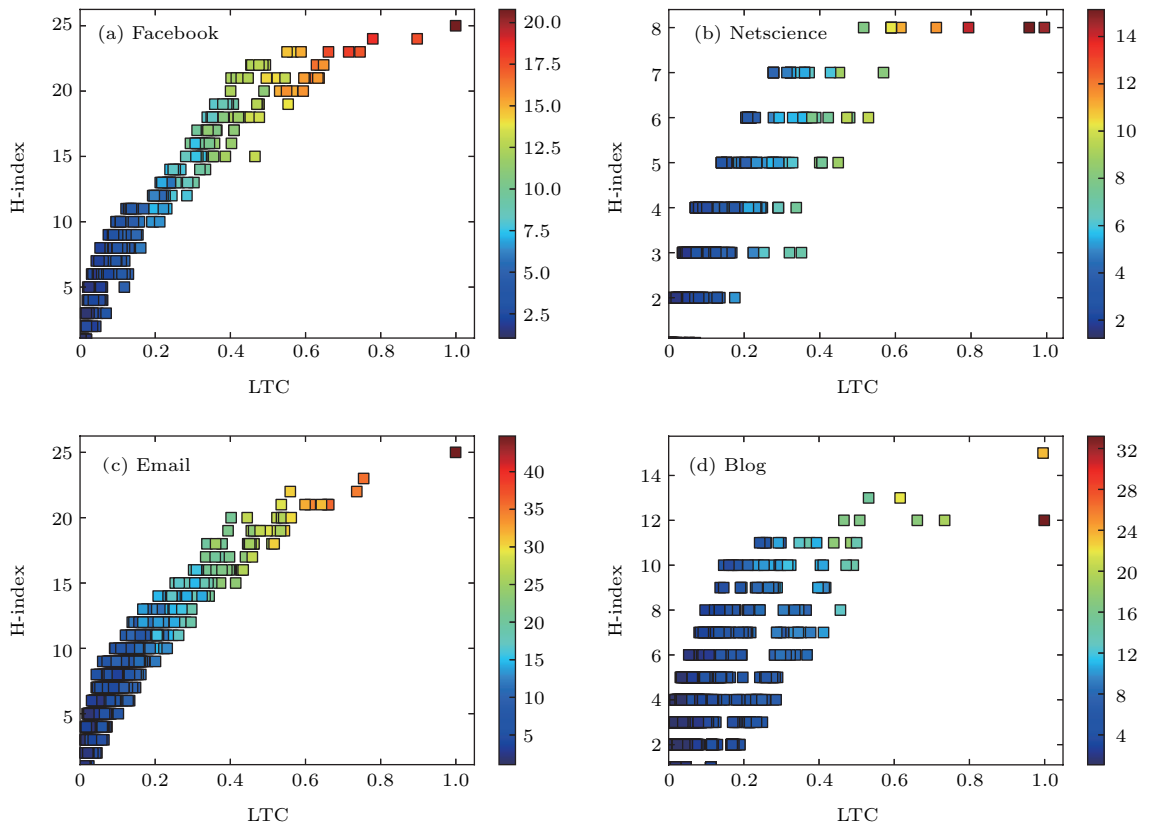


图5 (网刊彩色) LTC和H指数在4个代表网络中的影响力度量效果对比

Fig. 5. (color online) Effect comparison of LTC and H-index in 4 representative networks.

4.3 鲁棒性分析

复杂网络中节点的影响力不仅体现在节点的传播能力上, 对于一些网络通过攻击网络中的少量节点可以使整个网络分崩离析最终“瘫痪”. 因此, 移除节点对于网络鲁棒性的影响也是衡量节点影响力的一个重要方面. 通过移除网络中部分节点, 考虑网络结构的变化来衡量被移除的那部分节点的重要性. 网络的变化越大, 说明这部分节点越重要. 用不同指标对节点影响力进行降序排序, 然后按序移除 i/n 比例的节点, 用 $\sigma(i/n)$ 表示网络剩余节点中属于最大联通集的比例. 若 $\sigma(i/n)$ 值越小, 说明网络中孤立的单个节点或小团体越多, 则移除 i/n 比例的那部分节点越重要. 与节点的传播影响力度量方式不同, 度量节点对网络鲁棒性的重要性更加倾向于介数值大的那些在网络中处在“繁忙”位置的节点, 移除网络中处于紧密连接位置的节点

并不会造成网络的分裂, 而移除那些“桥梁”节点会使整个网络分裂为多个部分, 导致网络崩溃. 此时节点与邻居节点之间的三角结构数量无疑起到的作用是负面的.

根据上述分析, 我们对 LTC 指标中的参数进行修正, (2) 式中 $S(x) = (1+x)^{-1}$, 即一个节点与其邻居节点之间三角结构数量越多, 则移除这个节点对于网络的破坏性越小. α 的取值为 0, 与节点的传播能力不同, 邻居节点的度值此时起到的作用并不明显, 在度量节点的位置重要性时并不需要考虑邻居节点的度值. 画出 i/n 与 $\sigma(i/n)$ 在二维坐标上的曲线, 从而可以分析节点移除后的影响 [33].

图 6 所示为不同规模和不同节点平均聚类系数的 4 个真实复杂网络中按各指标降序排序后, 按序移除 (i/n) 比例的节点与剩余节点中占网络中最大联通集的比例 $\sigma(i/n)$ 关系, 图 6 横坐标表示移除比例, 纵坐标表示 $\sigma(i/n)$ 值.

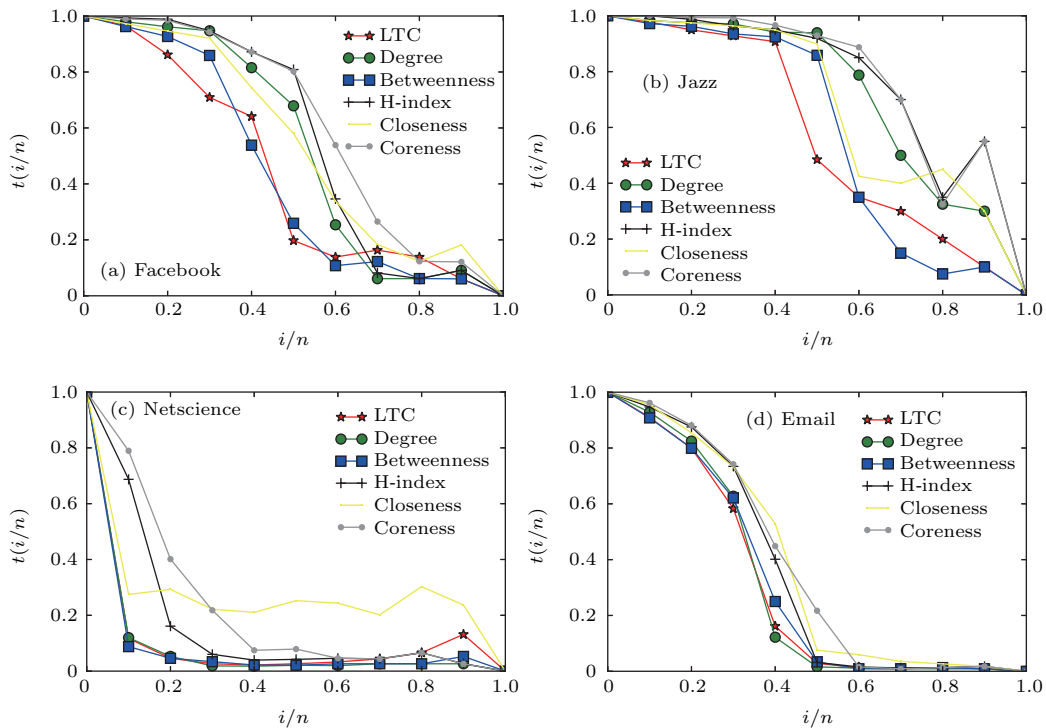


图 6 (网刊彩色) 移除 i/n 比例节点与剩余节点中最大联通集的比例关系图

Fig. 6. (color online) Ratio of giant component to remaining nodes after removing i/n nodes on four representative networks.

如图 6(a) 所示, Facebook 网络中 LTC 指标在移除 10% 部分左右节点后, $\sigma(i/n)$ 值率先迅速下降, 而其他指标要移除 30% 左右才开始明显下降, 说明在 Facebook 网络中 LTC 指标对节点重要性的度量效果比其他指标更好, 所识别的大影响力节点

对网络的鲁棒性影响更加显著. 如图 6(b) 所示, 在 Jazz 网络中和在 Facebook 网络中拥有类似的结果. 但是在移除网络中接近一半数量的节点后, $\sigma(i/n)$ 值才会开始下降, Jazz 网络自身的鲁棒性较好, 网络中的核心节点的数量较多. 如图 6(c) 所示, 在

Netscience网络中只需要移除很少的一部分节点该网络就已经四分五裂. Netscience网络的鲁棒性较差, 其网络核心节点较少, 且一旦这部分节点遭到攻击整个网络就会瘫痪. 在Netscience网络中, 度、介数和LTC指标的节点影响力度量要比H值、核数和接近中心性指标要好.

Email网络的规模相对于其他3个网络规模较大, 但是其平均聚类系数却最低. 如图6(d)所示, 在Email网络中各指标效果较为接近. 由于平均聚类系数较小, 整个网络中节点连接较为稀疏, 因此攻击网络中少量节点并不能够立即使网络陷入瘫痪. 在其他真实网络中, 实验结果与这4个代表网络情况类似.

综合分析实验结果, 在多个真实网络上, 介数和LTC指标都拥有较好的节点影响力度量效果, 介数和LTC指标能够很好地挖掘网络中的核心节点和连接不同团体的“桥梁”节点. 然而介数的算法时间复杂度要远大于LTC指标. 度中心性在不同网络中效果差异较大, 在Netscience和Email网络中度中心性的效果能够接近LTC指标和介数的效果, 而在Facebook和Jazz网络中度中心性的效果较差, 度中心性在不同网络中的适用性较差. 其他指标在所有网络中效果都不如前者, 所以LTC指标相比于其他指标能够更加快速、准确地识别复杂网络中高影响力的节点.

5 结 论

度量复杂网络中节点影响力是复杂网络研究的一个关键科学问题. 本文巧妙地利用三角结构这一局部特征去区分节点在网络中的全局位置. 核心区域通常由多个联系紧密的节点组成, 而边缘区域节点之间连接则相对稀疏. 节点与其邻居节点形成的三角结构数量越多, 则他们连接越紧密, 三角结构数量能反映节点在网络中所处的位置. 同时为了避免将“伪核心”(局部联系紧密的小团体)视为核心区域, 进一步考虑邻居节点的度, 即节点所处位置的周边环境, 在此基础上提出了一种新颖的节点影响力度量指标LTC.

在多个公开真实复杂网络上进行了大量实验, 通过SIR模型验证了LTC指标对节点的传播影响力度量更准确, 且时间复杂度较低, 可以应用于大规模复杂网络. 通过节点删除法对三角结构在度

量节点对整个网络鲁棒性的重要程度上所起作用进行了探索, 节点删除法的实验结果也表明了LTC指标具有较好效果.

分析多个节点影响力和传播的关系时, 多个影响力大的节点同时作为种子节点进行感染时, 感染范围可能会有重叠区域, 所以如何结合节点影响力和网络结构分析多个节点的影响力, 从而使影响力最大化是进一步的研究方向.

参考文献

- [1] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [2] Barabási A L, Albert R 1999 *Science* **286** 509
- [3] Newman M E J, Girvan M 2004 *Phys. Rev. E* **69** 026113
- [4] Klemm K, Serrano M Á, Eguíluz V M, Miguel M S 2012 *Sci. Rep.* **2** 292
- [5] Motter A E, Lai Y C 2002 *Phys. Rev. E* **66** 065102
- [6] Lü L Y, Zhang Y C, Yeung C H, Zhou T 2011 *PloS One* **6** e21202
- [7] Pei S, Makse H A 2013 *J. Stat. Mech. Theory Exp.* **2013** 12002
- [8] Pastor-Satorras R, Vespignani A 2002 *Phys. Rev. E* **65** 036104
- [9] Morone F, Makse H A 2015 *Nature* **524** 65
- [10] Bonacich P 1972 *J. Math. Sociol.* **2** 113
- [11] Chen D, Lü L, Shang M S, Zhou T 2012 *Physica A: Stat. Mech. Appl.* **391** 1777
- [12] Min L, Liu Z, Tang X Y, Chen M, Liu S Y 2015 *Acta Phys. Sin.* **64** 088901 (in Chinese) [闵磊, 刘智, 唐向阳, 陈矛, 刘三妍 2015 物理学报 **64** 088901]
- [13] Fowler J H, Christakis N A 2008 *BMJ* **337** a2338
- [14] Newman M E J 2005 *Social Networks* **27** 39
- [15] Sabidussi G 1966 *Psychometrika* **31** 581
- [16] Palla G, Barabási A L, Vicsek T 2007 *Nature* **446** 664
- [17] Chen D B, Gao H, Lü L Y, Zhou T 2012 *PloS One* **8** e77455
- [18] Zhao Z Y, Yu H, Zhu Z L, Wang X F 2014 *Chin. J. Comput.* **37** 753 (in Chinese) [赵之滢, 于海, 朱志良, 汪小帆 2014 计算机学报 **37** 753]
- [19] Su X P, Song Y R 2015 *Acta Phys. Sin.* **64** 020101 (in Chinese) [苏晓萍, 宋玉蓉 2015 物理学报 **64** 020101]
- [20] Han Z M, Wu Y, Tan X S, Duan D G, Yang W J 2015 *Acta Phys. Sin.* **64** 058902 (in Chinese) [韩忠明, 吴杨, 谭旭升, 段大高, 杨伟杰 2015 物理学报 **64** 058902]
- [21] Zhang J X, Chen D B, Dong Q, Zhao D B 2016 arXiv **1602** 00070
- [22] Berkhin P 2005 *Internet Mathematics* **2** 73
- [23] Kleinberg J M 1999 *JACM* **46** 604
- [24] Li Q, Zhou T, Lü L Y, Chen D B 2014 *Physica A: Stat. Mech. Appl.* **404** 47
- [25] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Nat. Phys.* **6** 888
- [26] Pei S, Muchnik L, Andrade J J S, Zheng Z M, Hernán A M 2014 *Sci. Rep.* **4** 5547
- [27] Liu J G, Ren Z M, Guo Q 2013 *Physica A: Stat. Mech. Appl.* **392** 4154

- [28] Zeng A, Zhang C J 2013 *Phys. Lett. A* **377** 1031
- [29] Lü L Y, Zhou T, Zhang Q M, Stanley H E 2016 *Nat. Commun.* **7** 10168
- [30] Hethcote, Herbert W 2000 *SIAM Rev.* **42** 599
- [31] Pastor S R, Castellano C, Van M P, Vespignani A 2015 *Rev. Mod. Phys.* **87** 925
- [32] Shu P, Wang W, Tang M, Do Y 2015 *Chaos* **25** 063104
- [33] Iyer S, Killingback T, Sundaram B, Wang Z 2013 *PLoS One* **8** e59613

An efficient node influence metric based on triangle in complex networks*

Han Zhong-Ming^{1)2)†} Chen Yan¹⁾ Li Meng-Qi¹⁾ Liu Wen¹⁾ Yang Wei-Jie¹⁾

1) (Beijing Technology and Business University, Beijing 100048, China)

2) (Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing 100048, China)

(Received 10 May 2016; revised manuscript received 15 June 2016)

Abstract

Influential nodes in large-scale complex networks are very important for accelerating information propagation, understanding hierarchical community structure and controlling rumors spreading. Classic centralities such as degree, betweenness and closeness, can be used to measure the node influence. Other systemic metrics, such as k-shell and H-index, take network structure into account to identify influential nodes. However, these methods suffer some drawbacks. For example, betweenness is an effective index to identify influential nodes. However, computing betweenness is a high time complexity task and some nodes with high degree are not highly influential nodes. Presented in this paper is a simple and effective node influence measure index model based on a triangular structure between a node and its neighbor nodes (local triangle centrality (LTC)). The model considers not only the triangle structure between nodes, but also the degree of the surrounding neighbor nodes. However, in complex networks the numbers of triangles for a pair of nodes are extremely unbalanced, a sigmoid function is introduced to bound the number of triangles for each pair of nodes between 0 and 1. The LTC model is very flexible and can be used to measure the node influence on weighted complex networks. We detailedly compare the influential nodes produced by different approaches in Karata network. Results show that LTC can effectively identify the influential nodes. Comprehensive experiments are conducted based on six real complex networks with different network scales. We select highly influential nodes produced by five benchmark approaches and LTC model to run spreading processes by the SIR model, thus we can evaluate the efficacies of different approaches. The experimental results of the SIR model show that LTC metric can more accurately identify highly influential nodes in most real complex networks than other indicators. We also conduct network robustness experiment on four selected networks by computing the ratio of nodes in giant component to remaining nodes after removing highly influential nodes. The experimental results also show that LTC model outperforms other methods.

Keywords: complex network, node influence, triangle, key node

PACS: 89.75.Hc

DOI: 10.7498/aps.65.168901

* Project supported by the National Natural Science Foundation of China (Grant No. 61170112), the Research Fund Project of the Ministry of Education of Humanities and Social Science, China (Grant No. 13YJC860006), and the Scientific Research Common Program of Beijing Municipal Commission of Education, China (Grant No. KM201410011005).

† Corresponding author. E-mail: hanzm@th.btbu.edu.cn