

一种新的网络传播中最有影响力的节点发现方法*

胡庆成[†] 尹龔燊 马鹏斐 高旻 张勇 邢春晓

(清华大学计算机系, 信息技术研究院, 清华信息科学与技术国家实验室, 北京 100084)

(2013年1月25日收到; 2013年3月25日收到修改稿)

在复杂网络的传播模型研究中, 如何发现最具影响力的传播节点在理论和现实应用中都有重大的意义. 目前的研究一般使用节点的度数、紧密度、介数和 K-shell 等中心化指标来评价影响力, 这种方法虽然简单, 但是由于它们仅利用了节点自身的内部属性, 因而在评价影响力时精确度并不高, 普遍性适用性较弱. 为了解决这个问题, 本文提出了 KSC (K-shell and community centrality) 指标模型. 此模型不但考虑了节点的内部属性, 而且还综合考虑了节点的外部属性, 例如节点所属的社区等. 然后利用 SIR (susceptible-infected-recovered) 模型对传播过程进行仿真, 实验证明所提出的方法可以更好地发现最具有影响力的节点, 且可适用于各种复杂网络. 本文为这项具有挑战性研究提供了新的思想和方法.

关键词: 复杂网络, 最具影响力的节点, 社区划分, 中性化测量

PACS: 01.75.+m, 05.90.+m, 89.75.-k, 89.70.Hj

DOI: 10.7498/aps.62.140101

1 引言

现实世界中的诸多系统都以复杂网络形式存在, 比如互联网、社会系统、计算网络、生物网络和社交网络等. 理解复杂网络的拓扑结构和功能是目前国际研究的热点. 在很多科学领域中, 都使用网络来表示系统中成员之间的关系, 如在社交网络中用节点代表人, 边代表人与人之间的联系. 由于这些网络具有很高的复杂性, 因此被称为“复杂网络 (complex network)”^[1-3], 它已成为当前最重要的多学科交叉研究领域之一. 在复杂网络中, 找出最具影响力的节点^[4-9]或意见领袖^[10]在理论和现实应用中都有重大的意义, 例如有效地控制疾病传播^[11]、流言散布^[12,13]、计算机病毒扩散, 还可以传播新产品^[14]、新思想^[15]以推进社会化进程.

在社会网络分析中, 常用“中心性 (centrality)”来测量最有影响力的节点. 文献 [16, 17] 利用度中心性 (degree centrality) 来测量最有影响力的节点, 在符合幂律的非均匀网络中, 度数较大的 hub 节点的传播影响力应该相对较大, 这也是目标免疫和熟人免疫策略的基本依据. 文献 [18] 利用了

介数中心性 (betweenness centrality) 来评价影响力, 介数中心性是以经过某个节点的最短路径的数目来刻画节点的重要性. 文献 [19] 提出 PageRank 算法来评价网页的重要性, 它认为一个网页 (节点) 的重要性取决于其前向链接的数量与质量. 文献 [20, 21] 利用紧密度指标 (closeness centrality) 刻画某个节点到其他节点的难易程度. 文献 [22] 提出通过 K-shell (ks)^[23] 分解来确定最具有影响力的单源节点.

本文认为节点的影响力不只是由节点内部属性决定, 而且其外部属性紧密相关, 这与哲学范畴的内因和外因思想一致. 内部属性如度数、紧密性、介数等中心化指标, 外部属性如所属社区的大小、社区内关系紧密度等. 社区^[24,25]是一个有共同爱好的, 或者位于一个共同地方的、或者同在某个工作场所的、或者具有家庭联系的群体^[26]. 社区内的节点联系紧密, 社区间的节点联系松散.

本文提出了 KSC (K-shell and community centrality) 指标模型. 此模型不但考虑了节点的内部属性, 而且还综合考虑了节点的外部属性, 例如节点所属的社区特性等. 然后利用 SIR (susceptible-

* 国家重点基础研究发展计划 (批准号: 2011CB302302) 和国家高技术研究发展计划 (批准号: 2012AA09A408) 资助的课题.

[†] 通讯作者. E-mail: hqc10@mails.tsinghua.edu.cn

infected-recovered) 模型对传播过程进行仿真, 实验证明本文提出的方法可以更好地发现最具有影响力的节点. 本文为这项具有挑战性研究提供了新的思想和方法.

2 相关研究

在社会网络中, 发现最具有影响力的节点可以帮助我们有效地控制疾病传播、流言散布、计算机病毒扩散, 还可以传播新产品、新思想以推进社会化进程.

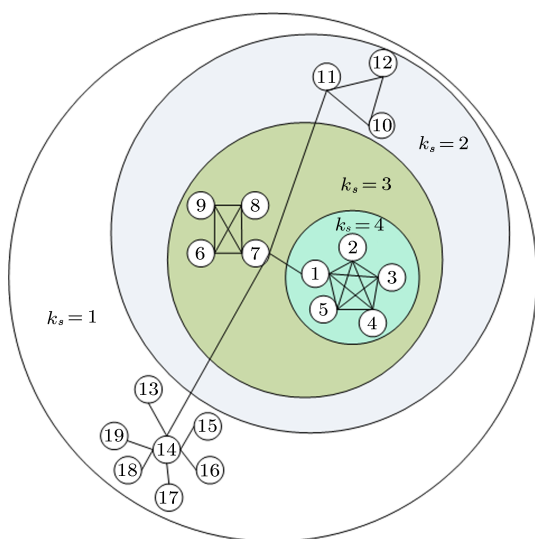


图1 网络分析样例图

2.1 度中心化

度中心化^[27,28]是一种最简单的方法, 用于描述在静态网络中节点所产生的直接影响力. 设网络 $G = (V, E)$ 具有 $n = |V|$ 个节点和 $m = |E|$ 条边, 节点 v 的度数是指与节点 v 连接的节点个数, 用 $C_d(v)$ 表示

$$C_d(v) = d(v), \quad (1)$$

其中 $d(v)$ 称为该节点的度. 如图1所示, 尽管节点14度数最大为7, 但作为初始节点, 它最终的传播范围并不一定是最广泛的, 因为它的邻居节点的度数都很小.

2.2 介数中心化

介数中心化刻画了网络中节点对于信息流动

的影响力. 节点 v 的介数^[18] 定义为

$$C_b(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma(v)}, \quad (2)$$

σ_{st} 表示源节点 s 到目标节点 t 的最短路径数, $\sigma(v)$ 表示节点 s 经过节点 v 到目标节点 t 的最短路径数.

2.3 紧密度

紧密度^[20,21] 用于刻画节点到达其他节点的难易程度:

$$C_c(v) = \sum_{t \in V \setminus s} \frac{1}{d_G(s, t)}, \quad (3)$$

$d_G(s, t)$ 表示源节点 s 到节点 t 的最短距离.

2.4 K-shell 分解方法

K-shell 给出了网络中节点重要性的一种粗粒化的划分, 用 $C_{ks}(v)$ 表示. 网络边缘节点的 K-shell 为1, 然后往内像剥洋葱一样层层进入网络的核心. 首先去除网络中度数小于 k 的所有节点及其连边; 如果在剩下的节点中仍有度值小于 k 的节点, 那么就继续去除这些节点, 直至网络中剩下的节点的度值不小于 k . 依次取 $k = 1, 2, 3, \dots$, 就得到了该网络的 K-shell 分解. 具体过程可参考图一. Kitsak 等^[22] 实验研究表明, 对于单个传播源感染率低的情形, 度数大或者高介数的节点不一定是最有影响力的节点, 而通过 K-shell 分解分析确定的网络核心节点 (即 K-shell 值大的节点) 才是最有影响力的节点; 其研究表明, K-shell 分解方法是一种比较好的影响力节点识别方法, 可以更好地预测疾病等的传播. 当传染病在网络的 K-shell 大的节点爆发时, 病毒总是可以在网络核心通过许多种路径开始感染其他部分, 无论该节点度数的的大与小, 这个结论都是有效的. 这些通路的存在反过来也意味着如果以一个随机节点为源爆发疫情, 有高 K-shell 值的节点更有可能早于其他节点被感染 (疾病预测).

2.5 社区划分算法

社区^[24,25] 是一个有共同爱好的、或者位于一个共同地方的、或者同在某个工作场所的、或者具有家庭联系的群体^[26]. 社区内的节点联系紧密, 社区间的节点联系松散. 社区划分 (community detecting) 的概念前期在 Lin 的论文^[29] 中提出. 如图1, 网络划分为4个社区 (节点1—5, 6—9, 10—12, 13—19). 社区是复杂网络信息传播中最普遍和

最重要的拓扑属性之一, 信息在同一社区内传播迅速, 而在社区之间传播缓慢. Kernighan-Lin (KL) 算法^[30]、快速 Newman (FN) 算法^[31] 和 Guimera-Amaral (GA) 算法^[32] 是经典的复杂网络社区划分算法.

本文使用了 FN 算法. FN 算法是 Newman 提出的基于局部搜索的快速社区划分算法. 其优化目标是极大化网络模块性 (modularity) 评价函数 Q ^[26]. FN 算法属于聚合算法, 它初始假定所有节点各自成为一个社区, 然后每一步根据原有连接合并两个社区, 计算 Q 值并记录下来, 直到所有社区都被合并成为一个社区. 最后根据记录下的 Q 值, 找到最优的划分结果.

3 算法模型

我们认为节点的传播力不但与节点的内部自身属性有关, 如度数、紧密度、介数、K-shell 等, 还与它的外部环境属性相关, 如节点所处外部环境: 社区大小, 社区关联紧密度等. 本文提出节点的影响力是由内因与外因同时决定的这一新颖思想和方法, 即 KSC 中心化指标模型. 模型符合哲学内因与外因对事物同样具有决定作用这一思想理念.

复杂网络 $G = (V, E)$ 中, 节点 v_0 的 KSC 值定义如下:

$$KSC(v_0) = \alpha f_{\text{internal}}(v_0) + \beta f_{\text{external}}(v_0), \quad \alpha + \beta = 1, \quad (4)$$

其中 $f_{\text{internal}}(v_0)$ 为节点的内部影响力, $f_{\text{external}}(v_0)$ 为节点的外部影响力, α 为内部影响因子, β 为外部影响因子, 且满足 $\alpha + \beta = 1$, α 和 β 可以根据实际网络结构与功能设定.

$f_{\text{internal}}(v_0)$ 定义如下:

$$f_{\text{internal}}(v_0) = \frac{K(v_0)}{\max_{v \in V}(K(v))}, \quad (5)$$

其中, $K(v)$ 为节点内部属性值, 可取度数、介数、紧密度和 K-shell 等节点自身属性, $\max_{v \in V}(K(v))$ 为归一化因子.

$f_{\text{external}}(v_0)$ 定义如下:

$$f_{\text{external}}(v_0) = \frac{\sum_{c \in C} d(v_0, c) |c|}{\max_{v \in V} \left(\sum_{c \in C} d(v, c) |c| \right)}, \quad (6)$$

其中 C 为 FN 算法划分的社区集合, $d(v_0, c)$ 为节点 v_0 在社区 c 中的邻居节点的总个数, $|c|$ 为社区 c 的

大小, $\max_{v \in V} \left(\sum_{c \in C} d(v, c) |c| \right)$ 是归一化因子.

本文实验中 α 与 β 可以根据网络的拓扑结构设定; 其中 $f_{\text{internal}}(v_0)$ 选取 K-shell 为内部属性, 也可以为度数、紧密度和介数等, $f_{\text{external}}(v_0)$ 选取为社区属性, 俗语“物以类聚, 人以群分”说明社区内部节点之间有共同的兴趣爱好等, 相似性更大.

表 1 给出了图 1 各节点的各种指标, 如节点 14 度数最大, 但是由于处在网络的边缘, 所以最终影响力为 2, 而节点 7 虽然度数为 6, 但是它是其他几个社区连通的枢纽, 结构洞^[33,34] 概念中重要节点. 节点 1—5 虽然 K-shell 相同, 但是节点 1 是 2—5 连接其他节点的桥接关系. 从表 1 中可以看出各种方法得到的最具影响力节点排名各不相同.

表 1 计算图一中节点的传播力

v	$C_d(v)$	$C_{ks}(v)$	$C_c(v)$	$C_b(v)$	$KSC(v)$	$F(t)$
1	5	4	0.4615385	0.3660131	0.85	3
2	4	4	0.3396226	0	0.8363636	2
3	4	4	0.3396226	0	0.8363636	2
4	4	4	0.3396226	0	0.8363636	2
5	4	4	0.3396226	0	0.8363636	2
6	3	3	0.4	0	0.6068182	2
7	6	3	0.6	0.7581699	0.6579545	3
8	3	3	0.4	0	0.6068182	2
9	3	3	0.4	0	0.6068182	2
10	2	2	0.3050847	0	0.3909091	1
11	3	2	0.4186047	0.2091503	0.4045455	2
12	2	2	0.3050847	0	0.3909091	1
13	1	1	0.3461538	0	0.2227273	1
14	7	1	0.5142857	0.5686275	0.475	2
15	1	1	0.3461538	0	0.2227273	1
16	1	1	0.3461538	0	0.2227273	1
17	1	1	0.3461538	0	0.2227273	1
18	1	1	0.3461538	0	0.2227273	1
19	1	1	0.3461538	0	0.2227273	1

4 模型及实验分析

4.1 SIR 模型

SIR 模型在现实社会网络中已广泛应用于疾病传播、信息传播及谣言传播的研究. 为了验证本文提出的模型, 我们采用 SIR 模型^[35-38] 来模拟仿真传播过程, 并将其结果与我们的模型实验结果做比较.

SIR 模型中有三种状态, 易感染状态 S (susceptible), 感染状态 I (infected), 免疫状态 (recovered).

当个体处于感染状态时,以 β 的概率感染处于易感染状态邻居个体,感染状态的节点以 γ 的概率恢复为免疫状态.

如果 β 值较大,节点传播能力很强,节点将很快感染整个网络,从而很难区分单个个体的重要性;较小的 β 能在有限时间内更好地显示出感染范围.本文中我们设定较小的 $\beta = 0.04$.

4.2 实验数据

考虑到不同的社会网络类型所代表不同的网络拓扑结构特性,我们选取了 4 个真实社会网络数据集^[39] 进行分析比较,表 2 给出各个网络的属性特征: 1) Blogs 网络数据^[40], MSN 博客空间中交流的关系网络; 2) Netscience 合著关系网络^[41], 选择了其中最大连通子图 379 个作者的关系网络; 3) Router 网络,数据是由 Rocketfuel Project^[42] 收集的互联网中路由连接层级关系网络; 4) Email 网络^[43], 是洛维拉·依维尔基里大学成员邮件通信的关系网络. 我们的模型也可应用于其他类型的复杂网络.

其中 n 是节点数, m 为边数, $\langle k \rangle$ 表示网络中平均度数据, $\max(k)$ 为节点中最大度数, d 为节点之间最短路径的平均数, $\max(C_{ks})$ 为网络中最大的 K-shell, C 为整个复杂网络被划分的社区数.

表 2 现实社会网络的属性情况

网名	n	m	$\langle k \rangle$	$\max(k)$	d	$\max(C_{ks})$	C
Blogs	3982	6803	3.42	189	6.227	7	47
Netscience	379	914	4.82	34	6.061	8	19
Router	5022	6258	2.49	106	6.393	7	75
Email	1133	5451	9.62	71	3.716	11	10

4.3 实验效果

我们应用章节 4.1 中所提到的 SIR 模型,对提出的 KSC 指标与度、紧密性、介数和 K-shell 中心化指标进行了分析比较. 实验模拟传播过程中,每次只选取网络中的一个节点作为初始传播节点,设定较短传播时间 ($t = 10$),对每个节点进行 1000 次重复实验取均值,最终感染与恢复的节点总数定义为影响力 $F(t)$.

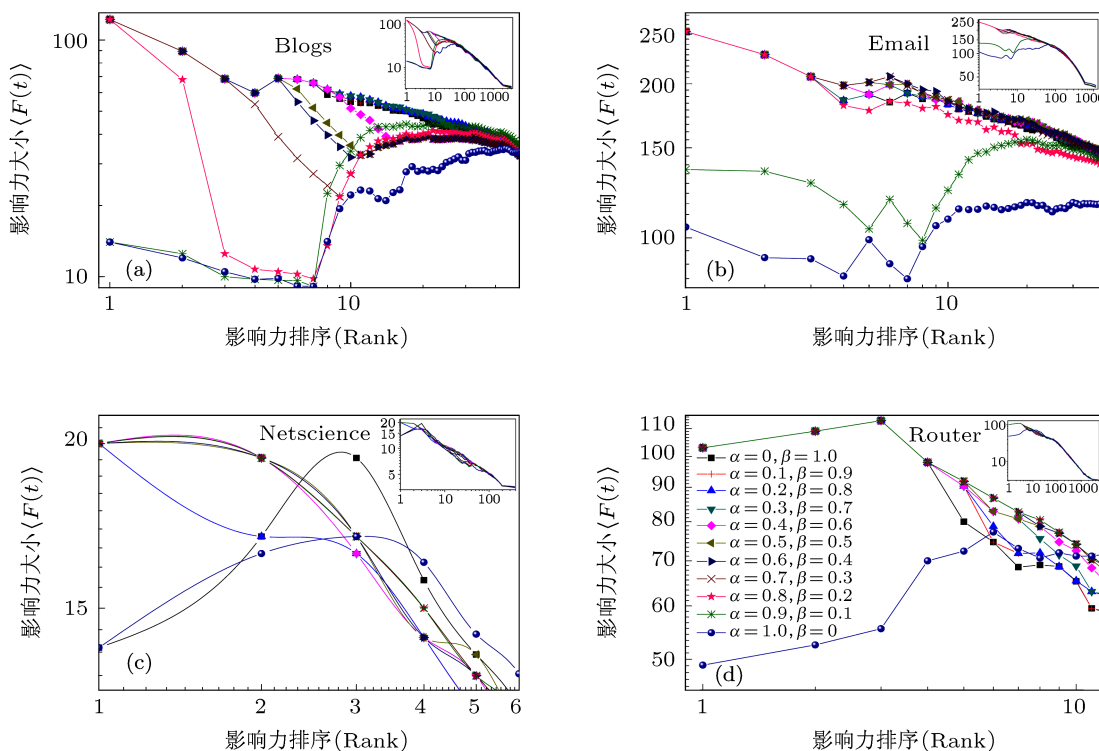


图 2 KSC 模型中内部影响因子 α 与外部影响因子 β 取不同对传播影响力指标的影响分析

图 2 表示了在 KSC 模型中选取不同的内部影响因子 α 与外部影响因子 β ($\alpha + \beta = 1$) 对结果的影响程度. 哲学范畴中事物的运动和变化,是由它

本身固有的内部矛盾引起的,又是同它所处的一定的外在条件相联系的,内因和外因在事物发展中的地位 and 作用是不同的. 这与模型中内部影响因子 α

与外部影响因子 β 对传播影响假设完全一致. 如在图 2(c) 中局部放大了最具影响力的前 10 位, 当 $0 \leq \alpha \leq 0.2, 1 \geq \beta \geq 0.8$, 或 $1 \geq \alpha \geq 0.8, 0 \leq \beta \leq 0.2$ 时可以看到, 影响力曲线波动较大, 影响力排名靠后的节点影响反而大. 即只是单纯的考虑内部属性或外部属性对发现最具有影响的节点来说准确性, 适用性较差.

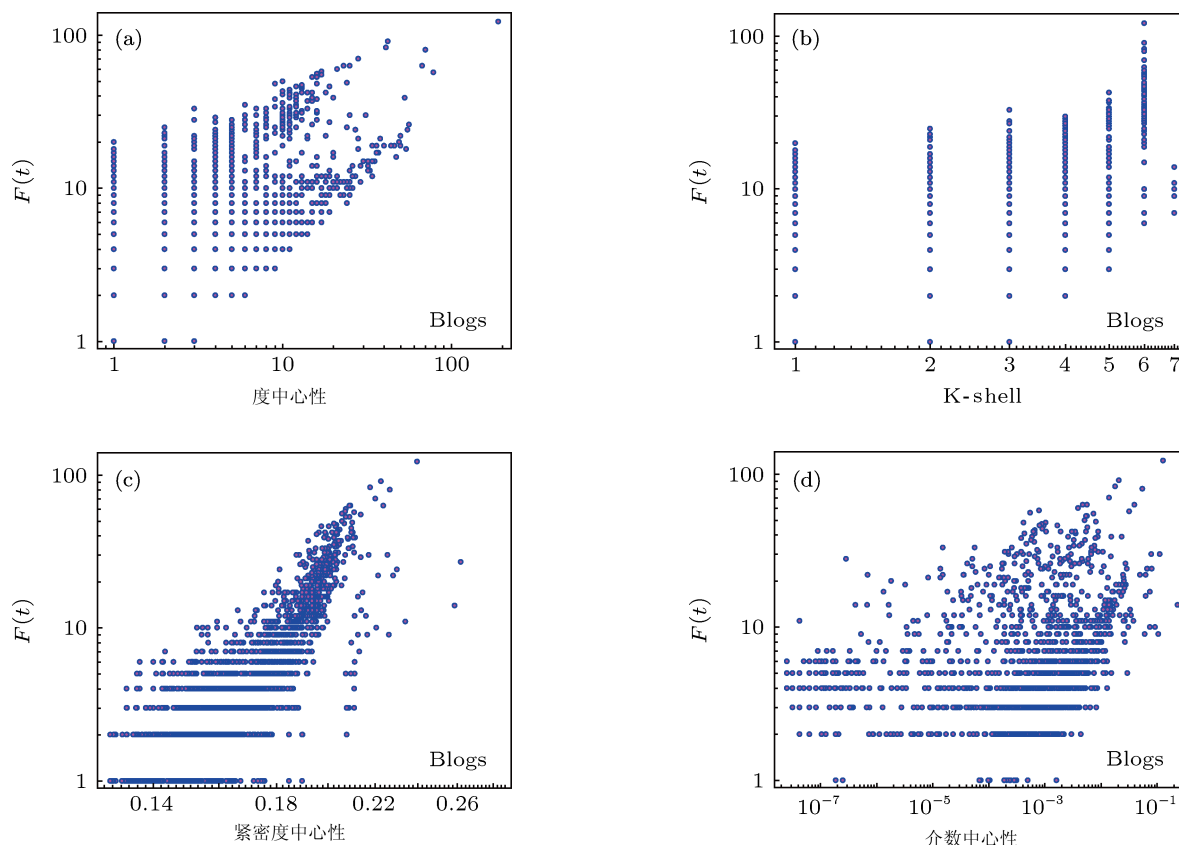
经实验分析得出四种不同拓扑结构网络的内、外影响因子经验取值情况, 如表 3 所示. 可以看出在不同类型的网络结构中, 内、外因所起的权重是不同的, 本文实验中的 α, β 根据表 3 中取较好的经验值, 这从另一方面证明了所提出模型的合理性及普遍性.

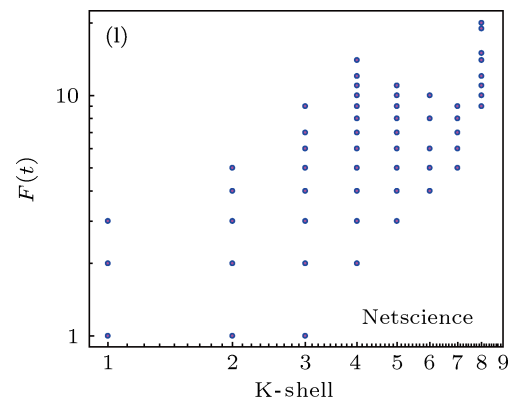
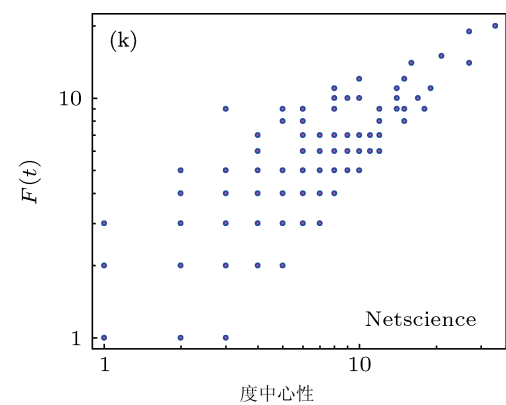
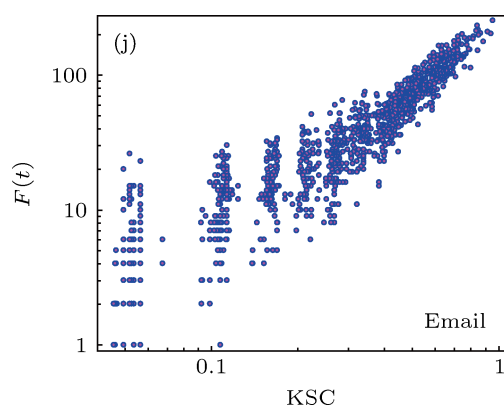
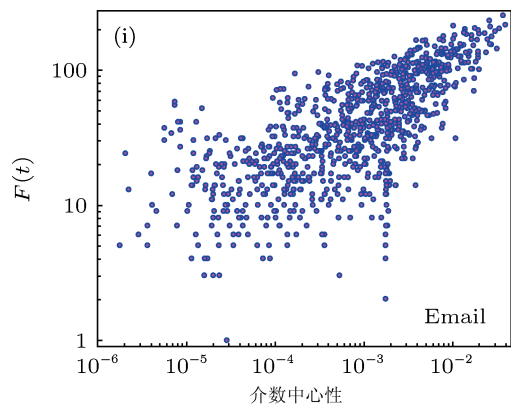
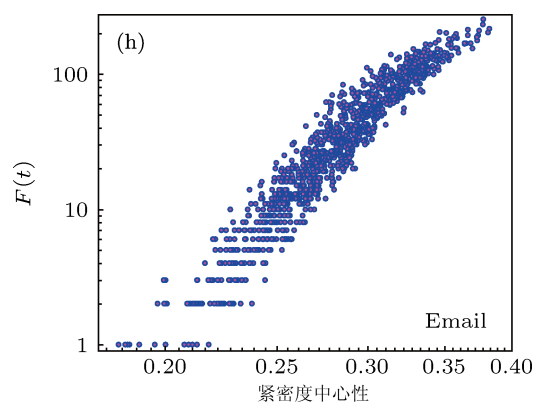
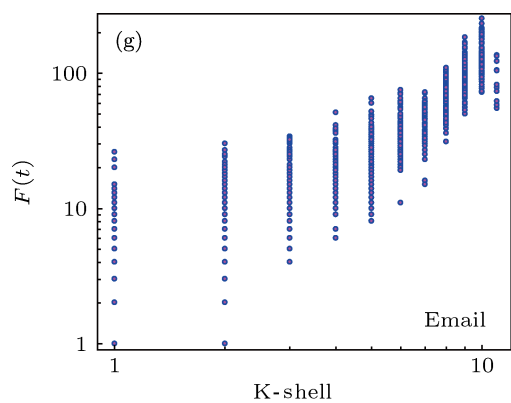
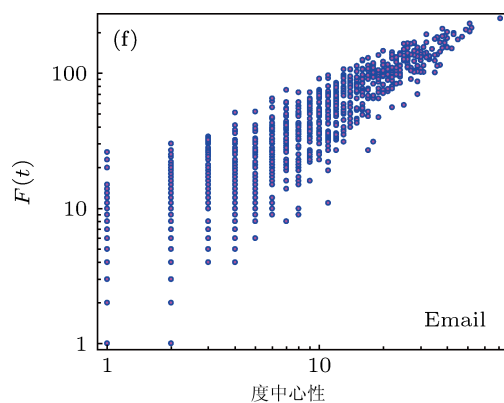
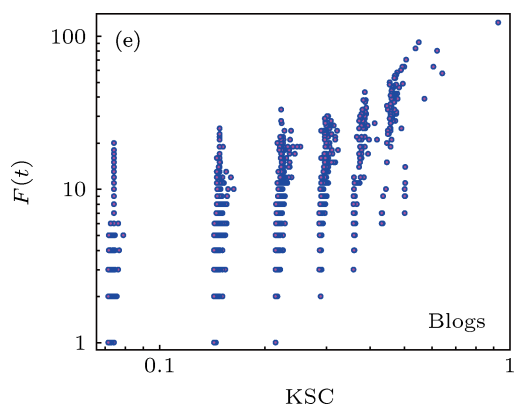
图 3 所表示的是在不同的网络中中心度指标与实际影响力之间的关系.

在 Blogs 网络中, 度, K-shell, 介数指标与影响力 $F(t)$ 关系都不是很明显, 例如在介数指标结果中, 一些介数大的节点 $F(t)$ 比一些介数很小的节点影响力还要低, 但 KSC 指标可以较好地地区分出最具影响力的节点. Email 网络中, 所有指标表现都较好, 紧密度与 KSC 更为理想, 这与网络的拓扑结构有关系. 在 Netscience 网络中, 介数衡量结果最差, 紧密度, K-shell 与节点 $F(t)$ 影响力关系也不明显, KSC 的效果要优于度数. 在路由网络 (Router) 中 K-shell 与 KSC 结果接近, 其他指标结果很难得到传播力较大的节点. 总之, 在不同的网络中传统的中心度指标各有优缺点, 但是我们提出的 KSC 整体效果最好, 都能有效地地区分出最具有影响力的节点, 通用性较强.

表 3 内、外影响因子取值情况

网络名称	内部影响因子取值范围 (α)	外部影响因子取值范围 (β)	较好经验取值	
			α	β
Blogs	0.1—0.4	0.9—0.6	0.2	0.8
Netscience	0.3—0.7	0.7—0.3	0.4	0.6
Router	0.3—0.8	0.7—0.2	0.7	0.3
Email	0.1—0.8	0.9—0.2	0.5	0.5





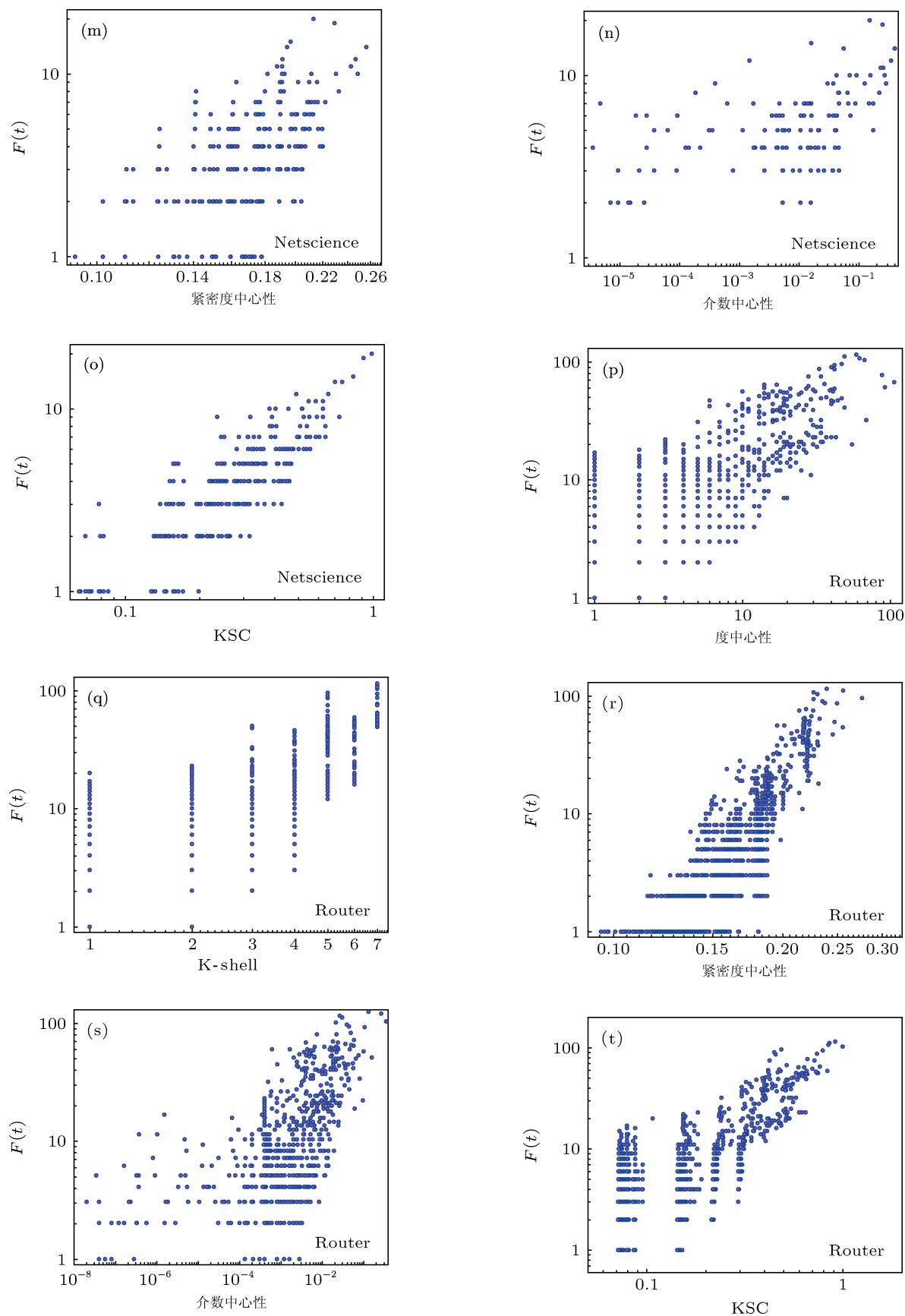


图3 节点影响力分析对比, 列代表5种不同中心化指标, 行表示为4种不同的复杂网络环境, 传播影响力 $F(t)$ ($t=10$), 每个初始节点重复运行1000次取均值

对于单个传播源的情形,最具有影响力的节点并非度数最大的节点或者介数最大的节点,而是 K-shell 最大的节点,这一结论在 Kitsak 等^[22]已在《Nature Physics》指出.图 4 比较了 KSC 与 K-shell 的实验效果, KSC 比 K-shell 增加了外部属性的影

响因素.例如在 Email 网络中,当 K-shell 为 10 时,传播影响力 $F(t)$ ($t = 10$) 并不稳定,而且变化范围较大,并且影响力随着 KSC 值的增大呈现增大趋势.在 4 种复杂网络中明显可以看出当 KSC 一定时 $F(t)$ 相对稳定,颜色变化范围较小.

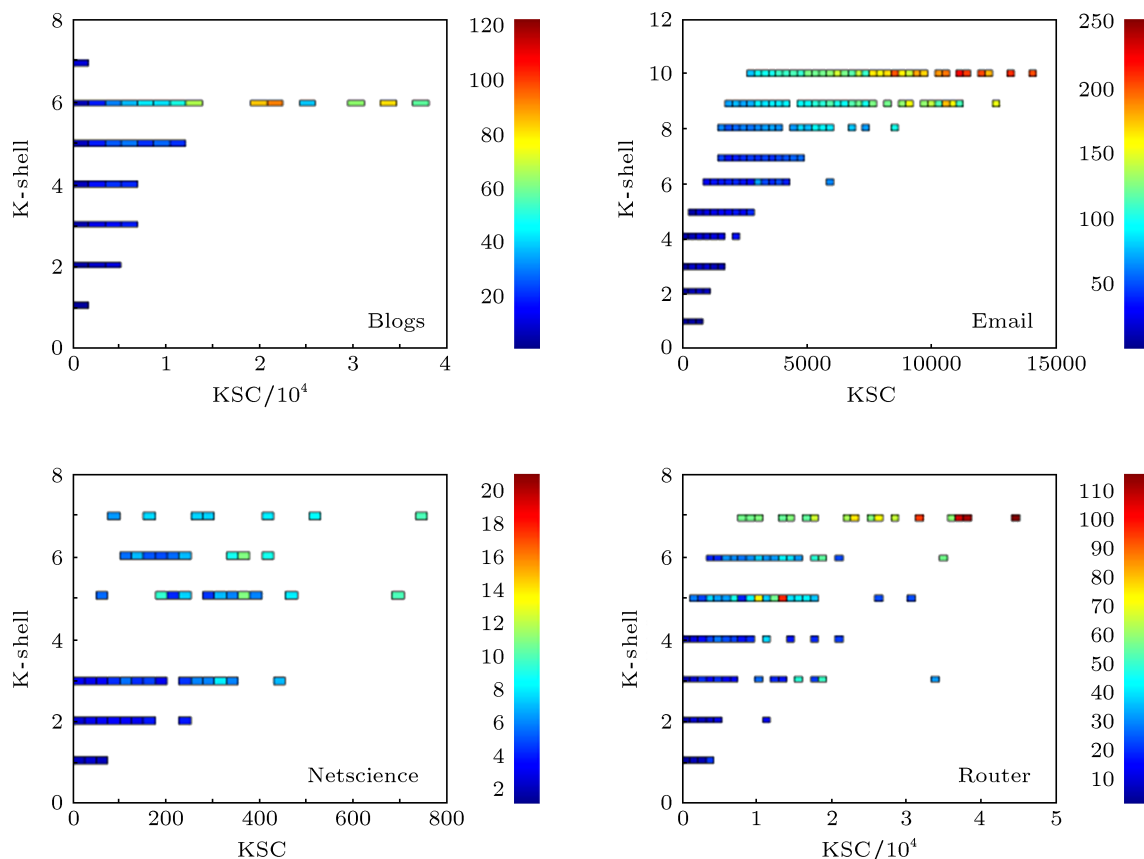


图 4 KSC 与 K-shell 指标在 4 种不同复杂网络中的分析比较, 横轴表示 KSC 的指标值, 纵轴表示 K-shell 的指标值, 颜色坐标代表 SIR 模型仿真出的实际影响力大小 $F(t)$

图 5 比较了 KSC 与介数的实验效果. 可看出在 Email 网络中, KSC 与介数表现正向同步较好, 都能较好地发出最具影响力的节点, 但在其他三个网络环境中 KSC 表现明显优于介数指标. 实验结果表明 KSC 方法明显优于已有的 4 种典型方法. 由此可见, 考虑节点的外部属性对于找出最有影响力的节点意义重大, 且具有较强的通用性, 验证了我们所提出模型的合理性与普遍性.

如果某个节点按该指标得出的排名越靠后, 其实际影响力 $F(t)$ 应该随之减小.

如图 6 所示, 在 Blogs 网络中, KSC 曲线的某一点 (x, y) , 横坐标表示此点用 KSC 指标值得出的排名为 x , 纵坐标表示此点在 SIR 仿真的实际影响力为 y . 理论上, 一个好的指标方法应该表现为向右倾斜单调下降的曲线. 因为对于一个好的指标方法,

从图 6 结果中分析, 可以看到已有的 4 种典型方法在不同网络中影响力均有所波动, 这些方法在不同网络中各有优劣, 通用性不强, 特别是在影响力前 10 中波动范围最为明显. 指标值小的传播影响力反而比指标值大的传播影响力要大, 特别是在 Blogs 网络中, 已有的 4 种指标方法结果都不太精确, 而我们提出的 KSC 指标几乎在所有的网络中都符合向右倾斜单调下降的理论曲线.

本实验充分论证了节点影响力不但由内部属性决定, 而且还与其外部属性密切相关, 说明我们提出的 KSC 模型的合理性及普遍适用性.

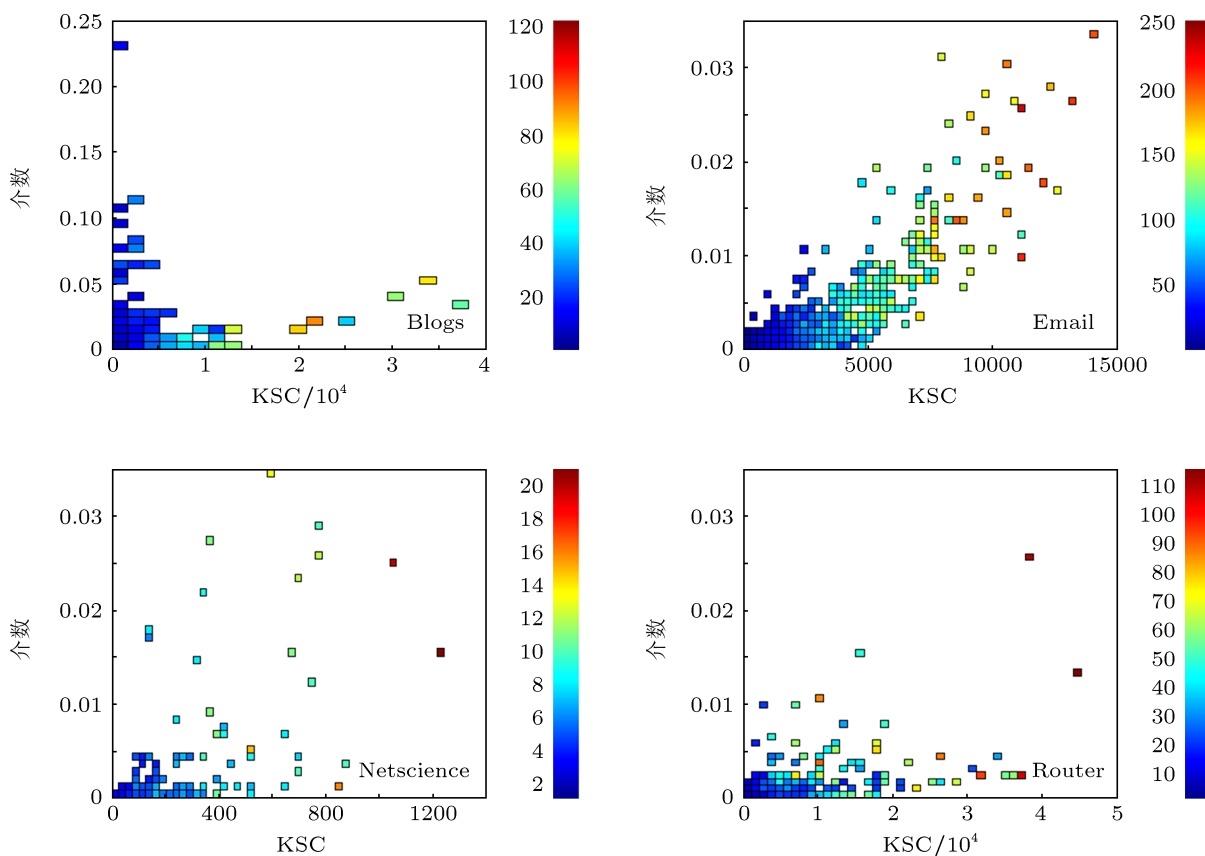


图5 KSC与介数指标在4种不同复杂网络中的分析比较,横轴表示KSC的指标值,纵轴表示介数的指标值,颜色坐标代表SIR模型仿真出的实际影响力大小 $F(t)$

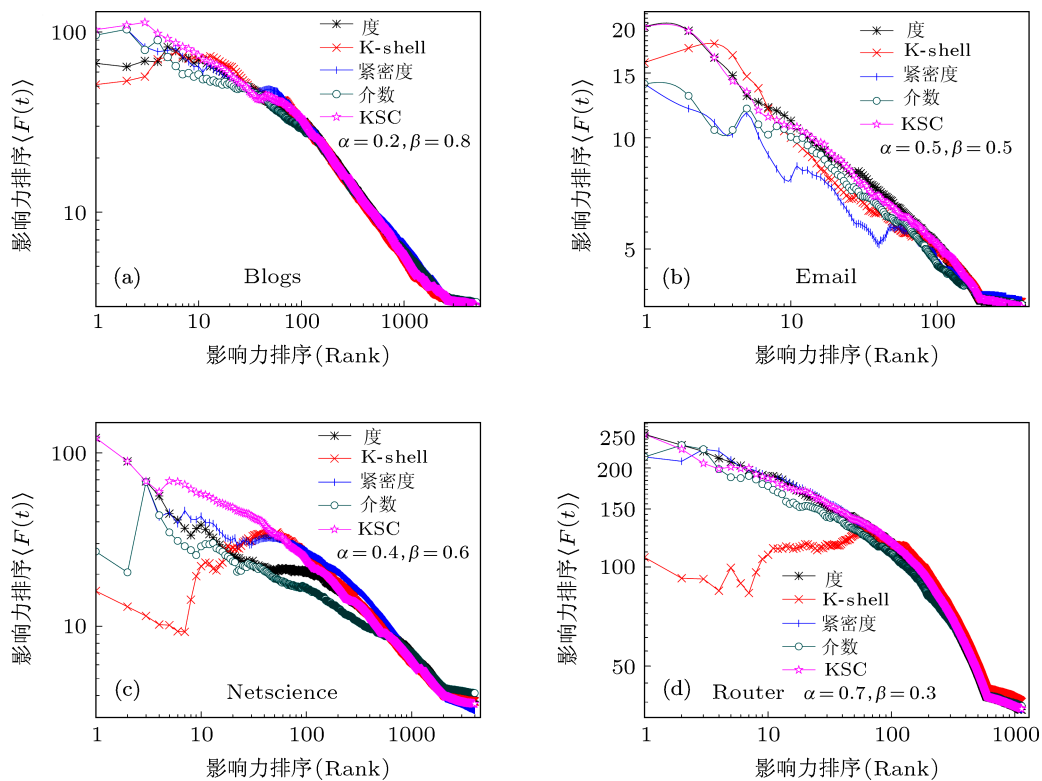


图6 5种指标的影响力 $F(t)$ ($t=10$)前Top-K的均值结果,横轴代表按各种指标值的排序,纵轴表示影响力 $F(t)$,不同颜色的曲线代表不同的方法

5 结论

在复杂网络中发现最具影响力的节点,可以帮助我们提高新知识、新产品的传播效率,同时可以有效地制定相应的策略来阻止疾病及谣言传播.复杂网络中鉴别最具影响力的节点一直以来是研究的热点与难点,我们提出了 KSC 中心化指标模型.实验选取现实生活中常见的四种复杂网络博客网、邮件网、路由网和科学合著网,通过 SIR 模型模拟节点传播过程,并对各种中心化指标得出的影响力

进行排序分析,验证了方法的高效性和正确性.

与传统的度、紧密度、介数和 K-shell 方法相比,传统方法在不同网络中各有优劣,通用性不强.而我们提出的 KSC 模型综合考虑了节点的内部属性与外部属性,实验证明用此方法鉴别节点的影响力要精确得多,适用范围更大.我们所提出的模型为这项具有挑战性研究提供了新的思想和方法.

感谢审稿人对本文提出的宝贵意见和建议,使得文章逻辑更加完整合理.

- [1] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [2] Barabási A L, Albert R 1999 *Science* **286** 509
- [3] Barabási A L, Albert R, Jeong H, Bianconi G 2000 *Science* **287** 2115a
- [4] Pastor-Satorras R, Vespignani A 2002 *Phys. Rev. E* **65** 036104
- [5] Kempe D, Kleinberg J, Tardos E 2003 *Proc. 9th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining* New Washington, DC, USA, August 24–27, 2003 p137
- [6] Gomez-Rodriguez M, Leskovec J, Krause A 2010 *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* Washington, DC, USA, July 25–28, 2010 p1019
- [7] Budak C, Agrawal D, El Abbadi A 2011 *Proceedings of the 20th International Conference on World Wide Web* Hyderabad, India, March 28–April 1, 2011 p665
- [8] Mislove A, Marcon M, Gummadi K P, Druschel P, Bhattacharjee B 2007 *Proceedings of the ACM SIGCOMM 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* Kyoto, Japan, August 27–31, 2007 p29
- [9] Yuan W G, Liu Y, Cheng J J 2013 *Acta Phys. Sin.* **62** 038901 (in Chinese) [苑卫国, 刘云, 程军军 2013 物理学报 **62** 038901]
- [10] Weng J, Lim E P, Jiang J, He Q 2010 *Proceedings of the Third International Conference on Web Search and Web Data Mining, WSDM 2010* New York, NY, USA, February 4–6, 2010 p261
- [11] Xu D, Li X, Wang X F 2007 *Acta Phys. Sin.* **56** 1313 (in Chinese) [许丹, 李翔, 汪小帆 2007 物理学报 **56** 1313]
- [12] Naveed N, Gottron T, Kunegis J, Alhadi A C 2011 *Proceedings of the 20th ACM International Conference on Information and Knowledge Management* Koblenz, Germany, June 14–17, 2011 p1
- [13] Gu Y R, Xia L L 2012 *Acta Phys. Sin.* **61** 238701 (in Chinese) [顾亦然, 夏玲玲 2012 物理学报 **61** 238701]
- [14] Swamynathan G, Wilson C, Boe B, Almeroth K, Zhao B Y 2008 *Proceedings of the First Workshop on Online Social Networks ACM* New York, USA, August 18, 2008 p1
- [15] Mislove A, Marcon M, Krishna P G, Druschel P, Bhattacharjee B 2007 *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement (IMC'07)* ACM New York, USA, October 24–26, 2007 p29
- [16] Albert R, Jeong H, Barabasi A L 2000 *Nature* **406** 378
- [17] Pastor-Satorras R, Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200
- [18] Freeman L C 1977 *Sociometry* **40** 35
- [19] Brin S, Page L 1998 *Comput. Netw. ISDN Syst.* **30** 107
- [20] Opsahl T, Agneessens F, Skvoretz J 2010 *Soc. Networks* **32** 245
- [21] Sabidussi G 1966 *Psychometrika* **31** 581
- [22] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Tech. Rep. Phys. Soc.* **1001** 5285
- [23] Carmi S, Havlin S, Kirkpatrick S, Shavitt Y, Shir E 2007 *Proc. Natl. Acad. Sci. USA* **104** p1150
- [24] Girvan M, Newman M E J 2002 *Proc. Natl. Acad. Sci. USA* **99** 7821
- [25] Fortunato S 2010 *Phys. Rep.* **486** 75
- [26] Newman M E J, Girvan M 2004 *Phys. Rev. E* **69** 026113
- [27] Freeman L C 1979 *Social Networks* **1** 215
- [28] Sabidussi G 1966 *Psychometrika* **31** 581
- [29] Lin K 1970 *Bell Syst. Tech. J.* **49** 291
- [30] Newman M E J 2004 *Phys. J. B* **38** 321
- [31] Newman M E J 2004 *Phys. Rev. E* **69** 066133
- [32] Guimera R, Amaral L A N 2005 *Nature* **433** 7028
- [33] Burt R 2004 *Am. J. Soc.* **110** 349
- [34] Granovetter M 1973 *Am. J. Soc.* **78** 1360
- [35] Roy M Anderson, Robert M May 1992 *Infectious Diseases of Humans: Dynamics and Control.* (New York: Oxford University Press) p66
- [36] Diekmann O, Heesterbeek J A P 2001 *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation* (New York: Wiley Series in Mathematical & Computational Biology)
- [37] Hethcote H W 2000 *Soc. Industr. Appl. Math.* **42** 599
- [38] Bernoulli D, Blower S 2004 *Rev. Med. Virol.* **14** 275
- [39] Chen D B, Lu L Y, Shang M S, Zhang Y C, Zhou T 2012 *Physica A* **391** 1777
- [40] Xie N 2006 *M. S. Dissertation* (Bristol: University of Bristol)
- [41] Newman M E J 2006 *Phys. Rev. E* **74** 36104
- [42] Spring N, Mahajan R, Wetherall D, Anderson T 2004 *Acm. Sigcomm. Comp. Commu. Rev.* **32** 3
- [43] Guimera R, Danon L, Diaz-Guilera A, Giral F, Arenas A 2003 *Phys. Rev. E* **68** 065103

A new approach to identify influential spreaders in complex networks*

Hu Qing-Cheng[†] Yin Yan-Shen Ma Peng-Fei Gao Yang
Zhang Yong Xing Chun-Xiao

(*Research Institute of Information Technology Tsinghua National Laboratory for Information Science and Technology Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*)

(Received 25 January 2013; revised manuscript received 25 March 2013)

Abstract

In the research of the propagation model of complex network, it is of theoretical and practical significance to detect the most influential nodes. Global metrics such as degree centrality, closeness centrality, betweenness centrality and K-shell centrality can be used to identify the influential spreaders. Each of these approaches is simple but has a low accuracy. We propose K-shell and community centrality (KSC) model. This model considers not only the internal properties of nodes but also the external properties, such as the community which these nodes belong to. The susceptible-infected-recovered model is used to evaluate the performance of KSC model. The experimental result shows that our method is better to detect the most influential nodes. This paper comes up with a new idea and method for the study in this field.

Keywords: complex networks, influential node, community detecting, centrality

PACS: 01.75.+m, 05.90.+m, 89.75.-k, 89.70.Hj

DOI: 10.7498/aps.62.140101

* Project supported by National Basic Research Program of China (Grant No. 2011CB302302) and the National High Technology Research and Development Program of China (Grant No. 2012AA09A408).

[†] Corresponding author. E-mail: hqc10@mails.tsinghua.edu.cn