

网络集聚性对节点中心性指标准确性的影响

宋玉萍 倪静

Effect of variable network clustering on the accuracy of node centrality

Song Yu-Ping Ni Jing

引用信息 Citation: *Acta Physica Sinica*, 65, 028901 (2016) DOI: 10.7498/aps.65.028901

在线阅读 View online: <http://dx.doi.org/10.7498/aps.65.028901>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2016/V65/I2>

您可能感兴趣的其他文章

Articles you may be interested in

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.65.028901028902>

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>

负荷作用下相依网络中的级联故障

[Load-induced cascading failure in interdependent network](#)

物理学报.2015, 64(4): 048901 <http://dx.doi.org/10.7498/aps.64.048901>

基于局部路由策略的复杂网络拥塞控制

[Congestion control in complex network based on local routing strategy](#)

物理学报.2014, 63(24): 248901 <http://dx.doi.org/10.7498/aps.63.248901>

基于节点间依赖度的社团结构划分方法

[Partitioning community structure in complex networks based on node dependent degree](#)

物理学报.2014, 63(17): 178901 <http://dx.doi.org/10.7498/aps.63.178901>

网络集聚性对节点中心性指标准确性的影响*

宋玉萍† 倪静

(上海理工大学管理学院, 上海 200093)

(2015年8月6日收到; 2015年10月8日收到修改稿)

节点中心性指标是从特定角度对网络某一方面的结构特点进行刻画的度量指标, 因此网络拓扑结构的改变会对节点中心性指标的准确性产生重要影响. 本文利用 Holme-Kim 模型构建可变集聚系数的无标度网络, 然后采用 Susceptible-Infective-Removal 模型进行传播影响力的仿真实验, 接着分析了节点中心性指标在不同集聚系数的无标度网络中的准确性. 结果表明, 度中心性和介数中心性的准确性在低集聚系数的网络中表现更好, 特征向量中心性则在高集聚网络中更准确, 而紧密度中心性的准确性受网络集聚系数的变化影响较小. 因此当网络的集聚系数较低时, 可选择度或者介数作为中心性指标进行网络节点影响力评价; 反之则选择紧密度指标或特征向量指标较好, 尤其当网络的集聚系数接近 0.6 时特征向量的准确性可以高达 0.85, 是度量小规模网络的较优选择. 另一方面, 传播过程的感染率越高, 度指标和介数指标越可靠, 紧密度和特征向量则相反. 最后 Autonomous System 实证网络的断边重连实验, 进一步验证了网络集聚性的改变会对节点中心性指标的准确性产生重要影响.

关键词: 网络科学, 节点中心性, 网络结构, 集聚性

PACS: 89.75.Fb, 02.10.Ox

DOI: 10.7498/aps.65.028901

1 引言

随着科学技术的迅猛发展, 社会的网络化现象越来越明显, 例如社交网络、电商网络、电力网络、交通网络^[1-3]都与日常生活密不可分, 甚至工业生产中也可以采用复杂网络解决相关的非线性动力学问题^[4]. 而且社会活动网络化的推进, 使网络结构变得越来越复杂, 网络的数据规模也日渐庞大^[5]. 因此如何在复杂结构和大数据背景下, 准确识别网络节点中心性从而判断节点的重要程度, 具有重大科学意义^[6,7].

网络的中心节点是指不仅能影响网络的结构, 如度分布、平均距离、连通性、集聚性、度相关性等, 同时也能够影响网络的抗毁性、传播、同步、控制等功能的一些特殊节点^[8]. 目前, 复杂网络中用来度量节点中心性的指标众多^[9-11]. 经典的中心性指标包括计算网络中节点的邻居个数的度中心性 (degree centrality), 度量网络中的节点对其

他节点施加影响的能力的紧密度中心性 (closeness centrality)^[12], 衡量个体社会地位的介数中心性 (betweenness centrality)^[13], 将单个节点的影响力看成是所有其他节点影响力的线性组合的特征向量中心性 (eigenvector centrality)^[14].

近年来, 研究者们为提高经典中心性指标的准确性并降低其计算复杂度, 对这些中心性方法进行了改进. 例如提高度、紧密度及介数等指标的准确性^[15-20], 降低特征向量的计算复杂度^[21], 以及探索新的度量指标如 k 核等^[22-27]. 这些新方法都是为了使复杂网络中节点的中心性指标更准确, 然而节点中心性指标都是从特定角度对于网络的某一方面的结构特点进行刻画, 如果目标网络的结构在该方面特征显著, 对应指标即可得到较好的效果^[28]. 现实世界的网络统计分析表明, 许多现实网络具有高集聚性^[29]. 例如生活社交网络中, 朋友之间往往都相互认识, 这样就成三角形结构. 三角形结构在实际社会网中代表了一种完全透明的社会关系: 你, 我和他三个人之间完全互相认识, 互相

* 上海市自然科学基金 (批准号: 14ZR1427800) 资助的课题.

† 通信作者. E-mail: violet910516@163.com

了解. 因而网络的集聚性高低通过网络中三角形的数量进行判断.

研究表明, 网络集聚性(即网络中三角形的数量)不仅会影响网络的传播、同步、控制等功能, 也对网络节点中心性准确性产生影响. 如Klemm等[30]研究发现集群动力学中节点的中心性是由网络拓扑结构和集群动力学机理共同决定的. Aral等[6]研究网络上的用户传播行为时, 发现中心节点具有高集聚特性. Centola[31]发现传播行为在高集聚类网络中速度更快. Bond等[32]以2010年美国大选为实例研究时发现Facebook用户的中心性与网络结构和传播机理两者都相关. Gao等[33,34]在研究非线性动力学行为时, 发现网络的拓扑结构会影响网络节点的动力学行为, 尤其是网络的集聚系数能有效刻画网络节点动力学行为的变化.

为了在特定集聚系数网络中探索更合适的中心性指标, 本文以网络集聚性对中心性指标的准确性影响研究为目标. 首先采用Holme-Kim (HK)模型[35]构建可变集聚系数的无标度网络, 从而得到不同的集聚性的网络; 然后采用Susceptible-Infective-Removal (SIR)传播模型[36]进行仿真得到节点影响力序列; 再用Kendall's Tau系数[37]来验证不同经典中心性指标与节点影响力序列的一致性, 分析网络结构对中心性指标的准确性影响; 并通过Autonomous System (AS)网络[38]的断边重连实验对结果进行验证; 最后总结当前面临的问题并展望可能的发展方向.

2 模型回顾

2.1 集聚系数的定义

假设无向网络是由 N 个节点和 M 条边所组成的. 其邻接矩阵表示为 $\mathbf{A} = (a_{ij})_{N \times N}$, 那么包含节点 i 的三角形数目为

$$E_i = \frac{1}{2} \sum_{j,k} a_{ij} a_{jk} a_{ik}. \quad (1)$$

当且仅当节点 i, j, k 连边构成三角形时, $a_{ij} a_{jk} a_{ik} = 1$, 否则 $a_{ij} a_{jk} a_{ik} = 0$. 因此网络的集聚系数[28]表示为

$$C = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{j,k} a_{ij} a_{jk} a_{ik}}{\sum_{j,k} a_{ij} a_{jk}}, \quad (2)$$

其中节点 i, j, k 两两互不相同. 显然集聚系数的取值范围为 $0 \leq C \leq 1$. 当 $C = 0$ 时, 表示网络中没有三角形结构; 当 $C = 1$ 时, 表示网络中任意三个节

点都能构成三角形结构.

2.2 可变集聚系数的无标度网络构建

可变集聚系数的无标度网络是由Holme和Kim提出的, 简称HK模型[35]. HK模型融合了小世界网络的高集聚特性[29,39]和BA网络的无标度[18,40]特点. 为了增加无标度网络的集聚系数, Holme和Kim在其基础上引入一个被称为三角形形成的新步骤. 具体来说就是: 首先初始化 m_0 个节点, 并让初始节点构成一个环, 以后每一步加入一个带有 m 条边的新节点 i . 新节点 i 先从已存在的节点 V 集中随机选择一个节点 j , 以概率 $P_A = k_j / (\sum_{v \in V} k_v)$ 进行优先连接, 其中 k_v 表示 V 中任意一个节点 v 的度. 如果新节点 i 与节点 j 成功连接, 接着从节点 j 的邻居中随机选择一个节点 r_j , 以概率 $P_t = M_t / (m - 1)$ 做三角形形成连接, 其中 M_t 为可变参数. 本文取 $m_0 = 4$, $m = 3$, $N = 10000$, $M_t \in [0, 2]$ 以0.1递增. 如图1(b)所示,

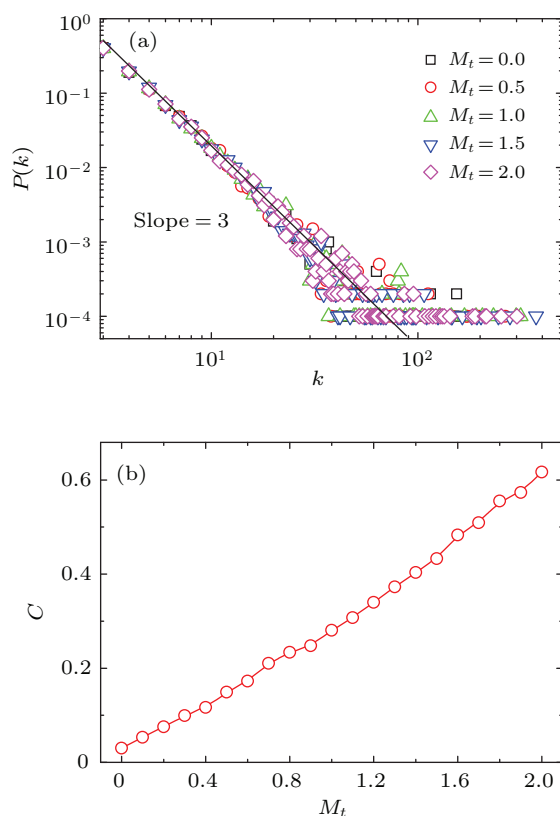


图1 可变集聚系数的无标度网络的特性 (a) 幂律分布图; (b) 随参数 M_t 变化的集聚系数

Fig. 1. The characteristics of scale-free networks with the tunable cluster: (a) Power-law degree distributions; (b) clustering coefficient with different parameters M_t .

随着 P_t (即 M_t) 的增大, 网络的集聚系数也会增大, $P_t = 0$ (即 $M_t = 0$) 对应于无标度网络.

本文构建得到的可变网络, 一方面具有经典无标度网络的幂律分布特性, 如图 1(a), 随着可变参数 M_t 的变化, 均保持斜率为 3 的幂律分布现象^[35,40]; 另一方面又具有小世界网络的高集聚系数特性^[28], 如图 1(b), 集聚系数随着可变参数 M_t 的增大而增大.

2.3 SIR 传播模型

SIR 传播模型是复杂网络中节点传播行为的经典模型^[36,41]. 由 N 个节点组成的复杂网络中, 每个节点可能处于三种状态 (易感, 感染, 免疫移出) 中的其中一种. 假定节点 j 遇到了感染者 i , 如果节点 j 是易感者, 那么它就以概率 β 被感染成为新的感染者; 如果节点 j 是感染者, 那么它经过 T_r 的时间以概率 δ 成为免疫移出者; 如果节点 j 已经是免疫移出者, 那么 j 遇到 i 之后不会再被感染. 假设网络中节点 i 为传染源, 将节点 i 的传播值定义为每次 SIR 仿真结束时, 网络中状态为感染或免疫移出的节点数, 独立重复仿真 T 次, 则节点 i 传播影响力可以表示为

$$\bar{X}_i = \frac{1}{T} \sum_{t=1}^T X_i(t), \quad (3)$$

其中 $X_i(t)$ 为节点 i 在第 t 次仿真时的传播值, $\bar{X}_i$ 为节点 i 在 T 次仿真中的平均传播值. 节点的传播值越大表明节点的传播影响力越大, 节点也就越重要. 于是对网络中的每个节点进行 SIR 传播仿真, 最终可以得到网络中节点的传播影响力序列 $(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N)$.

3 实验和仿真结果

3.1 HK 模型仿真结果分析

本文对构建的可变集聚系数的无标度网络, 采用 SIR 模型仿真得到网络中每个节点的传播影响力, 其中 SIR 模型参数 $\delta = 1$, $T_r = 1$, $T = 200$, $\beta \in [0, 0.3]$ 以 0.02 递增. 在可变参数 $M_t(t = \alpha)$ 对应的可变网络中, 采用传播率 β 进行仿真, 可以得到对应参数网络的节点传播影响力序列 $\text{Rank}_{(M_\alpha, \beta)}(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N)$. 接着根据经典的中心性指标, 依次对 $M_t(t = \alpha)$ 的网络计算出其度中心性指标序列 $\text{Degree}_{(M_\alpha)}(k_1, k_2, \dots, k_N)$, 介数中心性指标序列 $\text{Betweenness}_{(M_\alpha)}(b_1, b_2, \dots, b_N)$, 紧密度中心性指标

序列 $\text{Closeness}_{(M_\alpha)}(c_1, c_2, \dots, c_N)$ 和特征向量中心性指标序列 $\text{Eigenvector}_{(M_\alpha)}(e_1, e_2, \dots, e_N)$.

本文采用 Kendall's Tau 系数^[31] 衡量网络节点的度、紧密度、介数、特征向量与节点在 SIR 模型中的传播影响力之间的相关性, 即在可变参数 $M_t(t = \alpha)$ 对应的可变网络中, 将序列 $\text{Rank}_{(M_\alpha, \beta)}(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N)$ 分别与中心性指标序列 $\text{Degree}_{(M_\alpha)}(k_1, k_2, \dots, k_N)$, $\text{Betweenness}_{(M_\alpha)}(b_1, b_2, \dots, b_N)$, $\text{Closeness}_{(M_\alpha)}(c_1, c_2, \dots, c_N)$, $\text{Eigenvector}_{(M_\alpha)}(e_1, e_2, \dots, e_N)$ 做相关性分析. 相关性越强表明该指标越能准确度量节点的传播影响力. Kendall's Tau 系数用来描述两个序列之间相关性, 可以表示为

$$\tau = \frac{\sum_{i=1}^N \sum_{j=1}^N \text{sgn}[(x_i - x_j)(y_i - y_j)]}{N(N-1)}, \quad (4)$$

其中序列 X 和序列 Y 的元素个数都为 N , 并且 $x_i \in X$, $y_i \in Y$. 这里 $\text{sgn}(x)$ 为符号函数, 当 $x > 0$ 时其返回值为 1; 当 $x < 0$ 时其返回值为 -1, 当 $x = 0$ 时其返回值为 0. Kendall 系数 τ 的值越大, 表示两个序列之间的一致性越高, τ 最大取值为 1, 表示两个序列顺序完全一致.

根据 Kendall's Tau 系数检验. 我们发现不同集聚系数下中心性指标的准确性不同 (图 2), 不同感染率下中心性指标的准确性也有所不同 (图 3).

实验结果表明, 不同集聚系数的网络对不同中心性指标的准确性有着不同的影响. 根据图 2 的结果显示, 度指标和介数指标的准确性会随着网络集聚系数的增大而降低, 但度指标的变化较介数指标不显著; 特征向量指标则会随着集聚系数的增大提高准确性; 而紧密度指标的准确性受集聚系数变化的影响不大.

对衡量节点的邻居数目的度指标而言, 随着网络集聚系数的增大其准确性逐渐降低, 并且感染率越大, 下降趋势越明显; 网络的紧密度准确性受集聚系数的变化影响不大, 大概围绕在 0.78 左右, 但当疾病感染率小于 0.12 时, 紧密度的准确性会随着集聚系数的增大而降低; 介数指标的准确性随集聚系数的变化影响较为显著, 从集聚系数 $C \approx 0$ 增长到 $C \approx 0.6$ 的过程中, 介数的准确性从 0.6—0.7 降低至 0.3; 网络的集聚系数逐渐增大时, 特征向量指标则表现越来越准确, 当感染率较低的时候, 准确性提高不明显, 但是当感染率较高的时候, 准确性的提高就非常显著了.

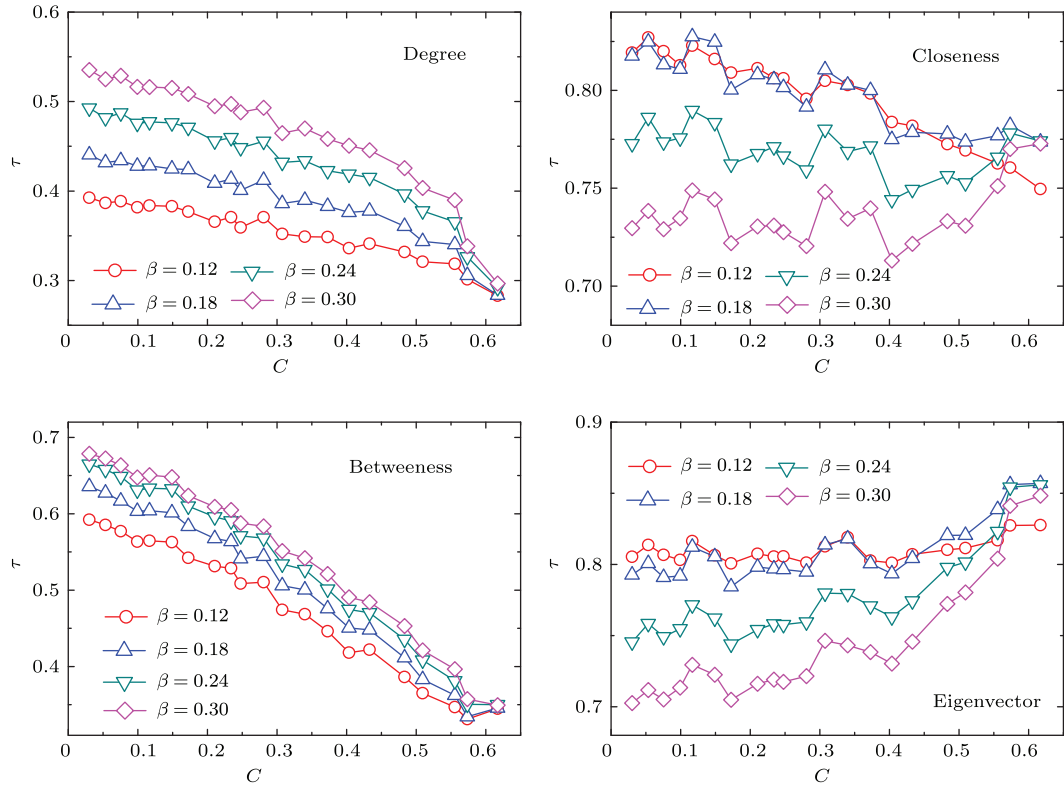


图2 不同集聚系数下中心性指标的准确性变化

Fig. 2. Variation of centrality measures accuracy with different clustering coefficient.

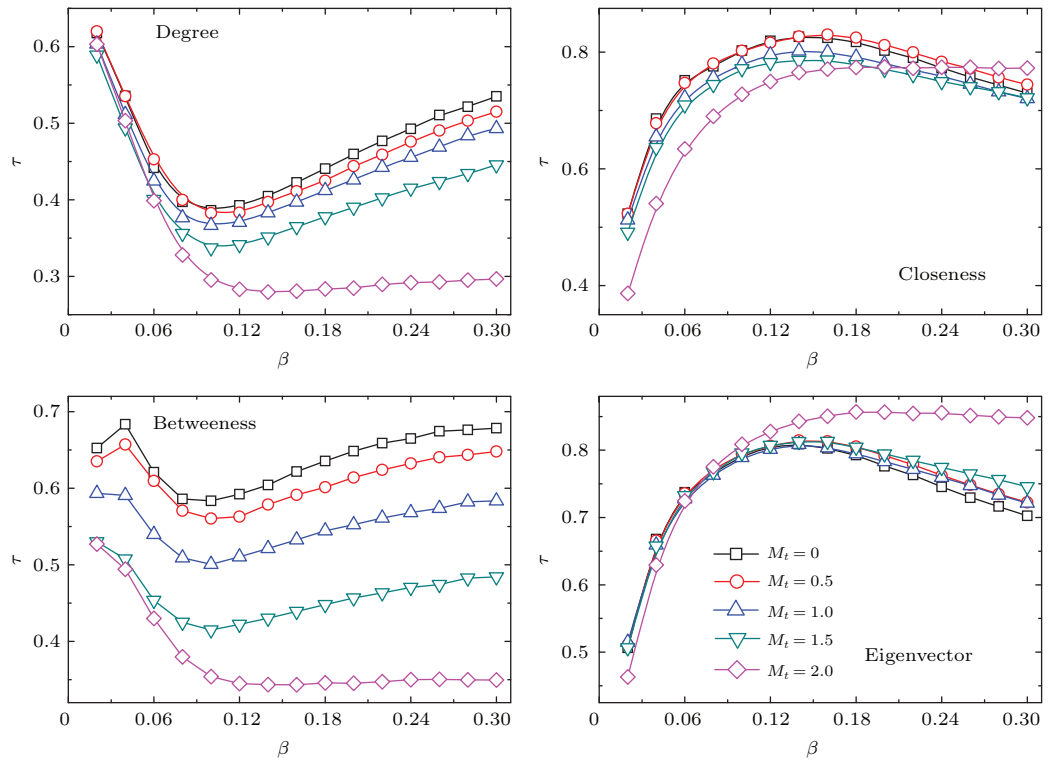


图3 不同感染率下中心性指标准确性变化

Fig. 3. Variation of centrality measures accuracy with different infectious rates.

图3的结果显示了不同的疾病感染率对中心性指标的准确性有不同的影响: 度指标会随着感染率的增大呈现出准确性先降后升的现象, 在降落过程中, 不同的集聚系数表现一致; 但上升过程中, 网络的集聚性越高上升越慢, 且准确性越差, $\beta = 0.1$ 时准确性最差; 紧密度指标和特征向量指标则正好相反, 随着感染率的增大指标准确性表现为先上升后下降, 集聚系数越高, 下降的可能性就越低, 且准确性越好, 均为 $\beta = 0.14$ 时最准确; 介数指标在感染率变化的过程中准确性的变化比较复杂, 当感染率 $\beta = 0.04$ 时, 介数指标的影响力评价准确性表现最好, 当 $\beta = 0.1$ 时表现最差, 另外, 随着感染率的变化, 介数指标在不同集聚系数的网络对应的指标准确性有明显的区分, 低集聚系数的网络准确性表现良好, 高集聚系数的网络在 $\beta > 0.1$ 时准确性表现非常不理想。

3.2 AS网络实证分析

为了进一步验证网络集聚性对中心性指标准确性的影响, 本文引入了AS网络做实证分析. 该网络是CAIDA 2005年的路由级拓扑测量数据获得的无向网络^[35], 网络规模为23752个节点, 58416条边。

本文通过对AS网络进行断边重连操作的方式改变网络的集聚性. 一次断边重连操作就是在原网

络中随机找寻2条边——如边ab(节点a与节点b相连)和边cd(节点c与节点d相连), 将原边去掉, 改变节点的连接获得新边——即边ad(节点a与节点d相连)和边bc(节点b与节点c相连). 如此进行 T_{rec} 次得到新的网络, 如图4所示, AS网络的初始集聚系数为0.60, 随着断边重连次数的增加, 集聚系数呈现出下降的趋势: 当 $T_{\text{rec}} = 2000$ 时, 集聚系数为0.53; 当 $T_{\text{rec}} = 20000$ 时, 集聚系数减小至0.31. 对AS网络进行 T_{rec} 次断边重连后的网络实施SIR传播仿真实验, 得到节点的传播影响力序列, 并分别与对应网络的中心性指标序列进行Kendall's Tau系数检验。

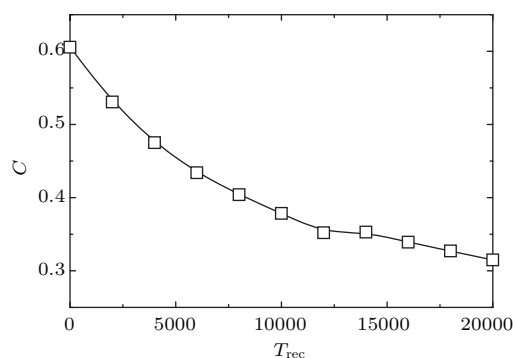


图4 AS网络断边重连次数 T_{rec} 与网络集聚系数 C 的关系

Fig. 4. The network cluster coefficient (C) after running T_{rec} times reconnection in the AS networks.

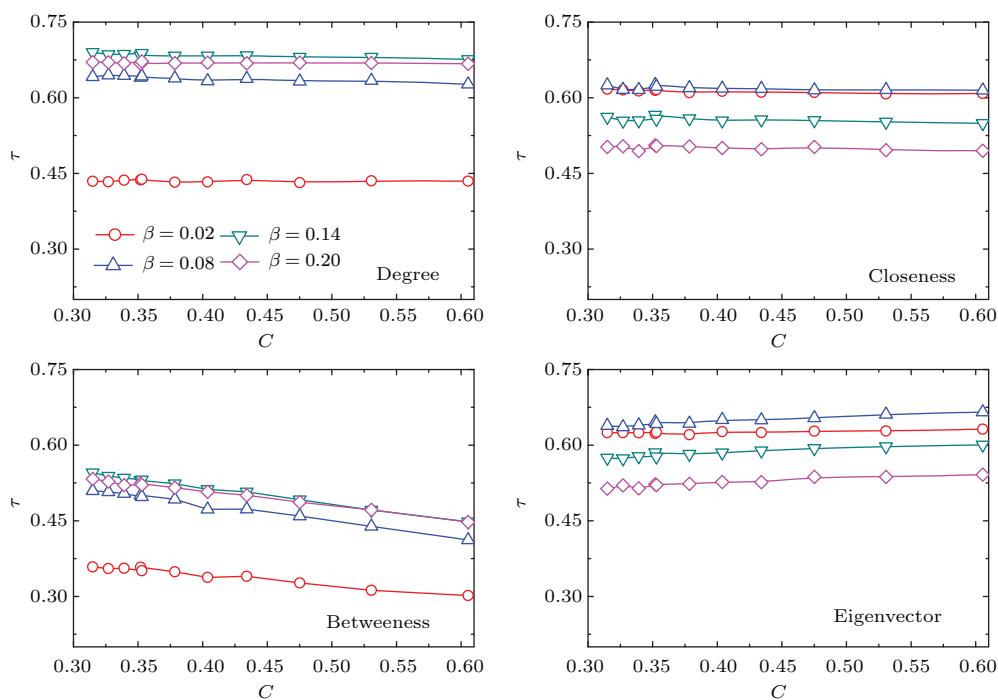


图5 AS网络断边重连后不同集聚系数下中心性指标准确性变化

Fig. 5. Variation of centrality measures accuracy with different clustering coefficient of re-connected AS networks.

根据图 5 的结果可以发现, AS 网络断边重连后得到的新网络中, 集聚系数的变化对中心性指标准确性的影响与可变网络中的影响类似: 在 0.3—0.6 的集聚系数范围内, 度指标准确性随着集聚系数的增大下降现象不显著, 感染率越高下降趋势越明显, 感染率越低度指标的准确性越差; 紧密度指标的准确性随着集聚系数的变化有轻微的抖动, 在 $C = 0.35$ 左右出现一个小峰值, 当网络的感染率在 0.08 左右时, 紧密度指标的准确性最高, 较高的感染率会降低指标的准确度; 介数指标准确性随着集聚系数的增大呈明显的下降现象, 当网络的集聚系数从 $C = 0.3$ 增加到 $C = 0.6$, 介数指标所对应于 SIR 模型得到的排序结果的 Kendall 相似值差将近 0.3, 也就是介数指标的准确性受网络的集聚系数变化的影响非常显著; 而特征向量指标则随着网络集聚系数的增大 Kendall 系数的值呈上升现象, 这也意味着网络的集聚性越高, 特征向量的准确性越高, 并且该网络在感染率为 0.08 左右时, 特征向量的准确性最高至 0.67. 因此, AS 实证网络进一步表明网络集聚性影响度、紧密度、介数及特征向量中心性指标的准确性.

4 结 论

文章根据 HK 模型构造的不同集聚系数的网络, 经过 SIR 仿真实验进行网络中节点的影响力评价, 采用 Kendall's Tau 系数讨论了不同网络集聚性对中心性指标的准确性影响, 并用 AS 网络进行实证分析. 分析表明, 基于 SIR 传播仿真得到的节点影响力评价体系中, 不同的网络集聚性对不同的中心性指标的准确性有着不同的影响.

因而, 当网络的集聚系数较低时, 可以根据需要选择度或者介数作为中心性指标进行网络节点影响力评价; 当网络的集聚系数较大时, 则要尽可能放弃度指标较低的时间计算复杂度优势, 转而选择紧密度指标或者特征向量指标. 就传播过程而言, 紧密度和特征向量的准确性在任何集聚系数的网络中都达到 0.7 以上, 尤其是特征向量在网络的集聚系数接近 0.6 时准确性可以高达 0.85 左右, 是小规模网络的较优选择. 另一方面, 传播过程的感染率越高, 度指标和介数指标越可靠, 紧密度和特征向量则相反.

文章通过改变网络的集聚系数来改变网络拓扑结构, 从而探索中心性指标的准确性. 然而文中探索的四种指标不能代表使用的所有指标, 其他指

标的准确性又会随着结构的变化发生何种变化; 而且网络结构不仅仅可以通过静态网络中的集聚性来显示, 也可以通过连通性、平均距离以及度相关性来反映, 在不同特性下, 中心性指标的准确性是否会有所不同; 在网络结构时刻改变的时变网络中^[42], 中心性指标的准确性又会发生什么样的变化; HK 模型构造的是无向网络, 如果是有向网络的结构变化对节点中心性指标的准确性又会如何影响; 传播模型的感染率也会影响准确性评价, 那特定感染率下如何选择更合适的中心性指标. 这些问题都有待进一步研究和解决.

参考文献

- [1] Albert R, Barabási A L 2002 *Rev. Mod. Phys.* **74** 47
- [2] Newman M E J 2003 *SIAM. Rev.* **45** 167
- [3] Lü L, Medo M, Yeung C H, Zhang Y C, Zhang Z K, Zhou T 2012 *Phys. Rep.* **519** 1
- [4] Gao Z K, Zhang X W, Jin N D, Norbert M, Jvrgen K 2013 *Phys. Rev. E* **88** 032910
- [5] Rong Z H, Tang M, Wang X F, Wu Z X, Yan G, Zhou T 2012 *Journal of Electronic Science and Technology* **34** 801 (in Chinese) [荣智海, 唐明, 汪小帆, 吴枝喜, 严钢, 周涛 2012 电子科技大学学报 **34** 801]
- [6] Aral S, Walker D 2012 *Science* **6092** 337
- [7] Zhao J, Yu L, Li J R, Zhou P 2015 *Chin. Phys. B* **24** 058904
- [8] Newman M E J 2010 *Networks An Introduction* (New York: Oxford University Press) p168
- [9] 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]
- [10] Ren X L, Lü L Y 2014 *Sci. Bull.* **13** 4 (in Chinese) [任晓龙, 吕琳媛 2014 科学通报 **13** 4]
- [11] Song B, Jiang G P, Song Y R, Xia L L 2015 *Chin. Phys. B* **24** 100101
- [12] Sabidussi G 1966 *Psychometrika* **31** 581
- [13] Goh K I, Oh E, Kahng B, Kim D 2003 *Phys. Rev. E* **67** 017101
- [14] Borgatti S P 2005 *Soc. Networks* **27** 55
- [15] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E, Makse H A 2010 *Nat. Phys.* **6** 888
- [16] Ren Z M, Shao F, Liu J G, Guo Q, Wang B H 2013 *Acta Phys. Sin.* **62** 128901 (in Chinese) [任卓明, 邵凤, 刘建国, 郭强, 汪秉宏 2013 物理学报 **62** 128901]
- [17] Chen D, Lv L, Shang M S, Zhang Y C, Zhou T 2012 *Physica A* **391** 1777
- [18] Wang J R, Wang J P, He Z, Xu H T 2015 *Chin. Phys. B* **24** 060101
- [19] Zhang J, Xu X K, Li P, Zhang K, Small M 2011 *Chaos* **21** 016107
- [20] Comin C H, Costa Lda F 2011 *Phys. Rev. E* **84** 056105
- [21] Poulin R, Boily M C, Mâsse B R 2000 *Soc. Networks* **22** 187

- [22] 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]
- [23] Garas A, Schweitzer F, Havlin S 2012 *New J. Phys.* **14** 083030
- [24] Zeng A, Zhang C J 2013 *Phys. Lett. A* **377** 1031
- [25] Travençolo B A N, Costa Lda F 2008 *Phys. Lett. A* **373** 89
- [26] Chen D B, Xiao R, Zeng A, Zhang Y C 2013 *EPL* **104** 68006
- [27] Lv L, Zhang Y C, Yeung C H, Zhou T 2011 *PloS one* **6** e21202
- [28] Ren Z M, Zeng A, Chen D B, Liao H, Liu J G 2014 *EPL* **106** 48005
- [29] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [30] Klemm K, Serrano M Á, Eguíluz V M, San Miguel M 2012 *Sci. Rep.* **2** 292
- [31] Centola D 2010 *Science* **329** 1194
- [32] Bond R M, Fariss C J, Jones J J, Kramer A D, Marlow C, Settle J E, Fowler J H 2012 *Nature* **489** 295
- [33] Gao Z K, Yang Y X, Fang P C, Jin N D, Xia C Y, Hu L D 2015 *Sci. Rep.* **5** 8222
- [34] Gao Z K, Fang P C, Ding M S, Jin N D 2015 *Experimental Thermal & Fluid Science* **60** 157
- [35] Holme P, Kim B J 2002 *Phys. Rev. E* **65** 026107
- [36] Pastor S R, Vázquez A, Vespignani A 2001 *Phys. Rev. Lett.* **87** 258701
- [37] Kendall M G 1938 *Biometrika* **30** 81
- [38] Papadopoulos F, Kitsak M, Serrano M Á, Boguñá M, Krioukov D 2012 *Nature* **489** 537
- [39] Zhang Z Z, Xu W J, Zeng S Y 2014 *Chin. Phys. B* **23** 088902
- [40] Barabási A L, Albert R 1999 *Science* **286** 509
- [41] Lu Y L, Jiang G P, Song Y R 2012 *Chin. Phys. B* **21** 100207
- [42] Holme P, Saramäki J 2012 *Phys. Rep.* **519** 97

Effect of variable network clustering on the accuracy of node centrality^{*}

Song Yu-Ping[†] Ni Jing

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

(Received 6 August 2015; revised manuscript received 8 October 2015)

Abstract

Measurements of node centrality are based on characterizing the network topology structure in a certain perspective. Changing the network topology structure would affect the accuracy of the measurements. In this paper, we employ the Holme-Kim model to construct scale-free networks with tunable clustering, and consider the four measurements of classical centrality, including degree centrality, closeness centrality, betweenness centrality and the eigenvector centrality. For comparing the accuracy of the four centrality measurements, we simulate the susceptible-infected-recovered spreading of the tunable clustering scale free networks. Experimental results show that the degree centrality and the betweenness centrality are more accurate in networks with lower clustering, while the eigenvector centrality performs well in high clustering networks, and the accuracy of the closeness centrality keeps stable in networks with variable clustering. In addition, the accuracy of the degree centrality and the betweenness centrality are more reliable in the spreading process at the high infectious rates than that of the eigenvector centrality and the closeness centrality. Furthermore, we also use the reconnected autonomous system networks to validate the performance of the four classical centrality measurements with varying cluster. Results show that the accuracy of the degree centrality declines slowly when the clustering of real reconnected networks increases from 0.3 to 0.6, and the accuracy of the closeness centrality has a tiny fluctuation when the clustering of real reconnected networks varies. The betweenness centrality is more accurate in networks with lower clustering, while the eigenvector centrality performs well in high clustering networks, which is the same as in the tunable clustering scale free networks. According to the spreading experiments in the artificial and real networks, we conclude that the network clustering structure affects the accuracy of the node centrality, and suggest that when evaluating the node influence, we can choose the degree centrality in the low clustering networks, while the eigenvector centrality and the closeness centrality are still in the high clustering networks. When considering the spreading dynamics, the accuracy of the eigenvector centrality and the closeness centrality is high, but the accuracy of the degree centrality and the betweenness centrality is more reliable in the spreading process at high infectious rates. This work would be helpful for deeply understanding of the node centrality measurements in complex networks.

Keywords: network science, node centrality, network structure, clustering coefficient

PACS: 89.75.Fb, 02.10.Ox

DOI: 10.7498/aps.65.028901

^{*} Project supported by the Natural Science Foundation of Shanghai, China (Grant No. 14ZR1427800).

[†] Corresponding author. E-mail: violet910516@163.com