

基于改进节点收缩法的复杂网络节点重要性评估

蔡晓楠, 郑中团

(上海工程技术大学 数理与统计学院, 上海 201620)

摘要: 识别网络中的重要节点对研究网络的拓扑结构及功能特性具有重要的实际应用价值。为了更好地挖掘复杂网络中的重要节点, 考虑到节点核心位置和网络拓扑结构变化对节点重要性评估的影响, 基于节点收缩法提出两类改进的复杂网络重要节点评估方法, 并进行仿真实验与比较分析。一方面, k -shell 值对节点位置可进行粗粒化评估, 将节点的 k -shell 值与网络中所有 k -shell 值之和的比作为节点收缩法得到的节点重要度 (IMC) 系数, 提出基于新的节点重要度 K -IMC 的改进算法; 另一方面, 利用网络结构熵刻画网络拓扑结构的变化, 结合收缩前后网络标准结构熵的变化, 提出基于又一新的节点重要度 E -IMC 的改进算法。在此基础上, 对两类改进的重要节点评估算法进行仿真实验, 运用 SIR 模型和鲁棒性测试对算法性能进行评价与分析。实验结果表明, 相较于原始节点收缩法, K -IMC 算法和 E -IMC 算法的重要节点排序结果均表现出更高的准确性。其中, E -IMC 算法相较 K -IMC 算法的准确性更高, K -IMC 算法相较 E -IMC 算法的运算效率更高。

关键词: 复杂网络; 节点收缩法; k -shell; 标准结构熵

DOI: 10.11907/rjdk.222201

中图分类号: O157.5

文献标识码: A

开放科学 (资源服务) 标识码 (OSID):

文章编号: 1672-7800(2024)002-0066-08



Evaluation of Node Importance in Complex Networks Based on Improved Node Contraction Method

CAI Xiaonan, ZHENG Zhongtuan

(School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: Identifying the important nodes in the network has important practical value for studying the topology and functional characteristics of the network. In order to better mine the important nodes in complex networks, considering the impact of the changes of node core location and network topology on the evaluation of node importance, two improved evaluation methods of important nodes in complex networks are proposed based on the node shrinkage method, and simulation experiments and comparative analysis are carried out. On the hand, combined with the characteristic that k -shell value can evaluate the coarse-grained location of nodes, the ratio of k -shell value of nodes to the sum of all k -shell values in the network is taken as the coefficient of node importance (IMC) obtained by node shrinkage method, and an improved algorithm based on the new node importance K -IMC is proposed; On the other hand, the change of network topology is described by network structure entropy. Combined with the change of network standard structure entropy before and after shrinkage, an improved algorithm based on another new node importance E -IMC is proposed. On this basis, simulation experiments are carried out on these two improved important node evaluation algorithms, and the performance of the algorithm is evaluated and analyzed by using SIR model and robustness test. The experimental results show that K -IMC algorithm and E -IMC algorithm show better accuracy in sorting important nodes compared with the original node shrinking method. In terms of accuracy, E -IMC algorithm is higher than K -IMC algorithm, and in terms of operational efficiency, K -IMC algorithm is better than E -IMC algorithm.

Key Words: complex network; node contraction method; k -shell; standard structure entropy

收稿日期: 2022-10-14

基金项目: 全国统计科学研究项目 (2018LY16, 2020LY080)

作者简介: 蔡晓楠 (1995-), 女, 上海工程技术大学数理与统计学院硕士研究生, 研究方向为复杂网络分析、应用随机过程; 郑中团 (1979-), 男, 博士, 上海工程技术大学数理与统计学院副教授, 研究方向为复杂网络与随机过程、机器学习与数据挖掘、区域可持续发展统计测度、应用统计与综合评价、风险模型与破产概率。本文通讯作者: 郑中团。

0 引言

从自然界到人类社会、从生态关系到社会关系都存在大量复杂系统,这些复杂系统可以用复杂网络表示。复杂系统中的联系与事物被抽象成复杂网络的边和点。复杂网络往往具有大量节点,节点与节点之间有着复杂的联系,往往各个节点本身都是非线性系统,且许多复杂网络常呈现出小世界性、无标度性与超家族性^[1-2]。复杂网络重要节点评估与识别是网络科学中的一个热点问题。重要节点是相比于网络中其他节点而言能够在更大程度上影响网络结构与功能的一些特殊节点^[3]。发现并找到网络中的重要节点,有助于提升网络的鲁棒性和抗毁性,并对网络上的动力学行为进行控制和预测^[4]。例如在电力网络中,哪些节点被攻击可以使整个电力网络更快瘫痪;在运输网络中,哪些节点更容易拥堵,从而造成出行不便的情况;在疫情传播网络中,哪些节点一旦被传染后疫情会快速扩散;在信息传播网络中,哪些节点的消息可以快速传播到其他节点。这些问题的核心就是识别复杂网络中的重要节点。因此,对节点进行重要性排序、找出网络中的重要节点具有较高的现实应用价值。

复杂网络重要节点评估方法研究方兴未艾^[5]。有些方法认为节点的度越大,节点越重要,如度中心性、半局部中心性方法等;有些方法认为位于几个链接区域的“桥点”很重要,如离心中心性、接近中心性、介数中心性方法等;有些方法基于随机游走的思想,认为一个节点被反复经过的概率越大则越重要,如特征向量中心性方法、PageRank算法、HITS算法等^[6-7];还有方法通过网络的连通性刻画节点的重要程度,如节点删除法、节点收缩法等^[8]。谭跃进等^[9]提出的节点收缩法改善了删除节点后带来的网络不连通性,并同时考虑到节点的度和节点位置两个影响因素;朱涛等^[10]综合考虑边权对节点重要度评估的影响,提出改进的适用于加权网络的节点收缩方法;王甲生等^[11]通过调节节点的重要度表示为节点自身重要度与其连边重要度的加权比例系数,分析节点连边重要度对节点重要度排序及节点相对重要性的影响;Kitsak等^[12]利用k-shell分解得到节点重要性排序指标,基于节点位置评估节点重要性,认为位于网络越中心处的节点重要性程度越高。k-shell算法无法精确地比较具有相同壳值节点的相对重要性,王凯莉等^[13]利用向量的形式来表示网络中节点自身壳值与其多阶邻居的壳值,提出多阶邻居壳数向量中心性方法。熵是系统的一种无序度量,谭跃进等^[14]利用标准网络结构熵的概念,阐述了网络结构熵与连接度分布的关系;胡钢等^[15]利用网络的结构熵来研究不同网络拓扑结构的演化规律。本文将融合节点位置和网络结构熵提出新的节点重要性度量,由此对节点收缩法进行改进。

综上,复杂网络重要节点评估方法已得到广泛、深入

的研究,在这些研究的基础上,本文将考虑节点核心位置和网络拓扑结构变化对节点重要性评估的影响,提出两类改进的节点收缩法,以期更好地挖掘复杂网络中的重要节点。具体地,在节点未收缩之前,通过引入k-shell值从全局属性对节点收缩法进行粗粒化的改进,或通过节点收缩后网络标准结构熵的变化对节点排序进行改进。在此基础上,对基于改进节点收缩法的重要节点评估算法进行仿真实验,运用SIR模型和鲁棒性测试对算法性能进行评价分析。

1 节点收缩法简介

复杂网络模型用图 $G=(V,E)$ 表示,其中 G 是一个无向无权的复杂网络,有 n 个节点和 m 条边,其中 $V=\{v_1, v_2, v_3, \dots, v_n\}$ 代表节点集合, $E=\{e_1, e_2, e_3, \dots, e_m\}$ 代表边的集合。节点收缩法^[11]是指将节点 v_i 及其相连的 k_i 个节点都与节点 v_i 相融合,用一个新节点代替这 k_i+1 个节点,将其他节点与 k_i+1 个节点相连的边变为与新节点相连。通过对比节点收缩后整个网络凝聚程度的变化,从而衡量一个节点的重要程度。

网络凝聚程度受两方面因素影响:一是节点之间的平均路径长度 L ,即网络中所有节点对之间最短距离的平均值;二是网络节点数目 n ^[9]。故网络凝聚度可以表示为节点数 n 与平均路径长度 L 乘积的倒数:

$$\partial_{(c)} = \frac{1}{n \cdot L} = \frac{n-1}{\sum_{i \neq j \in V} d_{ij}} \quad (1)$$

其中, n 大于2, d_{ij} 代表节点 v_i 与节点 v_j 之间的最短距离。

基于网络凝聚度得到的节点重要度(IMC)计算公式为^[9]:

$$IMC(v_i) = 1 - \frac{\partial(G)}{\partial(G^*v_i)} = \frac{n \cdot L(G) - (n - k_i) \cdot L(G^*v_i)}{n \cdot L(G)} \quad (2)$$

由式(2)可进一步看出,一个节点的邻接度越大,该节点的位置越重要,即越多的最短路径经过该节点,该节点越重要。 G 表示原始的复杂网络, G^*v_i 表示节点 v_i 收缩后的复杂网络, $L(G)$ 表示复杂网络 G 的平均路径长度, $L(G^*v_i)$ 表示节点 v_i 收缩后复杂网络的平均路径长度。

2 基于k-shell值的改进节点收缩法

2.1 节点位置对节点重要性评估的影响

采用节点收缩法对节点重要性进行评估时,虽然同时考虑到节点的度和节点位置对节点重要性的影响,但在实际的复杂网络中,节点位置这个因素并不能很好地发挥作用,因为网络平均距离的变化在一般情况下是很微小的^[10]。一个节点收缩后所造成的网络平均距离的最大变化量是2,即该网络所有最短距离的路径都要经过该节点。

但在大多数情况下,网络平均距离的变化 $L(G) - L(G^*v_i) \leq 2$ 。当一个边缘节点 v_a 具有较大的度,一个中间关键位置的节点 v_b 具有较小的度,最终也可能是 v_a 得到的IMC值较大,节点 v_a 的重要性高于节点 v_b 。

图1(a)为具有12个节点和12条边的无向无权网络 G ,图1(b)为节点 v_4 收缩后得到的无向无权网络 G^*v_4 ,图1(c)为节点 v_5 收缩后得到的无向无权网络 G^*v_5 ,图1(d)为节点 v_6 收缩后得到的无向无权网络 G^*v_6 。由于图 G 的拓

扑结构对称,相对称节点的IMC值也是相同的。表1为利用节点收缩法得到网络 G 的每个节点的IMC值。根据图1可以看到,虽然节点 v_4 的度是4,但其节点位置处于边缘,节点 v_5 的度是3,但是节点 v_5 可起到连接左右两部分桥梁的作用。若将图 G 看成一个信息传播网络或电力输送网络,攻击节点 v_5 要比攻击节点 v_4 对整个网络的影响更大。在节点的度和位置两个影响因素不减弱的情况下,需要基于网络全局属性,提升网络核心位置上节点的重要性。

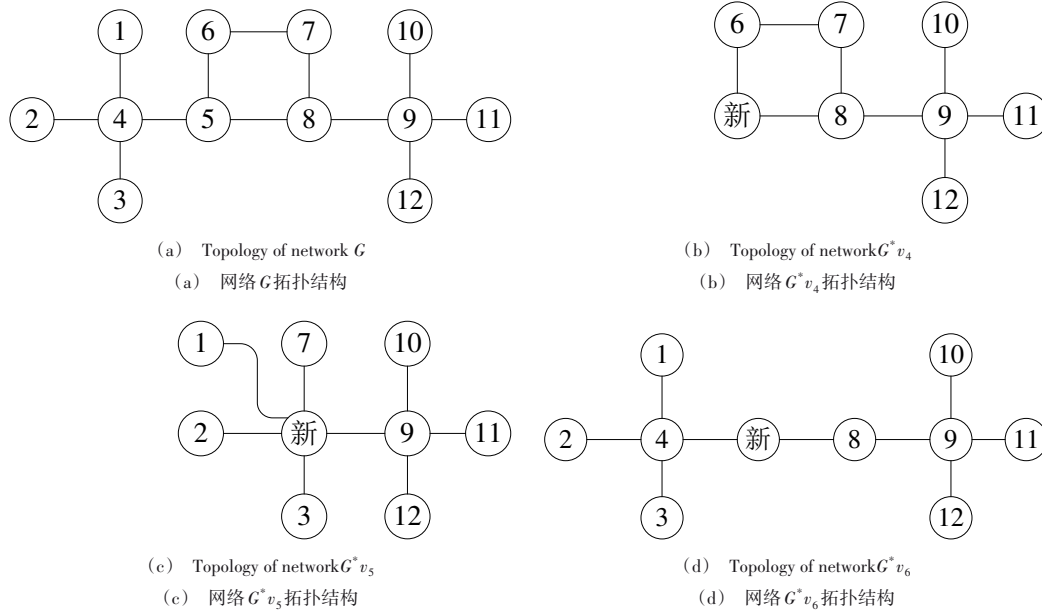


Fig. 1 Changes in network topology structure

图1 网络拓扑结构变化

Table 1 IMC value obtained by node concentrate method

表1 节点收缩法得到的IMC值

节点	1	2	3	4	5	6	7	8	9	10	11	12
度	1	1	1	4	3	2	2	3	4	1	1	1
IMC	0.112 9	0.112 9	0.112 9	0.484 6	0.379 0	0.165 5	0.165 5	0.379 0	0.484 6	0.112 9	0.112 9	0.112 9

2.2 基于k-shell值的改进节点收缩法原理与算例

k-shell算法在全局网络拓扑结构的层面给出判断节点在网络中位置的方法,但是会将很多节点分在相同层,导致排序效果不是很好,是一种粗粒化的节点划分方法。k-shell算法将网络节点一层层剥离后得到网络的核心节点,由该方法得到的核心节点不一定有很大的度,但是节点位置是比较重要的^[16]。k-shell值具有网络的全局属性,引入节点收缩法后,可以提高核心节点位置这一因素的影响力。首先要求出复杂网络中每个节点的k-shell值,再通过每个节点的k-shell值得到复杂网络有几层壳,即k-shell值可以分为几类。应将节点k-shell值与复杂网络每类k-shell值的和之比作为IMC的系数,故基于k-shell值的改进节点收缩法的节点重要度(K-IMC)计算公式为:

$$K - IMC(v_i) = \lambda_{v_i} \left(1 - \frac{\partial(G^*v_i)}{\partial(G)} \right) \quad (3)$$

其中, $\lambda_{v_i} = \frac{k_{s_i}}{k_{s_a} + k_{s_b} + \dots k_{s_p} + k_{s_q}}$, k_{s_i} 为节点 v_i 的 k-shell 值, $k_{s_a}, k_{s_b}, \dots, k_{s_p}, k_{s_q}$ 为复杂网络 G 的 k-shell 值, 相同的 k-shell 值只使用一次。 G 表示原始的复杂网络, G^*v_i 表示节点 v_i 收缩后的复杂网络。

分析图1的网络 G 可知,节点 $v_1, v_2, v_3, v_4, v_9, v_{10}, v_{11}, v_{12}$ 的k-shell值为1,节点 v_5, v_6, v_7, v_8 的k-shell值为2,故节点 $v_1, v_2, v_3, v_4, v_9, v_{10}, v_{11}, v_{12}$ 的 λ 为0.333 333,节点 v_5, v_6, v_7, v_8 的 λ 为0.666 667。用改进的节点收缩法得到的节点重要度结果如表2所示。根据K-IMC的值排序,节点 v_5 和节点 v_8 的重要性大于节点 v_4 与节点 v_9 ,改进算法的排序结果更符合客观情况。节点 v_5 和节点 v_8 在网络中处于关键位置,即比较脆弱的“桥点”,更容易造成网络不连通。如在疾病传染网络中,控制节点 v_5 和节点 v_8 可以较快地切断病毒传播;在电力输送网络中,维护节点 v_5 和节点 v_8 可以提高整个电力网络的稳定性。

Table 2 K-IMC value of improved node contraction method

表 2 改进节点收缩法的 K-IMC 值

节点	1	2	3	4	5	6	7	8	9	10	11	12
度	1	1	1	4	3	2	2	3	4	1	1	1
K-IMC	0.037 6	0.037 6	0.037 6	0.161 5	0.252 7	0.110 3	0.110 3	0.252 7	0.161 5	0.037 6	0.037 6	0.037 6

2.3 算法步骤及其复杂度

采用基于 k-shell 值的改进节点收缩法对节点重要性进行排序,算法步骤如下:

步骤 1:调用算法求解网络的最短路径矩阵,计算网络 G 的平均最短距离 L(G)。

步骤 2:计算原始网络 G 的网络凝聚度 $\partial(G)$ 。

步骤 3:节点收缩后,计算新网络 G^*v_i 的最短路径矩阵。

步骤 4:节点收缩后,计算新网络 G^*v_i 的网络凝聚度 $\partial(G^*v_i)$ 。

步骤 5:计算每个节点的 k-shell 值,使 k-shell 值与网络所有节点 k-shell 值去重后的和成比例,形成一个系数 λ 。

步骤 6:将形成的系数 λ 与改进前的 IMC 值相乘,得到 K-IMC 值。

在节点收缩法的改进过程中只是加入 k-shell 计算,该算法的核心是网络凝聚度计算,其复杂程度仍为 $O(n^3)$ 。

3 基于标准结构熵的改进节点收缩法

3.1 网络拓扑结构变化对节点重要性评估的影响

节点收缩法同时考虑节点的度和节点位置两方面影响因素,但忽略了网络收缩后拓扑结构的变化。在运用节点收缩法对节点重要性进行排序时,每当节点收缩,网络的拓扑结构就会变化,如表 3 为图 1 中各个节点的度。完全随机网络的节点重要性排序是相同的,而无标度网络的节点重要性排序是不同的。当网络的拓扑结构变化时,这种拓扑结构的变化就会对节点的 IMC 值产生影响,节点收缩后网络标准结构熵的变化表现了网络拓扑结构的变化,由此可通过网络标准结构熵对节点收缩法进行改进。

Table 3 Degree of nodes from figure 1

表 3 图 1 节点的度

节点的度	1	2	3	4	5
图 1(a)	$v_1, v_2, v_3, v_{10}, v_{11}$ v_{12}	v_6, v_7	v_5, v_8, v_4, v_9		
图 1(b)	v_{10}, v_{11} v_{12}	v_6, v_7, v_4 收缩后的 新节点	v_8	v_9	
图 1(c)	$v_1, v_2, v_3, v_{10}, v_{11}, v_7$			v_9	v_5 收缩后的新 节点
图 1(d)	$v_1, v_2, v_3, v_{10}, v_{11}$ v_{12}	v_8, v_6 收缩后的新 节点		v_4, v_9	

3.2 基于标准结构熵的改进节点收缩法原理与算例

无标度网络具有幂律分布的特点,存在极少数具有大量连接的核心节点和大多数具有少量连接的末梢节点,该

特点通常表现在节点的度分布上。用网络的标准结构熵来衡量网络的拓扑结构,网络结构熵与度分布的关系如同随机变量的数字特征与其概率分布的函数关系。熵是系统的一种无序度量,随机网络各节点的重要程度相同,网络被认为是无序的。如果网络是无标度网络,网络中有大量核心节点,节点的重要程度存在差异,说明网络是有序的^[17]。使用节点收缩法后,收缩后的网络与收缩前的网络熵差距越大,说明网络的拓扑结构变化越大,该节点对整体网络的影响越大^[18-20]。选择使用网络标准结构熵的目的是消除节点数目对熵的影响。网络的标准结构熵公式为:

$$\hat{E} = \frac{E - E_{\min}}{E_{\max} - E_{\min}} = \frac{-2 \sum_{i=1}^N I_i \cdot \ln I_i - \ln [4(N-1)]}{2 \ln N - \ln [4(N-1)]} \quad (4)$$

$$E = \sum_{i=1}^N I_i \cdot \ln I_i \quad (5)$$

$$I_i = \frac{k_i}{\sum_{j=1}^N k_j} \quad (6)$$

其中, E_{\max} 为复杂网络的最大网络结构熵,即 $\ln N$; E_{\min} 为复杂网络的最小网络结构熵,即 $\ln [4(N-1)]/2$ 。 k_i 为 v_i 的度, E 为网络结构熵, \hat{E} 为网络的标准结构熵。基于网络标准结构熵改进节点收缩法的节点重要度 (E-IMC) 计算公式为:

$$E - IMC = IMC + (\hat{E} - \hat{E}_{(G^*v_i)}) \quad (7)$$

其中, $\hat{E}_{(G^*v_i)}$ 为节点 v_i 收缩后的网络标准结构熵。

使用图 1 的网络,用改进的节点收缩法得到的节点重要度排序结果如表 4 所示。从表 3 可以看到,节点 v_5 收缩后的度分布变化比节点 v_4 收缩后的度分布变化更明显。由表 4 可知,节点 v_5 的 E-IMC 值为 0.899 76,大于节点 v_4 的 E-IMC 值 0.504 95,说明节点收缩后网络拓扑结构的变化确实会影响节点排序。若将图 G 看成一个铁路运输网络,节点 v_5 造成的拥堵要比节点 v_4 造成的拥堵对整个网络的影响更大,这也说明改进算法的排序结果更符合客观情况。

3.3 算法步骤及其复杂度

基于标准结构熵的改进节点收缩法对节点重要性评估的算法步骤如下:

步骤 1:调用算法求解网络的最短路径矩阵,计算网络 G 的平均最短距离 L(G)。

步骤 2:计算原始网络 G 的网络凝聚度 $\partial(G)$ 。

步骤 3:节点收缩后,计算新网络 G^*v_i 的最短路径

Table 4 E-IMC value of improved node contraction method

表4 改进节点收缩法的E-IMC值

节点	1	2	3	4	5	6	7	8	9	10	11	12
度	1	1	1	4	3	2	2	3	4	1	1	1
E-IMC	0.087 249	0.087 249	0.087 249	0.504 95	0.899 76	0.254 07	0.254 07	0.899 76	0.504 95	0.087 249	0.087 249	0.087 249

矩阵。

步骤4:节点收缩后,计算新网络 G^*v_i 的网络凝聚度 $\delta(G^*v_i)$ 。

步骤5:计算节点收缩前复杂网络的标准结构熵 \hat{E} 和节点收缩后复杂网络的标准结构熵 $\hat{E}_{(G^*v)}$ 。

步骤6:计算复杂网络的标准结构熵 \hat{E} 与 $\hat{E}_{(G^*v)}$ 的差。

步骤7:将两者的差与改进前的IMC值相加,得到E-IMC值。

在节点收缩法的改进过程中只是加入了网络的标准结构熵计算,该算法的核心依然是网络凝聚度计算,其复杂程度为 $O(n^3)$ 。

4 仿真实验与分析

4.1 数据准备

为评价改进的节点收缩法重要性排序算法的有效性,考察IMC值与K-IMC值排序结果对整体网络功能和结构的影响程度,使用两种常见方法对两个算法的排序结果进行验证。第一种方法是基于网络动力学模型评价算法优劣,第二种方法是用网络的鲁棒性评价算法优劣^[21]。选取<https://networkrepository.com/index.php>网站中两个经典的具有不同复杂程度的网络数据集Celegans Network与Euro-road Network,验证各个算法的排序结果对网络功能和结构的影响。两个网络的结构如图2、图3所示,两个网络的一些基本静态特征如表5所示。

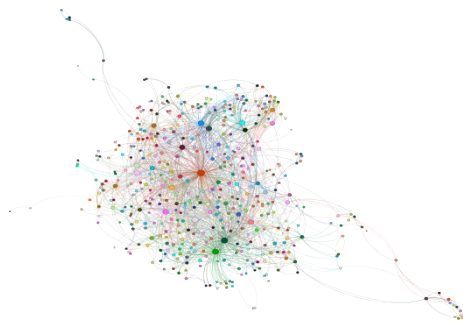


Fig. 2 Celegans Network structure

图2 Celegans Network结构

4.2 基于SIR模型的算法性能评估

使用SIR动力学模型对各个排序算法的结果进行验证。在SIR模型中,节点分为3种状态存在,且每个节点在同一时间只能为一种状态,这些节点分别为易感染节点S、感染节点I、恢复节点R^[22]。节点I在感染状态下,其邻居节点可能会以某种概率被节点I感染,这些节点被称为易



Fig. 3 Euroroad Network structure

图3 Euroroad Network结构

Table 5 Basic static characteristics of two networks

表5 两个网络基本静态特征

静态特征	节点数	连边数	平均度	网络直径	最大核值	聚类系数	图密度
Celegans Network	453	2 025	8.940	7	10	0.655	0.020
Euroroad Network	1 174	1 417	2.414	62	2	0.020	0.002

感染节点S,而被感染的节点I以某种概率恢复为节点R,并且恢复的节点不具再有传染能力,也不会再被感染。选择SIR中的全接触,即I状态的节点感染邻居时是选择所有S态的邻居。利用SIR传播模型对节点重要性进行评价的标准有两项:一是节点的平均感染状态,即达到稳定状态时感染态和恢复态的节点数量总和,数量越大,表示该节点越重要;二是传播速率,根据初始传播阶段的感染态和恢复态数量和达到稳态时所用时间来判断,数量越大,所用时间越短,表明该节点越重要。

采用的节点感染率和恢复率都为0.5,循环迭代次数为100。将度中心性(DC)、接近中心性(CC)、介数中心性(BC)、特征向量中心性(EC)、k-shell值法(K-shell)、熵值法(NSE)、节点收缩法(IMC)、基于k-shell值的节点收缩法(K-IMC)和基于网络结构熵(E-IMC)的节点收缩法共9种算法的结果进行排序,截取前20个重要节点作为感染节点,比较每个算法的感染速度和最大感染率,所得结果如图4、图5所示。

图4和图5中的(a)和(b),以循环迭代次数 t 为X轴,以排名前20的重要节点在 t 步中感染的节点数量占总数量的变化率 $F(t)$ 为Y轴。图4和图5中的(c),以每种算法的名称为X轴,以每种算法前20个重要节点所造成的最大感染率 M 为Y轴。在Celegans Network中,E-IMC算法得到的重要节点造成的感染率几乎是所有算法中最高的,K-IMC算法得到的关键节点造成的感染率比E-IMC算法略低,K-IMC算法和E-IMC算法得到的重要节点造成的感染率都比IMC算法高。在Euroroad Network中,可以更清晰地看到K-IMC算法和E-IMC算法得到的重要节点造成的

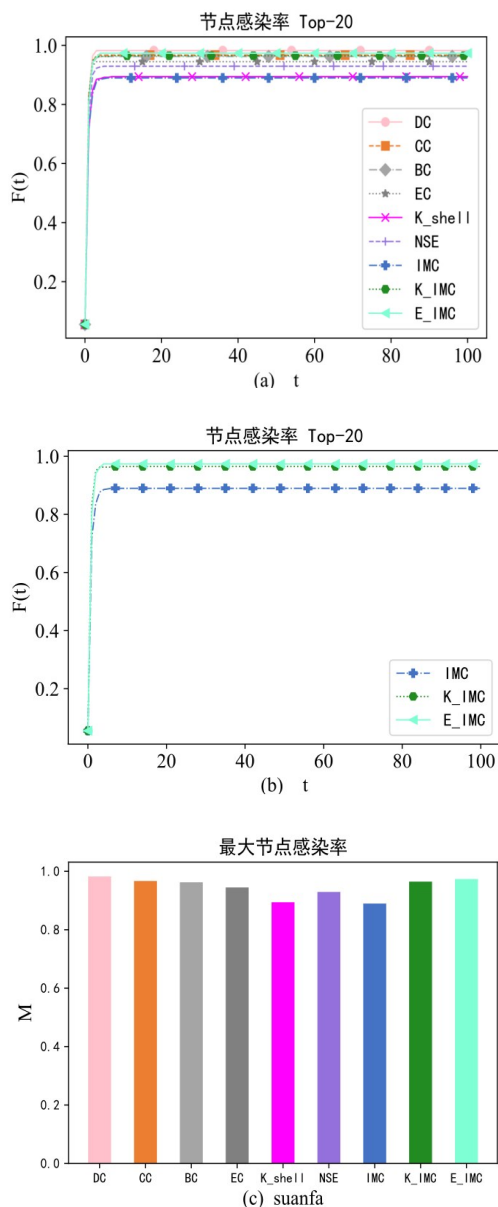


Fig. 4 Infection and maximum infection rate of the first 20 nodes in Celegans Network

图4 Celegans Network前20个节点感染情况与最大感染率

感染率都比IMC算法高。总而言之,K-IMC算法与E-IMC算法在两个网络上的感染率指标表明,K-IMC算法与E-IMC算法能够对网络中的重要节点进行有效评估。

4.3 基于鲁棒性测试的算法性能评估

在4.2节中通过节点传染性的强弱和快慢验证了重要节点排序的准确度,由于感染率和恢复率是一种概率,使用SIR模型验证的准确度也存在一定的不可重复性。为了更准确地验证改进算法的有效性,对上述两个网络进行网络鲁棒性测试。对节点的重要性从大到小进行排序,依次将一部分节点从网络中移除,观察网络连通度的变化。节点删除后,网络的连通度下降率越大,说明节点越重要。网络的鲁棒性也用连通度进行刻画。以删除的节点数目

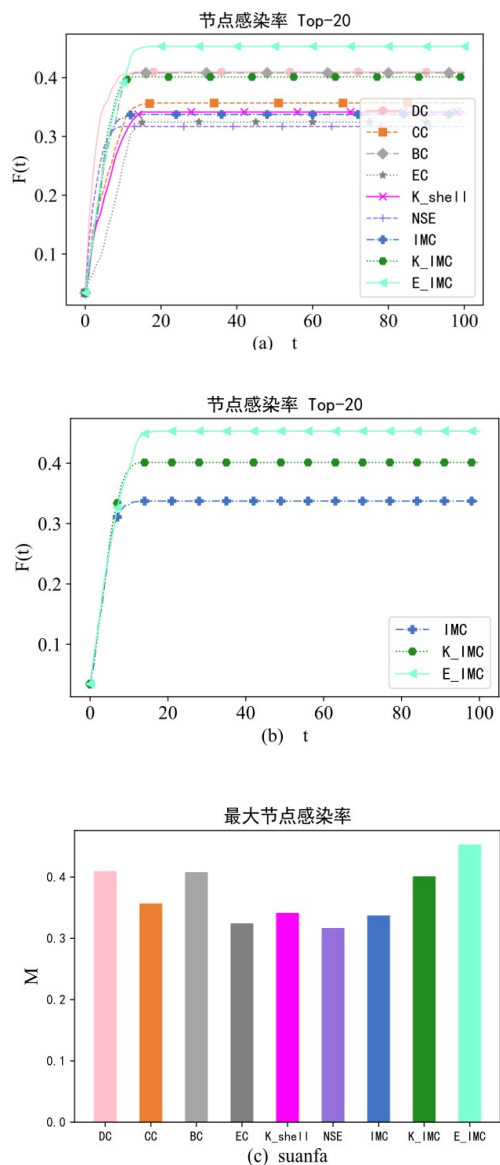


Fig. 5 Infection and maximum infection rate of the first 20 nodes in Euroroad Network

图5 Euroroad Network前20个节点感染情况与最大感染率

为X轴,以删除后的网络连通度 $P(n)$ 为Y轴。两种网络的鲁棒性测试结果如图6、图7所示。在Celegans Network中,E-IMC算法排序结果造成的连通度下降率是最高的,K-IMC算法排序结果造成的连通度下降率比IMC算法高;在Euroroad Network中,E-IMC算法和K-IMC算法排序结果造成的连通度下降率也比IMC算法高。实验结果表明,两种改进的节点收缩法排序结果使网络连通度下降较快,表明了改进算法具有更高的准确性和有效性。

4.4 两种改进的节点收缩法对比分析

本节只对两种改进的节点收缩法进行对比分析,表明其在准确性和运算效率上各有优势,形成互补。如图4、图5所示,利用网络标准结构熵改进的节点收缩法和基于k-shell值改进的节点收缩法排序都可以造成更多节点传染,

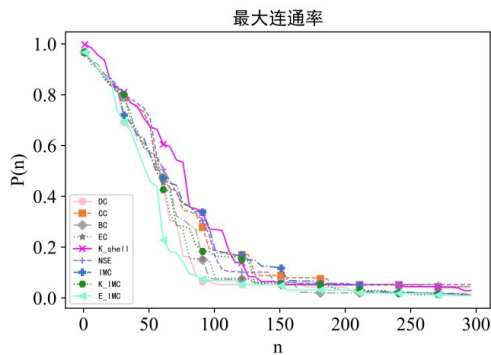


Fig. 6 Connectivity decline rate of Celegans Network
图6 Celegans Network 连通度下降率

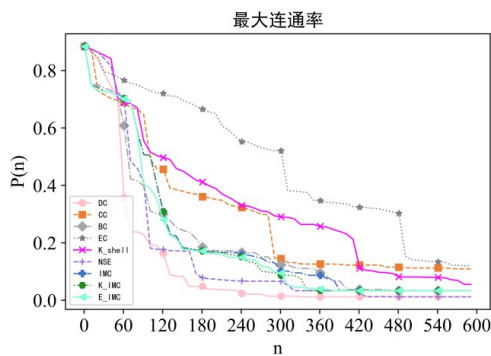


Fig. 7 Connectivity decline rate of Euroroad Network
图7 Euroroad Network 连通度下降率

或者可以使节点的感染率更快达到最高点;如图6、图7所示,按照基于网络标准结构熵改进节点收缩法的排序结果删除节点,网络的连通度下降率会更高,表明基于网络标准结构熵改进的节点收缩法排序具有更高的准确性。

在运行效率上,基于网络标准结构熵改进的节点收缩法排序所用时间较长,其运行效率低于基于k-shell值改进的节点收缩法。在Celegans Network中基于k-shell值改进的节点收缩法运行时间为40s,基于网络标准结构熵的节点收缩法运行时间为43.205s;在Euroroad Network中基于k-shell值改进的节点收缩法运行时间为384.819s,基于网络标准结构熵的节点收缩法运行时间为562.302s。k-shell值是粗粒化的节点划分,只计算没有节点收缩的复杂网络的k-shell值,而网络的标准结构熵在每个节点收缩后都要计算一遍,基于k-shell改进的节点收缩法比基于网络标准结构熵改进的节点收缩法排序结果运行时间更短,运行效率更高。

5 结语

结合节点核心位置和网络拓扑结构的变化,本文提出新的节点重要性度量:基于k-shell值的节点重要度(K-IMC)和基于网络标准结构熵的节点重要度(E-IMC),由此对节点收缩法进行改进,从而进行复杂网络的重要节点评估。基于两类改进节点收缩法的复杂网络重要节点评估

算法的核心是网络凝聚度计算,其复杂程度为 $O(n^3)$,并在Celegans Network和Euroroad Network两个复杂网络上进行了仿真实验,运用SIR模型与鲁棒性测试对算法的准确性和有效性进行评价与分析。

研究结果表明,相较于未改进的节点收缩法,基于k-shell值改进的节点收缩法和基于网络标准结构熵改进的节点收缩法在重要节点排序结果的准确性上均表现更好。通过对两类改进的节点收缩法进行对比分析,结果显示,基于E-IMC的算法相较于基于K-IMC的算法准确性更高,基于K-IMC的算法相较于基于E-IMC的算法运算效率更高。因此,在复杂网络的重要节点排序中,若更注重排序结果的准确性,可选择基于网络标准结构熵改进的节点收缩法,若更注重排序的运算效率,则可选择基于k-shell值改进的节点收缩法。

参考文献:

- [1] SI S K, SUN X J. Complex network algorithms and applications [M]. Beijing: National Defense Industry Press, 2015.
司守奎,孙玺菁. 复杂网络的算法与应用[M]. 北京:国防工业出版社, 2015.
- [2] WANG X F, LI X, CHEN G R. Complex network theory and application [M]. Beijing: Tsinghua University Press, 2006.
汪小帆,李翔,陈关荣. 复杂网络理论及其应用[M]. 北京:清华大学出版社, 2006.
- [3] REN X L, LYU L Y. Review of ranking nodes in complex networks [J]. Chinese Science Bulletin, 2014, 59(13): 1175-1197.
任晓龙,吕琳媛. 网络重要节点排序方法综述[J]. 科学通报, 2014, 59(13): 1175-1197.
- [4] BOCCALETTIAB S, BIANCONIC G, CRIADODE R, et al. The structure and dynamics of multilayer networks [J]. Physics Reports, 2014, 544(1): 1-122.
- [5] GUO C G. Algorithm research and design of vital nodes identification in complex networks [D]. Chengdu: University of Electronic Science and Technology of China, 2020.
郭春谷. 复杂网络中重要节点挖掘的算法研究与设计[D]. 成都:电子科技大学, 2020.
- [6] ZHU J F, CHEN D B, ZHOU T, et al. A survey on mining relatively important nodes in network science [J]. Journal of University of Electronic Science and Technology of China, 2019, 48(4): 595-503.
朱军芳,陈端兵,周涛,等. 网络科学中相对重要节点挖掘方法综述[J]. 电子科技大学学报, 2019, 48(4): 595-503.
- [7] JIU L, HUI F, SHENG Q. Identifying important nodes for temporal networks based on the ASAM model [J]. Physica A, 2022, 586: 1-13.
- [8] WANG K, MA H, HUANG R Y. Hierarchical structure analysis of social network based on important nodes removal method [J]. Computer Engineering, 2018, 44(2): 174-181.
王魁,马宏,黄瑞阳. 基于重要节点删除法的社会网络层次结构分析[J]. 计算机工程, 2018, 44(2): 174-181.
- [9] TAN Y J, WU J, DENG H Z. Evaluation method for node importance based on node contraction in complex networks [J]. Systems Engineering—Theory & Practice, 2006(11): 79-83.
谭跃进,吴俊,邓宏钟. 复杂网络中节点重要性评估的节点收缩方法[J]. 系统工程理论与实践, 2006(11): 79-83.

- [10] ZHU T, ZHANG S P, GUO R X, et al. Improved evaluation method for node importance based on node contraction in weighted complex network [J]. *Systems Engineering and Electronics*, 2009, 31(8): 1902-1905.
朱涛,张水平,郭戎潇,等.改进的加权复杂网络节点重要度评估的收缩方法[J].*系统工程与电子技术*, 2009, 31(8): 1902-1905.
- [11] WANG J S, WU X P, LIAO W, et al. Improved method of node importance evaluation weighted complex networks [J]. *Computer Engineering*, 2012, 38(10): 74-76.
王甲生,吴晓平,廖巍,等.改进的加权复杂网络节点重要度评估方法[J].*计算机工程*, 2012, 38(10): 74-76.
- [12] KITSAK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks [J]. *Nature Physics*, 2010, 6: 888-893.
- [13] WANG K L, WU C X, AI J, et al. Complex network centrality method based on multi-order K-shell vector [J]. *Acta Physica Sinica*, 2019, 68(19): 1-11.
王凯莉,郭春学,艾均,等.基于多阶邻居壳数的向量中心性度量方法[J].*物理学报*, 2019, 68(19): 1-11.
- [14] TAN Y J, WU J. Network structure entropy and its application to scale-free networks [J]. *Systems Engineering-Theory & Practice*, 2004(6): 1-3.
谭跃进,吴俊.网络结构熵及其在非标度网络中的应用[J].*系统工程理论与实践*, 2004(6): 1-3.
- [15] HU G, GAO H, XU X, et al. Importance identification method of complex network nodes based on importance transfer matrix [J]. *Acta Electronica Sinica*, 2020, 48(12): 2402-2408.
胡钢,高浩,徐翔,等.基于重要度传输矩阵的复杂网络节点重要性辨识方法[J].*电子学报*, 2020, 48(12): 2402-2408.
- [16] JING Z, JIE T, CONG M, et al. Panther: fast top-k similarity search on large networks [C]//Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015: 1445-1454.
- [17] WANG H. Research on vital nodes identification in complex network based on improved K-shell algorithm [J]. Shanghai: East China Normal University, 2019.
王环.基于改进K-shell算法的复杂网络关键节点识别研究[D].上海:华东师范大学, 2019.
- [18] ZHOU L N, CHANG X. Using adjacent structure entropy to determine vital nodes of hypernetwork [J]. *Computer Engineering and Application*, 2022, 58(8): 76-83.
周丽娜,常笑.利用邻接结构熵确定超网络关键节点[J].*计算机工程与应用*, 2022, 58(8): 76-83
- [19] WANG Q, HU S W, GUO J W, et al. Structure entropy of directed complex network based key node mining algorithm in software dynamic execution [J]. *Journal of Chinese Computer Systems*, 2019, 40(4): 884-889.
王倩,胡松旺,郭嘉伟,等.有向复杂网络结构熵的软件动态执行关键节点挖掘算法[J].*小型微型计算机系统*, 2019, 40(4): 884-889.
- [20] GENG J, SUN X, LI F, et al. Prediction method of important nodes and transmission lines in power system transactive management [J]. *Electric Power Systems Research*, 2022, 208: 1-8.
- [21] WANG A, GU Y J. Ranking and simulation of nodes importance of social network based on articulation point removal [J]. *Computer Simulation*, 2021, 38(3): 272-276.
王安,顾益军.基于割点移除的社交网络重要节点评估与仿真[J].*计算机仿真*, 2021, 38(3): 272-276.
- [22] ZHAO M, LI W, CAO J. Dynamics for an SIR epidemic model with nonlocal diffusion and free boundaries [J]. *Acta Mathematica Scientia*, 2021, 41(4): 1081-1106.

(责任编辑:黄健)