

基于改进随机游走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

收稿日期: 2023年1月23日; 录用日期: 2023年2月17日; 发布日期: 2023年2月23日

摘要

复杂系统可以抽象为复杂网络, 重要节点评估与识别是复杂网络的一个热点问题。针对网络拓扑结构和节点自身属性对有向复杂网络重要节点的影响, 提出基于改进随机游走的节点重要性评估方法。首先对节点的出度和入度分别附参数求出节点联合度数为节点质量, 并通过调节参数评估节点出度与入度对节点重要性的影响; 其次使用SimRank算法得任意两个节点相似值的倒数为引力模型的距离, 考虑节点间的拓扑结构; 最后通过相对路径数比值做引力模型的系数, 考虑节点间信息传播的影响效果。任意两节点的作用力构造引力矩阵, 将引力矩阵行归一化构造转移矩阵, 然后随机游走对节点进行排序。使用极大强连通性、极大弱连通性和脆弱性等评估指标在四个真实网络上进行实验对比, 结果表明, 提出的算法相比LeaderRank、PageRank、HITs等方法能更准确地评估节点的重要性。将复杂网络的多种特征进行融合, 新创建的重要节点评估方法可以运用在生物领域和经济贸易领域等。

关键词

有向复杂网络, 节点重要度, 节点相似性, 引力模型, 相对路径

Evaluation of Node Importance in Complex Networks Based on Improved Random Walk

Xiaonan Cai, Zhongtuan Zheng*

School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai

Received: Jan. 23rd, 2023; accepted: Feb. 17th, 2023; published: Feb. 23rd, 2023

Abstract

Complex systems can be abstracted as complex networks. The evaluation and identification of important nodes is a hot issue in complex networks Aiming at the influence of network topology

*通讯作者。

and node attributes on important nodes in directed complex networks, a node importance evaluation method based on improved random walk is proposed. Firstly, the joint degree of the node is calculated as the node quality by attaching parameters to the outgoing degree and incoming degree of the node respectively, and the influence of the outgoing degree and incoming degree of the node on the node importance is evaluated by adjusting parameters; Secondly, SimRank algorithm is used to get the reciprocal of the similarity value of any two nodes as the distance of the gravity model, considering the topology between nodes; Finally, the relative path number ratio is used as the coefficient of the gravity model to consider the effect of information transmission between nodes. The gravitational matrix is constructed by the force of any two nodes, and the transfer matrix is constructed by normalizing the rows of the gravitational matrix, and then the nodes are ranked by random walking. The experimental comparison is conducted on four real networks using the evaluation indicators of maximum strong connectivity, maximum weak connectivity and vulnerability. The results show that the proposed algorithm can more accurately evaluate the importance of nodes than the methods of Leader Rank, Page Rank, HITs, etc., integrate multiple features of complex networks, and the newly created evaluation method of important nodes can be applied in the biological field, economic and trade fields, etc.

Keywords

Directed Complex Network, Node Importance, Node Similarity, Gravitational Model, Relative Path

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

从自然界到人类社会、从生态关系到社会关系都存在着大量的复杂系统, 这些常见的复杂系统大多是由一系列基本单元以及单元们之间的关系所组成的。这些基本单元以及单元们之间的关系被抽象成复杂网络的点和边。复杂网络分为有向无权的复杂网络、有向有权的复杂网络、无向有权的复杂网络和无向无权的复杂网络, 复杂网络往往具有大量的节点, 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小世界性、无标度性、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一个重要的热点问题。

重要节点是相比于网络中的其他的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]。发现网络中的重要节点, 有助于提升网络的鲁棒性和抗毁性, 有助于对网络上的动力学行为进行控制和预测[4]。节点重要性评估在网络安全、信息传播、和疾病免疫等方面具有重大的实际应用价值。

经典度量复杂网络的方法有很多, 有学者认为节点的度越大节点越重要, 如度中心性、半局部中心性等; 有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、介数中心性等; 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; 还有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越重要, 如特征向量中心性、PageRank 算法、LeaderRank 算法、HITs 算法等[5] [6]。特征向量中心性和累计提名方法一般用在无向网络中, PageRank 算法和 LeaderRank 一般应用于有向网络中。HITs 算法、自动信息汇集算法, SALSA 算法都考虑节点的双重角色: 权威性和枢纽性, 并认为两者相互影响[3]。这些算法只考虑到了节点或网络某一特定性质, 没将节点和网络的多种性质进行融合考虑。

针对上面经典算法中出现的问题, 许多学者都提出了方法。段杰明[7]提出了一种结合网络结构局域

信息和标签扩散的节点排序算法。王凯莉等人[8]在 k-shell 分解法基础上考虑了节点自身壳值与其多阶邻居节点的壳值。马媛媛等人[9]提出了一种综合考虑节点的邻居数量和节点与邻居间亲密程度的节点重要度评估算法。张宪立等人[10]根据邻居节点的拓扑结构并结合万有引力定律, 提出一种基于改进重力中心性的复杂网络节点重要性评估方法。胡钢等人[11]基于节点间的最优路径长度、最优路径数目和信息传播率定义节点间的传输能力, 构建重要度传输矩阵, 最后综合节点局部和全局属性来评估节点的重要度大小。李秋晖等人[12]提出了基于引力模型及相对路径数的节点重要度评估算法。以上文献在评估节点重要性时, 未能充分考虑节点自身属性和网络拓扑结构对网络节点重要性的影响。

针对这些问题, 本文提出了基于引力模型的随机游走节点重要性的评估方法。首先, 通过调节节点出度和入度的比例做节点质量、以节点间相似值的倒数为引力模型中的距离、以节点间相对路径之比作为引力模型系数, 构造引力模型公式并度量网络节点间的相互作用力大小; 其次, 对形成的引力矩阵进行归一化得到可以随机游走的转移矩阵。本文提出的基于改进引力模型的随机游走复杂网络节点重要性评估的方法有以下优势和贡献:

一、该方法同时充分考虑了节点的出度和入度。连边的方向性反映节点的重要性, 只考虑节点的连入边, 不考虑节点的连出度, 不符合现实, 例如在信息传播中, 一个节点只接受信息, 不传播信息, 该节点是不重要的。讨论节点合适的出度和入度比重对节点属性的影响。

二、完全基于复杂网络的结构信息, 使节点间相关度最大化。考虑网络结构对节点重要性的影响, 在不同复杂网络中, 该方法具有更强的适用性。

三、节点间的相对路径之比考虑了四阶路径之内的所有传播信息, 充分考虑节点周围的局部信息, 比只考虑节点间的最短路径得到的排序结果更准确。

2. 基于引力模型的随机游走算法

根据复杂网络的边是否有向, 将网络分为有向网络和无向网络。构建有 N 个节点和 M 条边的有向无权网络模型, 用图 $G=(V,E)$ 表示, 网络节点合集为 $V=\{v_1, v_2, v_3, \dots, v_n\}$, 网络节点边的合集为 $E=\{e_1, e_2, e_3, \dots, e_m\}$, 每条边的权重都为 1。网络的邻接矩阵可以表示为 $A_{n \times n}=(a_{ij})$, 当一条边从节点 v_i 指向节点 v_j 时, $a_{ij}=1$; 如果没有节点 v_i 指向节点 v_j , 则 $a_{ij}=0$ 。在有向网络中, 节点的度根据节点边的方向分为出度和入度, k^{in} 表示节点的入度, k^{out} 表示节点的出度。

基于随机游走思想的方法, 解决有向复杂网络图的不连通问题, 保证算法的收敛效果, 构建合适的转移矩阵, 转移矩阵携带的信息对排序结果有重大的影响。为了准确的量化转移矩阵对节点重要性的影响, 融合节点属性、节点局部信息与网络结构到引力模型中, 构建两个节点的相互作用力, 通过作用力之间的大小表达节点间相互转移的可能性, 将任意两节点之间的相互力形成作用力矩阵, 最后通过行归一化形成转移矩阵。给网络添加背景点, 形成强连通图, 保证随机游走成功。

2.1. 引力模型

根据万有引力模型[13][14]知, 一种物质对其他物质产生的作用力其实也就是这个物质在其自身周围产生特定的场。任意两个物质之间的引力关系为:

$$F = G \frac{M_1 M_2}{r^2} \quad (1)$$

其中: M 表示物质自身的为质量。 r 可以是两个物体之间的最短距离。 G 可以系数。

1) 确定质量 M

节点的度[10]是与该节点直接相连接的其他节点的数目, 在 N 个节点的无权无向网络中, 节点 v_i 的

度表示为:

$$DC(v_i) = \sum_{j=1}^N A_{ij} \quad (2)$$

在 N 个节点的有向无权网络中, 节点 v_i 的出度和入度分别表示为:

$$k^{out}(v_i) = \sum_{j=1}^N e_{ij} \quad (3)$$

$$k^{in}(v_i) = \sum_{j=1}^N e_{ji} \quad (4)$$

其中: 节点 v_i 的出度 $k^{out}(v_i)$ 表示从节点 v_i 出发到达邻居节点 v_j 的连边数; 节点 v_i 的入度 $k^{in}(v_i)$ 表示从邻居节点 v_j 出发到达节点 v_i 的连边数。

其他经典随机游走的算法都只关注了节点的入度或出度, 本文通过系数 λ 调节节点的出度和入度同时对节点重要性的影响, 生成衡量节点属性的新指标: 综合度值, 综合度值的计算公式为:

$$k_i = \lambda k_i^{in} + (1 - \lambda) k_i^{out} \quad (5)$$

其中 k_i 表示有向网络中节点 v_i 的综合度, λ 表示调节节点出度与入度重要性的系数。

2) 确定距离 r

改进的引力模型在分析节点间相互作用大小时, 考虑节点与邻居间不同拓扑结构对节点的影响力不同, 将节点间相似值的倒数作为引力模型的距离[15]。SimRank 是一种有向图上基于图的拓扑结构信息来衡量任意两个对象间相似程度的模型, 得到的节点相似值具有复杂网络全局的信息。该算法的基本思想是: 如果两个对象被其相似的对象所引用, 那么这两个对象也相似[16]。该算法的基本公式为:

$$s(a, b) = \frac{c}{|I(a)||I(b)|} \sum_{i=1}^{|I(a)|} \sum_{j=1}^{|I(b)|} s(I_i(a), I_j(b)) \quad (6)$$

其中 $|I(v)|$ 表示节点 v 的入度, $I_i(v)$ 表示 v 的第 i 个邻居节点, c 为衰减因子, 且 $c \in (0, 1)$, $s(a, b)$ 表示两个节点的相似度。

3) 确定系数 G

在分析节点间相互作用大小时, 只考虑节点综合度数与节点间的相似性往往不够全面, 节点间的距离是不可忽视的。李秋晖[12]等人在无向网络中考虑相对路径数, 考虑二阶、三阶或更长阶的路径传递信息的能力。构造有向网络的作用力矩阵时, 只能构造出边链接转移到邻居节点的作用力, 在考虑节点通过一阶路径传递信息时, 也要考虑到通过其他阶路径传递信息, 不同路径传播信息的能力不同。通过这些相对路径数定义结构参数 G :

$$G = \sum_{l=1}^4 \frac{a_{i \rightarrow j}^l}{l^2 \cdot \sum_{k=1}^N a_{i \rightarrow k}^l} \quad (7)$$

其中: $a_{i \rightarrow j}^l$ 表示节点 v_i 与 v_j 间长度为 l 的路径数目, N 为网络节点数; $\sum_{k=1}^N a_{i \rightarrow k}^l$ 表示从节点 v_i 出发的所有 l 阶路径数目, l 阶路径数目用矩阵表示为 $A_{n \times n}^l$ 。

2.2. 确定转移矩阵和游走方式

本文重新定义网络中任意两节点间相互作用力大小为:

$$F_{i \rightarrow j} = \sum_{l=1}^4 \frac{a_{i \rightarrow j}^l}{l^2 \cdot \sum_{k=1}^N a_{i \rightarrow k}} \times \frac{\lambda k_i^{in} + (1-\lambda) k_i^{out}}{\text{simrank}(v_i, v_j)} \quad (8)$$

其中, $\text{simrank}(v_i, v_j)$ 为节点 v_i 与节点 v_j 的 SimRank 相似值。其他符号表达同公式(5)与公式(7)。路径越长节点间信息传递越差, 该算法只计算路径长度为 1、2、3、4 时的相对路径。由任意两点间的作用力生成作用力矩阵 $F_{n \times n}$ 。将作用力矩阵与有向网络的邻接矩阵对应元素相乘生成只有出边连接的矩阵 $M_{n \times n}$, 再按行对矩阵进行归一化得到转移矩阵 $Q_{n \times n}$, 归一化后需要保证矩阵行的和为 1, 所以使用每个节点的对应值除以该节点所在矩阵的行和即可。

有向复杂网络上随机游走, 可以选择与 PageRank 算法相同的方法, 加入跳转参数; 也可以选择 LeaderRank 算法[17] [18] [19]相同的方法, 加入背景点。这两种方法都保证了图的强连通性, 但是 LeaderRank 算法收敛速度更快, 排序结果唯一, 适用于大型复杂有向网络, 排序的准确性和抗噪声能力较高。转移矩阵做随机游走时, 选择添加背景点, 同 LeaderRank 算法一样构造强连通图。

2.3. 算法流程与步骤

基于引力模型的随机游走算法主要的流程图如图 1 所示, 该算法的步骤如下:

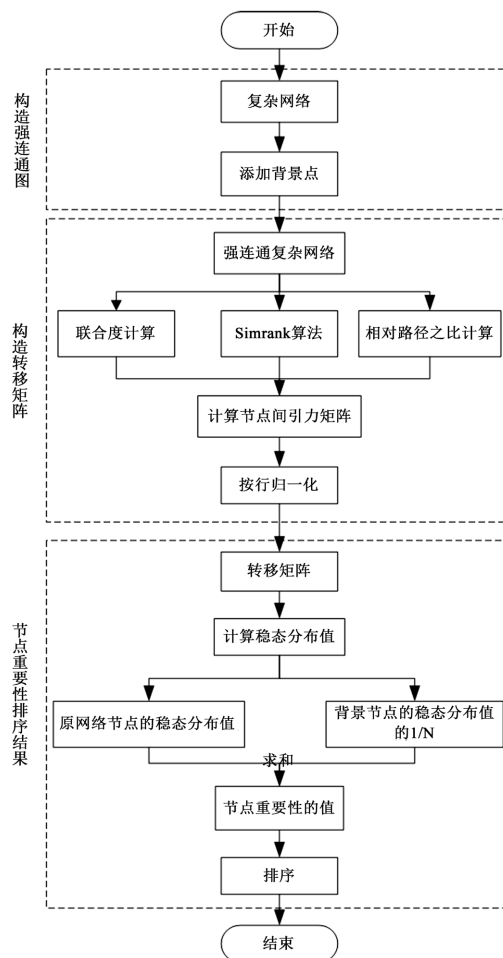


Figure 1. Step flow chart
图 1. 步骤流程图

输入: 有向无权复杂网络图 G 。

输出: 每个节点重要性的排序结果 $Y(v)$ 。

步骤一: 为网络图添加公共点, 并与每个节点双向链接, 生成强连通图 G^* 。

步骤二: 根据公式(5), 计算强连通图 G^* 所有节点的综合度 $k(v)$, 并通过调节 λ 的取值, 得到不同的综合度 $k(v)$ 。

步骤三: 根据公式(6), 计算强连通图 G^* 所有节点与其他节点的相似值, 形成节点相似性矩阵 $A_{sim(i,j)}$ 。

步骤四: 根据公式(7), 计算强连通图 G^* 所有节点与其他任意节点之间四阶路径之内的比, 形成节点系数矩阵 $A_{G(i,j)}$ 。

步骤五: 根据公式(8), 将得到节点综合度 $k(v)$ 节点相似性矩阵 $A_{sim(i,j)}$ 与节点系数矩阵 $A_{G(i,j)}$ 相结合得到节点间的作用力矩阵 $A_{F(i,j)}$ 。将生成的作用力矩阵 $A_{F(i,j)}$ 按行归一化, 生成转移矩阵 $Q_{n \times n}$ 。

步骤六: 根据幂律法求转移矩阵的稳态分布 $Y^*(v)$, 将公共节点的稳态值平均分给其他节点, 得到节点重要性的排序结果 $Y(v)$ 。

3. 仿真实验与分析

3.1. 实验数据

检验基于改引力模型的随机游走算法对网络节点重要度的评估效果, 本文选用四个真实网络数据进行相关实验。Enzymes-g1 网络[20] (简记为 Enzymes)、Friendship-network-data-2013 网络[21] (简记为 Friendship)、Celegans-Genetic 网络和与 CKM-Physicians-Innovation-Social 网络[22] (简记为 Celegans、Physicians), 数据均来源于网络数据库。表 1 为四个网络的基本统计特征, 其中: N 与 M 分别表示网络的节点数与边数; D 为网络直径; $\langle D \rangle$ 为网络的平均路径长度; $\langle d \rangle$ 为网络的图密度; $\langle k \rangle$ 为网络平均度; $\langle c \rangle$ 为网络的平均聚类系数。

Table 1. Basic statistical characteristics of network
表 1. 网络的基本统计特征

网络	Enzymes	Friendship	Celegans	Physicians
N	297	134	239	241
M	2345	668	313	1099
$\langle k \rangle$	7.896	4.985	1.744	4.56
D	14	12	6	9
$\langle D \rangle$	3.992	4.561	1.744	3.306
$\langle d \rangle$	0.027	0.037	0.006	0.019
$\langle c \rangle$	0.169	0.474	0.152	0.199

本文采用三个常用的指标评估节点的重要性, 即强连通性、弱连通性、脆弱性。其中在评判网络的连通性时既要考虑子团的数量, 也需要考虑最大连通子团的节点数量。仿真实验证明两个问题: 第一, 改进的算法比 PageRank 算法、LeaderRank 算法、HITs 算法具有更高的准确性; 第二, 不同出度和入度比值对节点综合度有影响, 从而影响重要节点的排序结果, 选取不同的 λ 值是为了找到比较合适的取值范围, 由公式(5)知, λ 取值不同, 作用力大小不同, 节点重要性的排序结果也不同, 寻找 λ 的最佳取值

范围。表 2 为计算节点重要性时 λ 的取值。

Table 2. Value of λ

表 2. λ 网的取值

名称	DSL_LR_0.05	DSL_LR_0.15	DSL_LR_0.25	DSL_LR_0.35	DSL_LR_0.45	DSL_LR_0.55	DSL_LR_0.65	DSL_LR_0.75	DSL_LR_0.85	DSL_LR_0.95
λ	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95

3.2. 强连通性评估指标

若 G 为有向图, 若 G 中存在一条以节点 v_i 为起点, 节点 v_j 为终点的有向路 P , 则称从节点 v_i 到节点 v_j 是可达的。如果 G 的任何两个顶点都是相互可达的, 则称图 G 是强连通的[23] [24] [25], 图 G 被称为强连通图。非强连通有向图的极大强连通子图, 称为强连通分量(strongly connected components), 如果一个强连通分量中不再能加入任何一个顶点, 则这个强连通分量是一个极大强连通分量。确定非强连通有向图每一个强连通分量的节点数, 按节点的个数确定强连通分量的规模。则极大强连通率计算公式为:

$$GS = \frac{n_g}{n} \tag{9}$$

其中: n 表示网络的节点数, n_g 表示删除部分节点后剩余网络中规模最大的强连通分量的节点数。 GS 越小, 说明极大强连通分量的节点越少, 网络被破坏的越厉害, 节点排序结果的准确性越高。

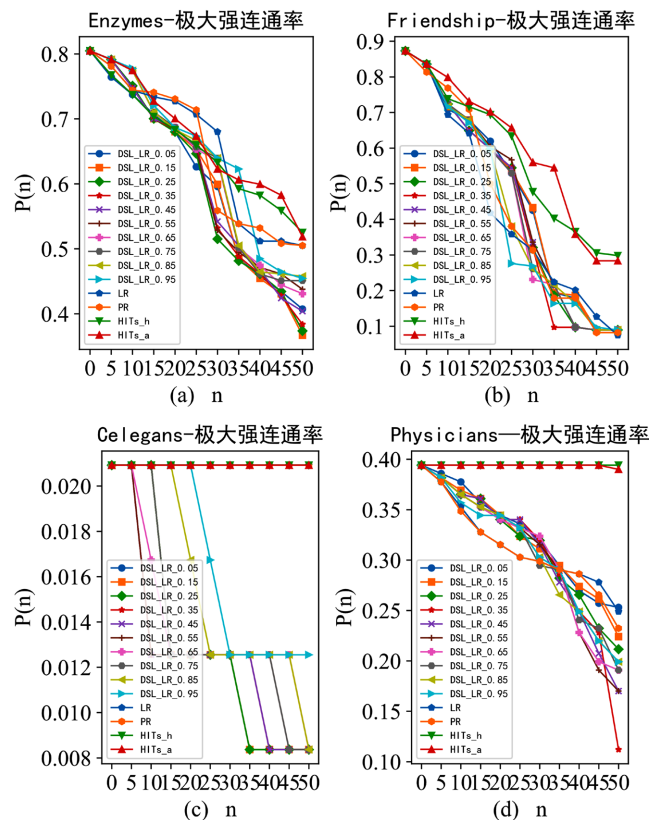


Figure 2. Change of strong connectivity after removing different node numbers
图 2. 移除不同节点数极大强连通率的变化情况

图 2 所示为四个真实网络的极大强连通率 GS 在不同攻击条件下随节点删除数目的变化情况。容易看出: 在 *Enzymes* 网络中, 改进算法的排序结果对网络的攻击效果明显好于其他经典算法。在 *Friendship* 网络中, 删除排序前 10 到前 25 的节点, 依据 LR 算法和改进算法的排序结果造成极大强连通率的变化是很相似的, 但是删除前 20 个节点到前 50 个节点时, λ 取值 0.35、0.45、0.55、0.65、0.75、0.85、0.95 的改进算法的排序结果造成极大强连通率的下降都要比其他算法快。在 *Celegans* 网络, λ 大部分取值的排序结果的准确性都要比其他经典算法好。在 *Physicians* 网络中, 虽然删除前 5 到前 25 个节点时, 改进算法的极大强连通率比其他算法高, 但是删除前 20 个节点到前 50 个节点时, 改进算法的极大强连通率比其他算法低, 且下降的速度也比较快, 改进算法呈现出较好的准确性。

复杂网络进行仿真实验中, 网络移除节点后会破碎成零散的网络, 这些零散的网络数量表明了网络的破碎程度, 所以在评判网络的连通性时也要考虑子团的数量。强连通分量数越多, 说明网络破碎的越厉害, 算法的效果越好。图 3 所示为四个真实网络在不同攻击条件下, 随节点删除数目的变化网络破碎成具有强连通性子网络数量情况。因为 λ 取值比较多, 图例不能很好的在图形中显示, 每删除 10 个节点, 从左到右的小条形图分别表示为 λ 取值 0.05、0.15、0.25、0.35、0.45、0.55、0.65、0.75、0.85、0.95 的改进、leaderRank 算法、PageRank 算法、HITs 的枢纽值和权威值排序结果造成的子网络数量。在 *Enzymes* 网络、*Celegans* 网络和 *Physicians* 网络中, λ 大部分取值的排序结果形成的强连通分量数都多于其他算法。*Friendship* 网络在删除前 20 个节点时, 排序效果与经典算法相似, 但删除前 40 或前 50 个节点时, 改进的算法显示出较好的准确性。

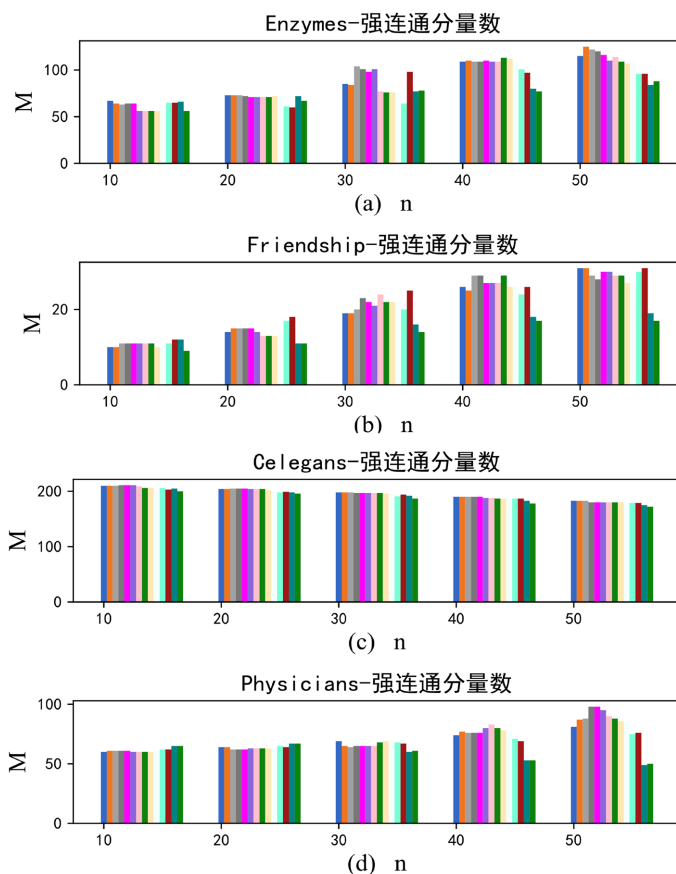


Figure 3. Change of the number of subnetworks with different node numbers removed
 图 3. 移除不同节点数子网络数量的变化情况

总体来看, 改进算法在极大强连通率和强连通分量数两个评估指标中都表现出较高的准确性, 且大部分的 λ 取值的改进算法都表现出较好的准确性。

3.3. 弱连通性评估指标

将有向图的所有有向边替换为无向边, 所得到的图称为原图的基图。如果一个有向图的基图是连通图, 则有向图是弱连通图[25] [26]。在对复杂网络进行攻击的仿真实验中, 网络受到攻击后会破碎成零散的网络子图, 按节点的个数确定弱连通分量的规模。则极大弱连通率计算公式为:

$$GW = \frac{n_g}{n} \quad (10)$$

其中: n 表示网络的节点数, n_g 表示删除部分节点后剩余网络中规模最大的弱连通分量的节点数。 GW 越小, 说明极大弱连通分量的节点越少, 网络被破坏的越厉害, 节点排序结果的准确性越高。

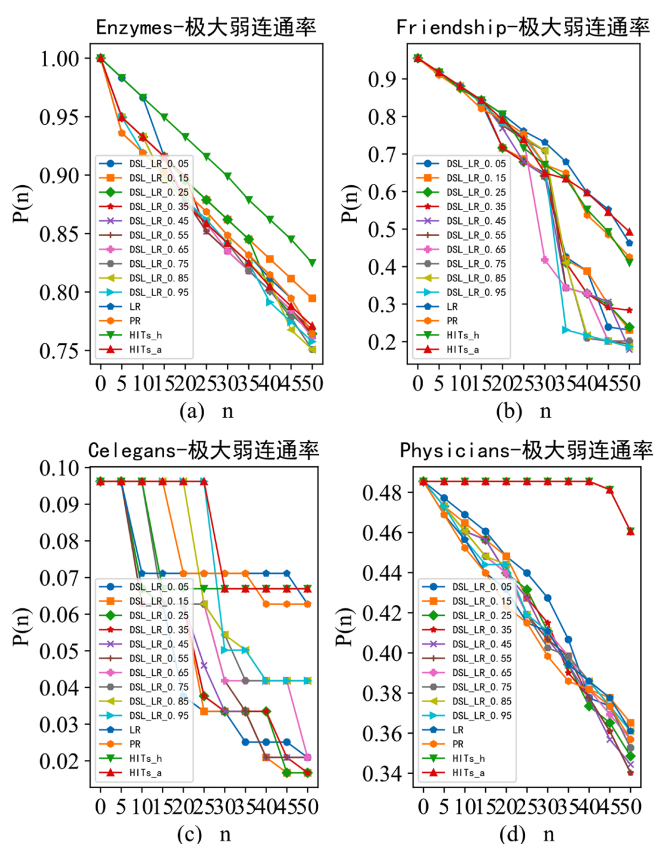


Figure 4. Change of maximal weak connectivity with different node numbers removed

图 4. 移除不同节点数极大弱连通率的变化情况

图 4 所示为四个真实网络的极大弱连通率 GW 在不同攻击条件下随节点删除数目的变化情况。容易看出: 在 Friendship、Celegans 网络中, 改进的算法对网络的攻击效果明显好于其他经典的算法。在 Enzymes 网络中, 删除排序前 20 的节点, 依据 HITs 权威值和改进算法的排序结果造成极大弱连通率的变化是很相似的, 但是删除前 20 个节点到前 50 个节点时, λ 取值 0.65、0.75、0.85、0.95 的改进算法的排序结果造成极大弱连通率的下降都要比其他算法快。但在 Physicians 网络, 改进算法的极大弱连通率比其他算法高, 改进效果略次于其他三个网络。但总体上, 改进算法的排序结果表现出更高的准确性。

与强连通性评估指标相同, 在评判网络的连通性时也要考虑符合弱连通性子团的数量, 弱连通分量数越多, 说明网络破碎的越厉害, 算法的效果越好。因为 λ 取值比较多, 图例不能很好的在图形中显示, 图 5 每种颜色和顺序的条形图表示的算法与图 3 一致。在 Enzymes 网络、Celegans 网络和 Friendship 网络中, λ 大部分取值的排序结果形成的弱连通分量数多于其他算法, 算法表现出较高的准确性。Physicians 网络只有在在删除前 20 和 30 个节点时, 排序效果与经典算法相似, 但删除前 10、40 或前 50 个节点时, 改进的算法又显示出较好的准确性。

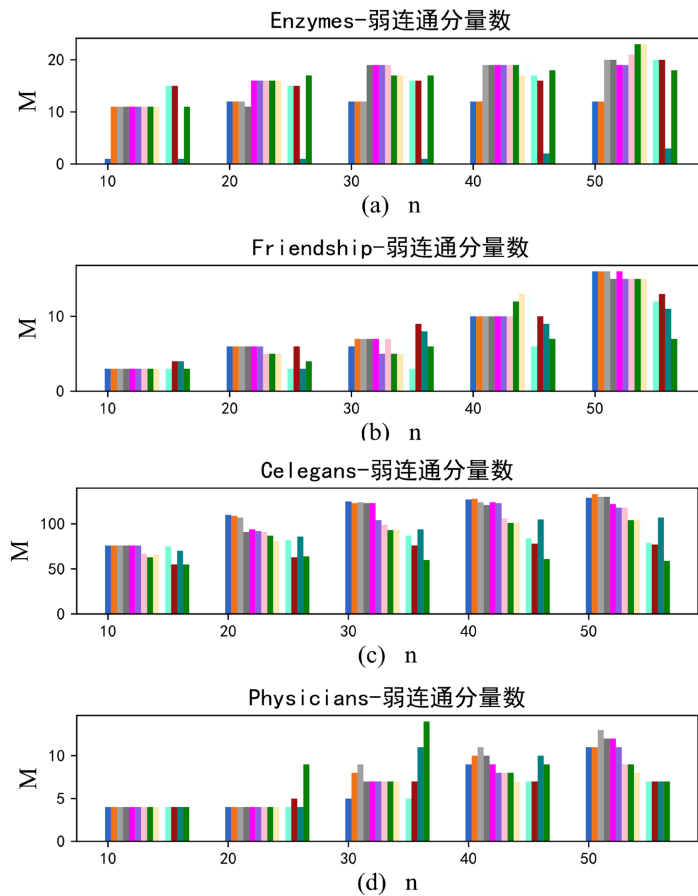


Figure 5. Change of the number of subnetworks with different node numbers removed
图 5. 移除不同节点数子网络数量的变化情况

总体来看, 改进算法在极大弱连通率和弱连通分量数两个评估指标中都表现出较高的准确性, 且大部分的 λ 取值的改进算法都表现出较好的准确性。

3.4. 脆弱性评估指标

使用网络的连通性衡量网络中节点失效或者遭受攻击后引起的整个网络连通度的下降, 也可以使用脆弱性衡量网络中节点失效或者遭受攻击后引起的整个网络系统性能的下降。节点脆弱性系数[7] V 是网络效率的变化情况, 计算公式如下:

$$V = \frac{E - E_i}{E} \quad (11)$$

其中, E 表示未删除节点时的网络效率, i 表示网络中的节点, E_i 表示移除节点 v_i 后的网络效率。

重要节点之所以能够在网络结构方面发挥关键性作用, 不仅是节点度数的优势, 也是与网络中其余节点间存在的拓扑关系。移除重要节点必定给网络结构造成更大的伤害, 这可以通过节点脆弱系数来表现。脆弱性系数越大, 表示节点移除后对网络的破坏力度越大。图 6 比较了经典算法与改进算法的节点脆弱系数的实验效果, 横坐标 n 表示的是移除节点的个数, 纵坐标 $p(n)$ 表示的是节点脆弱系数值。节点脆弱系数值随移除节点数 n 的变化在四个网络中, 都证实了改进算法表现出更高的准确性和有效性。大部分的 λ 取值的改进算法都表现出较好的准确性。

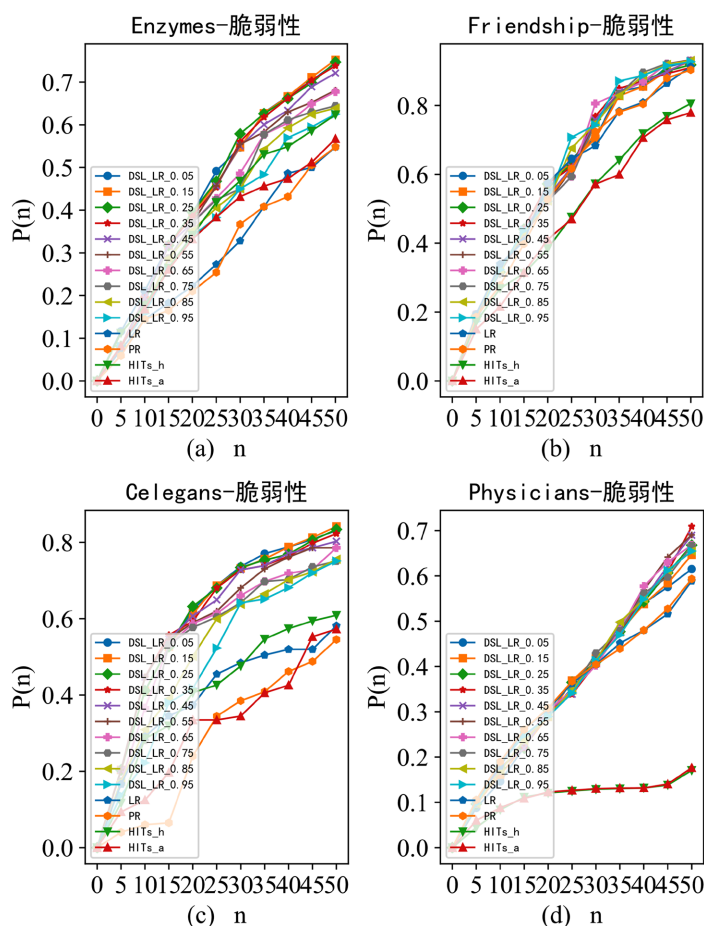


Figure 6. Change of vulnerability coefficient after removing different nodes
图 6. 移除不同节点脆弱系数的变化情况

4. 结束语

将节点自身属性、节点间的相似性和节点间的相对路径比到引力模型中, 构造引力矩阵, 再通过行归一化构造转移矩阵, 利用 LeaderRank 算法中的方法构造强连通网络, 保证稳态分布。在利用节点的综合度计算节点自身属性时, 调节出度和入度的参数 λ , 因此, 在验证改进算法比其他经典算法有更高准确性时, 也确定最优的参数 λ 。研究表明, 相比较 PageRank 算法、leaderRank 算法和 HITs 算法, 改进的算法具有更高的准确性。通过观察四个真实网络在三种评估指标中的结果可知, 当 λ 取值 0.25 到 0.65 时, 四个网络的强连通分量数和弱连通分量数几乎是最高的, 并且极大强连通率和极大弱连通率的变化幅度也比较大; 节点脆弱系数也比较高。因此, λ 取值 0.25 到 0.65 时, 改进的算法表现出更高的准确性。

基金项目

全国统计科学研究项目(No. 2020LY080)。

参考文献

- [1] 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: 1-23.
- [2] 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: 1-100.
- [3] 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., *et al.* (2014) The Structure and Dynamics of Multilayer Networks. *Physics Reports*, **544**, 1-122. <https://doi.org/10.1016/j.physrep.2014.07.001>
- [5] 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.
- [6] 赫南, 李德毅, 淦文燕, 朱熙. 复杂网络中重要性节点发掘综述[J]. 计算机科学, 2007, 34(12): 1-5.
- [7] 段杰明, 尚明生, 蔡世民. 基于自规避随机游走的节点排序算法[J]. 物理学报, 2015, 64(20): 1-8.
- [8] 王凯莉, 邬春学, 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): 235-245.
- [9] 马媛媛, 韩华, 瞿倩倩. 基于节点亲密度的重要性评估算法[J]. 计算机科学, 2021, 48(5): 140-146.
- [10] 张宪立, 唐建新. 一种新的复杂网络节点重要性评估方法[J]. 计算机工程, 2021, 47(2): 139-151.
- [11] 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.
- [12] 李秋晖, 韩华, 马媛媛, 等. 基于引力模型及相对路径数的节点重要度评估算法[J]. 计算机应用研究, 2022, 39(3): 764-769.
- [13] 何建军, 李仁发. 改进的随机游走模型节点排序方[J]. 计算机工程与应用, 2011, 47(12): 87-89.
- [14] 何建军. 复杂网络节点重要性评价研究[D]: [硕士学位论文]. 长沙: 湖南大学, 2010.
- [15] Janaswamy, R. (2003) Path Loss Predictions in the Presence of Buildings on Flat Terrain: A 3-D Vector Parabolic Equation Approach. *IEEE Transactions on Antennas & Propagation*, **51**, 1716-1728. <https://doi.org/10.1109/TAP.2003.815415>
- [16] 王林, 张一帆. 基于节点相似度的有向网络社团检测算法[J]. 软件与算法, 2017, 36(3): 19-22.
- [17] Li, Q., Zhou, T., Lv, L.Y., *et al.* (2014) Identifying Influential Spreaders by Weighted Leader Rank. *Physica A: Statistical Mechanics and Its Applications*, **404**, 47-55. <https://doi.org/10.1016/j.physa.2014.02.041>
- [18] 孙连, 李书琴, 刘斌. 基于加权 LeaderRank 的用户社交网络排序算法[J]. 计算机工程, 2019, 45(10): 196-202.
- [19] 刘建国, 任卓明, 郭强, 等. 复杂网络中节点重要性排序的研究进展[J]. 物理学报, 2013, 62(17): 1-10.
- [20] ENZYMES-G1. <https://networkrepository.com/ENZYMES-g1.php>
- [21] DATASET: High School Contact and Friendship Networks. <http://www.sociopatterns.org/datasets/high-school-contact-and-friendship-networks>
- [22] Datasets Released for Reproducibility. <https://manliodomenico.com/data.php>
- [23] 陈燕, 江克勤. 求强连通分量的几种算法的实现与分析[J]. 电脑知识与技术, 2011, 7(9): 2140-2142.
- [24] 曹雁锋, 张先伟. 一种强连通判定算法[J]. 计算机应用与软件, 2007, 24(4): 152-153.
- [25] 王小霞, 姜金平. L-双拓扑空间中几类弱连通性及其等价刻画[J]. 河南科学, 2013, 31(9): 1343-1345.
- [26] 王金山, 任蓓. 拓扑空间中的弱连通集[J]. 安徽机电学院学报, 2002, 17(2): 64-66.