

基于网络结构和信息传播的节点重要性识别方法

赵宇臣, 郭进利

(上海理工大学 管理学院, 上海 200093)

摘要:节点重要性评估是复杂网络研究的重要问题,对研究复杂网络的脆弱性和鲁棒性具有重要意义。通过考虑网络拓扑结构和信息传播,对局部网络间的拓扑结构进行量化,提出一种考虑节点度以及邻居节点的结构相似性的节点重要度评估 LSC 算法。基于 6 个经典实际网络和两个人工网络,以蓄意破坏的方式对网络进行攻击,通过极大联通系数、网络效率和网络节点单调性 3 个评估指标的实验结果对比,证明所提算法 LSC 相比于其他已有算法更能准确地评估网络节点的重要性。

关键词:复杂网络;鲁棒性;节点重要性;信息传播

中图分类号:TP393 **文献标志码:**A **文章编号:**1671-1807(2022)05-0326-06

随着科学信息技术的高速发展,复杂网络理论已经在人类社会中各个复杂系统取得巨大研究进展,人类的日常生活也越来越依靠复杂网络系统的稳定性^[1]。复杂网络的不同节点在网络中承担功能不尽相同,其中网络中某些重要节点对网络结构有较大影响,甚至对网络结构的稳定性和鲁棒性起到决定性作用^[2]。所以对网络重要节点的挖掘和排序不仅具有重要的理论价值,还具有显著的实践价值,如网络中社区发现、病毒疾病发现与控制等。

近些年来,关于如何对复杂网络中的重要节点进行识别和排序,目前国内外研究人员发展了很多方法,根据特点不同可以将其分为以下 4 类:

1) 基于邻居节点的排序方法。根据节点本身拓扑结构性质的度中心性(degree centrality, DC)进行排序是一种简单有效的局部排序算法^[3];根据节点中心性的 H 指数(H-index centrality, H)进行排序以及将网络进行分解通过节点的 K 核分解中心性进行排序(K-Shell decomposition centrality, KS)^[4]。

2) 基于路径的排序方法。根据一个节点位于其他节点最短路径上的介数中心性(betweenness centrality, BC)进行排序^[5];根据一个节点到其他节点距离之和的倒数的紧密中心(closeness centrality, CC)进行排序^[6]。

3) 基于特征向量的排序方法。根据一个节点

的邻居节点的数量和重要性,即节点的特征向量中心性(eigenvector centrality, EC)^[7]进行排序。

4) 基于节点删除和收缩的排序方法。通过判断删除网络中某个节点后对网络结构和功能的负向影响,从而反映节点对于网络的重要程度。(复杂网络中节点重要度评估的节点收缩方法)。

上述的节点重要性评价指标主要是基于节点自身性质来判断的,且判断的维度较为单一,无法普适于各种类型的网络之中。基于此,国内学者综合考虑这些因素,根据具体的研究背景,对现有的节点重要性评价标准进行改进。Gao 等根据节点及其邻居节点的性质,通过判断节点局部结构中心性进行排序^[8];Kitsak 等通过分析传播动力学中网络 SIR 模型,根据节点在网络中的传播效率判断节点的重要程度^[9];Wei 等利用马尔科夫链分析节点在网络中的动态性质,判断节点在网络中的重要程度^[10]。还有学者将多属性决策理论应用在网络节点重要性的判断中。Yang 等利用 Topsis 方法,综合评价单一指标的权重,提升节点重要性判断的准确性^[11];Lin 等通过 D-S 方法综合考虑多个中心性指标,提出了一种全新赋值的方法,判断节点的影响力大小^[12]。近些年来,许多学者提出基于邻居节点的结构特征的节点重要性排序算法。Gao 等通过分析节点邻域节点的联通子图数,判断节点的影响力^[13];Bea 等通过节点及其邻节点的重要性,进行

收稿日期:2021-10-19

基金项目:国家自然科学基金(71571119)。

作者简介:赵宇臣(1995—),男,河南郑州人,上海理工大学管理学院,硕士研究生,研究方向为复杂网络;郭进利(1960—),男,陕西西安人,上海理工大学管理学院,教授,博士,研究方向为网络科学与人类行为动力学。

过节点的网络边缘度值分析,判断节点重要性^[14]; Zareie 等通过分析邻节点结构特征的多样性,大大提升了节点重要度排序的准确性^[15]; Wang 等利用网络中节点与邻居节点的壳值,提出了多阶邻居壳数向量中心性算法,表示节点的重要性^[16]。

目前基于网络局部信息的节点重要性评估主要是针对节点的本身属性研究的,未充分考虑邻居节点的结构特征对节点重要性的影响。且现实生活中的大规模网络结构会随时间的迁移而变化,获取完整结果网络结构数据信息较为困难。因此通过全局网络结构信息定义节点的影响力具有一定的局限性。本文通过量化网络局部信息,基于网络节点及其邻节点的拓扑结构特征,提出一种新的网络节点重要性评估算法 LSC。并在 4 个实际网络和两个人工网络中进行实验验证,证明此算法能更加准确地评估节点的重要性。

$$\text{sim}(a,b) = \begin{cases} \frac{|N(a) \cap N(b)|}{\sqrt{(|N(a)| \times |N(b)|)}}, & \text{如果节点 } a \text{ 和 } b \text{ 不存在连边} \\ 1, & \text{如果节点 } a \text{ 和 } b \text{ 存在连边} \end{cases} \quad (1)$$

Salotn 指标与 Jaccard 指标相比优势在于,当处理节点的共同邻域时,Jaccard 指标没有考虑到每个共同邻域节点在量化过程中计算权值是不一样的,但在实际网络中,不同节点的权值会对网络重要节点的评价产生误差。所以在此基础上,引入 Salotn 系数作为判断网络邻域节点的相似性指标更加合理。

首先,在实际网络中节点的邻域节点数量越多,且邻域节点间的拓扑结构重合性越小,表明邻域节点的相似性越低,节点对于邻域节点的影响力越大,对整个网络来说更加重要。基于以上分析,提出一种基于节点本身度数和邻域节点相似性的节点重要性评价算法 LSC(i),表示为

$$\text{sim}(a,b) = \begin{cases} \frac{|N(a) \cap N(b)|}{\sqrt{(|N(a)| \times |N(b)|)}} & (2) \\ 1, \text{其他} \end{cases}$$

$$\text{LSC}(i) = \sum_{a,b \in N(i)} [1 - \text{sim}(a,b)] \quad (3)$$

其次,利用信息传播相关理论对 LSC 算法进行修正,当节点 i 的 LSC 值大于 0 时,节点 i 和其邻节点 j 之间的信息传播才有意义,故保留其 LSC 值;当 LSC 值小于或等于 0 时,LSC 取 0 值。例如在 SIR 模型中,当传播值大于两节点之间的阈值时,才可以进行传播;当传播值小于或等于两节点之间的阈值时,不进行传播。基于这点增加了 $f(X)$ 函数对 LSC 算法进行修正来保证算法的准确性,计算公式为

$$\text{LSC}(i) = \sum_{j \in N(i)} f[\text{LSC}(i) - \text{LSC}(j)] \quad (4)$$

1 算法改进

节点对网络的重要性程度不仅取决于节点本身的特性,还高度依赖于邻节点的拓扑结构。目前很多节点重要性的评价方法仅是在节点某些特征上进行拓展,研究其对网络的影响,忽略了网络中邻节点在结构上的差异。考虑到邻域节点的拓扑结构对节点影响力的影响,通过量化网络结构局部信息,提出了节点的邻域相似性,当邻节点的邻域相似性越低,网络对于节点的依赖性越高,节点对网络的影响力越大。基于以上分析,本文采用 Salotn 指标^[17]值定义邻居节点的相似性,即 $\text{sim}(a,b) = |N(a) \cap N(b)| / \sqrt{(|N(a)| \times |N(b)|)}$ 。如果邻居节点 a 和 b 之间不存在连边时, a 与 b 的相似性定义为 Salotn 指标值;如果 a 和 b 之间存在连边时,定义 $\text{sim}(a,b) = 1$,即

$$f(X) = \begin{cases} X, X > 0 \\ 0, X \leq 0 \end{cases} \quad (5)$$

综上分析,算法 LSC 综合考虑了节点本身度数和邻居节点的拓扑结构,并利用信息传播对节点的 LSC 值进行修正,当节点 i 的 LSC 值越大时, i 的度数越大且与其邻居节点相似性越小,即与邻节点的拓扑结构重合性越小,说明节点 i 对网络的重要性越高;反之,当节点 i 的 LSC 值越小时,对网络的重要性越低。可以利用节点的 LSC 值的大小对网络中节点的重要性进行判断。

2 评价标准

为验证算法的可行性,本文采用 3 种方法对节点重要性排序的算法进行度量。

2.1 网络效率法^[18-19]

通过移除利用算法识别出的重要节点,模仿对网络中重要节点进行蓄意攻击,比较移除节点前后对网络效率的影响。网络效率可以用来评价网络的联通性,当某些节点及其连接边被移除时,使得其他节点间最短路径变大,进而影响整个网络的平均路径长度。当网络效率变化越大,说明节点越重要;反之,说明节点越不重要。网络效率可以表示为

$$\eta = \frac{1}{N(N-1)} \sum_{i,j \in V} \eta_{ij} \quad (6)$$

$$\mu = \frac{\eta}{\eta_0} \quad (7)$$

式中： N 表示网络节点数； $\eta_{ij} = \frac{1}{d_{ij}}$ ， d_{ij} 表示节点 i 与节点 j 之间的最短距离，通过移除识别出的重要节点，模拟蓄意攻击网络节点的仿真效果； η_0 表示网络未受到攻击前的网络效率； η 表示移除节点后的网络效率，使用 $\mu = \eta/\eta_0$ 网络效率变化进行规范化处理，比值越小，说明移除的节点重要性更高。

2.2 极大联通系数法^[20]

按照节点重要度评价指标，将节点按重要度从大到小进行排序。移除一部分节点后，观察对网络极大连通子集的影响，计算公式为

$$G = R/N \tag{8}$$

式中： R 表示移除一部分节点后的网络极大连通子集的节点数； N 表示网络中节点总数。极大连通子集规模随着节点移除而变小的趋势越明显，表明采用该方式攻击网络节点的效果越好，识别出的节点对于网络的重要性越高。

2.3 单调性

节点单调性是区分节点传播能力和等级的重要指标之一，反映了在某指标下节点的得分情况，单调性指标值越大，节点获得相同得分的数量越少，算法的效果就越好，计算公式为

$$M(R) = \left[1 - \frac{\sum_{r \in R} n_r(n_r - 1)}{n(n - 1)} \right]^2 \tag{9}$$

式中： n 表示序列 R 中的节点数量； n_r 表示在等级 r 上的节点数量； M 的取值范围为 $[0, 1]$ ，其数值较大表示该方法具有较高的评估准确率。

3 实验数据收集

考虑到不同网络具有不同的拓扑结构，本文共使用 4 个真实网络数据集和 2 个人工网络数据集进行仿真实验。真实网络包括 karate 美国空手道俱乐部、Dolphins 海豚网络、Adjnoun 美国空军局域网络和 Celegansneural 网络；人工网络包括 WS 小世界网络 (WS) 和 BA 无标度网络，人工网络均通过 Python 软件自动生成，其中 BA 的参数为 $N=500$ ， $M=5$ ，WS 的参数为 $N=500$ ， $K=5$ ， $P=0.5$ ^[21-22]。4 个真实网络和 2 个人工网络的拓扑特性见表 1。

表 1 4 个真实网络和 2 个人工网络的拓扑特性

网络	节点	边	最大度	最小度	平均度
Karate	34	78	17	1	4.588
Dolphins	62	159	12	1	5.129
Adjnoun	112	425	49	1	7.589
Celegansneural	297	2 148	134	1	14.465
WS	500	1 000	9	2	4.0
BA	500	2 475	88	5	9.9

4 实验结果分析

为评价本文提出重要性节点评价算法 LSC 的性能，进行以下实验：①以蓄意攻击的方法移除一定比例 P 排名靠前的节点，模拟网络遭受蓄意攻击的极大连通子图规模和网络效率的变化，从而评价各个节点重要性排序的方法；②将 LSC 算法与基于局部相似性的 LLS 算法、基于节点位置信息的 K-shell 算法、Local Centrality 算法和同样采用局部信息的度的排序 Degree 算法进行单调性分析与比较，利用单调性检验不同中心性序列对于节点重要性的区分能力。

4.1 蓄意攻击效果

在蓄意攻击网络结构对网络极大连通系数 G 影响的实验中，通过采用 K-shell 指标、LC 指标、Degree 指标、LLS 指标和本文提出的 LSC 指标对 6 个真实网络和 2 个人工网络中排名靠前的节点进行移除，对网络影响结果如图 1 所示。在所有的 8 个网络中，从实验结果的整体趋势来看，LSC 指标导致网络的极大连通系数变小的走势最为明显，且 LSC 指标总是以最少的移除节点比例就能最先使网络的极大连通系数降到最低，说明综合考虑网络局部拓扑结构相似性和信息传播理论的网络节点识别算法 (LSC) 能更好地识别出关键节点。尤其在 实际网络 [图 1(b)] 和人工网络 [图 1(e)] 中表现得更为明显，LSC 指标在蓄意攻击过程中对网络极大连通影响明显优于其他算法指标。比较 LSC 指标和 LLS 指标，从 6 个网络实验结果来看，当移除相同比例的节点时，在多数情况下 LSC 指标可以使网络的连通性下降更大的幅度，说明 LSC 指标通过对算法的优化，相比于 LLS 指标在网络重要节点识别的准确性上有较大的提升。

本文采用网络效率比较 K 指标、LC 指标、D 指标、LLS 指标和本文提出的 LSC 指标的性能。利用上述的节点重要性评价指标删除一定比例排序靠前的节点，网络效率下降率 μ 随着网络连通性的变化而变化，连通性越差网络效率的下降趋势越明显。图 2 反映了不同节点重要性指标下删除一定比例节点后，网络效率的变化情况。由图 2 可知，随着网络节点移除比例的增加，LSC 指标导致网络效率下降的幅度最为显著，明显优于其他指标。特别是在图 2(d) Karate 网络中，从移除节点开始网络效率下降的幅度高于其他指标，一直保持到网络效率下降到 0，且随着移除节点的增加网络效率下降趋势

并没有出现负向波动,说明通过 LSC 算法识别的关键节点更加准确。

综上所述,利用 LSC 算法的节点排序方法破坏

网络拓扑结构时,网络效率和网络的极大连通系数变化得最为显著,说明 LSC 算法可以更加准确地识别出网络中的重要节点。

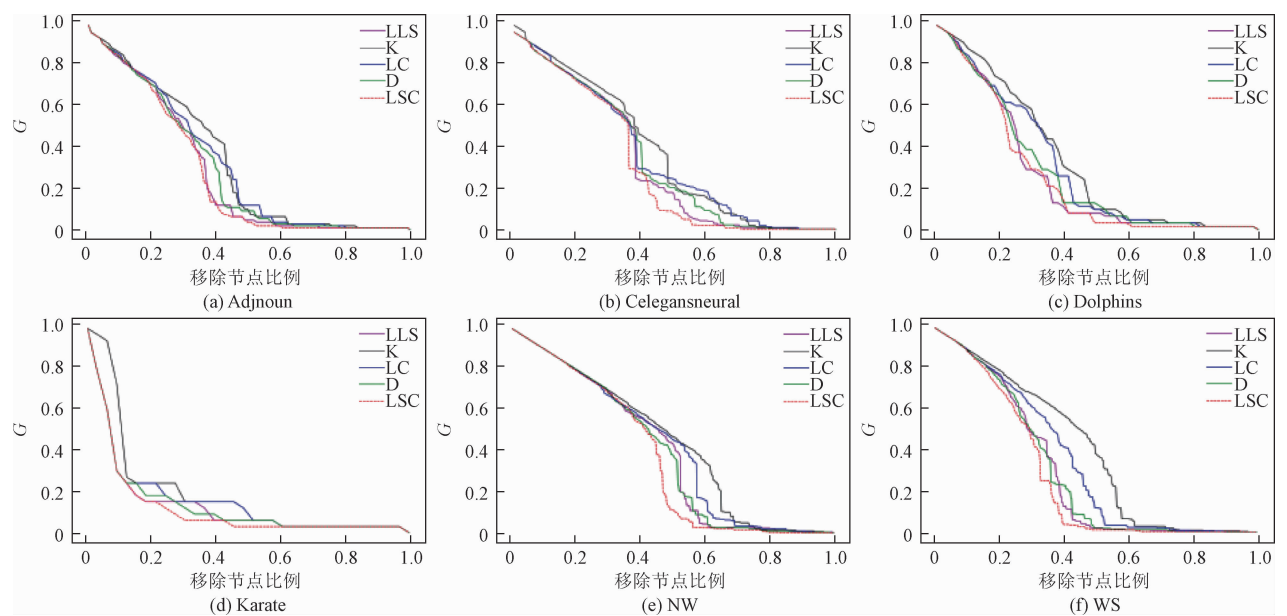


图 1 利用不同指标攻击网络重要节点后极大连通系数的变化

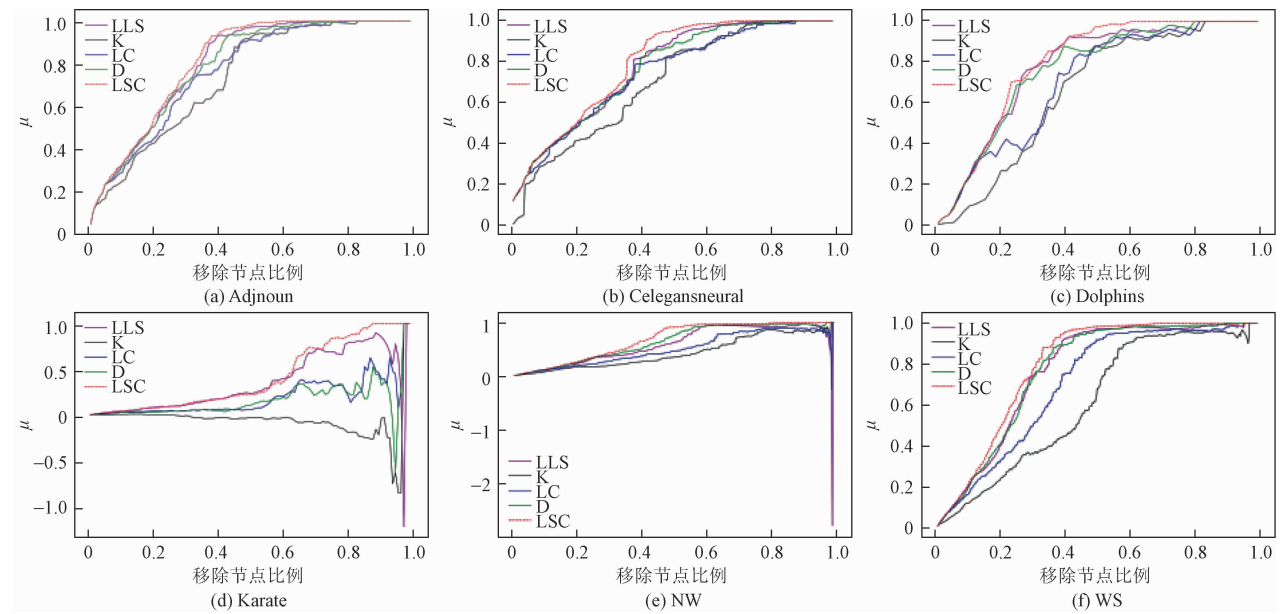


图 2 利用不同指标攻击网络重要节点后网络效率的变化

4.2 单调性

不同指标在各网络中得到网络节点单调性值见表 2。表 2 结果表明,LSC 算法除了在 Karate 网络中关键节点的单调性 LC 算法和 Degree 算法外,在其他网络中单调性均高于其他 4 种算法,单调性反映了在某指标下节点的得分情况,单调性指标值越大,节点重要性排序有较好的差异性,算法的效

果就越好。与 LLS 算法相比,LSC 算法在 4 个真实网络中的重要节点的单调性数据均优于 LLS 算法,说明可以更加准确地评估节点的重要度。

5 结语

复杂网络关键节点的准确识别具有重要的理论意义和实践价值,不仅对加快有益信息在网络中的传播、抑制病毒的爆发具有现实意义,还有助于提

表 2 不同指标在不同网络数据上的单调性值

网络	M(D)	M(LC)	M(K)	M(LLS)	M(LSC)
Karate	0.707 9	0.882 5	0.936 9	0.772 3	0.839 5
Dolphins	0.831 2	0.963 7	0.976 9	0.962 3	0.967 3
Adjnoun	0.866 1	0.977 4	0.997 4	0.978 9	0.985 3
Celegansneural	0.921 7	0.994 0	0.995 8	0.993 1	0.995 1

升网络的抗毁性和稳定性。在已有研究的基础上，本文综合考虑节点度以及邻居节点的结构相似性，利用信息传播的性质对算法进行修正，提出了基于网络节点结构和信息传播的复杂网络节点识别的新方法，并在 4 个真实网络和 2 个人工网络中进行仿真实验，验证算法的可行性。实验表明 LSC 指标在网络连通性、网络效率及节点的单调性 3 个方面均优于其他比较算法，说明 LSC 算法可以更好地识别出网络中的重要节点。同时也希望为后续的研究提供一定的参考。本研究所用的网络均为无向无权网络，如何将此方法应用于有向加权网络以及据时态变化的网络是后续的研究重点。

参考文献

[1] BARABA'SI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286: 509-512.

[2] SHIRAZ M, GANI A, KHOKHAR RH, et al. A review on distributed application processing frameworks in smart mobile devices for mobile cloud computing[J]. IEEE Communications Surveys & Tutorials, 2013, 15 (3): 1294-1313.

[3] BONACICH P F. Factoring and weighting approaches to status scores and clique identification[J]. Journal Mathematical Sociology, 1972, 2: 113 - 120.

[4] KITSACK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. Nature Physics, 2010, 6(11): 888-893.

[5] FREEMAN L C. A set of measures of centrality based on betweenness[J]. Sociometry, 1977, 40(1): 35-41.

[6] SABIDUSSI G. The centrality index of a graph[J]. Psychometrika, 1966, 31(4): 581-603.

[7] BONACICH P. Factoring and weighting approaches to status scores and clique identification [J]. Journal of Mathematical Sociology, 1972, 2(1): 113-120; 1-1.

[8] GAO S, MA J, CHEN Z M, et al. Ranking the spreading ability of nodes in complex networks based on local structure[J]. Physica a; Statistical Mechanics and ITS Applications, 2014, 403: 130-147.

[9] KITSACK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. Nature Physics, 2010, 6: 888-893.

[10] WEI D, DENG X, ZHANG X, et al. Identifying influential nodes in weighted networks based on evidence theory [J]. Physica A, 2013, 392: 2564-2575.

[11] YANG L, CAO J N, YUAN Y, et al. A framework for partitioning and execution of data stream applications in mobile cloud computing[J]. ACM SIGMETRICS Performance Evaluation Review, 2013, 40(4): 23-32.

[12] LIN X, WANG Y Z, XIE Q, et al. Task scheduling with dynamic voltage and frequency scaling for energy minimization in the mobile cloud computing environment[J]. IEEE Trans on Services Computing, 2015, 8 (2): 175-186.

[13] GAO C, WEI D, HU Y, et al. A modified evidential methodology of identifying influential nodes in weighted networks[J]. Physica A, 2013, 392: 5490-5500.

[14] BAE J, KIM S. Identifying and ranking influential spreaders in complex networks by neighborhood coresnes [J]. Physica A: Statistical Mechanics and ITS Applications, 2014, 395: 549-559.

[15] ZAREIE A, SHEIKHAHMADI A, JALILI M, et al. Influential node ranking in social networks based on neighborhood diversity[J]. Future Generation Computer Systems, 2019, 94: 120-129.

[16] WANG K, WU C X, AI J. Vector centrality measurement method based on multi-order neighborhood shell number [J]. Acta Physica Sinica, 2019, 68(19): 196402.

[17] RESTREPO J G, OTT E, HUNT B R. Characterizing the dynamical importance of network nodes and links [J]. Physical Review Letters, 2006, 97: 094102

[18] 叶青. 基于复杂网络理论的轨道交通网络脆弱性分析 [J]. 中国安全科学学报, 2012, 22(2): 122-126.

[19] ZHANG Y J, AYYUB B M, SAADAT Y, et al. A double-weighted vulnerability assessment model for metrorail transit networks and its application in Shanghai metro [J]. Future Generation Computer Systems, 2020, 29: 100358.

[20] YAN G, ZHOU T, HU B, et al. Efficient routing on complex networks [J]. Physical Review E, 2006, 73: 046108.

[21] NEWMAN M E J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality[J]. Physical Review E, 2001, 64: 016132.

[22] WATTS D J, STROGATZ S H. Collective dynamics of small-world networks[J]. Nature, 1998, 393: 440-442.

Node Importance Identification Method Based on Network Structure and
Information Dissemination

ZHAO Yuchen, GUO Jinli

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

Abstract: Node importance assessment is an important issue in complex network research, and it is important to study the vulnerability and robustness of complex networks. By considering the network topology and information dissemination, and quantifying the topology among local networks, a node importance evaluation LSC algorithm that considers the degree of nodes and the structural similarity of neighbor nodes was presented. Based on six classical real networks and two man-made networks, the network is attacked by means of intentional destruction. The experimental results of three evaluation indicators, i. e. , maximum connectivity factor, network efficiency and network monotonicity, show that the proposed algorithm LSC can more accurately evaluate the importance of network nodes than other existing algorithms.

Keywords: complex networks; robustness; node importance; information dissemination