引用格式：郭强, 欧阳, 江明珠, 等. 社交网络节点重要性识别研究进展 [J]. 电子科技大学学报, 2025, 54(1): 125-151.
GUO Q, OU Y, JIANG M Z, et al. Review of social network spreading influence nodes identification[J]. Journal of University of Electronic Science and Technology of China, 2025, 54(1): 125-151.

社交网络节点重要性识别研究进展

郭 强[1]，欧 阳[1,2]，江明珠[3]，刘建国[4*]

（1. 上海理工大学 管理学院，上海 200093；2. 上海哔哩哔哩科技有限公司 人工智能平台部，上海 200433；
3. 上海大学 管理学院，上海 200444；4. 上海财经大学 数字经济系，上海 200433）

摘要：准确识别社交网络中的节点重要性对于促进或抑制信息传播、遏制疾病传播具有重要意义，同时在精准营销和社会治理等领域也具有重要理论意义和应用价值。该文从 4 个角度对节点影响力识别算法进行总结和梳理，具体包括：基于微观局部结构、中观的社团结构、宏观全局结构及基于机器学习的算法。详细介绍了其中的代表性算法，并从不同层面分析了不同算法的优缺点。此外还总结了常用的传播动力学模型和评价指标。最后提炼了仍需解决的问题和未来可能的研究方向。

关键词：社交网络；节点重要性；社团结构；机器学习
中图分类号：TP311; N94 **文献标志码：**A DOI: 10.12178/1001-0548.2022067

Review of social network spreading influence nodes identification

GUO Qiang[1], OU Yang[1,2], JIANG Mingzhu[3], and LIU Jianguo[4*]

(1. Business School, University of Shanghai for Science & Technology, Shanghai 200093, China;
2. AI Platform, Shanghai Bilibili Technology Co., Ltd., Shanghai 200433, China; 3. Business School, Shanghai University, Shanghai 200444, China;
4. Department of Digital Economics, Shanghai University of Finance & Economics, Shanghai 200433, China)

Abstract: The identification of spreading influence node in social networks aims to uncover individuals or groups that can effectively promote information dissemination or have a significant impact on the network structure, which is helpful for deeply understanding the features of important nodes and their applications in the targeted marketing, rumor containment and so on. This review categorizes existing spreading influence node identification algorithms into four categories: Micro-structure-based (MI), mesoscopic-structure-based (ME), macro-structure-based (MA) and machine-learning-based (ML). It provides a detailed introduction to representative algorithms and analyzes the advantages and disadvantages of each type from different perspectives. Additionally, this review summarizes the commonly used propagation dynamics models and evaluation metrics in this research direction, and finally highlights urgent issues that need to be addressed and potential future research directions.

Key words: social networks; node importance; community structure; machine learning

社会系统中个体或群体之间错综复杂的交互关系可由社交网络抽象表示[1-2]。社交网络已成为信息传播[3-4]、传染病扩散[5-6]、链路预测[7-8]、在线用户声誉评估[9-10] 等领域的主要研究对象。节点重要性识别作为社交网络分析的研究热点之一，致力于发展社交网络节点影响力度量的理论与算法。由于社交网络中节点所处的位置和结构差异，各节点在网络中所起的作用不同[11-12]。因此，识别出社交网络的重要节点对于精准营销[13]、遏制舆情[14]、虚假信息验证[15] 具有重要意义。

由于社交网络的类型多样且规模庞大，加之时间和资源上的局限性。通过社会实验来衡量出每个节点的重要性是不切实际的。因此，研究者更倾向于利用节点属性与网络的结构属性来估计节点的重要性[16-17]。现有文献从不同角度对这一研究方向进行了回顾和梳理，文献 [18] 从网络拓扑结构和传播动力学的视角分类总结了节点重要性识别算法，并对各种算法的优缺点及适用场景进行了系统的梳

收稿日期：2022−03−07
基金项目：国家自然科学基金（72171150, 72371150）；中央高校基本科研业务费专项（2023110139）
作者简介：郭强，教授，主要从事大数据分析、社交网络、科学知识图谱分析等方面的研究。
*通信作者 E-mail: liujg004@ustc.edu.cn

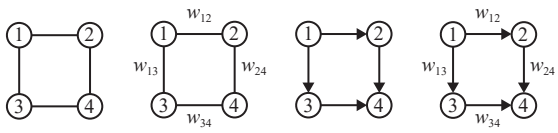
理。文献 [19] 对 30 余种具有代表性的节点重要性识别算法进行了综述，文献 [12] 详细地对节点重要性识别算法及评价指标进行了分类介绍，并在不同类型的网络上分别对比了各种方法的表现。文献 [20] 重点梳理了在社会网络分析中应用的节点重要性识别算法，文献 [21] 介绍了在中心性[22-23]、PageRank[24] 与 HITS[25] 基础上进行扩展和改进的算法。除了基于静态网络的算法外，由于时序网络能够更好地刻画复杂系统的动态特性，针对时序网络提出的节点重要性识别算法也越来越多[26-28]。文献 [29] 介绍了时序网络建模方法，并从网络拓扑结构、随机游走动力学以及机器学习这 3 个角度对时序网络节点重要性识别算法进行了梳理。文献 [30] 总结了增长网络、实时动态网络以及结构微扰或突变的时序网络中节点重要性识别算法面临的问题和挑战。

随着计算性能的不断改善、数据规模的爆炸式增长、跨学科合作的增加及需求的快速变化，近年来出现了众多融合新技术和新理念的算法。根据所使用的结构属性和算法特点，现有节点重要性识别算法可分为：基于微观局部结构（micro-structure-based, MI）、中观社团结构（mesoscopic-structure-based, ME）、宏观全局结构（macro-structure-based, MA）及基于机器学习（machine-learning-based, ML）4 类。具体地，为了提升基于微观局部结构（MI）算法的准确率，研究者不再局限于简单聚合邻居节点的结构信息，而是进一步考虑了局部网络中高阶邻居之间的关系[31-32]。社团结构属性能够反映出网络更深层的结构信息，有助于提高节点重要性识别的准确性并降低计算复杂度，逐渐成为越来越多研究者选择的切入点[33-35]。通过将宏观全局信息和微观局部信息进行结合以增强算法的泛化性能，这受到了越来越多学者的认可和关注[36-37]。此外，随着机器学习和网络科学研究的结合愈发紧密，越来越多基于机器学习（ML）的算法出现在节点重要性的研究中[38-39]。为了及时掌握节点的重要性研究进展及未来的发展方向，本文系统地介绍了当前具有代表性算法的细节及优缺点，并梳理了节点重要性研究中常用的传播动力学模型与评价指标。最后尝试提出这一研究方向仍需解决的问题和可能的研究方向。

1 相关定义

1.1 社交网络的定义

社交网络由节点和连边构成，其中节点代表个体或群体，连边则表示节点之间的关系。根据连边是否有权重和方向，社交网络可分为无向无权网络、无向有权网络、有向无权网络以及有向有权网络（如图 1 所示）。$G(V,E)$代表一个由$|V|=n$个节点与$|E|=m$条连边所构成的无权网络，$V=\{v_1, v_2,\cdots,v_n\}$为网络中的节点集合，$E=\{e_{ij}|i,j=1,2,\cdots,n\}$为网络中的连边集合，$\boldsymbol{A}=\{a_{ij}\}_{n\times n}$表示网络对应的邻接矩阵，当节点$i$与节点$j$相连时$a_{ij}=1$，否则$a_{ij}=0$。在无向网络中，连边$e_{12}=e_{21}$，在有向网络中此等式不成立。$G(V,E,W)$是一个由 n 个节点与 m 条连边构成的加权网络，其中 $\boldsymbol{W}$ 为连边权重矩阵，w_{ij}为连边e_{ij}的对应权重。



a. 无向无权网络 b. 无向无权网络 c. 有向无权网络 d. 有向无权网络

图 1　4 种社交网络图例

1.2 节点排序与影响力最大化问题

节点重要性识别问题可进一步细分为节点排序和影响力最大化。其中，节点排序是指用某种节点重要性识别函数$f(\cdot)$对网络中的所有节点进行重要性打分，根据得分对节点进行排序。影响力最大化是指当初始激活节点数为固定值k时，识别出k个节点作为种子节点集S，使得在种子节点集S作为传播源时所取得的影响力大于或等于由其他任意k个节点组成的种子节点集的影响力，其数学表示为：

$$S^*=\arg\max\{\sigma(S)\} \tag{1}$$

式中，$\sigma(S)$为节点集S的影响力。

为详细展示现有研究细节，本文首先在表 1 中列出了部分算法的科学问题、所用数据、主要发现等详细信息。

表 1　节点重要性识别算法名称、所属类别、科学问题、所用数据集、传播模型及主要发现

方法	类别	科学问题	所用数据集	传播模型	主要发现
Pbga[40]	MI	影响力最大化	GrQc[41], HepTh[41], Enron[42], DBLP[43], LiveJournal[43], QQ[44], Macau Weibo[45], NoLA Facebook[46]	SIR	节点重要性可以基于微观局部结构属性进行近似估计

续表

方法	类别	科学问题	所用数据集	传播模型	主要发现
Spreading strength (SS)[47]	MI	节点排序	Guntella08[41], GrQc[41], CondMat[41], HepTh[41], Facebook[48], PGP[49], Protein[49], PowerGrid[50], US Air[51]	SIR	节点对于其邻居节点的间接影响同样是反应节点影响力的重要指标
Local centrality (LC)[52]	MI	节点排序	Blog[53], Netscience[54], Router[55], Email[56]	SIR	高阶邻居度信息可以提高度中心性的准确性
Neighborhood centrality (NC)[57]	MI	节点排序	HepTh[41], PGP[49], Router[55], Email[56], Hamster[58], Astro Physics[59]	SIR	考虑更高阶邻居信息不一定会获得更好的性能，使用二阶邻居信息时准确率和效率最为平衡
Local structure similarity (LSS)[60]	MI	影响力最大化	GrQc[41], Routers[55], Hamster[58], Polblogs[61]	SIR, SI	利用局部结构属性能比距离指标更准确的识别多个重要节点
Vote rank[62]	MI	影响力最大化	CondMat[41], Berkstan[42], YouTube[43], Notre DAME[63]	SIR, SI	VoteRank的性能与已排序节点的数量相关
Cluster rank[64]	MI	节点排序	Delicious[45], SM	SIR	节点的局部集聚系数越小，节点未来的度值越大
Local structure centrality (LSC)[65]	MI	节点排序	PGP[49], Email[56], Twitter, Blog	SIR	节点重要性与二阶邻居集聚系数的正相关关系有助于识别节点重要性
V-community (Vc)[66]	ME	节点排序	GrQc[41], Facebook[48], Protein[50], Netscience[54]	SIR	考虑节点所连社团个数识别重要节点
Community-based centrality (cbc)[67]	ME	节点排序	Facebook[48], PowerGrid[50], Router[54], Metabolic, Email, Blogcatalog	SIR	社团的规模及邻居节点在各社团的分布是社团层面反应节点重要性的重要指标
Community-based mediator (cbm)[68]	ME	节点排序	Karate[69], American football network[70], Dolphin[71], Airport, Internet	SIR	考虑社团内部和社团之间的连边密度可以在较低计算复杂度基础上准确识别重要节点
Community-hole index (CHR)[72]	ME	节点排序	GrQc[40], Weibo[73], Arxiv[74], Amazon	SIR	节点所属社团的重要性度量节点的重要性
Omc[35]	ME	节点排序	GrQc[41], Facebook[48], Netscience[54]	SIR	通过将具有重叠社团结构的网络划分为局部网络和全局网络识别重要节点
Network global structure-based centrality (NGSC)[37]	MA	影响力最大化	Netscience[54], Advogato[75], Odlis[76]	SIR	网络连通片与网络密度对于节点重要性识别算法的性能具有重要影响
Gravity centrality (GC)[77]	MA	节点排序	HepTh[41], PGP[49], Blogs[53], Netscience[54], Router[55], Email[56], TAP[78], Y2H[78], Facebook	SIR	引力模型可用于节点重要性识别且获得较高的准确性
C_{effg}[79]	MA	节点排序	GrQc[41], Netscience[54], Jazz, EEC, Email, PB, Facebook, US Air, Physicians, PDZBase, Haggle, Infectious	SI	将引力中心性的欧式距离替换为有效距离可以提高准确性
Dynamic-sensitive (DS)[80]	MA	影响力最大化	Erdos, Protein[51], Router[55], Email contact[81]	SIR, SI	节点重要性不仅与结构属性相关，还与传播动力学相关
Influence capacity[82]	MA	节点排序	Enron[42], PGP[49], Blog[53], Netscience[54], Email[56], Karate[69], Dolphin[71], Jazz, Twitter, Facebook	SIR	K-核分解法中属于同一层节点被移除的顺序可以区分同层节点的重要性差异
Link entropy[83]	MA	节点排序	HepTh[41], PGP[49], Netscience[54], Router[55], Email[56], Hamster[58], Astro Physics[59], Email contact[81], AS	SIR	K-核分解法中位于网络核心层的节点与其他层节点连边多样性较高时该节点为重要节点
θ method[84]	MA	节点排序	P2P[41], PGP[49], Email[56], AS	SIR	K-核分解法中同层节点到核心层节点的距离能够区分同层节点的重要性差异
Multi-centrality predictors[85]	ML	节点排序	CondMat[41], GrQc[41], HepTh[41], PGP[49], PowerGrid[51], Astro Physics[59], Hamster[58], Advogato[75], Adolescent, AS, Brightkite, Email, Epinions, Euroroad, Facebook, GitHub, Guntella, Googleplus, IMDB, OpenFlights, Twitch, Twitter Stanford, US Airports, WikiTalk	SIR	MI算法易将处于局部连边密度较高但位于网络边缘层的节点识别为重要节点，而MA方法可以进行纠正
P&c[86]	ML	节点排序	Email[87], NetHEPT[88], Epinions, WikiVote	SIR	集成学习的思想可用于提高K-核分解与PageRank算法的泛化性能
Influence deep learning (IDL)[89]	ML	节点排序	Sina Weibo, Epinions, WikiVote	IC	图卷积神经网络在节点重要性识别上有较大的发展潜力

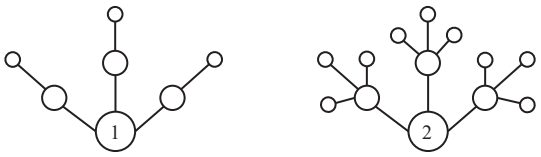
2 基于微观局部结构的算法

在信息爆炸时代，微博、推特等在线社交平台的用户规模庞大且交互关系复杂多变。直接利用全局结构属性识别节点重要性会耗费大量的计算资源与时间。文献 [40] 基于渗流理论[90] 揭示了传播过程的成核特性并通过实验验证了利用微观局部结构属性估计节点重要性的可行性。近年来，如何仅使用网络中的局部结构属性对节点进行重要性识别是重要研究方向[60, 91] 之一。

度中心性[22] 是一个经典的基于微观结构的 MI 算法。在无向网络中，度中心性的数学定义为：

$$\mathrm{DC}(i)=\frac{k_i}{n-1} \tag{2}$$

式中，n为网络中节点的总数；k_i表示节点i的度值。由式 (2) 可知，度中心性只考虑了目标节点的一阶邻居数量。但社交网络中的传播过程是链式的，一位用户将信息传播给朋友之后，该用户朋友的朋友（高阶邻居）也会有一定概率被这一信息所影响。在图 2 中，虽然节点 1 与 2 具有相同的度中心性值，但节点 2 的邻居的邻居数量（二阶邻居数量）远远大于节点 1，因此节点 1 与 2 之间的重要性实际上是有差异的，而度中心性却无法对其进行区分。



a. 节点 1　　b. 节点 2

图 2　度中心性值相同但二阶邻居数量不同的节点

为了克服这一缺陷，最直观的办法就是将邻居的邻居，甚至邻居的邻居的邻居等高阶邻居节点的度信息考虑进来。文献 [52] 提出了基于高阶邻居度的算法，实验发现该算法能够取得比度中心性更高的准确性。虽然高阶邻居的度信息有助于节点重要性识别，文献 [57] 分析了核数存在分辨率过低和难以识别并不具备中心地位的“假核团”的缺陷，提出了通过度量各壳层间连接的多样性来排除“假核团”的 role of neighbors （RN）算法，发现邻居节点的阶数与算法的准确性并非呈正相关关系。当涉及的邻居节点阶数超过 3 阶时，算法性能会达到饱和状态，即在计算复杂度上升的同时准确率不变，RN 算法的计算公式为：

$$\mathrm{NC}_i^l(a)=r_i+a\sum\nolimits_{j\in\Gamma_i}r_j+a^2\sum\nolimits_{z\in\Gamma_j\setminus i}r_Z+ a^{l-1}\sum\nolimits_{l\in\Gamma_{l-1}\setminus i}r_l \tag{3}$$

式中，r为中心性指标；a为 0～1 之间的自由参数；Γ_i为节点i的一阶邻居集合；l表示节点的第l阶邻居。由式 (3) 可知，算法假定邻居节点与目标点距离越近，其对目标节点重要性的贡献就越大，即邻居节点对于目标节点重要性的贡献存在差异。局部邻居贡献度算法（local neighbor contribution, LNC）[31] 将节点自身的重要性与邻居节点对该节点重要性的贡献两部分进行结合以区分相同度值节点的重要性差异。文献 [47] 认为即便是在无向网络中，节点i对于节点j的影响与节点j对i的影响不同。这种差异可以通过两个相连节点的非共同邻居节点数量来进行衡量。其原因在于：每个节点的邻居组成不同，当节点i影响了节点j之后，节点i就有机会通过节点j影响与节点i间接相连的节点；反过来，节点j影响了i之后，就有机会通过节点i影响与节点j间接相连的节点，节点i对于节点j的影响为：

$$c_{ij}=1+|e_{jl}|\left[1+\frac{|d_{i,j,2}|}{2^2}\right]^a \quad (l\notin\Gamma_i, l\neq i) \tag{4}$$

式中，$|d_{i,j,2}|$表示节点i与节点j之间长度为 2 的路径数量；a为自由参数；若节点i与节点j有连边，则$|e_{ij}|=1$，否则$|e_{ij}|=0$。

网络的微观局部结构属性除了节点度值外，还包括邻居节点之间的连接关系。邻居节点之间的连接关系能够有效反映出目标节点潜在的信息传播广度，对节点重要性识别同样具有重要意义。具体地，邻居节点之间的连接关系强弱可以通过局部聚类系数进行量化，其表达形式为：

$$c_i=\frac{2\left|\{e_{jv}|j,v\in\Gamma_i\}\right|}{k_i(k_i-1)} \tag{5}$$

式中，目标节点的度大于 1，即$k_i>1$。

文献 [93] 提出桥接度算法识别对于全局关联起到重要作用的节点和连边。文献 [64] 揭示了具有较小局部聚类系数的节点在未来能够与更多的节点建立联系，并提出了一种同时考虑目标节点邻居度值与邻居之间交互信息的算法，其表达形式为：

$$s_i=f(c_i)\sum\nolimits_{j\in\Gamma_i}(k_j^{\mathrm{out}}+1) \tag{6}$$

式中，$f(c_i)$为关于节点i聚类系数的函数；$k_j>1$。其基本思想为：当节点的局部聚类系数大时，信息

难以扩散到网络的其他部分；当邻居节点度越大且目标节点的局部聚类系数越小时，节点越有可能将信息传播给网络中更多节点，所以节点就越重要。文献 [65] 基于节点二阶邻居聚类系数与节点重要性的正相关关系提出了 LSC 算法，公式为：

$$\mathrm{LSC}(i)=\left(\sum_{j\in\Gamma_i} aN(j)+(1-a)\sum_{v\in\Gamma_j^2} c_v\right) \tag{7}$$

式中，$N(j)=\left|\Gamma_j^2\right|$为节点 j的一阶和二阶邻居数之和；$a\in[0,1]$为自由参数。虽然自由参数的存在使算法更加灵活，但调参需要测试，会产生时间成本。文献 [94] 引入了“熵”来计算分配给局部聚类系数及节点度值的权重，计算公式为：

$$\mathrm{DCC}(i)=\alpha I_D(i)+\beta I_c(i) \tag{8}$$

式中，$I_D(i)=k(i)+\sum_{j\in\Gamma_i} k(j)$表示与节点 i的度相关信息；而$I_c(i)=\mathrm{e}^{-c_i}\sum_{j\in\Gamma_i} c_j$表示节点 i的局部集聚系数信息，c_j越大，e^{-c_i}就越小；a和β为参数。文献 [95] 提出了一种同时考虑节点度、局部聚类系数与节点影响力的算法，定义为：

$$\mathrm{NPC}(i)=\frac{k_i}{\mathrm{cc}(i)+\dfrac{1}{k_i}}\sum_{j\in\Gamma_i^2} c_j \tag{9}$$

式中，$\mathrm{cc}(i)$为节点 i的接近中心性；$\Gamma_j{}^2$为节点 j的一阶和二阶邻居节点集合。

虽然引入节点的局部聚类系数能够区分相同度值节点的重要性。但基于局部聚类系数的算法未考虑节点的全局位置，易将局部连接紧密但处于网络边缘位置的节点识别为重要节点。

在解决影响力最大化问题中，MI 基于微观结构的影响力最大化算法所面临的主要挑战是如何避免种子节点之间的影响力重叠[96, 98-100]。以度中心性为例，由于社交网络通常具有异质性，仅根据一阶邻居节点数量来选择种子节点，容易得到聚集在网络某一局部位置的种子节点集。

针对影响力重叠问题，有研究者通过度量节点邻域结构特征的相似性使得所选种子节点尽量分散在网络的各个部分。文献 [97] 将种子节点之间的距离与共同邻居占比作为选择种子节点的约束条件。通过最小化种子节点的共同邻居占比并控制种子节点之间的距离来避免影响力重叠。但计算节点之间的距离具有较高的计算复杂度，文献 [60] 仅利用节点之间的共同邻居数量，提出了 local structure similarity （LSS）方法。首先，LSS 将网络中度值最大的节点设为初始候选节点，然后以候选节点集中所有节点共同的一阶和二阶邻居为范围，寻找与所有候选节点局部结构相似性小于一个给定阈值r的节点加入候选节点集。最后按照加入的先后顺序选择前k个候选节点作为最终的种子节点集。LSS 所使用的局部结构相似性计算公式为：

$$S_{ij}=\frac{\left|\Gamma_i\cap\theta_j\right|}{k_i} \tag{10}$$

式中，Γ_i为节点 i的一阶邻居集合；当节点 i与节点 j直接相连时，θ_j为节点 j的一阶邻居集合；当节点 j为节点 i的二阶邻居时，θ_j表示节点 j的一阶和二阶邻居集合。虽然通过约束种子节点之间的局部结构相似性能够缓解影响力重叠问题，但基于局部结构相似性的影响力最大化算法依赖于初始参数的设定。以 LSS 算法为例，LSS 在选择初始种子节点时仅考虑了节点的度信息，没有将其他有价值的局部结构属性考虑在内。

受投票机制的启发，文献 [62] 设计了 VoteRank 算法。VoteRank 算法在初始化阶段赋予每个节点相同的投票值和得分，即(V_{i0},S_{i0})，然后根据式 (11) 更新每轮投票后节点的得分从而聚合高阶邻居的信息：

$$S_i=\sum_{j\in\Gamma_i} V_j \tag{11}$$

每轮总得分最高的节点会被选为该轮的种子节点，并且该节点在下一轮的得分会变为 0，其邻居节点的投票值也会按一定比例降低，以缓解种子节点的影响力重叠。VoteRank 算法在初始阶段将各节点的投票值均设为 1，这使得在同一轮投票中往往无法区分同度值节点重要性的差异。文献 [101] 认为投票过程中每个节点的投票能力取决于其局部重要性，进而对投票迭代机制进行了改进。文献 [102] 提出应从个体和群体的角度分别衡量节点的传播影响力。从个体角度出发，该方法基于社会从众理论认为每个节点对于其邻居节点的吸引力是不同的，量化吸引力的公式为：

$$\mathrm{AP}(i,j)=\begin{cases}\dfrac{\left|\Gamma_i^{\mathrm{out}}\right|}{\sum_{v\in\Gamma_i^{\mathrm{out}}}\left|\Gamma_v^{\mathrm{in}}\right|} & \sum_{v\in\Gamma_j^{\mathrm{out}}}\left|\Gamma_v^{\mathrm{in}}\right|\neq 0\\ \dfrac{1}{\left|\Gamma_j^{\mathrm{out}}\right|} & \sum_{v\in\Gamma_j^{\mathrm{out}}}\left|\Gamma_v^{\mathrm{in}}\right|=0,\Gamma_j^{\mathrm{out}}\neq\varnothing\end{cases} \tag{12}$$

式中，Γ_i^{out}为节点 i的出度邻居的集合。从群体角度出发，以节点所处社团的规模来区分节点的重要性，计算公式为：

$$\mathrm{IP}(i,j)=\begin{cases}0 & N_{\mathrm{Com}_i}=N_{\mathrm{Com}_j}\\ \dfrac{\left|N_{\mathrm{Com}_i}\right|}{\mathrm{Max}(N_{\mathrm{Com}_v})} & v\in V, N_{\mathrm{Com}_i}\neq N_{\mathrm{Com}_j}\end{cases} \tag{13}$$

式中，N_{Com_i}表示节点i所属社团的规模。节点的投票能力由式 (14) 计算得到：

$$S_i=\sum\nolimits_{j\in\Gamma_i}(\mathrm{AP}(i,j)+\mathrm{IP}(i,j)) \tag{14}$$

在此基础上，文献 [102] 为了进一步降低影响力重叠的问题，从个体和群体的角度提出了两种选择种子节点的策略：1）当节点被选为种子节点时，该节点的邻居就会被删除，且该节点不会再进行投票；2）当所选节点所处社团与种子节点所属社团紧密性较强时，该节点不会被选为种子节点。文献 [103] 认为，这种差异可以通过邻居节点的 K-核值来进行区分，并基于邻居 K-核值[104] 对 VoteRank 进行了改进：

$$S_i=\sum\nolimits_{j\in\Gamma_i}V_i(1-a)C_{nc}(j)+V_i\times a \tag{15}$$

式中，$C_{nc}(j)$为节点i的邻居j的 K-核值；$a\in[0,1]$为自由参数。文献 [105] 利用信息熵来区分各节点的投票值差异，提出了 EnRenew 算法。EnRenew 根据式 (16) 迭代计算节点的信息熵，并每次选择信息熵最大的节点加入种子节点集：

$$S_i=-\sum\nolimits_{j\in\Gamma_i}p_{ij}\log p_{ij} \tag{16}$$

$$p_{ij}=\frac{k_i}{\sum_{l\in\Gamma_j}k_l} \tag{17}$$

当节点被选取后，该节点的l阶邻居的信息熵会按式 (18) 进行不同程度的衰减以避免影响力重叠：

$$S_{j^{l-1}j^l}=\frac{1}{2^{l-1}}\frac{S_{j^{l-1}j^l}}{E_{\langle k\rangle}} \tag{18}$$

$$E_{\langle k\rangle}=-\langle k\rangle\frac{1}{\langle k\rangle}\log\frac{1}{\langle k\rangle} \tag{19}$$

式中，j^l为节点j的l阶邻居节点；$\langle k\rangle$为网络的平均度。文献 [32] 在 VoteRank 算法的基础上考虑节点的邻居数量以及连边权重，拓展了该方法在加权网络中的使用，提出了 WVoteRank 算法，表达式为：

$$S_i=\sqrt{(|\Gamma_i|\sum\nolimits_{j\in\Gamma_i}V_jw_{ij})} \tag{20}$$

式中，V_j为节点j的投票值；w_{ij}为节点i和节点j之间的连边权重。

由于未涉及节点之间的距离信息，基于 VoteRank 的方法具有较高的效率。但是如何为各节点的初始投票值进行分配，以及选取种子节点后如何对邻居节点的投票值和投票得分进行更新仍存在一定的改进空间。

节点的重要性不仅与结构属性相关，与信息传播机制同样存在相关性。折扣度方法[106] 在独立级联模型基础上，根据目标节点邻居中种子节点的数量对节点度进行打折以降低影响力重叠。文献 [107] 假设当节点具有较高概率影响其邻居节点时，该节点具有较大的潜在影响力。在考虑信息传播机制的情况下，从概率的角度出发，以目标节点高阶邻居被感染的概率之和作为目标节点的重要性指标，公式为：

$$R_i=\sum_{l=1}^{3}\sum_{j\in\Gamma_i^l}P(j,l) \tag{21}$$

式中，$P(j,l)$表示节点i的第l阶邻居节点j被感染的概率，计算公式为：

$$P(j,l)=1-F_s(j,l) \tag{22}$$

$$F_s(j,l)=\prod_{v\in\Gamma_i^{l-1}}[1-P(j,l-1)\beta] \tag{23}$$

式中，i为目标节点；j为已感染节点；l为节点i的第l阶邻居节点；β为感染概率；$P(v,0)=1$。

基于传播机制的算法同时考虑了节点的结构属性与信息传播机制，更贴近实际情况。但现实中不同传播事件的信息传播机制存在差异，这限制了传播动力学算法的应用场景。

基于局部信息的 MI 算法为大规模社交网络中的节点重要性识别提供了解决方案。具体可分为 5 类：基于高阶邻居信息的方法、基于局部聚类系数的方法、基于局部相似性的方法、基于 VoteRank 的方法及基于传播机制的方法。代表性方法与以上 5 类方法的优缺点分别在表 2 和表 3 中列出（表中 r 表示迭代数）。虽然 MI 方法已经取得了出色的表现，但还存在如下挑战。

1）针对节点传播影响力排序问题。虽然基于高阶邻居度信息的算法计算复杂度较低，但邻居节点之间的联系未考虑在内，难以区分具有相同度值节点的重要性。基于局部聚类系数的算法改善了这一缺陷。值得注意的是，基于高阶邻居度信息的算法与基于局部聚类系数的算法均假设重要节点为处于局部连接较为紧密位置的节点，没有考虑社团层面与宏观层面信息。这导致以上两类算法在连边密

度较高的网络中，容易将处于网络边缘位置的节点识别为重要节点。因此，如何改善基于局部信息的方法在连边密度较高网络中的表现是一个值得研究的重要方向。

表 2　基于微观局部结构属性的节点重要性识别方法

方法	优势	劣势	时间复杂度
DC[22]	简单易于理解，计算复杂度低	未能充分考虑目标节点邻域内的拓扑信息；节点重要性识别过于粗粒度	$O(n)$
LocalCentrality[52]	利用邻居度值提高了度中心性对于节点影响力的区分能力	除了节点度之外没有考虑其他的结构信息	$O(\langle k\rangle n^2)$
LNC[31]	在保证较低计算复杂度的情况下取得了相较于度中心性、介数中心性更高的准确性	不适用于随机网络	$O(\langle k\rangle n)$
VoteRank[62]	计算复杂度低，比同样基于迭代的PageRank、LeaderRank方法准确度更高	未考虑不同节点投票能力的差异	$O(n)$
AIRank[102]	分别从个体和群体的角度考虑节点的影响力，提高了VoteRank算法区分节点影响力的能力	当社团间的连边紧密度较低时，准确度较差	$O(n)$
NCRank[103]	基于节点所处的位置区分了节点的投票能力，提升了VoteRank的准确性	调整自由参数会耗费较多时间	$O(n)$
EnRenew[105]	基于信息熵区分了节点不同的初始投票值，并对不同阶数邻居节点设置衰减机制，提升了VoteRank的准确性	计算复杂度相对于VoteRank要更高	$O(m+n+r\log(n)+\frac{m^2}{n^2})$
DynamicRank[107]	从传播概率的角度切入，准确度高于LeaderRank方法	需要确定一个合适的传播概率	$O(n)$

表 3　基于微观的局部结构属性节点重要性识别方法

方法流	相关工作	优势	劣势
基于高阶邻居度算法	LNC[31]; SS[47]; LC[52]; NC[57];	高效且可解释性较强	未考虑邻居节点之间的交互信息
基于局部聚类系数算法	CR[64]; LSC[65]; DCC[94]; Centrality[95]	节点重要性区分粒度比度中心性更细	不适用于连边密度较高的网络
基于局部相似性算法	LSS[9]; HC[91]; DegreeDistance[97];	确保种子节点分散在网络的各部分	准确性易受初始条件的影响
基于VoteRank算法	VoteRank[62]; AIRank[102]; NCRank[103]; EnRenew[105]	不需要考虑节点之间的距离信息	算法易受节点初始状态和重要性衰减机制的影响
基于传播机制算法	DD[106]; DynamicRank[107]	同时考虑节点的结构信息和传播机制	算法依赖于特定的传播机制

2）针对节点影响力最大化问题，基于局部相似性的算法以节点之间的结构特征相似性作为约束条件来避免重要节点影响力重叠[60]，但该类方法易受初始设定的影响，如何减少对初始条件的依赖值得进一步探讨。基于 VoteRank 的算法在考虑邻居节点贡献差异的基础上以迭代的方法来选择种子节点，其主要挑战在于如何量化邻居节点的贡献差异。基于传播动力学的算法是否在不同传播机制下能够保证较强的泛化性能是值得研究的另一个重要问题。

3　基于中观社团结构的算法

社团结构是社交网络的重要结构特征之一[70, 108]，网络中的社团由一组连边密度较高的节点子集所构成。同一社团节点间的连边较为密集，属于不同社团节点间的连边则较为稀疏。社团规模、社团内外部连边密度、所连社团个数等属性能够反映出网络更深层的结构信息[109]，有助于帮助研究人员进一步了解网络的结构与信息传播的机制。如要将社团 1 产生的信息传输给社团 2 的成员，则必须要有同时连接社团 1 和社团 2 的成员参与。此时，扮演中间角色的成员对于信息的传播就起到了非常重要的桥梁作用。文献 [66] 将节点所连接社团个数考虑在内提出了 Vc 指标，挖掘出了单靠一种中心性方法无法发现的重要节点。但由于 Vc 指标对于社团划分的结果依赖性较强，且不同社团发现算法对于同一网络进行社团划分的结果存在差异[107]，导致 Vc 指标存在不稳定性。为此，文献 [67] 进一步考虑目标节点所连接社团的规模以及目标节点邻居在各个社团的分布情况并提出了 CbC 方法，其数学定义如下：

$$\mathrm{CbC}_i=\sum_{q=1}^{c}k_{iq}\frac{N_{\mathrm{Com}_h}}{n} \tag{24}$$

式中，k_{iq}表示节点i的邻居节点中属于社团q的节点数量；c为网络中社团的总个数；N_{Com_h}为社团h中的节点总数；n 为网络中的节点数。文献 [68] 基于

节点在社团内外部的连边密度，根据节点随机游走到各个社团的熵以及节点的度值来识别节点的重要性。其基本思想为：若目标节点与所属社团内部与外部的节点都紧密连接，该节点就为重要节点，计算公式如下：

$$\mathrm{CbM}(i)=H(i)\frac{\frac{k_j}{n}}{\sum_{j=1}k_j} \tag{25}$$

式中，$H(i)$即为节点随机游走到网络中各社团的信息熵值。文献 [111] 通过引入节点一阶邻居所属社团规模与所连社团个数，提出了改进的接近中心度（improved closeness centrality, ICC）方法，其计算公式为：

$$\mathrm{ICC_i}=\mathrm{cc}(i)\frac{N_{\mathrm{Com_i}}}{n}+\sum_{w\in W_i}\mathrm{Max}\{cc(j)\}\frac{N_{\mathrm{Com}w}}{n}\quad j\in w,a_{ij}=1 \tag{26}$$

式中，$\mathrm{cc}(i)$为节点i的接近中心性；W_i为除了节点i所属社团外，节点i所连接社团的集合。

从社团结构网络划分的角度，文献 [34] 通过构建仅由社团内连边构成的网络，分别衡量了节点对于社团内部和外部的影响力。通过实验发现：当网络的社团结构较强时，在局部网络上进行节点重要性识别比在全局网络上要更准确；当网络的社团结构较弱时，在全局网络上进行节点重要性识别的结果更优。在现实当中，一个节点很有可能同时属于多个社团，即网络中具有重叠社团结构。这些同时属于多个社团的节点可以被视为社团间的桥梁，往往能够扩大信息的传播范围。因此，文献 [34] 提出了一种适用于具有重叠社团结构的重要节点识别框架，如图 3 所示。

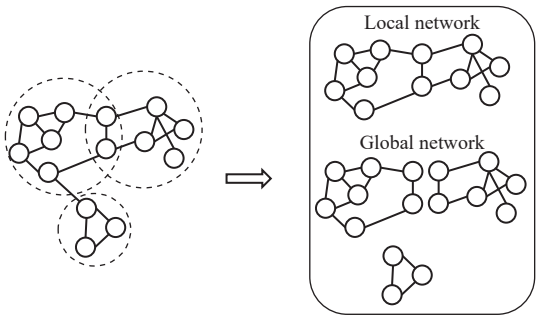


图 3 基于社团结构的重要节点识别框架

为识别具有重叠社团结构网络中的重要节点，文献 [112] 利用 BIGCLAM 模型[113] 检测出网络的重叠社团结构，然后根据目标节点的一阶和二阶邻居节点所连社团个数与网络约束系数来选择种子节点集，其表达形式为：

$$\mathrm{OC}(i)=\frac{\sum_{j\in\Gamma_i}\sum_{v\in\Gamma_j}10^{-C_v}\mathrm{Com}_n(v)}{\mathrm{Max}\{\mathrm{OC}\}} \tag{27}$$

式中，C_v为节点v的网络约束系数，可用于衡量节点通过形成结构洞施加的约束；$\mathrm{Com}_n(v)$是节点v的邻居节点所连社团个数。

上述工作证明了社团结构对于识别社交网络的节点重要性具有重要作用。但随着更多社团结构特征被引入，算法的复杂度同样会上升。因此，筛选出更有效的社团特征是保证算法效果的基础。此外，虽然基于社团结构进行网络划分为节点重要性识别提供了一个新的视角，但是社交网络的社团划分结果会随着社团划分算法的表现出现变动，算法的性能容易产生波动。

除了能反映网络拓扑结构的信息外，社团结构信息有助于提升节点影响力最大化算法的效率。文献 [114] 基于次模函数理论[115] 假设：对网络进行社团划分之后，独立地在各社团中进行节点重要性识别，最后获得的种子节点集影响力与基于整个网络进行节点重要性识别得到的影响力近似相等。文献 [14] 以爬山算法[116] 作为实验对像，先利用可并行计算的 Louvain 算法[117] 对网络进行社团划分，然后在每个社团中独立进行节点重要性识别，最后将获得的重要节点进行合并得到种子节点集。结果证明：这一策略能够在保证准确性的基础上大幅度降低算法的计算复杂度。文献 [118] 提出了一种以社团为基础进行节点重要性识别的算法，称为 CGA 算法。算法使用动态规划算法选取能够带来最大重要性增益的目标社团以缩小选择范围，然后对目标社团内部的节点进行重要性识别。社团C_i中节点能够带来的最大重要性增益ΔR_i由式 (28) 计算得到：

$$\Delta R_i=\max\{R_i(I_{k-1}\cup v)-R_i(I_{k-1})|c\in\mathrm{Com}_i\} \tag{28}$$

式中，I_{k-1}是包含$k-1$个重要节点的集合；$R_i(I_{k-1})$为节点集I_{k-1}的影响力。选择目标社团的策略如下：

$$R[u,k]=\max[R[u-1,k],R[u,k-1]+\Delta R_c] \tag{29}$$

$$R[u,0]=0,R[0,k]=0 \tag{30}$$

式中，u表示当前共挖掘了的社团数量；$R[u-1,k]$是在前$u-1$个社团中挖掘第k个种子节点带来的影响力。由式 (29) 可知，当从前$u-1$个社团中挖掘第

k个种子节点带来的影响力小于在社团u中挖掘第k个种子节点带来的影响力时，该算法会选择社团u；反之，会在前$u-1$个社团中来挖掘。这一策略相较于同样基于贪婪算法的 MixGreedy[92] 方法具有较低的计算复杂度，但在准确度上有所降低。考虑传播机制，文献 [119] 将传播过程分为两个阶段：1）种子节点集S向一阶邻居集$N(S)$进行传播；2）种子节点的一阶邻居集Γ_j向各社团内部的非种子节点进行传播。第一个阶段，$N(S)$中的节点被种子节点集S传播的概率为：

$$P_i(S)=1-\prod_{j\in\Gamma_i\cap S}(1-p_{ji}) \tag{31}$$

式中，p_{ji}为节点i感染节点j的概率。第二阶段，$N(S)$中的节点在各社团内部独立进行传播，这一阶段的影响力为：

$$f(S)=\sum\nolimits_{\mathrm{Com}_i\in\mathrm{Com}} f(S,S',\mathrm{Com}_i) \tag{32}$$

$$f(S,S',\mathrm{Com}_i)=\sum\nolimits_{v\in\mathrm{Com}_i} P_v(S,S',\mathrm{Com}_i) \tag{33}$$

式中，$P_v(S,S',\mathrm{Com}_i)$为处于社团i的节点v最终被激活的概率；Com_i为社团集合。基于加权级联模型，种子节点集S的最终影响力为：

$$g(S)=|N(S)|+a|\mathrm{NC}(S)| \tag{34}$$

式中，$\mathrm{NC}(S)$代表S的邻居节点集合；$|N(S)|$代表种子节点集合S的邻居节点个数；a为自由参数。

基于传播机制的中观社团结构（ME）方法能够从更全面的角度去度量节点的重要性，但其局限性与基于传播动力学的基于局部结构（MI）方法类似，即随着传播机制的改变，算法性能也会发生波动。

社交网络中的各个社团并不是一个社团的简单复制。各个社团的规模、连边密度等属性均存在差异。文献 [120] 认为相较于社团间的信息传递行为，更多的信息传递行为发生在社团内部。因此，相对于小规模社团中的节点，大规模社团的节点能够将信息进行更大范围的传播。通过社团规模筛选出目标社团，在每个目标社团中，根据节点的度值、相似性总和以及是否为枢纽节点这 3 个因素来选择候选节点集，最后再从候选节点集中选择种子节点。这一策略虽然降低了计算复杂度，但社团规模并非影响社团内部节点重要性的唯一因素。如社团间的连边同样也是反映节点重要性的一个关键因素。文献 [72] 认为在信息传播的过程当中，影响力较大的节点往往分布在对于信息传播较为重要的社团当中，并基于结构洞理论，将社团的重要性分为内部重要性和外部重要性：

$$C_{r_i}=a\times\mathrm{CI}_i^{\mathrm{out}}+\beta\times\mathrm{CI}_i^{\mathrm{in}} \tag{35}$$

式中，α与β为自由参数；$\mathrm{CI}_i^{\mathrm{in}}$为节点$i$所属社团的内部连边密度；$\mathrm{CI}_i^{\mathrm{out}}$考虑的是节点$i$所属社团在网络中的重要性，即将各社团视为节点，度量各社团的重要性。由结构洞理论可知，在网络中与多个社团相连的节点往往比仅与单个社团相连的节点对信息传播具有更强的控制力，因此在考虑社团重要性的基础上也需要考虑节点所连接社团的数量，文献 [72] 提出了基于社团重要性的方法：

$$\mathrm{CHR}_i=C_{r_i}B_{r_i} \tag{36}$$

$$B_{r_i}=\frac{\sum_{j\in\Gamma_i^{\mathrm{out}}}e_{ij}C_{r_j}}{\sum_{t\in\Gamma_i^{\mathrm{in}}}e_{it}} \tag{37}$$

式中，C_{r_j}为节点j的社团重要性；Γ_i^{out}为与节点i属于不同社团的邻居节点集合；Γ_i^{in}为与节点i属于同一社团的邻居节点集合。当目标节点属于具有较重要的社团且与较多重要社团连接时，节点就会被赋予较大的重要性。文献 [121] 基于社团结构将寻找种子节点的过程划分为 3 个阶段：1）在每个社团内部，基于节点度和节点连接社团数量获得候选种子节点集；2）在候选种子节点集中，根据节点所连社团的平均规模、所连社团个数以及度值再次筛选以缩小搜索范围；3）基于独立级联模型，利用贪婪算法找出最终的种子节点集。在第一阶段，以社团为单位识别出核心节点集S_c与边缘节点集S_b，即度值较大的节点以及与多个社团连接的节点。第二阶段，根据式 (38) 分别对核心节点集和边缘节点集进一步筛选：

$$\mathrm{CI}(i)=\begin{cases}k_i+N_{\mathrm{Com}_i}+\dfrac{\mathrm{Avg}N_{\mathrm{Com}_i}}{3} & i\in S_b\\ k_i+\dfrac{N_{\mathrm{Com}_i}}{2} & i\in S_c\end{cases} \tag{38}$$

式中，N_{Com_i}为节点i所属社团的规模；$\mathrm{Avg}N_{\mathrm{Com}_i}=\dfrac{\sum_{j\in W}N_{\mathrm{Com}j}}{|W|}$为节点$v$所连社团的平均规模。最后，利用贪婪算法从所得的节点集中寻找种子节点。该方法通过预筛选来减少贪婪算法所要考虑的节点范围从而降低计算复杂度。

社团结构属性在提升节点重要性排序准确性及种子节点选择效率上起到了重要的作用。然而，这

些方法仍然存在以下挑战。

针对节点重要性排序问题，并不是考虑越多社团属性算法的性能就越好。因此，设计此类算法时，在保证算法准确性的情况下，也要保证算法的计算效率。针对节点重要性排序和影响力最大化问题。由于现实中社交网络的真实社团结构往往是未知的，而不同的社团划分算法对于同一个网络划分的结果往往存在差异。因此，如何增强基于中观社团结构 ME 算法对于社团划分算法结果差异的抗扰动能力仍然是一个挑战。此外，基于中观社团结构 ME 算法往往会用到社团本身的结构属性，但目前关于社团结构属性与节点重要性之间关系的研究还较少，缺乏理论依据。最后，真实的社交网络中存在重叠社团结构，但大多数 ME 方法是基于非重叠社团结构设计的。因此，如何将现有的算法拓展到具有重叠社团结构的网络中，是有待进一步解决的问题。本节代表性算法的优缺点信息总结在了表 4 和表 5 中，其中，K为种子节点总数，M为社团个数，T_p为计算属于社团p的一个节点的度值所需的时间，n'和m'分别表示备选节点和连边数量。

表 4　基于社团结构属性的节点重要性识别算法

方法	优势	劣势	时间复杂度
Vc[66]	能够发现靠单一中心性无法识别的重要节点	过于依赖社团划分算法，表现不稳定	$O(n)$
CbC[67]	考虑了社团规模以及各社团内的邻居数量，降低了Vc指标的不稳定性	除社团规模外，没有考虑其他社团结构属性	$O(n\langle k\rangle)$
CbM[68]	考虑了节点在社团内外连边的密度，准确性高	计算复杂度高	$O(mn\langle k\rangle)$
CGA[118]	以社团为单位识别重要节点集，降低了计算复杂度	相较于其他算法，该方法在大规模网络中的计算复杂度仍然较高	$O(\mathrm{MKT}_p + KN_{\mathrm{Com}_p}T_p)$
CoFIM[119]	将传播过程分为两个阶段，在确保准确性的情况下降低了计算复杂度	需要确定合适的传播概率和自由参数	$O(K^2nk_{\max})$
PHE[121]	通过预筛选缩小贪婪算法考虑的范围，降低了计算复杂度	在选择核心节点候选集时仅使用度值	$O(n\log n + n + Kn'm')$
CIM[120]	基于社团规模来缩小种子节点的选择范围	社团规模不是反映节点影响力的唯一因素	—
CHR[72]	基于结构洞理论考虑社团的重要性对于节点重要性的影响	无法适用于大规模网络	—
ICC[108]	利用社团结构信息提高了准确性	无法适用于大规模网络	—

表 5　3 类 ME 的算法优势与劣势

方法流	相关工作	优势	劣势
基于社团结构属性算法	MC[34]; Vc[66]; CbC[67]; CbM[68]; ICC[111]; OC[112]	社团结构属性有助于提高节点重要性识别的准确率和效率	对于社团划分算法的依赖性较强
基于传播动力学算法	CGA[118]; CoFIM[119]	同时考虑了社团结构与传播机制，提升了种子节点选择的效率	传播机制限制了其应用场景
基于社团重要性算法	CHR[72]; CIM[120]; PHG[121]	考虑了各社团对于传播的重要性差异；社团结构提升了种子节点选择的效率	如何准确度量社团的重要性缺乏理论依据

4　基于宏观全局结构的算法

考虑社交网络的全局结构信息，能够准确区分处于边缘层和核心层位置的节点。经典基于宏观全局结构的 MA 中心性方法包括介数中心性[23]、接近中心性[122]、特征向量中心性[123]。不同中心性方法对于节点重要性的定义不同。

介数中心性根据节点在网络所有最短路径中出现的次数来衡量节点的重要性，其数学表示如下：

$$\mathrm{BC}(i) = \sum_{v,u\in V} \frac{\sigma(v,u|i)}{\sigma(v,u)} \tag{39}$$

式中，$\sigma(v,u)$为节点v与节点u之间的最短路径数；$\sigma(v,u|i)$为节点v与u的最短路径经过节点i的次数。

接近中心性[122]将越接近于网络中心的节点定义为越重要的节点。若目标节点与网络中所有其他节点之间的平均最短路径长度越小，接近中心性就认为该节点越接近于网络的中心，计算公式为：

$$\mathrm{cc}(i) = \frac{n-1}{\sum_{j\neq i} d_{ij}} \tag{40}$$

式中，d_{ij}为节点i与节点j之间的最短路径长度。

特征向量中心性[123]基于邻居节点的重要性来衡量目标节点的重要性，计算公式为：

$$\mathrm{EC}(i) = a\sum_{j=1}^{n} a_{ij}\boldsymbol{x}_j \tag{41}$$

式中，a为比例常数；$\boldsymbol{X}=[x_1,x_2,\cdots,x_n]$为包含各节点重要性的向量。通过迭代更新$\boldsymbol{X}$，直到$\boldsymbol{X}=\boldsymbol{aAX}$，进而可以得到特征向量度量节点的重要性。

PageRank 算法是基于特征向量中心性思想的经典方法。最初用于对 Google 搜索引擎返回的结果进行排名[24]。由于初始版本的 PageRank 算法只能在强连通网络中收敛，后续引入返回概率以解决这一问题。但返回概率的确定需要经过大量的测试，这使得当 PageRank 用于社交网络时会变得比较低效。为此，文献 [45] 在 PageRank 的基础上提出了一种无参数形式，即 LeaderRank 算法。该算法通过在网络当中加入一个与所有节点均双向连接的 Ground 节点，确保网络成为强连通网络。具体地，在初始阶段，除了 Ground 节点外的所有节点 LR 值均为 1，Ground 节点 LR 值为 0，然后根据式 (42) 迭代更新各节点的 LR 值：

$$\mathrm{LR}_i(t)=\sum_{j=1}^{n+1}a_{ij}\frac{\mathrm{LR}_j(t-1)}{k_j^{\mathrm{out}}} \tag{42}$$

当该过程收敛后，将 Ground 节点的 LR 值均匀分配给网络中的每个节点。文献 [124] 认为 Ground 节点的得分不应均匀分配，入度较大的节点应获得更多的值。以社交网络为例，入度大的用户代表该用户受到了较多的关注，可以反映出该用户的重要性，改进后的 LR 值为：

$$\mathrm{LR}_i(t)=\sum_{j=1}^{n+1}w_{ij}\frac{\mathrm{LR}_j(t-1)}{\sum_{l=1}^{n+1}w_{jl}} \tag{43}$$

式中，$w_{ij}=\{k_i^{\mathrm{in}}\}^a$，$a$为自由参数。

基于迭代的算法主要思想是通过迭代来聚合高阶邻居节点的结构信息，比直接使用节点之间的距离更高效。但此类方法在量化邻居节点对目标节点的重要性贡献时主要依赖于节点度，未考虑其他结构信息。

通过模拟离散 SIR 模型，文献 [80] 以起始节点在时刻t产生的传播概率之和作为节点的重要性，提出了 DS 算法。DS 算法假定处于激活状态的节点以概率β激活处于未激活状态的节点，而被激活的节点有μ的概率恢复为未激活状态。节点在t时刻被激活的概率为：

$$x(t)-x(t-1)=\beta\boldsymbol{A}[\beta\boldsymbol{A}+(1-\mu)\boldsymbol{I}]^{t-1}x(0) \tag{44}$$

式中，$\boldsymbol{I}$为单位阵；$x(t)$则表示节点在 1 到t时刻被激活的累计概率。根据式 (44)，节点在 1 到t时刻被激活的累计概率为：

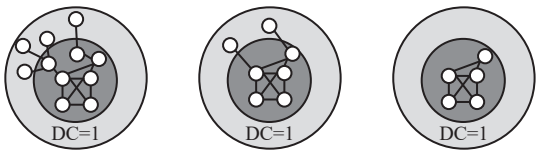
$$x(t)=\sum_{r=2}^{t}[x(r)-x(r-1)]+x(1) \tag{45}$$

式中，$H=\beta\boldsymbol{A}+(1-\mu)\boldsymbol{I}$，当节点$i$为初始激活节点时，令$\boldsymbol{e}_i=(0,\cdots,0,1,0,\cdots,0)^{\mathrm{T}}$，该向量中除了第$i$个元素为 1，其他元素均为 0。节点$i$的传播影响力为：

$$S_i(t)=\sum_{r=0}^{t-1}\beta\boldsymbol{A}\boldsymbol{H}^r\boldsymbol{e}_i \tag{46}$$

DS 算法不仅考虑了网络拓扑结构属性，同时也考虑了传播机制。

文献 [81] 发现节点在网络中所处的位置能比节点的度值更准确地反映出节点的重要性，提出了 K-核分解法。K-核分解法根据节点度从小到大递归移除节点，将节点划分到网络的各个层级当中。图 4 为 K-核分解法寻找属于最外层节点的过程示意图：1）首先剔除度值最小的节点，图 4a 中度值最小的节点为 1；2）在删除节点后的网络中，继续剔除度值为 1 的节点，如图 4b 所示；3）重复步骤 1）和步骤 2）直到网络中不存在度值为 1 的节点为止，即图 4c。



a. 步骤 1）　b. 步骤 2）　c. 步骤 3）

图 4　K-核分解示意图

K-核分解法的缺陷在于会把多个节点识别为同样的 k 核值[77, 104]。针对这一问题，研究者提出了多种解决方案。文献 [84] 利用目标节点到核心层各节点的最短路径长度来区分位于同一层级的节点的重要性，数学表达式为：

$$\theta(i|k_s)=(k_s^{\max}-k_s+1)\sum_{j\in J}d_{ij}\quad i\in S_{k_s} \tag{47}$$

式中，J为属于核心层的节点集合；S_{k_s}为 K-核值等于k_s的节点集合。文献 [125] 基于 K-核分解以及“富者越富”的思想[126] 提出了 RDA 算法，基本思想为：在资源分配时，网络中越重要的节点随着时间的推移往往会获得更多的资源。具体地，RDA 算法在初始化阶段赋予每个节点相同的资源，然后迭代地根据邻居节点的 K-核值，见式 (48)，将资源分配给邻居节点。当资源增益小于一个给定阈值

后停止迭代，这一策略保证了较低的计算复杂度。

$$R_{j\to i}(t+1)=\left(\frac{k_s(i)}{\sum_{u\in\Gamma_j}k_s(u)}a_{ij}\right)R_j(t) \tag{48}$$

式中，Γ_j为节点j的邻居节点集合；a_{ij}为节点邻接矩阵第i行和第j列的值；$R_{j\to i}(t+1)$表示$t+1$时刻节点j分配给节点i的资源；$R_i(t)$表示节点i在时刻t拥有的资源。在时刻$t+1$节点i所获得的资源为：

$$R_i(t+1)=\sum_{u\in\Gamma_j}R_{j\to i}(t+1) \tag{49}$$

文献 [36] 基于 IKs 算法，从邻居节点多样性、传播范围以及传播强度 3 个方面综合评价节点的影响力。首先基于香农熵计算各个节点邻居的多样性：

$$H_1(i)=-\sum_{j\in\Gamma_i}\frac{\mathrm{IK}_s(j)}{\mathrm{IK}_s(\Gamma_i)}\log\left(\frac{\mathrm{IK}_s(j)}{\mathrm{IK}_s(\Gamma_i)}\right) \tag{50}$$

式中，IK_s为改进 K-shell 分解法[127]；$\mathrm{IK}_s(\Gamma_i)$为节点i所有邻居节点的IK_s值，当目标节点邻居越多且越接近于核心层时，$H_1(i)$就越大。接着使用多样性[128] 来衡量目标节点的传播范围以及传播强度：

$$\mathrm{JSD}(j\in\Gamma_i)=H\left(\sum_{j\in\Gamma_i}\frac{1}{k_i}X_j\right)-\sum_{j\in\Gamma_i}\frac{1}{k_i}\times H(X_j) \tag{51}$$

式中，$X_j=(p_{j1},p_{j2},\ldots,p_{jk_s^{\max}})$为目标节点邻居在由$\mathrm{IK}_s$划分后各层邻居节点数的分布，基于式 (50) 和式 (51)，节点最终的重要性为：

$$\mathrm{DSC}(i)=\overline{\mathrm{IK}_s(\Gamma_i)}\times H_1(i)\times\mathrm{JSD}(j\in\Gamma_i) \tag{52}$$

式中，$\overline{\mathrm{IK}_s(\Gamma_i)}$为节点$i$邻居节点的$\mathrm{IK}_s$均值；$H_1$与 JSD分别由式 (51) 与 (52) 计算得到。

虽然属于同层节点的 K-核值相等，但是这些节点的度值存在差异。文献 [129] 将节点的度值以及邻居节点的 K-核值同时考虑以区分同一层级节点的影响力。基本思想为：若节点与越多接近或处于核心层的节点相连，该节点就越重要，公式为：

$$C_{nc}(i)=\sum_{j\in\Gamma_i}k_s(j) \tag{53}$$

文献 [130] 聚焦于具有最大 K-核值的节点，其中 Γ_j为节点j的邻居节点集合。利用邻居节点的 K-核值之和来识别网络中的重要节点，计算公式为：

$$\mathrm{INK}(i)=\sum_{j\in\Gamma_i}k_s(j)^a \tag{54}$$

式中，a为自由参数。当$a>1$时，k 核越大的节点发挥的作用越大。

除了节点度的差异外，在 K-核分解迭代的过程当中，同一层节点可能在不同的迭代轮数被移除，那些后移除的节点比先移除的节点更靠近网络的核心层。文献 [82] 利用迭代因子来区分同一层级上节点重要性的差异，其数学表达式为：

$$\delta(i)=k_s(i)\times\left(1+\frac{\mathrm{iter}(i)}{\mathrm{iter}(m)}\right) \tag{55}$$

式中，iter(i)为节点i被删除的轮次；iter(m)为迭代的总轮数。进一步同时考虑节点度值以及邻居节点的信息来识别重要节点：

$$\mathrm{IC}(i)=\delta(i)k_i+\sum_{j\in\Gamma_i}\delta(j)k_j \tag{56}$$

根据节点被移除的顺序以及节点的 K-核值，文献 [131] 将邻居节点分为了 Upper （K-核值大于目标节点的邻居节点记为e^u）、Equal Upper （K-核值相同，删除顺序与目标节点的删除顺序相同或后于目标节点，记为e^{eu}）、Equal Lower （K-核值相同，删除顺序与目标节点的删除顺序相同或先于目标节点，记为e^{el}）以及 Lower （K-核值小于目标节点的邻居节点，记为e^l）4 类，基于不同类型的邻居数量来区分同一层级节点的重要性：

$$K_S^{\mathrm{CN}}=\alpha e^u+\beta e^{eu}+\gamma e^{el}+\mu e^l \tag{57}$$

式中，α、β、γ以及μ为自由参数；e^u为属于 Upper 类的邻居节点数量。

除了无法区分同层节点的重要性差异，K-核分解法在不同类型的网络中泛化性较弱。文献 [83] 通过在大量真实网络上的实验发现：根据核心层节点与其他层节点之间的连边多样性可以将核心层节点分为 true-core 群和 core-like 群。其中，true-core 群指处于核心层的节点拥有最高重要性，此类核心层节点往往与网络中其他层节点的连边多样性较大；core-like 群指处于核心层的节点不具有最高的重要性，此类节点与其他层节点的连边多样性较小，从而导致信息传播只局限在小范围内。连边多样性为：

$$H_{k_s}=-\frac{1}{\ln L}\sum_{k_s'=1}^{k_s^{\max}}r_{k_s,k_s'}\ln r_{k_s,k_s'} \tag{58}$$

式中，L为网络中最大的 K-核值；$r_{k_s,k_s'}$是k_s层节点和k_s'层节点之间的平均连边强度。文献 [132] 通过解析证明了 K-核与度信息和 coreness 等指标的关联关系，并且提出用 K-指数方法识别节点重要性。在不同真实网络上的数值分析表明，h 指数可以准确度量节点的传播影响力。

同时考虑节点的全局重要性和局部重要性能够

更全面地反映出节点在网络中的重要性。文献 [37] 指出，在网络连通强度和网络密度不同的情况下，基于全局属性和基于局部属性的节点重要性识别方法各有所长。在网络连通性较强的情况下，K-核分解能够取得较好的表现，而在网络较为稀疏的情况下，邻居度中心性能够取得更好的表现[133]。基于这一发现提出了一种同时考虑网络全局和局部结构的节点重要性识别方法：

$$\mathrm{NGSG}(i)=ak_s(i)+\sum_{j\in\Gamma_i}(ak_s(j)+bk_j) \tag{59}$$

式中，$k_s(i)$为节点i的 K-核值；a和b为根据网络连通强度和网络密度进行调整的自由参数。虽然通过调整参数，该方法可以在网络结构特点不同的情况下取得较为稳定的表现，但确定合适的参数却是一个较为耗时的过程。为此，文献 [134] 在此基础上提出了一种能够取得相近表现的无参数形式，降低了计算复杂度，公式为：

$$k_sd(i)=\sum_{j\in\Gamma_i}(k_s(i)+k_s(j))+\lambda(k_i+k_j) \tag{60}$$

式中，$\lambda=\dfrac{k_s}{d}$。文献 [135] 引入节点度信息和节点的 K-核值来度量节点自身的重要性与邻居节点对于目标节点重要性的贡献。文献 [136] 同样将节点重要性分为全局重要性和局部重要性两方面。通过式 (61) 计算节点 K-核值的熵，以规避 K-核分解方法无法区分同一层节点重要性差异的缺陷并将其作为节点的全局重要性。

$$E_i=-\sum_{j=1}^{k_s^{\max}}p_i(x_j)\log_2 p_i(x_j) \tag{61}$$

$$p_i(x_j)=\frac{|x_j|}{\sum_{j=1}^{k_s^{\max}}x_j} \tag{62}$$

式中，$x_j=\{1,2,\cdots,k_s^{\max}\}$为节点$i$邻居的 K-核值，$p_i(x_j)$为节点$i$的邻居节点属于第$j$层的概率。文献 [136] 认为目标节点与邻居节点的相似性越高，目标节点对于邻居节点的归属度就越高，从而就越容易影响邻居节点，因此通过邻居节点相似性来衡量节点的局部重要性，公式为：

$$B_i=\sum_{j\in\Gamma_i}s(i,j) \tag{63}$$

$$s(i,j)=\frac{2w_{ij}+\sum_{t\in\Gamma_i\cap\Gamma_j}w_{it}w_{jt}}{\sqrt{(1+\sum_{t\in\Gamma_i}w_{it}^2)(1+\sum_{t\in\Gamma_j}w_{jt}^2)}} \tag{64}$$

$$w_{ij}=\frac{|\Gamma_i\cap\Gamma_j|}{|\Gamma_i\cup\Gamma_j|} \tag{65}$$

最后通过将全局重要性和局部重要性加权求和得到节点最终的重要性：

$$\mathrm{Influence}(i)=aE_i+bB_i \tag{66}$$

式中，a与b为自由参数，且$a+b=1$，如何确定合适的参数是该算法的一个主要问题。

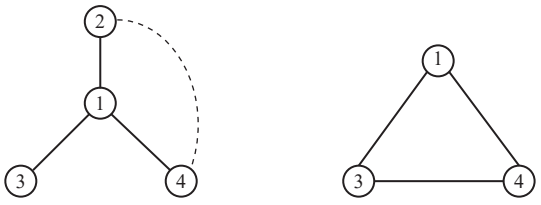
虽然通过结合节点的全局结构和局部结构信息可以改善算法的泛化性能，但赋予不同结构特征的权重需要实验确定，这使得该类方法无法考虑较多的属性。

文献 [137] 提出的结构洞理论指出在网络当中占据结构洞位置的节点在信息传播过程中具有重要作用。图 5 中，相较于图 5b 中 3 个相互连接的节点，图 5a 包含一个结构洞节点 1，其原因在于节点 2 与节点 3 之间若要传播信息必须经过节点 1。节点形成结构洞所受到的约束系数由式 (67) 计算得到：

$$C_i=\sum_{j\in\Gamma_i}(p_{ij}+\sum_q p_{iq}p_{qj})^2\quad q\neq i,j \tag{67}$$

$$p_{ij}=\frac{a_{ij}}{\sum_{j\in\Gamma_i}a_{ij}} \tag{68}$$

式中，a_{ij}为邻接矩阵中第i行第j列的值；p_{ij}为节点i为维持与节点j的邻居关系所投入的精力占总精力的比例。该指标考虑了目标节点的度值以及目标节点与邻居节点的共同邻居，将那些度值较大且与邻居节点形成封闭三角形数量较少的节点定义为更有可能构成结构洞的节点。



a. 1 个结构洞点　　b. 3 个相互连接的节点

图 5　结构洞[138]

约束系数仅考虑了目标节点与其一阶邻居之间的关系，文献 [138] 改进了约束系数公式，进一步考虑了二阶邻居节点的信息，提出了 N-Burt 方法。该方法更准确地反映了节点与邻居节点之间的关系，计算公式为：

$$p'_{ij}=\frac{Q(j)}{\sum_{v\in\Gamma_i}Q(v)} \tag{69}$$

式中，$Q(j)$为节点j的邻居度之和。文献 [139] 基于结构洞理论和网络连通性分别度量了节点的局部重要性和全局重要性，提出了 CumulativeRank（CR）方法。对于局部重要性，由于网络约束系数只考虑一阶邻居信息，使该指标的区分能力较弱，因此引入了改进后的约束系数，公式为：

$$\mathrm{INCC}_i = \sum\nolimits_{j\in\Gamma_i}\left(p'_{ij} + \sum_{k=1,k\neq i,j} p'_{ik}p'_{kj}\right) \tag{70}$$

对于节点的全局重要性，基于节点移除后对网络结构产生的影响来进行度量：

$$R_i = \min\left\{\frac{|i| + m(G-i)}{w(G-i)}, w(G-i)\geqslant 2\right\} \tag{71}$$

式中，$|i|$为移除节点i对网络联通性带来的影响；$m(G-i)$为移除节点i后的最大连通片大小；$w(G-i)$为移除节点i后网络连通片的个数，最后通过将局部结构重要性和全局结构重要性相加获得节点的最终重要性：

$$\mathrm{CR}_i = \frac{\mathrm{INCC}_i}{\sqrt{\sum_{j=1}^{n}\mathrm{INCC}_j}} + \frac{\mathrm{TC}_i}{\sqrt{\sum_{j=1}^{n}\mathrm{TC}_j}} \tag{72}$$

式中$\mathrm{TC}_i = \dfrac{(R_i - R_i^{\min})}{(R_i^{\max} - R_i^{\min})}$为经过最大最小值归一化后的全局重要性，该方法在考虑了局部和全局结构的情况下，保证了较低的计算复杂度。文献 [140] 综合考虑节点的k_s值及其之间的距离，提出了 GSM 方法，数学表示为：

$$\mathrm{GSM}(i) = \mathrm{e}^{\frac{k_s(i)}{n}}\sum_{j\in n, j\neq i}\frac{k_s(j)}{d_{ij}} \tag{73}$$

式中，节点的自身影响力由节点的 K-核值来反映，而节点的全局重要性则取决于节点与网络中其他节点之间的距离以及其他节点的 K-核值。

文献 [77] 基于物理学当中的引力方程，将节点的 K-核值视为质量（Mass），最短路径视为距离，提出了 GC 算法：

$$G(i) = \sum\nolimits_{j\in\Gamma_i}\frac{k_s(i)k_s(j)}{d_{ij}^2} \tag{74}$$

文献 [141] 通过将 K-核值替换为度值，距离替换为目标节点到网络中所有其他节点的最短路径长度对 GC 进行了改进，提出了重力模型（gravity model, GM），公式为：

$$\mathrm{GM}(i) = \sum_{i\neq j}\frac{k_i k_j}{d_{ij}^2} \tag{75}$$

由于 GM 模型需要计算目标节点到所有其他节点之间的最短路径长度，其计算复杂度较高。为解决这一问题，文献 [141] 设计了一种基于局部信息的 GM 模型，公式为：

$$\mathrm{LGM}(i) = \sum_{d_{ij}\leqslant R, j\neq i}\frac{k_i k_j}{d_{ij}^2} \tag{76}$$

式中，R为搜索半径。文献 [142] 在 GC 算法的基础上引入了多层节点的结构属性，其中质量由不同的中心性加权求和得到，而权重通过计算信息熵来获得。文献 [79] 认为现实中节点i到节点j之间的距离不一定总相等，但 GC 中的欧式距离则存在距离相等的假设。通过使用有效距离[143] 对欧式距离进行替换，对 GC 进行了改进，公式为：

$$C_{\mathrm{EffG}}(i) = \sum_{j=1,j\neq i}^{n}\frac{k_i k_j}{D_{j|i}^2} \tag{77}$$

式中，$D_{j|i}$为节点i到节点j的有效距离，定义为：

$$D_{j|i} = 1 - \log_2(P_{j|i}) \tag{78}$$

$$P_{j|i} = \frac{a_{ij}}{k(i)}, (i\neq j) \tag{79}$$

式中，a_{ij}为邻接矩阵中第i行第j列的值，若节点i与j之间存在多条路径，默认使用最短路径。文献 [144] 在衡量节点质量的同时考虑了节点的度信息、K-核值以及特征向量中心性。

该节介绍的 MA 算法可分为 4 类：基于迭代的算法、基于 K-核分解的算法、基于全局和局部结构算法及基于引力模型算法。在当前信息和数据爆炸式增长的时代，社交网络规模不断地扩大，使得 MA 算法的使用场景较为局限。具体地，MA 算法还面临对节点重要性排序问题。基于 K-核分解的算法更多的关注点在于如何区分同层节点的重要性差异。现有研究已经指出，在网络的核心层存在着 core-like 节点，该类节点虽然具有较高的 K-核值，但其重要性较低。识别出此类 core-like 节点可以提升节点重要性识别的准确性。

其次，基于引力模型的方法涉及到了距离的计算，具有较高的计算复杂度，如何提升基于引力模型算法的效率仍然是一项挑战。

最后，越来越多的研究者通过结合全局结构属性和局部结构属性来提升算法的泛化性能，但其局限性在于权重需要通过大量实验进行人为确定，当涉及多个属性时，调整该类方法的参数会产生较大的计算复成本。本节介绍的代表性算法及 4 类算法优缺点见表 6 与表 7。

表 6　基于网络全局结构属性的节点重要性识别方法

方法	优势	劣势	时间复杂度
BC[23]	能够识别网络中对于信息流具有较强控制力的节点	计算复杂度高，不适用于大规模网络	$O(n^3)$
CC[122]	利用节点间的距离反映出节点之间的接近程度	计算复杂度高，不适用于大规模网络	$O(n^3)$
PageRank[24]	利用迭代的思想聚合邻居节点的信息，计算复杂度较低	当含有出度为0的节点时，难以收敛	$O(m)$
LeaderRank[45]	对PageRank进行了改进，无需调整参数，准确度高于PageRank	只适用于有向网络	$O(m)$
K-shell[81]	考虑了节点在网络中所处的位置，能够比度中心性更准确地发现重要节点	难以区分同层节点重要性的差异	$O(n)$
CR[139]	基于结构洞理论和网络连通性分别度量节点的局部和全局重要性，比网络约束系数更全面	无法适用于同质网络	$O(n^2+nk^2)$
LGI[136]	利用熵区分了具有相同K-shell值的节点重要性差异，并考虑了节点之间的相似性	难以确定权重参数	$O(n^2+m)$
IKs[127]	更细粒度地将节点划分到不同的层级中，提高了K-核分解的区分能力	不适用于随机网络	$O(n)$
GSM[140]	考虑了节点自身重要性及与网络中所有其他节点之间的关系	不适用于大规模网络	$O(n^2)$

表 7　部分 MA 方法及其优缺点

方法流	相关工作	优点	缺点
基于迭代的算法	PageRank[24]; LeaderRank[45]; EC[123]; WleaderRank[127]	通过迭代方式聚合高阶邻居信息比直接利用节点距离更高效	邻居节点对于目标节点的重要性贡献主要采用节点度进行量化，未考虑其他结构属性
基于K-核分解算法	DSC[36]; IC[82]; Link entropy[83]; θ[84]; RAD[125]; IKs[127]; C_{nc}[129]; CN[131]	考虑到了节点的全局位置信息，在连通性较强的网络中表现比MI方法更好	具有最高K-核值的节点不一定具有最高的重要性
基于全局和局部结构算法	NGSC[37]; ksd[131]; Influence[136]	具有较强的泛化性能	确定权重计算复杂度高
基于引力模型算法	GC[76]; CeffG[79]; DCC[94]; GM[141]	同时考虑了节点的位置与节点之间交互的信息	距离计算具有较高的复杂度

5　基于机器学习的算法

在各学科交叉发展的大环境下，机器学习与社交网络的结合越来越紧密[145]。究其原因，一方面是因为现实中存在着大量需要用网络结构表示的数据，社交网络的研究有助于机器学习研究人员更有效地利用网络结构数据去解决实际问题[146]。另一方面，机器学习方法能够挖掘出网络拓扑结构更深层的信息，从而为社交网络的研究提供支持[147]。近年来，社交网络的众多分支中都开始出现基于机器学习的方法，如节点分类[148-149]、链路预测[150-151]、网络统计特征提取[152] 等。作为网络科学的核心研究问题之一，节点重要性研究中也开始应用机器学习方法，主要分为基于统计机器学习模型（statistical-machine-learning-based, SMLB）和基于深度学习（deep-learning-based, DBL）两类。

与中心性方法不同，支持向量机、决策树、逻辑斯蒂回归、线性回归等经典机器学习模型可同时根据节点的多维结构特征来完成分类或回归任务。相较于仅凭一种中心性指标，机器学习模型能更全面地考虑节点多方面的结构信息。但不同结构属性对于模型精度的贡献有所不同，因此选择哪些属性作为模型的输入就成为了提高 SMLB 方法准确性的关键。文献 [85] 基于支持向量机模型，选择了度中心性、一阶邻居度中心性、二阶邻居度中心性、接近中心性、K-核值、PageRank 以及特征向量中心性进行实验。通过大量实验发现：两种互补的中心性指标结合之后能够取得比单一中心性更高的识别准确率。其中，度中心性与特征向量中心性互补；一阶邻居度中心性与特征向量中心性、接近中心性及 K-核值互补。文献 [153] 利用主成分分析法[154] 检验了度中心性、介数中心性、拉普拉斯中心性[155] 以及网络约束系数 7 个指标对于节点重要性的贡献率。发现在 7 个不同的网络中，拉普拉斯中心性与网络约束系数对于节点重要性均有显著贡献。文献 [19] 基于结构洞理论，将网络约束系数、介数中心性、等级度、效率[116]、网络规模、PageRank 值以及聚类系数这 7 个指标作为节点的特征，使用 ListNet 算法[156] 对节点进行排序。该方法虽然综合使用多种指标提高了节点重要性识别的准确率，但计算复杂度非常高，无法适用于大规模网络。文献 [157] 将节点重要性识别转化为分类任务。在选择特征时，将能够反映节点局部和全局属

性的 9 种中心性指标以及 SIR 模型中的传播概率 β作为输入，使用朴素贝叶斯、随机森林、支持向量机等分类模型来完成节点重要性识别。在建模的过程中，由于直接将 SIR 模型模拟得到的结果为连续值，根据式 (80) 对标签进行了调整：

$$L_i = \frac{S_i - S_{\min}}{|S|} + 1 \tag{80}$$

式中，S_i为由 SIR 模型得到的节点i的传播规模；$S_{\max}$为所有节点中最大的传播规模；$S_{\min}$为所有节点中最小的传播规模；$|S| = \frac{S_{\max} - S_{\min}}{N}$，$N$表示节点影响力值集合的大小，即标签的个数。文献 [158] 提出了一种结合网络嵌入算法和分类机器学习模型来补充种子节点集的框架 InfEmb。基本流程为：1）利用网络嵌入算法将节点映射为低维向量；2）基于正负样本训练分类模型；3）将未加入种子节点集的l个属于正样本可能性最高的节点补充到种子节点集中。其中，正样本是指已知的种子节点，而负样本是根据节点的度值来进行选择的，度值越小被选为负样本的概率越大。这一策略的最大短板在于需要预先确定若干个种子节点。文献 [159] 认为处于局部结构紧密的节点能更高效地传播信息[81]，节点之间的路径长度无法有效反映出节点的局部结构紧密性，为了计算节点之间的欧式距离，文献 [159] 利用 DeepWalk 算法[160] 将网络中的节点映射为低维向量，并引入了 K-核值，考虑节点在网络中所处的位置提出了 NCL 方法，计算公式为：

$$\mathrm{NCL}(i) = \sum_{j \in \Gamma_i} k_s(i) \times \mathrm{e}^{-|x_i - x_j|^2} \tag{81}$$

式中，$k_s(i)$为节点i的 K-核值；x_i为由 DeepWalk 算法获得的节点i的低维向量表示。

除了网络拓扑结构属性外，还可以根据不同的使用场景考虑节点的非拓扑结构属性。文献 [161] 考虑了 Twitter 用户的发文数量、转发数量等数据，使用线性回归模型来衡量 Twitter 用户的重要性。为了提高 K-核分解抗干扰能力较弱的问题，文献 [85] 利用集成学习[162] 的思想，首先基于原始网络生成多个扰动后的网络，再分别计算原始网络和扰动后网络中节点的重要性得分，最后以不同网络中节点的得分均值来对节点进行重要性排序，有效增强了 PageRank 及 K-核分解算法的抗扰动能力，并且几乎没有增加计算复杂度。

使用传统机器学习模型进行节点重要性识别有两个主要的劣势：1）在建模之前需要进行繁琐的特征选择。当所选的特征需要用到网络全局结构属性获得时，该类方法的计算复杂度会大大增加；2）统计机器学习模型持续学习能力较弱，当数据发生变化需要重新训练整个模型。随着深度学习的快速发展，端到端的深度学习方法逐渐成为越来越多研究者进行重要节点识别的工具，其原因在于：基于深度学习的方法（DBL 方法）利用的是节点有限阶邻居的信息来获取节点的低维向量表示以训练模型，无须进行繁琐的特征工程且精度更高。

图结构数据不像图片、音频等欧式数据，可直接使用卷积神经网络或循环神经网络进行训练。针对图结构数据的特点，研究人员开发出了图神经网络。具体地，图神经网络可分为图循环神经网络和图卷积网络及空间–时序图神经网络[163]。在节点重要性识别中，图卷积神经网络因其较强的性能而被广泛使用。从节点嵌入表示角度，图卷积神经网络可进一步分为直推式（transductive）学习算法和归纳式（inductive）学习算法，直推式学习是指基于给定的网络学习每个节点的向量表示，当网络结构改变后，需要重新进行学习；归纳式学习是指通过在特定网络训练后，对未知节点也可以完成低维向量表示。文献 [164] 基于类比的思想将卷积神经网络应用到图结构数据中，采用图标注的方法，将在网络中处于相似位置的节点映射为相近的表示。基本流程为：1）根据中心性指标来获取每个节点固定大小的邻域；2）对邻域网络进行排序编号；3）进行参数共享。图卷积网络（graph convolutional network, GCN）方法[165] 利用网络的归一化拉普拉斯矩阵作为参数，经过特征变换与聚合获得节点的低维向量表示，该方法的设计简单且被广泛应用。与上述直推式学习的网络嵌入算法不同，GraphSAGE 算法[166] 使用随机游走的思想来获取节点的邻域网络，然后利用聚合函数来聚合邻居节点的结构信息，从而学习节点的向量表示，利用该方法训练的模型能够应用于未知网络。

文献 [89] 首先基于用户的行为日志采用随机游走的方法生成具有固定大小的子网络，其中用户的行为日志中记录了用户的关注、转发及评论行为。每一次随机游走会以目标节点或目标节点邻居中的活跃邻居作为起点，然后根据基于用户行为日志计算的连边权重迭代地遍历起始点的邻居，在遍历过程中每个节点都有一定的概率跳回到起始节点，当

遍历的邻居数达到了预设子网络的大小时则停止遍历。接着，将子网络输入预训练好的嵌入表示层中将子网络的拓扑结构编码为低维向量。为了使模型更关注用户在潜在空间中的相对位置而不是绝对位置，采用式 (82) 对用户嵌入表示进行归一化：

$$y_{u_i}=\frac{x_{u_i}-u_i}{\sqrt{\sigma_i^2-\varepsilon}} \tag{82}$$

式中，u_i和σ_i分别为用户i嵌入表示的均值和方差；ε为用于保证数值稳定的较小值。接着将用户的低维向量表示输入 GCN 中以训练影响力用户识别模型，框架见图 6。

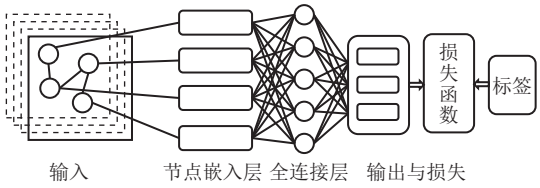


图 6　IDL 框架图

文献 [167] 利用广度优先算法获取每个节点固定大小的邻域网络，将邻域网络对应的归一化化拉普拉斯矩阵以及由目标节点度中心性、中介中心性、接近中心性以及聚类系数组成的特征向量同时输入到图卷积神经网络中学习节点的表示，最后利用全连接层来学习节点重要性识别模型。在使用 SIR 模型生成标签时，由于使用不同的传播概率β产生的标签存在差异且会对模型的训练造成一定影响，文献 [167] 使用辨识力指标选择了合适的传播概率：

$$D=\frac{\mathrm{XH}-\mathrm{XL}}{N(H-L)} \tag{83}$$

式中，根据各节点集合的影响力总和，将节点分为高影响力群体、低影响力群体、最高影响力群体和最低影响力群体。XH为高影响力群体的影响力总和；XL为低影响力群体的影响力总和；H为最高影响力；L为最低影响力；N为高影响力群体占比。文献 [38] 基于广度优先准则，以度值为标准选取每个节点对应的邻域网络，并按度值与邻居节点阶数对邻域网络中的节点重新编码，接着按规则如式 (84) 所示，转换邻域网络对应的邻接矩阵获得每个节点的输入矩阵，最后使用卷积神经网络训练节点重要性回归模型。通过实验发现当训练集与测试集的网络结构相似或训练集网络的平均度较低时，该方法能够取得更出色的表现，框架见图 7。

$$B_u=\begin{cases} a_{0j}k_j & i=0,j=1,2,\cdots,L-1 \\ a_{i0}k_i & i=1,2,\cdots,L-1,j=0 \\ k_i & i=j=0,1,2,\cdots,L-1 \\ a_{ij} & \text{其他} \end{cases} \tag{84}$$

式中，a_{ij}为邻接矩阵中第i行第j列的值；k_i为节点i的度值。

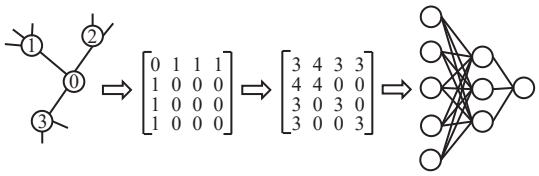


图 7　RCNN 框架图

文献 [39] 利用强化学习算法寻找网络中的重要节点。从网络连通性的角度出发，设定目标函数：

$$\mathrm{Rc}(v_1,v_2,\cdots,v_n)=\frac{1}{n}\sum_{i=1}^{n}\frac{\sigma(G|\{v_1,v_2,\cdots,v_i\})}{\sigma(G)} \tag{85}$$

式中，n 为网络中的节点总数；$\sigma(G)$表示原始网络的连通性。该式度量了去除所识别节点之后网络连通性的变化情况，$\mathrm{Rc}(v_1,v_2,\ldots,v_n)$越小则表明识别出来的节点在网络当中更重要。其思路是将节点重要性识别变成一个马尔可夫决策过程，即通过一系列的状态、动作以及奖励与环境进行交互。其中，环境为所分析的目标网络，状态为残差网络，动作为移除或激活所识别的节点，奖励是降低Rc。为了解决状态和动作表示的问题，使用了图嵌入的方法，并通过引入一个Q函数来决定所采取动作的质量，该方法通过在 200 000 个随机生成的小型人工网络训练之后可以在真实网络上取得良好的表现。

机器学习模型在处理多维特征数据方面的出色表现使其在节点重要性识别领域存在较大的发展潜力。该节介绍的 ML 算法可分为基于统计机器学习的算法与基于深度学习的算法。算法仍面临以下挑战。

1）社交网络用户的重要性不仅取决于其在网络中的结构属性，还与节点的自身属性，如年龄、性别、职业等属性相关。当前的研究多基于节点的拓扑结构属性进行节点重要性识别，对节点自身属性的关注度仍然较少。

2）现有 ML 方法均为有监督学习，意味着需要基于标签进行学习。而现实中，各个节点的影响力往往都是未知的，这就导致在训练此类模型时需要依赖于特定的传播动力学模型来生成标签，如 SIR 模型[46]、IC 模型[168-169]、LT 模型[170] 等。当网络规模较大时，使用以上传播动力学模型生成标签会

耗费大量的时间和计算资源。

3）当训练集网络的结构与测试集网络的结构存在显著差异时，此类模型的表现往往波动较大。因此，如何使用小规模网络训练此类模型且得到较强的泛化能力是未来 ML 算法需要解决的重要问题。

4）当前算法使用的是一组节点结构属性来训练出一个具有较高准确率的模型。但随着属性的增加，模型的计算复杂度也会大大提高。因此如何在准确性和效率之间权衡也是未来需要解决的问题。该节介绍的代表性算法和 2 类算法优劣势见表 8 和表 9。

表 8　基于机器学习的节点重要性识别方法

方法	类型	优势	劣势	机器学习模型
InfluenceRank[156]	经典机器学习	同时考虑了网络拓扑结构属性和社交网络用户的互动信息	特征选择过于繁琐，难以确保所选特征能够最准确地反映节点重要性	Liner Regression
InfEmb[158]	经典机器学习	基于DeepWalk算法获得节点的低维表示，能够根据给定的部分种子节点样本来找出其余种子节点	当给定的正样本过少时，该策略无法准确找出其余种子节点；在选择负样本时仅考虑了节点的度值	DeepWalk、SVM
P&C[85]	集成学习	利用集成学习的思想增强了节点重要性识别方法的抗扰动性能	计算复杂度过高，且依赖于网络扰动算法	Bagging
IDL[88]	深度学习	基于图卷积神经网络获得节点的拓扑低维表示，并结合实际应用场景考虑节点的非拓扑结构属性	仅适用于社交网络	GCN
InfGCN[162]	深度学习	将图卷积神经网络学习得的低维向量表示和节点中心性结合	计算复杂度过高，无法适用于大规模网络	GCN
RCNN[38]	深度学习	仅利用节点度来转换节点的表示，计算复杂度相对于其他基于深度学习模型的方法更低	需要保证训练网络和预测网络的结构相似	CNN
FINDER[39]	强化学习	利用强化学习的思想来进行节点重要性识别	训练网络集较为庞大，计算复杂度高	Reinforcement Learning

表 9　基于深度学习的节点重要性识别方法

方法流	相关工作	优势	劣势
SMLB	文献[148]; 文献[19]; 文献[157]; InfEmb[158]; NCL[159]; InfluenceRank[156]; P&C[85]	可通过数据学习得到每个属性的权重，可解释性较强	需要较为耗时的特征工程，无法持续学习 当训练集分布和测试集分布差异较大时，模型的性能较为不稳定，容易过拟合，可解释性较差。当训练集分布和测试集分布差异较大时，模型的性能较为不稳定
DLB	IDL[88]; InfGCN[162]; RCNN[38]; FINDER[39]	训练模型前无需特征工程，可持续学习，自定义损失函数	

6　传播动力学模型

为了测试节点重要性识别算法的有效性，研究人员需要获得网络中节点的真实重要性作为对照。但在解决实际问题时，节点的真实重要性难以获取或未知。因此，通常采用传播动力学模型来模拟真实的传播过程。由于在不同类型的社交网络中，各节点的重要性会同时受到节点结构信息、节点行为和传播特点等多方面因素的影响[171]。如在疾病传播过程中，部分传染病在治愈后仍有可能重复感染，同时也有部分传染病在恢复后就不会再次感染。针对上述两种不同传播特点，使用 SIS 模型和 SIR 模型模拟出的节点重要性不同。某些节点在 SIS 模型下被认为是重要节点，但可能在 SIR 模型下为非重要节点。因此，在选择传播动力学模型时，传播特点是一个重要因素。文献 [172] 通过对比基于介数中心性、度中心性、h-index 以及接近中心性的传播与免疫过程发现：基于介数中心性的疾病传播范围最广且在无标度网络中的免疫效果最好。但在其他网络中，基于度中心性的模型的免疫效果最好。其结果表明网络结构特点对于传播过程同样存在重要影响。在舆情传播中，信息源的可靠性影响着个体的重要性，即个体的重要性一定程度上取决于其发布信息的可靠性。此外，个体在不同线上平台的活跃度与行为也存在差异，即个体可能只在其常用的社交在线平台上发表言论，而在其他社交平台上不发布任何信息。因此，选择不同社交平台衡量得出的个体重要性会存在差异，应根据节点的行为特点，选择合适的社交网络与传播动力学模型。综上，为确保实验结果的可靠性，需要细致分析与探讨目标场景下节点的行为、传播特点等因

素对于节点重要性的影响，从而选择合适的传播动力学模型。本节简要介绍用来测试节点重要性识别算法的常用传播动力学模型 SIS 模型[173]、SIR 模型[46]、独立级联模型[92]、线性阈值模型[92] 以及这些模型的拓展。

6.1 SIS 模型

SI 模型假定节点被感染后，其状态就不会发生任何转换。SIS 模型在 SI 模型的基础上考虑了被感染个体再次转变为易感染个体的情况。在 SIS 模型当中，每个节点都会处于易感染状态 S（susceptible）和感染状态 I （infected）两种状态中的一种。与 SI 模型不同的是：在传播过程中每个被感染节点会以一定概率重新转变为易感节点，并且有再次被感染的风险。当网络中没有新增的易感节点时，传播过程停止，网络中处于感染状态节点的数量为种子节点集的传播能力。

6.2 SIR 模型

SIR 模型在 SI 模型的基础上，增加了免疫状态，即基于 SIR 模型模拟传播过程时，网络中的节点会处于易感染状态 S （susceptible）、感染状态 I （infected）以及免疫状态 R （recovered）3 种状态中的一种。其中，处于易感染状态的节点会以概率 β 被处于感染状态的邻居节点所感染，而处于感染状态的节点会以概率 r 转化为免疫状态，处于免疫状态的节点不会再次被感染。最后，当传播过程达到稳态后，处于感染状态的节点数和处于免疫状态的节点数为种子节点的影响力。

6.3 线性阈值模型

线性阈值模型被用于有向网络。在线性阈值模型中，有向网络的每条连边都会被赋予权重，如 w_{ij} 表示节点 i 指向节点 j 的一条连边的权重，其中 $w_{ij}=\dfrac{1}{k_j}$。该权重代表节点 i 在节点 j 的所有入邻居当中的重要性。在传播初始阶段，有一小部分节点处于激活状态。在每一时刻 t，若一个未激活节点与所有处于激活状态的入邻居连边权重之和大于传播阈值 θ，该节点会被激活，否则不会被激活。当网络中没有新的激活节点增加时，传播停止。

6.4 独立级联模型

独立级联模型同样被用在有向网络中，该模型将每个节点的状态分为激活（active）或未激活（inactive）。在传播模拟过程当中，每个处于激活状态的节点都会以一定概率 p 激活其处于未激活状态的邻居节点。若一个处于未激活状态的节点有多个处于激活状态的邻居节点，这些邻居节点则会按随机顺序依次独立的尝试激活该节点，当没有新的被激活节点时，传播过程停止。

6.5 加权独立级联模型

加权级联模型可以被看作是独立级联模型的一个扩展[174]，该模型同样将每个节点的状态分为激活（active）或未激活（inactive）。在传播模拟过程中，假设节点 i 与节点 j 互为邻居节点，节点 i 在第 t 轮传播过程中被激活，而节点 j 在第 t 轮传播过程中仍然属于未激活节点，那么在第 $t+1$ 轮传播过程中，节点 j 有 $\dfrac{1}{k_j}$ 的概率被节点 i 激活，若节点 i 在 $t+1$ 轮传播过程中有 l 个被激活的邻居节点，节点 j 在第 $t+1$ 轮传播过程中有 $1-(1-1/k_i)^l$ 的概率被激活。

6.6 从众意识级联模型

在现实世界当中，从众心理在传播过程中同样扮演着重要角色。如在社交网络中，用户会根据周围用户对信息认同的情况来选择自己是否跟从。文献 [175] 提出了考虑从众心理的传播模型，将群众对目标个体的认同程度考虑在内。在传播模拟过程中，假设节点 u 与 v 互为邻居节点，节点 i 在第 i 轮传播过程中被激活，而节点 j 在第 t 轮传播过程中仍然属于未激活节点，那么在第 $t+1$ 轮传播过程中节点 j 被节点 i 激活的概率为：

$$1-\prod_{j\in\Gamma_i}(1-\Phi(j)\Omega(i)) \tag{86}$$

式中，$\Phi(i)$ 为节点 i 的影响力；$\Omega(j)$ 为节点 j 的被用户认同的程度。

7 评价指标

7.1 平均影响力

利用传播动力学模型获得每个节点的重要性后，通过将由不同节点重要性识别方法识别出的前 $pn(p\in(0,1))$（n 为节点总数）个节点的平均影响力进行比较[64, 94]，可以检测出各算法在进行节点重要性识别上的表现，其计算公式为：

$$\mathrm{SI}(a)=\frac{\sum_{j\in S}\sigma(j)}{pn} \tag{87}$$

式中，S 为所选节点集合；$\sigma(i)$ 为节点 i 的影响力。除了平均影响力之外，影响规模 $F(t)$ 可以反映出节点重要性识别方法所选择的种子节点的影响力随时间的变化，计算公式为：

$$F(t)=\frac{n_{I(t)}+n_{R(t)}}{n} \tag{88}$$

式中，$n_{I(t)}$为时刻t处于感染状态的节点数量；$n_{R(t)}$为时刻t处于恢复状态的节点数量。

7.2 不精确性

不精确性指标$\varepsilon(p)$[176]通过比较节点重要性识别方法找出的前$pn(p\in[0,1])$个最有影响力节点与pn个真正的最有影响力节点之间的平均影响力差距来检验算法的性能。网络中任意节点i的传播效率M_i定义为被节点i影响的节点数量。为了计算不精确性值，首先需要根据每个节点的传播效率对网络中所有节点进行排序，并选择pn个传播效率最高的节点组成节点集$\delta_{\text{eff}}(p)$，然后选择由节点重要性识别方法x识别出来的pn个最有影响力节点组成节点集$\delta_x(p)$。接着分别计算这两组节点集中节点的平均影响力$M_x(p)$和$M_{\text{eff}}(p)$，最后根据式 (89) 计算不精确性值：

$$\varepsilon_x(p)=1-\frac{M_x(p)}{M_{\text{eff}}(p)} \tag{89}$$

当不精确性值$\varepsilon_x(p)$越接近于 0 时，代表由重要节点识别方法x识别出来的节点集的平均影响力与真正的最有影响力节点集的平均影响力越相近。

7.3 传播规模相对差异

传播规模相对差异$\Delta_y(p)$[177]指由两种不同的节点重要性识别方法识别出来的前$pn(p\in[0,1])$个最有影响力节点的传播总规模相对差距，其数学公式为：

$$\Delta_y(p)=\frac{S_y-S_x}{S_x} \tag{90}$$

式中，S_y为由节点重要性识别方法y识别出来的节点集的传播总规模。当$\Delta_y(p)>0$时，表示由方法y识别出来的节点集的总体影响力要大于由方法x识别出来的节点集的总体影响力。

7.4 肯德尔相关系数

肯德尔相关系数可以用来衡量两个有序列表之间的相似性，是测试节点重要性识别方法性能时常用的指标之一。假设有两个有序列表A和B，每个列表中都包含n个元素，列表A和B中的第i个元素可以组成一个元素对(A_i,B_i)，那么第j个元素对为(A_j,B_j)，当两个排序列表的任意两个元素对排名相同时，即$A_i>A_j$且$B_i>B_j$或$A_i<A_j$且$B_i<B_j$时，则这两个元素对就被认为是一致的，反之则不一致，其数学定义为：

$$\tau=\frac{2(C-D)}{k(k-1)} \tag{91}$$

式中，C为两个有序列表中一致对的数量；D为两个有序列表中非一致对的数量；k为每个有序列表中包含的元素数量。肯德尔系数越接近 1 时，就表明两个有序列表越相似。常用的评价指标中与肯德尔τ相关系数具有相同功能的还有 Jaccard 相关系数[137]：

$$J_c=\frac{|X(c)\cap Y(c)|}{|X(c)\cup Y(c)|} \tag{92}$$

式中，$X(c)$为重要节点识别方法所选择的种子节点集；$Y(c)$为实际中最有影响力的种子节点集。

7.5 单调性

单调性[82]用于衡量节点重要性排序的唯一性，其计算公式为：

$$M(X)=\left(1-\frac{\sum_{i\in I}n_i(n_i-1)}{n(n-1)}\right)^2 \tag{93}$$

式中，n_i为被分配到等级i的节点数量；n为网络中节点的总数；$M(X)$的取值位于 0～1 之间，该值越接近于 1 则表示更少的节点被分配到同一个等级上。许多将节点划分到同一层级的方法往往单调性值较低。

互补累计分布函数[35, 131]与单调性指标的功能相同，同样用于衡量节点排名的唯一性，但是互补累计分布函数描述的是节点在不同排名上的分布情况，数学表达式为：

$$\text{CCDF}(Z)=\text{prob}(Z>z)=1-\text{CDF}(z) \tag{94}$$

式中，CDF(z)指节点的排名小于或等于z的概率。

8 结束语

节点重要性识别研究因其广泛的应用性吸引了众多不同学科研究者的注意力。近年来，基于微观局部结构属性和基于宏观全局结构属性的算法保持着较快的发展速度，不断有新算法和改进的算法被提出。此外，利用中观的社团结构信息和机器学习模型来进行节点重要性识别的算法也越来越多。为了及时掌握节点重要性研究的发展动态，本文从社交网络的视角出发，重点总结了基于微观局部结构、基于社团结构、基于宏观全局结构以及基于机器学习 4 类算法。由于基于社团结构和机器学习的算法受到了越来越多研究者的重视但仍缺少系统的

综述文献。本文着重介绍了上述两类方法的设计思路、优缺点及发展所遇到的瓶颈，为感兴趣的研究者提供参考依据。此外，本文还总结了这一研究方向中常用的传播动力学模型及评价指标。

已有研究表明仅靠单一算法难以在所有类型的网络中均取得较为稳定的表现。对影响算法性能的网络拓扑结构属性进行深入理解，可以帮助使用者在实际应用时选择适合的方法。当网络规模较大且连边较为稀疏时，基于中观结构社团 ME 方法是一个较为合适的选择，其原因在于 ME 方法无需考虑网络的全局结构信息，仅通过局部结构对节点重要性进行估计。当网络规模适中且连边密度较高时，由于基于微观结构的 MI 算法聚焦于微观结构容易忽略节点在网络中所处的位置，尝试基于中观社团结构 ME 或基于宏观网络结构的 MA 方法是一个有效的选择。当网络具有较为清晰的社团结构时，基于中观社团结构的 ME 方法可以帮助使用者利用更多的结构属性来识别重要节点。此外，当拥有节点的多种属性信息时，基于机器学习的 ML 方法能够帮助研究者筛选出较为重要的结构属性以识别重要节点。本文提到算法的优势和劣势见表 10。虽然这一研究方向目前取得了巨大进展，但仍存在一些尚待解决的问题。

表 10　MI 算法、ME 算法、MA 算法及 ML 算法的优劣势

类别	优势	劣势
基于微观结构MI	仅利用节点的局部结构属性估计节点重要性，使得大型网络中的节点重要性识别变得可行	容易将处于网络边缘但局部连接紧密的节点识别为重要节点
基于中观社团ME	有助于提高影响力最大化问题中种子节点集选择的效率; 社团结构属性可以提高中心性方法的准确性	对于社团划分算法依赖性较强
基于宏观结构的MA	有助于纠正MI算法的误差; 在连边密度较强的网络中效果较好	计算复杂度较高; 考虑多属性的MA方法需要人为确定权重
基于机器学习的ML	可自动计算每个属性的权重; 相较于传统算法，能够同时考虑较多节点属性	算法效率方面有待改善; 模型准确性易受训练集和测试集数据分布差异的影响

1）节点重要性识别方法的表现会因为网络拓扑结构变化而产生波动，即同一种方法在不同类型网络中进行节点重要性识别存在性能差异。虽然采用的评价指标类似，但由于缺乏专用于测试的统一数据集，各类节点重要性识别方法通常在不同的数据集和传播动力学模型下进行测试，从而导致算法性能的真实差异难以准确度量，给方法的推广落地造成了阻碍。

2）时序网络与静态网络不同，其节点间的连边关系会随时间推移而发生变化。时序网络中存在部分节点重要性识别方法是以静态网络为基础进行的扩展和延伸，如度中心性、介数中心性以及接近中心性等。基于静态网络设计的方法虽然在静态网络中表现的性能良好，但是这些方法在考虑时间维度和网络拓扑结构演化的情况下如何进一步扩展到时序网络中进行应用是一个值得研究的问题。

3）除了网络的拓扑结构属性外，在解决问题时根据应用场景考虑节点的非结构属性对节点重要性排序同样具有重要意义。如在社交网络中，用户的发文频率、评论数量、互动情况、活跃粉丝数量等属性也是反映用户重要性的重要信息。但目前的节点重要性识别方法多聚焦于网络拓扑结构。如何将节点的其他属性与网络拓扑结构充分结合以更有效全面地解决实际问题是未来有待研究的问题。

4）虽然基于机器学习的节点重要性识别方法展现出了巨大发展潜力，但目前该类方法采用的都是监督学习策略，即在训练时需要用到节点的重要性作为标签，但获得这些标签需要依靠传播动力学模型进行模拟。在面对小规模网络时，利用传播动力学模型得到标签是可行的。但当网络规模较大时，这种策略会耗费大量的时间。此外，由于该类方法需要给定训练集网络进行学习，然后才能在目标网络上进行节点重要性识别。当训练网络的结构与目标网络的结构存在较大差异时，该类方法的表现会受到较大影响。因此，如何确保训练出来的模型具有较强的泛化性能是基于机器学习的节点重要性排序方法需要解决的重要问题。

参 考 文 献

[1] MORONE F, MAKSE H A. Influence maximization in complex networks through optimal percolation[J]. Nature, 2015, 524(7563): 65-68.

[2] ZHOU T. Progresses and challenges in link prediction[J]. iScience, 2021, 24(11): 103217.

[3] WATTS D J, DODDS P S. Influentials, networks, and

public opinion formation[J]. Journal of Consumer Research, 2007, 34(4): 441-458.
[4] MUTHUKRISHNA M, SCHALLER M. Are collectivistic cultures more prone to rapid transformation? computational models of cross-cultural differences, social network structure, dynamic social influence, and cultural change[J]. Personality and Social Psychology Review, 2020, 24(2): 103-120.
[5] JIA J S, LU X, YUAN Y, et al. Population flow drives spatio-temporal distribution of COVID-19 in China[J]. Nature, 2020, 582(7812): 389-394.
[6] BERTOZZI A L, FRANCO E, MOHLER G, et al. The challenges of modeling and forecasting the spread of COVID-19[J]. Proceedings of the National Academy of Sciences of the United States of America, 2020, 117(29): 16732-16738.
[7] LIBEN-NOWELL D, KLEINBERG J. The link-prediction problem for social networks[J]. Journal of the American Society for Information Science and Technology, 2007, 58(7): 1019-1031.
[8] LÜ L Y, ZHOU T. Link prediction in complex networks: A survey[J]. Physica A: Statistical Mechanics and Its Applications, 2011, 390(6): 1150-1170.
[9] LIU X L, LIU J G, YANG K, et al. Identifying online user reputation of user–object bipartite networks[J]. Physica A: Statistical Mechanics and Its Applications, 2017, 467: 508-516.
[10] DAI L, GUO Q, LIU X L, et al. Identifying online user reputation in terms of user preference[J]. Physica A, 2018, 494: 403-409.
[11] LOU T C, TANG J. Mining structural hole spanners through information diffusion in social networks[C]// Proceedings of the 22nd International Conference on World Wide Web. New York: ACM, 2013: 825-836.
[12] LÜ L Y, CHEN D B, REN X L, et al. Vital nodes identification in complex networks[J]. Physics Reports, 2016, 650: 1-63.
[13] HUANG H M, SHEN H, MENG Z Q, et al. Community-based influence maximization for viral marketing[J]. Applied Intelligence, 2019, 49(6): 2137-2150.
[14] BORGE-HOLTHOEFER J, MORENO Y. Absence of influential spreaders in rumor dynamics[J]. Physical Review E, 2012, 85(2): 026116.
[15] CAMPAN A, CUZZOCREA A, TRUTA T M. Fighting fake news spread in online social networks: Actual trends and future research directions[C]//Proceedings of the IEEE International Conference on Big Data. New York: IEEE, 2017: 4453-4457.
[16] DAS K, SAMANTA S, PAL M. Study on centrality measures in social networks: A survey[J]. Social Network Analysis and Mining, 2018, 8(1): 13.
[17] OU Y, GUO Q, LIU J G. Identifying spreading influence nodes for social networks[J]. Frontiers of Engineering Management, 2022, 9(4): 520-549.
[18] 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): 178901.
LIU J G, REN Z M, GUO Q, et al. Node importance ranking of complex networks[J]. Acta Phys Sin, 2013, 62(17): 178901.
[19] 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.
REN X L, LVY L Y. Review of ranking nodes in complex networks[J]. Chinese Science Bulletin, 2014, 59(13): 1175-1197.
[20] 韩忠明, 陈炎, 刘雯, 等. 社会网络节点影响力分析研究[J]. 软件学报, 2017, 28(1): 84-104.
HAN Z M, CHEN Y, LIU W, et al. Research on node influence analysis in social networks[J]. Journal of Software, 2017, 28(1): 84-104.
[21] LIU J Q, LI X R, DONG J C. A survey on network node ranking algorithms: Representative methods, extensions, and applications[J]. Science China Technological Sciences, 2021, 64(3): 451-461.
[22] FREEMAN L C. Centrality in social networks conceptual clarification[J]. Social Networks, 1979, 1(3): 215-239.
[23] FREEMAN L C. A set of measures of centrality based on betweenness[J]. Sociometry, 1977, 40(1): 35.
[24] BRIN S, PAGE L. The anatomy of a large-scale hypertextual Web search engine[J]. Computer Networks and ISDN Systems, 1998, 30: 107-117.
[25] KLEINBERG J M. Authoritative sources in a hyperlinked environment[C]//Proceedings of the 9th Annual ACM-SIAM Symposium on Discrete Algorithms. California: Society for Industrial and Applied Mathematics. [S.l.]: ACM, 1999: 604-632.
[26] 杨剑楠, 刘建国, 郭强. 基于层间相似性的时序网络节点重要性研究[J]. 物理学报, 2018, 67(4): 272-280.
YANG J N, LIU J G, GUO Q. Node importance idenfication for temporal network based on inter-layer similarity[J]. Acta Physica Sinica, 2018, 67(4): 272-280.
[27] YIN R R, GUO Q, YANG J N, et al. Inter-layer similarity-based eigenvector centrality measures for temporal networks[J]. Physica A: Statistical Mechanics and Its Applications, 2018, 512: 165-173.
[28] 郭强, 殷冉冉, 刘建国. 基于 TOPSIS 的时序网络节点重要性研究[J]. 电子科技大学学报, 2019, 48(2): 296-300.
GUO Q, YIN R R, LIU J G. Node importance identification for temporal networks via the TOPSIS method[J]. Journal of University of Electronic Science and Technology of China, 2019, 48(2): 296-300.
[29] 陈诗, 任卓明, 刘闯, 等. 时序网络中关键节点的识别方法研究进展[J]. 电子科技大学学报, 2020, 49(2): 291-314.
CHEN S, REN Z M, LIU C, et al. Identification methods of vital nodes on temporal networks[J]. Journal of University of Electronic Science and Technology of China, 2020, 49(2): 291-314.
[30] 任卓明. 动态复杂网络中节点影响力的研究进展[J]. 物理学报, 2020, 69(4): 18-26.
REN Z M. Node influence of the dynamic networks[J]. Acta Physica Sinica, 2020, 69(4): 18-26.
[31] DAI J Y, WANG B, SHENG J F, et al. Identifying influential nodes in complex networks based on local neighbor contribution[J]. IEEE Access, 2019, 7: 131719-

131731.
[32] SUN H L, CHEN D B, HE J L, et al. A voting approach to uncover multiple influential spreaders on weighted networks[J]. Physica A: Statistical Mechanics and Its Applications, 2019, 519: 303-312.
[33] GALVÃO V, MIRANDA J G V, ANDRADE R F S, et al. Modularity map of the network of human cell differentiation[J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, 107(13): 5750-5755.
[34] GHALMANE Z, EL HASSOUNI M, CHERIFI C, et al. Centrality in modular networks[J]. EPJ Data Science, 2019, 8(1): 15.
[35] GHALMANE Z, CHERIFI C, CHERIFI H, et al. Centrality in complex networks with overlapping community structure[J]. Scientific Reports, 2019, 9(1): 10133.
[36] ZAREIE A, SHEIKHAHMADI A, JALILI M. Influential node ranking in social networks based on neighborhood diversity[J]. Future Generation Computer Systems, 2019, 94: 120-129.
[37] NAMTIRTHA A, DUTTA A, DUTTA B, et al. Best influential spreaders identification using network global structural properties[J]. Scientific Reports, 2021, 11(1): 2254.
[38] YU E Y, WANG Y P, FU Y, et al. Identifying critical nodes in complex networks via graph convolutional networks[J]. Knowledge-Based Systems, 2020, 198: 105893.
[39] FAN C J, ZENG L, SUN Y Z, et al. Finding key players in complex networks through deep reinforcement learning[J]. Nature Machine Intelligence, 2020, 2: 317-324.
[40] HU Y Q, JI S G, JIN Y L, et al. Local structure can identify and quantify influential global spreaders in large scale social networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2018, 115(29): 7468-7472.
[41] LESKOVEC J, KLEINBERG J, FALOUTSOS C. Graph evolution[J]. ACM Transactions on Knowledge Discovery from Data, 2007, 1(1): 2.
[42] LESKOVEC J, LANG K J, DASGUPTA A, et al. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters[J]. Internet Mathematics, 2009, 6(1): 29-123.
[43] YANG J, LESKOVEC J. Defining and evaluating network communities based on ground-truth[J]. Knowledge and Information Systems, 2015, 42(1): 181-213.
[44] 任晓龙, 朱燕燕, 王思云, 等. 在线社交网络结构与区域经济关联性研究[J]. 电子科技大学学报, 2015, 44(5): 643-651.
REN X L, ZHU Y Y, WANG S Y, et al. Online social network analysis and the relation with regional economic development[J]. Journal of University of Electronic Science and Technology of China, 2015, 44(5): 643-651.
[45] LVY L Y, ZHANG Y C, YEUNG C H, et al. Leaders in social networks, the Delicious case[J]. PLoS One, 2011, 6(6): e21202.
[46] HETHCOTE H W. The mathematics of infectious diseases[J]. SIAM Review, 2000, 42(4): 599-653.
[47] YU S B, GAO L, XU L D, et al. Identifying influential spreaders based on indirect spreading in neighborhood[J]. Physica A: Statistical Mechanics and Its Applications, 2019, 523: 418-425.
[48] MCAULEY J, LESKOVEC J. Learning to discover social circles in ego networks[J]. Advances in Neural Information Processing Systems, 2012, 1: 539-547.
[49] BOGUÑÁ M, PASTOR-SATORRAS R, DÍAZ-GUILERA A, et al. Models of social networks based on social distance attachment[J]. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 2004, 70(5 Pt 2): 056122.
[50] JEONG H, MASON S P, BARABÁSI A L, et al. Lethality and centrality in protein networks[J]. Nature, 2001, 411(6833): 41-42.
[51] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. Nature, 1998, 393(6684): 440-442.
[52] CHEN D B, LÜ L Y, SHANG M S, et al. Identifying influential nodes in complex networks[J]. Physica A: Statistical Mechanics and Its Applications, 2012, 391(4): 1777-1787.
[53] XIE N. Social network analysis of blogs[D]. Bristol: University of Bristol, 2006.
[54] NEWMAN M E J. Finding community structure in networks using the eigenvectors of matrices[J]. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 2006, 74(3 Pt 2): 036104.
[55] SPRING N, MAHAJAN R, WETHERALL D. Measuring ISP topologies with rocketfuel[J]. ACM SIGCOMM Computer Communication Review, 2002, 32(4): 133-145.
[56] GUIMERÀ R, DANON L, DÍAZ-GUILERA A, et al. Self-similar community structure in a network of human interactions[J]. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 2003, 68(6 Pt 2): 065103.
[57] LIU Y, TANG M, ZHOU T, et al. Identify influential spreaders in complex networks, the role of neighborhood[J]. Physica A: Statistical Mechanics and Its Applications, 2016, 452: 289-298.
[58] KUNEGIS J. KONECT: The Koblenz network collection[C]//Proceedings of the 22nd International Conference on World Wide Web. Brazil: ACM, 2016: 1343-1350.
[59] NEWMAN M E. The structure of scientific collaboration networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2001, 98(2): 404-409.
[60] LIU J G, WANG Z Y, GUO Q, et al. Identifying multiple influential spreaders via local structural similarity[J]. EPL (Europhysics Letters), 2017, 119(1): 18001.
[61] BARABASI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286(5439): 509-512.
[62] ZHANG J X, CHEN D B, DONG Q, et al. Identifying a set of influential spreaders in complex networks[J]. Scientific Reports, 2016, 6: 27823.

[63] ALBERT R, JEONG H, BARABÁSI A L. Diameter of the world-wide web[J]. Nature, 1999, 401: 130-131.
[64] CHEN D B, GAO H, LÜ L Y, et al. Identifying influential nodes in large-scale directed networks: The role of clustering[J]. PLoS One, 2013, 8(10): e77455.
[65] 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.
[66] 赵之滢, 于海, 朱志良, 等. 基于网络社团结构的节点传播影响力分析[J]. 计算机学报, 2014, 37(4): 753-766.
ZHAO Z Y, YU H, ZHU Z L, et al. Analysis of node communication influence based on network community structure[J]. Chinese Journal of Computers, 2014, 37(4): 753-766.
[67] ZHAO Z Y, WANG X F, ZHANG W, et al. A community-based approach to identifying influential spreaders[J]. Entropy, 2015, 17(4): 2228-2252.
[68] TULU M M, HOU R H, YOUNAS T. Identifying influential nodes based on community structure to speed up the dissemination of information in complex network[J]. IEEE Access, 2018, 6: 7390-7401.
[69] ZACHARY W W. An information flow model for conflict and fission in small groups[J]. Journal of Anthropological Research, 1977, 33(4): 452-473.
[70] GIRVAN M, NEWMAN M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99(12): 7821-7826.
[71] LUSSEAU D, SCHNEIDER K, BOISSEAU O J, et al. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations[J]. Behavioral Ecology and Sociobiology, 2003, 54(4): 396-405.
[72] WANG Y F, YAN G H, MA Q Q, et al. Identifying influential nodes based on vital communities[C]// Proceedings of the IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/ CyberSciTech). New York: IEEE, 2018: 314-317.
[73] TANG L, LIU H. Relational learning via latent social dimensions[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2009: 817-826.
[74] PAN R K, SARAMÄKI J. The strength of strong ties in scientific collaboration networks[J]. EPL (Europhysics Letters), 2012, 97(1): 18007.
[75] MASSA P, SALVETTI M, TOMASONI D. Bowling alone and trust decline in social network sites[C]// Proceedings of the 8th IEEE International Conference on Dependable, Autonomic and Secure Computing. New York: IEEE, 2009: 658-663. .
[76] ZENG A, ZHANG C J. Ranking spreaders by decomposing complex networks[J]. Physics Letters A, 2013, 377(14): 1031-1035.
[77] MA L L, MA C, ZHANG H F, et al. Identifying influential spreaders in complex networks based on gravity formula[J]. Physica A: Statistical Mechanics and Its Applications, 2016, 451: 205-212.
[78] KUMAR A, SNYDER M. Protein complexes take the bait[J]. Nature, 2002, 415(6868): 123-124.
[79] SHANG Q Y, DENG Y, CHEONG K H. Identifying influential nodes in complex networks: Effective distance gravity model[J]. Information Sciences, 2021, 577: 162-179.
[80] LIU J G, LIN J H, GUO Q, et al. Locating influential nodes via dynamics-sensitive centrality[J]. Scientific Reports, 2016, 6: 21380.
[81] KITSAK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. Nature Physics, 2010, 6: 888-893.
[82] WANG Z X, ZHAO Y, XI J K, et al. Fast ranking influential nodes in complex networks using a k-shell iteration factor[J]. Physica A: Statistical Mechanics and Its Applications, 2016, 461: 171-181.
[83] LIU Y, TANG M, ZHOU T, et al. Core-like groups result in invalidation of identifying super-spreader by k-shell decomposition[J]. Scientific Reports, 2015, 5: 9602.
[84] LIU J G, REN Z M, GUO Q. Ranking the spreading influence in complex networks[J]. Physica A: Statistical Mechanics and Its Applications, 2013, 392(18): 4154-4159.
[85] BUCUR D. Top influencers can be identified universally by combining classical centralities[J]. Scientific Reports, 2020, 10(1): 20550.
[86] TIXIER A J P, ROSSI M E G, MALLIAROS F D, et al. Perturb and combine to identify influential spreaders in real-world networks[C]//Proceedings of the Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. New York: ACM, 2019: 73-80.
[87] KLIMT B, YANG Y M. The enron corpus: A new dataset for email classification research[M]//Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2004: 217-226.
[88] PAL S K, KUNDU S M, MURTHY C A. Centrality measures, upper bound, and influence maximization in large scale directed social networks[J]. Fundamenta Informaticae, 2014, 130(3): 317-342.
[89] WANG F, SHE J H, OHYAMA Y, et al. Deep-learning-based identification of influential spreaders in online social networks[C]//Proceedings of the IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society. New York: IEEE, 2019: 6854-6858.
[90] DOROGOVTSEV S N, GOLTSEV A V, MENDES J F F. Critical phenomena in complex networks[J]. Reviews of Modern Physics, 2008, 80(4): 1275-1335.
[91] BAO Z K, LIU J G, ZHANG H F. Identifying multiple influential spreaders by a heuristic clustering algorithm[J]. Physics Letters A, 2017, 381(11): 976-983.
[92] KEMPE D, KLEINBERG J, TARDOS É. Maximizing the spread of influence through a social network[C]//

Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '03. New York: ACM, 2003: 137-146.
[93] CHENG X Q, REN F X, SHEN H W, et al. Bridgeness: A local index on edge significance in maintaining global connectivity[J]. Journal of Statistical Mechanics: Theory and Experiment, 2010, 2010(10): P10011.
[94] YANG Y Z, WANG X, CHEN Y, et al. A novel centrality of influential nodes identification in complex networks[J]. IEEE Access, 2020, 8: 58742-58751.
[95] BERAHMAND K, BOUYER A, SAMADI N. A new centrality measure based on the negative and positive effects of clustering coefficient for identifying influential spreaders in complex networks[J]. Chaos, Solitons & Fractals, 2018, 110: 41-54.
[96] FAN T L, LÜ L Y, SHI D H, et al. Characterizing cycle structure in complex networks[J]. Communications Physics, 2021, 4: 272.
[97] SHEIKHAHMADI A, ALI NEMATBAKHSH M, SHOKROLLAHI A. Improving detection of influential nodes in complex networks[J]. Physica A: Statistical Mechanics and Its Applications, 2015, 436: 833-845.
[98] ZHOU T, LÜ L Y, ZHANG Y C. Predicting missing links via local information[J]. The European Physical Journal B, 2009, 71(4): 623-630.
[99] MACQUEEN J. Some methods for classification and analysis of multiVariate observations[C]//Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: Statistical Laboratory of the University of California, 1967: 281-297.
[100] LIU J, XIONG Q Y, SHI W R, et al. Evaluating the importance of nodes in complex networks[J]. Physica A: Statistical Mechanics and Its Applications, 2016, 452: 209-219.
[101] LI Y X, YANG X Z, ZHANG X W, et al. An improved voterank algorithm to identifying a set of influential spreaders in complex networks[J]. Frontiers in Physics, 2022, 10: 955727.
[102] ZHANG W, YANG J, DING X Y, et al. Groups make nodes powerful: Identifying influential nodes in social networks based on social conformity theory and community features[J]. Expert Systems with Applications, 2019, 125: 249-258.
[103] KUMAR S, PANDA B S. Identifying influential nodes in social networks: Neighborhood Coreness based voting approach[J]. Physica A Statistical Mechanics and Its Applications, 2020, 553: 124215.
[104] MAJI G, MANDAL S, SEN S. A systematic survey on influential spreaders identification in complex networks with a focus on K-shell based techniques[J]. Expert Systems with Applications, 2020, 161: 113681.
[105] GUO C G, YANG L W, CHEN X, et al. Influential nodes identification in complex networks via information entropy[J]. Entropy, 2020, 22(2): 242.
[106] CHEN W, WANG Y J, YANG S Y. Efficient influence maximization in social networks[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2009: 199-208.
[107] CHEN D B, SUN H L, TANG Q, et al. Identifying influential spreaders in complex networks by propagation probability dynamics[J]. Chaos, 2019, 29(3): 033120.
[108] DONG G G, WANG F, SHEKHTMAN L M, et al. Optimal resilience of modular interacting networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2021, 118(22): e1922831118.
[109] CANTWELL G T, NEWMAN M E J. Mixing patterns and individual differences in networks[J]. Physical Review E, 2019, 99: 042306.
[110] PAN Y, LI D H, LIU J G, et al. Detecting community structure in complex networks via node similarity[J]. Physica A: Statistical Mechanics and Its Applications, 2010, 389(14): 2849-2857.
[111] ZHAO Z J, GUO Q, YU K, et al. Identifying influential nodes for the networks with community structure[J]. Physica A: Statistical Mechanics and Its Applications, 2020, 551: 123893.
[112] WEI H, PAN Z S, HU G Y, et al. Identifying influential nodes based on network representation learning in complex networks[J]. PLoS One, 2018, 13(7): e0200091.
[113] YANG J, LESKOVEC J. Overlapping community detection at scale: A nonnegative matrix factorization approach[C]//Proceedings of the 6th ACM International Conference on Web Search and Data Mining. New York: ACM, 2013: 587-596.
[114] HALAPPANAVAR M, SATHANUR A V, NANDI A K. Accelerating the mining of influential nodes in complex networks through community detection[C]//Proceedings of the ACM International Conference on Computing Frontiers. New York: ACM, 2016: 64-71.
[115] GALSTYAN A, COHEN P. Cascading dynamics in modular networks[J]. Physical Review E, 2007, 75(3): 036109.
[116] DOMINGOS P, RICHARDSON M. Mining the network value of customers[C]//Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2001: 57-66.
[117] BLONDEL V D, GUILLAUME J L, LAMBIOTTE R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008, 2008(10): 10008.
[118] WANG Y, CONG G, SONG G J, et al. Community-based greedy algorithm for mining top-K influential nodes in mobile social networks[C]//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2010: 1039-1048.
[119] SHANG J X, ZHOU S B, LI X, et al. CoFIM: A community-based framework for influence maximization on large-scale networks[J]. Knowledge-Based Systems, 2017, 117: 88-100.
[120] CHEN Y C, ZHU W Y, PENG W C, et al. CIM: community-based influence maximization in social

networks[J]. ACM Transactions on Intelligent Systems and Technology, 2014, 5(2): 1-31.
[121] QIU L Q, JIA W, YU J F, et al. PHG: A three-phase algorithm for influence maximization based on community structure[J]. IEEE Access, 2019, 7: 62511-62522.
[122] SABIDUSSI G. The centrality index of a graph[J]. Psychometrika, 1966, 31(4): 581-603.
[123] BONACICH P. Factoring and weighting approaches to status scores and clique identification[J]. The Journal of Mathematical Sociology, 1972, 2(1): 113-120.
[124] LI Q, ZHOU T, LÜ L Y, et al. Identifying influential spreaders by weighted LeaderRank[J]. Physica A: Statistical Mechanics and Its Applications, 2014, 404: 47-55.
[125] MA S J, REN Z M, YE C M, et al. Node influence identification via resource allocation dynamics[J]. International Journal of Modern Physics C, 2014, 25(11): 1450065.
[126] D'SOUZA R M, BORGS C, CHAYES J T, et al. Emergence of tempered preferential attachment from optimization[J]. Proceedings of the National Academy of Sciences of the United States of America, 2007, 104(15): 6112-6117.
[127] LIU Z H, JIANG C, WANG J Y, et al. The node importance in actual complex networks based on a multi-attribute ranking method[J]. Knowledge-Based Systems, 2015, 84: 56-66.
[128] LIN J. Divergence measures based on the Shannon entropy[J]. IEEE Transactions on Information Theory, 1991, 37(1): 145-151.
[129] BAE J, KIM S. Identifying and ranking influential spreaders in complex networks by neighborhood coreness[J]. Physica A: Statistical Mechanics and Its Applications, 2014, 395: 549-559.
[130] LIN J H, GUO Q, DONG W Z, et al. Identifying the node spreading influence with largest k-core values[J]. Physics Letters A, 2014, 378(45): 3279-3284.
[131] LI C, WANG L, SUN S W, et al. Identification of influential spreaders based on classified neighbors in real-world complex networks[J]. Applied Mathematics and Computation, 2018, 320: 512-523.
[132] LÜ L Y, ZHOU T, ZHANG Q M, et al. The H-index of a network node and its relation to degree and coreness[J]. Nature Communications, 2016, 7: 10168.
[133] NAMTIRTHA A, DUTTA A, DUTTA B. Weighted kshell degree neighborhood method: An approach independent of completeness of global network structure for identifying the influential spreaders[C]//Proceedings of the 10th International Conference on Communication Systems & Networks. New York: IEEE, 2018: 81-88.
[134] MAJI G. Influential spreaders identification in complex networks with potential edge weight based k-shell degree neighborhood method[J]. Journal of Computational Science, 2020, 39: 101055.
[135] WANG F F, SUN Z J, GAN Q, et al. Influential node identification by aggregating local structure information[J]. Physica A: Statistical Mechanics and Its Applications, 2022, 593: 126885.
[136] MA T H, LIU Q, CAO J, et al. LGIEM: Global and local node influence based community detection[J]. Future Generation Computer Systems, 2020, 105: 533-546.
[137] BURT R S, KILDUFF M, TASSELLI S. Social network analysis: Foundations and frontiers on advantage[J]. Annual Review of Psychology, 2013, 64: 527-547.
[138] 苏晓萍, 宋玉蓉. 利用邻域“结构洞”寻找社会网络中最具影响力节点[J]. 物理学报, 2015, 64(2): 20101.
SU X P, SONG Y R. Leveraging neighborhood “structural holes” to identifying key spreaders in so cial networks[J]. Acta Physica Sinica, 2015, 64(2): 20101.
[139] ZHANG D Y, WANG Y, ZHANG Z X. Identifying and quantifying potential super-spreaders in social networks[J]. Scientific Reports, 2019, 9(1): 14811.
[140] ULLAH A, WANG B, SHENG J F, et al. Identification of nodes influence based on global structure model in complex networks[J]. Scientific Reports, 2021, 11(1): 6173.
[141] LI Z, REN T, MA X Q, et al. Identifying influential spreaders by gravity model[J]. Scientific Reports, 2019, 9(1): 8387.
[142] YAN X L, CUI Y P, NI S J. Identifying influential spreaders in complex networks based on entropy weight method and gravity law[J]. Chinese Physics B, 2020, 29(4): 048902.
[143] BROCKMANN D, HELBING D. The hidden geometry of complex, network-driven contagion phenomena[J]. Science, 2013, 342(6164): 1337-1342.
[144] LI Z, HUANG X Y. Identifying influential spreaders by gravity model considering multi-characteristics of nodes[J]. Scientific Reports, 2022, 12(1): 9879.
[145] BELKIN M, NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003, 15(6): 1373-1396.
[146] PENG C, WANG X, PEI J, et al. A survey on network embedding[J]. IEEE Transactions on Knowledge & Data Engineering, 2018, 31(5): 833-852.
[147] SILVA T C, ZHAO L. Network-based high level data classification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2012, 23(6): 954-970.
[148] JIE B, ZHANG D Q, WEE C Y, et al. Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification[J]. Human Brain Mapping, 2014, 35(7): 2876-2897.
[149] HALL M, FRANK E, HOLMES G, et al. The WEKA data mining software[J]. ACM SIGKDD Explorations Newsletter, 2009, 11(1): 10-18.
[150] ZHANG M H, CHEN Y X. Link prediction based on graph neural networks[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montreal: ACM, 2018: 5171-5181.
[151] CHEN J Y, ZHANG J, XU X H, et al. E-LSTM-D: A deep learning framework for dynamic network link prediction[J]. IEEE Transactions on Systems, Man, and

Cybernetics: Systems, 2021, 51(6): 3699-3712.
[152] SACCHET M D, PRASAD G, FOLAND-ROSS L C, et al. Elucidating brain connectivity networks in major depressive disorder using classification-based scoring[C]//Proceedings of the IEEE 11th International Symposium on Biomedical Imaging. New York: IEEE, 2014: 246-249.
[153] 胡钢, 徐翔, 张维明, 等. 基于主成分分析的网络节点重要性指标贡献评价[J]. 电子学报, 2019, 47(2): 358-365.
HU G, XU X, ZHANG W M, et al. Contribution analysis for assessing node importance indices with principal component analysis[J]. Acta Electronica Sinica, 2019, 47(2): 358-365.
[154] MOORE B. Principal component analysis in linear systems: Controllability, observability, and model reduction[J]. IEEE Transactions on Automatic Control, 1981, 26(1): 17-32.
[155] QI X Q, DUVAL R D, CHRISTENSEN K, et al. Terrorist networks, network energy and node removal: A new measure of centrality based on Laplacian energy[J]. Social Networking, 2013, 2(1): 19-31.
[156] CAO Z, QIN T, LIU T Y, et al. Learning to rank: From pairwise approach to listwise approach[C]//Proceedings of the 24th International Conference on Machine Learning. New York: ACM, 2007: 129-136.
[157] ZHAO G H, JIA P, HUANG C, et al. A machine learning based framework for identifying influential nodes in complex networks[J]. IEEE Access, 2020, 8: 65462-65471.
[158] IVANOV S, DURASOV N, BURNAEV E. Learning node embeddings for influence set completion[C]// Proceedings of the IEEE International Conference on Data Mining Workshops. New York: IEEE, 2018: 1034-1037.
[159] 杨旭华, 熊帅. 利用网络表征学习辨识复杂网络节点影响力[J]. 小型微型计算机系统, 2021, 42(2): 418-423.
YANG X H, XIONG S. Identification of node influence using network representation learning in complex network[J]. Journal of Chinese Computer Systems, 2021, 42(2): 418-423.
[160] PEROZZI B, AL-RFOU R, SKIENA S. DeepWalk: Online learning of social representations[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2014: 701-710.
[161] NARGUNDKAR A, RAO Y S. InfluenceRank: A machine learning approach to measure influence of Twitter users[C]//Proceedings of the International Conference on Recent Trends in Information Technology. New York: IEEE, 2016: 1-6.
[162] BREIMAN L. Bagging predictors[J]. Machine Learning, 1996, 24(2): 123-140.
[163] WU Z H, PAN S R, CHEN F W, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24.
[164] NIEPERT M, AHMED M, KUTZKOV K. Learning convolutional neural networks for graphs[EB/OL]. [2023-11-10]. http://arxiv.org/abs/1605.05273v4.
[165] ZHANG H Q, LU G Q, ZHAN M M, et al. Semi-supervised classification of graph convolutional networks with Laplacian rank constraints[J]. Neural Processing Letters, 2022, 54(4): 2645-2656.
[166] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. California: ACM, 2017: 1025-1035.
[167] ZHAO G H, JIA P, ZHOU A M, et al. InfGCN: Identifying influential nodes in complex networks with graph convolutional networks[J]. Neurocomputing, 2020, 414: 18-26.
[168] GOLDENBERG J, LIBAI B, MULLER E. Talk of the network: A complex systems look at the underlying process of word-of-mouth[J]. Marketing Letters, 2001, 12(3): 211-223.
[169] GOLDENBERG J , LIBAI B . Using complex systems analysis to advance marketing theory development: modeling heterogeneity effects on new product growth through stochastic cellular automata[J]. Academy of Marketing Science Review, 2001, 9(3): 1-18.
[170] GRANOVETTER M. Threshold models of collective behavior[J]. American Journal of Sociology, 1978, 83(6): 1420-1443.
[171] CHAHARBORJ S S, NABI K N, FENG K L, et al. Controlling COVID-19 transmission with isolation of influential nodes[J]. Chaos, Solitons & Fractals, 2022, 159: 112035.
[172] WEI X, ZHAO J C, LIU S, et al. Identifying influential spreaders in complex networks for disease spread and control[J]. Scientific Reports, 2022, 12: 5550.
[173] COHEN J E. Infectious diseases of humans: Dynamics and control[J]. JAMA, 1992, 268(23): 3381.
[174] PALLA G, DERÉNYI I, FARKAS I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. Nature, 2005, 435(7043): 814-818.
[175] LI H, BHOWMICK S S, SUN A X. CINEMA: Conformity-aware greedy algorithm for influence maximization in online social networks[C]//Proceedings of the 16th International Conference on Extending Database Technology. New York: ACM, 2013: 323-334.
[176] ZHAO X Y, HUANG B, TANG M, et al. Identifying effective multiple spreaders by coloring complex networks[J]. EPL (Europhysics Letters), 2014, 108(6): 68005.
[177] KNIGHT W R. A computer method for calculating Kendall's tau with ungrouped data[J]. Journal of the American Statistical Association, 1966, 61(314): 436-439.

编　辑　叶　芳