

## 基于多领域复杂网络拓扑结构的节点重要度评价方法

刘雁, 饶元

(西安交通大学软件学院社会智能与复杂数据处理实验室, 陕西西安 710049)

**摘要:** 节点的重要度评价对复杂网络上节点的传播影响力具有重要的理论意义和应用价值, 但传统的基于网络位置的方法并未考虑多维指标特征对网络节点重要度的影响, 导致在大型网络的节点重要度评价中, 一般节点的排序结果精度不高. 为此在深入剖析经典的混合度分解算法以及传统重要性排序算法缺陷的基础上, 结合网络节点的全局特征和局部特征对节点进行重要度影响分析, 并将三度影响力原则融入节点的局部特征, 提出一种适用于无向网络的基于多领域复杂网络拓扑结构下的节点重要度评价方法, 即基于聚集系数和邻居特征的混合分解方法(CNMD). 在社交网络、电子邮件网络、协作网络等 10 个领域数据集上的实验结果表明, 相比于 MDD、Eksd 和 MCDWE 等算法, CNMD 方法排序结果的分辨率分别达到了 92.44%、99.99%、98.68% 等, 在 10 个领域数据集上的平均分辨率为 98.73%, 最高分辨率为 99.99%, 最低分辨率为 92.44%, 明显优于对比算法, 可以更有效地应用于大型复杂网络中节点重要度的快速评价与计算.

**关键词:** 复杂网络; 多领域; 全局特征; 局部特征; 三度影响力

**中图分类号:** TP391 **文献标识码:** A **doi:** 10.3969/j.issn.0253-2778.2019.07.003

**引用格式:** 刘雁, 饶元. 基于多领域复杂网络拓扑结构的节点重要度评价方法[J]. 中国科学技术大学学报, 2019, 49(7): 533-543.

LIU Yan, RAO Yuan. A new method node importance evaluation based on multi-domain topology characteristics in complex networks[J]. Journal of University of Science and Technology of China, 2019, 49(7): 533-543.

### A new method node importance evaluation based on multi-domain topology characteristics in complex networks

LIU Yan, RAO Yuan

(Lab of Social Intelligent and Complex Data Processing, College of software, Xi'an Jiaotong University, Xi'an 710049, China)

**Abstract:** Many efforts have been made to evaluate node importance in complex networks. However, some traditional methods based on node position in networks do not take into consideration the influence derived from multiple domain topology features, which leads to the low evaluation precision about node importance. To solve this problem, based on a deep analysis of such traditional methods as mixed degree decomposition (MDD) algorithm, a new method, named cluster and neighbor mixed decomposition method(CNMD), is proposed, which combines the global and local features of the complex network

**收稿日期:** 2018-09-21; **修回日期:** 2018-12-04

**基金项目:** 国家自然科学基金(61741208), 教育部“云数融合科教创新”基金项目(2017B00030), 中央高校基本科研业务费(ZDYF2017006), 2018年中央高校建设世界一流大学(学科)和特色发展引导专项资金(PY3A022), 2018年西安市碑林区科技项目(GX1803), 2019年教育部社科重大项目(18JZD022), 2019年深圳市科技创新项目(JCYJ20180306170836595)资助.

**作者简介:** 刘雁, 女, 1993年生, 硕士生, 研究方向: 复杂网络节点重要性评估. E-mail: ly77xs@163.com.

**通讯作者:** 饶元, 博士/副教授, E-mail: yuanrao@163.com.

topology structure, and adopts in kinds of three-degree influence principle to represent the local features of the node. Extensive experiments on ten kinds of network datasets in different field show that the average resolution, the lowest and the highest resolution of all experimental datasets are 98.73%, 92.44% and 99.99%, respectively, which is obviously better than traditional methods, like MDD, Eksd and MCDWE algorithms. Therefore, CNMD method is not only suitable for multi-scale undirected network topology, but also applicable for evaluating node importance under all circumstances.

**Key words:** complex networks; multi-domain; global features; local features; three degrees of influence

## 0 引言

随着网络的快速发展,各个领域通过复杂网络的形式更加清晰的研究信息传播网络中的动力学机制,复杂网络中节点重要性的研究已经成为各领域的研究核心.例如在传染病的爆发过程中,通传染病的爆发过程中,及时控制感染源、隔离相关易感人群可以减缓灾情的扩散和蔓延<sup>[1]</sup>;社交网络中“意见领袖”对舆情传播的方向和效率会产生重要的影响<sup>[2]</sup>;在通信网络中,各节点发生故障概率相等的条件下,一旦关键节点发生故障,就会导致整个网络无法通信造成巨大损失<sup>[3]</sup>,因此复杂网络中节点重要性的研究具有重要的研究价值与意义<sup>[4-5]</sup>.

传统的节点重要性的评价方法有两种:节点的显著性和网络的破坏性.从节点的显著性的角度来看,节点的属性指标越高,节点的显著性越突出,则节点越重要;从节点对网络造成的破坏性角度来看,删除节点或节点集及其连边后,会形成很多个不连通的网络,因此对网络造成的破坏性影响越大,则节点越重要.

目前,研究节点重要性度量的方法大致可以分为基于网络局部属性的指标、基于网络全局属性的指标、基于网络位置属性的指标、基于随机游走的节点重要性排序这 4 类指标.第一种基于网络局部属性的指标考虑了节点自身的度信息和该节点的邻居节点信息,如度指标、节点的聚集系数,这类指标计算简单,时间复杂度低,可用于大型网络.第二种基于网络全局属性的指标通常有特征向量、节点的紧密度、节点的介数、网络路径等.这类指标一般准确性比较高,但是时间复杂度也高,不适用于大型网络的研究.第三种是基于网络位置属性的指标,通常有  $k$ -shell 分解法以及基于  $k$ -shell 的各种改进算法<sup>[6-7]</sup>,这种方法考虑节点的全局位置信息,且计算复杂度低,因此该方法的使用较为广泛.第四类是基于随机游走的节点重

要性排序方法,典型的方法包括:PageRank、LeaderRank、HITS 等,这类评价指标不仅考虑了个体本身的连接度,同时也考虑了某个节点周围邻居节点影响力大小对其影响力的贡献.由于该指标时间复杂度低,较适用于大型网络,而且比紧密度、介数更能准确识别在信息传播过程中最具有影响力的节点.

传统的节点重要度评价方法存在以下不足:

(I)未结合多维指标特征全面地对网络中节点的重要度进行评价.

(II)适用于大型网络的节点重要度的计算方法,对一般节点排序精度不高.

针对以上不足,本文对现有评价算法进行深入分析,根据节点在网络中的多种特征,提出一个细粒度的评价节点重要度模型,通过模型中参数的学习,分析影响节点重要性的多种因素及其影响权重.

本文的创新之处在于融合了节点的全局特征和局部特征等多个因素,并根据节点的三度影响力原则从网络结构的角度对局部特征进行细粒度评估;通过模型参数学习,本文深入分析了复杂无向网络中节点的重要度评价因素,包括节点的全局重要度、节点的局部邻居影响力、节点的局部聚集系数.

## 1 相关工作

### 1.1 基于全局特征的算法

近年来,复杂网络中基于网络位置的节点重要度量方法由于计算复杂度小,因此使用较为广泛.典型的方法有  $k$ -shell 方法以及改进方法.

Kitsak 等<sup>[8]</sup>基于  $k$ -shell 分解思想( $k$ -shell decomposition),通过剥离网络中每层度数较小的节点,计算出节点的核数,核数较大的点节点的影响力越大. $k$ -核分解法优点在于方法简单且计算复杂度低,研究表明, $k$ -核分解法适用于大规模网络下的节点重要性的计算,但使用该方法对节点的排序结果区分度不大.

Zeng 等<sup>[9]</sup>针对  $K_s$  值相同的节点无法判别节点重要性的缺点,进一步提出了混合度分解法 (mixed degree decomposition method, MDD) 算法. MDD 算法不仅考虑使用  $k$ -shell 分解法移除后剩余节点的信息,还考虑了被移除节点的度,解决了  $K_s$  值相同的节点重要度的计算,更适用于大型网络研究. 简单地说,节点的核值越大,节点在网络中越处于核心地位. MDD 算法如图 1 所示.

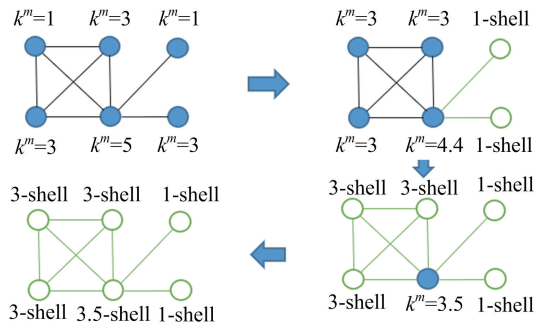


图 1 MDD 分解法示意图

Fig. 1 The core procedure of MDD algorithm

Ahmad<sup>[10]</sup>在  $k$ -shell 的基础上设定了 3 个指标,提出了一种特殊的分层度量核值的方法,该方法在真实世界和人工网络上进行的实验,结果表明该方法比其他方法能更准确地对节点的影响进行排序.

### 1.2 基于局部特征的算法

节点的邻居影响力是节点的局部特征,其中最常见指标是度. 度指标定义为节点的邻居数目,节点度的值越大,即节点的邻居数目越多,则节点可能有更大的影响力. 研究发现,网络中节点的重要性不但与自身的信息具有一定的关系,而且与该节点邻居节点的度也存在一定的关联,即节点的度及其邻居节点的度越大,节点就越重要. 罗仕龙<sup>[11]</sup>在  $k$ -shell 基础上,考虑节点位置和邻域对节点的影响力,提出了一种基于节点位置和邻域的多属性重要性排序方法,计算复杂度为  $O(N)$ . 利用  $k$ -shell 分解法迭代信息进一步区分节点的位置,这种方法比之前的排序方法更高效、准确、合理.

大量研究者通过结合节点的度信息和计算节点的聚集系数度量节点的重要度<sup>[12]</sup>, 研究结果发现,节点的影响力与节点的聚集系数相关联,且聚集系数较高的节点的与节点的影响力并非成正比,甚至起负面作用. 节点聚集系数示意图如图 2 所示. 图 2 是一个具有 4 个节点的网络实例,在 3 种结构下,深色节点的聚类系数分别为 0, 1/3 和 1.

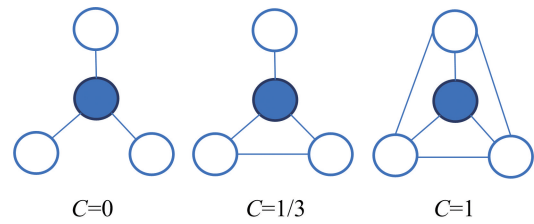


图 2 节点聚集系数示意图

Fig. 2 The representation of node clustering coefficient

### 1.3 三度影响力

Fowler 等<sup>[13]</sup>提出的三度影响力原则认为互相连接的节点之间会互相产生影响,且节点之间的影响力存在三度强连接关系,即节点的邻居节点、节点的邻居节点的邻居节点(二度)、节点二度邻居节点的邻居节点(三度). 如果节点的邻居关系超过三度以上的连接,节点之间的影响力会急剧下降,甚至消失. 节点的三阶邻居示意如图 3 所示.

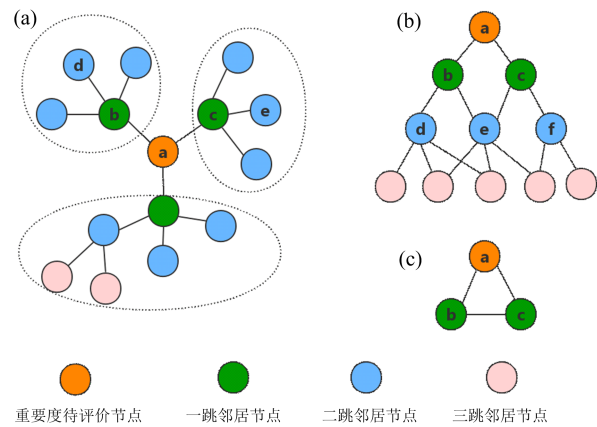


图 3 节点三阶邻居示意图

Fig. 3 The reflection about three-order neighbor of node

## 2 模型构建

给定一个无向复杂网络  $G=(V, E)$ ,  $V$  是节点集合,  $V=\{v_1, v_2, \dots, v_n\}$ ,  $E$  是无向边的集合. 本文结合网络节点的全局特征和局部特征对节点重要度的影响,提出了一种基于多领域复杂网络的拓扑结构下节点重要度评价方法,以计算节点在无向复杂网络中的重要度. 模型分别从全局特征和局部特征对节点的重要度进行评估.

### 2.1 全局特征

$k$ -shell 分解法是通过分析节点在整个网络中的位置,处于核心的节点在网络中更重要,但该方法赋予大量节点相同的值,导致无法区分这些  $K_s$  值相同节点的重要性. MDD 分解法针对这个问题不仅考虑移除后剩余节点的信息,而且考虑被移除节

点的度信息. 本文依据 MDD 方法来计算节点的在网络中的全局特征, 全局重要度的计算公式为

$$G(i) = k^r + \lambda \cdot k_i^e \quad (1)$$

式中,  $G(i)$  表示节点  $v_i$  的全局重要度,  $k^r$  代表残差度, 即连接到剩余节点的链接数量,  $k_i^e$  代表耗尽度, 即连接到已移除节点的链接的数量,  $\lambda$  代表可调节的参数, 值在 0—1 之间, 此处,  $\lambda$  设置为 0.7.

全局重要度的算法描述见算法 2.1.

### 算法 2.1 全局重要度量算法

输入:  $G=(V, E), V=\{v_1, v_2, \dots, v_n\}$ .

输出: 全局重要性值  $G(i)$ .

步骤:

根据  $G$  初始化矩阵  $M$ ;

WHILE SUM(( $M, 2$ ), 1)  $\neq 0$

    求出所有节点的度  $\text{deg}(i)$ ;

    求出所有节点中最小的度  $\min(\text{deg})$ ;

    FOR each  $v_i \in n$

        IF  $\text{deg}(i) = \min(\text{deg})$

    THEN 删除矩阵中编号为  $i$  的连边;

    ELSE continue;

        END IF

END FOR

    求出剩余节点的度  $\text{deg\_now}(i)$ ;

    FOR each  $v_i \in n$

        IF  $\text{deg\_now}(i) = 0$

    THEN break;

        ELSE

$k^r = \text{deg}(i)$ ;

$k_i^e = \text{deg}(i) - \text{deg\_now}(i)$ ;

    结合公式(1)计算  $G(i)$  的值;

        END IF

    END FOR

END WHILE

说明: 算法 2.1 中, 矩阵  $M$  表示  $G$  的无向边邻接矩阵, 其维度为  $N \times N$ , 若两节点之间有连边, 矩阵相应的行列元素为 1, 否则为 0; SUM(( $M, 2$ ), 1) 表示矩阵  $M$  中是否存在非零元素, 当其值不为 0 则代表网络中节点之间存在连边.

## 2.2 局部特征

MDD 混合分解法考虑的只是节点自身在网络中所处的位置, 并没有考虑节点所处的局部环境. 本文进一步将三度影响力原则融入节点的局部特征中, 从节点的三阶邻居影响力和节点的三阶聚集系数这两个局部特征细粒度地衡量节点在网络中的局部重要度.

节点的三阶邻居的算法描述见算法 2.2.

### 算法 2.2 三阶邻居寻找算法

输入:  $G=(V, E), V=\{v_1, v_2, \dots, v_n\}$ .

输出: 一阶邻居矩阵  $\text{Neighbor}_{\text{one}}$ , 二阶邻居矩阵  $\text{Neighbor}_{\text{two}}$ , 三阶邻居矩阵  $\text{Neighbor}_{\text{three}}$ .

步骤:

根据  $G$  初始化矩阵  $M$ , 初始化三个零值矩阵  $\text{Neighbor}_{\text{one}}, \text{Neighbor}_{\text{two}}, \text{Neighbor}_{\text{three}}$ ;

FOR each  $v_i \in n$

    寻找节点  $v_i$  的一阶、二阶、三阶邻居节点集  $N_1, N_2, N_3$ ;

        FOR each  $x_i \in N_1$

$\text{Neighbor}_{\text{one}}(i, :) = N_1$ ;

        END FOR

        FOR each  $y_i \in N_2$

$\text{Neighbor}_{\text{two}}(i, :) = N_2 \setminus N_1 \neq i$ ;

        END FOR

        FOR each  $z_i \in N_3$

$\text{Neighbor}_{\text{three}}(i, :) = N_3 \setminus N_2 \setminus N_1 \neq i$ ;

        END FOR

    END FOR

说明: 算法 2.2 中,  $M$  表示  $G$  的无向边邻接矩阵, 若两节点之间有连边, 矩阵相应的行列元素为 1, 否则为 0;  $\text{Neighbor}_{\text{one}}(i, :), \text{Neighbor}_{\text{two}}(i, :), \text{Neighbor}_{\text{three}}(i, :)$  分别表示一阶、二阶、三阶邻居矩阵中节点  $v_i$  的最终邻居集合.

#### 2.2.1 局部三阶影响力

根据邻居节点的重要性反馈, 如果节点的邻居很重要, 则这个节点重要性可能高. 由于很多方法将节点的重要度平均贡献给邻居节点<sup>[13]</sup>, 实际上节点的直接邻居与间接邻居对节点产生的直接影响和间接影响是不同的, 并且信息在经过每一个邻居后会产生衰减, 因此本文依据节点的一阶节点的直接影响力以及二阶和三阶邻居的间接影响力赋予不同的权重, 而且只考虑每阶邻居中最大影响力的节点对研究节点的影响力. 局部三阶邻居重要度的计算公式为

$$\text{LN}(i) = \sum_{i=1}^3 \alpha_i \cdot \max(G(j_i)) \quad (2)$$

式中,  $\text{LN}(i)$  代表节点  $v_i$  的局部三阶邻居影响力,  $j_i$  分别为节点的三阶邻居节点,  $\alpha_i$  分别代表三阶邻居、二阶邻居、三阶邻居的权重.

节点的三阶邻居度量的算法描述见算法 2.3:

### 算法 2.3 节点三阶邻居度量算法

输入:  $G=(V, E), V=\{v_1, v_2, \dots, v_n\}$ , 全局重要性值  $G(i)$ , 一阶邻居重要性值  $\text{neighbor}_{\text{one}}$ , 二阶邻居重要性值  $\text{neighbor}_{\text{two}}$ , 三阶邻居重要性值  $\text{neighbor}_{\text{three}}$ .

输出:  $n$  个节点局部三阶邻居重要性值  $LN(i)$ .

步骤:

根据  $G$  初始化矩阵  $M$ ;

FOR each  $v_i \in n$

IF  $v \in \text{neighbor}_{\text{one}}$

THEN  $\text{neighbor}_{\text{one}} = \max(G(v))$ ;

END IF

IF  $v \in \text{neighbor}_{\text{two}}$

THEN  $\text{neighbor}_{\text{two}} = \max(G(v))$ ;

END IF

IF  $v \in \text{neighbor}_{\text{three}}$

THEN  $\text{neighbor}_{\text{three}} = \max(G(v))$ ;

END IF

利用  $\text{neighbor}_{\text{one}}$ 、 $\text{neighbor}_{\text{two}}$ 、 $\text{neighbor}_{\text{three}}$  的值, 根据公式(2)更新节点  $v_i$

的  $LN(i)$  值;

END FOR

说明: 算法 2.3 中,  $M$  表示  $G$  的无向边邻接矩阵, 若两节点之间有连边, 矩阵相应的行列元素为 1, 否则为 0.

### 2.2.2 局部三阶聚集系数

节点的聚集系数不仅考虑节点的邻居, 还从层内考虑邻居之间的连接紧密程度. 节点的聚集系数的计算公式为

$$c(i) = \text{edge}(i) / \left( \frac{\text{deg}(i) \cdot (\text{deg}(i) - 1)}{2} \right) \quad (3)$$

式中,  $c(i)$  代表节点  $v_i$  的一阶邻居的聚集系数,  $\text{deg}(i)$  代表节点  $v_i$  的度,  $\text{edge}(i)$  代表节点邻居节点之间的相连的边.

局部聚类系数等于节点的邻居节点之间连边的数量与邻居节点之间可以连边的最大数量之比, 即

$$C(v_i) = \frac{2 | \{e_{jk} : v_j, v_k \in N_{v_i}, e_{jk} \in E\} |}{k_i(k_i - 1)} \quad (4)$$

式中,  $k_i$  为节点  $v_i$  指向其他节点的连边数量与其他节点指向  $v_i$  的连边数量的. 节点的局部聚集系数用于衡量节点的邻居节点之间联系的紧密程度.

由于节点的聚集系数越大, 节点的影响力越小, 故本文采用类似 Min-Max 标准化方法, 将聚集系数的取值与节点影响力正比化, 即

$$C(i) = \frac{\max(C(v)) - \min(C(v))}{c(i) - \min(C(v))} \quad (5)$$

式中,  $C(i)$  代表节点  $v_i$  的聚集系数,  $c(i)$  代表节点  $v_i$  的一阶邻居的聚集系数,  $\max(C(v))$  代表节点集  $V$  中聚集系数的最大值,  $\min(C(v))$  代表节点集  $V$  中聚集系数的最小值.

局部三阶聚集系数的计算公式为

$$LC(i) = \sum_{j=1}^3 \alpha_j \cdot \max(C(j_i)) \quad (6)$$

式中,  $LC(i)$  表示节点的局部聚集系数,  $j_i$  分别为节点的三阶邻居节点,  $\alpha_j$  分别代表三阶邻居、二阶邻居、三阶邻居的权重.

节点的三阶聚集系数度量的算法描述如算法 2.4.

### 算法 2.4 节点三阶聚集系数度量算法

输入:  $G=(V, E)$ ,  $V=\{v_1, v_2, \dots, v_n\}$ , 节点的聚集系数  $C(i)$ , 一阶邻居重要性值  $\text{neighbor}_{\text{one}}$ , 二阶邻居重要性值  $\text{neighbor}_{\text{two}}$ , 三阶邻居重要性值  $\text{neighbor}_{\text{three}}$ .

输出:  $n$  个节点局部三阶聚集系数  $LC(i)$ .

步骤:

FOR each  $v_i \in n$

IF  $v \in \text{neighbor}_{\text{one}}$

THEN  $\text{neighbor}_{\text{one}} = \max(C(v))$ ;

END IF

IF  $v \in \text{neighbor}_{\text{two}}$

THEN  $\text{neighbor}_{\text{two}} = \max(C(v))$ ;

END IF

IF  $v \in \text{neighbor}_{\text{three}}$

THEN  $\text{neighbor}_{\text{three}} = \max(C(v))$ ;

END IF

利用  $\text{neighbor}_{\text{one}}$ 、 $\text{neighbor}_{\text{two}}$ 、 $\text{neighbor}_{\text{three}}$  的值, 根据公式(6)更新节点  $v_i$  的  $LC(i)$  值;

END FOR

### 2.2.3 节点局部重要度

节点的局部重要度为局部三阶邻居影响力和局部三阶聚集系数之和, 即

$$LI(i) = LN(i) + LC(i) \quad (7)$$

式中,  $LI(i)$  代表节点  $v_i$  的局部重要度,  $LN(i)$  代表节点  $v_i$  的局部邻居影响力,  $LC(i)$  代表节点  $v_i$  的局部三阶聚集系数.

### 2.3 节点重要度

综合以上节点全局特征和局部特征, 节点的重要度用节点全局重要度和局部重要度之和表示, 即

$$I(i) = G(i) + LI(i) \quad (8)$$

式中,  $I(i)$  代表节点  $v_i$  的综合重要度,  $G(i)$  代表节点  $v_i$  的全局重要度,  $LI(i)$  代表节点  $v_i$  的局部重要度. 节点重要度算法描述如算法 2.5.

### 算法 2.5 节点重要度度量算法

输入: 全局重要性值  $G(i)$ , 节点的三阶局部聚集系数  $LC(i)$ , 节点的三阶邻居重要性  $LN(i)$ .

输出:  $n$  个节点重要性值  $I(i)$ .

步骤:

FOR each  $v_i \in n$

根据公式(7)计算节点局部重要性  $LI(i)$ ;

根据公式(8)计算节点重要性  $I(i)$ ;

END FOR

### 3 实验和评估

#### 3.1 实验数据和实验方法

##### 3.1.1 实验数据

为了更好地验证 CNMD 算法的有效性,本文分别选取基础设施网络、社交网络、动物网络、传染病

网络、邮件网络、协作网络、代谢网络、通信网络等 5 个完全不同领域的现实无向网络数据进行算法验证.数据集的具体特征描述如表 1 所示.其中,  $D_1 \sim D_{10}$  为 10 个数据集的编号,  $M_1 \sim M_5$  为 5 个领域的编号,  $N$  代表网络节点总数,  $E$  代表网络中节点之间的连边数,  $\langle k \rangle$  代表网络平均度大小,  $\langle c \rangle$  代表聚集系数;  $L$  为节点间平均最短路径长度. Facebook 社交网络<sup>[14]</sup> 拓扑结构如图 4 所示, E-mail 网络拓扑结构图如图 5 所示.

表 1 10 种真实网络的拓扑特征

Tab. 1 Topological features of ten kinds of real networks

数据集	网络类型	领域编号	领域	方向	网络名称	N	E	$\langle k \rangle$	$\langle c \rangle$	L
$D_1$	基础设施	$M_1$	基础设施	无向	Contiguous USA <sup>[17]</sup>	49	107	4.367	0.507	4.367
$D_2$	动物网络	$M_2$	社交	无向	Dolphin <sup>[18]</sup>	62	159	4.176	0.654	2.636
$D_3$	人际社交	$M_2$	社交	无向	Jazz <sup>[19]</sup>	198	2 742	27.697	0.633	2.235
$D_4$	传染病网络	$M_3$	代谢	无向	Infectious <sup>[20]</sup>	410	2 765	13.49	0.436	3.631
$D_5$	邮件网络	$M_5$	通信	无向	Email <sup>[21]</sup>	1 133	5451	9.622	0.254	3.606
$D_6$	社交网络	$M_2$	社交	无向	Facebook <sup>[22]</sup>	4 039	88 234	43.691	0.617	3.693
$D_7$	协作网络	$M_4$	协作	无向	CA-GrQc <sup>[23]</sup>	5 242	14 496	5.531	0.687	6.049
$D_8$	代谢网络	$M_3$	代谢	无向	Reactome <sup>[24]</sup>	6 327	147 547	43.949	0.67	4.214
$D_9$	通信网络	$M_5$	通信	无向	AS <sup>[25]</sup>	6 474	12 572	4.293	0.414	3.705
$D_{10}$	加密通信网络	$M_5$	通信	无向	PGP <sup>[26]</sup>	10 680	24 316	4.554	0.44	7.463

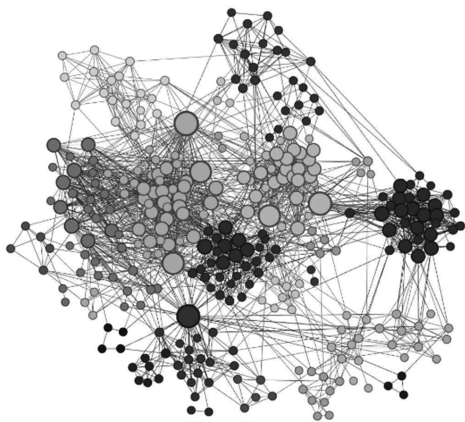


图 4 Facebook 网络拓扑结构图

Fig. 4 The topology of Facebook network

本文选取的 10 个数据集分别从 5 个领域对网络拓扑信息进行统计.具体的 5 个领域的数据集信息见表 2.①基础设施网络是物理基础设施网络.例如道路网络(RO),航空公司连接网络(OF)和电网(UG);②社交网络是指互相接触具有联系的网络,动物网络是动物之间的接触网络,与人类社交网络相当;③代谢网络是模拟代谢途径的网络;④协作网

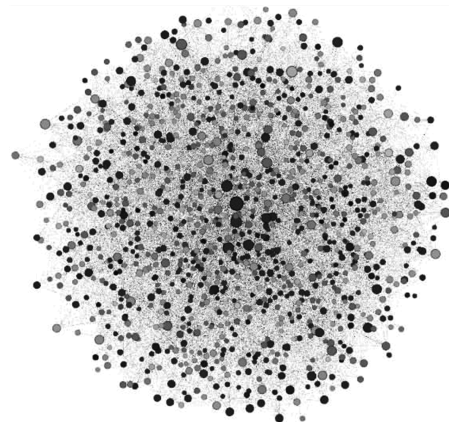


图 5 E-mail 网络拓扑结构图

Fig. 5 The topology of E-mail network

络是各节点之间合作完成作品构成的网络;⑤通信网络包含表示节点之间的消息构成的网络,如电子邮件网络.

##### 3.1.2 对比算法

将本文提出的方法 CNMD 方法和以下的 4 种方法在 10 个不同领域的数据集上进行实验.

(I) degree. 度量邻居节点数的方法,是最基

础的度量方法。

(II) MDD. Zeng 等<sup>[9]</sup>提出的基于  $k$ -shell 分解法的改进算法,该模型比  $k$ -shell 算法取得了更好的节点分类效果,解决了传统大量节点具有相同的  $k$ -shell 值而无法区分节点重要度的问题。

(III) Eksd<sup>[15]</sup>. 针对  $k$ -shell 分解法分辨率不高的问题,提出一种基于  $k$ -shell 和半局部信息的节点重要度的计算方法,该模型提高了排序结果的分辨率。

(IV) MCDWE<sup>[10]</sup>. 一种基于  $k$ -shell 方法改进的融合度信息和核信息评价网络中重要传播节点的方法,该模型重点解决了大量低核节点无法区分节点重要度的问题。

表 2 10 个数据集下各算法的分辨率

Tab. 2 The resolution of five kinds of algorithms under different data sets

数据集	M/ degree	M/ MDD	M/ Eksd	M/ MCDWE	M/ CNMD
$D_1$	0.697 3	0.845 0	0.854 1	0.924 2	0.993 2
$D_2$	0.831 2	0.908 1	0.927 3	0.962 3	0.995 8
$D_3$	0.926 8	0.980 1	0.987 2	0.998 2	0.999 4
$D_4$	0.927 8	0.980 4	0.987 6	0.998 9	0.999 8
$D_5$	0.887 4	0.923 3	0.926 9	0.964 7	0.999 9
$D_6$	0.913 3	0.909 2	0.919 6	0.967 8	0.924 4
$D_7$	0.746 0	0.794 0	0.796 7	0.892 4	0.986 8
$D_8$	0.940 1	0.955 9	0.956 7	0.980 6	0.987 6
$D_9$	0.502 2	0.519 7	0.520 2	0.757 7	0.992 6
$D_{10}$	0.619 3	0.667 3	0.674 2	0.709 3	0.993 2

### 3.2 分辨率指标

Joonhyun<sup>[16]</sup>提出了一种评价不同方法排序所得结果分辨率的指标,即 Monotonicity. 在排序结果中,具有相同排序结果的节点数量越少,那么排序结果的区分度和分辨率就越高. Monotonicity 指标的定义为

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

式中,  $R$  代表通过不同算法得到的排序结果,  $n$  代表网络中节点的总数,  $n_r$  是具有相同排名  $r$  的关系数.  $M(R)$  代表在最终的排序结果中具有相同排序值的节点在总节点中所占的比例。

### 3.3 实验结果及分析

本文选取 degree、MDD、Eksd、MCDWE 四个指标与 CNMD 值进行比较,通过实验验证 CNMD 方法对节点进行排序的结果的效果. 表 3 展示了 degree、MDD、Eksd、MCDWE、CNMD 五种度量方法的 Monotonicity 指标值。

表 3 5 个领域下各算法的分辨率精度

Tab. 3 The resolution precision of five kinds of algorithms in various fields

编号	数据集	M/ degree	M/ MDD	M/ Eksd	M/ MCDWE	M/ CNMD
$M_1$	$D_1$	0.697 3	0.845 0	0.854 1	0.924 2	0.993 2
$M_2$	$D_2, D_3, D_6$	0.890 4	0.932 5	0.944 7	0.976 1	0.973 2
$M_3$	$D_4, D_8$	0.934 0	0.968 1	0.972 1	0.989 7	0.993 7
$M_4$	$D_7$	0.746 0	0.794 0	0.796 7	0.892 4	0.986 8
$M_5$	$D_5, D_9, D_{10}$	0.669 6	0.703 4	0.707 1	0.810 5	0.995 2

#### 3.3.1 算法精度

本文在通过实验结果中节点排序的分辨率验证本算法的性能. 观察图 6 中的数据可以发现,在 10 个相同数据集上使用 degree、MDD、Eksd、MCDWE、CNMD 方法进行排序结果分辨率计算, CNMD 方法在所有数据集上,分辨率的结果为 99.32%、99.58%、99.94%、99.98%、99.98%、99.99%、92.44%、98.68%、98.76%、99.26%、99.32%,均在 92% 以上,最高的分辨率为 99.99%。使用 degree 方法在 10 个网络中进行排序得到的结果分辨率最差; degree 指标只考虑节点的局部信息,因此分辨率较低; MDD 指标考虑节点在全局中的位置,因此 MDD 指标稍好于 degree 值,但由于 MDD 方法存在参数选择问题,不同的参数值对排序结果有较大的影响,因此分辨率不是很高; Eksd 指标得到的排序结果是一种粗粒度的划分方法,导致最终结果中有大量的节点具有相同的排序值,分辨率精度略高于 MDD 方法. 在所有对比方法中 MCDWE 方法的精度的最高,平均为 91.56%。从实验结果可以明显得出,本文方法在社交网络中的结果最好. CNMD 方法在考虑了节点的全局特征和局部特征,在不同的实验数据集上的排序结果平均具有最高的分辨率,仅在社交网络中略低于 MCDWE 方法。

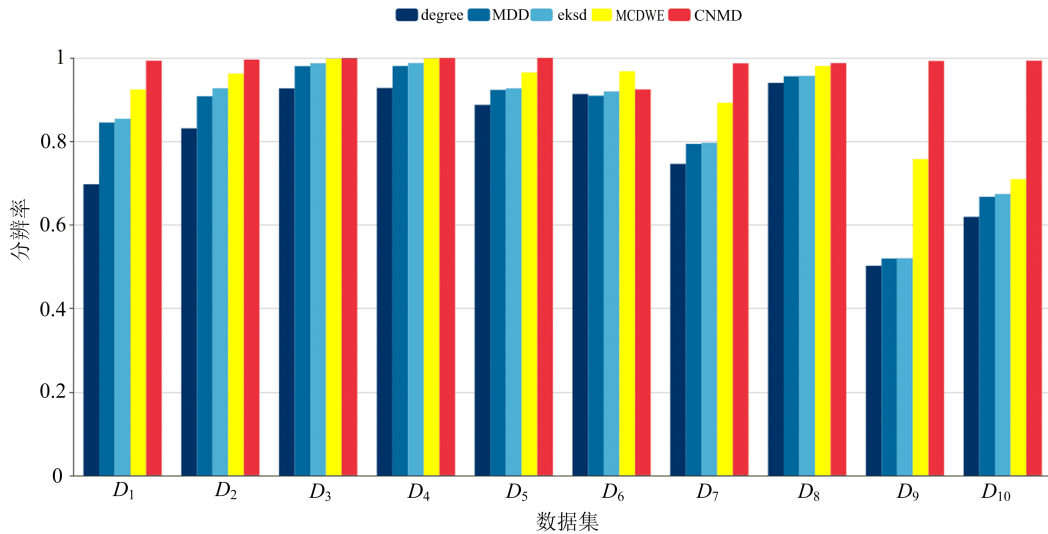


图 6 10 个数据集下不同算法分辨率结果图

Fig. 6 The resolution of five kinds of algorithms under different data sets

在 10 个不同数据集下 CNMD 算法相比其他 4 种算法的分辨率提升精度的对比如图 7 所示. 本算法在 10 个不同数据量的数据集上与 degree、MDD、Eksd、MCDWE 算法的分辨率结果进行比较. 从图 7 结果来看, 本算法相比其他算法提升精度较高的前 4 组数据为  $D_9$ 、 $D_{10}$ 、 $D_1$ 、 $D_7$  数据集,  $D_9$  数据集网络节点个数为 6 474, CNMD 算法相比其他 4 种算法分辨率精度分别提升了 97.65%、90.99%、90.81%、31%, 在所有数据集中精度提升最大.  $D_{10}$  数据集下 CNMD 算法相比其他 4 种算法精度提升了 60.37%、48.84%、47.32% 和 40.03%.

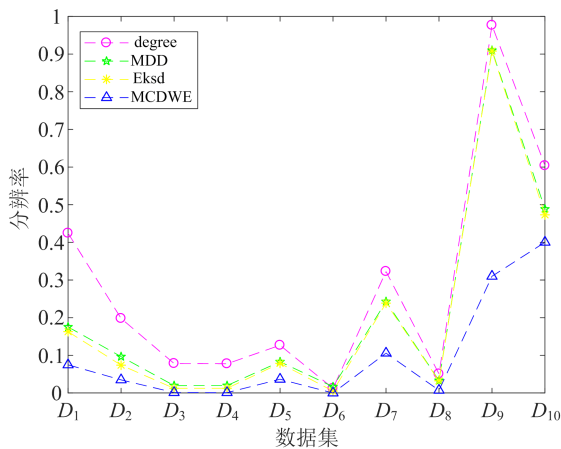


图 7 CNMD 算法相比其他算法提升精度对比图

Fig. 7 The improvement resolution of CNMD compared with other algorithms

### 3.3.2 不同领域算法分辨率的影响

由上可知, 本文提出的 CNMD 算法在 10 个数据集上的精度较高, 本节将进一步对 5 个不同领域

的数据集进行算法验证. 通过对 5 个不同领域中的 10 个数据集分别对同一领域下各算法运行的精度值取平均值, 5 个不同领域的网络在 degree 算法、MDD 算法、Eksd 算法、MCDWE 算法、CNMD 算法 5 种方法的对比值下, 分辨率精度对比情况如图 8 所示.

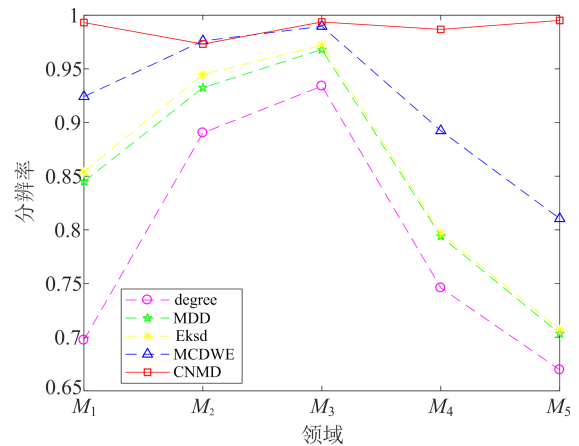


图 8 不同领域与算法分辨率关系图

Fig. 8 The resolution of five kinds of algorithms under different domains

由图 8 可知, 本文提出的 CNMD 在各领域中的分辨率精度高达 97% 以上, 均适用于这 5 种网络. 其他的 4 种算法在  $M_2$ 、 $M_3$  领域的分辨率较高, 但是在  $M_1$ 、 $M_4$ 、 $M_5$  领域下的分辨率明显降低; 对比 4 种度量方法, MCDWE 算法精度略低于 CNMD 算法, 在  $M_2$  领域中占明显优势, 而且明显高于 Eksd、MDD、degree 三种方法. 其中, Eksd 算法精度略高于 MDD 算法, degree 算法的精度在 4 种方法中精



度最低. 总的来看, CNMD 算法在这 5 个领域中都具有较高的分辨率.

### 3.3.3 网络节点数量对算法分辨率的影响

本文从网络的最基本的拓扑特征(网络中包含节点的数量)来分析该特征与 CNMD 算法分辨率之间的关系. 随着网络中节点数据量的增大, 本文提出的 CNMD 算法的分辨率较为稳定, 而且分辨率最高. 网络节点的数量从 49 到 10 680 不等, 其中,  $D_7$  数据集网络节点个数为 5 242, 相比其他 4 种算法精度提升了 32.28%、24.28%、23.86%、10.58%. 总体来看,  $D_9$ 、 $D_{10}$ 、 $D_7$  数据集的网络节点数据量大于 6 000, 在网络节点数目小于 6 000 节点时, 各算法的分辨率均比较高, 当节点数大于 6 000 节点时, degree、MDD、Eksd 三种算法的分辨率精度明显下降, 而且相比 CNMD 算法, 4 种算法的分辨率精度平均只有 73.36%, CNMD 算法的精度可达 99.11%, 而且算法精度提升的最多, 可见 CNMD 方法更适用于大规模数据的网络结构. 在 10 个不同的数据大小的数据集下, 5 种算法的分辨率精度与网络节点数的变化趋势如图 9 所示.

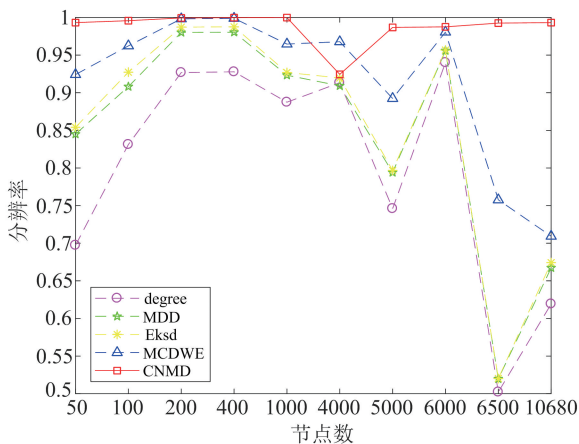


图 9 网络节点数与算法分辨率关系图

Fig. 9 The relationship between the number of nodes and the resolution of different algorithms

### 3.3.4 网络平均度对算法分辨率的影响

网络平均度反映整个网络的稳定程度, 因此本文进一步从网络平均度的特征分析与 CNMD 算法之间的关系. 本算法在不同网络的平均度的大小下, 与 degree、MDD、Eksd、MCDWE 算法的分辨率结果进行比较. 10 个网络的平均度大小从 4.176 到 43.691 不等, 当网络平均度从 4.176 增长到 9.622 时, 5 种算法的分辨率随之增大, 并且平均度为 9.622 时 CNMD 算法的分辨率最大, 此时

为 99.99%, 在 13.49 到 27.697 之间各算法的精确度保持平稳, 随后在网络平均度大于 27.697 之后, MDD 算法、Eksd 算法、CNMD 算法的精度开始下降, 但 degree 方法在网络平均度大于 13.49 之后一直趋于平稳. 总体来看, 本文提出的 CNMD 算法在不同网络平均度大小的情况下, 算法分辨率精度最高, 可见 CNMD 方法更适用于不同平均度规模的网络. CNMD 与其他 4 种算法不同网络平均度下的算法对比情况如图 10 所示.

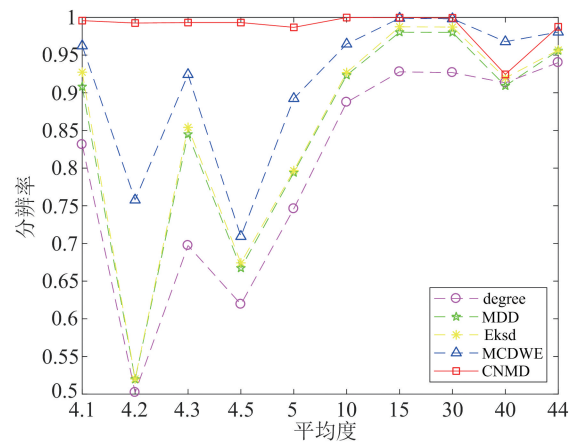


图 10 网络平均度与算法分辨率关系图

Fig. 10 relationship between network average degree and the resolution of different algorithms

### 3.3.5 网络聚集系数对算法分辨率的影响

现有的节点聚集系数研究只考虑一阶邻居之间的密集程度, 本文进一步结合三度影响力原则, 对节点的 3 阶聚集系数与分辨率之间的关系进行详细分析. 本算法在不同网络的聚集系数的大小下, 与 degree、MDD、Eksd、MCDWE 算法的分辨率结果进行比较. 10 个网络的聚集系数大小从 0.254 到 0.687 不等, 在网络聚集系数从最小值 0.254 增大到 0.507 时, 本文提出的算法 CNMD 有下降趋势, 但分辨率精度下降的幅度较小, 本文分析 CNMD 时分辨率较低的原因可能与数据集有关. 其他 4 种算法的分辨率随着网络聚集系数的增加, 频繁起伏波动较大, 可以发现 degree 算法、MDD 算法、Eksd 算法、MCDWE 算法对不同聚集程度网络的节点分辨率太低. 由此可见, 使用节点三阶聚集系数对不同类型网络节点有较强的分辨率. CNMD 方法更适用于不同聚集系数规模的网络. CNMD 与其他 4 种算法不同网络聚集系数下的算法对比情况如图 11 所示.

### 3.3.6 网络平均最短路径对算法分辨率的影响

在 10 个不同网络节点间平均最短路径的数据

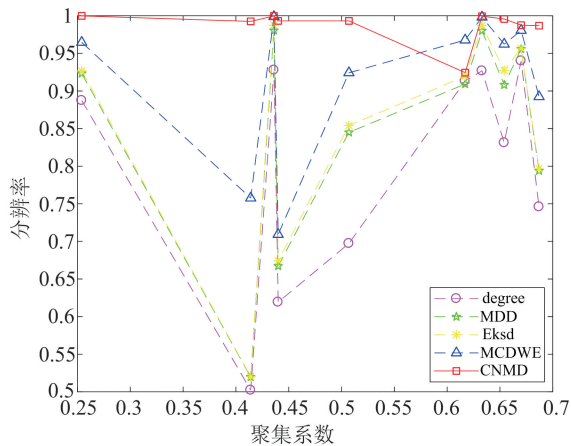


图 11 网络聚集系数与算法分辨率关系图

Fig. 11 The relationship between network aggregation coefficient and the resolution of different algorithms

集下,本文算法的分辨率都显著高于其他 4 种算法.当节点间平均最短路径变化时,本文算法在节点平均最短路径不同取值下,分辨率基本稳定在 92% 以上,最高达到了 99.99%,这也证明本文提出的算法比其他算法更适用于节点平均最短路径变化较大的网络拓节点,虽然在节点平均最短路径为 3.7 时分辨率整体较低,但这并不影响整体趋势.

由图 12 可得,其他 4 种算法的分辨率精度与节点间平均最短路径长度并非呈线性关系,本文提出的算法 CNMD 随着节点间平均最短路径的增加更加趋于稳定.

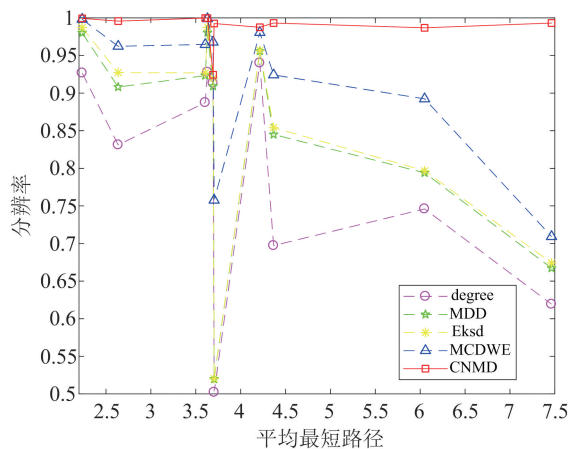


图 12 网络平均最短路径长度与算法分辨率关系图

Fig. 12 The relationship between the average shortest path length of the network and the resolution of different algorithms

本文从网络的 4 种拓扑结构信息分析不同特征对本算法的影响程度,通过 6 组实验表明,本文提出的 CNMD 算法在各种特征的变化下,算法的分辨率精度都在 92% 以上,而且均比较稳定,适用于多种

不同特征的网络.

## 4 结论

在多领域复杂网络的节点重要性研究中,已有的大部分研究是针对特定领域分析节点重要性,采用经典的 MDD 算法在  $k$ -shell 分解法的基础上进一步考虑被移除节点的度信息,区分大量相同核数的节点,但 MDD 算法仅考虑节点的全局特征,未从全局和局部对网络中的节点重要性进行全面考量.本文首先提出节点在网络中的全局特征和局部特征,进一步在局部特征的基础上融合三度影响力原则,提出一种基于多领域复杂网络的度空间尺度下节点重要度评价方法,从网络拓扑结构的角度进行建模,对节点的邻居节点量化分析.在对 5 个不同邻域下的 10 个不同的网络拓扑结构的数据集进行分析后,大量实验结果表明,本文所提出的模型在节点重要度的分辨率优于其他代表性同类模型或算法,本算法在不同邻域的数据集下分辨率都较高,更适用于多领域网络.由于现实网络节点的规模通常较大,这要求算法具有更高的大数据处理能力,所以本文下一步的工作将集中减少计算过程的时间复杂度.

## 参考文献(References)

- [1] WANG W, TANG M, EUGENE S H, et al. Unification of theoretical approaches for epidemic spreading on complex networks [J]. Reports on Progress in Physics, 2017, 80(3): 036603.
- [2] DE A, BHATTACHARYA S, GANGULY N. Demarcating endogenous and exogenous opinion diffusion process on social networks[C]// International World Wide Web Conference. Lyon, France: ACM, 2018: 549-558.
- [3] 孔江涛,黄健,龚建兴,李尔玉.基于复杂网络动力学模型的无向加权网络节点重要性评估[J].物理学报, 2018, 67(9): 255-271.
- [4] 饶元,吴连伟,张君毅.跨媒介舆情网络环境下信息传播机制研究与进展[J].中国科学:信息科学, 2017, 47(12): 1623-1645.
- [5] 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.
- [6] SHEIKHAHMADI A, NEMATBAKHSHE M A. Identification of multi-spreader users in social networks

- for viral marketing[J]. *Journal of Information Science*, 2017, 43(3): 412-413.
- [7] TIAN L, BASHAN A, SHI D N, et al. Articulation points in complex networks [J]. *Nature Communications*, 2017, 8: 14223(1-9).
- [8] KITSACK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. *Nature Physics*, 2010, 6(11): 888-893.
- [9] ZENG A, ZHANG C J. Ranking spreaders by decomposing complex networks[J]. *Physics Letters A*, 2013, 377(14): 1031-1035.
- [10] ZAREIE A, SHEIKHAHMADI A. A hierarchical approach for influential node ranking in complex social networks [J]. *Expert Systems with Applications*, 2018, 93: 200-211.
- [11] 罗仕龙, 龚凯, 唐朝生, 等. 加权网络中基于冗余边过滤的k-核分解排序算法[J]. *物理学报*, 2017, 66(18): 344-353.
- [12] 任卓明, 邵凤, 刘建国, 等. 基于度与集聚系数的网络节点重要度度量方法研究[J]. *物理学报*, 2013, 62(12): 128901(1-5).
- REN Z M, SHAO F, LIU J G, et al. Node importance measurement based on the degree and clustering coefficient information [J]. *Acta Physica Sinica*, 2013, 62(12): 128901(1-5).
- [13] FOWLER J H, CHRISTAKIS N A. Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham heart study [J]. *BMJ*, 2008, 337: a2338.
- [14] 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(04): 753-766.
- [15] 谢越. 一种基于K-shell和半局部信息的节点重要性排序方法[J]. *现代计算机*, 2018, (7): 51-54, 74.
- [16] BAE J, KIM S. Identifying and ranking influential spreaders in complex networks by neighborhood coeness[J]. *Physica A: Statistical Mechanics & Its Applications*, 2014, 395(4): 549-559.
- [17] KNUTH D E. *The Art of Computer Programming* [M]. Vol. 4, Addison-Wesley, 2008.
- [18] LUSSEAU D, SCHNEIDER K, BOISSEAU O J, et al. The bottleneck dolphin community of Doubtful Sound features a large proportion of long-lasting associations[J]. *Behavioral Ecology and Sociobiology*, 2003, 54(4): 396-405.
- [19] GLEISER P M, DANON L. Community structure in jazz[J]. *Advance in Complex Systems*, 2003, 6(4): 565-573.
- [20] ISELLA L, STEHLÉ J, BARRAT A, et al. What's in a crowd? Analysis of face-to-face behavioral networks [J]. *Journal of Theoretical Biology*, 2011, 271(1): 166-180.
- [21] GUIMERÀ R, DANON L, DIAZ-GUILERA A. Self-similar community structure in a network of human interactions[J]. *Physical Review E*, 2003, 68(6): 065103(1-4).
- [22] MCAULEY J, LESKOVEC J. Learning to discover social circles in ego networks[C]// *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe, USA: Curran Associates Inc., 2012: 539-547.
- [23] LESKOVEC J, KLEINBERG J, FALOUTSOS C. Graph evolution: Densification and shrinking diameters [J]. *ACM Transactions on Knowledge Discovery from Data*, 2007, 1(1): No. 2(1-41).
- [24] JOSHI-TOPE G, GILLESPIE M, VASTRIK I, et al. Reactome: A knowledgebase of biological pathways [J]. *Nucleic Acids Research*, 2005, 33(1): D428-D432.
- [25] LESKOVEC J, KLEINBERG J, FALOUTSOS C. Graphs over time: Densification laws, shrinking diameters and possible explanations[C]// *International Conference on Knowledge Discovery and Data Mining*. ACM, 2005, [2018-05-18] <http://pdfs.semanticscholar.org/6ea6/00cb73a2b75e9b5b0515f2a3dee3dd56cf1f.pdf>.
- [26] BOGUNA M, PASTOR-SATORRAS R, DIAZ-GUILERA A, et al. Models of social networks based on social distance attachment[J]. *Physical Review E*, 2004, 70(5): 056122.