

DOI: 10.11992/tis.201811012

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.tp.20190521.1449.003.html>

可拓聚类的科教人际网络节点重要性动态分析方法

严家萌^{1,2}, 许立波², 李兴森³, 庞超逸², 董瑞辰⁴

(1. 浙江大学 工程师学院, 浙江 杭州 310015; 2. 浙江大学 宁波理工学院 计算机与数据工程学院, 浙江 宁波 315100; 3. 广东工业大学 可拓学与创新方法研究所, 广东 广州 510006; 4. 北京邮电大学 国际学院, 北京 100876)

摘要: 目前大多数研究对复杂社会网络关键节点影响力的识别都是静态的, 缺乏动态变化的分析。采用可拓聚类方法对动态变化下的科教人际网络进行量化分析, 首先以多属性决策法计算每个节点重要性, 再利用变异系数权重法计算得该节点综合重要性量值, 之后划分等级并取标准正域和正域区间, 利用可拓关联函数计算每个节点与每个等级的关联度, 关联度值最大的等级即为该节点对应等级, 最后分析同一社会网络节点在不同时间点的重要性等级变化。可拓聚类方法尝试从动态上对网络节点重要性进行把握, 最后通过实例验证了该方法的有效性。

关键词: 复杂网络; 节点重要性; 多属性; 可拓学; 可拓聚类; 可拓理论; 物元; 关联函数

中图分类号: TP301.6 **文献标志码:** A **文章编号:** 1673-4785(2019)05-0915-07

中文引用格式: 严家萌, 许立波, 李兴森, 等. 可拓聚类的科教人际网络节点重要性动态分析方法 [J]. 智能系统学报, 2019, 14(5): 915-921.

英文引用格式: YAN Jiameng, XU Libo, LI Xingsen, et al. Dynamic analysis method of importance of science and education interpersonal network nodes based on extension clustering[J]. CAAI transactions on intelligent systems, 2019, 14(5): 915-921.

Dynamic analysis method of importance of science and education interpersonal network nodes based on extension clustering

YAN Jiameng^{1,2}, XU Libo², LI Xingsen³, PANG Chaoyi², DONG Ruichen⁴

(1. Polytechnic Institute, Zhejiang University, Hangzhou 310015, China; 2. School of Computer and Data Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo 315100, China; 3. Research Institute of Extension and Innovation Methods, Guangdong University of Technology, Guangzhou 510006, China; 4. Department of International School, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: At present, the identification of the influence of key nodes in complex social network is usually static and does not involve the analysis of dynamic changes. The extension clustering method is used to quantitatively analyze the interpersonal network of science and education under dynamic changes. First, the importance of each node is calculated by multi-attribute decision-making method. Then the comprehensive importance of the node is calculated by the coefficient of variation weight method. It is then classified, and the standard positive domain and the positive domain are acquired. The extension correlation function is used to calculate the degree of association between each node and each level. The level with the highest correlation value is the corresponding level of the node. Finally, the importance level of the same social network node at different time points is analyzed. The extension clustering method aims to dynamically determine the importance of network nodes. Finally, the effectiveness of the method is verified using an example.

Keywords: complex network; node importance; multi-attribute; extension; extension clustering; extension theory; matter element; correlation function

收稿日期: 2018-11-20. 网络出版日期: 2019-05-22.

基金项目: 国家自然科学基金项目 (61572022); 浙江省自然科学基金项目 (LY16G010010, LY18F020001).

通信作者: 许立波. E-mail: Xu_libo@163.com.

随着技术的进步, 交流的频繁, 人们的社会活动日益网络化, 各种各样的复杂网络随之充斥着我们的生活。复杂网络^[1]具有结构复杂、网络进

化、连接复杂、节点多样和动力学复杂性等特征,以及小世界、集群和幂律 (power law) 等现象。复杂网络拓扑图上各节点的复杂性,也决定了网络中每个节点的重要性程度存在差异,寻找出这些网络中的关键节点,有针对性地分析其性质并进行有效利用有着重要意义^[2]。例如:对社交网络中热点问题的挖掘,科研合作网中影响力因子认定,交通网络拥堵路段识别,电力网络用电高峰管理,信息传播机制的发现,舆论倾向性的研究,犯罪网络的侦察等都离不开节点重要性这个基本问题。发现重要的节点,有助于参透舆论热点和预测信息传播规律,以及把握扩散方向和对关键的认定,趋利避害为人们所利用。

在复杂网络中,对节点影响力重要性的评估得到了广泛而深入的研究,这些研究从不同角度对节点的重要性刻画给出了可行性分析与探索。目前,经典的关键节点中心性测量指标可被用来对节点的传播影响力进行评估,包括度中心性 (degree centrality)^[3]、介数中心性 (betweenness centrality)^[4]、紧密度指标 (closeness centrality)^[5]、K-Shell (KS 指标)^[6]等,总之,这些都是基于属性和网络位置的方法。文献[7]利用度中心性来测量最有影响力的节点,在符合幂律的非均匀网络中度数较大的节点的传播影响力相对较大。文献[8]利用介数中心性来评价社会网络影响力,以其经过某个节点的最短路径的数目来刻画节点的重要性。文献[9-10]利用紧密度指标刻画某个节点到其他节点的难易程度。文献[11-12]提出通过 K-Shell (KS) 分解来确定最具有影响力的单源节点。总体看来,这些评估方法虽然可行有效,但大都是静态地刻画和测量复杂网络节点重要性,而对其动态变化趋势和演化效果却极少提及。

可拓学^[13]是通过定性和定量相结合的手段去智能化处理矛盾问题的新兴学科,近几年来在模式识别^[14]、决策分析^[15-16]、问题求解等研究领域有着广泛的应用。其中可拓集理论是研究事物动态性和可变性的有力工具,据此本文提出一种基于可拓聚类的复杂人际网络节点重要性分析方法,尝试从动态上对网络节点影响力变化趋势做出判断。

1 节点重要性属性

设复杂网络 $G=(V,E)$ 是一个无向无权图,其中 $V=\{v_1, v_2, \dots, v_n\}$ 是网络中所有节点的集合, $|V|=N$; $E=\{e_1, e_2, \dots, e_m\}$ 是节点间边的集合,

$|E|=M$, $A=a_{ij}=1$, 其中 $A=a_{ij}=1$ 表示节点 i 和节点 j 相连, 否则 $a_{ij}=0$ 。在复杂网络中,由于对节点重要性的衡量标准和角度不同,节点中心性指标的定义也各有不同^[17],这里对所使用的几个节点重要性指标给出以下定义。

1.1 度中心性

在静态网络中,节点度是指与其直接连接的节点个数。度中心性用于描述节点在整个网络中所产生的影响力,用 $D(c)$ 表示为,即

$$D(c)=d(v) \quad (1)$$

式中: v 为节点; $d(v)$ 称为该节点的度。在复杂社会网络中,度中心性定义表明了一个节点对其相邻节点的影响能力,数值越大,在网络中越重要^[2]。

1.2 介数中心性

介数中心性刻画了网络中节点对于信息流动的影响力^[18],反映了相应的节点在整个网络中的作用,介数中心性的值越高,则该节点的影响力越大,相应地也就越重要。节点介数定义为网络中任意两个节点间最短路径总数与最短路径中经过该节点的路径的数目的比例的倒数,节点 v 的介数 $B(c)$ 定义为

$$B(c)=\sum \frac{S_{jk}(i)}{S_{jk}} \quad (2)$$

式中: S_{jk} 表示源节点 j 到目标节点 k 的最短路径数; $S_{jk}(i)$ 表示节点 j 和 k 之间经过节点 i 的最短路径数。复杂网络中介数中心性定义认为,如果一个节点是网络中其他节点对之间通信的必经之路,则其在网络中必具有重要地位^[4]。

1.3 紧密度

在一个社会网络中,紧密度^[9]用于刻画节点到达其他节点的难易程度,其值越大,表明到达其他节点越容易,节点越居于网络中心位置,相应地也就越重要,具体用 $C(c)$ 表示,即

$$C(c)=\frac{N-1}{\sum_{j=1}^n d_{ij}} \quad (3)$$

式中: d_{ij} 表示以节点 i 为起点,以 j 为终点的最短路径中所含边的数量。则节点 i 的紧密度中心性可以表示为网络中节点总数与该节点到其他所有节点距离之和的比值。

1.4 介数中心性

K-Shell 是图论里一个经典的概念,是一种粗粒化的节点重要性分类方法。操作上 K-Shell 相当于剥去最外面一层壳,首先去除网络中度数小于 k 的所有节点及其连边;如果在剩下的节点中

仍有度值小于 k 的节点,那么就继续去除这些节点,直至网络中剩下的节点的度值不小于 k 。依次取 $k=1, 2, \dots, n$, 就得到了该网络的 K-Shell 分解^[18], 依此类推,直到复杂社会网络中所有的节点都被赋予 K-Shell 值。

1.5 重要性属性权重

节点多属性不同指标从不同的角度探讨了节点在复杂网络中的重要程度,但由于实际的网络尤为复杂,仅依赖于单一指标来判断某个节点在网络中的影响力具有很大的片面性^[2]。而如果把某些单一指标简单叠加或者平均,并不能保证取得良好的效果,因为各个指标在度量节点重要性时权重有所不同,侧重点各异。这就需要进行多属性关联考虑,将多个重要性评价指标作为该节点的属性,通过计算各个属性的权重比例,最后得到该节点的综合重要性量值。

权系数反映了单个样本对整体结果的重要程度,在复杂社会网络中,动态分析节点重要性变化时,会出现影响力量值整体同方向变化情况,此时便会影响对结果的判断。因此,根据网络节点重要性量值的变化程度来设置各自的权值,比如权值、主观赋权值等更加合理。变异系数法是直接利用各属性所包含的信息通过计算得到的,为了消除各属性值的不同影响,需要用各项属性值的变异系数来衡量各项属性取值的差异程度。各项属性值的变异系数公式为

$$v_j = \frac{\sigma_j}{\bar{x}_j}, \quad j = 1, 2, \dots, m \quad (4)$$

式中: v_j 为第 j 项属性的变异系数; σ_j 为第 j 项属性的标准差; \bar{x}_j 为第 j 项属性的平均值。各项属性的权重为

$$W_j = \frac{v_j}{\sum_{j=1}^m v_j} \quad (5)$$

所有权值向量 W_j 满足:

$$W = (W_1, W_2, \dots, W_m), \quad \sum_{j=1}^m W_j = 1$$

根据多属性量值的计算,将每个节点在对应的属性上的取值规范化,即

$$p_{ij} = \frac{q_{ij}}{\max_i [q_{ij}]}, \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m \quad (6)$$

式中 q_{ij} 为第 i 个节点第 j 个属性重要性值。根据多属性计算的重要性量值及变异系数权重计算各节点综合重要性,即

$$k(i) = \sum_{j=1}^n w_j p_{ij} \quad (7)$$

2 算法流程

算法 人际网络可拓聚类动态算法

输入 网络图 $G = (V, E)$; 多属性 a_i ; 变异系数权重 W_j ; 综合关联度 $k(i)$; 等级划分及标准正域和正域区间;

输出 各节点重要性等级, 关联度变化。

1) 根据原始数据构建人际关系网络图, 每个个体人作为网络节点, 人与人熟识与否决定是否有关;

2) 根据式 (1)~(3) 包括度中心性、介数中心性、紧密度、K-Shell 值计算每个节点的各项重要性属性值;

3) 根据节点的各项重要性属性值, 按式 (4)、式 (5) 采用变异系数法计算各重要性属性权重;

4) 根据式 (7) 计算每个节点的每个重要性属性值与权重的乘积, 得到该节点的综合重要性度量值;

5) 划分等级 (L、M、H), 根据实际需要取每个等级的标准正域和正域区间;

6) 计算每个节点重要性值关于每个等级的关联度, 关联度值最大的等级为该节点的重要性等级;

7) 在第二时间点上, 对变化后的节点网络, 重复 3)~6)。

8) 比较两个时点每个节点重要性等级变化, 并给出变化论域。

3 可拓模型构建

3.1 建模机制

可拓聚类模型首先将根据多属性计算各节点重要性, 结合变异系数权重值求得综合重要性, 之后根据实际需要划分等级个数, 统计每个等级节点重要性的取值范围构成标准正域区间, 而在全体数据中的取值范围为正域区间。然后通过关联函数计算每个节点与每个等级的关联度, 取关联度值最大的等级为该节点等级。最后得到各节点不同时间点的关联度, 进而从动态的角度来分析节点重要性变化情况。

3.2 物元建模

以网络节点 O_i 为对象, c_i 为节点属性特征, O_i 关于 c_i 的量值 v_i 构成的有序三元组 $M_i = (O, c_i, v_i)$ 作为描述节点属性量值的基本元, 称为一维物元; 而网络节点 O 的 n 个特征 c_1, c_2, \dots, c_n 及 O 关于 $c_i (i = 1, 2, \dots, n)$ 对应的量值 $v_i (i = 1, 2, \dots, n)$ 所构

成的阵列^[19]

$$M_1 = \begin{bmatrix} O_i & c_1 & v_1 \\ & c_2 & v_2 \\ & \vdots & \vdots \\ & c_n & v_n \end{bmatrix} = (O_i, c_i, v_i)$$

称为 n 维物元。其中, $v_r (r=1, 2, \dots, n)$ 为节点 O 关于重要性特征 $c_i (i=1, 2, \dots, n)$ 的量值, 即各个节点关于各属性的量值。

3.3 标准正域和正域

根据需要划分节点重要性为高、中、低 3 个等级 (L、M、H), 分别统计出各等级取值范围作为标准正域区间, 全局取值范围作为正域区间。

$$M_2 = (O_d, c_j, v_j) = \begin{bmatrix} O_d & c_1 & v_1 \\ & c_2 & v_2 \\ & \vdots & \vdots \\ & c_n & v_n \end{bmatrix} = \begin{bmatrix} O_d & d_1 & \langle a_1, b_1 \rangle \\ & d_2 & \langle a_2, b_2 \rangle \\ & \vdots & \vdots \\ & d_n & \langle a_n, b_n \rangle \end{bmatrix}$$

式中: O_d 表示网络节点重要性; $d_r (r=1, 2, \dots, n)$ 表示对应的等级; $\langle a_{Oxr}, b_{Oxr} \rangle (x=1, 2; r=1, 2, 3)$ 表示节点 O 对于等级特征的量值范围即标准正域, 对应的 $\langle a_{xr}, b_{xr} \rangle (x=1, 2; r=1, 2, 3)$ 即为正域。

3.4 可拓距

对于节点重要性, 取实轴上的任一节点 x 与实域中的任一区间 $X = \langle a, b \rangle$ 的可拓距^[20]为

$$\rho_m(x, x_0, X) = \left| x - \frac{a+b}{2} \right| - \frac{b-a}{2} = \begin{cases} a-x, & x \leq \frac{a+b}{2} \\ x-b, & x \geq \frac{a+b}{2} \end{cases} \quad (8)$$

根据固定节点 x_0 在区间的位置的不同, 分为左侧可拓距、中点可拓距和右侧可拓距。给定区间 $X = \langle a, b \rangle$, 当 $x_0 \in (a, \frac{a+b}{2})$, 则 x 与区间 X_0 关于 x_0 左侧距为

$$\rho_l(x, x_0, X) = \begin{cases} a-x, & x \leq a \\ \frac{b-x_0}{a-x_0}(x-a), & x \in \langle a, x_0 \rangle \\ x-b, & x \geq x_0 \end{cases} \quad (9)$$

当 $x_0 \in (\frac{a+b}{2}, b)$, 则 x 与区间 X_0 关于 x_0 右侧距为

$$\rho_r(x, x_0, X) = \begin{cases} a-x, & x \leq x_0 \\ \frac{a-x_0}{b-x_0}(b-x), & x \in \langle x_0, b \rangle \\ x-b, & x \geq b \end{cases} \quad (10)$$

3.5 关联函数

反映各节点重要性量值与其标准正域的隶属程度的函数称为关联函数^[21]。设区间 $X_0 = \langle a, b \rangle$, $X = \langle c, d \rangle$ 且 $X_0 \subset X$, 则 x 关于 x_0 及区间 X_0 和 X 的关联函数^[20]为式 (11), 其中, $\rho(x, x_0, X)$ 为可拓距。

$$k(x) = \begin{cases} \frac{\rho(x, x_0, X)}{D(x, x_0, X_0, X)}, & D(x, x_0, X_0, X) \neq 0 \text{ 且 } x \in X \\ -\rho(x, x_0, X_0) + 1, & D(x, x_0, X_0, X) = 0 \text{ 且 } x \in X_0 \\ 0, & D(x, X_0, X) = 0, x \notin X_0 \text{ 且 } x \in X \\ \frac{\rho(x, x_0, X)}{D(x, x_0, X_0, \hat{X})}, & D(x, x_0, X, \hat{X}) \neq 0 \text{ 且 } x \in R-X \\ -\rho(x, x_0, \hat{X}) - 1, & D(x, x_0, X, \hat{X}) = 0 \text{ 且 } x \in R-X \end{cases} \quad (11)$$

3.6 节点重要性变化可拓学解释

可拓集理论可用来描述复杂人际网络节点重要性, 尝试刻画节点重要性的动态变化方向。设 $\tilde{E}(T) = \{(u, y, y') | u \in U, y = k(u) \in R; T_U u \in T_U U, y' = T_k k(T_U u \in R)\}$ 为论域 U 上的一个可拓集^[13], $y = k(u)$ 为 $\tilde{E}(T)$ 的关联函数, 当 $T = (T_k, T_U) = (e, e)$ 时, 根据关联函数值把论域分为正域、负域和零界, 即: $V_+ = \{u | u \in U, k(u) > 0\}$, $V_- = \{u | u \in U, k(u) < 0\}$, $V_0 = \{u | u \in U, k(u) = 0\}$ 。当 $T = (T_k, T_U) \neq (e, e)$ 时, 把论域 U 划分为关于变换 T 的正可拓、负可拓、正稳定和负稳定域, 分别为

$$\begin{aligned} V_+(T) &= \{u | u \in U, y = k(u) \leq 0; T_u(u) \in U, y' = T_k k(T_u u) > 0\} \\ V_-(T) &= \{u | u \in U, y = k(u) \geq 0; T_u(u) \in U, y' = T_k k(T_u u) < 0\} \\ V_+(T) &= \{u | u \in U, y = k(u) > 0; T_u(u) \in U, y' = T_k k(T_u u) > 0\} \\ V_-(T) &= \{u | u \in U, y = k(u) < 0; T_u(u) \in U, y' = T_k k(T_u u) < 0\} \end{aligned}$$

4 实验和分析

4.1 数据集

实验数据来自 Freeman 观测的由美国国家科学基金会赞助的计算机会议参加者构成的社会网络, 该网络呈现出复杂网络的基本特征。会议的参加者都是社会网络研究领域从事新兴学科的研究者, 在利用图这一数据结构抽象时, 通过熟识与否来定义网络的边取值, 选取其中 32 个人构成的网络数据作为实验数据, 并在两个时间点上进行了测量其网络变化, 两个时间点间距 9 个月。

4.2 节点重要性及权重系数计算

根据前文选取多属性测量节点重要性的分析, 分别计算度中心性 (式 (1))、介数中心性 (式 (2))、紧密度 (式 (3)) 以及 K-shell 分解的重要性量值, 并标准化 (式 (6)), 再根据式 (4)、式 (5) 计算各个属性变异系数权重, 最后根据式 (7) 取其值为该节点的综合重要性量值, 如表 1 所示。表 2 为第二个时间节点的各属性重要性量值及综合重要性量值。

表1 第一时间节点重要性及变异系数

Table 1 First time node importance and variation coefficient

n_i	a_1	a_2	a_3	a_4	$k(i)$
n_1	1.00	0.45	1.00	0.74	0.66
n_2	0.87	0.26	0.89	0.68	0.51
n_3	0.90	0.26	0.91	0.74	0.53
n_4	0.81	0.15	0.84	0.74	0.44
n_5	0.84	0.17	0.86	0.74	0.46
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n_{28}	0.81	0.21	0.84	0.71	0.47
n_{29}	0.87	0.25	0.89	0.74	0.51
n_{30}	0.65	0.15	0.74	0.58	0.38
n_{31}	0.87	0.33	0.89	0.68	0.55
n_{32}	0.52	0.03	0.67	0.52	0.27
W_j	0.20	0.55	0.11	0.14	1.00

表2 第二时间节点重要性及变异系数

Table 2 Second time node importance and variation coefficient

n_i	a_1	a_2	a_3	a_4	$k(i)$
n_1	1.00	0.27	1.00	0.87	0.56
n_2	0.97	0.24	0.97	0.87	0.54
n_3	0.90	0.16	0.91	0.87	0.47
n_4	0.97	0.20	0.97	0.87	0.52
n_5	0.84	0.07	0.86	0.87	0.41
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n_{28}	0.94	0.18	0.94	0.87	0.50
n_{29}	0.97	0.20	0.97	0.87	0.52
n_{30}	0.84	0.12	0.86	0.77	0.42
n_{31}	1.00	0.17	1.00	0.87	0.51
n_{32}	0.87	0.14	0.89	0.81	0.45
W_j	0.17	0.57	0.11	0.15	1.00

4.3 等级划分及标准正域与正域

为探究节点重要性变化情况,在分析节点重要性时分为3个等级(高(High)、中(Middle)、低(Low))。根据多属性和变异系数权重计算每个节点综合重要性量值,取其最低20%为低影响力节点,最高20%为高影响力节点,中间60%为中间节点。在每个等级量值中取标准正域区间,在全局量值上取正域区间,结果见表3、4。

表3 第一时间节点影响力等级划分及(标准)正域

Table 3 First time node importance level and (standard) positive domain

正域	标准正域		
—	L	M	H
<0.00,1.00>	<0.16,0.32>	<0.34,0.54>	<0.55,0.66>

表4 第二时间节点影响力等级划分及(标准)正域

Table 4 Second time node importance level and (standard) positive domain

正域	标准正域		
—	L	M	H
<0.00,1.00>	<0.23,0.39>	<0.41,0.51>	<0.52,0.56>

4.4 关联度

对3个等级中的每个等级运用关联函数公式计算每个节点与每个等级的关联度(见式(8)~式(11))。这里以节点2为例说明计算过程,因 $k_2(p) = 0.51$,有

$$k_{2L}(x) = \frac{\rho(x, x_0, X)}{D(x, x_0, X_0, X)} = \frac{0.51 - x}{\rho(x, x_0, X) - \rho(x, x_0, X_0)} = \frac{0.51 - 0.32}{(0.00 - 0.51) - (0.51 - 0.32)} = -0.27$$

$$k_{2M}(x) = \frac{\rho(x, x_0, X)}{D(x, x_0, X_0, X)} = \frac{0.34 - x}{\rho(x, x_0, X) - \rho(x, x_0, X_0) + a_{012} - b_{012}} = \frac{0.34 - 0.51}{(0.00 - 0.51) - (0.34 - 0.51) + 0.34 - 0.54} = 0.31$$

$$k_{2H}(x) = \frac{0.55 - 0.51}{(0.00 - 0.51) - (0.55 - 0.51)} = -0.07$$

最后列表,如表5、表6所示。

表5 第一时间点各等级上的关联度

Table 5 Relevance at each level of the first time

n_i	L	M	H
n_1	-0.34	-0.15	0.17
n_2	-0.27	0.31	-0.07
n_3	-0.28	0.35	-0.04
n_4	-0.21	0.19	-0.20
n_5	-0.23	0.22	-0.16
\vdots	\vdots	\vdots	\vdots
n_{28}	-0.24	0.24	-0.15
n_{29}	-0.27	0.31	-0.07
n_{30}	-0.14	0.07	-0.31
n_{31}	-0.29	-0.02	0.00
n_{32}	0.34	-0.21	-0.51

表6 第二时间点各等级上的关联度
Table 6 Relevance at each level of the second time

n_i	L	M	H
n_1	-0.12	-0.08	0.07
n_2	-0.11	-0.05	0.04
n_3	-0.07	0.12	-0.10
n_4	-0.10	-0.02	0.00
n_5	-0.02	0.00	-0.21
\vdots	\vdots	\vdots	\vdots
n_{28}	-0.18	0.18	-0.04
n_{29}	-0.20	-0.02	-0.00
n_{30}	-0.07	0.02	-0.19
n_{31}	-0.19	0.20	-0.02
n_{32}	-0.12	0.08	-0.13

4.5 节点重要性动态变动结果

根据以上计算结果,每个节点最大关联度值对应等级即为该节点所属等级,列表见表7(T_1 为第一时间点, T_2 为第二时间点)。从表7中能直观地看出每个节点的等级变化情况,这里根据可拓集理论,给出了每个节点的可拓论域变化类型。表7中,“稳定”表示节点影响力重要性等级不变,“可拓”表示其等级有所改变,“负”表示节点重要性下降,“正”表示上升。例如,对于 n_2 节点,在第一时间点高等级关联度只有-0.07,而到第二时间点为0.04,所以其由中间等级M(middle)变为高等级H(high),即为可拓域,关联度由负变为正,所以称之为正可拓域,亦即说明在人际网络中经过一段时间后其节点重要性提高了,影响力变大。

表7 节点重要性变化情况
Table 7 Change in node importance

n_i	T_1		T_2		$V_i(T)$
	关联度	等级	关联度	等级	
n_1	0.17	H	0.07	H	负稳定域
n_2	-0.07	M	0.04	H	正可拓域
n_3	0.35	M	0.12	M	负稳定域
n_4	-0.20	M	0.00	H	正可拓域
n_5	0.22	M	0.00	M	负稳定域
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n_{28}	0.24	M	0.18	M	负稳定域
n_{29}	-0.07	M	0.00	H	正可拓域
n_{30}	0.07	M	0.02	M	负稳定域
n_{31}	0.00	H	-0.02	M	负可拓域
n_{32}	-0.21	L	0.08	M	正可拓域

5 结束语

人际网络可以采用图这一数据结构抽象,然后计算每个节点在多个重要性属性上的值,根据变异系数权重法取得该节点的综合重要性,此方法克服了单一属性的片面性。而可拓聚类方法,首先通过聚类分析划分节点集合若干等级,构造标准正域和正域区间,根据关联函数计算每个节点与每个等级集合的关联度,关联度值最大的等级即为该节点所属重要性等级,最后在两个时间点上比较等级变化情况,给出可拓论域解释。利用可拓聚类方法尝试对节点重要性变化的动态把握,使网络节点及各指标均能以形式化的模型更直观地展现。接下来工作将会在更大数据集和更复杂类型网络进一步验证可拓聚类方法的有效性,以及对其方法本身进行改进等。

参考文献:

- [1] ADAMIC L A, HUBERMAN B A, BARABÁSI A L, et al. Power-law distribution of the world wide web[J]. *Science*, 2000, 287(5461): 2115.
- [2] 于会, 刘尊, 李勇军. 基于多属性决策的复杂网络节点重要性综合评价方法[J]. *物理学报*, 2013, 62(2): 020204. YU Hui, LIU Zun, LI Yongjun. Key nodes in complex networks identified by multi-attribute decision-making method[J]. *Acta physica sinica*, 2013, 62(2): 020204.
- [3] ALBERT R, JEONG H, BARABÁSI A L. Error and attack tolerance of complex networks[J]. *Nature*, 2000, 406(6794): 378–382.
- [4] FREEMAN L C. A set of measures of centrality based on betweenness[J]. *Sociometry*, 1977, 40(1): 35–41.
- [5] WEHMUTH K, ZIVIANI A. DACCER: distributed assessment of the closeness centrality ranking in complex networks[J]. *Computer networks*, 2013, 57(13): 2536–2548.
- [6] KITSACK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. *Nature physics*, 2010, 6(11): 888–893.
- [7] PASTOR-SATORRAS R, VESPIGNANI A. Epidemic spreading in scale-free networks[J]. *Physical review letters*, 2001, 86(14): 3200–3203.
- [8] FREEMAN L C. Centrality in social networks conceptual clarification[J]. *Social networks*, 1979, 1(3): 215–239.
- [9] OPSAHL T, AGNEESSENS F, SKVORETZ J. Node centrality in weighted networks: generalizing degree and shortest paths[J]. *Social networks*, 2010, 32(3): 245–251.
- [10] SABIDUSSI G. The centrality index of a graph[J]. *Psychometrika*, 1966, 31(4): 581–603.
- [11] CLAUSET A, SHALIZI C R, NEWMAN M E J. Power-

- law distributions in empirical data[J]. *SIAM review*, 2009, 51(4): 661–703.
- [12] CARMÍ S, HAVLIN S, KIRKPATRICK S, et al. A model of Internet topology using k-shell decomposition[J]. *Proceedings of the national academy of sciences of the United States of America*, 2007, 104(27): 11150–11154.
- [13] 杨春燕, 蔡文. 可拓学 [M]. 北京: 科学出版社, 2014.
- [14] 杨国为, 王守觉. 模式可拓识别及其神经网络模型 [J]. *哈尔滨工业大学学报*, 2006, 38(7): 1129–1132.
- YANG Guowei, WANG Shoujue. Pattern extension recognition and its neural network model[J]. *Journal of Harbin Institute of Technology*, 2006, 38(7): 1129–1132.
- [15] XU Libo, LI Xingsen, PANG Chaoyi. Uncertain multiattribute decision-making based on interval number with extension-dependent degree and regret aversion[J]. *Mathematical problems in engineering*, 2018, 2018: 6508636.
- [16] XU Libo, LI Xingsen, SHAO Junkai, et al. Extension dependent degree method with mapping transformation for three-parameter interval number decision making[J]. *Mathematical problems in engineering*, 2018, 2018: 1831086.
- [17] 王晓磊, 杨岳湘, 何杰. 基于多重属性的 P2P 网络节点重要性度量方法 [J]. *计算机应用*, 2014, 34(S2): 7–10, 19.
- WANG Xiaolei, YANG Yuexiang, HE Jie. P2P network node importance measurement method based on multi-attribute[J]. *Journal of computer applications*, 2014, 34(S2): 7–10, 19.
- [18] 胡庆成, 尹龔燊, 马鹏斐, 等. 一种新的网络传播中最有影响力的节点发现方法 [J]. *物理学报*, 2013, 62(14): 140101.
- HU Qingcheng, YIN Yanshen, MA Pengfei, et al. A new approach to identify influential spreaders in complex networks[J]. *Acta physica sinica*, 2013, 62(14): 140101.
- [19] 蔡文, 杨春燕, 陈文伟, 等. 可拓集与可拓数据挖掘 [M]. 北京: 科学出版社, 2008: 6.
- [20] 杨春燕. 可拓创新方法 [M]. 北京: 科学出版社, 2017: 3.
- [21] 蔡文, 杨春燕. 可拓学的基础理论与方法体系 [J]. *科学通报*, 2013, 58(13): 1190–1199.
- CAI Wen, YANG Chunyan. Basic theory and methodology on Extenics[J]. *Chinese science bulletin*, 2013, 58(13): 1190–1199.

作者简介:



严家萌, 男, 1987 年生, 硕士研究生, 主要研究方向为数据挖掘、深度学习。



许立波, 男, 1976 年生, 讲师, 博士, 主要研究方向为人工智能、智能信息处理。



李兴森, 男, 1968 年生, 教授, 博士, 中国人工智能学会理事, 中国人工智能学会可拓工程专业委员会副主任、秘书长, 浙江省创造学研究会常务理事, 国际 MCDM、IEEE 学会会员, 主要研究方向为可拓学、智能知识管理与数据挖掘。承担及参加省部级以上项目 8 项, 其中国家级项目 6 项。发表学术论文 60 余篇, 出版专著 2 部。