

文章编号: 2095-2163(2022)03-0189-11

中图分类号: TP301.6

文献标志码: A

有偏随机游走改进吸收中心性识别关键节点

宁阳¹, 宁晴², 王志峰³

(1 天津电子信息职业技术学院 计算机与软件技术系, 天津 300350; 2 北京联合大学 信息服务重点实验室, 北京 100101; 3 天津职业技术师范大学 信息技术工程学院, 天津 300222)

摘要: 针对无向网络吸收中心性算法中未考虑信息传递的有偏问题, 提出一种基于节点的度信息有偏随机游走, 改进吸收中心性识别关键节点的算法。首先在无向网络中利用节点的度信息, 构建信息传递的转移概率矩阵并定义吸收节点, 通过到达吸收节点的平均首次可达时间, 衡量节点的重要性; 将该算法向有向网络进行扩展研究, 结合有向网络节点的出度、入度信息, 重新构建信息转移概率矩阵, 识别关键节点。基于 SIR 传播模型和 Kendall tau 相似性, 在真实的无向、有向网络中设计了3组对比实验, 与度中心性、介数中心性、接近中心性、改进前吸收中心性算法、PageRank、LeaderRank、Pro_PageRank、DPRank等算法对比, 无向网络中的改进算法与其它中心性算法识别结果平均相似性达 0.869, 有向网络中基于节点度、入度、出度信息的识别结果, 平均相似性达 0.893、0.867、0.854。实验证明, 在识别关键点中改进后的算法能够起到有效的作用。

关键词: 复杂网络; 关键节点; 无向网络; 有向网络; 吸收中心性; 有偏随机游走

Improved absorption centrality based on biased random walk to identify key nodes

NING Yang¹, NING Qing², WU Zhifeng³

(1 Department of computer and software technology, Tianjin Electronic Information College, Tianjin 300350, China;

2 Key Laboratory of Information Service Engineering, Beijing Union University, Beijing, 100101, China;

3 College of Information Technology Engineering, Tianjin University of Technology and Education, Tianjin 300222, China)

[Abstract] Aiming at the problem that the information transmission bias is not considered in the Absorption Centrality algorithm of undirected networks, an improved algorithm for identifying key nodes based on the biased random walk of degree information is proposed. Firstly, the degree information of nodes is used to construct the transition probability matrix of information transmission in undirected network, and the absorption node is defined. The importance of the node is measured by the average first arrival time of the absorption node. The algorithm is extended to the directed network, and the information transfer probability matrix is reconstructed to identify the key nodes. Based on SIR propagation model and Kendall tau similarity, three groups of comparative experiments were designed in a real undirected and directed network, which were compared with Degree Centrality, Betweenness Centrality, Closeness Centrality, Absorption Centrality before improvement, PageRank, LeadRank, Pro_PageRank and DPRank algorithm, the average similarity between the improved algorithm in undirected network and other centrality algorithms is 0.869, and that based on node degree, in degree and out degree information in directed network is 0.893, 0.867 and 0.854 respectively. Experimental results show that the improved algorithm can effectively identify the key nodes.

[Key words] complex network; key node; undirected network; directed network; absorption centrality; biased random walk

0 引言

在电力交通、社交、生物、政治经济等领域, 复杂网络中的关键节点识别具有实际应用意义。如: 复杂网络中的关键节点识别, 可以在电力交通网络^[1]中挖掘枢纽节点, 进行重点保护, 达到降低经济风险的目的; 在社交网络中最有影响力的人, 能够被准确的挖掘出来^[2], 使广告效益得到最大化的同时, 可以有助于流言传播的控制; 在蛋白质网络中可以对

关键蛋白质进行挖掘, 这些蛋白质在机体再生和生存时是不能或缺的^[3]; 在生物网络中, 为医药发展提供有价值的理论和方法。

复杂网络重要节点识别这一问题, 从网络局部、全局信息方面考虑, 可以将其划分为基于邻居节点的结构化指标、基于路径的结构化指标、基于节点移除及收缩的中心性指标、基于迭代寻优的中心性指标进行节点重要性研究^[4]。在无向网络, 度中心性方法^[5]、介数中心性方法^[6]、接近中心性方法^[7]等,

基金项目: 天津市自然科学基金青年科学基金项目(18JCQNJC04700); 国家自然科学基金青年项目(61601331)。

作者简介: 宁阳(1995-), 女, 硕士, 助教, 主要研究方向: 复杂网络、数据挖掘; 宁晴(1990-), 女, 硕士研究生, 主要研究方向: 图像识别、目标检测; 王志峰(1974-), 男, 博士, 教授, 硕士生导师, 主要研究方向: 演化计算、数据挖掘、机器学习。

收稿日期: 2021-12-03

都是比较经典的关键点识别方法。度中心性方法存在节点的度值与重要性不一定成正比问题,而介数中心性方法和接近中心性方法具有较高的时间复杂度,研究人员通常与典型算法进行比较,改进算法的有效性^[8]。在有向网络,使用迭代寻优的中心性指标 PageRank 算法^[9],以及在此基础上改进的关键点识别方法(如:引入背景节点的 LeaderRank 算法^[10]、倾向于入度(出度)节点大的邻居节点转移的 Pro_PageRank^[11]、DPRank 算法^[12]等等),均存在参数问题。加权网络中,利用节点的边权信息,重新定义边权的节点移除及收缩方法^[13]进行关键节点识别,其算法时间复杂度随着网络规模的增加而增加。融合其他理论,基于证据理论^[14]和 TOPSIS^[15]、灰色关联度分析、层次分析法^[16]、相对熵^[17]等方法进行多属性决策,对各指标的权重重新分配,来综合判别关键节点;也可与社区划分、图着色理论^[18]、投票算法、聚类算法^[19]等相结合,用于识别一组关键节点。此外,文献[20]中提出了一种将游走相似和表征节点重要性进行叠加的方法,对节点度、邻接节点的属性以及节点在网络中的位置加以融合,基于的随机游走是等概率的,但未对实际网络中游走的倾向性进行考虑;文献[8]利用理想状态下等概率的随机游走计算首次到达时间,提出吸收中心性算法,其中未将随机游走向节点转移的概率纳入研究机制。

综上所述,本文在文献[8]的基础上进行扩展研究,结合邻居节点的度(出度、入度)信息,在无向网络和向有向网络中考虑信息传播的倾向性,利用随机游走首次可达时间构建转移概率矩阵,用以识别网络中的关键节点。

1 相关概念

1.1 网络表示

本文将一个具体网络抽象为一个图,该图由一个点集 V 和边集 E 组成, $G = (V, E)$ 。顶点数记为 $N = |V|$, 边数记为 $M = |E|$ 。本文将原始数据用邻接矩阵加以表示, $A = (a_{ij})_{N \times N}$ 为图 G 的邻接矩阵, 是一个 N 阶方阵。其中, a_{ij} 为第 i 行第 j 列上的元素, 定义如下:

$$a_{ij} = \begin{cases} 1, & \text{节点 } i \text{ 和节点 } j \text{ 有边} \\ 0, & \text{节点 } i \text{ 和节点 } j \text{ 无边} \end{cases}$$

无向网络中,刻画节点属性时使用 v_i 的度 k_i , 定义为与节点直接相连边的数目,即:

$$k_i = \sum_{j=1}^N a_{ij}$$

有向网络中,节点 i 的度包括入度和出度。 v_i 的入度 k_i^{in} (出度 k_i^{out}) 表示其它节点指向 v_i (v_i 指向其它节点) 边的数量,即 $k_i^{in} = \sum_{j=1}^N a_{ji}$, $k_i^{out} = \sum_{j=1}^N a_{ij}$ 。

1.2 随机游走

随机游走是一个很有用的工具,借助这个工具可以对复杂网络的结构进行研究;有时也将随机游走模型称为布朗运动,其是一种理想的数学状态。随机游走可以理解为无法基于过去的行为对后续的发展步骤和方向进行预测。假设当前节点有 4 个邻居节点,该节点向其周围邻居节点的转移概率均为 $1/4$,这是一种无限制的等概率随机游走。若将节点之间转移的倾向性考虑进来,就成为了一种无限制的有偏随机游走。

1.3 关键节点识别算法

根据网络是否具有方向性,将关键节点识别算法应用的网络类型划分为无向与有向两种。在无向网络中,衡量节点的重要性,主要基于邻居节点的信息和路径;有向网络中主要通过迭代寻优的方法,改进 PageRank 算法,利用节点的点权信息扩展到加权网络,以及将多属性进行融合,结合局部信息和全局信息进行关键节点识别。

1.3.1 无向网络关键节点识别

度中心性 (Degree Centrality) 考虑节点的局部信息,认为节点的重要性取决于中心性,其高低取决于节点度的值。在无向图中,公式(1)为节点 i 的度中心性:

$$DC(i) = \frac{k_i}{N-1} \quad (1)$$

式中, N 为节点总个数, $DC(i)$ 为 v_i 的度中心性值。

介数中心性 (Betweenness Centrality) 方法,对节点的全局信息进行考虑。认为信息传播时,经过的是最短路径。刻画节点的重要性时,采用每个节点的最短路径数目。但该节点不一定是网络中度最大或者是网络的拓扑中心。公式(2)为 v_i 的介数中心性:

$$BC(i) = \frac{2}{(N-1)(N-2)} \sum_{j < k} \frac{n_{jk}(i)}{n_{jk}} \quad (2)$$

式中, $n_{jk}(i)$ 为经过 v_i 的 v_j 和 v_k 之间的最短路径数; n_{jk} 为 v_j 和 v_k 之间的最短路径总数; $BC(i)$ 为 v_i 的介数中心性值。

接近中心性 (Closeness Centrality) 是衡量节点重要性指标时的最短路径。借助网络对其它节点施

加影响的能力来反映节点,反映网络的全局结构,该方法只能应用于连通网络中。 v_i 的接近中心性如公式(3):

$$CC(i) = \frac{N - 1}{\sum_{j=1}^N d_{ij}} \quad (3)$$

式中: $CC(i)$ 为节点 i 的接近中心性值。

1.3.2 有向网络关键节点识别

PageRank 是一种可以按照重要性对网页排序的方法。一个页面是否质量高,在于其输入是否有许多高质量的页面。也就是说如果有很多概率比较大的节点指向同一个节点,则通过某个节点到达该节点的概率也会高。为了使悬挂节点的问题得以解决,PageRank 算法中每一个节点均以一定的几率跳转到网络中的任一节点,经过不断迭代,使每个节点的 PageRank 值趋于稳定,如公式(4):

$$P_i(t_n) = \frac{1 - c}{N} + c \cdot \sum_{j=1}^N \frac{a_{ji}}{k_j^{out}} p_j(t_{n-1}) \quad (4)$$

式中: P_i 为节点 i 的 PageRank 值, $1 - c$ 是跳转概率,也称阻尼系数,通常 $c = 0.85$ 。

LeaderRank 在解决出度为 0 的悬挂节点问题时,添加了一个虚拟背景节点 g , 使网络中所有节点与其两两双向连接,当迭代达到稳定,将虚拟背景节点的权值平均分配给每一个节点。该指标值与节点重要性成正比。如公式(5):

$$LeaderRank_i(t_n) = \sum_{j=1}^{N+1} \frac{a_{ji}}{k_j^{out}} LeaderRank_j(t_{n-1})$$

$$LR_i = LeaderRank_i(t_c) + \frac{LeaderRank_g(t_c)}{N} \quad (5)$$

其中, t_c 表示稳态时刻。

Pro-PageRank 方法是在节点转移过程中,更倾向沿着权重大的边游走。边的权重通过出度节点的入度邻居信息衡量。即节点转移时更倾向于沿着入度大的节点进行,如公式(6):

$$p(j, i) = \frac{k_i^{in}}{\sum_{l \in \Gamma^{out}(j)} k_l^{in}} \cdot a_{ji} \quad (6)$$

$$Pro_PageRank_i(t_n) = \frac{1 - c}{N} +$$

$$c \left[\sum_{j \in \Gamma^{in}(i)} p(j, i) \cdot Pro_PageRank_j(t_{n-1}) \right]$$

DPRank 方法考虑两步转移节点信息,节点更

倾向于出度大的节点转移。DP 方法构建的转移概率矩阵如公式(7):

$$p(i, j) = \frac{k_j}{\sum_{l \in \Gamma(i)} k_l} \cdot a_{ij}$$

$$p(i, j) = \frac{k_j^{out}}{\sum_{l \in \Gamma^{out}(i)} k_l^{out}} \cdot a_{ij} \quad (7)$$

$$P_{ij} = \begin{cases} \frac{1 - c}{N} + c \cdot p(i, j), & \sum_{l \in \Gamma^{out}(i)} k_l^{out} \neq 0 \\ \frac{1}{N}, & \text{otherwise} \end{cases}$$

1.4 评价指标

1.4.1 Kendall tau 距离

Kendall tau 距离^[21]用来计算两个排序列表之间成对分歧数量,即两个完整列表 σ 和 τ , $K(\sigma, \tau)$ 表示两个列表之间的差异性。

$$K(\sigma, \tau) = |\{(i, j) : i < j, \sigma(i) < \sigma(j) \text{ but } \tau(i) > \tau(j)\}| \quad (8)$$

$K \in [0, 1]$, K 值越大,则相似性越小。Kendall 距离归一化处理得到:

$$K' = 1 - \frac{2K}{N(N - 1)}$$

将其用于比较一个序列与另一个类似标准答案的排序序列的相似性,得出排序序列有效性, K' 值与两个列表之间相似性成正比。 σ 列表与 τ 列表计算 Kendall tau 距离相似值的二元组集合,表 1 中分别列出列表的所有二元组组合。当 $\sigma(i) < \sigma(j)$ 且 $\tau(i) > \tau(j)$ 或者 $\sigma(i) > \sigma(j)$ 且 $\tau(i) < \tau(j)$ 时 $K = 1$, 否则 $K = 0$ 。

例: $\sigma = \{1, 2, 3, 4\}$, $\tau = \{1, 3, 2, 4\}$, 其中 σ 表示标准序列。

表 1 σ 列表与 τ 列表二元组集合

Tab. 1 Two tuple sets of σ list and τ list

变量	二元组集合					
σ	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)	(3,4)
τ	(1,3)	(1,2)	(1,4)	(3,2)	(3,4)	(2,4)
K	0	0	0	1	0	0

由上表可知 K 值,进而得到两个列表之间的相似性: $K' = 1 - 2 * 1/4 * 3 = 0.833$

1.4.2 SIR 传播模型

本文使用 SIR 传播模型^[21],模拟信息在网络中的传播过程,用于表示节点的传播能力。在典型的传染病模型中,可将 N 个节点的状态分为如下 3 类:

(1) S (易染状态):在初始条件下将所有节点均定义为该状态,该节点被邻居节点感染的概率为 β ;

(2) I (感染状态):说明该节点已经感染某种病毒,已成为传染源,感染其直接易感邻居节点的概率为 β ;

(3) R (移除状态):感染状态节点以 β 概率将邻居易感节点转变为感染状态后,以 γ 概率变为 R 移除状态,不再具有感染能力和易感特性。

本文采用 SIR 为多源感染模型,初始时刻假设网络中 I 的节点为 K 个,其余均为 S 个体。一个单位时间内,所有处于 I 节点以 β 概率将其周围邻居节点感染之后,以 γ 值为 1 的概率转变为 R ,在不存在易感节点时达到稳定状态。此时对处于 R 的节点个数进行统计,用于衡量初始感染状态节点的传播能力,记为 Ft 。为减少 β, γ 参数带来的随机性,每个节点计算 100 次,利用平均值进行计算。对网络中的所有节点均执行以上操作,节点传播能力的大小即为网络中节点的重要度。

2 有偏随机游走改进吸收中心性算法

基于网络中的随机游走原理,利用到达吸收节点时间提出的吸收中心性方法,由于其在研究随机游走机制时,仅考虑理想状态下的等概率随机游走,并没有考虑网络中信息传播的复杂性问题。针对于此,本文将信息流在随机游走时向节点转移的概率纳入研究。在无向网络中,节点信息倾向于度大的节点转移;在有向网络中,节点的转移分别倾向于沿着出度、入度大的节点,将节点的重要性进行分析对比。

2.1 算法原理

吸收中心性方法(AS)是一种随机游走机制,这种机制是基于网络中等概率的,是利用网络中节点的吸收过程提出的一种识别重要节点的方法。吸收节点是在网络中任意选择一个节点 i ,对网络中其他节点等概率的随机游走到节点 i 的首次到达时间进行计算。该方法采用信息流的效率评估标准是首次到达时间,该节点的重要性衡量指标,是网络中其它节点到吸收节点的平均首次到达时间。平均吸收时间越短,其它节点到该节点的距离就越小,节点越重要。算法公式(9):

$$\begin{aligned} P_{ij} &= A_{ij} \cdot \frac{1}{k_{v_i}} \\ P_i^r &= (I - P) \cdot \text{drop}(i, i) \\ T &= (P_i^r)^{-1} \\ MFT_i &= \frac{1}{N-1} \sum_{j=1}^N T_{ij} \\ IMP(i) &= \frac{\max(MFT)}{MFT_i} \end{aligned} \quad (9)$$

式中: P_{ij} 是根据节点度信息构建的转移概率矩阵; k_{v_i} 为 v_i 的度; I 是单位矩阵; P_i^r 为 $I - P$ 矩阵的子矩阵,即移除第 i 行第 i 列生成的矩阵。对矩阵求逆,得到各节点到 v_i 的首次可达时间, T_{ij} 表示节点 j 随机游走到 v_i 的首次可达时间,进而对首次到达时间计算平均值 MFT 。使用最大值对节点的重要性标准化衡量。

本文提出对算法的改进工作,是在无向网络中根据邻居节点的度信息重新设置边权。邻居节点的度越大,信息越倾向于向该方向进行转移,为此提出了 GAS_D 算法;在有向网络中,基于邻居节点的出度和入度信息,分别提出了 GAS_D_{out} 、 GAS_D_{in} 算法。

无向网络重新构建转移概率矩阵,采用文献[8]提出的设置吸收节点的方法,衡量节点重要性,定义为 GAS_D :

$$P_{ij} = A_{ij} \cdot W_{ij} = A_{ij} \cdot \frac{k_j}{\sum_{j \in \Gamma(i)} k_j} \quad (10)$$

式中: $\Gamma(i)$ 为 i 的邻居节点集合; P_{ij} 为转移概率矩阵的值; W_{ij} 为节点 i, j 之间边的权值。

在有向网络中,本文定义的 GAS_D_{out} 中心性指标中节点间的转移概率矩阵如式(11):

$$P_{ij} = A_{ij} \cdot \bar{W}_{ij} = A_{ij} \cdot \frac{k_j^{out} + 1}{\sum_{j \in \Gamma^{out}(i)} (k_j^{out} + 1)} \quad (11)$$

考虑网络节点的方向性,该指标中存在设置的吸收节点。由于吸收节点和出度邻居节点间存在直接连边,为避免转移概率为 0 的情况,采用每个节点的出度均加 1,同时避免了分母为 0 的问题。

在有向网络中,公式(12)为本文定义的 GAS_D_{in} 中心性指标中节点间的转移概率矩阵。在吸收节点的邻居节点均为边界节点时, v_i 向每个邻居节点随机游走的概率均为 $1/k_{out}$,该指标等同于等概率的随机游走。

$$P_{ij} = A_{ij} \cdot \bar{W}_{ij} = A_{ij} \cdot \frac{k_j^{in}}{\sum_{j \in \Gamma^{out}(i)} k_j^{in}} \quad (12)$$

为了方便计算分析,按照最大值,将各节点的中心性值归一化处理,如式(13):

$$IMP'(i) = IMP(i) / \max(IMP) \quad (13)$$

2.2 案例分析

本节通过计算实例,对算法的计算过程进行解释,如图 1 所示。其中包含 9 个节点 12 条边的无向网络,假设吸收节点为节点 1,将其代入式(9)、

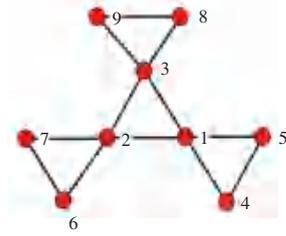


图 1 案例网络

Fig. 1 Case network

式(10),计算所有节点有偏随机游走到吸收节点 1 的首达时间。各中心性方法的中心性值见表 2。

表 2 各中心性方法所有节点的中心性值

Tab. 2 Centrality values of all nodes in each centrality method

节点	DC	BC	CC	AS	GAS_D	归一值
1	4	0.083 3	24	2.7	4.9	1
2	4	0.083 3	24	2.7	4.9	1
3	4	0.083 3	24	2.7	4.9	1
4	2	0.055 6	0	1	1	0.203
5	2	0.055 6	0	1	1	0.203
6	2	0.055 6	0	1	1	0.203
7	2	0.055 6	0	1	1	0.203
8	2	0.055 6	0	1	1	0.203
9	2	0.055 6	0	1	1	0.203

3 实验分析

根据上述分析,为验证本文提出基于有偏随机游走改进的吸收中心性算法(GAS_D、GAS_Dout、GAS_Din),识别关键点的有效性,对无向网络(karate、USAir、911terriset、Email)、有向网络(Freemans、Organisational、Celegansneural)进行仿真实验,实验数据集的网络参数见表 3。

表 3 中, n 表示节点数; m 表示边数; $\langle k \rangle$ 表示节点平均度; k_{max} 表示节点最大度; $\langle d \rangle$ 表示网络节点间最短路径平均数; C 表示聚类系数,用来评估节点聚集成团的集聚程度; R 表示同配系数,用来反映邻接节点间的度相关性; H 表示异质系数,衡量网络中节点度分布的异质性。

实验环境为: Intel (R) Core (TM) i3 - 2367M

CPU @ 1.40 GHz、4 GB 内存、Windows7、python3.2.7 32 位,使用 Python 语言实现算法。共设计 3 组对比实验,对算法的有效性和准确性进行验证。

实验内容如下:

(1)对比分析本文提出的改进算法与其它中心性算法中心性值的相关性。单调性表现越明显,算法之间的相关性越好。

(2)采用数学计算,度量 Kendall tau 距离计算相关系数。以本文提出的改进算法为标准,对该算法与其它方法之间的相似性量化分析,相似性越大说明算法越有效。

(3)基于 SIR 多源传播感染模型,在不同感染概率下,对各算法识别出的 TopK 节点在稳定状态下平均感染的节点数进行比较。

表3 数据集参数
Tab. 3 Parameters of datasets

网络	n	m	$\langle k \rangle$	k_{max}	$\langle d \rangle$	C	R	H
karate	34	78	4.59	17	2.41	0.571	-0.476	0.562
USAir	332	2126	12.81	139	2.74	0.625	-0.208	0.539
911terriset	62	148	4.77	22	2.98	0.436	-0.070	0.558
NetScience	379	914	4.82	34	6.04	0.741	-0.081	0.391
Email	1133	5451	9.62	71	3.61	0.220	0.078	0.452
Freemans_EIES_3	32	442	27.625	60	1.559	-	-0.406	0.85
Organisational	77	2228	57.870	130	1.605	-	-0.026	0.850
Celegansneural	297	2359	15.886	139	3.072	-	-0.226	0.606

GAS_D 算法在 5 个无向网络中,各算法的中心性值与其它中心性算法之间的相似性分析图,以及 Kendall tau 的相似性系数如图 2 所示。与其它 3 种中心性算法相比,本文提出的改进算法 GAS_D 与改进前的 AS 算法相关性最强,单调性表现最明显, kendall tau 的相似性值最大,与 DC、CC 算法的相关性次之。通过对单调性表现进行观察,GAS_D 与 CC 具有更明显的相关性,与 BC 算法的相关性最不明显。GAS_D 与 DC、CC 算法共同考虑了节点的度信息、节点间的最短路径,表现出了明显的相关性。在 5 个不同的数据集下,本文提出算法与改进前 AS、DC、BC、CC 算法的平均相似性见表 4。karate 数据集包含 32 个节点,且各算法识别出的 Top10 节点差别不大,故而基于 SIR 传播模型,进一步讨论其他 4 个数据集各算法识别出的 Top10 节点在不同感染概率下达到稳定状态的感染节点数占节点总数的百分比(式(14)),以此衡量 Top10 节点的感染能力(如图 3 所示)。

$$f(t_c) = \frac{R(t_c)}{N} \quad (14)$$

在 4 个数据集中均可以明显看出,当感染概率足够大的情况下,各方法识别出的 Top 结果在稳定状态下均可以覆盖整个网络;各中心性算法识别出的 Top10 节点不同的感染概率 $\beta \in (0, 1]$ 下,达到稳定状态的结果趋势一致,没有表现出明显差异,说明了算法的有效性;在感染概率 $\beta = 0.5$ 左右,911terriset、Netscience 数据集中,改进的 GAS_D 算法均优于改进之前的 AS 算法,感染节点数更多,说明本文算法的有效性,并且优于改进之前的算法。

文献[8]提出的吸收中心性方法并没有在有向网

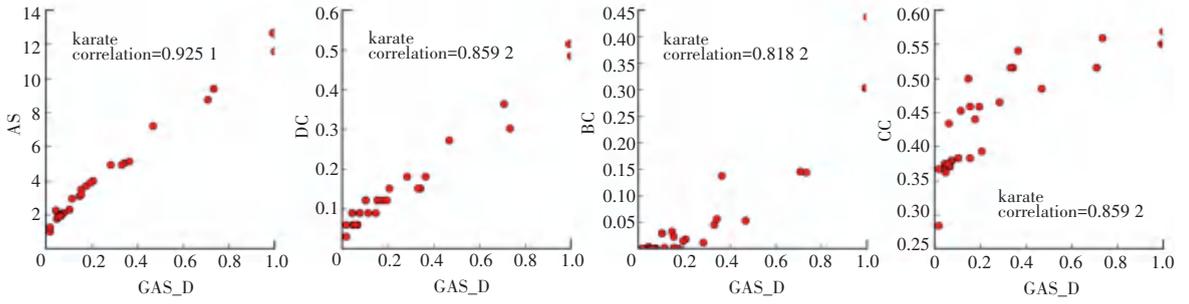
络中进行拓展研究,本文将其拓展方法记为 ZAS,并与提出的有偏随机游走改进之后的 GAS_Dout、GAS_Din 算法,在 3 个真实的网络进行对比研究,各算法之间的相关性及 Kendall tau 相似性系数如图 4~图 6 所示。在 Freeman、Organisational、Celegansneural 数据集中,ZAS 算法识别结果更接近于 PageRank、LeaderRank;GAS_Din 算法识别结果更接近于 Pro_PageRank 算法;GAS_Dout 算法识别结果更接近于 DPRank。在 3 个数据集下,ZAS、GAS_Din、GAS_Dout 分别与 PageRank、LeaderRank、Pro_PageRank、DPRank 的相似性见表 4。

由表 4 中数据可见,ZAS、PageRank、LeaderRank 均为等概率的向邻居节点进行转移,而 GAS_Din、Pro_PageRank 更倾向于向入度大的节点进行转移,GAS_Dout、DPRank 更倾向于向出度大的节点进行转移,故算法之间表现出更强的相关性。Top10 节点在不同感染概率下达到稳定状态下的感染节点率如图 7 所示,各算法识别结果的总体趋势相同,一定程度说明了 ZAS、GAS_Din、GAS_Dout 算法的有效性,GAS_Dout 的识别结果也明显优于 GAS_Din、ZAS 算法,GAS_Dout 的准确性更高。

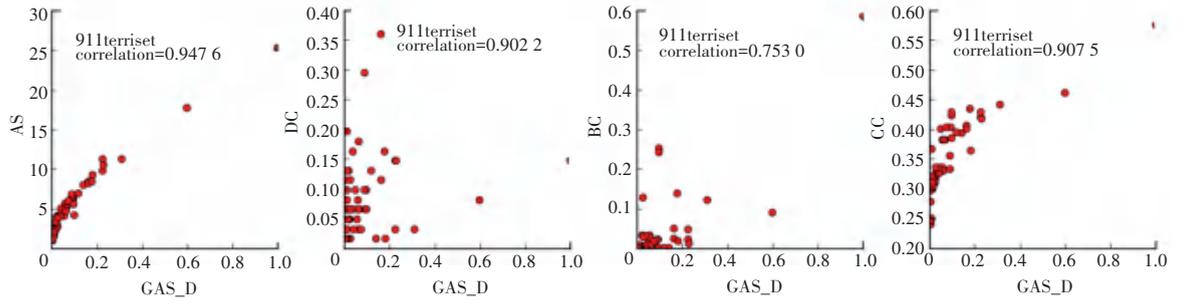
表 4 本文提出算法与其他算法识别结果的相似性对比

Tab. 4 The similarity comparison between the recognition results of the proposed algorithm and other algorithms

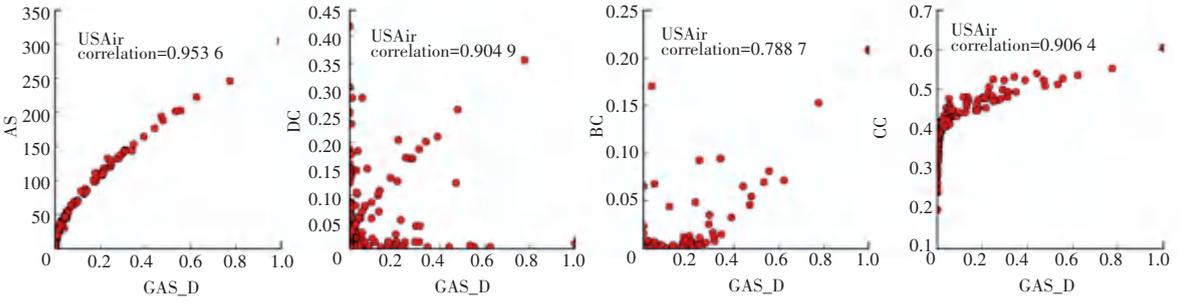
方法	AS	DC	BC	CC	平均相似性
GAS_D	0.928	0.884	0.777	0.886	0.869
方法	PR	LR	PPR	DPRank	平均相似性
ZAS	0.932	0.929	0.899	0.811	0.829
GAS_Din	0.856	0.869	0.925	0.818	0.867
GAS_Dout	0.803	0.813	0.842	0.959	0.854



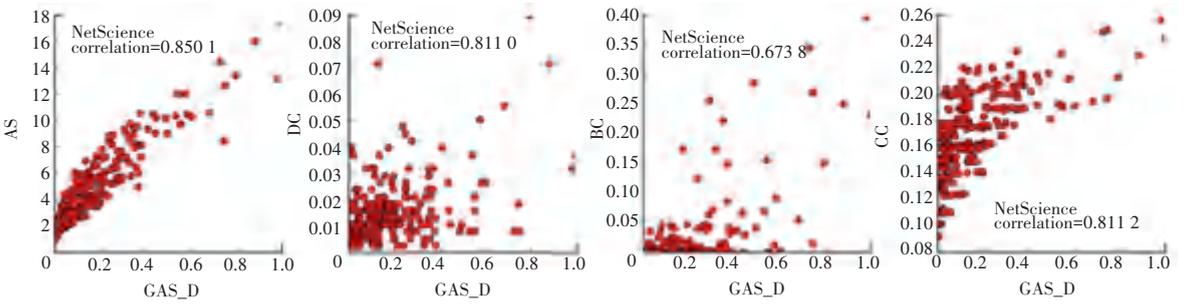
(a) karate 网络



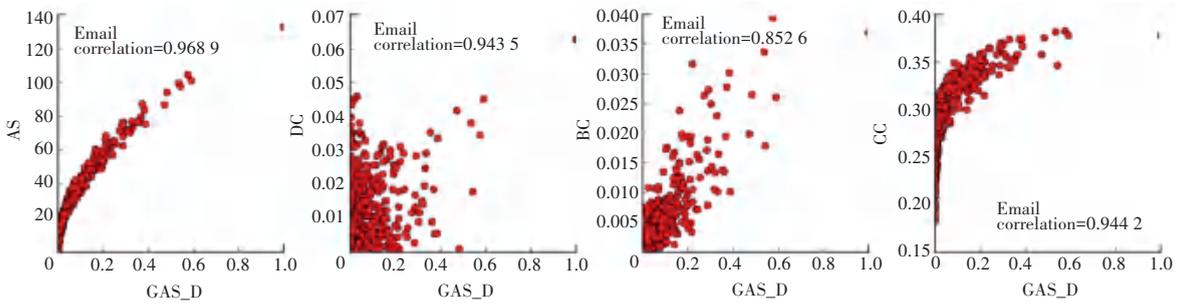
(b) 911terriset 网络



(c) USAir 网络



(d) NetScience 网络



(e) Email 网络

图2 不同网络中 GAS_D 算法与其它中心性算法的相关性及 Kendall tau 相关系数

Fig. 2 Correlation between GAS_D algorithm and other centrality algorithms and Kendall tau correlation coefficient in different networks

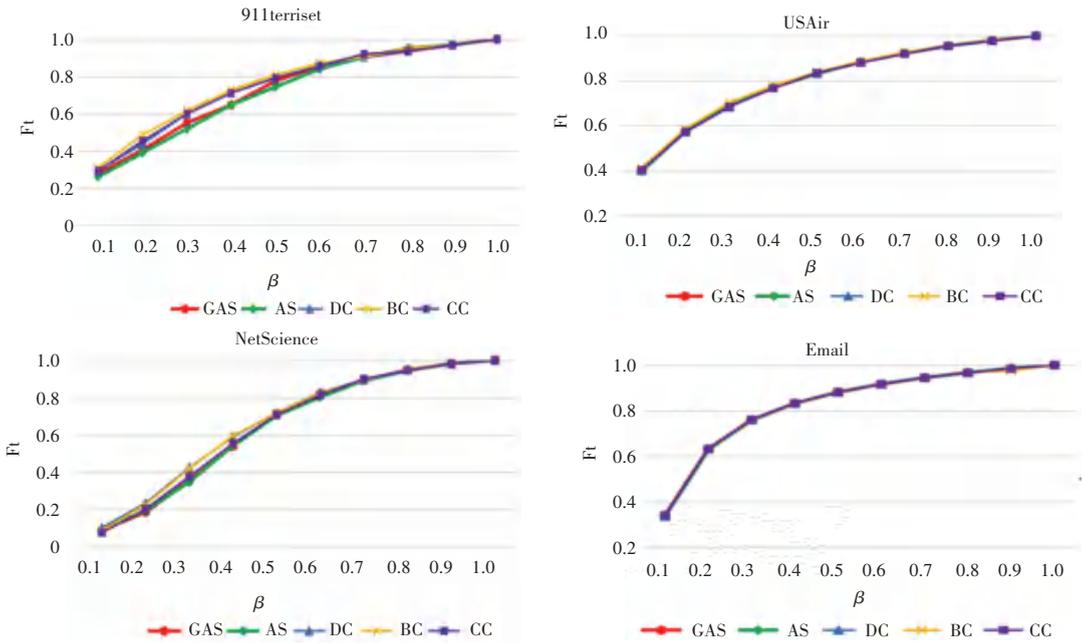
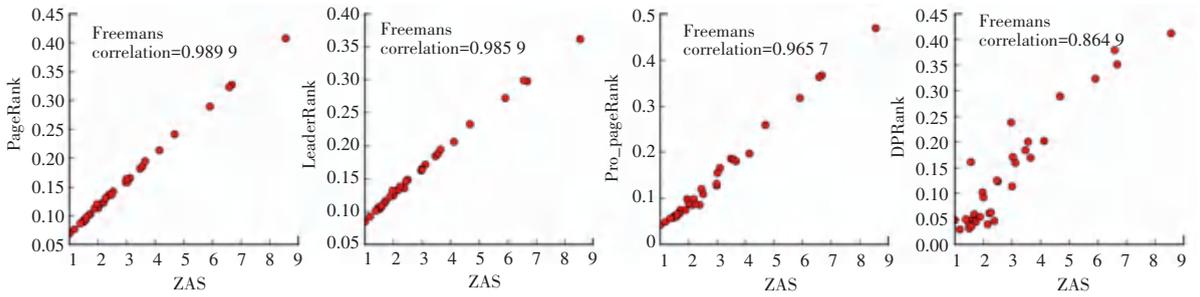
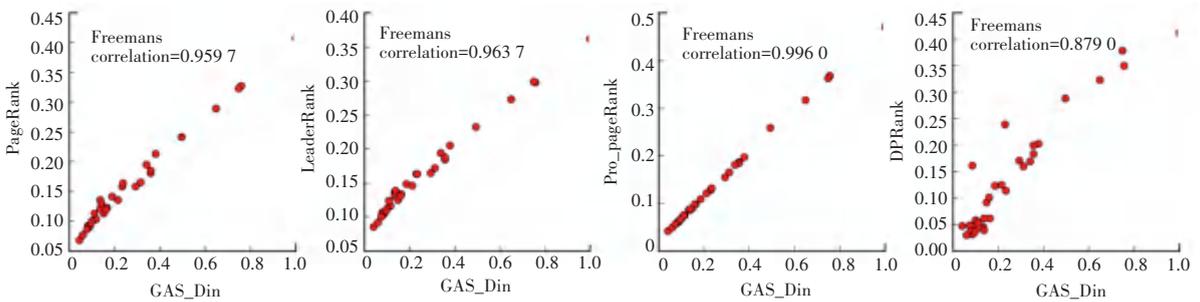


图3 无向网络中各中心性算法识别出的 Top10 节点在不同感染概率下平稳状态值

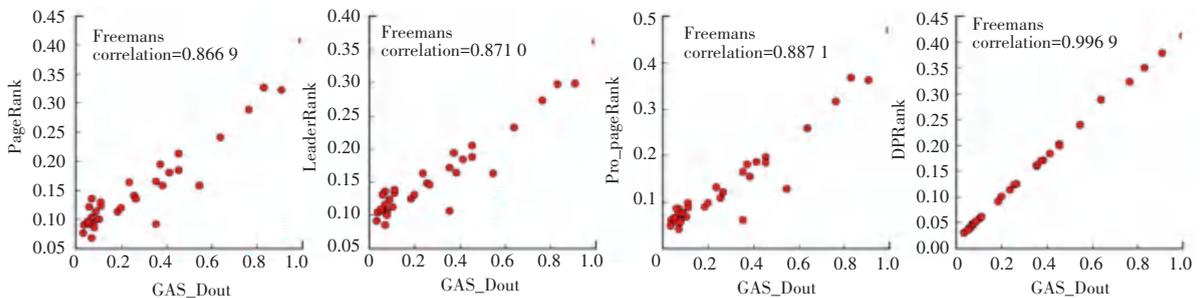
Fig. 3 The stationary state values of top 10 nodes identified by centrality algorithms under different infection probabilities in undirected networks



(a) ZAS 与其它中心性算法



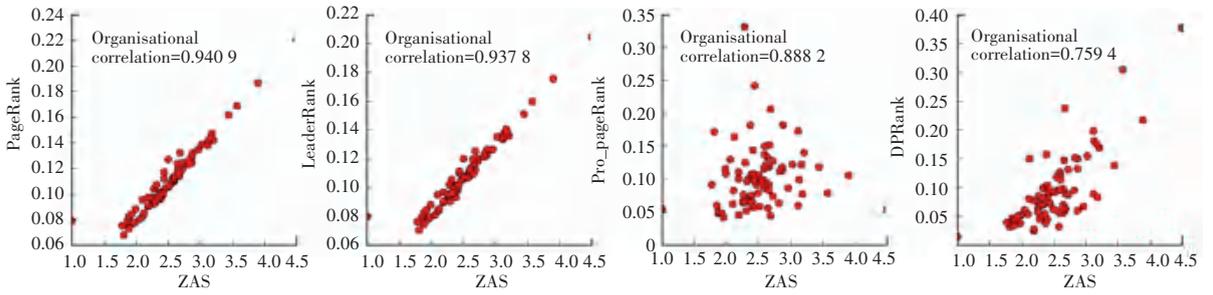
(b) GAS_Din 与其它中心性算法



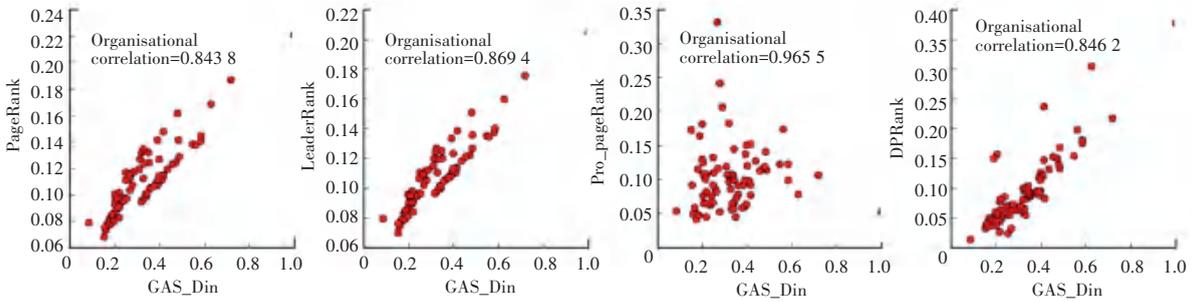
(c) GAS_Dout 与其它中心性算法

图4 Freemans 中 ZAS、GAS_Dout、GAS_Din 算法与其它中心性算法的相关性及 Kendall tau 相关系数

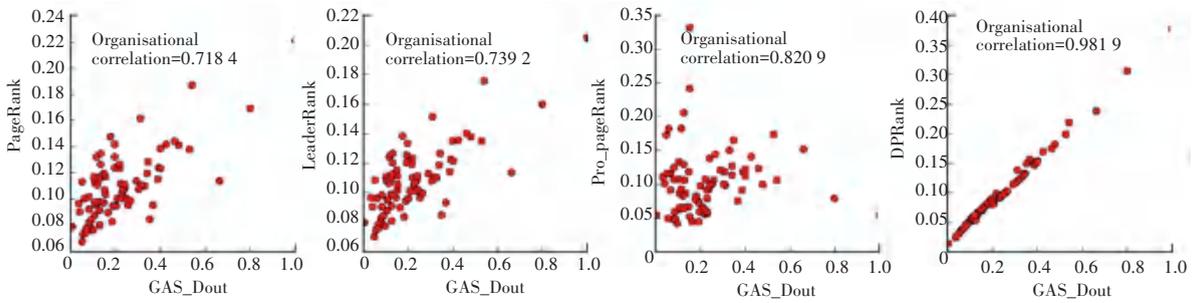
Fig. 4 Correlation between ZAS、GAS_Dout、GAS_Din algorithm and other centrality algorithms and Kendall tau correlation coefficient in Freemans



(a) ZAS 与其它中心性算法



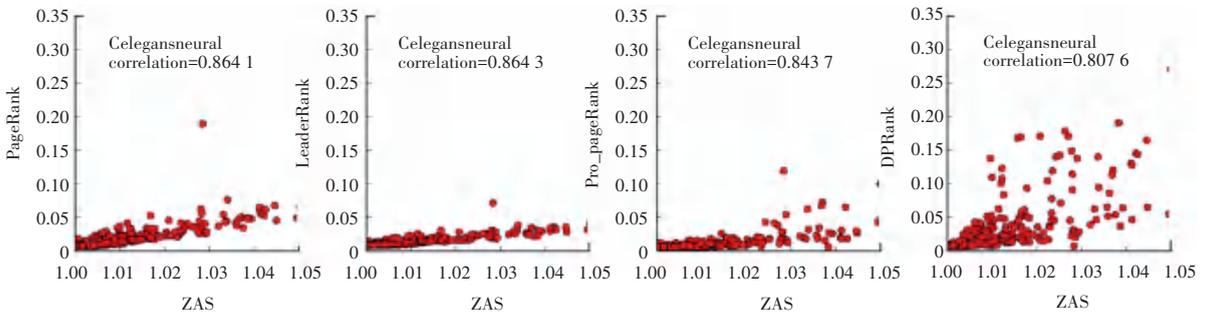
(b) GAS_Din 与其它中心性算法



(c) GAS_Dout 与其它中心性算法

图5 Organisational 中 ZAS、GAS_Dout、GAS_Din 算法与其它中心性算法的相关性及 Kendall tau 相关系数

Fig. 5 Correlation between ZAS、GAS_Dout、GAS_Din algorithm and other centrality algorithms and Kendall tau correlation coefficient Organisational



(a) ZAS 与其它中心性算法

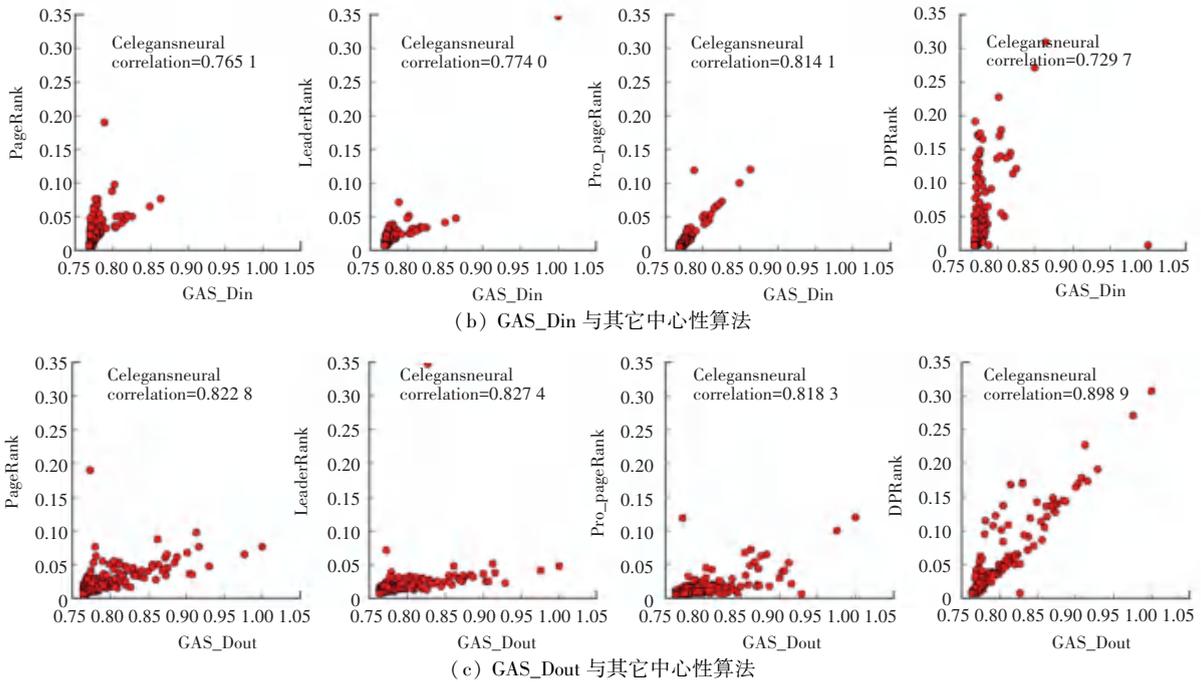


图6 Celegansneural 中 ZAS、GAS_Dout、GAS_Din 算法与其它中心性算法的相关性及 Kendall tau 相关系数

Fig. 6 Correlation between ZAS、GAS_Dout、GAS_Din algorithm and other centrality algorithms and Kendall tau correlation coefficient in Celegansneural

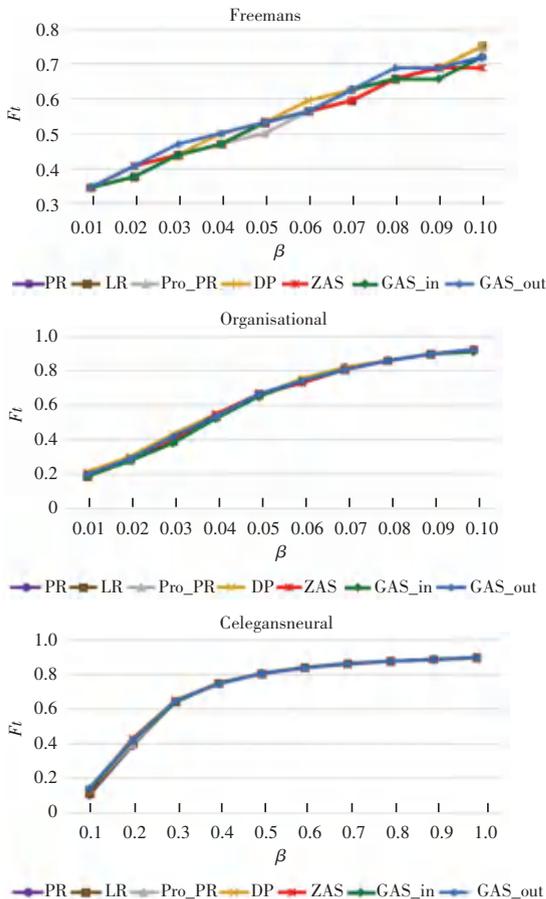


图7 有向网络中各中心性算法识别出的 Top10 节点在不同感染概率下平稳状态值

Fig. 7 The stationary state values of top 10 nodes identified by centrality algorithms under different infection probabilities in directed networks

4 结束语

本文基于无向网络中等概率的吸收中心性方法, 基于节点的度信息, 提出适用于无向网络的 GAS_D 算法进行不等概率有偏的关键节点识别, 并与节点的出度信息结合, 将 AS 算法拓展到有向网络, 结合邻居节点的出度和入度信息, 提出 GAS_Dout、GAS_Din 算法。通过真实的无向网络和有向网络设计仿真实验证明: 在无向网络中, 本文提出的考虑现实网络中信息转移倾向性的改进算法优于改进之前的算法; 在有向网络中, 考虑邻居节点的出度节点信息对信息的扩散更有利, 本文算法在有向网络中的扩展, 也解决了 PageRank 及其改进算法的参数问题。后续工作将边值的权重考虑在内进行进一步的拓展研究。

参考文献

- [1] LI M, ZHANG Q, DENG Y. Evidential identification of influential nodes in network of networks[J]. Chaos, Solitons and Fractals. 2018, 117: 283-296.
- [2] 王丹, 张海涛, 刘雅姝, 等. 微博舆情关键节点情感倾向分析及思想引领研究[J]. 图书情报工作. 2019, 63(4): 15-22.
- [3] 彭秀芬. 生物网络关键节点识别方法研究进展[J]. 生物学杂志. 2017, 34(4): 104-109.
- [4] LIAO H, MARIANI M S, MEDO M, et al. Ranking in evolving complex networks[J]. Physics Reports. 2017, 689: 1-54.
- [5] WASSERMAN S, FAUST K. Social network analysis: Methods and applications.[J]. Contemporary Sociology. 1995, 91(435): 219-220.
- [6] ZACHARY W W. An Information Flow Model for Conflict and

- Fission in Small Groups[J]. Journal of Anthropological Research. 1977, 3(4): 452-473.
- [7] FREEMAN L C. Centrality in social networks conceptual clarification [J]. Social Networks. 1978, 1(3): 215-239.
- [8] ZHANG T, LIANG X. A Novel Method of Identifying Influential Nodes in Complex Networks Based on Random Walks[J]. Journal of Information & Computational Science. 2014, 11(18): 6735-6740.
- [9] BRIN S, PAGE L. The anatomy of a large-scale hypertextual Web search engine[J]. Computer Networks and ISDN Systems. 1998, 30(1): 107-117.
- [10] LV L, ZHANG Y C, YEUNG C H. Leaders in Social Networks, the Delicious Case[J]. Plos One. 2011, 6(6): e21202.
- [11] WANG X S, MA J C, BI K Y, et al. A Improved PageRank Algorithm Based on Page Link Weight [J]. Algorithms and Architectures for Parallel Processing. 2014, 8630: 720-727.
- [12] LIU M, XIONG Z, MA Y, et al. DPRank centrality: Finding important vertices based on random walks with a new defined transition matrix[J]. Future Generation Computer Systems. 2018, 83: 376-389.
- [13] 张旭,袁旭梅,袁继革.基于加权改进节点收缩法的供应链网络节点重要度评估[J].计算机应用研究.2017,34(12):3801-3805.
- [14] BIAN T, DENG Y. A new evidential methodology of identifying influential nodes in complex networks [J]. Chaos, Solitons & Fractals. 2017, 103: 101-110.
- [15] WU X G. Identify influential nodes in complex networks based on modified TOPSIS [C]// 2017 36th Chinese Control Conference (CCC), Dalian, 2017:1474-1479.
- [16] BIAN T, HU J, DENG Y. Identifying influential nodes in complex networks based on AHP [J]. Physica A: Statistical Mechanics and its Applications. 2017, 479: 422-436.
- [17] FEI L, DENG Y. A new method to identify influential nodes based on relative entropy [J]. Chaos, Solitons & Fractals. 2017, 104: 257-267.
- [18] GUO L, LIN J, GUO Q, et al. Identifying multiple influential spreaders in term of the distance - based coloring [J]. Physics Letters A. 2016, 380(7): 837-842.
- [19] BAO Z, LIU J, ZHANG H. Identifying multiple influential spreaders by a heuristic clustering algorithm[J]. Physics Letters A. 2017, 381(11): 976-983.
- [20] 宋琛,尹波,张贤坤.基于叠加随机游走的复杂网络节点重要度评估方法[J].信息工程大学学报.2016,17(6):730-734.
- [21] 郑苏苏,付晓东,岳昆,等.基于 Kendall tau 距离的在线服务信誉度量方法[J].计算机研究与发展.2019,56(4):210-220.

(上接第 185 页)

4 结束语

本文针对 Hex 棋在 UCT 算法中所得结果准确度不够精确的问题,提出了一种结合 Hex 棋棋型策略改进的 UCT 算法。该算法对模拟过程采取了一系列的棋型策略,使得 UCT 模拟阶段的模拟准确度大幅提升,有效的提高了 Hex 棋的搜索效率与博弈水平。以此算法开发的 Hex 棋博弈程序获得了 2021 年全国大学生计算机博弈大赛 Hex 棋项目一等奖。虽然改进的算法的博弈水平有所提升,但还

存在不够精准的问题,因此结合神经网络,进行强化学习等成为为下一步的研究方向。

参考文献

- [1] 徐心和,邓志立,王骄,等.机器博弈研究面临的各种挑战[J].智能系统学报,2008(4):288-293.
- [2] 计算机博弈大赛.Hex 棋游戏[EB/OL].<http://computergames.caii.cn/jsgz19.html>, 2014.
- [3] HUANG S, ARNESON B, HAYWARD R B, et al. MoHex 2.0 a pattern-based MCTS Hex player[C]// International Conference on Computers and Games, Yokohama, Japan, Springer, 2013.
- [4] BRODERICK ARNESON R B H P. Monte Carlo Tree Search in Hex[J]. Trans. on Comput'1 Intel. and AI in Games, 2010, 2(4): 251-257.

(上接第 188 页)

3 结束语

传统的 K-means 聚类算法在应用于图像分割时,聚类中心的随机选取存在着算法迭代次数多,收敛时间长的问题。本文提出了一种结合人工萤火虫群优化算法与 K-means 聚类算法用于裂缝图像分割。该算法在初始化聚类中心之前,利用人工萤火虫群优化算法的全局寻优能力,快速逼近全局最优的聚类中心,然后在此基础上使用 K-means 聚类分割图像。将本文算法和传统的 K-means 聚类算法进行对比试验,结果表明本文算法可以很好地分割出裂缝骨架,提高 K-means 算法的抗干扰性,减少达到收敛的迭代次数,保证了沥青路面裂缝分割结果的准确性。

参考文献

- [1] 邵闯,王生怀,邹春龙,等.基于混合粒子群优化的 2 维 Otsu 路面裂缝图像阈值分割方法[J].湖北汽车工业学院学报,2019,33(3):53-57.
- [2] 韦春桃,余俊辰,赵平,等.基于自适应阈值的细小裂缝与微灰度差异裂缝自动检测方法[J].中外公路,2019,39(1):58-63.
- [3] 刘晟.基于改进的局部阈值分割的阴影路面裂缝提取方法[J].无线互联科技,2018,15(20):112-113.
- [4] 瞿中,鞠芙蓉,陈思琪.结构森林边缘检测与渗流模型相结合的混凝土表面裂缝检测[J].计算机科学,2018,45(11):288-291,311.
- [5] 刘丰年,朱平哲.基于相关特征约束的分水岭路面裂缝图像分割算法[J].微型机与应用,2016,35(3):96-99.
- [6] 李鹏,李强,马味敏,等.基于 K-means 聚类的路面裂缝分割算法[J].计算机工程与设计,2020,41(11):3143-3147.
- [7] LU Guofeng, ZHAO Qiancheng, LIAO Jianguo, et al. Pavement crack identification based on automatic threshold iterative method [C]// International Conference on Electronics and Information Engineering, 2017:320-325.