

# 基于不等概率叠加随机游走关键点识别

宁 阳,武志峰,张 策

(天津职业技术师范大学 信息技术工程学院,天津 300222)

**摘 要:**关键节点识别是网络科学的重要研究内容,在医学、社会学、网络安全、电力交通、政治与经济学领域有重要研究意义。当前流行的关键点识别算法的原理是通过考虑局部范围和全局范围网络节点的特性衡量节点中心性,结合节点自身及邻居节点贡献进行关键节点识别。存在识别有效性低和时间复杂度高的问题,不能在大规模网络中扩展。针对等概率叠加随机游走关键点识别方法没有考虑随机游走倾向性问题,采用节点相似性构造转移概率矩阵的方法,开展了不等概率叠加随机游走进行关键点识别的研究。通过在无向网络中与度中心性、介数中心性、接近中心性、等概率叠加随机游走评估方法间进行比较,各中心性算法与 SIR 模型的相关性比较的实验,证明基于不等概率叠加随机游走能以较高的精度进行网络中关键点识别,并且优于等概率叠加随机游走方法。

**关键词:**Jaccard 相似度;叠加随机游走;关键点识别;SIR 传播模型;Kendall tau 距离

中图分类号:TP393.02

文献标识码:A

文章编号:1673-629X(2020)08-0199-07

doi:10.3969/j.issn.1673-629X.2020.08.035

## Key Point Recognition Based on Unequal Probability Superposition Random Walk

NING Yang, WU Zhi-feng, ZHANG Ce

(School of Information Technology Engineering, Tianjin University of Technology and Education, Tianjin 300222, China)

**Abstract:**Key node identification is an important research content of network science, which has important research significance in the fields of medicine, sociology, network security, power transportation, politics and economics. The principle of the current key point recognition algorithm is to measure the centrality of nodes by considering the characteristics of local and global network nodes, and to identify key nodes by combining the contributions of nodes themselves and their neighbors. There are some problems such as low recognition efficiency and high time complexity, which cannot be extended in large-scale networks. Aiming at identifying key points of equal probability stacked random walk without considering the tendency of random walk, we use the method of constructing transition probability matrix by node similarity to carry out the research of identifying key points of different probability stacked random walk. By comparing the evaluation methods of degree centrality, closeness centrality, betweenness centrality and equal probability stacking random walk in undirected networks, and comparing the correlation of each centrality algorithm with SIR model, it is concluded that random walk based on unequal probability stacking can identify key points in networks with high accuracy. And it is superior to the conclusion of equal probability superposition random walk method.

**Key words:**Jaccard similarity; superimposed random walk; key node identification; SIR spread model; Kendall tau distance

## 0 引言

网络科学研究的是看起来互不相同的复杂网络之间共性和处理的普适方法。网络科学中研究的问题来源于各种实际网络,网络中的关键点识别是网络科学的重要研究内容之一。根据传播动力学的研究形成的理论和方法更好地认识不同网络上的传播行为之间的

联系与区别。关键点识别的研究在不同的领域具有重要意义,例如在社会网络中找到最有影响力的人可以控制流言的传播,疾病传播过程中找到易感人群,对疾病进行有效的预防和控制,城市交通系统、电力系统中找到关键枢纽进行重点维护,降低经济损失风险等。有效地评价和衡量网络中节点的重要性,借助图论的

收稿日期:2019-09-15

修回日期:2020-01-17

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

作者简介:宁 阳(1995-),女,硕士研究生,CCF 会员(D8877G),研究方向为复杂网络、链路预测、数据挖掘;武志峰,教授,硕导,CCF 会员(07407M),研究方向为演化计算、数据挖掘、机器学习。

概念和术语,将具体实际问题抽象为图,得到网络的拓扑性质,将多学科融合在一起作为研究方向,具有广泛的应用价值。

首先介绍提出不等概率叠加随机游走的相关工作,然后介绍不等概率叠加随机游走方法,构建不等概率转移矩阵,进行叠加。随机游走,通过相似和衡量节点重要性。通过对 3 个数据集的仿真实验,与度中心性、介数中心性、接近中心性和等概率叠加随机游走进行比较,并与 SIR 标准传播模型<sup>[1-2]</sup>进行 Kendall tau 距离<sup>[3-4]</sup>相关性分析,验证该方法的有效性。

## 1 相关工作

复杂网络中关键点识别的排序方法很多,文献[5]综述了关键点识别问题和方法,并对其进行分类,描述了最重要的进展和最新技术。近年来,学者通常在衡量节点的常用指标上进行改进。无向网络中几个常用衡量节点重要性的指标包括度值、介数、接近数、 $k$ -壳值和特征向量<sup>[6-10]</sup>。有向网络中经典的两个算法是 Kleinberg 提出的 HITS 算法<sup>[11]</sup>及 Page 和 Brin 提出的 PageRank 算法<sup>[12]</sup>。度中心性通过衡量一个节点的度,度越大节点越重要,即与节点在网络中所处的位置有关;介数中心性是经过某一节点整个网络中所有节点对之间的最短路径的数量反映节点重要性指标;接近中心性通过计算当前节点到网络中其他所有节点的距离在某种程度上反映节点重要性; $k$ -壳值通过不断地在原始网络中去除度值为 1, 2, 3... 的节点及其连边,从而进行节点重要性识别的一种推广的度值指标;特征向量中心性既考虑了邻居节点的数量又考虑了邻居节点的重要性<sup>[1]</sup>。HITS 算法通过刻画节点的权威性和枢纽性指标,衡量节点重要性。PageRank 算法通过指向当前节点的数量和质量衡量节点重要性。文献[13]提出通过节点间的相互作用力构建随机游走模型中的概率转移矩阵,从物理学的角度考虑网络中边的实际意义。文献[14]考虑节点度及邻居节点拓扑重合度,获取节点两步内的邻居节点信息,通过计算节点之间相似度衡量节点重要性。文献[15]提出半局部中心性方法,解决大规模网络中时间消耗大的问题,仅考虑了节点两步转移到达节点的数量及节点度。文献[16]考虑节点二阶邻居节点,节点更有远见的转移程度(出度)大的节点,改进 PageRank 指标计算转移概率矩阵的平稳分布衡量节点重要性。文献[17]结合节点中心性和边的中心性指标在无向网络中重新定义转移概率矩阵衡量节点重要性。文献[18]提出了一种叠加游走相似和表征节点重要性的方法,考虑节点度、邻接节点的属性以及节点在网络中的位置,基于等概率的随机游走,但未考虑实际网络中游走的倾

向性。

针对上述问题,以及受现有研究的启发,考虑节点之间的相似性<sup>[19]</sup>,进行有偏的随机游走,考虑两步到达节点的概率,提出一种基于不等概率叠加随机游走的重要节点识别算法。该方法将 Node2 vec<sup>[20]</sup>中提到的随机游走方法 2nd order 随机游走与有叠加效应的局部随机游走指标<sup>[18-19]</sup>相结合,考虑节点之间的相似性,同时考虑不等概率随机游走。2nd order 随机游走下一步的选择不再是等概率随机的,而是以不等概率进行随机游走。这里引入不等概率的随机游走,考虑有限次的游走步数、节点度的信息、节点之间的相似性作为不等概率随机游走的依据,进行无向网络中的关键节点识别。

## 2 不等概率叠加随机方法

文中提出的基于不等概率叠加随机游走关键点识别方法主要包括 3 个阶段,其中 2.1 节描述不等概率随机游走转移矩阵的构建,根据 Jaccard 指标<sup>[19,21]</sup>计算每个节点和网络中其他节点的相似性,考虑任意两个节点间有直接连边,但是无共同邻居的情况,采用文献[18]叠加随机游走通过相似和衡量节点重要度提到的节点度分之邻接矩阵值来计算。通过归一化处理,以此来构建概率转移矩阵。2.2 节描述叠加效应的局部随机游走,根据有叠加效应的局部随机游走指标<sup>[19]</sup>,在局部随机游走指标(local random walk, LRW)<sup>[22]</sup>的基础上,将  $t$  步及其以前的结果加总得到叠加效应的局部随机游走相似性。这种方法可以增大邻近目标节点的点与目标节点相连的机会。2.3 节基于相似和的叠加随机游走<sup>[18]</sup>,根据叠加效应的局部随机游走指标,累加各行相似度,从而根据对应节点的累加相似和进行重要节点识别。

### 2.1 构建不等概率随机游走转移矩阵

随机游走(random walk)指基于过去的表现,无法预测将来的发展步骤和方向,下一步的选择是随机的,一般来讲,节点通过存在边到达相邻节点的概率是相同的,到达非邻居节点的概率是 0,即等概率的随机选择到达存在边的节点。

但实际中,从一个节点到其他邻居节点的概率不是均匀随机的,而是有偏的,所以文中提出基于不等概率进行随机游走,采用 Jaccard 指标衡量网络中节点之间的相似性,这里不仅考虑了存在边的邻居节点,同时也考虑了部分非邻居节点,即两步转移能够到达的节点。同时对于节点间存在直接连边,但是两个节点没有共同邻居,造成的 Jaccard 指标衡量相似性不足的问题,文中基于文献[18]考虑节点度的信息,作为相邻节点间转移概率。

将一个具体网络抽象为一个由点集  $V$  和边集  $E$  组成的图  $G = (V, E)$ 。顶点数记为  $N = |V|$ , 边数记为  $M = |E|$ 。两种常见的表示图的基本结构是邻接矩阵 (adjacency matrix) 和邻接表 (adjacency list)。文中将原始数据转化为邻接矩阵, 图  $G$  的邻接矩阵  $A = (a_{ij})_{N \times N}$  是一个  $N$  阶方阵, 第  $i$  行第  $j$  列上的元素  $a_{ij}$  定义<sup>[1]</sup>如下:

无权无向图:

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

设置初始状态, 将一个 walker 放置在无向无权图  $G$  任意节点  $i$ , 构造随机游走模型, walker 每一步根据节点之间相似性, 以不等概率  $p$  到达邻居节点, 同时, walker 以  $p$  到达两步转移节点。游走的每一步方向都与已经发生的事件无关, walker 经过的路线是一条马尔可夫链。

算法主要步骤如下:

(1) 使用邻接矩阵表示图的基本结构, 得到每个节点的度信息;

(2) 使用 Jaccard 指标计算节点相似性, 两节点直接相连, 没有共同邻居利用节点度信息进行计算; Jaccard 指标是在共同邻居的基础上考虑两端节点度的影响, 提出的衡量节点相似性指标。

$$s_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|} \quad (2)$$

对于节点  $v_x$ , 邻居集合为  $\Gamma(x)$ ,  $s_{xy}$  为节点  $v_x, v_y$  的相似性。

$$s'_{xy} = \begin{cases} 1 - \frac{1}{k_x} \cdot a_{xy}, & \{(x, y) \mid \Gamma(x) \cap \Gamma(y) = \emptyset\} \\ \text{Jaccard}(x, y), & \text{其他} \end{cases} \quad (3)$$

其中,  $k_x$  是节点  $x$  的度,  $a_{xy}$  是邻接矩阵中对应的值,  $s'_{xy}$  为改进后用于衡量节点相似性的指标, 概率标准化处理<sup>[23]</sup>, 得到一步和两步转移同时构成的转移概率矩阵。上述相似性指标标准化处理, 寻找足够少差异较大能够代表整个网络的节点, 使转移概率矩阵中各行元素之和均为 1。

$$c'_i = \frac{1 - c_i}{\sum_{i=1}^n c_i} \quad (4)$$

其中,  $c_i$  表示概率矩阵中的元素,  $c'_i$  表示概率标准化后对应位置的值。

## 2.2 基于叠加效应的局部随机游走

刘伟平和吕琳媛<sup>[22]</sup>考虑有限次的随机游走, 提出一种基于网络局部随机游走的相似性指标 (LRW), 在 LRW 基础上, 将  $t$  步及其以前的结果加总得到 SRW

的值, 即:

$$s_{xy}^{\text{SRW}}(t) = \sum_{l=1}^t s_{xy}^{\text{LRW}}(l) = q_x \sum_{l=1}^t \pi_{xy}(l) + q_y \sum_{l=1}^t \pi_{yx}(l) \quad (5)$$

设各个节点的初始资源分布为  $q_x$ , 基于  $t$  步随机游走的相似性为:

$$s_{xy}^{\text{SRW}}(t) = q_x \cdot \pi_{xy}(t) + q_y \cdot \pi_{yx}(t) \quad (6)$$

文中采用刘伟平和吕琳媛提出的一种与度分布一致性的初始资源,  $q_x = k_x/M$ , 其中  $M$  作为网络的总边数, 对每一对节点对都相同, 计算过程忽略不计。  $\pi_{xy}(t)$  表示 walker 从节点  $x$  经过  $t$  步游走到节点  $y$  的转移概率矩阵, 一般的  $k$  步转移概率矩阵正好是一步转移概率矩阵的  $k$  次方, 可以证明  $k$  步转移概率矩阵中各行元素之和都是 1。

## 2.3 基于相似和的叠加随机游走

相似度矩阵中的值代表对应节点之间的相似度, 每一行代表当前节点与其他所有节点的相似度, 采用文献<sup>[18]</sup>提出的相似和概念衡量节点重要性。累加各行相似度, 得到基于相似和的叠加随机游走相似性指标, 将其用作网络中关键节点识别。

公式如下所示:

$$\text{Sum}_y(t) = \sum_{x=1}^n s_{xy}^{\text{SRW}}(t) \quad (7)$$

算法流程如图 1 所示。

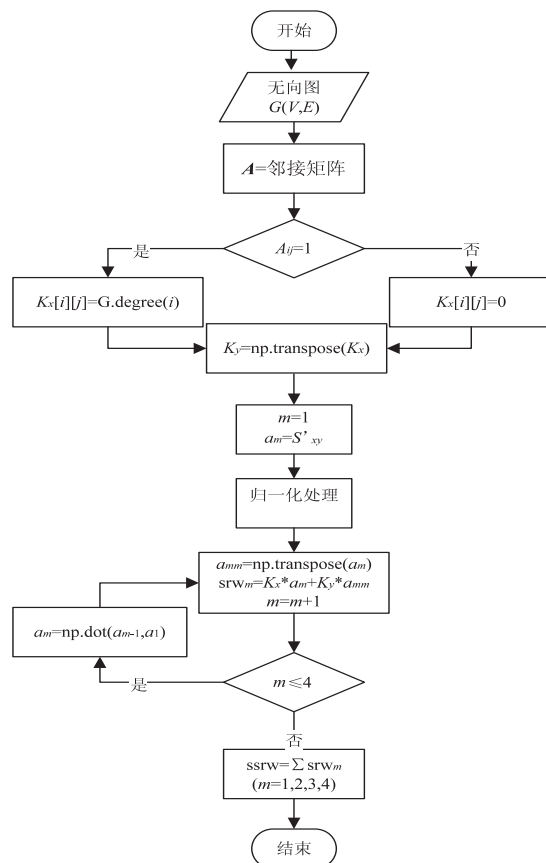


图 1 算法流程



### 3 实验结果与分析

#### 3.1 SIR 传播模型

文中使用 SIR 传播模型<sup>[1-2]</sup>得到的排序作为标准排序结果,在典型的传染病模型中, $N$ 个节点的状态可分为3类:

$S$ :易染状态,初始条件下所有节点均为易染状态,该节点以 $\beta$ 的概率被邻居节点感染。

$I$ :感染状态,感染某种病毒作为传染源的节点,该个体以 $\beta$ 概率感染其邻居节点。

$R$ :移除状态,也称免疫状态或恢复状态,感染状态节点以 $\beta$ 概率感染邻居易感节点后,以 $\gamma$ 概率变为 $R$ 移除状态,不再具有感染能力和易感特性。

采用单源感染模型,初始时刻,假设网络中只有一个节点处于感染状态,其余个体均处于易感状态,一个单位时间内,所有处于感染状态的节点以 $\beta = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$ 的概率感染其周围邻居节点,之后以 $\gamma = 1$ 的概率变为移除状态,统计达到稳定状态时,即不存在易感节点,统计处于移除状态节点的个数,用于衡量初始感染状态节点的传播能力。为减少 $\beta$ 、 $\gamma$ 参数带来的随机性,每个节点计算1 000次,利用平均值进行计算。对网络中所有节点均执行相同操作,根据节点的传播能力得到网络中节点重要度排序。其中, $\langle k \rangle$ 为平均度, $\langle k^2 \rangle$ 为平均度方。

#### 3.2 Kendall tau 距离

Kendall tau 距离<sup>[3]</sup>计算两个排序列表之间成对分歧数量,即两个完成列表 $\sigma$ 和 $\tau$ , $K(\sigma, \tau)$ 表示两个列表之间的差异性:

$$K(\sigma, \tau) = \frac{|\{(i, j) : i < j, \sigma(i) < \sigma(j) \text{ but } \tau(i) > \tau(j)\}|}{\binom{n}{2}} \quad (8)$$

$K \in [0, 1]$ ,  $K$ 值越大,相似性越小。Kendall 距离

表1 网络拓扑结构的统计特征

网络	$n$	$m$	$\langle k \rangle$	$k_{\max}$	$\langle d \rangle$	$C$	$R$	$H$
911terriset	62	148	4.77	22	2.98	0.44	-0.06	0.56
netscience	379	914	4.82	34	6.04	0.74	-0.08	0.39
USAir97	332	2 126	12.81	139	2.74	0.63	-0.21	0.54

9/11 恐怖袭击是伊斯兰恐怖组织在美国发动的四次有组织的恐怖袭击,19 名恐怖分子劫持四架客机,该无向网络结构共包含 62 个节点,148 条边,每个节点代表一名恐怖分子,包含劫持客机的 19 名恐怖分子及其同谋者。节点之间的边代表恐怖分子之间的联系。网络结构如图 3 所示,其中 4 次劫机行动中所在飞机的恐怖分子编号为 1-19(A11:1-2-3-4-5(飞行员)、UA175:5-6-7-8(飞行员)-9-10、AA77:11-12-13(飞行员)-14-15)、UA93:16-17-18(飞行员)-

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, $K'$ 值越大,两个列表之间相似性越大。

例: $\sigma = \{1, 2, 3, 4\}$ ,  $\tau = \{1, 3, 2, 4\}$ ,  $\sigma$  列表与  $\tau$  列表二元组集合如图 2 所示。

二元组						
$\sigma$	(1,2)	(1,3)	(1,4)	(2,3)	(2,4)	(3,4)
$\tau$	(1,3)	(1,2)	(1,4)	(3,2)	(3,4)	(2,4)
$K$	0	0	0	1	0	0
$K = 1, K' = 1 - 2 * 1/4 * 3 = 0.833$						

图2  $\sigma$  列表与  $\tau$  列表二元组集合

#### 3.3 仿真实验

根据上述分析,为了验证文中提出的基于不等概率叠加随机游走关键点识别方法的有效性,对 9/11 恐怖袭击网络<sup>[24]</sup>、科研合作网<sup>[25]</sup>、USAir97 数据集进行仿真实验,使用 Python 语言对算法进行实现,并与度中心性、介数中心性、接近中心性、等概率叠加随机游走方法进行比较;对 3 个数据集基于 SIR 模型计算标准排序结果,计算各中心性算法与 SIR 的 Kendall tau 距离的差异;分析算法的时间复杂度。表 1 给出了这 3 个网络的拓扑结构统计特征,其中  $n$  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<sup>[1,26]</sup>。可以看出 3 个网络具有明显的小世界性,即高聚类系数和短的平均最短路径长度,同时 USAir97 拥有明显大度节点且具有高异质性,因此具有无标度网络的特征。

19<sup>[16]</sup>,利用文中提出的算法,得到如图 4 所示的节点重要度相似和排名,与介数中心性、度中心性、接近中心性指标、等概率叠加随机游走比较,结果如表 2 所示。

在四次恐怖劫持中,不等概率叠加随机游走识别同节点度、接近中心性指标、等概率叠加均识别出 3 名飞行员(5、8、13),节点介数排序识别出 1 名,飞行员需要花费更多的资源培养,在劫持客机中是重要人员。上述算法在识别出的 Top10 节点中,度中心性可以识

别出 19 名直接参与劫机的 8 名成员,其他方法次之;在 Top15 的节点中,度中心性、接近中心性和不等概率叠加随机游走均识别出 19 名直接参与劫机的 10 名成员,高于介数中心性、等概率叠加随机游走识别出的节

点个数;Top19 节点中,基于不等概率叠加随机游走可以识别出 19 名直接参与劫机的 14 名成员,其他方法次之。从表 4 可知,文中提出的方法与标准排序结果相似性最大,显著提高了排序精度。

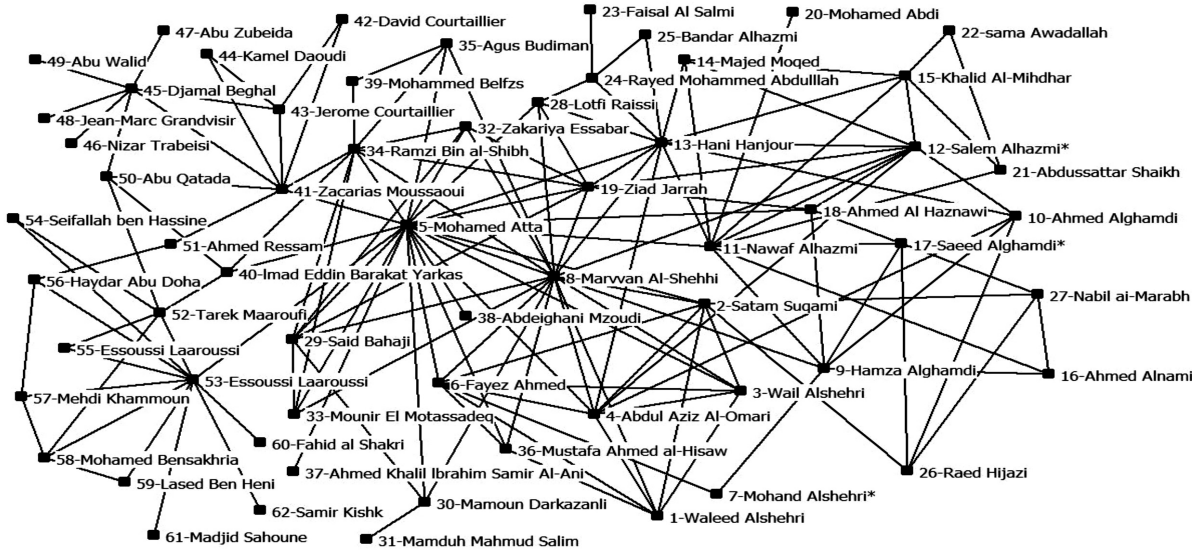


图 3 恐怖袭击网络结构

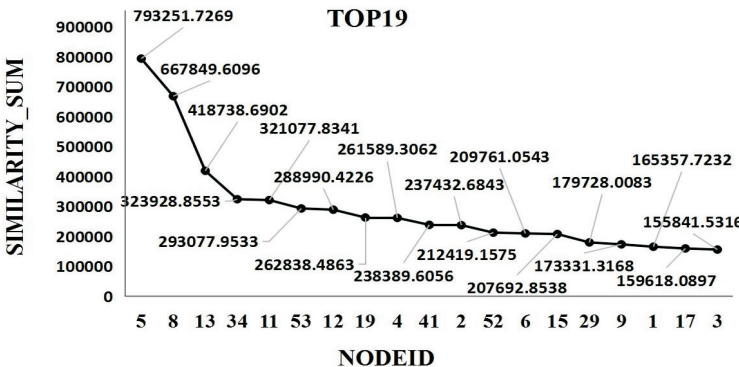


图 4 恐怖袭击网络重要度排名 Top19 的节点相似和

表 2 节点重要度比较(1)

重要性	DC	BC	CC	文献[9]	文中
1	5	5	5	13	5
2	8	53	8	8	8
3	13	41	13	5	13
4	53	11	11	41	34
5	11	45	34	19	11
6	34	13	41	12	53
7	12	8	53	43	12
8	19	2	4	53	19
9	4	34	19	34	4
10	2	24	2	4	41
11	41	30	18	11	2
12	6	52	28	29	52
13	29	50	40	32	6

续表 2

重要性	DC	BC	CC	文献[9]	文中
14	45	15	6	28	15
15	9	6	3	15	29
16	1	40	29	58	9
17	15	28	32	42	1
18	3	4	30	44	17
19	52	9	33	45	3
Top10	8	5	7	6	7
Top15	10	7	10	8	10
Top19	13	9	10	8	14

科研合作网络中网络的节点表示曾在网络科学领域发表过论文的科学家,连边表示合作关系。文中考虑无向无权网络,此数据集包含 1 589 个节点,2 742 条边,其中 128 个孤立节点,共包含 268 个连通集,文

中提取极大连通子图包含 379 个节点,974 条边。采用基于不等概率叠加随机游走关键点识别算法选取重要性排名前 5% 的 18 个节点与度中心性、介数中心性、接近中心性、等概率叠加随机游走方法进行比较,结果如表 3 所示,文中提出的算法能够通过 Top5% 的节点两步转移覆盖节点率 85%,低于介数中心性排序方法,高于节点度中心性、接近中心性和等概率叠加随机游走方法。如表 5 所示,文中方法排序结果与标准排序之间相似性次于度中心性,高于等概率叠加随机游走排序结果,能够有效地识别出网络中的关键节点。

USAir97 网络节点表示机场,边表示航空路线,该数据集包含 332 个节点,2 126 条边,通过中心性算法找到交通通畅性影响较大的机场,进行重点维护。表 3 通过比较不同中心性方法得到的 Top10 节点,不等概率叠加随机游走方法与其他中心性方法得到 Top10 的差异较小,与度中心性识别重要节点相同,介数中心性差异为 4 个,与接近中心性差异为 3 个,等概率叠加随机游走方法差异为 1 个。从表 4 可知,文中方法与 SIR 模型之间 Kendall tau 距离相似性明显高于其他中心性方法,证明了该方法的有效性,排序精度高于其他中心性方法。

表 3 节点重要度比较(2)

重要性	DC	BC	CC	文献[9]	文中
1	34	79	79	34	34
2	35	151	282	35	35
3	79	517	151	79	79
4	55	282	757	55	217
5	217	217	302	220	55
6	220	35	152	217	220
7	282	757	35	282	97
8	54	302	132	54	328
9	151	132	760	328	151
10	328	204	1 124	97	282
11	97	152	758	655	655
12	152	34	759	266	517
13	517	47	1 179	267	152
14	655	220	1 343	269	54
15	757	31	204	128	70
16	1 146	308	47	698	757
17	128	73	517	1 146	128
18	302	655	72	517	302
Top5%	83%	92%	71%	78%	85%

表 4 USAir97 节点重要度比较

重要性	DC	BC	CC	文献[9]	文中
1	118	118	118	118	118
2	261	8	261	261	261
3	255	261	67	255	255
4	182	201	255	152	182
5	152	47	204	182	152
6	230	182	182	230	230
7	166	255	47	166	166
8	67	152	248	67	67
9	112	313	166	112	112
10	201	13	112	147	201

表 5 各中心性算法与 SIR 模型 Kendall tau 距离相似性比较

网络	DC	BC	CC	文献[9]	文中
911terriset	0.878	0.712	0.880	0.784	0.882
netscience	0.796	0.660	0.670	0.744	0.763
USAir97	0.881	0.774	0.894	0.852	0.895

对比分析各算法时间复杂度,度中心性算法时间复杂度为  $O(n)$ ,介数中心性时间复杂度为  $O(n^3)$ ,接近中心性、等概率叠加随机游走,文中算法时间复杂度为  $O(n^2)$ 。文中算法在时间复杂度一致的情况下,较等概率叠加随机游走方法关键点识别的准确性有很大提高。

## 4 结束语

针对复杂网络中关键点识别问题,提出了一种基于不等概率叠加随机游走的评估方法,引入 Jaccard 相似性指标进行不等概率随机游走,结合叠加游走计算相似和在无向网络中进行关键点识别,综合考虑节点间的相似性、两步转移到达节点、节点度信息及网络中所处位置等信息。实验证明,基于不等概率叠加随机游走相似和在不增加时间复杂度的情况下,可以较高精度有效地识别关键节点。综合实际网络特点,不等概率随机游走更好地考虑了随机游走特点。在一定程度上说明不等概率随机游走在无向网络关键点识别是有效的,下一步是将不等概率随机游走与点权相结合,并将其扩展到有向加权网络,进行更深入研究。

## 参考文献:

- [1] 汪小帆,李翔,陈关荣.网络科学导论[M].北京:高等教育出版社,2012.
- [2] LI C, WANG L, SUN S, et al. Identification of influential spreaders based on classified neighbors in real-world complex networks[J]. Applied Mathematics and Computation,

- 2018,320:512–523.
- [3] 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): 884–894.
- [4] DWORK C, KUMAR R, NAOR M, et al. Rank aggregation methods for the Web[C]//Proceedings of the 10th international conference on world wide web. Hong Kong: ACM, 2001:613–622.
- [5] LÜ L, CHEN D, REN X L, et al. Vital nodes identification in complex networks[J]. Physics Reports, 2016, 650:1–63.
- [6] 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.
- [7] FREEMAN L C. A set of measures of centrality based on betweenness[J]. Sociometry, 1977, 40(1):35–41.
- [8] FREEMAN L C. Centrality in social networks conceptual clarification[J]. Social Networks, 1978, 1(3):215–239.
- [9] STEPHENSON K, ZELEN M. Rethinking centrality: methods and examples[J]. Social Networks, 1989, 11(1):1–37.
- [10] KITSACK M, GALLOS L K, HAVLIN S, et al. Identification of influential spreaders in complex networks[J]. Nature Physics, 2010, 6(11):888–893.
- [11] JON K. Authoritative sources in a hyperlinked environment[J]. Journal of the ACM, 1999, 46(5):604–632.
- [12] PAGE L, BRIN S, MOTWANI R, et al. The pagerank citation ranking: bringing order to the web[J]. Stanford InfoLab, 1999, 9(1):1–14.
- [13] 何建军,李仁发. 改进的随机游走模型节点排序方法[J]. 计算机工程与应用, 2011, 47(12):87–89.
- [14] 阮逸润,老松杨,王竣德,等. 基于领域相似度的复杂网络节点重要度评估算法[J]. 物理学报, 2017, 66(3):383–392.
- [15] CHEN D, LÜ L, SHANG M, et al. Identifying influential nodes in complex networks[J]. Physica A: Statistical Mechanics and Its Applications, 2012, 391(4):1777–1787.
- [16] 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.
- [17] LIU M, MA Y, CAO Z, et al. ECP-Rank: a novel vital node identifying mechanism combining PageRank with link prediction index[J]. Physica A: Statistical Mechanics and Its Applications, 2018, 512:1183–1191.
- [18] 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.
- [19] 吕琳媛,周 涛. 链路预测[M]. 北京:高等教育出版社, 2013.
- [20] GROVER A, LESKOVEC J. Node2vec: scalable feature learning for networks[C]//Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 2016:855–864.
- [21] JACCARD P. Etude comparative de la distribution florale dans une portion des Alpes et des Jura[J]. Bulletin of the Torrey Botanical Club, 1901, 37(142):547–597.
- [22] LIU W, LÜ L. Link prediction based on local random walk[J]. Europhysics Letters, 2010, 89(5):58007.
- [23] 李 星,钟志农,李 洋. 一种随机游走中心性的快速算法[J]. 计算机应用研究, 2013, 30(8):2337–2340.
- [24] KREBS V E. Uncloaking terrorist networks[J]. First Monday, 2002, 7(4):1–14.
- [25] NEWMAN M E J. Finding community structure in networks using the eigenvectors of matrices[J]. Physical Review E, 2006, 74(3):36104.
- [26] 周 漩,张凤鸣,惠晓滨,等. 基于熵的复杂有向网络异质性度量方法[J]. 系统工程, 2011, 29(8):123–126.