

复杂网络中基于引力模型的关键节点识别方法

徐清, 刘士虎*

云南民族大学数学与计算机科学学院, 云南 昆明

收稿日期: 2024年8月10日; 录用日期: 2024年9月2日; 发布日期: 2024年9月12日

摘要

基于引力模型的关键节点识别方法是常见的一类识别方法, 但该类方法存在以下两方面不足, 一是衡量节点“质量”时考虑的因素单一, 二是所需的运行时间代价较高。通过优化这两方面不足, 本文提出了一种改进的引力模型识别方法。首先, 综合考虑节点的度、H指数中心性和聚类系数以衡量节点“质量”, 并利用信息熵对节点“质量”进行区分。然后, 重新考虑节点的影响范围以减少所提方法的运行时间代价。最后, 将本文提出的方法与六种基准方法在八个数据集上进行比较。实验结果表明, 本文提出的方法在单调性、鲁棒性和准确性方面均表现出明显优势。与基于引力模型的识别方法相比, 本文提出方法的运行时间可以降低18.97%~87.65%。

关键词

复杂网络, 引力模型, 信息熵, 关键节点

A Method for Identifying Key Nodes Based on Gravity Model in Complex Networks

Qing Xu, Shihu Liu*

School of Mathematics and Computer Science, Yunnan Minzu University, Kunming Yunnan

Received: Aug. 10th, 2024; accepted: Sep. 2nd, 2024; published: Sep. 12th, 2024

Abstract

The key node identification method based on gravity model is a common type of identification method, but this type of method has the following two shortcomings: the first is the single factor considered when measuring “mass” of the node; the second is the high cost of required running time. Therefore, this paper proposes an improved gravity model identification method by optimizing these two shortcomings. Initially, “mass” of the node is measured by comprehensively considering

*通讯作者。

the degree, the H-index centrality and the clustering coefficient, and the information entropy is used to distinguish “mass” of the node. Then, the influence range of the node is being reconsidered to reduce the runtime cost of the proposed method. Finally, the method proposed in this paper is compared with six benchmark methods on eight datasets. Experimental results show that the proposed method has obvious advantages in monotonicity, robustness and accuracy. The running time of the proposed method in this paper can be reduced by 18.97% - 87.65% compared with the identification method based on the gravity model.

Keywords

Complex Network, Gravity Model, Information Entropy, Key Nodes

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

一般而言, 现实生活中的许多实体都可以抽象为复杂网络, 即把每一个实体抽象为一个节点, 把实体与实体之间的联系抽象为边[1]。例如在航空网络[2]中, 可以将机场抽象为节点, 将机场之间的航线抽象为边。类似的, 还有电力网络[3] [4]、社交网络[5] [6]等。识别出复杂网络中的关键节点, 不仅对于研究复杂网络的结构和功能具有重要的理论作用, 而且对人们的日常生活具有重要的指导意义。例如, 在谣言传播网络[7]中, 及时识别和控制能够迅速传播信息的关键节点, 可以有效地控制虚假信息的传播。因此, 识别出复杂网络中的关键节点已成为网络科学领域研究人员关注的研究热点之一, 复杂网络中的关键节点识别方法不断被提出。

到目前为止, 许多经典的关键节点识别方法已经被提出, 例如基于节点局部信息的度中心性[8]方法, 该方法认为邻居节点数越多的节点越关键。基于节点路径信息的介数中心性[9]方法, 该方法通过计算节点在网络中作为其他节点对间最短路径中介的次数来评估其关键程度。基于节点位置信息的 K-Shell 分解[10]方法, 该方法认为越位于网络核心位置的节点越关键。这些经典的关键节点识别方法通常只考虑了节点某一方面的信息, 以至于不能有效识别出复杂网络中的关键节点。因此, 近些年来, 综合考虑节点不同信息的关键节点识别方法不断被提出, 例如 Meng 等人[11]通过综合考虑节点的局部信息和位置信息, 提出了一种新颖的关键节点识别方法。基于物理学中引力公式提出的引力模型识别方法是其中常见的方法之一。

受万有引力定律的启发, Ma 等人[12]通过将万有引力公式推广到复杂网络的关键节点识别问题中, 提出了引力中心性方法。该方法利用 K-Shell 分解方法求得的 K-Shell 值代替引力公式中的质量, 用节点对间的最短路径长度代替引力公式中的距离, 考虑到计算所有节点对间最短路径长度的运行时间代价是较高的, 所以引力中心性方法只考虑了节点 3 阶内的邻居节点。类似的, 基于万有引力定律的启发, Li 等人[13]提出了一种新的识别方法, 称为局部引力模型。该方法利用节点的度代替引力公式中的质量, 用节点对间的最短路径长度代替引力公式中的距离, 并且只考虑小于网络平均最短路径长度一半以内的节点对。相较于引力中心性方法, 局部引力模型方法的性能有了提升。然而, 在衡量节点“质量”时, 局部引力模型方法仅考虑了节点的度, 这种单一因素的考虑在一定程度上限制了其性能。对于局部引力模型中仅考虑节点的度这一不足, Zhang 等人[14]对此进行改进并提出了拉普拉斯引力中心性方法, 该方法在衡量节点“质量”时, 不仅考虑了节点自身的度, 而且还考虑了邻居节点的度, 故该方法的性能得到了

进一步提升。总体来说, 拉普拉斯引力中心性方法只是利用了节点的度, 考虑的因素依旧单一。

基于上述分析, 为解决基于引力模型的关键节点识别方法存在的两方面不足, 即一是衡量节点“质量”时考虑的因素单一, 二是所需的运行时间代价较高, 本文提出了一种改进的引力模型方法。该方法同时考虑了节点的度、H 指数中心性和聚类系数, 并重新考虑了节点的影响范围。为评估所提方法的性能, 本文在八个数据集中进行仿真实验。实验结果表明, 与现有的六种关键节点识别方法相比, 本文提出的方法具有良好的单调性、鲁棒性以及准确性。并且与需要计算节点对间最短路径长度的引力模型识别方法相比, 本文方法的运行时间代价显著降低。

2. 基础知识

一般的, 无向无权的复杂网络可以表示为 $G=(V, E)$, 其中 $V=\{v_1, v_2, \dots, v_n\}$ 和 $E=\{(v_i, v_j) | v_i, v_j \in V\}$ 分别表示节点集与边集, $|V|=n$ 和 $|E|=m$ 表示节点的总数和边的总数。其邻接矩阵可以表示为 $A=(a_{ij})_{n \times n}$, 对于任意两个节点 v_i 和 v_j , 如果 $(v_i, v_j) \in E$, 则 $a_{ij}=1$, 否则 $a_{ij}=0$ 。 $K(v_i)=\sum_{j=1}^n a_{ij}$ 表示节点 v_i 的度, $cc(v_i)=\frac{2e(v_i)}{K(v_i)(K(v_i)-1)}$ 表示节点 v_i 的聚类系数, $e(v_i)$ 表示节点 v_i 的 1 阶邻居节点之间存在的边数, 当 $K(v_i)=1$ 时, $cc(v_i)=0$ 。

2.1. 接近中心性(CC)

接近中心性[15]方法根据节点对间的最短路径长度衡量节点的关键程度, 其计算公式为

$$CC(v_i)=\frac{n-1}{\sum_{j \neq i} d_{ij}}, \quad (1)$$

其中 d_{ij} 表示节点 v_i 和 v_j 之间的最短路径长度。

2.2. H 指数中心性(HC)

H 指数中心性[16]方法根据节点 1 阶邻居节点的度衡量节点的关键程度, 即假设节点 v_i 至少有 h 个 1 阶邻居节点的度大于或等于 h , 则节点 v_i 的 H 指数中心性值 $H(v_i)=h$ 。

2.3. K-Shell 分解(KS)

K-Shell 分解[10]方法根据节点在网络中的位置衡量节点的关键程度, 其为节点分配 K-Shell 值的步骤如下: 第一步, 计算所有节点的度, 找出度为 1 的节点并移除, 标记这些移除节点的 K-Shell 值为 1。第二步, 更新剩余节点的度, 检查是否还有度为 1 的节点。如果有, 则接着移除度为 1 的节点, 直至所有剩余节点的度都大于 1, 并标记这些移除节点的 K-Shell 值为 1; 如果没有, 则依次增加节点的度和 K-Shell 值并重复上述过程, 直至网络中所有的节点都被移除。

2.4. 引力中心性(GC)

引力中心性方法[12]通过结合节点的 K-Shell 值和最短路径长度值衡量节点的关键程度, 其计算公式如下

$$GC(v_i)=\sum_{j \neq i, d_{ij} \leq 3} \frac{KS(v_i) \times KS(v_j)}{d_{ij}^2}, \quad (2)$$

其中 $KS(v_i)$ 表示节点 v_i 的 K-Shell 值。

2.5. 拉斯引力中心性(LGC)

拉普拉斯引力中心性[14]方法通过结合节点的拉普拉斯中心性值和最短路径长度值衡量节点的关键程度, 其计算公式如下

$$LGC(v_i) = \sum_{j \neq i, d_{ij} \leq R} \frac{LC(v_i) \times LC(v_j)}{d_{ij}^2}, \quad (3)$$

$$LC(v_i) = K^2(v_i) + K(v_i) + 2 \sum_{v_j \in N_1(v_i)} K(v_j), \quad (4)$$

其中, R 为网络平均最短路径长度 $\langle d \rangle$ 的一半, 即 $R = \frac{1}{2} \langle d \rangle$, $LC(v_i)$ 表示节点 v_i 的拉普拉斯中心性值, $N_1(v_i)$ 表示节点 v_i 的 1 阶邻居。

2.6. 拉普拉斯能量中心性(LEC)

拉普拉斯能量中心性[17]方法根据节点的度及节点形成的三角形个数衡量节点的关键程度, 其计算公式为

$$LC(v_i) = K^3(v_i) + 3K^2(v_i) - 2K(v_i) + \sum_{v_j \in N_1(v_i)} \{3K^2(v_j) + 3K(v_j)\} - 6\Delta(v_i), \quad (5)$$

其中 $\Delta(v_i)$ 表示以节点 v_i 为顶点形成的三角形个数。

2.7. 信息熵

信息熵是信息论中广泛应用的基本概念之一, 可用于度量信息的信息量。对于任意的离散随机变量 $X = \{x_1, x_2, \dots, x_n\}$, 其概率分布为 $P = \{p_1, p_2, \dots, p_n\}$, 则随机变量 X 的信息熵可以通过以下公式计算, 即

$$E(X) = -\sum_{i=1}^n p_i \log_2 p_i, \quad (6)$$

其中, 信息熵的值越大, 表示其拥有的信息量越多。近年来, 信息熵作为有效的信息量度量指标, 不断被推广应用于复杂网络中关键节点识别问题的研究中[18]-[20]。

3. 提出方法

由于前人提出或改进的引力模型识别方法在衡量节点“质量”时考虑的因素比较单一, 因此, 本文通过考虑节点的度、H 指数中心性和聚类系数以改善这一问题。同时, 本文重新考虑节点的影响范围, 即只考虑节点 2 阶内的邻居节点。通过综合考虑多种因素及节点的影响范围, 本文提出一种改进的引力模型关键节点识别方法(Improved Gravity Model, 简称 IGM)。

3.1. 综合考虑多种因素

节点的度仅考虑了节点的 1 阶邻居个数, 而 H 指数中心性忽略了节点自身的度, 即对于一些节点自身度较大而 1 阶邻居节点度较小的节点, 会造成识别结果不准确的问题。此外, 无论是在计算节点度还是 H 指数中心性的过程中, 都只考虑了邻居节点的个数而没有考虑邻居节点之间的连接情况。与此相反, 聚类系数可以很好的衡量节点的邻居节点之间的连接情况。但近些年的研究表明, 高聚类系数值的节点在信息传播过程中会造成信息的重复传播, 从而影响信息传播的速度。因此, 本文将通过综合考虑

节点的度、H 指数中心性以及聚类系数, 对引力模型中衡量节点“质量”时考虑的因素单一这一问题进行改进, 改进后节点 v_i 的“质量”表示为

$$C(v_i) = \frac{K(v_i)}{K_{\max}} + \frac{H(v_i)}{H_{\max}} + \frac{1 - cc(v_i)}{1 - \min(cc)}, \quad (7)$$

其中, $K_{\max} = \max\{K(v_1), K(v_2), \dots, K(v_n)\}$, $H_{\max} = \max\{H(v_1), H(v_2), \dots, H(v_n)\}$, $\min(cc) = \min\{cc(v_1), cc(v_2), \dots, cc(v_n)\}$ 。为进一步区分不同节点的“质量”, 利用信息熵对求得的值再次进行区分。基于此, 节点 v_i 的“质量”进一步表示为

$$E(v_i) = - \sum_{v_j \in \Psi(v_i)} p(v_j) \log_2 p(v_j), \quad (8)$$

其中, $p(v_j) = \frac{C(v_j)}{\sum_{v_j \in \Psi(v_i)} C(v_j)}$, $\Psi(v_i) = N_1(v_i) \cup \{v_i\}$ 。

3.2. 改进的引力模型关键节点识别方法

在利用引力模型求节点的影响力值时, 如何衡量节点的影响范围一直是一个难题, 大部分学者将节点的影响范围定义为网络平均最短路径长度的一半。然而, 计算一个网络的平均最短路径长度所需的运行时间代价是较高的。通过对近些年关键节点识别的相关论文进行阅读, 发现许多学者在考虑节点的影响范围时, 通常考虑的是节点 2 阶内的邻居节点。因此, 本文在考虑节点的影响范围时, 只考虑节点 2 阶内的邻居节点。最终, 节点 v_i 的影响力值表示为

$$IGM(v_i) = \sum_{j \neq i, d_{ij} \leq 2} \frac{E(v_i) \times E(v_j)}{d_{ij}^2}. \quad (9)$$

4. 实验设置

这一部分将介绍本文使用的八个数据集以及三个评估指标。其中, 本文所有的实验, 均是在表 1 描述的实验环境中运行的。

Table 1. Description of experiment environment
表 1. 实验环境描述

参数项	参数值
操作系统	Windows 10
运行内存	16.0 GB
中央处理器	Intel(R) Core(TM) i5-1135G7
编程语言	Python 3.11

4.1. 实验数据集

本文采用以下八个数据集作为实验数据集, 即 Adjnoun [21]: 小说《大卫科波菲尔》中形容词和名词构成的网络; Enron [22]: 安然公司员工之间的电子邮件通讯网络; Netscience [21]: 网络科学领域中作者之间的合著关系网络; Celegans [21]: 秀丽隐杆线虫的代谢网络; Crime [22]: 由罪犯之间的联系构成的犯罪网络; Email [23]: 罗维拉-威尔吉利大学成员之间的电子邮件通讯网络; Air [23]: 由美国联邦航空

局国家飞行数据构建的航空网络; Health [24]: 由青少年之间的联系构成的网络。八个数据集的基本信息如表 2 所示, 其中, $\langle cc \rangle$ 表示网络平均聚类系数。

Table 2. The basic information of eight datasets
表 2. 八个数据集的基本信息

数据集	n	m	$\langle d \rangle$	$\langle cc \rangle$
Adjnoun	112	425	2.5356	0.1728
Enron	143	623	2.9670	0.4339
Netscience	379	914	6.0419	0.7412
Celegans	453	2025	2.6638	0.6465
Crime	829	1473	5.0400	0.0058
Email	1133	5451	3.6060	0.2202
Air	1226	2408	5.9290	0.0675
Health	2539	10455	4.5594	0.1467

4.2. 评估指标

这一部分将介绍三个评估指标, 分别是基于单调性的评估指标、基于网络连通性的评估指标以及基于传染病模型的评估指标。

4.2.1. 基于单调性的评估指标

单调性函数[25]可以用于评估不同方法的节点区分能力, 其计算公式为

$$M(R) = \left\{ 1 - \frac{\sum_{r \in R} n_r \times (n_r - 1)}{n \times (n - 1)} \right\}^2, \quad (10)$$

其中, R 是根据节点的影响力值从大到小将节点进行排序的排序等级列表, n_r 表示在排名等级 r 上的节点数, M 的取值范围是 $[0, 1]$ 。当 M 值为 0 时, 表明所有节点的影响力值是相同的, 即所有节点的排序等级都是相同的; 当 M 值为 1 时, 表明所有节点的影响力值都是不同的, 即所有节点的排序等级都是不同的。 M 值越接近于 1, 表明对应方法的节点区分能力越强。

4.2.2. 基于网络连通性的评估指标

按照排序等级列表 R 依次移除一定的比例节点后, 网络连通子图的数量和最大连通子图的规模可以用于衡量节点识别方法的鲁棒性[26]。记网络连通子图的总数为 ξ , 则最大连通子图的规模表示为

$$\tau = \frac{M_c}{n}, \quad (11)$$

其中, $M_c = \max\{N_1, N_2, \dots, N_\xi\}$ 表示最大连通子图包含的节点总数, N_ξ 表示对应连通子图包含的节点总数。当 ξ 的值越大, τ 的值越小时, 表明对应识别方法的鲁棒性越强。

4.2.3. 基于传染病模型的评估指标

识别方法的准确性可以用 SI 模型[27]进行评估。SI 模型将节点分为两种状态, 即易感状态和感染状态。在每一步感染过程中, 处于感染状态的节点以概率 β 感染其易感状态的邻居, 标记第 t 步时处于感染状态的总节点数为 $F(t)$ 。在传染过程中, 如果感染率 β 过低, 则感染状态节点只能以较低的概率感染其周围易感状态节点, 从而只有少量节点被感染。相反, 如果感染率 β 过高, 则所有节点最终都会被感

染, 使得难以区分出各个节点的真实感染能力。本文与文献[27]取值相同的参数值, 即 $\beta = 0.1$, $t = 10$ 。在相同步长内, 如果 $F(t)$ 值越大, 表示对应方法识别出的节点感染能力越强, 即识别出的节点越准确。

5. 实验结果分析

本文的实验结果分析主要包括四个部分, 分别是单调性分析、鲁棒性分析、准确性分析以及运行时间分析, 接下来将依次对这四个部分进行详细的分析。

5.1. 单调性分析

表 3 列出了七种方法在八个数据集的单调性值, 其中, 每个数据集上最高的单调性值被加粗表示。根据表 3 可知, 除了在 Netscience 数据集上, IGM 方法总是可以取得最高的单调性值, 并且在 Netscience 数据集上的单调性值也是位于第二的。KS 方法是一种基于节点位置信息的粗粒化识别方法, 故该方法将许多影响力不同节点分配到同一层中, 导致该方法在不同数据集上的单调性值都是最低的。从表 3 可以观察到, 相比于 KS 方法, HC 方法在节点区分能力上有一定的提升。LEC 方法不仅考虑了节点自身与邻居节点的度, 而且还考虑了节点的邻居互为邻居的情况。因此, LEC 方法的节点区分能力得到进一步提升, 并且在 Netscience 数据集上取得了最高的单调性值。CC 方法的单调性值基本都大于 0.99, 比起 KS 和 HC 方法, 该方法的节点区分能力是更强的。GC 方法同时考虑了节点的位置信息和路径信息, LGC 方法同时考虑了节点自身以及邻居节点的度和路径信息, 故这两种方法的节点区分能力相对来说也是比较好的。总而言之, 相较于其他六种方法, IGM 方法的节点区分能力是更强的。

Table 3. Monotonicity values for seven methods on eight datasets

表 3. 七种方法在八个数据集的单调性值

数据集	$M(CC)$	$M(KS)$	$M(HC)$	$M(GC)$	$M(LGC)$	$M(LEC)$	$M(IGM)$
Adjnoun	0.9837	0.5990	0.8110	0.9990	0.9997	0.9994	0.9997
Enron	0.9872	0.7245	0.8331	0.9998	0.9998	0.9986	0.9998
Netscience	0.9928	0.6421	0.6825	0.9946	0.9950	0.9972	0.9950
Celegans	0.9900	0.6962	0.7331	0.9967	0.9983	0.9971	0.9986
Crime	0.9982	0.4327	0.5651	0.9990	0.9986	0.9836	0.9994
Email	0.9988	0.8088	0.8583	0.9999	0.9999	0.9986	0.9999
Air	0.9992	0.3772	0.4916	0.9992	0.9997	0.9912	0.9997
Health	0.9994	0.5245	0.7986	0.9999	0.9999	0.9994	0.9999

5.2. 鲁棒性分析

图 1 给出了移除一定比例节点后 ξ 值的变化曲线, 其中横轴表示移除节点的比例, 纵轴表示 ξ 值。从图 1 中可以观察到, 在大多数情况下, IGM 方法总是可以取得最大的 ξ 值, 尤其是在 Adjnoun、Netscience、Crime 和 Air 数据集上, 可以看到明显的优势。在剩余的四个数据集上, 虽然不同方法之间的区别没有前面四个数据集明显, 但总体来说 IGM 方法依旧是更有优势的。例如在 Celegans 数据集上, 当移除前 40% 比例的节点时, IGM 方法的优势也是明显的。根据图 1 可知, 在大多数情况下, HC 方法获得连通子图数量是大于其它五种对比方法的。例如在 Air 数据集上, 当移除 20%~70% 比例的节点时, HC 方法获得的连通子图数量是大于其它五种对比方法的。相对来说, GC 和 LGC 方法总是位于中间的位置, 而 CC、KS 和 LEC 这三种方法获得的连通子图数量是较少的。例如在 Enron 数据集上, 当移除前 50% 比例的节

点时, KS 方法获得的连通子图数量总是最少的; 在 Celegans 数据集上, 当移除前 80% 比例的节点时, CC 和 LEC 方法在大多数情况下总是获得最小的 ξ 值。结合表 2, 可以观察到对于 $\langle d \rangle$ 较大的网络, CC 方法获得的连通子图数量总是较少的。

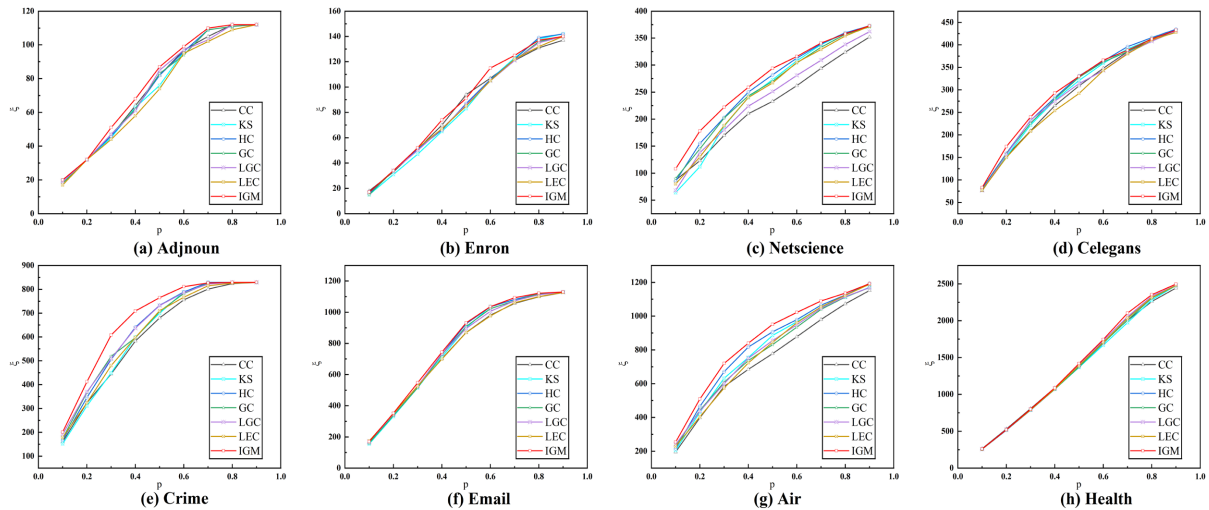


Figure 1. The changing curve of ξ value

图 1. ξ 值变化曲线

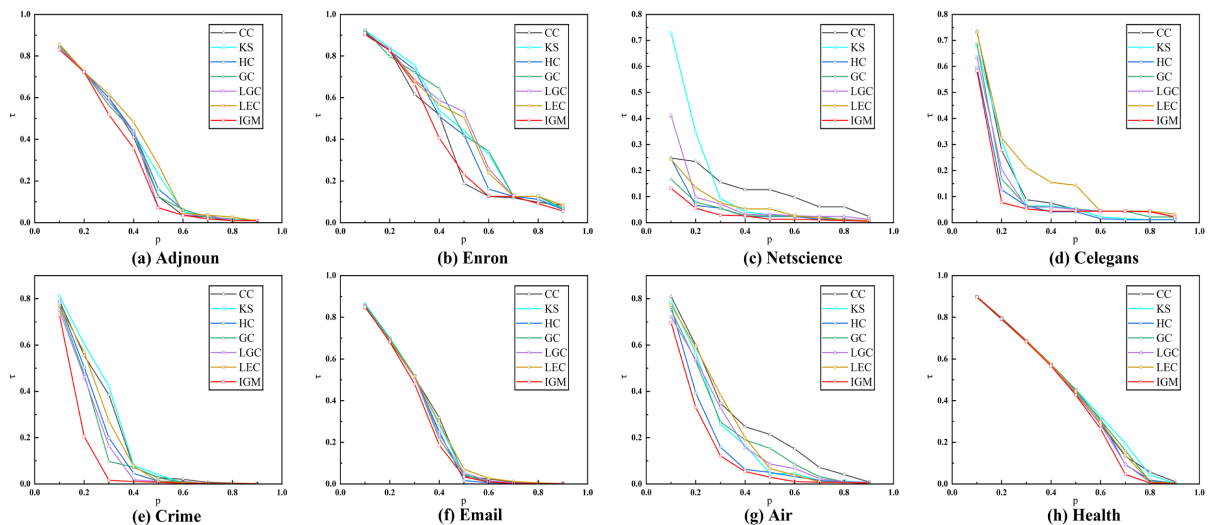


Figure 2. The changing curve of τ value

图 2. τ 值变化曲线

图 2 给出了移除一定比例节点后 τ 值的变化曲线, 其中横轴表示移除节点的比例, 纵轴表示 τ 值。从图 2 中可以观察到, 在 Adjnoun、Netscience、Crime 和 Air 数据集上, IGM 方法总是可以取得最小的 τ 值。例如, 在 Netscience 数据集上, 当移除节点比例为 10% 时, IGM 方法获得了最小的 τ 值, 并且与最大 τ 值之间的 τ 值差为 0.5963。与图 1 中的情况类似, HC、GC 和 LGC 方法总是位于中间的位置, 而 CC、KS 和 LEC 这三种方法的效果较差。根据图 1 和图 2 可知, 按一定比例移除 IGM 方法识别出的节点, 不仅可以快速增加连通子图的数量, 而且还可以使得最大连通子图的节点总数达到最少, 即说明 IGM 方法的鲁棒性是最强的。

5.3. 准确性分析

在本实验中, 选取每种方法影响力值最大的前 20 个节点为最初的感染节点, 将每种方法独立运行 100 次后, 取每一步的平均值并标记为 $\langle F(t) \rangle$, 该平均值即为每一步最终的感染节点数。图 3 给出了七种方法在八个数据集上的感染曲线, 其中横轴表示步长, 纵轴表示感染节点总数。从图 3 中可以观察到, IGM 方法在大多数情况下都可以取得最大的 $\langle F(t) \rangle$ 值, 即说明 IGM 方法识别出节点的感染能力是最强的, 也即 IGM 方法识别出的节点是更准确的。例如, 在 Health 数据集上, IGM 方法比其它对比方法存在明显的优势, 并且与最低的 $\langle F(t) \rangle$ 值之间的差值分别为 32, 116, 208, 286, 376, 378, 343, 263, 195, 129。相对来说, 在大多数数据集上, CC、GC 和 LGC 方法的 $\langle F(t) \rangle$ 值总是位于中间的位置。在 Netscience 和 Crime 数据集上, KS 方法识别出节点的感染能力是更弱的, 因为在大多数情况下, KS 方法获得的 $\langle F(t) \rangle$ 值总是更小的。在 Enron 数据集上, HC 方法识别出节点的感染能力是更弱的。在 Email 和 Health 数据集上, LEC 方法识别出节点的感染能力是更弱的。总而言之, 相比于其它六种节点识别方法, IGM 方法识别出的节点是更准确的。

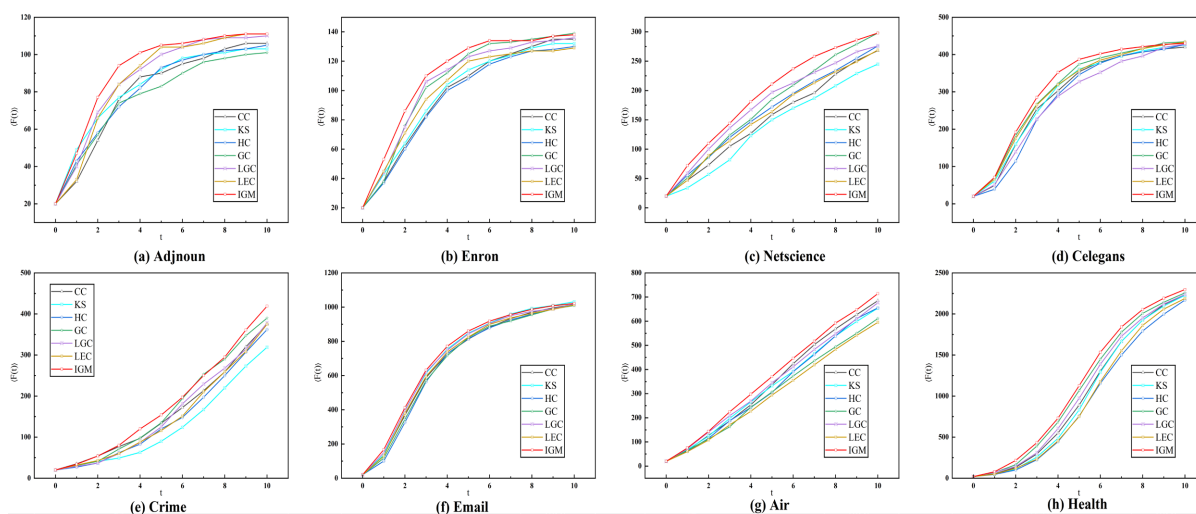


Figure 3. The changing curve of $\langle F(t) \rangle$ value

图 3. $\langle F(t) \rangle$ 值变化曲线

5.4. 运行时间分析

在本实验中, 将每种方法独立运行 100 次后, 取其运行时间的平均值作为每种方法的最终运行时间。表 4 列出了七种方法在八个数据集上的最终运行时间, 运行时间单位为秒。根据表 4 可知, IGM 方法的运行时间总是位于中间的位置, 即 IGM 方法的运行时间不是最高的, 也不是最低的。HC、KS 和 LEC 方法相对来说运行时间较短, 且三种方法均在 1 秒以内完成。然而, HC 和 KS 方法的节点区分能力较弱。虽然 CC、GC 和 LGC 方法的节点区分能力比 HC 和 KS 方法强, 但这三种方法的运行时间相对来说是更高的。尤其是 CC 和 LGC 方法, 因为它们需要计算所有节点对间的最短路径长度, 这是比较费时间的。从表 4 中可以观察到, 随着数据集规模的扩大, CC 和 LGC 这两种方法的运行时间是快速增加的, 尤其是在 Health 数据集上, CC 方法的运行时间是 IGM 的 5.6 倍, LGC 方法的运行时间是 IGM 的 8.1 倍。此外, 通过计算可得, 在所有数据集上与 GC 和 LGC 方法相比, IGM 方法的运行时间可以降低 18.97%~87.65%。综合考虑, IGM 方法的运行时间在可接受的范围内。

Table 4. The running time of the seven methods on eight datasets**表 4.** 七种方法在八个数据集上的运行时间

数据集	CC	KS	HC	GC	LGC	LEC	IGM
Adjnoun	0.0286	0.0033	0.0041	0.0324	0.0341	0.0086	0.0203
Enron	0.0453	0.0057	0.0054	0.0453	0.0520	0.0120	0.0248
Netscience	0.1633	0.0216	0.0233	0.0987	0.2490	0.0315	0.0706
Celegans	0.4041	0.0108	0.0131	0.5080	0.6250	0.0422	0.2157
Crime	1.0309	0.0155	0.0165	0.4633	1.5494	0.0363	0.2169
Email	2.4928	0.0280	0.0337	1.9417	3.4798	0.0848	0.5733
Air	2.5880	0.0215	0.0250	0.6702	3.2100	0.0539	0.4368
Health	11.0311	0.0598	0.0722	3.5867	15.8230	0.1564	1.9547

6. 结论

本文通过利用度、H 指数中心性和聚类系数衡量节点的“质量”，并利用信息熵对其进行区分，同时将节点的影响范围限制为其 2 阶内邻居节点，提出了改进的引力模型识别方法 IGM。在八个数据集上的实验结果表明，IGM 方法在单调性、鲁棒性以及准确性方面的性能是更优的。尽管本文提出的方法在运行时间上略高于 KS、HC 和 LEC 方法，但 KS 和 HC 方法通常会将相同的影响力值分配给多个节点，从而不能有效区分节点。但比起需要计算最短路径长度的 CC 和 LGC 方法，IGM 方法的运行时间是更短的，尤其是在数据规模比较大的网络中。此外，在衡量节点“质量”时，IGM 方法考虑了三种因素，未来的研究将扩展到考虑节点更多的因素，以确保其影响力得到更加合适的衡量。

基金项目

国家自然科学基金资助项目(61966039)。

参考文献

- [1] 周建林, 樊瑛, 狄增如. 复杂网络进展[J]. 北京师范大学学报(自然科学版), 2023, 59(5): 691-704.
- [2] Su, M., Luan, W., Li, Z., Wan, S. and Zhang, Z. (2019) Evolution and Determinants of an Air Transport Network: A Case Study of the Chinese Main Air Transport Network. *Sustainability*, **11**, Article 3933. <https://doi.org/10.3390/su11143933>
- [3] Li, G., Zhang, L., Wang, Y. and Kang, Z. (2023) Critical Node Identification Method of Power Grid Based on the Improved Entropy Weight Method. *Electronics*, **12**, Article 2439. <https://doi.org/10.3390/electronics12112439>
- [4] Li, J., Lin, Y. and Su, Q. (2024) Identifying Critical Nodes of Cyber-Physical Power Systems Based on Improved Adaptive Differential Evolution. *Electric Power Systems Research*, **229**, Article 110112. <https://doi.org/10.1016/j.epsr.2024.110112>
- [5] Kou, J., Jia, P., Liu, J., Dai, J. and Luo, H. (2023) Identify Influential Nodes in Social Networks with Graph Multi-Head Attention Regression Model. *Neurocomputing*, **530**, 23-36. <https://doi.org/10.1016/j.neucom.2023.01.078>
- [6] Jain, S. and Sinha, A. (2023) TriBeC: Identifying Influential Users on Social Networks with Upstream and Downstream Network Centrality. *International Journal of General Systems*, **52**, 275-296. <https://doi.org/10.1080/03081079.2023.2194642>
- [7] Yao, X., Gu, Y., Gu, C. and Huang, H. (2022) Fast Controlling of Rumors with Limited Cost in Social Networks. *Computer Communications*, **182**, 41-51. <https://doi.org/10.1016/j.comcom.2021.10.041>
- [8] Chen, D., Lü, L., Shang, M., Zhang, Y. and Zhou, T. (2012) Identifying Influential Nodes in Complex Networks. *Physica A: Statistical Mechanics and Its Applications*, **391**, 1777-1787. <https://doi.org/10.1016/j.physa.2011.09.017>
- [9] Newman, M.E.J. (2005) A Measure of Betweenness Centrality Based on Random Walks. *Social Networks*, **27**, 39-54.

- <https://doi.org/10.1016/j.socnet.2004.11.009>
- [10] Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E., *et al.* (2010) Identification of Influential Spreaders in Complex Networks. *Nature Physics*, **6**, 888-893. <https://doi.org/10.1038/nphys1746>
- [11] Meng, L., Xu, G., Yang, P. and Tu, D. (2022) A Novel Potential Edge Weight Method for Identifying Influential Nodes in Complex Networks Based on Neighborhood and Position. *Journal of Computational Science*, **60**, Article 101591. <https://doi.org/10.1016/j.jocs.2022.101591>
- [12] Ma, L., Ma, C., Zhang, H. and Wang, B. (2016) Identifying Influential Spreaders in Complex Networks Based on Gravity Formula. *Physica A: Statistical Mechanics and Its Applications*, **451**, 205-212. <https://doi.org/10.1016/j.physa.2015.12.162>
- [13] Li, Z., Ren, T., Ma, X., Liu, S., Zhang, Y. and Zhou, T. (2019) Identifying Influential Spreaders by Gravity Model. *Scientific Reports*, **9**, Article No. 8387. <https://doi.org/10.1038/s41598-019-44930-9>
- [14] Zhang, Q., Shuai, B. and Lü, M. (2022) A Novel Method to Identify Influential Nodes in Complex Networks Based on Gravity Centrality. *Information Sciences*, **618**, 98-117. <https://doi.org/10.1016/j.ins.2022.10.070>
- [15] Lü, L., Chen, D., Ren, X., Zhang, Q., Zhang, Y. and Zhou, T. (2016) Vital Nodes Identification in Complex Networks. *Physics Reports*, **650**, 1-63. <https://doi.org/10.1016/j.physrep.2016.06.007>
- [16] Lü, L., Zhou, T., Zhang, Q. and Stanley, H.E. (2016) The H-Index of a Network Node and Its Relation to Degree and Coreness. *Nature Communications*, **7**, Article No. 10168. <https://doi.org/10.1038/ncomms10168>
- [17] Zhao, S. and Sun, S. (2023) Identification of Node Centrality Based on Laplacian Energy of Networks. *Physica A: Statistical Mechanics and Its Applications*, **609**, Article 128353. <https://doi.org/10.1016/j.physa.2022.128353>
- [18] Xi, Y. and Cui, X. (2023) Identifying Influential Nodes in Complex Networks Based on Information Entropy and Relationship Strength. *Entropy*, **25**, Article 754. <https://doi.org/10.3390/e25050754>
- [19] Tong, T., Dong, Q., Sun, J. and Jiang, Y. (2023) Vital Spreaders Identification Synthesizing Cross Entropy and Information Entropy with Kshell Method. *Expert Systems with Applications*, **224**, Article 119928. <https://doi.org/10.1016/j.eswa.2023.119928>
- [20] 吴亚丽,任远光,董昂,等. 基于邻域 K-shell 分布的关键节点识别方法[J]. 计算机工程与应用, 2024, 60(2): 87-95.
- [21] Newman, M.E.J. (2006) Finding Community Structure in Networks Using the Eigenvectors of Matrices. *Physical Review E*, **74**, Article 036104. <https://doi.org/10.1103/physreve.74.036104>
- [22] Rossi, R. and Ahmed, N. (2015) The Network Data Repository with Interactive Graph Analytics and Visualization. *Proceedings of the AAAI Conference on Artificial Intelligence*, **29**, 4292-4293. <https://doi.org/10.1609/aaai.v29i1.9277>
- [23] Kunegis, J. (2013) KONECT: The Koblenz Network Collection. *Proceedings of the 22nd International Conference on World Wide Web*, Rio de Janeiro, 13-17 May 2013, 1343-1350. <https://doi.org/10.1145/2487788.2488173>
- [24] Moody, J. (2001) Peer Influence Groups: Identifying Dense Clusters in Large Networks. *Social Networks*, **23**, 261-283. [https://doi.org/10.1016/s0378-8733\(01\)00042-9](https://doi.org/10.1016/s0378-8733(01)00042-9)
- [25] Zheng, J. and Liu, J. (2023) A New Scheme for Identifying Important Nodes in Complex Networks Based on Generalized Degree. *Journal of Computational Science*, **67**, Article 101964. <https://doi.org/10.1016/j.jocs.2023.101964>
- [26] Liu, S. and Gao, H. (2022) The Self-Information Weighting-Based Node Importance Ranking Method for Graph Data. *Entropy*, **24**, Article 1471. <https://doi.org/10.3390/e24101471>
- [27] Fan, W., He, Y., Han, X. and Feng, Y. (2021) A New Model to Identify Node Importance in Complex Networks Based on DEMATEL Method. *Scientific Reports*, **11**, Article No. 22829. <https://doi.org/10.1038/s41598-021-02306-y>