Advances in Applied Mathematics 应用数学进展, 2022, 11(5), 2559-2571
Published Online May 2022 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2022.115271

Hans汉斯

河北省南宫市COVID-19网络重构及拓扑特征分析

董改琴，裴　鑫，李明涛*

太原理工大学数学学院，山西　太原

收稿日期：2022年4月16日；录用日期：2022年5月11日；发布日期：2022年5月19日

摘　要

2019年新型冠状病毒肺炎(COVID-19)大流行是一场严重的全球公共卫生突发事件。隔离病例、切断传播途径已经成为目前控制疫情的主要有效措施。由于溯源工作的复杂性，病例溯源信息统计数据不完整。在本研究中，我们提取了卫生健康委员会报告的确诊病例的行程轨迹相关数据，并构建了初始的疾病传播网络，基于链路预测算法构建最可能的疾病传播网络。在重构前后网络的基础上，比较重构前后的不同拓扑特征，获得重要病例和感染途径。结果表明，重构后的网络具有较高的聚类系数，较短的平均路径长度，即大多数病例无法直接连通，而是通过少量病例进行接触，这是疾病在较短时间内迅速传播的原因。此外，通过对疾病传播网络中心性指标的计算和分析，我们发现青亭路和天地名城小区的2例病例是网络传播的重要病例。我们的研究结果表明，尽快隔离重要病例，切断重要传播路径，将能够为控制新冠肺炎疫情作出重大贡献。

关键词

新型冠状病毒肺炎，疾病传播网络，链路预测，网络重构，拓扑特征

COVID-19 Network Reconstruction and Topology Characteristic Analysis in Nangong City, Hebei Province

Gaiqin Dong, Xin Pei, Mingtao Li*

College of Mathematics, Taiyuan University of Technology, Taiyuan Shanxi

Received: Apr. 16th, 2022; accepted: May 11th, 2022; published: May 19th, 2022

*通讯作者。

文章引用: 董改琴, 裴鑫, 李明涛. 河北省南宫市 COVID-19 网络重构及拓扑特征分析[J]. 应用数学进展, 2022, 11(5): 2559-2571. DOI: 10.12677/aam.2022.115271

Abstract

Coronavirus Disease 2019 (COVID-19) pandemic is a grave global public health emergency. Isolating cases and cutting off the route of transmission have become the main and effective measures to control the epidemic thus far. Owing to the complexity of traceability work, the statistical data of case tracing information are incomplete. In this study, we extract data associated with the itinerary trajectories of confirmed cases reported by Health Commission and then construct the initial disease transmission network. We establish the most likely disease transmission network based upon link prediction algorithms. On the basis of the network before and after reconstruction, we compare the different topological characteristics of the network to obtain important cases and infection routes. The results reveal that, the reconstructed network has higher clustering coefficient as well as shorter average path length, i.e. most cases cannot be connected but are contacted through a small amount of cases, signifying that the reason for the rapid spread of the disease in a short period of time. In particular, by calculating and analyzing the centrality index of the disease transmission network, we find that two cases live in Qingting road and Tiandimingcheng community respectively were important cases of network transmission. Our results suggest that isolating important cases and cutting off important connections as soon as possible, which will be able to significantly contribute to the control of COVID-19.

Keywords

COVID-19, Disease Transmission Network, Link Prediction, Network Reconstruction, Topological Characteristic

1. 引言

新型冠状病毒肺炎是一种严重的急性感染性肺炎，简称为新冠肺炎，可以在人与人之间进行传播[1] [2]。截至 2020 年 4 月，中国的疫情基本得到控制[3]。自 2020 年 4 月以来，中国一些地区多次出现散发病例引起的聚集性疫情。例如，2021 年 1 月 2 日，河北省石家庄市卫生健康委员会通报了新冠确诊病例(http://wsjkw.hebei.gov.cn)。第二天，邻近的南宫市也确诊了一例新冠阳性患者，随后确诊病例逐渐增加，直到一月底疫情才得以控制。而由于疫情溯源工作复杂，导致统计到的数据不完整。因此，寻找完整的疾病传播网络，是有待解决的问题，链路预测为解决此问题提供了新的角度。

链路预测是通过计算网络中两个没有接触的病例之间产生链接的可能性，来补全疾病传播网络。近年来，复杂网络的链路预测在传染病研究中得到了广泛的应用。例如：吕琳媛等在文献[4]中对链路预测的精确度指标和链路预测算法等进行了综述；樊洁茹等[5]利用随机分块模型的链路预测算法来预测网络中丢失的边和错误的边，得到更完整的活羊调运网络，为控制该疾病的传播和制定有效的防控措施提供了理论依据；Kaya 等[6]提出一种基于年龄序列的链路预测算法，表明了病人的年龄与疾病的传播有一定的相关性；McCoy 等[7]在现有的新冠肺炎的数据集上，利用链路预测算法有效地提取出与新型冠状病毒高度相关的药物；Folino 等[8]运用一种基于链路预测的方法，考虑网络节点之间的结构相似性来识别患者未来可能会患的疾病。总之，链路预测包含对未知连边的预测，也包含对未来连边的预测，在网络缺

失边的预测上有十分重要的意义。

网络的拓扑特征指标能够通过收集到的数据集来量化出网络结构信息，对于这方面的应用也有很多。例如：2016 年 Relun 等[9]在国家和社区层面计算了生猪交易网络的度、密度、聚类系数和平均路径长度，为预防和控制生猪的未来疾病的入侵提供了很好的方法；2017 年 Lichoti 等人[10]通过计算度、模块化和聚类系数来描述关键的网络属性，表明整个生猪交易网络是比较稀疏的；2021 年 Jing 等人[11]通过计算真实网络和重构网络的平均度、网络密度和聚类系数等拓扑特征，得出在重构后的网络指标表现良好；以及 2021 年 Kuang 等人[12]在研究基本拓扑特征后，又通过计算各种中心性指标，来找出重要节点和连边。简言之，网络的拓扑特征对研究网络的性质有很重要的意义。

通过链路预测来补全网络对于研究新冠肺炎病例的传播途径有重大意义。本文首先运用基于局部信息的相似性指标进行链路预测，选择出 AUC (Area under the receiver operating characteristic curve)值较高的资源分配指标(RA 指标)来重构网络，然后通过对拓扑指标进行分析来研究重构前后网络的特征，给出有效的关键病例和路径，为制定快速的防控措施提供了理论依据。

2. 初始网络构建

2.1. 数据来源

自 2019 年 12 月以来，新型肺炎疫情在中国蔓延。例如，河北省石家庄市于 2021 年 1 月 2 日发现确诊病例，河北省多个城区也相继发现病例。此后，南宫市 3 日也出现了新冠肺炎确诊病例，直到 27 日，南宫市的新冠肺炎疫情才得到控制。本文使用的数据为中国南宫市(2021 年 1 月 3 日至 1 月 27 日)确诊病例的信息。河北省卫生健康委报告确诊病例 68 例。有关这些病例的数据可在官方每日报告(http://wsjkw.hebei.gov.cn)中获得，包括年龄、性别、与其他已知病例的关系、确诊病例的诊断日期、密切接触者、地理位置和旅行轨迹。由于大规模追溯性调查难以在短时间内进行，导致统计数据不完整。

2.2. 数据预处理

将通报到的信息和数据进行整理，清理了主要信息中不可用的信息。例如，新冠病毒的潜伏期为 3~7 天，最长不超过 14 天，病毒在潜伏期内也是具有传染性的，于是只保留病例被确诊前的 1~14 天的行程轨迹。

数据的预处理即求出所需地理位置的经纬度，进行距离的计算。先从百度地图数据库提取出所有所需地点的名称和位置信息，后使用以下方程(1)计算位置 A 和位置 B 之间的距离 AB，$Longitude.A$ 和 $Longitude.B$ 是位置 A 和 B 的经度；$Latitude.A$ 和 $Latitude.B$ 是位置 A 和 B 的纬度，R 是赤道半径[13]。

$$\begin{aligned}\Gamma &= \sin\left(Latitude.A\right)*\sin\left(Latitude.B\right)*\cos\left(Longitude.A-Longitude.B\right)\\ &\quad+\cos\left(Latitude.A\right)*\cos\left(Latitude.B\right),\\ AB &= R*\arccos\left(\Gamma\right)*\pi/180\end{aligned}\tag{1}$$

2.3. 初始疾病传播网络构建

将疾病传播网络抽象为无向网络，该网络共有 68 个病例，以官方通报数据的时间先后为序，对南宫市这波疫情开始至结束(即 2021 年 1 月 3 日至 1 月 27 日)的所有病例进行编号。在疾病传播网络中病例看作网络的节点，病例间的接触看作网络的边，并做出如下假设：

1) 根据官方通报的接触信息，有具体的直接接触信息的，则两节点之间存在一条连边；
2) 病例间有共同居住或 14 天内有来往的亲密接触者，节点之间存在一条连边；
3) 在 14 天内两病例去过同一个地方，如酒店、商场和公司等，将视为有连边；

4) 根据距离，位于南屯村的 56 号节点未给出具体的出行路线与密切接触者，于是连边以两地之间的距离为依据进行连接，规定距离在 3.5 公里之内的两例病例进行连接(3.5 公里是该病例所在位置与其他病例所在位置的最小值) [14]。

根据上述假设，将疾病传播网络的 246 条连边记录在邻接表中，然后运用 R 语言的 igraph 包的可视化图形工具绘制南宫市的疾病传播网络，如图 1 所示。

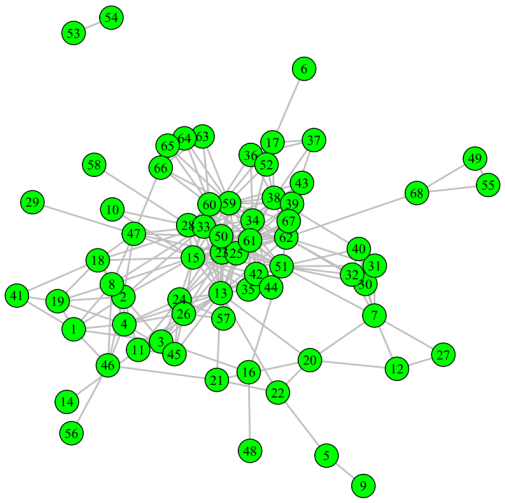


Figure 1. The disease transmission network graph of Nangong City
图 1. 南宫市疾病传播网络

3. 链路预测

3.1. 数据集的划分和评价指标

本文研究病例接触的无向无加权网络，$G(V,E)$为疾病传播网络，V 表示病例节点集合，E 表示链接集合，用U表示包含所有$U=|V|(|V|-1)/2$个可能链接的集合，用$|V|$表示集合 V 中元素个数。假设给定一种链路预测算法并得到未连接的链接的相似值，将这些相似值进行从大到小排序，排在越前面表示该连边在网络中出现的概率越大。为了测试算法是否准确，将链接 E 随机分为测试集E^T和训练集E^P，满足$E=E^T+E^P$且$E^P\cap E^T=\varnothing$，并将属于 U 但不属于 E 的链接定义为不存在的链接。

对于数据集的划分，选择随机抽样，这种算法保证了被选择到训练集和测试集中的链接是随机的，没有人为因素的干扰。通过吕琳媛等人[4]在综述中的叙述，本文采用从整体上衡量算法精确度的指标 AUC 来衡量指标的精确度。其中 n 表示 n 次循环抽取，且有 n' 次 E^P 中链接的相似值大于不存在的链接的相似值，n'' 次两者相似值相等，具体定义如下：

$$\mathrm{AUC}=\frac{n'+0.5n''}{n} \tag{2}$$

3.2. 基于局部信息的相似性指标

针对南宫市的疾病传播网络，本文基于网络拓扑结构的局部相似性指标对网络进行预测。下表 1 给出经典的十种局部相似性指标及其定义，两节点设为v_x和v_y，定义v_x邻居的集合为$\Gamma(x)$，v_y邻居的集合为$\Gamma(y)$，k_x是节点 x 的邻居的个数。

Table 1. Similarity index based on local information
表 1. 基于局部信息的相似性指标

相似性指标	定义	相似性指标	定义
CN 指标[15]	$S_{xy}=\lvert\Gamma(x)\cap\Gamma(y)\rvert$	Salton 指标[16]	$S_{xy}=\frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{\sqrt{k_x k_y}}$
Jaccard 指标[17]	$S_{xy}=\frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{\lvert\Gamma(x)\cup\Gamma(y)\rvert}$	SØrensen 指标[18]	$S_{xy}=\frac{2\times\lvert\Gamma(x)\cap\Gamma(y)\rvert}{k_x+k_y}$
HDI 指标[19]	$S_{xy}=\frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{\max\{k_x,k_y\}}$	HPI 指标[20]	$S_{xy}=\frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{\min\{k_x,k_y\}}$
LHI-I 指标[21]	$S_{xy}=\frac{\lvert\Gamma(x)\cap\Gamma(y)\rvert}{k_x k_y}$	AA 指标[22]	$S_{xy}=\sum_{z\in\Gamma(x)\cap\Gamma(y)}\frac{1}{\log k_z}$
RA 指标[23]	$S_{xy}=\sum_{z\in\Gamma(x)\cap\Gamma(y)}\frac{1}{k_z}$	PA 指标[24]	$S_{xy}=k_x k_y$

3.3. 相似性指标的算法描述

根据疾病传播网络的结构信息，本文采用基于局部信息的相似性指标进行分数值计算，并采用 AUC 作为评价指标，对上述指标的预测精确度进行对比分析，从而找出预测效果最优的指标。在实验过程中，我们以 90%的比例随机地抽取训练集网络进行预测。算法流程主要分为：邻接表和邻接矩阵的转化；相似值矩阵的求解；AUC 值的计算。详细的算法描述见表 2。

Table 2. Algorithm flow
表 2. 算法流程

算法 1：划分数据集
输入：邻接矩阵(*net*)、比率(0.9) (训练集 *train* 元素个数与测试集 *test* 个数的比值)
1. 根据数据，统计所有的边并将其放入邻接表；
2. 随机选择边；
3. 判断所选边的两端端点是否可达，如果是，就把它放到 *test* 中；
4. 重复步骤 2 和 3。
输出：*train* 和 *test*

算法 2：相似值计算
输入：*train* 和 *test*
基于相似度指标的定义(CN，Salton，Jaccard，SØrensen，HPI，HDI，LHN-I，AA，RA，PA)，利用训练集中的网络拓扑信息，测试集与不存在的边根据表 1 计算集合中所有节点构成的连通边的相似度值矩阵。
输出：*sim*

算法 3：AUC 值计算
输入：*train*、*test* 和 *sim*
1. 只保留了测试集和不存在边集之间的相似值；
2. 取两个集合(*test* 和 *non*)的上三角矩阵及相应的相似值得分；
3. 通过相应的相似值得分的比较，得到 n' 和 n''；
4. 通过公式(2)计算得到 AUC 值。
输出：AUC

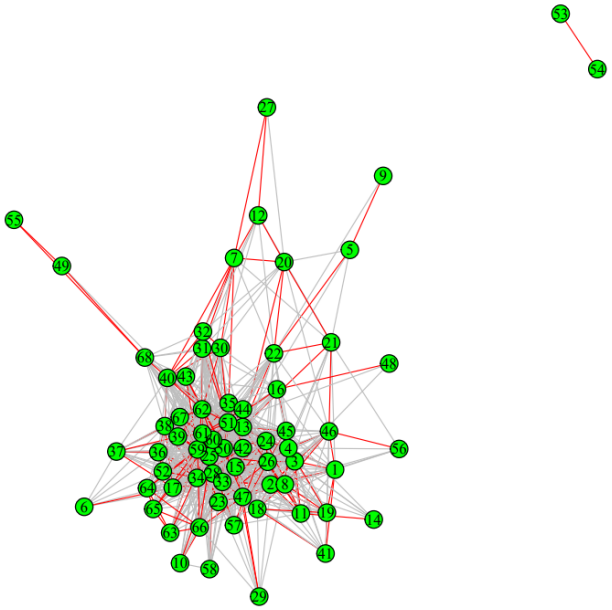


Figure 2. 2021 COVID-19 transmission network reconstruction graph
图 2. 2021 年新冠疾病传播重构网络

3.4. 南宫市新冠肺炎疾病传播网络重构

疾病传播网络的网络重构[25]表示针对现有的疾病传播网络中连边不全问题，运用链路预测的方法，通过对网络的拓扑结构以及节点属性的分析来挖掘网络中隐藏的关系。即对未知节点对之间预测的分数值大于 0 的边的连接以及已知链接中可能出现的错误链接的修正，得到更加真实，更能反映病例之间相互影响关系的网络结构。具体做法如下：

通过上一节的算法流程，计算出十种基于局部信息的相似性指标的 AUC 值，见表 3，RA 指标的预测精确度明显高于其他九种指标。于是，运用 MATLAB 软件，采用计算得到的最优链路预测指标 RA 对网络进行预测，得到测试集和不存在的链接的分数值。若两病例之间的分数值为 0，那么这两病例的连边概率为 0，不进行连边；若分数值大于 0，那么将两病例进行连接。依此，就得到一个新的重构完的邻接表，运用 R 语言就得到重构完的疾病传播网络，共有 68 个节点，818 条连边，如图 2 所示，图中红色的边是初始网络的边，灰色的为新添加的边。与初始网络相比，重构完的网络多了 300 多条边。由于在实际调查的过程中，家庭内部传染经常被忽略，行踪轨迹也不可能记录完全，导致缺失很多实际的病例接触。

Table 3. Comparison of link prediction AUC index
表 3. 链路预测 AUC 指标比较

指标	AUC 值	指标	AUC 值
CN	0.9380	HDI	0.9432
Salton	0.9635	LHN-1	0.9380
Jaccard	0.9634	AA	0.9659
Sørensen	0.9535	RA	**0.9748**
HPI	0.9561	PA	0.7514

4. 网络拓扑特征分析

通过链路预测算法，补充了网络中缺失的 372 条连边，构造出疾病传播的重构网络图。本节对重构前和重构后的疾病传播网络进行拓扑特征分析，选取网络密度[26]、平均路径长度、聚类系数[27]、同配系数[28]、节点度中心性[29]、节点接近中心性[30]、节点特征向量中心性[31]和边的介数中心性[32] [33]等统计特征指标研究疾病传播网络的拓扑性质。以上所说的拓扑特征指标运用 R 语言(4.0.3 版本)中的 igraph 包来拓扑分析。

4.1. 疾病传播网络的全局拓扑特征

表 4 给出了本文所研究的相关的拓扑特征的定义以及疾病传播网络重构前后的值，$|V|$表示节点数，$|E|$表示连边数，ρ 表示网络密度，$\langle d\rangle$表示网络的平均路径长度，c 表示网络的聚类系数，r 为同配系数。箭头表示重构网络与初始网络相比，各指标的变化情况，↑表示数值增加，↓表示数值降低。

图 3 给出了疾病传播的原始网络的度分布图[34]，其中横坐标 k 为度值，纵坐标 $p(k)$ 表示度为 k 的节点数比上整个网络节点数的值。图中可以看出，疾病传播网络的度分布是近似地遵循幂的形式[35]，即 $p(k)\propto k^{-\alpha}$，α 为幂指数，说明了疾病传播网络具有异质性和无标度网络特性。具体表现为对于大多数病例具有较少的接触病例，只有少数病例具有较多的病例接触。了解到原始的疾病传播网络拥有这一特性之后，在此基础上，我们详细分析和对比网络在重构前与重构后的描述性指标。

Table 4. Relevant definition and value of topological features
表 4. 拓扑特征的相关定义及值大小

拓扑特征(符号表示)	重构前的值	重构后的值	定义
节点数$\|V\|$	68	68	网络中总的病例数。
链接$\|E\|$	246	818↑	网络中病例之间的连边数。
网络密度 ρ	0.0538	0.1793↑	网络中病例之间的连接数量与所有可能的链接数量之间的比值，用于衡量网络中的病例是如何交织在一起。
平均路径长度$\langle d\rangle$	2.0831	1.6260↓	网络中全部病例对之间的平均距离。
聚类系数 c	0.5018	0.6819↑	网络中所有闭三元组的数量占所有连通三元组的总量之比，用于衡量病例在整个传播网络聚集性的一种趋势。
同配系数 r	−0.1433	−0.0280↑	相互连接的两个病例度的 Pearson 相关系数，用于衡量病例与具有相似(不相似)度的其他病例连接的趋势。

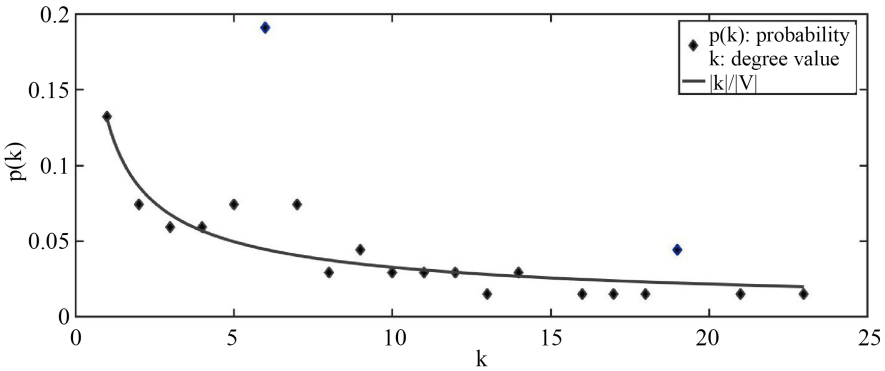


Figure 3. Degree distribution graph of the original network of disease transmission
图 3. 疾病传播原始网络的度分布图

南宫市病例传播的原始网络的网络密度为0.0538，对于一个病例数为68的网络来说，这个密度是相对偏小的。而重构网络的密度达到0.1793，比原始网络的密度增加了2倍。一方面，由数据监测发现，天一酒店、华座超市、信发商厦、信和商厦、联通公司和凤岗这几处为统计过程中出现频率最高的地方，过半数病例都在这几个地方活动，表明病例接触比较密切；另一方面，疾病传播网络的链路预测就是通过预测缺失边来补全网络，即通过补全缺失边导致网络密度增高。

从表4可知，原始网络与重构网络的同配系数均为负，即拥有较多接触者的病例趋向于与有较小接触者的病例连接。而重构前后的网络的异配性却存在较大的差异，重构后的网络的异配性是明显低于重构前的网络，且同配系数是接近于0的。由于在南宫市1月3日发现确诊病例开始，政府出台了一系列的政策和防控措施；1月7日，南宫市的天地名城小区、天一和院小区和凤岗办事处列为中风险地区；1月8日，要求全市居民居家7天；1月10日，政府发出工作地过年的通知。这一系列的防控措施使得疾病传播网络的疫情传播从1月中旬开始，就有了明显的确诊病例的下降，即网络的传播速度很小。这也证实了该网络就是非相关的网络，即同配系数趋于0。

对比重构前后的疾病传播网络，重构后的网络具有更高的聚类系数和更短的平均路径长度，这表示大多数病例不能直接相互连接，而是通过少量链接到达，这使得新冠肺炎疫情的前期，疾病能够在南宫市快速传播，这种拓扑性质也使得人群的持续感染，符合实际。在实际采取措施时，只能通过对接触者进行核酸检测进行快速甄别患者以达到控制传播的目的。

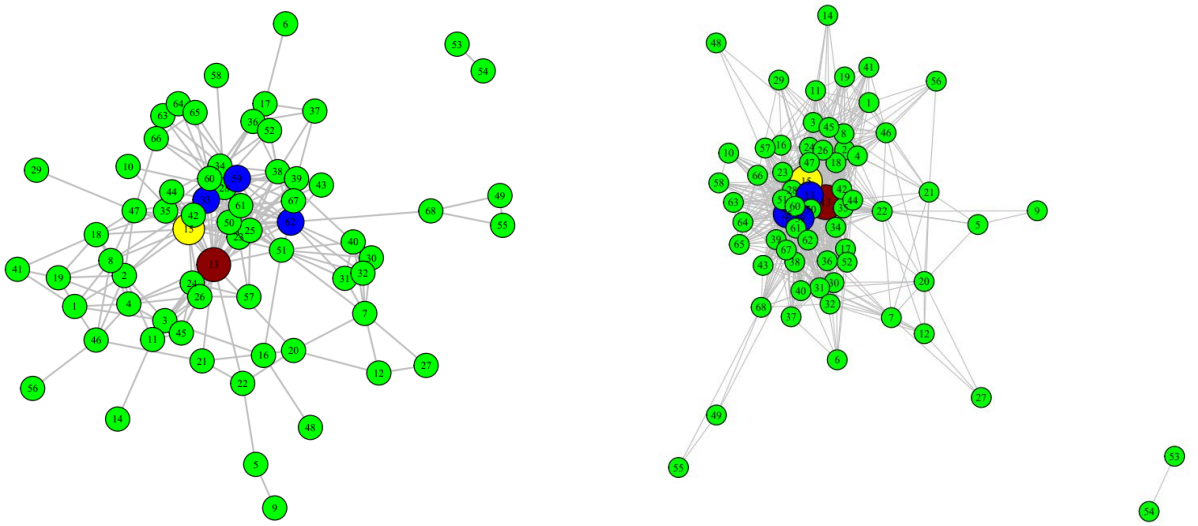
4.2. 疾病传播网络的重要病例

疾病传播网络中，寻找关键病例对于控制疫情传播是非常重要的。这类型的关键节点的分析可以有利于快速寻找病例的聚集地，为进一步控制疫情的传播节省了大量的时间和精力。测量节点在网络中的重要程度可以利用中心性指标进行分析。一般而言，度中心性分析是最简单和最直接的手段，而对于有相同数量接触者的病例来说，他们对于整个疾病传播网络的影响力与节点在网络中所处的位置有很大的关系。本文利用三种不同的中心性指标来寻找网络的关键节点，即度中心性指标(病例接触的越多，那么病例越重要)；接近中心性(病例到网络中其他病例的距离的平均值越小，那么中心度越高)和特征向量中心性(病例的邻居节点越重要，则病例越重要)。

表5给出了重构前和重构后网络的三种中心性指标值(从大到小排列，括号内为该节点对应的中心性值)。由表可得，重构前后，节点 v_{13} 的度中心性和接近中心性指标的值都是最高的，而特征向量中心性指标中最高的节点是 v_{15}，v_{13} 仅次于 v_{15}。通过分析通报的数据，发现排在前十的病例都曾去过信发商厦和信和商厦，且这两个商厦之间仅有一条街道(胜利街)作为分界线，周围有多个小区，医院等人群聚集地，这些都导致人员围绕这两个商厦大量流动；此外，1月10号确诊的 v_{13} 病例，是轨迹描述中最早去信和和信发商厦的确诊病例，其次是 v_{15} 病例。由上表可知，v_{13} 和 v_{15} 这两例病例的三类中心性值都是最高的，表明：这两例病例是此次局部疫情的重要性节点；信和商厦和信发商厦是此次南宫市疫情的部分聚集地。于是在实际的措施采取时，可以针对这两地相关人员进行核酸检测来尽快缩小勘测范围；也可以看出，重构前后网络的重要性节点是没有发生大幅度改变的。表明之前围绕初始网络采取的一系列的隔离措施是有效的，这也是此次疫情在不到一个月的时间迅速控制的原因。从图4~6分别给出三种指标的疾病传播的无向网络图，左边部分为原始网络的中心性指标图，右边为重构网络的中心性指标图，图中最大的红色节点即为中心性指标最高的病例，依次按大小为排在第2位(黄色)和在第3~5 (蓝色)的病例。从图中也可以很容易观察到疾病传播网络中找到的重要性节点都在图中相对中心的位置，表明这些节点与其他节点的联系相对紧密，再一次证实这两处地点(信和商厦、信发商厦)为疫情的聚集地。

Table 5. Centrality index
表 5. 中心性指标

网络重构前			网络重构后		
度中心性	接近中心性	特征向量中心性	度中心性	接近中心性	特征向量中心性
$v_{13}(23)$	$v_{13}(0.0040)$	$v_{15}(1.0000)$	$v_{13}(52)$	$v_{13}(0.0047)$	$v_{15}(1.0000)$
$v_{15}(21)$	$v_{15}(0.0039)$	$v_{13}(0.9903)$	$v_{15}(51)$	$v_{15}(0.0046)$	$v_{13}(0.9961)$
$v_{33}(19)$	$v_{33}(0.0039)$	$v_{61}(0.9881)$	$v_{33}(49)$	$v_{33}(0.0046)$	$v_{25}(0.9650)$
$v_{59}(19)$	$v_{25}(0.0038)$	$v_{25}(0.9584)$	$v_{25}(47)$	$v_{25}(0.0045)$	$v_{33}(0.9526)$
$v_{62}(19)$	$v_{51}(0.0038)$	$v_{59}(0.9529)$	$v_{59}(46)$	$v_{50}(0.0045)$	$v_{59}(0.9519)$
$v_{61}(18)$	$v_{59}(0.0038)$	$v_{62}(0.9450)$	$v_{62}(46)$	$v_{51}(0.0045)$	$v_{50}(0.9456)$
$v_{25}(17)$	$v_{61}(0.0038)$	$v_{33}(0.9388)$	$v_{50}(45)$	$v_{59}(0.0045)$	$v_{60}(0.9418)$
$v_{51}(16)$	$v_{62}(0.0038)$	$v_{50}(0.8541)$	$v_{60}(45)$	$v_{60}(0.0045)$	$v_{62}(0.9412)$
$v_{50}(14)$	$v_{50}(0.0037)$	$v_{51}(0.8050)$	$v_{61}(45)$	$v_{61}(0.0045)$	$v_{61}(0.9373)$
$v_{60}(14)$	$v_{60}(0.0037)$	$v_{60}(0.7903)$	$v_{51}(44)$	$v_{62}(0.0045)$	$v_{28}(0.9014)$

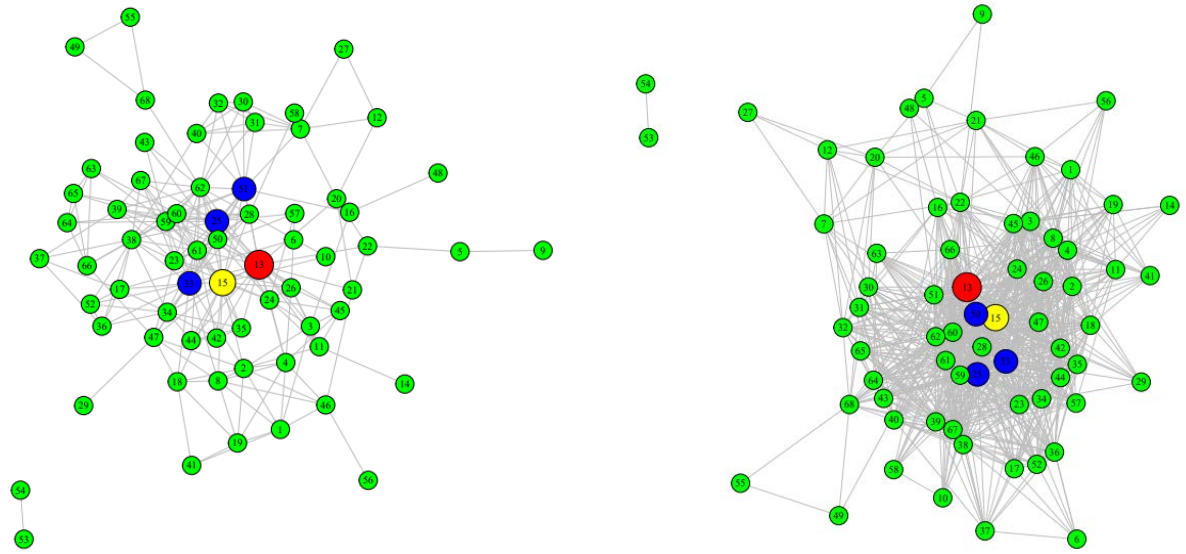


(a) (b)

Figure 4. Degree centrality index. (a) Degree centrality index before network reconstruction, (b) Degree centrality index after network reconstruction
图 4. 度中心性指标分布图。(a) 网络重构前的度中心性，(b) 网络重构后的度中心性

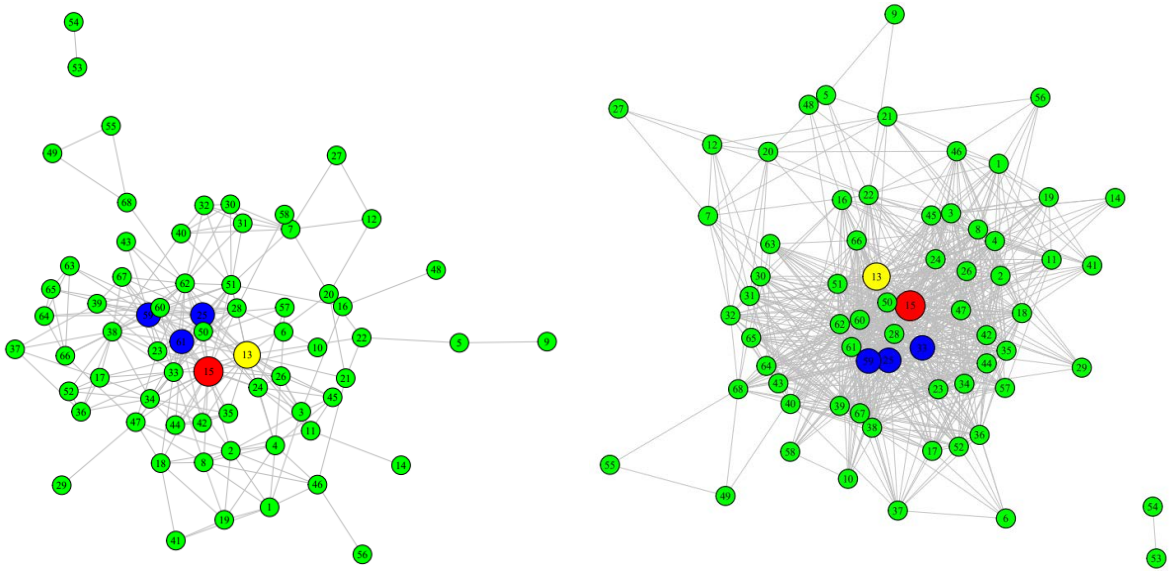
4.3. 疾病传播网络的重要连边

此次南宫市疫情有明显的聚集性特点，存在病例之间的单位传播、家庭聚集传播、医院传染和社区传染等。以家庭聚集为例，家庭 A 和家庭 B 都患有新冠肺炎，那么连接 A 和 B 的那条链接在网络中的“桥梁”作用是至关重要的，否则，网络就是由一个个孤立的“块”组成。因此，对于网络中的边也与节点有类似的重要意义。下面用边的介数中心性来衡量网络中边的重要程度，其具体含义为最短路径中经过这条路径的数目占最短路径的总数目的比值。



(a) (b)

Figure 5. Closeness centrality index. (a) Closeness centrality index before network reconstruction, (b) Closeness centrality index after network reconstruction
图 5. 接近中心性指标分布图。(a) 重构前接近中心性指标，(b) 重构后接近中心性指标



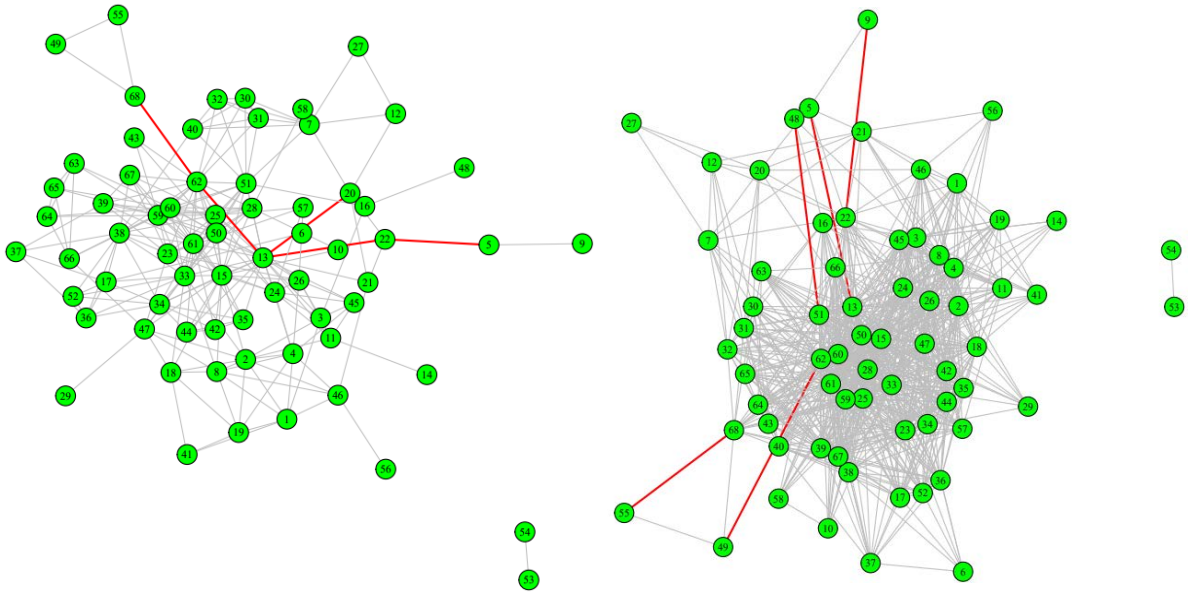
(a) (b)

Figure 6. Eigenvector centrality index. (a) Eigenvector centrality index before network reconstruction, (b) Eigenvector centrality index after network reconstruction
图 6. 特征向量中心性指标分布图。(a) 重构前特征向量中心性指标，(b) 重构后特征向量中心性指标

通过计算，重构前的网络边的介数中心性值从大到小排列，排在前面的五条边依次是$\{62\to 68,13\to 22,5\to 22,13\to 20,13\to 62\}$，而重构后的网络根据边的介数中心性值进行排序，依次是$\{48\to 51,5\to 13,49\to 62,9\to 22,55\to 68\}$。结合实际的连边数据，得出连边的两端节点(病例)分为两种类型：中心性指标较高的节点(重要节点)，如v_{13}，v_{62}等；中心性指标较低的病例(边缘节点)，如v_{49}，v_{68}等，这些病例分别位于石家庄村、西乞家庄村、青亭路、交警家属等，这些病例所处位置都相对比较偏

远。这两类病例的连边都是疾病传播网络的边缘节点与中心性节点的重要“桥梁”，若是剔除这些连边，会出现多个孤立点，疾病传播网络的连通性将会有较大的变化。表明通过介数中心性指标找到的重要连边是合理的。图 7 给出了重构前后边的 5 条具有高的介数中心性值的边的无向网络图，图中红色标记的边为值最大的五条边。从图中，很容易看出，这些连边都是位于网络中心的节点和边缘节点的连边，证实了我们所找到的连边的重要性。

通过对节点和连边的中心性指标的分析和比对，较容易找出疾病传播网络的关键病例和关键链接。在实际的采取措施中，就针对这些病例以及聚集地给予相应的防控措施，对于关键连边，尽早切断一切传播路径，达到进一步快速控制疫情继续传播的目的。



(a) (b)

Figure 7. Edge betweenness centrality index. (a) Edge betweenness centrality index before network reconstruction, (b) Edge betweenness centrality index after network reconstruction

图 7. 边的介数中心性指标分布图。(a) 重构前边的介数中心性指标，(b) 重构后边的介数中心性指标

5. 总结

2021 年初，石家庄市发现新冠肺炎确诊病例，导致邻近的南宫市连续出现确诊病例。本研究以南宫市为例，分析疾病传播网络。首先，利用提取的数据构建疾病传播网络；其次，基于 COVID-19 病例的联系数据，通过 AUC 值对比，选取出资源分配指标(RA 指标)进行链路预测，最终构建出疾病传播网络；最后，对重构前和重构后的网络进行拓扑指标的对比和分析，得出重构后网络较初始网络有更高的网络密度，更高的聚类系数以及更短的平均路径，表明大部分病例不能直接连接，而是通过少量边到达，这也是使得疾病在短时间内迅速扩散的原因。在此基础上，本文深入研究重构前后的重要节点和重要连边的变化，分别对节点的度中心性、紧密中心性和特征向量中心性进行计算，得出重构前后网络的重要性节点未发生变化，即第 13 号病例和第 15 号病例，根据接触信息现实，这两例病例是最早出现在聚集地的病例，证实所求出的重要性节点的正确性。对于重要连边，选择基于连通性的介数中心性指标进行计算，分别给出重构前后的 5 条重要链接，并证明若剔除这些连边，网络的连通性会发生很大改变。

这些结果为新冠肺炎的防控提供了很好的建议。比如，在新冠溯源调查时，通过快速控制重要节点，

即对这些病例以及密切接触者采取隔离措施，以及通过快速切断这些有较高连通性的连边来使得疫情迅速控制在一个可控的范围内。综上，疾病传播的重构网络不仅适用于南宫市新冠疫情，也为研究其它地方的新冠疫情甚至其他疫病提供了一个很好的理论依据。

致　谢

本文作者衷心感谢审稿人的意见；同时感谢其余两位作者对我的培养。

基金项目

本文受到国家自然科学基金(批准号：11801398；12101443)和山西省应用基础研究面上青年项目(批准号：201801D221024；20210302124260)的资助。

参考文献

[1] Kraemer, M.U.G., Yang, C.H., Gutierrez, B., *et al*. (2020) The Effect of Human Mobility and Control Measures on the COVID-19 Epidemic in China. *Science*, **368**, 493-497. https://doi.org/10.1126/science.abb4218

[2] Wu. J.T., Leung, K. and Leung, G.M. (2020) Nowcasting and Forecasting the Potential Domestic and in China: A Modelling Study. *The Lancet*, **395**, 689-697. https://doi.org/10.1016/S0140-6736(20)30260-9

[3] Yang, Z.F., Zeng, Z.Q., Wang, K., *et al*. (2020) Modified SEIR and AI Prediction of the Epidemics Trend of COVID-19 in China under Public Health Interventions. *Journal of Thoracic Disease*, **12**, 165. https://doi.org/10.21037/jtd.2020.02.64

[4] 吕林媛. 复杂网络链路预测[J]. 电子科技大学学报, 2010, 39(5): 651-661.

[5] 樊洁茹, 孙向东, 戴琪, 等. 基于链路预测的全国活羊调运网络重构[J]. 中北大学学报(自然科学报), 2015, 36(3): 276-281.

[6] Kaya, B. and Poyraz, M. (2015) Age-Series Based Link Prediction in Evolving Disease Networks. *Computers in Biology and Medicine*, **63**, 1-10. https://doi.org/10.1016/j.compbiomed.2015.05.003

[7] Mccoy, K., Gudapati, S., He, L., *et al*. (2021) Biomedical Text Link Prediction for Drug Discovery: A Case Study with COVID-19. *Pharmaceutics*, **13**, 794. https://doi.org/10.3390/pharmaceutics13060794

[8] Folino, F. and Pozzuti, C. (2012) Link Prediction Approaches for Disease Networks. *International Conference on Information Technology in Bio- and Medical Informatics*, Springer, Berlin, 99-108. https://doi.org/10.1007/978-3-642-32395-9_8

[9] Relun, A., Grpsnois, V. and Sanchez-vizcafno, J.M. (2016) Spatial and Functional Organization of Pig Trade in Different European Production Systems: Implications for Disease Prevention and Control. *Frontiers in Veterinary Science*, **3**, Article No. 4. https://doi.org/10.3389/fvets.2016.00004

[10] Lichoti, J.K., Davies, J., Maru, Y., *et al*. (2017) Pig Traders' Networks on the Kenya-Uganda Border Highlight Potential for Mitigation of African Swine fever Virus Transmission and Improved ASF Disease Risk Management. *Preventive Veterinary Medicine*, **140**, 87-96. https://doi.org/10.1016/j.prevetmed.2017.03.005

[11] Jing, F.S., *et al*. (2021) Reconstructing the Social Network of HIV Key Populations from Locally Observed Information. *AIDS Care*, 1-8. https://doi.org/10.1080/09540121.2021.1883514

[12] Kuang, J. and Scoglio, C. (2021) Layer Reconstruction and Missing Link Prediction of Multilayer Network with Maximum A Posteriori Estimation.

[13] Per, X., Jin, Z., *et al*. (2019) Detection of Infection Sources for Avian Influenza A (H7N9) in Live Poultry Transport Network during the Fifth Wave in China. *IEEE Access*, **7**, 155759-155778. https://doi.org/10.1109/ACCESS.2019.2949606

[14] Kurscheid, J., Stevenson, M., Durr, P.A., *et al*. (2017) Social Network Analysis of the Movement of Poultry to and from Live Bird Markets in Bali and Lombok, Indonesia. *Transboundary and Emerging Diseases*, **64**, 2023-2033. https://doi.org/10.1111/tbed.12613

[15] Yang, X.H., *et al*. (2016) Link Prediction Based on Local Community Properties. *International Journal of Modern Physics B*, **30**, Article ID: 1650222. https://doi.org/10.1142/S0217979216502222

[16] Salton, G. and Mcgill, M.J. (1983) Introduction to Modern Information Retrieval. McGraw-Hill, Auckland.

[17] Hwang, C.M., Yang, M.S. and Hung, W.L. (2018) New Similarity Measures of Intuitionistic Fuzzy Sets Based on the

Jaccard Index with Its Application to Clustering. *International Journal of Intelligent Systems*, **33**, 1672-1688.

[18] Sorensen, T. (1948) A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons. *Biologiske Skrifter*, **5**, 1-34.

[19] Zhou, T., Lu, L.Y. and Zhang, Y.C. (2009) Prediction Missing Links via Local Information. *The European Physical Journal B—Condensed Matter and Complex Systems*, **71**, 623-630. https://doi.org/10.1140/epjb/e2009-00335-8

[20] Ravasz, E., Somera, A.L., Mongru, D.A., *et al*. (2002) Hierarchical Organization of Modularity in Metabolic Network. *Science*, **297**, 1551-1555. https://doi.org/10.1126/science.1073374

[21] Wang, W.Q., Zhang, Q.M. and Zhou, T. (2012) Evaluating Network Models: A Likelihood Analysis. *EPL* (*Europhysics Letters*), **98**, 28004. https://doi.org/10.1209/0295-5075/98/28004

[22] Adamic, L.A. and Adar, E. (2003) Friends and Neighbors on the Web. *Social Networks*, **25**, 211-230. https://doi.org/10.1016/S0378-8733(03)00009-1

[23] Qu, Q., Jin, Y.D., Zhou, T., *et al*. (2007) Power-Law Strength-Degree Correlation from Resource-Allocation Dynamics on Weighted Networks. *Physical Review E*, **75**, Article ID: 021102. https://doi.org/10.1103/PhysRevE.75.021102

[24] Barabasi, A.L. and Albert, R. (1999) Emergence of Scaling in Random Networks. *Science*, **286**, 509-512. https://doi.org/10.1126/science.286.5439.509

[25] Zhang, X.Q., *et al*. (2021) Multiplex Network Reconstruction for the Coupled Spatial Diffusion of Infodemic and Pandemic of COVID-19. *International Journal of Digital Earth*, **14**, 401-423. https://doi.org/10.1080/17538947.2021.1888326

[26] Battision, F., Nicosia, V. and Latora, V. (2014) Structural Measures for Multiplex Networks. *Physical Review E*, **89**, 1-16. https://doi.org/10.1103/PhysRevE.89.032804

[27] 卢鹏丽, 董瑞, 曹乐. 聚类系数指标对复杂网络鲁棒性的影响分析[J]. 兰州理工大学学报, 2019, 45(3): 101-107. https://doi.org/10.1103/PhysRevLett.89.208701

[28] Newman, M.E.J. (2002) Assortative Mixing in Networks. *Physical Review Letters*, **89**, Article ID: 208701.

[29] 王锋, 许梁煌, 郑玉芳. 基于 m 阶邻居节点的复杂网络关键节点评估[J]. 福州大学学报(自然科学版), 2019, 47(2): 237-243.

[30] Sabidussi, G. (1966) The Centrality Index of a Graph. *Psychometrika*, **31**, 581-603. https://doi.org/10.1007/BF02289527

[31] Agryzkov, T., Tortosa, L., Vicent, J.F. and Wilson, R. (2019) A Centrality Measure for Urban Networks Based on the Eigenvector Centrality Concept. *Environment and Planning B*: *Urban Analytics and City Science*, **46**, 668-689. https://doi.org/10.1177/2399808317724444

[32] Yan, G., Zhou, T., Hu, B., *et al*. (2006) Efficient Routing on Complex Networks. *Physical Review E*, **73**, Article ID: 046108.

[33] 王军进, 刘家国, 李竺珂. 基于复杂网络的供应链企业合作关系研究[J]. 系统科学学报, 2021, 29(3): 110-130.

[34] Tlaie, A., Leyva, I., Sevilla-Escoboza, R. and Vera-Avila, V.P. (2019) Dynamical Complexity as a Proxy for the Network Degree Distribution. *Physical Review E*, **99**, Article ID: 012310. https://doi.org/10.1103/PhysRevE.99.012310

[35] 周涛, 肖伟科, 等. 网络集团度的幂律分布[J]. 复杂系统与复杂性科学, 2007, 4(2): 10-17.