

# 团伙犯罪中基于 PageRank 的嫌疑人犯罪影响力分析

张俊豪

(郑州警察学院 图像与网络侦查系, 郑州 450003)

**摘要:**如何准确评估嫌疑人的影响力是侦破团伙犯罪的关键,为此,在 PageRank 算法的基础上,结合团伙犯罪的网络结构特性,提出团伙犯罪中嫌疑人犯罪影响力评估模型.该模型首先结合已掌握的作案信息将团伙犯罪网络转化为有向加权网络,再根据中间中心度、接近中心度两大网络特性确定罪犯影响力转移矩阵,最后该模型以实际的案例为实验数据得出具体的嫌疑人犯罪影响力,并以两种成熟的模型为对比模型,实现结果显示该模型的评估结果与实际案件结果拟合度较高,且准确率也高于对比模型,验证了该模型具有较高的准确性和可行性.

**关键词:**PageRank;网络结构特性;团伙犯罪;嫌疑人影响力

**中图分类号:**TP393.08

**文献标志码:**A

**文章编号:**1000-2367(2024)02-0081-08

近年来,随着通信技术的发展,团伙犯罪的作案手段也随之发生了巨大改变,呈现出结构性、隐秘性、动态性等特点,这给公安机关的侦破工作带来了一定的挑战,特别是近些年频发的涉黑团伙犯罪、经济团伙犯罪以及网络团伙犯罪.研究表明,目前的犯罪团伙呈现出一定的社会网络特性,所以很多学者将社会网络分析方法应用于团伙犯罪案件的分析中,主要涉及两大领域.第一,团伙犯罪网络组织结构的研究,比如赵迪等<sup>[1]</sup>通过研究犯罪嫌疑人的地域特征构建共同犯罪网络模型,并对共同犯罪现象及其原因进行探究;ZHANG 等<sup>[2]</sup>充分研究了人口拐卖案件中犯罪成员的人际关系网络;同时,MCILLWAIN<sup>[3]</sup>的研究也表明犯罪人之间的人际关系是团伙犯罪的一个决定性因素;也有研究成果表明目前的地域性犯罪是团伙犯罪的一个重要性倾向,并以此分析地域性犯罪的性质和特征<sup>[4-6]</sup>;还有一些通过分析网络结构特征,比如网络密度、聚集系数、中心性分析团伙犯罪网络<sup>[7-10]</sup>.第二,团伙犯罪网络中的人员以及人员关系研究,比如孙萌<sup>[11]</sup>通过 PageRank 算法对犯罪人的影响力以及用户关系强度进行研究;FREEMAN<sup>[12]</sup>对犯罪网络中的每一个成员进行了研究,得到每一成员对整个网络的影响程度;李瑞东<sup>[13]</sup>通过数据挖掘技术设计了犯罪成员的组织关系模型.

在团伙犯罪侦查中,能够有效识别关键人物,进而掌握犯罪组织网络结构是其中最为重要的一环.目前,关于团伙犯罪网络结构的研究虽然层出不穷,但是真正能够落地实施的却不多,究其原因主要有二:一是实验数据有限,算法的有效性难以保证;二是研究方法过于局限,算法的可移植性较差.因此,本文在大数据背景下,提出团伙犯罪中基于 PageRank 算法以及网络结构特性的嫌疑人犯罪影响力评估模型(CIS-PNSC, a model for assessing the criminal influence of suspects in gang crimes based on PageRank algorithm and network structure characteristics).该模型主要用于识别团伙犯罪中的关键人物,并对锁定的犯罪嫌疑人进行

**收稿日期:**2022-11-02;**修回日期:**2022-12-16.

**基金项目:**河南省科技攻关项目(212102210531);河南省高等学校重点科研项目计划(23B630017);郑州警察学院 2023 年基科费项目(2023TJJBKY038).

**作者简介(通信作者):**张俊豪(1990—),男,河南洛阳人,郑州警察学院中队长(讲师),研究方向为智慧警务、大数据技术、网络安全,E-mail:gadx\_zjh@163.com.

**引用本文:**张俊豪.团伙犯罪中基于 PageRank 的嫌疑人犯罪影响力分析[J].河南师范大学学报(自然科学版),2024,52(2):81-88.(Zhang Junhao.Analysis of the influence of suspects in gang crimes based on PageRank[J].Journal of Henan Normal University(Natural Science Edition),2024,52(2):81-88.DOI:10.16366/j.cnki.1000-2367.2022.11.02.0004.)

重要性排序.该嫌疑人犯罪影响力评估模型首先将团伙犯罪网络转化为有向加权网络,其次根据网络结构特性确定犯罪影响力转移矩阵,最后根据 PageRank 算法计算出团伙犯罪中嫌疑人的犯罪影响力.

## 1 研究方法

### 1.1 PageRank 算法

PageRank 算法将万维网看成一幅看似杂乱而有章可循的有向图,网页对应节点,超链接对应边,然后利用随机过程的相关知识计算得出网页的权值<sup>[14]</sup>.如图 1 是由 5 个网页构成的一幅网络有向图.

PageRank 算法的核心思想可总结为如下三点.

(1)链接即肯定:在万维网中,当一个网页 A 含有很多连入链接时,表明网页 A 是被其他网页肯定的,在计算网页权值时,其他网页会将自己的权威值分配给自己所指向的网页 A.

(2)链接即贡献:在万维网中,任何一个网页会平均将权威值分配给自己所指向的网页,同理,任何一个网页的权威值由所有指向自身网页的权威值决定.

(3)权威确定权威:若万维网中某一权威网页 A 指向另一网页 B 时,网页 B 也将变成权威网页,因为在计算网页 B 的权威值时,权威网页 A 对网页 B 的贡献值会远远大于其他指向网页 B 的非权威网页的贡献值.

PageRank 算法的核心公式如式(1)所示:

$$P = (1 - d)ee^T/n + dU^T P, \quad (1)$$

式(1)中, $e$  为单位矩阵, $P$  代表网页的权威向量值, $d$  代表阻尼系数, $U$  是根据网页的链接关系得到的转移矩阵.

在本文中将以 PageRank 算法为基础,衡量团伙犯罪网络中的嫌疑人犯罪影响力.将犯罪组织网络中的某个犯罪嫌疑人  $i$  的影响力设为  $P_i$ ,那么所有犯罪嫌疑人的用户影响力则为  $\{P_i, P_j, \dots, P_n\}$ ,对应的犯罪影响力可建模成为一个向量  $P = [P_i, P_j, \dots, P_n]^T$ ,故 PageRank 算法的核心公式可表示为:

$$P = (1 - d)ee^T/n + d \begin{pmatrix} u_{ii} & u_{ij} & \cdots & u_{in} \\ u_{ji} & u_{jj} & \cdots & u_{jn} \\ \vdots & \vdots & & \vdots \\ u_{ni} & u_{nj} & \cdots & u_{nn} \end{pmatrix} \begin{pmatrix} P_i \\ P_j \\ \vdots \\ P_n \end{pmatrix}, \quad (2)$$

在式(2)中,对于每一个  $k \in [i, n]$ ,都必须满足:

$$\sum_{k=i}^n u_{kt} = 1, \quad (3)$$

在本模型中, $u_{kk}$  的值都设定为 0.

### 1.2 点度中心度

现实世界中,为了最大程度地保护自己,很多犯罪嫌疑人都是实行单向联系,所以团伙犯罪组织网络往往都是一个有向网络.在一个犯罪网络中,如果一个犯罪嫌疑人与其他犯罪嫌疑人联系较为紧密,或者说该犯罪嫌疑人能够同时指向多个其他组织成员,那么该犯罪嫌疑人就会处于权力的中心,在此可用点度中心度进行衡量,如下式所示:

$$C_i = d_i, \quad (4)$$

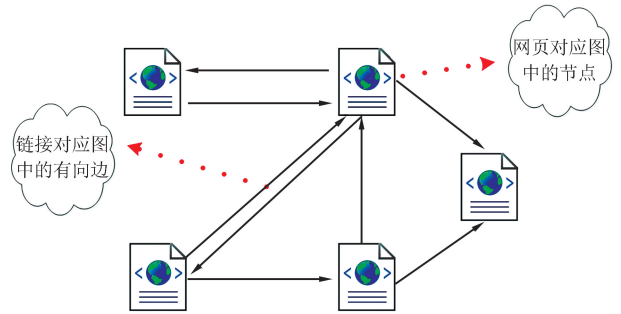


图1 由5个网页构成的网络有向图

Fig.1 A directed graph of a network consisting of 5 web pages

在式(4)中  $C_i$  表示的是犯罪嫌疑人  $i$  的点度中心度,  $d_i$  表示所有指向犯罪嫌疑人  $i$  的链接. 在构造实际的犯罪组织网络时, 要以信息的逆流向为准判别犯罪嫌疑人的关系指向, 比如犯罪嫌疑人  $B$  是通过犯罪嫌疑人  $A$  获取作案信息的, 那么在关系网络中,  $B$  应是指向  $A$  的, 即  $A$  应将自己的犯罪影响力贡献于  $B$ .

### 1.3 中间中心度

中间中心度衡量的是犯罪嫌疑人的桥梁作用. 在有向的犯罪组织网络中, 如果某犯罪嫌疑人处于其他两个人的最短联系路径上, 那么该犯罪嫌疑人对其他两人来说就比较重要. 如果该犯罪嫌疑人处于其他任意两个犯罪嫌疑人的最短路径上的频率越高, 那么该犯罪嫌疑人的“桥梁作用”就越强, 相对来说, 对组织的影响力就越大. 在此, 以式(5)表示某犯罪嫌疑人的中间中心度.

$$B_k = \sum_{\substack{i=1 \\ j \neq i}}^n S(i \rightarrow k \rightarrow j), \quad (5)$$

式(5)中,  $B_k$  表示犯罪嫌疑人  $k$  的中间中心度,  $S(i \rightarrow k \rightarrow j)$  表示犯罪嫌疑人  $i$  经过  $k$  与其他任意犯罪嫌疑人的最短路径数目.

### 1.4 接近中心度

接近中心度衡量的是犯罪嫌疑人影响力的波及程度, 在有向犯罪组织网络中, 如果某犯罪嫌疑人能够以较短的距离和其他许多人进行联系, 那么该犯罪嫌疑人也是比较重要的, 接近中心度的计算如式(6)所示:

$$F_k = \sum_{\substack{j=1 \\ j \neq k}}^n D(k \rightarrow j), \quad (6)$$

式(6)中,  $F_k$  表示犯罪嫌疑人  $k$  到该网络中所有犯罪嫌疑人的最短路径长度之和,  $D(k \rightarrow j)$  是指犯罪嫌疑人  $k$  到犯罪嫌疑人  $j$  的最短路径长度.

本模型使用狄克斯特拉算法(Dijkstra)求取最短路径, 该算法是由荷兰计算机科学家狄克斯特拉提出<sup>[15]</sup>. 采用的是贪心算法的策略, 在有向加权图中, 能够求取起始点到其他任意节点的最短距离.

## 2 基于 PageRank 算法以及网络结构特性的犯罪嫌疑人影响力评估模型

在本文中, CIS-PNSC 模型主要是对 PageRank 算法中转移矩阵  $U$  的改进, 因为该转移矩阵  $U$  是决定犯罪嫌疑人影响力的根本要素. 本模型主要通过衡量团伙犯罪网络中的三大网络特性进行犯罪影响力的分配, 即根据三大网络特性确定转移矩阵  $U$ . 另外 PageRank 算法容易发生主题漂移现象, 所以在本模型中, 加入了犯罪嫌疑人的主题相关性, 以此避免此现象.

### 2.1 有向加权网络的确定

在实际的犯罪组织网络中, 不同嫌疑人之间的关系不可能是完全相同的, 因此可以根据一些重要因素(比如个人背景、通信内容、联系方式等)计算犯罪嫌疑人之间的相似度, 然后根据相似度对有向的犯罪组织网络进行边加权, 将其变为一个有向加权网络. 不同的犯罪类型有着不同的测量维度, 比如地域性犯罪, 籍贯是一个重要属性; 黑社会组织犯罪, 从业种类是一个重要属性; 电信诈骗犯罪, 联系频率可能是一个重要属性等等. 在本模型中, 用式(7)进行衡量犯罪嫌疑人相似度的计算.

$$S(i, j) = \cos \theta = \frac{A_i * A_j}{|A_i| |A_j|}, \quad (7)$$

式(7)中, 采用余弦相似度  $S(i, j)$  评估犯罪嫌疑人之间的相似度.  $A_i$  表示的是犯罪嫌疑人  $i$  的个人信息向量(不同的案件对应着不同的测量属性). 因为在有向加权网络中, 用户之间的距离越近表示关系越亲密, 所以需要用户对用户的相似度进行转化, 转化为用户的关系强度. 在实际的加权过程中, 采用式(8)对用户关系的权值进行转化计算.

$$S' = e^{-S(i, j)}, \quad (8)$$

式(8)中,  $S'$  表示犯罪嫌疑人之间的关系强度, 同时  $S'$  是有向的, 比如在黑社会犯罪网络中, 犯罪嫌疑人  $B$  获取重要信息的主要渠道是通过犯罪嫌疑人  $A$ , 那么  $A$  对  $B$  的关系就越近, 相似度就越高, 那么  $S'$  的值就要越小, 反之  $B$  对  $A$  则不然. 有向加权网络的确定也有效地避免了 PageRank 算法的主题漂移现象.

## 2.2 转移矩阵的确定

点度中心度、中间中心度以及接近中心度是本模型确定转移矩阵的主要因素,设图2为某个团伙犯罪组织网络中的一部分,根据图2可说明犯罪嫌疑人之间的影响力分配原则。

原始的 PageRank 算法在考虑用户影响力分配时,主要的参考依据是点度中心度.在本模型中,主要根据中间中心度、接近中心度两大特性,再结合点度中心度改变  $u_{ki}$  的原始值,挖掘出团伙犯罪组织网络中真正的幕后大佬。

在图2中,犯罪嫌疑人  $D$  会将自己的犯罪影响力分给  $C$ 、 $B$  和  $F$ ,那么犯罪嫌疑人  $D$  的分配原则应是根据  $C$ 、 $B$  和  $F$  在犯罪网络中的重要性进行分配.所以在本模型中定义相对网络中心度  $T_{ki}$  表示犯罪嫌疑人  $k$  相对  $i$  的网络重要程度.在此,利用加权融合的方法确定犯罪嫌疑人  $k$  对  $i$  的相对网络中心度,  $T_{ki}$  的计算方法如式(9)所示:

$$T_{ki} = \alpha B'_{ki} + \beta F'_{ki}, \quad (9)$$

其中,  $B'_{ki}$  表示的是犯罪嫌疑人  $k$  对犯罪嫌疑人  $i$  的相对中间中心度,计算方法如式(10)所示:

$$B'_{ki} = \frac{B_k}{\sum_{j \in M(i)} B_j}. \quad (10)$$

在式(9)中,  $F'_{ki}$  表示的是犯罪嫌疑人  $k$  对犯罪嫌疑人  $i$  的相对接近中心度,计算方法如式(11)所示:

$$F'_{ki} = \frac{F_k}{\sum_{j \in M(i)} F_j}, \quad (11)$$

在式(10)、(11)中  $M(i)$  是指犯罪嫌疑人  $i$  所指向往的嫌疑人集合.在图2中,假设犯罪嫌疑人  $D$  指向了犯罪嫌疑人  $B$ 、 $C$  和  $F$ ,即  $M(D)$  是  $\{B, C, F\}$ .

$T_{ki}$  中,  $\alpha$  和  $\beta$  的确定可根据层次分析法确定<sup>[16]</sup>.此时,假设犯罪嫌疑人  $B$  在本团伙犯罪组织网络中的中间中心度为6,接近中心度为15;犯罪嫌疑人  $C$  在本团伙犯罪组织网络中的中间中心度为8,接近中心度为12;犯罪嫌疑人  $F$  在本团伙犯罪组织网络中的中间中心度为3,接近中心度为18.那么犯罪嫌疑人  $B$  相对犯罪嫌疑人  $D$  的相对中间中心度为  $B'_{BD} = 6/(6+8+3)$ ,犯罪嫌疑人  $B$  相对犯罪嫌疑人  $D$  的相对接近中心度为  $F'_{BD} = 15/(15+12+18)$ ,则犯罪嫌疑人  $B$  相对犯罪嫌疑人  $D$  的相对网络中心度为  $T_{BD} = \alpha B'_{BD} + \beta F'_{BD}$ .此时  $u_{ij}$  的计算如式(12)所示:

$$u_{ij} = \frac{T_{ji}}{\sum_{t \in M(i)} T_{ti}}. \quad (12)$$

## 2.3 犯罪影响力的计算

在图2中,假设  $B$  相对犯罪嫌疑人  $D$  的相对网络中心度为  $T_{BD} = 0.4$ ,犯罪嫌疑人  $C$  相对犯罪嫌疑人  $D$  的相对网络中心度为  $T_{CD} = 0.25$ ,犯罪嫌疑人  $F$  相对犯罪嫌疑人  $D$  的相对网络中心度为  $T_{FD} = 0.55$ ,那么犯罪嫌疑人  $D$  则将自己的影响力的  $0.40/(0.40+0.25+0.55)$  贡献给犯罪嫌疑人  $B$ ,同理,影响力的21%贡献给犯罪嫌疑人  $C$ ,影响力的46%贡献给犯罪嫌疑人  $F$ .此时  $u_{DB} = 0.33$ ,  $u_{DC} = 0.21$ ,  $u_{DF} = 0.46$ ,它们的和为1.

## 2.4 CIS-PNSC 模型构建的主要步骤

根据以上描述,CIS-PNSC 模型构建的主要步骤如下:

- (1)首先根据已获得的犯罪组织信息,结合犯罪团伙特点,形成团伙犯罪组织的网络拓扑结构;
- (2)根据已掌握的所有犯罪嫌疑人的作案信息,通过犯罪嫌疑人之间关系强度  $S'$  的计算方法确定该团伙犯罪组织网络的用户关系权值;
- (3)根据网络结构特性,即点度中心度、中间中心度以及接近中心度确定转移矩阵;

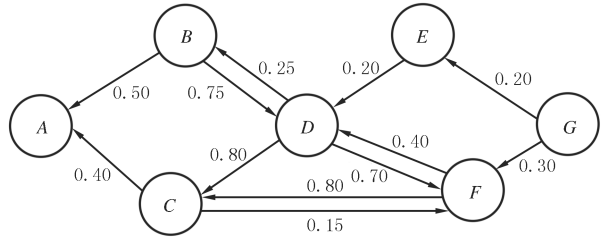


图2 犯罪组织网络

Fig.2 Criminal organization network



(4)根据 PageRank 算法计算得到犯罪团伙中嫌疑人的犯罪影响力。

### 3 实验结果及分析

#### 3.1 数据来源

本文的实验数据来源于 2019 年某市的一起特大涉黑案件,该犯罪团伙共有 253 名涉案人员,其中有 43 名主要涉案人员,共计涉案 39 件。根据用户关系权值可以得出该团伙犯罪组织网络的拓扑结构,具体包含 253 个节点,30 456 条有向边。在取参数时,考虑到主要涉案人员在犯罪团伙中的重要程度和影响力作用,认为参数  $\alpha$  大于  $\beta$ ,即该嫌疑人在犯罪网络中的相对中间中心度的权重大于相对接近中心度的权重,根据作案信息以及层次分析法,得出本模型的主要参数,如表 1 所示。

理论上, $d$  的取值在 0 至 1 之间,若  $d$  的取值趋近于 1,那么迭代的次数会成倍增长,效率极其低下,而人为控制的比重也将会下降。本文经过计算, $d$  的取值为 0.850 时,既增加了人为的控制比重,也将迭代次数控制在 140 次左右,符合 PageRank 算法的建议。

表 1 模型参数

Tab. 1 Model parameters

参数	$\alpha$	$\beta$	$d$
数值	0.628	0.372	0.850

#### 3.2 实验评价指标

##### 3.2.1 Precision@N

在信息检索或者评估分类中,倘若不考虑结果的排名顺序,召回率和准确率是最常用的评价指标,而在有序的排名结果中,用户关注的往往是那些靠前的排名结果,比如在团伙犯罪中,公安机关关注更多的往往是那些幕后黑手——实际的操盘者,此时依靠召回率与准确率评价其排名结果无法满足实际需要。这时 Precision@N( $P@N$ )就是一个很好的补充评价指标,其关注的是排名结果较为靠前的准确率。 $P@N$  的计算如式(13)所示:

$$P@N = \frac{A_N \cap B_N}{N}, \quad (13)$$

式(13)中, $A_N \cap B_N$  表示模型  $A(B)$  对犯罪嫌疑人影响力进行大小排名得到的前  $N$  名用户的交集数量,由于本文涉及的嫌疑人较多, $N$  的取值分别为 10,20,30,40,50,60,70,80。

##### 3.2.2 斯皮尔曼等级系数

斯皮尔曼等级系数是衡量不同算法相关性的一个重要参数,斯皮尔曼等级系数常用作解决分析两个变量之间相关性的问题,这是由英国统计学家斯皮尔曼根据通过积差相关的基础知识计算得出的<sup>[17]</sup>。在本文中,斯皮尔曼等级系数主要用来衡量不同模型得出的嫌疑人犯罪影响力结果的相关性。斯皮尔曼等级系数的计算方法如式(14)所示:

$$\rho = 1 - \frac{6 \sum_{i=1}^N (x_i - y_i)^2}{N(N^2 - 1)}, \quad (14)$$

在式(14)中, $\rho$  代表最终得到的变量关系值,即斯皮尔曼等级系数,在本文中,当  $\rho > 0$  时,表示两个模型所得的排名结果正相关,当  $\rho < 0$  时,表示两个模型所得的排名结果负相关,当  $|\rho| = 1$  时,表示两个模型所得的排名结果完全正相关,或者完全负相关。 $\rho$  值越大,表示两个模型所得的排名结果相关性越强。 $x_i$  和  $y_i$  分别代表不同模型求取得  $N$  个犯罪嫌疑人的影响力值。

通过斯皮尔曼等级系数可以评估本文模型与对比模型评估结果的相关性。斯皮尔曼等级系数与  $P@N$  评价指标的主要区别在于,前者从全局的角度出发评估实验结果,分析模型之间的相关性;而后者是从局部的角度出发分析模型的准确率。

#### 3.3 结果分析

本文基于 Hadoop 对实验数据进行分析处理,Hadoop 是由 Apache 基金会开发的开源、可靠稳定、可扩展、低成本分布式系统基本架构,以文献[18-19]中提到的模型为对比模型进行实验的对比分析。3 种模型计算所得的嫌疑人犯罪影响力的实验结果如表 2 所示。

根据实验结果,3种模型的相关性如表3所示。

从表3中可以看出,3个模型的相关性都呈现正相关性,而且 $\rho$ 均大于0.4,说明本模型与已有的较为成熟的模型具有一定的吻合度,所以也具有一定的可行性。其中CIS-PNSC与文献[18]的相关性最高,这是因为CIS-PNSC模型与文献[18]都是在PageRank的基础上进行改进的。

根据实际的办案结果,可得到真实的嫌疑人犯罪影响力排名,本文以实际办案得到排名为基线模型,以本文模型以及文献[18-19]为对比模型。对比模型所得结果在 $P@N$ 指标下的表现如表4所示。

根据表4可知,本模型在前80名犯罪影响力的排名准确率能够维持在70%以上,说明了本模型的准确性是达标的,对比文献[18-19],本模型的准确率是较高的,是优于对比模型的。随着 $N$ 值的增大,准确率在下降,但仍能维持在70%以上。根据计算结果可知,当 $N$ 为120时,本模型的 $P@N$ 值最小,为65%,能够达到基本要求;当 $N$ 的值大于120时, $P@N$ 值开始逐渐恢复增加。特别需要说明的是,本模型预测得到的前三名犯罪嫌疑人和实际办案结果是相吻合的,分别为16号犯罪嫌疑人(实际为该市城中村村主任)、84号犯罪嫌疑人(实际为该区刑侦大队大队长)和187号犯罪嫌疑人(实际为该市某房地产公司副董事长),对比文献[18-19]的预测结果,前3名与实际结果均不相符。

表4 以实际结果为基线模型的 $P@N$ 值测试结果

Tab. 4 Test results of  $P@N$  values based on actual results as the baseline model

$P@N$	CIS-PNSC 模型	文献[18]	to 文献[18]+%	文献[19]	to 文献[19]+%
P10	0.80	0.50	60.00%	0.70	14.29%
P20	0.75	0.50	50.00%	0.70	7.14%
P30	0.70	0.45	55.56%	0.65	7.69%
P40	0.75	0.50	50.00%	0.70	7.14%
P50	0.78	0.43	81.40%	0.62	25.81%
P60	0.72	0.43	67.44%	0.52	38.46%
P70	0.72	0.40	80.00%	0.50	44.00%
P80	0.70	0.38	84.21%	0.50	40.00%

为显示本模型对犯罪影响力评估的准确性,本文根据实际作案信息以及犯罪嫌疑人的实际重要性排名,对实际犯罪影响力进行和本模型同等数量级的转换,以此对比本文模型和文献[18-19]中的模型。由于犯罪嫌疑人较多,则在253名犯罪嫌疑人中采用分段随机抽样方法选取25名犯罪嫌疑人的影响力进行可视化,如图3所示。

根据图3可知,本文提出的模型能够较好拟合实际犯罪影响力排名结果,文献[18-19]拟合度较差。

根据 $P@N$ 和斯皮尔曼等级系数两个指标的评价结果验证了CIS-PNSC模型的可行性和准确率。CIS-PNSC模型在面对大样本数据集时,计算结果的准确率可观,且模型运行需要的时间并不长;相较于文献[18-19]提出的两种成熟模型,CIS-PNSC模型得到的犯罪嫌疑人影响力和实际结果拟合度较高,具有较好的鲁棒性,可以基本满足实际公安工作的需要。

表2 3种模型的实验结果

Tab. 2 Experimental results of three models

嫌疑人	CIS-PNSC	文献[18]	文献[19]
嫌疑人1	0.353 6	0.215 4	0.387 9
嫌疑人2	0.245 3	0.352 4	0.458 9
嫌疑人3	0.020 0	0.100 0	0.089 6
嫌疑人4	0.365 4	0.211 1	0.358 9
⋮	⋮	⋮	⋮
嫌疑人249	0.214 5	0.325 6	0.18 9
嫌疑人250	0.023 5	0.126 0	0.120 1
嫌疑人252	0.157 8	0.368 9	0.548 3
嫌疑人253	0.425 6	0.323 3	0.254 1

表3 3种模型的相关性

Tab. 3 Correlation of three models

类别	CIS-PNSC	文献[18]	文献[19]
CIS-PNSC	1.00	0.69	0.43
文献[18]	0.69	1.00	0.77
文献[19]	0.43	0.77	1.00

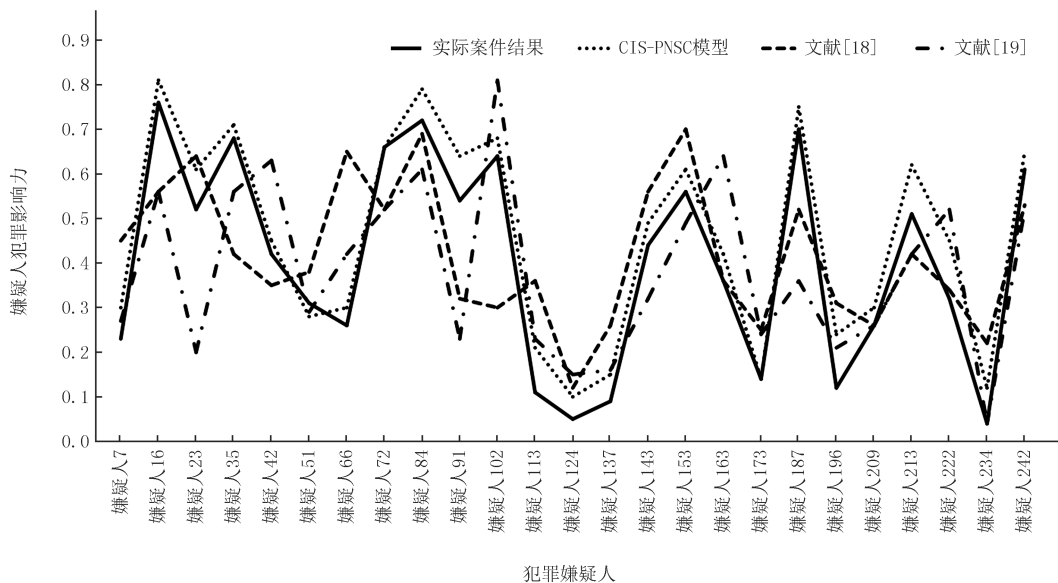


图3 犯罪影响力对比图

Fig.3 Comparison of criminal influence

## 4 结 论

本文在 PageRank 算法的基础上,结合团伙犯罪组织的网络结构特性,提出了团伙犯罪的嫌疑人犯罪影响力评估模型.在该模型中,首先通过犯罪分子之间的关系确定了犯罪分子之间的关系强度,符合现实逻辑,同时避免了 PageRank 的主题漂移现象;其次本模型能够结合网络结构特性,计算得到犯罪嫌疑人的影响力分配因子;最后本模型以某一黑社会犯罪团伙为实验对象,同时结合实际的结案情况以及两种对比模型,验证了本模型具有较高的准确率和可行性.在后续的研究过程中,仍需对部分参数进行优化,尽可能地挖掘犯罪网络的其他特性,减少对作案信息的依赖,以期能够做到提前研判.

## 参 考 文 献

- [1] 赵迪,陈鹏,江欢,等.基于犯罪人地域特征的共同犯罪网络及影响因素研究[J].地理与地理信息科学,2022,38(5):57-64.  
ZHAO D, CHEN P, JIANG H, et al. Co-offending network and influencing factors based on offenders' regional characteristics[J]. Geography and Geo-Information Science, 2022, 38(5): 57-64.
- [2] ZHANG S X, CHIN K L, MILLER J. Women's participation in Chinese transnational human smuggling: a gendered market perspective [J]. Criminology, 2007, 45(3): 699-733.
- [3] MCILLWAIN J S. Organized crime: a social network approach [J]. Crime, Law and Social Change, 1999, 32(4): 301-323.
- [4] MATTHEWS J L, MATLOCK T. Understanding the link between spatial distance and social distance [J]. Social Psychology, 2011, 42(3): 185-192.
- [5] 李国正,韩文硕,艾小青,等.社会关系网络重构与流动人口创业活动:作用机理与政策含义[J].人口与经济,2021(3):1-17.  
LI G Z, HAN W S, AI X Q, et al. Social networks restructuring and entrepreneurial activities of floating population: mechanism and policy implications [J]. Population & Economics, 2021(3): 1-17.
- [6] 黄忠良,翁文国.典型拐卖妇女犯罪团伙网络特征分析[J].清华大学学报(自然科学版),2020,60(4):328-333.  
HUANG Z L, WENG W G. Social network analysis of typical women trafficking networks [J]. Journal of Tsinghua University (Science and Technology), 2020, 60(4): 328-333.
- [7] SCHAEFER D R. Youth co-offending networks: an investigation of social and spatial effects [J]. Social Networks, 2012, 34(1): 141-149.
- [8] 王子杰.黑社会性质组织复杂犯罪网络分析[J].铁道警察学院学报,2021,31(3):50-55.  
WANG Z J. Analysis of complex criminal networks of underworld organizations [J]. Journal of Railway Police College, 2021, 31(3): 50-55.
- [9] MALM A. Networks of collaborating criminals: assessing the structural vulnerability of drug markets [J]. Journal of Research in Crime and Delinquency, 2011, 48(2): 271-297.

- [10] 马方.犯罪网络分析:社会网络分析在有组织犯罪研究中的应用[J].西南政法大学学报,2012,14(2):34-43.  
MA F.Criminal network analysis:social network analysis in the study of organized crime[J].Journal of Southwest University of Political Science and Law,2012,14(2):34-43.
- [11] 孙萌.面向社交网站的数据挖掘应用研究:用户关系分析[D].南京:南京邮电大学,2014.
- [12] FREEMAN L C.Centrality in social networks conceptual clarification[J].Social Networks,1978,1(3):215-239.
- [13] 李瑞东.基于数据挖掘技术的犯罪预警模型设计与实现[D].沈阳:沈阳建筑大学,2014.
- [14] 钟寒,张鸿洲,尹德春,等.基于 PageRank 算法的团伙犯罪中嫌疑人排名[J].广西师范大学学报(自然科学版),2019,37(3):79-86.  
ZHONG H,ZHANG H Z,YIN D C,et al.Ranking of suspects in Gang crime based on PageRank[J].Journal of Guangxi Normal University(Natural Science Edition),2019,37(3):79-86.
- [15] 祝国明.基于 Dijkstra 的多源点最短路径求解算法的设计与分析[J].电脑知识与技术,2021,17(16):177-178.  
ZHU G M.Design and analysis of the shortest path algorithm for multi-source points based on Dijkstra[J].Computer Knowledge and Technology,2021,17(16):177-178.
- [16] 杜栋,庞庆华,吴炎.现代综合评价方法与案例精选[M].2 版.北京:清华大学出版社,2008.
- [17] 张文耀.用斯皮尔曼系数衡量网络的度相关[D].合肥:中国科学技术大学,2016.
- [18] 董伟,陶金虎.融合 PageRank 与评论情感倾向的在线健康社区用户影响力研究[J].图书情报工作,2021,65(11):14-23.  
DONG W,TAO J H.Research on the user's influence in online health community based on page rank and emotional tendency[J].Library and Information Service,2021,65(11):14-23.
- [19] JU C H,GU Q Y,FANG Y,et al.Research on user influence model integrating personality traits under strong connection[J].Sustainability,2020,12(6):2217.

## Analysis of the influence of suspects in gang crimes based on PageRank

Zhang Junhao

(Department of Image and Network Investigation, Zhengzhou Police College, Zhengzhou 450003, China)

**Abstract:** How to accurately evaluate the influence of suspects is the key to detecting gang crime. Therefore, based on the PageRank algorithm and the network structure characteristics of gang crime, a model for evaluating the influence of suspects in gang crime is proposed. The model first converts the gang criminal network into a directed-weighted network based on the known crime information, and then determines the criminal influence transfer matrix according to the two network characteristics of middle centrality and proximity centrality. Finally, the model uses actual cases as experimental data to obtain specific suspects'criminal influence, and uses two mature models as comparison models. The implementation results show that the evaluation results of the model have a high degree of fitting with the actual case results, and the accuracy is also higher than the comparison models, verifying the high accuracy and feasibility of the model.

**Keywords:** PageRank; network structure characteristics; gang crime; the influence of suspects

[责任编辑 陈留院 赵晓华]