

基于社会网络的犯罪组织关系挖掘

杨莉莉, 杨永川

(中国人民公安大学研究生部, 北京 100038)

摘要: 在对犯罪组织进行图形化构建的基础上, 利用社会网络方法对犯罪组织关系进行挖掘。犯罪组织关系挖掘包含根据社会网络的中心性指标提出一种犯罪组织重点人员判定方法和挖掘犯罪组织成员间的关系。实验结果显示, 犯罪组织关系挖掘方法具有较好的性能和挖掘效果。

关键词: 社会网络; 犯罪组织; 重点人员; 组织关系

Crime Organization Relation Mining Based on Social Network

YANG Li-li, YANG Yong-chuan

(Graduate Department, Chinese People's Public Security University, Beijing 100038)

【Abstract】 Crime organization relations are mined using social network based on crime organization map. The crime organization relations mining includes two aspects, one is proposing a method to determine the important persons of crime organization using centrality index of social network, the other is mining the relation of crime organization members. Experimental results show that the methods of crime organization mining have good performance and effect.

【Key words】 social network; crime organization; important person; organization relation

1 概述

随着社会的发展, 越来越多的犯罪呈现集团化和团伙化, 有组织犯罪已成为犯罪结构的主要形式。犯罪组织涉及的犯罪成员众多, 若要在众多的犯罪成员当中辨别出犯罪成员头目, 分析犯罪成员间的关系, 对于调查人员来说并不是一件易事, 因此, 这种案件的调查应另辟新径。由于犯罪组织是社会网络的一种表现形式, 具有社会网络的基本特性, 因此可利用社会网络方法对犯罪组织关系的挖掘进行研究。

许多研究人员对不同情况的社会网络关系进行了分析。文献[1]对社会网络中最有影响的节点进行了研究, 文献[2]提出在社会网络分析中任何时间任何环境都有效的中心度计算方法, 文献[3]提出在大规模通话信息中测量中心度的方法。这些算法大都是针对大规模的网络而言, 由于公安机关所要处理的犯罪组织规模并没有上述情况那样庞大, 因此本文针对规模较小的犯罪组织, 提出了基于社会网络的犯罪组织关系挖掘方法。犯罪组织关系挖掘包括对犯罪组织重点人员的判定以及犯罪组织成员间关系的分析。

2 犯罪组织图形化构建

犯罪组织关系挖掘的基础是对犯罪组织图形化构建, 犯罪组织图形的构建实则为犯罪组织成员间通信关系的构建。犯罪组织的图形化构建以单点分析算法为基础, 在通信数据库中寻找与被调查账号通信的其他账号, 再以这些账号为出发点逐层挖掘与这些账号通信的其他账号, 并将这些账号间的通信关系以图形化的方式显示给调查人员。单点分析算法设计如下:

输入 初始账号 a , 查询开始时间 $time1$, 查询结束时间 $time2$ ($time1 < time2$), 最大返回数量限制 n

输出 与该账号直接通信记录的结果集

(1) 在通信数据库中选择符合(信息发送方账号= a)且($time1 \leq$ 记录时间 $\leq time2$)的所有节点, 将节点信息复制到通信记录结果集中;

- (2) 在通信数据库中选择符合(信息接收方账号= a)且($time1 \leq$ 记录时间 $\leq time2$)的所有节点, 将节点信息复制到通信记录结果集中;
- (3) 将结果集中与 a 通信所有节点进行分组, 取排名前 n 条分组;
- (4) 记录 n 条分组中各节点的通信关系, 若两者之间有通信, 则在两者之间建立一条边, 边的权值表示联系的频度;
- (5) 将结果集返回。

单点分析算法仅实现了与一个账号直接通信的其他账号的挖掘, 若要完成犯罪组织的构建, 还需对挖掘出的账号再进行单点分析, 挖掘与之相关联的其他账号。犯罪组织图形化构建算法流程如图1所示。

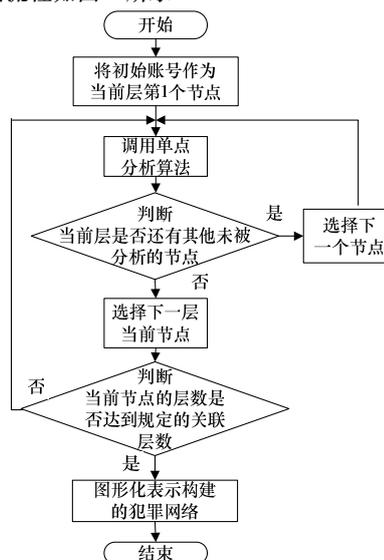


图1 犯罪组织构建算法流程

作者简介: 杨莉莉(1982—), 女, 博士研究生, 主研方向: 计算机犯罪侦查, 数字取证; 杨永川, 教授、博士生导师

收稿日期: 2009-01-12 **E-mail:** yl1830@yahoo.com.cn

3 犯罪组织关系挖掘

犯罪组织关系挖掘包含 2 个方面的内容:

- (1) 犯罪组织重点人员的判定。
- (2) 犯罪组织成员间关系的挖掘。

本文根据社会网络中心性指标提出一种犯罪组织重点人员的判定方法,此外,还对犯罪组织成员间的关系进行了分析。

3.1 犯罪组织重点人员判定

犯罪组织通常由重点人员及一般人员组成。重点人员是指具有影响力的人物,他负责决策的制定、信息的传递、成员间的联系等。一般人员主要负责犯罪的实施,这样的成员对犯罪组织的影响往往较小。在社会网络结构图中,节点的点度中心度、中间中心度以及接近中心度是量化网络中节点中心性的 3 个指标,但为了便于网络中的节点与其他网络中节点进行比较,通常采用相对点度中心度 C_{RD_i} 、相对中间中心度 C_{RB_i} 、相对接近中心度 C_{RC_i} 对节点的中心性进行测量。结合 C_{RD_i} 、 C_{RB_i} 、 C_{RC_i} 3 种中心性指标,本文提出一种犯罪组织重点人员判定算法,算法描述如下:

输入 犯罪组织成员节点表 node[i]

输出 重点人员信息表 person[i]

算法步骤:

- (1) 读取犯罪组织成员节点信息 node[i];
- (2) if node[i] 为空, then 转入步骤(5);
- (3) 计算当前节点的 C_{RD_i} 、 C_{RB_i} 、 C_{RC_i} 以及 $(C_{RD_i} \times C_{RB_i} \times C_{RC_i})$ 值,将计算后的结果分别输入到 person[i] 中的“相对点度中心度”、“相对中间中心度”、“相对接近中心度”及“综合指数”项中;
- (4) $i=i+1$, 转入步骤(2);
- (5) 将重点人员信息表按照“综合指数”由高到低进行排序;
- (6) 返回重点人员信息表 person[i]。

算法中的“综合指数”代表犯罪组织中成员重要性的 大小,在犯罪组织结构图中,由于叶子节点的 C_{RB_i} 为 0,因此节点的“综合指数”的值也为 0,图中叶子节点不可能为犯罪组织的重点人员。“综合指数”的确定不同于文献[4]中的方法,优点在于分析中可以忽略掉大量值为 0 的叶子节点,使得重点人员节点凸显出来,在一定程度上缩短了分析时间。犯罪组织重点人员判定算法给出了犯罪组织中重点人员的大致范围,但由于在判断犯罪组织重点人员过程中,调查人员并不希望重点人员的确定完全由程序自动实现,在判断的过程中还希望加入其他人为因素人工判研。因此,本文按照“综合指数”由高到低进行排序,并将重点人员信息表 person[i] 返回给调查人员,由调查人员根据案情的需求,结合表中的各种参数进行二次判研。

3.2 犯罪组织成员关系挖掘

犯罪组织成员关系是指在犯罪组织中各成员间的地位和支配关系,对犯罪组织成员关系的挖掘,是为了判断犯罪组织成员间是否存在联系以及成员间的支配关系,这对于公安机关分析案情具有至关重要的作用。对于一般规模的网络,若节点的 C_{RD_i} 、 C_{RB_i} 、 C_{RC_i} 越大,则认为该节点对于犯罪组织的影响越大,对他人的支配程度也越大,以此为基础,本文提出犯罪组织成员关系挖掘方法,方法过程如下:

在犯罪组织节点表中选取当前节点 n ,并找到 n 的相邻节点集合 N ,对于 N 中的任一节点 s , $s \in N$,都有:

(1) 相对点度中心度比较

- 1) 若 $C_{RD_n} > C_{RD_s}$, 则构建一条 n 与 s 之间的边,方向为 $n \rightarrow s$, 选择 N 中下一节点重新比较。
- 2) 若 $C_{RD_n} < C_{RD_s}$, 则构建一条 n 与 s 之间的边,方向为 $s \rightarrow n$, 选择 N 中下一节点重新比较。
- 3) 若 $C_{RD_n} = C_{RD_s}$, 则转入步骤(2)。

(2) 相对中间中心度比较

- 1) 若 $C_{RB_n} > C_{RB_s}$, 则构建一条 n 与 s 之间的边,方向为 $n \rightarrow s$, 选择 N 中下一节点转入步骤(1)。
- 2) 若 $C_{RB_n} < C_{RB_s}$, 则构建一条 n 与 s 之间的边,方向为 $s \rightarrow n$, 选择 N 中下一节点转入步骤(1)。
- 3) 若 $C_{RB_n} = C_{RB_s}$, 则转入步骤(3)。

(3) 相对接近中心度比较

- 1) 若 $C_{RC_n} > C_{RC_s}$, 则构建一条 n 与 s 之间的边,方向为 $n \rightarrow s$, 选择 N 中下一节点转入步骤(1)。
- 2) 若 $C_{RC_n} < C_{RC_s}$, 则构建一条 n 与 s 之间的边,方向为 $s \rightarrow n$, 选择 N 中下一节点转入步骤(1)。
- 3) 若 $C_{RC_n} = C_{RC_s}$, 则两者之间无边产生。

对于犯罪组织节点表中的每一个节点都要进行上述分析,并将犯罪组织成员间的关系图形化表示。在犯罪组织关系图中,边表示节点之间存在着支配关系,边的起点是终点的领导者,节点的出度越大,该节点在组织中的影响力就越大。

4 实验分析

4.1 实验数据与实验设计

本文以某一实验通信数据库中一个月时间内的通信数据为数据源进行实验,分析出与 10 个指定被调查账号的通信信息,利用犯罪组织图形化构建算法分别构建出以这 10 个账号为基础的犯罪组织。实验分为 2 部分,均在 P4 2.0 GHz, 1 GB, Windows XP Professional 的环境下进行。实验 1 对比文献[4]中的社团核心挖掘算法,分析本文重点人员挖掘算法的性能。实验 2 利用已被构建的某一个犯罪组织,验证犯罪组织成员关系挖掘的效果。

4.2 实验 1: 重点人员挖掘算法的性能

实验利用犯罪组织图形化构建算法构建了 10 个犯罪组织,以人工分析出的犯罪组织重点人员为参照,分别利用文献[4]中的社团核心挖掘算法和本文中犯罪组织重点人员判定算法对这 10 个犯罪组织的重点人员进行挖掘,分析比较各算法的准确率和执行时间。2 种算法的准确率与挖掘时间的对比分别如表 1、图 2 所示。由表 1 可以看出,2 种算法的平均准确率都在 87% 以上,都能有效地通过各自的算法挖掘出犯罪组织的重点人员,与文献[4]中的算法相比,本文算法在犯罪组织重点人员挖掘方面具有更高的准确性。

表 1 2 种算法的准确率对比表

犯罪组织编号	文献[4]方法准确率/(%)	本文方法准确率/(%)
1	87.23	89.12
2	84.69	86.13
3	89.35	88.43
4	86.24	87.19
5	90.31	91.87
6	91.51	92.24
7	88.45	88.06
8	87.06	88.29
9	89.17	89.76
10	83.26	85.53

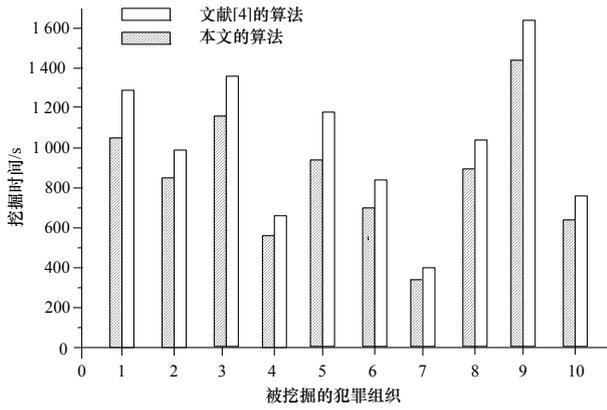


图2 2种算法挖掘时间的比较

在图2中, x 轴代表被挖掘的第 $i(1 \leq i \leq 10)$ 个犯罪组织, y 轴代表挖掘时间, 从图中可以看出, 对10个犯罪组织重点人员挖掘的过程中, 本文中重点人员判定算法运行的时间小于文献[4]中社团核心挖掘算法运行的时间。对于规模越大的犯罪组织, 2种算法执行的时间越长, 但由于本文中的重点人员挖掘算法可以忽略大量叶子节点, 2种算法的挖掘时间差值也随着组织规模的扩大而增加。因此, 在犯罪组织规模逐渐增大的情况下, 本文重点人员判定算法的时间要小于文献[4]中的社团核心挖掘算法。

4.3 实验2: 犯罪组织关系挖掘的效果

在对上述10个犯罪组织挖掘重点人员后, 为了说明犯罪组织关系挖掘的效果, 本文对其中的一个犯罪组织进行成员关系的挖掘。犯罪组织图形化构建如图3所示, 图中的边表示2个节点之间有通信关系, 边的权值为通信的频度。

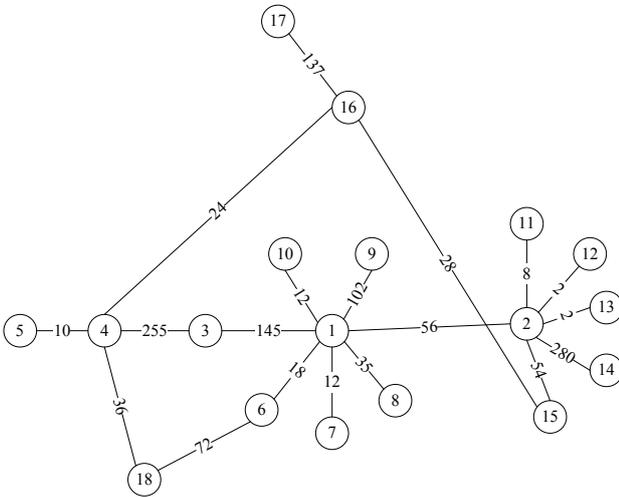


图3 犯罪组织图

根据犯罪组织成员关系挖掘方法, 该犯罪组织的成员关系如图4所示。图4以阶梯树的形式将犯罪组织成员间的关系表示出来, 节点1、节点2、节点4与图中1/4左右的人员都有直接联系, 且这3个节点的出度均大于入度, 由此可判断这3个节点的领导能力要大于其他节点。节点4直接支配节点3、节点5、节点16、节点18, 但节点4不被任何节点

所控制, 可以初步判定该犯罪组织由2个犯罪团伙组成, 领导者分别为节点1和节点4, 节点3、节点15、节点18隶属于2个团伙。由程序自动分析出的犯罪组织成员关系与人工分析后的结果基本一致, 但在实际中, 节点8、节点12、节点13并不是犯罪组织中的成员, 这是因为在构建犯罪组织及组织关系挖掘过程中, 尚未考虑成员间的通信内容、频度等因素, 在分析过程中, 只要两者有通信联系, 就将其纳入到犯罪组织中, 导致了部分结果的偏差。后续工作将考虑影响犯罪组织的因素, 对算法继续进行改进。

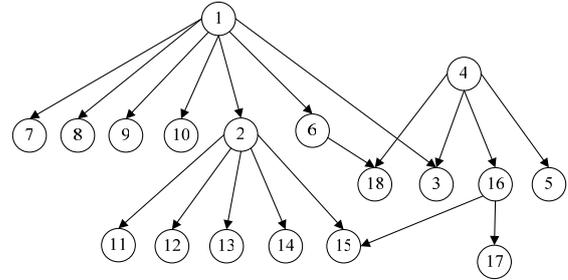


图4 犯罪组织成员关系

5 结束语

社会网络方法是研究犯罪组织关系的有效手段, 在对犯罪组织图形化构建的基础上, 本文利用社会网络方法对犯罪组织关系进行了研究, 包括犯罪组织重点人员的判定, 以及犯罪组织成员关系的挖掘。在对实际犯罪组织分析中, 分析结果与实际调查结果基本一致, 但由于未考虑到影响犯罪组织的其他因素, 因此造成与实际结果存在偏差。下一步工作将结合影响犯罪组织的因素, 对算法进一步改进, 减少分析结果与实际结果之间的误差。

参考文献

- [1] Estevez P A, Vera P, Saito K. Selecting the Most Influential Nodes in Social Networks[C]//Proceedings of International Joint Conference on Neural Networks. Orlando, Florida, USA: [s. n.], 2007: 12-17.
- [2] Santos E E, Pan Long, Arendt D, et al. An Effective Anytime Anywhere Parallel Approach for Centrality Measurements in Social Network Analysis[C]//Proceedings of 2006 IEEE International Conference on System, Man, and Cybernetics. [S. l.]: IEEE Press, 2006: 8-11.
- [3] Kiss C, Scholz A, Bichler M. Evaluating Centrality Measures in Large Call Graphs[C]//Proceedings of the 8th IEEE International Conference on E-Commerce Technology and the 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services. Washington D. C., USA: IEEE Computer Society Press, 2006.
- [4] 温粉莲, 唐常杰, 乔少杰, 等. 挖掘被监控社团核心的最短路径[EB/OL]. (2008-05-16). http://www.paper.edu.cn/downloadpaper.php?serial_number=200607-42.

编辑 索书志