

# 社交网络分析在公共安全领域的应用

邵莹侠,冯是聪

北京明略软件系统有限公司, 北京 102218

## 摘要

社交网络分析技术是一种通用有效的研究社会人员之间复杂关系模式的方法。以公安领域为背景,首先介绍了社交网络分析理论,然后详细阐述该技术在公安领域的3个应用案例,包括犯罪团伙挖掘、核心成员识别和串并案挖掘,以期为读者了解社交网络分析技术在公共安全领域的应用提供帮助。

## 关键词

社交网络分析;公共安全;犯罪团伙

中图分类号:TP311

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2017018

## *Applications of social network analysis in public security*

SHAO Yingxia, FENG Shicong

MiningLamp Software System Co., Ltd., Beijing 102218, China

### *Abstract*

Social network analysis (SNA) is a general and effective approach of studying the complex relationship patterns among social members. Public security field was focused. Firstly, the theory of SNA was introduced, and then three applications of applying SNA in public security were described, including crime gang mining, core criminal member identification, and serial and joint cases analysis. It's beneficial to readers to know about the capability of SNA in public security fields.

### *Key words*

social network analysis, public security, criminal gang

## 1 引言

近几年来,在线社交媒体(online social media)的迅速发展,微博、人人网、Facebook、Twitter等社交服务商的出现,给社交网络分析(social network analysis, SNA)技术的发展提供了充足的土壤。曾经只能通过调查问卷的方式收集几百到几千人的社交网络数据,如今通过公司合作或者数据交换等方式,可轻易获得千万到百亿级别的社交网络数据。为了处理和分析大数据环境下的社交网络数据,工业界和学术界都给社交网络分析技术注入了新鲜血液,提出了许许多多新的技术方案和指标。

社交网络分析主要研究人在社会网络中的位置及其信息在整个社会网络中传播形式等内容,通过关系分析,发现网络系统背后潜在的信息及价值,辅助决策者进行决策。如前文所述,随着社交媒体的出现,SNA技术变得越来越成熟,越来越完善。如今,它已经被应用到各个领域。比如,在社交媒体领域,服务提供商通过社交网络分析技术分析用户间好友或关注网络,可以发现网络中具有影响力的用户,进行广告投放,或者发现网络中的群体进行特定主题的内容推荐。在电子商务领域,利用用户的消费信息,可以得到用户与购买物品间的购物网络,进而抽取用户间的共同购买网络。在共同购买网络上,利用社交网络分析方法可以发现特定用户群体,从而推荐相应的商品。

同样地,社交网络分析技术作为一种社会研究的科学方法也为公安执法部门的智慧警务建设提供了一种有利手段。在实际的警务工作中,公安干警对案件的研判会利用犯罪嫌疑人既有的复杂社会关系,

对犯罪嫌疑人的可能动向、嫌疑性大小进行分析。与社交网络分析相同,整个研判过程可以使用网络结构进行建模。网络中的一个节点可以表示一个人、一个公司、一个组织,而边则代表着二者之间不同的信息交互类型,例如买卖双方的毒品交易、恐怖分子之间的通信联系、歹徒与受害者之间的联系等。由此构建的社交网络图,首先通过可视化技术,既可在视觉上直观地体现出犯罪嫌疑人之间、犯罪嫌疑人与受害人之间、犯罪嫌疑人与其亲人朋友之间的联系,又将抽象的案件分析具体化、数据化;然后,利用社交网络分析技术的定量分析,能够刻画出案件中的关键人物、团伙间的协作流程,为警员的案件侦破或者复盘提供丰富多角度的信息。

然而,由于公安领域内的信息化建设相对于互联网起步较晚,许多成熟的分析技术还并未真正运用到实战之中,社交网络分析技术就是一个实例。通过前面的简单分析可知,公安执法部门日常工作的服务对象还是以人为主,通过人与人之间的关系侦察办案,为社交网络分析技术的应用提供了充分的条件。下文笔者将讨论社交网络分析技术中的基本理论和方法,随后通过几个真实场景给出社交网络分析在公安领域进行实战的可能性。

## 2 社交网络分析理论及方法

社交网络分析理论是研究社会生活中人与人之间的网络如何形成、网络中的行为模式以及信息在网络中如何传递的一门学科。它重点关注人在网络中所处的位置和人的行为受整个网络的影响。从方法论角度出发,社交网络分析提供了一套面向网络的量化分析方法,能对网络中的关系模式、节点影响力等内容进行定量讨论。

一张社交网络主要包括两类基本元素，分别是节点和边。其中节点可以为现实生活中的人、组织、地点、物品等各类实体，边则为实体间的不同关系，比如人与人之间的好友、血缘关系，人和公司的从属关系，人与物品的拥有关系等。面向社交网络的分析输出内容主要有两方面：定量分析结果和可视化展示。下面从这两方面介绍相关的技术手段。

## 2.1 定量分析

定量分析的结果类型多种多样，包括网络中的顶点分析、链接分析、社区分析等内容。

顶点分析主要研究节点在网络中所处位置的重要性。通过量化的重要性反映该节点在网络中信息传播的影响力，影响越大的节点就越重要。在社交网络分析中主要通过中心性 (centrality)<sup>[1]</sup>对节点的重要性进行测量。经过多年的研究，已经提出了多种不同的中心性定义，包括 PageRank、超文本敏感标题搜索 (hypertext-induced topic search, HITS)、中介 (betweenness) 中心性、紧密 (closeness) 中心性、基于度 (degree) 的中心性等。每一种中心性对网络中信息流的传播做了不同的假设，从不同角度刻画一个节点在网络中的重要程度。以基于度的中心性为例，一个节点的度越大，即该节点直接联系的人越多，则认为其越重要。这个衡量指标适用于对事件型信息的刻画。以发表学术文章为例，两个作者若发表同一个文章，则互相之间产生一条边，那么在这样一个社交网络下，一个人的合作者越多，即网络中度越大，就说明该人在此学术界中影响力很大。另外，定量分析还包括对网络中群体的刻画，即期望发掘网络中具有相似特征的群体，以

便进行定向深入分析。

链接分析的主要研究对象则是网络中的边。类似顶点的研究，关于边的重要性的定量分析有边中心性的概念。除此之外，一个重要的研究方向是网络中的链接预测，即根据网络的全局拓扑结构及相关属性，预测节点间边存在的可能性。目前预测的方法主要分为两类。第一类方法是基于内容的特征工程，建立预测模型。它主要利用节点上的属性构造特征向量，比如年龄、性别、职业、兴趣爱好标签等。第二类方法是考虑网络结构特征构建预测模型。在生成特征向量时，除了考虑基本属性特征以外，还会引入共同好友数、ego-network 信息等特征，增加预测模型的维度特征。

社区分析则是关注群体在网络中的分布情况。此项技术在互联网应用中，经常用于发现具有相同兴趣爱好的人群，为互联网上的定向广告推荐提供重要的信息。针对此问题的方法同样可以分为两类：第一，基于特征工程的聚类算法，类似链接预测问题中的方法，对每个节点提取特征，利用聚类算法对顶点进行聚类；第二，一般认为属于同一社区的成员之间都存在一定的联系，于是有了如下的经验总结，社区成员间的边要比社区间的边更加稠密。基于上述经验总结，研究界提出了模块化 (modularity) 的概念，即模块内的稠密度 (density) 比模块间的高，具体定义可参考参考文献[2]。

## 2.2 可视化展示

可视化展示直接把网络的拓扑结构及不同属性的关系以点和线的方式呈现给用户，同时，用户可以通过交互式的方式探索和分析该网络。目前，针对社交网络的可视化研究同样可以分为两类。第一类是针对

局部图或者小规模图的展示,便于用户观察全图的局部,比如一个顶点的邻居网络,抑或是某个感兴趣的群体。关于此类小规模图的可视化布局已经有成熟的算法,比如Spring布局、Force-based布局等。第二类是针对大规模网络可视化问题研究相关技术方案。现实的社交网络规模越来越大,像Facebook、新浪微博等社交网络包含几亿个节点。传统的图布局算法无法处理新的大规模网络。最新的研究尝试利用图嵌入(graph embedding)方法,通过保留图结构中的邻近性,把图嵌入二维空间中<sup>[31]</sup>。而图嵌入方法通过优化,可以获得高效性能,从而支持大规模网络数据的可视化展示,为用户处理和分析大规模网络提供一种强有力的方法。

### 3 社交网络分析技术在公安领域的应用

本节主要探讨社交网络分析技术在公安领域的具体应用场景。

#### 3.1 犯罪团伙挖掘

随着社会的发展,犯罪手段和技术也在不断地演进,呈现出团伙作案的趋势。在一个犯罪团伙内部,成员们分工明确,通过相互协作实现隐蔽作案的目的。比如,电信网络诈骗团伙内部,有人负责编撰故事,有人负责跟受骗者沟通,有人负责洗钱;在涉毒团伙中,有人专门购买原材料,有人专门制毒,有人负责贩毒,还有人专门洗钱;在贩婴团伙中,有人负责寻找婴儿,有人运输,有人找买家,有人收账。总之,团伙往往分工明确,在办案过程中仅仅打掉其中一个环节或某几个环节,效果都不尽如人意,团伙仍有可能重新组织人手继续犯

罪。因此,明确团伙成员、分析清楚团伙内部成员的地位等级,将给执法办案提供极大的帮助,同时也是需要解决的一大难题。

社交网络分析为上述这一难题的解决带来了希望。前文提到,社区发现是社交网络分析的一项基本内容。而基于以往的经验了解到,犯罪团伙间都存在或多或少的联系,团伙内部的关系与团伙外部的关系有着显著的差异,比如,团伙的结构相对稀疏,但外围的人员与外界关系频繁等。根据上述经验,设置合理的模块优化目标<sup>[2]</sup>,可以利用社区发现的思想得到一个粗略的群体,然后再利用精细的分类方法,比如基于图核函数的聚类算法,从群体中分离出小团体。

举个具体的例子,在实际工作中,可以结合社会人员的出行轨迹、通信情况、住宿记录等信息建立多种关系,并形成一张异构网络。针对得到的异构网络,利用经典的图划分算法得到一系列高内聚性的群体。对于群体内部有较多(如大于3人)犯罪前科的人,则可以初步判定此群体为嫌疑犯罪团伙。针对感兴趣的犯罪团伙,再通过提取顶点特征,利用 $k$ -means算法对群体内顶点进行分类,形成精细的团伙。其中,顶点特征可以包括最近轨迹行为时间、通信频繁程度等。最后,根据团伙最新的活跃轨迹筛选出活跃度较高的群体,供业务人员进行研判分析。

综上,通过SNA技术,能有效地从海量数据中筛选出可疑度较高的团伙,从而降低了警员挖掘潜在团伙的人力成本。

#### 3.2 犯罪团伙中的核心成员识别

在分析出犯罪团伙以后,需要进一步识别团伙中各个成员的角色和地位。比如处于聚簇边缘、与外界联系相对较紧密

(与其他聚簇相连)的节点,可能是联络人,也可能是新入伙的人;在聚簇中心的可能是组织中的核心人物。所谓“擒贼先擒王”,首先识别出网络中的核心人员,警方再根据分析结果,选择特定的突破口,对这一组织实施打击粉碎行动。这里利用节点重要指标可以迅速定位挖掘犯罪团伙中的核心成员。前文已经提到过多个中心性衡量指标,这里具体介绍一下每类指标的含义及其应用条件。

首先,基于度定义的中心性 $C_{deg}$ 是最流行且最直接的一种定义。此中心性即相应节点的度数,一个节点关联的边越多就越重要。进一步,度中心性可以解释为统计了节点周围路径长度为1的路径数目。在此基础上,一个直接的泛化就是利用节点周围长度为 $K$ 的路径数目表示节点的重要性,称为 $K$ -path中心性。从路径长度定义类型的不同角度,出发, $K$ -path中心性也产生了很多变种:基于距离的 $K$ -path中心性、边独立的 $K$ -path中心性和节点独立的 $K$ -path中心性。从这类中心性指标定义出发,可以看出它们对信息传播的一个假设是信息以 $K$ 长度的路径进行传播。

相应地,一种流行的假设是信息的传播与节点间的最短路径有关。一般认为路径越短,信息传递越快。基于最短路径定义的中心性主要有紧密中心性和中介中心性。紧密中心性<sup>[4]</sup>可以描述为一个节点到网络中各个节点的平均最短路径距离的倒数,则节点 $v$ 的紧密中心 $c(v)$ 可按式(1)计算:

$$c(v) = \frac{n-1}{\sum_{\{v' \in V\}} d(v, v')} \quad (1)$$

其中, $n$ 表示网络中节点数目, $d(v, v')$ 表示节点 $v$ 和 $v'$ 之间的最短路径距离。紧密中心性是一个基于节点的单源最短路径距离分布的衡量指标,形象地描述了节点在网络中所处位置的重要程度。紧密中心性越大

的节点到其他节点的平均最短路径距离越小,处于网络的越中央,越能迅速地将信息传递给其他节点,同时也能更快速地接收其他节点的信息。

中介中心性的具体定义如下:

$$C_b(u) = \sum_{s \neq u \neq t \in V} \frac{\delta_{st}(u)}{\delta_{st}} \quad (2)$$

其中, $C_b(u)$ 表示节点 $u$ 的中介中心性, $\delta_{st}$ 表示节点 $s$ 和 $t$ 之间最短路的数目, $\delta_{st}(u)$ 表示节点 $s$ 和 $t$ 之间经过节点 $u$ 的最短路的数目。由定义可知,中介中心性反映了一个点所在的最短路径数目在全局所有最短路径对之间的一个占比。中介中心性越大,全网中将有越多的信息通过该点进行传递。

其次,以通路(walk)的概念为基础,也产生了多种相应的中心性概念。其中,Katz中心性用节点出发的通路数目表示,形式化表示为 $e_i^T \left( \sum_{j=1}^{\infty} (\beta A)^j \right) \mathbf{1}$ ,其中 $e_i$ 表示一个第 $i$ 个位置为1、其他位置为0的列向量。 $\beta$ 表示一个大于零的惩罚系数,反映了路径越长,其对中心性的贡献越小。在Katz中心性基础上,产生了Bonacich中心性,

其形式化为 $e_i^T \left( \frac{1}{\beta} \sum_{j=1}^{\infty} (\beta A)^j \right) \mathbf{1}$ ,并允许 $\beta$ 取负值。Katz中心性和Bonacich中心性都是Hubbell中心性的特殊形式。Hubbell中心性可以表示为 $e_i^T \left( \sum_{j=1}^{\infty} (X)^j \right) \mathbf{y}$ ,其中, $X$ 表示矩阵, $\mathbf{y}$ 表示向量。

最后,为了从网络的全局信息衡量一个点的重要性,提出了基于信息流的节点中心性定义。其中,PageRank、HITS等用来衡量一个网页节点在整个网络中的重要性的指标,就是经典的基于信息流的中心性定义。

上述多种中心性定义,从不同角度对网络中的节点重要性进行了刻画,根据实际网络情况,选择合理的指标对网络中的

重要节点进行识别。比如,在一个涉恐网络中,领导人物的命令要以最快的速度传播到整个网络中,那么紧密中心性将很好地识别出这类信息发布角色的重要节点。以美国“9·11”恐怖袭击的犯罪网络<sup>[5]</sup>为例,作者以新闻报道资料为基础构建了“9·11”袭击中的19个恐怖分子构成的关系网络。然后利用紧密中心性和中介中心性等中心性指标识别出了网络中的关键节点——Mohamed Atta,而事后证实确认该关键节点即该团伙的领导者。此例子给社交网络分析技术在公共安全中的应用提供了强有力的证据。

### 3.3 串并案挖掘

串并案是公安机关提高办案效率的一个有效途径,它是指把多个可能相关的案件一起进行研判分析。这里的相关包括同一伙犯罪成员或者是相似的作案手段等。通过对多个案件的串并分析,不仅能起到信息互补的作用,而且能解决不少冤假错案。社交网络分析有利于将同一犯罪团伙所为的案件进行串并。若独立地分析多个由同一犯罪团伙所为的各个案件,警员无法有效洞察其背后千丝万缕的联系。结合社交网络分析技术同样能给案件侦破提供线索。

以美国某州的警局针对该地区一系列便利店的抢劫案为例,介绍如何利用社交网络分析进行并案,从而抓捕了多个之前未被发现的犯罪嫌疑<sup>[6]</sup>。他们首先把单独的案件中涉及的人、事、物映射到一张网络上,然后在调查中发现某个正在被调查的人A与其他辖区的有案底的人B存在联系。于是,他们以这两人(A和B)的姓名为种子,利用类似标签传播的算法,构造了一张未知网络,即网络中出现了多个未被关注的人员。对这个未知网络进一步分析,发

现该有案底人员A又与多个青少年有频繁的联系,于是就把之前未被人视为同一伙人作案的抢劫案给串联起来了。最后,通过可视化展现的方式,对多张网络进行浏览和管理,为干警破案提供了快速有效的线索。

## 4 结束语

公安领域利用网络分析进行侦查办案的历史由来已久。传统的方式是将涉案人员以图钉和连线的方式在黑板上展示出来。近年来,社交网络分析技术的发展,不仅将传统的方式数字化,利用计算机就能快速分析案情,而且使得干警能够处理和分析更大规模网络,从而提高办案的效率和质量。除了社交网络分析技术,时序分析、数据挖掘、机器学习等多种其他大数据分析技术同样能够提升公安的侦查能力,从而极力保证社会更加稳定、和谐地发展。

## 参考文献:

- [1] FREEMAN L C. Centrality in social networks conceptual clarification [J]. *Social Networks*, 1978, 1(3): 215-239.
- [2] NEWMAN M E J. Fast algorithm for detecting community structure in networks[J]. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2004, 69(6): 066133.
- [3] TANG J, LIU J Z, ZHANG M, et al. Visualizing large-scale and high-dimensional data[C]// *The 25th International Conference on World Wide Web*, April 11-15, 2016, Montréal, Québec, Canada. New York: ACM Press, 2016: 287-297.
- [4] 邵蕊侠, 崔斌, 马林, 等. 一种基于 Sketch 的

Top-k 紧密中心性快速搜索算法[J]. 计算机学报, 2016, 39(10): 1965-1978.

SHAO Y X, CUI B, MA L, et al. A fast sketch-based approach of Top-k closeness centrality search on large networks[J]. Chinese Journal of Computers, 2016, 39(10): 1965-1978.

[5] KREBS V E. Mapping networks of terrorist cells[J]. Connections, 2002, 24(3): 43-52.

[6] JOHNSON J A, REITZEL J D, NORWOOD B F, et al. Social network analysis: a systematic approach for investigating[J]. FBI Law Enforcement Bulletin, 2013, 82(3): 1.

#### 作者简介



**邵懿侠** (1988-), 男, 博士, 北京明略软件系统有限公司技术经理, 主要研究方向为大规模图计算优化、图挖掘应用以及复杂网络分析等。目前在SIGMOD、VLDB、TKDE等国际一流学术会议和期刊上发表10余篇相关学术论文。曾获得2014年谷歌博士奖学金和微软学者等称号。



**冯是聪** (1973-), 男, 博士, 北京明略软件系统有限公司联合创始人兼CTO, 中国中文信息学会 (CIPS) 理事, 中国计算机学会 (CCF) 大数据专家委员会委员, 上海市数据科学重点实验室 (复旦大学) 数据科学家, 中国人民大学信息学院企业硕士导师, 上海交通大学大数据与商业创新实验室客座研究员。主要研究方向为大数据、云计算、计算广告学、信息检索等。

收稿日期: 2017-01-09