

# 基于异构信息网络的假币犯罪信息分析系统<sup>①</sup>

陈贞虹<sup>1,2</sup>, 李 慧<sup>2</sup>, 许舒人<sup>2</sup>, 叶 丹<sup>2</sup>

<sup>1</sup>(中国科学院大学, 北京 100049)

<sup>2</sup>(中国科学院软件研究所 软件工程技术研究开发中心, 北京 100190)

**摘 要:** 传统的犯罪查询的查询条件是文本信息, 查询结果是有序的文档列表, 这种方式无法展示结果之间的关系. 基于异构信息网络以信息网络的形式重构假币犯罪信息数据, 构建了假币犯罪信息网络, 使用人名消歧的技术建立假币犯罪信息网络中嫌疑人之间的关系, 并使用排序学习方法研究假币犯罪信息网络中的节点相关性问题的, 设计并实现了假币犯罪信息分析系统, 通过以实体对象为查询项和网络图为查询结果的方式解决假币犯罪数据的查询问题.

**关键词:** 异构信息网络; 元路径; 犯罪数据挖掘; 人名消歧; 排序学习

## Counterfeit Money Crimes Information Analysis System Based on Heterogeneous Information Networks

CHEN Zhen-Hong<sup>1,2</sup>, LI Hui<sup>2</sup>, XU Shu-Ren<sup>2</sup>, YE Dan<sup>2</sup>

<sup>1</sup>(University of Chinese Academy of Sciences, Beijing 100049, China)

<sup>2</sup>(Technology Center of Software Engineering, Institute of Software, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** The query condition of traditional criminal query is a text, and query results are ordered documents lists. However, the way is not conducive to present the relationship of the results and help the police find clues of cases. Therefore, this paper reconstructs counterfeit crime data in the form of information network based on heterogeneous information network information, and constructs a counterfeit money crimes information network. We apply name disambiguation to build relationships of the suspects in counterfeit crime information network, and study relevancy problem of counterfeit money crimes information network using learning to rank methods. In this paper we design and implement the counterfeit money crimes information analysis system. To resolve the query problem of the criminal data, we use the entity as a query term and the network graph as the query results.

**Key words:** heterogeneous information network; meta-path; crime data mining; name disambiguation; learning to rank

## 1 引言

随着犯罪行为的智能化和高科技化, 犯罪人员的组织化和职业化趋势也越来越明显, 假币犯罪发案数呈现增长态势, 犯罪规模不断升级, 假币犯罪有产业化和团伙化的特点. 传统的犯罪查询的查询条件是一段文本, 查询结果是有序的文档列表. 对于假币犯罪信息分析来说, 这种查询方式无法展示查询结果之间的关系, 对于具有团伙作案特性的假币犯罪来说, 这种查询方式并不适合. 本文采用以实体对象为查询项,

网络图为查询结果的方式解决假币犯罪分析中数据查询问题, 帮助警员分析假币犯罪信息, 寻找破案线索.

为了提供更加高效的假币犯罪信息查询, 本文对原始假币犯罪数据进行重构, 建立假币犯罪信息网络, 基于假币犯罪信息网络研究假币犯罪信息的查询分析与可视化. 随着信息网络结构和关系复杂化的发展, 一个信息网络内部的会有多种类型的节点, 节点之间的关系也会呈现出多样化的趋势, 为了区别单一同构的信息网络, 韩家炜等学者提出将此种信息网络称为

<sup>①</sup> 基金项目: 国家自然科学基金(61379044); 国家科技支撑课题(2015BAH18F02)

收稿时间: 2016-03-19; 收到修改稿时间: 2016-04-18 [doi: 10.15888/j.cnki.csa.005457]

异构信息网络<sup>[1]</sup>. 与同构信息网络相比, 存在多种节点和关系类型的异构信息网络模型更加适合含有三种关键数据类型, 即案件、嫌疑人和假币的假币犯罪数据. 因此本文基于异构信息网络建立假币犯罪信息网络模型, 研究了嫌疑人姓名消歧和假币犯罪信息相关性分析两项关键技术, 并在此基础上设计和实现了假币犯罪信息分析系统.

## 2 相关工作

本文基于异构信息网络以信息网络的形式重构假币犯罪数据, 构建了假币犯罪信息网络, 使用人名消歧的技术建立假币犯罪信息网络中嫌疑人之间的关系, 并使用排序学习方法研究假币犯罪信息网络中的节点相关性. 因此, 本节从异构信息网络、人名消歧以及排序学习三个方面介绍本文的相关工作.

### 2.1 异构信息网络

网络分析的研究已经从单一的网络结构逐渐向异构网络结构发展, 并成为新的研究热点, 早期的异构信息的研究包括 RankClus<sup>[1]</sup>和 NetClus<sup>[2]</sup>. 但是 RankClus 针对的是只有两种节点类型的异构信息网络, 而 NetClus 是基于星形异构信息网络的研究, 在应用上都有一定的局限性. 随后 Yizhou Sun 提出用元路径<sup>[3]</sup>表示异构信息网络丰富的语义信息, 同时还定义了一种新的基于元路径的相似性测量方法 PathSim<sup>[3]</sup>, 研究异构信息网络中同种对象之间的相似性. 此后有很多异构信息网络的研究都基于元路径展开. 文献[4]在学术研究网络中基于异构信息网络的元路径提取网络拓扑特征预测学者之后的合作关系. 综上, 异构信息网络的元路径分析方法是应用最广泛的分析方法, 因此本文选用基于元路径的分析方法研究基于异构信息网络构建的假币犯罪信息网络.

### 2.2 人名消歧

人名消歧的研究中应用较为广泛的是基于聚类的人名消歧. 基于聚类的人名消歧主要使用人物的属性信息进行聚类, 有效的人物属性可以提高聚类的结果的准确性. 基于聚类的人名消歧中的重要问题是如何确定类的个数, 文献[5,6]采用交叉验证、排列检测、重采样以及惩罚似然等方法来确定类的个数. 此外, 文献[7,8]采用了自底向上的凝聚式层次聚类方法, 能够不事先确定簇的个数. 本文采用自底向上的凝聚式层次聚类方法解决嫌疑人姓名消歧问题.

### 2.3 排序学习

排序学习方法是在信息检索问题中, 利用机器学习的方法, 基于特征集合, 自动学习得到用于检索的最终排序函数. 已有的排序学习算法可以根据训练样例的不同分为三类: Pointwise、Pairwise 和 Listwise 方法. Pointwise 方法, 如 Pranking with ranking 算法<sup>[9]</sup>, 采用一个查询对应的单一文档作为训练样例; Pairwise 方法, 如 Ranking SVM 算法<sup>[10,11]</sup>, 采用查询对应的文档对作为训练样例; 而 Listwise 方法, 如 ListNet 算法<sup>[12]</sup>, 采用查询对应的文档序列作为训练样例. 已有研究表明, 一般来说 Pairwise 和 Listwise 方法优于 Pointwise 方法<sup>[13,14]</sup>. 由于 Listwise 方法的算法复杂度太高, 本文选用应用更为成熟的 Pairwise 方法中的 Ranking SVM 算法.

## 3 假币犯罪信息网络模型

为了给假币犯罪数据提供网络化的查询方式, 本文基于异构信息网络建立假币犯罪信息网络模型.

### 3.1 相关定义

定义 1. 信息网络<sup>[3]</sup>(Information network): 信息网络被定义为有向图  $G = \langle V, E \rangle$  的形式, 有一个对象类型映射函数  $\phi: V \rightarrow A$  和一个链接类型映射函数  $\varphi: E \rightarrow R$ , 即每个对象  $v \in V$  属于一个特定的对象类型  $\phi(v) \in A$ , 每一个关系  $e \in E$  属于特定的关系类型  $\varphi(e) \in R$ . 当对象类型  $|A| > 1$  或关系类型  $|R| > 1$  时, 该信息网络称为异构信息网络, 否则为同构信息网络.

为了明确异构信息网络中的对象类型和关系类型, 使用网络模式来描述异构信息网络的元结构.

定义 2. 网络模式<sup>[3]</sup>(Network schema): 在异构信息网络  $G = (V, E)$  中, 有对象类型映射  $\varphi: V \rightarrow A$  和链接类型映射  $\phi: E \rightarrow R$ , 网络模式是在对象类型  $A$  上定义的有向图, 边为来自  $R$  的关系, 表示为  $T_G = (A, R)$ .

定义 3. 元路径<sup>[3]</sup>(Meta-path): 元路径  $P$  是一条定义在网络模式图  $T_G = (A, R)$  上的路径, 表示为  $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \dots \xrightarrow{R_j} A_{j+1}$ , 它定义了对象类型  $A_1$  和  $A_{j+1}$  之间的复合关系  $R = R_1 \circ R_2 \circ \dots \circ R_j$ , 其中  $\circ$  表示关系的复合操作.

元路径描述了网络模式中两类对象之间的连接路径, 表示两类节点在元层次上存在的关系, 不同的元路径代表了对象类型之间可以通过不同的节点类型和

关系建立不同的关系。

### 3.2 假币犯罪信息网络

定义4. 假币犯罪信息网络: 假币犯罪信息网络是异构信息网络的实例, 被定义为有向图  $G = \langle V, E \rangle$  的形式, 有一个对象类型映射函数  $\varphi: V \rightarrow A$  和一个关系类型映射函数  $\phi: E \rightarrow R$ , 其中节点类型  $A = \{\text{Case, Person, Money}\}$ , 即案件(Case, C), 嫌疑人(Person, P)和假币(Money, M)三种对象, 关系类型  $R = \{\text{involve, involved in, related by, identical, similar}\}$ , 六种关系类型描述如下:

- ① **Involve**: 指的是该案件所抓获的犯罪嫌疑人;
- ② **Involved in**: 指的是该犯罪嫌疑人所涉及的案件;
- ③ **Relate**: 指的是在一个案件中所缴获的假币;
- ④ **Related by**: 指的是该假币是在哪一个案件当中被缴获的;
- ⑤ **Identical**: 两条同名嫌疑人记录指向同一嫌疑人;
- ⑥ **Similar**: 通过已有假币相似度分析程序计算相似度, 并建立关联关系。

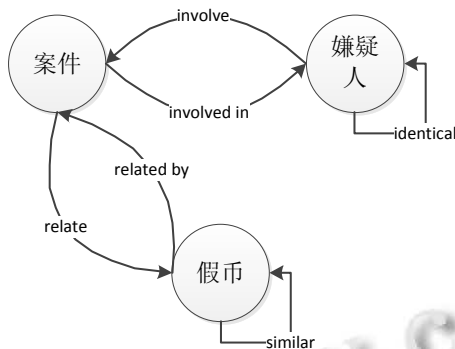


图1 假币犯罪信息网络的网络模式

根据定义4, 假币犯罪信息网络的网络模式如图1所示。基于假币犯罪信息网络的网络模式, 假币犯罪信息网络的元路径实例如图2所示, 其中元路径嫌疑人—案件—嫌疑人(PCP)表示人和人之间通过参与了相同的案件建立了一条路径, 路径中包括人和案件之间的“involved in”关系以及案件和人之间的“involve”关系。另一条元路径案件—假币—假币—案件(CMMC)表示两个案件之间通过关联了非常相似的假币建立了一条路径, 路径中包括案件和假币之间的“relate”关系, 假币和假币之间的“similar”关系以及

假币和案件之间“related by”关系。因此, 不同的元路径具有不同的语义来描述节点之间的相关度。

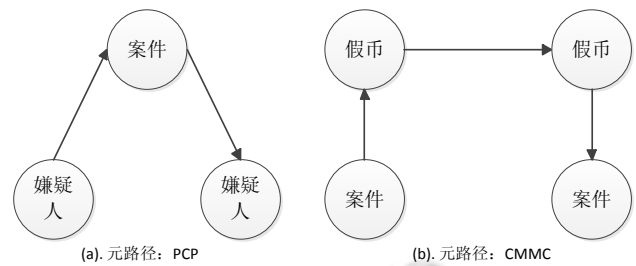


图2 假币犯罪信息网络的元路径实例

## 4 关键技术研究

本节介绍假币犯罪信息分析系统中使用的两个关键技术, 嫌疑人姓名消歧和假币犯罪相关性分析。嫌疑人姓名消歧用于构建假币犯罪信息网络的嫌疑之间的关系, 应用于假币犯罪分析系统的构建假币犯罪信息网络模块。假币犯罪信息相关性分析是基于假币犯罪信息网络提取节点网络拓扑结构特征, 同时使用排序学习的方法融合节点的属性信息和拓扑特征分析假币犯罪信息相关性, 应用于假币犯罪分析系统的数据分析模块。

### 4.1 嫌疑人姓名消歧

为了解决假币犯罪中嫌疑人记录重复录入导致的同一嫌疑人所犯的案件无法建立关联关系的问题, 本文使用人名消歧技术判断同名嫌疑人是否指向同一人。对于被认定为同一嫌疑人的记录不进行合并, 而是建立关联关系, 即假币犯罪信息网络中的嫌疑人之间的“identical”关系记为同一嫌疑人关系。

#### 4.1.1 嫌疑人属性

经过业务分析, 选择了28个属性作为嫌疑人消歧中使用的属性, 并将这28个属性分为四类:

- 1) 同一性判定属性: 此类属性具有较强的同一性判定功能, 如手机号、QQ号、银行账户号等。
- 2) 重要属性: 若该类属性相同, 则两条同名嫌疑人记录有较大可能性指向同一嫌疑人, 如现住址、户籍地址、绰号等。
- 3) 一般属性: 一般的常规属性, 如性别、民族和血型等。
- 4) 模糊属性: 警员在录入过程存在主观性误差的属性, 如体型、脸型、身高和口音等属性。

#### 4.1.2 两阶段聚类模型

本文采用自底向上的凝聚式层次聚类方法, 并分为两个阶段:

① 阶段一: 该阶段仅使用同一性判定属性. 初始状态下每一个同名的嫌疑人各为一个簇, 发现具有任一相同的同一性判定属性则合并成一个簇, 并输出若干个簇作为该阶段的聚类结果.

② 阶段二: 该阶段使用重要属性、一般属性和模糊属性三类属性计算两个嫌疑人之间的相似度, 对于不同类别的属性赋予不同的权重. 使用公式(1)计算嫌疑人  $p_x$  和嫌疑人  $p_y$  的属性  $k$  的相似度.

$$d_k(p_x, p_y) = \begin{cases} 1 & p_x \text{ 的属性 } k \text{ 和 } p_y \text{ 相同} \\ 0 & p_x \text{ 和 } p_y \text{ 任一属性 } k \text{ 为未知} \\ -1 & p_x \text{ 的属性 } k \text{ 和 } p_y \text{ 不同} \end{cases} \quad (1)$$

分别使用公式(2)(3)(4)计算嫌疑人  $p_x$  和嫌疑人  $p_y$  之间的重要属性, 一般属性以及模糊属性的相似性.

$$impSim(p_x, p_y) = \sum_{k \in A_{imp}} d_k(p_x, p_y) \quad (2)$$

$$norSim(p_x, p_y) = \sum_{k \in A_{nor}} d_k(p_x, p_y) \quad (3)$$

$$fuzSim(p_x, p_y) = \sum_{k \in A_{fuz}} d_k(p_x, p_y) \quad (4)$$

其中  $A_{imp}$ ,  $A_{nor}$ ,  $A_{fuz}$  分别表示重要属性集合, 一般属性集合以及模糊属性集合. 使用公式(5)计算嫌疑人  $p_x$  和嫌疑人  $p_y$  之间的相似度.

$$suspectSim(p_x, p_y) = \alpha_{imp} impSim(p_x, p_y) + \alpha_{nor} norSim(p_x, p_y) + \alpha_{fuz} fuzSim(p_x, p_y) \quad (5)$$

其中  $\alpha_{imp}$ ,  $\alpha_{nor}$ ,  $\alpha_{fuz}$  分别表示重要属性, 一般属性和模糊属性的权重. 将阶段一的输出的若干个簇作为该阶段的输入, 并使用公式(6)计算簇  $C_i$  和簇  $C_j$  之间的相似度.

$$sim(C_i, C_j) = \max_{p_x \in C_i, p_y \in C_j} suspectSim(p_x, p_y) \quad (6)$$

迭代过程中具有最大相似度的两个簇合并成一个簇, 当任意两个簇之间最大相似度小于一个给定的阈值  $t$  时停止聚类.

## 4.2 假币犯罪信息相关性分析

假币犯罪信息在本文中以假币犯罪信息网络的形式组织, 因此假币犯罪信息的相关性分析即为假币犯罪信息网络中节点的相关性分析.

### 4.2.1 节点相关度排序关键信息

传统的排序模型常常只关注节点的属性特征, 或

者只关注节点的拓扑特征, 但是只关注其中一种特征, 往往会造成结果的置信度不够, 本文将融合以上两类特征来决定排序结果.

#### 1) 节点之间的拓扑特征

节点的拓扑特征使用异构信息网络的元路径分析方法提取, 使用文献[4]的四个基于元路径的网络度量方法, 分别是路径数(path count)、归一化路径数(normalized path count)、随机游走(random walk)以及对称随机游走(symmetric random walk)作为拓扑特征的度量方法. 根据文献[3], 较短的元路径能够较好的度量相关度, 一条长的元路径甚至会造成干扰. 因此本文将选用长度不大于 4 的元路径, 使用传统的广度优先搜索算法, 在网络模式图上搜索出所有符合条件的元路径. 对于每一个类型的节点拓扑特征数为元路径的个数乘以度量方法个数.

#### 2) 节点的属性特征

节点的属性特征直接从节点的属性中通过业务分析选取, 部分属性特征列举如下:

- ① 案件: 案发时间, 案发地点, 简要案情等.
- ② 嫌疑: 居住地, 户籍地址, 是否惯犯等.
- ③ 假币: 冠字号, 年版, 币种, 厚度等.

### 4.2.2 节点相关性排序框架

图 3 是节点相关性排序的框架图, 训练数据为已经标注好的节点相关性数据, 选用 Ranking SVM 算法训练排序模型, 使用训练集的数据训练排序模型. 测试数据的结构和训练数据相同, 对测试集中的每一条记录, 使用排序模型计算待分析节点的得分获得相关性结果, 其得分是一个在 0 到 1 之间的小数, 得分越高说明两个节点越相关.

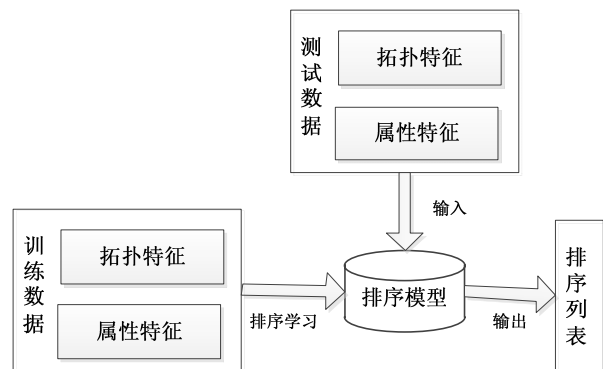


图 3 节点相关度排序框架示意图

### 5 假币犯罪信息分析系统的设计与实现

为了解决假币犯罪信息的查询问题，基于第三节介绍的假币犯罪信息网络模型构建了假币犯罪信息分析系统。假币犯罪信息分析系统提供假币犯罪数据的查询分析与可视化服务，帮助警员分析假币犯罪数据。假币犯罪信息分析系统的系统结构图如图 4 所示。构建假币犯罪信息网络模块从原始的假币犯罪数据存储的关系型数据库以及一些文本文件中抽取假币、案件和嫌疑人的属性信息和关系信息，同时使用 4.1 节介绍的嫌疑人姓名消歧算法解决原始嫌疑人建立嫌疑人关系，构建假币犯罪信息网络；数据存储模块持久化假币犯罪信息网络；数据分析模块是假币犯罪分析系统中的核心模块，负责处理用户查询，从数据存储模块存储的假币犯罪信息网络中查询出和查询节点相关并且符合用户定义的查询反胃的所有节点和关系，组成查询结果网络，使用 4.2 节介绍的假币犯罪信息相关性分析技术分析查询项与查询结果之间的相关性，返回结合相关性分析结果的查询结果网络；数据展示模块是假币犯罪分析系统与用户直接交互的模块，收集用户查询条件和查询结果网络的可视化展示。



图 4 假币犯罪信息分析系统结构图

#### 5.1 构建假币犯罪信息网络模块

构建假币犯罪信息网络模块负责从数据源，即关系型数据库中抽取关系和属性，构建假币犯罪信息网络并将结果传入数据存储模块。根据假币犯罪信息网

络关系的生成方式的不同分为以下三类：

- ① 原关系型数据库的关系和属性：嫌疑人和案件以及案件和假币之间的四种关系类型可以从原关系型数据库中直接抽取，同时对象的属性信息也从原数据库中一同抽取，用于数据分析。
- ② 假币和假币的 identical 关系：通过已有假币相似度分析程序计算相似度建立关联关系。
- ③ 嫌疑人和嫌疑人的 similar 关系：根据嫌疑人姓名消歧算法建立联系。

#### 5.2 数据存储模块

假币犯罪分析系统需要持久化的数据是假币犯罪信息网络，Neo4J 是图数据库，适合存储网络形式的数据，在社交网络领域已经有了成熟的应用，同时 Neo4J 能处理具有高达数十亿规模顶点和边的图数据，同时兼容完全 ACID 事务属性<sup>[15]</sup>。本文选用 NoSQL 数据库中的图数据库 Neo4J 作为假币犯罪分析系统的数据库。假币犯罪信息网络的数据库模式图如图 5 所示，和假币犯罪信息网络的结构完全一致，同时还存储了节点的属性信息，因此使用 Neo4J 图数据库存储假币犯罪信息网络，降低模型转换的代价，提高查询效率。

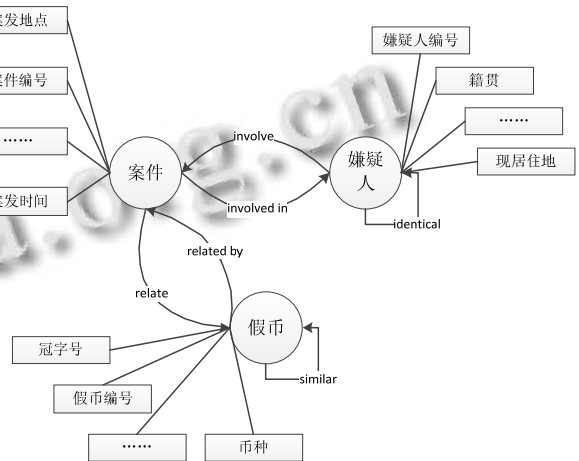


图 5 假币犯罪信息网络的数据库模式图

#### 5.3 数据分析模块

数据分析模块的主要功能是处理用户的查询请求，返回查询结果网络，分为用户查询处理，相关性分析，属性特征计算和拓扑特征计算四个小模块。

图 6 是数据分析的流程图，展示了数据分析模块中的四个小模块是如何协作完成数据分析的工作。首先用户查询处理模块接收用户的查询条件，根据用户

指定的查询节点以及限定的查询犯罪从数据存储模块存储的获取所有符合查询条件的节点，并生成查询结果网络，然后将生成的查询结果网络传送给相关性分析模块。相关性分析模块将查询结果网络分解为查询和待分析的相关性节点的二元组 $\langle Q, R \rangle$ ，同时调用拓扑特征计算模块计算该二元组的拓扑特征和属性特征计算模块计算属性特征，这两个模块将计算结果返回给相关性分析模块，由相关性分析模块汇总二元组的拓扑特征和属性特征，调用 4.2 节介绍的节点相关性排序模型计算相关性，然后将结果返回给用户查询处理模块，最后由用户查询处理模块将相关性信息加入查询结果网络，最后返回查询结果网络。

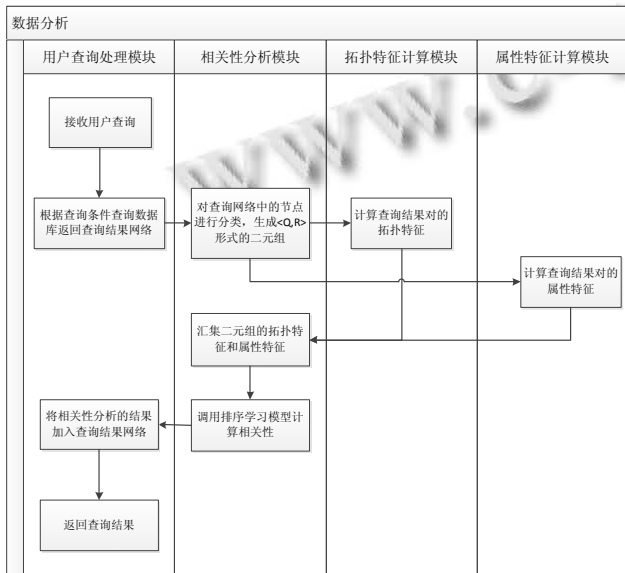


图 6 数据分析流程图

### 5.4 数据展示模块

数据展示模块负责用户交互，包括收集用户查询条件以及查询结果网络的可视化。用户查询条件包括指定查询实体，以及搜索的路径长度。

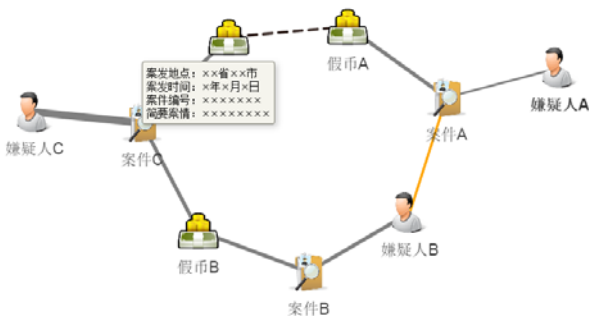


图 7 查询结果网络示例

图 7 是查询结果网络的示例，使用不同的图片表示不同类型的节点，用边的颜色和形状来区分不同的关系，同时使用悬浮窗口显示节点的属性信息，并根据结果节点和查询节点之间的相关性分析结果决定该节点显示的尺寸。

## 6 结语

本文将大量的犯罪数据以网络的形式组织起来，建立了假币犯罪信息网络，使用排序学习的方法融合属性信息和拓扑信息解决假币犯罪信息网络中的节点相关度排序问题，并在此基础上设计了假币犯罪信息分析系统，提供查询分析与可视化服务帮助警员进行假币犯罪分析和并案分析。在下一步工作中，我们将基于假币犯罪信息网络上挖掘假币犯罪团伙信息。

### 参考文献

- 1 Sun Y, Han J, Zhao P, et al. Rankclus: Integrating clustering with ranking for heterogeneous information network analysis. Proc. of the 12th International Conference on Extending Database Technology: Advances in Database Technology. ACM. 2009. 565–576.
- 2 Sun Y, Yu Y, Han J. Ranking-based clustering of heterogeneous information networks with star network schema. Proc. of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM. 2009. 797–806.
- 3 Sun Y, Han J, Yan X, et al. Pathsims: Meta path-based top-k similarity search in heterogeneous information networks. VLDB'11. 2011.
- 4 Sun Y, Barber R, Gupta M, et al. Co-author relationship prediction in heterogeneous bibliographic networks. 2011 International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE. 2011. 121–128.
- 5 Chen Y, Martin J. Cu-comsem: Exploring rich features for unsupervised web personal name disambiguation. Proc. of the 4th International Workshop on Semantic Evaluations. Association for Computational Linguistics. 2007. 125–128.
- 6 Elmacioglu E, Tan YF, Yan S, et al. Psnus: Web people name disambiguation by simple clustering with rich features. Proc. of the 4th International Workshop on Semantic Evaluations. Association for Computational Linguistics. 2007.

- 268–271.
- 7 Chen Y, Lee SYM, Huang CR. Polyuhk: A robust information extraction system for web personal names. 2nd Web People Search Evaluation Workshop (WePS 2009), 18th WWW Conference. 2009.
  - 8 Gong J, Oard D. Determine the entity number in hierarchical clustering for web personal name disambiguation. 2nd Web People Search Evaluation Workshop (WePS 2009), 18th WWW Conference. 2009.
  - 9 Crammer K, Singer Y. Pranking with Ranking. *Advances in Neural Information Processing Systems*, 2002, 14: 641–647.
  - 10 Herbrich R, Graepel T, Obermayer K. Large margin rank boundaries for ordinal regression. *Advances in Neural Information Processing Systems*, 1999: 115–132.
  - 11 Joachims T. Optimizing search engines using clickthrough data. *Proc. of the eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. 2002. 133–142.
  - 12 Cao Z, Qin T, Liu TY, et al. Learning to rank: from pairwise approach to listwise approach. *Proc. of the 24th International Conference on Machine Learning*. ACM. 2007. 129–136.
  - 13 Zhang M, Kuang D, Hua G, et al. Is learning to rank effective for Web search? *SIGIR 2009 Workshop: Learning to Rank for Information Retrieval*. 2009. 641–647.
  - 14 Liu TY. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 2009, 3(3): 225–331.
  - 15 Eifrem E. Neo4J-the benefits of graph databases. 2009. <http://www.oscon.com/oscon2009/public/schedule/detail/8364>.