(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 111291229 B
(45)授权公告日 2023.10.31

(21)申请号 202010071390.7

(22)申请日 2020.01.21

(65)同一申请的已公布的文献号
申请公布号 CN 111291229 A

(43)申请公布日 2020.06.16

(73)专利权人 中国科学院计算技术研究所
地址 100080 北京市海淀区中关村科学院南路6号

(72)发明人 刘盛华 石川 程学旗 李香峰 沈华伟 刘财政

(74)专利代理机构 北京律诚同业知识产权代理有限公司 11006
专利代理师 祁建国

(51)Int.Cl.
*G06F 16/901*(2019.01)
*G06F 16/9035*(2019.01)
*G06Q 40/04*(2012.01)

(56)对比文件
CN 110400220 A,2019.11.01
CN 107832964 A,2018.03.23
CN 109947814 A,2019.06.28
CN 109753797 A,2019.05.14
CN 110490730 A,2019.11.22
CN 109710754 A,2019.05.03
US 9787640 B1,2017.10.10
US 2017149814 A1,2017.05.25
杨莉、薛耀文、高慧敏.金融网络中资金异常流动监测的可视化支持研究.《计算机技术与发展》.2008,192-198.
P. Dickinson;H. Bunke;A. Dadej;M. Kraetzl.Median graphs and anomalous change detection in communication networks.《Final Program and Abstracts on Information, Decision and Control》.2002, 20-25.
官赛萍、靳小龙、贾岩涛、王元卓、程学旗.面向知识图谱的知识推理研究进展.《软件学报》.2018,2966-2994.
郑剑、周艳丽、刘聪.面向IaaS云平台的用户异常行为检测方法.《江西理工大学学报》.2016, 68-73.
杨冬梅等.金融网络中洗钱资金异常转移路径的经济成本模型.《系统工程理论与实践》.2006,(第05期),25-31.

审查员 胡燕清

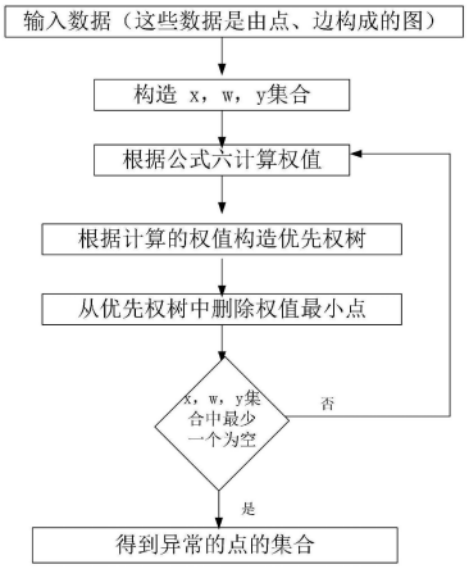
权利要求书3页 说明书8页 附图2页

(54)发明名称

一种稠密多部子图的检测方法及系统

(57)摘要

本发明提出一种基于稠密多部子图的检测方法及系统,包括:步骤1、根据链式特征中的信息流动,构建交易网络的多部图,根据预设的账户间信息流动阈值筛选该多部图,得到该多部图中的稠密子图;步骤2、以固定账户存在超阈值的信息流且在中间账户中保留低于阈值的权重为约束条件,生成该稠密子图中节点子集的异常值;步骤3、根据该异常值,输出该多部图中存在异常行为的节点子集作为异常行为检测结果。本发明通过具有有效性和鲁棒性和良好的可扩展性。




CN 111291229 B

1.一种基于稠密多部子图的检测方法，其特征在于，包括：

步骤1、根据链式特征中的信息流动，构建交易网络的多部图，根据预设的账户间信息流动阈值筛选该多部图，得到该多部图中的稠密子图；

步骤2、以固定账户存在超阈值的信息流且在中间账户中保留低于阈值的权重为约束条件，生成该稠密子图中节点子集的异常值；

步骤3、根据该异常值，输出该多部图中存在异常行为的节点子集作为异常行为检测结果；

其中，步骤1中该多部图$\mathcal{G}=(\mathcal{V},\mathcal{E})$，$\mathcal{V}=\mathcal{X}\cup\mathcal{W}\cup\mathcal{Y}$，其中$\mathcal{W}$是银行或者消息炒作团体的内部账户的集合，$\mathcal{X}$和$\mathcal{Y}$是银行或者消息炒作团体的外部的集合，其中$\mathcal{X}$是对银行净转入的账户集合或者是发布消息的团体集合，$\mathcal{Y}$是接受银行净转出的账户集合或者最终转发消息的集合，多部图中的边对应于从$\mathcal{X}$到$\mathcal{W}$以及从$\mathcal{W}$到$\mathcal{Y}$的权重，对于$i,j\in\mathcal{V}$，边(i，j)∈ε表示账户i将钱或者信息转给j；该稠密子图的节点子集$\mathcal{S}=\mathcal{A}\cup\mathcal{M}\cup\mathcal{C}$，其中$\mathcal{A}\subseteq\mathcal{X}$，$\mathcal{M}\subseteq\mathcal{W}$，$\mathcal{C}\subseteq\mathcal{Y}$；

该步骤2具体为：

根据该节点子集中为从$v_i$到$v_j$的总权重$e_{ij}$，权重使用转账金额或者消息转发的数量来衡量，得到节点$v_i\in\mathcal{M}$关于节点子集$\mathcal{S}$的总入度$d_i^-$和出度值$d_i^+$：

$$d_i^+(\mathcal{S})=\sum_{v_j\in\mathcal{C}\wedge(i,j)\in\varepsilon}e_{ij},d_i^-(\mathcal{S})=\sum_{v_k\in\mathcal{A}\wedge(k,i)\in\varepsilon}e_{ki} \quad \text{（公式一）}$$

得到一个中间账户关于节点子集$\mathcal{S}$的总加权出度和入度的最小值和最大值：

$$f_i(\mathcal{S})=min\{d_i^+(\mathcal{S}),d_i^-(\mathcal{S})\},\forall v_i\in\mathcal{M} \quad \text{（公式二）}$$

$$q_i(\mathcal{S})=max\{d_i^+(\mathcal{S}),d_i^-(\mathcal{S})\},\forall v_i\in\mathcal{M} \quad \text{（公式三）}$$

其中$d_i$，$d_i^+$，$d_i^-$分别表示节点自身的度、节点的出度和入度；

从节点子集$\mathcal{A}$通过中间账户子集$\mathcal{M}$转账到另一个子集$\mathcal{C}$的资金流或者信息流的异常值为：

$$g(\mathcal{S})=\frac{1}{|\mathcal{S}|}\sum_{v_i\in\mathcal{M}}f_i(\mathcal{S})-\lambda\left(q_i(\mathcal{S})-f_i(\mathcal{S})\right) \quad \text{（公式四）}$$

$$=\frac{1}{|\mathcal{S}|}\sum_{v_i\in\mathcal{M}}(1+\lambda)f_i(\mathcal{S})-\lambda\,q_i(\mathcal{S}) \quad \text{（公式五）}$$

其中λ≥0，且λ为资金转入转出不平衡的损失，$f_i(\mathcal{S}),q_i(\mathcal{S})$表示节点的出度和入度的最小值和最大值，$\forall v_i\in\mathcal{M}$，公式二中的$f_i(\mathcal{S})$是从源账户子集$\mathcal{A}$到目的账户子集$\mathcal{C}$所能够通过中间账户$v_i\in\mathcal{M}$的最大流量，$q_i(\mathcal{S})-f_i(\mathcal{S})$为完成转账后$v_i$节点中的账户余额或者转发信息后的权重，异常度量$g(\mathcal{S})$为子集$\mathcal{S}$中的每个账户在洗钱的过程获得的利润或者转发获得的收益。

2.如权利要求1所述的稠密多部子图的检测方法，其特征在于，该步骤3包括：

步骤31、为$\mathcal{S}$中的所有节点构建优先级树$\mathcal{T}$来寻找最大化公式四中目标函数的$g(\mathcal{S})$对应的子集，以更新替换$\mathcal{S}$，定义分配给节点$v_i$的权重为：

$$w_i(\mathcal{S}) = \begin{cases} f_i(\mathcal{S}) - \frac{\lambda}{1+\lambda} q_i(\mathcal{S}), & if\ v_i \in \mathcal{M} \\ d_i(\mathcal{S}), & if\ v_i \in \mathcal{A} \cup \mathcal{C} \end{cases}$$ （公式六）

其中$d_i(\mathcal{S})$为节点本身的出度或者入度；

步骤32、从子集$\mathcal{S} = \mathcal{A} \cup \mathcal{M} \cup \mathcal{C}$开始，从优先级树$\mathcal{T}$中权重最小的节点v，在集合$\mathcal{S}$中删除对应的节点v，在优先级树中更新以v为邻居节点的权重$w_i$，根据公式四或者公式五得到$g(\mathcal{S})$；

步骤33、重复步骤31和32，直到$\mathcal{A}$、$\mathcal{M}$、$\mathcal{C}$中至少有一个为空，输出$g(\mathcal{S})$最大时的集合$\mathcal{S}$，作为该异常行为检测结果。

3.一种基于稠密多部子图的检测系统，其特征在于，包括：

模块1、根据链式特征中的信息流动，构建交易网络的多部图，根据预设的账户间信息流动阈值筛选该多部图，得到该多部图中的稠密子图；

模块2、以固定账户存在超阈值的信息流且在中间账户中保留低于阈值的权重为约束条件，生成该稠密子图中节点子集的异常值；

模块3、根据该异常值，输出该多部图中存在异常行为的节点子集作为异常行为检测结果；

模块1中该多部图$\mathcal{G} = (\mathcal{V}, \mathcal{E})$，$\mathcal{V} = \mathcal{X} \cup \mathcal{W} \cup \mathcal{Y}$，其中$\mathcal{W}$是银行或者消息炒作团体的内部账户的集合，$\mathcal{X}$和$\mathcal{Y}$是银行或者消息炒作团体的外部的集合，其中$\mathcal{X}$是对银行净转入的账户集合或者是发布消息的团体集合，$\mathcal{Y}$是接受银行净转出的账户集合或者最终转发消息的集合，多部图中的边对应于从$\mathcal{X}$到$\mathcal{W}$以及从$\mathcal{W}$到$\mathcal{Y}$的权重，对于$i,j \in \mathcal{V}$，边(i，j)∈ε表示账户i将钱或者信息转给j；该稠密子图的节点子集$\mathcal{S} = \mathcal{A} \cup \mathcal{M} \cup \mathcal{C}$，其中$\mathcal{A} \subseteq \mathcal{X}$，$\mathcal{M} \subseteq \mathcal{W}$，$\mathcal{C} \subseteq \mathcal{Y}$；

该模块2具体为：

根据该节点子集中为从$v_i$到$v_j$的总权重$e_{ij}$，权重使用转账金额或者消息转发的数量来衡量，得到节点$v_i \in \mathcal{M}$关于节点子集$\mathcal{S}$的总入度$d_i^-$和出度值$d_i^+$：

$$d_i^+(\mathcal{S}) = \sum_{v_j \in \mathcal{C} \wedge (i,j) \in \varepsilon} e_{ij}, d_i^-(\mathcal{S}) = \sum_{v_k \in \mathcal{A} \wedge (k,i) \in \varepsilon} e_{ki}$$ （公式一）

得到一个中间账户关于节点子集$\mathcal{S}$的总加权出度和入度的最小值和最大值：

$$f_i(\mathcal{S}) = min\{d_i^+(\mathcal{S}), d_i^-(\mathcal{S})\}, \forall v_i \in \mathcal{M}$$ （公式二）

$$q_i(\mathcal{S}) = max\{d_i^+(\mathcal{S}), d_i^-(\mathcal{S})\}, \forall v_i \in \mathcal{M}$$ （公式三）

其中$d_i$，$d_i^+$，$d_i^-$分别表示节点自身的度、节点的出度和入度；

从节点子集$\mathcal{A}$通过中间账户子集$\mathcal{M}$转账到另一个子集$\mathcal{C}$的资金流或者信息流的异常值为：

$$g(\mathcal{S}) = \frac{1}{|\mathcal{S}|} \sum_{v_i \in \mathcal{M}} f_i(\mathcal{S}) - \lambda(q_i(\mathcal{S}) - f_i(\mathcal{S}))$$ （公式四）

$$= \frac{1}{|\mathcal{S}|} \sum_{v_i \in \mathcal{M}} (1 + \lambda) f_i(\mathcal{S}) - \lambda q_i(\mathcal{S})$$ （公式五）

其中λ≥0，且λ为资金转入转出不平衡的损失，$f_i(\mathcal{S}), q_i(\mathcal{S})$表示节点的出度和入度的最小

值和最大值，$\forall v_i \in \mathcal{M}$，公式二中的$f_i(\mathcal{S})$是从源账户子集$\mathcal{A}$到目的账户子集$\mathcal{C}$所能够通过中间账户$v_i \in \mathcal{M}$的最大流量，$q_i(\mathcal{S}) - f_i(\mathcal{S})$为完成转账后$v_i$节点中的账户余额或者转发信息后的权重，异常度量$g(\mathcal{S})$为子集$\mathcal{S}$中的每个账户在洗钱的过程获得的利润或者转发获得的收益。

4.如权利要求3所述的稠密多部子图的检测系统，其特征在于，该模块3包括：

模块31、为$\mathcal{S}$中的所有节点构建优先级树$\mathcal{T}$来寻找最大化公式四中目标函数的$g(\mathcal{S})$对应的子集，以更新替换$\mathcal{S}$，定义分配给节点$v_i$的权重为：

$$w_i(\mathcal{S}) = \begin{cases} f_i(\mathcal{S}) - \frac{\lambda}{1+\lambda} q_i(\mathcal{S}), & if\ v_i \in \mathcal{M} \\ d_i(\mathcal{S}), & if\ v_i \in \mathcal{A} \cup \mathcal{C} \end{cases} \quad \text{（公式六）}$$

其中$d_i(\mathcal{S})$为节点本身的出度或者入度；

模块32、从子集$\mathcal{S} = \mathcal{A} \cup \mathcal{M} \cup \mathcal{C}$开始，从优先级树$\mathcal{T}$中权重最小的节点v，在集合$\mathcal{S}$中删除对应的节点v，在优先级树中更新以v为邻居节点的权重$w_i$，根据公式四或者公式五得到$g(\mathcal{S})$；

模块33、重复模块31和32，直到$\mathcal{A}$、$\mathcal{M}$、$\mathcal{C}$中至少有一个为空，输出$g(\mathcal{S})$最大时的集合$\mathcal{S}$，作为该异常行为检测结果。

# 一种稠密多部子图的检测方法及系统

## 技术领域

[0001] 本发明涉及数据挖掘领域,特别涉及一种稠密多部子图的检测方法及系统。

## 背景技术

[0002] 随着互联网的发展,互联网应用获得了飞速发展,社交媒体也获得了飞速发展,目前社交媒体已超越搜索引擎,成为互联网第一大流量来源,二者占比分别为46%和40%。而随着技术的发展,恶意话题操作以及洗钱等也成为了不法分子的谋取暴力的工具。恶意话题炒作是一些团伙通过相互转发信息来炒作某个话题,从而获得舆论影响力、宣传推广等目的;互洗钱是将从非法来源获得的资金转化为合法资金的过程。图成为一种常见数据应用到许多科学和工程中,图可以表示成这样一种结构,即图G=(V,E)是一对集合:一组顶点V表示实体和一组边E表示实体之间的关系或连接。在计算机科学中,网络包含节点和边缘;而在社会科学中,相应的术语则是行为者和关系,在本文中这两个术语具有同等意义。如果用图中的顶点表示参与活动的人,用边表示消息或者人与人之间的关联。那么当发起媒体炒作或者洗钱行为时,会在特定时间或者特定常见下,参与活动的人之间产生一个多部稠密子图。用户之间的相互关注关系就构成了社交网络图,其常见的存储格式是邻接矩阵或者拉普拉斯矩阵,邻接矩阵如图1所示,当两个节点之间右边相连时,对应的位置置为1,如果两个节点之间无边相连,对应的位置置为0,如果是有权图,对应位置置为权值,无边相连对应位置置为极大的值,但是其特征几乎相同。

[0003] 图2显示了一个洗钱的案例,包含从源账户到中间账户到目标账户的两步资金转移流程。为了隐藏资金的真实来源和去向,洗钱者往往通过多层中间账户(可以是银行内或银行间转账)隐秘的将脏钱从源账户转移到目的账户。银行转账日志中只记录了自己银行中的账户的交易记录,包括从外部账户转入、从银行账户转出和银行账户之间的转账记录。由于不太可能从每个银行得到转账日志,因此洗钱检测问题通常集中在如何利用单个银行的转账日志上。实际上,现有的度量和算法通常足以使用单个银行的交易记录或者来自多个银行的联合交易记录来进行洗钱检测。下文使用“银行”来代指进行洗钱检测的银行或者多个银行集合。一般而言,洗钱流程包括三步:a)在银行开户若干中间账户;b)从其他银行开户的源账户转账到中间账户;c)将钱从中间账户汇集到若干不同的目的地账户。洗钱行为具有两个主要特征,第一个特征是密集转账。洗钱者创建了一个密集的高容量转账子图,无论是在资金流入银行还是流出银行的时候。这是因为欺诈账户的数量有限,并且需要在短时间内将大量资金转入银行并转出银行,从而产生了密集的高容量转账子图。第二个特征是中间账户的账户余额基本为零。中间账户在洗钱过程中充当了资金桥的作用:大部分流入的资金都将被转出,从而使得流入流出资金基本相等,账户余额为零。这是因为洗钱者留在中间账户中的钱会有被检测和冻结的风险。因此,欺诈者往往在中间账户中留下尽可能少的钱。

[0004] 当前对于多部稠密子图检测的方法包括:

[0005] 第一是是基于规则的分类。这些规则基于本体的专家系统来检测可疑交易;使用

基于规则设计的贝叶斯网络来评估客户的交易行为的风险指数。

[0006] 第二是基于机器学习算法来检测。这些方法包括SVM、决策树、RBF神经网络等。

[0007] 第三是通用的基于图的异常检测算法。这些检测方法主要基于图来检测洗钱行为。具体包括研究特征向量中的模式，基于消息传播以及基于稠密子图等。

[0008] 以上这些方法都没有捕捉洗钱行为中的异常信息或者活动链，也不提供理论保证，更容易被犯罪分子攻击，同时受到类不平衡问题的影响，适应性有限。此外，大多数现有的检测方法忽略了这些行为中的链式特征，也忽略了账户之间的复杂依赖关系，导致较低的检测准确率，容易被犯罪分子规避。稠密子图和密度子张量检测算法已被应用于图欺诈检测，但这些算法只考虑了一跳交易上的密度。尽管可以处理链式交易，但它需要大量的真实标记数据来完成模型的训练，而这种标记数据很少，并且使用特定标记数据可能使模型发生过拟合而降低了鲁棒性。

## 发明内容

[0009] 针对现有技术的不足，本发明提出一种基于稠密多部子图的检测方法，包括：

[0010] 步骤1、根据链式特征中的信息流动，构建交易网络的多部图，根据预设的账户间信息流动阈值筛选该多部图，得到该多部图中的稠密子图；

[0011] 步骤2、以固定账户存在超阈值的信息流且在中间账户中保留低于阈值的权重为约束条件，生成该稠密子图中节点子集的异常值；

[0012] 步骤3、根据该异常值，输出该多部图中存在异常行为的节点子集作为异常行为检测结果；

[0013] 其中，步骤1中该多部图$\mathcal{G}=(\mathcal{V},\mathcal{E})$，$\mathcal{V}=\mathcal{X}\cup\mathcal{W}\cup\mathcal{Y}$，其中$\mathcal{W}$是银行或者消息炒作团体的内部账户的集合，$\mathcal{X}$和$\mathcal{Y}$是银行或者消息炒作团体的外部的集合，其中$\mathcal{X}$是对银行净转入的账户集合或者是发布消息的团体集合，$\mathcal{Y}$是接受银行净转出的账户集合或者最终转发消息的集合，多部图中的边对应于从$\mathcal{X}$到$\mathcal{W}$以及从$\mathcal{W}$到$\mathcal{Y}$的权重，对于$i,j\in\mathcal{V}$，边(i，j)∈ε表示账户i将钱或者信息转给j；该稠密子图的节点子集$\mathcal{S}=\mathcal{A}\cup\mathcal{M}\cup\mathcal{C}$，其中$\mathcal{A}\subseteq\mathcal{X}$，$\mathcal{M}\subseteq\mathcal{W}$，$\mathcal{C}\subseteq\mathcal{Y}$；

[0014] 该步骤2具体为：

[0015] 根据该节点子集中为从$v_i$到$v_j$的总权重$e_{ij}$，权重使用转账金额或者消息转发的数量来衡量，得到节点$v_i\in\mathcal{M}$关于节点子集$\mathcal{S}$的总入度$d_i^-$和出度值$d_i^+$：

[0016] $$d_i^+(\mathcal{S})=\sum_{v_j\in\mathcal{C}\wedge(i,j)\in\mathcal{E}}e_{ij},\ d_i^-(\mathcal{S})=\sum_{v_k\in\mathcal{A}\wedge(k,i)\in\mathcal{E}}e_{ki} \quad \text{（公式一）}$$

[0017] 得到一个中间账户关于节点子集$\mathcal{S}$的总加权出度和入度的最小值和最大值：

[0018] $$f_i(\mathcal{S})=min\{d_i^+(\mathcal{S}),d_i^-(\mathcal{S})\},\forall v_i\in\mathcal{M} \quad \text{（公式二）}$$

[0019] $$q_i(\mathcal{S})=max\{d_i^+(\mathcal{S}),d_i^-(\mathcal{S})\},\forall v_i\in\mathcal{M} \quad \text{（公式三）}$$

[0020] 其中$d_i$，$d_i^+$，$d_i^-$分别表示节点自身的度、节点的出度和入度；

[0021] 从节点子集$\mathcal{A}$通过中间账户子集$\mathcal{M}$转账到另一个子集$\mathcal{C}$的资金流或者信息流的异

常值为：

[0022] $$g(\mathcal{S}) = \frac{1}{|\mathcal{S}|}\sum_{v_i\in\mathcal{M}} f_i(\mathcal{S}) - \lambda\left(q_i(\mathcal{S}) - f_i(\mathcal{S})\right)$$ （公式四）

$$= \frac{1}{|\mathcal{S}|}\sum_{v_i\in\mathcal{M}} (1+\lambda) f_i(\mathcal{S}) - \lambda q_i(\mathcal{S})$$ （公式五）

[0023] 其中λ≥0，且λ为资金转入转出不平衡的损失，$f_i(\mathcal{S}), q_i(\mathcal{S})$表示节点的出度和入度的最小值和最大值，$\forall v_i \in \mathcal{M}$，公式二中的$f_i(\mathcal{S})$是从源账户子集$\mathcal{A}$到目的账户子集$\mathcal{C}$所能够通过中间账户$v_i \in \mathcal{M}$的最大流量，$q_i(\mathcal{S}) - f_i(\mathcal{S})$为完成转账后$v_i$节点中的账户余额或者转发信息后的权重，异常度量$g(\mathcal{S})$为子集$\mathcal{S}$中的每个账户在洗钱的过程获得的利润或者转发获得的收益。

[0024] 所述的稠密多部子图的检测方法，该步骤3包括：

[0025] 步骤31、为$\mathcal{S}$中的所有节点构建优先级树$\mathcal{T}$来寻找最大化公式四中目标函数的$g(\mathcal{S})$对应的子集，以更新替换$\mathcal{S}$，定义分配给节点$v_i$的权重为：

[0026] $$w_i(\mathcal{S}) = \begin{cases} f_i(\mathcal{S}) - \frac{\lambda}{1+\lambda} q_i(\mathcal{S}), & if\ v_i \in \mathcal{M} \\ d_i(\mathcal{S}), & if\ v_i \in \mathcal{A}\cup\mathcal{C} \end{cases}$$ （公式六）

[0027] 其中$d_i(\mathcal{S})$为节点本身的出度或者入度；

[0028] 步骤32、从子集$\mathcal{S} = \mathcal{A}\cup\mathcal{M}\cup\mathcal{C}$开始，从优先级树$\mathcal{T}$中权重最小的节点v，在集合$\mathcal{S}$中删除对应的节点v，在优先级树中更新以v为邻居节点的权重$w_i$，根据公式四或者公式五得到$g(\mathcal{S})$；

[0029] 步骤33、重复步骤31和32，直到$\mathcal{A}$、$\mathcal{M}$、$\mathcal{C}$中至少有一个为空，输出$g(\mathcal{S})$最大时的集合$\mathcal{S}$，作为该异常行为检测结果。

[0030] 本发明还提出了一种基于稠密多部子图的检测系统，包括：

[0031] 模块1、根据链式特征中的信息流动，构建交易网络的多部图，根据预设的账户间信息流动阈值筛选该多部图，得到该多部图中的稠密子图；

[0032] 模块2、以固定账户存在超阈值的信息流且在中间账户中保留低于阈值的权重为约束条件，生成该稠密子图中节点子集的异常值；

[0033] 模块3、根据该异常值，输出该多部图中存在异常行为的节点子集作为异常行为检测结果；

[0034] 模块1中该多部图$\mathcal{G} = (\mathcal{V}, \mathcal{E})$，$\mathcal{V} = \mathcal{X}\cup\mathcal{W}\cup\mathcal{Y}$，其中$\mathcal{W}$是银行或者消息炒作团体的内部账户的集合，$\mathcal{X}$和$\mathcal{Y}$是银行或者消息炒作团体的外部的集合，其中$\mathcal{X}$是对银行净转入的账户集合或者是发布消息的团体集合，$\mathcal{Y}$是接受银行净转出的账户集合或者最终转发消息的集合，多部图中的边对应于从$\mathcal{X}$到$\mathcal{W}$以及从$\mathcal{W}$到$\mathcal{Y}$的权重，对于$i,j \in \mathcal{V}$，边(i，j)∈ε表示账户i将钱或者信息转给j；该稠密子图的节点子集$\mathcal{S} = \mathcal{A}\cup\mathcal{M}\cup\mathcal{C}$，其中

$\mathcal{A}\subseteq\mathcal{X}$，$\mathcal{M}\subseteq\mathcal{W}$，$\mathcal{C}\subseteq\mathcal{Y}$；

[0035] 该模块2具体为：

[0036] 根据该节点子集中为从$v_i$到$v_j$的总权重$e_{ij}$，权重使用转账金额或者消息转发的数

量来衡量，得到节点$v_i \in \mathcal{M}$关于节点子集$\mathcal{S}$的总入度$d_i^-$和出度值$d_i^+$：

[0037] $$d_i^+(\mathcal{S}) = \sum_{v_j \in \mathcal{C} \wedge (i,j) \in \varepsilon} e_{ij}, d_i^-(\mathcal{S}) = \sum_{v_k \in \mathcal{A} \wedge (k,i) \in \varepsilon} e_{ki} \quad \text{（公式一）}$$

[0038] 得到一个中间账户关于节点子集$\mathcal{S}$的总加权出度和入度的最小值和最大值：

[0039] $$f_i(\mathcal{S}) = min\{d_i^+(\mathcal{S}), d_i^-(\mathcal{S})\}, \forall v_i \in \mathcal{M} \quad \text{（公式二）}$$

[0040] $$q_i(\mathcal{S}) = max\{d_i^+(\mathcal{S}), d_i^-(\mathcal{S})\}, \forall v_i \in \mathcal{M} \quad \text{（公式三）}$$

[0041] 其中$d_i, d_i^+, d_i^-$分别表示节点自身的度、节点的出度和入度；

[0042] 从节点子集$\mathcal{A}$通过中间账户子集$\mathcal{M}$转账到另一个子集$\mathcal{C}$的资金流或者信息流的异常值为：

[0043] $$g(\mathcal{S}) = \frac{1}{|\mathcal{S}|} \sum_{v_i \in \mathcal{M}} f_i(\mathcal{S}) - \lambda\left(q_i(\mathcal{S}) - f_i(\mathcal{S})\right) \quad \text{（公式四）}$$

$$= \frac{1}{|\mathcal{S}|} \sum_{v_i \in \mathcal{M}} \left(1 + \lambda\right) f_i(\mathcal{S}) - \lambda\, q_i(\mathcal{S}) \quad \text{（公式五）}$$

[0044] 其中$\lambda \geq 01$，且$\lambda$为资金转入转出不平衡的损失，$f_i(\mathcal{S}), q_i(\mathcal{S})$表示节点的出度和入度的最小值和最大值，$\forall v_i \in \mathcal{M}$，公式二中的$f_i(\mathcal{S})$是从源账户子集$\mathcal{A}$到目的账户子集$\mathcal{C}$所能够通过中间账户$v_i \in \mathcal{M}$的最大流量，$q_i(\mathcal{S}) - f_i(\mathcal{S})$为完成转账后$v_i$节点中的账户余额或者转发信息后的权重，异常度量$g(\mathcal{S})$为子集$\mathcal{S}$中的每个账户在洗钱的过程获得的利润或者转发获得的收益。

[0045] 所述的稠密多部子图的检测系统，该模块3包括：

[0046] 模块31、为$\mathcal{S}$中的所有节点构建优先级树$\mathcal{T}$来寻找最大化公式四中目标函数的$g(\mathcal{S})$对应的子集，以更新替换$\mathcal{S}$，定义分配给节点$v_i$的权重为：

[0047] $$w_i(\mathcal{S}) = \begin{cases} f_i(\mathcal{S}) - \frac{\lambda}{1+\lambda} q_i(\mathcal{S}), & if\ v_i \in \mathcal{M} \\ d_i(\mathcal{S}), & if\ v_i \in \mathcal{A} \cup \mathcal{C} \end{cases} \quad \text{（公式六）}$$

[0048] 其中$d_i(\mathcal{S})$为节点本身的出度或者入度；

[0049] 模块32、从子集$\mathcal{S} = \mathcal{A} \cup \mathcal{M} \cup \mathcal{C}$开始，从优先级树$\mathcal{T}$中权重最小的节点v，在集合$\mathcal{S}$中删除对应的节点v，在优先级树中更新以v为邻居节点的权重$w_i$，根据公式四或者公式五得到$g(\mathcal{S})$；

[0050] 模块33、重复模块31和32，直到$\mathcal{A}$、$\mathcal{M}$、$\mathcal{C}$中至少有一个为空，输出$g(\mathcal{S})$最大时的集合$\mathcal{S}$，作为该异常行为检测结果。

[0051] 本发明与现有技术相比的优点在于：

[0052] (1)本发明提出了多部子图行为的新异常度量：本发明提出了一种新的度量来检测密集的多步流量异常，并验证了其检测多部子图的有效性；同时能够提供理论保证。

[0053] (2)本发明具有有效性和鲁棒性：本发明在各种多部图的拓扑结构下的性能优于最先进的算法，当使用更多欺诈性账户，更长的传输链时，本发明仍然可以有效检测对抗性的异常行为。

[0054] (3)本发明具有良好的可扩展性：本发明的算法复杂度和图中边的数量成近似线

性的关系(即转账记录数),有良好的可扩展性,因此非常适合银行快速增长的业务。

## 附图说明

[0055] 图1是无权图的邻接矩阵图。

[0056] 图2是洗钱的案例示意图。

[0057] 图3是系统工作流程图。

## 具体实施方式

[0058] 为让本发明的上述特征和效果能阐述的更明确易懂,下文特举实施例,并配合说明书附图作详细说明如下。

[0059] 本文所提出的一种稠密多部子图的检测方法和装置。在应用到洗钱行为检测时,具体包括:我们使用多部图来建模银行中的资金流动,并为转账行为定义新的异常度量。较高的度量值表明通过图中存在通过若干固定账户的大量资金流,而且不会在中间账户中留下太多资金。相反,正常(即诚实)账户并不总是一致地将钱转移到特定账户,也不会立即清空中间账户的余额,因此具有较低的度量值。本方法通过优化所设计的度量值来搜索欺诈性账户,同时近似贪心的优化源,中间和目的账户的子集。此外,本方法为检测结果的近似最优性提供理论保证,给出了欺诈者可以在不被检测到洗钱行为的情况下能够转移金额的上限。

[0060] 本发明具体实施例如下:

[0061] (1)本发明中关于实施实例中图的定义。用图$\mathcal{G}=(\mathcal{V},\mathcal{E})$是表示交易网络的三部图。定义$\mathcal{V}=\mathcal{X}\cup\mathcal{W}\cup\mathcal{Y}$,其中$\mathcal{W}$是银行的内部账户的集合,$\mathcal{X}$和$\mathcal{Y}$是银行外部账户的集合,其中$\mathcal{X}$是对银行净转入的账户集合,$\mathcal{Y}$是接受银行净转出的账户集合。多部图中的边对应于从$\mathcal{X}$到$\mathcal{W}$以及从$\mathcal{W}$到$\mathcal{Y}$的资金转账。对于$i,j\in\mathcal{V}$,边(i,j)∈ε表示账户i将钱转给j。由于许多转账可能发生在一条边上,因此每条边可以代表多次转账。$e_{ij}$是从$v_i$到$v_j$的转账总金额。

[0062] (2)少数账户中的大量资金流动构成了这个三部图中的一个稠密子图,本方法要评估由节点子集$\mathcal{S}=\mathcal{A}\cup\mathcal{M}\cup\mathcal{C}$构成的稠密子图的异常值,其中$\mathcal{A}\subseteq\mathcal{X},\mathcal{M}\subseteq\mathcal{W},\ \mathcal{C}\subseteq\mathcal{Y}$)以便检测大量的资金转移行为。

[0063] 定义$e_{ij}$为从$v_i$到$v_j$的总转账金额,并定义节点$v_i\in\mathcal{M}$关于节点子集$\mathcal{S}$的总(加权)入度和出度值:

[0064] $$d_i^+(\mathcal{S})=\sum_{v_j\in\mathcal{C}\wedge(i,j)\in\varepsilon}e_{ij},d_i^-(\mathcal{S})=\sum_{v_k\in\mathcal{A}\wedge(k,i)\in\varepsilon}e_{ki}\qquad(公式一)$$

[0065] 定义一个中间账户关于节点子集$\mathcal{S}$的总加权出度和入度的最小值和最大值:

[0066] $$f_i(\mathcal{S})=min\{d_i^+(\mathcal{S}),d_i^-(\mathcal{S})\},\forall v_i\in\mathcal{M}\qquad(公式二)$$

[0067] $$q_i(\mathcal{S})=max\{d_i^+(\mathcal{S}),d_i^-(\mathcal{S})\},\forall v_i\in\mathcal{M}\qquad(公式三)$$

[0068] 其中$d_i$,$d_i^+$,$d_i^-$分别表示节点自身的度,节点的出度和入度。

[0069] 定义洗钱的异常值:从节点子集$\mathcal{A}$通过中间账户子集$\mathcal{M}$转账到另一个子集$\mathcal{C}$的资金流的异常值是:

[0070] $$g(\mathcal{S}) = \frac{1}{|\mathcal{S}|}\sum_{v_i\in\mathcal{M}} f_i(\mathcal{S}) - \lambda\left(q_i(\mathcal{S}) - f_i(\mathcal{S})\right) \quad \text{（公式四）}$$

$$= \frac{1}{|\mathcal{S}|}\sum_{v_i\in\mathcal{M}} (1+\lambda) f_i(\mathcal{S}) - \lambda\, q_i(\mathcal{S}) \quad \text{（公式五）}$$

[0071] 其中λ≥0是常系数，将λ定义为资金转入转出不平衡的损失，用于量化洗钱者因单位的盈余或赤字（伪装成本）而遭受的损失程度，可以通过经验得到或者有专家给出。$f_i(\mathcal{S}), q_i(\mathcal{S})$表示节点的出度和入度的最小值和最大值，$\forall v_i \in \mathcal{M}$。公式二中的$f_i(\mathcal{S})$是从源账户子集$\mathcal{A}$到目的账户子集$\mathcal{C}$所能够通过中间账户$v_i \in \mathcal{M}$的最大可能流量。$q_i(\mathcal{S}) - f_i(\mathcal{S})$为完成转账后$v_i$节点中的“账户余额”，可以看做是洗钱的损耗，因为洗钱者更希望将中间账户的账户余额清零。“剩余金额”是中间账户的盈余或赤字（即通过和子集$\mathcal{A}$，$\mathcal{C}$之外的账户交易所产生的），这些“剩余金额”可以看作洗钱者为了逃避检测的所进行的对抗伪装行为。λ的解释和我们的度量：我们将λ定义为资金转入转出不平衡的损失，这是一个常数系数，用于量化洗钱者因单位的盈余或赤字（伪装成本）而遭受的损失程度。现在我们的异常度量$g(\mathcal{S})$可以被解释为子集$\mathcal{S}$中的每个账户可以在洗钱的过程获得的利润（收益减去成本）。

[0072] (3)本方法提出了一种近似贪婪的算法，通过为$\mathcal{S}$中的所有节点构建优先级树来寻找最大化公式(4)中目标函数的对$g(\mathcal{S})$应的子集。定义分$\mathcal{S}$配给节点$v_i$的权重（即优先级）为：

[0073] $$w_i(\mathcal{S}) = \begin{cases} f_i(\mathcal{S}) - \frac{\lambda}{1+\lambda} q_i(\mathcal{S}), & if\ v_i \in \mathcal{M} \\ d_i(\mathcal{S}), & if\ v_i \in \mathcal{A}\cup\mathcal{C} \end{cases} \quad \text{（公式六）}$$

[0074] 其中$d_i(\mathcal{S})$节点本身的出度或者入度。我们还可以将节点的先验异常值添加到权重$w_i(\mathcal{S})$中。

[0075] (4)本发明的输入是一个三部图，用图$\mathcal{G} = (\mathcal{V}, \mathcal{E})$表示交易网络的三部图。定义$\mathcal{V} = \mathcal{X}\cup\mathcal{W}\cup\mathcal{Y}$，其中$\mathcal{W}$是银行的内部账户的集合，$\mathcal{X}$和$\mathcal{Y}$是银行外部账户的集合，其中$\mathcal{X}$是对银行净转入的账户集合，$\mathcal{Y}$是接受银行净转出的账户集合。多部图中的边对应于从$\mathcal{X}$到$\mathcal{W}$以及从$\mathcal{W}$到$\mathcal{Y}$的资金转账。本发明的输出是最有可能涉嫌洗钱的节点子集。

[0076] (5)少数账户中的大量资金流动构成了这个三部图中的一个稠密子图，在本小节中给出本方法所使用的定义。本方法要评估由节点子集$\mathcal{S} = \mathcal{A}\cup\mathcal{M}\cup\mathcal{C}$构成的稠密子图的异常值，其中$\mathcal{A}\subseteq\mathcal{X}$，$\mathcal{M}\subseteq\mathcal{W}$，$\mathcal{C}\subseteq\mathcal{Y}$）以便检测大量的资金转移行为。

[0077] (6)根据公式六，计算图中节点的权重$w_i$，根据节点的权重构造优先级树$\mathcal{T}$.

[0078] (7)算法从子集$\mathcal{S} = \mathcal{A}\cup\mathcal{M}\cup\mathcal{C}$开始，从优先级树$\mathcal{T}$中权重最小的节点v，在集合$\mathcal{S}$中删除对应的节点v，在优先级树中更新v以为邻居节点的权重$w_i$，根据公式四或者公式五计算$g(\mathcal{S})$。

[0079] (8)重复步骤(6)和(7)，直到$\mathcal{A}$、$\mathcal{M}$、$\mathcal{C}$中最少有一个为空。

[0080] (9)得到使$g(\mathcal{S})$最大的集合$\mathcal{S}$，本方法结束。在步骤(7)中删除使优先树权重最小的节点，剩下的节点就可以保证$g(\mathcal{S})$最大。

[0081] 以下为与上述方法实施例对应的系统实施例，本实施方式可与上述实施方式互相配合实施。上述实施方式中提到的相关技术细节在本实施方式中依然有效，为了减少重复，这里不再赘述。相应地，本实施方式中提到的相关技术细节也可应用在上述实施方式中。

[0082] 本发明还提出了一种基于稠密多部子图的检测系统，包括：

[0083] 模块1、根据链式特征中的信息流动，构建交易网络的多部图，根据预设的账户间信息流动阈值筛选该多部图，得到该多部图中的稠密子图；

[0084] 模块2、以固定账户存在超阈值的信息流且在中间账户中保留低于阈值的权重为约束条件，生成该稠密子图中节点子集的异常值；

[0085] 模块3、根据该异常值，输出该多部图中存在异常行为的节点子集作为异常行为检测结果；

[0086] 模块1中该多部图$\mathcal{G}=(\mathcal{V},\mathcal{E})$，$\mathcal{V}=\mathcal{X}\cup\mathcal{W}\cup\mathcal{Y}$，其中$\mathcal{W}$是银行或者消息炒作团体的内部账户的集合，$\mathcal{X}$和$\mathcal{Y}$是银行或者消息炒作团体的外部的集合，其中$\mathcal{X}$是对银行净转入的账户集合或者是发布消息的团体集合，$\mathcal{Y}$是接受银行净转出的账户集合或者最终转发消息的集合，多部图中的边对应于从$\mathcal{X}$到$\mathcal{W}$以及从$\mathcal{W}$到$\mathcal{Y}$的权重，对于$i,j\in\mathcal{V}$，边(i，j)∈ε表示账户i将钱或者信息转给j；该稠密子图的节点子集$\mathcal{S}=\mathcal{A}\cup\mathcal{M}\cup\mathcal{C}$，其中

$\mathcal{A}\subseteq\mathcal{X}$，$\mathcal{M}\subseteq\mathcal{W}$，$\mathcal{C}\subseteq\mathcal{Y}$；

[0087] 该模块2具体为：

[0088] 根据该节点子集中为从$v_i$到$v_j$的总权重$e_{ij}$，权重使用转账金额或者消息转发的数量来衡量，得到节点$v_i\in\mathcal{M}$关于节点子集$\mathcal{S}$的总入度$d_i^-$和出度值$d_i^+$：

[0089] $$d_i^+(\mathcal{S})=\sum_{v_j\in\mathcal{C}\wedge(i,j)\in\varepsilon}e_{ij},\ d_i^-(\mathcal{S})=\sum_{v_k\in\mathcal{A}\wedge(k,i)\in\varepsilon}e_{ki}\qquad\text{（公式一）}$$

[0090] 得到一个中间账户关于节点子集$\mathcal{S}$的总加权出度和入度的最小值和最大值：

[0091] $$f_i(\mathcal{S})=min\{d_i^+(\mathcal{S}),d_i^-(\mathcal{S})\},\forall v_i\in\mathcal{M}\qquad\text{（公式二）}$$

[0092] $$q_i(\mathcal{S})=max\{d_i^+(\mathcal{S}),d_i^-(\mathcal{S})\},\forall v_i\in\mathcal{M}\qquad\text{（公式三）}$$

[0093] 其中$d_i$，$d_i^+$，$d_i^-$分别表示节点自身的度、节点的出度和入度；

[0094] 从节点子集$\mathcal{A}$通过中间账户子集$\mathcal{M}$转账到另一个子集$\mathcal{C}$的资金流或者信息流的异常值为：

[0095] $$g(\mathcal{S})=\frac{1}{|\mathcal{S}|}\sum_{v_i\in\mathcal{M}}f_i(\mathcal{S})-\lambda\left(q_i(\mathcal{S})-f_i(\mathcal{S})\right)\qquad\text{（公式四）}$$

$$=\frac{1}{|\mathcal{S}|}\sum_{v_i\in\mathcal{M}}\left(1+\lambda\right)f_i(\mathcal{S})-\lambda\,q_i(\mathcal{S})\qquad\text{（公式五）}$$

[0096] 其中λ≥0，且λ为资金转入转出不平衡的损失，$f_i(\mathcal{S}),q_i(\mathcal{S})$表示节点的出度和入度的最小值和最大值，$\forall v_i\in\mathcal{M}$，公式二中的$f_i(\mathcal{S})$是从源账户子集$\mathcal{A}$到目的账户子集$\mathcal{C}$所能够通过中间账户$v_i\in\mathcal{M}$的最大流量，$q_i(\mathcal{S})-f_i(\mathcal{S})$为完成转账后$v_i$节点中的账户余额或者转发信息后的权重，异常度量$g(\mathcal{S})$为子集$\mathcal{S}$中的每个账户在洗钱的过程获得的利润或者转发获得的收益。

[0097] 所述的稠密多部子图的检测系统，该模块3包括：

[0098] 模块31、为$\mathcal{S}$中的所有节点构建优先级树$\mathcal{T}$来寻找最大化公式四中目标函数的$g(\mathcal{S})$对应的子集，以更新替换$\mathcal{S}$，定义分配给节点$v_i$的权重为：

[0099] $$w_i(\mathcal{S}) = \begin{cases} f_i(\mathcal{S}) - \frac{\lambda}{1+\lambda} q_i(\mathcal{S}), & if\ v_i \in \mathcal{M} \\ d_i(\mathcal{S}), & if\ v_i \in \mathcal{A} \cup \mathcal{C} \end{cases}$$ （公式六）

[0100] 其中$d_i(\mathcal{S})$为节点本身的出度或者入度；

[0101] 模块32、从子集$\mathcal{S} = \mathcal{A} \cup \mathcal{M} \cup \mathcal{C}$开始，从优先级树$\mathcal{T}$中权重最小的节点v，在集合$\mathcal{S}$中删除对应的节点v，在优先级树中更新以v为邻居节点的权重$w_i$，根据公式四或者公式五得到$g(\mathcal{S})$；

[0102] 模块33、重复模块31和32，直到$\mathcal{A}$、$\mathcal{M}$、$\mathcal{C}$中至少有一个为空，输出$g(\mathcal{S})$最大时的集合$\mathcal{S}$，作为该异常行为检测结果。

[0103] 以上所述，仅为本发明部分具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本领域的人员在本发明揭露的技术范围内，可轻易想到的变化或替换，都应涵盖在本发明的保护范围之内。

| | 用户1 | 用户2 | 用户3 | 用户4 | 用户5 | 用户6 | 用户7 | 用户8 | 用户9 |
|---|---|---|---|---|---|---|---|---|---|
| 用户1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 |
| 用户2 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 用户3 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| 用户4 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 用户5 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 用户6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 用户7 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 用户8 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| 用户9 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |

图1

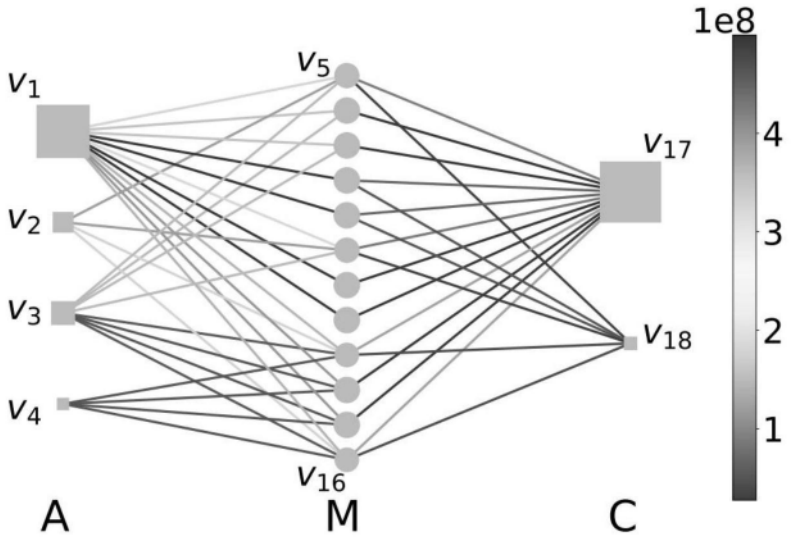



图2

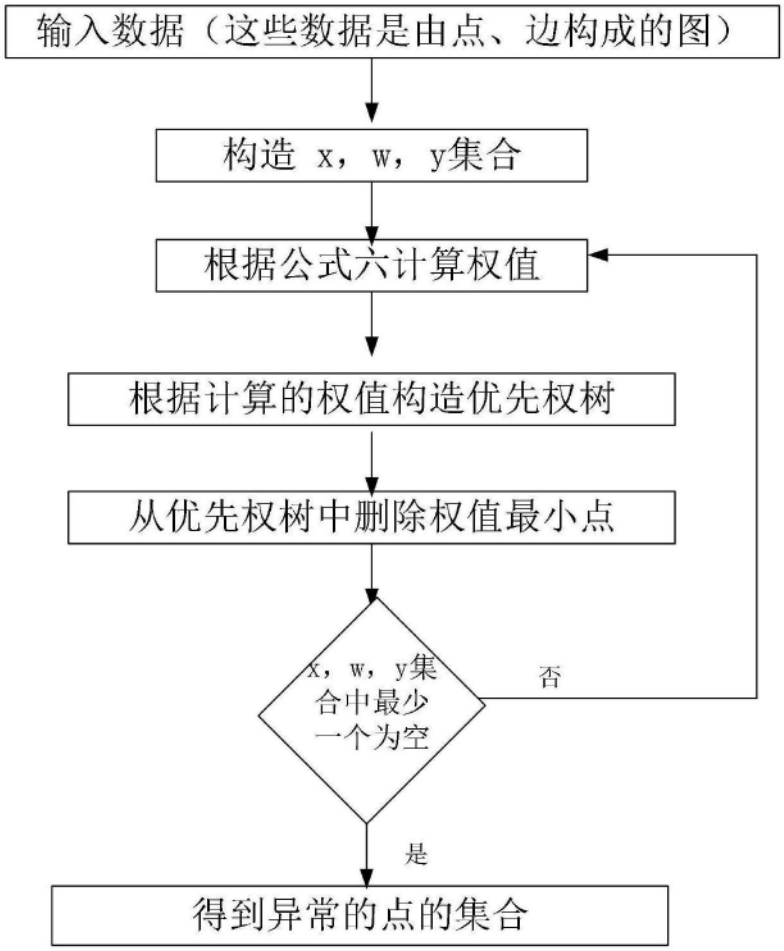

输入数据（这些数据是由点、边构成的图）
构造 x，w，y集合
根据公式六计算权值
根据计算的权值构造优先权树
从优先权树中删除权值最小点
x，w，y集合中最少一个为空
否
是
得到异常的点的集合

图3