FRAUD GANG IDENTIFICATION METHOD AND DEVICE

TECHNICAL FIELD

[0001] The present disclosure relates to the field of Internet technologies, and in particular, to fraud gang identification methods and devices.

BACKGROUND

[0002] In recent years, Internet frauds have become increasingly arrogant, especially gang crimes. Fraudulent criminal gangs can use the Internet platform to attract victims, and commit frauds in various ways. Fraudsters can change identities and register new accounts, or use a plurality of identities and register different accounts, and distribute frauds to different accounts. Consequently, it becomes more difficult to identify fraudulent crimes by using an anti-fraud system. In this background, to effectively prevent and control frauds, a gang identification model for detecting criminal gangs can be developed based on a relational network, so that crime gangs can be identified and cracked down powerfully.

SUMMARY

[0003] In view of this, one or more implementations of the present specification provide a fraud gang identification method and device to improve gang identification precision.

[0004] In particular, one or more implementations of the present specification are implemented by using the following technical solutions:

[0005] According to a first aspect, a fraud gang identification method is provided, including: constructing a relational network that includes a plurality of nodes; performing cluster discovery based on the relational network to obtain at least one fraud gang included in the relational network, each fraud gang including the plurality of nodes; determining a weak node from the nodes included in the fraud gang, the weak node being a node whose association with the fraud gang meets a weak association criterion; and removing the weak node from the fraud gang to identify a final target fraud gang.

[0006] According to a second aspect, a fraud gang identification device is provided, including: a network construction module, configured to construct a relational network that includes a plurality of nodes; a cluster processing module, configured to perform cluster discovery based on the relational network to obtain at least one fraud gang included in the relational network, each fraud gang including the plurality of nodes; a node determining module, configured to determine a weak node from the nodes included in the fraud gang, the weak node being a node whose association with the fraud gang meets a weak association criterion; and a pruning processing module, configured to remove the weak node from the fraud gang to identify a final target fraud gang.

[0007] According to a third aspect, a fraud gang identification device is provided, where the device includes a memory, a processor, and a computer instruction stored in the memory and executable by the processor, and the processor implements the following steps when executing the instruction: constructing a relational network that includes a plurality of nodes; performing cluster discovery based on the relational network to obtain at least one fraud gang included in the relational network, each fraud gang including the plurality of nodes; determining a weak node from the nodes included in the fraud gang, the weak node being a node whose association with the fraud gang meets a weak association criterion; and removing the weak node from the fraud gang to identify a final target fraud gang.

[0008] According to the fraud gang identification method and device provided in one or more implementations of the present specification, weak nodes are removed from the gang, that is, the nodes that are weakly associated with the gang are removed, the gang identification precision is increased, and the size of the identified gang is reduced, thereby improving the precision of gang identification.

BRIEF DESCRIPTION OF DRAWINGS

[0009] To describe the technical solutions in one or more implementations of the present specification or the prior art more clearly, the following briefly introduces the accompanying drawings for describing the implementations or the prior art. Clearly, the accompanying drawings in the following description are merely one or more implementations of the present specification, and a person of ordinary skill in the art can obtain other drawings based on the accompanying drawings without creative

efforts.

[0010] FIG. 1 is a flowchart illustrating a fraud gang identification method according to one or more implementations of the present specification;

[0011] FIG. 2 is a schematic diagram illustrating a relational network according to one or more implementations of the present specification;

[0012] FIG. 3 is a schematic diagram illustrating removing a link edge from a gang according to one or more implementations of the present specification;

[0013] FIG. 4 is a schematic diagram illustrating removing a weak node from a gang according to one or more implementations of the present specification;

[0014] FIG. 5 is a schematic diagram illustrating removing a weak node from a gang according to one or more implementations of the present specification;

[0015] FIG. 6 is a schematic diagram illustrating gang subdivision according to one or more implementations of the present specification;

[0016] FIG. 7 is a structural diagram illustrating a fraud gang identification according to one or more implementations of the present specification; and

[0017] FIG. 8 is a structural diagram illustrating a fraud gang identification according to one or more implementations of the present specification.

DESCRIPTION OF IMPLEMENTATIONS

[0018] To enable a person skilled in the art to better understand the technical solutions in one or more implementations of the present specification, the following clearly and completely describes the technical solutions in one or more implementations of the present specification with reference to the accompanying drawings in one or more implementations of the present specification. Clearly, the described implementations are merely some rather than all of the implementations of the present specification. Based on one or more implementations of the present specification, all other implementations obtained by a person of ordinary skill in the art without creative efforts shall fall within the scope of the present disclosure.

[0019] The fraud gang identification method according to one or more implementations of the present specification can be applied to identify fraud gangs, for example, gangs that commit fraudulent crimes based on an Internet platform.

[0020] FIG. 1 is a flowchart illustrating a fraud gang identification method. The method can include the following steps.

[0021] Step 100: Construct a relational network that includes a plurality of nodes.

[0022] In this step, the nodes in the relational network can be, for example, user accounts, or user equipment, or can be other types of nodes. The node can be treated or serve as a criminal individual in a gang crime.

[0023] Using user accounts as an example, transfer accounts of different users can be used as nodes. If there is a medium shared between/among different nodes, those nodes are treated as linked. For example, the shared medium can be a common device, a fingerprint, a certificate number, an associated account, Wi-Fi, an LBS, or the like used between accounts in a transfer transaction. If there is at least one shared medium between two nodes, an edge can be connected between the two nodes, and the edge is referred to as a link edge between the nodes.

[0024] Referring to the example relational network shown in FIG. 2, the network can include 15 nodes, and there is a link edge between nodes that have a shared medium. These nodes and link edges form a relational network.

[0025] In addition, it is worthwhile to note that each node in the relational network can be a node with at least a fraud risk. For example, fraudulent transactions have occurred on some nodes, and such nodes are confirmed as fraud nodes; and some nodes have shared a medium with the nodes that have been confirmed as fraud nodes, but no fraudulent transaction has been confirmed on these nodes, such nodes can be considered as nodes with a fraud risk/suspicion or suspected fraud nodes. In this example, a possible fraud gang can be detected in a relational network that includes confirmed fraud nodes or suspected fraud nodes.

[0026] Step 102: Perform cluster discovery based on the relational network to obtain at least one fraud gang included in the relational network, each fraud gang including the plurality of nodes.

[0027] In this step, a fraud gang included in the network can be detected based on the established relational network.

[0028] For example, a propagation tag clustering algorithm can be used to perform community discovery, to detect a fraud gang included in the relational network. Using FIG. 2 as an example, it is found through clustering that nodes 1 to 11 can be grouped into one gang and nodes 12 to 15 can be grouped into another gang.

[0029] Detection of a gang indicates a relatively strong correlation between the nodes included in the gang, for example, the nodes have shared a large number of shared media, or many transfer transactions have occurred on the nodes.

[0030] Step 104: Determine a weak node from the nodes included in the fraud gang, the weak node being a node whose association with the fraud gang meets a weak association criterion.

[0031] For example, a "weak association criterion" can be used to define weak nodes. The criterion can be determined based on the actual scenario of the fraud identification implementation or operation. Two examples of weak nodes are provided below, but the actual implementation is not limited thereto.

[0032] In an example, the "weak association criterion" can be that “the number of link edges to other nodes in a fraud gang is less than a predetermined edge number threshold”. Based on this criterion, in a gang detected in the relational network, if the number of link edges between one node and other nodes in the fraud gang is less than the predetermined edge number threshold, it can be determined that the node is a weak node meeting the weak association criterion.

[0033] Still referring to the example in FIG. 2, there is only one link edge between node 11 and the gang in which node 11 is located, that is edge “11-10”. Assuming that the edge number threshold is 1 and that the node whose link edge number is less than or equal to 1 is regarded as a weak node, then node 11 meets the weak association criterion. It can be determined that node 11 is a weak node.

[0034] In another example, the "weak association criterion" can be that “an edge weight of link edges to other nodes in a fraud gang is less than a predetermined weight threshold”. Based on this criterion, in a gang detected in the relational network, if an edge weight, for example, the edge weight can be an average weight value or the sum of a plurality of weight values, of link edges between one node and other nodes in the fraud gang is less than the predetermined weight threshold, it can be determined that the node is a weak node meeting the weak association criterion.

[0035] Still using FIG. 2 as an example, even if a node has a plurality of link edges to other nodes in the gang, the edge weights of the plurality of link edges are lower than a predetermined weight threshold value, and are confirmed as weak nodes. For example, a link edge exists between node 6 and each of node 7, node 8, and node 5, and each link edge can have a corresponding edge weight, which can be comprehensively determined based on factors such as the number of shared media among the nodes and the number of transfer transactions. The edge weight of one link edge can be used to measure an interaction frequency, an association strength, etc., between the two nodes corresponding to the link edge. For example, the edge weights

of the three link edges of node 6 can be averaged or summed, and the averaged or summed value can be referred to as the edge weight corresponding to node 6. If an edge weight corresponding to a node is lower than a predetermined weight threshold, it can be considered that the node meets a weak association criterion, and the node can be confirmed as a weak node.

[0036] In addition, before a weak node is confirmed, the gang link edge between the gangs can be removed from at least one fraud gang detected in the relational network. For example, in FIG. 2, assuming that nodes 1 to 11 can be grouped into one gang, and nodes 12 to 15 can be grouped into another gang, then the link edge between node 9 and node 13, which can be referred to as a gang link edge, that is, the two nodes connected by the gang link edge belong to different gangs, can be removed, and the link edge between node 2 and node 12 can be removed. Referring to the example in FIG. 3, two separate gangs are obtained after the gang link edges are removed.

[0037] Step 106: Remove the weak node from the fraud gang to identify a final target fraud gang.

[0038] In this step, weak nodes determined in step 104 are removed from each gang. In addition, the weak nodes can be removed in an iterative manner.

[0039] For example, referring to the examples in FIG. 4 and FIG. 5, in FIG. 4, the link edge between node 9 and node 1 can be removed based on the weak association criterion, which is equivalent to removing node 9 from the gang; and the link edge between node 11 and node 10 is also removed, which is equivalent to removing node 11 from the gang. Next, in FIG. 5, the determination is continued based on the weak association criterion, and node 10 is determined as a weak node, because node 10 also has only one link edge. After the node 10 is removed, the link edge between node 10 and node 5 can be removed in FIG. 5. After node 9, node 11, and node 10 have been removed, each of the remaining nodes has more than one link edge, and therefore is not a weak node.

[0040] In addition, in the above method of removing weak nodes in an iterative manner, all weak nodes in each fraud gang can be removed. In actual implementations, it is also possible that only some of the weak nodes are removed. For example, in FIG. 4, node 11 and node 9 can be removed, but node 10 can be retained. To some extent, the removal of some weak nodes can also improve the precision of the gang identification. Specifically, how many weak nodes are removed can be determined

based on the scenario of fraud gang identification operation. For example, an upper limit of the number of weak nodes to be removed can be predetermined.

[0041] According to the fraud gang identification method in this example, the weak nodes are removed from the gang, that is, the nodes that are weakly associated are removed from the gang, so that the precision of the gang identification is optimized, and the size of the gang is also optimized, which helps to improve the precision of the gang identification.

[0042] In addition, after the weak nodes are removed from the fraud gang, if the fraud gang meets the gang subdivision criterion, cluster discovery can be continued on the fraud gang, that is, the gang can be subdivided.

[0043] For example, the gang subdivision criteria include, but are not limited to, the two criteria listed below, which can be considered separately or together:

if the number of nodes included in the fraud gang is greater than a node number threshold, the fraud gang needs to be subdivided; or

if the fraud case concentration value of the fraud gang is lower than a predetermined case concentration threshold, the fraud gang needs to be subdivided. The fraud case concentration value can be, for example, a ratio between the number of fraudulent transactions executed by the nodes in the gang to the total number of transactions of the gang.

[0044] In FIG. 5, assuming that after the weak nodes are removed, the number of nodes of the gang of node 1 to node 8 still relatively large, and the number of nodes is greater than the node number threshold, then the propagation tag clustering algorithm can be used to perform gang subdivision. A weak node can be removed after the subdivision. For example, after the subdivision, the gang of node 1 to node 8 can be divided into two gangs: one gang includes node 1 to node 4, and the other gang includes node 5 to node 8, as shown in FIG. 6.

[0045] Through continuous optimization of the gang identification, the finally identified gang can be referred to as the target fraud gang. The target fraud gang is identified with a good precision. Parameters such as association strength and fraud case concentration value of the target fraud gang can be calculated and pushed to an anti-fraud strategy team, thereby improving the precision of gang crackdown.

[0046] To implement the above fraud gang identification method, one or more implementations of the present specification provide a fraud gang identification device. As shown in FIG. 7, the device can include a network construction module 71,

a cluster processing module 72, a node determining module 73, and a pruning processing module 74.

[0047] The network construction module 71 is configured to construct a relational network that includes a plurality of nodes.

[0048] The cluster processing module 72 is configured to perform cluster discovery based on the relational network to obtain at least one fraud gang included in the relational network, each fraud gang including the plurality of nodes.

[0049] The node determining module 73 is configured to determine a weak node from the nodes included in the fraud gang, the weak node being a node whose association with the fraud gang meets a weak association criterion.

[0050] The pruning processing module 74 is configured to remove the weak node from the fraud gang to identify a final target fraud gang.

[0051] In an example, the node determining module 73 is specifically configured to: if the number of link edges between a node and other nodes in a fraud gang is less than a predetermined edge number threshold, determine that the node is a weak node meeting a weak association criterion; or if an edge weight of a link edge between a node and another node in the fraud gang is lower than a predetermined weight threshold, determine that the node is a weak node meeting a weak association criterion.

[0052] In an example, as shown in FIG. 8, the device can further include a gang subdivision module 75, configured to continue to perform cluster discovery on the fraud gang if the fraud gang meets the gang subdivision criterion after the pruning processing module has removed the weak node from the fraud gang.

[0053] The devices or modules illustrated in the above implementations can be implemented by computer chips, entities, or products having a certain function. A typical implementation device is a computer in the form of a personal computer, a laptop computer, a cellular phone, a camera phone, a smart phone, a personal digital assistant, a media player, a navigation device, an e-mail transceiver, a game console, a tablet computer, a wearable device, or any combination of at least two of these devices.

[0054] For ease of description, the above device is described by dividing functions into various modules. Of course, during implementation of one or more implementations of the present specification, the functions of each module can be implemented in at least one of software or hardware.

[0055] The execution sequence of the steps in the process shown in the above figure is not limited to the sequence in the flowchart. In addition, the description of each step can be implemented in the form of software, hardware, or a combination of the software and the hardware. For example, a person skilled in the art can present the description of each step in the form of software code, namely computer executable instructions capable of implementing logical functions corresponding to the steps. When implemented in a software method, the executable instructions can be stored in a memory and executed by a processor in the device.

[0056] For example, corresponding to the above methods, one or more implementations of the present specification provide a fraud gang identification device. The device can include a processor, a memory, and a computer instruction that is stored in the memory and can run on the processor, and the processor executes the instruction to implement the following steps: constructing a relational network that includes a plurality of nodes; performing cluster discovery based on the relational network to obtain at least one fraud gang included in the relational network, each fraud gang including the plurality of nodes; determining a weak node from the nodes included in the fraud gang, the weak node being a node whose association with the fraud gang meets a weak association criterion; and removing the weak node from the fraud gang to identify a final target fraud gang.

[0057] It is also worthwhile to note that terms “include”, “comprise” or any other variant is intended to cover non-exclusive inclusion, so that processes, methods, commodities or devices that include a series of elements include not only those elements but also other elements that are not explicitly listed, or elements inherent in such processes, methods, commodities or devices. An element described by "includes a …" further includes, without more constraints, another identical element in the process, method, product, or device that includes the element.

[0058] A person skilled in the art should understand that one or more implementations of the present specification can be provided as a method, a system, or a computer program product. Therefore, one or more implementations of the present specification can use a form of hardware only implementations, software only implementation, or implementations with a combination of software and hardware. In addition, one or more implementations of the present specification can use a form of a computer program product that is implemented on one or more computer-usable storage media (including but not limited to a disk memory, a CD-ROM, an optical

memory, etc.) that include computer-usable program code.

[0059] One or more implementations of the present specification can be described in the general context of computer-executable instructions, for example, a program module. Generally, the program module includes a routine, a program, an object, a component, a data structure, etc., executing a specific task or implementing a specific abstract data type. One or more implementations of the present specification can also be practiced in distributed computing environments where tasks are performed by remote processing devices connected through a communications network. In a distributed computing environment, the program module can be located in both local and remote computer storage media including storage devices.

[0060] The implementations of the present specification are described in a progressive way. For same or similar parts of the implementations, mutual references can be made to the implementations. Each implementation focuses on a difference from the other implementations. Particularly, a system implementation is basically similar to a method implementation, and therefore is described briefly. For related parts, references can be made to related descriptions in the method implementation.

[0061] Specific implementations of the present specification are described above. Other implementations fall within the scope of the appended claims. In some situations, the actions or steps described in the claims can be performed in an order different from the order in the implementation and the desired results can still be achieved. In addition, the process depicted in the accompanying drawings does not necessarily require a particular execution order to achieve the desired results. In some implementations, multi-tasking and parallel processing can be advantageous.

[0062] The above descriptions are merely preferred implementations of one or more implementations of the present specification, and are not intended to limit one or more implementations of the present specification. Any modification, equivalent replacement, improvement, etc., made without departing from the spirit and principles of one or more implementations of the present specification shall fall within the protection scope of one or more implementations of the present specification.

CLAIMS

What is claimed is:

1. A computer-implemented method, comprising:

constructing, by a computing device having a processor, a relational network that includes a plurality of nodes and established link edges among the plurality of nodes, wherein at least one of the established link edges are established if there is a medium shared between at least two of each of the plurality of nodes, and wherein the shared medium includes common Wi-Fi signals shared by at least two of each of the plurality of nodes;

performing, by the computing device having the processor, cluster discovery based on the relational network to obtain at least one fraud gang included in the relational network, each fraud gang including nodes of the plurality of nodes;

determining, by the computing device having the processor, a weak node from the nodes included in the fraud gang, the weak node being a node whose association with other nodes in the fraud gang meets a weak association criterion, the weak association criterion includes one or more of an average value of weights of all link edges between the node and the other nodes in the fraud gang being smaller than a first threshold value or a sum of weights of all link edges between the node and the other nodes in the fraud gang being smaller than a second threshold value; and

removing, by the computing device having the processor, the weak node from the fraud gang to identify a final target fraud gang.

2. The method according to claim 1, wherein the determining the weak node from the nodes included in the fraud gang comprises:

in response to that a number of link edges between the node and the other nodes in the fraud gang is less than an edge number threshold, determining that the node is the weak node meeting the weak association criterion.

3. The method according to claim 1, wherein the determining the weak node from the nodes included in the fraud gang includes:

in response to that an edge weight of a link edge between a node and another node in the fraud gang is lower than a weight threshold, determining that the node is the weak node meeting the weak association criterion.

4. The method according to claim 1, wherein the removing the weak node from the fraud gang includes one or more of:

removing a gang link edge between different gangs from at least one fraud gang; or

removing some or all weak nodes from each fraud gang.

5. The method of claim 1, further comprising:

after the removing the weak node from the fraud gang and before the identifying the final target fraud gang, performing cluster discovery on the fraud gang in response to that the fraud gang meets a gang subdivision criterion.

6. The method according to claim 5, wherein the gang subdivision criterion includes:

a number of nodes included in the fraud gang is greater than a node number threshold; or

a fraud case concentration value of the fraud gang is lower than a case concentration threshold.

7. The method of claim 1, wherein the constructing the relational network includes determining a link edge between a first node and a second node of the plurality of nodes.

8. The method of claim 1, wherein the shared medium further includes one or more of a device, a fingerprint, a certificate number, an account number, a Wi-Fi location, or a location-based service account.

9. The method of claim 3, wherein the edge weight of the link edge is determined based on one or more of a number of shared media between the node and the other node or a number of transfer transactions between the node and the other node.

10. A device, comprising:

a network construction module, configured to construct a relational network that

includes a plurality of nodes linked to one another through established link edges, wherein at least one of the established link edges are established if there is a medium shared between at least two of each of the plurality of nodes, and wherein the shared medium includes common Wi-Fi signals shared by at least two of each of the plurality of nodes;

a cluster processing module, configured to perform cluster discovery based on the relational network to obtain at least one fraud gang included in the relational network, each fraud gang includes nodes of the plurality of nodes;

a node determining module, configured to determine a weak node from the nodes included in the fraud gang, the weak node being a node whose association with other nodes of the fraud gang meets a weak association criterion, the weak association criterion includes one or more of an average value of weights of all link edges between the node and the other nodes in the fraud gang being smaller than a first threshold value or a sum of weights of all link edges between the node and the other nodes in the fraud gang being smaller than a second threshold value; and

a pruning processing module, configured to remove the weak node from the fraud gang to identify a final target fraud gang.

11. The device according to claim 10, wherein the node determining module is configured to:

in response to that a number of link edges between the node and the other nodes in the fraud gang is less than an edge number threshold, determine that the node is the weak node meeting the weak association criterion; or

in response to that an edge weight of a link edge between a node and another node in the fraud gang is lower than a weight threshold, determine that the node is the weak node meeting the weak association criterion.

12. The device of claim 11, wherein the edge weight is determined based on one or more of a number of shared media between the node and another node or a number of transfer transactions between the node and the other node.

13. The device according to claim 10, further comprising:

a subdivision module, configured to continue to perform cluster discovery on the fraud gang if the fraud gang meets a gang subdivision criterion after the pruning

processing module has removed the weak node from the fraud gang.

14. A system, comprising:

a memory;

a processor; and

computer instructions stored in the memory, which, when executed by the processor, configures the processor to implement acts including:

constructing a relational network that includes a plurality of nodes linked to one another through established link edges, wherein at least one of the established link edges are established if there is a medium shared between at least two of each of the plurality of nodes, and wherein the shared medium includes common Wi-Fi signals shared by at least two of each of the plurality of nodes;

performing cluster discovery based on the relational network to obtain at least one fraud gang included in the relational network, each fraud gang including nodes of the plurality of nodes;

determining a weak node from the nodes included in the fraud gang, the weak node being a node whose association with other nodes in the fraud gang meets a weak association criterion, the weak association criterion including one or more of an average value of weights of all link edges between the node and the other nodes in the fraud gang being smaller than a first threshold value or a sum of weights of all link edges between the node and the other nodes in the fraud gang being smaller than a second threshold value; and

removing the weak node from the fraud gang to identify a final target fraud gang.

15. The system according to claim 14, wherein the determining the weak node from the nodes included in the fraud gang comprises:

in response to that a number of link edges between the node and the other nodes in the fraud gang is less than an edge number threshold, determining that the node is the weak node meeting the weak association criterion.

16. The system according to claim 14, wherein the removing the weak node from the fraud gang includes:

in response to that an edge weight of a link edge between a node and another node in the cluster of nodes is lower than a weight threshold, determining that the

node is the weak node meeting the weak association criterion.

17. The system of claim 14, further comprising:

after the removing the node from the fraud gang and before the identifying the final target fraud gang, performing cluster discovery on the fraud gang in response to that the fraud gang meets a gang subdivision criterion.

18. The system according to claim 17, wherein the gang subdivision criterion includes:

a number of nodes included in the fraud gang is greater than a node number threshold; or

a fraud case concentration value in the fraud gang is lower than a case concentration threshold.

19. The system of claim 16, wherein the edge weight of the link edge is determined based on one or more of a number of shared media between two linked nodes that are linked by the link edge or a number of transfer transactions between the two linked nodes.