SAS5801-2016

Minimizing Fraud Risk through Dynamic Entity Resolution and Network Analysis

Danielle Davis, Stephen Boyd, and Ray Ong, SAS Institute Inc., Cary, NC

ABSTRACT

Every day, businesses have to remain vigilant of fraudulent activity, which threatens customers, partners, employees, and financials. Normally, networks of people or groups perpetrate deviant activity. Finding these connections is now made easier for analysts with SAS® Visual Investigator (SAS® VI), an upcoming SAS® solution that ultimately minimizes the loss of money and preserves mutual trust among its shareholders. SAS Visual Investigator takes advantage of the capabilities of our new SAS® Viya™ server. Investigators can efficiently investigate suspicious cases across business lines, which has traditionally been difficult. However, the time required to collect, process and identify emerging fraud and compliance issues has been costly. Making proactive analysis accessible to analysts is now more important than ever. SAS Visual Investigator was designed with this goal in mind and a key component is the visual social network view. This paper discusses how the network analysis view of SAS Visual Investigator, with all its dynamic visual capabilities, can make the investigative process more informative and efficient.

INTRODUCTION

Today's fraud is like a moving target. It purposely flies under the radar and follows the path of least resistance. With companies automating more of their processes, the window to detect fraud is also shrinking. Traditional techniques often fail to identify fraudulent behavior. According to Gartner Research, basic monitoring for deviation will select the sole fraudsters but it falls short when looking at fraud rings. Also can be prone to a lot of false positives, which is not good for customer service.

SAS® VI is a solution to help your company detect and manage fraud. The social network view of SAS® VI can provide useful insight into large data sets along network, spatial, and time dimensions based on the interconnectedness of the subjects being analyzed. This social network analysis (SNA) can be a key component when looking for fraudulent activity.

FRAUD DETECTION

Why is fraud hard to detect? First, it is an uncommon pattern. It will always take the path of least resistance and the whole goal is to fly under the radar, to blend in. Therefore, there is rarely frequently occurring patterns where classical data mining techniques can be used. Second, today's fraud is typically very carefully organized crime. They do not operate independently, they are big organizations with multiple players and roles being played by different individuals. However, fraudsters try to blend into the environment and not behave different from others in order not to get noticed and to remain covered by non-fraudsters. Third, techniques and tricks fraudsters evolve over time. It looks to be an endless game of cat and mouse played between fraudsters and fraud fighters. Lastly, instead of setting up a massive, one-time strike, they can do a lot of smaller activities under the radar for the same monetary result. So how can we battle these issues? SAS® VI social network view can help ease the pain of detecting fraud.

SOCIAL NETWORK VIEWER

A social network view of your data is designed to analyze the data to identify relationships, and mining it for new information (such as the quality or effectiveness of a relationship). It identifies patterns of behavior that only appear suspicious when viewed across related accounts or attributes of an account. You can examine social structure and interdependencies (or work patterns) of individuals or organizations.

Let's breakdown the viewer into components that can assist in revealing fraud.
First, let's look at the viewer from a high level:

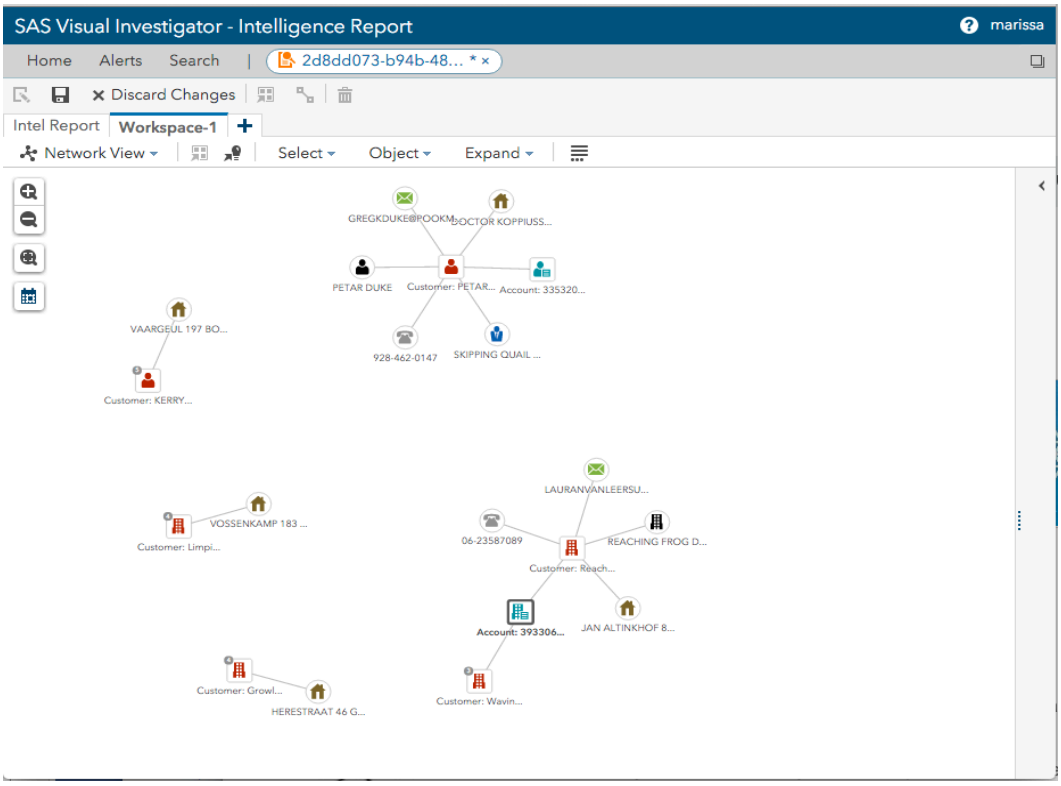

Figure 1 - SAS VI - Social Network View

SEARCH AND EXPANSION

In order to detect the unusual patterns of fraud, underlying fraud rings and fraud that crosses companies and/or product lines, a user must be able link nodes together. In other words, an analyst needs to see the relationships between information/knowledge entities. Since the network viewer is only as good as its data, SAS® Visual Investigator's network viewer is powered by data that has been processed through entity resolution and then stored and indexed in Elasticsearch. This provides the user the ability to have efficient searches for resolved document information that would have been otherwise overlooked.

The Entity Resolution Service enables SAS® Visual Investigator to identify unique entities across various source documents, which can have different data representing the same entity. Particularly in fraud detection scenarios, a Person entity, for example, can represent themselves with different first names, different last names, different phone numbers, different addresses, and different SSNs. Entity resolution seeks to find matches across a number of different compounds, each compound being defined as a set of columns (elements) that should determine uniqueness.

Elasticsearch storage provides quick retrieval time and allows the user to easily add new nodes to their existing network and discover relationships that were previously undetected. This also allows the analyst to know if a node has additional relationships that can be expanded on the network. This expansion can expand one or 2 levels as well as providing a complete community expansion.

Below is an example of the search feature of SAS® Visual Investigator in order to populate the network viewer with relational data. Once a user has a list of possibilities in his search, he/she can add these entities to a workspace. This is where the user can start to explore the relationships.

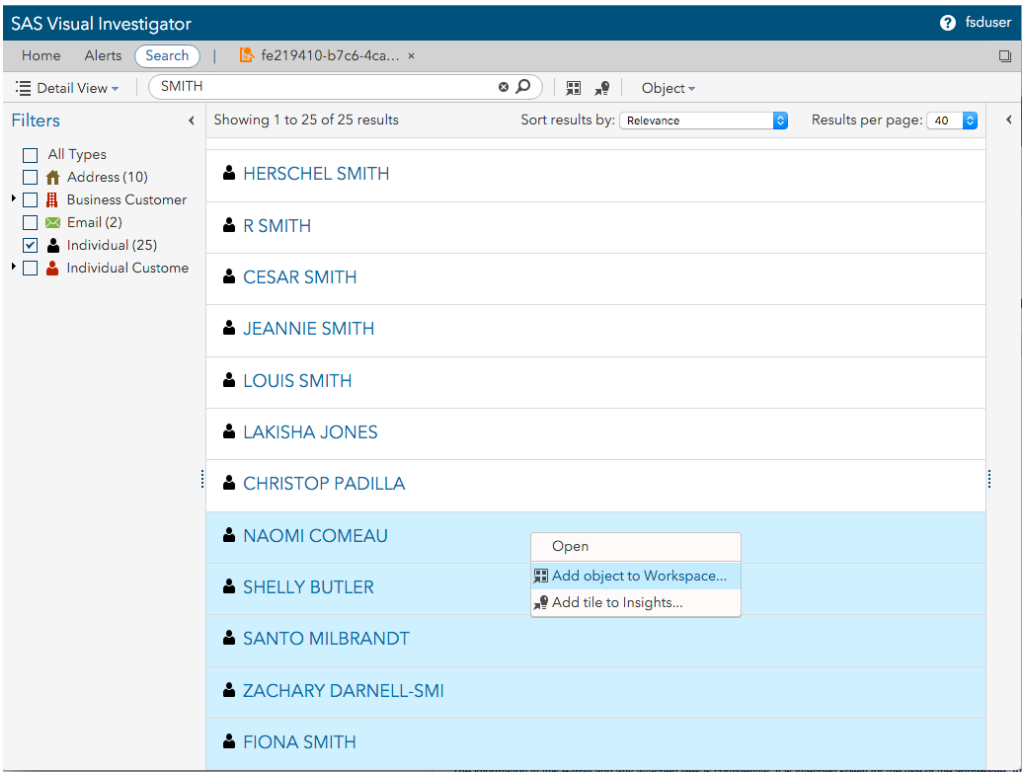

Figure 2. Search and Entity Resolution Built around the Network Viewer

Once the entities and documents are loaded in a workspace, the user can visually tell which nodes have further relationships that can be explored. These nodes contain a degree in the upper left corner to indicate the number of additional direct relationship they have. These nodes can be expanded in the network to reveal the additional information.

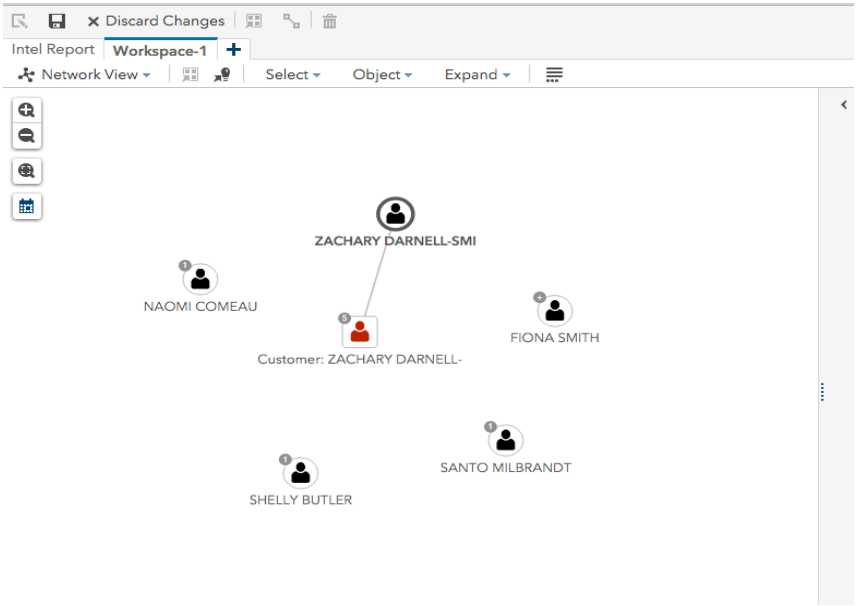

Figure 3. Five Nodes in the Network Have More Direct Relationships

This functionality allows a user to begin to understand the relationships that might otherwise go unnoticed. SAS® Visual Investigator adds another level of analytical power by using SAS® Viya™ and allowing a user to expose the entire community of a particular node. In other words, using analytical techniques to expose nodes that are highly interconnected to the node selected.

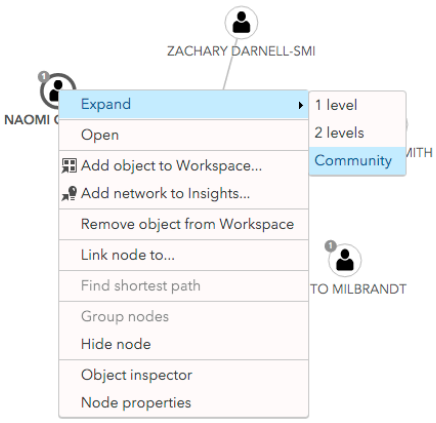

Figure 4. Expanding a Node

NODE AND LINK PROPERTIES AND ANNOTATION

Part of what makes a network viewer essential to fraud detection is its ability to help the analyst tell/track the fraud pattern in the network. This requires customization of the network nodes and links.

The network view in SAS® Visual Investigator allows user to change many different properties for a node including its size, color, shape, and icon.

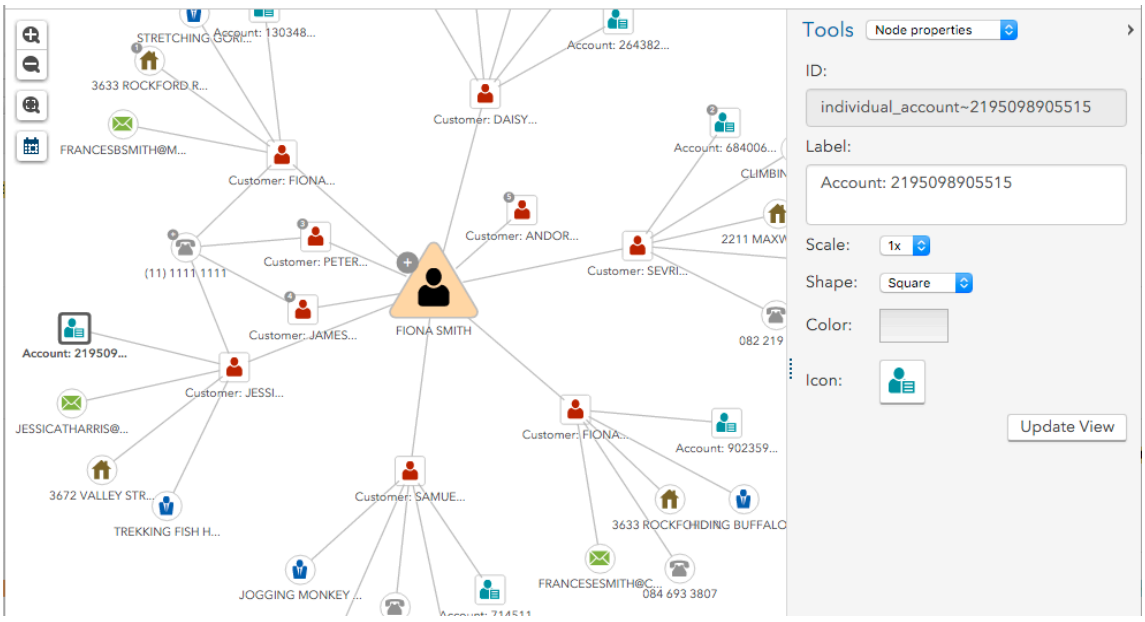

Figure 5. Node Properties

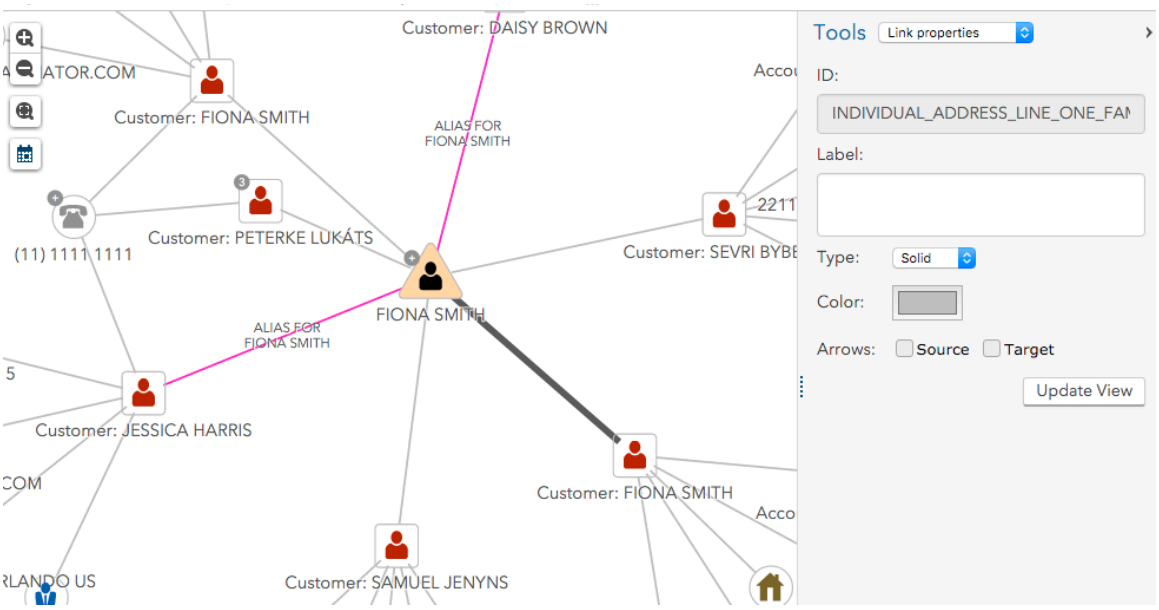

Figure 6. Link Properties

Node annotation is another feature that can assist an analyst in a number of different ways. These annotations are added based on the value of a specific attribute on the entity or document. This can be beneficial with the following:

- To help further document the investigation of the potential fraudulent activity.
- To alert the analyst that the entity has exceeded a threshold.
- To further define a node in more specific terms. For example, if there is a claim the annotations could inform the analyst of the type of claim, claim status, and insurer for the claim.

The annotations can be placed on 8 different locations of the nodes and can be in the form of text or icons.

Figure 7. Icon Decoration

GROUPING NODES

Another key feature in helping the analyst understand the story of the fraudulent activities is the ability to group nodes. The grouping functionality not only allows you to “clean up” you network without discarding important information it allows to view the contents of these nodes through our “object inspector”, which shows detailed information about a node.

If you look back at the links that are referenced in Figure 6, we see that they reference two nodes that are aliases for another person. A user can group all of those nodes together so that it helps minimize the complication of the network.

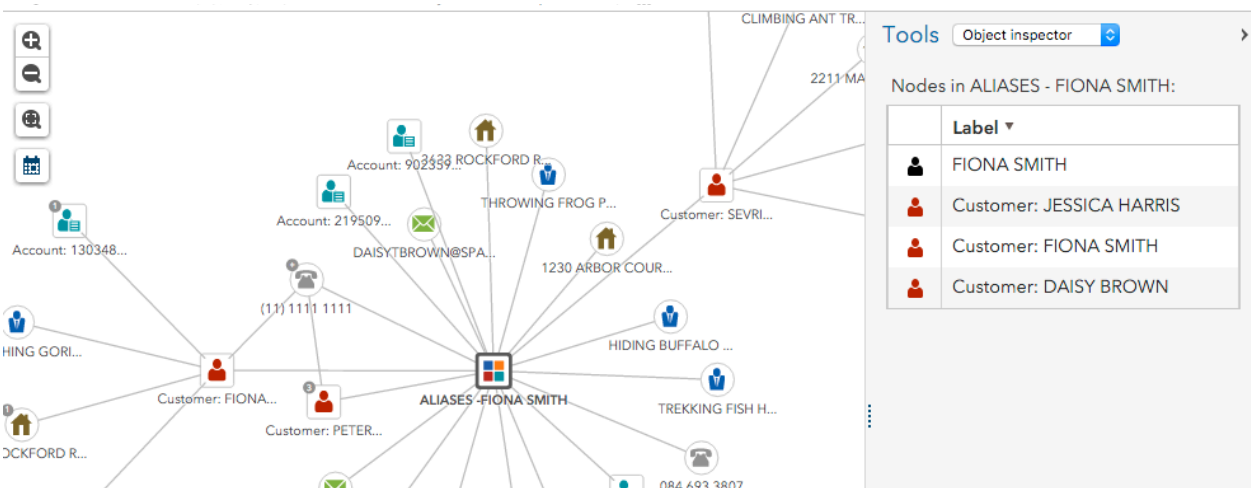

Figure 8. Grouping Nodes

NODE SELECTION

When networks become large, it is very cumbersome and tedious to find specific nodes. SAS® Visual Investigator provides a tool to easily find nodes that are of the same type and an option to search for a string in any of the name labels.

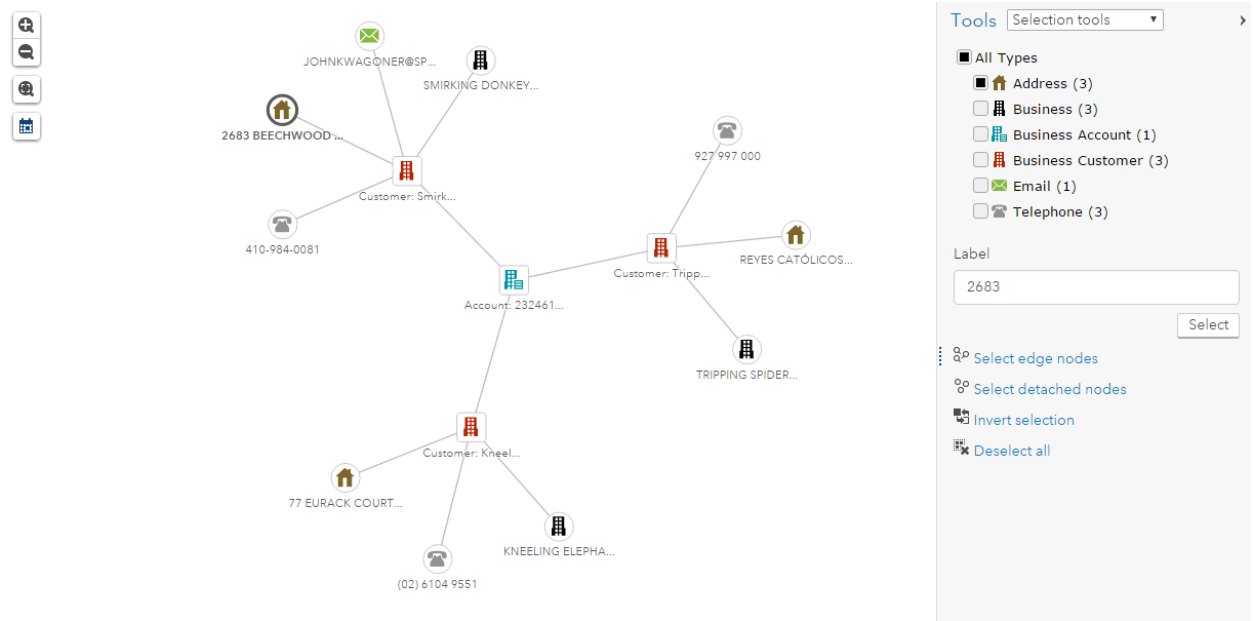

Figure 9. Node Selection

NETWORK LAYOUT

Some fraud rings can be quite large and this can require your social network to get quite large and adding new information to an existing network can become tedious when it comes to laying out all the nodes. SAS® Visual Investigator's network view allows the user to layout and re-adjust their network in a number of different ways.

1. Initial layout tries to ensure that nodes are not overlapping and minimizes the amount of link crossing. This applies to new nodes coming into the network as well as new nodes coming in from an expansion. The entire network might need to be adjusted depending on how many nodes are being added.
2. If a user has spent quite a bit of time adjusting individual nodes, they can also choose to only adjust the nodes that have not been manually moved.
3. A user also has the ability to choose to adjust only the nodes they have selected in the network.
4. Advanced features are also available for over all fine-tuning of the network.

The network viewer can also handle networks with thousands of nodes by reducing the details so that the analyst can see the overall view of the network as well zoom into the details.

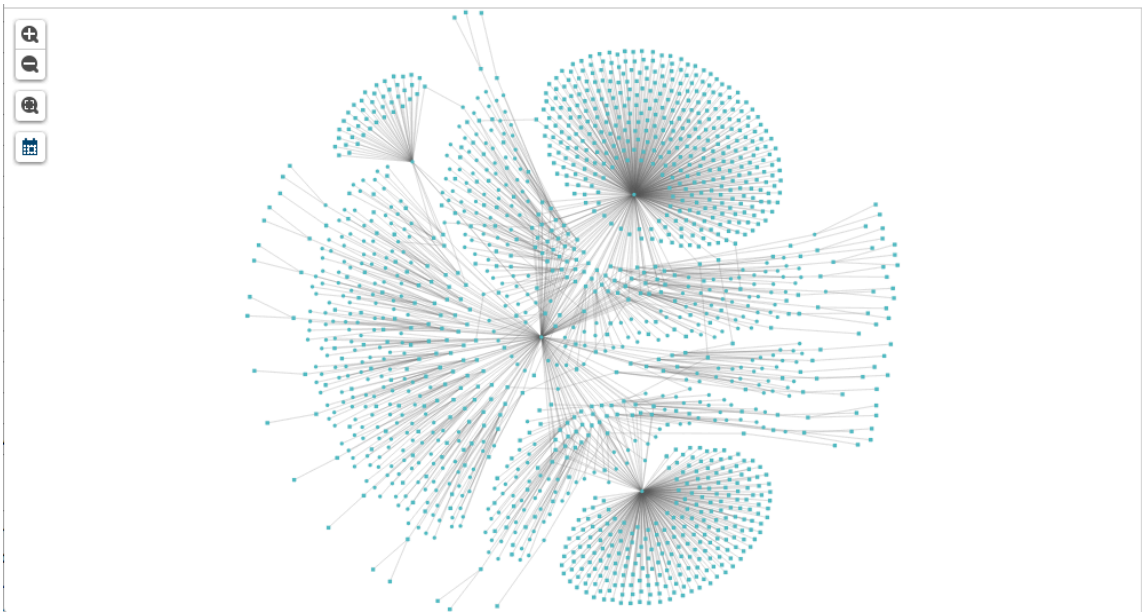

Figure 10. Large Networks

NETWORK ANALYTICS

Social Network Analysis can also involve some statistical analytics that can assist with understand the important players in a network. Taking advantage of SAS® Viya, SAS® Visual Investigator's network viewer provides several key network analytics that can help answer questions:

- How highly connected is an entity within a network?
- What is an entity's overall importance in a network?
- How central is an entity within a network?
- How does information flow within a network?

TIMELINE

Another key feature to network analysis is to understand when particular entities started their role in the network. The ability to look at patterns over time. The network viewer contains a time slider component that can provide a simple view of communications that happened within a given time-frame and a continuous temporal view that includes the history before the time-frame defined. Let's review both options.

The diagram below depicts the time slider in a "no-history" mode:

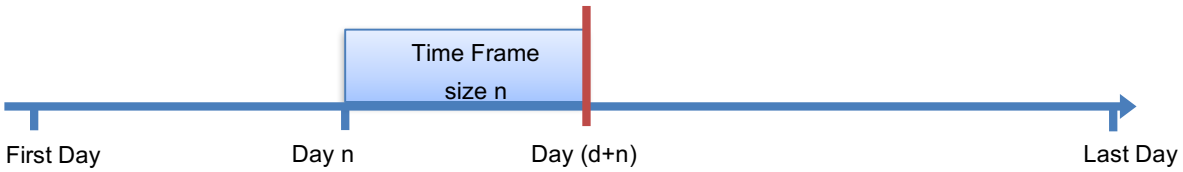

Figure 11. Time Slider in "no history" Mode

In this example, day d is the first day that the visualization is showing and the current time frame is [d, d+n]. Only communications inside the current time frame are calculated and displayed, and only communications within that time-frame are considered active.

The following figure shows the time-slider in the "no-history" mode:

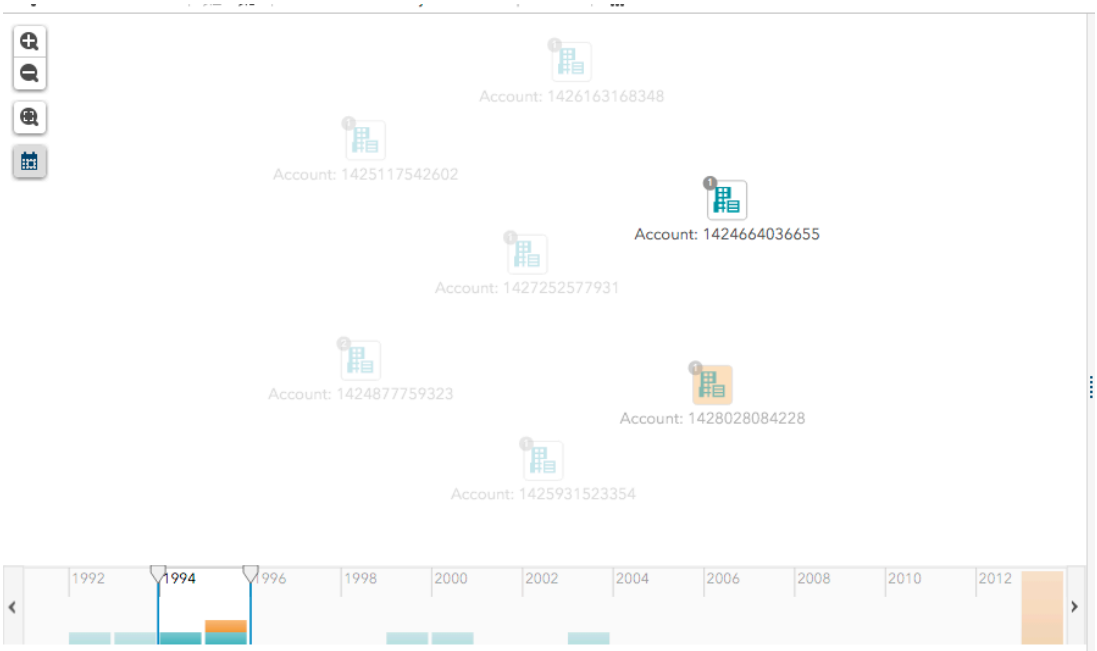

Figure 12. Time Slider with "no-history" Mode

After calculating the full range of dates, the timeline will determine what is the best interval to "bucket" date values in. Above you see that the range of data was over 10 years so that the intervals are yearly buckets. Notice that the nodes that are grayed out have start dates outside of our selected view, whether they started before the date range or after. The nodes that are grayed out but also have a muted orange background are nodes that had end dates within our range. In other words, they had an end date in 1995.

The next diagram shows the time slider in the “with history” mode:

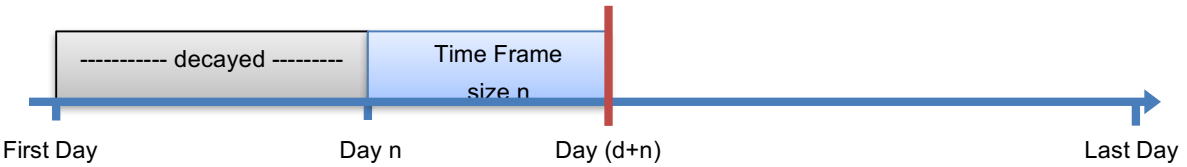

Figure 13. Time Slider in "with history" Mode

This time frame window allows users to foresee the activities happening inside the time frame after the current day. Thus, day d is the current day that the visualization is showing and the current time frame is [d, d+n]. All communications through day (d+n) are calculated and displayed, and if a communication takes place before or on day d, it is active. Below is an example of the time-slider using the “with history” feature:

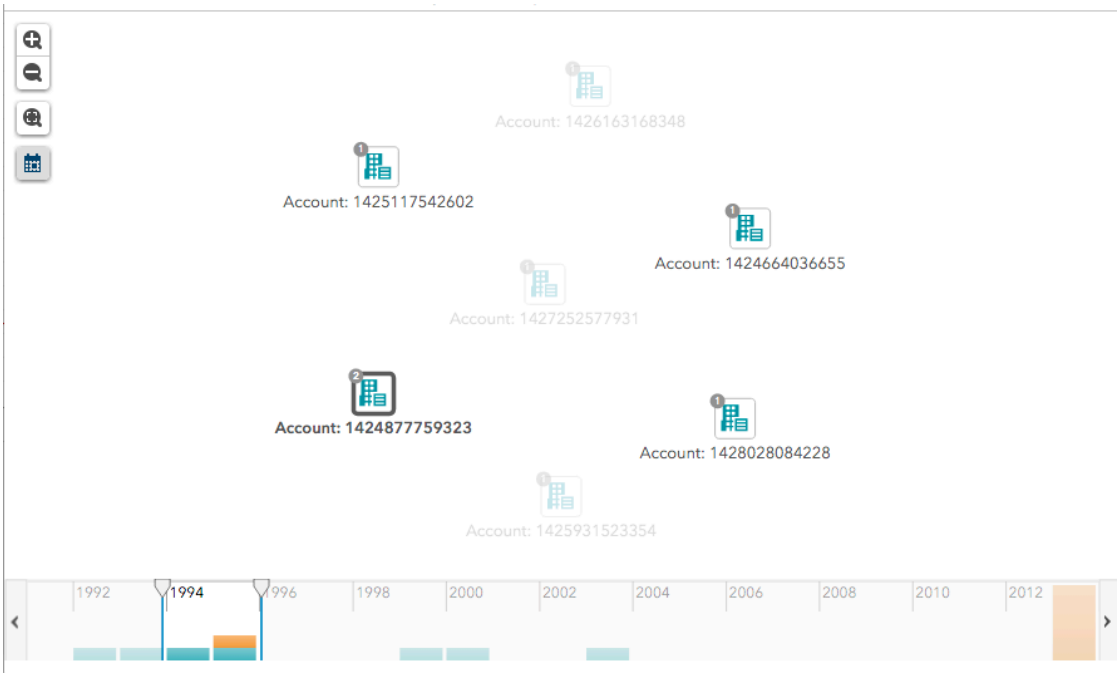

Figure 14 . Time Slider Show "with history" Feature

Notice that the nodes that are grayed out have not come into play yet. The have start dates after 1995. However, all other nodes have start dates within or before the range of time we have in view. None of the nodes are grayed-out and orange since the left-handed slide has not passed an orange bar in the timeline. Hence, the end date is not in the past yet.

The timeline also allows the user to zoom in on a bar and see the actual dates that fell into that bucket and those nodes the network will reflect that zoomed in view.

CONCLUSION

Social network analysis is an important part of a layered prevention approach to fraud and compliance discovery. A complete fraud solution involves incorporating fraud and compliance modeling to prevent fraudulent activity from surfacing in line, and social network analysis to discover new emerging patterns or changes in known fraud and compliance patterns. This combination provides a solution that can quickly update the fraud and compliance alert system with higher quality rule sets for triggering potential fraud and decreasing false positives or customer friction.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Danielle Davis
100 SAS Campus Drive
Cary, NC 27513
SAS Institute Inc.
Danielle.Davis@sas.com
http://www.sas.com