Implementing social network analysis for fraud prevention

ABOUT THIS PAPER

Fraud continues to rise and impose enormous costs on businesses, demanding the need for effective fraud detection technologies and methods.

The use of social network analysis (SNA) to combat fraud is slowing gaining acceptance, enabling fraud investigators to uncover and prevent increasingly sophisticated fraud schemes. With SNA, investigators can detect fraud patterns within and across product lines, overcoming the limitations of more traditional silo approaches to fraud detection and analysis.

Learn more in this paper about what SNA involves and how it might be advantageous for your business.

Fraud detection and analysis has traditionally involved a silo approach. Rarely does an investigator look across product lines to identify fraudulent connections. However, with the introduction of social network analysis (SNA), investigators are now able to detect data patterns within and across product lines as a potential crime ring or group is developing, saving companies from losses as the crime ring further develops. This paper introduces key ideas and concepts related to SNA, suggests an SNA approach for fraud prevention, provide a case study, and describes some of the limitations of SNA.

Social network analysis

The use of social network analysis for combating fraud is slowly gaining acceptance within a range of sectors, primarily in financial services, telecommunications and public organizations. Anti-money laundering, identity fraud, network fraud, denial of service attacks and terrorist financing are some of the areas of fraud where SNA could be used to significantly improve fraud detection. SNA techniques and tools have been deployed in landmark cases like tracing terrorist funding after 9/11 attacks by FinCEN and insider trading cases identified by the Australian Securities and Investment Commission.

For most business though, profit loss continues to be a key reason to invest in fraud detection. More than 15% of income loss for medium sized businesses in Germany is due to fraud, corruption, and defalcation. This is the second largest area of loss after theft, burglary, and assault [Corporate Trust 2009].

Financial crimes are continuously evolving into more complex systems of attack on businesses and, therefore, the technologies financial institutions use to detect and stop these crimes from occurring need to evolve. In Germany, there were more than 4,100 reported cases of check fraud in 2002. By 2009, this dropped to only a little more than 600 cases [BKA, 2002&2009]. In contrast, reported fraud cases related to card payments have risen 345% just from 2007 to 2009 [BKA, 2008&2009]. With the recent economic crisis and obvious changes in technology, it is important to be more vigilant regarding fraud detection.

Increasingly sophisticated fraudsters are able to easily slip behind risk-score based analysis to avoid detection and, to overcome this issue, organizations need to better understand the dynamics and patterns of fraud and fraud networks. This is where the visual and analytical capabilities of SNA can help the fraud prevention function to effectively detect and prevent fraud originating from web-based and other more traditional business channels.

In general, SNA is a “data mining technique that reveals the structure and content of a body of information by representing it as a set of interconnected, linked objects or entities.” [Mena, 2003].

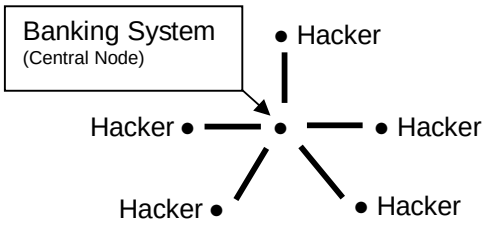
The perfect combination of advances in knowledge management, visualization techniques, data availability and increased computing power enabled the steady rise of SNA as an interdisciplinary investigative technique in a wide array of sectors. Unlike other analytical techniques like statistics that are based on the notion of independence of subjects, SNA can provide useful insight into large datasets along network, spatial and time dimensions based on the interconnectedness of the subjects being analyzed.

The perfect combination of advances in knowledge management, visualization techniques, data availability and increased computing power enabled the steady rise of SNA as an interdisciplinary investigative technique in a wide array of sectors.

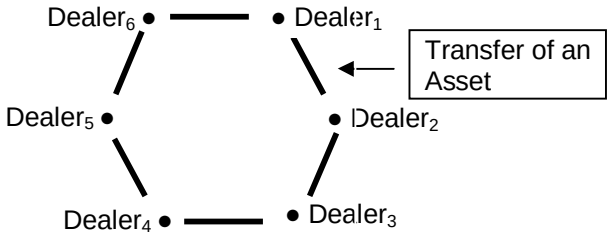
Social network structure

At a basic level, a social network consists of *nodes* (vertices) that are connected to other related nodes by *links* (relationships). The connection between two nodes is called an *edge*. If all the nodes in a social network are connected to each other, it is called a *fully-connected* network. A *path* refers to a collection of nodes that are connected by a link. The diagrams below show three simple variations of a financial fraud social network.

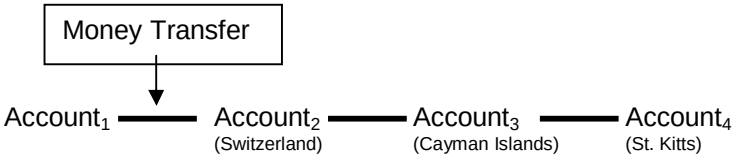
SNA measures



Denial of Service-Hacker Attack (Star)



Networking Fraud Ring (Circle)



Money Laundering (Chain)

Density:
Density is the general level of linkage among the social network nodes. It is defined as the number of edges in a portion of a social network to the maximum number of edges that theoretically make up the social network [Wassermann and Faust 1995]. The following formula is used to calculate the relative density of a portion of a social network. C is the number of observed edges and n the total number of nodes in the social network. n(n-1)/2 gives us the theoretical maximum number of edges that are possible in a given social network.

Density measures can take any value between [0,1], with 0 representing the least density and 1 the maximum density.

$$\text{Density} = \frac{C}{n(n-1)/2}$$

Density measures are extremely useful in determining potential fraud hotspots in retail banking from a maze of account transactions and applied control measures. Credit card transaction monitoring and money-laundering are potentially two areas where density metrics could trigger the necessity for deeper investigations.

Density measures are extremely useful in determining potential fraud hotspots in retail banking from a maze of account transactions and applied control measures. Credit card transaction monitoring and money-laundering are potentially two areas where density metrics could trigger the necessity for deeper investigations.

Centrality

Centrality is the measure of closeness of a node to the center(s) of high activity in a network [Chan & Leibowitz, 2006] and implies the structural importance of the node in the network. Centrality can be measured by *degree*, *closeness* and *betweenness*.

Degree is the direct count of the number of connections a node has to other nodes. A higher degree value compared to peers denotes higher influence in the network.

Closeness focuses on the overall closeness of a node to all other nodes in the network [Wassermann & Faust, 1997]. This measure goes beyond what *degree* offers and emphasizes the geodesic distance between the nodes. Chan & Leibowitz, 2006 opine that the *closeness* of a node gives it quicker access to all other nodes in the network.

Betweenness measures the extent of a node's placement on the shortest path between other pairs of nodes in the network [Ehrlich & Carboni, working paper undated]. One can use centrality to identify the nodes that are pivotal to the success of the fraud network and, in turn, focus resources on investigating these high return suspects.

Other measures

Other measures that are commonly used in SNA are *sub-structures, structural holes* and *clustering-coefficient.* These measures are used for network classification and network path prediction and a detailed discussion of them is beyond the scope of this article.

SNA fraud prevention approach

For an effective fraud prevention strategy using social network analysis, the following methodology could be adopted. This methodology was developed based on the works of Chan & Leibowitz, 2006 and Karamon, et al., 2008, as well as the authors' own expertise in the fraud prevention area.

1. Build a social network(s) that encompasses the whole organization. Configuring SNA systems to produce something of this magnitude could be cumbersome, and the social network visibility on a graph could be cluttered. To avoid this, it is advised that expert knowledge be used to create broad subsets of the network without loosing the logical linkages between the independent networks.
2. Select the set of nodes that are flagged and reproduce the network of all possible links only pertaining to these nodes. The s*nowball* method is used to build cluster (s) recursively until no further nodes are identified. The e*go-centric* method is similar to the *snowball* method, but the analyst arbitrarily chooses to limit the size and span of the structural sample.
3. Observe visual patterns/links and generate measures for network features.
4. If needed apply mitigation measures based on the inference from SNA measures.

SNA sample implementation: dealership fraud network

Fictional case scenario: Bank Gade & Kirchner specializes in financing the fleets of various car dealers across Germany. The interest it collects from the dealers is its primary source of income. The interest rates it charges increase progressively at the end of 3, 6, 12, 24 and 36 months for each financed account, as shown in the table below.

Month	3	6	12	24	36 on-wards
Interest rate	1%	3%	4%	5%	6%

The bank identified that the yearly interest intakes had been steadily decreasing and investigated the matter further. It was discovered that most dealers closed their accounts just before the end of the 3rd and the 6th month. Further investigation showed that the cars associated with the closed accounts appeared again in new accounts opened by other dealers following a short time gap. It became obvious that there was an elaborate system of collaboration between the dealers to avoid paying higher interest rates.

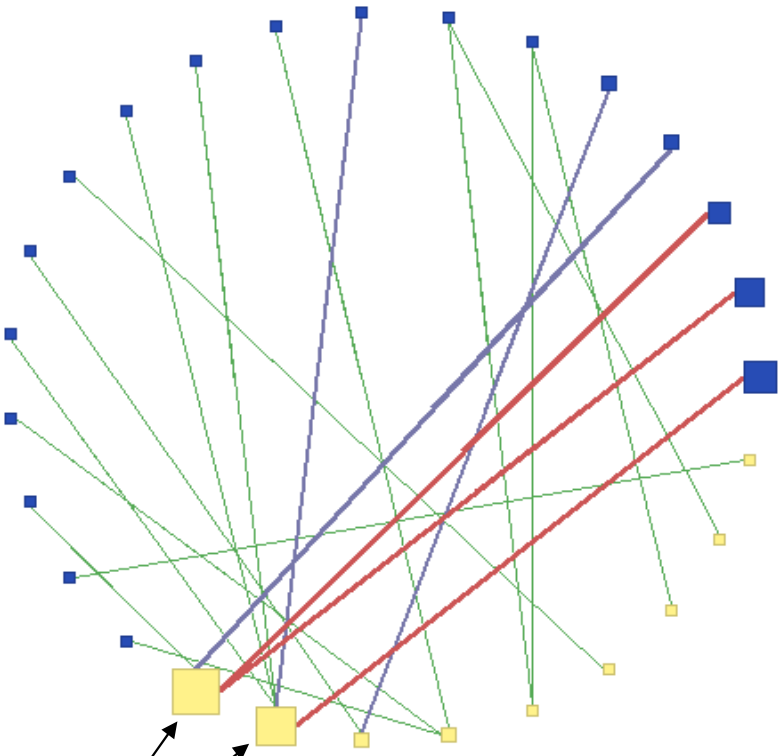
The bank identified that the yearly interest intakes had been steadily decreasing and investigated the matter further... It became obvious that there was an elaborate system of collaboration between the dealers to avoid paying higher interest rates.

The suspect dealers (nodes) were selected based on the criteria mentioned above and their linkages were recursively expanded using the snowball method to include all their accounts and the cars involved. After diffusing the account and car nodes, the fraudulent relationships between dealers were identified as clusters. Density measures were applied to the network component to identify the clusters with high fraudulent activity. The identified clusters were subjected to centrality measures to identify key actors within each cluster. As a mitigating measure the bank took a proactive step to confront the involved dealers and break up the identified fraud clusters. Future financing of all vehicles that fell into the above pattern were refused financing. By confronting only the large and obvious actors that caused most of the fraud, the bank respected genuine dealer to dealer transactions that fell into the same category.

The figures below depict the fraudulent buying and selling dealers that formed the fraud network. Analysis for this case study was done using SAS Enterprise Miner 4.2, but the current software SAS has developed is a product called SAS Social Network Analytics, which provides a more advanced user interface, profiling engine, and large scale analytics.

By confronting only the large and obvious actors that caused most of the fraud, the bank respected genuine dealer to dealer transactions that fell into the same category.

Identified major fraudulent dealers (Buyers)



Component Social Network

Identified major fraudulent dealers (Sellers)

Total detected accounts = 844
Accounts within the identified major fraud dealers = 635(approx. 75%)

Limitations of SNA

Applying SNA does come with some limitations regarding data and data processing, proactively fighting fraud and regulatory barriers. Data remodeling is required so that the effectiveness of SNA does not deteriorate as the volume of data observed increases. Database query languages like SQL though quite efficient cause significant overheads due to the joint operation performed on extremely large datasets that could increase investigation time. Transactional systems' data needs to be remodelled to avoid disambiguation, and improve data consolidation and aggregation for enhanced data availability [Goldberg and Wong 1998].

SNA inherently is retrospective in nature, i.e., one can only react to a fraudulent instance after it has occurred. Associating a transaction at runtime to SNA significantly slows transaction time, which is not desirable for most organizations. Unlike other data analytics tools and techniques, SNA is not a modelling technique. The expert knowledge of associations learned from visually observing a fraud network could be used as knowledge input to create improved analytical models.

SNA is also affected by differing regulations in various jurisdictions. Data protection laws and other regulatory measures could severely hamper SNA and could render most investigations futile as they run afoul with authorities. It is normally recommended that SNA investigation workflow processes should be checked for cross-border regulatory conformity. Recent high profile money laundering investigations have hit roadblocks due to bank secrecy legislation in Switzerland and Liechtenstein.

Software

The software applications that enable SNA are either desktop, client-based or enterprise- level server based implementations. Organizations can choose either of these two variants based on implementation complexity. Desktop applications like UCINET, etc., can be used for small scale ad hoc investigations. Enterprise-level systems like SAS Fraud Framework, which contains an SNA server, can perform extract, transform and load (ETL) operations on data and provide automation, scheduling and case management capabilities that help organizations implement end-to-end SNA based fraud prevention.

About the authors

Ms. Kirchner (christen.kirchner@cgi.com) is a lead consultant with CGI. Most recently, she has been advising an international bank within risk decision analytics and has developed SAS solutions for the calculation of its loan loss reserve. Mr. Gade (joseph.gade@cgi.com) is also a lead consultant with CGI. He advises global clients in the banking and telecom sectors in the areas of fraud prevention, as well as governance, risk and compliance (GRC).

ABOUT CGI

At CGI, we're in the business of satisfying clients. Since our founding in 1976, we've operated upon the principles of sharing in clients' challenges and delivering quality services to address them.

A leading IT and BPS provider, CGI has a strong base of 31,000 professionals operating in 125 offices worldwide. Through these offices, we offer local partnerships and a balanced blend of global delivery options to ensure clients receive the optimal combination of value and expertise required for their success.

We define success by helping our clients achieve superior performance and gain competitive advantage

CONTACT

CGI Information Systems and Management Consultants (Deutschland) GmbH
Niederkasseler Lohweg 175 – 40547 Düsseldorf – Germany
Tel.: +49 211/5355 – 0

Bibliography

1. Jesus Mena (2003). *Investigative Data Mining for Security and Criminal Detection.*

2. Kelvin Chan and Jay Leibowitz (2006). “The synergy of social network analysis and knowledge mapping: A case study,” *Int. J. Management and Decision Making*, Vol. 7, No. 1, 2006, 19.

3. Kate Ehrlich and Inga Carboni, working paper, IBM Watson Research Center.

4. S. Wasserman and K. Faust (1994). *Social Network Analysis: Methods and Applications.* Cambridge University Press.

5. H. Goldberg and R. Wong (1998). “Restructuring Transactional Data for Link Analysis in the FinCEN AI System.” AAAI Technical Report FS-98-01,
www.aaai.org.

6. Jun Karamon, Yutaka Matsuo and Mitsuru Ishizuka (2008). “Generating Useful Network-based Features for Analyzing Social Networks.” Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (2008).

7. [BKA. 2009, 2008 & 2002], Polizeiliche Kriminalstatistik (2009, 2008 & 2002).

8. [Corporate Trust 2009] Gefahrenbarometer 2010| Sicherheitsrisiken für den deutschen Mittelstand (2009).