Social Network Analysis – Extracting its Potential in Health Care Fraud Detection

Ross Kaplan
Solutions Architect for Health Care Fraud
Fraud and Financial Crimes Practice
SAS Institute Inc.
October, 2011
The National Health Care Anti-Fraud Association (NHCAA) estimates conservatively that 3% of all health care spending—or $68 billion—is lost to health care fraud (100 times credit card fraud estimates).

Other estimates by government and law enforcement agencies place the loss due to health care fraud as high as 10 percent of our nation’s annual health care expenditure—or a staggering $226 billion—each year.

…the potential losses to [healthcare] fraud and corruption could be at least €30-100 billion across Europe.  [$39 - $132 billion]
U.S. Department of Health and Human Services
“The President’s 2011 budget requests $1.7B for fighting fraud at the Department of HHS...The President’s budget will support investments in cutting-edge technology and techniques that allow for the analysis of potential fraud with unprecedented speed.”
Feb. 2010

Medicine & Health
CMS mounts offensive against home health, DME fraud
Oct 2008

The Washington Post
Sibeliaus Pledges to Fight Health Care Fraud
President Barack Obama’s choice to head the Health and Human Services Department...called for a crackdown on medical fraud as part of any health care overhaul. Mar 2009
Health Care Fraud: Paying the Price

• What does $70 - $230 billion buy these days?
  • Health Care premiums for half of the uninsured prior to March 23, 2010 in the United States
  • One quarter of the subprime mortgage losses reported by U.S. financial institutions in 2008
  • Enough to quadruple the World Health Organization’s budget and control malaria in Africa

• Bottom Line
  • Losses to fraud, waste, and abuse are too large for any payer (private or governmental) to leave on the table.
  • Payers have this money in-hand and do not have to pay it out to fraudsters.
The Perfect Storm Developing

- **Fraud, Waste & Abuse Perpetrators**
  - Far more sophisticated – organized, patient, sharing of rules
  - Leveraging multiple channels (providers & facilities) at the same time
  - Continuously evolving fraud strategies

- **Current Health Care Fraud Systems**
  - Most current detection systems act on claim level data alone
  - Investigations limited to individual members, providers and facilities
  - Focus on rules based approaches (linear and limited to known schemes)
  - Insufficient evidence to guide investigation

- **Current Health Care Fraud Operations**
  - Limited to 3rd party systems and rules
  - No real proactive steps taken to combat fraud, waste, and abuse
  - Inefficiencies driven by sheer amounts of data and disparate sources
Solution: A Comprehensive Fraud Framework

Automate the Application of Rules, Anomaly Detection, and Predictive Models to Linked, Risk-Scored Data

- **Goals**
  - Discover previously undetected entities & networks.
  - Build extensions to already-identified cases.
  - Reduce dependence on IT / informatics resources.
  - Provide desktop access to detailed data sources.
  - Reduce false positive rates.
  - Increase ROI per analyst / investigator.
Moving the Bar: Advanced Analytics are Required

Using a Hybrid Approach for Fraud, Waste & Abuse Detection

**Rules**
- Coding rules to filter fraudulent claims and behaviors
  - Examples:
    - CPT upcoding / correct coding
    - Value of payments for procedure exceeds threshold
    - Daily provider billing exceeds possible

**Anomaly Detection**
- Detect individual and aggregated abnormal patterns vs. peer groups
  - Examples:
    - Ratio of $ / procedure exceed norm
    - # procedures / provider exceeds norm
    - # patients from outside surrounding area exceeds norm

**Predictive Models**
- Predictive assessment against known fraud cases
  - Examples:
    - Upcoding behavior similar to known fraudster
    - Predicted diagnosis does not match actual
    - Similar provider / network growth rate (velocity)

**Social Network Analysis**
- Knowledge discovery through associative link analysis
  - Examples:
    - Provider association to known fraud
    - Linked members with similar suspicious behaviors
    - Suspicious referrals to linked providers

**Target Identification**

**Hybrid Approach** - Proactively applies combination of all 4 approaches at member, provider, facility, and network levels
Analytic Engine

Analytic Approach: Unsupervised Methods

- Use when no target exists
- Examine current behavior to identify outliers and abnormal transactions that are somewhat different from ordinary transactions
- Include univariate and multivariate outlier detection techniques, such as peer group comparison, clustering, trend analysis, etc...

Provider is not only an outlier, also shows extreme variation for average number of services submitted per attending provider.
Analytic Engine

Analytic Approach: Supervised Methods

- Use when a known target (fraud) is available
- Use historical behavioral information of known fraud to identify suspicious behaviors similar to previous fraud patterns
- Include parametric and nonparametric predictive models, such as generalized linear model, tree, neural networks, etc...
Fraud Analytics Components
## Limitations of standard analytical approaches

<table>
<thead>
<tr>
<th>Analytical Approach</th>
<th>Limitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rules identify known behaviors</td>
<td>While these approaches collectively provide an effective fraud analytic framework, they are all <strong>single entity focused</strong> approaches. As such, their effectiveness is limited when multiple entities are involved in <strong>collusion</strong> or single entities defraud multiple parties.</td>
</tr>
<tr>
<td>Anomaly Detection supports identification of abnormal behaviors and outliers</td>
<td></td>
</tr>
<tr>
<td>Predictive models help refine rules and identify new variables</td>
<td></td>
</tr>
</tbody>
</table>
Social Network Analysis

- Network Analysis is focused on the relations between actors while propensity or segmentation models (data mining) are based on individual actor attributes.

- Social Network Analysis (SNA) aims to identify relationships and their strength among subjects belonging to a communication network.

- Map of ties among actors provide a useful framework to identify role played by each individual within the network.

- Networks allow to define customized offering based on network roles as well as monitoring new product adoption by studying information diffusion within the network.
Social Network Analysis (SNA) aims to identify relationship among a group of individuals.

- A Network consists of one or more **Nodes** (could be persons, organizations, groups, nations)
- Connected by one or more **Ties** (could be one or more relationships)
- That form distinct, analyzable patterns (can study **patterns of relationships OR ties**)

![Diagram of a social network](image)
Sample Applications of SNA in Other Industries

- Identify market influencers for target marketing campaigns

- Identify trendsetters when new services or technologies are introduced, who influences adoption most?

- Identify telecommunications activity – who calls whom and how does it affect customer churn activity?

- Fraud in auto insurance – falsifying car accidents with no witnesses to drive revenue to certain body shops

- Fraud in government welfare benefits – cycling the same children through several day care benefits providers
Social Networks – Health Care Fraud

- Social Networks are an important paradigm within which to study relational ties.

- Fraud and Abuse is relational and inter-organizational schemes may be studied at network levels.

- Network can be a valuable tool (examples):
  - Patient-centered fraud
  - Cycling same patients for same services
  - Multi-party provider fraud, to escape outlier detection on each individual.
Analytic Engine

Social Network Analysis (Link Analysis)

• Network scoring
  – Rule and analytic-based
  – Scoring of networks & nodes
  – Suppression when appropriate

• Analytic measures of association help users know where to look in network
  – Net-CHAID for local area of interest (node) in the network
  – Density, Beta-Index (network)
  – Risk ranking with hypergeometric distribution, degree, closeness, betweenness, eigenvector, clustering coefficients (node)

• Modularity (sub-network)
Ties normalization
- summarize voice calls, sms and mms into a single ‘Intensity’ measure, representing ties strength between two nodes (msisdn)

Network splitting
- split Operator network into sub-networks using (for example) site location as splitting criteria

Community
- identify communities within sub-networks

Community measures
- computing community measures as density, average connection, diameter and structure

Node measures
- computing centrality, neighbourhood and betweenness measures

Community Sampling
- perform a stratified cluster sampling to extract a subset of nodes for identifying roles

Roles
- roles are identified using Factor and Cluster analysis on node measures

Score (Roles)
- scoring rule based on euclidean distance from cluster seeds to assign role to each network node (Operator and other operators nodes)

Conjoint Analysis
- SNA Role tuning using behavioural variables (handset, 2G and 3G services usage, etc): this applies only to Operator’s nodes

Analytical core process is a complex structured approach based on statistical and methodological steps
Scenario parameters application to SNA analytical core

**SNA Analytical core process**

Changing parameters in a scenario generates different results

Parameters applied to: **Connections normalization**

**Analysis Domain:**
Net segment → Population, time horizon, ties

Network splitting

Communities + SIM attributes

- **Community measures**
- **SNA role identification**

Parameters applied to:
- **Nodes Sampling**
- **Clustering Roles**
- **Roles Setting**
- **Discriminant Analysis**

Metadata, Process metadata, Control rules, models parameter

SNA Periodical Management

Metadata
SNA Analytical core process

- Normalization Connections
  - View of statistical distributions for type (voice, sms, mms)
  - Elimination of too high connected nodes
  - Min & max threshold to filter outliers for single kind of traffic (voice, mms, sms)
  - Min threshold to filter low number of connections
  - Min threshold to filter short duration calls
  - Weight rules for kind of traffic
  - Filters on net segment
  - Duration and time interval choices

- Node sampling
  - View of community sampling results
  - Sample size definition
    - Percentage and kind of sample

- Clustering roles
  - Cluster number definition
  - Results visualization
    - Cluster Means and Cluster Profiles

- Roles Setting
  - Roles setting by user interface

Discriminant analysis

- Variables selection for Conjoint Analysis
- Format setting for conjoint variables
  - Binning (continuous variables), Grouping: (categorical variables)
  - Missing value managing
- Weight setting to variables occurrences
  - Score assignment to occurrences combinations
  - Utility and preferences coefficients calculations
Social Network Views
Social Network-Doctor Shopping example
Social Network-DME example

http://fraud02.nas.sas.com:8280/healthcare/index.jsp?_sasapp=Social+Network+Analysis+2.1+Healthcare - Windows Internet Explorer pro

SAS® Social Network Analysis

Alert: DME PROVIDER NET 1 - 301 - NETWORK

Details Social Network Analysis

Legend

Icons Colors
Owner
SoftLink
Physician
DME Supplier
Address

Time Slider

Cumulative
Marginal

01Jul07
24Jun08
Social Network-Based Alerts

The network displays multiple doctors using similarly named DME suppliers at the same time for the same equipment.

Note the similar names (soft match) between the DME owners:
- 3_Tom Green
- 3_Richard Jones, Mary Green, Thomas Green

Note the similar names (soft match) between the 3 DME suppliers all with separate addresses:
- 2_King Medical Supply
- 2_King Health Supplies
- 3_King’s Medical Equipment

The network displays multiple doctors using similarly named DME suppliers at the same time for the same equipment.
Ad Hoc Reporting

Associative Cluster Analysis (ACA)
ACA: Detail View of Potential Collusion
ACA: Detail View of Potential Doctor Shopping
Display Relationship Among Parties

Identify / Display Relationships Among Parties Colluding in Suspicious Activity

Note: Example uses fictitious data and is for demonstration purposes only.
Summary – Value of Adding Social Network Analysis to the Mix

- Simultaneously employ multiple methods to optimize detection (more cases) and investigation (less false positives) of Fraud, Waste, and Abuse
- Identify anomalies and outliers even when fraudsters are trying to “spread the activity around”
- Understand which relationships are statistically relevant
- Prioritize claims for investigation
- Access claims / other detail independently and without programming
- Come to insights, course-of-action, and resolution significantly faster
- Spot the fastest-growing category of fraud - collusion