Graph Analysis for Detecting Fraud, Waste, and Abuse in Healthcare Data

Juan Liu, Eric Bier, Aaron Wilson, Tomo Honda, Sricharan Kumar,
Leilani Gilpin, John Guerra-Gomez and **Daniel Davies**
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA 94304

Abstract

Detection of fraud, waste, and abuse (FWA) is an important yet difficult problem. In this paper, we describe a system to detect suspicious activities in large healthcare claims datasets. Each healthcare dataset is viewed as a heterogeneous network of patients, doctors, pharmacies, and other entities. These networks can be large, with millions of patients, hundreds of thousands of doctors, and tens of thousands of pharmacies, for example. Graph analysis techniques are developed to find suspicious individuals, suspicious relationships between individuals, unusual changes over time, unusual geospatial dispersion, and anomalous networks within the overall graph structure. The system has been deployed on multiple sites and data sets, both government and commercial, to facilitate the work of FWA investigation analysts.

Introduction

Healthcare expenditures in the United States exceed $2 trillion a year. Driven by the market size, healthcare has become an important and fast growing application domain for data analytics. McKinsey's influential report on Big Data Analytics (McKinsey Corp 2012) lists healthcare as the most promising application domain. One significant problem of healthcare is the loss of healthcare expenditures to fraud, waste, and abuse (FWA). The Institute of Medicine (IOM) estimates the annual loss to FWA in the healthcare domain to be $75 Billion (PWC 2012). Healthcare-related programs such as Medicaid, Medicare, and Medicare Part C and D contribute significantly, representing more than half of the total. The magnitude of the fraud problem has attracted many efforts from the healthcare industry, the data analytics industry, and research communities to develop fraud detection systems.

Despite the substantial financial significance, the fraud detection problem is still far from being solved. While the vast amount of healthcare data (insurance claims, health records, clinical data, provider information, etc.) offers tantalizing opportunities, it also poses a series of technical challenges. From a data representation view, healthcare datasets are often large and diverse. Furthermore, they evolve dynamically over time. The complexity of the problem calls for a rich set of techniques to examine healthcare data from multiple perspectives.

Traditional fraud detection approaches start from domain knowledge to design a set of fraud detection rules and watch out for violations of these rules. This methodology works well, but its performance is inherently limited by subject matter expert knowledge, which can be inaccurate and incomplete. Furthermore, new fraud patterns are constantly invented to circumvent the built-in fraud detection rules. A different alternative, thriving due to the recent advances of machine learning and big-data infrastructure, is the data-driven methodology that identifies normal patterns from real data and detects deviations from the norm. This approach is more flexible, but computationally intense, as the search space for fraud is vast. We advocate a combined approach, where domain knowledge is used to guide the search, while data-driven machine learning methods do the rigorous computing to improve upon expert intuition to achieve better accuracy and flexibility.

To develop our overall system and the graph analysis algorithms described here we work with collaborators in Medicaid organizations and Xerox Services (which provides review and auditing services to a number of government healthcare programs and private sector health insurance companies). Our tool, known as the Xerox Program Integrity Validator (XPIV), has been deployed on multiple sites and is in use by fraud analysts in their investigation practice. The tool provides two broad categories of functionalities: (1) Automated Screening, which enables an analyst to focus attention on a small list of suspect providers, as opposed to a prohibitively large set, and (2) Interactive Drill-down, where the analyst starts from a suspicious individual or activity (as singled out by the automated screening components) and interacts with the system to navigate through data items and collect evidence to build an investigation case. The two categories have quite different technical foci: Automated Screening (1) focuses on algorithmic design for detecting diverse forms of anomalies, and Interactive Drill-down (2) focuses on database indexing/caching for fast data retrieval and user interface design for intuitive user-system interaction. For the conciseness of this paper, we do not attempt to describe the complete XPIV system, but only describe a particular subset of techniques, namely graph analysis, to detect suspicious activities and relationships. Other components of XPIV, such as outlier detection, temporal analysis, duplicate detection, and Interactive Drilldown are left out of the scope of this paper and may be discussed in follow-up publications.

Our graph analytic techniques fall into the category of Automated Screening tools, which augment analysts' abilities by helping them focus on a concise list of suspect providers. In particular, the proposed graph analytic techniques were designed with input from fraud analysts in order to detect providers whose behavior is suspect with respect to their interaction with other providers and beneficiaries in the claims database. Our graph analytic methods are the first of their kind that allow fraud analysts to detect network based fraud. Detecting network based fraud was not previously possible, because earlier systems in use by our analysts were limited to SQL-like rule-based queries.

Currently, we are moving toward more rigorous quantitative measures of system performance. However, we note that performance metrics are extremely hard to develop for real deployed systems of this size, due to the compounding effect of two factors: the significant cost of investigation, and the extreme class imbalance in which a few fraud cases are buried in the sea of regular cases. For this reason, in this paper, we resort to empirical validation, reporting cases of findings and ball-park recovery dollar amounts. As future work, we will be working with our collaborators to integrate user feedback, such as confirmation or dismissal of red-flagged cases. This will enable us to report more accurate estimates of system performance in the deployed environment. However, despite the fact that we have limited precision/recall results, our system is being widely used by the analysts for the reasons mentioned in the previous paragraph. This underlines the value of the network analytics methods that are presented in this paper.

Focus: Graph Analysis: In this paper, we describe our research effort on building graph analysis techniques. Each dataset is represented as a large and heterogeneous graph, where nodes represent millions of patients, hundreds of thousands of providers such as doctors, hospitals, and pharmacies, and edges represent billions of claimed services, medications and supplies involving multi-entity relationships among them. We apply graph analysis techniques to this dataset. Graph analysis, originally rooted in network science and graph theory, has been extended to a variety of applications such as communication networks, bioinformatics, and operations research. The recent decade has seen a rapid adoption of graph-based techniques to analyze large scale social interactions such as the World Wide Web (WWW) and social media such as Facebook, Twitter, and LinkedIn. We demonstrate that the very same set of techniques can be extended to analyze healthcare data for the detection of FWA.

We look for four types of anomalies in the graph:

- *Suspicious individuals.* We examine each individual entity (patient, provider, pharmacy, etc) based on its attributes.
- *Suspicious relationships in the graph.* While the previous type focuses on individual attributes, this type focuses on pairwise relationships. While individuals may appear perfectly normal, each out-of-norm relationship warrants a red flag.
- *Anomalous temporal changes and geospatial characteristics in the graph.* Our analysis couples graph analysis with temporal and geospatial analysis to look for unusual temporal changes or unusual geospatial distributions.
- *Structures in the graph.* Graph techniques can reveal structure, including clusters of doctors referring to each other or a heavily-connected group of individuals associated with narcotics transactions. We use graph structure analysis techniques to identify anomalous structures.

The sections to follow provide a few concrete examples of graph analysis techniques for FWA detection. Loosely speaking, graph analysis techniques fall under two categories. The first category, known as *the ego-net approach*, focuses on individual nodes and distills features from a node's local neighborhood. Features include for instance degree and entropy of local connectivities. We have developed ego-net approaches to examine narcotics relationships and temporal/spatial characteristics of patient flow between pharmacies/providers. The second category analyzes *the global structure* of the healthcare relation network and looks for communities sharing a common abnormal practice, or tight-knit communities that are anomalous in their aggregated statistics. The structural approach can identify fraud networks such as collusion networks and/or organized crime. The two categories combined together encompass both the local and the global characteristics.

Due to HIPAA restrictions (HIPAA 1996) and other business constraints, we cannot disclose full details such as personal health information (PHI) and business identities. Instead we present a high-level description, with all sample results anonymized.

Analysis of Narcotics Relationship Graphs

In this section, we illustrate graph analysis methods to detect suspicious individuals and suspicious relationships using a concrete example of narcotics use/prescription/sales. Narcotics is of concern because of the growing abuse of the medication and illicit drug trafficking. In recent years, narcotics have grown to be used recreationally by large parts of the population, and they are highly addictive (Epstein 1989). Despite federal efforts to restrict narcotics prescriptions, narcotics abuse continues to be a problem. In addition, narcotics can be illegally sold at a very high value because of the high demand and limited supply. Many people who abuse narcotics illicitly obtain them from patients with legitimate prescriptions (Radnofsky and Walker 2014), so it is important to track the individual patients that are obtaining large amounts of narcotics, as well as the doctors and pharmacies that are facilitating such diversion.

Our dataset consists of three types of entities: patient, doctor, and pharmacy. It is equivalent to a heterogeneous graph with three types of nodes. For each pairwise relationship (patient-doctor, patient-pharmacy, doctor-pharmacy), we produce a bipartite graph. Figure 1 visualizes doctor-pharmacy relationships in a real-world healthcare dataset. Red nodes are doctors, and blue nodes are pharmacies. To

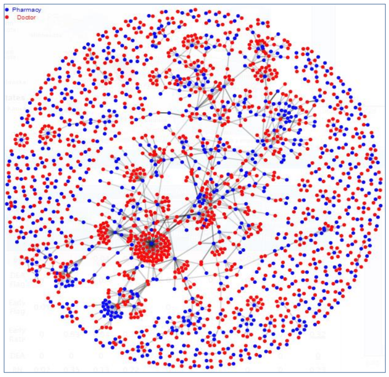


Figure 1: Bipartite graph visualizing the doctor-pharmacy relationship regarding narcotics prescription and sales.

avoid over-crowding the graph, we only visualize the top 3,000 nodes and the top 5,000 edges in terms of their narcotics amount. We use Fruchterman-Reingold, a physics-based layout, to reveal clusters of doctors and pharmacies who are connected together by heavy narcotics transactions. The graph exhibits clear patterns. For instance, it has long been suspected by fraud analysts that doctors with questionable narcotics prescription practice gravitate towards pharmacies bad at gate-keeping. In the graph, we clearly see this pattern in the provider clustering. While the system computes and displays the graph almost instantaneously, it would take an analyst many hours to perform this kind of analysis manually.

Approach: To automate detection of suspicious entities, we have designed a set of features, associated with aggregated statistics in the bipartite graphs. Given a node n and its 1-hop neighborhood $\mathcal{N}$, we have

- degree: $|\mathcal{N}|$, the number of nodes in the neighborhood;
- weight: the aggregated total number or total amount of claims that a node is associated with;
- entropy ratio: how evenly the node associates with entities in its neighborhood, in terms of total number of claims or total amount. Mathematically

$$ER_n = \frac{1}{\log(|\mathcal{N}|)} \sum_{k \in \mathcal{N}} p_k \log \frac{1}{p_k},$$

where p_k is the percentage of node n's business with neighbor k out of its total business. The summation term is the empirical entropy, measuring the dispersion of n's business among its neighborhood $\mathcal{N}$. The entropy is further divided by $\log(|\mathcal{N}|)$ to normalize to the range $[0, 1]$. If n evenly distributes its business among $\mathcal{N}$, the entropy ratio is 1. If in contrast, n does most of its business with one neighbor, the dispersion is very skewed, resulting in an entropy ratio close to 0.

Figure 2 lists the different anomalies that we look for in the relation graph. The anomalies fall into three categories: individual-level anomalies (labeled as "I"), anomalies at the relationship (edge) level (labeled as "R"), and anomalies with unexplainable medical behavior (labeled as "B"). They are shown in red, green, and blue fonts respectively. Individual-based anomalies of interest include: (I1) who are the heavy consumers of narcotics, and where do they get their drugs from; (I2) which doctors prescribe a lot of narcotics and to whom; (I3) which pharmacies sell a lot of narcotics and to whom. These questions are easy to answer based on degree and weight features.

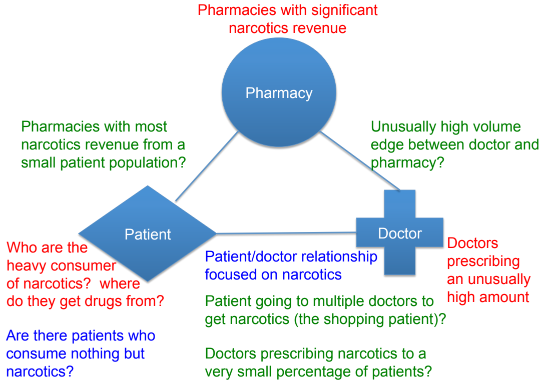

Pharmacies with significant narcotics revenue
Pharmacy
Pharmacies with most narcotics revenue from a small patient population?
Unusually high volume edge between doctor and pharmacy?
Patient
Doctor
Who are the heavy consumer of narcotics? where do they get drugs from?
Patient/doctor relationship focused on narcotics
Doctors prescribing an unusually high amount
Patient going to multiple doctors to get narcotics (the shopping patient)?
Are there patients who consume nothing but narcotics?
Doctors prescribing narcotics to a very small percentage of patients?

Figure 2: Anomalies in a narcotics relation graph

Anomalous relationships may include: Unusually focused relationship such as (R1) where a pharmacy's narcotics sales come from an unusually small number of patients and prescribing doctors; (R2) a doctor directs heavy narcotics sales to several pharmacies; and (R3) a doctor prescribes narcotics to only a few patients. High concentration between nodes can be interpreted as potential collusion. The Entropy ratio feature can be used here.

A consequence of this analysis is the ability to quickly detect fraudulent characteristics that are of interest to our users. For example, our users commonly look for (R4) "shopping patients", i.e. a patient visits a large number of doctors in order to get narcotics prescriptions. By sifting through millions of beneficiaries our algorithm can save analysts hours of manual search time.

Behavioral anomalies are those that are not justified by medical practice. These include (B1) if a patient consumes nothing but narcotics; and (B2) whether a patient/doctor relationship is focused on narcotics alone. In order to quantify these metrics, we also incorporate the patient and doctors claims outside of narcotics, and find the percentage of narcotics by dollar amount and number of total claims.

Anonymized cases under investigation: Our dataset contains medical and pharmacy claims from a state Medicaid program. It consists of roughly 64 million claim lines from 5.2 million patients, over 52,000 doctors, and nearly 9,000 pharmacies. We focus on Schedule II narcotics defined by the US Controlled Substances Act (CSA 1970). Examples of Schedule II drugs include Morphine, Oxycodone, and Fentanyl. Within the dataset, our graph analysis techniques have identified numerous suspicious activities. All findings are currently being investigated by the State's Program Integrity analysts. Here we give a few examples.

Patient P36641 is the top narcotics consumer in 2013, totaling an amount exceeding $400,000. He/she gets Fentanyl prescriptions entirely from Doctor D25542. Doctor D25542 is also the top prescribing doctor for narcotics. He/she is currently under active investigation by Medicaid's Program Integrity Office. The same analysis on 2012 data points to a top prescribing doctor who is now convicted.

Patient P96274 visits 26 different doctors for prescriptions of Methadone, Hydromorphone, and Fentanyl. The total is less than $10,000, but street value can be 50X higher.

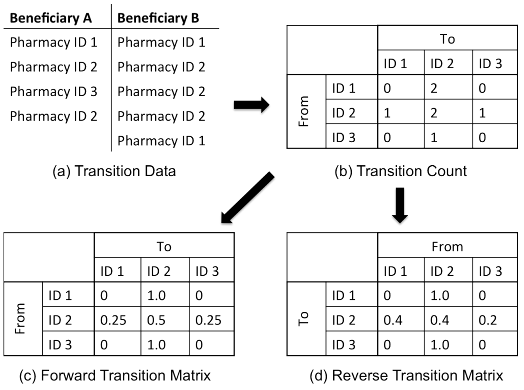
Pharmacy RX13230 has annual narcotics sales of $220,000, out of which $161,000 comes from a single doctor (Doctor D19848) for a single patient P90594. This unusually strong relationship is under investigation.

The detection of narcotics diversion can be extended to other diversion problems in healthcare with a high re-sale value, such as durable medical equipment and diabetes supplies. The same anomaly detection techniques, described here, are applicable in these domains.

Temporal/Geo-spatial Reasoning

Temporal Analysis: Interesting insights can be obtained by exploring the dynamic property of a healthcare graph. We analyzed graph's temporal characteristics to find several types of anomalies. These anomalies include sink vertices, source vertices, and heavy links. Sink vertices represent providers who attract patients from other providers at unusually high rates. Source vertices are providers who can't keep their own patients. Heavy links are graph edges where unusually strong business relationships occur. Note that these types of anomalous providers are currently manually detected by healthcare investigators. This approach will automate that effort and aid investigators to systematically search for these outliers.

We analyze the temporal characteristics by representing claims as a discrete time sequence of providers for each patient and computing transitional probabilities using Maximum Likelihood Estimation (Lee, Judge, and Zellner 1968). By comparing these transition probabilities to a baseline, we can identify source, sink, and heavy links. In this analysis, we consider the time sequence for pharmacies separately from the time sequence for physicians. Figure 3 demonstrates the computation of transitional (forward and backward) matrices from data using a simplified example. Figure 3 (a) shows the raw input data, which translates into transition counts in Figure 3 (b). By normalization with respect to the from- or the to- of the transition, one gets the forward and backward transition probabilities (Figure 3 (c,d)).



Beneficiary A	Beneficiary B
Pharmacy ID 1	Pharmacy ID 1
Pharmacy ID 2	Pharmacy ID 2
Pharmacy ID 3	Pharmacy ID 2
Pharmacy ID 2	Pharmacy ID 2
	Pharmacy ID 1

(a) Transition Data

		To		
		ID 1	ID 2	ID 3
From	ID 1	0	2	0
	ID 2	1	2	1
	ID 3	0	1	0

(b) Transition Count

		To		
		ID 1	ID 2	ID 3
From	ID 1	0	1.0	0
	ID 2	0.25	0.5	0.25
	ID 3	0	1.0	0

(c) Forward Transition Matrix

		From		
		ID 1	ID 2	ID 3
To	ID 1	0	1.0	0
	ID 2	0.4	0.4	0.2
	ID 3	0	1.0	0

(d) Reverse Transition Matrix

Figure 3: Sample Computation for Transition Matrices.

Anonymized cases under investigation: Our analysis shows that most patients return to the same pharmacy repeatedly and rarely deviate from their pattern. More than 80% of prescriptions are filled at the same pharmacy where the previous prescriptions are filled. By comparison to this baseline, two different types of source are detected by our algorithms. The first type of source tends to lose patients to another specific pharmacy. For example, our analysis identified two pharmacies where 85% of the source's business is later transferred to the sink. This is particularly unusual given that these two pharmacies are 500 miles apart. An example of this kind is worth further investigation to determine if the business relationship between the source and sink represent truly fraudulent behavior. Interestingly, some pharmacies with prior fraud convictions have shown up to be anomalous again for this analysis. The second type of source consists of pharmacies who spread their patients to many different pharmacies. These source pharmacies may not necessarily be involved in FWA activities, but could be losing customers due to poor quality of service.

Geospatial Analysis: Geo-spatial data are another useful source of information for anomaly detection. We assume that most patients visit physicians and pharmacies in their local cities. Note there are many benign but infrequent reasons why patients might visit providers far from home, e.g., a) sickness or injury during travel, and b) visiting specialists like a surgical oncologist for special treatment. We focus on outlier detection methods using aggregated statistics as features to help remove the effect of these rare events.

We compute the geographical distance between the physician-pharmacy pair and derive an empirical cumulative distribution function (cdf) (Mason 1982). Typically the empirical cdf increases sharply over distance. For example, a pharmacy or physician's business relationships are 50% within a 10-mile radius, 80% within a 20-mile radius, 90% within a 30-mile radius, and so on. The dashed lines in Figure 4 show a set of cdfs at different percentiles. We apply DBSCAN (Ester et al. 1996), a clustering algorithm, to the empirical distributions to define the baseline. Cdfs that deviate drastically from the norm are identified as anomalies.

Anonymized cases under investigation: The thick black line in Figure 4 shows an anomalous cdf of a pharmacy, where 42% of its business comes from a physician over 400 miles away. In addition to the long distances traveled by visiting patients, the fact that all long distance prescriptions come from this single physician is abnormal, which could be an interesting finding in its own right.

Discovering Latent Networks of Providers Sharing Anomalous Practices

In this section we discuss the discovery of heterogeneous provider communities that share anomalous business practices. In particular, we consider extracting communities of prescribing providers that are participating in anomalous drug sales. Within such a community, each individual provider's specialty will determine the kinds and quantity of the prescriptions they write. A cardiologist's prescriptions will be composed of a high proportion of heart disease re-

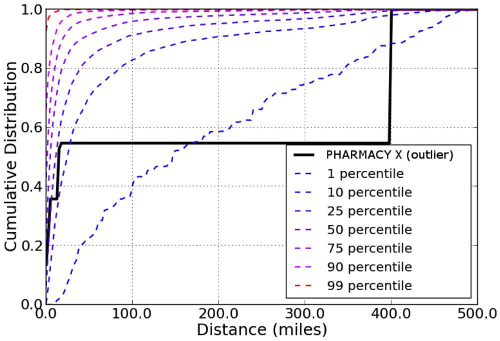


Figure 4: Example of Geospatial Anomaly

lated medications whereas an oncologist will tend to prescribe a high proportion of chemotherapy drugs. We aim to simultaneously discover provider types while detecting when the prescription behaviors of heterogeneous provider communities are anomalous. For instance, a hypothetical cardiologist and oncologist may be interacting with a pharmacy to sell narcotics to addicted patients. While the majority of their individual prescription sales are consistent with their types, composed of heart disease and chemotherapy drugs respectively, the narcotics sales represent a shared deviation from those types. In order to find these communities we need a concrete definition of a provider's type and a means of exploiting this type definition to find anomalous communities in our graph.

Approach: We model providers as documents and employ Latent Dirichlet Allocation (LDA) to extract a set of business topics (Blei, Ng, and Jordan 2003). The basic generative process is illustrated in Figure 5. Each provider in the network is a document w_p composed of a bag of words. Each word in a provider document is one of several Hierarchical Ingredient Code List (HICL) codes that identify the compounds of drugs prescribed by a provider. We define the count of a word within a document to be the total reimbursement amount. Given a set of documents our goal is to extract a set of business profiles, or topics, that explain the kinds of prescription combinations that are typical across providers. LDA defines a joint distribution over the set of provider documents and was used to infer the set of business topics. The distribution of provider documents is defined to be,

$$P(\mathbf{w_p}|\alpha,\beta) = \int P(\theta_p|\alpha) \left(\prod_{v\in V} \prod_{n=1}^{N_v} \sum_{z_{v,n}} P(z_{v,n}|\theta_p) P(w_{v,n}|\beta, z_{i,j}) \right).$$

Here the parameter θ encodes the mixture proportions over topics, β is the collection of prescription topics, $z_{p,v,n}$ indicates which topic generated word $w_{p,v,n}$, and α is a prior parameter controlling the sparsity of θ. We have slightly modified the typical expression for LDA to explicitly specify the word types v and their corresponding counts N_v. We will exploit this later when introducing our community anomaly score. By modeling provider documents as a mixture of topics, LDA more easily models the distinct categories of beneficiaries treated by specialists. We find this flexibility is integral to accurately modeling pharmacies which receive beneficiaries from many different specialties.

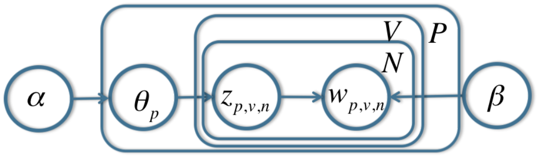


Figure 5: A plate model depicting the distribution over P provider documents. Each provider document decomposes into a set of words $w_{p,v,n}$ which is the nth instance of word type v in provider document p. Associated with each word is a variable $z_{p,v,n}$ indicating which topic $\beta_k \in \beta$ was responsible for generating the word. The parameters θ_p are the topic proportions which determine how frequently a provider uses a specific topic.

Given models of individual business practices we seek to define a notion of community. We define a collection of communities to be a decomposition of the graph into a set of connected components. In order to differentiate communities, we introduce the important concept of a community "color". The color of a community is a subset of word types, represented by a sparse indicator vector, identifying the community's anomalous drug sales. To identify these colors we begin by decomposing the provider documents into two sets. One set contains the collection of prescriptions made within the community. The second set contains the prescriptions sent to members outside of the community (within other connected components of the graph). Determining the community color is a problem of selecting the optimal subset of within-community prescription events that, once removed, maximize the community score. The community score is defined simply to be the log-likelihood of the community members after removing the indicated words. This score is given by the following equation,

$$P(\mathbf{C}|\alpha,\beta,\phi) = \prod_{p\in C} \int P(\theta_p|\alpha)$$
$$\left(\prod_{v\in V} \prod_{n=1}^{N_v - \phi_i N_{c,p,v}} \sum_{z_{p,v,n}} P(z_{p,v,n}|\theta_p) P(w_{p,v,n}|\beta, z_{p,v,n}) \right),$$

where we have introduced the indicator variables ϕ_v and the within-community word counts $N_{c,p,v}$. Here $N_{c,p,v}$ refers to the number of times word type v was communicated by provider p to members of the community. In order to set the indicator variables we developed a simple greedy procedure that selects the highest scoring indicator at each step subject to a bound on the number of non-zero values. The procedure removes the subset of words that are most poorly modeled by the available topics.

Finding the optimal decomposition of the graph is a computationally demanding task due to the large number of possible graph colorings. Therefore, we propose a simple agglomerative clustering procedure that seeks to iteratively improve the joint likelihood score by merging adjacent communities. To accomplish this we define a merge score that compares the log-likelihood of the current communities against the log-likelihood of the merged community. Using this score we designed a simple greedy agglomeration procedure that is guaranteed to find a partitioning of the graph. The result is an efficient search for an approximate solution to the optimal community coloring problem. As output we

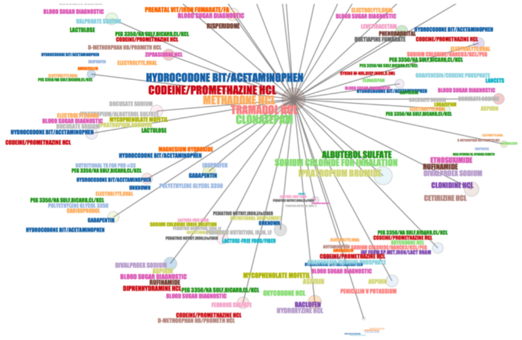


Figure 6: Anomalous communities discovered by the analysis.

receive the collection of communities and their corresponding colors for presentation to analysts.

Anonymized cases under investigation: In Figure 6 we show example results generated by our community discovery procedure. Given a network of approximately 74,000 providers with more than 900,000 prescription relationships our algorithm discovers 900 communities of varying sizes. In the figure we illustrate a subset of discovered communities with a particular outlier at the center of the star. Each node in the depicted graph represents a community discovered by our algorithm. The size of the node represents the degree of anomalousness where larger nodes are communities with highly-anomalous shared behaviors (measured by the increase in the log likelihood after removing the indicated words). Each community is described by its most anomalous drug sales. In this experiment we have constrained the set of indicators such that only five words can be selected during the search. We see that the central community, composed of 17 medical providers, is responsible for over $170,000 of anomalous narcotics sales (e.g., Hydrocodone, Codeine, Methadone, Tramadol, and Clonazepam). This fact is extracted directly from the high dimensional data set and represents valuable information for a FWA analyst. The goal of the visualization is to make this information immediately accessible to the analyst and to highlight the *reason* that the community stands out. Our experience shows that simple rankings of communities by scores is not sufficient to promote analyst exploration whereas carefully designed interpretable outputs provide an accessible representation of the analysis.

Discovering Anomalous Structure in the Graph

In this section, we report our work-in-progress on a non-parametric approach to discovering anomalous communities in the medical network. We assume that we are given an arbitrary input graph G with nodes being entities such as providers, hospitals, pharmacies and patients, and the edge attributes reflecting the strength of interaction between the nodes. For concreteness, in this paper we consider the specific case of referral networks where the graph G is comprised of provider nodes, and the links between nodes a and b represent the total number of referrals between providers a and b. Figure 7 shows an example referral network from a real-world dataset using the Group-in-a-Box visualization (Rodrigues et al. 2011) to highlight communities.

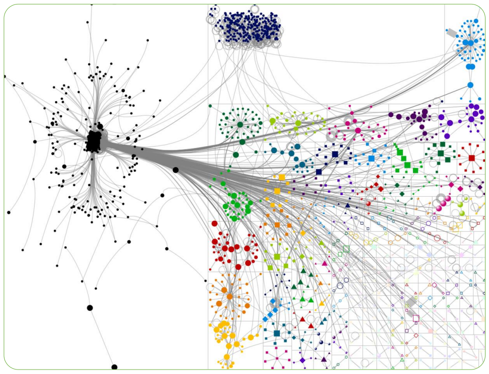


Figure 7: An Example Referral Network using the Group-in-a-Box to highlight communities.

Given this input graph G, we are interested in identifying subsets of communities that are anomalous. We do this in a three stage process:

1. Identification of communities in G
2. Extraction of features characterizing these communities
3. Identification of anomalous communities using these multivariate feature representations of these communities

We discuss each of these steps in detail in the sequel.

Community extraction: As a first step, we extract tight-knit communities in the graph G. Community detection in a graph is a widely studied problem in the network data mining literature. However, most of the popular methods such as graph partitioning, hierarchical clustering and spectral clustering are concerned with partitioning the graph into disjoint sets of tight-knit nodes (Fortunato 2010). These partitioning methods however are not a good fit in our particular context of medical networks for the reason that the entire graph G need not be partitionable into tight-knit communities; rather we expect a few pockets of tightly-knit communities interspersed in the graph.

As a consequence of this observation, we developed an agglomeration-based partitioning scheme that only identifies the small pockets of tight-knit communities as opposed to completely partitioning the set of nodes into disjoint subsets. The proposed agglomeration scheme works by building communities one node at a time in a greedy fashion, and adding nodes to the communities while ensuring that the communities remain tightly-knit. We denote the set of communities extracted from G by $\bar{C} = \{C_1, C_2, \ldots, C_k\}$.

After extracting the set of communities $\bar{C}$ via the proposed agglomeration scheme, we check to see if any of the extracted communities are anomalous. We do this in two steps. As a first step, we check the case where the very existence of communities is anomalous. To check this case, we compute the ratio of the total number of nodes in $\bar{C}$ relative to the total number of nodes in G. Conceptually this ratio is similar to the well-known graph modularity metric proposed by Newman (Newman 2006), except that this ratio is defined based on nodes, and the graph modularity metric is defined on edges. If the ratio is very small, it indicates that G is a network that is largely community-free, and we therefore declare that all the discovered communities in $\bar{C}$ are anomalous. On the other hand, if the ratio of the number of nodes in $\bar{C}$ relative to G is moderate to large, then we conclude that the presence of a community in G does not indicate that the community is anomalous.

Feature extraction: In the event of the latter scenario being true, we extract several features that are of interest in order to characterize each of these communities and subsequently we look for communities that are anomalous with respect to the extracted feature sets. In this paper, we consider the following sets of features to characterize any given community C_i in the referral network:

- Community size: the number of nodes in C_i.
- Community density: the ratio of the total number of edges in C_i relative to the number of nodes.
- Average dollar amount: the ratio of the total dollar amount associated with the referral edges in C_i relative to the number of nodes.
- Average anomaly score: Independent of network analytics, we compute anomaly scores for all providers based on marginal statistics and compute the average anomaly score of a community C_i as the average of the anomaly scores of all providers in the community in order to detect if a community has an abnormal concentration of anomalous providers.

Anomaly detection: In our final step, we extract anomalous communities using these features by feeding the features through an off-the-shelf anomaly detection method for multivariate data. In this paper, we use the iForest anomaly detection algorithm (Liu, Ting, and Zhou 2008), which is currently the state-of-the-art. The iForest method detects anomalies based on the difficulty of isolating a point from the rest of the points using randomly generated classification trees. The intuition is that an outlier point is far easier to isolate than normal points.

Anonymized cases under investigation: We applied the described procedure to a referral network with about 60,000 providers. On running our agglomeration-based partitioning algorithm, we discovered a total of 2,432 communities. These 2,432 communities accounted for about 40,000 providers, or about 66% of the total nodes in the network. Thus, the presence of communities in this network is not anomalous.

Subsequently, we extract community size, density, average dollar amount and average anomaly score as features for anomaly detection. On running iForest, we discovered a total of 34 anomalous communities. Five communities were flagged because of their large size. Each of these communities had in excess of 200 providers/community, while a majority of the communities had an average of about 10 providers. 10 other communities were flagged for high density, another 12 were flagged for high dollar amount and the remaining 7 were flagged for high anomaly scores. An interesting observation was that some of the communities were anomalous with respect to more than one feature. For instance, one particular community which had about 400 providers also had an abnormally high density.

Graph Analytics in Real-world FWA Detection

We have deployed our analytics system to support several business applications to detect fraud, waste, abuse and other kinds of inappropriate billing. These applications include Provider Review, Cost Containment, Recovery Services, and Pre-pay Detection.

The goal of *Provider Review* is to find providers (doctors, hospitals, clinics, etc.) who are billing inappropriately and who will be the most valuable to audit, judging by the amount billed, the degree to which the billing is inappropriate, and other factors such as the extent to which patient health is endangered. Analysis aims to maximize a value function over providers or sets of providers.

The goal of *Cost Containment* is to find a proposed change to the current claim payment rules that is likely to result in increased efficiency, decreased cost, or improved healthcare outcomes. These opportunities focus less on individual providers, patients, or claims and instead focus on a set of these. Here analysis aims to find billing patterns that are common and expensive but inappropriate.

The goal of *Recovery Services* is to find individual claims where more money was paid than should have been and then to contact the associated providers and get money back. For example, a Recovery Services call center may ask a provider to refund money if the provider was accidentally paid twice for the same service, or if another insurance company should have been billed first. Analytics for Recovery Services focuses on over-billing that can be proven easily and then tries to find as many instances as possible.

The goal of *Pre-pay Detection* is to identify inappropriate claims before the provider is paid for those claims. For any given claim, a pre-pay algorithm determines if the claim should be rejected, sent to a human analyst for further study, or processed normally.

We work with teams that provide services organized around the business applications mentioned above. In that work, we use our deployed system to provide analytics reports and interactive software that can be shared with analysts performing provider reviews, cost containment, and recovery. Our partner teams, through their interaction with the deployed system, give us feedback on algorithms, reports, and software, allowing us to improve them iteratively. In addition, improvements made to support one team often support others. Our analytics have already been used to find many overpayments including Provider Review and Cost Containment cases with a potential value of *several million dollars* and recoverable claims with a potential value of roughly *a million dollars per month.*

Our graph analytics support three of the four kinds of service. For Provider Review, one way to find suspicious providers is to look at the graph of relationships between providers, such as patient referrals and shared patients. If providers are colluding to defraud the system, that will show up in this graph. Likewise, providers and patients may collude to bill insurance payers for drugs or supplies and then sell them on the street. In Cost Containment, a provider billing too much for one patient will often bill too much for other patients as well. Patterns in the provider-patient network, then, can uncover systemic over-billing that can be addressed by a rule change. In Pre-pay Detection, when making a decision about a new claim, the algorithm can look at patient-provider, provider-provider, and patient-patient relationships together with information about particular providers, patients, and claim features that have been associated with over-billing in the past to recommend human review of some claims.

Evaluation Challenges

As we create our algorithms, we evaluate them against three levels of benefit: (1) productivity benefits, (2) human-level quality on results with a reduced detection time, and (3) greater-than-human level quality on results. Our evaluations run from informal, such as user testimonials, to formal, such as calculation of precision in finding overpayments.

For example, one Xerox partner wrote "Using these technologies will improve the selection of audit targets which has a direct impact to revenue on these contingency based contracts." Another Xerox partner wrote "Interesting flag. ... So it has a high positive hit rate at first pass." and also wrote "In the first 5 minutes I identified a possible referral ..." and also "Without [this tool], it would have been very difficult and quite time consuming to do this research".

On the more formal side, we have been fortunate to have analysts who are willing to go through large results sets, including thousands of flagged health care claims, to see which are or are not recoverable. For example, after several iterations of improving duplicate detection, we were able to get 100 percent precision on a first result set based on criteria set by the analysts. As these evaluations indicate, our tools and algorithms have been able to improve user productivity and allow users to produce results that were difficult or time-consuming to produce previously.

These statements speak to the impact of the system from the point of view of analysts. As our program continues to develop we plan to augment this analysis to include additional measurements of system quality. For instance, a crucial measurement in fraud detection is the rate of case identification for individual analysts. An ideal system increases this rate.

Our initial evaluations, though preliminary, suggest that our tool successfully improves work flows. Our future effort will determine the magnitude of this improvement. In addition, in coordination with our business partners we continue to construct larger sets of ground truth data that are crucial for preliminary evaluation of new analytics. We expect, in coming years, to establish empirically the robustness of our deployed system.

Conclusion

This paper presents our work on developing graph analysis techniques and applying them to real-world healthcare datasets to look for fraud, waste, and abuse activities. We represent the healthcare relationship using heterogeneous graphs and identifying anomalous individuals, relationships, and communities by analyzing the local and global characteristics of the graphs. Our work has identified investigation targets totaling millions of dollars of potential recovery for our collaborators at Xerox Services.

Our future work will take several forms. First, we plan to extend our graph analysis techniques to scan incoming claim streams fast enough to intercept suspicious claims before they are paid. This early detection requires the graph analysis algorithms to be optimized for memory and computation, running quickly on large graphs. In addition, we plan to add additional feedback loops to our system, so that actions taken by users of our technologies become input to the algorithms. This will enable a rigorous performance evaluation on the detection precision. At the same time, the algorithms will learn from the suspicious activities that users explore and mark, and the results of audits, investigations and recoveries. Finally, we will allow users to configure the analytics so that it is easy to tune them to the needs of specialists and repeat successful analyses on new data sets.

Acknowledgments. We'd like to acknowledge technical contributions from our colleagues John Hanley, Alan Bell, Sureyya Tarkan, Alex Brito, Ming Yang, and Nick Briggs.

References

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3:993–1022.

CSA. 1970. Title 21 CFR 1308.12. US Department of Justice, Drug Enforcement Adminstration, http://www.deadiversion.usdoj.gov/21cfr/cfr/1308/1308-12.htm.

Epstein, R. 1989. Drug wars in the united states. *British Medical Journal* 299(6710):1275–1276.

Ester, M.; Kriegel, H.-P.; Sander, J.; and Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *The Second International Conference on Knowledge Discovery and Data Mining*, volume 96, 226–231.

Fortunato, S. 2010. Community detection in graphs. *Physics Reports* 486(3):75–174.

HIPAA. 1996. Health Insurance Portability and Accountability Act of 1996, Public Law 104-191.

Lee, T. C.; Judge, G. G.; and Zellner, A. 1968. Maximum likelihood and bayesian estimation of transition probabilities. *Journal of the American Statistical Association* 63(324).

Liu, F. T.; Ting, K. M.; and Zhou, Z.-H. 2008. Isolation forest. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, 413–422. IEEE.

Mason, D. M. 1982. Some characterizations of almost sure bounds for weighted multidimensional empirical distributions and a glivenko-cantelli theorem for sample quantiles. *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete* 59(4):505–513.

McKinsey Corp. 2012. Big data: the next frontier for innovation, competition, and productivity. *McKinsey Global Institute Report*.

Newman, M. 2006. Modularity and community structure in networks. *Proc. Natl Acad Sci USA* 103(23).

PWC. 2012. The price of excess: identifying waste in healthcare spending. *PricewaterhouseCoopers (PWC) Health Research Institute Report*.

Radnofsky, L., and Walker, J. 2014. DEA restricts narcotic pain drug prescriptions. *Wall Street Journal*.

Rodrigues, E. M.; Milic-Frayling, N.; Smith, M.; Shneiderman, B.; and Hansen, D. 2011. Group-in-a-box layout for multi-faceted analysis of communities. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, 354–361.