Graph Analysis for Detecting Fraud, Waste, and Abuse in Health-Care Data

Juan Liu, Eric Bier, Aaron Wilson, John Alexis Guerra-Gomez, Tomonori Honda, Kumar Sricharan, Leilani Gilpin, and Daniel Davies

■ *Detection of fraud, waste, and abuse (FWA) is an important yet challenging problem. In this article, we describe a system to detect suspicious activities in large health-care data sets. Each health-care data set is viewed as a heterogeneous network consisting of millions of patients, hundreds of thousands of doctors, tens of thousands of pharmacies, and other entities. Graph-analysis techniques are developed to find suspicious individuals, suspicious relationships between individuals, unusual changes over time, unusual geospatial dispersion, and anomalous network structure. The visualization interface, known as the network explorer, provides a good overview of data and enables users to filter, select, and zoom into network details on demand. The system has been deployed on multiple sites and data sets, both government and commercial, and identified many overpayments with a potential value of several million dollars per month.*

Health-care expenditures in the United States exceed $2 trillion a year. Driven by the market size, health care has become an important and fast growing application domain for data analytics. McKinsey's influential report on big data analytics (Manyika et al. 2011) lists health care as the top most promising application domain. One significant problem of health care is the loss of health-care expenditures to fraud, waste, and abuse (FWA). Figure 1 lists the amount of improper payment in U.S. government expenditure. In 2012, improper payments totaled about $120 billion. Health-care-related programs such as Medicaid, Medicare fee for service (FFS) and parts C and D contribute significantly, representing more than half of the total. A separate report from the Institute of Medicine (IOM) estimates the annual loss to FWA in the health-care domain to be $750 billion (Institute of Medicine 2012). The magnitude of the fraud problem has attracted many efforts from the health-care industry, the data-analytics industry, and research communities to develop fraud-detection systems.

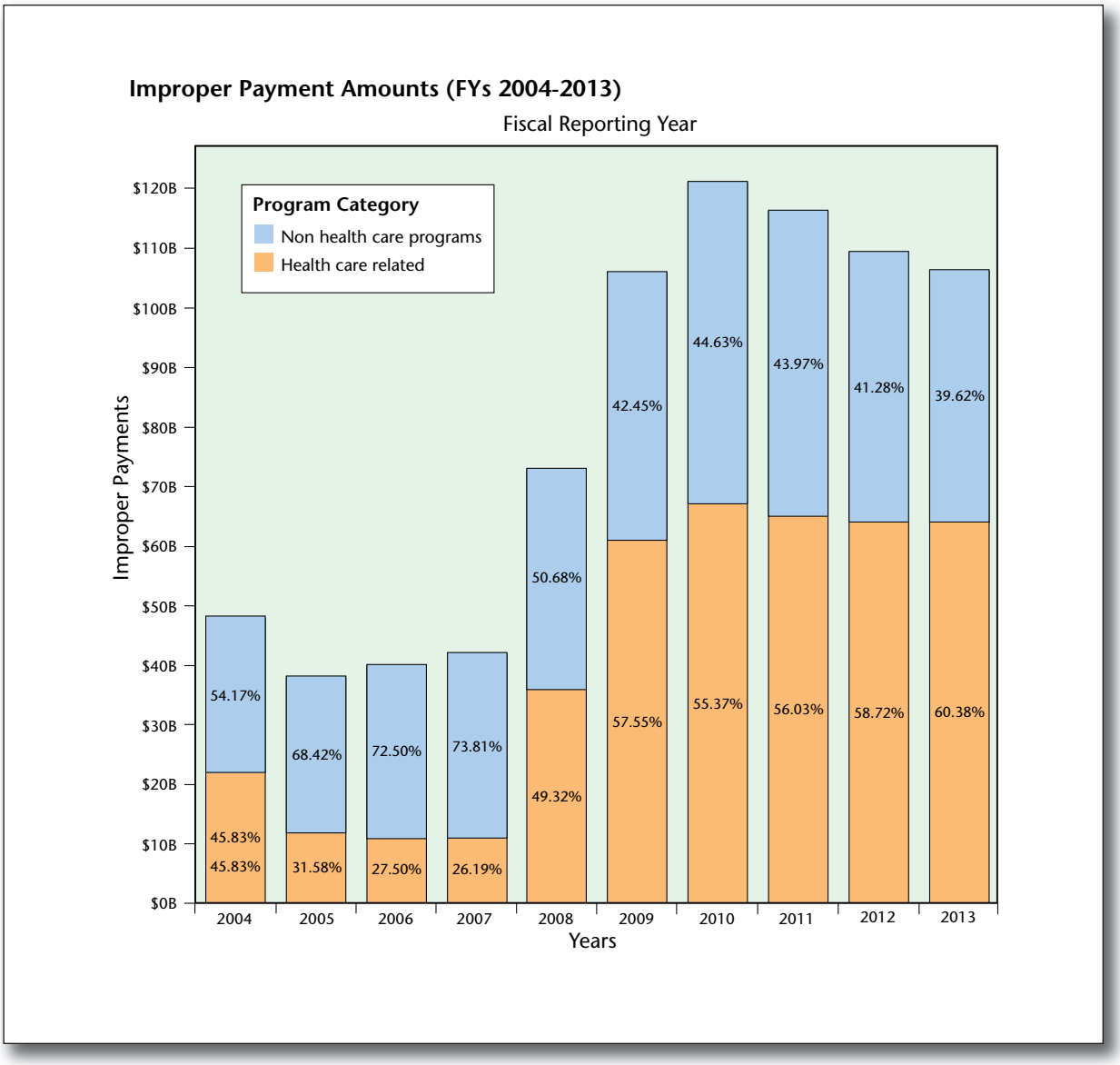

Figure 1. Improper Payments in Government Expenditure.

Health-care-related programs include federal and state government programs such as Medicaid, Medicare Advantage (Part C), Medicare FFS, and Medicare Prescription Drug Benefit (Part D). Non-health-care programs include Earned Income Tax Credit (EITC), Pell Grants, Public Housing/Rental Assistance, Retirement, Survivors and Disability Insurance (RSDI), School Lunch, Supplemental Nutrition Assistance Program (SNAP), Supplemental Security Income (SSI), Unemployment Insurance (UI), and other programs. *Source:* www.paymentaccuracy.gov.

Despite the substantial financial significance, the fraud-detection problem is still far from being solved. While health-care data (insurance claims, health records, clinical data, provider information, and others) offers tantalizing opportunities, it also poses a series of technical challenges. From a data representation view, health-care data sets are often large and diverse. It is common to see a state's Medicaid program or a private health-care insurance program having hundreds of millions of claims per year, involving millions of patients and hundreds of thousands of providers of various types, for example, physicians, pharmacies, clinics and hospitals, and laboratories. Any fraud-detection system needs to be able to handle the large data volume and data diversity. Furthermore, health-care programs and perpetrators are in the fraud arms race and constantly make adjustments to gain a competitive advantage over the other side. Data patterns from both sides are dynamic. The complexity of the problem calls for a rich set of techniques to examine health-care data.

Health-care financials are complex, involving a multitude of providers (physicians, pharmacies, clinics and hospitals, and laboratories), payers (insurance plans), and patients. To design a good fraud-detection system, one must have a deep understanding of the financial incentives of all parties. Terry L. Leap's book on medical billing (Leap 2011) provides a nice overview of each type of providers, their financial practices, and common fraud activities. Starting from domain knowledge, auditors and investigators have designed fraud-detection rules to watch out for false claims. The rules cover diverse behaviors, for instance, total sum of medical service hours claimed for any given day close to or exceeding 24 hours, work hours during short holiday weeks or under adverse weather conditions staying high, unusually high proportion of home visits, and billing of service to patients who are already deceased. Readers may refer to Rebecca S. Busch's *Healthcare Fraud: Auditing and Detection Guide* (Busch 2012), for a comprehensive list of example rules. This methodology of comparing data against a predefined rule set is widely adopted in auditing practice and works effectively, 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 outliers deviating from the norm. Various approaches have been developed for outlier detection, such as density-based approaches to identify points in low probability regions, proximity-based approaches to identify points isolated from others, subspace-based methods to identify rare classes, and supervised or semisupervised learning methods to learn differences between normal and abnormal classes (Aggarwal 2012). Compared to their rule-based counterparts, data-driven approaches are 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, we work with Xerox Services, which provides review and auditing services to a number of government health-care 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 article, 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 and duplicate detection, are left out of the scope of this article and may be discussed in follow-up publications.

Graph Analysis

We have worked with fraud analysts to understand real needs in their investigation effort. One common concern that the analyst would like to have help with is to detect organized crimes. Recent years have witnessed crime rings migrating from illegal drug trafficking to the safer and far more lucrative business of perpetrating frauds against health care. The National Health Care Anti-Fraud Association (NHCAA) has reported that, in Florida alone, government Medicare and Medicaid programs and private health-insurance companies have lost hundreds of millions of dollars in recent years to criminal rings.[1] Collusion among dishonest practitioners has also become common, with fraudulent activities such as self-referrals, false or unnecessary referrals, and kickbacks. Patients may be involved as well. A 2013 investigative series by the Center for Investigative Reporting and CNN[2] uncovered rampant overbilling in California's publicly funded drug rehabilitation system and fraud schemes such as clinics recruiting homeless to pose as patients. It is reported that California paid $94 million in fiscal years 2012 and 2013 to clinics that have shown signs of deception or questionable billing.

To detect crime rings and collusion networks such as the ones mentioned above, we need graph-analytics methods to examine data points in relation to others.

Compared to anomaly detection schemes, which focus on attributes of individuals, examining the relational aspects provides a new perspective. Our system is the first of its kind that allows fraud analysts to detect network-based fraud. Each data set is represented as a large and heterogeneous graph, where nodes represent millions of patients and hundreds of thousands of providers, such as doctors, hospitals, and pharmacies, and edges represent billions of claimed services, medications, and supplies involving multientity relationships among them. In this article, we describe the technical components.

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 and social media such as Facebook, Twitter, and LinkedIn. We have extended these techniques to analyze health-care data for FWA detection. In particular, we look for four types of anomalies in the graph: (1) suspicious individuals, (2) suspicious relationships, (3) anomalous temporal changes and geospatial characteristics, and (4) structures.

Suspicious individuals. We examine each individual entity (patient, provider, pharmacy, and so on) 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 that 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 and spatial characteristics of patient flow between pharmacies and providers. The second category analyzes the global structure of the health-care 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 or organized crime. The two categories combined together encompass both the local and the global characteristics. We also briefly describe network explorer, a visualization and user interface that supports an eagle-eye view of the entire network and an interactive drill down of suspicious nodes within its local ego-net.

Automated screening faces the technical challenge of balancing false alarms and missed detection. An accurate characterization of performance, such as a ROC curve or precision or recall metric, would be great. However, we note that the performance metrics are extremely hard to measure due to the high cost of investigation and the extreme class imbalance. Each investigation case bears a cost, ranging from $200 for a simple desk audit to $20,000 for a typical crime investigation. The approach of subsampling a set and labeling each data point to obtain the precision or recall metric, though common in academic studies, is infeasible here, as it would require labeling a very large set (due to class imbalance in which a few fraud cases are buried in the sea of regular cases) and hence incur a prohibitively large investigation cost. This is an inherent drawback of the fraud-detection application domain. For practical business reasons, we have designed our system to produce high precision and low recall. We resort to empirical validation, reporting cases of findings and ballpark recovery dollar amounts. Though still preliminary, our system is being widely used by the analysts to focus their investigation effort. This underlines the value of the network analytics methods presented in this article. Currently we are working with our collaborators to integrate user feedback, such as confirmation or dismissal of red-flagged cases to obtain more rigorous precision metrics.

Due to HIPAA restrictions[3] 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, and sales. Narcotics is of concern because of the growing abuse of medications and illicit drug trafficking. In recent years, narcotics have grown to be used recreationally, 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 data set 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 2 visualizes doctor-pharmacy relation-

ships in a real-world health-care data set. Red nodes are doctors, and blue nodes are pharmacies. To avoid overcrowding the graph, we only visualize the top 3000 nodes and the top 5000 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 practices gravitate toward pharmacies bad at gatekeeping. 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; and 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 3 lists the different anomalies that we look for in the relation graph. The anomalies fall into three categories: individual-level anomalies (labeled I), anomalies at the relationship (edge) level (labeled R), and anomalies with unexplainable medical behavior (labeled B). They are shown in various shades of gray or (if in color) red, green, and blue fonts respectively. Individual-based anomalies of interest include: (I1) who are the heavy consumers of narcotics, and where they get 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.

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.

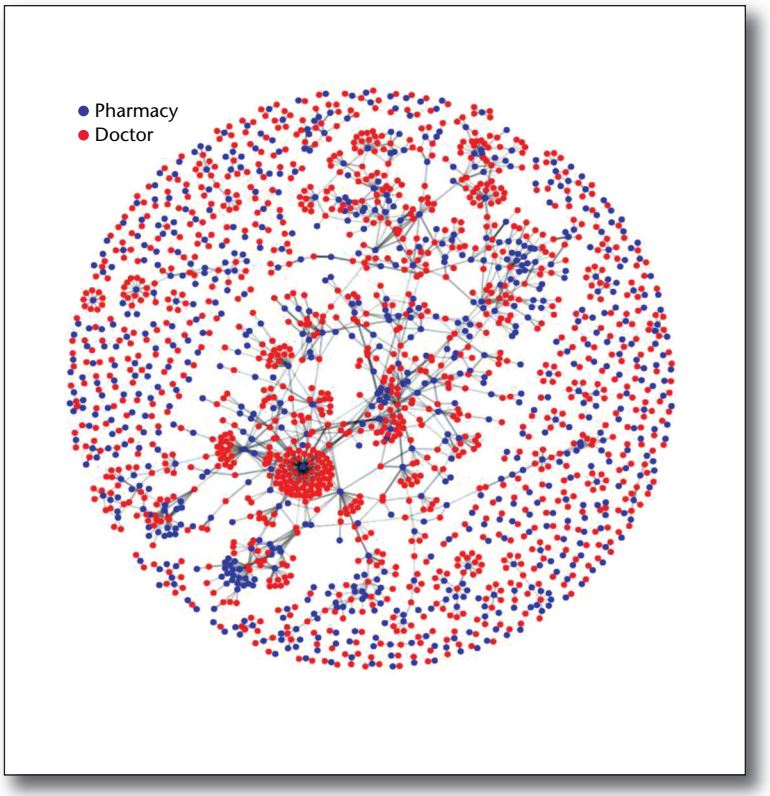

Figure 2. Bipartite Graph.

This graph visualizes the doctor (red/ light gray) – pharmacy (blue / dark gray) relationship regarding narcotics prescriptions and sales.

A consequence of this analysis is the ability quickly to detect fraudulent characteristics that are of interest to our users. For example, our users commonly look for (R4) "shopping patients," that is, 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 patients' 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 data set contains medical and pharmacy claims from a state Medicaid program. It consists of roughly 64 million claim lines from 5.2 million patients,

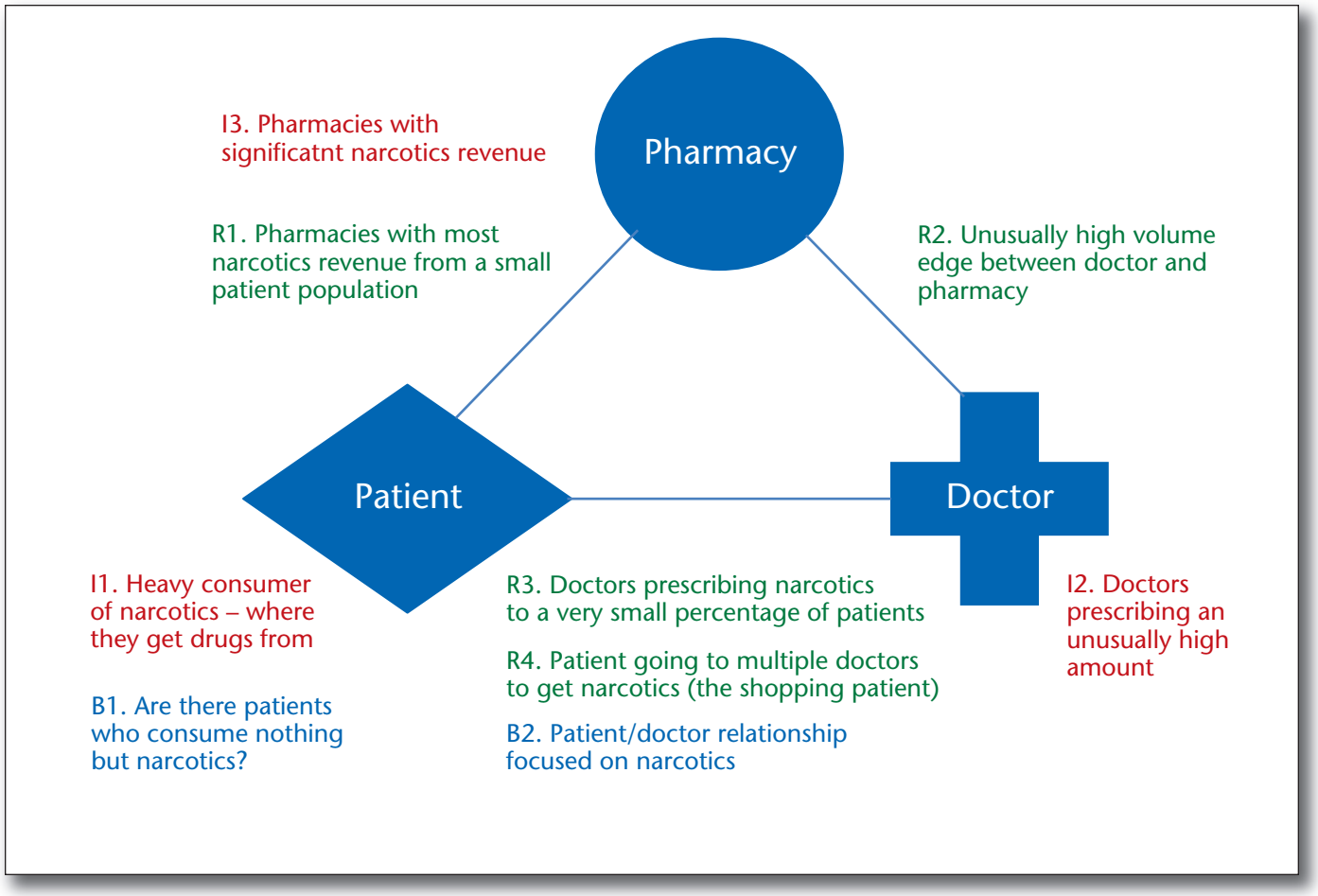

Figure 3. Anomalies in a Narcotics Relation Graph.

more than 52,000 doctors, and nearly 9,000 pharmacies. We focus on Schedule II narcotics defined by the U.S. Controlled Substances Act.[4] Schedule II drugs are from drug classes such as opiates, stimulants, and depressants, which are all highly abusable. Examples include morphine, oxycodone, and fentanyl. Within the data set, 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 or 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 50 times 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 health care with a high resale value, such as durable medical equipment and diabetes supplies. The same anomaly-detection techniques, described here, are applicable in these domains.

Temporal and Geospatial Reasoning

Interesting insights can be obtained by exploring the dynamic property of a health-care graph. We analyzed the 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 patients. Heavy links are graph edges where unusually strong

business relationships occur. In current practice, fraud analysts manually search for these types of anomalous providers through SQL-like queries. Our approach automates the effort and aids investigators to identify these outliers systematically.

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) on empirical observation data. By comparing these transition probabilities to a baseline, we can identify source, sink, and heavy links.

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 percent of prescriptions are filled at the same pharmacy where the previous prescriptions are filled. By comparison to this baseline, two different types of source nodes are detected by our algorithms. The first type of source nodes tends to lose patients to another specific pharmacy (that is, a sink node). For example, our analysis identified two pharmacies where 85 percent 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 whether the business relationship between the source and sink represents 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. Nevertheless such abnormal patterns are worth investigators' attention.

Geospatial Analysis

Geospatial data are another useful source of information for anomaly detection. We assume that most patients visit physicians and pharmacies in their local cities. This is especially true for the Medicaid population since Medicaid is designed to cover patients that have economic constraints or are physically immobile. Occasionally there are many benign reasons why patients might visit providers far from home, for example, (1) sickness or injury during travel, and (2) 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's or physician's business relationships are 50 percent within a 10-mile radius, 80 percent within a 20-mile radius, 90 percent 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 density-based clustering algorithm, to the empirical distributions to define the baseline. Cdfs that are similar to each other are grouped together, while 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 percent of its business comes from a physician more than 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 the provider writes. A cardiologist's prescriptions will be composed of a high proportion of heart-disease-related 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. We call these communities out-of-specification. 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. In this graph two providers are connected if one of them has sent at least one prescription to the other. Edges are labeled by a vector of features computed as a function of the prescription events. Unfortunately, we cannot observe a provider's type, nor the communities to which the provider belongs. We develop a probabilistic algorithm that discovers out-of-spec communities, ranks them and their members, and outputs a

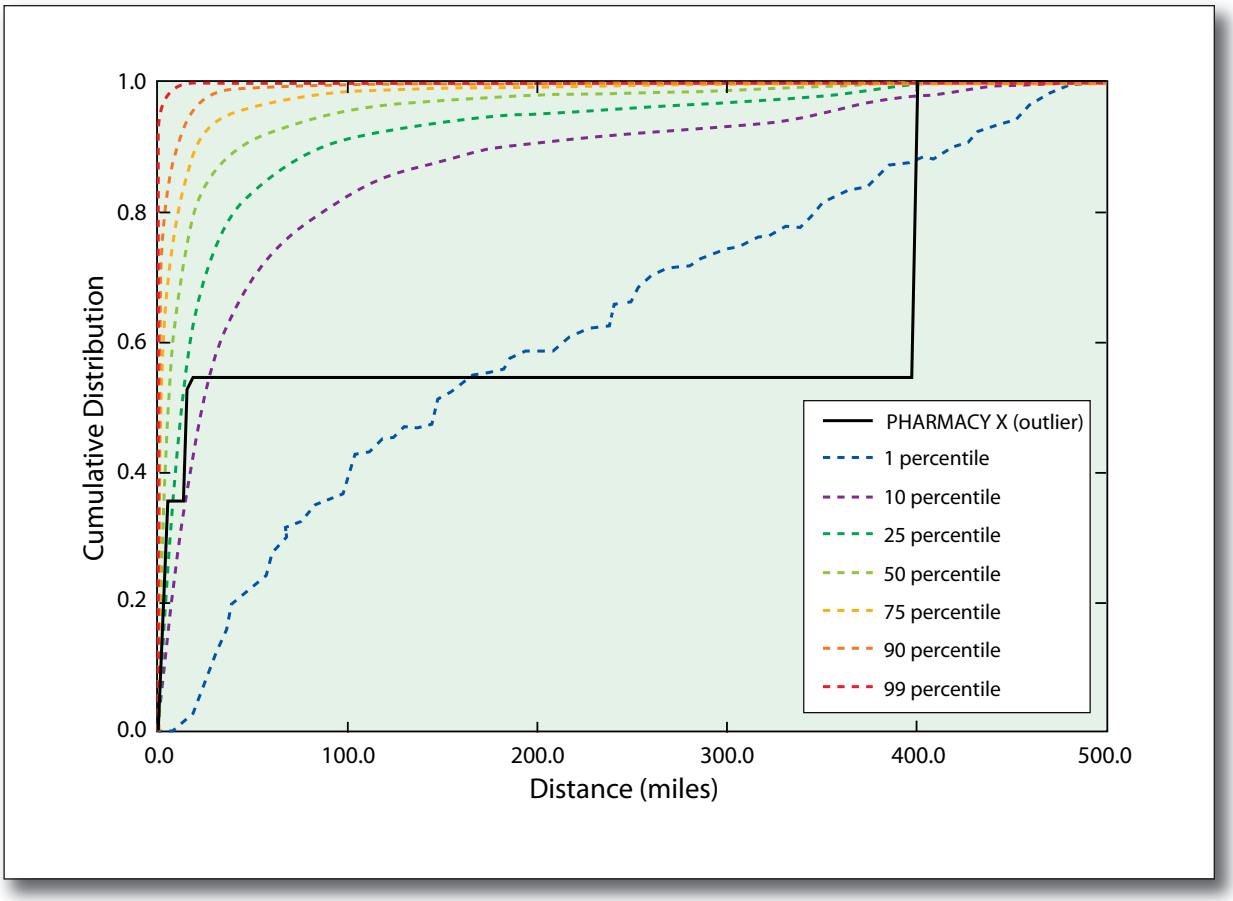

Figure 4. Example of Geospatial Anomaly.

description of the community's unusual behaviors.

Our model of out-of-spec communities represents the types of individual providers, the community each provider belongs to, the global prescription behaviors generated by these types, and community prescription behaviors generated by community members. To do this we build upon the Latent Dirichlet Allocation model (Blei, Ng, and Jordan 2003) by introducing the concept of communities, and community prescription behaviors. Each provider in the graph is associated with both a type and a community. Provider types are associated with distributions over the global prescription behaviors. These distributions provide a basis for discovering anomalous communities. Each provider is also associated with a community; a community is a connected component of the provider graph. Together a provider's type and community impose a distribution over the prescription behaviors. The set of global prescription behaviors and the community prescription behaviors are distributions over the edge features. Conditioned on a provider's type and community, the data observed in the edges is assumed to be generated from a mixture of global behaviors (representing what is typical) and community behaviors (representing what is peculiar to the community). We implemented an inference algorithm that infers each part of this representation. Given the data we discover the set of provider types, each provider's community memberships, the set of global behaviors, and the community prescription behaviors. Intuitively, communities with members whose adjacent edges are generated from the community behavior distributions are considered anomalous. According to this intuition we developed an algorithm that ranks the discovered communities and the community members.

Estimating the latent variables in the graph is a computationally demanding task. There are a lot of variables here: provider type, community memberships, prescribing behaviors of individual providers, and communities. 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 solu-

tion. As output we receive the collection of communities and their corresponding latent variable estimates for presentation to analysts.

We have integrated our model with two tools that our users can employ to investigate the discovered communities. The first is the network explorer (NE), which provides a high-level view of the community set. We will describe NE features in more detail later in the article. The second tool is the Group-in-a-Box visualization (Rodrigues et al. 2011) to provide visual cues of why communities and their members are anomalous. Figure 5 shows an example view of anomalous latent communities. Communities with low anomaly scores are removed in order to present a less cluttered visualization. The high-level view clearly highlights community members that stand out with respect to their peers. Larger nodes (red dots) represent more anomalous community members. Our users can garner further details on demand by clicking a node to zoom into a finer graph structure.

Anonymized Cases Under Investigation

Given a network of approximately 74,000 providers with more than 900,000 prescription relationships, our algorithm discovers 900 communities of varying sizes. In order to evaluate the quality of the discovered communities we presented our partners with a list of 40 providers. Our algorithm ranks both the communities and community members. To generate our list of providers we took the top five community members from the top 11 communities. This is a small set of providers, but our evaluation is limited by the very high cost of evaluating examples (our partners estimate up to two hours of time are needed per provider). We asked our partners to identify whether a member of the list represented FWA and to provide feedback on the providers. In order to produce these labels our partners could make use of a historical case database that stores information about past analysis. Of the set we presented to our partners, they identified nine individuals as being representative of FWA. Seven of these providers were previously identified, through weeks of painstaking search, by partners. Importantly, two of the discovered providers became new cases. This was interesting given that our analysis was performed on historical data. Our algorithms are able to analyze the data in hours and provide illustrating evidence so that the analysts can build a case with ease in a day's worth of effort.

Discovering Anomalous Structure in the Graph

In this section, we report our work in progress on a nonparametric 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 article we consider the specific case of referral networks where the graph G is composed of provider nodes, and the links between nodes a and b represent the total number of referrals between providers a and b. 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; and (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 tight-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 $\bar{C}$ through the proposed agglomeration scheme, we check to see whether 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 (2006), except that this ratio is defined based on nodes, and the graph-modularity metric is defined on edges. On one hand, 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 $\bar{C}$

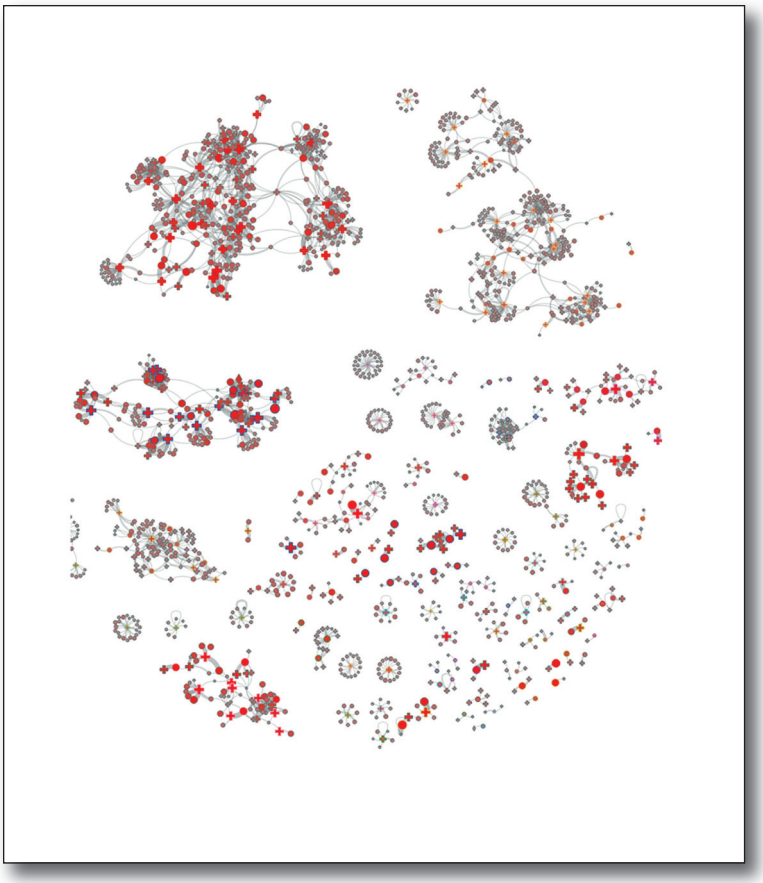

Figure 5. Anomalous Communities Discovered by the Analysis.

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 article, we consider the following sets of features to characterize any given community C_i in the referral network: community size, community density, average dollar amount, and average anomaly score. Community size is the number of nodes in C_i. Community density is the ratio of the total number of edges in C_i relative to the number of nodes. Average dollar amount is the ratio of the total dollar amount associated with the referral edges in C_i relative to the number of nodes. For the 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 whether 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 article, 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 2432 communities. These 2432 communities accounted for about 40,000 providers, or about 66 percent 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 or communities, while a majority of the communities had an average of about 10 providers. Ten 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.

Visualization: The Network Explorer

Network visualization is one of the most difficult challenges of information visualization. Many techniques have been proposed for small networks (with hundreds of nodes or less), but they are ill-suited for analyzing our medical network of tens of thousands of providers. To address the challenge, we have developed a new visualization tool, the network explorer, to support the FWA detection task. Figure 6 shows the network explorer's data overview interface. The overview enables users to get an eagle-eye picture of the network and get an overall idea of how many nodes (providers, patients, or pharmacies) are suspicious, whether and how the nodes form clusters, and their respective cluster sizes. The network explorer uses a rank-by-relevance framework to visualize a subset of nodes (for example, 10,000) to control the visualization complexity and focus visualization resources only on noteworthy nodes. Highly suspi-

cious nodes identified by the anomaly-detection schemes described in the earlier sections are highlighted. Clustering provides visual structure that can be easily interpreted by the users.

In the network explorer, users navigate the network through filtering and selection. The right-hand panel in figure 6 demonstrates the filtering mechanism. One can filter the network by edge properties (such as the number of referrals, as shown by the first filtering bar) or by node properties (such as the anomaly score based on how rare a procedure is performed, aka procedure fraction anomaly score, as shown in the second filtering bar). This filtering mechanism allows users to query the network directly.

Given an identified suspect node, the user may need to explore the suspect's network connections. In fact the ability to do local network exploration is the second most requested requirement from our users (next only to the overview). For instance, one of our users was interested in understanding which pharmacies were driving the business from a specific prescriber, and identifying which other prescribers were heavy customers of said pharmacy. To meet this requirement, the network explorer includes an egocentric mode that lets users select one or more target nodes and interactively explore their ego-net, that is, the *k*-hop neighborhood for small *k*. All of the features available in the overview mode also work in the egocentric mode, including filtering, dynamic group-in-a-box, and rank-by-relevance. Figure 7 demonstrates the progressive drill down and visualization with increasingly finer granularity. In this example, the user has clicked a node to jump into the gray cluster on the top-left corner of figure 7a. By doing this, only the 83 nodes in the cluster are shown in the screen (figure 7b). Then the user reruns the clustering algorithm to discover the finer subcluster structure shown in figure 7c and finally zooms into a subcluster with only 27 nodes (figure 7d). The node navigator bar on the left of figure 7 shows the users has jumped into a cluster two times. This zooming-in ability, as well as zooming out by clicking on the open space or the node navigator bar, allows the user to navigate the network freely and effectively.

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 prepay detection.

The goal of provider review is to find providers (doctors, hospitals, clinics, and others) 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 health-care 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 the 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 overbilling that can be proven easily and then tries to find as many instances as possible.

The goal of prepay detection is to identify inappropriate claims before the provider is paid for those claims. For any given claim, a prepay algorithm determines whether 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 overbilling that can be addressed by a rule change. In prepay detection, when making a decision about a new claim, the algorithm can look at patient-provider, provider-provider, and patient-

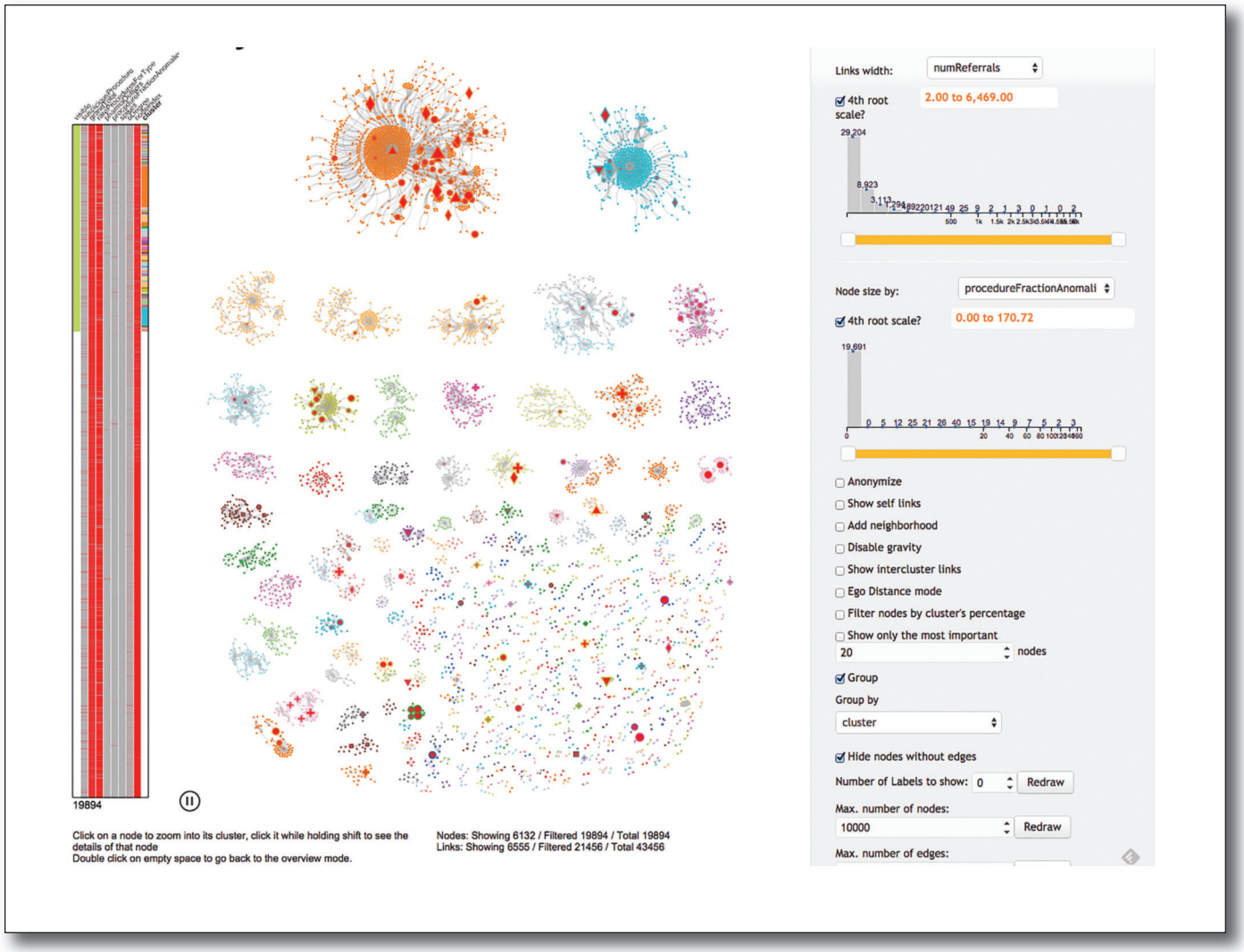

Figure 6. Network Explorer Main Interface with 10,000 Nodes.

patient relationships together with information about particular providers, patients, and claim features that have been associated with overbilling in the past to recommend human review of some claims.

Discussion: Deployment and Evaluation

We have applied modular programming principles to design our anomaly-detection components as independent and interchangeable algorithmic pieces that can be composed or rewired to operate on diverse data sets. For instance, the same set of graph feature extraction algorithms described in the Narcotics Network section can be applied to data sets regarding diabetic supply or durable medical equipment to identify potential perpetrators. The modularity and reusability have significantly reduced deployment cost. The remaining major deployment workload is on data ingestion and schema adaptation. Data ingestion needs to take into account the multiple data representations coexisting in different parts of the data. For instance, providers are often identified using National Provider Identifier (NPI) numbers, but pharmacies, which are also one type of provider, are sometimes identified using their NABP (National Association of Board of Pharmacy) numbers. Likewise, patient identity may be represented by several insurance plan/program numbers. U.S. medical practice is transitioning its diagnosis code system from ICD-9 to ICD-10, and hence almost all data sets see a mixture of both. We customize ETL (Extract/Transfer/Load) code to unify the differences and convert each data set's native data representation to the data-set-independent representation that our algorithm modules expect. Depending on data complexity, the

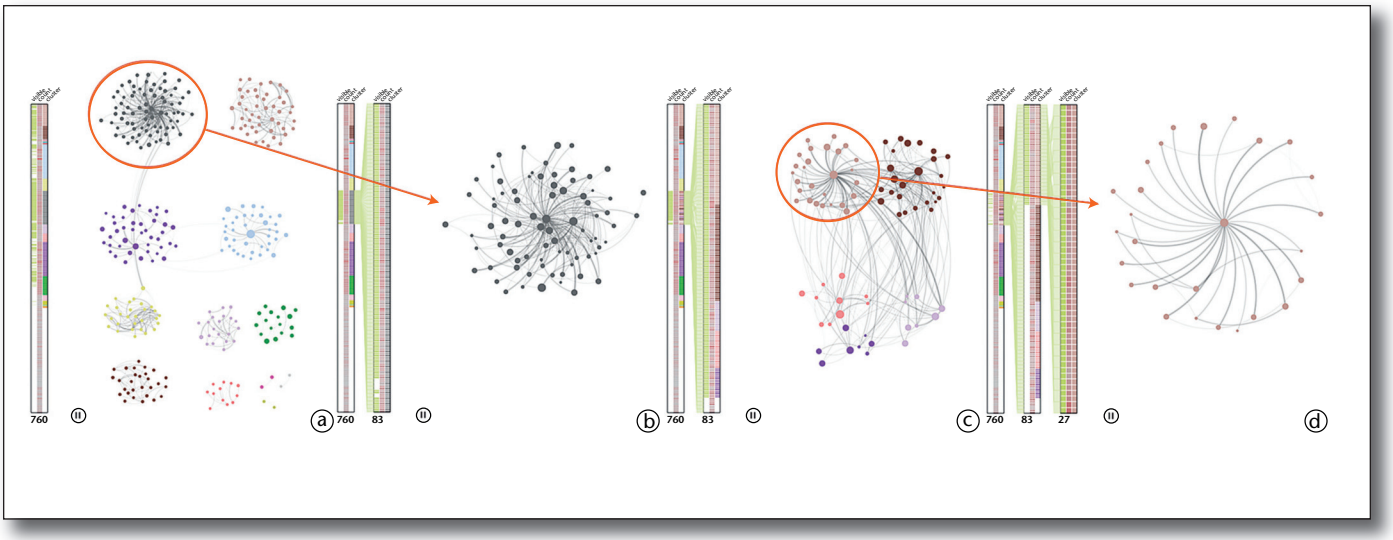

Figure 7. Users Can Jump into Clusters to Focus Their Exploration into Specific Parts of the Network.

data ingestion and schema adaptation effort may amount to days of work. Amortized over the lifetime of the XPIV, the deployment cost is fairly low.

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 another example, using relationships in a narcotics graph, we produced a list of providers with high anomaly scores and made them available to a group of analysts for review. Of the 39 providers reviewed so far, 8 were already under review and 1 more was written up for review, for 9/39 = 23 percent suspicious cases. Of the remaining cases, 11 were in provider categories that are out of scope for review at the time. Eight more were of low materiality based on predetermined thresholds defined by the payer. Such providers can easily be filtered out of future reports. Removing these 19 providers from our current set would give us an updated ratio of 9/20 or 45 percent. Considering a list of recommended providers still takes time and effort for our analysts, but it appears to be a valuable new source of candidate providers and reduces effort compared to manually constructed queries and reports. 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 article presents our work on developing graph-analysis techniques and applying them to real-world health-care data sets to look for fraud, waste, and abuse activities. We represent the health-care 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 of 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 acknowledge technical contributions from our colleagues John Hanley, Alan Bell, Sureyya Tarkan, Alex Brito, Ming Yang, and Nick Briggs.

Notes

1. See NHCAA. 2011: The Challenge of Health Care Fraud: Consumer Alerts: The Impact of Health Care Fraud on You. Washington, DC: The National Health Care Anti-Fraud Association (www.nhcaa.org/resources/health-care-anti-fraud-resources/the-challenge-of-health-care-fraud.aspx).

2. Infographics on the 2013 video series Rehab Racket produced by CIR and CNN are available from revealnews.org.

3. The Health Insurance Portability and Accountability Act of 1996, Public Law 104-191, is available at www.cms.gov/Regulations-and-Guidance/HIPAA-Administrative-Simplification/HIPAAGenInfo/Downloads/HIPAALaw.pdf.

4. Title 21 CFR 1308.12. U.S. Department of Justice, Drug Enforcement Administration, available at www.ecfr.gov.

References

Aggarwal, C. C. 2012. *Outlier Analysis*. Berlin: Springer.

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

Busch, R. S. 2012. *Healthcare Fraud: Auditing and Detection Guide*. New York: Wiley.

Epstein, R. 1989. Drug Wars in the United States. *British Medical Journal* 299(6710): 1275–1276. dx.doi.org/10.1136/bmj.299.6710.1275

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 *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining,* 226–231. Menlo Park, CA: AAAI Press.

Fortunato, S. 2010. Community Detection in Graphs. *Physics Reports* 486(3): 75–174. dx.doi.org/10.1016/j.physrep.2009.11.002

Institute of Medicine. 2012. *Best Care at Lower Cost: Transformation of Health System Needed to Improve Care and Reduce Costs*. Report, September 6. Washington, DC: National Academies of Sciences, Engineering, Medicine. iom.nationalacademies.org/Reports/2012/Best-Care-at-Lower-Cost-The-Path-to-Continuously-Learning-Health-Care-in-America/Press-Release.aspx#sthash.fvodsvbX.dpuf

Leap, T. L. 2011. *Phantom Billing, Fake Prescriptions, and the High Cost of Medicine*. Ithaca, NY: Cornell University Press.

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): 1162–1179. dx.doi.org/10.1080/01621459.1968.10480918

Liu, F. T.; Ting, K. M.; and Zhou, Z.-H. 2008. Isolation Forest. In *Proceedings of the Eighth IEEE International Conference on Data Mining,* ICDM'08, 413–422. Piscataway, NJ: Institute for Electrical and Electronics Engineers. dx.doi.org/10.1109/icdm.2008.17

Manyika, J.; Chui, M.; Brown, B.; Bughin, J.; Dobbs, R.; Roxburgh, C.; Byers, A. H. 2011. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute Report (May). New York: McKinsey & Company.

Mason, D. M. 1982. Some Characterizations of Almost Sure Bounds for Weighted Multidimensional Empirical Distributions and a Glivenko-Cantelli Theorem for Sample Quantiles. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 59(4): 505–513. dx.doi.org/10.1007/BF00532806

Newman, M. 2006. Modularity and Community Structure in Networks. In *Proceedings of the National Academy of Science*s USA 103(23): 8577–8582. dx.doi.org/10.1073/pnas.0601602103

Radnofsky, L., and Walker, J. 2014. DEA Restricts Narcotic Pain Drug Prescriptions. *Wall Street Journal:* 22 August.

Rodrigues, E. M.; Milic-Frayling, N.; Smith, M.; Shneiderman, B.; and Hansen, D. 2011. Group-in-a-Box Layout for Multifaceted Analysis of Communities. In *Proceedings of the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and Social Computing (SocialCom),* 354–361. Piscataway, NJ: Institute for Electrical and Electronics Engineers.

Juan Liu is the lead research scientist at Medallia. She was previously at Palo Alto Research Center, which is where the research for this article was conducted.

Eric Bier is principal scientist at Palo Alto Research Center.

Aaron Wilson is a research scientist and manager at Palo Alto Research Center.

John Alexis Guerra-Gomez is a research scientist at Yahoo Labs.

Tomonori Honda is a senior data scientist at Inflection.com.

Kumar Sricharan is a research scientist at Palo Alto Research Center .

Leilani Gilpin is a Ph.D. student at the Massachusetts Institute of Technology.

Dan Davies is a research engineer at Palo Alto Research Center.