Anomaly Detection in Opioid Prescription Using Graph Based Approach

Ramesh Paudel
Tennessee Technological University
Cookeville, Tennessee 38501
Email: rpaudel42@students.tntech.edu

***Abstract*—Every year, billions of dollars are lost due to fraud in the U.S. health care system and prescription fraud accounts substantial amount to that monetary loss. Health care claims are complex as they involve multiple parties including service providers, insurance subscribers, and insurance carriers. Medicare is susceptible to fraud because of this complexity. To build a comprehensive fraud detection system, one must take into consideration all of the financial practices involved among the associated parties. This paper is focused on graph-based analysis of CMS provided Medicare prescriber summary data to look for anomalies in the relationships and transactions carried by prescriber. We create graphs for prescriber, drug prescribed, and payment he/she received under the Medicare part D prescription drug program. We then demonstrate the potential effectiveness of applying graph-based approach to the problem of discovering anomalies and potential fraud scenarios.**

***Index Terms*—Heathcare Fraud, Anomaly Detection, Graph Based Anomaly, Prescription Fraud**

I. Introduction

Centers of Medicare and Medicaid Services (CMS) reported that U.S. health care spending reached $3 trillion or $9,523 per person in 2014. The total health care spending in 2014 accounted for 17.5% of the nations Gross Domestic Product and is expected to rise to 20.1% by 2025 [1]. Unfortunately, roughly one-third of health care spending can be attributed to fraud, waste, and abuse [2]. Because of this significant financial loss, there is a need to build better fraud detection mechanisms.

Medicare Part D is the part of Medicare that provides optional outpatient prescription drug coverage to Medicare beneficiaries. Prescription drug abuse is a serious and growing problem. Three classes of drugs were identified as being susceptible to abuse by prescription shoppers namely: opioids, benzodiazepines and psychostimulants [3]. Overdoses of prescription painkillers called opioids are among the leading causes of accidental death in the United States [4]. On Oct 26, 2017, US government declared opioid crisis a public health emergency. According to the Centers for Disease Control and Prevention (CDC) [4], in 2014 almost 2 million Americans abused or were dependent on prescription opioid.

With the rise in prescription drug abuse, concerns about Medicare fraud, particularly prescriber fraud, have increased. There are basically two types of health-care fraud operations [5]:

Hit and Run: Fly-by-night operators who steal millions in a relatively short period, then vanish.

Steal a little all the time: Perpetrators who work to ensure fraud goes unnoticed and bill fraudulently over a long period of time. The provider may hide false claims within large batches of valid claims and, when caught, will claim it an error, repay the money, and continue the behavior.

Prescription fraud is defined as the illegal acquisition of prescription drugs for personal use or profit [6]. And they could be observed in numerous ways: from dishonest providers, organized criminals, colluding patients, and patients who misrepresent their eligibility for health insurance coverage [7]. Typical prescription fraud scenario that involve patients is called doctor shopping where patient consults many physicians in order to obtain multiple prescriptions of drugs in excess of their own therapeutic need [8]. Other scenario of prescriptions shopping be [9]:

- Contradicting drug prescription (e.g. sleeping tablets versus stimulative tablets).
- Visiting a diversity of doctors for similar types of drugs.
- Excessive drug quantities over a set period.
- Sudden changes in prescription behaviors.
- Recurrent large temporal gaps after getting lots of drugs.

In this paper, we introduce an approach for discovering prescriber fraud involving opioid using *graph-based data mining*. If we consider the *entities* involved in the Medicare part D prescriber public use data as *nodes*, and the *relationships* and *transactions* between the entities involved as *edges*, we can represent the entire process as a *graph*. Using a known graph-based anomaly detection approach, we will show how anomalies that are potentially fraudulent can be discovered in data representing prescriber summary. To be more comprehensive in our approach we will focus on anomalies related to suspicious individual based on their attribute as well as suspicious community. To empirically validate our proposed approach, we will apply the publicly available GBAD [10] and SCAN [11] tool on CMS provided Medicare Part D Prescriber Public Use File (PUF) [12] and Medicare Part D Prescriber Summary [13]. GBAD will be used to find anomalies related to suspicious individual as this tool provide us a ability to find anomalies in attributed graph while SCAN tool will be used to find community (or relationship) based anomalies since it looks for anomalies based on how each individual node

share their neighbor. As Medicare part D prescriber public use data have relationship among provider, drug prescribed, city they live in, drug cost, claims filed, physician specialty, and total payment received, SCAN tool will be a perfect fit to analyze these relationships. Also GBAD system has been successfully applied to detect anomalous activities in Medicare claim files of diabetic patient [14] and using it on prescriber anomaly detection should be valid choice. For this work, we will specifically target prescriber with opioid prescription. A list of opioid drug names considered for the study are given in table 1.

SN	Drug Name
1	Buprenorphine
2	Codeine
3	Fentanyl
4	Hydromorphone
5	Methadone
6	Morphine
7	Oxycodone
8	Tramadol

TABLE I
NAME OF OPIOID DRUGS TAKEN INTO CONSIDERATION

The next section of this paper presents existing research on detecting health care fraud. We will then follow that with a description of the dataset. After which, we will briefly discusses the tools used, followed by a experimentation. We will then conclude with our results,discussion and future directions for this work.

II. RELATED WORK

Most of the research in health care fraud is focused on statistical analysis and the use of machine learning algorithms like clustering, k-nearest neighbor, decision trees, neural networks, etc. But, compared to the extent of financial loss in the health care sector, the research to date has been minimal.

One of the method is to use statistical and auditing methods [15] in existing Business Intelligence technology for detecting fraud in a states Medicaid payment system. [16] propose a supervised fraud detection system for the Chilean health care system which uses a committee of multilayer perceptron neural networks (MLP) for each one of the entities involved in the fraud/abuse problem: medical claims, affiliates, medical professionals and employers. Decision trees [17] is used for detecting insurance subscribers fraud for the Health Insurance Commission (HIC) of Australia while [18] apply process-mining techniques to gather clinical-instance data to construct a model that identifies service provider fraud for the NHI in Taiwan. [19] look at the data beyond the transaction level and build upon [5] fraud type classifications to predict the likelihood of fraud.

All of the above approaches focus on health care fraud in general. However, there are research that focus specifically in prescription fraud. [6] proposed an online/offline prescription fraud detection model in Social Security Agency in Turkey with false positive rate of 6%. [8] propose the use of a k-Nearest Neighbor (kNN) algorithm with an optimized non-Euclidean distance metric using a genetic algorithm. Their approach is focused on detecting two types of fraud schemes: inappropriate practice of service providers and doctor-shoppers. [20] describes a methodology for identifying and ranking candidate audit targets from a database of prescription drug claims. The relevant audit targets may include various entities such as prescribers, patients and pharmacies, who exhibit certain statistical behavior indicative of potential fraud and abuse over the prescription claims during a specified period of interest. [3], [21] and [9] focus on detecting doctor-shoppers in Medicare Australia. [3] captures the temporal (explicit) and spatial (postcode-based) aspects of consumers prescription behaviors to detect doctor-shoppers whereas [21] examines sequential prescription patterns from either a global or a localized view. All of these approach need expert knowledge to design a set of rules, and the anomaly is detected by observing the deviation from such rules. The performance of these approaches is limited by the availability of domain experts. However, [9] propose an unsupervised fraud detection system called UNISIM that can efficiently extract the hidden consumer patterns with respect to their temporal prescription behaviors.

None of the above techniques are targeting the *relational* aspect of the involved entities - something appropriate for a graph-based approach. [22] construct the social network of provider and use network features to distinguish between fraudulent and non-fraudulent providers. In earlier work, we introduced an approach for discovering health care fraud using graph-based data mining [14] . We considered the entities involved in the process of medical claims as nodes, and the relationships and transactions between the entities involved as edges, and represented the entire process as a graph. Using a graph-based anomaly detection approach called GBAD, we showed how anomalies that are potentially fraudulent can be discovered in data representing health care transactions. [23] propose a graph-analysis technique called Xerox Program Integrity Validator (XPIV) to find fraud in health care by using an ego-net approach to examine narcotics relationships and temporal-spatial characteristics of patients migrating between pharmacies and providers; as well as the global structure of the health care relationship network to look for communities sharing a common abnormal practice. In preliminary work, they are able to identify millions of dollars lost in fraud for potential recovery. Though they use a graph-based approach in detecting anomalies, their work is focused on overall narcotic relationships while our proposed approach will only targets anomalies on opioid prescriber network. Their approach focus on egonet which is in weighted graph. We will be using subdue based anomaly detection on static attributed graph and SCAN to discover communities, hubs, and outliers.

III. PRESCRIBER DATA

The dataset used for this research came from Center for Medicare and Medicaid Services (CMS). The first dataset used is Medicare Part D Prescriber Public Use File (PUF) 2015 [12] and Medicare Part D Prescriber Summary 2015

[13]. The Part D Prescriber Public Use File (PUF) provides information on prescription drugs prescribed by individual physicians and other health care providers and paid for under the Medicare Part D Prescription Drug Program. It contains all 2015 prescriptions aggregated at the physician and drug level, as well as the physician information like name, credentials, gender, complete address etc. For each prescriber and drug, the dataset includes the total number of prescriptions that were dispensed (including original prescriptions and any refills), total 30-day standardized fill counts, total days supply for these prescriptions, and the total drug cost. Part D Prescriber Summary is the aggregated information at the prescriber-level (i.e. one summary record per NPI) that includes enhanced prescriber demographic information beyond what is provided in the Part D Prescriber PUF detail.

More detailed information about the data can be found at [12] and [13]. For the purpose of this study, we will use the prescriber from the state of Florida who dispensed opioid prescription on 2015. Opioid drug taken under consideration are shown in table I. It should be noted that there is nothing that limits us to this particular subsample. It was an arbitrary choice for validating our proposed approach, and we will be expanding our dataset in the future with more samples.

IV. Approach

We employ two different approach to detect anomalies in opioid prescriber. First, is use to detect individual anomalies and second is used to detect anomalies in community. GBAD tool is use to detect individual anomalies and SCAN is used for detecting anomalies in community. The detail discussion about these tools is given below.

A. Graph-Based Anomaly Detection

In order to lay the foundation for this effort, we hypothesize that a real-world, meaningful definition of a graph-based anomaly is an unexpected deviation to a normative pattern. The importance of this definition (which we more formally define below) lies in its relationship to any deceptive practices that are intended to illegally obtain or hide information [10].

Definition IV.A. A labeled graph *G* = (*V*,*E*,*F*), where *V* is the set of vertices (or nodes), *E* is the set of edges (or links) between the vertices, and the function *F* assigns a label to each of the elements in *V* and *E*.

Definition IV.B. A subgraph *SA* is anomalous in graph *G* if (0 ¡ d(*SA*,*S*) ¡ *TD*) and (P(*SA*|*S*) ¡ *TP*), where P(*SA*|*S*) is the probability of an anomalous subgraph *SA* given the normative pattern *S* in *G*. *TD* bounds the maximum distance (d) an anomaly *SA* can be from the normative pattern *S*, and *TP* bounds the maximum probability of *SA*.

Definition IV.C. The score of an anomalous subgraph *SA* based on the normative subgraph *S* in graph *G* is d(*SA*,*S*) * P(*SA*|*S*), where the smaller the score, the more anomalous the subgraph.

The advantage of graph-based anomaly detection is that the relationships between entities can be analyzed for structural oddities in what could be a rich set of information, as opposed to just the entities attributes. However, graph-based approaches have been prohibitive due to computational constraints, because graph-based approaches typically perform subgraph isomorphisms, a known NP-complete problem. Yet, in order to use graph-based anomaly detection techniques in a real-world environment, we need to take advantage of the structural/relational aspects found in dynamic, streaming data sets.

In order to test our approach, we will use the publicly-available GBAD test suite , as defined by [10]. Using a greedy beam search and a minimum description length (MDL) heuristic, GBAD first discovers the best subgraph, or normative pattern, in an input graph. The MDL approach is used to determine the best subgraph(s) as the one that minimizes the following:

$$M(S,G) = DL(G|S) + DL(S),$$

where G is the entire graph, S is the subgraph, DL($G|S$) is the description length of G after compressing it using S, and DL(S) is the description length of the subgraph. The complexity of finding the normative subgraph is constrained to be polynomial by employing a bounded search when comparing two graphs. Previous results have shown that a quadratic bound is sufficient to accurately compare graphs in a variety of domains [10].

For more details regarding the GBAD algorithms, the reader can refer to [10]. In summary, the key to the GBAD approach is that anomalies are discovered based upon small deviations from the norm (e.g., insider threat, identity theft, etc.) not outliers, which are based upon significant statistical deviations from the norm.

B. SCAN: A Structural Clustering Algorithm for Networks

SCAN (Structural Clustering Algorithm for Networks) [11] detects clusters, hubs and outliers in networks. It clusters vertices based on a structural similarity measure. Let $G = V, E$ be a undirected and unweighted graph, where V is a set of vertices and E is set of edges, SCAN either assign each vertex in V to the community C, hub H or an outliers O based on how they share neighbors. Doing so makes sense when you consider the detection of communities in large social networks. Two people who share many friends should be clustered in the same community.

In order to find clusters, SCAN first detects a special node that has a lot of neighbor nodes with highly dense connections, called core. Once SCAN finds core, it expands a cluster from the core by assigning directly structure-reachable nodes to the same cluster as the core. For e.g. is u is the core node, all directly structure-reachable nodes $D[u]$ are assigned to the same cluster with u. It then recursively expands the cluster by checking whether each node v in the cluster is (1) included in $D[u]$ and (2) node v is core. If so, SCAN assigns nodes in $D[v]$ to the same cluster as node u. These expanded directly structure-reachable nodes (i.e. $D[v]$) expanded from a member node v of the cluster (i.e. $D[u]$) are called structure-reachable nodes of node u. If node $v \epsilon D[u]$ is not core, it does not expand

the cluster from node v. All nodes in a cluster, except the core node, are called border nodes. SCAN recursively finds cores and expands the clusters from the cores until there are no undiscovered cores in the structure-reachable nodes of node u. The collection of structure-reachable nodes of node u, which are composed of cores and borders belongs to the same cluster as node u. Formally, the cluster that has node u is defined as follows:

Definition (CLUSTER). The cluster by node u, denoted by $C[u]$, is defined as $C[u] = w\epsilon D[v] : v\epsilon C[u]$, where $C[u]$ is initially set to $C[u] = u$.

After termination of cluster expansion, SCAN randomly selects a new node from the nodes that have yet to be checked. SCAN continues this procedure until there are no undiscovered cores. Finally, SCAN identifies non-clustered nodes (i.e. nodes that belong to no cluster) as hubs or outliers.

Definition (HUB AND OUTLIER). Assume node u does not belong to any cluster. $u\epsilon H$ *iff* node v and w exist in neighborhood of u, $N[u]$ such that $C[v] \neq C[w]$. Otherwise $u\epsilon O$.

To know detail about the algorithm please refer the original SCAN paper [11].

V. Experiment

Our experimental setup consists of parsing the required data from the Medicare Part D Prescriber Public Use File (PUF) and Part D Prescriber Summary table, constructing a single graph that contains the data for each prescriber from both the tables, and processing the resulting graph with a graph-based anomaly detection tool. In order to create the graph input file, we will create a parser (written in the python programming language) that will read the prescriber data and build the graph.

A. Graph Input File

We limited our anomaly detection to only opioid prescriber in Florida. The choice of population and drug was arbitrary and was done to ensure that we are examining prescriber with similar demographics and characteristics. For GBAD tool, we need to create a graph for each individual prescriber. There are 22,372 opioid prescriber in the dataset resulting in 22,372 examples or subgraph, each representing a opioid prescriber, for a total of 497,282 vertices, 474,910 edges. We will call this graph "individual prescriber graph". Figure 1 shows the sample of an individual prescriber graph used for anomaly detection using GBAD. From figure 1, one can see that each prescriber is represented by a *provider* node, where a provider *prescribe* a drug. Each of the drug is represented by a node. For e.g if a provider prescribe a drug called "Hydrocodone", then the drug a node *hydrocodone* is linked to *provider* node with an edge called *prescribe*. The statistics like the number of claim filed, total beneficiary, total drug cost, total opioid claim, total day supply etc. are hanged in as the leaf node of a graph. For each statistics, the value is bucketized by calculating standard deviation and mean and dividing them into five categories (mid, high, low, mid-high, and mid-low).

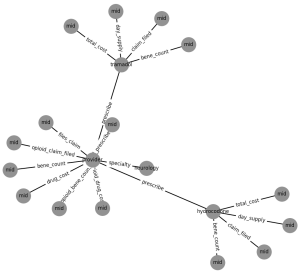

Fig. 1. GBAD Graph for Single Prescriber

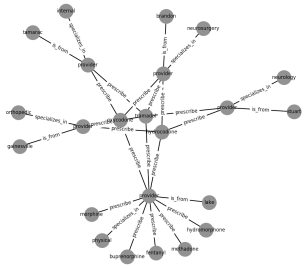

Fig. 2. Snapshot of SCAN Graph

For SCAN tool, we need a graph that represent the community. So, we build another graph that best represents the relationship between prescriber so that we will be able to find the anomaly related to community. We will call this a "prescriber community graph". Figure 2 shows the snapshot of a prescriber community graph, where the prescriber are linked by the city they live in, drugs they prescribe, or specialization they share. Thus, in the prescriber community graph, shown in figure 2, these relationship are represented as "provider" - "is_from" - "city", "provider" - "prescribe" - "drug" and "provider" - "specializes_in" - "specialization". Each city node have a city name as a label (like "Orlando", "Miami" etc.) , drug node have a drug name as a label (like "Hydrocodone", "Tramadol" etc.) and specialization node have a specialty as a label (like "Neurosurgery", "General Medicine" etc.) It should be noted that these are just visualizations, as the actual graph input files are just ASCII text files.

B. Individual Anomaly Detection

Running GBAD on the individual prescriber graph, Figure 3, shows the discovered normative pattern while running probabilistic algorithm of GBAD, which tells us that a "provider" prescribe "hydrocodone" and all the attribute values are "mid". We will discuss about various types of anomaly found using GBAD in these section.

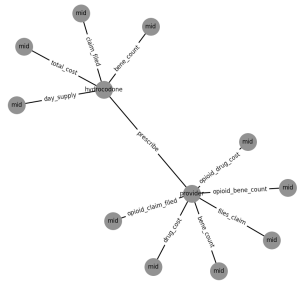

Fig. 3. Normative Pattern using Probabilistic Algorithm

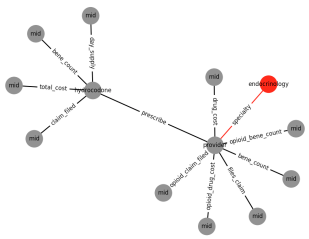

Fig. 4. Unexpected Edges & Vertices Anomaly

1) Unexpected Edges and Vertices: GBAD discover anomalous substructures with an extra vertex and edge in the individual prescriber graph as shown in Figure 4 while using probabilistic approach (one of several algorithms available in GBAD) with the amount of change (TD) set to 2 and

probability (TP) set to 1. The anomalies in each of the figures are depicted using a red vertex to represent the anomalous existence of a vertex and a red line to represent the anomalous existence of an edge. Further inspection of the data confirms that Figure 4 is anomalous. There are 24 endocrinologist in the dataset and almost all of them prescribe "tramadol". In addition to "tramadol", 9 of them prescribe "hydrocodone". But this specific provider has prescribed "hydrocodone" to the high number of beneficiaries with the total cost adding upto thousands of dollars while other endocrinologist prescribe "hydrocodone" to very few beneficiaries resulting the total cost to only few hundred dollars. While this does not necessarily confirm any malicious intent from the provider, but does raise the red flag because it does not follow the normal prescription behavior of a endocrinologist (as "tramadol" is the usual opioid they prescribe).

2) Missing Edges and Vertices: Running MPS algorithm on the individual prescriber graph with the amount of change (TD) set to 0.4, Fig. 5, shows the discovered normative pattern and Fig. 6 shows one of the anomaly discovered. Anomaly in this case can be described as a normative patterns missing two vertices and two edges. Further inspection of the data confirms that Figure 6 is anomalous. In Part D Prescriber Summary data, if the number of beneficiary are less than 11 they are suppressed meaning the field is left empty. This specific provider has opioid beneficiary count field empty which resulted in missing edge (opioid_bene_count & opioid_drug_cost) from the normative pattern. The interesting fact about this provider is that, beside having so few beneficiary, number of claims and opioid prescription rate are still in the range "mid". Again, this does not necessarily confirm any malicious intent, but does raise the red flag because it does not follow the normal provider behavior of having "mid" value for opioid beneficiary count and opioid drug cost.

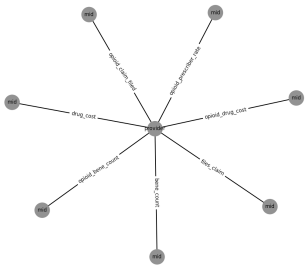

Fig. 5. Normative Pattern using MPS Algorithm

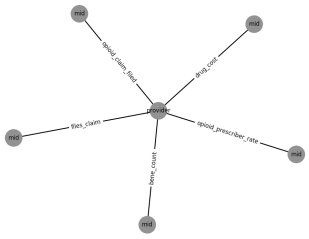

Fig. 6. Missing Edges & Vertices Anomaly

C. Community Based Anomaly Detection

Running SCAN (with $\epsilon = 0.25$ and $\mu = 8$) on the prescriber community graph, Figure 7, shows discovered clusters, hubs, and outliers. Each node in a single cluster are represented by the same color while hub nodes are in red color and outliers nodes are in gray color. SCAN finds 9 clusters and 14 hub nodes. Fig. 8 shows the snapshot of SCAN result that shows 3 clusters along with hubs and outliers connected to these 3 clusters. The detail study of clusters provide the knowledge on close knit community in opioid prescription behavior, while hub which is basically the shared node between different cluster give information about the most common drug, city or provider specialty and outliers are the node that are loosely link or not linked to any cluster or hubs. Closely inspecting fig. 8, three clusters are the cluster of provider from three cities Miami, Eglin Air Force Base (AFB) and Ponte Vedra. Observing even further into Eglin AFB cluster, almost all of the provider who are from to Eglin "specializes_in" "Emergency Medicine" and "prescribe" "hydrocodone". The outlier hanging off this cluster have provider who "specializes_in" "otolarangology", and "Student in an Organized Health Care Education/Training Program".

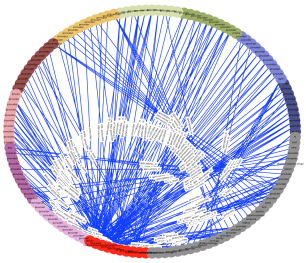

Fig. 7. Results of SCAN using $\epsilon = 0.25$ and $\mu = 8$

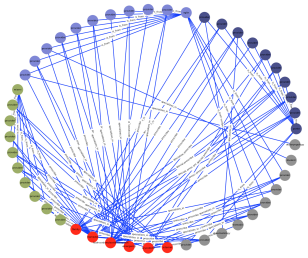

Fig. 8. Snapshot of SCAN Results Showing 3 Clusters

VI. Conclusion

In this paper, using a known graph-based anomaly detection approach, we showed how anomalies that are potentially fraudulent can be discovered in data representing opioid prescriber. We represented the prescriber data as a graph where the entities involved in opioid prescription are nodes, and the relationships and transactions between the entities involved are edges. For this work, we specifically target the opioid prescriber from Florida from 2015 and demonstrate the proof of concept of graph-based anomaly detection to the problem of discovering anomalies in opioid prescriber network. Prescriber network can be mapped as a traditional graph problems and using graph-based approach, anomalies that are difficult to be discovered by statistical approach can be discovered using these graph based tools.

Using the prescriber summary data without access to individual claim file, we were not able to represent all the relationship in prescriber network and our graph became shallow. This limited the performance of SCAN tool. Also individual prescriber graph has few relationship so running GBAD on it, discovered normative and anomalous patterns are star shaped representing attribute of specific node. Conceptually these normative and anomalous patterns are similar to the normal pattern that can be discovered by statistical approach. In future, we plan to use individual claim file which will have extra relationships and transactions. This will provide GBAD and SCAN with more edges representing the relationship in the data to uncover more anomalous activities.

REFERENCES

[1] CMS, "National health expenditure projections 2015-2025," 2016, [Online; accessed 8-November-2016]. [Online]. Available: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsProjected.html

[2] R. Kelley, "Where can $700 billion in waste be cut annually from the us healthcare system," *Ann Arbor, MI: Thomson Reuters*, vol. 24, 2009.

[3] K. S. Ng, Y. Shan, D. W. Murray, A. Sutinen, B. Schwarz, D. Jeacocke, and J. Farrugia, "Detecting non-compliant consumers in spatio-temporal health data: A case study from medicare australia," in *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*. IEEE, 2010, pp. 613–622.

[4] C. for Disease Control and Prevention, "Prescription opioid overdose data," 2017, [Online; accessed 5-November-2017]. [Online]. Available: https://www.cdc.gov/drugoverdose/data/overdose.html

[5] M. K. Sparrow, *License to steal: Why fraud plagues America's health care system*. Westview Press Boulder, CO, 1996.

[6] K. D. Aral, H. A. Güvenir, İ. Sabuncuoğlu, and A. R. Akar, "A prescription fraud detection model," *Computer methods and programs in biomedicine*, vol. 106, no. 1, pp. 37–46, 2012.

[7] D. Thornton, M. Brinkhuis, C. Amrit, and R. Aly, "Categorizing and describing the types of fraud in healthcare," *Procedia Computer Science*, vol. 64, pp. 713–720, 2015.

[8] H. He, W. Graco, and X. Yao, "Application of genetic algorithm and k-nearest neighbour method in medical fraud detection," in *Asia-Pacific Conference on Simulated Evolution and Learning*. Springer, 1998, pp. 74–81.

[9] M. Tang, B. S. U. Mendis, D. W. Murray, Y. Hu, and A. Sutinen, "Unsupervised fraud detection in medicare australia," in *Proceedings of the Ninth Australasian Data Mining Conference-Volume 121*. Australian Computer Society, Inc., 2011, pp. 103–110.

[10] W. Eberle and L. Holder, "Anomaly detection in data represented as graphs," *Intelligent Data Analysis*, vol. 11, no. 6, pp. 663–689, 2007.

[11] X. Xu, N. Yuruk, Z. Feng, and T. A. Schweiger, "Scan: a structural clustering algorithm for networks," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007, pp. 824–833.

[12] CMS, "Medicare provider utilization and payment data: 2015 part d prescriber," 2017, [Online; accessed 7-November-2017]. [Online]. Available: https://data.cms.gov/Medicare-Part-D/Medicare-Provider-Utilization-and-Payment-Data-201/3z4d-vmhm

[13] CMS/OEDA, "Medicare provider utilization and payment data: Part d prescriber summary table cy2015," 2017, [Online; accessed 7-November-2017]. [Online]. Available: https://data.cms.gov/Medicare-Part-D/Medicare-Provider-Utilization-and-Payment-Data-Par/qywy-pajd

[14] R. Paudel, W. Eberle, and D. Talbert, "Detection of anomalous activity in diabetic patients using graph-based approach," in *FLAIRS Conference*, 2016.

[15] L. Copeland, D. Edberg, A. K. Panorska, and J. Wendel, "Applying business intelligence concepts to medicaid claim fraud detection," *Journal of Information Systems Applied Research*, vol. 5, no. 1, p. 51, 2012.

[16] P. A. Ortega, C. J. Figueroa, and G. A. Ruz, "A medical claim fraud/abuse detection system based on data mining: A case study in chile." *DMIN*, vol. 6, pp. 26–29, 2006.

[17] G. J. Williams and Z. Huang, "Mining the knowledge mine," in *Australian Joint Conference on Artificial Intelligence*. Springer, 1997, pp. 340–348.

[18] W.-S. Yang and S.-Y. Hwang, "A process-mining framework for the detection of healthcare fraud and abuse," *Expert Systems with Applications*, vol. 31, no. 1, pp. 56–68, 2006.

[19] D. Thornton, R. M. Mueller, P. Schoutsen, and J. van Hillegersberg, "Predicting healthcare fraud in medicaid: a multidimensional data model and analysis techniques for fraud detection," *Procedia technology*, vol. 9, pp. 1252–1264, 2013.

[20] V. S. Iyengar, K. B. Hermiz, and R. Natarajan, "Computer-aided auditing of prescription drug claims," *Health care management science*, vol. 17, no. 3, pp. 203–214, 2014.

[21] B. S. U. Mendis, D. W. Murray, A. Sutinen, M. Tang, and Y. Hu, "Enhancing the identification of anomalous events in medicare consumer data through classifier combination," in *The 6th International Workshop on Chance Discovery (IWCD6)*, 2011, p. 1.

[22] V. Chandola, S. R. Sukumar, and J. C. Schryver, "Knowledge discovery from massive healthcare claims data," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 1312–1320.

[23] J. Liu, E. Bier, A. Wilson, T. Honda, K. Sricharan, L. Gilpin, J. Guerra-Gomez, and D. Davies, "Graph analysis for detecting fraud, waste, and abuse in healthcare data." in *AAAI*, 2015, pp. 3912–3919.