FLAIRS 2017 – Special Track in Health Care Informatics

DETECTION OF ANOMALOUS ACTIVITY IN DIABETIC PATIENTS USING GRAPH-BASED APPROACH

Ramesh Paudel, William Eberle and Doug Talbert

Department of Computer Science

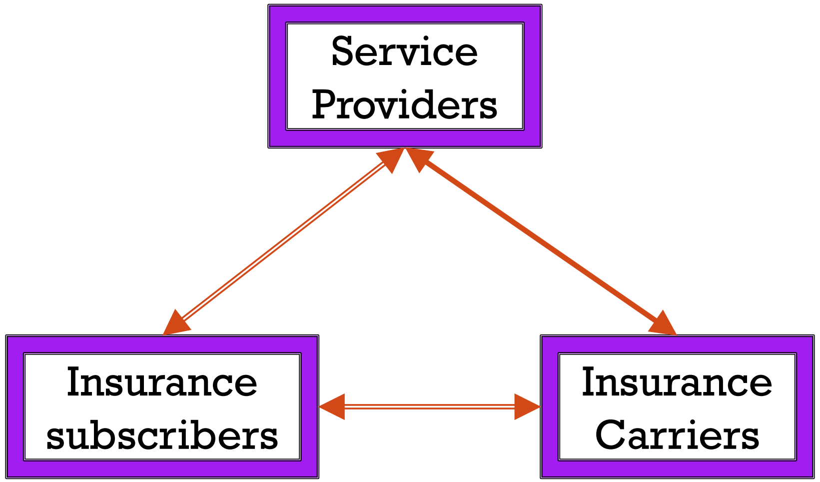
Tennessee Technological University

rpaudel42@students.tntech.edu, weberle@tntech.edu, and dtalbert@tntech.edu

Tennessee TECH

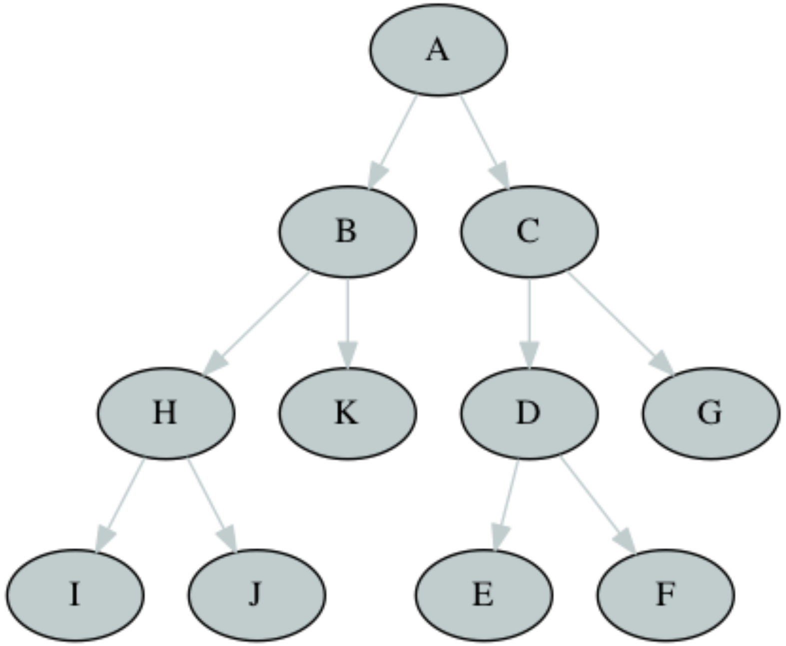
PROBLEM

- The total health care spending in 2014 accounted for 17.5% of the nation's GDP and is expected to rise to 20.1% by 2025.
- Roughly 1/3rd of health care spending can be attributed to fraud, waste, and abuse.
- Health care claims are complex because they involve multiple parties including service providers, insurance subscribers, and insurance carriers.
- In this paper, we introduce an approach for discovering health care fraud using *graph-based data mining*.

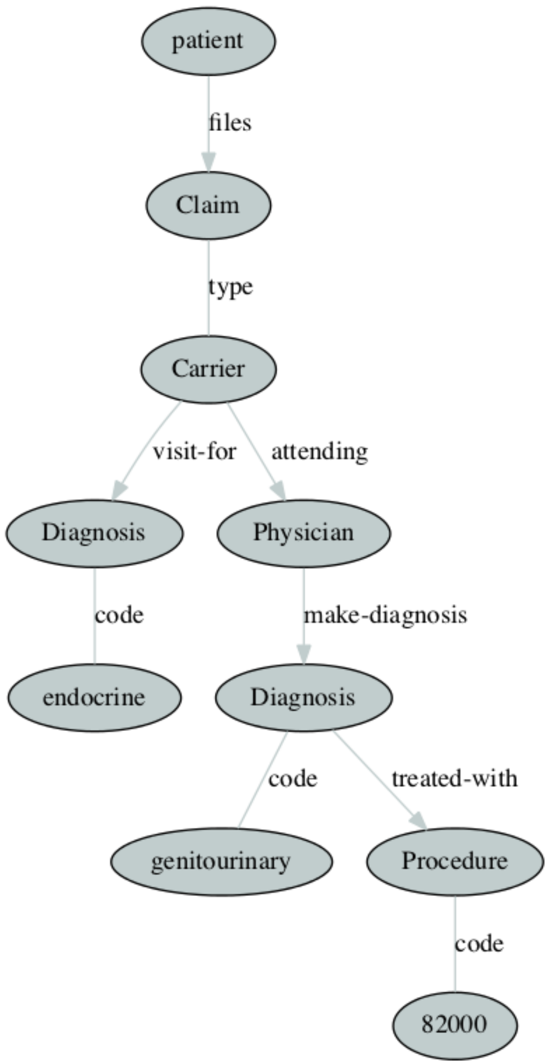


OVERVIEW

- Representing Medicare claim data as a graph will provide an intuitive and efficient method for detecting anomalies.
- To evaluate our hypothesis, we use the Graph-Based Anomaly Detection (GBAD) tool (www.gbad.info).
- We include the relationship between the patient, claims, physician, diagnosis, and what procedure was performed to treat that diagnosis as features for anomaly detection.



GRAPHS

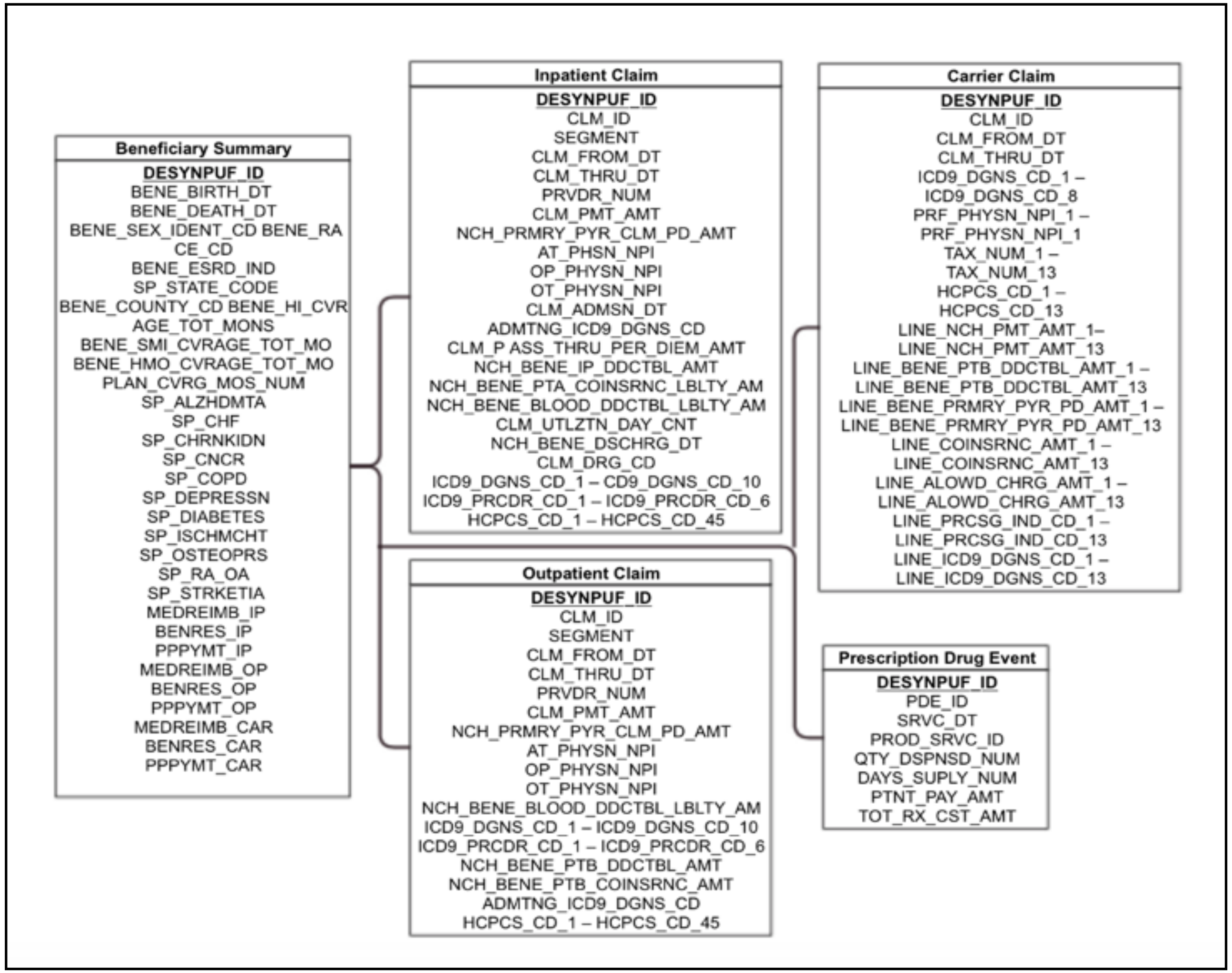


- If we consider the *entities* involved in the process of medical claims as *nodes*, and the *relationships* and *transactions* between the entities involved as *edges*, we can represent the entire process as a *graph*.
- Example:
  - Patient "files" a claim of the types "ip", "op", or "carrier".
  - Each claim can have an admitting diagnosis represented by a "visit-for" edge.
  - Each claim can also have a "Physician" who "make-diagnosis" - "Diagnosis", and it is "treated-with" - "Procedure".

DATA SET: DE-SYNPUF

- The dataset used for this research is the CMS Linkable 2008–2010 Medicare Data Entrepreneurs' Synthetic Public Use File (DE-SynPUF) dataset.
- Did not use all of the data
  - Focuses on a subset of involving diabetic patients in the state of Tennessee from 2009.

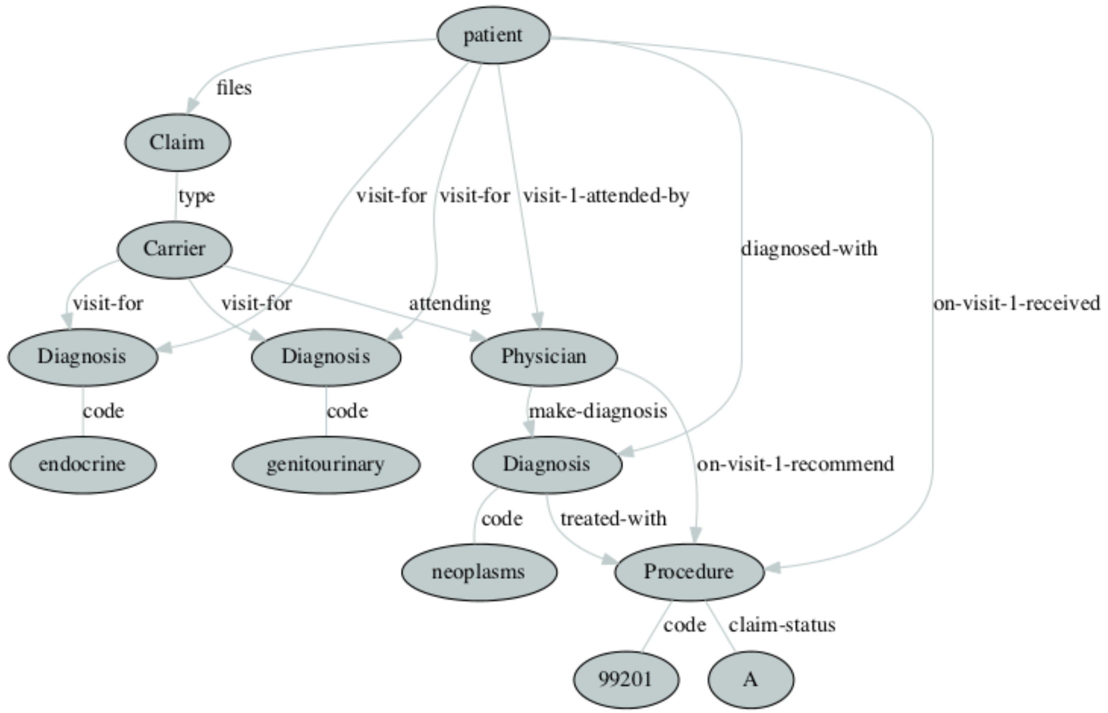
SN	Data Type	Number of rows
1	Beneficiaries	343,644
2	Carrier Claims	4,741,335
3	Inpatient Claims	66,773
4	Outpatient Claims	790,790
5	Prescription Drug Events	5,552,421



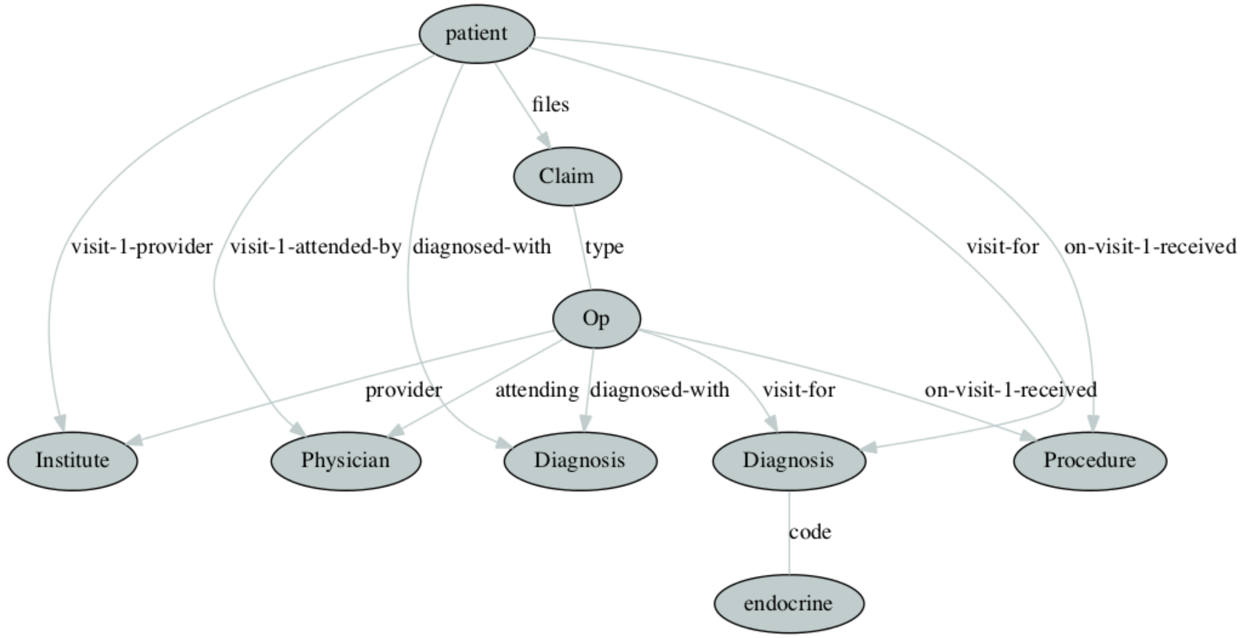
https://www.cms.gov/research-statistics-data-and-systems/downloadable-public-use-files/synpufs/de_syn_puf.html

GRAPH REPRESENTATION OF MEDICARE CLAIMS

Carrier claim graph has a total of 21,082 vertices and 32,214 edges representing 572 diabetic patient



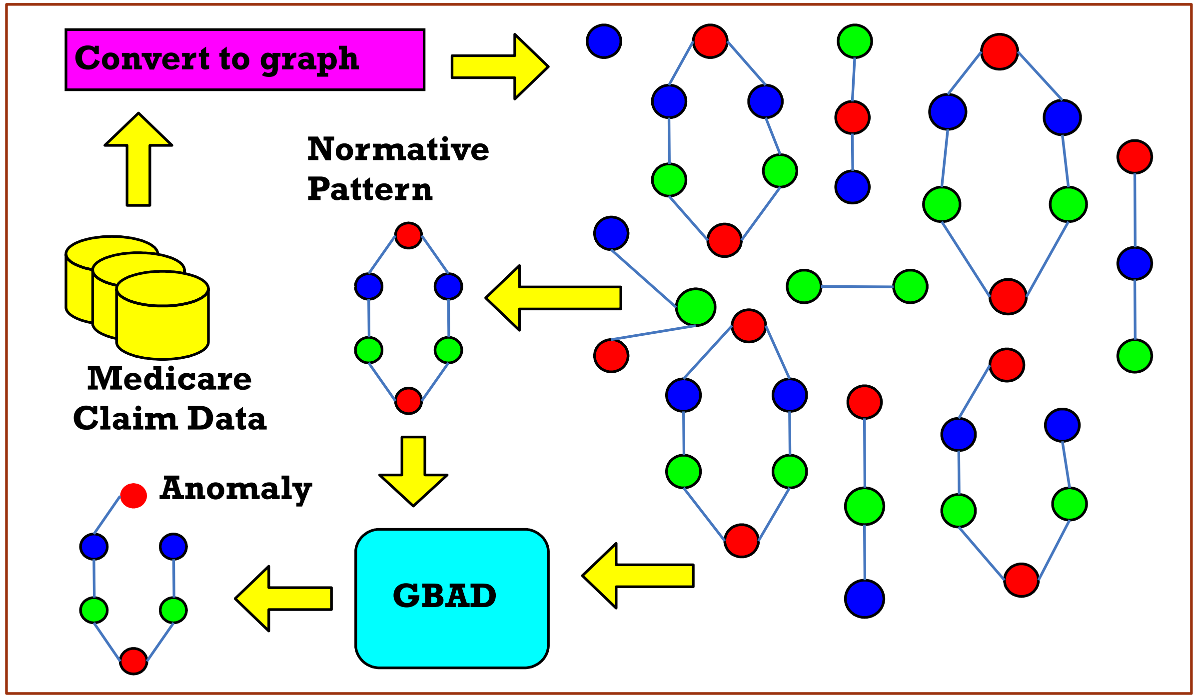
Ip-op claim graph has total of 1,469 vertices and 2,139 edges representing 62 diabetic patient



A visual representation of a carrier and ip-op claim subgraph (consisting of data for 1 patient each)

GRAPH-BASED ANOMALY DETECTION

- GBAD (www.gbad.info)
- Find normative highly-compressing pattern S
- Find closely-matching instances S_A of S
  - Missing nodes/edges (gathered along the way)
  - Additional nodes/edges (search a bit further)
  - Modified labels among structural matches
- $Pr(S_A)$ = # particular S_A / # all S_A's
- Anomaly score = $Pr(S_A) * dist(S_A,S)$



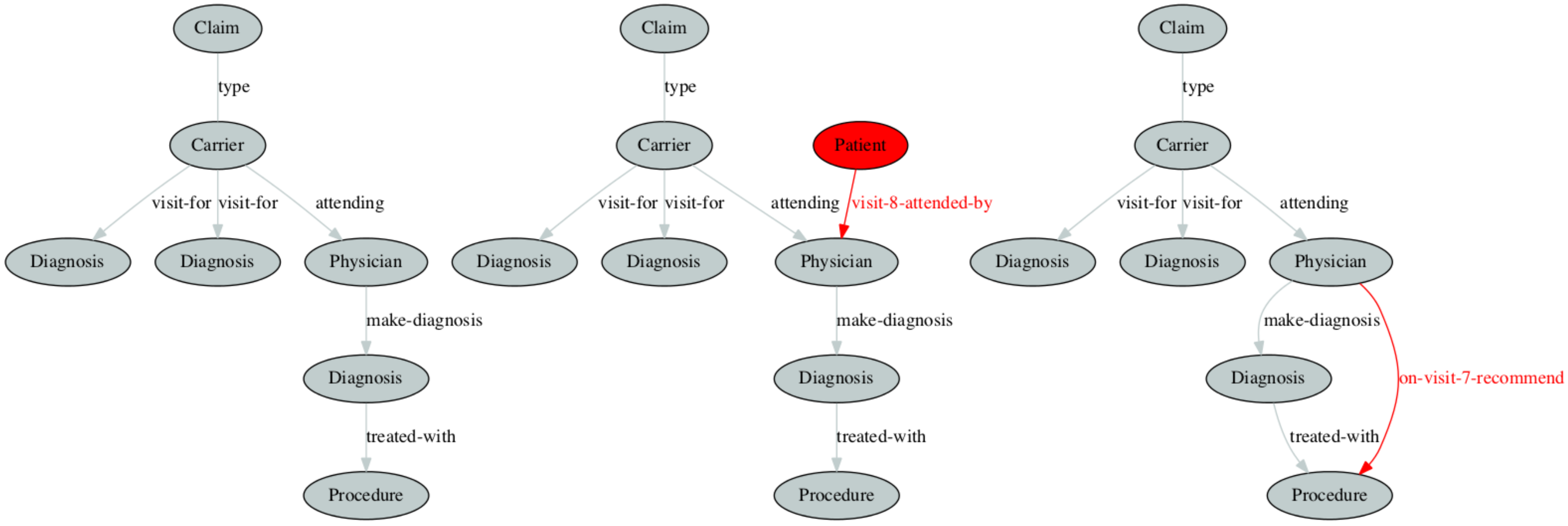
KNOWN FRAUD SCENARIOS

- ***Identity Theft:*** Stealing identity information and using it to submit fraudulent bills
- ***Phantom Billing:*** Billing for services that are not actually performed.
- ***Unbundling:*** Billing each stage of a procedure as if it were a separate treatment.
- ***Upcoding:*** Billing costlier services than the one actually performed.
- ***Bill Padding:*** Providing medically excessive or unnecessary services to a patient.
- ***Duplicate Billing:*** Submitting similar claims more than once.
- ***Kickbacks:*** A negotiated bribery in which a commission is paid to the bribe-taker
- ***Doctor shopping:*** Patient consults many physicians in order to obtain multiple prescriptions

EXPERIMENTAL RESULTS

- Discover several interesting anomalies...

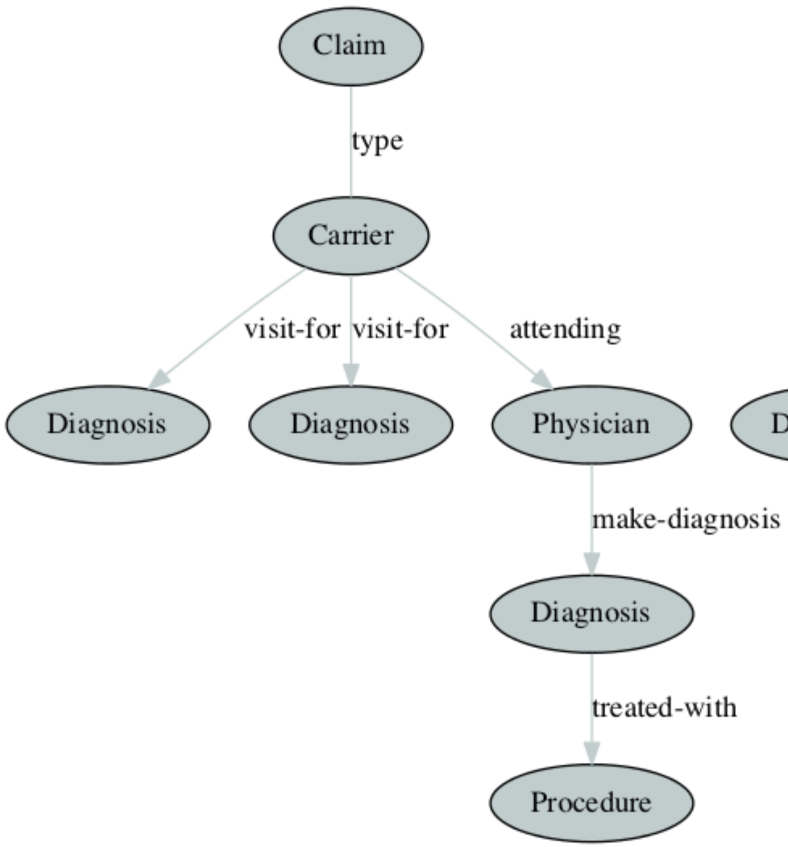
- ***Example #1***: Unusually high number of visits which could be a *doctor shopper* or a case of *identity theft*.
- ***Example #2***: Recommending the same procedure multiple times, which could be a case of *phantom billing* or *duplicate billing*.



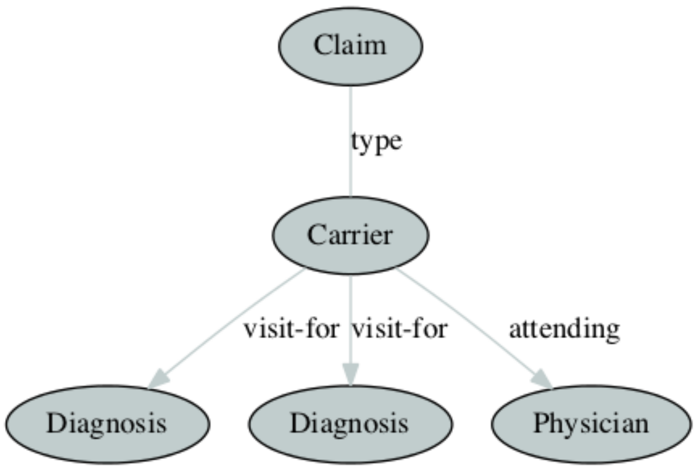
Normative Pattern   Unusual visit   Unusual procedure

EXPERIMENTAL RESULTS

- ***Example #3***: Missing node and edges.
  - Patient receives same procedures for treating the same diagnosis.
  - A potential scenario of *phantom billing*, or perhaps even the scenario of a *kickback* where the physician and patient are involved in filing fake claims.



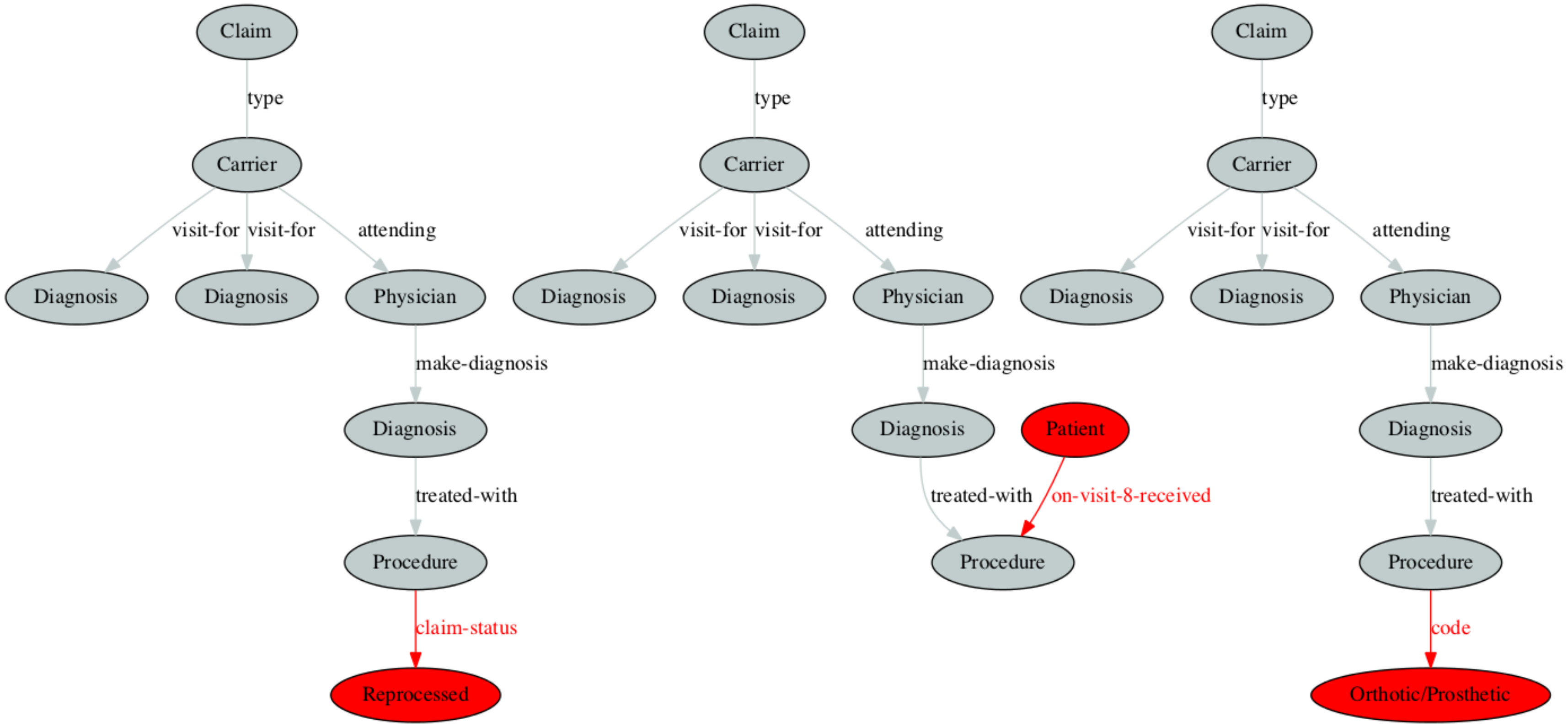
Normative Pattern



Missing Diagnosis

EXPERIMENTAL RESULTS

- ***Example #4:*** Usual claim status (Reprocessed).
- ***Example #5:*** Duplicate procedure on multiple visits.
- ***Example #6:*** Unusual procedure performed (Orthotic/Prosthetic)



SUMMARY

- We demonstrate how anomalies that are potentially fraudulent can be discovered using a graph based approach on data representing health care transactions.
- Unsupervised approach that exploits attributes of entities and as well as the relationships between entities for discovering anomalies.

FUTURE WORK

- Extend this approach to the entire Medicare claim dataset.
- Incorporate prescription drug claims
  - This will provide us with even more information as to potential fraudulent activities in the health care industry.
- Scalability
  - Investigate graph-partitioning and graph-sampling approaches that could handle streaming data.
- Phenotype discovery
  - Patients with certain diseases may have certain phenotypic groups based on their comorbidity characteristics because they require totally different management and treatment paths.
  - We plan to further investigate these phenotypic groups.

THANK YOU

QUESTION??