UNSUPERVISED ANOMALY DETECTION IN FINANCIAL TRANSACTION GRAPHS

Coen Valk

Agenda

- Problem Statement
- Method
- Datasets
  - *Don2Com*
  - *Project Caviar - Canadian Drug Trafficking Dataset*
  - *Bitcoin Transaction Dataset*
- Algorithms
  - *Direct Neighbor Outlier Detection Algorithm (DNODA)*
  - *Community Neighbor Algorithm (CNA)*
  - *Global Outlier Detection Algorithm (GLODA)*
  - *OddBall*

Problem

- What attributes do illegitimate transactions possess?
- How do we find which transactions are illegitimate?
- How accurate is unsupervised anomaly detection in financial networks?

Method

- Apply all three algorithms to all original datasets
- Observe which nodes and edges are consistently flagged as anomalous
- Intentionally add anomalous nodes to each graph
  - *10%*
  - *25%*
  - *50%*
  - *80%*
  - *100%*
- Observe how many of the anomalous nodes are flagged

Datasets - Don2Com

- Federal Election Commission data
- Weighted graph of donations from individuals to committees and political candidates
- Also used by the 2010 paper detailing the OddBall algorithm
- Can we scope out meddling in elections?

Datasets – Project Caviar

- Dataset detailing links between drug trafficking deals
- Show the limits of an unsupervised algorithm
  - *All these transactions would be considered illegitimate or anomalous in a graph with larger context*

Datasets - Bitcoin Transaction Dataset

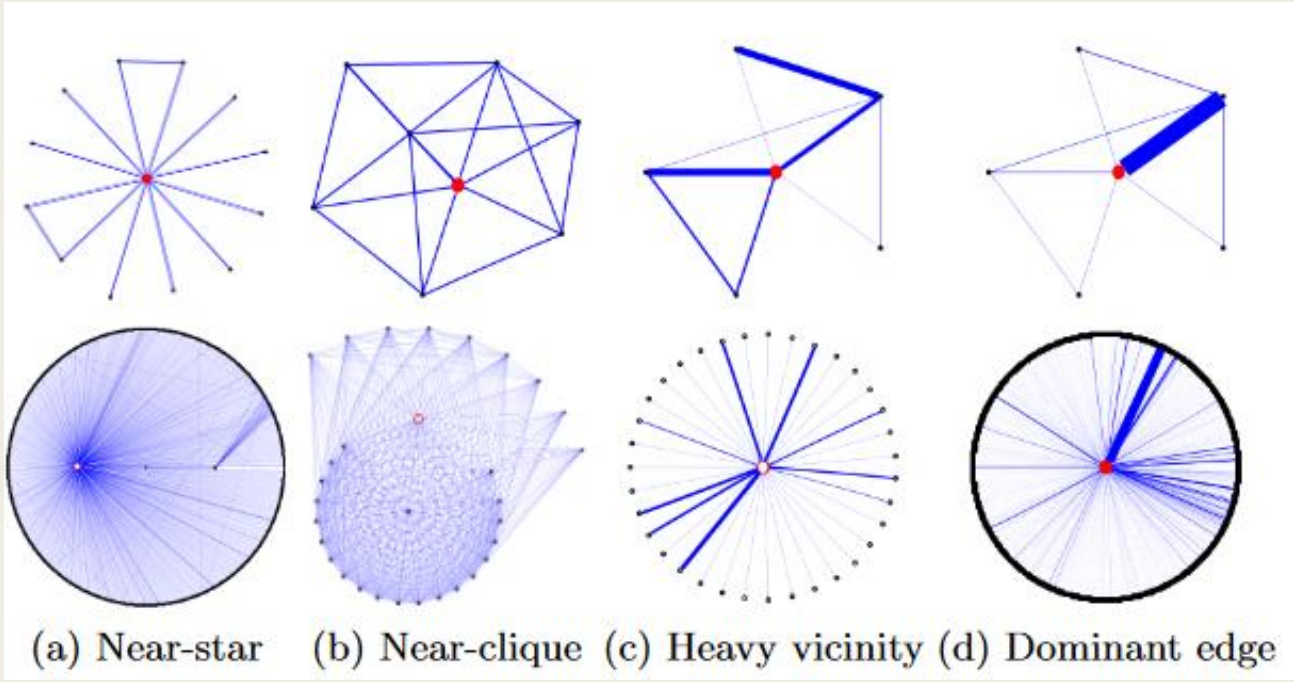
- Large weighted dataset of transactions over bitcoin blockchain
- Show the “ideal scenario” of anomaly detection
  - *Network contains both legitimate and illegitimate nodes*
- Stress test the performance of each algorithm
  - *Dataset contains several million transactions*

Algorithms – DNODA, CNA, GLODA

- Direct Neighbor Outlier Detection Algorithm (DNODA)
  - *Determine whether a node is an outlier based on the data of direct neighbors of that node*
- Community Neighbor Algorithm (CNA)
  - *Determine whether a node is an outlier based on the data of direct neighbors of that node*
  - *Several methods for community detection, as we've seen in class*
- Global Outlier Detection Algorithm (GLODA)
  - *Determine whether a node is an outlier based on the data of all other nodes in the graph*

Algorithms - OddBall

- Identifies 4 different kinds of anomalous vertices
  - *Near-star*
  - *Near-clique*
  - *Heavy vicinity*
  - *Dominant edge*




QUESTIONS?

References

- Akoglu, Leman, et al. “Oddball: Spotting Anomalies in Weighted Graphs.” *Advances in Knowledge Discovery and Data Mining Lecture Notes in Computer Science*, 2010, pp. 410–421., doi:10.1007/978-3-642-13672-6_40.
- Vengertsev, Dmitry, and Hemal Thakkar. *Anomaly Detection in Graph: Unsupervised Learning, Graph-Based Features and Deep Architecture*.
- Coscia, Michele and Viridiana Rios (2012). Knowing Where and How Criminal Organizations Operate Using Web Content. CIKM, 12 (October – November).