Network-based Anomaly Detection for Insider Trading

Sameera Bammidi

Divya Vajja

What is Insider Trading?

- Types of Insider Trading.
- Form - 4 with US Securities and Exchange Commission (SEC).
- The Electronic Data Gathering, Analysis, and Retrieval System (EDGAR).
- Hundreds of millions of dollars in profits.

Data

- MySQL database prepared from the data scraped from Insider Monkey and Google Finance.
- Got this data from our instructor and Teaching Assistant.

Total Companies	12,485
Total Insiders	70,408
Total Sale Transactions	757,194
Total purchase Transactions	311,013

Table 1: Statistics

Anomaly detection by Pairwise comparison

- Processed the data to construct networks based on the trading trends of each pair of insiders.
- Constructed graphs and extracted connected components from the graphs to analyze for anomalies.
- Worked on Purchase and sale networks independently.

Similarity based approach

Network	Nodes	Connected Components
Sale	1476	530
Purchase	1360	380

Table 2: Network Statistics (based on S)

- If a unique pair of insiders traded on at least 5 common dates compute similarity score.
- Similarity measure of X_c and Y_c of a company C:

$$S\,(X_C, Y_C) = \frac{\left(\sum_{i=1}^{|X_C|} \sum_{j=1}^{|Y_C|} I(x_i, y_j)\right)^2}{(|X_C| \times |Y_C|)}$$

I() = 1, if x_i=y_j, 0 otherwise

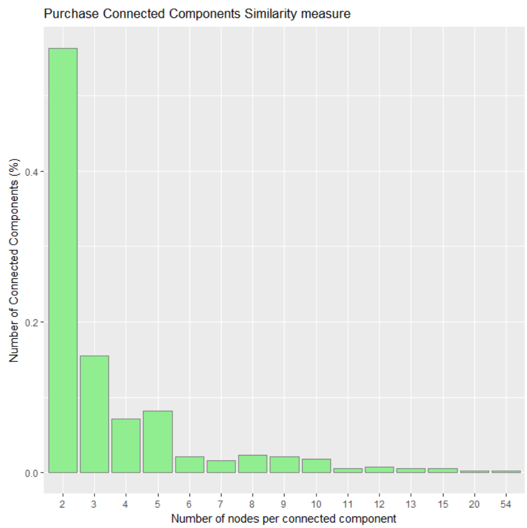
- Threshold > 0.5
- Constructed egonets, number of nodes V_u and number of edges E_u of an ego node u.
- Egonets are analyzed for discovering anomalies.
- Plotted V_u against E_u for all the egonets across all companies.
- Computed outlier scores for each ego node, to measure the deviation of each node u from the power law.

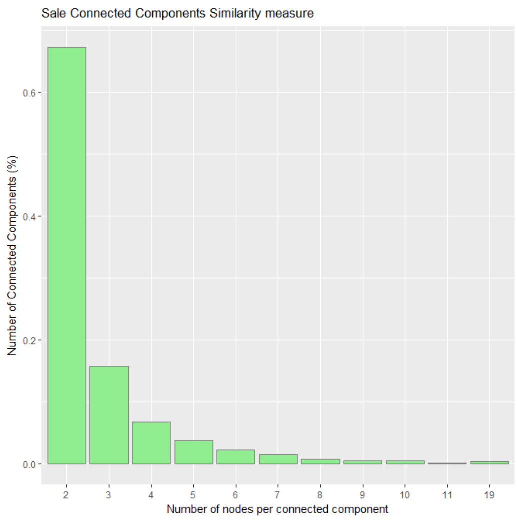
$$\text{Score}\,(u) = \frac{\max(E_u, f(V_u))}{\min(E_u, f(V_u))} \times (\log(|E_u - f(V_u)| + 1))$$

- Computed the Local Outlier Factor (LOF)
- Set value of K to 5
- Computed

TotalOutlierScore(u) = Score(u) + LOF(u)

Similarity based approach





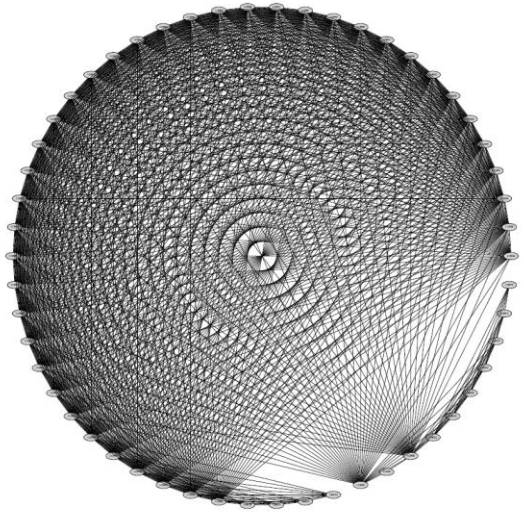


Figure 1(a): Purchase: Connected Component of International Speedway Corp Class A and Class B

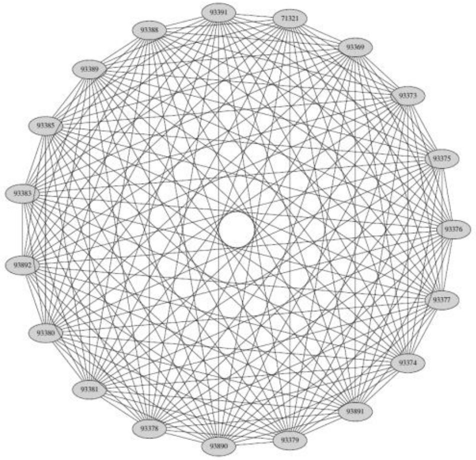


Figure 1(b): Sale: Connected Component of Vantiv. Inc

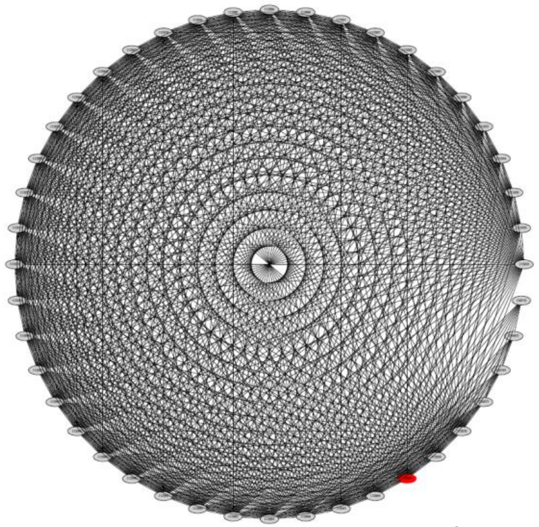


Figure 2(a): Purchase: Egonet of insider with highest outlier score of International Speedway Corp Class A and Class B

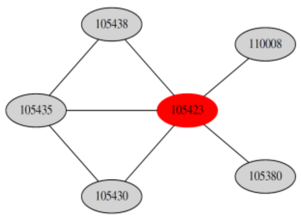


Figure 2(b): Sale: Egonet of insider with highest outlier score of WINN-Dixie Stores, Inc.

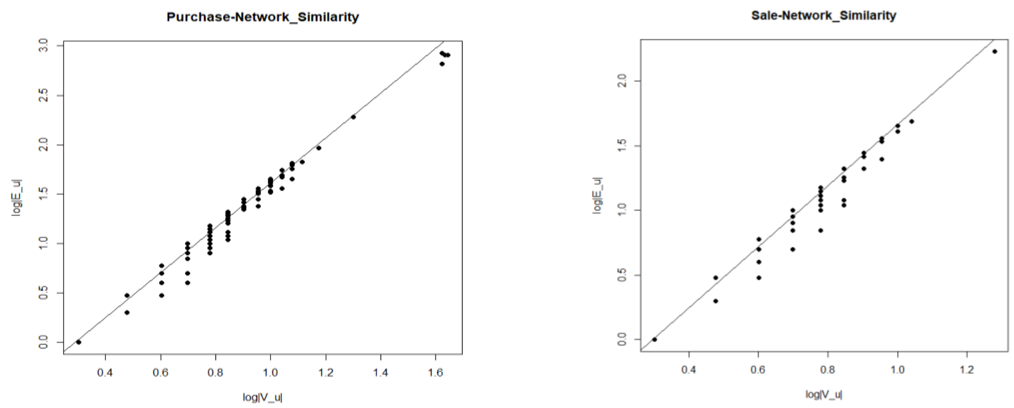
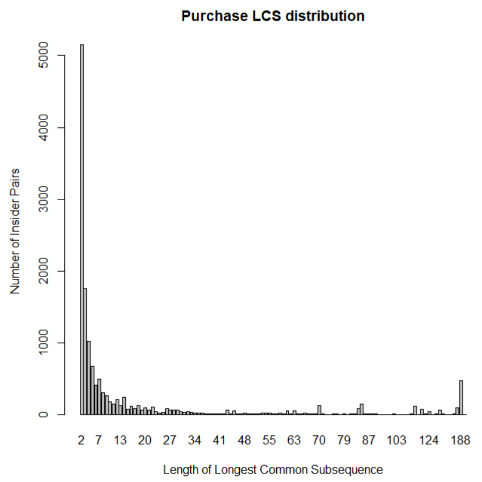


Figure 3: Power Law Fitting (Similarity based): (a) purchase, (b) sale

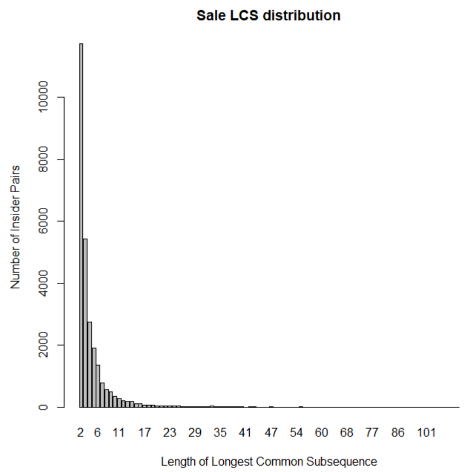
Longest Common Subsequence based approach

- To account for the temporal ordering of trade dates
- If a unique pair of insiders in a company shared a sub-sequence of dates of length at least t, add an edge
- Constructed graphs and extracted connected components from them
- Chose the threshold t based on the distribution of length of longest common subsequence among the traders in both purchase and sale networks
- Considered insiders with t > 10 for purchase network and t > 5 for sale network
- Constructed egonets and analyzed them for discovering anomalies as done for similarity.
- Plotted V_u against E_u for all the egonets across all companies.
- Computed outlier scores, LOF and total outlier scores for each ego node.

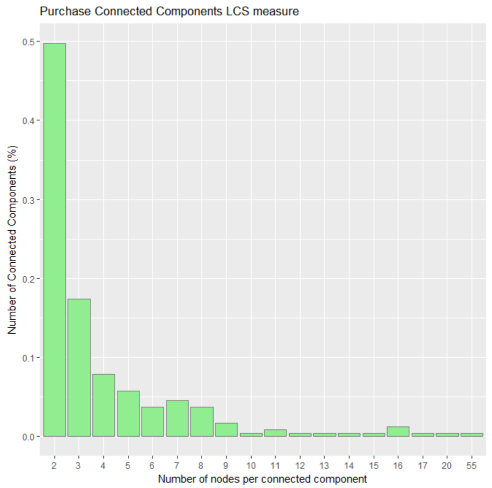


Network	Nodes	Connected Components
Sale	3819	1099
Purchase	977	241

Table 3: Network Statistics (LCS-based)



Longest Common Subsequence based approach



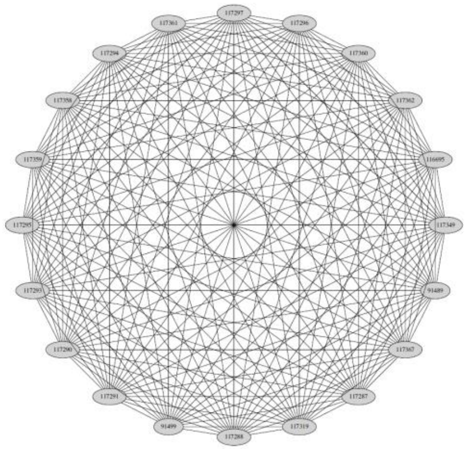


Figure 4(a): Purchase (LCS-based): Connected Component of Hyster-Yale Materials Handling Inc.

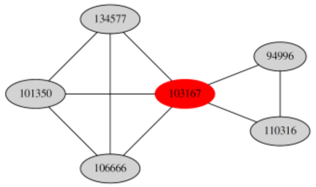


Figure 5(a): Purchase (LCS-based): Egonet of insider with highest outlier score of First Mid-Illinois Bancshares

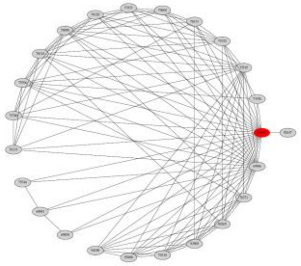
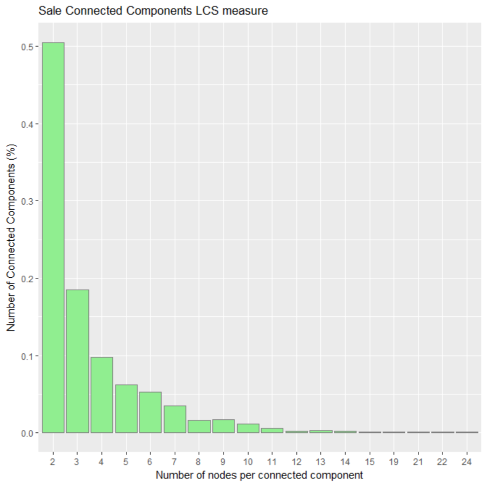


Figure 5(b): Sale (LCS-based): Egonet of insider with highest outlier score of General American Investors



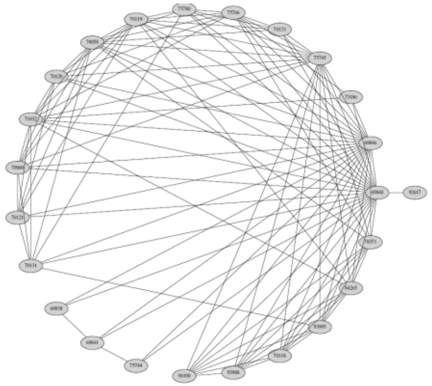


Figure 4(b): Sale (LCS-based): Connected Component of General American Investors

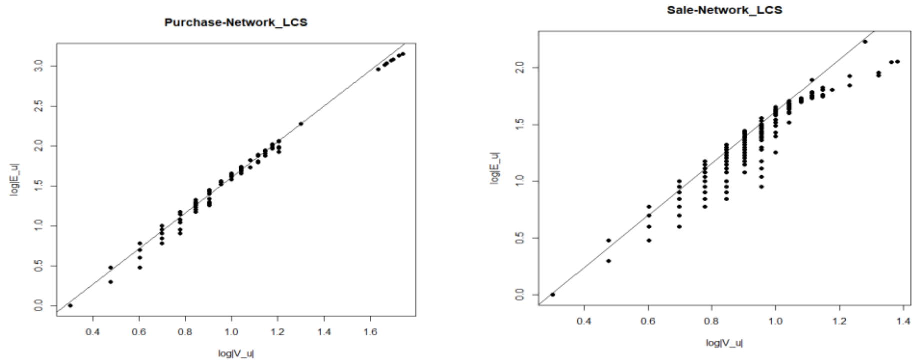


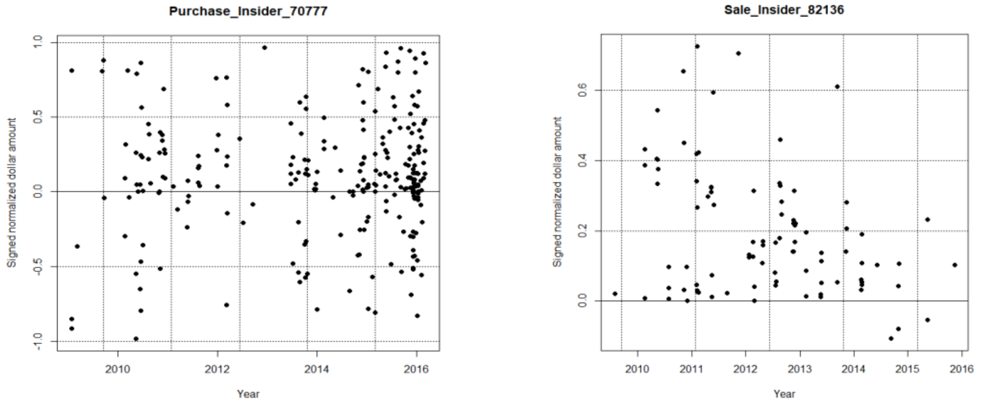
Figure 6: Power Law Fitting (LCS-based): (a) purchase, (b) sale

Evaluation with signed normalized dollar amount

- Quantified the above results by looking at the profit made during the sequence of dates
- Computed the signed normalized dollar amount
- Profit made in two scenarios
  - Buy stocks at a price lower than the closing price
  - Sell stocks higher than the closing price of the stock on that day
- Ranges from -1 to 1

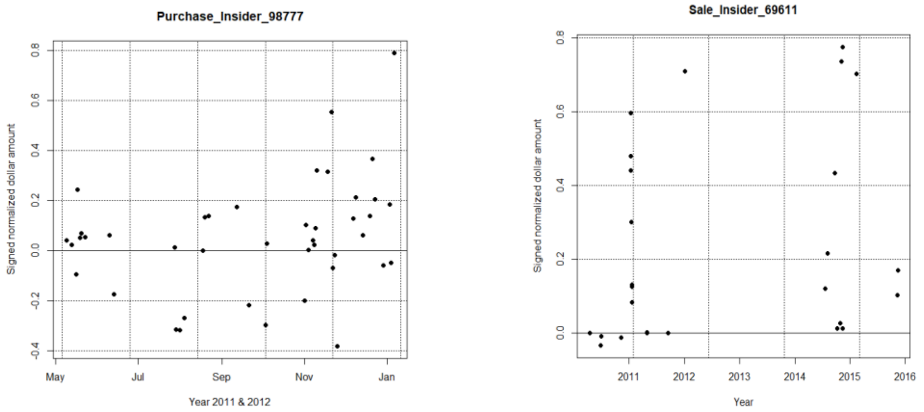
$$R = \frac{Transaction\ Price\ X\ \sum Shares\ Traded}{Dollar\ Volume}$$

Dollar Volume DV = Total number of shares traded on a day x market closing price of the stock on that day



(a) (b)

Figure 7: Time series of Signed Normalized Dollar Amount (LCS-based) egonets (a) Purchase network. Total Outlier Score is 1.054004865 which ranked 70. (b) Sale network. Total Outlier score is 1.013247896 and ranked least of all insiders. Filtered for R > 0.09 and looked for insiders with more number of trades.



(a) (b)

Figure 8: Time series of Signed Normalized Dollar Amount (Everyone-based) (a) Purchase network. (b) Sale network. Found interesting cases which were not found by LCS-based approach. Filtered for R > 0.09 and looked for insiders with more number of trades.

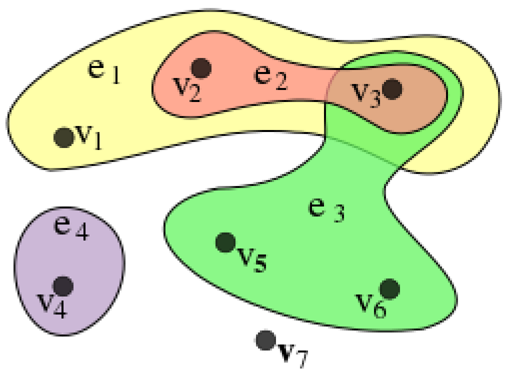
HyperGraph

- Generalization of a graph in which an edge can join any number of vertices
- Represented as a pair H = (V,E)
- Where V is a set of elements called nodes or ***vertices*** and E is a set of non-empty subsets of V called ***hyperedges*** or edges

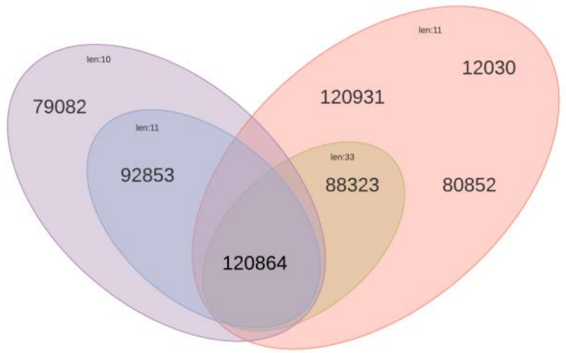
$V=\{v_1,v_2,v_3,v_4,v_5,v_6,v_7\}$

$E=\{e_1,e_2,e_3,e_4\} =\{ \{v_1,v_2,v_3\} , \{v_2,v_3\} ,$

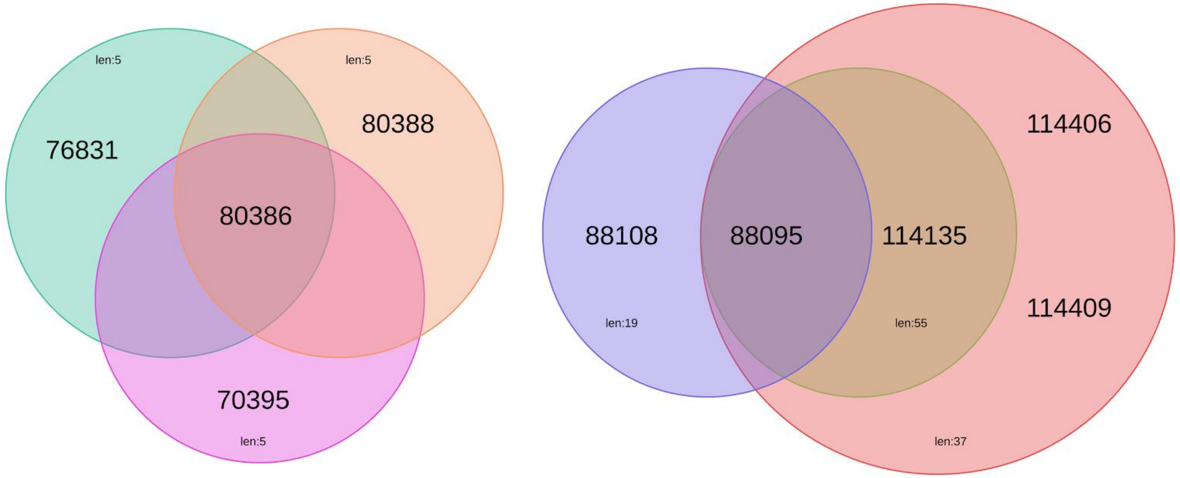
$\{v_3,v_5,v_6\} , \{v_4\}\}$



Hypergraphs for Insider data



Hypergraph for Ticker-QNBC for purchase



Hypergraph for Ticker-CUK for sale

Thank You!