Exploring Large Scale Insider Trading Data: Network Patterns & Discoveries

Acar Tamersoy, Georgia Tech, tamersoy@gatech.edu
Bo Xie, GT
Stephen Lenkey, SEC
Bryan Routledge, CMU
Polo Chau, GT
Sham Navathe GT

Summary

How do corporate insiders really trade? Does the CEO of a company trade differently than the CFO? We performed the first academic, large-scale analysis of the full insider trading data from SEC, from 1986 to 2012, totaling more than 12M transactions, among 370K insiders. We found that insiders form tightly-connected clusters in which trade related information might propagate.

Insiders and (Illegal) Insider Trading

Financial regulators are interested in applying **data mining** techniques to detect illegal trades among insiders (e.g., CEO, directors), by analyzing their **Form 4 filings.**

We performed the **first, large-scale** academic study of the complete Form 4 filings from SEC.

Insiders engage in **illegal insider trading** when they exploit their roles and use **nonpublic inside information** to profitably trade for their companies' stock.

Form 4 Dataset

SEC requires insiders to disclose their trades within 2 days via Form 4, publicly available from SEC's EDGAR system (www.sec.gov/edgar.shtml).

We analyzed all forms from January **1986** to August **2012**.

Insiders	370,627
Companies:	15,598
Transactions:	12,360,325
Sale transactions:	3,206,175
Purchase transactions:	1,206,038

Each form contains insider's name + company + role in the company (from CEO to Retired), transaction date and type (we focused on Purchases and Sales), etc.

Patterns, Observations, & Analysis

We conjecture that some insiders share nonpublic inside information with each other. We build **insider networks** where nodes are insiders and edges connect insiders trading similarly.

Our **similarity function** takes as input the transaction times of two insiders of the same company and returns a similarity scale based on the transaction timings.

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

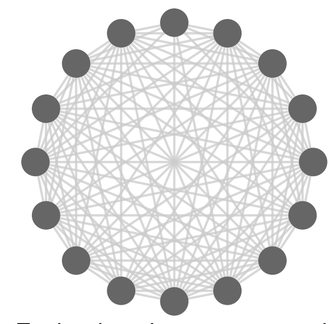
I(x, y) = 1 if x = y
I(x, y) = 0 o/w

We compute a similarity value for each pair of insiders (X_C, Y_C) of company C. If both insiders traded at least h_z times and their similarity value is at least h_m, we include nodes and an edge for these insiders to our network.

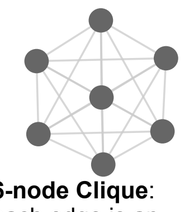
Network	Nodes	Edges	Connected Components
Sale	1630	1473	623
Purchase	1678	2656	489

$h_z = 5$
$h_m = 0.5$

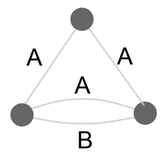
Connected Components



Each edge above corresponds to an Electrical Utilities Company



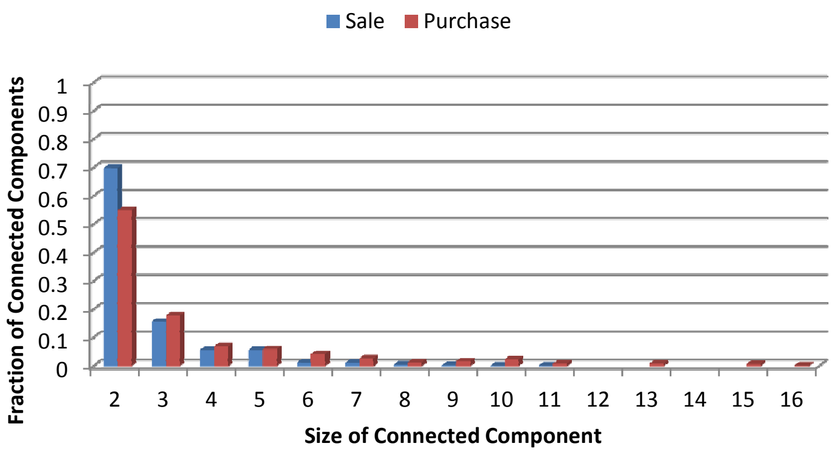
6-node Clique: each edge is an electrical company



Triangle: company A: biotech; company B: medical supplies

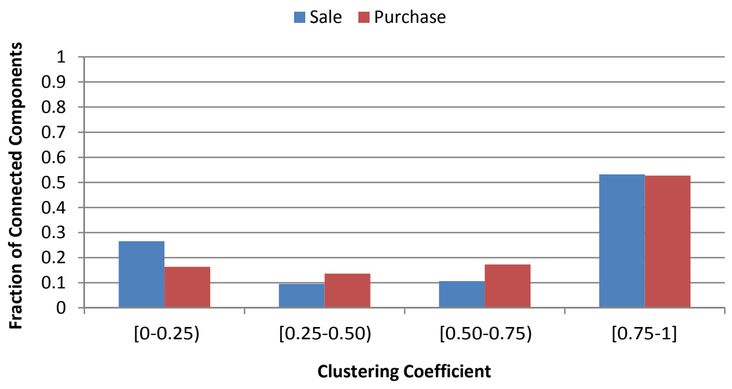
Chain: each edge is an electrical utilities company

Sizes of Components



Insiders form **small clusters.**

Density of Components



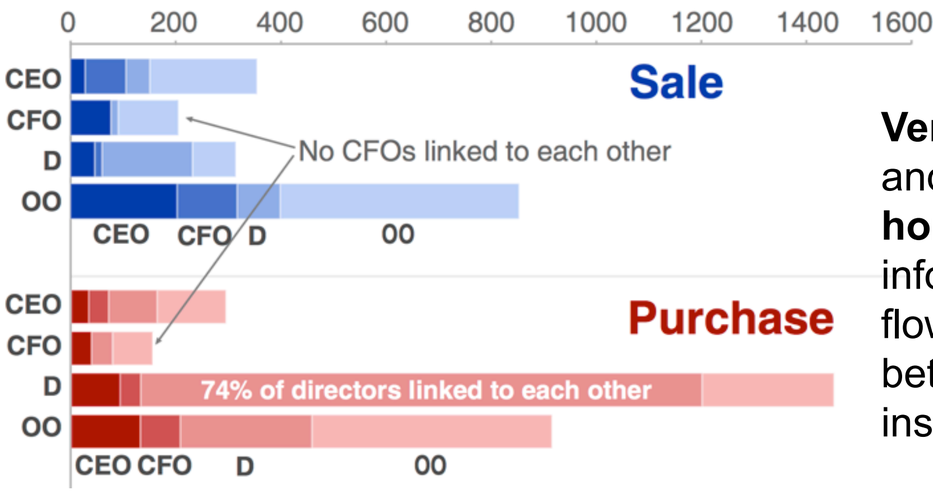
Tightly-connected clusters: trade-related information may propagate easily.

Number of Companies in Components

	Number of Companies						
	1	2	3	4	5	6	7
Sale	96.8%	2.7%	-	0.3%	-	-	0.2%
Purchase	97.5%	2.5%	-	-	-	-	-

Trade-related information flow about multiple companies is **not** likely to occur between insiders.

Roles of Insiders in Components



Vertical and **horizontal** information flow between insiders.

Discussion of Case Studies

Insiders from the **same family** trade similarly, ~7% of the directly connected insiders share the same last names.

All insiders in the chain below belong to the same investment firm, who may be **acting on behalf of** the firm.

Company **A** Retail — Company **B** Retail — Company **C** Semiconductors — Company **C** Semiconductors — Company **C** Semiconductors — Company **D** Semiconductors