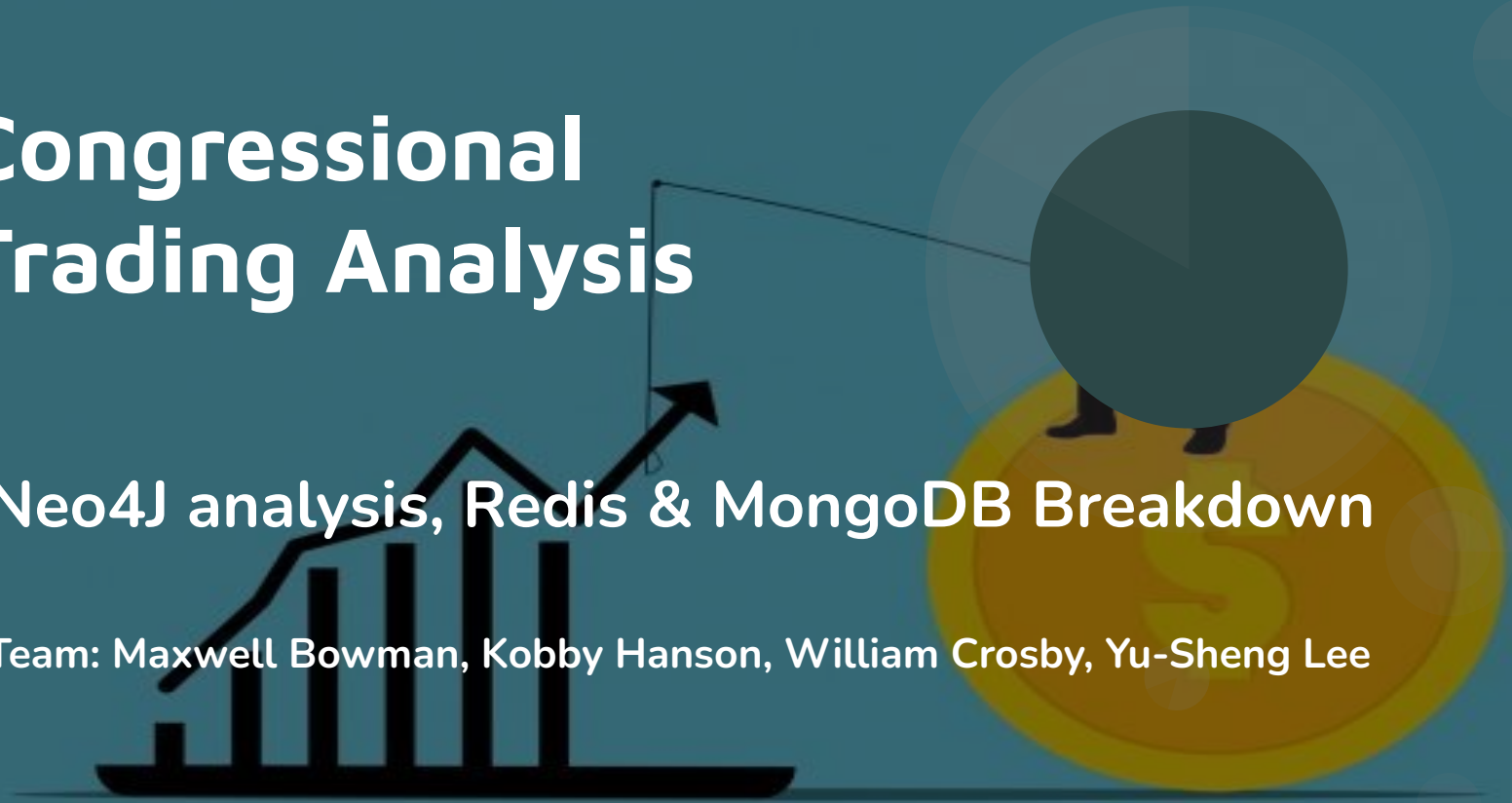
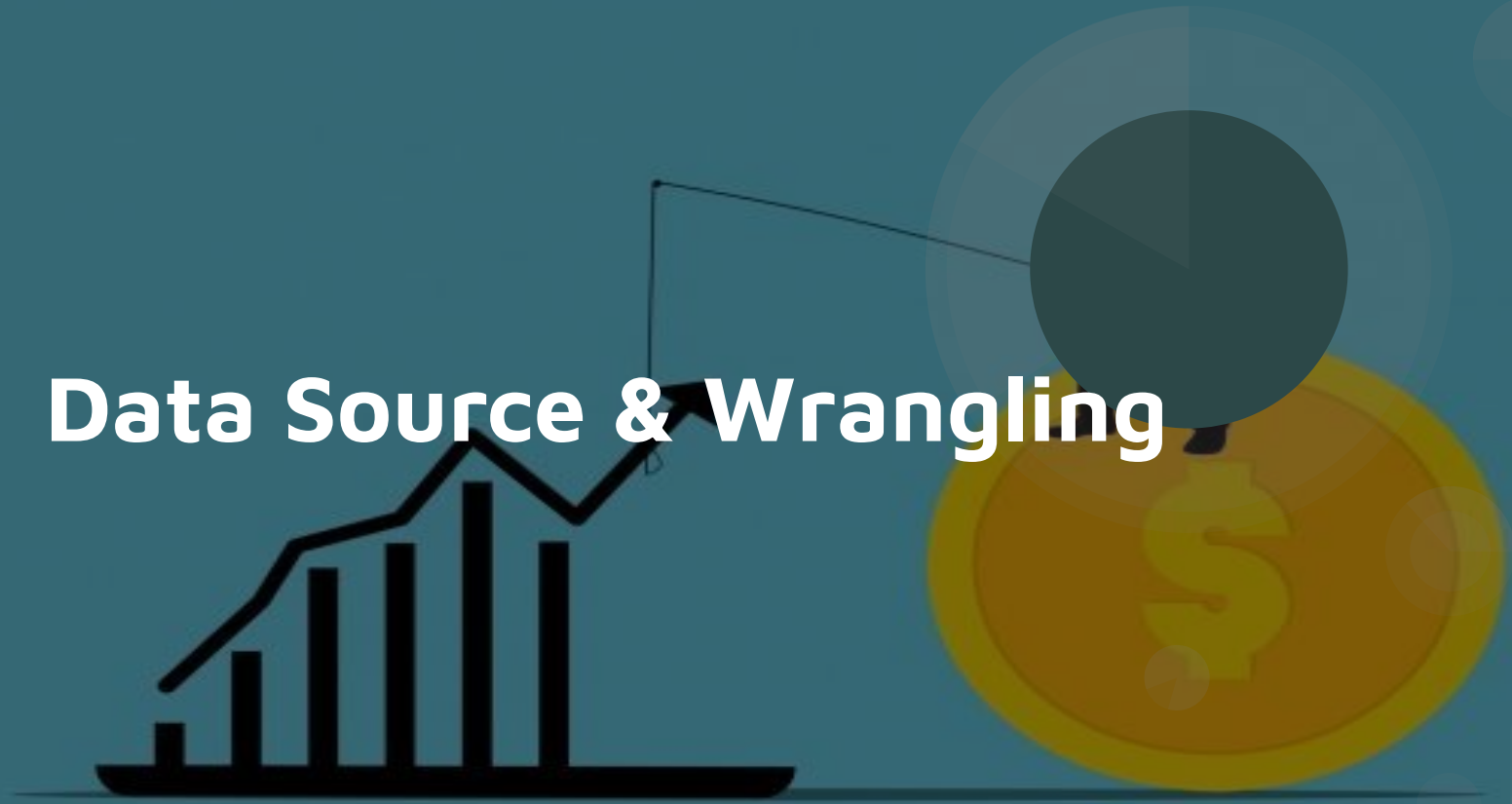
Congressional Trading Analysis



Neo4J analysis, Redis & MongoDB Breakdown

Team: Maxwell Bowman, Kobby Hanson, William Crosby, Yu-Sheng Lee

Data Source & Wrangling



Data Source

Source: House Stock Watcher API (https://housestockwatcher.com/api)

Data: US House representative financial disclosures (CSV/JSON)

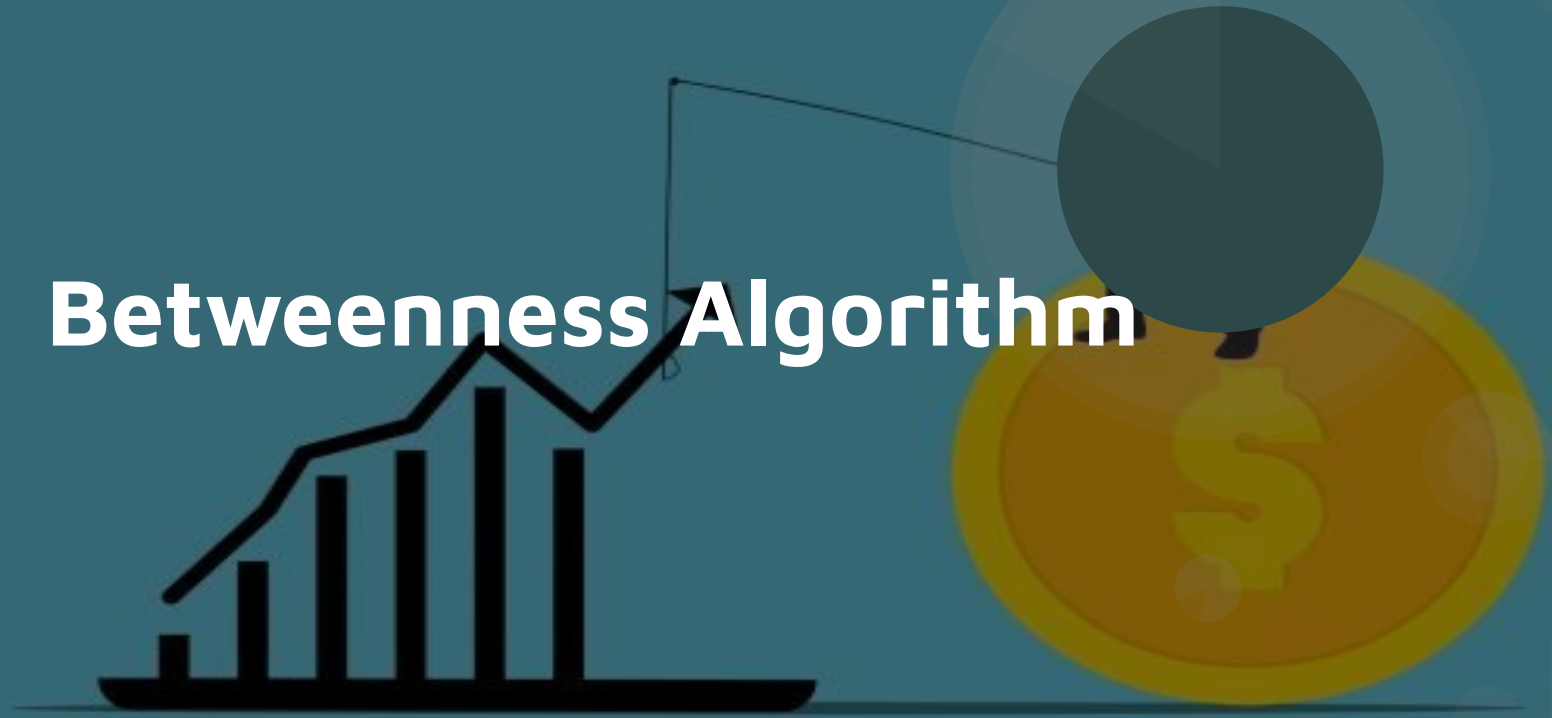
Includes: Trade timing, ownership, asset (ticker, description, sector, industry), transaction type, amount (range), capital gains (> $200), representative info, source link

Limitations: Amount ranges, data inconsistencies (blank industry/sector, transaction dates)

disclosure_year	disclosure_dat	transaction_da	owner	ticker	asset_descripti	type	amount	representative	district	state	ptr_link	cap_gains_over	industry	sector	party
2021	10/4/2021	9/27/2021	joint	BP	BP plc	purchase	$1,001 - $15,000	Virginia Foxx	NC05	NC	https://disclosures-	FALSE	Integrated oil Comp	Energy	Republican
2021	10/4/2021	9/13/2021	joint	XOM	Exxon Mobil Corpor	purchase	$1,001 - $15,000	Virginia Foxx	NC05	NC	https://disclosures-	FALSE	Integrated oil Comp	Energy	Republican
2021	10/4/2021	9/10/2021	joint	ILPT	Industrial Logistics	purchase	$15,001 - $50,000	Virginia Foxx	NC05	NC	https://disclosures-	FALSE	Real Estate Investm	Real Estate	Republican
2021	10/4/2021	9/28/2021	joint	PM	Phillip Morris Intern	purchase	$15,001 - $50,000	Virginia Foxx	NC05	NC	https://disclosures-	FALSE	Farming/Seeds/Mill	Consumer Non-Dur	Republican
2021	10/4/2021	9/17/2021	self	BLK	BlackRock Inc	sale_partial	$1,001 - $15,000	Alan S. Lowenthal	CA47	CA	https://disclosures-	FALSE	Investment Banker:	Finance	Democrat
2021	12/1/2021	12/1/2021	joint	AXP	American Express C	purchase	$1,001 - $15,000	Aston Donald McEa	VA04	VA	https://disclosures-	FALSE	Finance: Consumer	Finance	Democrat
2021	12/1/2021	11/30/2021		KPLTW	Katapult Holdings I	purchase	$1,001 - $15,000	Austin Scott	GA08	GA	https://disclosures-	FALSE			Republican
2021	12/1/2021	11/18/2021		AMD	Advanced Micro De	sale_full	$50,001 - $100,000	Thomas Suozzi	NY03	NY	https://disclosures-	FALSE	Semiconductors	Technology	Democrat
2021	12/1/2021	11/18/2021		AAPL	Apple Inc	sale_full	$50,001 - $100,000	Thomas Suozzi	NY03	NY	https://disclosures-	FALSE	Computer Manufac	Technology	Democrat
2021	12/1/2021	11/24/2021		MSFT	Microsoft Corporati	purchase	$50,001 - $100,000	Thomas Suozzi	NY03	NY	https://disclosures-	FALSE	Computer Software	Technology	Democrat

Data Cleaning

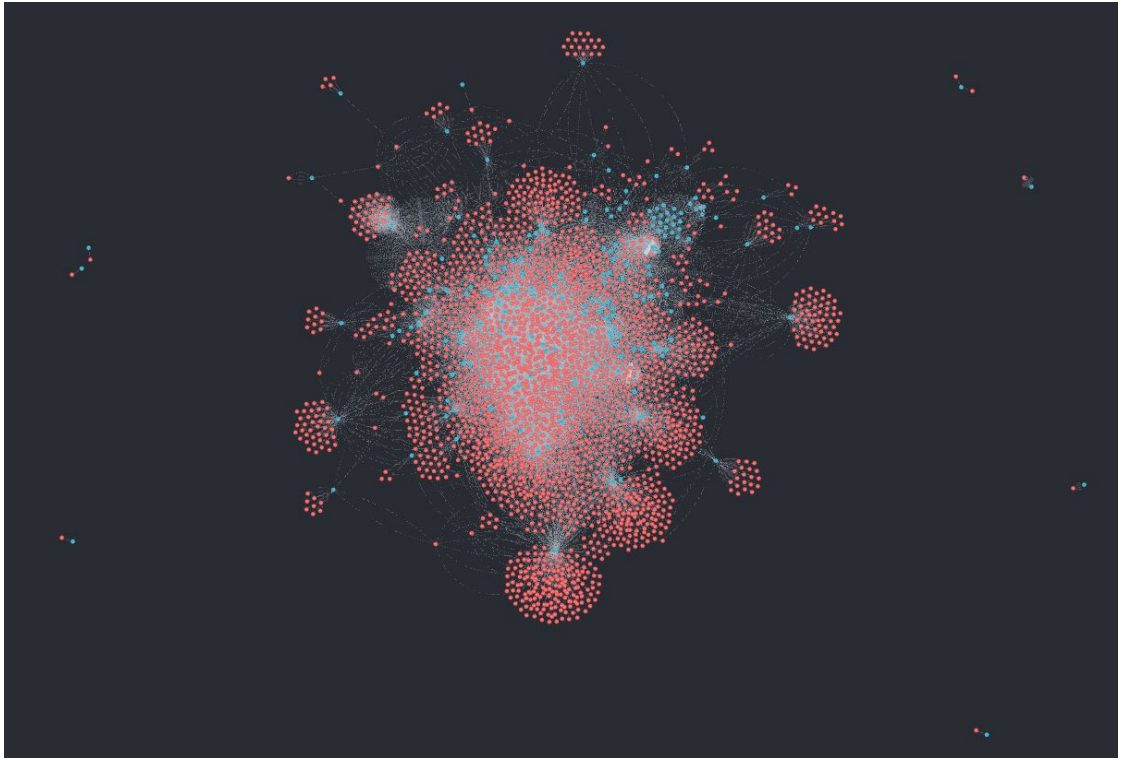
Betweenness Algorithm



Representatives and Stock Tickers

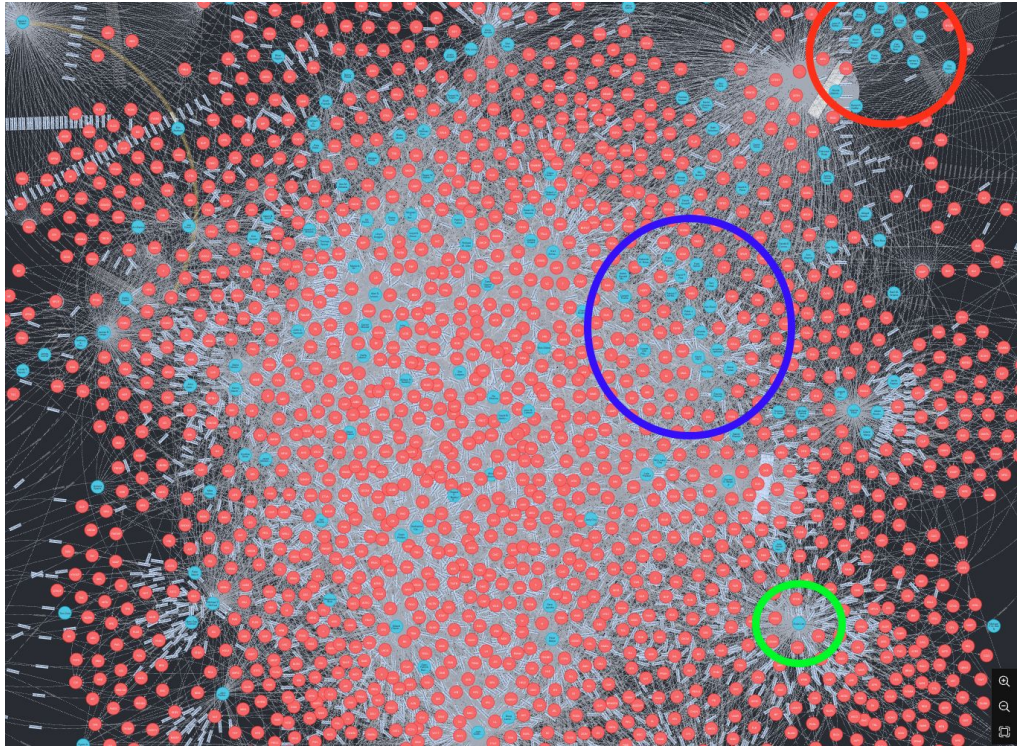
The network of representatives (blue nodes) and stock tickers (red nodes) highlights outliers and sub networks or clusters of interest

Analyzing the areas of high and low betweenness could inform network characteristics and dynamics between politicians, stock selection, and purchasing behavior individually or as groups



Pathways and Clusters

- High betweenness blue nodes (politicians) could indicate potentially a greater influence within the network by a politician
- Clusters of politicians isolated from most of the network
- Nodes with high betweenness between stock tickers and low betweenness to other politicians



Why Betweenness?

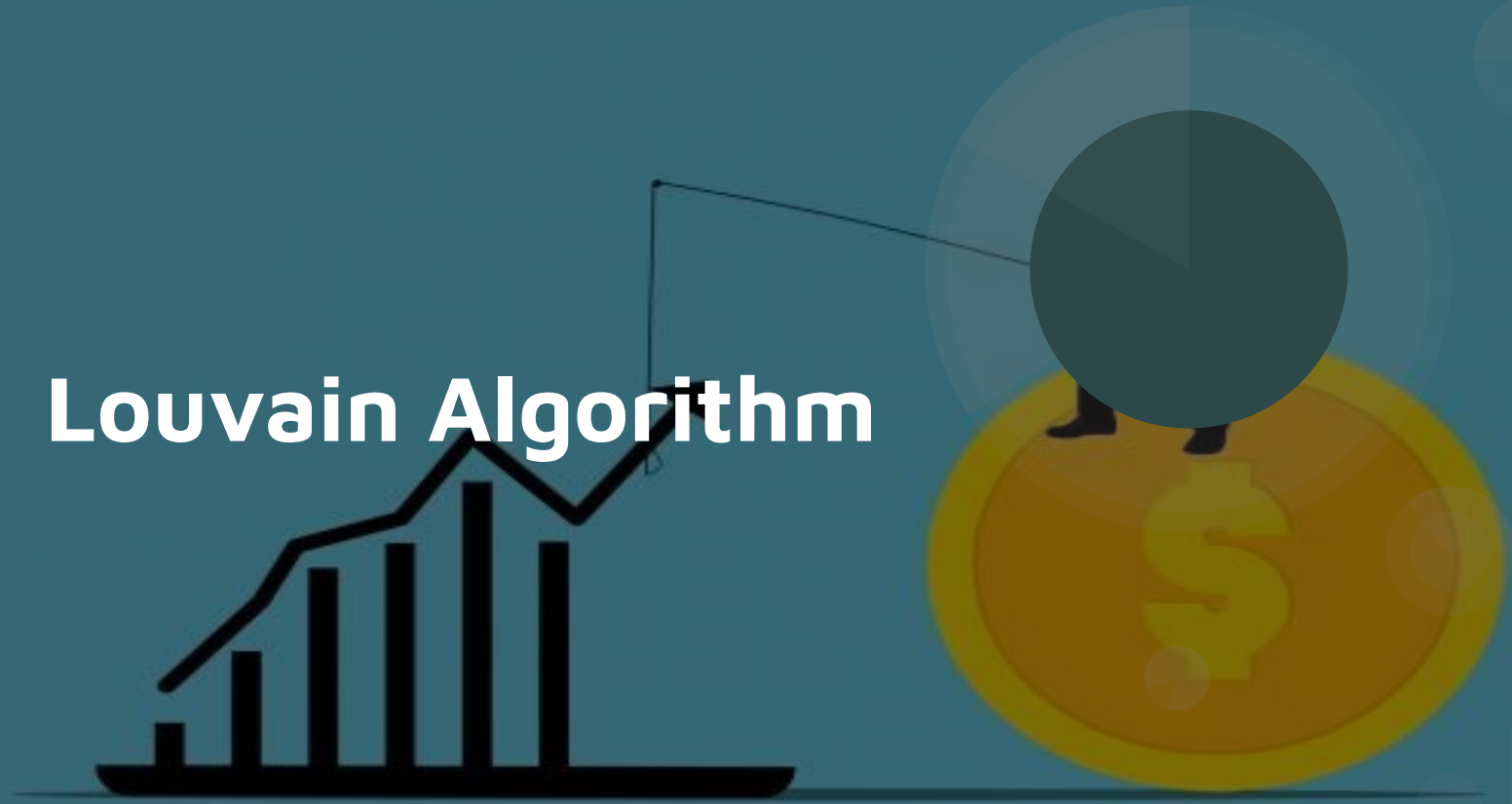
Story with this specific algo

- Helpful for quickly identifying clusters of relationships for further analysis, not so helpful for anything involving multiple variables and time

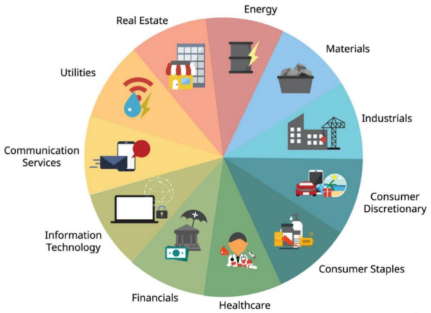
Usability nuances

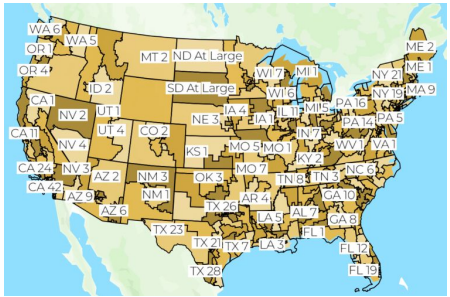
- Neo4J web browser has some customization limitations compared to other IDEs
- The ability to change node size and colors for multiple variables can be a nifty feature

Louvain Algorithm

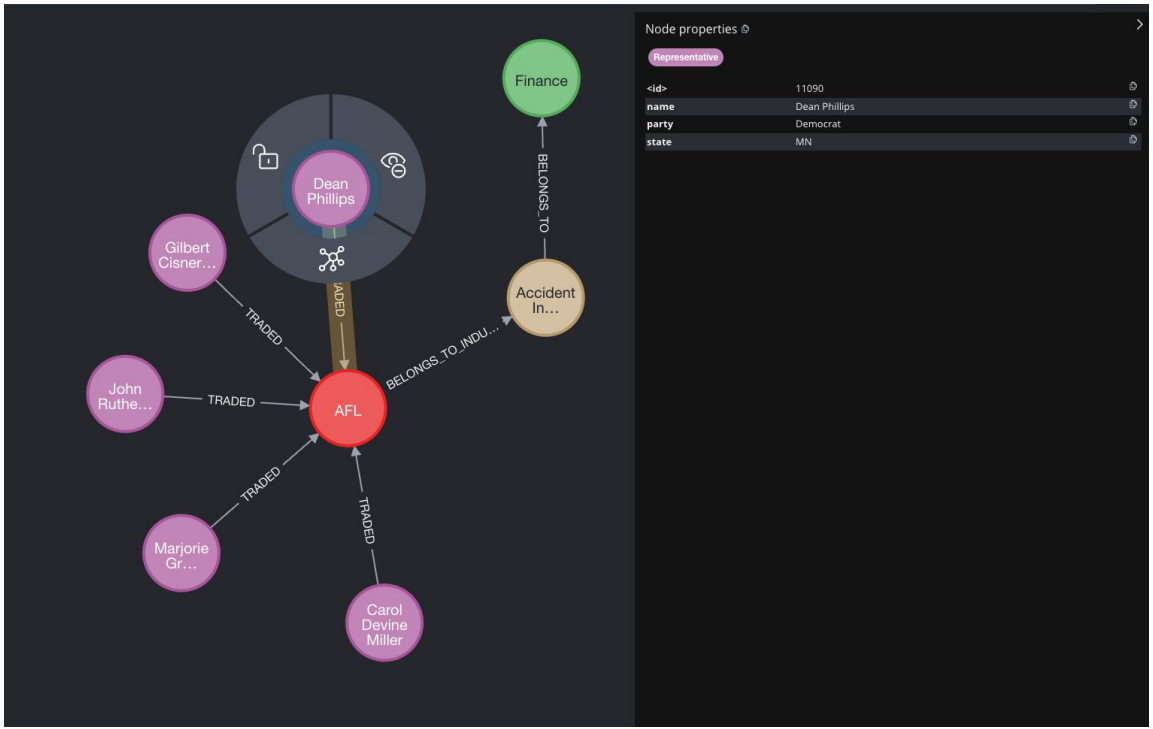


Application 2: Using Louvain to Cluster Transactions by industries and Parties

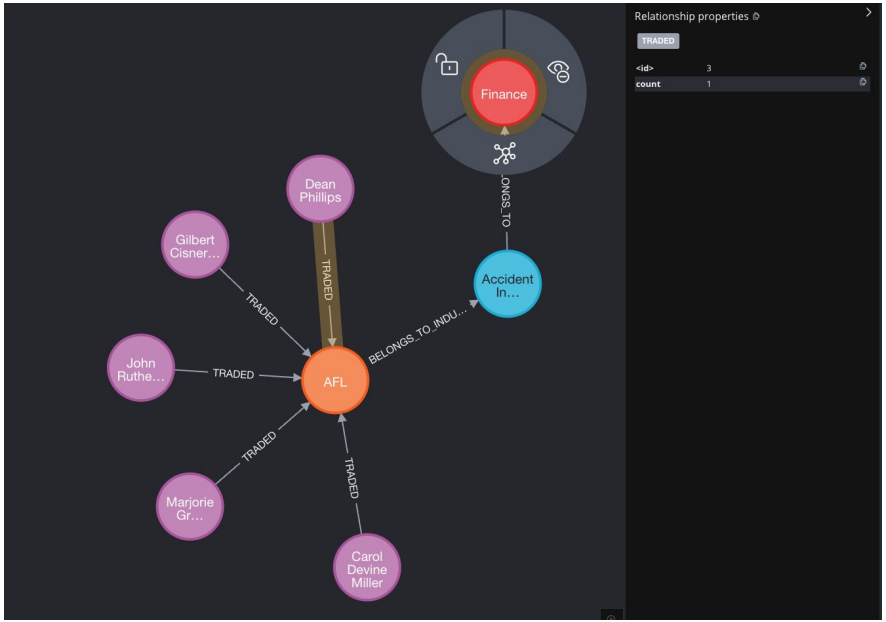




Graph Design - Politician as Nodes



Graph Design - [Traded] as edges, Industries, Sectors



- Politician
- [Traded]
- Company Ticker
- [Belongs to industry]
- Industry
- [Belongs to sector]
- Sector

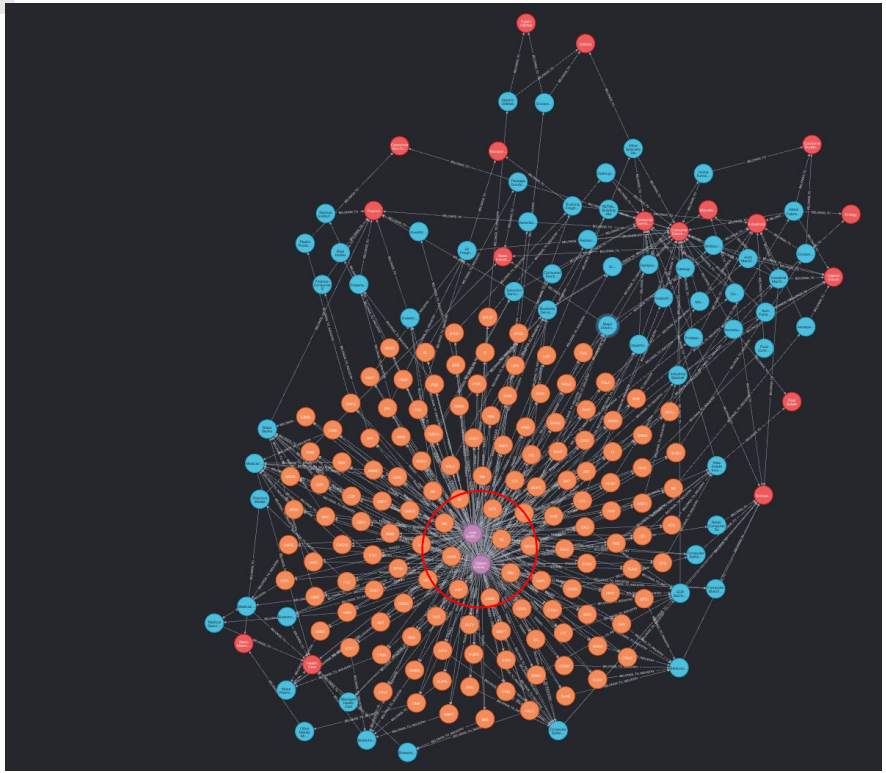
Louvain Modularity Overview

- One of the community detection algorithms
- It compares the relationship weights and densities to assess how well a node is assigned to a group
- Why Louvain?
  - Customize the weighting method.
  - Create groups

How Louvain is applied in this context

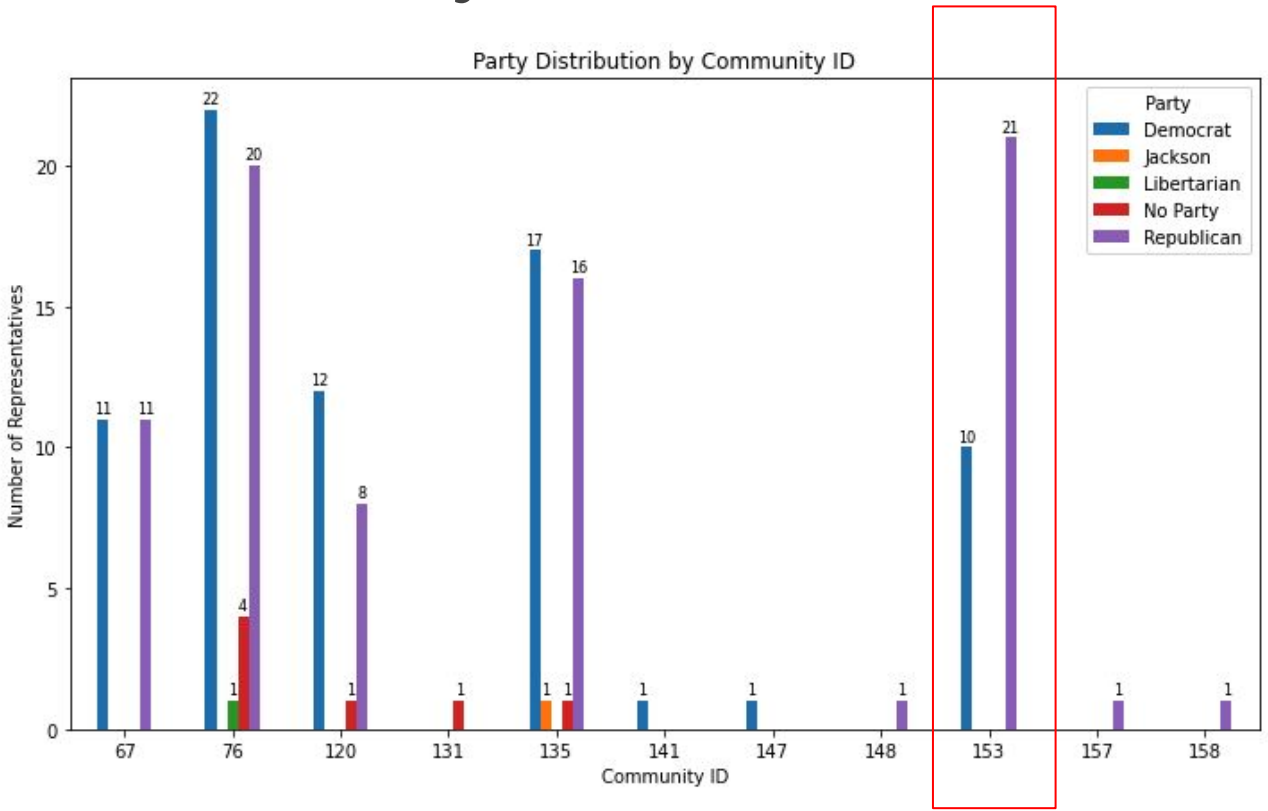
- Create **[similar_to]** Edges between representatives
  - Party and District Weighting
    - If same **party**, then similarity score +5
    - If same **state**, then similarity score +5
  - Industries and Sectors:
    - transaction count from [**traded**]
- Call Louvain stream: [similar_to] edges as the weight

Snapshot of two representatives

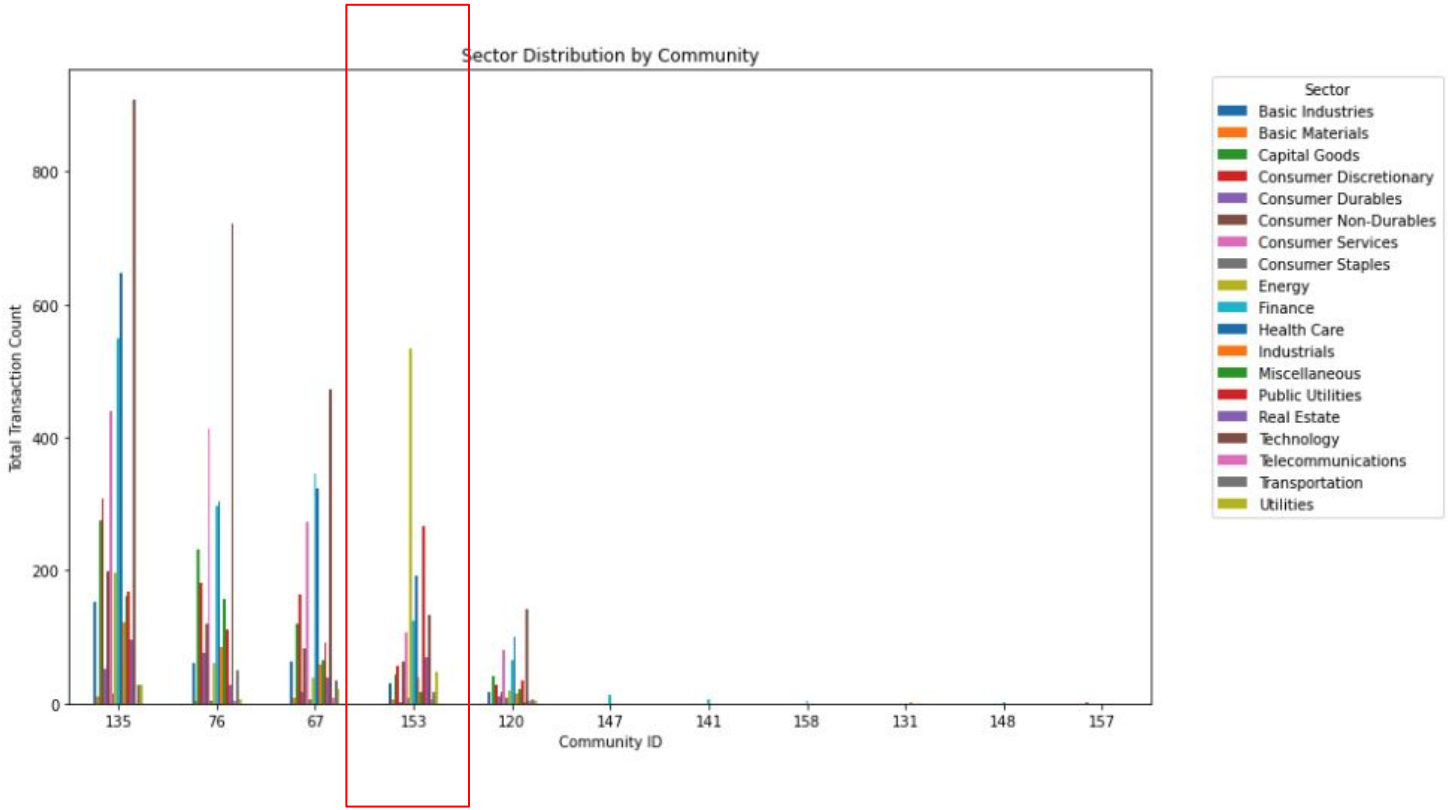


	rep1	rep2	similarity
0	Gilbert Cisneros	Josh Gottheimer	3462
1	Dean Phillips	Gilbert Cisneros	1996
2	Josh Gottheimer	Kathy Manning	1696
3	Josh Gottheimer	Kurt Schrader	1492
4	Josh Gottheimer	Peter Meijer	1322

Result - Party



Result - Sector Distribution by Community



Page Rank Algorithm

Why PageRank?

Why we chose it

- Useful for identifying influential individuals or stocks in trading networks

What we saw

- Congressional trading is dominated by tech, health care, and financial stocks.
- Cross-party participation in the trading of these stocks

Anomalies we observed

- Rank Sink is a potential issue if representatives trade heavily in stocks not connected to others

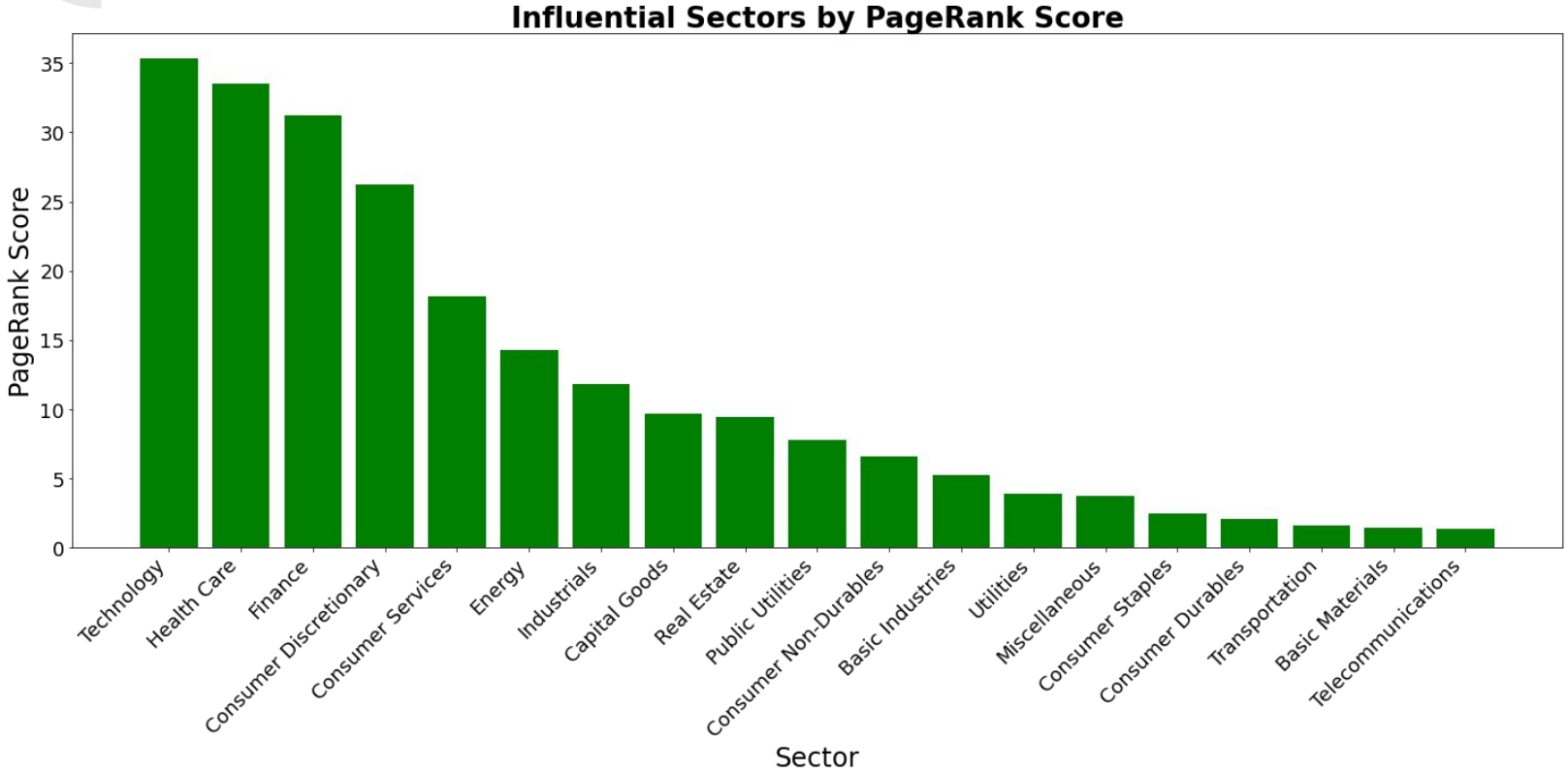
PageRank Overview

- Used to measure the relative importance of a node within a network
- Importance is based on the number and quality of the connections to other nodes
- Recursive importance helps identify indirect influence
- Anomalies can indicate outlier behavior or concentration of influence

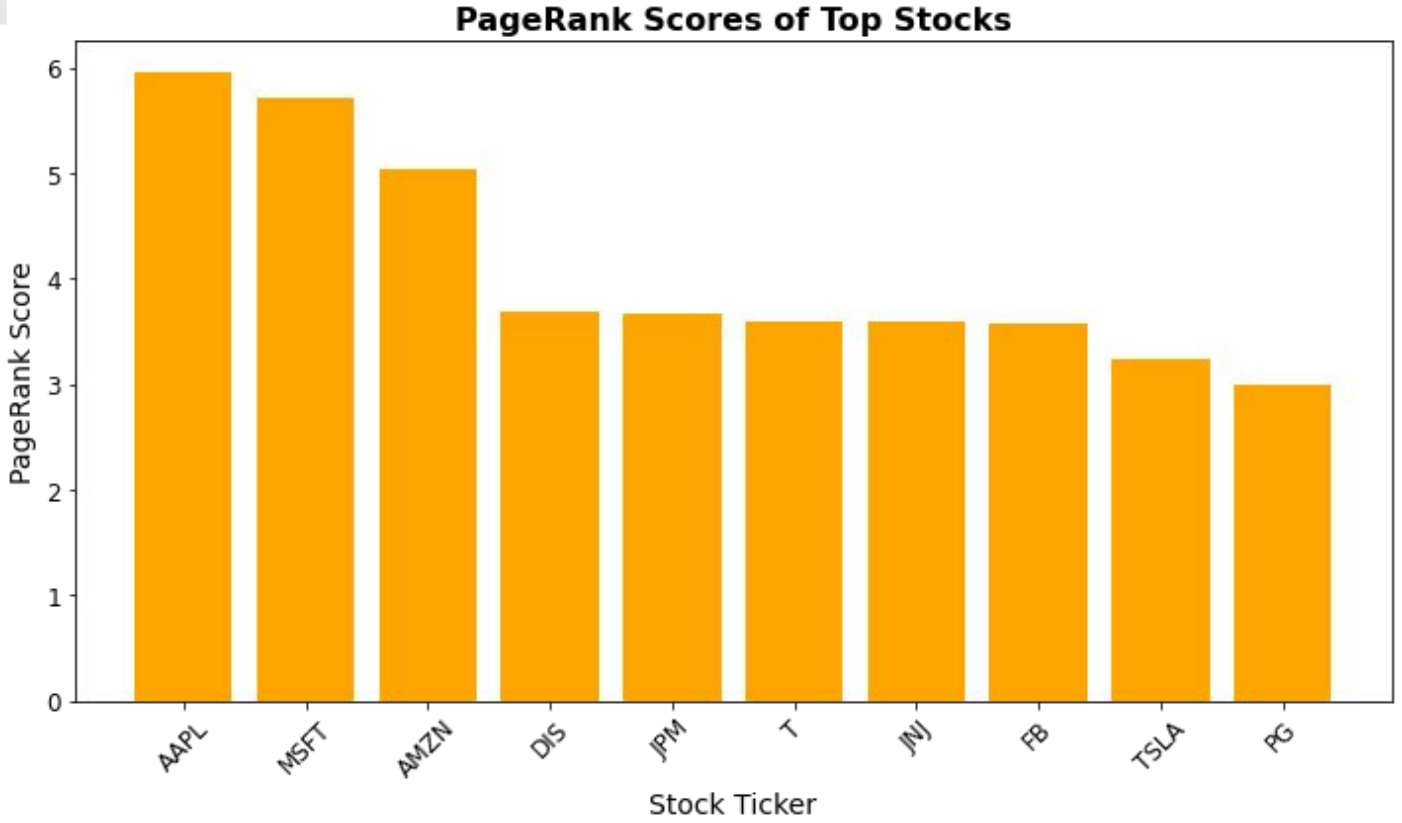
How PageRank is applied in this context

- Measure the most important stocks traded by representatives
  - Weighted by number and size of transactions
    - Analysis across all representatives
    - Analysis by party (Democrat vs Republican)
    - Analysis by stock
    - Analysis by sector

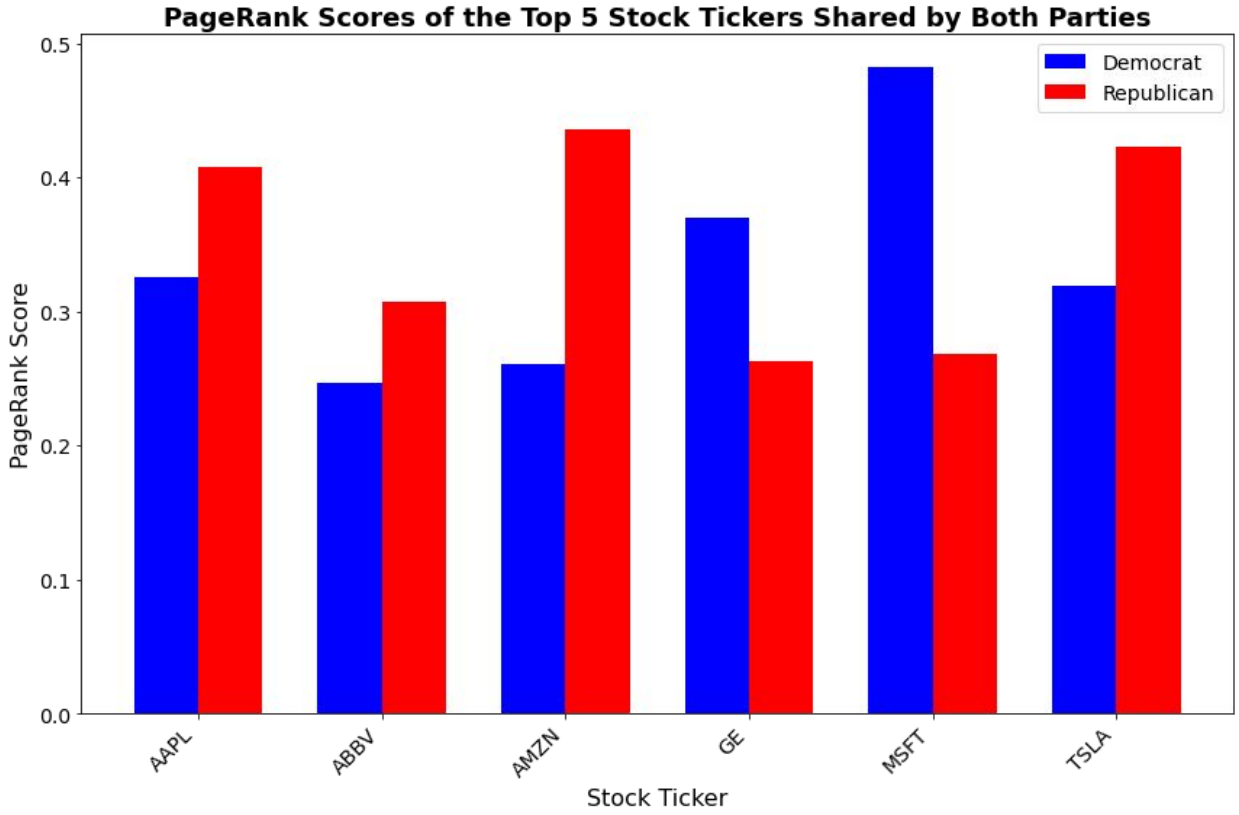
PageRank Results - Sector Influence



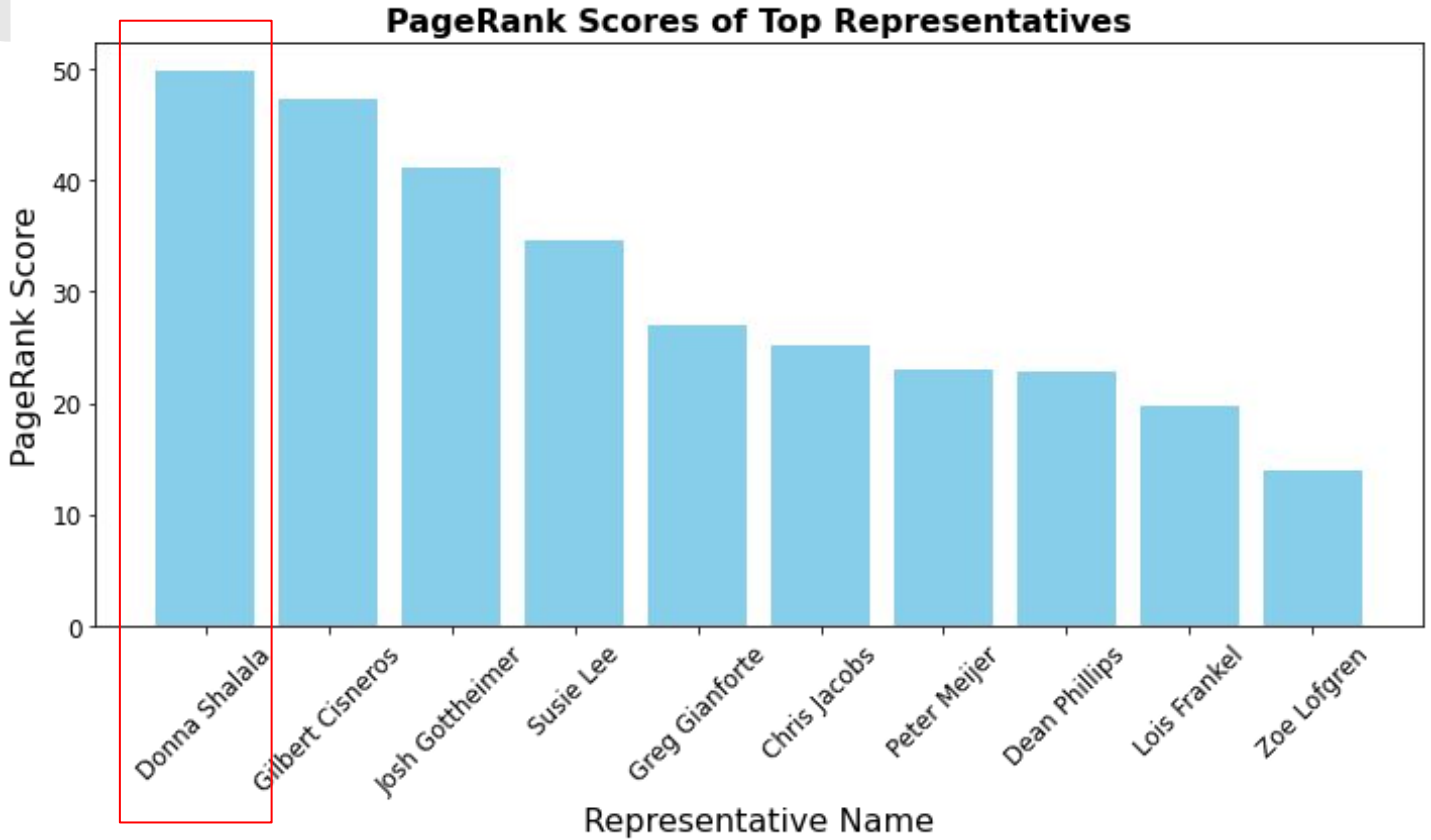
PageRank Results - Stock Influence



PageRank Results – Top 5 Stocks in Common



PageRank Results – Rank Sink Example

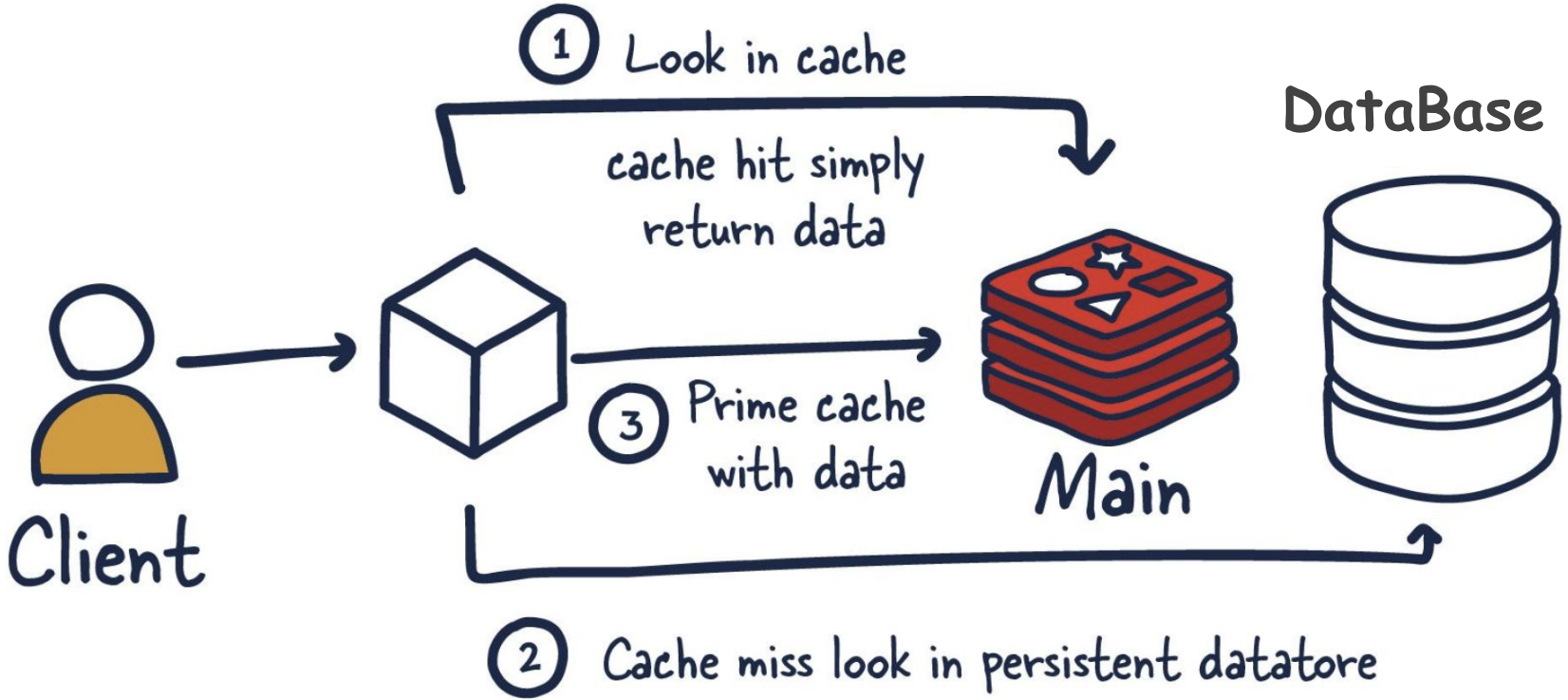


Why relational database is not a good fit?

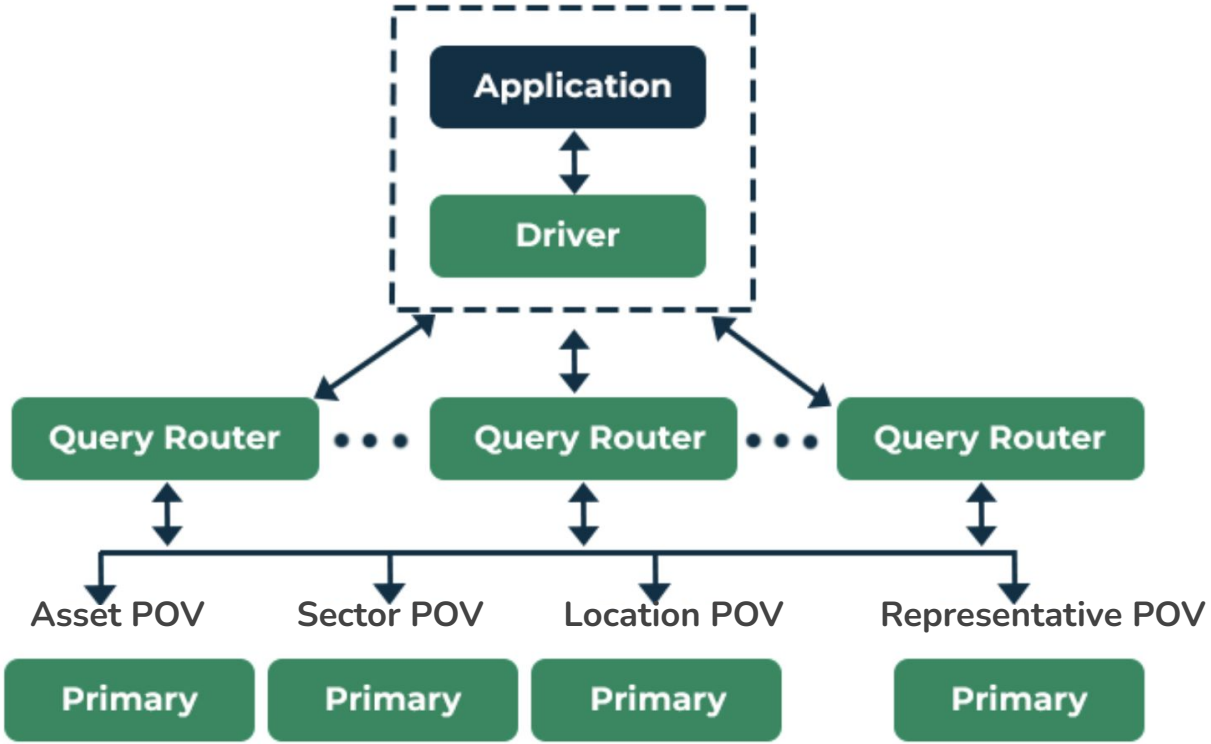
- Many relationships to analyze
- Built-in community detection algorithms in GDS
- Flexible Schema Change in Neo4J
- Easier visualization in Neo4J for relationships

Congress Trading Scenarios

Redis Scenario - Daily trade volume board



MongoDB Scenario



Thank You - QnA

Reference & Image Sources:

Data: https://housestockwatcher.com/api

Data Cleaning: https://thecustomizewindows.com/2021/02/what-does-data-cleansing-mean/

Industries: https://www.vecteezy.com/vector-art/10505835-a-stock-market-sector-is-a-group-of-stocks-that-have-a-lot-in-common-which-is-classify-by-the-global-industry-classification-standard-or-gics

Parties: https://wonderopolis.org/wonder/why-are-there-political-parties

Insider trading: https://www.insidermonkey.com/blog/inside-information-about-the-massive-insider-trading-probe-968/

Political Districts: https://www.govtrack.us/congress/members

Redis: https://architecturenotes.co/p/redis