



# Project 3 Presentation

## DATASCI 205 Summer 2025

Michael Strommer, Leo Lazzarini, Shikha Sharma

*Supporting Financial Asset Allocation with Graph Databases*

August 5, 2025

[GitHub Project3](#)



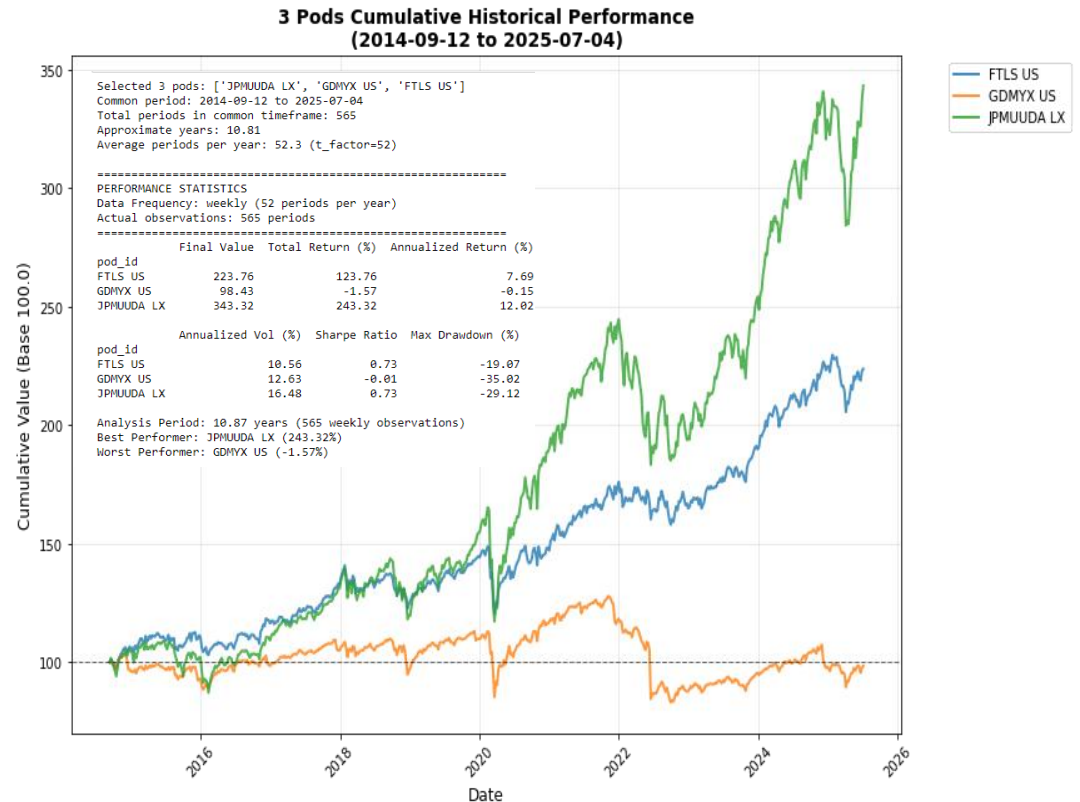
DATASCI 205 | Project 3

# Business Case



# The traditional approach

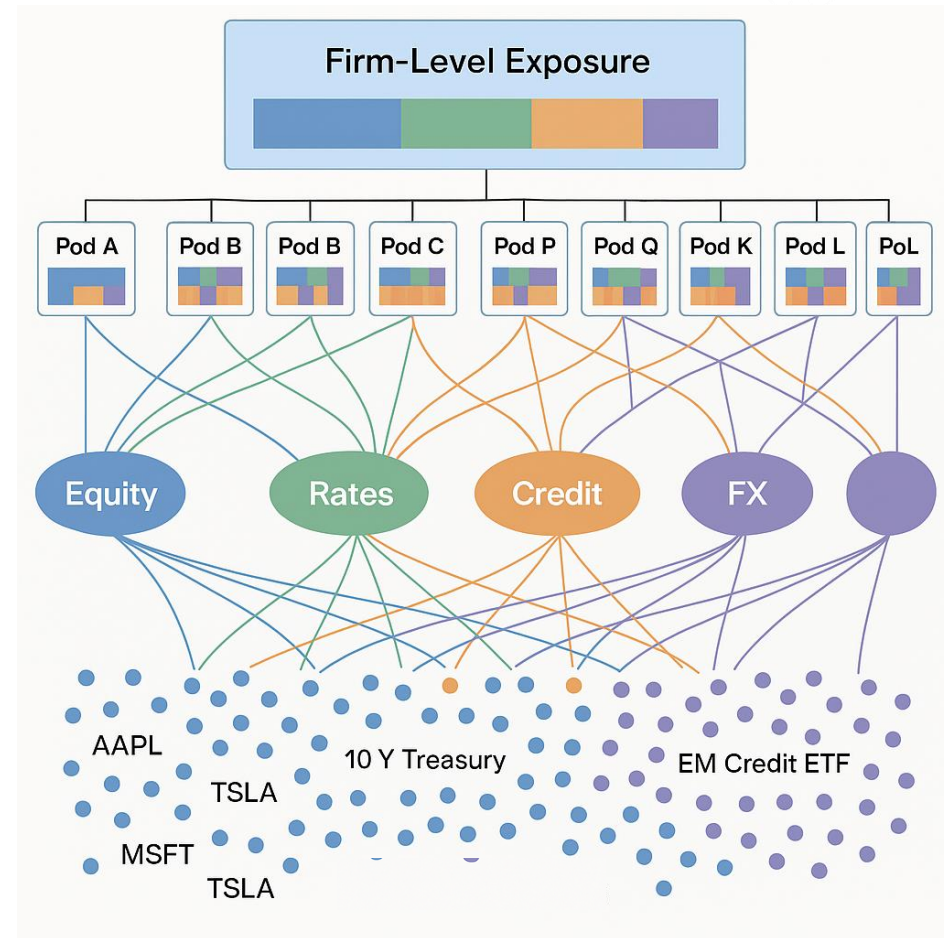
- Traditional pod selection relies on Sharpe ratios, performance trends, and correlations.
- These methods miss hidden common exposures and non-linear risk clusters.



Source: Bloomberg

# SML Hedge Fund: Challenge

- **Goal:** Smart capital allocation across investment pods
- **Problem:** Hidden overlapping exposure to Common Factors
- **Approach:** Graph network may reveal structure and diversification gaps

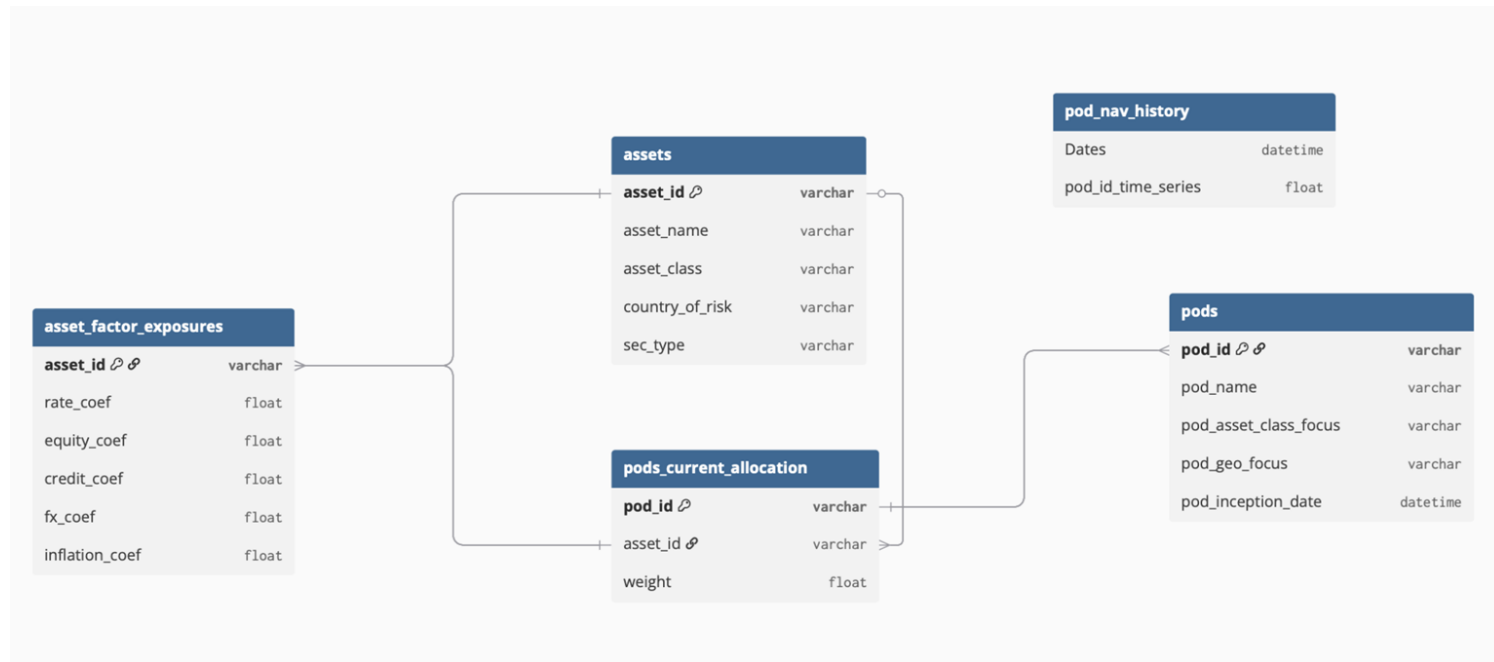




DATASCI 205 | Project 3

# Graph Design & Algorithms

# Graph Design 1: Data Pre Processing



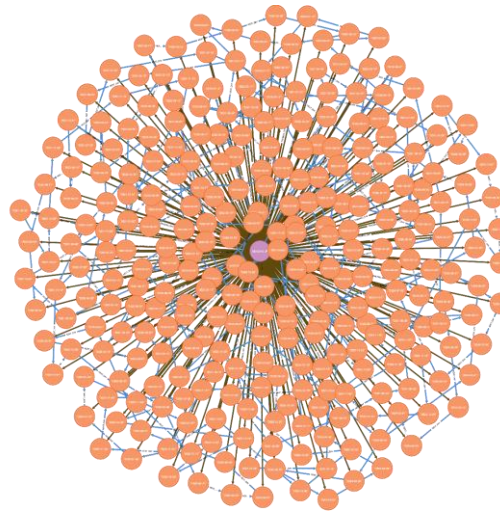
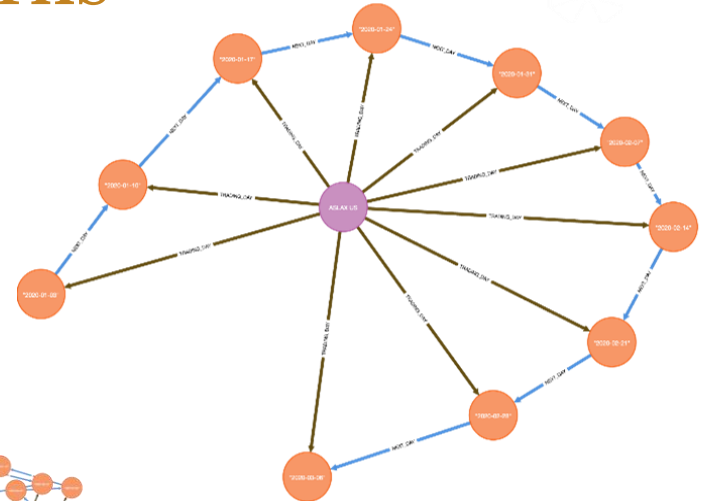
Steps: 1. Download .xlsx from Bloomberg 2. Create tables 3. Load data 4. Wrangle



# Graph Design 1: Investment Pod Weekly Returns

- Nodes:
  - Investment\_Pod
  - Pod\_Trading\_Day
- Relationship:
  - Trading\_Day
  - Next\_Day
- Labels:
  - Pod\_id,
  - date,
  - Weekly\_return
- Dates:
  - 1/1/2020 - 4/7/2025

ASLAX US  
AB SEL US LNG/SHRT-A



0431214C JP  
DAIWA JP EQ NEW GRWTH D SMA



# Graph Data Science Algorithms: Weekly Returns

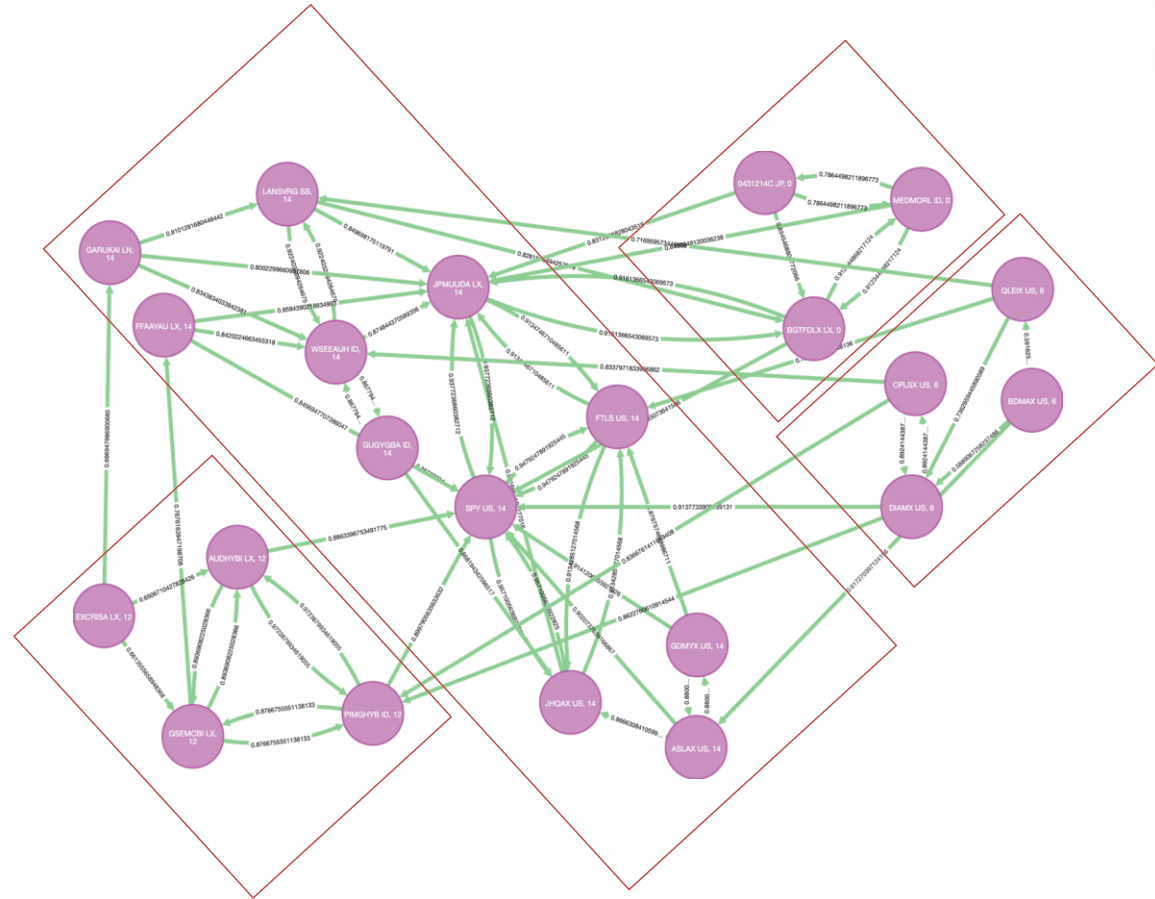


- **Pearson Similarity Correlation** Algorithm was used to identify similar pairs
- **Louvain Modularity** Algorithm was used to identify communities
- **Linear Regression** to identify pods that have similar weekly return slopes for the same dates

## Reason for not choosing Relational Database:

- Time Series Data – not suited for Relational Databases
- Cluster Analysis – better suited for Graph Databases

# Pearson plus Louvain Algorithm: Results



Business Case Conclusion: Diversify – don't invest in same cluster. Split investment among clusters.

# Graph Data Science Algorithms: Do the results make sense?

What is similar about these investment pods?  
(from internet search)

Community 0:

Long Term Equity  
Region: Global, US

Community 12:

High Yield Equity  
Region: Global and European

Community 14:

Long Term Total Return  
Region: Europe, US, Sweden

Community 18:

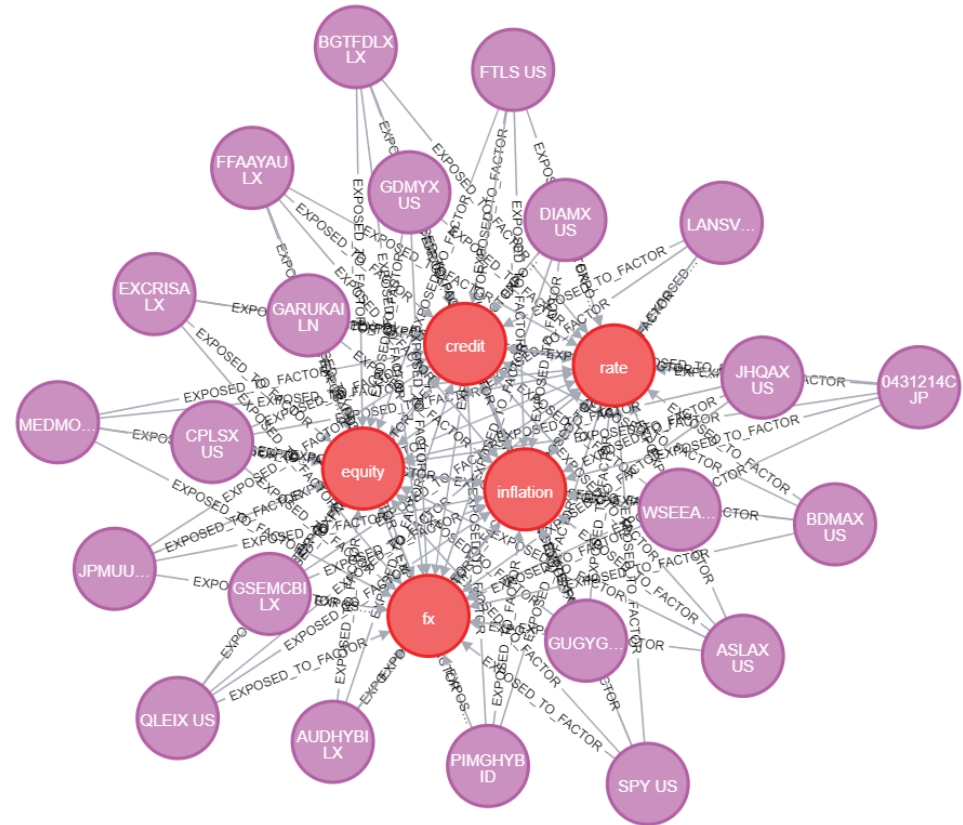
Long Short Equity  
Region: Global

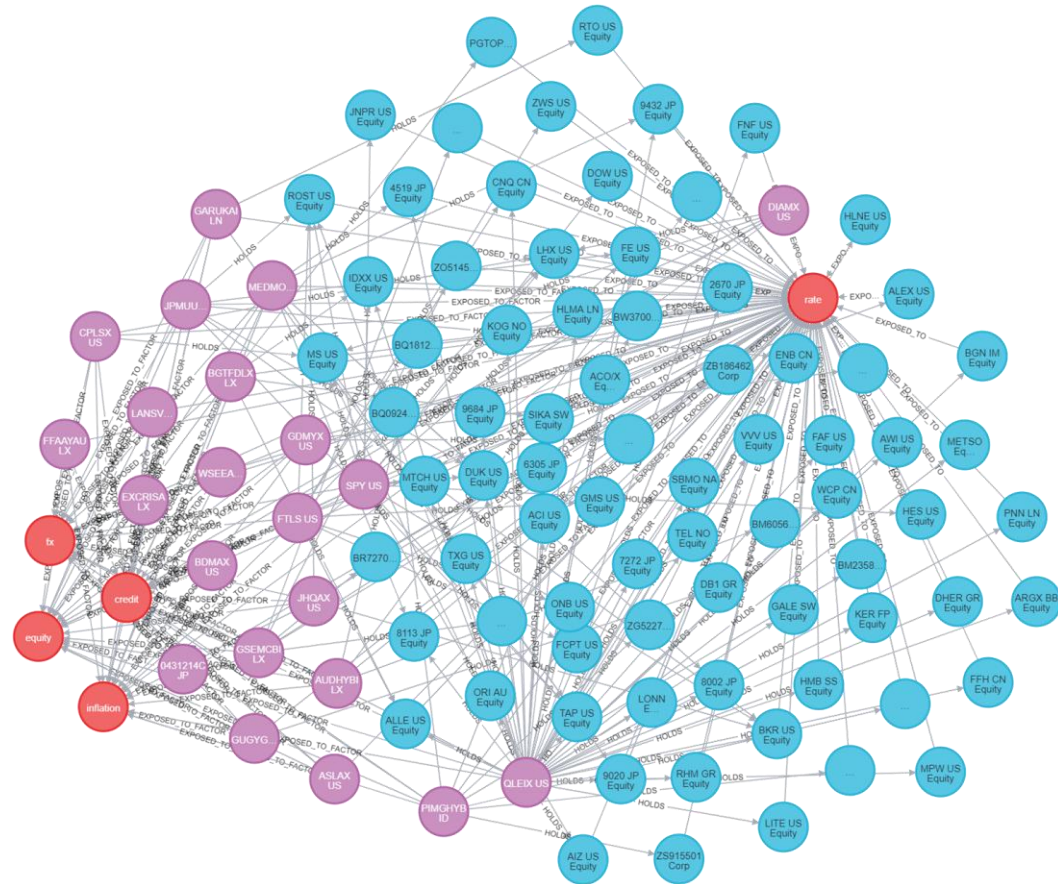
	POD ID	Community
0	0431214C JP	0
1	BGTFDLX LX	0
2	MEDMORL ID	0
3	AUDHYBI LX	12
4	EXCRISA LX	12
5	GSEMCBI LX	12
6	PIMGHYB ID	12
7	ASLAX US	14
8	FFAAYAU LX	14
9	FTLS US	14
10	GARUKAI LN	14
11	GDMYX US	14
12	GUGYGBA ID	14
13	JHQAX US	14
14	JPMUUDA LX	14
15	LANSVRG SS	14
16	SPY US	14
17	WSEEAUH ID	14
18	BDMAX US	18
19	CPLSX US	18
20	DIAMX US	18
21	QLEIX US	18

# Graph Design 2: Factor Exposure

Investment Pods have different degrees of exposure to **Factors**

All of them have some exposure to the factors at the center



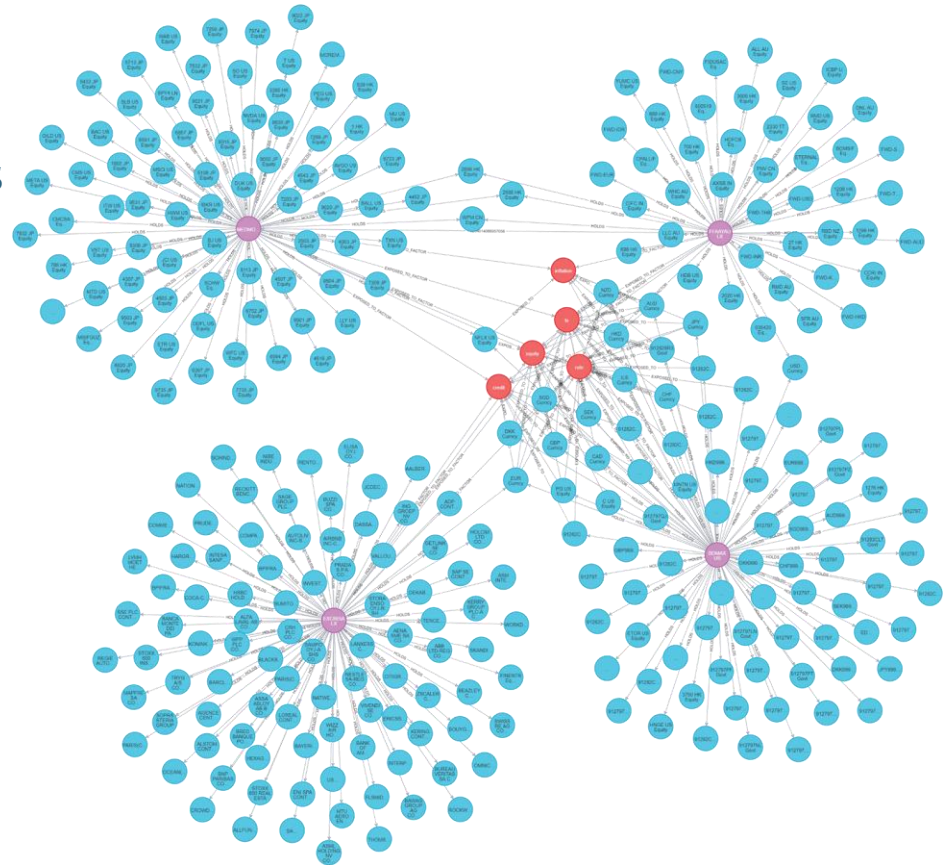


# Graph Design 2: Louvain Algorithm

4 detected investment communities  
Each investment pod shown represents  
a different community

## Inv. Pods:

- Mediolanum Morgan Stanley Global Equities
- Fidelity Asia Pacific: Equities Fund
- Blackrock Global Long / Short Fund
- Exane Fund Europe Long / Short



# Technology Review



**Graph**



**Document**

- Large, stored datasets
- Denormalized data
- Flexible Schema
- Great analytics



**Key-Value**

- In-memory
- Unique key
- Very fast, real time
- Similar to Python Dictionaries



# MongoDB in our Business Case

- **Store and index:** economic policy official speeches and transcripts for fast retrieval and NLP analysis (e.g., tone, topic modeling, sentiment).
- **Ingest and archive** audio files with metadata for audio-to-text processing pipelines.
- **Organize financial statements**, earnings transcripts, and press releases across all portfolio companies in a flexible format.
- **Maintain records** of regulatory filings (10-K, 8-K, etc.) linked to ticker, sector, and pod.
- **Future:** Use MongoDB to feed LLM-based internal search tools for research analysts and PMs to query any unstructured dataset.

# Redis in our Business Case

## Real-Time Signal Processing

- **Track** live market signals (e.g., rates, spreads, vol, liquidity metrics) and map them to portfolio or pod exposures using streaming speed.
- **Maintain fast-access** factor exposure matrices that update intraday and are referenced by live dashboards or trading systems.
- **Power** event-driven triggers (e.g., spread widening > threshold → notify trader or execute hedge).
- **Store real-time stop-loss metrics**, like rolling betas or VaR updates, for each pod or firm-wide.
- **Future:** Use Redis to broadcast market **regime shifts** or **macro alert** events to downstream algo engines or trader UIs.

# Final Conclusion

- **Graph Algorithms** worth further exploration
- Further analyses and back tests needed
- In the future
  - Add **Sentiment Analysis** (MongoDB)
  - **Link risk exposure** to market signals (Redis)



## DATASCI 205 | Project 3

# References

- [Diversify Your Stock Portfolio with Graph Analytics](#)
- [GitHub Project3](#)