

The background is a dark navy blue. In the top-left corner, there are two overlapping geometric shapes: a blue parallelogram and a light green parallelogram. In the bottom-left corner, there is a circular inset showing a close-up of a green printed circuit board (PCB) with various electronic components. In the top-right corner, there is a faint, light gray pattern of concentric, stepped lines resembling a topographical map or a stylized circuit.

# The Influence of Tweets on the Stock Market Price

By: Tan Dai Ngo, Naomi Phan, Dharnesh  
Somasundaram, Yashwanthkrishna Nagharaj



## Sources of Our Data

- **Tweets:** X (<https://api.x.com/2>)
- 2024/01/01 - 2025/11/20
- **Influencers:** "elonmusk",  
"realDonaldTrump",  
"CathieDWood", "jimcramer",  
"michaeljburry", "RayDalio"
- **Stock Prices:** Yahoo Finance  
(python library: yfinance)
- **Top 10 Tech Stocks:**  
NVIDIA, Apple, Microsoft,  
Alphabet, Amazon, Broadcom,  
Meta Platforms, Taiwan  
Semiconductor, Tesla, Tencent

## Data Collection Challenges

- **1000 results per request** (every 15 minutes)

- Must keep track of the **next pagination token**
- Only **861 out of 15,194 (5.67%)** tweets are related to the top 10 tech stocks from the 6 influencers

### MONTHLY POST CAP USAGE ⓘ



15,194 Posts pulled of 15,000

Resets on December 15 at 00:00 UTC

**Top up 10,000 Posts for \$100**

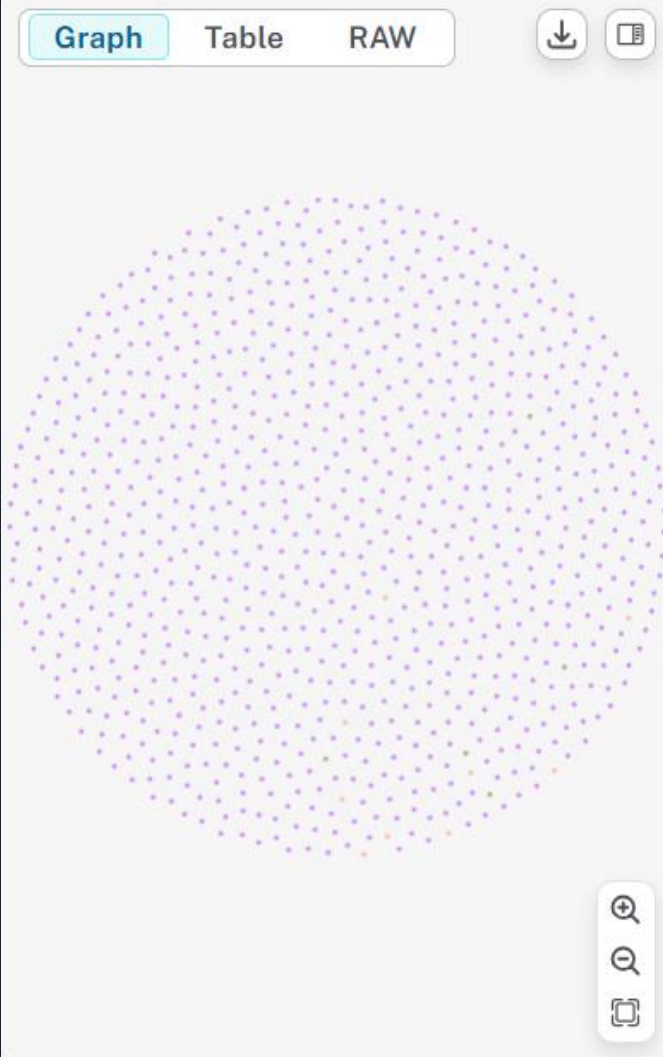
*Increase your Post cap limit by 10,000 Posts!*



# Knowledge Graph Structure

```
(:Person {id, username, name}  
  -[:POSTED]-> (:Tweet {id, text,  
                    created_at,  
                    like_count,  
                    retweet_count,  
                    reply_count,  
                    quote_count,  
                    lang})
```

```
(:Tweet)  
  -[:MENTIONS]->
```



## Results overview

Nodes (875)



**\* (875)**

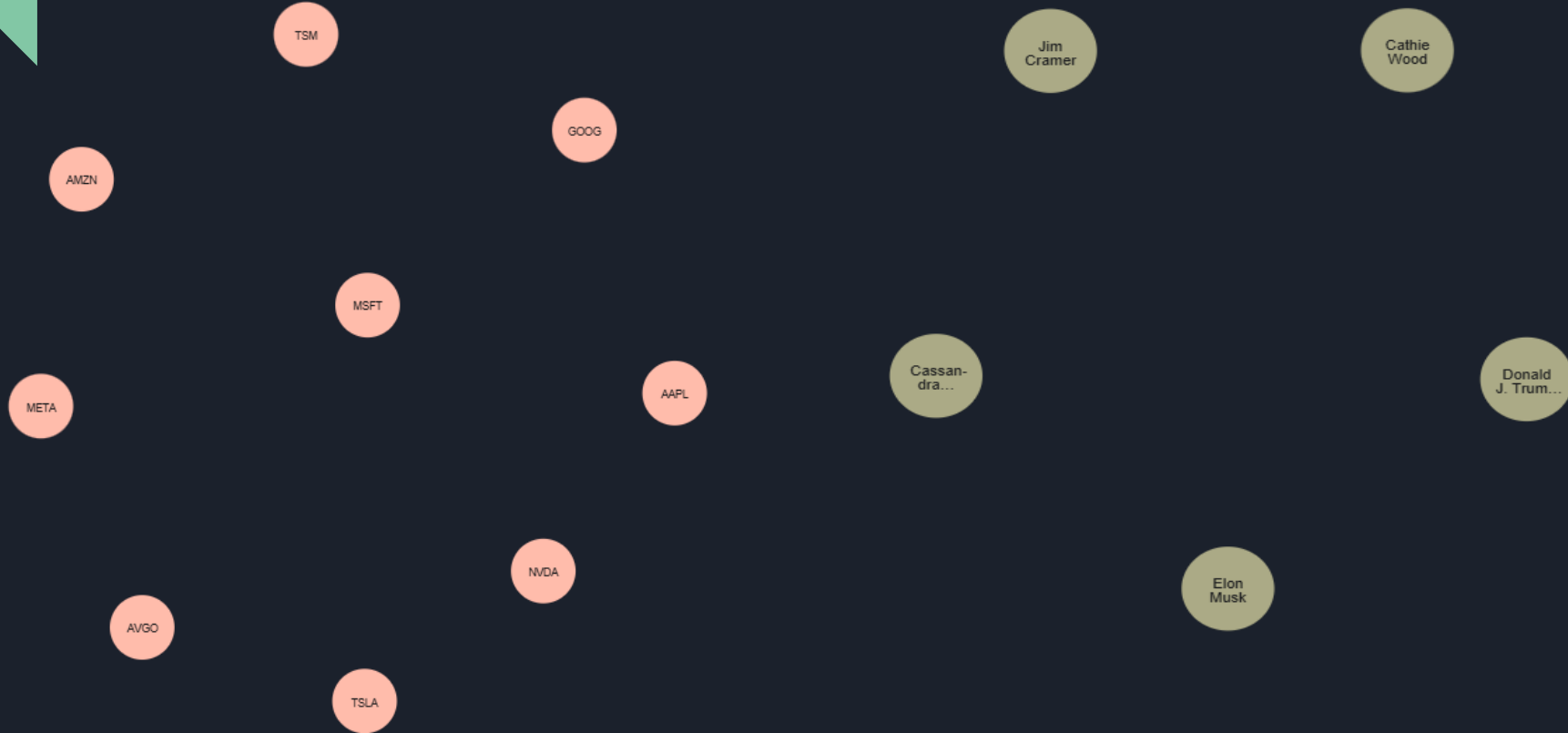
**Person (5)**


**Stock (9)**

**Tweet (861)**



# Stocks & Influencer Graphs





# Stock Influencer Tweets Database

## Purpose

- Stores all influencer tweets related to tech stocks
- Provides social sentiment + tweet frequency for each stock

## Key Tables

- **users** → influencer details (id, username, name)
- **tweets** → tweet text, timestamp, engagement metrics
- **stocks** → list of 10 tracked tech stock symbols
- **tweet\_stocks** → links tweets to mentioned stocks (many-to-many)

## Why It's Important

- Enables daily sentiment scoring using FinBERT
- Tracks tweet volume over time
- Allows correlation between influencer activity & stock price movement



# Stock Prices Databases (Combined)

The main database containing all historical stock price data

Includes daily OHLC values (open, high, low, close, adj close, volume)

Covers 10 major tech companies

Serves as the source for creating both:

- **StocksTraining.db** (2024 data)
- **StocksTesting.db** (2025 data)

Used for computing returns, volatility, and merging with sentiment.



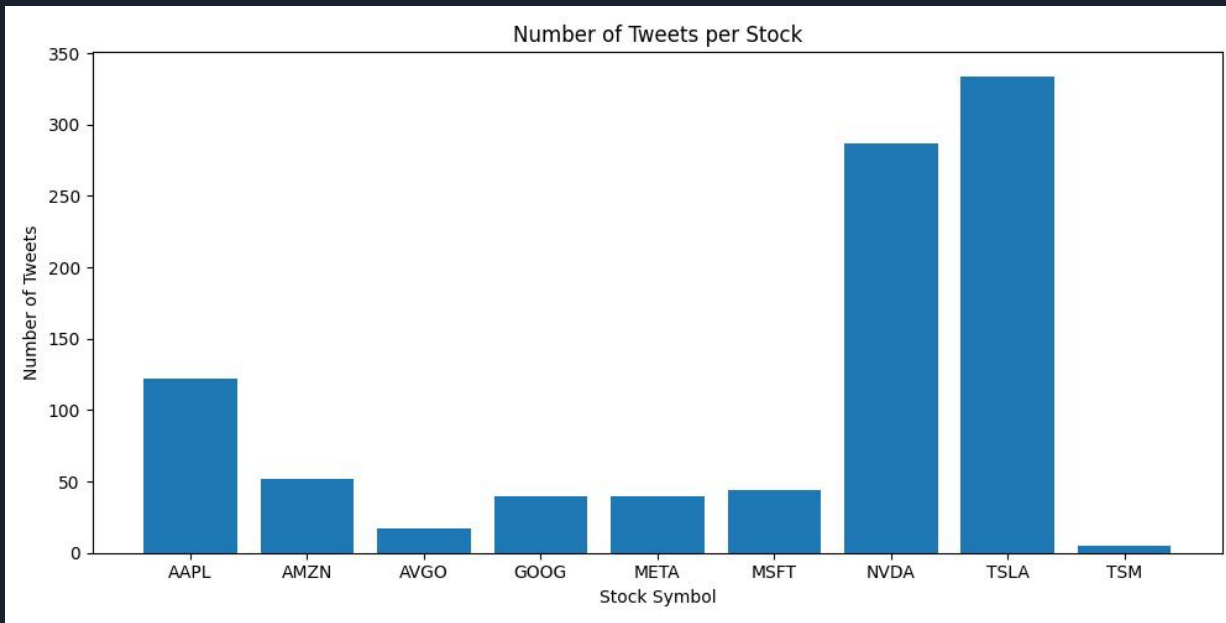
# How All Databases Work Together

**Tweets DB** → gives

- influencer tweets
- daily sentiment score
- tweet frequency per stock

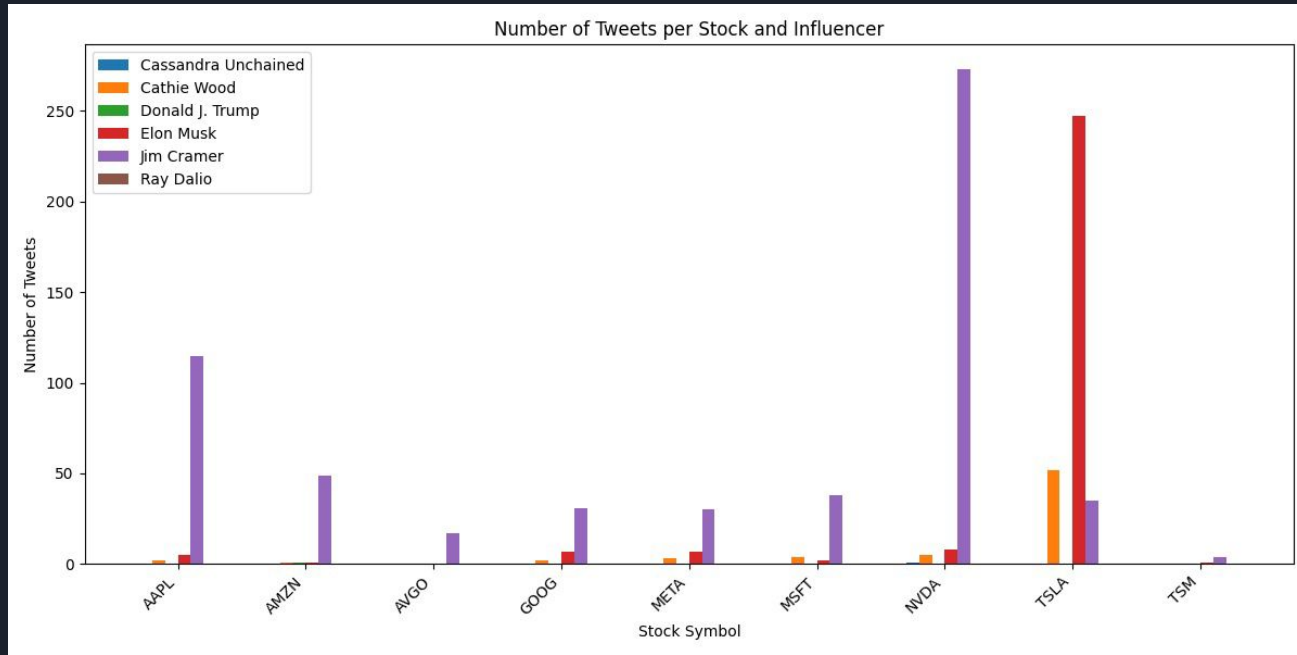
**Stock Prices DB** → gives

- daily returns
- price change percentages



### Number of Tweets per Stock

TSLA and NVDA have the highest tweet counts, AAPL is moderate, and the rest show low engagement.



### Tweets per Stock and Influencer

Jim Cramer tweets the most across all stocks, while Elon Musk focuses almost entirely on TSLA; others show minimal activity.

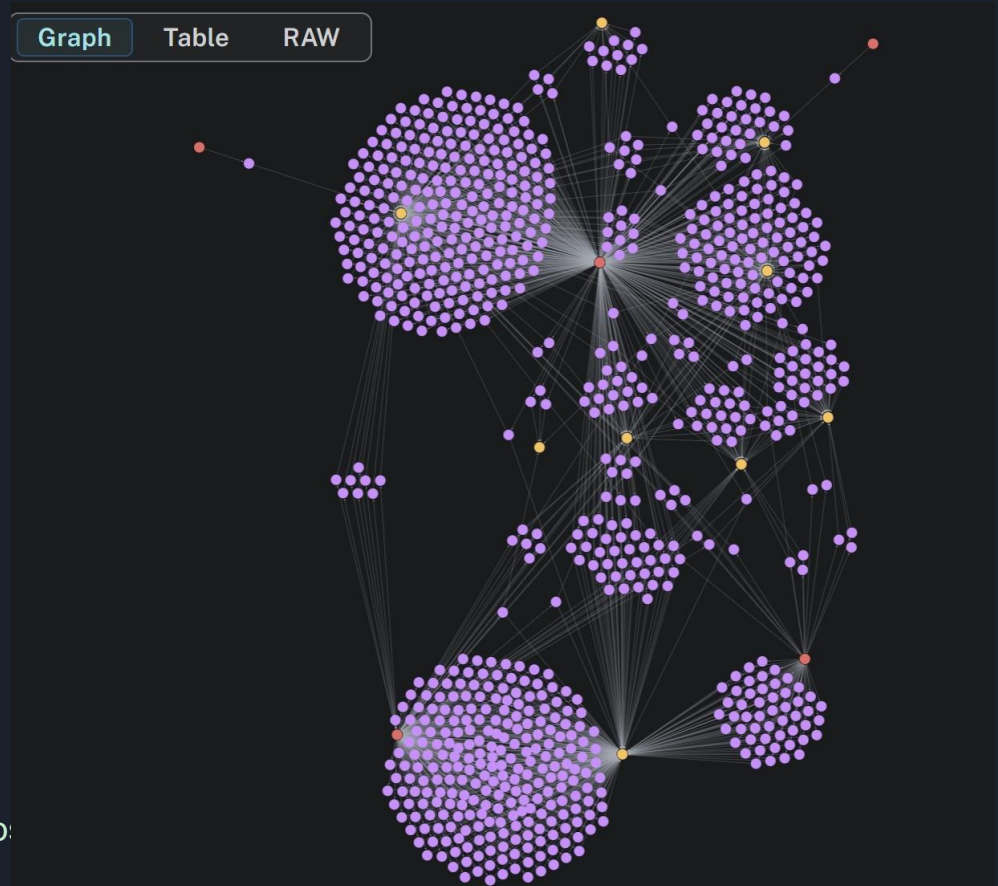


```
MATCH (p:Person) - [post:POSTED] ->
(t:Tweet) - [m:MENTIONS] -> (s:Stock)
```

```
RETURN p, post, t, m;
```

## Full Graph

- **Red nodes:** Influencers (Person)
- **Purple nodes:** Tweets
- **Yellow nodes:** Stocks
- Shows **complete network**
- Includes **POSTED + MENTIONS** relationship.

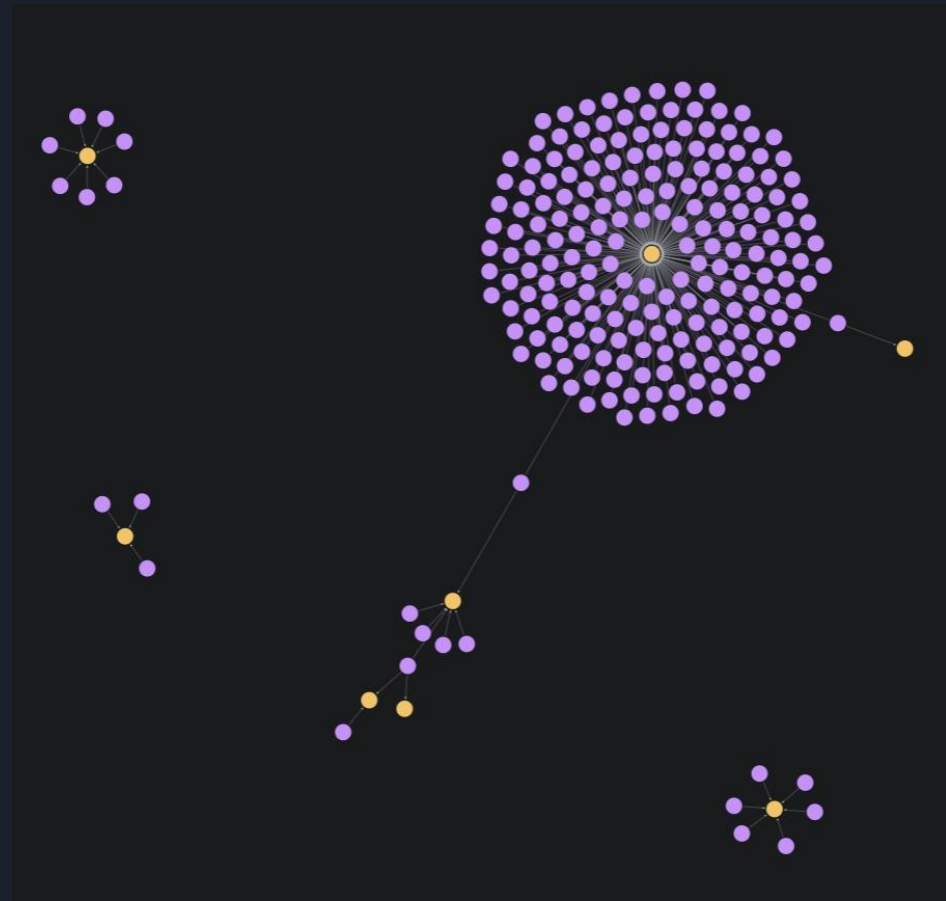




```
MATCH (t:Tweet) - [m:MENTIONS] -> (s:Stock)
WHERE t.retweet_count > 500
RETURN t, m, s;
```

### Retweet count

- Only **purple tweet nodes** with **> 500 retweets**
- Connected **yellow stock nodes**
- Highlights **viral tweets**

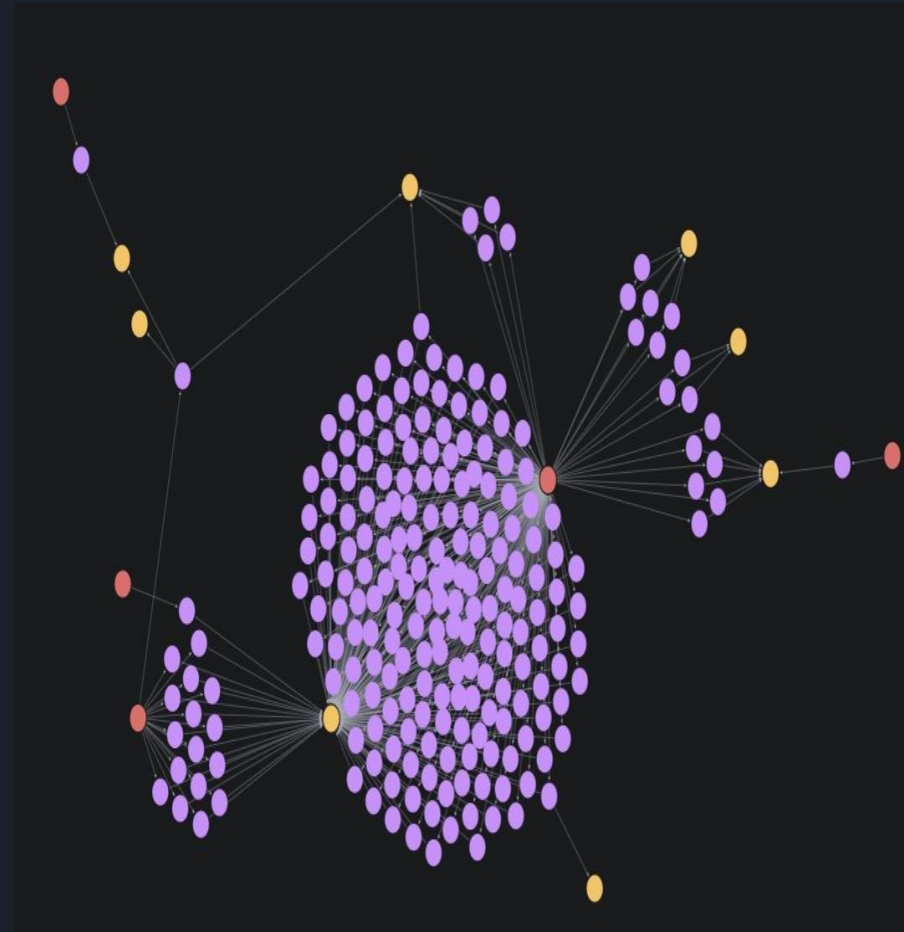


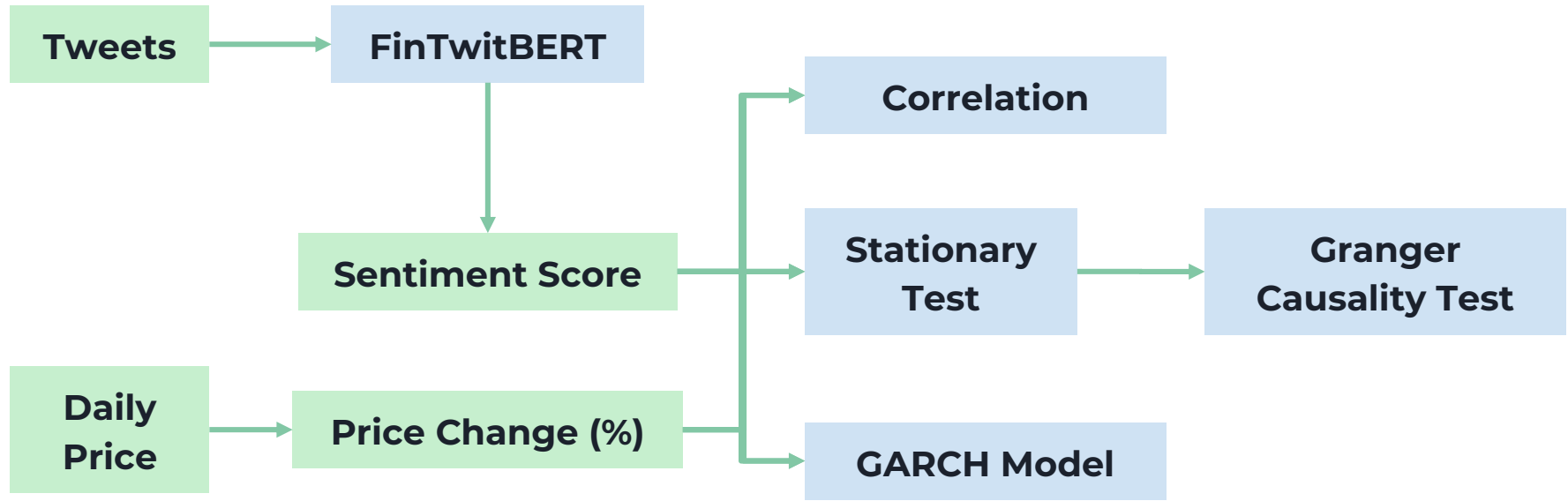


```
MATCH (p:Person) - [post:POSTED] ->
(t:Tweet) - [m:MENTIONS] -> (s:Stock)
WHERE t.retweet_count > 500
RETURN p, post, t, m, s;
```

## Person Retweet Count

- **Red influencer nodes** responsible for high-retweet tweets
- **Purple tweets** with > 500 retweets
- **Yellow stocks** mentioned in those viral posts





## Tweets Sentiment and Stock Price Change Correlation

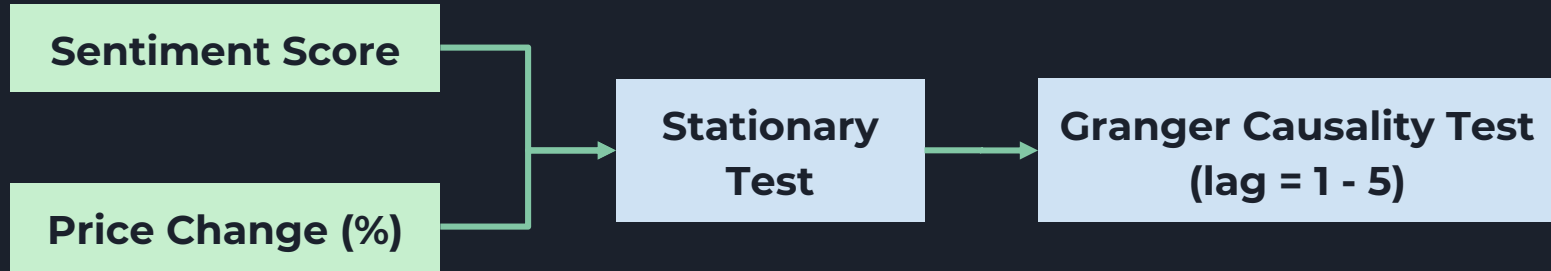
Ticker	lag_0	lag_1	lag_2	lag_3	lag_4	lag_5
AAPL	-0.03	0.05	-0.01	-0.01	-0.01	0.07
AMZN	-0.02	0.02	0.02	0.01	-0.01	0.04
AVGO	0.07	-0.07	0.03	-0.02	-0.04	0.00
GOOG	0.07	0.01	0.01	-0.02	0.00	-0.09
META	0.08	0.11	0.00	0.00	0.06	0.01
MSFT	0.01	0.03	0.02	0.00	0.06	0.07
NVDA	0.01	0.03	-0.01	0.05	0.02	0.05
TSLA	0.05	0.06	0.00	-0.02	0.00	0.01
TSM	-0.02	0.01	-0.03	-0.02	-0.05	0.05

Overall, the market  
tend to react in the  
**same day**

**META** has the  
highest correlation  
with tweet  
sentiment

# Granger Causality Test

Is tweet sentiment a helpful variable in predicting stock price change?



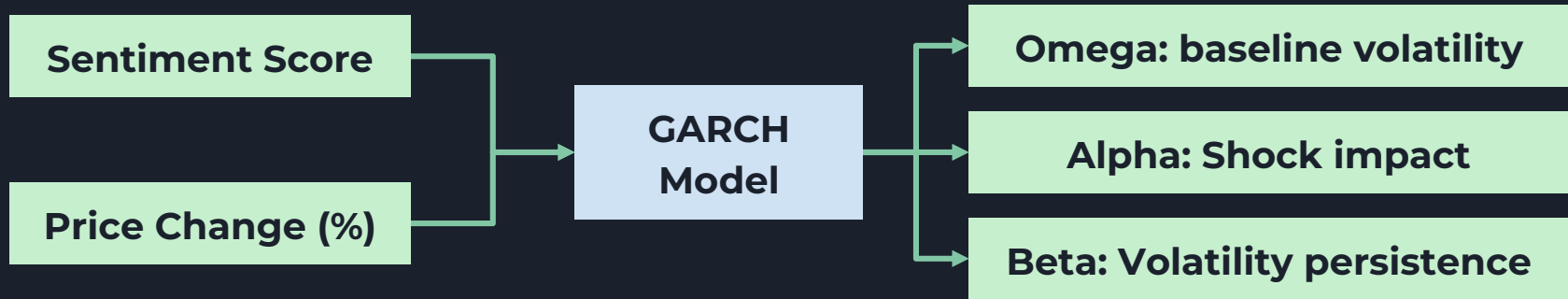
- **Null Hypothesis (H0):** Adding Sentiment Score **does not improve prediction accuracy** of Stock Price Changes
- **Alternative Hypothesis (H1):** Adding Sentiment Score **improves prediction accuracy** of Stock Price Changes

**Result:** Fail to reject the null hypothesis at all lags for all stocks

# GARCH Model

(Generalized Autoregressive Conditional Heteroskedasticity Model)

Does daily tweet sentiment affect stock volatility?



- How strongly markets react to sudden shocks?
- How long uncertainty persists?
- Whether sentiment contributes to volatility
- Differences between stocks' volatility patterns



# GARCH Model Result

	mu	omega	alpha	beta	interpretation
AAPL					Mostly stable during 2024-25
AMZN		1.59	0.59		AMZN reacts to news strongly, but quickly calms down afterwards
AVGO				0.59	High volatility stock
GOOG	0.17	2.21			Mostly stable during 2024-25
META				0.7	High volatility stock
MSFT				0.89	High volatility stock
NVDA	0.31			0.91	Most volatile
TSLA				0.96	Most volatile
TSM	0.2	4.18			Mostly stable during 2024-25



# Analysis Challenges

- 01** Small data size, short time frame
- 02** Data manipulation
- 03** Statistical/Assumption testings



“

Though we see sentiment shows **meaningful correlations** with short-term returns, but it is **not a reliable standalone predictor**.

Sentiment plays a more important role in **explaining market volatility than returns**.

”



Thank you for your  
attention!

