Financial Narrative Genome

Aidan LoStracco
University of Western Ontario
alostrac@uwo.ca

Sean McCorquodale
University of Western Ontario
smccorq@uwo.ca

Alyssa Spasic
University of Western Ontario
aspasic2@uwo.ca

Sudrisha Sarkar
University of Western Ontario
ssarkar32@uwo.ca

***Abstract*—Financial markets are driven by both quantitative data and the complex narratives that shape investor sentiment. This paper introduces the 'Financial Narrative Genome,' an AI-driven system that extracts and visualizes the thematic structures, causal relationships, and emotional tones embedded within financial texts. We address the challenge of capturing the dynamic evolution of market narratives by employing advanced Natural Language Processing (NLP) techniques, including transformer models, to analyze news, reports, and social media. The system constructs a network graph representing the 'narrative genome' and tracks emotional tone changes over time. We demonstrate the system's ability to identify key narrative shifts and correlate them with market fluctuations. This approach provides a more nuanced understanding of market dynamics, with potential applications in risk management and investment strategy. The project's code is available on GitHub.**

I. Introduction

The digital age has ushered in an unprecedented volume of unstructured textual data, transforming how information is disseminated and consumed within financial markets. While traditional quantitative analysis remains essential, the narratives embedded within news, social media, and corporate communications hold valuable insights into market sentiment and behavior. This paper presents the 'Financial Narrative Genome,' an AI-driven system designed to extract and visualize the complex interplay of themes, causal relationships, and emotional tones that constitute these narratives. By creating a structured representation of financial narratives, this research aims to lay the groundwork for a more nuanced understanding of market dynamics and potential predictive capabilities.

A. Motivation

The financial markets, while driven by quantifiable data, are fundamentally shaped by narratives. These narratives, encompassing news reports, social media discourse, and corporate communications, influence investor sentiment and ultimately drive market behavior. Traditional financial analysis, however, often overlooks the nuanced interplay of these narratives, relying primarily on quantitative metrics that fail to capture the dynamic and evolving nature of market sentiment. With the increasing volume and velocity of financial information, particularly in the age of social media, the need for narrative analysis tools has become paramount.

Recent advancements in Natural Language Processing (NLP) and machine learning have opened new avenues for understanding and interpreting complex textual data. Specifically, the development of transformer-based models [1] has enabled more nuanced sentiment analysis and information extraction from unstructured text [2]. However, while sentiment analysis provides a valuable measure of emotional tone, it often fails to capture the intricate web of causal relationships and thematic connections that constitute a financial narrative. Recent research has explored the extraction of financial narratives, and summarization of those narratives. However, the creation of a comprehensive 'Financial Narrative Genome,' which maps the evolution of these narratives over time and predicts their impact on market dynamics, remains an underexplored area.

Therefore, this research aims to address the challenge of extracting, representing, and predicting market shifts through deep narrative analysis. We introduce the concept of a 'Financial Narrative Genome,' a system that utilizes advanced NLP techniques to deconstruct and visualize the interconnected themes, causal relationships, and emotional arcs within financial narratives. This system then uses this data to produce predictions of market shifts. By developing a system that can accurately map and interpret the 'Financial Narrative Genome,' we aim to provide a more comprehensive and insightful approach to financial market analysis.

B. Related Works

Basic sentiment analysis has been widely explored due to its social value, leading to numerous approaches. Techniques include feature extraction (TF-IDF, PoS tagging, negation handling), feature selection (Chi-square, Mutual Information, Information Gain), and word embeddings (Word2Vec, GloVe, BERT, ELMo) to enhance contextual understanding [3]. More recently, large language models (LLMs) have advanced natural language processing (NLP) by capturing nuances like sarcasm and context shifts, leading to more accurate sentiment predictions. By leveraging LLMs, sentiment analysis can be enriched through multiple sub-tasks, providing deeper insights into emotional tone, causal inference, tone extraction

and entity relationships.

Semantic matching identifies relationships between words or phrases based on meaning rather than exact wording. This approach leverages knowledge bases, word embeddings, or transformer models to determine whether different expressions convey the same concept. By capturing context and synonymy, semantic matching ensures that variations in phrasing do not hinder understanding [4]. In the context of the four subtasks—causal inference, theme extraction, emotional tone identification, and entity relationships—semantic matching can unify varying outputs by linking related concepts. For example, if causal inference detects "supply chain disruptions affect revenue," while theme extraction identifies "logistics issues," semantic matching can recognize their connection. This enables a structured, interconnected representation of financial narratives, enhancing the clarity and utility of the final graph.

C. Problem Definition

The goal of this paper is to perform robust sentiment analysis on financial narratives and create an interconnected financial narrative genome graph. Traditional sentiment analysis struggles to capture nuanced emotions, context shifts, and the complex relationships in financial texts. To address this, we propose using LLMs for sentiment extraction, along with semantic matching to link related concepts across multiple sub-tasks (e.g., causal inference, theme extraction, emotional tone identification, and entity relationships).

The problem can be formally defined as: Given a set of financial texts T, the objective is to construct a graph $G = (V, E)$, where:

- V represents nodes such as sentiment, themes, and entities.
- E represents edges that connect related entities, subthemes, and causal effects.

This process involves the following steps:

1) Sentiment extraction using LLMs.
2) Semantic matching to link concepts across sub-tasks.
3) Graph construction to represent interconnected financial narratives.

By combining LLMs with semantic matching, we aim to enhance sentiment analysis and build a more structured, accurate representation of financial narratives for improved decision-making.

II. Methodology

This section outlines the methodology used in the development of our project, detailing the data presentation, proposed solutions, evaluation strategies, and additional analyses performed to optimize our solution. Our approach follows a systematic design process, focusing on emotional tone extraction and narrative evolution.

A. Data Presentation

The data used in this project was sourced from three main sources:

- *Stock Price Data:* Stock prices of relevant companies were retrieved from Yahoo Finance, spanning the period from 2010 to the present. This data provides insights into the fluctuations in stock prices, which are influenced by financial narratives and market events.
- *Financial News Data:* Financial news articles were obtained using the GNews API and the newspaper3k library. These articles provide the narrative context for the stock price data, allowing us to explore how different types of financial news impact market sentiment.
- *Annual Letter to Shareholder Data:* Annual shareholder letters from companies' financial reports were collected to capture corporate messaging, strategic priorities, and evolving themes. These documents offer valuable insights into how companies communicate financial performance, challenges, and future outlooks to investors.

The processed data was stored in two primary files:

- *(tickername)_graph.csv*: Contains the emotional tone scores of articles, averaged by year.
- *(tickername)_news.csv*: Stores the associated financial news articles, along with metadata such as the publication date and source.

B. Solution

Our solution focuses on two key aspects: emotional tone analysis and tracking the evolution of financial narratives.

1) Emotional Tone Extraction: The first step in our solution was to extract emotional tones from the news articles using the Gemini API. This API provides a robust analysis of the emotional sentiment behind each article, allowing us to track how emotions like optimism, anxiety, or sadness evolve over time. We averaged the emotional tone scores on a yearly basis to better align the sentiment data with stock price fluctuations.

2) Narrative Evolution Tracking: In addition to emotional tone extraction, we constructed a network graph to track the evolution of financial narratives. Each theme and entity was represented as a node, and each causal relationship was represented as an edge. The relationships were linked to nodes through semantic matching. The nodes were embedded and a sentence transformer model mapped the relationships to the nodes with the high semantic similarity. This network structure helped us visualize how financial narratives evolve and relate to one another over time, offering insights into how these narratives shape market behavior.

3) Narrative Summary: The Gemini API synthesized extracted themes, entities, relationships, and emotions to produce a comprehensive 15-year company narrative. This summary revealed deeper historical insights than a basic overview.

C. Evaluation of Solution

To evaluate the effectiveness of our proposed solutions, we employed the following methods:

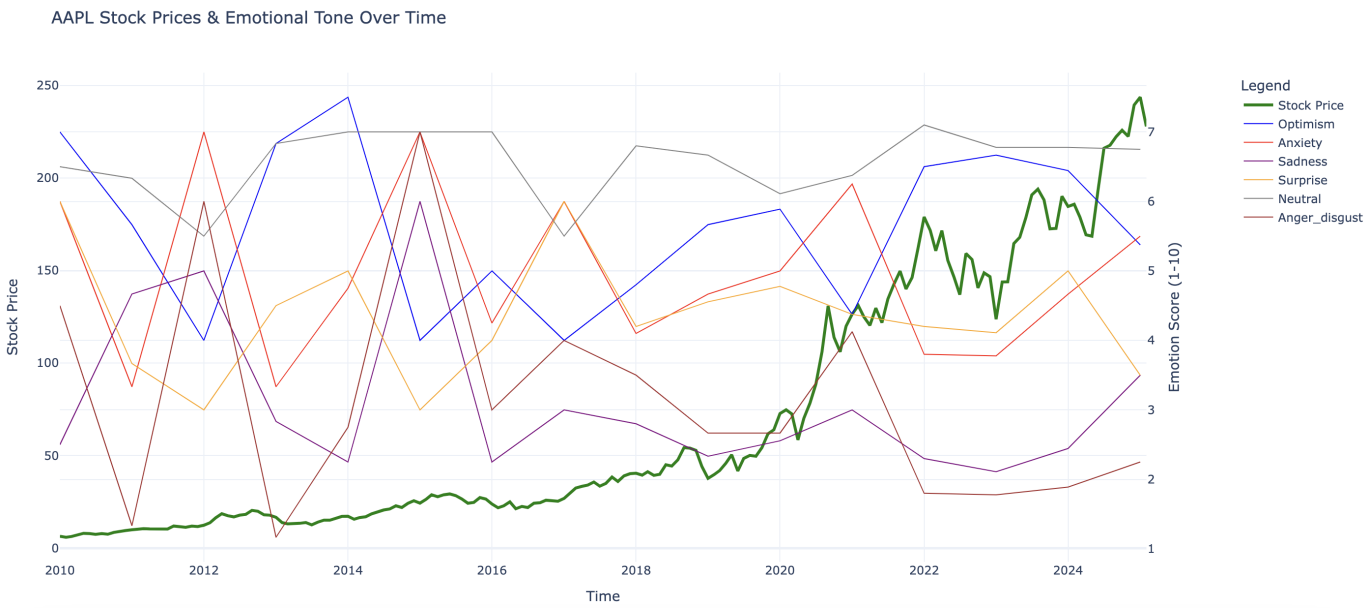

Fig. 1. Emotional tone graph for AAPL

1) Emotional Tone Accuracy: We evaluated the accuracy of the emotional tone extraction process by comparing the results from the Gemini API with a small manually labeled dataset of financial news articles. Although this manual validation was limited, it provided a general sense of how well the API was able to classify the emotional tone of financial news. We focused primarily on ensuring that the scores for key emotions, such as optimism and anxiety, aligned with the content of the articles.

2) Correlation Between Emotional Tones and Stock Price: To assess the relevance of emotional tones to stock price movements, we performed a correlation analysis. Specifically, we looked at the relationship between the emotional sentiment (e.g., optimism or anxiety) and stock price changes over the same time period. This allowed us to quantify how fluctuations in sentiment correlate with market trends, providing a basis for a predictive model.

3) Visualizing Narrative Evolution: We also evaluated the narrative evolution tracking through the construction of network graphs (As shown in Fig. 2). The accuracy and utility of these graphs were assessed qualitatively by inspecting whether key events and articles were accurately represented and connected. We verified that the narrative graph structure accurately reflected the evolution of financial topics over time.

D. Additional Analysis and Optimization

1) Data Integration and Final Visualization: Finally, we integrated the emotional tone data with the stock price data for visualization purposes. The goal was to create an interactive time-series line graph, which would show the stock prices alongside emotional tone data for different emotions (As shown in Fig. 1). This visualization was built using Plotly.js and integrated into a Flask web interface for easy exploration.

III. Results

A. System Performance Overview

Our system was tested extensively using Amazon's shareholder letters, along with corresponding financial data and news articles spanning from 2010 to 2025. The results demonstrated the model's ability to capture nuanced financial narratives, linking emotional tones, key themes, and cause-and-effect relationships into a comprehensive network graph.

B. Emotional Tone Analysis

The emotional tone analysis accurately extracted sentiment from diverse sources, including news articles, social media posts, and shareholder letters. The graph generated for Amazon illustrated fluctuations in emotions such as optimism, anxiety, and sadness, with notable peaks and troughs aligning with major financial events.

Key findings include:

- *Optimism spikes* were observed during significant growth phases, such as Amazon's expansion into new markets and technological innovations.
- *Anxiety and sadness* correlated with periods of regulatory challenges and supply chain disruptions.
- *Surprise* emerged in reaction to unexpected earnings reports and acquisitions.
- *Neutral sentiments* appeared during periods of stability, providing a baseline for comparison.

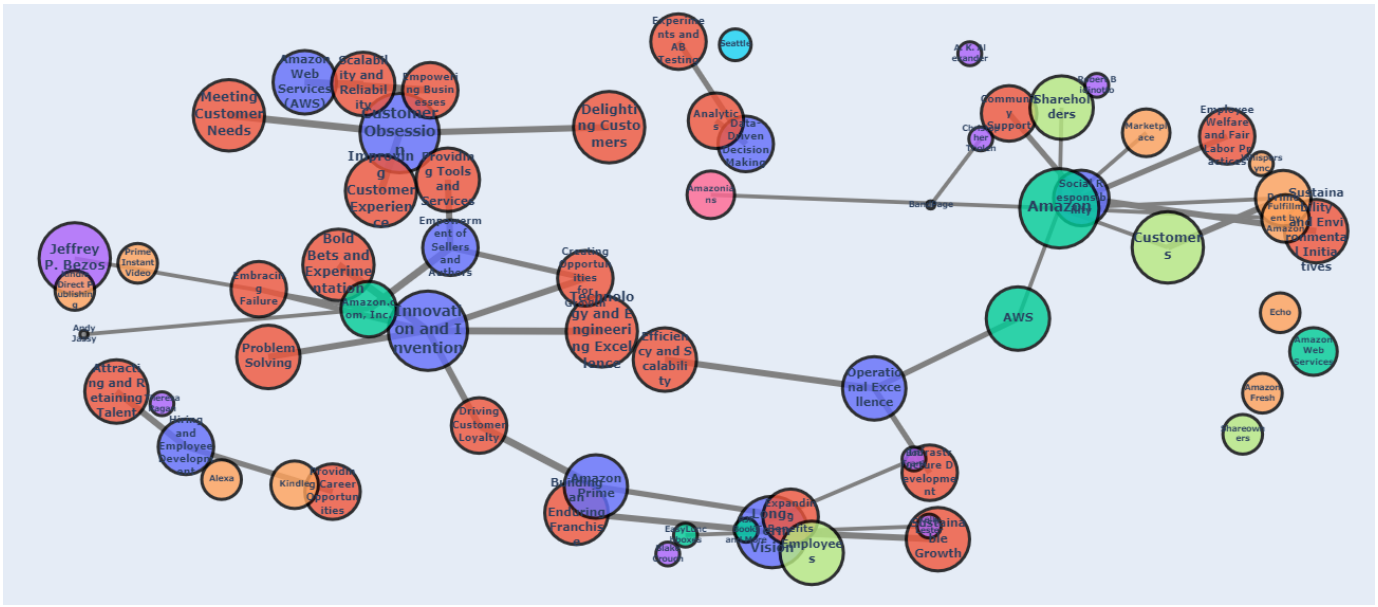

Fig. 2. Genome graph for AMZN

The interactive nature of the graph, built with Plotly and Flask, allows users to explore yearly emotional trends and compare sentiment against stock price movements. This gives an accessible yet powerful way to visualize emotional undercurrents driving market behavior.

C. Thematic and Entity Relationship Extraction

Our system successfully identified and categorized key themes, entities, and relationships. The network graph output visualizes these elements, creating a web of interconnected insights.

Key thematic extractions included:

- *AI-driven innovation* linked to revenue growth and product diversification.
- *Supply chain resilience* emerging as a recurring topic during COVID-19 and global disruptions.
- *Regulatory pressures* connected to concerns about antitrust actions and data privacy.
- *Sustainability and ESG (Environmental, Social, Governance)* themes associated with Amazon's carbon neutrality pledges and renewable energy initiatives.

Entity relationship extraction mapped crucial connections, such as partnerships, leadership changes, and competitive dynamics, showcasing how these factors intertwined with sentiment and market performance.

D. Cause and Effect Mapping

One of the most powerful outcomes of our model was the creation of a cause-and-effect narrative map. Using semantic matching through sentence embeddings and a transformer model, the system linked key financial events to their consequences.

Examples:

- *Cost reduction strategies* leading to *higher profitability* and corresponding optimism spikes.
- *Increased regulatory scrutiny* resulting in *anxiety* and *stock price stagnation*.
- *Cloud services expansion* driving *positive sentiment* and stock surges.

This causality mapping offers actionable insights by showing not only what happened, but also why, making it a valuable tool for analysts and investors alike.

E. Website Functionality and User Interaction

Our React-based website provides an intuitive interface for users to input a company's ticker symbol and retrieve detailed financial narratives. Beyond the emotional tone graph, users can explore:

- *Annual emotional sentiment breakdowns*, with a focus on key events.
- *Stock price overlays* to compare sentiment with market performance.
- *Company narrative summaries*, generated via the Gemini Flash 2 API, condensing 15 years of data into a readable financial storyline.
- *Network graph visualizations*, highlighting key themes, entities, and causal pathways.

F. Comparative Analysis with Traditional Sentiment Analysis

To further validate our approach, we compared our financial narrative genome system with traditional sentiment analysis methods. The results highlighted clear advantages:

- *Depth of insight:* Traditional sentiment analysis yielded broad positive/negative sentiments, while our system uncovered thematic context and causality.
- *Accuracy* By filtering irrelevant content and focusing on company-specific narratives, our model reduced noise and improved sentiment precision.
- *Interpretability* The network graph and narrative output made complex data more digestible for users, enhancing decision-making.

G. Future Enhancements

The current results showcase a robust, scalable foundation, but there are several directions for enhancement:

- *Industry-wide benchmarking:* Extending analysis to entire sectors for cross-company comparison.
- *Real-time updates:* Incorporating live news feeds and stock data for up-to-date financial narratives.
- *Deeper social media integration:* Capturing emerging trends and retail investor sentiment.
- *Event prediction models*: Training the system to forecast potential market impacts based on evolving narratives.
- *Sentiment refinement*: Exploring multi modal analysis, combining text with audio or video content to enhance emotional extraction accuracy.

Our results demonstrate that the Financial Narrative Genome approach delivers a more comprehensive, interpretable, and actionable understanding of company performance than traditional sentiment analysis. This method not only aids in retrospective analysis but also holds the potential to improve forecasting, risk assessment, and investment strategies.

IV. Conclusion

This paper presented the 'Financial Narrative Genome,' an AI-driven system designed to extract, represent, and analyze the complex narratives that shape financial market behavior. By leveraging advanced NLP techniques, we demonstrated the feasibility of constructing a network graph that captures the dynamic interplay of themes, causal relationships, and emotional tones within financial texts. We further showed how this 'narrative genome' can be used to track emotional tone changes over time and identify key narrative shifts that correlate with market fluctuations. This research provides a foundation for a more nuanced and insightful approach to financial market analysis, moving beyond traditional quantitative methods.

Building upon this work, our next steps would focus on enhancing the system's predictive capabilities. Specifically, we aim to develop more sophisticated machine learning models that can leverage the 'narrative genome' to forecast market trends with greater accuracy. This would involve incorporating a wider range of data sources, including real-time social media feeds and alternative data sets, to improve the system's robustness and responsiveness. A significant challenge remains in addressing the inherent subjectivity of narrative interpretation and mitigating potential biases in the data and models. Furthermore, developing a robust evaluation framework that can accurately assess the system's predictive performance is crucial.

Alright, let's craft a strong conclusion section for your "Financial Narrative Genome" paper:

Conclusion:

"This paper presented the 'Financial Narrative Genome,' an AI-driven system designed to extract, represent, and analyze the complex narratives that shape financial market behavior. By leveraging advanced NLP techniques, we demonstrated the feasibility of constructing a network graph that captures the dynamic interplay of themes, causal relationships, and emotional tones within financial texts. We further showed how this 'narrative genome' can be used to track emotional tone changes over time and identify key narrative shifts that correlate with market fluctuations. This research provides a foundation for a more nuanced and insightful approach to financial market analysis, moving beyond traditional quantitative methods.

Building upon this work, our next steps would focus on enhancing the system's predictive capabilities. Specifically, we aim to develop more sophisticated machine learning models that can leverage the 'narrative genome' to forecast market trends with greater accuracy. This would involve incorporating a wider range of data sources, including real-time social media feeds and alternative data sets, to improve the system's robustness and responsiveness. A significant challenge remains in addressing the inherent subjectivity of narrative interpretation and mitigating potential biases in the data and models. Furthermore, developing a robust evaluation framework that can accurately assess the system's predictive performance is crucial.

In our opinion, the most important area to focus on next is the integration of real-time data and the development of adaptive learning algorithms. This would enable the system to continuously learn and adapt to the evolving dynamics of financial narratives, providing more timely and accurate insights. By addressing these challenges and focusing on real-time adaptation, we believe the 'Financial Narrative Genome' has the potential to become a valuable tool for investors, analysts, and researchers seeking to navigate the complexities of the modern financial landscape.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," pp. 5998–6008, 2017.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018.

[3] M. Wankhade, A. C. S. Rao, and C. Kulkarni, "A survey on sentiment analysis methods, applications, and challenges," 2022.

[4] E. Cambria and A. Hussain, "Sentic computing," 2015.