

Integrating Market Microstructure Mechanisms into Financial Knowledge Graphs: A Semantic Modeling Approach

Yuxuan Zhao ^{1,*}

¹ Guangdong University of Finance & Economics, Guangzhou, China

* Correspondence: Yuxuan Zhao, Guangdong University of Finance & Economics, Guangzhou, China

Abstract: The rapid evolution of financial markets necessitates advanced data representation techniques that can capture both macro-level entity relationships and micro-level trading dynamics. Traditional Financial Knowledge Graphs (FKGs) have primarily focused on static or slow-moving data, such as corporate hierarchies, industry classifications, and supply chain links. However, these structures often fail to incorporate the high-frequency, granular insights provided by market microstructure mechanisms, including order flow toxicity, liquidity fluctuations, and price discovery processes. This research proposes a novel semantic modeling approach to integrate market microstructure mechanisms into a unified Financial Knowledge Graph framework. By developing a specialized ontology that defines classes for limit order book events, participant behaviors, and execution patterns, the study enables the transformation of raw high-frequency trading data into structured semantic triples. The methodology involves a multi-stage process: first, the design of a temporal-aware ontology; second, the extraction of microstructure features using signal processing techniques; and third, the fusion of these features into an existing knowledge graph using graph embedding algorithms. Experimental results demonstrate that the integrated model significantly improves the representation of market states and enhances the predictive power of the graph for short-term volatility and liquidity risk. The findings suggest that semantic modeling provides a robust bridge between qualitative financial knowledge and quantitative market mechanics. This integration allows for a more holistic understanding of market fragility and the interconnectedness of financial assets under high-frequency conditions. The study concludes that incorporating microstructure dynamics into knowledge graphs is essential for the next generation of intelligent financial monitoring and algorithmic decision-support systems, offering a scalable solution for real-time market analysis.

Keywords: Financial Knowledge Graph; Market Microstructure; Semantic Modeling; Ontology Engineering; High-Frequency Data

Received: 31 January 2026

Revised: 20 March 2026

Accepted: 30 March 2026

Published: 05 April 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Research Background and Motivation

The evolution of financial data management has witnessed a significant transition from traditional relational databases to graph-based representations, driven by the increasing complexity and interconnectedness of modern financial systems [1, 2]. Relational databases, while effective for structured and static datasets, often struggle to capture the dynamic, multi-relational nature of financial markets, where entities and their interactions evolve rapidly. This limitation becomes particularly pronounced in the context of electronic markets, where high-frequency trading and granular transaction data generate intricate webs of relationships that demand more flexible and semantically rich modeling approaches [3]. Existing financial knowledge frameworks predominantly focus on macro-level entities, such as institutions, instruments, and regulatory structures, often overlooking the microstructure dynamics that underpin market behavior [4]. This disconnect underscores the need for a unified semantic framework capable of bridging the gap between high-level abstractions and the fine-grained realities of modern trading

environments [5]. By integrating market microstructure mechanisms into financial knowledge graphs, it becomes possible to achieve a more holistic representation of financial ecosystems, enabling deeper insights and more robust analytical capabilities.

1.2. Problem Statement and Objectives

Current financial knowledge graphs exhibit significant limitations in their ability to capture the granular, intraday dynamics inherent to market microstructure. While these graphs effectively model static relationships between corporate entities, economic indicators, and broader financial systems, they often fail to incorporate the temporal and transactional nuances that define high-frequency trading environments [4, 6]. This gap restricts their utility in analyzing the intricate interplay of order flows, price movements, and liquidity variations that occur within shorter timeframes. The primary objective of this research is to develop a semantic framework that bridges this divide, enabling the integration of market microstructure variables into financial knowledge graphs [7, 8]. By facilitating the querying and analysis of these variables alongside corporate and economic entities, the proposed approach aims to enhance the representational fidelity and analytical capabilities of financial knowledge graphs in dynamic market contexts.

2. Theoretical Framework and Literature Review

2.1. Foundations of Financial Knowledge Representation

Financial knowledge representation has undergone significant evolution, transitioning from static ontological frameworks to more dynamic and adaptive graph-based structures. Traditional approaches often relied on rigid taxonomies and hierarchical schemas to model financial entities such as assets, institutions, and transactions, emphasizing predefined relationships and static categorizations [2]. While these methods provided foundational clarity, they struggled to accommodate the complexity and fluidity of real-world financial systems, particularly in capturing temporal dynamics and contextual variability [9]. Recent advancements have shifted toward the use of semantic networks and knowledge graphs, which enable the integration of heterogeneous data sources and the representation of multi-dimensional relationships [10, 11]. Dynamic graph structures, in particular, allow for the modeling of evolving interactions, such as fluctuating market linkages or adaptive trading behaviors, by incorporating temporal and event-driven elements [12]. This progression reflects a broader trend toward more flexible and expressive frameworks capable of addressing the intricate interdependencies inherent in financial ecosystems.

2.2. Market Microstructure Theory and Data Dynamics

Market microstructure theory provides a foundational framework for understanding the mechanisms governing financial markets at a granular level [3, 8]. Central to this framework is the limit order book, which serves as a dynamic repository of buy and sell orders, organized by price and time priority. The bid-ask spread, representing the difference between the highest bid and lowest ask prices, encapsulates the immediate cost of liquidity and reflects underlying market conditions such as volatility and order flow [6]. Order flow imbalance, defined as the net difference between buy and sell orders over a given interval, offers insights into market sentiment and price pressure [9]. These objective properties exhibit structured patterns that lend themselves to semantic categorization, enabling the integration of microstructure dynamics into financial knowledge graphs. By capturing these mechanisms semantically, it becomes possible to enhance the interpretability and interoperability of financial data within broader analytical systems [13].

3. Semantic Modeling Methodology

3.1. Microstructure Ontology Design

The proposed microstructure ontology is designed to capture the intricate dynamics of trading events and their interactions within financial markets. Central to this ontology are classes such as `'OrderEvent'`, `'LiquidityProvider'`, and `'PriceImpact'`, which collectively model the flow of market activity. The `'OrderEvent'` class represents discrete actions initiated by market participants, such as buy or sell orders, cancellations, and modifications. This class is linked to financial instruments through the `'Instrument'` property, ensuring that each event is contextually tied to specific assets. The `'LiquidityProvider'` class encapsulates entities that supply liquidity, such as market makers or institutional participants, and includes attributes that characterize their role, such as bid-ask spread contributions and order book depth adjustments.

The hierarchical structure of the ontology is designed to reflect the layered complexity of market microstructure. At the top level, trading events are categorized into broader types, such as market orders and limit orders, with subclasses capturing finer distinctions. Properties such as `'Timestamp'`, `'Volume'`, and `'Price'` are used to detail the temporal and quantitative aspects of each event. The `'PriceImpact'` class quantifies the effect of trading activity on asset prices, incorporating attributes like volatility measures and price elasticity [7]. Relationships between classes are defined using semantic links, such as `'affects'` and `'initiatedBy'`, which connect trading events to liquidity providers and financial instruments. This design enables a granular representation of market interactions while maintaining semantic coherence across the ontology.

3.2. Data Acquisition and Preprocessing Pipeline

The process of cleaning and normalizing high-frequency message data is critical for ensuring the reliability of semantic extraction in financial knowledge graphs. Initially, raw tick-level data, which often contains irregularities such as missing values, duplicate entries, and out-of-sequence timestamps, undergoes rigorous preprocessing. Timestamp synchronization is a central challenge, as data streams from multiple sources may exhibit varying temporal resolutions and clock skews [6]. To address this, all timestamps are aligned to a unified temporal framework, typically using a high-precision reference clock. Following synchronization, tick-level data is aggregated into semantically meaningful intervals, such as fixed time windows or event-driven periods, to reduce noise and enhance interpretability [11]. Aggregation involves calculating representative metrics, such as average price, trade volume, and bid-ask spreads, which serve as inputs for subsequent semantic modeling. This pipeline ensures that the processed data is both temporally coherent and semantically enriched, forming a robust foundation for downstream analytical tasks.

3.3. Triple Extraction and Graph Construction

The process of converting numerical market signals into Resource Description Framework (RDF) triples involves a structured methodology that bridges quantitative data with semantic representations. At its core, this approach leverages threshold-based logic to identify significant market conditions and event sequences, which are then encoded as subject-predicate-object triples. Numerical signals, such as price movements, trading volumes, or volatility metrics, are first analyzed to detect predefined thresholds or patterns indicative of specific market phenomena. For instance, a price increase exceeding a certain percentage within a defined time window may be interpreted as a "price surge" event, while a simultaneous rise in trading volume could establish a correlated relationship.

Once these conditions are identified, the system asserts semantic relationships by mapping the detected events to domain-specific ontologies. For example, a "price surge" event might be represented as a triple where the subject is the financial instrument, the predicate denotes the type of event, and the object specifies the magnitude or classification

of the surge [6]. Sequential event detection further enables the construction of temporal relationships, such as identifying causality or co-occurrence between events [5]. This layered approach ensures that the resulting RDF graph encapsulates both the quantitative dynamics of market behavior and their semantic interpretations, facilitating downstream analytical tasks.

4. Implementation and Knowledge Fusion

4.1. Temporal Graph Embedding Techniques

Temporal graph embedding techniques are instrumental in capturing the dynamic nature of financial knowledge graphs by representing their evolving structure in a continuous vector space. These methods leverage the temporal dimension of market data to encode both structural and temporal dependencies. A key component of this process is the incorporation of time-decay functions, which assign greater importance to recent events while gradually diminishing the influence of older interactions. This ensures that the embeddings remain sensitive to the most current market conditions, reflecting the transient nature of financial systems.

The embedding process typically begins by defining a temporal adjacency matrix, where edges between nodes are weighted based on their temporal proximity and relevance. Time-decay functions, such as exponential decay $w(t) = e^{-\lambda t}$, are applied to these weights to prioritize recent interactions. Additionally, temporal random walk strategies are employed to sample sequences of nodes, ensuring that the temporal order of events is preserved during the embedding generation. These sampled sequences are then processed using neural embedding models, such as skip-gram or graph neural networks, to produce low-dimensional representations that encode both topological and temporal features [1, 5].

By integrating time-aware mechanisms, these embeddings enable downstream tasks, such as link prediction and anomaly detection, to account for the dynamic and time-sensitive nature of financial markets. This approach enhances the ability of financial knowledge graphs to adapt to evolving market conditions and provide actionable insights.

4.2. Integrating Micro and Macro Financial Data

The integration of microstructure data into macro-level financial knowledge graphs necessitates a robust framework to address the inherent challenges of entity resolution and semantic consistency across differing data scales. Entity resolution is particularly critical, as microstructure datasets often feature granular, transaction-level entities such as individual trades, order book entries, and market participants, which must be reconciled with broader macro-level entities like corporations, sectors, and economic indicators [7]. This process involves the identification and linking of equivalent entities across datasets, ensuring that micro-level nodes and edges are accurately mapped to their macro-level counterparts without introducing redundancy or semantic ambiguity.

Maintaining semantic consistency across scales further requires the alignment of ontologies and schema definitions used in both graphs [8]. Microstructure data typically emphasizes temporal precision and transactional dynamics, while macro-level graphs prioritize aggregated trends and structural relationships. To harmonize these perspectives, a semantic mapping approach is employed, wherein shared attributes, such as timestamps, asset identifiers, and market classifications, serve as anchor points for integration. Additionally, hierarchical modeling techniques are applied to preserve the nested relationships between micro and macro entities, ensuring that localized market behaviors are contextualized within broader economic frameworks.

The fusion process is iterative, leveraging automated matching algorithms alongside domain expertise to refine mappings and resolve discrepancies. By systematically bridging the granular and aggregated dimensions of financial data, the resulting unified graph facilitates enhanced analytical capabilities, enabling insights that span both microstructure mechanisms and macroeconomic phenomena.

5. Experimental Results and Evaluation

5.1. Graph Structural Analysis

The structural analysis of the integrated financial knowledge graph reveals significant enhancements in connectivity and density following the incorporation of market microstructure data. Quantitative evaluation of node degree distribution demonstrates a shift towards higher average degrees, indicating that the graph's nodes are more interconnected compared to traditional financial networks. This increased connectivity reflects the ability of the integrated graph to capture nuanced relationships between entities, such as traders, assets, and market events. Furthermore, the clustering coefficient analysis highlights a notable rise in local cohesiveness, suggesting that the graph exhibits stronger community structures. These findings imply that the inclusion of microstructure mechanisms facilitates the emergence of densely connected subgraphs, which are critical for modeling complex financial interactions. Overall, the enriched graph topology underscores the value of integrating microstructure data in enhancing the semantic representation of financial networks, thereby supporting more robust analyses and predictive modeling.

5.2. Predictive Performance and Case Studies

The evaluation of the model's predictive performance focuses on its ability to identify and forecast market anomalies, such as flash crashes and liquidity droughts, by leveraging the semantic structure of the financial knowledge graph. The model demonstrates a capacity to detect precursors to these anomalies by analyzing specific semantic paths that capture intricate relationships between market participants, instruments, and events. For instance, in the case of a flash crash, the model identifies a sequence of heightened trading activity, abrupt shifts in order book depth, and correlated sentiment changes across social media and news sources. These interconnected signals, represented as semantic paths within the graph, provide early warning indicators of impending price dislocations.

Similarly, the model's performance in identifying liquidity droughts highlights its ability to integrate diverse data sources, such as historical trade volumes, bid-ask spreads, and institutional order flows. By tracing patterns of diminishing liquidity through the graph, the system uncovers latent vulnerabilities that precede significant market stress. A case study involving a prolonged liquidity drought revealed that the model successfully flagged a sequence of diminishing counterparty interactions and widening spreads, which aligned with subsequent price stagnation and volatility spikes.

These findings underscore the utility of semantic modeling in capturing the complex interplay of market microstructure elements. By mapping these interactions onto a graph-based framework, the model not only enhances anomaly detection but also provides interpretable pathways that elucidate the underlying mechanisms driving market disruptions. This capability positions the approach as a valuable tool for both predictive analytics and the proactive management of systemic risks in financial markets.

6. Discussion

6.1. Interpretability of Semantic Market Links

The semantic modeling approach offers a significant advantage in terms of interpretability compared to traditional black-box machine learning models [9]. By leveraging the graph structure inherent in financial knowledge graphs, this method provides a transparent representation of the relationships and dependencies between market entities. Unlike black-box models, which often obscure the reasoning behind their outputs, the semantic approach explicitly encodes the connections and interactions that drive market behavior. For instance, the graph structure can reveal how specific market events, such as price shocks or liquidity imbalances, propagate through interconnected entities, thereby elucidating the underlying drivers of market volatility. This transparency not only enhances the explainability of the model's outputs but also facilitates a deeper

understanding of complex market dynamics. Furthermore, the ability to trace semantic links across the graph enables researchers and practitioners to identify causal pathways and assess the impact of individual factors on market outcomes. As a result, the semantic approach aligns more closely with the interpretability requirements of financial decision-making and regulatory frameworks.

6.2. Scalability and Computational Challenges

Maintaining a high-frequency financial knowledge graph presents significant computational challenges, particularly in balancing semantic depth with real-time processing efficiency. The integration of market microstructure mechanisms necessitates the continuous ingestion and semantic modeling of vast quantities of high-frequency data, which often arrive in sub-second intervals [2]. This demands robust infrastructure capable of handling large-scale data streams while preserving the nuanced relationships and contextual information required for meaningful knowledge representation [2]. However, increasing semantic depth—such as encoding complex interdependencies between market participants and instruments—introduces substantial computational overhead, as it involves more intricate reasoning and graph updates.

A critical trade-off emerges between the granularity of semantic modeling and the system's ability to process and update the graph in real time. While deeper semantic representations enhance the graph's utility for advanced analytics and decision-making, they can significantly slow down processing speeds, particularly when dealing with high-dimensional data. Optimizing this balance requires innovative approaches, such as parallelized graph processing, dynamic indexing, and selective pruning of less relevant data, to ensure scalability without compromising the graph's semantic integrity.

7. Conclusion

7.1. Summary of Research Contributions

This research successfully demonstrates the integration of market microstructure mechanisms into financial knowledge graphs, offering a novel semantic modeling approach to bridge the gap between trading mechanics and entity relationships. By leveraging the intricate dynamics of market microstructure, the proposed framework captures the multifaceted interactions between trading participants, instruments, and market conditions, enabling a more granular and interconnected representation of financial ecosystems. Unlike traditional methods that often treat trading mechanics and financial entities as isolated components, this approach emphasizes their interdependence, providing a unified semantic layer that enhances interpretability and analytical depth.

The novelty of this work lies in its ability to model the interplay between microstructural elements, such as order flows, price formation, and liquidity dynamics, within a graph-based structure. This semantic modeling approach not only enriches the descriptive power of financial knowledge graphs but also facilitates advanced analytical capabilities, such as the identification of emergent patterns and systemic risks. By addressing the limitations of existing methodologies, this research contributes to the growing body of work aimed at integrating domain-specific mechanisms into knowledge representation frameworks, paving the way for more robust and actionable insights in financial analysis.

7.2. Practical Implications and Future Work

The integration of market microstructure mechanisms into financial knowledge graphs presents significant opportunities for financial institutions, particularly in the domains of risk management and algorithmic trading. By leveraging the semantic modeling framework proposed in this study, institutions can enhance their ability to identify and mitigate systemic risks through improved representation and analysis of market dynamics. The structured incorporation of microstructure data enables a more granular understanding of liquidity patterns, order flow imbalances, and price formation

processes, which are critical for developing robust risk assessment models. Furthermore, the framework supports the optimization of algorithmic trading strategies by providing a semantic foundation for real-time decision-making, allowing algorithms to adapt dynamically to evolving market conditions.

Future research should explore the integration of sentiment analysis derived from financial news and media into the microstructure graph. This additional layer of contextual information could enrich the graph's predictive capabilities by linking market behaviors with external sentiment drivers. Developing methodologies to process unstructured textual data and map sentiment indicators onto graph nodes and edges would be a valuable extension of this work. Additionally, investigating the interplay between sentiment dynamics and microstructure variables could yield insights into how psychological factors influence market efficiency and volatility. Such advancements would further solidify the utility of financial knowledge graphs in addressing complex challenges within modern financial systems.

References

1. S. Yuan, "Mechanisms of High-Frequency Financial Data on Market Microstructure," *Modern Economics & Management Forum*, vol. 6, no. 4, pp. 569–572, 2025.
2. K. Schlepper, H. Hofer, R. Riordan, and A. Schrimpf, "The market microstructure of central bank bond purchases," *J. Financ. Quant. Anal.*, vol. 55, no. 1, pp. 193–221, 2020.
3. B. Oza and A. Behnaz, "Using knowledge graphs for enabling collaborative financial market data analytical processes," *Int. J. Complex. Appl. Sci. Technol.*, vol. 1, no. 2, pp. 142–154, 2024.
4. R. Ren, *Modeling and Simulation of Order Book Dynamics: A Study on Financial Market Microstructure*, Doctoral dissertation, Cornell University, 2023.
5. S. Yuan, "Conceptual modeling and semantic relations in the construction of financial knowledge graphs," *Econ. Manag. Innov.*, vol. 3, no. 1, pp. 64–70, 2026.
6. X. Jiao, Z. Li, C. Xu, Y. Liu, W. Liu, and J. Bian, "Microstructure-empowered stock factor extraction and utilization," arXiv preprint arXiv:2308.08135, 2023.
7. A. Hota, "Generative AI-powered portfolio optimization through retrieval-augmented generation and contextual financial knowledge graphs," in *2025 Int. Conf. Electron. Comput., Commun. Netw. Autom. Technol. (ICEC2NT)*, pp. 1–4, Sep. 2025.
8. M. K. Brunnermeier, "Prices, price processes, volume and their information-A survey of the market microstructure literature," SSRN, 2010.
9. G. Rao, T. Lu, L. Yan, and Y. Liu, "A hybrid LSTM-KNN framework for detecting market microstructure anomalies: Evidence from high-frequency jump behaviors in credit default swap markets," *J. Knowl. Learn. Sci. Technol.*, vol. 3, no. 4, pp. 361–371, 2024.
10. N. Sharma, S. Manohar, and A. S. Rao, "Decoding market voices: Beyond the charts analysis of cryptocurrency market microstructure," *F1000Research*, vol. 15, p. 274, 2026.
11. G. Maitrier, *The Hidden Complexity of Prices: Exploring Microstructural Mechanisms in Financial Markets*, Doctoral dissertation, Institut Polytechnique de Paris, 2025.
12. C. A. Lehalle, "Market microstructure knowledge needed for controlling an intra-day trading process," arXiv preprint arXiv:1302.4592, 2013.
13. K. A. Dayri, *Market Microstructure and Modeling of the Trading Flow*, Doctoral dissertation, Ecole Polytechnique X, 2012.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.