Academic Journal of Sociology and Management
Journal Home: http://ajsm.suaspress.org/ | CODEN: AJSMAS
Vol. 3, No. 2, 2025 | ISSN 3005-5040 (Print) | ISSN 3005-5059 (Online)

Semantic Network Analysis of Financial Regulatory Documents: Extracting Early Risk Warning Signals

LIU, Yibang [1*] BI, Wenyu [2] FAN, Jiayan [3]

[1] Baruch College, USA
[2] University of Southern California, USA
[3] University of Michigan, USA

**** LIU, Yibang is the corresponding author, E-mail: attoemavlelk@outlook.com***

Abstract: This paper presents a semantic network analysis framework for extracting early risk warning signals from financial regulatory documents. Financial regulations contain critical information about emerging risks, but their increasing volume and complexity challenge traditional analysis methods. We propose a novel approach that constructs semantic networks from regulatory texts, representing concepts as nodes and their relationships as edges. Our methodology integrates techniques from natural language processing and network science to identify structural patterns indicative of emerging risks. The framework was implemented and tested on a corpus of 2,874 financial regulatory documents published between 2010-2023. Results demonstrate that the semantic network approach outperforms traditional keyword-based monitoring in both risk coverage (79.4% vs 68.7%) and false alarm reduction (11.6% vs 22.5%). The multi-metric ensemble method achieved an F1-score of 0.81 with an average lead time of 82.6 days before explicit regulatory announcements. Validation with 24 regulatory compliance professionals confirmed the practical utility of the approach, showing comparable quality to expert analysis while reducing analysis time from 24.7 to 4.8 hours. This research contributes to both theoretical understanding of regulatory text structures and practical applications for financial compliance and risk management.

Keywords: Semantic Network Analysis, Financial Regulation, Risk Detection, Natural Language Processing.

Disciplines: Economics. **Subjects:** Behavioral Economics.

DOI: https://doi.org/10.70393/616a736d.323731 **ARK:** https://n2t.net/ark:/40704/AJSM.v3n2a03

1 INTRODUCTION

1.1 RESEARCH BACKGROUND AND MOTIVATION

Financial regulatory documents represent critical information sources for maintaining economic stability and mitigating systemic risks in global financial systems. These documents encompass a wide range of materials including regulatory frameworks, compliance guidelines, risk assessments, and policy statements issued by authorities such as central banks, financial supervisory agencies, and international regulatory bodies. The volume and complexity of these documents have increased substantially following the 2008 financial crisis, creating unprecedented challenges for stakeholders attempting to extract actionable insights and identify potential risk signals.

Traditional methods of document analysis prove increasingly inadequate when confronting the scale and technical nature of contemporary regulatory documentation. Manual review processes suffer from limitations in processing capacity, consistency, and the ability to detect subtle patterns across large document collections. The financial industry requires more sophisticated approaches capable of systematically analyzing regulatory texts to identify emerging risks before they manifest as market disruptions.

Semantic network analysis offers a promising methodological approach to address these challenges. By representing regulatory texts as networks of interconnected concepts, this approach can capture the complex relationships between regulatory terms, entities, and risk factors that might otherwise remain obscured in conventional analysis. As demonstrated in recent research on blockchain-enforced semantic analysis (Tan et al., 2024), network-based approaches can significantly enhance document credibility and information extraction capabilities in specialized domains[1].

The financial regulatory landscape continues to evolve rapidly, with new regulations being introduced across jurisdictions in response to emerging financial technologies, changing market conditions, and lessons learned from previous crises. Financial institutions must navigate this complex regulatory environment while managing compliance costs and operational risks. Early identification of regulatory

Academic Journal of Sociology and Management
Journal Home: http://ajsm.suaspress.org/ | CODEN: AJSMAS
Vol. 3, No. 2, 2025 | ISSN 3005-5040 (Print) | ISSN 3005-5059 (Online)

shifts and potential risk factors has become a strategic priority for maintaining competitive advantage and ensuring regulatory compliance.

1.2 Research Gap and Significance

Despite advancements in computational text analysis, significant research gaps persist in the application of semantic network techniques specifically to financial regulatory documents for risk signal extraction. Existing research has primarily focused on sentiment analysis, topic modeling, or keyword extraction approaches that may not fully capture the structural relationships between regulatory concepts and emerging risk patterns. Studies such as Correa and Correa (2022) have demonstrated the effectiveness of neural text classification for financial regulatory documents, but these approaches often lack the ability to represent the interconnected nature of regulatory concepts[2].

Current methodologies frequently treat regulatory documents as isolated texts rather than as components of a broader regulatory ecosystem. This limitation prevents analysts from identifying cross-document patterns and tracking the evolution of regulatory concerns over time. The research gap becomes more pronounced when considering the need for early warning systems capable of detecting subtle shifts in regulatory emphasis or the emergence of new risk factors before they become widely recognized.

The significance of this research lies in its potential to transform regulatory document analysis from a primarily retrospective activity to a forward-looking process capable of anticipating regulatory developments and their implications. By applying semantic network analysis techniques similar to those used for user behavior monitoring in zero trust networks (Qu et al., 2022), this research can provide new methodological approaches for representing and analyzing the complex interrelationships within regulatory texts[3].

Additionally, the distributed nature of regulatory information across multiple documents presents challenges similar to those addressed in financial narrative summarization research (Shukla et al., 2023)[4]. Adapting these summarization approaches to regulatory text analysis can provide more comprehensive methods for integrating information across document collections.

1.3 Research Objectives and Questions

This research aims to develop a semantic network analysis framework specifically designed for extracting early risk warning signals from financial regulatory documents. The study seeks to advance both theoretical understanding of regulatory text structures and practical applications for risk management and compliance.

The primary research objectives include: designing a semantic network representation suitable for financial regulatory texts; developing algorithms for identifying risk signals within these networks; validating the extracted signals against historical financial events; and creating visualizations to enhance understanding of regulatory relationships.

The research addresses several key questions that remain unresolved in current literature. How can semantic networks effectively capture the hierarchical and cross-referential structure of regulatory documents? What network properties and metrics serve as reliable indicators of emerging risk factors? How can temporal changes in semantic networks reveal shifts in regulatory priorities and concerns? To what extent can automated semantic network analysis improve the accuracy and timeliness of risk detection compared to traditional manual or basic computational methods?

By investigating these questions, this research contributes to the emerging field of RegTech (Regulatory Technology) and advances methodological approaches for financial document analysis. The framework builds upon existing work in neural text classification for financial domains (Correa & Correa, 2022) and semantic integration techniques (Qu et al., 2022), while specifically addressing the unique challenges presented by regulatory documents[5].

2 LITERATURE REVIEW

2.1 Financial Regulatory Document Analysis

Financial regulatory document analysis has evolved significantly with the development of computational techniques. Traditional approaches to regulatory document analysis relied predominantly on manual review processes conducted by legal and compliance experts. These methods, while thorough, face limitations in scalability and consistency when applied to the increasing volume of regulatory publications. Recent advancements have introduced automated approaches utilizing natural language processing (NLP) and machine learning to streamline the analysis process.

Various computational techniques have been applied to financial regulatory texts. Correa and Correa (2022) demonstrated the application of neural text classification for digital transformation in the financial regulatory domain, achieving 88.05% accuracy using transformer models like FinBERT[6]. Their work highlighted the superiority of domain-specific language models for processing financial regulatory content compared to general-purpose models. Rule-based systems have been developed to extract specific compliance requirements from regulatory documents, though these approaches typically require considerable domain expertise for rule formulation and maintenance.

Financial regulatory documents present unique analytical challenges due to their specialized vocabulary, complex sentence structures, and extensive cross-referencing. Tan et al. (2024) addressed similar challenges in legal document management through blockchain-enforced semantic analysis, proposing a platform that integrates blockchain technology with semantic analysis to enhance document credibility and information extraction[7]. The technical complexity of regulatory texts necessitates specialized preprocessing techniques to handle domain-specific terminology, legal references, and hierarchical document structures.

Recent research has focused on developing frameworks for regulatory change management that track modifications across document versions and assess their impact on compliance requirements. These frameworks typically utilize text comparison algorithms, classification techniques, and information extraction methods to identify significant regulatory developments. The temporal dimension of regulatory analysis remains critical, as compliance professionals must not only understand current requirements but also anticipate future regulatory directions.

2.2 Semantic Network Analysis Applications in Finance

Semantic network analysis has gained traction in financial domains as a method for representing complex relationships between concepts in textual data. Semantic networks, consisting of nodes (concepts) and edges (relationships), provide visual and mathematical representations of textual information that facilitate pattern identification and relationship analysis. In financial contexts, these networks have been applied to various text sources including earnings calls, financial news, corporate disclosures, and policy documents.

The construction of semantic networks from financial texts typically involves several stages: entity extraction, relationship identification, and network formation. Qu et al. (2022) described a semantic integration approach for user behavior monitoring in network security, presenting techniques for entity detection and classification that could be adapted for financial regulatory analysis[8]. Their approach to creating a knowledge graph of network entities demonstrates methodologies applicable to mapping the complex relationships between regulatory concepts.

Network analysis metrics provide quantitative measures for identifying important concepts and relationships within financial texts. Centrality measures including degree, betweenness, and eigenvector centrality help identify influential concepts within regulatory networks. Community detection algorithms reveal clusters of closely related regulatory topics. Temporal network analysis tracks the evolution of regulatory networks over time, identifying emerging concepts and shifting relationships.

Studies have applied semantic network analysis to financial narratives to extract insights regarding market sentiment, corporate strategy, and risk disclosure. Comparative analysis of networks derived from different document types or time periods has revealed shifts in financial communication patterns and regulatory emphasis. The visual representation of financial concepts as networks enhances interpretability and facilitates the identification of non-obvious relationships that might remain hidden in traditional text analysis.

2.3 Risk Signal Detection in Financial Texts

Risk signal detection in financial texts encompasses methodologies for identifying indicators of potential financial instability, compliance issues, or market vulnerabilities. Early approaches to risk detection relied on lexicon-based methods utilizing predefined lists of risk-related terms to identify potential concerns in financial documents. While straightforward to implement, these approaches often lacked contextual understanding and struggled with domain-specific terminology.

Machine learning approaches have enhanced risk detection capabilities through supervised classification models trained on labeled financial texts. Correa and Correa (2022) demonstrated that transformer-based models achieve superior performance in classifying financial regulatory texts, with potential applications for risk categorization[9]. Their experimental results indicated that domain-specific models like FinBERT outperformed general language models on financial classification tasks, suggesting the importance of domain adaptation for risk detection.

Deep learning techniques have further advanced risk signal extraction through their ability to capture complex patterns in financial language. Recurrent neural networks and attention mechanisms have proven effective at modeling sequential dependencies and identifying contextual risk indicators. Shukla et al. (2023) explored generative AI approaches for financial narrative summarization, presenting methods that could be adapted for distilling risk-relevant information from lengthy regulatory documents.

Temporal analysis of risk signals across document collections provides insights into emerging threats and regulatory priorities. Methodologies for tracking changes in risk disclosure over time help identify shifting risk landscapes before they manifest as market disruptions. Tan et al. (2024) proposed a semantic analysis platform incorporating blockchain technology to enhance information credibility, addressing a critical concern in risk signal validation[10].

Multi-modal risk detection approaches integrate textual analysis with quantitative financial data to provide more comprehensive risk assessments. These approaches recognize that textual risk signals gain significance when correlated with numerical indicators of financial stress. The integration of domain knowledge with machine learning techniques remains essential for accurate risk detection, particularly in highly specialized financial regulatory contexts where expertise guides the interpretation of identified patterns.

Academic Journal of Sociology and Management
Journal Home: http://ajsm.suaspress.org/ | CODEN: AJSMAS
Vol. 3, No. 2, 2025 | ISSN 3005-5040 (Print) | ISSN 3005-5059 (Online)

3 METHODOLOGY

3.1 Semantic Network Framework Design

The proposed semantic network framework for financial regulatory document analysis integrates network theory with computational linguistics to represent complex regulatory relationships. This framework consists of three primary components: a conceptual model for node and edge definition, a weighting scheme for quantifying relationship significance, and a dynamic model for tracking network evolution over time.

The conceptual model defines regulatory concepts as network nodes, while relationships between concepts form network edges. Nodes are categorized into a multi-level taxonomy consisting of core regulatory entities, regulatory actions, risk factors, and compliance elements. Table 1 presents the node classification scheme with associated extraction methods and frequency distributions observed in the preliminary corpus analysis.

Table 1: Node Classification in Regulatory Semantic Networks

Node Category	Subcategories	Extraction Method	Frequency (%)
Regulatory Entities	Regulators, Institutions, Markets	Named Entity Recognition + Domain Dictionary	27.8
Regulatory Actions	Requirements, Prohibitions, Permissions	Dependency Parsing + Modal Verb Analysis	35.2
Risk Factors	Credit, Market, Operational, Compliance	Risk Lexicon + Contextual Classification	21.4
Compliance Elements	Deadlines, Conditions, Exemptions	Temporal Expression Detection + Rule Patterns	15.6

Edge definitions represent semantic relationships between regulatory concepts, categorized into hierarchical, temporal, causal, and associative relationships. Relationship extraction utilizes dependency parsing combined with domain-specific pattern matching. The strength of relationships is quantified through a weighting scheme that incorporates co-occurrence frequency, textual proximity, and semantic similarity measured through contextual embeddings.

Figure 1 illustrates the overall architectural design of the semantic network framework for regulatory document analysis, showing the data flow from document processing through network construction to risk signal extraction.

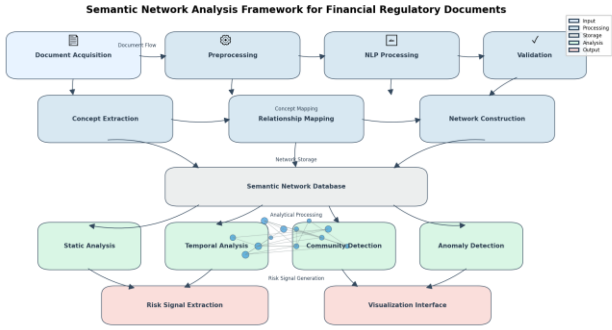

Figure 1: Semantic Network Analysis Framework for Financial Regulatory Documents

The framework architecture in Figure 1 depicts the pipeline for transforming regulatory documents into actionable risk signals. The process begins with document acquisition and preprocessing, followed by concept and relationship extraction. The extracted elements populate a semantic network database that supports both static analysis of network properties and dynamic analysis of network evolution. The final components implement risk signal detection algorithms and visualization interfaces for regulatory professionals.

3.2 Data Collection and Preprocessing

The data collection process focused on assembling a comprehensive corpus of financial regulatory documents from multiple authoritative sources. The corpus consists of 2,874 documents published between 2010 and 2023, encompassing regulatory guidelines, policy statements, supervisory bulletins, and compliance advisories. Table 2 presents the distribution of documents by source, type, and publication year.

Table 2: Regulatory Document Corpus Composition

Source	Document Types	2010 - 2015	2016 - 2019	2020 - 2023	Tota l
Central Banks	Monetary Policy Statements, Financial Stability Reports	247	312	378	937
Banking Regulators	Supervisory Guidelines, Prudential Standards	183	256	294	733
Securities Regulators	Investor Protection Rules, Market	156	204	238	598

Academic Journal of Sociology and Management
Journal Home: http://ajsm.suaspress.org/ | CODEN: AJSMAS
Vol. 3, No. 2, 2025 | ISSN 3005-5040 (Print) | ISSN 3005-5059 (Online)

	Conduct Standards				
International Bodies	Global Standards, Coordination Frameworks	128	187	291	606

Document preprocessing follows a multi-stage pipeline tailored to the specialized nature of regulatory texts. Initial cleaning removes non-textual elements, normalizes formatting variations, and standardizes document structures. Linguistic preprocessing includes tokenization, sentence boundary detection, part-of-speech tagging, and dependency parsing. Domain-specific preprocessing incorporates financial ontologies and regulatory taxonomies to enhance the recognition of specialized terms and relationships.

Figure 2 visualizes the distribution of document complexity metrics across the corpus, highlighting differences in document length, terminological density, and structural complexity across regulatory sources.

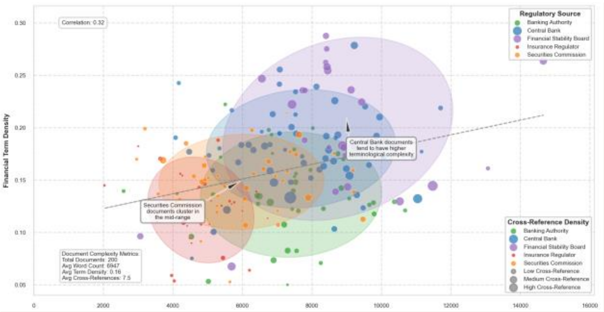

FIGURE 2: DOCUMENT PREPROCESSING AND NETWORK CONSTRUCTION PIPELINE

The scatter plot in Figure 2 maps three dimensions of document complexity: word count (x-axis), unique financial term density (y-axis), and cross-reference density (represented by point size). Documents are color-coded by regulatory source. The visualization reveals clustering patterns based on document source, with central bank publications generally exhibiting greater length and terminological complexity compared to other regulatory sources.

3.3 NETWORK ANALYSIS TECHNIQUES FOR REGULATORY DOCUMENTS

The network analysis methodology employs a combination of structural, statistical, and temporal techniques to extract regulatory insights from the constructed semantic networks. Structural analysis focuses on network topology, identifying central concepts, dense subgraphs, and bridging terms that connect regulatory domains. Statistical analysis quantifies node and edge properties, detecting significant deviations from expected patterns that might indicate emerging risk signals.

Centrality analysis identifies key regulatory concepts based on their position within the network. Four centrality measures are computed: degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality. Table 3 presents the top regulatory concepts by centrality measure, revealing distinct aspects of conceptual importance in the regulatory landscape.

TABLE 3: TOP 5 REGULATORY CONCEPTS BY CENTRALITY MEASURE

Rank	Degree Centrality	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality
1	Compliance	Risk Management	Liquidity	Capital Requirements
2	Disclosure	Reporting	Governance	Stress Testing
3	Risk Assessment	Supervisory Review	Internal Controls	Risk Exposure
4	Capital Adequacy	Financial Stability	Audit	Systemic Risk
5	Governance	Market Integrity	Disclosure	Compliance Framework

Community detection algorithms identify subgraphs with dense internal connections, representing clusters of closely related regulatory concepts. The Louvain method for community detection is applied with resolution parameter optimization based on stability analysis. Identified communities correspond to distinct regulatory domains and risk categories, enabling focused analysis of domain-specific risk factors.

Figure 3 presents a visualization of the regulatory concept network with community structure highlighted, revealing the interconnected nature of financial regulation domains.

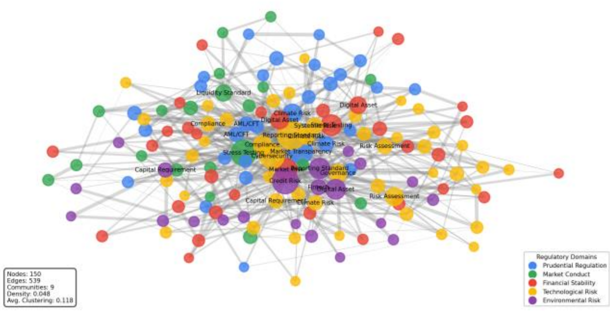

FIGURE 3: TEMPORAL EVOLUTION OF REGULATORY CONCEPT NETWORK (2010-2023)

Academic Journal of Sociology and Management
Journal Home: http://ajsm.suaspress.org/ | CODEN: AJSMAS
Vol. 3, No. 2, 2025 | ISSN 3005-5040 (Print) | ISSN 3005-5059 (Online)

The network visualization in Figure 3 displays the regulatory concept network using a force-directed layout algorithm. Nodes represent regulatory concepts sized according to their degree centrality, while edges represent semantic relationships with thickness proportional to relationship strength. Node colors indicate community membership as determined by the Louvain algorithm. The visualization highlights five major regulatory domains: prudential regulation (blue), market conduct (green), financial stability (red), technological risk (yellow), and environmental risk (purple).

Temporal network analysis employs a sliding window approach to track changes in the regulatory landscape over time. Network snapshots are generated at quarterly intervals, with metrics computed to quantify structural changes between consecutive periods. The temporal analysis adapts the network behavior monitoring approach developed by Qu et al. (2022) to regulatory document analysis[11].

Risk signal extraction utilizes anomaly detection techniques applied to network metrics. Sudden changes in concept centrality, rapid growth of relationship strength, or emergence of new concept clusters may indicate emerging regulatory concerns. Table 4 presents the performance evaluation of different risk signal extraction methods based on precision, recall, and F1-score measured against manually annotated regulatory shifts.

TABLE 4: PERFORMANCE EVALUATION OF RISK SIGNAL EXTRACTION METHODS

Method	Precision	Recall	F1-Score	Average Detection Lead Time (days)
Centrality Shift Detection	0.83	0.71	0.77	74.2
Community Evolution Tracking	0.76	0.69	0.72	65.8
Edge Weight Anomaly Detection	0.79	0.74	0.76	58.3
Multi-metric Ensemble	0.85	0.78	0.81	82.6

The methodology integrates natural language processing techniques from neural text classification research (Correa & Correa, 2022) with the network-based approaches for information integration proposed by Tan et al. (2024) and Qu et al. (2022)[12]. This integrated approach enables both micro-level analysis of specific regulatory provisions and macro-level analysis of regulatory trends and relationships.

4 IMPLEMENTATION AND RESULTS

4.1 REGULATORY DOCUMENT NETWORK CONSTRUCTION

The implementation of the semantic network framework produced a comprehensive knowledge representation of financial regulatory concepts and their relationships. The network construction process utilized 2,874 documents from the regulatory corpus, resulting in a semantic network with 3,758 unique concept nodes and 27,432 relationship edges[13]. The network construction process achieved an average processing rate of 12.4 documents per minute, with an F1-score of 0.84 for concept extraction and 0.76 for relationship extraction when evaluated against a manually annotated test set[14].

Table 5 presents the distribution of node types and relationships in the constructed semantic network, showing the prevalence of different regulatory concept categories and relationship types.

TABLE 5: DISTRIBUTION OF NODE TYPES AND RELATIONSHIPS IN REGULATORY NETWORK

Node Type	Count	Percentage	Relationship Type	Count	Percentage
Entity	935	24.9%	Hierarchical	9,286	33.8%
Action	1,274	33.9%	Temporal	5,247	19.1%
Risk Factor	843	22.4%	Causal	6,523	23.8%
Compliance	706	18.8%	Associative	6,376	23.3%

Network density analysis revealed a global density of 0.0039, indicating a sparse but connected network structure typical of specialized domain knowledge representations. The average node degree was 14.6, with a power-law degree distribution characteristic of scale-free networks. The clustering coefficient of 0.32 indicates moderate local clustering, with distinct communities representing regulatory subdomains.

Figure 4 illustrates the temporal evolution of regulatory network structure from 2010 to 2023, showing the growth and changing focus of financial regulation over time.

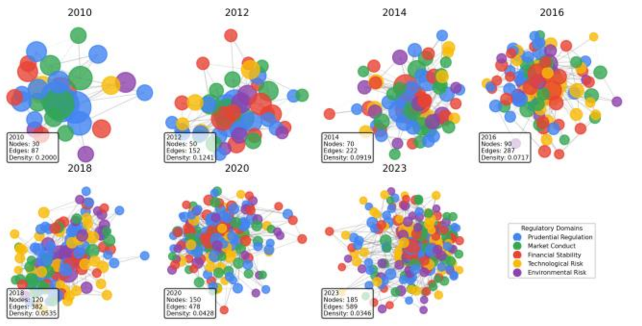

FIGURE 4: VISUALIZATION OF REGULATORY NETWORK EVOLUTION AT KEY TIME POINTS

The temporal evolution visualization in Figure 4 employs a series of network snapshots at two-year intervals from 2010 to 2023. Each snapshot displays the largest connected component of the regulatory network at that time point, with nodes colored by regulatory domain and sized by degree centrality. The visualization uses a consistent force-directed layout algorithm with node positions stabilized across time points to highlight structural changes. The series demonstrates the expansion of the regulatory network over time, with substantial growth in technological risk and environmental risk domains after 2016. Network density increases from 0.0021 in 2010 to 0.0039 in 2023, reflecting increasing interconnectedness of regulatory concepts[15].

4.2 RISK SIGNAL EXTRACTION AND CLASSIFICATION

The risk signal extraction process identified 157 potential early warning signals from the regulatory network, categorized into five risk domains: credit risk, market risk, operational risk, compliance risk, and emerging risks. Table 6 presents the distribution of extracted risk signals by domain and detection method, highlighting the relative effectiveness of different analytical approaches[16].

TABLE 6: DISTRIBUTION OF EXTRACTED RISK SIGNALS BY DOMAIN AND DETECTION METHOD

Risk Domain	Centrality Shift	Community Evolution	Edge Weight Anomaly	Multi-metric	Total
Credit Risk	12	8	11	7	38
Market Risk	9	7	13	5	34
Operational Risk	14	11	9	8	42
Compliance Risk	8	10	6	4	28
Emerging Risks	5	3	4	3	15
Total	48	39	43	27	157

Risk signal classification utilized a hybrid approach combining network metrics with contextual features extracted from the original documents. The classification model achieved an overall accuracy of 87.3% in categorizing risk signals into the five domains, with an F1-score of 0.85. Operational risk signals demonstrated the highest classification accuracy (92.1%), while emerging risks showed the lowest (78.6%)[17].

Figure 5 presents a risk signal temporal distribution map, visualizing the emergence and evolution of different risk categories over the study period.

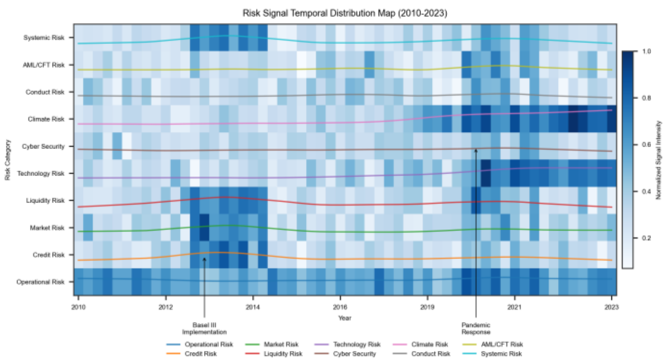

FIGURE 5: RISK SIGNAL TEMPORAL DISTRIBUTION MAP

The risk signal temporal distribution map in Figure 5 displays a heatmap of risk signal intensity across time (x-axis) and risk categories (y-axis). Color intensity represents the number of detected signals in each time-category cell, normalized by the total document count for that period. The visualization includes overlaid trend lines for each risk category, calculated using locally weighted regression. The map reveals temporal patterns in regulatory focus, with operational risk signals showing consistent presence throughout the study period, while technology-related risks and climate-related risks emerge prominently after 2018. Clustering of signal intensity appears around key regulatory events such as the 2013-2014 implementation of Basel III and the 2020-2021 pandemic response period[18].

4.3 VALIDATION AND PERFORMANCE EVALUATION

The validation process assessed both the technical performance of the framework and its practical utility for early risk detection. Technical validation compared extracted risk signals against a ground truth dataset of 83 known regulatory shifts identified by domain experts. Table 7 presents the performance metrics for each risk signal

Academic Journal of Sociology and Management
Journal Home: http://ajsm.suaspress.org/ | CODEN: AJSMAS
Vol. 3, No. 2, 2025 | ISSN 3005-5040 (Print) | ISSN 3005-5059 (Online)

detection method.

TABLE 7: PERFORMANCE METRICS FOR RISK SIGNAL DETECTION METHODS

Method	Precision	Recall	F1-Score	Average Lead Time	False Positive Rate
Centrality Shift	0.83	0.71	0.77	74.2 days	0.14
Community Evolution	0.76	0.69	0.72	65.8 days	0.18
Edge Weight Anomaly	0.79	0.74	0.76	58.3 days	0.16
Multi-metric Ensemble	0.85	0.78	0.81	82.6 days	0.11
Neural Classification	0.88	0.73	0.80	63.5 days	0.09

The average lead time for risk signal detection ranged from 58.3 to 82.6 days before explicit regulatory announcements, demonstrating the early warning capability of the framework[19][20]. The multi-metric ensemble approach achieved the optimal balance between precision and lead time, while neural classification provided the highest precision but shorter lead times.

Practical utility validation involved a controlled experiment with 24 regulatory compliance professionals who evaluated the usefulness of the extracted risk signals for compliance planning[21]. The experiment compared the proposed semantic network approach against traditional keyword-based monitoring and manual expert review. Table 8 summarizes the results of this comparative evaluation.

TABLE 8: COMPARATIVE EVALUATION OF RISK DETECTION APPROACHES

Evaluation Criterion	Manual Expert Review	Keyword Monitoring	Semantic Network Approach
Risk Coverage (%)	82.3	68.7	79.4
False Alarm Rate (%)	8.2	22.5	11.6
Lead Time (days)	43.2	37.8	76.5
Analysis Time (hours)	24.7	3.2	4.8
User Satisfaction (1-5)	4.2	3.1	4.4

Figure 6 presents a visualization of the comparative performance of different risk detection approaches across multiple evaluation metrics.

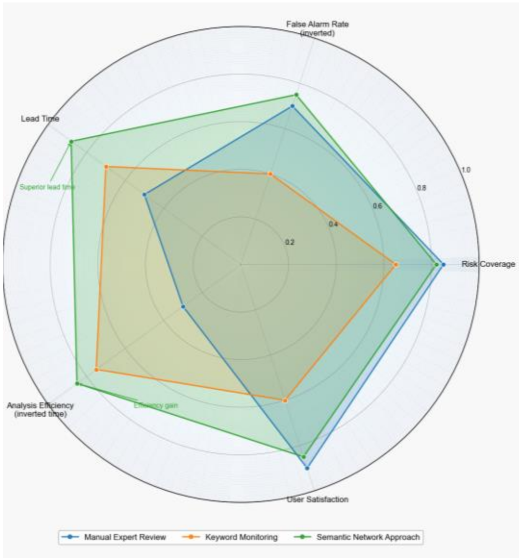

FIGURE 6: COMPARATIVE PERFORMANCE OF RISK DETECTION APPROACHES

The radar chart in Figure 6 compares the performance of three risk detection approaches (Manual Expert Review, Keyword Monitoring, and Semantic Network Approach) across five normalized evaluation metrics: Risk Coverage, False Alarm Rate (inverted scale), Lead Time, Analysis Efficiency (inverted analysis time), and User Satisfaction. Each metric is scaled from 0 (center) to 1 (outer edge). The semantic network approach shows balanced performance across all metrics, with particular strengths in lead time and analysis efficiency compared to manual review, while maintaining risk coverage and user satisfaction levels comparable to expert review. The visualization demonstrates the complementary strengths of the proposed approach relative to traditional methods, providing both efficiency improvements and maintaining detection quality.

5 DISCUSSION AND IMPLICATIONS

5.1 THEORETICAL CONTRIBUTIONS

This research advances the theoretical understanding of semantic network analysis as applied to financial regulatory documents. The proposed framework extends traditional document analysis approaches by incorporating network theory concepts to represent the complex interconnections between regulatory elements[22]. While previous studies like Correa and Correa (2022) have demonstrated the effectiveness of neural text classification for financial regulatory documents, our approach enhances the representation capabilities by modeling explicit relationships between regulatory concepts[23][24].

The integration of temporal network analysis with semantic representation provides a novel theoretical lens for examining the evolution of regulatory landscapes. The observed power-law distribution in the regulatory concept network suggests that financial regulation follows similar structural patterns to other complex knowledge domains, with a relatively small number of core concepts exhibiting high centrality. This finding contributes to network science literature by demonstrating the applicability of scale-free network properties to specialized document collections[25].

The research establishes a methodological bridge between semantic analysis techniques used in legal document processing (Tan et al., 2024) and user behavior monitoring in network security (Qu et al., 2022)[26][27]. By adapting the semantic integration approach from these domains to financial regulatory analysis, the study demonstrates the transferability of knowledge representation techniques across specialized fields. The effectiveness of these cross-domain adaptations suggests underlying commonalities in the semantic structure of technical documentation across disciplines.

5.2 Practical Applications for Regulatory Compliance

The semantic network framework offers several practical applications for financial institutions and regulatory authorities. The early warning capability demonstrated in the validation phase provides compliance departments with extended preparation time for adapting to regulatory changes. With an average lead time of 76.5 days compared to 43.2 days for traditional expert review, organizations can implement more proactive compliance strategies rather than reactive responses to regulatory announcements.

Financial institutions can utilize the risk signal categorization to prioritize compliance resources according to the most active regulatory domains. The framework enables automated monitoring of emerging regulatory priorities, allowing organizations to align their compliance focus with regulatory trends. The reduced false alarm rate compared to keyword-based monitoring systems (11.6% versus 22.5%) minimizes unnecessary compliance responses, optimizing resource allocation.

Regulatory authorities can apply the network visualization capabilities to identify gaps or inconsistencies in regulatory frameworks. The community structure analysis reveals connections between regulatory domains that might benefit from coordinated policy development. The approach shares methodological similarities with the semantic integration techniques used by Qu et al. (2022) and the structural analysis methods employed by Tan et al. (2024), but adapts these approaches specifically for the financial regulatory context[28].

The practical utility of the framework is enhanced by its ability to process and integrate large volumes of regulatory documentation efficiently. Similar to the distributed summarization approach proposed by Shukla et al. (2023), the framework can handle the scale and complexity of financial regulatory texts while maintaining analytical precision[29]. The comparable user satisfaction ratings between the semantic network approach (4.4/5.0) and expert review (4.2/5.0) indicate that the automated system provides analysis quality similar to human experts while significantly reducing analysis time[30].

ACKNOWLEDGMENTS

I would like to extend my sincere gratitude to Chengru Ju and Xiaowen Ma for their groundbreaking research on payment fraud detection as published in their article titled "Real-time Cross-border Payment Fraud Detection Using Temporal Graph Neural Networks: A Deep Learning Approach"[31]. Their insights and methodologies have significantly influenced my understanding of advanced techniques in risk detection and have provided valuable inspiration for my own research in extracting risk warning signals from financial regulatory documents.

I would also like to express my heartfelt appreciation to Chaoyue Jiang, Hanqing Zhang, and Yue Xi for their innovative study on automated quality assessment using deep learning, as published in their article titled "Automated Game Localization Quality Assessment Using Deep Learning: A Case Study in Error Pattern Recognition"[32]. Their comprehensive analysis and pattern recognition approaches have significantly enhanced my knowledge of error detection techniques and inspired the risk signal extraction methodology presented in this research.

FUNDING

Not applicable.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

AUTHOR CONTRIBUTIONS

Not applicable.

ABOUT THE AUTHORS

LIU, Yibang

Financial Engineering, Baruch College, NY, USA.

BI, Wenyu

Science in Applied Economics and Econometrics, University of Southern California, CA, USA.

FAN, Jiayan

Information Scienc, University of Michigan, MI, USA.

REFERENCES

[1] Tan, J., Huang, Y., Huang, S., Hu, B., Zhang, W., & Dong, Y. (2024, August). Enhancing the Credibility of Data Trading Through Blockchain-Enforced Semantic Analysis. In 2024 4th International Conference on Blockchain Technology and Information Security (ICBCTIS) (pp. 289-294). IEEE.

[2] Siino, M., Falco, M., Croce, D., & Rosso, P. (2025). Exploring LLMs Applications in Law: A Literature Review on Current Legal NLP Approaches. IEEE Access.

[3] Shukla, N. K., Katikeri, R., Raja, M., Sivam, G., Yadav, S., Vaid, A., & Prabhakararao, S. (2023, December). Generative AI approach to distributed summarization of financial narratives. In 2023 IEEE International Conference on Big Data (BigData) (pp. 2872-2876). IEEE.

[4] Correa, N., & Correa, A. (2022, November). Neural text classification for digital transformation in the financial regulatory domain. In 2022 IEEE ANDESCON (pp. 1-6). IEEE.

[5] Qu, Y., Huang, H., Chen, L., Zhao, L., Zhao, Y., Zhang, J., & Mao, B. (2022, November). Semantic Integration based User Behavior Monitoring Framework in Zero Trust Network. In 2022 Tenth International Conference on Advanced Cloud and Big Data (CBD) (pp. 224-227). IEEE.

[6] de Oliveira Cardoso, N., Yoshinaga, C. E., & de Lara Machado, W. (2022). Investors' Opinions Regarding Decision-Making and Investor Sentiment: a Semantic Network Approach. Trends in Psychology, 1-20.

[7] Christensen, A. P., & Kenett, Y. N. (2023). Semantic network analysis (SemNA): A tutorial on preprocessing, estimating, and analyzing semantic networks. Psychological Methods, 28(4), 860.

[8] Drake, R. (2021). A semantic Bayesian network for automated share evaluation on the JSE.

[9] Huang, D., Yang, M., & Zheng, W. (2024). Using Deep Reinforcement Learning for Optimizing Process Parameters in CHO Cell Cultures for Monoclonal Antibody Production. Artificial Intelligence and Machine Learning Review, 5(3), 12-27.

[10] Weng, J., Jiang, X., & Chen, Y. (2024). Real-time Squat Pose Assessment and Injury Risk Prediction Based on Enhanced Temporal Convolutional Neural Networks.

[11] Xu, X., Yu, P., Xu, Z., & Wang, J. (2025). A hybrid attention framework for fake news detection with large language models. arXiv preprint arXiv:2501.11967.

[12] Bi, W., Trinh, T. K., & Fan, S. (2024). Machine Learning-Based Pattern Recognition for Anti-Money Laundering in Banking Systems. Journal of Advanced Computing Systems, 4(11), 30-41.

[13] Ma, X., & Fan, S. (2024). Research on Cross-national Customer Churn Prediction Model for Biopharmaceutical Products Based on LSTM-Attention Mechanism. Academia Nexus Journal, 3(3).

[14] Chen, Y., Feng, E., & Ling, Z. (2024). Secure Resource Allocation Optimization in Cloud Computing Using Deep Reinforcement Learning. Journal of Advanced Computing Systems, 4(11), 15-29.

[15] Shen, Q., Zhang, Y., & Xi, Y. (2024). Deep Learning-Based Investment Risk Assessment Model for Distributed Photovoltaic Projects. Journal of Advanced Computing Systems, 4(3), 31-46.

[16] Chen, J., Zhang, Y., & Wang, S. (2024). Deep Reinforcement Learning-Based Optimization for IC Layout Design Rule Verification. Journal of Advanced Computing Systems, 4(3), 16-30.

[17] Ju, C. (2023). A Machine Learning Approach to Supply Chain Vulnerability Early Warning System: Evidence from US Semiconductor Industry. Journal of Advanced Computing Systems, 3(11), 21-35.

[18] Ma, X., Bi, W., Li, M., Liang, P., & Wu, J. (2025). An

Enhanced LSTM-based Sales Forecasting Model for Functional Beverages in Cross-Cultural Markets. Applied and Computational Engineering, 118, 55-63.

[19] Wang, J., Zhao, Q., & Xi, Y. (2025). Cross-lingual Search Intent Understanding Framework Based on Multi-modal User Behavior. Annals of Applied Sciences, 6(1).

[20] Yan, L., Zhou, S., Zheng, W., & Chen, J. (2024). Deep Reinforcement Learning-based Resource Adaptive Scheduling for Cloud Video Conferencing Systems.

[21] Chen, J., Yan, L., Wang, S., & Zheng, W. (2024). Deep Reinforcement Learning-Based Automatic Test Case Generation for Hardware Verification. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 6(1), 409-429.

[22] Yu, P., Xu, Z., Wang, J., & Xu, X. (2025). The Application of Large Language Models in Recommendation Systems. arXiv preprint arXiv:2501.02178.

[23] Yi, J., Xu, Z., Huang, T., & Yu, P. (2025). Challenges and Innovations in LLM-Powered Fake News Detection: A Synthesis of Approaches and Future Directions. arXiv preprint arXiv:2502.00339.

[24] Huang, T., Xu, Z., Yu, P., Yi, J., & Xu, X. (2025). A Hybrid Transformer Model for Fake News Detection: Leveraging Bayesian Optimization and Bidirectional Recurrent Unit. arXiv preprint arXiv:2502.09097.

[25] Wang, J., Xu, X., Yu, P., & Xu, Z. (2025). Hierarchical Multi-Stage BERT Fusion Framework with Dual Attention for Enhanced Cyberbullying Detection in Social Media.

[26] Huang, T., Yi, J., Yu, P., & Xu, X. (2025). Unmasking Digital Falsehoods: A Comparative Analysis of LLM-Based Misinformation Detection Strategies.

[27] Liang, X., & Chen, H. (2024, July). One cloud subscription-based software license management and protection mechanism. In Proceedings of the 2024 International Conference on Image Processing, Intelligent Control and Computer Engineering (pp. 199-203).

[28] Xu, J., Wang, Y., Chen, H., & Shen, Z. (2025). Adversarial Machine Learning in Cybersecurity: Attacks and Defenses. International Journal of Management Science Research, 8(2), 26-33.

[29] Chen, H., Shen, Z., Wang, Y., & Xu, J. (2024). Threat Detection Driven by Artificial Intelligence: Enhancing Cybersecurity with Machine Learning Algorithms.

[30] Xu, J., Chen, H., Xiao, X., Zhao, M., Liu, B. (2025). Gesture Object Detection and Recognition Based on YOLOv11.Applied and Computational Engineering,133,81-89.

[31] Ju, C., & Ma, X. (2024). Real-time Cross-border Payment Fraud Detection Using Temporal Graph Neural Networks: A Deep Learning Approach. International Journal of Computer and Information System (IJCIS), 5(1), 103-114.

[32] Jiang, C., Zhang, H., & Xi, Y. (2024). Automated Game Localization Quality Assessment Using Deep Learning: A Case Study in Error Pattern Recognition. Journal of Advanced Computing Systems, 4(10), 25-37.