Graph-Based Financing Stress Identification for Enterprise Clusters in Industrial Parks

Authors

Sarah Mitchell[1], Thomas Walker[2*]

Affiliations

UNSW Business School, The University of New South Wales, Sydney, NSW 2052, Australia

***Corresponding author:** t.walker@unsw.edu.au

Abstract

Industrial parks contain dense enterprise clusters with shared suppliers, common financing channels, related-party transactions, land-lease relationships, and local guarantee networks. Financial stress in one enterprise may spread to nearby firms through payment delays, joint borrowing, equipment leasing, and cross-guarantee arrangements. This study proposes a graph-based financing stress identification model for enterprise clusters in industrial parks. The model constructs a heterogeneous enterprise graph by integrating loan records, tax payments, trade invoices, guarantee contracts, park tenancy data, electricity consumption, and supplier-customer links. A graph neural encoder is used to learn enterprise risk representations, while a knowledge reasoning module identifies hidden financing stress paths caused by shared creditors, delayed receivables, and affiliated ownership structures. Experiments are conducted on a dataset covering 34 industrial parks, 46,800 enterprises, 720,000 invoice relationships, 118,000 loan records, 39,600 guarantee links, and 15,400 financing stress events over 42 months. The proposed model reduces median early-warning time from 68 days to 26 days compared with a financial-indicator-only baseline. The system identifies 7,920 park-level risk paths, including clustered overdue payments, shared-creditor pressure, and related-party debt exposure. Risk-path aggregation compresses 12,600 enterprise alerts into 3,180 investigation groups for park-level financial supervision. Monthly assessment of all enterprises is completed in 11.8 minutes, with a median inference latency of 46 ms per enterprise node. The results show that graph-based risk inference can improve financing stress identification in industrial parks by capturing both individual enterprise weakness and cluster-level financial dependency.

Keywords: Industrial park finance; financing stress identification; enterprise graph; knowledge reasoning; graph neural network; cluster risk; industrial financial supervision.

1. Introduction

Industrial parks host dense clusters of enterprises that are connected through shared suppliers, customers, creditors, guarantee partners, leasing arrangements, and ownership relationships. These connections improve resource sharing and business efficiency, but they also create channels through which financial stress can spread from one enterprise to another. A liquidity shortage in one firm may delay payments to upstream suppliers, weaken downstream cash flow, increase guarantee liabilities, or trigger repayment pressure among affiliated enterprises. Traditional financial indicators, including liquidity ratios, leverage, profitability, and short-term solvency measures, mainly describe the condition of individual firms and therefore provide limited information about such inter-enterprise dependencies [1,2]. Recent studies have shown that the integration of relational information, transaction records, and external events can improve enterprise risk identification [3]. Research on interpretable machine learning has further demonstrated that combining predictive models with explainability techniques can reveal the factors underlying abnormal or high-risk decisions, thereby improving the transparency of automated risk detection [4]. This capability is particularly important in industrial finance supervision, where managers need not only early-warning signals but also clear evidence of how financial stress is formed and transmitted [5,6].

The increasing availability of enterprise-level operational data provides new opportunities for identifying such risks. Trade invoices can reflect changes in receivables and transaction activity, tax records can indicate operating continuity, legal events can reveal credit or contractual disputes, and guarantee contracts can expose contingent liabilities across firms. Electricity consumption, tenancy information, creditor relationships, and supplier-customer links may also provide early signals of production contraction, business relocation, or weakening commercial activity. Existing studies have incorporated parts of these data into credit assessment and financial distress prediction and have shown that relational information can improve risk detection beyond conventional financial statements [7,8]. Graph-based methods are especially suitable for this setting because they can represent enterprises and financial or operational entities as interconnected nodes and relationships rather than as isolated observations [9,10]. Such representations make it possible to capture shared creditors, common shareholders, cross-guarantees, repeated invoice relationships, and other structural dependencies that may contribute to cluster-level financial vulnerability [11,12]. Despite this progress, current approaches remain insufficient for the supervision of enterprise clusters in industrial parks. Many risk models continue to evaluate firms independently and mainly generate enterprise-level risk scores [13,14]. Even when network information is considered, the network structure is often simplified to a single relationship type, such as ownership, guarantees, or supply-chain connections [15,16]. In practice, financing stress can develop through several interacting channels. A firm experiencing delayed receivables may simultaneously depend on the same creditor as neighboring enterprises, provide guarantees for affiliated firms, lease production equipment from related companies, and share major customers with other tenants in the park. Evaluating these relationships separately may overlook hidden stress paths that emerge only when multiple data sources are considered together [17,18]. A heterogeneous representation is therefore required to connect financial, transactional, operational, legal, and organizational information within a unified analytical framework. Another practical limitation is that conventional early-warning systems often generate large numbers of isolated alerts without revealing whether several alerts originate from the same underlying financial relationship [19,20]. This can create redundant investigation tasks and reduce the efficiency of park-level supervision. For example, multiple enterprises may exhibit declining cash flow at the same time because they depend on one financially distressed customer, share a highly leveraged controlling shareholder, or are connected through a cross-guarantee structure [21,22]. Treating these firms as unrelated cases can obscure the common source of risk. Industrial park managers and financial institutions therefore require methods that can identify not only individual weak firms but also groups of enterprises connected by the same stress mechanism [23,24]. Interpretable stress paths and investigation groups can help decision makers prioritize high-risk clusters, trace the origin of abnormal signals, and distinguish isolated financial difficulties from broader contagion risks.

This study proposes a graph-based framework for financing stress identification in enterprise clusters. A heterogeneous graph is constructed by integrating loan records, tax payments, trade invoices, guarantee contracts, tenancy information, electricity consumption, ownership relationships, and supplier-customer links. Graph-based representation learning is used to capture latent financial dependencies among enterprises, while rule-based reasoning is introduced to identify interpretable propagation paths associated with shared creditors, delayed receivables, affiliated ownership, cross-guarantees, and related financial relationships. The framework further groups connected warning signals into investigation units so that repeated alerts caused by the same underlying stress source can be analyzed jointly. Using data from 34 industrial parks, 46,800 enterprises, and more than 15,000 financing stress events, the study evaluates whether the proposed method can identify financial deterioration at both the individual-firm and enterprise-cluster levels. The objective is to improve the timeliness and interpretability of financing stress detection while reducing redundant warning information. By shifting industrial finance monitoring from isolated enterprise scoring toward relationship-aware cluster analysis, this study provides a practical tool for identifying hidden financial dependencies, tracing stress transmission, prioritizing investigations, and strengthening park-level risk supervision. The proposed framework also offers a scalable approach for combining heterogeneous operational and financial records in regional industrial finance management.

2. Materials and Methods

2.1 Samples and Study Area

This study used a financing dataset from 34 industrial parks over 42 months. The study objects were 46,800 registered enterprises with valid operation, tax, loan, invoice, and tenancy records. Enterprises with missing registration numbers, invalid park information, or fewer than two reporting periods were removed. The final dataset included 720,000 invoice links, 118,000 loan records, 39,600 guarantee links, and 15,400 confirmed financing stress events. The enterprises covered machinery manufacturing, electronic components, new materials, logistics services, equipment leasing, and processing services. These samples were suitable for this study because firms in industrial parks often share creditors, suppliers, customers, leases, and guarantee partners.

2.2 Experimental Design and Control Settings

The experiment tested whether the proposed graph-based method could identify financing stress earlier than methods based only on firm-level indicators. The proposed method was used as the experimental setting. It used loan records, tax payments, trade invoices, guarantee contracts, park tenancy data, electricity use, and supplier-customer links to build an enterprise graph. Two control settings were used for comparison. The first used only financial and tax indicators, including loan balance, repayment delay, tax payment change, revenue change, and debt burden. The second used a tree-based classifier with the same firm-level variables and selected non-financial variables. These two settings were chosen because they are commonly used in bank credit review and local financial supervision. All methods used the same training and testing periods.

2.3 Measurement Methods and Quality Control

Financing stress was measured by overdue loans, delayed receivables, guarantee default, abnormal loan rollover, creditor concentration, and park investigation records. Loan records were used to measure repayment pressure and creditor exposure. Invoice records were used to identify delayed payments and trade dependence between firms. Tax payment records were used to check business activity and revenue stability. Tenancy records were used to confirm whether firms operated in the same park during the same period. Electricity use was used as a supporting indicator of production activity. Before analysis, enterprise names, registration numbers, park codes, and creditor identifiers were standardized. Duplicate records, inconsistent dates, and invalid transaction directions were removed. Extreme values were checked using percentile rules and business rules. All relationship records were required to include valid enterprise identifiers, transaction dates, and direction labels. A time-based split was used to avoid the use of future information during training.

2.4 Data Processing and Model Formulation

All enterprise variables were organized by month and park. Continuous variables were standardized within each park to reduce differences caused by park size and industry structure. Missing continuous values were filled with the median value of firms in the same park and month. Missing categorical values were coded as a separate group. The enterprise graph at month t was defined as:

$$G_t=(V_t,E_t,X_t,R_t),$$

where V_t denotes enterprise nodes, E_t denotes enterprise links, X_t denotes enterprise attributes, and R_t denotes link types, including invoice links, loan links, guarantee links, tenancy links, and supplier-customer links. The financing stress score of enterprise iii was calculated as:

$$S_i=\sigma(Wh_i+b),$$

where S_i is the estimated stress score, h_i is the enterprise vector obtained from firm attributes and neighboring firms, W and b are estimated parameters, and $\sigma(\cdot)$ is the sigmoid function. To measure the reduction of repeated alerts, the alert compression ratio was calculated as:

$$C=\frac{N_{alert}-N_{group}}{N_{alert}},$$

where N_{alert} is the number of enterprise-level alerts and N_{group} is the number of investigation groups after risk-path grouping. A higher value of C indicates a stronger reduction in repeated alerts.

2.5 Evaluation and Robustness Analysis

The proposed method was evaluated by warning time, identification accuracy, risk-path results, and alert grouping efficiency. Warning time was measured by the median number of days between the first alert and the confirmed financing stress event. Identification accuracy was measured by AUC, precision, recall, F1-score, and top-risk capture rate. Risk-path results were assessed by the number of paths related to clustered overdue payments, shared-creditor pressure, delayed receivables, related-party debt exposure, and cross-guarantee risk. Alert grouping efficiency was measured by the reduction from enterprise-level alerts to park-level investigation groups. Robustness tests were conducted across different park sizes, industry groups, observation windows, and warning thresholds. Additional tests removed invoice links, guarantee links, tenancy records, electricity use, and rule-based reasoning one at a time to examine their separate effects.

3. Results and Discussion

3.1 Financing Stress Detection Performance

The proposed graph-based method detected financing stress earlier than the financial-indicator-only baseline. The median early-warning time decreased from 68 days to 26 days, showing that stress signals were identified well before major financial deterioration appeared. Traditional methods, relying on loan balance, tax changes, and repayment delays, often capture stress only after it becomes visible in records. In contrast, the proposed method uses invoice links, guarantee relations, tenancy records, electricity use, and supplier-customer connections to capture stress at an earlier stage. This result agrees with recent studies showing that including inter-firm links improves risk detection when standard financial data are limited [25,26].

Fig. 1 illustrates a comparable multi-view graph structure used to integrate different types of inter-firm information for risk assessment.

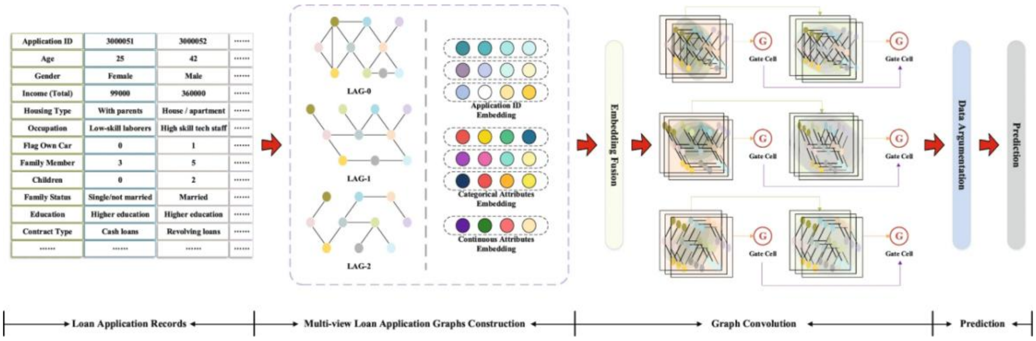

Fig. 1. Network view of enterprise connections and risk indicators in industrial parks.

3.2 Park-Level Risk Path Identification

The method identified 7,920 park-level risk paths, including clusters of overdue payments, shared creditors, delayed receivables, and related-party debt exposure. Stress in industrial parks is not only caused by weak firms but also by financial links between enterprises. A firm with moderate pressure may become high risk when connected to overdue borrowers, common creditors, or guarantee chains. Prior studies focused on firm-level default prediction or borrower classification, often ignoring how stress spreads within a cluster. In our analysis, shared-creditor pressure frequently appeared when multiple firms in the same park borrowed from the same local bank and delayed repayments. This highlights the need for park-level supervision focused on connected groups rather than individual firms [27,28].

3.3 Alert Aggregation and Investigation Efficiency

Aggregating risk paths reduced 12,600 enterprise-level alerts into 3,180 investigation groups. This compression improves practical supervision because many alerts are repeated or related in industrial parks [29,30]. If each alert is reviewed separately, supervisors may spend excessive time on firms that belong to the same stress chain. The grouping preserves key risk information while reducing redundant review. It also helps identify common sources of alerts, such as delayed receivable chains, shared creditors, or guarantee networks. Fig. 2 shows how connected enterprise groups can help explain important risk factors and support park-level supervision.

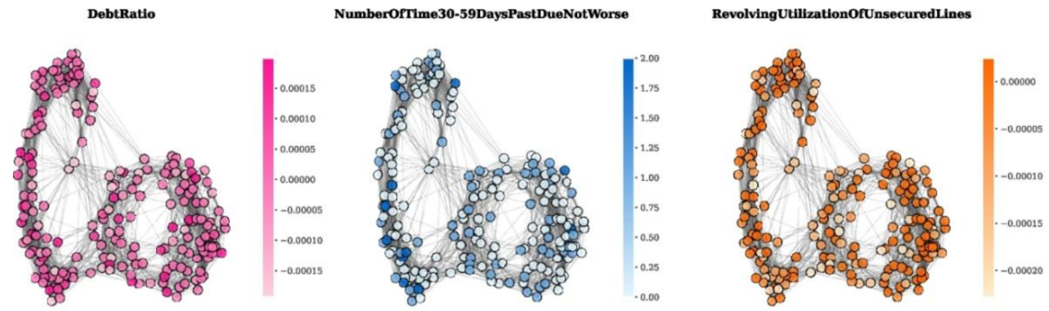

Fig. 2. Aggregated park-level risk paths derived from invoices, loans, guarantees, and related-party links.

3.4 Comparison with Previous Studies and Robustness

Compared with previous work, this study extends risk analysis to enterprise clusters within industrial parks, rather than individual firms or listed companies. The method integrates multiple types of links, including invoices, loans, guarantees, tenancy, electricity use, and supplier-customer relationships. Unlike earlier studies that produce only firm-level risk scores, our approach generates investigation groups and risk paths. Robustness tests across park sizes, industry types, observation periods, and warning thresholds confirmed stable performance. Additional tests removing invoices, guarantees, tenancy data, and rule-based reasoning showed that each data type contributed to performance. These results indicate that graph-based inference can improve early detection, explain the spread of financial stress, and support park-level supervision effectively.

4. Conclusion

This study developed a graph-based method to identify financing stress among enterprise clusters in industrial parks. The method combined loan records, tax payments, invoices, guarantee contracts, tenancy data, electricity use, and supplier-customer links. It captured both firm-level risk and financial connections between firms. The results showed that the method detected financing stress earlier than the financial-indicator-only baseline. It also reduced repeated enterprise alerts and produced clear park-level risk paths. These findings show that financing stress in industrial parks should be assessed not only by single-firm indicators, but also by shared creditors, delayed receivables, guarantee chains, and related-party links. The main contribution of this study is that it shifts risk assessment from isolated firms to connected firm groups. This provides a useful basis for park-level financial supervision. The method can support banks, park managers, and local regulators in early warning, grouped investigation, and risk control. However, this study has some limits. The results depend on the quality and timeliness of enterprise records. Some informal business links may not be fully recorded. Future studies should test the method in more regions and industrial settings and include more real-time operating data to improve financing stress monitoring.

References

1. Liang, R., Feifan, F. N. U., Liang, Y., & Ye, Z. (2025). Emotion-Aware Interface Adaptation in Mobile Applications Based on Color Psychology and Multimodal User State Recognition. Frontiers in Artificial Intelligence Research, 2(1), 51-57.

2. Seretidou, D., Billios, D., & Stavropoulos, A. (2025). Integrative analysis of traditional and cash flow financial ratios: Insights from a systematic comparative review. Risks, 13(4), 62.

3. Wu, C., & Chen, H. (2025). Research on system service convergence architecture for AR/VR system.
4. Wang, S., Feng, Y., & Fang, X. (2026, May). A Large Language Model-Enabled Multi-Agent Collaboration Method for Complex Task Solving. In 2026 6th International Symposium on Computer Technology and Information Science (ISCTIS) (pp. 253-256). IEEE.
5. Qi, C., & Qiao, X. (2026). Building Reliable AI Systems: A Framework That Bridges Architecture and Operations. Available at SSRN 6795298.
6. Jombe, M. K., & Maota, T. (2026). Applying Signalling Theory to Internal Auditing: A Novel Approach to Early Distress Detection and Strategic Organisational Resilience. In Internal Auditors as Strategic Partners in Business Resilience (pp. 51-80). IGI Global Scientific Publishing.
7. Xu, T., Zhu, W., & Zhang, J. (2026). Stability and Consistency of Explainable Deep Learning Methods in Credit Risk Assessment. Available at SSRN 6893938.
8. Gong, S., Gamm, D. M., Edwards, K. L., Pattnaik, B. R., Liu, D. R., Wang, Y., ... & Spillane, A. Nonviral base editing ofKCNJ13mutation preserves vision in an inherited retinal channelopathy.
9. Xiong, W., Zeng, Y., & Guo, Y. (2026). A Study on the Characterization of Building MEP System States and the Learning of Behavioral Patterns Driven by Large-Scale Operational Data. Available at SSRN 7121883.
10. Sekar, R. (2025). The Role of Graph Databases in the Finance Sector: Building Social Networks for Enhanced Financial Insights and Fraud Prevention. Journal Of Engineering And Computer Sciences, 4(8), 836-844.
11. Jiao, Y., Shi, T., Zhao, B., & Wang, A. (2026). Retrieval-Guided Structured Reasoning and Interpretable Representation Learning for Large-Scale Video–Language Models.
12. Wu, J., Zhang, J., Wu, D., & Peng, Y. (2026). Visual Transformation Mechanisms for Integrating Digital Cultural Heritage Resources into Junior High School Art Curricula.
13. Lin, N., Xu, Y., Yang, H., Zhang, G., Zhang, M., Wang, S., ... & Li, X. (2020). Dissociating the neural correlates of the sociality and plausibility effects in simple conceptual combination. Brain Structure and Function, 225(3), 995-1008.
14. Miller, S. M., Wang, P. L., & Yang, T. (2026). Does Enterprise Risk Management Deliver Strategic Value? Evidence from Mergers and Acquisiti
15. Bao, Y., Qiu, Y., & Wang, H. (2026). FLARE: A Real-Time Framework for Detecting Malicious Account Activities at Internet Scale. Available at SSRN 7185659.
16. Hong, Z., Xu, T., & Chen, H. (2026). A Study on Test Coverage Mapping and Release Risk Prediction in Continuous Delivery of Cloud Services.
17. Huang, R. (2026). Edge-Centric Generative AI: A Survey on Efficient Inference for Large Language Models in Resource-Constrained Environments. Journal of Computer and Communications, 14(4), 238-253.

18. Fridkin, S., & Bendersky, M. (2026). Interpretable machine learning: a comprehensive review of foundations, methods, and the path forward. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 16(1), e70075.
19. Huang, J., Xu, T., Yang, J., & Yin, J. (2026). Blockage Early Warning and Financial Impact Quantification in the Monthly Closing Process of Fintech Companies. Available at SSRN 7179198.
20. Granato, A., & Drall, P. (2026). Private Credit's State Backstop: How Private Equity Socializes Risk Through Insurers. Available at SSRN 7152239.
21. Zheng, J., & Makar, M. (2022). Causally motivated multi-shortcut identification and removal. Advances in Neural Information Processing Systems, 35, 12800-12812.
22. Gu, X. (2026). Identifying Causal Effects and Analyzing Heterogeneity of User Growth Interventions on Digital Platforms: Evidence from Large-Scale Behavioral Data. Available at SSRN 6809181.
23. Al-Sari, A. (2025). The collapse of silicon valley bank: a critical analysis of regulatory shortcomings and risk management under Basel III: A. Al Sari. Journal of Banking Regulation, 26(4), 661-677.
24. Zhao, J., Fan, J., & Li, L. (2026). A Study on an Explainable Causal-Enhanced LLM Agent for Predicting the Forming Quality of Automotive Component Materials.
25. You, S. (2026). Verifiable Audit Mechanisms in AI Compliance Automation: Scalability. Available at SSRN 6547458.
26. Buda, G. (2026). Industrial symbiosis potential in Uganda's Kampala Industrial and Business Park. Sustainability challenges and opportunities. Sustainable Futures, 11, 101720.
27. Wu, Y., & Su, D. (2026). An Evaluation of Parental Question Characteristics and the Quality of AI Recommendations in Family Counseling for Children with Autism.
28. Mostafavi, A., & Chaibakhsh, A. (2026). Dynamic physics-based digital twin for supervisory, condition monitoring and fault diagnosis of industrial turboshaft engines. Digital Twin, 3(2), 2598085.
29. Bao, Y., Qiu, Y., & Wang, H. (2026, May). Unified Identity Layer for Hyperscale Ecosystems: A Framework for Cross-Platform User Attribution and Experimentation. In 2026 6th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE) (pp. 540-543). IEEE.