Journal of Civil Engineering and Urban Planning (2025) DOI: 10.23977/jceup.2025.070306
Clausius Scientific Press, Canada ISSN 2616-3969 Vol. 7 Num. 3

Study on Risk Identification and Prediction Mechanism of Major Engineering Group Events in Complex Social Environment

Chunyi Yu[1,a,*], Yiran Lv[1,b]

[1]*School of Intelligent Construction, Chongqing Jianzhu College, Chongqing, 400072, China*
[a]*yuchunyi@cqjzc.edu.cn*, [b]*278171026@qq.com*
**Corresponding author*

Keywords: Risk Identification, Prediction Mechanism, Major Engineering Group Events, Complex Social Environment, Social Network Analysis, System Dynamics, Agent-Based Modelling

Abstract: Under the complex social environment of accelerating urbanization and deep adjustment of interest pattern, mass incidents of major projects are on the rise, and the traditional static risk assessment model is difficult to meet the needs of dynamic risk evolution. In this study, the whole chain risk identification and prediction mechanism is constructed, which integrates social network analysis (SNA), natural language processing (NLP), system dynamics (SD) and agent-based modeling (ABM), and the interaction network of stakeholders is deconstructed by SNA to identify key influencers and vulnerable nodes. By tracking the emotional evolution and topic changes of public opinion through NLP, this study aims to capture early risk signals and predict potential issues. The SD-ABM fusion model is constructed, which simulates the accumulation of public dissatisfaction and policy response feedback at the macro level, and depicts the herd effect and social psychological amplification mechanism of individual protest decision-making at the micro level, forming a "recognition-early warning-response" closed loop. Taking the PX chemical project in X city as a case, the empirical study based on 25,443 pieces of social media data and 200 household questionnaires shows that the model accurately predicts the incident probability (the baseline scenario reaches 68%), and effectively evaluates the differentiated effects of intervention strategies such as popular science propaganda and economic compensation. The research verifies the ability of the fusion model to capture nonlinear risk evolution and sudden disturbance, and provides a forward-looking decision support tool for the social risk management of major projects from passive response to active prediction.

1. Introduction

As the strategic support of national economic development, major projects are often accompanied by significant social externalities. In recent years, with the acceleration of urbanization and the deep adjustment of social interest pattern, mass incidents caused by major projects are on the rise. This kind of incident not only causes project stagnation and economic losses, but also may lead to systemic social risks, threatening public safety and government credibility [1]. The superposition of complex social environment further magnifies the uncertainty of risk transmission [2]. At present, the society is characterized by diversification of interest subjects, fragmentation of information dissemination and easy intensification of social emotions. The traditional prevention and control model based on single interest appeal or static risk assessment has been difficult to adapt to the dynamic and complex social environment, and it is urgent to build a more forward-looking and systematic risk management framework [3].

Most studies rely on the qualitative induction of historical cases or the quantitative analysis of a single dimension, lacking the dynamic description of social complexity such as stakeholder interaction network and information dissemination mechanism [4]; Most risk prediction models adopt statistical regression or traditional machine learning methods, which makes it difficult to capture the influence of nonlinear relationship and sudden disturbance on risk evolution. The results of risk identification and prediction often lag behind the development of events and do not form a closed loop with intervention strategies, which leads to the break of the "identification-early warning-response" chain [5].

The purpose of this study is to break through the "linear thinking" trap of traditional risk management, build a full chain risk management mechanism for major engineering group events in a complex social environment, and deconstruct the interaction mode of stakeholders and the evolution path of public opinion and emotion through social network analysis (SNA) and natural language processing (NLP) technologies; Combining system dynamics (SD) and agent-based modeling (ABM), a dynamic prediction framework considering social psychological amplification effect is constructed.

2. Risk Identification Mechanism

2.1. Deconstruction of Stakeholder Interaction

Risk identification is the basic link, focusing on extracting risk signals from structured (relational network) and unstructured (text data) data [6]. SNA quantifies the characteristics of network structure and identifies key influencers and vulnerabilities by constructing stakeholder networks involved in major projects (nodes represent governments, contractors, residents, NGOs, etc., while edges represent cooperative relations, conflict relations or information flow).

Degree centrality measures the number of direct connections of nodes and identifies the role of "hub". The formula is:

$$C_D(v) = \frac{\deg(v)}{n-1} \tag{1}$$

Where $\deg(v)$ represents the degree of node v (number of connected edges) and n represents the total number of network nodes. The larger the value, the more central the node is in information dissemination, but its failure may lead to chain risks.

Median centrality evaluates the ability of nodes to control resource flow and identifies the role of "bridge". The formula is:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{2}$$

Where σ_{st} represents the shortest path number from node s to t, and $\sigma_{st}(v)$ is the shortest path number passing through v. A high value indicates that the node is the key to conflict coordination, but its blocking may aggravate group opposition.

2.2. Analysis on the Evolution of Public Opinion and Emotion

NLP extracts texts from social media and news comments, carries out emotional analysis and theme modeling, tracks public mood fluctuation and topic evolution, and finds protest signals early [7].

The emotional score calculation is based on BERT, which quantifies the emotional polarity of the text [8]. The formula is:

$$S_{doc} = \frac{1}{N}\sum_{i=1}^{N} polarity(w_i) \tag{3}$$

Where S_{doc} represents the emotional score of the document (ranging from -1 to 1), N is the number of words, and $polarity(w_i)$ is the emotional value of the word w_i (negative value means negative). A continuous score below the threshold indicates risk accumulation.

The topic evolution model uses LDA (Latent Dirichlet Allocation) to dynamically track the topic changes. The probability of topic generation is:

$$P(topic|doc) = \sum_{w} P(topic|w)P(w|doc) \tag{4}$$

The abrupt change of subject probability distribution can be regarded as a precursor of risk.

3. Risk Prediction Mechanism

3.1. Macro-risk Transmission of SD Modeling

The measurement mechanism focuses on simulating the risk evolution path, portrays the macro-system feedback through SD, simulates the micro-subject behavior through ABM, and then integrates the two to capture the social psychological amplification effect.

SD defines variables such as public dissatisfaction, project progress and government response efficiency through causal cycle diagram, describes feedback loop with differential equation, and simulates risk accumulation threshold. Dynamic equation of public dissatisfaction is shown in below:

$$\frac{dM(t)}{dt} = \alpha I(t) - \beta R(t) - \gamma M(t) \tag{5}$$

Among them, $M(t)$ is the public dissatisfaction at t moment (0-1 standardized value), $I(t)$ is the event impact, $R(t)$ is the effectiveness of government mitigation measures (0-1 value), and α,β,γ is the weight parameter. α represents shock sensitivity (from NLP emotional score), β represents policy response coefficient, and γ represents natural attenuation rate (reflecting social forgetting effect). When $M(t)$ exceeds the threshold, the probability of triggering group events rises.

3.2. ABM Simulates Microscopic Interaction and Amplification Effect

ABM defines heterogeneous subjects, endows them with attributes (risk perception, conformity) and behavior rules (protest decision), and simulates the emergence of risks caused by local interaction through NetLogo platform.

The protest probability function of the subject is based on the theoretical planning behavior (TPB), and the protest probability of the resident subject i is defined as:

$$P_i = \frac{1}{1+e^{-(aS_i + bI_i + cN_i)}} \tag{6}$$

Among them, S_i is individual satisfaction (dissatisfaction mapping from SD), I_i is information exposure, N_i is neighbor influence (neighbor protest ratio), and a,b,c is weight parameter. The value is calibrated by survey data, and the conformity effect is reflected when c is high.

The macro variables of SD are taken as the environmental input of ABM, and the micro results of ABM are fed back to adjust the SD parameters, forming a closed loop. For example, the data is updated once a month, and the NLP emotional score drives SD's I_i and ABM to simulate the output risk probability as an early warning signal.

4. Empirical Analysis and Result Discussion

4.1. Data Source and Preprocessing

Take the "PX Chemical Project in X City" which has caused widespread social controversy as a case. Through the collection of multivariate data in the early stage of the project (from the project publicity to the outbreak of large-scale protests), an empirical analysis is made. The time span is 6 months (T1-T6). Two networks, namely "information exchange" and "trust relationship", have been constructed through a questionnaire survey of 200 residents in five communities around the project. There are 25,443 posts and comments with keywords such as "PX Project" and "X Chemical Industry" crawled from the local forum "X Post Bar" and Sina Weibo. Project and policy data include project publicity, EIA report, government press conference draft, etc. All data have been cleaned, denoised and standardized.

4.2. Risk Identification Result Score

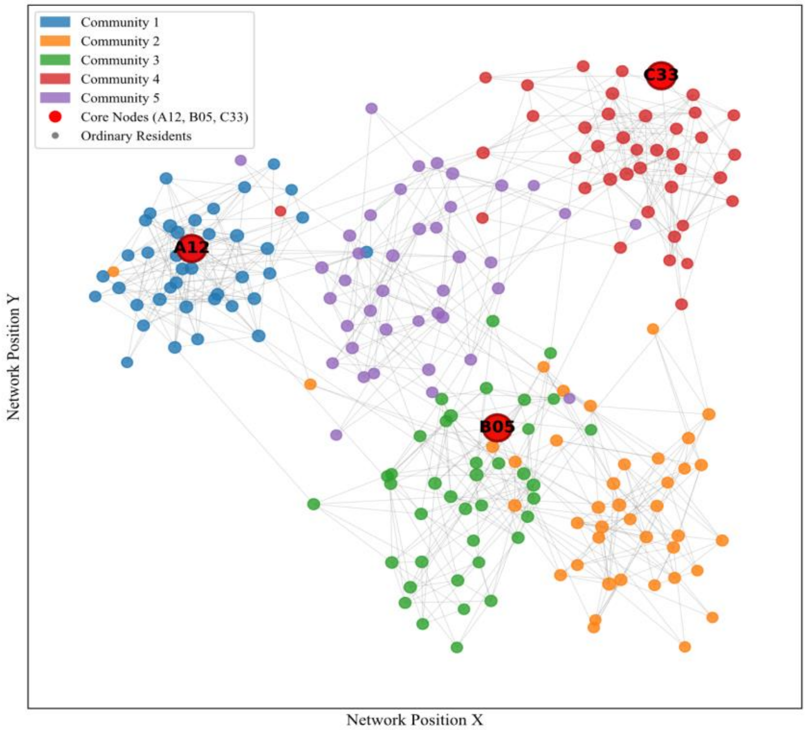

Figure 1: Community residents' information exchange network diagram

The information exchange network index of community residents is constructed and calculated through SNA (see Figure 1 and Table 1). The network presents an obvious community structure with several core figures.

Table 1: Key node SNA indicators (partial)

Node ID	Identity	Degree centrality	Intermediate centrality	Risk role judgment
A12	Former community teacher, opinion leader	0.85	0.12	Key nodes of information diffusion
B05	Representatives of affected merchants	0.45	0.31	Connecting different community groups is the key to risk transmission
C33	Ordinary residents	0.20	0.08	Edge node
...	...	...	...	...

The analysis shows that nodes A12 and B05 occupy key positions in the network. Once negative information is received and spread by A12, or B05 transmits antagonism among different groups, it will greatly accelerate the risk diffusion. This suggests that risk prevention and control should focus on these key figures and conduct targeted communication.

NLP technology is used to analyze the sentiment of text data and model the LDA theme, and the results are shown in Figure 2. The evolution of public opinion clearly shows the path of risk escalation. The emotional deterioration in March was mainly due to the spread of environmental risk information, indicating that the risk identification signal appeared for the first time. In May, the topic turned to "compensation and resettlement", which showed that the interest demands were concrete and the social risks were further aggravated. NLP has successfully captured the dynamic changes of risk focus and provided early input for forecasting.

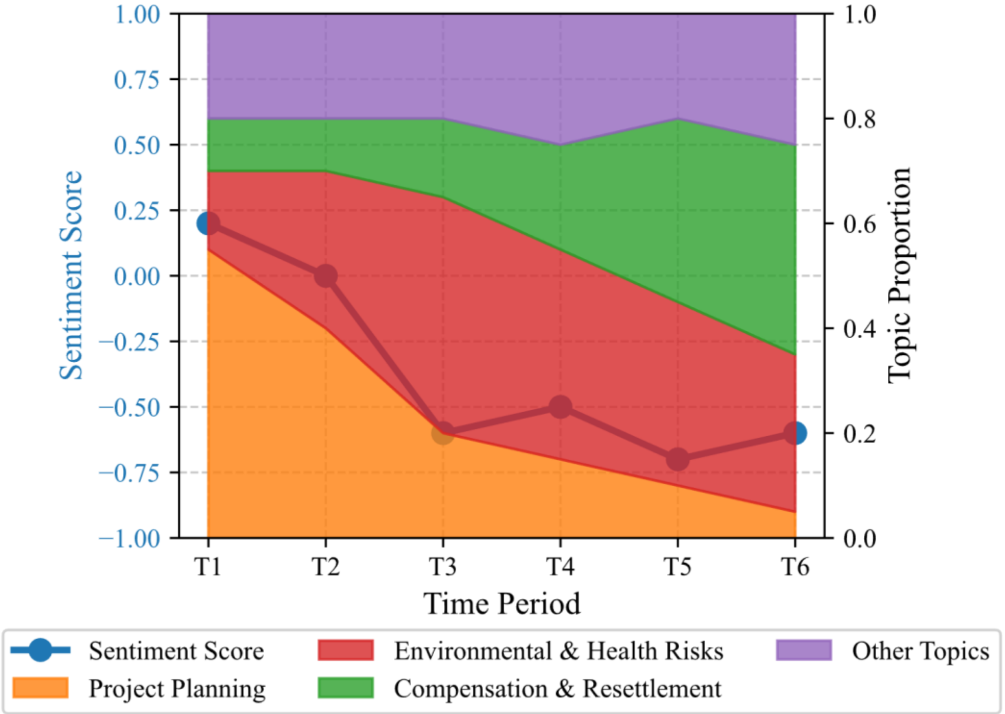

Figure 2: Trend of public opinion and emotion and evolution of main theme proportion

4.3. Verification and Results of Risk Prediction Model

Taking the above recognition results as input, it is substituted into SD-ABM fusion model for simulation and prediction. Taking T3 as the starting point of prediction, the probability of group events in the next three months (T4-T6) is simulated.The change of public dissatisfaction $M(t)$ simulated by SD model is shown in Figure 3. Without effective intervention (baseline scenario), the dissatisfaction will exceed the early warning threshold of 0.7 at the end of T5. However, after the introduction of "strengthening popular science propaganda" (policy scenario 1) in April, the upward trend of dissatisfaction slowed down.

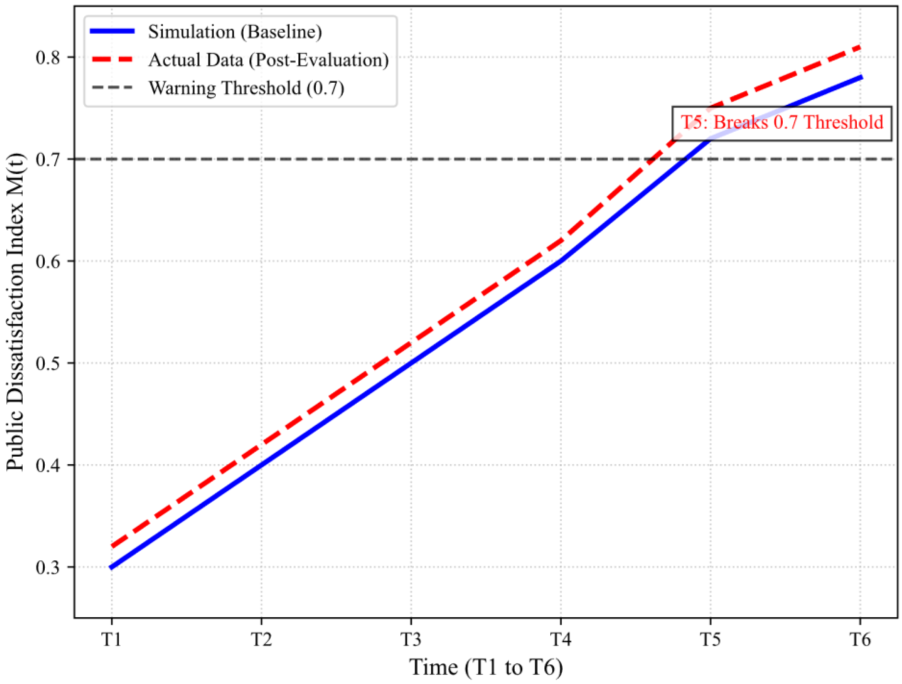

Figure 3: Comparison between SD simulation results of public dissatisfaction and actual situation

Based on the macro dissatisfaction environment of SD output, ABM runs 50 Monte Carlo simulations, and obtains the probability of group events in different policy scenarios (see Table 2).

Table 2: Prediction result of group event outbreak probability at T6 time

Simulation scene	Describe	Average outbreak probability	95% confidence interval
Baseline scene	No additional intervention	68%	[62%, 74%]
Policy scenario 1	T4 Strengthen the Popularization of Popular Science	45%	[39%, 51%]
Policy scenario 2	T4 strengthens popular science+T5 announces optimized compensation scheme	22%	[17%, 27%]

The simulation result of the baseline scene (with a high probability of 68%) is basically consistent with the historical fact that a large-scale mass incident actually occurred in T6 period, which proves the effectiveness of the prediction mechanism. The simulation results clearly quantify the effects of different policy interventions. Only popular science propaganda (scenario 1) can reduce the risk, but it can't be fundamentally resolved; The comprehensive measures combined with economic compensation (Scenario 2) can significantly reduce the risk probability to an acceptable level. This provides a scientific policy pre-evaluation tool for decision makers. In the ABM simulation, it is observed that when the key network nodes are endowed with high negative information dissemination tendency, the risk probability will be significantly higher than the independent prediction of SD model, which confirms the magnifying effect of micro-interaction on macro-risk.

5. Conclusion

At the level of risk identification, the combination of SNA and NLP realizes multi-dimensional dynamic monitoring from "who is spreading" to "what is spreading", which makes the risk traceability more accurate. At the level of risk prediction, SD-ABM fusion model overcomes the limitations of a single model, which can not only simulate macro trends, but also capture micro-emergent behaviors, and make "sand table deduction" for different intervention strategies, which significantly improves the foresight of early warning and the support of decision-making. Social risk management of major projects should change from passive response to active prediction. It is suggested to establish an intelligent early warning platform for social risks of major projects, integrate multi-source data, run this model in real time, and realize dynamic assessment and visual early warning of risk levels, thus winning valuable time for making accurate and differentiated intervention strategies. Future research will cooperate with the actual project in depth to obtain more detailed data to optimize the model parameters. At the same time, we can further explore the introduction of advanced technologies such as graph neural network (GNN) to describe the co-evolution of network and public opinion more precisely.

Acknowledgements

This work is Supported by the Science and Technology Research Program of Chongqing Municipal Education Commission(Grant No.KJZD-K202304303) (Grant No.KJQN202404311) and Chongqing Education Science Planning Project + Teaching Reform Research Special General Project + "Digital Intelligence +" Engineering Costing Professional Curriculum System Reform and Practice under the Background of New Engineering (K24ZG3230237).

References

[1] Yunze, Z. and Tingyu, X. (2023) The Spread of Collective Panic and Risk Governance Under Multiple Emergencies. Journal of Soochow University (Philosophy and Social Sciences Edition), 44(03): 162-169.

[2] Xiaokun, W. and Wanyi, Li. (2023) Internet-Based Social Governance Driven by Both Risk and Technology: Core Issues and Emerging Trends. Dongyue Forum, 44, 75-83.

[3] Xiaocui, Z., Zhao, K., Ling, T., Bo, H. and Xi, Z. (2023) Research on Predicting Terrorism Risk Shifts Based on Combinatorial Weighting. Guangxi Science, 30, 89-99.

[4] Jidi, Z., Rujie, Q., Lei, W. and Zhiqiang, D. (2022) Research on the Evolution of Social Risks and Prevention Strategies in Urban Digital Transformation. E-Government, 6, 111-124.

[5] Ming, G. Renhui H.(2021) The Influence of Group Emotion on Risk-Type Environmental Mass Incidents: An RDEU Evolutionary Game Model Approach. Journal of University of Electronic Science and Technology of China (Social Sciences Edition), 23, 1-9.

[6] Lixin, Y., Ziqiang, L. and Chengning, Z. (2024) Research on Spatio-Temporal Evolution Analysis Methods for User Sentiment in Short Videos During Sudden Natural Disaster Events. Library and Information Work, 68, 77-91.

[7] Jinbao, S., Mingxing, Y. and Lijun, D. (2024) Research on Co-evolution Mechanisms and Guidance Strategies for Three-dimensional Online Public Opinion Information Transmission Based on Information Authentication and Sentiment Analysis. Research on Information Communication, 31, 42-54.

[8] Xiao, S. and Kailong, Y. (2024) Theme Mining, Evolution, and Implications of Online Public Opinion on Emergencies Based on Weibo Hot Search Data. Information Technology and Management Application, 3, 98-112.