Academic Journal of Management and Social Sciences
ISSN: 2958-4396 | Vol. 16, No. 3, 2026

Knowledge Graph Construction and Risk Prediction for Early Warning of Individual Extreme Violence

Xin Liu [a], Yifan Wang [b], Fanliang Bu *

School of Information Network Security, People's Public Security University of China, Beijing, China
*** Corresponding author:** Fanliang Bu (Email: bufanliang@sina.com), [a] 2820335115@qq.com, [b] 2024111026@stu.ppsuc.edu.cn

Abstract: Personal extreme violence is sudden, hidden and highly destructive, and its early identification and early warning have become a key task for public security governance. However, the existing methods based on text analysis and knowledge graph still have obvious shortcomings in risk semantic identification, conceptual system construction and risk assessment interpretability. For this reason, this article is aimed at the early warning of personal extreme violence behaviour, based on the risk ontology library, and then builds a knowledge graph, and puts forward an interpretable risk prediction framework ROKEF. The experimental results show that the method is significantly superior to the representative models of KGGen, GraphRAG and RAKG in terms of key semantic recognition such as high-risk behaviours, extreme speech and psychological abnormalities, and the generated spectrum also performs better in structural integrity and risk correlation. At the same time, the knowledge graph and interpretable forecasting framework based on the ontology library can effectively support the structured representation, quantitative evaluation and traceability analysis of risks. The research results show that this method can provide important theoretical value and practical support for intelligent early warning of individual extreme violence, intelligent public security construction and social security governance.

Keywords: Individual Extreme Violence; Knowledge Graphs; Risk Prediction; Interpretability.

1. Introduction

Today's global public security governance is facing new challenges in the digital age. Personal extreme violence has become a key problem in social governance because of its suddenness, concealment and serious social harm. Governments have paid attention to the capacity building of "smart policing" and "active early warning". China's "14th Five-Year Plan Outline" also clearly proposed to "promote the intelligence of social governance[1] ", emphasising the construction of a public security early warning system based on big data. However, under the background of the mass and fragmentation of Internet information, the traditional method of relying on manual intelligence analysis and post-event response is inefficient, and it is difficult to realise the transformation from "post-disposal" to "pre-warning".

With the development of information technology, knowledge graph, as an important way of structured knowledge expression, has shown significant advantages in intelligent search[2], recommendation system[3], question and answer system [4]and other fields. In recent years, knowledge graph construction technology [5]has gradually extended from the general field to high-risk scenarios such as public security[6], mental health[7], crime prediction[8], and its value in complex event modelling and semantic understanding is becoming more and more prominent. However, the early warning of extreme violence against individuals is still challenging: such behaviours have the characteristics of concealment, suddenness and complexity, making it difficult for traditional intelligence analysis, social media monitoring and behaviour identification methods to achieve early identification and dynamic assessment.

The core of knowledge graph construction is to convert natural language texts into structured semantic networks. Its key difficulties include entity identification, relational extraction, and unified representation of complex semantics. The early method mainly relied on the grammar rules [9]written by experts to match patterns[10]. Although it has a certain accuracy, it is expensive and has poor scalability, and it is difficult to adapt to different fields and language styles. After that, the open information extraction [10](OpenIE) paradigm emerged, which improved the degree of automation of extraction by extracting triplet facts without supervision. However, the granularity of the results was inconsistent, the normalisation was difficult, and the noise was high, which limited its application in high-risk fields.

With the development of deep learning and pre-training language models (PLMs)[12], the knowledge graph has entered the end-to-end generation stage. Grapher[13] and other methods regard the mapping of text to graphs as a sequence generation problem, generating entities and relationships through large models, which improves efficiency and generalisation. However, such methods generally lack explicit modelling of the graph structure and rely on a large amount of high-quality training data.

In order to further improve the accuracy and consistency of graph generation, researchers began to introduce external knowledge or multi-modal information, such as retrieval enhanced generation (RAKG[14]) to solve the problem of long text forgetting through the "pre-entity" mechanism, and multi-modal embedding methods (such as OTKGE[15]) through optimal transmission. The alignment of different modal knowledge. However, these methods are still mainly aimed at general scenarios, with limited support for the identification of extreme violence. There are still the following problems: lack of a professional risk concept system embedded in psychology, criminology and other fields, and it is difficult to establish a knowledge structure with risk perception ability; it is difficult to identify implicit psychological states and threatening remarks. With risk behaviour, it is difficult to build a chain of behaviour with early warning value; the graph is mostly at the level of fact

description, and there is a lack of an explainable risk assessment mechanism, and it is difficult to move from "what happened" to "why it happened".

In response to the above shortcomings, this paper proposes to build a knowledge graph and explainable risk prediction framework ROKEF based on the risk ontology. On the basis of drawing on the advantages of RAKG and other methods, the framework systematically introduces risk ontology construction, risk perception information extraction and risk assessment modules based on graph neural networks to realise a complete link from data to knowledge and from knowledge to decision-making. The contributions of this article include:

1)Build a risk ontology library for individual extreme violence, systematise and structure risk factors in psychology and sociology, and build a knowledge graph based on the risk ontology base;

2) Design a risk-perceiving entity and relationship extraction mechanism to prompt the project to guide the large model to identify the semantics of key risks;

3) Propose a risk propagation and quantitative model based on graph neural networks to realise explainable risk assessment and traceability analysis.

This research realises the deep integration of knowledge graph construction and risk assessment, providing transparent, interpretable and operable intelligent early warning tools for public security scenarios.

2. METHOD

This paper proposes to build a knowledge graph based on the risk ontology and an interpretable risk prediction framework ROKEF, including the construction of the risk ontology base[16], the construction of knowledge graph based on the ontology, risk assessment and quantisation. As shown in Figure 1, the overall framework consists of three core stages: risk ontology library construction, risk perception knowledge graph generation, and graph neural network risk assessment. The framework generates a knowledge graph that meets the needs of downstream tasks, extracts the structure and semantic characteristics of the graph, and conducts risk learning with the support of the attention mechanism[17], while realising the interpretability of the results with risk calibration and traceability mechanisms.

2.1. Construction of Risk Ontology Library

This article is based on the public judgement containing keywords such as "intentional homicide", "explosion", "crime of endangering public safety by dangerous methods" and "retaliation against society" in the China Referee Document Network, as well as the reports of Xinhua News Agency and the People's Daily on major violent cases, collecting a total of 213 case data. Each of the data contains information about time, place, person and key behaviour. Based on theoretical systems such as the Violence Risk Assessment Guide (VRAG[18]) and the Behavioural Threat Assessment Model (BTAM[19]), this paper systematically summarises the concepts and relationships in the data and constructs a domain risk ontology.

The risk ontology library contains a total of 9 parent concepts, which are further subdivided into 22 subconcepts, covering key risk factors such as speech, behaviour, psychological characteristics, social situation and historical background. At the same time, the ontology library defines 10 types of relationships, which are used to express the key semantic associations in the chain of risk behaviour, such as "publishment", "implementation", "expression", "intensification", etc., which can provide a unified semantic specification for the subsequent knowledge graph construction.

2.2. Build a Knowledge Graph Based on The Risk Ontology

This stage aims to build a knowledge graph that conforms to the risk ontology system from the original text, including text blocking, entity identification, entity disambiguation, relationship extraction and other steps.

Text dynamic blocking and vectorization: In order to ensure semantic integrity, this article adopts a dynamic block strategy with sentences as the boundary, and divides the document set D into several text blocks $T=\{text_1, text_2, \ldots\ldots, text_i\}$. Subsequently, the BGE-M3 vector model is used to vectorise each text block to reduce the processing cost of large models and improve the accuracy of named entity identification.

$$T = DocSplit(D) \tag{1}$$

$$\text{s.t.}\ \forall i \neq j, text_i \cap text_j = \emptyset \wedge \bigcup_{i=1}^{n} text_i = D \tag{2}$$

$$V_T = \{\, \vec{v}_i \mid \vec{v}_i = Vect(text_i), text_i \in T \,\} \tag{3}$$

Entity identification based on prompt engineering : According to the characteristics of risk scenarios, this article designs a specific instruction template for the large language model, focussing on guiding the model to pay attention to key information such as people, dangerous goods, violent organisations, psychological state and threatening remarks. LLM recognises entities paragraph by paragraph on each text block and generates types and descriptions for them. The relevant process can be formalised as:

$$Pre_{entity_i} = \text{NER}(text_i) \tag{4}$$

$$Pre_{entity} = \bigcup_{i} Pre_{entity_i} \tag{5}$$

Among them, $text_i$ is a text block, and Pre_{entity_i} is an entity obtained in advance through LLM.

Solid vectorisation and disambiguation combination: After vectorisation of the identified entities, this paper groups similar entities into the same entity set according to the similarity threshold. First, use formula (6) to screen potential similar entities, and then complete the entity merger according to formula (7), and finally generate a unique and semantically consistent entity node.

$$Sim(e_i) = \{\, e_j \mid e_j \in Pre_{\boldsymbol{entity}}, VectJudge(e_i, e_j) = 1 \,\} \tag{6}$$

$$Same(e_i) = \{\, e_j \mid e_j \in Sim(e_i), SameJudge(e_i, e_j) = 1 \,\} \tag{7}$$

Relationship construction and ternary authenticity judgement : After the entity is de-weighted, this paper uses the vector retrieval method to search for text fragments that are highly related to the entity in the original text (Formula 8). Then, the retrieval content and entity context will be input into LLM, and the model will generate a candidate triplet. Then use LLM as a judge to evaluate the authenticity of the triplet (Formula 9), and only keep the highly credible relationship entries.

$$retriever_{V_T}(e) = \left\{ text_i \middle|_{retriever(e, \boldsymbol{v}_j, threshold)=1}^{text_i \in T} \right\} \quad (8)$$

$$rel(e_i) = LLM_{rel}\left(entity_1, retriever_{text}(e_i), retriever_{kg}(e_i)\right) \quad (9)$$

Through the above steps, a structured map conforming to the risk ontology system can be obtained, providing unified and high-quality input data for subsequent risk assessment..

2.3. Risk Assessment Based on Graph Neural Network

In order to realise the transformation from static structure description to risk prediction, this paper constructs a risk assessment model based on graph neural networks, inputs the generated knowledge graph into the neural network, outputs individual risk scores, and provides interpretable traceability paths.

Node type characteristic representation : This paper maps the node type into a 7-dimensional distribution vector, and constructs the risk type characteristics in combination with the domain a priori risk weight. Let the node set in Figure G be V, and the node mapping type function is $\phi: V \to T$, where $T = \{\text{PERPETRATOR,WEAPON, ...,PERSON}\}$ is a predefined 7 types, and the preset risk weight vector is $\mathbf{w} \in \mathbb{R}^7$, then the node $\mathbf{f}_{\text{type}} \in \mathbb{R}^7$ The calculation of the type characteristics is as follows:

$$\mathbf{f}_{\text{type}}^{(i)} = \frac{\sum_{v \in V} \mathbb{I}(\phi(v) = t_i) \cdot w_i}{|V|} \quad (10)$$

Where $\mathbb{I}(\cdot)$ is the indicator function, and $|V|$ is the total number of nodes. This feature also reflects the node category distribution and risk weight, which helps to improve the risk sensitivity to key nodes.

Features of the topological structure of the figure : In order to express the risk propagation structure of the graph, this paper extracts 6-dimensional topological characteristics, including the number of nodes $|V|$, the number of edges $|E|$, graph density $\rho(G) = \frac{|E|}{|V|(|V|-1)}$, average $\bar{d}(G)$, maximum $d_{max}(G)$ and weak connectivity index $C_{\text{weak}}(G)$ (the connected value is 1, otherwise it is 0). Among them, S_* is the normalisation factor, and its calculation method is as shown in formula (11).

$$\mathbf{f}_{\text{struct}} = \left[\frac{|V|}{S_v}, \frac{|E|}{S_e}, \rho(G), \frac{\bar{d}(G)}{S_d}, \frac{d_{max}(G)}{S_{d_{max}}}, C_{\text{weak}}(G) \right] \quad (11)$$

Semntic risk characteristics: In order to reflect the risk intensity in the semantics of the text, this article constructs a collection of three types of keywords: high, medium and low and gives weight. The text of the node $\mathbf{f}_{\text{semantic}} \in \mathbb{R}^3$ is composed of names and descriptions, and the semantic risk characteristics are calculated as follows:

$$\mathbf{f}_{\text{semantic}}^{(c)} = \frac{\sum_{v \in V} \sum_{k \in \mathcal{K}_c} \mathbb{I}\big(k \in \text{text}(v)\big) \cdot w(k)}{|V|} \quad (12)$$

Among them, $\mathcal{K}_c$ is the keyword set, and $w(k)$ is its corresponding weight.

Relationship risk characteristics: The type of edge also reflects the nature of risky behaviour. This paper divides three types of risk relationship sets and extracts four-dimensional relationship characteristics:

$$\mathbf{f}_{\text{relation}} = \left[\frac{\sum_{r \in E} w_{\text{high}}(r)}{|E|}, \frac{\sum_{r \in E} w_{\text{medium}}(r)}{|E|}, \frac{\sum_{r \in E} w_{\text{safe}}(r)}{|E|}, min\left(\frac{|E|}{|V|}, 1\right) \right] \quad (13)$$

Among them, $w_*(r)$ is the weight of relation r

Feature coding and attention weighting: The multimodal risk characteristics of nodes and edges are input into the multi-layer perception machine encoder, and mapped to a more distinguishable hidden space through formulas (14)-(16), where $\mathbf{x} \in \mathbb{R}^{d_{\text{input}}}$ is the input feature, and $z \in \mathbb{R}^{d_{\text{hidden}}/4}$ is the feature after coding:

$$\mathrm{h}_1 = \text{Dropout}(\text{ReLU}(\text{BatchNorm}(\mathrm{W}_1\mathrm{x} + \mathrm{b}_1))) \quad (14)$$

$$\mathrm{h}_2 = \text{Dropout}(\text{ReLU}(\text{BatchNorm}(\mathrm{W}_2\mathrm{h}_1 + \mathrm{b}_2))) \quad (15)$$

$$\mathrm{z} = \text{ReLU}(\mathrm{W}_3\mathrm{h}_2 + \mathrm{b}_3) \quad (16)$$

Then use the attention mechanism to assign weights to different dimensions:

$$\alpha = \sigma(\mathrm{W}_{a2}\tanh(\mathrm{W}_{a1}\mathrm{z} + \mathrm{b}_{a1}) + \mathrm{b}_{a2}) \quad (17)$$

$$\mathrm{z}_{\text{att}} = \alpha \odot \mathrm{z} \quad (18)$$

Among them, σ is the sigmoid function, and α is the attention weight.

Risk prediction and calibration mechanism: The final risk score of the weighted characteristics is output by the predictor $\hat{y}$.

$$\hat{y} = \sigma\big(\mathbf{W}_p \mathbf{z}_{\text{att}} + b_p\big) \quad (19)$$

Considering the imbalance of high and low-risk samples and the possible deviation of the predicted value from the expected distribution, this paper designs a rule-based risk calibration function. Integrate node types, semantic characteristics and relational characteristics to construct risk indicators (CRI), and calibrate them according to formula (20) to put large high-risk samples and suppress low-risk noise, so that the final score is more in line with the needs of early warning scenarios.

$$\text{CRI} = 0.4 \cdot \mathbf{f}_{\text{node}} + 0.4 \cdot \mathbf{f}_{\text{semantic}} + 0.2 \cdot \mathbf{f}_{\text{relation}} \quad (20)$$

Risk traceability and explainability: In order to realise the interpretability of the results, this paper measures the source of risk according to the node contribution. Node contribution is calculated in combination with node risk type weight and keyword contribution coefficient:

$$\text{Contribution}(v) = min\left(w_{\text{type}}\big(\phi(v)\big) + \sum_{k \in \mathcal{K}_{\text{high}}} \mathbb{I}\big(k \in \text{text}(v)\big) \cdot \lambda, 1.0 \right) \quad (21)$$

The final output includes the overall risk score, the most contributing node Top-K and its high-risk relationship path, which provides a decision-making basis for public security early warning.

3. Experiment

This chapter aims to verify the effectiveness of the model proposed in this article in risk quantification tasks. The content of the experiment mainly includes data source and preprocessing, baseline model setting, evaluation index definition and experimental result analysis.

3.1. Data Sources and Preprocessing

In view of the special nature of the research object, this article selects case data from authoritative channels such as press releases and public judgements of China Referee Document Network, covering typical personal extreme violence such as knife murder, driving collision, arson and so on. Each record contains information such as the summary of the case, the characteristics of the subject of the behaviour, the background of the incident and the final consequences.

In order to construct a control data set, this article selects the daily life records of 100 low-risk subjects, including professional activities, hobbies, family socialising and other contents. This kind of data reflects typical normal behaviour patterns, which helps the model to learn the semantic difference boundaries of high and low-risk texts, thus enhancing the ability to identify risks.

All texts have been uniformly preprocessed, including the setting of specific named entity identification prompts [20]and relationship extraction prompts to highlight key information such as violent behaviour, extreme speech, psychological state, etc. This article uses Qwen2.5-72B [21]to complete the extraction of time, place and subject-predicate structure, and uses BGE-M3[22] for sentence vector coding. Under the guidance of the risk ontology library, the model can accurately extract entities and relationships from the original text.

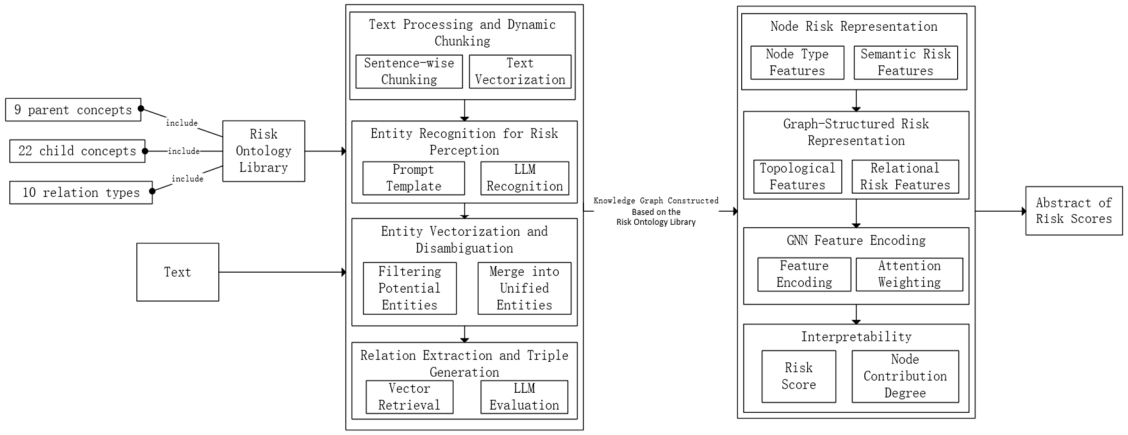

Figure 1. Knowledge graph construction and interpretable risk prediction framework

3.2. Baseline Models

In order to comprehensively verify the advantages of the method in this article, the following three representative knowledge graph construction methods are selected as the comparative baseline:

KGGen[23] was proposed by the Stanford Trusted Artificial Intelligence Research Team, which can automatically convert unstructured text into a structured knowledge network containing entities and relationships based on language models and clustering algorithms. The tool supports direct call through the Python library, is easy to operate, and is suitable for knowledge extraction of general text.

The GraphRAG[24] framework proposed by Microsoft combines the knowledge graph with the retrieval enhancement mechanism, and realises document-level semantic modelling through the three steps of graph indexing, graph retrieval and graph enhancement. GraphRAG improves the semantic acquisition ability of long texts through structured entity relationships.

RAKG was proposed by the Shanghai Artificial Intelligence Laboratory to realise the retrieval and enhanced extraction of long texts through the "pre-entity" mechanism. The framework can effectively alleviate the problem of "long-term forgetting" in the understanding of long texts in large models, and has advantages in physical disambiguation [25], relationship alignment and knowledge integration.

The above three methods represent the mainstream paradigms in the field of current knowledge graph construction: end-to-end generation, graph enhancement retrieval and retrieval enhancement extraction, which can be used as a comparative verification of the superiority of the method in this article.

3.3. Evaluation Metrics

In order to quantitatively evaluate the quality of knowledge graph generation and its support for risk assessment tasks, this paper adopts the following indicators:

Entity Density (ED):

$$ED = \frac{N_e}{\mathrm{D}} \tag{22}$$

$N_e$represents the number of entities, and D represents the total number of words in the text where the entity is located. The more entities, the stronger the ability to extract semantic information.

Relation Richness (RR):

$$RR = \frac{N_r}{D} \tag{23}$$

N_r represents the number of terds or attribute relationships in the graph. The larger the value, the more complex the relationship network and the more adequate the information retention.

Entity Fidelity (EF):

$$EF = \frac{1}{N_e}\sum_{i=1}^{N_e} LLMJudge_{entity}(e_i, retriever_{V_T}(e_i)) \quad (24)$$

Use LLM as a judge to assign a credibility score of 0 to 1 to each entity to measure the reliability of the entity.

Relation Fidelity (RF):

$$RF = \frac{1}{N_r}\sum_{i=1}^{N_r} LLMJudge_{rel}(e_i, retriever_{V_T}(e_i), retriever_{V_{kg}}(e_i)) \quad (25)$$

Similarly, LLM determines the authenticity of the relationship and outputs the credibility of 0 to 1, reflecting the extraction quality of the relationship.

Measure the performance of knowledge graph in downstream risk identification based on high and low risk labels based on entity credibility, relationship credibility and manual labelling. The higher the accuracy rate, the better the retention effect of the graph on extreme semantics.

3.4. Experimental Results

The experimental results of the four models on the data set proposed in this article are shown in Table 1, among which the bold font represents the optimal effect. We will further show the detailed results below.

3.4.1. Entity Density and Relationship Richness

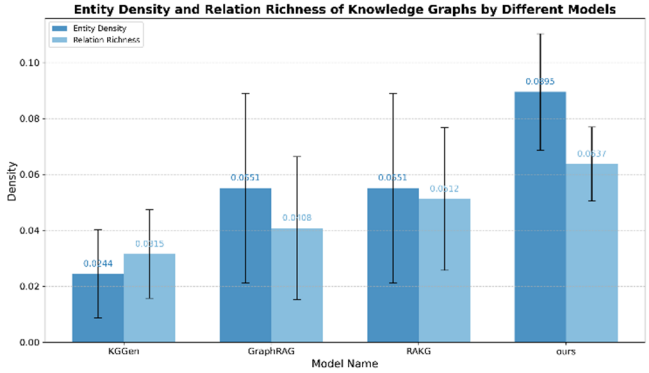

Figure 2. Entity density and relationship richness of knowledge graph built by different models

After text blocking, vectorisation, pre-entity extraction and entity disambiguation, this article guides LLM through a specific prompt template, which can output high-quality triplets under the premise of maintaining format specifications. The results of the entity density and relationship richness of the knowledge graph constructed by different models are shown in Figure 2. This method is significantly higher than the number of identifications of extreme behavioural semantics, psychological characteristics related entities and key risk relationships compared to the model, which shows that the risk ontology library and prompt template can effectively improve the risk-related semantics. Take the ability.

3.4.2. The Confidence of Entities and Relationships

The credibility of the entities and relationships of the knowledge graph built through different models is shown in Figure 3. In the figure, it can be seen that the credibility of the model entities and relationships proposed in this paper is higher than that of other models.

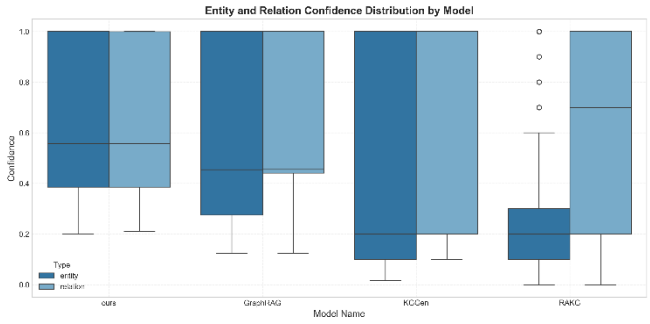

Figure 3. The credibility distribution of different models to construct knowledge graph

3.4.3. The Accuracy Rate of Risk Score and Label Correspondence

This article stipulates that the risk score corresponding to high-risk labels is between 0.7-1, while the risk score corresponding to low-risk labels is between 0-0.2. In terms of overall performance, the method in this paper shows a clear separation of the risk score of high and low risk samples: the score of high-risk samples is concentrated in 92.7%–96.3%, with an average value of 93.8%; the score of low-risk samples is distributed at 4.8%–12.3%, with an average value of 8.9%; the average difference between high and low risks reaches 84. 9%, the decision-making boundary is clear, and the variance within the category is small. In contrast: the abnormal rate of GraphRAG high-risk samples reaches 40.7%, and the distribution of low-risk samples is concentrated in 0.48–0.484, which is difficult to distinguish the categories; KGGen high and low-risk mean is close (the difference is only 0.0006), and the classification ability is close to random; RAKG has high-risk samples In the case of extremely low scores and high scores of low-risk samples, some sample scores overlap, as shown in Figure 4.

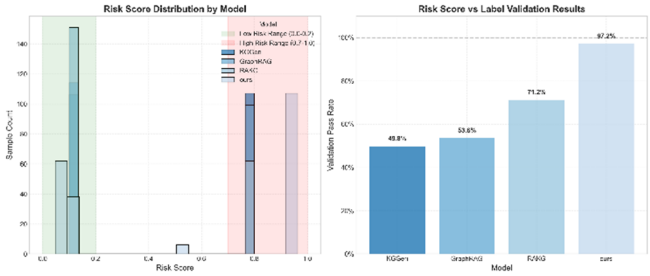

Figure 4. Risk score and label corresponding accuracy distribution analysis of different models

Table 1. The performance results of the four models

Model Name	Entity Density	Relation Richness	Accuracy of Scores and Label Correspondence
KGGen	0.0244	0.0315	49.8%
GraphRAG	0.0511	0.0408	53.6%
RAKG	0.0511	0.0512	71.2
Ours	**0.0895**	**0.0537**	**97.2%**

The above results show that the risk ontology library, prompt template and GNN risk assessment module built in

this paper have significant advantages, and they are stable in extreme semantic capture and risk identification tasks.

4. Conclusion

In response to the problems of insufficient risk semantic structure, difficult to portray the key behaviour chain and lack of interpretability of risk assessment in the early identification of individual extreme violence, this study has built a complete technical framework covering risk knowledge modelling, risk information extraction and risk quantitative evaluation, and put forward knowledge based on the risk ontology library. Graph Construction and Explainable Risk Forecasting Framework (ROKEF). The experimental results show that the framework is significantly superior to the KGGen, GraphRAG and RAKG methods in terms of risk entity identification, extreme semantic capture and relational network quality. In the risk prediction task, the average difference between high and low risk samples reaches 84.9%, and the node contribution can effectively reveal the key wind. The source of risk and the chain of behaviour verify the validity and interpretability of the model. Although good results have been achieved, the method in this article still has limitations in terms of risk ontology dependence on expert experience, scarcity of extreme event data and single text modality. Its scope of application is mainly aimed at individual early warning scenarios with sufficient text and clear risk characteristics. Future research can be further expanded from multimodal risk graph, timing risk modelling, weak supervision and migration learning, and interpretable diagram reasoning to improve the robustness and cross-scene application ability of the model. In general, the ROKEF framework proposed in this paper provides a feasible and practical technical path for the structured representation and risk warning of personal extreme violence.

References

[1] Meng, T. G., & Zhao, J. (2018). Big data-driven intelligent social governance: Theoretical construction and governance system. *E-Government*, (8), 2–11. https://doi.org/10.16582/ j. cnki. dzzw.2018.08.001.

[2] Gao, L., Zhang, H. C., & Yang, L. (2018). Research on intelligent information search technology based on knowledge graph and semantic computing. *Information Studies: Theory & Application*, *41*(7), 42–47. https://doi.org/ 10.16353/j. cnki. 1000-7490.2018.07.009.

[3] Zhao, Y. H., Liu, L., Wang, H. L., et al. (2023). A survey of knowledge graph-based recommendation systems. *Journal of Frontiers of Computer Science and Technology*, *17*(4), 771–791.

[4] Zhang, H. Y., Wang, X., Han, L. F., et al. (2023). Research on question answering system integrating large language model with knowledge graph. *Journal of Frontiers of Computer Science and Technology*, *17*(10), 2377–2388.

[5] Liu, Q., Li, Y., Duan, H., et al. (2016). A survey on knowledge graph construction techniques. *Journal of Computer Research and Development*, *53*(3), 582–600.

[6] Yuan, F. (2019). Research on anomalous event detection based on public safety knowledge graph [Master's thesis]. Institute of Automation, Chinese Academy of Sciences.

[7] Liao, Y. G., Lin, M. M., & He, W. (2018). A scientific knowledge graph of Chinese college students' psychological research in recent two decades: A visual analysis based on CiteSpace V. *Journal of Southwest University (Social Sciences Edition)*, *44*(2), 94–103, 192–193. https://doi.org/10. 13718/j. cnki. xdsk.2018.02.011.

[8] Tang, D. Q., Shi, W. Q., & Zhang, B. Y. (2018). Research on crime prediction algorithm based on multimodal information feature fusion. *Computer Applications and Software*, *35*(7), 221–225, 262.

[9] Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*(2), 199–220.

[10] Suchanek, F. M., Kasneci, G., & Weikum, G. (2007). Yago: A core of semantic knowledge. In *Proceedings of the 16th International Conference on World Wide Web* (pp. 697–706).

[11] Zhao, J., Liu, K., Zhou, G. Y., et al. (2011). Open information extraction from text. *Journal of Chinese Information Processing*, *25*(6), 98–110.

[12] Ai, J., Bai, S., Yang, S., et al. (2023). A versatile vision-language model for understanding, localization, text reading, and beyond [Preprint]. arXiv:2308.12966.

[13] Melnyk, I., Dognin, P., & Das, P. (2021). Grapher: Multi-stage knowledge graph construction using pretrained language models. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*.

[14] Zhang, H., Si, J., Yan, G., et al. (2025). RAKG: Document-level retrieval augmented knowledge graph construction [Preprint]. arXiv:2504.09823.

[15] Cao, Z., Xu, Q., Yang, Z., et al. (2022). Otkge: Multi-modal knowledge graph embeddings via optimal transport. *Advances in Neural Information Processing Systems*, *35*, 39090–39102.

[16] Ylönen, M., & Aven, T. (2023). A framework for understanding risk based on the concepts of ontology and epistemology. *Journal of Risk Research*, *26*(6), 581–593.

[17] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

[18] Rice, M. E., Harris, G. T., & Lang, C. (2013). Validation of and revision to the VRAG and SORAG: The Violence Risk Appraisal Guide—Revised (VRAG-R). *Psychological Assessment*, *25*(3), 951.

[19] Vossekuil, B. (2002). *The final report and findings of the Safe School Initiative: Implications for the prevention of school attacks in the United States*. Diane Publishing.

[20] Lyu, K., Zhao, H., Gu, X., et al. (2024). Keeping llms aligned after fine-tuning: The crucial role of prompt templates. *Advances in Neural Information Processing Systems*, *37*, 118603–118631.

[21] Bai, S., Chen, K., Liu, X., et al. (2025). Qwen2.5-vl technical report [Preprint]. arXiv:2502.13923.

[22] Chen, J., Xiao, S., Zhang, P., et al. (2024). Bge m3-embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation [Preprint]. arXiv:2402.03216.

[23] Mo, B., Yu, K., Kazdan, J., et al. (2025). Kggen: Extracting knowledge graphs from plain text with language models [Preprint]. arXiv:2502.09956.

[24] Edge, D., Trinh, H., Cheng, N., et al. (2024). From local to global: A graph rag approach to query-focused summarization [Preprint]. arXiv:2404.16130.

[25] Werlen, L. M., & Henderson, J. (2022). Graph refinement for coreference resolution. In *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 2732–2742).