IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol.006, LIIR413
DOI: https://doi.org/10.52731/liir.v006.413

A Knowledge Graph Approach for Analyzing Player Social Media Reviews

Ziyu Zhao *, Junichi Fukumoto *

Abstract

In recent years, as the gaming industry has developed, both the number of games and players have continuously increased. However, in recent years, numerous conflicts have arisen between game developers and players, leading to significant consequences for both parties. These incidents stem from a lack of effective communication between the two sides. Meanwhile, player communities have amassed a wealth of authentic reviews posted by actual players, reflecting their most genuine opinions about the games. To address this issue, the present study conducts sentiment analysis on reviews from player communities and proposes the use of graph neural networks (GNN) along with a knowledge graph constructed from gaming wikis to uncover the deep-seated reasons behind the various emotions expressed by players. Additionally, a pre-trained large model is employed to better understand player feedback, thereby enabling game developers to establish more effective communication with their player base.
Keywords: Game reviews, Graph Neural Network, Knowledge graph, Sentiment analysis

1 Introduction

In recent years, the gaming industry has grown quickly, with more games and more players worldwide. Players are no longer just receiving game content—they now use online platforms to share their thoughts and feelings about game mechanics, content, and gameplay [1]. Player communities have become important places for honest feedback and discussion.

However, most game developers still use traditional surveys and qualitative research, which creates a communication gap with players and sometimes leads to conflicts. Previous studies have mostly looked at game design and marketing, but have not deeply analyzed real player emotions and feedback. This gap can hurt user experience and a game's reputation.

To address this, our study introduces a method using graph neural networks (GNN) and knowledge graphs. We build a knowledge graph from game wikis and combine it with data from player reviews. Using sentiment analysis and a large pre-trained model, we better understand player emotions and their causes. Our approach offers three main benefits:

- Data fusion: Knowledge graphs combine in-game information and player reviews for better analysis.

- Deeper sentiment insights: GNN and big models find deeper reasons behind player emotions.

- Real-world validation: Experiments on real community data help close the gap between de-

* Ritsumeikan university, Kyoto, Japan

velopers and players.

Overall, this research aims to give the gaming industry a better tool for understanding user feedback, which can improve games and player experiences.

2 Related Work

The methodology of emotion attribution analysis has evolved through distinct stages, transitioning from human-engineered feature dependency to knowledge-augmented reasoning. Early approaches relying on sentiment dictionaries [2] employed rule-based matching of predefined emotional lexicons (e.g., "disappointment" or "excitement") to detect explicit affective expressions. While effective in tracking emotion fluctuations during specific events, these approaches fail to decode causal relationships embedded in non-emotional vocabulary (e.g., linking "graphical lag" to "sadness" requires domain-specific knowledge of game mechanics). Hybrid neural architectures (e.g., biLSTM-CNN [3]) advanced contextual keyword extraction through temporal modeling (e.g., identifying "high latency"), yet their feature fusion mechanisms lacked explicit encoding of domain knowledge, hindering the establishment of deep logical chains connecting "design elements → user experience → emotional responses." Although Large Language Models (LLMs) have revolutionized abstract emotion mapping via pretrained semantic understanding [4][5], their text-driven, unimodal explanations remain prone to hallucination and lack verifiable causality in multi-hop reasoning (e.g., validating how "skill cooldown settings" systematically influence "player churn rates"). Knowledge Graph (KG)-enhanced methods [6] mitigate factual biases through structured knowledge constraints, but conventional retrieval strategies (e.g., beam search) yield fragmented entity-relation snippets, resulting in disjointed explanations (e.g., isolating "weapon attributes" from "tutorial design"). To address these limitations, our proposed GNN-RAG framework leverages graph neural network-driven dense subgraph reasoning to extract multi-level semantic topologies from KGs (e.g., cross-module causal chains like "boss HP → weapon upgrade cost → frustration peaks"). By providing LLMs with deeply structured reasoning scaffolds, this approach elevates emotion attribution from shallow keyword matching to systematic, verifiable explanations, overcoming the dual constraints of knowledge depth and explanatory coherence inherent in existing methodologies.

3 Approach

3.1 Data

The dataset used in this study was collected from Bilibili (https://www.bilibili.com/), a leading Chinese online video-sharing platform. Bilibili hosts a wide variety of user-generated content, including video game playthroughs, reviews, and discussions. The platform features a highly active user community, where registered users can post comments and engage in discussions on both video and post pages. Bilibili operates primarily as a website, and all user interactions are conducted in Chinese.

For this research, we focused on user comments related to the game *Delta Force*. A custom web crawler was implemented to systematically retrieve relevant content from Bilibili. The crawler utilized the game's title as a keyword to search for associated videos and forum-style posts on the platform. Comments were then extracted from both video pages and related discussion threads. The data collection was conducted over the period from the official release of *Delta Force* (March 15, 2024) through May 10, 2025. During this period, the crawler was executed regularly to ensure comprehensive coverage of newly uploaded or popular contentThe final dataset comprises approximately 100,000 user comments written in Chinese, each annotated with its timestamp and basic metadata of the corresponding content (e.g., video or post title, identifier, and publication time). All data were obtained from publicly accessible content and contain no personally identifiable information.

This dataset provides a valuable resource for analyzing user sentiment and perceptions regarding *Delta Force* within the Chinese gaming community. By focusing on organically generated discourse from Bilibili, it enables a realistic understanding of how players interact with and respond to the game over time.

3.2 Emotion Classify

Player comments in gaming communities often express emotions in subtle and indirect ways–such as through sarcasm, game-specific slang, or shared cultural references. These characteristics make it difficult for traditional sentiment classification methods, including lexicon-based approaches and pre-trained models like BERT, to accurately capture the underlying emotional intent.

To address this challenge, we adopt a large language model enhanced through prompt engineeri ng to perform emotion classification. This approach enables the model to better interpret nuance d emotional expressions by leveraging contextual understanding. We evaluated its performance on a manually annotated dataset of 400 player comments, focusing on the emotion category of j oy. Compared with a standard BERT-based classifier, the prompt-based model achieved an F1 s core that is 22.8 percentage points higher, as shown in Table 1.

Beyond classification accuracy, emotion detection plays a crucial role in enhancing the knowled ge graph we build from user comments. By incorporating emotional attributes–such as positive or negative sentiment–into entity relationships, we enable emotion-aware reasoning over player feedback. This enriched graph structure supports downstream tasks such as detecting dissatisfact ion patterns or recommending targeted game improvements, which would be difficult to achieve using emotion-agnostic models.

Table 1: Performance comparison on 4o-mini, 4o-mini with prompt, and BERT. F1 values marked with ↑ indicate the best among models for each emotion.

Method	4o-mini with prompt			4o-mini			BERT		
	F1	Recall	AUC	F1	Recall	AUC	F1	Recall	AUC
Joy	**66.7**	58.1	75.1	**66.7**	74.4	75.3	43.9	95.2	55.0
Regret	**58.5↑**	76.0	83.5	35.5	32.0	63.0	–	–	–
Disgust	50.0	41.7	70.2	**50.0**	41.7	70.2	–	–	–

3.3 Build Knowledge Graph

To construct the knowledge graph, we first crawled wiki pages related to the game, collecting a total of 40,634 articles. A locally deployed LLaMA-7B model was used to extract entity–relation–entity triplets from these texts, which were then stored in a Neo4j graph database.

Using a knowledge graph offers key advantages in representing and analyzing complex information in gaming contexts. Compared to flat or sequential data formats, the graph structure explicitly models relationships between game elements—such as characters, weapons, missions, maps, and narrative events—allowing for structured reasoning. Most importantly, this structure supports the modeling of emotion-related causal chains. For example, by linking user comments expressing frustration to specific game events (e.g., a boss battle or patch update), and then connecting those events to upstream mechanics (e.g., weapon nerfs or difficulty settings), the graph enables emotion-aware causal reasoning.

Such reasoning chains are difficult to capture using traditional text-based methods. The knowledge graph thus provides a foundation not only for sentiment classification, but also for identifying the deeper causes of emotional responses, making it a powerful tool for understanding player experience at a systemic level.

3.4 GNN

Traditional player sentiment analysis often looks at reviews as isolated text and cannot capture the complex connections between game elements, leading to a shallow understanding of player concerns.

Graph Neural Networks (GNN) solve this problem by using knowledge graphs to represent relationships between things like characters, mechanics, and storylines. By combining player sentiment with this structured information, GNN can:

- Find patterns in how players feel about related game features

- Trace negative feedback back to the specific game elements causing it

- Show how one problem might affect opinions on other parts of the game

GNN also offers better explanations than black-box models, because the graph clearly shows how player feelings are connected to game details. This helps developers understand not just what players dislike, but why, and how problems are related—making it easier to fix the real issues.

3.4.1 GNN Model Training Procedure

The system first extracts all entities, relations, and triples from Neo4j, mapping them to dense indices. The graph is constructed using node and edge lists (edge_index, edge_type), following the PyTorch Geometric Data structure.

The model backbone is a three-layer attention GNN, with a trainable embedding per entity and per relation (entity_embedding, relation_embedding). For each batch, all positive triples (h,r,t) are scored, and negative samples (h,r,t′)(are generated by corrupting the tail. The triple score is calculated as:

$$score(h,r,t) = \sum_{d} e_h^d \cdot r_r^d \cdot e_t^d$$

The link prediction loss is:

$$L_{link} = -\frac{1}{N}\sum_{i=1}^{N} \log \sigma \backslash Big\left(score_{pos}^{i} - score_{neg}^{i} \backslash Big\right)$$

Where σ is the sigmoid function, $score_{pos}$ and $score_{neg}$ are positive/negative triple scores.

3.4.2 GNN Reasoning Procedure

During inference, for each input comment, the system first locates relevant entities in the knowledge graph by substring matching using the function. For each identified entity, all possible paths with a maximum length of three are enumerated via the find_3_hop_paths function. Each candidate path is then evaluated according to two criteria: emotion matching and semantic similarity.

For emotion matching, each relation r_i along the path is compared to the target emotion's keyword set KW_e(EMOTION_RELATION_KEYWORDS). If the relation matches any keyword in this set, it is assigned a high score; otherwise, it receives a lower score. The overall emotion score for the path is computed as the average score along the path:

$$score_{path}^{emo} = (1/k)\sum_{i=1}^{k} s_i$$

$$s_i = \begin{cases} 0.8, r_i \in KWe \\ 0.1, otherwise \end{cases}$$

And k denotes the path length.

For the semantic similarity score, the comment is segmented into a keyword set, and all keywords as well as all path relations are encoded by a pretrained semantic model (m3e). The semantic score for the path is defined as the maximum cosine similarity between any path relation embedding r_i and any comment keyword embedding h_w:

$$score_{path}^{sem} = \max_{i=1..k} \max_{w \in KW(C)} \cos(r_i, h_w)$$

Where KW(C) denotes the set of segmented comment keywords.

Finally, the overall path score is calculated as a weighted sum of the emotion score and the semantic score (the latter multiplied by 10 for balancing scale as in the implementation):

$$score_{final} = w_e \cdot score_{path}^{emo} + w_s \cdot 10 \cdot score_{path}^{sem}$$

Where we and ws are weighting factors for emotion and semantic components, respectively All candidate paths are ranked according to $score_{final}$, and the top K results are selected as the final inference and explanation outputs.

3.5 LLM Explanation

To enhance the interpretability and persuasiveness of the final sentiment analysis explanations, this work further integrates the subgraphs (i.e., inference paths) generated by GNN reasoning, the original video descriptions or user post content, and the predicted sentiment category as joint input to a large language model (LLM). A custom prompt is constructed to guide the LLM to generate more detailed and causally coherent explanations regarding the expressed sentiment. The procedure is as follows:

First, the high-confidence relationship triples inferred by the GNN, together with the original review text and sentiment classification results, are aggregated as structured input. Then, a multi-step prompting strategy is designed, instructing the LLM to sequentially perform information filtering, causal reasoning, and conclusion generation based on the supplied data. The model is guided to answer questions such as "What, or why, did this review exhibit such sentiment?", ultimately providing logically organized explanations for users.

4 Analysis

Figure 1 shows an example comparing KG-RAG and GNN-RAG in explaining why players feel disgust. For the review "It's all copied, boring Tencent games", the traditional KG-RAG method only finds a simple one-step fact ("Tencent Games → [belong] → Tencent.inc"), which does not explain the real reason for the negative sentiment.

In contrast, GNN-RAG gives much better reasoning and explanations. First, it finds direct causes like "Tencent Games → [Forced monetization] → Player" (where "Forced monetization" is an emotion keyword). GNN-RAG also identifies several multi-step paths, and can tell which ones are more semantically related to the review, such as "Tencent Games → [Subsidiary] → Delta Force".

Next, the system chooses a multi-step chain like "Tencent Games → [Subsidiary] → Delta Force → [Integration] → Battlefield." This path matches the player's use of the word "copied," showing the model can link a player's frustration to deeper issues in the game's design.

Overall, as seen in Table 2, GNN-RAG can find the right reasoning chains for many negative reviews, whether about unfair monetization, security issues, or cheating. This approach combines emotion matching and multi-step reasoning, helping developers understand not just what the problem is, but why players feel that way.

Table 2: Analysis of Player Emotions, GNN Reasoning Chains

Review	**GNN Reasoning Chain**	**Emotion**
It's all copied, boring Tencent games	Tencent Games → [Subsidiary] → Delta Force → Integration → Battlefield	Disgust
You should absolutely avoid Tencent games	Tencent Games → [Forced monetization] → Player	Disgust
Tencent games have too many cheaters	Tencent Games → [Security vulnerability] → Insufficient anti-cheat measures	Disgust

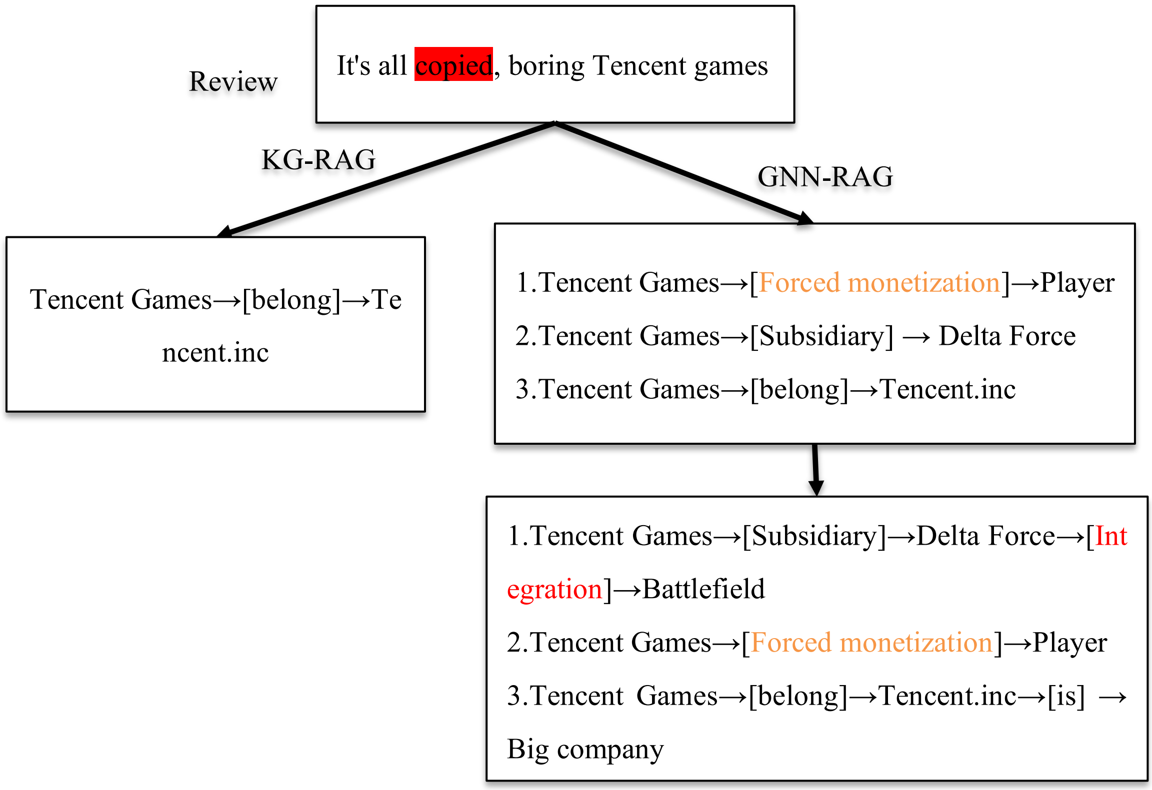

Figure 1: The information found by KG-RAG and GNN-RAG

5 Conclusion

We have demonstrated that leveraging a knowledge graph in tandem with a refined large language model can significantly enhance the analysis of player social media reviews. Beyond sentiment classification, the use of GNN-based reasoning chains and BFS-guided path exploration provides a transparent view into how contextual relationships drive specific emotional responses. This interpretability is crucial for game developers and community managers seeking to pinpoint the root causes of user feedback, whether it is anxiety over in-game resources or dissatisfaction with gameplay changes.

Our results confirm that incorporating knowledge graph structures can surpass traditional deep learning baselines in both accuracy and explainability. Future work includes extending the knowledge graph to cross-lingual or multimodal data, as well as exploring more advanced LLMs for finer-grained sentiment and topic detection. By uniting sentiment analysis with robust graph-based reasoning, this framework lays the groundwork for more comprehensive and interpretable analytics in online gaming communities and beyond.

6 References

[1] Kong Shao-hua. Research on user information behavior massively multiplayer online game virtual community—an example of netease's massively multiplayer online game fantasy westward journey. Information Sciences, 2013.

[2] Fujio Toriumi, Takeshi Sakaki, and Mitsuo Yoshida. Social emotions under the spread of covid-19 using social media. The Japanese Society for Artificial Intelligence, 35(4):F-K45 1–7, 2020.

[3] Mehrimah Amirpour and Reza Azmi. Prfashion24: A dataset for sentiment analysis of fashion products reviews in persian, 05 2024.

[4] Xiaoyu Zhang, Yishan Li, Jiayin Wang, Bowen Sun, Weizhi Ma, Peijie Sun, and Min Zhang. Large language models as evaluators for recommendation explanations. In Proceedings of the 18th ACM Conference on Recommender Systems, RecSys '24, page 33–42, New York, NY, USA, 2024. Association for Computing Machinery.

[5] Junichiro Niimi. Dynamic sentiment analysis with local large language models using majority voting: A study on factors affecting restaurant evaluation. 07 2024.

[6] Jiashuo Sun, Chengjin Xu, Luminyuan Tang, Saizhuo Wang, Chen Lin, Yeyun Gong, Heung-Yeung Shum, and Jian Guo. Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph, 07 2023.

[7] Claudiu Creanga and Liviu P. Dinu. ISDS-NLP at SemEval-2024 task 10: Transformer based neural networks for emotion recognition in conversations. In Atul Kr. Ojha, A. Scza Doğruöz, Harish Tayyar Madabushi, Giovanni Da San Martino, Sara Rosenthal, and Aiala Rosá, editors, Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), pages 649–654, Mexico City, Mexico, June 2024. Association for Computational Linguistics.