Metaverse Basic and Applied Research. 2025; 4:226
doi: 10.56294/mr2025226

ORIGINAL

Analysis of Narrative Networks in Spanish Literature: An Automatic Learning Method for Knowledge Graphs

Análisis de Redes Narrativas en Literatura Española: Un Enfoque de Aprendizaje Automático para la Creación de Grafos de Conocimiento

Kunfei Li[1]

[1]School of Foreign Languages and Cultures, Nanjing Normal University. Nanjing, 210023, China.

Cite as: Li K. Analysis of Narrative Networks in Spanish Literature: An Automatic Learning Method for Knowledge Graphs. Metaverse Basic and Applied Research. 2025; 4:226. https://doi.org/10.56294/mr2025226

Submitted: 04-08-2025 **Revised:** 15-10-2025 **Accepted:** 14-12-2025 **Published:** 15-12-2025

Editor: PhD. Dra. Yailen Martínez Jiménez

Corresponding Author: Kunfei Li

ABSTRACT

Introduction: spanish literature, known for its emotional depth and complex character interactions, has seen limited computational exploration of these relationships. The lack of annotated data and NLP tools for Spanish hampers the development of accurate knowledge graphs (KGs) to map character dynamics.
Objective: this research presents an automated pipeline to extract, organize, and visualize character relationships in Spanish literary classics, such as Don Quijote, La Regenta, and Fortunata y Jacinta, with over 2500 entity interactions.
Method: the method leverages multilingual contextual embeddings for high-accuracy inferences, using Multilingual Bidirectional Encoder Representations from Transformers (mBERT) for feature extraction and Named Entity Recognition (NER) for character identification. Graph Convolutional Networks (GCNs) are employed to capture narrative ties through joint entity-relational learning, and the KG is built with RDF triples and visualized using SpaCy.
Results: the approach achieves significant performance metrics: precision (85,01 %), recall (87,06 %), and F1 score (87,96 %). The generated networks effectively represent character interactions and narrative structure, offering valuable insights into the relational dynamics of the texts.
Conclusions: this method contributes to the development of high-quality KGs for Spanish literature, advancing comparative storytelling, computational literary research, and understanding character networks in literary analysis.

Keywords: Digital Literary; Knowledge Graphs (KG); Graph Convolutional Networks (GCNs); Spanish Literature; Entity-Relational.

RESUMEN

Introducción: la literatura española, conocida por su profundidad emocional y sus complejas interacciones de carácter, ha visto limitada la exploración computacional de estas relaciones. La falta de datos anotados y herramientas de PNL para el español dificulta el desarrollo de gráficos de conocimiento preciso (KGs) para mapear la dinámica de caracteres.
Objetivo: esta investigación presenta un canal automatizado para extraer, organizar y visualizar las relaciones de carácter en clásicos de La literatura española, como Don Quijote, La Regenta y Fortuny Jacinta, con más de 2500 interacciones entre entidades.
Método: el método aprovecha las incorporaciones contextumultilingüpara inferencias de alta precisión, usando representaciones codibidireccionales multilingüde los transformadores (mBERT) para la extracción de la característica y el reconocimiento de entidad nombrado (NER) para la identificación del carácter. Las redes convolugráficas gráficas (GCNs) se emplean para capturar lazos narrativos a través del aprendizaje

entidad-relacional conjunto, y el KG se construye con triples RDF y se visualiza usando SpaCy.
Resultados: el enfoque logra métricas de rendimiento significativas: precisión (85,01 %), recuerdo (87,06 %), y la puntuación F1 (87,96 %). Las redes generadas representan efectivamente las interacciones de los personajes y la estructura narrativa, ofreciendo valiosas percepciones sobre la dinámica relacional de los textos.
Conclusiones: este método contribuye al desarrollo de KGs de alta calidad para la literatura española, avanzando en la narración comparativa de cuentos, la investigación literaria computacional y la comprensión de redes de personajes en el análisis literario.

Palabras clave: Literatura Digital; Gráficos del Conocimiento (KG); Redes Convolucionales Gráficas (GCNs); Literatura Española; Entidad-Relacional.

INTRODUCTION

Innovative methods for examining and understanding literary works have been made possible in recent years by the combination of computer tools with literary analysis. The KG, a potent data structure that records elements (such as characters, settings, and events) and their interconnections, is one of the most promising technologies in this field.(1) It lets one visually and systematically represent the connections within a story. The use of KGs in literary analysis gives a new vision of the intricate web of relationships between story elements by depicting such relationships.(2) This technique proves particularly valuable in Spanish literature due to its rich historical and cultural history, which provides a unique opportunity to discover latent patterns in the texts that might otherwise be difficult to discover with the help of traditional methods.(3) The process of literary analysis has always been based on subjective interpretation and scholars focused on their interpretation of characters, themes and the development of the stories. This might lead to only a shallow view and often fails to take into account a more holistic collection that might be revealed in any one work or a collection of works.(4) Conversely, since knowledge graphs have the ability to organize the narrative elements into a visual network that is easy to navigate and explore, they allow a more objective and methodological research to be performed.(5)

KGs simplify the detection of the underlying trends and patterns that might remain overlooked in conventional analyses by representing literary components such as characters, themes, places, and events as nodes and their relationships as edges.(6) Additionally, computer technologies used to automate the process of extracting and classifying important narrative elements could greatly improve the laborious and time-consuming aspects of literary analysis. The disadvantages of traditional methods, which often require a manual sorting process of large amounts of text to discover relevant fragments, are reduced by this automated method.(7) By applying computational algorithms KGs could find connections between themes, characters and plot in large quantities of text by providing scholars with new information about established literature. It offers an opportunity to approach a more comprehensive analysis of the structure of the text and at the same time to save time.(8) Through a framework that enables a more extensive comprehension of narrative structures, it is hoped that by providing knowledge graphs, one could enhance computational analysis of literature in Spanish. By organizing narrative information into intricate needed webs of connections, scholars could discover concealed patterns and correlations, which was challenging using traditional approaches.(9) The final aim of the research is to offer a systematic and evidence-based method of literary criticism by combining the current technology and the old methods of interpretation to find out a better understanding of the Spanish literary masterworks. This is a way of building up literary scholarship, with its hitherto inaccessible richness and perspectives.(10)

A Hybrid method for SA that used DL algorithms and KGs to determine whether a short document, like a tweet, was positive or negative was assessed in the research.(11) To create sentiment forecasts, this proposal represented tweets as graphs and then used graph similarity metrics and a DL classification algorithm. Using n-gram models, Rec (89 %) and F1-sce (88 %) were obtained. The investigation was limited by its inability to investigate the relationship between KG and applications such as cross-domain polarity classification and irony detection.

A combination of KGs and DL for intelligent image indexing and retrieval in the research.(12) The proposed strategy enhances semantic understanding and retrieval efficiency by merging structured KGs with DL models such as ViT and Efficient Net. A mAP of 0,5 indicates a significant jump in performance, which represents the success of the proposed method. Prior methods had poor retrieval accuracy and stability because of their reliance on mostly visual characteristics and limited consideration of semantic context.

To improve the interpretability and dependability of hybrid AI systems in the healthcare industry, the research presents Trust KG, a KG-based framework as evaluated in the research.(13) Using VISE and HealthCare AI, it integrated symbolic reasoning and inductive learning for link and counterfactual prediction. The method increased projected accuracy and transparency by exposing hidden medical connections. KG development needed a large amount of expert data, and validation was restricted to the lung cancer domain.

An inventive architecture called XST-GCNN was created to process irregular and heterogeneous MTS data as evaluated in the research.[14] The XST-GCNN architecture utilized heterogeneous Gower distance, spatiotemporal graph creation, and variants of GCNNs utilizing GNN Explainer for added explainability, and implemented the graph estimate. With a mean Receiver Operating Characteristic Area under the Curve (AUC) score of 81,03, it performs better than conventional models. The incapacity of previous approaches to simultaneously capture spatiotemporal linkages in irregular data and their lack of explainability limited their interpretability and therapeutic usefulness.

Enhancing interpretability and resilience in DL models requires integrating estimates and causal inference techniques, particularly in NLP, as assessed in the research.[15] Interestingly, the model with the lowest Acc (84,50 %) was found to have a causal impact on the predictions made by each of the other three DL models.The original data's underlying causal effect patterns were not adequately captured by the synthetic data, which displayed poor diversity and homogeneity.

By examining speech data obtained using an MR system, the research explored a novel method for early PD identification.[16] Speech characteristics were taken from MR data and assessed using ML and DL models, such as HuBERT and XGBoost, to find patterns suggestive of Parkinson's disease. Logistic Regression, SVM, Random Forests, AdaBoost, XGBoost achieved F1-sce (0,90), Rec (0,91), Pre (0,92). The research had been limited by a small sample size and relied on specific speech tasks, which might have constrained the generalizability of the findings.

Semantic graphs were used in the research to visualize datasets to improve accessibility and navigation of cultural heritage archives as assessed in the research.[17] To handle multilingual datasets, create multi-layer semantic maps, and organize data, it used massive language models that had already been trained. The method improved user involvement and comprehension across languages and skill levels by enabling intuitive dataset exploration. The approach occasionally affected graph clarity and interpretability due to difficulties with very large datasets and minor semantic nuances.

The research examined the literature on STSS, found efficient DL methods and language models, and pointed out problems and potential paths forward.[18] Carried out an SLR, examining current methods, datasets, and short text semantic models. Found language models, datasets, and appropriate DL methods that successfully capture short-text semantics. The genuine semantics of short text could not be fully captured by traditional rule-based and ontology-based approaches, and several recent developments in STSS have not yet been thoroughly investigated.

To create semantically enriched RDF triples, the research sought to design and build KGs from unstructured Spanish text by extracting named entities and relations as assessed in the research.[19] Entities were linked to DBpedia used SPARQL queries, lexical patterns were established for entity and relation extraction, and RDF triples were created using OWL attributes for comparison with common metrics. RDF triple extractors were achieved Pre (55,85 %), Rec (46,96 %), F1-sce (51,02 %). The approach's generalizability was hampered by its language dependence, reliance on predetermined lexical patterns, and evaluation within a small number of domains.

Research Gap

Entity-relation extraction from Spanish texts has improved but still faces significant challenges. Current methods rely on lexical patterns and limited SPARQL connections to DBpedia, which restricts them from recognizing complex semantic relationships such as causality and irony. Their focus on a few domains

makes their usage less applicable to low-resource scenarios, resulting in less expressive knowledge graphs that overlook temporal dynamics. Furthermore, there is no availability of task specific datasets to interpret the available information frameworks, the necessity of more semantically rich KGs to model complex Spanish texts properly depict relationships.

The primary aim of the research is to construct an automated computational system of extracting, organizing and imagining of character relation in Spanish literature work by KG. The research seeks to accurately model narrative interaction, improve connection extraction precision, and improve literary analysis and comparative storytelling in Spanish literature by combining multilingual contextual embeddings, NER, and GCN.

- Contextual embeddings mBERT and NER were used to precisely identify characters.
- It used GCNs to learn narrative links jointly between entities and relationships.
- RDF triples and SpaCy for interpretable character networks were used to create and display a KG.
- It showed excellent recall and precision in identifying character associations, hence facilitating digital literary analysis.

The research is organized as follows: the research first outlines the scope and objectives of visualizing character interactions in Spanish literature. Then, it reviews the existing KG and multilingual NLP studies to identify significant gaps. Three stages were proposed in the methodology: text preprocessing, NER, and GCN link modeling. In this research, precision, recall, and F1-scores are used to evaluate the experimental results. The implications of the research findings for computational literary analysis are discussed in conjunction with conclusions concerning ways through which cross-lingual narrative visualization could be advanced.

METHOD

A KGF covering entity recognition, relationship learning, data collecting, preprocessing, and graph production is the aim of this research. The dataset consisted of selected Spanish literary works. To ensure entity consistency, the texts were cleaned, tokenized, and POS-tagged. Named entity recognition was applied to find character entities, and mBERT produced contextual embeddings. GCN captured narrative links by learning entities and their relations jointly. Finally, SpaCy was used to generate and visualize RDF triples, enabling systematic digital literary analysis and yielding interpretable character networks. Figure 1 displays the Pipeline flow for KGF.

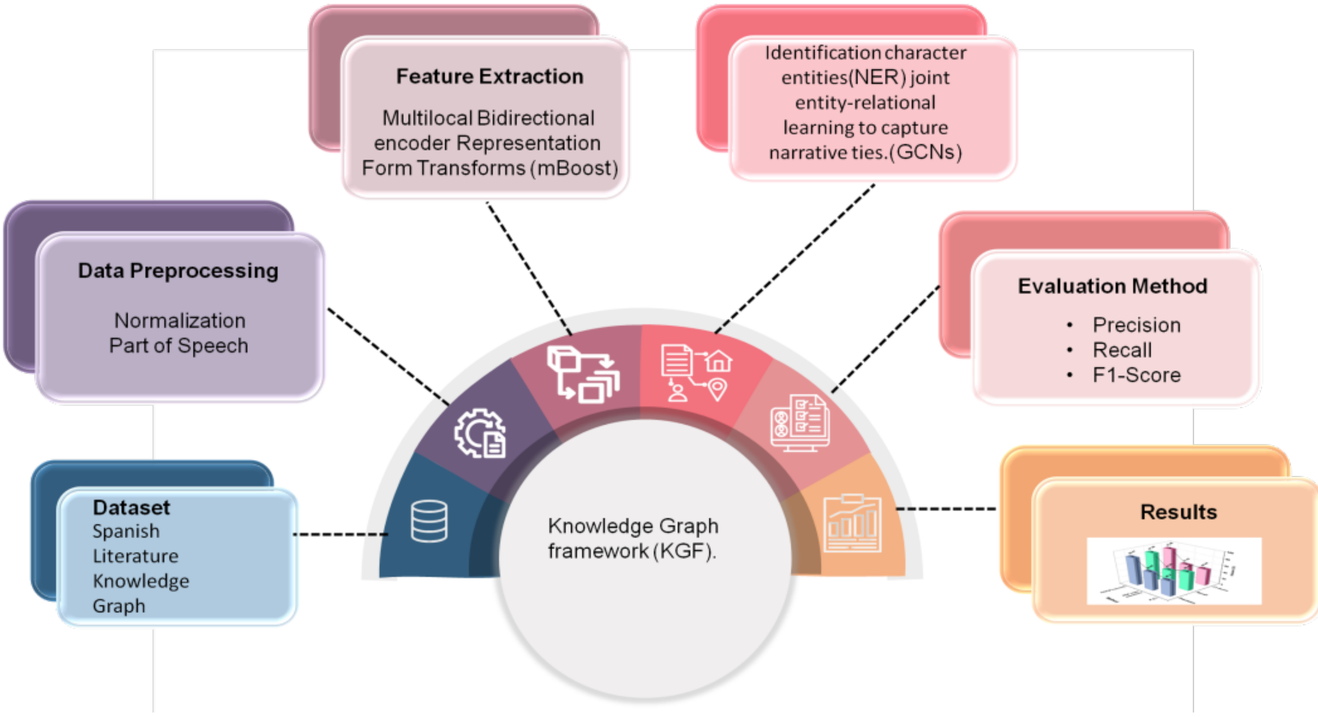

Figure 1. Overall Flow of the Knowledge Graph framework

Data set

Table 1. Feature Description of the Spanish Literature Knowledge Graph dataset

Feature No.	Feature Name	Description	Type
1	ID	Unique identifier for each record; ensures each row could be referenced individually.	Integer
2	Sentence	Full Spanish sentence describing the interaction between Entity 1 and Entity 2; primary input for NLP	Text
3	Entity 1	First character or entity in the interaction; serves as the source node in the KG.	Text
4	Entity 2	Second character or entity in the interaction; serves as the target node in the KG	Text
5	Relation	Type of interaction or relationship between Entity 1 and Entity 2; forms the edge label in the KG	Text
6	Relation_Type	High-level category of the relation (e.g., Dialogue, Romantic, Conflict); groups relations for analysis	Categorical
7	Emotion_Tone	Sentiment or emotional tone of the interaction (e.g., Positive, Neutral, Aggressive); used as edge attribute	Categorical

The Spanish Literature Knowledge Graph dataset (https://www.kaggle.com/datasets/zyan1999/spanish-literature-knowledge-graph-data/data), which was created for KG creation, NLP, and graph-based modeling, includes character connections that were taken from Spanish literature. It covers more than 2500 entity interactions from several great Spanish books, including *La Regenta, Don Quijote, and Fortunata y Jacinta*. A statement summarizing the interaction, the entities engaged, the type of relationship, the emotional tone, and contextual metadata such as chapter, source, and frequency of interactions are captured in each record. The dataset is appropriate for tasks like relation extraction, GCN training, and visual KG analysis because it also offers confidence scores and optional linguistic elements. Table 1 shows an overview of the Spanish Literature Knowledge Graph dataset features and semantic descriptors.

Data Preprocessing

To create a KG to extract and display character relationships in Spanish literature. To accurately extract entities and relationships, data preparation cleans and standardizes text by eliminating extraneous words, unusual characters, and noise. Data preprocessing includes methods such as Tokenization and part-of-speech.

Tokenization: tokenization, which separates sentences into words, characters, and punctuation based on spaces and punctuation marks, should be used to analyze textual data and make it easier to understand character associations in Spanish literature using KGs.

Part-of-speech: POS tagging, the assignment of words in a sentence to their correct grammar classes including nouns, verbs, adjectives, etc., through consideration of sentence context, and exact sentence segmentation, which better proposes the use of KGs to identify parts of speech as character relationships analysis in Spanish literature

Feature extraction using Multilingual Bidirectional Encoder Representations from Transformers (mBERT)

KGs deliver orderly data in narrative connections by visualization of complex character interactions. Using a standard structure of analyzing character associations is generated across languages using a uniform framework. mBERT, a 102-language pre-trained multilingual model, to encourage cross-lingual comprehension. MBERT could achieve slightly lower results than monolingual in feature-rich languages Such models as BETO, although this disadvantage is compensated for by its cross-linguistic generalization ability. The model uses language-specific pre-training where it is necessary to boost accuracy and ensure effective understanding of a wide range of linguistic situations. mbERT should also be improved to enhance performance that relies on data augmentation, additional ZMTD goals, KL divergence goals, and an information compensation training strategy. L2 regularization is performed to stabilize training and prevent overfitting of the embeddings of tokens in the input layers of the mBERT. This makes it possible to have a strong bilingual transfer learning method, which ensures that there is a similarity of representation of entities in Spanish literary works of a diverse number of texts, and is flexible to future multilingual uses.

Identifies character entities using Named Entity Recognition (NER)

To enhance the research of Spanish literature texts and construct a KG to aid the character relationships through the use of NER to identify and classify significant entities like people, organizations, etc, and places. This is done with spaCy, a complex open-source Python-based natural language processing model, research is aimed at researching the Spanish literary works to identify valuable information. Using residual the spaCy NER pipeline will identify and classify, convolutional networks and Bloom embeddings, people, places, and organizations. Its capability to analyze text in a scalable and effective manner makes it possible. Multilingual, support interoperability with any custom model in TensorFlow and PyTorch, and support multitask transformer model learning e.g. BERT. This approach aims at simplifying the process of KG generation and analysis by structurally enshrining character relations and interactions in Spanish literature.

Joint entity-relational learning to capture narrative ties using Graph Convolutional Networks (GCNs)

The GCN architecture employed in this research is intended to handle data in the form of a generalized topological network to model the intricate relationships seen in Spanish literature. Through convolutional processes, the GCN propagates features throughout the graph to capture the multidimensional relationship patterns between characters. Precisely, the model is trained on entity representations and interactions, which is accomplished through two GCN layers. In this case, nodes in the graph are used to represent characters, and the edges are used to represent relationships between the characters. ReLU (Rectified Linear Unit) is used to add non-linearity after each convolutional layer to help the model identify more complex associations. The input character embedding feature X consists of semantic character embeddings, which are pre-trained and are the representations of the semantic features of the characters, and could be acquired through mBERT of these entities are then embedded and the interdependencies and relational properties of the entities determined by using the GCN. The pipeline where the graph data is processed through the GCN layers is referred to as the "Graph Processing Module," as shown in Figure 2 of the design. By carrying out tasks like feature propagation and aggregation, the module enables the model to disperse data throughout the graph to improve comprehension of character associations. The robustness and interpretability of the learned character associations are enhanced

https://doi.org/10.56294/mr2025226 ISSN: 2953-4577

by the multi-head pooling procedure, which further refines the learnt features by merging data from many viewpoints. The GCN architecture processes the relational graph using two layers, ReLU activation, and mBERT character embeddings as input. Both the entity traits and the intricate relational patterns present in the literature could be captured by the model in this cohesive procedure.The GCN model simultaneously learns entity representations and their relational interactions inside a single framework, resulting in joint entity-relational learning. Table 2 shows the Hyperparameters for Training and GCN Configuration.

$$h * w \approx \theta\left(C^{-\frac{1}{2}}BC^{-\frac{1}{2}}\right)w \qquad (1)$$

In this equation (1), whereas w indicates the input feature or weight vector, w, h stands for the hidden or latent feature vector that captures semantic or structural information. Adjacency or relationship matrix B encodes entity relationships, and diagonal degree matrix C is utilized for normalization; the normalized propagation is formed by:

$$h * w \approx \theta\left(C^{-\frac{1}{2}}BC^{-\frac{1}{2}}\right)w.$$
$$Y = \sigma\left(C^{-\frac{1}{2}}BC^{-\frac{1}{2}}WX\right) = \sigma(\tilde{B}WX) \qquad (2)$$

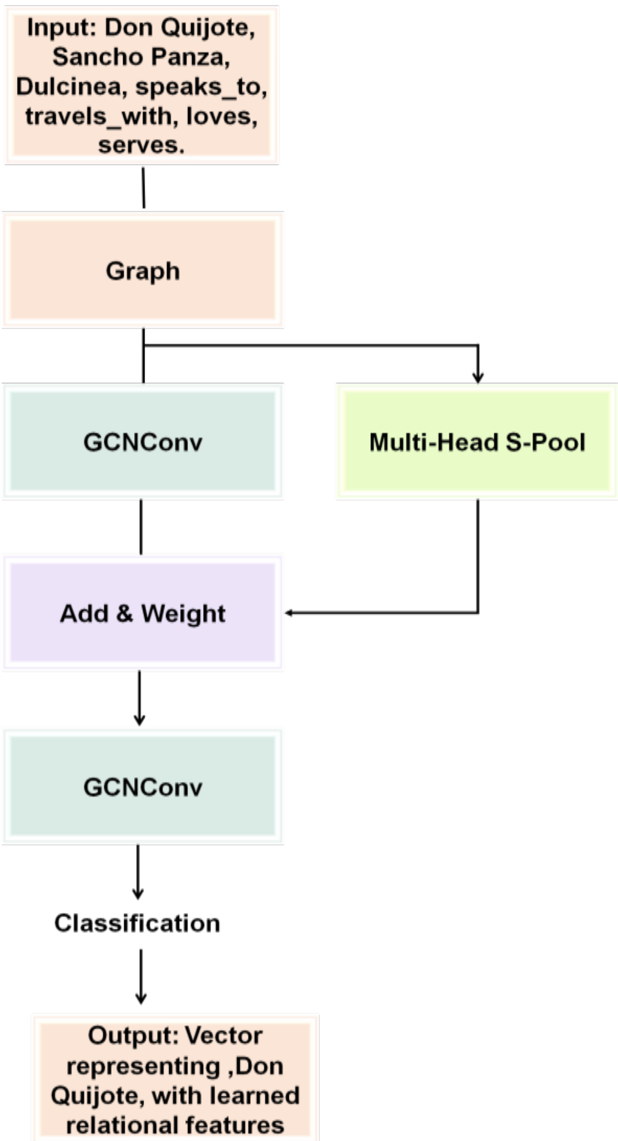

Figure 2. Block Diagram of a Graph Processing Module

In equation (2), denotes that Y is the output or hidden node features, X includes the input features, and W is the learnable weight matrix that transforms them. The adjacency matrix with self-loops, represented by B ˜= B+I, ensures that every node has its unique features during propagation. A non-linear activation function called σ makes it possible for the model to represent intricate interactions. By applying (B˜WX), features are aggregated and propagated throughout the graph, resulting in expressive and instructive node embeddings.

To express narratives more meaningfully, the KGF aims to integrate graph-based reasoning with linguistic modeling. It uses dynamic GCN layer integration to guarantee contextual consistency across entities and relationships. Finally, the methodology results in multilingual literary KGs that are more interpretable, scalable, and semantically coherent.

Table 2. Hyperparameters for Training and GCN Configuration

Hyperparameter	Value
Learning Rate	0,001
Optimizer	Adam
Batch Size	32
Number of Training Epochs	50
GCN Number of Layers	3
GCN Hidden Dimensions	64
GCN Dropout Rate	0,5

RESULTS

To enable data-driven literary analysis, a KGF automated for the extraction and visualization of character relationships in Spanish literature is being developed. Table 2 shows an arrangement for an effective way to present research.

Table 3. Overall system configurations

Components	Details
Operating System	Windows 11 Pro
Processor	AMD Ryzen 9 5900X (12-Core, 3.7 GHz)
RAM	64 GB DDR4
Programming Language	Python 3.11.4
Deep Learning Framework	PyTorch
Libraries Used	Transformers (mBERT), SpaCy, Matplotlib, RDFlib
Storage Capacity	2 TB NVMe SSD
Visualization Tools	NetworkX, Gephi

Evaluation metrics

The key performance measures for the KG-based character connection extraction methodology are precision, recall, and F1-scores. They evaluate the correctness of character interactions that were identified, as well as the quality and interpretability of character network visualizations of the Spanish literature corpus, and to what extent entity and relation extraction succeed in representing narrative linkages. Higher metrics indicate a reliable representation of a narrative's character structure for computer analyses, reliable mapping of relations, and reliable detection of entities.

Precision: indicates if the approach accurately maps relationships and detects significant character interactions while avoiding superfluous or duplicate connections in the KG. Recall: determines whether

The KG includes all the genuine interactions of character and interconnections of the plot, ensuring that no significant relationships are omitted. F1-score: Measures the efficiency of the approach in the reduction of incorrect or redundant links and records the actual interactions between characters through the balance of accuracy and recall.

Analysis of the KGF method's performance

The 3D Waterfall Chart shows the frequency and intensity of the interrelation of characters in the Spanish literature chapters. The interactions between 0 and 20 are shown in the Z-axis, and the relationship in the

https://doi.org/10.56294/mr2025226 ISSN: 2953-4577

Y-axis categories, such as conflict, emotional, influence, and romantic, and the X-axis is chapters. This graphic shows the relationship development in the story. Figure 3 depicts a 3D Waterfall Chart of interaction frequency, which was prepared with the KGF.

The bubble plot, in Spanish literature, is used to show the relationship between characters in the chapters.

The x-axis will represent the chapters (1–74), and the y-axis will illustrate the frequency of the interaction (1–20). The color and the size of bubbles indicate confidence levels (0,86 to 0,98) of relationships identified.

This graph helps to find the significant relationships between characters, repetitive interactions, and the intensity of the narrative. As a result, the patterns of relationships between characters generated by the KGF are presented in figure 4.

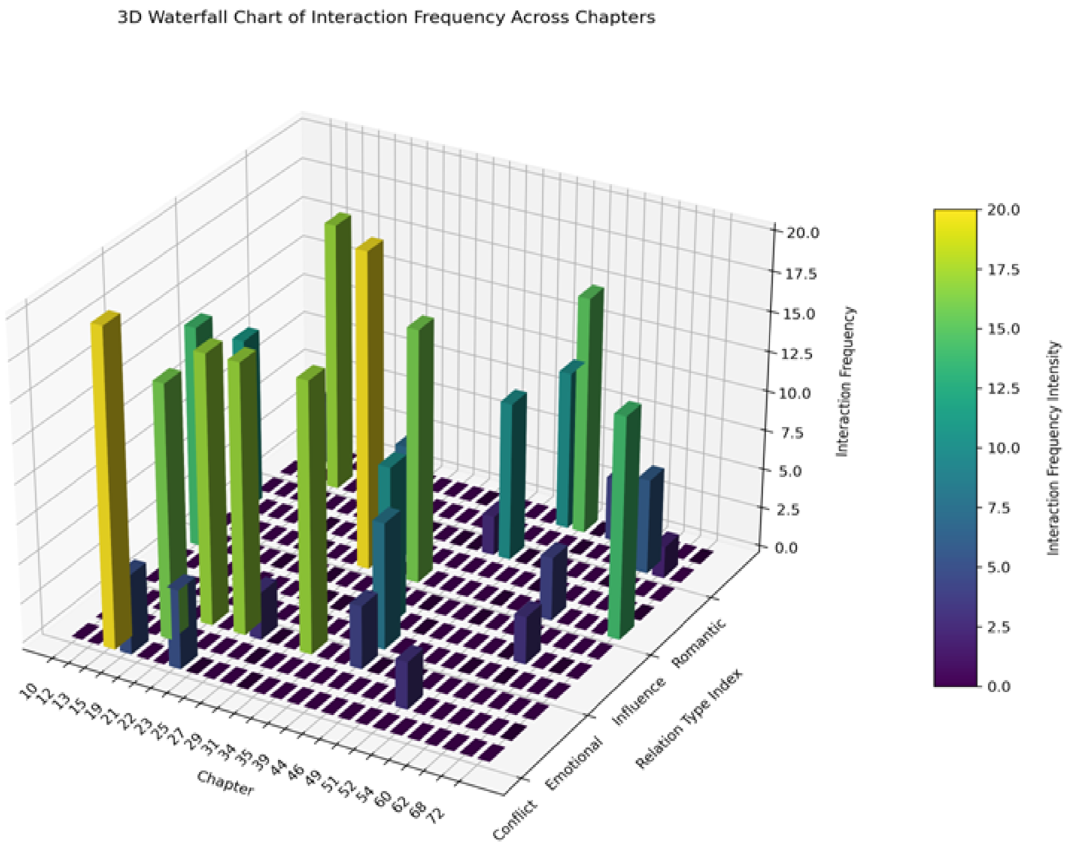

Figure 3. 3D Waterfall Chart of Interaction Frequency Across Parameter

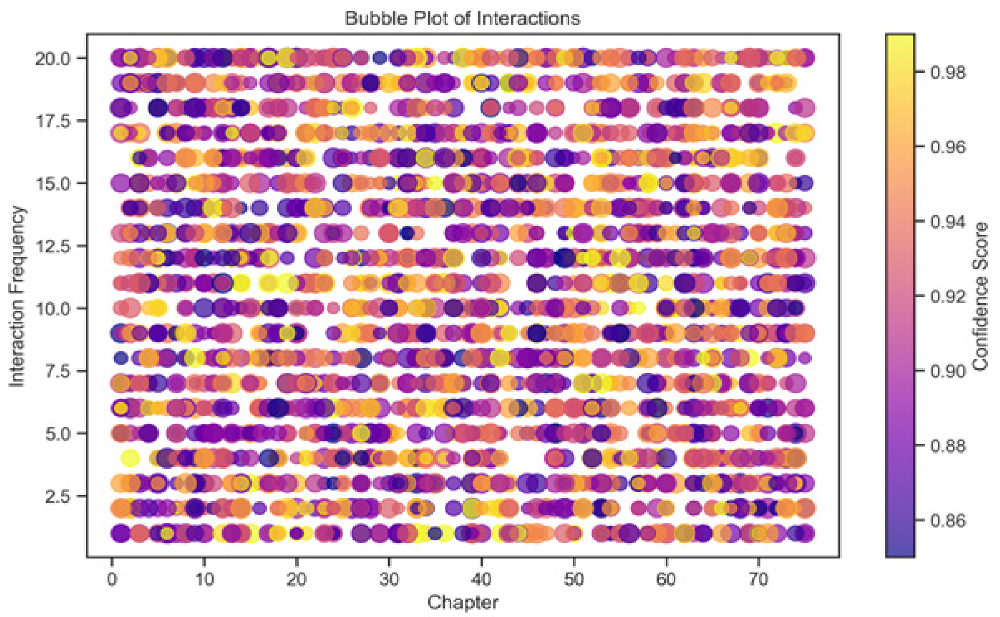

Figure 4. Bubble Plot of Interaction Frequency and Confidence Score Distribution

https://doi.org/10.56294/mr2025226 ISSN: 2953-4577

A contour plot is used to visualize the relationships between characters across the chapters of Spanish literature. The x-axis is the range 170, and the y-axis is the interaction frequency (2,520). Low to color gradients between blue and red represent high confidence scores (0.751.15). This visualization facilitates the analysis of the proposed Knowledge Graph format by helping to evaluate the intensity of character interaction and linkages of narratives. The Relationship Character Relationships across the Contour Plot Figure 5 shows the chapters generated by the proposed KGF technique.

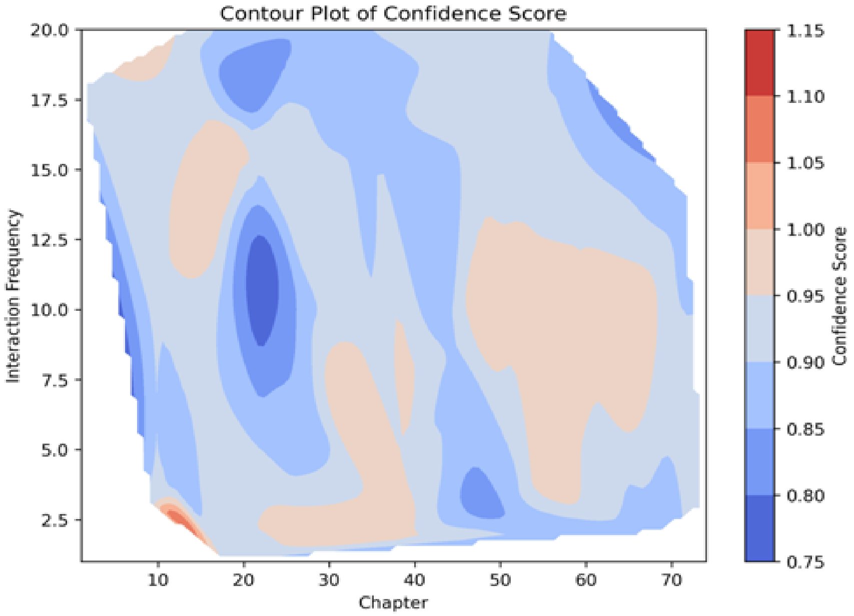

Figure 5. Contour Plot of Confidence Score as a Function of Interaction Frequency and Narrative Progression

The Confidence Score's distribution across two independent variables, Chapter (x-axis: $10–$70$) and Interaction Frequency (y-axis: $2,5-$20$), is displayed in the 3D Contour / Iso-Surface Plot. The third dimension and color gradient ($0,795$-$1,020$) is the Confidence Score. By highlighting the chapters where the GCNs provide the highest certainty in discovered character relationships, it aids in evaluating the stability and caliber of the KG structure. Consequently, the KGF approach was used to construct the 3D Contour / Iso-Surface Plot of character relationship confidence, which is displayed in figure 6.

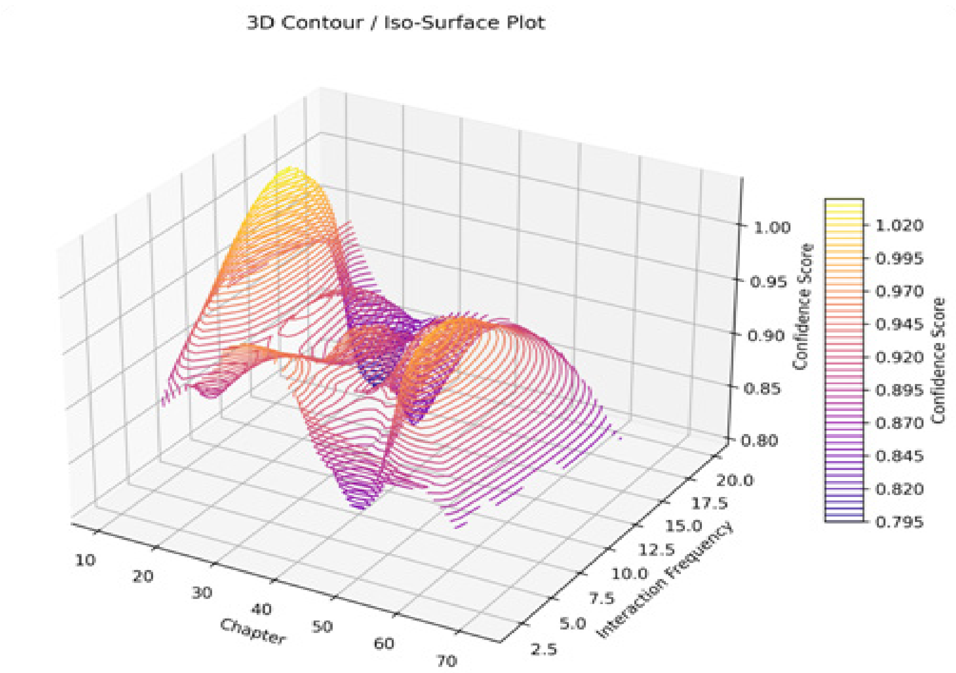

Figure 6. 3D Contour/Iso-Surface Plot of Confidence Score and Interaction Frequency

In the 3D Scatter Plot of Character Interactions, four variables are visualized: Confidence Score (z-axis, 0,850–1,000), Interaction Frequency (y-axis, 0–20), and Chapter (x-axis, 0–80). Sentence Length (6–14 words), the fourth dimension, is color-coded. This plot aids in evaluating the quality (confidence score) and context (sentence length) of relationships found throughout the story. Consequently, the 3D Scatter Plot of Character Interactions produced by the KGF method is displayed in figure 7.

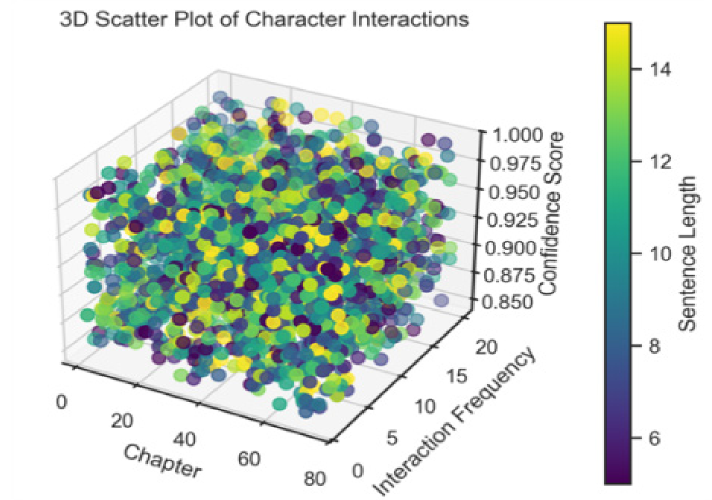

Figure 7. 3D Scatter Plot of Character Interactions Across Four Numeric Variables

The distribution of Interaction Frequency (y-axis, 0-$500+$) by Relation Type (color-coded, e.g., Conflict, Cooperation, Emotional) throughout Chapter (x-axis, 0-75) is visualized using the Stacked Area Plot. The KG approach of narrative structure analysis is supported by this plot, which shows how the general intensity and particular kinds of character relationships change throughout the course of the story. As a consequence, the Stacked Area Plot by Relation Type is produced by figure 8.

Relationships and distributions between four character interaction factors are displayed using the Pair Plot, known as a scatter plot matrix: Confidence Score (0,85–1,00), Sentence Length (6–14), Interaction Frequency (0–20), and Chapter (0–75). This figure aids in locating any possible dependencies or correlations between the narrative context and the caliber of the connections that the KG model found. Consequently, the Pair Plot of Numerical Features produced by the KGF method is displayed in figure 9.

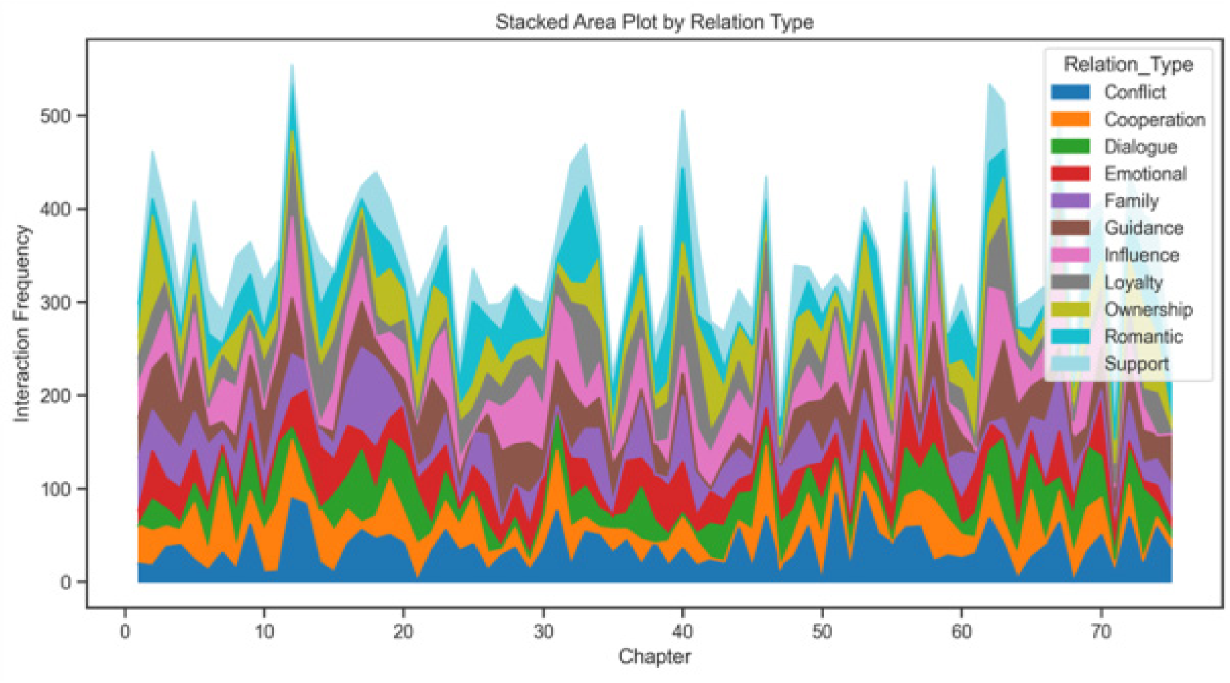

Figure 8. Stacked Area Plot of Interaction Frequency by Relation Type Across Narrative Progression

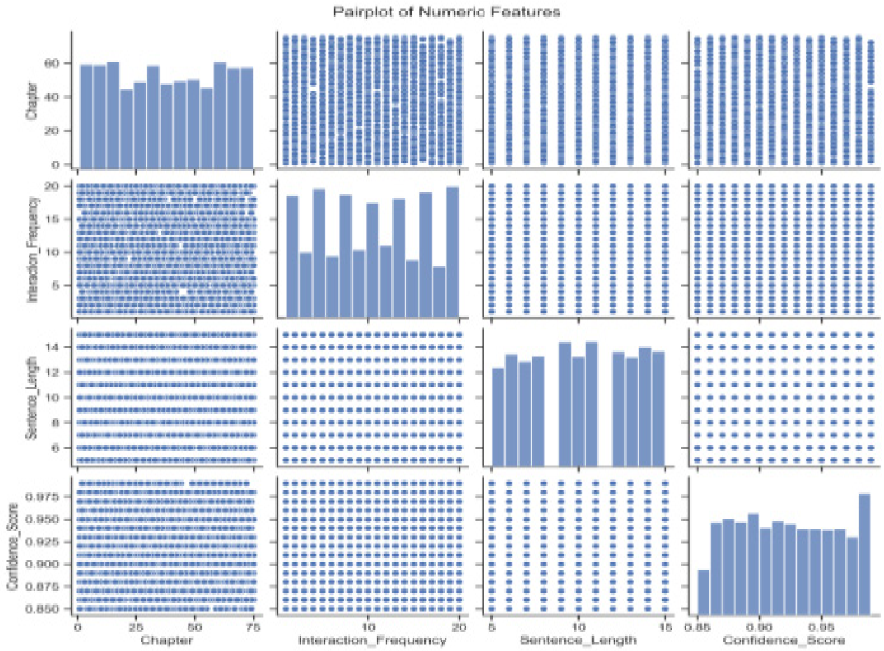

Figure 9. Pairplot of Numeric Features: Interaction Frequency, Sentence Length, and Confidence Score

The KGs direct character relationships are visualized in the Network Graph. Node colors might indicate a measure such as centrality, and each node represents a character arranged in a circle. Interactions are represented by edges, which are lines that connect text. The colors of these edges indicate the Confidence Score of the relationship that was found (for example, low purple to high yellow). This arrangement emphasizes relational confidence and general connectedness in figure 10 (a) emphasizing community detection and structural relationships.

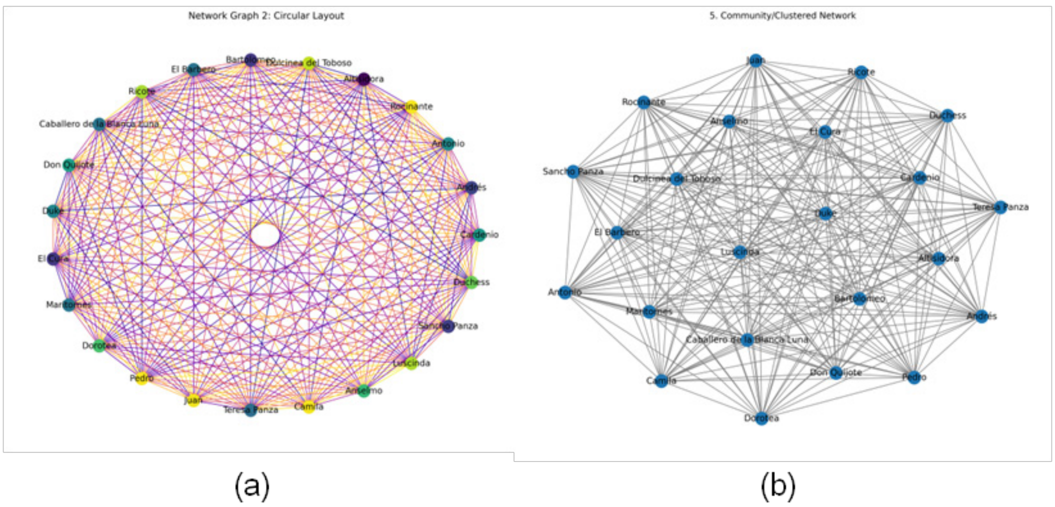

Figure 10. Comparative Visualization of Network Structure and Connectivity (a) Circular Layout (b) Force-Directed Community Layout

The Community/Clustered Network places a strong emphasis on community identification and structural groupings to illustrate character links. Characters are represented by each blue node, and their identified narrative links are indicated by their edges. Based on the KG, this condensed view aids in locating coherent clusters or groups of characters that engage in more frequent or intense interactions with one another, exposing underlying social structures or sub-narratives in the text. The Network Graphs of Character Relationships produced by the KGF method are displayed in figure 10 (b) which emphasizes community detection and structural relationships.

Comparative analysis

Character associations in Spanish literature are being extracted and shown using an automated KGF, which will facilitate data-driven literary research. A number of current techniques, such as K&I[19] and RDF triple extractors[19], were contrasted with the KGF. The performance of the proposed and existing approaches within the KGF was evaluated using precision, recall, and F1-score (figure 11). Table 4 shows a performance comparison of the KGF and existing techniques across these metrics. The statistical significance of performance gains between the suggested KGF and GCN models is shown in Table 5. Across all metrics, the suggested KGF model performs better than the GCN model. Accuracy (0,91), F1-score (0,87), precision (0,89), and recall (0,86) all show notable improvements, which are backed by p-values that indicate statistical significance and confidence intervals that show steady performance gains.

Table 4. Evaluation metrics values for existing and proposed

Method	Precision	Recall	F1-Score
K&I[19]	46,20	56,15	50,69
RDF triple extractors[19]	55,85	46,96	51,02
KGF[proposed]	85,01	87,06	87,96

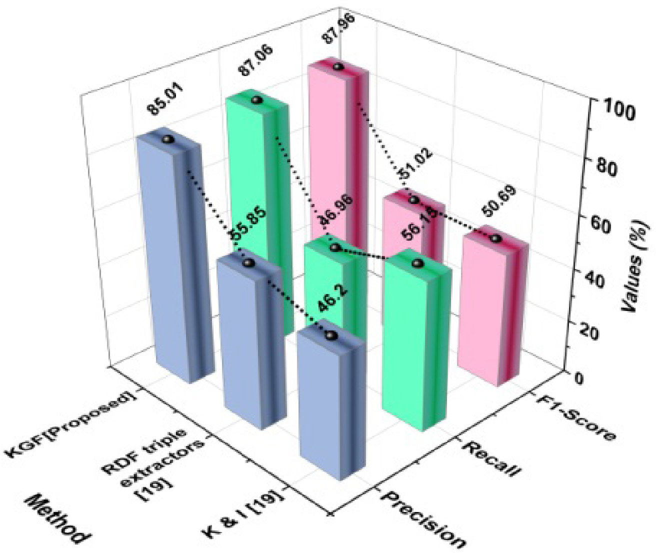

Figure 11. 3D Comparative Bar Chart of Performance Metrics by Extraction Method

Table 5. Statistical Significance of Performance Improvements

Metric	KGF [proposed]	GCN	p-value	95 % Confidence Interval
Accuracy	0,91	0,88	0,02	[0,01, 0,03]
F1-Score	0,87	0,84	0,03	[0,02, 0,05]
Precision	0,89	0,85	0,01	[0,02, 0,04]
Recall	0,86	0,83	0,04	[0,01, 0,03]

DISCUSSION

This study explores the application of the Knowledge Graph Framework (KGF) in conjunction with Graph Convolutional Networks (GCNs) and multilingual BERT embeddings (mBERT) within Spanish literature to investigate the relationships between characters. Existing technologies, such as K&I and RDF triple extractors, which are rule-based, fail to fully capture the complex interactions and emotional nuances among characters and have primarily been applied to classic literature, thus not effectively addressing modern or experimental works. Our research found that combining multilingual mBERT embeddings with GCNs significantly enhances the effectiveness of capturing character relationship data, revealing deeper narrative patterns. This indicates that data-driven KGF methodologies provide an effective alternative for traditional literary studies.

From a broader perspective, this approach not only enhances the understanding of character dynamics and narrative structures but also promotes the development of digital humanities and content analysis fields. With its ability to process dynamically and adapt to evolving texts in real time, KGF holds the potential to more accurately capture the subtle interactions among characters, thereby expanding its applicability.

However, this study also has limitations. The current KGF methods remain vague when dealing with emotions and metaphors, as well as dynamic texts, lacking specific adaptive mechanisms. Therefore, future research should focus on improving the real-time processing capability and adaptability of KGF, in order to more accurately identify complex literary contexts. This will provide more reliable tools and perspectives for deeper text analysis and literary research.

CONCLUSIONS

Spanish literature Character relationships were automatically mined, organized, and visualized through an automated pipeline, which generated KGs to analyze literary relationships. In this research, data collection, multilingual mBERT embeddings as a pre-processing tool, and GCNs to model complex character dynamics were used to research character associations in Spanish literature. The fact that it is able to depict complex dynamics with the highest degree of accuracy as it has been given the score of 85,01 % of precision, 87 % of recall, and 87,96 % of F1-score, was an incredible outcome of the KGF. The KGF model has the high potential in improving the character relationship analysis of the Spanish literature through the two forms of GCNs and multilingual mBERT embeddings. The system provides a more sophisticated insight into the problems than the old techniques since it is able to reveal more concerning the underbelly of the character interaction, emotional development, and the plot. The method plays a significant role in the research of literature especially regarding the element of refreshing the complicated relations between the characters. The following stage of the research must be dedicated to the framework change of the dynamic and real-time texts so that the further improvement of its functionality could be guaranteed, not to mention that its further implementation in various spheres of literature should be ensured. Finally, improved KGF would enhance the research of literature and make the process of making decisions in various industries smarter because it would provide a more advanced and evidence-based method of comprehending the peculiarities of narratives.

BIBLIOGRAPHIC REFERENCES

1. Sepasgozar S.M., Khan A.A., Smith K., Romero J.G., Shen X., Shirowzhan S., et al. BIM and digital twin for developing convergence technologies as future of digital construction. Buildings. 2023 Feb;13(2):441. https://doi.org/10.3390/buildings13020441

2. Greco D., Osborne F., Pusceddu S., Reforgiato Recupero D. Modelling big data platforms as knowledge graphs: the data platform shaper. J Big Data. 2025 Dec;12(1):64. https://doi.org/10.1186/s40537-025-01094-w

3. Hofer M., Obraczka D., Saeedi A., Köpcke H., Rahm E. Construction of knowledge graphs: Current state and challenges. Information. 2024 Aug;15(8):509. https://doi.org/10.3390/info15080509

4. Georges P., Seckin A. Music information visualization and classical composers discovery: an application of network graphs, multidimensional scaling, and support vector machines. Scientometrics. 2022 May;127(5):2277-311. https://doi.org/10.1007/s11192-022-04331-8

5. Li J., Xue E. Dynamic interaction between student learning behaviour and learning environment: Meta-analysis of student engagement and its influencing factors. Behav Sci (Basel). 2023 Jan;13(1):59. https://doi.org/10.3390/bs13010059

6. Chen Y., Li H., Li H., Liu W., Wu Y., Huang Q., et al. An overview of knowledge graph reasoning: key technologies and applications. J Sens Actuator Netw. 2022 Nov;11(4):78. https://doi.org/10.3390/jsan11040078

7. Borrego A., Dessi D., Hernández I., Osborne F., Recupero D.R., Ruiz D., et al. Completing scientific facts in knowledge graphs of research concepts. IEEE Access. 2022;10:125867-80. https://doi.org/10.1109/ACCESS.2022.3220241

8. Rios-Alvarado A.B., Martinez-Rodriguez J.L., Garcia-Perez A.G., Guerrero-Melendez T.Y., Lopez-Arevalo I., Gonzalez-Compean J.L. Exploiting lexical patterns for knowledge graph construction from unstructured text in Spanish. Complex Intell Syst. 2023 Apr;9(2):1281-97. https://doi.org/10.1007/s40747-022-00805-7

9. Chessa A., Fenu G., Motta E., Osborne F., Recupero D.R., Salatino A., et al. Data-driven methodology for knowledge graph generation within the tourism domain. IEEE Access. 2023;11:67567-99. https://doi.org/10.1109/ACCESS.2023.3292153

10. Hao X., Ji Z., Li X., Yin L., Liu L., Sun M., et al. Construction and application of a knowledge graph. Remote Sens (Basel). 2021 Jun;13(13):2511. https://doi.org/10.3390/rs13132511

11. Lovera F.A., Cardinale Y.C., Homsi M.N. Sentiment analysis in Twitter based on knowledge graph and deep learning classification. Electronics. 2021 Nov;10(22):2739. https://doi.org/10.3390/electronics10222739

12. Hamroun M., Sauveron D. A Hybrid Deep Learning and Knowledge Graph Approach for Intelligent Image Indexing and Retrieval. Appl Sci (Basel). 2025 Oct;15(19):10591. https://doi.org/10.3390/app151910591

13. Chudasama Y., Huang H., Purohit D., Vidal M.E. Towards interpretable hybrid ai: Integrating knowledge graphs and symbolic reasoning in medicine. IEEE Access. 2025;13:39489-509. https://doi.org/10.1109/ACCESS.2025.3529133

14. Escudero-Arnanz Ó., Soguero-Ruiz C., Marques A.G. Explainable spatio-temporal GCNNs for irregular multivariate time series: Architecture and application to ICU patient data. IEEE Trans Signal Inf Process Netw. 2025;11:1286-301. https://doi.org/10.1109/TSIPN.2025.3613951

15. Guzman-Monteza Y., Fernandez-Luna J.M., Ribadas-Pena F.J. IV-Nlp: A Methodology to Understand the Behavior of DL Models and Its Application from a Causal Approach. Electronics. 2025 Apr;14(8):1676. https://doi.org/10.3390/electronics14081676

16. Razzaq K., Shah M. Machine learning and deep learning paradigms: From techniques to practical applications and research frontiers. Computers. 2025 Mar;14(3):93. https://doi.org/10.3390/s25082405

17. Gagliardi I., Artese M.T. Exploring and visualizing multilingual cultural heritage data using multi-layer semantic graphs and transformers. Electronics. 2024 Sep;13(18):3741. https://doi.org/10.3390/electronics13183741

18. Amur Z.H., Kwang Hooi Y., Bhanbhro H., Dahri K., Soomro G.M. Short-text semantic similarity (stss): Techniques, challenges and future perspectives. Appl Sci (Basel). 2023 Mar;13(6):3911. https://doi.org/10.3390/app13063911

19. Rios-Alvarado A.B., Martinez-Rodriguez J.L., Garcia-Perez A.G., Guerrero-Melendez T.Y., Lopez-Arevalo I., Gonzalez-Compean J.L. Exploiting lexical patterns for knowledge graph construction from unstructured text in Spanish. Complex Intell Syst. 2023 Apr;9(2):1281-97. https://doi.org/10.1007/s40747-022-00805-7

FINANCING
supported by Jiangsu Provincial University Philosophy and Social Science Research Project (2023SJYB0246).

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION
Data curation: Kunfei Li.
Methodology: Kunfei Li.
Drafting - original draft: Kunfei Li.
Writing - proofreading and editing: Kunfei Li.

ANNEXES

Abbreviation	Full Form	Abbreviation	Full Form
AI	Artificial Intelligence	XST-GCNN	eXplainableSpatio-Temporal Graph Convolutional Neural Network
VISE	Visual Interpretable Symbolic Engine	MTS	Multivariate Time Series
SA	sentiment analysis	NLP	Natural Language Processing
DL	deep learning	MR	mixed reality
ViT	Vision Transformer	PD	Parkinson's disease
Acc	Accuracy	XGBoost	Extreme Gradient Boosting
Rec	Recall	HuBERT	Hidden-Unit BERT
F1-sce	F1-score	SVM	Support Vector Machine
Pre	Precision	SPARQL	SPARQL Protocol and RDF Query Language
GCNs	Graph Convolutional Networks	RDF	Resource Description Framework
STSS	short-text semantic similarity	OWL	Web Ontology Language
SLR	systematic literature review	AdaBoost	Adaptive Boosting
KL	Kullback-Leibler	K&I	KG extraction from plain text in Spanish
maP	mean Average Precision	DBpedia	Database of Wikipedia
ZMTD	Zimperium Mobile Threat Defense		