

# Beyond Topology: Generative Node Importance Estimation via Structure-Guided Semantic Reasoning

Kuofei Fang, Siyan Wu, Guo Yu, Bin Wu\*

Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia,  
Beijing University of Posts and Telecommunications  
{coffeyf, aiyana, ygbro666, wubin}@bupt.edu.cn

## Abstract

Node Importance Estimation (NIE) in Knowledge Graphs (KGs) aims to quantify the significance of entities, serving as a pivotal instrument for deciphering the latent mechanisms of social dynamics. However, existing methods are typically confined to supervised paradigms and rely heavily on topological aggregation, resulting in limited generalization capability. To address these challenges, we propose GenNIE, the first end-to-end generative reasoning framework for NIE. Specifically, GenNIE leverages Large Language Models (LLMs) integrated with topological information to generate precise importance scores for entities in KGs. Furthermore, GenNIE introduces a Global-Structural Graph Perception mechanism to empower the LLMs with holistic graph cognition. Extensive experiments demonstrate the performance superiority of GenNIE and its robust generalization across diverse domains. Our code is available at <https://github.com/CoffeyF/GenNIE.git>.

## 1 Introduction

Knowledge Graphs are real-world semantic networks endowed with rich textual attributes, as illustrated in Figure 1. Large-scale KGs such as Freebase (Bollacker et al., 2008) and DBpedia (Lehmann et al., 2015) have proven to be indispensable resources, serving as a fundamental tool for Computational Social Science (CSS) to empower intelligent decision-making.

Node Importance Estimation (NIE) stands as a fundamental task in KG applications, aiming to quantify the significance of entities and identify critical nodes by leveraging the rich knowledge embedded in KGs (Cao et al., 2025). It is widely applied across diverse disciplines, including economics (Battiston et al., 2012), sociology (Wang et al., 2019; Zhao et al., 2022), and biological

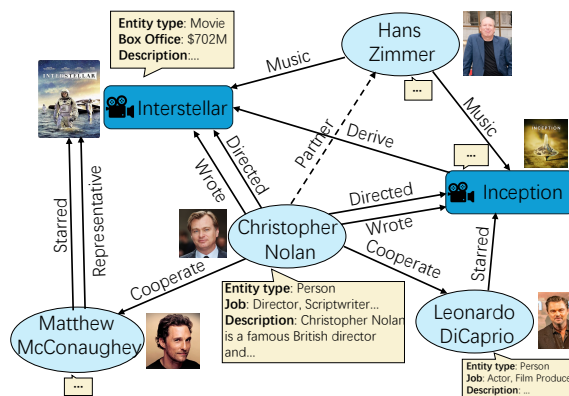


Figure 1: An example knowledge graph on film industry.

sciences (Barabási et al., 2011). Early works relied on calculating centrality metrics to estimate node importance. However, these metrics often diverge from the actual importance of entities in real-world KGs and become computationally impractical on large-scale networks. Subsequent studies employed Graph Neural Networks (GNNs) and their variants to capture the correlation between node importance and topological patterns through neighborhood aggregation. Recent advancements have incorporated textual attributes to refine node importance estimation. However, these methods typically operate under a supervised paradigm that overfits domain-specific structural and semantic patterns, which results in poor generalization. Furthermore, the intricate fusion of these modalities often leads to semantic entanglement, hindering effective reasoning.

To address these challenges, we propose GenNIE, an end-to-end generative reasoning framework that reformulates the NIE task from traditional topological aggregation into a generative inference paradigm.

Specifically, GenNIE introduces an LLM-adaptive topology injection mechanism. By employing a Mixture-of-Experts (MoE) projector and structure-guided cross-attention, we integrate

\* Corresponding author.

relation-aware node embeddings derived from the graph encoder within the LLM’s semantic space.

Furthermore, unlike general KG tasks, NIE requires a comprehensive perception of global topology. Crucially, the injected structural information must be highly discriminative to distinguish nodes at varying importance levels. Meanwhile, the LLM must internalize a robust scoring standard derived from both global textual and topological information. To achieve this, GenNIE features a global-structural graph perception mechanism. Globally, a learnable soft prompt is utilized to encapsulate overall scoring patterns in a knowledge graph during fine-tuning. Structurally, we devise a self-supervised contrastive learning strategy. By aligning the embeddings of nodes with similar importance, this strategy explicitly reconstructs the topological importance manifold of entities within the feature space.

Finally, the learnable soft prompt, injected structural information, and textual attributes are composed as input, upon which the LLM reasons to generate importance scores. The framework is optimized through Hybrid Parameter-Efficient Fine-Tuning (Hybrid PEFT) with LLM backbone frozen.

We conduct extensive experiments on four real-world KGs, demonstrating the effectiveness of GenNIE. Moreover, the transfer experiments verify its superior generalization capability across domains. Finally, analysis confirms that the graph perception mechanism significantly augments the LLM’s sensitivity to complex topological structures. Our contributions are summarized as follows:

- We propose GenNIE, an end-to-end generative reasoning framework for NIE, which demonstrates robust generalization capabilities compared to existing supervised methods.
- We introduce a graph perception mechanism to augment the LLM’s sensitivity to topological structures on NIE tasks.
- We demonstrate GenNIE’s practical utility through a case study on historical character influence, underscoring its potential for CSS.

## 2 Related Work

**Node Importance Estimation.** Early research on NIE primarily approximates node importance by calculating topological statistics (Ren and Lü, 2014) such as Degree Centrality (Albert et al.,

2000), K-shell (Kitsak et al., 2010), and PageRank (Page et al., 1999). These traditional methods heavily rely on pure topological statistics, completely discarding rich textual attributes. Furthermore, their computational costs scale prohibitively with the growing size of modern knowledge graphs.

Subsequently, GNN-based methods have been applied to the NIE task. **GENI** (Park et al., 2019) utilizes graph attention mechanisms to aggregate neighboring importance scores. **MultiImport** (Park et al., 2020) infers latent node importance through multi-signal fusion. **HIVEN** (Huang et al., 2022) captures local structural patterns inherent in graphs. **RGTN** (Huang et al., 2021) deeply fuses semantic attributes with structural features. **SKES** (Chen et al., 2024b) quantifies the difference between a node and a sufficiently uninformative reference. **MVPL** (Ma et al., 2026) introduces multi-view contrastive learning. **AGCL** (Shu et al., 2025) combines attention mechanisms with contrastive learning to model importance characteristics. **LI-CAP** (Zhang et al., 2025) leverages label-informed sampling strategies to incorporate real-world prior knowledge. **EASING** (Chen et al., 2025) models informative distributions to regularize uncertainty for unannotated data. **LENIE** (Lin et al., 2025) integrates LLMs to enhance text attributes. While these GNN-based approaches attempt to fuse textual attributes and topological structures, their inherent lack of deep language understanding often leads to severe semantic entanglement. In contrast, our method bypasses convoluted fusion mechanisms by directly leveraging LLMs to comprehend textual prompts and explicitly injecting topological features into the semantic space.

### Large Language Models on Graph Mining.

With the rapid advancement of LLMs (Minaee et al., 2025), they have been widely applied to graph mining tasks including reasoning, node classification, and link prediction (Jin et al., 2024; Wang et al., 2025). For instance, **GraphLLM** (Chai et al., 2025), **GraphGPT** (Tang et al., 2024), and **GNP** (Tian et al., 2024) enhance LLMs with graph embedding on node classification and link prediction tasks. To further enhance reasoning, **GraphWiz** (Chen et al., 2024a) employs direct preference optimization on graph-specific instruction data, and **MuseGraph** (Tan et al., 2026) utilizes compact graph descriptions to facilitate instruction tuning.

### 3 Preliminaries

**Definition 1 (Knowledge Graph).** A knowledge graph is formally defined as  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ . Here,  $\mathcal{V}$  denotes the set of nodes (entities) enriched with textual attributes,  $\mathcal{R}$  denotes the set of relation types, and  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{R} \times \mathcal{V}$  is the set of triplets.

**Definition 2 (Node Importance).** The importance of a node  $v_i \in \mathcal{V}$  is quantified by a numerical score  $s_i \in \mathbb{R}^+$ . This non-negative scalar reflects the intrinsic significance, popularity, or centrality of the entity within the knowledge graph.

**Definition 3 (Node Importance Estimation).** Given a KG  $\mathcal{G}$  and a subset of nodes  $\mathcal{V}_L \subset \mathcal{V}$  associated with known importance scores  $\mathcal{S}_L = \{s_i \mid v_i \in \mathcal{V}_L\}$ , the goal of NIE is to learn a mapping function  $\Phi : \mathcal{V} \rightarrow \mathbb{R}^+$ . This function aims to accurately predict the latent importance scores for the remaining unlabeled nodes.

## 4 Methodology

As shown in Figure 2, GenNIE is a framework that uses LLMs to generate node importance scores. First, we introduce adaptive topology injection to integrate aggregated node structural information into the semantic space. Then, we introduce the graph perception mechanism to enhance the holistic topology perception capability of the LLMs through supervised contrastive learning and a learnable soft prompt. Finally, we adopt the Hybrid PEFT strategy to jointly optimize the overall architecture in an end-to-end manner.

### 4.1 Topological Information Injection

#### 4.1.1 Node Information Aggregation

Firstly, we utilize the Node2Vec algorithm (Grover and Leskovec, 2016) to get the initial embeddings for all nodes in knowledge graph, denoted as  $\mathbf{H}^{(0)} = \{\mathbf{h}_v^{(0)} \mid v \in \mathcal{V}\}$ . Then, we aggregate structural information from the  $L$ -hop local subgraph of each entity to capture its intrinsic topological characteristics. Considering that different types of relations (predicates) contribute unequally to the importance of an entity, we employ a predicate-aware attention mechanism (Huang et al., 2021) to explicitly model these relation distinctions. Specifically, we iteratively update node features by stacking  $L$  predicate-aware multi-head attention layers. For the  $k$ -th attention head at the  $l$ -th layer, the normalized attention weight  $\alpha_{(u,v),r}^{k,l}$  for node  $v$  re-

garding neighbor  $u$  under relation  $r$  is defined as:

$$\alpha_{(u,v),r}^{k,l} = \text{Softmax}_{\substack{u' \in \mathcal{N}_v \\ r' \in \mathcal{R}_{u'v}}} \left( \frac{(\mathbf{W}_Q^{k,l} \mathbf{h}_v^{(l)})^\top (\mathbf{W}_K^{k,l} \mathbf{h}_{u'}^{(l)})}{\sqrt{d_k}} + [\mathbf{W}_{RP}^l \mathbf{p}_r]_k \right), \quad (1)$$

where  $\mathcal{N}_v$  is the neighbor set of  $v$ , and  $\mathcal{R}_{u'v}$  is the set of relations connecting  $u'$  and  $v$ .  $\mathbf{W}_{(\cdot)}$  represents the learnable projection matrices, and  $\mathbf{p}_r$  is the learnable embedding vector for relation  $r$ .  $[\cdot]_k$  is the scalar bias corresponding to the  $k$ -th head. We add a self-loop to each node with a relation type  $r_{\text{self}}$ , ensuring  $v \in \mathcal{N}_v$ .

Based on the normalized attention weights, we perform a weighted aggregation over the heterogeneous neighborhood. The subsequent feature update follows the standard Transformer encoder architecture, which explicitly employs residual connections and a Feed-Forward Network (FFN). We abstract this general non-linear transformation as:

$$\mathbf{h}_v^{(l+1)} = \text{TransformerBlock} \left( \mathbf{h}_v^{(l)}, \{\alpha_{(u,v),r}^{k,l}\} \right), \quad (2)$$

After  $L$  stacked layers, we obtain the structural embedding  $\mathbf{h}_v^{\text{struct}} = \mathbf{h}_v^{(L)}$ .

#### 4.1.2 LLM-Adaptive Node Topology Injection

Directly concatenating aggregated node embeddings with textual tokens often hinders effective reasoning due to the inherent modality gap between structural and semantic spaces. To bridge this gap, we design an LLM-adaptive topology injection strategy comprising an MoE projector and a structure-guided cross-attention mechanism.

**MoE Projector.** To achieve fine-grained modality alignment, we introduce an MoE projector that utilizes multiple expert networks to capture distinct feature subspaces within the topology-semantic mapping. Specifically, the projector comprises  $N$  expert networks  $\{\mathcal{E}_i\}_{i=1}^N$ , where each expert  $\mathcal{E}_i$  is implemented as a two-layer MLP with GELU activation. Given the structural embedding  $\mathbf{h}_v^{\text{struct}}$ , we compute the routing probabilities  $\mathbf{g}$  via a gating network to derive the active expert indices  $\mathcal{T}$  (Top- $m$ ). The final output  $\mathbf{h}_v^{\text{proj}}$ , which aligns with the LLM hidden dimension, is generated through weighted aggregation of the selected experts:

$$\mathbf{g} = \text{Softmax}(\mathbf{R}_g \mathbf{h}_v^{\text{struct}}), \quad \mathcal{T} = \text{Top-}m(\mathbf{g}),$$

$$\mathbf{h}_v^{\text{proj}} = \sum_{i \in \mathcal{T}} \frac{g_i}{\sum_{j \in \mathcal{T}} g_j} \cdot \mathcal{E}_i(\mathbf{h}_v^{\text{struct}}), \quad (3)$$

where  $\mathbf{R}_g$  is the learnable gating matrix, and  $g_i$  denotes the gating probability of the  $i$ -th expert.

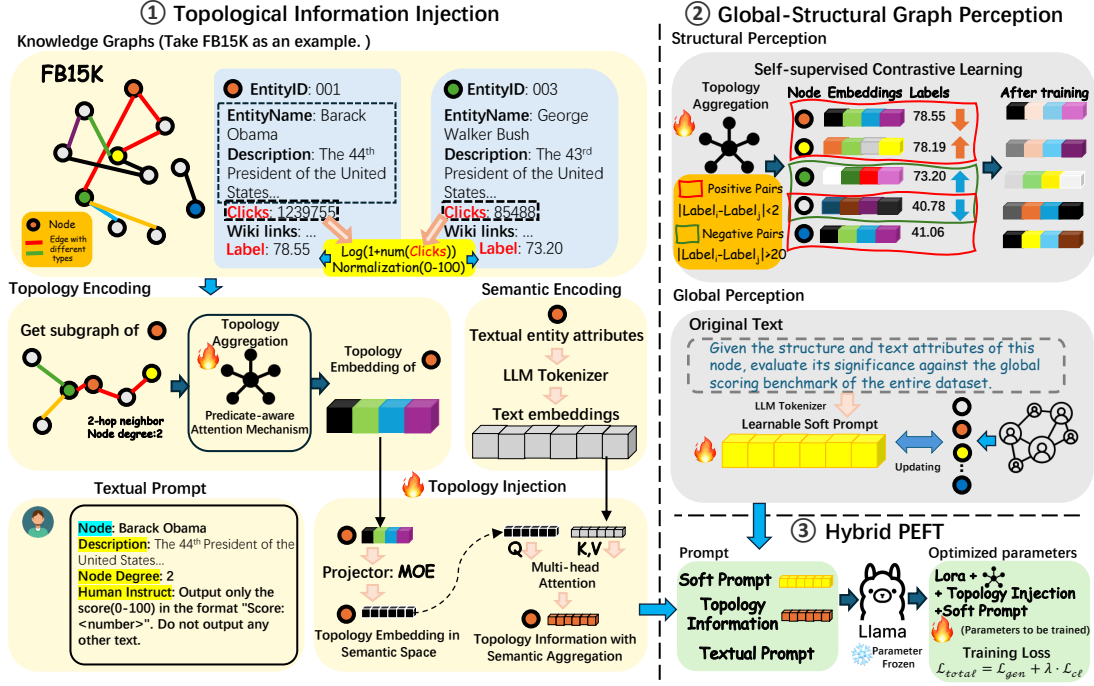


Figure 2: The overall architecture of GenNIE. Given a knowledge graph, it first obtains nodes embeddings and injects them into semantic space. Then with the help of graph perception mechanism, we enhance the LLMs’ capability of topology perception. Finally, we construct the input and train the model with hybrid fine-tuning.

**Structure-Guided Cross-Attention.** To achieve deep integration between the projected topological features and the entity’s textual attributes, we employ a structure-guided cross-attention mechanism. First, we leverage the pre-trained embedding layer of the LLM to encode the target entity’s textual attributes, yielding a sequence of text embeddings  $\mathbf{H}_v^{text}$ . Subsequently, we treat the structural embedding as the query to attend to textual features. The cross-modal fusion is computed as:

$$\mathbf{A} = \text{Softmax} \left( \frac{(\mathbf{U}_Q \mathbf{h}_v^{proj})^\top (\mathbf{U}_K (\mathbf{H}_v^{text})^\top)}{\sqrt{d_{head}}} \right), \quad (4)$$

$$\mathbf{h}_v^{aligned} = (\mathbf{U}_V (\mathbf{H}_v^{text})^\top) \mathbf{A}^\top,$$

where  $\mathbf{U}_{(\cdot)}$  represent the learnable projection matrices, and  $d_{head}$  is the dimension of the attention head. Through this mechanism, the model automatically aggregates relevant semantic features guided by the entity’s topological information. The final output  $\mathbf{h}_v^{aligned}$  is directly utilized in the subsequent LLM inference process.

## 4.2 Global-Structural Graph Perception

Although the adaptive topology injection effectively captures local semantics, the LLM still lacks macro-perception capabilities beyond the local scope. To address this, we introduce Global-Structural Graph Perception.

### 4.2.1 Global Graph Perception

Recent studies (Zhang et al., 2024b; Mao et al., 2022) demonstrate that incorporating learnable continuous vectors during the training can efficiently adapt models to specific downstream tasks. Inspired by this, we design a hybrid tuning strategy by introducing a set of global learnable prefixes at the beginning of the instruction sequence, named the learnable soft prompt, which serves as a set of globally shared task-level parameters. During training, these parameters are continuously optimized via gradient backpropagation across all samples. This mechanism enables the soft prompt to aggregate local signals from all nodes. Formally, for a given node  $v$ , we construct the hybrid input sequence as follows:

$$\mathbf{P}_{input} = [\mathbf{P}_{global}, \mathbf{h}_v^{aligned}, \mathbf{E}_v^{text}], \quad (5)$$

where  $\mathbf{P}_{global}$  denotes the soft prompt, and  $\mathbf{E}_v^{text}$  represents the token embeddings of the node’s textual attributes and the NIE task instructions.

### 4.2.2 Structural Graph Perception

To further enhance the structural distinguishability of LLMs regarding node importance levels and prevent representation degeneration of the graph encoder caused by text modality dominance, we de-

Dataset	Nodes	Edges	Edge Types	Target Node	Importance Metric	Nodes with Importance
<b>FB15K</b>	14,951	592,213	1,345	Entity	Page Views	14,105 (94.0%)
<b>TMDB5K</b>	76,926	359,780	12	Movie	Popularity	4,802 (6.2%)
<b>MUSIC10K</b>	22,986	80,272	8	Artist	Artist Hotness	4,214 (18.3%)
<b>DBLP</b>	249,903	2,428,250	6	Paper	Citation Count	100,000 (40.0%)

Table 1: Statistics of the knowledge graph datasets. **Target Node** specifies the particular category of entities subject to importance ranking. **Importance Metric** denotes the ground-truth standard used for evaluation.

sign a label-aware supervised contrastive learning strategy during the training process.

First, we define node pairs in the training batch with an absolute difference in ground truth importance scores below a specific threshold  $\delta_{pos}$  as *positive samples*. In NIE tasks, nodes with similar importance inherently possess similar topological characteristics. Conversely, node pairs with an absolute difference exceeding a threshold  $\delta_{neg}$  are treated as *negative samples*. Recognizing that the degree of difference varies among negative pairs (e.g., a gap of 10 points versus 50 points represents completely different hierarchical spans), we categorize negative samples into tiers based on the score gap and assign differentiated weights. Accordingly, we utilize a weighted InfoNCE loss function to perform adaptive feature optimization:

$$z(u, v) = \exp(\text{sim}(\mathbf{h}_u^{\text{struct}}, \mathbf{h}_v^{\text{struct}}) / \tau),$$

$$\mathcal{L}_{cl} = -\frac{1}{|\mathcal{P}|} \sum_{(i,j) \in \mathcal{P}} \log \frac{z(i, j)}{z(i, j) + \sum_{r \in \mathcal{M}_i} \omega_{i,r} \cdot z(i, r)}, \quad (6)$$

where  $\mathcal{P}$  denotes the set of positive node pairs,  $\mathcal{M}_i$  represents the set of negative samples for node  $i$ , and  $\tau$  is the temperature coefficient. Notably,  $\omega_{i,r}$  is a tier-based adaptive weight that increases in a step-wise manner with the absolute score difference  $|y_i - y_r|$ , enforcing stricter separation for pairs with larger hierarchical gaps.

### 4.3 Model Training and Optimization

Before training, we perform logarithmic smoothing and normalization on the raw entity importance labels  $s_i$  to derive the target score  $y_i$ :

$$y_i = 100 \cdot \frac{\log(s_i + 1) - \log(s_{\min} + 1)}{\log(s_{\max} + 1) - \log(s_{\min} + 1)}, \quad (7)$$

where  $s_{\min}$  and  $s_{\max}$  denote the minimum and maximum importance scores in the training set, respectively. This transformation maps metrics of varying magnitudes into the  $[0, 100]$  percentile interval, a range to which LLMs are semantically sensitive.

For each entity, we assemble the hybrid input sequence  $\mathbf{P}_{input}$  as described in Section 4.2.1 and

feed it into the LLM. During the training phase, we employ the teacher forcing strategy for end-to-end optimization. We serialize the target score  $y_i$  into a discrete token sequence and minimize the following generative cross-entropy loss  $\mathcal{L}_{gen}$ :

$$\mathcal{L}_{gen} = -\frac{1}{T} \sum_{t=1}^T \log P(y_t | \mathbf{P}_{input}, y_{<t}; \Theta), \quad (8)$$

where  $T$  is the target sequence length,  $y_t$  is the ground truth token at step  $t$  given the historical context  $y_{<t}$ , and  $\Theta$  denotes all learnable parameters.

Finally, we incorporate the structural contrastive loss  $\mathcal{L}_{cl}$  proposed in Section 4.2.2 into the joint optimization objective:

$$\mathcal{L}_{total} = \mathcal{L}_{gen} + \lambda \cdot \mathcal{L}_{cl}. \quad (9)$$

Where  $\lambda$  serves as a balancing hyperparameter.

## 5 Experiments

In this section, we conduct extensive experiments on real-world datasets to evaluate the performance of GenNIE. We detail the experimental setup in Section 5.1. Subsequently, from Section 5.2 to Section 5.6, we sequentially analyze the performance of GenNIE, generalization capabilities, ablation studies, the effectiveness of the graph perception mechanism, and the parameter sensitivity of the framework.

### 5.1 Experimental Settings

#### 5.1.1 Dataset Description

We conducted experiments on four publicly available knowledge graphs of distinct domains and process the graphs following previous work (Huang et al., 2021). Please refer to Appendix A.1 for descriptions of these datasets. The detailed statistics of nodes and edges are presented in Table 1.

#### 5.1.2 Baseline Methods

To evaluate GenNIE, we compare it against four categories of baselines. (1) **Traditional Network Analysis**: PageRank and Personalized

Method	FB15K		TMDB5K		MUSIC10K		DBLP	
	NDCG@100	SPEARMAN	NDCG@100	SPEARMAN	NDCG@100	SPEARMAN	NDCG@100	SPEARMAN
PR	0.8400 ± 0.0103	0.3497 ± 0.0188	0.8387 ± 0.0102	0.6247 ± 0.0130	0.7992 ± 0.0140	0.1772 ± 0.0189	0.3310 ± 0.0000	-0.1040 ± 0.0000
PPR	0.8411 ± 0.0112	0.3500 ± 0.0192	0.8495 ± 0.0083	0.6856 ± 0.0098	0.7981 ± 0.0109	0.1890 ± 0.0228	0.3330 ± 0.0000	0.0510 ± 0.0000
LR	0.9407 ± 0.0076	0.7837 ± 0.0044	0.7507 ± 0.0337	0.3569 ± 0.0321	0.8702 ± 0.0171	0.3566 ± 0.0199	0.5380 ± 0.0000	0.3120 ± 0.0000
MLP	0.9246 ± 0.0065	0.7075 ± 0.0109	0.7192 ± 0.0270	0.2751 ± 0.0100	0.7229 ± 0.0244	0.0349 ± 0.0523	-	-
GCN	0.8778 ± 0.0147	0.4660 ± 0.0288	0.8620 ± 0.0233	0.6588 ± 0.0471	0.8854 ± 0.0199	0.4440 ± 0.0326	-	-
RGCN	0.8547 ± 0.0189	0.2859 ± 0.0264	0.8656 ± 0.0171	0.7282 ± 0.0240	0.8974 ± 0.0087	0.5377 ± 0.0097	-	-
GraphSAGE	0.9491 ± 0.0081	0.8076 ± 0.0088	0.8297 ± 0.0322	0.5467 ± 0.0318	0.9076 ± 0.0111	0.4591 ± 0.0195	-	-
GENI	0.9174 ± 0.0139	0.7710 ± 0.0094	0.8522 ± 0.0132	0.7345 ± 0.0219	0.8872 ± 0.0065	0.4928 ± 0.0262	0.6020 ± 0.0000	0.4120 ± 0.0000
RGTN	0.9587 ± 0.0034	0.8370 ± 0.0101	0.8996 ± 0.0191	0.7480 ± 0.0164	0.9126 ± 0.0114	0.4590 ± 0.0237	0.6430 ± 0.0000	0.4380 ± 0.0000
HIVEN	-	-	0.9102 ± 0.0000	0.7931 ± 0.0000	0.8807 ± 0.0000	0.5523 ± 0.0000	0.7475 ± 0.0000	0.5323 ± 0.0000
SKES	-	-	<b>0.9420 ± 0.0000</b>	<b>0.8230 ± 0.0000</b>	0.8940 ± 0.0000	0.6020 ± 0.0000	0.6740 ± 0.0000	0.4830 ± 0.0000
LICAP	0.9409 ± 0.0086	0.7839 ± 0.0130	0.8851 ± 0.0195	0.7306 ± 0.0178	0.9130 ± 0.0133	0.4717 ± 0.0331	-	-
MVPL	0.9586 ± 0.0072	0.8260 ± 0.0020	0.9157 ± 0.0041	0.8018 ± 0.0046	-	-	-	-
LENIE	0.9629 ± 0.0037	0.8427 ± 0.0036	0.9133 ± 0.0118	0.7851 ± 0.0167	0.9395 ± 0.0032	0.6553 ± 0.0164	-	-
InstructGLM	0.8554 ± 0.0110	0.2150 ± 0.0045	0.8715 ± 0.0085	0.5850 ± 0.0072	0.8340 ± 0.0125	0.1880 ± 0.0090	0.4550 ± 0.0080	0.0520 ± 0.0055
GraphWiz	0.5215 ± 0.0055	-0.0296 ± 0.0030	0.6120 ± 0.0068	-0.0671 ± 0.0040	0.6250 ± 0.0095	-0.0450 ± 0.0035	0.3126 ± 0.0075	-0.0150 ± 0.0020
GenNIE(1.1B)	<b>0.9755 ± 0.0025</b>	0.9081 ± 0.0020	0.9098 ± 0.0030	0.7765 ± 0.0028	<u>0.9592 ± 0.0052</u>	<u>0.6881 ± 0.0045</u>	<b>0.9696 ± 0.0022</b>	0.5695 ± 0.0018
GenNIE(3B)	<u>0.9750 ± 0.0035</u>	<b>0.9204 ± 0.0015</b>	<u>0.9188 ± 0.0025</u>	0.8092 ± 0.0015	<b>0.9635 ± 0.0040</b>	<b>0.7364 ± 0.0038</b>	0.9680 ± 0.0020	<b>0.5881 ± 0.0016</b>

Table 2: Performance comparison on node importance estimation. The best results are highlighted in **bold**, and the second-best results are underlined.

PageRank (PPR). (2) **Machine Learning:** Logistic Regression (LR) and Multi-Layer Perceptron (MLP). (3) **GNN-based Methods:** We include general encoders (GCN (Kipf and Welling, 2017), RGCN (Schlichtkrull et al., 2018), GraphSAGE (Hamilton et al., 2017)) and specialized frameworks (GENI (Park et al., 2019), RGTN (Huang et al., 2021), LICAP (Zhang et al., 2025), HIVEN (Huang et al., 2022), MVPL (Ma et al., 2026), LENIE (Lin et al., 2025), SKES (Chen et al., 2024b)). (4) **LLM-based Graph Frameworks:** We select representatives from two dominant paradigms and tailor them to the specific requirements of the NIE task. **InstructGLM** (Ye et al., 2024) represents the *structure-serialization paradigm*, transforming topology into text for instruction tuning. **GraphWiz-DPO** (Chen et al., 2024a), a graph foundation model, is enhanced with CoT and Few-Shot strategies (Chen et al., 2024c) to represent the *prompt-guided reasoning paradigm*, leveraging its intrinsic reasoning capabilities to derive scores. The inclusion of the LLM-based frameworks aims to verify whether methods designed for general graph mining tasks can be effectively adapted to the precise numerical estimation challenges inherent to NIE, and specific adaptation for them are detailed in Appendix A.2.

### 5.1.3 Evaluation Metrics

Following previous works, we employ two metrics to evaluate model performance on the NIE task. Normalized Discounted Cumulative Gain (**NDCG@K**) evaluates the ranking quality of the top-K nodes, while the Spearman correlation coefficient (**SPEARMAN**) measures the global ranking

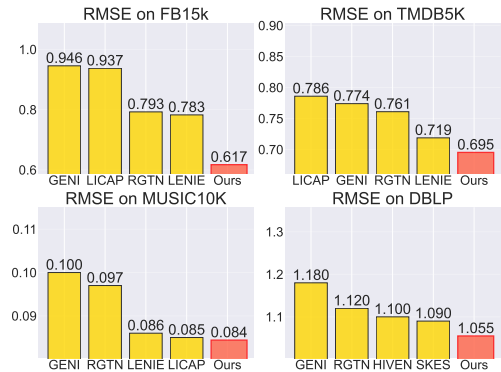


Figure 3: Performance of RMSE on each method.

consistency. Please refer to Appendix A.3 for the detailed definitions of these metrics.

### 5.1.4 Implementation Details

Please refer to Appendix A.4 for detailed implementation parameters and training settings.

## 5.2 Performance Analysis

We evaluate GenNIE instantiated with TinyLlama-1.1B (Zhang et al., 2024a) (sharing the identical architecture with Llama) and Llama-3.2-3B (Grattafiori et al., 2024) against various baselines. The results are presented in Table 2.

Overall, GenNIE achieves superior performance across most datasets, demonstrating its exceptional capability in the NIE task. Notably, GenNIE consistently maintains an NDCG@100 score above 0.90 across all datasets, indicating its stability and effectiveness in mining important entities within knowledge graphs. Furthermore, GNN-based methods generally outperform traditional and ML-based approaches, validating the GNNs’ capability in ag-

gregating topological features. In contrast, the modified LLM-based frameworks exhibit suboptimal performance on the NIE task. Despite being enhanced with prompt optimization, GraphWiz is unable to establish a reliable standard for node importance ranking. Similarly, InstructGLM, relying on graph-to-text serialization, struggles to effectively capture complex topological information. Regarding model scale, the 3B version of GenNIE generally outperforms the 1.1B variant, and the detailed impact of parameter size will be discussed in Section 5.6.

We also adopt the Root Mean Square Error (RMSE) to assess the deviation between the predicted and ground-truth importance scores. To ensure consistency with established benchmarks, we compute RMSE directly on the log-transformed importance values rather than the normalized  $[0, 100]$  interval. We select a subset of representative baselines to compare with GenNIE-Llama3B, and the results are illustrated in Figure 3. These results indicate that GenNIE achieves superior predictive accuracy, demonstrating that our generative approach driven by semantic reasoning can more effectively capture the underlying scoring patterns.

### 5.3 Generalization Analysis

Method	FB15K-MUSIC10K		FB15K-TMDB5K	
	NDCG@100	SPEARMAN	NDCG@100	SPEARMAN
RGTN	0.7096	-0.0284	0.4135	-0.1266
GENI	0.7297	0.0287	0.6545	0.0113
LENIE	0.7336	0.0168	0.5607	0.0934
<b>GenNIE (Ours)</b>	<b>0.9171</b>	<b>0.5907</b>	<b>0.8172</b>	<b>0.5519</b>

Table 3: Zero-shot transfer evaluation results.

To verify the generalization capability of GenNIE, we conducted zero-shot transfer experiments. Specifically, we trained the model on FB15K dataset and evaluated its performance directly on the other test datasets without additional training. For unseen relation types in the target datasets, their embeddings are randomly initialized during transfer. We compared GenNIE-Llama3B with representative GNN-based models.

As shown in Table 3, although the performance of GenNIE shows a slight decline compared to the in-domain supervised setting, it still maintains a remarkable level. In contrast, the performance of GNN-based methods is unsatisfactory. The results indicate that they fail to produce a meaningful ranking order when the domain changes. This experiment proves that unlike traditional GNNs

that overfit to dataset-specific topological distributions, GenNIE utilizes LLMs for semantic reasoning, demonstrating strong generalization capability. A more in-depth analysis of the generalization performance is provided in Appendix A.5.

### 5.4 Ablation Study

Variant	FB15K		MUSIC10K	
	NDCG@100	SPEARMAN	NDCG@100	SPEARMAN
w/o Contrastive Learning	0.9518	0.8313	0.9434	0.6551
w/o Soft Prompt	0.9634	0.8077	0.9145	0.5271
w/o Cross-Attention	0.9385	0.7920	0.9110	0.5050
w/o Textual Attributes	0.8780	0.6826	0.7553	0.1012
w/ MLP Projector	0.9512	0.8377	0.9490	0.5925
w/o Topological Injection	0.7162	0.1104	0.6700	0.0913
<b>GenNIE (Full)</b>	<b>0.9755</b>	<b>0.9081</b>	<b>0.9592</b>	<b>0.6881</b>

Table 4: Ablation study on FB15K and MUSIC10K. "w/o" denotes the removal of a specific component.

To investigate the contribution of each component in GenNIE, we conducted ablation studies on TinyLlama-1.1B with FB15K and MUSIC10K. Experiment results are presented in Table 4. Specifically, the *w/o Topological Injection* variant completely removes the GNN encoder and the MoE projector, relying solely on text prompts. The *w/o Textual Attributes* variant provides only structural embeddings to the LLM without any text descriptions. For the *w/o Cross-Attention* variant, the MoE projector is retained for dimensional alignment while the structure-guided cross-attention mechanism is removed, directly concatenating the projected topological embeddings before the text prompt. The *w/ MLP Projector* variant replaces the MoE architecture with a standard MLP of equivalent capacity for modality alignment.

The significant performance drop upon the *w/o Topological Injection* variant and the *w/o Textual Attributes* variant confirms that the importance of the entities in KGs is jointly determined by the synergy between topology and semantic information. Furthermore, the exclusion of the cross-attention mechanism validates the effectiveness of deeply fusing topological signals with textual features. Additionally, the superiority of the MoE projector over the MLP baseline demonstrates that the multi-expert architecture is more effective at capturing highly diverse topological patterns. Finally, the ablation of the perception mechanism highlights its necessity, which will be discussed in detail in Section 5.5.

### 5.5 Effectiveness of Graph Perception

**Global Graph Perception.** To demonstrate the effectiveness of learnable soft prompts, we calcu-

Group	NDCG@100			SPEARMAN		
	w/o Prompt	GenNIE (Full)	Improv.	w/o Prompt	GenNIE (Full)	Improv.
High	0.9118	0.9454	+3.7%	0.6729	0.7381	+9.7%
Medium	0.8507	0.9081	+6.7%	0.6261	0.7503	+19.8%
Low	0.7919	0.8868	+12.0%	0.5073	0.7131	+40.6%

Table 5: Analysis of the effectiveness of the learnable soft prompt across different node centrality groups.

lated the betweenness centrality of all nodes in the TMDB5K test set. Based on these centrality scores, we categorized the nodes into three groups: High (top 20%), Medium (middle 60%), and Low (bottom 20%). We then randomly sampled 150 nodes from each group to compare the performance of the *w/o Soft Prompt* variant against the full GenNIE(1.1B) model. The results are presented in Table 5. We observe that the performance improvement is most significant in the low betweenness group (e.g., +40.6% in Spearman). This result indicates that the soft prompt effectively compensates for the lack of structural information in less central nodes, demonstrating its capability to capture and inject global topological context into the model.

**Structural Graph Perception.** To investigate the enhancement of contrastive learning on structural graph perception, we sorted the nodes in the TMDB5K test set by their ground-truth importance scores and divided them into three equal groups. From each group, we randomly selected 150 nodes. Subsequently, we performed K-means clustering on the neighbor-aggregated representations  $\mathbf{h}_v^{struct}$  of these selected nodes and calculated the Normalized Mutual Information (NMI) and Adjusted Rand Index (ARI) to evaluate the discriminative capability of structural encodings across different importance levels. We compared the *w/o Contrastive Learning* variant against the full GenNIE (1.1B) model. Furthermore, to verify the contribution of contrastive learning to model generalization, we conducted zero-shot transfer experiments. The results are presented in Table 6.

Model	Clustering (TMDB5K)		FB15K-TMDB5K	
	NMI	ARI	NDCG@100	SPEARMAN
w/o Contrastive Learning	0.312	0.328	0.635	0.198
GenNIE (Full)	<b>0.622</b>	<b>0.642</b>	<b>0.795</b>	<b>0.547</b>

Table 6: Analysis of Contrastive Learning (CL).

The results demonstrate that contrastive learning significantly enhances the graph encoder’s topological discriminability. While instruction tuning improves scoring, it risks diluting structural perception. GenNIE mitigates this trade-off by preserving

discriminative patterns through contrastive learning, thereby ensuring robust generalization across diverse distributions.

## 5.6 Parameter Sensitivity Analysis

Model	FB15K	TMDB	Music	DBLP
Qwen2.5-0.5B	0.8271	0.7105	0.4992	0.5113
Qwen2.5-3B	0.9109	0.7700	0.6963	0.5960
TinyLlama-1.1B	0.9081	0.7765	0.6881	0.5695
Llama-3.2-3B	0.9204	<u>0.8092</u>	<u>0.7364</u>	0.5881
Llama-3-8B	<u>0.9277</u>	<b>0.8369</b>	<b>0.7985</b>	<u>0.6102</u>
Gemma-2-2B	0.9274	0.7907	0.7069	0.5965
Gemma-2-9B	<b>0.9319</b>	0.8038	0.7358	<b>0.6135</b>

Table 7: SPEARMAN of GenNIE with different LLMs.

To investigate the impact of LLMs’ parameter scales and architectures on GenNIE, we conducted experiments across a diverse set of representative open-source models, including **Qwen2.5** (Team et al., 2024b), **Llama** series (Grattafiori et al., 2024), and **Gemma-2** (Team et al., 2024a). By analyzing the Spearman coefficients in Table 7, we observe that while increasing the parameter scales generally contributes to performance gains, the lightweight models still exhibit remarkable effectiveness. Furthermore, the Llama series exhibits relatively better performance. Overall, apart from the 0.5B model, GenNIE remains effective across various LLMs, proving its architectural robustness.

## 6 Case Study

To demonstrate GenNIE’s potential in CSS, particularly for scenarios with scarce resources, we conducted a case study to assess the importance of historical characters. We constructed the knowledge graph by structuring raw historical corpora, which allows for capturing comprehensive character information. We selected the official biographical history *Records of the Three Kingdoms* as our corpus, which provides authoritative records of the Eastern Han period in China.

Tilly (2002) demonstrates that the influence of a historical character is closely correlated with the extent of their participation in collective events. Guided by this, we designed an ontology for Document-level Event Extraction (DEE) (Liu et al., 2023) and utilized LLMs to extract structured event records. Historical characters in extracted records are serve as nodes, and edges are established between characters who co-participate within the same event record. Subsequently, we performed LLMs to perform entity disambiguation and en-

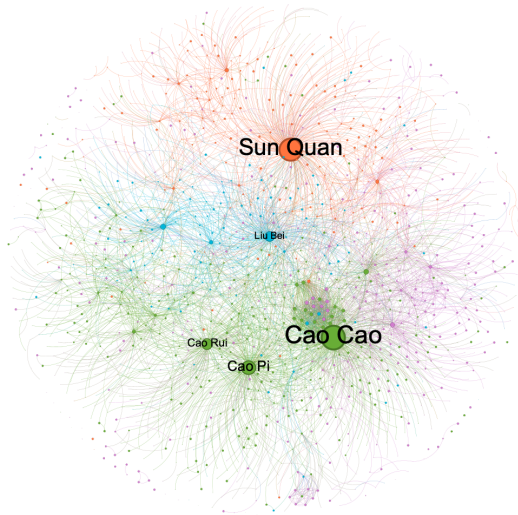


Figure 4: Visualization of the character KG. Distinct colors represent different political factions.

riched character attributes using the CBDB (Harvard University et al., 2019). Finally, we constructed the KG with statistics as table 8.

Dataset	Nodes	Edges	Edge Types
Three Kingdom character KG	1,678	13,921	17

Table 8: Statistics of the constructed knowledge graph.

Given the shared focus on character importance, we directly applied the Music-10K-trained GenNIE-Llama-8B to our KG. Figure 4 visualizes the constructed knowledge graph, where the node size corresponds to the predicted importance score.

As shown in Figure 4, we can clearly observe the importance of the characters, where key faction leaders are distinctly highlighted. This demonstrates the effectiveness of GenNIE in applications within a specific cultural context. The elaborate case study is provided in **Appendix A.6**.

## 7 Conclusion

In this paper, we proposed GenNIE, a generative NIE framework based on structure-guided semantic reasoning. Through topology injection and graph perception mechanism, it achieves a deep integration of topology and semantics. Extensive experiments demonstrate GenNIE’s superior performance and its generalization capability, proving that it can serve as a powerful tool for sociological analysis.

## Limitations

Current KG-based approaches for Node Importance Estimation (NIE) are predominantly predicated on static knowledge graph. Consistent with this established research paradigm, our work models entity interactions as a unified structural whole. However, in real-world scenarios, the importance of an entity is inherently time-dependent and dynamic rather than constant. In future work, we plan to develop a specialized benchmark for Temporal Node Importance Estimation in historical contexts, aiming to provide more robust analytical tools for the Computational Social Science (CSS) community.

## Ethical Considerations

All artifacts utilized in this study, including the four benchmark datasets and the LLMs, are open-source and employed strictly for academic research. The Records of the Three Kingdoms text resides in the public domain. Our derived data, encompassing the knowledge graph and node importance rankings, contains no copyrighted material or personally identifying information.

## Acknowledgments

This work is supported by the National Natural Science Foundations of China under Grant (62372060).

## References

- Réka Albert, Hawoong Jeong, and Albert-László Barabási. 2000. [Error and attack tolerance of complex networks](#). *nature*, 406(6794):378–382.
- Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo. 2011. [Network medicine: a network-based approach to human disease](#). *Nature reviews genetics*, 12(1):56–68.
- Stefano Battiston, Michelangelo Puliga, Rahul Kaushik, Paolo Tasca, and Guido Caldarelli. 2012. [Debtank: Too central to fail? financial networks, the fed and systemic risk](#). *Scientific reports*, 2(1):541.
- Thierry Bertin-Mahieux, Daniel PW Ellis, Brian Whitman, and Paul Lamere. 2011. [The million song dataset](#).
- Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. [Freebase: a collaboratively created graph database for structuring human knowledge](#). In *Proceedings of the 2008 ACM SIGMOD International Conference on Management*

- of Data, SIGMOD '08, page 1247–1250, New York, NY, USA. Association for Computing Machinery.
- Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. [Translating embeddings for modeling multi-relational data](#). In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.
- Lu Cao, Cangfeng Ding, Lerong Ma, Zhaoyao Yan, Hao You, and Anqi Hong. 2025. [Advances in node importance ranking based on graph neural networks](#). *Journal of Frontiers of Computer Science & Technology*, 19(4).
- Ziwei Chai, Tianjie Zhang, Liang Wu, Kaiqiang Han, Xiaohai Hu, Xuanwen Huang, and Yang Yang. 2025. [Graphllm: Boosting graph reasoning ability of large language model](#). *IEEE Transactions on Big Data*, pages 1–9.
- Nuo Chen, Yuhan Li, Jianheng Tang, and Jia Li. 2024a. [Graphwiz: An instruction-following language model for graph computational problems](#). In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, page 353–364, New York, NY, USA. Association for Computing Machinery.
- Yankai Chen, Yixiang Fang, Qiongyan Wang, Xin Cao, and Irwin King. 2024b. [Deep structural knowledge exploitation and synergy for estimating node importance value on heterogeneous information networks](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(8):8302–8310.
- Yankai Chen, Taotao Wang, Yixiang Fang, and Yunyu Xiao. 2025. [Semi-supervised node importance estimation with informative distribution modeling for uncertainty regularization](#). In *Proceedings of the ACM on Web Conference 2025*, WWW '25, page 3108–3118, New York, NY, USA. Association for Computing Machinery.
- Zhikai Chen, Haitao Mao, Hang Li, Wei Jin, Hongzhi Wen, Xiaochi Wei, Shuaiqiang Wang, Dawei Yin, Wenqi Fan, Hui Liu, and Jiliang Tang. 2024c. [Exploring the potential of large language models \(llms\) in learning on graphs](#). *SIGKDD Explor. Newsl.*, 25(2):42–61.
- Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, and 542 others. 2024. [The llama 3 herd of models](#). *Preprint*, arXiv:2407.21783.
- Aditya Grover and Jure Leskovec. 2016. [node2vec: Scalable feature learning for networks](#). In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 855–864, New York, NY, USA. Association for Computing Machinery.
- Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. [Inductive representation learning on large graphs](#). In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Harvard University, Academia Sinica, and Peking University. 2019. [China biographical database \(CBDB\)](#).
- Chenji Huang, Yixiang Fang, Xuemin Lin, Xin Cao, Wenjie Zhang, and Maria Orłowska. 2022. [Estimating node importance values in heterogeneous information networks](#). In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pages 846–858.
- Han Huang, Leilei Sun, Bowen Du, Chuanren Liu, Weifeng Lv, and Hui Xiong. 2021. [Representation learning on knowledge graphs for node importance estimation](#). In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, page 646–655, New York, NY, USA. Association for Computing Machinery.
- Bowen Jin, Gang Liu, Chi Han, Meng Jiang, Heng Ji, and Jiawei Han. 2024. [Large language models on graphs: A comprehensive survey](#). *IEEE Transactions on Knowledge and Data Engineering*, 36(12):8622–8642.
- Thomas N. Kipf and Max Welling. 2017. [Semi-supervised classification with graph convolutional networks](#). In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings*. OpenReview.net.
- Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse. 2010. [Identification of influential spreaders in complex networks](#). *Nature Physics*, 6:888–893.
- Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. 2015. [Dbpedia – a large-scale, multilingual knowledge base extracted from wikipedia](#). *Semantic Web*, 6(2):167–195.
- Xinyu Lin, Tianyu Zhang, Chengbin Hou, Jinbao Wang, Jianye Xue, and Hairong Lv. 2025. [Node importance estimation leveraging llms for semantic augmentation in knowledge graphs](#). *Knowledge-Based Systems*, 330:114521.
- Qi Liu, Zhen Luan, Kunlong Wang, Ye Zou, Bing Liu, and Yang Zhang. 2023. [Document-level event extraction - a survey of methods and applications](#). *Journal of Physics: Conference Series*, 2504(1):012008.
- Siqi Ma, Yang Fang, Weidong Xiao, and Xiang Zhao. 2026. [Node importance estimation via multi-view graph prompting](#). *IEEE Transactions on Network Science and Engineering*, 13:2022–2037.

- Yuning Mao, Lambert Mathias, Rui Hou, Amjad Almahairi, Hao Ma, Jiawei Han, Scott Yih, and Madian Khabsa. 2022. [UniPELT: A unified framework for parameter-efficient language model tuning](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6253–6264, Dublin, Ireland. Association for Computational Linguistics.
- Shervin Minaee, Tomas Mikolov, Narjes Nikzad, Meysam Chenaghlu, Richard Socher, Xavier Amatriain, and Jianfeng Gao. 2025. [Large language models: A survey](#). *Preprint*, arXiv:2402.06196.
- Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford infolab.
- Namyong Park, Andrey Kan, Xin Luna Dong, Tong Zhao, and Christos Faloutsos. 2019. [Estimating node importance in knowledge graphs using graph neural networks](#). In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19*, page 596–606, New York, NY, USA. Association for Computing Machinery.
- Namyong Park, Andrey Kan, Xin Luna Dong, Tong Zhao, and Christos Faloutsos. 2020. [Multiimport: Inferring node importance in a knowledge graph from multiple input signals](#). In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '20*, page 503–512, New York, NY, USA. Association for Computing Machinery.
- Luming Qiu. 2015. *Political Power and Family Networks during the Wei-Jin Transition*. Shanghai Classics Publishing House, Shanghai.
- Xiaolong Ren and Linyuan Lü. 2014. Review of ranking nodes in complex networks. *Chinese Science Bulletin*, 59(13):1175–1197. (in Chinese).
- Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. [Modeling relational data with graph convolutional networks](#). In *The Semantic Web*, pages 593–607, Cham. Springer International Publishing.
- Jian Shu, Yiling Zou, Hui Cui, and Linlan Liu. 2025. [Node importance evaluation in heterogeneous network based on attention mechanism and graph contrastive learning](#). *Neurocomputing*, 626:129555.
- Yanchao Tan, Hang Lv, Pengxiang Zhan, Shiping Wang, and Carl Yang. 2026. [Graph-oriented instruction tuning of large language models for generic graph mining](#). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(1):155–169.
- Jiabin Tang, Yuhao Yang, Wei Wei, Lei Shi, Lixin Su, Suqi Cheng, Dawei Yin, and Chao Huang. 2024. [Graphgpt: Graph instruction tuning for large language models](#). In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '24*, page 491–500, New York, NY, USA. Association for Computing Machinery.
- Gemma Team, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, Johan Ferret, Peter Liu, Pouya Tafti, Abe Friesen, Michelle Casbon, Sabela Ramos, Ravin Kumar, Charline Le Lan, Sammy Jerome, and 179 others. 2024a. [Gemma 2: Improving open language models at a practical size](#). *Preprint*, arXiv:2408.00118.
- Qwen Team and 1 others. 2024b. Qwen2 technical report. *arXiv preprint arXiv:2407.10671*, 2(3).
- Yijun Tian, Huan Song, Zichen Wang, Haozhu Wang, Ziqing Hu, Fang Wang, Nitesh V. Chawla, and Panpan Xu. 2024. [Graph neural prompting with large language models](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(17):19080–19088.
- Charles Tilly. 2002. *Stories, identities, and political change*. Rowman & Littlefield.
- Feng Wang, Jinhua She, Yasuhiro Ohyama, and Min Wu. 2019. [Deep-learning-based identification of influential spreaders in online social networks](#). In *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, volume 1, pages 6854–6858.
- Shijie Wang, Jiani Huang, Zhikai Chen, Yu Song, Wenzhuo Tang, Haitao Mao, Wenqi Fan, Hui Liu, Xiaorui Liu, Dawei Yin, and Qing Li. 2025. [Graph machine learning in the era of large language models \(llms\)](#). *ACM Trans. Intell. Syst. Technol.*, 16(5).
- Ruosong Ye, Caiqi Zhang, Runhui Wang, Shuyuan Xu, and Yongfeng Zhang. 2024. [Language is all a graph needs](#). In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 1955–1973, St. Julian's, Malta. Association for Computational Linguistics.
- Peiyuan Zhang, Guangtao Zeng, Tianduo Wang, and Wei Lu. 2024a. [Tinyllama: An open-source small language model](#). *Preprint*, arXiv:2401.02385.
- Renrui Zhang, Jiaming Han, Chris Liu, Peng Gao, Aojun Zhou, Xiangfei Hu, Shilin Yan, Pan Lu, Hongsheng Li, and Yu Qiao. 2024b. [Llama-adapter: Efficient fine-tuning of language models with zero-init attention](#). *Preprint*, arXiv:2303.16199.
- Tianyu Zhang, Chengbin Hou, Rui Jiang, Xuegong Zhang, Chenghu Zhou, Ke Tang, and Hairong Lv. 2025. [Label informed contrastive pretraining for node importance estimation on knowledge graphs](#). *IEEE Transactions on Neural Networks and Learning Systems*, 36(3):4462–4476.
- Jiaqi Zhao, Yuting Wei, and Bin Wu. 2022. [Analysis of the social network and the evolution of the influence of ancient chinese poets](#). *Social Science Computer Review*, 40(4):1014–1034.

## A Appendices

### A.1 Dataset Description

We evaluate our framework on four diverse real-world knowledge graph datasets: **FB15K** (Bordes et al., 2013), **TMDB5K** (Park et al., 2019), **Music-10K** (Park et al., 2019), and **DBLP** (Chen et al., 2024b). Detailed statistics are summarized in Table 1.

**FB15K.** FB15K (Bordes et al., 2013), a widely adopted subset of Freebase (Bollacker et al., 2008), comprises rich general knowledge triplets. In our experimental setting, we leverage entity descriptions from Wikidata as semantic information and define the ground-truth node importance based on the 30-day page view count of the corresponding Wikipedia pages.

**TMDB5K.** Derived from the TMDB 5000 dataset, TMDB5K encompasses entities including movies, actors, crew members, and production companies. For this dataset, we specifically evaluate the importance of movie nodes, employing the *popularity* score provided in the original source as the ground-truth label.

**Music-10K.** Sampled from the Million Song Dataset (Bertin-Mahieux et al., 2011), Music-10K models the relationships between artists and songs. We focus on assessing the importance of artist nodes, utilizing the *artist hotness* metric as the ground truth.

**DBLP.** DBLP is an academic citation graph representing papers and their citation relationships. In this dataset, we target paper nodes and adopt the citation count of each paper as the quantitative measure of its importance.

### A.2 Model Adaptation

To align the capabilities of Large Language Models with the specific requirements of Node Importance Estimation (NIE), we performed targeted adaptations on two representative frameworks: **InstructGLM** (Ye et al., 2024) and **GraphWiz** (Chen et al., 2024a). While these models were originally designed for general graph reasoning or generation tasks, they require specific modifications to bridge the gap between semantic understanding and precise importance quantification. The specific adaptation schemes are described as follows.

**InstructGLM.** InstructGLM (Ye et al., 2024) introduces a rigorous instruction-tuning framework tailored for graph machine learning. By linearizing local subgraphs into natural language descriptions, it enables LLMs to comprehend complex structural information. Originally, this framework was primarily deployed for classical graph tasks such as node classification and link prediction. We selected InstructGLM as a representative baseline because it exemplifies the paradigm of transforming graph structures into textual sequences, thereby leveraging instruction fine-tuning (SFT) to enhance performance on graph-specific tasks.

To adapt InstructGLM for the NIE task, we maintain its original model architecture but reconfigure the input-output format for Supervised Fine-Tuning. Specifically, we incorporate the textual attributes of the target node into the input and serialize its surrounding local subgraph structure into a natural language prompt. Crucially, we redefine the generation target: instead of predicting a class label, the model is trained to generate the normalized importance score of the central node. Consistent with the original training protocol, we fine-tuned the model using a mixture of subgraph scopes ranging from 1-hop to 3-hop neighbors. During the inference phase, we evaluated the model using prompts constructed from 1-hop, 2-hop, and 3-hop neighborhoods separately, and reported the optimal performance achieved across these settings.

**GraphWiz.** GraphWiz (Chen et al., 2024a) is a family of open-source language models specifically instruction-tuned for explicit graph reasoning tasks. Unlike general-purpose LLMs, GraphWiz is optimized to understand structural dependencies and solve complex graph problems. To further elevate its capability in quantifying node importance, we adopt the *LLM-as-Predictor* paradigm highlighted by Chen et al. (2024c), which demonstrates that combining LLMs with Chain-of-Thought (CoT) and Few-Shot prompting significantly enhances performance on graph tasks such as node classification. We selected GraphWiz as a baseline to represent the paradigm of leveraging in-context learning (ICL) on specialized graph LLMs, distinguishing it from the supervised fine-tuning (SFT) paradigm employed by InstructGLM.

Following (Chen et al., 2024c), we designed a structure-aware CoT prompt strategy. First, to construct a representative few-shot context, we employed a stratified sampling strategy: we selected

one representative node instance from each 20-point score interval (e.g., 0-20, 20-40, ..., 80-100) to ensure the model perceives the full spectrum of importance distribution. Second, for the target node, we linearized its neighborhood into textual descriptions, enabling the model to process topological context semantically.

An illustration of the specific prompt template used for the FB15K dataset is shown in Figure 7.

### A.3 Evaluation Metrics

To comprehensively assess the performance of our framework in terms of both ranking quality and numerical prediction precision, we employ three standard metrics: Normalized Discounted Cumulative Gain (NDCG), Spearman’s Rank Correlation Coefficient (Spearman), and Root Mean Square Error (RMSE).

**NDCG@K.** Normalized Discounted Cumulative Gain is widely used to evaluate the quality of ranking, particularly emphasizing the correctness of top-ranked items. Given a list of nodes ranked by their predicted importance scores, the Discounted Cumulative Gain at position  $K$  (DCG@K) is calculated as:

$$DCG@K = \sum_{i=1}^K \frac{rel_i}{\log_2(i+1)}, \quad (10)$$

where  $rel_i$  denotes the ground-truth importance score of the node located at the  $i$ -th position in the predicted ranking list. To normalize this metric across different lists, we divide DCG@K by the Ideal DCG (IDCG@K), which is obtained by ranking nodes strictly according to their ground-truth scores:

$$NDCG@K = \frac{DCG@K}{IDCG@K}. \quad (11)$$

We report NDCG@K to demonstrate the model’s ability to identify the most significant nodes.

**Spearman Correlation.** To evaluate the monotonic relationship between the predicted ranking and the ground-truth ranking over the entire test set, we utilize the Spearman Rank Correlation Coefficient ( $\rho$ ). Let  $rg_i$  and  $rp_i$  denote the rank positions of the  $i$ -th node in the ground-truth list and the predicted list, respectively. The coefficient is computed as:

$$\text{Spearman} = \frac{\sum_{i=1}^N (rg_i - \bar{r}g)(rp_i - \bar{r}p)}{\sqrt{\sum_{i=1}^N (rg_i - \bar{r}g)^2} \sqrt{\sum_{i=1}^N (rp_i - \bar{r}p)^2}}, \quad (12)$$

where  $N$  is the total number of test nodes, and  $\bar{r}g, \bar{r}p$  are the mean ranks. A higher Spearman correlation indicates a stronger alignment between the predicted global order and the actual importance order.

**RMSE.** Root Mean Square Error is employed to measure the precision of the regression values. It quantifies the average magnitude of the error between the predicted importance scores and the ground-truth scores:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (13)$$

where  $y_i$  and  $\hat{y}_i$  represent the ground-truth importance score and the predicted score for the  $i$ -th node, respectively. Lower RMSE values indicate higher accuracy in numerical estimation.

### A.4 Implementation Details

All experiments were implemented using Python 3.10.19 and the Deep Graph Library (DGL) version 2.4.0 (CUDA 12.1). The training process was conducted on a single NVIDIA L20 Tensor Core GPU with 48GB of VRAM. We employed the AdamW optimizer with a weight decay of  $1e-4$  and a consistent batch size of 32 across all model variants. To ensure parameter efficiency, we utilized Low-Rank Adaptation (LoRA) with a rank of  $r = 16$ , a scaling factor  $\alpha = 32$ , and a dropout rate of 0.05. The learning rates were empirically tuned based on model scales: set to  $1e-3$  for GenNIE-Llama-1.1B,  $5e-4$  for GenNIE-Llama-3B, and  $3e-4$  for GenNIE-Llama-8B.

Regarding the hyperparameters, the balancing coefficient  $\lambda$  in the joint loss function was set to 0.5, which yielded the optimal performance on the FB15K validation set. For the contrastive learning module, the temperature coefficient  $\tau$  was fixed at 0.1. To construct informative sample pairs, we set the threshold for positive sample definition to  $\delta_{pos} = 2.0$  and the threshold for negative sample identification to  $\delta_{neg} = 10.0$ . Additionally, to facilitate faster convergence, the soft prompt embeddings were initialized using the semantic representation of the instruction: *Estimate the importance score for this node based on the graph structure and text information.* To prevent overfitting, an early stopping mechanism was applied with a patience of 20 epochs based on the validation set performance.

## A.5 Generalization Analysis

To further investigate the factors influencing the generalization capability of GenNIE, we analyze whether the performance drop during transfer is primarily caused by unseen relation types. Specifically, we pre-train the model on FB15K and evaluate it on Music-10K under three distinct settings. The *Zero-shot* setting is identical to the generalization analysis described in Section 5.3. The *Relation-Tuned* setting freezes all parameters except for the learnable relation embeddings after pre-training on FB15K, and then fine-tunes them on Music-10K. The *Fixed-Random* setting trains the model on Music-10K but keeps the randomly initialized relation embeddings frozen, isolating the effect of relation representations.

Model Setting	NDCG@100	SPEARMAN
GenNIE (Zero-shot)	0.9171	0.5907
GenNIE (Relation-Tuned)	0.9317	0.6385
GenNIE (Fixed-Random)	0.9352	0.6721
GenNIE (Full Supervised)	0.9635	0.7364

Table 9: Performance comparison of different transfer settings on Music-10K.

As shown in Table 9, fine-tuning only the relation embeddings brings limited improvement over the zero-shot baseline and still underperforms the *Fixed-Random* setting. This indicates that the performance gap in transfer learning is not solely dominated by relation semantics, but rather stems fundamentally from the adaptation of topological encoders and global perception modules to new graph structures.

## A.6 Case Study

To demonstrate the generalization capability and practical utility of our framework in processing unstructured raw corpora, we conducted a comprehensive case study centered on the *Records of the Three Kingdoms (Sanguozhi)*. Unlike standard benchmark datasets, this case study involves the complete pipeline from ontology construction and information extraction to importance estimation and historical analysis.

### A.6.1 Dataset Background and Ontology Design

The *Records of the Three Kingdoms (Sanguozhi)* serves as the foundational corpus for this case study. It chronicles the tripartite division of China between the states of *Wei*, *Shu*, and *Wu* following the

collapse of the *Eastern Han* Dynasty. This period is arguably one of the most celebrated and culturally significant eras in East Asian history, widely popularized by the 14th-century historical novel *Romance of the Three Kingdoms* and countless subsequent cultural adaptations.

We selected this specific historical corpus as our experimental testbed not only for its cultural prominence but primarily for the inherent complexity of its social and political fabric. The narrative presents a densely interwoven network of wavering alliances, military conflicts, and intricate kinship relations, offering an ideal scenario to evaluate the model’s ability to disentangle complex topological structures. Furthermore, the classical text is rich in implicit semantics, where critical power shifts are often embedded in subtle narrative details rather than explicit statements, necessitating deep semantic understanding beyond simple keyword matching.

To capture these nuances, we designed a specialized ontology for document-level event extraction (Figure 6). Distinct from generic schemas, our ontology explicitly defines complex historical relations (e.g., *Military Alliance*, *Political Appointment*) and event types, facilitating the extraction of structured knowledge from narrative texts.

### A.6.2 LLM-based Construction Pipeline

We implemented a robust pipeline to transform the raw historical text into a structured knowledge graph:

**Event Extraction via Few-shot Prompting.** Leveraging the semantic capability of Large Language Models, we employed a few-shot prompting strategy to extract event records. The prompt template, illustrated in Figure 7, guides the LLM to identify participants and their roles within specific historical contexts.

**Entity Disambiguation and Resolution.** A significant challenge in processing ancient Chinese texts is the multiplicity of names (e.g., courtesy names/Zi, pseudonyms, and official titles). For instance, "Cao Cao" is frequently referred to as "Cao Mengde" or "Prime Minister Cao." To address this ambiguity and ensure accurate entity alignment, we employed an LLM to perform disambiguation across all extracted historical figures. Specifically, we sequentially input each entity mention into the model, instructing it to normalize these diverse

appellations into their corresponding canonical primary names based on historical knowledge.

**Topology and Attribute Enrichment.** We constructed the graph topology by treating entities co-occurring in the same event record as connected edges. To further enhance node semantics, we performed entity linking with the China Biographical Database (CBDB). By aligning our extracted entities with CBDB records, we enriched the nodes with verified biographical attributes (e.g., birth/death years, ancestral home), transforming the graph from a simple interaction network into a metadata-rich historical prosopography.

### A.6.3 Visualization and Historical Analysis

To visually assess the efficacy of our model in capturing historical prominence and structural relationships, we applied GenNIE to estimate the importance scores of all 1678 historical characters in the constructed graph. The resulting network visualization, generated using Gephi, is presented in Figure 8. In this visualization, node size is directly proportional to the predicted importance score, and node color encodes the political faction: **Wei (Green)**, **Wu (Orange)**, **Shu (Blue)**, and **Qun (Purple)**, where *Qun* represents various independent warlords and factions outside the three primary kingdoms.

#### Community Structure and Factional Dynamics.

The visualization reveals a clear community structure consistent with the geopolitical landscape of the era. The four color-coded factions form distinct, densely connected clusters, with significant inter-factional edges illustrating the intense diplomatic and military interactions characterizing the period. Notably, the **Qun (Purple)** faction appears more dispersed around the periphery, reflecting the fragmented nature of lesser warlords who were gradually absorbed by the three major powers. A striking observation is the disproportionate size of the **Wei (Green)** community, which is significantly larger and denser than the others. This topological feature is historically grounded: Chen Shou, the author of the *Records of the Three Kingdoms*, was an official of the *Jin* Dynasty, which succeeded the *Wei* regime. To establish political legitimacy, the *Jin* court viewed *Wei* as the orthodox successor to the *Han* Dynasty (Qiu, 2015). Consequently, the historical records naturally contain a more extensive and detailed chronicling of *Wei* figures and events, leading to a richer representation in our

constructed graph.

**Historical Prominence of Key Figures.** The predicted importance scores accurately reflect historical consensus on key characters.

- **The Three Founding Emperors: Cao Cao (Wei, Green), Sun Quan (Wu, Orange), and Liu Bei (Shu, Blue)** emerge as the largest hubs within their respective communities. Cao Cao, the de facto founder of *Wei*, was a brilliant strategist and statesman who unified northern China. Sun Quan established a stable regime in the Jiangdong region and reigned for over half a century. Liu Bei, known for his benevolence and persistence, founded the *Shu* state in the southwest. Their paramount node sizes in the graph perfectly align with their foundational roles in history.
- **Dynastic Continuity:** The high centrality of successors such as **Cao Pi** (first emperor of *Wei*) and **Cao Rui** within the *Wei* cluster indicates the successful consolidation and transfer of power within the *Cao* clan, further underscoring the stability and dominance of the *Wei* regime.

This alignment between computational results and qualitative historical knowledge demonstrates GenNIE’s capability to extract and quantify latent historical significance from unstructured texts.

### A.6.4 Implications for Computational Social Science

This case study demonstrates the transformative potential of our proposed method in Computational Social Science (CSS). By automating the extraction of structured knowledge from unstructured classical texts and modeling them as quantifiable networks, our approach empowers researchers to:

- **Quantitatively Validate Historical Narratives:** Our method provides a data-driven framework to statistically verify qualitative historical consensus, such as the hegemonic stability provided by central leadership figures like Cao Cao, transforming subjective historical interpretation into rigorous empirical evidence.
- **Uncover Latent Social Structures:** Beyond prominent leaders, our topological analysis identifies "marginal figures"—individuals

with sparse biographical records who nevertheless serve as critical structural bridges or information brokers in specific historical events, revealing the hidden mechanisms of social mobilization.

- **Decode Complex Geopolitical Dynamics:** By mapping the dense interactions within and between factions, our framework exposes latent patterns of political alliance and conflict resolution that are difficult to discern through close reading alone, offering a scalable lens to study the evolution of power structures in pre-modern societies.

**Task:** Estimate the importance score (0-100) for nodes in a knowledge graph. The importance score reflects how central or significant a node is in the graph.

**Consider factors such as:**

- Number and diversity of connections (degree centrality)
- Role as a hub or bridge between different domains
- Information richness and description quality
- Domain significance and general knowledge importance
- Position in the graph structure (central vs. peripheral)

**IMPORTANT:** The importance score must be a number between 0 and 100, with exactly 2 decimal places (e.g., 85.23, 42.67, 3.14).

**Examples:**

Example 1:

Entity: Barack Obama

Description: Barack Hussein Obama II is an American politician who served as the 44th president of the United States from 2009 to 2017.

Reasoning: This is a highly prominent person (Barack Obama) with significant historical or cultural impact. Such entities typically have extensive connections in knowledge graphs, appearing in many relationships across different domains, which explains the high importance score.

Importance Score: 92.50

Example 2:

Entity: Machine Learning

Description: Machine learning is a method of data analysis that automates analytical model building.

Reasoning: This entity (Machine Learning) is important within its domain, with significant connections to related entities. It serves as a key node but may be more specialized than the most central entities.

Importance Score: 75.30

Example 3:

Entity: Python Programming

Description: Python is a high-level programming language known for its simplicity and readability.

Reasoning: The brief description suggests this entity has limited information, which often correlates with moderate importance. It likely has some connections but is not a central hub in the knowledge graph.

Importance Score: 58.70

Example 4:

Entity: Local Restaurant

Description: A small family-owned restaurant serving traditional cuisine.

Reasoning: The minimal description indicates limited information and connections. This entity likely represents a specialized or less prominent node with fewer relationships in the graph.

Importance Score: 35.20

Example 5:

Description: A minor entity with limited connections in the knowledge graph.

Reasoning: The lack of a specific name and minimal description indicates this is a peripheral entity with very few connections. It may be a minor concept or entity with limited relevance in the overall graph structure.

Importance Score: 12.80

**Target Node:**

Entity: Albert Einstein

Description: Albert Einstein was a German-born theoretical physicist who is widely held to be one of the greatest and most influential scientists of all time.

**Neighbor Context:**

This node is connected to 5 neighbor(s):

- Physics: The study of matter, motion, and behavior...
- Theory of Relativity: A fundamental theory in physics...
- Nobel Prize: An international award...
- Quantum Mechanics: A fundamental theory in physics...
- Princeton University: A private research university...

Note: The number and types of neighbors can indicate the node's connectivity and centrality.

**Response Requirements:** Please provide your response following the "Chain-of-Thought" process below:

Step 1: Topological Analysis Analyze the provided Neighbor Context. Does the node connect to diverse or high-value neighbors? Does it act as a hub or a bridge?

Step 2: Semantic Analysis Evaluate the Target Node's description. Is the entity fundamentally important in its domain? Is the description rich and informative?

Step 3: Synthesis & Scoring Combine the topological and semantic evidence. Compare this node to the examples provided to estimate its relative position.

Final Answer:

Reasoning: [Your step-by-step deduction based on the above 3 steps]

Importance Score: [The final score between 0.00 and 100.00, exactly 2 decimal places]

Figure 5: Illustration of the Structure-aware CoT Prompt Design.

```

Three Kingdoms Historical Event Extraction Ontology (
  Politics (
    description: "Focuses on power structure transitions such as regime changes, institutional reforms, coups, and revolutions, covering the establishment of ruling systems, power succession mechanisms, and sudden political turbulence."
    Regime_Change (
      description: "Describes events where ruling systems are replaced due to power struggles, social contradictions, or external interventions, often accompanied by war, coups, or revolutions. Examples include the establishment of the Qin Dynasty, the abdication of the Qing Emperor, etc."
      arguments: {
        Year: { required: true, to_be_disambiguated: false },
        majorPerson: { required: true, to_be_disambiguated: true },
        majorPerson2: { required: false, to_be_disambiguated: true },
        majorPerson3: { required: false, to_be_disambiguated: true },
        establishedRegime: { required: true, to_be_disambiguated: true },
        deadRegime: { required: false, to_be_disambiguated: true }
      }
      examples: [
        {
          Year: "1804",
          majorPerson: "Napoleon I",
          majorPerson2: "Louis XVI",
          majorPerson3: "",
          establishedRegime: "First French Republic",
          deadRegime: "First French Empire"
        }
      ]
    }
    Institutional_Reform (
      description: "Covers systematic adjustments to laws, bureaucracies, economies, or governance models, such as the imperial examination system replacing the recommendation system, or the Qing Dynasty issuing the Qianliu Order."
      arguments: {
        Year: { required: true, to_be_disambiguated: false },
        nameOfPolicy: { required: true, to_be_disambiguated: false },
        majorPerson: { required: true, to_be_disambiguated: true },
        majorPerson2: { required: false, to_be_disambiguated: true },
        majorPerson3: { required: false, to_be_disambiguated: true },
        country: { required: false, to_be_disambiguated: true },
        field: { required: false, to_be_disambiguated: false }
      }
      examples: [
        {
          Year: "356",
          nameOfPolicy: "Shang Yang's Reforms",
          majorPerson: "Shang Yang",
          country: "State of Qin",
          majorPerson2: "Duke Xiao of Qin",
          majorPerson3: "",
          field: "Economy"
        }
      ]
    }
    Coup_Revolution (
      description: "A coup involves seizing power in the short term through violent means (e.g., Xuanwu Gate Incident), while a revolution establishes a new system by systematically overturning the old order (e.g., French Revolution)."
      arguments: {
        Year: { required: true, to_be_disambiguated: false },
        type: { required: false, to_be_disambiguated: true },
        majorPerson: { required: true, to_be_disambiguated: true },
        majorPerson2: { required: false, to_be_disambiguated: true },
        majorPerson3: { required: false, to_be_disambiguated: true },
        country: { required: true, to_be_disambiguated: true }
      }
      examples: [
        {
          Year: "220 AD",
          type: "Coup",
          majorPerson: "Cao Pi",
          majorPerson2: "Emperor Xian of Han",
          majorPerson3: "",
          country: "Wei"
        }
      ]
    }
    Strategic_Confrontation (
      description: "Consists of wit and strategic confrontations, including intellectual games such as stratagems, instigating defection, espionage activities, etc."
      arguments: {
        Year: { required: true, to_be_disambiguated: false },
        stratagem: { required: true, to_be_disambiguated: true },
        targetPerson: { required: true, to_be_disambiguated: true },
        strategyType: { required: true, to_be_disambiguated: false },
        strategyContent: { required: false, to_be_disambiguated: false },
        result: { required: false, to_be_disambiguated: false }
      }
      examples: [
        {
          Year: "208 AD",
          stratagem: "Zhou Yu",
          targetPerson: "Cao Cao",
          strategyType: "Fire Attack",
          strategyContent: "Interlocking ships fire attack",
          result: "Successful"
        }
      ]
    }
    Mentor_Friend_Relation (
      description: "Important mentorship relations, friendship relations, or academic exchanges, reflecting knowledge inheritance and interpersonal networks."
      arguments: {
        Year: { required: true, to_be_disambiguated: false },
        personA: { required: true, to_be_disambiguated: true },
        personB: { required: true, to_be_disambiguated: true },
        relationType: { required: true, to_be_disambiguated: false },
        content: { required: false, to_be_disambiguated: false },
        impact: { required: false, to_be_disambiguated: false }
      }
      examples: [
        {
          Year: "184 AD",
          personA: "Liu Bei",
          personB: "Guan Yu",
          relationType: "Sworn Brotherhood",
          content: "Oath of the Peach Garden",
          impact: "Lifelong friendship"
        }
      ]
    }
  )
  Character_Interaction (
    description: "Records important interaction relationships between historical figures, including ruler-minister relations, alliances, betrayals and surrenders, strategic confrontations, etc."
    Ruler_Minister_Relation (
      description: "Important interactions between rulers and ministers, including political relationship changes such as conferment of titles, appointment and dismissal, rewards and punishments, and ruler-minister dialogues."
      arguments: {
        Year: { required: true, to_be_disambiguated: false },
        ruler: { required: true, to_be_disambiguated: true },
        minister: { required: true, to_be_disambiguated: true },
        relationType: { required: true, to_be_disambiguated: false },
        content: { required: false, to_be_disambiguated: false },
        location: { required: false, to_be_disambiguated: false }
      }
      examples: [
        {
          Year: "208 AD",
          ruler: "Cao Cao",
          minister: "Xun Yu",
          relationType: "Conferment of Title",
          content: "Conferment the title of Marquis of Wansui Ting",
          location: "Xuchang"
        }
      ]
    }
    Alliance_Formation (
      description: "The establishment of political or military alliances, including marriage alliances, military alliances, political cooperation, etc."
      arguments: {
        Year: { required: true, to_be_disambiguated: false },
        alliancePartyA: { required: true, to_be_disambiguated: true },
        alliancePartyB: { required: true, to_be_disambiguated: true },
        allianceType: { required: true, to_be_disambiguated: false },
        originalMaster: { required: false, to_be_disambiguated: true },
        newMaster: { required: false, to_be_disambiguated: true },
        majorPersonA: { required: false, to_be_disambiguated: true },
        majorPersonB: { required: false, to_be_disambiguated: true },
        purpose: { required: false, to_be_disambiguated: false }
      }
      examples: [
        {
          Year: "208 AD",
          alliancePartyA: "Liu Bei",
          alliancePartyB: "Sun Quan",
          allianceType: "Military Alliance",
          originalMaster: "Liu Bei",
          newMaster: "Sun Quan",
          majorPersonA: "Liu Bei",
          majorPersonB: "Sun Quan",
          purpose: "Resist Cao Cao"
        }
      ]
    }
    Betrayal_Surrender (
      description: "A shift in political stance, including actions like defection, surrender, changing sides, etc."
      arguments: {
        Year: { required: true, to_be_disambiguated: false },
        traitor: { required: true, to_be_disambiguated: true },
        originalMaster: { required: true, to_be_disambiguated: true },
        newMaster: { required: true, to_be_disambiguated: true },
        betrayalType: { required: true, to_be_disambiguated: false },
        reason: { required: false, to_be_disambiguated: false }
      }
      examples: [
        {
          Year: "208 AD",
          traitor: "Cai Mao",
          originalMaster: "Liu Bao",
          newMaster: "Cao Cao",
          betrayalType: "Surrender",
          reason: "Forced by circumstances"
        }
      ]
    }
  )
  Economy (
    description: "Analyzes the expansion of trade networks and the scramble for resources, revealing the interaction laws between productive forces and production relations."
    Trade (
      description: "Refers to events involving the formation, interruption, or reconstruction of cross-regional resource circulation paths, such as the Silk Road connecting Eurasian silk and spice trade through camel caravans, or the triangular trade system established by colonists during the Age of Discovery reshaping global resource distribution."
      arguments: {
        startYear: { required: true, to_be_disambiguated: false },
        endYear: { required: true, to_be_disambiguated: false },
        tradersA: { required: true, to_be_disambiguated: true },
        majorPersonA: { required: false, to_be_disambiguated: true },
        majorPersonB: { required: false, to_be_disambiguated: true },
        tradersB: { required: true, to_be_disambiguated: true },
        majorPersonB: { required: false, to_be_disambiguated: true },
        tradingContent: { required: false, to_be_disambiguated: false }
      }
      examples: [
        {
          startYear: "-114",
          endYear: "127",
          tradersA: "China",
          majorPersonA: "Zhang Qian",
          majorPersonB: "",
          tradersB: "Countries in West Asia",
          tradingContent: "Silk"
        }
      ]
    }
    Agricultural_Policy (
      description: "Policy adjustments related to ancient agriculture, such as the Tunian system (military-agricultural colonies), the Juntian system (equal-field system), and other land and agricultural management systems."
      arguments: {
        Year: { required: true, to_be_disambiguated: false },
        policyName: { required: true, to_be_disambiguated: false },
        implementer: { required: true, to_be_disambiguated: true },
        region: { required: false, to_be_disambiguated: false },
        policyContent: { required: false, to_be_disambiguated: false },
        effect: { required: false, to_be_disambiguated: false }
      }
      examples: [
        {
          Year: "196 AD",
          policyName: "Tunian System",
          implementer: "Cao Cao",
          region: "Xuchang",
          policyContent: "Military and civilian Tunian",
          effect: "Increased grain production"
        }
      ]
    }
  )
)

```

Figure 6: Schematic overview of the Three Kingdoms Historical Event Extraction Ontology. The ontology is structured hierarchically into eight primary domains: Politics, Character Interaction, Economy, Culture & Academia, Military, Environment, Global Civilization, and Society. These domains further categorize specific historical event types (e.g., Alliance Formation, Key Battle, Regime Change), defining the semantic scope and necessary arguments (such as participants, temporal slots, and locations) required for structured event extraction from unstructured historical texts.

Please extract historical events from the following classical Chinese text of the *Records of the Three Kingdoms* and return them in the specified JSON format. Note:

- 1.The event type must use one of the following types: {event\_type\_list}
- 2.Strictly adhere to the argument requirements in the ontology definition.
- 3.For classical Chinese content, argument values can be:
  - Entities such as names, locations, and times found in the original text.
  - Information inferred based on historical background and context.
  - Content where classical Chinese concepts are translated into modern understanding.
- 4.Temporal information can be inferred as concrete years based on reign titles, header information, etc. (e.g., "Xingping 1st Year" can be inferred as 194 AD).
- 5.Use formal historical appellations for person names (e.g., Cao Cao, Liu Bei, Sun Quan, etc.; do not use titles such as "Taizu").
- 6.For mandatory arguments, if they cannot be inferred from the text or context, discard the event.

The output format must be entirely consistent with the example.

Ontology Definition: {Ontology}

Example Format:

```
{
  "doc_id": "sanguo_001",
  "text": "兴平元年春，太祖自徐州还。初，太祖父嵩，去官后还谿，董卓之乱，避难琅邪，为陶谦所害，故太祖志在复讎东伐。夏，使荀彧、程昱守鄄城，复征陶谦，拔五城，遂略地至东海。还过郯，谦将曹豹与刘备屯郯东，要太祖。太祖击破之，遂攻拔襄贲，所过多所残戮。",
  "events": [
    {
      "event_type": "关键战役",
      "arguments": {
        "Year": "兴平元年",
        "location": "徐州",
        "majorPerson": "曹操",
        "majorPerson1": "陶谦",
        "participateA": "曹操",
        "participateB": "陶谦"
      }
    },
    {
      "event_type": "军事征服",
      "arguments": {
        "startYear": "兴平元年",
        "endYear": "兴平元年",
        "majorPerson": "曹操",
        "participateA": "曹操",
        "participateB": "陶谦"
      }
    },
    {
      "event_type": "君臣关系",
      "arguments": {
        "Year": "兴平元年",
        "ruler": "曹操",
        "minister": "荀彧",
        "relationType": "委任",
        "content": "守鄄城",
        "location": "鄄城"
      }
    }
  ]
}
```

Please return solely the extraction results in JSON format; do not include any additional content.

Figure 7: The instruction prompt template used for few-shot historical event extraction. The prompt includes explicit extraction rules, placeholders for input text/ontology, and a concrete few-shot example to guide the LLM's output structure. Note: To ensure historical authenticity and preserve linguistic integrity, the classical Chinese text within the example remains untranslated.

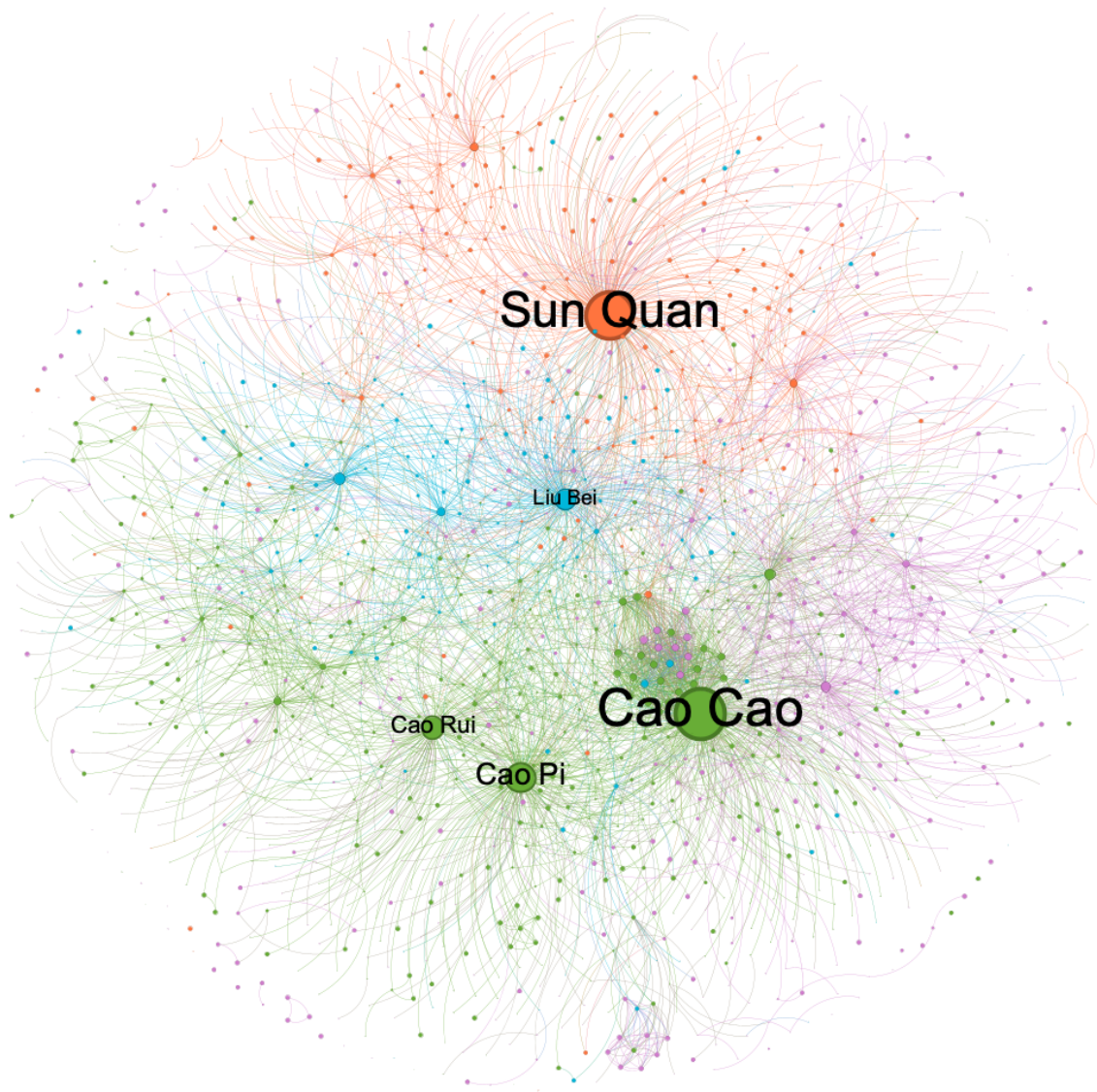


Figure 8: **Visualization of historical figure influence and factional structure in the *Records of the Three Kingdoms* knowledge graph.** Node sizes are proportional to the importance scores predicted by GenNIE, and colors represent political factions: **Wei (Green)**, **Wu (Orange)**, **Shu (Blue)**, and **Qun (Purple)**. The graph topology highlights the centrality of major leaders and the complex interplay between warring states.