# Statistical Network Analysis for LLMs via Knowledge Graphs

Lin Tang[1], Xiaohang Zhang[1], Xuezhou Zhang[2], Xiaowei Yu[3], Huimin Cheng[1*]

[1]Department of Biostatistics, Boston University, MA, USA
[2]Faculty of Computing and Data Sciences, Boston University, MA, USA
[3]Department of Computer Science, Missouri University of Science and Technology, MO, USA
[*]Corresponding Author: Huimin Cheng (huimin23@bu.edu)

September 16, 2025

**Abstract**

Knowledge Graphs (KGs)-enhanced LLMs have emerged as promising approaches to address fundamental limitations in language models, including hallucinations, inconsistent responses, and lack of explainability. Since KGs are inherently networks, the methods of statistical network analysis provide a powerful toolkit. Yet these research areas rarely interact, hindering potential advancement in both fields. This paper bridges this gap by identifying two critical, bidirectional research directions. First, from statistics to LLM applications: while advanced statistical methods exist for community detection and network analysis, KG applications like GraphRAG predominantly use simple algorithms (e.g., Leiden), overlooking the rich statistical toolkit that could substantially improve performance. Second, from LLM applications to new statistical methods: existing statistical link prediction methods assume homogeneous edge types and cannot handle the heterogeneous relationships inherent in knowledge graphs, motivating the development of new statistical approaches tailored to KG structures. Our comprehensive review covers KG construction from text, community detection for retrieval-augmented generation, link prediction for completion, and emerging directions, including uncertainty quantification and extensions to multi-layer, hypergraph, and temporal networks. This paper is the first to systematically connect a problem-rich domain (LLMs) with a solution-rich domain (statistics), unlocking significant potential new research opportunities.

*Keywords:* LLM, KG, statistical network analysis, community detection, link prediction, graph alignment

# 1 Introduction

Large Language Models (LLMs) such as GPT-5, Grok-4, and Gemini have transformed natural language processing and artificial intelligence, enabling remarkable progress in machine translation, summarization, information retrieval, and question answering (Qin et al., 2024; Minaee et al., 2024). However, despite their impressive performance, LLMs face several fundamental limitations. First, LLMs hallucinate: they confidently produce fluent but unsupported or false statements because next-token prediction is not equivalent to fact retrieval (Lin et al., 2021; Ji et al., 2023; Li et al., 2023; Banerjee et al., 2025). Second, LLMs exhibit inconsistent responses, often providing different answers to identical or semantically equivalent questions due to their reliance on probabilistic generation rather than access to a stable, authoritative knowledge base (Krügel et al., 2023; Ahn and Yin, 2025). Third, while LLMs excel at language understanding and generation, they frequently struggle with complex reasoning tasks that require multi-step logical inference or understanding intricate relationships between entities across multiple domains (Li et al., 2025b; Patil and Jadon, 2025). Finally, the lack of explainability makes it nearly impossible to trace the reasoning process behind an LLM's response, creating a significant barrier to deployment in applications where transparency and accountability are paramount (Palikhe et al., 2025). Together, these issues motivate integrating LLMs with external, structured representations that (i) constrain generations to verifiable facts, (ii) support compositional queries, and (iii) expose provenance for inspection.

Knowledge Graphs (KGs) offer a promising approach to mitigate these limitations by providing a structured, factual foundation that complements the generative strengths of LLMs (Lewis et al., 2020; Agrawal et al., 2023; Pan et al., 2024; Lavrinovics et al., 2025). A KG is a structured representation of entities (e.g., drugs, diseases, genes) and their relationships (e.g., “treats,” “causes”), typically modeled as a graph with heterogeneous node and edge types. KGs enable the retrieval of precise, verified data to ground LLM responses. For instance, by supplying an LLM with relevant facts extracted from a KG, the risk of hallucinations is substantially reduced (Agrawal et al., 2023; Pusch and Conrad, 2024; Kau et al., 2024; Li et al., 2025a), as the model shifts from relying on its parametric memory to synthesizing coherent answers from provided context. This grounding also promotes consistency, ensuring that responses to equivalent queries remain stable and reliable over time (Pan et al., 2024; Ibrahim et al., 2024). Moreover, the explicit relational structure of KGs facilitates deeper reasoning, allowing LLMs to handle complex queries involving multi-hop inferences or nuanced entity interactions (Chen et al., 2024; Chakraborty, 2024). Finally, KGs enhance explainability, as the retrieved information can be traced back to verifiable sources (Abu-Rasheed et al., 2024). For example, in biomedical applications, an LLM might suggest that drug A may be repurposed to treat disease B. A KG constructed from biomedical literature and databases can provide supporting evidence through intermediate relationships, such as drug A targets protein C, and protein C is implicated in disease B. Such reasoning is difficult for LLMs trained only on text, but becomes tractable when statistical or algorithmic methods are applied to structured KGs.

A prominent example of this integration is GraphRAG (Graph-enhanced Retrieval-Augmented Generation) (Edge et al., 2024), which leverages community detection algorithms to partition knowledge graphs into meaningful clusters before retrieval and generation. Community detection serves as a crucial step in GraphRAG, enabling more targeted and contextually relevant information retrieval. Community detection has been extensively studied in the statistical network analysis literature, with methods including stochastic block models (Bickel and Chen, 2009; Abbe, 2018) and their variants (Airoldi et al., 2008;

Karrer and Newman, 2011; Jin et al., 2024b), spectral clustering (Rohe et al., 2011), and modularity optimization (Newman, 2006). However, due to the disciplinary gap between computer science (CS) and statistics, applications in KG-enhanced LLMs often default to CS-oriented algorithms like Leiden (Traag et al., 2019) and Louvain (Blondel et al., 2008), overlooking the rich statistical toolkit. This represents a significant opportunity: investigating how established statistical community detection methods can be adapted and optimized for knowledge graph structures to enhance LLM performance.

Similarly, link prediction for KGs, the task of inferring missing relationships between entities, plays a crucial role in KG completion and reasoning. Existing link prediction methods in CS have been specifically designed for KGs, incorporating techniques like knowledge graph embeddings (Bordes et al., 2013; Wang et al., 2014; Lin et al., 2015; Ji et al., 2015; Sun et al., 2019a) and neural relational models (Schlichtkrull et al., 2018; Vashishth et al., 2020a). However, the rich tradition of link prediction methods in statistical network analysis, including graphon models (Zhang et al., 2017; Chandna et al., 2022), latent space models (Hoff et al., 2002), exponential random graph models (Snijders et al., 2002; Robins et al., 2007), and matrix factorization approaches (Nickel et al., 2011), has seen limited application to KGs due to a fundamental incompatibility: most statistical methods assume homogeneous edge types, while KGs inherently contain heterogeneous relationships. This gap motivates another promising research direction: developing new statistical methods motivated by the practical needs of KGs in LLM applications, such as new graphon models that account for edge heterogeneity for the link prediction task.

While existing surveys have separately explored the benefits of KGs for LLMs (Agrawal et al., 2023; Pan et al., 2024; Jin et al., 2024a; Kau et al., 2024) and the methods of statistical network analysis (Goldenberg et al., 2010; Rácz and Bubeck, 2017; Borsboom et al., 2021; Sengupta, 2025; Kim, 2025), a critical gap remains in connecting these two domains. They have not explicitly linked the sophisticated tools of statistical network analysis with the challenges and opportunities in using KGs to ground and enhance LLMs. This paper aims to bridge that gap by providing a focused, bidirectional perspective. We explore how established and emerging methods in statistical network analysis can advance KG-based LLM applications, and conversely, how the practical needs of KGs can motivate new developments in statistical network analysis. Through this synthesis, we hope to inspire novel research directions and foster a deeper collaboration between the statistics and computer science communities.

The remainder of this paper is organized as follows. In Section 2, we first review how to construct KG from textual data. Although the primary focus of this paper is the synergy between statistical network analysis and KG-driven LLM applications, KG construction is an indispensable prerequisite. In practice, high-quality, domain-specific KGs are rarely available and must be built from unstructured sources like documents or reports. For instance, before statistical link prediction models can be applied to complete a graph, that graph must first exist. Similarly, in applications like GraphRAG for text summarization, a KG is first extracted from the long-form text to provide a structured foundation for the LLM. Therefore, understanding this foundational text-to-KG step is crucial, as the quality and structure of the resulting graph directly determine the feasibility and effectiveness of the downstream statistical analyses we discuss. In Section 3, we review how community detection is applied in KGs for LLM enhancement. Community detection is a widely studied area in statistical network analysis, with extensive theoretical foundations and methodological developments. However, how we can further leverage these statistical advances to help LLM systems remains unclear and underexplored. In Section 4, we address the critical task of completing and refining the KG through link prediction. Here, we review current approaches and pinpoint the need for new statistical models that can handle the complex, heterogeneous relationships inherent in knowledge graphs. In Section 5, we discuss some other emerging possibilities, including graph alignment for hallucination detection, uncertainty quantification for LLM answers using pairwise graph distance, and extensions to more complex graph structures such as multi-layer graphs, hypergraphs, and time-varying graphs.

# 2 Knowledge Graph Construction

This section provides a comprehensive review of knowledge graph construction methodologies, examining the fundamental pipeline architecture, taxonomic organization of existing approaches, and empirical evaluation of different techniques. Our analysis focuses on methods that extract factual knowledge from text, excluding approaches that primarily target schema construction or ontology learning from structured sources.

## 2.1 Knowledge Graph Construction Pipeline

The KG construction pipeline transforms raw text into a coherent graph through the following standard sequential stages, where preprocessing and postprocessing steps are optional, while named entity recognition and relation extraction steps are key to KG construction.

**Step 1: Preprocessing.** This foundational step refines input text to mitigate ambiguities. Key sub-tasks include coreference resolution (e.g., linking "he" to "Steve Jobs") and abbreviation expansion (e.g., "AAPL" to "Apple Inc."), facilitating accurate downstream extraction. These processes ensure cleaner data for downstream tasks, reducing errors in entity and relation detection.

**Step 2: Named Entity Recognition (NER).** NER involves the identification and classification of entity mentions within the text, such as persons, organizations, locations, or dates, which form the nodes of the knowledge graph. This process must accurately detect entity boundaries, resolve ambiguous mentions, and assign appropriate semantic types according to a predefined schema or ontology. Ambiguity often complicates this task: "John Smith" may refer to a person or a company, "May" can denote a person's name or the month, and "Washington" could represent a name, a city, or a state. Misclassification at this stage can cascade into downstream errors, making NER one of the most critical and error-prone components of the pipeline.

**Step 3: Relation Extraction (RE).** Following entity identification, RE focuses on detecting and classifying semantic relationships between the identified entities, generating triples (e.g., entity-relation-entity) that serve as the edges in the knowledge graph. This task is particularly challenging, as it requires understanding not only lexical patterns that signal relationships but also broader contextual, syntactic, and semantic structures to determine roles and directionality.

**Step 4: Post-processing.** In this final phase, the extracted entities and relations are refined and integrated into a cohesive graph structure. Key processes include entity linking (mapping extracted mentions to canonical entries in established knowledge bases like Wikidata (Vrandečić and Krötzsch, 2014) or DBpedia (Lehmann et al., 2015) to ensure uniqueness); entity resolution and deduplication (merging variant representations, such as "Apple" and "Apple Inc." using similarity metrics and contextual analysis while preserving aliases); and relation canonicalization (standardizing diverse expressions, e.g., unifying "founded by", "established by", or "created by" to a consistent "founder" relation in the schema).

## 2.2 Taxonomy of KG Construction Methods

We organize knowledge graph construction methodologies according to their fundamental approaches to named entity recognition and relation extraction, the two core components of the construction pipeline. Our classification identifies four primary method categories: rule-based, traditional machine learning (ML), deep learning, and LLMs-based.

Rule-based methods represent early paradigms in knowledge graph construction, employing hand-crafted linguistic patterns, grammars, and domain-specific rules to perform entity recognition and relation extraction without requiring training data. Early prominent implementations include domain-specific tools such as MedLEE (Friedman et al., 1995) and MetaMap (Aronson, 2001) in biomedicine, which combine rule-based parsing with specialized dictionaries to identify clinical concepts. General-domain frameworks like GATE's JAPE (Cunningham et al., 2000) and Stanford's TokensRegex (Chang and Manning, 2014) provide flexible rule-crafting environments for extraction tasks. Notable successes include rule-based systems dominating clinical NLP competitions, such as the 2014 i2b2 de-identification challenge where top-performing systems employed primarily pattern-matching approaches (Stubbs et al., 2015). Recent work continues this tradition with specialized applications like the R-MIMO model (Chen et al., 2023) for geological knowledge graphs, which uses expert-defined grammar rules to extract complex fact-condition statements with domain-specific precision. Rule-based approaches have advantages, include high interpretability, elimination of labeled training data requirements, and low computational overhead. However, these benefits are offset by poor scalability to diverse textual domains, substantial manual effort for rule design, and limited robustness to linguistic ambiguity and variation (Zhong et al., 2023; Zhang et al., 2025).

Traditional ML methods are another type of early methods, bridge rule-based and neural approaches, using supervised or semi-supervised algorithms with hand-engineered features for NER and RE. The key task now became feature engineering: designing and extracting informative signals from text to serve as input for a statistical classifier. These methods dominated pre-2018 literature but remain relevant in resource-constrained or explainable KG construction. NER often employs models like Conditional Random Fields (CRF) (Lafferty et al., 2001) or Hidden Markov Models (HMM) (Rabiner, 2002), incorporating features such as part-of-speech tags, word shapes, and gazetteers. For RE, techniques include Support Vector Machines (SVM) (Cortes and Vapnik, 1995) for classification or bootstrapping (e.g., distant supervision) (Brin, 1998; Agichtein and Gravano, 2000; Mintz et al., 2009) to expand labeled data from seed patterns. While effective in structured domains, these methods suffer from feature engineering overhead and sensitivity to noisy data, leading to their displacement by neural approaches in most contemporary applications. In this paper, we maninly focus on the deep learning and LLM-based methods.

### 2.2.1 Deep Learning Approaches

The taxonomy classifies DL methods based on their primary task (separate NER, separate RE, or joint NER-RE) and underlying architectures. Architectures are grouped into RNN-based (for sequential dependencies), CNN-based (for local feature extraction), Transformer-based (for contextual embeddings), GNN-based (for structural relationships), and hybrid/other (combinations or specialized variants).

**NER Only Methods.** Deep learning methods for NER primarily focus on sequence labeling and context modeling. Early RNN-based approaches, such as BiLSTM-CRF (Huang et al., 2015), used bidirectional recurrence to capture sequential context from both directions, while employing a CRF decoding layer to enforce valid tag structures Huang et al. (2015); Ma and Hovy (2016). These models achieved strong performance on benchmarks like CoNLL-2003 with F1 scores around 90-92%. CNN-based approaches, such as ID-CNN (Strubell et al., 2017) and character-level CNNs (Chiu and Nichols, 2016), extracted local contextual features efficiently and were particularly effective for handling out-of-vocabulary words Strubell et al. (2017); Gridach (2017). Transformer-based architectures, exemplified by BERT (Devlin et al., 2019), TENER (Yan et al., 2019a), and LUKE (Yamada et al., 2020), became dominant from 2019 onward, leveraging large-scale pre-training and fine-tuning to produce superior contextual representations, achieving F1 scores of about 93% or higher on CoNLL-2003. This shift improved over RNNs by better handling long-range dependencies without recurrence. In parallel, graph-based models employed graph convolutional networks (GCNs) with dependency-aware variants incorporating syntactic dependencies to improve NER performance (Jie and Lu, 2019). For handling nested or overlapping entities, span-based or span-graph methods apply GNNs over span graphs and yield strong results (Fei et al., 2021; Wan et al., 2022). Hybrid models (Ma and Hovy, 2016; Lin et al., 2019; Wang et al., 2020a; Song et al., 2020; Sun and Bhatia, 2021) that combined RNNs, CNNs, and external knowledge (e.g., gazetteers) further pushed NER accuracy up to 94% and better handled more complex datasets.

**RE Only Methods.** Separate RE models typically assume entities are pre-identified and focus on classifying relations between entity pairs. RNN-based methods used BiLSTM (Zhou et al., 2016; Lee et al., 2019) or BiGRU encoders (Jat et al., 2018), often enhanced with attention, to capture sentence-level dependencies for supervised RE, achieving F1 scores of 80-85% on SemEval-2010. For distantly supervised settings, models like BiGRU (Jat et al., 2018) handled noisier data on datasets like NYT (Riedel et al., 2010), though with lower performance. CNN-based models, such as PCNN (Zeng et al., 2015), applied convolution and pooling over sentences with entity position embeddings, achieving an average precision of 78.3% on top extracted relation instances on NYT. Graph-based methods extended RE beyond sentence boundaries. By modeling entities and their interactions with GNNs, models such as GCN over Pruned Dependency Trees (Zhang et al., 2018) and AGGCN (Guo et al., 2019) achieved about 85% F1 on SemEval-2010 and 50-60% F1 on DocRED. With the rise of PLMs, Transformer-based approaches like BERT (Devlin et al., 2019) and RoBERTa (Liu et al., 2019) variants quickly became the new standard, leveraging pre-training to handle long-range dependencies and document-level contexts with F1 scores of 88-90% on SemEval-2010 and around 60-65% on more challenging document-level benchmarks like DocRED (Soares et al., 2019; Ye et al., 2020; Zeng et al., 2020b; Zhao et al., 2024; Cabot and Navigli, 2021; Zhou et al., 2021). Hybrid designs, such as attention-enhanced CNN/RNN models or multi-instance learning frameworks, offered flexibility in handling noisy training data, performing strongly on large-scale distantly supervised datasets like NYT (Han et al., 2018; Surdeanu et al., 2012).

**Joint NER and RE Methods**. Joint models have gained increasing attention as they mitigate error propagation by performing entity recognition and relation extraction in a unified framework. RNN-based models, such as CopyMTL (Zeng et al., 2020a), OrderRL (Zeng et al., 2019) and RSAN (Yuan et al., 2021), introduced parameter-sharing mechanisms and copy strategies to simultaneously identify entities and relations, achieving F1 scores of about 72-84% on NYT, and F1 scores of about 56-82% on WebNLG. Li et al. (2017) integrated convolutional encoders with stacked Bi-LSTM-RNN to extract entity and relation features in parallel, reaching around 84% F1 on entity recognition and about 71% in relation extraction of the ADE biomedical task. Transformer-based methods currently dominate this space. Generative frameworks such as TANL (Paolini et al., 2021), TDEER (Li et al., 2021a), UniRel (Tang et al., 2022), and DEEPSTRUCT (Wang et al., 2022a), along with iterative encoder-based approaches like ITER (Hennen et al., 2024), reformulate joint extraction as sequence-to-sequence learning, allowing overlapping triples and achieving 90-95% F1 on NYT and WebNLG. GNN-based models, such as GraphRel (Fu et al., 2019) and RIFRE (Zhao et al., 2021), employ graph structures to capture complex interdependencies and overlapping relations, reporting about 62% and 93% F1 on NYT. Hybrid architectures further advanced the field: TPLinker (Wang et al., 2020b) employed token-pair linking to extract triples, K-BERT (Liu et al., 2020) leveraged knowledge-enhanced BERT representations, and span-based frameworks like SpERT (Eberts and Ulges, 2019) modeled entities as spans and relations as span pairs, yielding robust performance across datasets.

### 2.2.2 LLMs-Based Approaches

Approaches to leveraging LLMs for NER and RE tasks fall into two broad categories: prompt-based (in-context learning without parameter updates) and fine-tuning-based (adapting model weights).

Prompt-based methods can be broadly classified into four categories: basic prompting, advanced reasoning prompts, ensemble prompting, and domain-guided prompts. **Basic Prompting**: This category includes zero-shot and few-shot methods. (1) Zero-shot prompting directly instructs the LLM to extract entities or relations without providing examples. For instance, Ghanem and Cruz (2024b,a) evaluated zero-shot text-to-knowledge graph (T2KG) methods across LLaMA-2, Mistral, and Starling, finding that raw zero-shot performance provides a fast baseline for KG construction. Iterative zero-shot pipeline (Carta et al., 2023) that decompose extraction into repeated prompt/refinement steps (GPT-3.5) also report strong entity and triplet extraction performance on domain corpora, similar to tools like KGGen (Mo et al., 2025) that use language models for direct text-to-KG generation. Likewise, ChatIE (Wei et al., 2023a), a multi-turn zero-shot framework instantiated with ChatGPT that converts IE into staged QA, outperforms single-turn prompting and, on some datasets, matches or exceeds few-shot and supervised baselines. (2) Few-shot prompting extends this by including a handful of annotated examples in the prompt. GPT-NER (Wang et al., 2023a) reframes NER as a text generation problem with entity markers and self-verification, showing that few-shot prompting achieves accuracy comparable to supervised models, particularly in low-resource settings. Similarly, Wadhwa et al. (2023) demonstrated that GPT-3, when prompted in a few-shot manner, could extract relation triples from datasets such as CoNLL04 and NYT with performance surprisingly close to supervised baselines. Ashok and Lipton (2023) introduced Prompt-NER, which leverages carefully crafted prompts with in-context examples (including entity explanations) for robust NER in low-resource domains and cross-domain settings.

**Advanced Reasoning: Chain-of-Thought (CoT).** CoT prompting, popularized by Wei et al. (2022), encourages the model to think step by step before producing a final output, which has been adapted for RE tasks requiring multi-hop reasoning. For instance, Xue et al. (2024) introduced AutoRE, a system that decomposes RE into multiple reasoning subtasks—first predicting candidate relations, then identifying head entities, and finally extracting full triples—thereby improving performance on complex benchmarks like DocRED. In NER, Wang et al. (2023b) introduced a self-verification strategy within GPT-NER, where the model checks its own extracted entities against the type definition, reducing hallucinations.

**Ensemble Prompting**: Ensemble Prompting seeks to mitigate prompt sensitivity and variance by combining multiple prompts or models. Pitis et al. (2023) used boosted prompt ensembles that iteratively construct few-shot prompts by selecting "hard" examples from a small dataset. Aggregating the outputs from this ensemble outperforms single-prompt baselines on reasoning tasks. For NER, Islam et al. (2025) introduced a prompt ensemble for reliable medical entity recognition from electronic health records using LLMs like GPT-4o, with improved performance and reliability through embedding-based similarity and majority voting. For RE, Sivarajkumar et al. (2024) evaluated ensemble prompts that combine various strategies (e.g., prefix, CoT) via majority voting for zero-shot clinical NLP tasks, including relation extraction, to balance strengths and improve accuracy. As an alternative to manual prompt design, heuristic-based search algorithms can automate prompt optimization (Cui et al., 2025). Although frameworks like APE (Zhou et al., 2022), TextGrad (Yuksekgonul et al., 2024), and AutoPDL (Spiess et al., 2025) are designed to discover a single, highly-optimized prompt or agent configuration for a given task, they can generate multiple optimized variants whose outputs can then be ensembled for higher accuracy (Tonolini et al., 2024). Zhu et al. (2024) extended this approach to multi-agent ensembles in their AutoKG framework, where different LLMs (e.g., GPT-4, ChatGPT) are prompted separately, and their outputs are reconciled into consistent triples. Similarly, Lu and Wang (2025) advanced this paradigm with the KARMA framework, leveraging multi-agent LLMs for automated KG enrichment through entity discovery, relation extraction, and conflict resolution via multi-layer verification to reduce discrepancies. Ensemble prompting thus provides a practical strategy to reduce the variance of LLM outputs and improve robustness in KG construction pipelines.

**Domain-Specific Enhancements: Guidance-Based Prompting**. Domain-guided prompting tailors information extraction for knowledge graph construction by embedding domain-specific constraints, ontologies, or guidelines into prompts for LLMs. This approach ensures that the generated outputs align closely with specialized schemas and requirements in various fields. For instance, taxonomy-driven methods leverage hierarchical taxonomies to facilitate domain-specific KG construction in scientific applications (Pan et al., 2025). Similarly, the LLM-ACNC (Liu et al., 2025) framework employs guided prompting with LLMs to build KGs from aerospace requirement texts, capturing complex relationships in engineering documents. In the clinical domain, prompting techniques have been used to construct patient journey KGs directly from unstructured consultation notes and documentation, enabling better tracking of healthcare pathways (Al Khatib et al., 2025). Furthermore, multi-step guided prompting assists in structuring KGs for nuclear fusion energy, where LLMs iteratively extract and organize domain-specific entities and relations from scientific literature (Loreti et al., 2025). Prompt-based methods can also generate domain-specific KGs solely from the internalized knowledge in LLMs' parameters, without relying on external corpora, by using schema-guided procedures (Parović et al., 2025; Bi et al., 2024b). In the service domain, the BEAR system (Yu et al., 2023) applies LLM-guided prompting to construct comprehensive KGs, automatically generating prompts from a domain ontology to ensure alignment with specialized schemas. Likewise, the application of prompt engineering in machining processes, where LLMs are guided to build process KGs for intelligent planning, integrating domain-specific terminologies and relations (Xu et al., 2025).

Another key category involves fine-tuning methods, which adapt LLM parameters using task-specific, labeled datasets to improve accuracy in NER and RE tasks for KG construction. This is particularly effective in specialized domains like biomedicine and bioinformatics, where instruction tuning can adapt general-purpose LLMs to achieve superior performance (Monajatipoor et al., 2024; Keloth et al., 2024; Ghanem and Cruz, 2025). For example, models like BioBERT (Lee et al., 2020) and CinicalBERT (Alsentzer et al., 2019), fine-tuned from BERT (Devlin et al., 2019), have demonstrated strong performance in the biomedical domain (Tinn et al., 2023; Li et al., 2022a). Similarly, models such as REBEL (Cabot and Navigli, 2021) illustrate how end-to-end language generation can be fine-tuned for relation extraction, simplifying triplet prediction in KG pipelines and adapting well to domain-specific needs.

## 2.3 Benchmark Datasets and Evaluation for KG Construction

**Datasets.** Table 1 surveys widely used benchmarks for text-to-KG construction across three families: (i) general-domain corpora (e.g., CoNLL-2004 (Roth and Yih, 2004), SemEval-2010 Task 8 (Hendrickx et al., 2019), NYT (Riedel et al., 2010), WikiReading (Hewlett et al., 2016), WebNLG (Gardent et al., 2017), WIKI-TIME (Yan et al., 2019b), DocRED (Yao et al., 2019b), FewRel 2.0 (Gao et al., 2019), Re-TACRED (Stoica et al., 2021), and Text2KGBench variants (Mihindukulasooriya et al., 2023)); (ii) domain-specific datasets in biomedicine, law, and finance (ADE (Gurulingappa et al., 2012), DDI (Herrero-Zazo et al., 2013), ChemProt (Krallinger et al., 2017), SciERC (Luan et al., 2018), CUAD (Hendrycks et al., 2021), FinRED (Sharma et al., 2022)); and (iii) multilingual resources (ACE-2005 (Walker et al., 2006), SMiLER (Seganti et al., 2021)). These resources differ markedly in schema complexity, task formulation (sentence- vs. document-level joint NER+RE, temporal RE, clause extraction, ontology-guided KG generation), supervision regime (manual vs. distantly supervised; few-shot vs. large-scale), and scale, collectively providing complementary test beds for evaluating end-to-end KG construction pipelines.

**Evaluation metrics.** Evaluating KG construction quality requires different metrics depending on the approach used. We categorize evaluation metrics into the following groups: (1) Canonical metrics. Traditional supervised extraction uses Precision (fraction

Table 1: Common Benchmark Datasets for Knowledge Graph Construction.

| Dataset | Domain | Entity Types | Relation Types | Train/Test Sizes | Year | Primary Tasks |
|---|---|---|---|---|---|---|
| ***General Domain*** | | | | | | |
| CoNLL-2004 (Roth and Yih, 2004; Gupta et al., 2016) | General | 4 | 5 | 922/288 | 2004 | Joint NER+RE |
| SemEval-2010 Task 8 (Hendrickx et al., 2019) | General | N/A | 19 | 8k/2.7k | 2010 | Multi-way RE |
| NYT (Riedel et al., 2010) | General | 3 | 24 | 56k/5k | 2010 | Distantly Supervised RE |
| WikiReading (Hewlett et al., 2016) | General(Document-level) | N/A | 884 | 14.9M/3.7M | 2016 | Large-scale RE |
| WebNLG (Gardent et al., 2017) | General | 15 | 171 | 5k/0.7k | 2017 | RE |
| WIKI-TIME (Yan et al., 2019b) | General(Temporal) | 3 | 57 | 97.6k/40k | 2019 | Temporal RE |
| DocRED (Yao et al., 2019b) | General(Document-level) | 6 | 96 | 3008/700 | 2019 | Joint NER+RE (Doc-level) |
| FewRel 2.0 (Gao et al., 2019) | General(Few-shot) | N/A | 100 | 56k/14k | 2019 | Few-shot RE |
| Re-TACRED (Stoica et al., 2021) | General | 17 | 40 | 58.5k/13.4k | 2021 | RE |
| Text2KGBench (Wikidata-TeKGen)(Mihindukulasooriya et al., 2023) | General | 10 ontologies | Variable | 13k | 2023 | Ontology-driven KG Gen |
| Text2KGBench (DBpedia-WebNLG)(Mihindukulasooriya et al., 2023) | General | 19 ontologies | Variable | 4.8k | 2023 | Ontology-driven KG Gen |
| ***Specific Domain*** | | | | | | |
| ADE (Gurulingappa et al., 2012) | Biomedical | 2 | 1 | 4.3k/1k | 2012 | Joint NER+RE |
| DDI (Herrero-Zazo et al., 2013) | Biochemical | 1 | 5 | 25.3k/5.7k | 2013 | RE |
| ChemProt (Krallinger et al., 2017) | Biochemical | 2 | 10 | 19.5k/16.9k | 2017 | RE |
| SciERC (Luan et al., 2018) | Scientific | 6 | 7 | 1.8k/0.55k | 2018 | Joint NER+RE |
| CUAD (Hendrycks et al., 2021) | Legal | N/A | 41 | 10.48k/2.62k | 2021 | Clause Extraction |
| FinRED (Sharma et al., 2022) | Finance | 4 | 29 | 5.7k/1k | 2022 | Joint NER+RE |
| ***Multi-lingual*** | | | | | | |
| ACE-2005 (Walker et al., 2006) | General | 7 | 6 | 1k/0.5k | 2005 | Joint NER+RE, Event Extraction |
| SMiLER (Seganti et al., 2021) | General | 3 | 36 | 733k/15k | 2021 | Joint NER+RE |

of extracted triples that are correct), Recall (fraction of gold-standard triples recovered), and F1-score (harmonic mean of both). These are computed at the triple level, often with micro- or macro-averaging (Yao et al., 2019b). To handle lexical variations, evaluations often use relaxed matching through entity normalization or semantic similarity measures like G-BERTScore (Saha et al., 2021) rather than strict string matching. G-BERTScore extends the text generation metric BERTScore for graph-matching by treating graphs as sets of edges. It finds the optimal assignment between the edges of a predicted and a ground-truth graph, treating each edge as a sentence and scoring pairs with BERTScore (Zhang* et al., 2020). From this optimal assignment, it computes precision, recall, and a final F1-score. (2) Ranking-based metrics. Large-scale noisy training data (e.g., NYT), where evaluating all possible outputs is infeasible, necessitates ranking-based evaluation. Precision@K (P@K) assesses accuracy of top-K predictions. Mean Reciprocal Rank (MRR $= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\text{rank}_i}$) evaluates ranking of correct triples (Jiang et al., 2019; Bastos et al., 2021). Hit@K measures the proportion of instances with at least one correct triple in the top-K.

**Empirical results.** We empirically benchmarked three KG construction frameworks on CoNLL04 (Roth and Yih, 2004), ADE (Gurulingappa et al., 2012), and SciERC (Luan et al., 2018) datasets: a supervised fine-tuned model (ITER (Hennen et al., 2024)), two existing popular LLM-based frameworks (KGGen (Mo et al., 2025) with few-shot prompting, and CodeKGC (Bi et al., 2024b) with guidance-based prompting). The results, presented in Table 2, reveal a key trade-off between high performance on structured benchmarks and flexibility in real-world scenarios. To ensure a fair comparison, we report the G-BERTscore F1 (Saha et al., 2021) for the LLM-based frameworks and the Lenient F1-score for ITER.

Table 2: Empirical Results of KG Construction Frameworks

| **Dataset** | **ITER** | **KGGen** | **CodeKGC** |
|---|---|---|---|
| CoNLL04 | 0.7750 | 0.3113 | 0.6909 |
| ADE | 0.8750 | 0.5226 | 0.8977 |
| SciERC | 0.5070 | 0.4130 | 0.5282 |

As expected, the supervised ITER framework generally achieved the highest F1-scores, excelling on datasets with fixed schemas and labeled training data. In contrast, LLM-based frameworks offer superior zero- or few-shot capabilities, ideal for extracting knowledge from free-form text without predefined schemas or training data. Our experiments highlight several practical considerations for selecting a framework: (1) Choose supervised methods for tasks with abundant labeled data and fixed schemas to maximize accuracy. In contrast, LLM-based approaches are better suited for exploratory tasks, unlabeled data, or domains requiring schema flexibility. (2) Successfully deploying LLMs requires significant effort in prompt engineering to ensure structured outputs and post-processing to standardize relations and filter hallucinations. (3) While high-performance proprietary models like GPT-4 deliver strong results, their API costs and higher latency make cost-effective open-source or fine-tuned models more practical for scalable production environments.

# 3 Community Detection for GraphRAG

Retrieval-augmented generation (RAG) systems augment large language models with external knowledge to reduce hallucination and improve factual accuracy. Standard RAG implementations retrieve text chunks based on embedding similarity—an approach that works adequately for simple factual queries but fails when questions require synthesizing information across documents or understanding document-level themes. GraphRAG (Edge et al., 2024) addresses this limitation by introducing an intermediate graph representation layer between raw text and the language model. The GraphRAG pipeline extracts entities and relationships from documents, constructs a knowledge graph, applies community detection, and generates summaries of the resulting communities using LLMs. At query time, responses are generated by retrieving community summaries rather than isolated passages.

The use of knowledge graphs and community summaries has made GraphRAG highly influential, leading to multiple extensions. For instance, CommunityKG-RAG (Chang and Zhang, 2024) shifts the focus from retrieval and question answering to fact verification by treating communities as reasoning units and exploiting their internal coherence for zero-shot fact-checking. ArchRAG (Wang et al., 2025) develops an attributed, community-based hierarchical index that enriches entities with attribute information and supports efficient online retrieval, thereby improving accuracy while reducing token cost. Applications in specific domains have also emerged. In healthcare, Jiang et al. (2025) adapts the GraphRAG framework to augment electronic health records with graph-based community retrieval for interpretable, reasoning-enhanced clinical predictions like mortality and readmission risk assessment. In the legal domain, Zhai (2025) introduces LawGraph to handle temporal and hierarchical legal norms through versioned entity modeling, ensuring deterministic representations of evolving statutes and superior performance in legal information retrieval and analysis.

All such methods highly depend on a crucial step: community detection over the knowledge graph. This raises two fundamental statistical questions. First, how does the choice of community detection algorithm affect GraphRAG's performance? Second, how does the granularity of partitioning, that is, the number of detected communities, influence GraphRAG's performance? Addressing these questions is essential for understanding the statistical foundations of GraphRAG and for guiding principled methodological improvements.

## 3.1 Influence of Different Community Detection Methods

The effectiveness of GraphRAG-related research highly depends on the quality of community detection. Well-defined communities lead to coherent summaries, efficient retrieval, and facilitated multi-hop reasoning, while poorly detected communities result in incoherent summaries and loss of important distinctions. Initial GraphRAG implementations used the Leiden algorithm (Traag et al., 2019) primarily for computational efficiency. Nevertheless, recent work by Wang et al. (2025) demonstrated that the choice of detection method significantly impacts downstream performance, showing that spectral clustering and attribute-based algorithms can outperform Leiden in certain scenarios.

While these results highlight the importance of community detection methodological choice, they (Wang et al., 2025) only compared three methods: spectral clustering, Leiden, and their proposed attributed community detection by exploiting both links and the attributes of nodes. In statistics, community detection has been analyzed much more broadly, and a wide range of probabilistic and model-based methods remain unexamined in the context of GraphRAG.

Classical methods include spectral clustering based on graph Laplacians (Rohe et al., 2011), modularity maximization (Newman, 2006), and likelihood-based approaches under the stochastic block model (SBM) and its extensions (Bickel and Chen, 2009; Abbe, 2018). More sophisticated approaches have addressed degree heterogeneity and mixed memberships, including the degree-corrected SBM (Karrer and Newman, 2011), mixed-membership SBM (Airoldi et al., 2008), degree-corrected mixed membership (DCMM) model (Jin et al., 2024b). Other advances include tight community detection (Deng et al., 2024), which allows for scattered nodes outside any community, and network-adjusted covariates methods (Hu and Wang, 2024) that integrate node attributes with network topology.

## 3.2 Influence of Different Number of Communities

The number of detected communities is another important factor for GraphRAG. When there are too many communities, the graph is broken into very small groups. This increases computational cost and token usage, and often splits apart topics that should be summarized together. When there are too few communities, by contrast, distinct topics may be merged into a single group, which makes the summaries less informative and can reduce retrieval accuracy. It has been shown that different number of communities significantly impact the accuracy of GraphRAG (Edge et al., 2024). This raise the practical question: how do you choose the right number? While this question has been widely studied in statistical network analysis, existing GraphRAG implementations largely ignore this rich statistical literature, representing a significant missed opportunity.

**Statistical opportunities: leveraging community number selection to enhance GraphRAG.** The statistics literature provides three principled categories of methods that could be integrated into GraphRAG to refine community partitioning, thereby improving retrieval accuracy, summary coherence, and overall efficiency in knowledge-graph-based LLM applications. (1) **Cross-validation methods** adapt machine learning techniques to networks by splitting data and testing out-of-sample performance. Chen and Lei (2018) introduced Network Cross-Validation (NCV) for stochastic block models, while Li et al. (2020) improved this with edge-sampling cross-validation that works better for sparse networks by splitting edges rather than node pairs. (2) **Information criteria and likelihood-based methods** are designed specifically for model-based approaches like stochastic block models. WANG and BICKEL (2017) developed likelihood ratio tests based on log likelihood statistics, analyzing their asymptotic properties even under model misspecification. Later work extended this to degree-corrected SBMs (Yan et al., 2014) and overlapping community models (Latouche et al., 2014). (3) **Goodness-of-fit tests** provide another way to determine the appropriate number of communities by assessing how well a fitted model captures network structure. The first formal GOF test for SBMs was proposed by Lei (2016), using the largest singular value of the residual adjacency matrix to detect model misfit. Hu et al. (2021) improved this with entrywise deviation tests that handle linear degree growth. For DC-SBMs, Zhang and Amini (2023) introduced adjusted chi-square tests, and Karwa et al. (2024) developed Monte Carlo methods for finite-sample settings.

# 4 Link Prediction in LLMs

Link prediction has become a critical component for enabling LLMs to reason over structured knowledge and mitigate hallucination. In the context of KGs, link prediction aims to infer missing edges between entities, thereby enriching the graph and providing factual grounding for downstream reasoning tasks. This capability underpins a wide range of application scenarios. For instance, in biomedical KGs, link prediction can suggest new drug–disease or gene–phenotype associations, offering hypotheses for experimental validation. In open-domain question answering, predicted links fill in incomplete relations, allowing LLMs to navigate multi-hop reasoning chains more effectively. Moreover, in retrieval-augmented generation pipelines, link prediction supports knowledge alignment and contextual expansion by surfacing plausible but implicit relations that may not be explicitly encoded in text corpora. These scenarios illustrate the importance of link prediction as both a knowledge augmentation mechanism and a reasoning scaffold for LLMs.

**Unique challenges**. Link prediction on knowledge graphs presents unique challenges that distinguish it from traditional graph analysis. Unlike regular graphs where nodes are homogeneous and edges represent simple connectivity, nodes have distinct types (e.g., person, gene, drug, organization), and edges represent typed relations (e.g., treats, causes, located in). In regular graphs, link prediction can rely on structural heuristics like shortest path length (Liben-Nowell and Kleinberg, 2003), common neighbors (Newman, 2001), or Jaccard similarity (Jaccard, 1901) between node neighborhoods. Knowledge graphs demand more sophisticated reasoning because predicting links often requires multi-hop logical inference across different relation types. For example, if the facts (A, "father of", B) and

(B, “father of”, C) exist in the knowledge graph, the system should be able to infer the new relationship (A, “grandfather of”, C).

**Existing link prediction methods for KGs.** To address these challenges, a variety of link prediction methods have been developed specifically for knowledge graphs. These methods can be grouped into four primary categories: (1) Embedding-based methods, which map entities and relations to vector spaces for scoring potential links; (2) Path-based and rule-based methods, which use graph traversal, logical rules, or meta-paths rather than (or in addition to) embeddings. (3) Graph deep learning-based methods, which aggregate neighborhood information to capture structural dynamics; (4) LLM-based methods.

## 4.1 Embedding-Based Methods

Embedding-based methods form the foundational category, embedding entities and relations into low-dimensional spaces to predict links via distance or similarity functions, effectively handling heterogeneity by modeling diverse relational patterns.

(1) **Additive function over embedding**: These methods are also known as translational approaches. The TransE model (Bordes et al., 2013) was the first to introduce the translational-distance framework for knowledge graph embeddings. In this approach, each relation is modeled as a translation vector in a continuous space, enforcing the principle that for a valid triple, the embedding of the head entity plus the relation vector should be close to the embedding of the tail entity, i.e., $\mathbf{h}+\mathbf{r} \approx \mathbf{t}$, where $\mathbf{h}$, $\mathbf{r}$, and $\mathbf{t}$ are the embeddings of the head entity, relation, and tail entity, respectively. This works well for asymmetric or directional relations (e.g., “capital of”), but breaks down for symmetric ones (e.g., “married to” or “similar to”), where the relation should hold bidirectionally, this would require $\mathbf{r} = 0$. In this case, entities involved in symmetric relations will have nearly identical embeddings,

losing their individual distinctiveness. TransE also struggles with one-to-many, many-to-one, and many-to-many relations, as it compels diverse entities linked by the same relation to converge in embedding space (e.g., multiple cities under "has city" must approximate the same translated position from a country), leading to oversimplified and imprecise modeling of relational cardinality.

To address these challenges, several extensions such as TransH (Wang et al., 2014), TransR (Lin et al., 2015), TransD (Ji et al., 2015) have been proposed. TransH (Wang et al., 2014) uses relation-specific hyperplanes to vary entity positions and normalize distances, preventing embedding collapse; TransR (Lin et al., 2015) employs projection matrices to map entities into separate relation spaces for better heterogeneity handling; and TransD (Ji et al., 2015) adds dynamic, entity-type-aware mappings. Later, RotatE (Sun et al., 2019b) model represented relations as rotations in the complex vector space, enabling it to model symmetry, anti-symmetry, inversion, and composition. HAKE (Zhang et al., 2020) further incorporated polar coordinates for hierarchies, using modulus for inter-level distinctions and phase for intra-level patterns.

(2) **Multiplicative function over embedding**: RESCAL (Nickel et al., 2011) computes scores as $\mathbf{h}^T\mathbf{R}\mathbf{t}$ , where $\mathbf{h} \in \mathbb{R}^d$ and $\mathbf{t} \in \mathbb{R}^d$ are embedding vectors of entities $h$ and $r$, $\mathbf{R} \in \mathbb{R}^{d\times d}$ is an embedding matrix of relation $r$. Higher scores indicate a more likely triple. To alleviate computational burden, DistMult (Yang et al., 2015) restricts $\mathbf{R}$ to a diagonal form and computes triple scores as $\mathbf{h}^T\text{diag}(\mathbf{r})\mathbf{t}$, where $\mathbf{r} \in \mathbb{R}^d$ is a vector. However, this symmetric scoring function cannot distinguish between $(h, r, t)$ and $(t, r, h)$, making it unsuitable for asymmetric relations. ComplEx (Trouillon et al., 2016) addresses this limitation by extending entities and relations to complex vector space, using the scoring function $\text{Re}(\mathbf{h}^T\text{diag}(\mathbf{r})\overline{\mathbf{t}})$ where $\overline{\mathbf{t}}$ denotes complex conjugate, enabling modeling of asym-

metric relations. SimplE (Kazemi and Poole, 2018) uses dual embeddings per entity (head and tail), computing scores as $\frac{1}{2}(\mathbf{h}_h^T \mathbf{r} \mathbf{t}_t + \mathbf{h}_t^T \mathbf{r}^{-1} \mathbf{t}_h)$, where $\mathbf{h}_h$ and $\mathbf{h}_t$ are head and tail embeddings of entity $h$, and $\mathbf{r}$ and $\mathbf{r}^{-1}$ are forward and inverse relation embeddings. This design naturally handles both symmetric and asymmetric relations by allowing entities to have different representations based on their position in triples. TuckER (Balažević et al., 2019) achieves maximum expressiveness by treating the KG as a 3-dimensional tensor and applying Tucker decomposition, which factorizes this tensor into a core tensor $\mathcal{W}$ and three factor matrices: $\mathcal{W} \times_1 \mathbf{h} \times_2 \mathbf{r} \times_3 \mathbf{t}$ computes tensor products along each mode (entities and relations), where $\times_i$ denotes tensor contraction along the $i$-th dimension. This approach provides theoretical completeness by being able to represent any possible interaction pattern between entities and relations, with most existing linear models emerging as special cases of this general framework.

(c) **Neural network function over embedding:** In addition to additive and multiplicative functions, other neural network-based methods learns how the head, relation, and tail embeddings interact. Early methods (Socher et al., 2013) employ a straightforward multi-layer perceptron architecture that concatenates entity and relation embeddings as input: $\text{MLP}([\mathbf{h}; \mathbf{r}; \mathbf{t}])$, where $[;]$ denotes concatenation. ConvE (Dettmers et al., 2018) uses 2D convolutions over embeddings to predict missing links in knowledge graphs. The key is to reshape and stack head entity and relation embeddings into a 2D input matrix, apply 2D convolution to capture local interaction patterns, then compute scores with tail entities. ConvE's main limitation is that its reshaping strategy segregates head entity and relation features into separate regions of the 2D input, preventing many feature combinations from interacting during convolution. InteractE (Vashishth et al., 2020b) addresses this by expanding, permuting, and reshuffling embeddings before convolution, ensuring that each entity feature can interact with multiple relation features.

## 4.2 Graph Deep Learning Methods

To leverage the semantically rich graph neighborhood, several approaches have adapted graph deep learning methods to multi-relational graphs for KG link prediction. Broadly, these methods can be grouped into three main families: GCN-based methods, GAT-based methods, and graph Transformer-based methods.

(1) **GCN-based methods** form the earliest and most fundamental category. The R-GCN proposed by Schlichtkrull et al. (2018) was the first to adapt GCNs for relational graphs by using relation-specific weight matrices in the aggregation step:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_r(i)} \frac{1}{c_{i,r}}\mathbf{W}_r^{(l)}\mathbf{h}_j^{(l)} + \mathbf{W}_0^{(l)}\mathbf{h}_i^{(l)}\right),$$

where $\mathbf{h}_i^{(l)}$ is the embedding of entity $i$ at layer $l$, $\mathcal{N}_r(i)$ are neighbors under relation $r$, $c_{i,r}$ is a normalization factor (e.g., degree-based), $\mathbf{W}_r^{(l)}$ is the relation-specific weight, $\mathbf{W}_0^{(l)}$ handles self-loops, and $\sigma$ is an activation (e.g., ReLU). Since R-GCN uses a separate weight matrix $\mathbf{W}_r^{(l)}$ for each relation type $r$, the number of parameters grows linearly with the number of relations. To address this limitation, CompGCN (Vashishth et al., 2020a) introduces an explicit composition function $\phi(\cdot,\cdot)$ that combines an entity embedding and a relation embedding before message passing. $\mathbf{h}_i^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_r(i)} \frac{1}{c_{i,r}}\mathbf{W}_r^{(l)}\,\phi(\mathbf{h}_j^{(l)}, \mathbf{r}^{(k)}) + \mathbf{W}_0^{(l)}\mathbf{h}_i^{(l)}\right)$, where the relation embeddings are updated separately via a linear transformation $\mathbf{r}^{(k+1)} = \mathbf{W}_{\text{rel}}^{(k)}\mathbf{r}^{(k)}$. This reduces parameters while enabling richer entity-relation interactions.

A parallel research direction integrates geometric principles from translational embedding models into GCN architectures. For example, instead of simply aggregating neighbor information, TransGCN (Cai and Wang, 2019) enforces that entity embeddings should satisfy translational relationships with their neighbors. SACN (Shang et al., 2019) further advances this category by integrating structure-aware encoding with translational decoding. It employs a weighted graph convolutional network (WGCN) encoder that aggregates knowledge graph node structure, attributes, and relation types using learnable weights to control neighbor information in local aggregation. The WGCN output serves as input to the Conv-TransE decoder, similar to ConvE but preserving translational properties between entities and relations, thereby enhancing relational geometry modeling and prediction accuracy through convolutional patterns.

(2) **GAT-based models**: While GCN-based methods aggregate neighbor information with uniform weights, GAT-based models learn to assign different importance to different neighbors, depending on both relation type and context. KBGAT (Nathani et al., 2019) was the first to adapt graph attention to the knowledge graph setting. It incorporates both entity embeddings and relation embeddings into the attention mechanism, so that different neighbors under different relations are weighted dynamically. The attention coefficient from neighbor $j$ to target node $i$ under relation $r$ is computed as:

$$\alpha_{ij}^{(r)} = \text{softmax}_{j,r}(\text{LeakyReLU}(\mathbf{a}^T[\mathbf{W}_e\mathbf{h}_i \| \ \mathbf{W}_r\mathbf{h}_r \| \ \mathbf{W}_e\mathbf{h}_j]))$$

where the softmax is computed over all neighbor-relation pairs $(j, r)$ such that $j \in \mathcal{N}_r(i)$ (neighbors of $i$ connected via relation $r$), where $\mathbf{h}_i, \mathbf{h}_j$ are entity embeddings, $\mathbf{r}$ is the relation embedding, $\mathbf{W}_e, \mathbf{W}_r$ are transformation matrices, and $\|$ denotes concatenation. The normalized coefficient $\alpha_{ij}^{(r)}$ reflects how much attention entity $i$ should pay to neighbor $j$ under relation $r$. The updated embedding of node $i$ at $(t+1)$th iteration is: $\mathbf{h}_i^{(t+1)} = \sigma\left(\sum_{j\in N_i}\sum_{r\in R_{ij}} \alpha_{ij}^{(r)}\mathbf{W}_1\left[\mathbf{h}_i^{(t)} \parallel \mathbf{h}_j^{(t)} \parallel \mathbf{g}_r^{(t)}\right]\right)$, where $N_i$ is the set of neighbors of node $i$, $R_{ij}$ is the set of relations connecting $i$ and $j$, $\mathbf{W}_1$ is a linear transformation matrix applied to the concatenated vector.

RGAT (Busbridge et al., 2019) refine this approach by incorporating relation-specific attention parameters rather than a shared attention mechanism. In RGAT, the attention coefficient is: $\alpha_{ij}^{r} = \text{softmax}_j(\mathbf{a}_r^T[\mathbf{W}_r\mathbf{h}_i|\mathbf{W}_r\mathbf{h}_j])$ where $\mathbf{a}_r$ and $\mathbf{W}_r$ are relation-specific parameters learned for each relation type $r$. MRGAT (Li et al., 2021b) further introduced multi-relational attention by using separate attention heads for different relation types. Multiple heads act like an ensemble, averaging out noise or suboptimal solutions across heads, leading to more consistent gradients and faster convergence. DisenKGAT (Wu et al., 2021) improves upon standard GAT-based approaches by explicitly factorizing relation embeddings into $K$ independent latent semantic components, each capturing distinct semantics (e.g., one for family feature, another for career). Beyond these, several other models (Li et al., 2022b; Wei et al., 2024) extend the GAT paradigm to more challenging heterogeneous settings.

(3) **Graph Transformer-based methods**. Unlike GCNs or GATs, which are inherently local and limited by the number of layers, Transformer-based architectures capture long-range dependencies more effectively. In these models, entity and relation embeddings are updated by full-graph attention rather than local neighborhood aggregation. Hu et al. (2020) introduced the HGT, which handles node and edge heterogeneity simultaneously by using type-specific projections in self-attention. For a target node $i$ of type $\tau_i$ and a neighbor node $j$ of type $\tau_j$, queries, keys, and values are projected as $\mathbf{q}_i = \mathbf{h}_i W_Q^{\tau_i}, \quad \mathbf{k}_j = \mathbf{h}_j W_K^{\tau_j}, \quad \mathbf{v}_j = \mathbf{h}_j W_V^{\tau_j}$. The attention coefficient between $i$ and $j$ is computed as $\alpha_{ij} = \frac{\exp(\mathbf{q}_i^\top \mathbf{k}_j/\sqrt{d})}{\sum_{k\in\mathcal{N}(i)}\exp(\mathbf{q}_i^\top \mathbf{k}_k/\sqrt{d})}$. Relphormer (Bi et al., 2023) extends this by combining subgraph sampling with structure-enhanced self-attention, where sampled subgraphs are encoded via $\alpha_{ij} = \frac{\exp(\mathbf{q}_i^\top \mathbf{k}_j/\sqrt{d}+\mathbf{s}_{ij})}{\sum_k \exp(\mathbf{q}_i^\top \mathbf{k}_k/\sqrt{d}+\mathbf{s}_{ik})}$, with $\mathbf{s}_{ij}$ as structural biases (e.g., shortest path distance). LHGNN (Nguyen et al., 2023) further generalizes the framework to graphs without explicit type labels by inferring latent node and relation types through clustering in

embedding space, using soft assignments in attention to adapt to implicit heterogeneity.

## 4.3 LLM-based methods

One line of work adapts standard Transformer encoders to knowledge graphs by treating triples $(h, r, t)$ as input tokens and applying self-attention across sequences of triples. KG-BERT (Yao et al., 2019a) is the first model to apply BERT (Devlin et al., 2019) to KG embedding. In KG-BERT, entities and relations are converted to natural language descriptions, forming sentences like "[CLS] head entity [SEP] relation [SEP] tail entity [SEP]". This sequence is fed into BERT, which is first pretrained with masked language modeling and next sentence prediction on triples, and then fine-tuned for link prediction. While effective, KG-BERT treats each triple in isolation. It cannot directly leverage multi-hop structure (paths or subgraphs). To address this limitation, StAR (Wang et al., 2021) reformulate relation paths as sequences. For example, to infer the nationality of Barack Obama, it encodes the path Obama → Honolulu (born_in) → United States (located_in) so the Transformer can reason over multi-hop connections. HittER (Chen et al., 2021) extends this idea by hierarchically encoding entire subgraphs. LMKE (Wang et al., 2022b) further leverages pre-trained language models to embed entities and relations directly from their textual descriptions. For example, it learns that "Marie Curie" relates to "Physics" and "Chemistry" from her Wikipedia description, enabling better predictions even for rare entities.

Another line of research adapt LLMs like GPT, T5, and LLaMA to KGs through prompt-based approaches. The key idea is to textualize nodes, relations, and subgraphs so that powerful pre-trained LLMs can be applied to link prediction tasks. LPNL (Bi et al., 2024a) introduces a two-stage heterogeneous subgraph sampling strategy to reduce the size of graph neighborhoods before converting them into prompts. In the first stage, it creates $h$-hop ego-subgraphs around a source node using normalized degree-based sampling. The second stage ranks nodes in the subgraph via personalized PageRank score, and selects the top-k nodes with the highest PPR scores as the anchor nodes. These anchor nodes enrich the textual descriptions of the source node. Specifically, for a node $v$, its description $d(v)$ integrates the anchors as a simple phrase like "$v$: [description] is related with [list of anchors and their descriptions]," implicitly embedding graph topology (e.g., key connections and centrality) into readable text without needing to explicitly describe edges or the full graph structure. Finally, these enriched descriptions, are assembled into prompts using the template $T(s, R, C) = q(R) + d(s) + \sum d(c_i)$, where $q(R)$ is a relation-specific query (e.g., "Which candidate is linked to the source?"), $d(s)$ describes the source, and $\sum d(c_i)$ concatenates candidate descriptions. This setup allows LLMs to perform link prediction while managing token limits effectively.

However, LPNL is limited in explicit multi-hop reasoning, i.e., it samples only 2-3 hops for context, without structured mechanisms like CoT to infer long-range dependencies or step-by-step path traversal. KG-LLaMA (Shu et al., 2024) directly addresses this by specializing in multi-hop tasks (up to 5 hops), using path extraction and CoT prompts to enable explicit, reasoned inference over extended connections, which LPNL's local sampling approach does not handle as effectively. Example Prompt (with CoT) is "Below is the detail of a knowledge graph path. Is Node 1 connected with Node 3? Answer the question by reasoning step-by-step. Choose from: 1. Yes 2. No. Input: Node1 has relation1 with Node 2, and Node 2 has relation 2 with Node 3." Following work include KICGPT (Wei et al., 2023b), and KoPA (Zhang et al., 2024b).

## 4.4 Path- and Rule-Based Methods

**Path-based methods** infer missing links in a KG by examining multi-hop paths connecting a head entity and a tail entity. Unlike models that rely solely on embeddings, path-based approaches can explicitly explain the reasoning behind a prediction. The first major path-based method was the Path Ranking Algorithm (PRA) (Lao and Cohen, 2010). The algorithm has three main phases: path enumeration, feature computation via random walks, and supervised ranking for prediction. (a) In the path enumeration phase, PRA enumerates a set of edge-labeled path types $P = R_1, R_2, \dots, R_\ell$, where each $R_i$ is a relation type, and $\ell$ is a small bound (e.g., 3-4 hops to avoid explosion). (b) For each path type $P$, PRA computes a probability $\mathbb{P}(h, t|P)$, representing the likelihood of reaching tail $t$ from head $h$ via a random walk $P$. (c) Finally, a logistic regression classifier is trained on a feature matrix $\mathbf{X} \in \mathbb{R}^{n\times p}$ constructed from the enumerated path types. Each row $\mathbf{X}_i$ corresponds to a training entity pair $(h_i, t_i)$ sampled from the dataset, comprising both positive instances (observed triples with binary label $y_i = 1$) and negative instances (unobserved entity pairs with label $y_i = 0$). Each column $j$ represents a distinct path type $P_j$, where the feature value $\mathbf{X}_{ij} = \mathbb{P}(h_i, t_i|P_j)$. The logistic regression model learns parameter $\hat{\beta} \in \mathbb{R}^p$.

Following path-based methods can be grouped into two main subcategories: (1) **Random walk–based**: A number of works (Gardner et al., 2013; Wang et al., 2013; Gardner et al., 2014; Gardner and Mitchell, 2015) have proposed techniques to improve the scalability and accuracy of PRA. (2) **Reinforcement learning (RL)–based**: Unlike PRA's exhaustive enumeration of fixed path templates, RL-based methods train an agent to learn how to walk paths in a knowledge graph that are predictive of missing links. DeepPath (Xiong et al., 2017) formulates link prediction as a Markov Decision Process where an agent starts at a source entity and learns to select relations step-by-step until reaching a target entity. The state includes the current entity and known tail, while actions correspond to outgoing relations. The agent receives positive rewards for successfully reaching the correct target. However, DeepPath assumes the target entity is known during training, limiting its applicability to realistic completion tasks where the tail entity must be discovered. MINERVA (Das et al., 2018) addresses this limitation by reformulating the task as query answering. Given only a head entity and query relation, the agent must discover the correct tail by walking through the graph. The state includes the current entity and query relation, with rewards provided only upon reaching correct tail entities. This sparse-reward formulation is more challenging to train but enables practical link prediction where target entities are unknown. M-Walk (Shen et al., 2018) and SSRL (Ma et al., 2022) further enhance exploration and training efficiency.

**Rule-based methods**: (1) **Classical rule mining**. Classical rule-based methods infer missing facts by explicitly searching for Horn clause rules in the knowledge graph. A Horn clause is a logical implication, e.g., if a person is born in a city and that city is located in a country, then the person has that nationality. The first widely adopted method for mining such rules is AMIE (Galárraga et al., 2013), with its improved version AMIE+ (Galárraga et al., 2015). AMIE enumerates candidate rules, such as relation 1 (e.g., born in) and relation 2 (e.g., located in) imply relation 3 (e.g., nationality). Each rule is then evaluated using two measures: (a) Support which count the absolute frequency of observations where both condition (relation 1 and relation 2) and the outcome (relation 3) occur together (b) Confidence is the conditional probability that the outcome holds given the condition, calculated as the ratio of support to the frequency of the condition alone, measuring rule reliability like precision in classification. AnyBURL (Meilicke et al., 2019) improved efficiency by sampling-based rule exploration, producing competitive rules even on large graphs. Advanced variants like RuDiK (Ortona et al., 2018) extends the expressive power of rule mining systems by moving beyond Horn clauses. Recent extensions, such as eXpath (Sun et al., 2024), integrate ontological closed path rules to not only mine but also explain link predictions, (2) Neural network rule learning. Recent work use differentiable neural network to construct rules in an end-to-end differentiable manner, learning optimal structures directly from data, such as DRUM (Sadeghian et al., 2019) and RNNLogic (Qu et al., 2021).

## 4.5 Link Prediction in Statistics

In traditional statistical network analysis, link prediction methods typically regard edges as random variables, with link probabilities determined by latent variables. For example, graphon models (Zhang et al., 2017; Chandna et al., 2022) assume that edges are generated according to a latent position model, with $P(A_{ij} = 1) = f(u_i, u_j)$ for latent variables $u_i, u_j \in [0, 1]$. Despite many successful applications of these methods, they cannot be directly applied to KGs due to the presence of node features, heterogeneous node types, and diverse edge types. This raises a key question: how can current statistical random graph models be extended to KGs? In what follows, we envision a possible model which generalize graphons to heterogeneous KGs, incorporating node types and node features.

**Problem setup.** Let the KG be denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$, where $\mathcal{V} = 1, \ldots, n$ is the set of nodes, $\mathcal{E} \subseteq \binom{\mathcal{V}}{2} \times \mathcal{R}$ is the set of labeled edges (with at most one label per unordered pair $i, j$), and $\mathcal{R} = 1, \ldots, R$ is the set of relation types. Each node $i \in \mathcal{V}$ has an observed type $k_i \in 1, \ldots, K$ (e.g., "drug", "disease") and covariate $x_i \in \mathbb{R}^p$ (e.g., text embedding). Define the edge label $E_{ij} \in 0 \cup \mathcal{R}$, where $E_{ij} = r > 0$ if $(i, j, r) \in \mathcal{E}$ and $E_{ij} = 0$ if no edge exists between $i$ and $j$. The node type node $i$ is denoted as $k_i$.

**Potential model assumptions applied to KGs.** (1) Latent positions: Each node $i$ has a node-specific latent position $\xi_i \in [0,1]^d$, drawn independently as $\xi_i \overset{\text{i.i.d.}}{\sim} \text{Unif}([0,1]^d)$, where $d \geq 1$ is the latent dimension. (2) Covariate generation: Observed covariates are noisy realization of the effective latents: $x_i = \Gamma_{k_i}\xi_i + \epsilon_i$, where $\Gamma_{k_i} \in \mathbb{R}^{p \times d}$ is a type-specific projection matrix (assumed full rank for invertibility), and $\epsilon_i \overset{\text{i.i.d.}}{\sim} \text{subG}(\sigma^2 I_p)$ (sub-Gaussian with variance proxy $\sigma^2$). $\Gamma_{k_i}$ permits dimensionality reduction for high-dimensional embeddings. (3) Link functions: For each pair $i, j$ with types $(k_i, k_j)$ and each relation $r$,

$$\mathbb{P}(E_{ij} = r | \xi_i, \xi_j, k_i, k_j) = \frac{\exp(f_r(\xi_i, \xi_j, k_i, k_j))}{\sum_{r'=0}^{R} \exp(f_{r'}(\xi_i, \xi_j, k_i, k_j))},$$

where $E_{ij} \in \{0, 1, ..., R\}$ with 0 indicating no edge, and $f_r$ is a relation-specific link function. To ensure theoretical tractability, additional smoothness and identifiability conditions might be required.

# 5 Discussions

Beyond community detection and link prediction, statistical network analysis offers additional avenues to enhance LLM applications. In this section, we explore emerging opportunities, including graph alignment for hallucination detection, uncertainty quantification (UQ) for LLM answers using pairwise graph distance, and extensions to more complex graph structures such as multi-layer graphs, hypergraphs, and time-varying graphs.

**Graph alignment for fact verification.** Graph alignment techniques represent a promising direction for verifying LLM outputs against KGs, particularly in detecting hallucinations, unsupported or false statements generated by the model. For instance, FactAlign (Rashad et al., 2024) constructs a KG from a given source and generated text, and uses word embeddings to align individual claim triples with source triples to identify factual misalignments, categorizing them as hallucinations if the similarity is low.

The graph alignment problem has been extensively studied in statistics and ML. From a statistical perspective, graph alignment can be framed as a problem of estimating an unknown correspondence (permutation or partial matching) between two sets of nodes that preserves relational patterns under noise. Classical methods include the quadratic assignment problem (Lawler, 1963; Loiola et al., 2007) and its relaxations (spectral (Feizi et al., 2019) and Frank–Wolfe (Frank et al., 1956) approaches), optimal transport–based methods such as Gromov–Wasserstein (GW) (Xu et al., 2019) and fused GW (Vayer et al., 2020) that align structural and feature information.

However, applications to fact verification remain limited, creating opportunities for statisticians to adapt these methods to this new setting. Unique challenges include: (1) Uncertainty quantification. Most alignment methods produce a single point estimate of the mapping. A statistical treatment could produce confidence sets for alignments or probabilistic match scores, enabling more principled decision thresholds for hallucination classification. (2) Partial and noisy correspondence. Unlike classical network alignment, we expect many unmatched nodes and edges (since LLM outputs may contain novel entities). Methods that allow partial matching, unbalanced optimal transport, or sparsity-regularized mappings can better capture this setting. (3) Adapting alignment for heterogeneous structures: Classical alignment methods are often designed for simple, homogeneous graphs. Their direct application to KGs is limited because they ignore the rich, typed information in nodes and edges. The opportunity here is to develop new statistical alignment models tailored to this complexity.

**Uncertainty quantification for LLM outputs.** Another promising direction is to use KGs to quantify uncertainty in LLM outputs. A recent example is KG-UQ (Yuan et al., 2025), which treats the dispersion of responses in a structured space as a proxy for model uncertainty. The procedure is straightforward: (1) sample multiple responses to a given prompt; (2) construct a KG from each response; and (3) compute a pairwise distance matrix across KGs. A response's "confidence" is defined as its average distance to other responses, and the prompt-level uncertainty is the mean confidence across all responses. Intuitively, when the induced KGs are mutually dissimilar, the model is less certain. This framework opens rich statistical opportunities, as the core step, measuring distances between graphs, has been extensively studied across statistics, ML, and network science. The literature include structural distance measures (graph edit distances (Riesen and Bunke, 2009; Sanfeliu and Fu, 2012), spectral methods based on Laplacian eigenvalues (Wilson and Zhu, 2008)), distribution-based approaches (graph kernels (Borgwardt and Kriegel, 2005; Vishwanathan et al., 2010), optimal transport methods like Gromov-Wasserstein distances (Xu et al., 2019) and fused GW (Vayer et al., 2020)), and information-theoretic measures (graph entropy (Dehmer, 2008), mutual information (Anand et al., 2011)).

The application of graph distances to UQ of LLM outputs presents promising statistical opportunities. First, one can try different ways of measuring the distance between graphs and see which ones give uncertainty scores that best match reality. It is also possible to combine several distance measures—similar to ensemble learning—to get a more reliable and stable estimate of uncertainty. Second, researchers can design new alignment methods tailored to knowledge graphs. Unlike simple graphs, KGs have nodes with attributes, typed edges, and often incomplete overlap. Developing methods that handle heterogeneity, partial matching, and noisy triples can lead to more accurate and semantically meaningful uncertainty estimates. Finally, because KGs for long text can be large and computationally expensive to compare, there is a need for scalable solutions such as sampling schemes, sketching, or low-rank approximations that make sophisticated distances like optimal transport or graph edit distance feasible at scale.

**Extensions to complex graph structures.** The majority of statistical network models and KG applications assume a static, simple graph structure with nodes and pairwise edges. However, real-world knowledge often involves more complex relationships that are better represented by multi-layer graphs, hypergraphs, or time-varying graphs.

(1) **Multi-layer graph** is a network model with multiple interconnected layers, where each layer represents a different type of relationship, social circle, or functional mode among the same or different sets of nodes. The term refers to two main types of networks: multiplex networks, where the same set of nodes appear in every layer with different interactions, and networks of networks, where nodes can exist in different layers but are not necessarily the same set (Kivelä et al., 2014). Recent applications have demonstrated the potential of multi-layer knowledge graphs in enhancing LLM capabilities, such as TravelRAG (Song et al., 2024) and OrthographRAG (Tata et al., 2025), which employ multiple layers to represent orthographic and semantic relationships. Meanwhile, there is also a growing literature on statistical and ML methodology for multi-layer networks, including community detection across layers (Wilson et al., 2017; Lei et al., 2020; Jing et al., 2021; Zhang et al., 2024a), and link prediction methods that leverage cross-layer information (Jafari et al., 2021; Yang et al., 2025). Applying and extending these approaches to multi-layer knowledge graphs is a promising direction.

(2) **Hypergraph** is a generalization of a standard graph where edges (called hyperedges) can connect any number of nodes, not just pairs, enabling the representation of complex multi-way relationships that cannot be adequately captured by pairwise connections. In KG contexts, hypergraphs are essential because many facts are inherently n-ary relationships, e.g., “drug A treats disease B via mechanism C” or “gene X is associated with

phenotype Y in population Z under condition W". Recognizing this potential, recent years have seen increasing interest in knowledge hypergraphs (Fatemi et al., 2023; Dou et al., 2025) and their integration with LLMs. Notable examples include HyperGraphRAG (Luo et al., 2025) and HyperG (Huang et al., 2025), which leverage hypergraph structures to enhance knowledge representation and reasoning capabilities. To fully realize the potential of these applications, statistical methods for hypergraphs, such as community detection (Ke et al., 2019; Zhen and Wang, 2023; Xu et al., 2023), and generative models such as stochastic block hypergraph models (Pister and Barthelemy, 2024), latent space models (Turnbull et al., 2024), and hypergraphons (Balasubramanian, 2021), which can be used to quantify the probability of hyperedge formation and hyperedge prediction. In particular, community detection holds great promise for identifying topical clusters within knowledge hypergraphs. Applying and extending these methods to knowledge hypergraphs remains largely unexplored and represents a promising direction.

(3) **Time-varying KGs**, a.k.a temporal KGs, are essential for capturing the dynamic nature of real-world knowledge, where facts evolve over time, relationships change, and historical context significantly impacts reasoning and inference (Cai et al., 2022, 2024). Recent applications have begun integrating temporal KGs with LLMs for enhanced reasoning capabilities (Saxena et al., 2021; Dhingra et al., 2022). From a statistical perspective, extensive methodological frameworks exist for analyzing time-varying networks, including community detection (Zhang and Cao, 2017; Matias and Miele, 2017), and temporal link prediction algorithms (Kim et al., 2018). Integrating these dynamic statistical models is a critical next step toward building more sophisticated AI systems that can understand and reason about the evolution of knowledge over time.

# References

Abbe, E. (2018). Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18(177):1–86.

Abu-Rasheed, H., Weber, C., and Fathi, M. (2024). Knowledge graphs as context sources for llm-based explanations of learning recommendations. In *2024 IEEE Global Engineering Education Conference (EDUCON)*, pages 1–5. IEEE.

Agichtein, E. and Gravano, L. (2000). Snowball: Extracting relations from large plain-text collections. In *Proceedings of the fifth ACM conference on Digital libraries*, pages 85–94.

Agrawal, G., Kumarage, T., Alghamdi, Z., and Liu, H. (2023). Can knowledge graphs reduce hallucinations in llms?: A survey. *arXiv preprint arXiv:2311.07914*.

Ahn, J. J. and Yin, W. (2025). Prompt-reverse inconsistency: Llm self-inconsistency beyond generative randomness and prompt paraphrasing. *arXiv preprint arXiv:2504.01282*.

Airoldi, E. M., Blei, D., Fienberg, S., and Xing, E. (2008). Mixed membership stochastic blockmodels. *Advances in neural information processing systems*, 21.

Al Khatib, H. S., Mittal, S., Rahimi, S., Marhamati, N., and Bozorgzad, S. (2025). From patient consultations to graphs: Leveraging llms for patient journey knowledge graph construction. In *2025 IEEE Conference on Artificial Intelligence (CAI)*, pages 410–415. IEEE.

Alsentzer, E., Murphy, J. R., Boag, W., Weng, W.-H., Jin, D., Naumann, T., and McDermott, M. (2019). Publicly available clinical bert embeddings. *arXiv preprint arXiv:1904.03323*.

Anand, K., Bianconi, G., and Severini, S. (2011). Shannon and von neumann entropy of random networks with heterogeneous expected degree. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 83(3):036109.

Aronson, A. R. (2001). Effective mapping of biomedical text to the umls metathesaurus: the metamap program. In *Proceedings of the AMIA Symposium*, page 17.

Ashok, D. and Lipton, Z. C. (2023). Promptner: Prompting for named entity recognition. *arXiv preprint arXiv:2305.15444*.

Balasubramanian, K. (2021). Nonparametric modeling of higher-order interactions via hypergraphons. *Journal of Machine Learning Research*, 22(146):1–35.

Balažević, I., Allen, C., and Hospedales, T. (2019). Tucker: Tensor factorization for knowledge graph completion. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5185–5194.

Banerjee, S., Agarwal, A., and Singla, S. (2025). Llms will always hallucinate, and we need to live with this. In *Intelligent Systems Conference*, pages 624–648. Springer.

Bastos, A., Nadgeri, A., Singh, K., Mulang, I. O., Shekarpour, S., Hoffart, J., and Kaul, M. (2021). Recon: relation extraction using knowledge graph context in a graph neural network. In *Proceedings of the Web Conference 2021*, pages 1673–1685.

Bi, B., Liu, S., Wang, Y., Mei, L., and Cheng, X. (2024a). Lpnl: Scalable link prediction with large language models. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 3615–3625. Association for Computational Linguistics.

Bi, Z., Chen, J., Jiang, Y., Xiong, F., Guo, W., Chen, H., and Zhang, N. (2024b). Codekgc: Code language model for generative knowledge graph construction. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 23(3):1–16.

Bi, Z., Cheng, S., Chen, J., Liang, X., Xiong, F., and Zhang, N. (2023). Relphormer: Relational graph transformer for knowledge graph representations. *Neurocomputing*.

Bickel, P. J. and Chen, A. (2009). A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.

Borgwardt, K. M. and Kriegel, H.-P. (2005). Shortest-path kernels on graphs. In *Fifth IEEE international conference on data mining (ICDM'05)*, pages 8–pp. IEEE.

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., et al. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, 1(1):58.

Brin, S. (1998). Extracting patterns and relations from the world wide web. In *International workshop on the world wide web and databases*, pages 172–183. Springer.

Busbridge, D., Sherburn, D., Cavallo, P., and Hammerla, N. Y. (2019). Relational graph attention networks. In *arXiv preprint arXiv:1904.05811*.

Cabot, P.-L. H. and Navigli, R. (2021). Rebel: Relation extraction by end-to-end language generation. In *Findings of the association for computational linguistics: emnlp 2021*, pages 2370–2381.

Cai, B., Xiang, Y., Gao, L., Zhang, H., Li, Y., and Li, J. (2022). Temporal knowledge graph completion: A survey. *arXiv preprint arXiv:2201.08236*.

Cai, L., Mao, X., Zhou, Y., Long, Z., Wu, C., and Lan, M. (2024). A survey on temporal knowledge graph: Representation learning and applications. *arXiv preprint arXiv:2403.04782*.

Cai, L. and Wang, W. Y. (2019). Transgcn: Coupling transformation assumptions with graph convolutional networks for link prediction. In *Proceedings of the 10th International Joint Conference on Knowledge Graphs*, pages 131–138.

Carta, S., Giuliani, A., Piano, L., Podda, A. S., Pompianu, L., and Tiddia, S. G. (2023). Iterative zero-shot llm prompting for knowledge graph construction. *arXiv preprint arXiv:2307.01128*.

Chakraborty, A. (2024). Multi-hop question answering over knowledge graphs using large language models. *arXiv preprint arXiv:2404.19234*.

Chandna, S., Olhede, S. C., and Wolfe, P. J. (2022). Local linear graphon estimation using covariates. *Biometrika*, 109(3):721–734.

Chang, A. X. and Manning, C. D. (2014). Tokensregex: Defining cascaded regular expressions over tokens.

Chang, R.-C. and Zhang, J. (2024). Communitykg-rag: Leveraging community structures in knowledge graphs for advanced retrieval-augmented generation in fact-checking. *arXiv preprint arXiv:2408.08535*.

Chen, K. and Lei, J. (2018). Network cross-validation for determining the number of communities in network data. *Journal of the American Statistical Association*, 113(521):241–251.

Chen, Q., Yao, H., Li, S., Li, X., Kang, X., Lai, W., and Kuang, J. (2023). Fact-condition statements and super relation extraction for geothermic knowledge graphs construction. *Geoscience Frontiers*, 14(5):101412.

Chen, R., Jiang, W., Qin, C., Rawal, I. S., Tan, C., Choi, D., Xiong, B., and Ai, B. (2024). Llm-based multi-hop question answering with knowledge graph integration in evolving environments. *arXiv preprint arXiv:2408.15903*.

Chen, S., Liu, X., Gao, J., Jiao, J., Zhang, R., and Ji, Y. (2021). Hitter: Hierarchical transformers for knowledge graph embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10395–10407. Association for Computational Linguistics.

Chiu, J. P. and Nichols, E. (2016). Named entity recognition with bidirectional lstm-cnns. *Transactions of the association for computational linguistics*, 4:357–370.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Cui, W., Zhang, J., Li, Z., Sun, H., Lopez, D., Das, K., Malin, B. A., and Kumar, S.

(2025). Automatic prompt optimization via heuristic search: A survey. *arXiv preprint arXiv:2502.18746*.

Cunningham, H., Mayard, D., and Tablan, V. (2000). Jape: a java annotation patterns engine second edition ed. *Sheffield: University of Sheffield*.

Das, R., Dhuliawala, S., Zaheer, M., Vilnis, L., Durugkar, I., Krishnamurthy, A., Smola, A., and McCallum, A. (2018). Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning. In *International Conference on Learning Representations*.

Dehmer, M. (2008). Information processing in complex networks: Graph entropy and information functionals. *Applied Mathematics and Computation*, 201(1-2):82–94.

Deng, J., Yang, X., Yu, J., Liu, J., Shen, Z., Huang, D., and Cheng, H. (2024). Network tight community detection. In *Forty-first International Conference on Machine Learning*.

Dettmers, T., Minervini, P., Stenetorp, P., and Riedel, S. (2018). Convolutional 2d knowledge graph embeddings. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*. AAAI Press.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186.

Dhingra, B., Cole, J. R., Eisenschlos, J. M., Gillick, D., Eisenstein, J., and Cohen, W. W. (2022). Time-aware language models as temporal knowledge bases. *Transactions of the Association for Computational Linguistics*, 10:257–273.

Dou, C., Zhang, Y., Jin, Z., Jiao, W., Zhao, H., Zhao, Y., and Tao, Z. (2025). Enhancing llm generation with knowledge hypergraph for evidence-based medicine. *arXiv preprint arXiv:2503.16530*.

Eberts, M. and Ulges, A. (2019). Span-based joint entity and relation extraction with transformer pre-training. *arXiv preprint arXiv:1909.07755*.

Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., Truitt, S., Metropolitansky, D., Ness, R. O., and Larson, J. (2024). From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*.

Fatemi, B., Taslakian, P., Vazquez, D., and Poole, D. (2023). Knowledge hypergraph embedding meets relational algebra. *Journal of Machine Learning Research*, 24(105):1–34.

Fei, H., Zhang, Y., Ren, Y., and Ji, D. (2021). A span-graph neural model for overlapping entity relation extraction in biomedical texts. *Bioinformatics*, 37(11):1581–1589.

Feizi, S., Quon, G., Recamonde-Mendoza, M., Medard, M., Kellis, M., and Jadbabaie, A. (2019). Spectral alignment of graphs. *IEEE Transactions on Network Science and Engineering*, 7(3):1182–1197.

Frank, M., Wolfe, P., et al. (1956). An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110.

Friedman, C., Hripcsak, G., DuMouchel, W., Johnson, S. B., and Clayton, P. D. (1995). Natural language processing in an operational clinical information system. *Natural Language Engineering*, 1(1):83–108.

Fu, T.-J., Li, P.-H., and Ma, W.-Y. (2019). GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1409–1418, Florence, Italy. Association for Computational Linguistics.

Galárraga, L. A., Teflioudi, C., Hose, K., and Suchanek, F. M. (2013). Amie: Association rule mining under incomplete evidence in ontological knowledge bases. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 413–422. ACM.

Galárraga, L. A., Teflioudi, C., Hose, K., and Suchanek, F. M. (2015). Fast rule mining in ontological knowledge bases with amie+. *The VLDB Journal*, 24(6):707–730.

Gao, T., Han, X., Zhu, H., Liu, Z., Li, P., Sun, M., and Zhou, J. (2019). Fewrel 2.0: Towards more challenging few-shot relation classification. *arXiv preprint arXiv:1910.07124*.

Gardent, C., Shimorina, A., Narayan, S., and Perez-Beltrachini, L. (2017). Creating training corpora for nlg micro-planning. In *55th Annual Meeting of the Association for Computational Linguistics, ACL 2017*, pages 179–188. Association for Computational Linguistics (ACL).

Gardner, M. and Mitchell, T. M. (2015). Efficient and expressive knowledge base completion using subgraph feature extraction. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1488–1498. Association for Computational Linguistics.

Gardner, M., Talukdar, P. P., Kisiel, B., and Mitchell, T. (2013). Improving learning and inference in a large knowledge-base using latent syntactic cues. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 833–838. Association for Computational Linguistics.

Gardner, M., Talukdar, P. P., Krishnamurthy, J., and Mitchell, T. (2014). Incorporating vector space similarity in random walk inference over knowledge bases. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 397–406. Association for Computational Linguistics.

Ghanem, H. and Cruz, C. (2024a). Enhancing knowledge graph construction: Evaluating with emphasis on hallucination, omission, and graph similarity metrics. In *International Knowledge Graph and Semantic Web Conference*, pages 32–46. Springer.

Ghanem, H. and Cruz, C. (2024b). Fine-tuning vs. prompting: evaluating the knowledge graph construction with llms. In *3rd International Workshop on Knowledge Graph Generation from Text (Text2KG) Co-located with the Extended Semantic Web Conference (ESWC 2024)*, volume 3747, page 7.

Ghanem, H. and Cruz, C. (2025). Fine-tuning or prompting on llms: evaluating knowledge graph construction task. *Frontiers in Big Data*, 8:1505877.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., Airoldi, E. M., et al. (2010). A survey of statistical network models. *Foundations and Trends® in Machine Learning*, 2(2):129–233.

Gridach, M. (2017). Character-level neural network for biomedical named entity recognition. *Journal of biomedical informatics*, 70:85–91.

Guo, Z., Zhang, Y., and Lu, W. (2019). Attention guided graph convolutional networks for relation extraction. *arXiv preprint arXiv:1906.07510*.

Gupta, P., Schütze, H., and Andrassy, B. (2016). Table filling multi-task recurrent neural network for joint entity and relation extraction. In *Proceedings of COLING 2016, the*

*26th International Conference on Computational Linguistics: Technical Papers*, pages 2537–2547.

Gurulingappa, H., Rajput, A. M., Roberts, A., Fluck, J., Hofmann-Apitius, M., and Toldo, L. (2012). Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. *Journal of biomedical informatics*, 45(5):885–892.

Han, X., Yu, P., Liu, Z., Sun, M., and Li, P. (2018). Hierarchical relation extraction with coarse-to-fine grained attention. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 2236–2245.

Hendrickx, I., Kim, S. N., Kozareva, Z., Nakov, P., Séaghdha, D. O., Padó, S., Pennacchiotti, M., Romano, L., and Szpakowicz, S. (2019). Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. *arXiv preprint arXiv:1911.10422*.

Hendrycks, D., Burns, C., Chen, A., and Ball, S. (2021). Cuad: An expert-annotated nlp dataset for legal contract review. *arXiv preprint arXiv:2103.06268*.

Hennen, M., Babl, F., and Geierhos, M. (2024). Iter: Iterative transformer-based entity recognition and relation extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 11209–11223.

Herrero-Zazo, M., Segura-Bedmar, I., Martínez, P., and Declerck, T. (2013). The ddi corpus: An annotated corpus with pharmacological substances and drug–drug interactions. *Journal of biomedical informatics*, 46(5):914–920.

Hewlett, D., Lacoste, A., Jones, L., Polosukhin, I., Fandrianto, A., Han, J., Kelcey, M., and Berthelot, D. (2016). Wikireading: A novel large-scale language understanding task over wikipedia. *arXiv preprint arXiv:1608.03542*.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098.

Hu, J., Zhang, J., Qin, H., Yan, T., and Zhu, J. (2021). Using maximum entry-wise deviation to test the goodness of fit for stochastic block models. *Journal of the American Statistical Association*, 116(535):1373–1382.

Hu, Y. and Wang, W. (2024). Network-adjusted covariates for community detection. *Biometrika*, 111(4):1221–1240.

Hu, Z., Dong, Y., Wang, K., and Sun, Y. (2020). Heterogeneous graph transformer. In *Proceedings of The Web Conference (WWW)*, pages 2704–2710.

Huang, S., Li, H., Gu, Y., Hu, X., Li, Q., and Xu, G. (2025). Hyperg: Hypergraph-enhanced llms for structured knowledge. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1218–1228.

Huang, Z., Xu, W., and Yu, K. (2015). Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*.

Ibrahim, N., Aboulela, S., Ibrahim, A., and Kashef, R. (2024). A survey on augmenting knowledge graphs (kgs) with large language models (llms): models, evaluation metrics, benchmarks, and challenges. *Discover Artificial Intelligence*, 4(1):76.

Islam, K., Nipu, A. S., Wu, J., and Madiraju, P. (2025). Llm-based prompt ensemble for reliable medical entity recognition from ehrs. *arXiv preprint arXiv:2505.08704*.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bull Soc Vaudoise Sci Nat*, 37:547–579.

Jafari, S. H., Abdolhosseini-Qomi, A. M., Asadpour, M., Rahgozar, M., and Yazdani, N. (2021). An information theoretic approach to link prediction in multiplex networks. *Scientific Reports*, 11(1):13242.

Jat, S., Khandelwal, S., and Talukdar, P. (2018). Improving distantly supervised relation extraction using word and entity based attention. *arXiv preprint arXiv:1804.06987*.

Ji, G., He, S., Xu, L., Liu, K., and Zhao, J. (2015). Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)*, pages 687–696.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., and Fung, P. (2023). Survey of hallucination in natural language generation. *ACM computing surveys*, 55(12):1–38.

Jiang, H., Cui, L., Xu, Z., Yang, D., Chen, J., Li, C., Liu, J., Liang, J., Wang, C., Xiao, Y., et al. (2019). Relation extraction using supervision from topic knowledge of relation labels. In *IJCAI*, pages 5024–5030.

Jiang, P., Xiao, C., Jiang, M., Bhatia, P., Kass-Hout, T., Sun, J., and Han, J. (2025). Reasoning-enhanced healthcare predictions with knowledge graph community retrieval. In *The Thirteenth International Conference on Learning Representations*.

Jie, Z. and Lu, W. (2019). Dependency-guided lstm-crf for named entity recognition. *arXiv preprint arXiv:1909.10148*.

Jin, B., Liu, G., Han, C., Jiang, M., Ji, H., and Han, J. (2024a). Large language models on graphs: A comprehensive survey. *IEEE Transactions on Knowledge and Data Engineering*.

Jin, J., Ke, Z. T., and Luo, S. (2024b). Mixed membership estimation for social networks. *Journal of Econometrics*, 239(2):105369.

Jing, B.-Y., Li, T., Lyu, Z., and Xia, D. (2021). Community detection on mixture multilayer networks via regularized tensor decomposition. *The Annals of Statistics*, 49(6):3181–3205.

Karrer, B. and Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 83(1):016107.

Karwa, V., Pati, D., Petrović, S., Solus, L., Alexeev, N., Raič, M., Wilburne, D., Williams, R., and Yan, B. (2024). Monte carlo goodness-of-fit tests for degree corrected and related stochastic blockmodels. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 86(1):90–121.

Kau, A., He, X., Nambissan, A., Astudillo, A., Yin, H., and Aryani, A. (2024). Combining knowledge graphs and large language models. *arXiv preprint arXiv:2407.06564*.

Kazemi, S. M. and Poole, D. (2018). Simple embedding for link prediction in knowledge graphs. *Advances in neural information processing systems*, 31.

Ke, Z. T., Shi, F., and Xia, D. (2019). Community detection for hypergraph networks via regularized tensor power iteration. *arXiv preprint arXiv:1909.06503*.

Keloth, V. K., Hu, Y., Xie, Q., Peng, X., Wang, Y., Zheng, A., Selek, M., Raja, K., Wei, C. H., Jin, Q., et al. (2024). Advancing entity recognition in biomedicine via instruction tuning of large language models. *Bioinformatics*, 40(4):btae163.

Kim, B., Lee, K. H., Xue, L., and Niu, X. (2018). A review of dynamic network models with latent variables. *Statistics surveys*, 12:105.

Kim, J. (2025). Basic issues and challenges of statistical network analysis. *Communications for Statistical Applications and Methods*, 32(1):107–123.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *Journal of complex networks*, 2(3):203–271.

Krallinger, M., Rabal, O., Akhondi, S. A., Pérez, M. P., Santamaría, J., Rodríguez, G. P., Tsatsaronis, G., Intxaurrondo, A., López, J. A., Nandal, U., et al. (2017). Overview of the biocreative vi chemical-protein interaction track. In *Proceedings of the sixth BioCreative challenge evaluation workshop*, volume 1, pages 141–146.

Krügel, S., Ostermaier, A., and Uhl, M. (2023). Chatgpt's inconsistent moral advice influences users' judgment. *Scientific Reports*, 13(1):4569.

Lafferty, J., McCallum, A., and Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Lao, N. and Cohen, W. W. (2010). Relational retrieval using a combination of path-constrained random walks. *Machine Learning*, 81(1):53–67.

Latouche, P., Birmelé, E., and Ambroise, C. (2014). Model selection in overlapping stochastic block models. *Electronic Journal of Statistics*, 8:762–794.

Lavrinovics, E., Biswas, R., Bjerva, J., and Hose, K. (2025). Knowledge graphs, large language models, and hallucinations: An nlp perspective. *Journal of Web Semantics*, 85:100844.

Lawler, E. L. (1963). The quadratic assignment problem. *Management science*, 9(4):586–599.

Lee, J., Seo, S., and Choi, Y. S. (2019). Semantic relation classification via bidirectional lstm networks with entity-aware attention using latent entity typing. *Symmetry*, 11(6):785.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., and Kang, J. (2020). Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S., et al. (2015). Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic web*, 6(2):167–195.

Lei, J. (2016). A goodness-of-fit test for stochastic block models. *The Annals of Statistics*, pages 401–424.

Lei, J., Chen, K., and Lynch, B. (2020). Consistent community detection in multi-layer network data. *Biometrika*, 107(1):61–73.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., et al. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33:9459–9474.

Li, F., Zhang, M., Fu, G., and Ji, D. (2017). A neural joint model for entity and relation extraction from biomedical text. *BMC bioinformatics*, 18(1):198.

Li, H., Appleby, G., Alperin, K., Gomez, S. R., and Suh, A. (2025a). Mitigating llm hallucinations with knowledge graphs: A case study. *arXiv preprint arXiv:2504.12422*.

Li, J., Cheng, X., Zhao, W. X., Nie, J.-Y., and Wen, J.-R. (2023). Halueval: A large-scale hallucination evaluation benchmark for large language models. *arXiv preprint arXiv:2305.11747*.

Li, J., Wei, Q., Ghiasvand, O., Chen, M., Lobanov, V., Weng, C., and Xu, H. (2022a). A comparative study of pre-trained language models for named entity recognition in clinical trial eligibility criteria from multiple corpora. *BMC medical informatics and decision making*, 22(Suppl 3):235.

Li, T., Levina, E., and Zhu, J. (2020). Network cross-validation by edge sampling. *Biometrika*, 107(2):257–276.

Li, X., Luo, X., Dong, C., Yang, D., Luan, B., and He, Z. (2021a). Tdeer: An efficient translating decoding schema for joint extraction of entities and relations. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, pages 8055–8064.

Li, Z., Liu, H., Zhang, Z., Liu, T., and Xiong, N. N. (2022b). Learning knowledge graph embedding with heterogeneous relation attention networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33(8):3961–3973.

Li, Z., Zhao, H., Chang, B., and Sui, Z. (2021b). Mrgat: Multi-relational graph attention network for knowledge graph completion. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10271–10277.

Li, Z.-Z., Zhang, D., Zhang, M.-L., Zhang, J., Liu, Z., Yao, Y., Xu, H., Zheng, J., Wang, P.-J., Chen, X., et al. (2025b). From system 1 to system 2: A survey of reasoning large language models. *arXiv preprint arXiv:2502.17419*.

Liben-Nowell, D. and Kleinberg, J. (2003). The link prediction problem for social networks. In *Proceedings of the twelfth international conference on Information and knowledge management*, pages 556–559.

Lin, H., Lu, Y., Han, X., Sun, L., Dong, B., and Jiang, S. (2019). Gazetteer-enhanced attentive neural networks for named entity recognition. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6232–6237.

Lin, S., Hilton, J., and Evans, O. (2021). Truthfulqa: Measuring how models mimic human falsehoods. *arXiv preprint arXiv:2109.07958*.

Lin, Y., Liu, Z., Sun, M., Liu, Y., and Zhu, X. (2015). Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29.

Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., and Wang, P. (2020). K-bert: Enabling language representation with knowledge graph. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 2901–2908.

Liu, Y., Hou, J., Chen, Y., Jin, J., and Wang, W. (2025). Llm-acnc: Aerospace requirement texts knowledge graph construction utilizing large language model. *Aerospace*, 12(6):463.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Loiola, E. M., De Abreu, N. M. M., Boaventura-Netto, P. O., Hahn, P., and Querido, T. (2007). A survey for the quadratic assignment problem. *European journal of operational research*, 176(2):657–690.

Loreti, A., Chen, K., George, R., Firth, R., Agnello, A., and Tanaka, S. (2025). Automated construction of a knowledge graph of nuclear fusion energy for effective elicitation and retrieval of information. *arXiv preprint arXiv:2504.07738*.

Lu, Y. and Wang, J. (2025). Karma: Leveraging multi-agent llms for automated knowledge graph enrichment. *arXiv preprint arXiv:2502.06472*.

Luan, Y., He, L., Ostendorf, M., and Hajishirzi, H. (2018). Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. *arXiv preprint arXiv:1808.09602*.

Luo, H., Chen, G., Zheng, Y., Wu, X., Guo, Y., Lin, Q., Feng, Y., Kuang, Z., Song, M., Zhu, Y., et al. (2025). Hypergraphrag: Retrieval-augmented generation via hypergraph-structured knowledge representation. *arXiv preprint arXiv:2503.21322*.

Ma, X. and Hovy, E. (2016). End-to-end sequence labeling via bi-directional lstm-cnns-crf. *arXiv preprint arXiv:1603.01354*.

Ma, Y., Burns, O., Wang, M., Li, G., Du, N., El Shafey, L., Wang, L., Shafran, I., and Soltau, H. (2022). Knowledge graph reasoning with self-supervised reinforcement learning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pages 1670–1679.

Matias, C. and Miele, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(4):1119–1141.

Meilicke, C., Wudage Chekol, M., Ruffinelli, D., and Stuckenschmidt, H. (2019). Anytime bottom-up rule learning for knowledge graph completion. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI-19)*, pages 4136–4142. International Joint Conferences on Artificial Intelligence Organization.

Mihindukulasooriya, N., Tiwari, S., Enguix, C. F., and Lata, K. (2023). Text2kgbench: A benchmark for ontology-driven knowledge graph generation from text. In *International semantic web conference*, pages 247–265. Springer.

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., and Gao, J. (2024). Large language models: A survey. *arXiv preprint arXiv:2402.06196*.

Mintz, M., Bills, S., Snow, R., and Jurafsky, D. (2009). Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 1003–1011.

Mo, B., Yu, K., Kazdan, J., Mpala, P., Yu, L., Cundy, C., Kanatsoulis, C., and Koyejo, S. (2025). Kggen: Extracting knowledge graphs from plain text with language models. *arXiv preprint arXiv:2502.09956*.

Monajatipoor, M., Yang, J., Stremmel, J., Emami, M., Mohaghegh, F., Rouhsedaghat, M., and Chang, K.-W. (2024). Llms in biomedicine: A study on clinical named entity recognition. *arXiv preprint arXiv:2404.07376*.

Nathani, D., Chauhan, J., Sharma, C., and Kaul, M. (2019). Learning attention-based embeddings for relation prediction in knowledge graphs. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4710–4723.

Newman, M. E. (2001). Clustering and preferential attachment in growing networks. *Physical review E*, 64(2):025102.

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582.

Nguyen, T.-K., Liu, Z., and Fang, Y. (2023). Link prediction on latent heterogeneous graphs. In *Proceedings of the ACM Web Conference 2023*, pages 2653–2663.

Nickel, M., Tresp, V., and Kriegel, H.-P. (2011). A three-way model for collective learning on multi-relational data. pages 809–816.

Ortona, S., Meduri, V. V., and Papotti, P. (2018). Rudik: Robust discovery of positive and negative rules in knowledge bases. In *Proceedings of the 34th IEEE International Conference on Data Engineering (ICDE 2018)*, pages 1168–1179. IEEE.

Palikhe, A., Yu, Z., Wang, Z., and Zhang, W. (2025). Towards transparent ai: A survey on explainable large language models. *arXiv preprint arXiv:2506.21812*.

Pan, H., Zhang, Q., Adamu, M., Dragut, E., and Latecki, L. J. (2025). Taxonomy-driven knowledge graph construction for domain-specific scientific applications. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 4295–4320.

Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., and Wu, X. (2024). Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 36(7):3580–3599.

Paolini, G., Athiwaratkun, B., Krone, J., Ma, J., Achille, A., Anubhai, R., Santos, C. N. d., Xiang, B., and Soatto, S. (2021). Structured prediction as translation between augmented natural languages. *arXiv preprint arXiv:2101.05779*.

Parović, M., Li, Z., and Du, J. (2025). Generating domain-specific knowledge graphs from large language models. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 11558–11574.

Patil, A. and Jadon, A. (2025). Advancing reasoning in large language models: Promising methods and approaches. *arXiv preprint arXiv:2502.03671*.

Pister, A. and Barthelemy, M. (2024). Stochastic block hypergraph model. *Physical Review E*, 110(3):034312.

Pitis, S., Zhang, M. R., Wang, A., and Ba, J. (2023). Boosted prompt ensembles for large language models. *arXiv preprint arXiv:2304.05970*.

Pusch, L. and Conrad, T. O. (2024). Combining llms and knowledge graphs to reduce hallucinations in question answering. *arXiv preprint arXiv:2409.04181*.

Qin, L., Chen, Q., Feng, X., Wu, Y., Zhang, Y., Li, Y., Li, M., Che, W., and Yu, P. S. (2024). Large language models meet nlp: A survey. *arXiv preprint arXiv:2405.12819*.

Qu, M., Chen, X., Xhonneux, L.-P., Bengio, Y., and Tang, J. (2021). Rnnlogic: Learning logic rules for reasoning on knowledge graphs. In *International Conference on Learning Representations (ICLR)*.

Rabiner, L. R. (2002). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Rácz, M. Z. and Bubeck, S. (2017). Basic models and questions in statistical network analysis.

Rashad, M., Zahran, A., Amin, A., Abdelaal, A., and AlTantawy, M. (2024). Factalign: Fact-level hallucination detection and classification through knowledge graph alignment. In *Proceedings of the 4th Workshop on Trustworthy Natural Language Processing (TrustNLP 2024)*, pages 79–84.

Riedel, S., Yao, L., and McCallum, A. (2010). Modeling relations and their mentions without labeled text. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 148–163. Springer.

Riesen, K. and Bunke, H. (2009). Approximate graph edit distance computation by means of bipartite graph matching. *Image and Vision computing*, 27(7):950–959.

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social networks*, 29(2):173–191.

Rohe, K., Chatterjee, S., and Yu, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, pages 1878–1915.

Roth, D. and Yih, W.-t. (2004). A linear programming formulation for global inference in natural language tasks.

Sadeghian, A., Armandpour, M., Ding, P., and Wang, D. Z. (2019). Drum: End-to-end differentiable rule mining on knowledge graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 15321–15331.

Saha, S., Yadav, P., Bauer, L., and Bansal, M. (2021). Explagraphs: An explanation graph generation task for structured commonsense reasoning. *arXiv preprint arXiv:2104.07644*.

Sanfeliu, A. and Fu, K.-S. (2012). A distance measure between attributed relational graphs for pattern recognition. *IEEE transactions on systems, man, and cybernetics*, (3):353–362.

Saxena, A., Chakrabarti, S., and Talukdar, P. (2021). Question answering over temporal knowledge graphs. *arXiv preprint arXiv:2106.01515*.

Schlichtkrull, M., Kipf, T. N., Bloem, P., Van Den Berg, R., Titov, I., and Welling, M. (2018). Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer.

Seganti, A., Firląg, K., Skowronska, H., Satława, M., and Andruszkiewicz, P. (2021). Multilingual entity and relation extraction dataset and model. In *Proceedings of the 16th conference of the european chapter of the association for computational linguistics: Main volume*, pages 1946–1955.

Sengupta, S. (2025). Statistical network analysis: Past, present, and future. In *Frontiers of Statistics and Data Science*, pages 153–179. Springer.

Shang, C., Tang, Y., Huang, J., Bi, J., He, X., and Zhou, B. (2019). End-to-end structure-aware convolutional networks for knowledge base completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3060–3067.

Sharma, S., Nayak, T., Bose, A., Meena, A. K., Dasgupta, K., Ganguly, N., and Goyal, P. (2022). Finred: A dataset for relation extraction in financial domain. In *Companion Proceedings of the Web Conference 2022*, pages 595–597.

Shen, Y., Chen, J., Huang, P.-S., Guo, Y., and Gao, J. (2018). M-walk: Learning to walk over graphs using monte carlo tree search. In *Advances in Neural Information Processing Systems*, volume 31, pages 6787–6798.

Shu, D., Chen, T., Jin, M., Zhang, C., Du, M., and Zhang, Y. (2024). Knowledge graph large language model (kg-llm) for link prediction. In *Proceedings of the 16th Asian Conference on Machine Learning (ACML 2024)*, volume 260 of *Proceedings of Machine Learning Research*, pages 143–158. PMLR.

Sivarajkumar, S., Kelley, M., Samolyk-Mazzanti, A., Visweswaran, S., and Wang, Y. (2024). An empirical evaluation of prompting strategies for large language models in zero-shot clinical natural language processing: algorithm development and validation study. *JMIR Medical Informatics*, 12:e55318.

Snijders, T. A. et al. (2002). Markov chain monte carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2):1–40.

Soares, L. B., FitzGerald, N., Ling, J., and Kwiatkowski, T. (2019). Matching the blanks: Distributional similarity for relation learning. *arXiv preprint arXiv:1906.03158*.

Socher, R., Chen, D., Manning, C. D., and Ng, A. (2013). Reasoning with neural tensor networks for knowledge base completion. In *Advances in neural information processing systems*, pages 926–934.

Song, C. H., Lawrie, D. J., Finin, T., Mayfield, J., et al. (2020). Gazetteer generation for neural named entity recognition. In *FLAIRS*, pages 298–302.

Song, S., Yang, C., Xu, L., Shang, H., Li, Z., and Chang, Y. (2024). Travelrag: A tourist attraction retrieval framework based on multi-layer knowledge graph. *ISPRS International Journal of Geo-Information*, 13(11):414.

Spiess, C., Vaziri, M., Mandel, L., and Hirzel, M. (2025). Autopdl: Automatic prompt optimization for llm agents. *arXiv preprint arXiv:2504.04365*.

Stoica, G., Platanios, E. A., and Póczos, B. (2021). Re-tacred: Addressing shortcomings of the tacred dataset. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 13843–13850.

Strubell, E., Verga, P., Belanger, D., and McCallum, A. (2017). Fast and accurate sequence labeling with iterated dilated convolutions. *arXiv preprint arXiv:1702.02098*, 138.

Stubbs, A., Kotfila, C., and Uzuner, Ö. (2015). Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/uthealth shared task track 1. *Journal of biomedical informatics*, 58:S11–S19.

Sun, Q. and Bhatia, P. (2021). Neural entity recognition with gazetteer based fusion. *arXiv preprint arXiv:2105.13225*.

Sun, Y., Shi, L., and Tong, Y. (2024). expath: Explaining knowledge graph link prediction with ontological closed path rules. *arXiv preprint arXiv:2412.04846*.

Sun, Z., Deng, Z.-H., Nie, J.-Y., and Tang, J. (2019a). Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*.

Sun, Z., Deng, Z.-H., Nie, J.-Y., and Tang, J. (2019b). Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*.

Surdeanu, M., Tibshirani, J., Nallapati, R., and Manning, C. D. (2012). Multi-instance multi-label learning for relation extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 455–465. Association for Computational Linguistics.

Tang, W., Xu, B., Zhao, Y., Mao, Z., Liu, Y., Liao, Y., and Xie, H. (2022). Unirel: Unified representation and interaction for joint relational triple extraction. *arXiv preprint arXiv:2211.09039*.

Tata, V., Bouchamaoui, Z., and Bhaskara, N. V. (2025). OrthographRAG: Enhancing clinical decision making with multi-level knowledge graphs. In *ICML 2025 Generative AI and Biology (GenBio) Workshop*.

Tinn, R., Cheng, H., Gu, Y., Usuyama, N., Liu, X., Naumann, T., Gao, J., and Poon, H. (2023). Fine-tuning large neural language models for biomedical natural language processing. *Patterns*, 4(4).

Tonolini, F., Aletras, N., Massiah, J., and Kazai, G. (2024). Bayesian prompt ensembles: Model uncertainty estimation for black-box large language models. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 12229–12272.

Traag, V. A., Waltman, L., and Van Eck, N. J. (2019). From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12.

Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., and Bouchard, G. (2016). Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080. PMLR.

Turnbull, K., Lunagómez, S., Nemeth, C., and Airoldi, E. (2024). Latent space modeling of hypergraph data. *Journal of the American Statistical Association*, 119(548):2634–2646.

Vashishth, S., Sanyal, S., Nitin, V., and Talukdar, P. (2020a). Composition-based multi-relational graph convolutional networks. In *International Conference on Learning Representations*.

Vashishth, S., Sanyal, S., Nitin, V., and Talukdar, P. (2020b). Interacte: Improving convolution-based knowledge graph embeddings by increasing feature interactions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3009–3016.

Vayer, T., Chapel, L., Flamary, R., Tavenard, R., and Courty, N. (2020). Fused gromov-wasserstein distance for structured objects. *Algorithms*, 13(9):212.

Vishwanathan, S. V. N., Schraudolph, N. N., Kondor, R., and Borgwardt, K. M. (2010). Graph kernels. *The Journal of Machine Learning Research*, 11:1201–1242.

Vrandečić, D. and Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.

Wadhwa, S., Amir, S., and Wallace, B. C. (2023). Revisiting relation extraction in the era of large language models. In *Proceedings of the conference. association for computational linguistics. meeting*, volume 2023, page 15566.

Walker, C., Strassel, S., Medero, J., and Maeda, K. (2006). Ace 2005 multilingual training corpus. *(No Title)*.

Wan, J., Ru, D., Zhang, W., and Yu, Y. (2022). Nested named entity recognition with span-level graphs. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 892–903.

Wang, B., Shen, T., Long, G., Zhou, T., Wang, Y., and Chang, Y. (2021). Structure-augmented text representation learning for efficient knowledge graph completion. In *Proceedings of the Web Conference 2021 (WWW '21)*, pages 1737–1748. ACM.

Wang, C., Liu, X., Chen, Z., Hong, H., Tang, J., and Song, D. (2022a). Deepstruct: Pre-training of language models for structure prediction. *arXiv preprint arXiv:2205.10475.*

Wang, S., Fang, Y., Zhou, Y., Liu, X., and Ma, Y. (2025). Archrag: Attributed community-based hierarchical retrieval-augmented generation. *arXiv preprint arXiv:2502.09891.*

Wang, S., Sun, X., Li, X., Ouyang, R., Wu, F., Zhang, T., Li, J., and Wang, G. (2023a). Gpt-ner: Named entity recognition via large language models. *arXiv preprint arXiv:2304.10428.*

Wang, S., Sun, X., Li, X., Ouyang, R., Wu, F., Zhang, T., Li, J., and Wang, G. (2023b). Gpt-ner: Named entity recognition via large language models. *arXiv preprint arXiv:2304.10428.*

Wang, W. Y., Mazaitis, K., Lao, N., Mitchell, T., and Cohen, W. W. (2013). Efficient inference and learning in a large knowledge base: Reasoning with extracted information using a locally groundable first-order probabilistic logic. In *Proceedings of the 2013 Conference on Uncertainty in Artificial Intelligence (UAI).*

Wang, X., He, Q., Liang, J., and Xiao, Y. (2022b). Language models as knowledge embeddings. In *Proceedings of the Thirty-First International Joint Conference on Artificial*

*Intelligence (IJCAI-22)*, pages 2291–2297. International Joint Conferences on Artificial Intelligence Organization.

Wang, X., Jiang, Y., Bach, N., Wang, T., Huang, Z., Huang, F., and Tu, K. (2020a). Automated concatenation of embeddings for structured prediction. *arXiv preprint arXiv:2010.05006*.

Wang, Y., Yu, B., Zhang, Y., Liu, T., Zhu, H., and Sun, L. (2020b). Tplinker: Single-stage joint extraction of entities and relations through token pair linking. *arXiv preprint arXiv:2010.13415*.

WANG, Y. R. and BICKEL, P. J. (2017). Likelihood-based model selection for stochastic block models1. *The Annals of Statistics*, 45(2):500–528.

Wang, Z., Zhang, J., Feng, J., and Chen, Z. (2014). Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI conference on artificial intelligence*, volume 28.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837.

Wei, W., Song, Y., and Yao, B. (2024). Enhancing heterogeneous knowledge graph completion with a novel gat-based approach (gath). *arXiv preprint arXiv:2408.02456*.

Wei, X., Cui, X., Cheng, N., Wang, X., Zhang, X., Huang, S., Xie, P., Xu, J., Chen, Y., Zhang, M., et al. (2023a). Chatie: Zero-shot information extraction via chatting with chatgpt. *arXiv preprint arXiv:2302.10205*.

Wei, Y., Huang, Q., Zhang, Y., and Kwok, J. T. (2023b). Kicgpt: Large language model with knowledge in context for knowledge graph completion. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 8667–8683, Singapore. Association for Computational Linguistics.

Wilson, J. D., Palowitch, J., Bhamidi, S., and Nobel, A. B. (2017). Community extraction in multilayer networks with heterogeneous community structure. *Journal of Machine Learning Research*, 18(149):1–49.

Wilson, R. C. and Zhu, P. (2008). A study of graph spectra for comparing graphs and trees. *Pattern Recognition*, 41(9):2833–2841.

Wu, J., Shi, W., Cao, X., Chen, J., Lei, W., Zhang, F., Wu, W., and He, X. (2021). Disenkgat: Knowledge graph embedding with disentangled graph attention network. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management (CIKM)*, pages 2140–2149.

Xiong, W., Hoang, T., and Wang, W. Y. (2017). Deeppath: A reinforcement learning method for knowledge graph reasoning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 564–573. Association for Computational Linguistics.

Xu, H., Luo, D., Zha, H., and Duke, L. C. (2019). Gromov-wasserstein learning for graph matching and node embedding. In *International conference on machine learning*, pages 6932–6941. PMLR.

Xu, Q., Qiu, F., Zhou, G., Zhang, C., Ding, K., Chang, F., Lu, F., Yu, Y., Ma, D., and Liu, J. (2025). A large language model-enabled machining process knowledge graph construction method for intelligent process planning. *Advanced Engineering Informatics*, 65:103244.

Xu, S., Zhen, Y., and Wang, J. (2023). Covariate-assisted community detection in multi-layer networks. *Journal of Business & Economic Statistics*, 41(3):915–926.

Xue, L., Zhang, D., Dong, Y., and Tang, J. (2024). Autore: Document-level relation extraction with large language models. *arXiv preprint arXiv:2403.14888*.

Yamada, I., Asai, A., Shindo, H., Takeda, H., and Matsumoto, Y. (2020). Luke: Deep contextualized entity representations with entity-aware self-attention. *arXiv preprint arXiv:2010.01057*.

Yan, H., Deng, B., Li, X., and Qiu, X. (2019a). Tener: adapting transformer encoder for named entity recognition. *arXiv preprint arXiv:1911.04474*.

Yan, J., He, L., Huang, R., Li, J., and Liu, Y. (2019b). Relation extraction with temporal reasoning based on memory augmented distant supervision. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1019–1030.

Yan, X., Shalizi, C., Jensen, J. E., Krzakala, F., Moore, C., Zdeborová, L., Zhang, P., and Zhu, Y. (2014). Model selection for degree-corrected block models. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(5):P05007.

Yang, B., Yih, S. W.-t., He, X., Gao, J., and Deng, L. (2015). Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the International Conference on Learning Representations (ICLR) 2015*.

Yang, R., Ali, M. A., Wang, H., Chen, J., and Wang, D. (2025). Luster: Link prediction utilizing shared-latent space representation in multi-layer networks. In *Proceedings of the ACM on Web Conference 2025*, pages 2476–2487.

Yao, L., Demeter, R., and Riedel, S. (2019a). KG-BERT: BERT for Knowledge Graph Completion.

Yao, Y., Ye, D., Li, P., Han, X., Lin, Y., Liu, Z., Liu, Z., Huang, L., Zhou, J., and Sun, M. (2019b). Docred: A large-scale document-level relation extraction dataset. *arXiv preprint arXiv:1906.06127*.

Ye, D., Lin, Y., Du, J., Liu, Z., Li, P., Sun, M., and Liu, Z. (2020). Coreferential reasoning learning for language representation. *arXiv preprint arXiv:2004.06870*.

Yu, S., Huang, T., Liu, M., and Wang, Z. (2023). Bear: Revolutionizing service domain knowledge graph construction with llm. In *International Conference on Service-Oriented Computing*, pages 339–346. Springer.

Yuan, Y., Tao, L., Lu, H., Khushi, M., Razzak, I., Dras, M., Yang, J., and Naseem, U. (2025). Kg-uq: Knowledge graph-based uncertainty quantification for long text in large language models. In *Companion Proceedings of the ACM on Web Conference 2025*, pages 2071–2077.

Yuan, Y., Zhou, X., Pan, S., Zhu, Q., Song, Z., and Guo, L. (2021). A relation-specific attention network for joint entity and relation extraction. In *International joint conference on artificial intelligence*. International Joint Conference on Artificial Intelligence.

Yuksekgonul, M., Bianchi, F., Boen, J., Liu, S., Huang, Z., Guestrin, C., and Zou, J. (2024). Textgrad: Automatic" differentiation" via text. *arXiv preprint arXiv:2406.07496*.

Zeng, D., Liu, K., Chen, Y., and Zhao, J. (2015). Distant supervision for relation extraction via piecewise convolutional neural networks. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 1753–1762.

Zeng, D., Zhang, H., and Liu, Q. (2020a). Copymtl: Copy mechanism for joint extraction of entities and relations with multi-task learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9507–9514.

Zeng, S., Xu, R., Chang, B., and Li, L. (2020b). Double graph based reasoning for document-level relation extraction. *arXiv preprint arXiv:2009.13752*.

Zeng, X., He, S., Zeng, D., Liu, K., Liu, S., and Zhao, J. (2019). Learning the extraction order of multiple relational facts in a sentence with reinforcement learning. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 367–377.

Zhai, H. (2025). Law graphrag: An advanced legal question-answering system. In *2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA)*, pages 1407–1410. IEEE.

Zhang, J. and Cao, J. (2017). Finding common modules in a time-varying network with application to the drosophila melanogaster gene regulation network. *Journal of the American Statistical Association*, 112(519):994–1008.

Zhang, J., Wang, J., and Wang, X. (2024a). Consistent community detection in inter-layer dependent multi-layer networks. *Journal of the American Statistical Association*, 119(548):3141–3151.

Zhang, L. and Amini, A. A. (2023). Adjusted chi-square test for degree-corrected block models. *The Annals of Statistics*, 51(6):2366–2385.

Zhang*, T., Kishore*, V., Wu*, F., Weinberger, K. Q., and Artzi, Y. (2020). Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.

Zhang, Y., Chen, Z., Guo, L., Xu, Y., Zhang, W., and Chen, H. (2024b). Making large language models perform better in knowledge graph completion. In *Proceedings of the 2024 ACM Multimedia Conference (ACM MM '24)*. ACM / IW3C2.

Zhang, Y., Levina, E., and Zhu, J. (2017). Estimating network edge probabilities by neighbourhood smoothing. *Biometrika*, 104(4):771–783.

Zhang, Y., Qi, P., and Manning, C. D. (2018). Graph convolution over pruned dependency trees improves relation extraction. *arXiv preprint arXiv:1809.10185*.

Zhang, Y., Sui, X., Pan, F., Yu, K., Li, K., Tian, S., Erdengasileng, A., Han, Q., Wang, W., Wang, J., et al. (2025). A comprehensive large-scale biomedical knowledge graph for ai-powered data-driven biomedical research. *Nature Machine Intelligence*, pages 1–13.

Zhang, Z., Cai, J., Zhang, Y., and Wang, J. (2020). Learning hierarchy-aware knowledge graph embeddings for link prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 3065–3072.

Zhao, K., Xu, H., Cheng, Y., Li, X., and Gao, K. (2021). Representation iterative fusion based on heterogeneous graph neural network for joint entity and relation extraction. *Knowledge-Based Systems*, 219:106888.

Zhao, X., Deng, Y., Yang, M., Wang, L., Zhang, R., Cheng, H., Lam, W., Shen, Y., and Xu, R. (2024). A comprehensive survey on relation extraction: Recent advances and new frontiers. *ACM Computing Surveys*, 56(11):1–39.

Zhen, Y. and Wang, J. (2023). Community detection in general hypergraph via graph embedding. *Journal of the American Statistical Association*, 118(543):1620–1629.

Zhong, L., Wu, J., Li, Q., Peng, H., and Wu, X. (2023). A comprehensive survey on automatic knowledge graph construction. *ACM Computing Surveys*, 56(4):1–62.

Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., and Xu, B. (2016). Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 2: Short papers)*, pages 207–212.

Zhou, W., Huang, K., Ma, T., and Huang, J. (2021). Document-level relation extraction with adaptive thresholding and localized context pooling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14612–14620.

Zhou, Y., Muresanu, A. I., Han, Z., Paster, K., Pitis, S., Chan, H., and Ba, J. (2022). Large language models are human-level prompt engineers. In *The eleventh international conference on learning representations*.

Zhu, Y., Wang, X., Chen, J., Qiao, S., Ou, Y., Yao, Y., Deng, S., Chen, H., and Zhang, N. (2024). Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities. *World Wide Web*, 27(5):58.