SPIREX: Improving LLM-based relation extraction from RNA-focused scientific literature using graph machine learning

Emanuele Cavalleri
University of Milan, Italy
emanuele.cavalleri@unimi.it

Mauricio Soto-Gomez
University of Milan, Italy
mauricio.soto@unimi.it

Ali Pashaeibarough
University of Milan, Italy
ali.pashaeibarough@unimi.it

Dario Malchiodi
University of Milan, Italy
dario.malchiodi@unimi.it

Harry Caufield
Lawrence Berkeley National Lab, USA
jhc@lbl.gov

Justin Reese
Lawrence Berkeley National Lab, USA
justaddcoffee@gmail.com

Christopher J. Mungall
Lawrence Berkeley National Lab, USA
CJMungall@lbl.gov

Peter N. Robinson
Charité University, Berlin, Germany
peter.robinson@bih-charite.de

Elena Casiraghi
University of Milan, Italy
elena.casiraghi@unimi.it

Giorgio Valentini
University of Milan, Italy
giorgio.valentini@unimi.it

Marco Mesiti
University of Milan, Italy
marco.mesiti@unimi.it

ABSTRACT

Relation extraction from scientific literature to align with a domain ontology is a well-known challenge in natural language processing, particularly critical in precision medicine. The advent of large language models (LLMs) has enabled the development of new and effective approaches to this problem. However, the extracted relations can be prone to problems (e.g., hallucination) that must be minimized. In this paper, we present the initial development of SPIREX, an extension of the SPIRES-based system designed to extract triples from scientific literature involving RNA molecules. Our system leverages schema constraints in the formulation of LLM prompts and utilizes graph machine learning on our RNA-based knowledge graph, RNA-KG, to assess the plausibility of the extracted triples. RNA-KG comprises more than 12.5M edges representing various types of relationships involving RNA molecules.

VLDB Workshop Reference Format:
Emanuele Cavalleri, Mauricio Soto-Gomez, Ali Pashaeibarough, Dario Malchiodi, Harry Caufield, Justin Reese, Christopher J. Mungall, Peter N. Robinson, Elena Casiraghi, Giorgio Valentini, and Marco Mesiti. SPIREX: Improving LLM-based relation extraction from RNA-focused scientific literature using graph machine learning. VLDB 2024 Workshop: LLM+KG.

VLDB Workshop Artifact Availability:
The experiments have been realized by using the following datasets: (schema, docs, and manual annotations) https://zenodo.org/records/11393776; (RNA-KG): https://zenodo.org/doi/10.5281/zenodo.10078876.

Proceedings of the VLDB Endowment. ISSN 2150-8097.

1 INTRODUCTION

Ribonucleic acid (RNA) is essential in the central dogma of molecular biology, functioning as the intermediary between DNA and proteins, the fundamental building blocks of life. Beyond its traditional role in protein synthesis, RNA is involved in various cellular processes, including gene regulation and catalysis, emphasizing its critical importance in understanding the complexities of biological systems. RNA-KG [8] is an ontology-based knowledge graph (KG) that represents both coding and non-coding RNA molecules and their interactions with other biomolecular data, pathways, abnormal phenotypes, and diseases. This KG supports the study and discovery of RNA's biological roles. RNA-KG includes around 12.5M edges derived from over 60 public data sources, enabling the exploration of RNA molecules and the development of innovative graph algorithms for knowledge discovery in data science.

Manually ingesting triples into a KG by expert curators is a time-consuming and costly process. Thus, tools are strongly needed to support the domain experts in extracting biological entities and their relationships from plain texts by exploiting *Relation Extraction* (RE) approaches [11] to identify triples containing the entity mentions, their relationship, and then grounding them according to the classes and relationships made available by domain ontologies. Standard supervised RE techniques involve training models according to task-oriented corpora that are not always so easy to identify in some contexts, like the biomedical domain. Recent general-purpose large language models (LLMs) [3], like GPT-3, appear to obtain very good performances by applying prompt engineering techniques that determine reliable examples of the task to be carried out [38]. However, these techniques can produce incorrect statements due to hallucinations [13, 22] that are not acceptable in sensitive areas like precision medicine.

To face these issues, the integration of LLMs with KGs and knowledge bases appears very promising. Indeed, they provide a structured representation of the entities and relationships available in a given domain and offer rich contextual information that the LLM

can exploit in the extraction process. The utility of this integration has been recently proved through the SPIRES (Structured Prompt Interrogation and Recursive Extraction of Semantics) system [6] that exploits a knowledge schema (expressed in terms of LinkML [28]) to specify the context and the examples of interesting links. This has a positive impact on the performances of LLMs by defining more effective interacting prompts. Additionally, SPIRES allows the grounding of concepts in a variety of bio-ontologies, such as Open Bio Ontology (OBO) Foundry ontologies [21].

In this paper, we enhance the integration of SPIRES with KGs by considering the latter for the semi-automatic validation of the statements extracted from scientific documents. Indeed, when the KG already contains many facts, it can be exploited to evaluate the plausibility of the extracted triples according to the current knowledge of the domain. This is an important feature to reduce the manual efforts of the domain curators in evaluating and accepting the automatically extracted facts.

To reach this goal, we introduce *SPIREX*, a system designed to extract reliable triples from scientific papers by exploiting: *i*) the SPIRES functionalities for extracting triples involving RNA interactions; and *ii*) RNA-KG for assessing the plausibility of the extracted triples because of its wide coverage of the interactions involving RNA molecules. Initial experimental results on a manually curated testbed of 100 scientific texts are promising.

The main contributions of this paper are: *i*) the use of RNA-KG schema in combination with SPIRES for an effective extraction of triples involving RNA molecules from scientific documents; *ii*) the use of heterogeneous graph representation techniques for representing RNA-KG in the latent space and their adoption for evaluating the plausibility of the extracted triples; *iii*) a wide (even if preliminary) evaluation of SPIREX for extracting meaningful triples from RNA specific documents (the extracted triples are validated according to the ground truth stored in RNA-KG). By exploiting the proposed notion of plausibility we can identify triples that can be accepted without expert curators' validation and triples for which a second check is strictly needed.

In the reminder, Section 2 discusses related work in relation extraction and introduces the characteristics of RNA-KG, SPIRES, and link prediction approaches in KGs. Then, Section 3 details the characteristics of SPIREX. Section 4 reports our experimental results. Section 5 contains our conclusions.

2 RELATED WORK

A knowledge graph is an abstract representation of the knowledge of a given domain represented in terms of the individuals existing in the domain and their relations. More formally, a KG is a 4-tuple (N, E, N_T, E_T), where N is the set of typed nodes representing real-world entities (the available types are contained in N_T). The set E represents the typed edges between nodes, i.e. $E \subseteq N \times E_T \times N$, where E_T represents the predicate that can exist among entities in the considered domain. A triple $(s, p, o) \in E$ represents the existence of the relationship/predicate p between a subject s and an object o. A KG is *ontology-based* when the type of nodes and edges that can contain is compliant with the constraints imposed by an Ontology $\mathcal{O}$ and the structural relationships available in $\mathcal{O}$ are included in KG. The RE problem from a text T according to an ontology-based knowledge graph KG consists of: *i*) the extraction of the triples/facts (e_{i_1}, r_i, e_{i_2}), where e_{i_1} and e_{i_2} are entity mentions and r_i is the predicate representing the relationships between them identified in T; and, *ii*) the grounding of these mentions according to the classes and predicates available in the Ontology $\mathcal{O}$ to provide an unambiguous representation of the facts. In the remainder of the section, we provide a description of RNA-KG, an ontology-based KG representing the interactions involving coding and non-coding RNA molecules, different approaches for relation extraction, and discuss link prediction models that can be exploited for evaluating the plausibility of triples (s, p, o) in a KG.

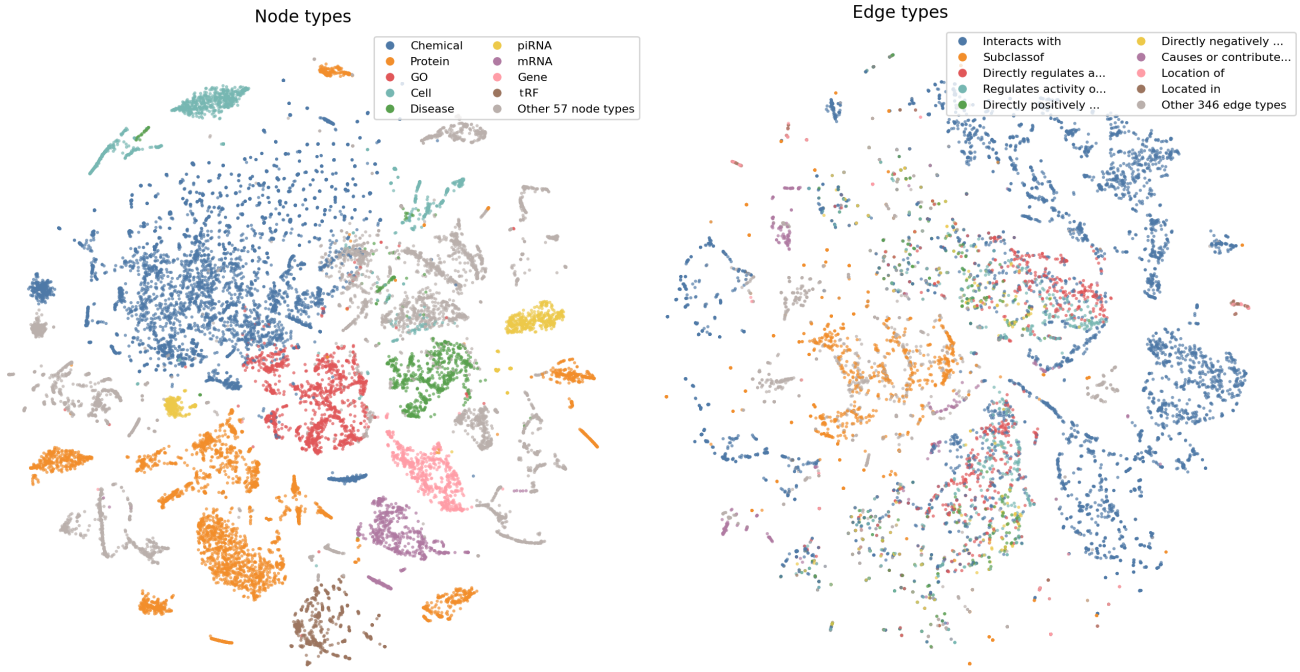
RNA-KG. It is the first KG encompassing biological knowledge about RNAs gathered from more than 60 public databases, integrating functional relationships with genes, proteins, chemicals, and ontologically grounded biomedical concepts. The current release of RNA-KG has a single component containing around 670K nodes and 12.5M edges and can be queried via SPARQL endpoint at https://RNA-KG.anacleto.di.unimi.it. Nodes are usually mapped to reference biomedical vocabularies and ontologies such as NCBI Gene Entrez identifiers for uniquely identifying genes and many kinds of non-coding RNAs (ncRNAs), Human Phenotype Ontology (HPO) for phenotypes, Monarch merged disease ontology (Mondo) for diseases, and Gene Ontology (GO) for annotating ncRNAs. Moreover, all the possible interactions are represented through the Relation Ontology (RO). This ensures common semantics for the different relationships that are extracted from the sources.

Figure 1 shows the t-SNE representation of an embedding of the nodes/edges in RNA-KG, obtained by using the GRAPE implementation of node2Vec with CBOW [16], with walk length equal to 5. Figure 1a shows how the embedding of the node type is able to effectively identify the similarities among the nodes of the same type, thus capturing their function in the network. On the other hand, Figure 1b depicts the edge embedding, which captures the similarity between edges with the only exception of the `interacts with` and `regulates activity of` relations which seem to overlap several other edge types. This fact is not so surprising considering that the `interacts with` predicate is used to denote a generic relation.

Figure 2 shows an excerpt of the kind of relationships that are available in RNA-KG among the considered entities (a more detailed representation is reported in the meta-graph presented in [7]) that will be used in this paper for relation extraction.

Relation extraction. Relation extraction from textual documents is a crucial task in natural language processing (NLP) that involves identifying and categorizing semantic relationships between entities within text [11, 31]. The SoTa techniques in RE have evolved significantly, driven by advancements in machine learning and the availability of large annotated corpora.

Traditionally, supervised learning has dominated the field of RE. These techniques rely on large datasets annotated with entity and relation labels. Models such as Support Vector Machines (SVMs), Conditional Random Fields (CRFs), and more recently, deep learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have been used to tag tokens and predict relationships. The use of pre-trained language models such as BERT [12] has further improved the accuracy of these models by leveraging contextual word embeddings. Despite their



(a) Node types embedding. **(b) Edge types embedding.**

Figure 1: Bidimensional view of RNA-KG embeddings.

success, these models require extensive labeled data, which can be expensive and time-consuming to produce.

Recent advancements have seen a shift towards sequence-to-sequence (Seq2Seq) RE models (e.g., [30, 43], and REBEL [20]). In this paradigm, the problem is reframed as a text generation task where the input sequence (text) is mapped to an output sequence (relations). This approach leverages models like Transformer-based architectures, which have been highly effective in capturing complex dependencies in text. By linearizing relations between entities as target strings, these models can generate structured information directly from the input text. This methodology has shown promise in reducing the need for extensive feature engineering and making the model more adaptable to different domains.

To mitigate the data annotation bottleneck, distant supervision methods (e.g., [27, 33, 34]) and semi-supervised learning methods (e.g., knowItAll [14], TextRunner [2], and OLLIE [25]) have been explored. Distant supervision techniques automatically generate training data by aligning text with existing knowledge bases, assuming that any sentence containing two entities related in the knowledge base expresses that relation. However, this approach often introduces noise due to incorrect assumptions. Semi-supervised methods, on the other hand, exploit bootstrapping algorithms for automatically generating labeled data. The advantages of these approaches are to reduce the efforts in generating labeled data and make use of freely available unlabelled data. Techniques such as co-training, self-training, and generative adversarial networks have been utilized to enhance the robustness and accuracy of RE models with limited labeled data [31].

All the previously presented approaches require to consider a specific corpus for training the model to the considered task. However, the identification of task-specific corpus of big size in the bio-medical domain is not always feasible and the generation of the models is a time and money consuming activity. On the other hand, general purposes LLMs (like GPT3, GPT4, LLamas) trained on billions of data are increasingly available on the web (encompassing also bio-medical documentation) and expose high capacity of reasoning. According to recent studies [38], they can be exploited for the RE problem and can achieve performances comparable with fully supervised models by exploiting prompts enhanced with a few shot examples of the task to be carried out. However, these techniques have shown different limitations, such as generating incorrect statements due to hallucinations (inaccurate, nonsensical, or irrelevant output in the given context) [22] and insensitivity to negations [13], that cannot be tolerated in sensitive domains like precision medicine. Integrating KGs and ontologies into RE systems has become increasingly popular in addressing these issues. KGs provide a structured representation of entities and their relationships, offering rich contextual information that can enhance the extraction process. Ontology-based systems use predefined schemas to guide the extraction, ensuring that the identified relationships adhere to a specified structure. Tools like SPIRES [6] (described in the next section) leverage such frameworks to improve the precision and reliability of extracted triples. These systems benefit from the logical consistency and domain-specific insights encoded in the ontologies, making them particularly useful for specialized fields like biomedicine.

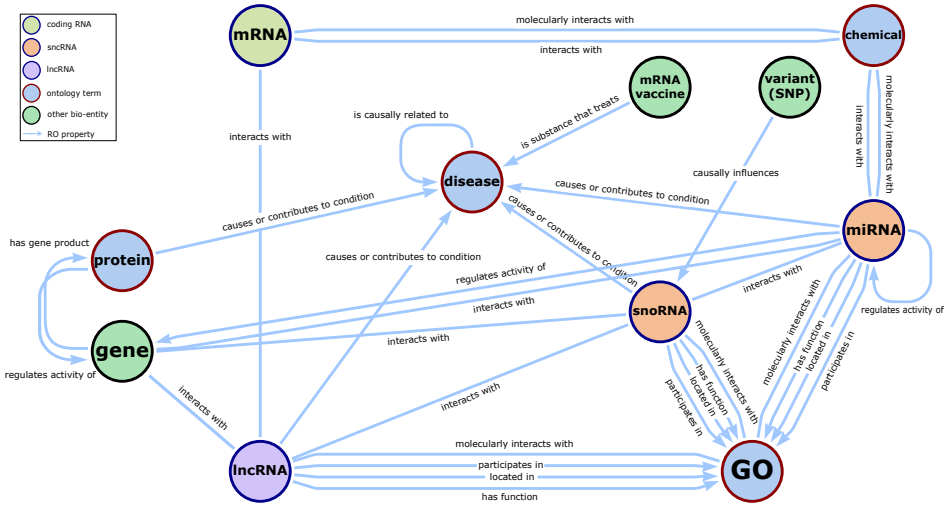


Figure 2: An excerpt of RNA-KG schema.

SPIRES. It is a recently proposed approach to information extraction that creates and refines prompts to maximize the effectiveness of general-purpose LLMs by exploiting domain knowledge encapsulated through a schema expressed in LinkML [28]. By identifying and extracting relevant information from an input text, it adopts zero-shot or few-shot learning to identify and extract relevant entities and relationships among them, which are then normalized and grounded through ontologies and vocabularies. SPIRES can be used across a variety of domains and does not require specific training/-tuning on a domain. SPIRES adopts an engineering approach for creating prompts for interacting with an LLM to improve the quality of the generated responses through the use of domain-specific schema. In this way, technical challenges for generative AI (e.g., constructing comprehensive real-world knowledge and improving the accuracy of automated responses) can be addressed.

The linkML schema specification contains the relevant classes of entities and relationships in the specified domain. Classes can also include attributes (e.g., name, type, and list of synonyms) to enrich entity description. The LinkML schema is automatically processed to generate a list of prompts through which SPIRES interacts with a LLM. Each prompt of the list is submitted to the LLM for collecting information that is exploited for completing the following prompt by eventually considering the bio-ontologies (e.g., for changing a protein symbol with the corresponding identifier in an ontology). This recursive refinement process improves the quality of the information gathered through the LLM.

Link prediction in Knowledge Graphs. Link prediction in KGs is a critical task that aims to infer missing relationships between entities, thereby enhancing the graph's completeness and utility.

Traditional approaches to link prediction rely on heuristic methods [44], such as common neighbors, preferential attachment, and the Adamic-Adar index [1], which leverage the structural properties of the graph. Recent advancements have been increasingly focused on embedding-based methods [41]. These methods involve learning low-dimensional representations of entities and relationships by capturing the graph's semantic and structural information. In this context three approaches are mostly used: Random-walk (RW) based models, Graph Neural networks (GNNs), and Relation-learning neural models [4].

The first kind of models exploits walks across the graph to rely on the "distributional hypothesis"[1], firstly exploited in word2vec [26] to capture the semantic similarity of words, and then extended to capture the similarity between graph nodes [16]. To adapt the word2vec strategy on the graph-embedding tasks, RW-based models collect each node context by running n walks from the node itself and then train a word2vec neural model, to recognize the context given the node (e.g., Skipgram) or viceversa (e.g., CBOW). Once trained, the network weights are used as node embeddings. The most popular and effective RW-based graph embedding techniques are deepWalk [32] and node2vec [16], which differ for the RW strategy. deepwalk applies a standard first-order RW, whereas node2vec leverages a second-order RW to bias the walk and promote exploration or exploitation. Both approaches are scalable and can work with huge graphs when the GRAPE [5] implementation is considered. GNNs leverage DNNs to process graphs, using, e.g., convolutional filters [23] to direct supervised feature learning in node-neighborhoods [18]. These models present all the advantages provided by supervised learning and can eventually integrate attention mechanisms of any sort, from the standard one [37] to transformers attention [42]; however, their low scalability is still hampering its application on large graphs (e.g., KGs).

Relation-learning models [4, 40] have been specifically developed for working with heterogeneous graphs like KGs and generating a latent space where the different kinds of relationships are "optimally" represented. These models rely on the use of contrastive learning techniques to project entities (subject s and object

[1]The distributional hypothesis was originally proposed in linguistics [15, 19]. It assumes that "linguistic items with similar distributions have similar meanings", from which it follows that words (elements) used and occurring in the same contexts tend to purport similar meanings [19].

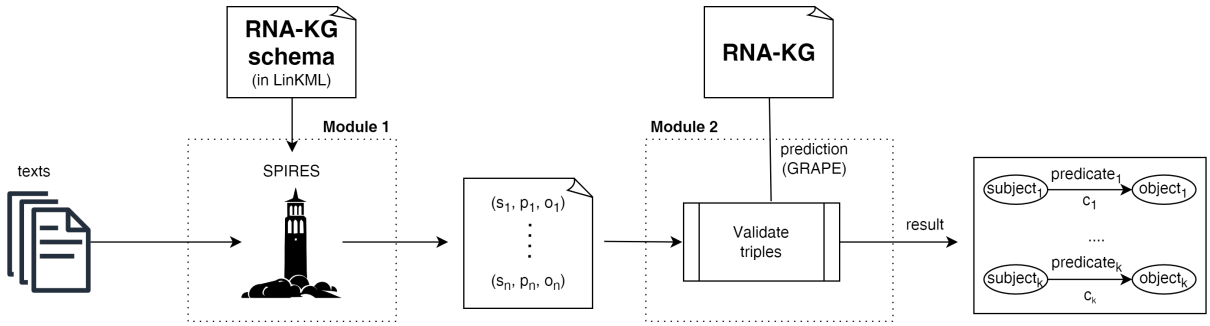


Figure 3: The SPIREX architecture.

o nodes) and the predicate p between them into low-dimensional latent spaces that preserve the relationships between entities and predicates. This is achieved by assigning a score to each (s, p, o) triple, which is maximized for true triples and minimized for "corrupted triples", i.e. triples not truly existing in the graph. TransE [4] is probably the most used relation-learning technique due to its promising results. It computes a score based on the key assumption that the vector associated with a predicate, **p**, is a translation of the subject entity vector, **s**, to the object entity vector, **o**. Based on this assumption the model optimizes a margin-based ranking loss function enforcing that for each existing triple, the sum of the embeddings of the subject and predicate are as close as possible to the embedding of the object: $\mathbf{s} + \mathbf{p} \approx \mathbf{o}$. Besides being effective and efficient in practical scenarios [5], the embedding strategy of TransE has nice mathematical properties that allow improving over other relation-learning techniques, e.g., DistMult [40].

3 THE SPIREX SYSTEM

As shown in the architecture in Figure 3, SPIREX is composed of two modules: the SPIRES module is used for extracting the triples from scientific abstracts. Then, an embedding of RNA-KG is used to validate the generated triples and score their level of plausibility.

The meta-graph in Figure 2 has been translated into LinkML and used by SPIRES as a template to guide the extraction of relationships from plain text. Figure 4 shows an excerpt of a LinkML template for extracting protein to disease relationships. In LinkML, each class can be associated with properties that SPIRES uses to refine the prompt, thereby guiding the backend general-purpose LLM to extract more accurate entities. For example, the *ProteinDiseaseRelationship* class (Figure 4a) specifies an "annotations" property containing examples of potential subject, predicate, and object for a triple (each LinkML *Triple* comprises a subject, a predicate, and an object). In *ProteinToDiseaseRelationship*, subjects are entities of type *Protein*, predicates are entities of type *causes or contributes to condition*, and objects are entities of type *Disease*. To collect *ProteinToDiseaseRelationship* entities, we encapsulate them in a *TextWithTriples* LinkML core class. The predicates "causes or contributes to condition" are grounded to the corresponding RO property (`RO:0003302`) by exploiting the "pattern" and "annotations" properties (Figure 4b). Figure 4c illustrates the representation of the *Protein* class in LinkML. Classes can be enhanced by adding properties such as "synonyms" and "sequence". The `multivalued` specification indicates that the value of "synonyms" is a list of attributes. Protein entities are grounded using the PRotein Ontology (PRO) whereas diseases are grounded to Mondo and HPO terms.

Starting from our LinkML schema, SPIRES generates a list of prompts specific to the RNA domain according to which the entities and the relationships contained in a text are extracted by considering the schema constraints. Moreover, SPIRES adopts bio-ontologies of our domain (details in [9]) for producing source and target identifiers according to the RNA-KG identification scheme and RO predicates.

Example 1. *Consider the text related to "protein-causes-disease":*

> Alpha-1 antitrypsin (AAT) deficiency is a common cause of liver disease. SERPINA1 enzyme inhibitor is also involved in different pulmonary diseases.

By exploiting the LinkML specification in Figure 4c the protein Alpha-1 antitrypsin (AAT) is identified, whereas the liver and pulmonary diseases are identified through an analogous LinkML file. Then, the LLM is asked to populate the class ProteinToDiseaseRelationship (Figure 4a) with triples whose subject is a protein and object is a disease. Predicates for these triples have to be compliant to the specified relationship: "causes or contributes to condition". This leads to the identification of the following triples:

(1) *AAT - causes or contributes to condition - liver disease*
(2) *SERPINA1 - causes or contributes to condition - pulmonary disease*

For grounding entities and relationships, SPIRES exploits the Ontology Access Kit (OAK [29]) which looks up ontology elements such as labels, definitions, relationships, and aliases. Specifically, AAT and SERPINA1 are recognized to refer to the same term (`PR:000014678`*), the predicate is grounded to* `RO:0003302` *property, liver disease is grounded to* `MONDO:0005154`*, and pulmonary disease is grounded to* `MONDO:0005275`*. Whenever an entity cannot be grounded, the prefix AUTO: is associated with the entity mentioned (e.g.,* `AUTO:gynecological_cancer`*). SPIRES outputs the following grounded triples that suit the provided schema:*

(1) `PR:000014678` – `RO:0003302` – `MONDO:0005154`
(2) `PR:000014678` – `RO:0003302` – `MONDO:0005275`

The validation of new potential triples derived from the SPIRES module can be performed by evaluating their plausibility according to the content of RNA-KG. This process can be modeled as a link prediction task of new triples on RNA-KG. To this end, in SPIREX

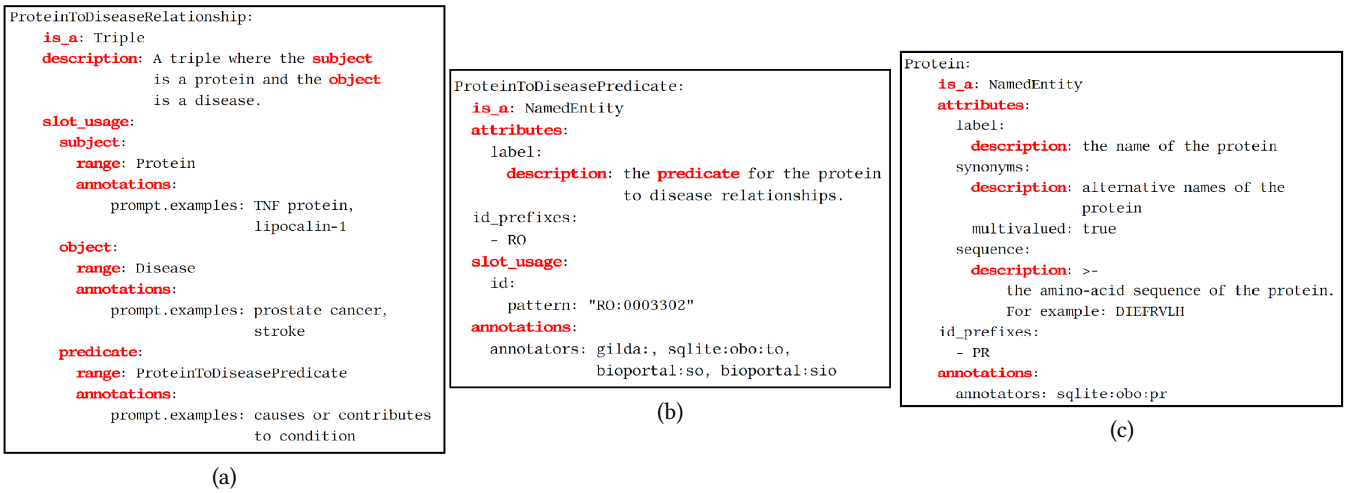


Figure 4: LinkML key classes for protein-disease interaction.

we first use a graph representation learning technique for creating a latent representation of RNA-KG. The computed entity embeddings are then used as input of a classification model that is trained to recognize existent and not-existent edges. At this stage, the plausibility of a triple (s, p, o) is assessed by using the embedding of the triple as input to the trained model to get the probability of its existence (its plausibility).

Specifically, the following three embedding algorithms have been considered: deepWalk, node2vec and TransE. The first two are RW-based models developed for homogeneous graphs, whereas the last one is tailored for working with heterogeneous graphs. The classification model is a Random Forest that outputs a score in $[0, 1]$, with a high value favoring the existence of a link, and vice versa. In this way, the predicted probability of a link can be considered as a score that quantifies the plausibility of a triple to be included according to the already existing RNA-KG content.

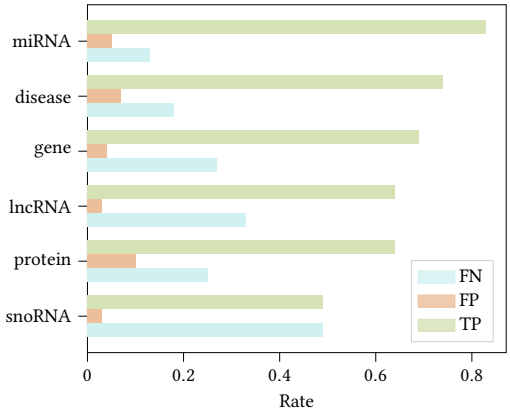
Example 2. *Consider the two triples extracted through SPIRES in previous example. By exploiting the TransE embedding of RNA-KG and the Random Forest classifier, the following evaluation of plausibility can be generated.*

(1) `PR:000014678` – `RO:0003302` – `MONDO:0005154` – *Plausibility* 0.35
(2) `PR:000014678` – `RO:0003302` – `MONDO:0005275` – *Plausibility* 0.89

In this case, we can conclude that the second triple is "more plausible" and can be considered for direct inclusion in the KG. Conversely, the first triple, with a lower plausibility score, requires manual verification by a domain expert before it can be confirmed.

4 EXPERIMENTAL RESULTS

Experiments have been realized for both modules of SPIREX. For the first module, we evaluated the prediction accuracy of SPIRES in extracting triples (by considering a set of manually annotated documents). We also compared SPIRES with base LLMs to verify the advantage of using LinkML in the specification of the domain schema. For the second module, we first assessed the performance

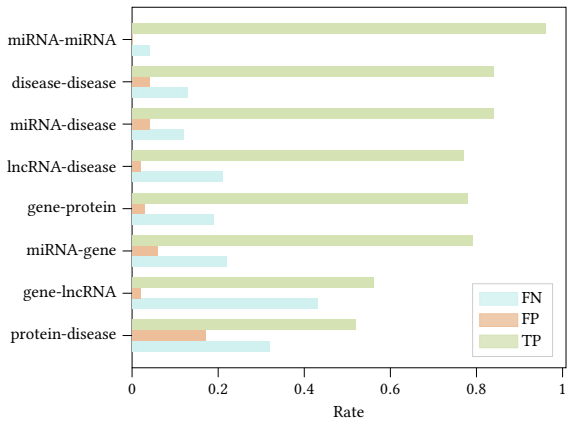


	# Paragraphs	TP	FP	FN	F-score	Precision	Recall
miRNA	42	238	14	36	0.90	0.94	0.87
disease	75	312	31	77	0.86	0.91	0.80
gene	45	199	11	79	0.82	0.95	0.72
lncRNA	19	76	4	39	0.78	0.95	0.66
protein	46	155	24	62	0.78	0.87	0.71
snoRNA	11	17	1	17	0.66	0.94	0.50

Figure 5: RE results for SPIRES-GPT4t. For the sake of readability, the entities involved in relationships appearing in at least 10 paragraphs are reported. Grounding is realized with HGNC, PRO, Mondo, HPO, and RO.

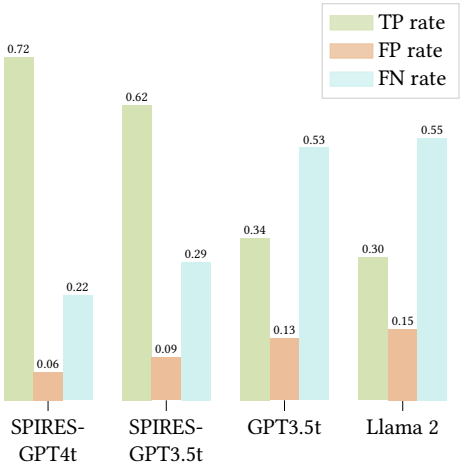
of three state-of-the-art link-prediction models when applied on RNA-KG views; they were finally used to evaluate the plausibility of triples extracted through SPIRES according to RNA-KG.

SPIRES prediction accuracy and comparison with base LLMs. A corpus of 100 scientific articles related to RNA molecules and their interactions was gathered from PubMed, ResearchGate, and Google Scholar. Starting from them, we identified abstracts, discussions, or specific subsections within the domain of interest. They were manually annotated with the entities and the kinds of interactions that can be extracted from them (reported in Figure 2). For evaluating



	# Paragraphs	TP	FP	FN	F-score	Precision	Recall
miRNA-miRNA	1	24	0	1	0.98	1.0	0.96
disease-disease	18	66	3	10	0.92	0.96	0.87
miRNA-disease	37	145	7	20	0.92	0.95	0.88
lncRNA-disease	9	36	1	10	0.90	0.97	0.82
gene-protein	21	90	3	22	0.88	0.97	0.80
miRNA-gene	24	78	6	15	0.88	0.93	0.84
gene-lncRNA	8	30	1	23	0.72	0.97	0.57
protein-disease	29	65	21	40	0.68	0.76	0.62

Figure 6: RE results for SPIRES-GPT4t for each category of interaction involving entities mentioned in Figure 5 and presenting at least 25 occurrences.



	TP	FP	FN	F-score	Precision	Recall
SPIRES-GPT4t	559	44	170	0.84	0.93	0.77
SPIRES-GPT3.5t	497	70	232	0.77	0.88	0.68
GPT3.5t	287	105	442	0.50	0.73	0.39
Llama 2	256	127	473	0.46	0.67	0.35

Figure 7: Comparing SPIRES, Llama, and GPT in the RE task on the manually annotated dataset.

the obtained predictions, we have used standard metrics (precision, recall, and F-score) by considering the True Positive (TP), False Positive (FP), and False Negative (FN) according to the manually tagged paragraphs.

Figure 5 shows the results obtained using SPIRES (with the GPT4t engine) for extracting relations involving a given entity. The y-axis reports the entities that are present in relationships appearing in at least 10 paragraphs, while the x-axis shows the true positive TP, FP, and FN rates. We observe that precision values consistently exceed recall values due to the higher number of FPs compared to FNs. We achieved high precision on the 42 documents involving miRNAs because these molecules are always specified using recognizable patterns and no synonyms (e.g., the "miR-" substring followed by numbers and the "-3p" or "-5p" substring). Conversely, relations involving snoRNA molecules are often undetected because they act in biomolecular complexes such as "SNHG8 interacts with miR-152/c-MET". These relations are often misinterpreted by an LLM, as it fails to recognize that "miR-152/c-MET" consists of two distinct entities, which would need the extraction and grounding of two separate relations ("SNHG8 interacts with miR-152" and "SNHG8 interacts with c-MET").

Figure 6 illustrates results obtained using SPIRES (with the GPT4t engine) for each category of interaction involving entities presented in Figure 5 and mentioned in at least 25 relations. A consistent trend is evident where the TP rate is higher than both the FP and FN rates. The only exceptions are gene-lncRNA and protein-disease relations, where the FN rate is higher compared to the other relations. These types of relations are often undetected because they are expressed in complex ways, leading to inaccurate entity recognition and subsequent grounding. For instance, the interchangeable use of symbols like "/" and "," (e.g., "lncRNAs mapped to chromosome 8q.24, such as CACS11, CCAT family, and PVT1, promote CRC progression by interacting with proteins to stimulate myc or other Wnt target gene expression at the posttranslational level"). Additionally, mapping proteins to PRO is challenging when textual information is sparse or ambiguously expressed. For instance, the mention of "PMP-22" solely as "myelin 22" instead of "peripheral myelin protein 22" can lead to inaccurate grounding [9]. Despite this, the overall precision remains remarkably high and, in biomedicine, this is preferable because it prioritizes certainty over ambiguity.

To assess the performance of SPIRES we compared it with OpenAI GPT (ver. GPT-3.5-t) and Llama 2 [36] (ver. llama-2-70b-chat). As back-end LLM of SPIRES, we considered both GPT3.5t and GPT4t. Regarding the prompt to be used with the base LLM system, we considered a simple one requesting to extract triples from the considered text with an explicit request for mapping the extracted concepts to appropriate terminologies. Given that both OpenAI GPT and Llama 2 caution that the ontology identifiers provided are hypothetical and might not align with actual identifiers in the ontologies, and considering the general community advice against relying on IDs from an LLM [17], we decided to substitute the grounding process with our manually curated look-up tables [8].

As shown in Figure 7, SPIRES outperforms OpenAI GPT-3.5t or Llama 2 alone both in terms of precision and recall. The histogram points out a high increment in TP rate and a decrease in FP and FN rates when adopting SPIRES for extracting relations that adhere to a specified schema within texts. Furthermore, our results are comparable to the named entity extraction analysis involving diseases and chemicals on the *BC5CDR* corpus presented in [6]. By adopting GPT-4t in SPIRES, the recall metric improves due to a lower FN rate, positively affecting the F-score. Our results also show a slight

View	KG embedding	F1	AUC	Precision	Recall
RNA-KG$_{\text{miRNA}}$	deepwalk	0.8577/0.8537	0.8190/0.8009	0.7866/0.7855	**0.9430/0.9349**
	node2vec	**0.8605/0.8499**	**0.8387/0.8145**	**0.8005/0.7973**	0.9303/0.9102
	TransE	0.8524/0.8470	0.8108/0.7920	0.7871/0.7855	0.9295/0.9192
RNA-KG$_{\text{miRNA-10}}$	deepwalk	0.8383/0.8257	0.8051/0.7839	0.8044/0.8005	0.8750/0.8525
	node2vec	**0.8531/0.8454**	0.8315/0.8156	0.8152/0.8126	**0.8947/0.8809**
	TransE	0.8437/0.8223	**0.8333/0.8052**	**0.8160/0.8092**	0.8734/0.8357
RNA-KG$_{\text{RNA-10}}$	deepwalk	0.8429/0.8342	0.8192/0.8010	0.8021/0.7992	**0.8886/0.8727**
	node2vec	0.8189/0.7990	0.7984/0.7737	0.8077/0.8013	0.8304/0.7967
	TransE	**0.8470/0.8362**	**0.8291/0.8112**	**0.8128/0.8094**	0.8842/0.8649
RNA-KG$_{\text{piRNA-10}}$	deepwalk	**0.8801/0.8794**	**0.8524/0.8328**	**0.8111/0.8115**	**0.9620/0.9597**
	node2vec	0.8639/0.8614	0.8367/0.8154	0.8000/0.7998	0.9389/0.9335
	TransE	0.8258/0.8176	0.7719/0.7436	0.7754/0.7727	0.8832/0.8680
RNA-KG$_{\text{piRNA}}$	deepwalk	**0.8752/0.8749**	**0.8378/0.8163**	**0.7991/0.7989**	**0.9674/0.9667**
	node2vec	0.8563/0.8518	0.8291/0.8057	0.7972/0.7956	0.9250/0.9166
	TransE	0.8228/0.8131	0.7747/0.7467	0.7789/0.7754	0.8719/0.8546

Table 1: Link prediction performance metrics for different prediction models in the train/validation sets.

improvement in the F-score with respect to the one reported in [6] because in the RE task we first extract entities, but only a few of them are involved in relations, (thus, a subset of entities are erroneously discarded by SPIRES). Moreover, in our domain, relations are often presented in clusters rather than separately (e.g., "miR-155-5p plays a critical role in various physiological and pathological processes such as hematopoietic lineage differentiation, immunity, inflammation, viral infections, cancer, cardiovascular disease, and Down syndrome."). If SPIRES detects one relation correctly, likely also the others in the cluster are correctly identified. Additionally, apart from proteins, chemicals, and diseases, many other entities are well-specified in the text (as mentioned, in case of miRNAs, they follow a precise pattern) and this leads to an increase in the TP rate. Finally, in [6] diseases were grounded considering only MeSH terms. By contrast, both Mondo and HPO are used as annotators which, as the authors suggest, led to improve the performances.

Evaluation of link prediction models on RNA-KG. To evaluate the capabilities of the prediction module of SPIREX, we considered five distinct views of RNA-KG and generated their vectorial representations using the deepwalk, node2vec, and TransE models. These representations are then used to train a prediction model to estimate the link plausibility.

Each view is a subgraph of RNA-KG and is defined based on a schema that is a subset of the one in Figure 2, i.e., it includes entities and relations (e.g., the relations between miRNA sequences and diseases) that focus on predicting specific relations between RNA molecules and other entities (e.g., diseases or GO terms).

We considered the possibility that training predictive models on data from public sources focusing on specific RNAs, GO terms, or diseases (e.g., miRCancer, which focuses on miRNA-cancer interactions) could introduce bias. However, we believe that using RNA-KG for generating views mitigates this risk because it has been developed by the integration of more than 60 sources covering a wide variety of RNAs, GO terms, and diseases.

For each view, we extract a subset of triples that defines the test set used to assess the link-prediction task. To create the test set, two different strategies are considered. In the first strategy, the links extracted from a specific data source are eliminated from the view and used as test set. In the second strategy, a sample corresponding to the 10% of the triples in the view is eliminated from the view and used as test set. In both strategies, graph connectivity is guaranteed according to a connected Monte-Carlo hold-out policy [5]. Using these strategies, the following views have been generated:

(1) RNA-KG$_{\text{miRNA}}$. It contains ∼1.1M triples involving the following kinds of entities: miRNA, disease, and gene. The test is constructed according to the first strategy and corresponds to the triples extracted from the source RNAdisease [10].
(2) RNA-KG$_{\text{miRNA-10}}$. Analogously to the previous case, it contains ∼1.1M the triples involving genes, miRNAs, and diseases, but the test set is obtained according to the second strategy.
(3) RNA-KG$_{\text{RNA-10}}$. It contains ∼6M triples involving the following kinds of entities: miRNA, lncRNA, protein, circRNA, disease, gene, and GO terms. By applying the second strategy, 10% of the triples is left for the test set.
(4) RNA-KG$_{\text{piRNA-10}}$. It contains ∼2M triples involving the following kinds of entities: variant (SNP), piRNA, disease, lncRNA, miRNA, gene, and GO terms. The test set is obtained according to the second strategy.
(5) RNA-KG$_{\text{piRNA}}$. It contains ∼2M triples involving the same kinds of entities of the previous view. The test set is constructed according to the first strategy excluding the source piRBase [39].

For the creation of the predictive model, 80% of each view is used as a training set, while a validation set is composed of the remaining 20% of the view (positive examples) plus the same amount of negative samples that random triples not occuring in RNA-KG.

DeepWalk, node2vec, and TransE were used for creating the embeddings of the training samples, which were then fed to a Random Forest classifier trained for link prediction. The values of the hyperparameters for the embedding and prediction models were

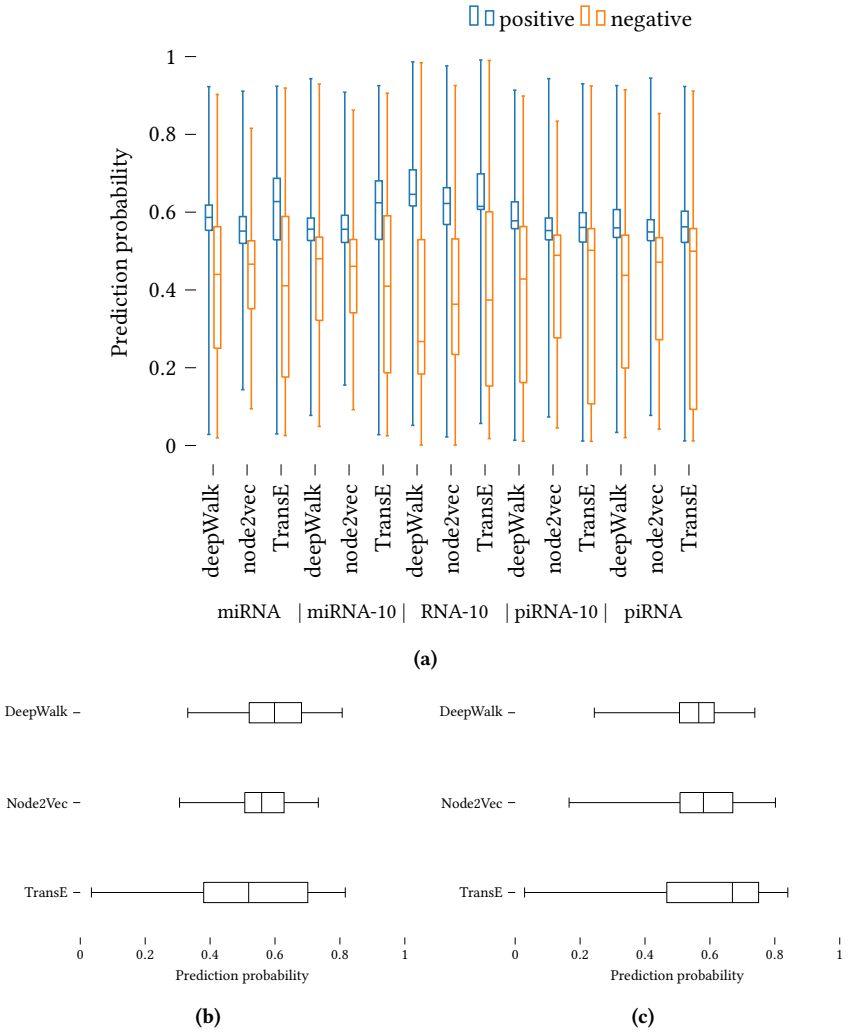


Figure 8: (a) Distribution of the link-prediction probabilities for the positive/negative edges of the validation set. (b) Distribution of the link-prediction probabilities in the test for the view RNA-KG$_{\text{miRNA}}$. (c) Distribution of the link-prediction probabilities in the test for the view RNA-KG$_{\text{miRNA-10}}$.

set at the default values proposed in GRAPE [5]. Table 1 reports the performance of the link prediction task based on the different graph representation models for each view in the training and validation sets; in all the settings, the predictive models attain a good level of prediction capabilities. Moreover, Figure 8a shows the probability distribution of the predictions of each model in all the views for the positive and negative samples of the validation set. As we can see, the models can successfully discriminate between positive and negative examples in the validation set. Finally, each model has been trained on the whole view and then applied to the test samples to obtain the link probabilities.

For the RNA-KG$_{\text{miRNA}}$ and RNA-KG$_{\text{miRNA-10}}$ views, Figures 8b-8c depict the probability distribution of the model predictions in the test sets. We can note that most of the (positive) relations in the test set are identified. The same behavior has been observed also for the other views that are not reported here for space constraints.

According to these results, the predictive module can be used to validate the relations to include in RNA-KG and hence to assess the plausibility of SPIRES predictions.

Evaluation of the plausibility of SPIREX predictions. The analysis conducted in the previous section is further extended to evaluate the ability of the predictive module to assess the plausibility of the triples extracted by SPIRES. To this end, true positive triples extracted from our manually curated dataset have been considered. Figure 9a shows the distribution of the probabilities predicted on the miRNA-disease triples extracted by SPIRES according to the different representation models computed on RNA-KG$_{\text{miRNA-10}}$. Green boxes represent the distributions of probabilities computed over miRNA-disease links that are already included in RNA-KG view, whereas red boxes represent the distributions of probabilities computed over edges that are missing in RNA-KG. We highlight

the capabilities of the models to correctly classify almost all the links already present in the RNA-KG view. They also discriminate between plausible and implausible new triples, offering a potential validation tool.

An histogram of the triples predicted by SPIRES on the RNA-$KG_{miRNA\text{-}10}$ view according to deepwalk predictions is provided in Figure 9b. In this case, green/red bars represent the probability of links that are already included in the view (*included*) or that are new triples that could be introduced in the view (*not included*). Note that a large amount of (non yet included in RNA-KG) triples extracted by SPIRES has been independently "validated" by using RNA-KG, and can therefore be considered as plausible new candidates for miRNA-disease relationships.

On the other hand, some triples proposed by SPIRES have a low confidence according to the predictive module. These edges can be considered "uncertain" given that they are not confirmed by an independent edge prediction method that exploits the RNA-KG topological characteristics. Analogous behavior has been observed for the other views (not reported here for lack of space).

A potential reason for errors in predictions, particularly for triples included in RNA-KG but showing a probability lower than 0.5, might be the presence of relationships between miRNAs and overrepresented disease categories in RNA-KG (e.g., miRNA-cancer relationships like prostate cancer, pancreatic cancer, sarcoma, colorectal cancer, breast cancer, and hepatocellular cancer). The high frequency of miRNA connections to these diseases in RNA-KG might lead to a situation where the model, being trained on numerous miRNA links to these prevalent diseases, struggles to distinguish specific disease associations accurately. Consequently, it becomes challenging for the ML algorithm to determine whether a miRNA is genuinely involved in a specific disease. To mitigate these issues in future work, we would like to incorporate additional features for distinguishing the different miRNA-disease relationships, like specific biological pathways and molecular features that characterize specific diseases.

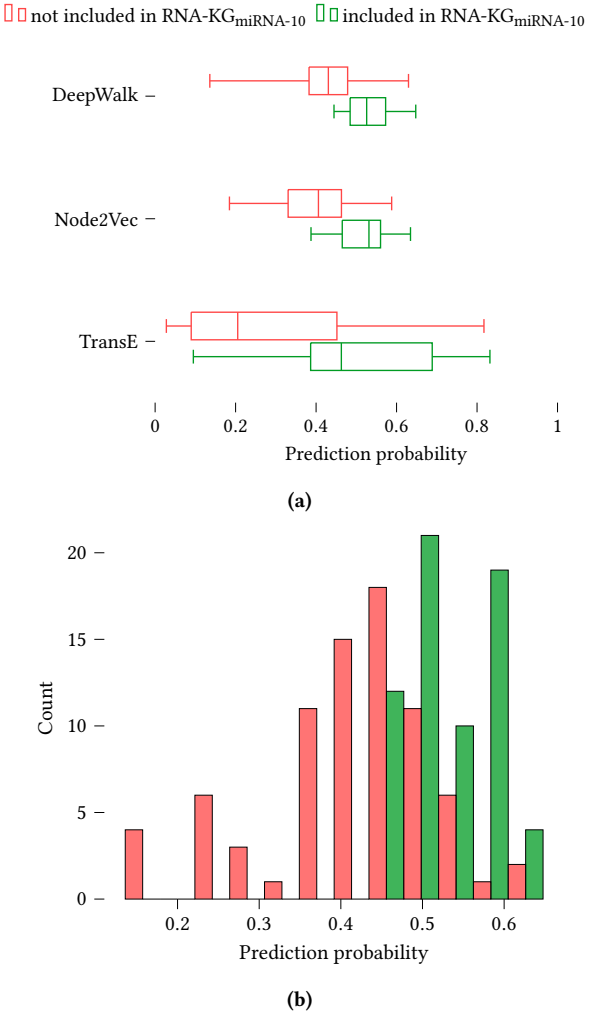


Figure 9: (a) Probability distribution of the miRNA-disease edges extracted by SPIRES and validated according to different prediction models on the RNA-$KG_{miRNA\text{-}10}$ view. 'Green'/'red' bars represent triples 'included'/'not included' in the RNA-KG view. (b) Histogram of the triples predicted by SPIRES on the RNA-$KG_{miRNA\text{-}10}$ view according to the prediction probabilities obtained by deepWalk.

5 CONCLUDING REMARKS

In this paper we have described the initial steps in the design and development of the SPIREX system for the extraction of meaningful triples from scientific papers that exploit RNA-KG as a gold standard for checking the plausibility of the extracted triples. The initial experimental results are encouraging of the effectiveness of the proposed tool. At the current stage, we have used basic link prediction models for assessing the relationship's plausibility according to the KG's current state.

In future work, a much more accurate evaluation strategy that also considers the performance of one-class approaches and the evaluation of domain experts will be investigated. Moreover, we wish to develop a web environment for supporting expert curators in the extraction of meaningful facts from the scientific literature by exploiting the SPIREX approach and thus reducing the user effort in validating the extracted knowledge. In this direction, we will also evaluate the usability of the developed system by enrolling expert curators in the field. Finally, we wish to verify the application of the proposed approach also to other biomedical contexts that exploit different KGs and the use of RAG techniques [24] to improve the quality of the RE approach in the same spirit of [35].

ACKNOWLEDGMENTS

This research was in part supported by the "National Center for Gene Therapy and Drugs based on RNA Technology", PNRR-NextGeneration EU program [G43C22001320007] and in part by the MUSA - Multilayered Urban Sustainability Action - Project, funded by the PNRR-NextGeneration EU program ([G43C22001370007], Code ECS00000037).

REFERENCES

[1] Lada A Adamic and Eytan Adar. 2003. Friends and neighbors on the Web. *Social Networks* 25, 3 (2003), 211–230. https://doi.org/10.1016/S0378-8733(03)00009-1

[2] Michele Banko, Michael J. Cafarella, Stephen Soderland, Matt Broadhead, and Oren Etzioni. 2007. Open information extraction from the web. In *Proc. of 20th Int'l Conf. on Artifical Intelligence*. IJCAI, Hyderabad, India, 2670–2676.

[3] Rishi Bommasani et al. 2021. On the Opportunities and Risks of Foundation Models. *CoRR* abs/2108.07258 (2021).

[4] Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. 2011. Learning structured embeddings of knowledge bases. In *Proc. of AAAI conference on artificial intelligence*, Vol. 25. AAAI, San Francisco, California, 301–306.

[5] Luca Cappelletti et al. 2023. GRAPE for fast and scalable graph processing and random-walk-based embedding. *Nature Computational Science* 3, 6 (June 2023), 552–568. https://doi.org/10.1038/s43588-023-00465-8

[6] J Harry Caufield et al. 2024. Structured Prompt Interrogation and Recursive Extraction of Semantics (SPIRES): a method for populating knowledge bases using zero-shot learning. *Bioinformatics* 40, 3 (02 2024), btae104. https://doi.org/10.1093/bioinformatics/btae104

[7] Emanuele Cavalleri et al. 2023. A Meta-Graph for the Construction of an RNA-Centered Knowledge Graph. In *Bioinformatics and Biomedical Engineering*. Springer Nature Switzerland, Cham, 165–180. https://doi.org/10.1007/978-3-031-34953-9_13

[8] Emanuele Cavalleri et al. 2023. RNA-KG: An ontology-based knowledge graph for representing interactions involving RNA molecules. arXiv:2312.00183 [cs.CE]

[9] Emanuele Cavalleri and Marco Mesiti. 2024. On the extraction of meaningful RNA interactions from Scientific Publications through LLMs and SPIRES. In: 8th Int'l workshop on Data Analytics solutions for Real-LIfe APplications..

[10] Jia Chen et al. 2022. RNADisease v4.0: an updated resource of RNA-associated diseases, providing RNA-disease analysis, enrichment and prediction. *Nucleic Acids Research* 51, D1 (2022), D1397–D1404. https://doi.org/10.1093/nar/gkac814

[11] Kartik Detroja, C.K. Bhensdadia, and Brijesh S. Bhatt. 2023. A survey on Relation Extraction. *Intelligent Systems with Applications* 19 (2023), 200244. https://doi.org/10.1016/j.iswa.2023.200244

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs.CL]

[13] Allyson Ettinger. 2020. What BERT Is Not: Lessons from a New Suite of Psycholinguistic Diagnostics for Language Models. *Transactions of the Association for Computational Linguistics* 8 (2020), 34–48. https://doi.org/10.1162/tacl_a_00298

[14] Oren Etzioni et al. 2004. Web-scale information extraction in knowitall: (preliminary results). In *Proc. of 13th Int'l Conf. on World Wide Web*. ACM, New York NY USA, 100–110. https://doi.org/10.1145/988672.988687

[15] Charles C Fries. 1954. Meaning and linguistic analysis. *Language* 30, 1 (1954), 57–68.

[16] Aditya Grover and Jure Leskovec. 2016. Node2vec: Scalable Feature Learning for Networks. In *Proc. of the 22nd ACM SIGKDD Intl Conf. on Knowledge Discovery and Data Mining*. ACM, San Francisco, California, USA, 855–864. https://doi.org/10.1145/2939672.2939754

[17] Tudor Groza et al. 2024. An evaluation of GPT models for phenotype concept recognition. *BMC Medical Informatics and Decision Making* 24, 1 (2024), 30. https://doi.org/10.1186/s12911-024-02439-w

[18] William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *Proc. of 31st Int'l Conf. on Neural Information Processing Systems*. NIPS, Long Beach, CA, USA, 1025–1035.

[19] Zellig S Harris. 1954. Distributional structure. *Word* 10, 2-3 (1954), 146–162.

[20] Pere-Lluís Huguet Cabot and Roberto Navigli. 2021. REBEL: Relation Extraction By End-to-end Language generation. In *Findings of the Association for Computational Linguistics*. ACL, Punta Cana, Dominican Republic, 2370–2381.

[21] Rebecca Jackson et al. 2021. OBO Foundry in 2021: operationalizing open data principles to evaluate ontologies. *Database* 2021 (10 2021). https://doi.org/10.1093/database/baab069

[22] Ziwei Ji et al. 2023. Survey of Hallucination in Natural Language Generation. *Comput. Surveys* 55, 12 (March 2023), 1–38. https://doi.org/10.1145/3571730

[23] Thomas Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of ICLR'17*. ICLR, Toulon, France.

[24] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems* 33 (2020), 9459–9474.

[25] Mausam, Michael Schmitz, Stephen Soderland, Robert Bart, and Oren Etzioni. 2012. Open Language Learning for Information Extraction. In *Proc. of 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. ACM, Jeju Island Korea, 523–534.

[26] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and Their Compositionality. In *Proc. of 26th Int'l Conf. on Neural Information Processing Systems*. NIPS, Harrahs and Harveys, Lake Tahoe, 3111–3119.

[27] Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proc. of Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*. ACL, Suntec, Singapore, 1003–1011.

[28] Sierra Moxon et al. 2021. The Linked Data Modeling Language (LinkML): A General-Purpose Data Modeling Framework Grounded in Machine-Readable Semantics. In *Int'l Conf. on Biomedical Ontologies*. ICBO, Bozen-Bolzano, Italy. https://ceur-ws.org/Vol-3073/paper24.pdf

[29] Chris Mungall et al. 2023. INCATools/ontology-access-kit: v0.5.24. https://doi.org/10.5281/zenodo.10277632

[30] Tapas Nayak and Hwee Tou Ng. 2019. Effective Modeling of Encoder-Decoder Architecture for Joint Entity and Relation Extraction. In *AAAI Conference on Artificial Intelligence*. AAAI, New York, USA. https://api.semanticscholar.org/CorpusID:208248243

[31] Sachin Pawar, Girish K. Palshikar, and Pushpak Bhattacharyya. 2017. Relation Extraction : A Survey. arXiv:1712.05191 [cs.CL]

[32] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. DeepWalk: Online Learning of Social Representations. In *Proc. of 20th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. ACM, New York USA, 701–710. https://doi.org/10.1145/2623330.2623732

[33] Pengda Qin, Weiran Xu, and William Yang Wang. 2018. Robust Distant Supervision Relation Extraction via Deep Reinforcement Learning. In *Proc. of 56th Annual Meeting of the Association for Computational Linguistics*. ACL, Melbourne, Australia, 2137–2147. https://doi.org/10.18653/v1/P18-1199

[34] Chris Quirk and Hoifung Poon. 2017. Distant Supervision for Relation Extraction beyond the Sentence Boundary. In *Proc. of 15th Conference of the European Chapter of the Association for Computational Linguistics*. ACL, Valencia, Spain, 1171–1182.

[35] Darya Shlyk, Tudor Groza, Stefano Montanelli, Emanuele Cavalleri, and Marco Mesiti. 2024. REAL: A Retrieval-Augmented Entity Linking Approach for Biomedical Concept Recognition. In *The 23rd ACL workshop BioNLP*. ACL, Bangkok, Thailand.

[36] Hugo Touvron et al. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv:2307.09288 [cs.CL]

[37] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. 2017. Graph attention networks. *stat* 1050, 20 (2017), 10–48550.

[38] Somin Wadhwa, Silvio Amir, and Byron C. Wallace. 2023. Revisiting Relation Extraction in the era of Large Language Models. In *Proc. of Int'l Conf. of the Association for Computational Linguistics*, Vol. 2023. ACL, Toronto, Canada, 15566–15589. https://doi.org/10.18653/v1/2023.acl-long.868

[39] Jiajia Wang et al. 2021. piRBase: integrating piRNA annotation in all aspects. *Nucleic Acids Research* 50, D1 (Dec. 2021), D265–D272. https://doi.org/10.1093/nar/gkab1012

[40] Bishan Yang, Wen tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. In *Int'l Conf. on Learning Representations*. ICLR, San Diego, CA, USA. https://api.semanticscholar.org/CorpusID:2768038

[41] Carl Yang, Yuxin Xiao, Yu Zhang, Yizhou Sun, and Jiawei Han. 2022. Heterogeneous Network Representation Learning: A Unified Framework With Survey and Benchmark. *IEEE Transactions on Knowledge and Data Engineering* 34, 10 (Oct. 2022), 4854–4873. https://doi.org/10.1109/tkde.2020.3045924

[42] Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J Kim. 2019. Graph transformer networks. *Advances in neural information processing systems* 32 (2019).

[43] Xiangrong Zeng, Daojian Zeng, Shizhu He, Kang Liu, and Jun Zhao. 2018. Extracting Relational Facts by an End-to-End Neural Model with Copy Mechanism. In *Proc. of the 56th Annual Meeting of the Association for Computational Linguistics*. ACL, Melbourne, Australia, 506–514. https://doi.org/10.18653/v1/P18-1047

[44] Muhan Zhang and Yixin Chen. 2018. Link prediction based on graph neural networks. In *Proc. of 32nd Int'l Conf. on Neural Information Processing Systems*. NIPS, Montréal CANADA, 5171–5181.