Explainable representations for relation prediction in knowledge graphs

Rita T. Sousa[1] , **Sara Silva**[1] , **Catia Pesquita**[1]

[1]LASIGE, Faculdade de Ciências da Universidade de Lisboa

{risousa,sgsilva,clpesquita}@ciencias.ulisboa.com

Abstract

Knowledge graphs represent real-world entities and their relations in a semantically-rich structure supported by ontologies. Exploring this data with machine learning methods often relies on knowledge graph embeddings, which produce latent representations of entities that preserve structural and local graph neighbourhood properties, but sacrifice explainability. However, in tasks such as link or relation prediction, understanding which specific features better explain a relation is crucial to support complex or critical applications.

We propose SEEK, a novel approach for explainable representations to support relation prediction in knowledge graphs. It is based on identifying relevant shared semantic aspects (i.e., subgraphs) between entities and learning representations for each subgraph, producing a multi-faceted and explainable representation.

We evaluate SEEK on two real-world highly complex relation prediction tasks: protein-protein interaction prediction and gene-disease association prediction. Our extensive analysis using established benchmarks demonstrates that SEEK achieves significantly better performance than standard learning representation methods while identifying both sufficient and necessary explanations based on shared semantic aspects.

1   Introduction

Knowledge Graphs (KGs) (Ehrlinger and Wöß 2016) are a representation of factual information about entities in the real world and how they relate to each other, having been widely used to support various applications including machine learning (ML) (Hogan et al. 2021). Particularly, in scientific domains, KGs have become highly relevant because they allow for the description and linking of information about entities based on ontologies (Staab and Studer 2010), allowing the description of complex natural phenomena that are not easily captured in mathematical form (Nicholson and Greene 2020).

In recent years, KG embedding methods (Wang et al. 2017) have become increasingly popular to bridge the gap between the complex representations a KG affords and the vectorial representations most ML methods take as input, since they map KGs into low-dimensional spaces preserving syntactic and structural properties. KG embeddings are popularly employed in link prediction via a scoring function or as features for supervised learning (Portisch, Heist,

and Paulheim 2022). However, this represents a significant trade-off: KG embeddings sacrifice the full and rich interpretability offered by KGs, especially when structured by rich ontologies, for the more simple to process latent representations (Palmonari and Minervini 2020). The effectiveness and usefulness of KG embeddings approach hinges on the crucial assumption that KG embeddings serve as semantically meaningful representations of the underlying entities. To validate such an assumption, KG embedding methods would need to be explainable (i.e., they would need to afford a human-understandable description of the logic, behavior or factors that influence the representation learning process), but in the vast majority of cases they are not. This is a fundamental requirement to ensure the scientific validity of KG embeddings, or any artificial intelligence (AI) method, as a tool that can be used to uncover new knowledge, help understand the mechanisms underlying natural phenomena, and distinguish meaningful predictions from spurious correlations (Barredo Arrieta et al. 2020).

In this work, we focus specifically on the problem of predicting a relation between KG entities that is not defined in the KG. Predicting relations such as protein-protein interactions (PPI) or gene-disease associations (GDA) by exploring KGs and ontologies has been the focus of extensive research in the biomedical domain. Both algorithmic (Zhang and Tang 2016; Hoehndorf, Schofield, and Gkoutos 2011; Asif et al. 2018) and ML approaches (Kulmanov et al. 2021) have been employed to achieve this with success, with KG embeddings particularly excelling at the task (Chen, Wang, and Hu 2019; Ieremie, Ewing, and Niranjan 2022; Alshahrani et al. 2017). However, understanding the nature of these relations requires discerning which aspects of the KG have the most influence on a prediction. This empowers users not only in assessing the reliability of the model itself but also in potentially elucidating the phenomena underlying the relation. For example, if we were to explain the interaction between the proteins *Protransforming growth factor α* and *Disks large homolog 2*, generating an explanation based on the fact that they both perform the very specific function *MAPK cascade*, we would likely increase trust as well as highlight a relevant aspect for interaction. In contrast, a very general explanation, such as the fact that both proteins are present in the *plasma membrane* would contribute to neither.

We propose SEEK (Shared Explainable Embeddings for

Knowledge graphs), a novel method for generating explainable KG embeddings that represent entity pairs for relation prediction. The intuition behind this is that an entity pair can be represented by combining embeddings that represent each of their shared semantic aspects, rather than simply combining their respective embeddings. This technique explores the rich semantics of the ontology to identify the shared semantic aspects between related entities based on computing their disjoint common ancestors. Then, these pair embeddings are used to train a supervised ML model for relation prediction. SEEK is fundamentally different from link prediction methods since it produces representations of pairs of entities based on shared semantic aspects, whereas link prediction methods rely on learning representations of individual KG entities and apply a scoring function to estimate the likelihood of triples.

Given a prediction, our method explains it by computing the importance of each shared semantic aspect in identifying it. Inspired by (Watson et al. 2021; Rossi et al. 2022a) we consider that an explanation includes two complementary views: the set of semantic aspects that, if absent from an entity pair, would render the model incapable of generating that prediction (i.e., necessary explanations); the set of semantic aspects that, if shared by any entity pair, would prompt the model to produce that prediction (sufficient explanations). Since SEEK explains specific predictions rather than the global mechanism of the model, it consequently falls under the category of local post-hoc explanation methods as proposed by (Guidotti et al. 2018).

We evaluate the effectiveness of SEEK in two different tasks, PPI prediction and GDA prediction. Predicting PPI is a crucial task in molecular biology (Li et al. 2021; Hu et al. 2021), and several KG embedding-based methods have been employed to tackle it (Kulmanov et al. 2019; Smaili, Gao, and Hoehndorf 2019; Kulmanov et al. 2021; Kulmanov et al. 2019; Xiong et al. 2022). Due to the high costs and challenges involved in experimentally determining PPI, computational methods can be used to identify protein pairs that are likely to interact, which are subsequently validated through experimental assays rendering the process more efficient. Likewise, predicting the relation between genes and diseases is essential to understand disease mechanisms and identifying potential biomarkers or therapeutic targets (Eilbeck, Quinlan, and Yandell 2017). Once again, computational approaches to identify the most promising associations to be further validated are commonly employed, with recent approaches applying KG embedding methods (Alshahrani et al. 2017; Smaili, Gao, and Hoehndorf 2019; Nunes, Sousa, and Pesquita 2021). However, opaque methods such as KG embeddings are unable to provide explanations behind each prediction. Explanatory mechanisms can elucidate the potential mechanisms behind the predicted relation, which can be helpful to determine the type of experimental procedure that should be applied to confirm the precited relation but also to identify data biases that can result in misclassification and should be grounds to discard the candidate pair. Our extensive experiments show that our method produces useful explanations besides improving performance over state-of-the-art embedding methods.

Our main contributions are the following:

- We propose SEEK, a novel method for generating explainable KG embeddings that represent entity pairs for relation prediction.

- We develop extensions of popular KG embedding methods implementing SEEK.

- We design explanation methods that quantify the importance of specific KG semantic aspects for specific relation predictions.

- We report extensive experimental results demonstrating that SEEK is able to produce effective explanations for relation prediction as well as generally improving predictive performance on multiple models and biomedical datasets.

2   Problem overview

We define a KG as a labeled directed graph $KG = (V, E, R)$ where V is the set of vertices that represent entities, R is the set of relations and E is the set of edges that connect vertices through relations. Our particular focus is on ontology-rich KGs with ontologies defined using Web Ontology Language (OWL) (Grau et al. 2008) since biomedical ontologies are typically developed in OWL or have an OWL version. These are frequently found in scientific fields like biomedicine. In these KGs, ontologies are typically used to describe individual instances, while the instances themselves are usually flat with no connections between them. Consequently, there will be two types of vertices in the KG: those that correspond to individual entities and those that correspond to ontology classes, as well as two types of edges: those that relate ontology classes to each other, and those that link individuals to the classes that describe them. For example, using OWL 2 whose constructs correspond to SROIQ(D), we can indicate that a protein P carries out a function F described in the Gene Ontology (GO) by declaring the axiom $P \sqsubseteq \exists hasFunction.F$. KG embedding methods are then able to learn representations of biomedical entities by exploring the links that connect an entity to the ontology classes that describe it, as well as the structure of the ontology itself.

Our objective is to learn a relation between two KG entities, a pair, when the relation itself is not explicitly defined in the KG, using embeddings as inputs for a supervised ML algorithm. This is a fundamentally distinct task from link prediction, where the training set relations are part of the KG. To tackle this relation prediction task, common approaches typically employ three steps: (1) generate embeddings for each entity in the KG; (2) aggregate the embeddings of each entity in a pair into a single representation; (3) use these aggregated representations as input to a supervised learning algorithm to learn a relation prediction model (Sousa, Silva, and Pesquita 2021; Celebi et al. 2019). This generates non-explainable predictions since KG embeddings are, of course, non-explainable, as each dimension does not represent any particular meaning, which poses a serious limitation to the use of KG embeddings in a scientific setting.

Moreover, this particular formulation results in two over-simplifications, which may limit its effectiveness and usefulness. Firstly, it relies on the aggregation of individual embeddings to represent a pair of entities, instead of directly learning an embedding that represents the pair. One should clarify that simply representing the pair as yet another entity on the KG would not be a viable solution, as it would limit the applicability of the approach to pairs seen at representation learning time and thus fail to generalize to novel pairs. Secondly, it focuses on creating an overall representation of each entity, rather than capturing the different semantic aspects that may contribute to the relation we aim to predict. In large KGs, it is not uncommon for entities to be described according to multiple semantic aspects, but only a few may be relevant for the prediction of a particular relation. In a previous study (Sousa, Silva, and Pesquita 2020), it was demonstrated that not all branches of the GO are equally important for predicting PPIs.

The problem we tackle is then two-fold: (1) to generate latent representations that represent an entity pair directly and (2) to generate latent representations that are amenable to explanation and can capture the relevant semantic aspects for relation prediction.

3 Related work

3.1 Knowledge graph embeddings

KG embedding methods represent KG entities and their relations in a lower-dimensional space preserving the KG semantic information as much as possible. These embeddings have been employed as features in a variety of downstream tasks, such as link prediction, triple classification, or entity typing. KG embeddings have been successfully employed in a number of scientific applications, with particular success in the life sciences (Mohamed, Nounu, and Nováček 2021; Kulmanov et al. 2021; Chen, Wang, and Hu 2019; Ieremie, Ewing, and Niranjan 2022). There are several types of KG embeddings, including translational models, semantic matching models, or random walk-based KG embedding approaches.

Translational methods use distance-based scoring functions. TransE (Bordes et al. 2013) is a well-known approach that assumes the vector of the head entity plus the relation vector should be close to the vector of the tail entity if a relation holds between two entities. However, TransE only handles one-to-one relationships. To overcome this limitation, TransH (Wang et al. 2014) introduces a unique relation-specific hyperplane for each relationship.

Semantic matching models rely on scoring functions based on similarity, which can represent the underlying meaning of entities and relationships in vector spaces. An example of such a method is DistMult (Yang et al. 2015), which uses tensor factorization to create vector embeddings for entities and diagonal matrices for relationships.

Random walk-based embedding techniques perform walks in the graph to produce a corpus of sequences that is given as input to a neural language model (Mikolov et al. 2013) to learn a latent low-dimensional representation of each entity within the corpus of sequences. RDF2Vec (Ris-toski and Paulheim 2016) is used to learn embeddings over RDF graphs.

More recently, KG embedding approaches that tailor representations by considering specific aspects of a KG have been proposed. EL (Kulmanov et al. 2019) and BoxEL (Xiong et al. 2022) embeddings are geometric approaches that consider the logical structure of the ontology. OWL2Vec* (Chen et al. 2021) is very similar to RDF2Vec, but it was designed to learn embeddings of OWL ontologies, which are used to represent knowledge in a more expressive and formal way than RDF graphs. OPA2Vec (Smaili, Gao, and Hoehndorf 2019) considers the lexical portion of the KG, specifically the labels of entities, when generating triples.

3.2 Explainable artificial intelligence techniques

The scientific community has long recognized the potential of AI as a tool for scientific discovery, with ML, pattern mining, and reasoning playing crucial roles in several steps of the scientific process (Mjolsness and DeCoste 2001). Despite this, the vast majority of scientific projects that utilize AI do not prioritize explainability (Roscher et al. 2020). In the biomedical domain, the complexity of both the data and the natural phenomena under study emphasizes the importance of domain knowledge to support explainability (Holzinger et al. 2017). A knowledge-enabled explainable AI (XAI) system includes a representation of the domain knowledge specific to the application field. This knowledge is explored for generating explanations that are both comprehensible to users and contextually aware of the mechanistic functioning of the AI system and the knowledge it employs (Chari et al. 2020).

XAI aims to address several key objectives, including promoting algorithmic fairness, detecting potential biases or issues in training data, ensuring that algorithms function as intended, and bridging the gap between the ML community and other scientific disciplines (Gilpin et al. 2018). According to (Barredo Arrieta et al. 2020), XAI approaches can be classified into two types: models that are transparent by design, such as decision trees, linear models, and genetic programming models (Mei et al. 2022), or post-hoc explainability techniques that are used to improve the interpretability of models that are not transparent by design. Post-hoc explainability techniques can be categorized as either model-specific or model-agnostic if they are applicable to any ML model. Post-hoc techniques may include visual explanations, explanations by example, explanations by simplification, or feature relevance explanations.

KG embeddings are not explainable, and there is no widely accepted methodology to effectively explain the predictions of KG embeddings (Palmonari and Minervini 2020). CRIAGE (Pezeshkpour, Tian, and Singh 2019) and Kelpie (Rossi et al. 2022b) have made striding efforts towards explaining link prediction based on KG embeddings by identifying the fact to add into or remove from the KG that affects the prediction for a target fact. Betz *et al.* (Betz, Meilicke, and Stuckenschmidt 2022) also propose a post hoc method that uses adversarial attacks on KG embedding models to identify triples that serve as logical explanations

specific predictions. These works differ fundamentally from ours by focusing on single facts about each entity, whereas we focus on shared aspects between entity pairs. Additionally, all of these works face the computational challenge posed by having to retrain the KG model after removing a single fact to explain each prediction, and devise heuristic approaches to minimize this aspect. Our approach does not require retraining the model. Instead, we generate explanations by identifying shared semantic aspects and making predictions with the trained model. ExCut (Gad-Elrab et al. 2020) is another approach that uses KG embeddings to identify clusters of entities and then combines it with rule-mining methods to learn interpretable labels.

4 Methods

4.1 Overview

SEEK is a novel approach that generates explainable vector representations of KG entity pairs to support relation prediction tasks with minimal loss in performance.

Figure 1 shows an overview of the SEEK approach. In the first step, the KG is transformed into an RDF graph, which facilitates the subsequent processing. Representations for each ontology class are then learned using a KG embedding method. Notably, SEEK is agnostic to the specific KG embedding method employed and can accommodate a broad range of techniques.

The second step is concerned with identifying the shared semantic aspects between the entities of the pair, which are determined by computing the disjoint common ancestors of all classes related to them. The identification of these semantic aspects is essential for the subsequent generation of accurate and meaningful explanations. Having identified the relevant semantic aspects, the final representations of entity pairs are then generated by aggregating the embeddings of the shared semantic aspects.

In the third and final step, supervised learning methods are employed to learn a relation prediction model taking as input the pair embeddings. This model is then used to generate explanations by adopting a perturbation-inspired approach where the contribution of each semantic aspect to the final prediction is assessed in terms of its sufficiency and necessity. The necessary explanations provide insights into the semantic aspects that are necessary for a particular decision to be made, while the sufficient explanations reveal the aspects that are sufficient to support a particular decision. These explanations enable a more thorough understanding of predicted relations and which KG aspects influence it and can be invaluable in identifying potential biases or errors.

4.2 Generating the RDF graph and learning embeddings

Ontology-rich KGs are typically defined in OWL. However, the majority of graph processing and analysis tools require RDF graphs. Therefore, the initial step is to convert the KG into an RDF graph following the guidelines provided by the W3C[1]. The conversion process involves transforming simple axioms directly into RDF triples, such as subsumption
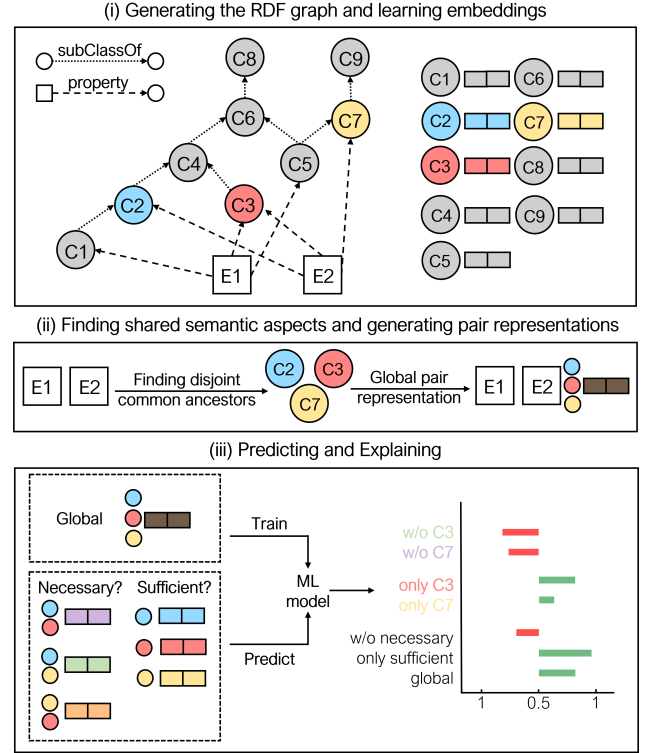
---

[1]https://www.w3.org/TR/owl2-mapping-to-rdf/

Figure 1: Overview of SEEK with the main steps: (i) generating the RDF graph and learning embeddings (ii) finding shared semantic aspects and generating pair representations (iii) predicting and explaining.

axioms or data and annotation properties associated with an entity. Multiple triples are created for more complex axioms involving class expressions, which usually require blank nodes. The relations between entities and the ontology classes describing them are usually stored in annotation files in the biomedical domain. These annotations are processed into object properties. After conversion, the nodes in the RDF graph represent ontology classes or individuals, and the edges represent named relations. Finally, we employ a KG embedding method to learn latent representations of all the ontology classes in the KG.

4.3 Finding shared semantic aspects and generating pair representations

To generate a representation for an entity pair we explore the concept of semantic aspect (i.e., a subgraph of the KG that captures a specific perspective of the domain). We propose to represent a pair of KG entities by the set of semantic aspects they share, unfolding their relationship into different dimensions each based on a shared aspect. We define the shared semantic aspects as the set of disjoint common ancestors computed over the set of classes that describe each entity.

Let us take two entities e_1 and e_2 and their set of linked classes C_1, C_2. To compute the set of disjoint shared as-
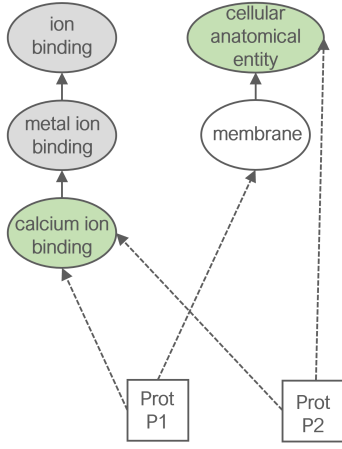
Figure 2: A GO KG subgraph to represent the shared semantic aspects of two entities. Green classes represent the disjoint common ancestors of proteins P1 and P2. Grey classes represent the remaining common ancestors.

pects, we first compute the disjoint common ancestors of C_1 and C_2. Following (Couto and Silva 2011), we define that a_1 and a_2 are disjoint common ancestors of a class c if $c \sqsubseteq a_1$, $c \sqsubseteq a_2$, $a_1 \not\sqsubseteq a_2$ and $a_2 \not\sqsubseteq a_1$. We first compute C_a, the set of common ancestors between the two sets C_1 and C_2, and then filter this set to include only the disjoint common ancestors, each of which represents a shared semantic aspect. The shared semantic aspects of two sets only include indirect common ancestors if they do not subsume other common ancestors. Considering the example in Figure 2, the shared semantic aspects of proteins P1 and P2 correspond to *calcium ion binding* and *cellular anatomical entity*.

To represent an entity pair we take the embeddings of each class in the shared semantic aspects set and aggregate them using simple operators such as the Hadamard product, the sum, the average or the L1-norm.

4.4 Predicting and explaining

After obtaining the vector representations, we use supervised ML algorithms to learn relation prediction models and ultimately produce explanations for predicted relations. First, we train our model using the global representation of the pair, generated by aggregating all shared semantic aspect embeddings. Then, for each prediction we want to explain, we generate multiple representations that differ by the presence or absence of a semantic aspect. To understand which semantic aspects are necessary for the prediction, we generate representations that remove each aspect in turn (see Algorithm 1), whereas to understand which aspects are sufficient for the prediction, we generate representations that include a single aspect (see Algorithm 2). A semantic aspect is considered necessary for a prediction if the predicted class changes when it is removed. Likewise, a semantic aspect is considered sufficient for a prediction if the predicted class does not change when it is the only aspect considered.

We define an explanation as the set of the most relevant shared semantic aspects identified as necessary or sufficient. A necessary explanation is a shared semantic aspect that, when removed from the pair representation, causes the classifier to change its prediction. A sufficient explanation is a shared semantic aspect that, when used alone to represent a pair, causes the classifier to maintain its prediction. A relation may be explained by multiple necessary and sufficient explanations.

This approach is similar to how saliency XAI methods inject perturbations in the feature space to capture the importance of features. However, it addresses a significant challenge that perturbation or modification-based methods face, including those that aim to explain KG embeddings (Pezeshkpour, Tian, and Singh 2019; Rossi et al. 2022b), which is the need to relearn representations after performing the modification to the data. SEEK avoids this hurdle since it is based on composite representations of ontology classes, which are easy to modify and do not require retraining since the ontology itself is never altered, so the learned class embeddings remain fixed.

The final explanation can be represented as a chart where sufficient and necessary shared semantic aspects are presented alongside their impact on the prediction. In Figure 1, both C3 and C7 are necessary to support the prediction since, without either of them, the prediction value changes when compared to the prediction obtained for the global representation. C3 is also a sufficient aspect since it can single-handedly produce a prediction that agrees with the global one. The explanation can be further enriched with the prediction of the global approach, a prediction made with all sufficient shared semantic aspects, and a prediction made without any of the necessary shared semantic aspects, all predictions including their respective likelihood.

Algorithm 1 Generation of necessary explanations

Input: the entity pair (e_1, e_2);
the KG embedding model K;
the relation prediction model M;
Output: the set of disjoint shared aspects that are necessary for explaining the prediction

1: $N \leftarrow empty$
2: $D \leftarrow$ GET DISJOINT SHARED ASPECTS$((e_1, e_2))$
3: $E \leftarrow$ GET EMBEDDINGS(K, D)
4: $v \leftarrow$ AGGREGATE(E)
5: $p \leftarrow$ PREDICT(M, v)
6: **for** $d \in D$ **do**
7:     $e' \leftarrow E.delete(d)$
8:     $v' \leftarrow$ AGGREGATE(e')
9:     $p' \leftarrow$ PREDICT(M, v')
10:    **if** $p \neq p'$ **then**
11:        $N.append(d)$
     return N

Algorithm 2 Generation of sufficient explanations

Input: the entity pair (e_1, e_2);
the KG embedding model K;
the relation prediction model M;
Output: the set of disjoint shared aspects that are sufficient for explaining the prediction

1: $S \leftarrow empty$
2: $D \leftarrow$ GET DISJOINT SHARED ASPECTS$((e_1, e_2))$
3: $E \leftarrow$ GET EMBEDDINGS(K, d)
4: $v \leftarrow$ AGGREGATE(E)
5: $p \leftarrow$ PREDICT(M, v)
6: **for** $d \in D$ **do**
7:   $v' \leftarrow$ GET EMBEDDING(K, d)
8:   $p' \leftarrow$ PREDICT(M, v')
9:   **if** $p == p'$ **then**
10:     $S.append(d)$
    return S

5 Experimental Results

5.1 Experimental setup

We evaluate SEEK on two biomedical relation prediction tasks: predicting PPIs and predicting GDA. Both tasks are grounded on ontology-rich KGs, where PPI employs the GO and GDA is based on the Human Phenotype Ontology (HP). Additionally, prior studies have shown that different branches of these ontologies have varying impacts on achieving precise predictions (Sousa, Silva, and Pesquita 2020).

Our work targets relation prediction tasks cast as a classification task that takes as input entity pairs and a KG backboned by an ontology. Ontologies are arranged in a directed acyclic graph, where ontology classes are connected by subclass relations such that each class is more specific than its ancestor. Moreover, these relationships are transitive, meaning they inherit all ancestors to the root. The data used are described in the following sections.

Table 1: Statistics for each task regarding classes, nodes, and edges. Positive and negative pairs correspond to the number of positive and negative relations.

	PPI	GDA
Ontology classes	50422	15656
Literals and blank nodes	462874	443489
Instances	6738	4523
Annotations	349500	160009
Positive Pairs	23571	8189
Negative Pairs	23571	8189

Protein-protein interaction prediction The target relations to predict are obtained from the STRING database (Szklarczyk et al. 2020), one of the largest PPI databases that integrate physical interactions and functional associations between proteins from various sources. We filtered the protein pairs to include only pairs that met the following criteria: (i) each protein must be annotated with

the GO, (ii) interactions must be extracted from curated databases or experimentally determined, and (iii) interactions must have a confidence score above 0.950. The PPI dataset contains 23571 interacting protein pairs as well as 23571 negative pairs derived from random negative sampling of the same set of proteins.

The GO KG is used to describe proteins and is built by integrating the GO (Consortium 2021) and protein annotation data (Huntley et al. 2014). The GO defines a hierarchy of classes that describe protein functions and their relationships. It can be represented as a graph where nodes are GO classes and edges define relationships between them (e.g., is_a; $part\_of$; has_part; $regulates$; $negatively\_regulates$ and $positively\_regulates$), being the majority is_a relations. The three domains of GO (biological processes, molecular functions, and cellular components) are represented as separate root ontology classes since they do not share any common ancestor. A GO annotation is a statement about the function F of a protein P, and it is added in the KG as an assertion $\langle P, hasFunction, F \rangle$. In GO KG, nodes represent proteins or GO classes, and edges represent links between GO classes or annotations. Table 1 describes the statistics about PPI data.

Gene-disease association Prediction The target relations to predict are obtained from DisGeNET (Piñero et al. 2019). We follow the approach in (Nunes, Sousa, and Pesquita 2021), which excludes associations whose sources are used to create some of the ontology annotations. Moreover, each gene and disease must have at least one HP annotation. This resulted in a balanced dataset with a total of 16378 gene-disease pairs.

In this experiment, we employ the HP KG comprising the HP (Köhler et al. 2020) and HP annotation data to describe genes and diseases. HP characterizes the phenotypic abnormalities in human hereditary diseases concerning five semantic aspects, namely phenotypic abnormalities, mode of inheritance, clinical course, clinical modifier and frequency. Regarding the HP annotations, they link genes and diseases to HP classes and are added in the KG in the same fashion as in the PPI experiment. The statistics about GDA data are also shown in Table 1.

Models SEEK is independent of the KG embedding method and of the supervised ML algorithm. For our experiments, we implemented five representative KG embeddings covering translational, semantic matching and random walk-based methods: RDF2Vec (Ristoski and Paulheim 2016), OWL2Vec* (Chen et al. 2021), TransE (Bordes et al. 2013), TransH (Wang et al. 2014) and distMult (Yang et al. 2015). RDF2Vec and OWL2Vec* are path-based approaches adapted to RDF graphs that employ neural language models over random walks on the graph. TransE and TransH are translational distance embedding approaches that exploit distance-based scoring functions. distMult is a semantic matching approach that exploits similarity-based scoring functions.

To generate a pair representation, we use the average as the aggregation which ensures that the values of each dimension remain within the distribution. In the case of necessary

explanations, removing one similar semantic aspect will result in a very similar aggregated representation, revealing that the semantic brings little novel information for the prediction (since a similar semantic aspect is still considered). In the case of sufficient explanations, semantic aspects are evaluated independently.

As supervised ML algorithms, we employ two ensemble methods, Random Forest (RF) (Breiman 2001) and eXtreme Gradient Boosting (XGB) (Chen and Guestrin 2016), and a neural network-based method, Multilayer Perceptron (MLP) (Rumelhart, Hinton, and Williams 1986).

5.2 Results and Discussion

Performance evaluation   To assess our method, we compare the relation prediction performance of our pair representations against the state-of-the-art approach of entity vector aggregation using representative KG embedding methods, supervised ML algorithms and the Hadamard operator. We do not compare SEEK to other KG embedding explanation methods such as Kelpie or CRIAGE because they learn embeddings that target link prediction, whereas SEEK learns embeddings to serve as features for supervised ML. We evaluate the predictive performance of our approach against our baselines for each task using 10-fold cross-validation. For each partition, the precision (Pr), recall (Re) and weighted average f1-score (F1) are computed, and we report the median of the obtained scores (Table 2) and statistical significance of the observed differences.

The results demonstrate that SEEK outperforms the baseline in all cases but one for PPI prediction, while achieving similar or improved scores for GDA. Curiously, the performance of translational methods shows a marked improvement when using SEEK, likely due to the fact that these methods struggle with learning entity representations, but not ontology class representations.

To better understand the differences between the pair representations obtained using the baselines and the ones obtained using SEEK, we plot the RDF2Vec embeddings using t-SNE (Van der Maaten and Hinton 2008), a nonlinear dimensionality reduction technique that is particularly well-suited for visualizing high-dimensional data (Figure 3). These plots show that our pair representations decrease the overlap between positive and negative pairs and thus are likely to be capturing more meaningful representations.

Effectiveness of explanations   The effectiveness of the explanations is measured based on how predictive performance varies under two scenarios: when pairs are represented without the *necessary* shared semantic aspects; when pairs are represented by *sufficient* shared semantic aspects only. Table 3 presents the results obtained for the PPI and GDA tasks using the two best performing KG embedding methods.

In the necessary scenario, we extract the necessary explanations for all correctly predicted relations and produce an ablated representation that does not include any of the necessary shared semantic aspects. The performance variation, in terms of precision (Pr), recall (Re) and F1-score (F1), is measured as the difference in predictive performance between the global representation and the ablated represen-

(a) Baseline for PPI (b) SEEK for PPI
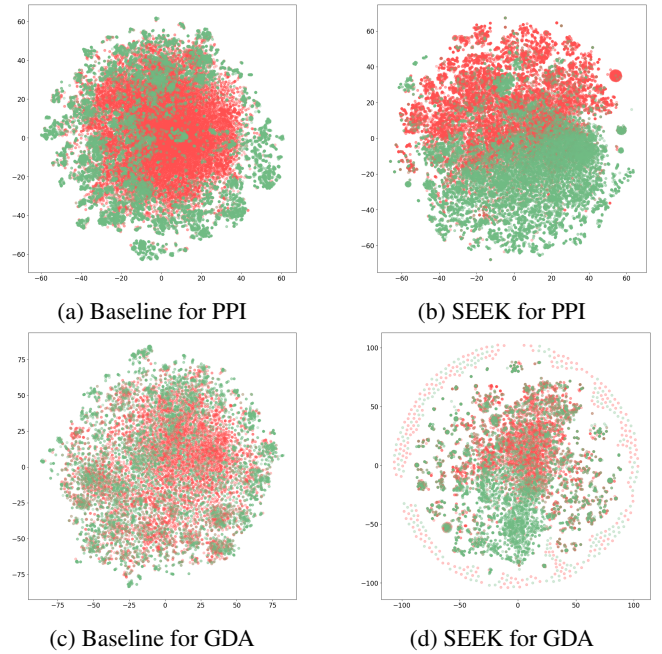
(c) Baseline for GDA (d) SEEK for GDA

Figure 3: t-SNE plots comparing SEEK to the baseline using RDF2Vec. Positive pairs in green and negative pairs in red.

tation. The more negative Δ Pr, Δ Re or Δ F1 are, the more effective are the necessary explanations.

In the sufficient scenario, we extract the sufficient explanations for all incorrectly predicted relations and produce an ablated representation that only includes the sufficient shared semantic aspects. The performance variation is also measured as the difference in predictive performance between the global representation and the ablated representation, but in this case the performance of the global representation is actually zero for all scores, since this is only applied to incorrectly predicted relations. A higher Δ value indicates increased effectiveness.

Explanation length   The lengths of the explanations, as measured by the number of shared semantic aspects that compose them, are presented in Tables 4 and 5. In both tasks, the length of necessary explanations is markedly lower than the length of sufficient explanations, highlighting that for many relations there are no necessary shared semantic aspects. When comparing the shown results to the original number of shared semantic aspects, 9.1 ($\pm$ 6.5) for PPI and 8.5 ($\pm$ 11.0) for GDA, we can verify that sufficient explanations amount to roughly 30% of shared semantic aspects. These sizes are congruent with the number of objects (7 $\pm$ 2) humans are able to hold in short-term memory according to cognitive studies (Miller 1956).

Examples of explanations   Table 6 presents explanations for four protein pairs chosen randomly from the PPI dataset. Each pair represents each of the four possible outcomes: a true positive, a false positive, a true negative, and a false negative.

Table 2: Medians of precision, recall, and weighted average f1-score (Pr, Re, F1) comparing our approach SEEK to the baseline when coupled with different supervised ML methods for PPI and GDA prediction. SEEK performance values are underlined when improvements are statistically significant with p-value < 0.05 for the Wilcoxon test against the baselines.

| | | PPI Prediction | | | | | | GDA Prediction | | | | | |
| | | Baseline | | | SEEK | | | Baseline | | | SEEK | | |
		Pr	Re	F1	Pr	Re	F1	Pr	Re	F1	Pr	Re	F1
RDF2Vec	XGB	0.905	0.917	0.910	0.920	0.910	0.915	0.736	0.708	0.724	0.772	0.626	0.719
	RF	0.921	0.881	0.902	0.922	0.892	0.910	0.783	0.673	0.740	0.787	0.625	0.723
	MLP	0.897	0.907	0.902	0.908	0.924	0.917	0.700	0.705	0.696	0.730	0.645	0.703
OWL2Vec*	XGB	0.890	0.881	0.888	0.933	0.925	0.929	0.700	0.664	0.688	0.780	0.647	0.728
	RF	0.913	0.832	0.875	0.922	0.915	0.919	0.730	0.618	0.690	0.780	0.662	0.737
	MLP	0.872	0.865	0.869	0.934	0.923	0.931	0.648	0.676	0.650	0.749	0.642	0.720
distMult	XGB	0.897	0.905	0.902	0.914	0.910	0.912	0.718	0.668	0.704	0.764	0.649	0.722
	RF	0.904	0.860	0.884	0.910	0.897	0.905	0.745	0.636	0.706	0.766	0.637	0.716
	MLP	0.894	0.894	0.896	0.881	0.895	0.888	0.731	0.681	0.715	0.768	0.589	0.698
TransE	XGB	0.642	0.613	0.638	0.914	0.912	0.913	0.526	0.509	0.524	0.755	0.650	0.721
	RF	0.590	0.542	0.583	0.908	0.900	0.905	0.505	0.474	0.502	0.765	0.640	0.719
	MLP	0.250	0.500	0.333	0.882	0.899	0.890	0.500	1.000	0.333	0.779	0.555	0.694
TransH	XGB	0.642	0.614	0.637	0.921	0.918	0.919	0.511	0.493	0.510	0.767	0.651	0.726
	RF	0.586	0.551	0.579	0.912	0.908	0.910	0.500	0.453	0.494	0.770	0.642	0.720
	MLP	0.250	0.500	0.333	0.915	0.920	0.920	0.000	0.000	0.333	0.735	0.665	0.711

Table 3: Explanation efectiveness measured based on the precision (Pr), recall (Re) and weighted average f1-score (F1) variation for PPI and GDA prediction.

| | | PPI Prediction | | | | | | GDA Prediction | | | | | |
| | | RDF2Vec | | | OWL2Vec* | | | RDF2Vec | | | OWL2Vec* | | |
		MLP	XGB	RF	MLP	XGB	RF	MLP	XGB	RF	MLP	XGB	RF
w/o necessary	ΔPr	-0.157	-0.109	-0.106	-0.095	-0.099	-0.089	-0.291	-0.296	-0.326	-0.265	-0.332	-0.269
	ΔRe	-0.137	-0.120	-0.153	-0.145	-0.131	-0.129	-0.329	-0.220	-0.277	-0.353	-0.208	-0.329
	ΔF1	-0.148	-0.113	-0.125	-0.117	-0.113	-0.107	-0.264	-0.225	-0.273	-0.270	-0.256	-0.260
only sufficient	ΔPr	0.932	1.000	0.973	0.981	1.000	0.988	0.957	0.969	0.893	0.921	0.986	0.917
	ΔRe	0.959	1.000	0.888	0.927	1.000	0.942	0.737	0.905	0.779	0.777	0.993	0.869
	ΔF1	0.950	1.000	0.945	0.954	1.000	0.967	0.898	0.964	0.896	0.885	0.993	0.925

Table 4: Explanation average length (Avg) and standard deviation (Std) for PPI prediction.

| | | RDF2Vec | | OWL2Vec* | |
		Avg	Std	Avg	Std
sufficient	MLP	5.6	3.9	5.3	3.5
	XGB	6.2	3.9	6.3	4.1
	RF	5.6	3.7	5.9	3.7
necessary	MLP	0.4	1.1	0.3	1.0
	XGB	0.4	1.1	0.3	1.0
	RF	0.4	1.3	0.3	1.1

Table 5: Explanation average length (Avg) and standard deviation (Std) for GDA prediction.

| | | RDF2Vev | | OWL2Vec* | |
		Avg	Std	Avg	Std
sufficient	MLP	5.6	7.1	5.5	7.8
	XGB	6.0	8.3	6.0	9.4
	RF	5.6	7.7	5.7	8.6
necessary	MLP	0.6	1.5	0.6	1.1
	XGB	0.5	1.3	0.5	1.2
	RF	0.7	1.7	0.7	1.4

The first pair in our analysis consists of paxillin[2] and integrin α-4[3]. There is strong evidence for their interaction (Han et al. 2001) since integrin α-4 binds tightly to paxillin, leading to increased cell migration and an altered cytoskeletal or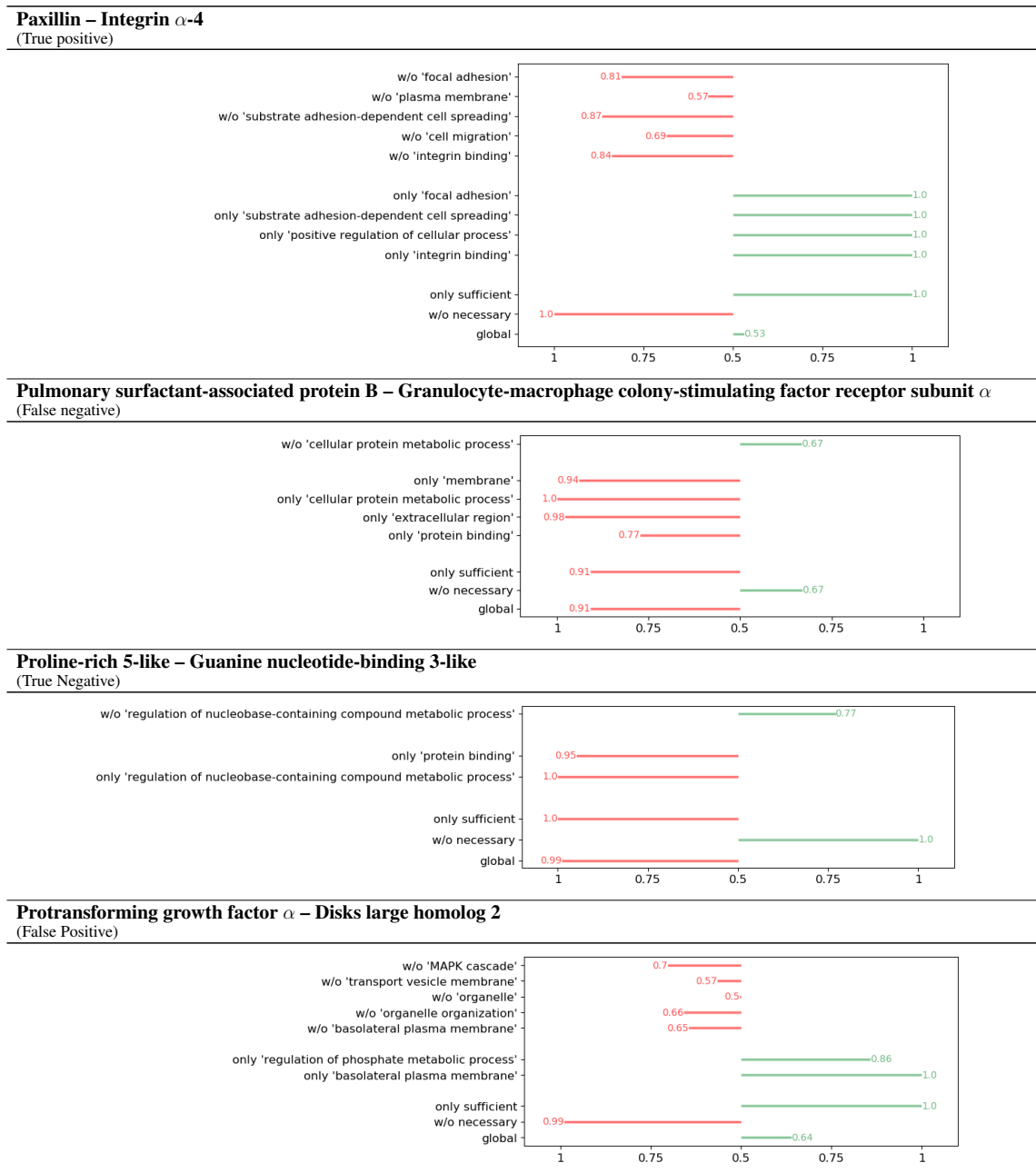ganization that results in reduced cell spreading. Our explanations identify several aspects that are necessary and/or sufficient to explain the interaction and that strongly correlate with the known evidence: focal adhesion, substrate adhesion-dependent cell spreading, cell migration and integrin binding.

---

[2]https://www.uniprot.org/uniprot/P49023

[3]https://www.uniprot.org/uniprot/P13612

Table 6: Explanations of PPI prediction models for four randomly selected pairs. For each pair, we provide a bar chart using different sets of disjoint common ancestors to represent the pair. On the x-axis, each bar represents the likelihood returned by the MLP model of the predicted class being correct. Classes are represented by colors (red for class 0 and green for class 1).

Paxillin – Integrin α-4
(True positive)

w/o 'focal adhesion'	0.81
w/o 'plasma membrane'	0.57
w/o 'substrate adhesion-dependent cell spreading'	0.87
w/o 'cell migration'	0.69
w/o 'integrin binding'	0.84
only 'focal adhesion'	1.0
only 'substrate adhesion-dependent cell spreading'	1.0
only 'positive regulation of cellular process'	1.0
only 'integrin binding'	1.0
only sufficient	1.0
w/o necessary	1.0
global	0.53

Pulmonary surfactant-associated protein B – Granulocyte-macrophage colony-stimulating factor receptor subunit α
(False negative)

w/o 'cellular protein metabolic process'	0.67
only 'membrane'	0.94
only 'cellular protein metabolic process'	1.0
only 'extracellular region'	0.98
only 'protein binding'	0.77
only sufficient	0.91
w/o necessary	0.67
global	0.91

Proline-rich 5-like – Guanine nucleotide-binding 3-like
(True Negative)

w/o 'regulation of nucleobase-containing compound metabolic process'	0.77
only 'protein binding'	0.95
only 'regulation of nucleobase-containing compound metabolic process'	1.0
only sufficient	1.0
w/o necessary	1.0
global	0.99

Protransforming growth factor α – Disks large homolog 2
(False Positive)

w/o 'MAPK cascade'	0.7
w/o 'transport vesicle membrane'	0.57
w/o 'organelle'	0.5
w/o 'organelle organization'	0.66
w/o 'basolateral plasma membrane'	0.65
only 'regulation of phosphate metabolic process'	0.86
only 'basolateral plasma membrane'	1.0
only sufficient	1.0
w/o necessary	0.99
global	0.64

The proteins Pulmonary surfactant-associated protein B[4] and ganulocyte-macrophage colony-stimulating factor receptor subunit α[5] make up the second pair. Although the proteins share some necessary and/or sufficient semantic aspects, they are very general; therefore, the model does not predict the interaction. However, according to the litera-

ture, they are likely involved in the same pulmonary disease (Trapnell, Whitsett, and Nakata 2003). Both proteins are poorly described under the GO, which can explain why the relation prediction model fails.

The third pair includes the proline-rich 5-like protein[6] and the guanine nucleotide-binding 3-like protein[7]. The model

predicts this as a negative pair, and the explanations confirm this, with the removal of the necessary shared semantic aspect resulting in a positive prediction. No interaction is known between these two proteins.

The Protransforming growth factor (TGF) α[8] and the Disks large homolog 2 (Dlg2)[9] compose the last pair and correspond to a false positive. The explanations highlight their participation in the MAPK cascade (central signaling pathways that regulate a wide variety of stimulated cellular processes, including proliferation, differentiation, apoptosis and stress response) as well as their co-location in the basolateral plasma membrane. It is intriguing to note that although there is no known interaction between these proteins, there is evidence of an interaction between highly similar proteins: TGF-β is regulated by Dlg5 and both proteins activate the MAPK cascade (Sezaki et al. 2013). We hypothesize this is not in fact a true negative pair but a still unknown PPI erroneously used as a negative example through random negative sampling.

6 Conclusion

Existing KG embedding methods are not explainable, which hinders their application in complex and critical domains. This is especially challenging in relation prediction, where understanding which KG semantic aspects are more relevant for a relationship between two KG entities can provide insightful knowledge about its mechanisms and help distinguish meaningful predictions from spurious correlations.

To address this challenge, we propose SEEK, a novel approach for learning and explaining representations of KG entity pairs based on their shared semantic space for relation prediction. Its explanatory mechanism is based on generating perturbed representations to identify the relevant semantic aspects of the KG that explain a relation; and since it does not require retraining of representations, it is particularly efficient. We evaluate SEEK on protein-protein interaction prediction and gene-disease association prediction, two complex and core tasks in the biomedical domain. SEEK clearly outperforms state-of-the-art learning representation methods in performance, while generating explanations that can identify critical factors driving biological phenomena.

In future work, we will conduct user studies with biomedical domain experts to evaluate SEEK explanations and also improve explanations by investigating the minimal set of shared semantic aspects required to adequately explain a relation.

Acknowledgments

References

Alshahrani, M.; Khan, M. A.; Maddouri, O.; Kinjo, A. R.; Queralt-Rosinach, N.; and Hoehndorf, R. 2017. Neuro-symbolic representation learning on biological knowledge graphs. *Bioinformatics* 33(17):2723–2730.

Asif, M.; Martiniano, H. F. M. C. M.; Vicente, A. M.; and Couto, F. M. 2018. Identifying disease genes using machine learning and gene functional similarities, assessed through gene ontology. *PLOS ONE* 13(12):1–15.

Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; Chatila, R.; and Herrera, F. 2020. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion* 58:82–115.

Betz, P.; Meilicke, C.; and Stuckenschmidt, H. 2022. Adversarial explanations for knowledge graph embeddings. In Raedt, L. D., ed., *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, 2820–2826. International Joint Conferences on Artificial Intelligence Organization. Main Track.

Bordes, A.; Usunier, N.; Garcia-Durán, A.; Weston, J.; and Yakhnenko, O. 2013. Translating embeddings for modeling multi-relational data. In *Proceedings of NIPS 2013*, 2787–2795. Red Hook, NY, USA: Curran Associates Inc.

Breiman, L. 2001. Random forests. *Machine learning* 45(1):5–32.

Celebi, R.; Uyar, H.; Yasar, E.; Gumus, O.; Dikenelli, O.; and Dumontier, M. 2019. Evaluation of knowledge graph embedding approaches for drug-drug interaction prediction in realistic settings. *BMC bioinformatics* 20(1):1–14.

Chari, S.; Gruen, D. M.; Seneviratne, O.; and McGuinness, D. L. 2020. Foundations of explainable knowledge-enabled systems. In *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*. IOS Press. 23–48.

Chen, T., and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. New York, NY, USA: Association for Computing Machinery.

Chen, J.; Hu, P.; Jimenez-Ruiz, E.; Holter, O. M.; Antonyrajah, D.; and Horrocks, I. 2021. OWL2Vec*: Embedding of OWL ontologies. *Machine Learning* 1–33.

Chen, K.-H.; Wang, T.-F.; and Hu, Y.-J. 2019. Protein-protein interaction prediction using a hybrid feature representation and a stacked generalization scheme. *BMC Bioinformatics* 20(1):308.

Consortium, G. 2021. The gene ontology resource: enriching a GOld mine. *Nucleic Acids Research* 49(D1):D325–D334.

Couto, F. M., and Silva, M. J. 2011. Disjunctive shared

---

[8]https://www.uniprot.org/uniprot/P01135

[9]https://www.uniprot.org/uniprot/Q15700

information between ontology concepts: application to gene ontology. *Journal of biomedical semantics* 2:1–16.

Ehrlinger, L., and Wöß, W. 2016. Towards a definition of knowledge graphs. *SEMANTiCS (Posters, Demos, SuCCESS)* 48(1-4):2.

Eilbeck, K.; Quinlan, A.; and Yandell, M. 2017. Settling the score: variant prioritization and mendelian disease. *Nature Reviews Genetics* 18(10):599–612.

Gad-Elrab, M. H.; Stepanova, D.; Tran, T.-K.; Adel, H.; and Weikum, G. 2020. Excut: Explainable embedding-based clustering over knowledge graphs. In *The Semantic Web–ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part I*, 218–237. Springer.

Gilpin, L. H.; Bau, D.; Yuan, B. Z.; Bajwa, A.; Specter, M.; and Kagal, L. 2018. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, 80–89. IEEE.

Grau, B. C.; Horrocks, I.; Motik, B.; Parsia, B.; Patel-Schneider, P.; and Sattler, U. 2008. OWL 2: The next step for OWL. *Journal of Web Semantics* 6(4):309–322.

Guidotti, R.; Monreale, A.; Ruggieri, S.; Turini, F.; Giannotti, F.; and Pedreschi, D. 2018. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)* 51(5):1–42.

Han, J.; Liu, S.; Rose, D. M.; Schlaepfer, D. D.; McDonald, H.; and Ginsberg, M. H. 2001. Phosphorylation of the integrin alpha-4 cytoplasmic domain regulates paxillin binding. *Journal of Biological Chemistry* 276(44):40903–40909.

Hoehndorf, R.; Schofield, P. N.; and Gkoutos, G. V. 2011. PhenomeNET: a whole-phenome approach to disease gene discovery. *Nucleic Acids Research* 39(18):e119.

Hogan, A.; Blomqvist, E.; Cochez, M.; d'Amato, C.; Melo, G. d.; Gutierrez, C.; Kirrane, S.; Gayo, J. E. L.; Navigli, R.; Neumaier, S.; et al. 2021. Knowledge graphs. *ACM Computing Surveys (CSUR)* 54(4):1–37.

Holzinger, A.; Biemann, C.; Pattichis, C. S.; and Kell, D. B. 2017. What do we need to build explainable AI systems for the medical domain? *arXiv preprint arXiv:1712.09923*.

Hu, L.; Wang, X.; Huang, Y.-A.; Hu, P.; and You, Z.-H. 2021. A survey on computational models for predicting protein–protein interactions. *Briefings in Bioinformatics* 22(5):bbab036.

Huntley, R. P.; Sawford, T.; Mutowo-Meullenet, P.; Shypitsyna, A.; Bonilla, C.; Martin, M. J.; and O'donovan, C. 2014. The GOA database: gene ontology annotation updates for 2015. *Nucleic Acids Research* 43(D1):D1057–D1063.

Ieremie, I.; Ewing, R. M.; and Niranjan, M. 2022. TransformerGO: predicting protein–protein interactions by modelling the attention between sets of gene ontology terms. *Bioinformatics*.

Kulmanov, M.; Liu-Wei, W.; Yan, Y.; and Hoehndorf, R. 2019. EL embeddings: geometric construction of models for the description logic EL++. *Proceedings of the Twenty-*

Eighth International Joint Conference on Artificial Intelligence.

Kulmanov, M.; Smaili, F. Z.; Gao, X.; and Hoehndorf, R. 2021. Semantic similarity and machine learning with ontologies. *Briefings in Bioinformatics* 22(4):bbaa199.

Köhler, S.; Gargano, M.; Matentzoglu, N.; Carmody, L. C.; Lewis-Smith, D.; Vasilevsky, N. A.; and Danis, D. e. a. 2020. The Human Phenotype Ontology in 2021. *Nucleic Acids Research* 49(D1):D1207–D1217.

Li, D.; Li, D.; Wang, C.; and Chen, Y. 2021. Network embedding method based on semantic information. In *2021 3rd International Conference on Advanced Information Science and System (AISS 2021)*, 1–6.

Mei, Y.; Chen, Q.; Lensen, A.; Xue, B.; and Zhang, M. 2022. Explainable artificial intelligence by genetic programming: A survey. *IEEE Transactions on Evolutionary Computation* 1–1.

Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Miller, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review* 63(2):81.

Mjolsness, E., and DeCoste, D. 2001. Machine learning for science: state of the art and future prospects. *science* 293(5537):2051–2055.

Mohamed, S. K.; Nounu, A.; and Nováček, V. 2021. Biological applications of knowledge graph embedding models. *Briefings in Bioinformatics* 22(2):1679–1693.

Nicholson, D. N., and Greene, C. S. 2020. Constructing knowledge graphs and their biomedical applications. *Computational and structural biotechnology journal* 18:1414–1428.

Nunes, S.; Sousa, R. T.; and Pesquita, C. 2021. Predicting gene-disease associations with knowledge graph embeddings over multiple ontologies. In *ISMB Annual Meeting - Bio-Ontologies*.

Palmonari, M., and Minervini, P. 2020. Knowledge graph embeddings and explainable AI. *Knowledge Graphs for Explainable Artificial Intelligence: Foundations, Applications and Challenges* 47:49.

Pezeshkpour, P.; Tian, Y.; and Singh, S. 2019. Investigating robustness and interpretability of link prediction via adversarial modifications. *NAACL-HLT*.

Piñero, J.; Ramírez-Anguita, J. M.; Saüch-Pitarch, J.; Ronzano, F.; Centeno, E.; Sanz, F.; and Furlong, L. I. 2019. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research* 48(D1):D845–D855.

Portisch, J.; Heist, N.; and Paulheim, H. 2022. Knowledge graph embedding for data mining vs. knowledge graph embedding for link prediction–two sides of the same coin? *Semantic Web* 13(Preprint):1–24.

Ristoski, P., and Paulheim, H. 2016. RDF2Vec: RDF graph embeddings for data mining. In *International Semantic Web Conference*, 498–514. Springer.

Roscher, R.; Bohn, B.; Duarte, M. F.; and Garcke, J. 2020. Explainable machine learning for scientific insights and discoveries. *Ieee Access* 8:42200–42216.

Rossi, A.; Firmani, D.; Merialdo, P.; and Teofili, T. 2022a. Explaining link prediction systems based on knowledge graph embeddings. In *Proceedings of the 2022 International Conference on Management of Data*, 2062–2075.

Rossi, A.; Firmani, D.; Merialdo, P.; and Teofili, T. 2022b. Kelpie: an explainability framework for embedding-based link prediction models. *Proceedings of the VLDB Endowment* 15(12):3566–3569.

Rumelhart, D. E.; Hinton, G. E.; and Williams, R. J. 1986. Learning representations by back-propagating errors. *Nature* 323(6088):533–536.

Sezaki, T.; Tomiyama, L.; Kimura, Y.; Ueda, K.; and Kioka, N. 2013. Dlg5 interacts with the TGF-beta receptor and promotes its degradation. *FEBS Letters* 587(11):1624–1629.

Smaili, F. Z.; Gao, X.; and Hoehndorf, R. 2019. OPA2Vec: combining formal and informal content of biomedical ontologies to improve similarity-based prediction. *Bioinformatics* 35(12):2133–2140.

Sousa, R. T.; Silva, S.; and Pesquita, C. 2020. Evolving knowledge graph similarity for supervised learning in complex biomedical domains. *BMC Bioinformatics* 21(1):6.

Sousa, R. T.; Silva, S.; and Pesquita, C. 2021. evoKGsim+: a framework for tailoring knowledge graph-based similarity for supervised learning. In *ESWC 2021 Poster and Demo Track*.

Staab, S., and Studer, R. 2010. *Handbook on ontologies*. Springer-Verlag.

Szklarczyk, D.; Gable, A. L.; Nastou, K. C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N. T.; Legeay, M.; Fang, T.; Bork, P.; Jensen, L. J.; and von Mering, C. 2020. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research* 49(D1):D605–D612.

Trapnell, B. C.; Whitsett, J. A.; and Nakata, K. 2003. Pulmonary alveolar proteinosis. *New England Journal of Medicine* 349(26):2527–2539.

Van der Maaten, L., and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9(11).

Wang, Z.; Zhang, J.; Feng, J.; and Chen, Z. 2014. Knowledge Graph Embedding by Translating on Hyperplanes. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 1112–1119. AAAI Press.

Wang, Q.; Mao, Z.; Wang, B.; and Guo, L. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering* 29(12):2724–2743.

Watson, D. S.; Gultchin, L.; Taly, A.; and Floridi, L. 2021. Local explanations via necessity and sufficiency: Unifying theory and practice. In *Uncertainty in Artificial Intelligence*, 1382–1392. PMLR.

Xiong, B.; Potyka, N.; Tran, T.-K.; Nayyeri, M.; and Staab,

S. 2022. Faithful Embeddings for EL++ Knowledge Bases. In *International Semantic Web Conference*, 22–38. Springer.

Yang, B.; tau Yih, W.; He, X.; Gao, J.; and Deng, L. 2015. Embedding entities and relations for learning and inference in knowledge bases.

Zhang, S.-B., and Tang, Q.-R. 2016. Protein–protein interaction inference based on semantic similarity of gene ontology terms. *Journal of Theoretical Biology* 401:30–37.