Resilience in Knowledge Graph Embeddings

Arnab Sharma
Data Science Group (DICE), Heinz Nixdorf Institute, Paderborn University, Germany

N'Dah Jean Kouagou
Data Science Group (DICE), Heinz Nixdorf Institute, Paderborn University, Germany

Axel-Cyrille Ngonga Ngomo
Data Science Group (DICE), Heinz Nixdorf Institute, Paderborn University, Germany

Abstract

In recent years, knowledge graphs have gained interest and witnessed widespread applications in various domains, such as information retrieval, question-answering, recommendation systems, amongst others. Large-scale knowledge graphs to this end have demonstrated their utility in effectively representing structured knowledge. To further facilitate the application of machine learning techniques, knowledge graph embedding models have been developed. Such models can transform entities and relationships within knowledge graphs into vectors. However, these embedding models often face challenges related to noise, missing information, distribution shift, adversarial attacks, etc. This can lead to sub-optimal embeddings and incorrect inferences, thereby negatively impacting downstream applications. While the existing literature has focused so far on adversarial attacks on KGE models, the challenges related to the other critical aspects remain unexplored. In this paper, we, first of all, give a unified definition of *resilience*, encompassing several factors such as generalisation, in-distribution generalization, distribution adaption, and robustness. After formalizing these concepts for machine learning in general, we define them in the context of knowledge graphs. To find the gap in the existing works on resilience in the context of knowledge graphs, we perform a systematic survey, taking into account all these aspects mentioned previously. Our survey results show that most of the existing works focus on a specific aspect of resilience, namely robustness. After categorizing such works based on their respective aspects of resilience, we discuss the challenges and future research directions.

2012 ACM Subject Classification Computing methodologies → Reasoning about belief and knowledge

Keywords and phrases Knowledge graphs, Resilience, Robustness

Digital Object Identifier 10.4230/TGDK.3.2.1

Category Survey

Funding This work has been supported by the Ministry of Culture and Science of North Rhine-Westphalia (MKW NRW) within the project SAIL under the grant no NW21-059D, the project WHALE (LFN 1-04) funded under the Lamarr Fellow Network programme by the Ministry of Culture and Science of North Rhine-Westphalia (MKW NRW), the European Union's Horizon Europe research and innovation programme under grant agreement No 101070305, and by the German Federal Ministry of Research, Technology and Space (BMFTR) within the project KI-OWL under the grant no 01IS24057B.
Arnab Sharma: SAIL under the grant no NW21-059D
N'Dah Jean Kouagou: European Union's Horizon Europe research and innovation programme under grant agreement No 101070305, WHALE (LFN 1-04) funded under the Lamarr Fellow Network programme
Axel-Cyrille Ngonga Ngomo: KI-OWL under the grant no 01IS24057B

Received 2024-09-24 **Accepted** 2025-03-11 **Published** 2025-10-15

1 Introduction

In recent years, there has been significant progress in the construction and application of knowledge graphs (KGs). Many KGs, including Freebase [14], DBpedia [4], YAGO [91], and NELL [20], have been developed and successfully implemented in various real-world applications. Due to

their effectiveness in knowledge representation, KGs now find applications in domains such as information retrieval [32], question answering [42], and recommendation systems [99], amongst others. A KG serves as a structured depiction of knowledge, organized as a multi-relational graph where nodes denote entities or concepts, and edges signify relationships between them [53]. Knowledge therein is represented using assertions – model statements (which could in some cases be real-world facts) – in the form of triples denoted as (h, r, t), where h and t correspond to the head and tail entities respectively, and r represents the relationship between them. For instance, the fact "Biden is the president of USA" can be represented in a KG as $(\texttt{Biden}, \texttt{presidentOf}, \texttt{USA})$.

Knowledge Graph Embedding (KGE) involves transforming the entities and relations within a KG into vectors [15, 35, 103, 121, 124]. This transformation makes computational operations more feasible, allowing machine learning and deep learning techniques to be applied to extract insights from the KG. Consequently, an effective KGE model should aim to preserve the properties and semantics inherent in the original KG. Based on the type of KGE models, entities and relations are commonly embedded in d-dimensional vector spaces $\mathbb{V}$ such as $\mathbb{R}^d$ (real numbers) [15], $\mathbb{C}^d$ (complex numbers) [97], or even $\mathbb{H}^d$ (quaternions) [21].

Despite their effectiveness in capturing complex relationships between entities of KGs and facilitating various downstream tasks, KGE models can be vulnerable to *adversarial manipulations* [11,12,80,108,119,122]. Since these models rely heavily on the observed connections in a given graph, noise or missing information can lead to sub-optimal embeddings and potentially incorrect inferences. For instance, the presence of incorrect triples (in the sense of non-conformity with an ontology, or wrong assertions) might lead to poorly performing KGE models on certain downstream tasks [57, 129]. To this end, a deliberate attack or the presence of noise can both equally degrade the performance. KGE models might also struggle to generalize to out-of-distribution or unseen data, e.g., when the underlying data distribution changes or when encountering new entities or relations. Since KGs often contain sensitive and critical information pertaining to individuals or organizations, this might give rise to potential security vulnerabilities. For instance, an attacker might subtly alter the relation between entities or introduce fictitious entities and relationships that distort the model's understanding of the graph and make the KGE model learn *poisoned* embeddings. Such adversarial attacks on KGE models can take various forms, such as adding, deleting, or modifying triples within the knowledge graph, where such perturbations are often minimal and crafted to exploit vulnerabilities in the embedding process. Due to the usage of KGE models in various downstream tasks, such adversarial attacks can cause potential disruptions in these tasks, for instance, in

1. **Question answering**, adversarial modifications can cause KGE models to produce incorrect or manipulated answers or fail to retrieve relevant information,
2. **Recommendation systems**, the embeddings can be poisoned to promote certain items unfairly, leading to biased or irrelevant recommendations,
3. **Information extraction**, adversarial perturbations can result in inaccurate extractions of facts, affecting downstream applications like content summarization or data integration, amongst others.

Therefore, to reliably use KGE models in downstream tasks, there is a need to develop models that can function without any potential disruption even in the presence of such adversarial conditions.

Although the aforementioned challenges pose potential threats to the use of KGE models in critical downstream tasks, current efforts to deal with these challenges still remain infancy. The existing literature mostly contains works addressing challenges related to noisy data, distribution shifts, and adversarial attacks in the context of graph neural networks [24, 40, 127]. So far, works considering KGE models mostly focused on performing adversarial attacks on them [11,12,119,122]. The core idea behind these attacks is to target specific facts and modify the KGE model to either

increase or decrease their plausibility scores. These scores reflect the likelihood of a fact being true: higher scores imply higher probability, while lower scores imply lower probability. For instance, if (`Biden`, `PresidentOf`, `USA`) is selected as the target triple, one type of adversarial attack would be to make the underlying KGE model assign a low plausibility score to it. In this case, such attacks are typically dealt with via a min-max optimization function, where the objective is to minimize the inclusion/deletion of adversarial/existing triples in/from the underlying KG [119]. Simultaneously, the attacker aims to maximize the objective function, which involves either increasing or decreasing the plausibility of a targeted fact being true.

Since KGs are used in many safety-critical environments, safeguarding sensitive information and preserving user privacy are paramount considerations in deploying KGE models in real-world settings. Furthermore, we need to enable KGE models to adapt to dynamic environments and evolving data distributions to enhance their resilience to concept drift and temporal changes. Therefore, in this work, we first of all propose the concept of *resilience* in the context of ML, and further extend the definition for KGE models. We aim to bridge the gap in resilience literature on KGE from a holistic perspective that considers the diverse facets of robustness, adaptability, distribution shift, and consistency, amongst others. By addressing these aspects comprehensively, researchers can propel the development of resilient KGE models that not only excel in performance metrics but also demonstrate stability and reliability in real-world applications. Note that, our resilience definition is quite generic, i.e., it does not depend on any specific application domain. Precisely, we give a generic formal definition of resilience in ML models considering (i) generalization consistency, (ii) distribution adaptation, (iii) in-distribution generalization, (iv) robustness, and (v) missing entry handling. We then discuss these aspects of resilience in the context of KGE models. To this end, we survey the works on KGE models considering the aforementioned aspects of resilience. Specifically, we provide a survey of works studying the resilience of KGE models in any of the aspects from (i)–(v). After discussing these works, we highlight possible challenges and suggest future work directions.

Note that this paper provides two-faceted contributions. After exploring existing literature on KGE models, we recognize the need for a holistic definition of resilience in embedding models. Therefore, in this work, we first introduce a formal definition of resilience, considering five aspects. Thereafter, we discuss the related works in this context. In this sense, our paper is not purely a survey, rather, it combines a conceptual framework that defines and evaluates resilience in KGE models with a survey of different notions of resilience. Additionally, we propose a comparative analysis of existing methods and explore potential challenges and future works.

This paper is organized as follows. Section 2 formalizes the notions of KGs and KGE models. The definition of resilience is given in Section 3. Section 4 describes the methodology regarding the collection of papers. Existing works discussing aspects of resilience are presented in Section 5. Section 6 presents and discusses different aspects of robustness. Section 10 highlights existing challenges and potential future work directions, and Section 11 concludes the paper.

2 Foundations

A knowledge graph is a collection of assertions that describe a domain of interest. In this paper, we consider knowledge graphs composed of *triples* $(h, r, t) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}$, where $\mathcal{E}$ is a discrete set of entities and $\mathcal{R}$ is a discrete set of relations. Therefore, KGs are representations of information in a discrete space. More formally, a KG is defined as a set of triples $\mathcal{G} := \{(h, r, t) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}\}$, where $\mathcal{E}$ and $\mathcal{R}$ stand for a set of entities and a set of relations [7,35]. To facilitate downstream applications, KGE algorithms have been developed to represent a KG in a continuous, low-dimensional vector space. Given a KG $\mathcal{G} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$, the goal of a KGE model is to learn continuous vector

representations for entities and relation types in $\mathcal{G}$ such that these representations can be used to recover all the facts in $\mathcal{G}$. Most KGE approaches are tailored towards link prediction [21, 53], i.e., their scoring function is $\phi_\Theta : \mathcal{E} \times \mathcal{R} \times \mathcal{E} \to \mathbb{R}$, where Θ denotes parameters and often comprises $\mathbf{E}$, $\mathbf{R}$, and additional parameters (e.g., affine transformations, batch normalizations, convolutions). Given an assertion in the form of a triple $(h, r, t) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}$, a prediction $\hat{y} := \phi_\Theta(h, r, t)$ signals the likelihood of (h, r, t) being true [35, 103, 121]. Therefore, KGE models are learned in such a way that the scoring function assigns a higher score to the triples that exist in the KG compared to the non-existing ones.

Let $\mathbb{V}$ denote a normed-division algebra, e.g. $\mathbb{R}, \mathbb{C}, \mathbb{H}$, or $\mathbb{O}$ [6, 34, 97, 116, 125]. A KGE model of a KG comprises entity embeddings $\mathbf{E} \in \mathbb{V}^{|\mathcal{E}| \times d_e}$ and relation embeddings $\mathbf{R} \in \mathbb{V}^{|\mathcal{R}| \times d_r}$, where d_e and d_r are the size of the embedding vectors. Note that some KGE models represent entities or relations as matrices or higher-dimensional tensors, e.g., RESCAL [65, 76]. Throughout this paper, we will focus on vector representations for entities and relations and denote embedding vectors with bold fonts, for instance, the embedding of h, r, and t will be denoted as $\mathbf{h}$, $\mathbf{r}$, and $\mathbf{t}$, respectively. Since KGs contain triples which represent the existing facts only, to learn a KGE model effectively, non-existing facts, i.e., *negative* facts often need to be incorporated into the learning process. For that, a technique called negative sampling is used to generate a number of false facts or negative triples. To this end, Bordes et al. [15] proposed a negative sampling technique by perturbing an entity in a randomly sampled triple from the KG. In this setting, a triple $(h, r, t) \in \mathcal{G}$ is considered as a positive example, whilst $\{(h, r, x) | x \in \mathcal{E} \wedge (h, r, x) \notin \mathcal{G}\} \cup \{(x, r, t) | x \in \mathcal{E} \wedge (x, r, t) \notin \mathcal{G}\}$ is regarded as the set of possible candidate negative triples corresponding to (h, r, t). During training, k negative triples are constructed for every correct triple.

3 Resilience

As mentioned beforehand, resilience is a term that is frequently used when engineering systems, more specifically in the context of building fault-tolerant systems [90]. In those systems, *resilience refers to the ability of a system to maintain its functionality and performance in the face of faults, failures, disruptions, or adverse conditions.* In other words, a resilient system is capable of detecting, mitigating, and recovering from faults or failures, ensuring continuous operation and minimal impact on its overall performance and availability. Therefore, in [10], the authors defined the resilience of a system using

1. **availability**, i.e., the readiness for correct service,
2. **reliability**, i.e., the probability of performing correctly for a period of time,
3. **safety**, i.e., the robustness against adversarial manipulations,
4. **integrity**, i.e., the absence of improper system altercation, and
5. **maintainability**, i.e., the ability to undergo modifications and repairs.

Note that, due to the lack of resilience definition in the ML literature, we use this as our starting point. While the typical definition of resilience in fault-tolerant systems provides a useful help for understanding resilience in the machine learning domain, it needs to be extended and adapted to account for the unique characteristics, challenges, and considerations inherent in machine learning models and systems. More specifically, for ML models, resilience cannot be defined by using these parameters directly since they do not capture the typical data-driven workflow that is used in ML. For this, we need to consider other factors such as *consistent* performance in a distribution, or when a *distribution shift* occurs, robustness, stability, amongst others.

This paper makes a two-fold contribution in addressing this gap. First, after analyzing existing works on KGE models, we identify the necessity for a holistic definition of resilience tailored to embedding models. To this end, we propose a formal definition of resilience, considering aspects

that encapsulate the ability of a model to generalize, adapt, and maintain stable performance under varying conditions. Second, we provide a structured discussion of related works that align with these resilience principles, offering insights into existing approaches and their limitations. As mentioned beforehand, our work is not purely a survey; rather, it combines theoretical formalization with a comprehensive review, bridging the gap between conceptual understanding and practical advancements in resilient knowledge representation.

To define resilience, we start with some basic formalization. Let us consider an ML model as a function f which takes as input x coming from a specific distribution $\mathcal{D}$. We define two types of distributions from where the data might come, the source distribution which is defined as $\mathcal{D}_s$ from where the training data comes, and the target distribution $\mathcal{D}_t$ on which the model would typically operate. The sets of values corresponding to the distributions $\mathcal{D}_s$ and $\mathcal{D}_t$ can be defined as $\mathcal{X}_s$ and $\mathcal{X}_t$, respectively. $\mathcal{H}(\mathcal{D}_s, \mathcal{D}_t)$ defines a divergence measure between the two distributions $\mathcal{D}_s$ and $\mathcal{D}_t$. Furthermore, we define $\mathcal{L}_f$ as the *loss function* of the model measuring the model's performance on a set of data instances. Note that, the loss function can be of any type, however, our definition is independent of it. An ML model f is said to be resilient if it conforms to the following constraints:

Generalization consistency corresponds to the ability of the model to generalize consistently across different distributions of data. This can be formally defined as

$$\exists\ \epsilon > 0 \text{ s.t. } \forall\ \mathcal{D}_s, \mathcal{D}_t,\ |\mathbb{E}_{\mathcal{D}_s}(\mathcal{L}_f) - \mathbb{E}_{\mathcal{D}_t}(\mathcal{L}_f)| \leq \epsilon, \tag{1}$$

where ϵ defines a threshold that basically bounds the difference between the average losses on the training data distribution $\mathcal{D}_s$ and the target data distribution $\mathcal{D}_t$. Xu et al. [112] defined this as the robustness property of the learning algorithms where they argued that a robust algorithm should achieve similar performance on the training and testing data that are *close* in some sense; which basically corresponds to the robust optimization problem. However, in connection to resilience, we define this as the *consistency* property over the generalization of the model f. To this end, we simply say that the loss occurring on the data instances taken from the target distribution might differ only by a threshold ϵ from the loss occurring on the instances of the source distribution. Note that, in that sense, this definition could also be termed as *out of distribution generalization*, since this captures how well a model can perform when the inputs are out-of-distribution compared to the training dataset. Furthermore, in this definition, we are not concerned with whether a model f achieves high accuracy or low loss on the training data; with generalization consistency, we aim to signify that the model's performance should not vary drastically between the training and test data distributions.
In the context of knowledge graph embedding models, generalization consistency refers to the model's ability to meaningfully construct embeddings for unseen entities or relations, and accurately predict missing links between entities based on the learned patterns from the training data. Assuming $\mathcal{L}_{\phi_\Theta}$ is a loss function which can be used to train the parameterized embedding model ϕ_Θ, generalization consistency can be defined as

$$\exists\ \epsilon > 0 \text{ s.t. } \forall\ \mathcal{D}_\mathcal{G}, \mathcal{D}_{\mathcal{G}'},\ |\mathbb{E}_\mathcal{G}(\mathcal{L}_{\phi_\Theta}) - \mathbb{E}_{\mathcal{G}'}(\mathcal{L}_{\phi_\Theta})| \leq \epsilon, \tag{2}$$

where $\mathcal{D}_\mathcal{G}$ and $\mathcal{D}_{\mathcal{G}'}$ refer to the distribution of the training knowledge graph's data and that of the test knowledge graph's data, respectively.

Distribution adaption corresponds to the model's ability to adapt to a target domain (i.e., test data distribution) without significantly compromising its performance as achieved on the source domain (i.e., training data distribution). This can be defined as follows

$$\forall\ \epsilon > 0, \exists\ \delta > 0 \text{ s.t. } \forall\ \mathcal{D}_s, \mathcal{D}_t,\ \mathcal{H}(\mathcal{D}_s, \mathcal{D}_t) \leq \delta \Rightarrow |\mathbb{E}_{\mathcal{D}_s}(\mathcal{L}_f) - \mathbb{E}_{\mathcal{D}_t}(\mathcal{L}_f)| \leq \epsilon, \tag{3}$$

where $\mathcal{H}(\mathcal{D}_s, \mathcal{D}_t)$ defines any divergence measure such as maximum mean discrepancy (MMD) [89], Kullback-Leibler (KL) divergence [59], or Wasserstein distance [106]. Informally, if the distributions $\mathcal{D}_s$ and $\mathcal{D}_t$ are different with a bound δ, then the average prediction losses on the data instances in these distributions must not differ more than ϵ. Note that, the distributional mismatch between the training and test data has been studied in many settings, for instance, in [13, 39, 54, 82, 88, 115] and as pointed out by the authors in [1] most of these works assume the *covariate shift* where only the distribution of class labels differs between the training and test distributions. There exist some works such as [9, 29, 41] which consider shift of generic data distributions, however, none of them consider this as part of the resilience of ML models. For KGE models, distribution adaptation refers to a model's ability to adjust its parameters to account for changes in a given knowledge graph. When new entities, relation types, or new links are added to (or removed from) a given knowledge graph, the resulting graph data distribution might deviate from the initial one. In this case, the KGE model's adaptation to this distribution change can be formally defined as follows.

$$\forall\ \epsilon > 0, \exists\ \delta > 0 \text{ s.t. } \forall\ \mathcal{D}_{\mathcal{G}}, \mathcal{D}_{\mathcal{G}'},\ \mathcal{H}(\mathcal{D}_{\mathcal{G}}, \mathcal{D}_{\mathcal{G}'}) \leq \delta \Rightarrow |\mathbb{E}_{\mathcal{D}_{\mathcal{G}}}(\mathcal{L}_{\phi_\Theta}) - \mathbb{E}_{\mathcal{D}_{\mathcal{G}'}}(\mathcal{L}_{\phi_\Theta})| \leq \epsilon. \tag{4}$$

Unlike generalization consistency, which assumes stable data conditions, distribution adaptation ensures that the model can adjust to new distributions without significant performance degradation. In other words, generalization consistency ensures stability across similar distributions, while distribution adaptation guarantees stability on dynamic or shifted distributions.
Note that, in the context of graphs, a distribution shift refers to a change in the statistical distribution of the graph data. This can manifest in different ways, such as

1. **Node feature distribution shift** which occurs when the distribution of node attributes or features changes over time or across different subsets of the graph. For example, in a knowledge graph representing entities and their attributes (e.g., people and their professions), a node attribute shift could involve changes in the distribution of professions among individuals over time or across different subsets of the graph. Nodes may furthermore be added to or removed from the knowledge graph, leading to changes in the overall node distribution. This could happen, for instance, when new entities are discovered or when outdated entities are removed from the knowledge graph.
2. **Node degree shift** which happens when some relationships between entities are removed (e.g., two entities that were previously friends are no longer friends) or added, e.g., (an entity gets married to another entity). It could also be the case that new entities are introduced but with little to zero links to other entities in the graph. When such changes in relationships between entities are significant, the average degree of nodes in the considered knowledge graph might also shift.
3. **Edge feature distribution shift** which refers to changes in the properties or attributes associated with the relationships (edges) between nodes in the knowledge graph. For example, in a knowledge graph representing relationships between entities (e.g., co-authorship relationships between researchers), an edge attribute shift could involve changes in the publication venues or collaboration patterns over time. New relationships may further be established or existing relationships may be removed from the knowledge graph, leading to changes in the edge distribution. This could occur due to the emergence of new relationships or the obsolescence of existing ones.
4. **Graph structure shift** which involves alterations in the overall structure or topology of the knowledge graph, including changes in connectivity patterns between nodes, changes in node/edge attributes (e.g., many entities and relationships in the reference knowledge graph now have textual descriptions), and changes in entity type hierarchies. For example, in a

knowledge graph representing hierarchical relationships (e.g., taxonomy or ontology), changes in the hierarchy or the addition of new branches can lead to structural shifts. Changes to the schema or ontology of the knowledge graph, such as the addition, modification, or removal of entity types, relationship types, or property types, can also constitute graph structure shifts. These changes may reflect updates in domain knowledge or evolving data modeling requirements.

To make the distinction between generalization consistency and distribution adaption more concrete in the context of KGE models, we consider the following example. Consider a recommendation system based on a knowledge graph. Generalization consistency would ensure that the embeddings trained on historical data remain effective for predicting new links in the same dataset. However, distribution adaptation would be required if the dataset undergoes significant changes, such as the inclusion of new user demographics or shifting product categories.

In-distribution generalization corresponds to the model's ability to perform consistently across different instances or subsets of a data distribution. Typically, this distribution could be the target distribution $\mathcal{D}_t$ where the data instances come from the model deployment phase. Formally, consistency can be defined as follows,

$$\exists\ \epsilon > 0 \text{ s.t. } \forall\ \mathcal{S}, \mathcal{S}' \subseteq \mathcal{D}_{\mathcal{G}},\ |\mathbb{E}_{\mathcal{S}}(\mathcal{L}_f) - \mathbb{E}_{\mathcal{S}'}(\mathcal{L}_f)| \leq \epsilon. \tag{5}$$

Here we enforce that, for any two non-empty subsets $\mathcal{S}$ and $\mathcal{S}'$ from the distribution $\mathcal{D}_{\mathcal{G}}$, the expected losses achieved on the two sets differ at most only by some parameter ϵ [1]. If the observed differences between the losses are statistically significant (e.g., greater than ϵ), it indicates that the model's performance varies consistently across different subsets of the data, suggesting potential limitations or biases in the model. On the other hand, if the observed differences are not statistically significant, it suggests that the model's performance remains consistent across subsets, providing greater confidence in its resilience. Furthermore, this measure of consistency is different from the generalization consistency in the sense that herein we consider uniform performance across different sub-spaces of the same distribution space, whereas in case of generalization consistency, two different distributions are considered.

In-distribution generalization for the KGE models refers to the model's ability to maintain consistent performance across different instances or subsets of the knowledge graph data distribution, particularly when deployed in real-world applications where the distribution of incoming triples may vary. In other words, the KGE model should demonstrate resilience to variations in the distribution of knowledge graph data encountered during deployment, ensuring that its performance remains reliable and predictable across different scenarios. This consistency is crucial for maintaining the effectiveness and reliability of the model in real-world applications where the knowledge graph would evolve over time or across different contexts.

Robustness focuses on the model's stability with respect to some small changes in the input. In the literature, two versions of robustness are generally considered, namely *local* and *global* robustness [47, 85]. Informally, local robustness corresponds to a single point x, and requires any points within a specific distance Δ to x to be classified as the same as the former. More formally, this can be defined with respect to a data point x as

$$\exists\ \epsilon > 0 \text{ s.t. } \forall\ x',\ ||x - x'||_p \leq \Delta \Rightarrow ||f(x) - f(x')||_p \leq \epsilon. \tag{6}$$

[1] Note that, here we have considered a strong notion of consistency, however, a weaker notion can also be chosen where the subsets must follow some specific rules.

On the other hand, Seshia et al. [85] defined global robustness considering all the points within a specific distribution $\mathcal{D}$. In other words, for every point x within a considered distribution, any other point x' which is within Δ distance from x should be classified as the same class as x. This can be formally defined as

$$\exists\ \epsilon > 0 \text{ s.t. } \forall\ x, x' \in \mathcal{D},\ ||x - x'||_p \leq \Delta \Rightarrow ||f(x) - f(x')|| \leq \epsilon. \tag{7}$$

Note that, in the literature, robustness is more often associated with the idea of local robustness for a single point or a set of points. Thus, in defining resilience, we would primarily consider the local robustness property of ML models. Note that, herein, $f(x)$ could be a single integer or could also be a probability. This would depend on the type of the underlying model.
In the context of KGE models, we can adapt the concept of local robustness to refer to the model's ability to produce consistent embeddings for entities or relations that are similar in the graph structure. Informally, local robustness in this context would correspond to: any entity (respectively, relation) within a specific neighborhood of an entity h (respectively a relation r), defined by a distance metric, should have an embedding that is similar to that of h (respectively r). More formally, for an entity or a relation x in the knowledge graph, and for any other entity or relation x' within a specific distance Δ of x in the graph structure, the embeddings produced by the KGE model, say $\mathbf{x}$ and $\mathbf{x}'$ should be similar, with their distance in the embedding space bounded by ϵ. Given a knowledge graph $\mathcal{G}$, this idea of robustness can be formally defined as

$$\exists\ \epsilon > 0 \text{ s.t. } \forall\ x, x' \in \mathcal{G}, d_{\mathcal{G}}(x, x') \leq \Delta \Rightarrow d_{Emb}(\mathbf{x}, \mathbf{x}') \leq \epsilon, \tag{8}$$

where $d_{\mathcal{G}} : \mathcal{G} \times \mathcal{G} \rightarrow \mathbb{R}_+$ is a distance on the graph $\mathcal{G}$, e.g., Adamic-Adar index, Katz similarity, or Common Neighbors, and d_{Emb} a distance function in the embedding space, e.g., the Euclidean distance. ϵ is a threshold that limits the allowable difference between embeddings to ensure local robustness.
The above definition of robustness concerns the functionality of the embedding models in generating robust embeddings. However, we require further robustness notion encompassing the KGE model as well as the scoring function together. To this end, first of all, we define the adversarial robustness.
Adversarial robustness for KGE models refers to a model's ability to maintain its performance and produce reliable predictions in the presence of worst-case perturbations intentionally crafted to degrade its functionality. Herein, we define the robustness property considering the KGE model+scoring function whereas the previous robustness definition (Equation 8) considers solely the KGE models. The perturbations considered can be applied to the symbolic KG or directly to the embedding space and are designed to maximize the model's predictive errors. Formally, adversarial robustness by considering the *symbolic KG* can be described as

$$\exists\ \epsilon > 0 \text{ s.t. } \forall\ \mathcal{G}, \mathcal{G}', \Psi(\mathcal{G}, \mathcal{G}') \leq \Delta \implies \frac{\mathbb{E}_{\mathcal{G}}(\mathcal{L}_{\phi_\Theta})}{\mathbb{E}_{\mathcal{G}'}(\mathcal{L}_{\phi_\Theta}) + \eta} \approx \epsilon, \tag{9}$$

where $\Psi(\mathcal{G}, \mathcal{G}')$ denotes the structural similarity of two graphs, $\mathbb{E}_{\mathcal{G}}(\mathcal{L}_{\phi_\Theta})$, $\mathbb{E}_{\mathcal{G}'}(\mathcal{L}_{\phi_\Theta})$ denote the expected loss of the embedding model ϕ_Θ on the graphs $\mathcal{G}$ and $\mathcal{G}'$, respectively, and η is an infinitesimal number (e.g., $\eta = 10^{-8}$, i.e., a small but non-zero scalar value). ϵ in the ideal case should be close to 1. $\Psi(\mathcal{G}, \mathcal{G}')$ could be defined using graph isomorphism [49], or sub-graph similarity [81] matching technique such as graphlet similarity, frequent subgraph mining, or global graph similarity techniques. Such a measure could be decided based on the specific domain. Note that, the above definition considers changes in the KG, however, this could be extended considering perturbations performed on the embedding. Formally, adversarial robustness by considering the embedding can be described as

$$\exists\ \epsilon > 0 \text{ s.t. } \forall\ \mathbf{E}, \mathbf{E}' : \Psi(\mathbf{E}, \mathbf{E}') \leq \Delta \implies \frac{\mathbb{E}_{\mathbf{E}}(\mathcal{L}_{\phi_\Theta})}{\mathbb{E}_{\mathbf{E}'}(\mathcal{L}_{\phi_\Theta}) + \eta} \approx \epsilon. \tag{10}$$

Here $\mathbf{E}$ and $\mathbf{E}'$ represent the original and perturbed embeddings, respectively. $\Psi(\mathbf{E}, \mathbf{E}')$ measures the similarity or distance between $\mathbf{E}$ and $\mathbf{E}'$. This can be defined as $||\mathbf{E} - \mathbf{E}'||_p$ (e.g., $p = 1$ for the L1 norm and $p = 2$ for the L2 norm) or $\frac{\mathbf{E} \cdot \mathbf{E}'}{||\mathbf{E}||, ||\mathbf{E}'||}$ (cosine similarity) or $\sum_i \mathbf{E}_i \log\left(\frac{\mathbf{E}_i}{\mathbf{E}'_i}\right)$ (KL divergence).

Note that Equations (8)–(10) give a generic notion of adversarial robustness which can be extended by considering the case where the aim is to degrade the score of a specific triple, i.e., $\phi_\theta(h, r, t)$ by doing δ changes on the KG $\mathcal{G}$ or on the embedding space $\mathbf{E}$. The existing works on adversarial robustness of KGE models, while lacking a formal definition, focus on this specific notion [75, 86, 87, 111, 128].

Non-adversarial robustness corresponds to the ability of a KGE model (including its scoring function) to be invariant to a certain level of noise present in a KG. More specifically, the performance of a robust KGE model should not degrade considerably when noise is prevalent in KG. Consider $\mathcal{G}$ as a clean KG and $\mathcal{G}'$ as a noisy KG, where the latter is obtained by adding δ amount of noise to the KG $\mathcal{G}$, i.e., $\mathcal{G}' = \mathcal{G} + \delta$. [2] Then the robustness can be defined as

$$\frac{\mathbb{E}_{\mathcal{G}}(\mathcal{L}_{\phi_\Theta})}{\mathbb{E}_{\mathcal{G}'}(\mathcal{L}_{\phi_\Theta}) + \eta} \approx 1, \tag{11}$$

where $\mathbb{E}_{\mathcal{G}}(\mathcal{L}_{\phi_\Theta})$, $\mathbb{E}_{\mathcal{G}'}(\mathcal{L}_{\phi_\Theta})$ denote the expected loss of the embedding model ϕ_Θ on the graphs $\mathcal{G}$ and $\mathcal{G}$, respectively, and η is an infinitesimal number. This implies that the performance of the KGE models should remain almost the same even when δ amount of noise is present in $\mathcal{G}$. Note that we assume the expected loss not to be zero, as it is often the case in most machine learning tasks. Note that the primary difference between adversarial and non-adversarial robustness lies in the nature of the perturbations. Adversarial perturbations are crafted with intent and precision, targeting the model's weaknesses, while non-adversarial perturbations are accidental and random, reflecting real-world data imperfections. Additionally, adversarial robustness is critical for security-focused applications to protect against malicious attacks, whereas non-adversarial robustness is essential for ensuring reliability in a real-world environment.

Stability w.r.t. incomplete input deals with the model's ability to handle missing values, more specifically, maintain accurate predictions despite the presence of missing values in the input features. We can express this as follows:

$$\exists\ \epsilon > 0 \text{ s.t. } \forall\ x, x^*, |x| > |x^*| \ \wedge\ \text{Sim}(x, x^*) \leq \delta \Rightarrow ||f(x) - f(x')|| \geq \epsilon. \tag{12}$$

Herein $\text{Sim}(x, x^*)$ measures the similarity between two vectors x and x^* with unequal numbers of elements (i.e., $|x| > |x^*|$). One such similarity measure could be the cosine similarity, which is often used for comparing the similarity between vectors in high-dimensional spaces. The cosine similarity measures the cosine of the angle between two vectors and is defined as the dot product of the vectors divided by the product of their magnitudes. Elements that are missing in one vector but present in the other are effectively treated as zeros in the dot product. Another approach is to use measures that explicitly handle missing values, such as the Jaccard similarity or the Pearson correlation coefficient, with the imputation of missing values.

[2] Here, + is not the usual addition, but a perturbation operator instead.

In the context of KGs, this aspect of resilience deals with the ability of a KGE model to maintain stable predictions despite missing nodes, edges, or attributes in the input knowledge graph. To this end, we define stability to incomplete input formally as follows:

$$\exists\ \epsilon > 0 \text{ s.t. } \forall\ \mathcal{G}, \mathcal{G}^*, |\mathcal{G}| > |\mathcal{G}^*| \wedge \text{Sim}(\mathcal{G}, \mathcal{G}^*) \leq \delta \Rightarrow ||\mathbb{E}_{\mathcal{G}}(\mathcal{L}_{\phi_\Theta}) - \mathbb{E}_{\mathcal{G}'}(\mathcal{L}_{\phi_\Theta})|| \leq \epsilon. \tag{13}$$

where $\mathcal{G}$ is the original, complete knowledge graph, and $\mathcal{G}'$ is the incomplete KG with missing nodes, edges, or attributes. $\text{Sim}(\mathcal{G}, \mathcal{G}^*)$ measures the structural similarity between the two graphs. For KGs, Sim() could be graph edit distance, i.e., number of node/edge insertions, deletions, or modifications required to transform $\mathcal{G}$ into $\mathcal{G}'$, or Jaccard similarity over entity/relation sets, amongst others. Note that, this definition implies that the expected loss of the embedding model ϕ_Θ on the graphs $\mathcal{G}$ and $\mathcal{G}$, $\mathbb{E}_{\mathcal{G}}(\mathcal{L}_{\phi_\Theta})$, $\mathbb{E}_{\mathcal{G}'}(\mathcal{L}_{\phi_\Theta})$ respectively, should not change more than a fixed threshold ϵ. The similarity measure can be quite flexible and will potentially depend on the domain (for e.g., image classification, graph data, and others), and hence, we do not fix the Sim() function. Depending on this function and the domain of the application, the bound δ will also change, however, not drastically.

4 Paper Collection Methodology

To discuss resilience in KGs and KGE models, in this work, we further review existing works in this domain. While doing a literature survey of such works, we adhere to specific inclusion criteria for compiling papers for our review. If a paper satisfies any or many of the following criteria, it is considered for inclusion:

1. the paper introduces or discusses the overarching concept of any related aspect outlined in Section 3.
2. the paper proposes an approach, study, or tool/framework aimed at developing resilient or robust KGE models.
3. the paper introduces a set of measurement criteria applicable for defining resilience of KGE models or KGs.

We briefly discuss some papers focusing solely on using KGs to make resilient systems, however, we do not delve into detail on this. To comprehensively gather papers across various research domains, we initiated our search process by employing precise keyword queries on prominent scientific databases such as Google Scholar, DBLP, and arXiv. The keywords that we searched for are detailed in the **Keywords** column in Table 1. We conducted searches across the three repositories until 23.09.2024, aiming to encompass a broad spectrum of literature. The specifics of the paper collection outcomes are outlined in Table 1. It is observed that the papers obtained from Google Scholar and arXiv were subsets of those gathered from DBLP. Therefore, we solely present the results obtained from DBLP. Furthermore note that apart from the papers that discuss resilience in KGE models, or in KGs, we also report the results here where any of the aspects of resilience (as described in Section 3) are discussed in the body of the paper.

Note that there exist a number of surveys discussing primarily two aspects of resilience proposed in this paper, such as robustness of deep learning models [28, 74, 113], language models [27, 48]; distribution adaption [64]. However, none of them give a definition of resilience considering the notions that we describe above and discuss the related works encompassing these aspects individually. Moreover, such a study is not done considering the KGE models. The closest work to ours are the works related to GNNs. Existing literature considering GNNs is quite vast and furthermore, there already exist surveys discussing the robustness [31, 114] and some other aspects of resilience of graph neural networks [123, 131]. Discussing the works related to the resilience of GNNs would extend this paper to a much greater extend. Therefore, we do not consider the GNNs in this paper.

Table 1 Paper query results. Here, "Body" represents the main content of a paper. Numbers correspond to the number of articles where the keyword occurred more than once.

Keywords	Title	Body
resilience in knowledge graphs	4	0
resilience in knowledge graph embedding models	2	6
generalization consistency in/of knowledge graphs	0	0
generalization consistency in/of knowledge graph embedding models	0	0
domain adaption in/of knowledge graph embedding models	0	2
distribution shift of knowledge graphs	1	0
in-distribution generalization of knowledge graph embedding models	1	0
robustness of knowledge graph embedding models	8	18

5 Resilience in KGE

The existing works considering resilience in KGE models mainly focus on a specific aspect, that is building KGE models that are resilient against the noise present in the KGs. To this end, there exist a number of such contributions [75, 86, 87, 111, 128]. These works consider if the performance of the model does not degrade with noise present in the KG, then the underlying KGE model is said to be resilient. This, however, is not resilience based on the definition provided in Section 3 where we defined resilience as a multi-faceted term that takes into account many aspects. Based on our definition of resilience, the work on resilience to safeguard against *noises* in the KGs mostly aligns with the definition of robustness, more specifically, the non-adversarial robustness (as defined in Equation 11). Therefore, we categorize this line of work related to resilience against noise as part of the non-adversarial robustness.

Moreover, there exist some works which concentrate on constructing resilient systems leveraging KGs [3, 30, 58, 117]. For instance, the works in [117] focus on building a KG-based risk assessment framework to improve the resiliency of supply chain management. A KG is built in [58] from the natural disaster data to improve the disaster management department's resilience towards such incidents. A similar sort of study is done in [3] to employ a resilient management system in case a crisis happens in a city. To assess the resiliency of the cyber-physical system for a water management system, Dagnas et al. [30] utilized the KG as a modeling graph.

Apart from the works mentioned earlier, no other works could be found that consider resilience (as we defined in this work) as part of KGs or KGE models, therefore, in this work, we further survey the existing literature by considering individual aspects of resilience as described in Equation (1)–(13). Figure 1 shows the categorization of the works found corresponding to the aspects of resilience that we described in Section 3. There exist works that focus on improving the generalizability of the KGE models [50, 55, 62, 107], or focusing on the logical consistency of the ontological rules [33, 37, 38, 51, 84], however, no work exists discussing generalization consistency, distribution adaption, and stability to incomplete input of KGE models. A very recent work by Zhu et al. [133] discusses the in-distribution generalization aspect and this is the only work that we could find related to this aspect of resilience.

More importantly, all the works found in regards to resilience in KGE can be distributed along two fields of robustness, namely adversarial and non-adversarial. There exists work such as by Zhu et al. [132] where they proposed to use KG to tackle the distribution shift problem for the *few-shot* learning approach. More specifically, by using KGs, the aim is to capture the semantic relationship between different categories of instances. Despite the data samples originating from

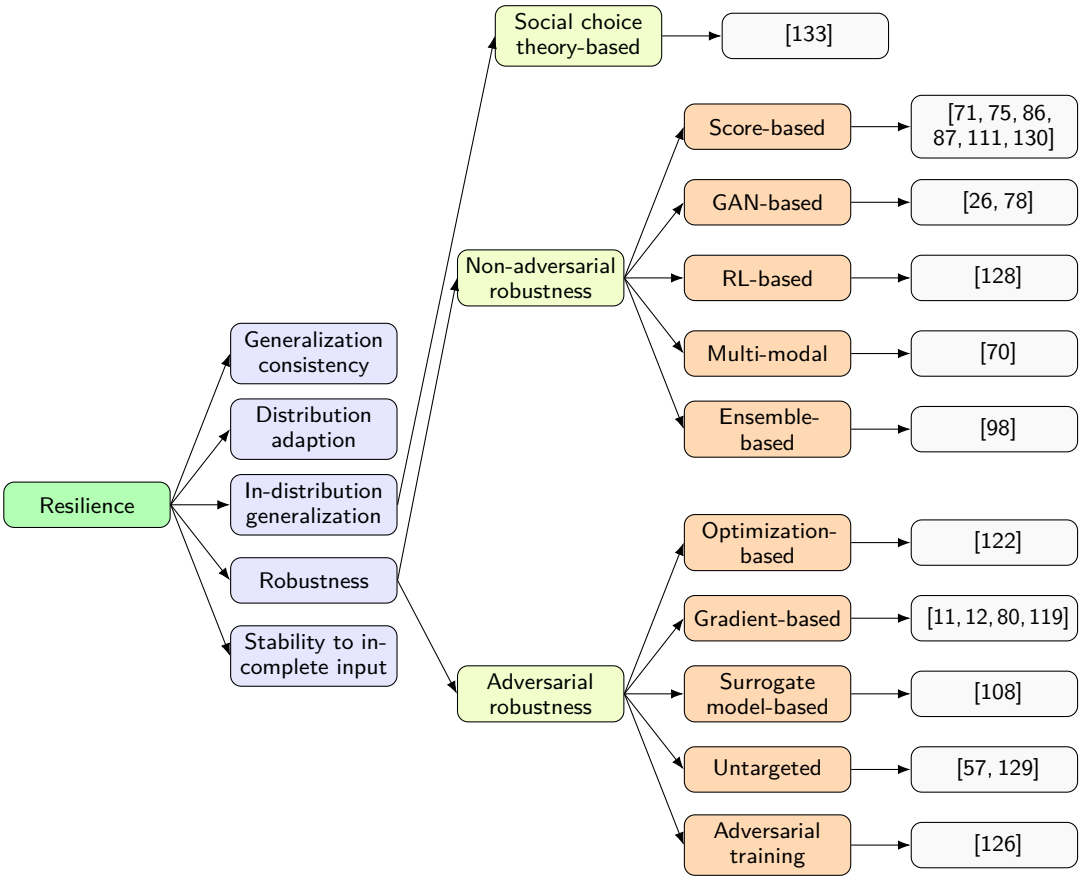

Figure 1 Categorisation of different works based on the underlying approaches in the context of resilience for knowledge graph embedding models.

diverse distributions, they frequently possess shared auxiliary knowledge, along with prior semantic relationships between classes. For this, KG can be used to find out when such a distribution shift occurs and help the underlying model to adapt.

Note that there exists *temporal* KG which dynamically evolves over time [56]. The idea therein is to model the temporal information in KGs to keep track of how different assertions/facts evolve over time. For this, the KG is defined as $\mathcal{G}_t := \{(h, r, t, t') \in \mathcal{E} \times \mathcal{R} \times \mathcal{E} \times \mathcal{T}\}$, where t' basically points to the timestamp for a specific fact. For instance, "(Barack Obama, President_of, USA, [2009-2017])" is an assertion associated with a timestamp [2009-2017] for which it is true. Temporal KGs are called dynamic KGs since they are not static and evolve with the addition of new timestamps corresponding to assertions, and there are KGE models that attempt to learn embeddings for such graphs, such as [2, 52, 92, 94, 120]. Although such KGE models learn to map entities and relations into an embedding space that changes over time, we do not expect the learned embeddings to adapt to the distribution shift of the KG. This is because such KGE approaches assume that, alike typical KGs, the distribution of temporal KGs does not change and only a new timestamp is added by replacing the old timestamp, for instance, (h, r, t, t') is replaced by (h, r, t, t^*). The underlying KGE models therefore only need to adapt based on this newly added timestamp. Therefore, we do not consider such works under distribution adaption aspect of resilience, and a survey on temporal KGE models can be found here [101].

Since there do not exist any works considering the aspects of resilience apart from robustness and in-distribution generalization, in the following, we first describe the related works encompassing the two areas that exist regarding the resilience of KGE.

6 Robustness

The concept of robustness in KGE models, as discussed earlier, can be divided into two main types: adversarial and non-adversarial robustness. As mentioned in Section 3, adversarial robustness concerns the model's ability to withstand intentional attacks, where malicious entities modify the knowledge graph (KG) to compromise the KGE model's performance. In contrast, non-adversarial robustness deals with the model's resilience against noise and inconsistencies naturally present in KGs, without any malicious intent. In the following subsections, we provide a detailed survey of the current research related to these two areas of robustness in KGE models. We begin with adversarial robustness, followed by an exploration of non-adversarial robustness techniques.

6.1 Adversarial Robustness

Despite its significance, the existing works on adversarial robustness in KGE models are in their infancy. The majority of studies focus on generating adversarial examples to deliberately manipulate the knowledge graph and assess the vulnerability of KGE models [11,12,80,108,119,122]. These works proposed methods for *attacking* the KGE models by generating adversarial examples to study the robustness of the existing KGE models. Below we categorize them based on the approaches used to perform such attacks.

6.1.1 Optimization Approach

One of the earlier works in this area [122] introduced a data poisoning attack strategy, aiming to alter the score of a target triple (h_t, r_t, t_t) by modifying the KG. To achieve the poisoning goal, they assumed the attacker had a fixed budget (for instance, like Δ in Equation 8) in terms of the number of changes that could be made on the KG. To this end, they have given two attack strategies, namely direct and indirect attack.

- **Direct attack.** The direct attack involves identifying the necessary perturbations, such as adding or removing triples, to achieve the attacker's objective – for example, reducing the likelihood of a target fact (h_t, r_t, t_t) being true. This process starts by determining the embedding shift ϵ required for either the head entity $\mathbf{h}_t$ or the tail entity $\mathbf{t}_t$ of the target triple to ensure that the new score $\phi'_\Theta(h_t, r_t, t_t)$, learned on the adversarially modified KG, is lower than the original score $\phi_\Theta(h_t, r_t, t_t)$. Potential perturbations are evaluated and ranked based on a scoring metric, guiding the selection of the most effective changes. The top M perturbations are then chosen using an optimization technique, taking into account the attacker's budget and constraints.
- **Indirect attack.** Performing a direct attack that involves shifting the embeddings of the target triple might be detected by using some kind of sanity check. Hence, to make the attack stealthy, the authors in [122] proposed indirect attack which involves shifting the embedding of the entities which are some k-hops away from the target triple (h_t, r_t, t_t). The changes would then propagate to the required embedding shifting of the target triple.

The adversarial attacks described above are performed considering the KG itself, and therefore, it adheres to Definition 9. While direct and indirect adversarial attacks on knowledge graph embeddings leverage optimization techniques to identify the most impactful perturbations, there are drawbacks when using such an approach in this process. One key limitation is that KGs have a highly discrete and complex structure, making it difficult to navigate the search space effectively using straightforward optimization.

6.1.2 Gradient-based and Attribution Attacks

Gradient-based approaches have emerged as a more effective alternative to performing adversarial attacks on the KGE models compared to simple optimization approaches. Note that the adversarial attack performed herein follows the definition of Equation 10. By leveraging the continuous embedding space of KGE models, gradient-based methods allow for a more effective exploration of potential perturbations. These approaches identify influential triples or paths in the KG by analyzing the gradient of the model's loss function with respect to the embeddings, enabling targeted modifications that maximize the attack's impact. Unlike optimization-based methods, gradient-based approaches offer computational advantages by operating in a lower-dimensional, continuous space, albeit with limitations in their applicability to specific types of KGE models.

Building on the ideas of direct and indirect attacks that target specific triples or entities in the knowledge graph, Pezeshkpour et al. [80] followed a typical gradient-based approach to find out the most influential neighboring triple (h'_t, r'_t, t_t) of the target triple (h_t, r_t, t_t), the removal $(\mathcal{G} \setminus \{(h'_t, r'_t, t_t)\})$ or addition $(\mathcal{G} \cup \{(h'_t, r'_t, t_t)\})$ of which would maximize the attack objective which can be defined as

$$\underset{(h'_t, r'_t)}{\operatorname{argmax}} \phi_\Theta(h_t, r_t, t_t) - \phi'_\Theta(h_t, r_t, t_t),$$

where $\phi'_\Theta(h_t, r_t, t_t)$ defines the score when trained on either $\mathcal{G} \setminus \{(h'_t, r'_t, t_t)\}$ or $\mathcal{G} \cup \{(h'_t, r'_t, t_t)\}$. However, searching for such a h'_t, r'_t is computationally expensive since the size of the search space is $|\mathcal{E}| \times |\mathcal{R}|$ (*number of entities in* $\mathcal{G}$ $\times$ *number of relations in* $\mathcal{G}$). Therefore, unlike the previous work [122], the authors herein modified the objective function by performing the search in the embedding domain, i.e., in the continuous space which gives the embedding for the optimal head and relation as $\mathbf{h}'_t, \mathbf{r}'_t$. Thereafter, an autoencoder is used to get h'_t, r'_t from $\mathbf{h}'_t, \mathbf{r}'_t$. However, one of the drawbacks of this approach is that it could only be used for multiplicative KGE models and moreover it does not take into account the nature of the KGE model being attacked.

Bhardwaj et al. [12] proposed a poisoning attack on KGE models by leveraging the inductive capabilities of these models, encapsulated through relationship patterns such as symmetry, inversion, and composition within a knowledge graph. Their approach aims to either decrease or increase the model's confidence in predicting a target triple h_t, r_t, t_t. For instance, if the attacker's goal is to decrease the score, they aim to ensure that $\phi_\Theta(h_t, r_t, t_t) > \phi'_\Theta(h, r, t)$, where ϕ'_Θ is the model learned on the KG modified with the addition of adversarial triples, referred to as *decoy triples.* These decoy triples are selected based on the inductive relation patterns that the KGE model captures. For example, if there exists a target triple h, r, t composed of $h_t, r_1, \bar{t}$ and $\bar{t}, r_2, t_t$, an additive model that captures the symmetry relationship can be exploited, such that $\mathbf{r}_1 + \mathbf{r}_2 = \mathbf{r}$. The model then selects a relation $\mathbf{r}_t$ as the target relation, minimizing the Euclidean distance $|\mathbf{r}_t - (\mathbf{r}_1 + \mathbf{r}_2)|$. By doing so, the method identifies the relation that strongly captures the symmetry. Once the target relation is chosen, two decoy triples are added in the form of h, r_1, t^* and t^*, r_2, t'. These added triples manipulate the inductive properties of the KGE model, indirectly decreasing the score of the original target triple h, r, t. By exploiting the underlying inductive patterns that KGE models learn, such as symmetry and composition, this approach makes the target triple less likely to be predicted as true.

Bhardwaj et al. [11] further extended their approach by employing instance attribution methods from the domain of interpretable machine learning to carry out data poisoning attacks on KGE models. The aim of these attacks remains similar to their previous work: reducing the likelihood of the target triple (h_t, r_t, t_t) being correctly predicted by the KGE model. They specifically defined the attacker's capability as the ability to make a single change (either by removing or adding a triple) within the neighborhood of the target triple. The neighborhood is constructed

based on triples that share either the subject or object of the target triple, formally defined as $\mathcal{H} = (h_n, r_n, t_n) \mid h_n \in h_t, t_t \vee t_n \in h_t, t_t$. To identify which triple should be manipulated, they introduced an *influence score* $\mathcal{I}((h_t, r_t, t_t), (h, r, t))$. This score measures the effect that a particular training triple (h, r, t) has on the model's prediction for the target triple (h_t, r_t, t_t). A larger influence score indicates that removing the triple (h, r, t) would significantly reduce the predicted score for (h_t, r_t, t_t). However, directly retraining the KGE model for each triple removal is computationally expensive. To tackle this, the authors adopted techniques from interpretable machine learning, specifically using similarity metrics in the embedding space.

You et al. [119] recently proposed a model-agnostic, semantic, and stealthy data poisoning attack on KGE models, addressing several aspects: black-box attack, semantically preserving poisoning, and stealthiness by ensuring good performance for clean triples. Unlike previous works, their approach focuses on inserting *indicative paths* rather than individual triples to maximize the prediction probability of a target poisoned triple. The attack goal can be formalized as

$$\max_{\hat{T}} \phi_{\Theta}(h_t, r_t, t_t),$$

where $\hat{T}$ is the set of triples in the indicative path. In their approach, the key idea is to add *indicative paths* that comprise more than one triple, which encourages the KGE model to predict the malicious fact as true. They translate the relation of the malicious fact into a sequence of relations using a *path template*. For example, a path template $p_{h_t \to t_t}$ could be $h_t \xrightarrow{r_1} e \xrightarrow{r_2} t_t$, where r_1, r_2 is a relation template, and e is an entity satisfying certain semantic constraints. The steps involve using the Path Ranking Algorithm (PRA) to generate candidate relation paths. Next, they leverage *semantic constraints* by selecting entities for the indicative paths that adhere to the domain and range constraints of the relations involved. The selection is carried out using a gradient-based search technique to find the indicative paths that maximize the prediction score for the target triple h_t, r_t, t_t. By ensuring that the added paths align with semantic constraints and maximize the plausibility of the malicious triple, their approach not only remains stealthy but also effectively biases the model's predictions towards the attacker's objective. This method is validated through extensive evaluations on benchmark datasets, demonstrating its effectiveness in achieving a high attack success rate under various opaque-box settings.

6.1.3 Surrogate Model-based Attack

Building on the approaches discussed in the previous sections, where gradient-based and attribution-based methods target specific triples or entities in the knowledge graph, surrogate model-based attacks introduce an alternative perspective. Instead of directly manipulating the embeddings or leveraging inductive patterns, these attacks employ an intermediate surrogate model to simulate the behavior of the original KGE model. By doing so, they enable the attacker to optimize adversarial manipulations in a more computationally efficient manner, particularly for downstream tasks where KGE models are used to answer user queries.

Xi et al. [108] introduced ROAR, an attack strategy designed to attack KGE models through both knowledge graph poisoning and *query misguiding*. ROAR particularly focuses on downstream applications where KGEs provide answers to user queries. The goal of the attack is to manipulate the response to a specific query by poisoning the knowledge graph in a manner that maximizes the probability of the targeted fact being true. The attack begins by generating a *surrogate knowledge graph* $\mathcal{G}'$ from the original one. This surrogate graph is used to build a *surrogate knowledge graph reasoner*, which consists of a surrogate embedding function ϕ' and a transformation function ψ. These functions are trained on a set of question-answer pairs sampled from $\mathcal{G}'$. The challenge here is that directly searching for poisoning facts that make the targeted fact true in the discrete space of the knowledge graph is computationally expensive.

To overcome this, the authors first employ *latent space optimization*. They search for an *anchor entity* connected to the target fact and identify facts in the embedding space which, when added, increase the probability of the targeted fact. These potential additions to the graph are gathered in a set of embeddings $\{\mathbf{h}_i, \mathbf{r}_i, \mathbf{t}_i\}_{i=1}^{N}$. Next, the effectiveness of adding each potential fact is assessed using a *fitness score*, which indicates how much each fact's addition would increase the plausibility of the target fact. Based on this score, the top n_g facts are selected for addition to the knowledge graph. This selection process ensures that only the most influential facts are included in the poisoning attack, thereby maximizing the impact on the targeted queries. This two-step process of latent space optimization followed by fitness-based selection makes ROAR a highly adaptable and effective adversarial attack against KGEs, especially in scenarios where downstream applications rely on the knowledge graph for query resolution.

6.1.4 Untargeted Attack

Apart from the adversarial attacks primarily focusing on making the KGE model perform badly on a specific triple, there exists a type of attack aiming to downgrade the overall accuracy of KGE models. This is referred to as untargeted attacks [57, 129], and so far only a few works have considered this. To this end, Zhao et al. [129] proposed a logic-rule-driven framework for conducting untargeted adversarial attacks on knowledge graph embeddings. The key idea herein is to perform adversarial additions or deletions that can systematically degrade overall model performance. To achieve this, the authors exploit logic rules that summarize global structural patterns in a KG. First, they use NCRL, a neural rule learning method [25], to extract high- and low-confidence rules from the graph. Based on these rules, they design two attack strategies, namely adversarial deletion and addition. In adversarial deletion, triples that strongly support high-confidence rules are removed, breaking reliable structural dependencies and preventing the model from learning accurate regularities. In contrast, in the case of addition, low-confidence rules are deliberately corrupted into non-existing rules which are then used to generate noisy triples. This then distorts the KG's semantics and encourages the model to capture misleading patterns. Therefore, the attacks do not focus on a specific target fact or triple, rather aim to disrupt the overall performance of the KGE models in the underlying tasks.

Based on a similar idea of performing untargeted attacks, Kapoor et al. [57] studied the robustness of KGE models considering three different *attack surfaces*, namely graph, parameter, and the labels. To this end, they first consider the knowledge graph perturbation, wherein a subset of triples from the KG is randomly modified by replacing either the head entity or the relation with another from the graph. This introduces structural inconsistencies without introducing new entities or relations. In parameter perturbation, embedding vectors are considered where the noise vectors are added directly to a subset of entity or relation embeddings during training. This is similar to the adversarial attacks like [5, 60, 67] where an attacker gains limited access to model parameters and subtly corrupts the representation space. Finally, in label perturbation, the label vectors used in training are inverted, flipping positives to negatives and vice versa.

6.1.5 Adversarial Training

Most of the works studied adversarial attack approaches for knowledge graph embedding models. To this end, we could find only the work by Zhang et al. [126] that focused on developing a defence approach against such attacks. They proposed a two-fold approach to improve the robustness of KGE models against adversarial perturbations. Firstly, by considering the adversarial training approach using GAN, the approach uses a generator–discriminator setup where the generator proposes adversarial perturbation triples and the discriminator learns to distinguish true from

perturbed triples. This forces the KGE model to become more resilient by directly training on adversarially crafted negatives. In the second step, to filter malicious triples from the graph, the authors propose subgraph-based detection methods. They focus on subgraphs around target triples, apply link prediction scores, and compare outputs of models trained on different subgraph partitions. This approach generates candidate completions from clean subgraphs to identify likely adversarial additions.

6.2 Non-adversarial Robustness

While adversarial robustness focuses on defending against malicious attacks, non-adversarial robustness concerns the model's resilience to naturally occurring noise and inconsistencies in KGs as defined in Equation 11. Real-world KGs are often incomplete, contain errors, and exhibit conflicting information due to the diverse sources from which they are constructed. A robust KGE model should be able to handle these imperfections without significantly compromising its performance. Several approaches have been proposed to improve the robustness of KGE models under noisy KGs, ranging from confidence score-based methods to GAN-based frameworks, reinforcement learning techniques, multi-modal approach, and ensemble approach [26, 70, 71, 75, 86, 87, 98, 111, 128, 130]. Below we discuss the existing works considering these approaches.

6.2.1 Confidence Score-based Approaches

Confidence score-based approaches have been proposed to enhance the robustness of the KGE models by quantifying the reliability of each triple within the KG. These methods assign a confidence score, trustworthiness value, or distance-based measure to each triple, allowing the model to prioritize more reliable data during training [71, 75, 86, 87, 111, 130]. The confidence scores guide the learning process, helping the model to distinguish between correct and noisy triples, thus reducing the impact of inaccuracies present in real-world KGs. In this section, we discuss several works that introduce different mechanisms for computing and utilizing confidence scores to improve the robustness of KGE models. These mechanisms range from local and global confidence scores to trustworthiness evaluations and distance-based assessments.

6.2.1.1 Local and Global Confidence Score

Xie et al. [111] introduced one of the earliest methods to address noise in knowledge graphs by developing KGE models that are robust to such noise. They proposed a novel approach known as the confidence-aware knowledge representation learning (CKRL) framework, which assigns a confidence score to each triple in the KG. This score indicates the correctness and significance of each triple, allowing the model to prioritize more reliable triples during learning. Their model builds upon the translation-based KGE approach, specifically utilizing TransE [16], as the scoring function ϕ_Θ. The standard margin-based ranking loss function [22] was modified to incorporate the confidence scores of triples. The revised objective function aims to minimize the impact of noisy triples by giving higher importance to more reliable triples. Specifically, they introduced the confidence-aware loss function:

$$\sum_{(h,r,t)\in S^+} \sum_{(h,r,x)\in S^-} [\gamma + \phi_\Theta(h,r,t) - \phi_\Theta(h,r,x)] \cdot C(h,r,t),$$

where γ is the margin, and S^+, S^- are the sets of positive and negative triples, respectively. Here, $C(h,r,t)$ is the confidence score for the triple h,r,t. A higher confidence score signals that the model should prioritize this triple during training. In essence, triples with lower scores are weighted less, which are essentially considered as noisy.

The computation of the confidence score $C(h, r, t)$ involves two components: local and global confidence scores as described below.

- **Local confidence score.** This score evaluates how well a triple conforms to the translation assumption within the KGE model. The triple's quality is updated iteratively during training. If a triple does not align with the translation rule, its confidence decreases by a geometric rate α. Conversely, if it does align, the confidence increases at a constant rate β. This iterative adjustment ensures that the confidence scores reflect the quality of triples over time.
- **Global confidence score.** Global confidence scores assess a triple's reliability by analyzing its broader structural context in the knowledge graph (KG). It consists of prior path confidence (PP), and adaptive path confidence (AP). PP measures how often a relation co-occurs with multi-step paths connecting the same entities. If similar paths frequently support the relation, PP is high. AP learns semantic similarity between a relation and its multi-step paths using embeddings. If a path relates to the target relation, AP is computed as high.

By combining these scores, CKRL effectively learns embeddings while simultaneously detecting and mitigating the influence of noise in the KG. This pioneering work laid the foundation for later developments in confidence-aware KGE models.

In a later work, Shan et al. [86] argued that the confidence score mechanism proposed by Xie et al. [111] could lead to the *zero loss problem.* This issue occurs when the negative triples sampled during training quickly fall outside the margin in the ranking loss function, resulting in zero loss. When this happens, the negative triples cease to contribute to refining the model's embeddings, leading to slow convergence, reduced accuracy, and diminished effectiveness in detecting noise within the knowledge graph. To address this problem, Shan et al. introduced a novel *confidence-aware negative sampling method.* They proposed a mechanism to assign a confidence score not just to positive triples but also to negative triples, with the goal of identifying high-quality negative triples that could contribute more significantly to the model's learning process. The key idea is to incorporate the confidence scores of negative triples into the training process.

Shao et al. [87] extended the confidence score-based methods by introducing a novel framework called DSKRL (Dissimilarity-Support-Aware Knowledge Representation Learning) to handle noise in KGs more effectively. Their approach incorporates two main components: triple dissimilarity and triple support, leveraging both structural and auxiliary information in KGs. While the former measures how well the entities and relations in a triple match, using entity hierarchical types and relation paths, the latter combines local and dynamic path support to assess a triple's credibility. After computing both the dissimilarity estimator and triple support, they are combined to improve the noise resilience in KGE models.

6.2.1.2 Trustworthiness Score

While confidence score-based methods focus on quantifying the reliability of individual triples through local and global assessments, trustworthiness score approaches extend this concept by leveraging semantic information and structural properties of entities and relations within the Knowledge Graph (KG). These methods aim to evaluate entities' inherent credibility and associations, refining the training process to prioritize trustworthy information. More specifically, such approaches differ from traditional confidence scores by incorporating additional semantic and contextual cues, such as entity types, descriptions, and path-based correlations. This integration allows for a more nuanced understanding of the data, enabling KGE models to better handle noise and inconsistencies in real-world KGs.

In [130], Zhao et al. proposed TransT, a method to compute the *trustworthiness value* of a triple by leveraging *entity types* and *descriptions*. The key idea is that certain entity types are more credible for specific relations. For example, a living entity (e.g., `/people/person`) is a more suitable subject for `was_born_in` than a non-living one (e.g., `/book/written_work`). TransT quantifies trustworthiness using two components:

- **Entity type trustiness (TT)** measures type compatibility for a relation:

$$TT(h,r,t) = \frac{1}{Z} \sum_{(h_i,t_i)\in\mathcal{T}(r)} \exp(-d(h_i,r,t_i)),$$

  where $\mathcal{T}(r)$ is the set of valid type pairs, and d measures alignment between types and relations.
- **Entity description trustiness (DT)** captures semantic consistency using cosine similarity:

$$DT(h,r,t) = \cos(\mathbf{d}_h + \mathbf{r}, \mathbf{d}_t),$$

where $\mathbf{d}_h$ and $\mathbf{d}_t$ are entity description embeddings, and $\mathbf{r}$ is the relation vector. The final trustworthiness score is a weighted combination:

$$T(h,r,t) = \alpha \cdot TT(h,r,t) + \beta \cdot DT(h,r,t),$$

where α and β control the contributions of the two factors. This trustworthiness score is integrated into the knowledge graph embedding model, prioritizing reliable triples during training.

While TransT focuses on assessing trustworthiness at the entity level, leveraging type compatibility and semantic descriptions, Ma et al. [71] take a structural approach with PTrustE by evaluating path trustworthiness and triple embeddings. Instead of relying solely on entity-level attributes, PTrustE incorporates path-based reasoning to detect noisy triples, capturing both local and global structural features within the knowledge graph. More specifically, given a triple (h,r,t), PTrustE first searches all paths between the head entity h and the tail entity t. Each path consists of a series of intermediate entities and relations, which are then used to compute both local and global trustworthiness scores. Specifically, two types of trustworthiness are introduced. local triple trustworthiness and global triple trustworthiness. In the absence of connecting paths, the confidence score of the triple relies more heavily on the local trustworthiness score derived from triple embeddings rather than path-based features. More specifically, PTrustE evaluates whether h and t are structurally disconnected or if they exist in separate KG components. The triple is likely to be erroneous if the entities belong to isolated graph fragments. In such cases, embedding-based similarity and logical constraints from the KG are used to assess plausibility, rather than path-based reasoning.

PTrustE focuses on detecting noise in KGs by leveraging path trustworthiness and probabilistic logic, since it primarily aims to filter out incorrect triples before embedding learning. An alternative approach to handling noisy triples is to directly modify the training objective rather than discarding them outright. Nayyeri et al. [75] introduced a modification to the marginal ranking loss function to handle noisy data in knowledge graphs (KGs), particularly focusing on incorrect triples. Their approach does not build on the previous confidence score-based works but instead introduces a distance-based strategy to identify and manage noisy triples effectively. In their method, the authors define separate objective functions for positive and negative triples and then combine them into a unified loss function. One key component of their approach is a distance function, which intuitively measures the likelihood of a triple being correct or noisy. During the optimization process, this distance is constrained to lie within the range $[0,\gamma]$, where γ serves as a discriminator that separates positive and negative triples. A probability function is employed to assign a score

based on the computed distance. A high probability indicates a high likelihood of the triple being incorrect (noisy), whereas a lower probability suggests a higher confidence in the triple's correctness. The objective is to minimize the overall loss by maximizing the likelihood of correct triples and minimizing the likelihood of noisy ones.

To summarize, confidence score-based methods improve the performance of KGE models by adjusting the influence of noisy triples during training. These methods vary in how they compute and integrate confidence scores, leveraging different aspects of local consistency, global structural reasoning, and adaptive loss functions. Early approaches like CKRL [111] introduced confidence-aware learning by assigning local and global confidence scores to triples, refining embeddings iteratively. However, CKRL suffered from the zero-loss problem, where negative triples quickly became uninformative. To mitigate this, Shan et al. [86] proposed a confidence-aware negative sampling strategy, dynamically selecting high-quality negative triples to improve training effectiveness. Expanding beyond embeddings, DSKRL [87] integrated semantic knowledge, such as entity types and relation paths, to compute confidence scores, improving robustness against inconsistencies but requiring additional structured information. Alternatively, PTrustE [71] introduced a path-based trustworthiness framework, assessing global and local triple reliability through correlation networks and probabilistic logic. While effective in structured graphs, PTrustE is computationally expensive and less applicable to sparse KGs. A distinct approach was taken by Nayyeri et al. [75], who modified the ranking loss function to incorporate a distance-based confidence score, adjusting training weights dynamically instead of explicitly assigning confidence scores. This method avoids reliance on heuristic scoring functions but requires fine-tuning distance thresholds for optimal performance. Embedding-based approaches are computationally efficient but can struggle with noisy negatives, whereas semantic-aware models (DSKRL) improve interpretability but depend on auxiliary knowledge. Path-based trustworthiness methods (PTrustE) enhance global reasoning but introduce high complexity, and distance-based confidence models (Nayyeri et al.) provide a principled alternative at the cost of hyperparameter sensitivity. The optimal choice, therefore, depends on dataset characteristics, noise levels, and computational constraints.

6.2.2 GAN-based Approaches

While confidence score and trustworthiness score approaches address noise by quantifying the reliability of triples or entities based on structural and semantic properties, Generative Adversarial Network (GAN)-based approaches adopt a more dynamic mechanism. These methods introduce an adversarial framework to detect and mitigate noise in Knowledge Graphs (KGs) by simultaneously learning to generate and classify noisy triples. More specifically, GAN-based approaches leverage adversarial training to refine the embeddings by detecting and mitigating the impact of noisy triples during training. By generating synthetic noisy triples and training the model to differentiate between true and noisy triples, these methods ensure that the learned embeddings remain robust.

NoiGAN [26] extends the idea of confidence score proposed in Section 6.2.1. They argued, similar to the previously described approaches, that using only the confidence score as an indication of how well a triple fits to the KGE model might lead to bias and uncertainty. Therefore, the confidence score $C(h, r, t)$ in this work is learned by using a *generator* and *discriminator* as a generative adversarial network (GAN). More specifically, they proposed a learning framework inspired by the adversarial training [8, 63, 73] methods. In the GAN framework, NoiGAN consists of two main components: a generator and a discriminator. The generator is designed to generate noisy triples, while the discriminator is trained to distinguish between true and noisy triples, ultimately computing the confidence score for each triple. During training, the KGE model uses this confidence score as a guiding signal to eliminate noisy data.

Given a true triple (h, r, t), the generator generates a noisy triple (h', r, t') from an initially generated negative sample candidate set $\mathcal{N}(h, r, t)$. This is achieved through a neural network that takes as input the embedding vectors of the triple (h', r, t') and outputs a probability indicating the plausibility of the triple being noisy. More formally, the generator aims to maximize the expected reward:

$$R_G = \sum_{(h,r,t)} \mathbb{E}_{(h',r,t') \sim G(\cdot|(h,r,t);\Theta_G)}[\log f_D(h', r, t')],$$

where $f_D(h', r, t')$ is the probability predicted by the discriminator that the generated triple (h', r, t') is true. The generator uses reinforcement learning to generate triples that can effectively fool the discriminator. The discriminator, on the other hand, acts as a noisy triple classifier. It aims to distinguish between true triples and noisy triples generated by the generator.

Apart from the embedding models, the GAN-based approach is also used in KG-based systems such as in the entity-alignment approach. To this end, Pie et al. [78] propose an approach to make robust cross-lingual entity alignment between KGs by incorporating noise detection into the alignment process using a generative adversarial network (GAN)-based approach [46]. The model consists of a Graph Neural Network (GNN) for entity embedding and a Generative Adversarial Network (GAN) for noise detection. The GAN therein consists of a generator G and a discriminator D. The generator generates fake entity pairs, while the discriminator assigns a trust score $T(e_1, e_2)$ to distinguish correct and noisy pairs. To align entities across KGs, a margin-based ranking loss is used to bring correct entity pairs closer together and push noisy pairs further apart.

6.2.3 Reinforcement Learning Approaches

Reinforcement Learning (RL) approaches take a different perspective by formulating noise detection and triple selection as a decision-making problem. RL-based methods focus on improving the robustness of KGE models by systematically identifying and removing noisy triples before the training process begins. This proactive approach ensures that the KGE models are trained on cleaner datasets, leading to more reliable embeddings. A recent work by Zhang et al. [128] proposes a multi-task reinforcement learning (RL) framework to make the KGE models robust by identifying and removing noisy triples from the training dataset. Unlike previous approaches that directly train on noisy datasets, this method first cleans the dataset before the training process, ensuring that the KGE models are learned on a noise-free graph. The authors define the state, action, reward, and the objective of the RL framework in the following manner.

State. Each state in RL is represented as the set of triples that have already been selected as clean and the current triple that is under consideration. Mathematically, the state at time step t can be defined as $s_t = (T_{\text{selected}}, (h, r, t))$, where T_{selected} is the set of triples that have already been marked as clean up to time t, and (h, r, t) is the triple being evaluated.

Action. The RL agent takes an action to either select or reject the triple (h, r, t). The action space A consists of binary decisions, $A = \{0, 1\}$, where 1 indicates selecting the triple as clean, and 0 indicates rejecting it.

Reward. The reward function is designed based on the scoring functions of multiple KGE models like TransE, DistMult, ConvE, or RotatE, along with a heuristic term that encourages the model to select more triples. The reward R for a set of selected triples T_{selected} is calculated as

$$R = \frac{1}{|T_{\text{selected}}|} \sum_{(h,r,t) \in T_{\text{selected}}} \phi_\Theta(h, r, t) + \alpha \frac{|T_{\text{selected}}|}{|T_{\text{total}}|},$$

where α is a hyperparameter, and $|T_{\text{total}}|$ is the total number of triples in the KG.

Objective. The aim of the RL model is to maximize the expected reward by selecting those triples that exhibit higher plausibility.

The authors highlight that this approach has the potential drawback of filtering out a large number of triples, which could include some correct triples. However, the RL framework's use of scoring functions from different KGE models helps to mitigate this by making decisions based on the inferred relationships and plausibility scores.

6.2.4 Multi-modal Knowledge Representation

Multi-modal methods aim to combine information from different knowledge sources to better capture the semantics and context of entities and relations within the knowledge graph. This integration enables the model to mitigate the effects of noise in a single modality by relying on complementary information from other modalities. To this end, the work closest to the idea of robustness of KGE models is done by Lu et al. [70] where they propose multi-modal knowledge representation learning (MMKRL) to generate robust KGE models. The idea therein is to use several knowledge such as textual knowledge, entity description, visual knowledge to generate the embedding [95, 105, 110]. MMKRL essentially consists of two main modules: knowledge reconstruction and adversarial training, where the knowledge reconstruction module aligns and integrates various knowledge embeddings to reconstruct multi-modal knowledge graphs, while the training module enhances robustness and performance using adversarial strategies.

6.2.5 Ensemble Approaches

Ensemble-based approaches combine multiple models trained on diverse subgraphs of the Knowledge Graph (KG). This strategy leverages the principle that an ensemble of learners can outperform individual models, especially in the presence of noise or inconsistencies in the data. By aggregating predictions from multiple models, ensemble-based approaches mitigate the impact of errors or biases present in a single model. Wan et al. [98] proposed an ensemble-based approach to enhance the robustness of the KGE models. Their method involves generating a set of diverse subgraphs from a given KG $\mathcal{G}$ and training an individual base learner for each subgraph.

Due to the complexity of KGs, traditional graph sampling methods are not directly applicable. To address this, Wan et al. employ a random walk-based approach [69] to sample meaningful subgraphs. The random walk process starts by selecting an initial fact (h, r, t) uniformly at random from the KG $\mathcal{G}$. Then, the random walk samples a neighbor of the current node, following the relations in the KG. This sampling continues until a predefined boundary condition L (e.g., a maximum path length or number of nodes) is met. After executing multiple random walks, a set of subgraphs $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_n\}$ is generated. For each subgraph $\mathcal{G}_i$, a shallow KGE model ϕ_{Θ_i} is trained independently to obtain entity and relation embeddings. The model's goal is to learn an embedding function $\phi_{\Theta_i}(h, r, t)$ that maximizes the plausibility of triples in the subgraph. The final ensemble model combines the outputs of these n base learners. Let ϕ_{ensemble} represent the final embedding function, which is defined as a weighted combination of the individual models:

$$\phi_{\text{ensemble}}(h, r, t) = \sum_{i=1}^{n} \alpha_i \phi_{\Theta_i}(h, r, t),$$

where α_i is the weight assigned to model ϕ_{Θ_i} based on its prediction performance. The weights α_i are determined by an uncertainty measure, which reflects the predictive capability of each model on its corresponding subgraph. For example, the uncertainty can be calculated using entropy or variance in the predictions. The robustness of this ensemble approach is then evaluated by injecting noise into the KG $\mathcal{G}$. Wan et al. demonstrate that their ensemble model performs significantly better than individual KGE models in the presence of noisy triples.

6.3 Comparative Analysis and Future Directions

The previous sections explored various approaches to enhance the robustness of KGE models when KGs contain noisy triples. These approaches differ in how they detect and mitigate noise, with some focusing on explicit confidence estimation, others leveraging adversarial learning, and some incorporating external multimodal information. In this section, we contrast these methods, analyze their respective strengths and limitations, and propose potential hybrid strategies to further enhance resilience. To better understand how different robustness approaches complement or compete with one another, first of all, we categorize them based on key aspects such as information used for noise handling, adaptability to different types of noise, and computational complexity as summarized in Table 2.

Table 2 Comparison of different noise-robust KGE approaches.

Approach	Key Mechanism	Strengths	Limitations
Confidence score	Assigns confidence scores to triples based on local/global plausibility	Adaptive to structured noise, interpretable	Struggles with adversarial noise, requires careful calibration
Trust score	Uses entity type and path-based information to determine trustworthiness	Strong semantic reasoning, robust to inconsistencies	Relies on well-defined entity types, limited adaptability
GAN	Generator-discriminator model to filter noise iteratively	Dynamically adapts to different noise patterns	Training instability, risk of mode collapse
RL-based	Reinforcement learning selects reliable triples pretraining	Generalizes well, avoids overfitting to noise	Filtering errors can lead to knowledge loss
Multi-modal	Uses text and images to supplement KG information	Effective for missing or ambiguous data	Requires external data sources, computationally expensive
Ensemble	Aggregates predictions from multiple KGE models	Improves generalization and robustness	Computational overhead, limited effect in adversarial settings

Furthermore, given the strengths and limitations of individual methods, a promising direction is to develop hybrid approaches that integrate complementary techniques. Below, we suggest three strategies to enhance robustness by combining different noise-handling mechanisms.

Confidence score with GAN-based noise correction. Confidence-based methods provide an effective first step in detecting structured noise, while GAN-based filtering adapts dynamically to unstructured noise. A potential hybrid model could first use a confidence estimator, such as CKRL, to assign preliminary confidence scores to triples. These confidence scores help distinguish between highly reliable triples and those suspected as being noisy. Once the confidence scores are assigned, the high-confidence triples can be fed directly into a standard KGE training process to learn robust embeddings from cleaner data. Meanwhile, the low-confidence triples, which are more likely to contain noise, are passed into a GAN-based filtering mechanism. The GAN consists of a generator that produces synthetic noise and a discriminator that learns to distinguish between correct and incorrect triples. During training, the discriminator iteratively refines its decision boundary by learning from both real and generated noisy triples.

RL-based filtering with multi-modal learning. RL-based models excel at identifying and removing highly noisy triples, making them suitable to use as an initial data-cleaning step before training a multi-modal KGE model. The combination of RL and multi-modal learning allows for more effective noise reduction while leveraging complementary knowledge sources to enhance embeddings. For instance, an RL agent could be trained to evaluate triples based on a reward function that incorporates multiple KGE scoring functions. The agent iteratively selects high-confidence triples while discarding unreliable ones. Once the RL agent filters out noisy triples, a multi-modal KGE model is trained exclusively on the cleaned dataset. This model integrates information from textual descriptions, entity attributes, and visual embeddings to improve the quality of entity and relation representations. The multi-modal embeddings can further be used to refine the RL-based filtering in a feedback loop. If a previously filtered triple gains support from external modalities (e.g., a missing relation is inferred via textual descriptions), it may be reintroduced into the knowledge graph.

Ensemble learning with path-based trustworthiness scores. Ensemble learning enhances robustness by aggregating predictions from multiple KGE models, while path-based trustworthiness scoring ensures that models are weighted based on their reliability in capturing meaningful entity-relation patterns. Instead of training a single KGE model, multiple models are trained on different subgraphs generated through random walks, clustering-based sampling, or relation-specific partitions. Thereafter, each entity pair in the KG is evaluated based on the reliability of intermediate paths connecting them. Approaches like PTrustE [71] could be used to score paths based on semantic consistency, redundancy, and coherence with established entity-type constraints. The final embedding for a given entity or relation is determined by aggregating the predictions from the ensemble models, weighted according to their path-based trustworthiness scores. Models that perform better on structurally supported paths contribute more to the final representation. These hybrid approaches offer promising directions for improving the robustness of KGE models.

7 Robustness of KG-based Systems

There are some works that do not directly discuss the robustness of the KGE models; however, they consider the KG-driven systems, such as entity linking [72], cross-lingual entity alignment [78], knowledge-grounded dialogue system [100], improving the robustness of the facts of KG [109].

Mao et al. [72] propose a robust entity linking method that tackles 3 aspects, namely, inefficient graph encoders, the need for negative sampling, and catastrophic forgetting in semi-supervised learning. To improve the graph encoders therein, they use relational attention to update the entity features. Furthermore, the authors prove that negative samples are unnecessary in entity linking. It adopts a symmetric negative-free alignment loss to align entity pairs without generating negative samples thereby removing the need for negative samples, and aligning entity pairs with the loss function. Finally, to mitigate catastrophic forgetting, the approach stores previously learned embeddings and selectively reviews them during each training iteration. This approach allows the model to maintain alignment accuracy without retraining on all previous data. The evaluation has shown state-of-the-art results with improved robustness.

Pei et al. [78] propose REA (Robust Entity Alignment), a method for cross-lingual entity alignment between noisy knowledge graphs (KGs). Existing entity alignment models assume clean labeled data, but in real-world scenarios, labeled entity pairs often contain errors that degrade the alignment quality. REA first encodes the structure of knowledge graphs using a Graph Neural Network (GNN). The GNN processes entities and their relationships within each knowledge graph, learning meaningful embeddings that capture the structural similarities between

entities, even if they exist in different languages. For example, if "Eiffel Tower" in an English knowledge graph has the same connections as "Tour d'Eiffel" in a French knowledge graph, their embeddings should be similar. The approach further introduces a trust score for each labeled entity pair. The trust score acts as a measure of confidence, determining how reliable a given entity alignment is. REA also uses a margin-based ranking loss function. This function ensures that correctly aligned entity pairs have their embeddings placed closer together, while incorrect pairs are pushed further apart in the learned space. The noise detection module within REA, which operates using an adversarial training framework, continuously updates the trust scores based on newly identified errors. In turn, the noise-aware entity alignment module adjusts the entity embeddings based on these refined trust scores. This iterative learning process ensures that the model becomes increasingly accurate, filtering out noise while improving the quality of entity alignment. An extensive evaluation on real-world multilingual knowledge graph datasets such as DBP15K, DWY100K [93] show that REA outperforms state-of-the-art methods (e.g., GCN-Align [104], MuGNN [19]) in noisy settings. REA provides a robust approach for integrating multilingual knowledge graphs, ensuring high-quality entity alignment despite label noise.

Wang et al. [100] introduce an entity-based contrastive learning framework, named EnCo, to enhance the robustness of knowledge-grounded dialogue (KGD) systems. Given a dialogue context $C = \{u_1, u_2, \ldots, u_{n-1}\}$ consisting of utterances u_i and an external knowledge set $K = \{(h_1, r_1, t_1), \ldots, (h_m, r_m, t_m)\}$ comprising knowledge triples where h_i, r_i, and t_i represent the head entity, relation, and tail entity respectively, the goal of a KGD system is to generate a response u_n based on C and K. The authors aim to enhance the robustness of KGD models to handle real-world perturbations, including semantic-irrelevant (e.g., misspellings, paraphrasing) and semantic-relevant (e.g., incorrect entity replacements) perturbations. To this end, they leverage contrastive learning to improve robustness by constructing positive and negative samples and training the model to recognize semantic similarities and differences.

Xiao et al. [109] address the problem of evaluating the robustness of *outstanding* facts (OFs) derived from KGs. An OF is defined as a statement highlighting how an entity stands out based on specific attributes when compared to its peers. Consider a KG $\mathcal{G}$ containing information about universities and their employees, including attributes like gender. An OF from this KG might state: "At the American Council on Education (ACE), only 31% of the employees are male." This statement could suggest a notable gender disparity at ACE-affiliated institutions. However, the robustness of this fact needs to be evaluated by considering the broader context and possible data variations. To formalize this, Xiao et al. introduce the concept of robustness by analyzing how the "strikingness" of an OF changes under various perturbations. Let $\mathcal{S}(f)$ denote the strikingness of an OF f in a given context. The goal is to ensure that $\mathcal{S}(f)$ remains consistent even when the context or data changes slightly. The authors propose two types of perturbations to evaluate this:

- **Entity perturbation.** It assesses the robustness of an OF by replacing its context entity c with a similar entity c'. Formally, let c represent the context entity in the OF f. We replace c with c', where c' is chosen based on its similarity to c. The similarity between entities c and c' is computed as

$$\mathrm{Sim}(c, c') = \frac{|N(c) \cap N(c')|}{|N(c) \cup N(c')|},$$

  where $N(c)$ and $N(c')$ are the sets of neighbors of c and c', respectively.
- **Data perturbation.** It involves modifying the KG by adding or altering edges, thereby changing the peer entity set of the OF. Formally, the relevance of a data perturbation is quantified using a head-tail relevance function, which measures the semantic connection of the newly added edges to the original fact. Given an added edge (h', r', t'), the head-tail relevance function $\mathcal{R}(h', r', t')$ evaluates whether the modification preserves the context's semantic integrity.

The robustness of an OF is then defined by the expected strikingness $\mathbb{E}_{p(\mathcal{P})}[\mathcal{S}(f)]$ over a perturbation relevance distribution (PRD) $p(\mathcal{P})$. This method of evaluating robustness relates to earlier discussions in the literature on robustness, specifically, similar to our proposed robustness formalization in Equation 10. Much like ensuring that KGE models are resilient against adversarial attacks and noise (e.g., as described in works like Xie et al. [111] and Shan et al. [86]), evaluating OFs for robustness ensures that their interpretations remain valid across different contexts.

8 Robustness Improvement Using Knowledge Graphs

Note that, similar to using KGs to improve the resilience of several systems, there also exist a number of works that use KGs [83, 118] and KGE models [61] to improve the robustness of ML models. However, the notion of robustness therein pertains to the effectiveness of performing the underlying tasks. Below we describe some of them.

Multi-object detection. Lang et al. [61] propose the use of KGEs to develop more robust multi-object detection models. The main idea is to use KGEs to incorporate semantic knowledge into object detection, aiming to achieve more structured and semantically grounded predictions. Traditional object detection models often use a one-hot encoding approach, treating object classes as discrete and unrelated. This method maximizes inter-class distances but ignores the semantic relationships between different object types. The authors therein introduce a new formulation where they replace these learnable class prototypes with fixed object type embeddings derived from knowledge graphs. Specifically, the object detector learns to map visual features into a semantic embedding space, using either word embeddings (like GloVe) [79] or embeddings derived directly from knowledge graphs using any standard KGE models. In their evaluation, this approach demonstrated more semantically grounded misclassifications, meaning the errors made by the model were often more contextually appropriate. Additionally, their evaluation on benchmark datasets showed that KGE-based models matched or even outperformed traditional one-hot methods, particularly in challenging object detection benchmarks.

Deep learning. Radtke et al. [83] propose using KGs to enhance deep learning models for fault diagnostics in prognostics and health management (PHM). They introduce a KG-enhanced deep learning approach to incorporate domain-invariant knowledge, improving model robustness and generalization. The method leverages the structure of KGs to encode semantic information hierarchically and combines this with supervised contrastive learning to create a more stable feature representation. Experimental results demonstrate that this approach increases the model's ability to handle domain shifts, making fault diagnostics more resilient across varying conditions.

Recommender system. Yang et al. [118] propose knowledge graph contrastive learning (KGCL) to suppress noise and enhance item representations in recommender systems. Their approach addresses challenges such as *long-tail entity distributions* and *noisy, topic-irrelevant connections* in Knowledge Graphs (KGs). More specifically, to improve robustness, KGCL generates two perturbed views of the KG, $\mathcal{G}_1$ and $\mathcal{G}_2$, by randomly dropping edges. This introduces structural perturbations, allowing the model to learn robust embeddings by contrasting entity representations across different views. KGCL then employs contrastive learning to maximize agreement between the same entity's embeddings in different views while minimizing similarity with other entities that are not close.

9 In-distribution Generalization

The work by Zhu et al. [133] is the only work that contributes to this aspect of resilience. Therein, they define this as *predictive multiplicity*, a phenomenon where multiple models with similar accuracy make conflicting predictions for the same query. The authors conduct an empirical study on multiple KGE models and datasets to measure predictive multiplicity. For each KGE algorithm, they train multiple models with different random initializations and hyperparameters. They then select a set of "competing" models – those whose link prediction performance is virtually the same as a best baseline model (within a small tolerance ϵ, e.g. 1% difference in Hits@K). Using this set of models, the authors evaluate how often their predictions diverge. For each test query (a partially specified triple such as $(h, r, ?)$), they compare the model's top-ranked results. If one model's top answer is different from another's, that query is counted as a conflicting case. The ambiguity metric is computed as the percentage of test queries with any such conflict, and discrepancy reflects the maximum disagreement rate among the models. These metrics provide a quantitative measure of predictive multiplicity for the link prediction task.

After measuring the extent of conflicting predictions, the authors apply ensemble voting methods to combine model outputs. Each model in the competing set produces a ranked list of candidate entities for a query. The authors apply three voting schemes to aggregate these rankings into one result,

1. majority voting which picks the candidate that appears as the top choice for most models,
2. borda voting, which assigns points based on rank position (e.g., a candidate gets more points for being ranked 1st, slightly fewer for 2nd, and so on, across all models) and then selects the candidate with the highest total points, and
3. range voting, which uses the actual prediction scores from each model (rescaled to a common range) and sums them up for each candidate.

These methods generate an aggregated ranking intended to reflect a consensus. The impact of aggregation is assessed by recomputing the ambiguity and discrepancy metrics on the combined ranking, and by checking the standard accuracy metrics (Hits@K) to ensure that the ensemble prediction is still performing well.

10 Challenges and Future Works

Future research in the domain of resilience on knowledge graphs and KGE models presents a number of possibilities to improve different aspects of resilience that we defined in this work. We can envisage works aiming at developing KGE models considering generalization consistency, distribution adaptation, and in-distribution generalization, amongst others. We describe future work directions in more detail in the following.

Generalization under Distribution Shift. One promising avenue for future work is the development of resilient KGE models that can adaptively adjust to changes in the underlying data or graph structure. Traditional KGE models often assume static or stationary environments, which may not hold in dynamic or evolving KGs. Future research could explore dynamic embedding techniques that continuously update entity and relation embeddings to capture temporal or contextual changes, for instance, when new entities or relations are added to the KG. Additionally, integrating uncertainty modeling and probabilistic reasoning mechanisms into KGE models could enhance their resilience to noisy or uncertain data. Some existing works attempt to quantify uncertainty in KGE models, such as *probabilistic soft logic-based methods* in [23] and confidence-aware embedding techniques in [66]. The work in [23] employs probabilistic soft logic to generate confidence scores capturing structural and assertional

uncertainties, enabling the model to provide confidence-based predictions when new entities and relations are introduced. Similarly, [66] defines a KG as uncertain when each assertion is associated with a confidence score, which is integrated into the KGE learning process to adjust predictions dynamically. These studies establish an important foundation for making KGE models aware of distribution shifts by incorporating uncertainty estimation.

To further enhance robustness against distribution shifts, conformal prediction could be incorporated into KGE models. Conformal prediction provides a mathematically sound framework for quantifying the uncertainty of model predictions by constructing prediction sets that offer guaranteed coverage probabilities [44]. Instead of generating point estimates, KGE models could output prediction intervals for link prediction tasks, ensuring that the true answer is included within a certain confidence level. Note that there already exist some works incorporating this technique in KGE models, such as [134] where the authors apply conformal prediction theory, which enables uncertainty-aware answer set prediction by ensuring that the correct answer is included within a generated answer set with probabilistic guarantees. This shows the potential of utilizing conformal prediction in KGE models in dealing with resilience. For instance, adaptive conformal prediction techniques could be applied to KG completion tasks, where the KGE model dynamically updates its uncertainty estimates as new data arrives. When distribution shifts occur, such as new entities being introduced or relationships evolving, the conformal predictor could adjust its confidence intervals accordingly. This is particularly useful in real-world applications, where decision-making systems rely on KGE models and require calibrated confidence scores for each prediction.

Besides uncertainty-aware learning and conformal prediction, other techniques could be explored to increase the robustness of KGE models in dynamic environments such as:

Bayesian knowledge graph embeddings. Instead of learning fixed embeddings, a Bayesian approach would model entity and relation embeddings as probability distributions (e.g., using Gaussian distributions) [96], allowing the model to express uncertainty in predictions explicitly. This would be particularly effective in scenarios where distribution shifts occur.

Meta-learning for KGE adaptation. A meta-learning framework could be designed to quickly adapt KGE models when distribution shifts occur. Few-shot learning techniques, such as Model-Agnostic Meta-Learning (MAML) [43], could be used to train KGE models to generalize across different graph structures with minimal re-training. This would be beneficial in dynamic knowledge graphs, where new domains or unseen entities frequently appear.

Contrastive learning for distribution shift detection. Contrastive learning techniques could be integrated to detect and quantify shifts in graph structures [45]. By learning embedding distances between past and present snapshots of a KG, models can determine when a significant shift has occurred and retrain embeddings accordingly. This approach could also be combined with self-supervised learning, enabling KGE models to update embeddings without requiring extensive labeled data.

Incorporating conformal prediction, Bayesian embeddings, meta-learning, and contrastive learning into KGE models could significantly enhance their ability to handle distribution shifts and noisy data. Note that, the list of approaches mentioned here is not exhaustive, and there could be further techniques to tackle the aforementioned problems. Future research should delve into exploring more such approaches and find out explore how these techniques can be efficiently integrated into KGE pipelines while ensuring computational scalability and real-time adaptation capabilities.

Adversarial and non-adversarial robustness. Another possible research direction is that of resilience of KGE models against adversarial attacks and manipulations, i.e., developing KGE models that are adversarially robust. As mentioned beforehand, real-world KGs might suffer

from adversarial attacks where adversaries may attempt to exploit vulnerabilities in the KG or KGE models to inject false information, manipulate inference results, or disrupt system functionality. There have already been some works to this end, however, all of them focus on developing methods to perform targeted attacks, i.e., considering a specific fact to add or remove from the KG and thereby making the KGE model learn based on the attackers' goal. To this end, only the KG has been considered as a possible attack surface. However, there can be other possibilities, for instance, the parameters of the already trained KGE model can be attacked. Such kind of attacks are often prevalent in the ML domain and termed Trojan attacks [67, 68, 102] where the attacker aim to make the model learn their objective either by generating inputs with certain *triggers*, or by changing the already trained model's parameters. For KGE models, such Trojan attacks could correspond to the modification of the entries of learned embedding vectors so as to achieve a specific attacker's objective. Apart from considering targeted attacks by taking into account different attack surfaces, it would also be needed to consider performing non-targeted attacks [57], where the idea is to simply disrupt the performance of the underlying KGE models by introducing noise in the KGs or in the KGE models. As mentioned previously, there are already some works which considered such kind of attacks. However, more sophisticated attack approaches to this end could be explored.

Additionally, the targeted attacks so far have been considered only for a specific type of task, namely link prediction tasks. KGE models are used in many critical downstream application tasks [32, 42, 99], and hence, more research is needed to understand how to perform adversarial attacks on such KGE-based tasks. This basically opens up a number of different attack surfaces along with the need to explore different attack dimensions, including non-targeted attacks.

While several works considered adversarial attacks on KGE models, a much-needed direction to be focused on is the development of defence mechanisms against such attacks that can detect and mitigate adversarial attacks in real-time, thereby enhancing the overall robustness of KG-based applications. This would include developing graph-based anomaly detection algorithms to identify and mitigate adversarial attacks or abnormal patterns in the KG, performing adversarial training of KGE algorithms, developing certified guaranteed methods to build robust KGE models, and so on. Furthermore, defence mechanisms should be extended to combat non-targeted attacks, effectively addressing noise, or incompleteness inherent in KGs. This entails the creation of robust data integration and ensemble algorithms capable of handling diverse and noisy information from various sources. Moreover, exploring techniques for automated error detection, correction, and data validation within KGs could significantly enhance their quality and reliability over time.

Recently, the works to combine large language models (LLMs) with KGEs are gaining popularity [77]. There is a potential that by augmenting LLMs to KGEs, one could achieve improved robustness. Leveraging the semantic richness of natural language representations encoded in LLMs, such as BERT [36] or GPT [17], may enhance the understanding and representation of entities and relations in the KG. This integration could potentially mitigate the impact of noisy or incomplete KGs on downstream tasks.

Stability to incomplete inputs. Existing works primarily focus on handling missing data through imputation or data augmentation rather than explicitly ensuring robustness against missingness. The lack of standardized evaluation benchmarks and theoretical formulations of stability further hinders progress in this area. Therefore, to this end, we first of all require a benchmark to evaluate different approaches. Afterward, we require suitable graph similarity metrics depending on different domains, moving beyond standard metrics like graph edit distance or Jaccard similarity. Furthermore, we can envisage using adversarial training [18] where missing elements are simulated during training to improve resilience. To this end, *dropouts* for KGEs, randomly

removing nodes/edges during training could also help. Knowledge distillation technique is quite useful in learning where a model trained on complete KGs transfers knowledge to one dealing with incomplete KGs. Apart from the training techniques, adapting loss functions could also be considered. For instance, a stability-aware loss function could be designed that explicitly penalize drastic embedding changes due to missing data, ensuring bounded divergence within ϵ. To this end, furthermore, a consistency regularization technique, where models minimize differences in predictions from full vs. incomplete graphs could also be used.

Finally, resilience has already been vastly explored in fault-tolerant systems, therefore, interdisciplinary approaches that draw insights from fields such as network science, complex systems theory, and resilience engineering could provide valuable perspectives and methodologies for enhancing the resilience of KGs and KGE models. By leveraging principles from these domains, researchers can develop holistic, multi-faceted strategies for improving the reliability and robustness of KG-based systems in diverse application domains.

11 Conclusion

In this work, we explored the resilience of knowledge graph embedding models, addressing their ability to withstand and adapt to various challenges such as noise, adversarial attacks, and dynamic changes in the underlying knowledge graphs. While significant research has been conducted on robustness, particularly adversarial robustness, there is a pressing need to consider a more comprehensive notion of resilience. This broader understanding includes aspects such as generalization consistency, distribution adaption, and performance stability under diverse real-world conditions. A key finding of this survey is that while adversarial robustness has received considerable attention, with various strategies to perform attacks on KGE models, resilience in non-adversarial contexts is equally critical. Models must not only defend against malicious interventions but also maintain their reliability in the presence of natural noise and inconsistencies prevalent in real-world KGs. The surveyed works on non-adversarial robustness primarily focus on mitigating the effects of noise by incorporating confidence-aware learning and enhanced negative sampling strategies. However, these approaches often overlook the dynamic nature of KGs, particularly temporal and evolving KGs, where distribution shifts are inevitable. Addressing such shifts through adaptive retraining mechanisms remains an open challenge. Moreover, ensuring in-distribution generalization across diverse application domains is essential. KGE models must be able to operate effectively even with incomplete input data, which is a common scenario in real-world applications. Achieving this consistency demands that future research goes beyond traditional robustness frameworks, integrating novel methodologies from graph neural networks, reinforcement learning, and explainable AI to enhance both the adaptability and transparency of KGE models. In conclusion, while much progress has been made in improving the robustness of KGE models, a more holistic approach to resilience – incorporating adaptability, consistency, and robustness in the face of both adversarial and natural challenges – will be key to unlocking the full potential of these models in real-world, dynamic, and noisy environments.

References

1 Alekh Agarwal and Tong Zhang. Minimax regret optimization for robust machine learning under distribution shift. In Po-Ling Loh and Maxim Raginsky, editors, *Conference on Learning Theory, 2-5 July 2022, London, UK*, volume 178 of *Proceedings of Machine Learning Research*, pages 2704–2729. PMLR, 2022. URL: `https://proceedings.mlr.press/v178/agarwal22b.html`.

2 Saadullah Amin, Stalin Varanasi, Katherine Ann Dunfield, and Günter Neumann. Lowfer: Low-rank bilinear pooling for link prediction. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July*

2020, Virtual Event, volume 119 of *Proceedings of Machine Learning Research*, pages 257–268. PMLR, 2020. URL: http://proceedings.mlr.press/v119/amin20a.html.

3 Amin Anjomshoaa, Hannah Schuster, Johannes Wachs, and Axel Polleres. From data to insights: constructing spatiotemporal knowledge graphs for city resilience use cases. In *Second International Workshop On Linked Data-driven Resilience Research 2023*, 2023.

4 Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. Dbpedia: A nucleus for a web of open data. In *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007*, volume 4825 of *Lecture Notes in Computer Science*, pages 722–735. Springer, 2007. doi:10.1007/978-3-540-76298-0_52.

5 Jiawang Bai, Baoyuan Wu, Yong Zhang, Yiming Li, Zhifeng Li, and Shu-Tao Xia. Targeted attack against deep neural networks via flipping limited weight bits. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL: https://openreview.net/forum?id=iKQAk8a2kM0.

6 Ivana Balažević, Carl Allen, and Timothy M Hospedales. Hypernetwork knowledge graph embeddings. In *Artificial Neural Networks and Machine Learning–ICANN 2019: Workshop and Special Sessions: 28th International Conference on Artificial Neural Networks, Munich, Germany, September 17–19, 2019, Proceedings 28*, pages 553–565. Springer, 2019.

7 Ivana Balažević, Carl Allen, and Timothy M Hospedales. Tucker: Tensor factorization for knowledge graph completion. *arXiv preprint arXiv:1901.09590*, 2019.

8 Mislav Balunovic and Martin T. Vechev. Adversarial training and provable defenses: Bridging the gap. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL: https://openreview.net/forum?id=SJxSDxrKDr.

9 Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Mach. Learn.*, 79(1-2):151–175, 2010. doi:10.1007/S10994-009-5152-4.

10 Christian Berger, Philipp Eichhammer, Hans P. Reiser, Jörg Domaschka, Franz J. Hauck, and Gerhard Habiger. A survey on resilience in the iot: Taxonomy, classification, and discussion of resilience mechanisms. *ACM Comput. Surv.*, 54(7):147:1–147:39, 2022. doi:10.1145/3462513.

11 Peru Bhardwaj, John D. Kelleher, Luca Costabello, and Declan O'Sullivan. Adversarial attacks on knowledge graph embeddings via instance attribution methods. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 8225–8239. Association for Computational Linguistics, 2021. doi:10.18653/v1/2021.emnlp-main.648.

12 Peru Bhardwaj, John D. Kelleher, Luca Costabello, and Declan O'Sullivan. Poisoning knowledge graph embeddings via relation inference patterns. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 1875–1888. Association for Computational Linguistics, 2021. doi:10.18653/v1/2021.acl-long.147.

13 Steffen Bickel, Michael Brückner, and Tobias Scheffer. Discriminative learning for differing training and test distributions. In Zoubin Ghahramani, editor, *Machine Learning, Proceedings of the Twenty-Fourth International Conference (ICML 2007), Corvallis, Oregon, USA, June 20-24, 2007*, volume 227 of *ACM International Conference Proceeding Series*, pages 81–88. ACM, 2007. doi:10.1145/1273496.1273507.

14 Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250, 2008. doi:10.1145/1376616.1376746.

15 Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26, 2013.

16 Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2787–2795, 2013. URL: https://proceedings.neurips.cc/paper/2013/hash/1cecc7a77928ca8133fa24680a88d2f9-Abstract.html.

17 Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information*

Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, 2020. URL: https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html.

18 Liwei Cai and William Yang Wang. KBGAN: Adversarial learning for knowledge graph embeddings. In Marilyn Walker, Heng Ji, and Amanda Stent, editors, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1470–1480, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi:10.18653/v1/N18-1133.

19 Yixin Cao, Zhiyuan Liu, Chengjiang Li, Juanzi Li, and Tat-Seng Chua. Multi-channel graph neural network for entity alignment. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1452–1461, 2019.

20 Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr., and Tom M. Mitchell. Toward an architecture for never-ending language learning. In Maria Fox and David Poole, editors, *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010*, pages 1306–1313. AAAI Press, 2010. doi:10.1609/AAAI.V24I1.7519.

21 Ines Chami, Adva Wolf, Da-Cheng Juan, Frederic Sala, Sujith Ravi, and Christopher Ré. Low-dimensional hyperbolic knowledge graph embeddings. *arXiv preprint arXiv:2005.00545*, 2020. arXiv:2005.00545.

22 Wei Chen, Tie-Yan Liu, Yanyan Lan, Zhiming Ma, and Hang Li. Ranking measures and loss functions in learning to rank. In Yoshua Bengio, Dale Schuurmans, John D. Lafferty, Christopher K. I. Williams, and Aron Culotta, editors, *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a meeting held 7-10 December 2009, Vancouver, British Columbia, Canada*, pages 315–323. Curran Associates, Inc., 2009. URL: https://proceedings.neurips.cc/paper/2009/hash/2f55707d4193dc27118a0f19a1985716-Abstract.html.

23 Xuelu Chen, Muhao Chen, Weijia Shi, Yizhou Sun, and Carlo Zaniolo. Embedding uncertain knowledge graphs. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 3363–3370. AAAI Press, 2019. doi:10.1609/AAAI.V33I01.33013363.

24 Yongqiang Chen, Yonggang Zhang, Yatao Bian, Han Yang, Kaili Ma, Binghui Xie, Tongliang Liu, Bo Han, and James Cheng. Learning causally invariant representations for out-of-distribution generalization on graphs. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL: http://papers.nips.cc/paper_files/paper/2022/hash/8b21a7ea42cbcd1c29a7a88c444cce45-Abstract-Conference.html.

25 Kewei Cheng, Nesreen K. Ahmed, and Yizhou Sun. Neural compositional rule learning for knowledge graph reasoning. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL: https://openreview.net/forum?id=F8VKQyDgRVj.

26 Kewei Cheng, Yikai Zhu, Ming Zhang, and Yizhou Sun. Noigan: Noise aware knowledge graph embedding with adversarial learning. In *ICLR 2020 Conference*, 2020. URL: https://api.semanticscholar.org/CorpusID:226951634.

27 Arijit Ghosh Chowdhury, Md Mofijul Islam, Vaibhav Kumar, Faysal Hossain Shezan, Vaibhav Kumar, Vinija Jain, and Aman Chadha. Breaking down the defenses: A comparative survey of attacks on large language models. *CoRR*, abs/2403.04786, 2024. doi:10.48550/arXiv.2403.04786.

28 Joana C Costa, Tiago Roxo, Hugo Proença, and Pedro RM Inácio. How deep learning sees the world: A survey on adversarial attacks & defenses. *IEEE Access*, 2024.

29 Gabriela Csurka. Deep visual domain adaptation. In *22nd International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, SYNASC 2020, Timisoara, Romania, September 1-4, 2020*, pages 1–8. IEEE, 2020. doi:10.1109/SYNASC51798.2020.00013.

30 Romain Dagnas, Michel Barbeau, Joaquin Garcia-Alfaro, and Reda Yaich. Resilience assessment of multi-layered cyber-physical systems. In *IFIP Networking 2024-IOCRCI*, 2024.

31 Enyan Dai, Tianxiang Zhao, Huaisheng Zhu, Junjie Xu, Zhimeng Guo, Hui Liu, Jiliang Tang, and Suhang Wang. A comprehensive survey on trustworthy graph neural networks: Privacy, robustness, fairness, and explainability. *Mach. Intell. Res.*, 21(6):1011–1061, 2024. doi:10.1007/s11633-024-1510-8.

32 Jeffrey Dalton, Laura Dietz, and James Allan. Entity query feature expansion using knowledge base links. In Shlomo Geva, Andrew Trotman, Peter Bruza, Charles L. A. Clarke, and Kalervo Järvelin, editors, *The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '14, Gold Coast , QLD, Australia - July 06 - 11, 2014*, pages 365–374. ACM, 2014. doi:10.1145/2600428.2609628.

33 Thomas Demeester, Tim Rocktäschel, and Sebastian Riedel. Lifted rule injection for relation embeddings. In Jian Su, Xavier Carreras, and Kevin Duh, editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 1389–1399. The Association for Computational Linguistics, 2016. doi:10.18653/V1/D16-1146.

34 Caglar Demir, Diego Moussallem, Stefan Heindorf, and Axel-Cyrille Ngonga Ngomo. Convolutional hypercomplex embeddings for link prediction. In *Asian Conference on Machine Learning*, pages 656–671. PMLR, 2021. URL: https://proceedings.mlr.press/v157/demir21a.html.

35 Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

36 Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics, 2019. doi:10.18653/V1/N19-1423.

37 Jianfeng Du, Kunxun Qi, and Yuming Shen. Knowledge graph embedding with logical consistency. In Maosong Sun, Ting Liu, Xiaojie Wang, Zhiyuan Liu, and Yang Liu, editors, *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data - 17th China National Conference, CCL 2018, and 6th International Symposium, NLP-NABD 2018, Changsha, China, October 19-21, 2018, Proceedings*, volume 11221 of *Lecture Notes in Computer Science*, pages 123–135. Springer, 2018. doi:10.1007/978-3-030-01716-3_11.

38 Jianfeng Du, Kunxun Qi, Hai Wan, Bo Peng, Shengbin Lu, and Yuming Shen. Enhancing knowledge graph embedding from a logical perspective. In Zhe Wang, Anni-Yasmin Turhan, Kewen Wang, and Xiaowang Zhang, editors, *Semantic Technology - 7th Joint International Conference, JIST 2017, Gold Coast, QLD, Australia, November 10-12, 2017, Proceedings*, volume 10675 of *Lecture Notes in Computer Science*, pages 232–247. Springer, 2017. doi:10.1007/978-3-319-70682-5_15.

39 Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard S. Zemel. Fairness through awareness. In Shafi Goldwasser, editor, *Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8-10, 2012*, pages 214–226. ACM, 2012. doi:10.1145/2090236.2090255.

40 Shaohua Fan, Xiao Wang, Chuan Shi, Peng Cui, and Bai Wang. Generalizing graph neural networks on out-of-distribution graphs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(1):322–337, 2024. doi:10.1109/TPAMI.2023.3321097.

41 Uriel Feige, Yishay Mansour, and Robert E. Schapire. Learning and inference in the presence of corrupted inputs. In Peter Grünwald, Elad Hazan, and Satyen Kale, editors, *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, volume 40 of *JMLR Workshop and Conference Proceedings*, pages 637–657. JMLR.org, 2015. URL: http://proceedings.mlr.press/v40/Feige15.html.

42 David A. Ferrucci, Eric W. Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John M. Prager, Nico Schlaefer, and Christopher A. Welty. Building watson: An overview of the deepqa project. *AI Mag.*, 31(3):59–79, 2010. doi:10.1609/AIMAG.V31I3.2303.

43 Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 1126–1135. PMLR, 2017. URL: http://proceedings.mlr.press/v70/finn17a.html.

44 A. Gammerman, V. Vovk, and V. Vapnik. Learning by transduction. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, UAI'98, pages 148–155. Morgan Kaufmann Publishers Inc., 1998.

45 Saurabh Garg, Amrith Setlur, Zachary C. Lipton, Sivaraman Balakrishnan, Virginia Smith, and Aditi Raghunathan. Complementary benefits of contrastive learning and self-training under distribution shift. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL: http://papers.nips.cc/paper_files/paper/2023/hash/26f96550613971371c5d07f37f0e06c0-Abstract-Conference.html.

46 Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial nets. In Zoubin Ghahramani, Max Welling, Corinna Cortes, Neil D. Lawrence, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 2672–2680, 2014. URL: https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html.

47 Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL: http://arxiv.org/abs/1412.6572.

48 Shreya Goyal, Sumanth Doddapaneni, Mitesh M. Khapra, and Balaraman Ravindran. A survey of adversarial defenses and robustness in nlp. *ACM Comput. Surv.*, 55(14s), July 2023. doi:10.1145/3593042.

49 Martin Grohe and Pascal Schweitzer. The graph isomorphism problem. *Commun. ACM*, 63(11):128–134, October 2020. doi:10.1145/3372123.

50 Shu Guo, Quan Wang, Bin Wang, Lihong Wang, and Li Guo. SSE: semantically smooth embedding for knowledge graphs. *IEEE Trans. Knowl. Data Eng.*, 29(4):884–897, 2017. doi:10.1109/TKDE.2016.2638425.

51 Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Jointly embedding knowledge graphs and logical rules. In Jian Su, Xavier Carreras, and Kevin Duh, editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 192–202. The Association for Computational Linguistics, 2016. doi:10.18653/V1/D16-1019.

52 Peng He, Gang Zhou, Hongbo Liu, Yi Xia, and Ling Wang. Hyperplane-based time-aware knowledge graph embedding for temporal knowledge graph completion. *J. Intell. Fuzzy Syst.*, 42(6):5457–5469, 2022. doi:10.3233/JIFS-211950.

53 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. Knowledge graphs. *ACM Computing Surveys (CSUR)*, 54(4):1–37, 2021. doi:10.1145/3447772.

54 Jiayuan Huang, Alexander J. Smola, Arthur Gretton, Karsten M. Borgwardt, and Bernhard Schölkopf. Correcting sample selection bias by unlabeled data. In Bernhard Schölkopf, John C. Platt, and Thomas Hofmann, editors, *Advances in Neural Information Processing Systems 19, Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 4-7, 2006*, pages 601–608. MIT Press, 2006. URL: https://proceedings.neurips.cc/paper/2006/hash/a2186aa7c086b46ad4e8bf81e2a3a19b-Abstract.html.

55 Nitisha Jain, Trung-Kien Tran, Mohamed H. Gad-Elrab, and Daria Stepanova. Improving knowledge graph embeddings with ontological reasoning. In Andreas Hotho, Eva Blomqvist, Stefan Dietze, Achille Fokoue, Ying Ding, Payam M. Barnaghi, Armin Haller, Mauro Dragoni, and Harith Alani, editors, *The Semantic Web - ISWC 2021 - 20th International Semantic Web Conference, ISWC 2021, Virtual Event, October 24-28, 2021, Proceedings*, volume 12922 of *Lecture Notes in Computer Science*, pages 410–426. Springer, 2021. doi:10.1007/978-3-030-88361-4_24.

56 Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and S Yu Philip. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE transactions on neural networks and learning systems*, 33(2):494–514, 2021. doi:10.1109/TNNLS.2021.3070843.

57 Sourabh Kapoor, Arnab Sharma, Michael Röder, Caglar Demir, and Axel-Cyrille Ngonga Ngomo. Robustness evaluation of knowledge graph embedding models under non-targeted attacks. In Edward Curry, Maribel Acosta, María Poveda-Villalón, Marieke van Erp, Adegboyega K. Ojo, Katja Hose, Cogan Shimizu, and Pasquale Lisena, editors, *The Semantic Web - 22nd European Semantic Web Conference, ESWC 2025, Portoroz, Slovenia, June 1-5, 2025, Proceedings, Part I*, volume 15718 of *Lecture Notes in Computer Science*, pages 264–281. Springer, 2025. doi:10.1007/978-3-031-94575-5_15.

58 Seonhyeong Kim and Young-Woo Kwon. Construction of disaster knowledge graphs to enhance disaster resilience. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 6721–6723, 2022. doi:10.1109/BigData55660.2022.10021017.

59 Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

60 Keita Kurita, Paul Michel, and Graham Neubig. Weight poisoning attacks on pretrained models. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 2793–2806. Association for Computational Linguistics, 2020. doi:10.18653/v1/2020.acl-main.249.

61 Christopher Lang, Alexander Braun, and Abhinav Valada. Robust object detection using knowledge graph embeddings. In Björn Andres, Florian Bernard, Daniel Cremers, Simone Frintrop, Bastian Goldlücke, and Ivo Ihrke, editors, *Pattern Recognition - 44th DAGM German Conference, DAGM GCPR 2022, Konstanz, Germany, September 27-30, 2022, Proceedings*, volume 13485 of *Lecture Notes in Computer Science*, pages 445–461. Springer, 2022. doi:10.1007/978-3-031-16788-1_27.

62 Chengjiang Li, Yixin Cao, Lei Hou, Jiaxin Shi, Juanzi Li, and Tat-Seng Chua. Semi-supervised entity alignment via joint knowledge embedding model and cross-graph model. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 2723–2732. Association for Computational Linguistics, 2019. doi:10.18653/V1/D19-1274.

63 Jintang Li, Jiaying Peng, Liang Chen, Zibin Zheng, Tingting Liang, and Qing Ling. Spectral adversarial training for robust graph neural network. *IEEE Trans. Knowl. Data Eng.*, 35(9):9240–9253, 2023. doi:10.1109/TKDE.2022.3222207.

64 Jian Liang, Ran He, and Tieniu Tan. A comprehensive survey on test-time adaptation under distribution shifts. *Int. J. Comput. Vis.*, 133(1):31–64, 2025. doi:10.1007/S11263-024-02181-W.

65 Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29, 2015.

66 Qi Liu, Qinghua Zhang, Fan Zhao, and Guoyin Wang. Uncertain knowledge graph embedding: an effective method combining multi-relation and multi-path. *Frontiers Comput. Sci.*, 18(3):183311, 2024. doi:10.1007/S11704-023-2427-Z.

67 Yingqi Liu, Shiqing Ma, Yousra Aafer, Wen-Chuan Lee, Juan Zhai, Weihang Wang, and

Xiangyu Zhang. Trojaning attack on neural networks. In *25th Annual Network and Distributed System Security Symposium, NDSS 2018, San Diego, California, USA, February 18-21, 2018*. The Internet Society, 2018. URL: https://www.ndss-symposium.org/wp-content/uploads/2018/02/ndss2018_03A-5_Liu_paper.pdf.

68 Yunfei Liu, Xingjun Ma, James Bailey, and Feng Lu. Reflection backdoor: A natural backdoor attack on deep neural networks. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part X*, volume 12355 of *Lecture Notes in Computer Science*, pages 182–199. Springer, 2020. doi:10.1007/978-3-030-58607-2_11.

69 László Lovász. Random walks on graphs. *Combinatorics, Paul erdos is eighty*, 2(1-46):4, 1993.

70 Xinyu Lu, Lifang Wang, Zejun Jiang, Shichang He, and Shizhong Liu. MMKRL: A robust embedding approach for multi-modal knowledge graph representation learning. *Appl. Intell.*, 52(7):7480–7497, 2022. doi:10.1007/S10489-021-02693-9.

71 Jiangtao Ma, Chenyu Zhou, Yanjun Wang, Yifan Guo, Guangwu Hu, Yaqiong Qiao, and Yong Wang. Ptruste: A high-accuracy knowledge graph noise detection method based on path trustworthiness and triple embedding. *Knowl. Based Syst.*, 256:109688, 2022. doi:10.1016/J.KNOSYS.2022.109688.

72 Xin Mao, Wenting Wang, Yuanbin Wu, and Man Lan. Are negative samples necessary in entity alignment?: An approach with high performance, scalability and robustness. In Gianluca Demartini, Guido Zuccon, J. Shane Culpepper, Zi Huang, and Hanghang Tong, editors, *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021*, pages 1263–1273. ACM, 2021. doi:10.1145/3459637.3482232.

73 Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual adversarial training: A regularization method for supervised and semi-supervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 41(8):1979–1993, 2019. doi:10.1109/TPAMI.2018.2858821.

74 Awais Muhammad and Sung-Ho Bae. A survey on efficient methods for adversarial robustness. *IEEE Access*, 10:118815–118830, 2022. doi:10.1109/ACCESS.2022.3216291.

75 Mojtaba Nayyeri, Sahar Vahdati, Emanuel Sallinger, Mirza Mohtashim Alam, Hamed Shariat Yazdi, and Jens Lehmann. Pattern-aware and noise-resilient embedding models. In Djoerd Hiemstra, Marie-Francine Moens, Josiane Mothe, Raffaele Perego, Martin Potthast, and Fabrizio Sebastiani, editors, *Advances in Information Retrieval - 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 - April 1, 2021, Proceedings, Part I*, volume 12656 of *Lecture Notes in Computer Science*, pages 483–496. Springer, 2021. doi:10.1007/978-3-030-72113-8_32.

76 Maximilian Nickel, Volker Tresp, Hans-Peter Kriegel, et al. A three-way model for collective learning on multi-relational data. In *Icml*, volume 11, pages 3104482–3104584, 2011.

77 Jeff Z. Pan, Simon Razniewski, Jan-Christoph Kalo, Sneha Singhania, Jiaoyan Chen, Stefan Dietze, Hajira Jabeen, Janna Omeliyanenko, Wen Zhang, Matteo Lissandrini, Russa Biswas, Gerard de Melo, Angela Bonifati, Edlira Vakaj, Mauro Dragoni, and Damien Graux. Large language models and knowledge graphs: Opportunities and challenges. *TGDK*, 1(1):2:1–2:38, 2023. doi:10.4230/TGDK.1.1.2.

78 Shichao Pei, Lu Yu, Guoxian Yu, and Xiangliang Zhang. Rea: Robust cross-lingual entity alignment between knowledge graphs. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2175–2184, 2020. doi:10.1145/3394486.3403268.

79 Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1532–1543. ACL, 2014. doi:10.3115/V1/D14-1162.

80 Pouya Pezeshkpour, Yifan Tian, and Sameer Singh. Investigating robustness and interpretability of link prediction via adversarial modifications. In *1st Conference on Automated Knowledge Base Construction, AKBC 2019, Amherst, MA, USA, May 20-22, 2019*, 2019. URL: https://openreview.net/forum?id=Hkg7rbcp67.

81 Natasa Przulj. Biological network comparison using graphlet degree distribution. *Bioinform.*, 26(6):853–854, 2010. doi:10.1093/BIOINFORMATICS/BTQ091.

82 Joaquin Quionero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D. Lawrence. *Dataset Shift in Machine Learning*. The MIT Press, 2009.

83 Maximilian-Peter Radtke, Marco Huber, and Jürgen Bock. Increasing robustness of data-driven fault diagnostics with knowledge graphs. In *Proceedings of the Annual Conferenceof the PHM Society 2023*. PHM Society, 2023.

84 Tim Rocktäschel, Sameer Singh, and Sebastian Riedel. Injecting logical background knowledge into embeddings for relation extraction. In Rada Mihalcea, Joyce Yue Chai, and Anoop Sarkar, editors, *NAACL HLT 2015, The 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Denver, Colorado, USA, May 31 - June 5, 2015*, pages 1119–1129. The Association for Computational Linguistics, 2015. doi:10.3115/V1/N15-1118.

85 Sanjit A. Seshia, Ankush Desai, Tommaso Dreossi, Daniel J. Fremont, Shromona Ghosh, Edward Kim, Sumukh Shivakumar, Marcell Vazquez-Chanlatte, and Xiangyu Yue. Formal specification for deep neural networks. In Shuvendu K. Lahiri and Chao Wang, editors, *Automated Technology for Verification and Analysis - 16th International Symposium,*

ATVA 2018, Los Angeles, CA, USA, October 7-10, 2018, Proceedings, volume 11138 of *Lecture Notes in Computer Science*, pages 20–34. Springer, 2018. doi:10.1007/978-3-030-01090-4_2.

86 Yingchun Shan, Chenyang Bu, Xiaojian Liu, Shengwei Ji, and Lei Li. Confidence-aware negative sampling method for noisy knowledge graph embedding. In Xindong Wu, Yew-Soon Ong, Charu C. Aggarwal, and Huanhuan Chen, editors, *2018 IEEE International Conference on Big Knowledge, ICBK 2018, Singapore, November 17-18, 2018*, pages 33–40. IEEE Computer Society, 2018. doi:10.1109/ICBK.2018.00013.

87 Tianyang Shao, Xinyi Li, Xiang Zhao, Hao Xu, and Weidong Xiao. DSKRL: A dissimilarity-support-aware knowledge representation learning framework on noisy knowledge graph. *Neurocomputing*, 461:608–617, 2021. doi:10.1016/J.NEUCOM.2021.02.099.

88 Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of statistical planning and inference*, 90(2):227–244, 2000.

89 Alexander J. Smola, Arthur Gretton, Le Song, and Bernhard Schölkopf. A hilbert space embedding for distributions. In Marcus Hutter, Rocco A. Servedio, and Eiji Takimoto, editors, *Algorithmic Learning Theory, 18th International Conference, ALT 2007, Sendai, Japan, October 1-4, 2007, Proceedings*, volume 4754 of *Lecture Notes in Computer Science*, pages 13–31. Springer, 2007. doi:10.1007/978-3-540-75225-7_5.

90 Lorenzo Strigini. Fault tolerance and resilience: Meanings, measures and assessment. In Katinka Wolter, Alberto Avritzer, Marco Vieira, and Aad P. A. van Moorsel, editors, *Resilience Assessment and Evaluation of Computing Systems*, pages 3–24. Springer, 2012. doi:10.1007/978-3-642-29032-9_1.

91 Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*, pages 697–706, 2007. doi:10.1145/1242572.1242667.

92 Jinze Sun, Yongpan Sheng, Ling Zhan, and Lirong He. TKGR-RHETNE: A new temporal knowledge graph reasoning model via jointly modeling relevant historical event and temporal neighborhood event context. In Biao Luo, Long Cheng, Zheng-Guang Wu, Hongyi Li, and Chaojie Li, editors, *Neural Information Processing - 30th International Conference, ICONIP 2023, Changsha, China, November 20-23, 2023, Proceedings, Part V*, volume 14451 of *Lecture Notes in Computer Science*, pages 331–343. Springer, 2023. doi:10.1007/978-981-99-8073-4_26.

93 Zequn Sun, Qingheng Zhang, Wei Hu, Chengming Wang, Muhao Chen, Farahnaz Akrami, and Chengkai Li. A benchmarking study of embedding-based entity alignment for knowledge graphs. *Proceedings of the VLDB Endowment*, 13(11):2326–2340, 2020. URL: http://www.vldb.org/pvldb/vol13/p2326-sun.pdf.

94 Xiaoli Tang, Rui Yuan, Qianyu Li, Tengyun Wang, Haizhi Yang, Yundong Cai, and Hengjie Song. Timespan-aware dynamic knowledge graph embedding by incorporating temporal evolution. *IEEE Access*, 8:6849–6860, 2020. doi:10.1109/ACCESS.2020.2964028.

95 Xing Tang, Ling Chen, Jun Cui, and Baogang Wei. Knowledge representation learning with entity descriptions, hierarchical types, and textual relations. *Inf. Process. Manag.*, 56(3):809–822, 2019. doi:10.1016/J.IPM.2019.01.005.

96 Armin Toroghi and Scott Sanner. Bayesian inference with complex knowledge graph evidence. In Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan, editors, *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*, pages 20550–20558. AAAI Press, 2024. doi:10.1609/AAAI.V38I18.30040.

97 Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080. PMLR, 2016. URL: http://proceedings.mlr.press/v48/trouillon16.html.

98 Guojia Wan, Bo Du, Shirui Pan, and Jia Wu. Adaptive knowledge subgraph ensemble for robust and trustworthy knowledge graph completion. *World Wide Web*, 23(1):471–490, 2020. doi:10.1007/S11280-019-00711-Y.

99 Hongwei Wang, Fuzheng Zhang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. Multi-task feature learning for knowledge graph enhanced recommendation. In Ling Liu, Ryen W. White, Amin Mantrach, Fabrizio Silvestri, Julian J. McAuley, Ricardo Baeza-Yates, and Leila Zia, editors, *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, pages 2000–2010. ACM, 2019. doi:10.1145/3308558.3313411.

100 Jiaan Wang, Jianfeng Qu, Kexin Wang, Zhixu Li, Wen Hua, Ximing Li, and An Liu. Improving the robustness of knowledge-grounded dialogue via contrastive learning. In Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan, editors, *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*, pages 19135–19143. AAAI Press, 2024. doi:10.1609/AAAI.V38I17.29881.

101 Jiapu Wang, Boyue Wang, Meikang Qiu, Shirui Pan, Bo Xiong, Heng Liu, Linhao Luo, Tengfei Liu, Yongli Hu, Baocai Yin, and Wen Gao. A survey on temporal knowledge graph completion: Taxonomy, progress, and prospects. *CoRR*, abs/2308.02457, 2023. doi:10.48550/arXiv.2308.02457.

102 Jie Wang, Ghulam Mubashar Hassan, and Naveed Akhtar. A survey of neural trojan attacks and defenses in deep learning. *CoRR*, abs/2202.07183, 2022. arXiv:2202.07183.

103 Meihong Wang, Linling Qiu, and Xiaoli Wang. A survey on knowledge graph embeddings for

link prediction. *Symmetry*, 13(3):485, 2021. doi:10.3390/SYM13030485.

104 Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. Cross-lingual knowledge graph alignment via graph convolutional networks. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 349–357, 2018. doi:10.18653/V1/D18-1032.

105 Zikang Wang, Linjing Li, Qiudan Li, and Daniel Zeng. Multimodal data enhanced representation learning for knowledge graphs. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019. doi:10.1109/IJCNN.2019.8852079.

106 L. N. Wasserstein. Markov processes over denumerable products of spaces describing large systems of automata. *Problems of Information Transmission*, 5:47–52, 1969.

107 Junkang Wu, Wentao Shi, Xuezhi Cao, Jiawei Chen, Wenqiang Lei, Fuzheng Zhang, Wei Wu, and Xiangnan He. Disenkgat: Knowledge graph embedding with disentangled graph attention network. In Gianluca Demartini, Guido Zuccon, J. Shane Culpepper, Zi Huang, and Hanghang Tong, editors, *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021*, pages 2140–2149. ACM, 2021. doi:10.1145/3459637.3482424.

108 Zhaohan Xi, Tianyu Du, Changjiang Li, Ren Pang, Shouling Ji, Xiapu Luo, Xusheng Xiao, Fenglong Ma, and Ting Wang. On the security risks of knowledge graph reasoning. In Joseph A. Calandrino and Carmela Troncoso, editors, *32nd USENIX Security Symposium, USENIX Security 2023, Anaheim, CA, USA, August 9-11, 2023*, pages 3259–3276. USENIX Association, 2023. URL: https://www.usenix.org/conference/usenixsecurity23/presentation/xi.

109 Hanhua Xiao, Yuchen Li, Yanhao Wang, Panagiotis Karras, Kyriakos Mouratidis, and Natalia Rozalia Avlona. How to avoid jumping to conclusions: Measuring the robustness of outstanding facts in knowledge graphs. In Ricardo Baeza-Yates and Francesco Bonchi, editors, *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024*, pages 3539–3550. ACM, 2024. doi:10.1145/3637528.3671763.

110 Ruobing Xie, Stefan Heinrich, Zhiyuan Liu, Cornelius Weber, Yuan Yao, Stefan Wermter, and Maosong Sun. Integrating image-based and knowledge-based representation learning. *IEEE Transactions on Cognitive and Developmental Systems*, 12(2):169–178, 2019. doi:10.1109/TCDS.2019.2906685.

111 Ruobing Xie, Zhiyuan Liu, Fen Lin, and Leyu Lin. Does william shakespeare REALLY write hamlet? knowledge representation learning with confidence. In Sheila A. McIlraith and Kilian Q. Weinberger, editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 4954–4961. AAAI Press, 2018. doi:10.1609/AAAI.V32I1.11924.

112 Huan Xu and Shie Mannor. Robustness and generalization. In Adam Tauman Kalai and Mehryar Mohri, editors, *COLT 2010 - The 23rd Conference on Learning Theory, Haifa, Israel, June 27-29, 2010*, pages 503–515. Omnipress, 2010. URL: http://colt2010.haifa.il.ibm.com/papers/COLT2010proceedings.pdf#page=511.

113 Jiarong Xu, Junru Chen, Siqi You, Zhiqing Xiao, Yang Yang, and Jiangang Lu. Robustness of deep learning models on graphs: A survey. *AI Open*, 2:69–78, 2021. doi:10.1016/j.aiopen.2021.05.002.

114 Jiarong Xu, Junru Chen, Siqi You, Zhiqing Xiao, Yang Yang, and Jiangang Lu. Robustness of deep learning models on graphs: A survey. *AI Open*, 2:69–78, 2021. doi:10.1016/J.AIOPEN.2021.05.002.

115 Ziyu Xu, Chen Dan, Justin Khim, and Pradeep Ravikumar. Class-weighted classification: Trade-offs and robust approaches. *CoRR*, abs/2005.12914, 2020. arXiv:2005.12914.

116 Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*, 2014.

117 Yi Yang, Chen Peng, En-Zhi Cao, and Wenxuan Zou. Building resilience in supply chains: A knowledge graph-based risk management framework. *IEEE Transactions on Computational Social Systems*, pages 1–9, 2023. doi:10.1109/TCSS.2023.3334768.

118 Yuhao Yang, Chao Huang, Lianghao Xia, and Chenliang Li. Knowledge graph contrastive learning for recommendation. In Enrique Amigó, Pablo Castells, Julio Gonzalo, Ben Carterette, J. Shane Culpepper, and Gabriella Kazai, editors, *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022*, pages 1434–1443. ACM, 2022. doi:10.1145/3477495.3532009.

119 Xiaoyu You, Beina Sheng, Daizong Ding, Mi Zhang, Xudong Pan, Min Yang, and Fuli Feng. Mass: Model-agnostic, semantic and stealthy data poisoning attack on knowledge graph embedding. In Ying Ding, Jie Tang, Juan F. Sequeda, Lora Aroyo, Carlos Castillo, and Geert-Jan Houben, editors, *Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023*, pages 2000–2010. ACM, 2023. doi:10.1145/3543507.3583203.

120 Mei Yu, Jiujiang Guo, Jian Yu, Tianyi Xu, Mankun Zhao, Hongwei Liu, Xuewei Li, and Ruiguo Yu. TBDRI: block decomposition based on relational interaction for temporal knowledge graph completion. *Appl. Intell.*, 53(5):5072–5084, 2023. doi:10.1007/S10489-022-03601-5.

121 Mohamad Zamini, Hassan Reza, and Minou Rabiei. A review of knowledge graph completion. *Information*, 13(8):396, 2022. doi:10.3390/INFO13080396.

122 Hengtong Zhang, Tianhang Zheng, Jing Gao, Chenglin Miao, Lu Su, Yaliang Li, and Kui Ren.

Data poisoning attack against knowledge graph embedding. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 4853–4859. ijcai.org, 2019. doi:10.24963/ijcai.2019/674.

123 Kexin Zhang, Shuhan Liu, Song Wang, Weili Shi, Chen Chen, Pan Li, Sheng Li, Jundong Li, and Kaize Ding. A survey of deep graph learning under distribution shifts: from graph out-of-distribution generalization to adaptation. *CoRR*, abs/2410.19265, 2024. doi:10.48550/arXiv.2410.19265.

124 Shuai Zhang, Yi Tay, Lina Yao, and Qi Liu. Quaternion knowledge graph embeddings. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 2731–2741, 2019. URL: https://proceedings.neurips.cc/paper/2019/hash/d961e9f236177d65d21100592edb0769-Abstract.html.

125 Shuai Zhang, Yi Tay, Lina Yao, and Qi Liu. Quaternion knowledge graph embeddings. *Advances in neural information processing systems*, 32, 2019.

126 Yuxiao Zhang, Qingfeng Chen, Xinkun Hao, Haiming Pan, Qian Yu, and Kexin Huang. Defense against adversarial attack on knowledge graph embedding. *Emerging Trends in Cybersecurity Applications*, page 441, 2023.

127 Zeyang Zhang, Xin Wang, Ziwei Zhang, Haoyang Li, Zhou Qin, and Wenwu Zhu. Dynamic graph neural networks under spatio-temporal distribution shift. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL: http://papers.nips.cc/paper_files/paper/2022/hash/2857242c9e97de339ce642e75b15ff24-Abstract-Conference.html.

128 Zhao Zhang, Fuzhen Zhuang, Hengshu Zhu, Chao Li, Hui Xiong, Qing He, and Yongjun Xu. Towards robust knowledge graph embedding via multi-task reinforcement learning. *IEEE Trans. Knowl. Data Eng.*, 35(4):4321–4334, 2023. doi:10.1109/TKDE.2021.3127951.

129 Tianzhe Zhao, Jiaoyan Chen, Yanchi Ru, Qika Lin, Yuxia Geng, and Jun Liu. Untargeted adversarial attack on knowledge graph embeddings. In Grace Hui Yang, Hongning Wang, Sam Han, Claudia Hauff, Guido Zuccon, and Yi Zhang, editors, *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2024, Washington DC, USA, July 14-18, 2024*, pages 1701–1711. ACM, 2024. doi:10.1145/3626772.3657702.

130 Yu Zhao, Huali Feng, and Patrick Gallinari. Embedding learning with triple trustiness on noisy knowledge graph. *Entropy*, 21(11):1083, 2019. doi:10.3390/E21111083.

131 Lei Zheng, Pei Quan, Yong Shi, and Lingfeng Niu. A brief survey of distribution robust graph neural networks. *Procedia Computer Science*, 242:1281–1286, 2024. 11th International Conference on Information Technology and Quantitative Management (ITQM 2024). doi:10.1016/j.procs.2024.08.140.

132 Yongchun Zhu, Fuzhen Zhuang, Xiangliang Zhang, Zhiyuan Qi, Zhi-Ping Shi, Juan Cao, and Qing He. Combat data shift in few-shot learning with knowledge graph. *Frontiers Comput. Sci.*, 17(1):171305, 2023. doi:10.1007/S11704-022-1339-7.

133 Yuqicheng Zhu, Nico Potyka, Mojtaba Nayyeri, Bo Xiong, Yunjie He, Evgeny Kharlamov, and Steffen Staab. Predictive multiplicity of knowledge graph embeddings in link prediction. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, Florida, USA, November 12-16, 2024*, pages 334–354. Association for Computational Linguistics, 2024. doi:10.18653/V1/2024.FINDINGS-EMNLP.19.

134 Yuqicheng Zhu, Nico Potyka, Jiarong Pan, Bo Xiong, Yunjie He, Evgeny Kharlamov, and Steffen Staab. Conformalized answer set prediction for knowledge graph embedding. *CoRR*, abs/2408.08248, 2024. doi:10.48550/arXiv.2408.08248.