Leveraging Disease-Specific Topologies and Counterfactual Relationships in Knowledge Graphs for Inductive Reasoning in Drug Repurposing

Çerağ Oğuztüzün
Department of Computer Science
Center for AI in Drug Discovery
Case Western Reserve Univesity
Cleveland, OH 44106, USA
cxo147@case.edu

Zhenxiang Gao
Center for AI in Drug Discovery
Case Western Reserve Univesity
Cleveland, OH 44106, USA
zxg306@case.edu

Hui Li
Center for AI in Drug Discovery
Case Western Reserve Univesity
Cleveland, OH 44106, USA
hxl1529@case.edu

Rong Xu *
Center for AI in Drug Discovery
Case Western Reserve Univesity
Cleveland, OH 44106, USA
rxx@case.edu

Abstract

Drug repurposing offers a cost-effective strategy to accelerate drug development by identifying new therapeutic uses for approved medications. However, it poses significant challenges for complex diseases with poorly understood mechanisms of action. Addressing these diseases requires the efficient integration of new data while minimizing retraining time, prompting us to develop **domain-specific graph augmentation techniques that support semi-inductive reasoning.** We discovered that leveraging counterfactual relationships derived from disease-specific topological structures significantly enhances model performance. Based on this insight, we integrated counterfactual relationships as an augmentation method and an initialization step in our knowledge graph (KG) link prediction training process. We introduce **KGïA**, an **i**nductive **KG a**ugmentation method that utilizes counterfactual relationships based on disease-specific topologies. By aligning augmentation with the intrinsic topological features of disease entities, KGïA enhances the KG in a domain-specific manner, facilitating the discovery of a broader range of novel drug candidates tailored to specific diseases. Our biomedical KG comprises 1,614,801 triples and 100,563 biomedical entities, including 30,006 diseases, constructed from **6 biomedical datasets** and enriched through Natural Language Processing (NLP) relation extraction. Extensive experiments on this comprehensive KG using **5 augmented architectures** demonstrate that semi-inductive reasoning significantly improves generalizability (up to a **24× increase in Mean Reciprocal Rank (MRR)**) and that augmented models outperform state-of-the-art KG-based drug repurposing methods (up to a **32% improvement in MRR**). Additionally, in an Alzheimer's Disease (AD) case study, our model identified up to **5 mechanism categories** compared to **2 in the baseline**, highlighting its enhanced capability to uncover diverse drug candidates.

*Corresponding author.

AI for New Drug Modalities at NeurIPS 2024.

1 Introduction

Drug repurposing is the process of finding new applications for existing, approved drugs; It offers a time and cost-efficient approach to drug development Ehrlinger & Wöß (2016); Zhu et al. (2020). Knowledge graph (KG)-based approaches for drug repurposing leverage machine learning to identify drug candidates that are safer and more effective. By integrating multiple databases, KGs can identify semantic connections, which enhances the selection process for appropriate treatments by incorporating multiple databases and literature Li et al. (2022); Bonner et al. (2022); Nicholson & Greene (2020).

There are several KG-based drug repurposing models available, including KG-Predict Gao et al. (2022), which infers new drug-disease interactions based on heterogeneous feature interactions between entity and relation embeddings in a biomedical KG. However, as with other KG-based drug repurposing methods, KG-Predict's lack of inductive inference capabilities means it cannot reason about unseen entities or relations, resulting in prolonged training times whenever new data points are added to the KG. Current methods are categorized as either inductive or transductive, depending on whether they identify underlying patterns to predict unseen samples (inductive) or build prediction models for observed samples only (transductive) de la Fuente et al. (2023). Inductive reasoning offers several advantages over transductive reasoning: It avoids data leakage issues common in transductive feature generation de la Fuente et al. (2023), excels with large datasets and clear hypotheses by discovering solutions and novel findings Shin (2019), and generalizes across different vocabularies in Knowledge Graphs (KGs) Galkin et al. (2023). Existing work in biomedical link prediction and drug discovery has predominantly used transductive reasoning, whereas inductive reasoning is less common de la Fuente et al. (2023) due to the biomedical domain's complex, heterogeneous, and often incomplete data, which complicates its application McCoy et al. (2021). Additionally, many current link prediction models perform well in transductive settings but struggle in more realistic inductive settings where new nodes are introduced, making it difficult to properly evaluate and benchmark inductive reasoning approaches Chatterjee et al. (2022). Our work sets itself apart from current inductive reasoning models for drug-target interactions, such as NeoDTI Wan et al. (2019), MolTrans Huang et al. (2021), HyperAttentionDTI Zhao et al. (2022a), and EEG-DTI Kirino et al. (2019), by applying inductive reasoning to both nodes and relations, predicting drug-disease interactions, and employing disease-specific graph augmentation.

Additionally, there are diseases whose mechanism of action is not well understood, or whose factors are too numerous Chung et al. (2023). In complex diseases, such as Alzheimer's Disease (AD), cardiovascular diseases, psychiatric conditions, and stroke, a lack of understanding of therapeutic pathways posed a challenge to the design of trials Krishnamurthy et al. (2022). Drug repurposing for those diseases is harder than repurposing for diseases that are well known and well researched Kulkarni et al. (2023). For such diseases, augmenting the graph in an informative way can improve the quality of drug candidates predicted by our models Ozkan et al. (2023). There are efforts for this augmentation task such as Graph Substructure Networks (GSNs) Bouritsas et al. (2022) which pass additional structural descriptors to message-passing functions generated by counting subgraph isomorphisms. The authors have observed that different datasets and problems require different topological substructures, so the choice will likely depend on the problem, as we will illustrate in the drug repurposing case. For example in social networks, lower-order cliques are common as they depict groups of small tight-knit circles and show how people naturally form close relationships Hanneman & Riddle (2005). Similarly, for drug repurposing, we know that the principle of looking for patterns or motifs can reveal important information about the disease's behavior.

We are motivated by the challenge of addressing diseases with poorly understood mechanisms of action, as well as the need to efficiently incorporate new information about these diseases while minimizing re-training time. Domain-specific graph augmentation can help unravel complex disease mechanisms, while inductive reasoning allows us to integrate new data with reduced re-training effort. We introduce **KGïA**, a **KG**-based drug repurposing model with both **i**nductive and **a**ugmentative capabilities. Regarding the first case, the key contribution of KGïA's inductive approach is to identify new or unforeseen drug repurposing opportunities without requiring additional training. Secondly, we enhance the quality of predicted drug candidates by augmenting the biomedical KG—specifically tailored to the topological properties of the disease we are repurposing drugs for. We introduce equivariance preserving and counterfactual edges to the biomedical KG we have constructed using various high-quality public biomedical datasets and literature.

2 Methods

2.1 Problem Formulation

2.1.1 Knowledge Graph Definition

We define KGs as labeled, directed graphs with multiple, distinct edges between nodes Rossi et al. (2021). Formally, a KG is represented as KG $= (V, L, F)$, where V is a set of vertices, each symbolizing an entity. L is a collection of labels, each indicating a type of relation. $F \subseteq V \times L \times V$ is a set of directed edges that depict relationships as triples $\langle h, r, t \rangle$, with h as the head entity, r as the relation label, and t as the tail entity.

2.1.2 Link Prediction in Graphs

The link prediction task in the context of KG reasoning answers queries $\langle h, r, ? \rangle$. The problem involves estimating the likelihood of a specific link $\langle h, r, t \rangle$ being true but not observed in the current graph structure. Given a query $\langle h, r, ? \rangle$, our goal is to rank all nodes $v \in V$ in the inference graph, denoted as KG $= (V, L, F)$, such that the probability $P(t = v \mid h, r)$ of correctly identifying a true tail entity t is maximized. For drug repurposing, the link prediction problem can be specialized to predict new $Drug \xrightarrow{\text{treats}} Disease$ triples Doshi & Chepuri (2022). This involves estimating the likelihood of a given drug being effective in treating a disease that it is not currently known or approved to treat, based on the knowledge encoded in the KG Islam et al. (2023).

In the *transductive link prediction* setting, the model is trained and inferred on the same graph $G_{\text{train}} = G_{\text{inf}}$. The model is expected to predict new links between entities that were already part of the graph during the training phase but does not generalize to unseen entities or relations. Where $G_{\text{train}} = (V_{\text{train}}, L_{\text{train}}, F_{\text{train}})$ and $G_{\text{inf}} = (V_{\text{inf}}, L_{\text{inf}}, F_{\text{inf}})$: $V_{\text{train}} = V_{\text{inf}}$, $L_{\text{train}} = L_{\text{inf}}$, $F_{\text{train}} \subset F_{\text{inf}}$.

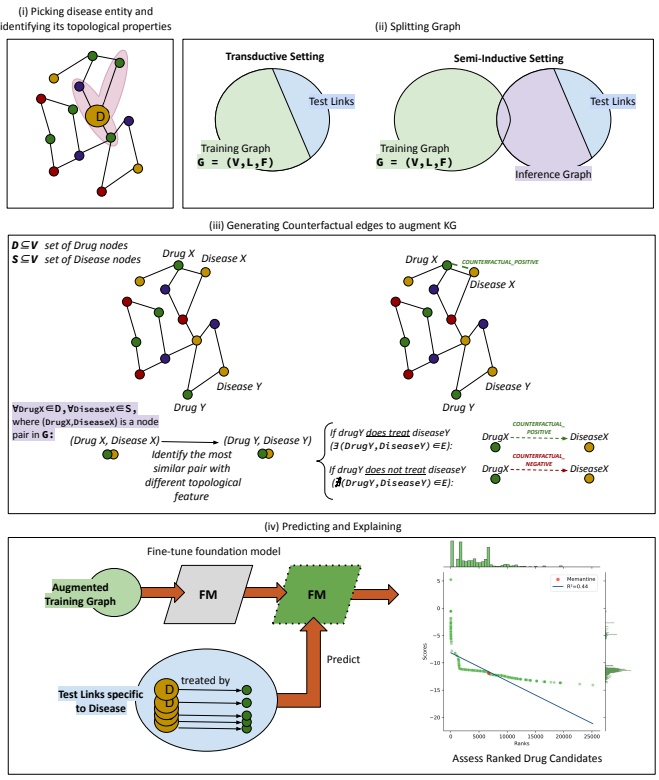
In the *semi-inductive setting*, the model is trained on a graph where it may not have seen all the entities (V) and relations (L) during training, but it is expected to generalize to unseen entities ($V_{\text{inf}} \backslash V_{\text{train}}$) and relations ($L_{\text{inf}} \backslash L_{\text{train}}$) in the same graph during inference Vapnik (1999). This means there's a partial overlap between the training and inference graphs, particularly with some new entities or relationships being introduced at inference time that were not present during training such that: $V_{\text{train}} \subseteq V_{\text{inf}}$, $L_{\text{train}} \subseteq L_{\text{inf}}$, $F_{\text{train}} \subseteq F_{\text{inf}}$.

2.2 System Overview

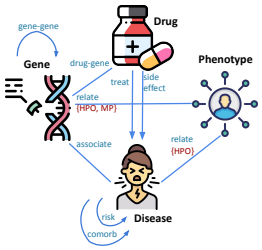
Figure 1 presents an overview of this approach. In the first step, a disease entity of interest for drug repurposing is selected from the biomedical KG. The topological properties of this disease node are investigated in Section 3.3.1. An understanding of the appropriate manipulation for augmenting the graph with counterfactual edges is developed. The second step involves deciding on the dataset split, with options for either a transductive or semi-inductive setting. More details on splits are presented in Section 2.5.2. In the third step, the training set is augmented based on the manipulation chosen in the first step. For each drug-disease pair in the graph, the goal is to determine how the topological structure influences the link (non-)existence for each node pair. This involves identifying the most similar pair of nodes with a differing topological feature and enriching the graph with `COUNTERFACTUAL_POSITIVE` and `COUNTERFACTUAL_NEGATIVE` links. This enrichment aims to train models to recognize not only existing but also potentially non-effective treatments. Further details on the augmentation method are provided in Section 2.4. In the fourth and final step, the foundation model is fine-tuned using the augmented graph. This fine-tuned model is then used to score and rank each drug in the graph for potential treatment of the selected disease node in the first step. Lastly, the top-ranked drug candidates are assessed and interpreted using domain knowledge. The code is available at: https://github.com/ceragoguztuzun/KGIA.

2.3 Foundation Model

Foundation models are versatile AI systems engineered to generate a broad spectrum of outputs. These models can function independently or serve as the underlying framework for various other applications Jones et al. (2023). In their inspiring work Galkin et al. (2023) employ double equivariance by modeling it as the *transferable invariance of relation interactions*. In their view, representing relations



(a) Overview of KGïA with main steps



(b) Data schema of the biomedical knowledge graph

Figure 1: **(a)** (i) Picking disease entity and identifying its topological properties, (ii) splitting graph Galkin et al. (2022), (iii) generating counterfactual edges to augment KG, (iv) predicting drug candidates, explaining the intrinsic ranking mechanism of the model, analyzing drug mechanisms and evidence levels of drug candidates.

as a graph of relations is a double equivariant approach (see A.3 for further details). Consequently, learning relational representations is independent of identitites, but dependent on the interaction of relations and entities. We utilize ULTRA as our state-of-the-art foundation model to achieve inductive reasoning, adapting its application in the downstream task of identifying new drug repurposing opportunities. In A.1, we provide the relation embedding process in Algorithm 1, and the entity embedding and link prediction process in Algorithm 2.

2.4 Graph Augmentation

We propose the following graph augmentation technique with two key strengths: **(i)** It incorporates both counterfactual relationships and domain-specific graph structures. **(ii)** It maintains the model's invariant properties regarding double equivariance (Proof provided in A.3, Theorem 1).

2.4.1 Counterfactual Relationships

Zhao et al. (2022b) defined counterfactual links to augment graph structure of training data in their innovative rule-based approach. These links represent hypothetical outcomes between a pair of nodes that are not directly observable in the real-world data. The counterfactual links originate from the pivotal question in causal inference: *"Would the connection between the nodes persist if the observed graph structure were altered?"*. The process relies on causal models structured around three elements: the context, the manipulation, and the outcome Van der Laan & Petersen (2007); Johansson et al. (2016).

Inspired by this work, for each drug-disease node pair in our KG, we ask the question: *"Would Drug X treat Disease X if their selected topological graph structure became different from observation?"*. In our scenario of link prediction, we consider *the data of a drug-disease node pair* as the context, *the characteristics of the graph's selected topological structure* as the manipulation, and *the presence or absence of a* `TREAT` *link* as the outcome.

For each drug-disease node pair $(dr_x, di_x) \in V \times V$ in our KG $G = (V, L, F)$, we find the most similar drug-disease node pair (dr_y, di_y) in G measured by $h(\cdot)$ which measures the Euclidian distance between a pair of nodes, with the opposite manipulation of the original node pair. The matrix M denotes manipulations on node pairs, with $M_{i,j} = 1$ if the drug dr_x and disease di_x share the same manipulation state and $M_{i,j} = 0$ otherwise. We use their outcome value to define counterfactual edges. These types of covariate matching are widely used to estimate manipulation effects Johansson et al. (2016); Alaa & Van Der Schaar (2019).

We define two cases for the counterfactual link depending on the counterfactual answer, which are named to be distinguished from observed edges. This augmented graph represents a broader spectrum of possible interactions. `COUNTERFACTUAL_POSITIVE` links strengthen the hypothesis that dr_x could potentially treat di_x, even if they didn't have the same topologic structure and `COUNTERFACTUAL_NEGATIVE` links are used to train our model to recognize not only existing but also potential non-effective treatments. The counterfactual edges are defined as the following:

$$\forall (dr_x, di_x) \in V \times V,$$
$$\begin{cases} dr_x \xrightarrow{\text{COUNTERFACTUAL\_POSITIVE}} di_x & \text{if } dr_y \xrightarrow{\text{TREAT}} di_y, \\ dr_x \xrightarrow{\text{COUNTERFACTUAL\_NEGATIVE}} di_x & \text{if } dr_y \not\xrightarrow{\text{TREAT}} di_y. \end{cases}$$

where
$$(dr_y, di_y) = \underset{dr_y, di_y \in V}{\operatorname{argmin}} \; h((dr_x, di_x), (dr_y, di_y)) | M_{dr_y, di_y} = 1 - M_{dr_x, di_x}$$

We experiment with a range of topological properties as manipulations, creating a different augmented training data for each manipulation. Our goal is to select manipulations that highlight properties potentially similar to those of a specific disease node of interest. We experimented with the manipulations: *Katz Centrality, Common Neighbours, Louvain, K-Core, Propagation* Newman (2018). Their definitions are reproduced in the appendix (A.1).

2.5 Experiment Setting

To evaluate our model's performance as a framework for KG reasoning in our experiments, we ask:

1. Does our model outperform a state-of-the-art drug repurposing model trained from scratch?
2. Does inductive generalization outperform transductive reasoning?
3. Does augmenting with a specific manipulation have any benefits?
4. *(Ablation Study)* Does introducing either counterfactual positive or negative edges provide any benefit?

2.5.1 Dataset

We constructed a KG by extracting interactions among genes, drugs, diseases, phenotypes, side effects, and symptoms from various high-quality public biomedical datasets (Figure 1b). We sourced gene functions from the Mouse Genome Informatics (MGI) Eppig et al. (2017) and Human Phenotype Ontology (HPO) databases Robinson & Mundlos (2010), and gene interactions from protein-protein interaction (PPI) networks Safari-Alighiarloo et al. (2014). Drug-side effect associations were obtained from the SIDER database Kuhn et al. (2016). Drug-gene associations were obtained from the DrugBank database Wishart et al. (2008). Disease-phenotype associations and disease-gene interactions were obtained from the HPO and DisGeNET databases Piñero et al. (2015), respectively. Further, drug-disease, disease-risk, and disease-comorbidity knowledge were mined using natural language processing (NLP) techniques from our previous work Xu & Wang (2013); Xu et al. (2013, 2014). The resulting KG comprises 100,563 entities across 4 types and 1,614,801 triplets across 10 edge types. Our KG is freely available at our GitHub repository.

2.5.2 Setup

Data splitting is conducted in two frameworks: transductive and semi-inductive, both utilizing 70% of the total data for training. For both settings, we used the same training set. The data split is performed by randomly selecting triples from the graph. To address the high computational cost of data augmentation, we implemented hold-out validation. This approach avoids augmenting each fold in k-fold cross-validation individually while still ensuring robust evaluation through multiple runs with varied initializations (see Figure 6 for details). In clinical research, model validation often involves testing on external datasets, which aligns more closely with the hold-out method, as suggested and employed by Eertink et al. (2022); Liu et al. (2024); Crowley.

The semi-inductive setting features a disjoint inference graph with potentially new entities and relation types, combining validation and test data. This graph includes 212,266 unique triples not present in training data. The validation set encompasses 10,535 unseen entities, while the test set contains 17,926 unseen entities, with 7,591 entities unique to the test set. While our model can infer new relation types, no unseen relation types are present in the validation or test sets for this study. For the case study, we ensured that neither the training nor validation sets contained triples containing Alzheimer's Disease. Figure 5 details the number of augmentation relations for each manipulation strategy. Due to an imbalance between the counts of `COUNTERFACTUAL_POSITIVE` and `COUNTERFACTUAL_NEGATIVE` relations, we employed random downsampling on the majority group's triples to match with the numbers of the smaller group Noor et al. (2022).

2.5.3 Baseline

We use KG-Predict Gao et al. (2022) as our state-of-the-art KG-based drug repurposing computational framework baseline. Through the use of stacked CompGCN Vashishth et al. (2019) layers and InteractE Vashishth et al. (2020), entities and relations are embedded. As shown in their paper, they surpass DeepWalk Perozzi et al. (2014), matrix factorization approaches, TransE Bordes et al. (2013), DistMult Yang et al. (2014), ConvE Dettmers et al. (2018), RotatE Sun et al. (2019). In addition to comparing generalizability, we compare the drugs repurposed by KG-Predict and our model in our case study.

2.5.4 Evaluation Protocol

We employ Mean Rank (MR), Mean Reciprocal Rank (MRR) and Hits@k as our evaluation metrics. As part of zero-shot inference, we use the inference graph and the test set of triples. For fine-tuning, we continue training the model on the training split, maintaining the checkpoint for the best MRR in the validation set. Whenever we run zero-shot experiments, we run them once since the results are deterministic, and we do not use augmentation as it only makes a difference in fine-tuned scenarios. We report an average of five runs for each fine-tuning run, and convert the average of MR into an integer by rounding up.

A case study of repurposing for Alzheimer's Disease uses test split queries to perform link prediction to find drug candidates that have a high probability of having a `TREAT` relation with the disease node corresponding to Alzheimer's Disease (7845 such queries). By using domain knowledge, we examine drug candidates for each group to determine if the augmentation methods benefit disease-specific use cases (Figure 4a, Table 4, Table5).

3 Results

3.1 Model Performance

Table 1 showcases the models' performance across transductive and semi-inductive settings, focusing on their proficiency in prioritizing 'true' triples without pinpointing specific diseases or drug repurposing tasks. In the transductive scenario, the vanilla ULTRA foundation model outperforms KG-Predict, while augmented models demonstrate superior MRR and Hits@K metrics. Similarly, in the semi-inductive setting, augmented models exhibit comparable MR values but greatly improved MRR and Hits@K, highlighting our augmentation technique's effectiveness in improving link prediction tasks' generalizability and performance.

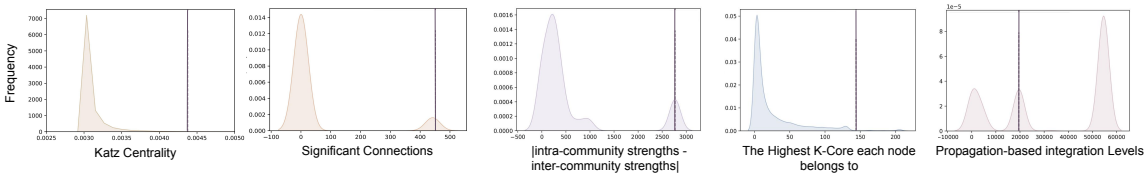


Figure 2: Distributions of topological structures for augmentation manipulations. The node corresponding to Alzheimer's Disease is annotated.

Table 1: Evaluation metrics of models under Transductive and Semi-Inductive settings. We use KG-Predict as our primary benchmark, and non-augmented ULTRA models (zero-shot and fine-tuned) as secondary benchmarks.

	Transductive					Semi-Inductive				
Model	MR	MRR	Hits@1	Hits@3	Hits@10	MR	MRR	Hits@1	Hits@3	Hits@10
KG-Predict	18049	0.0118	**0.0080**	0.0111	0.0152	-	-	-	-	-
ULTRA 0-shot	49748	0.0009	0.0002	0.0008	0.0018	55	0.2958	0.1964	0.3449	0.4656
ULTRA fine-tuned	**3935**	0.0178	0.0003	0.0185	0.0415	**11**	0.6351	0.4824	0.7404	0.9290
KGïA (Katz Augmented)	14448	0.0302	0.0002	**0.0328**	0.0849	13	0.7456	0.6251	0.8512	**0.9376**
KGïA (Common Neighbors Aug.)	27002	0.0035	0.0002	0.0042	0.0084	17	0.5856	0.4245	0.6964	0.8923
KGïA (Louvain Augmented)	20295	0.0209	0.0001	0.0169	0.0487	23	0.8060	0.7437	0.8527	0.9108
KGïA (K-Core Augmented)	10808	0.0227	0.0001	0.0197	0.0669	15	**0.8380**	**0.7736**	**0.8939**	0.9364
KGïA (Propagation Augmented)	4801	**0.0338**	0.0002	0.0271	**0.0857**	20	0.8167	0.7444	0.8781	0.9280

3.2 Impact of Positive and Negative Counterfactual Edges

In our ablation study, which aimed to analyze the impact of counterfactual positive and negative edges, we present the results in Table 2. In the transductive setting, we observed that adding only counterfactual positive or negative edges led to better performance than adding both together. However, in the semi-inductive setting, the combination of both edge types yielded better results. This suggests that in the transductive setting, the model may be more sensitive to the introduction of additional information, and using either positive or negative counterfactual edges alone enhances performance without adding unnecessary complexity. In contrast, the semi-inductive setting, which lacks overlapping entities and relations, benefits more from the combined use of both edge types.

3.3 Case Study

We applied our method to repurpose drugs for Alzheimer's Disease (AD), selecting AD for its significant genetic diversity Chung et al. (2023), which serves as a test for our augmentation technique's ability to encapsulate the various potential causes of AD. Our study aimed to rank potential drug candidates $Drug \xrightarrow{\text{TREAT}} AD$, focusing on analyzing the top 5 ranked drug candidates.

3.3.1 Preliminary analysis of topological properties

To begin the case study with step (i) in Figure 1, we analyze AD's topological properties. We plot the distributions of each manipulation choice, and we annotate the values of the entity corresponding to AD in Figure 2 so that we can determine where AD falls within the distribution of the biomedical KG. AD exhibits a Katz centrality score of 0.004, relatively high among the graph's nodes. By computing intra-community strength through the diagonal of the membership matrix M and inner-community strength by summing each row minus the diagonal, it's evident from Figure 2 that AD's intra-community strength greatly surpasses its inter-community strength. This suggests AD may act as a bridge between communities. Further analysis reveals AD is part of the highest k-core of 144, and has a propagation-based integration level of 19520, typical for nodes within the KG.

3.3.2 Evaluation

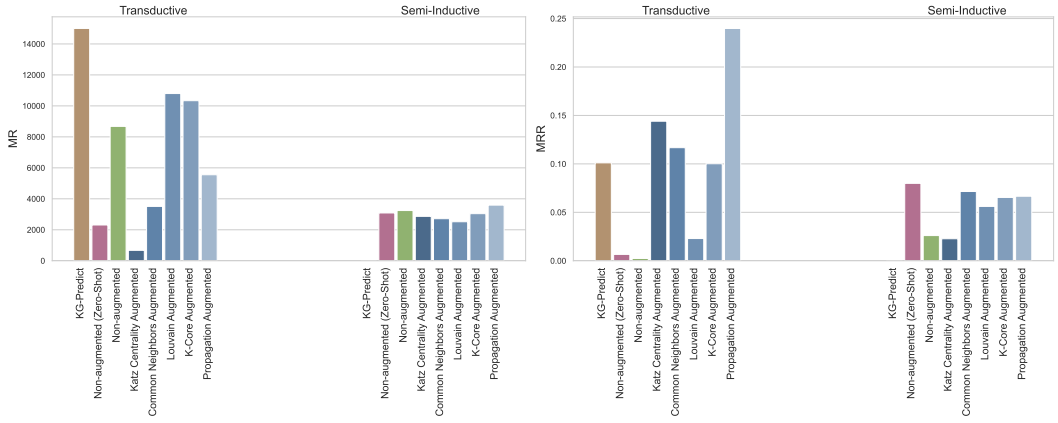
In Figure 3, we compare the MR and MRR metrics of various models, with KG-Predict serving as our primary benchmark and non-augmented models (both zero-shot and fine-tuned) as secondary baselines. In the transductive setting, Katz centrality, Common Neighbors, and Propagation augmented

Table 2: Ablation study on the effects of positive and negative counterfactual edge augmentations. Positive augmentation only introduces `COUNTERFACTUAL_POSITIVE` edges, and negative augmentation only introduces `COUNTERFACTUAL_NEGATIVE` edges. The reported values are the averages of all augmented models.

Augmentation	MR	MRR	Hits@10
Transductive			
Positive Augmentation	**8286**	**0.0298**	**0.0805**
Negative Augmentation	10724	0.0269	0.0744
Both	15471	0.0222	0.0589
Semi-Inductive			
Positive Augmentation	20	0.6666	0.8883
Negative Augmentation	**14**	0.7397	0.9051
Both	18	**0.7584**	**0.9210**

models outperform others in MR metrics (Figure 3a). While KG-Predict outperforms non-augmented ULTRA in terms of MRR, augmented models, such as Katz, Common Neighbors, and Propagation, outperform (Figure 3b). As for the semi-inductive setting, all augmented models, excluding Propagation, demonstrate superior performance regarding MR (Figure 3a). As for MRR, Common Neighbors, Louvain, K-Core, and Propagation also exhibit strong performance (Figure 3b). Synthesizing these insights with Hit@K values (Table 3), the **Propagation augmented model** emerges as the top model for the transductive setting, and the **Common Neighbors augmented model** for semi-inductive setting.

The semi-inductive results for the AD case study differ significantly from the general case due to the strict formatting of triples, such as $Drug \xrightarrow{\text{TREAT}} AD$ (7845 such queries). This limited data variety increases sparsity, unlike the general setting, where flexible formats (as described in Section 2.5.2) allow richer data interactions and better model performance. The case study's narrower focus inherently presents challenges.



(a) MR Comparison across Models

(b) MRR Comparison across Models

Figure 3: MR and MRR values for the case study in Transductive and Semi-Inductive settings.

3.3.3 Analysis of Drug Candidates

In Figure 4b,c, we assess the models' internal consistency by assessing the ranks and scores given to each drug candidate. The curvature differences reflect how the model's certainty in its predictions changes across the ranking spectrum in each setting, and this is influenced by how the models were trained and how they generalize from seen to unseen data. In the transductive setting (Figure 4b) where the model has access to all entities and relations during training, we observe a concave curve. In the semi-inductive setting (Figure 4c), the model is based on a partial view of the network, leading to a steeper decline in confidence as the ranks increase. When we look at both the Propagation augmented

Table 3: Hits@k values for the case study in Transductive and Semi-Inductive settings.

	Transductive			Semi-Inductive		
Model	Hits@1	Hits@3	Hits@10	Hits@1	Hits@3	Hits@10
KG-Predict	0	0	0	-	-	-
ULTRA 0-shot	0.00459	0.00459	0.00739	**0.07559**	0.07610	0.08107
ULTRA fine-tuned	0.00140	0.00166	0.00166	0.00025	0.03403	0.10121
KGïA (Katz Augmented)	0.03697	0.19401	**0.31217**	0.01651	0.01740	0.04442
KGïA (Common Neighbors Aug.)	0.04283	0.18917	0.20637	0.04066	**0.11447**	**0.14800**
KGïA (Louvain Augmented)	0.00688	0.03238	0.04576	0.05010	0.05048	0.05354
KGïA (K-Core Augmented)	0.09433	0.10057	0.11421	0.05851	0.06068	0.06386
KGïA (Propagation Augmented)	**0.23429**	**0.23811**	0.24551	0.06661	0.05163	0.07075

model and the Common Neighbors augmented model, the density plot shows a peak skewed to the left, suggesting a tendency to assign higher ranks to triples. Yet, the scatters corresponding to ranked triples convey that many predictions are tied to negative logit scores, highlighting a lack of confidence in those outcomes. On the other hand, the Non-Augmented models maintain a moderate level of confidence throughout their rankings.

We listed the drug candidates for each model in Figure 4a, as well as each drug's evidence level and mechanism category. Figure 4a is a high-level summary of the details presented in Table 4 (transductive setting) and Table 5 (semi-inductive setting). An in-depth discussion of drug candidates is presented in Section A.4.

4 Discussion and Conclusion

Addressing diseases with complex, poorly understood mechanisms requires efficiently integrating new data while minimizing re-training time. Domain-specific graph augmentation helps unravel these mechanisms, and inductive reasoning allows for easy updates with minimal re-training. To address these challenges, our model KGïA introduces an innovative method for enhancing knowledge graphs (KGs) that supports semi-inductive reasoning. Table 1 shows that inductive reasoning significantly improves MRR metrics compared to transductive models, demonstrating superior generalization. This improvement is largely due to the KG's heterogeneous connectivity, where semi-inductive models excel in capturing and generalizing across diverse structural contexts.

We leveraged counterfactual relationships derived from disease-specific topological structures. Table 1 reveals that this approach has tripled the MRR and increased Hits@K. The results demonstrate that domain-specific augmentation leverages domain knowledge effectively within the model, yielding a broader array of novel drug candidates tailored to specific diseases. For Alzheimer's Disease (AD), augmented models predict drugs with 5 distinct mechanisms of action, compared to just 2 with non-augmented models (Figure 4a). This increased diversity could aid in better understanding disease mechanisms and support more personalized prescription options.

We observed that the effectiveness of a manipulation strategy depends on the position of the disease entity within the distribution of manipulation metric values of all entities in the graph. Specifically, when the disease entity of interest does not reside at the extremes of this distribution, the chosen manipulation strategy enhances the graph more informatively. In our case study focusing on AD, we found that the Propagation and Common Neighbours methods—both centrally positioned within their respective distributions (see Figure 2)—were the most effective models for transductive and semi-inductive reasoning, respectively (Figure 3). Central entities contribute to more robust outcomes due to their superior connectivity. In contrast, entities with topological structure values at the extremes of the distribution reduce the consistency of our augmentation. Even for diseases with limited existing knowledge, such as AD, identifying a topological structure within the KG that reflects the overall structure and selecting it for manipulation leads to more effective graph augmentation.

Limitations of our work include scalability issues, as identifying counterfactual edges between all drug-disease pairs increases training time when applied to large KGs. Additionally, our evaluation of drug candidates relies on specific metrics that may not fully reflect their practicality. Future work will validate our model across diverse diseases and manipulation strategies. This study lays a

solid foundation for using more complex graph structures to develop domain-specific augmentation methods that align with the observable topological structures of the entities of interest.

Acknowledgments

This work has been supported by NIH National Institute of Aging, USA R01 AG057557, R01 AG061388, R56 AG062272, National Institute on Alcohol Abuse and Alcoholism, USA (grant no. R01AA029831), National Eye Institute, USA (EY029297), National Institute on Drug Abuse, USA (UG1DA049435, CTN-0114), American Cancer Society Research Scholar, USA Grant RSG-16-049-01-MPC, The Clinical and Translational Science Collaborative (CTSC) of Cleveland, USA (UL1TR002548-01). The authors declare no competing interests. Ethics approval and consent to participate is not applicable.

References

Ahmed Alaa and Mihaela Van Der Schaar. Validating causal inference models via influence functions. In *International Conference on Machine Learning*, pp. 191–201. PMLR, 2019.

Joseph Martin Alisky. Intrathecal corticosteroids might slow alzheimer's disease progression. *Neuropsychiatric disease and treatment*, 4(5):831–833, 2008.

Stephen Bonner, Ian P Barrett, Cheng Ye, Rowan Swiers, Ola Engkvist, Andreas Bender, Charles Tapley Hoyt, and William L Hamilton. A review of biomedical datasets relating to drug discovery: a knowledge graph perspective. *Briefings in Bioinformatics*, 23(6):bbac404, 2022.

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.

Giorgos Bouritsas, Fabrizio Frasca, Stefanos Zafeiriou, and Michael M Bronstein. Improving graph neural network expressivity via subgraph isomorphism counting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):657–668, 2022.

Ayan Chatterjee, Robin Walters, Giulia Menichetti, and Tina Eliassi-Rad. Improving inductive link prediction through learning generalizable node representations. 2022.

Jaeyoon Chung, Nathan Sahelijo, Toru Maruyama, Junming Hu, Rebecca Panitch, Weiming Xia, Jesse Mez, Thor D Stein, Alzheimer's Disease Neuroimaging Initiative, Andrew J Saykin, et al. Alzheimer's disease heterogeneity explained by polygenic risk scores derived from brain transcriptomic profiles. *Alzheimer's & Dementia*, 2023.

Ryan Crowley. Graph representation learning for child mental health prediction.

IV Damulin, DA Stepkina, and AB Lokshina. Neuromidin in mixed vascular and alzheimer's dementia. *Zhurnal Nevrologii i Psikhiatrii Imeni SS Korsakova*, 111(2):40–43, 2011.

Jesus de la Fuente, Guillermo Serrano, Uxía Veleiro, Mikel Casals, Laura Vera, Marija Pizurica, Antonio Pineda-Lucena, Idoia Ochoa, Silve Vicent, Olivier Gevaert, et al. Towards a more inductive world for drug repurposing approaches. *arXiv preprint arXiv:2311.12670*, 2023.

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.

Siddhant Doshi and Sundeep Prabhakar Chepuri. A computational approach to drug repurposing using graph neural networks. *Computers in Biology and Medicine*, 150:105992, 2022.

Jakoba J Eertink, Martijn W Heymans, Gerben JC Zwezerijnen, Josée M Zijlstra, Henrica CW de Vet, and Ronald Boellaard. External validation: a simulation study to compare cross-validation versus holdout or external testing to assess the performance of clinical prediction models using pet data from dlbcl patients. *EJNMMI research*, 12(1):58, 2022.

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

Janan T Eppig, Cynthia L Smith, Judith A Blake, Martin Ringwald, James A Kadin, Joel E Richardson, and Carol J Bult. Mouse genome informatics (mgi): resources for mining mouse genetic, genomic, and biological data in support of primary and translational research. *Systems Genetics: Methods and Protocols*, pp. 47–73, 2017.

Mikhail Galkin, Max Berrendorf, and Charles Tapley Hoyt. An Open Challenge for Inductive Link Prediction on Knowledge Graphs. mar 2022. URL http://arxiv.org/abs/2203.01520.

Mikhail Galkin, Xinyu Yuan, Hesham Mostafa, Jian Tang, and Zhaocheng Zhu. Towards foundation models for knowledge graph reasoning. *arXiv preprint arXiv:2310.04562*, 2023.

Jianfei Gao, Yangze Zhou, Jincheng Zhou, and Bruno Ribeiro. Double equivariance for inductive link prediction for both new nodes and new relation types. *arXiv preprint arXiv:2302.01313*, 2023.

Zhenxiang Gao, Pingjian Ding, and Rong Xu. Kg-predict: A knowledge graph computational framework for drug repurposing. *Journal of biomedical informatics*, 132:104133, 2022.

Robert A Hanneman and Mark Riddle. Introduction to social network methods, 2005.

Kexin Huang, Cao Xiao, Lucas M Glass, and Jimeng Sun. Moltrans: molecular interaction transformer for drug–target interaction prediction. *Bioinformatics*, 37(6):830–836, 2021.

Xingyue Huang, Miguel Romero, Ismail Ceylan, and Pablo Barceló. A theory of link prediction via relational weisfeiler-leman on knowledge graphs. *Advances in Neural Information Processing Systems*, 36, 2024.

Bradley T Hyman, Paul J Eslinger, and Antonio R Damasio. Effect of naltrexone on senile dementia of the alzheimer type. *Journal of Neurology, Neurosurgery & Psychiatry*, 48(11):1169–1171, 1985.

Md Kamrul Islam, Diego Amaya-Ramirez, Bernard Maigret, Marie-Dominique Devignes, Sabeur Aridhi, and Malika Smaïl-Tabbone. Molecular-evaluated and explainable drug repurposing for covid-19 using ensemble knowledge graph embedding. *Scientific Reports*, 13(1):3643, 2023.

Fredrik Johansson, Uri Shalit, and David Sontag. Learning representations for counterfactual inference. In *International conference on machine learning*, pp. 3020–3029. PMLR, 2016.

Elliot Jones, Sabrina Küspert, Nicolas Moës, and Connor Dunlop. Explainer: What is a foundation model?, Jul 2023. URL https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Eiji Kirino, Yayoi Hayakawa, Rie Inami, Reiichi Inoue, and Shigeki Aoki. Simultaneous fmri-eeg-dti recording of mmn in patients with schizophrenia. *PLoS One*, 14(5):e0215023, 2019.

Nithya Krishnamurthy, Alyssa A Grimshaw, Sydney A Axson, Sung Hee Choe, and Jennifer E Miller. Drug repurposing: a systematic review on root causes, barriers and facilitators. *BMC health services research*, 22(1):970, 2022.

Michael Kuhn, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. The sider database of drugs and side effects. *Nucleic acids research*, 44(D1):D1075–D1079, 2016.

VS Kulkarni, V Alagarsamy, VR Solomon, PA Jose, and S Murugesan. Drug repurposing: an effective tool in modern drug discovery. *Russian Journal of Bioorganic Chemistry*, 49(2):157–166, 2023.

Xiling Lei, Jing Yu, Qi Niu, Jianhua Liu, Patrick C Fraering, and Fang Wu. The fda-approved natural product dihydroergocristine reduces the production of the alzheimer's disease amyloid-β peptides. *Scientific reports*, 5(1):16541, 2015.

Zongren Li, Qin Zhong, Jing Yang, Yongjie Duan, Wenjun Wang, Chengkun Wu, and Kunlun He. Deepkg: an end-to-end deep learning-based workflow for biomedical knowledge graph extraction, optimization and applications. *Bioinformatics*, 38(5):1477–1479, 2022.

Feiyan Liu, Liangzhi Li, Xiaoli Wang, Feng Luo, Chang Liu, Jinsong Su, and Yiming Qian. Mhgrl: An effective representation learning model for electronic health records. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 11272–11282, 2024.

Kevin McCoy, Sateesh Gudapati, Lawrence He, Elaina Horlander, David Kartchner, Soham Kulkarni, Nidhi Mehra, Jayant Prakash, Helena Thenot, Sri Vivek Vanga, et al. Biomedical text link prediction for drug discovery: a case study with covid-19. *Pharmaceutics*, 13(6):794, 2021.

Christopher Morris, Martin Ritzert, Matthias Fey, William L Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and leman go neural: Higher-order graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 4602–4609, 2019.

Mark Newman. Networks, 2nd edn oxford. *UK: Oxford University Press.[Google Scholar]*, 2018.

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

Shagofah Noor, Omid Tajik, and Jawad Golzar. Simple random sampling. 1:78–82, 12 2022. doi: 10.22034/ijels.2022.162982.

Elif Ozkan, Remzi Celebi, Arif Yilmaz, Vincent Emonet, and Michel Dumontier. Generating knowledge graph based explanations for drug repurposing predictions. In *14th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences*, pp. 22–31, 2023.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 701–710, 2014.

Janet Piñero, Núria Queralt-Rosinach, Alex Bravo, Jordi Deu-Pons, Anna Bauer-Mehren, Martin Baron, Ferran Sanz, and Laura I Furlong. Disgenet: a discovery platform for the dynamical exploration of human diseases and their genes. *Database*, 2015:bav028, 2015.

Peter N Robinson and Stefan Mundlos. The human phenotype ontology. *Clinical genetics*, 77(6): 525–534, 2010.

Andrea Rossi, Denilson Barbosa, Donatella Firmani, Antonio Matinata, and Paolo Merialdo. Knowledge graph embedding for link prediction: A comparative analysis. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 15(2):1–49, 2021.

Nahid Safari-Alighiarloo, Mohammad Taghizadeh, Mostafa Rezaei-Tavirani, Bahram Goliaei, and Ali Asghar Peyvandi. Protein-protein interaction networks (ppi) and complex diseases. *Gastroenterology and Hepatology from bed to bench*, 7(1):17, 2014.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *The semantic web: 15th international conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, proceedings 15*, pp. 593–607. Springer, 2018.

Michael Serby, Richard Resnick, Barbara Jordan, John Adler, June Corwin, and John P Rotrosen. Naltrexone and alzheimer's disease. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 10(3-5):587–590, 1986.

Hyoung Seok Shin. Reasoning processes in clinical reasoning: from the perspective of cognitive psychology. *Korean journal of medical education*, 31(4):299, 2019.

Balasubramaniam Srinivasan and Bruno Ribeiro. On the equivalence between positional node embeddings and structural graph representations. *arXiv preprint arXiv:1910.00452*, 2019.

Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. *arXiv preprint arXiv:1902.10197*, 2019.

Mark J Van der Laan and Maya L Petersen. Causal effect models for realistic individualized treatment and intention to treat rules. *The international journal of biostatistics*, 3(1), 2007.

Vladimir N Vapnik. An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5):988–999, 1999.

Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha Talukdar. Composition-based multi-relational graph convolutional networks. *arXiv preprint arXiv:1911.03082*, 2019.

Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, Nilesh Agrawal, and Partha Talukdar. Interacte: Improving convolution-based knowledge graph embeddings by increasing feature interactions. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 3009–3016, 2020.

Fangping Wan, Lixiang Hong, An Xiao, Tao Jiang, and Jianyang Zeng. Neodti: neural integration of neighbor information from a heterogeneous network for discovering new drug–target interactions. *Bioinformatics*, 35(1):104–111, 2019.

David S Wishart, Craig Knox, An Chi Guo, Dean Cheng, Savita Shrivastava, Dan Tzur, Bijaya Gautam, and Murtaza Hassanali. Drugbank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic acids research*, 36(suppl_1):D901–D906, 2008.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

Rong Xu and QuanQiu Wang. Large-scale extraction of accurate drug-disease treatment pairs from biomedical literature for drug repurposing. *BMC bioinformatics*, 14:1–11, 2013.

Rong Xu, Li Li, and QuanQiu Wang. Towards building a disease-phenotype knowledge base: extracting disease-manifestation relationship from literature. *Bioinformatics*, 29(17):2186–2194, 2013.

Rong Xu, Li Li, and QuanQiu Wang. driskkb: a large-scale disease-disease risk relationship knowledge base constructed from biomedical text. *BMC bioinformatics*, 15:1–13, 2014.

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.

Qichang Zhao, Haochen Zhao, Kai Zheng, and Jianxin Wang. Hyperattentiondti: improving drug–protein interaction prediction by sequence-based deep learning with attention mechanism. *Bioinformatics*, 38(3):655–662, 2022a.

Tong Zhao, Gang Liu, Daheng Wang, Wenhao Yu, and Meng Jiang. Learning from counterfactual links for link prediction. In *International Conference on Machine Learning*, pp. 26911–26926. PMLR, 2022b.

Yongjun Zhu, Chao Che, Bo Jin, Ningrui Zhang, Chang Su, and Fei Wang. Knowledge-driven drug repurposing using a comprehensive drug knowledge graph. *Health Informatics Journal*, 26(4): 2737–2750, 2020.

Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal Xhonneux, and Jian Tang. Neural bellman-ford networks: A general graph neural network framework for link prediction. *Advances in Neural Information Processing Systems*, 34:29476–29490, 2021.

A Appendix

A.1 Manipulation Definitions

We experimented with the following manipulations Newman (2018), where A is the adjacency matrix for graph G:

- **Katz Centrality**: $M = ((I - \beta A)^{-1} - I \geq \theta)$, θ is a threshold value and $\beta = \min\left(\frac{1}{\lambda_{\max}}, 0.003\right)$
- **Common Neighbours:** $M = ((A + A^2 + A^3 > 0) \times (A + A^2 + A^3 > 0)^T \geq k)$, k is a threshold value.
- **Louvain:** $M = (LL^T)$, L is the membership matrix derived from applying Louvain to A.
- **K-Core:** $M = (KK^T)$, K is the membership matrix derived from the k-core labels of the nodes in A.
- **Propagation:** $M = (PP^T)$, P is the membership matrix derived from the propagation clustering community labels of the nodes in A.

A.2 Embedding Process

The relation embedding process is shown in Algorithm 1, and the entity embedding and link prediction process is shown in Algorithm 2. The foundation model captures interactions by creating a graph of relations $G_r = (R, R_{fund}, E_r)$ to map out how different types of relations interact. In G_r, all $r \in R$ are unique relation types in the original graph G, and two nodes are linked if the corresponding types of relations in G share a common entity. $R_{fund} = \{$tail-to-tail edges, head-to-head edges, head-to-tail edges, tail-to-head edges$\}$ which are the four types of links in G_r that describe the basic topological structure of a directed graph.

Specifically, in our graph of relations, G_r, each node represents a unique relation type from the original graph, G. Two nodes in G_r are connected if their corresponding relation types in G share a common node. Depending on the directionality of these connections in G, the connections in G_r are classified as either tail-to-tail, head-to-head, head-to-tail, or tail-to-head.

Embeddings are obtained using conditional representations Zhu et al. (2021), which are demonstrably more expressive and practical than unconditional GNN encoders Huang et al. (2024); Gao et al. (2023). By doing so, node representations are conditioned on query relations.

In Algorithm 1 and 2, **AGG** is a summation aggregation function and **MSG** is a multiplication messaging function.

Algorithm 1 Relation Graph Embedding

Input: Original graph G, Query $(h, q, ?)$, Dimension d
Output: d-dimensional node representations N_q of G_r

function COMPUTEREPR(q, G_r, d)
    for each node u in G_r **do**
        $N_q[u] \leftarrow (u == q)?\mathbf{1}^d : \mathbf{0}^d$
    return N_q

function MESSAGEPASSING(G_r, N_q)
    $t \leftarrow 0$
    repeat
        for each node v in G_r **do**
            $h_v[t+1] \leftarrow$ UPDATE$(h_v[t], \text{AGG}(\{\text{MSG}(h_w[t], r) \mid w \in neighbors(v), r \in R_{\text{fund}}\}))$
        $t \leftarrow t + 1$
    until convergence
    return $h_v[t]$

$G_r \leftarrow$ ConstructGr(G)
$N_q \leftarrow$ ComputeRepr(q, G_r, d)
$N_q \leftarrow$ MessagePassing(G_r, N_q)

Algorithm 2 Entity Embedding and Link Prediction

Input: Graph G, Query $(h, q, ?)$, Conditional relation representations R_q, Dimension d
Output: Logits $p(h, q, v)$ for each node v in G

function INITGRAPH(G, h, q, R_q, d)
    $h_0 \leftarrow$ Vector($|G|, d, 0$) ▷ Zero vector for each node
    $h_0[h] \leftarrow R_q[q]$ ▷ Set query vector for head node
    return h_0

function MESSAGEPASSING(G, h_0, R_q, d)
    $t \leftarrow 0$
    $R_t \leftarrow$ TransRelations(R_q) ▷ Apply 2-layer MLP
    repeat
        for each node v in G **do**
            $h_v[t+1] \leftarrow$ UPDATE$(h_v[t],$ AGG$(\{$MSG$(h_w[t], R_t[r]) \mid w \in neighbors(v)\}))$
        $t \leftarrow t + 1$
    until convergence
    return $h_v[t]$

$h_0 \leftarrow$ InitGraph(G, h, q, R_q, d)
$h_v \leftarrow$ MessagePassing(G, h_0, R_q, d)
$logits \leftarrow$ ComputeLogits(h_v)

A.3 Preserving Equivariance

A.3.1 Invariance and Equivariance

In inductive inference, neural networks are trained to discern and generalize invariant features within a graph to accommodate novel data. Xu et al. (2018); Morris et al. (2019) has revealed a notable parallel between the Weisfeiler-Lehman (WL) test and the operations of graph message passing neural networks, which facilitate convolution-like functionalities on graphs. A standard GNN updates the hidden state $h_v^{(k)}$ of a node v at layer k as: $h_v^{(k)} = \phi\left(h_v^{(k-1)}, f\left(\left\{h_u^{(k-1)} : u \in N(v)\right\}\right)\right)$ Kipf & Welling (2016), where ϕ denotes a transformation operation, f denotes an aggregation operation, $N(v)$ denote the neighboring nodes from previous layer $k-1$. Whereas, a 1-WL test iteration represents the label $l_v^{(k)}$ of node v as: $l_v^{(k)} = g\left(l_v^{(k-1)}, \text{sort}\left\{l_u^{(k-1)} : u \in N(v)\right\}\right)$ Xu et al. (2018), where g is an injective function. Xu et al. (2018) show by induction that (Theorem 3 in their paper) there always exists an injective function φ that $h_v^{(k)} = \varphi\left(l_v^{(k)}\right)$, which show that WL's 1-dimensional form 1-WL is analogous to GNN's neighbor aggregation. Showing the existence of the injective function φ is stating that for any given unique label $l_v^{(k)}$, there is a unique embedding $h_v^{(k)}$ in the GNN. It implies that these embeddings are independent of node IDs, similar to the WL test's ability to create unique labels for nodes based on their structure, rather than arbitrary properties like node IDs. Hence, these studies have underscored the benefits for inductive graph tasks from assuming node IDs are arbitrary, suggesting that models should maintain consistency despite permutations of node IDs. This concept, known as *permutation equivariance*, implies that the network does not rely on node IDs for embedding but rather employs the graph's structural topology for representation Srinivasan & Ribeiro (2019).

Extending this to multi-relational graphs such as KGs, the notion of learnable invariances extends to the joint permutations of node IDs and relation IDs, a concept referred to as *double permutation equivariance* or *double equivariance* defined by Gao et al. (2023). If you consider a graph that is processed by an RGCN (Relational Graph Convolutional Network) Schlichtkrull et al. (2018), the goal is for the RGCN to learn representations of nodes and relations from their structure rather than their labels or order. Given we have a permutation operation π, which shuffles the order of nodes and the types of relations between them. Permutation invariance in multi-relational graphs like RGCNs can be captured by demonstrating that the aggregation function f we defined previously is

permutation-invariant concerning the representations of neighboring nodes of v through relation type r as $(N^r(v))$, then permutation invariance is established which can be defined as the aggregation function f_r:

$$f_r(\{h_u^{(k-1)}, w_r \quad : \quad u \quad \in \quad N^r(v)\}) \quad = \quad f_r(\{h_{\pi(u)}^{(k-1)}, w_r \quad : \quad \pi(u) \quad \in \quad \pi(N^r(v))\}) \quad (1)$$

Theorem 1. *Graphs with counterfactual edge augmentation preserve double equivariance.*

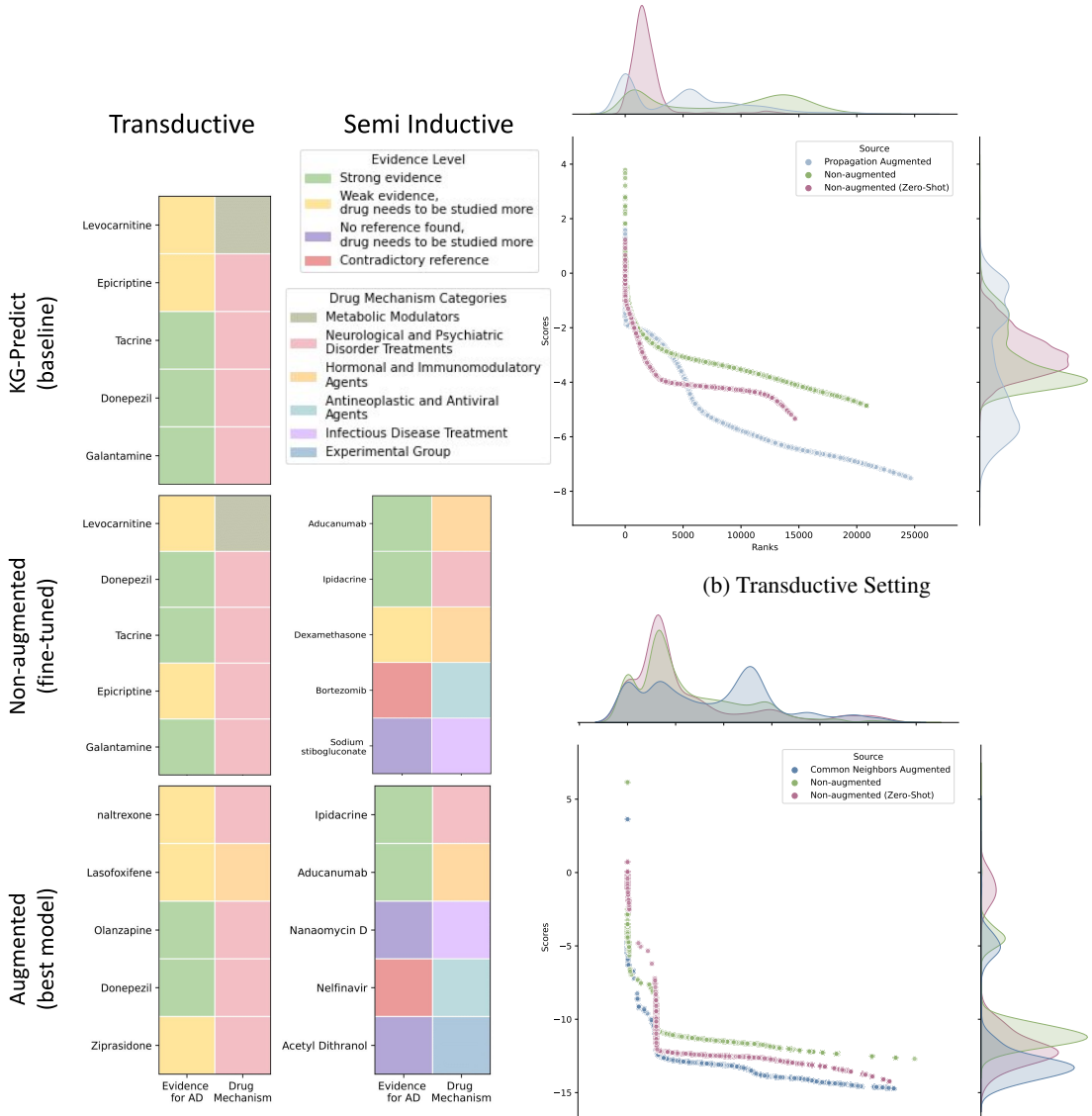
$$f_r(\{h_u^{(k-1)}, w_r \quad : \quad u \quad \in \quad N_{cf}^r(v)\}) \quad = \quad f_r(\{h_{\pi(u)}^{(k-1)}, w_r \quad : \quad \pi(u) \quad \in \quad \pi(N_{cf}^r(v))\}) \quad (2)$$

Proof. Let $G = (V, L, F)$ be the original KG, and $G' = (V, L \cup L_{cf}, F \cup F_{cf})$ represent the graph augmented with counterfactual edges. N_{cf}^r is the neighborhood of node v in G'. The aggregation function for relation r at layer k is denoted as f_r, which aggregates features from the neighborhood $N_{cf}^r(v)$ of a node v through relation r. Initially defined in Equation 1, the relation-based aggregation function f_r is permutation-invariant.

Since the addition of counterfactual edges L_{cf}, F_{cf} to the graph G to form G' simply extends the neighborhood $N_{cf}^r(v)$ without altering the inherent permutation-invariant property of f_r, the aggregation function will process $N_{cf}^r(v)$ and $\pi(N_{cf}^r(v))$ identically. Therefore, f_r preserves double equivariance in the presence of counterfactual edges, as the permutation of node IDs does not affect the outcome of the aggregation function. □

A.4 Analysis of Drug Candidates

In the transductive setting, treatments for neurological and psychiatric disorders are prioritized, leading to a range of drugs with varying evidence levels for repurposing in AD. The transductive setting however, is not optimal for identifying more novel repurposing strategies. The non-augmented model presents the same drug candidates as the baseline model KG-Predict, except in a different order. In contrast, our augmented model recommends additional candidates that show weak evidence of potential repurposing for AD, differing from both the KG-Predict baseline and the non-augmented model. In the semi-inductive setting, as the model was not trained on the complete set of entities and relations, the drug candidates are considerably more diverse than in the transductive setting. We find that the top two drug candidates (Aducanumab, Ipidacrine) are identical for both the non-augmented and augmented models, exhibiting strong evidence in the mechanisms of hormonal immunomodulatory agents and treatments for neurological and psychiatric disorders (Table 5). The augmented model has identified a greater number of drugs with no existing references, indicating these require further investigation.



(a) Top 5 Drug Candidates proposed across models

(b) Transductive Setting

(c) Semi-inductive Setting

Figure 4: **(a)** A high-level summary of the evidences and drug mechanisms provided in Table 4 and Table 5. Best models for the augmented setting are Propagation augmented model for the transductive setting, and Common Neighbors augmented model for the semi-inductive setting. **(b)** and **(c)** Rankings and scores as predicted by models, assessing internal consistency within models based on the case study. The best augmented models and non-augmented baselines are shown. Each scatter indicates a drug candidate ranked by the model, with logit scores.

Table 4: Evidence and mechanism of the top 3 drug candidates for Alzheimer's Disease, predicted by Transductive Model. The full version of this table and the tables for other models' drug candidates are presented in https://github.com/ceragoguztuzun/KGIA.

Drug Candidates	Evidence for AD	Drug Mechanism
KG-Predict		
Levocarnitine	No clinical trial for levocarnitine itself, but clinical trial NCT02955706 for its derivative Acetyl-L-carnitine (L-carnitine is a synonym for levocarnitine)	Prevent and treat a lack of carnitine in patients with kidney disease on dialysis. A carrier of long chain fatty acids to mitochondria for energy.
Epicriptine	No reference, but with indication in Alzheimer's disease on Drug Central website and its analog dihydroergocristine is proved to delay progression of Alzheimer's disease Lei et al. (2015)	Nootropic to treat signs and symptoms of an idiopathic decline in mental capacity. Mainly due to the agonistic activity on dopamine receptor
Tacrine	First drug approved for Alzheimer's disease	Treat AD. A powerful acetylcholinesterase inhibitor (AChE)
Non-augmented		
Levocarnitine	No clinical trial for levocarnitine itself, but clinical trial NCT02955706 for its derivative Acetyl-L-carnitine (L-carnitine is a synonym for levocarnitine)	Prevent and treat a lack of carnitine in patients with kidney disease on dialysis. A carrier of long chain fatty acids to mitochondria for energy.
Donepezil	NCT02787746(completed), NCT04661280(recruiting)	Treat AD. Inhibiting the acetylcholinesterase enzyme
Tacrine	First drug approved for Alzheimer's disease	Treat AD. A powerful acetylcholinesterase inhibitor (AChE)
KGïA (Propagation Aug)		
naltrexone	Serby et al. (1986); Hyman et al. (1985) (but show no significant improvement)	Treat alcohol dependence and block the effects of exogenously administered opioids. It is a pure opiate antagonist and has little or no agonist activity.
Lasofoxifene	Selective estrogen receptor modulators as brain therapeutic agents (further study is needed)	An estrogenic and antiestrogenic drug used in both vitro and in vivo to treat osteoporosis and vaginal atrophy.
Olanzapine	NCT00015548(completed, olanzapine is used in combination with other two drugs)	The drug is used to treat schizophrenia, bipolar disorder, and agitation associated with these disorders. It inhibits multiple neuronal receptors, including D1, D2, D3, and D4.

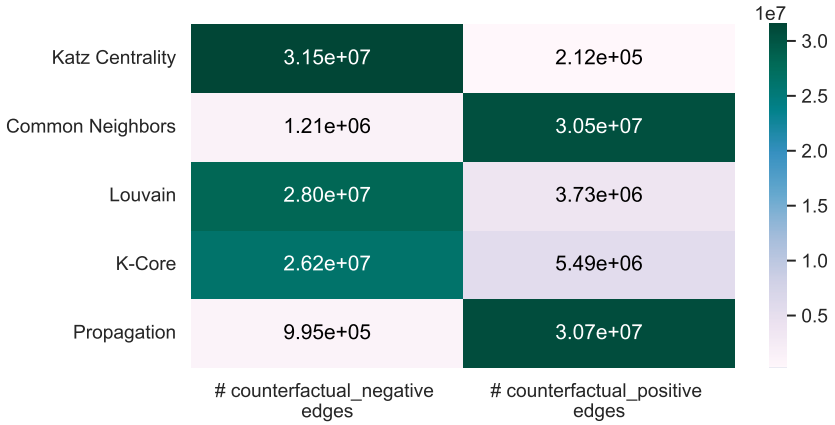


Figure 5: The number of counterfactual edges that were found in the knowledge graph, for each choice of manipulation.

Table 5: Evidence and mechanism of the top 3 drug candidates for Alzheimer's Disease, predicted by Semi-Inductive Model. The full version of this table and the tables for other models' drug candidates are presented in https://github.com/ceragoguztuzun/KGIA

.

Drug Candidates	Evidence for AD	Drug Mechanism
Non-augmented		
Aducanumab	NCT04241068 (active, not recruiting), NCT02477800(terminated with results)	Treat AD. It is a human immunoglobulin gamma 1 (IgG1) monoclonal antibody which selectively targets and binds aggregated soluble oligomers and insoluble fibril conformations of Aβ plaques.
Ipidacrine	Also named Neiromidin or Neuromidin(translated from Russian) Damulin et al. (2011).	Treat memory disorders of different origins. A reversible acetylcholinesterase inhibitor
Dexamethasone	Alisky (2008) (It is a case report and hypothesis, needs further validation)	Inflammatory disorders such as asthma, endocrine disorders, and rheumatoid arthritis are treated with glucocorticoids that inhibit pro-inflammatory signals and promote anti-inflammatory ones.
KGïA (Common Neighbors Aug)		
Ipidacrine	Also named Neiromidin or Neuromidin(translated from Russian) Damulin et al. (2011).	Treat memory disorders of different originsa. A reversible acetylcholinesterase inhibitor
Aducanumab	NCT04241068 (active, not recruiting), NCT02477800(terminated with results)	Treat AD. It is a human IgG1 monoclonal antibody which selectively targets and binds aggregated soluble oligomers and insoluble fibril conformations of Aβ plaques.
Nanaomycin D	No reference	An enantiomer of the antibiotic kalafungin, is a natural product found in Streptomyces and Streptomyces rosa.

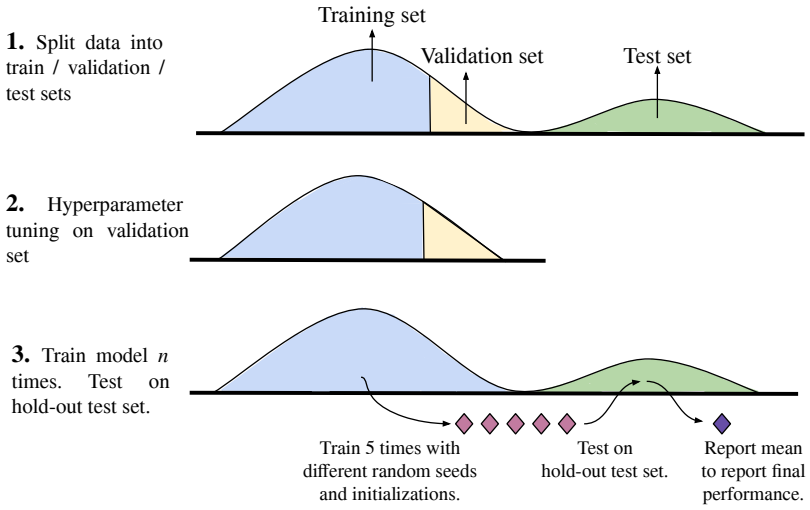


Figure 6: Overview of our evaluation process: (i) The dataset was divided into a training set and a fixed test set. (ii) Hyperparameter tuning was conducted using the validation set. (iii) Each model was trained five times with different random seeds and initializations, while evaluation was consistently carried out on the same fixed test set. Performance metrics were collected from each of the five runs, and the average of these metrics was reported as the final performance result.