Graph Representation Learning for Predicting Drug-Target Associations in Complex Human Disease Networks

Myrthe Remmers
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
contactmyrthe@uab.edu

Galasso Trevisani
Department of Computer Science, University of New Hampshire, Durham, NH, USA.
galassot@unh.edu

Abstract

Predicting drug-target associations in complex human disease networks requires models that can integrate molecular interaction topology, disease-module structure, and heterogeneous pharmacological evidence without collapsing them into a single similarity score. Existing supervised drug-target interaction models often treat diseases as label contexts rather than as structured network perturbations, which limits their ability to distinguish shared targets from disease-specific vulnerabilities. This manuscript develops a graph representation learning framework that conditions drug-target association scoring on disease-network topology. The framework represents drugs, targets, disease modules, pathways, and anatomical or functional contexts as typed nodes and relations; propagates information through relation-specific message passing; and uses a disease-context gate to modulate target embeddings. A bilinear association head produces calibrated probabilities, while the training objective combines positive-unlabeled learning, relation-aware regularization, and negative sampling that controls for target popularity and literature bias. The contribution is methodological: it specifies how disease modules enter representation learning, how graph topology affects target prioritization, and how evaluation should separate predictive discrimination from mechanistic plausibility. The framework also identifies failure modes that are specific to complex disease networks, including literature-driven edge sparsity, confounding by drug promiscuity, and leakage across disease splits. The resulting design provides a technically coherent basis for prioritizing drug-target associations in network medicine and for interpreting model outputs as evidence-ranked hypotheses rather than validated therapeutic claims.

Keywords

graph representation learning; drug-target association; disease networks; network medicine; heterogeneous graphs; calibration

1. Introduction

Complex human diseases are rarely explained by a single molecular lesion. They emerge from distributed perturbations across protein interaction, regulatory, metabolic, and tissue-specific networks, and their clinical phenotypes reflect the overlap of multiple disease modules rather than the behavior of isolated genes [1]. The human disease network formalizes

this observation by linking disorders through shared disease genes, revealing modular organization and comorbidity patterns that are not apparent from gene-level annotation alone [2]. Subsequent analyses of incomplete interactomes have shown that disease modules can be topologically close even when their gene sets differ, and that network separation provides information about disease relationships beyond semantic or phenotypic similarity [3]. These findings have direct consequences for drug-target association prediction: a target that is central to one disease module may be marginal in another, and a drug that modulates a shared hub may produce benefits or adverse effects that depend on the disease context.

Computational drug-target interaction prediction has developed along several lines. Early approaches integrated chemical and genomic spaces to infer bipartite drug-target relationships [4], and supervised formulations used known interactions with chemical and genomic similarities to score unseen pairs [5]. Network integration methods later incorporated heterogeneous evidence, including drug similarity, target similarity, and disease associations, to support drug repositioning [6]. These approaches established the value of relational data, but they often operate on homogeneous similarity graphs or on bipartite projections in which the disease is represented only as a downstream label. The result is a representation that can estimate whether a drug binds a target but may not distinguish whether that target is relevant to a particular disease network.

Graph representation learning offers a natural formalism for the problem. Graph convolutional networks aggregate local neighborhood information and can be trained end-to-end with node-level or edge-level objectives [8], while graph attention mechanisms learn relation-dependent importances among neighbors [9]. Drug databases and interaction resources provide the typed entities and edges needed for such models [7], but the availability of data does not by itself resolve the modeling problem. A drug-target graph is not a generic citation network: edges have different biological meanings, negative evidence is rarely observed, and the same target can participate in multiple disease modules with different topological roles. A representation learned from the full graph may therefore encode popularity and annotation density rather than disease-specific mechanism.

This manuscript addresses that gap by specifying a disease-context-aware graph representation learning framework for drug-target association prediction in complex human disease networks. The central contribution is not a new benchmark result but a technically developed modeling strategy that treats disease modules as first-class graph contexts. The framework combines typed graph construction, relation-specific message passing, disease-context gating, bilinear association scoring, and calibrated uncertainty. It also develops an evaluation logic that distinguishes discrimination from mechanistic plausibility and that controls for common confounders in biomedical graph learning. The discussion is deliberately framed around design assumptions and derivable consequences, because claims about predictive superiority or clinical utility require empirical evaluation that is outside the scope of the present analysis.

2. Problem Formulation and Biological Constraints

2.1 Disease modules as contextual variables

A disease network can be represented as a graph whose nodes denote genes, proteins, pathways, or phenotypes and whose edges denote physical, regulatory, or functional relations. For a disease c, let M_c be the set of nodes assigned to its disease module. The module is not a fixed biological object; it depends on evidence thresholds, tissue context, and the granularity

of pathway annotation. This uncertainty matters for prediction. If M_c is defined too broadly, the model will learn generic essentiality; if it is defined too narrowly, the model will overfit to a small set of literature-curated genes. The framework therefore treats M_c as a set of candidate contexts and uses graph topology, rather than membership alone, to construct a disease representation.

2.2 Heterogeneous graph construction

The input graph contains multiple node types and relation types. Drug nodes carry chemical, structural, and pharmacological attributes. Target nodes carry sequence, domain, and pathway attributes. Disease nodes carry phenotype and module annotations. Pathway, tissue, and functional-context nodes provide intermediate biological structure. Edges are typed: drug-target binding, drug-side-effect, target-pathway membership, pathway-disease association, protein-protein interaction, and disease-disease similarity are not interchangeable. Formally, the graph is defined as a tuple:

$$G = (V, E, R, X, Z)$$

Here V is the node set, E is the edge set, R is the set of relation types, X maps nodes to feature vectors, and Z maps edges to relation-specific attributes such as confidence scores or evidence counts. The tuple is intentionally general: it permits the model to use curated databases, predicted interactions, and text-mined associations while requiring that their provenance be encoded in Z rather than flattened into a binary edge. This design choice is important because a high-confidence physical interaction and a low-confidence co-occurrence assertion should not exert the same influence on a drug-target score.

2.3 Community structure and polypharmacology

Disease modules and drug mechanisms both exhibit community structure. A drug may affect several targets within a pathway community, and a disease may involve several communities that are weakly connected. Network-community formulations have been used to identify synergistic combinations from herbal medicines and complex systems [10], illustrating how community-level organization can inform combination-oriented prediction. That line of work is relevant because it treats multi-target effects as structured rather than as independent edges. The present framework adopts the same general premise, namely that association prediction should respect module-level organization, but it focuses on drug-target associations conditioned on human disease networks rather than on herbal combination discovery.

The biological implication is that a target should not be scored solely by its similarity to known targets of a drug. A target embedded in a disease module may be more plausible if it lies on a path between the drug's known targets and the disease module, or if it occupies a bottleneck that connects two disease-relevant communities. Conversely, a high-degree hub may be statistically salient but mechanistically nonspecific. The representation learning problem is therefore to preserve both local chemical similarity and global disease-network position without allowing either to dominate.

3. Graph Representation Learning Framework

3.1 Node and edge encodings

Each node receives an initial feature vector. Drug features can include molecular fingerprints, learned chemical embeddings, and pharmacological annotations. Target features can include sequence-derived embeddings, domain annotations, and expression summaries.

Disease features can include phenotype ontology terms and module-level statistics. Because the graph is heterogeneous, node features are projected into a common latent space by type-specific linear maps. This allows message passing to compare, for example, a drug node with a pathway node through a typed relation while retaining type-specific information in the projection.

The framework is inductive in spirit. New drugs or targets can be embedded from their attributes and local neighborhoods without requiring retraining of all node parameters, following the logic of inductive neighborhood aggregation [11]. Earlier transductive embedding methods such as node2vec demonstrated that neighborhood-preserving random walks can capture structural roles [12], but they do not naturally incorporate typed edges or disease contexts. The proposed model therefore uses message passing rather than walk-based factorization as the primary representation mechanism.

3.2 Relation-specific message passing

At each layer, a node updates its representation by aggregating messages from neighbors, with separate transformations for each relation type. The update is:

$$h_i' \;=\; \sigma(\, W_0\, h_i \;+\; \sum_\{r \in R\} \sum_\{j \in N_r(i)\}\, \alpha_ij\hat{}(r)\, W_r\, h_j \,)$$

Here h_i is the current representation of node i, h_i' is the updated representation, σ is a nonlinear activation, W_0 is a self-transformation, R is the set of relation types, N_r(i) is the set of neighbors of i under relation r, W_r is a relation-specific weight matrix, and α_ij^(r) is an attention coefficient. The self-transformation preserves information that would otherwise be washed out by neighborhood averaging. Relation-specific weights prevent protein-protein interactions from being treated identically to drug-target binding or pathway membership.

The attention coefficient is computed as:

$$\alpha_ij\hat{}(r) \;=\; softmax_j(\, a_r\hat{}T\, [\, W_r\, h_i \,||\, W_r\, h_j \,||\, e_ij\hat{}(r)\,]\,)$$

In this expression, a_r is a learnable attention vector for relation r, || denotes concatenation, and e_ij^(r) is the edge attribute vector. The softmax is taken over neighbors j in N_r(i). The edge attributes allow the model to distinguish strong from weak evidence and to represent uncertainty in predicted interactions. Without this term, a literature-curated edge and a low-confidence inferred edge would contribute equally, which is undesirable in biomedical graphs where annotation density varies widely.

Relation-aware message passing is related to graph convolutional and graph attention architectures [8,9], but the present formulation differs in two respects. First, edge attributes are included directly in attention, so provenance affects propagation. Second, the model is trained with a disease-context objective rather than only a link-prediction objective. These choices connect representation learning to the biological question: not merely whether a drug and target are adjacent in the graph, but whether the association is coherent within a disease module.

3.3 Disease-context gating

A disease context is represented by pooling the embeddings of its module nodes. Let M_c be the set of nodes in the module for disease c and s_c be the context vector:

$$s_c \;=\; (1/|M_c|) \sum_\{k \in M_c\}\, h_k$$

The context vector alone is not sufficient, because different targets within the same disease module may play different roles. A gating function therefore modulates each target representation according to its compatibility with the disease context:

$$\gamma_(i,c) = \sigma(\, q\text{^}T \tanh(\, V\, h_i + U\, s_c + b\,)\,)$$

Here γ_(i,c) is a scalar gate for node i under disease context c, q is a learnable vector, V and U are weight matrices, and b is a bias. The gate is applied multiplicatively to the node representation before association scoring. A high gate indicates that the target's local embedding and the disease-module aggregate are mutually consistent; a low gate suppresses targets whose graph neighborhood is poorly aligned with the disease context.

This gating mechanism is a response to a specific failure mode in graph-based drug-target prediction. A target can have many known interactions and therefore receive high attention in the graph, yet be irrelevant to the disease under study. Conversely, a less-studied target may be topologically positioned between two disease-relevant modules and should not be discarded merely because it has low degree. The gate introduces disease dependence directly into the scoring function and prevents a single global target embedding from serving all diseases.

3.4 Association scoring and objective

For a drug d, target t, and disease context c, the association score is defined as a bilinear function of their context-adjusted embeddings:

$$p(d,t|c) = \sigma(\, z_d\text{^}T\, diag(w_c)\, z_t + b_d + b_t\,)$$

Here p(d,t|c) is the predicted probability of a disease-relevant drug-target association, z_d and z_t are the drug and target embeddings after context adjustment, w_c is a disease-specific weighting vector, diag(w_c) forms a diagonal matrix, and b_d and b_t are drug and target biases. The diagonal form is a deliberate compromise between expressiveness and parameter efficiency. A full bilinear tensor would allow arbitrary drug-target-disease interactions but would be prone to overfitting in sparse disease modules. The diagonal weighting captures disease-specific feature importance while keeping the number of parameters manageable.

Training data consist of positive drug-target-disease tuples and unlabeled or sampled negative tuples. Because negative evidence is rarely recorded, the objective should be interpreted as positive-unlabeled learning rather than fully supervised classification. A suitable objective is:

$$L = -\sum_(i \in D)\, [\, y_i\, log\, p_i + (1 - y_i)\, log(1 - p_i)\,] + \lambda||\Theta||\text{^}2$$

In this expression, D is the set of training tuples, y_i is the label, p_i is the predicted probability p(d,t|c) for tuple i, λ is a regularization strength, Θ denotes all trainable parameters, and ||Θ||^2 is the squared parameter norm. The regularization term controls capacity, but it does not address sampling bias. Negative sampling must therefore be designed to avoid trivial negatives that are absent from the database only because they have not been studied.

3.5 Calibration and uncertainty

A probabilistic score is useful only if it is calibrated. Graph neural networks can produce overconfident predictions when trained on sparse positive-unlabeled data. A simple post hoc calibration applies temperature scaling to the logit:

$$p_cal = \sigma(\, logit(p) \,/\, T\,)$$

Here p is the raw predicted probability, logit(p) is its log-odds, T is a positive temperature parameter, and p_cal is the calibrated probability. Temperature scaling does not change the ranking of predictions, but it makes probability statements more interpretable and allows decision thresholds to be set according to acceptable false-positive rates. Calibration should be assessed on disease-level splits rather than on random edge splits, because random splits can leak information across diseases with overlapping modules.

4. Methodological Assessment and Technical Discussion

4.1 Baselines and evaluation logic

A credible evaluation must compare the proposed framework against models that isolate different sources of information. Suitable baselines include similarity-based drug-target interaction models [4,5], heterogeneous network integration methods [6], graph convolutional networks without disease gating [8], graph attention networks without edge provenance [9], and graph neural network models that use molecular graphs but not disease modules [15,16]. The comparison should be organized as an ablation hierarchy: removing disease-context gating tests whether module structure adds information; removing relation-specific weights tests whether typed edges matter; removing edge attributes tests whether evidence provenance matters; and removing calibration tests whether probability estimates are reliable.

Evaluation should not rely solely on area under the receiver operating characteristic curve. In imbalanced drug-target graphs, precision-recall curves and early-enrichment metrics are more informative because they reflect the practical task of ranking a small number of candidates. Moreover, random edge splits can overstate performance. A drug-target pair may share a target with a training pair, or a disease may share module members with another disease in the training set. Disease-level and scaffold-level splits are therefore necessary to estimate generalization to unseen diseases and unseen chemical series. The machine learning literature on drug-target interaction prediction has emphasized the importance of task-specific evaluation [17], and graph neural network reviews have highlighted the sensitivity of performance to neighborhood sampling and depth [18]. These considerations apply directly to disease-network prediction, where the relevant generalization unit is the disease context rather than the individual edge.

4.2 Sensitivity and failure modes

The framework has several identifiable failure modes. First, literature bias can create edges that reflect research attention rather than biological causality. A target that has been studied extensively may appear central in the graph even if its disease relevance is weak. Edge attributes and confidence weighting reduce this risk but do not eliminate it. Second, drug promiscuity can produce high scores for many targets because the drug embedding dominates the bilinear term. Regularization and target-specific biases help, but a more direct remedy is to evaluate ranking within a drug rather than across drugs. Third, disease-module definition is sensitive to annotation source and threshold. If modules are defined from protein interaction databases, they inherit the coverage and curation biases of those databases [20]. If they are defined from pathway databases, they may miss context-specific signaling relationships. The framework should therefore report results under multiple module definitions rather than a single curated set.

A fourth failure mode is over-smoothing. As message passing depth increases, node representations can become indistinguishable across neighborhoods, particularly in dense disease modules. This is especially problematic when a drug has many targets and the model

aggregates over a large neighborhood. Residual connections, relation-specific normalization, and depth-limited propagation can mitigate over-smoothing, but the appropriate depth is an empirical question. A fifth failure mode is leakage through disease-disease similarity edges. If two diseases share curated genes, a random split can place one disease in training and a related disease in testing, making the prediction task easier than it appears. Grouped splits and explicit reporting of overlap are necessary.

4.3 Mechanistic interpretation versus statistical association

The model output is a probability of a disease-relevant drug-target association, not a mechanism. A high score may reflect a shared pathway, a common chemical substructure, or a literature co-occurrence pattern. Mechanistic interpretation requires additional evidence: pathway enrichment, expression in relevant tissue, direction of effect, and experimental perturbation data. The framework can support interpretation by exposing attention weights and gate values, but these quantities should be treated as diagnostics rather than causal explanations. For example, a high attention weight from a drug to a target may indicate that the target resembles known drug targets, not that the drug physically binds it.

This distinction is important for translational claims. Drug-target prediction models are often evaluated by their ability to recover known interactions, but clinical relevance depends on target engagement, disease stage, tissue exposure, and safety. A model that ranks a target highly because it is central in a disease module may still fail if the target is not druggable or if modulation produces toxicity. The proposed framework should therefore be embedded in a broader evidence pipeline that includes chemical tractability, target safety, and experimental validation. The contribution of graph representation learning is to prioritize hypotheses, not to replace mechanistic or clinical assessment.

5. Implications for Network Medicine and Drug Repositioning

The framework has implications for how disease networks are used in drug discovery. First, it treats disease modules as contexts that shape target relevance, rather than as labels attached to drugs. This is closer to the biology of complex disease, where the same target can have different effects depending on the affected module. Second, it makes edge provenance explicit. Biomedical graphs are assembled from heterogeneous evidence sources, and ignoring provenance can produce models that learn database artifacts. Third, it supports drug repositioning by allowing a drug's known targets to be evaluated against a new disease module without requiring that the drug-disease pair have a direct edge in the graph. The model can thus generate hypotheses for diseases that are topologically related to known indications but not annotated as such.

There are practical constraints. Disease modules are incomplete, and many rare diseases have few curated genes. In such cases, the model may default to generic essentiality or pathway-level similarity. A useful design choice is to allow partial pooling: diseases with sparse modules borrow context from related diseases, but the gate is regularized so that borrowing does not overwhelm disease-specific evidence. Another constraint is computational. Typed message passing over a large biomedical graph requires efficient neighborhood sampling and mini-batch construction. The framework is compatible with inductive neighborhood sampling [11] and with relation-specific batching, but the engineering details affect both runtime and statistical validity. Finally, the model should be assessed for fairness across disease areas. If training data are concentrated in common diseases, predictions for rare diseases may be poorly calibrated even when overall discrimination appears acceptable.

6. Conclusion

This manuscript developed a graph representation learning framework for predicting drug-target associations in complex human disease networks. The framework represents drugs, targets, diseases, pathways, and functional contexts as a typed graph; propagates information through relation-specific message passing with edge-aware attention; constructs disease-context vectors from module nodes; and uses gating and bilinear scoring to produce disease-conditioned association probabilities. It also specifies a positive-unlabeled objective and a calibration step, and it articulates an evaluation logic based on disease-level splits, ablation of topological components, and separation of predictive discrimination from mechanistic plausibility.

The main argument is that disease-network structure should enter the representation learning problem directly, not as a post hoc filter. Disease modules determine which targets are contextually relevant, and typed edges determine which evidence should influence propagation. The framework's technical value lies in making these dependencies explicit and in identifying failure modes that arise from literature bias, drug promiscuity, module definition, over-smoothing, and leakage. These are not incidental implementation issues; they shape what a graph model can learn from biomedical data. The proposed design provides a coherent basis for hypothesis prioritization in network medicine, while remaining clear that association probabilities are evidence-ranked predictions rather than validated therapeutic mechanisms.

References

1. Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. Nature Reviews Genetics, 12(1), 56–68. https://doi.org/10.1038/nrg2918

2. Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., & Barabási, A.-L. (2007). The human disease network. Proceedings of the National Academy of Sciences, 104(21), 8685–8690. https://doi.org/10.1073/pnas.0701361104

3. Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., & Barabási, A.-L. (2015). Uncovering disease-disease relationships through the incomplete interactome. Science, 347(6224), 1257601. https://doi.org/10.1126/science.1257601

4. Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., & Kanehisa, M. (2008). Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics, 24(13), i232–i240. https://doi.org/10.1093/bioinformatics/btn162

5. Bleakley, K., & Yamanishi, Y. (2009). Supervised prediction of drug-target interactions using bipartite graphs and chemical/genomic spaces. Bioinformatics, 25(18), 2397–2403. https://doi.org/10.1093/bioinformatics/btp400

6. Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., Peng, J., Chen, L., & Zeng, J. (2017). A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nature Communications, 8, 573. https://doi.org/10.1038/s41467-017-00680-8

7. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Pon, A., Knox, C., & Wilson, M.

(2018). DrugBank 5.0: A major update to the DrugBank database for 2018. Nucleic Acids Research, 46(D1), D1074–D1082. https://doi.org/10.1093/nar/gkx1037

8. Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations.

9. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. International Conference on Learning Representations.

10. HerbSyner_Finder: a network community-based model for identifying synergistic combinations from herbal medicines and complex systems. Targetome. 2026;2(2):0-0. doi: 10.48130/targetome-0026-001

11. Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in Neural Information Processing Systems, 30, 1024–1034.

12. Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 855–864). https://doi.org/10.1145/2939672.2939754

13. Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics, 34(13), i457–i466. https://doi.org/10.1093/bioinformatics/bty294

14. Himmelstein, D. S., Lizee, A., Hessler, C., Brueggeman, L., Chen, S. L., Hadley, D., Green, A., Khankhanian, P., & Baranzini, S. E. (2017). Systematic integration of biomedical knowledge prioritizes drugs for repurposing. eLife, 6, e26726. https://doi.org/10.7554/eLife.26726

15. Tsubaki, M., Tomii, K., & Sese, J. (2019). Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. Bioinformatics, 35(2), 309–318. https://doi.org/10.1093/bioinformatics/bty535

16. Nguyen, T., Le, H., Quinn, T. P., Nguyen, T., Le, T. D., & Venkatesh, S. (2021). GraphDTA: Predicting drug-target binding affinity with graph neural networks. Bioinformatics, 37(8), 1140–1147. https://doi.org/10.1093/bioinformatics/btaa921

17. Chen, R., Liu, X., Jin, S., Lin, J., & Liu, J. (2018). Machine learning for drug-target interaction prediction. Molecules, 23(9), 2208. https://doi.org/10.3390/molecules23092208

18. Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. AI Open, 1, 57–81. https://doi.org/10.1016/j.aiopen.2021.01.001

19. Wang, Y., You, Z.-H., Yang, S., Yi, H.-C., Chen, Z.-H., & Zhang, K. (2020). A deep learning-based method for drug-target interaction prediction based on long short-term memory neural network. BMC Medical Informatics and Decision Making, 20(Suppl 2), 49. https://doi.org/10.1186/s12911-020-1052-0

20. Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2021). The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Research, 49(D1), D605–D612. https://doi.org/10.1093/nar/gkaa1074