Integrating Graph Neural Networks and Molecular Interaction Data for Discovering Novel Therapeutic Targets in Autoimmune Disorders

Lilla Totino
Department of Computer Science, George Mason University, Fairfax, VA, USA.
totinolilla@gmu.edu

Abstract

Autoimmune disorders present a difficult target-discovery problem because their molecular mechanisms are distributed across interacting proteins, pathways, and immune cell contexts rather than localized to a small number of isolated genes. Molecular interaction networks encode this distributed structure, but current computational approaches often treat target prioritization as a ranking problem over individual nodes, thereby underusing relational context, interaction type, and evidence reliability. This manuscript develops a methodological framework that integrates multi-relational molecular interaction data with graph neural networks to discover candidate therapeutic targets in autoimmune disorders. The framework represents genes, proteins, diseases, and drugs as a heterogeneous graph and uses relation-specific message passing with attention to learn context-dependent node embeddings. Heterogeneous evidence from protein interaction databases, pathway repositories, drug-target resources, and gene-set collections is fused through an edge-confidence function, allowing the model to distinguish high-quality from noisy or indirect interactions. Target discovery is formulated as probabilistic link prediction between molecular entities and disease phenotypes, trained with a discriminative objective and regularized to limit overfitting. Because candidate ranking is sensitive to model variance and incomplete interactome coverage, an uncertainty-aware ranking rule penalizes candidates with unstable ensemble predictions. The paper specifies the architectural components, objective functions, evidence-fusion logic, and evaluation conditions required for a reproducible target-discovery pipeline. The resulting analytical framework provides a disciplined route from heterogeneous molecular interaction data to experimentally testable autoimmune target hypotheses and clarifies where graph learning adds value beyond conventional network-proximity and differential-expression analyses.

Keywords

graph neural networks; molecular interaction networks; autoimmune disorders; therapeutic target discovery; uncertainty quantification; network medicine

1. Introduction

Autoimmune disorders are characterized by chronic immune activation against self-antigens and by substantial clinical and molecular heterogeneity across patients and tissues. Target-discovery efforts commonly begin with differential expression, genome-wide association signals, or candidate-gene curation, but these approaches often identify isolated genes whose biological significance depends on the surrounding interaction context. A gene that appears differentially expressed may be a downstream consequence rather than a causal mediator, and a receptor that is statistically associated with disease may be a poor therapeutic

target if it participates in broad essential pathways. Molecular interaction data provide a way to place candidate genes within a relational structure that includes physical binding, pathway co-membership, regulatory relationships, drug-target annotations, and disease-gene associations. The central computational challenge is therefore not merely to score genes but to learn which regions of a multi-relational interaction graph are most plausibly relevant to an autoimmune disease phenotype.

Graph representation learning offers a natural formalism for this problem. Graph convolutional networks [1], graph attention networks [2], and message-passing neural networks [3] have demonstrated that node representations can be learned from local neighborhood structure and node attributes. Graph-based models have also been used to predict polypharmacy side effects from molecular interaction networks [4], showing that relational learning can capture biologically meaningful dependencies beyond pairwise similarity. Network-community methods have been developed to identify synergistic combinations from herbal medicines and complex systems [5], and such work illustrates the value of community-level structure in complex molecular systems. However, community detection alone does not provide a supervised, uncertainty-aware ranking of individual therapeutic targets, nor does it naturally incorporate heterogeneous edge types and evidence quality.

Network medicine provides a complementary conceptual foundation. Disease phenotypes can be viewed as perturbations of molecular interaction networks [6], disease modules can be identified through network proximity in the incomplete interactome [7], and the human disease network links disorders through shared disease genes [8]. These studies establish that relational context matters, but they also expose limitations of proximity-based reasoning: interactions are incomplete, tissue-specific, and unevenly validated. For drug-target interaction prediction, deep learning models such as DeepDTA [9] and network integration approaches [10] have improved the ability to infer molecular associations from chemical and genomic spaces. Yet autoimmune target discovery requires a formulation that acknowledges disease heterogeneity, noisy interaction evidence, and the distinction between a statistically strong association and a therapeutically plausible intervention point.

This paper makes a methodological contribution rather than reporting experimental findings. It develops a graph neural network framework that integrates multi-relational molecular interaction data for autoimmune target discovery. The framework has three connected components. First, it represents genes, proteins, diseases, and drugs as a heterogeneous graph with typed edges and node features. Second, it uses relation-specific message passing with attention and an edge-confidence fusion function to learn representations that reflect both interaction type and evidence reliability. Third, it formulates target prioritization as a probabilistic link-prediction problem and ranks candidates with an uncertainty penalty derived from ensemble variance. The paper also specifies the evaluation logic required to test such a framework, including held-out disease-gene associations, external network validation, and prospective experimental triage. The argument is deliberately analytical: it derives design consequences from the structure of the biological problem and from established properties of graph learning, without asserting empirical performance that has not been established.

2. Problem Formulation and Biological Rationale

The input to the framework is a multi-relational molecular interaction graph. Nodes represent molecular entities, disease phenotypes, and drugs; edges represent typed relations

such as physical protein-protein interaction, pathway co-membership, co-expression, drug-target binding, and disease-gene association. Formally, the graph is defined as follows.

$$G = (V, E, R), E = \cup_\{r \in R\} E_r, X \in \mathbb{R}\^\{|V| \times d\}$$

Here, V is the set of nodes, R is the set of relation types, E_r is the set of edges of relation type r, and X is the node feature matrix with d features per node. The union over relation types indicates that the same node pair may be connected by more than one relation, which is common in molecular data; for example, two proteins may both physically interact and share a pathway. Node features include expression summaries, functional annotations, subcellular localization, and sequence-derived descriptors where available. Disease nodes are associated with phenotype identifiers, and drug nodes carry mechanism-of-action and target annotations.

The biological rationale for graph-based target discovery rests on the module structure of autoimmune disease. Disease-associated genes are not randomly distributed in the interactome; they tend to cluster in connected or proximally connected regions that correspond to immune signaling, cytokine regulation, antigen presentation, and tissue-specific inflammatory processes [6-8]. A candidate target is therefore plausible if it is embedded in a disease-relevant module, if it connects multiple evidence types, and if its perturbation is unlikely to disrupt unrelated essential functions. This reasoning motivates a link-prediction formulation: the model learns to distinguish disease-gene associations that are consistent with the graph context from those that are not. The output is a score for each candidate molecular entity with respect to a disease node, not a deterministic causal claim.

The problem also has important boundary conditions. The interactome is incomplete, and absent edges are not equivalent to absent interactions. Edge evidence is heterogeneous: a curated physical interaction differs from a text-mined association, and a pathway co-membership may reflect functional relatedness rather than direct binding. Autoimmune diseases are tissue-specific and cell-type-specific, whereas many interaction databases aggregate evidence across tissues. A target that is highly ranked in a generic interactome may be irrelevant in the relevant immune cell subset. The framework therefore treats edge confidence and tissue context as explicit modeling concerns rather than incidental preprocessing details.

3. Multi-Relational Graph Neural Network Architecture

The core representation learner is a relational graph neural network. Each node begins with its feature vector, and each layer updates the node representation by aggregating messages from neighbors under each relation type. The update rule is:

$$h_i\^(l+1) = \sigma(W_0\ h_i\^(l) + \sum_\{r \in R\} \sum_\{j \in N_i\^r\}\ c_\{i,r\}\^(- 1)\ W_r\ h_j\^(l))$$

In this expression, h_i^(l) is the representation of node i at layer l, W_0 is a self-transformation matrix, W_r is a relation-specific transformation matrix, N_i^r is the set of neighbors of i connected by relation r, c_{i,r} is a normalization constant such as the degree of i under relation r, and σ is a nonlinear activation function. The relation-specific weights are essential because physical binding, pathway co-membership, and co-expression carry different biological semantics. A single shared weight matrix would force the model to treat these relations as interchangeable, which would blur the distinction between direct molecular contact and statistical co-regulation.

Attention is used within each relation type to weight neighbors according to their relevance to the target node. For a node i and a neighbor j under relation r, the attention coefficient is:

$$\alpha_ij\text{^}r \; = \; softmax_j(\; LeakyReLU(a_r\text{^}T \; [W_r \; h_i \; || \; W_r \; h_j]) \;)$$

Here, α_ij^r is the normalized attention weight, a_r is a relation-specific attention vector, || denotes concatenation, and the softmax is taken over j∈N_i^r. Attention allows the model to emphasize neighbors that are more informative for the node's context. For example, in a cytokine signaling module, a receptor node may have many interaction partners, but only a subset is relevant to the disease phenotype. Fixed averaging would dilute that signal; attention can learn to concentrate on informative edges, provided that the training objective and regularization prevent it from overfitting to spurious correlations.

After L layers, the model produces final node embeddings h_i for molecular entities and u_d for disease nodes. Target discovery is treated as a link-prediction problem between a molecular node i and a disease node d. The score is computed with a bilinear form:

$$s_{i,d} = h_i^T M u_d + b_d$$

where M is a learnable bilinear interaction matrix and b_d is a disease-specific bias. The bilinear form is more expressive than a dot product because it can capture asymmetric and relation-specific compatibility between molecular and disease embeddings. The probability that a candidate target is associated with the disease is p_{i,d}=σ(s_{i,d}). The model is trained with a discriminative objective that treats observed disease-gene associations as positive examples and sampled non-associations as negative examples:

$$L \; = \; -\sum_\{(i,d) \in D\text{^}+\} \; log \; \sigma(s_\{i,d\}) \; - \; \sum_\{(i,d) \in D\text{^}-\} \; log(1 - \sigma(s_\{i,d\})) \; + \; \lambda||\theta||_2\text{^}2$$

In this loss, D^+ is the set of positive disease-gene pairs, D^- is the set of negative pairs, θ denotes all learnable parameters, and λ controls L2 regularization. Negative sampling must be designed carefully. Random non-associations may be trivially easy to distinguish, whereas hard negatives drawn from disease-adjacent but non-associated genes provide a more meaningful training signal. The loss does not by itself guarantee biologically valid rankings; it optimizes discrimination under the chosen sampling distribution. This is why the evaluation design must include held-out associations and external network evidence rather than only internal classification accuracy.

The architecture has several deliberate design choices. Two or three message-passing layers are typically sufficient for molecular interaction graphs because deeper propagation can over-smooth node representations and make local neighborhoods indistinguishable. Residual connections and dropout are used to preserve node-specific information and reduce overfitting. Relation-specific transformations increase parameter count, but this cost is justified by the need to distinguish biological relation types. The attention mechanism is constrained to operate within relation types rather than across all edges, which prevents a high-degree generic relation from overwhelming a sparse but specific relation.

4. Integration of Heterogeneous Molecular Interaction Evidence

The graph is constructed from multiple sources, each with different reliability and coverage. Drug-target and mechanism annotations can be drawn from DrugBank [11]. Protein-protein association data can be assembled from STRING [12], while curated physical and genetic interactions can be supplemented with BioGRID [13]. Pathway membership can

be represented using KEGG [14], and gene-set collections such as MSigDB [15] provide functional modules and hallmark signatures that are useful for constructing co-membership or co-expression edges. These sources are not interchangeable. A curated physical interaction from BioGRID is different from a text-mined association in STRING, and a KEGG pathway edge represents shared pathway membership rather than direct binding. The graph must preserve these distinctions as relation types or edge features.

Edge confidence is modeled explicitly because noisy interactions can dominate message passing. For an edge between nodes i and j, the confidence weight is:

$$w_{ij} = \sigma(\beta_0 + \sum_{s=1}^{S} \beta_s \varphi_s(x_{ij}^s))$$

Here, S is the number of evidence sources, x_{ij}^s is the feature value for source s, φ_s is a transformation or indicator function for that source, β_s are parameters that determine the contribution of each source, and β_0 is a bias term. The sigmoid maps the combined evidence to a bounded confidence weight. In a fully supervised setting, β can be learned from validated interactions. In a data-limited setting, β can be set from prior reliability estimates or calibrated through cross-validation on held-out edges. The important point is that confidence is not assumed to be uniform: an edge supported by multiple curated sources should have a different influence than an edge supported only by text mining.

The confidence weight is incorporated into message passing by multiplying the neighbor contribution by w_{ij}. This changes the aggregation so that low-confidence edges contribute less to the node representation. The same logic can be applied at the scoring stage, where a candidate's ranking is marginalized over edge reliability rather than treated as certain. This is especially important in autoimmune target discovery because many disease-associated genes are connected through indirect or context-dependent interactions. A model that treats all edges equally may learn a dense immune module that is biologically plausible but therapeutically non-specific. Confidence weighting encourages the model to rely on evidence that is more likely to reflect a real molecular relationship.

The integration also raises a representation issue. Different data sources may have different coverage across immune cell types. A static graph aggregates evidence across tissues, but autoimmune pathology is often tissue-specific. The framework therefore allows node features to include tissue-specific expression summaries and allows edge features to include context indicators where such data are available. When tissue-specific data are missing, the model should not infer context-specificity from the absence of evidence. This distinction between missing evidence and negative evidence is a fundamental boundary condition for graph-based target discovery.

5. Target Prioritization and Uncertainty-Aware Ranking

The output of the graph neural network is a score for each candidate target-disease pair. A simple ranking by p_{i,d} is inadequate because scores from a single model can be unstable, especially when the graph is sparse or when the candidate lies near a decision boundary. Graph neural networks are also sensitive to random initialization, neighborhood sampling, and changes in edge composition. An ensemble of models trained with different initializations or bootstrapped edge sets provides a distribution of scores for each candidate. The uncertainty-aware ranking rule is:

$$R_i = \mu_i - \gamma \sqrt{((1/M) \sum_{m=1}^{M} (s_i^{(m)} - \mu_i)^2)}$$

In this expression, $s_i^{(m)}$ is the score assigned to candidate i by ensemble member m, M is the number of ensemble members, μ_i is the mean score across members, and γ is a nonnegative risk-aversion parameter. The square-root term is the standard deviation of the ensemble scores, and subtracting it penalizes candidates whose predictions are unstable. When $\gamma=0$, the rule reduces to ranking by mean score. When γ is large, the rule favors candidates with lower variance even if their mean score is slightly lower. This is appropriate for experimental target triage because a small set of stable, mechanistically coherent candidates is more useful than a long list of unstable predictions.

The ranking rule must be coupled with an evaluation protocol that distinguishes discrimination from calibration. Discrimination asks whether true disease-gene associations receive higher scores than non-associations; calibration asks whether predicted probabilities correspond to observed frequencies. In target discovery, calibration is often more important than raw discrimination because experimental follow-up is expensive. A model that ranks well but is overconfident in the wrong probability range can mislead prioritization. The evaluation logic should therefore include held-out disease-gene associations, leave-one-disease-out testing, external validation on independent interaction networks, and comparison against network-proximity and matrix-factorization baselines. Graph neural network reviews [16] provide a useful guide to architectural choices, but they also emphasize that performance depends on graph construction and evaluation design.

Prior work on compound-protein interaction prediction offers additional lessons. End-to-end graph-and-sequence models have shown the value of learning joint representations [17], while studies of realistic drug-target interaction prediction have highlighted the gap between random splits and time-split or scaffold-split evaluation [18]. Bipartite local models [19] and integrated chemical-genomic space models [20] demonstrate that relational information improves prediction when the graph is constructed carefully. For autoimmune target discovery, the analogous requirement is that evaluation splits must respect disease biology. Randomly splitting disease-gene associations can leak information because genes associated with one autoimmune disease may be associated with another. A leave-one-disease-out split is more demanding and more clinically relevant, but it is also harder because the model must generalize to a disease node with few or no training associations.

The framework should also produce candidate-level explanations. Attention weights and edge-confidence contributions can be aggregated to show which interaction types and evidence sources drove a candidate's score. This does not prove causality, but it allows experimentalists to assess whether the model's reasoning is biologically plausible. A candidate that scores highly because of a single curated physical interaction is easier to interpret than a candidate that scores highly because of many weak, text-mined associations. Explanation is therefore not a cosmetic addition; it is a quality-control mechanism for target triage.

6. Technical Discussion and Boundary Conditions

The proposed framework is not a substitute for experimental validation, and its outputs should be interpreted as prioritized hypotheses. Several failure modes are predictable. First, incomplete interactome coverage can cause true associations to appear as non-associations. This is not merely a data problem; it affects the training objective because negative sampling may select genes that are actually associated but not yet annotated. Second, edge reliability is heterogeneous, and confidence weights learned from validated interactions may not transfer to autoimmune-specific interactions. Third, graph neural networks can over-smooth

representations in dense modules, making many nodes in an immune cluster appear similar. Relation-specific message passing and residual connections mitigate this risk but do not eliminate it. Fourth, tissue and cell-type specificity are often lost in aggregated interaction networks. A target that is central in a generic protein interaction network may be expressed at low levels in the relevant immune compartment.

Alternative approaches have complementary strengths. Network proximity methods are interpretable and require less training data, but they depend strongly on network construction and do not naturally incorporate multiple relation types. Matrix factorization methods scale well and capture global associations, but they may miss local neighborhood structure. Classical machine learning on node features can be effective when the graph is unreliable, but it discards relational information. The graph neural network approach is most appropriate when the graph is informative, when edge types matter, and when the goal is to learn context-dependent representations rather than hand-engineered proximity scores. The choice of relation-specific attention is justified when multiple interaction types are present; if only a single curated relation is available, a simpler graph convolutional network may be preferable.

The framework also has implications for study design. Prospective validation should compare model-ranked candidates against a matched control set of genes that are disease-adjacent but not top-ranked. Assays should be chosen that test the proposed mechanism, not merely expression changes. For autoimmune disorders, relevant assays may include cytokine production, immune cell activation, receptor-ligand binding, or tissue-specific organoid models. The model's uncertainty estimate can be used to select candidates for high-throughput screening versus lower-throughput functional validation. This is a practical advantage of uncertainty-aware ranking: it aligns computational triage with experimental resource allocation.

7. Conclusion

This manuscript has developed a methodological framework for integrating graph neural networks and molecular interaction data for therapeutic target discovery in autoimmune disorders. The framework represents genes, proteins, diseases, and drugs as a multi-relational graph, learns node embeddings with relation-specific message passing and attention, fuses heterogeneous evidence through an edge-confidence function, and ranks candidate targets with an uncertainty-penalized ensemble score. The approach addresses a central limitation of current target-discovery pipelines: the tendency to score individual genes without accounting for relational context, interaction type, and evidence reliability. By formulating target discovery as probabilistic link prediction and by making uncertainty explicit, the framework produces candidate rankings that are more suitable for experimental triage than single-score rankings. The paper also specifies the evaluation conditions required to test the framework, including held-out disease associations, leave-one-disease-out validation, external network validation, and prospective experimental assessment. The main contribution is analytical and design-oriented: it clarifies how multi-relational graph learning can be structured for autoimmune target discovery, where its assumptions are most defensible, and which failure modes must be addressed before computational predictions can be interpreted as therapeutic hypotheses. Future work should focus on tissue-specific graph construction, calibration of edge confidence across data sources, and prospective validation of uncertainty-aware rankings in immune cell models.

References

1. Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

2. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. arXiv preprint arXiv:1710.10903.

3. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural message passing for quantum chemistry. In Proceedings of the 34th International Conference on Machine Learning (pp. 1263-1272). PMLR.

4. Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics, 34(13), i457-i466.

5. 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

6. Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. Nature Reviews Genetics, 12(1), 56-68.

7. 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.

8. 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.

9. Öztürk, H., Özgür, A., & Ozkirimli, E. (2018). DeepDTA: deep drug-target binding affinity prediction. Bioinformatics, 34(17), i821-i829.

10. Luo, Y., Zhao, X., Zhou, J., Yang, J., Zhang, Y., Kuang, W., ... & Zeng, J. (2017). A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. Nature Communications, 8(1), 573.

11. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., ... & Assempour, N. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Research, 46(D1), D1074-D1082.

12. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., ... & von Mering, C. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research, 47(D1), D607-D613.

13. Oughtred, R., Rust, J., Chang, C., Breitkreutz, B. J., Stark, C., Willems, A., ... & Tyers, M. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Science, 30(1), 187-200.

14. Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research, 28(1), 27-30.

15. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Systems, 1(6), 417-425.

16. Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., ... & Sun, M. (2020). Graph neural networks: A review of methods and applications. AI Open, 1, 57-81.

17. 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.

18. Pahikkala, T., Airola, A., Pietilä, S., Shakyawar, S., Szwajda, A., Tang, J., & Aittokallio, T. (2015). Toward more realistic drug-target interaction predictions. Briefings in Bioinformatics, 16(2), 325-337.

19. Bleakley, K., & Yamanishi, Y. (2009). Supervised prediction of drug-target interactions using bipartite local models. Bioinformatics, 25(18), 2397-2403.

20. 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.