OPEN ACCESS Freely available online

ISSN: 2153-0602
Journal of
Data Mining in
Genomics & Proteomics

Perspective

Predictive Modeling of MicroRNA mRNA Interactions Using Graph Neural Networks

Michael Anderson*

Department of Systems Biology, Harvard Medical School, Boston, USA

DESCRIPTION

MicroRNAs (miRNAs) are small, non-coding RNA molecules that regulate gene expression post-transcriptionally by binding to complementary sequences in messenger RNAs (mRNAs), leading to mRNA degradation or translation inhibition. This regulatory mechanism plays a crucial role in numerous biological processes such as development, cell differentiation, apoptosis, and immune response. Dysregulation of miRNA mRNA interactions has been implicated in various diseases, particularly cancers, cardiovascular conditions, and neurological disorders. Therefore, accurately identifying miRNA mRNA interactions is essential for understanding gene regulatory networks and uncovering disease mechanisms. Traditional experimental methods like luciferase reporter assays, CLIP-Seq, and microarray-based platforms are labor-intensive, time-consuming, and often limited in scope. To address these limitations, computational prediction methods have gained traction, particularly those utilizing advanced machine learning and deep learning models. Among them, Graph Neural Networks (GNNs) have recently emerged as a powerful approach for modeling biological interactions due to their unique ability to learn from graph-structured data, making them highly suitable for predicting miRNA mRNA interactions. In essence, miRNA mRNA interaction networks naturally form bipartite graphs, where nodes represent miRNAs and mRNAs, and edges denote potential or validated regulatory interactions. Unlike traditional models that treat features independently, GNNs are designed to exploit the topology of such graphs, learning feature representations not only from node attributes (like sequence data, secondary structures, or expression levels) but also from the connections among nodes. This capacity enables GNNs to capture both local and global interaction patterns within the network, resulting in more accurate and biologically relevant predictions.

To apply GNNs in this domain, researchers typically begin by constructing an interaction graph based on known miRNA mRNA pairs obtained from databases such as miRTarBase, TargetScan, or miRDB. Each miRNA and mRNA node is annotated with features that may include nucleotide sequences, thermodynamic binding scores, expression profiles, or secondary structure predictions. This graph is then used to train a GNN model, such as a Graph Convolutional Network (GCN), Graph Attention Network (GAT), or GraphSAGE, where the model learns low-dimensional embeddings for each node that preserve both their intrinsic features and structural context. The embeddings are subsequently used to predict whether a given miRNA mRNA pair is likely to interact. One of the significant advantages of using GNNs for this task is their inductive learning capability, allowing the model to generalize to unseen data. This is particularly useful when predicting interactions involving novel miRNAs or mRNAs not present in the training dataset. Moreover, GNNs can integrate heterogeneous information from multiple sources, making them ideal for multi-modal biological data. For instance, integrating expression data from RNA-Seq experiments, sequence conservation scores, or epigenetic markers into the graph enables the model to leverage a more comprehensive biological context.

Another compelling aspect of GNNs is their capacity to identify indirect or higher-order relationships between miRNAs and mRNAs. In biological systems, regulation is rarely linear; a miRNA may indirectly affect an mRNA through intermediate regulatory nodes. GNNs can propagate information across several layers of neighbors in the graph, thereby capturing such complex dependencies that traditional machine learning methods often overlook. This deep propagation feature is critical for understanding gene regulation at a systems level. While GNNs have demonstrated remarkable success in miRNA mRNA interaction prediction, challenges remain. One is the interpretability of deep learning models. Biologists often require not just predictions but also explanations of why a particular interaction is predicted. Recent advances in explainable AI (XAI) for graph models, such as GNNExplainer or GraphLIME, are helping to address this by identifying important substructures or node features that drive predictions. By highlighting the nucleotide motifs or network pathways most influential in a

Correspondence to: Michael Anderson, Department of Systems Biology, Harvard Medical School, Boston, USA, Email: manderson@harvardsys.us

Received: 24-Feb-2025, Manuscript No. JDMGP-25-29291; Editor assigned: 26-Feb-2025, Pre QC No. JDMGP-25-29291 (PQ); Reviewed: 12-Mar-2025, QC No. JDMGP-25-29291; Revised: 18-Mar-2025, Manuscript No. JDMGP-25-29291 (R); Published: 26-Mar-2025, DOI: 10.35248/2153-0602.25.16.375

Citation: Anderson M (2025). Predictive Modeling of MicroRNA mRNA Interactions Using Graph Neural Networks. J Data Mining Genomics Proteomics.16: 375.

given prediction, these tools bridge the gap between computational inference and biological insight.

Another challenge is the imbalance in biological datasets, where the number of true miRNA mRNA interactions is significantly lower than the number of non-interactions. This class imbalance can bias learning algorithms. To mitigate this, techniques such as synthetic oversampling, cost-sensitive learning, or focal loss functions are applied during model training. Additionally, incorporating negative samples (i.e., randomly paired miRNA mRNA combinations assumed not to interact) must be done carefully to avoid introducing noise or bias. From a practical standpoint, GNN-based predictive modeling holds immense promise in biomedical applications. For example, identifying miRNA mRNA interactions specific to a cancer subtype could uncover novel biomarkers or therapeutic targets. In personalized medicine, patient-specific expression profiles can be integrated into the graph to predict individualized regulatory interactions, facilitating precision diagnostics or treatment strategies. Furthermore, such models can help identify cross-species regulatory conservation, aiding in translational research from model organisms to humans.