

---

# A NOVEL GRAPH TRANSFORMER FRAMEWORK FOR GENE REGULATORY NETWORK INFERENCE

---

**Binon Teji**

Network Reconstruction & Analysis (NETRA) Lab ,  
Department of Computer Applications,  
Sikkim University, Sikkim, India, 737102  
bteji.20pdca01@sikkimuniversity.ac.in

**Swarup Roy**

Network Reconstruction & Analysis (NETRA) Lab,  
Department of Computer Applications,  
Sikkim University, Sikkim, India, 737102  
sroy01@cus.ac.in

## ABSTRACT

The inference of gene regulatory networks (GRNs) is a foundational stride towards deciphering the fundamentals of complex biological systems. Inferring a possible regulatory link between two genes can be formulated as a link prediction problem. Inference of GRNs via gene coexpression profiling data may not always reflect true biological interactions, as its susceptibility to noise and misrepresenting true biological regulatory relationships. Most GRN inference methods face several challenges in the network reconstruction phase. Therefore, it is important to encode gene expression values, leverage the prior knowledge gained from the available inferred network structures and positional informations of the input network nodes towards inferring a better and more confident GRN network reconstruction. In this paper, we explore the integration of multiple inferred networks to enhance the inference of Gene Regulatory Networks (GRNs). Primarily, we employ autoencoder-based embeddings to capture gene expression patterns directly from raw data, preserving intricate biological signals. Then, we embed the prior knowledge from GRN structures transforming them into a text-like representation using random walks, which are then encoded with a masked language model, BERT, to generate global embeddings for each gene across all networks. Additionally, we embed the positional encodings of the input gene networks to better identify the position of each unique gene within the graph. These embeddings are integrated into graph transformer-based model, termed GT-GRN, for GRN inference. The GT-GRN model effectively utilizes the topological structure of the ground truth network while incorporating the enriched encoded information. Experimental results demonstrate that GT-GRN significantly outperforms existing GRN inference methods, achieving superior accuracy and highlighting the robustness of our approach.

## 1 Introduction

The key aspect of Systems Biology seeks to understand the bigger picture of the complex biological systems that focuses to extract the relevant biological information of an organism at the cellular level. Biological components, such as genes, interact with each other to reconstruct gene regulatory networks (GRNs) from observational gene expression data [1]. This process unveils a complex web of interactions, shedding light on the underlying patterns that govern gene regulation. Inference of GRNs from gene expression data is crucial for understanding the multifactorial molecular interaction patterns among genes. A gene network consists of interlinked genes, where the expression of each gene influences the activity of other genes in the network [2]. An effective approach to describing gene regulatory networks (GRNs) involves the use of graphical and mathematical modeling, often grounded in graph-theoretic formalism to capture the complex interactions between genes. Formally, a GRN is represented as a network of nodes and edges, where the nodes represent genes, and edges represent the regulatory interactions between them [3]. GRN inference involves predicting or computationally guessing the connections among macromolecules by analyzing their relative expression patterns.

Technologies such as DNA Microarray [4], Single-Cell RNA Sequencing (scRNA-seq) [5], and Single-Nucleus RNA Sequencing (snRNA-seq) [6] have revolutionized transcriptomics by offering diverse and detailed insights into gene expression. Although each of these technologies has its unique strengths, they also come with certain limitations.

Common limitations include noisy gene expression data, which often complicates the inference process. Furthermore, the dynamic and non-linear nature of gene-gene regulatory interactions present a significant challenge, as traditional or linear methods often fail to capture the complex relationships comprehensively. In addition, gene regulatory networks tend to be sparse, further reducing the overall accuracy of inference methods. In the case of single-cell technologies, significant dropout events introduce a large number of zero counts in the expression matrix, adding another layer of complexity [7]. These challenges highlight the need for more robust and sophisticated approaches to effectively analyze and interpret gene expression data.

A plethora of supervised and unsupervised GRN inference methods have been developed to uncover intricate relationships within gene networks [8, 9]. Early efforts focused on simpler techniques, such as correlation analysis, mutual information-based methods, differential equations, and more. Significant effort has been devoted to understand the intricate relationships within gene networks [10, 11, 12]. However, many methods come with inherent limitations. Therefore, a more reliable GRN inference technique may prove more effective in this case. Numerous intelligent computational techniques have been developed to address this issue.

Over the years, significant progress has been made in developing approaches that incorporate machine learning techniques to infer the network from gene expression [13]. A recent trend in network-science has gained significant momentum in modeling graph-based applications powered by graph neural networks (GNNs). For example, Wang et al., propose GRGNN[14] method to reconstruct GRNs from gene expression data in a supervised and semi-supervised framework. The problem is formulated as a graph classification problem for GRN inference on DREAM5 benchmarks. Q-GAT[15] proposes quadratic neurons based on dual attention mechanism for GRN inference. The model is validated by introducing adversarial perturbations to the gene expression data on *E. coli* and *S. cerevisiae* datasets. Huang et al.,[16] propose a GNN based model called MIGGRI for GRN inference using spatial expression images that capture gene regulation from multiple images.

Another line of work also concentrates on Transformer-based architectures for GRN inference. TRENDY [17] leverages transformer models to construct a pseudo-covariance matrix as part of the WENDY [18] framework. Rather than generating gene regulatory networks (GRNs) from scratch, it enhances existing inferred GRNs. However, it does not incorporate additional structural side information into the inference process. STGRNS [19] is an interpretable transformer-based method for inferring GRNs from scRNA-seq data. This method only considers two genes for gene regulation excluding the possibility of indirect regulation for prediction. GRN-Transformer [20] utilizes multiple statistical features extracted from scRNA-seq data and uses inferred GRN extracted from a single inference algorithm PIDC[21].

Despite the advancements and superior performance of several methods, notable limitations persist. A primary concern is that many approaches rely solely on a single source of information, that is, gene expression data for inferring gene regulatory networks (GRNs), which is often insufficient for accurate and conclusive network prediction. Furthermore, some methods fail to utilize existing knowledge from previously inferred GRNs, limiting their potential to build upon existing insights. Moreover, several techniques overlook the integration of topological information, which is crucial for capturing the structural properties essential for robust GRN inference.

Rather than focusing solely on each aforementioned issue, our approach adopts an integrated perspective—driven by the intuition that combining multiple complementary sources of information, beyond gene expression alone, can enhance the quality of GRN inference. We propose *GT-GRN*, a novel approach that integrates the strengths of both unsupervised inference methods and supervised learning frameworks. Our method combines outcomes from the available inference techniques to minimize method-specific biases, ultimately deriving a more realistic and biologically meaningful GRN. To integrate multiple networks, rather than relying on GNN-based methods, which often suffer from over-smoothing when stacking multiple layers. We adopt a state-of-the-art NLP-based unsupervised approach that effectively captures and integrates information across networks. *GT-GRN* leverages the latest advancements in Graph Transformer models to enhance GRN inference. *GT-GRN* integrates three distinct representations derived from input expression networks: (1) topological features, which capture the structural properties of the network; (2) gene expression values, which are crucial for identifying gene interactions; and (3) the positional importance of genes, which reflects their functional relevance within the network. By fusing the multi-modal embeddings from diverse perspectives, our framework improves both the interpretability and predictive power of inferred GRNs, making it a robust solution for various biological applications. *GT-GRN* is advantageous in multiple sense.

1) *Multi-Networks Integration*: A key challenge in supervised Gene Regulatory Network (GRN) inference is the absence of ground-truth networks. True GRNs are often incomplete or unavailable, so we must rely on inferred networks as proxies. However, using a single inferred network can introduce bias or overlook critical interactions. We incorporate multiple networks inferred by different methods, harnessing their complementary strengths. While various inference models exist, each with its own set of advantages and limitations, combining these diverse sources allows us to leverage

their shared strengths. This approach helps mitigate methodological bias, ultimately enhancing the confidence and accuracy of our GRN predictions.

2) *Gene Expression Embedding*: Capturing meaningful representations of gene expression data through advanced embedding techniques can provide a richer understanding of the underlying regulatory mechanisms and improve GRN inference.

3) *Graph Transformer Frameworks*: Traditional GNNs rely on local message-passing mechanisms to infer graph structures. However, adopting graph transformer-based frameworks offers a more effective encoding strategy by leveraging global attention mechanisms, enabling better capture of complex regulatory relationships in GRNs.

The contributions of our present work are listed below:

- We capture the quantitative characteristics of gene expression profiles through an autoencoder that learns biologically meaningful latent representations, effectively summarizing complex gene activity patterns while preserving essential regulatory signals. (**Section 3.1**).
- We introduce a method to consolidate prior knowledge from multiple inferred GRNs by converting networks into text-like sequences, enabling a BERT-based masked language model to learn global gene embeddings that integrate structural information across all networks. (**Section 3.2**).
- We propose a novel framework, *GT-GRN*, which leverages attention mechanisms within a Graph Transformer model to learn rich gene embeddings by integrating multi-source data—including gene expression profiles, structured inferred knowledge, and graph positional encodings from the input graph. These unified gene embeddings effectively capture the underlying biological relationships between genes, facilitating enhanced Gene Regulatory Network (GRN) inference (**Section 3.3**).
- We demonstrate that *GT-GRN* effectively advances cell-type-specific gene regulatory network (GRN) reconstruction. Moreover, the superior quality of the learned embeddings enables their successful application to cell type annotation tasks, highlighting the model’s robustness and generalizability.

## 2 Related Work

With decades of effort within the research community dedicated to deciphering gene regulatory relationships from gene expression data, numerous methods have been proposed for reconstructing GRNs[22]. Traditional approaches include regression-based techniques [23] and mutual information-based methods such as ARACNE [24], MRNET [25], and CLR [26], which assess statistical dependencies between genes to infer potential regulatory interactions. Supervised machine learning methods have also been explored for GRN inference. Support Vector Machines (SVMs) have been utilized to reconstruct biological networks through local modeling approaches [27]. Extensions such as CompareSVM [28] and GRADIS [29] further leverage classification-based frameworks to enhance network prediction accuracy.

With the advent of deep learning, more powerful and data-driven models have emerged. For instance, Daoudi et al. [30] proposed a deep neural network (DNN) model to infer GRNs from multifactorial experimental data. Turki et al. [31] integrated both supervised and unsupervised learning techniques to perform link prediction on time-series gene expression data. Mao et al. [32] introduced a 3D convolutional neural network (CNN) model utilizing single-cell transcriptomic data, employing a novel labeling trick to enhance performance. GNE [33], a graph-based deep learning framework, unified known gene interactions and expression profiles to robustly infer GRNs in a scalable manner. Other works, Teji et al.,[34, 35] uses synthetic data using to evaluate various embedding models to evaluate for GRN inference in link prediction setup.

Recent approaches have increasingly explored the utility of graph neural networks (GNNs) to capture the structural and relational complexity of biological systems. DeepRIG [36] emphasizes learning global regulatory structures by embedding entire graphs using a graph autoencoder, thereby capturing comprehensive latent representations. GMFGRN [37] applies graph convolutional networks (GCNs) to factorize single-cell RNA-seq data into gene and cell embeddings, which are then used in a multilayer perceptron (MLP) for interaction prediction. GNNLink [38] frames GRN inference as a link prediction problem by employing a GCN-based interaction graph encoder to capture and infer potential regulatory dependencies between genes.

Despite these advancements, many GRN inference methods continue to rely on a single source of data or even topological information, often emphasizing local or pairwise gene interactions. This narrow focus can limit the depth and breadth of biological insights. Although deep learning approaches—such as MLPs, CNNs, and GNNs—have enhanced inference performance, they frequently process each data modality in isolation, missing opportunities for deeper integration.

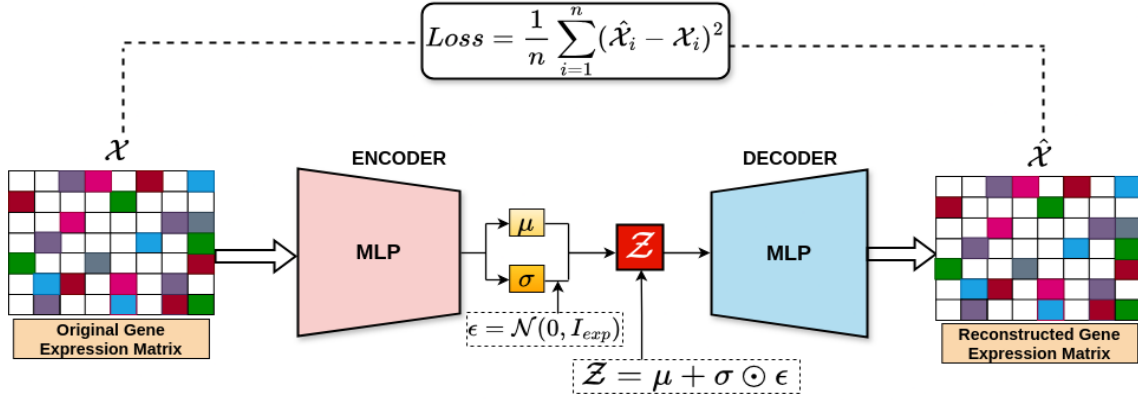


Figure 1: Gene Expression Embeddings via Variational Autoencoder (VAE)

Present research is a step forward that proposes a novel graph transformer framework integrated with multiple sources of biological information for GRN inference. Unlike conventional methods that rely on convolution-based architectures, our approach utilizes graph-based attention mechanisms to effectively model complex regulatory relationships. A key strength of this framework lies in its ability to fuse diverse embeddings derived from gene expression data, input graph structures, and both existing and previously inferred regulatory networks. By jointly leveraging these complementary sources within a unified, the model enables more accurate and biologically meaningful inference of gene regulatory networks.

### 3 Materials and Methodology

This section outlines the methodology of the proposed *GT-GRN* framework for gene regulatory network (GRN) inference, followed by a description of the datasets used for evaluation. The framework is composed of three key modules: (1) encoding gene expression profiles as embedding features using unsupervised deep learning; (2) extracting global gene embeddings through multi-network integration; and (3) capturing graph positional encodings from the input network structure. These complementary representations—gene expression embeddings, prior gene representations, and graph positional encodings—are fused to enhance GRN interaction prediction within a graph transformer model. The effectiveness of our approach is then evaluated using publicly available gene expression datasets.

#### 3.1 Gene Expression Feature Encoding

Understanding the dynamics of gene expression data involves unraveling the patterns and relationships embedded within large-scale datasets. As profiling technologies rapidly advance, gene expression data has become increasingly complex and high-dimensional. In such settings, traditional linear models often fall short in capturing the intricate patterns within the data. Unsupervised deep learning methods hold significant potential for uncovering meaningful biological signals and patterns from gene expression data. By leveraging such methods to non-linearly encode expression profiles, we can extract richer and more informative representations that capture the underlying biological complexity. Here, we employ variational autoencoder (VAE) [39] to encode gene expression data into compact, meaningful feature representations. Figure 1 illustrates the overview VAEs for encoding gene expressions. Variational Autoencoders (VAEs) are a probabilistic deep generative class of neural networks designed to reconstruct input data by learning a compressed, low-dimensional representation that effectively characterizes the input. Using VAE for gene expression encoding, we can efficiently capture complex expression dynamics and generate compact feature representations for further analysis and downstream tasks. VAEs are a powerful framework for unsupervised learning and generally comprise two interconnected components: an *encoder* and a *decoder*.

- **Encoder** : maps the input gene expression matrix  $\mathcal{X}$  to a latent representation space  $\mathcal{Z}$ . It approximates the posterior distribution  $\alpha(\mathcal{Z}|\mathcal{X})$  using a neural network. The encoder outputs the parameters, i.e., the mean and variance of a multivariate Gaussian distribution  $\beta(\mathcal{Z}|\mathcal{X})$ , that serves as an approximation of the true posterior  $\alpha(\mathcal{Z}|\mathcal{X})$ . This process captures the underlying biological variability and regulatory patterns among genes.

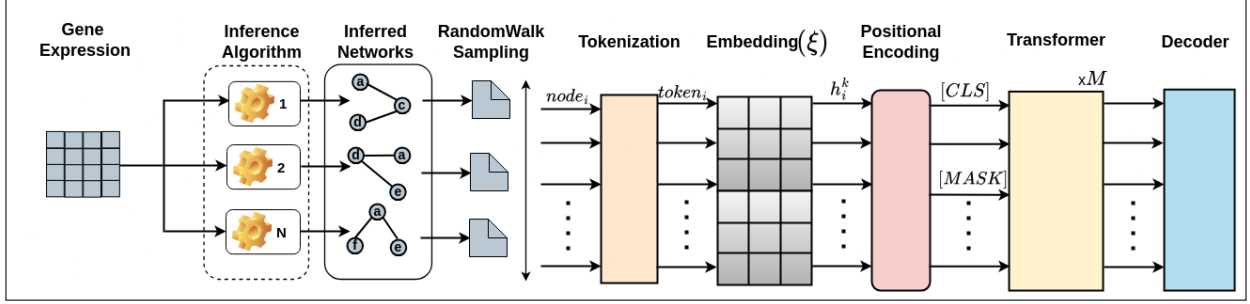


Figure 2: **Global Gene Embeddings via Multi-Network Integration.** Gene expression data is processed through inference algorithms to generate networks. Every network is sampled via random walks to produce node sequences. Each sequence starts with a special [CLS] token. These sequences are tokenized and embedded, then passed to a transformer model resulting in global embedding of each node or gene considering the context of all input networks in a joint learning setting. The model is trained by masking nodes in sequences and predicting them. Final node/gene embeddings are extracted from the embedding layer.

- **Decoder** : takes a sample  $\mathcal{Z}$  from the latent space and maps it back to the original gene expression space, generating a reconstructed matrix  $\hat{\mathcal{X}}$ . This is modeled by the likelihood  $\alpha(\mathcal{X}|\mathcal{Z})$ , which represents the probability of generating the observed gene expression profiles  $\mathcal{X}$  given the latent variables  $\mathcal{Z}$ . The decoder, implemented via a neural network, learns to reconstruct biologically plausible gene expression patterns from the learned latent representations.

During training, the VAE aims to learn the parameters of the encoder and the decoder network parameters by maximizing the Evidence Lower Bound (ELBO) which is given by :

$$\text{ELBO} = \mathbb{E}_{\beta(\mathcal{Z}|\mathcal{X})}[\log \alpha(\mathcal{X}|\mathcal{Z})] - \text{KL}[\beta(\mathcal{Z}|\mathcal{X})||\alpha(\mathcal{Z})] \quad (1)$$

where,  $\mathbb{E}_{\beta(\mathcal{Z}|\mathcal{X})}[\log \alpha(\mathcal{X}|\mathcal{Z})]$  is the reconstruction term that reconstructs the input gene expression data  $\mathcal{X}$  given the latent representation  $\mathcal{Z}$ .  $\text{KL}[\beta(\mathcal{Z}|\mathcal{X})||\alpha(\mathcal{Z})]$  is the KL (Kullback-Leibler) divergence that quantifies the distance between the approximate posterior  $\beta(\mathcal{Z}|\mathcal{X})$  and the prior distribution  $\alpha(\mathcal{Z})$ .

Since sampling from learned distributions is inherently non-differentiable, it hinders the use of gradient-based optimization during backpropagation. To address this, the reparameterization trick introduces a differentiable transformation by expressing the random variable as  $\mathcal{Z} = \tau(\mathcal{X}, \epsilon)$ , where,  $\tau$  is a deterministic function and  $\epsilon$  is an auxiliary noise variable drawn from a fixed, independent distribution. The above problem can be re-written as:

$$\mathcal{Z} \sim \beta(\mathcal{Z}|\mathcal{X}^{(i)}) = \mathcal{N}(\mathcal{Z}; \mu^{(i)}, \sigma^{2(i)} I_{exp})$$

$$\mathcal{Z} = \mu + \sigma \odot \epsilon; \text{reparameterization trick}$$

Where,  $\epsilon \sim \mathcal{N}(0, I_{exp})$ ,  $\odot$  is the element-wise product and  $I_{exp}$  is the identity matrix, which serves as the covariance matrix.

### 3.2 Global Embeddings via Multi-Network Integration of the inferred GRNs

We integrate multi-network information to understand gene interactions as prior knowledge. Integrating data from diverse inference methods, each with unique strengths and limitations, provides a holistic and reliable view of the network. This approach overcomes the shortcomings of relying on a single method, enabling robust downstream analysis. Figure 2 illustrates the workflow.

#### 3.2.1 Unsupervised Network Integration via Random Walks and Transformers

We present an unsupervised learning approach for integrating multiple networks to generate global embeddings. Let a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  represent an inferred network, where  $\mathcal{V}$  denotes the set of nodes and  $\mathcal{E}$  represents the set of edges. The graph is characterized by its adjacency matrix  $\mathcal{A} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ , where each entry  $\mathcal{A}_{ij}$  reflects the relationship between

nodes  $v_i$  and  $v_j$ . Specifically,  $\mathcal{A}_{ij} = 1$  if an edge  $e_{v_i, v_j} \in \mathcal{E}$  connects  $v_i$  and  $v_j$ , and  $\mathcal{A}_{ij} = 0$  otherwise, indicating no direct connection.

We consider a collection of  $C$  networks,  $\{G_1, G_2, \dots, G_C\}$ , sharing the same set of  $n$  nodes but differing in the number of edges in each network. We capture the structural information of the networks by converting them into text-like sequences using random walks, similar to node2vec [40]. The walks are encoded through an embedding matrix  $\xi \in \mathbb{R}^{n \times d_n}$ , where  $n$  is the size of the vocabulary (total nodes across all networks) and  $p$  is the desired embedding dimension. Positional encodings are used to account for node order as described in [41]:

$$PE_{seq}(pos, i) = \begin{cases} \sin\left(\frac{pos}{10000^{i/p}}\right), & \text{if } i \bmod 2 = 0, \\ \cos\left(\frac{pos}{10000^{(i-1)/p}}\right), & \text{otherwise.} \end{cases} \quad (2)$$

Here,  $PE_{seq}(pos, i)$  represents the  $i$ -th coordinate of the position encoding at sequence position  $pos$ . These encodings are concatenated with the original input features or the embedding matrix.

The embedding matrix and the decoding layer are initialized with uniform random values, while the transformer layer is initialized using Xavier’s initialization [42]. During training, all parameters are updated. Each sequence begins with a special classification token  $[CLS]$ , while other tokens correspond to node-specific vectors from the embedding matrix. The final hidden state of the  $[CLS]$  token for a given sequence serves as the sequence representation.

### 3.2.2 Masked Language Learning with BERT

We utilize the Masked Language Modeling (MLM) approach as implemented in BERT [42]. At its core, this method employs a transformer encoder composed of  $M$  identical blocks. Each block includes a self-attention mechanism followed by a feedforward neural network, as described in [41].

Let  $F = [f_1, f_2, f_3, \dots, f_n]$  denote an input sequence of  $n$  tokens, where each token is represented by a  $p$ -dimensional vector. A self-attention layer processes this sequence using the following transformation:

$$\text{Attention}_{seq}(Q_{seq}, K_{seq}, V_{seq}) = \text{softmax}\left(\frac{Q_{seq}K_{seq}^\top}{\sqrt{d_{n_{k_{seq}}}}}\right)V_{seq}, \quad (3)$$

where,  $Q_{seq} = W_{q_{seq}}F$ ,  $K_{seq} = W_{k_{seq}}F$ , and  $V_{seq} = W_{v_{seq}}F$ . Here,  $W_{q_{seq}}$ ,  $W_{k_{seq}}$ , and  $W_{v_{seq}}$  are learnable matrices that project the input into query, key, and value spaces of node sequences, respectively.  $d_{n_{k_{seq}}}$  is the dimension of the key vectors.

The feedforward layer, applied independently to each token, performs the transformation:

$$FFN(\xi) = \text{ReLU}(\xi W_{seq1} + b_{seq1})W_{seq2} + b_{seq2}, \quad (4)$$

where  $W_{seq1}$ ,  $W_{seq2}$  are learnable matrices, and  $b_{seq1}$ ,  $b_{seq2}$  are bias vectors.  $FFN$  is the feed-forward network and  $\xi$  is the global gene embeddings.

The MLM task involves masking a random subset of input tokens and predicting their identities based on the remaining context. This encourages the model to capture bidirectional contextual relationships within sequences. Specifically, we mask 20% of the tokens (representing nodes) in each sequence and train the model to recover the masked tokens using a cross-entropy loss function:

$$L_{bert} = - \sum_{b=1}^B \sum_{ln=1}^{Ln} \sum_{cl=1}^{Mc} 1_{\{b, ln \in \text{mask}\}} \cdot y_{b, ln, cl} \log p_{b, ln, cl}, \quad (5)$$

where,  $B$  is the batch size,  $Ln$  is the sequence length and  $Mc$  is the number of classes (total number of possible tokens that can be predicted).  $y_{b, ln, cl}$  is a binary indicator equal to 1 if the correct class of token  $ln$  in batch  $b$  is  $cl$ , and  $p_{b, ln, cl}$  is the predicted probability for this classification. The final embeddings are extracted from the embedding layer represented as  $\xi$ .

This enables the model to learn rich contextual embeddings for nodes, capturing both structural and positional relationships within the networks.

### 3.3 Graph Transformer for GRN inference

After deriving features from available GRNs and gene expression data, we utilize the Graph Transformer (GT) to learn comprehensive representations by injecting the underlying regulatory structure. Since GT is specifically designed to model complex dependencies in graph-structured data, it effectively captures gene-gene interactions based on attention mechanisms, making it well suited for GRN-based representation learning. *GT-GRN* is illustrated in Figure 3.

#### 3.3.1 Graph Positional Encodings

NLP-based Transformers are supplied with Positional Encodings. At the heart of GT, graph positional encodings hold a special place which is important for encoding node positions. From the available graph structure, we make use of Laplacian eigenvectors and use them as graph positional encoding ( $PE_{graph}$ ) information. This is helpful to encode distance-aware information, i.e., nearby nodes would have similar positional features and vice versa. Eigenvectors are defined as the factorization of the graph Laplacian matrix:

$$\Delta = I_{graph} - D^{-\frac{1}{2}} \mathcal{A} D^{-\frac{1}{2}} = U \Lambda U^T \quad (6)$$

where  $\mathcal{A}$  is the input adjacency matrix of  $n \times n$  dimensions,  $I_{graph}$  is the identity matrix of size  $n \times n$ .  $D$  is the degree matrix,  $\Lambda$  is the eigenvalues and  $U$  are the eigenvectors. We then use the  $p$  smallest significant eigen-vectors of a node as its positional encoding, which is denoted by  $\lambda_i$  for node  $i$ .

#### 3.3.2 Input to *GTGRN*

The input to the Graph Transformer (GT) layer is the graph structure  $\mathcal{G}$  and its associated features  $h$ . The features  $h$  are constructed as a combination of gene expression embeddings ( $\mathcal{Z}$ ), global gene embeddings ( $\xi$ ), and graph positional embeddings ( $\lambda$ ), which are derived from the graph  $\mathcal{G}$ .

- For gene expression embeddings, each gene  $\mathcal{Z}_i \in \mathbb{R}^{d_n \times 1}$  is passed through a linear projection layer to embed it into a  $d$ -dimensional space.

$$\hat{\mathcal{Z}}_i^0 = S^0 \mathcal{Z}_i + s^0 \quad (7)$$

where,  $S^0 \in \mathbb{R}^{d \times d_n}$  is a learnable weight matrix,  $s^0 \in \mathbb{R}^d$  is the learnable bias vector, and  $\hat{\mathcal{Z}}_i^0 \in \mathbb{R}^d$  is the projected embedding.

- For the global gene embeddings, each vector  $\xi_i \in \mathbb{R}^{d_n \times 1}$  is also embedded into a  $d$  dimensional space using a separate linear projection layer. The transformation is defined as:

$$\hat{\xi}_i^0 = T^0 \xi_i + t^0 \quad (8)$$

where,  $T^0 \in \mathbb{R}^{d \times d_n}$  is a learnable weight matrix,  $t^0 \in \mathbb{R}^d$  is the learnable bias vector.

- The graph positional encodings are extracted from the input graph  $\mathcal{G}$ . For a particular gene's positional encoding  $\lambda_i \in \mathbb{R}^{d_n}$  is embedded into  $d$ -dimensional space using a linear projection layer which is given by:

$$\lambda_i^0 = U^0 \lambda_i + u^0 \quad (9)$$

where,  $U^0 \in \mathbb{R}^{d \times d_n}$  is a learnable weight matrix,  $u^0 \in \mathbb{R}^d$  is a learnable bias vector.

Finally, the gene expression embeddings  $\hat{\mathcal{Z}}^0 \in \mathbb{R}^{n \times d}$ , global gene embeddings  $\hat{\xi}^0 \in \mathbb{R}^{n \times d}$ , and graph positional embeddings  $\lambda^0 \in \mathbb{R}^{n \times d}$  are each projected into a shared  $d$ -dimensional space through separate linear transformation layers. These projected representations are then summed element-wise (often called fusion by summation) to form the final node features  $h \in \mathbb{R}^{n \times d}$ .

$$h = \hat{\mathcal{Z}}^0 + \hat{\xi}^0 + \lambda^0 \quad (10)$$

where,  $h$  is the unified, per-gene embedding that fuses its expression profile, its topological position in the regulatory network, and a dataset-wide global context. This final representation  $h$  is then injected into the GT layer along with the adjacency matrix  $\mathcal{A}$  of the input graph  $\mathcal{G}$ .

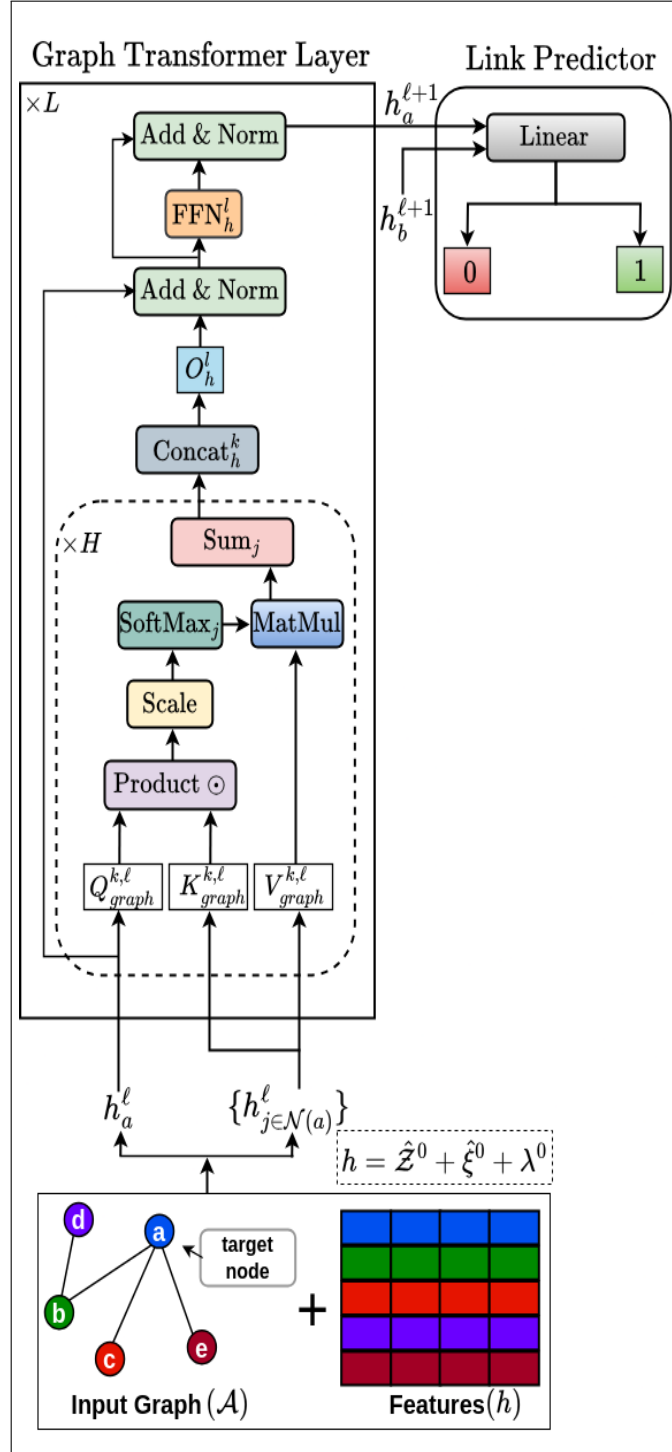


Figure 3: **Architecture diagram of GT-GRN.** Graph Transformer layer operates on the input graph  $\mathcal{A}$  with its corresponding features  $h$ . The input features consists of gene expression embeddings, graph positional encodings and the global gene embeddings. It operates to compute the node embeddings for a particular node  $a$  after passing through multiple layers  $L$  to produce representations at the next layer  $h_a^{\ell+1}$ . Link prediction is done via a separate link predictor module that takes two node embeddings say  $h_a^{\ell+1}$  and  $h_b^{\ell+1}$  to predict a link between them.



### 3.3.3 Graph Transformer Layer

The node update in the GT at layer  $\ell$  is defined as follows:

$$\hat{h}_i^{\ell+1} = O_h^\ell \left\|_{k=1}^H \left( \sum_{j \in \mathcal{N}(i)} w_{ij}^{k,\ell} V_{graph}^{k,\ell} h_j^\ell \right) \right. \quad (11)$$

$$\text{where, } w_{ij}^{k,\ell} = \text{softmax}_j \left( \frac{Q_{graph}^{k,\ell} h_i^\ell \cdot K_{graph}^{k,\ell} h_j^\ell}{\sqrt{d_k}} \right), \quad (12)$$

and  $d_k = d/H$ ,  $k = 1$  to  $H$  denotes number of attention heads, and  $\left\| \right.$  denotes the contatenation of the number of heads..  $Q_{graph}^{k,\ell}, K_{graph}^{k,\ell}, V_{graph}^{k,\ell} \in \mathbb{R}^{d_k \times d}$ ,  $O_h^\ell \in \mathbb{R}^{d \times d}$ , The attention outputs  $\hat{h}_i^{\ell+1}$  are then passed to a Feed Forward Network (FFN) preceded and succeeded by residual connections and normalization layers as:

$$\hat{\hat{h}}_i^{\ell+1} = \text{Norm} \left( h_i^\ell + \hat{h}_i^{\ell+1} \right) \quad (13)$$

$$\hat{\hat{h}}_i^{\ell+1} = W_{graph2}^\ell \text{ReLU} \left( W_{graph1}^\ell \hat{\hat{h}}_i^{\ell+1} \right) \quad (14)$$

$$h_i^{\ell+1} = \text{Norm} \left( \hat{\hat{h}}_i^{\ell+1} + \hat{h}_i^{\ell+1} \right) \quad (15)$$

where  $W_{graph1}^\ell \in \mathbb{R}^{2d \times d}$ ,  $W_{graph2}^\ell \in \mathbb{R}^{d \times 2d}$ , and  $\hat{h}_i^{\ell+1}, \hat{\hat{h}}_i^{\ell+1}$  are the intermediate representations. Norm could be either Layer-Norm[43] or BatchNorm[44].

### 3.3.4 Link Prediction with learned representations

The final module is designed to predict edges between nodes using the learned node representations  $h^{\ell+1}$  obtained from the GT layer. The module takes as input the embeddings  $h^{\ell+1} \in \mathbb{R}^{n \times d}$ , where  $n$  is the number of nodes and  $d$  is the embedding dimension, along with an edge index representing node pairs. For each edge  $(i, j)$ , the embeddings of the source node  $h_i^{\ell+1}$  and the destination node  $h_j^{\ell+1}$  are extracted and concatenated to form a feature vector. This vector is passed through a decoder network consisting of a multi-layer perceptron (MLP) with a hidden layer, ReLU activation, and an output layer, which reduces the concatenated vector to a scalar. The scalar represents the predicted likelihood of an edge between the nodes  $i$  and  $j$ . By utilizing the updated node embeddings  $h^{\ell+1}$ , this module effectively learns to identify and score potential edges in the graph.

Next, we discuss the experimental setup used to demonstrate the superiority of *GT-GRN*.

## 3.4 Experimental Setup

The performance of *GT-GRN* is evaluated on Linux based NVIDIA RTX A3000 GPU as the computing machine. The deep learning libraries used here are pytorch<sup>1</sup>, dgl<sup>2</sup>, scikit-learn<sup>3</sup>, Pytorch Geometric<sup>4</sup>.

## 3.5 Datasets

We establish our findings on two single-cell RNA-sequencing (scRNA-seq) human cell types: human embryonic stem cells (hESC) [45] and mouse embryonic stem cells (mESC) [46] from BEELINE [47] study. The cell-type-specific ChIP-seq ground-truth networks are used as a reference for these datasets. Additionally, we use the synthetic expression

<sup>1</sup><https://pytorch.org/>

<sup>2</sup><https://www.dgl.ai/>

<sup>3</sup><https://scikit-learn.org/stable/>

<sup>4</sup><https://pytorch-geometric.readthedocs.io/en/latest/index.html#>

profiles generated using GeneNetWeaver (GNW) [48], a simulation tool developed for DREAM<sup>5</sup> along with their corresponding ground-truth networks. The details of the datasets have been discussed in Table 1.

Table 1: Dataset statistics

Species/Cell Types	Type	Source	$\mathcal{V}$	$\mathcal{E}$
Yeast	Microarray	GNW	4000	11323
hESC-500	scRNA-seq	BEELINE	910	3940
mESC-500	scRNA-seq	BEELINE	1120	20923
hESC-1000	scRNA-seq	BEELINE	1410	6139
mESC-1000	scRNA-seq	BEELINE	1620	30254

### 3.5.1 Preprocessing of raw data

We preprocess the raw scRNA-seq data using an established method [47] to handle redundancy. We filter out low-expressed genes and prioritized the variable ones. Primarily, the genes expressed in less than 10% of cells were removed. Then, we computed the variance and P-values for each gene, selecting those with P-values below 0.01 after Bonferroni correction. Gene expression levels were log-transformed for normalization. This yielded a feature matrix  $\mathcal{X} \in \mathbb{R}^{n \times m}$ , where  $n$  is the number of genes and  $m$  is the number of cells. Furthermore, we adopt the approach of Pratapa et al.[47] to assess performance across different network sizes. Specifically, we rank genes by variance and select the most variable transcription factors (TFs), along with the top 500 and 1000 genes with the highest variability.

### 3.6 Baseline methods

We evaluate the efficacy of *GT-GRN* against the existing baselines methods commonly used for inferring GRNs are shown in Table 2.

## 4 Results & Analysis

We report results using both single-cell and microarray gene expression datasets, selecting representative methods from each major category of gene regulatory network (GRN) inference techniques against *GT-GRN*. These include mutual information-based methods, feature selection approaches, ensemble tree-based models, and graph neural network frameworks. This diverse selection allows us to comprehensively evaluate performance across different inference paradigms, ensuring a balanced comparison that highlights the strengths and limitations of each method.

### 4.1 GRN Inference via Full Network Reconstruction

Fundamentally, Gene Regulatory Network (GRN) inference aims to reconstruct the entire regulatory network, capturing the full complexity of gene interactions. By striving for complete network reconstruction, it seeks to reveal the intricate web of regulatory relationships among all involved genes, reflecting the true, comprehensive regulatory architecture [55]. We evaluate the effectiveness of *GT-GRN* alongside existing methods designed for GRN inference. Following the similar motivation, we report these results for *GT-GRN* and its baseline methods in Figure 4. From the figure, the results of the full network reconstruction, highlight the comparative performance of different methods. For the scRNA dataset, the performance of *GT-GRN* remains consistently higher with minor variations for all datasets. This suggests that the *GT-GRN* method is robust for different cell types and sequencing depths.

For the *Yeast* (microarray) dataset, the *GT-GRN* method significantly outperforms all other network inference methods, indicating its superior ability to capture gene interactions. Baseline methods such as ARACNE, BC3NET, C3MTC, C3NET, and MRNET perform at similar levels. In general, *GT-GRN* appears to be the most effective method for the microarray dataset, while for scRNA expression profiles the method demonstrate stable performance across conditions in terms of AUROC score.

### 4.2 GRN Inference via Link Prediction

There exist many benchmark methods that treat GRN inference as a link prediction problem, focusing on identifying only a limited subset of interactions. We report results using this conventional approach. Table 3 reports the performance results for the candidate datasets in terms of *Area Under the Receiver-Operating Characteristic Curve* (AUROC) and *Area Under the Precision-Recall Curve* (AUPRC) metrics. To assess the predictive performance of *GT-GRN*, we

<sup>5</sup>Dialogue on Reverse Engineering Assessment and Methods

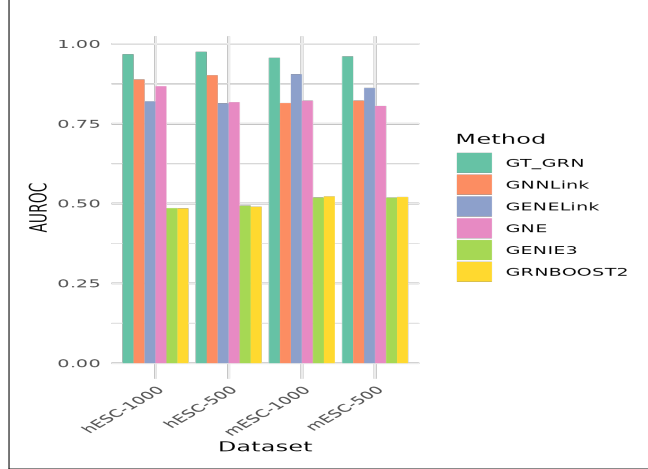
Table 2: Summary of Gene Regulatory Network (GRN) Inference Methods classified by Category.

Category	Method	Description	Source
Graph Neural Network	GNNLink [38]	Uses a GCN-based interaction graph encoder to capture gene expression patterns.	<a href="https://github.com/sdesignates/GNNLink">https://github.com/sdesignates/GNNLink</a>
	GENELink [49]	Leverages a graph attention network (GAT) to infer GRNs via attention mechanisms.	<a href="https://github.com/zpliulab/GENELink">https://github.com/zpliulab/GENELink</a>
	GNE [33]	Uses an multi-layer perceptron (MLP) to encode gene expression profiles and network topology for predicting gene regulatory links.	<a href="https://github.com/kckishan/GNE">https://github.com/kckishan/GNE</a>
Mutual Information	ARACNE [24]	Infers networks based on Adaptive Partitioning (AP) and Mutual Information (MI) approach.	<a href="https://bioconductor.org/packages/release/bioc/html/minet.html">https://bioconductor.org/packages/release/bioc/html/minet.html</a>
	BC3NET [50]	An ensemble technique derived from the C3NET algorithm that employs a bagging approach for inferring GRNs.	<a href="https://cran.r-project.org/web/packages/bc3net/index.html">https://cran.r-project.org/web/packages/bc3net/index.html</a>
	C3MTC [51]	Infers gene interaction networks where edge weights are defined by corresponding mutual information values.	<a href="https://cran.r-project.org/web/packages/bc3net/index.html">https://cran.r-project.org/web/packages/bc3net/index.html</a>
	C3NET [52]	Leverages mutual information estimates and a maximization step to efficiently capture causal structural information in the data.	<a href="https://cran.r-project.org/web/packages/c3net/index.html">https://cran.r-project.org/web/packages/c3net/index.html</a>
Feature Selection	MRNET [25]	Applies a series of supervised gene selection procedures using the MRMR (maximum relevance/minimum redundancy) principle.	<a href="https://bioconductor.org/packages/release/bioc/html/minet.html">https://bioconductor.org/packages/release/bioc/html/minet.html</a>
Ensemble Tree-Based	GRNBOOST2 [53]	A fast GRN inference algorithm using Stochastic Gradient Boosting Machine (SGBM) regression.	<a href="https://github.com/aertslab/arboreto">https://github.com/aertslab/arboreto</a>
	GENIE3 [54]	A classic GRN inference algorithm using Random Forest (RF) or ExtraTrees (ET) regression.	<a href="https://github.com/aertslab/arboreto">https://github.com/aertslab/arboreto</a>

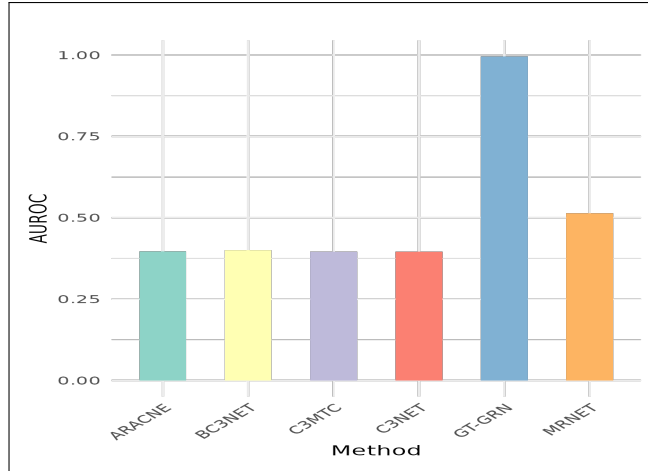
present a comparison with various baseline methods. From the table it is clearly evident that the *GT-GRN* consistently outperforms other methods in all datasets, achieving the highest AUROC and excelling in AUPRC, particularly for the mESC-1000 and mESC-500 datasets. GNE emerges as a strong contender, especially in the hESC-1000 dataset, where it achieves the highest AUPRC 0.9042 and the AUROC of 0.9025, indicating its effectiveness in balancing precision and recall. GENELink maintains strong performance across most datasets, while GNNLink performs well in AUROC but lags in AUPRC. In contrast, GENIE3 and GRNBOOST2 show consistently lower scores, indicating challenges in handling these complex datasets. Notably, GNE outperforms all methods in both AUROC and AUPRC for the hESC-1000 dataset, highlighting its scalability. For the Yeast-4000 dataset, *GT-GRN* demonstrates superior performance compared to its baseline counterparts. Overall, *GT-GRN* proves to be the most reliable across datasets, while GNE stands out for its precision, making them the top choices for biological network predictions in this study.

### 4.3 Impact of Hyperparameters

Additionally, we aim to investigate the role of various hyperparameters (HPs) that influence the overall performance of the models. Optimal selection of HPs is a difficult task and time-consuming activity. We analyze the behavior of the learning models compared to *GT-GRN* by selecting key hyperparameters from each baseline model to assess their impact on the overall performance of each model. Table 4 describes the parameteric configurations that we tuned for each learning model with their respective explanations. We summarize the overall impact of hyperparameter tuning for different models using a boxplot, as shown in Figure 5. This visualization presents a comprehensive comparison of our model, *GT-GRN*, against several baseline approaches in terms of AUROC performance across multiple datasets.



(a)



(b)

Figure 4: **Full network reconstruction performance of various methods for different datasets in terms AUROC score.** (a) *BEELINE's* scRNA-seq datasets and (b) *GNW's* Yeast dataset.

A key observation is that *GT-GRN* consistently achieves higher median AUROC scores with notably lower variance across datasets, highlighting its robustness and reliability under varying experimental conditions. Each model was tuned with its respective optimal hyperparameters, yet *GT-GRN* exhibits both stability and effectiveness across settings, clearly outperforming the existing baselines on all candidate datasets. In contrast, *GNNLink* shows a higher variance, particularly in the hESC-500 dataset, where its performance fluctuates significantly. *GENELink* and *GNE* display relatively stable performances, though both fall short of the superior AUROC achieved by *GT-GRN*. In particular, the mESC-1000 dataset highlights the clear dominance of *GT-GRN*, with its AUROC surpassing that of all other methods by a substantial margin.

## 5 Application of *GT-GRN* on Cell-Type Classification

Cell-type-specific gene regulatory networks (GRNs) are crucial for defining transcriptional states during development, with each cell-type being characterized by a unique set of active transcription factors (TFs). These GRNs offer an unbiased method for studying gene regulation, providing valuable insights into the mechanisms driving cellular diversity. In this context, we explore the effectiveness of *GT-GRN* in reconstructing cell-type-specific GRNs, with the goal of cell-type annotation. To achieve this, we apply *GT-GRN* to single-cell RNA sequencing (scRNA-seq) data from over

Table 3: AUROC and AUPRC scores for different methods across the various scRNA-seq datasets.

Dataset	Method	AUROC	AUPRC
mESC-1000	<i>GT-GRN</i>	<b>0.9483</b>	<b>0.8990</b>
	GNNLink	0.8833	0.8660
	GENELink	0.9133	0.8103
	GNE	0.8984	0.8925
mESC-500	<i>GT-GRN</i>	<b>0.9402</b>	<b>0.8853</b>
	GNNLink	0.8768	0.8331
	GENELink	0.9057	0.8004
	GNE	0.8378	0.8416
hESC-500	<i>GT-GRN</i>	<b>0.8793</b>	0.5932
	GNNLink	0.8251	0.4542
	GENELink	0.8618	0.5581
	GNE	0.8402	<b>0.8466</b>
hESC-1000	<i>GT-GRN</i>	0.8784	0.8604
	GNNLink	0.8442	0.5011
	GENELink	0.8657	0.5610
	GNE	<b>0.9025</b>	<b>0.9042</b>

Table 4: **Tuned hyper-parameters for different models.** *Epochs* sees the entire dataset during training, *Learning Rate* controls the pace at which algorithm updates, *Output Dimensions* refers to the final layer’s shape from which the predictions are made, *Attention Heads* mechanism computes attention scores over the input elements and aggregates information. *Layers* refer to the number of transformation steps that processes inputs and passes the transformed output to the next layer.

Model	Epochs	Learning Rate	Output Dimensions	Attention Heads	Layers
<i>GT-GRN</i>	-	0.001, 0.003, 0.0005	-	2,4,8	4,6,8
GNNLink	100, 200, 300	0.001, 0.005, 0.01	128, 256, 512	-	-
GENELink	10, 20, 30	0.001, 0.003, 0.0005	128, 256, 512	-	-
GNE	10, 20, 30	0.001, 0.005, 0.01	128, 256, 512	-	-

8,000 human peripheral blood mononuclear cells (PBMCs8k), sourced from 10X Genomics<sup>6</sup>. The data is preprocessed using the Scanpy framework[56], ensuring efficient handling and analysis of the single-cell data. For the ground truth network, we utilize the hTFtarget database[57], which integrates ChIP-seq data, transcription factor binding sites, and epigenetic modification information. This comprehensive resource provides detailed insights into gene regulation and TF-target interactions, making it an invaluable tool for studying gene regulatory mechanisms.

In order to perform cell-type annotation, it is essential to first reconstruct the GRN. This involves generating embeddings that represent the cell types based on their gene regulatory networks. These embeddings serve as the foundation for cell-type classification, enabling accurate annotation of the cell types based on their unique regulatory patterns. By leveraging the power of *GT-GRN* in inferring cell-type-specific GRNs, we aim to advance the cell-type annotation in scRNA-seq data, ultimately improving our understanding of the complex regulatory landscapes that define cellular identities.

### 5.1 *GT-GRN* for PBMC Network Reconstruction

We investigate GRN inference using *GT-GRN* by formulating it as a network regeneration problem on the PBMC dataset. First, we filter the data by removing genes expressed in fewer than 5% of cells and discarding cells that express fewer than 200 genes. Next, we normalize the total counts per cell to 10,000, ensuring comparability across cells. To further refine the data, we apply MAGIC[58] imputation, which reduces noise and improves expression patterns. Finally, we perform a logarithmic transformation to improve interpretability and optimize the data for downstream analysis.

<sup>6</sup><https://www.10xgenomics.com/datasets/8-k-pbm-cs-from-a-healthy-donor-2-standard-2-1-0>

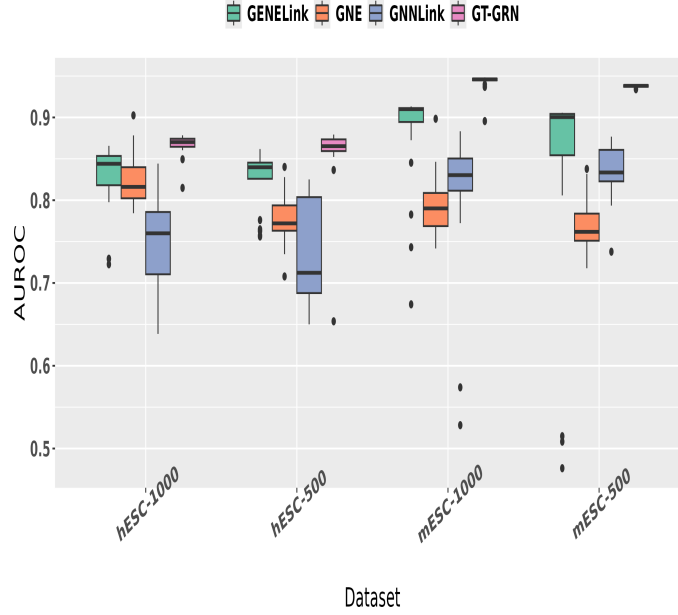


Figure 5: **Overall hyper-parameter tuning plot for various models.**

After preprocessing, we employ *GT-GRN* to reconstruct the PBMC gene regulatory network and compare its performance against existing baseline methods. Table 5 reveals the comparison of PBMC’s hTFTarget (Gold standard) and generated networks reveals key structural differences and predictive performance variations. *GT-GRN* achieves the highest AUROC (0.9769) while maintaining balanced connectivity and clustering, making it the best-performing model. GENELink and GNNLink exhibit dense connectivity, high clustering, and shorter path length but are highly disassortative, indicating a strong preference for high-degree nodes connecting to low-degree ones. GNE, with lower connectivity and clustering, results in longer path lengths and the lowest AUROC (0.7596) but retains some structural similarities to the Gold network. The Gold standard itself maintains moderate connectivity and a sparse clustering structure, serving as a key benchmark. We also measure the quality of the generated graph network characteristics from the candidate methods with the input network (hTFTarget) using a single measurement score using Pearson correlation coefficient. The results show that all methods exhibit strong correlation, with GNE (0.9992) achieving the highest agreement, followed closely by *GT-GRN* (0.9838). However, GNNLink and GENELink report the similar score of 0.9817. These insights highlight the trade-offs between network structure and predictive performance, guiding model selection for biological network analysis.

Further, we analyze the degree-distribution plot of the generated networks in comparison to the input PBMC’s hTFTarget network. Figure 6 describes the log-log degree distribution plot compares the degree distributions of the input (Gold) and generated networks (*GT-GRN*, GENELink, GNNLink, and GNE). The Gold network follows a natural decay which is scale-free in degree distribution, while *GT-GRN* shows a similar trend with slight deviations. GNE displays a more scattered pattern, indicating the similar tailed degree-distribution. GENELink and GNNLink exhibit significantly higher maximum degrees, i.e., these generated networks generate more high-degree nodes than the input network.

Next, we assess how well our embeddings capture the community structure by clustering gene embeddings. We utilize the Leiden algorithm [59] to cluster the resultant gene embeddings. Figure 7 presents a uniform manifold approximation and projection (UMAP) dimensional reduction of gene representations for various methods. The methods compared include Gene Expression data, GENELink, GNNLink, GNE, and *GT-GRN*. The visualization clearly shows the distinct community structures produced by GNNLink, and *GT-GRN* embeddings that indicates effective preservation of biological modules.

## 5.2 GT-GRN for Cell-Type Annotation

We further investigate these embeddings for cell-type annotation task. We manually annotate cell-types for gene expression data and focus on four cell-types with highest number of cells, *CD4+ T cells*, *CD14+ Monocyte cells*, *CD8+ T cells*. We train a three-layered multi-layered perceptron classifier for annotating cell-types. The classifier is

Table 5: **Network Characteristics Comparison of PBMC’s hTFTarget and Generated Networks with AUROC Score.** *Maximum degree* computes the degree over all vertices. *Assortativity* is the pearson correlation of degrees of connected nodes. *Triangle Count* denotes the connection between two nodes. *Clustering Coefficient* measure of the tendency of nodes in a network to form triangles. *Characteristic Path Length* represents the average shortest path length between all nodes pairs in a network. *PCC* is the Pearson Correlation Coefficient between hTFTarget and generated network characterisits.

Model	Maximum Degree	Assortativity	Triangle Count	Clustering Coefficient	Characteristic Path Length	AUROC	PCC
Gold	1837	-0.4762	9596	0.0060	2.7392	-	-
GT-GRN	2593	-0.6614	207579	0.0255	2.1934	0.9769	0.9838
GENELink	3999	-0.9867	5671530	0.0389	1.9731	0.8810	0.9817
GNNLink	3999	-0.9867	5671530	0.0389	1.9731	0.8467	0.9817
GNE	924	-0.1872	4014	0.0094	2.9621	0.7596	0.9992

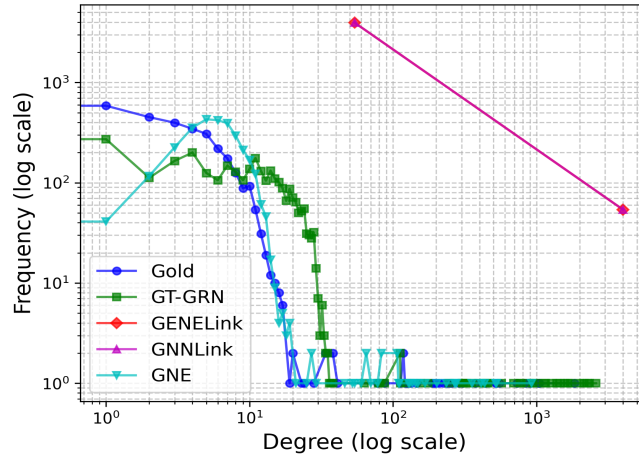


Figure 6: **Log-Log Degree Distribution plot of the input PBMC’s hTFTarget Network and Generated Network for different models.**

trained using in multi-class classification setting in five-fold cross validation setting. We benchmark *GT-GRN* against GENELink, GNNLink and GNE methods. Figure 8 demonstrate that *GT-GRN* effectively captures the cell-types using gene representation classification setup.

Next, we delve into the individual contributions of each embedding modality within the *GT-GRN* framework through a systematic ablation study on the PBMC dataset. This dataset provides a biologically rich and diverse single-cell

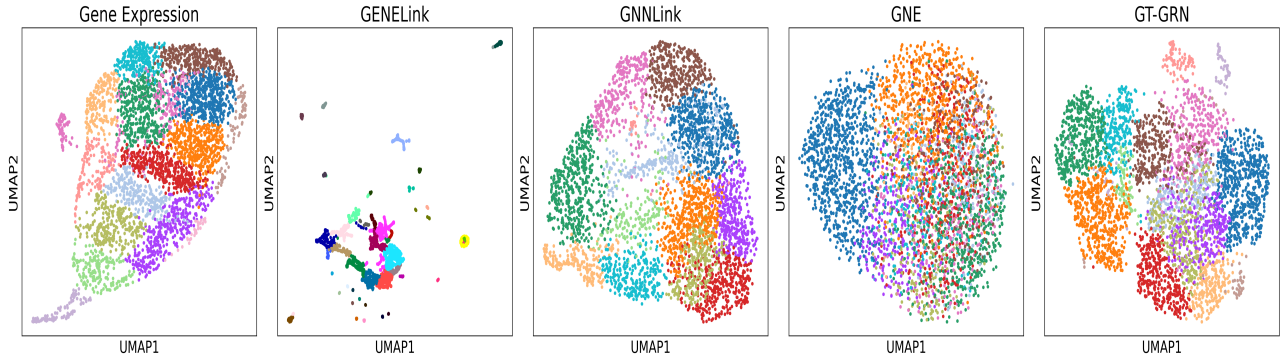


Figure 7: **UMAP visualization of genes representations for PBMC’s dataset according to different methods.**



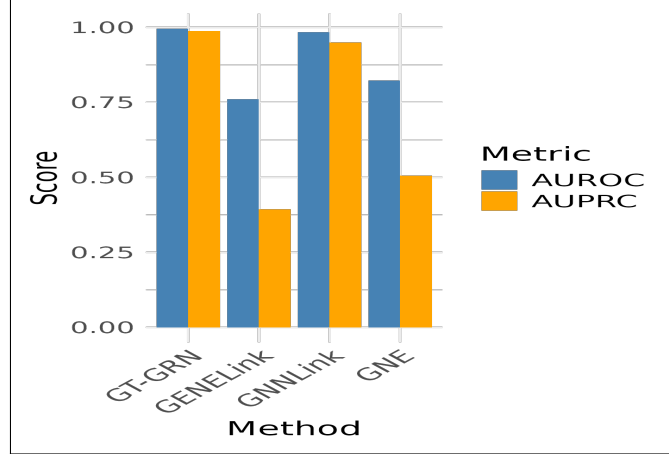


Figure 8: **Cell-Type Classification.** AUROC and AUPRC score of *GT-GRN*, *GENELink*, *GNNLink*, *GNE* based embeddings in annotation cell-types using MLP classifier in 5 fold cross-validation setting.

expression landscape, making it an ideal benchmark for evaluating the role of each embedding component in GRN inference.

### 5.3 Ablation Studies

To assess the overall efficacy and robustness of *GT-GRN*, we performed a comprehensive ablation study that systematically examines the individual and combined contributions of its embedding modalities: structural positional encodings, global embeddings, and gene expression embeddings. This experiment is crucial, given the multi-component nature of our framework, as it allows us to systematically assess the role and efficacy of each component in the GRN inference process under a link prediction setup. We organize our baselines into three categories: uni-modal, bi-modal, and tri-modal, where the prefixes "uni", "bi", and "tri" denote the number of information sources used. This study establishes the necessity of each module in the overall architecture, demonstrating that omitting any single modality leads to a significant performance drop—thereby justifying the integration of all components for optimal GRN reconstruction.

- *Structural Positional Encodings* : In this uni-modal baseline, we have used the  $PE_{graph}$  of the input network which is then fed into the *GT-GRN*. Here, each gene positional encodings is of length 512. This vector is fed into the model to predict possibility of link with other gene vectors based on the input link information.
- *Global Embeddings* : This is a uni-model baseline, that extracts the global knowledge of inferred GRNs in a multi-network integration framework using BERT model. The output length of each gene embedding here is 512. The vector is given as input into the model to estimate the likelihood of forming links with other gene vectors for link inference.
- *Gene Expression Embeddings* : In this uni-modal baseline, raw gene expression is converted into a embedding vector to capture the latent information using auto-encoder model. The length of the embedding vector for each gene is 512 which is then used to estimate the link likelihood using *GT-GRN* framework.
- *Structural Positional Encodings + Global Embeddings* : This multi-modal approach combines the structural positional encodings of the input network with global embeddings derived from the BERT-based multi-network integration framework. The fused representation is used in the *GT-GRN* model to enhance link prediction performance.
- *Structural Positional Encodings + Gene Expression Embeddings* : This baseline integrates the structural positional encodings of the input network with gene expression embeddings obtained through an autoencoder. The combined vector representation is fed into the *GT-GRN* model to infer potential links between genes.
- *Global Embeddings + Gene Expression Embeddings*

In this setup, global embeddings capturing inferred GRN knowledge are combined with gene expression embeddings. The resulting representation is used to predict gene interactions within the *GT-GRN* framework.



- *Structural Positional Encodings + Global Embeddings + Gene Expression Embeddings* : This comprehensive multi-modal approach fuses all three embeddings—structural positional encodings, global embeddings, and gene expression embeddings—to provide a richer representation for link inference. This integrated approach aims to leverage complementary information from multiple modalities to enhance predictive performance.

Table 6 reports the ablation study results from different feature sets for *GT-GRN* using AUROC scores. Among unimodal representations, Global Embeddings achieve the highest AUROC of 0.8860, followed by Gene Expression Embeddings 0.8693 and Structural Positional Encodings with AUROC score of 0.8480, indicating that global information is the most informative for link inference.

Bimodal combinations improve performance, with Global Embeddings + Structural Positional Encodings 0.8843 and Global Embeddings + Gene Expression Embeddings 0.8841 performing best. The trimodal combination of all three embeddings achieves the highest AUROC 0.8877, demonstrating that integrating multiple modalities provides the most effective representation for GRN inference.

Overall, the study highlights the importance of multi-modal integration, with Global Embeddings playing a key role in enhancing predictive performance.

Table 6: Ablation study in terms of different features for *PBMC*’s dataset for *GT-GRN* framework.

Modal	Feature sets	AUROC
Unimodal	Structural Positional Encodings	0.8480
	Global Embeddings	0.8860
	Gene Expression Embeddings	0.8693
Bimodal	Structural Positional Encodings + Global Embeddings	0.8843
	Structural Positional Encodings + Gene Expression Embeddings	0.8666
	Global Embeddings + Gene Expression Embeddings	0.8841
Trimodal	Structural Positional Encodings + Global Embeddings + Gene Expression Embeddings	<b>0.8877</b>

## 6 Conclusion

In this work, we proposed a novel gene regulatory network (GRN) inference framework, *GT-GRN*, which leverages Graph Transformers to infer regulatory links by incorporating graph-based techniques. Our approach begins by generating gene embeddings through an autoencoder. We then integrate prior network knowledge from known GRNs using an NLP-based BERT model, where these graphs are converted into sequences to extract contextual embeddings. Additionally, we incorporate graph positional information to enhance the inference process.

Through extensive experiments, *GT-GRN* demonstrates superior link prediction performance compared to baseline methods. We further assess the quality of the generated embeddings by evaluating their community structure, showing that *GT-GRN* effectively supports cell-type annotation in real PBMC gene expression datasets. Our ablation study reveals that the combination of gene expression data, global gene context, and positional information significantly contributes to improved GRN inference precision. While the current work focuses on accurate reconstruction of Gene Regulatory Networks (GRNs) using multi-modal embeddings, ongoing efforts are directed toward extending the framework to prioritize disease-associated genes. By leveraging the inferred GRNs, the goal is to identify key regulatory hubs and pathways that may play pivotal roles in disease development and progression.

## 7 Acknowledgement

This work is supported by IDEAS-TIH, ISI-Kolkata research project (funded by Department of Science & Technology (DST), Govt. of India) at the Network Reconstruction and Analysis (NETRA) Lab, Department of Computer Applications, Sikkim (Central) University.

## References

- [1] Pau Bellot, Catharina Olsen, Philippe Salembier, Albert Oliveras-Vergés, and Patrick E Meyer. Netbenchmark: a bioconductor package for reproducible benchmarks of gene regulatory network inference. *BMC bioinformatics*, 16:1–15, 2015.
- [2] Alberto de la Fuente. What are gene regulatory networks? In *Handbook of research on computational methodologies in gene regulatory networks*, pages 1–27. IGI Global, 2010.
- [3] Pietro Hiram Guzzi and Swarup Roy. *Biological network analysis: Trends, approaches, graph theory, and algorithms*. Elsevier, 2020.
- [4] Radha Shyamsundar, Young H Kim, John P Higgins, Kelli Montgomery, Michelle Jorden, Anand Sethuraman, Matt van de Rijn, David Botstein, Patrick O Brown, and Jonathan R Pollack. A dna microarray survey of gene expression in normal human tissues. *Genome biology*, 6:1–9, 2005.
- [5] Aleksandra A Kolodziejczyk, Jong Kyoung Kim, Valentine Svensson, John C Marioni, and Sarah A Teichmann. The technology and biology of single-cell rna sequencing. *Molecular cell*, 58(4):610–620, 2015.
- [6] Rashel V Grindberg, Joyclyn L Yee-Greenbaum, Michael J McConnell, Mark Novotny, Andy L O’Shaughnessy, Georgina M Lambert, Marcos J Araúzo-Bravo, Jun Lee, Max Fishman, Gillian E Robbins, et al. Rna-sequencing from single nuclei. *Proceedings of the National Academy of Sciences*, 110(49):19802–19807, 2013.
- [7] Divyanshu Talwar, Aanchal Mongia, Debarka Sengupta, and Angshul Majumdar. Autoimpute: Autoencoder based imputation of single-cell rna-seq data. *Scientific reports*, 8(1):16329, 2018.
- [8] Vân Anh Huynh-Thu and Guido Sanguinetti. Gene regulatory network inference: an introductory survey. *Gene regulatory networks: Methods and protocols*, pages 1–23, 2019.
- [9] Monica Jha, Swarup Roy, and Jugal K Kalita. Prioritizing disease biomarkers using functional module based network analysis: a multilayer consensus driven scheme. *Computers in Biology and Medicine*, 126:104023, 2020.
- [10] Swarup Roy, Dhruba K Bhattacharyya, and Jugal K Kalita. Reconstruction of gene co-expression network from microarray data using local expression patterns. *BMC bioinformatics*, 15:1–14, 2014.
- [11] Softya Sebastian, Swarup Roy, and Jugal Kalita. A generic parallel framework for inferring large-scale gene regulatory networks from expression profiles: application to alzheimer’s disease network. *Briefings in Bioinformatics*, 24(1):bbac482, 2023.
- [12] Softya Sebastian, Swarup Roy, and Jugal Kalita. Network-based analysis of alzheimer’s disease genes using multi-omics network integration with graph diffusion. *Journal of Biomedical Informatics*, 164:104797, 2025.
- [13] Gregory T Dewey and David J Galas. Gene regulatory networks. In *Madame Curie Bioscience Database [Internet]*. Landes Bioscience, 2013.
- [14] Juexin Wang, Anjun Ma, Qin Ma, Dong Xu, and Trupti Joshi. Inductive inference of gene regulatory network using supervised and semi-supervised graph neural networks. *Computational and structural biotechnology journal*, 18:3335–3343, 2020.
- [15] Hui Zhang, Xuexin An, Qiang He, Yudong Yao, Feng-Lei Fan, and Yueyang Teng. Quadratic graph attention network (q-gat) for robust construction of gene regulatory networks. *arXiv preprint arXiv:2303.14193*, 2023.
- [16] Yuyang Huang, Gufeng Yu, and Yang Yang. Miggr: A multi-instance graph neural network model for inferring gene regulatory networks for drosophila from spatial expression images. *PLOS Computational Biology*, 19(11):e1011623, 2023.
- [17] Xueying Tian, Yash Patel, and Yue Wang. Trendy: Gene regulatory network inference enhanced by transformer. *bioRxiv*, pages 2024–10, 2024.
- [18] Yue Wang, Peng Zheng, Yu-Chen Cheng, Zikun Wang, and Aleksandr Aravkin. Wendy: Covariance dynamics based gene regulatory network inference. *Mathematical Biosciences*, 377:109284, 2024.
- [19] Jing Xu, Aidi Zhang, Fang Liu, and Xiujun Zhang. Stgrns: an interpretable transformer-based method for inferring gene regulatory networks from single-cell transcriptomic data. *Bioinformatics*, 39(4):btad165, 2023.

- [20] Hantao Shu, Fan Ding, Jingtian Zhou, Yexiang Xue, Dan Zhao, Jianyang Zeng, and Jianzhu Ma. Boosting single-cell gene regulatory network reconstruction via bulk-cell transcriptomic data. *Briefings in Bioinformatics*, 23(5):bbac389, 2022.
- [21] Thalia E Chan, Michael PH Stumpf, and Ann C Babbie. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell systems*, 5(3):251–267, 2017.
- [22] Keiichi Mochida, Satoru Koda, and Komaki Inoue. Statistical and machine learning approaches to predict gene regulatory networks from transcriptome datasets. *Frontiers in Plant Science*, 9:421043, 2018.
- [23] Anne-Claire Haury, Fantine Mordelet, Paola Vera-Licona, and Jean-Philippe Vert. Tigress: trustful inference of gene regulation using stability selection. *BMC systems biology*, 6:1–17, 2012.
- [24] Adam A Margolin, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla Favera, and Andrea Califano. Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. In *BMC bioinformatics*, volume 7, pages 1–15. Springer, 2006.
- [25] Patrick E Meyer, Kevin Kontos, Frederic Lafitte, and Gianluca Bontempi. Information-theoretic inference of large transcriptional regulatory networks. *EURASIP journal on bioinformatics and systems biology*, 2007:1–9, 2007.
- [26] Jeremiah J Faith, Boris Hayete, Joshua T Thaden, Ilaria Mogno, Jamey Wierzbowski, Guillaume Cottarel, Simon Kasif, James J Collins, and Timothy S Gardner. Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS biology*, 5(1):e8, 2007.
- [27] Kevin Bleakley, Gérard Biau, and Jean-Philippe Vert. Supervised reconstruction of biological networks with local models. *Bioinformatics*, 23(13):i57–i65, 2007.
- [28] Zeeshan Gillani, Muhammad Sajid Hamid Akash, MD Matiur Rahaman, and Ming Chen. Comparesvm: supervised, support vector machine (svm) inference of gene regularity networks. *BMC bioinformatics*, 15:1–7, 2014.
- [29] Zahra Razaghi-Moghadam and Zoran Nikoloski. Supervised learning of gene-regulatory networks based on graph distance profiles of transcriptomics data. *NPJ systems biology and applications*, 6(1):21, 2020.
- [30] Meroua Daoudi and Souham Meshoul. Deep neural network for supervised inference of gene regulatory network. In *Modelling and Implementation of Complex Systems: Proceedings of the 5th International Symposium, MISC 2018, December 16-18, 2018, Laghouat, Algeria*, pages 149–157. Springer, 2019.
- [31] Turki Turki, Jason TL Wang, and Ibrahim Rajikhan. Inferring gene regulatory networks by combining supervised and unsupervised methods. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 140–145. IEEE, 2016.
- [32] Guo Mao, Zhengbin Pang, Ke Zuo, and Jie Liu. Gene regulatory network inference using convolutional neural networks from scrna-seq data. *Journal of Computational Biology*, 30(5):619–631, 2023.
- [33] Kishan Kc, Rui Li, Feng Cui, Qi Yu, and Anne R Haake. Gne: a deep learning framework for gene network inference by aggregating biological information. *BMC systems biology*, 13:1–14, 2019.
- [34] Binon Teji, Jayanta K Das, Swarup Roy, and Dinabandhu Bhandari. Predicting missing links in gene regulatory networks using network embeddings: A qualitative assessment of selective embedding techniques. In *Intelligent Systems: Proceedings of ICMIB 2021*, pages 143–154. Springer, 2022.
- [35] Binon Teji, Swarup Roy, Devendra Singh Dhami, Dinabandhu Bhandari, and Pietro Hiram Guzzi. Graph embedding techniques for predicting missing links in biological networks: An empirical evaluation. *IEEE Transactions on Emerging Topics in Computing*, 12(1):190–201, 2023.
- [36] Jiacheng Wang, Yaojia Chen, and Quan Zou. Inferring gene regulatory network from single-cell transcriptomes with graph autoencoder model. *PLoS Genetics*, 19(9):e1010942, 2023.
- [37] Shuo Li, Yan Liu, Long-Chen Shen, He Yan, Jiangning Song, and Dong-Jun Yu. Gmfrn: a matrix factorization and graph neural network approach for gene regulatory network inference. *Briefings in Bioinformatics*, 25(2):bbad529, 2024.
- [38] Guo Mao, Zhengbin Pang, Ke Zuo, Qinglin Wang, Xiangdong Pei, Xinhai Chen, and Jie Liu. Predicting gene regulatory links from single-cell rna-seq data using graph neural networks. *Briefings in Bioinformatics*, 24(6):bbad414, 2023.
- [39] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [40] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.
- [41] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.

- [42] Jacob Devlin. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [43] Jimmy Lei Ba. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.
- [44] Sergey Ioffe. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [45] Ye Yuan and Ziv Bar-Joseph. Deep learning for inferring gene relationships from single-cell expression data. *Proceedings of the National Academy of Sciences*, 116(52):27151–27158, 2019.
- [46] J Gray Camp, Keisuke Sekine, Tobias Gerber, Henry Loeffler-Wirth, Hans Binder, Malgorzata Gac, Sabina Kanton, Jorge Kageyama, Georg Damm, Daniel Seehofer, et al. Multilineage communication regulates human liver bud development from pluripotency. *Nature*, 546(7659):533–538, 2017.
- [47] Aditya Pratapa, Amogh P Jalihal, Jeffrey N Law, Aditya Bharadwaj, and TM Murali. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature methods*, 17(2):147–154, 2020.
- [48] Thomas Schaffter, Daniel Marbach, and Dario Floreano. Genenetweaver: in silico benchmark generation and performance profiling of network inference methods. *Bioinformatics*, 27(16):2263–2270, 2011.
- [49] Guangyi Chen and Zhi-Ping Liu. Graph attention network for link prediction of gene regulations from single-cell rna-sequencing data. *Bioinformatics*, 38(19):4522–4529, 2022.
- [50] Ricardo de Matos Simoes and Frank Emmert-Streib. Bagging statistical network inference from large-scale gene expression data. *PloS one*, 7(3):e33624, 2012.
- [51] Ricardo de Matos Simoes and Frank Emmert-Streib. Influence of statistical estimators of mutual information and data heterogeneity on the inference of gene regulatory networks. *PLoS One*, 6(12):e29279, 2011.
- [52] Gökmen Altay and Frank Emmert-Streib. Inferring the conservative causal core of gene regulatory networks. *BMC systems biology*, 4:1–13, 2010.
- [53] Thomas Moerman, Sara Aibar Santos, Carmen Bravo González-Blas, Jaak Simm, Yves Moreau, Jan Aerts, and Stein Aerts. Grnboost2 and arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*, 35(12):2159–2161, 2019.
- [54] Vân Anh Huynh-Thu, Alexandre Irrthum, Louis Wehenkel, and Pierre Geurts. Inferring regulatory networks from expression data using tree-based methods. *PloS one*, 5(9):e12776, 2010.
- [55] Binon Teji, Swarup Roy, Pietro Hiram Guzzi, and Dinabandhu Bhandari. Application of generative graph models in biological network regeneration: A selective review and qualitative analysis. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 5786–5792. IEEE, 2024.
- [56] F Alexander Wolf, Philipp Angerer, and Fabian J Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19:1–5, 2018.
- [57] Qiong Zhang, Wei Liu, Hong-Mei Zhang, Gui-Yan Xie, Ya-Ru Miao, Mengxuan Xia, and An-Yuan Guo. htftarget: a comprehensive database for regulations of human transcription factors and their targets. *Genomics, proteomics & bioinformatics*, 18(2):120–128, 2020.
- [58] David Van Dijk, Roshan Sharma, Juozas Nainys, Kristina Yim, Pooja Kathail, Ambrose J Carr, Cassandra Burdziak, Kevin R Moon, Christine L Chaffer, Diwakar Pattabiraman, et al. Recovering gene interactions from single-cell data using data diffusion. *Cell*, 174(3):716–729, 2018.
- [59] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12, 2019.