Graph rewiring for long range-aware protein learning

Ali Hariri
EPFL
Lausanne, Switzerland
ali.hariri@epfl.ch

Pierre Vandergheynst
EPFL
Lausanne, Switzerland
pierre.vandergheynst@epfl.ch

Abstract

Peptides and proteins are biomolecules that exist in a broad spectrum of size, structure, and function. Both structure and function are defined by the underlying sequence of amino acids, causing the polyamide to take three-dimensional conformations when in solution. Despite significant efforts and advances in function and conformation prediction, there remains a critical need for computational methods to accurately infer protein function from sequence and structure. Recent advancements in Graph Neural Networks have shown promise in learning the sequence and structure of proteins. However, they fail to capture essential long-range dependencies inherent in the complex and dynamic three-dimensional structures of proteins, leading to issues including oversquashing and oversmoothing. Here, we explore solutions to the challenge of capturing long-range dependencies in graph representations of polyamides, focusing on latent nodes and graph rewiring techniques. While graph rewiring enhances information flow between distant nodes, latent nodes enable the concentration of global information. Our unified framework combines these approaches to address the limitations of current methods, offering insights into protein function and regulation. Through experimental analysis, we demonstrate the efficacy of our proposed methods in capturing long-range dependencies.

1 Introduction

Proteins are biomacromolecules that serve as essential components within cells and play critical roles in nearly every biological process, such as catalyzing metabolic reactions and transporting molecules. Protein functions determine health outcomes and the progression of diseases, hence predicting the functional properties of proteins is vital for developing new drug therapies. Protein design has emerged as an integral aspect of pharmaceutical research, seeking to better understand the design principles that form a basis for the structure and functions of proteins. This would enable the discovery of proteins with properties that are key for therapeutic and technological applications. Graph Neural Networks (GNNs) [14, 2] have emerged as a powerful tool for learning structural representations of proteins and biomolecules [15, 5]. Despite their success, GNNs exhibit clear limitations when confronted with long-range dependencies due to the phenomena known as **oversquashing** and **oversmoothing** [20] which diminish the expressive power of graph-based architectures. Such dependencies are fundamental aspects of proteins' structural and functional complexity, as inferred from their residue interactions. These interactions play pivotal roles in stabilizing tertiary structures, facilitating ligand binding, and orchestrating allosteric regulation [9, 12, 8]. Understanding long-range dependencies is crucial for deciphering protein folding mechanisms, predicting protein structures from sequences, and designing novel therapeutics targeting protein-protein interactions. Therefore, it is crucial to account for such dependencies when modeling proteins with GNNs, rising the need for more expressive architectures that would account for distant interactions.

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.

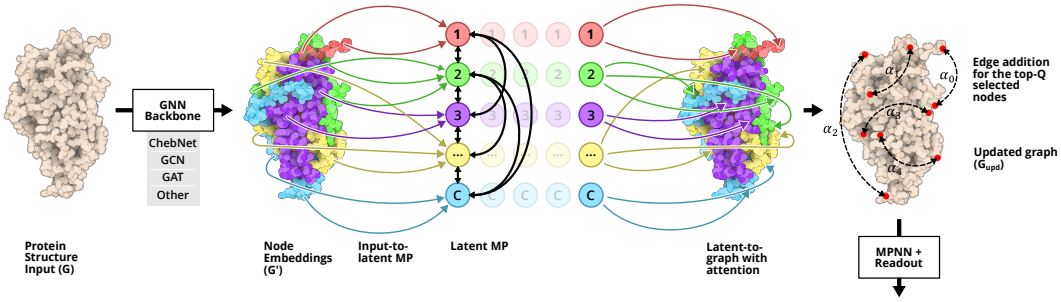

Figure 1: Pipeline of our rewiring framework with the different steps detailed in Section 3

To tackle the aforementioned limitations of GNNs, a set of approaches has been proposed, mainly through the use of virtual nodes to reduce the commute time between any two given nodes [3], or by changing the graph topology to allow for a better flow of information by optimizing some properties related to graph bottlenecks: this is known as **graph rewiring** which has been recently investigated [16, 11, 1]. Despite the aforementioned recent advances in tackling long-range dependencies in graphs, those are still under-explored for learning on protein structures. In this work we shed light on the power of long-range techniques in representing protein structures for diverse tasks at the global (protein) and local (residue) levels. More specifically, we propose a GNN-based rewiring scheme that defines a set of trainable latent nodes to cover different regions of the protein. Our model then uses the latent nodes as mediators to rewire the graph by attending distant nodes through attention-based edge addition.

2 Background

Graph Neural Networks have become a key component in computational biology due to their ability to represent complex molecular surfaces and learn useful interactions among the atoms in those systems. A notable use-case is protein function prediction. In [7], Gliborijevic et.al present a graph-based architecture that takes as input a protein structure and a sequence from a pre-trained language model. The model predicts the function of the protein and the key residues in the sequence for that function. Another important application is protein structure comparison which is crucial for structural homology discovery and other downstream structure-based analysis. For this purpose, GraSR was introduced in [19] as a graph contrastive model to better learn global and local geometric features of residues. Recent studies on molecular graphs discuss the importance of combining local semantics carrying potentially critical information about graph substructures with graph-level features summarizing its global topology [18, 13].

3 Graph neural networks for long-range dependencies in proteins

Consider a protein represented as a graph $G = (X, A)$ whereby $X \in \mathbb{R}^{n \times d}$ is the node feature matrix and $A \in \mathbb{R}^{n \times n}$ is the adjacency matrix. We provide a novel use-case based on attention-based rewiring of protein structures and show its ability to boost the performance on graphs with long-range dependencies. This approach is motivated by the fact that edge addition in general results in an increased spectral gap and hence reduced bottlenecks in the network [4]. Adapting a pragmatic approach on which edges to add is crucial to prevent oversmoothing by adding too many unnecessary connections. We summarize the main steps of our architecture shown in Figure 1.

Step 1: Intra-atoms message passing. We begin by covering the local neighborhoods of the protein, so we update the node representations based on their surrounding neighbors through a graph convolution operation Θ_{conv} resulting in new atom embeddings $h'_{loc} \in \mathbb{R}^{n \times f}$.

Step 2: Protein-to-Latent Nodes message passing. We propose to fuse the local graph information into a compressed global representation through message passing between the local embeddings h'_{loc} and a set of c trainable latent nodes (LN) forming a graph $G_c \in \mathbb{R}^{c \times f}$ such that $c << n$. Given the

embedding $\{h'_{loc}\}_i$ for each node i in G and the initial latent embedding z_q for a node $q \in [1, c]$, we perform an input-to-latent message passing:

$$z_q^{(l+1)} = M_l(z_q^l, \sum_{j \in \mathcal{N}(q)} AGG_l(z_q^l, h_j^l)) \tag{1}$$

In this scenario, the input graph nodes in G and the latent nodes are considered to be part of a bipartite graph. Each node in G is mapped to a latent node q, hence two or more nodes in G can be connected to the same latent node. A total of $|E| = n$ edges need to be added, alleviating some computational cost in comparison with a Transformer where each node attends all other nodes. Instead, we use a Graph Attention Network (GAT) [17] to aggregate messages towards the latent components.

Step 3: Intra-LN message passing. The latent nodes then exchange information through fully connected message passing i.e each latent node attends all other C nodes as described below:

$$z_q^{(l+1)} = M_l(z_q^l, \sum_{j \in [1,c]} AGG_l(z_q^l, z_j^l)) \tag{2}$$

This step ensures that latent nodes covering different ranges of the protein have interacted, providing a solid global operator for protein representation.

Step 4: Attention-based rewiring through latent nodes. This methodology is based on the idea that distant nodes need to communicate for a more effective global representation at the graph-level, hence it uses attention as metric to quantify the importance of the interaction of two distant nodes using the latent nodes as a mediator. In contrast to previous methods where rewiring is based on pre-processed measures [16, 6], we perform rewiring in an end-to-end fashion in analogy to more recent work [1, 10] . To proceed, we use an additional GAT to perform message passing from the updated latent nodes back to the input graph nodes $\{h'_{loc}\}_i$ as described in Eq.3. At this stage, we select the the top Q nodes with the highest attention scores for edge addition among them. Finally, we perform message passing on the updated graph using the new adjacency matrix $G_{rew} = (X, A')$ s.t $A' = A + Q_{top}$. A pseudo-code is given in Algorithm 1.

$$h_{loc}^{(l+1)} = N_l(h_{loc}^l, \sum_{j \in \mathcal{N}(h)} \alpha \times AGG_l(h_{loc}^l, z_j^l)) \tag{3}$$

Readout: For the final classification task, we combine the information from the global summation of latent node embeddings z_{out} and that of the rewired graph embeddings h_{rew} which are given by Eq.4. The final step consists of feeding the embeddings to a Multi-layer Perceptron (MLP) to get the readout of the task. We find experimentally that combining the embeddings h_{rew} from the rewired graph with the latent node embeddings scaled by a hyperparameter λ provides the best performance as show in 5

$$h_{rew} = P_l(h_{loc}^{l+1}, \sum_{j \in \overline{\mathcal{N}}(h)} Top_Q(\alpha) \times AGG_l(h_{loc}^{l+1}, h_j^l)) \tag{4}$$

$$Readout = MLP(h_{rew} + \lambda\ z_{out})\ ;\ z_{out} = \sum_{i=1}^{q} z_i \tag{5}$$

4 Experiments

The experiments in this study were conducted on several datasets to evaluate the model's performance across different tasks. For protein 3D structures, the model was validated on the EnzymeClass task (predicting reaction types from the Enzyme Commission database) and binding site identification using the PDBBind2020 dataset. Additionally, the model was tested on peptide datasets from the Long Range Graph Benchmark (LRGB), focusing on graph classification and regression, where amino acids were represented as nodes in peptide structures.

5 Results

Results on different protein-level and atom-level tasks are summarized in Tables 1 and 2, respectively. It is shown as expected that adding multiple latent nodes to cover different regions of the proteins boosts the performance relative to using the backbone alone. Further details on training is found in Appendix A. For the PDBBind dataset, we disregard the GCN-latent architecture given that as a node classification task, and in contrast to the graph-level ones, we do not directly use the latent nodes in the classifier, but rather only the messages propagating back from them.

Attention-based rewiring advantage: We highlight the improvement obtained by the rewiring framework when using the GCN backbone. On all the benchmark datasets, GCN-Rewire performs better than both the GCN backbone and the GCN model with multiple virtual nodes. On the Peptide dataset, it surpasses other rewiring frameworks such as FOSR and LASER in addition the Graph Transformers GPS and Exphormer. The advantage is especially shown on the binding site detection task in the PDBBind dataset, which is the only atomic-level task. This can be explained by the success of this method in attending distant residues whose communication potentially determines the overall protein's function. By doing so, it provides a solid combination of local and distant neighboring features on the one hand through both features of h_{local} and h_{rew} and the global features on the other hand through the latent node features z_q. We evaluated the model under different parameters and it is found empirically that the best performance saturates around $K = 6$ latent nodes and $Q = 8$ newly rewired nodes.

Model	**Peptide-function**	**Peptide-structure**
	Test AP ↑	Test MAE ↓
GCN	0.5930 ± 0.0023	0.3496 ± 0.0013
GPS Transformer	0.6535 ± 0.0041	**0.2500 ± 0.0005**
Exphormer	0.6527 ± 0.0043	-
FOSR	0.4629 ± 0.0071	0.3078 ± 0.0026
LASER	0.6447 ± 0.0033	0.3151 ± 0.0006
Transformer + PE	0.6326 ± 0.0126	**0.2529 ± 0.0016**
GCN-Latent	0.6211 ± 0.0059	0.2723 ± 0.0040
GCN-Rewire	**0.6670 ± 0.0024**	**0.2660 ± 0.0043**

Table 1: Performance of our rewiring framework on the LRGB datasets against numerous baselines.

Model	**EnzymesClass**	**PDBBind**
Type	Graph-level	Node-level
	Test Acc ↑	Test AUROC ↑
GCN	73.33 ± 1.06	62.65 ± 0.13
GCN-Latent	74.22 ± 0.50	N/A
GCN-Rewire	**75.80 ± 0.92**	**66.53 ± 0.34**

Table 2: Performance of our rewiring framework on protein structure datasets.

6 Conclusion

In this work, we have proposed a unified framework to alleviate the phenomenon of oversquashing that GNNs exhibit on proteins when dealing with long-range dependencies. The main components of the framework include the use of latent nodes that cover different regions of proteins and a novel use-case these latent nodes as a mediator for rewiring. We show that latent nodes can enhance the performance on a given GNN backbone for the datasets under consideration. We also evaluate our extension on the same dataset and show the additional boost it provides. Future work can make use of this methodology to further extend its expressive power by trying more pragmatic approaches to define the input to latent connections. It could also be interesting to explore the latent space exhibited by the latent nodes for interpretability.

References

[1] Adrián Arnaiz-Rodríguez, Ahmed Begga, Francisco Escolano, and Nuria M Oliver. DiffWire: Inductive Graph Rewiring via the Lovász Bound. In Bastian Rieck and Razvan Pascanu, editors, *Proceedings of the First Learning on Graphs Conference*, volume 198 of *Proceedings of Machine Learning Research*, pages 15:1–15:27. PMLR, 09–12 Dec 2022.

[2] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203*, 2013.

[3] Chen Cai, Truong Son Hy, Rose Yu, and Yusu Wang. On the connection between mpnn and graph transformer. *International Conference on Machine Learning*, 2023.

[4] Hau Chan and Leman Akoglu. Optimizing network robustness by edge rewiring: a general framework. *Data Mining and Knowledge Discovery*, 30:1395–1425, 2016.

[5] Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*, 2020.

[6] Jhony H Giraldo, Konstantinos Skianis, Thierry Bouwmans, and Fragkiskos D Malliaros. On the trade-off between over-smoothing and over-squashing in deep graph neural networks. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 566–576, 2023.

[7] Vladimir Gligorijević, P Douglas Renfrew, Tomasz Kosciolek, Julia Koehler Leman, Daniel Berenberg, Tommi Vatanen, Chris Chandler, Bryn C Taylor, Ian M Fisk, Hera Vlamakis, et al. Structure-based protein function prediction using graph convolutional networks. *Nature communications*, 12(1):3168, 2021.

[8] M Michael Gromiha and S Selvaraj. Comparison between long-range interactions and contact order in determining the folding rate of two-state proteins: application of long-range order to folding rate prediction. *Journal of molecular biology*, 310(1):27–32, 2001.

[9] M Michael Gromiha and Samuel Selvaraj. Importance of long-range interactions in protein folding. *Biophysical chemistry*, 77(1):49–68, 1999.

[10] Benjamin Gutteridge, Xiaowen Dong, Michael M Bronstein, and Francesco Di Giovanni. Drew: Dynamically rewired message passing with delay. In *International Conference on Machine Learning*, pages 12252–12267. PMLR, 2023.

[11] Kedar Karhadkar, Pradeep Kr. Banerjee, and Guido Montúfar. Fosr: First-order spectral rewiring for addressing oversquashing in gnns. In *The Eleventh International Conference on Learning Representations, ICLR*, 2023.

[12] Daisuke Kihara. The effect of long-range interactions on the secondary structure formation of proteins. *Protein Science*, 14(8):1955–1963, 2005.

[13] Xiao Luo, Wei Ju, Meng Qu, Yiyang Gu, Chong Chen, Minghua Deng, Xian-Sheng Hua, and Ming Zhang. Clear: Cluster-enhanced contrast for self-supervised graph representation learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[14] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008.

[15] Kristof T Schütt, Huziel E Sauceda, P-J Kindermans, Alexandre Tkatchenko, and K-R Müller. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24), 2018.

[16] Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M Bronstein. Understanding over-squashing and bottlenecks on graphs via curvature. *arXiv preprint arXiv:2111.14522*, 2021.

[17] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. *International Conference on Learning Representations*, 2018.

[18] Hanchen Wang, Jean Kaddour, Shengchao Liu, Jian Tang, Joan Lasenby, and Qi Liu. Evaluating self-supervised learning for molecular graph embeddings. *arXiv preprint arXiv:2206.08005*, 2022.

[19] Chunqiu Xia, Shi-Hao Feng, Ying Xia, Xiaoyong Pan, and Hong-Bin Shen. Fast protein structure comparison through effective representation learning with contrastive graph neural networks. *PLoS computational biology*, 18(3):e1009986, 2022.

[20] Lingxiao Zhao and Leman Akoglu. Pairnorm: Tackling oversmoothing in {gnn}s. In *International Conference on Learning Representations*, 2020.

A Dataset Summary

We report the properties of different datasets in the Table below. For training on all datasets, we used the AdamW optimizer with the learning rates tabulated below. All experiments are done on an NVIDIA Titan-RTX GPU and we report an average of 3 runs.

Dataset	# of graphs	Loss Function	hidden dim	Learning rate
Peptide-func	15,535	Cross-Entropy	300	0.0001
Peptide-struc	15,535	MSE	300	0.0001
PDBBind	2839	Cross-Entropy	150	0.0005
EnzymesClass	15,603	Cross-Entropy	150	0.0005
Structural Similarity	994	MSE	150	0.0005

Table 3: Dataset description

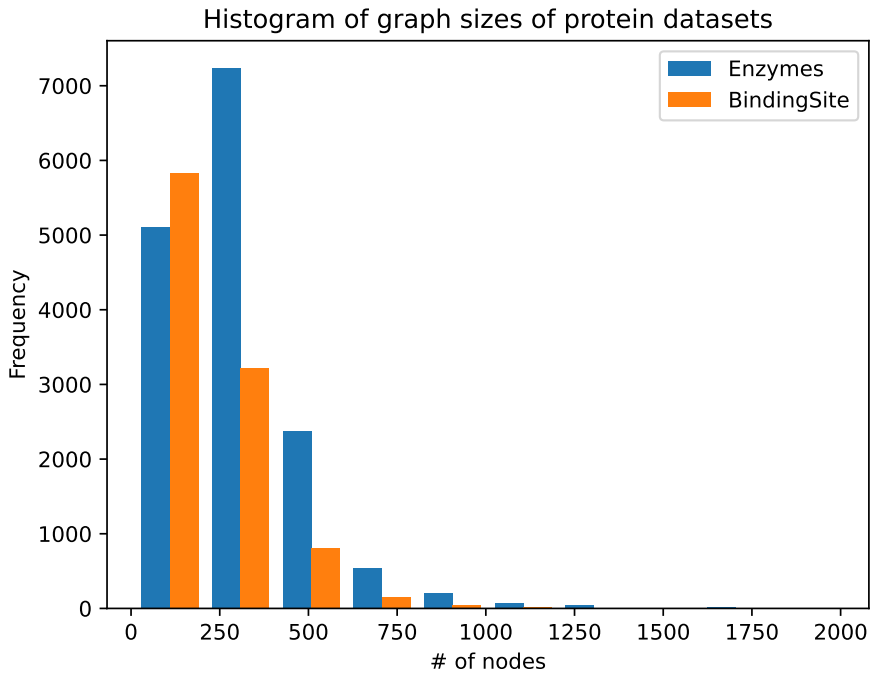

Figure 2: Number of nodes across different ProteinShake Datasets

B Definitions

Effective resistance (ER) in a graph measures the resistance between two nodes when an electrical current is passed through the edges. It quantifies how well-connected or isolated the nodes are and is commonly used in network analysis to assess the flow of information or current within the graph.

Commute time (CT) in a graph represents the expected time it takes for a random walk or particle to travel between two nodes, starting from one and reaching the other. It is a measure of the efficiency of traversal within the graph and finds applications in various fields, such as computer science and transportation planning.

Algorithm 1 Attention-based rewiring through latent nodes

Input: Graph $G = (X, A)$
Initialize: $\theta \leftarrow \theta_0, \phi \leftarrow \phi_0, V_c \leftarrow rand(C, F)$
repeat
  $G_{i+1} \leftarrow \mathrm{GNN}_\theta(G_i)$ *Backbone graph convolution*
until i==Z
$V_c \leftarrow G_Z$
$\phi(V_c) \leftarrow V_c$ *Fully-connected latent MP*
$G_Z \leftarrow \phi(V_c); \alpha \in \mathbb{R}^E$ *GAT from latent to input*
$E_{att} \leftarrow max_\alpha^Q$ *Select top-Q attention values*
$E_{upd} \leftarrow concat(E_G, E_{upd})$ *Update adjacency matrix*
repeat
  $G_{j+1} \leftarrow \mathrm{GNN}_\gamma(G_Z)$ *updated GCN*
until j==2