

Project Proposal: Implementing Graph Attention Networks

Valle, Ruy
rrv4@illinois.edu

Soni, Kushagra
soni14@illinois.edu

March 19, 2023

1 Introduction

Protein interactions play a crucial role in various biological processes, such as signal transduction, metabolism, and immune response. Identifying protein interactions can provide valuable insights into these processes, and can also lead to the discovery of new drug targets. However, experimental methods for identifying protein interactions are often time-consuming and expensive. In recent years, there has been increasing interest in using machine learning methods to predict protein interactions from protein sequences and structures. In this project, we propose to implement Graph Attention Networks (GAT) for predicting protein interactions. We will try to replicate the results which were received by the original authors of Graph Attention Networks [1]

2 Objectives

The main objective of this project is to develop a machine learning model that can accurately predict protein interactions. Specifically, we aim to achieve the following objectives:

- Develop a GAT model that can take protein sequences and structures as input and predict protein interactions.
- Evaluate the performance of the GAT model on benchmark datasets for protein interaction prediction.
- Compare the performance of the GAT model with other state-of-the-art machine learning models for protein interaction prediction.

3 Methodology

The GAT model will be implemented using Python and the PyTorch deep learning library. The input to the model will be a graph where nodes represent amino acids and edges represent physical interactions between the amino acids. The graph will be constructed from the protein sequence and structure using established methods. The GAT model will use attention mechanisms to weight the importance of neighboring nodes when aggregating information for each node. The model will be trained using cross-entropy loss and optimized using the Adam optimizer.

4 Evaluation

The performance of the GAT model will be evaluated on benchmark datasets for protein interaction prediction [2], such as the Protein-Protein Interaction Extraction (PPI) dataset that consists of graphs corresponding to different human tissues. The dataset contains 20 graphs for training, 2 for validation and 2 for testing. To construct the graphs, we will use the preprocessed data provided by [3]. The evaluation metrics will include precision, recall and F1-score. The performance of the GAT model will be compared with other state-of-the-art machine learning models for protein interaction prediction, such as the Graph Convolutional Networks (GCN) [4] and the Deep Graph Infomax (DGI) [5].

5 Timeline

- Week 1-2: Literature review and data collection
- Week 3-4: Graph construction and feature extraction
- Week 5-7: GAT model implementation and optimization
- Week 8-9: Model evaluation and comparison with other models
- Week 10-11: Results analysis and report writing
- Week 12: Final presentation and submission

6 Conclusion

The successful implementation of the GAT model for predicting protein interactions will have important implications for the field of computational biology. The model could be used to predict protein interactions for new proteins and provide valuable insights into biological processes. The proposed project will also provide valuable experience in implementing deep learning models for graph-structured data.

References

- [1] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations (ICLR)*, 2018.
- [2] Marinka Zitnik, Monica Agrawal, and Jure Leskovec. Predicting combinatorial drug synergy in multi-dimensional cancer immunotherapies with genomics and distributed computing. *Pacific Symposium on Biocomputing*, 22:410–421, 2017.
- [3] William L. Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- [4] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.
- [5] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Deep graph infomax. In *International Conference on Learning Representations*, 2019.