

Resumo

The proposed thesis project aims to investigate graph-based approaches, with a particular focus on Graph Neural Networks (GNNs), for predicting disease-associated genes within human Gene Regulatory Networks (GRNs). Complex diseases often arise from perturbations in gene regulation, making the identification of disease genes essential for diagnostics, therapeutics, and precision medicine. Traditional methods struggle to capture the heterogeneous and interconnected nature of biological networks, whereas GNNs provide a powerful framework to learn from graph structure and features. Beyond predictive modeling, the project will also explore classical graph analysis techniques — such as centrality, community detection, and visualization — to complement the predictions with interpretable insights into the structure and dynamics of GRNs. By combining GNN-based prediction with network exploration, this project seeks to deliver both accurate predictive tools and a deeper understanding of biological regulatory networks.

Objetivos

The main objectives of this thesis are:

- Develop and evaluate GNN architectures for predicting disease genes from GRNs.
- Explore network analysis techniques (e.g., centrality, motifs, clustering, visualization) to identify candidate disease genes and network structures.
- Integrate predictive and exploratory approaches to enhance interpretability and biological relevance.
- Benchmark predictive models against baseline graph and machine learning methods.

Plan

1. Conduct a literature review on disease gene prediction, GRNs, GNNs in bioinformatics, and graph exploration methods.
2. Collect and preprocess GRN data (e.g., ENCODE, TRRUST) and disease-gene associations (e.g., DisGeNET).
3. Perform network exploration using centrality measures, community detection, and graph visualization.
4. Design, implement, and train GNN models (e.g., GCN, GAT, GraphSAGE) for disease gene prediction.
5. Evaluate predictive and exploratory results using metrics such as accuracy, F1-score, ROC-AUC, and biological case studies.
6. Combine insights from GNN models and network analysis to produce interpretable results.
7. Deliverables: a working GNN pipeline, exploratory analysis toolkit, thesis manuscript, and a potential scientific paper.

Inovação

The innovation of this project lies in combining predictive deep learning methods (GNNs) with classical graph/network exploration techniques. This dual perspective provides both predictive accuracy and interpretability, offering novel insights into disease gene discovery and the structural properties of gene regulatory networks.

Referências

1. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P. S. (2021): A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24.
2. Zhang, Xiao-Meng, Cui, Yuzhong (2021): Graph Neural Networks and Their Current Applications in Bioinformatics. *Frontiers in Genetics*, 12, 690049.
3. Karlebach, G., Shamir, R. (2008): Modelling and Analysis of Gene Regulatory Networks. *Nature Reviews Molecular Cell Biology*, 9, 770–780.
4. Chan, T. E., Stumpf, M. P. H., Babbie, A. C. (2017): Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. *Cell Systems*, 5(3), 251–267.
5. Chatzianastasis, S., Korkontzelos, I., Kanavos, A. (2023): Explainable Multilayer Graph Neural Networks for Cancer Gene Prediction. *Briefings in Bioinformatics*, 24(2), bbac588.