Detection and Characterization of Evolutionary Patterns Through Graph Neural Networks

Mariana Meireles[1] and Carsten Fortmann-Grote[1]

1 Motivation

Graph Neural Networks (GNNs), a branch of deep learning designed for non-Euclidean data, have emerged as a powerful tool for analyzing interconnected information.

The rapid accumulation of biological network data has opened new avenues for understanding complex biological systems. GNNs are particularly well-suited for analyzing these intricate networks, as they are able to capture the inherent non-linear relationships and hierarchical structures present in biological data [1].

This work focuses on applying GNNs to detect and characterize evolutionary patterns in *Pseudomonas fluorescens* SBW25, a common model organism in microbial evolution [2].

2 Methods

Public graph databases Starting from a de-novo hybrid assembly and manual genome annotation [3], we curated a graph database [4] containing all known genomic features, their products and functions as well as cross references to domain specific databases, such as the protein database Uniprot, the KEGG database of metabolic pathways, and orthology databases. In this way, our graph is integrated in a public knowledge graph enabling to enrich our graph representation and improve the contextual understanding for the GNN algorithms of evolutionary processes across different biological domains and scales.

Node level, edge level, graph level The next steps of our work will involve exploring different approaches leveraging GNN architectures.

Considering the three different levels in which a graph can be analysed - the node level, link or edge level and the graph level - we propose the following approaches to capture the complexities of *Pseudomonas fluorescens*' evolutionary processes: The node level classification could help us classify genetic elements based on their evolutionary significance, potentially identifying genes or regulatory regions that are hotspots for adaptive changes. Edge prediction could be employed in forecasting potential genetic interactions and evolutionary relationships, possibly predicting future adaptive mutations or gene transfer events . The graph level of analysis can be used to generate synthetic graph structures, which may allow us to predict evolutionary trajectories and test hypotheses about adaptive processes.

References

[1] Xiao-Meng Zhang et al. Graph neural networks and their current applications in bioinformatics. *Frontiers in Genetics*, 12:690049, 2021.

[2] Paul B Rainey and Michael Travisano. Adaptive radiation in a heterogeneous environment. 394(6688):69–72.

[3] Carsten Fortmann-Grote, Eric Hugoson, Joanna Summers, Loukas Theodosiou, and Paul B. Rainey. Genome Update for Pseudomonas fluorescens Isolate SBW25. 0(0):e00637–22.

[4] Carsten Fortmann-Grote and Paul B. Rainey. From genome annotation to knowledge graph: The case of pseudomonas fluorescens sbw25. In *30th Conference on Intelligent Systems for Molecular Biology 2022*, Bio-Ontologies COSI. International Society for Computational Biology, Zenodo.

Extended Abstract.

[1] Max Planck Institute for Evolutionary Biology; | July 24, 2024