Assessing the Impact of Graph Embedding Refinements on IDP Classification

Nenad Vilendečić , Milana Grbić, Dragan Matić,

Faculty of Natural Sciences and Mathematics, University of Banja Luka

Intrinsically disordered proteins (IDPs) play crucial roles in various biological processes, yet their identification remains a challenge in bioinformatics. Traditional IDP classification methods primarily rely on sequence-based features or secondary structural information. However, recent studies suggest that integrating data from protein-protein interaction (PPI) networks can improve IDP prediction by leveraging the complex interdependencies between proteins.

Building on our previous research, this study aims to enhance IDP classification by refining graph embeddings and integrating their attributes with additional biological features. Specifically, we propose an improved training strategy for generating node embeddings within the PPI network, capturing more informative structural representations. Additionally, we will explore the combination of these enhanced embeddings with physicochemical and biochemical attributes derived from protein sequences.

For classification, we will employ machine learning models such as Support Vector Classifier (SVC) and Random Forest (RF), evaluating their ability to distinguish IDPs from non-IDPs. A comprehensive feature selection process will be conducted to determine the most informative attributes for classification.

By improving the quality of graph embeddings and integrating diverse biological features, we aim to develop a more robust IDP prediction framework. The anticipated findings could lead to a deeper understanding of IDP functionality and their interactions within cellular processes, contributing to more accurate bioinformatics models for protein disorder prediction.