Predicting Intrinsically Disordered Proteins through Combined Sequence and Network-Based Features

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 methods of IDP classification primarily rely on sequence-based features or secondary structural information. However, recent advancements suggest that integrating data from protein-protein interaction (PPI) networks can enhance the accuracy of IDP prediction by leveraging the complex interdependencies between proteins. This study aims to develop a binary classification model to predict whether a given protein is intrinsically disordered (IDP) or not, using a set of yeast proteins and their associated PPI network.

The proposed approach integrates features from protein sequences, including physicochemical and biochemical properties, with advanced graph embeddings derived from the node2vec algorithm. Building on last year’s work, this study explores an improved version of node2vec embeddings to capture more sophisticated structural information within the PPI network. Additionally, a comprehensive feature selection process will be employed to identify the most relevant attributes that contribute to the classification task.

For classification, we will utilize well-established machine learning models, such as Support Vector Classifier (SVC) and Random Forest (RF), and evaluate their performance in distinguishing IDPs from non-IDPs. By incorporating both sequence-based and network-based features, we aim to improve the predictive accuracy and offer insights into the role of protein interactions in the disordered state.

While the results of this approach are still under investigation, we anticipate that combining network embeddings with protein sequence features will provide a more robust framework for IDP prediction. The success of this model could lead to a better understanding of IDP functionality and their interactions within cellular processes.