Multi-layer Network Analysis for Herb-Symptom-Protein Interaction Prediction

Introduction:

The prediction of interactions between herbs, symptoms, and proteins plays a crucial role in various domains such as drug discovery, traditional medicine, and biomedical research. In this project, we propose a multi-layer network analysis approach to predict these interactions. By leveraging graph theory concepts and neural network modeling, we aim to uncover hidden patterns and relationships within complex biological systems.

Data Collection and Preprocessing:

Our project begins with the collection of diverse datasets containing information about herb, symptom, protein, herb-protein, and symptom-protein interactions. These datasets are loaded into the environment

- (datahs.csv)
- (dataht.csv)
- (datass.csv)
- (datatt.csv)
- (datahh.csv)

To construct a comprehensive representation of the interaction network, we utilize the NetworkX library in Python. This allows us to create a multi-layer graph where nodes represent herbs, symptoms, and proteins, and edges denote interactions within and between layers.

Feature Extraction:

An essential step in our approach is the extraction of meaningful features from the multi-layer graph. To improve computational efficiency and focus on relevant interactions, we apply thresholding techniques to simplify the graph. Symptom nodes are reduced based on degree and betweenness centrality thresholds, resulting in a more manageable representation of the network.

For each pair of nodes in the herb-symptom and herb-protein layers, we calculate a variety of features including Common Neighbors (CN), Jaccard Coefficient (JC), Shortest Path Length (SL), and others (17 local indices and 2 global indices). These features serve as valuable inputs for our predictive model, capturing the intricate relationships between herbs, symptoms, and proteins.

Data are saved in ‘build/hs_result.csv’ and ‘build/ht_result.csv’

Model Training:

With the extracted features in hand, we merge the datasets for herb-symptom and herb-protein interactions into a final dataset (**final_df**). The labels for our prediction task are derived from this dataset, specifically the target variable **final_df['t']**, representing the protein interactions.

Our X is final_df.drop('t' , axis=1) with one-hot 'h' column. [1255968 rows × 40 columns]
and y is final_df['t']

Utilizing the TensorFlow library, we define a neural network model architecture suitable for our prediction task. The model consists of multiple dense layers with varying numbers of neurons, designed to capture complex patterns in the input data. We specify the model parameters, including the learning rate, batch size, and number of epochs, to optimize the training process.

Results and Evaluation:

After training the neural network model, we evaluate its performance on the prediction task. The model achieves an accuracy of 73% on test data, demonstrating its ability to effectively predict interactions between herbs and proteins.

Model is saved in 'build/73model.h5' and 'build/73model.keras'

Conclusion:

In conclusion, our project showcases the potential of multi-layer network analysis in predicting interactions between herbs, symptoms, and proteins. By leveraging graph-based features and neural network modeling, we can gain valuable insights into the complex interplay within biological networks. Future work may focus on further optimizing the model and exploring additional features to enhance predictive performance and advance our understanding of biological systems.