



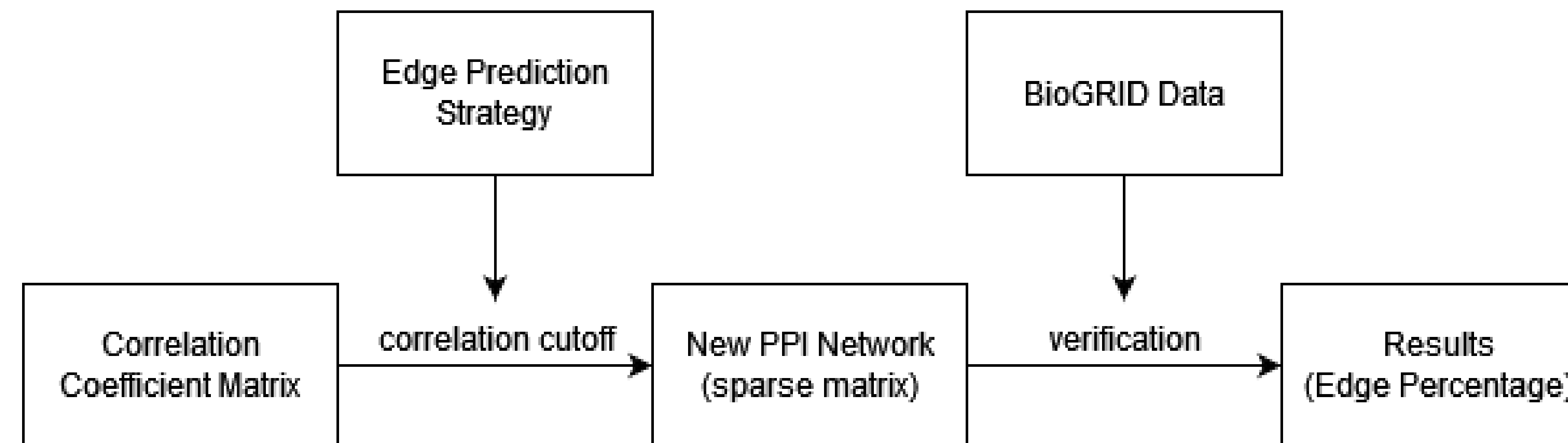
Project Summary

The project's purpose is to explore computational strategies to improve the prediction accuracy of edges in a biological network. More specifically, this project builds upon previous research on the protein-protein interaction (PPI) network of the yeast species *Saccharomyces cerevisiae*. To accomplish the project's goals, various edge prediction strategies were explored and their prediction accuracies were validated against biological wet lab data.

User Interface

Users have the option of running the edge prediction software interactively, with all image results displayed inline, or via a command-line interface which outputs the results to image files.

Project Architecture



From a high-level perspective, the network edge prediction strategies are used to generate new PPI networks containing the predicted edges. The predictive accuracies of these new networks are then evaluated by comparing the networks with the BioGRID data to determine the percentage of all predicted network edges that are correct.

Edge Prediction Strategies

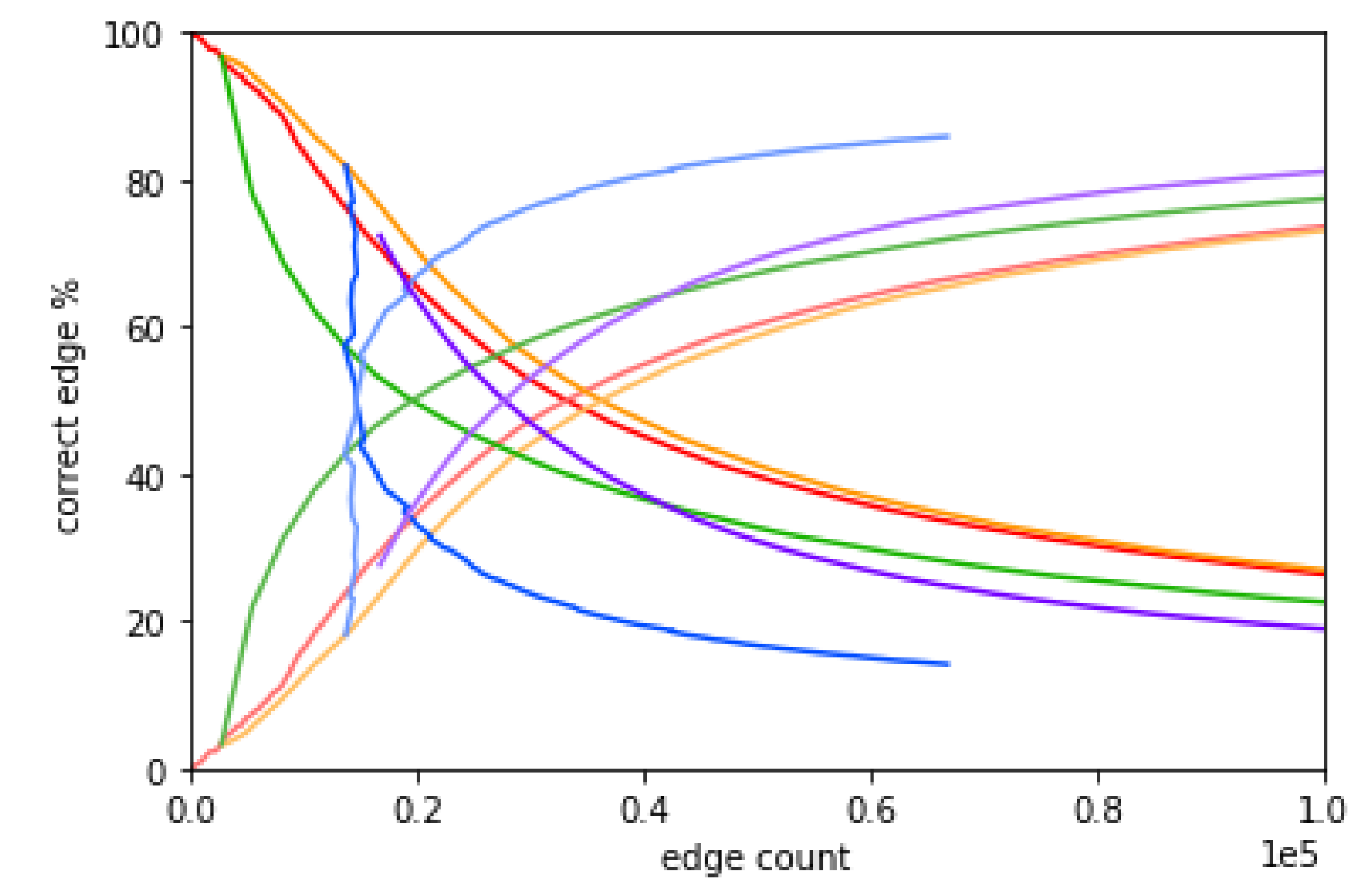
Edge prediction was performed on a matrix containing correlation coefficients for all edges in the PPI network; this correlation matrix was obtained from prior research. All strategies involved generating sparse matrices by applying a correlation cutoff to the correlation matrix; the strategies differ only in how the correlation cutoff is selected. In some cases, a single cutoff is applied to the entire correlation matrix. However, in other cases, the chosen cutoff is based on some property of the PPI network, such as the global average of node degrees.

BioGRID Verification

Prediction results were validated against data from BioGRID, a curated database of biological information including PPI data. The PPI data in BioGRID represents the best currently available information, so the efficacies of the different edge prediction strategies are evaluated on the percentage of edges in the generated networks that are also present in the BioGRID data.

Results

The BioGRID verification results for each of the prediction strategies were plotted as a function of the total edge count for the generated networks. This allowed for a more direct comparison of the strategies.



Plot of verified and unverified network edges for the various edge prediction strategies.

Conclusion

This project extends previous efforts to improve the quality of a PPI network using computational means. The obtained results reveal the relative efficacies of different edge prediction strategies and suggest potential future avenues of research into improving the quality and accuracy of biological networks.