

Predicting synthetic genetic interactions in yeast through machine-learning analysis of protein interaction data

Sri R. Paladugu¹, Shan Zhao^{1,2}, Animesh Ray¹, Alpan Raval^{1,3,*}

1. Keck Graduate Institute of Applied Life Sciences, Claremont, USA

2. Department of Physics, University of Rochester, Rochester, USA

3. School of Mathematical Sciences, Claremont Graduate University, Claremont, USA

*E-mail: araval@kgi.edu

Abstract

A synthetic genetic interaction (i.e., synthetic sick or lethal) is an extension of the single gene dispensability concept: two genes are said to participate in a synthetic genetic interaction if knocking out either one of the two genes alone will not result in any observable change in the phenotypic response (growth in this case) of the organism, but knocking out both the genes simultaneously would result in a slow growth or lethal phenotype. Identification of synthetic sick and lethal (SSL) gene pairs is important for determining functional relationships between genes and has important therapeutic implications. Such pairs have been most extensively identified in *Saccharomyces cerevisiae*, although even in yeast most synthetic lethal pairs have not yet been found due to the time and cost necessary for an exhaustive experimental determination of all SSL genes pairs [1, 2, 3]. Under the circumstances, computational methods for predicting SSL gene pair candidates are desirable. Our goal here is to explore the extent to which synthetic sick and lethal interactions can be computationally predicted using properties of protein interaction datasets alone. To efficiently guide the experimental discovery of synthetic genetic pairs, we have used support vector machines (SVM) for the classification of the data. The SVM is trained on a set of synthetic lethal and non-synthetic lethal pairs found from large scale genetic screens. Various graph-theoretical properties (local as well as global) of two proteins in a protein interaction network are fed as inputs to the SVM classifier, which is schematically represented in Figure 1. The output of the SVM classifier is a score corresponding to the propensity of a particular gene pair to be synthetically lethal. Most predictive inputs were determined by training the support vector machine separately using each of the inputs (Figure 2). Our results indicate that protein interaction network data can be used to identify synthetic lethal pairs with simultaneously high sensitivity and specificity (both close to 90%), suggesting a novel computational route towards elucidation of the synthetic lethal interactome. Our methods make use of a recent compendium of literature-curated data on protein-protein interactions and genetic interactions in yeast [4]. The methods may be extended to other interaction types and also to other organisms for which large scale protein interaction network data is available. On a different level, our results demonstrate that the protein interaction network embodies sufficient information for predicting genetic interaction - a surprising result because genetic interaction or lack thereof, do not necessarily imply a physical interaction between their encoded proteins.

References

- [1] Tong, A.H., et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* 294: 2364-2368.
- [2] Pan X., et al. (2004) A robust toolkit for functional profiling of the yeast genome. *Mol. Cell* 16: 487-496.
- [3] Wong, S.L., et al. (2004) Combining biological networks to predict genetic interactions. *Proc Natl Acad Sci U S A*, 101(44):15682-15687.
- [4] Reguly, T., et al. (2006) Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*. *Journal of Biology*, 5:11.

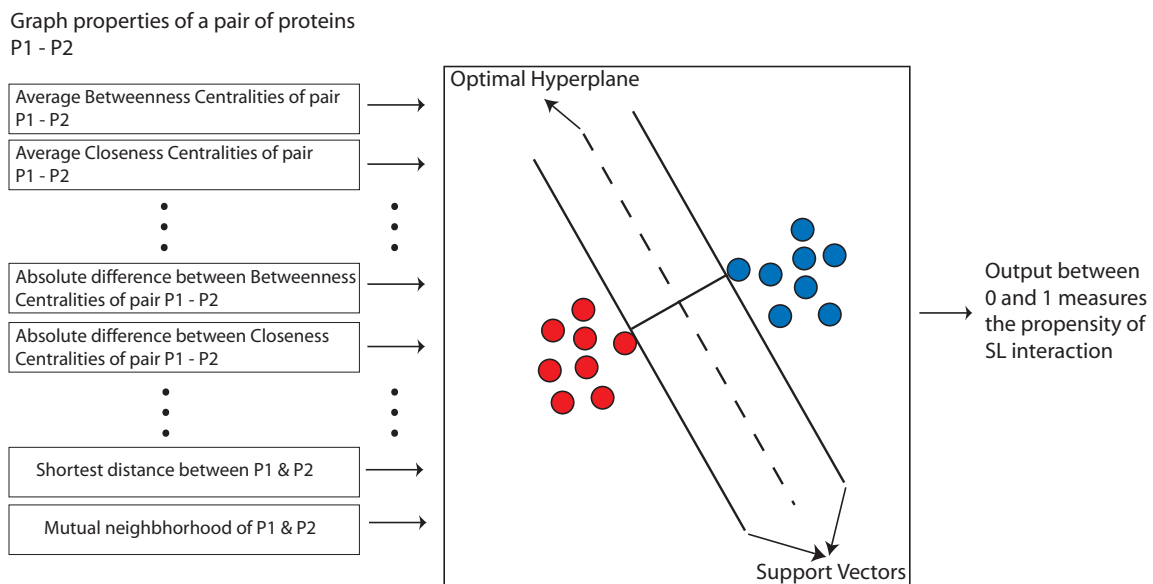


Figure 1: A schematic diagram representing the SVM classifier with various graph theoretic properties fed as inputs.

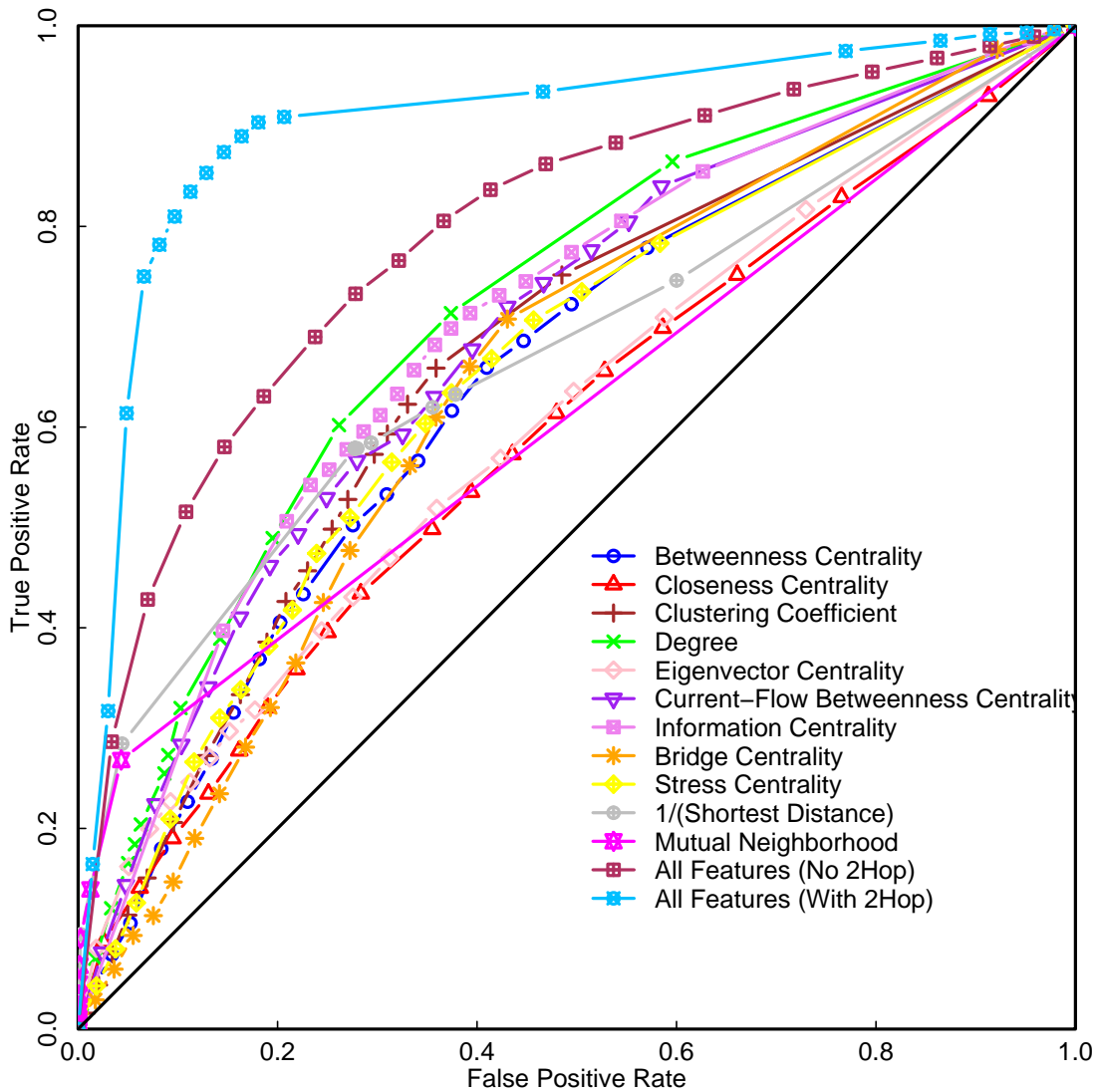


Figure 2: ROC curves for SVM classifiers trained on literature curated and high throughput data using single predictor variables. The diagonal line indicates random prediction. ROC curve for the SVM classifier trained using all the input features is also shown in the figure.