**Identifying Probable Protein-Protein Interactions Related to those of Asthma and Allergy Using the Diffusion Kernel**

Dominic K. Bett, and Ananda Mohan Mondal\*

*Department of Mathematics and Computer Science,*

*Claflin University, Orangeburg, SC 29115*

\*Corresponding Author: [amondal@claflin.edu](mailto:amondal@claflin.edu)

***Abstract* –** Many diseases including cancer, diabetes, and asthma occur due to over-expression or suppression of certain proteins. These proteins come to physical contact with each other due to biochemical events in a process called Protein-Protein Interaction (PPI). Studies have been done on how the PPI networks directly influence the development and progression of diseases. However, little focus has been placed on neighborhood proteins in the PPI network that do not physically interact with each other but have a higher likelihood to interact than the actual PPIs in the network. Identifying these missing PPIs would complete the network biomarker for a disease. In the present study, we seek to predict probable missing PPIs related to asthma and allergy.

PPI network for asthma and allergy used for present analysis is developed by overlaying the differentially expressed proteins on genome-wide PPI network. The PPI network for asthma and allergy is composed of 1,425 PPIs with 84 proteins. Genome-wide PPI data are obtained from STRING database and differentially expressed proteins are obtained from SABiosciences of Qiagen. In our proposed approach, protein interactions for asthma and allergy are represented in a two dimensional space called a Laplacian matrix. In this study, Laplacian matrix is an 84 by 84 square matrix, where an element is 1 if two proteins interact, otherwise 0. The diagonal value for a protein is the negative of number of interactions with the protein. The Laplacian matrix is then used to evaluate a Diffusion Kernel, which assigns similarity weights to all possible PPIs for 84 unique proteins associated with asthma and allergy. The PPIs that do not belong to the set of actual PPIs, but have greater kernel values than the actual interactions are predicted to be probable PPIs.

From our experiment, we determined a set of probable missing PPIs for asthma and allergy. The resulting prediction, together with the actual PPIs, enables us to establish more comprehensive network of proteins that cause asthma and allergy disease. In future, we will investigate the relationship between our proposed protein interaction network and their biological functionality.

*Keywords:- Laplacian Matrix; brute force method; Diffusion Kernel; PPI score; average degree; diffusion score.*

# Introduction

# Datasets Preparation

A set of Asthma and Allergy data is required in this study. Genome-wide PPI data are obtained from STRING database and differentially expressed proteins are obtained from SABiosciences of Qiagen. We considered the human genome belonging to Asthma and Allergy disease that consists of 1,425 scored PPIs with 84 proteins.

1. *Scored PPI Subnetwork*

PPI scores range from 150 to 999, 150 indicating minimum interaction, and 999 indicating maximum interaction. The PPIs are ordered according to their scores and Subnetworks generated. We considered networks containing PPIs with scores greater than 500, 600, 700, 800, and 900 and 950. Networks with higher PPI scores are likely to show more correlation. Hypothetically, higher PPI scores in a network increases the probability of missing PPIs.

1. *Average Degree*

The average degree of each network is calculated by dividing the number of PPIs by the number of proteins in the network. The average degree describes how connected the network is. More connected networks have higher degrees

# Methodology

1. *Laplacian Matrix*

# Results and discussion

# Conclusion and Future Remarks

##### Acknowledgment

##### References