VAWKUM Transactions on Computer Sciences
http://vfast.org/journals/index.php/VTCS@ 2022, ISSN(e): 2308-8168, ISSN(p): 2411-6335
Volume 10, Number 1, January-June, 2022 *pp:57-67*

Pathway-disease Association Prediction Based on Graph Regularized Logistic Matrix Factorization (PDA-GRLMF)

Ali Ghulam

Information Technology Centre, Sindh Agriculture University, Tandojam, Pakistan

Keywords: *Disease, Pathways, Bipartite, disease-genes, GRLMF.*

Journal Info:
Submitted:
June 15, 2022
Accepted:
September 16, 2022
Published:
September 20, 2022

Abstract Complex alterations to the cellular machinery occur as a result of diseases. There are distinctive patterns associated with a disease in the gene expression profile of the affected cells. As a result, these profiles can be used to extract additional biological information about an illness, which helps us better identify and evaluate disease risks. Human pathway-disease interaction research is a recurrent area of interest for the biomedical community. Finding the processes or connections between diseases and pathways can be aided by this association. This paper provides an overview of human pathway and human disease, with the accuracy of disease identification has been less than satisfactory. In predicting disease-pathway interactions, this study suggests a computer model. In this research study we proposed the Graph Regularized Logistic Matrix Factorization (GRLMF) method for pathway-disease association prediction. A cutting-edge computational model called the PDA-GRLMF disease-pathway association model can predict probable pathway-disease associations. The model can also assist pathologists in comprehending the relationships between disease-pathway linkages, therapies, and outcomes. In order to increase the association between disease variation and new molecular correlations between genetic mutations, we carried out a pathway-based investigation. On the basis of shared gene interactions among pathways-disease, we created a biological network, and then we used network analysis to try and understand how a disease constructed the pathway-pathway network and then disease-disease network. To merge the gathered biological data, which was based on the pair similarity of sequence expression weights, we employed the heterogeneous network of pathway-disease relationships. The ROC (AUC) score achieved for the best prediction results was 0.8018%, and the precision-recall curve had two classes. These findings suggest that our strategy outperforms previously suggested methods in terms of scientific performance. By contrasting them with established connections and conducting a literature search, we projected relationships between pathogen, DD, and disease-pathway.

***Correspondence Author Email Address:**
garahu@sau.edu.pk

1 Introduction

In recent years, researchers have become increasingly interested in regular computational biology and bioinformation research. Numerous studies have investigated research papers in computational biology and bioinformation that report notable gains over standard algorithms via competitions. In addition to gene cutting associations, we have reviewed lot other networks, such as those which are formed by disease-pathway and gene ontological annotations. A number of recent studies have examined the boundaries, weight is given to the characteristics of the form, reliability, and computerization experiments [1]. Over the years, an enormous amount of research has been carried out in an attempt to diseases affect the cellular machinery. Researchers have analyzed these data to search for a correlation between disease, disease pathway association, and gene-disease association [2]. Ample evidence [3] exists to support the hypothesis that structure using with proteins, shared genes with in cell pathways. Disease-pathway association prediction based on computational methods deserves more research attention. Lot of bioinformatician researchers are able to correctly predict the diagnosis based on the type of analysis that has been done. Few studies have investigated the impact of hypothesis about disease-pathway associations, arguing for possible explanations for our predictions.

The importance of finding disease-pathways association based on specific proteins related datasets, a particular disorder protein are highlighted by the discovery of these pathways as protein complexes in which the disorder is found. Computational methods assist in finding disease-disease, disease-gene, and disease-pathway connections through the use of disease-disease network, disease-gene network, and disease-pathway network. We have thus investigated new cutting-edge techniques for identifying diseases and discovered that those with broken pathways perform far worse. And so we can conclude that it's possible that not just the network's connectivity, but also its overall structure, may be required for detecting disease. Our approach, however, shows how new methods for larger network structures like small subgraphs could be developed.

Most scholars seem to agree that accurate prediction with the use of complex network study. However, there continues to be debate about patients with complex disorders, such as cancer with multiple pathways [4]. Giugno et al. [5] made rules of association between classification gene expression period. Many scientists conducted pathways to evaluate the predefined genes (i.e. pathways) as markers of disease [6, 7]. In integrating gene-set enhancement study with help vector machine, Kim etc. [8] introduced pathway-based markers. Lee et al. [9] identified key genes on pathway-grading pathways. The search space decreased significantly, and hypothesis testing was reduced, and correlation predictions were observed in a GWAS field of research. Pathway based methods for the study of the genomic experiment profiling of genes for the detection of moderate but clear disease changes in expression was first developed [10]. Our current knowledge of pathways, which is very far from complete and clear, indicates that disease-related proteins associate each other, and improved through the development of association networks. In this section, we focus on recent methods for pathway-based identification of disease links, in particular methods that integrate GWAS with other omic data [11]. This study is concerned with primary objectives are to study gene regulatory networks and apply these findings to network-based applications, such as protein structure prediction, gene disease prioritization, and a broad-association network-based genome study[12]. New insights into the etiology, classification, and common biological mechanisms of diseases may be found if disease markers can be linked to the underlying biology. There is a quantitative approach with systematic recognition of genetic disease, pathway association, and specific interaction linking diseases in order to study disease relationships. Learn more about how disease-pathway association [13]. This study attempts to establish the connection between gene and pathway, almost all of the above strategies are used to identify disease-related pathways. Additional elements, such as interactions, were later proposed as possible solutions. An important component of the pathway is the physical entities that exist within it (genes are defined as nodes in the network model). Another important feature of this network model is that the interactions of gene and protein interactions, as well as the dynamics of these interactions, are included [14, 15]. The evidence points to candidate genes. However [16], the role of genomic data is still poorly understood. Toward this end, the strategy is to target candidate genes based on their relatedness to previously identified disease genes. Because of this, there is a need for the integration

of multiple data sources, each of which holds information that is different.

2 Method and Datasets

2.1 Datasets collections

This study used pathway datasets from the Reactome pathway database and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database as its subjects. These open-source databases offer a computing environment for biological reactions in networks chosen to produce predetermined human biological pathways. Also, 30,3800 unknown association datasets that included pathways, illnesses, and connected genes were used in the current investigation. We used 72 disease and 116 pathways were identified once the data were standardized. The majority of the datasets extracted from the CTD database as shown in Table 1. The Online Mendelian Inheritance in Man (OMIM) database provided the disease datasets, which were then assessed based on disease similarity and CTD.

Table 1. Data Source

Data Source	Websites
OMIM	https://www.omim.org/downloads/
KEGG	https://www.genome.jp/kegg/pathway.html
Reactome	https://reactome.org/ CTD http://ctdbase.org/
MESH	https://www.ncbi.nlm.nih.gov/mesh

2.2 Pathway-Pathway Network

We looked at the results of recent research that have route information and pathways linked to genes. We conducted all analyses using node to node interaction than created pathway network which contain on node activity a pathway in the current study, while edges indicate the genes. It was discovered that the route and genes were positively correlated. If two pathways contain the same genes, they are linked. Before we analyze the data, it would be wise to pathway-pathway network as shown in Figure 1. A pathway is represented as a set of blue rectangular nodes, while edges connecting pathways are shown as sets of black lines. We used input datasets after pre-process and removed the similarity then CSV format, which we used. This method takes into account information about diseases and routes, as well as semantic similarities between pathways. We configure and fix the P to be equal to p1, p2,..., pn, where n is the number of pathways. Rows and columns are represented by N*N for the same dimension. Additional variables were derived from adjacency matrix Aij) can be described as a comparable matrix Simpi, pj matching to a beginning based on the comparation association between two routes nodes, and it is presented as follows:

$$A(i,j) = \begin{cases} \frac{1}{0} if sim(Pi, Pj)^{3q} & \text{Others} \end{cases} \quad (1)$$

2.3 Disease-Disease Network

The goal of the current study was to examine the connection between diseases and the associated genes, which were found in benchmark datasets from the CTD database. All known phenotypic and illness gene correlations can be explored in a single graph-theoretic framework using a network of disorders and disease genes linked by known disorder-gene associations, revealing the shared genetic ancestry of many diseases. A DD relationship network constructed around common genes has been made to identify every pair of diseases (Figure 3). For determining the similarity value, the common gene's importance was taken into consideration. For the graph network, we combined techniques for qualitative and quantitative research. A fresh approach with a modest intention of connotation complexity was also applied. As a result, managing the wide are research direction strategy takes a lot of effort and time. We looked at this

strategy's usability. We measured the shortest path between two nodes and estimated the characteristics of known illness nodes. A bipartite graph that consists of two separate groups of nodes was created. The first collection relates to all recognized genetic illnesses, whereas the second set relates to all recognized disease genes in the human genome. The most comprehensive and current database of all known disease genes and the disorders they cause is OMIM, notwithstanding the fact that this history adds some biases and the disease gene record is far from complete. We constructed the disease-causing bipartite network. (Center) Circles and rectangles represent disorders and disease genes, respectively, in a limited portion of OMIM-based disorder-disease gene relationships. If mutations in a disease gene cause the particular ailment, a link is established between the disorder and the disease gene. The size of a circle reflects how many genes are involved in the associated ailment, and the color reflects which disorder class the disease falls under. The disease-disease bipartite graph's projection, which shows that two disorders are related if the same gene is implicated in both (left). The number of genes linked to both diseases is inversely correlated with a link's width. In a DD network, we also examined the connections between disease nodes and edges as well as the disease network relationships. The undirected graph G = (d, d) was taken into consideration. Here, d stands for the groups of diseases (nodes), and the edges between them are determined by how similar they are. The number of diseases that the two genes are frequently linked to determines the width of a relationship. There is a complete disease bipartite network available as Figure 2. We connected two diseases, di and dk, to the DD graph if they share at least one gene. The total number of genes is the determining factor in weight.

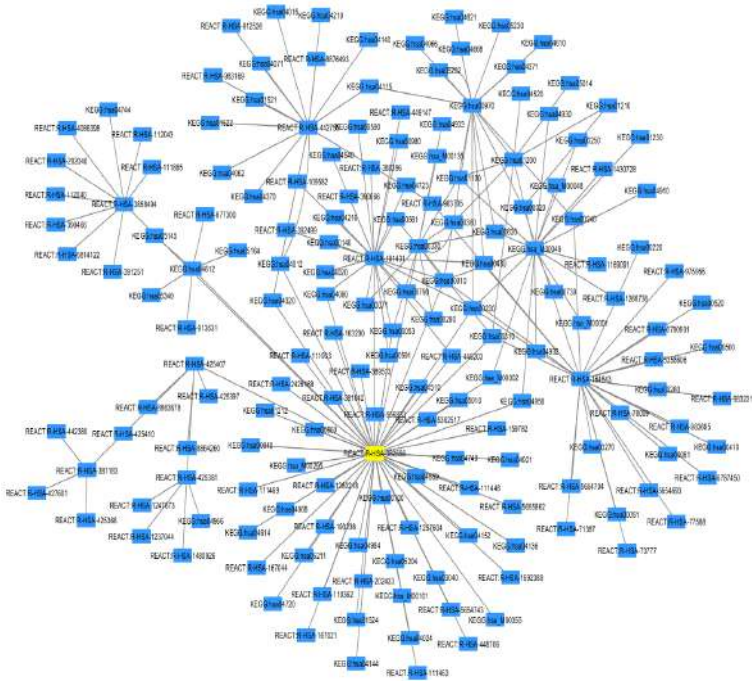

Figure 1. Pathway-Pathway Network

$$DD_{ik} = \sum_{j=1}^{|g|} DG_{ij} \quad DG_{kj} \Rightarrow DD_{ik} \tag{2}$$

(3)

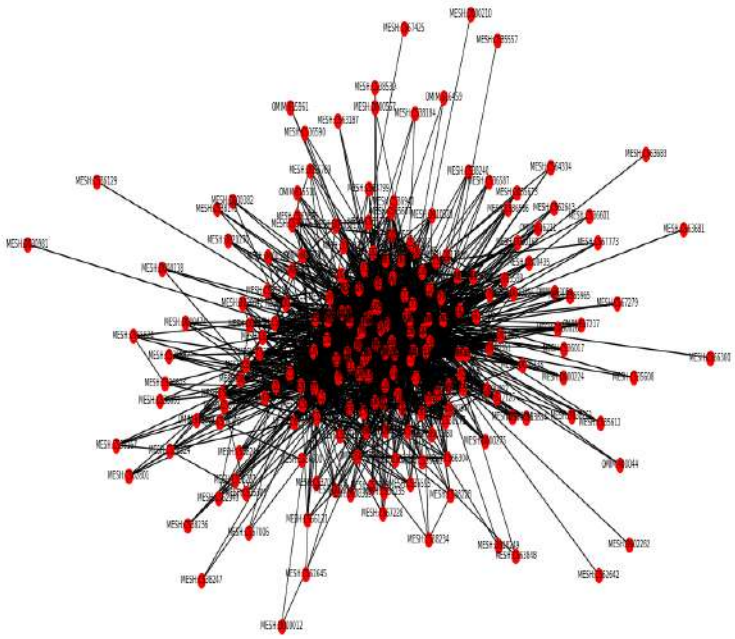

Figure 2. Disease-Disease Network

2.4 Graph Regularized Matrix Factorization Association

We analyzed the relationship between pathway-disease connotation matrix, Y R nm, can be approximated linearly by the equation Y W.DT, where W R nf, D R mf, we than f is the quantity of latent features in W (in pathway) and D. (in diseases). The selection of W and D, given a data matrix Y, must minimized the renovation error between Y and W DT. The squared error, often known as the euclidean distance with regard to the frobenius norm, is the most popular error function among those that have been suggested. Hence, the issue can be expressed as follows:

$$min_{W,D} \left\| Y - WD^T \right\| F^2 \tag{4}$$

In W or D alone, but not in both of them together, the objective function in Eq. 3 is convex. We can incorporate linear and graph regularization parameters to avoid overfitting. Whereas graph regularization parameters minimized the reserve between latent feature vectors of two neighboring pathways and disease, linear regularization terms minimized the norms of both W and H. Hence, the goal function is:

$$min_{W,D} \left\| Y - WD^T \right\| F^2$$
$$+\lambda_a Tr\left(W^T L_b W \right)$$
$$+\lambda_a Tr\left(D^T L_c D \right)$$

We proposed algorithm Graph Regularized Matrix Factorization (GRLMF)[17, 18], we offered a framework model as shown Figure 3. The singular value decomposition (SVD), we can extract U R nf , R ff and V R

mf from Y. We then initialize W and D as W = U and D = V ., respectively. In each round, W and D were updated using alternating least squares. By designating J as the objective function of Eq. 4, we can update W and D by setting J, W = 0 and J D = 0.

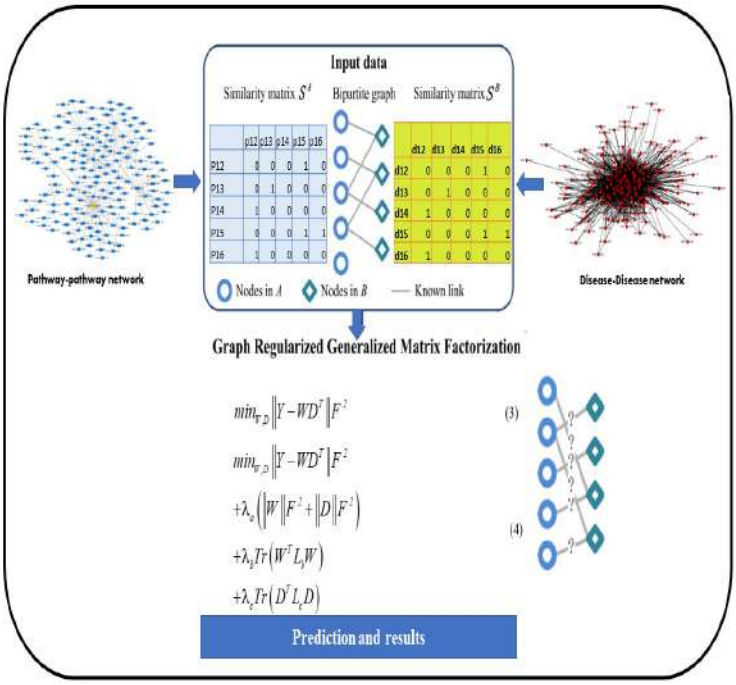

Figure 3. Proposed Framework model.

3 Performance Evaluation

We evaluated the performance of the proposed GRMF-based technique to three cutting-edge pathway-disease prediction methods in order to assess its effectiveness. The GRMF method's predictive accuracy is measured by computing the area under the ROC curve. AUC ranges from 0 to 1, with higher values indicating greater predictive ability. The performance is equivalent to a random forecast if the value is equal to 0.5.

The remaining pathways (such there is no relevant evidence) are classified as unlabeled nodes, whereas all the known pathways associated with the disease d1 are defined as labelled nodes. A labelled node is considered a positive sample if, given a threshold, its outcome prediction is larger than. A node is regarded as a negative sample if the outcome prediction of an unlabeled node is less than. The true positive rates (TPR, or sensitivity), as well as the false positive rates (FPR, or 1-specificity), were calculated by:

$$\mathbb{N}. \tag{5}$$

Sensitivity is the proportion of positive samples that are correctly identified among all the positives, while specificity is the proportion of negative samples that are successfully identified among all the negatives.

4 Results

We therefore partitioned our association matrix Y into six pieces for each repeat and removed each portion one at a time as the test set while using the final five components as the practice set. Consider all unidentified pathway-disease relationship pairs to be candidate samples. Following the application of GMRF, the

test sample scores were compared to all of the candidate sample scores. We did this procedure five times in order to improve the validity. The 5-fold cross validation for the GRLMF method's ROC curve and estimated AUC score 0.80% as shown figure 4(a) and we then AUPR score achieved 0.20% are shown in Figure 4(b).

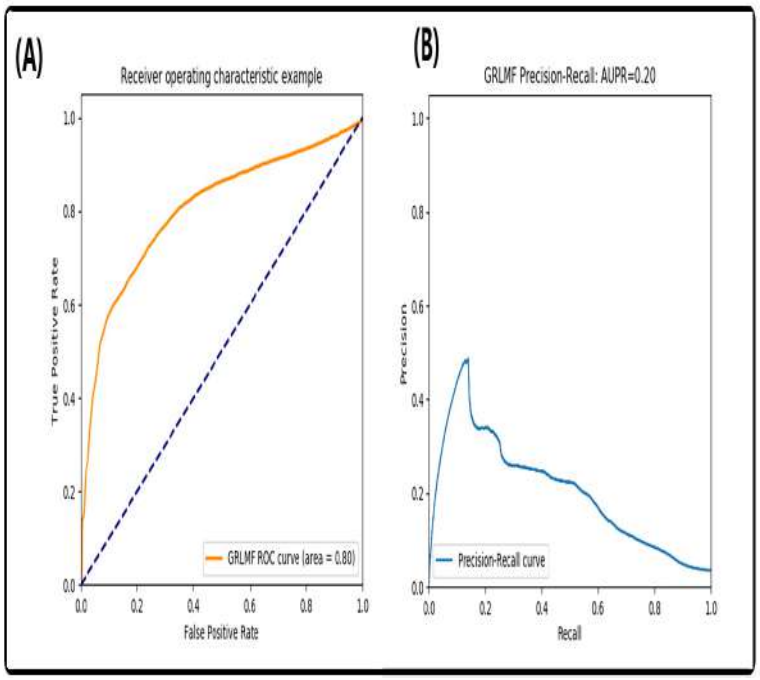

Figure 4. Figure 4. (a) The performance of ROC(AUC) score of DPA-GRLMF and (b) DPA-GRLMF PR checking performance metrics' performance metrics.

4.0.1 DPA-GRLMF

Table 2. 10 top rank pathway-disease association prediction results with their indication

Disease name	Pathway ID's	Indication	Obtained score
Adjustment Disorders	REACT:R-HSA-5610785	KeggDB	0.43897338425341337
Adult-onset citrullinemia type 2	REACT:R-HSA-5610785	KeggDB	0.4388230309612839
Adjustment Disorders	REACT:R-HSA-6804759	KeggDB	0.4203687018132794
Adult-onset citrullinemia type 2	REACT:R-HSA-6804759	KeggDB	0.42017391084911027
Alzheimer Disease, Familial, 3,	REACT:R-HSA-5610785	KeggDB	0.41347416254132074
Adjustment Disorders	REACT:R-HSA-375165	KeggDB	0.413359517299722
Adult-onset citrullinemia type 2	REACT:R-HSA-375165	KeggDB	0.41314826084760786
ALAZAMI-YUAN SYNDROME	REACT:R-HSA-5610785	KeggDB	0.4069040411266833
Adjustment Disorders	REACT:R-HSA-2022090	KeggDB	0.4054251773181184
Adult-onset citrullinemia type 2	REACT:R-HSA-2022090	KeggDB	0.4051954923241435

The analysis was based on 5-fold cross validations as achieved the 10 disorders and their connections to disease cancer are listed in Table 2. The results analysis consists 98% of those forecasts existed accurate in total. In the Table 2 demonstrates how we applied DPA-GRLMF to tumors with disease-pathways that included certain human pathways. It is possible to forecast dysregulated interactions and disease-related pathways using the DPA-GRLMF approach. We examined the factors linked to different diseases and made

predictions about probable routes that might be connected. Also, we achieved the good percentages in the table 2 represent using as the training sample cancer disease and activity of pathways connected to a certain class of pathways. Then, using cancer as a test sample, unknown correlations between cancer were discovered. Results of the investigation of 21 human illness-pathways, including the main cancer pathways, are shown in Table 2. Due to space restrictions, we concentrated on the outcomes of particular cancer types. The thorough, top-notch relationships between diseases based on pathways can also be used to examine crucial issues in systems medicine, like drug repurposing. Two medications, prednisone and folic acid, may have new indications, which may provide prospective possibilities for the treatment of complex disorders, according to a thick subgraph in our network.

4.1 DPA-GRLMF

Most scholars seem to agree that Graph Regularized Matrix Factorization algorithm which is best for network prediction. However, there continues to be debate about successfully transition from the transition matrix. There are two different methods to find the transition matrix. One method is the traditional method, and the other method, called Laplace normalization, uses the Laplacian concept. A limited number of studies have addressed regarding algorithm of ranking [19] known as the random restart walk (RWR) was used in the previous study [20] to aid in the prioritization of candidate genes [21, 22]. In order to find out whether or not the random walker has reached the desired node, we can calculate the probabilities for all the nodes in the graph. In this article, we suggest a novel method for link prediction called Graph Regularized Logistic Matrix Factorization (PDA-GRLMF) to find probable links in biological bipartite networks. Initially, in order to take use of the latent patterns hiding the observable linkages, we develop a generalized matrix factorization model. In particular, the neighborhood information of each node can be acquired adaptively, and it can be taken into consideration when learning the latent representation for each node.

4.2 A process of looking for biological pathways with high-performance using computational similarity

A comparative analysis of structures and composition of biological pathways will enable us to explain the roles of newly discovered routes, understand evolutionary features and identify missing pathway components. A method for comparing biological pathways and searching for similarities has been developed. The comparison determines the differences of each pair of XML paths. A ranking system is used to identify the pathway's similarities against those in the route repository. The search for similarity. The method is applied with the Condor high performance computer environment to achieve relatively good results [23].

4.3 Mechanism-based methods

Mechanism-based methods are more likely to assign genes and other biomolecules to different diseases, and then they compare sets of genes or other biomolecules to identify their overlap. Based on the paragraphs below, we can group the methods based on the association of diseases with the sets: gene-based, pathway-based, expression-based, network-based, and additional biomolecules (such as other biomolecules used in addition to genes). A number of studies have shown that gene-pathway association. However, important questions regarding gene-pathway association remain unanswered. Although the method by Goh et al. has been successful in the past, it is unable to discover disease relationships that do not share genes. Concretely, for instance, two diseases may have genes in common (such as g(A) and g(B), as well as a separate set of genes (e.g., g(A) and g(B)) that belong to a shared pathway. In other words, measures of gene similarity must go beyond known gene associations. In this case, the data may be integrated into the analysis by way of incorporating additional data, or diseases may be connected to a larger number of genes. There is a growing body of research on this particular instance, however, a better approach may be to identify disease-pathway associations. Previous studies have demonstrated genome-wide association studies (GWAS) and pathway-based similarity measures

was recently conducted by Lewis and colleagues [24] GWAS was used to investigate disease similarity at different levels. A recent line of research has established shows that pairwise similarity between diseases is higher when pathways, rather than genes, are compared. Current theories hypothesize that deploy pathway-based approaches could yield more new disease-disease relations because of the low number of existing known disease-pathway relationships. A number of studies have explored the relationship between gene and pathway, such as empirical evidence has supported the claim by Li and Agarwal [25] suggests that it is possible to conclude that two disease-gene and gene-pathway associations exist by first conducting pathway analysis to identify significant pathways and then investigating disease-gene and gene-pathway associations to see if any links can be made. Many researchers use other techniques to strengthen gene-disease associations.

5 Future Direction

Previous research has largely overlooked challenges associated with their shortfalls and discussed the possible formation of computational models for human pathway research in the future. To supplement the findings, we included information on the prediction [26]. In this research, we suggest the use of the GRLMF, a novel matrix factorization-based technique for identifying potential linkages in biological bipartite networks. Several different models have therefore been developed for calculating possible drug-target interactions on a large scale. Even though current computational models have greatly helped to advance the field of drug-target interactions, network and machine learning techniques have their own shortcomings. The network can be used to determine the drugs' interactions in the body. We're also looking into network-based pathways and strategies for pathways that are tailored to individuals based on genome sequence, tumour characterization, and cancer marking [27]. It is now known that lncRNAs play an important role in almost every stage of the life cycle of cells, both directly and indirectly. In short, this suggests that lncRNAs are altered and dysregulated as a result of a variety of human diseases, so it is not surprising that they contain mutations and are dysregulated [28].

6 Conclusion

Recently, computational techniques to forecast probable linkages for various biomedical bipartite networks have been presented. Existing techniques, however, frequently rely on the coverage of established linkages, which can be problematic when working with new nodes that lack established link information. Moreover, the continuous intercommunication between genes results in complex pathways. For certain disease, the candidate genes of infections have a productive relationship in the context of protein or organic pathways. Therefore, this study attempted to investigate the disease relationship of seed pathway nodes that are based on shared genes [29, 30]. Finally, we provide the following summary of our contributions. Then, we created a GRLMF model that uses neighborhood data to determine each node's latent component. Second, our model could learn the neighbor information for each node adaptively and further support the prediction of probable links rather than using similarity matrices derived from externally associated databases with predefined metrics. Third, we carry out extensive experiments that show how successful the suggested GRGMF approach.

Author Contributions

A.G.: Conceptualization, Methodology, Software, Data curation, Writing- Original draft preparation, Visualization, Investigation. Supervision, Software, Validation, Writing- Reviewing and Editing

Compliance with Ethical Standards

It is declare that all authors don't have any conflict of interest. It is also declare that this article does not contain any studies with human participants or animals performed by any of the authors. Furthermore,

informed consent was obtained from all individual participants included in the study.

Funding Information

FMS acknowledges the support of NSF grant CHE-1111111.

Author Information

ORCID:
Author 1 name: AAAA-BBBB-CCCC-DDDD
Author 2 name: EEEE-FFFF-GGGG-HHHH

References

[1] L. Eronen and H. Toivonen, “Biomine: predicting links between biological entities using network models of heterogeneous databases,” BMC Bioinformatics, vol. 13, no. 1, p. 119, 2012.

[2] T. Gaudelet, N. Malod-Dognin, J. Sanchez-Valle, V. Pancaldi, A. Valencia, and N. Przulj, “Unveiling new disease, pathway, and gene associations via multi-scale neural networks,” arXiv [q-bio.QM], 2019.

[3] M. Agrawal, M. Zitnik, and J. Leskovec, “Large-scale analysis of disease pathways in the human interactome,” in Biocomputing 2018, 2018.

[4] H. Lee and M. Shin, “Mining pathway associations for disease-related pathway activity analysis based on gene expression and methylation data,” BioData Min., vol. 10, no. 1, p. 3, 2017.

[5] R. Giugno, A. Pulvirenti, L. Cascione, G. Pigola, and A. Ferro, “MIDClass: Microarray data classification by association rules and gene expression intervals,” PLoS One, vol. 8, no. 8, p. e69873, 2013.

[6] F. Emmert-Streib and G. V. Glazko, “Pathway analysis of expression data: deciphering functional building blocks of complex diseases,” PLoS Comput. Biol., vol. 7, no. 5, p. e1002053, 2011.

[7] Y. Liu and Z. Hu, “Identification of collaborative driver pathways in breast cancer,” BMC Genomics, vol. 15, p. 605, 2014.

[8] M. List, N. Alcaraz, and R. Batra, “De Novo pathway-based classification of breast cancer subtypes,” Methods Mol. Biol., vol. 2074, pp. 201–213, 2020.

[9] E. Lee, H.-Y. Chuang, J.-W. Kim, T. Ideker, and D. Lee, “Inferring pathway activity toward precise disease classification,” PLoS Comput. Biol., vol. 4, no. 11, p. e1000217, 2008.

[10] P. Khatri, M. Sirota, and A. J. Butte, “Ten years of pathway analysis: current approaches and outstanding challenges,” PLoS Comput. Biol., vol. 8, no. 2, p. e1002375, 2012.

[11] Y. Liu and M. R. Chance, “Pathway analyses and understanding disease associations,” Curr. Genet. Med. Rep., vol. 1, no. 4, pp. 230–238, 2013.

[12] D. Yu, M. Kim, G. Xiao, and T. H. Hwang, “Review of biological network data and its applications,” Genomics Inform., vol. 11, no. 4, pp. 200–210, 2013.

[13] Y. Li and P. Agarwal, “A pathway-based view of human diseases and disease relationships,” PLoS One, vol. 4, no. 2, p. e4346, 2009.

[14] K. Wang, M. Li, and M. Bucan, “Pathway-based approaches for analysis of genomewide association studies,” Am. J. Hum. Genet., vol. 81, no. 6, pp. 1278–1283, 2007.

[15] S. Köhler, S. Bauer, D. Horn, and P. N. Robinson, "Walking the interactome for prioritization of candidate disease genes," Am. J. Hum. Genet., vol. 82, no. 4, pp. 949–958, 2008.

[16] Y. Li and J. C. Patra, "Integration of multiple data sources to prioritize candidate genes using discounted rating system," BMC Bioinformatics, vol. 11 Suppl 1, no. S1, p. S20, 2010.

[17] Z.-C. Zhang, X.-F. Zhang, M. Wu, L. Ou-Yang, X.-M. Zhao, and X.-L. Li, "A graph regularized generalized matrix factorization model for predicting links in biomedical bipartite networks," Bioinformatics, vol. 36, no. 11, pp. 3474–3481, 2020.

[18] Z. Gao, Y.-T. Wang, Q.-W. Wu, J.-C. Ni, and C.-H. Zheng, "Graph regularized L2,1-nonnegative matrix factorization for miRNA-disease association prediction," BMC Bioinformatics, vol. 21, no. 1, p. 61, 2020.

[19] C. Wu, R. Gao, D. Zhang, S. Han, and Y. Zhang, "PRWHMDA: Human microbe-disease association prediction by random walk on the heterogeneous network with PSO," Int. J. Biol. Sci., vol. 14, no. 8, pp. 849–857, 2018.

[20] X. Shen, Y. Chen, X. Jiang, X. Hu, T. He, and J. Yang, "Predicting disease-microbe association by random walking on the heterogeneous network," in 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2016.

[21] S. Zou, J. Zhang, and Z. Zhang, "A novel approach for predicting microbe-disease associations by bi-random walk on the heterogeneous network," PLoS One, vol. 12, no. 9, p. e0184394, 2017.

[22] A. Ghulam, X. Lei, M. Guo, and C. Bian, "Disease-pathway association prediction based on random walks with restart and PageRank," IEEE Access, vol. 8, pp. 72021–72038, 2020.

[23] H. N. Kim et al., "Effectiveness of direct-acting antiviral therapy in patients with human immunodeficiency virus-hepatitis C virus coinfection in routine clinical care: A multicenter study," Open Forum Infect. Dis., vol. 6, no. 4, p. ofz100, 2019.

[24] S. N. Lewis, E. Nsoesie, C. Weeks, D. Qiao, and L. Zhang, "Prediction of disease and phenotype associations from genome-wide association studies," PLoS One, vol. 6, no. 11, p. e27175, 2011.

[25] J.-A. R. Sarmiento, A. Lao, and G. A. Solano, "Pathway-based human disease clustering tool using self-organizing maps," in 2017 8th International Conference on Information, Intelligence, Systems Applications (IISA), 2017.

[26] K. Jiang, Y. Huang, and J. Robertson, "A method of biological pathway similarity search using high performance computing," Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., vol. 2009, pp. 4933–4936, 2009.

[27] X. Chen, B. Ren, M. Chen, Q. Wang, L. Zhang, and G. Yan, "NLLSS: Predicting synergistic drug combinations based on semi-supervised learning," PLoS Comput. Biol., vol. 12, no. 7, p. e1004975, 2016.

[28] X. Chen, C. C. Yan, X. Zhang, and Z.-H. You, "Long non-coding RNAs and complex diseases: from experimental results to computational models," Brief. Bioinform., p. bbw060, 2016.

[29] I. Baskara-Yhuellou and J. Tost, "The impact of microRNAs on alterations of gene regulatory networks in allergic diseases," Adv. Protein Chem. Struct. Biol., vol. 120, pp. 237–312, 2020.

[30] L. Hood, J. R. Heath, M. E. Phelps, and B. Lin, "Systems biology and new technologies enable predictive and preventative medicine," Science, vol. 306, no. 5696, pp. 640–643, 2004.