Journal of Advanced Research in Microelectronics and VLSI
Volume 7, Print Issue 1 - 2024, Pg. No. 10-16
Peer Reviewed Journal

Research Article

Comparative Analysis of Differential Networks in Biological System Between Diseased and Normal Conditions

Pratibha Soram[1], Harjit Kaur[2]

[1,2]Assistant Professor, Department of Computer Science & Engineering, PCTE Institute of Engineering & Technology, Punjab, India.

I N F O

Corresponding Author:
Pratibha Soram Department of Computer Science & Engineering, PCTE Institute of Engineering & Technology, Punjab, India.
E-mail Id:
pratibhasoram@pcte.edu.in
Orcid Id:
https://orcid.org/0009-0001-2945-8424
How to cite this article:
Soram P , Kaur H. Comparative Analysis Of Differential Networks In Biological System Between Diseased And Normal Conditions. J Adv Res Microelec VLSI 2024; 7(1): 10-16.
Date of Submission: 2024-03-17
Date of Acceptance: 2024-04-18

A B S T R A C T

Comprehension of biological networks and their comparison is one of the key issues in computational biology. It is essential to find out the differences in the characteristics and topological structures of biological networks between affected and healthy tissues to understand any diseases better. Data concerning interactions occurring within a cell can be found in biological networks. All the biological networks are dependent with one another inside the cell and form a network of networks determining the cell's overall behavior. I have identified the deviating gene(s) in my study using a variety of network parameters such as degree, degree distribution, rewiring edges, average path length, clustering co-efficient, scale free networks and power distribution. R programming language was utilized to do the experimental analysis.

Keywords: Gene Regulatory Networks (GRNs), Discretization, Transfer Entropy, Comparative Analysis.

Introduction

Purpose of the Research

Comparing two biological networks—normal or benign and diseased or malignant—is the aim of this study. To obtain the relevant biological network, we first examine the existing inference methods on gene expression datasets. We also conduct a thorough analysis of the biological network comparison techniques that are currently available. Finally, we perform a variety of network comparison analyses, such as calculating absolute degree differences, analyzing rewiring patterns within the network, comparing the average path length, and comparing the clustering co-efficient across the systems.

Various techniques such as Conditional Mutual Information, Transfer Entropy, and Dynamic Bayesian Network can be employed for GRN inference. However, we have used Transfer Entropy in our study to infer the necessary GRNs.

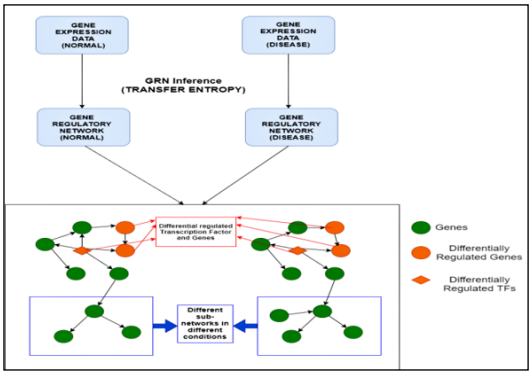

Figure 1.Overview of the research

Objective of the Research

The primary aim of this study is to obtain the inferred network by utilizing Transfer Entropy, an established inference approach. Secondly, to perform post inference analysis of the disease networks by comparing normal vs. disease

networks. To identify the noteworthy distinctions between the two networks' topologies is the third goal so as to rank the nodes of the networks based on their higher degree differences. And lastly, employing a variety of network comparison approaches to pinpoint the deviating gene(s) between the two networks.

What Are Gene Regulatory Networks (Grns)?

GRN is a gene interaction network that describes how genes are activated or repressed, and therefore contains information about which proteins are produced in a cell at a particular time. In order to identify patterns that distinguish between cohorts of affected and healthy patients, Gene Regulatory Networks are frequently employed to examine the changes in genetic conditions between the two.

Discretization

Discretization is commonly used to convert continuous data to discrete ones. Some of the measures for inferring causality in GRN like Interaction Information, Transfer entropy etc., requires discrete data as input and hence the available continuous data needs to be converted to discrete time series data. There are three types of Discretization methods discussed in this project viz. Equal Width Discretization (EWD), Equal Frequency Discretization (EFD) and Global Equal Width Discretization (GWED).

Transfer Entropy

Transfer Entropy (TE) was first introduced by Thomas Schreiber in 2000. The GRN Inference method utilized to obtain the networks for this research is Transfer Entropy. Further explanation of Transfer Entropy is discussed in Chapter 4.

Literature Review

A network-based technique called Disease-Gene Association (DGA) was created by Suratanee et al. to determine the association score between a query gene and a newly conceivable range of disorders. Two new graphlet-based measures, REConstruction Rate (REC) and REC Graphlet Degree (RGD), were proposed by Martin et al. RGD detects the subset of nodes with the most topological variation, while REC calculates the rate of graphlet similarity between various network states. Differential Gene Regulatory Networks (DiffGRNs) are a method that Kim et al. devised to find substantial variations in the gene controls of various GRNs under dissimilar circumstances. Koschutzki et al. compared two biological networks using a variety of centrality metrics, including degree, eccentricity, closeness, etc. In order to do a differential gene regulatory network analysis, Dirk et al. examined many centrality measures, including degree, shortest path betweenness, Katz status index, and PageRank.

Gene Regulatory Networks (GRNs) represent regulatory interactions between genes with a graph model, where a node and an edge indicate a gene and functional relationships between genes respectively. GRNs can be computationally approximated by reverse engineering from gene expression data, which is called GRN inference.

GRN inference can be mainly categorised into three: (i) Correlation-based, (ii) Bayesianbased, and (iii) Regression-based inference. Correlation-based GRN inference, which is the simplest approach of reconstructing GRN, builds a co-expression network by measuring interactions between genes with correlation coefficients, and the connections of the network are determined by a threshold. Correlation-based GRNs are undirected graphs, and only pairwise relationships between genes are considered. In contrast to the correlation-based GRNs, Bayesian-based approaches can capture a regulatory direction between genes by approximating conditional probabilities of genes. Additionally, Bayesian-based methods can incorporate previous biological knowledge with prior probability and handle a missing value problem. However, due to their computational complexity, Bayesian-based GRNs are not practical for handling vast amounts of genomic data, and it is difficult to reliably predict the distribution using a small number of samples. Regression-based GRN inference builds a directed graph for GRN by taking into account the multivariate effects of genes that control a gene based on a regression model. In the evaluation of GRN inference, regression-based inference has demonstrated exceptional performance with ensemble learning. Because correlation-based inference is easy to implement and has an understandable representation, it has been the primary method used in the majority of DiNA studies to create gene regulatory networks from gene expression data.

Research Methodology

The following is the proposed project's execution plan:

1. Collecting datasets from NCBI
2. Using R studio platform, discretizing the datasets
3. Implementing Transfer Entropy to obtain the two GRNs
4. Performing comparative analysis using various network parameters
5. Analyzing the results

In order to obtain the necessary networks for additional analysis in our project, we will first obtain the datasets and discretize them. Then, we will apply the inference technique i.e., Transfer Entropy. Network characteristic parameters, which are covered in more detail in Chapter 4, are used to compare the two networks.

Implementation

Dataset Description

The (disease) expression data containing 54675 genes is obtained from NCBI. The (normal) expression data containing 4774 genes is also obtained from NCBI.

Table 1.Description of the datasets used

Sample	GSM3318887(Disease), GSM3100633(Normal)
Organism	Homo Sapiens
Cell Type	Prostate Cancer cells

The following section contains the implementation part of our work.

Discretization

The following techniques are used for pre-processing continuous time series data.

Equal Width Discretisation (EWD)

In EWD, the continuous time series data is sorted in any order and then the range of observed variables are divided into k equally size bins, where k is a parameter supplied by the user. If a variable x is observed to have values bounded by x_{min} and x_{max} then this method computes the bin width, δ = and constructs bin boundaries, or thresholds, at x_{min}+i , where i=1,...,k-1. The method is applied to each continuous feature independently. It makes no use of instance class information and is thus an unsupervised discretization method.

Equal Frequency Discretization (EFD)

In EFD, the continuous time series data is partitioned into |xi| intervals, each having the same number, m/$|x_i|$, of data points. As a result, the intervals can have different sizes. If the $|x_i|$ intervals have equal frequency, then the computation of entropy is straightforward: log| |. On the other hand, the resulting entropy must be estimated if one of the bins is denser than the rest. This discretization method is one of the most efficient methods.

Global Equal Width Discretization (GEWD)

Although GEWD and EWD have similarities, GEWD considers a distinct range between the minimum and highest value. While we use the maximum and minimum of the complete dataset in GEW, we simply use the local maximum and minimum as the range in EWD.

Transfer Entropy

Schreiber developed the notion of Transfer Entropy (TE) between two processes which measures the uncertainty reduction in determining the future state of a process by learning the past and current states of other processes. Supposing that two systems generates events, we define an entropy rate which is the amount of additional information required to represent the value of the next observation of one of the systems:

$$h_1 = -\sum_{x_{n+1}} p(x_{n+1}, x_n, y_n) log_a\, p(x_{n+1}|x_n, y_n) \quad \ldots\ldots (1)$$

Supposing that value of observation x_{n+1} was not dependent on the current observation y_n:

$$h_2 = -\sum_{x_{n+1}} p(x_{n+1}, x_n, y_n) log_a\, p(x_{n+1}|x_n) \quad \ldots\ldots (2)$$

Now, the quantity h_1 represents the entropy rate for the two systems, and h_2 represents the entropy rate assuming that x_{n+1} is independent of y_n. Thus, we get transfer entropy as follows:

$$h_2 - h_1 = -\sum_{x_{n+1}, x_n, y_n} p(x_{n+1}, x_n, y_n) log_a\, p(x_{n+1}|x_n) + \sum_{x_{n+1}, x_n, y_n} p(x_{n+1}, x_n, y_n) log_a\, p(x_{n+1}|x_n, y_n) = \sum_{x_{n+1}, x_n, y_n} p(x_{n+1}, x_n, y_n) log_a\, (p(x_{n+1}|x_n, y_n)|\, p(x_{n+1}|x_n)) \quad \ldots\ldots (3)$$

Because Transfer Entropy has an inherent asymmetry in it, there are two equations for it.

$$T_{J\to I} = \sum_{x_{n+1}, x_n, y_n} p(x_{n+1}, x_n, y_n) log_a\, (p(x_{n+1}|x_n, y_n)|\, p(x_{n+1}|x_n)) \quad \ldots\ldots (4)$$

$$T_{I\to J} = \sum_{y_{n+1}, x_n, y_n} p(y_{n+1}, x_n, y_n) log_a\, (p(y_{n+1}|x_n, y_n)|\, p(y_{n+1}|x_n)) \quad \ldots\ldots (5)$$

Substituting:

$$p(x_{n+1}|x_n, y_n) = p(x_{n+1}, x_n, y_n)|\, p(x_n, y_n)$$

$$p(x_{n+1}|x_n) = p(x_{n+1}, x_n)|\, p(x_n)$$

Equations (4) & (5) becomes:

$$T_{J\to I} = \sum_{x_{n+1}, x_n, y_n} p(x_{n+1}, x_n, y_n) log_a\, (p(x_{n+1}|x_n, y_n).\, p(x_n)|\, p(x_{n+1}|x_n)) \quad \ldots\ldots (6)$$

$$T_{I\to J} = \sum_{y_{n+1}, x_n, y_n} p(y_{n+1}, x_n, y_n) log_a\, (p(y_{n+1}|x_n, y_n)|\, p(x_n, y_n).p(y_{n+1}|x_n)) \quad \ldots\ldots (7)$$

"TE is more adequate for determining the direction of inf. flow between two coupled processes."[Kaiser, A. and Schreiber, T., 2002]. One can obtain an inferred causal network of genes by determining the TE between each pair of genes. One can then use a heuristic approach to distinguish between indirect and direct causal relationships

If I=J, then the value is assigned as 0, else for each I≠J, the value is assigned the TE value. After this, max and min of the values in the matrix is obtained to find the threshold value as: $\text{threshold} = \min + \frac{\max - \min}{2}$

For every value TE, the value is compared with the resultant threshold value. If TE<threshold, then assign 0. Else TE is assigned 1. Thus, the final matrix that is produced, which is made up of discrete values, will be the necessary network.

Network Parameters Used

Some of the commonly used network parameters to characterize any real life networks are discussed as follows:

Degree (Connectivity)

It is the number of edges a node has to other nodes in the network. It is the measure of connectivity of a node in the network. In directed network, each of the nodes has two types of degrees:

i) In-degree (k_{in}), also known as incoming edges, and

ii) Out-degree (k_{out}), also known as outgoing edges.

In the above Diseased network, $k_{in}(A)=2$, $k_{in}(B)=1$, $k_{in}(C)=1$, $k_{in}(D)=3$, $k_{in}(E)=0$, $k_{out}(A)=1$, $k_{out}(B)=2$, $k_{out}(C)=1$, $k_{out}(D)=1$, $k_{out}(E)=2$. In the above Normal network, $k_{in}(A)=2$, $k_{in}(B)=1$, $k_{in}(C)=3$, $k_{in}(D)=2$, $k_{in}(E)=0$, $k_{out}(A)=1$, $k_{out}(B)=2$, $k_{out}(C)=1$, $k_{out}(D)=1$, $k_{out}(E)=2$.

Degree Distribution

Degree distribution (DD) is the fraction of the number of nodes with degree k to the total number of nodes in the network. It is denoted as:

$$P_{deg}(k) = n(k)/n$$

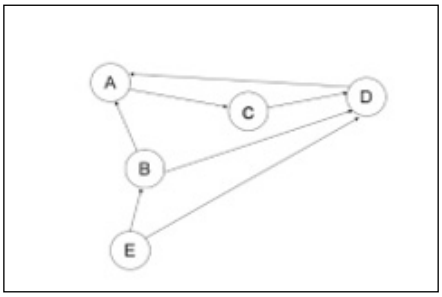

Figure 2.(a) Sample directed(Diseased) network

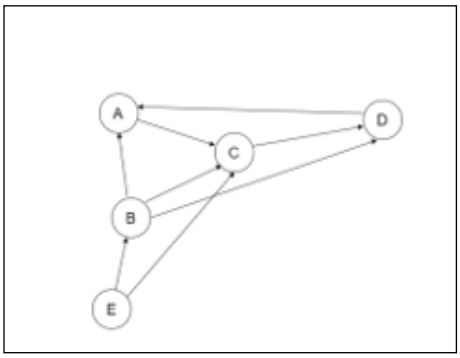

Figure 3.(b) Sample directed(Normal) network

where n(k) is the total number of nodes with degree k, and n is the total number of nodes in the network.

But, in a directed graph, the degree distribution becomes a two-dimensional distribution, so $P_{deg}(k_{in}, k_{out})$ is the fraction of the nodes in the graph with in-degree k_{in} and out-degree k_{out} . So, the total degree distribution will become, $P^{tot}_{deg}(k_{tot})$, where $k_{tot}=k_{in}+k_{out}$.

So, for the given above diseased network, total degree k_{tot} will become: $k_A^{tot}=3$, $k_B^{tot}=3$, $k_C^{tot}=2$, $k_D^{tot}=4$, $k_E^{tot}=2$. Similarly, for the normal network, total degree ktot will become: $k_A^{tot}=3$, $k_B^{tot}=3$, $k_C^{tot}=4$, $k_D^{tot}=3$, $k_E^{tot}=2$.

Hence, the Degree Distribution (DD) for the normal network can be obtained as $P_{deg}(2)=2/5=0.4$, $P_{deg}(3)= 2/5 = 0.4$, $P_{deg}(4)= 1/5 = 0.2$, and the Degree Distribution (DD) for the normal network can be obtained as $P_{deg}(2)= 1/5 = 0.2$, $P_{deg}(3)= 3/5 = 0.6$, $P_{deg}(4)= 1/5 = 0.2$.

Rewiring Edges

It is the percentage of edges that got rewired i.e., the edges that got deleted or added in the affected network in comparison to the normal counterpart. It can also be defined as the indicator of similarity or dissimilarity between the two networks.

Three parameters X, Y and Z is used to measure the percentage edge deletion and percentage edge addition:
Percentage Edge Deletion = $(Y/X) \times 100$

Percentage Edge Addition = $(Z/X) \times 100$

where X is the total number of edges present in the normal network, Y is the total number of edges present in the normal but not in the affected network, and Z is the total number of edges present in the affected but not in the normal network.

In the above figure, we can see that the total number of edges present in the normal network is 8, so X=8. There are two edges (e5 and e8) which are present in the normal but not in the diseased network, so Y=2. And finally there is one edge (e9) which is present in the diseased but not in the normal network, so Z=1. Hence, Percentage Edge Deletion= (2/8)×100 =25% and, Percentage Edge Addition= (1/8) × 100=12.5%.

Average Path Length (APL)

APL of a network is the average distance of all pairs of its nodes[23]. It is also the measure of efficiency or mass transport on a network, and is calculated by using the following formula,

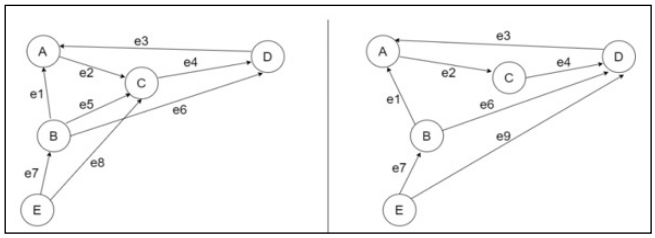

Figure 4.Re-wiring structures of the sample normal network and diseased network

$$l_g = \frac{1}{n(n-1)} \sum_{i \neq j} d(v_i, v_j)$$

where n is the total number of nodes, and $d(v_i, v_j)$ is the distance between the vertices v_i and v_j.

The average distance between any couples of nodes may also highlight interesting properties of a given graph. APL is a measure of the efficiency of information or mass transport on a network. It offers a measure of a network's overall navigability. In the above diseased network, d(A,B)=0,

d(A,C)=1, d(A,D)=2, d(A,E)=0, d(B,A)=1, d(B,C)=2, d(B,D)=3, d(B,E)=0, d(C,A)=2, d(C,B)=0, d(C,D)=1, d(C,E)=0, d(D,A)=1, d(D,B)=0, d(D,C)=2, d(D,E)=0, d(E,A)=2, d(E,B)=1, d(E,C)=3, d(E,D)=1. All the distances adds to 22 and since there are 5 nodes therefore the APL, l_g = ×22= 1.1

In the above normal network, d(A,B)=0, d(A,C)=1, d(A,D)=2, d(A,E)=0, d(B,A)=1, d(B,C)=1, d(B,D)=1, d(B,E)=0, d(C,A)=2, d(C,B)=0, d(C,D)=1, d(C,E)=0, d(D,A)=1, d(D,B)=0, d(D,C)=2, d(D,E)=0, d(E,A)=2, d(E,B)=1, d(E,C)=1, d(E,D)=2. All the distances adds to 22 and since there are 5 nodes therefore the APL, l_g = × 18=0.9

Clustering Co-efficient (CC)

CC is the measure of the degree of connectedness among the neighbouring nodes of that node. Clustering Co-efficient characterizes the overall tendency of nodes to form clusters or groups. The CC of a network is the sum of all the CC of all the nodes in the network. The CC of a node is given by the following formula: $C_i = \frac{2Li}{ki\,(ki-1)}$, where k_i is the degree of node i, and L_i is the number of edges between ki neighbors of node i. So, the CC of the entire network is given by: $C = \frac{1}{N}\sum_{i=1}^{N} Ci$, where N is the total number of nodes in the entire network.

In the above diseased network, for node A: K_A=3, L_A=2 and therefore C_A=0.667. Similarly, for node B: K_B=3, L_B=2 and therefore C_B=0.667, for node C: K_C=2, L_C=1 and therefore CC=1, for node D: K_D=4, L_C=1 and therefore C_D=0.167 and finally for node E: K_E=2, L_E=1 and therefore C_E=1. Thus, the CC of the entire network is, C= (0.667 + 0.667 + 1 + 0.167 +1) = 0.70.

In the above normal network, for node A: K_A=3, L_A=3 and therefore C_A=0.667. Similarly, for node B: K_B=3, L_B=2 and therefore C_B=0.667, for node C: K_C=4, L_C=1 and therefore CC=1, for node D: K_D=3, L_C=3 and therefore C_D=0.167 and finally for node E: K_E=2, L_E=1 and therefore C_E=1. Thus, the CC of the entire network is, C= (1 + 0.667 + 0.167 +1 + 1) = 0.87.

Scale Free Networks & Power Law Distribution

A scale-free network is a complex network or a connected graph with the property that the number of links originating from a given node exhibits a power law distribution. Real life networks including biological networks exhibit scale-free architecture, where connectivity among the nodes ideally follows a power law degree distribution. According to the power law distribution, the number of high degree nodes are relatively lower as compared to low degree nodes. Another important component of scale-free networks is its clustering coefficient. It has been observed that real networks have a much higher clustering coefficient. Another feature of a scale-free real network is low average path length.

Results And Analysis

The real-time datasets are first acquired from NCBI, and discretization techniques are then used to transform the continuous data into discrete data. Secondly, we use the GRN inference technique known as Transfer Entropy to obtain the necessary networks—Diseased and Normal—for further investigation. To avoid inference biasness and to produce a concurrent observation, Transfer Entropy method is used for making the comparative study.

After obtaining the results, we have taken top 20 genes and proceeded with the comparative analysis.

Comparative Analysis of the Networks

To find the deviating gene(s) between the two networks and to uncover any notable changes between their topologies, comparative analysis is performed. Using the network parameters from Chapter 4, we apply them to an actual diseased network and compare its results with those of the normal network.

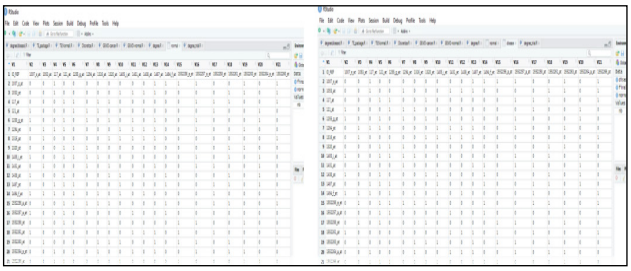

Figure 5.Resulted GRNs (Normal and Diseased) after applying inference (Transfer Entropy) technique to the datasets

Degree Distribution (DD) Comparison

We have displayed the degree distribution inferred for real networks of both diseased and normal GRN in figure 5.1.1. The power law distribution is satisfied by the normal network, as shown by the figure, where the DD value for the higher degree (17) is less than that of the lower degrees (15, 16). However, the DD value for the higher degree (17) in the disease network is greater than the lower degree (15), resulting that the disease network does not satisfy the power law distribution.

Network Re-wiring structures analysis

We conducted an experiment to determine the values of the percentages of edge addition and percentage of edge deletion. The following table shows the result of our experiment.

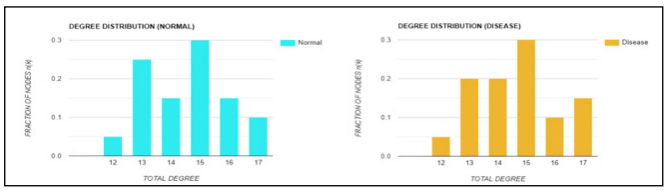

Figure 6.Degree Distribution for Normal Vs Diseased network

It may be noted that the lesser the percentage value of edge deletion and edge addition, the better is the inference technique, as this indicates how close the diseased network is to its normal counterpart.

Average Path Length (APL) Comparison

The average path length of the two inferred networks is displayed in the above figure. Moreover, we see that the APL in the normal network is higher than in the diseased network, suggesting that the nodes in the diseased network have greater connection than in the normal network

Table 2.Percentage of Edge Deletion and Addition between Normal and Diseased Networks

Edge	Percentage Value
Percentage Edge Deletion = $(Y/X) \times 100$	1.37457
Percentage Edge Addition = $(Z/X) \times 100$	2.061856

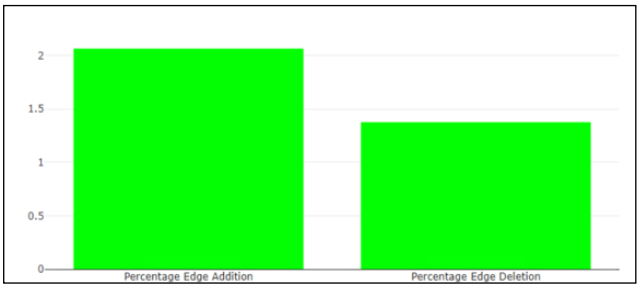

Figure 7.PED & PEA for the Diseased and Normal networks

Clustering Co-efficient Comparison

The results of the two inferred networks' clustering coefficients are displayed in the above figure. It can be noticed that the normal network has a higher clustering coefficient than the diseased network, indicating that some of the causal edges between the neighboring nodes in the diseased network have been eliminated, which has decreased the diseased network's clustering coefficient.

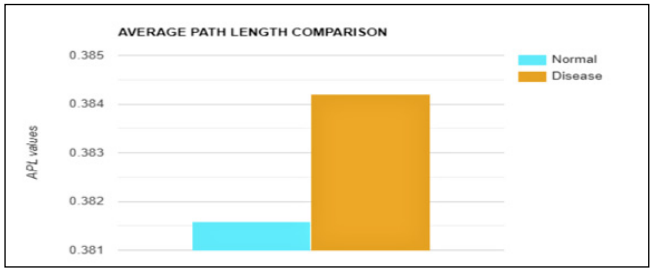

Figure 8.Average Path Length for Normal versus Diseased Networks

Discussion and Conclusion

It has been observed that the normal network exhibits a power law distribution and is more akin to a scale-free

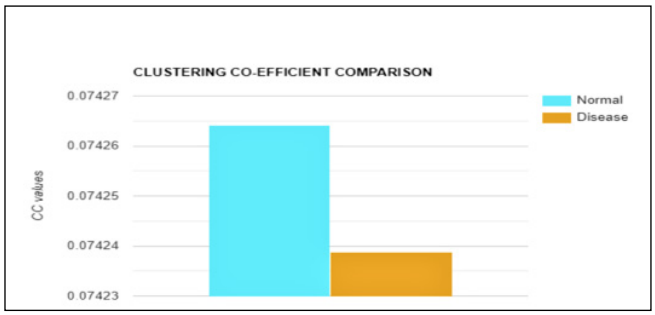

Figure 9.Clustering Co-efficient for Normal versus Diseased Networks

actual network than the diseased one. Additionally, we computed the clustering coefficient and average path length of the normal and diseased networks to conduct a comparative network analysis. Our result also indicates that there is a higher similarity between the normal and the diseased network.

References

1. Da Ruan, Alastair Young, and Giovanni Montana. Differential analysis of biological networks. BMC bioinformatics, 16(1):327, 2015
2. Valeria Fionda. Biological network analysis and comparison: mining new biological knowledge. Open Computer Science, 1(2):185–193, 2011.
3. Swarup Roy, Dipankar Das, Dhrubajyoti Choudhury, Gunenja G Gohain, Ramesh Sharma, and Dhruba K Bhattacharyya. Causality Inference Techniques for In-Silico Gene Regulatory Network. In Mining Intelligence and Knowledge Exploration,pages 432–443. Springer, 2013.
4. Lichtblau Y, Zimmermann K, Haldemann B, Lenze D, Hummel M and Leser U (2017) 'Comparative assessment of differential network analysis methods,' Briefings in Bioinformatics, Vol. 18, No. 5, pp.837–850.
5. Chai, L.E., Loh, S.K., Low, S.T., Mohamad, M.S., Deris, S. and Zakaria, Z., 2014. A review on the computational approaches for gene regulatory network construction. Computers in biology and medicine, 48, pp.55-65.
6. Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., Aderhold, A., Bonneau, R., Chen, Y. and Collins, J.J., 2012. Wisdom of crowds for robust gene network inference. Nature methods, 9(8), p.796.

7. Banf, M. and Rhee, S.Y., 2017. Computational inference of gene regulatory networks: approaches, limitations and opportunities. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1860(1), pp.41-52.
8. Hill, S.M., Heiser, L.M., Cokelaer, T., Unger, M., Nesser, N.K., Carlin, D.E., Zhang, Y., Sokolov, A., Paull, E.O., Wong, C.K. and Graim, K., 2016. Inferring causal molecular networks: empirical assessment through a community-based effort. Nature methods, 13(4), p.310.
9. Apichat Suratanee and Kitiporn Plaimas. Network-based association analysis toinfer new disease-gene relationships using large-scale protein interactions. PloSone, 13(6):e0199435, 2018.
10. Alberto JM Martin, Calixto Dominguez, Sebastián Contreras-Riquelme, David S Holmes, and Tomas Perez-Acle. Graphlet based metrics for the comparison of gene regulatory networks. PloS one, 11(10):e0163497, 2016.
11. Youngsoon Kim, Jie Hao, Yadu Gautam, Tesfaye B Mersha, and Mingon Kang. Diffgrn: differential gene regulatory network analysis. International Journal of Data Mining and Bioinformatics, 20(4):362–379, 2018.
12. Dirk Koschützki and Falk Schreiber. Comparison of centralities for biological networks. In German Conference on Bioinformatics, pages 199–206. Citeseer, 2004.
13. T. Schreiber. Measuring information transfer. Physical review letters, 85(2):461, 2000.
14. Dirk Koschützki and Falk Schreiber. Centrality analysis methods for biological networks and their application to gene regulatory networks. Gene regulation and systems biology, 2:GRSB– S702, 2008.
15. Leo Katz. A new status index derived from sociometric analysis. Psychometrika, 18(1):39– 43, 1953.
16. Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.
17. Gennaro Gambardella, Maria Nicoletta Moretti, Rossella De Cegli, Luca Cardone, Adriano Peron, and Diego Di Bernardo. Differential network analysis for the identification of condition-specific pathway activity and regulation. Bioinformatics, 29(14):1776–1785, 2013.
18. Xiao-Fei Zhang, Le Ou-Yang, Xing-Ming Zhao, and Hong Yan. Differential network analysis from cross-platform gene expressiondata. Scientificreports, 6:34112, 2016.
19. Kaiser, A. and Schreiber, T., 2002. Information transfer in continuous processes. Physica D: Nonlinear Phenomena, 166(1-2), pp.43-62.
20. J. Dougherty, R. Kohavi, M. Sahami, et al. Supervised and unsupervised discretization of continuous features. In Machine learning: proceedings of the twelfth international conference, volume 12, pages 194–202, 1995.
21. Y. Yang and G. I. Webb. Discretization for naive-bayes learning: managing discretization bias and variance. Machine learning, 74(1):39–74, 2009.
22. P. E. Meyer, F. Lafitte, and G. Bontempi. minet: Ar/bioconductor package for inferring large transcriptional networks using mutual information. BMC bioinformatics, 9(1):461, 2008.
23. Mark EJ Newman. The structure and function of complex networks. SIAM review, 45(2):167– 256, 2003.
24. Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small world' networks. nature, 393(6684):440, 1998.
25. Reka Albert. Scale-free networks in cell biology. Journal of Cell Science, 118(21):4947–4957, 2005.
26. Albert-L´aszl´oBarab´asiandM´artonP´osfai.Network-Science.Cambridgeuniversity press, 2016.
27. Gang, X., Zhong, L., Xiwei, L., Fenghua, Z. and Dong, S., 2012. Service Science Management and Engineering: Theory and Applications.
28. Marinescu, D.C., 2016. Complex systems and clouds: a self-organization and selfmanagement perspective. Morgan Kaufmann.
29. Perez, C. and Germon, R., 2016. Graph Creation and Analysis for Linking Actors: Application to Social Data. In Automating Open Source Intelligence (pp. 103-129). Syngress.