DGPPIsAS :A Dynamic Global PPIs Alignment System

Sarwar Kamal[1], Shuxiang Xu[2], Sonia Farhana Nimmy[3], Mohammad Ibrahim Khan[4]
[1,4] Chittagong University of Engineering and Technology, Chittagong, Bangladesh,
[2] University of Tasmania, Australia

[3]BGC Trust University Bangladesh, BGC Biddha Nagar, Chandanaish, Chittagong, Bangladesh

Summary
In the view of Computational Biology, Computational Chemistry, Chemical Engineering, Molecular Biology, Biochemistry and Genetics Engineering, investigating proteins networks of separate species in a significant way is definitely one of the most important problems in ongoing evolutionary and systems biology research and experiments. PPIs networks enable to check the sequence of proteins in a unique fashion. Dynamic PPIs Alignment System (DPPIsAS) is an algorithm based system that enables to find out the proteins associated in a certain network. DPPIsAS used Protein Road Discovery (PRD) to determine the similar proteins that shares the interactions. After PRD, Protein Road Maintenance (PRM) is assured under several steps. How the proteins interact is checked by Canonical Correlation Analysis (CCA). Finally, the results are depicted based on the closeness, betweenness, average distance, and degree and edge betweenness.

Keywords:
Protein Road Discovery (PRD), Protein Road Maintenance (PRM), Canonical Correlation Analysis (CCA). DPPIsAS.

1. Introduction

In the view of Computational Biology, Computational Chemistry, Chemical Engineering, Molecular Biology, Biochemistry and Genetics Engineering, investigating proteins networks of separate species in a significant way is definitely one of the most important problems in ongoing evolutionary and systems biology research and experiments. PPIs networks enable to check the sequence of proteins in a unique fashion. In that case the biological network is very helpful because it provides appropriate information towards information interchange. PPIs networking are great to share the genes features of cell as feedback or transformation mode of network. Alignments of Protein sequences are also important subjects of PPIs networking as pair wise local or global alignment or multiple alignments.

PPI network is a formation of a graph that is an undirected weightless graph with set of vertices and edges. It is as similar as G (V, E) where, V denotes all the proteins of the chains or networks and E indicates the relationship between proteins as edges. Information loaded into DNA segments or genes transform though RNA to ultimate goal of proteins. A reliable and powerful network system is very essential to represents all the data set and information of PPIs. Protein-Protein Interaction Networks (PPIN) is precisely useful due to its presence helps to define all functionalities of living cell of plants, animals and insects. All the organic functionalities inside the cell are controlled by proteins. But it is very interesting that a single protein does not perform any duties of cell, instead a groups of protein used to perform the duties by making cooperation with each other by creating a bond which the bond is works as the path of networking. Million cells of contain millions of proteins. So their bonding of each other creates a very complex networking among them.

Proteins-Protein Interactions manage various transcriptional, signaling and metabolic mechanisms in cells [1]**.** Some research have narrated the PPIs in graphical representation [2-3]. The prime target of network alignment is to predict the best mapping among all nodes in given networks based on the similarity of the constituent molecules and their related proteins. Networks alignment and sequencing may be imposed for predicting conserved functional modules [4-5]. The network isomorphism outline tells that alignment is a NP-hard problems and sometimes it is really difficult to manage the interaction among proteins [6-9]. In this regards some heuristics models also developed to make the alignment easier and computationally feasible [10-15].

1.1 Why Global Network Alignment over Local Network Alignment

Various research have had addressed network alignment complication as before [8],[9],[10] and some distinction is noticed. Earlier we have defined that PPIs network is similar with undirected graph as G(V,E). for each relation in edges, the graph may be weightless or weighted. The prime objective of network alignment is to determine one or heterogeneous scaling between the terminals of the source network and for each scaling the respective set of marked edges. The scaling of source network may be partial (local) or complete (global).

Manuscript received February 5, 2015
Manuscript revised February 20, 2015

1.2 Local Network Alignment

Local Network Alignment is mapping technique that does not select all the nodes from the source network; instead a small segment of network will be select from source networks. Local network alignment enables to determine collective, unconnected nodes of homologous sub-graph from whole source input networks. As for example, suppose there are two source protein networks as P_1 and P_2, if a protein r_1 from network P_1 is scaled with and r_2 from network P_2, then r_1 and r_2 indicate to the common node to the common sub-graph as local network alignment and their exist a common edge between r_1 and r_2. Greater part of the local network alignment research follows the scaling as sequence similarity as the degree of metamorphic relation. But the pivotal drawback and difficulties is the duplicate overlap between scaled source protein networks. To address the problem Kolley et al **[16]** illustrate how BLAST homogenous scores and PPIs network information can use to fix conserved activity motifs and the algorithm is named as PathBLAST which is one of the pioneer algorithms in Local Protein Network Alignment. An update of PathBLAST is design **[19]** named as NetworkBLAST-M, addressed to fix preserved protein factors in different species. Consequently, the authors from **[18]** concentrate to reduce the duplication and deletion and the algorithm is named as MaWISh [Maximum Weight Induced Sub-graph]. The recent development on local network alignment **[17]** designed for multiple species under module structures which named as Graemlin.

1.3 Global Network Alignment

The process which scale complete input protein sequences to fix the best matching part of the given sequences is called Global Network Alignment (GNA). This always selects a unique scaling throughout the whole part of the source network sequences despite if there were locally minimal in few parts of the input networks. On the other hand, a Local Network Alignment (LNA) can select any locally best scaling for every local sub-area of similarity, even if there exist overlapping and mutually irregular. The important goal of Global Network Alignment is to fix an extensive scaling so that the length of respective similar sub-network is capitalized. In GNA, every node in source network is either matched to some terminal of the given network or completely gap as Indel (insert or delete). The similarity between Global Sequence Alignment (GSA) and Global Network Alignment (GNA) is that GSA continuously compares the genome segments to check the mutation between input sequences and GNA is used to check the interaction and to measure the mutation among distinct species. In later section various Global Network Alignment algorithm have been described. Some are PATH [20], GA [21], NATALIE [22], NetAlignBP, NetAlignMR [23], and PISwap [24] all focus on GNA and all of them only address the pairwise alignment problem. In our previous work [25] we also checked for local sequence alignment for DNA. Consequently, our last work [26] on global PPIs alignment is also a good analysis for global protein network alignment.

This research work has some important significance. Firstly, we have verified the similar proteins network interaction with protein road discovery (PRD) analysis. After PRD, it is very essential to maintain the interactions information for future road tracking. Protein Road Maintenance (PRM) method permits to keep the tracking of the system. PRD and PRM process are performed inside local proteins networks. Thirdly, Canonical Correlation Analysis (CCA) assesses the relationship amongst protein networks. Finally, over all global networks are measured using Destination Sequenced Protein Vectors (DSPV). This paper organized with seven sections including introduction. Section 2 narrates basics on global networks algorithms. Subsequent section describes the protein road discovery analysis. Section 4 illustrates protein road maintenance. Next section depicts the global protein networks inter-relation. Experimental results have showed at section 6. Last but not the least, concluding remarks have been described at section 7.

2. Fundamentals of Global Networks Algorithms

It is feasible to consider all source protein data set and their interactions as a formation and relations like formal graph. To define a graph (G) in mathematically it is easy to define that a graph is a mathematical object consisting of nodes, vertices or points (V) and **edges**, links or arcs (E) between pairs of nodes. V is also defied by $V(G)$ and V_G and E is defined as $E(G)$ and E_G ; $E \subseteq V \times V$. The protein networks graph size can be demonstrated as parameters: $n = |V|$, $m = |E|$ where n denotes the total number of vertices and m is the total number of edges. Suppose for a protein network graph G contains m edges then the degree of the graph is $\Sigma \deg(v) = 2m = 2|E|$. For a specific input protein data set it is possible to represents the proteins as a **complete graph K_n**, which is a simple graph constructed by n vertices with each connected to each of the others by an edge. Individual vertex in K_n has degree $n - 1$. Simultaneously, bipartite graph helps to clearly visualize the global multiple alignment. For any **bi-partite graph $K_{m,n}$** formed by two set of vertices, demonstrated with m vertices in one set and n in the other. Each of the m vertices is connected to each of the n vertices. Standalone network maintain an edge between two protein vertices if and only if the respective proteins pair interact with each other. In this regards for k PPIs networks, the complete

network or graph set are $G_k = \{G_1, G_2, G_3, \ldots\ldots\ldots\ldots G_k\}$ and $G_{i=} (V_i, E_i)$

3. Protein Road Discovery

The Protein Road Discovery (PRD) is an interactive process that dynamically established relations among proteins. It is designed specifically for use in multi-hop proteins network. Using PRD, the Protein network is completely self-organizing and mechanical. Network nodes interact to interchange chemical reactions and signal to allow relationship over multiple "hops" between nodes. As protein nodes in the network move about or join or leave the network, and as interactions among Proteins such as sources of interference change, all road finding are automatically determined and maintained by the PRD process. Since the number or sequence of intermediate protein hops needed to reach any destination protein nodes may change at any time, the resulting protein network structure may be quite rich and rapidly changing. The PRD process allows protein nodes to automatically discover a source road across multiple protein network hops to any destination in the global protein network alignment. Information exchange among proteins sent then carries in proteins where information keeps the track with sequence.

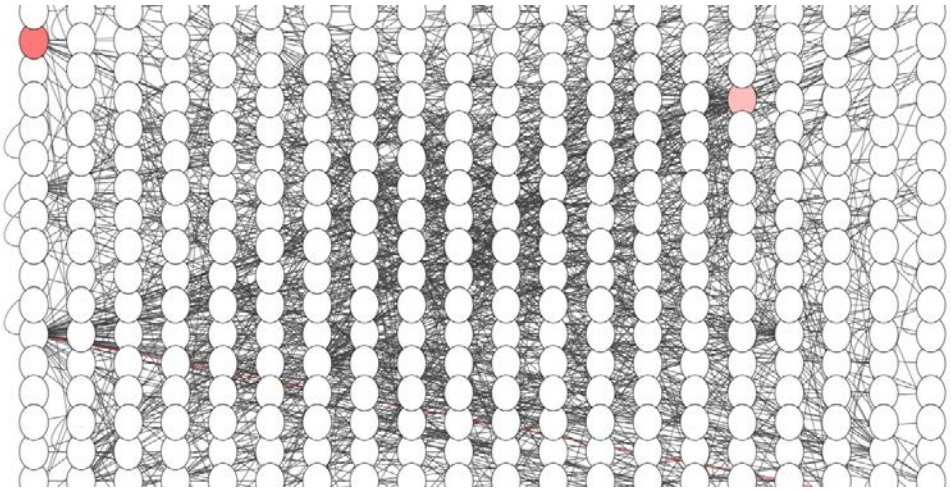


Figure 1: Large protein networks with thousand of nodes

A protein network may have huge protein nodes. Suppose a protein network (Figure 1) comprise thousands of protein nodes. Protein Road Discovery is the process by which a source node **S** interacts to a destination node **D** and creates a path form source to destination (Figure 2). Basically PRD is used when a source node or any protein node wants to make a relationship with any other protein nodes that is a destination node. A protein node first keeps the track of the destination node by placing information of destination in source node header and desired path cost. Generally, S will find a perfect source road by searching its Road Cache of road previously learned. However if it fails, PRD automatically find a new road to destination D.

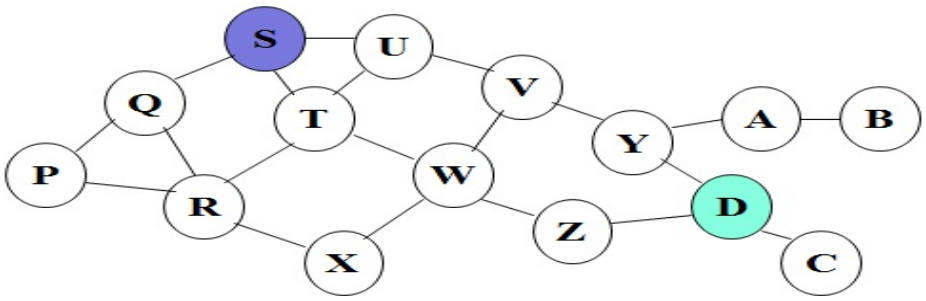


Figure 2: Interactions among proteins, started from source S to destination D

To start the Road Discovery, **S** initiates Road Request (RREQ) information to its neighbor nodes and all network nodes accept this interaction initiated request in the range of S. Each and every RREQ message determines the initiator and target of the Road Discovery, and also contains a distinct request id, determined by the initiator of the request. Consequently, each RREQ also maintain a record list that contains the address of each internal node through which this fixed copy of the RREQ message has been forwarded (Figure 3).

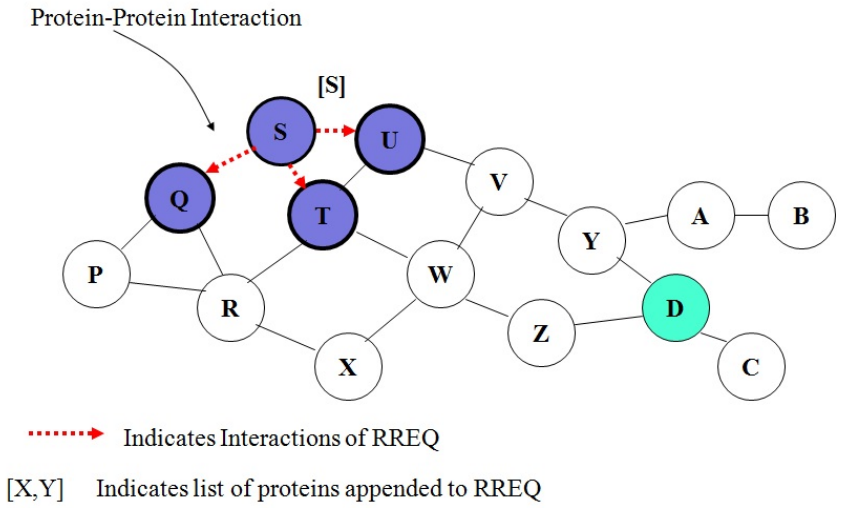


Figure 3: Broadcasting the messages from source nodes to remaining nodes where Source S propagate the interactions among neighbors nodes.

The process is continuous until the propagation reached to destination node (Figure 4). Each source RREQ contains source proteins node id, current positions of source node, destination id, last known sequence number and broad cast id. Besides, initiate a timer to get wait reply. Intermediate nodes check the source id and broadcast id of RREQ. If already get the response it will discard the response, if not, set up a reverse road for the source node and increase the RREQ hope count. Finally, broad cast the RREQ to all of its neighbors.

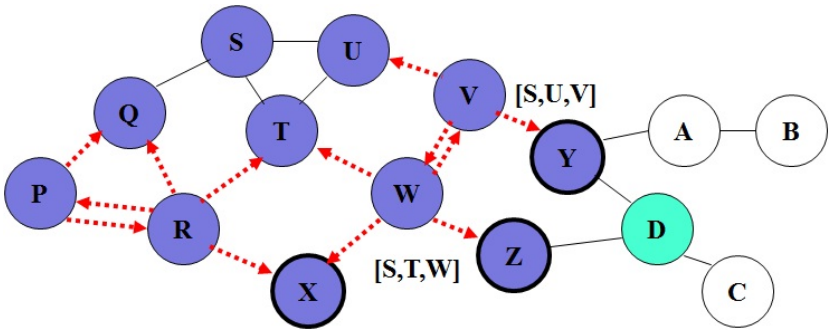
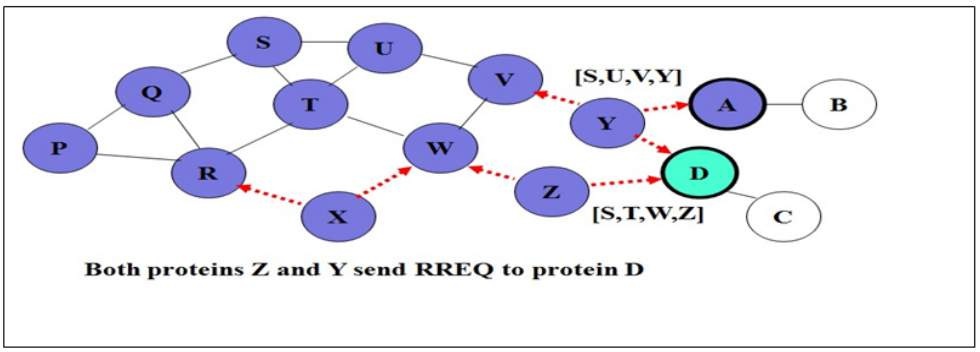


Figure 4: Here protein T receives interactions RREQ from W and R, however T does not interact with any other protein because T has already forwarded RREQ once.

When another node receives a RREQ and if it is the target of the road discovery, it returns a Road Reply (RREP) message to the starting node of road discovery.

Whenever starting node receives a RREP, it caches this road in its Road Cache (RC) for transferring information to the destination (Figure 5(a) and 5(b)).



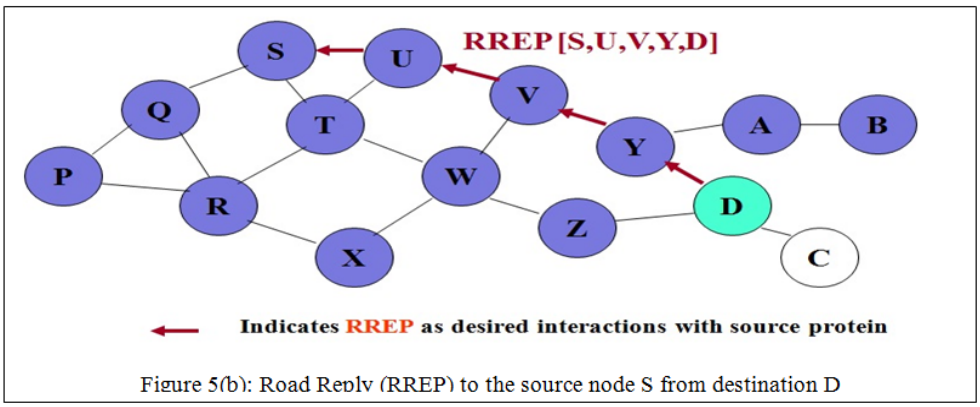


Figure 5(b): Road Reply (RREP) to the source node S from destination D

4. Protein Road Maintenance

Protein Road Maintenance (PRM) is the method that source node **S** is able to identify, during using a source road to destination **D**, if the network structure has altered such that it can no longer use its road to the destination **D** due to the previous road no longer works. When Road Maintenance notice a source road is broken, source **S** can try to maintain any other road it happens to know to **D**, or can fixed Road Discovery again to find an alternative road. Road Maintenance is used only when source **S** is actually sending information to destination **D**.

5. PPIs Network Correlation Measures

The relationships among proteins or PPIs can be possible to estimates. The best estimation is the Canonical Correlation Analysis (CCA) [27]. Various proteins vectors as *a* and *b* in r and t dimensional space. Suppose X and Y are set of proteins. X= (x_1, x_2, x_3, x_4,. . . . . ., x_p) and Y= (y_1,y_2,. . . . . .y_q). The correlation between X and Y is ρ. So ρ=cor ($a^{/}Xb^{/}Y$). So we get in equation 1

$$\rho = \frac{a^{/}\sum_{XY} b}{\sqrt{a^{/}\sum_{XX} a}\sqrt{b^{/}\sum_{YY} b}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Where

$\sum_{XX} = \mathrm{E}[(X-\mu_X)(X-\mu_X)^{/}]$ and $\sum_{YY} = \mathrm{E}[(Y-\mu Y)(Y-\mu Y)^{/}]$ are the covariance of X and Y respectively. $\sum_{XY} = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)^{/}]$ is the cross covariance of X and Y.

Two protein nodes X and Y always interact with parent node *j*. the neighbors proteins x_1, x_2, x_3,……..x_p and y_1,…..y_q. The interaction among x(j) and y(j) as cross correlation $\rho_{xy}(j)$ in equation 2

$$\hat{\rho}(j) = \frac{X(j)^T \sum_{XY} \hat{y}(j)}{\sqrt{X(j)^T \Sigma_{XX} X(j)}\sqrt{Y(j)^T \Sigma_{YY} \hat{Y}(j)}} \quad \ldots.(2)$$

The individual interaction are narrated in equation 3 to 5

$$\sum_{XX} = \frac{1}{n}\sum (X(j)-\mu_X)(X(j)-\mu_X)^T \text{.....(3)}$$

$$\sum_{YY} = \frac{1}{n}\sum (Y(j)-\mu_Y)(Y(j)-\mu_Y)^T \text{.....(4)}$$

$$\sum_{XY} = \frac{1}{n}\sum (X(j)-\mu_X)(Y(j)-\mu_Y)^T \text{.....(5)}$$

Where

$$\mu_X = \frac{1}{n}\sum X(j)$$

And

$$\mu_Y = \frac{1}{n}\sum Y(j)$$

6. Results and Implementation

We develop a system based on Java to measure the interactions. Consequently, we use Cytoscape [**28**], a proteins analysis tool for PPIs measurement. The interactions are very complex to measure and proteins always change its positions to get proper interaction. We use eukaryotes baker's yeast and human data set for PPIs networks alignment (Section 6.1). On the other hand, prokaryotic PPI networks, of bacteria and viruses, respectively illustrated subsequently (Section 6.2 and Section 6.3). We illustrate that DPPIsAS expose large networks that are the parts of total network. We also align proteins of same function. The interactions among proteins are clearly visible at above figures. We use Java springs to enable such relationships among all protein data. At the same time we have compare results of our system with Cytoscape. We significantly noticed that our system outperforms Cytoscape at various aspects especially at time and space complexity. Besides, we minimize the complexities in some other sides such as easy PPI, grade measurements, protein orientation identification and structure formation. Cytoscape is a generalized tool and it is difficult to handle all functionalities rapidly. Some molecular identification plays important role in human body as well as animals and plants nutrition's. To measure instant activity and actions, Cytoscape is slow and imperfect.

6.1 Global PPIs alignment for Yeast-Human

We use our system to align yeast *S.cerevisiae* global PPIs network [29]. Similarly, global protein network of human [30]. The yeast has 2, 45,788 interactions amongst 37,876 proteins and the latter has 3,67,1,456 interactions amongst 49,141 proteins. To design alignments, we execute each and every possible 2^7 = 128 combinations of the seven topological and sequence measures such as Average shortest path to number of directed nodes, Closeness centrality versus self loops, Clustering coefficient to number of direct edge, Edge Count to neighborhood Connectivity, In degree versus out degree, Neighborhood Connectivity versus Topological Connectivity, Neighborhood to closeness connectivity certain genes and Partner of Multi-edge Connectivity to neighborhoods connectivity. To account for a possible randomness in the algorithm caused by randomly breaking ties, we run each of the 32 tests 30 times and compute the statistics. The maximum edge effect is 33.55% including 7383 PPIs alignment among 21000 proteins. The alignment is performed by scoring proteins interactions pair values. The measurements of these interactions are vital part of the alignment. We termed it as sequencing 1 (Table 1). Nonetheless, applying only signatures does not satisfy all possible combinations and results to dissimilar matchings for dissimilar executions with mean edge correctness (EC) (equation 6) of 29.88% and the standard deviation of 0.55% for 555 executions.

$$EC = \frac{|\{(u,v) \in E_1 \wedge (f(u), f(v)), \in E_2\}|}{|E_1|} \text{.....(6)}$$

Here there are two global alignment networks as $G_1(V_1;E_1)$ and $G_2(V_2;E_2)$, where $V_1 < V_2$, as a total injective function $f : V1 \rightarrow V2$. The experimental result shows that such alignment is not always perfect due to lots of interactions (Figure 6).

It is also proven that only BLAST edge value does not satisfy the relationships with different networks for different executions with the mean edge correctness of 23.45% and standard deviation of 0.55% for 555 executions. BLAST shows best outcomes for sequencing with the EC values of 0.12% and it constructed by 4563 aligned interactions amongst 7893 proteins. We termed this as sequencing 2. Besides, when we impose protein signatures, relationship degrees, protein cluster coefficients and calculated BLAST scores, we noticed that the alignment are 99.98% similar amongst 555 executions with edge corrections 34.77% grouped by 7893 aligned interactions amongst 9987 proteins. We termed this as sequencing 3.

In Addition, we measure structural nature of sequencing 1, 2 and 3 by testing the amplitude of their largest common connected sub-sequences *(LCCSSs)*. LCCSSs are the largest connected string graph where each of the aligned protein networks has fixed size. We verify this network due to the preferences of large network alignment and contiguous sub-sequences instead of a number of small disconnected network regions. The extent of the LCCSSs in Sequencing 1 is 5,345 nodes and 9,536 edges, that is

about 87:7% and 31:5% of the yeast's nodes and edges, respectively. The LCCSSs disclose by Sequencing 2 has 6,345 nodes and 8,234 edges. Sequencing 3, has the LCCSSs with 8,853 nodes and 12,678 edges.

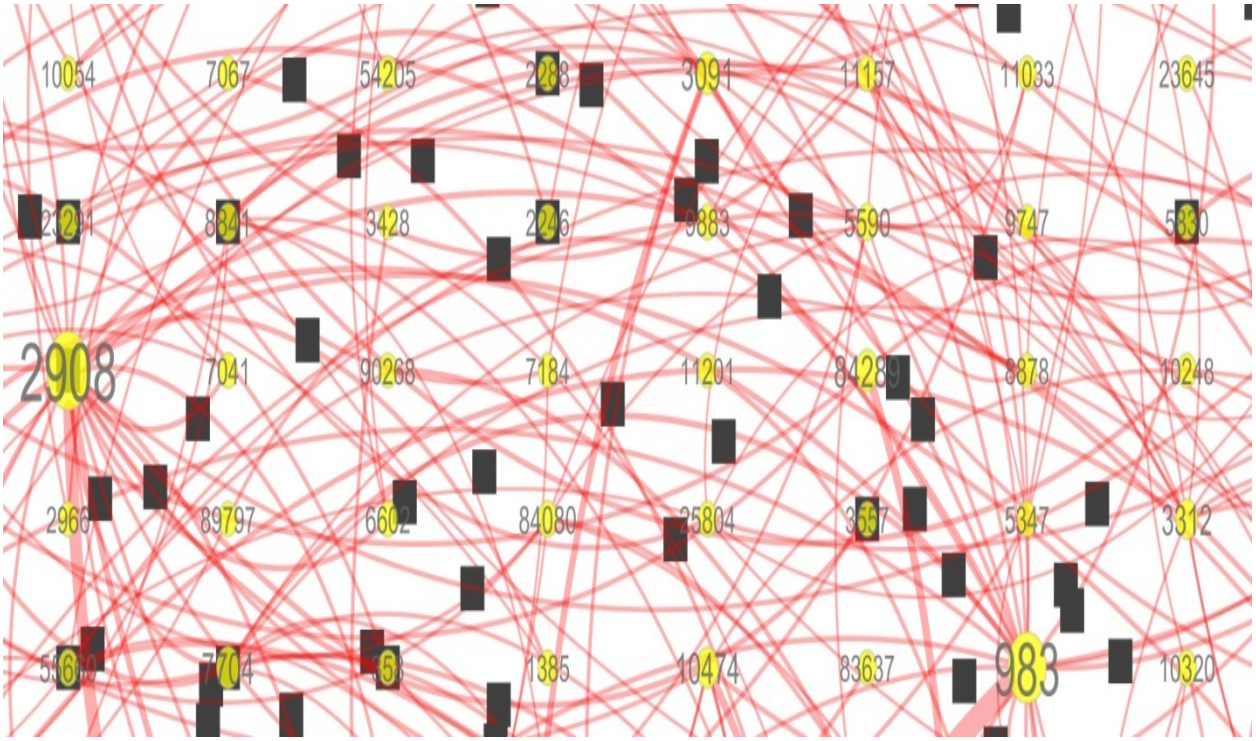


Figure 6: Protein Networks with huge number of edges.

	Sequencing 1	Sequencing 2	Sequencing 3
>=1	66:45% (10^{-11})	23:55% (10^{-6})	19:33% (10^{-9})
>=2	76:23% (10^{-2}	23:51% (10^{-4})	27:56% (10^{-9})
>=3	78:97% (10^{-5}	45:24% (10^{-3})	8:07% (10^{-9})
>=4	1.55%	6:12% (10^{-4})	7:04% (10^{-9})
>=5	2.33%	9:71% (10^{-6})	2:97% (10^{-9})
>=6	12.45%	1:97% (10^{-9})	0.23:71% (10^{-9})

Table 1: Sequencing Edge correctness

We design proper prediction to prepare "Biological Process" forecasting for botanical and zoological protein data set. We have taken all biological data to verify the yeast-human alignment to determine protein pairs where one protein is annotated with other protein in biological annotated terms. In this regards we easily moves the annotations from the annotated protein to the one that is not annotated. The experiments for data forecasting and predictions for "Atomic and Biological Function" and "Nuclear Component" are made in the same directions. We design "Biological Process" predictions for 235 anthropological proteins, "Atomic and Biological Function" predictions for 345 human proteins and "Nuclear Component" predictions for 897 human proteins. Similarly, the experiments under yeast data set, we construct "Biological Process" predictions for 896 proteins, "Atomic and Molecular Function" guess for 1367 proteins and "Cellular Component" guess for 987 proteins.

6.2 Aligning Bacterial PPI networks

We measure the uplifted backbone for practical interconnected networks of *E. coli* that merge with good features of empirical Protein-Protein Interactions and summing data [31]. The data set contains 9,564 interconnected relations amongst 4,321 proteins. We assess the top estimation of *C. jejuni* proteins network constructed by 14,987 interconnected relations amongst 2,678 proteins generated by yeast-2-hybrid analysis [32]. We have compare M-GRAAL and our system and noticed that our system perform 30% better and accurate result than that of M-GRAAL. We have collected all the protein sequences and Gene Ontology annotation data for these bacteria from the National Center for Biotechnology Information (NCBI) and European Bioinformatics Institute (EMBL-EBI) website of December 2014.

6.3 Aligning viral PPI networks

We also measure herpesviral protein-protein interaction in global networks of five herpesviruses: as first one is varicella-zoster virus (VZV), second one Kaposis sarcoma- associated herpes virus (KSHV), third one is herpes simplex virus 1 (HSV-1), next one is murine cytomegalovirus (mCMV) and finally Epstein-Barr virus (EBV) [33]. However, these global networks may have some limitations in the sense of proper interaction and maintainers. We have designed state of the arts structural global protein networks alignment and compare with M-GRAAL algorithm and noticed that our developed system perform far better result in the matter of path discovery.

7. Conclusion

Sequencing global proteins networks of several species is anticipated to be a helpful automated system or tool. DGPPIS allows comparative analysis of different data of different species. Besides, it enables to discover the pathway in large global networks. PPIs are very complex matter due to huge number of interconnected proteins. In the light of upcoming growth of gigantic number of molecular and other functional network data set, protein sequencing and interactions methods are anticipated to become progressively priceless in elaborating our consideration and regulation huge global proteins networks.. In future, we will measure the Mutual information test (MIT) score for global protein networks and gene regulatory networks. MIT score tells the actual nature of the gene and protein networks. Gene regulatory network is a complex molecular structure that generated from gene expressions. MIT score will strictly check the gene expressions formation and topology.

References

[1] A. Zhang, Protein Interaction Networks: Computational Analysis. New York, NY, USA: Cambridge University Press, 1st edition, 2009.

[2] A.L.Barabasi, Z.N.Oltvai, Network biology: understanding the cell's functional organization. Nat Rev Genet 5: 101–113,2004.

[3] M.E. Cusick, N.Klitgord, M.Vidal, D.E.Hill, Interactome: gateway into systems biology. Hum Mol Genet 14 Spec No. 2: R171–181,2005.

[4] R.Sharan R, S.Suthram, R.M.Kelley ,T. Kuhn ,S. McCuine , Conserved patterns of protein interaction in multiple species. Proc Natl Acad Sci USA 102: 1974–1979,2005.

[5] D.Park , IsoBase: a database of functionally related proteins across PPI networks. Nucleic Acids Res 39: 295–300,2011.

[6] G.Klau, A new graph-based method for pairwise global network alignment. BMC Bioinformatics 10: S59,2009.

[7] F.Ay ,SubMAP: aligning metabolic pathways with sub-network map- pings.J Comput Biol 18: 219–235,2011.

[8] J.Flannick,Græmlin: general and robust alignment of multiple large interaction networks. Genome Res 16: 1169–1181,2006.

[9] C.S.Liao, IsoRankN: spectral methods for global alignment of multiple protein networks. Bioinformatics 25: i253–258,2009.

[10] O.Kuchaiev, N. Przulj, Integrative network alignment reveals large region of global network similarity in yeast and human. Bioinformatics 27: 1390–1396,2011.

[11] V.Memisevic, N.Przulj,C-GRAAL: common-neighbors-based global GRAph ALignment of biological networks. Integr Biol 4: 734–743,2012.

[12] G.Ciriello,.AlignNemo: alocal network alignment method to integrate homology and topology. PLoS ONE 7: e38107,2012

[13] H.T.Phan, M.J.Sternberg, PINALOG: a novel approach to align protein interaction networks–implications for complex detection and function prediction.Bioinformatics 28: 1239–1245,2012.

[14] P.Csermely, T.Korcsmaros, H.J.Kiss, G.London, R.Nussinov Structure and dynamics of molecular networks: A novel paradigm of drug discovery: com prehensive review. Pharmacol Ther 138: 333–408,2013.

[15] R.Sharan, S.Suthram, R.M.Kelley T.Kuhn, S.McCuine , Conserved patterns of protein interaction in multiple species. Proc Natl Acad Sci USA 102: 1974–1979, 2005.

[16] B.P. Kelley et al. Pathblast: a tool for alignment of protein interaction networks. Nucleic Acids Res, 32(Web Server issue):W83-8, 2004.

[17] J. Flannick, A. Novak, B.S. Srinivasan, H.H. McAdams, and S. Batzoglou. Graemlin: general and robust alignment of multiple large interaction networks. Genome Res, 16(9):1169-81, 2006.

[18] M. Koyuturk, A. Grama, and W. Szpankowski., Pairwise local alignment of protein interaction networks guided by models of evolution, Proc of the 9th International Conference on Research in Computational Molecular Biology (RECOMB), 2005.

[19] R. Sharan, S.Suthram, R.M.Kelley,T. Kuhn,S. McCuine, P.Uetz, T.Sittler,R.M. Karp, and T.Ideker,Conserved patterns of protein interaction in multiple species. Proceedings of the National Academy of Sciences of the United States of America, 102(6), 1974–1979,2005.

[20] M. Zaslavskiy, F.Bach, J.P.Vert, A path following algorithm for the graph matching problem. IEEE Trans Pattern Anal Mach Intell,31(12):2227-2242,2009.

[21] M.Zaslavski, F.Bac J.P.Vert, Global alignment of protein-protein interaction networks by graph matching methods. Bioinformatics, 25(12):i259-i267, 2009.

[22] G.Klau, A new graph-based method for pairwise global network alignment. BMC Bioinformatics 2009, 10(Suppl 1):S59,2009.

[23] M. Bayati, M.Gerritsen, D.F.Gleich, A.Saberi, Y.Wang, Algorithms for large, sparse network alignment problems,Ninth IEEE International Conference on Data Mining IEEE, 705-710,2009.

[24] L. Chindelevitch L, C.S.Liao, B.Berger, Local optimization for global alignment of protein interaction networks. Pac Symp Biocomput ,123-132,2010.

[25] M. I. Khan, M.S. Kamal, RSAM: An Integrated Algorithm For Local Sequence Alignment, Archives Des Sciences, Vol 66, No. 5; May 2013.
[26] M.I.Khan and M.S.Kamal, Chapman-Kolmogorov equations for Global PPIs with Discriminant-EM, International Journal of Biomathematics, Vol. 7, No. 5 , World Scientific Publishing Company ,DOI: 10.1142/S1793524514500533, 2014.
[27] H. Hotelling, Relations between two sets of variates. Biometrika 28: 321-377,136.
[28] P.Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B.Schwikowski, T. Ideker, Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks, Genome Res.13: 2498-2504,2003.
[29] S.Collins, P.Kemmeren, X.Zhao, J.Greenblatt, F.Spencer, F.Holstege, J.Weissman, and N.Krogan, Toward a comprehensive atlas of the physical interactome of saccharomyces cerevisiae. Molecular and Cellular Proteomics, 6:3, 439–450, 2007.
[30] P.Radivojac, K.Peng, W.T.Clark, B.J.Peters, A.Mohan, S.M.Boyle, M.S.D, An integrated approach to inferring gene-disease associations in huma, 2008.
[31] J.M.Peregrin-Alvarez, X. Xiong, C.Su, and J.Parkinson,,The modular organization of protein interactions in escherichia coli. PLoS Comput Biol, 5(10), 2009.
[32] J.Parrish, J.Yu, G.Liu, J.Hines, J.Chan, B.Mangiola, H. Zhang, S.Pacifico, F.Fotouhi,,V. DiRita, IT.deker, P.Andrews, and R.Finley, A proteome- wide protein interaction map for campylobacter jejuni. Genome Biology, 8(7), R130,2007.
[33] F.Fossum, C.Friedel,,S.Rajagopala, ,B. Titz, A.Baiker, T.Schmidt, T.Kraus, T.Stellberger, C.Rutenberg, S.Suthram, S.Bandyopadhyay, D.Rose, A.von Brunn, , M.Uhlmann, C.Zeretzke, , Y.A.Dong, H. Boulet, M.Koegl, S.M.Bailer, U. Koszinowski, T.Ideker, P.Uetz, R.Zimmer, and J.Haas, Evolutionarily conserved herpesviral protein interaction networks. PLoS Pathog, 5,2009,

Md.Sarwar Kamal received the B.Sc (Hons) in Computer Science and engineering from University of Chittagong Bangladesh in 2009 and M.Sc (Engineering) from Chittagong University of Engineering and Technology at 2014. He is working as lecturer at BGC Trust university of Bangladesh. His research interest are Bioinformatics and Data Mining.

Dr.Shuxiang Xu is currently a lecturer of School of Engineering and ICT, University of Tasmania, Australia. He received a BSc from University of Electronic Science and Technology of China (1986), China, a MSc from Sichuan Normal University (1989), China, and a PhD in Computing from University of Western Sydney (2000), Australia. He received an OPRA award from the Australian government in 1996 to research his Computing PhD. His current interests include the theory and applications of Artificial Neural Networks, Genetic Algorithms, and Data Mining.

Sonia Farhana Nimmy is Faculty member of BGC Trust University Bangladesh. She has completed her M.Sc in Computer Science and Engineering from University of Chittagong at 2013. Her research interest includes Bioinformatics and Data Mining.

Dr. Mohammad Ibrahim Khan received the B.S. degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET), Bangladesh in 1999. He received M.S. degree in Computer Science and Engineering from the same University in 2002. He received his Ph.D. degree in Computer Science and Engineering from Jahangirnagar University in 2010. Since 1999, he has been serving as a faculty member in the Department of Computer Science and Engineering at Chittagong University of Engineering & Technology (CUET), Chittagong, Bangladesh.. His research interest includes Digital Image Processing, Graph Theory, Cryptography, Digital Watermarking, Multimedia Systems, and Digital Signal Processing.