ISSN 2278-3091
Volume 8, No.4, July – August 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse42842019.pdf
https://doi.org/10.30534/ijatcse/2019/42842019

Alignment of Protein Interaction Networks and Disease Prediction: A Survey

Anooja Ali[1], Vishwanath R. Hulipalled[2], S. S. Patil[3], Raees Abdulkader[4]
[1]REVA University, India, anoojaali@reva.edu.in
[2]REVA University, India, vishwanath.rh@reva.edu.in
[3]University of Agricultural Sciences, India, spatiluasb@gmail.com
[4]IBM India Pvt Limited, India, raeeskader@gmail.com

ABSTRACT

Proteins interact each other to perform many cellular activities. These interactions can be considered as Protein Protein Interaction networks (PPI). Interacting proteins form protein complexes. Mapping nodes between networks is denoted as alignment. The main intention of network alignment approach is to identify the protein complexes, which in turn helps to identify the functionality of protein complexes in various cellular systems. These interactome units form the conserved pathways between the networks. So network alignment requires lot of attention and several algorithms and techniques have been proposed to address this. The study of PPI is widely recognized to know more about the underlying complex disease because proteins associated with any disease get connected and form subgraphs or pathways. In this paper, the authors compared the various aligners, the performance evaluation metrics, the common databases used for PPI evaluation and the importance of PPI network in biomedical research.

Key words : Alignment, Biological Similarity, Complex Diseases, Network, Protein Complexes, Topological Similarity.

1. INTRODUCTION

Biomolecules are produced by the cells of a living being. The major biomolecules that play an important role in any living organism are proteins. Proteins are responsible for several functionalities for the existence of life. Proteins are made of linear sequence of amino acids. Proteins interact each other and this interaction can be marked as a network, also known as PPI network [1]. In PPI network proteins corresponds to vertices and the interactions corresponds to edges.

Studies on the structure of biological networks have obtained wide range of significance. The exponential increase in biological data provides lot of challenges to research strategies. Comparison between two networks can be done by graph alignment [2]. Network alignment is the equivalent to subgraph isomorphism [3]. Subgraph isomorphism is NP complete, means the performance of the algorithms developed so far were not completely accurate. If interaction between proteins are for a short period of time to perform any specific biological activity and later they dissipate, they are transient PPI [4]. Interactions may vary according to several factors like stimuli, time and cellular features. Interactions also vary according to the location of protein. Proteins while interacting with other proteins perform multiple functionalities [5].

Similarity between two graph structures can be identified by evaluating their topology [6]. This process can be referred as network alignment. Network alignment is divided into two categories, Global and local network alignment. Global network alignment increases the number of nodes by mapping nodes from one network to another network. . Local network alignment considers substructure mapping and global alignment considers the entire network and starts mapping [7].

Fig.1 shows the comparison between local network alignment and global network alignment. Shaded region indicates different alignment of nodes and are differentiated by different dashed lines. Alignment varies according to the methodology of alignment and the networks under consideration. Network alignment concentrates on evaluating topological or sequential similarity between sequences. Several wide advances have been proposed in past few years, most of the approaches are derived from the previous ones and few of them are widely different. The parameters that were used to evaluate solutions are varying over the years.

The knowledge of protein protein interactions can be used to diagnose diseases like auto immune disorders and cancer [8]. Several studies proved that rather than considering individual molecules, considering the interactions as a network is effective for many complex multi genetic diseases [9]. The

genes that are associated will have similar biological processes. Identifying the candidate genes based on these molecular interactions in PPI network helps to separate healthy and disease causing genes [10]. These hypothesis and relationships recommend an innovative paradigm for the analysis of complicated mutagenic diseases and cancers.

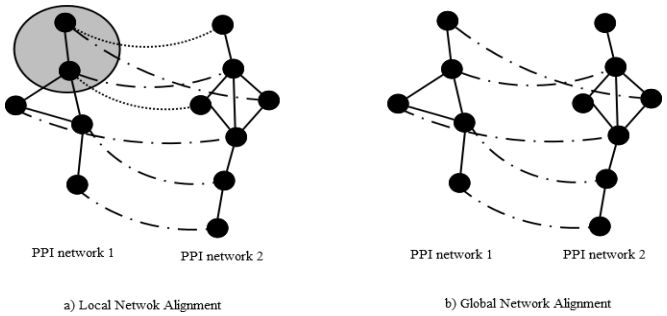

Figure 1: Local network alignment versus Global network alignment. (a) Shaded region indicates those nodes have different alignment (b) the two networks are aligned to a common subnetwork.

In this paper, the authors address the definition of network alignment and the assessment of networks with respect to topological and biological measures. The rest of the paper is organized as follows: Next section begins with the definition of network alignment, followed by the evaluation of different measures for alignment. Following this, the evaluation of popular aligners are performed and the role of network alignment in diagnosing complex diseases are discussed. The authors conclude the manuscript by addressing the various open research problems.

2. NETWORK ALIGNMENT

Network alignment aims at finding the similarity between two networks, *N1* and *N2*. Each network is a set of vertices and edges, $N1 = (V1, E1)$, $N2 = (V2, E2)$. The interaction between protein of the two network is denoted by edge $(u1, u2)$. Network alignment aims at calculating an injective function, f represented as

$$f(u1,u2)=\{u2, \text{ where } u1 \in N1 \text{ and } u2 \in N2\} \quad (1)$$

If it is possible to perform alignment for every node of the first network with a node in the second network, then the alignment is considered as completely defined. If alignments cannot be made for all the nodes, then the problem of network alignment is partially defined.

2.1 Topological Assessment

There are several matrix to evaluate the topology of network. Edge Correctness (EC) and Symmetric Substructure Score (S^3) are the common matrices. EC aligns edges from the first network to the second network. EC aligns edges from the first network to the second network. The percentage of edges aligned is represented in (2).

$$EC(N1, N2, f) = \frac{|f(E1) \cap f(E2)|}{|E1|} \quad (2)$$

S^3 is completely based on the composite graph structure obtained by integrating the two networks. It is independent of the network population. The common topological structures of PPI network is clique structure [11]. Cliques are fully dependent on connected subgraph. The drawback of this approach is that fully connected subgraphs cannot be satisfied in all cases with protein complexes. Dense subgraph is the second topological structure with the preliminary concept that in any closely compacted network, interactions within a protein complex are robust than the interactions between complexes [12]. Clique and dense subgraph assent with protein subgraph with priority and these methods penalizes for missing interactions. Combining the predictions based on overlap score and topological methods identifies the excelling prediction.

2.2 Biological Assessment

Topological assessment does alignment by identifying a matching node. Considering the shortcomings of topological assessments mentioned above, it is not good to rely only on topological assessment. Biological assessment is done with respect to the functional similarity between two proteins that are aligned. Functional similarity is evaluated by comparing the functionality of either the nodes or proteins.

PPI network is subdivided into various modules based on functionality or biological similarity [13]. If the topological and functional similarities are present, then they can be merged under a particular functional module [14]. This functional module is also called as larger node or as a super protein node. The entire PPI network is viewed in various levels. Fusion happens at each level between super protein nodes of the next level. Fusion combines gene ontology information by union. Merging of protein at each level is the main concern.

3. GLOBAL NETWORK ALIGNMENT

Global alignment aligns two different networks and derives a common sub network. Pairwise network alignment form the base for global network alignment. All the primary approaches identify motif by considering Basic Local Alignment Search Tool, called as BLAST bit scores and PPI network information [15]. Local network alignments fall of due to inconsistency. The main category of global aligners are IsoRank aligners and GRAAL based aligners.

3.1 IsoRank Aligners

IsoRank aligners are based on the compatibility between sequence order and topological structure [16]. It is based on

functional similarity and the similarity between proteins is estimated using a pattern similar to Google's PageRank algorithm [17]. Ranking is based on the number of links to that page. Consider 3 pages, P1, P2 and P3. PageRank will be transferred from one page to another page, if it is the target of outgoing link. Consider the link from P2 to P1 and P3 to P1. Rank calculation is given in (3).

$$\text{Rank}(P1) = \text{Rank}(P2) + \text{Rank}(P3) \quad (3)$$

IsoRank has the advantage of finding match for nodes that does not have a proper match. IsoRank aligners cannot categorize k-regular graphs. The possibility of over fitting increases in IsoRank with the increase in the number of true positives. IsoRankN generates aligned clusters of multiple networks based on spectral clustering [18].

3.2 GRAAL Aligners

Graph Alignment Aligners (GRAAL) are based on topological similarity. GRAAL approach is similar to seed and extend approach of BLAST. It identifies a seed vertex and from the seed it aligns vertices greedily on radius measurement. GRAAL aligners has different category, each with a set of advantages and disadvantages. Table I summarizes the different aligners, their methodologies and drawbacks. The different variants in GRAAL family are Hungarian GRAAL (H-GRAAL), Matching based Integrative GRAAL (MI-GRAAL) and Common Neighbors based GRAAL (C-GRAAL). In GRAAL aligners, any induced subgraph is indicated as a graphlet and the each graphlet consist of two to a maximum of five nodes. Graphlet orbits are automorphic. Fig.2 indicates the 73 orbits of graphlet.

Table I: Summary of different GRAAL Aligners

Aligner	Methodology	Drawback
GRAAL [7].	Topological similarity, Graphlet degree signature.	Cannot perform efficient vertex pairing
H-GRAAL [19].	Graphlet degree signature, Hungarian algorithm, Topological similarity	Higher runtime
MI-GRAAL [20].	Combination of multiple techniques	Complex while integrating various matrices
C-GRAAL [21].	Biological similarity, Matching based on common neighbor	Less effective vertex alignment

3.3 Other Aligners

When there are only two input networks, most of the existing aligners work pairwise. Several aligners are developed to perform alignment in cases where the number of input networks are more than two. In this section the authors discuss various aligners like Gr□mlin [22], PROPER [23], MAGNA [24] and SMETANA [25].

Gramiln is General and Robust Alignment of multiple interaction networks. Gr□mlin is the only algorithm that consider network data as insufficient for alignment. Gr□mlin is unique, because it need phylogenic information also for alignment. Gr□mlin is a two-step aligner and in first step pairwise scoring function for edges and nodes are developed. In second step iterations are done on initial alignment and continued till the final alignment is obtained. Gr□mlin consider four events for each node. They are protein duplication, protein deletion, protein mutation and paralog mutation.

PROPER is PROtein protein interaction network alignment based on PERcolation. The algorithm is based on the presence of identical pattern or motif across different species. This constitutes the sequence similarity. PROPER is a two-step algorithm where in the first step considers the sequence similarity. The highest sequence similarity pair is considered as the seed set for percolation algorithm. In the second step it considers the topological structure and seed sets generated from the initial step. This is the map percolation step and it aligns the remaining couples. MAGNA is the first aligner to use genetic algorithm to solve network alignment. MAGNA is maximizing the accuracy in Global network alignment. The principle of genetic algorithm states that the crossover of parents will generate a population of alignments and the fittest alignment will survive [26]. The major contribution of MAGNA is the crossover function. MAGNA consider permutations and alignments are mutually the same.

SMETANA is Semi- Markov random walk scores Enhanced by consistency Transformation for Accurate Network Alignment. The key feature of SMETANA is the calculation of maximum expected alignment. SMETANA is also a two phase aligner. In the initial step semi-Markov random walk model is developed and node correspondence scores are calculated. Similarity score matrix shows the possibility of aligning nodes of two networks. Nodes are aligned from one network to another if the neighbors are similar. If the neighbors of a node *n1* in network *N1* are similar to the neighbors of the node *n2* in network *N2*, then *n1* and *n2* can be mapped. Let *N (n1)* and *N (n2)* denotes the neighbors of *n1* and *n2* respectively. The quality of mapping between *n1* and *n2* can be represented as (4).

$$R(n1,n2) = \sum_{i \in N(n1)} \sum_{j \in N(n2)} \frac{1}{|N(i)N(j)|} Rij \quad (4)$$

Similarity between the nodes can be expressed with a cost function α [27]. Mapping, R can be expressed as in (5) and A corresponds the Eigen vector. The normalized sequence similarity is E. Few approaches integrates topological and biological similarity.

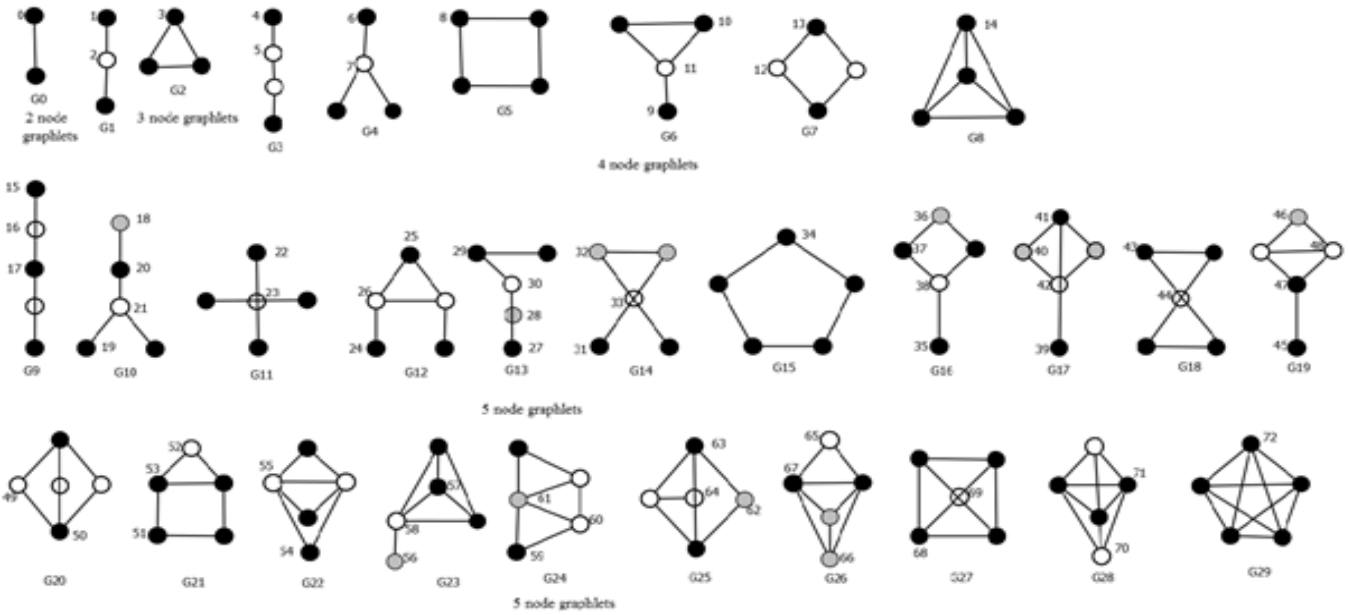

Figure 2: G0 to G29 indicates graphlets ranging from 2 nodes to 5 nodes.The 73 different orbits are labelled.

$$R = \alpha AR + (1-\alpha)E \quad (5)$$

Each protein in the PPI network is considered as a vector and if the neighbors of two proteins have the same degree, then these proteins are assumed to be similar. The biological similarity between proteins is based on the expect value (*E*-value).

4. PPI NETWORK AND DISEASES

PPI network consist of a group of proteins that interact each other in pathway analysis and functional complexes [28]. In case of cancer, cancer genes cluster together and they always co-occur in network. These disease genes form highly interconnected proteins and they take up tangential position of interactome [29]. PPI network demonstrates the specificity between proteins. Every PPI comprises of seed protein or candidate protein with direct neighbors [30].

The difference between healthy and disease states can be explored using PPI. Identifying the candidate gene for a disease along with their interaction with other protein plays a key role in finding the phenotype- genotype associations. So it is advised that the advanced way to know further about the disease is to investigate the gene candidate and its interacting partners [31]. Any mutation on the interacting protein create similar phenotype. Gene candidates estimated using PPI can be used to formulate the genetic backbone of the disease.

To identify the potential drug target of a disease, rather than considering highly connected protein, the less connected nodes in the network are more sensitive and need to be considered with high priority for discovering the target drug [32]. The correlation between genotype and phenotype is highly sophisticated, considering the network of interconnected genes will be an excellent hypothesis in identifying the molecular pathway of the disease phenotype. It is obvious to conclude that the genes which are closely associated in a network will have comparable biological suit. The possibility of more computational advent to differentiate disease and healthy genes are progressing.

Approaches for disease detection using PPI network falls in three categories. (1) Neighborhood clustering method- Neighbors of a node will be related in terms of topological or functional similarity and they will fall under the same cluster. Seed and extent method can be incorporated without clustering [33]. (2) Diffusion based techniques- Random walk is used for seed protein and the walker moves randomly to any protein neighbor. Frequency of visiting nodes in the network corresponds to the rank of protein. (3) Learning methods- These methods are mainly concentrated on graphlet degree signature or neural embedding. They initially capture all the known protein disease interaction and then work as downstream predictor by learning the representations as input.

4.1 Application of PPI in Cancer Detection

A precise systematic methodology for evaluating cancer proteins can yield several biological knowledge to disclose the molecular influence in cancer. The metastatic and non-metastatic tumors were classified using PPI network and gene expression panels. Cancer proteins has the tendency to interact with more proteins and they will become the prominent central hub in a network there by increasing the disease gene's participation and centrality. Cancer proteins have a high dimension of structural realm which helps them to be a part of many protein interaction. The global and local network features of the cancer proteins were studied in [34]. This study confirms that the structure of cancer and non-cancer genes are widely variant. So it can be concluded that the structure of protein network changes in the progression of cancer. By investigating the topological structures it is noticed that the essential protein or control protein will have higher centrality and higher Betweenness.

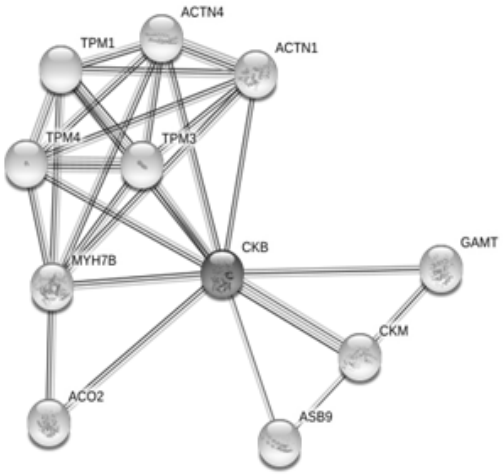

Figure 3: Interactions made by proteins CKB, ACTN1 and ACTN4 as essential proteins.

In figure.3 CKB, ACTN1 and ACTN4 are essential proteins. These hub genes are used to construct PPI. α-Actinins appear in various isoforms. Among these ACTN1 and ACTN4 increases the motility of cancer cells. So they are closely related to cancer malignasies [35]. It is observed that these proteins act as hub nodes in PPI and get co-expressed. Creatine Kinase B (CKB) is closely linked to cancer and there by having maximum edges with other nodes.

The distance between the essential proteins will be lesser and will have weaker clustering co-efficient [36]. The interacting proteins will be co expressed and form protein complexes. A new method for essential protein discovery outperform the regular measures like Betweenness centrality, Degree centrality, Closeness centrality, Subgraph centrality and Eigenvector centrality [37]. The research on identifying whether protein complex can be considered as essential protein is in progress. Few researches considered that essential proteins are also called as hub. Essential proteins will always interact with majority of neighbors. Few other researches pointed that essential protein complexes will also have essentiality feature. Few protein complexes are already known while others are obtained by considering the interaction of essential proteins within PPI network.

The connectivity patterns of the proteins associated with diseases are widely unexplored. It is found that for 60% of diseases, the proteins linked to the disease are over represented or dominant in any higher order network [38]. This gives insight to the fact that they have similar structural characteristics.

5. PPI DATASETS

Network alignment problem is NP hard in nature. The lack of a proper guaranteed aligner increases the complexity of alignment. Also, there is no final measure of alignment. Presence of noise in the dataset affect topological similarity. There are several databases available to perform alignment between networks. The different databases are NAPA bench [39], IsoBase [40], Protein Reference Database (HPRD) [41] and BioGRID [42]. IsoBase is the commonly used dataset for global network alignment. Table 2 denotes the four main species that are commonly under study. This is summarized based on Isobase dataset.

Table 2: Summary of Isobase dataset on four major species.

Species	No: of Nodes	No: of Edges
Saccharomyces cerevisiae (Yeast)	6,659	38,109
Drosophila melanogaster (Fly)	14,098	26,726
Homo sapiens (Human)	22,369	43,757
Mus musculus (Mouse)	24,855	452

Presence of noise is more in real time dataset. This leads to missing PPI and hinders network alignment. IsoBase can group proteins that are related to each other by some functional similarity. IsoBase generates IsoBase clusters. Most of the real world PPI have many shortcomings like missing interactions and they are error prone. So then came the necessity of a synthetic dataset. NAPA bench overcomes these issues by knowing the true alignment.

6. OPEN RESEARCH PROBLEMS

Literature survey indicates that there are a few areas for future work. The authors figure out a few open research problems about network alignment. Exploring these problems can possibly give better results. Research on PPI is sometimes bound to the dataset under consideration.

6.1 Weightage for Similarity Feature

Existing network aligners concentrate on heuristic algorithm to develop alignment. If we concentrate on topological alignment only, when the location of protein varies, the functionality also varies. Most of the aligners work in two steps with initially measuring sequence similarity using BLAST bit scores or E values. So the challenge is to develop an aligner that consider evaluation in terms of topological and biological similarities.

6.2 Parameterization for Different type of Network

For different network the evaluation is bound to find the correct parameterization for the alignment of these networks. The pre requisite for few aligners is the knowledge about the network. One such aligner is Gr□mlin. It need advance phylogenic information. Due to this reason it is not possible to compare any aligners which doesn't require advance information with them. The performance of topological matrices are advancing. This should not be at the cost of biological measures.

6.3 Larger Datasets

Most of the existing network alignment works are concentrated mainly on smaller networks. Comparison between larger networks is still an unexplored research area. Any pairwise alignment methods can be extended by greedy techniques for larger datasets.

6.4 Non-Availability of Positive Samples

The number of positive samples available for a disease is less than the number of unknown or negative samples. This will create an imbalance while training the model for any classification problem. All the existing databases provide information about candidate protein for a specific disease [42]. Non disease genes are very rarely present in any database.

6.5 Limiting the Degree of Interaction of Nodes in PPI

If two proteins X and Y have similar partners for interaction (A, B, C). If X and Y share only few interaction interfaces with binding sites of A, B and C. According to Triadic Closure Principle, X and Y interact each other, if the path connecting X and Y have length of two [43]. If the nodes have high degree of interaction, multiple shortcuts will be present in the network. To avoid bias due to multiple degree of nodes, few researches suggested to restrict interaction degree of every node by L3. Research is progressing to finalize the limiting degree of every node in PPI.

7. CONCLUSION

In this paper, the authors performed a survey on the aligners for PPI network. Following this the biological and topological matrices were evaluated. Later the popular state of art aligners are demonstrated with their methodologies and drawbacks. This review was focused on the importance of PPI network in biomedical research. Interpreting the topological structure of PPI by identifying the candidate gene and its interaction can act as a target for treatment of several auto immune disorders and cancer. Disease pathways are crumbled and moderately planted in PPI network. So learning the disease pathways over the set of predefined protein association can provide new advances in disease protein discovery.

Any research on network alignment aims to improve the performance evaluation metric and the time complexity. Datasets plays an important role in the performance of aligners. Identifying the protein species in these well studied species helps to uncover the complexes present in poor studied species. Creating a framework for measuring the different aligners and the evaluation metric along with the data set to be used will definitely help the biologist to categorize the best aligner to use.

REFERENCES

1. Elmsallati, Ahed, Abdulghani Msalati, and Jugal Kalita. **Index-based network aligner of protein-protein interaction networks,** *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 15.1 (2018): 330-336. https://doi.org/10.1109/TCBB.2016.2613098
2. Elmsallati, Ahed, Connor Clark, and Jugal Kalita. **Global alignment of protein-protein interaction networks: A survey**, *IEEE/ACM transactions on computational biology and bioinformatics* 13.4 (2016): 689-705. https://doi.org/10.1109/TCBB.2015.2474391
3. Cook, Stephen A. **The complexity of theorem-proving procedures**, *Proceedings of the third annual ACM symposium on Theory of computing*. ACM, 1971. https://doi.org/10.1145/800157.805047
4. Bhowmick, Sourav S., and Boon Siew Seah. **Clustering and summarizing protein-protein interaction networks: A survey**, *IEEE Transactions on Knowledge and Data Engineering* 28.3 (2016): 638-658. https://doi.org/10.1109/TKDE.2015.2492559
5. *Zhao, Bihai, et al.* **A new method for predicting protein functions from dynamic weighted interactome networks**, *IEEE transactions on nanobioscience* 15.2 (2016): 131-139. https://doi.org/10.1109/TNB.2016.2536161
6. Shui, Yong, and Young-Rae Cho. **Alignment of PPI Networks Using Semantic Similarity for Conserved Protein Complex Prediction,** *IEEE transactions on nanobioscience* 15.4 (2016): 380-389. https://doi.org/10.1109/TNB.2016.2555802
7. Kuchaiev, Oleksii, et al. **Topological network alignment uncovers biological function and phylogeny**, *Journal of the Royal Society Interface* 7.50 (2010): 1341-1354. https://doi.org/10.1098/rsif.2010.0063

8. Ali, Anooja, et al. **A review of aligners for protein protein interaction networks**, 2017 *2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*. IEEE, 2017. https://doi.org/10.1109/RTEICT.2017.8256879
9. Chuang, Han□Yu, et al**. Network□based classification of breast cancer metastasis,** *Molecular systems biology* 3.1 (2007): 140. https://doi.org/10.1038/msb4100180
10. Ideker, Trey, et al. **Discovering regulatory and signaling circuits in molecular interaction networks,** *Bioinformatics*18.suppl_1 (2002): S233-S240. https://doi.org/10.1093/bioinformatics/18.suppl_1.S233
11. Spirin, V., & Mirny, L. A. **Protein complexes and functional modules in molecular networks**, *Proceedings of the National Academy of Sciences of the United States of America*, *100*(21), 12123–12128. doi:10.1073/pnas.2032324100 https://doi.org/10.1073/pnas.2032324100
12. Ori, Alessandro, et al. **Spatiotemporal variation of mammalian protein complex stoichiometries**, *Genome biology* 17.1 (2016): 47. https://doi.org/10.1186/s13059-016-0912-5
13. Ji, Junzhong, et al. **Detecting functional modules based on a multiple-grain model in large-scale protein-protein interaction networks**, *IEEE/ACM transactions on computational biology and bioinformatics* 13.4 (2016): 610-622. https://doi.org/10.1109/TCBB.2015.2480066
14. Cho, Young-Rae, Woochang Hwang, and Aidong Zhang. **Efficient modularization of weighted protein interaction networks using k-hop graph reduction**, *Sixth IEEE Symposium on BioInformatics and BioEngineering (BIBE'06)*. IEEE, 2006. https://doi.org/10.1109/BIBE.2006.253347
15. Kelley, Brian P., et al. **PathBLAST: a tool for alignment of protein interaction networks**, Nucleic *acids research*32.suppl_2 (2004): W83-W88. https://doi.org/10.1093/nar/gkh411
16. Singh, Rohit, Jinbo Xu, and Bonnie Berger. **Global alignment of multiple protein interaction networks with application to functional orthology detection**, *Proceedings of the National Academy of Sciences* 105.35 (2008): 12763-12768. https://doi.org/10.1073/pnas.0806627105
17. Page, Lawrence, et al. **The PageRank citation ranking: Bringing order to the we*b*.** Stanford InfoLab, 1999.
18. Liao, Chung-Shou, et al. **IsoRankN: spectral methods for global alignment of multiple protein networks**, *Bioinformatics*25.12 (2009): i253-i258. https://doi.org/10.1093/bioinformatics/btp203
19. Pržulj, Nataša. **Biological network comparison using graphlet degree distribution**, *Bioinformatics* 23.2 (2007): e177-e183. https://doi.org/10.1093/bioinformatics/btl301
20. Kuchaiev, Oleksii, and Nataša Pržulj. **Integrative network alignment reveals large regions of global network similarity in yeast and human**, *Bioinformatics* 27.10 (2011): 1390-1396. https://doi.org/10.1093/bioinformatics/btr127
21. Memišević, Vesna, and Nataša Pržulj. **C-GRAAL: Common-neighbors-based global GRAph Alignment of biological networks**, *Integrative Biology* 4.7 (2012): 734-743. https://doi.org/10.1039/c2ib00140c
22. Flannick, Jason, et al. **Graemlin: general and robust alignment of multiple large interaction networks,** *Genome research* 16.9 (2006): 1169-1181. https://doi.org/10.1101/gr.5235706
23. Kazemi, Ehsan, et al. **PROPER: global protein interaction network alignment through percolation matching**, *BMC bioinformatics* 17.1 (2016): 527. https://doi.org/10.1186/s12859-016-1395-9
24. Saraph, Vikram, and Tijana Milenković. **MAGNA: maximizing accuracy in global network alignment**, *Bioinformatics* 30.20 (2014): 2931-2940. https://doi.org/10.1093/bioinformatics/btu409
25. Sahraeian, Sayed Mohammad Ebrahim, and Byung-Jun Yoon. **SMETANA: accurate and scalable algorithm for probabilistic alignment of large-scale biological networks,** *PloS one* 8.7 (2013): e67995.
26. Kanehisa, Minoru, et al. **KEGG as a reference resource for gene and protein annotation**, *Nucleic acids research* 44.D1 (2015): D457-D462.
27. Song, Bo, et al. **Aligning Multiple PPI Networks with Representation Learning on Networks**, *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018.
28. Rouillard, Andrew D., et al. **The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins**, Database 2016 (2016).
29. Jonsson, Pall F., and Paul A. Bates. **Global topological features of cancer proteins in the human interactome**, Bioinformatics 22.18 (2006): 2291-2297.
30. Daisy, A., and R. Porkodi. **Classification of Human Cancer Diseases Gene Expression Profiles using Genetic Algorithm by Integrating Protein Protein Interactions Along with Gene Expression Profiles,** *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*. IEEE, 2018.
31. Oti, Martin, et al. **Predicting disease genes using protein–protein interactions**, *Journal of medical genetics* 43.8 (2006): 691-698.
32. Kann, Maricel G. **Advances in translational bioinformatics: computational approaches for the hunting of disease genes**, *Briefings in bioinformatics* 11.1 (2009): 96-110. https://doi.org/10.1093/bib/bbp048

33. Ata, Sezin Kircali, et al. **Integrating node embeddings and biological annotations for genes to predict disease-gene associations**, *BMC systems biology* 12.9 (2018): 138.
34. van Dam, Sipko, et al. **Gene co-expression analysis for functional classification and gene–disease predictions**, *Briefings in bioinformatics* 19.4 (2017): 575-592.
35. Wu, K., Yi, Y., Liu, F., Wu, W., Chen, Y., Zhang, W. **Identification of key pathways and genes in the progression of cervical cancer using bioinformatics analysis,** Oncology Letters 16.1 (2018): 1003-1009.
36. Zhou, Jun, et al. **Identification of the spinal expression profile of non-coding RNAs involved in neuropathic pain following spared nerve injury by sequence analysis,** *Frontiers in molecular neuroscience* 10 (2017): 91.
37. Li, Min, et al. **United complex centrality for identification of essential proteins from PPI networks**, *IEEE/ACM transactions on computational biology and bioinformatics* 14.2 (2017): 370-380. https://doi.org/10.1109/TCBB.2015.2394487
38. Agrawal, Monica, Marinka Zitnik, and Jure Leskovec. **Large-scale analysis of disease pathways in the human interactome,** *Pacific Symposium on Biocomputing*. Vol. 23. 2018.
39. Sahraeian, Sayed Mohammad Ebrahim, and Byung-Jun Yoon. **A network synthesis model for generating protein interaction network families**, *PloS one* 7.8 (2012): e41474.
40. Park, Daniel, et al. **IsoBase: a database of functionally related proteins across PPI networks**, *Nucleic acids research*39.suppl_1 (2010): D295-D300.
41. Keshava Prasad, T. S., et al. **Human protein reference database—2009 update**, *Nucleic acids research* 37.suppl_1 (2008): D767-D772.
42. Stark, Chris, et al. **BioGRID: a general repository for interaction datasets,** *Nucleic acids research* 34.suppl_1 (2006): D535-D539.
43. Luo, Ping, et al. **Disease gene prediction by integrating PPI networks, clinical RNA-Seq data and OMIM data**, *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 16.1 (2019): 222-232. https://doi.org/10.1109/TCBB.2017.2770120
44. Pech, Ratha, et al. **Link prediction via linear optimization**. *arXiv preprint arXiv: 1804.00124* (2018).