www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 9 Issue 04 April 2020, Page No. 24989-24993
ISSN: 2319-7242 DOI: 10.18535/ijecs/v9i04.4465

Application of Centrality Measures for Potential Drug Targets: A Review

[1]Akula Chandra Sekhar[*], [2]Ch. Ambedkar

[1]Avanthi Institute of Engineering and Technology, Cherukupally(V), Vizianagaram District, India
[2]SRK Institute of Technology, Vijayawada, India

Abstract:
Protein-Protein Interactions (PPI) have important role in drug binding with the Proteins called drug targets. For identifying the potential drug targets there are different techniques. In this paper we are presenting application of Centrality Measures for identifying the drug targets. Centrality measure indicates importance of node in the graph or network. Protein-Protein Interactions for proteins which are involved in a particular disease are identified and centrality measures will be calculated based on the graph built suing the PPI interactions. Further the nodes which are playing crucial role will be identified using the various centrality measures and these drug targets can be used for drug discovery of a particular disease through insilico docking studies.

Keywords: Protein-Protein Interactions, PPI, Centrality Measures, Drug Targets.

1. Introduction

Human Protein-Protein Interactions Data is freely available and much research is going on in finding the missing interactions also. Using those interactions, identify the potential drug targets by constructing the PPI network. There are various databases available which are maintaining the PPI interactions. In the Protein-Protein Interaction Network, proteins are represented as vertices and interactions among them are as edges. Proteins are the representatives of the biological networks and they are realized only if the relationship between essentiality and topological properties such as the centrality measures, clustering coefficients, degree distribution, and community structures of the network are studied. Network centralities are used to rank elements of a network according to a given importance concept.

Table 1: List of PPI Databases

Name of the Database	***Description***
STRING https://string-db.org/	PPI information about 5090 Organisms 2000 million interactions
The Molecular INTeraction Database (MINT) https://mint.bio.uniroma2.it/	647 Organisms 26344 interactions
BioGRID https://thebiogrid.org/	70 Organisms 1.37 million interactions
Human Protein Reference database (HPRD) http://www.hprd.org/	Human PPI 41,327 interactions
Database of Interacting Proteins (DIP) https://dip.doe-mbi.ucla.edu	834 Organisms 81923 interactions

For identifying the potential drug targets there are different techniques. In this paper, we are presenting the application of Centrality Measures for identifying the drug targets. Centrality measures indicate the importance of a node in the graph or network.

There are fourteen different graph centrality measures such as degree, radiality, Stress, eccentricity, closeness, centroid values, shortest-path betweenness, Eigenvector, Page Rank, current- flow closeness, Katz status index, hits-hubs, current-flow betweenness and hits-authority are used for PPI Networks for finding potential drug targets and are defined as follows [1-19]. The nodes which are playing crucial role will be identified using the various centrality measures.

Degree $C_{dev}(v) = \deg(v)$

Radiality

$C_{rad}(v) = \frac{\sum_{w \in v}(\Delta_G + 1 - dist(v,w))}{n-1}$

Stress $C_{str}(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \sigma_{st}(v)$

Eccentricity

$$C_{ecc}(v) = \frac{1}{max\{dist(v,w) : w \in v\}}$$

Shortest path closeness

$C_{cfc}(v) = \frac{n-1}{\sum_{t \neq v} p_{vt}(v) - p_{vt}(t)}$

$p_{vt}(t)$ Equals the potential difference.

Closeness $c_c(v) = \frac{1}{\sum_{u \in v} dist(u,v)}$

Shortest path Betweenness

$c_B(v) = \sum_{s \neq t \neq v \neq V} \frac{\rho_{st}(v)}{\rho_{st}}$

Eigen Vector λ $C_{IV} = AC_{IV}$

Katz status index $c_k = \sum_{k=1}^{\infty} \propto^k (A^t)\, \vec{1}$

Betweennes

$C_{cfb}(vertex) = \frac{1}{(n-1)(n-2)} \sum_{s,t \in v} T_{st}(vertex)$

Where $T_{st}(vertex)$ equals the division of electrical current running over vertex v in a network

HITS-Hubs $C_{hubs} = AC_{auths}$

Centroid

$$C_{cen}(v) = \text{minimum}\{f(v,w) : v\{v\}\}$$

where γ v (w) indicates the number of vertices that are nearer to v than to w and f (v, w) = γ v (w) – γ w (v).

Page Rank

$C_{pr} = dpC_{pr} + (1-d)\vec{1}$

Where d is the damping factor and P is the transition matrix

Hits-authority $C_{auths} = A^T C_{hubs}$

2. Methodology

Protein-Protein Interactions (PPI) have important role in drug binding with the Proteins called drug targets. To identify the potential drug targets for the proteins involved in a disease, first we need to collect various proteins or genes causing a disease. The genes/proteins can be collected from various literature, experimental results, knowledge mining etc.,. After data cleaning, identify the Human Protein-Protein Interactions from various databases, sources. Using those interactions, PPI network can be build. Potential or influential proteins can be identified with Centrality measures. Human Protein-Protein Interactions Data is freely available and much research is going on in finding the missing interactions also. Using those interactions, identify the potential drug targets by constructing the PPI network. There are various databases available which are maintaining the PPI interactions. The nodes which are playing crucial role will be identified using the various centrality measures.

Ambedkar et al [20] have applied centrality measures on a PPI network of Diabetes Mellitus (DM). For construction of the PPI Network, they have collected the proteins involved in the DM from Jensenlab with 1020 vertices and 2891 edges. The authors have reported top 10 proteins from all the fourteen centrality measures.

In Kalyani et al [21] the authors have PPI with data

collected from different data sources like DMBase, MalaCards and Jensen Group (Jensenlab). They have constructed the PPI with 1027 vertices and 2950 edges. They reported top 10 proteins from 14 centrality measures. They also find out the proteins based on clustering coefficient.

NB Muppalaneni et al [22] collected 345 genes from new drug targets database, 19 genes from Experimental evidence and 1135 genes causing Autism disorder using text mining from Jensenlab disease database. Finally, they have constructed Protien-Protien Interaction (PPI) network with 74 interactions after eliminating parallel edges, self-loops of 54 proteins. Based on the centrality measures they listed top 10 potential drug targets for autism disorder.

Arun, PV Parvati Sai, et al [23] have identified 10 drug molecules for the potential drug targets based on the centrality measures for Type II Diabetes.

Xiao et al [24] have constructed a yeast PPI network using dynamic gene expression profiles. 1167 proteins are found from 1285 essential proteins downloaded from various databases like SGD, MIPS, DEG and SGDP. But the authors have considered only 7 centrality measures only.

Zhan Tong et al [25] have constructed a PPI network for Hepatocellular carcinoma (HCC). They have collected the genes causing HCC from DisGeNET and Human Protein Atlas Database. Protein-Protein Interactions of those proteins are collected from BioGRID database. The network is analyzed based on the network characteristics and genetic dependency scores. Then, using Support Vector Machine, they find out the drug targets for HCC.

Izudheen et al [26] applied centrality measures in the PPI network built using the oncogenes collected from GeneSignDB, OMIM. After elimination of duplicates, self edges, they build the network with 1234 cancer proteins. They used overlapping community structure detecting algorithm to identify the communities which have a prominent role in many mutation pathways.

Estrada et al [27] have used centrality measures on Yeast PPI. They presented a new technique, transform the PPI into a clique and calculate the centrality measures. Based on the ranking of the centrality measures essential proteins can be selected.

Lei et al[28] proposed new technique called Hyperlink Induced Topic Search (HITS) to apply to PPI network based on Gene Ontology and Gene Expression data. In this work, Undirected network is transformed into bidirectional using the topological features and biological information. Then, they found the essential proteins based on the new algorithm HSEP.

Sharma et al[29] used MatlabBGL package to calculate the centrality measures of Yeast PPI and Human PPI. They observed radiality and pagerank can identify the most prominent nodes in the PPI networks.

Mistry et al[30] analyzed the yeast PPI from DIP database along with gene expression dataset. They used a new technique DiffSLC, which is a weighted coexpression based degree centrality and eigenvector centrality.

The drug targets identified using Graph Centrality measures can be used for drug discovery of a particular disease through insilico docking studies and can be analysed with [31]

3. Tools

There are various open source and free tools available for calculating centrality measures

- CentiBiN (https://omictools.com/centibin-tool)
- ArcGIS (https://www.arcgis.com)
- Graph-tool (https://graph-tool.skewed.de)
- Social Network Visualizer (SocNetV) (https://socnetv.org)
- CytoScape (https://cytoscape.org)

4. Conclusions

There are various techniques have been used to find out the different diseases related candidate genes for potential drug targets. In this study, we highlighted how Centrality measures can apply for finding the potential drug targets along with correlation and clustering coefficient. Protein-Protein Interactions for proteins which are involved in a particular disease are identified and centrality measures will be calculated based on the graph built using the PPI interactions. The nodes

which are playing crucial role will be identified using the various centrality measures. Further, these drug targets can be used for drug discovery of a particular disease through insilico docking studies.

References

[1.] Zhang A. Chapter 4: Basic Properties and Measurements of protein interaction network. Protein Interaction Networks Computational Analysis. Cambridge University Press 2009; 33-49.

[2.] Jeong H et al. *Nature* 2001, **411:**41-42. [PMID: 11333967]

[3.] Hegde SR et al. PLoS Comput Biol 2008; 4:e1000237. [PMID: 19043542]

[4.] Koschützki D, Schreiber F: Comparison of Centralities for Biological Networks. Proc German Conf Bioinformatics (GCB'04), Volume P-53 of LNI 2004:199-206

[5.] Kranthi T et al.. Mol BioSyst 2013; 9:2163-7. [PMID: 23728082]

[6.] Stelzl U et al. Cell 2005; 122:957-68.[PMID: 16169070]

[7.] Chen J et al. BMC Bioinformatics 2009; 10:73.[PMID: 19245720]

[8.] Newman, M.E.J. 2010. Networks: An Introduction. Oxford, UK: Oxford University Press.

[9.] Borgatti, Stephen P.; Everett, Martin G. (2006). "A Graph-Theoretic Perspective on Centrality". Social Networks (Elsevier)28 : 466–484. doi:10.1016/j.socnet.2005.11.005

[10.] Kann MG. Brief Bioinform 2007; 8:333–46.[PMID: 17638813]

[11.] Goh KI et al. Proc Natl Acad Sci USA. 2007; 104:8685-90. [PMID: 17502601]

[12.] Junker HB et al. BMC Bioinformatics 2006; 7:219 [PMID: 16630347]

[13.] Kleinberg JM: Authoritative Sources in a Hyperlinked Environment. Journal of the ACM 1999, 46(5):604-632.

[14.] Ortutay C et al. Nucleic Acids Res 2009; 37:622-8. [PMCID: PMC2632920]

[15.] Keshava Prasad TS et al. Nucleic Acids Res 2009; 37:D767-72. [PMID: 18988627]

[16.] Chatr-aryamontri A et al. Nucleic Acids Res 2007;D572-4. [PMID: 17135203]

[17.] Xenarios I et al. Nucleic Acids Res. 2000; 28:289-91. [PMCID: PMC102387]

[18.] Page L, Brin S, Motwani R, Winograd T: The Page Rank Citation Ranking: Bringing Order to the Web. Tech rep, Stanford Digital Library Technologies Project 1998.

[19.] http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture4/lecture4.html

[20.] Ambedkar, Chintagunta, et al. "Application of centrality measures in the identification of critical genes in diabetes mellitus." Bioinformation 11.2 (2015): 90.

[21.] Kalyani D, Muppalaneni NB, Ambedkar C, Reddi KK (2016) Identification of drug targets from integrated database of diabetes mellitus genes using protein-protein interactions. Application of computational intelligence to biology. Springer, pp 83–86

[22.] Muppalaneni, Naresh Babu, K. Lalitha, and Sasikumar Gurumoorthy. "Identification of Critical Genes in Autism Disorder Using Centrality Measures." Cognitive Science and Health Bioinformatics. Springer, Singapore, 2018. 113-121.

[23.] Arun, PV Parvati Sai, et al. "Computational Prediction of Ligands with Multiple Protein Targets Involved in Type II Diabetes." Cognitive Science and Health Bioinformatics. Springer, Singapore, 2018. 107-112.

[24.] Xiao Q, Wang J, Peng X, Wu FX, Pan Y. Identifying essential proteins from active PPI networks constructed with dynamic gene expression. BMC Genomics. 2015;16 Suppl 3(Suppl 3):S1. doi:10.1186/1471-2164-16-S3-S1

[25.] Tong, Z., Zhou, Y. & Wang, J. Identifying potential drug targets in hepatocellular carcinoma based on network analysis and one-class support vector machine. Sci Rep 9, 10442 (2019)

[26.] Izudheen, Sminu, et al. "Effect of community structures in protein–protein interaction network in cancer protein identification." CURRENT SCIENCE 118.1 (2020): 62.

[27.] Estrada, Ernesto, and Grant J. Ross. "Centralities in simplicial complexes. Applications to protein interaction networks." Journal of theoretical biology 438 (2018): 46-60.

[28.] Lei, Xiujuan, Siguo Wang, and Fangxiang Wu. "Identification of Essential Proteins Based on Improved HITS Algorithm." Genes 10.2 (2019): 177.

[29.] Sharma, Pooja, Dhruba K. Bhattacharyya, and Jugal K. Kalita. "Centrality analysis in PPI networks." 2016 International Conference on Accessibility to Digital World (ICADW). IEEE, 2016.

[30.] Mistry, Divya et al. "DiffSLC: A graph centrality method to detect essential proteins of a protein-protein interaction network." PloS one vol. 12,11 e0187091. 9 Nov. 2017, doi:10.1371/journal.pone.0187091

[31.] Rao, Gali Nageswara, et al. "A tool for the post data analysis of screened compounds derived from computer-aided docking scores." Bioinformation 9.4 (2013): 207.

Author Profile

Dr. Akular Chandra sekhar has obtained his M.Tech from Andhra University, Subsequently he earned his Doctorate in Computer Science and Engineering in the year 2012 from Acharaya Nagarjuna University. He has a total of 16 years. of experience in teaching, academic leadership and research. He has got 10 research papers to his credit and also one of the editor for "Application of computational intelligence to biology "(A Springer Series Book). He served several reputed institutions in the capacities of academic & administrative such as Principal, Director and Head Of The Department. He was Awarded as Best Teacher for the Academic Year 2009-10 from JNTU Kakinada. He has participated in 15 National Conferences,4 International conferences and 5 National Workshops. He organized 3 International conferences and 1 International conference

Mr. Ch. Ambedkar is working as Assistant Professor in the Department of Computer Science and Engineering at SRK Institute of Technology, Vijayawada, India. He did his M.Tech from Andhra University. Currently he is pursuing Ph.D. from JNTU Kakinada.