

Community Detection in Disease Gene Network

Pre-Thesis Seminar

S. Chatterjee

Supervisor: B. S. Sanjeev

Department of Applied Sciences
Indian Institute of Information Technology, Allahabad

1 June, 2024



Outline

- Introduction
- Previous work and reviews
- Pathogen Host Interactions
- Network Analysis and Centrality Indices
- Network Modularity
- Modularity based Community detection
- Epstein-Barr associated carcinomas
- COVID-19 DisGeNET Analysis
- Conclusions and future work

Introduction

Definition:

Network theory is a useful tool for analyzing complex systems. Applications of network theory include the World Wide Web, Internet, Gene regulatory networks, Metabolic networks, Social networks, etc.

Objective:

Our objective is to introduce the concept of general network biology and elucidate how network analysis can be used to represent and interrogate different aspects of the Pathogen-Host Interactions.

Introduction

- Nodes and edges are the building blocks of any network, also called a graph, where nodes represent variables and edges represent relationships between the variables. Using this concept, network analysis has been applied to several fields.
- In recent times, network analysis has also been applied extensively in biology, especially in computational and systems biology research.^{1,2}

¹Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet 2011, 12:56–68

²Califano A, Butte AJ, Friend S, Ideker T, Schadt E. Leveraging models of cell regulation and GWAS data in integrative network-based association studies. Nat Genet 2012, 44:841–847

Motivation

The human disease network

Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L (2007) Proc Natl Acad Sci USA 104:8685-8690

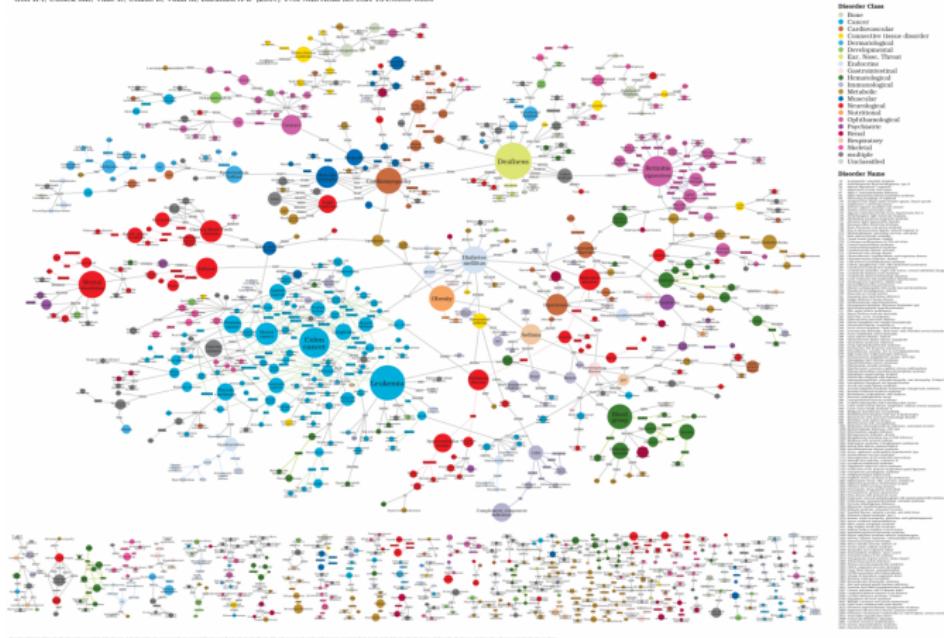


Figure: The Human Disease Network³.

³Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M. and Barabási, A.L., 2007. The human disease network. Proceedings of the National Academy of Sciences, 104(21), pp.8685-8690.

Disease module hypothesis

- The hypothesis states that proteins involved in the same disease have a tendency to interact with each other, forming ‘disease modules’.
- A topological module is a particularly dense area of a network in which the nodes have a higher proportion of links between the components of the module than to components outside that module.
- Functional modules are groups of nodes in a network neighborhood with similar function.
- Disease modules are defined as a group of network components that contribute to a cellular function and, when disrupted, lead to disease.

Types of network-based modules

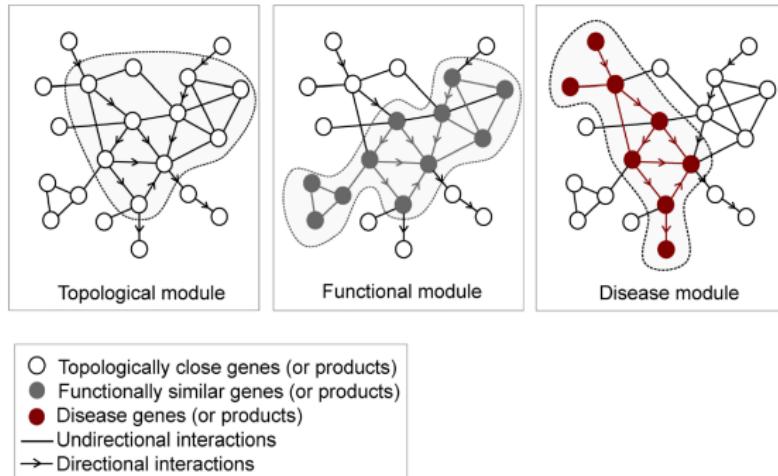


Figure: Review of network-based modules⁴.

⁴Barabási, A.L., Gulbahce, N. and Loscalzo, J., 2011. Network medicine: a network-based approach to human disease. *Nature reviews genetics*, 12(1), pp.56-68.

Network-based methodology

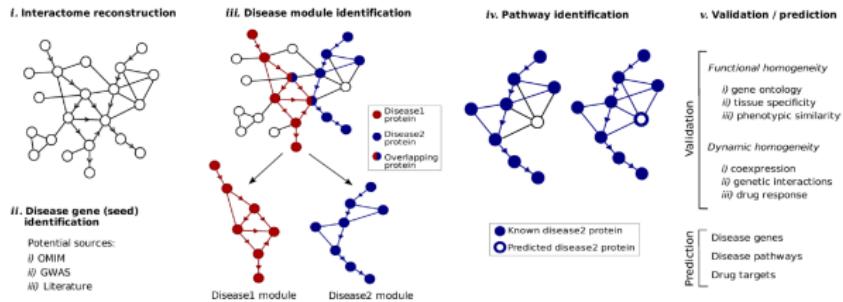


Figure: Schematic representation of methodology⁵.

⁵Barabási, A.L., Gulbahce, N. and Loscalzo, J., 2011. Network medicine: a network-based approach to human disease. *Nature reviews genetics*, 12(1), pp.56-68.

Previous Work and Reviews

- Comprehensive reviews have appeared on molecular networks and their applications.^{6,7,8,9}
- We focus specifically on network concepts in the context of the host-pathogen interactions, comprising many dynamic, multi-scalar processes.

⁶Veiga DF, Dutta B, Balazsi G. Network inference and network response identification: moving genome-scale data to the next level of biological discovery. Mol Biosyst 2010, 6:469–480

⁷Vital-Lopez FG, Memiševic V, Dutta B. Tutorial on biological networks. WIREs Data Mining Knowl Discov 2012, 2:298–325

⁸Chen B, Butte AJ. Network medicine in disease analysis and therapeutics. Clin Pharmacol Ther 2013, 94:627–629

⁹Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet 2004, 5:101–113.

Previous Work and Reviews

- Various studies suggest that interactions between host and pathogen molecular networks are not random but rather well orchestrated, directed events that can influence both host and pathogen behavior.¹⁰
- Given the intricate interactions that underlie immune function, the use of networks is not only suitable but also necessary for understanding host immune responses.¹¹
- A holistic view of pathogen-host interactions (PHIs) therefore requires a systematic and comprehensive study of host and pathogen molecular networks.

¹⁰Fraser ID, Germain RN. Navigating the network: signaling cross-talk in hematopoietic cells. *Nat Immunol* 2009, 10:327–331

¹¹Kidd BA, Peters LA, Schadt EE, Dudley JT. Unifying immunology with informatics and multiscale biology. *Nat Immunol* 2014, 15:118–127

Work-flow of Research

Specifically, we want to concentrate on the following five broad issues:

- Construction, identification and application of network centrality measures in biological networks.
- Topological analysis and community detection in disease gene networks.
- Identification of modules and hierarchical structure of networks.
- Synchronization of network-based modules in the network.
- Statistical validation and conclude our findings.

Introduction to Modularity in Biological Networks

- In general, the term ‘modularity’ refers to the compartmentalization and interrelation of the parts.
- Modularity can be mathematically represented and topologically studied to reveal unexpected structural features.
- Modularity reflects the concentration of edges within modules compared with random distribution of links between all nodes regardless of modules.

Network modularity based Louvain method

The Louvain method¹² based on network modularity (Q) is defined as:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (1)$$

where A_{ij} represents the edge weight between nodes i and j ; k_i and k_j are the sum of the weights of the edges attached to nodes i and j respectively; m is the sum of all of the edge weights in the graph; c_i and c_j are the communities of the nodes; and δ is Kronecker delta function ($\delta(x, y) = 1$ if $x = y$, 0 otherwise).

¹²Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, P10008

Introduction to Topological Communities

- Larger topological structures within networks are commonly explored in terms of communities (or modules).^{13,14}
- A community is loosely defined as a subgraph with high local link density, so that nodes within the community have a higher number of links to each other than to nodes outside the community. A large number of definitions of communities appear in the literature, as well as algorithms to detect them¹⁵.
- Various algorithms can also reveal hierarchical community structures.^{16,17}

¹³Girvan and Newman 2002

¹⁴Ravasz, Somera, et al. 2002

¹⁵Fortunato 2010

¹⁶Girvan and Newman 2002

¹⁷Ahn, Bagrow, et al. 2010

Brief outline of Louvain Algorithm

- The algorithm is based on 2 components applied alternatively:
 - ▶ local moving
 - ▶ contraction
- Each node u randomly checks for each of its neighbors v and how much the modularity would (+) or (-) when moving from u to v community.
- If an improvement is possible, u is moved into neighbor's community v that gives maximum modularity.
- After this local moving phase, the graph is contracted such that each community becomes a node in the contracted graph.

Louvain Algorithm

Algorithm 1: Louvain Algorithm for Community Detection

Input: Set of nodes and edges in the Graph

Output: Hierarchical layout of modular communities

```
1 G the initial network for Community Detection
2 repeat
3     Place each node of G in its own community
4     Save the modularity of this decomposition
5     while some nodes are moved do
6         for all nodes n of G do
7             c ← neighboring community maximizing modularity
8             if c results in a strictly positive increase then
9                 move n from its community to c
10            endif
11        endfor
12    endwhile
13    if the new modularity is greater than the initial then
14        end ← false
15        G = the network between communities of G
16    else
17        end ← true
18    end
19  endif
20 until;
```

Title:

Identification of Human Proteins vulnerable to Multiple Organisms. ^a ^b

Keywords: Centrality analysis, IIPs, Graph theory, Host pathogen interaction, Hubs.

^aWBSB-2016

^bS. Chatterjee and B.S. Sanjeev, Department of Applied Sciences, Indian Institute of Information Technology, Allahabad, 211012, UP, India

Abstract

- Background: The purpose of this study is to perform topographical analysis of large scale pathogen host interaction networks (PHI Networks) in order to identify proteins important for network topology and integrity.
- Methods: Various *graph theoretical* approaches were implemented to identify highly interacting hub and central proteins that are crucial for network integrity.

Overview

Table: Overview of PHI Networks¹⁸

Parameters	Nos.
Total no. of pathogens	182
No. of human proteins (H_p)	668
No. of pathogen proteins (P_p)	1,188
Total no. of proteins (H_p) + (P_p)	1,856
Total no. of interactions	3,905

¹⁸S. D. Tekir, T. Çakır, E. Ardiç, A. S. Sayılıbaşı, G. Konuk, M. Konuk, H. Sarıyer, A. Uğurlu, İ. Karadeniz, A. Özgür, F. E. Sevilgen, K. Ö. Ülgen, PHISTO: Pathogen-Host Interaction Search Tool, Bioinformatics., (Databases and ontologies):1357-1358, 2013.

Method

In our network analysis, we make use of several indicators of centrality to identify the most important vertices within a graph^{19, 20}. Herein, for our undirected graph Γ , such that $\Gamma_p = (V_p, E_p)$, V_p is the set of all nodes (proteins) of H_p (Human) and P_p (Pathogenic) and E_p is the set of corresponding edges (interactions) wherein $E_p \subseteq [V_p]^2$.

¹⁹N. T. Doncheva, Y. Assenov, F. S. Domingues, M. Albrecht, Topological analysis and interactive visualization of biological networks and protein structures, *Nature Protocols* 7:670-685, 2012

²⁰Y. Assenov, F. Ramírez, S. E. Schelhorn, T. Lengauer, M. Albrecht, Computing topological parameters of biological networks, *Bioinformatics*, 24(2):282-284, 2008

Our PHI Network

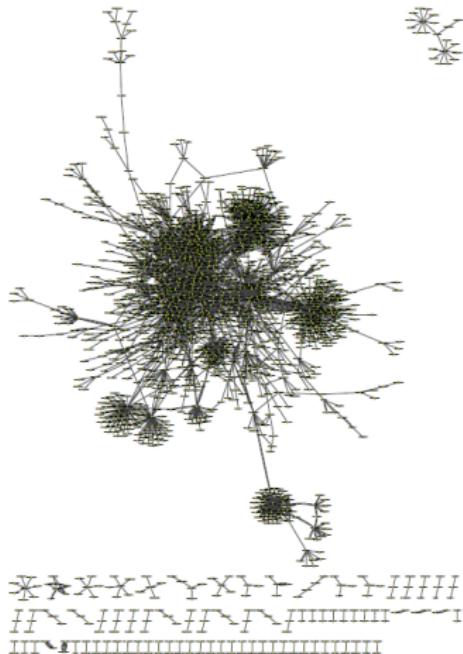
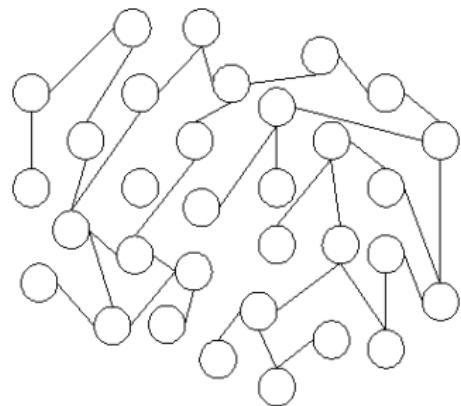


Figure: Pathogen Host Interactome Network.

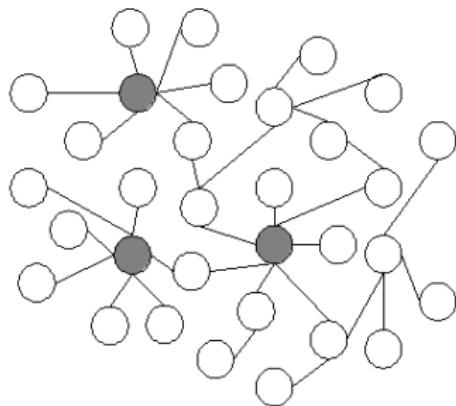
Random v/s Scale-free graphs

Erdős-Rényi random graphs



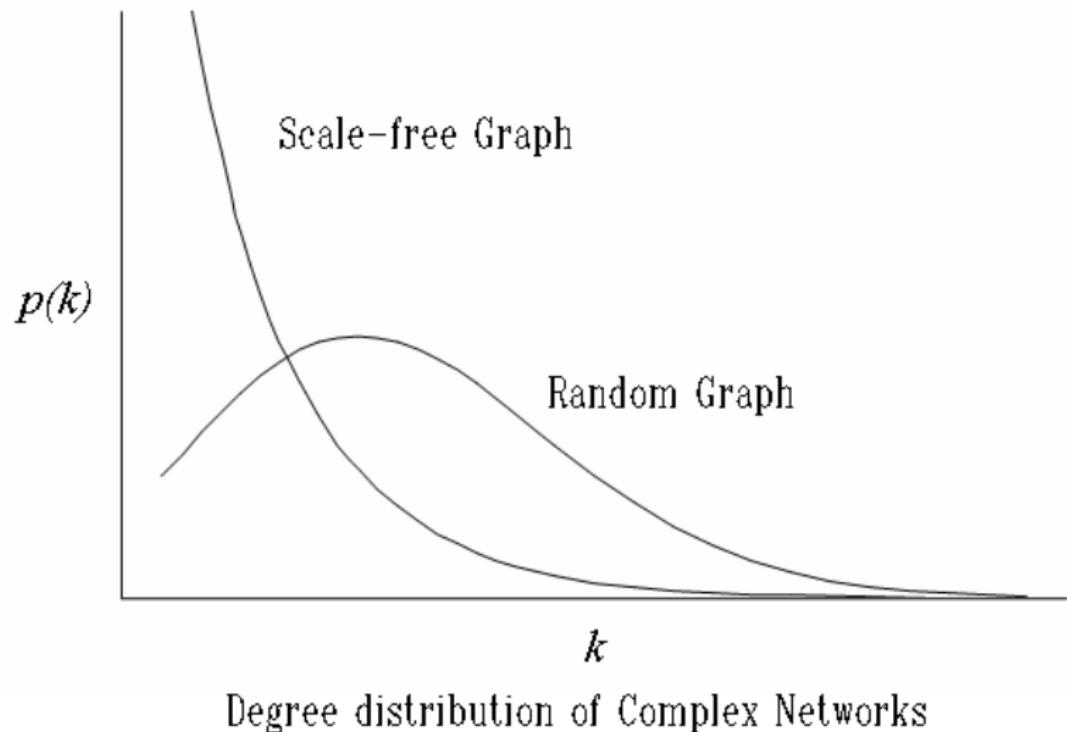
(a) Random network

Barabási-Albert scale-free graphs



(b) Scale-free network

Degree of distribution



Node Degree Distribution

Definition: The node degree distribution is an indicator of the number of nodes with a degree of k . In undirected networks, the node degree of a node n is the number of edges linked to n .

Node Degree Distribution

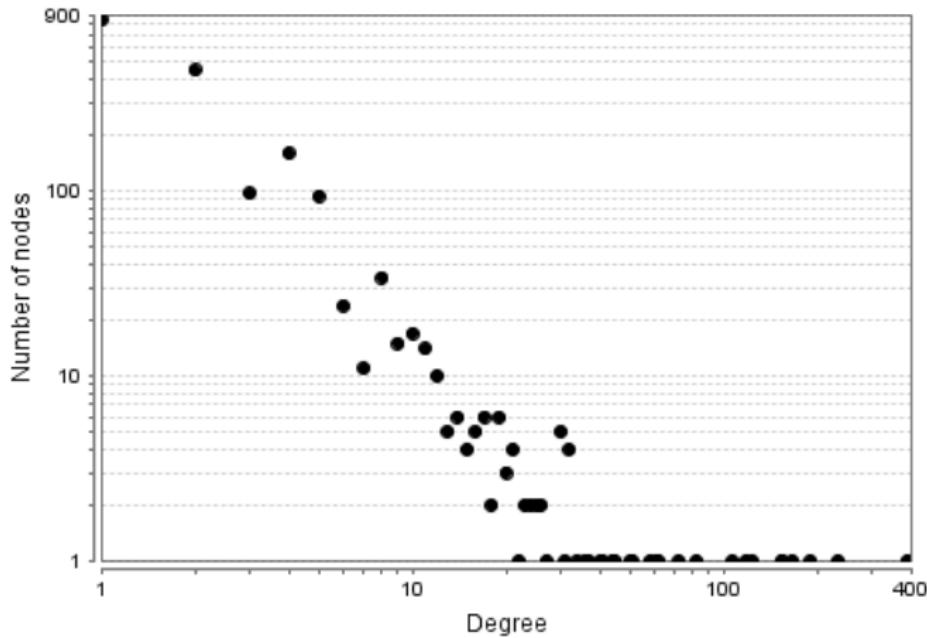


Figure: Node Degree Distribution

Topological Coefficient

Definition: Topological coefficient $T(n)$ of a node n with k_n neighbors is defined as the number of neighbors shared between a pair of nodes, k_n and k_m , divided by the number of neighbors of node k_n :

$$T(n) = \frac{\text{avg}(J(n, m))}{k_n} \quad (2)$$

where the value of $J(n, m)$ is the number of neighbors shared between the nodes n and m , plus one if there is a direct link between n and m and is defined for all nodes m that share at least one neighbor with n . It gives a relative measure for the extent to which a node shares its neighbors with other nodes.

Topological Coefficient

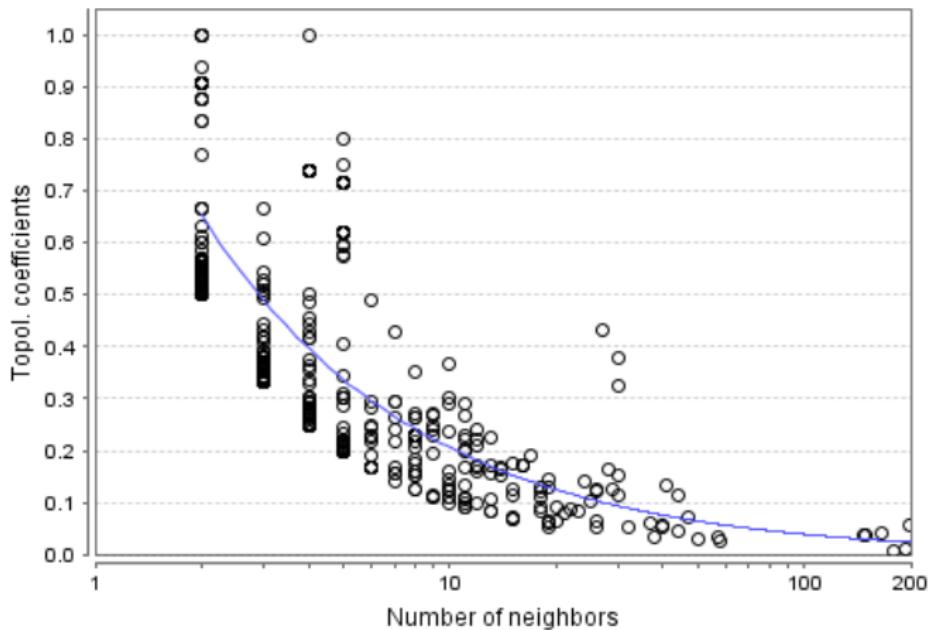


Figure: Topological Coefficients

Stress Centrality

Definition: The stress centrality of a node n is defined as the total number of shortest paths passing through n .

A node has a high stress centrality if it is traversed by a large number of shortest paths. This parameter is defined only for networks without multiple edges. In short, this index reflects how often a node lies on the geodesics between other nodes.

Stress Centrality

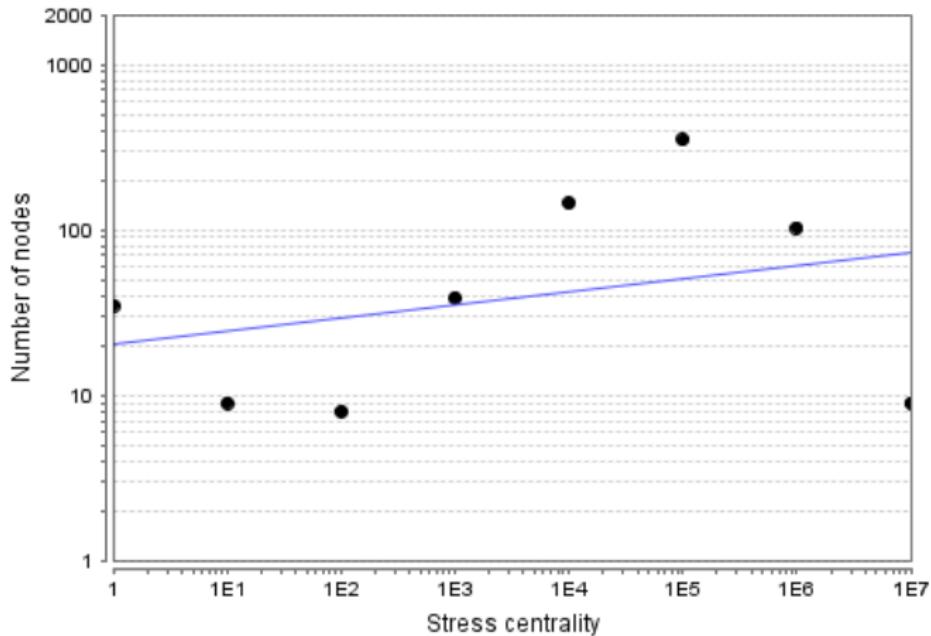


Figure: Stress Centrality Distribution

Results and Discussions

Table: Summary of Network Statistics of PHI Networks

Parameter	Statistics
Connected Components	82
Network Diameter	17
Shortest Paths	2,610,796 (75%)
Avg. number of Neighbors	3.357
Multi-edge node pairs	496

Results and Discussions

Table: Overview of Important Interacting proteins (IIPs)

Human Protein	Uniprot ID	Deg.	No. of Path.	
			Virus	Bact.
P53_Human	P04637	193	35	5
NFKB1_Human	P19838	180	6	4
GBLP_Human	P63244	40	17	3
TOX4_Human	O94842	23	12	3
PDIA1_Human	P07237	22	9	3
MYH9_Human	P35579	13	6	3
RAC1_Human	P63000	41	3	7
CCAR2_Human	Q8N163	18	13	2
ILF3_Human	Q12906	30	23	1

Conclusions

After analyzing the graph, we could pick out all the important interacting proteins (IIPs) in the human host that acted as hubs in the pathogen-host interactive graph and subsequently vulnerable to the pathogenic proteins. The pathogenic proteins were then short-listed and were linked to their respective pathogenic organisms/strains. We could list out 9 highly interacting hubs present in the human host that were involved with 560 pathogenic proteins. Herein, we have emphasized only on those human protein/ protein coding genes that were predominantly present in 5 or more viral pathogenic organisms/strains and also subsequently present in 2 or more bacterial organisms/strains barring a few ones *viz.* RAC1, CCAR2 and ILF3.

Association of IIPs with Disease Modules

- Diseases are not caused by a single gene, but are rather in an orchestrated consequence of some abnormality that involves a cascade of interactions encompassing several cellular components (proteins/protein coding genes).

Conclusions

- Topological analyses of bacterial PPI networks have revealed that they follow power-law distributions, i.e., they have few hubs and many peripheral proteins²¹ and similar to bacterial PPI networks.²² also follow scale-free distribution.
- Topological analysis of proteins involved in interaction of host and pathogen networks indicates that both bacterial and viral pathogens in general target hub and bottleneck proteins in host networks²³.

²¹Barabasi AL, Albert R. Emergence of scaling in random networks. *Science* 1999, 286:509–512.

²²Guiguemde WA, Shelat AA, Bouck D, Duffy S, Crowther GJ, Davis PH, Smithson DC, Connelly M, Clark J, Zhu F, et al. Chemical genetics of *Plasmodium falciparum*. *Nature* 2010, 465:311–315.

²³Durmus Tekir SD, Ulgen KO. Systems biology of pathogen-host interaction: networks of protein-protein interaction within pathogens and pathogen-human interactions in the post-genomic era. *Biotechnol J* 2013, 8:85–96.

Introduction

Background:

The Epstein–Barr virus (EBV) is the first herpesvirus associated with human cancers known to infect majority of the world population ^a. Epstein-Barr virus (EBV) is a ubiquitous human herpes virus associated with the development of many lymphoid and epithelial tumors.

^aEpstein, M.A., 1964. Virus particles in cultured lymphoblasts from burkitt's lymphoma. Lancet 1, 702–703.

EBV infection:

Epstein-Barr virus (EBV) is the primary cause of infectious mononucleosis (IM) and is associated with several malignancies. About 200,000 cancer cases globally per year are thought to be attributable to EBV. These oncogenic viruses have many complicated strategies that disrupt biological pathways in the infected host cells.

EBV-affecting B-cells and Epithelial cells

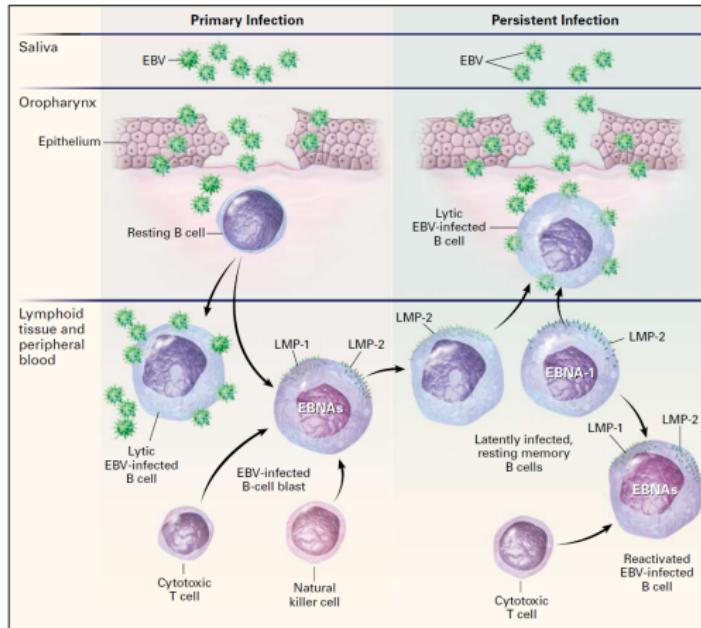


Figure 1. Model of Epstein-Barr Virus (EBV) Infection in Humans.

Figure: Epstein Barr Virus mostly affecting B-cells and Epithelial cells.²⁴

²⁴Cohen, J.I., 2000. Epstein-Barr virus infection. New England journal of medicine, 343(7), pp.481-492.

EBV-associated carcinomas

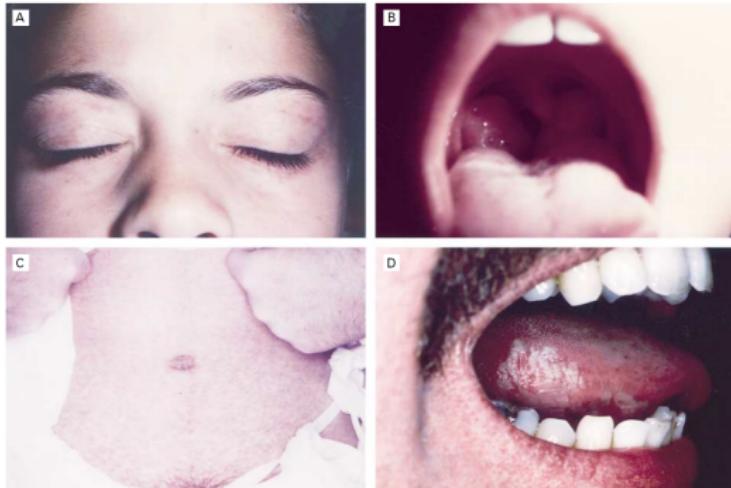


Figure 3. Clinical Findings in Epstein–Barr Virus (EBV) Infection.
Panel A shows petechiae of the eyelids with periorbital edema, and Panel B shows tonsillar enlargement in a patient with infectious mononucleosis. Panel C shows macular rash after ampicillin therapy in a patient with infectious mononucleosis. Panel D shows oral hairy leukoplakia in a patient with AIDS. Photographs courtesy of Maria Turner, M.D.

Figure: Epstein Barr Virus mostly affecting B-cells and Epithelial cells.²⁵

²⁵Cohen, J.I., 2000. Epstein–Barr virus infection. New England journal of medicine, 343(7), pp.481-492.

Review of EBV Vaccine Development

nature medicine

Explore content ▾

About the journal ▾

Publish with us ▾

Subscribe

[nature](#) > [nature medicine](#) > [research highlights](#) > article

RESEARCH HIGHLIGHT | 19 May 2022

Finding a vaccine for Epstein-Barr virus

Nanoparticle vaccines show promise in preclinical models; clinical evaluation is warranted to assess whether they could help reduce the burden of disease associated with EBV in humans.

Karen O'Leary 

Figure: Finding a vaccine for EBV.²⁶

²⁶O'Leary, K., 2022. Finding a vaccine for Epstein-Barr virus. *Nature medicine*.

Problem definition

Problem:

Epstein-Barr virus (EBV) is associated with several malignancies. Potential drug targeted therapies/EBV vaccine affecting critical biological processes/pathways can reduce the rate of infectious mononucleosis (IM) in various carcinomas.

Current status:

Trials to reduce the incidence of Hodgkin lymphoma, Burkitt lymphoma and other associated carcinomas are currently under development.

Objective:

Our objective is to implement network theory approaches and elucidate how community detection and functional enrichment can be used to represent and interrogate different aspects of EBV-associated carcinomas and find out which critical processes are being affected.

Types of cancer associated with Epstein-Barr virus

Table: Types of cancer associated with Epstein-Barr virus through DisGeNET (v7).

Sl.	Category	Associated Cancer Type	Reference [PMID]
1	GC	Malignant neoplasm of stomach	1342957
2		Stomach Neoplasms	1342957
3		Stomach Carcinoma	1342957
4		Benign neoplasm of stomach	1342957
5		Carcinoma in situ of stomach	1342957
6		Neoplasm of uncertain / unknown behavior of stomach	1342957
7	HL	Lymphoma, Non-Hodgkin	9547991
8		Hodgkin Disease	9547991
9		Lymphoma, Non-Hodgkin, Familial	9547991
10		Adult Hodgkin Lymphoma	9547991
11		Hodgkin lymphoma, lymphocyte depletion	9547991
12		Lymphocyte Rich Classical Hodgkin Lymphoma	10505767
13		Mixed Cellularity Hodgkin Lymphoma	10505767
14	BL	Nodular Lymphocyte Predominant Hodgkin Lymphoma	10505767
15		Burkitt Lymphoma	2824192
16		Burkitt Leukemia	2824192
17	NPC	African Burkitt's lymphoma	2824192
18		Cancer of Nasopharynx	8732865
19		Malignant neoplasm of nasopharynx	8732865
20	Syndrome	Lymphoproliferative Syndrome, Ebv-Associated, Autosomal	22409825
21		Chronic Lymphoproliferative Disorder of NK-Cells	6158759
22		X-linked immunodeficiency with magnesium defect	25313976

EBV-associated carcinomas

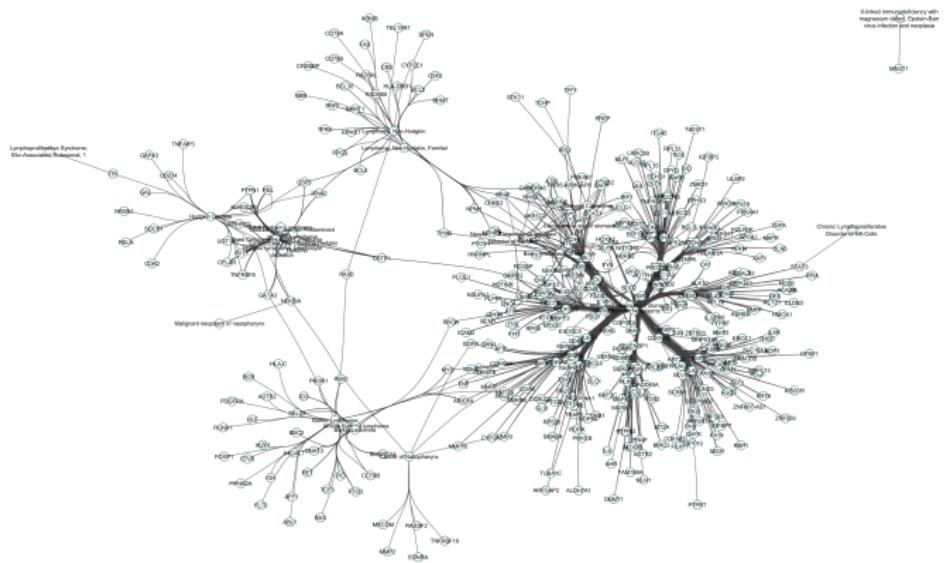


Figure: The Disease Gene Network of Epstein Barr Virus in various neoplastic processes (Gastric cancer (GC), Nasopharyngeal cancer (NPC), Hodgkin's lymphoma (HL) and Burkitt Lymphoma (BL))

Node degree distribution

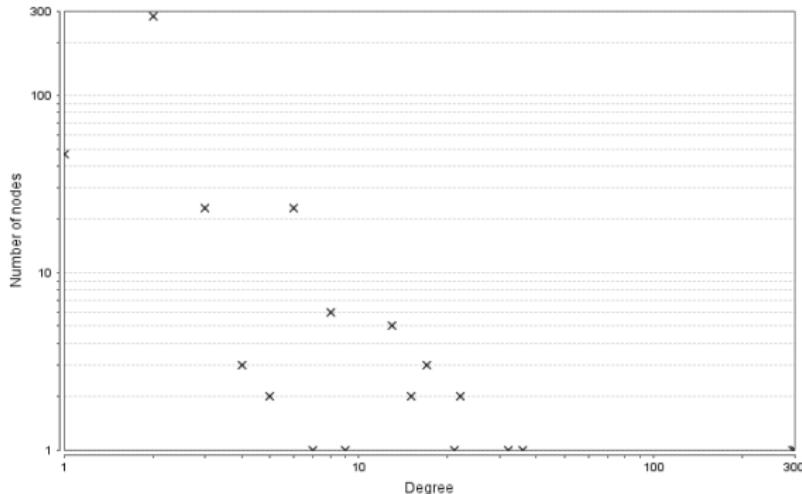


Figure: The node degree distribution of the EBV-associated cancers. Malignant neoplasm of stomach (in GC category) was the most affected carcinoma associated with EBV.

Network Modularity in EBV-linked Carcinoma Network

- Modularity is defined as the fraction of edges that fall within the given groups minus the expected fraction, if edges were distributed at random.²⁷
- A positive modularity signifies that the number of edges connected to nodes in a network within a community exceeds the expected value by chance randomly (range between -1/2 to 1). To identify significant communities Louvain method-based community detection was implemented, and critical processes were selected based on p-values representing their significance.²⁸
- The network modularity of the disease-gene network was found to be 0.359 which depicted a positive modular network structure.

²⁷Brandes Ulrik, Delling Daniel, G.M.G.R.H.M.N.Z., Dorothea, W., 2008. On modularity clustering. IEEE Transactions on Knowledge and Data Engineering 20, 172–188.

²⁸Sham, P.C., Purcell, S.M., 2014. Statistical power and significance testing in large-scale genetic studies. Nature Reviews Genetics 15, 335–346.

Network modularity of Carcinoma Network

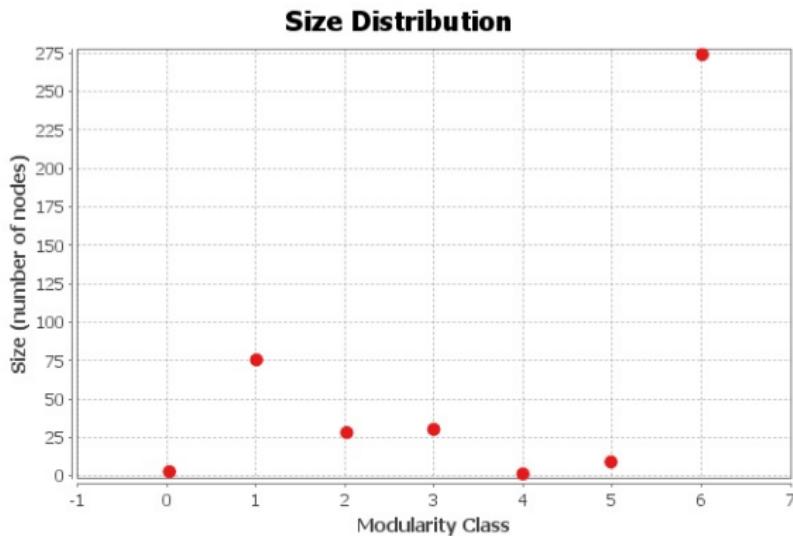


Figure: The network modularity of the disease-gene network in EBV-associated carcinomas was 0.359 signifying a positive structured modular network. The figure depicts the size distribution of the 7 modularity classes.

Community detection in modular networks

- Community detection is fundamental to biological network analysis for protein function annotation, disease gene prediction and studies based on targeted drug therapies.²⁹
- Such techniques have successfully been implemented for the identification of critical genes based on modularity, network modules and drug targeted therapies.³⁰
- We implemented Community Detection Application and Service framework (CDAPS) based on the integrated approach to identify and visualize large-scale multi-scale network communities.
- Community detection methods build hierarchical representations directly from the graph structure and we used Hierarchical community Decoding Framework (HiDeF), for network community detection in biological networks.³¹

²⁹Singhal, A., Cao, S., Churas, C., Pratt, D., Fortunato, S., Zheng, F., Ideker, T., 2020. Multiscale community detection in cytoscape. PLoS computational biology 16, e1008239.

³⁰Gulbahce, N., Lehmann, S., 2008. The art of community detection. BioEssays 30, 934–938.

³¹Zheng, F., Zhang, S., Churas, C., Pratt, D., Bahar, I., Ideker, T., 2021. Hidef: identifying persistent structures in multiscale ‘omics data. Genome biology 22, 1–15.

Network community detection methodology

- We performed community detection using the Louvain algorithm in the CDAPS framework with a maximum resolution parameter of less than 50 to gain insights even from relatively smaller communities.
- The community persistence threshold was kept at 5, to remove unstable clusters.

Result

This resulted in a hierarchical network with communities as nodes and their hierarchical relationships as edges. We fetched significant communities in the form of hierarchical modules in the overall network.

Over-representation Enrichment Methodology

- We employed the methodology of over-representation analysis to detect statistically significant gene ontology terms.
- Genes were mapped to known functional annotations to detect statistically significant enriched pathways for functional enrichment.³²
- The nodes were selected with minimum Jaccard index value³³ for overlap greater than 0.05 and the range of p-value³⁴ was kept less than 0.00001.

³²Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., Vilo, J., 2019. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic acids research 47, W191–W198.

³³Jaccard, P., 1912. The distribution of the flora in the alpine zone. 1. New phytologist 11, 37–50.

³⁴Storey, J.D., Tibshirani, R., 2003. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences 100, 9440–9445.

Hierarchical Network representation

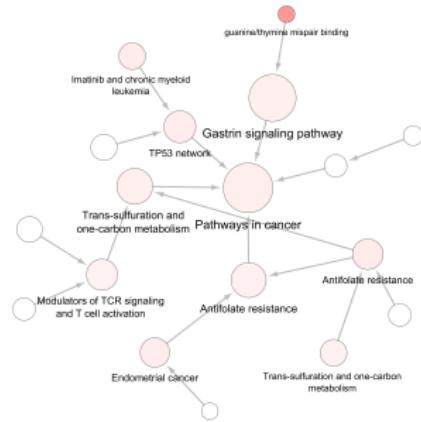


Figure: The figure depicts nodes in the Disease-Gene Network which represents its significance denoted by the size of the node (based on p-values). The respective biological processes and pathways are interlinked in a hierarchical network layout using over-representation analysis.

Over-representation Analysis

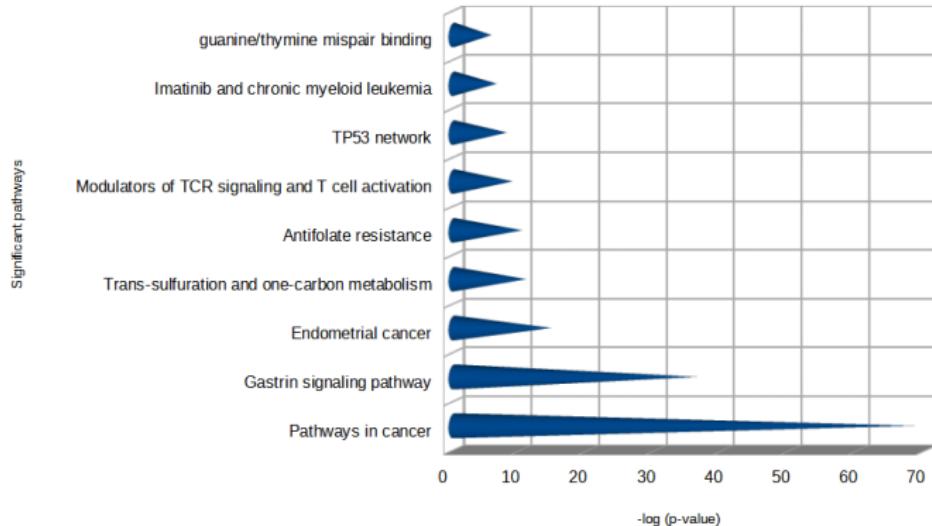


Figure: List of functionally enriched significant biological processes / pathways.

Results and Discussions

- The top 10 genes that were involved with EBV-associated carcinomas involving GC, NPC, HL and BL were identified as CASP10, BRAF, NFKBIA, IFNA2, GSTP1, CSF3, GATA3, UBR5, AXIN2 and POLE.
- We listed the community of genes involved with 9 key biological processes and critical pathways linked with the Epstein-Barr virus in the overall disease gene network.
- The critical processes mainly involved were associated with pathways in cancer, Gastrin signaling pathway, Endometrial cancer, Trans-sulfuration and one carbon metabolism, Antifolate resistance, Modulators of TCR signaling and T-cell activation, TP53 Network, Imatinib and chronic myeloid leukemia and Guanine / Thymine mispair binding.

Conclusions

- We report 9 key biological processes and disease pathways affected by Epstein-Barr virus (EBV) in the disease gene network. Using topological network analysis and community detection algorithm followed by functional enrichment, we identified the community of genes that were involved with respective biological processes involving carcinomas (GC, NPC, HL and BL).
- Further, it was observed that the ABL1 gene plays a vital role in 3 out of 9 biological processes based on over-representation analysis.
- Hence, the linkages between various biological processes infected with Epstein-Barr virus in causing carcinomas from the network of common genes present significant novel insights and therefore, we provide our case for clinical studies.

Introduction

Background:

Evidence from multiple studies has shown that the severity of COVID-19 is enhanced by comorbid and life-threatening diseases, and patients with comorbidities have been affected disproportionately ^a. Such patients are more susceptible to organ failure and mortality ^b

^azong2021intersection, sanyaolu2020comorbidity

^bsanyaolu2020comorbidity

COVID-19 infection:

Given the global spread of COVID-19, it is vital to investigate biological processes and pathways that are common between COVID-19 and other diseases and comorbid conditions. Such advances would ultimately contribute towards better therapeutic outcomes.

NIH LitCovid dataset

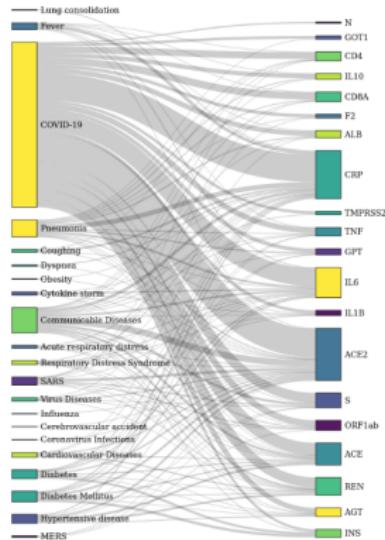


Figure: Genes that are frequently mentioned together with COVID-19.³⁵

³⁵<https://www.ncbi.nlm.nih.gov/research/coronavirus>

Problem definition

Problem:

Treating patients with life threatening diseases during COVID-19 pandemic has been quite challenging and therapeutic clinical applications used in various immunotherapies have led to the worsening of COVID-19 outcomes among patients [35].

Current status:

The molecular interplay involving overlapping relationships between COVID-19 and other diseases is not yet fully understood [12] and undergoing clinical investigation.

Objective:

Our objective is to implement network theory approaches and elucidate how community detection and functional enrichment can identify biological processes and pathways shared between COVID-19 and other diseases.

COVID-19 DisGeNET Network

(a) High degree genes			(b) High degree diseases		
Rank	Gene	Degree	Rank	Disease Name	Degree
1	VEGFA	1899	1	Neoplasms (unclassified)	535
2	BCL2	1456	2	Malignant Neoplasms	459
3	CTNNB1	1368	3	Primary Malignant Neoplasms	444
4	ALB	1198	4	Malignant Neoplasm of breast	372
5	COX2	875	5	Tumor cell Invasion	368
6	AGT	765	6	Breast Carcinoma	363
7	HLA-A	672	7	Liver Carcinoma	352
8	HMOX1	666	8	Carcinogenesis	347
9	FGF2	635	9	Neoplasm metastasis	344
10	COMT	622	10	Colorectal carcinoma	309
11	MET	594	11	Malignant neoplasm of prostate	284
12	PLG	586	12	Prostrate carcinoma	276
13	PCNA	581	13	Malignant neoplasm of lung	263
14	PARP1	565	14	Carcinoma of lung	262
15	AIMP2	555	15	Primary malignant neoplasm of lung	257
16	FBN1	552	16	Non-small cell Lung carcinoma	241
17	HSPA4	550	17	Tumor progression	235
18	GNAS	536	18	Alzheimer's Disease	226
19	STAT1	531	19	Malignant neoplasm of stomach	221
20	AHSA1	526	20	Stomach carcinoma	209
21	TGFBR2	502	21	Malignant neoplasm of colon	208
22	HSPA1B	502	22	Glioblastoma	198
23	DNMT1	496	23	Glioblastoma Multiforme	195
24	HSPA1A	458	24	Glioma	193
25	DPP4	451	25	Melanoma	188

Degree Distribution

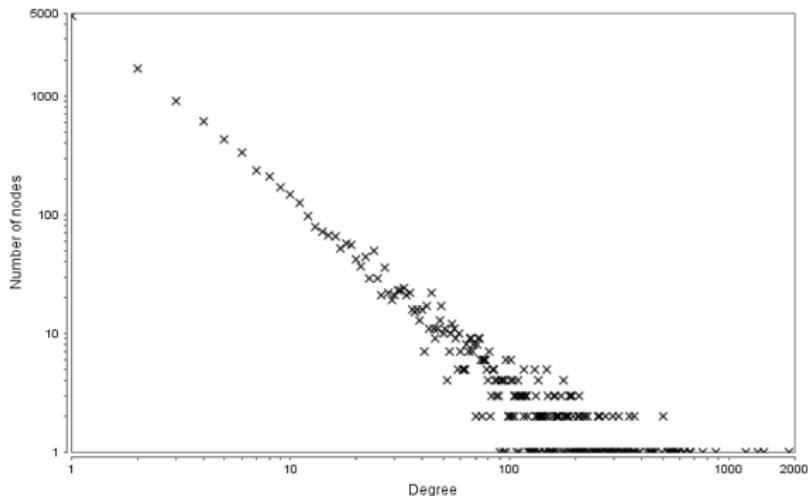


Figure: Very high degrees for only a fraction of nodes shows that only a smaller set of genes and diseases have high interactions with other nodes. For instance, Vascular Endothelial Growth Factor A (VEGFA) is known to be involved in 1899 diseases, while malignant neoplasms are associated with 459 genes. All the genes, taken from COVID DisGeNET data set, are known to be associated with SARS-CoV-2 too.

Network modularity

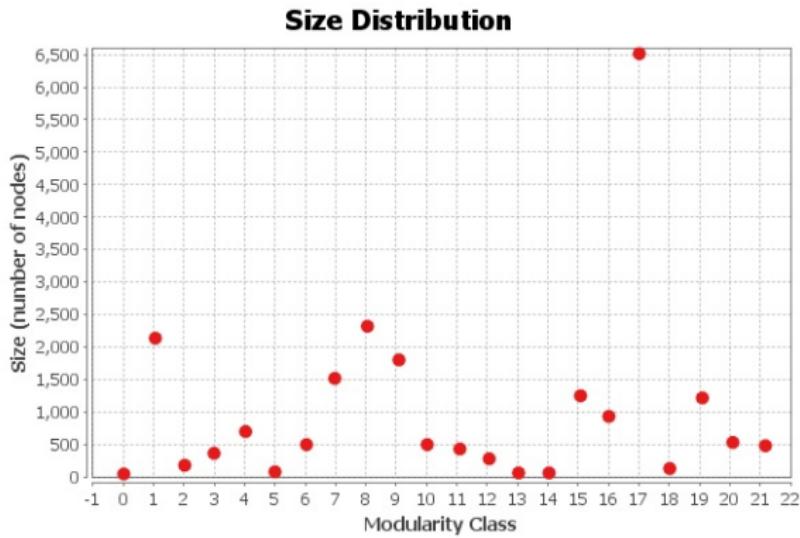


Figure: The network modularity of the disease-gene network signifying a positive structured modular network. The figure depicts the size distribution of the 22 modularity classes.

Over representation Analysis

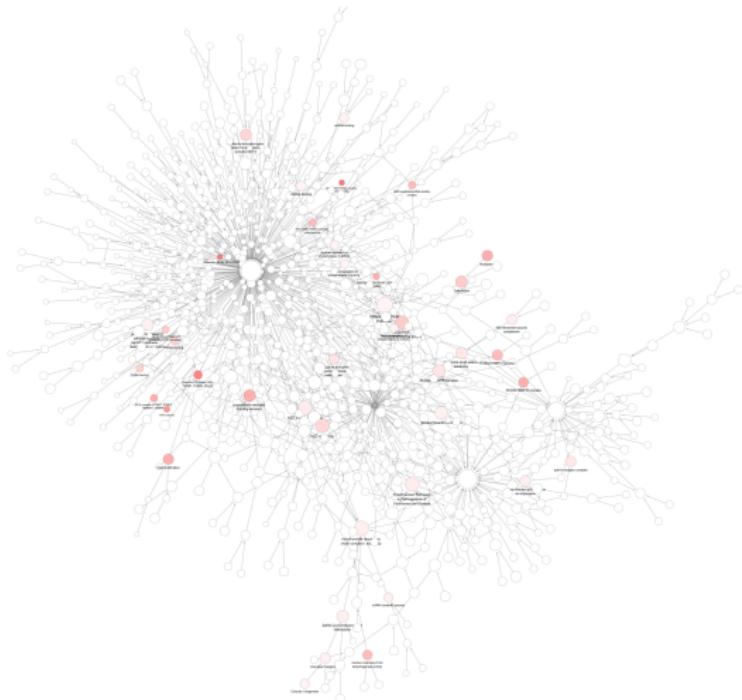


Figure: Hierarchical network layout of biological pathways.

Table: Significant biological processes and pathways. p-value threshold is less than 1.09E-06.

Rank	Gene Ontology / Biological process	P-value	-log(P)	Reference
1	Mitochondrial respiratory chain complex	8.71E-12	11.05998	GO:0032981
2	Ubiquitin E3 ligase	1.63E-10	9.787812	CORUM:622
3	Activin receptor complex	2.82E-10	9.549751	GO:0048179
4	Arylamine N-acetyltransferase activity	1.49E-09	8.826814	GO:0004060
5	Complex I biogenesis	1.49E-08	7.826814	REAC:R-HSA-6799198
6	FK506 binding	3.23E-08	7.490797	GO:0005528
7	HDL assembly	5.21E-08	7.283162	REAC:R-HSA-8963896
8	Integrated Cancer Pathway	3.76E-07	6.424812	WP:WP1971
9	ACE Inhibitor Pathway	5.50E-07	6.258848	WP:WP554
10	Hypothesized Pathways in CVD	5.72E-07	6.242604	WP:WP3668
11	40S ribosomal subunit, cytoplasmic	7.42E-07	6.129596	CORUM:305
12	TIM complex, mitochondrial	8.17E-07	6.087778	CORUM:623
13	EIF3 complex	8.17E-07	6.087778	CORUM:4399
14	VCB complex	8.25E-07	6.083546	GO:0030891
15	Acetylation	1.09E-06	5.962574	REAC:R-HSA-156582
16	SARS-CoV-2 mitochondrial interactions	1.23E-06	5.910095	WP:WP5038
17	snRNA binding	1.30E-06	5.886057	GO:0017069
18	Complex I biogenesis	1.42E-06	5.847712	REAC:R-HSA-6799198
19	Ribosomal small subunit assembly	1.50E-06	5.823909	GO:0000028
20	PLOD2-FKBP10 complex	1.65E-06	5.782516	CORUM:7000
21	Negative regulation of apoptotic pathway	1.73E-06	5.761954	GO:1902042
22	CRD-mediated mRNA stability complex	2.06E-06	5.686133	GO:0070937
23	TIM complex, mitochondrial	2.06E-06	5.686133	CORUM:623
24	ncRNA catabolic process	2.47E-06	5.607303	GO:0034661
25	Acetylation	2.89E-06	5.539102	REAC:R-HSA-156582
26	Angiotensin-mediated drinking behavior	3.28E-06	5.484126	GO:0003051
27	Positive regulation of phospholipase C	4.44E-06	5.352617	GO:0010863
28	DCS complex	4.90E-06	5.309804	CORUM:1288
29	snRNA binding	6.58E-06	5.181774	GO:0017069
30	Ubiquinone and terpenoid-quinone	6.71E-06	5.173277	KEGG:00130
31	Metanephric glomerular capillary formation	7.31E-06	5.136083	GO:0072277
32	CCT complex	7.62E-06	5.118045	CORUM:126
33	Cellular Heat acclimation	8.01E-06	5.096367	GO:0070370
34	Striated Muscle Contraction	8.70E-06	5.060481	REAC:R-HSA-390522
35	Protein N-linked glycosylation via asparagine	9.63E-06	5.016374	GO:0018279
36	Medium-chain-acyl-CoA dehydrogenase activity	9.84E-06	5.007005	GO:0070991
37	Interferon-stimulated gene factor 3 transcription complex	9.90E-06	5.004365	CORUM:60

Table: Community of genes from functionally enriched pathways along with their respective overlap representation ratio. The genes that are underlined represent mitochondrial genes.

Community	Genes	Overlap ratio	Community Size
1	ACAD9 AIFM1 BCS1L <u>NDUFA10</u> <u>NDUFAF1</u> <u>NDUFAF2</u> <u>NDUFB9</u>	0.075	426
2	CUL2 ELOB ELOC	0.6	27
3	ACVR2A ACVR2B ACVR1 ACVR1B	0.111	67
4	NAT2 NAT1 AAANAT	0.025	230
5	<u>NDUFA10</u> <u>NDUFAF1</u> <u>NDUFAF2</u> <u>NDUFB9</u>	0.067	69
6	FKBP1A FKBP3 FKBP7	0.231	15
7	A2M ABCA1 PRKACA	0.187	310
8	BCL2 CHEK2 MYC PTEN	0.057	816
9	ACE2 AGT KNG1	0.097	295
10	FBN1 FBN2 TGFBR2	0.088	355
11	RPS10 RPS17 RPS27	0.083	63
12	TIMM10 TIMM9	0.667	6
13	EIF3B EIF3J	0.667	5
14	CUL2 ELOB	0.5	8
15	NAT2 NAT1	0.4	72

Conclusions

- Our study strongly suggests that the patients with comorbid conditions such as those with various carcinomas, neoplasms and Alzheimer's disease, are the most vulnerable among all the disease classes infected by SARS-CoV-2 through community detection.
- Functional enrichment through over-representation analysis methodology was used to discover significant biological processes and pathways shared between COVID-19 and other diseases.
- Improving endothelial permeability and vasodilation with VEGF-targeted therapy could allow the repair of damaged vascular endothelium and may have an anti-inflammatory effect and improve oxygenation.
- Strategies involving treatment with anti-VEGF drugs such as bevacizumab besides VEGFR-mediated signaling therapies may deliver favorable outcomes.

Publications

- Over-representation analysis of angiogenic factors in immunosuppressive mechanisms in neoplasms and neurological conditions during COVID-19.
S. Chatterjee, B.S. Sanjeev, Microbial Pathogenesis, 185 (106386), pp. 1-9, Dec 2023. doi: 10.1016/j.micpath.2023.106386
- Community detection in Epstein-Barr virus associated carcinomas and role of tyrosine kinase in etiological mechanisms for oncogenesis.
S. Chatterjee, B.S. Sanjeev, Microbial Pathogenesis, 180 (106115), pp. 1-7, July 2023. doi: 10.1016/j.micpath.2023.106115
- Identification of Human Proteins vulnerable to multiple Organisms.
S. Chatterjee, B.S. Sanjeev, In Proceedings of International Conference on Bioinformatics and Systems Biology (BSB), IEEE pp. 1-4, 04-06 March 2016. doi: 10.1109/BSB.2016.7552164
- Role of Toll-like Receptors in the interplay between pathogen and damage associated molecular patterns.
S. Chatterjee, B.S. Sanjeev, In preparation.

Conclusions

- Interaction between a pair of host and pathogen proteins can be predicted based on availability of interacting domain pairs from relevant databases. Once all such pairwise interactions have been predicted, the results can be filtered to retain only hub genes in the host PPI network, as these genes are the ones most likely to be targeted by pathogen proteins to gain access to the host network³⁶.

³⁶ Arnold R, Boonen K, Sun MG, Kim PM. Computational analysis of interactomes: current and future perspectives for bioinformatics approaches to model the host-pathogen interaction space. Methods 2012, 57:508–518.

Appendix

Few recent studies have created some widely used PHI-databases:

- PHISTO³⁷ [<http://www.phisto.org>]
- PHI-base^{38, 39} [<http://www.phi-base.org>]
- HPIDB⁴⁰ [<http://agbase.msstate.edu/hpi/main.html>]
- CAPIH⁴¹ [<http://bioinfo-dbb.nhri.org.tw/capih>]

³⁷Durmus Tekir S, Cakir T, Ardic E, Sayilibas AS, Konuk G, Konuk M, Sariyer H, Ugurlu A, Karadeniz I, Ozgur A, et al. PHISTO: pathogen-host interaction search tool. Bioinformatics 2013, 29:1357–1358.

³⁸Winnenburg R, Baldwin TK, Urban M, Rawlings C, Kohler J, Hammond-Kosack KE. PHI-base: a new database for pathogen host interactions. Nucleic Acids Res 2006, 34:D459–D464.

³⁹M. Urban, R. Pant, A. Raghunath, A. G. Irvine, H. Pedro, K. E. Hammond-Kosack, The Pathogen-Host Interactions database (PHI-base): additons and future developments, Nucleic Acids Research., 43 (Database Issue):D645–655, 2015.

⁴⁰Kumar R, Nanduri B. HPIDB—a unified resource for host-pathogen interactions. BMC Bioinformatics 2010, 11(suppl 6):S16.

⁴¹Lin FK, Pan CL, Yang JM, Chuang TJ, Chen FC. CAPIH: a web interface for comparative analyses and visualization of host-HIV protein-protein interactions. BMC Microbiol 2009, 9:164.

Community Detection APplication and Service (CDAPS)

The figure shows the official documentation for CDAPS. The left side is a sidebar with a dark background containing links to 'CONTENTS', 'Installation', 'What's New', 'Quick Tutorial', 'Tally Attributes on Hierarchy', 'Columns', and 'Settings'. Below these are 'LINKS' to Cytoscape.org, Cytoscape App Store, Idekerlab.ucsd.edu, and GitHub. A code snippet box contains a JSON-like configuration file:

```
# No user data
ethiclabel: deva
topic: deva
region: global
type: image
```

A note below says: "Reach the right audience on a privacy-first ad network only for software devs: EthicalAds". At the bottom are links to "Read the Docs" and the version "v: v1.12.0".

The main content area has a header "CDAPS" and a subheader "Community Detection APplication and Service (CDAPS)". It describes CDAPS as performing multiscale community detection and functional enrichment for network analysis through a service-oriented architecture, integrating popular community detection algorithms and enrichment tools. All algorithms and tools run remotely on a dedicated server.

Under "Currently supported features:" is a bulleted list:

- Community detection algorithms: Louvain, Infomap, OSLOM, CLIQUO, HiDeF
- Functional enrichment tools: gProfiler, Enrichr, iQuery

To the right of the text is a screenshot of a Cytoscape window showing a network graph with nodes and edges, and a separate window showing a more detailed view of the network structure.

Figure: CDAPS framework.⁴²

⁴²<https://cdaps.readthedocs.io/>

Hierarchical community Decoding Framework (HiDeF)

The image shows two screenshots of the HiDeF documentation website. The left screenshot is the homepage, featuring a dark sidebar with navigation links like 'Tutorials', 'API', 'Read the Docs', 'Versions' (latest, stable), 'Downloads' (PDF, HTML, Epub), 'On Read the Docs', 'Project Home', 'Builds', 'Downloads', 'On GitHub', 'View', 'Edit', and 'Search'. It also includes a 'Search docs' bar and a footer with 'Hosted by Read the Docs - Privacy Policy'. The right screenshot shows the 'Welcome to HiDeF's documentation!' page, which includes a 'Search' bar, a GitHub edit link, and a main content area. The content area starts with a heading 'Welcome to HiDeF's documentation!' followed by a paragraph about the framework's purpose and how it integrates graph-based community detection and persistent homology. Below this is a section titled 'Installation' with a bullet point 'Welcome to HiDeF's documentation!'. Under 'Local installation of python package', there are two code snippets: 'python setup.py install' for source installation and 'pip install hidef' for pip installation.

Figure: HiDeF framework.⁴³

⁴³<https://hidef.readthedocs.io/>

Questions ??