# Uncovering Strategies of Viruses for Selection of Targets for Infection of Human Cells

Debomita Chakraborty[a], Nikhita Damaraju[a]

[a]*Department of Biotechnology, Indian Institute of Technology, Madras*

**Abstract**

Viral infection involves interactions between host and viral proteins. In this study, viral-host protein interactions for 53 out of the 77 human infecting viruses available in the Viruses.STRING database, are analyzed as a biological network problem. Three centrality measures: (i). degree centrality, (ii). closeness centrality, and (iii). subgraph centrality were calculated [1] in order to understand the viral-host interaction networks.

Three broad classes of viruses showed targeted human proteins having higher degree and closeness centralities and a low subgraph centrality than the average protein centrality of the nodes in the high confidence human PPI network.

The centrality measures for human proteins that interact with the viral proteins for Epstein-Barr virus (EBV) and Herpes Simplex Virus (HSV-1) do not show significant change before and after infection by the virus.

The type of human proteins that interact with host proteins differ with the class of virus as is observed on analyzing the most targeted proteins and their functional roles. However, no significant trend in centrality measures is seen among the highly targeted proteins.

## 1. Introduction

The mechanism underlying a viral infection involves the manipulation of host cellular processes for propagation. The interactions between viral and host proteins [2] lead to a disruption of the host protein network. A recently developed database Viruses.STRING [3] that consists of a set of 238 virus-host protein-protein interaction (PPI) networks, was added to the STRING repository that opened up the possibility of analyzing viral-host protein interactions as a biological network problem. In this study, the viral-host networks for 53 out of the 77 human infecting viruses available in the database, were analyzed. These viruses span 3 basic classes of human viruses based on their nucleic acid composition (Figure 1) according to the Baltimore classification of human viruses [4]:

### *1.1. Objectives*

The objective of this study is to understand the following key points for all the classes of viruses:

1. The network characteristics of the identified human proteins that interact with viral proteins in the host PPI network. [5]
2. The network characteristics of the human protein nodes in the viral-host protein interaction networks of EBV [6] and Herpes HSV-1 upon infection by the virus.
3. Functional analysis of human proteins that interact with viruses to uncover insights into differences and similarities between RNA [7]and DNA viruses while infecting human host cells. [8]

*Email addresses:* debomita02@gmail.com (Debomita Chakraborty), nikhita97@gmail.com (Nikhita Damaraju)

| | ds-DNA | ss-RNA (positive) | ss-RNA (negative) |
|---|---|---|---|
| 1 | Epstein-Barr virus | Dengue virus type 2 | Bunyavirus La Crosse |
| 2 | Hepatitis B virus | Dengue virus type 3 | Hantaan virus |
| 3 | Human adenovirus C | Dengue virus type 4 | Hendra virus |
| 4 | Human cytomegalovirus | Encephalomyocarditis virus | Human metapneumovirus |
| 5 | Human herpesvirus 1 | Hepatitis C virus | Human respiratory syncytial virus B |
| 6 | Human herpesvirus 2 | Hepatitis E virus | Influenza A virus |
| 7 | Human herpesvirus 6A | Human coronavirus 229E | Influenza B virus |
| 8 | Human herpesvirus 6B | Human hepatitis A virus | Influenza C virus |
| 9 | Human herpesvirus 8 type P | Human parechovirus 2 | Lake Victoria marburgvirus |
| 10 | Human papillomavirus type 16 | Human SARS coronavirus | Lassa virus |
| 11 | Human papillomavirus type 18 | Japanese encephalitis virus | Lymphocytic choriomeningitis virus |
| 12 | Human papillomavirus type 1 | Norwalk virus | Measles virus |
| 13 | Human papillomavirus type 4 | Poliovirus type 1 | Mumps virus |
| 14 | Human papillomavirus type 5 | Rubella virus | Rabies virus |
| 15 | Molluscum contagiosum virus | Semliki forest virus | Rift valley fever virus |
| 16 | Orf virus | Sindbis virus | Vesicular stomatitis Indiana virus |
| 17 | Varicella-zoster virus | West Nile virus | Zaire ebolavirus |
| 18 | Variola virus | | |
| 19 | Yaba monkey tumor virus | | |

Figure 1: Classification of the viruses from the string.Viruses database chosen for study

## 2. Methods

The analysis has been done using Pandas and Networkx libraries of Python. All codes and files generated have been uploaded on the GitHub repository: bt5240_project_viruses.

### *2.1. Extraction of data*

The virus-host interaction data is extracted from the viruses.STRING database for 53 viruses with the host as Homo sapiens. The human proteins that interact with viral proteins are filtered out for each virus using the identifier '9606' (Fig. 2). The human PPI data is then extracted from the STRING database and only those protein-protein interactions that have a combined score greater or equal to 900 are retained for analysis.

### *2.2. Network analysis using centrality measures*

The human proteins listed in the 53 viral-human interaction networks are located in the human PPI network. For each of these human proteins, an ego graph is constructed such that it includes only the nodes constituting its first, second and third neighbors in the human PPI network (Figure 3). The three centrality measures: (i). Degree centrality, (ii). Closeness centrality, and (iii). subgraph centrality, are calculated for each of the human proteins within the corresponding ego graph.

In the subsequent analysis, the viral-human interactome proteins are merged with the human PPI network to get a combined network (Figure 4). In this network, the ego graphs of each of the human proteins from the interactome, are constructed and the centrality measures of those proteins are calculated.

## 3. Results and Discussion

### *3.1. Analysis of human proteins used as targets across different classes of viruses*

It is known that DNA viruses are more prone to targeting immune specific proteins and cell cycle regulators while RNA viruses target RNA-transport and RNA-binding proteins [9]. Certain observations found in our analysis are coherent with existing literature:

1. The tumor suppressor protein p53 was found to be the most targeted protein across 19 DNA viruses. This is due to its dual role in initiating innate immunity based antiviral mechanisms in both tumor promoting and non-promoting viruses. [10] [11]
2. RNA-positive viruses target proteins belonging to the Polypyrimidine tract binding protein (PTPB) family. [12, 13]
3. Proteins targeted by RNA-negative viruses were found to belong to proteins in SFXN family that are crucial in the mitochondrial transport chain [14] No significant role of these proteins in viral infection has been found.

### *3.2. Centrality of interacting human proteins in the human PPI network*

#### *3.2.1. Degree centrality*

The cumulative frequency graph Figure 6 for the degree centrality of human proteins that constitute the interactome shows that most of these proteins have significantly higher degree centralities within the uninfected human PPI network than the average degree centrality of the network.

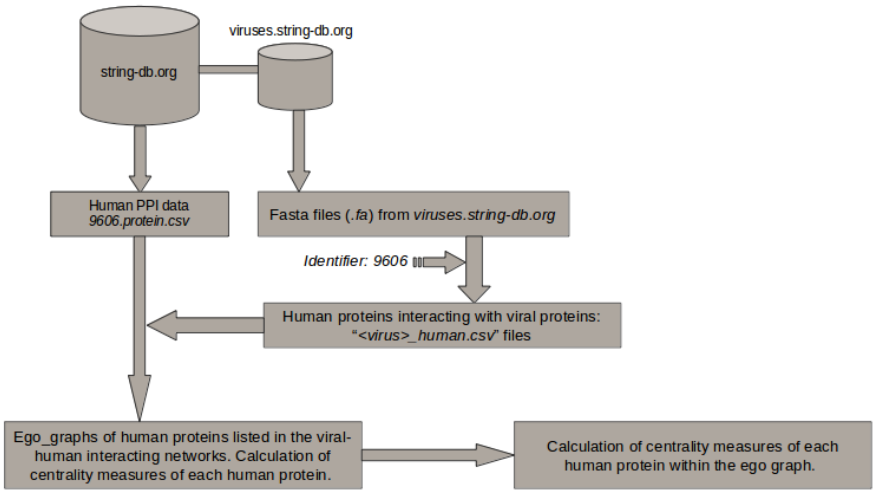


Figure 2: Data extraction from database and construction of graphs for analysis of centrality measures

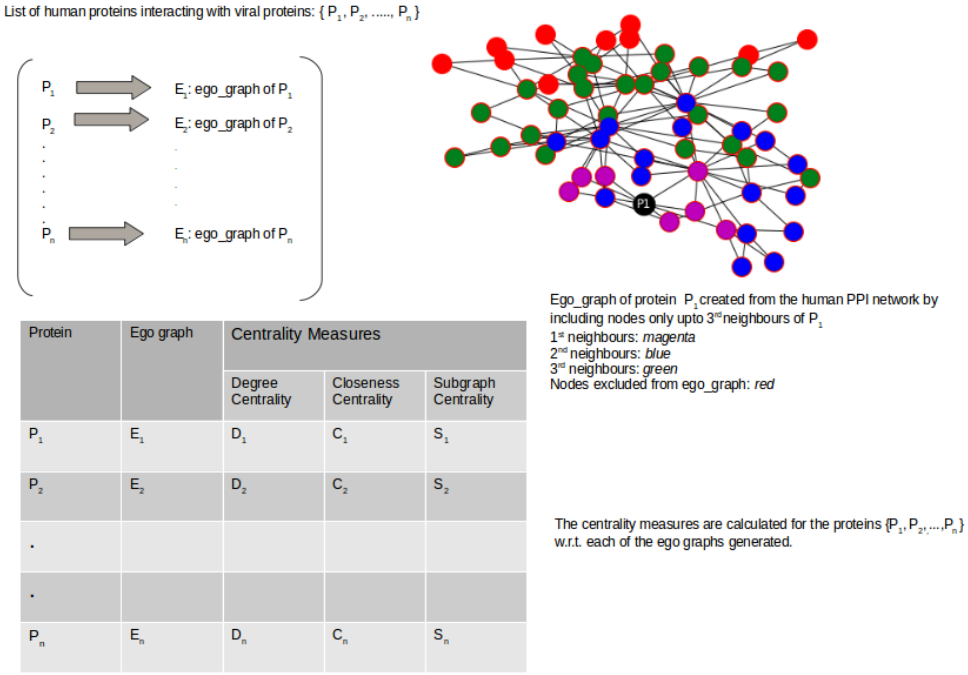


Figure 3: Ego graph construction and centrality measure calculations for analysis

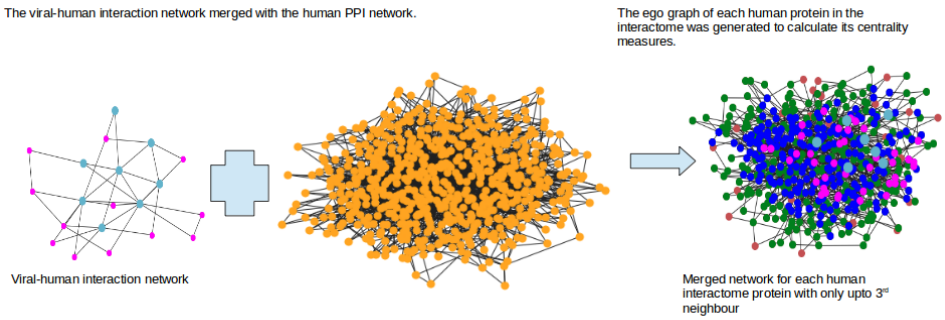


Figure 4: Viral-human interaction network is merged with the human PPI network to analyze the centrality of the human interacting proteins in this network. Viral nodes: *cyan*, human interacting proteins: *magenta*, 1st neighbours: *blue*, 2nd neighbours: *green*, 3rd neighbours: *red*

| DNA viruses | | |
|---|---|---|
| Protein name | occurrence | Protein description |
| TP53 | 8 | Tumor suppressor protein 53 |
| CDC25A | 5 | Tyrosine protein phosphatase which functions as a dosage-dependent inducer of mitotic progression |
| IRF3 | 5 | Interferon Regulatory Factor 3 plays an important role in innate immunity against DNA and RNA viruses |
| PTGS2 | 5 | PTGS2 is responsible for production of inflammatory prostaglandins |
| TERT | 4 | Active in progenitor and cancer cells. |

| RNA-positive viruses | | |
|---|---|---|
| Protein name | occurrence | Protein description |
| PTBP3 | 5 | polypyrimidine tract binding protein 3: aids in cell proliferation |
| PTBP2 | 5 | polypyrimidine tract binding protein 2: cell proliferation |
| NOG2 | 5 | GTPase that promotes cell proliferation |
| IFNB1 | 5 | Interferon Beta 1 precursor: has antiviral roles |
| IRAK3 | 5 | Interleukin-1 receptor-associated kinase 3 |

| RNA-negative viruses | | |
|---|---|---|
| Protein name | occurrence | Protein description |
| SFXN2 | 5 | Mitochondrial amino-acid transporter that mediates transport of serine into mitochondria |
| TLR4 | 4 | Cooperates with LY96 and CD14 to mediate the innate immune response |
| SFXN4 | 4 | Mitochondrial amino-acid transporter |
| CDSN | 4 | Important for the epidermal barrier integrity |
| SFXN3 | 4 | Mitochondrial serine transporter that mediates transport of serine into mitochondria |

Figure 5: Protein descriptions of the highly targeted human proteins for all 3 classes of viruses.

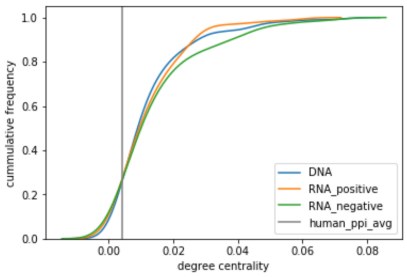


Figure 6: Cumulative frequency plot showing degree centralities for all human proteins interacting with viruses of the three classes: (i). DNA viruses, (ii). RNA positive viruses, and (iii). RNA negative viruses.

### *3.2.2. Closeness centrality*

Closeness centralities of all the interactome human proteins are significantly higher than the average closeness centrality of proteins in the human PPI network as seen in the plot Figure 7.

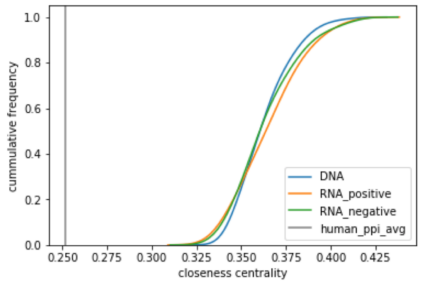


Figure 7: Cummulative frequency plot showing closeness centralities for all human proteins interacting with viruses

### *3.2.3. Subgraph centrality*

Subgraph centrality of a node measures the number of paths that start and end at it such that the path lengths carry lesser weights the further they traverse from the node. The subgraph centralities of the majority of the interactome human proteins are below the average subgraph centrality of nodes in the human PPI network.Figure 8

### *3.3. Comparison of centrality measures of human proteins across viral-human interactome and human PPI*

On comparing the values of the three centrality measures for the two viruses: (i). EBV,

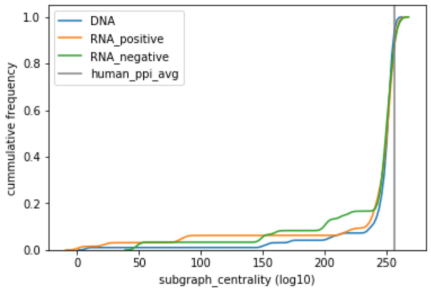


Figure 8: Plot showing cumulative subgraph centrality averaged over a log scale for all human proteins interacting with viruses

(ii). HHV1, it was observed that there was no significant change in the centrality measures upon introduction of the viral-host interaction network.(Figures 9 and 10)

## 4. Conclusion

Proteins targeted by the three classes of viruses were functionally different from each other and showed a significantly higher degree and closeness centralities but a lower value for subgraph centrality. The viral-host interactome and human PPI networks do not change the centralities of the interacting proteins.

## References


[1] K. Azhagesan, B. Ravindran, K. Raman, Network-based features enable prediction of essential genes across diverse organisms. PLoS, ONE 13 (12) (2018) 0208722.
URL https://doi.org/10.1371/journal.pone.0208722

[2] N. Gulbahce, H. Yan, A. Dricot, Viral perturbations of host networks reflect disease etiology, PLoS Comput Biol (2012).

[3] H. V. Cook, N. T. Doncheva, D. Szklarczyk, C. Mering, L. J. Jensen, Viruses.STRING: A Virus-Host Protein-Protein Interaction Database, Viruses, 2018.

[4] "Baltimore Classification of Viruses" (Website.) Molecular, Biology Web Book -, Retrieved on 2008 (8) 18.

[5] M. D. Dyer, Murali TM, Sobral BW. The landscape of human proteins interacting with viruses and other pathogens. PLoS Pathog (2008).

[6] M. Poorebrahim, A. Salarian, S. Najafi, M. F. Abazari, M. N. Aleagha, M. N. Dadras, S. M. Jazayeri, A. Ataei, V. Poortahmasebi, Regulatory network analysis of Epstein-Barr virus identifies functional modules and hub genes involved in infectious mononucleosis, Archives of Virology 162 (2017) 5–1309.

[7] P. O. Vidalain, F. Tangy, Virus-host protein interactions in RNA viruses, Microbes and Infection 12 (2010).

[8] S. Durmuş, Ü. KO, Comparative interactomics for virus-human protein-protein interactions: DNA viruses versus RNA viruses, FEBS Open Bio 1 (2017) 96–107.

[9] S. Durmuş, Ülgen KÖ. Comparative interactomics for virus-human protein-protein interactions: DNA viruses versus RNA viruses, FEBS Open Bio 1 (96–107) (2017) 4.

[10] C. Rivas, S. A. Aaronson, Munoz-Fontela C. Dual Role of p53 in Innate Antiviral Immunity, Viruses 1 (298–313).

[11] R. Aloni-Grinstein, M. Charni-Natan, H. Solomon, V. Rotter, p53 and the Viral Connection: Back into the Future ‡, Cancers (Basel) 2018 (10) (2018) 4.

[12] Identification of RNA-protein interaction networks involved in the norovirus life cycle.

[13] M. Niepmann, Evidence for an RNA chaperone func-


| prot_id | degree_centrality | degree_centrality in human ppi | closeness_centrality | closeness_centrality in human ppi | subgraph_centrality | subgraph centrality in human ppi |
|---|---|---|---|---|---|---|
| 9606.ENSP00000360266 | 0.02545661 | 0.02529607 | 0.39279608 | 0.39274419 | 1.38E+110 | 1.38E+110 |
| 9606.ENSP00000360266 | 0.02545661 | 0.02529607 | 0.39279608 | 0.39274419 | 1.38E+110 | 1.38E+110 |
| 9606.ENSP00000359531 | 0.01805898 | 0.01757126 | 0.36509819 | 0.36493611 | 3.17E+109 | 3.17E+109 |
| 9606.ENSP00000359531 | 0.01805898 | 0.01757126 | 0.36509819 | 0.36493611 | 3.17E+109 | 3.17E+109 |
| 9606.ENSP00000355231 | 0.00679844 | 0.00655242 | 0.3556461 | 0.35560335 | 1.04E+109 | 1.04E+109 |
| 9606.ENSP00000329967 | 0.00613223 | 0.00591518 | 0.36521704 | 0.3651777 | 8.15E+109 | 8.15E+109 |
| 9606.ENSP00000329967 | 0.00613223 | 0.00591518 | 0.36521704 | 0.3651777 | 8.15E+109 | 8.15E+109 |
| 9606.ENSP00000315768 | 0.01044831 | 0.01022395 | 0.35979368 | 0.35973554 | 2.61E+109 | 2.61E+109 |
| 9606.ENSP00000315768 | 0.01044831 | 0.01022395 | 0.35979368 | 0.35973554 | 2.61E+109 | 2.61E+109 |
| 9606.ENSP00000313007 | 0.02147201 | 0.02125808 | 0.37542519 | 0.37534413 | 1.84E+110 | 1.84E+110 |
| 9606.ENSP00000309555 | 0.00856315 | 0.00844999 | 0.3695379 | 0.36951254 | 7.34E+109 | 7.34E+109 |
| 9606.ENSP00000269305 | 0.04436316 | 0.04411903 | 0.41193061 | 0.41186932 | 2.33E+111 | 2.33E+111 |
| 9606.ENSP00000269305 | 0.04436316 | 0.04411903 | 0.41193061 | 0.41186932 | 2.33E+111 | 2.33E+111 |

Figure 9: Comparison of centrality measures for proteins targeted by HHV1

| prot_id | degree_centrality | degree_centrality in human ppi | closeness_centrality | closeness_centrality in human ppi | subgraph_centrality | subgraph centrality in human ppi |
|---|---|---|---|---|---|---|
| 9606.ENSP00000369373 | 0.006852926 | 0.35329174 | 4.75E+108 | 0.006592827 | 0.353187724 | 4.75E+108 |
| 9606.ENSP00000362994 | 0.004253568 | 0.34571415 | 3.19E+108 | 0.004116921 | 0.345683112 | 3.19E+108 |
| 9606.ENSP00000345702 | 0.003768844 | 0.350746634 | 1.20E+107 | 0.003631792 | 0.350707882 | 1.20E+107 |
| 9606.ENSP00000345206 | 0.009656652 | 0.363904555 | 5.08E+109 | 0.009424958 | 0.36384946 | 5.08E+109 |
| 9606.ENSP00000340330 | 0.023542001 | 0.381802582 | 6.36E+110 | 0.023447537 | 0.381769875 | 6.36E+110 |
| 9606.ENSP00000262238 | 0.014963142 | 0.373448928 | 3.18E+110 | 0.014861295 | 0.373396909 | 3.18E+110 |
| 9606.ENSP00000262105 | 0.010873462 | 0.372900236 | 1.68E+110 | 0.010757829 | 0.372866248 | 1.68E+110 |
| 9606.ENSP00000252996 | 0.012851986 | 0.360902648 | 2.31E+109 | 0.012718601 | 0.360853239 | 2.31E+109 |
| 9606.ENSP00000239938 | 0.010561423 | 0.360853693 | 8.24E+108 | 0.010344828 | 0.360797849 | 8.24E+108 |

Figure 10: Comparison of centrality measures for proteins targeted by EBV

tion of polypyrimidine tract-binding protein in picornavirus translation, Song, Yutong & Tzima, Eleni & Ochs, Kerstin & Bassili, Gergis & Trusheim, Heidi & Linder, Monica & T Preissner, Klaus 11 (2006) 1809–24.

[14] E. E. Mon, F. Y. Wei, R. N. R. Ahmad (2019).