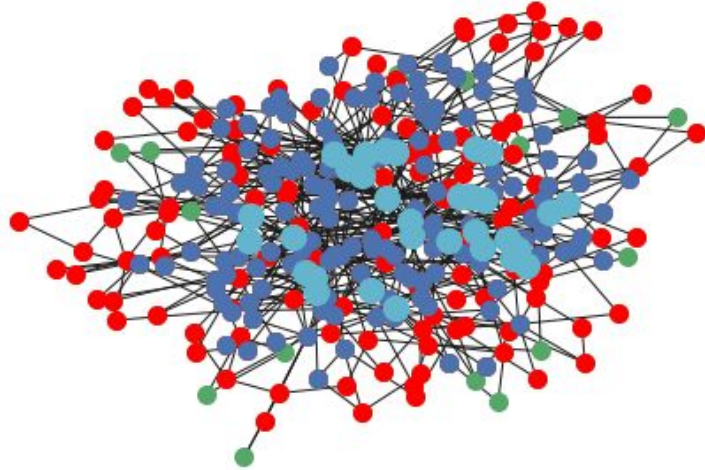# Uncovering **Strategies of Viruses** for Selection of **Targets for Infection** of Human Cells

Debomita Chakraborty (BT18D301)
Nikhita Damaraju (BS15B022)

Outline

Introduction

1

An overview of the motivation of the project and key concepts

Methodology

2

Methods used in the analysis of the data in each step

Results and Discussion

3

Analysis of results and coherence with existing literature

Challenges

4

Summary of key findings and challenges

Introduction

- Viral infection involves interaction between viral and human proteins

- Classes of viruses based on nucleic acid composition:
  - DNA viruses
  - RNA viruses

- Viral-host protein interactions as a biological network problem

Objectives

- Understanding centrality measures of human interactome proteins

- Change in centrality of human interactome proteins upon infection by Epstein Barr Virus (EBV) & Herpes Simplex Virus (HSV-1)

- Functional similarities/dissimilarities of human interactome proteins for DNA and RNA virus classes

Viruses.STRING Database

- Database of known and predicted protein-protein interactions

- More than 9 million proteins from 2031 organisms (Version 10.5)

- The interactions include physical and functional associations

Data Sources

Interactions in viruses.STRING are derived from five main sources:

Genomic Context Predictions
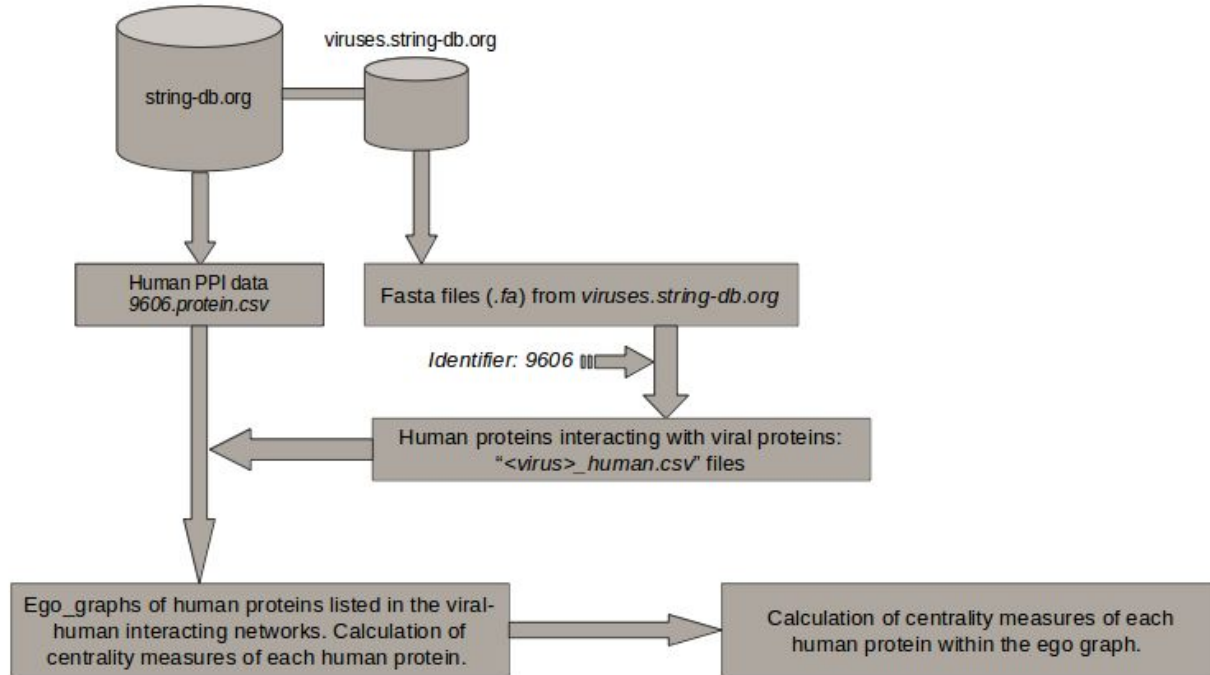
High-throughput Lab Experiments

(Conserved) Co-Expression

Automated Textmining

Previous Knowledge in Databases

Extraction of data

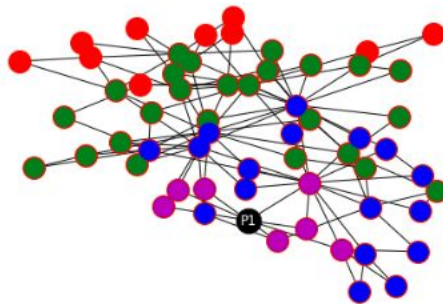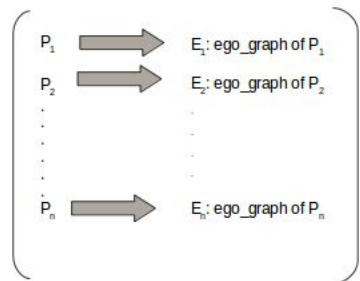|	ds-DNA	ss-RNA (positive)	ss-RNA (negative)
1	Epstein-Barr virus	Dengue virus type 2	Bunyavirus La Crosse
2	Hepatitis B virus	Dengue virus type 3	Hantaan virus
3	Human adenovirus C	Dengue virus type 4	Hendra virus
4	Human cytomegalovirus	Encephalomyocarditis virus	Human metapneumovirus
5	Human herpesvirus 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		

- Data for 51 organisms was downloaded
- Viruses were categorized into 3 broad classes-
  - DNA (double stranded)
  - RNA (positive single stranded)
  - RNA (negative single stranded)
- Human proteins interacting with each virus were extracted

Network Analysis Using Centrality Measures

List of human proteins interacting with viral proteins: { P_1, P_2, ..., P_n }

$$P_1 \Rightarrow E_1: \text{ego\_graph of } P_1$$
$$P_2 \Rightarrow E_2: \text{ego\_graph of } P_2$$
$$\vdots$$
$$P_n \Rightarrow E_n: \text{ego\_graph of } P_n$$

Ego_graph of protein P_1 created from the human PPI network by including nodes only upto 3rd neighbours of P_1

1st neighbours: *magenta*
2nd neighbours: *blue*
3rd neighbours: *green*
Nodes excluded from ego_graph: *red*

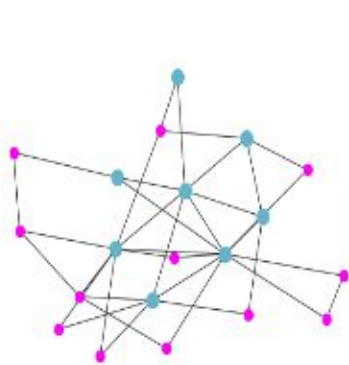| Protein	Ego graph	Centrality Measures		
		Degree Centrality	Closeness Centrality	Subgraph Centrality
P_1	E_1	D_1	C_1	S_1
P_2	E_2	D_2	C_2	S_2
.				
.				
P_n	E_n	D_n	C_n	S_n

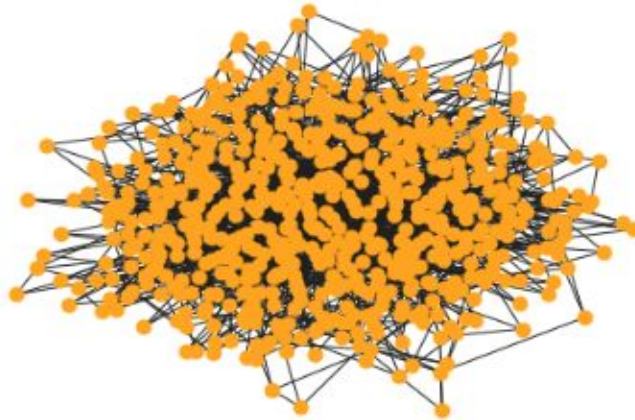The centrality measures are calculated for the proteins { P_1, P_2,...,P_n } w.r.t. each of the ego graphs generated.
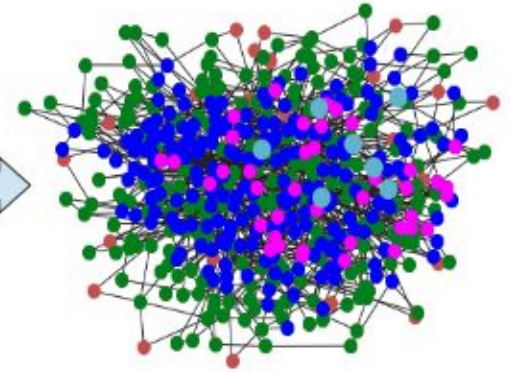
Viral-human + Human PPI Network

The viral-human interaction network merged with the human PPI network.

The ego graph of each human protein in the interactome was generated to calculate its centrality measures.

Viral-human interaction network

Merged network for each human interactome protein with only upto 3rd neighbour

Centrality measures

Centrality measure	Bootstrap test	Wilcoxon Rank-Sum test
Edge Clustering Coefficient Centrality	0	8
Betweenness Centrality	27	23
Load Centrality	27	24
Random Walk Betweenness Centrality	19	25
Information Centrality	19	26
Closeness Centrality	27	26
Degree Centrality	27	26
Harmonic Centrality	27	26
PageRank	27	26
Reaching Centrality	27	26
Subgraph Centrality	27	26
Eigenvector Centrality	27	27

Table shows the number of organisms in which a given measure was found to be significant (p-value <0.05). For further details on p-value computation, refer text.

https://doi.org/10.1371/journal.pone.0208722.t002

Centrality measures chosen were filtered based on:
- Choice in other papers analysing PPI networks for smaller organisms
- Computational feasibility to execute in current network
- Relevance to the undirected nature of graph for analysis

Measures chosen based on above criteria were: **degree centrality, closeness centrality** and **subgraph centrality**
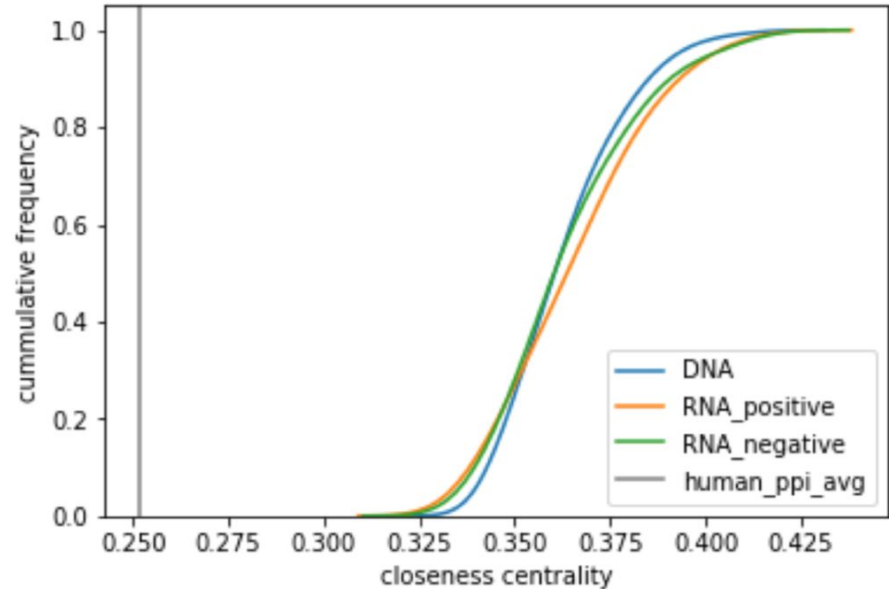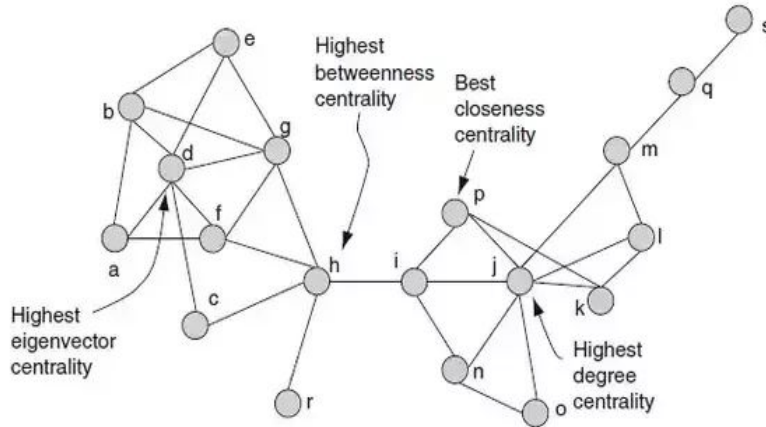
Degree centrality

- Number of neighbors of a node

- Highly connected nodes or "hubs" are more likely to be essential - centrality-lethality hypothesis

- Higher degree centrality nod are preferred

- Average human degree centrality: 0.004219
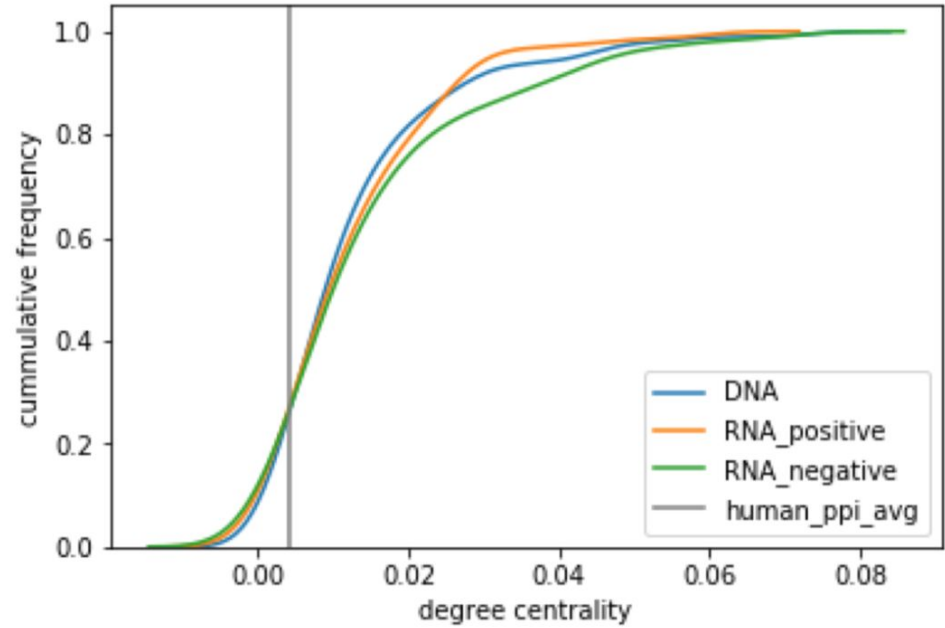
Closeness centrality

- Measure of reach to all other nodes in graph

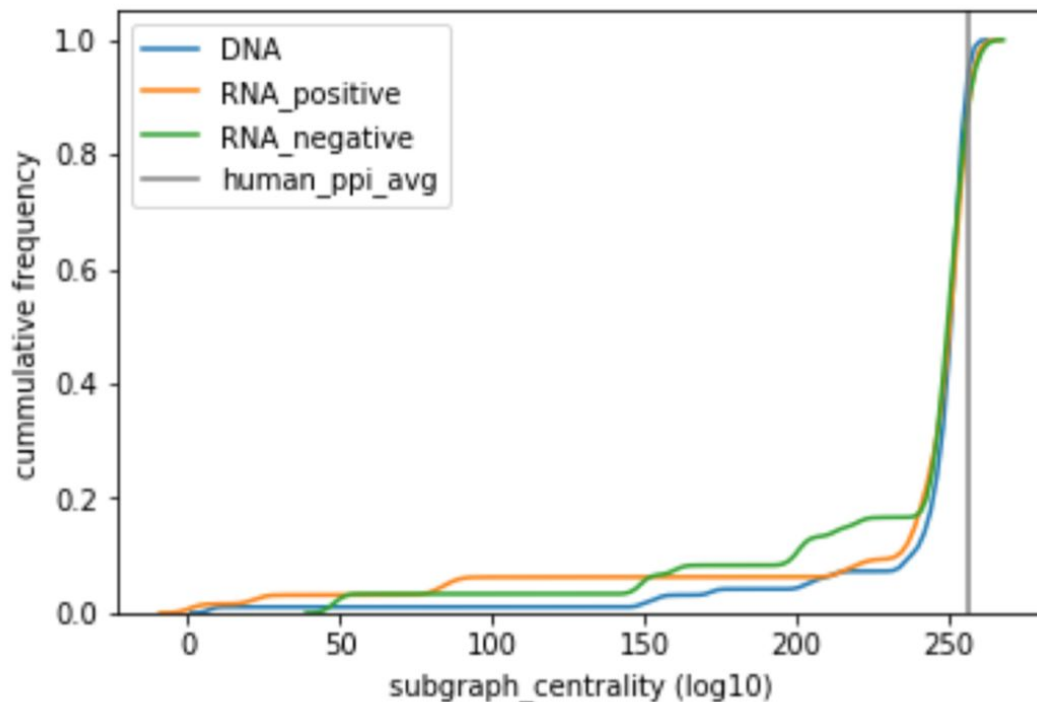- Higher closeness centrality nodes are preferred

- Average human closeness centrality: 0.2517

Subgraph centrality

- Participation of a node in all possible subgraphs of a network

- A decrease was observed for a large number of proteins

- Human average subgraph centrality: 3.18e + 111

Viral-host interactome vs human PPI

Centrality measures of interacting proteins before and after addition of interactome network did **not** change significantly

- Interacting proteins are not given a central status

- Addition of interactome network does not change the host network significantly

- No. of viral nodes added to the overall network is less in comparison to human PPI

Image source: Viruses. 2018 Oct; 10(10): 519.
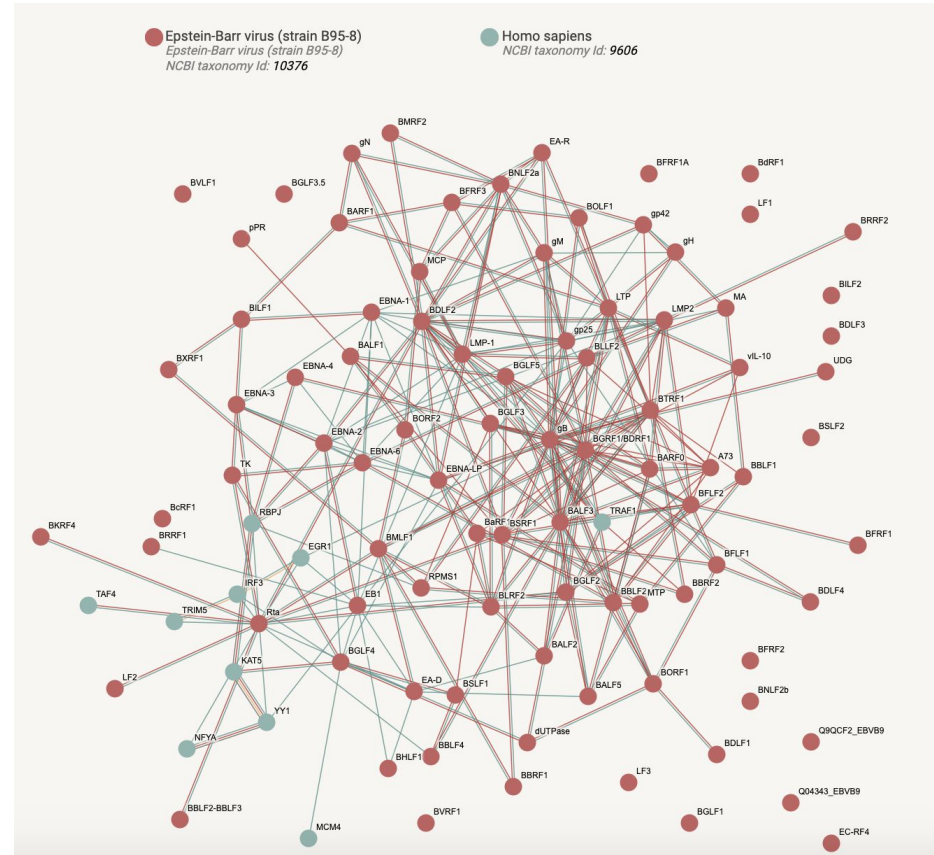
Functional analysis of target proteins

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

Challenges

- Small interactome network size

- Incomplete interaction data for most viruses except EBV and HSV-1

- Choice of suitable parameters of analysis for 53 viral networks

- Missing annotation for certain human proteins

- Missing annotation for most viral proteins

Thank You