Detection of spreader nodes and ranking of interacting edges in Human-SARS-CoV protein interaction network

Sovan Saha[1,*], Piyali Chatterjee[2], Subhadip Basu[3], Mita Nasipuri[4]

[1,*] Corresponding Author, Department of Computer Science & Engineering, Dr. Sudhir Chandra Sur Degree Engineering College, Kolkata

[2] Department of Computer Science & Engineering, Netaji Subhash Engineering College, Kolkata

[3, 4] Department of Computer Science & Engineering, Jadavpur University, Kolkata

Abstract: The entire world has recently witnessed the commencement of coronavirus disease 19 (COVID-19) pandemic. It is caused by a novel coronavirus (n-CoV) generally distinguished as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). It has exploited human vulnerabilities to coronavirus outbreak. SARS-CoV-2 promotes fatal chronic respiratory disease followed by multiple organ failure which ultimately puts an end to human life. No proven vaccine for n-CoV is available till date in spite of significant research efforts worldwide. International Committee on Taxonomy of Viruses (ICTV) has reached to a consensus that the virus SARS-CoV-2 is highly genetically similar to Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) outbreak of 2003. It has been reported that SARS-CoV has ~89% genetic similarities with n-CoV. With this hypothesis, the current work focuses on the identification of spreader nodes in SARS-CoV protein interaction network. Various network characteristics like edge ratio, neighborhood density and node weight have been explored for defining a new feature spreadability index by virtue of which spreader nodes and edges are identified. The selected top spreader nodes having high spreadability index have been also validated by Susceptible-Infected-Susceptible (SIS) disease model. Initially, the proposed method is applied on a synthetic protein interaction network followed by SARS-CoV-human protein interaction network. Hence, key spreader nodes and edges (ranked edges) are unmasked in SARS-CoV proteins and its connected level 1 and level 2 human proteins. The new network attribute spreadability index along with generated SIS values of selected top spreader nodes when compared with the other network centrality based methodologies like Degree centrality (DC), Closeness centrality (CC), Local average centrality (LAC) and Betweeness centrality (BC) is found to perform relatively better than the *existing-state-of-art*.

Keywords: Spreader nodes, Spreader edges, Spreadability Index, Severe Acute Respiratory Syndrome Coronavirus, Protein-Protein Interaction Network, Node Weight, Neighborhood Density, Edge Ratio, Susceptible-Infected-Susceptible model.

1. Introduction

The pandemic COVID-19 registered its first case on 31 December 2019 [1]. It laid its foundation in the Chinese city of Wuhan (Hubei province) [2]. Soon, it made several countries all over the world [3] as its victim by community spreading which ultimately compelled World Health Organization (WHO) to declare a global health emergency on 30 January 2020 [4] for the massive outbreak of COVID-19. Owing to its expected fatality rate, which is about 4%, as projected by WHO [5], researchers from nations, all over the world, have joined their hands to work together to understand the spreading mechanisms of this virus SARS-CoV-2 [6]–[9] and to find out all possible ways to save human lives from the dark shadow of COVID-19.

Coronavirus belongs to the family Coronaviridae. This single stranded RNA virus not only affects humans but also mammals and birds too. Due to coronavirus, common fever/flu symptoms are noted in humans followed by acute respiratory infections. Nevertheless, corona viruses like Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) have the ability to create global pandemic due to their infectious nature. Both of these corona viruses are the member of genus Betacoronavirus under Coronaviridae. SARS created a major outbreak in 2003 originating from Southern China. 774 deaths were reported among 8098 globally registered cases resulting in an estimated fatality rate of 14%–15% [10]. While MERS commenced in Saudi Arabia creating an endemic in 2012. The world witnessed 858 deaths among 2494 registered positive cases. It generated high fatality rate of 34.4% in comparison to SARS [11].

SARS-CoV-2 is under the same Betacoronavirus genus as that of MERS and SARS coronavirus [12]. It comprises of several structural and non-structural proteins. The structural proteins includes the envelope (E) protein, membrane (M) protein, nucleocapsid (N) protein and the spike (S) protein. Though SARS-CoV-2 has been identified recently but there is an intense scarcity of data as well as necessary information needed to gain immunity against SARS-CoV-2. Studies has revealed the fact that SARS-CoV-2 is highly genetically similar to SARS-CoV based on several experimental genomic analyses [9], [12]–[14]. This is also the reason behind the naming of SARS-CoV-2 by International Committee on Taxonomy of Viruses (ICTV) [15]. Due to this genetic similarity, immunological study of SARS-CoV may lead to the discovery of SARS-CoV-2 potential drug development.

In the proposed methodology, Protein-protein interaction network (PPIN) has been used as the central component in identification of spreader nodes in SARS-CoV. PPIN has been found to be very effective module for protein function determination [16]–[18] as well as in the identification of central/essential or key spreader nodes in the network [19]–[23]. Compactness of the network and its transmission capability is estimated using centrality analysis. Anthonisse *et al.* [23] proposed a new centrality measure named as Betweenness Centrality (BC) . BC is actually defined as the measure of impact of a particular node over the transmission between every pair of

nodes with a consideration that this transmission is always executed over the shortest path between them. It is defined as:

$$C_B^{(u)} = \sum_{s \neq u \neq t} \frac{\rho(s,u,t)}{\rho(s,t)}$$

where $\rho(s,t)$ is the total number of shortest paths from node s to node t and $\rho(s,u,t)$ is the number of those paths that pass through u. Another centrality measure, called as closeness centrality (CC), was defined by Sabidussi *et al.* [24]. It is actually a procedure for detecting nodes having efficient transmission capability within a network. Nodes which have high closeness centrality value are considered to have the shortest distance to all available nodes in the network. It can be mathematically expressed as:

$$C_C^{(u)} = \frac{|N_u| - 1}{\sum_{v \in V} dist(u,v)}$$

where $|N_u|$ denotes the number of neighbors of node u and $dist(u,v)$ is the distance of the shortest path from node u to node v. Two other important centrality measures: Degree centrality (DC) [19] and Local average centrality (LAC) [21] are also found to be very effective in this area of research. DC is considered to be the most simplest among the available centrality measures which only count the number of neighbors of a node. Nodes having high degree is said to be highly connected module of the network. It is defined as:

$$C_D^{(u)} = |N_u|$$

where $|N_u|$ hold the same meaning as state above. LAC is defined to be the local metric to compute the essentiality of node with respect to transmission ability by considering its modular nature, the mathematical model of which is highlighted as:

$$LAC(u) = \frac{\sum_{w \in N_u} deg_{C_u}^{w}}{|N_u|}$$

where C_u is the subgraph induced by N_u and $deg_{C_u}^{w}$ is the total number of nodes which are directly connected in C_u.

Due to high morbidity and mortality of SARS-CoV2, it has been felt that there is a pressing need to properly understand the way of disease transmission from SARS-CoV-2 PPIN to human PPIN [16]–[18]. But since SARS-CoV-2 PPIN is still not available, SARS-CoV PPIN is considered for this research study due to its high genetic similarity with SARS-CoV-2. In the proposed methodology, at first SARS-CoV-Human PPIN (up to level 2) is formed from the collected datasets [25]–[27]. Once it is formed, spreader nodes are identified in each of SARS-CoV proteins, its level 1 and level 2 of human network by the application of a new network attribute

i.e. spreadability index which is a combination of three terminologies: 1) edge ratio [28] 2) neighborhood density [28] and 3) node weight [29]. The detected spreader nodes are also validated by the existing SIS epidemic disease model [30]. Then the edges connecting two spreader nodes are ranked based on the average of spreadability index of spreader nodes themselves to access to the spreading ability of the corresponding edge. The ranked edges thus highlight the path of entire disease propagation from SARS-CoV to human level 1 and then from human level 1 to human level 2 proteins.

2. Methodology

Spreader nodes and edges both play a crucial role in transmission of infection from one part of the network to another. Generally in a disease specific PPIN models, at least two entities are involved: one is pathogen/Bait and the other is host/Prey [31]. In this research work, SARS-CoV takes the role of the former while human later. SARS-CoV first transmits infection to its corresponding interaction human proteins which in turn affect its next level of proteins. So, the transmission occurs through connected nodes and edges. Not all the nodes or proteins in the PPIN transmit infection. So, proper identification of nodes transmitting infection (spreader nodes) is required. It is simultaneously also true that the transmission is not possible without the edges connecting two spreader nodes. Thus these connecting edges are called spreader edges. The proposed methodology involves a proper study and assessment of various existing established PPIN features followed by the identification of spreader nodes, which has been also verified by SIS model.

2.1 Basic Network Characteristics

2.1.1 Ego Network

Ego network of node i (S_i) [28] is defined as the grouping of node i itself along with its corresponding level 1 neighbors and interconnections. N (S_i) [28] consists of the set of nodes which belong to the ego network, S_i i.e. $\{i\} \cup \Gamma(i)$.

2.1.2 Edge ratio

Edge ratio of node i [28] is defined by the following equation:

$$Edge\ ratio\ (i) = \frac{\left(\sum_{j\in(i)} |\Gamma\ (j) - N(S_i)|\right)+1}{\left(\frac{1}{2}\sum_{j\in\Gamma\ (j)} |\Gamma^{S_i}(J)|\right)+1} = \frac{E_{out}^{S_i}+1}{E_{in}^{S_i}+1}$$

where $E_{out}^{S_i}$ is the total number of outgoing edges from the ego network S_i and $E_{in}^{S_i}$ is the total number of interactions among node i's neighbors, respectively. Γ (i) denotes the level 1 neighbors of node i. S_i is considered to be Ego network. $\Gamma^{S_i}(j)$ denotes node j's neighbors which

belongs S_i. In the edge ratio, $E_{out}^{S_i}$ is positively related to the non-peripheral location of node i. Large number of interactions resulting from the ego network actually denotes that the node has high level of interconnectivity between its neighbors. On the other hand, $E_{in}^{S_i}$ is negatively related to the inter-module location of node i. It represents the fact that the interconnectivity between neighbors is usually connected to the number of structural holes available around the node. When neighbor's interconnectivity is low, the root or the central node i gain more control of the flow of transmission among the neighbors.

2.1.3 Jaccard Dissimilarity

Jaccard dissimilarity [32] of node i and j ($dissimilarity\ (i,j)$) is defined as:

$$dissimilarity\ (i,j) = 1 - sim(i,j) = \frac{|\Gamma(i) \cap \Gamma(j)\ |}{|\Gamma(i) \cup \Gamma(j)|}$$

where $|\Gamma(i) \cap \Gamma(j)\ |$ refers to the number of common neighbors of neighbors of i and j. $|\Gamma(i) \cup \Gamma(j)|$ is the total number of neighbors of i and j. Similarity between two nodes is determined by Jaccard similarity based on their common neighbors. The similarity degree between i and j is considered to be more when they have more number of common neighbors. Whereas, when dissimilarity between the neighbors of a node is high, it guarantees that the only common node among the neighbors is the central node, which is termed as structural hole situation [28].

2.1.4 Neighborhood Density

The neighborhood diversity [28] is defined as:

$$neighborhood_diversity\ (i) = \sum_{j,k \in (i)} dissimilarity(j,k)$$

When a node's neighborhood density reaches its greatest value, it reveals the fact that the neighbors have no other closer path. Hence, the neighbors should transmit or communicate through this node.

2.1.5 Node weight

Node weight w_v of node $v \in V$ in PPIN [29] is interpreted as the average degree of all nodes in $G_{V'}$, a sub-graph of a graph G_V. It is considered as another measure to determine the strength of connectivity of a node in a network. Mathematically, it is represented by

$$w_v = \frac{\sum_{u \in V''} \deg(u)}{|V''|}$$

where, V'' is the set of nodes in $G_{V'}$. $|V''|$ is the number of nodes in $G_{V'}$. And $deg(u)$ is the degree of a node $u \in V''$.

2.2 Spreadability Index and Spreader nodes

Spreadability index of node i is defined as the ability of node i to transmit an infection from one node to another. Mathematically it can be defined as:

$$Spreadability_index\ (i)\ = (Edge\ ratio\ (i)\ \times neighborhood_diversity\ (i)\ ) + Node\ weight\ (w_i)$$

Nodes having high spreadability index are termed as spreader nodes i.e. these nodes have the capability to transmit the infection to its maximum number of interconnected neighbors in a much short amount of time in comparison to the other nodes in PPIN.

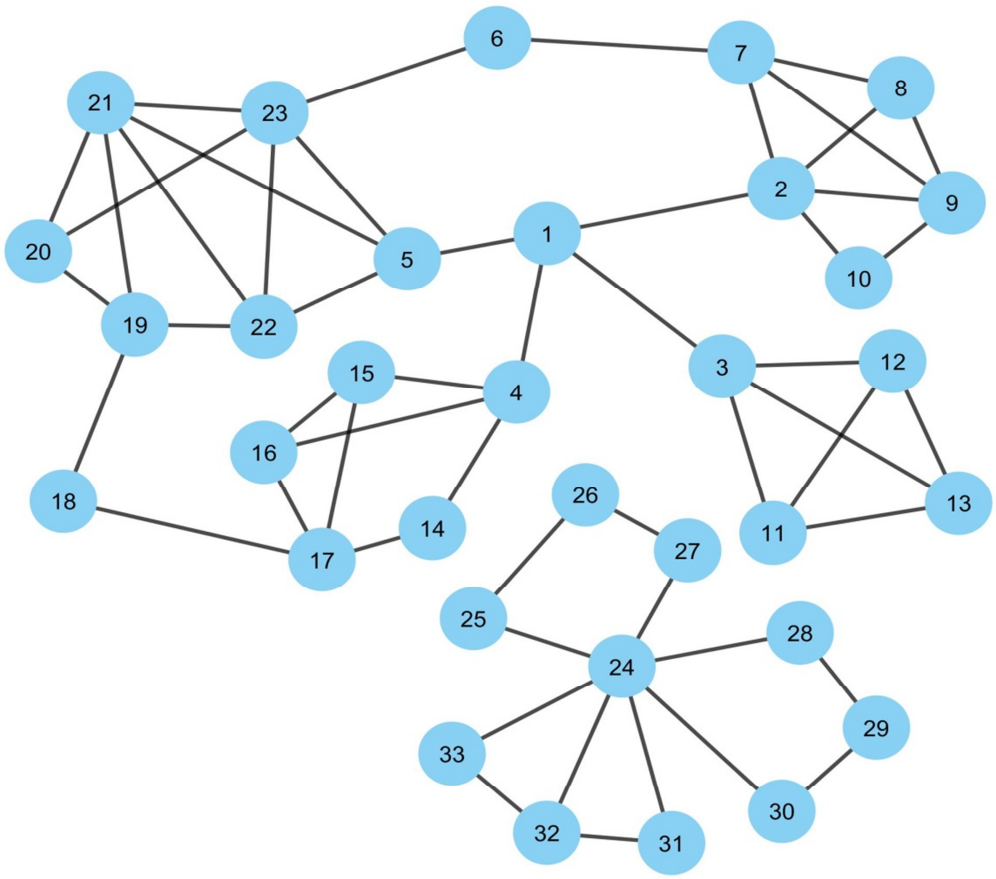

Figure 1 Synthetic PPIN1 consisting of 33 nodes and 53 edges where each node represents a protein and its interaction is represented by edges.

Figure 1 represents a sample PPIN where each protein is denoted as nodes while its interaction by edges. In this network, spreadability index is computed by using basic network characteristics as stated earlier which has been highlighted and compared to DC, BC, CC and LAC in Table 1 to Table 5.

Table 1 Computation of spreadability index of synthetic figure 1 along with computation of infection rate of selected top 10 spreader nodes by SIS model.

Rank	**Proteins**	$E_{out}^{S_i}$	$E_{in}^{S_i}$	**Edge Ratio**	**Neighborhood Density**	**Node Weight**	**Spreadability Index**	**Sum of SIS infection rate of top 10 nodes**
1	1	13	0	14.0	5.19	3.40	76.15	**2.46**
2	24	4	2	1.66	12.5	1.87	22.70	
3	4	6	1	3.50	3.63	2.40	15.11	
4	5	8	3	2.25	4.39	3.60	13.48	
5	19	6	2	2.33	3.8	2.80	11.66	
6	23	5	4	1.20	6.58	3.00	10.89	
7	17	4	1	2.50	3.33	2.00	10.33	
8	6	7	0	8.00	0.87	3.00	10.00	
9	2	4	4	1.00	6.84	2.83	9.68	
10	22	6	4	1.40	3.88	3.60	9.03	
11	25	7	0	8.00	0.71	3.00	8.71	_ _
12	27	7	0	8.00	0.71	3.00	8.71	
13	28	7	0	8.00	0.71	3.00	8.71	
14	30	7	0	8.00	0.71	3.00	8.71	
15	18	6	0	7.00	0.85	2.66	8.66	
16	20	7	2	2.66	1.78	3.50	8.26	
17	7	4	3	1.25	4.15	2.80	7.98	
18	21	3	6	0.57	6.66	3.33	7.13	
19	3	3	3	1.00	4.00	2.60	6.60	
20	16	4	2	1.66	2.06	2.75	6.19	
21	15	4	2	1.66	2.06	2.75	6.19	
22	31	6	1	3.50	0.75	3.33	5.95	
23	33	6	1	3.50	0.75	3.33	5.95	
24	32	4	2	1.66	1.75	2.75	5.66	
25	8	4	3	1.25	1.88	3.25	5.60	
26	14	6	0	7.00	0.40	2.66	5.46	
27	9	2	4	0.60	3.64	2.80	4.98	
28	10	5	1	3.00	0.50	3.00	4.50	
29	13	1	3	0.50	1.70	2.50	3.35	
30	11	1	3	0.50	1.70	2.50	3.35	
31	12	1	3	0.50	1.70	2.50	3.35	
32	29	2	0	3.00	0.00	1.33	1.33	
33	26	2	0	3.00	0.00	1.33	1.33	

Table 2 Computation of CC of synthetic figure 1 along with computation of infection rate of selected top 10 spreader nodes by SIS model.

Rank	Proteins	Closeness Centrality	Sum of SIS infection rate of top 10 nodes
1	1	0.085	1.94
2	5	0.083	
3	2	0.082	
4	4	0.082	
5	23	0.081	
6	3	0.081	
7	21	0.081	
8	22	0.081	
9	7	0.08	
10	15	0.08	
11	16	0.08	– –
12	19	0.079	
13	14	0.079801	
14	9	0.079602	
15	20	0.079602	
16	6	0.079602	
17	8	0.079404	
18	17	0.078818	
19	10	0.078624	
20	11	0.078049	
21	12	0.078049	
22	13	0.078049	
23	18	0.07767	
24	24	0.041558	
25	32	0.041237	
26	28	0.041237	
27	30	0.041237	
28	25	0.041237	
29	27	0.041237	
30	31	0.041184	
31	33	0.041184	
32	29	0.040921	
33	26	0.040921	

Table 3 Computation of BC of synthetic figure 1 along with computation of infection rate of selected top 10 spreader nodes by SIS model.

Rank	Proteins	Betweeness Centrality	Sum of SIS infection rate of top 10 nodes
1	1	269.1	2.2
2	2	117.93	
3	4	117.1	
4	3	114	
5	5	108	
6	24	57	
7	23	56.4	
8	19	45.56	
9	17	39.1	
10	7	36.9	
11	6	32.9	– –
12	18	32	
13	21	29.36	
14	22	20.53	
15	16	12.1	
16	15	12.1	
17	14	12.1	
18	28	7	
19	30	7	
20	25	7	
21	27	7	
22	20	6.63	
23	9	4.16	
24	32	1	
25	29	1	
26	26	1	
27	8	0	
28	11	0	
29	12	0	
30	13	0	
31	10	0	
32	31	0	
33	33	0	

Table 4 Computation of LAC of synthetic figure 1 along with computation of infection rate of selected top 10 spreader nodes by SIS model.

Rank	Proteins	Local Average Centrality	Sum of SIS infection rate of top 10 nodes
1	21	2.4	2.19
2	9	2	
3	22	2	
4	8	2	
5	11	2	
6	12	2	
7	13	2	
8	2	1.6	
9	23	1.6	
10	7	1.5	
11	3	1.5	– –
12	5	1.5	
13	16	1.33	
14	15	1.33	
15	20	1.33	
16	32	1.33	
17	19	1	
18	10	1	
19	31	1	
20	33	1	
21	24	0.57	
22	4	0.5	
23	17	0.5	
24	1	0	
25	14	0	
26	18	0	
27	6	0	
28	28	0	
29	29	0	
30	30	0	
31	25	0	
32	26	0	
33	27	0	

Table 5 Computation of DC of synthetic figure 1 along with computation of infection rate of selected top 10 spreader nodes by SIS model.

Rank	Proteins	Degree Centrality	Sum of SIS infection rate of top 10 nodes
1	24	7	2.3
2	2	5	
3	23	5	
4	21	5	
5	1	4	
6	7	4	
7	9	4	
8	3	4	
9	4	4	
10	17	4	
11	5	4	– –
12	22	4	
13	19	4	
14	8	3	
15	11	3	
16	12	3	
17	13	3	
18	16	3	
19	15	3	
20	20	3	
21	32	3	
22	10	2	
23	14	2	
24	18	2	
25	6	2	
26	28	2	
27	29	2	
28	30	2	
29	25	2	
30	26	2	
31	27	2	
32	31	2	
33	33	2	

In Figure 1, it can be observed that nodes 1, 24 are clearly the most essential spreaders. Node 1 connects the four densely connected modules of the network which turns this node to stand in the

first position having highest spreadability index by our proposed methodology. This node has been correctly ranked by all the methods except LAC and DC. Node 24 though has moderate edge ratio and node weight, but is one of the most densely connected modules itself in spite of getting isolated from the main network module of node 1. Moreover, node 24 has the highest neighborhood density. It establishes the fact that the only path of transmission for nodes 26, 27, 25, 28, 29, 30, 31, 32 and 33 is node 24. Thus if node 24 gets affected, then all the connected nodes with it will be immediately affected due to the lack of the connectivity of the neighbors with other central nodes. So, node 24 holds the second position with respect to spreadability index in our proposed methodology. Node 24 is not properly identified as the second most influential spreader nodes by the other methods. Further assessment of the remaining nodes highlights the fact that the performance of the new attribute spreadability index in our proposed methodology is relatively better in comparison to the others.

2.3 Validation of spreader nodes by SIS model

The SIS Epidemic Model [30] is used in this proposed methodology as a way of modeling SARS-CoV epidemic by classifying the proteins in SARS-CoV-human PPIN based on their disease/infection status. SIS actually refers to **S**usceptible, **I**nfected and **S**usceptible states which are generally considered as the three probable states of a protein in a PPIN. 1) **S** - The susceptible proteins who are not yet infected but are at risk for getting infected. In general every protein in PPIN is initially in susceptible state. 2) **I** - Infected proteins who are infected and are capable of transmitting the disease to other proteins. 3) **S** - proteins who have recovered and again become susceptible. Infection rate of the network, recovery rate (general assumption is any one out of the infected proteins gets recovered in one day) and total number of proteins are usually provided as input to SIS. If a protein gets infected and it has many neighbors, then any neighbor can get infected or may not be. So the final result is generated after 50 iterations for each infected protein. The total number of susceptible proteins after 50 iterations in the neighborhood of each infected protein divided by the total number of proteins in the network gives the infection capability of the infected protein. Thus the spreader nodes identified by spreadability index are validated by the infection rate as generated by SIS for them. It can be observed from Table 1 to Table 5 that the proposed methodology has the highest SIS infection rate of 2.46 (see Table 1) in comparison to others for their corresponding top 10 spreader nodes in the synthetic network as shown in Figure 1.

2.4 Ranking of Spreader edges

To show the ranking of interacting spreader edges, another synthetic PPIN 2 has been considered as network 1 while the previous synthetic PPIN 1 in section 2.2 (Figure 1) has been represented as network 2 in Figure 2. Node D, E and F are the selected top spreader nodes in network 1 by spreadability index in the same way as that of synthetic PPIN 1 in Figure 1 while to avoid the complexity in the diagram, top 5 nodes in network 2 (see Table 1) are selected as spreader nodes in network 2. Red colored edges are the interconnectivities between the nodes of network 1 while

black colored edges show the interconnectivity between nodes of network 2. Green colored spreader edges (i.e. edges connected with spreader nodes) show the interconnectivity between network 1 and 2. Ranking of a spreader edge measures the effectiveness of transmission ability of a spreader edge i.e. how many interactive nodes gets infected through that edge. Thus all the spreading edges are ranked based on the average of the spreadability index of its connected spreader nodes. The ranked spreader edges in Figure 2 are highlighted in Table 6.

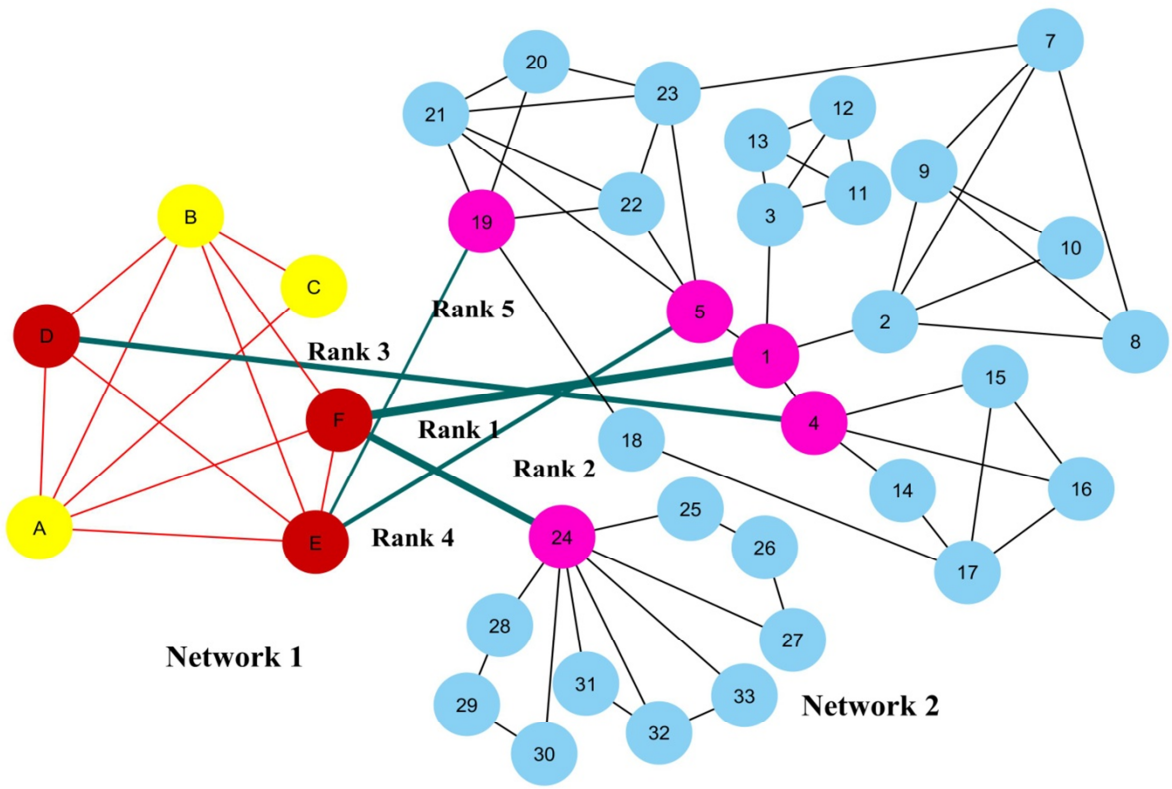

Figure 2 Ranking of spreader edges between two interconnected PPIN based on spreadability index. Thickness of the edges vary with the order of ranking.

Table 6 Ranking of spreader edges for network 1 and network 2 in Figure 2

	Spreader Edges				
Rank	**Spreader nodes in network 1**	**Spreader nodes in network 2**	**Spreadability Index of spreader nodes in network 1**	**Spreadability Index of spreader nodes in network 2**	**Ranking of spreader edges**
1	F	1	5.5	76.15	40.825
2	F	24	5.5	22.70	14.104
3	D	4	5.5	15.11	10.308
4	E	5	4.7	13.48	9.0919
5	E	19	4.7	11.66	8.1833

3. Experimental Result

The proposed methodology leads to the identification of spreader nodes and edges through a network characteristic, called spreader index which has been also checked and validated by SIS model. Initially the whole working module is implemented on synthetic networks as shown in Figure 1 and Figure 2 in section 2 and then on SARS-CoV-human dataset. To visualize the infection transmission from SARS-CoV to human, PPIN of the both needs to be formed for which three datasets have been curated. They are: 1) SARS-CoV PPIN [25] 2) SARS-CoV-Human PPIN [25] 3) Human PPIN [26], [27]. After the removal of self-loops and data redundancy, the final SARS-CoV PPIN consists of 17 interactions among SARS-CoV proteins along with the involvement of 7 unique proteins after the removal of self loops and data redundancy. Only the densely interconnected SARS-CoV proteins having direct connections (level 1) with human proteins are considered rather than the isolated ones. SARS-CoV-Human PPIN includes 118 interactions between SARS-CoV and human. It is used to fetch the level 1 interaction of human proteins for the corresponding SARS-CoV proteins in SARS-CoV PPIN. Human PPIN consists of 314384 interactions. It is utilized for getting the indirect interactions (level 2) of level 1 human proteins formed earlier. The application of the proposed methodology in SARS-CoV-human PPIN is highlighted in Figure 3.

In Figure 3A, at first SARS-CoV PPIN is displayed in which each protein is marked in red. Thereafter spreader nodes in SARS-CoV PPIN are identified by spreadability index which are denoted as blue nodes among red. Once the spreader nodes are active (in Figure 3B), it transmits the infection to its corresponding direct partners i.e. human level 1proteins (marked in green). In Figure 3C, spreader nodes are identified in SARS-CoV level 1 human proteins (marked in blue) and the disease will be transmitted further. The same will continue till SARS-CoV level 2 human proteins in which green nodes are the spreaders and thus the infection will penetrate further in human PPIN resulting in a significant fall in human immunity level followed by severe acute respiratory syndrome.

In Figure 4, the PPIN of SARS-CoV network has been highlighted. There are mainly 9 proteins which include E, M, ORF3A, ORF7A, S, N, ORF8A, ORF8AB, and ORF8B. The computed spreadability index of each of this protein and its corresponding validation by SIS model is highlighted in Table 7. It is also compared with other central/ influential spreader node detection methodologies like DC, CC, LAC and BC which are shown in Table 8, Table 9, Table 10 and Table 11 respectively. Similarly spreader nodes are also identified in SARS-CoV's level 1 neighbors and level 2 neighbors (see Figure 5 and Figure 6).

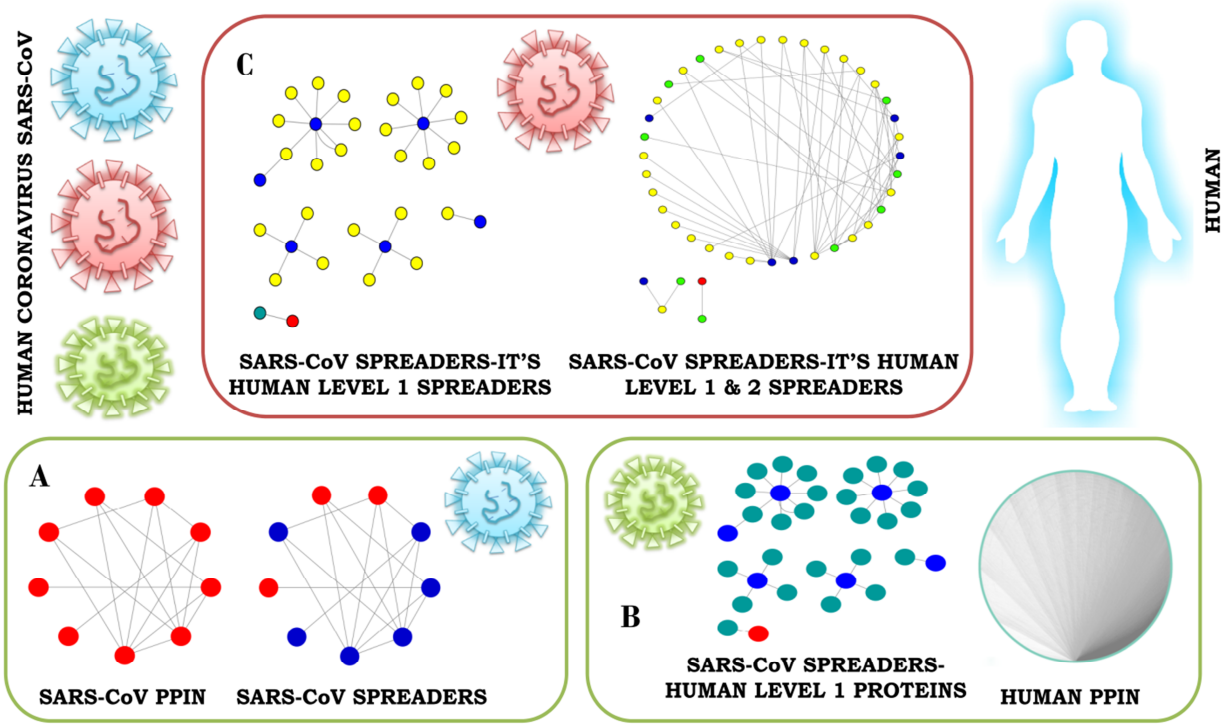

Figure 3. Mechanism of transmission of infection through spreader nodes. **A**. PPIN of SARS-CoV (red). Selection of spreader nodes in SARA-CoV PPIN (blue). **B.** Interaction of SARS-CoV spreaders with its level 1 corresponding proteins in human PPIN (green). **C.** Selection of spreaders in level 1 (yellow) and level 2 human proteins (green). Rest proteins in human PPIN are ignored to prevent overlap in the diagram.

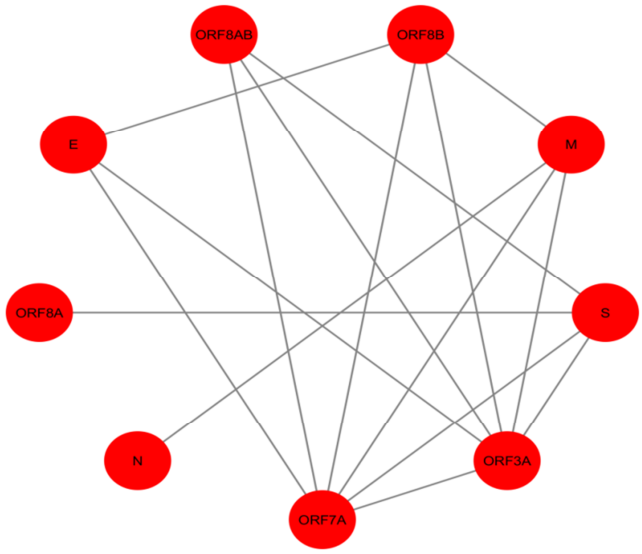

Figure 4. SARS-CoV PPIN consisting of 9 proteins which each protein is represented by a node and its corresponding interaction is highlighted by edges.

Table 7 Computation of spreadability index of SARS-CoV PPIN along with computation of infection rate of selected top 6 spreader nodes by SIS model.

Rank	Proteins	$E_{out}^{S_i}$	$E_{in}^{S_i}$	Edge Ratio	Neighborhood Density	Node Weight	Spreadability Index	SIS infection rate of top 6 nodes	Sum of SIS infection rate of top 6 nodes
1	M	7	3	2.0	3.845	3.4	11.090	1	2.935
2	S	6	3	1.75	4.047	3.2	10.283	0.2	
3	ORF8AB	7	3	2.0	1.785	4.0	7.5714	1	
4	ORF8B	5	5	1.0	3.464	3.8	7.2642	0.2	
5	E	7	3	2.0	1.428	4.0	6.8571	0.25	
6	ORF3A	2	8	0.333	9.249	3.428	6.5119	0.285	
7	ORF7A	2	8	0.333	9.25	3.428	6.5119	– –	– –
8	ORF8A	3	0	4.0	0.0	2.0	2		
9	N	3	0	4.0	0.0	2.0	2		

Table 8 Computation of degree centrality of SARS-CoV PPIN along with computation of infection rate of selected top 6 spreader nodes by SIS model.

Rank	Proteins	Degree Centrality [19]	SIS infection rate of top 6 nodes	Sum of SIS infection rate of top 6 nodes
1	ORF7A	6	0.285	1.82
2	ORF3A	6	0.285	
3	ORF8B	4	0.2	
4	M	4	0.6	
5	S	4	0.2	
6	E	3	0.25	
7	ORF8AB	3	– –	– –
8	N	1		
9	ORF8A	1		

Table 9 Computation of closeness centrality of SARS-CoV PPIN along with computation of infection rate of selected top 6 spreader nodes by SIS model.

Rank	Proteins	Closeness Centrality [24]	SIS infection rate of top 6 nodes	Sum of SIS infection rate of top 6 nodes
1	ORF7A	0.239	0.285	1.82
2	ORF3A	0.239	0.285	
3	ORF8B	0.224	0.2	
4	M	0.224	0.6	
5	S	0.224	0.2	
6	E	0.215	0.25	
7	ORF8AB	0.22		

8	N	0.196	– –	– –
9	ORF8A	0.196		

Table 10 Computation of local average centrality of SARS-CoV PPIN along with computation of infection rate of selected top 6 spreader nodes by SIS model.

Rank	**Proteins**	**Local Average Centrality** [21]	**SIS infection rate of top 6 nodes**	**Sum of SIS infection rate of top 6 nodes**
1	ORF7A	2.666	0.285	2.22
2	ORF3A	2.666	0.285	
3	ORF8B	2.5	0.2	
4	E	2	0.25	
5	ORF8AB	2	1	
6	S	1.5	0.2	
7	M	1.5	– –	– –
8	N	0		
9	ORF8A	0		

Table 11 Computation of betweeness centrality of SARS-CoV PPIN along with computation of infection rate of selected top 6 spreader nodes by SIS model.

Rank	**Proteins**	**Betweeness Centrality** [23]	**SIS infection rate of top 6 nodes**	**Sum of SIS infection rate of top 6 nodes**
1	M	14	0.6	1.82
2	S	14	0.2	
3	ORF7A	13.33	0.285	
4	ORF3A	13.33	0.285	
5	ORF8B	1.33	0.2	
6	E	0	0.25	
7	ORF8AB	0	– –	– –
8	N	0		
9	ORF8A	0		

Spreadability index definitely plays an important role in this proposed methodology. In fact, spreader nodes are successfully identified by the virtue of this scoring technique which covers all the aspects of transmitting infection from one node to another in a PPIN. It should be mentioned here that while identifying spreader nodes in SARS-CoV level 2 human proteins, it has been noted that the number of nodes are getting increased largely with the increment of successive levels. So, high, medium and low threshold [33] have been applied and the entire disease transmission through spreadability index is computationally assessed at each threshold.

Table 12 Network statistics of spreaders at three levels of thresholds

Threshold	SARS-CoV spreaders	SARS-CoV-s level 1 human spreaders	SARS-CoV-s level 2 human spreaders
High	6	24	9
Medium	6	24	22
Low	6	24	111

The network statistics of spreader nodes at each level of threshold is shown in Table 12. It can be observed that threshold application is only implemented at SARS-CoV level 2 human proteins not on others. This is because of the availability of very less number of nodes and edges. Only nodes and edges having extremely low spreadability index have been discarded at the first level.

Beside identification of spreader nodes, spreader edges are also identified. The ranked edges between SARS-CoV spreaders and its level 1 human spreaders is highlighted in Table 13 while the ranked edges between SARS-CoV s level 1 and level 2 human spreaders at high, medium and low threshold are highlighted in A-Table 1, A-Table 2 and A-Table 3 in the appendix respectively.

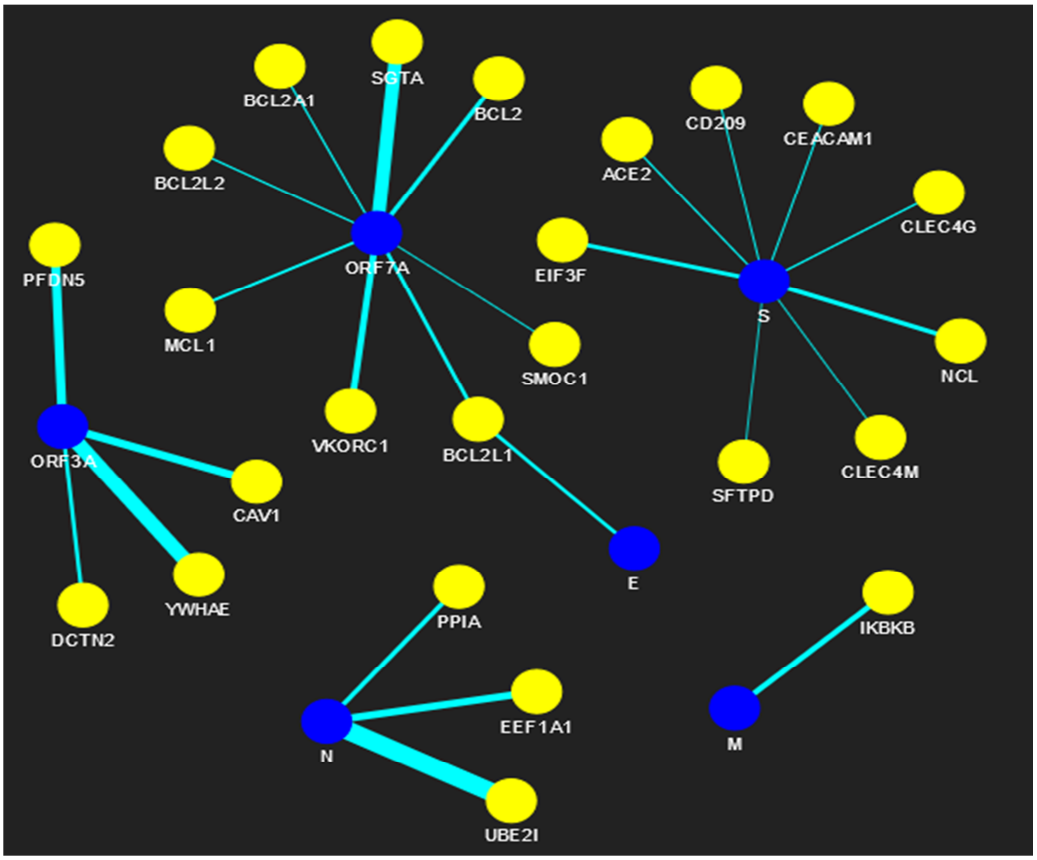

Figure 5 SARS-CoV-human PPIN (blue node represents SARS-CoV proteins while yellow node represents SARS-CoV s level 1 human spreaders)

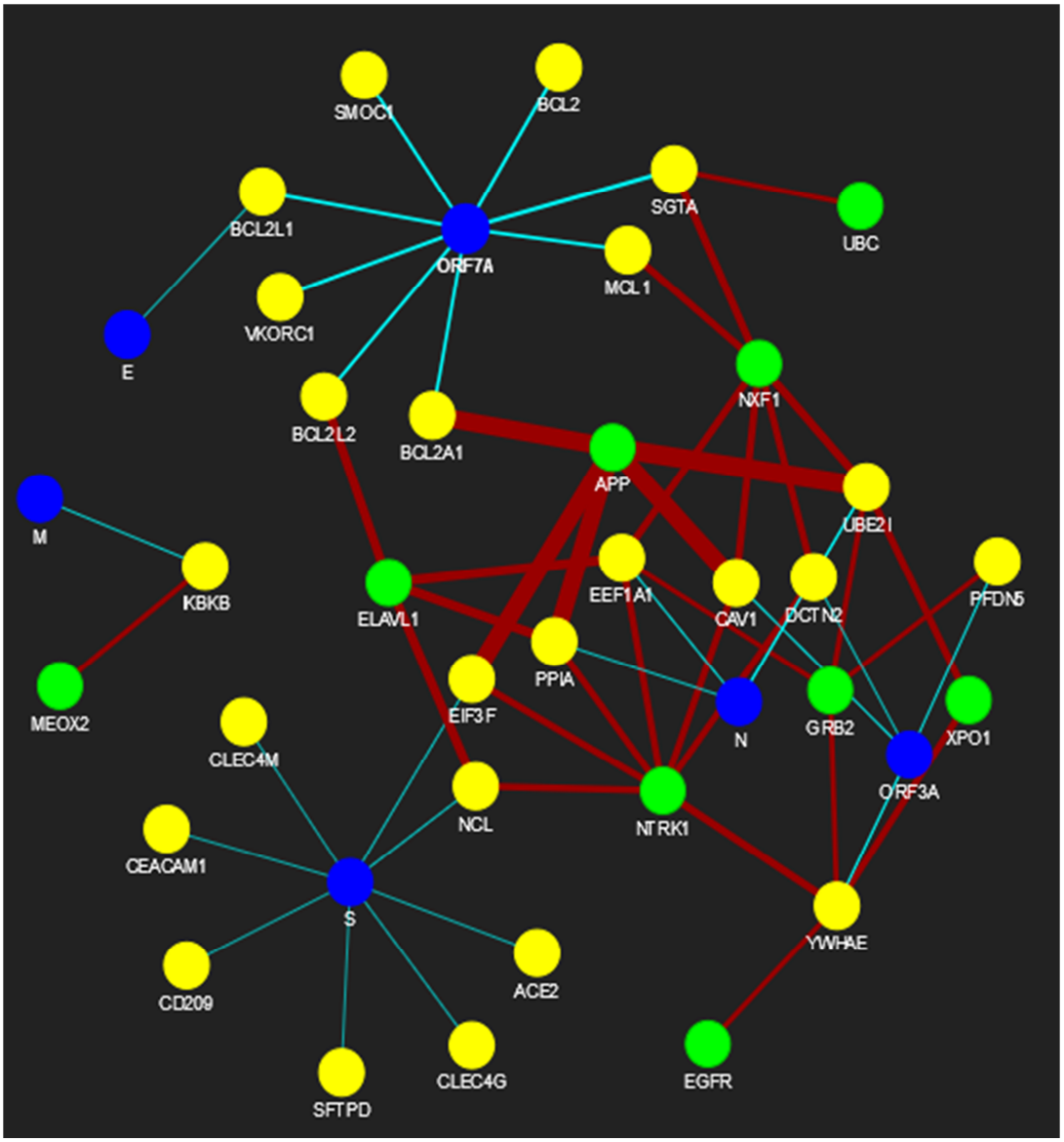

Figure 6 SARS-CoV-human PPIN (blue node represents SARS-CoV proteins while yellow and green nodes represent SARS-CoV s level 1 and level 2 human spreaders respectively)

Table 13 Ranked spreader edges between SARS-CoV spreaders and its level 1 human spreaders

	Spreader Edges		
Rank	**SARS-CoV Spreaders**	**SARS-CoV s level 1 human spreaders**	**Spreading ability of spreader edges**
1	N	UBE2I	679697.677
2	ORF3A	YWHAE	500684.2755
3	ORF7A	SGTA	428397.3206

4	ORF3A	PFDN5	273863.194
5	ORF3A	CAV1	264566.0653
6	N	EEF1A1	241407.2776
7	ORF7A	VKORC1	187916.2768
8	M	IKBKB	164728.3002
9	S	NCL	131643.7345
10	N	PPIA	125719.6427
11	S	EIF3F	119529.0273
12	ORF7A	BCL2	119299.092
13	ORF3A	DCTN2	92293.0019
14	E	BCL2L1	89404.47117
15	ORF7A	BCL2L1	89404.29855
16	ORF7A	MCL1	63953.80825
17	S	CLEC4G	27477.4133
18	ORF7A	BCL2L2	22974.97399
19	ORF7A	BCL2A1	22252.28441
20	S	ACE2	18775.88601
21	S	CEACAM1	14834.82402
22	S	CD209	12215.99362
23	ORF7A	SMOC1	6068.990602
24	S	CLEC4M	3844.528751
25	S	SFTPD	119.09278

The entire network view of SARS-CoV and human PPIN has been generated online under three circumstances:

1) All the nodes and edges are considered as spreader nodes and edges respectively and are ranked accordingly.

https://yu2qkp7gwoinjwsebyw0xw-on.drv.tw/www.graph_all.html/graph_all.html

2) Selected Spreader nodes and edges are highlighted for high threshold.

https://yu2qkp7gwoinjwsebyw0xw-on.drv.tw/www.high_threshold.com/graph_high_threshold.html

3) Selected Spreader nodes and edges are highlighted for medium threshold.

https://yu2qkp7gwoinjwsebyw0xw-on.drv.tw/www.medium_threshold.com/graph_medium_threshold.html

4) Selected Spreader nodes and edges are highlighted for low threshold.

https://yu2qkp7gwoinjwsebyw0xw-on.drv.tw/www.low_threshold.com/graph_low_threshold.html

In the above generated network views, the blue, yellow and green color represent SARS-CoV spreaders, its level 1 human spreaders and its level2 human spreaders respectively. The remaining nodes are in indigo.

4. Conclusion

Spreadability index is thus proved to be effective in the detection of spreader nodes and edges in SARS-CoV-human PPIN along with the cross validation by SIS model. Spreader nodes are the most critical nodes in the network which actually transmit the infection to its successors. Simultaneously it is also true if the spreader nodes are not connected with spreader edges infection transmission would not have been possible.

In a nutshell, it can be said that the proposed work exploits the possibility of understanding the entire disease propagation/transmission from SARS-CoV network to human network. It should be borne in mind that SARS-CoV2 is ~89% genetically similar to its predecessor SARS-CoV [34], [35]. It strongly reveals the fact that the human proteins chosen as spreaders of SARS-CoV might be the potential targets of SARS-CoV 2 also. If our future research work reveals this assumption, then it will definitely explore a new direction in the identification of essential drugs/vaccine for SARS-CoV2.

Acknowledgement

Authors are thankful to the CMATER research laboratory of the Computer Science Department, Jadavpur University, India, for providing infrastructure facilities during progress of the work. This project is partially supported by the CMATER research laboratory of the Computer Science and Engineering Department, Jadavpur University, India, and DBT project (No.BT/PR16356/BID/7/596/2016), Ministry of Science and Technology, Government of India.

References

[1] "World-Health-Organization Coronavirus disease (COVID-19) outbreak." https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (accessed Apr. 02, 2020).

[2] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health concern," *The Lancet*, vol. 395, no. 10223. Lancet Publishing Group, pp. 470–473, Feb. 15, 2020, doi: 10.1016/S0140-6736(20)30185-9.

[3] "World Map | CDC." https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/world-map.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Flocations-confirmed-cases.html (accessed Apr. 02, 2020).

[4] "Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)." https://www.who.int/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov) (accessed Apr. 02, 2020).

[5] "Statement on the meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus 2019 (n-CoV) on 23 January 2020." https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov) (accessed Apr. 02, 2020).

[6] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.

[7] D. L. Heymann, "Data sharing and outbreaks: best practice exemplified," *The Lancet*, vol. 395, no. 10223. Lancet Publishing Group, pp. 469–470, Feb. 15, 2020, doi: 10.1016/S0140-6736(20)30184-7.

[8] X. Liu and X.-J. Wang, "Potential inhibitors for 2019-nCoV coronavirus M protease from clinically approved medicines," *bioRxiv*, p. 2020.01.29.924100, Jan. 2020, doi: 10.1101/2020.01.29.924100.

[9] P. Zhou *et al.*, "A pneumonia outbreak associated with a new coronavirus of probable bat origin," *Nature*, vol. 579, no. 7798, pp. 270–273, Mar. 2020, doi: 10.1038/s41586-020-2012-7.

[10] "WHO | Update 49 - SARS case fatality ratio, incubation period." https://www.who.int/csr/sars/archive/2003_05_07a/en/ (accessed Apr. 06, 2020).

[11] "WHO | Middle East respiratory syndrome coronavirus (MERS-CoV)." https://www.who.int/emergencies/mers-cov/en/ (accessed Apr. 06, 2020).

[12] R. Lu *et al.*, "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding," *Lancet*, vol. 395, no. 10224, pp. 565–574, Feb. 2020, doi: 10.1016/S0140-6736(20)30251-8.

[13] M. Letko, A. Marzi, and V. Munster, "Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses," *Nat. Microbiol.*, vol. 5, no. 4, pp. 562–569, Feb. 2020, doi: 10.1038/s41564-020-0688-y.

[14] M. Hoffmann, H. Kleine-Weber, N. Krüger, M. Müller, C. Drosten, and S. Pöhlmann, "The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells," *bioRxiv*, p. 2020.01.31.929042, Jan. 2020, doi: 10.1101/2020.01.31.929042.

[15] "Naming the coronavirus disease (COVID-19) and the virus that causes it." https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it (accessed Apr. 06, 2020).

[16] S. Saha, A. Prasad, P. Chatterjee, S. Basu, and M. Nasipuri, "Protein function prediction from dynamic protein interaction network using gene expression data," *J. Bioinform. Comput. Biol.*, vol. 17, no. 4, Aug. 2019, doi: 10.1142/S0219720019500252.

[17] S. Saha, A. Prasad, P. Chatterjee, S. Basu, and M. Nasipuri, "Protein function prediction from protein-protein interaction network using gene ontology based neighborhood analysis and physico-chemical features," in *Journal of Bioinformatics and Computational Biology*, Dec. 2018, vol. 16, no. 6, doi: 10.1142/S0219720018500257.

[18] S. Saha, P. Chatterjee, S. Basu, M. Nasipuri, and D. Plewczynski, "Funpred 3.0: Improved protein function prediction using protein interaction network," *PeerJ*, vol. 2019, no. 5, p. e6830, May 2019, doi: 10.7717/peerj.6830.

[19] H. Jeong, S. P. Mason, A. L. Barabási, and Z. N. Oltvai, "Lethality and centrality in protein networks," *Nature*, vol. 411, no. 6833, pp. 41–42, May 2001, doi: 10.1038/35075138.

[20] S. Wuchty and P. F. Stadler, "Centers of complex networks," *J. Theor. Biol.*, vol. 223, no. 1, pp. 45–53, Jul. 2003, doi: 10.1016/S0022-5193(03)00071-7.

[21] M. Li, J. Wang, X. Chen, H. Wang, and Y. Pan, "A local average connectivity-based method for identifying essential proteins from the network level," *Comput. Biol. Chem.*, vol. 35, no. 3, pp. 143–150, Jun. 2011, doi: 10.1016/j.compbiolchem.2011.04.002.

[22] M. P. Joy, A. Brock, D. E. Ingber, and S. Huang, "High-betweenness proteins in the yeast protein interaction network," *J. Biomed. Biotechnol.*, vol. 2005, no. 2, pp. 96–103, Jun. 2005, doi: 10.1155/JBB.2005.96.

[23] J. M. Anthonisse, *The rush in a directed graph*, no. BN 9/71. Stichting Mathematisch Centrum, 1971.

[24] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966, doi: 10.1007/BF02289527.

[25] S. Pfefferle *et al.*, "The SARS-Coronavirus-host interactome: Identification of cyclophilins as target for pan-Coronavirus inhibitors," *PLoS Pathog.*, vol. 7, no. 10, 2011, doi: 10.1371/journal.ppat.1002331.

[26] "BioSNAP: Network datasets: Human protein-protein interaction network." https://snap.stanford.edu/biodata/datasets/10000/10000-PP-Pathways.html (accessed Apr. 05, 2020).

[27] M. Agrawal, M. Zitnik, and J. Leskovec, "Large-Scale Analysis of Disease Pathways in the Human Interactome," *bioRxiv*, p. 189787, Sep. 2017, doi: 10.1101/189787.

[28] N. Samadi and A. Bouyer, "Identifying influential spreaders based on edge ratio and neighborhood diversity measures in complex networks," *Computing*, vol. 101, no. 8, pp. 1147–1175, 2019, doi: 10.1007/s00607-018-0659-9.

[29] S. Wang and F. Wu, "Detecting overlapping protein complexes in PPI networks based on robustness," *Proteome Sci.*, vol. 11, no. Suppl 1, pp. 1–8, 2013, doi: 10.1186/1477-5956-11-S1-S18.

[30] N. T. J. Bailey, "The mathematical theory of infectious diseases and its applications. 2nd edition.," *Math. theory Infect. Dis. its Appl. 2nd Ed.*, 1975.

[31] S. Saha, K. Sengupta, P. Chatterjee, S. Basu, and M. Nasipuri, "Analysis of protein targets in pathogen–host interaction in infectious diseases: a case study on Plasmodium falciparum and Homo sapiens interaction network," *Brief. Funct. Genomics*, vol. 17, no. 6, pp. 441–450, Sep. 2017, doi: 10.1093/bfgp/elx024.

[32] P. Jaccard, "THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.," *New Phytol.*, vol. 11, no. 2, pp. 37–50, Feb. 1912, doi: 10.1111/j.1469-8137.1912.tb05611.x.

[33] Y. Zhang, H. Lin, Z. Yang, J. Wang, Y. Liu, and S. Sang, "A method for predicting protein complex in dynamic PPI networks," *BMC Bioinformatics*, vol. 17, no. S7, p. 229, Jul. 2016, doi: 10.1186/s12859-016-1101-y.

[34] J. F. W. Chan *et al.*, "Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan," *Emerg. Microbes Infect.*, vol. 9, no. 1, pp. 221–236, Jan. 2020, doi: 10.1080/22221751.2020.1719902.

[35] "China releases genetic data on new coronavirus, now deadly | CIDRAP." https://www.cidrap.umn.edu/news-perspective/2020/01/china-releases-genetic-data-new-coronavirus-now-deadly (accessed Apr. 10, 2020).

Appendix

A-Table 1 Ranked spreader edges between SARS-CoV s level 1 and level 2 human spreaders at high threshold

	Spreader Edges		
Rank	**SARS-CoV s level 1 and level 2 human spreaders**		**Spreading ability of spreader edges**
1	APP	UBE2I	9297903
2	APP	CAV1	8882769
3	APP	PPIA	8743925
4	APP	EIF3F	8737730
5	BCL2A1	APP	8640455
6	ELAVL1	EEF1A1	4284371
7	NCL	ELAVL1	4174604
8	PPIA	ELAVL1	4168684
9	BCL2L2	ELAVL1	4065937
10	NXF1	UBE2I	3653905
11	NXF1	SGTA	3402603
12	NTRK1	YWHAE	3275095
13	NXF1	CAV1	3238771
14	NXF1	EEF1A1	3215615
15	XPO1	UBE2I	3169805
16	DCTN2	NXF1	3066498
17	NTRK1	CAV1	3038977
18	NXF1	MCL1	3038159
19	NTRK1	EEF1A1	3015820
20	XPO1	YWHAE	2990790
21	NCL	NTRK1	2906052
22	NTRK1	PPIA	2900133
23	NTRK1	EIF3F	2893938
24	DCTN2	NTRK1	2866704
25	GRB2	UBE2I	2349446
26	GRB2	YWHAE	2170431
27	IKBKB	MEOX2	2061694
28	UBC	SGTA	2048943
29	YWHAE	EGFR	2023580
30	GRB2	PFDN5	1943610
31	GRB2	EEF1A1	1911156

A-Table 2 Ranked spreader edges between SARS-CoV s level 1 and level 2 human proteins at medium threshold

	Spreader Edges		
Rank	**SARS-CoV s level 1 and level 2 human spreaders**		**Spreading ability of spreader edges**
1	APP	UBE2I	9297902.664
2	APP	CAV1	8882768.796
3	APP	PPIA	8743924.63
4	APP	EIF3F	8737729.873
5	BCL2A1	APP	8640455.016
6	ELAVL1	EEF1A1	4284371.281
7	NCL	ELAVL1	4174603.596
8	PPIA	ELAVL1	4168683.646
9	BCL2L2	ELAVL1	4065936.721
10	NXF1	UBE2I	3653905.316
11	NXF1	SGTA	3402602.703
12	NTRK1	YWHAE	3275094.903
13	NXF1	CAV1	3238771.448
14	NXF1	EEF1A1	3215614.916
15	XPO1	UBE2I	3169805.311
16	DCTN2	NXF1	3066498.385
17	NTRK1	CAV1	3038976.693
18	NXF1	MCL1	3038159.191
19	NTRK1	EEF1A1	3015820.161
20	XPO1	YWHAE	2990789.654
21	NCL	NTRK1	2906052.476
22	NTRK1	PPIA	2900132.526
23	NTRK1	EIF3F	2893937.769
24	DCTN2	NTRK1	2866703.629
25	GRB2	UBE2I	2349446.25
26	GRB2	YWHAE	2170430.592
27	IKBKB	MEOX2	2061694.24
28	UBC	SGTA	2048942.842
29	YWHAE	EGFR	2023579.783
30	GRB2	PFDN5	1943609.51
31	GRB2	EEF1A1	1911155.85
32	UBC	EEF1A1	1861955.055
33	TP53	UBE2I	1829375.898
34	BAG3	UBE2I	1809920.855
35	NCL	GRB2	1801388.165
36	GRB2	PPIA	1795468.215

37	CAV1	EGFR	1787461.572
38	GOLGA2	UBE2I	1779690.372
39	EEF1A1	EGFR	1764305.041
40	DCTN2	GRB2	1762039.318
41	NCL	UBC	1752187.37
42	EIF3F	UBC	1740072.663
43	YWHAE	SHMT2	1738748.567
44	DCTN2	UBC	1712838.523
45	BCL2A1	GRB2	1691998.601
46	MCL1	UBC	1684499.329
47	TP53	YWHAE	1650360.24
48	EIF3F	EGFR	1642422.649
49	PROP1	UBE2I	1641256.177
50	BAG3	YWHAE	1630905.197
51	CDK2	UBE2I	1622979.319
52	DCTN2	EGFR	1615188.509
53	CUL3	YWHAE	1614438.165
54	TRIM27	UBE2I	1579133.993
55	CEACAM1	EGFR	1537728.445
56	HSP90AA1	YWHAE	1503676.376
57	TRAF6	UBE2I	1451472.666
58	CDK2	YWHAE	1443963.661
59	MCM2	YWHAE	1440253.354
60	BRCA1	UBE2I	1439839.399
61	HSP90AA1	SGTA	1431389.421
62	ESR1	UBE2I	1424801.897

A-Table 3 Ranked spreader edges between SARS-CoV s level 1 and level 2 human spreaders at low threshold

	Spreader Edges		
Rank	**SARS-CoV s level 1 and level 2 human spreaders**		**Spreading ability of spreader edges**
1	APP	UBE2I	9297902.664
2	APP	CAV1	8882768.796
3	APP	PPIA	8743924.63
4	APP	EIF3F	8737729.873
5	BCL2A1	APP	8640455.016
6	ELAVL1	EEF1A1	4284371.281
7	NCL	ELAVL1	4174603.596
8	PPIA	ELAVL1	4168683.646

9	BCL2L2	ELAVL1	4065936.721
10	NXF1	UBE2I	3653905.316
11	NXF1	SGTA	3402602.703
12	NTRK1	YWHAE	3275094.903
13	NXF1	CAV1	3238771.448
14	NXF1	EEF1A1	3215614.916
15	XPO1	UBE2I	3169805.311
16	DCTN2	NXF1	3066498.385
17	NTRK1	CAV1	3038976.693
18	NXF1	MCL1	3038159.191
19	NTRK1	EEF1A1	3015820.161
20	XPO1	YWHAE	2990789.654
21	NCL	NTRK1	2906052.476
22	NTRK1	PPIA	2900132.526
23	NTRK1	EIF3F	2893937.769
24	DCTN2	NTRK1	2866703.629
25	GRB2	UBE2I	2349446.25
26	GRB2	YWHAE	2170430.592
27	IKBKB	MEOX2	2061694.24
28	UBC	SGTA	2048942.842
29	YWHAE	EGFR	2023579.783
30	GRB2	PFDN5	1943609.51
31	GRB2	EEF1A1	1911155.85
32	UBC	EEF1A1	1861955.055
33	TP53	UBE2I	1829375.898
34	BAG3	UBE2I	1809920.855
35	NCL	GRB2	1801388.165
36	GRB2	PPIA	1795468.215
37	CAV1	EGFR	1787461.572
38	GOLGA2	UBE2I	1779690.372
39	EEF1A1	EGFR	1764305.041
40	DCTN2	GRB2	1762039.318
41	NCL	UBC	1752187.37
42	EIF3F	UBC	1740072.663
43	YWHAE	SHMT2	1738748.567
44	DCTN2	UBC	1712838.523
45	BCL2A1	GRB2	1691998.601
46	MCL1	UBC	1684499.329
47	TP53	YWHAE	1650360.24
48	EIF3F	EGFR	1642422.649
49	PROP1	UBE2I	1641256.177

50	BAG3	YWHAE	1630905.197
51	CDK2	UBE2I	1622979.319
52	DCTN2	EGFR	1615188.509
53	CUL3	YWHAE	1614438.165
54	TRIM27	UBE2I	1579133.993
55	CEACAM1	EGFR	1537728.445
56	HSP90AA1	YWHAE	1503676.376
57	TRAF6	UBE2I	1451472.666
58	CDK2	YWHAE	1443963.661
59	MCM2	YWHAE	1440253.354
60	BRCA1	UBE2I	1439839.399
61	HSP90AA1	SGTA	1431389.421
62	ESR1	UBE2I	1424801.897
63	COPS5	UBE2I	1402029.245
64	TP53	EEF1A1	1391085.498
65	CAV1	MOV10	1381399.319
66	ZDHHC17	EEF1A1	1380592.242
67	UBE2I	ATXN1	1362088.512
68	MOV10	EEF1A1	1358242.787
69	TMEM17	CAV1	1356149.641
70	CUL3	EEF1A1	1355163.423
71	HSPB1	UBE2I	1341884.753
72	MYC	YWHAE	1341253.761
73	LNX1	PFDN5	1333010.914
74	DCTN2	SHMT2	1330357.294
75	TP53	IKBKB	1314401.976
76	YWHAE	FBXO6	1313852.619
77	HGS	UBE2I	1294714.22
78	TP53	NCL	1281317.814
79	CUL3	IKBKB	1278479.9
80	YWHAE	VCP	1277472.997
81	TRAF2	YWHAE	1274995.25
82	TP53	BCL2	1268975.057
83	HSP90AA1	CAV1	1267558.165
84	BAG3	NCL	1261862.77
85	MDM2	UBE2I	1260258.833
86	YWHAZ	YWHAE	1255600.341
87	BAG3	BCL2	1249520.013
88	NCL	MOV10	1248475.102
89	CUL3	NCL	1245395.738
90	HSP90AA1	EEF1A1	1244401.634

91	CUL3	PPIA	1239475.788
92	TP53	BCL2L1	1239080.263
93	EWSR1	PPIA	1229282.205
94	FN1	YWHAE	1214083.458
95	TP53	MCL1	1213629.773
96	EP300	UBE2I	1213483.256
97	CUL3	DCTN2	1206046.891
98	CDK1	UBE2I	1205680.466
99	UBE2I	AGTRAP	1198756.302
100	HDAC1	UBE2I	1194236.516
101	BAG3	MCL1	1194174.73
102	IKBKE	UBE2I	1186760.92
103	LNX1	PPIA	1184869.618
104	CDK2	EEF1A1	1184688.919
105	YWHAE	ATXN1	1183072.854
106	MCM2	EEF1A1	1180978.612
107	PFDN5	SDCBP	1178978.137
108	TCF4	UBE2I	1174611.299
109	TP53	BCL2L2	1172650.939
110	HSP90AA1	IKBKB	1167718.111
111	CSNK2B	UBE2I	1166588.488
112	HSPB1	YWHAE	1162869.095
113	JUN	UBE2I	1162163.465
114	KDM1A	UBE2I	1155214.334
115	CDC37	UBE2I	1151160.054
116	TRIM23	UBE2I	1145591.625
117	HSP90AA1	NCL	1134633.949
118	HSP90AA1	BCL2	1122291.192
119	BMI1	UBE2I	1120564.971
120	CCDC8	YWHAE	1118799.628
121	CSNK2A1	CAV1	1117578.493
122	MYC	PFDN5	1114432.68
123	HSF2BP	UBE2I	1113581.235
124	FBXW11	YWHAE	1086642.037
125	SRC	YWHAE	1085833.441
126	NCL	CDK2	1074921.234
127	NCL	MCM2	1071210.927
128	PPIA	CDK2	1069001.284
129	CALM1	YWHAE	1067438.977
130	MCM2	PPIA	1065290.977
131	IKBKB	TRIM27	1064160.071

132	BCL2	CDK2	1062578.478
133	MCM2	EIF3F	1059096.22
134	TERF2	UBE2I	1055418.427
135	FBXO6	EEF1A1	1054577.877
136	UNC119	UBE2I	1052917.816
137	CAV1	VCP	1041354.787
138	BAG6	SGTA	1039029.262
139	TRAF2	CAV1	1038877.04
140	YWHAB	YWHAE	1038503.438
141	TRAF6	CAV1	1036338.798
142	CFTR	UBE2I	1035883.116
143	YWHAQ	YWHAE	1034131.766
144	AURKA	UBE2I	1032671.376
145	CLTC	UBE2I	1030687.678
146	KAT5	UBE2I	1015856.627
147	UBE2I	TSNAX	1010628.678
148	CUL7	YWHAE	1010074.36
149	ESR1	CAV1	1009668.029
150	WNK1	UBE2I	1009409.677
151	MCL1	CDK2	1007233.194
152	CREBBP	UBE2I	1001371.822
153	HDAC5	UBE2I	999897.9437
154	FOS	UBE2I	998888.38
155	UBE2I	SIAH1	998794.2497
156	WDYHV1	PFDN5	998338.5732
157	YWHAZ	EEF1A1	996325.5986
158	UBE2I	TSG101	994223.1527
159	HNRNPA1	UBE2I	994154.0026
160	UBE2I	LMNA	993727.7467
161	BTRC	YWHAE	993713.5665
162	SH3KBP1	UBE2I	992752.976
163	TRIP13	UBE2I	992337.1069
164	NCL	CSNK2A1	984654.2768
165	YWHAG	YWHAE	983125.5399
166	CUL1	YWHAE	983107.805
167	UBE2I	SUMO2	982210.8384
168	PTEN	UBE2I	981258.9481
169	TRIM28	UBE2I	979151.0849
170	SLX4	UBE2I	976097.5171
171	APC	YWHAE	975907.5724
172	UNK	YWHAE	973611.9526

173	CSNK2A1	EIF3F	972539.5696
174	MYC	NCL	972211.3346
175	CDC37	YWHAE	972144.3969
176	MYH9	UBE2I	970752.1564
177	HLA-B	SGTA	968992.1249
178	CAV1	PRKACA	968598.7454
179	AR	UBE2I	967692.3425
180	ATF2	UBE2I	966712.2472
181	YWHAB	SGTA	966216.483
182	SUMO1	UBE2I	965376.1591
183	COPS5	EEF1A1	963738.8454
184	AKT1	YWHAE	963671.5308
185	ABL1	YWHAE	962733.2864
186	MYC	BCL2	959868.5778
187	UBE2I	SMAD4	959192.5665
188	OBSL1	YWHAE	957100.0055
189	FN1	EEF1A1	954808.7158
190	FHL3	UBE2I	954696.496
191	DAXX	UBE2I	953684.4234
192	HDAC4	UBE2I	950274.9524
193	FUS	UBE2I	949535.1125
194	SUZ12	UBE2I	948330.2296
195	TRIP6	SGTA	946975.115
196	PRKDC	UBE2I	946021.6212
197	UBE2I	UBE2D2	945725.2466
198	STAT1	UBE2I	944739.8548
199	UBE2I	TRIM54	944513.0565
200	SIRT1	UBE2I	943590.5083
201	HSPA8	YWHAE	943349.4099
202	PRKAA2	UBE2I	939213.0138
203	TRAF2	IKBKB	939036.9856