

Finding Prediction of Interaction between SARS-CoV-2 and Human Protein: A Data Driven Approach

Moumita Ghosh

Jadavpur University <https://orcid.org/0000-0002-8640-3745>

Pritam Sil

Jadavpur University <https://orcid.org/0000-0002-0992-6892>

Anirban Roy

West Bengal Biodiversity Board <https://orcid.org/0000-0003-3786-7380>

Rohmatul Fajriyah

Islamic University of Indonesia <https://orcid.org/0000-0002-7132-3937>

Kartick Chandra Mondal (✉ kartickjgec@gmail.com)

Jadavpur University <https://orcid.org/0000-0003-3647-5799>

Research Article

Keywords: Biclusters, Association Rule, Protein-Protein Interactions, COVID-19

Posted Date: September 18th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-73045/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Journal of The Institution of Engineers (India): Series B on May 4th, 2021. See the published version at <https://doi.org/10.1007/s40031-021-00569-7>.

Abstract

COVID-19 pandemic defined a worldwide health crisis into a humanitarian crisis. Amid this global emergency, human civilization is under enormous strain since no proper therapeutic method is discovered yet. A wave of research effort has been put towards the invention of therapeutics and vaccines against COVID-19. Contrarily, the spread of this fatal virus has already infected millions of people and claimed many lives all over the world. Computational biology can attempt to understand the protein-protein interactions between the viral protein and host protein. Therefore potential viral-host protein interactions can be identified which is known as crucial information towards the discovery of drugs. In this paper, we have presented an approach for predicting novel interactions from maximal biclusters. Additionally, the predicted interactions are verified from biological perspectives. For this, we conduct a study on the gene ontology and KEGG pathway in relation to the newly predicted interactions.

1. Introduction

Protein-protein interactions (PPI) denote a complex network of reactions that take the responsibility to synchronize and execute the biological process at the cell level in all organisms. PPI dataset is of immense importance in molecular and system biology. Due to its power of revealing the infection mechanism by viral protein, nowadays it plays a crucial role in the study of drug discovery. Thus treatment optimization is enhanced. A large number of experimental and computational approaches [11] have attempted to the way of drug discovery following the PPI dataset as the computational approaches can infer the predicted interactions [6, 5]. Hence, analyzing the PPI dataset of SARS-CoV-2 and human would be useful.

Being a computer science researcher, we could show an efficient algorithmic approach so that the prediction of the protein-protein interaction could be done in an efficient way. Not only that, we expand our scope by incorporating a way to validate the predictions made. Any interaction dataset is easily represented by binary datasets which are easier to store in a compact form. Hence, we find it worthy to use Bimax [9] biclustering approach that could be used to generate biclusters [4]. Also, a systematic manner can find the association rules [1] from generated biclusters and finally the predicted interactions from the binary PPI dataset.

Multiple efforts have already made towards this direction of drug discovery. A very few research [3, 2] have been done specifically in case SARS-CoV-2 or COVID-19 disease. Multiple groups of researchers have studied the HIV-human protein interaction. [5, 7, 8] are a few notable contributions among many. The literature has shown that many such analyses are done over the PPI network following the classification based methodology. But it suffers from subjective judgment which is the underlying feature for the designing of a classifier. Specific to the context, for the PPI dataset, a classifier needs both positive and negative interaction data. Though positive cases are validated experimentally, there is no evidence of negative interactions. It is true that based on the quantity of positive data, classifier become reliable. At the same time, the selection of negative data is also important for performance consideration. Thus a good classifier needs proper exploitation of both positive and negative interaction data. This kind of barrier directs us to follow the approach of rule mining based methodology which has been successfully implemented in [5].

So, summarizing the above discussion, in this paper we have followed bi-max biclustering approach on the manually curated protein protein interaction dataset of SARS-COV-2 and human protein. We have shown the procedure for predicting interactions from biclusters and from the generated rules as well. Moreover, we relate our prediction to gene-ontology and KEGG pathway-based study using two well-known bioinformatics tool.

The organization of the remaining part of the paper is as follows. The detailed explanation of the followed approach is featured in section 2. In section 3, the description of SARS-CoV-2-Human PPI dataset is presented. Section 4 highlights the predicted interactions and briefly justify their biological relevance. Finally, section 5 concludes this paper.

2. Approach

In order to predict new interactions from the existing interaction set, we make use of the association rule mining [1] approach. It is consisting of two main sub-parts. These are finding frequent item-sets (FI) or patterns (FP) and generating association rules from those patterns. As finding FP is computationally expensive, we decide to focus only on the maximal frequent patterns [9]. All the supersets of such patterns are infrequent but information loss arises in the case of the subsets. This is because all subsets are frequent but supporting objects are not available. To combat this information loss while minimizing the computational cost, frequent closed itemset [5] (FCI)/ bicluster mining could be the most favorable way. A FI is called closed if there exists no superset S

of FI such that supporting objects of S = supporting objects of FI. Thus using the set of FCI, all FI across the column attributes along with their supporting row objects can be extracted. In consequence, this approach reduces the number of rules to be presented to the users without losing any information.

In this regard, it can be concluded that finding maximal biclusters from a binary dataset is synonymous to retrieving the FCI in frequent pattern mining problem. Thus the first subproblem is converted into the problem of maximal bicluster generation. The step of generation of maximal biclusters is followed by the step of rule generation and new interaction prediction. These two are the main subtasks here as shown in Figure 1 and are discussed below.

Biclustering [4] is a way to compute the cohesiveness among the elements of a dataset. Bimax [9] is a well-known biclustering algorithm for extracting the maximal biclusters from the binary datasets. It follows a simple divide-conquer approach and extracts all inclusion maximal biclusters, basically, submatrices of all 1s'. An inclusion maximal bicluster indicates that the bicluster is not contained completely in any other bicluster. Here we use bimax as it accomplishes results within comparable execution time and memory usage along with the other best-performing algorithms. We use the obtained biclusters to predict new interactions and further to generate rules. Additionally, these rules also do prediction for new interactions. A straight forward approach is used to generate the rules from the generated biclusters or maximal closed frequent pattern as explained in [5]. Pictorial representation of the process is shown in Figure 1. Here, we have considered and generated only a non-redundant set of association rules for the necessary knowledge prediction.

3. Preparation Of The Sars-cov-2-human Ppi Dataset

The experimental dataset is created by combining PPI interactions from two different sources viz, BioGRID repository [10] and a research article [2]. BioGRID is a conventional biological repository for protein-protein interaction, genetic interaction, etc. We found very less amount of interaction in the data repository to extract sufficient predicted information. For this, an extra prediction in the data preprocessing step is carried out. We predict that interactions may possible by the rule of transitivity. We come across the data of intra-organism SARS-CoV-2 protein interactions. We assume that a SARS-CoV-2 protein A interacts with other SARS-CoV-2 proteins X, Y, Z, and human proteins as well. Moreover, X, Y, and Z interact with some other human proteins. Hence, SARS-CoV-2 protein A would also interact with the human proteins that have interaction with X, Y, and Z.

Finally, our dataset consists of 28 SARS-CoV-2 proteins and 346 human proteins. The interaction data covers both inter-organism PPI in between human and SARS-CoV-2, as well as intra-organism PPI of SARS-CoV-2. Before applying the transitive-law for the SARS-CoV-2 proteins interaction data, we found 405 interactions. This prediction gives us a total of 1255 interaction data. In our dataset, 1 represents an interaction between the human protein and viral protein. This binary dataset is treated as an input to the bimax algorithm. The input dataset is given as supplementary material along with this article.

4. Results And Discussion

SARS-CoV-2-Human PPI dataset is denoted as HxC where the rows represent the human proteins and the columns represent the viral proteins. In this case, we get the clustered itemsets of SARS-CoV-2 proteins that have similar interaction pattern for a subset of objects of human proteins. We have also used the transposed form of the SARS-CoV-2-Human PPI dataset which is represented by the CxH. The aim of this transposed data is to obtain the clusters along with the human proteins for a subset of objects of SARS-CoV-2 proteins. Here, each bicluster specifies that with a subset of human proteins, a subset of SARS-CoV-2 proteins interacts and vice-versa.

Table 1 Statistics for the generated outputs

Dataset	# Biclusters	# Rules	Rules > 0.7 Confidence	Non-redundant Rules	# Predicted Interactions
HxC OLD	10	1	0	0	0
CxH OLD	9	1	0	0	0
HxC	34	30	9	5	48
CxH	34	20	11	7	8

We use both HxC and CxH as input to the bimax algorithm and summary of generated biclusters can be found in Table 1. It has listed the interaction information for two sets of both of the datasets, i.e before (HxC OLD, CxH OLD) and after (HxC, CxH) applying the rule of transition on the original known protein protein interaction information. The improved results for the updated datasets can be observed from Table 1.

4.1 Interactions prediction from biclusters

Interactions can be predicted from the obtained list of biclusters. Let us take an example of HxC dataset and consider two obtained biclusters in the form of $\langle IS_CP1 : IS_HP1 \rangle$ and $\langle IS_CP2 : IS_HP2 \rangle$ where $\langle IS_CPI : IS_HPI \rangle$ represents itemsets denoting i^{th} bicluster consisting of SARS-CoV-2 protein and human protein. To predict new interactions from the biclusters itself, we need to take the

advantage of set-difference operator. $|IS_CP1 - IS_CP2| \xrightarrow{\text{Predicts Interaction With}}$

$|IS_HP2 - IS_HP1|$. Similarly, $|IS_CP2 - IS_CP1| \xrightarrow{\text{Predicts Interaction With}}$

$|IS_HP1 - IS_HP2|$. Following this way, we obtain a list of predicted interactions, and Table 2 is showing two sets of examples of such a predicted list.

Table 2 Predicted interactions from biclusters

Obtained Biclusters	SARS-CoV-2 Protein	Human Protein
1a. $\langle \text{ORF8 ORF9C NSP7 NSP8 NSP12 : FAM162A LMAN2 HS2ST1 NAT14 RALA EMC1} \rangle$	ORF9C $\xrightarrow{\text{Interacts}}$	GLA AKAP8L
1b. $\langle \text{M ORF8 NSP7 NSP8 NSP12 : GLA HS2ST1 AKAP8L} \rangle$	M $\xrightarrow{\text{Interacts}}$	FAM162A LMAN2 NAT14 RALA EMC1
2a. $\langle \text{E ORF9B : FKBP10 AP3B1 BRD2 BRD4 CWC27 SLC44A2 ZC3H18 BAG5 CHMP2A CSDE1 DCTPP1 DPH5 MARK1 MARK2 MARK3 PTBP2 SLC9A3R1 TOMM70A} \rangle$	ORF9B $\xrightarrow{\text{Interacts}}$	RAB14 DDX21 G3BP2 LARP1 MOV10 PABPC1 PABPC4 RBM28 RPL36 RRP9 UPF1 ALG5 ARL6IP6 CLCC1 HMOX1 SUN2 TRIM59 VPS11 VPS39 CSNK2A2 CSNK2B FAM98A G3BP1 SNIP1
2b. $\langle \text{E N ORF3A NSP4 : RAB14 DDX21 G3BP2 LARP1 MOV10 PABPC1 PABPC4 RBM28 RPL36 RRP9 UPF1 AP3B1 BRD2 BRD4 CWC27 SLC44A2 ZC3H18 ALG5 ARL6IP6 CLCC1 HMOX1 SUN2 TRIM59 VPS11 VPS39 CSNK2A2 CSNK2B FAM98A G3BP1 SNIP1} \rangle$	N ORF3A NSP4 $\xrightarrow{\text{Interacts}}$	FKBP10 BAG5 CHMP2A CSDE1 DCTPP1 DPH5 MARK1 MARK2 MARK3 PTBP2 SLC9A3R1 TOMM70A

4.2 Prediction by generating rules from the biclusters

This section establishes the approach that we have already mentioned in section 2. CxH dataset gives rules between human proteins and HxC dataset gives rules between SARS-CoV-2 proteins which are true for corresponding SARS-CoV-2 and human protein, respectively. The obtained rules are filtered out in two steps. First, we filter the rules based upon the support and confidence values.

Being a sparse dataset, experimentally we fix the support values to a lower threshold as follows. For the CxH dataset, the minimum support count is kept to 4 with a 70% minimum confidence level. Similarly, for the HxC dataset, the minimum support count is 11 with the same minimum confidence level as 70%. Next, the redundancy is eliminated taking the most general rules. Then the interaction prediction is performed. From the final results, the HxC dataset predicts 48 unique and novel interactions. Similarly, from CxH dataset, 8 unique and novel interactions are predicted. To keep the prediction unambiguous, we discard the interactions with a lower confidence level in case of the similar predicted interactions. The predicted interactions obtained from both the datasets are merged and a PPI network is drawn as shown in Figure 2. We find 3 common interactions and thus finally 53 unique novel interactions are obtained from the dataset. The confidence level for each predicted interaction lies within 85% to 90% in most of the cases followed by 90% to 95% interval. For this, we have shown a histogram in Figure 2 along with the interaction network. It can be seen from the figure that SARS-CoV-2 proteins ORF8, NSP12, and E act as network hub. All SARS-CoV-2 proteins are highlighted in yellow color whereas the human proteins are in pink colour. The edge width representing an interaction is getting wider along with the increasing confidence level of the interaction. Edge level is showing the value of confidence level for each interaction.

4.3 Gene Ontology based justification of the predicted interactions

To justify the prediction, we have tried to find relevance with the help of biological interpretation. For this, we make use of DAVID (<http://david.abcc.ncifcrf.gov>), a freely available online bioinformatics repository that provides functional annotations for a large set of genes. Among the multiple information extracted from this tool, we opt for gene ontology (GO) based study and KEGG Pathway. From all the three domains (Biological Process, Cellular Component, and Molecular Function) covered by GO, we find the GO terms. Moreover, we extract the non-redundant informative GO terms, based upon the p-values (taken from DAVID generated result), by using another tool REVIGO (<http://revigo.irb.hr/>). It reveals the outliers from the list of submitted GO-terms via checking the semantic similarity and outputs a sorted list based upon dispensability value for each GO-term. More unique terms are having lesser dispensability.

4.3.1 For the predicted interactions obtained from the biclusters

As shown in Table 2, ORF9B is predicted to have interactions with 24 human proteins. Table 3 is showing GO-terms for verifying the predictions. It can be seen from the molecular function of predicted human proteins that these are related to identical binding activities. Study has shown that, ORF9B also has a functional property of *membrane binding*.

Table 3 Significant GO terms found in the human proteins that are predicted to interact with SARS-CoV-2 Protein ORF9B

GO-id	Term	% of Proteins	P value	Dispensability
GO Category: Biological Process (P-value < 1E-01, Dispensability < 0.6)				
GO:1903955	positive regulation of protein targeting to mitochondrion	12.5	5.11E-3	0
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	12.5	9.57E-3	0.019
GO:0006413	translational initiation	12.5	1.25E-3	0.106
GO:0007265	Ras protein signal transduction	8.33	8.40E-2	0.183
GO:0016192	vesicle-mediated transport	12.5	1.52E-2	0.216
GO:0006364	rRNA processing	12.5	2.89E-2	0.335
GO:0034058	endosomal vesicle fusion	8.33	1.12E-2	0.347
GO:0045070	positive regulation of viral genome replication	8.33	3.20E-2	0.445
GO:0008333	endosome to lysosome transport	8.33	4.76E-2	0.453
GO:0016239	positive regulation of macroautophagy	8.33	2.83E-2	0.468
GO:0045727	positive regulation of translation	8.33	6.42E-2	0.516
GO Category: Cellular Component (P-value < 1E-01, Dispensability < 0.6)				
GO:0010494	cytoplasmic stress granule	12.5	9.35E-04	0
GO:0016020	membrane	29.17	5.05E-2	0
GO:0005622	intracellular	25	2.30E-2	0.075
GO:0030123	AP-3 adaptor complex	8.33	1.3E-2	0.227
GO:0005829	cytosol	41.67	1.51E-2	0.327
GO:0005765	lysosomal membrane	12.5	4.63E-2	0.33
GO:0005730	nucleolus	20.83	2.10E-2	0.377
GO:0030897	HOPS complex	8.33	1.75E-2	0.417
GO:0031519	PcG protein complex	8.33	3.35E-2	0.491
GO:0030529	intracellular ribonucleoprotein complex	16.66	6.45E-04	0.54
GO Category: Molecular Function (P-value < 1E-01, Dispensability < 0.05)				
GO:0044822	poly(A) RNA binding	54.17	1.26E-09	0
GO:0000166	nucleotide binding	25	5.61E-05	0
GO:0004004	ATP-dependent RNA helicase activity	12.5	2.92E-3	0

GO:0003729	mRNA binding	12.5	1.04E-2	0
GO:0008494	translation activator activity	8.33	1.11E-2	0
GO:0008143	poly(A) binding	8.33	1.60E-2	0
GO:0005515	protein binding	70.83	2.10E-2	0
GO:0003676	nucleic acid binding	20.83	3.11E-2	0

Table 4 Significant GO terms found in the human proteins that are predicted to interact with SARS-CoV-2 Protein ORF8

GO-id	Term	% of Proteins	P value	Dispensability
GO Category: Biological Process (P-value < 1E-01, Dispensability < 0.3)				
GO:0000398	mRNA splicing, via spliceosome	14.28	2.57E-02	0
GO:0017148	negative regulation of translation	9.52	6.36E-02	0.205
GO Category: Cellular Component (P-value < 1E-01, Dispensability < 0.2)				
GO:0016020	membrane	33.33	1.16E-2	0
GO:0071013	catalytic step 2 spliceosome	9.52	8.24E-2	0
GO:0005737	cytoplasm	42.85	8.30E-2	0.15
GO Category: Molecular Function (P-value < 1E-01, Dispensability < 0.05)				
GO:0044822	poly(A) RNA binding	33.33	1.12E-2	0

4.3.2 For the predicted interactions obtained from the rules

Among the predicted interactions, SARS-CoV-2 protein ORF8 is found to have maximum predicted interactions with 21 human proteins. Followed by this, 16 and 8 numbers of human proteins have been predicted to interact with SARS-CoV-2 proteins NSP12 and E, respectively. Below, Table 4, 5, and 6 are summing up the informative GO-terms verifying the identical biological activities for viral proteins ORF8, NSP12, and E, respectively.

Table 5 Significant GO terms found in the human proteins that are predicted to interact with SARS-CoV-2 Protein NSP12

GO-id	Term	% of Proteins	p value	Dispensability
GO Category: Biological Process (p-value < 1E-04, Dispensability < 0.05)				
GO:0007077	mitotic nuclear envelope disassembly	35.71	2.88E-08	0
GO:0006409	tRNA export from nucleus	28.57	1.77E-06	0
GO:0010827	regulation of glucose transport	28.57	1.95E-06	0
GO:0075733	intracellular transport of virus	28.57	7.39E-06	0
GO:1900034	regulation of cellular response to heat	28.57	2.37E-05	0
GO:0031047	gene silencing by RNA	28.57	7.66E-05	0
GO:0016925	protein sumoylation	28.57	8.96E-05	0
GO Category: Cellular Component(p-value < 1E-01, Dispensability < 0.75)				
GO:0044613	nuclear pore central transport channel	21.43	3.65E-05	0
GO:0001527	microfibril	14.285	7.11E-03	0
GO:0005829	cytosol	42.85	7.06E-02	0.1
GO:0031012	extracellular matrix	21.42	1.82E-02	0.5
GO:0005643	nuclear pore	14.28	5.01E-02	0.7
GO:0005578	proteinaceous extracellular matrix	21.42	1.51E-02	0.7
GO category: Molecular Function(p-value <1E-01, Dispensability < 0.05)				
GO:0017056	structural constituent of nuclear pore	21.43	1.14E-04	0
GO:0005515	protein binding	92.86	2.65E-03	0
GO:0005487	nucleocytoplasmic transporter activity	14.29	1.68E-02	0
GO:0004386	helicase activity	14.29	6.35E-02	0
GO:0005178	integrin binding	14.29	7.80E-02	0.04
GO:0030023	extracellular matrix constituent conferring elasticity	14.29	3.84E-03	0.31
GO:0005201	extracellular matrix structural constituent	14.29	5.04E-02	0.4

While the activities are highly cohesive within each table data, each individual table has a distinct set of functions. Table 4 depicts significance GO terms like *membrane*, *poly(A) RNA binding*, etc, indicating the involvement of many human proteins in these. It is an important observation as these proteins are expected to have interaction with ORF8 that plays the main role in host-virus interaction. SARS-CoV-2 protein NSP12 is a multifunctional protein and mainly involved in the transcription and replication of viral RNAs. Table 5 is found to have many such common GO-terms related to RNA functions in the biological process. Similarly, from Table 6, it appears that many human proteins are involved in the molecular function of *protein transporter activity*. It says that these proteins are highly involved in transporting molecules across biological membrane. Hence, the prediction of the human proteins that interact with viral protein E is intuitive as it is a small membrane protein having a major role in the assembly of virions.

Table 6 Significant GO terms found in the human proteins that are predicted to interact with SARS-CoV-2 Protein E

GO-id	Term	% of Proteins	P value	Dispensability
GO Category: Biological Process (P-value < 1E-01, Dispensability < 0.7)				
GO:0006626	protein targeting to mitochondrion	42.86	2.38E-05	0
GO:0007605	sensory perception of sound	28.57	3.12E-02	0
GO:0072321	chaperone-mediated protein transport	28.57	1.90E-03	0.42
GO:0015031	protein transport	28.57	9.08E-02	0.68
GO Category: Cellular Component (P-value < 1E-03, Dispensability < 0.7)				
GO:0042719	mitochondrial intermembrane space protein transporter complex	42.86	9.03E-07	0
GO:0005739	mitochondrion	71.43	3.77E-04	0.28
GO:0005743	mitochondrial inner membrane	71.43	4.88E-06	0.65
GO:0042721	mitochondrial inner membrane protein insertion complex	28.57	9.87E-04	0.68
GO Category: Molecular Function (P-value < 1E-01, Dispensability < 0.5)				
GO:0042803	protein homodimerization activity	42.86	2.50E-2	0
GO:0008565	protein transporter activity	28.57	2.53E-2	0
GO:0005215	transporter activity	28.57	6.97E-2	0
GO:0008270	zinc ion binding	42.86	6.0E-2	0.06
GO:0051087	chaperone binding	28.57	2.94E-2	0.46

4.4 Justification on the predicted interactions using KEGG pathway

Along with the GO study, we also examine the KEGG pathway obtained from DAVID tool. KEGG pathway has importance in bioinformatics research in understanding genomes, biological pathways, disease, drugs, etc.

4.4.1 For the predicted interactions obtained from the biclusters

Here we would address the example of KEGG pathway enrichment for ORF9B. The tool reveals that the human proteins that are predicted to have interactions with ORF9B are related to the *Metabolic pathways*, pathways to *Measles*, *Herpes simplex infection*. Many of the proteins are related to the *B-signaling pathway*, *RNA degradation*, etc.

4.4.2 For the predicted interactions obtained from the rules

From the KEGG Pathway obtained using the tool, it has come by the probable interaction of ORF8 with human protein which indicates that multiple cellular activities may lead to Human T-Cell Leukemia Virus Infection. Human proteins that are predicted to interact with ORF8, are involved in the pathway of *HTLV-I infection*, *Apoptosis*, *Hepatitis C*, *Epstein-Barr virus infection*, *Insulin signaling pathway*, etc. For the predicted interactions with NSP12, the pathways are *Metabolic pathways*, *RNA transport*, etc. Similarly, SARS-CoV-2 protein E is expected to interact with human protein ALG11, IDE that are found to have involvement in *Metabolic pathways*, *Alzheimer's disease*, etc.

5. Conclusions

Our main aim here is to help in accelerating the procedure of designing the drug and hence an improved medication for COVID-19. For this, we exploit the simple binary SARS-CoV-2-Human PPI dataset. Formation of biclusters, association rule generation, and the

procedure for discovering novel interactions are shown here. Also, the predicted interactions are interpreted biologically for finding their relevance. It has been seen that multiple human proteins that have predicted interaction with a single viral protein share common biological activities. Our study has left scope for consideration of multiple interaction types. Directions of the interactions (viral to host or host to viral) are also crucial information to be examined and working with.

Declarations

Competing interests:

The authors declare no competing interests.

References

1. Rakesh Agarwal, Ramakrishnan Srikant, et al. Fast algorithms for mining association rules. In *of the 20th VLDB Conference*, pages 487–499, 1994.
2. David E Gordon, Gwendolyn M Jang, Mehdi Bouhaddou, Jiewei Xu, Kirsten Obernier, Kris M White, Matthew J O’Meara, Veronica V Rezelj, Jeffrey Z Guo, Danielle L Swaney, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, pages 1–13,
3. Abdul Arif Khan and Zakir Khan. Covid-2019 associated overexpressed Prevotella proteins mediated host-pathogen interactions and their role in coronavirus outbreak. *Bioinformatics*,
4. Sara C Madeira and Arlindo L Oliveira. Biclustering algorithms for biological data analysis: a *IEEE/ACM transactions on computational biology and bioinformatics*, 1(1):24–45, 2004.
5. Kartick Chandra Mondal, Nicolas Pasquier, Anirban Mukhopadhyay, C’elia da Costa Pereira, Ujjwal Maulik, and Andrea GB Tettamanzi. Prediction of protein interactions on HIV-1-human PPI data using a novel closure-based integrated approach. In *International Conference on Bioinformatics Models, Methods and Algorithms*, pages 164–173. SciTePress,
6. Anirban Mukhopadhyay, Ujjwal Maulik, and Sanghamitra A novel biclustering approach to association rule mining for predicting HIV-1–human protein interactions. *PLoS One*, 7(4), 2012.
7. Debasmita Pal, Anindita Sarkar Mondal, and Kartick Chandra Mondal. Knowledge discovery from HIV-1-human PPIs assimilating interaction keywords. In *2016 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*, pages 1–8. IEEE,
8. Debasmita Pal and Kartick Chandra Mondal. Predicting novel interactions from hiv-1- human ppi data integrated with protein signatures and go annotations. *International Journal of Bioinformatics Research and Applications (IJBRA)*,
9. Amela Prelić, Stefan Bleuler, Philip Zimmermann, Anja Wille, Peter Buchmann, Wilhelm Gruissem, Lars Hennig, Lothar Thiele, and Eckart A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics*, 22(9):1122–1129, 2006.
10. Stark, B. J. Breitkreutz, L. Reguly, T. and Boucher, and M. Breitkreutz, A. and Tyers. Biogrid: a general repository for interaction datasets. *Nucleic acids research*, 34(Database issue):D535–D539, 2006.
11. Hina Umbrin and Saba Latif. A survey on protein protein interactions (ppi) methods, databases, challenges and future directions. In *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, pages 1–6. IEEE,

Figures

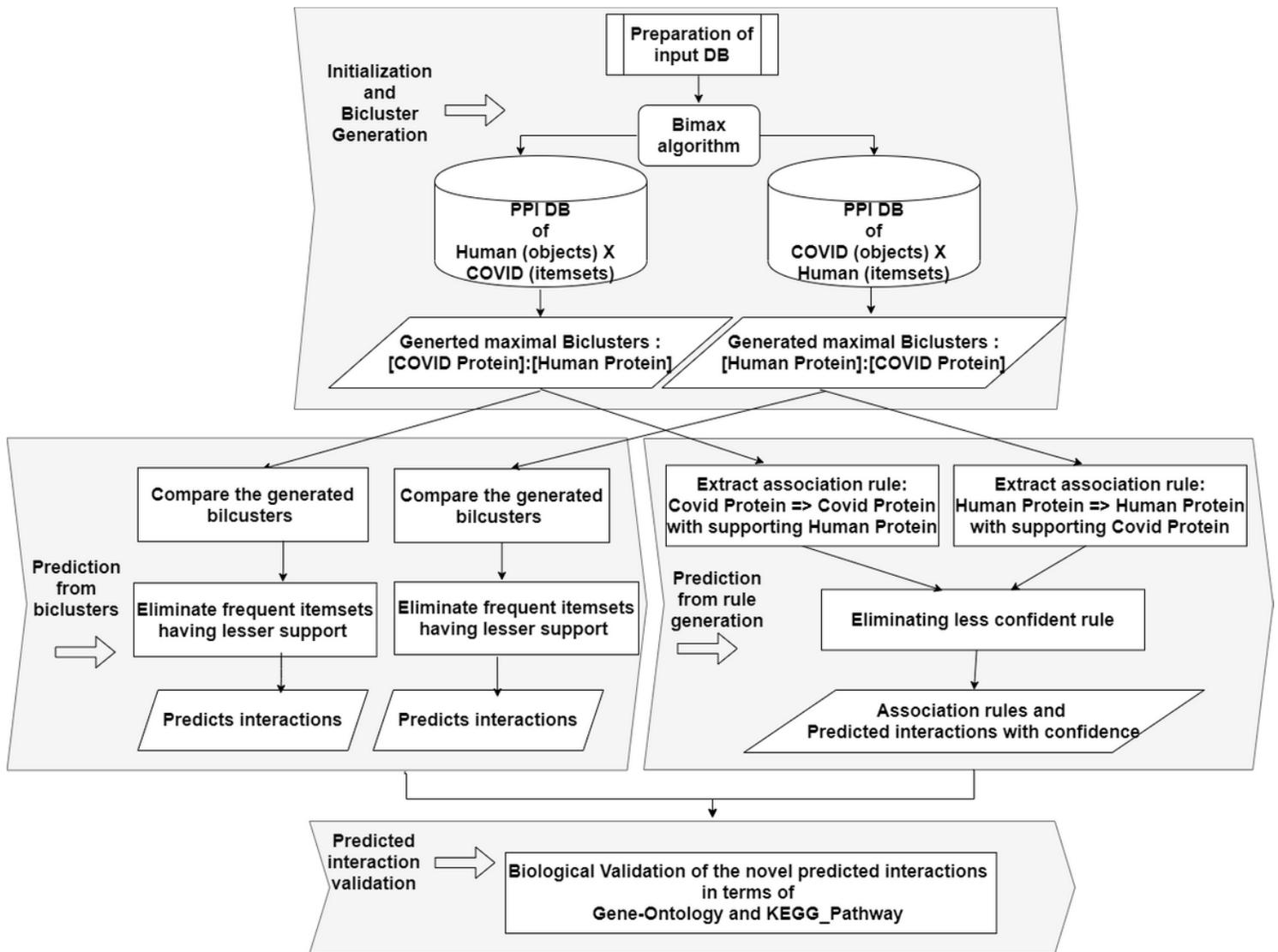


Figure 1

Outline of the approach used for the experimentation performed

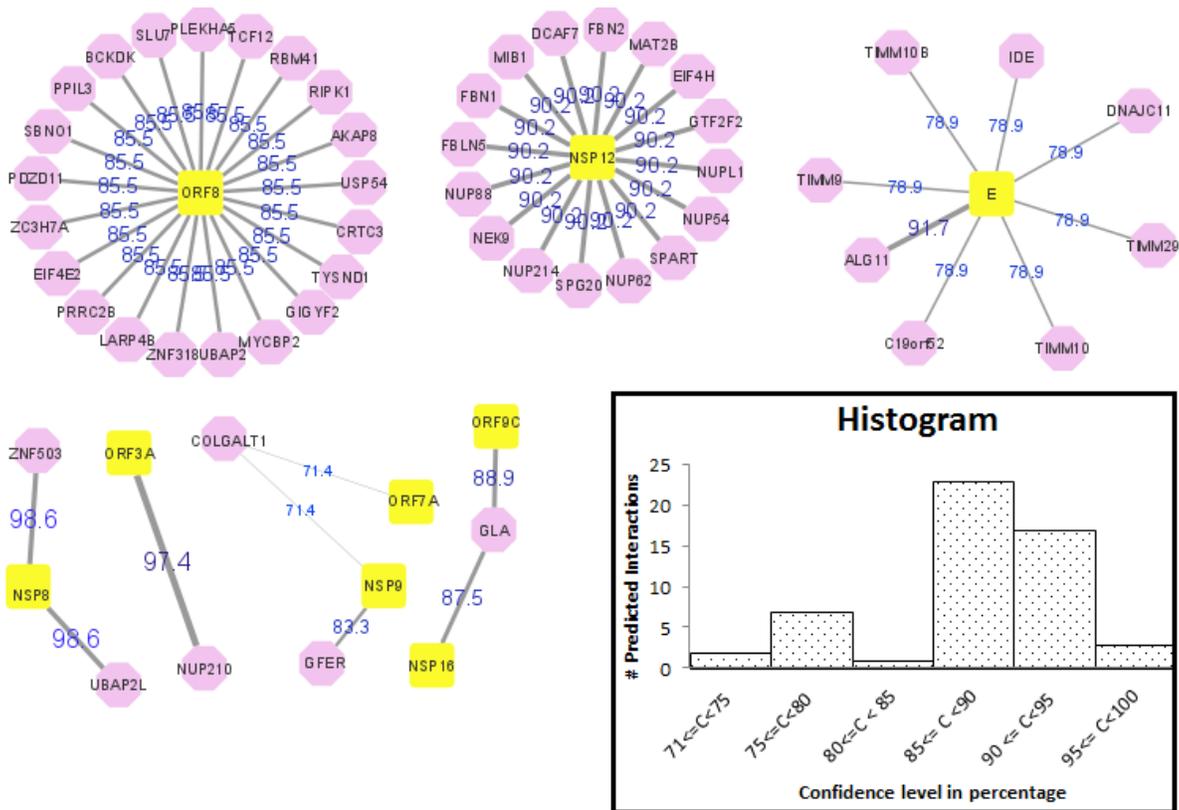


Figure 2

Network for the predicted PPIs along with histogram for the number of predicted interactions at the varying confidence level

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [PPICOVIDHUMAN.csv](#)
- [PPICOVIDHUMANT.csv](#)
- [FinalPredictedInteractions.xlsx](#)