

Identification of Significant Proteins in Coronavirus Disease 2019 Protein-Protein Interaction Using Principal Component Analysis and ClusterONE

Hilmi Farhan Ramadhani¹, Annisa¹, Wisnu Ananta Kusuma^{2*}

¹Department of Computer Science, Faculty of Mathematics and Natural Sciences, IPB University, Bogor 16680, Indonesia

²Department of Computer Science, Faculty of Mathematics and Natural Sciences, Tropical Biopharmaca Research Center, Bogor 16680, Indonesia

Article history:

Submission October 2020

Revised November 2020

Accepted December 2020

*Corresponding author:

E-mail:

ananta@apps.ipb.ac.id

ABSTRACT

Coronavirus Disease 2019 (COVID-19) will cause disease complications and organ damage due to excessive inflammatory reactions if left untreated. Computational analysis of protein-protein interactions can be carried out in various ways, including topological analysis and clustering of protein-protein interaction networks. Topological analysis can identify significant proteins by measuring the most important nodes with centrality measurements. By using Principal Component Analysis (PCA), the types of centrality measures were extracted into the overall centrality value. The study aimed to find significant proteins in COVID-19 protein-protein interactions using PCA and ClusterONE. This study used 57 proteins associated with COVID-19 to obtain protein networks. All of these proteins are homo sapiens organism. The number of proteins and the number of interactions from 57 proteins were 357 proteins and 1686 interactions. The results of this study consisted of two clusters; the best cluster was the first cluster with a lower p-value but had an average overall centrality value that closed to the second cluster. There are twenty important proteins in that cluster, and all of these proteins are related to COVID-19. These proteins are expected to be used in the process of discovering medicinal compounds in COVID-19.

Keywords: Centrality measurements, ClusterONE, COVID-19, Overall centrality, Protein-protein interaction, Significant protein

Introduction

In December 2019, a case of mysterious Pneumonia was first reported in Wuhan, Hubei, China [1]. The source of the transmission of this case is still uncertain, but the first case was linked to a fish market in Wuhan [2]. From December 18, to December 29, 2019, five patients were treated with Acute Respiratory Distress Syndrome (ARDS) [3]. In less than a month, the disease has spread to other provinces in China, Thailand, Japan, South Korea and other regions [4]. Initially the disease was temporarily named as 2019 novel coronavirus (2019-nCoV) [5], then World Health Organization (WHO) announced a new name on

February 11, 2020, namely Coronavirus Disease 2019 (COVID-19). This disease is caused by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) [6]. This virus can be transmitted from person to person and has spread widely in China and other countries [7]. Since January 2020, WHO has declared that the world is entering into a global emergency related to this virus [8]. This phenomenon is an extraordinary phenomenon that occurs on earth in the 21st century after occurring SARS in 2002 and Middle East Respiratory Syndrome (MERS) in 2012 [9].

Based on data collected by the Worldometer [10], in December 3, 2020, there were at least

How to cite:

Ramadhani, HF., Annisa, Kusuma, WA. (2020). Identification of Significant Proteins in Coronavirus Disease 2019 Protein-Protein Interaction Using Principal Component Analysis and ClusterONE. *Bioinformatics and Biomedical Research Journal* 3(2):25-34. doi: 10.11594/bbrj.03.02.04

64,985,212 total cases of COVID-19 that had been reported in 220 countries, and 1,502,253 people have died from this disease. In Indonesia, the average number of cases added since early September 2020 was 2,000 cases per day, with the largest addition of cases was on November 29, 2020 (6,267 cases). The number of cases of the Indonesian population who tested positive for the COVID-19 was 557,877 cases, while the cases that were still active were 93,324 cases and the number of Indonesians who died was 17,355 people.

If this continues, it is feared that there will be complications of the disease when exposed to this virus because the steps applied are still limited to the design. If that happens, the patient may experience damage to the organs affected by this virus [11]. This virus attaches to the host cell and will cause an excessive inflammatory reaction [12]. Inflammation is a local protective response caused by damage to the tissue [13]. In inflammation, protein can be a biomarker of damage to organs [12].

Therefore, this study will look for significant proteins that cause inflammation of COVID-19. Significant protein is an important medium that contains all the nutrients needed by the body in the process of growth. In carrying out their duties, proteins interact with other proteins [14]. By using an experimental approach, need a long time and considerable cost to identify significant proteins [15]. Protein-protein interaction (PPI) can be used to properly recognize protein function, disease genes and the relationship between diseases. PPI can be formed into a network of interactions known as protein interaction networks [16]. In this network, there are significant proteins that are the most influential proteins in the protein interaction network [17].

PPI is a graph depicting the topology of the interactions between proteins. The nodes in the PPI represent proteins, while the edges represent the relationships between the interacting proteins. Computational analysis of PPIs can be carried out in various ways, including topology analysis and clustering on the PPI network [18]. Topological analysis can identify significant proteins by measuring the most influential nodes with centrality measures [18] whereas clustering in PPIs is carried out to form protein complexes which are proteins having the same function. Protein complexes can be used to predict significant protein [19].

There are various types of centrality measurements. From the several types of centrality measure values, there are significant protein values on the centrality measures. By using Principal Component Analysis (PCA), the types of centrality measures were extracted into the overall centrality value. Although the dimensions of centrality measures are reduced, they still minimize the loss of information [20]. By the overall centrality value, it can be easier to compare best cluster. A good clustering method for finding protein complexes is the soft clustering method because it can detect overlaps in the resulting clusters [21]. Overlapping detection is important in the case of finding significant proteins because proteins are multi-functional.

There have been several previous studies that examined significant protein using centrality measures. Identification of significant protein in Diabetes Mellitus using three centrality measurements, namely betweenness centrality, closeness centrality, and degree centrality [22]. The same thing is done, but diabetes is more specific, namely Diabetes Mellitus Type 2 using Fuzzy C-Means to found significant protein clusters and PCA to rank significant protein in each resulting cluster based on the overall centrality value [23].

Other study in Parkinson's disease, by using degree centrality, betweenness centrality, closeness centrality, bridging centrality, radiality, and eigenvector centrality. By using Skyline Query and PCA methods, the results in this study there are 14 proteins with 12 proteins are related on Parkinson's disease [24]. Other study used ClusterONE to find biomarkers in cancer with graph reduction using the degree of centrality. ClusterONE is one of the soft clustering methods [25]. In the case of COVID-19, there has been research on developing methods to predict SARS-CoV-2 human PPIs and reported the most central proteins in using centrality measurements [26].

The objective of this study is to found significant proteins in the PPI COVID-19 using PCA and ClusterONE. PCA is used to get the overall centrality value and this value is used as a basis for graph reduction. The values of centrality measures that will be reduced to overall centrality are degree centrality, betweenness centrality, closeness centrality, subgraph centrality, eigenvector centrality, information centrality, and network centrality. After that, grouping on PPI to form protein com-

plexes using ClusteONE. The clusters from ClusterONE were analyzed to determine whether the protein complexes in the best clusters were related with COVID-19.

Methods

Data

Gene

The gene data used is the protein associated with COVID-19. Genes are obtained from two database sites, namely the OMIM site (<https://www.omim.org/>) and UniProt site (<https://www.uniprot.org/>), and previous research about the COVID-19 protein [27-29]. From these various sources, there are 57 proteins obtained. All of these genes are homo sapiens organism.

Protein-Protein Interaction

The gene data that has been obtained are used as input on the STRING site (<https://string-db.org/>) to obtain PPI data. The first steps to get PPI data are entering the protein and organism symbols. The protein symbol is obtained from the protein data that has been obtained and select the homo sapiens organism. After that select the appropriate gene symbol. If the protein interaction view appears, download the data by using the export feature on the site into tsv file.

The PPI data obtained must go through the preprocessing stage. There are three preprocessing stages, such as combining PPI data, cleaning data duplication (between nodes there is only one connecting edge) and eliminating PPIs in small subgraphs that are not connected to the main protein interaction network.

All the proteins obtained previously had interactions with other proteins, so that after going through the PPI data combining stage and cleaning the duplication of data, the number of proteins and the number of interactions from 57 proteins were 357 proteins and 1686 interactions. PPI data visualization after this stage can be seen in Figure 1.

Eliminating PPIs in small subgraphs because it could interfere with other centrality measures in the main network. After passing through the final preprocessing stage, there are 222 proteins and 1239 interactions. PPI data visualization after the preprocessing data stage can be seen in Figure 2.

Centrality Measurements

In this study, the topological analysis method used to determine significant protein from a graph

is centrality measures [30]. The values of centrality measures in network topology can be used to

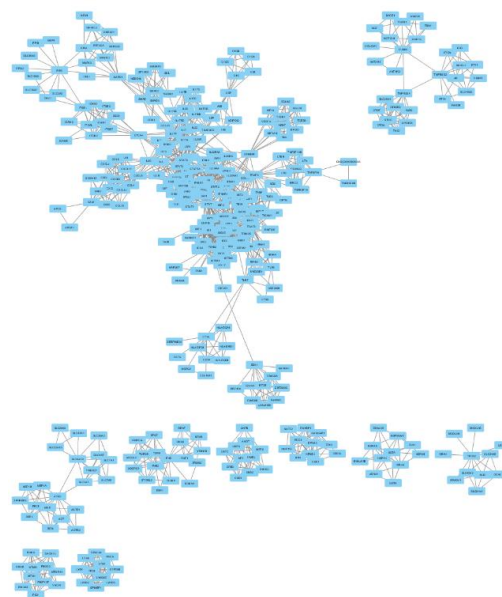


Figure 1. PPI data visualization after cleaning data duplication

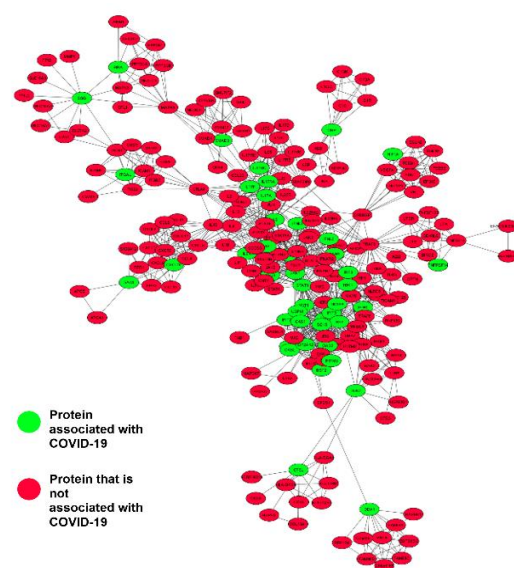


Figure 2. PPI data visualization after preprocessing

rank and measure the importance of nodes in a network based on their location [31]. In research to identify significant proteins, the method of centrality measures is the most widely used [32].

There are various types of centrality measurement. The value of centrality measures was obtained using the CytoNCA plugin on Cytoscape [33]. The plugin is more focused on looking for

the centrality measures value in the case of finding significant proteins. The values of centrality measures used in this study are as follows [34]:

Degree Centrality

Degree centrality is the simplest measure of value in centrality measures. Degree centrality can be found by counting all the edges that incident to a node. If the degree centrality value is greater, protein becoming the center of regulation. The equation for finding degree centrality value can be seen in Equation 1. $a_{u,v}$ is an adjacency matrix for node u and node v with value of 1 if connected and 0 otherwise.

$$DC(u) = \sum_v a_{u,v} \quad (1)$$

Betweenness Centrality

Betweenness centrality is the average of the shortest path that connects two nodes. If the betweenness centrality value obtained is greater, protein influences the interaction network between proteins. The equation for finding betweenness centrality can be seen in Equation 2.

$$BC(u) = \sum_s \sum_t \frac{\rho(s,u,t)}{\rho(s,t)}, s \neq t \neq u \quad (2)$$

$\rho(s, t)$ is the total shortest path between node s and node t , and $\rho(s, u, t)$ is the number of shortest paths between node s and node t passing through node u .

Closeness Centrality

Closeness centrality is the average shortest path in a node to access all nodes in the network. If the closeness centrality value obtained is greater, total distance to all proteins is getting less and position of the protein is getting central. The equation for finding closeness centrality can be seen in Equation 3.

$$CC(u) = \frac{N-1}{\sum_v d(u,v)} \quad (3)$$

$d(u, v)$ is the distance from node u to node v . Distance is the length of the shortest path of the two nodes. N is the number of nodes.

Subgraph Centrality

Subgraph centrality at a node is the number of subgraphs of the entire PPI network in which the nodes participate in subgraph creation. The equation for finding the value of subgraph centrality can be seen in Equation 4.

$$SC(u) = \sum_{l=0}^{\infty} \frac{\mu_l(u)}{l!} = \sum_{v=1}^N [\alpha_v(u)]^2 e^{\lambda_v} \quad (4)$$

$\mu_l(u)$ is the number of closed loops of length l . The length in this case is the number of edges which starts and ends at node u . $(\alpha_1, \alpha_2, \dots, \alpha_N)$ is

the orthonormal basis of R^N which are composed of eigenvectors of A associated with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_N$, where $\alpha_v(u)$ is the u^{th} component of α_v .

Eigenvector Centrality

The value of eigenvector centrality explains that the nodes that are connected to the more important nodes have a greater value. If the eigenvector centrality value obtained is greater, protein is more influential than other proteins. The equation for finding eigenvector centrality can be seen in Equation 5.

$$EC(u) = \alpha_{max}(u) \quad (5)$$

α_{max} is the eigenvector obtained from the largest eigenvalue of A . $\alpha_{max}(u)$ is the u^{th} component of α_{max} .

Information Centrality

Information centrality at a node is the harmonic mean lengths of the paths ending at that node. The equation can be seen in Equation 6.

$$IC(u) = \left[\frac{1}{N} \sum_v \frac{1}{l_{uv}} \right]^{-1} \quad (6)$$

$l_{uv} = (c_{uu} + c_{vv} - c_{uv})^{-1}$. Let D is a diagonal matrix of degree at each node and J is a matrix where all the elements are 1. Matrix $C = (c_{uv}) = [D - A - J]^{-1}$. For computational purposes, l_{uu} is infinite, thus $1/l_{uu} = 0$.

Network Centrality

Edge Clustering Coefficient (ECC) is one of the method that can be used to evaluate the importance of edges in a PPI network and can describe how close the two proteins are. The way to find the ECC value on an edge in the PPI can be calculated using Equation 7.

$$ECC(u, v) = \frac{z_{u,v}}{\min(d_u - 1, d_v - 1)} \quad (7)$$

$z_{u,v}$ is a possible triangle formed by node u and node v while d_u and d_v are the number of degree of node u and node v , respectively. After getting

all the *ECC* values on each edge in the PPI network, to find the network centrality value, add the *ECC* value of each edge connected to the node (Equation 8).

$$NC(u) = \sum_{v \in N_u} ECC(u, v) \quad (8)$$

Overall Centrality

Principal Component Analysis (PCA) is a method that can be used to transform high-dimensional data into lower dimensions without reducing or losing a lot of information in the data [35]. In other words, PCA maximizes diversity to retain the information on that data. The output from PCA is linear combinations with variations.

In this study, PCA was used to find the overall centrality value. The overall centrality value is a value that considers all the characteristics that exist in the optimal linear combination of the previous centrality values [36]. In other words, the overall centrality value is a value from the reduction dimensions of the centrality measures.

To get the overall centrality value, first look for the overall centrality equation using PCA. PCA will extract features from the centrality measures. The overall centrality value is used to express which centrality values contain the most relevant information in identifying influencing nodes in a graph [32]. Prior to feature extraction, each centrality measure is standardized using the StandardScaler so that it has a mean distribution of 0 and a standard deviation of 1 because the value of each centrality measure has a much different range. The standardization equation can be seen in Equation 9.

$$x_{u,v}' = \frac{x_{u,v} - \mu_v}{\sigma_v} \quad (9)$$

$x_{u,v}'$ is the value of standardization while $x_{u,v}$ is the u th data value of the attribute v . μ_v and σ_v are the mean and standard deviation of attribute v , respectively. Furthermore, to obtain linear combinations, it can be done by finding the eigenvector and eigenvalue values by using a covariance matrix, a matrix that explains the diversity of a number of attributes or variables. The stage in finding the overall centrality value namely, there is an $N \times 7$ matrix where N is the amount of protein data and 7 are the centrality values used in this study. Let S is the covariance

matrix of the data matrix of size $N \times 7$ and $v = (v_1, v_2, \dots, v_7)$ is the eigenvector of S which is obtained from the largest eigenvalue, then the overall centrality value for each protein can be calculated using Equation 10.

$$OC(u) = v_1 DC(u) + v_2 BC(u) + v_3 CC(u) + v_4 SC(u) + v_5 EC(u) + v_6 IC(u) + v_7 NC(u) \quad (10)$$

All proteins are then sorted based on the value of overall centrality. After that, reduce the graph to get the subgraph using the induced graph method. The subgraph is formed by the node with the highest overall centrality value and all the edges that incident to that node. The amount of protein taken for subgraph formation is 10% of the total protein in the main graph. Uptake of 10% protein refers to research [25]. A value of 10% is also obtained using the power-law distribution on the human PPI graph [37]. This study proves that the 10% protein with the highest centrality value had a significant difference from the remaining 90% protein [37]. These proteins are linked to most of the proteins in the graph compared to the remaining 90% of the proteins.

ClusterONE

The objective of ClusterONE is to found cohesive clusters. A cohesive cluster is a cluster that has a larger number of edges connecting the nodes in a cluster compared to the number of edges connecting the nodes in the cluster with nodes outside the cluster. ClusterONE uses a greedy approach to form a cohesive cluster that starts from the seed node. The first seed node is the node with the largest degree.

There are 5 stages for cluster formation. First, let $V_0 = \{v_0\}$ and set the step number $t = 0$. Second, calculate the cohesiveness of V_t and let $V_{t+1} = V_t$. Third, for every external vertex v incident on at least one boundary edge, calculate the cohesiveness of $V' = V_t \cup \{v\}$. If $f(V') > f(V_{t+1})$, let $V_{t+1} = V'$. Fourth, for every internal vertex v incident on at least one boundary edge, calculate the cohesiveness of $V'' = V_t \setminus \{v\}$. If $f(V'') > f(V_{t+1})$, let $V_{t+1} = V''$. Fifth, if $V_t \neq V_{t+1}$, increase t and return to step 2. Otherwise, declare V_t a locally optimal cohesive group.

ClusterONE can be done using the ClusterONE plugin in the Cytoscape [21]. The parameters used in this method are a minimum

size of 3. Size is the number of nodes in a cluster. With a set minimum size value of 3, the minimum number of nodes in a cluster is 3.

The results of ClusterONE are clusters with several criteria. These criteria are density, quality, and p-value. Density is the ratio of the number of edges on the cluster and the number of edges that may form in the cluster. Quality is the ratio of in-weight and the sum of in-weight and out-weight. P-values were generated from the Mann-Whitney U Test on in-weight and out-weight. The cluster results are valid if the cluster has a p-value of less than 0.05. A valid cluster is not a random cluster that happened to be found. Furthermore, cluster analysis is carried out by comparing the overall centrality and p-value values in each cluster. The selected cluster is the cluster with the highest overall centrality value and a p-value less than 0.05. The smaller the p-value, the more valid the cluster [21].

Validation of Significant Protein Results

When constructing PPIs, the protein data that became the input for STRING had interactions with other proteins that had no related on COVID-19. These proteins have the possibility to be selected as a significant protein based on the topology of the COVID-19 interaction network. Therefore, the protein that has been produced needs to be analysed for its effect on COVID-19.

Validation of protein results was carried out by conducting literature studies from previous research that discussed the effect of protein on COVID-19. If based on the literature study there is evidence that states protein has an important effect on COVID-19, it will be considered proven to have a significant effect. In addition, an analysis of the results of significant proteins was obtained, such as looking for the function of these significant proteins with previous literature studies.

Table 1. Centrality measurements

No	Gene	Centrality Measures						
		DC	BC	CC	SC	EC	IC	NC
1	ADIPOQ	5	0	0.3	41371.06	4.55E-04	3.4	5
2	AIFM1	2	0	0.2	13.1	1.92E-05	2.2	2
3	ALB	6	215	0.3	41812.57	4.58E-04	3.7	5
...
222	ZFYVE9	6	0	0.2	1105.43	7.12E-04	3.7	6

Result and Discussion

Centrality Measurements

The following are some examples of the results of the transformation from PPI data to centrality measures data. The examples can be seen in Table 1.

Overall Centrality

After obtaining the values of centrality measures, and standardizing the values of centrality measures, the highest eigenvalue obtained using PCA was 5.22 and the eigenvector component were $v_1 = 0.43$, $v_2 = 0.24$, $v_3 = 0.36$, $v_4 = 0.40$, $v_5 = 0.39$, $v_6 = 0.42$, and $v_7 = 0.37$. From these values it can be concluded that in determining the value of overall centrality, the value of centrality measures that has the greatest effect is the degree centrality, while the value of centrality measures that has the least effect is betweenness centrality. The explained variance ratio generated from the eigenvector is 74.28%. Explained variance explains how much the information can be attributed to each principal component.

After obtaining the eigenvector value, this value is substituted in Equation 10 to obtain the overall centrality value, then sort the proteins based on the highest to lowest overall centrality value. The results of the calculation and ordering of the overall centrality value can be seen in Table 2.

Table 2. Overall centrality

No	Gene	OC
1	ADIPOQ	11.28
2	AIFM1	8.89
3	ALB	7.74
...
222	ZFYVE9	-3.22

The protein used in the clustering stage was only 10% protein with the highest overall centrality value of the previous protein amount. Therefore, from 222 proteins, 22 proteins with the highest overall centrality were selected to form the subgraphs that were used in the clustering stage. In this subgraph, there are 124 interactions formed by 22 proteins.

ClusterONE

There are two protein complex clusters resulting from the clustering of PPI. The results of clustering can be seen in Figure 3. The results of clustering will produce gray, red, and yellow nodes. The gray nodes show the outliers, the red nodes show the proteins that are only in one cluster, and the yellow nodes show the proteins that are in more than one cluster. The results of clustering in this study, there were no outlier nodes.

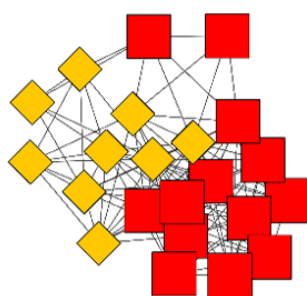


Figure 3. Clustering result

The detailed information on the resulting clusters can be seen in Table 3. Both clusters have a p-value of less than 0.05, it means that all clusters are valid clusters and not a random

cluster that happened to be found.

Table 3. Detail cluster information

Cluster	Size	Density	Quality	P-value
1	20	0.605	0.935	1.40E-08
2	11	0.600	0.452	0.013

From Table 3, the first cluster has a higher density and quality value compared to the second cluster, and also the first cluster has a much lower p-value than the second cluster. This means that the first cluster is better than the second cluster.

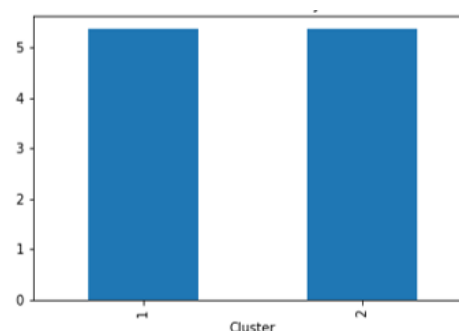


Figure 4. Average overall centrality of each cluster

Figure 4 shows the average value of overall centrality in each class. The first cluster has an average overall centrality value of 5.35 while the second cluster has an average overall centrality value of 5.36. With a p-value that is much lower than the second cluster, the first cluster has an average overall centrality value close to the second cluster. Based on the criteria already mentioned, the first cluster is a significant protein cluster for COVID-19.

Table 4. Influence protein on COVID-19.

No	Gene	Description
1	DDX58	Upregulation of this protein is under COVID-19 virus infection [38].
2	IFI6	This protein is related with an inflammatory response and others, among the top activated functional categories of broncho alveolar lavage (BAL) from patients with severe COVID-19. This protein was commonly found in BAL from patients with severe and mild COVID-19 [39].
3-8	IFIT1, IRF7, ISG15, MX1, OAS1, OAS2	These proteins could protect host cells from COVID-19 infection and can be considered as potential candidate for a drug targets in the treatments of COVID-19 [40].
9-11	IFIT3, MX2, STAT1	Epithelial cells infected by SARS-CoV-2 shared these proteins with human lung cells [41].
12-13	IRF3, TBK1	At least 3.5% of patients with life-threatening COVID-19 Pneumonia had known deficiency of these proteins [42].

14-16	IFNAR1, JAK1, TYK2	Multiplex gene expression analysis demonstrated upregulation of these proteins contrasting in SARS-CoV-2 critical patients [43].
17	IL6	This protein is one of the main mediators of inflammation and immune response initiated by infection. Elevated levels of IL6 were found in more than half of COVID-19 patients [44]. IL6 levels are related with inflammatory response, respiratory failure, mechanical ventilation and / or intubation, and the risk of death in COVID-19 patients [45, 46].
18	RSAD2	Two days after SARS-CoV-2 infection caused the interferon stimulated genes having the antiviral action. RASD is one of the antiviral action [47].
19	STAT2	STAT2 is an important protein related with COVID-19 and related to the immune system and respiratory organs, which emphasizes the fact that COPD ^a weakens these systems and makes patients more susceptible to developing COVID-19 [48].
20	STAT3	Dysfunctions of this protein induced by SARS-CoV-2 proteins may be the foundation of severe COVID-19 pathophysiology [49].

^aChronic Obstructive Pulmonary Disease (COPD) is a common lung disease in the form of emphysema and chronic bronchitis.

Validation of Significant Protein Results

In the first cluster, there are 20 proteins, namely STAT3, TYK2, IL6, STAT1, JAK1, STAT2, TBK1, RSAD2, OAS2, OAS1, MX2, MX1, ISG15, IRF7, IRF3, IFNAR1, IFIT3, IFIT1, IFI6, and DDX58. The influence of these proteins in COVID-19 can be seen in Table 4. Based on Table 4, all the proteins obtained are related to and have an effect on COVID-19.

Conclusion

In this study, by using topological analysis and clustering on protein networks, it is expected to obtain significant proteins that have an effect on COVID-19. In previous studies, some of the results were not optimal and many significant proteins had no effect on disease due to inappropriate methods.

The use of seven appropriate centrality measurements and overall centrality values to rank proteins will select proteins that had high centrality values. The ClusterONE method can detect overlaps in the resulting clusters, this method is good because the proteins are multi-functional. By combining these methods, this study reported 20 significant proteins and all of these proteins have an effect on COVID-19. These proteins are expected to be used in the process of discovering medicinal compounds in COVID-19.

For further research, research on the identification of significant proteins needed other data such as biological information data, namely gene ontology, subcellular localization, and others, and also PPI data have high false-positive values.

References

1. Lee PI, Hsueh PR (2020) Emerging threats from zoonotic coronaviruses-from SARS and MERS to 2019-nCoV. *J Microbiol Immunol Infect.* 53:365–367. doi: 10.1016/j.jmii.2020.02.001.
2. Rothan HA, Byrareddy SN (2020) The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J. Autoimmun.* 109:102433. doi: 10.1016/j.jaut.2020.102433.
3. Ren LL, Wang YM, Wu ZQ, Xiang ZC, Guo L, Xu T, et al. (2020) Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. *Chinese Med J.* 133:1015–1024. doi: 10.1097/CM9.0000000000000722.
4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 395:497–506. doi: 10.1016/S0140-6736(20)30183-5
5. Khaerunnisa S, Kurniawan H, Awaluddin R, Suhartati S, Soetjipto S (2020) Potential inhibitor of COVID-19 main protease (Mpro) from several medicinal plant compounds by molecular docking study. *Preprints.* 2020:1-14.
6. Gorbalenya AE, Baker SC, Baric RS, de Groot RJ, Drosten C, Gulyaeva AA, et al. (2020) The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol.* 5:536–544. doi: 10.1038/s41564-020-0695-z
7. Susilo A, Rumende CM, Pitoyo CW, Santoso WD, Yulianti M, Herikurniawan H, et al. (2020) Coronavirus disease 2019: Tinjauan literatur terkini. *Jurnal Penyakit Dalam Indonesia.* 7:1-45.
8. Buana DR (2020) Analisis perilaku masyarakat indonesia dalam menghadapi pandemi COVID-19 dan kiat menjaga kesejahteraan jiwa. *Jurnal Sosial Dan Budaya Syar-I.* 7:217–226.

9. Yuliana Y (2020) Corona virus diseases (Covid-19); sebuah tinjauan literatur. *Wellness Healthy Mag.* 2:187–192. doi: [10.30604/well.95212020](https://doi.org/10.30604/well.95212020).
10. Worldometer. Covid-19 Coronavirus Pandemic [Retrieved 12 Desember 2020]. Available from: <https://www.worldometers.info/coronavirus/>.
11. Rodríguez-Morales AJ, MacGregor K, Kanagarajah S, Patel D, Schlagenhauf P (2020) Going global – Travel and the 2019 novel coronavirus. *Travel Med Infect Dis.* 33:101578. doi: [10.1016/j.tmaid.2020.101578](https://doi.org/10.1016/j.tmaid.2020.101578).
12. Sukmana M, Yuniarti FA (2020) The pathogenesis characteristics and symptom of COVID-19 in the context of establishing a nursing diagnosis. *J. Kesehatan Pasak Bumi Kalimantan.* 3:21–28.
13. Ramadhani N, Sumiwi SA (2016) Aktivitas antiinflamasi berbagai tanaman diduga berasal dari flavonoid. *Farmaka.* 14:111–123.
14. Chang JW, Zhou YQ, Ul Qamar MT, Chen LL, Ding YD (2016) Prediction of protein–protein interactions by evidence combining methods. *Int J Mol Sci.* 17:1946. doi: [10.3390/ijms17111946](https://doi.org/10.3390/ijms17111946)
15. Acencio ML, Lemke N (2009) Towards the prediction of essential genes by integration of network topology, cellular localization and biological process information. *BMC Bioinformatic.* 10:1–18. doi: [10.1186/1471-2105-10-290](https://doi.org/10.1186/1471-2105-10-290)
16. Liu W, Wu A, Pellegrini M, Wang X (2015) Integrative analysis of human protein, function and disease networks. *Sci Rep.* 5:14344. doi: [10.1038/srep14344](https://doi.org/10.1038/srep14344)
17. Singh M, Singh G (2011) Cluster analysis technique based on bipartite graph for human protein class prediction. *Int J Comput Appl.* 20:22–27. doi: [10.5120/2414-3226](https://doi.org/10.5120/2414-3226)
18. Pan A, Lahiri C, Rajendiran A, Shanmugham B (2016) Computational analysis of protein interaction networks for infectious diseases. *Brief Bioinform.* 17:517–526.
19. Lei X, Yang X (2018) A new method for predicting essential proteins based on participation degree in protein complex and subgraph density. *PLOS ONE.* 13:e0198998. doi: [10.1371/journal.pone.0198998](https://doi.org/10.1371/journal.pone.0198998)
20. Jolliffe IT, Cadima J (2016) Principal component analysis: A review and recent developments. *Philos Trans R Soc A.* 374:1–16. doi: [10.1098/rsta.2015.0202](https://doi.org/10.1098/rsta.2015.0202)
21. Nepusz T, Yu H, Paccanaro A (2012) Detecting overlapping protein complexes in human protein–protein interaction networks. *Nat Methods.* 9:471–472. doi: [10.1038/nmeth.1938](https://doi.org/10.1038/nmeth.1938)
22. Usman MS, Kusuma WA, Afendi FM, Heryanto R (2019) Identification of significant proteins associated with diabetes mellitus using network analysis of protein–protein interactions. *Comput Eng Appl J.* 8:41–52.
23. Zulfikar AA, Diansyah MR, Rizqa AA, Kusuma WA (2018) Identification of significant protein Diabetes Mellitus Type 2 with Fuzzy C- Means and topological analysis. 2018 *1st Int Conf Bioinform Biotechnol Biomed Eng.* 1:1–5.
24. Diansyah MR, Kusuma WA, Annisa A (2019) Analysis of protein–protein interaction using skyline query on parkinson disease. 2019 *Int Conf Adv Comput Sci Inform Syst.* 2019:175–180. doi: [10.1109/ICACSIIS47736.2019.8979892](https://doi.org/10.1109/ICACSIIS47736.2019.8979892)
25. Rezaei-Tavirani M, Rezaei-Tavirani S, Mansouri V, Rostami-Nejad M, Rezaei-Tavirani M (2017) Protein-protein interaction network analysis for a biomarker panel related to human esophageal adenocarcinoma. *Asian Pac J Cancer Prev.* 18:3357–3363. doi: [10.22034/APJCP.2017.18.12.3357](https://doi.org/10.22034/APJCP.2017.18.12.3357)
26. Khorsand B, Savadi A, Naghibzadeh M (2020) SARS-CoV-2-human protein-protein interaction network. *Informatics in Medicine Unlocked.* 20:100413. doi: [10.1016/j.imu.2020.100413](https://doi.org/10.1016/j.imu.2020.100413).
27. Kermali M, Khalsa RK, Pillai K, Ismail Z, Harky A (2020) The role of biomarkers in diagnosis of COVID-19–A systematic review. *Life Sci.* 254:1–9. doi: [10.1016/j.lfs.2020.117788](https://doi.org/10.1016/j.lfs.2020.117788).
28. Kuleshov M, Clarke DJB, Kropiwnicki E, Jagodnik K, Bartal A, Evangelista JE, et al. (2020) The COVID-19 gene and drug set library. *SSRN Electronic J.* 1:1–1.
29. Zhang L, Guo H (2020) Biomarkers of COVID-19 and technologies to combat SARS-CoV-2. *Adv Biomarker Sci Technol.* 2:1–23. doi: [10.1016/j.abst.2020.08.001](https://doi.org/10.1016/j.abst.2020.08.001).
30. Zhu M, Gao L, Li X, Liu Z, Xu C, Yan Y, et al. (2009) The analysis of the drug-targets based on the topological properties in the human protein-protein interaction network. *J Drug Target.* 17:524–532. doi: [10.1080/10611860903046610](https://doi.org/10.1080/10611860903046610).
31. Mistry D, Wise RP, Dickerson JA (2017) DiffSLC: A graph centrality method to detect essential proteins of a protein-protein interaction network. *PLOS ONE.* 12:e0187091. doi: [10.1371/journal.pone.0187091](https://doi.org/10.1371/journal.pone.0187091)
32. Ashtiani M, Salehzadeh-Yazdi A, Razaghi-Moghadam Z, Hennig H, Wolkenhauer O, Mirzaie M, et al. (2018) A systematic survey of centrality measures for protein-protein interaction networks. *BMC Syst Biol.* 12:1–17. doi: [10.1186/s12918-018-0598-2](https://doi.org/10.1186/s12918-018-0598-2)
33. Tang Y, Li M, Wang J, Pan Y, Wu FX (2015) CytoNCA: A cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *BioSystems.* 127:67–72. doi: [10.1016/j.biosystems.2014.11.005](https://doi.org/10.1016/j.biosystems.2014.11.005).
34. Wang J, Li M, Wang H, Pan Y (2012) Identification of essential proteins based on edge clustering coefficient. *IEEE/ACM Transact Comput Biol Bioinform.* 9:1070–1080. doi: [10.1109/TCBB.2011.147](https://doi.org/10.1109/TCBB.2011.147).
35. Dash P, Nayak M, Das GP (2014) Principal Component Analysis using singular value decomposition for image compression. *Int J Comput Appl.* 93:21–27.
36. Lee GS, Djauhari MA (2012) An overall centrality measure: The case of U.S. stock market. *Int J Basic Appl Sci.* 12:99–103.
37. Dosztányi Z, Chen J, Dunker, AK, Simon I, Tompa P (2006) Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res.* 5:2985–2995. doi: [10.1021/pr060171o](https://doi.org/10.1021/pr060171o).
38. Blanco-Melo D, Nilsson-Payant BE, Liu WC, Møller R, Panis M, Sachs D, et al. (2020) SARS-CoV-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems. *BioRxiv.* 1–21. doi: [10.1101/2020.03.24.004655](https://doi.org/10.1101/2020.03.24.004655)
39. Shaath H, Vishnubalaji R, Elkord E, Alajezi NM (2020) Single-cell transcriptome analysis highlights a role for neutrophils and inflammatory macrophages in the pathogenesis of severe COVID-19. *Cells.* 9:1–19. doi: [10.3390/cells9112374](https://doi.org/10.3390/cells9112374).
40. Prasad K, Khatoon F, Rashid S, Ali N, AlAsmari AF, Ahmed MZ, et al. (2020) Targeting hub genes and pathways of innate immune response in COVID-19: A network biology perspective. *Int. J. Biol Macromol.* 163:1–8. doi: [10.1016/j.ijbiomac.2020.06.228](https://doi.org/10.1016/j.ijbiomac.2020.06.228)
41. Hachim MY, Al Heialy S, Hachim IY, Halwani R, Senok AC, Maghazachi AA, et al. (2020) Interferon-induced transmembrane protein (IFITM3) is upregulated explicitly in SARS-CoV-2 infected lung epithelial cells. *Front Immunol.* 11:1–9. doi: [10.3389/fimmu.2020.01372](https://doi.org/10.3389/fimmu.2020.01372)
42. Zhang Q, Liu Z, Moncada-Velez M, Chen J, Ogishi M, Bigio B, et al. (2020). Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science.* 370: eabd4570.
43. Hadjadj J, Yatim N, Barnabei L, Corneau A, Boussier J, Smith N, et al. (2020) Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science.* 369:718–724. doi: [10.1126/science.abc6027](https://doi.org/10.1126/science.abc6027).
44. Zhang Z-L, Hou Y-L, Li D-T, Li F-Z (2020) Laboratory findings of COVID-19: a systematic review and meta-analysis. *Scand J Clin Lab Invest.* 80(6):441–447. doi: [10.1080/00365513.2020.1768587](https://doi.org/10.1080/00365513.2020.1768587).
45. Aziz M, Fatima R, Assaly R (2020) Elevated interleukin-6 and severe COVID-19: A meta-analysis. *J Med Virol.* 92(11):2283–2285.
46. Herold T, Jurinovic V, Arnreich C, Lipworth BJ, Hellmuth JC, von Bergwelt-Baildon M, et al. (2020) Elevated levels of IL-6 and CRP predict the need for mechanical ventilation in COVID-19. *J Allerg Clin Immunol.* 146(1):128–136.e4. doi: [10.1016/j.jaci.2020.05.008](https://doi.org/10.1016/j.jaci.2020.05.008)

47. Gupta R (2020) The double-edged interferon riddle in COVID-19 pathogenesis. *Crit Care*. 24:1–2. doi: 10.1186/s13054-020-03337-z.
48. Sharma P, Pandey AK, Bhattacharyya DK (2021) Determining crucial genes associated with COVID-19 based on COPD Findings. *Comput Biol Med*. 128:104126. doi: 10.1016/j.compbiomed.2020.104126.
49. Matsuyama T, Kubli SP, Yoshinaga SK, Pfeffer K, Mak TW (2020) An aberrant STAT pathway is central to COVID-19. *Cell Death Differ*. 27:3209–3225. doi: 10.1038/s41418-020-00633-7