BAREKENG: Journal of Mathematics and Its Application
December 2022 Volume 16 Issue 4 Page 1453–1464
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol16iss4pp1453-1464

BICLUSTERING APPLICATION IN INDONESIAN ECONOMIC AND PANDEMIC VULNERABILITY

Wiwik Andriyani Lestari Ningsih [1], I Made Sumertajaya[2*], Asep Saefuddin[3]

[1,2,3]Department of Statistics, IPB University
Jl. Raya Dramaga, Bogor, West Java, 16680, Indonesia

[1]BPS-Statistics Indonesia
Jl. Dr. Sutomo 6-8, Jakarta, 10710, Indonesia

Corresponding author's e-mail: [2] imsjaya@apps.ipb.ac.id*

Abstract. *Biclustering is an analytical tool to group data from two dimensions simultaneously. The analysis was first introduced by Hartigan (1972) and applied by Cheng and Church (2000) to the gene expression matrix. The Cheng and Church (CC) algorithm is a popular biclustering algorithm and has been widely applied outside the field of biological data in recent years. This algorithm application in economic and Covid-19 pandemic vulnerability cases is exciting and essential to do in order to get an overview of the spatial pattern and characteristics of the bicluster of economic and COVID-19 pandemic vulnerability in Indonesia. This study uses secondary data from some ministries. Forming a bicluster using the CC algorithm requires determining the delta threshold so that several types of delta thresholds are formed to choose the best (optimum) using the evaluation of the average value of mean square residue (MSR) to volume ratios. The similarity of the optimum bi-cluster with the other is also seen based on the Liu and Wang index values. The 0.01 delta threshold is chosen as the optimum threshold because it produces the smallest average value of MSR to volume ratios (0.00032). Based on Liu and Wang Index values, the optimum threshold has a similarity level below 50% with other types of delta thresholds, so the threshold is the best unique threshold. The optimum threshold resulted in six biclusters (six spatial patterns). Most regions in Indonesia (11 provinces) tend to have low economic and COVID-19 pandemic vulnerability in the first spatial pattern characteristic variables.*

Keywords: *algorithm, biclustering, Liu and Wang index, MSR, pattern detection.*

Article info:

Submitted: 11th August 2022 *Accepted: 4th November 2022*

How to cite this article:
W. A. L. Ningsih, I. M. Sumertajaya and A. Saefuddin, "BICLUSTERING APPLICATION IN INDONESIAN ECONOMIC AND PANDEMIC VULNERABILITY", *BAREKENG: J. Math. & App.*, vol. 16, iss. 4, pp. 1453-1464, Dec., 2022.

1. INTRODUCTION

Biclustering is an analytical tool to group data from two dimensions simultaneously. This analysis was first introduced by Hartigan [1] and applied by Cheng and Church [2] to the gene expression matrix. The application of biclustering was initially mostly carried out in biological data to analyze gene expression microarray data. However, the application of biclustering has become popular outside the field of biological data in recent years. One of the exciting areas to study in the application of biclustering is the economic field associated with the COVID-19 pandemic that occurred in early 2020.

The occurrence of the COVID-19 pandemic had a significant impact on various aspects, one of which was the unstable economy. Economic instability is indicated by Indonesia's economic condition in 2020, experiencing a contraction of 2.07 percent (c-to-c). This instability can potentially cause regional economic vulnerability and is not an escape from the COVID-19 pandemic. The reflection of covid-19 pandemic vulnerability is the statistics of covid-19 cases, which have increased and fluctuated throughout 2020. There were 22138 Indonesians who died as an impact of COVID-19 throughout 2020. The United Nations (UN) made a standard measure to measure economic vulnerability in the form of an index, namely the Economic Vulnerability Index (EVI) [3]. Meanwhile, the National Institute of Environmental Health Sciences (NIEHS) creates a risk profile in the form of a Pandemic Vulnerability Index (PVI) score to measure COVID-19 pandemic vulnerability [4].

Various algorithms can perform biclustering. One popular and widely applied algorithm is the Cheng and Church (CC). [5] and [6] applied the CC algorithm to poverty cases in East Java (38 regions × 15 variables) and Central Java (35 regions × 15 variables). In addition, [7] applies the CC algorithm to food insecurity cases with a data matrix measuring 34 regions × 12 variables. The CC algorithm application in this previous study only used the intra-bicluster evaluation function. However, this study added the use of the inter-bicluster evaluation function. On the other hand, the CC algorithm is also interesting to apply to economic and COVID-19 pandemic vulnerability cases. Therefore, the main objective of this study is to apply the CC biclustering algorithm to cases of economic and COVID-19 pandemic vulnerability in Indonesia. In addition, this study will examine the results of the CC algorithm biclustering in this case. In this case, the biclustering application will produce a spatial pattern of Indonesia's economic and COVID-19 pandemic vulnerability. This study's results expect to provide benefits for policymakers through the results of the detection of spatial patterns obtained.

2. RESEARCH METHODS

2.1 Data

This study uses secondary data from the BPS-Statistics Indonesia, the Ministry of Health, the Ministry of Village, the Development of Disadvantaged Regions and Transmigration, and the Ministry of Environment and Forestry. The observation unit in this study is the region (province) in Indonesia, with the variables that make up the EVI and PVI indicators based on 2020 data as presented in Table 1 (except for the X3 and X16 variables). Due to the limited data obtained, the X3 variable is based on 2018, and the X16 variable is based on 2019.

Table 1. Research Variables Composing EVI and PVI Indicators Based on Subcomponents

Component	Variable (notation)
EVI Indicator	
Size	Population size (X1)
Location	Remoteness and underdeveloped areas (X2)
Environment	Population of the coastal area (X3)
Economic structure	Export concentration (X4)
	Share of agriculture, forestry, and fisheries (Category A) in GRDP (X5)
Trade shocks	Instability in the export of goods and services (X6)
Natural shocks	Instability of agricultural production (X7)
	Natural disaster victims (X8)

Component	Variable (notation)
PVI Indicator	
Infectious cases	Covid-19 infectious cases (X9)
Spread of disease	Covid-19 death rate (X10)
Population mobility	Daytime population (X11)
	Traffic volume average (X12)
Housing density	Average of total household members (X13)
Testing	Covid-19 testing (X14)
Social distancing	Social distancing score (X15)
Air pollution	Air quality index (X16)
Age distribution	Population aged 65 years and over (X17)
Comorbidities including premature death, smoking, diabetes, and obesity	Morbidity rate (X18)
	Residents without insurance (X20)
Adult residents (20 years and over) smoker (X19)	
Health disparities	
	Poor population (X21)
	Open unemployment rate (X22)
Hospital beds	Availability of hospital beds (X23)

A brief explanation of some of the variables in Table 1 is as follows:

1. Remoteness and disadvantaged areas using an approach of the inverse value of the Developing Village Index (IDM);
2. The population of the coastal area is estimated using the proportion of the number of villages located by the sea multiplied by the total population;
3. Export concentration using the percentage of exports of agricultural, forestry, and fishery product types;
4. Instability in the export of goods and services using an approach of the inverse value of the ratio of the contribution of total exports in 2020 divided by the previous year;
5. Instability of agricultural production using an approach of the ratio of the Category A GRDP contribution in 2020 divided by the previous year;
6. The daytime population is estimated by multiplying the total population by the proportion of the number of passenger cars and motorcycles;
7. The social distance score uses an approach of the ratio of domestic tourists (number of people traveling other than for work or school) per resident.

The inverse value is used so that the condition of those variables is in line with the concept of vulnerability in the other variables; the higher the value of that variable, the more tend vulnerability will occur. However, non-vulnerable conditions tend to occur along with the lower of those variables.

2.2 Research Stages

This study has three general stages: preparation, biclustering, and evaluation and selection of the bicluster. The following is a more detailed explanation regarding each step carried out.

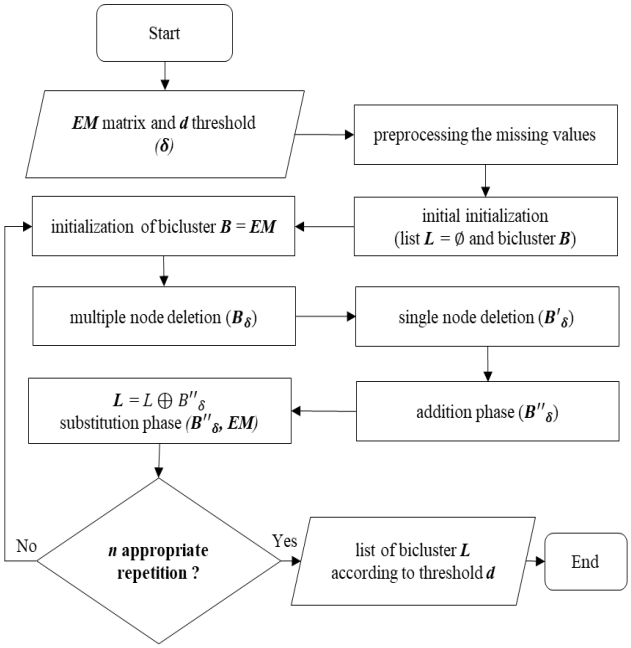
Stage 1 (Preparation)

1. Form a data matrix consisting of 34 provinces as rows and 23 EVI and PVI variables as columns,
2. Scale the data matrix using the standard normalization approach, and
3. Explore the data to see the initial characteristics using a heatmap.

Stage 2 (Biclustering)

Biclustering, in this study, uses the Cheng and Church (CC) algorithm. The classification of the CC algorithm is into an iterative greedy search group based on the evaluation size [8]. According to [9], the CC algorithm is a greedy algorithm that tries to find the maximum bicluster with high similarity. The purpose of the CC algorithm is to find biclusters with a Mean Square Residual/MSR which is smaller than a specified limit or tolerance value [10]. Therefore, it is necessary to determine the tolerance/threshold value before carrying out the biclustering process. This study tries several types of thresholds and will evaluate and choose a threshold that gives optimal bicluster results.

The CC algorithm illustration is as the flow chart in Figure 1. In general, through Figure 1, there are deletion, addition, and substitution phases [8]. Deletion and addition phases are iterative, ensuring that the MSR is greater than a predetermined threshold value. Meanwhile, the carried out of substitution phase is to prevent overlapping between the resulting biclusters.



Notation description:

J	:	column (variable)
I	:	row (variable)
EM	:	a matrix of size $I \times J$
δ	:	threshold (maximal mean square residue score)
$b_{Ij} = \frac{\sum_{i \in I} b_{ij}}{\lvert I \rvert}$	:	average in row
$b_{iJ} = \frac{\sum_{j \in J} b_{ij}}{\lvert J \rvert}$	:	average in column
$b_{IJ} = \frac{\sum_{i \in I, j \in J} b_{ij}}{\lvert I \rvert \lvert J \rvert}$	:	average across the bi-clusters
$RS_{IJ}(i,j) = b_{ij} - b_{Ij} - b_{iJ} + b_{IJ}$	:	residual score of the elements in the matrix
$H(I,J) = \sum_{i \in I, j \in J} \frac{RS_{ij}^2}{\lvert I \rvert \lvert J \rvert}$	:	mean square residu (MSR)
$d(i) = \frac{1}{\lvert J \rvert} \sum_{j \in J} RS_{ij}^2$	:	square residual score value from row
$d(j) = \frac{1}{\lvert I \rvert} \sum_{i \in I} RS_{ij}^2$	:	square residual score value from column

Figure 1. Cheng and Church Algorithm Flowchart (Modified from [8]**)**

The following is a description of each stage of the CC algorithm flowchart in Figure 1:

1. Preprocessing missing values → when there is a missing value in the matrix, the filling is with a random number;
2. Initial initialization (L = ∅ and bicluster **B**) → sets the initial conditions for the list $\boldsymbol{L}$ as an empty set and a bicluster denoted by $\boldsymbol{B}$;
3. Repetition of *n* times → find as many as n appropriate biclusters based on the threshold that has been determined;
4. Initialization of bicluster $\boldsymbol{B = EM}$ → sets the initial conditions for B bicluster, which is the initial matrix of the input data;
5. Multiple node deletion ($\boldsymbol{B_\delta}$) → deletes rows based on condition $d(i) > \alpha H(I,J)$ and deletes columns based on condition $d(j) > \alpha H(I,J)$, as long as it satisfies the condition $H(I,J) > \delta$;
6. Single node deletion ($\boldsymbol{B'_\delta}$) → deletes rows or columns based on the maximum condition of $d(i)$ or $d(j)$, as long as it satisfies the condition $H(I,J) > \delta$. Deletion of rows/columns aims to reduce MSR;
7. Addition phase ($\boldsymbol{B''_\delta}$) → adding columns based on condition $d(j) \leq H(I,J)$ and adding rows based on condition $d(i) \leq H(I,J)$, as long as it fulfills the MSR condition of output addition phase $\leq H(I,J)$;
8. $\boldsymbol{L} = L \oplus B''_\delta$ → sets the conditions for the list $\boldsymbol{L}$ as the bicluster result of the node addition process output. $L \oplus B''_\delta = (L \cup B''_\delta) \setminus (L \cap B''_\delta)$ → addition without any duplicates;
9. Substitution phase ($\boldsymbol{B''_\delta, EM}$) → replaces the elements in the input matrix contained in the newly discovered bicluster with a random number. It aims to prevent overlap between biclusters.

Stage 3 (Bicluster evaluation and selection)

Bicluster formed at each threshold type is evaluated using the average value of MSR to volume ratios. Select one type of threshold for further analysis, the characteristics of each bicluster. The spatial pattern of economic and COVID-19 pandemic vulnerability from the selected/optimum CC threshold results can then be described as a map. In addition, it is also seen how well each bicluster group at the optimum threshold has similarities with the bicluster group at other threshold types using the Liu and Wang index values.

According to [11], [12] the performance of the biclustering algorithm can be evaluated using the Mean Squared Residue (MSR) defined by Equation 1,

$$MSR_{(I,J)} = \frac{\sum_{i \in I} \sum_{j \in J} (b_{ij} - b_{iJ} - b_{Ij} + b_{IJ})^2}{I \times J} \quad (1)$$

where b_{IJ} is the average across the bi-clusters, b_{Ij} is the average in column j, b_{iJ} is the average in row i, $I \times J$ is the dimension of the bi-cluster, i.e., the size of the bi-cluster row (I) times the size of the bi-cluster column (J). $MSR_{(I,J)}$ represents the variation associated with the interaction between rows and columns in a bi-cluster [13]. According to [14], the quality of a bicluster will be better as the residual value (MSR) decreases and/or the volume (rows × column) of the bicluster increases and it is obtained by calculating the average value of MSR to volume ratios $\left(\frac{1}{n}\sum_{i=1}^{n}\frac{MSR_i}{Volume_i}\right)$. In addition, to see how well each optimal bicluster group (M_{opt}) will have similarities with every other bicluster group (M), we can use the size of Liu and Wang index defined by Equation 2 [15],

$$I_{Liu\&Wang}(M_{opt}, M) = \frac{1}{K_{opt}} \sum_{i=1}^{K_{\text{opt}}} \max\left(\frac{[G_i \cap G_j] + [C_i \cap C_j]}{[G_i \cup G_j] + [C_i \cup C_j]}\right) \quad (2)$$

where M_{opt} is defined as the bi-cluster group that has the smallest average value of mean square residue to volume ratios, K_{opt} is the number of bi-clusters in M_{opt}, $[G_i \cap G_j]$ is the number of rows (G) in M_{opt} that intersect with rows in M, $[C_i \cap C_j]$ is the number of columns (C) in M_{opt} that intersect with columns in M, $[G_i \cup G_j]$ is the number of combined rows (G) of $M_{opt}$and M, and $[C_i \cup C_j]$ is the number of combined columns (C) of M_{opt} and M.

3. RESULTS AND DISCUSSION

3.1. Data Exploration

The initial picture of the data used in this study is illustrated through the heatmap in Figure 2. The data in the image has gone through a scaling process using the standard normalization approach. It can be seen from the heatmap that there are several data variables for EVI (X1 to X8) and PVI (X7 to X23) in a particular province whose values tend to be highly positive (dark blue) and extremely negative (broken white). The province with a value that tends to be highly positive indicates that the province in the relevant variable tends to be vulnerable. On the other hand, the province with a value that tends to be highly damaging indicates that the province in the relevant variable tends to be less vulnerable (non-vulnerable). For example, in the PVI X12 variable (traffic volume average), DKI Jakarta Province has an extremely positive value (dark blue), indicating that DKI Jakarta Province tends to have a high COVID-19 pandemic vulnerability, especially in the traffic volume average indicator. Another example is the EVI X2 variable (remoteness and underdeveloped areas). DKI Jakarta Province has a highly negative value (broken white) which indicates that DKI Jakarta Province tends to have a low economic vulnerability, especially on indicators of remoteness and disadvantaged areas.

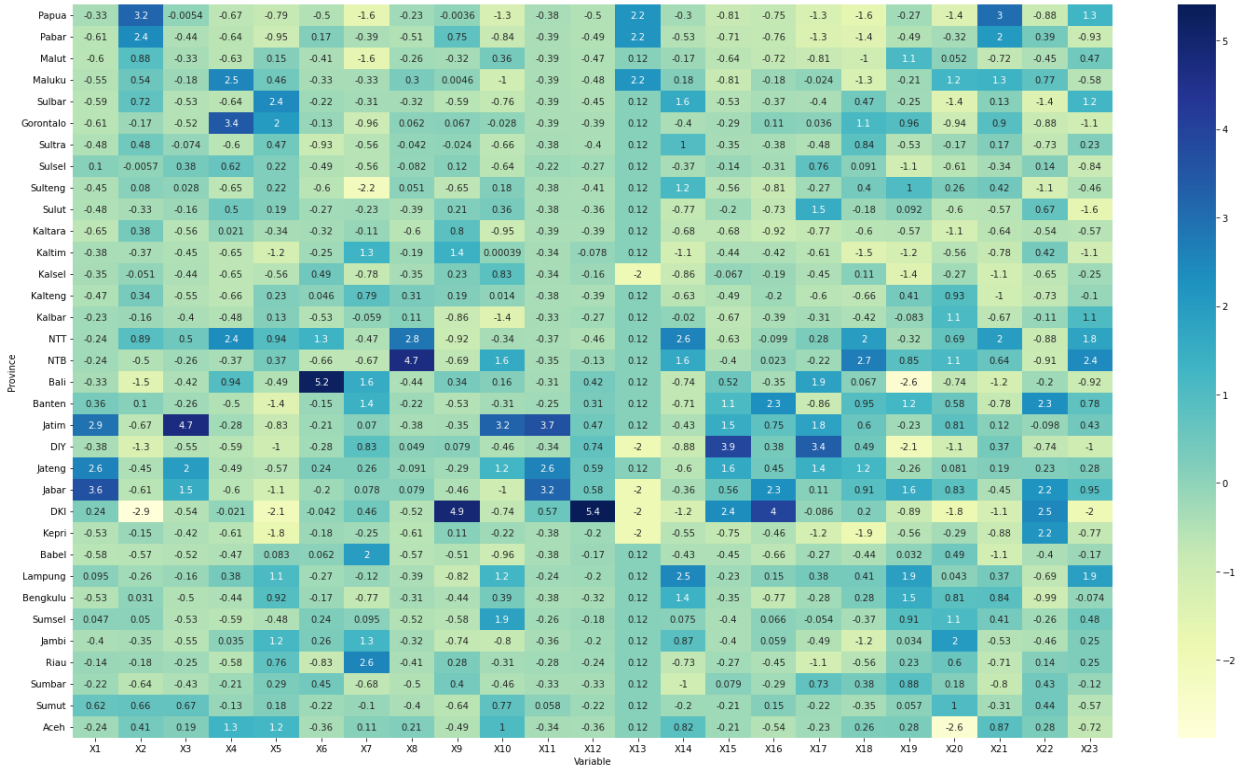


Figure 2. Heatmap Matriks Data *Scaling*

3.2. Biclustering Results

Bicluster analysis using the CC algorithm requires determining the delta threshold value. In this study, several thresholds were tried based on the MSR value of the scaled data matrix, namely 0.91. The initial test to determine the delta threshold is to use a range of values below 0.91 with multiples of 0.1 (0.1 to 0.9), and the number of biclusters obtained is two, which are at the delta threshold of 0.6 and 0.7. The trial was continued by looking at the number of biclusters formed in the delta threshold range below 0.7 with 0.01 sequences (0.01 to 0.70), and the number of biclusters tended to be homogeneous. Therefore, this study uses a value of 0.01 to 0.77 (0.02 sequences) so that there are 39 types of the delta threshold. Each type of threshold will produce an average value of MSR to volume ratios and the number of biclusters, as depicted in Figure 3.

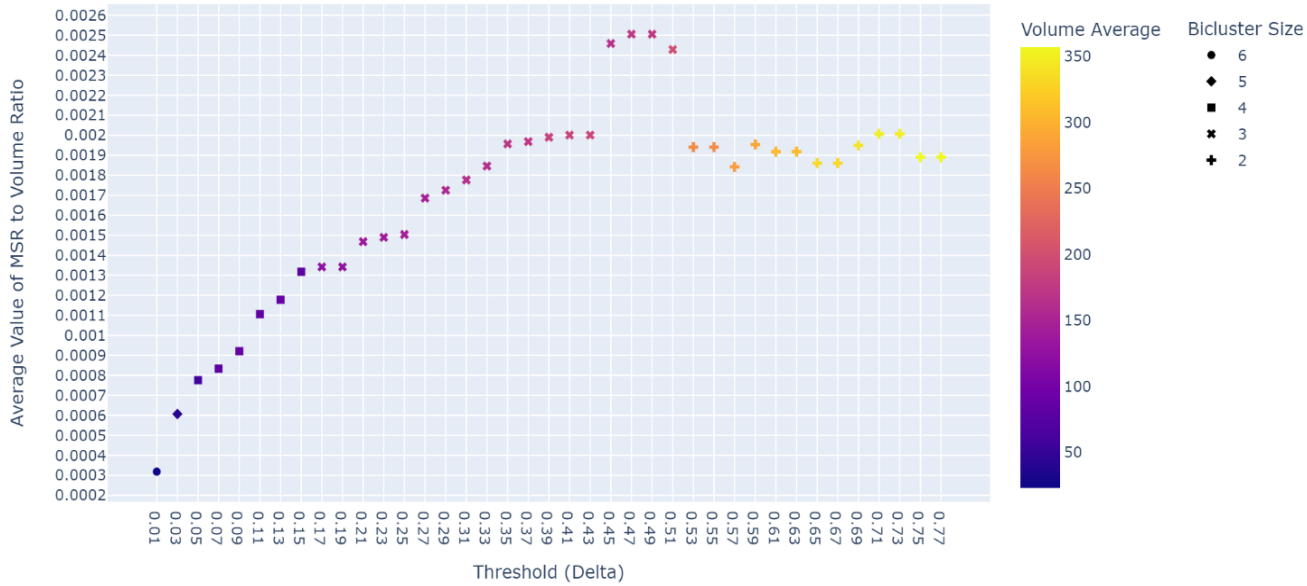


Figure 3. Heatmap of the Average Value of MSR to Volume Ratio and the Number of Biclusters According to Delta Threshold

It can be seen from Figure 3 that the 0.01 delta threshold has the smallest average value of MSR to volume ratio (0.00032), with the number of biclusters formed being 6. Meanwhile, the 0.49 and 0.47 delta thresholds have the highest average value of the MSR to volume ratio (0.00250), with the number of biclusters formed being 3, respectively. In general, it can be seen from the figure that the average value of the MSR to volume ratio tends to be worth higher (towards a dark blue color) when the delta thresholds are getting higher. The larger the average value of the MSR to volume ratio, the lower the quality of the bicluster formed. Therefore, the 0.01 delta threshold is chosen from the existing threshold types as the optimal threshold. Each bicluster of the selected delta threshold type is then analyzed for its characteristics.

Before further analyzing the characteristics of the selected delta threshold, it is necessary to identify the similarity of bicluster's goodness with the other biclusters in the other threshold type. The identification uses the Liu and Wang index values depicted in Figure 4.

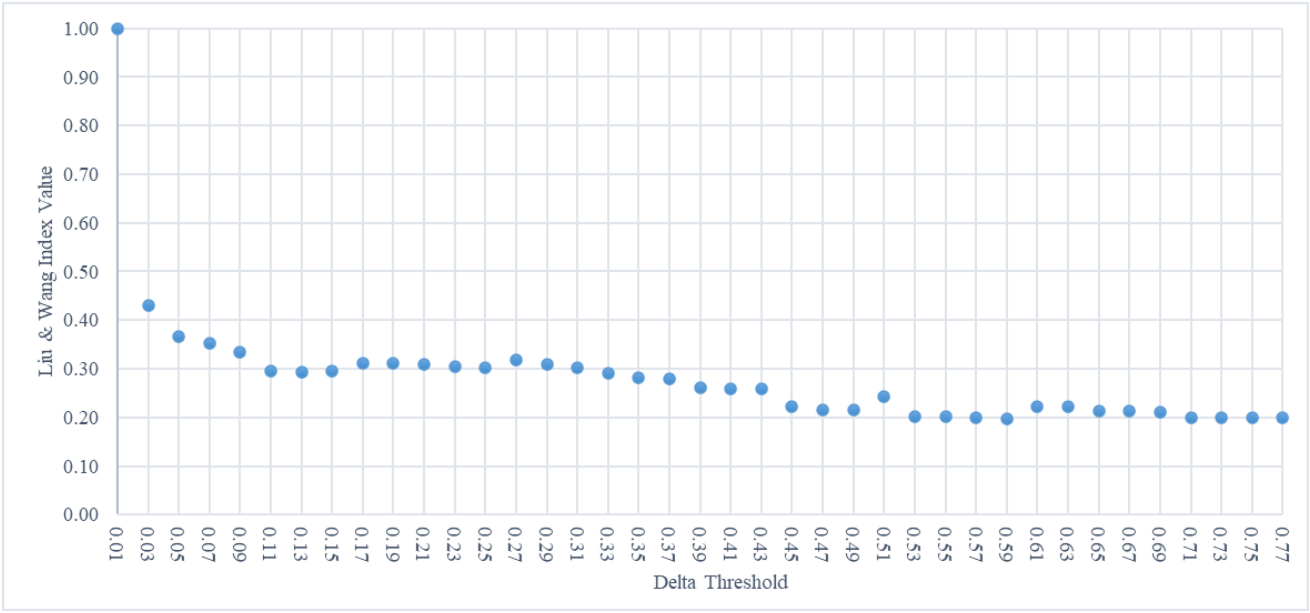


Figure 4. Scatter of Liu and Wang Index Values according to Delta Thresholds

The 1.00 value of the Liu and Wang index at the 0.01 delta threshold indicates that the bicluster group at that threshold is the optimal bi-cluster group (having the smallest average MSR to volume ratio). The Liu and Wang index values, closer to 1, indicate that the bicluster group at the relevant threshold has a higher level of similarity with the optimal bicluster group. It can be seen from Figure 4 that the bicluster group at the selected (optimal) threshold has a similarity level below 50% with the bicluster group at another threshold. It shows that the selected optimal threshold is the best unique threshold so that no other one can replace or have the same as the optimal one.

3.3. Optimal Biclustering Results

The membership of the optimal CC biclustering results (0.01 delta threshold) according to the bicluster formed in Table 2. It can be seen from the table that the membership of the region/province between the six biclusters formed does not have an intersection with each other (not overlapping). However, the membership of the variables has an overlap. For example, variable X1 (population size) is the membership of variables in biclusters 1, 3, and 5. Likewise, variables X11 (daytime population), X12 (traffic volume average), and X15 (social distancing score) are membership variables in bicluster 1, 2, and 3. In addition, not all of this study's variables of the biclustering variables membership are characteristic of the formed bicluster. Twelve variables are not members of the variables in the biclustering results: two EVI variables (X5 and X7) and ten PVI variables (X9, X10, X16, X17, X18, X19, X20, X21, X22, and X23).

Table 2. Membership of Optimal CC Biclustering Results According to Bicluster

Bicluster	Bicluster Size	Membership	
		Province (code)	Variable
1	11 × 6	14. 15. 17. 19. 61. 62. 65. 74. 75. 76. 82	X1. X3. X11. X12. X13. X15
2	5 × 5	16. 18. 63. 71. 73	X2. X8. X11. X12. X15
3	8 × 4	11. 21. 52. 53. 72. 81. 91. 94	X1. X11. X12. X15
4	3 × 4	32. 33. 36	X4. X8. X12. X14
5	4 × 3	13. 34. 51. 64	X1. X11. X14
6	3 × 3	12. 31. 35	X4. X6. X8

The existence of overlap in the variables membership between the six biclusters formed does not result in detecting different types of spatial patterns. However, each type of spatial pattern's characteristics certainly differ. The spatial pattern of the biclustering results is depicted as a map. as shown in Figure 5.

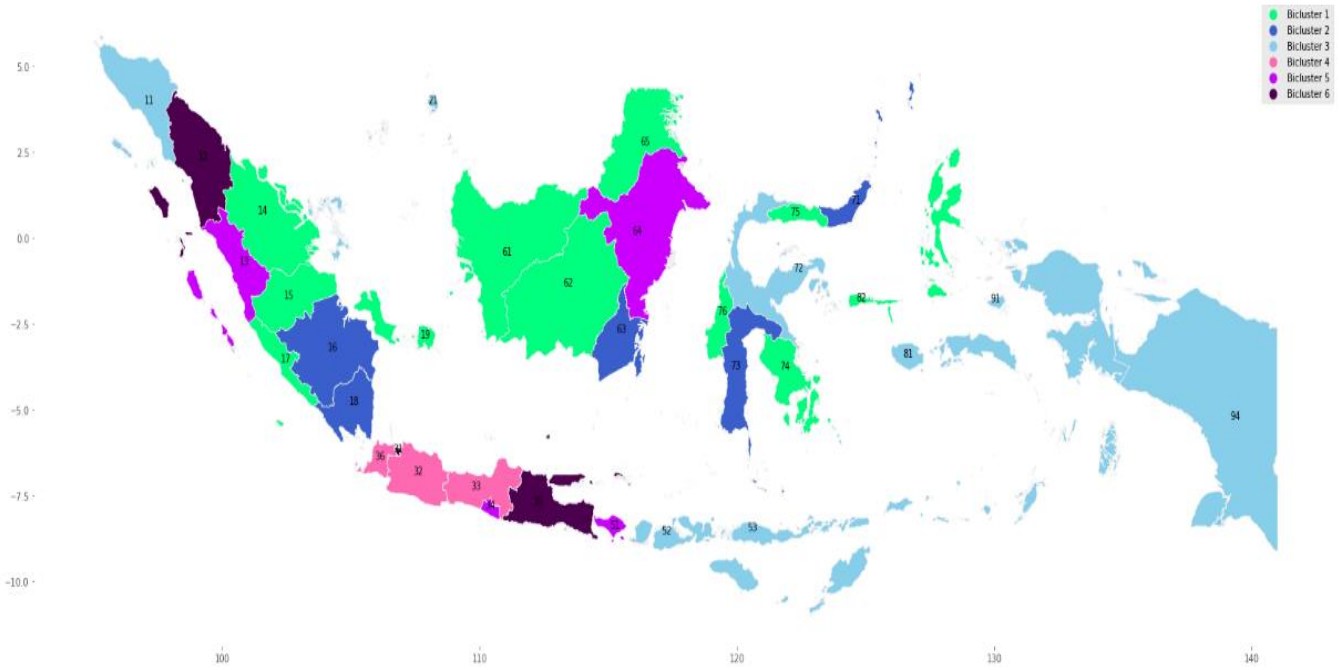


Figure 5. Map of Optimal CC Biclustering Results According to Spatial Pattern Type

Through the picture, there are six types of spatial patterns formed according to the number of biclusters produced. First, the light green area (bicluster 1), namely the provinces coded 14 (Riau), 15 (Jambi), 17 (Bengkulu), 19 (Bangka Belitung), 61 (Kalimantan Barat), 62 (Kalimantan Tengah), 65 (Kalimantan Utara), 74 (Sulawesi Tenggara), 75 (Gorontalo), 76 (Sulawesi Barat), and 82 (Maluku Utara). Second, the dark blue area (bicluster 2), namely the provinces coded 16 (Sumatera Selatan), 18 (Lampung), 63 (Kalimantan Selatan), 71 (Sulawesi Utara), and 73 (Sulawesi Selatan). Third, the areas that are light blue (bicluster 3), namely the provinces coded 11 (Aceh), 21 (Kepulauan Riau), 52 (NTB), 53 (NTT), 72 (Sulawesi Tengah), 81 (Maluku), 91 (Papua Barat), and 94 (Papua). Fourth, the areas that are colored pink (bicluster 4), namely the provinces coded 32 (Jawa Barat), 33 (Jawa Tengah), and 36 (Banten). Fifth, the areas that are colored purple (bicluster 5), namely the provinces coded 13 (Sumatera Barat), 34 (DIY), 51 (Bali), and 64 (Kalimantan Timur). Sixth, the areas that are dark purple (bicluster 6), namely the provinces coded 12 (Sumatera Utara), 31 (DKI Jakarta), and 35 (Jawa Timur).

Besides identifying spatial patterns through maps such as Figure 5, the variable characteristics for each type of spatial pattern can also be identified and described in Figures 6 to 12 as a bar chart. The bar chart illustrates the average value for each variable included in each type of spatial pattern (characteristic variable). For example, for the first spatial pattern (bicluster 1), it can be seen in Figure 6 that only six are characteristic variables in the area included in this spatial pattern.

Figure 6. First Spatial Pattern Characteristics Variable (Bicluster 1) of Optimal Threshold in CC Algorithm Results

Most of the variables, two EVI variables (X1 and X3) and three PVI variables (X11, X12, and X15) tend to be low vulnerable. In contrast. only one PVI variable X13 tends to be highly vulnerable. It indicates that the Riau, Jambi, Bengkulu, Bangka Belitung, Kalimantan Barat, Kalimantan Tengah, Kalimantan Utara, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, and Maluku Utara regions tend to have a low economic vulnerability on indicators X1 (population size) and X3 (population of the coastal area) and tend to have low pandemic vulnerability on indicators X11 (daytime population), X12 (traffic volume average), and X15 (social distancing score). However, the region also has a high pandemic vulnerability in the X13 indicator (average of total household members).

The characteristics of the variables for the second spatial pattern depicted in Figure 7 show that only five are characteristic variables in the area included in this spatial pattern.

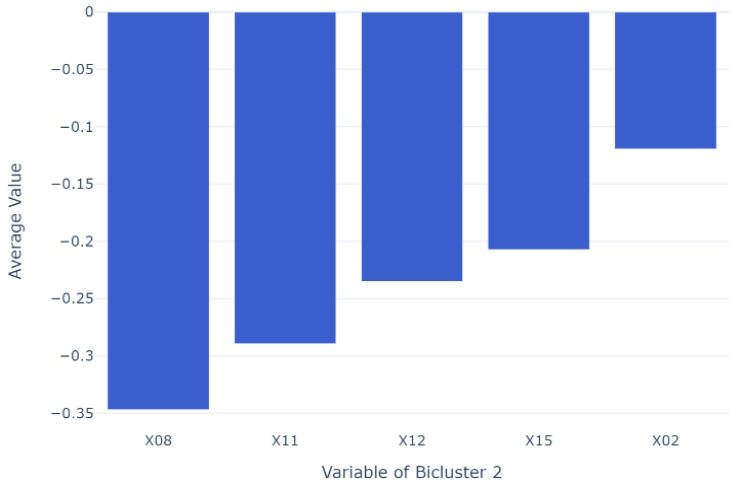


Figure 7. Second Spatial Pattern Characteristics Variable (Bicluster 2) of Optimal Threshold in CC Algorithm Results

It can be seen from Figure 7 that all of these characteristic variables tend to have low vulnerability in terms of the economy and the COVID-19 pandemic. which are two EVI indicators (X2 and X8) and three PVI indicators (X11, X12, and X15). It indicates that the regions of Sumatera Selatan, Lampung, Kalimantan Selatan, Sulawesi Utara, and Sulawesi Selatan tend to have a low economic vulnerability on indicators X2 (remoteness and underdeveloped areas) and X8 (natural disaster victims). The region also tends to have a low pandemic vulnerability on indicators X11 (daytime population), X12 (traffic volume average), and X15 (social distancing score).

We can see a pattern almost the same as the second spatial pattern in the third spatial pattern depicted in Figure 8. It can be seen from the figure that all the characteristic variables in the third spatial pattern tend to have low vulnerability from the economic and covid-19 pandemic side. which are one indicator EVI (X1) and three PVI indicators (X11. X12. and X15). It indicates that the Aceh. Kepulauan Riau. NTB. NTT. Sulawesi Tengah. Maluku. Papua Barat. and Papua regions tend to have low economic vulnerability in the

X1 indicator (population size). The region also tends to have low pandemic vulnerability in the X11 indicator (daytime population). X12 (traffic volume average). and X15 (social distancing score).

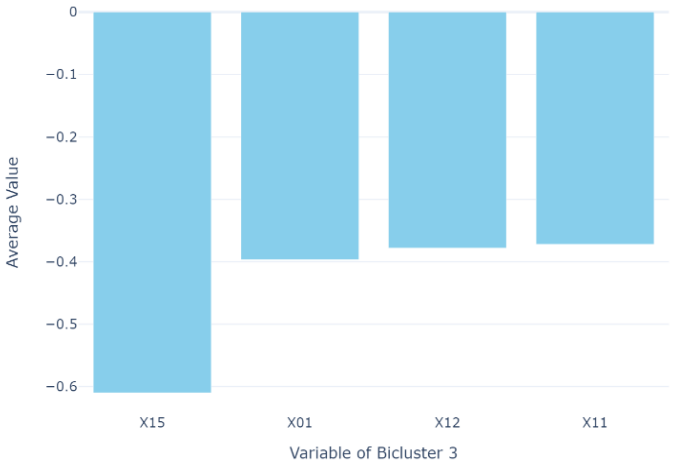


Figure 8. Third Spatial Pattern Characteristics Variable (Bicluster 3) of Optimal Threshold in CC Algorithm Results

Figure 9 illustrates the characteristics of the variables for the fourth spatial pattern. It can be seen from the figure that there are only four variables that describe the characteristics of this spatial pattern that are two EVI variables (X4 and X8) and two PVI variables (X12 and X14). Only one of the four characteristic variables in this spatial pattern tends to have a high vulnerability on the COVID-19 pandemic side. namely the X12 variable. It indicates that the regions of Jawa Barat, Jawa Tengah, and Banten tend to have a high pandemic vulnerability on the X12 indicator (traffic volume average). Meanwhile, the other three characteristic variables, namely the X4, X8, and X14, tend to have low vulnerability in the economy and the COVID-19 pandemic. It indicates that the Jawa Barat, Jawa Tengah, and Banten regions tend to have low economic vulnerability in the X4 (export concentration) and X8 (natural disaster victims) indicators and tend to have low pandemic vulnerability in the X14 indicator (covid-19 testing).

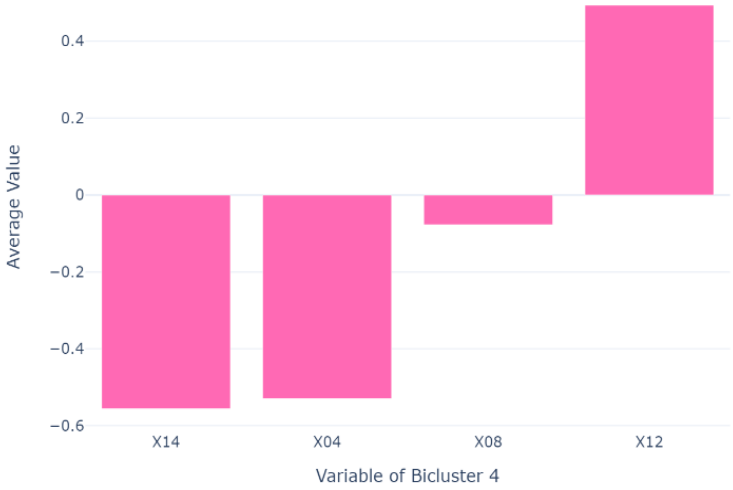


Figure 9. Fourth Spatial Pattern Characteristics Variable (Bicluster 4) of Optimal Threshold in CC Algorithm Results

Similar to the second and third spatial patterns. the fifth spatial pattern depicted in Figure 10 shows that all of the characteristic variables tend to have low vulnerability. Only three are characteristic variables in this fifth spatial pattern: one EVI variable (X1) and two PVI variables (X11 and X14). It indicates that the regions of Sumatera Barat, DIY, Bali, and Kalimantan Timur tend to have a low economic vulnerability on the X1 indicator (population size). In addition, the region also tends to have a low pandemic vulnerability on indicators X11 (daytime population) and X14 (covid-19 testing).

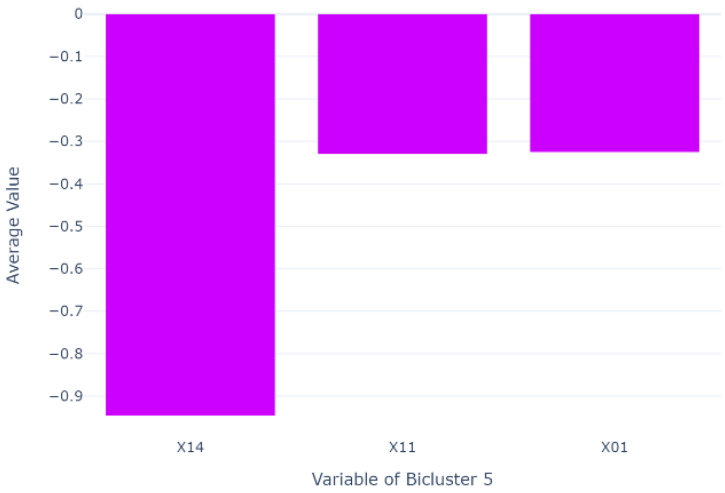


Figure 10. Fifth Spatial Pattern Characteristics Variable (Bicluster 5) of Optimal Threshold in CC Algorithm Results

A pattern is almost the same as the fifth spatial pattern. seen in the sixth spatial pattern depicted in Figure 11. It can be seen from the figure that all the characteristic variables in the sixth spatial pattern tend to have low vulnerability only from an economic perspective. which are three EVI variables (X4, X6, and X8). It indicates that the regions of Sumatera Utara, DKI Jakarta, and Jawa Timur tend to have a low economic vulnerability on indicators X4 (export concentration), X6 (instability in the export of goods and services), and X8 (natural disaster victims).

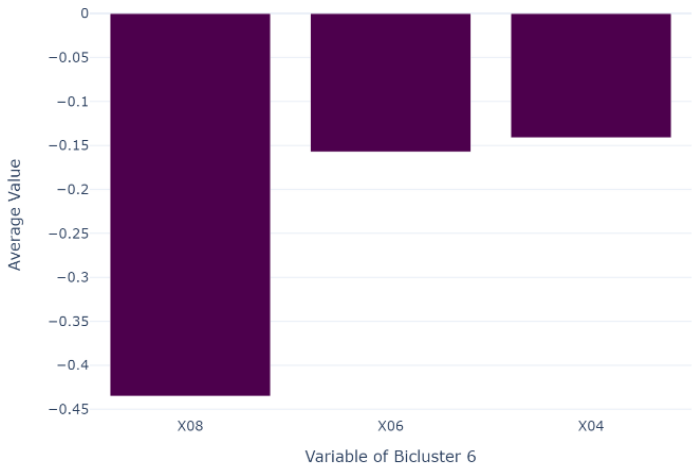


Figure 11. Sixth Spatial Pattern Characteristics Variable (Bicluster 6) of Optimal Threshold in CC Algorithm Results

The optimal threshold of CC biclustering algorithm results shows that most areas in Indonesia are in the first type of spatial pattern with the characteristics variables of EVI and PVI, most of which tend to have low vulnerability. In addition, most areas on Java Island are in the fourth type of spatial pattern with the characteristic variables of EVI and PVI, most of which tend to have low vulnerability. It indicates that most regions in Indonesia generally tend to have low economic and COVID-19 pandemic vulnerability in the first spatial pattern characteristic variables. In general, most areas on Java Island also tend to have low economic and COVID-19 pandemic vulnerability in the fourth spatial pattern characteristic variable.

4. CONCLUSIONS

This study has applied biclustering to economic and COVID-19 pandemic vulnerability cases in Indonesia using the Cheng and Church (CC) algorithm. In this case. the results of CC biclustering use a 0.01 delta threshold as the optimal threshold with the smallest average MSR to volume ratio value (0.00032). The optimal threshold has a similarity level below 50% with other types of delta thresholds (0.03 delta to on). The selected optimal threshold is the best unique threshold so that no other threshold can replace or have a similarity with the optimal threshold. The number of biclusters formed with the optimal threshold is six so there are six types of spatial patterns and different characteristics. Most regions in Indonesia, in general tend to have low economic and COVID-19 pandemic vulnerability in the first spatial pattern characteristic variables. In addition, most areas on Java Island also tend to have low economic and COVID-19 pandemic vulnerability in the fourth spatial pattern characteristic variables.

AKNOWLEDGEMENT

The authors thank BPS-Statistics Indonesia education and training center (Pusdiklat BPS) for funding this research.

REFERENCES

[1] J. A. Hartigan, "Direct clustering of a data matrix," *J. Am. Stat. Assoc.*, vol. 67, no. 337, pp. 123–129, 1972, doi: 10.1080/01621459.1972.10481214.
[2] Y. Cheng and G. M. Church, "Biclustering of expression data.," *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, vol. 8, pp. 93–103, 2000.
[3] United Nations, "EVI Indicators," 2011. ipb.link/un-evi (accessed Apr. 27, 2021).
[4] National Institute of Environmental Health Sciences, "Details for PVI Maps," 2020. ipb.link/niehs (accessed Apr. 27, 2021).
[5] B. Yuniarto and R. Kurniawan, "Understanding Structure of Poverty Dimensions in East Java: Bicluster Approach," *Signifikan J. Ilmu Ekon.*, vol. 6, no. 2, pp. 289–300, 2017, doi: 10.15408/sjie.v6i2.4769.
[6] C. A. Putri, R. Irfani, and B. Sartono, "Recognizing poverty pattern in Central Java using Biclustering Analysis," *J. Phys. Conf. Ser.*, vol. 1863, no. 1, 2021, doi: 10.1088/1742-6596/1863/1/012068.
[7] R. Novidianto and R. Irfani, "Bicluster CC Algoritm Analysis to Identify Patterns of Food Insecurity in Indonesia," *J. Mat. Stat. dan Komputasi*, vol. 17, no. 2, pp. 325–338, 2020, doi: 10.20956/jmsk.v17i2.12057.
[8] B. Pontes, R. Giráldez, and J. S. Aguilar-Ruiz, "Biclustering on expression data: A review," *J. Biomed. Inform.*, vol. 57, pp. 163–180, 2015, doi: 10.1016/j.jbi.2015.06.028.
[9] Nurmawiya and R. Kurniawan, "Pengelompokan Wilayah Indonesia Dalam Menghadapi Revolusi Industri 4.0 Dengan Metode Biclustering," pp. 790–797, 2020.
[10] P. A. Kaban, R. Kurniawan, R. E. Caraka, B. Pardamean, B. Yuniarto, and Sukim, "Biclustering method to capture the spatial pattern and to identify the causes of social vulnerability in Indonesia: A new recommendation for disaster mitigation policy," *Procedia Comput. Sci.*, vol. 157, pp. 31–37, 2019, doi: 10.1016/j.procs.2019.08.138.
[11] H. Cho and I. S. Dhillon, "Coclustering of human cancer microarrays using minimum sum-squared residue coclustering," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 5, no. 3, pp. 385–400, 2008, doi: 10.1109/TCBB.2007.70268.
[12] H. Ben Saber and M. Elloumi, "A Comparative Study of Clustering and Biclustering of Microarray Data[," *Int. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 6, pp. 93–111, 2014, doi: 10.5121/ijcsit.2014.6607.
[13] N. Kavitha Sri and R. Porkodi, "An extensive survey on biclustering approaches and algorithms for gene expression data," *Int. J. Sci. Technol. Res.*, vol. 8, no. 9, pp. 2228–2236, 2019.
[14] A. Chakraborty and H. Maka, "Biclustering of gene expression data using genetic algorithm," *Proc. 2005 IEEE Symp. Comput. Intell. Bioinforma. Comput. Biol. CIBCB '05*, vol. 2005, no. 2000, 2005, doi: 10.1109/cibcb.2005.1594893.
[15] X. Liu and L. Wang, "Computing the maximum similarity bi-clusters of gene expression data," *Bioinformatics*, vol. 23, no. 1, pp. 50–56, 2007, doi: 10.1093/bioinformatics/btl560.