J. Ris. & Ap. Mat. Vol. 03 No. 01 (2019) pp. 23-31
Jurnal Riset dan Aplikasi Matematika
e-ISSN: 2581-0154
URL: journal.unesa.ac.id/index.php/jram

PREDICTING FIGURE COALITION FOR 2019 INDONESIAN PRESIDENTIAL ELECTION USING MODIFIED MARKOV CLUSTERING ALGORITHM

ABSTRACT

This paper presents a new approach to ameliorate the Markov Cluster algorithm for predicting figure coalition for 2019 Indonesian Presidential Election. The proposed method is the modification of the Markov Clustering algorithm. First, 20 figures are collected to form a 20 x 20 matrix. Second, the entries of the matrix are scored by 0, 1, 2, or 3 concerning the number of positive comments from netizen towards the observed figures photo on Instagram. Third, we implemented the Markov Clustering to find the clusters that represent the number of coalitions. The MCL method is used in this research because the algorithm can be used to clustering large data with a high level of sparsity. The effectiveness of the proposed method is confirmed by comparing the prediction results with the actual coalition. The result is Markov Clustering method can be used to solve problems in the fields of politics with detail first coalition consist of seven members with the center is PRO and the other coalition, consist of six members with the center is JKW.

Keywords: Modified Markov Clustering, Coalition, Indonesian Presidential Election

ABSTRAK

Makalah ini menyajikan pendekatan baru untuk mempertajam akurasi dari algoritma Markov Clustering untuk memprediksi koalisi figur yang terbentuk dalam Pemilihan umum Presiden Indonesia 2019. Metode yang diusulkan adalah modifikasi dari algoritma Markov Clustering. Pertama, 20 figur dikumpulkan untuk membentuk matriks 20 x 20. Kedua, entri-entri dari matriks diberi nilai 0, 1, 2, atau 3 dengan mempertimbangkan jumlah komentar positif dari *netizen* terhadap foto figur yang diamati di Instagram. Ketiga, Modifikasi dari algoritma Markov Clustering diimplementasikan untuk menemukan kelompok yang mewakili jumlah koalisi figur. Modifikasi dari algoritma Markov Clustering digunakan dalam penelitian ini karena algoritma ini dapat digunakan untuk mengelompokkan data berukuran besar dengan tingkat *sparsity* yang tinggi. Keefektifan dari algoritma ini dikonfirmasi dengan membandingkan hasil prediksi koaliasi yang terbentuk dengan koalisi yang sebenarnya (aktual). Hasilnya menunjukkan bahwa modifikasi dari algoritma Markov Clustering dapat digunakan untuk menyelesaikan masalah di bidang politik, dengan koalisi pertama terdiri dari tujuh anggota dengan pusat adalah PRO dan koalisi lainnya terdiri dari enam anggota dengan pusat adalah JKW

Kata kunci: *Modified Markov Clustering*, Koalisi, Pemilihan Presiden Indonesia

1 Introduction

Indonesia is a country that adheres to democracy in carrying out the political system, as contained in Chapter 1 Verse 2 of the 1945 Constitution, which states that sovereignty is in the hands of the people and carried out according to the provisions of the Constitution. The

2010 *Mathematics Subject Classification*: 91C20, 68T10
Tanggal Masuk: 28-02-19; direvisi: 26-04-18; diterima: 29-04-18

democratic political system provides an opportunity for Indonesian people to participate in political life actively. One place to look at the activeness of the community in determining attitudes and issuing opinions on political issues is through social media. It is marked by the emergence of social media, such as Facebook, Twitter, Instagram, WhatsApp, and several other social media networks that people use to communicate, get information, and issue opinions that related to politics.

Towards the 2019 election, social media often provides political information to shape the mindset of the community regarding elections in 2019. Also, based on the information provided, the public freely expresses their opinions. It can be used to predict the coalition of political figures that will be formed ahead of the 2019 election by analyzing community responses to political figures. The coalition predictions of political figures that will be formed can be made by analyzing the interactions between the political figures. One method that can be used to analyze an interaction from a collection of objects is the clustering method.

There are various types of clustering methods, one of which is the Markov Clustering (MCL) method. This method was proposed in 2000 by Stijn Van Dongen through his dissertation [1]. This clustering method has been widely used in the field of bioinformatics to see the network of human protein interactions that cause humans to develop Schizophrenia [2], Cancer [3], HIV [4], Dengue Fever [5], and Herpes [6]. This method was developed by [7] to see semantic network of word association data. Based on the results of the related research, it can be seen that the MCL method can group data that has a large volume and has a high level of sparsity.

In this paper, we developed the MCL method into Modified MCL that will be used to solve problems in the political field, that is used to predict coalitions of political figures formed in the 2019 election and to identify the number of coalitions figures formed.

2 Research Methods

2.1 Data Description

The collected data consisted of the number of community comments and followers of 20 political figures in observation and then transform it into matrix form. The figures used are the names assumed by the two presidential candidate pairs with a high probability of being elected in 2019 which is presented in Table 1.

Table 1: Contesting Political Figure 2019

Number of Figure	Name of Figure
1	JKW
2	CAKIM
3	MD
4	MA
5	AIR
6	MOEL
7	TGB
8	DS
9	MR
10	CT
11	SM
12	SP
13	ABW
14	UAS
15	AHY
16	ZH

17	SS
18	AH
19	PRO
20	UNO

The construction of the interaction matrix in this study uses the following assumptions:

- A joint photo between two figures in one of the personal Instagram accounts of the figure.
- If there are no shared photos between the two figures on their respective Instagram accounts, then find their photos along with the (explore) feature using the hashtag (#). Example: #anies-gatot.
- The parts considered in calculating the value of the interaction matrix elements are the number of followers on the Instagram account of each character and the number of positive and negative comments, while neutral comments do not count. Table 2 shows the example result of grouping comments.

Table 2: Grouping Comments

Comment (in Indonesia)	Comment category
Kanan Pak JK (Wapres RI 2014-2019), kiri pak Prabowo (Wapres RI 2019-2024), wah perpaduan yang pas	Positive
Ksatria…:* semoga takdirmu menjaga dan memimpin kami di 2019 dan sepuluh thn kedepan #2019gantipresiden #2019presidenbaru	Negative
Saya sebagai anak bangsa kita dukung gubernur yang terpilih, jangan saling menghujat, mari kita dukung bersama… hidup @jokowi hehehe presidenku <3	Negative
Pak anies capres RI !!!	Not correlated

- The number of comments reviewed is limited to 250 comments.
- If the number of positive and negative comments found in the comment column has reached 50 comments even though 250 comments have not been read, then the comment reading process stops.
- If the number of positive and negative comments does not reach 50 out of 250 comments, then the denominator is the total positive and negative comments obtained.
- If the political figure observed does not have an Instagram account, then the followers' weight is zero.
- If there are no shared photos between two figures, either on a personal Instagram account of one of the characters or when searched using a hashtag (#). Then the weight of the comment is zero.

In this case, the assessment of positive, negative, and neutral comments from netizens is done manually by the researcher. The sample in table 2 shows the comments of netizens regarding the possibility of forming a coalition between JKW and PRO.

The following is a calculation formula for interaction matrix entries that will be formed: Let i and j denote two different figures, then

$$p_{ij} = \text{Followers}(i) + \text{Comments}(j) \tag{1}$$

where

$$\text{Followers}(i) = \frac{\text{NF}(i)}{\text{HF}} \text{x } 25\% \quad (2)$$

and

$$\text{Comments}(j) = \frac{\text{NC}(j)}{\text{TOC}} \text{x } 75\ \% \quad (3)$$

where $\text{NF}(i)$ represents the number of followers on i's Instagram account, $\text{NC}(j)$ represents the number of comments on j's photo who supports j to ally with i, TOC represents the total number of read comments, and HF represents the highest number of followers among all observed figures.

Based on the results of the calculation above, we will get four values, namely 0, 1, 2, and 3 which represent the level of interaction between two figures. Following are the details of each score:

The construction of the interaction matrix in this study uses the following assumptions:

- $0 < p_{ij} \leq 0.25$; $i \neq j$ represented with score 0.
- $0.25 < p_{ij} \leq 0.5$; $i \neq j$ represented with score 1.
- $0.5 < p_{ij} \leq 0.75$; $i \neq j$ represented with score 2.
- $0.75 < p_{ij} \leq 1$; $i \neq j$ represented with score 3.

In this case, if the value of the score obtained is greater, then the level of interaction between the two figures is higher. Here is given one example on how to calculate the entry of interaction matrix. Note that we consider the Instagram account of all figures on October 2018 and the maximum followers at that time is JKW, with 9.1 million followers. Considering two figures JKW and PRO. Let JKW denoted as i and PRO denoted as j. By our observation on PRO's Instagram account, we obtained:

- $\text{NC}(j)$ = Positive comments = 5
- Negative comments = 45
- Total of comments (TOC) = 50

Hence by Equation (2) and (3)

$$\text{Comments}(j) = \frac{\text{Positive comments}}{\text{Total of comments}} \times 75\% = \frac{5}{50} \times 75\% = 0.075$$

$$\text{Followers}(i) = \frac{\text{The number of followers}}{\text{Maximum follower from 20 figures}} \times 25\% = \frac{1}{9.1} \times 25\% = 0.02747$$

$$p_{ij} = \text{Followers}(i) + \text{Comments}(j)) = 0{,}075\ +\ 0{,}02747 = 0.10247$$

Because $0 \leq 0.10247 < 0.25$ then the score of interaction between two figures (JKW and PRO) is 0. This score represents that both of figures impossible to make a coalition.

2.2 Modified Markov Clustering

Markov Clustering (MCL) is a clustering method proposed by Stijn Van Dongen in 2000 [1]. In this paper, we modified the MCL algorithm on the normalization stage in order to form a stochastic matrix using the Max-Min normalization. Here is the algorithm of the Modified MCL:

1. Create the interaction matrix
2. Add Self Loops
3. Normalize the interaction matrix using Max-Min normalization.
4. Expand by taking the k-th power of the interaction matrix
5. Inflate by taking inflation of the resulting matrix with parameter r

6. Repeat steps 5 and 6 until a steady state is reached (convergence).

Here is the explanation of each stage of the Modified MCL algorithm:

1. Create the interaction matrix.
   The process of constructed the interaction matrix is given by the flowchart below.

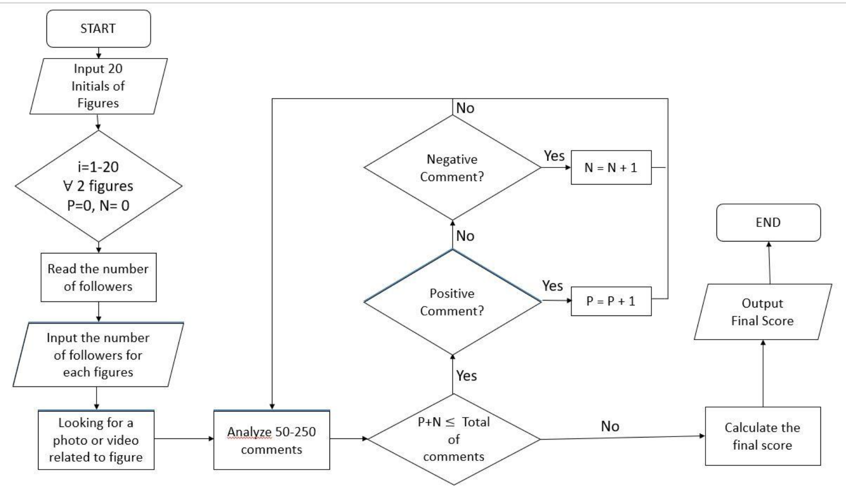

Figure 1: The Process of Constructed the Interaction Matrix

2. Add Self-Loops.
   Let A be a square matrix. The add self-loops is a function that map A to a square matrix $A + K$, where K is a diagonal matrix with $k_{ij} = 3$ for $i = j$. [1]
3. Normalize the interaction matrix using Max-Min normalization.
   Let A be a square matrix. The Max-Min normalization is a function that map A to a square matrix A', where the entries of A' are in the range [0,1]. if a_{ij} is the entry (i, j) of A, then the entry a'_{ij} of A',is given by the following formula: [8]

$$a'_{ij} = \frac{a_{ij} - \min a_j}{\min a_j - \min a_j} \tag{4}$$

   where $\max a_j$ and $\min a_j$, respectively, denote the maximum and minimum value in j-th column of the matrix A.
4. Expand function $\mathrm{Exp}(A, p)$
   Let A be a square matrix. The expand function $\mathrm{Exp}(A, p)$ is a function that map A to p-power operation of the matrix A. Mathematically it can be written as follows. [1]

$$\mathrm{Exp}(A, p) = A^p \tag{5}$$

   In this research the value of p is 2 because most of previous research used this value for parameter p. The expand function $\mathrm{Exp}(A, p)$ is done with the aim of increasing the flow between existing vertices and opening new potential flows.
5. Inflate function $\mathit{inflate}(A, r)$
   Let A be a square matrix. The inflate function $\mathit{inflate}(\mathrm{A}, \mathrm{r})$ is a function that map A to a square matrix A', where each element in every column of A' is obtained by the r-power of each element in every column of A then divided by the sum of the r-power of the elements in that column. Mathematically, the inflation value of a square matrix A can be written as follows. [1]

$$(\Gamma_r A)_{ij} = \frac{(a_{ij})^r}{\sum_{k=1}^{m}(a_{kj})^r}; i = 1, \dots, n; j = 1, \dots, n. \tag{6}$$

The value of r must be greater than one, and in most research the value of r is 2. The inflate function is responsible for both strengthening and weakening of current flows. That way in this research used the value of r = 2.

6. Convergence criteria.
The MCL process is said to have reached convergence if the global chaos value in the equation (8) is close to zero. If the global chaos value is close to zero, there is no significant change in the new cluster. The global chaos is the maximum of global chaos values in the k-column can be calculated by calculating the maximum value of the k-column divided by the sum of squares of the elements in the k-column. Mathematically, global chaos and chaotic values of a square matrix A can be written as follows. [1]

$$(chaos)_k = \frac{\max\{a_{ik}, a_{2k}, \dots, a_{nk}\}}{\sum_{i=1}^{n} a_{ik}^2} \tag{7}$$

where

$$globalchaos = \max\{(chaos)_k, k = 1, \dots, n\} < 10^{-3}. \tag{8}$$

3 Simulation and Discussion

The dataset used is a file with CSV format so it is needed syntax to open this file form. Syntax the one shown in Figure 1 to transform the form of data into numerical matrix so that it can be processed. Figure 2 is the interaction data display between figures who will coalition in 2019 presidential election represented by the matrix measuring 20 x 20. Based on the results of calculations from equations (1), (2), and (3) and fulfilling the assumptions in sub-chapter 2.1, it can be formed the interaction matrix in Figure 2.

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16
1	3	3	3	3	1	1	2	1	2	1	1	1	0	0	0	0
2	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	3	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
5	1	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
7	2	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
8	1	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
9	2	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0
10	1	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0
11	1	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0
12	1	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3

Figure 2: The Interaction Matrix

After getting a matrix representation from the dataset, the next step is grouping on the matrix dataset using Modified MCL. This method will form many clusters where in this study the obtained clusters represent the coalition between the political figures as given on Table 1. Here is the result of modified MCL where expansion parameter $p = 2$, inflation parameter $r = 2$, the maximum iteration is 100, and threshold value is 10^{-3}.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 3: Adjacency matrix

The matrix in Figure 3 is generated in the 20^{th} iteration. The number of clusters obtained is two clusters that represent figures coalition. The adjacency matrix in Figure 3 can be interpreted as undirected graph as given by Figure 4. In addition, matrix adjacency is a symmetry matrix.

Based on the theory, the correlation between figures represented by entry 1 in the adjacency matrix shows that the figure coalesces with each other. For example, a cluster with a center numbered 19, look at the row 19 in the matrix adjacency. For each entry with a value of 1 in each column corresponding to row 19, a graph with no directional direction can be formed with cluster 19 as shown in Figure 4. In the other side, it can be looked by considering entries with a value of 1 in each row corresponding to column 19. If the matrix entry is worth 1 in the same row and column and has a value of 0 for the other rows or columns, this indicates that the figure on the row does not form a coalition. For example, figures represented by numbers 5,6,8,10,11,12 and 18 only form loops against themselves.

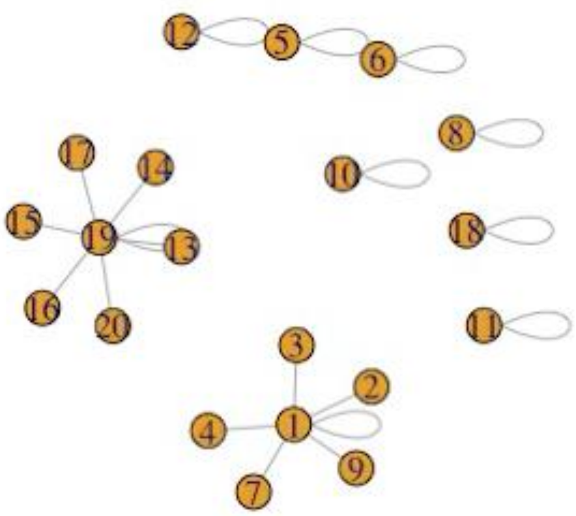

Figure 4: The Graph Output from MCL Process with Input Matrix from Figure 2

Based on Figure 3 and 4, it can be analyzed for the results of grouping data with the Modified MCL on the R produce many figures at each cluster that represent the number of figure in each coalition presented by Table 3 below:

Table 3: The Center and Members of Each Cluster

Number	Index	Number of Figure	Center of Cluster	Member of Cluster
1	1	6	1	1,2,3,4,7,9
2	0	7	-	5,6,8,10,11,12,18
3	9	7	19	13,14,15,16,17,19,20

From Table 3, we obtained two coalitions with cluster center are number 19 and 1. Beside that, there are seven figures who separately or they are not including both of two coalitions. The first coalition consist of seven members with the cluster center is number 19. The second coalition consist of six members with the cluster center is number 1. Moreover, based on Table 1 and Table 3, the representation of the names of figures is presented in Table 4 below.

Table 4: The Name of Figures that corresponding with Table 3

Center of Cluster	Member of Cluster
JKW	JKW, CAKIM, MD, MA, TGB, MR
-	AIR, MOEL, DS, CT, SM, SP, AH
PRO	ABW, UAS, AHY, ZH, SS, PRO, UNO

4 Conclusion

Based on the results of research that has been done, it can be concluded several things as follows:

- The Markov Clustering method can be used to solve problems in the fields of politics.
- The number of coalitions formed from interaction between figures in Indonesia as many as two coalitions.
- There are seven figures who are not coalition with any figure.
- The first coalition consist of seven members with the center is PRO. In the other coalition, consist of six members with the center is JKW.
- Based on the reality, JKW join with MA, and PRO join with UNO as president and vice president candidates. This finding corresponds with the results of this research that JKW and MA in the same coalition. Moreover, PRO and UNO also in the same coalition.

Reference

[1] S. V. Dongen, *Graph Clustering by Flow Simulation*, PhD Thesis. University of Utrecht, 2000.

[2] T. H. Sabel and A. Bustamam, ‘Clustering protein-protein interaction network of TP53 tumor suppressor protein using Markov clustering algorithm’, *IEEE Transactions on Reliability*, pp. 221-226, 2015.

[3] R. Ginanjar, A. Bustamam, and H. Tasman, ‘Implementation of regularized Markov clustering algorithm on protein interaction networks of schizophrenia's risk factor candidate genes’, *IEEE Transactions on Reliability*, pp. 297-302, 2016.

[4] D. Lestari, D. Raharjo, A. Bustamam, B. Abdillah, and W. Widhianto, ‘Application of clustering methods: Regularized Markov clustering (R-MCL) for analyzing dengue virus similarity’, *AIP Conf. Proc.*, vol. 1862, pp. 030130-1-030130-7, 2017.

[5] A. Bustamam, T. Siswantining, N. L. Febriyani, and I. D. Novitasari, ‘Protein sequences clustering of herpes virus by using Tribe Markov clustering (Tribe-MCL)’,

AIP Conf. Proc., vol. 1862, pp. 030150-1-030150-8, 2017.

[6] A. Bustamam, M. S. Wisnubroto, and D. Lestari, 'Analysis of protein-protein interaction network using Markov clustering with pigeon-inspired optimization algorithm in HIV (human immunodeficiency virus)', *AIP Conf. Proc.*, vol. 2023, pp. 020229-1-020229-6, 2018.

[7] J. Jaeyoung, M. Miyake, and H. Akama, 'Parameterization of nonlinear shallow water waves over sloping bottoms', *European Language Resources Association (ELRA)*, pp. 1428-1431, 2006.

[8] J. C. Ang, A. Mirzal, H. Haron, and H. Hamed, 'Supervised, Unsupervised, and Semi-Supervised Feature Selection: A Review on Gene Selection', *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 13, pp. 971-989, 2016.