Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 157 (2019) 31–37

Procedia Computer Science

www.elsevier.com/locate/procedia

4th International Conference on Computer Science and Computational Intelligence 2019 (ICCSCI), 12-13 September 2019

Biclustering Method to Capture the Spatial Pattern and to Identify the Causes of Social Vulnerability in Indonesia : A New Recommendation for Disaster Mitigation Policy

Puspita Anggraini Kaban[a,*], Robert Kurniawan[a], Rezzy Eko Caraka[b,c], Bens Pardamean[b,d], Budi Yuniarto[a], Sukim[a]

[a] *Computational Statistics Department, Polytechnic of Statistics - STIS, Jakarta 13330, Indonesia*

[b] *Bioinformatics and Data Science Research Center, Bina Nusantara University, Indonesia*

[c] *College of Informatics, Chaoyang University of Technology, Taichung City 41349, Taiwan (R.O.C.)*

[d] *Computer Science Department, BINUS Graduate Program Master of Computer Science Bina Nusantara University, Jakarta, Indonesia, 11480*

Abstract

Geographically, Indonesia is a meeting point of three continental plates. Scilicet, the Eurasian Plate, the Indo-Australian Plate, and the Pacific Plate. Therefore, Indonesia is part of the infamous volcanic zone called the ”Ring of Fire” and one of the areas prone to natural disasters such as volcanic eruptions, earthquakes, tsunamis, floods, and landslides. This study aims to capture the spatial pattern and identify the causes of social vulnerability in the districts/cities in Indonesia using the biclustering method. The data is extracted from the Indonesian National Socio-Economic Survey (SUSENAS) by BPS-Statistics in 2014. The biclustering result indicates that each district/city has its own social vulnerability characteristics and shows that the vulnerable aspects of each district/city are different. The adjacent observations tend to have social vulenrability characteristics. The results of this study can be used as a reference for national disaster mitigation policy in Indonesia.

Peer-review under responsibility of the scientific committee of the 4th International Conference on Computer Science and Computational Intelligence 2019.

Keywords: biclustering; disaster mitigation; social vulnerability; SoVI; spatial pattern

1. Introduction

Geographically, Indonesia is a meeting point of three continental plates. Scilicet, the Eurasian Plate, the Indo-Australian Plate, and the Pacific Plate. Therefore, Indonesia is part of the infamous volcanic zone called the ”Ring

* Corresponding author. Tel.: +62-857-6331-8138.
E-mail address: 15.8823@stis.ac.id

Peer-review under responsibility of the scientific committee of the 4th International Conference on Computer Science and Computational Intelligence 2019.
10.1016/j.procs.2019.08.138

of Fire" and one of the areas prone to natural disasters such as volcanic eruptions [1], earthquakes [2], tsunamis [3], floods [4,5], and landslides [6]. The National Disaster Management Agency Indonesia recorded 1,134 natural disasters during January-October 2018. These disasters bring negative impacts on a variety of scales in each area of the incident.

Although disasters with the same scale occur in different social groups, their effects are vary depending on the ability of each group to deal with disaster [7]. The ability of social groups to deal with disasters is related to their vulnerability to disasters themselves. O'Keefe, Westgate, and Wisner [8] define vulnerability as the level at which the community is at risk from natural disasters, the risk can be the quantity of disaster occurrence, the level of socio-economic, or socio-political capabilities of the society to survive and restore yourself from disaster. Cutter [9] defines vulnerability as the potential for loss when natural disasters occur. The vulnerability can be classified into two main categories, physical vulnerability in the form of risk of physical loss and damage or social vulnerability in the form of factors that increase or reduce the influence of physical damage [9,10].

The previous researches focused more on physical vulnerability than on social vulnerability [11,12,13]. Lately, social vulnerability is gaining more attention and widely researched throughout the world. Social vulnerability assessment provides broader insights about which community decision making influences people's differential experiences of disaster events [14]. The SoVI method is a popular method used to measure vulnerability in various research in the world as its implementation is easy to adapt and modify. However, the SoVI method has several disadvantages, including the uncertainty regarding the identified areas that are very vulnerable [15], the inability to identify the basic causes of social vulnerability, and the tendency to emphasize on variables with high variance leads to missing the demographic aspects with significant influences and low variances [16].

The SoVI method only generates a general (global) model for the scope of research (eg: country) and is not able to identify the specific (local) models of vulnerabilities that occur in the unit level of observation. Biclustering is a method of concurrent grouping on rows and columns in a matrix that represents data. The two-way grouping method was first introduced by Hartigan [17] as a method of clustering the rows and columns separately (not simultaneous). The term biclustering was later proposed by Cheng and Church [18]. Grouping data with the clustering method produces a global model because each row is considered to have the same characteristics (same conditions on columns) while the biclustering method produces a local model because each bicluster has different characteristics [19]. This study aims to capture the spatial pattern and identify the causes of social vulnerability in the districts/cities in Indonesia using the biclustering method. The biclustering method creates local models for each group of regions that has similar characteristics of social vulnerability. The results of grouping regions are then mapped to see the spatial pattern of social vulnerability.

2. Literature Review

Cutter et al [11] first performed quantitative social vulnerability assessment by building an index called social vulnerability index (SoVI). SoVI was originally developed as an analysis of social contributions to natural disasters on the counties in the United States in 1990. SoVI measured social inequality and space gaps as dimensions of social vulnerability. Social disparity took the form of social factors that influenced the sensitivity and responsiveness of a community group to disasters. Space gaps included the characteristics of the community and its environments, such as the level of urbanization, the level of population growth and economic vitality. SoVI was built using 42 normalized independent social variables. Principal component analysis (PCA) was implemented to transform all of the variables into several components. The index was constructed with 11 selected principal components which explained 76.4% of the data variance.

Nowadays, the SoVI method is adopted in many social vulnerability assessments arround the world. There are many modifications to the application of SoVI and some studies only adopt the stages of SoVI. In general, the components and variables that contruct the index are customized to suit the characteristics of each location. The SoVI method has also been applied in Indonesia [20,21,22,23,24,25]. Among all social vulnerability assessments in Indonesia, study by Siagian et al [22] was the first that covered the entire territory of Indonesia with districts/cities as observation units. They used 10 social variables that represent 4 main indicators of social vulnerability. The study concluded that there were 3 boosting aspects of social vulnerability condition in Indonesia. The aspects were : socioeconomic and infrastructure condition, gender and age distributions and population growth, and household structures in the population.

The biclustering method was originally implemented to analyze gene expression microarray data. Madeira and Oliveira [19] provided a comprehensive description of the biclustering method for further understanding. Yuniarto and Kurniawan [26] conducted a study that implemented of the biclustering method to analyze social data. The aim of the study is to map the structure of dimensional poverty in East Java Province. The study used Cheng and Church algorithm [18] and generated two biclusters that represented the regions with good poverty indicators and poor poverty indicators.

3. Materials and Methods

3.1. Data

We extract the data from Indonesian National Socio-Economic Survey (SUSENAS) by the Central Bureau of Statistics Indonesia in 2014. SUSENAS is a national survey to collect information/data in the fields of population, health, education, household structure, housing, consumption, and expenditure. It covers all 497 districts and cities in Indonesia. The social vulnerability variables used in this study is adapted from the prior research by Siagian et al [22]. Table 1 summarize all the social vulnerability indicators and its representative variable(s) used in this research.

Table 1: Indicators of social vulnerability and its representative variable(s)

Indicator	Variable
Age	Proportion of the population under the age of 5 (X_1), Proportion of the population over the age of 65 (X_2)
Gender	Proportion of female population (X_3), Proportion of households with a female head of households (X_4)
Income	Proportion of poor population (X_5)
Education	Proportion of illiterate population (X_6), Proportion of the population over the age of 15 with a low level of education(X_7)
Household Structure	Average number of household members (X_8)
Infrastructure	Proportion of households without electricity (X_9)
Population growth	Population growth rate (X_{10})

3.2. Methods

Cheng and Church [18] defined a bicluster as part of a row and part of a column that has a high similarity value. The similarity between objects in bicluster is measured by the mean squared residue (R). The purpose of this algorithm is to find bicluster(s) with mean squared residue smaller than the specified limit. Data is defined as a matrix $A = (X, Y)$ and a bicluster is represented as a sub-matrix (M, N) with $M \subset I$ and $N \subset J$. The sub-matrix generates mean squared residue which is smaller than δ, where δ is a predetermined parameter with a value greater than or equal to 0. The following is the equation for mean squared residue:

$$R(M,N) = \frac{1}{|M||N|} \sum_{m\epsilon M, n\epsilon N} \left(a_{mn} - \frac{1}{|N|}\sum_{n\epsilon N} a_{mn} - \frac{1}{|M|}\sum_{m\epsilon M} a_{mn} + \frac{1}{|M||N|}\sum_{m\epsilon M, n\epsilon N} a_{mn}\right)^2 \tag{1}$$

According to Chakraborty and Maha [27], a good bicluster has a small mean squared residue with maximum (large) dimension. Therefore the quality of the biclusters can be measured by calculating the average of ratio between residue (R) and volume (V). The following is the formula for measuring the quality of biclustering result with p bicluster(s):

$$\frac{R}{V} = \frac{1}{p}\sum_{k=1}^{p} \frac{R_k}{V_k} \tag{2}$$

3.3. Tools

The scripts are witten in R [28] and executed using RStudio version 1.0.143 [29]. We use package *biclust* [30] for biclustering analysis and package *ggplot2* [31] for radar charts visualization. We use QGIS version 2.18.13 [32] for spatial map.

4. Results and Discussion

The best parameter is determined by manual tuning. Table 2 summarizes the result of parameter tuning. It can be seen that even though δ limits the residue tolerance, δ value is not linear with R/V. The δ values $0.1 - 0.3$ show positive relationship with R/V. The R/V value increases as the δ value increases. On the contrary, the δ values $0.4-0.7$ show negative relationship with R/V. The R/V value decreases as the δ value increases. $\delta = 0.8$ produces maximum R/V. The δ values $0.9 - 1$ produces very small R/V values but produces only one bicluster. $\delta = 0.7$ is selected as the best parameter by considering the R/V value and the number of biclusters it produces. In the interval $0.4 - 0.7$, $\delta = 0.7$ produces the smallest R/V value and the convergent number of 5 biclusters. Table 3 sumarizes the result of biclustering with $\delta = 0.7$.Each observation is grouped in exactly one bicluster except the city of Yogyakarta that is not grouped in any bicluster. The majority of observations are grouped into bicluster 1.

Table 2: The result of parameter tuning

δ	Number of biclusters	R/V
.1	31	.0024
.2	15	.0032
.3	12	.0046
.4	8	.0069
.5	8	.0067
.6	5	.0061
.7	5	.0052
.8	5	.0153
.9	1	.0002
1	1	.0002

Table 3: The number of observations and variable(s) in each bicluster

Bicluster	Number of districts / cities	Number of Variables
1	432	10
2	35	7
3	17	5
4	9	8
5	3	10

Bicluster 1 with 432 observations contains all of the variables. The averages of variables in bicluster 1 approximately similar to the national averages, bicluster 1 represents the general condition of social vulnerability in Indonesia. Bicluster 2 with 35 observations contains 7 variables with the averages of variables X_5, X_6, and X_9 are higher than the national averages. Bicluster 3 with 17 observations contains 5 variables the averages of variables X_2, X_3, X_4, and X_8 are lower than the national averages. But the average of variable X_7 in bicluster 3 is significantly higher than the national average. Bicluster 4 with 9 observations contains 8 variables with the averages of variables X_2, X_3, X_6, and X_7 are lower than the national averages. The averages of variables X_5, X_8, and X_9 are higher than the national average. Bicluster 5 with 3 observations contains all of the variables. This bicluster indicates unique and critical condition. The averages of variables X_1, X_2, X_3, X_4, and X_8 are lower than the national averages. The averages of variables X_5, X_6, X_7, X_9, and X_{10} are higher than the national average. See figure 1 for radar charts representations of the variables in each bicluster. Variables description can be seen at table 1.

According to figure 2, the adjacent observations tend to be grouped into the same bicluster, this indicates that the spatial effects and social relations the regions affect the social vulnerability conditions of these regions. This tendency can be observed especially in the area of the island of Papua, the topography and accessibility conditions of the regions strongly affect the biclustering result. Bicluster 5, which indicators indicates the worst social vulnerability conditions

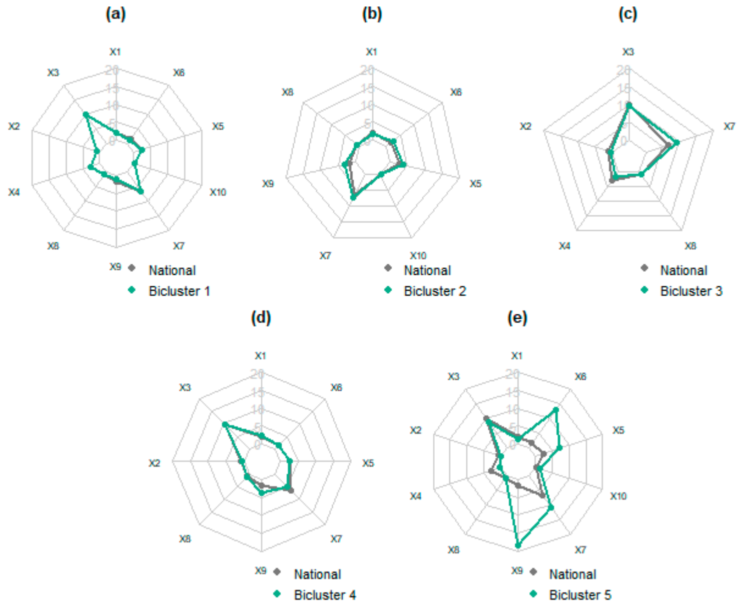

Fig. 1: (a) Radar chart for bicluster 1 and national; (b) Radar chart for bicluster 2 and national; (c) Radar chart for bicluster 3 and national; (d) Radar chart for bicluster 4 and national; (e) Radar chart for bicluster 5 and national

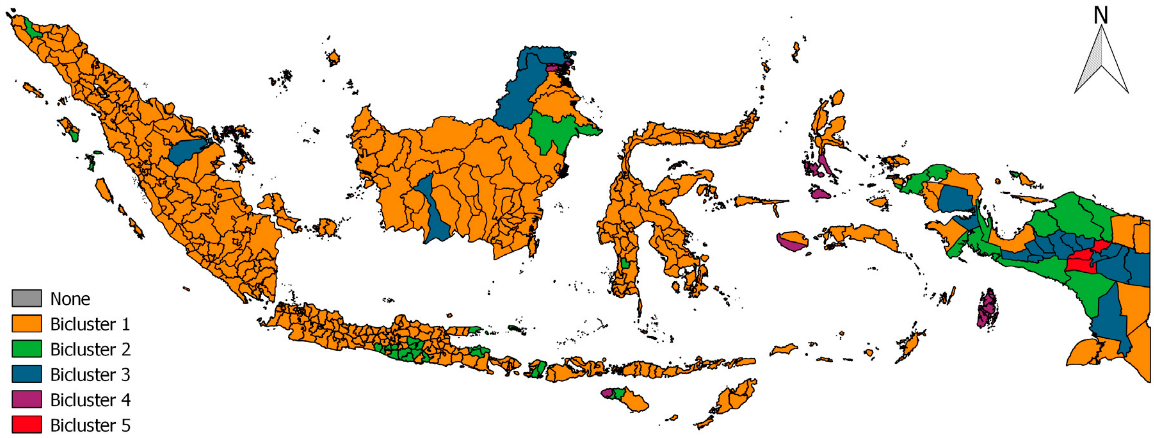

Fig. 2: Biclusters map of social vulnerability in Indonesia

in Indonesia, captures three adjacent districts in the mountainous area of Papua. Bicluster 4 captures the other districts in the same mountainous area. The indicators of bicluster 4 also indicates worse social vulnerability condition than national. The districts/cities in coastal areas of Papua have relatively better social vulnerability conditions and grouped in bicluster 2. This discovery substantiates the result of the study by Siagian et al[22] which also captures that the majority of districts/cities on the island of Papua are socially vulnerable to disasters.

5. Conclusion and Recommendation

The biclustering method is able to group the districts/cities into 5 biclusters and to show the spatial pattern of social vulnerability in Indonesia. The result identifies the local models of social vulnerability for each bicluster. These models show the most significant causes of social vulnerability for the districts/cities. National Disaster Management Agency set the disaster mitigation policies in the territory of Indonesia based on the location prioritization by type of disasters[33]. Priority locations are determined based solely on the number of lives and infrastructure exposed, the probability of occurrences over the next five years, and events affecting more than two provinces. The biclustering result indicates that each district/city has its own social vulnerability characteristic and shows that the vulnerable aspects of each region are different. Based on the results of this study, a more customized mitigation policy can be made for each district/city based on its social vulnerability characteristics. Since this work only presents a prototype workflow, for future research we suggest building a DSS framework that provides data capturing, model implementation and dashboards to monitor the social vulnerability condition in Indonesia. For the observations in bicluster 5 that indicates critical social vulnerability condition, we suggest future research to investigate deeper on the profiles of the observations (eg: topography, accessibility, culture, etc).

6. Acknowledgement

This work supported by Computational Statistics Department, Polytechnic of Statistics and Bioinformatics and Data Science Research Center, Bina Nusantara University, Indonesia

References

1. Oppenheimer, C.. Climatic, environmental and human consequences of the largest known historic eruption: Tambora volcano (indonesia) 1815. *Progress in physical geography* 2003;**27**(2):230–259.
2. Socquet, A., Hollingsworth, J., Pathier, E., Bouchon, M.. Evidence of supershear during the 2018 magnitude 7.5 palu earthquake from space geodesy. *Nature Geoscience* 2019;**12**(3):192.
3. Heidarzadeh, M., Muhari, A., Wijanarto, A.B.. Insights on the source of the 28 september 2018 sulawesi tsunami, indonesia based on spectral analyses and numerical simulations. *Pure and Applied Geophysics* 2019;**176**(1):25–43.
4. Caraka, R.E., Bakar, S.A., Tahmid, M., Yasin, H., Kurniawan., I.D.. Neurocomputing fundamental climate analysis. *Telkomnika* 2019; **17**(4):101–110.
5. Caraka, R.E., Tahmid, M.. *Statistika Klimatologi*. Yogyakarta: Mobius Graha Ilmu; 2019.
6. Syifa, M., Kadavi, P.R., Lee, C.W.. An artificial intelligence application for post-earthquake damage mapping in palu, central sulawesi, indonesia. *Sensors* 2019;**19**(3):542.
7. de Loyola Hummell, B.M., Cutter, S.L., Emrich, C.T.. Social vulnerability to natural hazards in brazil. *International Journal of Disaster Risk Science* 2016;**7**(2):111–122.
8. O'Keefe, P., Westgate, K., Wisner, B.. Taking the naturalness out of natural disasters. *Nature* 1976;**260**:566–567.
9. Cutter, S.L.. Vulnerability to environmental hazards. *Progress in human geography* 1996;**20**(4):529–539.
10. Alcántara-Ayala, I.. Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. *Geomorphology* 2002;**47**(2-4):107–124.
11. Cutter, S.L., Boruff, B.J., Shirley, W.L.. Social vulnerability to environmental hazards. *Social science quarterly* 2003;**84**(2):242–261.
12. Gall, M.. *Indices of Social Vulnerablity [sic] to Natural Hazards: A Comparative Evaluation*. Ph.D. thesis; University of South Carolina; 2007.
13. Flanagan, B.E., Gregory, E.W., Hallisey, E.J., Heitgerd, J.L., Lewis, B.. A social vulnerability index for disaster management. *Journal of homeland security and emergency management* 2011;**8**(1).
14. Schmidtlein, M.C., Deutsch, R.C., Piegorsch, W.W., Cutter, S.L.. A sensitivity analysis of the social vulnerability index. *Risk Analysis: An International Journal* 2008;**28**(4):1099–1114.
15. Tate, E.. Social vulnerability indices: a comparative assessment using uncertainty and sensitivity analysis. *Natural Hazards* 2012;**63**(2):325–347.
16. Wood, N.J., Burton, C.G., Cutter, S.L.. Community variations in social vulnerability to cascadia-related tsunamis in the us pacific northwest. *Natural Hazards* 2010;**52**(2):369–389.
17. Hartigan, J.A.. Direct clustering of a data matrix. *Journal of the american statistical association* 1972;**67**(337):123–129.
18. Cheng, Y., Church, G.M.. Biclustering of expression data. In: *Ismb*; vol. 8. 2000, p. 93–103.
19. Madeira, S.C., Oliveira, A.L.. Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 2004;**1**(1):24–45.
20. Utami, P.. Measuring social vulnerability in volcanic hazards: The case study of merapi volcano, indonesia, earth sciences. *MSc thesis, University of Bristol Bristol* 2008;.
21. Donovan, K., Suryanto, A., Utami, P.. Mapping cultural vulnerability in volcanic regions: The practical application of social volcanology at mt merapi, indonesia. *Environmental Hazards* 2012;**11**(4):303–323.
22. Siagian, T.H., Purhadi, P., Suhartono, S., Ritonga, H.. Social vulnerability to natural hazards in indonesia: driving factors and policy implications. *Natural hazards* 2014;**70**(2):1603–1617.

23. Maharani, Y.N., Lee, S., Ki, S.J.. Social vulnerability at a local level around the merapi volcano. *International Journal of Disaster Risk Reduction* 2016;**20**:63–77.
24. Hizbaron, D., Hadmoko, D., Mei, E., Murti, S., Laksani, M., Tiyansyah, A., et al. Towards measurable resilience: Mapping the vulnerability of at-risk community at kelud volcano, indonesia. *Applied geography* 2018;**97**:212–227.
25. Kurniawan, R., Siagian, T.H., Yuniarto, B., Nasution, B.I., Caraka, R.E.. Construction of social vulnerability index in indonesia using partial least squares structural equation modeling. *International Journal of Engineering Technology* 2018;**7**(4):6131–6136.
26. Yuniarto, B., Kurniawan, R.. Understanding structure of poverty dimensions in east java: Bicluster approach. *Signifikan: Jurnal Ilmu Ekonomi* 2017;**6**(2):289–300.
27. Chakraborty, A., Maka, H.. Biclustering of gene expression data using genetic algorithm. In: *2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*. IEEE; 2005, p. 1–8.
28. R Core Team, . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria; 2017. URL https://www.R-project.org/.
29. RStudio Team, . *RStudio: Integrated Development Environment for R*. RStudio, Inc.; Boston, MA; 2016. URL http://www.rstudio.com/.
30. Sebastian Kaiser Rodrigo Santamaria, T.K.M.S.R.T.L.Q.F.L., Troyer, E.D.. *biclust: BiCluster Algorithms*; 2018. R package version 2.0.1; URL https://CRAN.R-project.org/package=biclust.
31. Hadley Wickham Winston Chang, L.H.T.L.P.K.T.C.W.K.W.R.c.. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*; 2019. R package version 3.1.1; URL https://CRAN.R-project.org/package=ggplot2.
32. QGIS Development Team, . *QGIS Geographic Information System*. Open Source Geospatial Foundation; 2016. URL http://qgis.org.
33. Badan Nasional Penaggulangan Bencana, . *Rencana Nasional Penanggulangan Bencana 2015-2019*. 2014. Accessed 10 December 2018.; URL https://bnpb.go.id/uploads/24/buku-renas-pb.pdf.