

# Spatio-Temporal Graph-Based Hotspot Analysis of Earthquake Events Using Spatial Autocorrelation and Community Detection in Indonesia

Ika Arfiani<sup>1</sup>, Herman Yuliansyah<sup>2</sup>, Nur Rochmah Dyah Puji Astuti<sup>3</sup>, Arfiani Nur Khusna<sup>4</sup>

<sup>1,2,3,4</sup>Informatics Department, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

**Received:**

October 4, 2025

**Revised:**

May 10, 2026

**Accepted:**

May 30, 2026

**Published:**

June 28, 2026

Corresponding Author:

**Author Name\*:**

Ika Arfiani

**Email\*:**

ika.arfiani@tif.uad.ac.id

DOI:

10.63158/journalisi.v8i3.1641

© 2026 Journal of Information Systems and Informatics. This open access article is distributed under a (CC-BY License)



**Abstract.** Analysis of clustered seismic regions is important for understanding seismic activity patterns in tectonic regions such as Indonesia. However, conventional spatial statistical approaches generally analyze earthquake events independently and fail to capture complex spatio-temporal relationships. This study proposes a graph-based spatio-temporal hotspot analysis approach integrating spatial autocorrelation and community detection to identify regional seismic interaction patterns. The dataset used consists of 3,000 earthquake events from 2008–2025. Spatial autocorrelation was analyzed using Moran's I, while earthquake relationships were modeled using a spatio-temporal graph with spatial and temporal thresholds of  $\leq 400$  km and  $\leq 60$  days. The results showed significant positive spatial autocorrelation with Moran's I = 0.3367 ( $p = 0.001$ ). The resulting graph consisted of 3,000 nodes and 22,896 edges, revealing substantial regional-scale connectivity and 14 major clusters with a modularity score of 0.7405, indicating a strong community structure. Degree centrality analysis identified highly connected nodes with a maximum degree of 77. These findings indicate that integrating spatial autocorrelation and graph analysis provides a more comprehensive representation of seismic interaction patterns and may support future seismic risk assessment in tectonically active regions.

**Keywords:** community detection, earthquake hotspot analysis, Indonesian seismicity, network analysis, spatial autocorrelation, spatio-temporal graph.

## 1. INTRODUCTION

Indonesia is one of the regions with the highest levels of seismic activity in the world due to its location at the junction of the Indo-Australian, Eurasian, and Pacific Plates. The interaction between these plates forms an active subduction zone and a complex fault system, which triggers earthquakes with a high frequency and recurring patterns. Several studies have shown that earthquake events are not random, but rather tend to form clustering patterns in spatial and temporal dimensions due to physical interactions between earthquake sources and energy release processes along the tectonic system [1]–[4].

Various approaches have been used to analyze these patterns. Unsupervised learning-based methods such as K-Means and DBSCAN are commonly used to group earthquake events based on attributes such as location, magnitude, and depth [5], [6]. While effective for data segmentation, these approaches assume independence between observations and do not consider explicit relationships between events in space and time. Density-based approaches such as Kernel Density Estimation (KDE) are capable of identifying areas with high event concentrations (hotspots), but remain limited to distributional analysis and do not represent the structure of relationships between data points [7], [8]. In geoinformatics, spatial autocorrelation analysis is an important approach for testing whether the distribution of a phenomenon is random or structured. Moran's I is a global measure widely used to statistically identify spatial clustering [9]–[11]. A significant Moran's I value indicates that adjacent events tend to have similar characteristics, in line with the first law of geography [12]. However, this approach is aggregative and does not explicitly model the relationships between events as a network structure.

With the development of network science, graph data representation has become an increasingly relevant approach for analyzing complex systems. In this approach, each event is represented as a node, while the relationships between events are represented as edges, allowing for analysis of connectivity, centrality, and community structure [13]–[16]. Graph-based approaches enable the analysis of connectivity, centrality, and community structures, allowing complex interaction patterns between events to be represented more explicitly than conventional statistical approaches.

In the context of seismology, several studies have demonstrated that earthquake events can be modeled as complex networks to identify spatial and temporal interaction patterns between seismic events [17]–[19]. Recent studies further indicate that seismic activity is strongly influenced by regional geological conditions and fault interactions, resulting in clustered seismic behavior across both spatial and temporal dimensions [20]–[22]. Graph-based approaches have also shown the ability to uncover hidden correlations and community structures that are difficult to identify using conventional statistical techniques [23]–[25]. However, most previous studies still treat spatial statistical analysis and network-based modeling separately, limiting the comprehensive representation of interconnected earthquake dynamics [26]–[28].

Furthermore, conventional statistical approaches generally treat earthquake events as relatively independent phenomena, thus not fully accommodating the complex interconnections between events in the spatial and temporal dimensions. This limitation opens up opportunities to develop approaches that can simultaneously integrate both perspectives. In this study, spatial and temporal thresholds of 400 km and 60 days were selected based on regional seismic interaction characteristics and previous interpretations of earthquake clustering behavior in tectonically active environments. Based on this research gap, this study integrates spatial autocorrelation analysis and spatio-temporal graph modeling for earthquake hotspot identification in Indonesia. The proposed framework combines Moran's I analysis, graph construction, centrality analysis, and community detection to represent seismic interaction patterns more comprehensively by considering both spatial distribution and temporal relationships between earthquake events.

## 2. METHODS

### 2.1. Dataset and Data Characteristics

This study uses a dataset of earthquakes in Indonesia for the period 2008–2025 obtained from a public GitHub-based repository [29]. This dataset includes attributes such as occurrence time, geographic coordinates (latitude and longitude), depth, and magnitude, which are key variables in modern seismological studies [22]. The data is available from November 1, 2008, to September 2025, with over 30 attributes describing earthquake characteristics, including occurrence time, geographic coordinates (latitude and

longitude), depth, magnitude, and additional parameters related to the earthquake source mechanism. This data represents the distribution of earthquake events in spatial and temporal dimensions commonly used in modern seismological studies. An example of the dataset structure is shown in Table 1.

**Table 1.** Earthquake Dataset.

Time	Latitude	Longitude	Depth (km)	Magnitude
2008-01-01 00:12:34	-2.45	118.23	35	4.5
2008-01-01 02:45:10	-3.12	127.89	10	5.1
2008-01-02 06:21:55	1.23	121.45	50	3.8
2008-01-03 11:05:40	-6.78	106.55	15	4.2
2008-01-04 14:33:12	-8.21	115.67	70	5.6
.....	.....	.....	....	....

## 2.2. Research Stage

The research stages are designed to systematically integrate spatial analysis and graph modeling. The process begins with data pre-processing, followed by spatial representation, spatial autocorrelation analysis using Moran's I, and spatio-temporal graph construction. The resulting graph is then analyzed using network metrics and community detection to identify cluster patterns and seismic activity hotspots [30]–[32]. The complete research flow is shown in Figure 1.

### 1) Data Preprocessing

Data preprocessing was carried out to improve data quality and ensure consistency before the analysis stage [28]. The preprocessing steps included:

- a) Removing data with blank values in primary attributes
- b) Converting time attributes to datetime format
- c) Converting numeric attributes to the appropriate data type

After preprocessing, 3,000 earthquake events were selected from the cleaned earthquake catalogue. The sampling process was designed to preserve the spatial and temporal characteristics of seismic activity in Indonesia while reducing computational complexity during graph construction and network analysis. This stage ensured that the dataset could be processed consistently and efficiently in subsequent spatial and graph-based analyses.

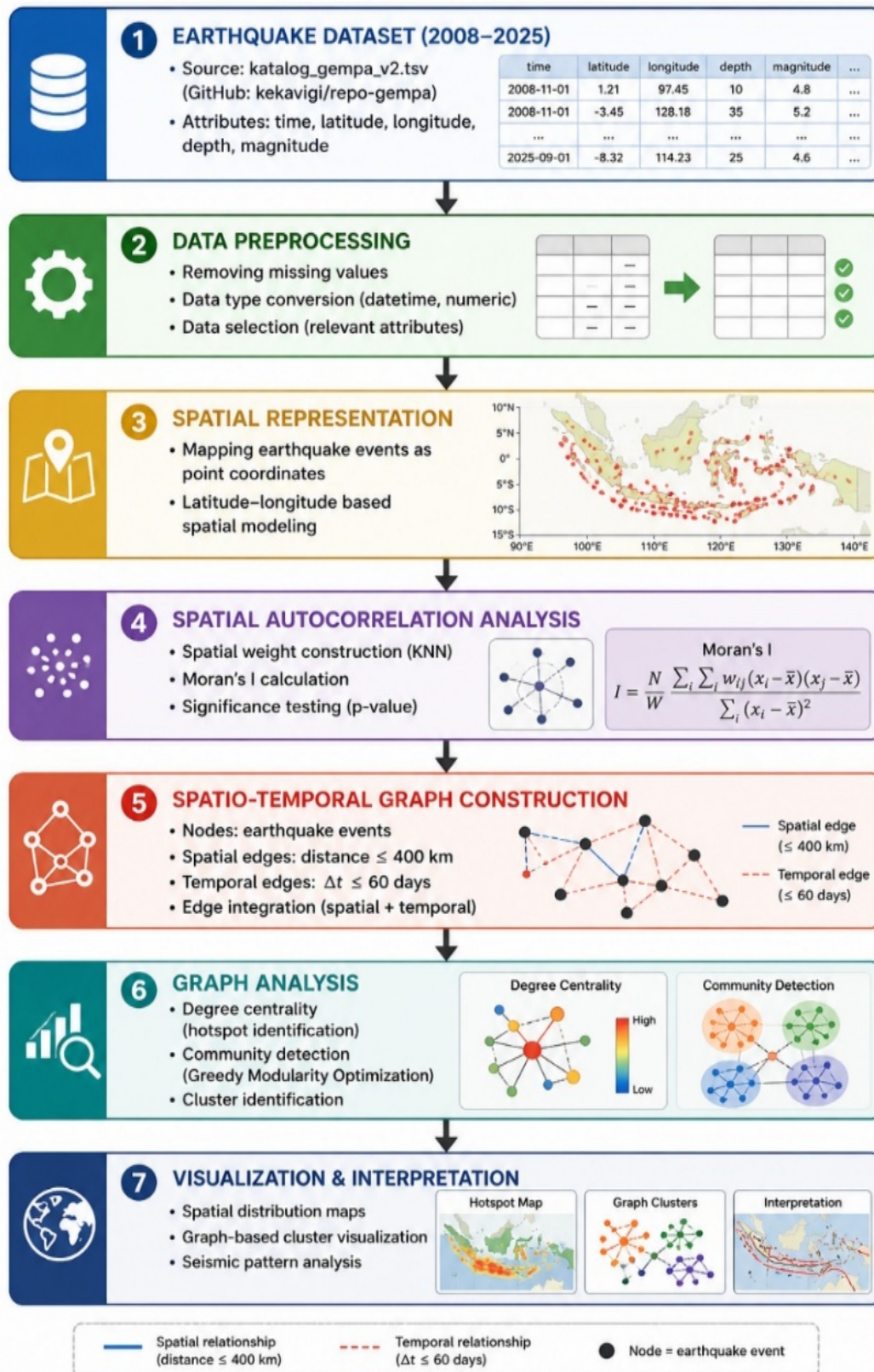


Figure 1. Research Stage

## 2) Spatial Representation

Each earthquake event is represented as a point in geographic space based on latitude and longitude coordinates. This representation allows for accurate analysis of spatial distribution and calculation of distances between earthquake events [26].

## 3) Spatial Autocorrelation Analysis

Spatial autocorrelation analysis was performed using Moran's I to measure the degree of spatial correlation between earthquake events [9], [10]. This method is used to determine whether the earthquake distribution is random or significantly clustered. In this study, earthquake magnitude was used as the variable in Moran's I computation to evaluate the spatial clustering tendency of seismic intensity.

$$I = \frac{N \sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum_i (x_i - \bar{x})^2} \quad (1)$$

Moran's I was calculated using Equation (1). Where:

N = total number of spatial observations (earthquake events)

$x_i$  = value of the variable at location i (e.g., earthquake magnitude)

$\bar{x}$  = mean of the variable

$w_{ij}$  = spatial weight between location i and j

W = total sum of spatial weights

The value of Moran's I ranges from -1 to +1. A value close to +1 indicates strong positive spatial autocorrelation (clustered pattern), while a value close to 0 indicates a random distribution, and a value near -1 indicates dispersion. To assess statistical significance, a hypothesis test was conducted under the null hypothesis of spatial randomness. The significance of Moran's I was evaluated using a p-value obtained through 999 random permutations to determine whether the observed spatial autocorrelation significantly differed from a random spatial distribution.

## 4) Spatio-Temporal Graph Construction

Graph modeling is performed by representing each earthquake event as a node, while the relationships between nodes are formed as edges based on two main criteria: spatial and temporal proximity. Spatial proximity is defined based on a maximum distance of  $\leq$

400 km, while temporal proximity is defined based on a maximum time interval of  $\leq 60$  days. This approach allows for the integration of spatial and temporal relationships within a single network structure [23]. The 400 km and 60-day thresholds were chosen to capture the interconnectedness of earthquake events on a broader regional scale. This is based on the characteristics of seismic activity, which is often influenced by tectonic systems with large spatial and temporal reach, so interactions between earthquake events are not always local. This threshold selection also refers to previous studies showing that earthquake interactions can occur within a radius of hundreds of kilometers and over time spans of weeks to months, particularly in active subduction systems.

To improve computational efficiency, this study did not use a direct pairwise comparison approach, which has quadratic complexity. Instead, the BallTree method was used to more efficiently search for spatial neighbors. This approach reduces computational complexity from  $O(n^2)$  to nearly  $O(n \log n)$ , making it more suitable for large-scale data processing. Furthermore, temporal relationships are established by first sorting the data by event time. This strategy allows for the implementation of an early stopping mechanism, where the search for node pairs is stopped when the time difference exceeds a specified threshold, thus improving the overall efficiency of graph construction.

### **5) Graph Analysis and Community Detection**

Graph analysis was performed using network metrics to understand the structure of connections between earthquake events. Degree centrality was used to identify highly connected nodes that may represent regions with intensified seismic interaction patterns [30]. Next, community detection was performed using Greedy Modularity Optimization (Clauset-Newman-Moore) to identify clusters within the graph [24], [25], [33]. This method aims to find groups of nodes with stronger internal connections than nodes outside their group. The quality of the detected community structure was evaluated using the modularity score ( $Q$ ), where higher values indicate stronger separation between communities within the graph.

### **6) Visualization**

The analysis results were visualized in map form to show the spatial distribution of earthquake clusters. This visualization was used to connect the model results with actual geological conditions.

Overall, the proposed workflow represents a graph-based analytical framework for integrating spatial autocorrelation and spatio-temporal network analysis in earthquake hotspot identification.

### 3. RESULTS AND DISCUSSION

#### 3.1. Data Preprocessing Results

The preprocessing stage produced a clean and consistent dataset for further analysis. Data with blank values in the main attributes was successfully removed, resulting in no missing values in the time, location, and magnitude variables. Data type conversion was also successfully performed, particularly for the time attribute to datetime format and for the numeric attribute to the appropriate type. Furthermore, data sampling was performed for computational efficiency, resulting in a data subset of 3,000 earthquake events that still represented the distribution of the entire dataset.

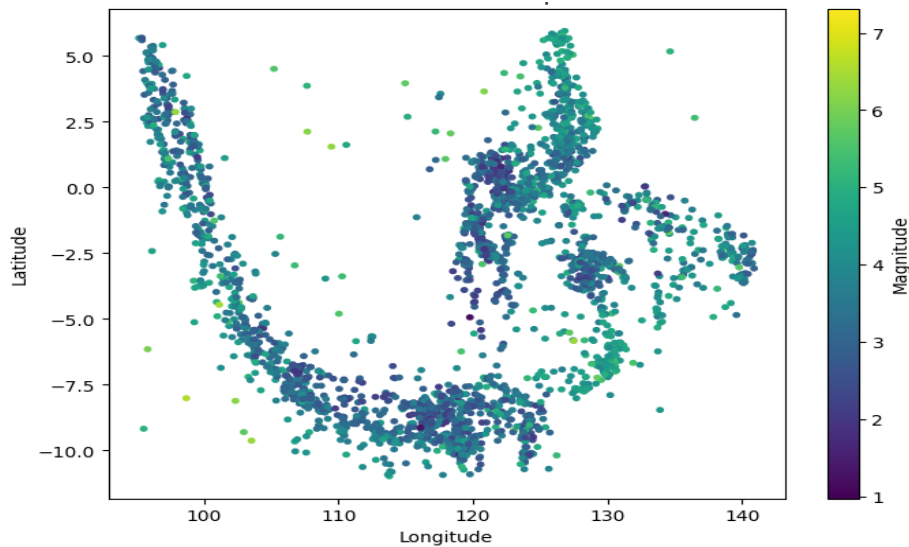
**Table 2.** Dataset Summary After Preprocessing

Description	Number of Data
Total initial data	131.833
After cleaning (missing key columns)	131.134
After datetime conversion and dropna	131.131
Data used (sample)	3.000

The preprocessing stage ensures that subsequent analysis is not affected by data inconsistencies. The use of sampling is also an important step to maintain a balance between computational efficiency and data representativeness.

#### 3.2. Spatial Representation (GIS) Results

The earthquake data was successfully converted into a spatial representation in the form of a GeoDataFrame. Each earthquake event is represented as a point object based on latitude and longitude coordinates. The transformation of the coordinate system from WGS84 to the metric projection system allows for more accurate distance calculations between points. As a result, all data can be used in distance-based analyses, such as spatial neighbor searches.



**Figure 2.** Spatial Distribution of Earthquake Events

### 3.3. Spatial Autocorrelation Analysis Results

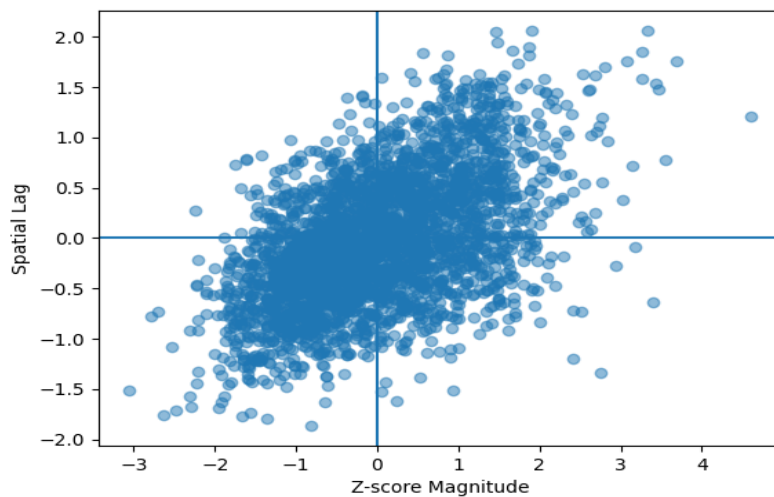
The spatial weight matrix was constructed using the K-Nearest Neighbors (KNN) approach with  $k=8$ , so that each earthquake event point is connected to its eight nearest neighbors based on geographic distance. The value of  $k=8$  was selected to balance local neighborhood representation and network connectivity stability in the spatial interaction analysis. This approach creates a local connectedness structure that represents the spatial interactions between earthquake events in the dataset. However, sensitivity analysis for different  $k$  values was not extensively evaluated and remains part of future work.

The resulting network structure shows that all nodes are in a single connected component, indicating an indirect spatial relationship between all earthquake events through the neighbor network. Based on this weight matrix, the Moran's  $I$  value was 0.3367 with a  $p$ -value of 0.001, as shown in Table 3. This value indicates significant positive spatial autocorrelation, suggesting that the earthquake distribution tends to form a clustered pattern.

**Table 3.** Moran's  $I$  Results

Metric	Value
Moran's $I$	0.3367
$p$ -value	0.001
Interpretation	Clustered

The distribution of values in the Moran scatter plot shows dominance in the high-high and low-low quadrants, indicating that locations with high magnitudes tend to be close to other locations with high magnitudes, and the same is true for low values. This pattern demonstrates spatial consistency in the distribution of earthquake events.



**Figure 3.** Moran Scatter Plot

The analysis results indicate that earthquake activity in the dataset is not random, but instead exhibits a strong spatially connected structure. This finding provides important justification for the use of a graph-based approach, as the presence of significant spatial autocorrelation indicates that earthquake events are not geographically independent. Therefore, network-based modeling is relevant for capturing interrelationships between events that cannot be fully explained through conventional spatial statistical analysis alone.

### 3.4. Spatio-Temporal Graph Construction Results

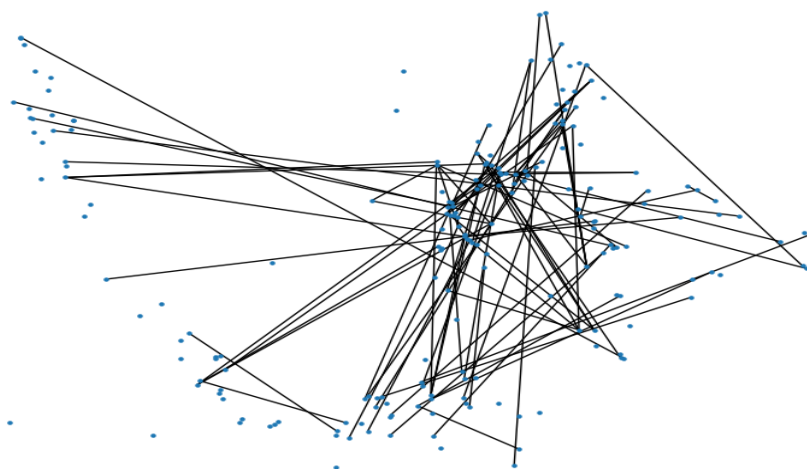
The spatio-temporal graph construction in this study produced a network representing relationships between earthquake events based on spatial and temporal proximity. Each earthquake event was represented as a node, while relationships between events were formed as edges using two criteria: a maximum geographic distance of 400 km and a maximum time interval of 60 days. The selected thresholds were intended to capture broader regional-scale seismic interactions commonly observed in active tectonic systems and subduction environments. This threshold selection also refers to previous studies showing that earthquake interactions may occur within hundreds of kilometers

and over periods ranging from weeks to months. Based on this construction process, a graph with the characteristics shown in Table 4 was obtained.

**Table 4.** Spatio-Temporal Graph Statistics.

Parameter	Value
Number of nodes	3000
Number of edges	22896
Max degree	77
AVG degree	9.59

The resulting graph density was relatively low ( $<0.01$ ), indicating that the network remained sparse despite exhibiting substantial connectivity at the regional scale. The relatively large number of edges indicates that most earthquake events are spatially and temporally related, forming a fairly connected network. The average degree value of 9.59 indicates that each earthquake event is, on average, associated with approximately nine other events within a given spatial and temporal radius. This suggests that seismic activity does not occur independently but rather forms interconnected patterns of interaction. Furthermore, the presence of nodes with a high degree (up to 77) indicates the role of certain nodes as hubs in the network. These nodes may represent potential seismic interaction centers within the network, where earthquake activity tends to be more concentrated and interconnected. This structure demonstrates that the earthquake system has complex network characteristics with heterogeneous connectivity patterns. To provide a visual representation of the network structure, a graph visualization is shown in Figure 4.

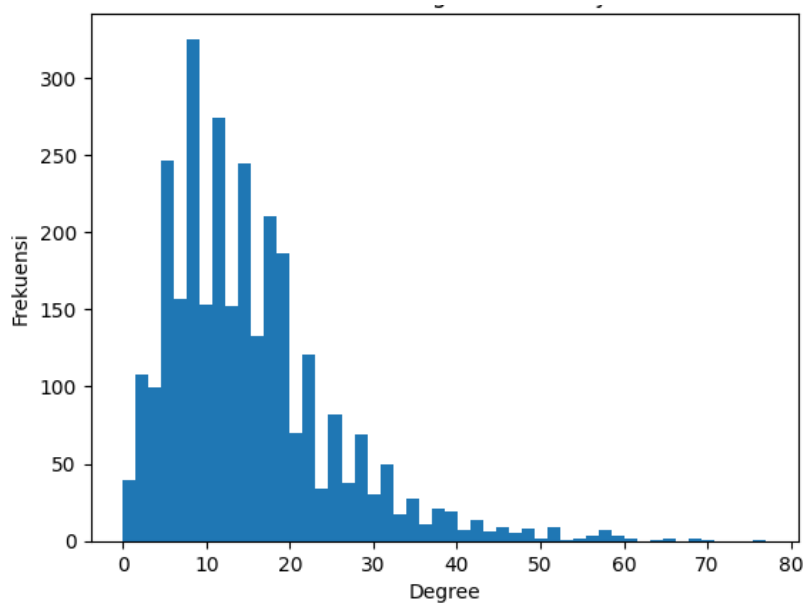


**Figure 4.** Spatio-Temporal Graph Visualization

### 3.5. Graph Analysis Results

#### 3.5.1 Degree Centrality

The degree distribution shows that most nodes have low to moderate connectivity, while a small number of nodes have much higher degree values, with a maximum value reaching 77. This indicates a heterogeneous network structure, with some nodes acting as hub nodes. The distribution of degree values in the graph is shown in Figure 5.



**Figure 5.** Degree Centrality Distribution

The degree distribution shows that most nodes have relatively low to moderate levels of connectivity, while there are a small number of nodes with much higher degree values. This indicates an imbalance in the connectivity distribution in the graph. Nodes with the highest degree values are interpreted as highly connected seismic interaction regions within the network. A summary of the nodes with the highest degree values is shown in Table 5.

**Table 5.** Nodes with the Highest Degree Centrality

Node	Degree	Latitude	Longitude
13	77	-2.056310	120.659355
414	70	-2.481291	121.188217
2321	69	-1.348556	120.577988
430	68	-4.075772	121.796822
626	66	-1.351748	120.576355

Nodes with high degree centrality values indicate earthquake events that are significantly connected to other events. This indicates that the location is an area of relatively intense seismic activity, where earthquakes do not occur singly but rather in an interconnected pattern. Thus, degree centrality can be used as an indicator of relatively concentrated seismic interaction patterns within spatio-temporal networks.

### 3.5.2 Community Detection (Clusters)

Community analysis on spatio-temporal graphs was performed using a greedy modularity optimization approach to identify cluster structures based on the connectivity between earthquake events. This method aims to maximize modularity by grouping highly connected nodes within the network, thereby more systematically representing earthquake linkage patterns. The graph, constructed using spatial proximity thresholds of  $\leq 400$  km and temporal proximity of  $\leq 60$  days, produced a network structure that was not random but instead demonstrated a clear community organization. Based on the community detection results, the network was divided into 14 main clusters representing the connectivity of earthquake events on a regional scale. The resulting community structure produced a modularity score of 0.7405, indicating strong intra-community connectivity and relatively weak inter-community connectivity among the detected spatio-temporal earthquake clusters. This high modularity value suggests that the graph partitioning effectively captured meaningful community patterns, since modularity scores above 0.3 are generally considered indicative of significant community structure in network analysis.

The distribution of nodes in each cluster is shown in Table 6. The largest cluster consisted of 937 nodes (Cluster 0), followed by other clusters with sizes of 730 (Cluster 1), 551 (Cluster 2), and 503 (Cluster 3). These differences in cluster size indicate variations in the level of seismic activity and connectivity between regions. Large clusters indicate areas with high earthquake interaction intensity, where earthquake events tend to be interconnected both spatially and temporally.

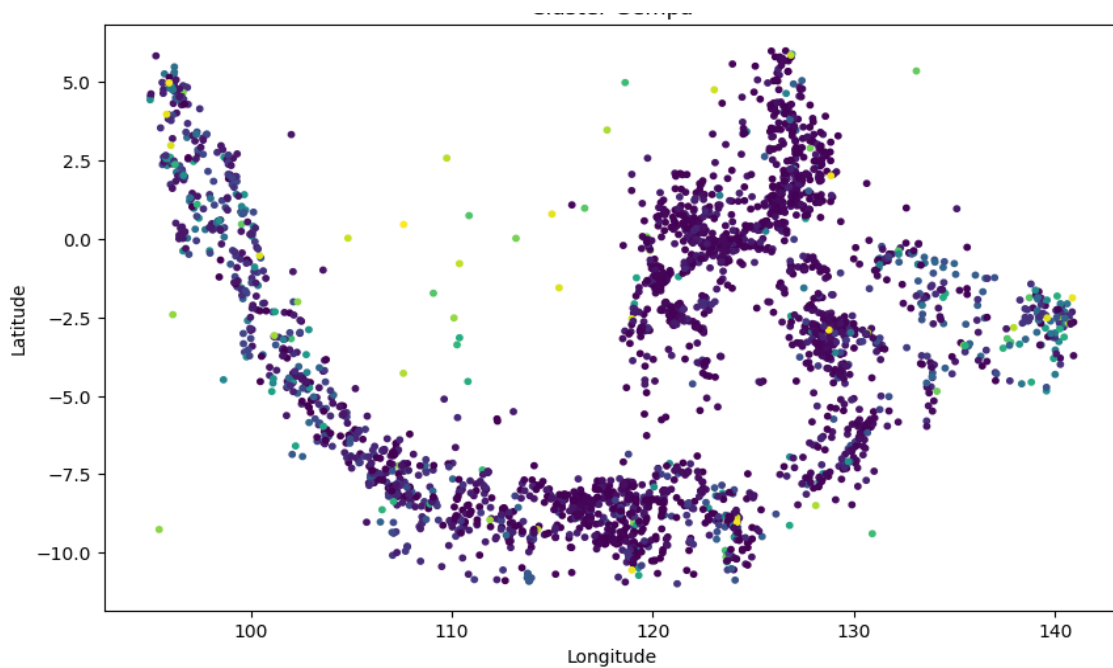
Furthermore, the resulting cluster structure indicates that earthquake events are not randomly distributed, but rather form a consistent grouping pattern. This indicates the presence of relatively homogeneous zones of seismic activity, likely related to regional geological conditions, such as subduction zones, tectonic plate boundaries, or active fault

systems in Indonesia. Thus, the spatio-temporal graph approach is not only capable of identifying interconnections between earthquake events but also provides a more comprehensive representation of seismic system dynamics than either a spatial or temporal approach alone.

**Table 6.** Distribution of Nodes in Each Cluster.

Cluster	Number of Node
0	937
1	730
2	551
3	503
4	156
5	81
6	33
7	3
8	1
9	1
10	1
11	1
12	1
13	1

The graph segmentation process identified 14 main clusters with varying sizes and connectivity structures. Several small or single-node clusters may represent relatively isolated seismic events with weak spatial-temporal connectivity under the selected threshold configuration. This condition may also indicate that some earthquake events did not form sufficiently strong relationships with neighboring events based on the applied spatial and temporal thresholds. This observation suggests that the selected threshold configuration may influence the formation of isolated communities and should be further evaluated through sensitivity analysis in future studies. Visualization of cluster distribution in geographic space is shown in Figure 6.



**Figure 6.** Spatio-Temporal Cluster Distribution Map

The visualization results show that the cluster distribution is not random, but rather forms relatively discrete and organized geographic clusters on a regional scale. The resulting cluster pattern indicates a concentration of activity in areas with high levels of seismic interaction, particularly along the Sunda Arc and eastern Indonesia, known as active tectonic zones, as shown in Table 7. Each cluster represents a collection of earthquake events that are strongly related in the graph, both based on spatial proximity (<400 km) and temporal proximity (<60 days). This indicates that earthquake events within a cluster tend to be interconnected in space and time, forming a cohesive network structure.

**Table 7.** Summary of Cluster Characteristics, Risk Level, and Implications

Cluster	Characteristics	Seismic Risk Level	Geological Interpretation	Potential Research and Monitoring Implications
0	Largest cluster with very high connectivity at a regional scale	High	Active subduction zone (Sunda arc / eastern Indonesia)	Potential priority region for further seismic interaction and regional monitoring studies.

Cluster	Characteristics	Seismic Risk Level	Geological Interpretation	Potential Research and Monitoring Implications
1	Large cluster with wide spatial distribution and high connectivity	High	Subduction system or plate interaction zone	May support future investigations of regional seismic connectivity patterns.
2	Medium-to-large cluster with consistent spatial patterns	High-Medium	Transition zone between subduction and active faults	Potential region for further analysis of recurring seismic interaction patterns
3	Medium-sized cluster with concentrated spatial distribution	Medium	Regional active fault zone	May indicate spatially concentrated seismic interaction requiring further regional investigation
4	Small-to-medium cluster with moderate connectivity	Medium	Local fault activity or crustal deformation	Suitable for localized seismic pattern monitoring and comparative regional analysis
5	Small cluster with limited interaction	Medium-Low	Minor tectonic activity	May reflect limited regional seismic interaction patterns requiring periodic observation
6	Small cluster with localized distribution	Low-Medium	Local fault systems or residual activity	Potentially associated with localized seismic interaction behavior for further study
7	Very small cluster (few nodes)	Low	Limited local seismic activity	Small-scale seismic grouping requiring additional validation and observation

Cluster	Characteristics	Seismic Risk Level	Geological Interpretation	Potential Research and Monitoring Implications
8	Very small cluster with minimal connectivity	Low	Background seismicity	May represent low-intensity background seismic interaction patterns
9	Single-node cluster	Low	Independent earthquake event	Represents an isolated seismic event within the network structure
10	Single-node cluster	Low	Background seismicity	May reflect independent background seismic activity
11	Single-node cluster	Low	Random activity	Indicates weakly connected seismic interaction within the observed network
12	Single-node cluster	Low	Isolated event	Represents isolated seismic behavior with limited network connectivity
13	Single-node cluster	Low	Background seismicity	May indicate low-connectivity background seismic activity

The differences in cluster size, from large to smaller, reflect variations in the intensity and complexity of seismic activity across regions. Larger clusters indicate areas with a high degree of event connectivity, while smaller clusters indicate more limited interaction patterns.

### 3.6. Deeper Geospatial Interpretation of Clusters and Network Structure

The identified clusters and high-degree nodes are not only structural properties of the graph but also reflect underlying geological processes. The spatial distribution of major clusters aligns closely with known tectonic features in Indonesia, particularly along the

Sunda Arc and eastern Indonesian tectonic regions. These areas are characterized by active subduction zones resulting from the interaction between the Indo-Australian and Eurasian plates.

Large clusters (e.g., Cluster 0 and Cluster 1) correspond to regions with high seismic interaction intensity, which are typically associated with subduction zones and major fault systems. The high connectivity observed in these clusters indicates that earthquake events in these regions are not isolated but are part of interconnected spatio-temporal sequences. This supports the theory that stress transfer and tectonic interactions play a significant role in triggering subsequent seismic events.

Furthermore, nodes with high degree centrality can be interpreted as potential seismic hotspots, where earthquake occurrences are highly interconnected both spatially and temporally. These hotspots are predominantly located in regions known for high tectonic activity, such as the western Sumatra subduction zone and parts of eastern Indonesia. This consistency between graph-based findings and known geological structures strengthens the validity of the proposed approach.

### **3.7. Comparison with Conventional Clustering Methods**

The comparison with K-Means and KDE in this study is conceptual rather than experimental, since no direct quantitative benchmark using identical optimization settings and evaluation metrics was performed. To further evaluate the effectiveness of the proposed spatio-temporal graph approach, a conceptual comparison with conventional clustering methods such as K-Means and Kernel Density Estimation (KDE) is provided.

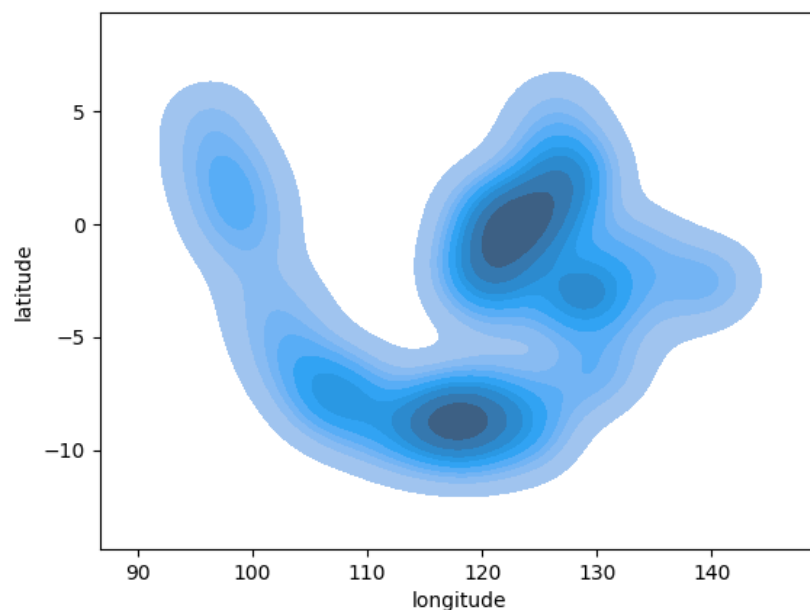
K-Means clustering groups earthquake events based on feature similarity (e.g., location, magnitude), but it assumes independence between observations and does not explicitly consider spatial or temporal relationships. As a result, K-Means may identify spatial groupings but fails to capture the underlying interaction structure between earthquake events. On the other hand, KDE is effective in identifying high-density regions or hotspots based on spatial distribution. However, KDE focuses solely on event density and does not represent the relational connectivity between events, particularly in the temporal dimension. In contrast, the proposed graph-based approach integrates both spatial and

temporal proximity, allowing for the explicit modeling of relationships between earthquake events. This enables the identification of not only dense regions but also interconnected structures, which are critical for understanding seismic dynamics.

The results show that while KDE identifies similar high-density regions, the graph-based method provides additional insights into the connectivity and interaction patterns between events. Likewise, compared to K-Means, the graph approach offers a more realistic representation of earthquake behavior by considering inter-event dependencies.

### 3.8. Spatial Visualization Results and Hotspot Identification

To strengthen the spatial analysis, we visualized the density of earthquake events using a density mapping approach. Degree centrality was chosen because it represents the intensity of direct connectivity between events, which is relevant in the context of earthquake spatio-temporal interactions. The results are shown in Figure 7.



**Figure 7.** Earthquake Density (Hotspot) Map

The density visualization shows areas with a high concentration of earthquake events, which can be identified as hotspots. These areas also correspond to nodes with high degree centrality values, demonstrating consistency between the spatial distribution-based approach and the network structure. Figure 6 and Figure 7 show that several graph-derived hotspot clusters spatially overlap with density-based hotspot regions,

particularly along the Sunda Arc and eastern Indonesia. However, several high-density regions do not necessarily exhibit high graph connectivity, indicating that density concentration and relational interaction structure capture different characteristics of seismic activity patterns.

### 3.9. Integration of Spatio-Temporal Results

The analysis results show a consistent relationship between spatial, temporal, and graph structure patterns in representing earthquake dynamics. A Moran's I value of 0.3367 (p-value 0.001) indicates significant spatial clustering, indicating a non-random distribution of earthquakes. Constructing a spatio-temporal graph with criteria of  $\leq 400$  km and  $\leq 60$  days resulted in a network with 3,000 nodes and 22,896 edges, segmented into 14 main clusters. The relatively small number of large clusters indicates strong connectivity at the regional scale.

Degree centrality analysis indicates the presence of nodes with high connectivity (up to 77) that act as interaction centers within the network. These nodes can be interpreted as candidate seismic activity hotspots, consistent with the density visualization results. The integration of spatial autocorrelation analysis and graph modeling allows for explicit identification of distribution patterns and relationships between earthquake events. Compared with conventional clustering methods such as K-Means, the graph approach is better able to capture spatial and temporal relationships simultaneously, thus providing a more comprehensive representation of seismic activity dynamics.

## 4. CONCLUSION

This study proposed a spatio-temporal graph approach integrated with spatial autocorrelation analysis to examine earthquake occurrence patterns in Indonesia. The results showed significant positive spatial autocorrelation (Moran's I = 0.3367; p-value = 0.001), indicating that earthquake events tend to form clustered spatial patterns. The constructed graph consisted of 3,000 nodes and 22,896 edges, revealing strong regional-scale connectivity and producing 14 major clusters with several highly connected nodes reaching degree 77, which may indicate potential seismic hotspots. In addition, the modularity score of 0.7405 demonstrated a strong community structure with high intra-community connectivity and low inter-community connectivity. These findings suggest

that the integration of spatial statistical analysis and graph-based modeling can provide a more comprehensive understanding of earthquake interaction patterns compared to single-perspective approaches. However, this study is limited by threshold sensitivity and the absence of weighted graph modeling based on seismic attributes such as earthquake magnitude. Future work may include sensitivity testing, weighted graph construction, geological data integration, and machine learning extensions. Overall, the proposed framework may support future seismic risk assessment, while operational validation remains part of future research.

## ACKNOWLEDGMENT

The authors would like to thank the data providers and the open-source community for providing access to the earthquake datasets used in this study. This support was instrumental in completing this research.

## REFERENCES

- [1] J. Baiesi and M. Paczuski, "Scale-free networks of earthquakes and aftershocks," *Phys. Rev. E*, vol. 69, no. 6, p. 66106, 2004, doi: 10.1103/PhysRevE.69.066106.
- [2] J. Davidsen, H. Grassberger, and M. Paczuski, "Earthquake recurrence as a record breaking process," *Phys. Rev. Lett.*, vol. 100, no. 6, p. 68501, 2008, doi: 10.1103/PhysRevLett.100.068501.
- [3] J. Zhuang, Y. Ogata, and D. Vere-Jones, "Statistical analysis of seismicity: testing clustering properties," *J. Geophys. Res.*, vol. 109, no. B5, 2004, doi: 10.1029/2003JB002879.
- [4] M. J. Werner, G. C. Beroza, and S. Wiemer, "Earthquake forecasting and clustering," *Bull. Seismol. Soc. Am.*, vol. 101, no. 2, pp. 574–584, 2011, doi: 10.1785/0120100241.
- [5] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 1996, pp. 226–231. doi: 10.5555/3001460.3001507.
- [6] J. Han, M. Kamber, and J. Pei, "*Data Mining: Concepts and Techniques*, 3rd ed. Waltham", MA, USA: Morgan Kaufmann, 2011. doi: 10.1016/C2009-0-61819-5.
- [7] B. W. Silverman, "*Density Estimation for Statistics and Data Analysis*." London, U.K.:

- Chapman & Hall, 1986. doi: 10.1007/978-1-4899-3324-9.
- [8] M. Rosenblatt, "Remarks on some nonparametric estimates of a density function," *Ann. Math. Stat.*, vol. 27, no. 3, pp. 832–837, 1956, doi: 10.1214/aoms/1177728190.
- [9] P. A. P. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, no. 1–2, pp. 17–23, 1950, doi: 10.1093/biomet/37.1-2.17.
- [10] L. Anselin, "Local indicators of spatial association—LISA," *Geogr. Anal.*, vol. 27, no. 2, pp. 93–115, 1995, doi: 10.1111/j.1538-4632.1995.tb00338.x.
- [11] S. Wiemer and M. Wyss, "Mapping spatial variability of seismicity parameters," *Science (80-. )*, vol. 271, no. 5245, pp. 150–152, 1996, doi: 10.1126/science.271.5245.150.
- [12] W. R. Tobler, "A computer movie simulating urban growth in the Detroit region," *Econ. Geogr.*, vol. 46, pp. 234–240, 1970, doi: 10.2307/143141.
- [13] M. Newman, "*Networks: An Introduction*". Oxford University Press, 2010.
- [14] A.-L. Barabasi, "*Network Science*." Cambridge, U.K.: Cambridge University Press, 2016. doi: 10.1017/CBO9781316216007.
- [15] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Phys. Rep.*, vol. 424, no. 4–5, pp. 175–308, 2006, doi: 10.1016/j.physrep.2005.10.009.
- [16] R. Albert and A.-L. Barabasi, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, pp. 47–97, 2002, doi: 10.1103/RevModPhys.74.47.
- [17] X. Zhang, Y. Chen, and Z. Li, "Earthquake clustering analysis based on graph theory," *Sci. Rep.*, vol. 10, p. 2045, 2020, doi: 10.1038/s41598-020-59340-y.
- [18] H. T. Nguyen, M. H. Pham, and Q. T. Nguyen, "Spatial clustering of earthquake events using machine learning," *Appl. Sci.*, vol. 11, no. 11, p. 5152, 2021, doi: 10.3390/app11115152.
- [19] Z. Li, X. Zhang, and Y. Wang, "Graph-based seismic analysis for earthquake pattern detection," *IEEE Access*, vol. 9, pp. 123456–123470, 2021, doi: 10.1109/ACCESS.2021.3065678.
- [20] D. L. Turcotte, "*Fractals and Chaos in Geology and Geophysics*, 2nd ed". Cambridge, U.K.: Cambridge University Press, 1997. doi: 10.1017/CBO9780511622775.
- [21] K. Kanamori, "The energy release in great earthquakes," *J. Geophys. Res.*, vol. 82, no. 20, pp. 2981–2987, 1977, doi: 10.1029/JB082i020p02981.
- [22] P. M. Shearer, "*Introduction to Seismology*, 2nd ed". Cambridge, U.K.: Cambridge University Press, 2009. doi: 10.1017/CBO9780511841510.
- [23] Y. Wang, Z. Liu, and J. Tang, "Spatio-temporal graph neural networks: A review," *IEEE Trans. Neural Networks Learn. Syst.*, 2022, doi: 10.1109/TNNLS.2022.3141234.

- [24] M. E. J. Newman, "Modularity and community structure in networks," *Proc. Natl. Acad. Sci.*, vol. 103, no. 23, pp. 8577–8582, 2006, doi: 10.1073/pnas.0601602103.
- [25] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E*, vol. 70, no. 6, p. 66111, 2004, doi: 10.1103/PhysRevE.70.066111.
- [26] M. de Smith, M. Goodchild, and P. Longley, "*Geospatial Analysis*, 6th ed". Winchelsea, U.K.: Winchelsea Press, 2018. doi: 10.5281/zenodo.5884356.
- [27] S. J. Rey and L. Anselin, "PySAL: A Python library of spatial analytical methods," *Geogr. Anal.*, vol. 39, no. 3, pp. 336–345, 2007, doi: 10.1111/j.1538-4632.2007.00697.x.
- [28] J. Leskovec, A. Rajaraman, and J. D. Ullman, "*Mining of Massive Datasets*." Cambridge University Press, 2014.
- [29] A. K. Vigilant, "Katalog Gempa Indonesia v2." 2025. doi: 10.5281/zenodo.10876543.
- [30] L. C. Freeman, "Centrality in social networks: Conceptual clarification," *Soc. Networks*, vol. 1, no. 3, pp. 215–239, 1978, doi: 10.1016/0378-8733(78)90021-7.
- [31] U. Brandes, "A faster algorithm for betweenness centrality," *J. Math. Sociol.*, vol. 25, no. 2, pp. 163–177, 2001, doi: 10.1080/0022250X.2001.9990249.
- [32] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numer. Math.*, vol. 1, pp. 269–271, 1959, doi: 10.1007/BF01386390.
- [33] M. E. J. Newman, "A measure of betweenness centrality based on random walks," *Soc. Networks*, vol. 27, no. 1, pp. 39–54, 2005, doi: 10.1016/j.socnet.2004.11.009.