

MALAYSIA'S DOMESTIC TOURISM NETWORK

(Rangkaian Pelancongan Domestik Malaysia)

MUHAMAD AIMAN HAKIM NOR KHALID, SAKHINAH ABU BAKAR
& FATIMAH ABDUL RAZAK*

ABSTRACT

Malaysia's domestic tourism can be as important as international tourism. This paper analyses Malaysia's domestic tourism network using graph theory. Data are obtained from the Domestic Tourism Survey 2023 by the Department of Statistics Malaysia. States are represented as vertices, while unweighted directed edges represent significant tourist flows between states. The study aims to identify tourism hubs through centrality measures and explore network communities using the Louvain algorithm and Fiedler spectral clustering. State performance is assessed using in-degree, out-degree, betweenness, closeness, and eigenvector centralities. States with higher in-degree centrality are shown to be popular destinations. The study also examines the Pearson correlation between centrality metrics and each state's gross domestic product (GDP) to assess the relationship between economic activity and domestic tourism. Findings show a strong positive correlation between betweenness centrality and GDP. Both the Louvain and Fiedler methods reveal clusters that are aligned with the spatial distribution of the states. These network insights provide useful information for tourism authorities in identifying key destination hubs, strengthening inter-state tourism corridors, and supporting regional tourism planning.

Keywords: network; centralities; community detection; domestic tourism

ABSTRAK

Pelancongan domestik Malaysia adalah sama penting seperti pelancongan antarabangsa. Artikel ini menganalisis rangkaian pelancongan domestik Malaysia menggunakan teori graf. Data yang digunakan diperoleh daripada Survei Pelancongan Domestik 2023 oleh Jabatan Perangkaan Malaysia. Negeri-negeri ditakrifkan sebagai bucu, manakala sisi berarah tanpa pemberat mewakili aliran pelancong yang signifikan antara negeri. Kajian ini bertujuan untuk mengenal pasti hab pelancongan melalui ukuran pemusatan serta meneroka komuniti rangkaian menggunakan algoritma Louvain dan pengelompokan spektrum Fiedler. Prestasi negeri dinilai berdasarkan pemusatan darjah masuk, darjah keluar, pengantaraan, kedekatan, dan vektor eigen. Negeri yang mempunyai pemusatan darjah masuk yang lebih tinggi dikenal pasti sebagai destinasi pelancongan yang popular. Kajian ini juga meneliti korelasi Pearson antara metrik pemusatan dan keluaran dalam negeri kasar (KDNK) setiap negeri bagi menilai hubungan antara aktiviti ekonomi dan pelancongan domestik. Dapatan kajian menunjukkan wujud korelasi positif yang kuat antara pemusatan pengantaraan dan KDNK. Kedua-dua kaedah Louvain dan Fiedler mendedahkan kelompok yang selari dengan kedudukan geografi negeri-negeri. Dapatan rangkaian ini memberikan maklumat yang berguna kepada pihak berkuasa pelancongan dalam mengenal pasti hab destinasi utama, mengukuhkan koridor pelancongan antara negeri, serta menyokong perancangan pelancongan serantau.

Kata kunci: rangkaian; pemusatan; pengesanan komuniti; pelancongan domestik

1. Introduction

The tourism sector is an important contributor to the economy (Mosbah *et al.* 2014). In Malaysia, it contributed at least 15% to the country's gross domestic product (GDP) in 2023 (Department

of Statistics Malaysia 2023c). Although tourism research and development initiatives have largely focused on international travel, domestic tourism also plays a significant role. It supports local businesses and distributes economic activity across different states.

Graph theory provides a useful approach for understanding travel movements within a country. In this approach, places such as cities or states are represented as vertices, while tourist movements between them are represented as edges. This enables the construction of a network and allows the connectivity among states to be examined. Previous studies have applied network analysis in tourism, but many have focused on other countries or on international travel. Other studies have examined the role of factors such as geographical distance, income levels, and transportation accessibility in influencing travel behaviour. However, relatively little research has explored the structural characteristics of domestic tourism networks within Malaysia.

To address this gap, this study uses data from the Domestic Tourism Survey 2023 by the Department of Statistics Malaysia (Department of Statistics Malaysia 2024). A domestic tourism network is constructed in which each state is represented as a vertex. A directed edge is added when there is a significant flow of tourists from one state to another.

2. Literature Review

Previous researchers have applied network analysis to study tourism performance. Shih (2006) recommended tourist services and facilities for 16 destinations based on their network positions. McKercher and Lau (2008) noted that tourist movement patterns are influenced by territoriality, journeys per day, stops per journey, commercial day tour participation, extra-destination travel, and multi-stop journey patterns. Miguéns and Mendes (2008) analysed worldwide tourist arrivals using a directed network. Guo *et al.* (2015) identified fluctuation patterns in inbound tourist flows in China. González-Díaz *et al.* (2015) examined changes in Spanish regional tourism networks by accommodation type, while Su Chung *et al.* (2017) studied tourist movement patterns in Europe through blog analysis.

Previous studies have also examined tourism networks using centrality measures. Lee *et al.* (2013) studied the degree, betweenness, and eigenvector centralities of 43 villages in Korea to determine the influence of core villages on surrounding tourism activities. Furthermore, Seok *et al.* (2021) examined international tourism flows from the perspective of social network analysis. The study revealed groupings based on continent, language, culture, economy, and population. A key finding was that out-degree, in-degree, betweenness, and eigenvector centralities were correlated with GDP and population. The study also found that the United States remained the most central country, followed by Canada, Australia, and Belgium. Meanwhile, China gained centrality, particularly in terms of betweenness, between 2002 and 2014.

In 2024, Zhang *et al.* (2024) applied machine learning and social network analysis to investigate the performance of border tourism in the Qinghai-Tibet Plateau, which includes China and neighbouring countries such as Pakistan, Nepal, and India. The study identified two key sub-networks, namely Xinjiang and Tibet, with Gar and Hotan serving as central connecting vertices. Tourism hotspots were found to cluster around transport corridors, reflecting the importance of infrastructure in shaping tourist flows. In addition, the study identified underdeveloped areas such as Zhongba, Burang, and Yadong, where targeted investment could further support tourism growth.

In the context of Malaysia's tourism, researchers have primarily focused on the relationship between tourism and the economy (Mura & Tavakoli 2014). In addition, scholars have placed greater emphasis on the sustainable development of tourism in Malaysia and the preservation of the local natural environment (Cao *et al.* 2022). Tourism networks have also been studied in the context of Chinese outbound tourists, where the tourism flow network was constructed using data on tourists from China (Cao *et al.* 2022). The study found that popular destinations among Chinese tourists include W.P. Kuala Lumpur, Melaka, Pulau Pinang, Langkawi, Tawau, and Semporna. Community detection was also performed to group cities in Malaysia based on

Chinese tourist flows. The findings showed the existence of communities based on city tours and island tours. Based on this network analysis, the study recommended enhancing the cultural and quality aspects of urban tourism products. However, studies on Malaysia's tourism network from the perspective of domestic tourism remain limited.

Therefore, it is of interest to analyse the performance of individual states in Malaysia's domestic tourism sector. This paper focuses on Malaysia's domestic tourism network. Previous research has motivated the analysis of state centralities in Malaysia's domestic tourism network, as well as the clustering of states through community detection to observe the resulting patterns. This study aims to identify the so-called hub or key state in Malaysia's domestic tourism network.

3. Materials and Methods

3.1. Dataset

This paper analyses data from the Domestic Tourism Survey (DTS) 2023 provided by the Department of Statistics Malaysia (Department of Statistics Malaysia 2024). The data consist of the number of tourists travelling from each state to every other state. The Department of Statistics Malaysia provides data on incoming and outgoing domestic tourists for 16 states in Malaysia, with the Federal Territories, or Wilayah Persekutuan (W.P.), divided into three regions, namely W.P. Kuala Lumpur, W.P. Putrajaya, and W.P. Labuan. Specifically, let m_{ij} denote the number of tourists from state i to state j .

According to the Department of Statistics Malaysia, a domestic visitor is defined as a Malaysian citizen or a resident who has lived in Malaysia for at least one year, including non-citizens, and who is either a tourist or an excursionist. A tourist refers to a person who travels within Malaysia for at least 24 hours, or equivalently one night, while an excursionist refers to a person who travels within Malaysia for less than 24 hours. A domestic tourism trip refers to a trip to a main destination within the visitor's country of residence, which in this study refers to Malaysia.

3.2. Adjacency Matrix

A network G consists of a set of vertices V and a set of edges E , written as $G = (V, E)$ (Razak & Expert 2021; Razak & Zamzuri 2021). Similarly, Malaysia's domestic tourism network can be defined as a graph $G = (V, E)$, where V is the set of states and federal territories in Malaysia, and E is the set of directed edges representing significant tourism movements from state i to state j . In this study, $|V| = 16$, indicating that the network consists of 16 vertices, each representing a state or federal territory, as defined in Eq. 1.

$$V = \{v_1, v_2, \dots, v_{16}\}. \quad (1)$$

To define E , the significance of the number of tourists travelling from state i to state j , denoted by m_{ij} , is considered. Let p denote the probability that a tourist departing from a given state chooses any particular destination state. Since tourists from one state may travel to any of the other 15 states or federal territories, a uniform probability assumption gives

$$p = \frac{1}{15}. \quad (2)$$

This means that, on average, each state is expected to contribute $\frac{1}{15}$ of its outgoing tourists to any other state. Next, let $q(i)$ be the total number of outgoing tourists from state i to all other

states. This is given by

$$q(i) = \sum_{j \in V, j \neq i} m_{ij}. \quad (3)$$

Thus, the expected number of tourists travelling from state i to state j under the uniform probability assumption is given by

$$\eta_{ij} = pq(i) = \frac{1}{15}q(i), \quad i \neq j. \quad (4)$$

State i is connected to state j by a directed edge if m_{ij} is considered significant. In general, the expected value η_{ij} may be used as the threshold (Kreinovich *et al.* 2008). However, in this study, a slightly lower threshold is adopted because some outgoing tourist flows are close to the expected value and may still represent meaningful interstate tourism movements.

Therefore, the threshold for significant outgoing tourist flow, denoted by $\mu(i)$, is calculated as 90% of the expected value, as follows:

$$\mu(i) = 0.90\eta_{ij}. \quad (5)$$

Using data from the Department of Statistics Malaysia (Department of Statistics Malaysia 2024), for each state i , any value of m_{ij} that is less than the threshold $\mu(i)$ in Eq. 5 is assigned a value of 0, indicating the absence of an edge. The remaining edges represent significant interstate tourism flows, as defined in Eq. 6.

$$E = \left\{ \overrightarrow{v_i, v_j} \mid m_{ij} \geq \mu(i), i, j \in V, i \neq j \right\}. \quad (6)$$

Based on this threshold, Malaysia's domestic tourism network is represented as a directed and unweighted graph with $|E| = 95$. The graph G is represented by an adjacency matrix A with entries a_{ij} , as shown in Eq. 7.

$$a_{ij} = \begin{cases} 1, & \text{if a significant tourist flow from state } i \text{ to state } j \text{ exists,} \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

In Section 5, the robustness of the 90% threshold is examined. Specifically, a sensitivity analysis is conducted by varying the parameter α around the baseline value of 0.90. In addition to $\alpha = 0.90$, seven alternative thresholds are considered, namely

$$\alpha \in \{0.75, 0.80, 0.85, 0.925, 0.95, 0.975, 1.00\}. \quad (8)$$

4. Results and Discussion

4.1. Malaysia's Domestic Tourism Network

4.1.1. Network visualisation

Figure 1 shows Malaysia's domestic tourism network. The network can be interpreted as a directed graph, where the arrows represent the direction of significant tourist flows from one state to another.

For example, significant tourist flows can be observed from W.P. Labuan to W.P. Putrajaya and Sabah, as well as from Terengganu to Kelantan and Pahang. Several states, such as W.P. Kuala Lumpur, Selangor, Johor, Perak and W.P. Putrajaya, are located near the centre of the network. This may indicate that these states have relatively high centrality values, including

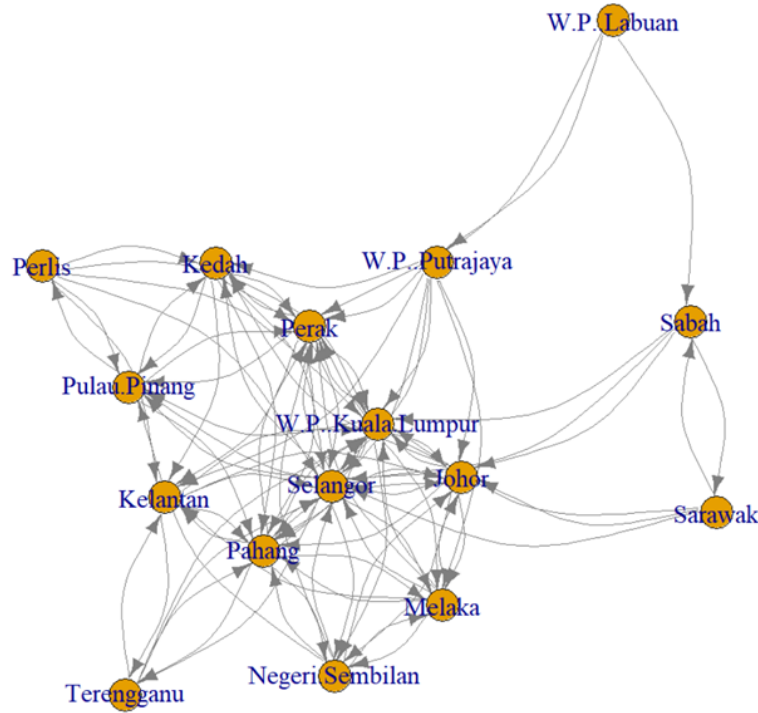


Figure 1: Malaysia's domestic tourism network

degree, closeness and betweenness centralities. W.P. Kuala Lumpur is particularly important because it is the capital city and contains various attractions, shopping centres and cultural sites.

In contrast, states such as Perlis, Terengganu and W.P. Labuan appear as more peripheral vertices due to their positions on the outer part of the network. This may suggest lower centrality values compared with states located closer to the centre of the network. W.P. Labuan, Sabah and Sarawak have fewer visible connections, suggesting more specialised tourism activities that may include ecotourism, adventure tourism and cultural experiences unique to Borneo (Ministry of Tourism Malaysia 2021). The northern states, such as Pulau Pinang, Perlis and Kedah, also appear closely connected, potentially reflecting a shared tourism circuit involving historical sites, beaches and cultural heritage (Ministry of Tourism Malaysia 2021).

Figure 1 also illustrates the in-degree and out-degree patterns of the states. The in-degree, denoted by $k^{\text{in}}(i)$, refers to the number of edges pointing inward to vertex i , whereas the out-degree, denoted by $k^{\text{out}}(i)$, refers to the number of edges pointing outward from vertex i . For instance, Sarawak has an in-degree of $k^{\text{in}}(\text{Sarawak}) = 1$ and an out-degree of $k^{\text{out}}(\text{Sarawak}) = 4$. This indicates that Sarawak receives a significant tourism connection from one state, while sending tourists to four different states.

Similarly, Selangor appears to function as a central hub, with relatively high in-degree and out-degree values. This suggests substantial tourist exchanges with many other regions. States such as W.P. Kuala Lumpur and W.P. Putrajaya, which are administrative and economic centres, also show important in-degree and out-degree patterns, reflecting their roles as major tourist destinations and transit points.

In contrast, more peripheral states such as Perlis and Terengganu show lower in-degree and out-degree values, indicating fewer direct tourism connections. This may reflect their focus on niche tourism markets or less extensive tourism connectivity compared with more central states (Ministry of Tourism Malaysia 2021). Overall, the network structure, with densely connected central states and more sparsely connected peripheral states, suggests a hierarchical pattern

of domestic tourism flows within Malaysia. This highlights the importance of central states in facilitating tourism movement and indicates the potential for developing under-connected regions to improve overall tourism connectivity and regional economic benefits.

The density of a network measures how closely connected the network is by comparing the actual number of edges with the maximum possible number of edges. In Malaysia's domestic tourism network, a density of approximately 0.396 indicates that about 40% of all possible interstate tourism flows are significant. This value suggests a moderate level of interaction among states in terms of domestic tourist movement. It also indicates that Malaysia has a reasonably connected domestic tourism structure, although not all possible interstate flows are equally strong. The moderate density may also suggest the presence of clustering among groups of states, which is further examined through the global clustering coefficient in Section 4.2.2.

4.1.2. Centrality

This section discusses the results obtained from the centrality measurements of Malaysia's domestic tourism network. The centrality measurements used in this article are in-degree, out-degree, closeness centrality, betweenness centrality and eigenvector centrality. Since Malaysia's domestic tourism network is a directed network, the direction of edges is considered in the definition of each centrality.

Table 1 shows the in-degree and out-degree centralities of states in the tourism network, while Figure 2 shows the network plot with different vertex sizes based on in-degree centrality. In-degree centrality indicates the popularity of a state. A higher in-degree centrality means that the state receives tourists from more states compared with states with lower in-degree values. The states with the highest in-degree centralities are Selangor, followed by W.P. Kuala Lumpur, Pahang and Perak. Selangor receives tourists from 13 states, while W.P. Kuala Lumpur receives tourists from 12 states. It is not surprising that Selangor and W.P. Kuala Lumpur are at the top of the list, as these two states are located in the central part of Malaysia. Moreover, Kuala Lumpur International Airport (KLIA) is located in Sepang, Selangor, while KL Sentral, the largest railway station, is located in W.P. Kuala Lumpur. In addition, W.P. Kuala Lumpur is highly visited because it is the capital of Malaysia and serves as a major transportation hub for tourists (Cao *et al.* 2022).

Table 1: In-degree and out-degree centralities

State	In-degree centrality	Out-degree centrality
Selangor	13	9
W.P. Kuala Lumpur	12	7
Pahang	11	8
Perak	11	6
Johor	9	8
Kelantan	7	5
Pulau Pinang	7	4
Kedah	6	6
Melaka	6	7
Negeri Sembilan	5	6
Terengganu	3	4
Sabah	2	4
Perlis	1	6
Sarawak	1	4
W.P. Putrajaya	1	8
W.P. Labuan	0	3

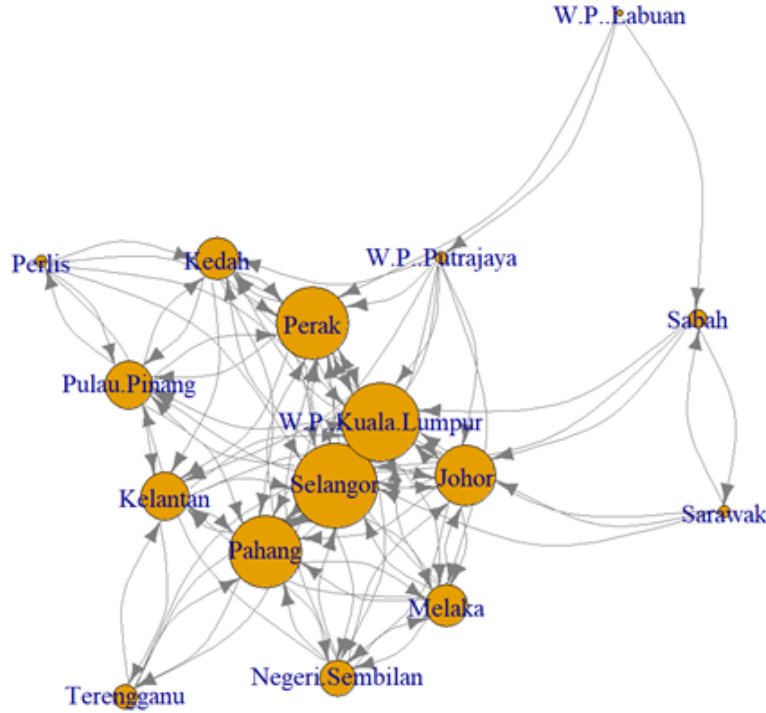


Figure 2: Malaysia's domestic tourism network with vertex size based on in-degree centrality

It is also interesting to observe the next states in the ranking. Pahang is ranked third, followed by Perak. Both Pahang and Perak receive tourists from 11 different states. A simple justification for the high in-degree centrality of Pahang is that its top tourist destinations in 2023 include Genting Highlands, Cameron Highlands and Pantai Teluk Cempedak (Department of Statistics Malaysia 2024). Perak also has high in-degree centrality, not because of the presence of mega malls, but due to its natural tourist attractions such as Pulau Pangkor, Teluk Batik and Lumut Waterfront (Department of Statistics Malaysia 2024). Therefore, based on in-degree centrality, Pahang and Perak can be considered favourite and popular tourism choices in Malaysia, in addition to the leading states of Selangor and W.P. Kuala Lumpur.

The in-degree distribution shown in Figure 3a provides insights into the connectivity patterns of Malaysia's domestic tourism network. The most common in-degree, or mode, of the distribution is $k^{\text{in}} = 1$, indicating that most states receive tourists from exactly one other state. The long tail of the distribution, with a maximum in-degree of 13 for a single state, suggests a power-law-like behaviour. This is a characteristic of networks in which a few states act as major hubs and receive disproportionately high tourism connections compared with other states.

Conversely, the out-degree distribution shown in Figure 3b appears to approximate a binomial distribution with $N = 15$ and $p = 0.396$. This indicates a more evenly distributed network, where states send tourists to several other states with varying probabilities. The value of $p = 0.396$ for the binomial distribution parameter is derived from the average out-degree, $\langle k^{\text{out}} \rangle$, divided by the maximum out-degree, $k_{\text{max}}^{\text{out}}$. The prominent modes of this distribution at $k^{\text{out}} = 4$ and $k^{\text{out}} = 6$ suggest that it is most common for states to send tourists to around four to six other states. This further emphasises the relatively dispersed nature of outgoing tourist flows in this network compared with the in-degree distribution.

Closeness centrality (CC) measures how close a vertex is to the rest of the network. In a

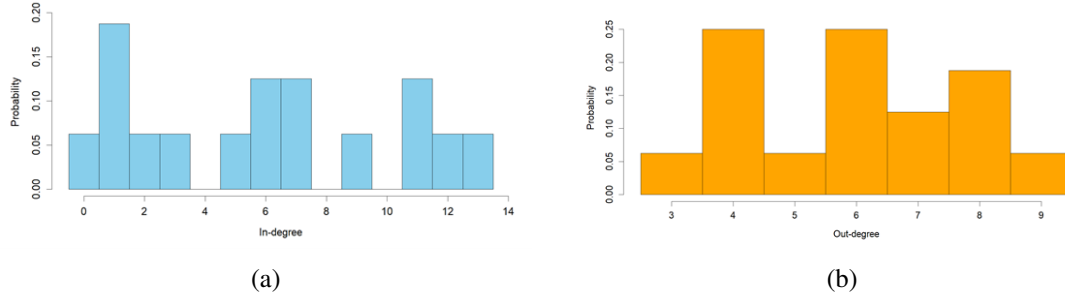


Figure 3: Plots of (a) in-degree and (b) out-degree distributions

graph $G = (V, E)$, a path P is defined as a sequence of vertices

$$P = (v_1, v_2, \dots, v_{k-1}, v_k), \quad (9)$$

such that each consecutive pair of vertices (v_i, v_{i+1}) is connected by an edge in E . The length of a path is defined as the number of edges contained in the path. The distance between two vertices u and v is the length of the shortest path connecting them in the graph (Goddard & Oellermann 2011).

For a directed network, the in-directed CC and out-directed CC are defined in Eq. 10 and Eq. 11, respectively:

$$CC_{in}(i) = \frac{1}{\sum_j d_{in}(i, j)}. \quad (10)$$

$$CC_{out}(i) = \frac{1}{\sum_j d_{out}(i, j)}. \quad (11)$$

Here, $\sum_j d_{in}(i, j)$ is the sum of distances from all other vertices to vertex i , considering incoming edges. Similarly, $\sum_j d_{out}(i, j)$ is the sum of the shortest distances from vertex i to all other vertices, considering outgoing edges (Putman *et al.* 2020; Abu Bakar *et al.* 2013). If no shortest path exists from vertex j to vertex i , then $d_{in}(i, j) = N$. Similarly, if no shortest path exists from vertex i to vertex j , then $d_{out}(i, j) = N$, where N is the total number of vertices.

For example, to calculate the out-directed CC of a state, the outgoing edges from that state are considered as the starting point. Then, the distances of tourist flows from that state to all other states are summed as in Eq. 11. If no shortest path exists from state i to state j , then $d_{out}(i, j) = 16$, since the total number of vertices is 16. This adjustment lowers the CC score to reflect disconnected paths (Freeman 1978). Thus, a vertex that is closer to the rest of the network will have a higher closeness centrality. If no path exists, the closeness centrality will be low.

Table 2 shows the out-directed closeness centrality of states in the tourism network, while Figure 4 shows the network plot with different vertex sizes based on out-directed closeness centrality. Out-directed CC provides interesting insights. W.P. Labuan, Sabah and Sarawak are ranked as the top three states. These three states, which are located in East Malaysia, are ranked higher than all states in Peninsular Malaysia. This indicates that tourists from East Malaysia make a significant contribution to Malaysia's domestic tourism by serving as important sources of tourist flows.

On the other hand, betweenness centrality (BC), as shown in Eq. 12, measures the impor-

Table 2: Out-directed closeness centrality

State	Out-directed closeness centrality
W.P. Labuan	0.500
Sabah	0.268
Sarawak	0.268
W.P. Putrajaya	0.231
Selangor	0.195
Johor	0.192
Pahang	0.190
W.P. Kuala Lumpur	0.190
Perak	0.188
Melaka	0.188
Negeri Sembilan	0.188
Kedah	0.188
Perlis	0.188
Pulau Pinang	0.183
Kelantan	0.183
Terengganu	0.181

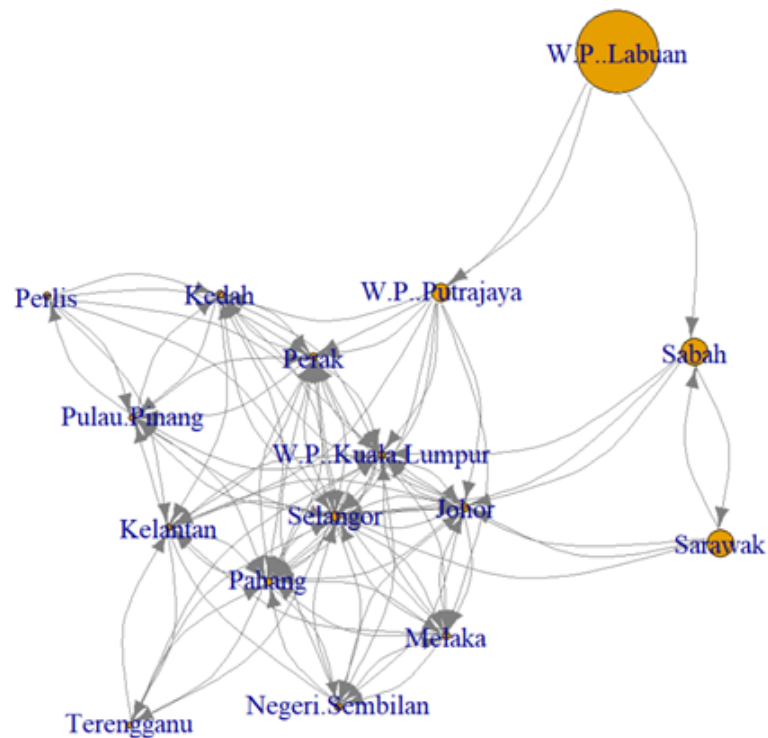


Figure 4: Malaysia's domestic tourism network with vertex size based on out-directed closeness centrality

tance of a vertex based on its position along the shortest paths between other pairs of vertices.

$$BC(i) = \sum_{s,t \in V} \frac{\sigma(s, t | i)}{\sigma(s, t)}, \quad (12)$$

where $\sigma(s, t)$ is the number of shortest paths from vertex s to vertex t , while $\sigma(s, t | i)$ is the number of shortest paths from vertex s to vertex t that pass through vertex i (Brandes 2008). Note that $\sigma(s, t)$ depends on the direction of edges, and (s, t) represents every possible pair of vertices in the graph, excluding vertex i . Since the direction of edges is considered, the number of possible shortest paths between any pair of vertices becomes limited.

In the domestic tourism network, states with high BC scores play crucial roles as bridges within the network. This means that tourists may need to pass through these states when travelling to other destinations. Table 3 shows the BC values of states in the tourism network, while Figure 5 shows the network plot with different vertex sizes based on BC. Selangor has the highest BC score, with a value of 0.110. This suggests that many tourist flows pass through Selangor when moving between other states, making it a crucial hub for domestic tourism flow. Selangor's central location in Peninsular Malaysia, together with its extensive infrastructure, major highways and proximity to W.P. Kuala Lumpur, makes it a natural crossroads for tourists (Ramlan & Rodzi 2023). It is also well connected to various regions, both within Peninsular Malaysia and to East Malaysia, through major airports and transportation networks.

Although some states are not geographically central, they remain important in terms of BC. For example, Johor, located in the southern part of Peninsular Malaysia, and Pulau Pinang, located in the northern part, are not centrally positioned within the country. Despite this, Pulau Pinang ranks second and Johor ranks fifth in terms of BC. These high scores indicate that Pulau Pinang and Johor are not only key destinations but also significant transit hubs in tourist flows. This may be due to their incoming connectivity, with Pulau Pinang and Johor receiving flows from seven and nine other states, respectively, making them pivotal points for tourists travelling across the region.

The top eight states with high in-degree and BC values, excluding Pulau Pinang, share several common characteristics. Geographically, each of these states is bordered by at least three neighbouring states, which facilitates their role as key vertices in the tourism network. Pahang stands out among these states due to its large area and adjacency to the highest number of states, namely Selangor, Negeri Sembilan, Johor, Perak, Terengganu and Kelantan. This extensive connectivity likely contributes to Pahang's role in tourist flows, as indicated by its BC score of 0.066. The significant number of connections allows these states to act as both destinations and transit points within the network.

Interestingly, Sabah is also highlighted as a significant state in terms of BC, despite its geographical separation from Peninsular Malaysia. Sabah ranks ninth, with a BC score of 0.012, surpassing five other states in Peninsular Malaysia. This suggests that although Sabah has relatively low in-degree, $k^{\text{in}}(\text{Sabah}) = 2$, and out-degree, $k^{\text{out}}(\text{Sabah}) = 4$, it still serves as an important transit point in the domestic tourism network. Sabah's role as a transit hub may be influenced by its connections to key air and sea routes, which facilitate the movement of tourists between different regions of Malaysia, including those travelling to or from Peninsular Malaysia and neighbouring countries (Samuel & Nor Ahmad 2020).

Perlis, the smallest state in Malaysia, has a very low BC score of 0.004, indicating its minimal role as a bridge in the tourism network. As the northernmost state in Peninsular Malaysia and a state bordering Thailand, Perlis is often visited directly for its natural and agricultural attractions (Ministry of Tourism Malaysia 2021). However, it does not serve as a significant transit point for tourists travelling between other Malaysian states.

Sarawak and W.P. Labuan, on the other hand, have BC scores of 0, indicating that these states do not serve as transit hubs within the domestic tourism network. This means that tourists visiting Sarawak and W.P. Labuan do not typically pass through these states when travelling

to other destinations within Malaysia. Instead, Sarawak and W.P. Labuan function more as endpoints in the tourism network. Their locations on the island of Borneo, separated from Peninsular Malaysia by the South China Sea, limit their roles as bridges in the domestic tourism network. Sarawak is known for its unique cultural heritage, diverse ecosystems and outdoor adventure activities, making it a distinct destination. W.P. Labuan is the only state with an in-degree equal to 0, which also explains its BC score of 0.

Table 3: Betweenness centrality

State	Betweenness centrality
Selangor	0.110
Pulau Pinang	0.070
Pahang	0.066
W.P. Kuala Lumpur	0.063
Johor	0.050
Perak	0.047
W.P. Putrajaya	0.022
Kelantan	0.016
Sabah	0.012
Melaka	0.010
Kedah	0.010
Negeri Sembilan	0.004
Perlis	0.004
Terengganu	0.002
W.P. Labuan	0
Sarawak	0

In addition, it is also useful to analyse the eigenvector centrality of each state in the tourism network. The eigenvector centrality of vertex i is given by x_i , which is the i^{th} entry of the leading eigenvector $\mathbf{x}$ in Eq. 14. This vector satisfies the discrete eigenvalue problem in Eq. 13.

$$A\mathbf{x} = \lambda_1\mathbf{x}, \quad (13)$$

where λ_1 is the largest eigenvalue of the adjacency matrix A defined in Eq. 7.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}. \quad (14)$$

A state with high eigenvector centrality is not only popular on its own, but is also connected to other popular destinations. This implies that the state either attracts tourists from influential destinations or sends tourists to them. Interestingly, Pahang has the highest eigenvector centrality. This indicates that Pahang has strong connectivity with other key tourist locations, such as Selangor and Johor, thereby enhancing its accessibility and attractiveness. Understanding eigenvector centrality can assist Pahang in planning tourism strategies, such as marketing campaigns, infrastructure development and partnerships with high-degree centrality states to boost tourism in other regions. Furthermore, Pahang has an advantage over other states due to its geographical location, as it is adjacent to six other states.

Table 4 and Figure 6 show that Melaka ranks seventh in terms of eigenvector centrality. This indicates that Melaka is important in the tourism flow because it is connected to other important states. Although Melaka is not one of the top states in terms of BC, ranking twelfth, it still plays

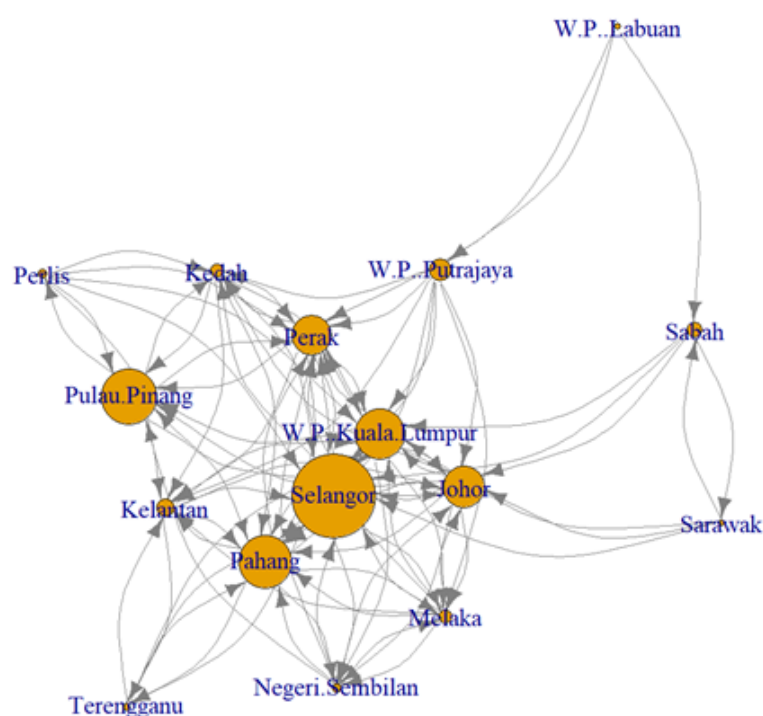


Figure 5: Malaysia's domestic tourism network with vertex size based on betweenness centrality

a crucial role in tourist flows into high-degree states.

It can also be observed that the eigenvector centrality scores of Sabah and Sarawak are close to 0, while the eigenvector centrality scores of W.P. Labuan and W.P. Putrajaya are equal to 0. These four states have outgoing edges to high-degree states, especially W.P. Putrajaya, which has eight outgoing edges. This may seem counterintuitive at first. However, if a state directs tourists to many high-degree states, this does not necessarily increase its own eigenvector centrality. Eigenvector centrality is more concerned with incoming connections from influential vertices. If a vertex sends tourists to high-degree vertices but does not receive tourists from vertices with high eigenvector centrality, it will have a low eigenvector centrality (Meghanathan 2004).

Table 5 summarises the rankings of all states across the four centrality measures and clarifies their structural roles in the domestic tourism network. Selangor stands out as the most dominant state. It ranks first in both in-degree and BC, and second in eigenvector centrality. This shows that Selangor is not only a popular destination but also a key bridge connecting other states.

Pahang also shows a strong position in the network. It ranks first in eigenvector centrality and third in both in-degree and BC. This suggests that Pahang is well connected to other influential states and plays an important bridging role. W.P. Kuala Lumpur and Perak rank highly in both in-degree and eigenvector centrality, confirming their importance as major tourism destinations.

Johor and Pulau Pinang have relatively high BC values. This indicates that they function as regional transit connectors in the southern and northern parts of Peninsular Malaysia, respectively. Kedah, Melaka and Negeri Sembilan show moderate rankings across most measures. These states are connected to major hubs, but they do not dominate the overall structure.

Kelantan and Terengganu rank lower in several measures. This reflects their more peripheral roles in the network. Sabah, Sarawak and W.P. Labuan record high out-directed closeness centrality but low rankings in other measures. This suggests that they mainly act as tourism sources rather than central hubs. Sarawak has a BC score of 0, which supports its role as an

Table 4: Eigenvector centrality

State	Eigenvector centrality
Pahang	1.000
Selangor	0.924
Perak	0.897
W.P. Kuala Lumpur	0.844
Pulau Pinang	0.686
Johor	0.680
Melaka	0.611
Negeri Sembilan	0.611
Kelantan	0.523
Kedah	0.495
Terengganu	0.369
Perlis	0.103
Sabah	5.438×10^{-17}
Sarawak	3.955×10^{-17}
W.P. Labuan	0
W.P. Putrajaya	0

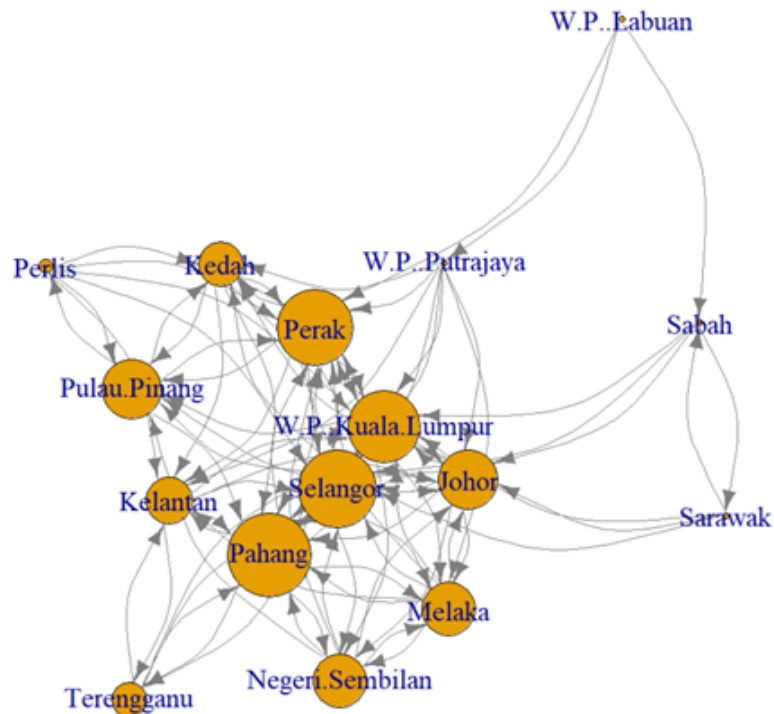


Figure 6: Malaysia's domestic tourism network with vertex size based on eigenvector centrality

Table 5: Comparative ranking of centrality measures and structural roles

State	Centrality rank				Structural role
	In-degree	Out-CC	BC	Eigenvector	
Selangor	1	5	1	2	Popular destination and main bridge
W.P. Kuala Lumpur	2	7	4	4	Popular destination
Pahang	3	8	3	1	Popular destination and main bridge
Perak	3	9	6	3	Popular destination
Johor	5	6	5	6	Southern transit connector
Pulau Pinang	6	14	2	5	Northern transit connector
Kelantan	7	14	8	9	Peripheral tourism state
Kedah	8	9	11	10	Connected destination linked to major states
Melaka	8	9	10	7	Connected destination linked to major states
Negeri Sembilan	10	9	12	7	Connected destination linked to major states
Terengganu	11	16	14	11	Peripheral tourism state
Sabah	12	2	9	13	Important source and occasional bridge
Perlis	13	9	12	12	Minimal bridge role
Sarawak	13	2	16	14	Important source and endpoint destination
W.P. Putrajaya	13	4	7	15	Administrative source-oriented state
W.P. Labuan	16	1	15	15	Important source

endpoint destination. W.P. Putrajaya shows moderate betweenness but low eigenvector centrality, indicating administrative importance with limited structural influence.

Overall, the results show that structural importance in the domestic tourism network is multidimensional. Some states act as main hubs and bridges, while others serve as connectors, sources or peripheral destinations. The comparative centrality results presented in Table 5 highlight the structural importance of each state in the domestic tourism network. However, structural importance alone does not explain whether highly central states are also economically stronger. To further understand this relationship, it is useful to examine how network centrality is associated with gross domestic product (GDP). By comparing centrality measures with GDP, the study can assess whether states that occupy more influential positions in the tourism network also demonstrate stronger economic performance.

4.1.3. Centrality and its relationship with gross domestic product

In the analysis of the domestic tourism network, it is useful to investigate the relationship between network centrality and gross domestic product (GDP). GDP is a measure of total economic output. It represents the monetary value of goods and services produced within an economy over a specific period, usually annually (Department of Statistics Malaysia 2023b). GDP can be expressed using the expenditure approach, as shown in Eq. 15:

$$GDP = C + G + I + (X - M), \quad (15)$$

where C represents private consumption, G represents government expenditure, I represents investment, X represents exports and M represents imports. The 2023 GDP data for each state are obtained from the Department of Statistics Malaysia and used to investigate the Pearson correlation coefficients between GDP and four centrality measures, namely in-degree, eigenvector, betweenness and out-directed closeness centralities (Department of Statistics Malaysia 2023a).

The relationship between in-degree centrality and GDP among Malaysian states provides insight into the association between domestic tourism attractiveness and economic output. In Figure 7, the Pearson correlation coefficient between in-degree centrality and GDP is 0.597, with a p-value of 0.0146. This indicates a statistically significant positive relationship between GDP and in-degree centrality, although the strength of the relationship is moderate. This suggests that states with higher GDP tend to have higher in-degree centrality, meaning that economically stronger states tend to attract tourists from more states.

Selangor has the highest in-degree centrality, with a value of 13. This implies that Selangor is a major destination for tourists from other states. Selangor also records the highest GDP, approximately RM406.1 billion, and correspondingly shows high attractiveness in the network. This association may be linked to its developed infrastructure, diverse tourism offerings and economic advantages. W.P. Kuala Lumpur follows Selangor closely. W.P. Kuala Lumpur has 12 in-degree links and the second highest GDP, approximately RM249.3 billion. As the capital city, W.P. Kuala Lumpur's high in-degree centrality is expected. Its significant GDP supports a variety of tourist activities and amenities, making it a central hub for domestic tourism as well as a transportation hub.

Sarawak presents an interesting case within Malaysia's domestic tourism network. Its in-degree is very low, with $k^{\text{in}}(\text{Sarawak}) = 1$, but it ranks fourth in terms of GDP, with a GDP of approximately RM142.4 billion. Sarawak, located on the island of Borneo, is one of the largest states in Malaysia and has substantial natural resources. Its high GDP can be attributed to several key economic sectors.

Sarawak is a major producer of oil and natural gas, which contributes significantly to its GDP (Department of Statistics Malaysia 2023a). The state is also rich in timber and palm oil plantations, which are important export sectors (Lau 2019). Despite its economic strength, Sarawak's low in-degree centrality indicates that it attracts fewer domestic tourism connections compared with other states. This may be partly due to its geographical separation from Peninsular Malaysia, which makes it less accessible for short or spontaneous domestic trips. Sarawak's attractions are primarily nature-based, such as national parks, caves and cultural heritage sites. Although these attractions are significant, they may appeal more to niche tourism markets, such as ecotourism, rather than mass domestic tourism (Mariapan *et al.* 2020).

In Figure 8, the Pearson correlation coefficient of 0.422 suggests a moderate positive association between eigenvector centrality and GDP. However, the relationship is not statistically significant at the 0.05 level, with a p-value of 0.104. This suggests that states with higher eigenvector centrality in the tourism network tend to have higher GDP, although the evidence is not strong enough to conclude a statistically significant relationship. This implies that being connected to important tourism hubs, such as Selangor and W.P. Kuala Lumpur, may contribute to or reflect economic strength. States with high eigenvector centrality may benefit from greater tourism activity, which can support local economies and contribute to GDP. High eigenvector centrality may also reflect more developed infrastructure and services, which support both tourism and broader economic activities.

The scatter plot in Figure 9 shows the relationship between BC and GDP for the Malaysian states. The Pearson correlation coefficient of 0.784, with a p-value of 0.000321, indicates a strong positive correlation between GDP and BC. This association is statistically significant at the 1% level, with $p < 0.01$, and remains significant at the 0.1% level, with $p < 0.001$. This suggests that states with higher GDP tend to exhibit higher BC in the domestic tourism network. BC measures the extent to which a state acts as an intermediary along the shortest paths between other states in the tourism flow. A higher BC suggests that a state serves as an important transit point in the tourism network. However, although the correlation is strong and

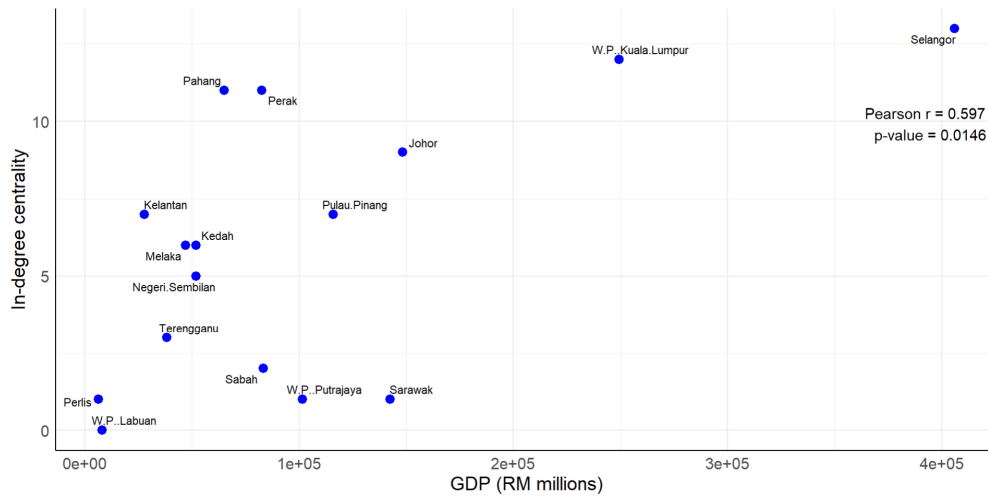


Figure 7: Scatter plot of in-degree centrality against GDP

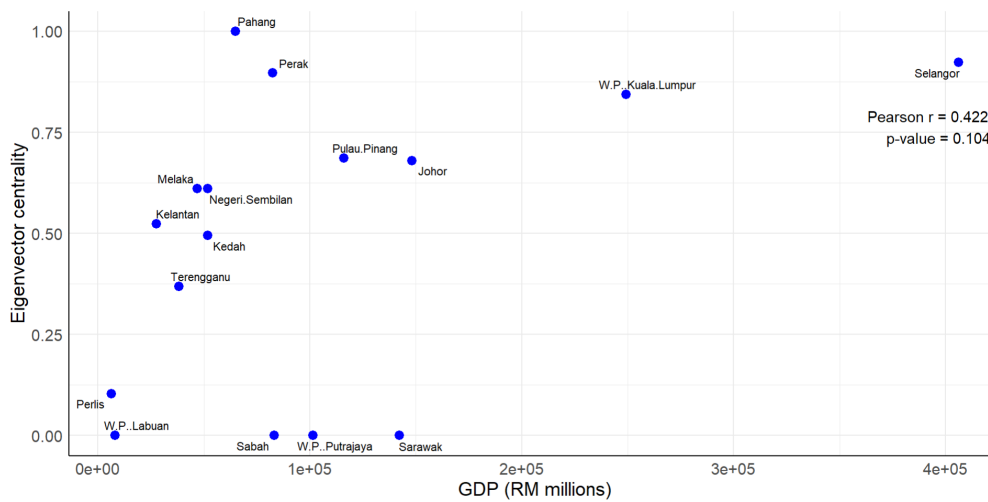


Figure 8: Scatter plot of eigenvector centrality against GDP

statistically significant, it does not imply a causal relationship.

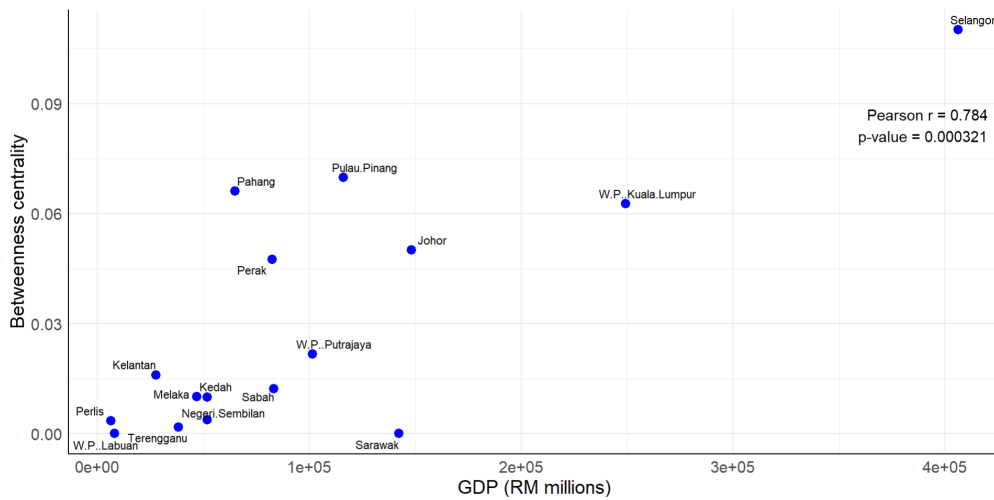


Figure 9: Scatter plot of betweenness centrality against GDP

W.P. Kuala Lumpur, the capital city, has both high BC and high GDP. This reflects its role as a central hub in Malaysia's tourism and economy. As a major urban centre, it not only attracts tourists directly but also serves as an important transit point for tourists travelling to other regions. The wider Klang Valley's transport infrastructure, including major rail connections and access to Kuala Lumpur International Airport in Selangor, further contributes to this centrality. Meanwhile, Johor, located at the southern tip of Peninsular Malaysia, shows a notable combination of moderate to high BC and GDP. It serves as a gateway between Malaysia and Singapore, which may significantly influence its centrality in tourism flows (Malaysian Investment Development Authority 2024). The state also benefits economically from cross-border tourism and trade, with attractions such as Legoland and Johor Bahru city centre drawing visitors.

Low BC indicates that a state is less frequently located on the shortest paths between other states. This suggests fewer opportunities for the state to benefit from transit-related economic activities, such as hospitality, retail and services that cater to travellers. For example, Kelantan and Perlis both have low BC and relatively low GDP. These states are less connected to the main domestic tourism routes compared with the more densely populated and economically developed west coast. This geographical positioning may limit their inclusion in major transit routes within the domestic tourism network. In addition, the east coast of Malaysia, including Kelantan and Terengganu, experiences a significant monsoon season from November to March, characterised by heavy rainfall and rough seas. This seasonality can affect tourist inflows because it limits beach activities and may disrupt travel and infrastructure (Ismail *et al.* 2020).

The Pearson correlation coefficient between out-directed closeness centrality and GDP is -0.196, indicating a weak negative correlation, as shown in Figure 10. The associated p-value is 0.466, suggesting that the relationship is not statistically significant at the 5% significance level. Although the coefficient is negative, the lack of statistical significance means that no reliable association can be inferred between GDP and out-directed closeness centrality.

Out-directed closeness centrality measures how quickly tourists can spread from a given state to all other states in the network. A higher value indicates that a state is more central in terms of sending tourists to other states. W.P. Labuan has the second smallest GDP but the highest out-directed closeness centrality. This suggests that despite its lower economic output, it plays an important role in distributing tourists to other states. This may be due to its strategic location and connections through ferry and flight services to other parts of Malaysia. In addition, Perlis has the smallest GDP and a relatively higher out-directed closeness centrality compared with several other states, indicating that Perlis is more significant as a source of

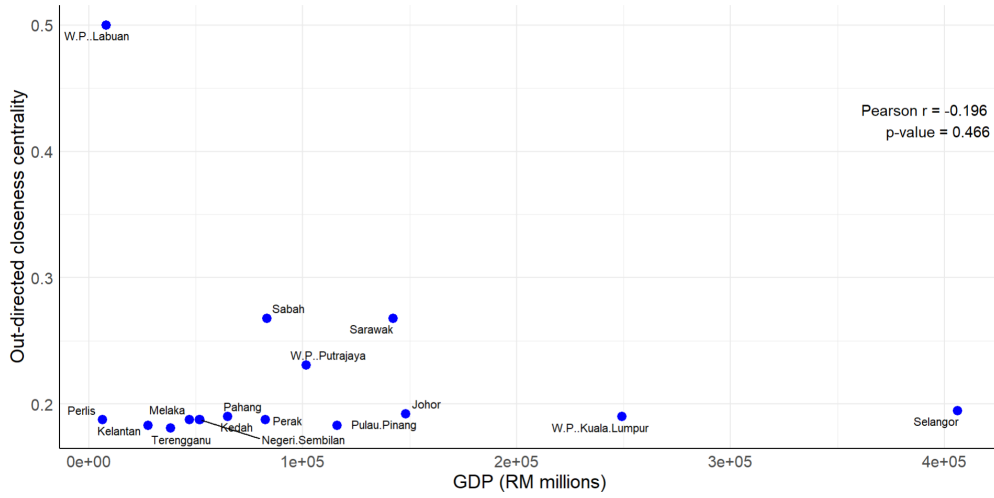


Figure 10: Scatter plot of out-directed closeness centrality against GDP

tourist flows than as a major destination. As the smallest state in Malaysia, Perlis may have fewer attractions to retain tourists for longer periods, encouraging tourist movement to nearby states.

The relationship between GDP and out-directed closeness centrality is weak and not statistically significant. The weak negative correlation suggests that states with higher GDP are not necessarily more central in sending tourists to other states. This indicates that economic strength may be more closely associated with a state's ability to attract tourists rather than its role in distributing tourists to other destinations. Taken together, these findings suggest that wealthier states in Malaysia function more as tourism destinations than as tourism sources within the domestic tourism network.

4.2. Community Detection

4.2.1. Undirected graph

The original tourism network is a directed network. For the purpose of community detection using the clustering coefficient, Louvain algorithm and spectral clustering method, the directed network is transformed into an undirected network. Malaysia's directed domestic tourism network is transformed into an undirected network by considering an edge between two states if at least one directed edge exists between them, as shown in Figure 11.

One cannot determine the direction of tourist flow by looking at the graph plot in Figure 11. The existence of an edge between two states simply means that there is a significant tourism connection between these two states. For instance, Melaka has strong tourism ties with Johor and Negeri Sembilan, without indicating the direction of tourist flow, but it has no strong tourism connections with Sabah and Sarawak.

After the transformation from a directed network G to an undirected network G' , the set of vertices remains the same. Thus, $G' = (V', E')$, where $V' = V$ consists of the 16 states and federal territories. However, the number of edges becomes $|E'| = 68$. The undirected edge set can be written as

$$E' = \{\{v_i, v_j\} \mid (v_i, v_j) \in E \text{ or } (v_j, v_i) \in E, i \neq j\}. \quad (16)$$

The graph G' can also be represented by an adjacency matrix A' , with entries a'_{ij} defined as

follows:

$$a'_{ij} = \begin{cases} 1, & \text{if there is a significant tourism connection between states } i \text{ and } j, \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

Unlike the adjacency matrix A in Eq. 7, the matrix A' is symmetric. The density of Malaysia's undirected domestic tourism network is 0.567. This means that more than half of all possible connections between states are realised, although the network is not fully saturated.

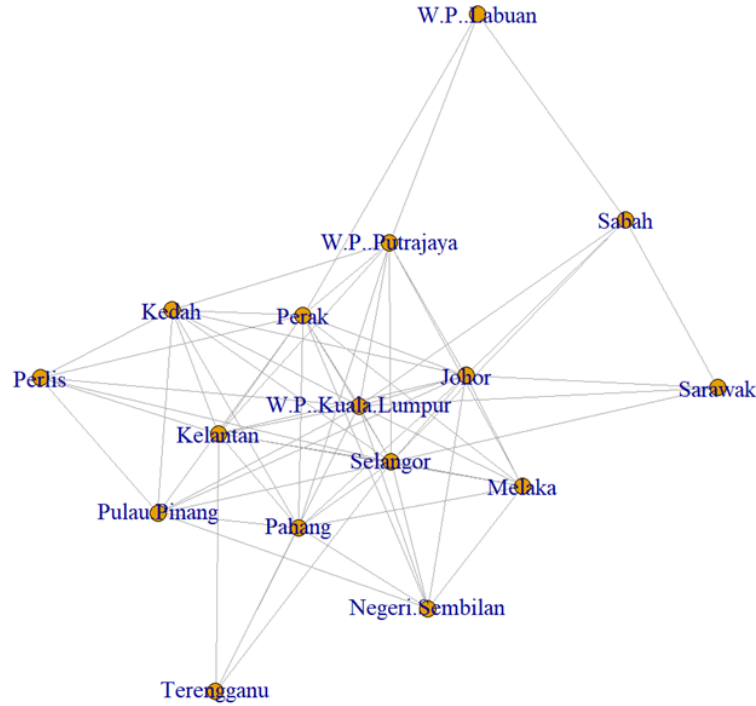


Figure 11: Malaysia's domestic tourism network as an undirected graph

4.2.2. Clustering coefficient

The clustering coefficient, specifically the global clustering coefficient, measures the likelihood that two vertices in a network with a common neighbour are also directly connected, thereby forming a triangle. In the context of Malaysia's domestic tourism network, the clustering coefficient represents the probability that two states connected to a common state are also directly connected to each other. A high clustering coefficient suggests that if two states are connected to the same tourism state, there is also a high tendency for a direct tourism connection to exist between them. This may indicate a strong regional tourism structure, where states not only attract visitors individually but also facilitate interstate tourism within closely connected groups.

The global clustering coefficient for Malaysia's domestic tourism network is 0.706. This indicates a high likelihood that two states connected to a common neighbouring state are also connected to each other. The density of Malaysia's undirected tourism network is 0.567. Since the global clustering coefficient is higher than the network density, Malaysia's domestic tourism network exhibits a high clustering property. This clustering can be advantageous for promoting regional tourism, as it facilitates travel within groups of closely connected states.

The local clustering coefficient (LCC) of a vertex can be defined as the probability that two neighbours of that vertex are also neighbours of each other. Mathematically, the LCC of vertex

i is calculated as

$$\text{LCC}(i) = \frac{2e_i}{k_i(k_i - 1)}, \quad (18)$$

where e_i is the number of edges between the neighbours of vertex i , and k_i is the degree of vertex i (Abu Bakar *et al.* 2014). If $k_i < 2$, then $\text{LCC}(i) = 0$ because no triangle can be formed.

The LCC values in Table 6 provide a local view of how individual states contribute to the overall clustering structure of the network. Terengganu and Sarawak have the maximum LCC values, equal to 1. Based on the LCC definition in Eq. 18, this indicates that Terengganu and Sarawak each form a complete subgraph, or clique, with their respective neighbours. A maximum LCC value of 1 means that all neighbours of the vertex are also connected to one another. In this network, two cliques of size five are identified, as implied by the LCC results and illustrated in Figure 12a and Figure 12b.

In addition, Negeri Sembilan, Perlis and Melaka are also among the top five states in terms of LCC values. These states contribute substantially to the global clustering coefficient because they increase the network's tendency to form complete triads. States with high LCC values have a dense pattern of tourism connections among their neighbouring states. This means that tourists frequently travel between these neighbouring states, creating tightly connected clusters. Interestingly, Terengganu and Sarawak, which have the maximum LCC values, each form clusters with three out of four of their respective neighbours based on the Louvain algorithm, as shown in Figure 13.

On the other hand, W.P. Labuan has the minimum LCC value of 0.333, indicating weaker connectivity among its neighbours and a smaller contribution to the global clustering coefficient. However, its presence in the network also highlights the diversity of connectivity patterns. This may indicate specialised connectivity that does not necessarily conform to the broader clustering pattern but serves a specific function, such as an administrative or source-oriented connection.

Table 6: Local clustering coefficient

State	Local clustering coefficient
Sarawak	1.000
Terengganu	1.000
Negeri Sembilan	0.952
Perlis	0.933
Melaka	0.929
Kedah	0.861
Pulau Pinang	0.857
W.P. Putrajaya	0.778
Kelantan	0.756
Pahang	0.745
Perak	0.667
Johor	0.652
Sabah	0.600
Selangor	0.560
W.P. Kuala Lumpur	0.560
W.P. Labuan	0.333

4.2.3. Louvain algorithm

The Louvain algorithm is a method for detecting communities in networks. It was developed by Blondel *et al.* (2008) and aims to maximise modularity Q , as shown in Eq. 19. Modularity Q

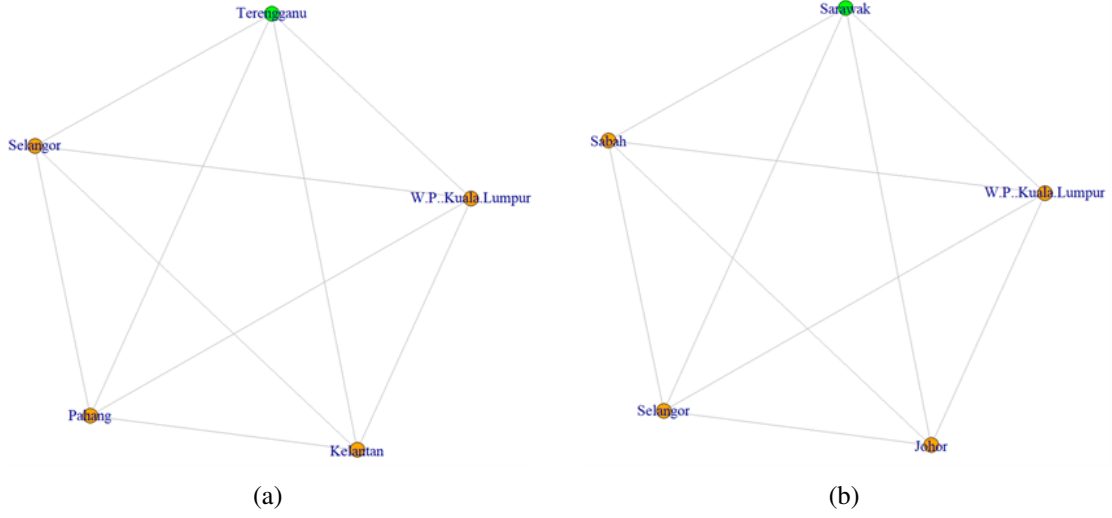


Figure 12: Examples of cliques in Malaysia's domestic tourism network: **(a)** clique formation implied by $LCC(\text{Terengganu}) = 1$ and **(b)** clique formation implied by $LCC(\text{Sarawak}) = 1$

measures the density of edges within communities compared with the density of edges between communities. In network analysis, Q is used to quantify the strength of the division of a network into clusters or modules.

$$Q = \frac{1}{2L} \sum_{i=1}^N \sum_{j=1}^N \left(A'_{ij} - \frac{k_i k_j}{2L} \right) \delta(x_i, x_j), \quad (19)$$

where A' is the adjacency matrix defined in Eq. 17, k_i is the degree of vertex i , k_j is the degree of vertex j , L is the total number of edges in the graph, x_i is the community label of vertex i , and $\delta(x_i, x_j)$ is the Kronecker delta function defined in Eq. 20 (Blondel *et al.* 2008). The function $\delta(x_i, x_j)$ acts as a filter that selects only pairs of vertices i and j that belong to the same community.

Intuitively, Eq. 19 compares the observed number of edges within a community with the expected number of edges between the same vertices under a randomly rewired configuration model that preserves the degree sequence. A configuration model is a random graph with a predefined degree sequence (Fosdick *et al.* 2018). The larger the value of Q , the stronger the community structure in the graph.

$$\delta(x, y) = \begin{cases} 1, & \text{if } x = y, \\ 0, & \text{otherwise.} \end{cases} \quad (20)$$

Based on the Louvain algorithm, Malaysia's domestic tourism network can be segmented into four modules or clusters. The states within the same cluster tend to share stronger tourism connections and, in some cases, geographical proximity. Figure 13 shows the resulting community structure.

The orange vertices represent the northern region cluster, which consists of Perlis, Kedah, Pulau Pinang and Perak. These states are located in the northern part of Peninsular Malaysia, making them geographically close and relatively accessible for interstate travel. This region includes popular tourist destinations such as Langkawi Island in Kedah, the historical city of George Town in Pulau Pinang and natural attractions in Perak.

The blue vertices represent the central and east coast cluster of Peninsular Malaysia. This

cluster consists of Kelantan, Terengganu, Pahang, Selangor, Negeri Sembilan and Melaka. The inclusion of these states suggests that domestic tourism flows in this cluster are shaped by both geographical continuity and strong interstate connectivity.

Meanwhile, the green vertices represent another cluster comprising Sabah, Sarawak, W.P. Kuala Lumpur and Johor. Although Sabah and Sarawak are geographically separated from Peninsular Malaysia, their grouping with W.P. Kuala Lumpur and Johor suggests that the Louvain algorithm captures tourism flow patterns rather than geographical proximity alone. This may reflect the role of W.P. Kuala Lumpur and Johor as important connection points in the broader domestic tourism network.

Finally, two out of the three federal territories, W.P. Labuan and W.P. Putrajaya, are assigned to the same cluster, represented by the yellow vertices. This indicates a distinct connectivity pattern for these federal territories within the domestic tourism network.

Based on the Louvain algorithm, Malaysia's domestic tourism network can be segmented into four modules or clusters. The states within the same cluster tend to share stronger tourism connections and, in some cases, geographical proximity. Figure 13 shows the resulting community structure.

The orange vertices represent the northern region cluster, which consists of Perlis, Kedah, Pulau Pinang and Perak. These states are located in the northern part of Peninsular Malaysia, making them geographically close and relatively accessible for interstate travel. This region includes popular tourist destinations such as Langkawi Island in Kedah, the historical city of George Town in Pulau Pinang and natural attractions in Perak.

The blue vertices represent the central and east coast cluster of Peninsular Malaysia. This cluster consists of Kelantan, Terengganu, Pahang, Selangor, Negeri Sembilan and Melaka. The inclusion of these states suggests that domestic tourism flows in this cluster are shaped by both geographical continuity and strong interstate connectivity.

Meanwhile, the green vertices represent another cluster comprising Sabah, Sarawak, W.P. Kuala Lumpur and Johor. Although Sabah and Sarawak are geographically separated from Peninsular Malaysia, their grouping with W.P. Kuala Lumpur and Johor suggests that the Louvain algorithm captures tourism flow patterns rather than geographical proximity alone. This may reflect the role of W.P. Kuala Lumpur and Johor as important connection points in the broader domestic tourism network.

Finally, two out of the three federal territories, W.P. Labuan and W.P. Putrajaya, are assigned to the same cluster, represented by the yellow vertices. This indicates a distinct connectivity pattern for these federal territories within the domestic tourism network.

The modularity score obtained from the Louvain optimisation is

$$Q = 0.0697448.$$

This value is relatively low compared with the commonly cited threshold of $Q \geq 0.3$ for strongly modular networks (Clauset *et al.* 2004). However, this result should not be interpreted as a methodological weakness. Instead, it reflects the structural characteristics of Malaysia's domestic tourism network.

A low modularity value indicates that the density of edges within the detected communities is not substantially higher than what would be expected under a random configuration model with the same degree distribution. In practical terms, this suggests that tourism flows are highly interconnected across states rather than confined within tightly separated regional blocs. Tourists frequently travel across regions, for example, between northern states and the central corridor, or between Peninsular Malaysia and Borneo. Such cross-regional flows increase the number of inter-community edges, thereby reducing modularity. Hence, the relatively small value of Q is consistent with a nationally integrated tourism network rather than a set of strongly segmented regional subsystems.

The Louvain algorithm still identifies four communities, but these groups should be interpreted as loose regional tendencies rather than sharply bounded modules. Nevertheless, the

northern region cluster, indicated by the orange vertices, shows a relatively strong intra-cluster connection due to geographical proximity and shared tourist attractions. Similarly, the central and east coast cluster, indicated by the blue vertices, and the Borneo-related cluster, indicated by the green vertices, also exhibit some degree of cohesion within each cluster. The inclusion of W.P. Labuan and W.P. Putrajaya in the same cluster may highlight the algorithm's ability to group vertices based on tourism flow patterns and administrative similarity, rather than geographical proximity alone.

Therefore, the modularity score provides insight into the nature of community structure in Malaysia's domestic tourism network. While the score indicates a limited level of modular separation, it also suggests that domestic tourism flows are broadly interconnected across regions. This provides a useful basis for further understanding the dynamics of domestic tourism flow between different regions.

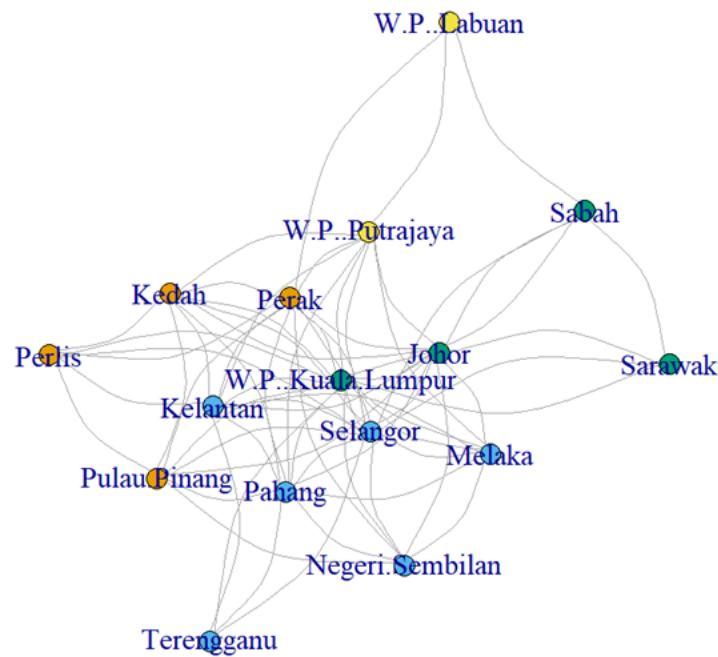


Figure 13: Four communities in Malaysia's domestic tourism network detected using the Louvain algorithm. Orange vertices represent the northern region cluster. Blue vertices represent the central and east coast cluster together with Melaka and Negeri Sembilan. Green vertices represent the Borneo cluster together with W.P. Kuala Lumpur and Johor. Yellow vertices represent the federal territories cluster

4.2.4. Spectral clustering

To further investigate cluster formation in Malaysia's domestic tourism network, the spectral clustering method is used. This approach has been widely studied due to its effectiveness and computational efficiency. Important graph properties, including community structure, can be revealed through the eigenvalues and eigenvectors of matrices naturally associated with graphs (Spielman 2012).

For a simple undirected graph, the Laplacian matrix L is defined by

$$L = K - A', \quad (21)$$

where K is the degree matrix and A' is the adjacency matrix defined in Eq. 17. The degree

matrix K is a diagonal matrix with entries

$$K_{ij} = \begin{cases} k_i, & \text{if } i = j, \\ 0, & \text{if } i \neq j, \end{cases} \quad (22)$$

where k_i is the degree of vertex i . Since Malaysia's domestic tourism network has been transformed into an undirected graph G' , the adjacency matrix A' of G' is used to construct L . Thus, the entries of L can be expressed as

$$l_{ij} = \begin{cases} k_i, & \text{if } i = j, \\ -1, & \text{if } i \sim j, \\ 0, & \text{otherwise.} \end{cases} \quad (23)$$

Since G' is a connected graph with 16 vertices, the Laplacian matrix L has 16 real eigenvalues λ_i satisfying

$$0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \dots \leq \lambda_{16}.$$

The second smallest eigenvalue of L , denoted by λ_2 , is known as the Fiedler value and provides a measure of how well connected the graph is (Fiedler 1973). The Fiedler vector is the eigenvector corresponding to the Fiedler value. In this study, bipartitioning is performed based on the signs of the entries of the Fiedler vector, where states with positive entries are grouped into one cluster and states with negative entries are grouped into another cluster. This process is applied recursively, following the approach discussed by Lucińska and Wierzchoń (2018). At each iteration, the bipartition is accepted only if each resulting cluster contains at least three states.

The spectral clustering results in Figure 14 are generally consistent with the Louvain algorithm results in Figure 13, with some differences. The northern states, namely Perlis, Kedah and Pulau Pinang, remain in one cluster. Similarly, the Borneo states, consisting of Sarawak and Sabah, are grouped with W.P. Labuan. The central and east coast states also remain largely within one module.

Furthermore, the Fiedler spectral clustering method is able to detect a community involving the federal territories. This is evident because W.P. Labuan and W.P. Putrajaya are grouped within the same broader cluster, together with the Malaysian Borneo states. This grouping may initially seem surprising due to geographical separation. However, it may reflect administrative and connectivity patterns that influence domestic tourism flows. W.P. Labuan, although geographically located in Borneo, operates under federal jurisdiction, similar to W.P. Putrajaya, the administrative centre of Malaysia. The inclusion of these territories in the same cluster may therefore indicate shared administrative or functional characteristics in the tourism network.

It is also useful to highlight that the spectral clustering method produces slightly different communities from the Louvain algorithm. Using the Fiedler spectral clustering method, the southern states, such as Johor, Melaka and Negeri Sembilan, are assigned to the same community, whereas this grouping is not fully observed in the Louvain algorithm. This implies that the Fiedler spectral clustering method is able to detect more localised spatial connections in the tourism flow network.

5. Sensitivity Analysis

This section evaluates whether the main findings depend strongly on the choice of the 90% threshold. The parameter α is varied from 0.75 to 1.00, and the structural and centrality outcomes are compared with the baseline value of $\alpha = 0.90$. For each $\alpha \in \{0.75, 0.80, 0.85, 0.925, 0.95, 0.975, 1.00\}$, the network is reconstructed and its structural and centrality properties are

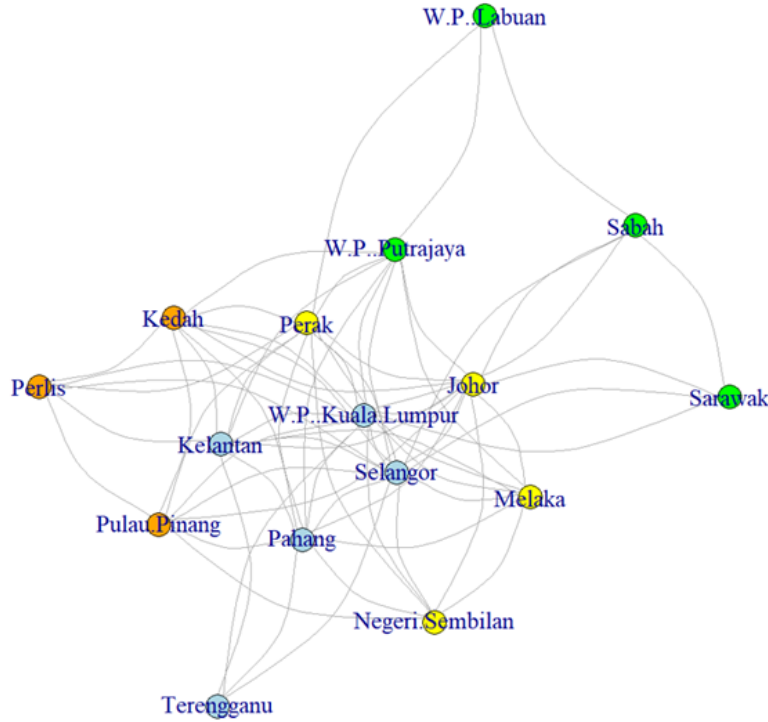


Figure 14: Four communities in Malaysia's domestic tourism network based on the Fiedler spectral method. Orange vertices represent the northern region cluster. Blue vertices represent the central and east coast cluster. Green vertices represent the Borneo cluster together with the federal territories. Yellow vertices represent the southern region cluster together with Perak

re-evaluated. This allows the study to assess whether small changes in the significance cut-off lead to substantial changes in network topology, community structure or vertex rankings. If the results remain stable across nearby thresholds, the baseline choice can be considered robust.

5.1. Edge Density Stability

Figure 15 shows the relationship between α and edge density. Edge density decreases smoothly and monotonically as α increases. There is no sharp drop or structural break at $\alpha = 0.90$. This suggests that the network remains locally stable around the baseline threshold.

Relative to the baseline, density increases by at most 11.76% at $\alpha = 0.75$ and decreases by 10.29% at $\alpha = 1.00$. The total variation remains within approximately $\pm 12\%$. Importantly, there is no discontinuity between $\alpha = 0.85$, $\alpha = 0.90$ and $\alpha = 0.925$. This indicates gradual adjustment rather than abrupt topological fragmentation. Thus, the 90% threshold lies within a stable density regime.

5.2. Centrality Ranking Stability

For each centrality ranking in Table 5, the Spearman rank correlation between the $\alpha = 0.90$ baseline and the rankings under alternative α thresholds is computed. Figure 16 presents the Spearman rank correlations for each centrality measure relative to the baseline. Spearman correlation measures the consistency of rankings across thresholds (Jiang *et al.* 2024). At $\alpha = 0.90$, all correlations are equal to 1 by construction.

When the threshold changes slightly to $\alpha = 0.85$ or $\alpha = 0.95$, the correlations remain very close to 1. This indicates strong local stability around the baseline. In-degree and eigenvector

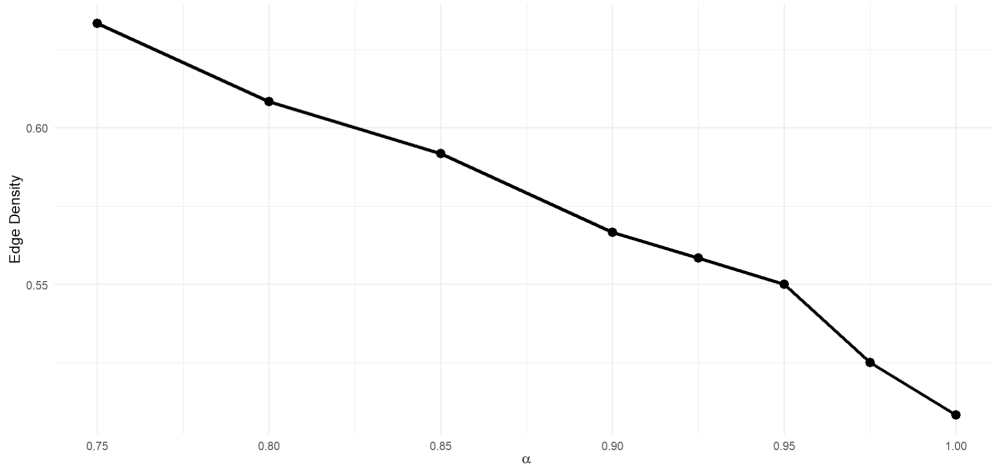


Figure 15: Plot of edge density against α

centralities show the highest robustness, with correlations remaining above 0.95 for most α values. This suggests that the identification of major hubs remains largely unchanged.

Out-directed closeness centrality is moderately sensitive at lower thresholds. At $\alpha = 0.75$ and $\alpha = 0.80$, the correlation decreases but remains above 0.88. This indicates that the overall ranking pattern is still preserved. BC is the most sensitive measure. Its correlation declines more noticeably at extreme thresholds, especially at $\alpha = 1.00$. This is expected because betweenness depends on shortest paths, which are more affected by changes in network sparsity.

To complement the correlation analysis, the mean absolute rank deviation (MARD) is computed. MARD measures the average absolute shift in ranking positions across thresholds (Wee *et al.* 2018). It is defined as

$$\text{MARD}(\alpha) = \frac{1}{N} \sum_{i=1}^N \left| r_i^{(\alpha)} - r_i^{(0.90)} \right|, \quad (24)$$

where N is the number of vertices, $r_i^{(\alpha)}$ is the rank of vertex i at threshold α , and $r_i^{(0.90)}$ is its baseline rank. Both ranks are obtained from Table 5. A MARD value of 0 indicates perfect ranking stability.

Table 7 shows that rank deviations remain small across most thresholds. Larger deviations appear mainly under extreme specifications, especially for betweenness and out-directed closeness at $\alpha = 1.00$. Overall, the ranking changes are limited in magnitude.

5.3. Community Detection Robustness

The robustness of community detection is examined using modularity and the number of detected clusters. Figure 17 shows the modularity trajectories for the Louvain and spectral methods. For the Louvain method, modularity at $\alpha = 0.90$ is approximately 0.07. Although it is slightly lower than the value at $\alpha = 0.85$, it increases again at $\alpha = 0.925$ and continues rising under stricter thresholds. Deviations range from -6% to +43% relative to the baseline. The trajectory changes gradually, and there is no structural collapse at $\alpha = 0.90$.

Spectral modularity lies within a narrow absolute range of approximately 0.04–0.07. Percentage deviations appear large, reaching +92% at $\alpha = 0.85$ and +85% at $\alpha = 1.00$. However, this is due to the small baseline value at $\alpha = 0.90$. In absolute terms, the changes remain modest.

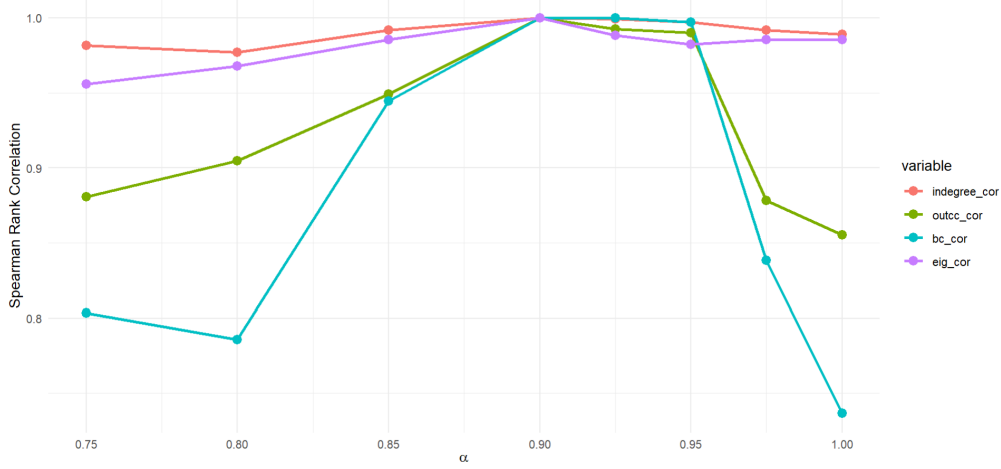


Figure 16: Plot of Spearman rank correlation for each centrality measure against α

Figure 18 shows the stability of the number of clusters. Spectral clustering consistently detects four communities across all eight thresholds. This indicates complete partition invariance under the tested threshold values.

Louvain clustering identifies three to four communities across most α values. At $\alpha = 0.90$, four clusters are detected. A larger contraction occurs only at $\alpha = 1.00$, where the number of clusters drops to two, with $\Delta k_L = -2$. Thus, major structural reconfiguration occurs only under the strictest cut-off.

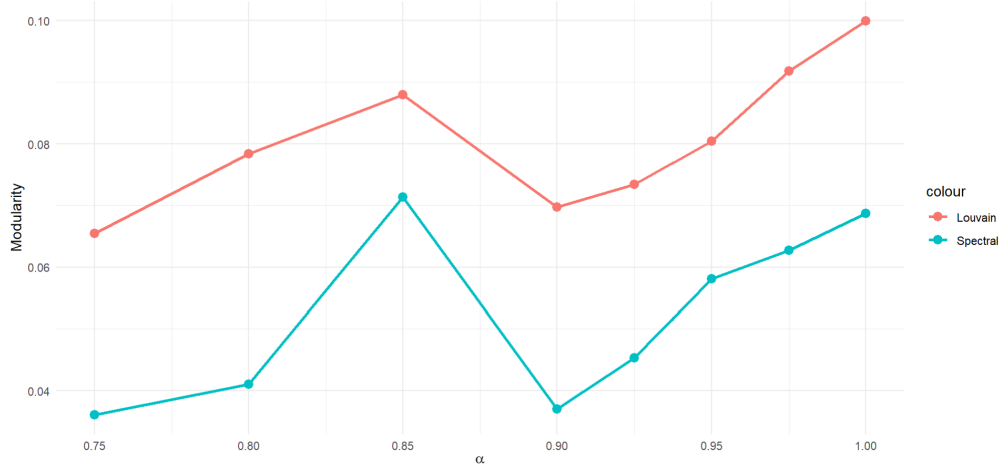


Figure 17: Plot of modularity score for Louvain and spectral methods against α

5.4. Summary of Quantitative Deviations

Table 7 summarises all structural and centrality deviations relative to the 90% baseline. Edge density varies within approximately $\pm 12\%$. Louvain modularity changes gradually without local instability, while spectral clustering preserves identical four-community partitions across all thresholds. Centrality rank deviations remain small, except under extreme sparsification. There is no abrupt change in density, centrality rankings, modularity or partition scale near the 90% threshold. Substantive structural shifts appear only under the extreme specification of

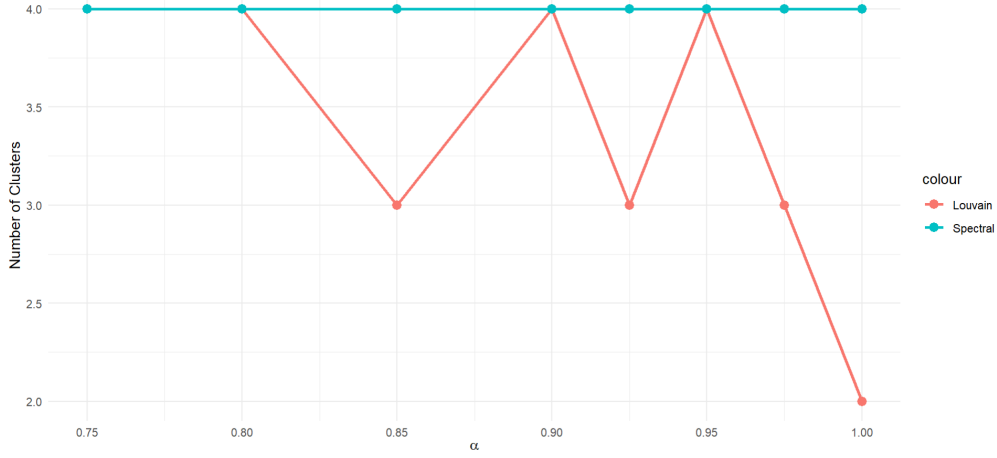


Figure 18: Plot of the number of clusters for Louvain and spectral methods against α

$\alpha = 1.00$. Overall, the principal structural patterns remain consistent across nearby threshold values.

Table 7: Structural and centrality deviations relative to the 90% threshold

α	Edges	Den. (%)	Modul. _L (%)	Modul. _S (%)	Δk_L	Δk_S	InDeg	OutCC	BC	Eig
0.75	76	11.76	-6.17	-2.39	0	0	0.625	1.500	2.500	1.000
0.80	73	7.35	12.33	10.87	0	0	0.750	1.500	2.438	1.000
0.85	71	4.41	26.14	92.84	-1	0	0.438	1.000	1.000	0.500
0.90	68	0.00	0.00	0.00	0	0	0.000	0.000	0.000	0.000
0.925	67	-1.47	5.24	22.58	-1	0	0.063	0.250	0.000	0.375
0.95	66	-2.94	15.37	57.06	0	0	0.188	0.375	0.125	0.625
0.975	63	-7.35	31.68	69.64	-1	0	0.375	1.688	1.625	0.500
1.00	61	-10.29	43.34	85.67	-2	0	0.500	1.813	2.250	0.375

Note: Deviations are computed relative to the 90% baseline network. Δk denotes the change in the number of detected communities. L and S denote the Louvain and spectral clustering methods, respectively.

6. Malaysia's Domestic Tourism Network: A Weighted Network Perspective

To extend the unweighted network framework in Section 3.2 to a weighted setting, the same vertex set V and edge set E are retained. The vertex set V represents the states and federal territories of Malaysia, while the edge set E represents significant directed tourism flows. The network remains directed, but each edge is now assigned a weight that reflects the relative strength of the corresponding tourism flow.

To account for large differences in tourist volumes across states, edge weights are normalised using the expected-flow benchmark (El Rhiauane *et al.* 2025). Under the uniform destination assumption, the expected flow η_{ij} represents the theoretical number of tourists travelling from state i to state j if destinations are chosen uniformly. Dividing the observed flow m_{ij} by η_{ij} therefore measures the relative strength of each edge in Malaysia's domestic tourism network compared with the expected baseline flow.

Formally, the weighted adjacency matrix $W = (w_{ij})$ is defined as

$$w_{ij} = \begin{cases} \frac{m_{ij}}{\eta_{ij}}, & \text{if } m_{ij} \geq \mu(i), \\ 0, & \text{otherwise,} \end{cases} \quad (25)$$

where m_{ij} denotes the observed number of tourists travelling from state i to state j , η_{ij} is the expected flow under the uniform destination assumption, and $\mu(i)$ is the threshold defined in Eq. 5.

Hence, the graph remains directed with the same edge structure as the unweighted network, but each edge now carries a weight. The value w_{ij} measures the strength of a tourism linkage relative to its probabilistic benchmark. A value of $w_{ij} > 1$ indicates that the observed flow exceeds the expected level, while a value close to 1 indicates that the observed flow is consistent with the baseline distribution.

This normalization serves several purposes. First, it aligns directly with the probabilistic model developed in this study, ensuring that the weighted network remains consistent with the theoretical framework. Second, it avoids the use of arbitrary threshold values in defining important tourism links. Edges are evaluated relative to their statistical expectation rather than by fixed cut-off values. Third, it reduces size bias across states. Larger states naturally generate more tourists, which may inflate raw flow counts. Scaling flows by their expected values controls for differences in total outgoing tourists and allows a more meaningful comparison across states.

6.1. Network Visualisation

Figure 19 illustrates the weighted and directed network of Malaysia's domestic tourism flows. The thickness of each edge represents the magnitude of the edge weight, which is defined as the ratio of the observed flow to the expected flow. Thicker edges correspond to larger weighted flows, indicating stronger tourism movements between the corresponding states.

The variation in edge thickness provides a visual indication of the intensity of tourism interactions. Edges with small weights appear as thin lines, representing relatively weak flows that are close to the expected level under the normalisation framework. In contrast, very thick edges highlight tourism routes where the observed flow substantially exceeds the expected baseline. These edges therefore represent dominant travel corridors in the domestic tourism system.

Several thick edges can be observed connecting major states in Peninsular Malaysia. In particular, strong connections appear around Selangor and W.P. Kuala Lumpur. Multiple thick edges both originate from and terminate at these vertices. This pattern indicates that these two regions are involved in large tourism exchanges with many other states. Their visual prominence in the network reflects the large weights associated with their incoming and outgoing flows.

Pahang also shows several moderately thick connections linking it to neighbouring states, such as Selangor, Perak and Kelantan. Although it is not as visually dominant as Selangor, the presence of these thicker edges suggests that Pahang functions as an important intermediary in the tourism network.

In the northern region, noticeable thick edges appear among Kedah, Pulau Pinang and Perlis. These edges reflect relatively strong regional tourism movements within the northern part of Peninsular Malaysia. The presence of these thicker links indicates that tourism exchanges within this region are stronger than many long-distance connections.

Sabah and Sarawak appear on the right side of Figure 19 and are connected to Peninsular Malaysia through several long edges. Some of these edges are relatively thick, particularly those linking them with W.P. Kuala Lumpur. This pattern suggests that W.P. Kuala Lumpur functions as a major gateway connecting tourism flows between Peninsular Malaysia and the Borneo states.

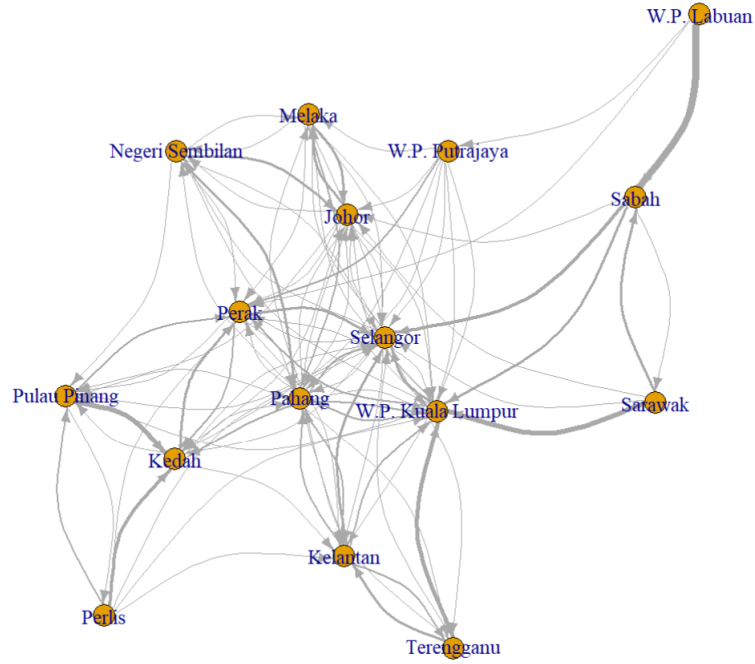


Figure 19: Malaysia's domestic tourism network as a weighted directed graph

Peripheral vertices such as W.P. Labuan and W.P. Putrajaya show mostly thin edges. However, W.P. Labuan is connected to Sabah by a relatively thick edge. The smaller thickness of most edges connected to these vertices indicates relatively lower tourism flow magnitudes compared with the major hubs.

Overall, the visualisation highlights how the distribution of edge thickness reveals the strength of tourism movements across the network. A small number of routes carry much larger flows than others, and these routes are concentrated around central hubs such as Selangor and W.P. Kuala Lumpur. The weighted representation therefore provides additional insight beyond the binary structure by showing not only which states are connected, but also how strong those connections are.

6.2. Centrality

The weighted and directed centrality results provide further insight into the structural roles of Malaysian states within the domestic tourism network. Since the weights are defined as the ratio of observed flow to expected flow, the measures reflect the strength of tourism linkages relative to the probabilistic benchmark.

The in-degree and out-degree definitions introduced in Section 4.1.1 are modified into in-strength and out-strength for the weighted directed network. In addition, the definitions of out-directed closeness centrality, betweenness centrality and eigenvector centrality are extended to incorporate edge weights.

In-strength and out-strength. In the weighted directed network, in-degree is replaced by in-strength. The in-strength of vertex i is calculated as

$$s^{\text{in}}(i) = \sum_j w_{ji}, \quad (26)$$

where w_{ji} represents the normalised tourism flow intensity from state j to state i (Sadhu *et al.* 2024). Similarly, out-degree is replaced by out-strength. The out-strength of vertex i is given by

$$s^{\text{out}}(i) = \sum_j w_{ij}, \quad (27)$$

where w_{ij} represents the normalised tourism flow intensity from state i to state j . Thus, in-strength measures the total weighted incoming tourism flow received by a state, whereas out-strength measures the total weighted outgoing tourism flow sent by a state.

Out-directed closeness centrality. In the weighted network, shortest paths incorporate edge costs, unlike the unweighted directed network in which the shortest path length is measured by the number of edges. Since larger weights represent stronger tourism flows, distances are defined to decrease as weights increase. Therefore, edge lengths are defined as the inverse of weights.

Recall the definition of a path in Eq. 9. For a directed path from state i to state j , consisting of consecutive vertices

$$i = v_0, v_1, v_2, \dots, v_k = j,$$

the weighted shortest path distance is defined as

$$d^{(w)}(i, j) = \min_{P_{ij}} \sum_{r=0}^{k-1} \frac{1}{w_{v_r, v_{r+1}}}, \quad (28)$$

where P_{ij} denotes the set of all directed paths from i to j , and $w_{v_r, v_{r+1}}$ is the normalised weight of the directed edge from vertex v_r to vertex v_{r+1} . Thus, the summation is taken over all consecutive vertex pairs (v_r, v_{r+1}) along a directed path, and the minimum is selected among all possible directed paths connecting i to j . If no directed path exists from i to j , the same convention as in the unweighted case is applied by assigning a penalty distance.

Following Das and Ram (2024), Eq.11 is extended to the weighted out-directed closeness centrality as follows:

$$CC_{\text{out}}^{(w)}(i) = \frac{1}{\sum_j d^{(w)}(i, j)}. \quad (29)$$

Betweenness centrality. In the weighted case, shortest paths are computed using the weighted distances $d^{(w)}(i, j)$ defined in Eq. 28. By incorporating the weight factor into Eq. 12, the weighted betweenness centrality is computed as

$$BC^{(w)}(i) = \sum_{s, t \in V} \frac{\sigma^{(w)}(s, t \mid i)}{\sigma^{(w)}(s, t)}, \quad (30)$$

where $\sigma^{(w)}(s, t)$ is the number of weighted shortest paths from vertex s to vertex t , and $\sigma^{(w)}(s, t \mid i)$ is the number of weighted shortest paths from vertex s to vertex t that pass through vertex i (Das & Ram 2024). In this formulation, a state has high weighted BC if it frequently lies along strong and efficient weighted tourism routes between other states.

Eigenvector centrality. The weighted eigenvector centrality is obtained by replacing the unweighted adjacency matrix A in Eq. 13 with the weighted adjacency matrix W . Thus, the weighted eigenvector centrality satisfies

$$W\mathbf{x}^{(w)} = \lambda_1^{(w)}\mathbf{x}^{(w)}, \quad (31)$$

where W is the weighted adjacency matrix defined in Eq. 25, $\lambda_1^{(w)}$ is the largest eigenvalue of W , and $\mathbf{x}^{(w)}$ is the corresponding leading eigenvector. The i^{th} entry of $\mathbf{x}^{(w)}$ gives the weighted eigenvector centrality of state i . A high value indicates that the state is strongly connected to other states that are also important in terms of weighted tourism flow.

$$\text{BC}^{(w)}(i) = \sum_{s,t \in V} \frac{\sigma^{(w)}(s, t \mid i)}{\sigma^{(w)}(s, t)}, \quad (32)$$

where $\sigma^{(w)}(s, t)$ is the number of weighted shortest paths from vertex s to vertex t , while $\sigma^{(w)}(s, t \mid i)$ is the number of weighted shortest paths from vertex s to vertex t that pass through vertex i (Brandes 2008). The ordered pair (s, t) represents every possible pair of vertices in the graph, excluding vertex i .

Eigenvector centrality. In the weighted network, the weighted adjacency matrix W replaces A in Eq. 13 (Ando *et al.* 2021).

$$W\mathbf{x}^{(w)} = \lambda_1^{(w)}\mathbf{x}^{(w)}. \quad (33)$$

Thus, influence depends not only on connections to central states, but also on the strength of those connections.

The comparative analysis of the centrality rankings in Table 5 and Table 8 reveals three clearly differentiated structural patterns across Malaysia's domestic tourism network. First, core hubs such as Selangor and Pahang exhibit strong ranking stability across both binary and weighted specifications. In the unweighted network, Selangor ranks first in in-degree, with $k^{\text{in}} = 13$, and first in BC, with a value of 0.110. Pahang ranks joint third in in-degree, with a value of 11, third in BC, with a value of 0.066, and first in eigenvector centrality.

In the weighted domestic tourism network, Selangor remains dominant, recording the second-highest in-strength, with a value of 29.633, the highest weighted BC, with a value of 0.176, and the maximum eigenvector centrality, with a value of 1.000. Pahang similarly maintains a top-tier position, with an in-strength value of 21.065, ranked third overall, a weighted BC value of 0.071, and an eigenvector centrality value of 0.847, ranked second overall. These consistently high rankings confirm that their prominence is not driven merely by the number of incoming links, but also by flows that substantially exceed the expected levels. They function simultaneously as major accumulation nodes, principal geodesic intermediaries and highly embedded vertices in the dominant eigenstructure of the network. Their importance is therefore both structurally central and magnitude-consistent.

Second, regional connectors such as Johor and Pulau Pinang display moderate sensitivity to weighting. In the unweighted network, Johor ranks fifth in in-degree and fifth in BC, while Pulau Pinang ranks sixth in in-degree and second in BC. After incorporating weights, Johor's weighted BC declines to 0.019, placing it outside the top tier, despite maintaining a solid in-strength value of 16.551. Pulau Pinang's weighted BC is 0.071, but it no longer holds the same relative prominence compared with Selangor, with a value of 0.176, and W.P. Kuala Lumpur, with a value of 0.119. This indicates that part of their brokerage role in the binary network was supported by structurally short paths that do not necessarily correspond to disproportionately

Table 8: Comparative ranking of centrality measures for the weighted domestic tourism network

State	Centrality rank			
	In-strength	Out-CC	BC	Eigenvector
Selangor	2	2	1	1
W.P. Kuala Lumpur	1	4	2	3
Pahang	3	10	3	2
Perak	4	3	5	4
Johor	6	6	9	5
Kedah	5	11	7	6
Kelantan	8	5	8	8
Melaka	9	15	10	7
Negeri Sembilan	11	1	10	10
Pulau Pinang	10	12	3	9
Terengganu	12	8	10	11
Sabah	7	13	6	13
Perlis	14	7	10	12
Sarawak	13	9	10	13
W.P. Putrajaya	15	14	10	13
W.P. Labuan	16	16	10	13

strong tourism flows. Their transit importance remains visible, but some of their shortest-path roles are reduced when flow intensity is explicitly modelled.

Third, peripheral and endpoint states such as Sarawak, W.P. Labuan and Perlis remain consistently marginal across both frameworks. In the unweighted network, these states occupy bottom-tier positions in in-degree and BC. In the weighted network, Sarawak records an in-strength value of 1.136, Perlis records 0.993, and W.P. Labuan records 0.000, while all three show zero or near-zero BC. Moreover, their eigenvector centralities are equal to 0 or close to 0, indicating the absence of strong incoming ties from influential states. This persistence confirms that their peripheral status is not an artefact of binary thresholding, but reflects structurally and quantitatively weak integration within the national tourism system.

Overall, the weighted domestic tourism network analysis refines, but does not overturn, the conclusions derived from the unweighted network. The dominant hubs identified in the binary structure, particularly Selangor and Pahang, remain central after accounting for flow magnitude. This demonstrates that their influence is supported by flows that significantly exceed the expected distribution. At the same time, weighting sharpens the hierarchy by distinguishing between structurally present but moderate connections and genuinely strong tourism corridors. Several states that exhibit moderate structural connectivity in the unweighted network show noticeably lower weighted influence. For instance, Negeri Sembilan has a relatively small weighted in-strength value of 6.370 despite being connected to multiple states structurally, suggesting that many of its links are close to the expected baseline flow. Similarly, Perlis records a very small in-strength value of 0.993 and near-zero BC, indicating that its connections are both weak and rarely used as transit routes.

In contrast, strong tourism corridors become more visible under weighting. Edges connecting major hubs such as Selangor, W.P. Kuala Lumpur and Pahang appear as thicker edges in the weighted network and correspond to large weighted strengths, including 30.830 for W.P. Kuala Lumpur and 29.633 for Selangor. These values highlight flows that significantly exceed the expected uniform distribution. The resulting structure therefore represents a magnitude-adjusted topology that more accurately captures meaningful centrality within Malaysia's domestic tourism network.

7. Limitations and Future Work

7.1. *Limitations of State-Level Aggregation*

This study constructs Malaysia's domestic tourism network at the state level. While this level of aggregation is consistent with official tourism statistics, it introduces certain limitations that should be acknowledged.

First, Malaysian states differ substantially in population size. More populous states, such as Selangor and Johor, naturally generate larger volumes of outbound travel. As a result, out-strength centrality may partly reflect population scale rather than purely structural influence within the tourism system. Similarly, highly urbanised or economically developed states may attract greater inflows partly because of their market size and infrastructure capacity.

Although the construction of the weighted domestic tourism network adjusts each state's outgoing flows relative to its own distribution of destinations, it does not account for differences in state population size. Larger states naturally generate more travellers because they have more residents. As a result, higher out-strength or in-strength centrality may partly reflect demographic scale rather than purely network position.

In other words, the normalisation procedure ensures that flows are evaluated relative to what is expected within each state's distribution of destinations, but it does not convert the data into per capita terms. Therefore, the centrality measures in this study capture the relative intensity and structural importance of tourism flows between states, not tourism activity per resident. The rankings should thus be interpreted as indicators of flow-based influence within the network, rather than measures of tourism participation adjusted for population size.

Future research could address this limitation by incorporating population-adjusted flows, for example by normalising tourism volumes on a per capita basis or by constructing a doubly standardised matrix that accounts for both origin and destination population size. A finer spatial resolution, such as district-level or metropolitan-level analysis, may also reveal sub-state heterogeneity that is masked under state-level aggregation.

Accordingly, while the present results provide a robust representation of inter-state tourism structure, caution is warranted when interpreting centrality rankings as indicators of relative tourism performance independent of demographic scale.

7.2. *Future Research Directions*

This study examines Malaysia's domestic tourism network for the year 2023. Although the single-year analysis provides a clear picture of the current structure, the proposed framework can be extended to a multi-year setting.

The weighted domestic tourism network can be constructed separately for each year using the same procedure. This would generate a sequence of directed weighted networks over time. Comparing these yearly networks would allow changes in centrality rankings, hub stability and regional connectivity patterns to be tracked.

Such an extension would be particularly useful for understanding post-pandemic recovery. By examining pre-pandemic, pandemic and post-pandemic years, it would be possible to observe whether dominant states such as Selangor and Pahang maintain their central roles or whether new hubs emerge. Changes in in-strength and eigenvector centrality could indicate shifts in destination attractiveness, while variation in BC may reveal the rise or decline of key transit corridors.

A longitudinal analysis would also allow the assessment of structural resilience. Stability in centrality rankings across years would suggest a robust tourism system, whereas large fluctuations might reflect vulnerability to external shocks such as the COVID-19 pandemic.

In addition, future research could incorporate transportation infrastructure networks. Combining tourism flows with domestic flight routes, highway connectivity or rail networks would help clarify the relationship between physical accessibility and tourism intensity. This approach

may reveal whether highly central tourism states derive their position mainly from infrastructural advantage or from broader economic and demographic factors.

Overall, extending the analysis across multiple years and integrating complementary network layers would provide a deeper understanding of long-term structural change and regional development within Malaysia's domestic tourism.

8. Conclusions

This paper analyses Malaysia's domestic tourism network using graph theory. Data from the 2023 Domestic Tourism Survey by the Department of Statistics Malaysia are used, with states represented as vertices and significant tourist flows between states represented as directed edges. The study identifies key tourism hubs based on centrality measures and analyses community structures using the Louvain algorithm and Fiedler spectral clustering method.

The results show that BC has a strong positive and statistically significant correlation with GDP, while in-degree and eigenvector centralities show moderate positive correlations with GDP. In contrast, out-directed closeness centrality shows a weak negative correlation with GDP. The strong correlation between BC and GDP suggests that states acting as major connectors within Malaysia's domestic tourism network tend to have higher economic output. However, this relationship should be interpreted as an association rather than evidence of causality. Selangor, W.P. Kuala Lumpur, Pahang, Johor and Pulau Pinang have relatively high BC values and high GDP levels, indicating that central positions in tourism flows are associated with stronger economic performance.

Out-directed closeness centrality reveals that the Borneo-related states and territories, namely W.P. Labuan, Sabah and Sarawak, occupy high positions in the ranking. Although the out-degrees of W.P. Labuan, Sabah and Sarawak are relatively lower than those of several states in Peninsular Malaysia, their closeness values indicate that they are structurally close to the rest of Malaysia's domestic tourism network in terms of outgoing paths. This implies that W.P. Labuan, Sabah and Sarawak play important roles as sources of domestic tourism flows. This pattern also suggests that the relatively higher GDP and stronger infrastructure of several Peninsular Malaysian states may make them attractive destinations for tourists from W.P. Labuan, Sabah and Sarawak.

The findings also provide several practical implications. Improving the accessibility of states with low in-degree centrality and low GDP may help strengthen domestic tourism. Interstate transportation links should be enhanced, including more affordable and frequent flights from Peninsular Malaysia to W.P. Labuan, Sabah and Sarawak, to make these destinations more accessible to domestic tourists. In addition, tourism authorities could leverage each state's unique natural and cultural assets through targeted marketing campaigns to attract both niche and general tourists, especially to the Borneo states. Investment in tourism-supporting infrastructure, including hotels, restaurants and local transport services, may also help increase tourism activity and support regional economic development.

Future studies may extend the network analysis across multiple years to examine structural changes in domestic tourism flows. Such longitudinal analysis would allow researchers to observe long-term trends and measure the recovery pattern of domestic tourism after the COVID-19 pandemic. Tracking centrality changes over time may also help policymakers evaluate whether tourism strategies successfully strengthen connectivity between states.

Community detection based on the Louvain algorithm and Fiedler spectral clustering method reveals four clusters of states. The optimised modularity score of $Q = 0.06974481$ suggests that Malaysia's domestic tourism network has some level of community structure, although the modularity is relatively weak since $Q < 0.3$. Nevertheless, both clustering methods reveal patterns that are broadly aligned with the spatial locations of the states, including the northern region, southern region, east coast region, central region and Borneo. This suggests that significant tourist flows between states in Malaysia are still influenced by geographical proximity and

neighbouring-state relationships, even in 2023.

Overall, Malaysia's domestic tourism network shows a dynamic pattern of interactions between states. Based on the analysis, Selangor, W.P. Kuala Lumpur and Pahang are identified as major hubs of Malaysia's domestic tourism in 2023 due to their strong performance in in-degree centrality, BC and eigenvector centrality. Nevertheless, every state has its own strengths and there remains room for improvement to boost domestic tourism in each state. The performance of each state can therefore be assessed from a network centrality perspective, including degree, betweenness, closeness and eigenvector centralities.

Acknowledgement

This research was funded by grants from the Ministry of Higher Education of Malaysia under FRGS Grant FRGS/1/2020/STG06/UKM/02/6 and FRGS/1/2020/STG06/UKM/02/8.

References

- Abu Bakar S., Taheri J. & Zomaya A.Y. 2013. Characterisation of essential proteins in proteins interaction networks. *Journal of Quality Measurement and Analysis* **9**(2): 11–26.
- Abu Bakar S.A., Taheri J. & Zomaya A.Y. 2014. Characterization of essential proteins based on network topology in proteins interaction networks. *AIP Conference Proceedings*, pp. 36–42. American Institute of Physics.
- Ando H., Bell M., Kurauchi F., Wong K.I. & Cheung K.F. 2021. Connectivity evaluation of large road network by capacity-weighted eigenvector centrality analysis. *Transportmetrica A: Transport Science* **17**(4): 648–674.
- Blondel V.D., Guillaume J.L., Lambiotte R. & Lefebvre E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**(10): P10008.
- Brandes U. 2008. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks* **30**(2): 136–145.
- Cao Z., Xu H. & Xian B.T.S. 2022. Chinese tourists in Malaysia: An analysis of spatio-temporal behavior based on tourism digital footprints. *IEEE Access* **10**: 122047–122060.
- Chung H.C., Chung N. & Nam Y. 2017. A social network analysis of tourist movement patterns in blogs: Korean backpackers in Europe. *Sustainability* **9**(12): 2251.
- Clauset A., Newman M.E.J. & Moore C. 2004. Finding community structure in very large networks. *Physical Review E* **70**(6): 066111.
- Das J. & Ram S. 2024. Exploring the relationship of transport network and land use patterns: An approach through weighted centrality assessment. *Urban, Planning and Transport Research* **12**(1).
- Department of Statistics Malaysia 2023a. Gross domestic product (GDP) by state 2023. <https://www.dosm.gov.my/portal-main/release-content/gross-domestic-product-gdp-by-state-20232023>.
- Department of Statistics Malaysia 2023b. Technical notes on GDP. <https://open.dosm.gov.my/publications/technical-notes/gdp-technotes>.
- Department of Statistics Malaysia 2023c. Tourism Satellite Account 2023. <https://bernama.com/en/news.php?id=2339713>.
- Department of Statistics Malaysia 2024. Domestic tourism survey 2023. <https://newss.statistics.gov.my/newss-portalx/ep/epFreeDownloadContentSearch.seam?cid=122034>.
- El Rhiaouane A., Oukhouya H., Guerbaz R., Belkhoutout K., Lmakri A., Fihri M. & El Afia A. 2025. Integrating ESG criteria in portfolio optimization: A Moroccan case study using Markowitz's theory and correlation network analysis. *Physica A: Statistical Mechanics and its Applications* **667**: 130521.
- Fiedler M. 1973. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal* **23**(98).
- Fosdick B.K., Larremore D.B., Nishimura J. & Ugander J. 2018. Configuring random graph models with fixed degree sequences. *SIAM Review* **60**(2): 315–355.
- Freeman L.C. 1978. Centrality in social networks: Conceptual clarification. *Social Networks* **1**(3): 215–239.
- Goddard W. & Oellermann O.R. 2011. Distance in graphs. *Structural Analysis of Complex Networks* (ed. M. Dehmer), pp. 49–72. Boston: Birkhäuser Boston.

- González-Díaz B., Gómez M. & Molina A. 2015. Configuration of the hotel and non-hotel accommodations: An empirical approach using network analysis. *International Journal of Hospitality Management* **48**: 39–51.
- Guo Y., Zhang J., Yang Y. & Zhang H. 2015. Modeling the fluctuation patterns of monthly inbound tourist flows to China: A complex network approach. *Asia Pacific Journal of Tourism Research* **20**: 1–12.
- Ismail N.I., Ariffin E.H., Yaacob R., Husain M.L. & Baharim N.B. 2020. The impact of seasonal monsoons on the morphology of beaches protected by barrier islands in Setiu, Terengganu, Malaysia. *Journal of Sustainability Science and Management* **15**(4): 120–129.
- Jiang J., Zhang X. & Yuan Z. 2024. Feature selection for classification with Spearman's rank correlation coefficient-based self-information in divergence-based fuzzy rough sets. *Expert Systems with Applications* **249**: 123633.
- Kreinovich V., Nguyen H.T. & Niwitpong S.A. 2008. Statistical hypothesis testing under interval uncertainty: An overview. *International Journal of Intelligent Technology and Applied Statistics* **1**(1): 1–32.
- Lau R. 2019. Bolstering Sarawak's trade outlook. *The Borneo Post*.
- Lee S.H., Choi J.Y., Yoo S.H. & Oh Y.G. 2013. Evaluating spatial centrality for integrated tourism management in rural areas using GIS and network analysis. *Tourism Management* **34**: 14–24.
- Lucińska M. & Wierzchoń S.T. 2018. Clustering based on eigenvectors of the adjacency matrix. *International Journal of Applied Mathematics and Computer Science* **28**(4): 771–786.
- Malaysian Investment Development Authority 2024. A gateway to investment opportunities in Malaysia's new southern nucleus. <https://www.mida.gov.my/mida-news/a-gateway-to-investment-opportunities-in-malaysias-new-southern-nucleus/>.
- Mariapan M., Lim E., Bidin S. & Zawawi A. 2020. Ecotourism in Malaysia: Current scenario. *Universiti Putra Malaysia*.
- McKercher B. & Lau G. 2008. Movement patterns of tourists within a destination. *Tourism Geographies* **10**(3): 355–374.
- Meghanathan N. 2004. Centrality measures. <https://www.jsums.edu/nmeghanathan/files/2015/08/CSC641-Fall2015-Module-2-Centrality-Measures.pdf>. Accessed 11 August 2024.
- Miguéns J.I.L. & Mendes J.F.F. 2008. Travel and tourism: Into a complex network. *Physica A: Statistical Mechanics and its Applications* **387**(12): 2963–2971.
- Ministry of Tourism Malaysia 2021. Malaysia tourism annual report 2021. <https://www.tourism.gov.my/files/uploads/annual-report-2021.pdf>.
- Mosbah A., Saleh M. & Al Khuja A. 2014. A review of tourism development in Malaysia. *European Journal of Business and Management* **6**(5), www.iiste.org.
- Mura P. & Tavakoli R. 2014. Tourism and social capital in Malaysia. *Current Issues in Tourism* **17**(1): 28–45.
- Putman K.L., Boekhout H.D. & Takes F.W. 2020. Fast incremental computation of harmonic closeness centrality in directed weighted networks. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 1018–1025. New York, NY: Association for Computing Machinery.
- Ramlan Y. & Rodzi N.H. 2023. A test of time in KL and Selangor. *Selangor Journal*.
- Razak F.A. & Expert P. 2021. Modelling the spread of COVID-19 on Malaysian contact networks for practical reopening strategies in an institutional setting. *Sains Malaysiana* **50**(5): 1497–1509.
- Razak F.A. & Zamzuri Z.H. 2021. Modelling heterogeneity and super spreaders of the COVID-19 spread through Malaysian networks. *Symmetry* **13**(10): 1954.
- Sadhu S., Namtirtha A., Malta M.C. & Dutta A. 2024. Normalized strength-degree centrality: Identifying influential spreaders for weighted network. *Social Network Analysis and Mining* **14**(1): 232.
- Samuel H. & Nor Ahmad N.A. 2020. Divergent dimensions of radicalization risk: Migration and violent extremism in Sabah, Malaysia. *Final Report to The United States Department of State's Bureau of Conflict and Stabilization Operations*.
- Seok H., Barnett G.A. & Nam Y. 2021. A social network analysis of international tourism flow. *Quality and Quantity* **55**(2): 419–439.
- Shih H.Y. 2006. Network characteristics of drive tourism destinations: An application of network analysis in tourism. *Tourism Management* **27**(5): 1029–1039.
- Spielman D. 2012. Spectral graph theory. *Combinatorial Scientific Computing* **16**.
- Wee H.L., Yeo K.K., Chong K.J., Khoo E.Y.H. & Cheung Y.B. 2018. Mean rank, equipercentile, and regression mapping of World Health Organization quality of life brief (WHOQOL-BREF) to EuroQoL 5 dimensions 5 levels (EQ-5D-5L) utilities. *Medical Decision Making* **38**(3): 319–333.

Zhang S., Yang Z. & Wang C. 2024. The border tourism hotspots network based on travelogues. *Heliyon* **10**(13): e33260.

Department of Mathematical Sciences

Faculty of Science and Technology

Universiti Kebangsaan Malaysia

43600 UKM Bangi

Selangor, MALAYSIA

*E-mail: aimanhakim4795@gmail.com, sakhinah@ukm.edu.my, fatima84@ukm.edu.my**

Received: 7 January 2026

Accepted: 14 March 2026

*Corresponding author