

**Communities and Key Players Detection in Small and Medium airports in Asian Countries and Japan**

**Report Submission for Web Analytics (EB5202)**

Team Members

Ang Wei Hao Max - A0178278L

Bharat Nagaraju - A0178258N

Khine Zin Win - A0163222U

Koo Chee Kit - A0178200N

 Li Xiaoguang - A0178450B

Lin JunLiang - A0178295M

Contents

[Introduction 2](#_Toc528644149)

[Chapter 1 Small and Medium airports in Asian Countries 2](#_Toc528644150)

[1.1. Community detection 2](#_Toc528644151)

[1.1. Key players 4](#_Toc528644152)

[Chapter 2 Domestic airports in Japan 5](#_Toc528644153)

[2.1. Data Collection 5](#_Toc528644154)

[2.1 Community Detection 6](#_Toc528644155)

[2.2 Algorithms Evaluation 9](#_Toc528644156)

[2.3 Key Player 9](#_Toc528644157)

[3 Conclusion 10](#_Toc528644158)

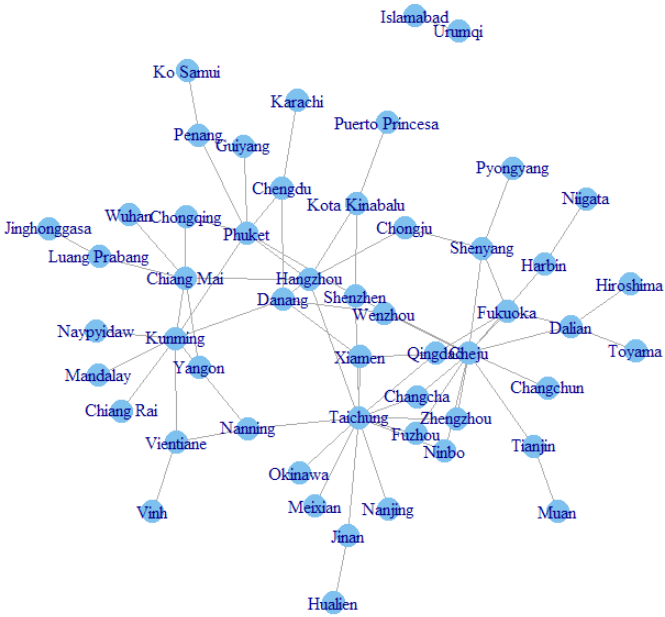
# Introduction

International airport is an indispensable hardware facility in any commercial metropolis or tourist city in the world. It is a key factor for local government in building a world logistics and tourism center.

This analysis is to detect the communities in the given network. That is, a subgroup in this network includes a set of airports that has a relatively large number of internal ties, and also relatively few ties from the group to other parts of the network. The benefit of the finding is conducive for the airport belonging governments to work closely together to form international corporations for building better logistics infrastructure. It also supports strategic planning for airports’ future developments.

Furthermore, this analysis is to discover the predominant actors in the communities, which play important role in the local network. The key players who have high centrality have bigger chances to become the future stars to handle great passenger volume and cargo flow like Singapore. It is very meaningful in the decision making process for local governments, especially in sense of business and investment.

# Chapter 1 Small and Medium airports in Asian Countries

This project utilizes the airline routes data for second tier airport web analysis. It looks at the potentials and tries to discover the future predominance in the regional airline network in Asian counties.

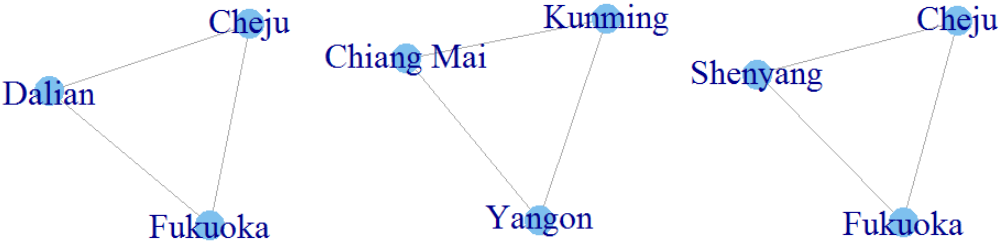
The project team purposely excludes developed regional air hubs and heavy traffic airports such as Singapore, Hong Kong and Shanghai etc. It specifically looks at the medium size and small airports, which have less than fifty international routes connecting the region.

52 airports with 172 pair flights in 12 Asian counties (Burma, China, Japan, Laos, Malaysia, North Korea, Pakistan, Philippines, South Korea, Taiwan, Thailand, and Vietnam) form the base data. Each of the airports has bilateral flights with its peer city. So it is non-directed connections linking the airports in the base data.

## Community detection

A clique is a maximally complete subgraph. It is a subset of nodes that have all possible ties among them, and one of the simplest types of cohesive subgroups.

In the base data, 3 maximum cliques are found. They are also the largest cliques. Each of them contains 3 air ports only. It demonstrates that, each airport in this small data set is relatively loose and not everyone is tightly connected to any other airports.



On the other hand, modularity is an important characteristic of a network, which is used in many community detection algorithms. It is a measure of the structure of the network, specifically the extent to which nodes, in our case airport, exhibit clustering where there is greater density within the clusters and less density between them (Newman 2006).

Modularity is very flexible and can be used in different fashions. The statistic can range from −1/2 to +1. The closer to 1, the more the network exhibits clustering with respect to the given node grouping.

In this concern, the team applies three different community detection algorithms to detect and locate the different communities in the given airport network.

**Walktrap Algorithm**

With Walktrap.community algorithm, 11 communities are detected from all the airports. The largest community consists of 17 airports of which most locate in China. The second largest community consists of 15 airports. More interestingly, for the rest of 9 communities detected, most of them are airport pairs, and only two of them consist of 3 airports.

**Leading Eigenvector Algorithm**

The algorithm of the Leading.eigenvector.community presents 6 findings in the community detection. Not only the number of communities detected is smaller than that from the Walktrap, but also the number of members in the biggest community is also less. Another major difference is, it tries to group the members into sub-groups. That is, instead of assigning small number of members to a separate community, most of the communities consist of at least 5 members.

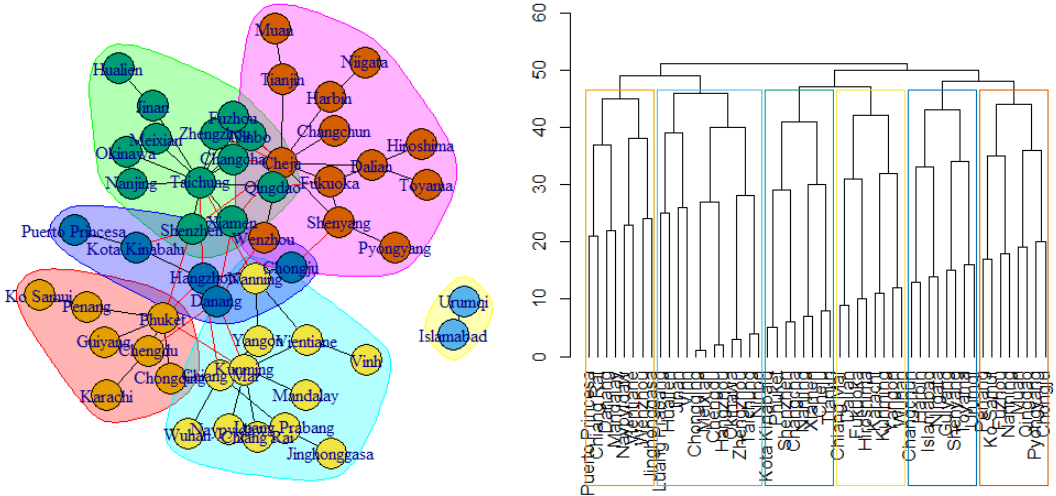
**Greedy optimization algorithm**

Further, below shows the community structure via greedy optimization of modularity. 9 communities in total are detected. Quite similar to the Walktrap algorithm, 4 of the communities only consist of two airports. The result of the greedy method seems sits in the middle of the between the Walktrap and Leading.eigenvector algorithm.

R can produce modularity score from each algorithm. When putting the modularity scores together from the 3 algorithms, we can see the difference. Even though the difference between the Leading.eigenvector to greedy optimization is not very obvious, it suggests that the Leading.eigenvector algorithm has done a better job at detecting subgroup structure.



For illustration of the discovery of communities, the results are plotted in two different formats.



For the next section, in identifying the key players in the communities, the team uses the results from the Leading.eigenvector community detection algorithm.

## Key players

Key players in the community are any airports perform a more important role amongst their peers. The recommendation is a heads-up to local governments to continue to expand airport facilities and various tourism and relating infrastructures such as rail and shipping facilities to win and hold profits.

Likewise to other industries, one of the most interesting finding in network analysis for airport and route is the prominent actors. It makes intuitive sense that a network member who is connected to many other members of the network is in a prominent position. For non-directed networks, this type of actor has high centrality, or it is in a central position.

There are dozens of centrality statistics available. In catching the key players in the predefined communities, the team applied 3 most common measures of centrality, which are degree, closeness and betweenness.

Taichung, Cheju and Kunming are found as stop players from the top 3 communities. Below table shows the statistics, from which we can see, even though they are the top layers of their subgroups, the statistics decrease along the size of communities.



In addition, the team searched other methods to identify key players. When searching for the optimal set of key players, a greedy search algorithm was originally proposed in Borgatti (2006). Weihua An and Yu-Hsin Liu revised the algorithm in multiple ways to enhance its usability and efficiency, and created an R Package, Keyplayer.

The basic idea of the algorithm is to select a set of nodes as seeds and then swap the selected nodes with unselected ones if the swap increases the group centrality, when the algorithm proceeds:

* Step 1. Select an initial candidate set C. The residual set is denoted as R.
* Step 2. Update the candidate set C.
* Step 3. Return the final set C and the centrality score.

The team applied this algorithm in the top 3 communities predefined. The findings are tabulated as below:



The results are aligned with the findings from the traditional statistics. Further, this package can easily produce and identify the key players with the given range, eg. front 2 players in each community. This provides an alternative to save human labor when reading the results.

# Chapter 2 Domestic airports in Japan

Based on flights from Japan, the goal is to identify domestic paths between source and destination of airports. This will help create an understanding of connecting flights domestically in Japan.

## Data Collection

Datasets were taken from openflights.org and joined using R sqldf(). Initial dataset created by routes.dat inner joining with airport.dat on common IATA number. This was further filter by country in (‘Japan’). To further analyse to specific flights in Japan, dataset used for community detection is filter by country in (‘Japan’) and destination in ('Haneda Airport','Narita International Airport', 'Kansai International Airport','Fukuoka Airport','New Chitose Airport'). The dataset is created by SQL, joining routes.dat and airport.dat from openflights.org.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **'data.frame': 66056 obs. of 16 variables:** | | | | |
| $ Airline\_ID  $ Source\_Airport\_ID  $ Codeshare  $ Stops | $ Source\_Airport  $ Des\_Airport\_ID  $ Des\_Airport | $ Equipment  $ name\_source  $ city\_source | $ Country  $ IATA  $ ICAO | $ name\_des  $ city\_des |

## Community Detection

**Community Detection Algorithms Used**

Community detection algorithms are used to identify significant community between destination cities in Japan (predominantly). Community structure and density can be observed from the network’s grouping of nodes. Overlapping community findings implies that dense inter-connection and sparse intra-connections between groups. 3 community detection algorithms are based on Japan’s top 5 busiest airports of 2017 (according to reports from Japan’s Ministry of Land, Infrastructure, Transport and Tourism). The top 5 airports are 'Haneda Airport (Tokyo)', 'Narita International Airport (Tokyo)', 'Kansai International Airport (Osaka)', 'Fukuoka Airport (Fukoka)' and 'New Chitose Airport (Sapporo)'.

**1. wt = walktrap.community(g1)**

Walktrap is a bottom-up approach that split the graph according to random walks. Random walks performed are about the same within a community as they will differ by a few edges only. Separate communities are formed from the results of the walk.

|  |
| --- |
| **Walktrap Plot by 5 busiest airport** |
|  |

**Figure 1**

**2. le = leading.eigenvalue.community(g1)**

The leading eigenvalue is a top-down approach that split the graph according to modularity. Each split is determined by evaluating the eigenvalue of the modularity mix.

|  |
| --- |
| **Leading Eigenvalue Plot by 5 busiest airport** |
|  |

**Figure 2**

**3. eb = cluster\_edge\_betweenness(g1)**

The edge betweeness measures the shortest path between nodes, gradually removing edge with high betweeness scores and removing the one with the highest score.

|  |
| --- |
| **Edge Betweeness Plot by 5 busiest airport** |
|  |

**Figure 3**

|  |  |
| --- | --- |
| Walktrap (Directed) | Leading.eigenvector (Directed) |
|  |  |
| Cluster\_Edge\_Betweeness (Directed) | Regions of Japan |
|  | https://upload.wikimedia.org/wikipedia/commons/c/c1/Regions_and_Prefectures_Japan.png |

**Figure 4**

## Algorithms Evaluation

Based on the plots above and in comparison with Regions of Japan Map (Figure 4), the 3 algorithms produce close representation of Japan and its airports relations. Japan’s busiest airports’ location (Tokyo, Osaka, Fukuoka and Sapporo) are central and are generally nodes with more links. The algorithms differ in their categorization of cluster and overlaps. However, there is no clear representation of the clusters.

Walktrap and leading.eigenvalue identify clusters that includes destination out of Japan as well, although leading.eigenvalue manages to capture only Japan’s region in its second cluster. Based on directed community (Figure 4), there is a clear indication of pointing towards the bigger airports by the others in the region.

**Modularity Evaluation**

|  |  |
| --- | --- |
| Walktrap.community | 0.09726697 |
| Leading.eigenvector.community | 0.2587694 |
| Cluster\_edge\_betweeness | 0 |

Modularity is a measure of cluster bond to each other. From the scores above, it is evident that leading.eigenvalue has the highest modularity and cluster\_edge\_betweeness having the least. This can be seen in the graphs above.

## Key Player

**Cliques Evaluation**

Largest clique can be found in the following results. Maximum clique stands at 5 and include the major cities in Japan. Sapporo, Fukuoka, Tokyo, Osaka, Sendai and Okinawa are the key player with centrality.

[[1]]

+ 5/34 vertices, named, from efa8f46:

[1] Sapporo Fukuoka Tokyo Osaka Sendai

[[2]]

+ 5/34 vertices, named, from efa8f46:

[1] Sapporo Fukuoka Tokyo Osaka Okinawa

# Conclusion

This project points out the subgroups of medium and small size airports in the region. The finding is for local government to re-adjust the local civil aviation policy for future developments. Members sit together in the same community, shall eye on and strengthen corporations amongst the airports. The goal is to grow together to a mega international airport alliance to push regional economy.

In this context, affiliates can work together closely in future in the areas of logistics, route planning, tourism and marketing etc to achieve win-win situation, for example 2 airports in Taiwan, 1 in Japan and the rest in China in the group of Changcha, Fuzhou, Hualien, Jinan, Meixian, Nanjing, Ninbo, Okinawa, Qingdao, Shenzhen, Taichung, Xiamen and Zhengzhou.

On the other hand, findings of key players in this project shows Taichung (Taiwan), Cheju (South Korea), Kunming (China) play important role individually in their communities in the connections with other airports in the region. They are the likely to grow and become future stars in the airport industry developments. For now, in addition to assisting the mega airports in the region, they need to expand their capabilities as international airports in phases by increasing number of flights and airport routes. Airport carrying capacity and infrastructure construction are all matters in the development of their role as important gateways for cargo flow and passengers.

For Japan domestic air industries, it can be observed that flights routes in Japan will be centralise towards the major cities in Japan. This is evident from the directed routes towards the centre of all three graphs. The graphs are good to point out the routes but will not be accurate on the geographic make-up of Japan’s region.