Topological Data Analysis on the Northeast Asian Air Route Network

Seyun Kim
Department of Civil and Environmental Engineering
Korea Advanced Institute of Science and Technology
Daejeon, South Korea
whataud@kaist.ac.kr

Yuyol Shin
Department of Civil and Environmental Engineering
Korea Advanced Institute of Science and Technology
Daejeon, South Korea
yuyol.shin@kaist.ac.kr

Yoonjin Yoon
Department of Civil and Environmental Engineering
Korea Advanced Institute of Science and Technology
Daejeon, South Korea
yoonjin@kaist.ac.kr

***Abstract*—Studying the properties of airport network of various countries based on complex network theory has been well researched over the past decade. However, there were only a few attempts made to utilize air route network with waypoints and air traffic service routes. In this study, three variations of networks – unweighted, distance-weighted, and demand-weighted air route networks of the rapidly growing Northeast Asian region are considered, and various node centrality measures are applied. Based on measured values, nodes sharing common characteristics are identified by mapper algorithm, one of the main approaches of topological data analysis. Mapper output network successfully grouped waypoints sharing common characteristics such as terminal nodes and nodes on busy air routes. Findings on the set of important waypoints provide key insights for effective regional airspace design.**

Keywords-air transport network; air route segment; complex network; weighted network; topological data analysis

I. INTRODUCTION

Air traffic continues to grow, and most of the rapid growth is caused by the travel demand in China. As high demand of air transportation leads to the increase of the complexity of the air transport system, there have been numerous researches which represented air transport system based on complex network theory [1-3], since complex network approach is one of the powerful ways to represent the system by connections of elements.

However, most of the air transport network researches modeled it as airport network having pairs of airports connected by hypothetical airline routes even though the actual air transport system has more elements and is much more complex. To overcome this limitation, several studies modeled air transport network as the air route network that consists of air navigational waypoints and tracks [4-6].

Since the air route is where air traffic goes by, air route network needs to be considered with spatial and traffic information. In this study, therefore, we modeled air route network of Northeast Asian regions containing China, Japan, and South Korea. By applying topological data analysis on calculated node centrality values from the Northeast Asian air route network, this study proposes a new approach to classify nodes and find critical nodes in the network.

II. AIR ROUTE SEGMENT NETWORK

A. Modeling Northeast Asian Air Transport Network

The Northeast Asian air transport network was modeled as network of air route segments, which consists of a track and connecting two waypoints. Air route segment network is represented as an undirected and connected graph $\mathrm{G} = (\mathrm{V}, \mathrm{E})$ defined by an adjacency matrix $A = (a_{ij}) \in R^{n \times n}$. Air route information used in the modeling was extracted from the Aeronautical Information Service (AIS) of China, Japan, and South Korea. The Northeast Asian air route segment network contains 1,459 waypoints (nodes) and 1,985 tracks (links) extracted from 545 air routes as shown in Fig. 1.

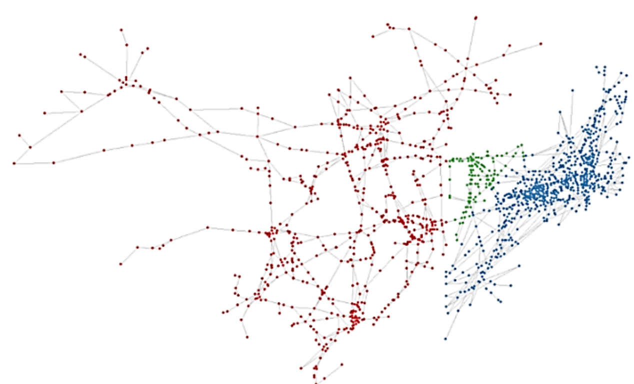

Figure 1. Northeast Asian air route segment network. Nodes of each national network are colored by red (China), blue (Japan), and green (South Korea).

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government (Ministry of Science and ICT) (NRF-2017R1E1A1A01076315), and financially supported by Korea Ministry of Land, Infrastructure and Transport(MOLIT) as 「Innovative Talent Education Program for Smart City」

B. Weighted Northeast Asian Air Route Segment Network

In air route segment network, naturally, some air routes connecting major airports are crowded, while others are not. In addition, the length of each air route segment varies from several meters to hundreds of kilometers. This study modeled weighted air route segment network by assigning weights to links in the Northeast Asian air route segment network to better describe features in the air transport network.

While the air route segment network as unweighted network is defined with the binary adjacency matrix, a weighted version is defined with the weighted adjacency matrix $A^w = (a_{ij}^w) \in R^{n\times n}$. Two types of weights – link distance and link-wise air traffic demand were considered. The distance-based network used geographical distance of each air route segment as link weight a_{ij}^w, and in the demand-based network, link-wise air traffic volume was used as weight. The air traffic volume was estimated based on the ADS-B based actual flight trajectory data sourced from the online flight tracking service 'FlightAware'. 16,734 flights between 114 busiest airports in China (50), Japan (50), and South Korea (14) on July 1, 2018 were included, which address 93.5% of total traffic in the region.

III. Methodology

In this study, node importance was measured by node centrality measures including degree, closeness, and betweenness. Based on node importance values, Topological Data Analysis (TDA) was applied to find nodes acting an important role in the network.

A. Node Centrality Measures

In the unweighted Northeast Asian air route segment network, node importance was measured by three node centralities including *degree*, *closeness*, and *betweenness* [7]. *Degree* of node i is the number of links connected to node i, which is defined as $k_i = \sum_j a_{ij}$. *Closeness* of node i is defined as $c_i = \frac{1}{\sum_j d_{ij}}$, where d_{ij} is the distance of the shortest path connecting the node i and all other node j $(i \neq j)$. *Betweenness* is the frequency of a node included in the shortest path between all possible connected pairs of nodes, and defined as $b_i = \sum_{j\neq k} \frac{n_{jk}(i)}{n_{jk}}$. n_{jk} is the number of shortest paths connecting the node *j* and *k*, and $n_{jk}(i)$ is the number of shortest paths between node *j* and *k* that include node *i*.

In the weighted networks, node importance could be measured with three more node centralities – strength, weighted closeness, and weighted betweenness [8]. Strength is a sum of weight of connected links to a node, which is most relevant in the demand-based network. This study utilized weighted closeness and weighted betweenness by applying link distance as link weight in the calculation of each measure since the shortest path extraction is based on link distance. Therefore, weighted closeness and weighted betweenness are calculated based on the distance-based weighted network.

B. Topological Data Analysis and Mapper Algorithm

Topological Data Analysis (TDA) is used to analyze the nodes in the air route segment network of Northeast Asian countries. TDA is a field that emerged from algebraic topology, which aims to analyze complex topological structures by using topological and statistical approaches [9-10]. Of the two main methods of TDA, the persistent homology and mapper, the mapper algorithm is applied to our dataset.

TDA mapper is a partial clustering algorithm that explains the continuity among the clusters. With the chosen filter function $\boldsymbol{f}$, the mapper algorithm can capture the local differences of data. The major components of the mapper are the filter function $\boldsymbol{f}$, and cover, which consists of the number of intervals, $\mathbf{S} = \{\mathrm{s}_1, \dots, \mathrm{s}_\mathrm{n}\}$, and overlap, $\mathbf{U} = \{\mathrm{u}_1, \dots, \mathrm{u}_\mathrm{n}\}$.

Given a set of N-dimensional data observations $\mathbf{D} \subset \mathbb{R}^N$, filter function $\boldsymbol{f}$ maps $\mathbf{D}$ into lower dimension n and, in most of the time, the number of dimensions for projected data is either 1 or 2. A filter function can be expressed as,

$$\boldsymbol{f}: \boldsymbol{D} \rightarrow (X_1, \dots, X_n). \quad (1)$$

Filter functions frequently implemented in other studies include t-SNE [11], principal components [12], eccentricity, and k nearest neighbor distance.

After the filter function projects the data into a lower dimension, the range of each dimension of the filter output is divided into s_i equal intervals. Then the overlap extends both ends of each interval by the multiplication of the interval length and u_i.

The last step of the mapper is to find the data contained in each interval of the cover, and perform clustering separately for the data in each interval in the original dimension, $\mathbb{R}^\mathrm{N}$. Finally, in the output network, each cluster is a node and if there exists any common data point in two different clusters, a link connects the nodes. For this step, any clustering algorithm can be used.

Fig. 2 demonstrates a simple application of mapper algorithm to point cloud data. In the example, the filter function is the x-coordinate and cover parameters, the number of intervals and percent overlap are 5 and 0.2, respectively. Since the mapper algorithm takes the concept of nerve and cover, it has topology-like advantages in data analysis application. First, the output is free of coordinates. Suppose that in the example in Fig. 2, x-coordinates range from 100 to 104.2. It still has the same output. Second is that it is robust against outliers and small deformations. Outliers would just be shown as separated nodes and the small deformations, or noises will be integrated into each node in the output network. The last advantage is strong visualization of the result. Through the visualization, we can decide the region, or parts of data to investigate and discover new knowledge of the underlying topology of data.

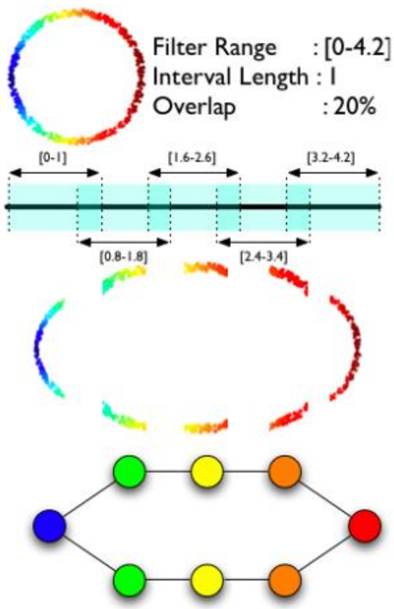

Figure 2. Mapper algorithm application example [9]

IV. Analysis Results

A. Node Centrality Distribution

Six node centrality values including degree, closeness, betweenness, strength, weighted closeness, and weighted betweenness of nodes in the Northeast Asian air route segment network are shown in scale in Fig. 3. Among unweighted centralities, one can observe that higher degree nodes are sporadically located inside the national airspace of China and Japan. On the other hand, nodes of higher closeness are concentrated around the East China Sea. Nodes with higher betweenness are located along the air routes connecting three national airspaces, identifying a limited number of nodes, or the 'pseudo-hubs', connecting three individual national networks.

In weighted centrality values, one can observe that top strength nodes are located on the domestic routes connecting major cities in China, such as Beijing and Guangzhou, exhibiting the evidence of air travel demand concentration among the major cities domestically. Top nodes of weighted closeness are located around South Korean airspace to radially decrease outward, in a similar fashion in the unweighted network. Nodes of higher weighted betweenness are found in the East China Sea and the Yellow Sea.

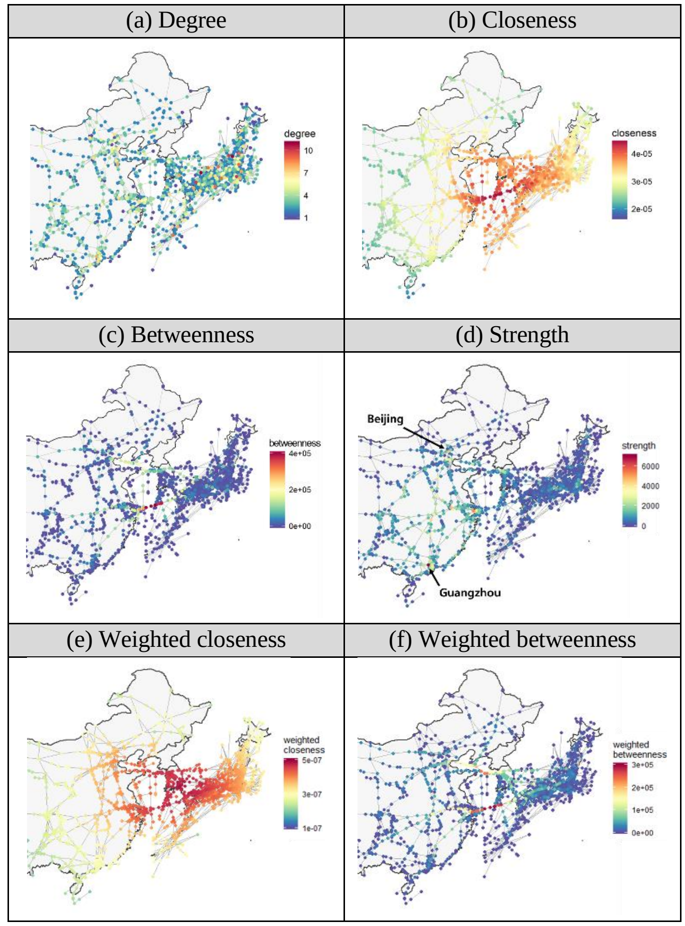

Figure 3. Six node centrality plots of the unweighted and weighted Northeast Asian Air Route Segment Network.

B. Mapper Results

In this study, the mapper algorithm was applied to six node centrality values of all 1,459 nodes in the Northeast Asian air route segment network. The mapper output network shown in Fig. 4 was obtained by using t-SNE for filter function, and DBSCAN algorithm for clustering. From the big connected graph shown in the center of Fig. 4, three branches (A, B, C) shown to have distinctive characteristics and their memberships were compared by coloring waypoints as shown in Fig. 5.

One can observe that a certain number of waypoints in A are terminal nodes in the Northeast Asian air route segment network. Those waypoints have a degree of 1 and show low closeness and betweenness values (Fig. 3). While locations of waypoints in B look similar to ones in A, waypoints in B share common characteristics that they are located in air routes for long-haul trips and have a degree of 2. On the other hand, waypoints in C are clearly found in busy air routes of domestic flights in each country as well as international flights from China to the rest of the region. Unconnected nodes in the mapper output network

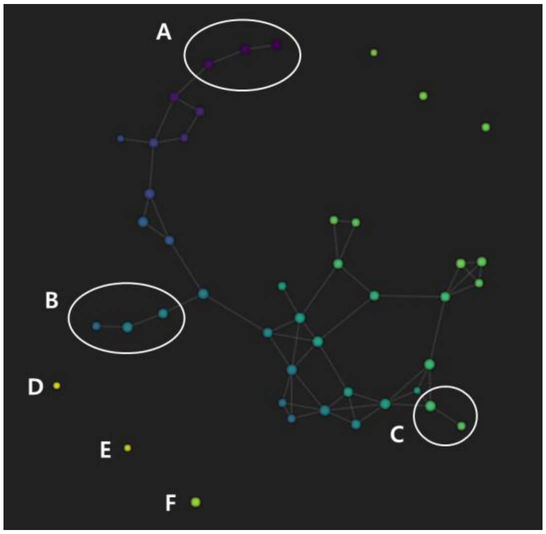

Figure 4. Mapper output network

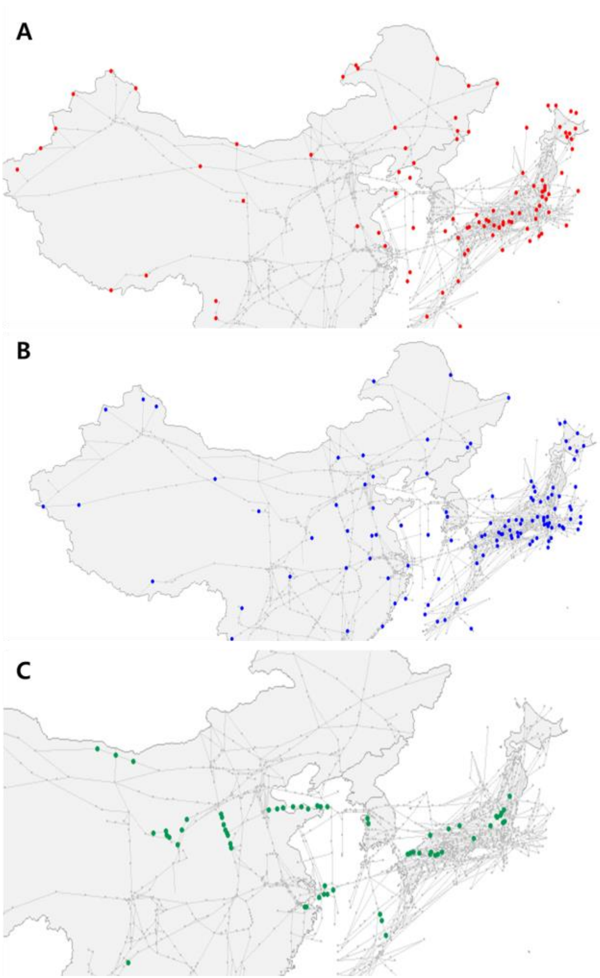

Figure 5. Clustered nodes from selected branches in the mapper output network represented on the Northeast Asian air route segment network.

including D, E, and F also have distinctive characteristics, especially for D and E having waypoints on the air route connecting national airspace, and F mainly including waypoints on well-used South Korean domestic air routes. Looking at the role of each waypoint in the regional air route network, TDA perspective analysis fits very well.

V. Conclusions

In this study, air transport network of the rapidly growing Northeast Asian region was modeled and analyzed based on the unweighted and weighted air route segment networks. Geographical distance was used for constructing distance-based network, and flight demand was used for demand-based network. The air route network approach has the strength in analyzing airspace status compare to traditional airport network approach.

Nodes with certain characteristics and critical nodes were identified by six different node centrality measures and TDA mapper algorithm. Mapper output network successfully grouped waypoints sharing common features such as terminal nodes and nodes on the busy and important air routes. Findings on these node clusters provide key insights for adjusting air route network in the Northeast Asian region. Our future research will focus on including more information from nodes and trying to apply various parameters and filter functions on TDA analysis.

References

[1] O. Lordan, J. M. Sallan, P. Simo, and D. Gonzalez-Prieto, "Robustness of the air transport network," Transp. Res. E: Logist. Transp. Rev., vol. 68, 2014, pp.155-163.

[2] M. G. Song, and G. T. Yeo. "Analysis of the air transport network characteristics of major airports," Asian J. Shipp. Logist., vol. 33, no. 3, 2017, pp.117-125.

[3] X. Sun, V. Gollnick, and S. Wandelt, “Robustness Analysis Metrics for Worldwide Airport Network: A Comprehensive Study,” Chin. J. Aeronaut., vol. 30, no. 2, 2017, pp.500-512.

[4] K.-Q. Cai, J. Zhang, W.-B. Du and X.-B. Cao, “Analysis of the Chinese Air Route Network as a Complex Network,” Chin. Phys. B, vol. 21, no. 2, 2012, 028903.

[5] X. Sun, and S. Wandelt, “Network Similarity Analysis of Air Navigation Route Systems,” Transp. Res. E: Logist. Transp. Rev., vol. 70, 2014, pp. 416-434.

[6] S. Kim, and Y. Yoon. "On node criticality of the Northeast Asian air route network", J Air Transp. Manag., vol. 80, 2019, 101693.

[7] L. C. Freeman, “Centrality in social networks conceptual clarification,” Soc. Netw., vol. 1, no. 3, 1978, pp. 215-239.

[8] T. Opsahl, and P. Panzarasa, “Clustering in weighted networks,” Soc. Netw., vol. 31, no. 2, 2009, pp. 155-163.

[9] P. Y. Lum et al., “Extracting insights from the shape of complex data using topology,” Sci. Rep., vol. 3, 2013, 1236.

[10] G. Singh, F. Memoli, and G. Carlsson, “Topological methods for the analysis of high dimensional data sets and 3d object recognition,” SPBG, 2007, pp. 91-100.

[11] L. Maaten, and G. Hinton, “Visualizing data using t-SNE,” J Mach. Learn. Res., vol 9, 2008, pp. 2579-2605.

[12] I. Jolliffe, “Principal component analysis,” International encyclopedia of statistical science, Berlin: Springer, 2011, pp. 1094-1096.