Passenger Flow Characteristics of Seoul Metropolitan Subway Network

Kang Won Lee, Jung Won Lee

***Abstract*—**Characterizing the network flow is of fundamental importance to understand the complex dynamics of networks. And passenger flow characteristics of the subway network are very relevant for an effective transportation management in urban cities. In this study, passenger flow of Seoul metropolitan subway network is investigated and characterized through statistical analysis. Traditional betweenness centrality measure considers only topological structure of the network and ignores the transportation factors. This paper proposes a weighted betweenness centrality measure that incorporates monthly passenger flow volume. We apply the proposed measure on the Seoul metropolitan subway network involving 493 stations and 16 lines. Several interesting insights about the network are derived from the new measures. Using Kolmogorov-Smirnov test, we also find out that monthly passenger flow between any two stations follows a power-law distribution and other traffic characteristics such as congestion level and throughflow traffic follow exponential distribution.

***Keywords*—**Betweenness centrality, correlation coefficient, power-law distribution, Korea traffic data base.

I. Introduction

A network is a collection of nodes and links. Regardless of fields, there are many systems consisting of components and their interactions, which can be considered as networks [1]. Internet, WWW, SNS, and transportation systems are some of the good examples. There have been numerous studies to find out dynamic behaviors of a network through network topology analysis.

A subway system can be also modeled as a network with stations as nodes and connections between two stations as links. There have been several studies to discover structural and flow characteristics of subway systems using network analysis methods. For example, Stoilova et al. [2] investigated 22 European subway systems. They defined new indicators such as degree of routing, density of route, and connectivity of route to analyze and classified the metro networks.

Betweenness centrality measure which was introduced first in social network analysis [1] has been used to identify the node importance and dynamic patterns of a network. Derrible et al. [3] compared various worldwide metro systems from the view of betweenness centrality and revealed the common trends at global level. Monterola et al. [4] demonstrated that the Singapore rapid transit system has a high correlation between betweenness centrality and passenger throughput.

Studies about subway networks using centrality measures also exist in Korea. Han et al. [5] carried out a network analysis on the Seoul subway system based on degree centrality, closeness centrality, and eigenvector centrality. However, they only focused on the rank of stations, which is insufficient to describe the entire dynamic network structure.

Since betweenness centrality simply takes account of the physical structure of subway networks, the passenger flows are missed out in the analysis. We, therefore, suggested a new measure called weighted betweenness centrality, which combines both actual passenger flows between stations and traditional betweenness centrality. By analyzing correlation coefficient between betweenness centrality and weighted betweenness centrality, we found out some interesting passenger flow characteristics of Seoul metropolitan subway system.

To fully identify and understand the characteristics of passenger flow of Seoul metropolitan subway network, some traffic characteristics are newly proposed in this study such as daily passenger flow between neighboring stations i and j, congestion level of each station, and throughflow traffic.

In this study we have introduced new betweenness centrality measure, which is weighted by amount of passenger flow. Traffic characteristics of passenger flow are also discussed. In this work we have attempted to summarize and present our previous works [6]-[8] of last few years about Seoul metropolitan subway network.

II. Seoul Metropolitan Subway Network

In our previous work [6], the characteristics and the efficiency of Seoul metropolitan subway network are analyzed. Since the subway network became the first in transportation mode share overtaking the bus system in 1996, the subway mode share has taken upturn and becomes 38.8% in 2013 as can be seen in the Fig. 1.

The Seoul metropolitan subway network consists of nine Seoul lines and seven suburban lines including New Bundang line, Incheon line, Suin line, Kyungui Central line, Kyungchun line, Airport line, Uijeongbu line, New Bundang line and Ever line. Seoul subway network has good transfer systems from one line to another and is also well connected to suburban subway networks.

Under the urban railroad construction plan of Seoul about 90 km of railroad route length (27% increase) will be added by the year 2025. In any place of Seoul, there will be a subway station within 10 minute's walking distance. In our previous work [6], we have analyzed the evolutionary topological change of Seoul metropolitan subway network between September, 2008 and

K. W. Lee is with Seoul National University of Science and Technology, Republic of Korea (e-mail: kwlee@snut.ac.kr).

September, 2013. We have found out followings.

- Number of lines and stations are increased by 45% (from 11 to 16) and 33% (from 371 to 493), respectively.
- Average shortest path length is increased by 22% (from 28.9 km to 35.1 km). This increase is due to expansion of Seoul metropolitan subway network to suburban area during this period, which can be verified through the enlargement of network diameter during this period. Network diameter is increased by 42% (from 143.3 km to 204.3 km).

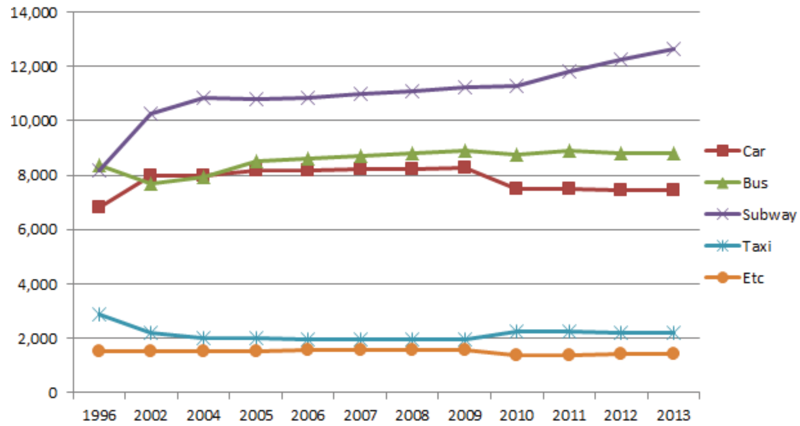

Fig. 1 Transportation Mode Share [6]

- Coverage, which is defined as the sum of the area surrounding the stations divided by the total served area, is increased by 26% (from 0.038 to 0.48). This means that Seoul metropolitan subway network underwent intensification stage as well as expansion stage during this period. It is turned out that coverage of Seoul metropolitan subway network is well above of subway network of rivalry cities of New York, Paris, London, Tokyo, etc.
- Clustering coefficients during this period remains the same around 0.02. The clustering coefficient is meaningless in subway network. However, if we consider other topological space such as P and R, it can have some meaning. In P-space, stations are expressed as nodes, but links exist when two nodes are on the same subway line and a direct route between them is available without any transfer. In R-space, nodes represent a subway lines. Links exist when it is possible to transfer from a line to another line directly, which means that a link exists if there is a transfer station between any two subway lines. The clustering coefficients are extremely high in P-space and R-space, 0.94 and 0.95, respectively.
- Efficiency of subway network can be calculated using following equations.

$$E_{glob}(G) = \frac{\sum_{i \neq j \in G} d_{ij}^{-1}}{\sum_{i \neq j \in G} l_{ij}^{-1}} \qquad (1)$$

In the above equation, dij is the shortest distance of track between station i and j. lij is the Euclidean distance between station i and j in the subway network G. Efficiency is just little bit increased by only 2.1% (from 0.729 to 0.745). Considering that the efficiency of Seoul metropolitan subway network in 2013 is just less by 25.5% compared to that of ideal network, which has direct paths between all stations, Seoul metropolitan subway network is a very efficient network. It is turned out that efficiency of Seoul metropolitan network is higher than those of networks of other cities such as Boston, Tokyo and Osaka.

III. Weighted Betweenness Centrality

A new measure called weighted between centrality is proposed in our previous work [7]. It is weighed by amount of monthly passenger flow between two stations and can be expressed as follows.

$$\text{WBC}(v) = \sum_{i \neq j \neq v} \left[\frac{F(i \to j)}{\sum_{k \neq l \neq v} F(k \to l)} \times \frac{\sigma(i \to j|v)}{\sigma(i \to j)} \right] \qquad (2)$$

$\boldsymbol{\sigma(i \to j)}$ is the number of shortest paths from station i to j. Usually it is 1, but it can be more than 1 if there are more than one shortest path between station i and j. $\boldsymbol{\sigma(i \to j|v)}$ is the number of shortest paths from station i to j containing station v at the same time. It is 1 if the path contains station v. Otherwise, it is equal to 0.

$F(i \to j)$ represents the number of passengers traveling from station *i* to *j* monthly. Every $(i \to j)$ pair has, when station *v* is contained in a specific shortest path, a different weight depending on the passenger flow amount between station *i* and *j*. Weighted betweenness centrality without the term $\sum F(i \to j)$ represents the number of on-board passenger at station *v*, which can be interpreted as congestion level in station

Computations of betweenness centrality and weighted betweenness centrality have been executed with Python. A library called Networkx has been used to construct the network weighted by traveling time in order to figure out the shortest paths, which are calculated using Dijkstra's algorithm. 2 minutes of transfer time are considered to calculate shortest path. In [7], we have calculated betweenness centrality and weighted betweenness centrality, respectively. It shows the top 20 stations of BC and WBC.

IV. Characteristics of Passenger Flow

In our previous study [7], we have shown that passenger flow between any two stations follows a skewed distribution through the analysis of correlation coefficient between BC(Betweenness Centrality) and WBC. It was also shown that $\boldsymbol{F(i \rightarrow j)}$ follows a power-law distribution.

$$\rho = \frac{cov(BC,WBC)}{\sigma_{BC} \cdot \sigma_{WBC}} \tag{3}$$

To fully identify and understand the characteristics of passenger flow of Seoul metropolitan subway network further traffic measures are required as follows.

- Traffic characteristics $\boldsymbol{Q(i \rightarrow j)}$, which represent daily passenger flow between neighboring stations *i* and *j*
- Congestion level of each station C_i, which can be determined as summation of through passenger of station *i*, passenger starting from station *i*, and passenger arriving at station *i*.
- Throughflow traffic, INF_i and $OUTF_i$. INF_i is defined as incoming-throughflow traffic amounts to station *i*. It can be determined as summation of incoming passenger amounts to station *i* from outside and its neighboring stations. $OUTF_i$ is defined as outgoing-throughflow traffic amounts from station *i*. It can be determined as summation of outgoing passenger amounts from station *i* to outside and its neighboring stations.

In our previous study [8], we have investigated the statistical properties of $\boldsymbol{Q(i \rightarrow j)}$, C_i, INF_i, and $OUTF_i$. Through these investigations and the result of $\boldsymbol{F(i \rightarrow j)}$ of previous study [7], we can get more meaningful insight about passenger flow characteristics of Seoul metropolitan subway network and its efficiency to handle passenger traffic. The data used in this study come from Korea Transportation DB [9]-[11].

V. Analysis Results

We have constructed CCDFs (Complimentary Cumulative Density Function) of $\boldsymbol{Q(i \rightarrow j)}$, C_i, INF_i, and $OUTF_j$, respectively. It is found out that $\boldsymbol{all\ of\ Q(i \rightarrow j)}$, C_i, INF_i, and $OUTF_i$ do not follow Power-law distribution through Kolmogorov-Smirnov test [8].

Fig. 2 shows the CCDF of $\boldsymbol{Q(i \rightarrow j)\ in\ Log - Log\ scale}$. Since the graph is not straight line., we can conclude that $\boldsymbol{Q(i \rightarrow j)}$ does not follow power-law distribution. And its P-value is less than 0.01. Other CCDF graphs of C_i, INF_i, and $OUTF_j$ are very similar to Figure 1 except the scale of variable.

To find out proper distributions $\boldsymbol{Q(i \rightarrow j)}$, C_i, INF_i, and $OUTF_i$ follow, we have used R-programming. For candidate distributions we have used Exponential, Lognormal, and Weibull distributions. Exponential distribution is a representative for short-tail distribution and lognormal for Long-tail distribution.

R, p = fit.distribution_compare('power_law', 'exponential', normalized_ratio=True)

R, p = fit.distribution_compare('power_law', 'lognormal', normalized_ratio=True)

R, p = fit.distribution_compare('power_law', 'weibull', normalized_ratio=True)

In [8] it was found out that Exponential distribution is the best fit for all of $\boldsymbol{Q(i \rightarrow j)}$, C_i, INF_i, *and* $OUTF_i$. Exponential distribution is a short-tail distribution compared to long-tail distribution of Power-law, whose probability decreases rapidly as value of variable grows. Power-law distribution is characterized by a highly heterogeneity. The fact that $\boldsymbol{Q(i \rightarrow j)}$, C_i, INF_i, *and* $OUTF_i$ follow Exponential distribution says that traffic flow of passenger in Seoul metropolitan subway network is relatively homogeneous. That is, there is no extra-ordinary high traffic of passenger flow compared to average in Seoul metropolitan subway network.

To get a more meaningful insight on the characteristics of passenger flow and efficiency of Seoul metropolitan subway network the statistical results should be investigated considering the topology of subway network and the spread of population. It is still under our investigation.

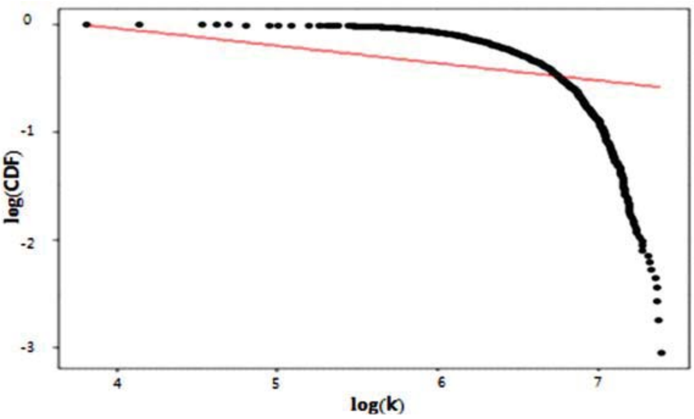

Fig. 2 CCDF of Q(i→j)

VI. Conclusion

The importance of metropolitan subway network in Seoul cannot be overemphasized. The transportation mode share of subway system in Seoul is absolutely higher than those of other transportation systems. In our previous works, we have went through the analysis of topological characteristics and efficiency of Seoul metropolitan subway network. Also, we have seen the evolutionary changes between September, 2008 and September, 2013. We could find out that Seoul metropolitan subway network went through expansion and intensification stages during this period. In our previous work, considering monthly passenger flow between any two stations we have also introduced the weighted betweenness centrality. It is a combined measure of topological characteristics and passenger traffic flow of subway network. In our previous work, we have also shown that the values of weighted betweenness centrality of stations and lines are pretty much different from those of traditional betweenness centrality, which does not consider passenger traffic flow. The subway network can be constructed using different topological spaces (L-space, P-space and R-space). L-space is a conventional one where node represents a station and link represents a connection between neighboring stations. The topological measures have different meanings depending on the network topological spaces. By combining and analyzing all the measures from three different spaces, we could have more clear understanding and direction of future planning of Seoul metropolitan subway network.

To identify the characteristics of passenger flow of Seoul metropolitan subway network and its efficiency, we have investigated the statistical properties of traffic characteristics such as $\boldsymbol{F(i \to j)}$, $\boldsymbol{Q(i \to j)}$, C_i, INF_i, and $OUTF_i$. For our analysis, we have used the Pearson correlation coefficient and Kolmogorov-Smirnov test for power-law. It is found that $\boldsymbol{Q(i \to j)}$, C_i, INF_i, and $OUTF_i$. do not follow Power-law distribution, but Exponential distribution. That is, there is no extra-ordinary high traffic of passenger flow compared to average passenger traffic in Seoul metropolitan subway network. These results are different from those of subway network of other cities such as Beijing, San Francisco and Boston. They follow Power-law distribution, which means their passenger traffics are pretty much biased. Further study is required to answer the question why passenger flow characteristics of Seoul metropolitan subway network do not follow power-law. To answer this question topological structure of subway network, passenger flow characteristics and population distribution in Seoul metropolitan area should be jointly considered.

It is still under study to get some meaningful insights of passenger flow characteristics in Seoul metropolitan subway network from our investigations. The results can be used as some guidelines for 2025 urban subway construction plan of Seoul metropolitan area.

References

[1] M. E. J. Newman. (2010). Networks: An Introduction, UK: Oxford University Press.
[2] S. Stoilova., & V. Stoev. (2015). An Application of the graph theory which examines the metro networks, Transport Problems, Volume 10, Issue 2, pp 35-48.
[3] S. Derrible. (2012, July). Network centrality of metro systems, PloS One, Volume 7, Issue 7, e40575, pp 1-10.
[4] C. P. Monterola., M. A. Ramli., G. L. K. Khoon., & T. H. G. Guang. (2014). A Method to Ascertain Rapid Transit System's Throughput Distribution Using Network Analysis," Procedia Computer Science, Volume 29, pp 1621-1630.
[5] K. H. Han., Y. S. Seo., & G. B. Park. (2015). Subway stations network analysis by using social network analysis, KSR, pp 34-40.
[6] S. Y. ZZang., & K.W. Lee. (2016, June). Characteristics and Efficiency Analysis of Evolutionary Seoul metropolitan Subway Network, Journal of the Korean Society for Railway, Volume 19, No.3, pp.388-396.
[7] J. W. Lee, & K. W. Lee. (2017). Analysis of Seoul Metropolitan Subwy network Characteristics using Network Centrality Measures, Journal of the Korean Society for Railway, Volume 20, No.3, pp.413-422.
[8] K.W. Lee, & J.W. Lee. (2018). network Betweenness Centrality and Passenger Flow Analysis of Seoul Metropolitan Subway Lines, Journal of the Society of Korea Industrial and Systems Engineering, Volume 41, No.2, (to appear).
[9] Korea Transport Database (Aug.20 2017). http://www.ktdb.go.kr.
[10] Seoul Open Data Plaza (Sep.17 2017). http://data.seoul.go.kr.
[11] Korean Statistical Information Service (Sep.17 2017). http://kosis.kr.