

Network Betweenness Centrality and Passenger Flow Analysis of Seoul Metropolitan Subway Lines

Kang Won Lee[†] · Jung Won Lee

Department of Industrial and Systems Engineering, Seoul National University of Science and Technology

서울 수도권 지하철망의 호선별 망 매개 중심성과 승객 흐름 분석

이강원[†] · 이정원

서울과학기술대학교 산업정보시스템공학과

Using network betweenness centrality we attempt to analyze the characteristics of Seoul metropolitan subway lines. Betweenness centrality highlights the importance of a node as a transfer point between any pairs of nodes. This 'transfer' characteristic is obviously of paramount importance in transit systems. For betweenness centrality, both traditional betweenness centrality measure and weighted betweenness centrality measure which uses monthly passenger flow amount between two stations are used. By comparing traditional and weighted betweenness centrality measures of lines characteristics of passenger flow can be identified. We also investigated factors which affect betweenness centrality. It is the number of passenger who get on or get off that significantly affects betweenness centrality measures. Through correlation analysis of the number of passenger and betweenness centrality, it is found out that Seoul metropolitan subway system is well designed in terms of regional distribution of population. Four measures are proposed which represent the passenger flow characteristics. It is shown they do not follow Power-law distribution, which means passenger flow is relatively evenly distributed among stations. It has been shown that the passenger flow characteristics of subway networks in other foreign cities such as Beijing, Boston and San Francisco do follow power-law distribution, that is, pretty much biased passenger flow traffic characteristics. In this study we have also tried to answer why passenger traffic flow of Seoul metropolitan subway network is more homogeneous compared to that of Beijing.

Keywords : Betweenness Centrality, Weighted Betweenness Centrality, Correlation Coefficient, Power-Law

1. 서론

현재 서울 수도권 지하철은 9개 노선의 서울 지하철과 12개 노선(공항, 경의중앙, 경춘, 분당, 수인, 신분당, 인천 1, 인천 2, 의정부, 자기부상, 에버, 경강)의 교외선으

로 구성되어 있으며 역의 개수도 567개에 이른다. 9개 노선의 서울 지하철은 우수한 환승 시스템을 갖추고 있으며 12개의 교외선들과도 잘 연결되어 있어서 모스크바, 동경, 뉴욕 지하철에 버금가는 세계에서 가장 큰 지하철 중의 하나가 되었다.

이러한 지하철 시스템은 각 역을 노드로 인접한 역 사이의 연결을 링크로 나타내 네트워크로 모형화할 수 있는데 이를 통해 지하철망의 위상적인 분석을 시도한 여러 연구가 있었다[8, 10, 12, 15]. 또한 네트워크의 중심성

(Centrality) 지표를 이용하여 지하철 망의 구조적 특성과 효율성을 조사한 여러 연구가 진행되었다. 일반적으로 네트워크의 중심성 분석에 사용되는 측도는 매개 노드 지표인 매개 중심성(Betweenness Centrality), 허브 노드 지표인 근접 중심성(Closeness Centrality), 그리고 차수 중심성(Degree Centrality) 등이 있다. 그런데 지하철 망의 특성상 단순히 노드들 간의 근접도 라든지 노드의 차수에 초점을 둔 근접 중심성과 차수 중심성은 지하철 망 분석에 큰 의미가 없어 보인다. 반면에 임의의 두 노드를 연결해주는 역할에 초점을 맞추는 매개 중심성은 지하철 망의 분석에 가장 많이 사용되는 중심성 측도다. 매개 중심성은 특정 노드가 네트워크에 존재하는 노드들의 모든 쌍(Pair)간 최단 경로에 얼마나 많이 존재하는가를 나타내는 지표다. 네트워크에서 각 노드의 중요성은 노드의 위치(Location) 보다는 임의의 두 노드 사이의 최단거리 상에 해당 노드가 존재 하느냐에 달려있다. 이 중요도는 지하철망의 경우에 매우 적절한 개념으로 매개 중심성으로 나타낼 수 있다. 특정 역은 그것이 중요한 위치(예로 중심 상업지역이나, 문화 오락 지역)에 있기 때문에 이용 빈도가 높을 수도 있지만 해당 역이 다양한 지역으로 가기 위한 사용 빈도가 높다면 그 중요도는 더 높을 수 있다. 또한 매개 중심성은 지하철 망에서 노드의 중요도를 평가 하는데도 이용 될 수 있는데 매개 중심성이 높은 노드에 고장이나 장애가 발생하여 이용 할 수 없게 된다면 네트워크 전체에 미치는 영향력이 가장 크게 되기 때문이다.

Derrible[3]은 지하철망의 역별 매개 중심성을 분석하고 지하철망의 크기가 증가할수록 각 역의 매개 중심성이 어떻게 변하는지를 조사하였다. 아울러 과포화 상태에 있는 역들의 부하를 줄이기 위하여 승객의 흐름을 재분배 하기 위한 기초 자료로 매개 중심성 지표를 조사하였다. Monterola 등[9]은 싱가포르 지하철 네트워크의 매개 중심성을 구하여 승객 효율(Passenger Throughput)과의 상관관계를 분석하였다. 승객 효율은 특정 역 v 를 지나는 승객들의 비율로 정의 된다. Haznagy 등[4]은 헝가리 지하철 네트워크의 매개 중심성과 근접 중심성을 구한 후 이를 토대로 지하철의 구조적 특성 등을 조사하였다. Cheng 등은[1] 지하철 네트워크의 중요 노드를 결정하고 네트워크를 개선하거나 고장에 대처하는 가장 효율적 대안을 조사하기 위하여 네트워크 매개 중심성 지표를 사용하였다. 아울러 새로운 중심성 지표인 승객 흐름 중심성(Commuter Flow Centrality)을 제안하였다. 한편 Xu 등[14]은 베이징 지하철 승객 흐름을 나타내주는 측도들을 제안하고 이들의 분포가 Power-law 분포

를 따르는지를 검증함으로써 지하철 망에서 승객 흐름의 편중 현상을 조사하였다. 그리고 Wang 등[13]도 샌프란시스코와 보스턴 URT(Urban Rail Transit)의 승객 흐름을 조사하여 이들 분포가 Power-law를 따르는 것을 보여 주었다. Power-law는 Web망 등[2] 시스템 내의 편중 현상을 조사하기 위해 사용되고 있다.

위에서 살펴본 기존의 매개 중심성을 이용한 연구들은 지하철 개별역들의 매개 중심성을 구하고 이로부터 지하철망의 특성과 효율성을 분석하였다. 그러나 지하철의 호선(Line)별로 호선별 매개 중심성을 구하여 이로부터 지하철망의 특성을 분석한 연구는 아직 없다. 역별로 산출한 매개 중심성에 비해 호선별 매개 중심성은 보다 거시적인 관점에서 지하철 망의 특성을 분석 할 수 있을 뿐더러 호선별로 망 이용에 대한 지표를 산출 할 수 있기 때문에 지하철 망 전체적인 관점에서 승객들의 흐름과 과포화 특성들을 파악 할 수 있어서 보다 효율적인 운영 및 관리가 가능하다. 다양한 역세권 특성이 지하철 이용 수요에 미치는 영향을 분석한 연구는[5] 있었지만 우리나라 지하철의 승객 흐름 분포의 편중 현상을 통계적으로 규명한 연구는 아직 존재 하지 않는다. 따라서 본 연구의 목적은 호선별로 매개 중심성 측도를 산출하고 이로부터 망의 효율적 운영 및 관리에 필요한 정보추출을 그 목적으로 삼았다. 아울러 수도권 지하철망의 승객 흐름 특성을 파악하기 위하여 인접한 두 역 사이의 월간 승객 흐름량, 역의 혼잡도, 유입 관통 흐름량, 유출 관통 흐름량 등 4가지 측도를 제시하였다. 그리고 수도권 지하철망의 승객 흐름에 과도한 편중 현상이 있는지를 파악하기 위하여 이들이 Power-law 분포를 따르는지 조사하였다.

서론에 이어 제 2장에서는 호선별 매개 중심성 측도를 산출하는 방법을 설명하였으며 제 3장에서는 계산과정과 결과를 분석하였다. 제 4장에는 승객 흐름 특성분석을 제 5장에 결론을 수록하였다.

2. 호선별 매개 중심성 측도

본 연구에서 사용한 호선별 매개 중심성은 본 연구자들이 선행 연구에서 정의한 역별 매개 중심성을 토대로 [7] 다음처럼 정의 할 수 있다.

서울 수도권 지하철 망의 역수를 N 개라 하면 모두 $N \cdot (N-1)$ 개의 최단 시간 경로가 존재한다. 이 경로들 중에 호선 K 를 지나는 경로들의 개수를 정규화(Normalizing)하여 나타낸 것을 호선 K 의 매개 중심성으로 정의한다. 즉, 호선 K 의 매개 중심성 $BC(K)$ 는 다음과 같이 표현된다.

$$BC(K) = \sum_{i \neq j} \left[\frac{1}{N \cdot (N-1)} \times \frac{\sigma(i \rightarrow j|K)}{\sigma(i \rightarrow j)} \right] \quad (1)$$

위 식에서 N 은 역의 수를 나타낸다. $\sigma(i \rightarrow j)$ 는 역 i 에서 j 로 가는 최단 경로의 개수를 나타내는데 대부분 1의 값을 가지며 두 역 사이에 최단 경로가 2개 이상 존재하면 2 이상의 값을 가질 수도 있다. 두 역 사이의 최단 시간 경로는 Dijkstra 알고리즘을 이용하여 구할 수 있다. $\sigma(i \rightarrow j|K)$ 는 역 i 에서 j 로 가는 최단 경로가 호선 K 를 통과하면 1 통과하지 않으면 0의 값을 갖는다. $N \cdot (N-1)$ 은 지하철 내에 있는 모든 역들 간 최단 시간 경로의 수를 나타내는데 매개 중심성 값을 정규화하기 위하여 사용된다.

매개 중심성은 식 (1)에서 볼 수 있듯이 정규화 상수 $N \cdot (N-1)$ 로 나뉘는데 이는 모든 (출발지, 도착지) 조합이 동일한 가중치($1/N \cdot (N-1)$)를 가짐을 의미한다. 예로 출발지가 강남역이고 도착지가 강남터미널역인 월간 승객 수는 출발지가 강남역이고 도착지가 인천역인 월간 승객 수의 350배에 이룸에도 이 두 조합은 같은 가중치를 갖는다. 매개 중심성 지표에 보다 현실적인 의미를 부여하기 위해서는 승객 흐름양이 많은 (출발지, 도착지) 조합에 보다 높은 가중치를 주어야 한다. 따라서 본 연구에서는 다음과 같이 호선별 가중 매개 중심성지표인 $WBC(K)$ 를 제안하였다.

$$WBC(K) = \sum_{i \neq j} \left[\frac{F(i \rightarrow j)}{\sum_{k \neq l} F(k \rightarrow l)} \times \frac{\sigma(i \rightarrow j|K)}{\sigma(i \rightarrow j)} \right] \quad (2)$$

식 (2)에서 $F(i \rightarrow j)$ 는 출발지가 역 i 이고 도착지가 역 j 인 월간 승객 흐름양을 나타내며 $\sum F(k \rightarrow l)$ 는 월간 승객 총 수를 나타낸다. 따라서 $WBC(K)$ 에서는 출발지와 도착지가 다른 경로는 월간 이용 승객 총 수에 따라 서로 다른 가중치를 갖게 된다. 이렇게 정의된 호선별 가중 매개 중심성의 경우 어떤 호선의 $WBC(K)$ 가 크다는 것은 실제로 그 호선을 통과하는 승객 수가 많다는 것을 의미하고 이는 그 호선에 사고가 발생하여 그 호선을 사용하지 못할 경우 전체 지하철 망에 미치는 영향도가 크다는 것을 나타낸다.

한편 K_n 을 호선 K 에 있는 역의 개수를 나타낸다고 하자. 각 호선별 매개 중심성은 각 호선에 존재하는 K_n 값이 클수록 커질 가능성이 높다. 따라서 이를 정규화하기 위하여 정규 가중 매개 중심성 $NWBC(K)$ 를 다음과 같이 정의 하였다.

$$NWBC(K) = WBC(K)/K_n \quad (3)$$

위와 같이 정의된 $NWBC(K)$ 는 각 호선별 역의 평균 매개 중심성을 나타낸다.

3. 호선별 매개 중심성 분석

2017년 현재 서울 수도권 지하철은 9개 노선의 서울 지하철과 12개 노선(공항, 경의중앙, 경춘, 분당, 수인, 신분당, 인천 1, 인천 2, 의정부, 자기부상, 예버, 경강)의 교외선으로 구성된다. 그런데 9호선과 신분당선, 의정부선 그리고 예버라인은 2013년 9월 이후로 승객 수송 관련 자료를 공개하지 않고 있기 때문에 본 연구에서는 실제적인 데이터가 존재하는 가장 최근 자료인 2013년 2월 자료를 사용하였다. 모두 17개 노선의 493개 역을 분석 대상으로 하였으며 데이터는 국가 교통 DB[6]와 서울 열린 데이터 광장[11]으로 부터 획득하였다.

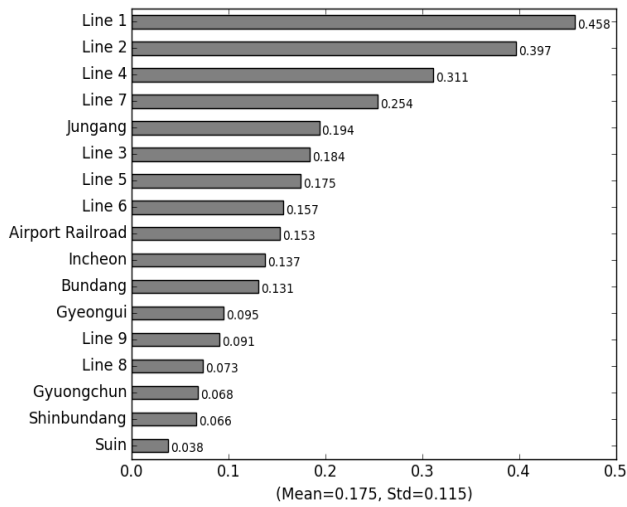
3.1 호선별 매개 중심성 계산

$BC(K)$, $WBC(K)$, $NWBC(K)$ 는 각각 위의 식 (1)~식 (3)을 통하여 구할 수 있다. 이중 호선별 가중 매개 중심성 $WBC(K)$ 의 계산과정을 살펴보면 다음과 같다.

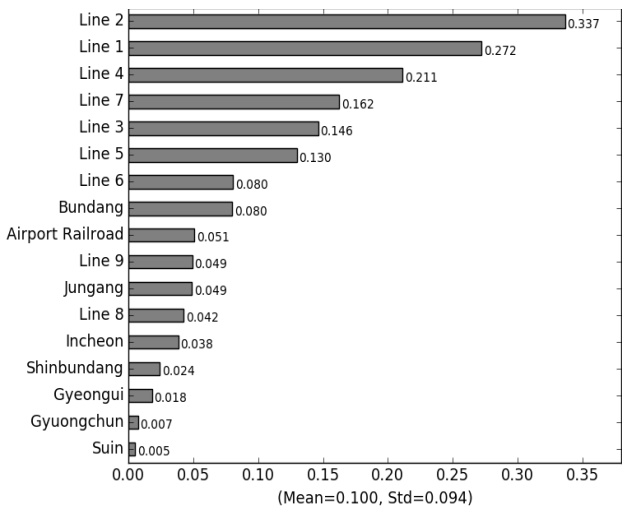
- 1) 임의의 두 역 i 와 j ($i \neq j$, $i, j = 1, 2, \dots, 493$) 사이의 최단 경로 $SPath(i \rightarrow j)$ 를 구한다. 본 연구에서는 Dijkstra Algorithm을 사용하였다. 모든 최단 경로의 수는 242,556 (= $493 \cdot 492$)개다. 최단 경로 계산은 시간을 기준으로 하였으며 Seoul Open Data Plaza에서 제공하는 인접한 두 역 사이의 시간 데이터와 모든 환승역들의 평균 환승시간인 2분을 사용하여 최단 경로를 계산하였다.
- 2) $SPath(i \rightarrow j)$ 가 호선 K 를 포함하면 $\sigma(i \rightarrow j|K)$ 를 1로 하고 포함하지 않으면 0으로 한다, $\sigma(i \rightarrow j)$ 는 역 i 에서 j 로 가는 최단 경로의 개수를 나타내는데 최단 경로 개수가 1개면 1 두 역 사이에 최단 경로가 m 개 존재하면 m 을 갖는다.
- 3) $F(i \rightarrow j)$ 를 역 i 에서 출발하여 역 j 까지 이동하는 월간 승객 수를 나타낸다고 하자. 이는 국가 교통 DB로부터 구한다.
- 4) 이제 식 (2)를 사용하여 $WBC(K)$ 를 계산한다.

3.2 결과

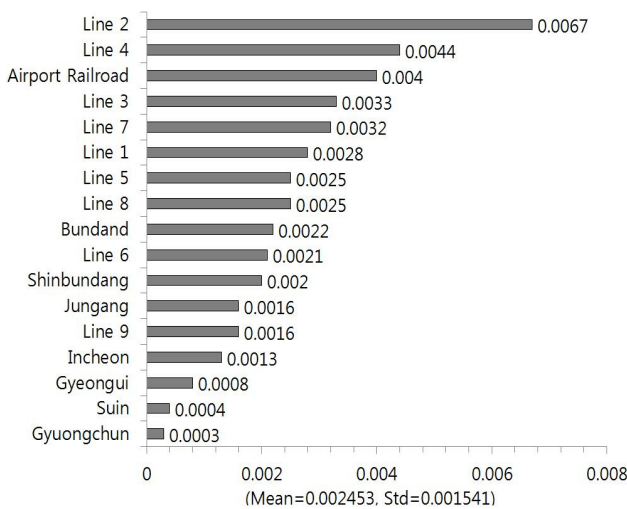
$BC(K)$, $WBC(K)$ 그리고 $NWBC(K)$ 를 계산하여 다음 <Figure 1>~<Figure 3>에 각각 나타냈다.



<Figure 1> BC(K) of Subway Lines



<Figure 2> WBC(K) of Subway Lines



<Figure 3> NWBC(K) of Subway Lines

3.3 분석

3.3.1 BC(K)와 WBC(K)

<Figure 1>과 <Figure 2>에서 볼 수 있듯이 BC(K)와 WBC(K)는 일치하지 않음을 볼 수 있는데 이 결과로부터 다음과 같은 분석을 할 수 있다.

- 첫째, WBC(K)가 BC(K)에 비해 항상 낮은 값을 갖는다. 실제로 BC(K)의 평균은 0.175이고 WBC(K)의 평균은 0.1로 모든 호선에 걸쳐 WBC(K)가 적은 값을 갖는다. 이는 BC(K) 값은 242,556(= 493 · 492)개의 모든 (출발역, 도착역)조합이 같은 가중치를 갖는데 비해서 WBC(K)는 (출발지, 도착지)조합에 따라 서로 다른 가중치인 월간 승객 흐름양 $F(i \rightarrow j)$ 를 갖는다. 일반적으로 WBC(K) 값이 BC(K) 값보다 적다는 사실은 대부분의 (출발지, 도착지)의 승객 수요가 평균보다 적어 WBC(K)의 대부분 가중치가 BC(K)의 가중치보다 적고 승객 수요가 매우 큰 값들이 소수 존재한다는 것을 의미한다. 즉, 중앙값(Median)이 평균값보다 적은 경우를 의미한다. 실제로 $F(i \rightarrow j)$ 의 중앙값은 97, 평균값은 802, 최댓값은 212,850, 왜도는 15.58(기준 0), 첨도는 474.29(기준 3)로 나타나 $F(i \rightarrow j)$ 가 어느 정도 편향(Skewed)된 분포를 따름을 확인할 수 있다.

- 둘째, 승객 흐름양을 고려하지 않고 지하철망의 구조만 고려할 경우 1호선의 매개 중심성이 가장 크게 나온 반면에 실제 242,556(= 493 · 492)개의 모든 (출발지, 도착지)조합의 월간 승객 흐름양을 고려하면 2호선의 매개 중심성이 가장 높다. 이는 1호선을 경유하는 승객 수보다 2호선을 경유하는 승객 수가 더 많음에 기인한 결과로 2호선이 순환선의 역할을 제대로 하고 있음을 나타낸다. 중앙선의 경우도 월간 승객 흐름양을 고려하지 않으면 7호선에 이어 5번째로 매개 중심성이 높은 반면에 월간 승객 흐름양을 고려하면 11번째로 순위가 떨어진다. 중앙선의 경우는 환승역이 22개로(<Table 1> 참조) 비교적 많아 BC(K)는 높은 값을 가질 수 있지만 중앙선을 경유하는 승객 수가 다른 호선에 비해 많지 않아 WBC(K) 값은 낮아진다.

앞의 2장 매개 중심성 측도에서 언급했듯이 매개 중심성 지표가 더 현실적이고 의미 있는 값을 갖기 위해서는 승객 흐름양이 많은 (출발역, 도착역) 조합에 보다 높은 가중치를 부여한 WBC(K)를 사용하는 것이 타당하다. <Figure 1>과 <Figure 2>에서 볼 수 있듯이 기존에 사용해온 BC(K)와 본 연구에서 제시한 WBC(K)는 일치하지 않음을 볼 수 있는데 이는 본 연구에서 사용한 WBC(K) 사용 정당성을 확인 시켜준다,

3.3.2 $WBC(K)$ 와 $NWBC(K)$

호선별 가중 매개 중심성을 각 호선에 있는 역의 개수로 나눈 정규 가중 매개 중심성 $NWBC(K)$ 는 <Figure 2>와 <Figure 3>에서 볼 수 있듯이 $WBC(K)$ 와 차이가 있다. 2호선은 $NWBC(K)$ 도 $WBC(K)$ 와 마찬가지로 가장 높게 나타났다. 2호선 역들의 평균 매개 중심성은 0.0067로 다른 호선들의 역들에 비해 월등히 높음을 알 수 있는데 이는 순환선의 역할을 맡고 있음에 기인한다. 1호선의 경우 $WBC(K)$ 는 두 번째로 높았지만 $NWBC(K)$ 는 여섯 번째로 순위가 많이 내려감을 볼 수 있다. 이는 1호선의 역의 개수가 98개로 가장 많기 때문으로 볼 수 있다. 다시 말하면 1호선의 가중 매개 중심성 $WBC(K)$ 가 2번째로 높은 이유는 지하철 네트워크에서 1호선이 갖는 중심적인 위치 때문이 아니라 역의 개수가 많아서 생긴 결과인데 이를 역의 개수로 정규화하면 $NWBC(K)$ 는 낮아 질 수밖에 없다. 그리고 4호선과 공항철도, 3호선 그리고 7호선 순으로 $NWBC(K)$ 가 높게 나타났다. 만약에 열차 차량수와 배차간격 등이 모든 호선에 걸쳐 동일하다고 가정하면 정규 가중 매개 중심성 $NWBC(K)$ 는 각 호선별 역의 평균 포화도(Congestion)를 나타낼 수 있다. 반대로 시간대별 승객 흐름량 데이터를 토대로 시간대별로 $NWBC(K)$ 를 계산할 수 있으면 이를 토대로 각 호선별 시간대별 배차 간격 등을 결정하는 기본 자료로 이용할 수 있다.

3.3.3 매개 중심성에 영향을 미치는 요인 분석

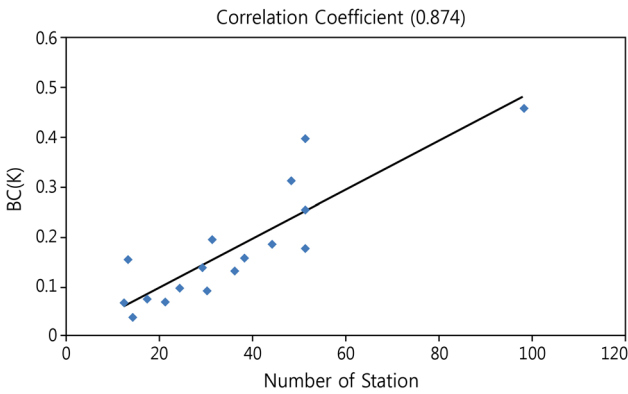
각 호선별 매개 중심성 지표에 영향을 미치는 요인을 살펴보기 위하여 호선별 역의 개수, 환승역의 수, 그리고 호선별 월간 승차 인원수, 하차 인원수들 간의 상관관계를 조사하였다. 다음 <Table 1>에 이들 자료를 요약 정리하였다. 환승역 수는 환승하는 노선이 여러 개 있을 경우 따로 합산하였다.

(1) 역수

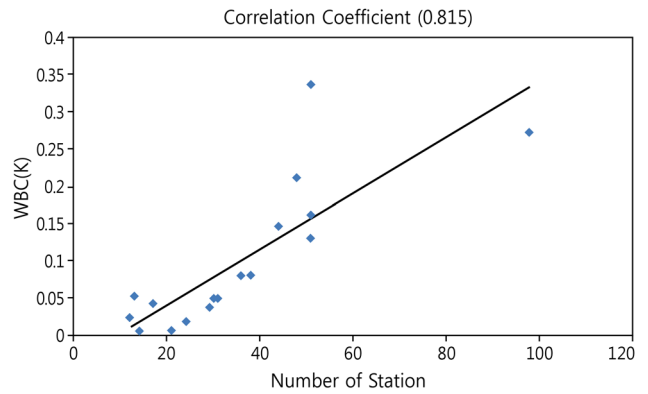
<Figure 4>와 <Figure 5>에서 보듯이 역수와 $BC(K)$, $WBC(K)$ 의 상관 계수는 각각 0.874, 0.815로 나타났다. 먼저 이는 역의 개수가 많다고 해서 $BC(K)$, $WBC(K)$ 값이 반드시 높다는 것을 보여주지는 않는다. 그렇지만 <Table 1>과 <Figure 4>에서 보듯이 $BC(K)$ 는 어느 정도 역의 수에 비례하는 것을 볼 수 있다. 이는 특정 노선의 역의 개수가 많을수록 특정 노선을 경유할 가능성이 커지기 때문이다. 반면에 $WBC(K)$ 는 역의 수가 51개인 2호선이 0.337로 역의 수가 98개인 1호선의 0.272에 비해 높다. 또한 역의 수가 51개인 5호선은 0.13으로 역의 수가 48개인 4호선의 0.211에 비해 상당히 낮다. 이는 $WBC(K)$ 의 계산과정에 승객 흐름량이 고려되어서 특정 노선에 역의 수가 많더라도 실제 승객 흐름량이 적으면 $WBC(K)$ 값은 낮아지기 때문이다.

<Table 1> Number of Station, Transfer Station and Monthly Passenger

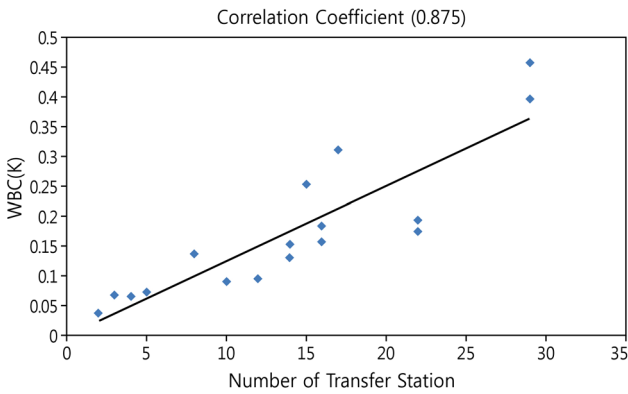
Lines	Number of Station	Number of Transfer Station	Number of Passenger (getting on)	Number of Passenger (getting off)	$BC(K)$	$WBC(K)$
1	98	29	35,597,591	35,395,053	0.458	0.272
2	51	29	39,383,435	39,849,153	0.397	0.337
3	44	16	16,951,069	17,143,072	0.184	0.146
4	48	17	23,700,298	23,711,558	0.311	0.211
5	51	22	15,939,347	15,790,443	0.175	0.130
6	38	16	8,264,266	8,127,304	0.157	0.080
7	51	15	19,109,615	19,030,012	0.254	0.162
8	17	5	4,245,330	4,173,503	0.073	0.042
9	30	10	6,066,783	5,999,204	0.091	0.049
Bundang	36	14	8,582,259	8,693,292	0.131	0.080
Airport	13	14	3,244,821	3,270,193	0.153	0.051
Jungang	31	22	3,460,679	3,421,951	0.194	0.049
Incheon	29	8	5,219,227	5,199,299	0.137	0.038
New Bundang	12	4	1,697,258	1,714,368	0.066	0.024
Kyungeui	24	12	1,669,608	1,633,937	0.095	0.018
Kyungchoon	21	3	913,633	896,371	0.068	0.007
Suin	14	2	508,602	505,111	0.038	0.005



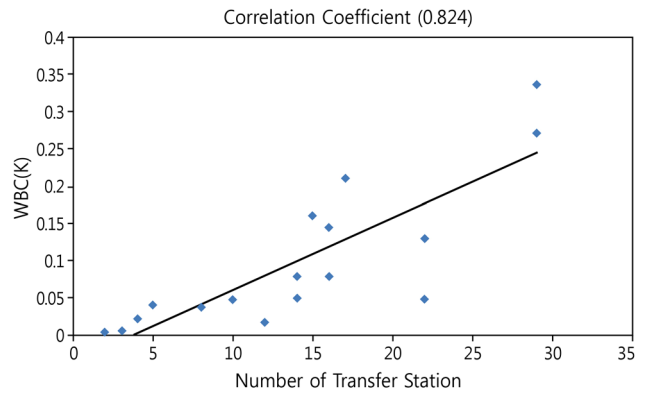
<Figure 4> Correlation between $BC(K)$ and Number of Stations



<Figure 5> Correlation between $WBC(K)$ and Number of Stations



<Figure 6> Correlation between $BC(K)$ and Number of Transfer Stations



<Figure 7> Correlation between $WBC(K)$ and Number of Transfer Stations

(2) 환승역 수

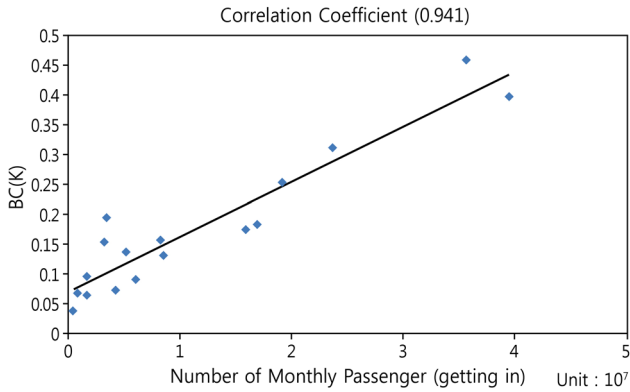
<Figure 6>과 <Figure 7>에서 보듯이 환승역 수와 $BC(K)$, $WBC(K)$ 의 상관 계수는 각각 0.875, 0.824로 나타났다. 이는 먼저 앞의 역 수와 마찬가지로 환승역의 개수가 많다고 해서 $BC(K)$, $WBC(K)$ 값이 반드시 높다는 것을 보여주지는 않는다. 그렇지만 <Table 1>과 <Figure 6>에서 보듯이 $BC(K)$ 는 어느 정도 환승역의 수에 비례하는 것을 볼 수 있다. 이는 특정 노선에 환승역의 개수가 많을수록 특정 노선을 이용할 가능성이 커지기 때문이다. 반면에 $WBC(K)$ 는 환승역의 수가 29개로 1, 2호선이 같지만 2호선이 0.337로 1호선의 0.272에 비해 높다. 또한 환승역의 수가 17개인 4호선은 0.211로 환승역의 수가 22개인 중앙선의 0.049에 비해 상당히 높다. 이는 앞의 역의 수와 마찬가지로 $WBC(K)$ 의 계산과정에 승객 흐름양이 고려되어서 특정 노선에 환승역의 수가 많더라도 실제 승객 흐름양이 적으면 $WBC(K)$ 값이 낮아지기 때문이다.

(3) 호선별 월간 승하차 인원

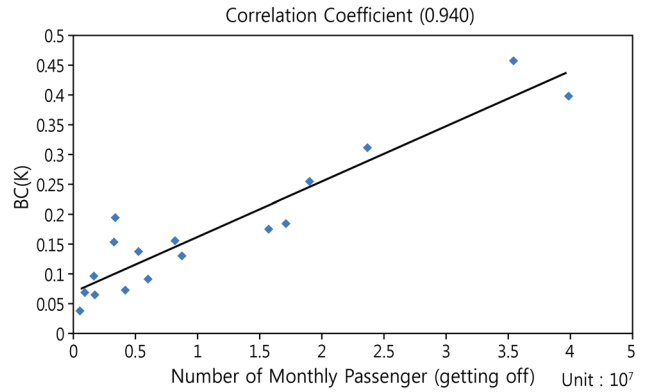
다음 <Figure 8>과 <Figure 9>에서 볼 수 있듯이 호선

별 월간 승하차 인원과 $BC(K)$ 의 상관계수는 각각 0.941과 0.940으로 높은 상관 계수를 갖는다. 일부 예외는 있지만 대부분의 승객들은 승차 역에서 일정시간 뒤에 하차하는 형태를 보이기 때문에 <Table 1>에서 보듯이 월간 승차 인원과 하차 인원은 거의 일치한다. 따라서 월간 승차인원과 하차인원의 매개 중심성 측도와의 상관계수는 거의 일치하는 결과를 보인다.

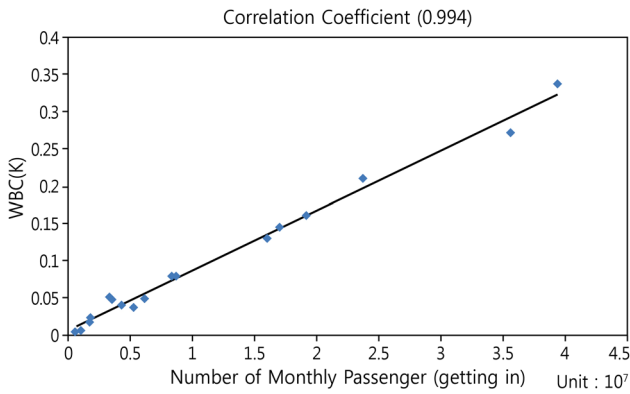
$BC(K)$ 의 계산에는 승객 흐름양이 전혀 고려되지 않는다는 점을 감안하면 이는 매우 높은 값이다. 즉, 대체로 매개 중심성이 더 높은 중요한 호선을 이용하는 승객들이 더 많고 반대로 매개 중심성이 낮은 호선을 이용하는 승객들의 양은 적다라는 것을 뜻한다. 이는 서울 수도권 지하철 망이 인구의 지역별 분포 대비 효율적으로 설계되었다는 것을 나타낸다. 만약에 매개 중심성이 낮은 호선에 승하차 인원이 많다면 이는 많은 승객들이 원하는 목적지를 가는데 여러 번의 환승을 해야만 하는 비효율성이 발생하고 반대로 매개 중심성이 높은 노선에 승하차 인원이 적다면 이는 필요 없는 자원이 과하게 사용되고 있음을 뜻하기 때문이다.



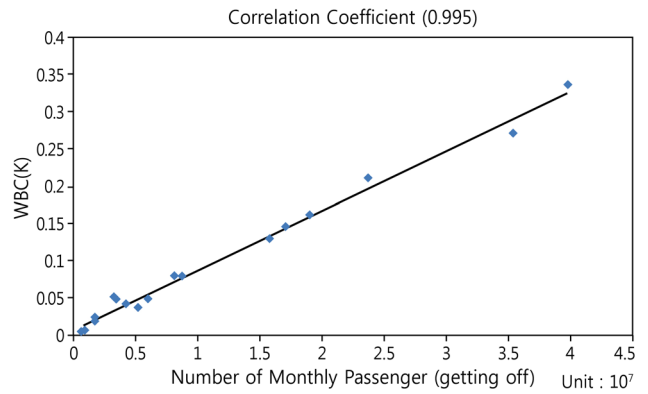
<Figure 8> Correlation between $BC(K)$ and Number of Monthly Passenger(getting on)



<Figure 9> Correlation between $BC(K)$ and Number of Monthly Passenger(getting off)



<Figure 10> Correlation between $WBC(K)$ and Number of Monthly Passenger(getting on)



<Figure 11> Correlation between $WBC(K)$ and Number of Monthly Passenger(getting off)

한편 호선별 월간 승하차 인원과 $WBC(K)$ 의 상관계수는 <Figure 10>과 <Figure 11>에서 보듯이 각각 0.994, 0.995로 거의 일치하는 결과를 보여준다. 물론 $WBC(K)$ 의 계산에는 월간 승하차 인원이 고려되기 때문에 $BC(K)$ 보다는 월간 승하차 인원과 상관 계수가 높은 값을 갖는 것이 당연하지만 상관 계수가 1과 거의 가깝다는 것은 매우 특이하다. 실제 해당 노선을 이용하는 승객 수는 해당 노선에서 승차나 하차하는 승객 수 뿐 만 아니라 다른 노선들에서 승차하여 해당 노선을 이용하는 승객 수들이 함께 고려되기 때문이다. 즉,

$$\begin{aligned} & \text{노선 } K \text{를 이용하는 월간 승객 수} \\ &= \text{노선 } K \text{의 월간 승차 승객 수} \\ &+ \sum_{L \neq K} (\text{노선 } L \text{의 역에서 승차하여 노선 } K \text{로 환승한} \\ & \text{월간 승객 수}) \end{aligned}$$

위 식에서 노선 K 의 월간 승차 승객 수는 노선 K 에 있는 역에서만 승차한 월간 승객 수를 나타낸다. 가중 매개 중심성인 $WBC(K)$ 의 계산에는 가중치로 노선 K 를 이

용하는 월간 승객 수가 사용되기 때문에 $WBC(K)$ 와 노선 K 를 이용하는 월간 승객 수는 1에 가까운 상관관계를 갖는 것이 당연하다. 그렇지만 <Figure 10>에서 보듯이 $WBC(K)$ 는 노선 K 의 월간 승차 승객 수와도 거의 1에 가까운 상관 계수를 갖는다. 이는 $\sum_{L \neq K}$ (노선 L 의 역에서 승차하여 노선 K 로 환승한 월간 승객 수)가 0이 아니기 때문에 노선 K 의 월간 승차 승객 수와 $\sum_{L \neq K}$ (노선 L 의 역에서 승차하여 노선 K 로 환승한 월간 승객 수)는 서로 정비례 관계에 있을 때 가능하다. 즉, 서울 수도권 지하철의 경우 월간 승차 승객 수가 많은 노선은 다른 노선의 역에서 승차하여 해당 노선으로 환승한 승객 수도 많고 월간 승차 승객 수가 적은 노선은 다른 노선의 역에서 승차하여 해당 노선으로 환승한 승객 수도 적음을 나타낸다.

서울 수도권 지하철의 경우에는 호선별 가중 매개 중심성 $WBC(K)$ 가 해당 노선의 월간 승차나 하차 승객 수에 의해 결정되는 것을 확인하였다. 그 이유는 앞에서 밝혔듯이 노선 K 의 월간 승차 승객 수와 다른 노선들의 역들에서 승차하여 노선 K 로 환승한 승객 수가 서로 정비

례하기 때문으로 보인다. 이러한 성질이 서울 수도권 지하철에만 해당되는 특이한 성질인지 아니면 외국 지하철의 경우에도 호선별 가중 매개 중심성 $WBC(K)$ 가 해당 노선의 월간 승차나 하차 승객 수에 의해 결정 되는지는 분명 하지 않다. 관련 자료들을 획득할 수 있으면 이는 좋은 추후 연구가 될 수 있을 것이다.

호선별 월간 총 이용 금액은 호선별 월간 총 승차 인원과 각 승객이 지불하는 이용금액으로부터 계산 가능하다. 만약 호선별로 각 승객이 지불하는 이용 금액의 분포가 동일하다고 가정하면 호선별 월간 총 이용 금액은 호선별 월간 총 승차 인원수에 의해 결정 된다. 가중 매개 중심성의 월간 승차 승객 수와 상관계수가 거의 1이라는 점을 고려하면 특정호선의 가중 매개 중심성 $WBC(K)$ 가 높으면 총 이용 금액도 높고 낮으면 총 이용 금액도 낮게 된다. 반대로 말하면 호선별 지하철 총 이용금액에 의해 호선별 가중 매개 중심성이 결정 된다고 이야기 할 수 있다.

4. 승객흐름 특성분석

제 3.3절의 $BC(K)$ 와 $WBC(K)$ 의 분석에서 $WBC(K)$ 의 계산을 위해 사용된 $F(i \rightarrow j)$ 가 어느 정도 편향된 분포를 갖는다는 것을 간접적으로 추론 할 수 있었다. 따라서 본 장에서는 한국 지하철의 승객 흐름 분포의 편향성이나 쏠림 현상이 서론에서 언급한 베이징이나 샌프란시스코, 그리고 보스톤의 경우처럼 심각한지를 조사하였다. 이를 위하여 먼저 승객 흐름을 나타내주는 4개의 지표를 제시하고 이들이 Power-law 분포를 따르는지를 조사하였다.

일반적으로 사용되는 지하철을 이용하는 승객 흐름을 나타내주는 네 개의 측도는 다음과 같다. 먼저 인접한 두 역 i 와 j (여로 i 가 삼성역이면 j 는 종합운동장역이거나 선릉역이고 i 가 건대입구역이면 j 는 구의역, 성수역, 어린이대공원역 혹은 뚝섬유원지역이다) 사이의 월간 승객 흐름 양과 역 i 의 혼잡도다. 혼잡도는 한 달간 역 i 를 경유, 역 i 에서 출발 혹은 역 i 에 도착하는 승객 수의 합으로 정의하였다. 그리고 역 i 의 유입 관통 흐름양(Throughflow)과 유출 관통 흐름양이다. 유입 관통 흐름양은 외부에서 역 i 로 들어오는 승객 수와 인접한 역에서 역 i 로 오는 승객 수의 합으로 유출 관통 흐름양은 역 i 에서 외부로 나가는 승객 수와 역 i 에서 인접한 역으로 나가는 승객 수의 합으로 정의된다. 이들 네 가지 척도는 외국 지하철들의 승객 흐름을 분석하기 위하여[8, 13] 이미 사용된 척도고 혼잡도는 Cheng 등이[1] 제시한 'Commuter Flow Centrality'와 같은 개념이다. 베이징[9]의 경우 인접한 지하철 역사이의 월간 승객흐름양, 유입, 유출 관통흐름양들이 Power-

law 분포를 따르는 것으로 나타났다. 이는 승객 흐름양의 관측치가 평균보다 매우 큰 값들이 다수 개 존재하며 중앙값이 평균값보다 아주 작은 편향된 분포를 따른다는 것을 나타내고 지하철 시스템의 승객 흐름에 매우 심한 불균형성이 있다는 것을 의미한다. 한편 Wang 등[12]은 샌프란시스코와 보스톤 URT의 일정 구간(Segment) 승객흐름의 분포를 조사하였다. 승객 흐름은 각 구간을 경유하는 최단 경로의 개수로 부터 구해지는데 양 도시 모두 승객 흐름 분포는 Power-law를 따르는 것으로 나타났다. 이는 양 도시 모두 승객 흐름이 이질적(Heterogeneous)으로 분포되어 있으며 매우 많은 승객 흐름을 갖는 구간들이 존재 한다는 것을 의미한다. 본 장에서는 한국의 경우 이들 네가지 측도가 Power-law 분포를 따르는지를 검증하였다. 앞의 호선별 매개 중심성 분석과 마찬가지로 9호선과 신분당선, 의정부선 그리고 에버라인은 2013년 9월 이후로 승객 수송 관련 자료를 공개하지 않고 있기 때문에 본 연구에서는 실제적인 데이터가 존재하는 가장 최근 자료인 2013년 2월 자료를 사용하였다. 모두 17개 노선의 493개 역을 분석 대상으로 하였으며 데이터는 국가 교통 DB[6]와 서울 열린 데이터 광장[11]으로부터 획득하였다.

Power-law 분포를 따르는지 검증하는 방법은 관측치의 PDF(Probability Density Function)나 CDF(Cumulative Density Function) 그리고 CCDF(Complimentary Cumulative Density Function)가 Log-Log Scale에서 기울기 1에서 3사이의 직선 식을 따르는지를 Kolmogorov-Smirnov 테스트를 통해 조사하는 것이다. 그런데 일반적으로 PDF에 비해 CDF나 CCDF가 Power-law 성질을 보다 명확하게 나타내준다고 알려져 있다. 본 연구에서는 Power-law 검증을 위해 CCDF를 사용하였다.

Kolmogorov-Smirnov 테스트 결과 p-value들이 모두 0.01 이하로 $N(i \rightarrow j)$ 와 INF_i , $OUTF_i$ 그리고 각 역의 혼잡도 C_i 가 모두 Power-law 분포를 따르지 않는 것으로 나타났다. 이는 서울 수도권 지하철은 승객 흐름이 평균치에 비해서 매우 큰 곳이 거의 존재하지 않는 비교적 균일한 승객 흐름 형태를 나타내는 것으로 볼 수 있다. 베이징의 경우와는 다르게 서울 지하철의 승객 흐름과 혼잡도가 Power-law 분포를 따르지 않는 이유는 베이징과 서울 지하철의 서로 다른 다음의 환경에 기인한 것으로 보인다.

첫째, 서울의 면적은 605.28km^2 이고 베이징은 $16,808\text{km}^2$ 로 서울의 27배에 이른다. 반면에 지하철역 수는 서울 수도권의 경우 463개고 베이징 수도권의 경우 318개로(본 연구의 데이터 획득시점인 2013년 9월 시점) 서울 수도권이 약 1.5배 더 많다. 따라서 서울이 단위 면적당 지하철역 수도 많고 역 간 거리도 훨씬 짧을 수밖에 없다. 이는 지하철 노선들이 베이징에 비해 서울권역에 보다 잘 골고루 분산되어 있기 때문에 특정 두 역사이의

흐름양이나 특정역의 혼잡도 그리고 관통 흐름양이 평균치에 비해 매우 큰 값이 존재할 가능성이 베이징에 비해 낮다.

둘째, 베이징의 경우 주거지역, 상업지역, 그리고 문화구역이나 관광구역이 매우 잘 구분되어서 특정 지역들로 지하철승객이 모이는 불균형 현상이 발생 하게 된다. 반면에 서울의 경우에는 이 지역 구분이 비교적 명확하지 않다. 예로 서울 강남이나 여의도 혹은 신촌의 경우 주거지역뿐 만 아니라 상업지역, 문화구역 그리고 관광 구역이 서로 혼재되어 있는 것을 알 수 있다. 그래서 서울의 경우에는 베이징의 경우처럼 특정 지역들로 지하철승객이 모이는 불균형 현상이 발생할 가능성이 비교적 낮다.

5. 결 론

본 연구에서는 서울 수도권 지하철 망의 호선별 분석을 네트워크의 매개 중심성 지표를 사용하여 시행하였다. 매개 중심성 지표로는 전통적인 매개 중심성 $BC(K)$ 와 가중 매개 중심성 $WBC(K)$, 그리고 정규 가중 매개 중심성 $NWBC(K)$ 를 사용하였다.

전통적인 매개 중심성은 1호선이 가장 높게 나온 반면에 승객들의 월간 흐름양을 고려한 가중 매개 중심성은 2호선이 가장 높게 나오는 등 $BC(K)$ 와 $WBC(K)$ 는 다소 차이가 발생하였다. $BC(K)$ 와 $WBC(K)$ 의 비교를 통하여 호선별 가중 매개 중심성 지표 사용의 정당성을 확인 할 수 있었다. 정규 가중 매개 중심성 측도 $NWBC(K)$ 는 일정 가정 하에서 각 호선별로 존재하는 역들의 평균 포화도를 나타낸다. 따라서 이 자료를 토대로 가공 작업을 거치면 호선별 시간대별 최적 배차 간격 등을 결정하는데 이용 할 수 있다.

한편 역의 수와 환승역의 수는 매개 중심성 지표에 어느 정도 영향을 미치지만 역의 수나 환승역의 수가 많다고 해서 매개 중심성 지표가 반드시 높지는 않았다. 반대로 호선별 월간 승차나 하차 승객 수는 매개 중심성 지표와 매우 밀접한 관계가 있는 것으로 나타났다. 특히 호선별 월간 총 승차나 총 하차 인원수는 호선별 가중 매개 중심성과 거의 1의 상관관계를 보이고 있다. 즉 서울 수도권 지하철의 호선별 가중 매개 중심성은 호선별 월간 총 승차나 총 하차 인원수에 의해 결정 된다. 이러한 성질이 서울 수도권 지하철 망에만 나타나는 특이한 현상인지 아니면 외국의 지하철 망에도 같은 현상이 발생하는지는 추후 연구과제로 남긴다. 또한 호선별 월간 총 승차나 총 하차 인원수와 매개 중심성과의 상관관계 분석을 통해 서울 수도권 지하철망은 인구의 지역별 분포 대비 효율적으로 설계되었음을 확인 할 수 있었다.

본 연구에서는 한국 지하철의 승객 흐름 특성을 파악하기 위하여 4가지 승객 흐름 측도를 제안하였다. 이 4가지 측도는 모두 Power-law 분포를 따르지 않는 것으로 나타났다. $BC(K)$ 와 $WBC(K)$ 의 분석에서 $F(i \rightarrow j)$ 가 어느 정도 편향성이 있다는 것을 간접적으로 추론할 수 있었지만 우리나라 지하철 망의 승객 흐름이 본 연구에서 살펴본 외국의 경우처럼 특정 역에 과도하게 편중 되어 있다는 사실을 통계적으로 확인 할 수 없었다.

Acknowledgement

This research was supported by Seoul National University of Science and Technology Research funds.

References

- [1] Cheng, Y.Y., Lee, R.K.W., Lim, E.P., and Zhu, F., Measuring centralities for transportation networks beyond structures, *Applications of Social Media and Social Network Analysis*, 2015, pp. 23-39.
- [2] Choi, S.H., A Study on the Development of a Simulator for Social Network in Organizations Using ARENA, *Journal of Society of Korea Industrial and Systems Engineering*, 2012, Vol. 35, No. 3, pp. 62-69.
- [3] Derrible, S., Network Centrality of Metro Systems, *PLoS One*, 2012, Vol. 7, No. 7, pp. 1-10.
- [4] Haznagy, A., Fi, I., Londo, A., and N'emeth, T., Complex Network analysis of Public Transportation Networks : A Comprehensive Study, *Models and Technology for Intelligent Transport Systems*, 2015, Budapest, Hungary, pp. 371-378.
- [5] Jo, H.W., Lee, S.H., and Shin, K.W., An Empirical Study on the Relationship between Subway Trips and Characteristics of Subway Catchment Area, *Journal of the Korea Academia-Industrial Cooperation Society*, 2010, Vol. 12, No. 12, pp. 5191-5198.
- [6] Korea Transportation Database, (Accessed 30 January 2018), <http://www.ktdb.go.kr>.
- [7] Lee, J.W. and Lee, K.W., Analysis of Seoul Metropolitan Subway Network Characteristics using Network Centrality Analysis, *Journal of the Korean Society for Railway*, 2017, Vol. 20, No. 3, pp. 413-422.
- [8] Leng, B., Zhao, X., and Zhang, X., Evaluating the Evolution of Subway Networks : Evidence from Beijing Subway Network, *A letters Journal Exploring The Frontier of Physics*, 2014, Vol. 105, e58004, pp. 1-6.

- [9] Monterola, C.P., Ramli, M.A., Khoon, G.A., and Guang, T.H., A Method to Ascertain Rapid Transit System's Throughput Distribution Using Network Analysis, *Procedia Computer Science*, 2014, Vol. 29, pp. 1621-1630.
- [10] Mouronte, M.L., Topological Analysis of the Subway Network of Madrid, *International Multi-Conference on Computing in the Global Information Technology*, 2014, Seville, Spain, pp. 9-13.
- [11] Seoul Open Data Plaza, (Accessed 30 January 2018), <http://data.seoul.go.kr>.
- [12] Stoilova, S. and Stoev, V., An Application of the Graph Theory Which Examines the Metro Network, *Transport Problems*, 2015, Vol. 10, No. 2, pp. 35-48.
- [13] Wang, J., Li, Y., Liu, J., He, K., and Wang, P., Vulnerability Analysis and Passenger Source Prediction in Urban Rail Transit System, *Plos One*, 2013, Vol. 8, No. 11, pp. 1-8.
- [14] Xu, Q., Mao, B.A., and Bai, Y., Network Structure of Subway Passenger Flow, *Journal of Statistical Mechanics : Theory & Experiment*, 2016, pp. 1-18.
- [15] Zang, S.Y. and Lee, K.W., Characteristics and Efficiency Analysis of Evolutionary Seoul Metropolitan Subway Network, *Journal of the Korean Society for Railway*, 2016, Vol. 19, No. 3, pp. 388-396.

ORCIDKang Won Lee | <http://orcid.org/0000-0002-5644-4490>Jung Won Lee | <http://orcid.org/0000-0003-1215-5838>