

# 네트워크 중심성 지표를 이용한 서울 수도권 지하철망 특성 분석

## Analysis of Seoul Metropolitan Subway Network Characteristics Using Network Centrality Measures

이정원 · 이강원\*

Jeong Won Lee · Kang Won Lee

**Abstract** In this study we investigate the importance of the subway station using network centrality measures. For centrality measures, we have used betweenness centrality, closeness centrality, and degree centrality. A new measure called weighted betweenness centrality is proposed, that combines both traditional betweenness centrality and passenger flow between stations. Through correlation analysis and power-law analysis of passenger flow on the Seoul metropolitan subway network, we have shown that weighted betweenness centrality is a meaningful and practical measure. We have also shown that passenger flow between any two stations follows a highly skewed power-law distribution.

**Keywords** : Betweenness Centrality, Closeness Centrality, Degree Centrality, Power-Law, Correlation Coefficient

**초 록** 본 연구에서는 네트워크 중심성 지표를 사용하여 지하철 네트워크의 개별 노드의 중요성을 분석하고 이로부터 한국 지하철 네트워크의 특성을 분석하였다. 중심성 측도로 매개, 근접 그리고 차수 중심성을 사용하였다. 본 연구에서는 기존에 제안된 매개 중심성 지표와 승객들의 실제 흐름양을 함께 고려한 가중 매개 중심성 지표를 새롭게 제안하였다. 그리고 본 연구에서 제안한 여러 중심성 지표들 사이의 상관관계를 조사함으로써 서울 수도권 지하철과 승객 흐름의 구조적 특성 등을 조사하였다. 아울러 승객들 흐름의 편중 현상을 조사하기 위하여 멱분포(Power-law) 분석을 수행하여 결과 분석의 신빙성을 더하였다.

**주요어** : 매개 중심성, 근접 중심성, 차수 중심성, 멱분포(Power-law), 상관계수

## 1. 서 론

서울은 세계에서 가장 인구 밀도가 높은 도시 중의 하나로 매우 선진화되고 정교한 대중 교통 시스템을 갖고 있다. 서울의 가장 대표적인 대중 교통 시스템은 지하철인데 1974년에 1호선 개통 이후 서울 수도권 지하철은 교통 정체 해소, 위성 도시 개발, 도심지 인구 분산 및 도시개발지원등의 목적으로 지난 30여 년 동안 많은 발전과 변화가 있었다. 서울 수도권 지하철의 수송 분담률은 1996년에 버스를 추월한 이후 2014년에는 39%에 이를 정도로 급속하게 팽창되어왔다. 2017년 현재 서울 수도권 지하철은 9개 노선의 서울 지하철과 12개 노선(공항, 경의중앙, 경춘, 분당, 수인, 신분당, 인천1, 인천2, 의정부, 자기부상, 에버, 경강)의 교외선으로 구성된다. 9개 노선의 서울 지하철은 우수한 환승 시스템을 갖추고 있으며 12개의 교외선들과도 잘 연결되어 있어서 모스크바, 동경, 뉴욕 지하철들에 버금가는 세계에서 가장 큰 지하철 중의 하나가 되었다.

많은 시스템들이 링크에 의해 연결되는 노드들로 구성되는 네트워크(Network)로 표현될 수 있다. 개별 라우터들 혹은 AS (Autonomous System)들로 구성된 Internet, WWW(World Wide Web), SNS(Social Network Service), 그리고 생물학적 (Biological) 네트워크 등이 좋은 예다. 교통 시스템도 그 중 하나로 항공, 고속도로, 그리고 지하철 등 다양한 운송 시스템들이 네트워크로 표현될 수 있다.

세계 주요 도시의 지하철의 효율성과 구조적 특성을 파악하기 위하여 지하철을 네트워크로 모형화하고 분석한 여러 연구가 진행되었다. Derrible 등은[1] 캐나다 토론토 지하철 네트워크를 커버리지(Coverage), 직접성(Directness) 그리고 연결성 (Connectivity) 관점에서 분석하고 세계 다른 나라의 그것들과 비교하였다. Latora 등은[2] 지하철 네트워크의 효율성을 분석하

기 위하여 전체 효율성(Global Efficiency)과 지역 효율성(Local Efficiency) 두 가지 측도를 제시한 후 이를 토대로 미국 보스턴 지하철 네트워크를 분석하였다. Mouronte 등은[3] 세 개의 서로 다른 네트워크 공간(L-space, P-space, R-space)을 사용하여 스페인 마드리드 지하철을 구조적으로 분석하였다. Stoilova 등은[4] 파리, 모스크바, 런던 등 세계 주요 지하철 네트워크 특성을 그래프 이론을 사용하여 비교분석하였다. Leng 등은[5] 중국 베이징 지하철의 진화과정을 확장 모드(Expand Mode)와 강화 모드(Intensify Mode) 두 개의 관점에서 살펴보았으며 Zang 등은[6] 과거 30여 년 서울 지하철의 발전과정을 네트워크의 구조적 측면과 효율성 측면에서 분석하였다.

위에서 살펴본 연구들은 주로 네트워크의 전체적인 관점에서 구조적 특성과 효율성을 조사하였다. 반면에 네트워크의 개별 노드들의 중요성을 다루는 네트워크 중심성(Network Centrality) 관련 연구도 활발하게 진행되었다. 네트워크 중심성은 크게 매개 노드 지표인 매개 중심성(Betweenness Centrality), 허브 노드 지표인 근접 중심성(Closeness Centrality), 그리고 차수 중심성(Degree Centrality)으로 세분되는데 이에 대한 구체적인 논의는 2장에서 하였다. Monterola 등은[7] 싱가포르 지하철 네트워크의 매개 중심성을 구하여 승객 효율(Passenger Throughput)과의 상관관계를 분석하였다. H'aznaghy 등은[8] 헝가리 지하철 네트워크의 매개 중심성과 근접 중심성을 구한 후 이를 토대로 지하철의 구조적 특성 등을 조사하였다. Cheng 등은[9] 지하철 네트워크의 중요 노드를 결정하고 네트워크를 개선하거나 고장에 대처하는 가장 효율적 대안을 조사하기 위하여 네트워크 중심성 지표를 사용하였다. 아울러 새로운 중심성 지표인 승객 흐름 중심성(Commuter Flow Centrality)을 제안하였다.

한편, 한국의 경우 1974년에 1호선 개통 이후 세계에서 대표적인 지하철로 성장하여 왔는데 네트워크를 이용하여 서울 지하철의 구조적 특성과 효율성을 분석한 연구는 드물다. 더구나 네트워크 중심성 지표를 사용하여 지하철 네트워크의 개별 노드의 중요성을 분석하고 이로부터 한국 지하철 네트워크의 특성을 분석한 논문은 전무하다. 본 논문에서는 9개의 서울 지하철 노선과 8개의 교외선을 포함한 서울 수도권 지하철을 네트워크 중심성 관점에서 분석하였는데 중심성 지표로 매개, 근접 그리고 차수 중심성을 사용하였다. 본 연구에서는 기존에 제안된 매개 중심성 지표와 승객들의 실제 흐름량을 함께 고려한 가중(Weighted) 매개 중심성 지표를 새롭게 제안하였다. 그리고 본 연구에서 제안한 여러 중심성 지표들 사이의 상관관계(Correlation Coefficient)를 조사함으로써 서울 수도권 지하철과 승객 흐름의 구조적 특성 등을 조사하였다. 아울러 승객들 흐름의 편중 현상을 조사하기 위하여 먹분포분석을 수행하여 결과 분석의 신빙성을 더하였다.

서론에 이어 2장에서는 네트워크 중심성 지표와 분석하고자 하는 서울 수도권 지하철 네트워크에 대하여 논하였다. 3장에서는 네트워크 중심성 지표인 매개, 근접, 차수 중심성을 유도하였으며 가중 매개 중심성 지표를 새롭게 제안하였다. 4장에서는 중심성 지표들 사이의 상관관계와 먹분포 분석을 수행하여 서울 수도권 지하철과 승객 흐름의 특성을 조사하였다. 마지막 5장에는 결론과 추후 연구를 수록하였다.

## 2. 네트워크 중심성 지표

네트워크의 전체적인 관점에서 위상적(Topological) 특성과 효율성을 조사하는 대신에 네트워크의 개별 노드들, 즉 개별 역의 특성을 정량화하기 위해 네트워크 중심성 지표를 사용한다. 네트워크 중심성 지표로는 크게 매개, 근접 그리고 차수 중심성 지표가 있는데 본 장에서는 이들에 대해 간단히 설명하였다. 아울러 승객들의 실제 흐름을 고려한 가중 매개 중심성을 본 연구에서 새롭게 제안하였다.

지하철은 네트워크  $G=(V, E)$ 로 나타낼 수 있는데  $V$ 는 노드로 표현되는 역들의 집합이고  $E$ 는 링크로 표현되는 인접한 두 역 사이의 연결 철로들의 집합이다.

### 2.1 매개 중심성(Betweenness Centrality: BC)

지하철 네트워크의 중심성 분석에 가장 자주 사용되는 매개 중심성은 특정 노드가 네트워크에 존재하는 노드들의 모든 쌍(Pair)간 최단 경로에 얼마나 많이 존재하는가를 나타내는 지표다. 일반적으로 이 지표는 각 노드의 중요성을 나타내기 위하여 사용된다. 예로 지하철에 100개의 역이 있다고 하면 모두  $_{100}C_2$ 의 최단 경로가 존재하는데 특정 역이  $_{100}C_2$ 의 조합 가운데 몇 번 존재하는가를 나타내는 지표로 이 값이 크면 클수록 특정 역의 중요도는 크다고 볼 수 있다. 매개 중심성은 또한 네트워크의 탄성력(Resilience)을 분석하는데 사용될 수 있는데 예로 높은 매개 중심성을 갖는 역이 고장 나거나 폐쇄된다면 지하철의 승객 흐름 패턴에 매우 큰 영향을 미친다.

지하철역  $v$ 의 매개 중심성  $BC(v)$ 는 다음의 식을 이용하여 구할 수 있다[7].

$$BC(v) = \sum_{i \neq j \neq v} \left[ \frac{1}{(n-1)(n-2)} \times \frac{\alpha(i \rightarrow j|v)}{\alpha(i \rightarrow j)} \right] \quad (1)$$

위 식에서  $n$ 은 역의 수를 나타낸다.  $\alpha(i \rightarrow j)$ 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 경로는 Dijkstra 알고리즘을 이용하여 구할 수 있다.  $\alpha(i \rightarrow j|v)$ 는 역  $i$ 에서  $j$ 로 가는 최단 경로 상에 역  $v$ 가 존재하지 않으면 0 존재하면 1의 값을 갖는다.  $(n-1)(n-2)$ 는 해당 역  $v$ 를 제외한 지하철 내에 있는 모든 역들 간 최단 경로의 수를 나타내는데 매개 중심성 값을 정규화하기 위하여 사용된다.

## 2.2 가중 매개 중심성(Weighted Betweenness Centrality: WBC)

2.1에서 정의한 매개 중심성은 식(1)에서 볼 수 있듯이 정규화 상수  $(n-1)(n-2)$ 로 나뉘는데 이는 모든 (출발지, 도착지) 조합이 동일한 가중치( $=1/(n-1)(n-2)$ )를 가짐을 의미한다. 예로 출발지가 강남역이고 도착지가 고속터미널역인 연간 승객 수는 출발지가 강남역이고 도착지가 인천역인 연간 승객 수의 350배에 이를에도 이 두 조합은 같은 가중치를 갖는다. 매개 중심성 지표가 더 의미 있는 값을 갖기 위해서는 승객 흐름량이 많은 ( $i \rightarrow j$ ) 조합에 보다 높은 가중치를 부여해야 타당하다. 따라서 본 연구에서는 다음과 같이 가중 매개 중심성지표인  $WBC(v)$ 를 제안하였다.

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

식(2)에서  $F(i \rightarrow j)$ 는 출발지가 역  $i$ 이고 도착지가 역  $j$ 인 월간 승객 흐름량을 나타내며  $\sum F(k \rightarrow l)$ 는 월간 승객 총 수를 나타낸다. 따라서  $WBC$ 에서는 출발지가 강남역이고 도착지가 고속터미널역인 경로의 가중치가 출발지가 강남역이고 도착지가 인천역인 경로의 가중치보다 350배 크게 된다. 어떤 역의  $WBC$ 가 크다는 것은 실제로 그 역을 통과하는 승객 수가 많다는 것을 의미하고 이는 그 역에 사고가 발생하여 그 역을 사용하지 못 할 경우 전체 지하철 망에 미치는 영향도가 크다는 것을 나타낸다. 따라서 지하철역의 유지 보수 시 각 역의  $WBC$  값을 토대로 한정된 인력과 자원을 할당해 나가는 것도  $WBC$ 의 좋은 활용 예가 될 수 있다.

## 2.3 근접 중심성(Closeness Centrality: CC)

근접 중심성은 네트워크의 특정 노드와 모든 다른 노드들 사이의 최단 거리들의 평균을 토대로 구해지는 지표다. 지하철 네트워크에서 특정 역의 근접 중심성  $CC(v)$ 가 크다는 것은 이 역에서 모든 다른 역으로 상대적으로 빠르게 이동 가능하다는 것을 나타내며 해당 역은 허브(Hub) 역일 가능성이 높다. 모든 역들의 근접 중심성 값이 비교적 일정하게 나오면 대부분의 역에서 모든 다른 역으로 이동하는데 소요되는 최단 거리의 평균이 비교적 일정하다는 것을 의미하기 때문에 지하철 망의 환승 시스템이 비교적 잘 구축되었다고 볼 수 있다. 그렇지 않고 근접 중심성 값에 편차가 심하면 환승 시스템에 문제가 있는 것으로 진단할 수 있다.

지하철 역  $v$ 의 근접 중심성  $CC(v)$ 는 최단 거리들의 평균의 역수로 구해지는데 다음 식을 이용하여 구할 수 있다.

$$CC(v) = \frac{n-1}{\sum_{w \neq v} d(v, w)} \quad (3)$$

위 식(3)에서  $d(v, w)$ 는 Dijkstra 알고리즘을 이용하여 구한 역  $v$ 와  $w$ 의 최단 거리를 나타내며  $n$ 은 지하철 네트워크의 총 역의 개수를 나타낸다.

## 2.4 차수 중심성(Degree Centrality: DC)

네트워크의 노드 차수는 노드의 링크 개수로 정의된다. 지하철 망의 경우에는 대부분 역의 차수는 2이고 환승역의 경우 2 이

상의 차수를 갖기 때문에 차수 중심성의 의미와 활용도는 높지 않다. 다만 본 연구에서는 차수가 높은 환승역이 매개 중심성과 근접 중심성도 높은지 등에 대한 상관관계를 분석하기 위하여 차수 중심성을 조사하였다.

지하철 역  $v$ 의 근접 중심성  $DC(v)$ 는 다음 식을 이용하여 구할 수 있다.

$$DC(v) = \sum_{w \neq v} \frac{z_{vw}}{n-1} \quad (4)$$

위 식(4)에서  $z_{vw}$ 는 역  $v$ 에서  $w$ 로의 직접 연결이 존재하지 않으면 0, 존재하면 1의 값을 가지며  $n$ 은 지하철 네트워크의 총 역의 개수를 나타낸다.

### 3. 서울 수도권 지하철 네트워크 중심성 지표 계산

본 장에서는 실 데이터를 사용하여 서울 수도권 지하철 네트워크의 매개, 가중 매개, 근접, 차수 중심성을 구하였다.

#### 3.1 데이터 수집과 계산 과정

본 연구를 위한 계산 과정으로는 크게 두 단계가 존재하였다. 첫째, 지하철 네트워크에서의 최단경로 계산이며 이는 시간을 기준으로 한다. 따라서 Seoul Open Data Plaza[10]에서 제공하는 인접한 두 역 사이의 시간 데이터와 모든 환승역들의 평균 환승시간인 2분을 사용하여 최단경로를 계산하였다. 둘째, 이렇게 구한 최단경로를 바탕으로  $BC(v)$ 와  $WBC(v)$ 를 계산하였다.  $BC(v)$ 의 경우 모든 최단경로의 가중치가 1이지만  $WBC(v)$ 의 경우 각 최단경로의  $F(i \rightarrow j)$ 를 토대로 가중치를 구하였다. 데이터는 Korea Transportation DB[11]로부터 월간 총 승객수인  $F(i \rightarrow j)$ 를 획득하여 사용하였다.

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  $BC$ 와  $WBC$  계산을 위한 조합의 수는  $242,556(=493 \cdot 492)$ 개다.

#### 3.2 네트워크 중심성 지표 계산

##### 3.2.1 매개 중심성(BC)

식(1)을 이용하여 서울 수도권 지하철의 역별 매개 중심성을 구하였다. 이중 상위 20개를 Fig. 1에 나타냈다. 왕십리역이 0.215로 가장 높는데 이는 승객들의 실제 흐름양을 고려하지 않을 경우에는 왕십리역의 중요도가 가장 높음을 나타낸다. 그 뒤로 사당역이 0.195, 신도림역이 0.177 순으로 나타났으며 전체 평균은 0.042다.

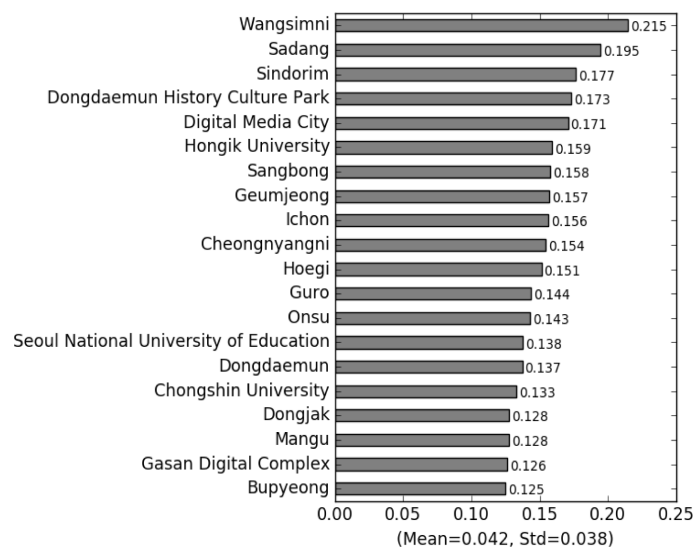


Fig. 1. Top 20 stations of BC.

### 3.2.2 가중 매개 중심성(WBC)

식(2)을 이용하여 서울 수도권 지하철의 역별 가중 매개 중심성을 구하였다. 이중 상위 20개를 Fig. 2에 나타냈다. 동대문역사문화공원역이 0.116으로 가장 높게 나타났는데 이는 승객들의 실제 흐름양을 고려할 경우 동대문역사문화공원역의 이용도가 가장 높다는 것을 말하며 만약 어떤 이유에서든 동대문역사문화공원역이 이용 불가능하면 서울 수도권 지하철에 가장 큰 피해가 발생한다는 것을 의미한다. 동대문역사문화공원역에 이어 신도림역, 교대역, 사당역, 왕십리역, 동대문역, 충무로역 등이 높게 나타났으며 전체 평균은 0.019로 나타났다. WBC 값이 높은 상위 20개 역중 10개 역은 BC 상위 20개 역에 포함되지 않아 둘 사이에 차이가 있는 것으로 나타났는데 이는 4장의 상관 분석에서 보다 자세히 논하였다.

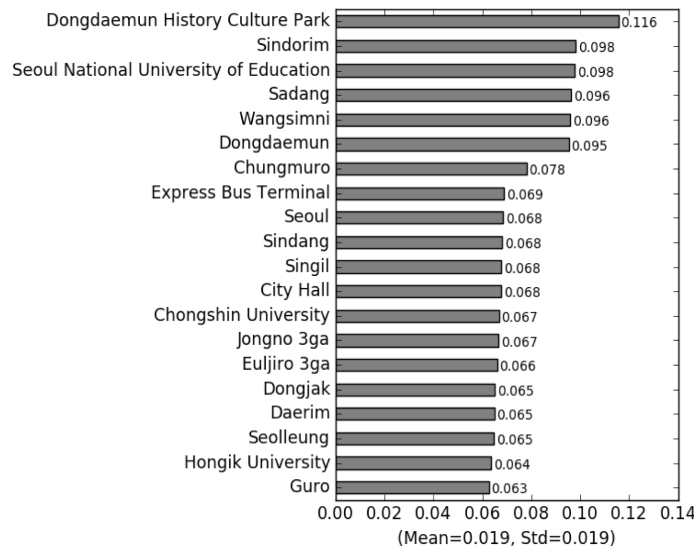


Fig. 2. Top 20 stations of WBC.

### 3.2.3 근접 중심성(CC)

식(3)을 이용하여 서울 수도권 지하철의 역별 근접 중심성을 구하였다. 이중 상위 20개를 다음 Fig. 3에 나타냈다. 그림에서 보듯이 근접 중심성은 상위 20개 역의 차이가 거의 없다. 전체 평균도 0.018로 가장 높은 서울역의 0.02571과 비교하여 큰 차이가 없으며 표준 편차도 0.005로 작은 값을 갖는다. 이는 서울 수도권 지하철 네트워크에는 강력한 허브(Hub) 역이 존재하지 않는다는 것을 의미한다. 아울러 대부분의 역에서 모든 다른 역으로 이동하는데 소요되는 최단 거리의 평균이 비교적 일정하다

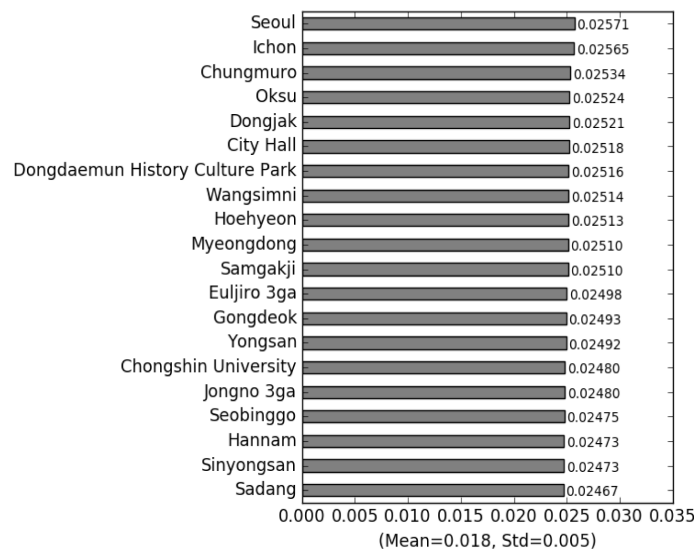


Fig. 3. Top 20 stations of CC.

는 것은 서울 수도권 지하철의 환승 시스템이 비교적 잘 구축되어 있다는 것을 의미한다. 근접 중심성이 높은 상위 20개 역중 10개 역은 WBC도 높은 것으로 나타나 어느 정도 상관관계 있어 보인다. 보다 구체적인 분석은 4장 상관관계 분석에서 수행하였다.

### 3.2.4 차수 중심성(DC)

식(4)을 이용하여 서울 수도권 지하철의 역별 차수 중심성을 구하였다. 이중 상위 20개를 다음 Fig. 4에 나타냈다. 이는 기본적으로 각 역에 존재하는 노선 수로부터 구해지는데 4개의 노선이 존재하는 공덕역과 왕십리역이 가장 높게 나타났고 그 뒤로 3개의 노선이 존재하는 홍대입구역, 디지털미디어시티역 등이 뒤를 잇고 있다. 특정 역의 차수가 높다는 것은 환승역을 의미하는데 차수가 가장 높은 공덕역은 역의 중요도를 나타내는 WBC 상위 20개 역에 포함되지 않고 상위권인 디지털미디어시티역과 김포공항역도 WBC 상위 20개 역에 포함되지 않고 있다. 이는 환승역이라고 해서 반드시 해당역의 중요도가 높다는 것을 의미하지 않는다는 것을 나타내며 4장 상관관계 분석에서 구체적인 상관관계 값을 제시하였다.

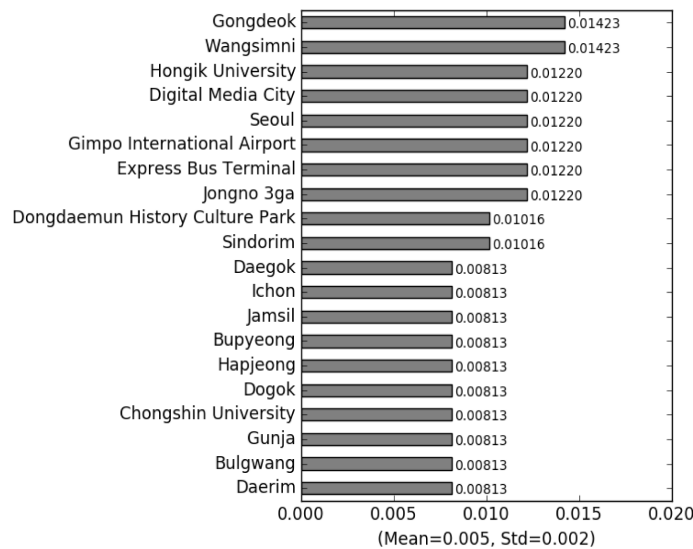


Fig. 4. Top 20 stations of DC.

## 4. 상관관계 분석

본 장에서는 네트워크 중심성 지표들 사이의 상관관계 분석을 수행하였다. 첫째, 본 연구에서 제안한 지표인 WBC의 유효성을 검증하기 위해 BC와 WBC 사이의 상관관계를 분석하였다. 만약 BC와 WBC 상관관계가 1에 가깝고 둘 사이의 직선 기울기가 1에 가깝다면 굳이 WBC를 사용할 필요가 없다. 그렇지 않다면 BC 대신에 실제 승객흐름을 고려한 WBC를 사용하는 것이 바람직하다. 둘째, BC와 WBC 그리고 DC와 CC사이의 상관관계 분석을 수행하였는데 이는 매개 중심성이 큰 노드들이 근접 중심성과 차수중심성도 큰지 여부를 확인하고 이로부터 서울 수도권 지하철의 특성을 일부 파악해 보았다.

### 4.1 BC와 WBC의 상관관계

BC와 WBC 사이의 상관관계를 다음 Fig. 5에 나타냈다. Pearson 상관계수는  $\rho=0.8$ 로 나타났으며 두 지표 사이에는 기울기가 0.4인 선형관계가 있는 것으로 나타났다.

상관관계가 0.8이고 두 지표 사이의 직선 기울기가 0.4라는 사실은 BC와 WBC값이 어느 정도 차이가 있다는 것을 나타낸다. WBC는 BC에 실제 승객흐름양을 가중한 것이기 때문에 승객흐름 분포의 특성이 상관관계에 영향을 미치리라 예측된다. 이를 확인하기 위하여 본 연구에서는  $F(i \rightarrow j)$  값으로 실제 데이터 대신에 최솟값인 0과 최댓값인 212,850 사이의 임의로 발생시킨 난수를 사용하여 WBC를 구하였다. 즉, 승객 흐름 분포를 최솟값인 0과 최댓값인 212,850 사이의 균일 분포(Uniform Distribution)로 가정하여 사용하였다. 이렇게 구한 WBC와 BC 사이의 상관관계를 다음 Fig. 6에 나타냈는데 Pearson 상관계수는 거의 1인  $\rho=0.999978$ , 두 지표 사이에는 기울기가 1인 선형관계가 있는 것으로 나타났다. 이는 승객 흐름 분포가 어떤 특정 구간에 편중되어 있지 않고 두 개의 역 사이에 승객 흐름양이 균일 분포를 따르는 경우에는(즉, 두 개의 역 사이의 승객 흐름양으로 균

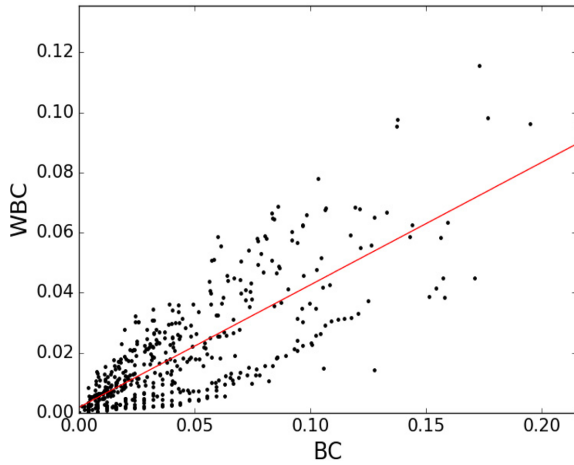


Fig. 5. Correlation between BC and WBC.

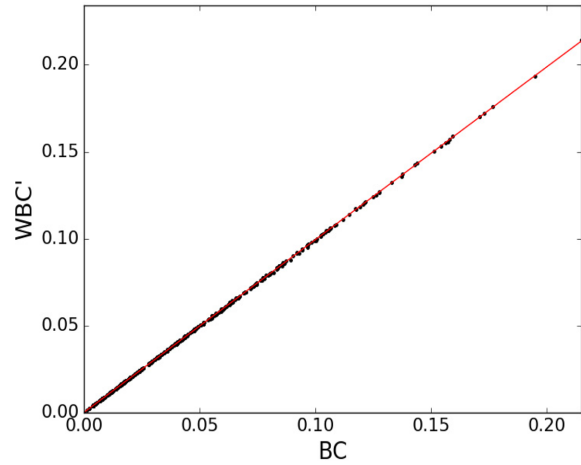


Fig. 6. Correlation between BC and WBC'.

일 분포에서 추출한 난수를 사용)  $WBC = BC$ 가 되어 굳이 WBC를 사용할 필요가 없음을 나타낸다.

$F(i \rightarrow j)$  값으로 1) 실제 데이터를 이용하는 경우와 ( $\rho = 0.8$ , 직선 기울기 = 0.4),  $F(i \rightarrow j)$  값으로 2) 최솟값인 0와 최댓값인 212,850 사이에 임의의 난수를 발생시켜 사용하는 경우에 ( $\rho = 0.999978$ , 직선 기울기 = 1) 상관계수와 직선의 기울기가 차이가 난다는 것은 다음의 사항들을 의미한다.

첫째, 가상의 예로든  $F(i \rightarrow j)$ 의 분포가 균일한 경우에는 Fig. 6에서보듯이 BC와 WBC가 거의 동일하여 굳이 승객 흐름을 고려한 WBC를 계산할 필요가 없다. 서울 수도권 지하철의 경우에는 비록 둘 사이에 상관관계가 존재하긴 하지만 Fig. 5에서 보듯이 두 개 값의 차이가 발생하여 승객 흐름을 고려한 WBC를 계산하여 사용할 필요가 있다. 앞의 3장에서도 서울 수도권 지하철에서 WBC값이 높은 상위 20개 역중 10개 역은 BC값이 높은 상위 20개 역에 포함되지 않아 둘 사이에 차이가 있는 것을 확인하였다.

둘째,  $F(i \rightarrow j)$ 의 분포가 균일 분포가 아닌 어떤 특정 분포를 따르는 것을 확인할 수 있다. 실 데이터를 이용하는 경우에는 대략  $WBC(v) = 0.4 \cdot BC(v)$ 가 되는데 대부분의 WBC값이 BC값보다 작다. BC값은 242,556(=493·492)개의 모든 (출발지, 도착지) 조합이 같은 가중치를 갖는데 비해서 WBC는 (출발지, 도착지) 조합에 따라 서로 다른 가중치를 갖는다. 일반적으로 WBC값이 BC값보다 작다는 사실은 대부분의 (출발지, 도착지)의 승객 수요가 평균보다 작아 WBC의 대부분 가중치가 BC의 가중치보다 작고 승객 수요가 매우 큰 값들이 소수 존재한다는 것을 의미한다. 즉, 중앙값(Median)이 평균값보다 매우 작은 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 매우 편향(Skewed)된 분포를 따름을 수치적으로도 확인할 수 있다.

분포가 편향된 대표적인 분포는 멱분포(Power-law)인데 자연이나 사회에서 나타나는 여러 현상들을 잘 나타내주기 때문에 최근 들어 많은 관심을 받고 있다. 멱분포는  $y = x^{-\alpha}$ 로 나타내지는데  $\alpha$ 는 대략 2와 3사이의 값을 갖는다. 이 분포는 평균에 비해 매우 큰 값들이 다수 개 존재하는 긴 꼬리 분포(Long Tail Distribution)로 평균값이 의미 없다. 실제로 (출발지, 도착지)의 월간 승객 수요  $F(i \rightarrow j)$ 의 평균값은 802명인데 반해 최댓값은 212,850명으로 나타났다.

앞에서  $F(i \rightarrow j)$ 가 매우 편향된 분포라는 것을 인지했기 때문에  $F(i \rightarrow j)$ 가 멱분포를 따르는지 통계적으로 검증하였다. 검증하는 방법은  $F(i \rightarrow j)$ 의 CCDF(Complimentary Cumulative Density Function)가 Log-Log Scale에서 기울기가 대략 2와 3 사이의 직선 식을 따르는지 조사하는 것이다. 다음 Fig. 7은 Log-Log Scale에서  $F(i \rightarrow j)$ 의 CCDF이며, Kolmogorov-Smirnov 테스트를 통해  $F(i \rightarrow j)$ 가 직선식을 따르며(p-value = 0.24) 기울기가 -3인 Power-law 분포를 따르는 것으로 확인되었다.

앞에서 보였듯이 월간 (출발지, 도착지) 승객 수를 나타내는  $F(i \rightarrow j)$ 가 균일한 분포를 따를 경우에는 BC와 WBC가 동일하여 굳이 승객 흐름을 고려한 WBC를 계산할 필요가 없다. 그러나 위에서 밝힌 것처럼  $F(i \rightarrow j)$ 가 균일한 분포가 아닌 매우 편향된 멱분포를 따르면 앞에서 설명한 것처럼 중앙값과 평균값이 매우 다르기 때문에 BC와 WBC의 값이 많이 다르게 된다. 이는 BC 대신에 (출발지, 도착지) 승객흐름양을 고려한 WBC 사용의 타당성을 한층 더 부각시킨다.

(출발지, 도착지)의 승객흐름양을 나타내는  $F(i \rightarrow j)$ 가 매우 편향된 멱분포를 따른다는 사실은 본 연구에서 새롭게 제안한 가중 매개 중심성 지표인 WBC 사용의 타당성을 입증하는데서 나온 부수적인 결과다. 이 결과는 지하철 전체 망에서 승객흐름이 매우 과도하게 편중되는 지역들이 다수 존재한다는 것을 의미하기 때문에 지하철 효율적 운영에 매우 방해되는 요인이다. 지하

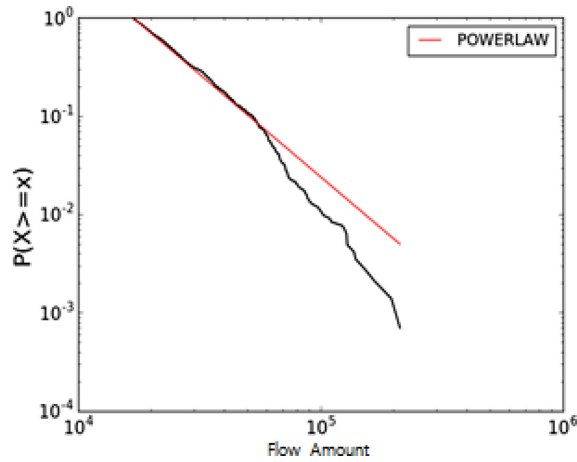


Fig. 7. CCDF of passenger flow.

철 승객흐름의 분포가 멱 분포를 따르는지의 연구는 최근 들어 베이징 지하철을 대상으로 수행된 바 있지만[12] 국내에서는 아직 시도되지 않았다. 본 연구에서 규명한 (출발지, 도착지) 승객흐름량의 특성만으로는 서울 수도권 지하철의 승객흐름 특성을 규명하는데 한계가 있다. (출발지, 도착지)의 승객흐름량 뿐만 아니라 인접한 두 역 사이의 실제 탑승(On-Board) 승객 수, 각 역을 지나가는 승객 수(Through flow), 일정 시간 동안 지하철을 이용하는 총 승객 수 중 특정 역을 경유하는 승객 수의 비율로 정의 되는 역의 효율성(Throughput) 등에 대한 자료를 바탕으로 승객흐름 특성을 보다 종합적으로 정확하게 파악할 수 있다면 서울 수도권 지하철에 대한 보다 거시적이고 정책적인 분석이 가능할 것이다.

#### 4.2 다른 중심성 지표 사이의 상관관계

BC와 WBC 그리고 DC와 CC사이의 Pearson 상관관계를 구하여 다음 Table 1에 정리하였고 Fig. 8과 Fig. 9에 상관관계도를 나타냈다.

먼저 BC와 DC 사이의 상관계수는 0.59로 나타나 강한 상관관계를 보여 주지 못한다. 이는 지하철 환승역이라고 해서 역의 중요도가 반드시 크지 않다는 사실을 나타낸다. 3장에서 살펴 본대로 차수가 높은 역들이 역의 중요도를 나타내는 BC 상위 20개

Table 1. Correlation coefficients between centrality measures.

	Closeness centrality (CC)	Degree centrality (DC)
Betweenness centrality (BC)	0.41	0.59
Weighted betweenness centrality (WBC)	0.65	0.63

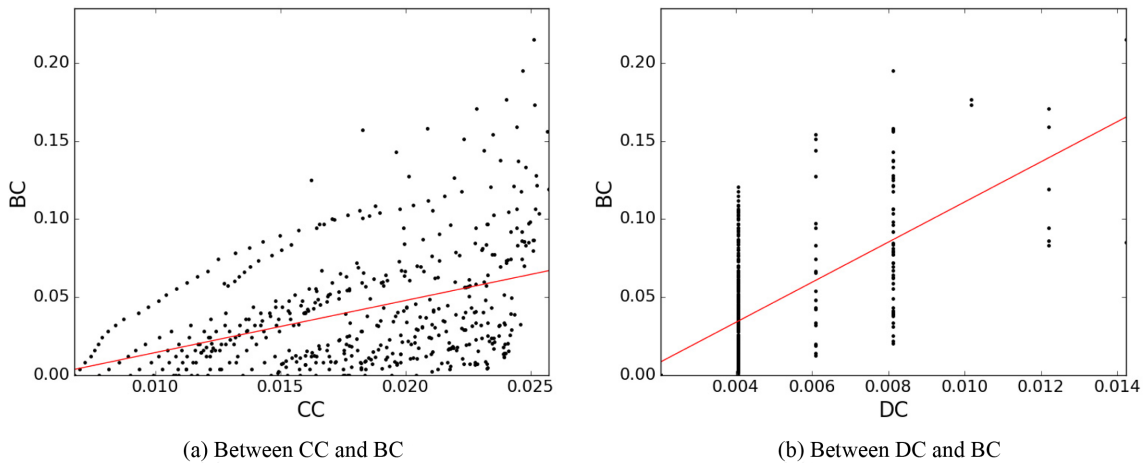


Fig. 8. Correlation with BC.



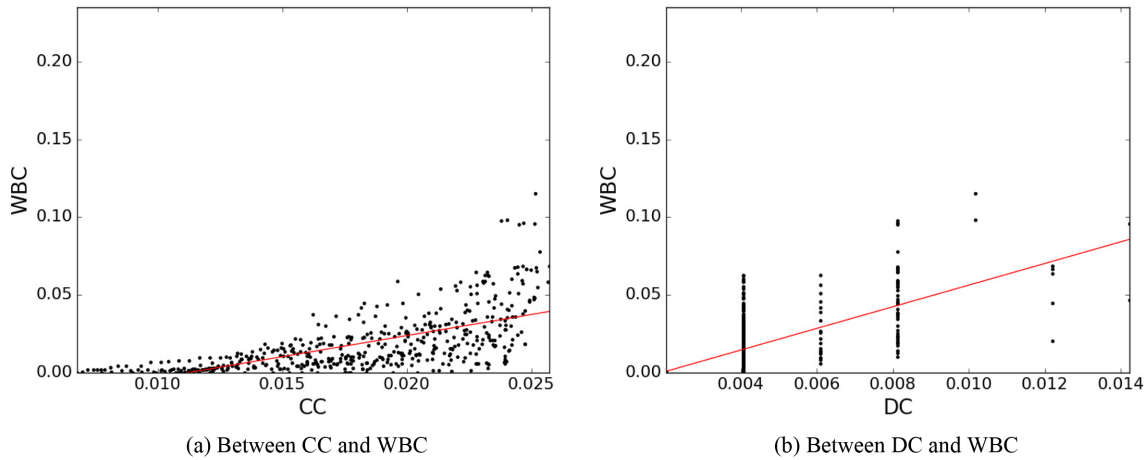


Fig. 9. Correlation with WBC.

역에 상당 부분 포함되지 않는다는 사실과 부합되는 결과다. 한편 BC와 CC 사이의 상관계수는 0.41로 매우 약한 상관관계를 보여 준다. 앞의 Fig. 3에서 볼 수 있듯이 CC는 지하철 네트워크에서 역의 위치에 따른 차이가 크지 않아 일정하다. 앞의 3장 차수 중심성 분석에서 언급했듯이 서울 수도권지하철의 효율적인 환승 시스템으로 인해 역의 위치에 상관없이 특정 역에서 모든 다른 역으로의 평균 최단 거리는 비교적 일정하게 나타나기 때문에 역의 위치에 따라 다른 값을 갖는 매개 중심성과는 낮은 상관관계를 가질 수밖에 없다.

WBC의 경우도 BC와 마찬가지로 CC와 DC 모두와 강한 상관관계를 나타내지 못한다. DC와 상관계수는 0.63으로 BC와 마찬가지로 지하철 환승역과 WBC로 나타나는 역의 중요도는 반드시 일치하지 않는다. 이는 차수 중심성이 가장 높은 공덕역이 역의 중요도를 나타내는 WBC 상위 20개 역에 포함되지 않고 차수 중심성이 매우 상위권인 디지털미디어시티역과 김포공항역도 WBC 상위 20개 역에 포함되지 않는다는 3장의 분석 내용과 일치하는 결과다. 한편 WBC와 CC 사이의 상관계수는 0.65로 BC보다는 다소 강한 상관관계를 보여준다. 이는 WBC는 승객흐름양을 고려하기 때문에 지하철 네트워크의 외곽보다 중심부에 있는 역들의 승객흐름양이 많아 WBC가 비교적 크게 되고 이 역들이 외곽에 있는 역들보다는 CC도 미미하게나마 다소 크기 때문에 두 지표 사이에 상관관계가 형성된 것으로 보인다.

## 5. 결 론

매개, 근접 그리고 차수 세 개의 네트워크 중심성 지표를 사용하여 서울 수도권 지하철 네트워크의 개별 역의 중요성을 분석하고 이로부터 서울 수도권 지하철 네트워크와 승객흐름 특성을 분석 하였다. 본 연구에서는 기존에 제안된 매개 중심성 지표인 BC와 승객들의 실제 흐름양을 함께 고려한 가중 매개 중심성 지표 WBC를 새롭게 제안하였다. 그리고 BC와 WBC의 상관관계 분석을 통하여 서울 수도권 지하철 승객 흐름의 편중 현상과 가중 매개 중심성인 WBC 사용의 유효성을 확인하였다. 그리고  $F(i \rightarrow j)$ 의 CCDF로부터 Kolmogorov-Smirnov 테스트를 통해  $F(i \rightarrow j)$ 가 멱분포(Power-law)를 따르는 것을 확인하였다. 아울러 BC와 WBC 그리고 DC와 CC사이의 Pearson 상관관계도 조사하였다.

(출발지, 도착지)의 승객흐름양을 나타내는  $F(i \rightarrow j)$ 가 멱분포를 따른다는 사실은 본 연구에서 새롭게 제안한 가중 매개 중심성 지표인 WBC 사용의 타당성을 입증하는데서 나온 부수적인 결과로 서울 수도권 지하철의 승객 흐름이 과도하게 일부 지역들에 편중되어 있다는 사실을 나타낸다. 본 연구에서 제안한 매개 중심성 지표인 WBC와 결합하여 서울 수도권 지하철 망에 대한 보다 심층적이고 정책적인 분석을 위해서는 (출발지, 도착지) 승객흐름양 뿐만 아니라 인접한 두 역 사이의 실제 탑승 승객 수, 각 역을 지나가는 승객 수, 일정 시간 동안 지하철을 이용하는 총 승객 수 중 특정 역을 경유하는 승객 수의 비율로 정의되는 역의 효율성 등에 대한 자료를 바탕으로 승객흐름 특성을 보다 정확하게 종합적으로 파악할 수 있어야 한다. 이 부분에 대한 연구는 본 논문의 범위를 훨씬 뛰어 넘는 것으로 추후 계속 연구로 수행 할 예정이다.

본 연구는 앞에서도 언급했듯이 민자 노선이 2013년 9월 이후 승객 수송 관련 자료를 공개하지 않기 때문에 데이터를 수집할 수 있는 가장 최근 자료인 2013년 2월 자료를 토대로 분석하였다. 추후에 최근 자료들을 확보할 수 있으면 보다 더 현실적이고 정확한 결과를 도출할 수 있을 것이다.

## 후 기

이 연구는 서울과학기술대학교 교내연구비의 지원으로 수행되었습니다.

## References

- [1] S. Derrible (2012) Network Centrality of Metro Systems, *PLoS One*, 7(7), e40575, pp. 1-10.
- [2] V. Latora, M. Marchiori (2001) Efficient Behavior of Small-World Networks, *Physical Review Letters*, 87(19), P.198701.
- [3] M.L. Mouronte (2014) Topological Analysis of the Subway Network of Madrid, *International Multi-Conference on Computing in the Global Information Technology*, Seville, Spain, pp. 9-13.
- [4] S. Stoilova, V. Stoev (2015) An Application of the Graph Theory Which Examines the Metro Network, *Transport Problems*, 10(2), pp. 35-48.
- [5] B. Leng, X. Zhao, X. Zhang (2014) Evaluating the Evolution of Subway Networks: Evidence from Beijing Subway Network, *Europhysics Letters*, 105(5), P.58004.
- [6] S.Y. Zang, K.W. Lee (2016) Characteristics and Efficiency Analysis of Evolutionary Seoul Metropolitan Subway Network, *Journal of the Korean Society for Railway*, 19(3), pp. 388-396.
- [7] C.P. Monterola, M.A. Ramli, G.L. Khoon, T.H. Guang (2014) A Method to Ascertain Rapid Transit System's Throughput Distribution Using Network Analysis, *Procedia Computer Science*, 29, pp. 1621-1630.
- [8] A. H'aznagy, I. Fi, A. Londo, T. N'emeth (2015) Complex Network Analysis of Public Transportation Networks: A Comprehensive Study, *2015 Models and Technology for Intelligent Transport Systems*, Budapest, Hungary, pp.371-378.
- [9] Y.Y. Cheng, R.K. W. Lee, E.P. Lim, F. Zhu (2015) Measuring Centralities for Transportation Networks beyond Structures, Applications of Social Media and Social Network Analysis, Springer International Publishing, Switzerland, pp. 23-39.
- [10] Seoul Open Data Plaza, <http://data.seoul.go.kr> (Accessed 30 January 2016)
- [11] Korea Transportation Database, <http://www.ktdb.go.kr> (Accessed 30 January 2016)
- [12] Q. XU, B.H. Mao, Y. Bai (2016) Network Structure of Subway Passenger Flow, *Journal of Statistical Mechanics: Theory & Experiment*, pp. 1-18.

(Received 4 April 2017; Revised 17 May 2017; Accepted 30 May 2017)

---

**Jeong-Won Lee:** iris3933@naver.com

Department of Industrial and Information Systems Engineering, Seoul National University of Science and Technology,  
232, Gongneung-ro, Nowon-gu, Seoul 01811, Korea

**Kang-Won Lee:** kwlee@snut.ac.kr

Department of Industrial and Information Systems Engineering, Seoul National University of Science and Technology,  
232, Gongneung-ro, Nowon-gu, Seoul 01811, Korea