

사회 네트워크 분석을 이용한 지하철역 네트워크 구조 분석

Subway Stations Network Structure Analysis by Using Social Network Analysis

한경훈*, 서영수*, 박근병*

Kyung Hoon Han*, Young Su Seo*, Geun Byung Park*

Abstract A social network is a social structure made up of a set of individuals or organizations linked to a variety of ties such as relationship with friends, business relations, communication frequencies, or providing information. The network is described in terms of nodes or points and the ties or links between them. This study is a structural analysis of subway stations network from the perspective of network science by defining that nodes are the subway stations and links are the movements of subway passengers. This study, using data which are passengers getting on and off subway trains within the Seoulmetro section for one year in 2013, analyzed the network characteristics of the subway stations on the basis of the notion of centrality, which is a concept used in the social network analysis(SNA).

Keywords : Subway stations network, Centrality, Social network analysis, Network characteristics, SNA

초 록 사회 네트워크(social network)는 개인이나 기관이 친구 관계, 거래 관계, 통신 빈도, 정보 제공 등과 같은 다양한 연결 관계에 의해 구성된 사회적 구조이다. 이 네트워크는 노드들(nodes) 혹은 점들(points)과 그들 사이의 관계(tie) 혹은 링크(link)에 의해 표현된다. 본 논문은 지하철역을 노드로 역간 승객들의 이동을 연결 관계로 정의한 지하철역 네트워크를 네트워크 과학 측면에서 구조적으로 분석한 것이다. 이번 논문은 2013년 1년 동안 서울메트로 구간 내 승객의 승·하차에 대한 데이터를 이용하였으며, 사회 네트워크 분석에서 활용되는 중심성 개념을 기준으로 지하철역의 네트워크 특성을 분석하였다.

주요어 : 지하철역 네트워크, 중심성, 사회 네트워크 분석, 네트워크 특성, SNA

1. 서 론

사회 네트워크(social network)는 사람들이 연결되어 있는 관계망으로 표현할 수 있다[1]. 관계는 친구/친족 관계, 정치적 관계, 거래 관계, 권위/권력 관계, 연인/원수 관계, 통신 빈도, 정보 제공 등 다양하게 정의될 수 있다. 이러한 관계망 구조는 사람들의 삶에 영향을 주며 그 구조에서 한 개인(행위자)이 어느 위치에 놓여 있는지에 따라 그 개인의 의식과 행동은 달라진다. 그리고 달라진 개인들은 구조에 변화를 가져온다. 이런 점에서 행위와 구조 사이에는 상호작용이 존재하며 이 상호작용은 구조는 행위를 강요하지만 동시에 그 강요 받는 행위에 의해 매 순간 다시 생산되는 존재라는 점을 설파하였던 영국의 사회학자 앤서니 기든스 (Anthony Giddens)가 고안한 개념인 “구조의 이중성(duality of structure)”을 구체적으로 접근하고 있다[2]. 마찬가지로 지하철도 네트워크 관점에서

* 서울메트로 정보관리처

접근해 보면 지하철역의 연결 구조와 사람들의 역 간 이동 패턴 사이에는 상호작용이 존재하며 이 상호작용은 지하철역의 연결망 구조가 사람들의 이동 패턴에 영향을 미치지만 동시에 사람들의 이동 패턴은 다시 지하철역의 연결망 구조를 재생성하거나 변화시키는 역할을 한다고 할 수 있다.

사회 네트워크 분석(social network analysis, SNA)은 다수의 점(노드, 꼭지점, 행위자, 구성원)과 이들을 연결하는 선(모서리, 이음)으로 구성된 망(network)에 대한 사회과학적·통계적 분석이다[3].

본 논문은 이러한 사회 네트워크 분석 관점에서 지하철역을 노드로 승객들의 역간 이동을 연결 선으로 간주하여 구조적 분석을 하였으며, 네트워크 과학에서 통용되고 있는 중심성(연결정도 중심성, 근접 중심성, 아이겐벡터 중심성) 개념을 활용하여 지하철역의 네트워크적 특성에 대해 고찰해 보았다.

2. 본 론

2.1 데이터 수집 및 분석 방법

2.1.1 데이터 수집 및 정제

서울메트로 구간 내에서 2013년 1년 동안 승객들의 승·하차에 대한 Raw Data(200GB, 일회권 제외)를 수집했으며, 파이썬(python)으로 수집한 데이터를 분석에 용이한 행렬 형태로 변환하였다. 행렬 형태의 데이터는 Table 1과 같으며 총 120×120 행렬이다. 첫 번째 열은 승차역, 첫 번째 행은 하차역에 해당되며 나머지 행과 열이 교차하는 셀이 이동 인원이다.

Table 1 Matrix data of passengers getting on and off

	150	151	152	...	타기관
150	328635	470636	812736	...	11535180
151	516509	100608	183749	...	5490085
152	879054	161462	154135	...	9816753

2.1.2 분석방법

본 논문에서는 지하철역 연결망을 계량 네트워크(valued network) 개념으로 분석하였다. 계량 네트워크는 방향에 대한 정보뿐만 아니라 관계의 질에 대한 정보를 담고 있다. 방향은 관계의 시작(sender)과 끝(receiver)에 대한 정보를 제공한다. 즉, 서울역에서 승차해서 사당역에 하차한 경우와 사당역에서 승차해서 서울역에 하차한 경우를 구분하기 위해서이다. 그리고 관계의 질은 역간 승객의 이동 빈도에 대한 정보를 제공한다.

이번 연구에서는 네트워크의 구조적 속성인 중심성(centrality)을 측정했다. 중심성은 어느 한 행위자가 전체 네트워크에서 중심에 위치하는 정도를 표현하는 지표로써, 지하철역 네트워크 내에서는 중심성 분석을 통해 어느 역이 얼마나 광범위하게 다른 역들과 연결되어 있는지를 나타내는 개념으로 사용하였다. 네트워크 분석을 하기 위해서 통계 소프트웨어인 R의 sna package를 이용하여 분석하였다.

2.2 분석 및 결과

2.2.1 서울메트로 2013년 승·하차 인원

서울메트로(1~4호선) 2013년 1~12월 일평균 승·하차 인원 및 순위는 Table 2와 같다.

Table2 The daily average number of passengers getting on and off

순 위	승 차		하 차	
	역 명	일 평균	역 명	일 평균
1	강남	105,295	강남	109,060
2	잠실	78,297	신림	72,073
3	신림	74,034	잠실	71,838
4	서울역(1)	73,855	홍대입구	70,320
5	홍대입구	66,699	삼성	65,033
6	신도립	65,385	신도립	65,004
7	삼성	63,261	서울역(1)	64,948
8	구로디지털	62,884	고속터미널	63,296
9	고속터미널	58,637	구로디지털	62,920
10	선릉	57,239	신촌	56,033

(출처: 서울메트로 홈페이지 수송통계 자료실)

2.2.2 연결정도 중심성(degree centrality)

연결정도 중심성은 한 노드가 다른 노드와 연결된 정도를 중심으로 보는 개념이며 연결된 노드가 많으나 적으냐의 여부가 절대적인 기준이 된다[4]. 연결정도 중심성은 직접적(direct)으로 연결되어 있는 노드의 수로 측정되기 때문에 국지적(local) 중심성의 의미가 강하다[5]. 연결정도 중심성은 식(1)로 표현된다.

$$D_i = \sum_{j=1}^N \frac{Z_{ij}}{N-1} \quad (1)$$

Z_{ij} : 노드 i에서 j로의 직접연결, N: 네트워크내 전체 노드의 수
 식(1)을 이용하여 연결정도 중심성을 구해보면 Table3과 같다.

Table 3 Degree centrality analysis result

순 위	승 차		하 차	
	역 명	중심성	역 명	중심성
1	강남	37118413	강남	38444197
2	잠실	27150964	신림	25346889
3	신림	26094440	잠실	24905999
4	서울역(1)	24426449	홍대입구	24497879
5	홍대입구	23249884	신도립	22860886
6	신도립	23016382	삼성	22770656
7	구로디지털	22300899	구로디지털	22278660
8	삼성	22231561	서울역(1)	21827377
9	선릉	20209973	고속터미널	21214269
10	고속터미널	19758461	신촌	19565119

지하철역 연결망은 방향성을 가지고 있기 때문에 나가는 방향(outward direction)은 승차, 들어오는 방향(inward direction)은 하차로 정의해서 분석하였다.

연결정도 중심성 결과를 보면 승차 중심성은 강남, 잠실, 신림 순으로, 하차 중심성은 강남, 신림, 잠실 순으로 나타났다. 어느 한 역의 연결정도 중심성은 다른 역과의 연결 빈도수로 측정되기 때문에 Table 2의 일평균 승·하차 순위와 큰 차이가 없음을 알 수 있다. 역간의 순위 차이는 일평균 승·하차 인원은 역간 연결관계가 아닌 단순히 게이트를 통과한 인원 개념으로 계산된 반면, 연결정도 중심성은 역간 연결관계가 중심이며 대각행렬 즉, 승차역과 하차역이 같은 닫힌 워크(closed walk)는 제외한다.

2.2.3 근접 중심성(closeness centrality)

근접 중심성은 한 노드가 다른 노드에 얼마만큼 가깝게 있는지를 나타내는 것으로 두 노드 사이의 거리(distance)가 핵심 개념이다[6]. 근접 중심성은 직접적으로 연결된 노드뿐만 아니라 간접적으로 연결된 모든 노드들 간의 거리를 계산하여 중심성을 측정하기 때문에 네트워크 전체의 총체적인 관계를 고려할 수 있는 글로벌(global) 중심성 측정이 가능하다[7].

근접 중심성은 식(2)로 표현된다.

$$C_i = (N-1) \left[\sum_{j=1}^N d(i,j) \right]^{-1} \quad (2)$$

$d(i,j)$: 노드 i에서 j에 이르는 최단경로의 길이, N:네트워크내 전체 노드의 수
 식(2)을 이용하여 근접 중심성을 구해보면 Table 5와 같다.

Table 5 Closeness centrality analysis result

순 위	승 차		하 차	
	역 명	중심성	역 명	중심성
1	서울역(1)	1167271	서울역(1)	1137475
2	신도림	1155496	홍대입구	1135330
3	홍대입구	1153548	강남	1133355
4	강남	1152571	신도림	1130713
5	종각	1147133	종각	1117037
6	고속터미널	1137530	고속터미널	1114188
7	구로디지털	1112079	신촌	1093476
8	신림	1109772	구로디지털	1091374
9	신촌	1107789	삼성	1081503
10	선릉	1097940	신림	1079967

지하철역 네트워크에서는 역들간의 실제 물리적인 거리가 아닌 승·하차 이동 빈도의 가중치를 역의 개념으로 사용하여 논리적 거리를 계산하였다. 즉, 역간 승객 이동 빈도가 많을수록 역간 거리가 가깝고 이동 빈도가 적을수록 거리는 먼 것이다.

근접 중심성 분석 결과 승·하차 중심성은 서울역(1)이 1순위, 종각이(일평균 인원에서는 승차 15위, 하차 22위, 연결정도 중심성에서는 승차 17위, 하차 22위) 5순위로 상승했다. 반면, 잠실은 승차 11위(9↓), 하차 13위(10↓)로 순위가 내려갔다. 1순위 서울역(1)은 전체

지하철역 네트워크에서 다른 역들과 최단 경로로 연결되어 있다는 것을 의미한다. 즉, 직접적인 연결뿐만 아니라 간접적인 연결에 있어서 서울역(1)이 전체 네트워크에서 글로벌 중심성을 보이고 있다.

2.2.4 아이겐벡터 중심성(eigenvector centrality)

아이겐벡터(Eigenvector) 중심성은 연결된 노드의 개수뿐만 아니라 연결된 노드가 얼마나 중요한지도 함께 고려함으로써 연결정도 중심성의 개념을 확장한 것이다[8]. 즉, 아이겐벡터 중심성은 자신과 연결된 이웃들의 중심성을 가중치로 하여 자신의 중심성을 판단하는 데 사용한다. 위세가 높은 기관과 많이 접촉할수록 그 기관의 위세가 높아진다는 개념이다.

아이겐벡터 중심성은 식(3)으로 표현한다.

$$V_i = \sum_{j=1}^N C_j Z_{ij} \quad (3)$$

C_j :노드 j의 중요도, Z_{ij} :노드 i에서 j로의 연결 벡터C는 $\lambda C = ZC$ 의 고유 방정식의 해이다. 가장 큰 고유값(λ)에 해당하는 고유벡터로부터 아이겐벡터 중심성을 도출해 보면 Table 6과 같다.

Table 6 Eigenvector centrality analysis result

순 위	승 차		하 차	
	역 명	중심성	역 명	중심성
1	강남	0.188964	강남	0.194959
2	서울역(1)	0.177135	홍대입구	0.171318
3	신도림	0.166401	서울역(1)	0.167083
4	홍대입구	0.163229	신도림	0.165731
5	종각	0.145428	고속터미널	0.143011
6	고속터미널	0.140399	종각	0.137951
7	신림	0.139323	신림	0.134494
8	잠실	0.134831	구로디지털	0.134037
9	구로디지털	0.13324	삼성	0.129416
10	삼성	0.124316	신촌	0.128105

아이겐벡터 중심성 분석 결과 승·하차 중심성은 강남이 1 순위를 나타낸다. 일평균 및 연결정도 중심성에서 10 위권 밖에 있던 종각이 5,6 위로 순위가 상승한 반면, 잠실은 승차 8 위(6↓), 하차 11(8↓)위로 순위가 내려갔다. 1 순위 강남은 연결정도 중심성이 높은 다른 역들과 많은 연결을 맺고 있다는 것을 알 수 있다.

2.2.5 지하철역 네트워크 시각화

지하철역 네트워크의 시각화를 위해 Fig 1은 R 소프트웨어의 rgl package 안에 있는 gplot3d 함수를 사용하였고 Fig 2는 네트워크 및 복잡계 시스템의 시각화에 사용되는 도구인 Gephi를 이용하여 강남역 중심으로 시각화하였다. 색깔은 1~4 호선을 구분하며 연결 빈도가 많은 역은 원의 크기가 크고 역간 이동 빈도는 선의 두께로 표현하였다.

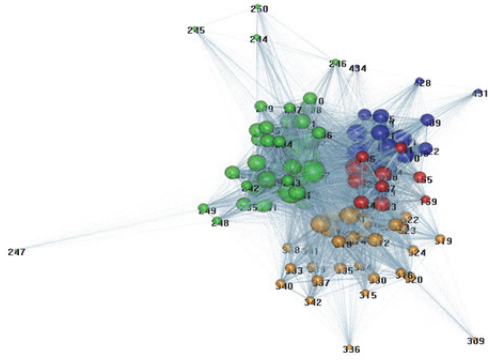


Fig. 1 3-dimension graph of subway stations network

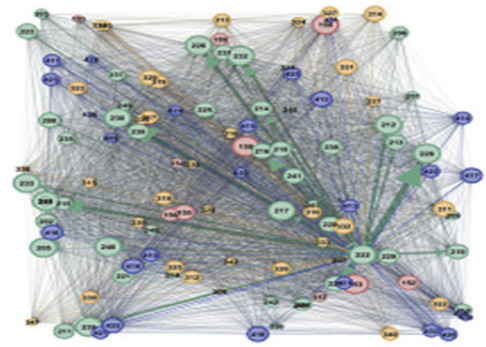


Fig. 2 Subway stations network visualization of using Gephi

3. 결론

본 논문은 사회 네트워크 분석 방법을 통해 지하철역 연결망의 구조적 특징 분석에 중점을 둔 것이다. 분석 결과 강남역은 승·하차 인원뿐만 아니라 중심성 분석에서도 높은 순위를 나타냈으며 이는 강남역이 지하철역 연결망에서 허브 역할을 한다고 볼 수 있다. 종각역은 일평균 승·하차 순위에서는 승차 15 위, 하차 22 위를 나타냈으나 근접 중심성에서는 승차 5 위(10↑), 하차 5 위(17↑), 아이겐벡터 중심성에서는 승차 5 위(10↑), 하차 6 위(16↑)로 순위가 상승했다. 이는 종각역이 직접적인 연결에 비해 상대적으로 연결정도가 많은 역들과 많은 연결관계를 가지고 있음을 알 수 있다. 반면, 잠실역은 일평균 승·하차 순위에서는 승차 2 위, 하차 3 위를 나타냈으나 근접 중심성에서는 승차 11 위(9↓), 하차 13 위(10↓), 아이겐벡터 중심성에서는 승차 8 위(6↓), 하차 11 위(8↓)로 순위가 내려갔다. 이는 잠실역이 직접적인 연결빈도는 높으나 상대적으로 연결정도가 많지 않은 역들 중심으로 연결되어 있음을 알 수 있다. 이러한 중심성에서 큰 편차를 보이는 역들에 대해서는 그 역들의 노선 특성에 대한 보다 세밀한 연구가 필요하다.

중심성 분석 결과 어느 하나의 중심성 지수가 높거나 혹은 낮아도 다른 중심성 지수가 낮거나 높은 역이 존재하며 이렇기 때문에 각 역들의 노선 특성을 보다 자세히 살펴보고 네트워크 측면에서 다양한 관점으로 짚어봐야 할 필요성이 있다.

본 연구의 한계점으로는 데이터 수집의 제한으로 서울메트로 구간만 분석하였으나 향후 데이터 범위를 확대하여 수도권 전체 지하철역에 대한 네트워크 구조 분석을 한다면 수도권 전체 노선에 반영 될 수 있는 일반화 된 결론을 도출할 수 있을 것이다.

더 나아가서는 중요한 사회적 변수들의 분포로 정의하는 분포 구조주의 개념을 통해 구체화 할 수도 있다. 지하철을 이용하는 승객들의 이동 패턴은 역 주변의 인구분포, 상권, 교통망, 주거지, 직장 등 다양한 변수들의 분포에 영향을 받는다. 이렇게 분포로부터 오는 구조적인 영향력을 ‘구조적 효과’ 라고 한다[9]. 간단한 모델 $S(t+1)=NS(t)$ 에서 N은 지하철역 연결망 행렬이고 S는 역 주변의 다양한 변수들의 분포 또는 승객들의 이동 특성을 나타낸다면, t시점에서 S의 특성과 연결망 행렬 N에 의해 t+1 시점에서 S의 패턴을 예측할 수 있을 것이다.

참고문헌

- [1] 손동원 (2002) 사회네트워크 분석, 경문사, 서울시 서대문구, pp.1.
- [2] Anthony Giddens (1986) The Constitution of Society, University of California Press, pp.25.
- [3] 허명희 (2012) R을 활용한 사회네트워크 분석, 자유 아카데미, 경기도 파주시, pp. 1.
- [4] 손동원 (2002) 사회네트워크 분석, 경문사, 서울시 서대문구, pp.95.
- [5] 손동원 (2002) 사회네트워크 분석, 경문사, 서울시 서대문구, pp.98.
- [6] 손동원 (2002) 사회네트워크 분석, 경문사, 서울시 서대문구, pp.95.
- [7] 손동원 (2002) 사회네트워크 분석, 경문사, 서울시 서대문구, pp.100~101.
- [8] 광기영 (2014) 소셜네트워크 분석, 청람, 서울시 마포구, pp.212.
- [9] 김용학 (2007) 사회 연결망 이론, 박영사, 서울시 금천구, pp.48.