
SNA in R 실습과정 분석 보고서

201401137 중어중문학과 김성수

201611539 통계학과 하성진



I. 개요

경부선의 주요 역(서울역, 부산역)을 제외한 허브역을 선정하기 위해 한국철도공사 역 별 승 하차 현황을 Social Network Analysis 기법을 이용하여 분석하였다. 사회연결망분석은 개인들 간의 상관관계를 통해 인간행위와 사회구조의 효과를 설명하고자 하는 방법론이다. 사회연결망 이론에서는 사람이나 조직간 관계를 노드(node)와 두 노드를 연결하는 선(edie)의 네트워크 형태로 표현하여 체계의 특성이나 행위를 분석한다[1]. 관계의 주체가 되는 행위자들은 Node로, 관계들은 Node사이를 연결하는 Edge로 나타낼 수 있으며, 중심성은 그래프 상에서 어떤 Node가 가장 중요한지를 살피는 척도이다. 이번 실습에서 사용한 각 측정지표의 정의 및 특성은 다음과 같다.

-연결정도 중심성(degree centrality)

연결정도 중심성은 한 Node에 연결된 모든 Egde의 개수(이 실습에서는 Weight의 합)로 중심성을 평가한다. 즉, 한 Node에 얼마나 많은 Egde들이 관계를 맺고 있는지를 기준으로, 노드의 중심 정도를 계량화 한 것이다.

-위세중심성(eigen centrality)

자신의 연결 정도 중심성으로부터 발생하는 영향력과 자신과 연결된 타인의 영향력을 합하여 결정한다. 즉, 연결된 노드의 중요성에 가중치를 뒤 노드의 중심성을 측정하는 방법이다.

-근접 중심성(closeness centrality)

각 노드 간의 거리를 근거로 중심성을 측정하는 방법으로, 연결 정도 중심성과는 달리 직접적으로 연결된 노드 뿐만 아니라 간접적으로 연결된 모든 노드 간의 거리를 합산해 중심성을 측정한 방법이다.

II. 분석 절차

1. 데이터

본 보고서에서 사용된 데이터는 공공데이터포털(<https://www.data.go.kr>)으로부터 수집하였으며, '한국철도공사_역 별 승 하차 현황'의 데이터셋을 이용하였다. 경부선에서의 데이터를 수집하기 위해서 경부선이 아닌 역은 제외하였으며, 추출된 데이터는 From to 데이터의 형식을 갖추어야 하기 때문에 아래와 같이 그림1과 같이 전처리하였다. From열은 링크의 시작 점인 'Source Node'가 되고, to가 위치한 열은 링크의 도착 점인 'Target Node', total은 링크의 가중치인 weight가 된다. 결과적으로 분석에 활용된 역은 서울, 동대구, 부산, 대전, 영등포, 천안아산, 수원, 구미, 평택, 구포, 경산, 조치원, 밀양, 김천, 왜관, 영동으로 총 16개 역의 승 하차 데이터를 이용하였다.

A	B	C
from	to	total
서울	영등포	121705
서울	수원	1335737
서울	평택	686491
서울	천안아산	1895049
서울	조치원	234574
서울	대전	2900921
서울	영동	67336
서울	김천	57260
서울	구미	478864

그림 1

2. 실습 절차

대략적인 실습의 절차는 다음과 같다. 수집된 승 하차 데이터를 이용하여 EDA분석을 수행하고, 그 결과로부터 승 하차 이동량으로부터 중요한 역을 정의한다. 그 이후, 사회관계망 분석을 수행하여 기존에 정의한 중요한 역과 사회관계망 분석의 다양한 측정지표와 비교하면서 그 의미를 찾아보고 중요한 역을 제외한 허브역을 도출한다. 이 실습에서는 각 철도역이 하나의 노드를 구성하고, 엣지를 이동경로로 구성하였다.

III. 분석 결과

1. EDA 분석

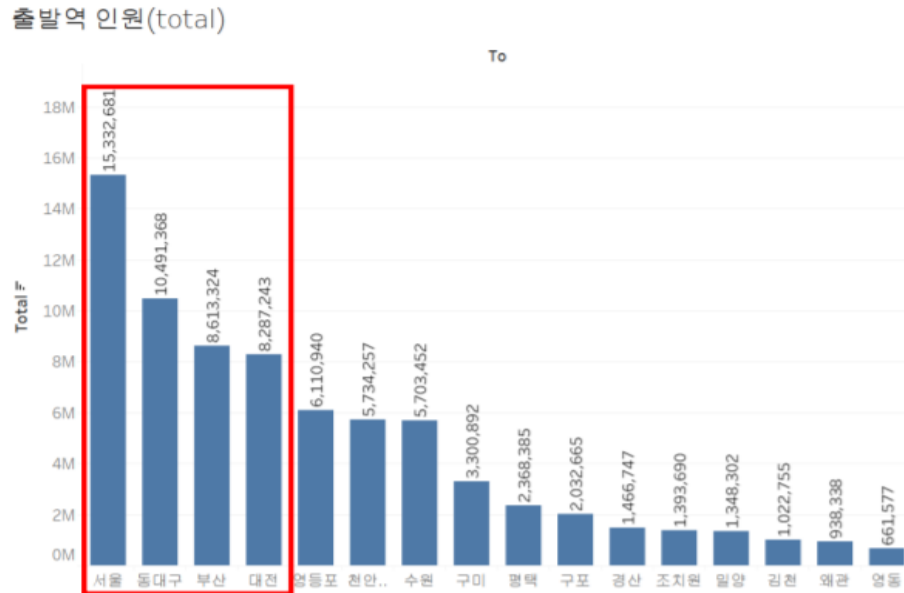


그림 2

그림 2는 경부선의 각 역에서 얼마나 많이 상차했는지를 나타내고 있다. 예상과 같이, 서울역, 동대구역, 부산역, 대전역에서 많은 사람이 상차를 하며 이는 하차인원의 총합을 보여주는 아래의 그림3의 순위와 일치한다.

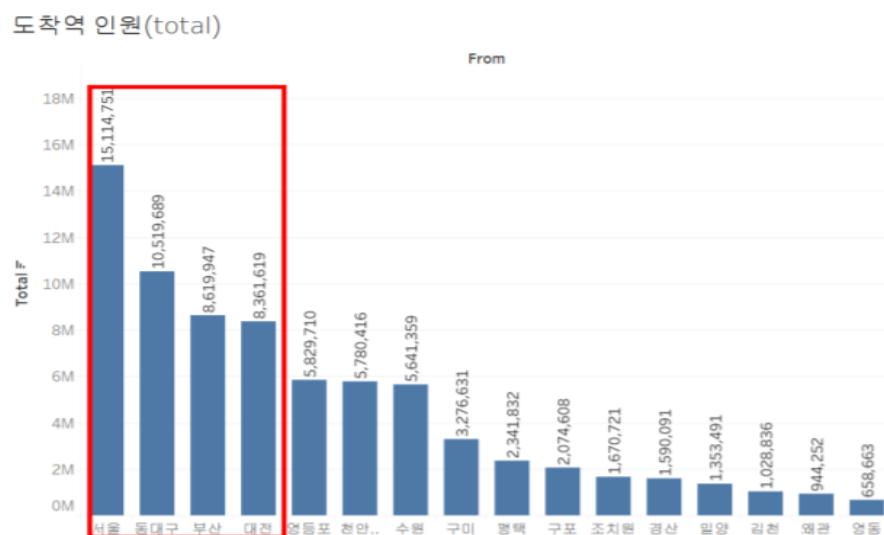


그림 3

그림 2와 그림 3을 통해 상하차가 가장 많이 일어나는 서울, 동대구, 부산, 대전을 경부선에서 가장 중요한 역으로 정의할 수 있다.

2. 사회연결망 분석

(1) 연결정도 중심성(degree centrality)

Rank	역명(Station)	연결정도 중심성(degree centrality)
1	서울	15114751
2	동대구	10519689
3	부산	8619947
4	대전	8361619
5	영등포	5829710
6	천안아산	5780416
7	수원	5641359
8	구미	3276631
9	평택	2341832
10	구포	2074608
11	조치원	1670721
12	경산	1590091
13	밀양	1353491
14	김천	1028836
15	왜관	944252
16	영동	658663

Table 1 Result of degree centrality

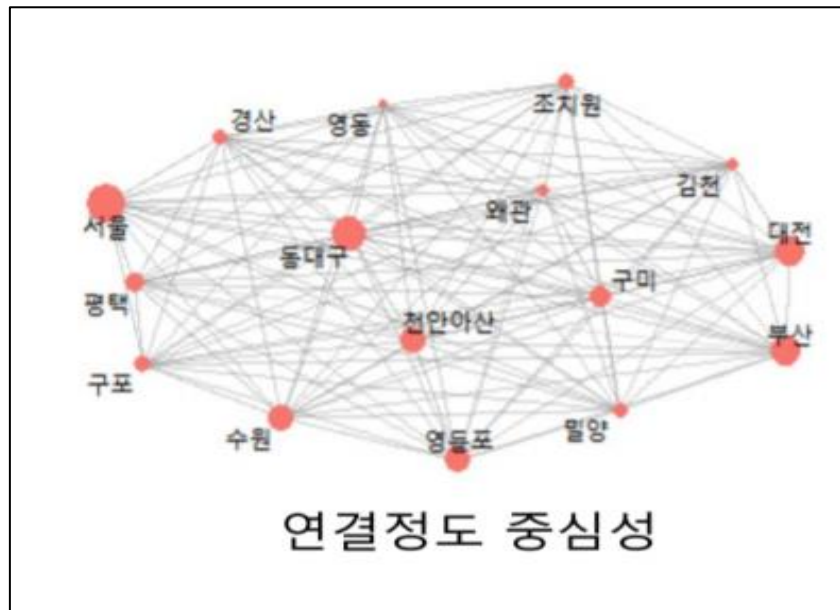


그림 4. 연결정도 중심성 경부선 연결망

다른 역과 연결되는 빈도수(weight 가중치)로 측정하며, 연결정도 중심성으로 측정했을 때 역의 순위는 그림 2,3과 같이 단순히 얼마나 승하차를 했는지를 나타내는 순위와 비슷하다는 것을 알 수 있다.

(2) 위세중심성(eigen centrality)

Rank	역명(Station)	위세중심성(centrality_eigen)
1	서울	1
2	부산	0.711
3	동대구	0.676
4	대전	0.614
5	천안아산	0.435
6	수원	0.383
7	영등포	0.346
8	구미	0.235

9	평택	0.161
10	구포	0.126
11	조치원	0.0962
12	경산	0.0930
13	밀양	0.0815
14	왜관	0.0627
15	김천	0.0616
16	영동	0.0441

Table 2 Result of eigen centrality

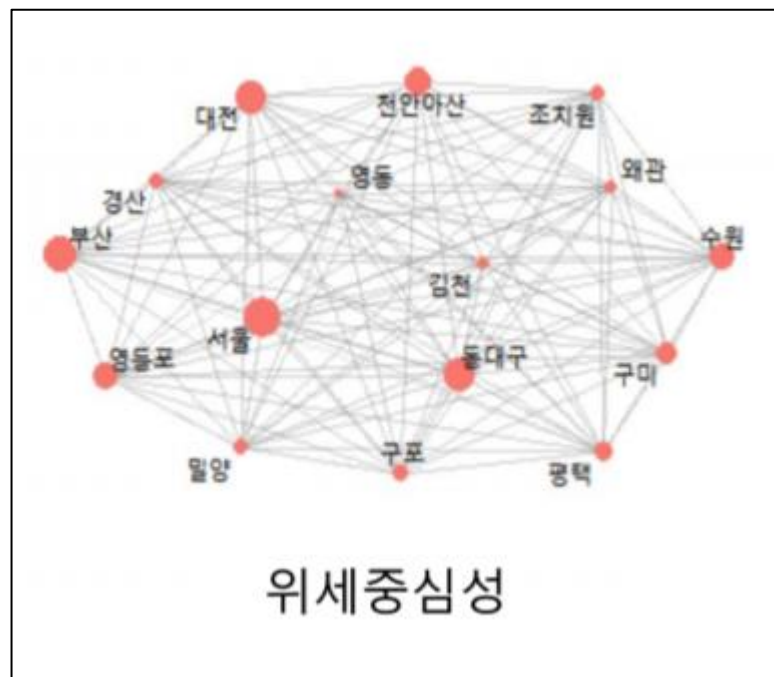


그림5. 위세 중심성 경북선 연결망

연결정도 중심성이 높은 다른 역과 연결이 많은 경우 위세 중심성이 상승하기 때문에 연결빈도가 낮더라도 상대적으로 중요한 역들과 연결이 되어있는 역일수록 순위가 높아질 것이다. 표2의 순위와 표1의 순위에서 눈여겨봐야 할 사실은 천안아산의 순위가 상승했다는 것이다. 이를 통해 천안아산역이 중요한 역과 연결이 되어있음을 알 수 있다.

(3) 근접 중심성(closeness centrality)

Rank	역명(Station)	근접 중심성(closeness centrality)
1	조치원	0.00000509
2	밀양	0.00000491
3	왜관	0.00000469
4	영등	0.00000466
5	경산	0.00000435
6	평택	0.00000434
7	김천	0.00000383
8	천안아산	0.00000353
9	구포	0.00000345
10	수원	0.00000269
11	영등포	0.00000254
12	부산	0.00000248
13	구미	0.00000245
14	서울	0.00000192
15	대전	0.000000185
16	동대구	0.00000181

Table 3 Result of eigen centrality

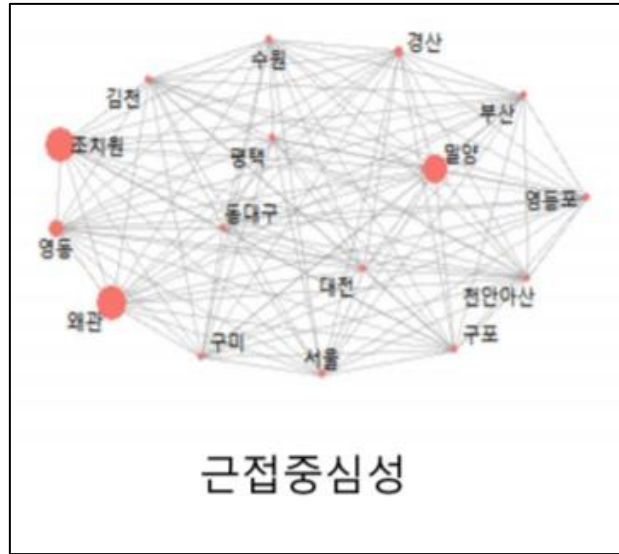


그림6. 근접 중심성 경부선 연결망

근접 중심성은 중요한 노드일수록 다른 노드까지 도달하는 경로가 짧을 것이라는 가정을 두어 계산하는 방식으로, 앞에서 계산했던 연결정도 중심성, 위세 중심성의 결과와 반대결과가 나왔다. 따라서 가장 많은 인구가 지나는 서울과 대전 동대구를 지나며, 이 역과 가장 가까운 역인 조치원, 왜관, 밀양 등의 순위가 높아졌다고 예상할 수 있다. 즉, 유의미한 분석이 아니라고 판단했다.

IV. 실습 결론

본 실습에서는 경부선에 있는 역을 대상으로 승 하차 데이터를 이용하여 사회 연결망 분석을 수행하고, 그 결과로 도출된 결과들을 통해 허브역을 도출하고자 하였다. 연결정도 중심성, 근접 중심성, 위세 중심성을 지표로 이루어졌으며, 연결 빈도가 상대적으로 낮더라도 EDA 분석을 통해 분석한 중요한 역(서울, 부산)과 연결이 가장 많은 천안 아산역이 경부선에서 허브 역 역할을 하는 것을 알 수 있었다.

해당 실습을 바탕으로 제안할 수 있는 방향은 다음과 같다. 허브 역인 천안아산 주변에 공간분석을 활용하여 주변 교통시설, 역사 내 시설을 통해 주요역의 인프라에 비해 부족한 천안아산역이 허브역으로 성장할 수 있도록 문제점을 분석하고 개선할 수 있을 것이다. 또 다양한 경부선 관련 데이터를 활용하여 경부선에 대한 다양한 주제에 대한 분석을 할 수 있을 것이다.