

Network Analysis of Seoul Subway Network

이예준 최수환

2021년 1학기 신입생세미나 중간에세이

Table of Contents

1. Introduction

2. Preprocessing Network Data

3. Centrality Analysis

A. Degree Centrality

B. Closeness Centrality

C. Betweenness Centrality

D. Eigenvector Centrality

E. More on eigenvector centrality

References

1. Introduction

수도권 지하철은 수도권 지역의 핵심 교통수단으로서 자리매김하고 있다. 본 에세이에서는 수도권 지하철 네트워크의 중심성(centrality)을 다양한 방법을 통해 분석한다. (예정: 나아가, 지하철 승하차객 수 데이터와 연관 지어 네트워크 분석을 진행한다.)

https://github.com/Wittgensteinian/Seoul_Metro에서 본 에세이에 관한 모든 코드를 찾아볼 수 있다.

2. Preprocessing Network Data

수도권 지하철 네트워크 데이터는 <http://gangwon.github.io/subway-data/>로부터 받아왔다. 이 데이터는 수도권 지하철의 역명과 역번호, 각 역 간의 이동 시간에 대한 정보를 담고 있다. 이 데이터를 바탕으로, 두 가지 다른 그래프를 만들어 분석에 활용했다. 하나는 역명을 노드로, 하나는 역번호를 노드로 삼는 그래프다.

역명을 노드로 삼는 그래프의 경우, 우리가 떠올리는 지하철 노선도와 가장 유사한 그래프 구조를 생성할 수 있다는 장점이 있다. 그러나 역과 역 사이의 최단거리(정확히는 최단시간)를 구하기가 복잡하다는 단점이 있다. 구체적으로 이야기하자면, 역명을 노드로 삼는 그래프에서 계산하는 최단 거리는 노선 간 환승 시간을 전혀 고려하지 못한다는 뜻이다. 예를 들면, 이 그래프에서는 2호선 서울대입구역부터 2호선 방배역까지의 거리와 4호선 남태령역까지의 거리가 동일하다(모든 노드 사이의 거리를 1이라 가정). 그러나 이는 환승 시간을 고려하지 않은 결과로, 실제 이동 시간과는 괴리가 있다.



그림 1: 사당역과 그 주변 역들

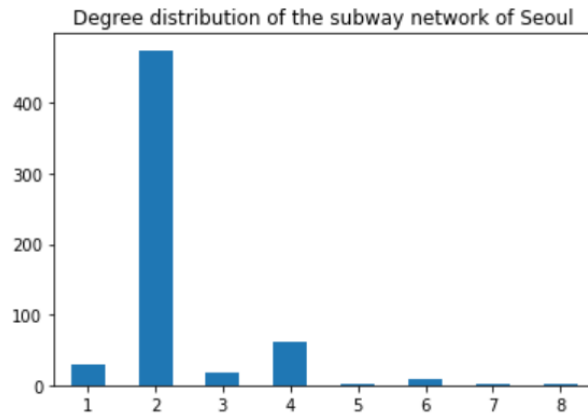
따라서 최단거리 계산을 위해 역번호를 노드로 삼는 그래프를 만들었다. 역번호를 노드로 삼는 그래프의 경우, 같은 역일지라도 노선이 다른 역(환승역)은 여러 개의 다른 노드로 취급한다. 같은 역이여도 노선이 다르다면 역번호 역시 다르기 때문이다. 예를 들면 2호선 사당역은 노드 '226'으로, 4호선 사당역은 노드 '433'으로 다르게 취급된다. 두 노드 사이의 거리에는 일괄적으로 '3'을 부여했다. 환승 시간을 3분으로 설정한 것과 같은 의미다.

본 에세이에서는 지하철 네트워크의 구조를 보존해야 하는 경우 역명을 기준으로, 최단거리를 계산해야 하는 경우 역번호를 기준으로 하는 그래프를 활용했다. 또한 거리를 계산해야 하는 경우 모든 엣지(edge)의 거리에 1을 부여하는 네트워크 거리(network distance)와 실제 시간(physical time)을 같이 활용했다.

3. Centrality Analysis

A. Degree Centrality

Degree란, 해당 노드와 인접한 노드의 개수를 의미한다. 역명을 기준으로 하는 그래프에서 총 노드의 개수는 598개이며, degree distribution은 다음과 같다. Degree centrality가 가장 높은 노드(degree = 8)는 공덕역이다. 두번째로 높은 노드(degree = 7)는 왕십리역이다.



B. Closeness Centrality

$$c_{Cl}(v) = \frac{1}{\sum_{u \in V} \text{dist}(v, u)}$$

Closeness centrality는 해당 노드에서 다른 노드까지의 거리의 합에 반비례하는 값으로 주어진다. 네트워크 거리를 따를 때는 이촌역(경의중앙선), 실제 시간을 따를 때는 서울역(4호선)이 제일 높은 centrality를 가진다.

Closeness centrality는 eccentricity를 활용하는 방법으로도 변형될 수 있다. Eccentricity란, 해당 노드에서 다른 노드까지의 최단거리의 최댓값이다. 따라서 closeness centrality를 eccentricity에 반비례하는 값으로 설정할 수 있다. 네트워크 거리를 따를 때는 노량진(1호선), 동작(4호선), 충신대입구(이수)(4호선), 남태령(4호선), 실제 시간을 따를 때는 충신대입구(이수)(4호선)가 제일 높은 centrality를 가진다.

Eccentricity의 최댓값을 그래프의 지름(diameter), 최솟값을 그래프의 반지름(radius)이라고 부른다. 네트워크 거리를 따를 때는 지름이 75, 반지름이 39이며, 실제 시간을 따를 때는 지름이 242, 반지름이 121이다.

C. Betweenness Centrality

$$c_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma(s, t | v)}{\sigma(s, t)}$$

Betweenness centrality는 두 노드를 잇는 최단 거리 중 해당 노드를

D. Eigenvector Centrality

E. More on eigenvector centrality

첫번째로, centrality가 공항철도의 역 혹은 공항철도와 환승되는 역에서

그렇다면 공항철도의 역들 중에서도 서울의 외곽에 위치한 김포공항의 centrality가 높게 나타나는 이유가 무엇일까?

Reference