

**붙임 4****[응모분야1] 데이터분석 보고서 작성방법 및 서식**

접수번호

※작성하지 않음

**「2025년 통계데이터 활용대회」 데이터분석 보고서**

제 목

네트워크 분석을 통한 생활권 구분 방법 제안

신 청 자 명

소속/직위

서울대학교/대학원생

성 명

정우진

휴대전화

010-3107-0752

전자우편

cowzin@snu.ac.kr

제출일

2025.07.01

# 〈네트워크 분석을 통한 생활권 구분 방법 제안〉

## 1. 배경

### □ 주제 선정

생활권이란 통근·통학·쇼핑·여가·친교·업무·공공서비스 등 주민들의 일상적인 생활 활동이 이루어지는 공간범위를 의미한다(국토연구원, 2022). 따라서 생활권 구분은, 지역 별 미래상과 발전전략, 관리방향 설정 등의 도시 계획의 측면에 있어서 중요한 기준이 되며, 서울시 등 각 지자체는 도시 내의 생활권을 나누어 추후 정책 수립의 기준으로 삼고 있다. 본 연구에서는, 네트워크 데이터 분석을 통하여 생활권을 구분하는 방법을 제시하고, 이를 서울시 인구 이주 데이터에 적용하여 그 실효성을 확인하고자 한다.

### □ 분석 필요성(문제점) 및 전략

상술한 생활권 구분의 중요성에 따라, 서울과 부산 등 지자체는 자체적으로 도시 내 생활권을 구분하였다. 생활권 구분에 대한 심도있는 분석이 이루어진 ‘2030 서울생활권계획’에 따르면, 해당 분석은 각계 각층의 전문가와 주민들이 참여하여 이루어졌으며, 지역 별 특성과 연계성을 심도있게 분석함으로서 이루어졌다. 우리는 다소 다르게, 인구이주 데이터의 출발지(구)와 도착지(구)를 노드(Node)로 가지고, 엣지(Edge)의 가중치로 인구 이주 수를 가지는 네트워크 데이터로 생각한다. 이를 네트워크 관점에서 해석함으로써, 우리는 각 서울 내 구간의 관계를 고려한 분석을 수행 할 수 있게 된다. 본 연구에서는, 네트워크 모델 기반의 생활권 구분 방법을 제시하고, 이를 서울시 데이터에 적용하여 본다. 더 나아가, 기존 생활권 분석에서 아직 제시하지 못하였던 ‘연령대별’ 생활권을 구분하고자 하며, 이 중 대표적으로 서울시 청년인구의 생활권을 제시한다. 제시된 방법론을 기존 구분 생활권과 비교함으로써 상호보완 목적으로 사용할 수 있고, 또한 매우 간단하기 때문에, 새로운 생활권 구분이나 인구별 생활권을 조사할 때 프로토타입으로 사용할 수 있다.

## 2. 데이터 분석

### □ 데이터 선정과 전처리

#### - 데이터 선정

본 연구에서는 MDIS에서 제공하는 2019-2024년 “국내인구이동통계-인구관련 연간자료”를 사용하였다. 이 데이터는 <표 1>과 같이, 연도 별로 대한민국 내 각 기초자치단체(읍면동)별로 세대 별 인구 유입과 인구 유출에 관련한 정보를

조사한 데이터로, 각 행은 하나의 세대에 해당한다.

〈표 5 : 국내인구이동통계-인구관련연간자료 변수명〉			
구분	항목명	구분	항목명
코드	전입행정구역_시도코드	코드	전출행정구역_읍면동코드
코드	전입행정구역_시군구코드	코드	전입사유코드
코드	전입행정구역_읍면동코드	코드	전입자1_세대주관계코드
문자	전입연도	숫자	전입자1_만연령
문자	전입월	코드	전입자1_성별코드
문자	전입일	코드	전입자2_세대주관계코드
코드	전출행정구역_시도코드	숫자	전입자2_만연령
코드	전출행정구역_시군구코드	코드	전입자2_성별코드
코드	전입자(3-10) 만연령	코드	... 그 외(전입자 3-10)

고다현(2023)의 분석에 따르면, 서울시 내 인구이주의 대부분은 기존 생활권을 크게 벗어나지 않는 범위 내에서 진행되었다. 우리는 이 사실을 이용하여, 인구이주 데이터를 서울 내 구의 생활권 구분 방법을 제안하고자 한다.

#### - 데이터 전처리

원 데이터를 바탕으로 연도별 서울시 내 인구 이주 데이터를 형성하기 위해, 〈표 1〉의 데이터에서 서울 내부 간에 이주한 데이터만을 모은 후, 〈표 2〉와 같이 추가적인 전처리를 진행하여, 노드와 엣지를 가진 네트워크 데이터 형식으로 변형하였다. 이를 통해 산출한 결과는 표 2에 나타나 있다. 추가적으로, 네트워크의 구조적 해석을 용이하게 하고, 노드들 간의 상호작용에 집중하고자 동일 구 내에서 이주한 데이터는 제외하였다(S. Wasserman and K. Faust 1994).

〈표 6 : 전처리 후 데이터 변수 설명〉		
구분	항목명	설명
코드	출발구	전출행정구역_시군구코드와 동일(서울시 구)
코드	도착구	전입행정구역_시군구코드와 동일(서울시 구)
코드	이주 수	$E_{pq} = \sum_{i=1}^N I(M_i \in G_{pq})w_i$ <p> <math>E_{pq}</math> : p에서 q로 이주한 인원 수.  <math>N</math> : 연도 별 총 서울 내 세대 이주 건수.  <math>M_i</math> : i 번째 세대 이주 사건.  <math>G_{pq}</math> : 출발지가 p이고 도착지가 q인 세대 별 이주 사건들의 집합  <math>w_i</math> : i번째 세대의 총 인원 수 </p>

### □ 데이터 분석(분석 프로세스, 분석방법 등)

#### - 개요

데이터 전처리를 통하여 얻어진 네트워크 데이터에 대해 Stochastic Block Model을 통하여 각 노드(구) 들에 대한 군집을 형성하고, 각 군집을 생활권으로 해석하고자 하였다.

#### - 모델 설명

Stochastic Block Model(SBM)은 네트워크 데이터를 분석 할 때 노드들을 여러 개의 군집으로 분류하기 위한 클러스터링 모델이다. 이 모델은 같은 군집에 속한 노드들끼리는 높은 확률로 연결되고, 다른 군집에 속한 노드 간에는 적은 확률로 연결된다고 가정한다. 구체적으로, 총 N개의 노드와 K개의 군집이 있다고 가정하자. 각 노드는 1부터 K까지의 군집에 속하게 된다.  $\theta$ 는 군집 간의 연결확률을 나타내는 행렬로, 각 원소는 두 군집 소속 노드간의 연결확률을 나타낸다. A는 관찰된 인접행렬로, 두 노드 사이에 엣지가 존재하는가의 여부를 나타내고, 기본적인 SBM 하에서는 0 또는 1을 가지는 NxN 행렬이 된다. SBM은 다음 모델을 가정한다.

$$1) P(A_{ij} = 1 | z_i = k, z_j = l) = \theta_{kl}$$

이 모델 하에서, 주어진 데이터에 대한 우도 혹은 사후확률을 최대화 시키는 군집 구조를 찾는 것이 모델의 목적이다. 모델을 통해서 구해진 군집이 구해졌다면, 군집에 포함된 노드(구) 사이에 상호작용이 크다고 해석 할 수 있다.

한편, 우리의 분석에서는, 각 엣지(이주 수)는 Binary 값을 가지지 않고, 이주인구 수를 나타내는 Count Data로 구성되며, 출발지와 도착지가 2)구분되는 Directed Weighted Network 형식을 지닌다. 따라서 우리는 각 엣지가 3)조건부 포아송분포를 따른다고 가정하고, 방향성이 있는 네트워크를 처리 할 수 있는 수정된 SBM을 사용한다. 분석을 위해서는 R 4.3.2와 ‘blockmodel’ 패키지의 BM\_poisson 함수를 사용하였다.

## - 분석 방법

### 1. 엣지(이주 수) 가중치 수정.

앞의 <표 2>에서 언급한 바와 같이, 우리는 분석 대상 데이터를 출발지와 도착지를 노드로 가지고, 이주인구 수를 엣지의 가중치로 두는 네트워크를 고려한다. 만약 이 네트워크의 엣지를 그대로 사용한다면, SBM을 통한 군집결과 단순히 거주 인구가 많은 노드들이 같은 군집으로 분류될 가능성이 있다. 따라서, 우리는 원 네트워크 데이터의 엣지에 다음의 <표 3>과 같은 보정을 적용하였다. 이러한 엣지 가중치 보정절차를 통하여, 군집에 있어 원 거주 인구의 영향을 배제하고자 하였다. 이러한 가중치 보정절차는 추후 분석에서 개선될 여지가 있다.

1) 즉 군집 소속이 주어졌을 때, 엣지의 연결은  $\theta$ 에 의존하는 베르누이 분포를 가진다고 가정한다.

2) 예를 들어, 강남구 -> 송파구 이주와 송파구 -> 강남구 이주 가중치는 다른 값을 가진다.

3) Count data는 일반적으로 포아송분포를 통하여 모델링 할 수 있다.

〈표 3 : 엣지 가중치 보정 절차〉	
1	기존의 전처리된 데이터의 인접행렬 A의 i번째 행 j번째 열의 원소는 i번째 구에서 j번째 구로 이동하는 이주 인구 수를 나타낸다.
2	A의 i번째 행에 속한 모든 원소를, i번째 구의 총 인구로 나누었다. 이러한 변환을 적용하면, i번째 행 j번째 열의 원소는 i번째에 사는 어떠한 사람이 해당 년에 j번째 구로 이주했을 확률을 나타낸다고 볼 수 있다. (각 행의 합은 해당 구 거주 인구가 다른 구로 이주했을 확률을 나타낸다.)
3	2.에서 구한 엣지의 가중치는 0~1 사이의 값을 가진다. 우리는 2.에서 산출한 행렬에 1000을 곱한 후 각 원소를 반올림하여 각 원소를 정수로 가지는 수정된 인접행렬을 만든 후, 가중치를 Count Data로 간주하였다. 이러한 경우, 모델설명 부분에서 언급하였던 BM_poisson 모델을 그대로 사용할 수 있다.

## 2. Stochastic Block Model(SBM) 적합.

1에서 보정된 엣지 인접행렬을 기반으로, 앞에서 언급하였던 SBM을 통해 서울시 생활권을 나눠보고자 하였다. 2019년부터 2024년까지의 서울시 인구이주 데이터를 각각 적합하였다. 이에서 더 나아가, 전체 연령 인구이주 데이터가 아닌, 청년(19-34세)들의 이주 데이터를 기반으로 모델을 적합하여 청년인구의 생활권 구분 역시 진행하였다.

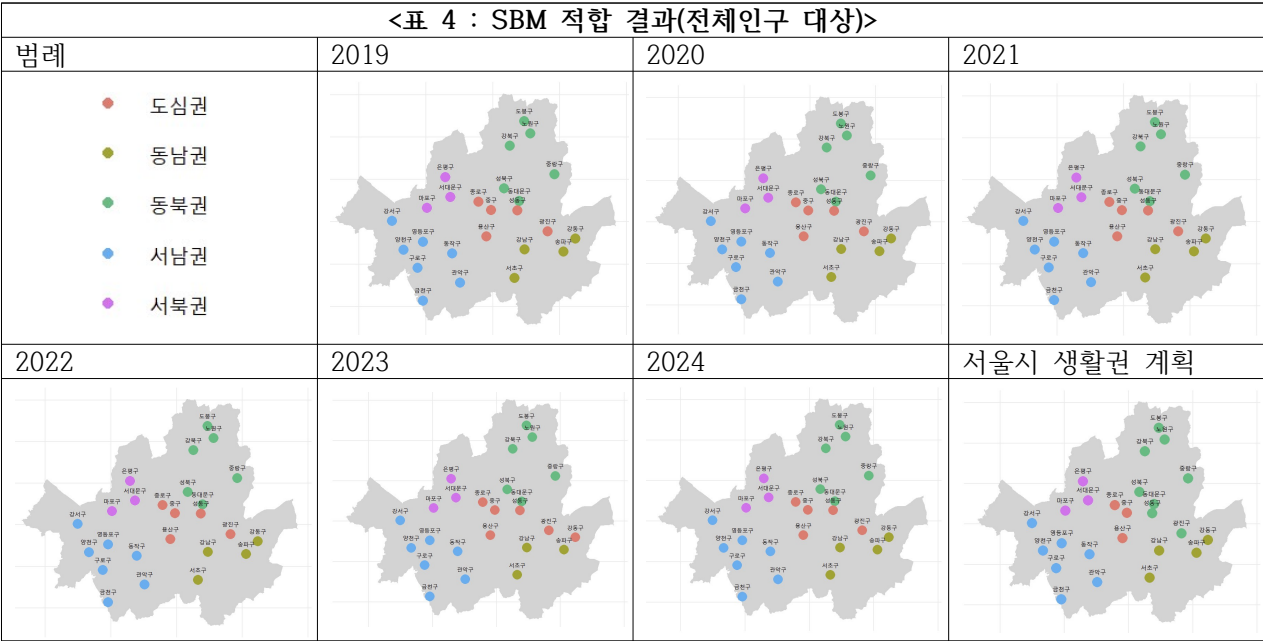
한편, SBM은 모델 적합 전 지정이 필요한 하이퍼파라미터로, 노드의 군집 수 K를 가진다. 일반적으로 K를 지정하는 데 사용되는 기준은 ICL(Integrated Completed Likelihood)로, 이는 각 K에서의 최적의 군집할당에 대한 우도(likelihood)에 군집 수에 대한 벌점항(penalty)가 더해진 형태이다. 우리는 연도별로 ICL에 대한 <sup>4)</sup>elbow point를 계산하여, 이를 K로 사용하였다. K는 분석대상 모든 연도에 대하여 5로 산출되었다.

### □ 분석 결과 및 해석

#### - 전체 인구이주 데이터에 대한 적합

〈표 4〉는 전체 인구 이주데이터에 대한 모델 나타낸 것이다. 편의상 범례는 ‘서울 생활권 계획’에서 사용하였던 범례를 그대로 사용하였다. 적합결과를 보면, SBM 적합결과는 2023년에 강동구가 동남권이 아닌 도심권으로 분류되는 것을 제외하면 일관적인 결과를 띈다. SBM 결과와 서울 생활권 계획의 분류를 비교하면, 서남권, 동남권, 서북권에 속하는 지역은 동일하다. 한편, 성동구와 광진구의 경우 SBM에 따르면 도심권으로 분류되고, 서울시 생활권 계획은 동북권으로 분류하고 있다.

4) Elbow point는 kneedle algorithm으로 구하였는데, 이는 기준 직선으로부터 제일 거리가 먼 점을 elbow point로 삼는 방법이다.



**- 전체 인구이주 데이터에 대한 클러스터 분석**

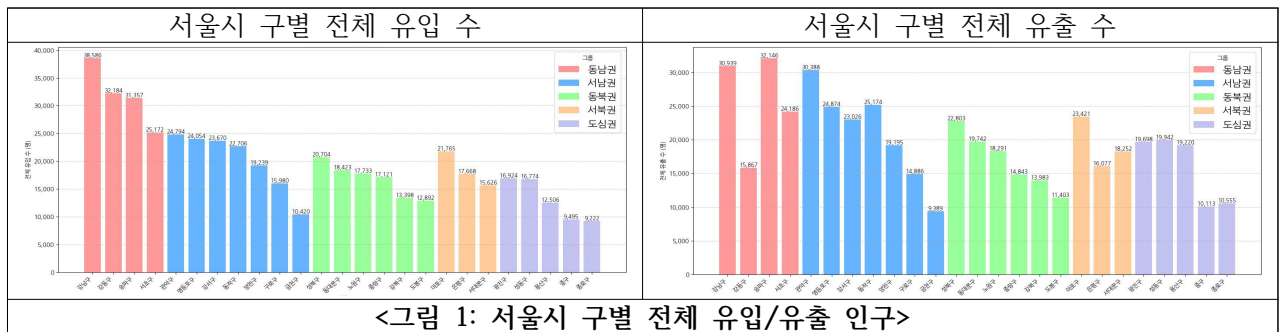
<표 4>의 결과를 바탕으로 나뉘어진 클러스터 결과는 아래의 <표5>와 같다. 이러한 클러스터별로 유입 및 유출인구, 네트워크 중심성 및 사회적 특성을 분석하였다.

<div> <b>&lt;표 5 : SBM 적합에 따른 서울 구별 클러스터 분류&gt;</b> </div>	
동남권	강남구 강동구 서초구 송파구
서남권	강서구 관악구 구로구 금천구 동작구 양천구 영등포구
동북권	강북구 노원구 도봉구 동대문구 성북구 중랑구
서북권	마포구 서대문구 은평구
도심권	광진구* 성동구* 용산구 중구 종로구
* 서울시 생활권 계획에서 서북권에 해당하는 구	

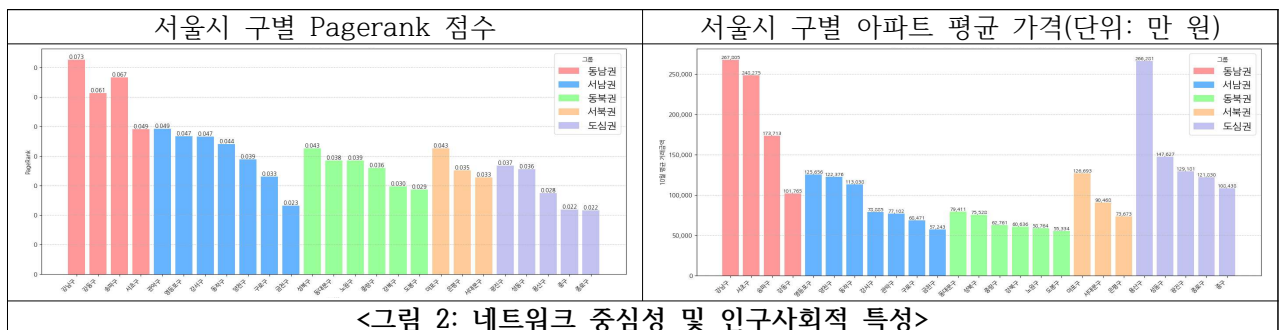
먼저 유입 및 유출인구의 경우 <그림 1>과 같이 전체 유입 및 유출인구를 분석하였으며, 강남구를 선두로 동남권의 유입인구가 다른 클러스터에 비해 유의미하게 높았다. 한편 유출 인구의 경우 유입인구에 비해 클러스터 별 차이가 적었으며, 서남권의 유출 인구가 많은 것이 눈에 띄었다.

중심성 및 인구사회적 특성을 분석한 결과는 <그림 2>와 같이 나타났다. 먼저 네트워크 중심성 지표 중 이주 인원 수를 반영할 수 있는 PageRank (S. Brin and L. Page, 1998) 지표를 살펴본 결과, 동남권에서 중심성이 유의미하게 높게 나타났으며, 외곽 지역으로 갈수록 점수가 낮아지는 경향을 보여 지리적 특성의 중요성을 확인할 수 있었다.

아파트 가격의 경우 도심권 중에서도 용산구의 아파트 가격이 유의미하게 높게 나타났는데, 이는 도심지역의 특성을 반영한 자연스러운 결과라 할 수 있다.



그러나 아파트 가격이 높음에도 불구하고 네트워크 중심성은 낮은 것으로 나타나, 용산구는 단순히 높은 가격만으로 이주 매력도가 높은 지역으로 평가하기는 어려웠다. 또한 동남권의 인구 과밀 현상은 이주 흐름을 통해서도 뚜렷하게 확인할 수 있었다.



한편 모델 적합결과에 대하여, 단순히 인접한 지역끼리 인구이주가 5) 많기 때문에 모델이 우연하게 주변 지역을 클러스터로 묶은 것이라고 비판 할 수 있다. 하지만, 모델의 적합결과는 연도와 관계없이 거의 유사한 양상을 보이며, 이는 거리 이외에 각 군집을 형성시키는 요인을 모델이 반영한 것이라고 해석 할 수 있다.

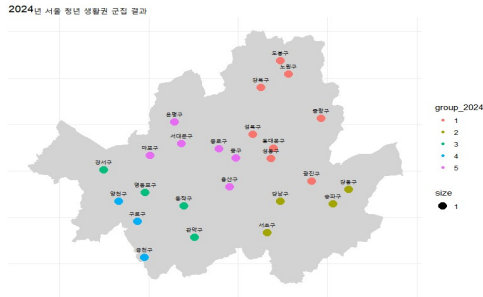
- 청년 인구이주 데이터에 대한 적합

기존의 생활권 구분은, 전체 인구만을 대상으로 했고 각 인구 특성 별로 달라지는 생활권을 제시하지 못하였다는 한계점이 있다. 이에 본 연구에서는 서울시 인구이주 데이터 중 청년(19~34세) 데이터만을 대상으로 상술한 모델을 적합시켜 청년인구에 대한 생활권 구분을 제시하고자 한다. <그림 3>은 6)청년인구를 대상으로 적합한 결과이며, 이를 바탕으로 나뉜 클러스터 결과는 <표 7>과 같다.

5) 인구이동을 설명하는 대표적 모델인 Gravity Model은 이주 유입량이 거리에 반비례한다고 가정한다.

6) 2024년 청년인구 데이터에 SBM을 적용할 때의 ICL의 elbow point는 5이므로, 5개의 군집으로 나누었다.



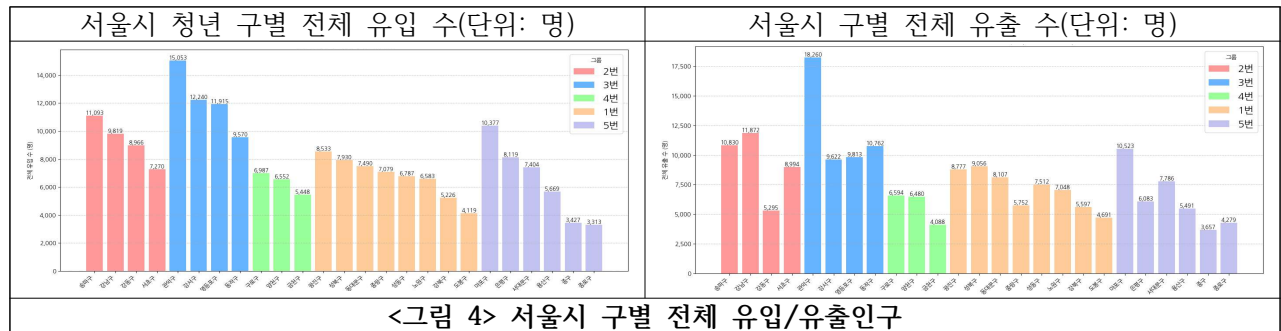


<그림 3: 24년 청년인구 이주 데이터 적합 결과>

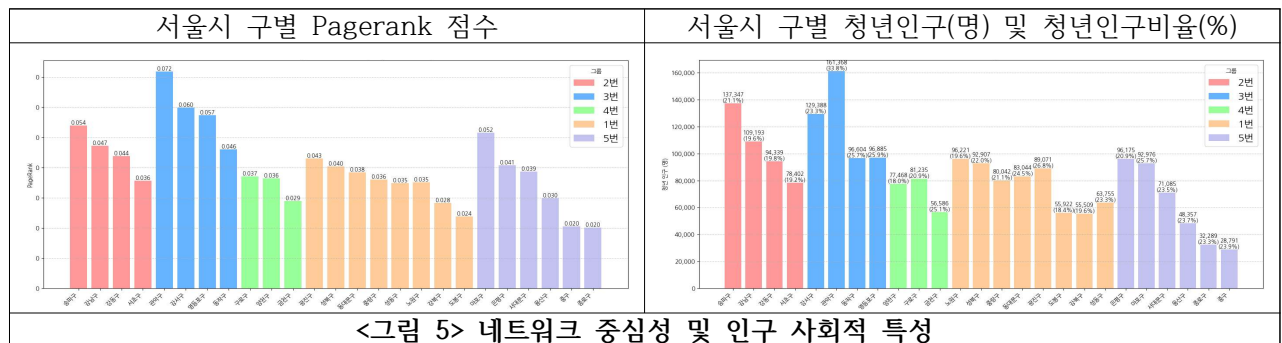
〈표 7: SBM 적합에 따른 서울 청년인구 구별 클러스터 분류〉	
1번 그룹	강북구 노원구 도봉구 동대문구 성북구 중랑구
2번 그룹	강남구 강동구 서초구 송파구
3번 그룹	강서구 관악구 동작구 영등포구
4번 그룹	구로구 금천구 양천구
5번 그룹	광진구 마포구 서대문구 성동구 용산구 은평구 중구 종로구

## - 청년인구 이주 데이터에 대한 클러스터 분석

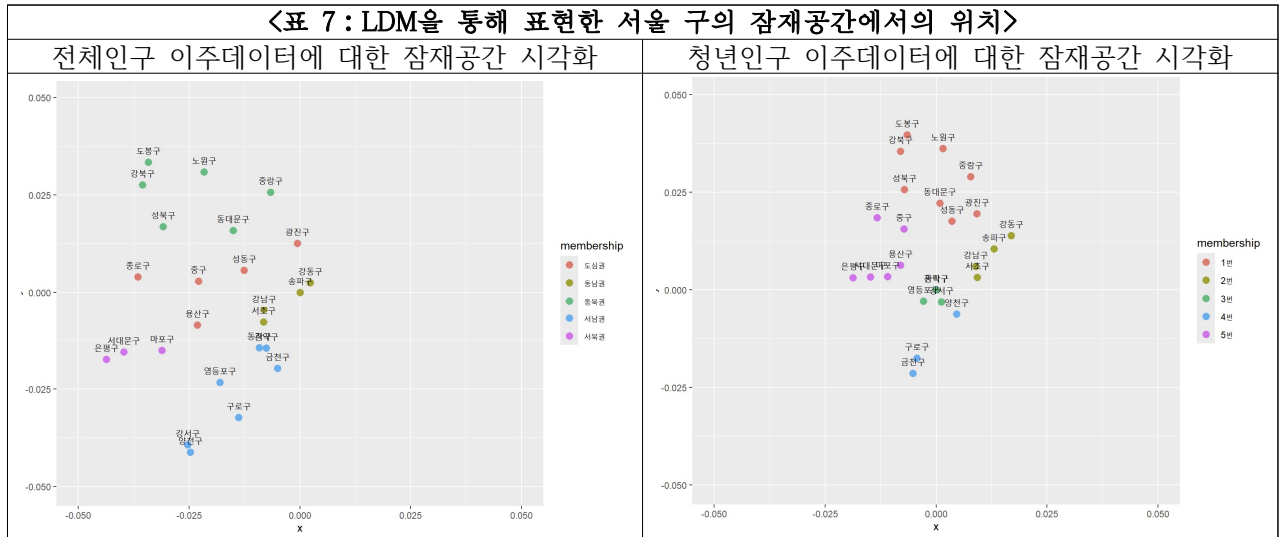
서울시 전체 인구 이주와는 달리 청년인구 이주의 경우 <그림 4>와 같이 3번 그룹의 유입과 유출인구의 양이 가장 많았다. 특히 관악구의 경우 유입은 물론 유출인구가 다른 구에 비해 유난히 높았다. 한편 4번 그룹의 경우는 유입과 유출인구가 가장 적은 범위에 들어가 서남권에서 4번 그룹과 구분된 것이 합리적 으로 보인다. 이는 <그림 5>의 청년인구와 청년인구 비율이 높은 3번 그룹의 특성을 잘 나타내고 있는 것으로 보인다.



PageRank 지표 또한 3번 그룹이 가장 높은 것으로 나타나, 청년 인구에게 매력적인 입주 지역으로 나타난다. 2번 그룹의 경우 동남권 생활권으로 평균보다 높은 유입 및 유출 수를 보이고 있으며, 중심성 지표와 인구수에서도 비슷한 경향을 보인다. 반면 1, 3, 5번 그룹은 청년인구의 수와 비율도 낮으며, 중심성 지표도 낮게 나타났다. 이러한 현상의 원인은 앞의 <그림 2>에서 살펴본 아파트 가격을 비롯한 인구 사회적 특성이 주요인인 것으로 분석된다.







### - 7) Latent Distance Model(LDM)을 이용한 추가분석

앞의 분석에서, (1) SBM을 이용한 전체인구 생활권 구분과 서울시 생활권 계획 간의 차이(성동구, 광진구), (2) SBM을 이용한 청년인구 생활권 구분과 전체인구 생활권 구분 간의 차이가 발견되었다. 이를 보다 자세히 분석하기 위해, 우리는 <표 7>과 같이 LDM 모델을 사용하여 각 구를 2차원 잠재공간 위로 시각화 하였다.

<표 7>의 좌측 전체인구 데이터에 대하여 구 별 잠재공간 시각화 한 것을 보면, 성동구와 광진구의 경우 잠재공간 위치 상 동북권과 도심권 둘 중 하나에 뚜렷하게 속한다기 보다는, 두 그룹 사이 비교적 중립적인 지점에 위치한다. 이는 왜 두 생활권 구분의 결과가 다르게 나왔는지를 설명한다.

<표 7>의 우측 청년인구 데이터에 대한 시각화를 보면, 전체인구 생활권에서 서북권으로 분류되었던 은평-서대문-마포와 도심권 종로-중구-용산 사이의 잠재공간 상 위치가 가깝게 나왔다. 이는 왜 기존의 서북권과 도심권이 5번 그룹으로 통합되었는지를 설명 할 수 있다. 또한, 기존에 모두 서남권으로 분류되었던 구로-금천 사이와 관악-영등포-동작이 잠재공간 상에서 뚜렷하게 분류되는 것은, 왜 청년인구 생활권 구분에서 기존의 서남권이 3번과 4번 그룹으로 분리되었는지를 설명한다. 이와 같이, SBM 결과와 LDM 결과를 모두 참고함으로써 모델 기반 생활권 구분에 대하여 보다 많은 정보를 얻을 수 있다.

7) Latent distance model(LDM)은 다음 모형을 가정한다 :  $Y_{ij} \sim \text{poisson}(\log(\beta_0 - \|z_i - z_j\|_2^2))$  여기서  $z$ 는 각 노드의 군집을 의미하며,  $Y$ 는 엣지의 가중치이다. 따라서, 연결강도가 강한 지역끼리 잠재 공간 상에서 가깝게 배치된다. 모델은 R의 cmdstanr 패키지를 통하여 구현하였다. 이때 SBM 적합과 동일한 인접행렬(A)을 사용하였다.

### 3. 분석 활용 전략

#### □ 기대효과

본 분석은 도시 계획에서 중요하게 고려되어야 하는 생활권을 네트워크 모델 기반으로 구분하는 방법을 제시하였다는 점에서 그 의의가 있다. 이 방법을 통해 구분된 생활권을 기존의 생활권 구분과 비교분석함으로써 상호보완적인 용도로 사용할 수 있으며, 아직 심도 있는 생활권 구분이 이루어지지 않은 지역의 경우 본 방법을 프로토타입으로 삼을 수 있다. 또한, 빠르고 간단하게 각 인구 특성 별로 맞춤형 생활권을 구분함으로써, 향후 인구 특성별로 세부적인 도시 정책 수립에 있어 참고할 수 있을 것으로 기대한다.

#### □ 방향제시

분석 결과를 바탕으로 다음과 같은 정책 및 연구 방향을 제안한다. 네트워크 모델을 통해 도출된 클러스터는 현재 생활권 계획과 상당 부분 일치하므로, 향후 이주 관리 정책은 생활권 계획을 반영할 필요가 있다. 특히 청년 인구의 이주는 일자리 창출과 지역의 지속 가능한 발전에 직결되므로, 청년 이주 흐름을 고려한 일자리 정책이 요구된다.

또한, 일부 클러스터가 기존 생활권 구획과 불일치하는 점은 생활권 구획의 재검토를 시사한다. 생활권 정책은 효과적일 수 있으나, 지나치게 생활권 중심으로 설계할 경우 오히려 고립과 지역 불균형을 초래할 수 있다. 특히 동남권의 인구 과밀 해소를 위해, 생활권을 연결하는 지하철 노선 신설을 고려할 필요가 있다. 생활권 중심지를 유기적으로 연결하면 이주 활성화와 고립 완화에 기여할 수 있을 것이다.

생활권은 고정된 개념이 아니며, 변화 가능성을 지속적으로 모니터링하고 이에 따라 정책을 유연하게 조정해야 한다. 이를 위해 이주 네트워크 분석은 유용한 도구이며, 연도별 이주 흐름 분석을 통한 후속 연구가 필요하다.

## 참고문헌

- [1] 고다현(2023). ‘응팔’ 덕선이는 왜 쌍문동을 떠났을까?:서울시 인구구조 변화에 따른 대응방안 모색. 서울연구원.
- [2] 국토연구원(2022). *생활권*. 국토연구원, (검색일: 2025년 6월 30일)
- [3] <https://www.krihs.re.kr/krihsLibraryDictionary/bbsView.es?pageIndex=6&num=87806>
- [4] Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.
- [5] Brin, S., & Page, L. (1998). *The anatomy of a large-scale hypertextual web search engine*. Computer Networks and ISDN Systems, 30(1-7), 107-117.