

그래프 풀링 기반 메타인구의 군집 및 계층 구조 해석

오석준*, 김민경**

Uncovering Latent Hierarchical Structures in Metapopulations via Graph Pooling

Seokjoon Oh*, Minkyung Kim**

요 약

최근 개별 노드 수준의 상호작용을 학습할 수 있는 그래프 신경망 연구는 활발히 이루어져 왔으나, 개별 노드보다 거시적인 메타인구의 영향력을 반영한 연구는 부족하다. 이에 본 연구는 서울시의 개별 POI(Point-of-Interest) 단위의 사용자 이용 시계열 데이터로 이루어진 그래프 신경망 위에, 계층적 그래프 풀링 기법을 적용하여 그래프 풀링이 메타인구 관점에서 제공할 수 있는 해석력을 조명하려 한다. 연구 결과, POI 단위 시계열 그래프는 풀링 레이어를 통과하며 기능적 유사성이 드러나는 중간 단계 슈퍼노드와 복잡한 시계열 패턴이 구분되는 상위 슈퍼노드로 군집하는 양상이 드러났다. 이는 그래프 풀링이 데이터에 내재한 메타인구의 계층적 특성을 인위적 구분 없이 재구성해 포착하는 잠재력이 있음을 시사한다. 본 연구는 그래프 풀링을 다양한 시공간 그래프의 응용 도메인 및 문제 해결에 확장할 수 있는 기반이 될 것으로 기대한다.

Abstract

Recent studies on Graph Neural Networks (GNNs) have been actively focused on representing individual node-level interactions. However, few studies have addressed the influence of higher-level metapopulation structures beyond these local interactions from a macroscopic perspective. To bridge this gap and examine the latent interpretability of graph pooling layer itself, this study applies a hierarchical graph pooling technique to a GNN model, utilizing user activity time series data at the individual Point-of-Interest (POI) level in Seoul. The results demonstrate that intermediate supernodes of POI-level time series data reflect functional similarity, while higher-level supernodes represent coarse-grained temporal patterns. This suggests that graph pooling layers have the potential to reconstruct and capture the hierarchical characteristics of metapopulations inherent in the data without pre-defined standards. We expect this study to facilitate the application of graph pooling approaches to various domains and downstream tasks, enhancing the interpretability of spatiotemporal GNN models.

Keywords

metapopulation, graph neural networks, graph pooling, hierarchical structures, temporal patterns

* 경희대학교 일반대학원 빅데이터응용학과
- ORCID: <https://orcid.org/0009-0007-6764-4141>
** 경희대학교 경영대학 빅데이터응용학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0001-8292-7851>

• Received: Aug. 25, 2025, Revised: Oct. 17, 2025, Accepted: Oct. 20, 2025
• Corresponding Author: Minkyung Kim
Department of Big Data Analytics, School of Management,
Kyung Hee University, Seoul 02447, Republic of Korea
Tel.: +82-2-961-0650, Email: minkkim@khu.ac.kr

I. 서 론

사회적 시스템은 다양한 개체들이 상호작용하는 과정에서 복잡하고 비선형적인 동역학을 보이며, 이러한 상호작용은 시간에 따라 시스템의 구조와 기능을 끊임없이 변화시킨다. 거시적 관점에서 보면, 개체들은 유사한 속성과 기능을 공유하는 하위 집단을 형성하며, 각 집단은 서로 다른 특성과 역할을 지닌 이질적 집합으로 구분될 수 있다. 이러한 하위 집단은 메타인구(Metapopulation)로 정의되며, 학문 분야 간 인용 네트워크[1], 온·오프라인 상의 소셜 네트워크[2], 인간 이동에 따른 모빌리티 네트워크[3] 등 다양한 사회적 시스템 전반에서 관찰된다.

메타인구 간의 관계와 상호작용은 네트워크(그래프) 구조로 표현될 수 있다. 이와 같은 네트워크 표현은 시스템의 전반적인 구조를 시각화하고, 개체 및 집단 간의 연결 패턴, 상호 영향력, 그리고 정보나 자원의 확산 경로를 정량적으로 분석하는 데 중요한 기반을 제공한다[2][4]. 가령, 메타인구 간의 상호 영향력을 모델에 반영함으로써, 학문적 경계를 뛰어넘는 학술논문의 인용 예측[1], 밈(meme)과 같은 다양한 소셜 미디어에 걸친 확산 예측[5] 및 비만과 같은 인구집단의 행태적 특성분석에 이르기까지[6], 메타인구 간 상호작용은 사회적 시스템 전반에 중요한 영향을 미친다[2].

최근에는 네트워크의 구조적 관계를 효과적으로 학습하기 위한 그래프 신경망(GNN, Graph Neural Network) 기반 접근이 활발히 연구되고 있다[7]. 하지만, 대부분의 GNN 연구는 개별 개체 간의 관계에만 초점을 두고 있어, 이종 집단 간 상호작용이 사회적 시스템에 미치는 잠재적 영향력 및 집단 간 계층 구조를 간과했다는 한계가 존재한다.

이에 본 연구는, 그래프의 전역 구조와 군집 특성을 학습 가능하도록 제안된 최근의 그래프 풀링(Graph pooling) 기법을 인간 이동패턴 기반 모빌리티 네트워크에 적용해 메타인구의 군집 및 계층 구조를 밝히려 한다. 이를 위해, 포스퀘어(Foursquare) 사용자의 POI(Point of Interest)와 체크인 시간 정보를 이용하여 그래프를 구축하고, 그래프 풀링을 통해 메타인구의 해석 가능성을 분석, 도출한다.

실험 결과, 그래프 풀링 기반 군집을 통해 군집

품질과 네트워크 구조적 해석력을 제고할 수 있음을 확인하였다. 개별 POI 단위 시계열 데이터에 계층적 그래프 풀링을 적용한 결과, 1단계에서는 유사한 기능적 단위를 내포하는 POI 군집(6개 클러스터)을, 2단계에서는 시계열적 주기성을 내포하는 상위 군집(3개 계층 클러스터)을 도출하였다. 이는 그래프 풀링을 통해 시공간 데이터가 내재한 메타인구 구조를 다면적으로 도출하고, 복합적 해석 가능성을 제공할 수 있음을 시사한다.

본 연구는 인간 이동패턴 기반 모빌리티 네트워크에 GNN 기반 그래프 풀링을 적용함으로써, 메타인구의 동역학적 해석 관점에서 풀링 레이어 자체가 지닌 잠재력을 분석하는 데 기여하였다. 이를 통해, 다양한 응용 도메인에 확장·적용하여 시사점을 도출할 수 있는 기반이 될 것으로 기대한다.

본 논문의 구성은 다음과 같다. 2장의 ‘관련 연구’에서는 메타인구를 표현하는 선행연구의 한계점 및 본 논문에서 다루는 그래프 풀링의 이론적 배경을 정리한다. 3장 ‘데이터 수집 및 전처리’에서는 연구에서 사용한 포스퀘어 데이터셋의 전처리 과정 및 네트워크 구축 방식을 설명한다. 이어 4장의 ‘실험방법 및 결과’에서는 앞서 구축한 네트워크에 GNN 기반 풀링을 적용하여 시계열 데이터 군집 품질 평가 및 군집 특징에 대한 실증평가 결과를 정리한다. 마지막으로 5장 ‘결론 및 향후 과제’에서는 연구의 주요 결과 및 시사점을 요약하고 향후 과제에 대해 논의하는 것으로 논문을 마무리한다.

II. 관련 연구

사회적 개체의 동역학을 포착하려는 기존의 연구는 주로 지역과 같은 독립적인 단위로 구분된 시계열 분석이나 군집화 방법론으로 다루어져 왔다. 가령, 감염병 확산 예측 분야에서는 각 지역에서 발생하는 감염자 수 증감에 대해 LSTM과 같은 외삽(Extrapolation) 기반 기법을 적용하거나[8], K-Means와 같은 군집화 기법을 활용해 지역 간 차이와 유사성을 평가했다[9]. 나아가, 도시 교통 분야에서도 지하철역별 승객 이용 시계열을 지역 단위로 구분해 군집을 분석하고, 승객 수를 예측하는 접근이 제안되었다[10]. 이러한 연구는 각 지역의 특성을 세

부적으로 이해할 수 있다는 장점이 있으나, 지역 간 종합적 상호작용과 같이 거시적인 집단의 파급 효과를 충분히 반영하지 못하는 한계가 존재한다.

이러한 맥락에서, 개별 노드 이상의 메타인구를 함께 고려할 수 있는 분석 프레임워크가 제안되어 [1][11] 메타인구 간 상호작용 및 영향력에 대한 설명은 가능하나, 계층구조를 도출하는 데에는 한계가 있다. 최근 그래프 신경망에서는 개별 노드 이상의 상위 레벨 노드 집합을 구축하는 그래프 풀링 기법이 제안되었다[12]. 즉, 입력 그래프의 노드를 선택적으로 축소, 병합하여 원 네트워크를 슈퍼노드 간 집합 표현으로 변환하여, 메타인구의 계층구조를 학습 가능케 한다.

노드를 축소하는 방식에 있어, 이러한 Clustering 기반 풀링은 노드 간 유사도 또는 구조적 정보를 기반으로 노드를 클러스터링한 후, 각 클러스터를 다음 계층의 슈퍼노드로 병합한다. DiffPool은 노드를 다음 계층의 군집에 확률적으로 할당(Assignment)하여 클러스터링하고, 슈퍼노드로 요약된 표현을 학습한다[13]. 이와 유사하게 MinCutPool은 그래프 부분집합의 최소 컷(Min-cut) 원리에 기반하여 정보 손실을 최소화하는 방향으로 구분이 뚜렷한 클러스터를 구성한다[14]. 이러한 기법들은 기존 네트워크의 노드 정보를 보존하면서도, 계층성이 있는 전역적인 구조로 그래프를 요약할 수 있다는 장점이 있다.

클러스터링 기반 풀링 기법은 노드 간 집단적 상호작용을 분석하는 메타인구 연구의 맥락에서, 누락되는 노드 없이 복잡한 노드 간 관계를 상위 수준의 계층적 집단 패턴으로 구조를 재구성하여 해석할 수 있다는 점에서 큰 잠재력을 가진다. 그러나 현재까지 그래프 풀링 기법은 정적 그래프의 노드 분류 문제와 연산 비용 절감을 위한 그래프 축소에 초점을 두고 연구되어 왔으며, 시공간의 통합적 특성에 기반하여 메타인구 간 계층적 구조를 고려한 시계열 예측 및 군집화 등의 문제 해결을 위해 적용한 사례는 상대적으로 매우 부족하다.

이에 본 연구는 서울시 POI별 시공간 데이터(체크인 횟수, POI 좌표 등)를 기반으로, 계층적 그래프 풀링을 통한 모빌리티 네트워크에서의 메타인구 군집 및 동역학에 대한 해석력을 조명하고자 한다.

III. 데이터 수집 및 전처리

3.1 데이터 전처리

본 연구는 메타인구 관점에서 서울시 내 인구의 다층적 상호작용을 모델링하기 위해, 선행연구에서 수집한 포스퀘어 체크인 데이터를 활용하였다[15]. 포스퀘어는 위치 기반 소셜 네트워크 서비스로 전 세계 사용자의 POI 방문 이력을 포함하고 있으며, 본 연구에서는 이를 유동인구의 이용이 활발하며, 사회적 맥락을 분석하기 좋은 서울시를 분석 대상으로 선정하였다. 이를 위해, 먼저 서울시의 위도, 경도 좌표를 기준으로 2012년 4월부터 2014년 1월까지 수집된 사용자 1,992명의 체크인 기록 124,186건을 필터링하여 이후의 전처리에 활용하였다. 필터링한 각 사용자의 체크인 데이터에는 방문 시각, 방문 장소의 이름(POI), 및 구체적인 좌표 정보가 포함된다.

이때, 세분화된 레벨로 제공된 포스퀘어의 POI 구분을 KSIC 대·중분류 수준의 산업 분류 기준을 참고하여[16] 상위 메타 카테고리로 재구성하였으며, 이를 통해 POI의 특성이 KSIC의 숙박 및 음식점업, 교육서비스 등 주요 생활·서비스 영역에서의 포괄적인 분류 체계와 일관성을 보이도록 하였다.

또한, 체크인 POI 데이터의 시간을 KST 기준으로 변환하고, 주기적 생활패턴을 포착할 수 있도록 각 POI별로 168(7일 × 24시간) 차원의 체크인 시계열 데이터를 생성하였다. 이때, 체크인 빈도가 지나치게 낮은 POI는 시계열 결측치가 다수 발생하여 통계적으로 유의미한 분석이 어려우므로, 빈도수를 기준으로 연속성이 보장되도록 시계열을 구성하였다. 더 나아가, 최상위 빈도의 POI만을 이용할 경우 특정 POI 메타에만 데이터가 편중될 수 있으므로, 대표성을 고려한 8개의 메타 카테고리를 대상으로 방문 빈도가 높은 38개의 POI별 시계열을 구성하였다. 이는 도시 내 생활 이동에 내재되어 있는 주중·주말 동안의 주기적 패턴을 반영하고, 출퇴근·식사·야간 활동 등 주요 변동성[3][15][17]을 1시간 단위 해상도로 명확히 포착함으로써, 다양한 메타인구 맥락을 다각도로 이해하기 위함이다. 최종 전처리한 데이터셋의 기초 통계 정보는 아래의 표 1과 같다.

표 1. 전처리된 포스퀘어 데이터셋의 기초통계 요약
Table 1. Basic statistics of the preprocessed Foursquare dataset

Category	Number of cases
Users	1,699
Check-ins	88,971
POIs	38
POI meta classification	8
Time stamp	168

3.2 모빌리티 네트워크 구축

모빌리티 네트워크의 전역적 구조와 군집 특성을 학습하기 위해, POI를 노드로, 주중 방문 시계열의 비선형적 상관성을 엣지의 가중치로 설정해 그래프를 구축하였다. 이때, 엣지의 정의는 시계열 유사성에 기반하여 그래프를 구축한 선행연구[18]의 전처리 방식에 따라 코렌트로피(Correntropy)를 사용하였다.

그래프 구축을 위해, 먼저 전체 POI 쌍의 코렌트로피를 계산하여 38×38 유사도 행렬을 생성한 뒤, 각 노드에서 자신을 제외한(대각성분 제외) 상위 4개(k=4)의 이웃을 선택하는 가중 k-NN 방식을 적용하였다. 이러한 방식은 체크인 데이터의 비선형적 특성을 포착하여, 메타인구의 전역적, 계층적 구조를 유동적으로 정의·관찰하기 위함이다.

IV. 실험방법 및 결과

4.1 그래프 풀링 모듈 구성

본 연구는 시계열 데이터에 계층적 풀링을 적용한 최신 HiGP[19] 모델로부터 그래프 풀링 레이어를 포함하는 GNN 기본 프레임워크를 참조하였다. HiGP는 기본적으로 GCN(Graph Convolutional Network) 구조에 이방성 메시지 패싱(Anisotropic message passing)이 가능한 Gate 기반 GraphAnisoConv 레이어를 적용한다. 풀링 모듈과 결합되는 GraphAnisoConv 층의 수는 연구에서 설정한 계층의 깊이를 반영하며, 노드 특징을 이웃 노드와 결합하여 메시지를 생성하고, 각 메시지는 sigmoid 기반 게이트를 통해 0과 1 사이의 중요도 값으로 조절된다. 이 게이트 값은 메시지의 각 성분에 원소별로 곱해져 정보의 흐름을 선

택적으로 강화하거나 억제하여 이웃 노드 간 영향력을 반영한다. 이후 skip connection과 활성화 함수를 거쳐 다음 계층으로 전달되며, 이러한 구조를 통해 HiGP는 노드 간 비선형적, 이방성 관계를 효과적으로 학습할 수 있다. 또한 그래프 축소를 위한 풀링 연산으로서 노드 클러스터링 기반 풀링인 MinCutPool에 Softmax[14]와 Gumbel-Softmax Trick[20]을 각각 적용한 결과를 비교하였다(그림 1 참조).

첫 번째, MinCutPool에 Softmax를 적용하여 노드를 다음 계층의 슈퍼노드로 풀링하는 방식은 아래 식 (1), (2)와 같다.

$$S = \text{softmax}(\bar{X}; \theta_{\text{GNN}}) \quad (1)$$

$$A^{\text{pool}} = S^T \tilde{A}; \quad X^{\text{pool}} = S^T \bar{X} \quad (2)$$

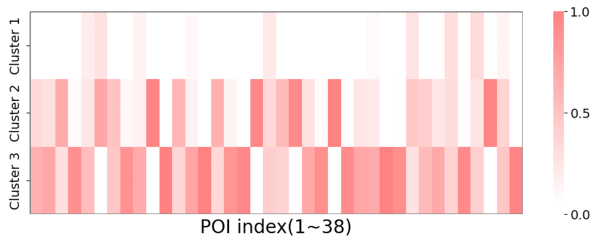
이때, 식 (1)의 S 는 이전 계층에서 GNN을 통해 생성된 노드 임베딩 $\bar{X} \in R^{N \times d}$ 을 입력으로 받아, softmax 함수를 통해 확률적으로 정규화된 할당 행렬 $S \in R^{N \times K}$ 을 의미한다. 즉, 식 (1)에서 생성된 행렬은 N 개의 노드를 K 개의 클러스터로 할당하는 soft cluster assignment matrix로서, 원 그래프의 노드가 단일 클러스터에 완전히 속하는 것이 아니라 $[0, 1]$ 범위의 확률 분포 하에서 부분적으로 여러 클러스터에 소속될 수 있는 soft assignment 구조를 갖는다. 생성된 할당 행렬 S 는 식 (2)를 통해 원 그래프의 인접 행렬과 노드 특성 행렬을 축소하는 데에 적용된다. 이때, $\tilde{A}$ 는 정규화된 인접 행렬을 의미하며, 연산된 A^{pool} 은 클러스터 간의 연결 관계를 나타내는 축소된 인접 행렬이다. 즉, A^{pool} 은 노드들 간의 연결 정보를 바탕으로 다음 계층의 슈퍼노드 간의 연결 구조를 간접적으로 연산한다. 또한, X^{pool} 은 각 클러스터의 특성을 원래 노드의 특성 가중합으로 연산한 노드 특성 행렬로, 다음 계층에서의 풀링된 슈퍼노드의 GNN 입력으로 사용된다. 이러한 구조는 GNN 모델 학습의 유연성을 제공하지만, 모델의 해석력 관점에서는 각 노드의 클러스터 소속이 모호하게 나타나 오히려 해석력이 떨어질 수 있다.

두 번째, 위의 식 (1) 대응하여 MinCutPool에 적용할 Gumbel-Softmax Trick은 아래 식 (3)과 같다.

$$y_i = \frac{\exp((\log(\pi_i) + g_i)/\tau)}{\sum_{j=1}^k \exp((\log(\pi_j) + g_j)/\tau)} \quad (3)$$

식 (3)에서 π_i 는 노드 i 가 군집에 할당될 사전 확률(logit)을, τ 는 입력 분포의 정밀도를 조절하는 temperature 파라미터로, 값이 작을수록 softmax 결과가 더 뚜렷해져 hard assignment에 가까운 분포를 형성한다. 따라서, MinCutPool에 Softmax를 적용한 할당행렬(S)에 비해 보다 명확한 클러스터 할당이 가능하므로 해석력이 높아, 본 연구에서는 MinCutPool에 Gumbel-Softmax Trick을 최종 적용하였다.

그림 1은 앞서 설명한 그래프 풀링 방식을 38개의 개별 POI 시계열 데이터에 적용해 얻어진 할당행렬(S)을 시각화한 결과이다.



(a) Softmax 적용 할당행렬(S) 시각화
(a) Visualization of assignment matrix (S) with Softmax applied



(b) Gumbel-Softmax 적용 할당행렬(S) 시각화
(b) Visualization of assignment matrix (S) with Gumbel-Softmax applied

그림 1. Softmax(a)와 Gumbel-Softmax(b) 적용에 따른 MinCutPool의 노드-클러스터 할당행렬(S) 시각화 비교

Fig. 1. Comparative visualization of node-to-cluster assignment matrices obtained from MinCutPool with Softmax (a) and Gumbel-Softmax (b) applied

그림 1(a)는 soft assignment 특성으로 인해 다수의 POI가 세 개의 클러스터에 확률적으로 분산되어 나타나는 반면, 그림 1(b)는 Gumbel-softmax trick의 적용으로 각 POI가 단일 클러스터에 소속되도록 수렴

하여, 보다 명확한 클러스터 구조가 형성됨을 확인할 수 있다. 결과적으로, 그래프 풀링을 통한 할당행렬은 모델의 학습 과정에서 외부의 군집 알고리즘 없이도 군집 및 계층 구조 도출이 가능함을 보여준다. 나아가, 비교를 통해 Gumbel-Softmax Trick이 클러스터 경계의 명확성과 직관적 해석 가능성 향상에 실질적으로 기여함을 알 수 있다.

4.2 실험 환경 설정

그래프 풀링 모델은 2계층 이하의 얇은 구조에서 우수한 성능을 보여주었다[21]. 이에, 실험에 활용한 네트워크 크기(N=38)를 고려하여 GraphAnisoConv 층과 MinCutPool 층이 교차하는 2단계의 계층적 군집화 구조(CONV→POOL→CONV→POOL)로 설계하였다.

그래프 풀링 연산을 위한 하이퍼파라미터는 다음과 같다: Gumbel-Softmax temperature는 0.5로 설정하였으며, 온도 감소율(temp_decay)은 0.995로 두어 학습이 진행됨에 따라 점진적으로 hard assignment에 수렴하도록 하였다. MinCut loss와 Orthogonality loss의 손실 가중치는 각각 $\lambda = 0.75$ 및 $\gamma = 1.25$ 를 설정하였다. 학습에는 Adam 옵티마이저를 적용하였으며, 학습률은 0.003, 총 200 에포크로 진행하였다.

4.3 군집 수에 따른 군집 품질의 변동성 분석

다운스트림 태스크(Downstream tasks)가 결합되지 않은 그래프 데이터에 대해, 계층적 풀링 모델의 중간 계층(첫번째 레이어) 군집 개수 변화에 대한 군집 품질의 변동성을 분석하였다. 이를 위해, 그래프 모듈러리티(Modularity)와 그래프 복원력 측정을 위한 원본 및 복원 그래프 간의 인접행렬 평균제곱오차(MSE)를 지표로 활용하였다. 모듈러리티는 각 군집 내 연결 밀도와 군집 간 연결 밀도의 상대적 차이를 기반으로 군집 결집도를 정량화한다. 인접행렬 평균제곱오차는 값이 낮을수록 군집화 및 풀링 과정에서 정보 손실이 적고, 원본 네트워크의 연결 패턴이 효과적으로 보존되었음을 의미한다. 이때, 중간 계층의 군집 개수는 5개에서 10개까지 변화시키며, 각 군집 수마다 20개의 시드를 적용하여 평균을 계산하였다.

표 2. 그래프 클러스터링 관점의 군집 품질 성능 비교

Table 2. Clustering quality performance comparison of graph clustering approaches

Approach	Modularity (Best)	Conductance	Coverage	Reconstruction MSE
K-means	0.0696 ± 0.0133 (0.1063)	0.6674 ± 0.0246	0.8625 ± 0.0168	0.5851 ± 0.0232
MinCutPool (Softmax)	0.0516 ± 0.1092 (0.3798)	0.4373 ± 0.1124	0.9387 ± 0.1277	0.2571 ± 0.0830
MinCutPool (Gumbel-Softmax)	0.2152 ± 0.0693 (0.3626)	0.4548 ± 0.0777	0.5708 ± 0.0546	0.1573 ± 0.0211

원본 노드 수가 38개인 그래프에서, 1차 폴링 이후 남기는 군집 수를 달리하며 실험한 결과(그림 2 참조), 전체 노드의 약 20% 이하의 비율에 해당하는 6~7개 군집으로 축소했을 때 두 지표 모두에서 평균적으로 뛰어난 성능을 보였다. 나아가, 중간 단계의 군집 수가 너무 많아질 경우에는 군집이 지나치게 세분화되어 그래프의 조직적 품질과 원 그래프를 설명할 수 있는 정보의 복원력 모두에 부정적인 영향을 미침이 드러난다. 이러한 결과는 그래프 폴링 과정에서 적정 군집 수의 선택이 품질 및 성능 모두에 중요함을 시사하며, 이후의 분석에서는 두 지표를 종합적으로 고려하여 중간 계층의 군집 수를 6으로 설정하였다.

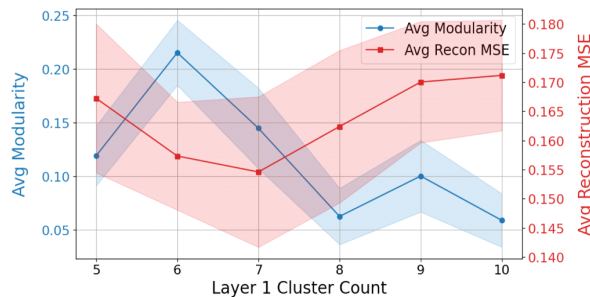


그림 2. 군집 수 변화에 따른 군집 모듈러리티 및 그래프 복원 성능의 변화

Fig. 2. Changes in clustering modularity and graph reconstruction performance with varying number of clusters

4.4 그래프 폴링 모델의 군집 품질 성능 비교

본 연구의 목표인, 데이터에 내재되어 있는 메타인구 간의 동역학적 해석을 위해, 도출된 군집 품질

의 평가는 매우 중요하다. 이를 위한 정량 지표로서, 군집의 정답 라벨이 없는 상황에 적합한, 네트워크 기반 모듈러리티, 컨덕턴스(Conductance), 커버리지(Coverage), 및 그래프 복원력(Reconstruction)을 종합적으로 고려하였다. 커버리지는 형성된 군집 내부의 연결 밀도를 나타내며, 컨덕턴스는 군집의 분리도를 평가하여 값이 낮을수록 군집이 평균적으로 외부와의 연결이 적어 경계가 선명함을 의미한다. 실험을 위한 그래프 클러스터링 기법으로, 앞서 구축한 MinCutPool 기반 모델에 Softmax와 Gumbel-Softmax를 각각 적용한 두 가지 변형과 전통적인 군집화 기법인 K-means를 사용하였으며, k는 최종 폴링 군집 수와 같은 3으로 설정하였다.

표 2는 위의 각 지표에 대해 20시드(seed)를 적용한 검증 결과의 평균값과 표준편차를 나타낸다. 표에서, Gumbel-Softmax를 적용한 MinCutPool 모델이 군집의 구조적 품질과 표현 학습 능력 모두에서 종합적으로 우수한 성능을 보여준다. 모듈러리티 지표에서 평균 0.215로 가장 높은 점수를 달성했으며, 노드 표현 학습의 품질을 나타내는 복원력 역시 가장 높아(오차평균 MSE: 0.1573), GNN 인코더가 노드의 특징 정보를 가장 효과적으로 학습했음을 시사한다. 이러한 성능 우위는, 모델이 각 노드를 특정 군집에 명확하게 할당하도록 강제하는 Gumbel-Softmax의 hard assignment 방식에 기인하는 것으로 분석된다. 이러한 이산적 선택 과정은 군집의 구조적 분리도를 높이는 동시에, 모델이 더 판별력 있는 노드 표현을 학습하도록 유도하여 모듈성과 복원력 성능을 동시에 향상시킨다. 반면, Softmax는 확률적인 할당을 통해 군집 경계가 모호해져, 커버리지 지표는 높게 나타났으나 모듈러리티

의 평균이 낮고 편차가 큰 불안정한 군집을 구성함을 확인할 수 있다.

K-means는 노드 피쳐 벡터의 유사도를 기반으로 군집화를 수행하며, 노드 간의 연결 관계, 즉 원 그래프의 위상 구조는 고려하지 않는다. 이러한 방법론적 한계는 모듈러리티와 복원력 모두에서 낮은 성능을 보여주었다. 이는 K-means가 시계열 피쳐를 기반으로 형성한 군집이 실제 네트워크 커뮤니티 구조와는 일치하지 않음을 의미하며, 가장 높은 복원력 오류(MSE) 값은 K-means가 노드의 확률적 할당과 같이 고차원의 노드 표현을 학습하는 기능을 수행하지 않음을 의미한다. 그림 3은 시계열 군집화에 있어 앞서 설명한 K-means의 군집 특성을 실증적으로 더 명확히 이해하기 위한 시각화 결과이다. K-means 결과에서 모듈러리티가 가장 높았던 시드의 군집화 결과, 38개 POI 중 체크인 빈도 상위 8개 POI 순위, 즉 지하철(Cluster 2), 카페·커피숍·회사·한식당·대학교·바·버스(Cluster 1)만이 다른 클러스터에 할당되었고, 이용 빈도가 상대적으로 낮은 나머지 30개 POI의 시계열 특성은 다소 간과되어 단일 군집(Cluster 3)에 모두 할당되는 결과가 나타났다. 이는 K-means 알고리즘이 시계열의 주기적 유사성보다는, 체크인 빈도에서 비롯되는 규모(scale)의 차이에 더 민감하게 반응해 군집을 형성하기 때문으로 보이며, 특히 표1에서 커버리지 점수가 높은 것 또한 대부분의 노드를 하나의 거대 군집으로 묶어버린 현상에 의한 것임을 추측할 수 있다.

MinCutPool은 결합된 다운스트림 태스크의 손실을 통해 군집을 학습하는 DiffPool[13]과 달리, MinCut 손실과 직교성 정규화를 통해 풀링 레이어 자체에서 그래프의 구조적 커뮤니티를 최적화하므로 본 연구의 목적에 더 부합하다. 한편, 그래프 구조적인 관점에서 작동 메커니즘이 근본적으로 다른 TopKPool[22] 및 SAGPool[23]과 같은 풀링 기법은 원 그래프에서 중요도가 낮은 노드를 제거하는 방식으로 작동하며[12], 이는 모든 노드를 보존해 다음 단계의 군집으로 표현하는 MinCutPool과 달리 그래프를 축소하는 방식에서 구조적인 차이를 야기한다. 이에, 본 연구에서는 전체 POI의 누락 없이 상위 계층의 군집 정보를 표현할 수 있는 MinCutPool에 기반하여 전체 POI 간의 상호 연결과

패턴을 이해하고자 하였다.

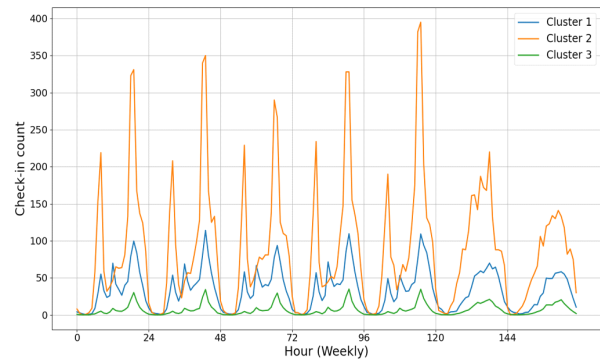


그림 3. K-means 클러스터링을 통한 시계열 그래프 군집 결과의 시각화

Fig. 3. Visualization of time series graph clustering results via K-means clustering

4.5 POI 방문 네트워크 그래프 풀링 결과

앞서 논의한 풀링 모듈의 시계열 데이터 해석력을 분석하기 위해, 4.4의 그래프 풀링 군집 품질 성능평가 결과 중 모듈러리티가 가장 높았던 시드(0.3626)의 계층적 군집 결과에 대한 실증적 분석을 진행하였다. 분석 결과, 인위적인 구분과는 달리, 첫 번째 하위 계층의 분류 구조에서 POI 간의 기능적 유사성이 반영된 계층적 군집 구조를 도출하였다.

그림 4는 POI 방문 네트워크의 풀링 결과로서, 할당행렬(S)을 통해 모델로부터 도출한 계층적 군집 구조를 도식화한 것이다. 먼저, 1단계 풀링을 통해 형성된 중간 계층의 6개 클러스터는 군집별로 메타 인구가 이용하는 POI의 기능적 유사성이 드러나는 양상을 보인다. 예컨대, Cluster 1은 아시안·이탈리안·일식 레스토랑, 쇼핑몰, 멀티플렉스 등 주로 점심에 이용하는 식당 및 저녁까지 이어지는 외식과 상업 활동을 중심으로 군집을 형성하였다. 또한, Cluster 5는 젊은 층에서 선호하는 오락실, 야구장, 영화관 등 구체적인 여가 장소가 중심을 이루고 있으며, Cluster 6은 지하철을 포함해 은행, 의류 매장, 패스트푸드점 등 역세권에서 흔히 볼 수 있는 상업 및 편의 시설을 드러낸다. 이러한 결과는 그래프 기반 풀링이 POI 간의 기능적 유사성과 시간적 이용 패턴을 효과적으로 반영하여, 메타인구 이동의 구조적 특성을 계층적으로 도출할 수 있음을 시사한다.

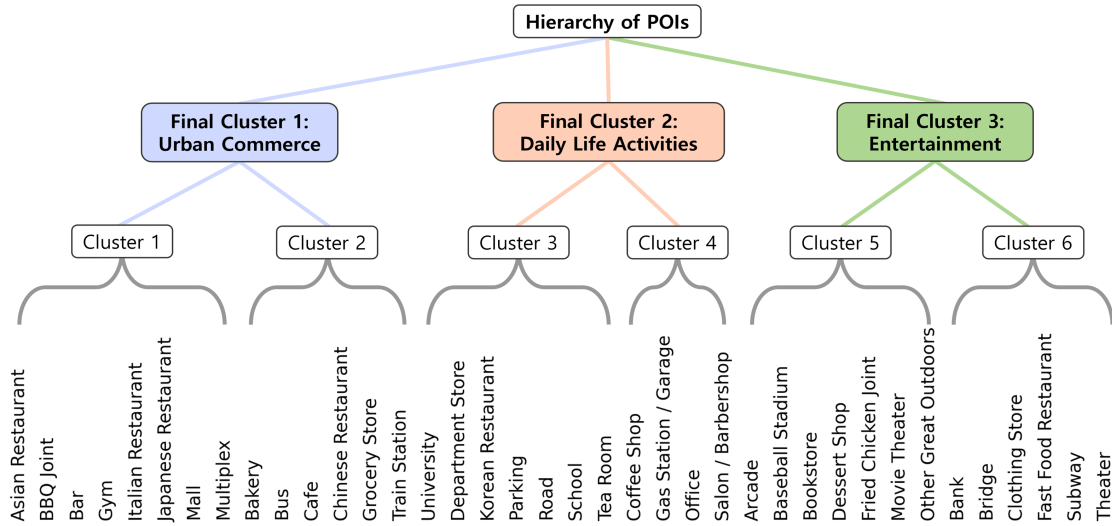


그림 4. 2 계층 그래프 풀링에 따른 POI의 계층적 군집 구조 시각화

Fig. 4. Visualization of the hierarchical clustering structure of POIs based on two-level graph pooling

나아가, 1단계 군집은 다시 2단계 풀링을 통해 3개의 상위 슈퍼노드로 풀링되었으며, 이때 각 군집은 인위적인 POI 메타 클래스 이상으로 거시적인 시계열적 특성이 드러나는 생활이동 패턴을 대표한다. 그림 5는 이를 시각화한 최종 계층의 그래프 풀링 결과로서, 세 가지 주요 패턴을 확인할 수 있다. 그림에서 Final Cluster 1은 주중과 주말 모두 점심 및 저녁 시간대를 중심으로 꾸준히 피크를 보이며, 외식, 쇼핑 등 복합적인 도심 활동이 집중되는 도시 상업 중심(Urban Commerce)의 생활이동 패턴을 대표한다. 또한, Final Cluster 2는 주중의 일상적인 시간대(오전, 점심, 오후)에 매우 뚜렷하고 규칙적인 피크(peak)를 보이며, 회사와 함께 직장인과 학생들이 일과 중에 자주 이용하는 한식당, 커피숍, 찻집 및 주유소, 주차장과 같은 기능적 POI가 묶여 일상적 활동(Daily Life Activities)을 대표한다. 마지막으로, Final Cluster 3은 점심을 거쳐 저녁 시간대에 가장 높은 피크를 보이는 시계열 패턴을 특징으로 한다. 이는 야구 경기나 영화 등 특정 이벤트에 따라 저녁까지 이용자가 점진적으로 집중되는 오락 활동(Entertainment)의 생활이동 패턴을 반영한다.

종합하여, 계층적 그래프 풀링을 통해 인위적 구분 없이도 데이터에 내재한 군집 단위의 특징을 도출할 수 있으며, 이는 네트워크 정보를 고려하지 않은 단일 계층의 군집화로는 발견하기 어려운 복잡한 시간 활동 패턴에 대한 해석력을 제공한다.

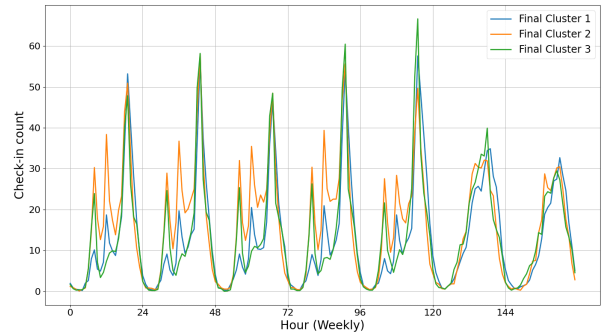


그림 5. 최종 풀링 결과의 시계열 그래프 군집

Fig. 5. Time series graph clustering of final pooling results

V. 결론 및 향후 과제

본 연구에서는 서울시의 포스퀘어 시공간 데이터를 이용해, 그래프 풀링 기반 GNN 모델이 잠재한 메타인구의 계층적·다차원적 패턴에 대한 해석력을 분석하였다. 특히, 시계열 체크인 데이터를 그래프 형태로 변환하고, MinCutPool 기반의 계층적 그래프 풀링 기법을 적용하여 서울시 내 POI 단위에서의 시계열이 지닌 잠재적 군집 구조를 도출하였다.

MinCutPool의 구조 분석 결과, 그래프 풀링 모델의 할당 행렬을 통해 모델이 학습한 원 노드의 군집 정보를 도출하였으며, 이는 별도의 군집분석을 수행하지 않고도 상위 메타인구의 군집 정보를 학습할 수 있음을 의미한다. 나아가, 네트워크 지표 기반 평가 및 그래프 구조적인 분석을 통해 군집

품질 및 폴링 기법의 구조적인 해석 가능성에 대한 차별성을 도출하였다. 최종적으로, 38개 POI 단위 그래프에 계층적 그래프 폴링을 단계적으로 적용한 결과를 실증적으로 분석하는 것으로, 서울시 유동인구의 생활이동 패턴이 드러나는 3개의 상위 슈퍼노드로 폴링되는 양상을 확인하였다. 이러한 결과는, 이동 데이터로부터 그래프 폴링을 통해 인위적 정의에서 벗어나 동적인 계층 구조를 학습, 도출할 수 있음을 보여주며, 다양한 관점의 메타인구 분석 도구로서 그래프 폴링이 지닌 잠재력을 뒷받침한다.

향후 연구 계획으로는, 그래프 폴링을 적용한 시공간 GNN을 여러 다운스트림 태스크에 확장, 적용하여 계층적인 메타인구 구조를 학습한 그래프 폴링 모델을 예측 및 분류 등의 모형에 적용하려 한다. 또한 교통, 상권, 감염병 확산 등 응용 도메인을 확장하여 범용적인 그래프 기반 메타인구 분석 프레임워크를 연구하고자 한다.

References

- [1] M. Kim, D. McFarland, and J. Leskovec, "Modeling affinity based popularity dynamics", *Proc. of the ACM International Conference on Information and Knowledge Management*, Singapore, pp. 477-486, Nov. 2017. <https://doi.org/10.1145/3132847.3132923>.
- [2] M. Kim, D. Paini, and R. Jurdak, "Real-world diffusion dynamics based on point process approaches: A review", *Artificial Intelligence Review*, Vol. 53, pp. 321-350, Jan. 2020. <https://doi.org/10.1007/s10462-018-9656-9>.
- [3] H. Oh and M. Kim, "Dynamics of COVID-19 spread based on human mobility patterns by epidemic stages in Seoul", *The Journal of Korean Institute of Information Technology*, Vol. 21, No. 4, pp. 153-160, Apr. 2023. <https://doi.org/10.14801/jkiit.2023.21.4.153>
- [4] M. Starnini, M. Frasca, and A. Baronchelli, "Emergence of metapopulations and echo chambers in mobile agents", *Scientific Reports*, Vol. 6, No. 1, 31834, Aug. 2016. <https://doi.org/10.1038/srep31834>.
- [5] L. Weng, F. Menczer, and Y. Ahn, "Virality prediction and community structure in social networks", *Scientific Reports*, Vol. 3, No. 1, 2522, Aug. 2013. <https://doi.org/10.1038/srep02522>.
- [6] M. White, M. Wellons, J. Rodriguez, J. Howard, K. Holliday, S. Gesell, E. Perrin, and J. Moody, "Family ties and health: Exploring the link between parents' social networks and child obesity", *Child Obesity*, Vol. 21, No. 7, pp. 569-579, Oct. 2025. <https://doi.org/10.1089/chi.2025.0003>.
- [7] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. Yu, "A comprehensive survey on graph neural networks", *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 32, No. 1, pp. 4-24, Jan. 2021. <https://doi.org/10.1109/TNNLS.2020.2978386>.
- [8] I. Sembiring, S. Wahyuni, and E. Sedyono, "LSTM algorithm optimization for COVID-19 prediction model", *Heliyon*, Vol. 10, No. 4, e26158, Feb. 2024. <https://doi.org/10.1016/j.heliyon.2024.e26158>.
- [9] M. Zubair, M. Iqbal, A. Shil, E. Haque, M. Hoque, and I. Sarker, "An efficient k-means clustering algorithm for analysing COVID-19", *International Conference on Hybrid Intelligent Systems*, pp. 422-432, Dec. 2020. https://doi.org/10.1007/978-3-030-73050-5_43.
- [10] Y. Park, Y. Choi, K. Kim, and J. Yoo, "Machine learning approach for study on subway passenger flow", *Scientific Reports*, Vol. 12, No. 1, 2754, Feb. 2022. <https://doi.org/10.1038/s41598-022-06767-7>.
- [11] M. Kim, D. Paini, and R. Jurdak, "Modeling stochastic processes in disease spread across a heterogeneous social system", *Proc. of the National Academy of Sciences*, Vol. 116, No. 2, pp. 401-406, Jan. 2019. <https://doi.org/10.1073/pnas.1801429116>.
- [12] D. Grattarola, D. Zambon, F. Bianchi, and C. Alippi, "Understanding pooling in graph neural networks", *IEEE Transactions on Neural Networks*

- and Learning Systems, Vol. 35, No. 2, pp. 2708-2718, Feb. 2022. <https://doi.org/10.1109/TNNLS.2022.3190922>.
- [13] Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling", Advances in Neural Information Processing Systems, Montréal, Canada, Vol. 31, Dec. 2018. <https://doi.org/10.48550/arXiv.1806.08804>.
- [14] F. Bianchi, D. Grattarola, and C. Alippi, "Spectral clustering with graph neural networks for graph pooling", Proc. of the International Conference on Machine Learning, Vienna, Austria, pp. 874-883, Nov. 2020. <https://doi.org/10.48550/arXiv.1907.00481>.
- [15] S. Oh, S. Joo, S. Kim, and M. Kim, "Understanding spatiotemporal human mobility patterns in Seoul based on POI metadata", The Journal of Korean Institute of Information Technology, Vol. 22, No. 9, pp. 111-122, Sep. 2024. <https://doi.org/10.14801/jkiit.2024.22.9.111>.
- [16] Korean Standard Industrial Classification (KSIC), Statistics Korea, <https://url.kr/jyqqgo>. [accessed: Aug. 01. 2025]
- [17] S. Oh, S. Joo, S. Kim, and M. Kim, "Decoding urban dynamics: Contextual insights from human meta-mobility patterns", Systems, Vol. 12, No. 8, 313, Aug. 2024. <https://doi.org/10.3390/systems12080313>.
- [18] A. Cini, I. Marisca, and C. Alippi, "Filling the G_ap_s: Multivariate Time Series Imputation by Graph Neural Networks", International Conference on Learning Representations, Online, Apr. 2022. <https://doi.org/10.48550/arXiv.2108.00298>.
- [19] A. Cini, D. Mandic, and C. Alippi, "Graph-based time series clustering for end-to-end hierarchical forecasting", Proc. of the International Conference on Machine Learning, Vienna, Austria, Vol. 235, pp. 8985-8999, Jul. 2024. <https://doi.org/10.48550/arXiv.2305.19183>.
- [20] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax", Proc. of the International Conference on Learning Representations, San Juan, Puerto Rico, Apr, May 2016. <https://doi.org/10.48550/arXiv.1611.01144>.
- [21] J. Wu, X. Chen, K. Xu, and S. Li, "Structural entropy guided graph hierarchical pooling", Proc. of the International Conference on Machine Learning, Baltimore, Maryland, Vol. 162, pp. 15960-15970, Jun. 2022. <https://doi.org/10.48550/arXiv.2206.13510>.
- [22] H. Gao and S. Ji, "Graph U-Nets", Proc. of the International Conference on Machine Learning, California, USA, Vol. 97, pp. 2083-2092, May 2019. <https://doi.org/10.48550/arXiv.1905.05178>.
- [23] J. Lee, I. Lee, and J. Kang, "Self-attention graph pooling", Proc. of the International Conference on Machine Learning, California, USA, Vol. 97, pp. 3734-3743, May 2019. <https://doi.org/10.48550/arXiv.1904.08082>.

저자소개

오 석 준 (Seokjoon Oh)



2024년 2월 : 경희대학교
응용영어통번역학과(문학사)
2024년 3월 ~ 현재 : 경희대학교
일반대학원 빅데이터응용학과
석박통합과정
관심분야 : 빅데이터, 딥러닝,
그래프 마이닝, 소셜 컴퓨팅

김 민 경 (Minkyung Kim)



2015년 6월 : The Australian
National University, School of
Computing(공학박사)
2022년 9월 ~ 현재 : 경희대학교
빅데이터응용학과 교수
관심분야 : 인공지능, 그래프
머신러닝, 소셜 컴퓨팅, 빅데이터