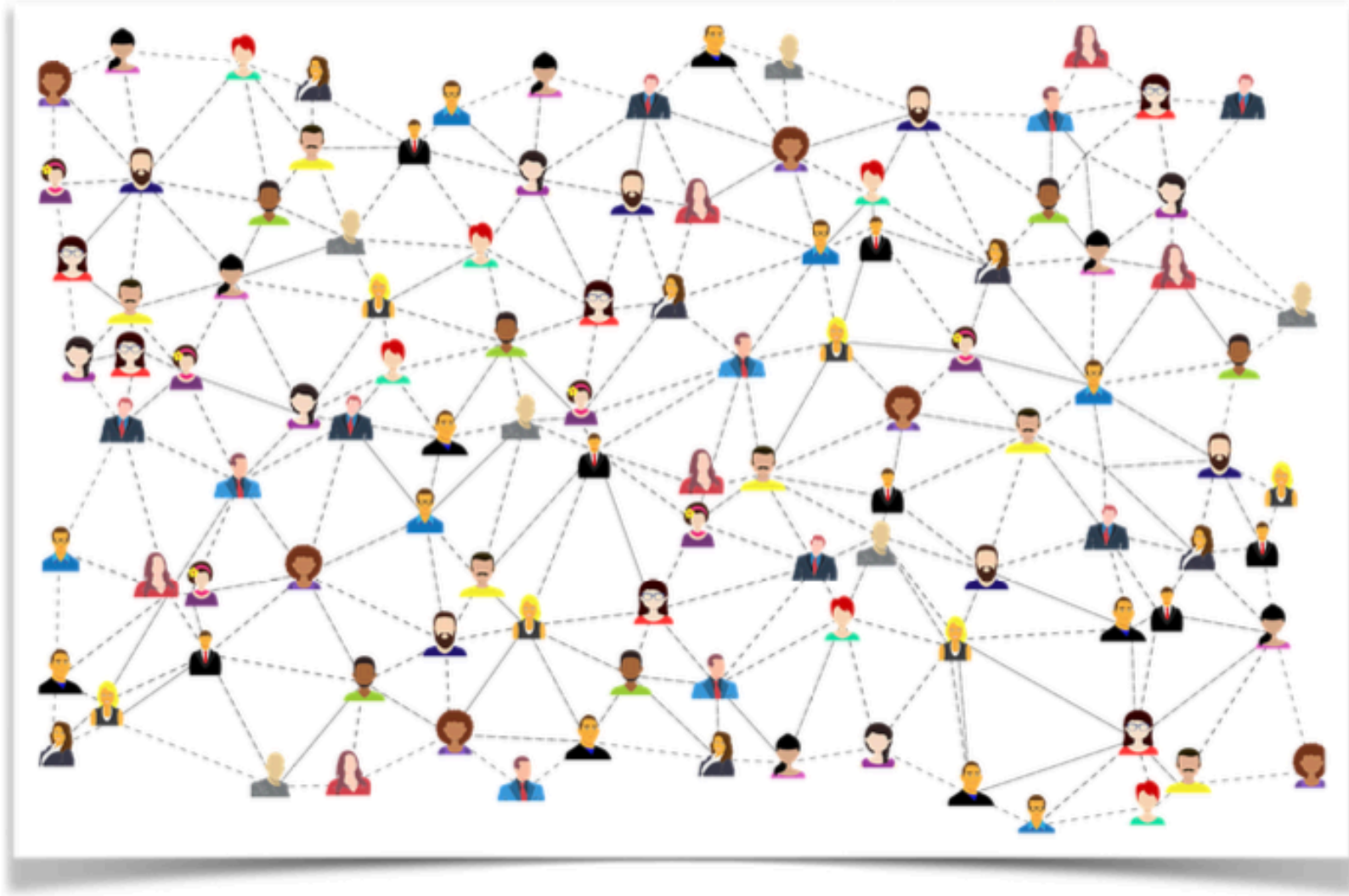


Abstract Subgraph-based GNN Explanation

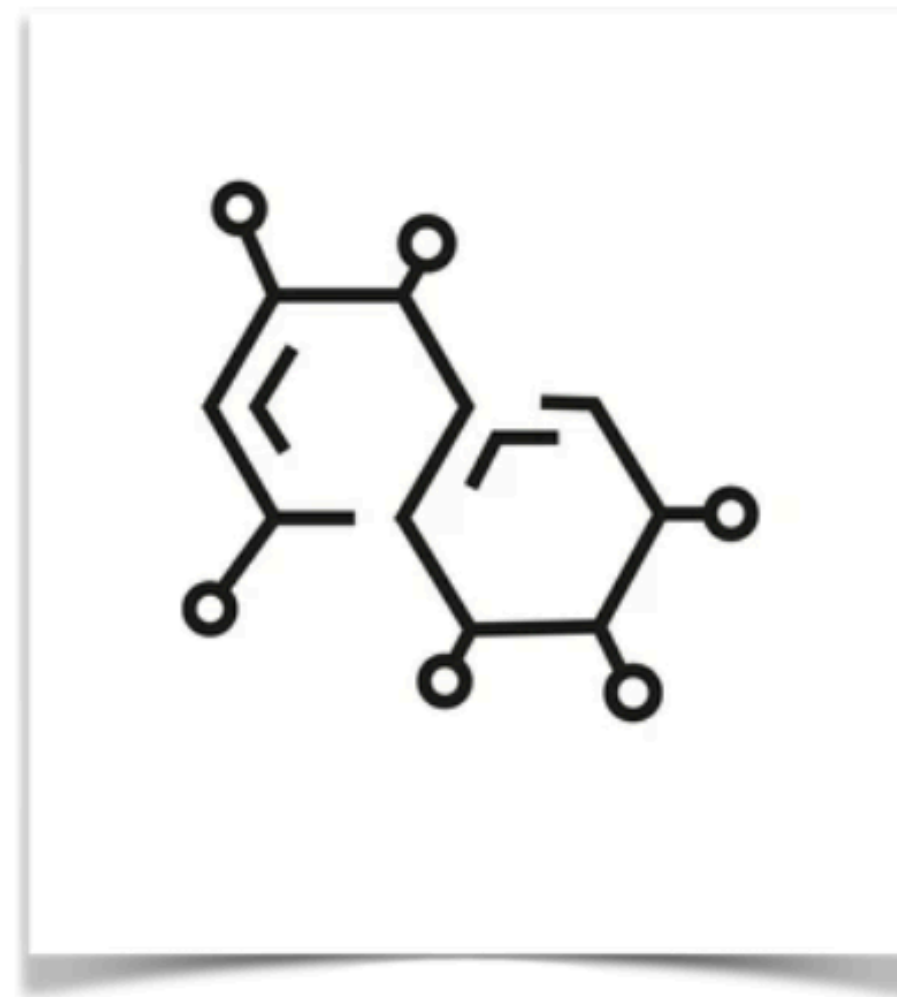
Seunghyun Park

Graph Data

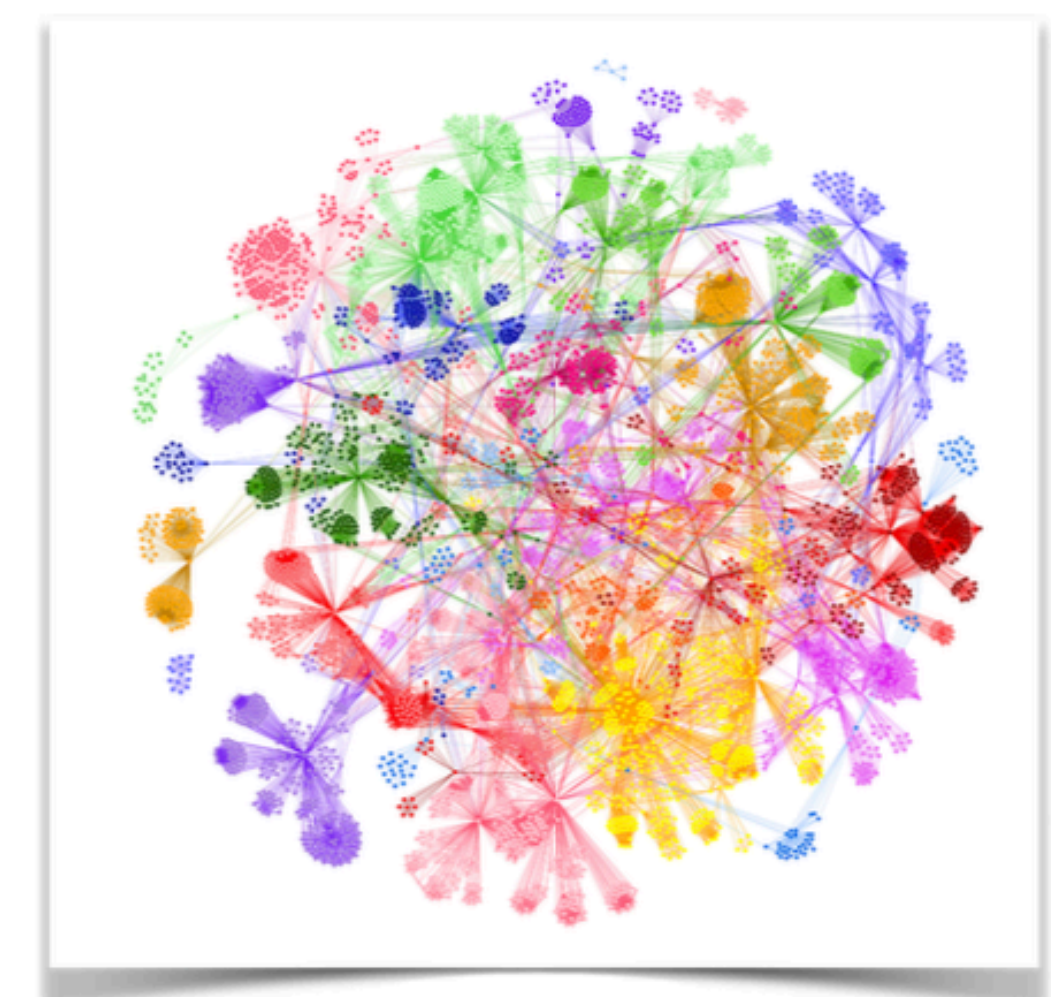
- 데이터 간의 관계를 표현할 수 있는 널리 사용되는 중요한 데이터



소셜 네트워크



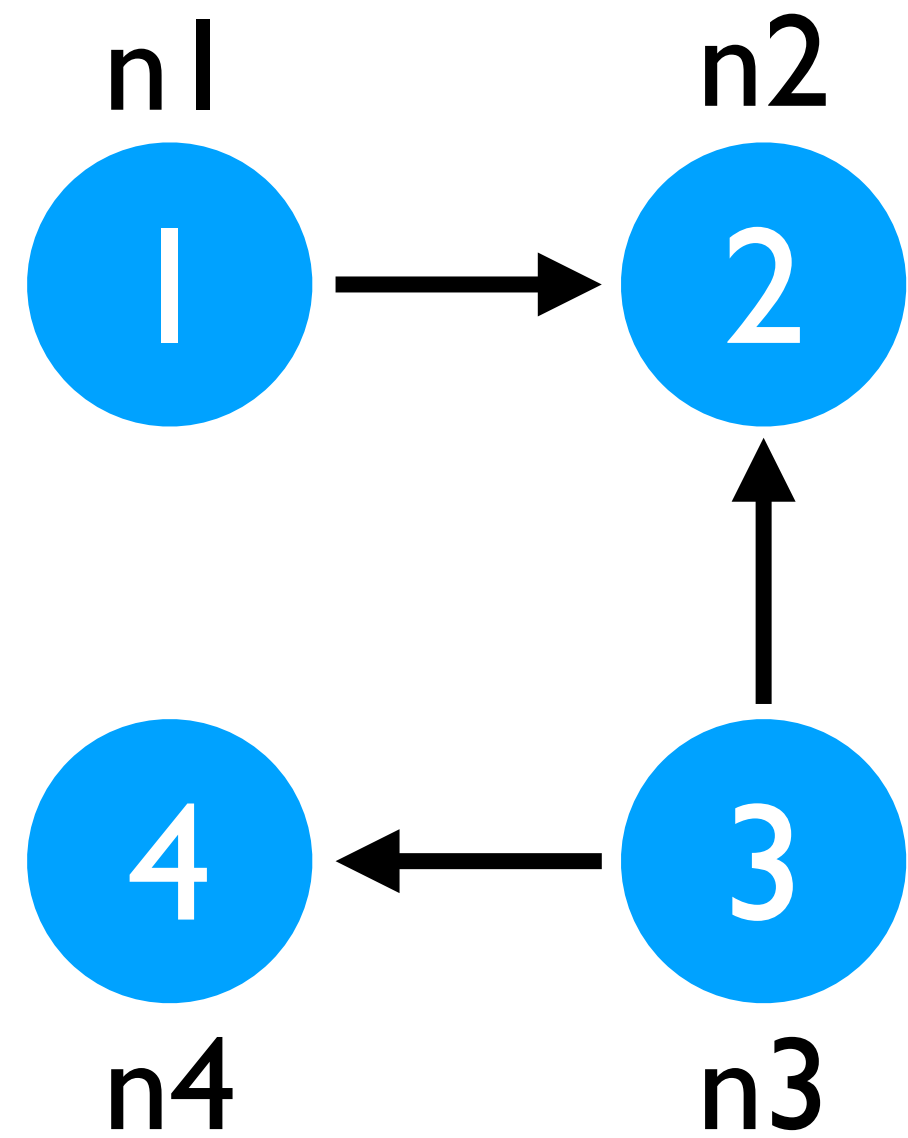
분자 구조



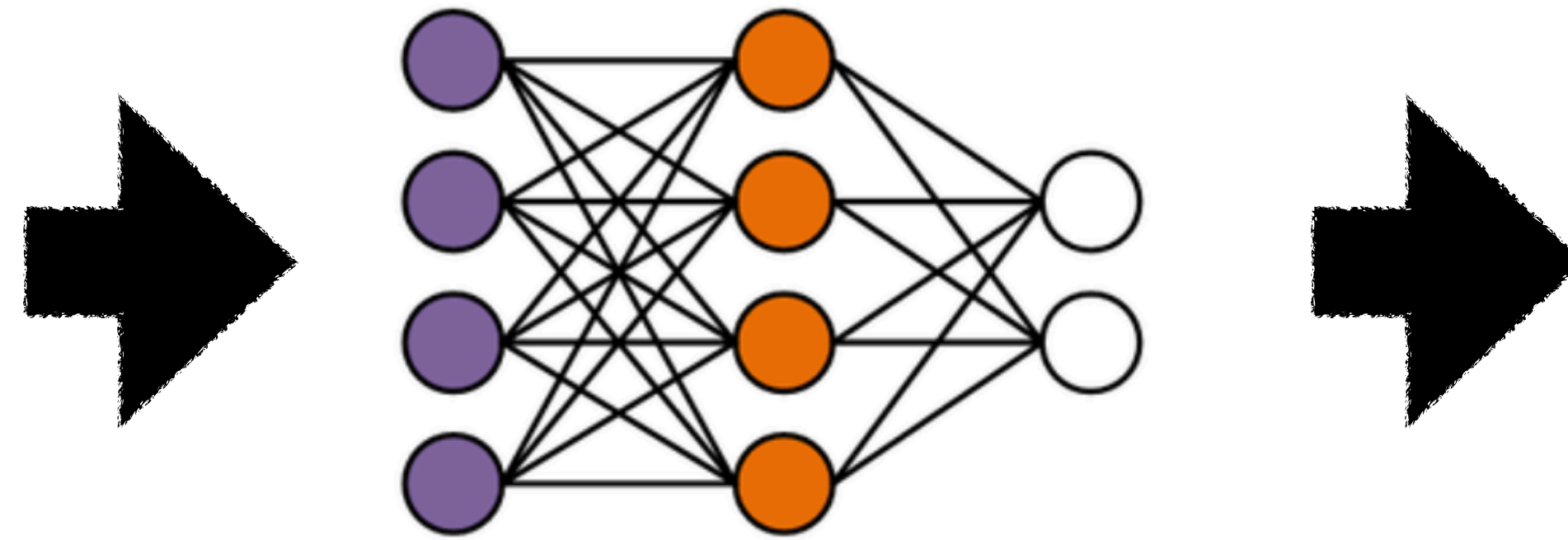
질병 분포

GNN : 가장 많이 사용하는 그래프 기계학습

- 학습된 GNN은 그래프 데이터가 들어오면 분류를 함



예시 그래프 데이터



학습된 GNN 모델

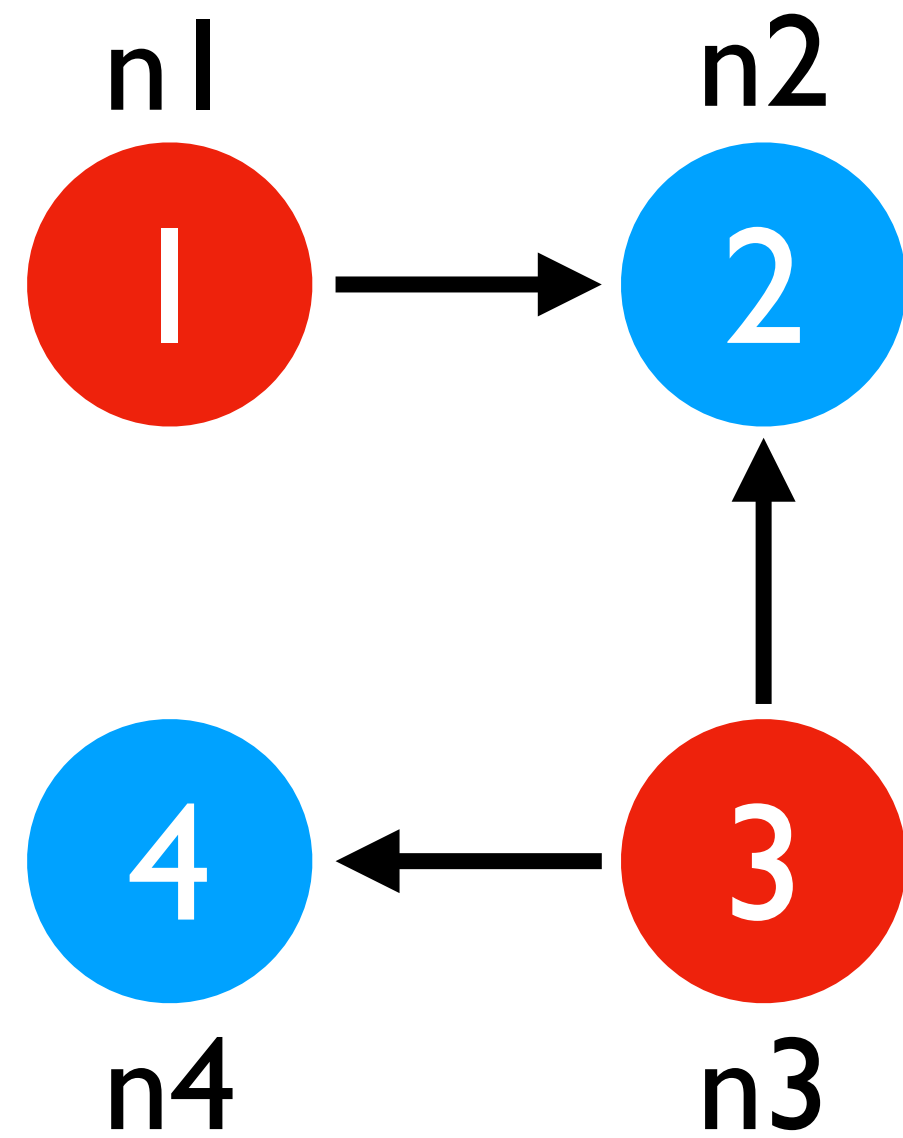
Label 1: {n1, n3}

Label 2: {n2, n4}

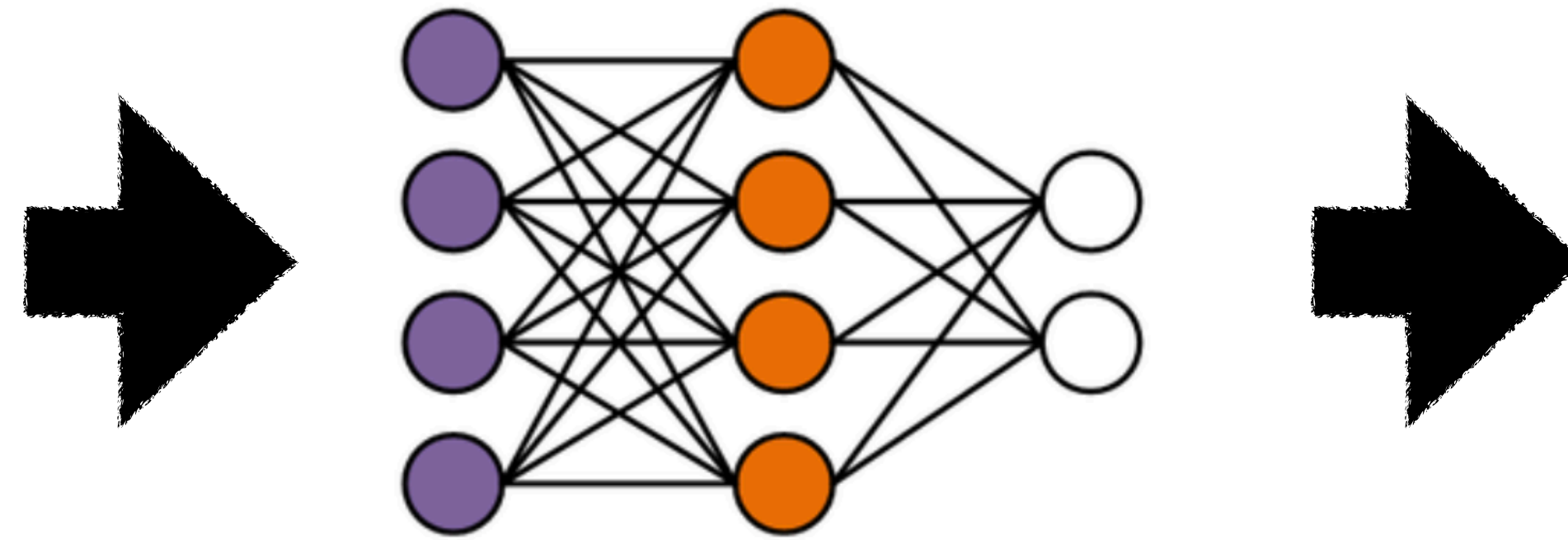
분류 결과

GNN : 가장 많이 사용하는 그래프 기계학습

- 학습된 GNN은 그래프 데이터가 들어오면 분류를 함



예시 그래프 데이터



학습된 GNN 모델

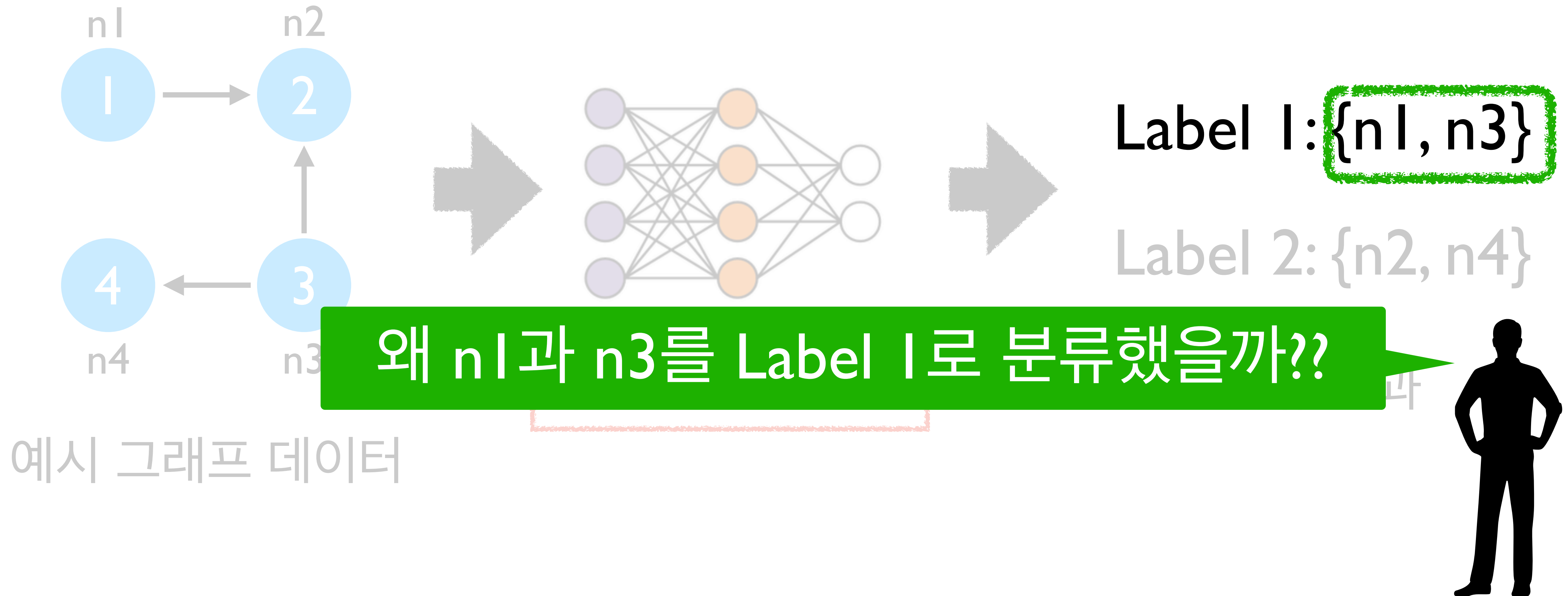
Label 1: {n1, n3}

Label 2: {n2, n4}

분류 결과

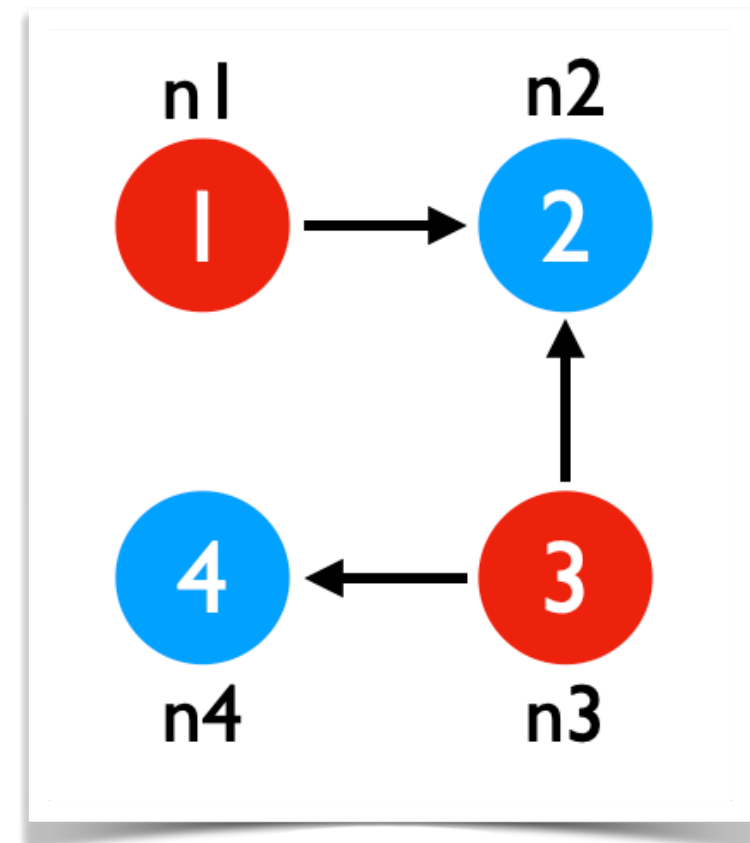
GNN 설명 기술의 필요성

- 사용자가 예측 이유를 궁금해할 수 있음

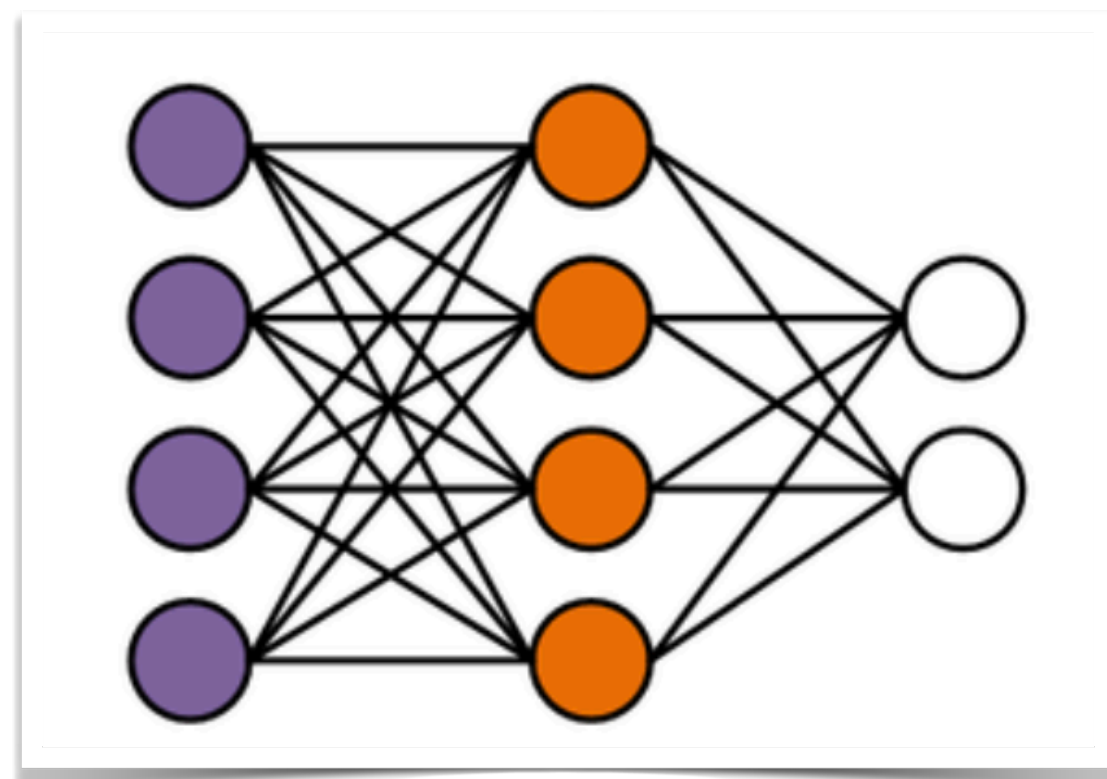


표준 설명 방식 : 서브그래프 제시

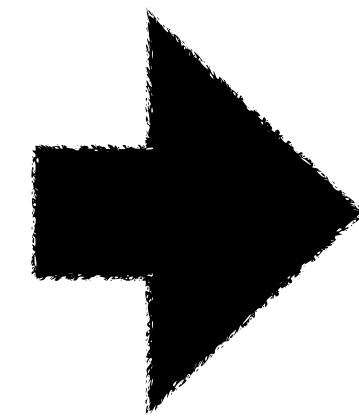
- GNN 설명 기술은 예측의 이유로 생각되는 서브그래프를 제시함



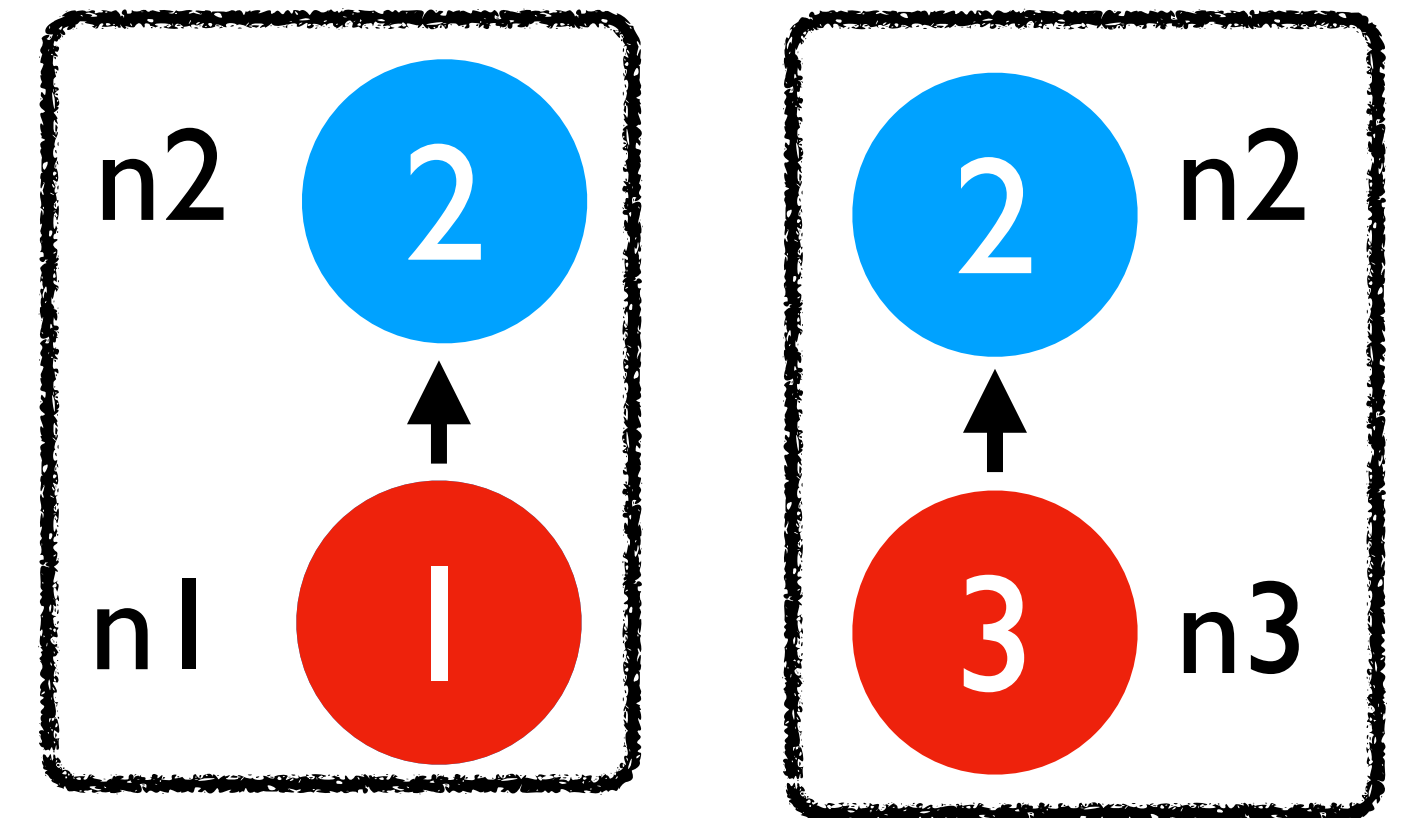
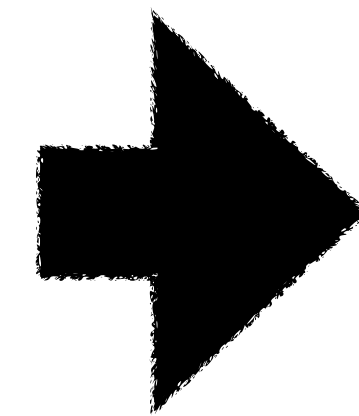
설명해야할 데이터



학습된 GNN 모델



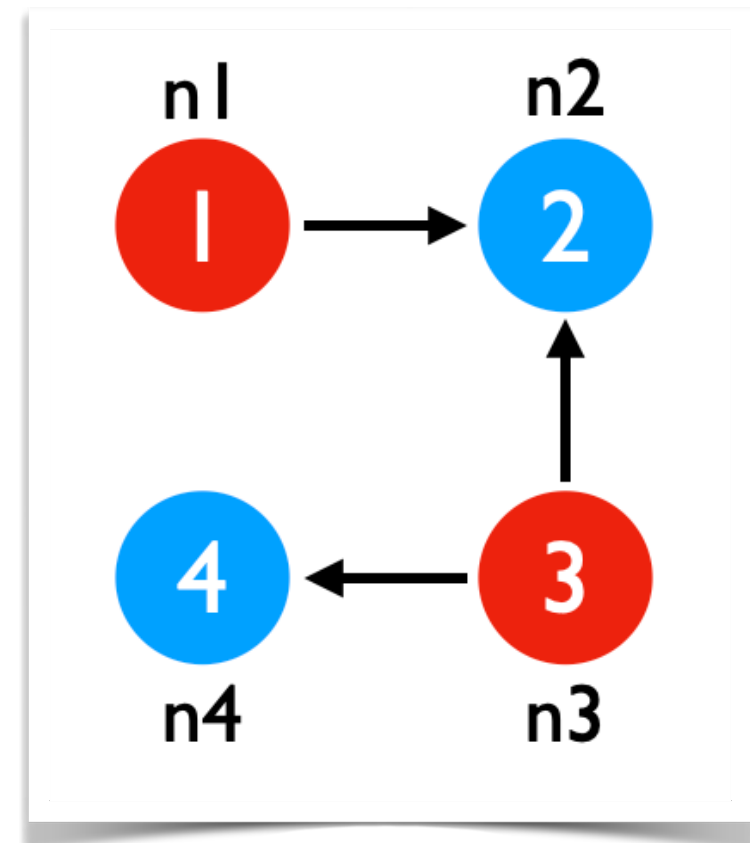
GNN
Explainer



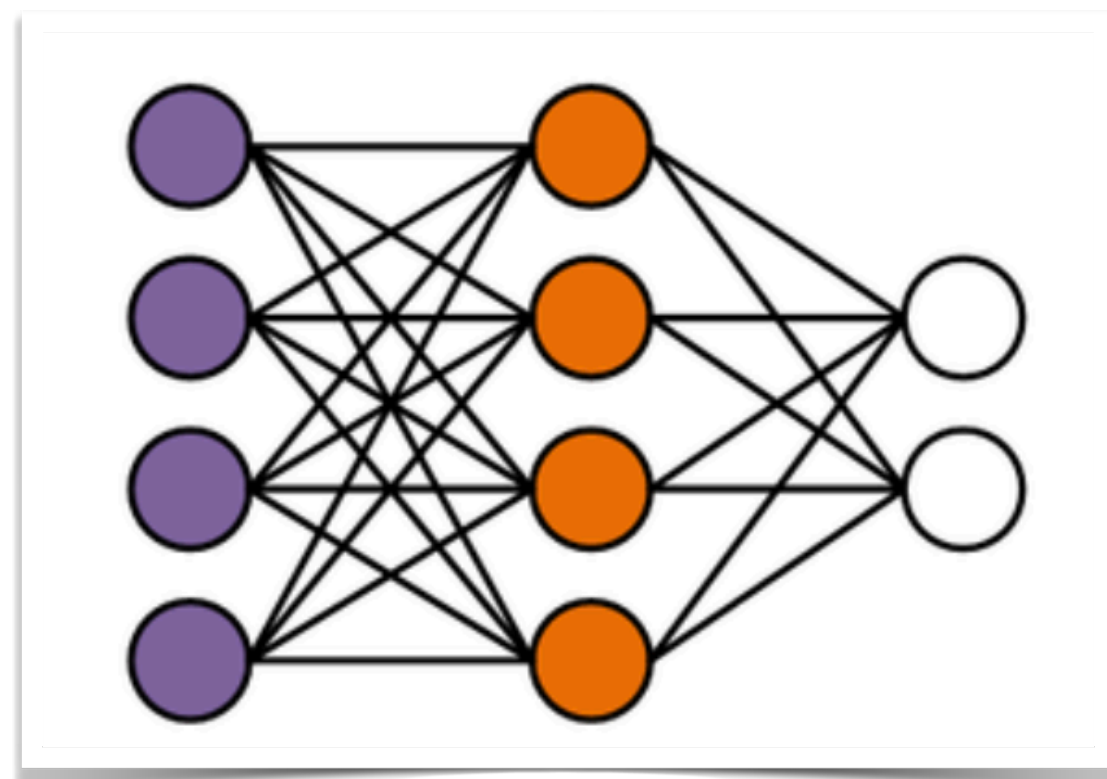
제공된 설명

표준 설명 방식 : 서브그래프 제시

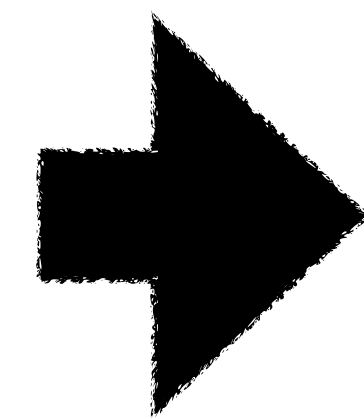
- GNN 설명 기술은 예측의 이유 **n3이 label 1로 분류된 이유에 대한 설명**



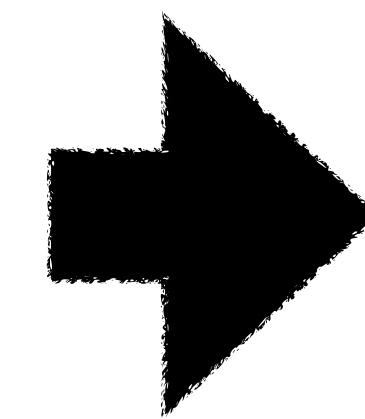
설명해야할 데이터



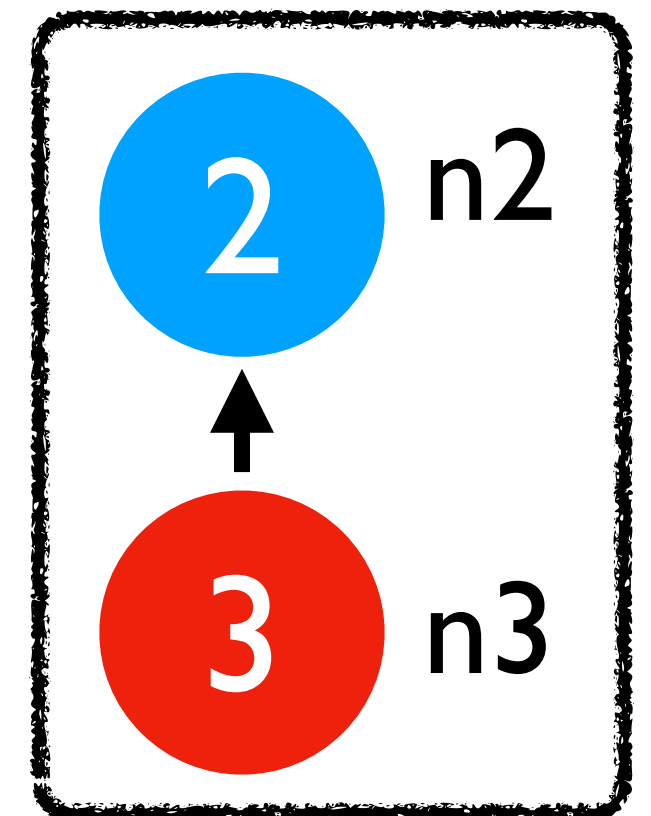
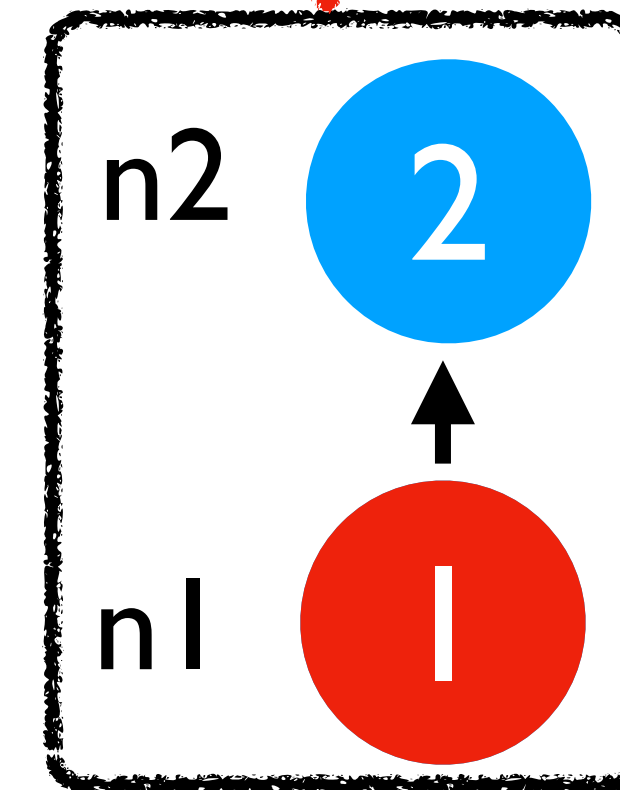
학습된 GNN 모델



**GNN
Explainer**



n1이 label 1로 분류된 이유에 대한 설명



제공된 설명

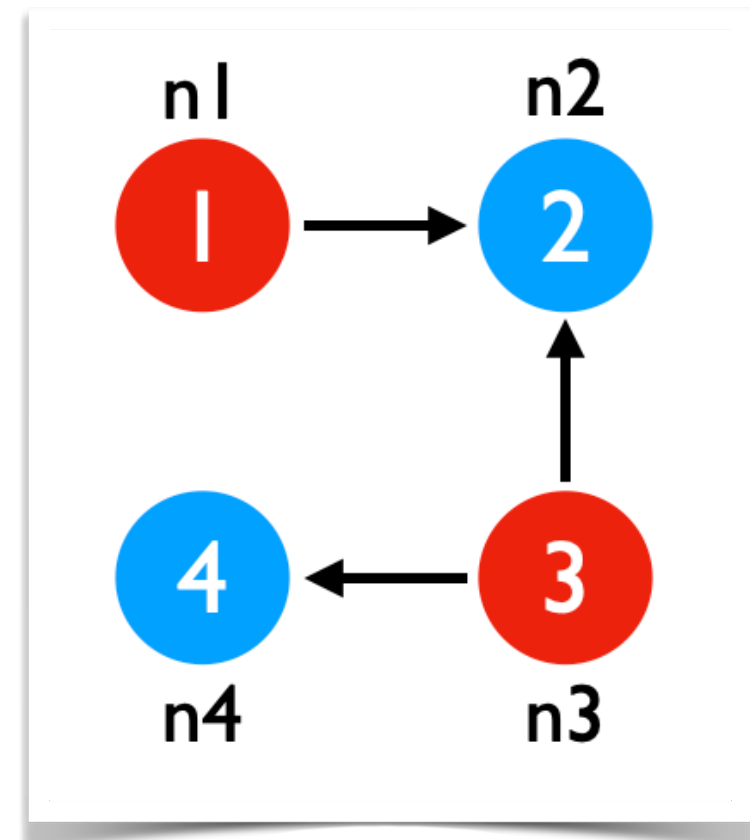
표준 설명 방법의 문제점

유저는 매번 새로운 설명을 이해해야 하는 부담이 있음

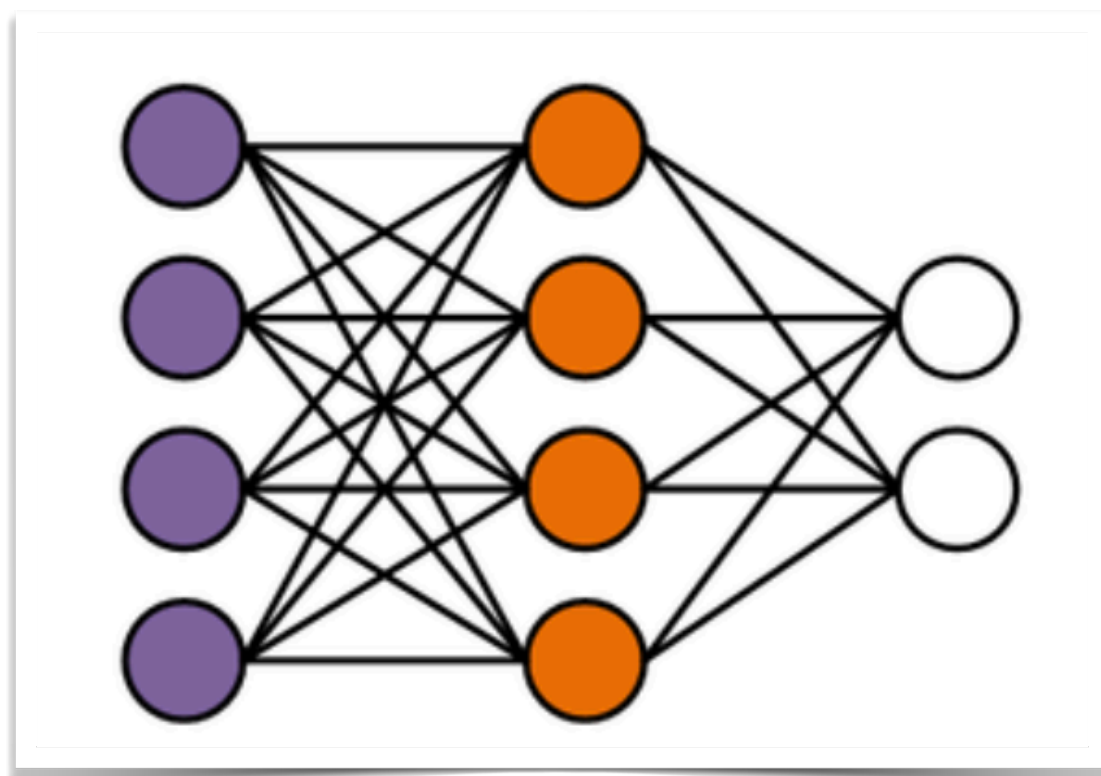
- GNN 설명 기술은 예측의 이유를 설명하는 방법

문제점

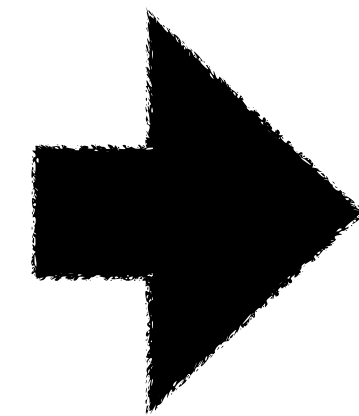
각 예측에 대해서 서로 다른 설명을 제시함



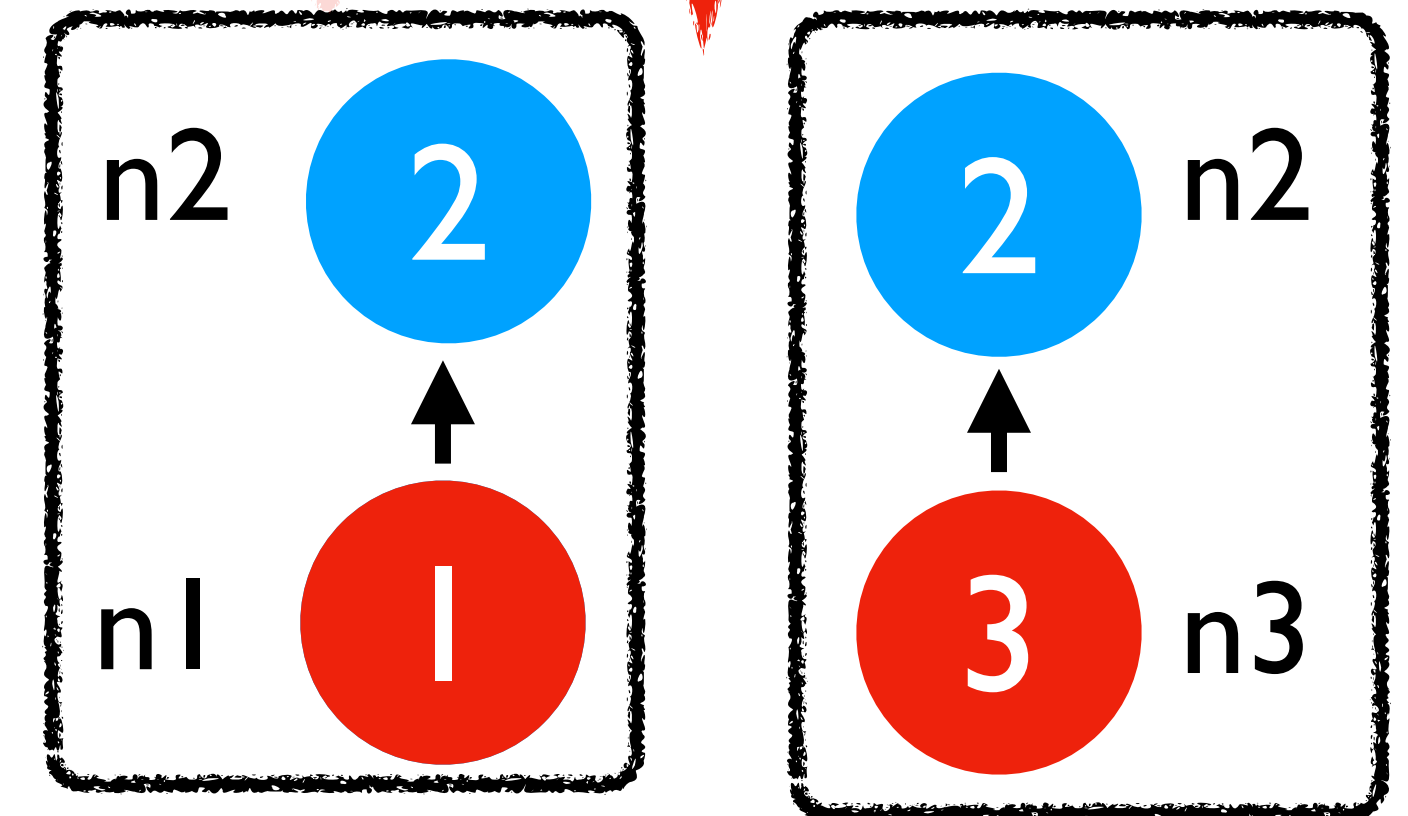
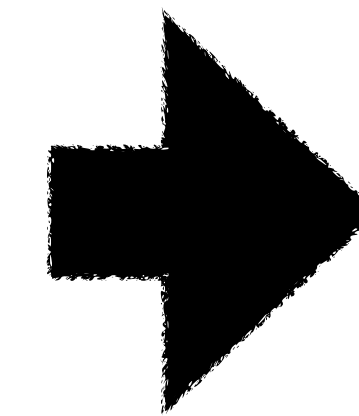
설명해야할 데이터



학습된 GNN 모델



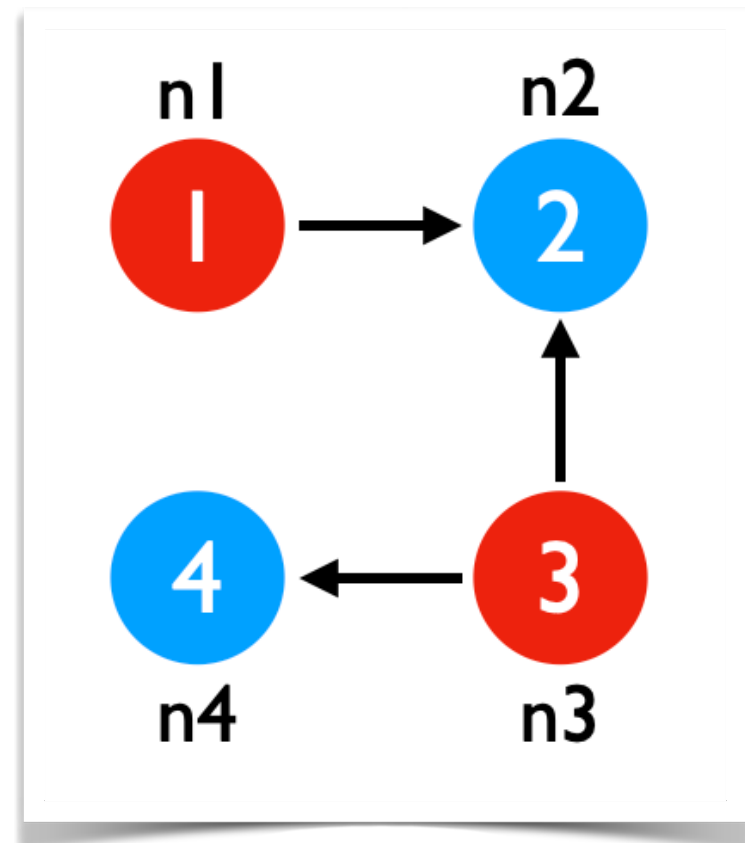
GNN
Explainer



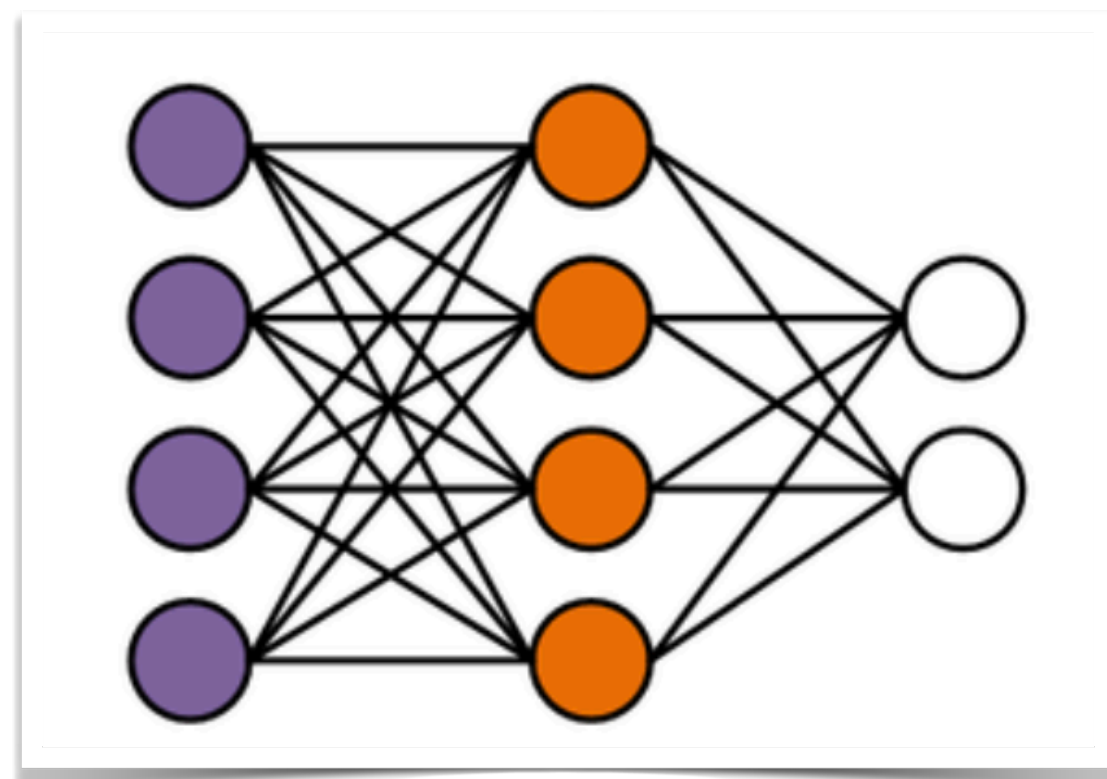
제공된 설명

우리의 아이디어 : 요약 서브그래프

- 쉽게 일반화 가능한 요약 서브그래프 설명 제시



설명해야할 데이터



학습된 GNN 모델

Successor의 feature가 2.5이하이면 GNN이 label 1로 분류하는거 같다

GNN
Explainer

$[-\infty, 2.5]$

$[-\infty, \infty]$

n1, n3에 대한
요약 서브그래프 설명