

Graph Neural Network

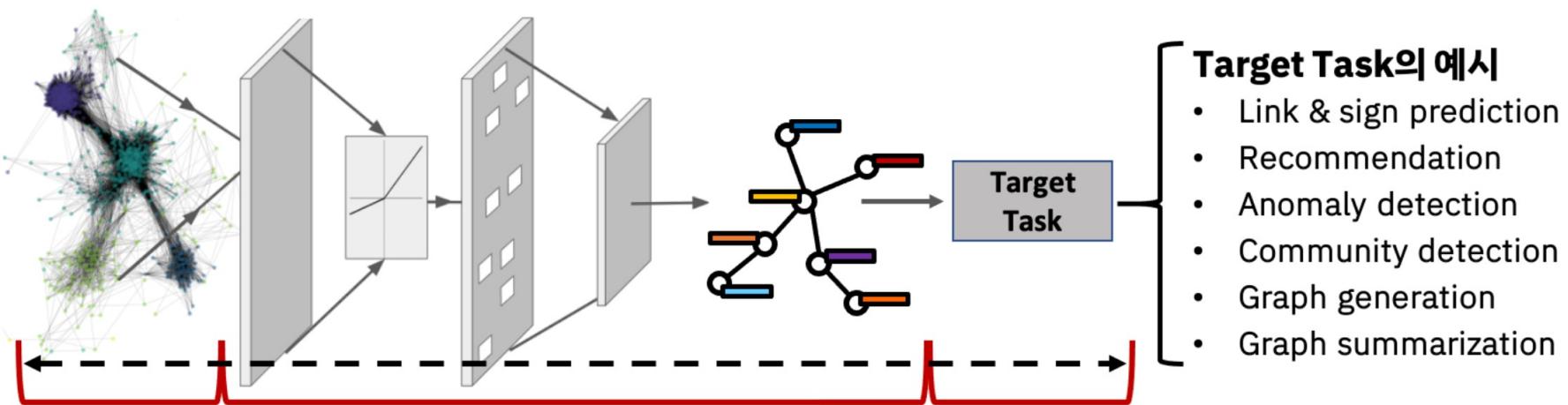
Jinhong Jung

Jeonbuk National University

jinhongjung@jbnu.ac.kr

Today's Topics

- Q1. 우리 주변에 어떠한 그래프 데이터가 있을까?
- Q2. 그래프 데이터를 어떻게 학습할 수 있을까?
- Q3. 그래프 데이터는 어떻게 활용 될 수 있을까?



Outline

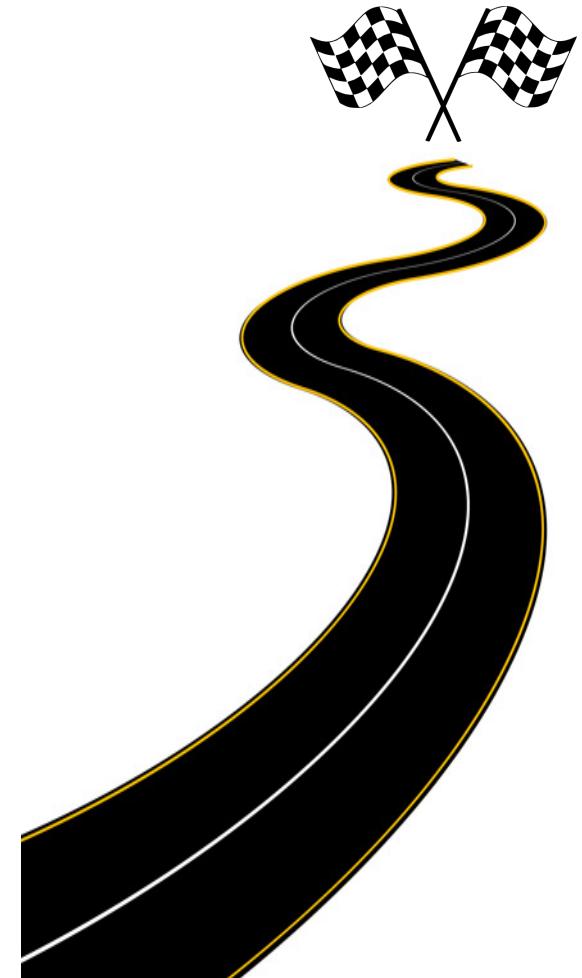
□ 실세계 그래프 데이터 소개

□ 그래프 뉴럴 네트워크 소개

□ 그래프 뉴럴 네트워크 발전 과정

□ 그래프 뉴럴 네트워크 응용

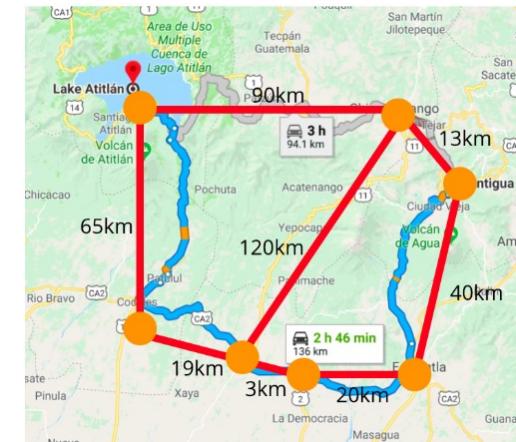
□ 데모: 그래프 정점 분류



그래프 (Graph)란?

□ 임의의 두 개체의 연결 관계를 표현하는 구조

- 예시: 친구 관계, 지도, 지하철 노선도, 전기회로

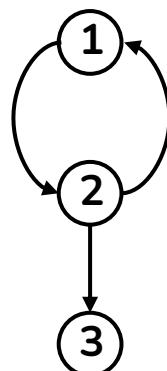


□ 그래프는 정점과 간선으로 표현

- 정점은 개체를, 간선은 개체 간의 연결 관계를 표현
- 종종 그래프를 네트워크 (network)라고도 함

□ 그래프 데이터는 인접 행렬의 형태로 저장

- 인접 행렬: 2차원 행렬로, 간선 $u \rightarrow v$ 에 대해 $A_{uv} = 1$ 로 저장 (없으면 0)



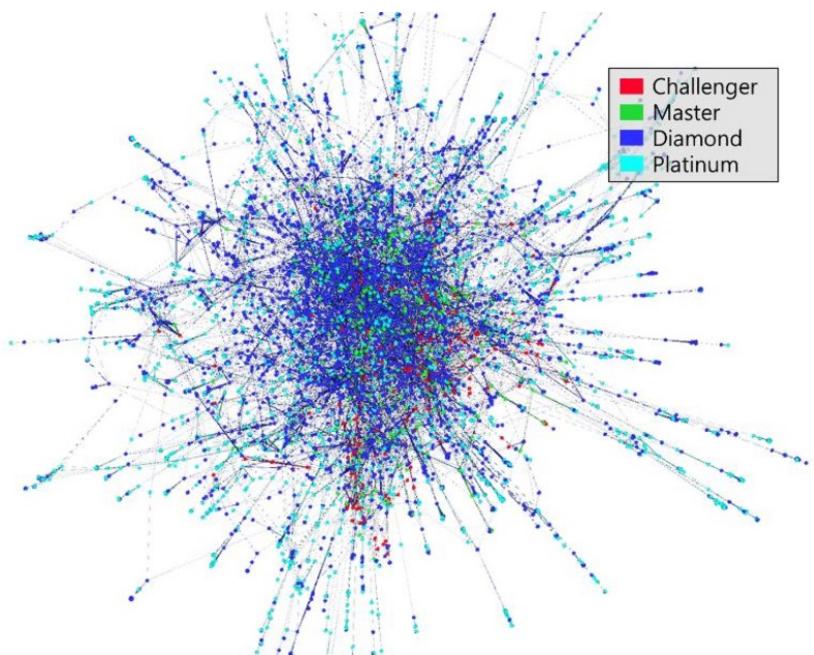
	1	2	3
1	0	1	0
2	1	0	1
3	0	0	0

인접 행렬 A

실세계 그래프 예시 (1)

□ 소셜 네트워크 (Social Network)

- 온/오프라인 상에서 친구 관계를 나타내는 그래프

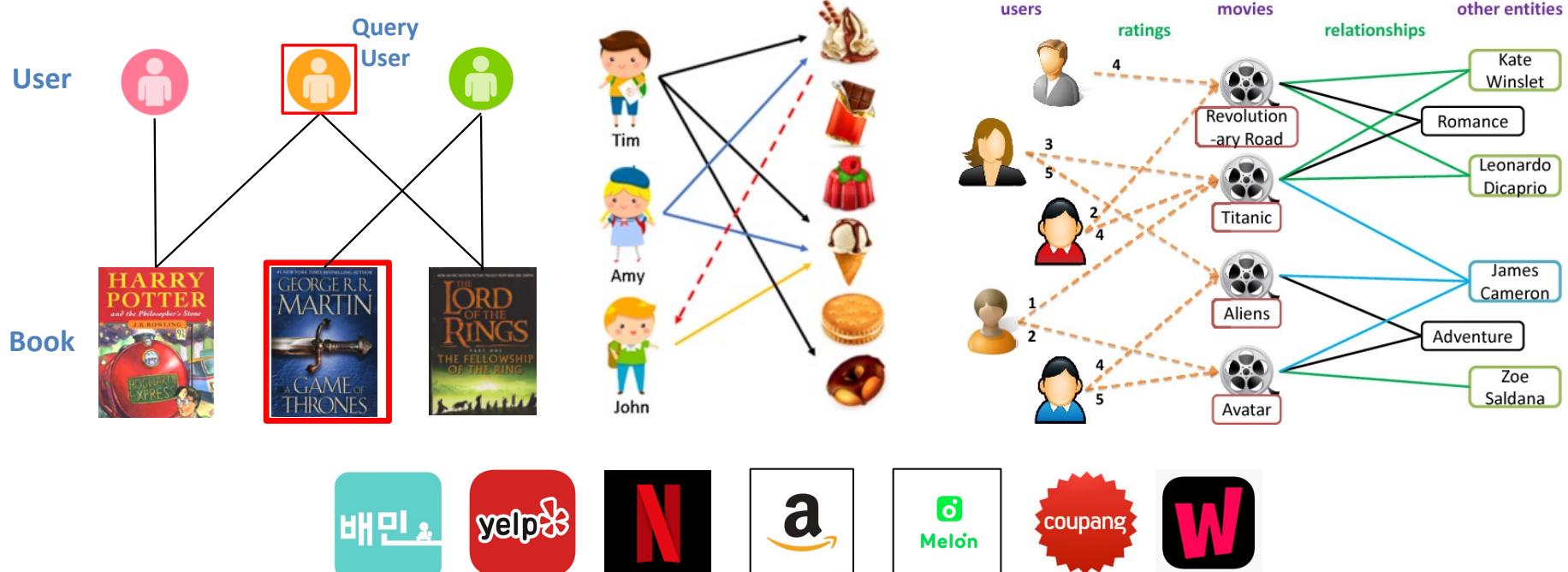


온라인 게임 상에서 고레벨 유저간의
소셜 네트워크

실세계 그래프 예시 (2)

□ 유저-아이템 네트워크 (User-Item Network)

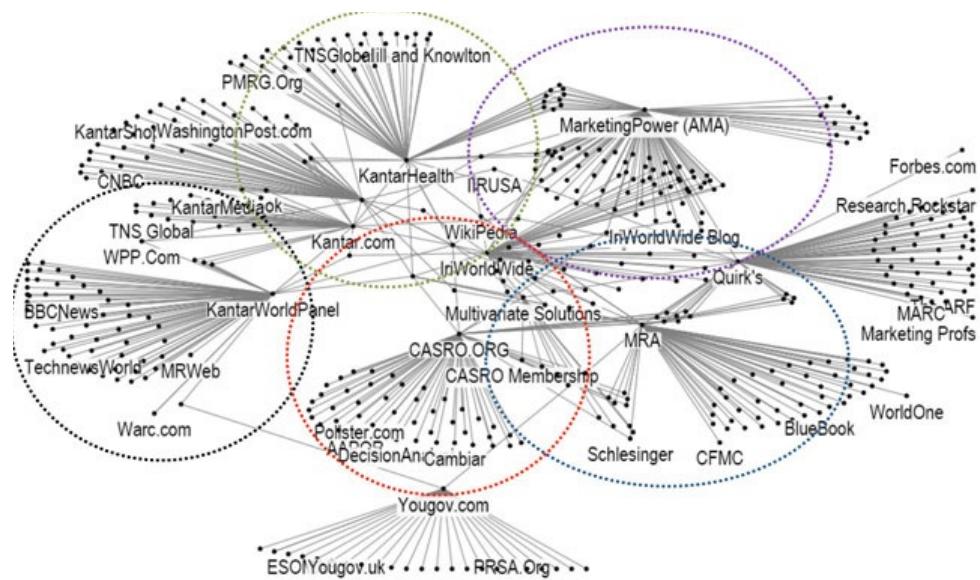
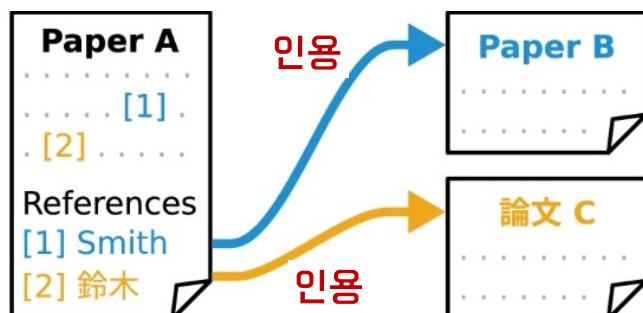
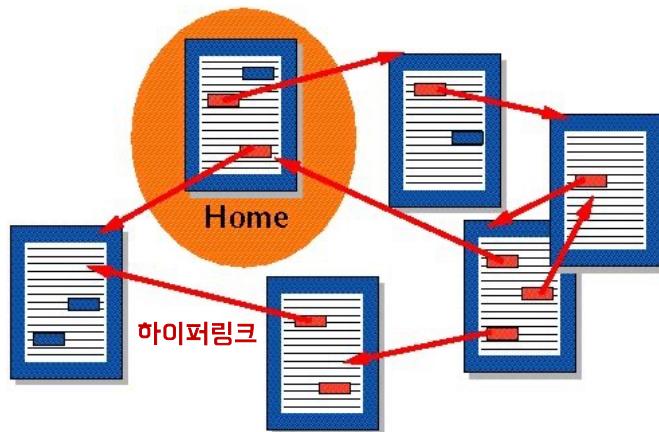
- 유저와 아이템의 상호작용 (상품 구매, 영화 평점 등)을 표현하는 그래프
- 아이템은 상품, 영화, 뉴스, 음악, 논문, 식당, 책 등 다양함



실세계 그래프 예시 (3)

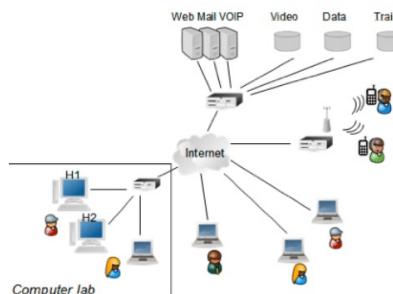
□ 문서 네트워크 (Document Network)

- 문서간의 연결 관계를 나타내는 그래프
- 웹에서는 하이퍼링크, 논문에서는 인용으로 연결 표현

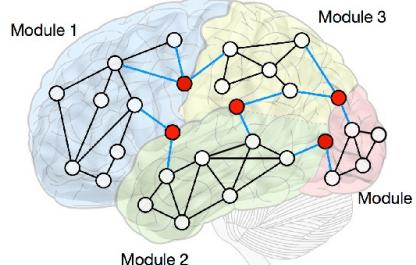


실세계 그래프 예시 (4)

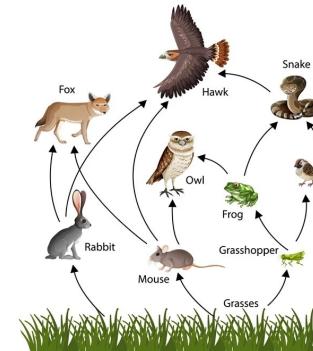
□ 다양한 실세계 복잡계가 그래프로 표현됨!



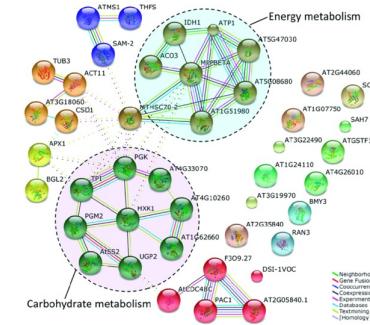
Machine



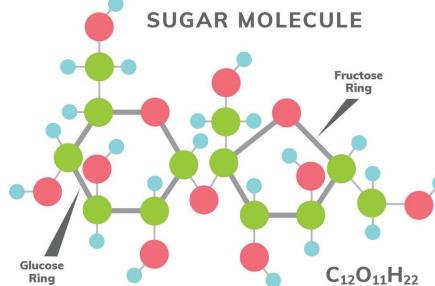
Brain



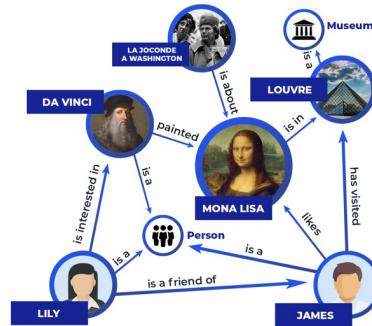
Food chain



Gene



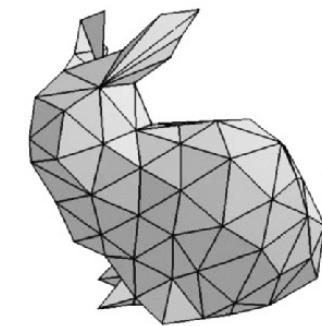
Molecule



Knowledge



Point cloud



3D mesh

그래프 데이터는 어떻게 학습 할 수 있을까?

Outline

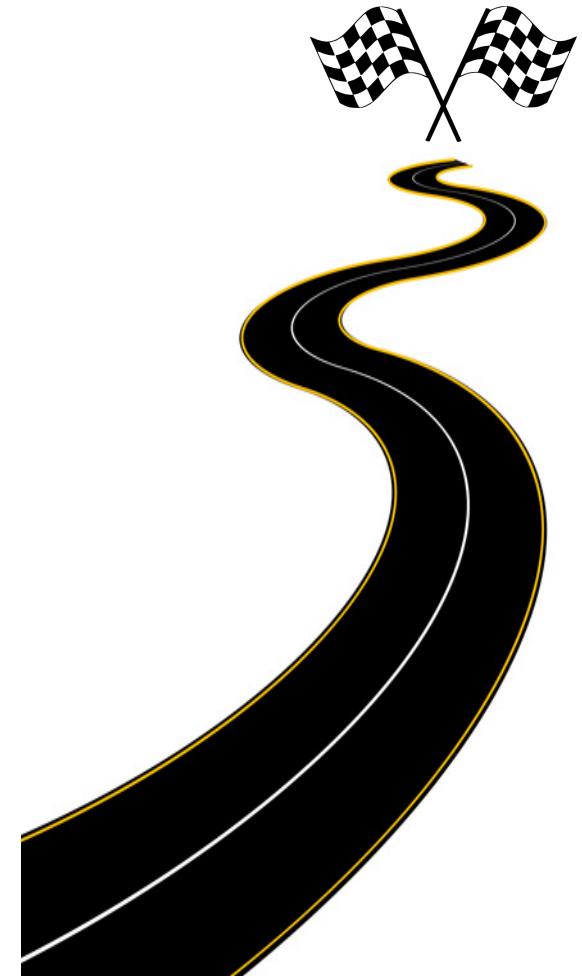
□ 실세계 그래프 데이터 소개

□ **그래프 뉴럴 네트워크 소개**

□ 그래프 뉴럴 네트워크 발전 과정

□ 그래프 뉴럴 네트워크 응용

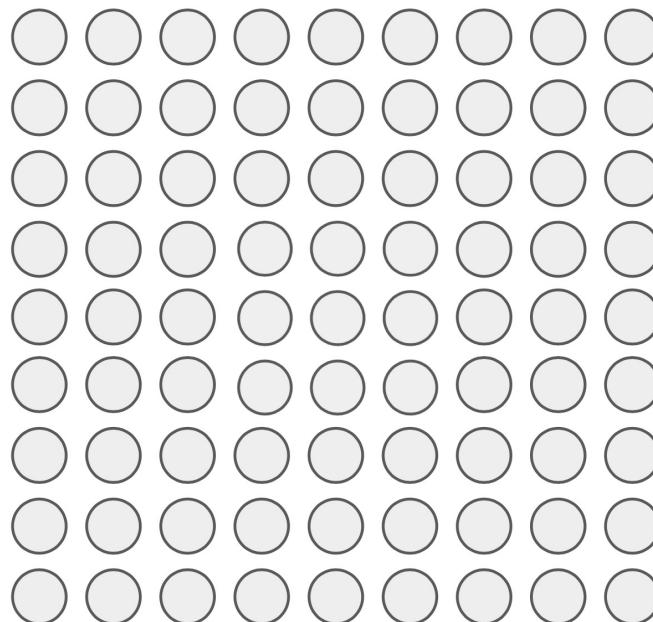
□ 데모: 그래프 정점 분류



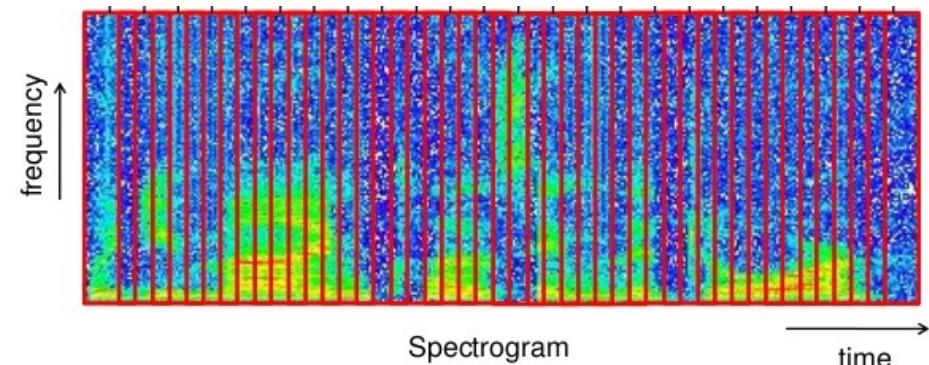
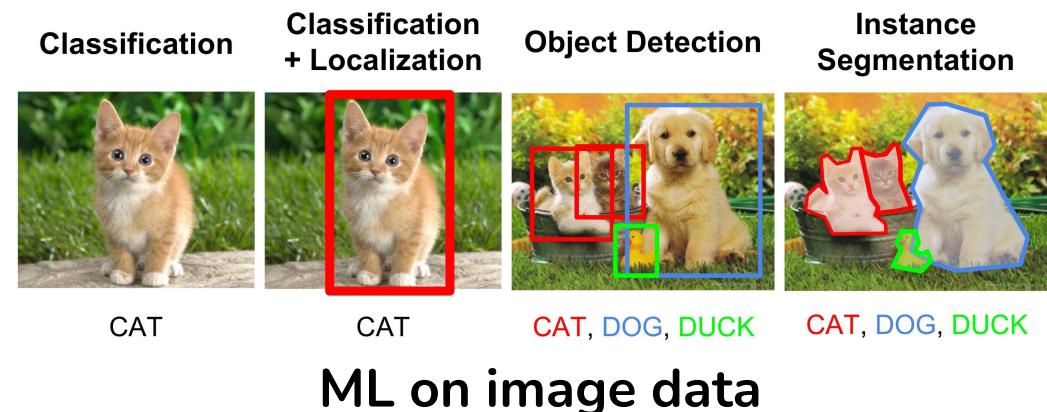
전통적인 데이터 학습 방법 (1)

□ 전통적인 딥 러닝 모델은 격자 (grid) 구조의 데이터에서 학습

- 예) 이미지, 비디오, 텍스트, 오디오 등



격자 구조
(Multi-dimensional array)

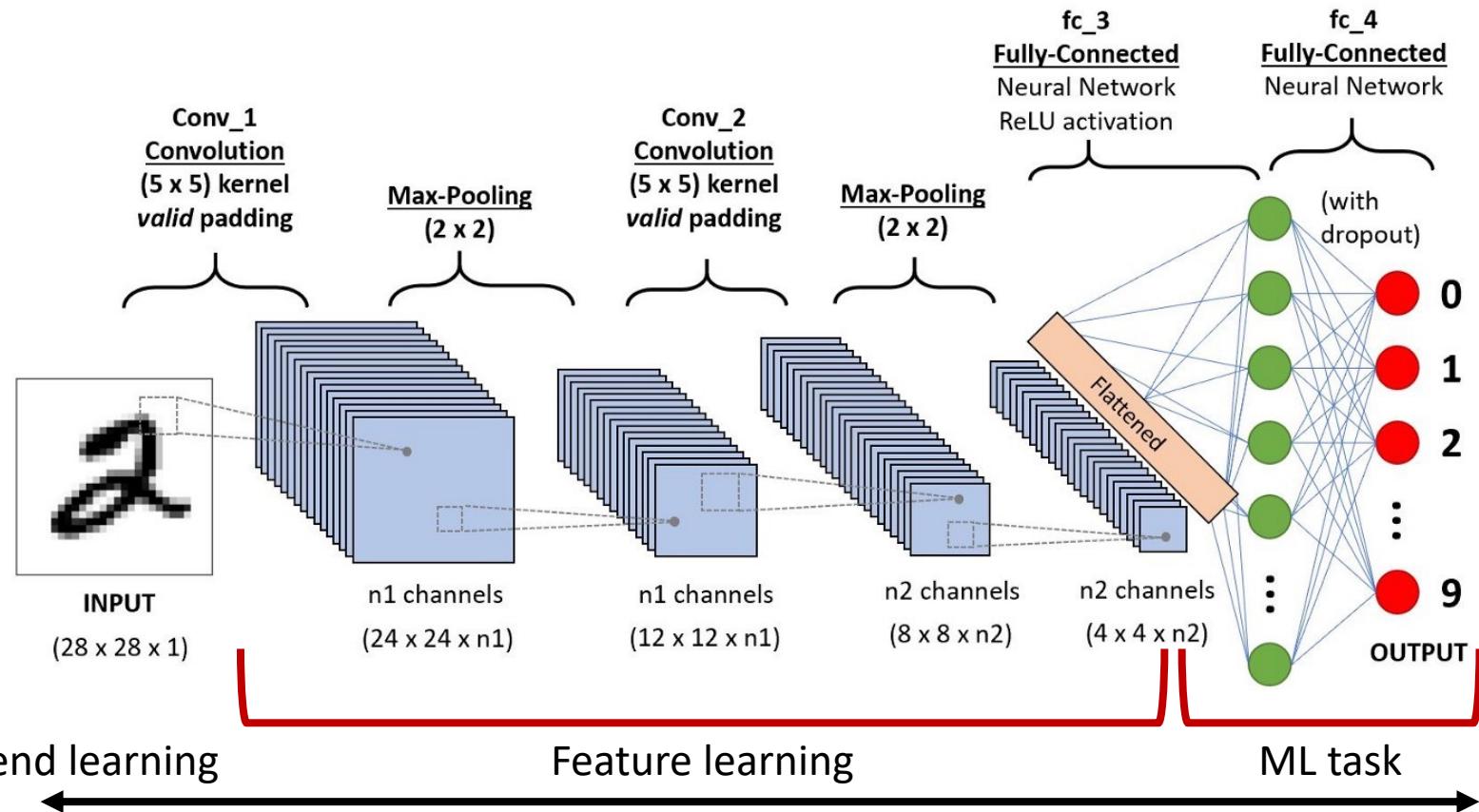


ML on audio data

전통적인 데이터 학습 방법 (2)

□ CNN은 격자 구조의 데이터에서 잘 작동함

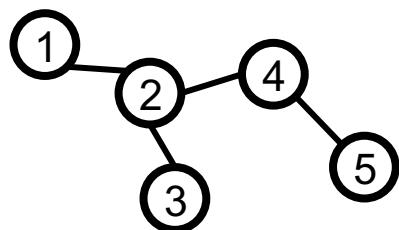
- 종단 학습 (end-to-end learning): hidden feature와 supervised task (예, 분류)를 동시에 학습 가능하여 좋은 성능을 보임



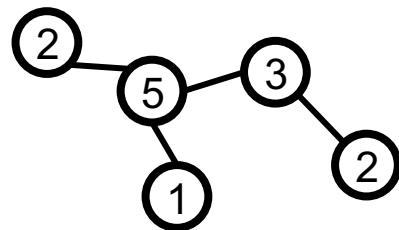
전통적인 데이터 학습 방법 (3)

▣ 그러나 전통적인 CNN은 그래프에서 잘 동작하지 않음!

- 동일한 그래프 구조이더라도 정점의 순서를 어떻게 매기느냐에 따라 인접 행렬의 모양이 다르게 됨
- 격자 구조에서는 합성곱 (convolution)이 명확하지만 그래프에서의 합성곱 적용 방법이 명확하지 않음

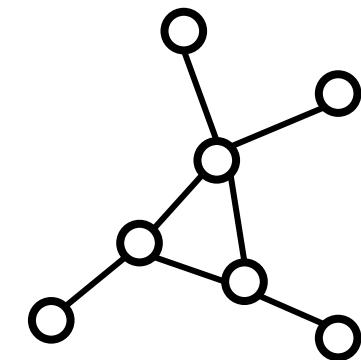
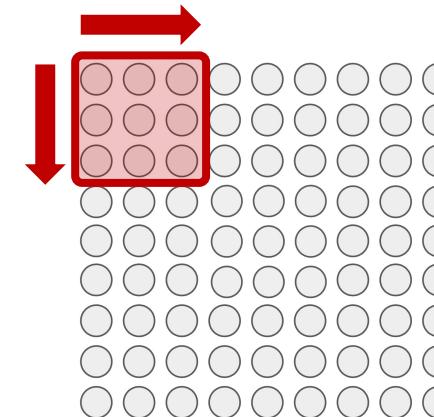


$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$



$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

그래프 구조는 같지만 인접 행렬은 다름

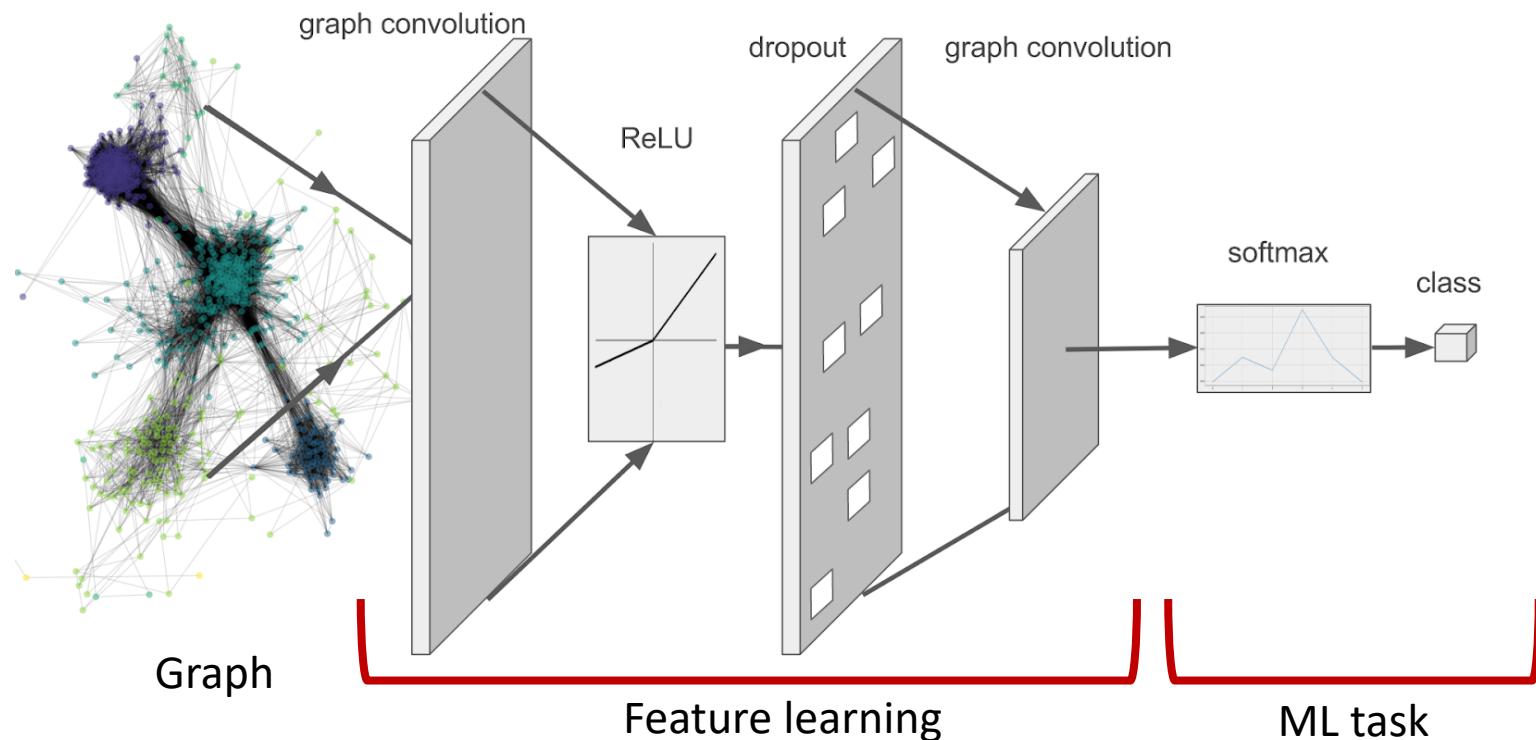


그래프에서 합성곱은 어떻게 해야 하는가?

Graph Convolution Network

□ 그래프 데이터 학습을 위한 그래프 뉴럴 네트워크 기법

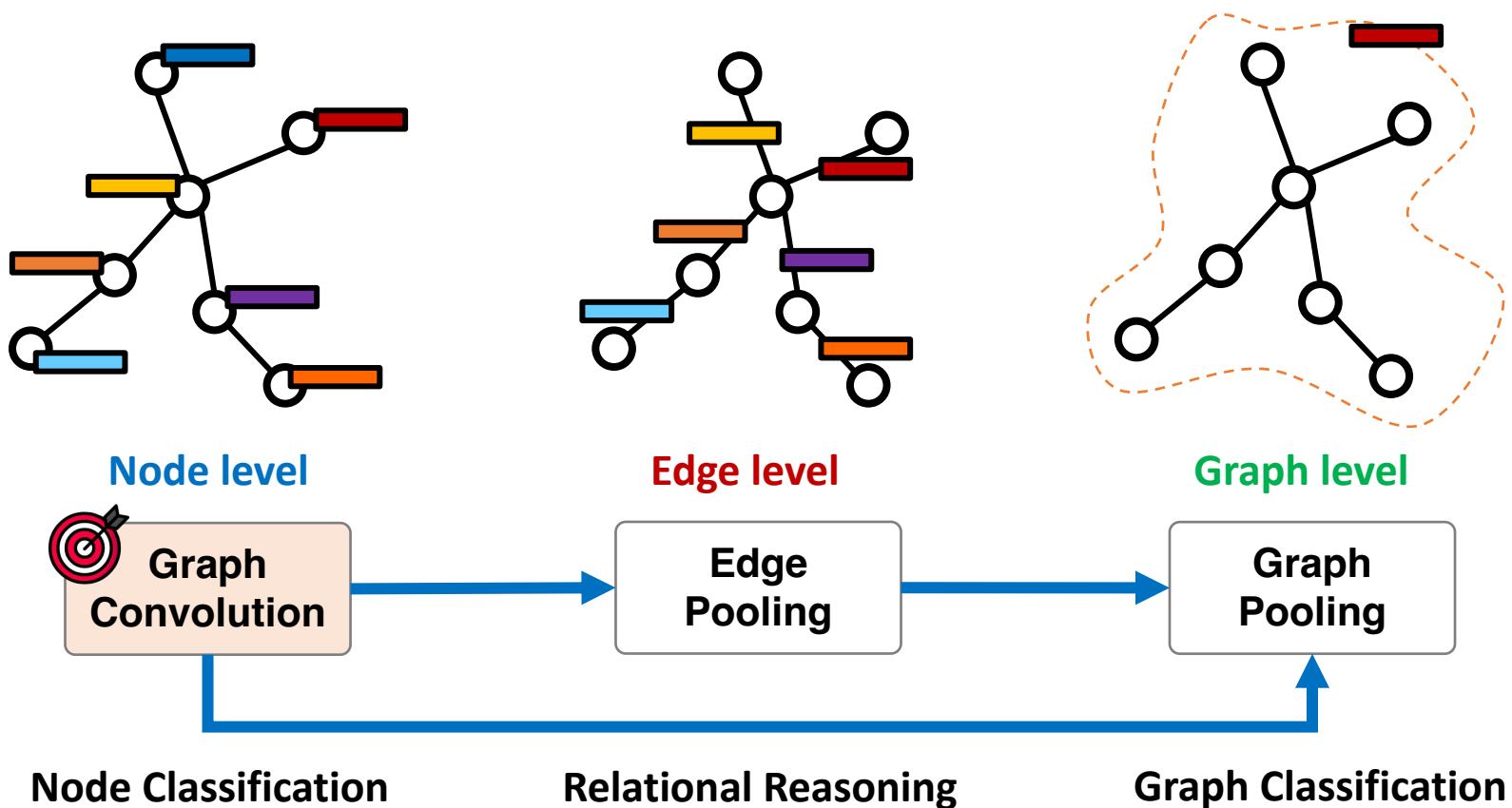
- 그래프에 특화된 합성곱 (graph convolution)을 기반으로 그래프 feature와 ML task를 **종단 학습**함



Representation Level

□ 그래프에서는 크게 세 가지 표현 레벨이 존재

- 정점, 간선, 그래프에 대한 feature vector가 학습 가능
- 본 강의에서는 정점에 대한 feature 학습 방법을 주로 살펴봄



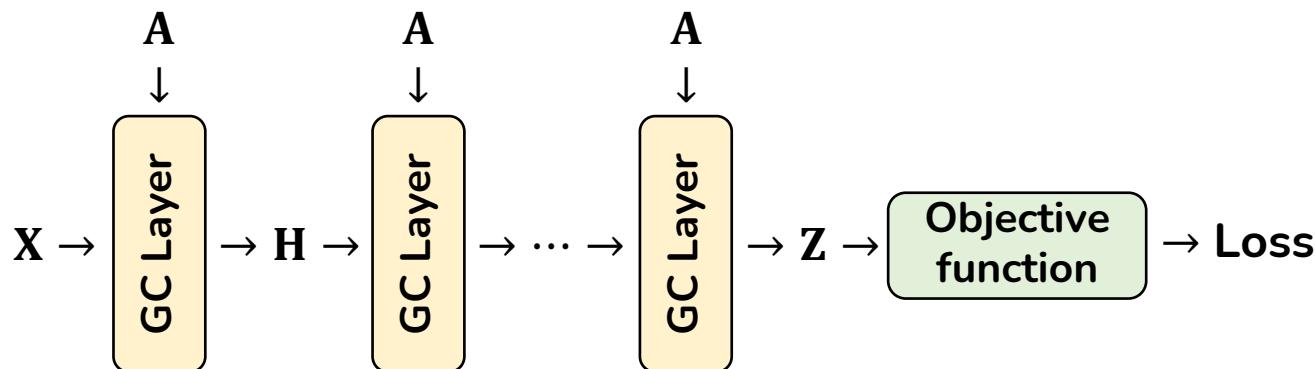
Graph Convolution Layer의 입출력

□ 입력

- 그래프의 인접 행렬 \mathbf{A} ($n \times n$ 행렬 - n 은 정점의 수)
- 초기 or 중간 단계 정점 feature 행렬 \mathbf{X} or \mathbf{H} ($n \times d$ 행렬 - d 는 feature의 차원)
 - Citation network에서의 예시: 초기 정점 feature는 각 정점이 논문이므로 feature는 논문의 초록 (abstract) 으로부터 추출되는 bag-of-words 벡터로 구성

□ 출력

- 그래프 합성곱으로 학습되는 정점 feature 행렬 \mathbf{H} ($n \times d$ 행렬 - d 는 feature의 차원)

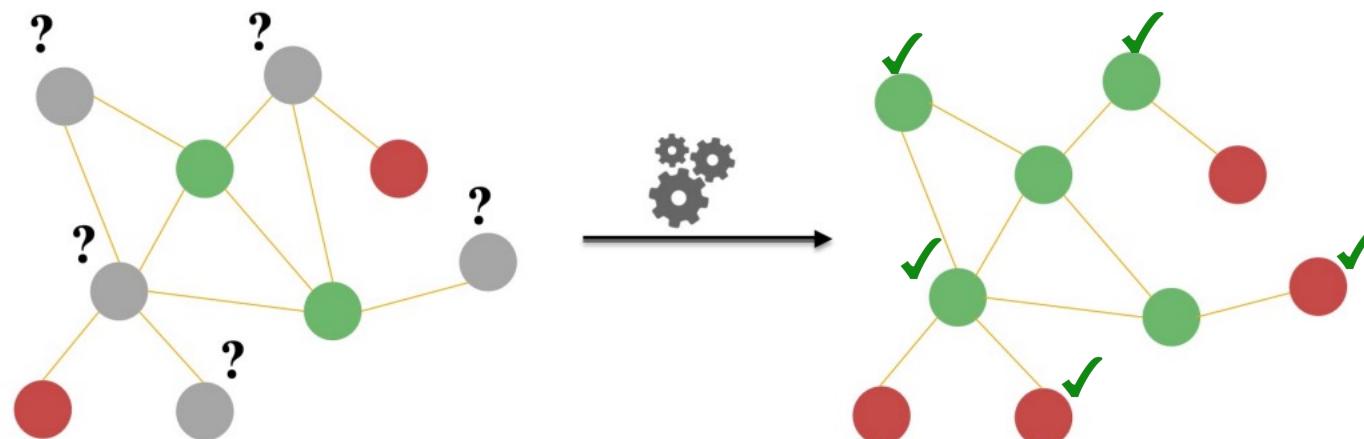


Graph Convolution Network

Graph ML 예시: 그래프 정점 분류

□ 정점에 레이블이 일부 주어졌을 때, 레이블이 없는 정점의 레이블을 예측하기

- 준지도 (semi-supervised) 학습의 일종
- Citation network에서 논문의 카테고리를 예측하는데 활용
- 레이블이 여러 개이기 때문에 multi-class cross entropy loss 사용

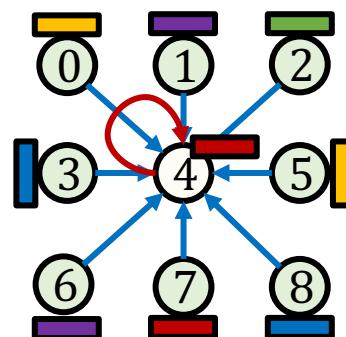
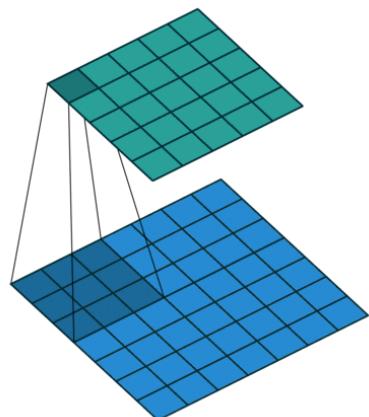


GCN [ICLR'17]

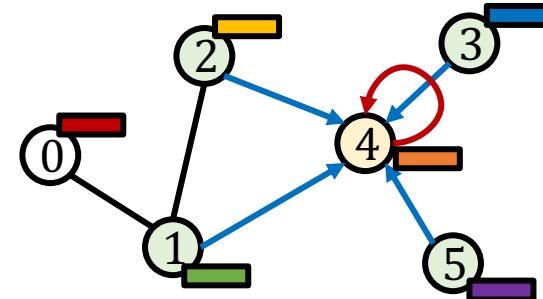
□ Main Idea of GCN

- CNN의 합성곱 (convolution)을 그래프 구조로 확장시키자!
- 격자 구조도 결국 그래프의 한 형태로 표현 할 수 있음
- 그래프에서도 목표 정점의 이웃 정점들의 feature 벡터를 취해 합성곱

Conventional convolution



Graph convolution



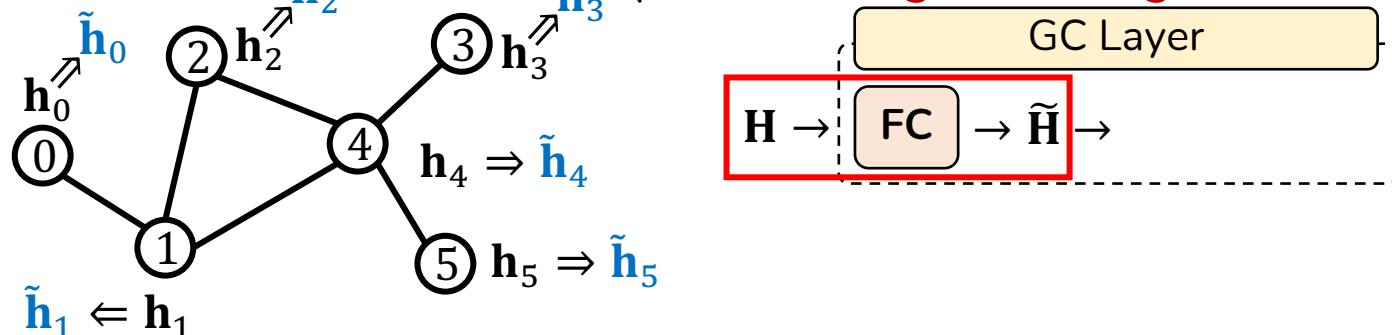
Graph Convolution의 단계 (1)

□ Step 1. 점점 feature 학습 단계 (feature transformation)

- \mathbf{h}_i 를 i 번째 정점의 feature 벡터라고 했을 때, 모든 정점에 대해 FC (fully-connected) 레이어를 통과시켜 학습

$$\tilde{\mathbf{h}}_i \leftarrow \text{FC}(\mathbf{h}_i) := \mathbf{h}_i \mathbf{W}$$

- \mathbf{W} 는 학습 가능한 $d \times d$ 행렬로 각 레이어 별로 있음
- 모든 정점에 대해 공유되어 사용됨 (CNN의 weight sharing과 같은 개념)



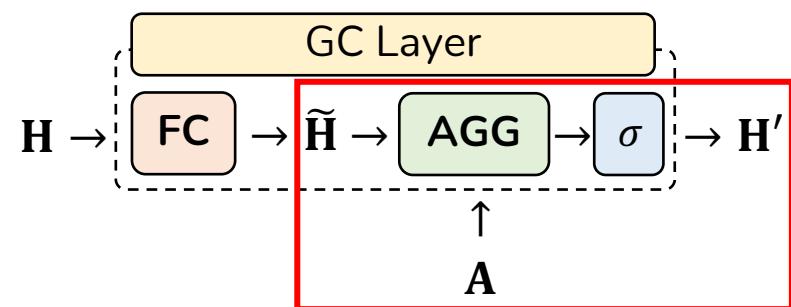
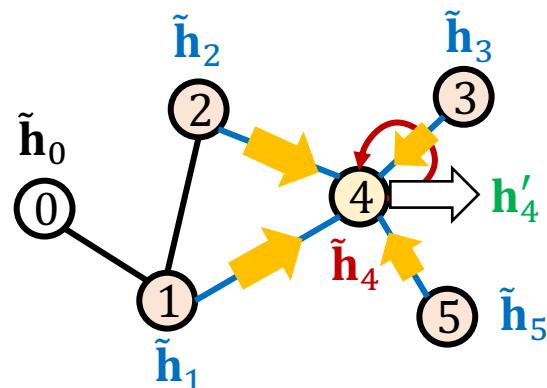
Graph Convolution의 단계 (2)

□ Step 2. 이웃 정점 feature 취합 단계 (aggregation)

- i 번째 정점의 feature에 대해, 이전 단계에서 구한 이웃 정점들의 feature 벡터와 자신의 벡터를 취합함

$$\mathbf{h}'_i \leftarrow \sigma \left(\sum_{j \in N_i} \frac{1}{c_{ij}} \tilde{\mathbf{h}}_j \right)$$

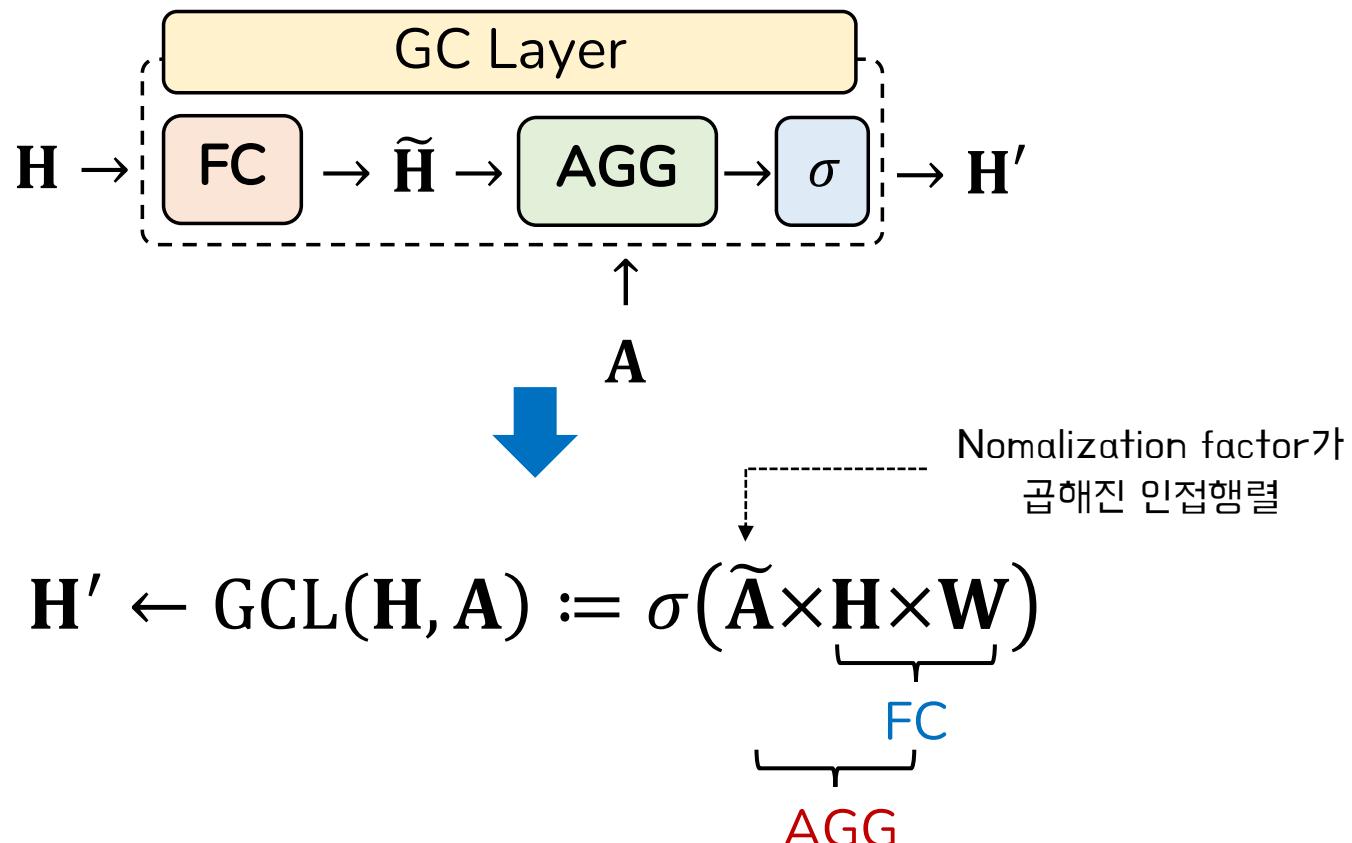
- N_i 는 정점 i 의 이웃 정점의 집합과 정점 i , σ 는 비선형 활성 함수 (예, ReLU)
- c_{ij} 는 normalization 상수 (레이어를 누적 시킴에 따라 계산되는 벡터값이 너무 커지지 않게 조절)



Graph Convolution의 행렬식

□ 그래프 합성곱 레이어의 행렬 표현 (증명은 생략)

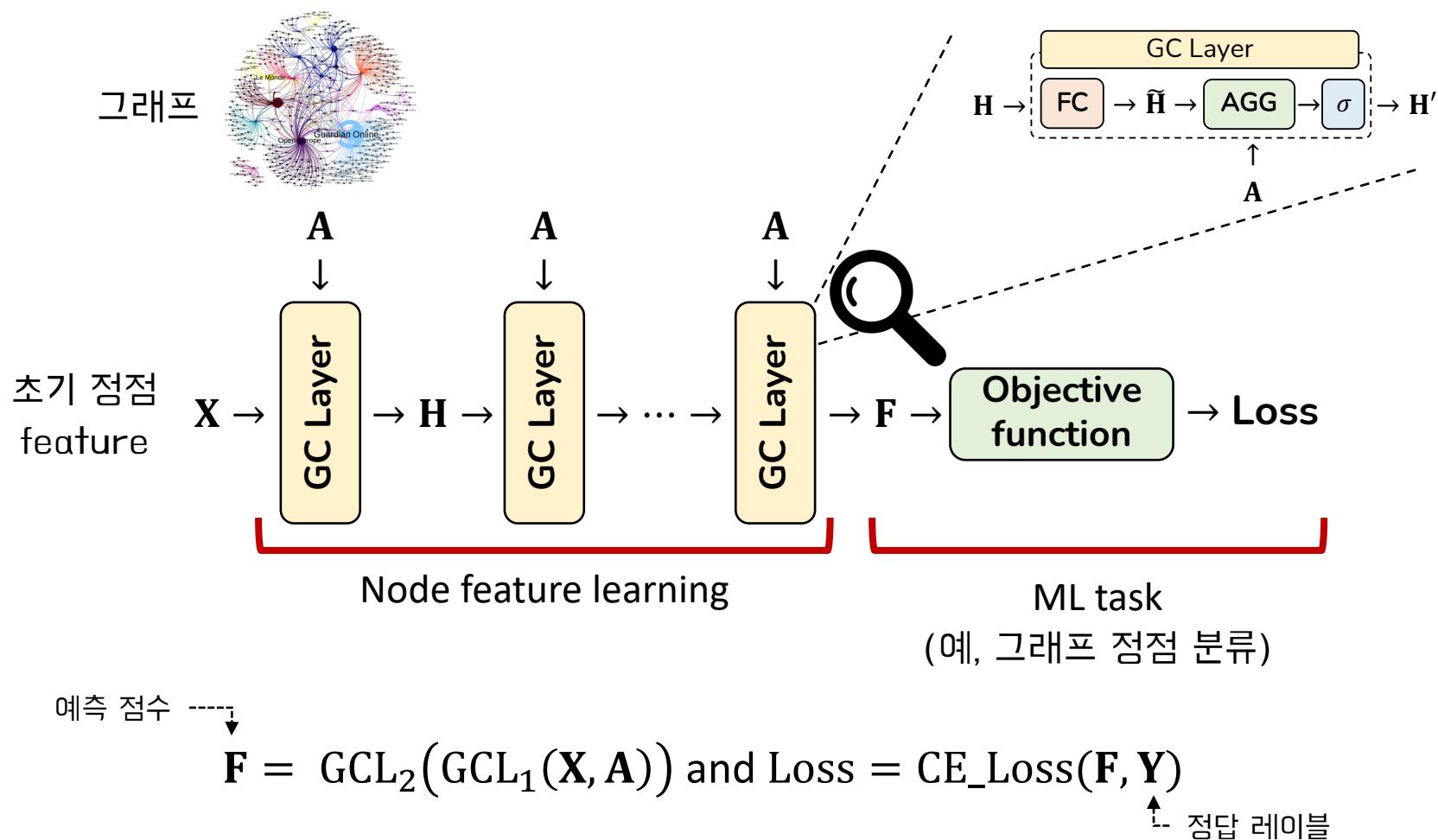
- 모든 정점에 대해 행렬로 묶어 표현하면 다음과 같음



Graph Convolution Network

□ 여러 개의 그래프 합성곱 레이어를 누적 시켜 구성

- 마지막 레이어의 결과는 ML task 목적 함수의 입력으로 사용



GCN 실험 결과

□ 그래프 정점 분류 정확도

- 2개의 GC 레이어를 활용한 GCN

Dataset	Type	Nodes	Edges	Classes	Features	Label rate
Citeseer	Citation network	3,327	4,732	6	3,703	0.036
Cora	Citation network	2,708	5,429	7	1,433	0.052
Pubmed	Citation network	19,717	44,338	3	500	0.003
NELL	Knowledge graph	65,755	266,144	210	5,414	0.001
Method		Citeseer	Cora	Pubmed	NELL	
ManiReg		60.1	59.5	70.7	21.8	
SemiEmb		59.6	59.0	71.1	26.7	
LP		45.3	68.0	63.0	26.5	
DeepWalk		43.2	67.2	65.3	58.1	
Planetoid*		64.7 (26s)	75.7 (13s)	77.2 (25s)	61.9 (185s)	
GCN		70.3 (7s)	81.5 (4s)	79.0 (38s)	66.0 (48s)	

Outline

□ 실세계 그래프 데이터 소개

□ 그래프 뉴럴 네트워크 소개

□ **그래프 뉴럴 네트워크 발전 과정**

□ 그래프 뉴럴 네트워크 응용

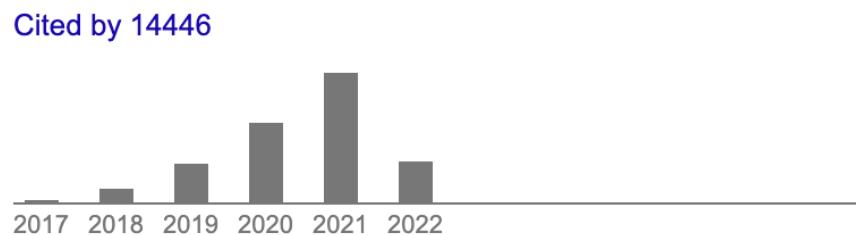
□ 데모: 그래프 정점 분류



그래프 뉴럴 네트워크의 발전

□ GCN의 발전

- GCN은 2017년에 처음 제안되어 기본적인 형태의 그래프 뉴럴 네트워크 기법으로 관련 연구 및 관심도가 꾸준히 증가



Semi-supervised classification with graph convolutional networks
TN Kipf, M Welling - arXiv preprint arXiv:1609.02907, 2016
Cited by 14446 Related articles All 19 versions

GCN 논문의 인용 추이



구글 트렌드
Graph neural network 검색어 결과

- 본 강의에서는 GCN이 어떤 방향으로 발전 중인지 “간략”하게 소개

GCN 발전: 모델 고도화 (1)

□ 기본 GCN의 각 단계에서 부분적으로 고도화를 시도

□ GraphSAGE [NeurIPS'17]

- 이웃 정점의 취합 단계를 단순 덧셈에서 학습 가능한 함수로 일반화 시킴

$$\mathbf{h}'_i \leftarrow \sigma \left(\sum_{j \in N_i} \frac{1}{c_{ij}} \mathbf{h}_j \mathbf{W} \right) \quad \rightarrow \quad \mathbf{h}'_i \leftarrow \sigma(\text{AGG}(\{\mathbf{h}_j\}) \times \mathbf{W})$$

GCNGraphSAGE

AGG함수는 학습 (예, LSTM)

□ GAT [ICLR'18]

- Attention 개념을 도입해 간선의 가중치를 학습

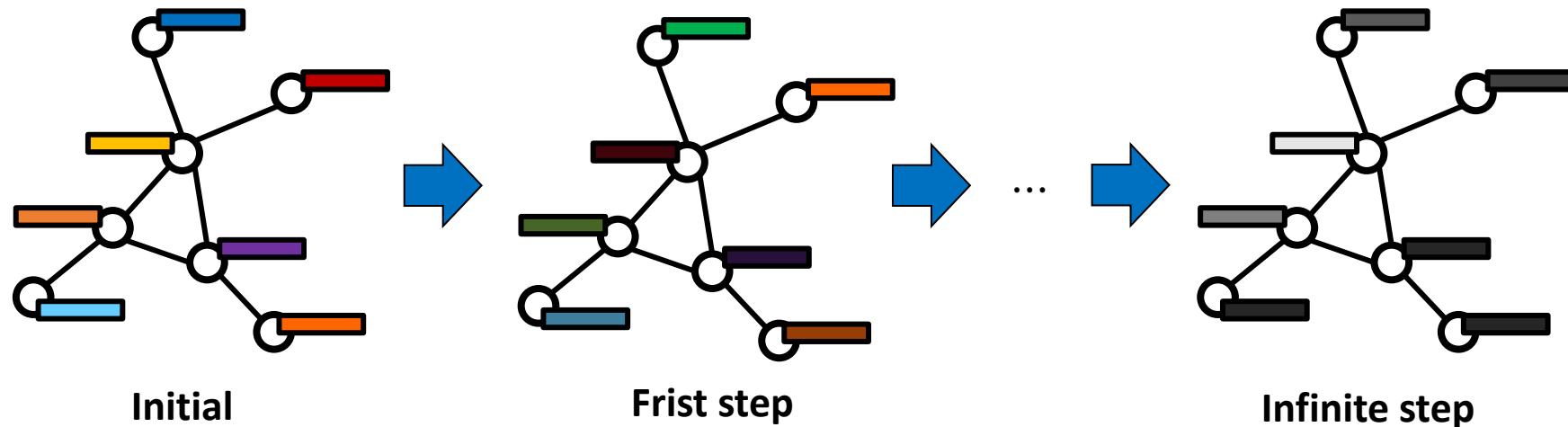
$$\mathbf{h}'_i \leftarrow \sigma \left(\sum_{j \in N_i} \frac{1}{c_{ij}} \mathbf{h}_j \mathbf{W} \right) \quad \rightarrow \quad \mathbf{h}'_i \leftarrow \sigma \left(\sum_{j \in N_i} \text{att}(\tilde{\mathbf{h}}_i, \tilde{\mathbf{h}}_j) \times \mathbf{h}_j \mathbf{W} \right)$$

GCNGAT

GCN 발전: 모델 고도화 (2)

□ GCN류는 네트워크를 깊게 쌓을 수록 성능이 안좋아짐

- 이를 over-smoothing issue라고 함
- 직관적으로는 정보를 섞으면 섞을 수록 유의미한 특징이 사라지는 현상
 - 이를 해결하기 위해 이전 정보들을 최대한 잘 남기면서 섞는 방법들이 제안됨



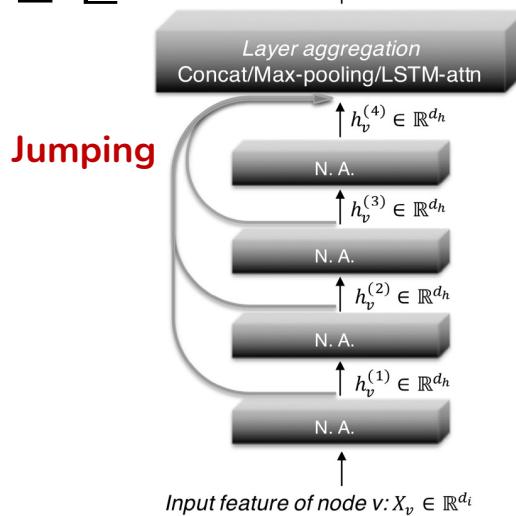
GCN 발전: 모델 고도화 (3)

□ JK [ICML'18]: 중간 결과값을 모두 섞는 방법

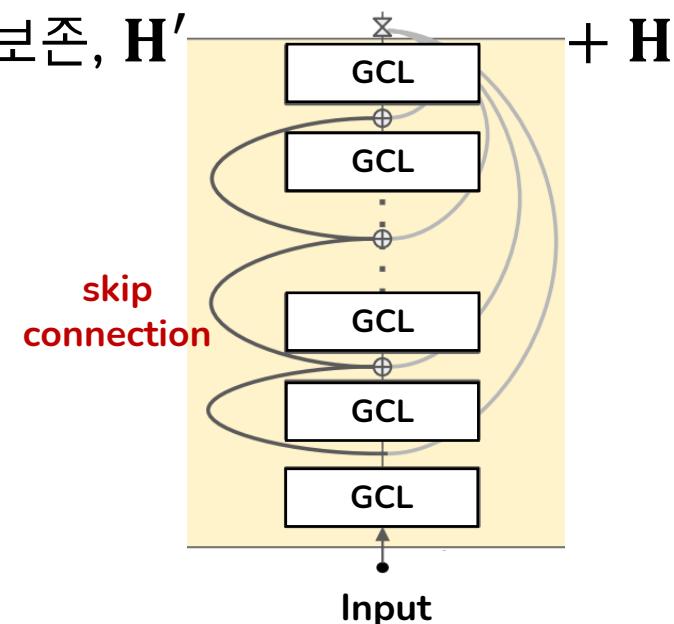
- 각 레이어의 결과값 (Knowledge)을 최종 단계로 Jumping시켜 섞음

□ ResGCN [ICCV'19]: residual (skip connection)을 활용하는 방법

- 이전 결과값을 기반으로 기전 정보를 보존, \mathbf{H}'



Jumping Knowledge (JK)



ResGCN

GCN 발전: 그래프 데이터 다양화

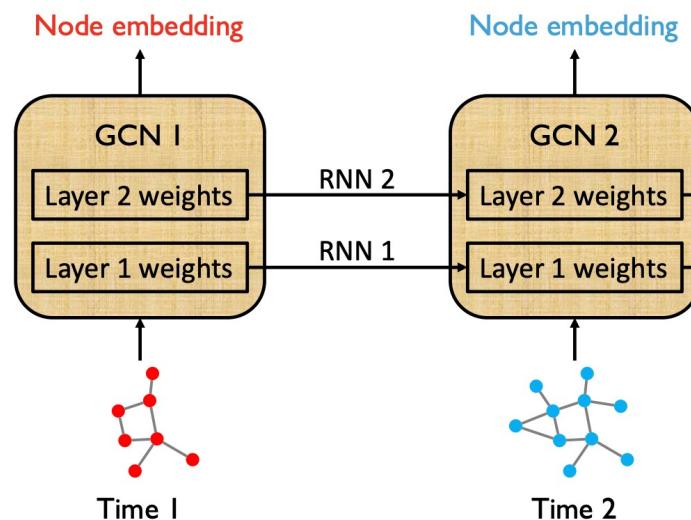
□ 다양한 그래프 종류로 그래프 특성에 맞춰 발전

■ 동적 그래프: 시간에 따라 정점/간선이 추가/삭제

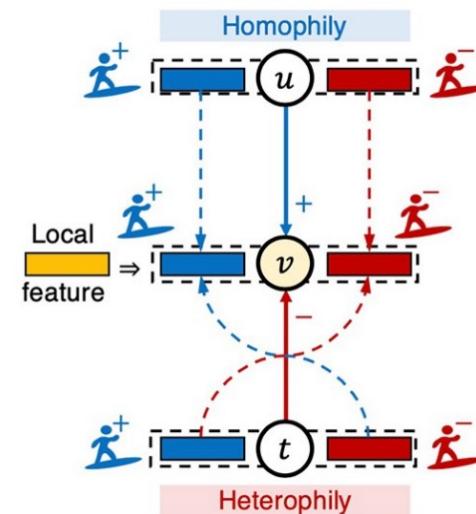
- EvolveGCN [AAAI'20]: GCN의 weight 행렬을 시간 순으로 RNN을 통과

■ 부호화된 그래프: 간선에 +/- 정보로 신뢰도를 표현

- SidNet [My Work]: 간선 부호를 처리할 수 있는 GCN 기법 개발



EvolveGCN



SidNet

GCN 발전: 다양한 분야로의 적용

□ 다양한 분야에서 발생하는 그래프 데이터/문제로 GCN 응용

3.1 Physics

3.2 Chemistry and Biology

3.3 Knowledge Graph

3.4 Recommender Systems

3.5 Computer Vision

3.6 Natural Language Processing

3.7 Generation

3.8 Combinatorial Optimization

3.9 Adversarial Attack

3.10 Graph Clustering

3.11 Graph Classification

3.12 Reinforcement Learning

3.13 Traffic Network

3.14 Few-shot and Zero-shot Learning

3.15 Program Representation

3.16 Social Network

3.17 Graph Matching

3.18 Computer Network

<https://github.com/thunlp/GNNPapers>

Outline

□ 실세계 그래프 데이터 소개

□ 그래프 뉴럴 네트워크 소개

□ 그래프 뉴럴 네트워크 발전 과정

□ **그래프 뉴럴 네트워크 응용**

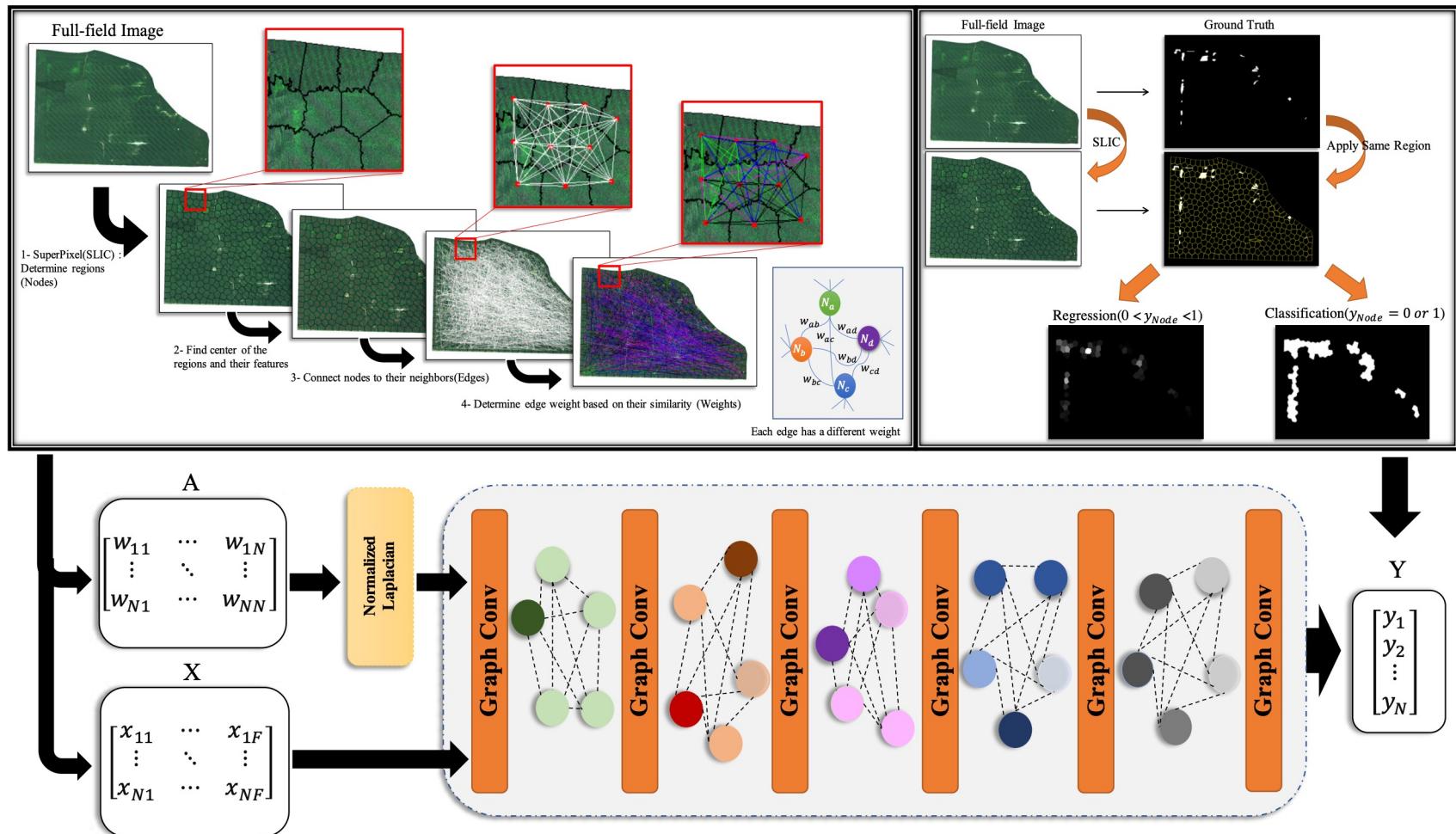
□ 데모: 그래프 정점 분류



농업 분야에서 GNN의 활용 예 (1)

■ 농업 지표면 영양 결핍 검출에 GNN 활용 [CVPR'21]

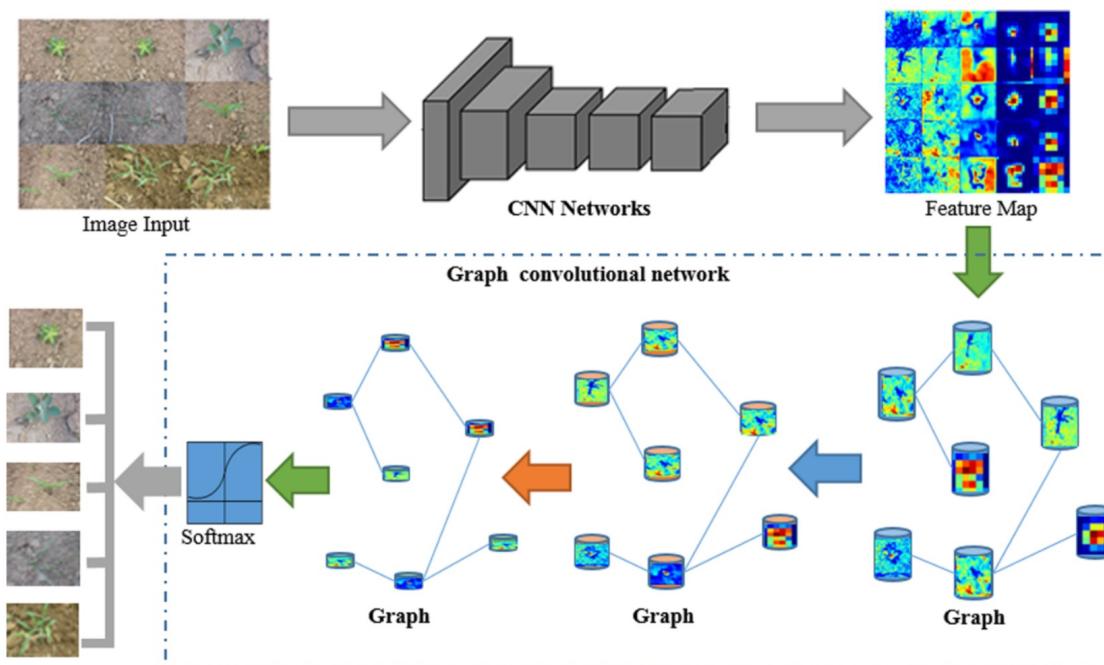
- 이미지 분할 기법으로 지대를 칸으로 나누고 칸을 점점으로 그래프 생성



농업 분야에서 GNN의 활용 예 (2)

□ 작물과 잡초 인식 [Computer and Electronics in Agriculture'20]

- CNN과 GCN의 결합으로 CNN으로 부터 얻은 영상의 feature를 기반으로 그래프를 생성 (영상 하나에 작물 또는 잡초 하나 있음)
 - 두 개의 feature의 거리를 측정하여 문턱값 이상인 경우 간선 할당

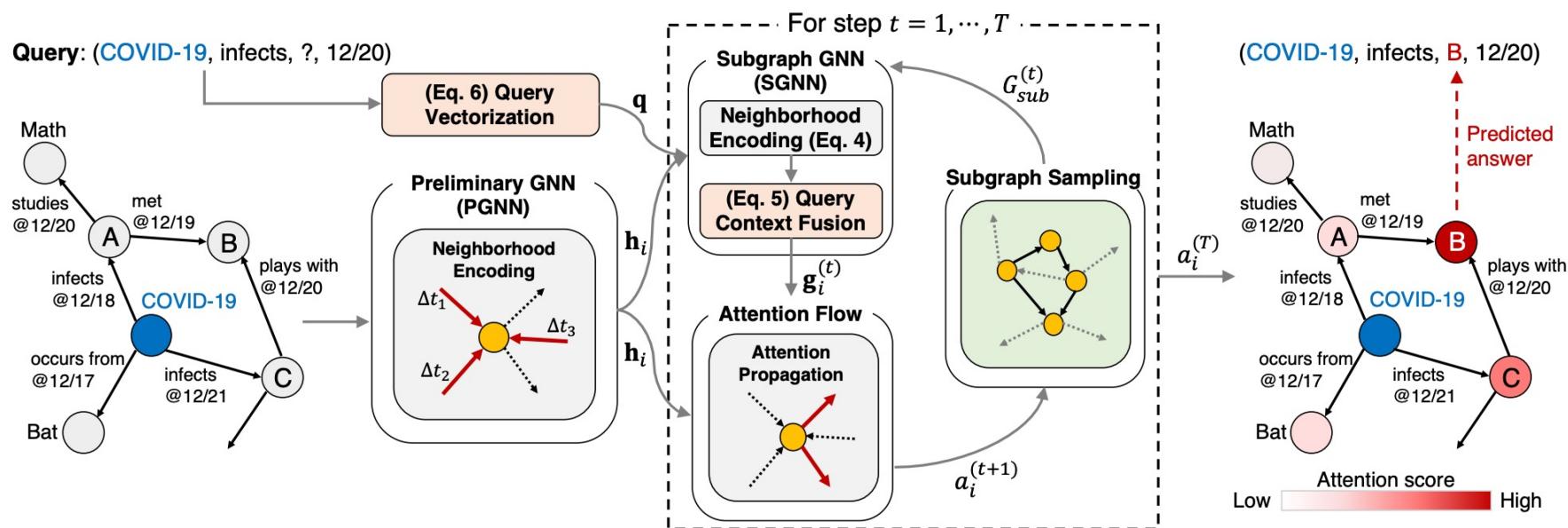


Methods	ACC
AlexNet	95.85
VGG16	97.54
ResNet-101	97.95
GCN-AlexNet (ours)	97.50
GCN-VGG16 (ours)	98.42
GCN-ResNet-101 (ours)	98.93

지식 그래프에서 GNN 활용 (1)

□ 시간 정보가 있는 지식 그래프에서 지식 추론 [KDD'21]

- Q. 20년 12월에 COVID-19에 감염된 사람은 누구인가?
 - 시간 정보를 포함하는 지식 그래프 상에서 GNN을 활용한 지식 추론

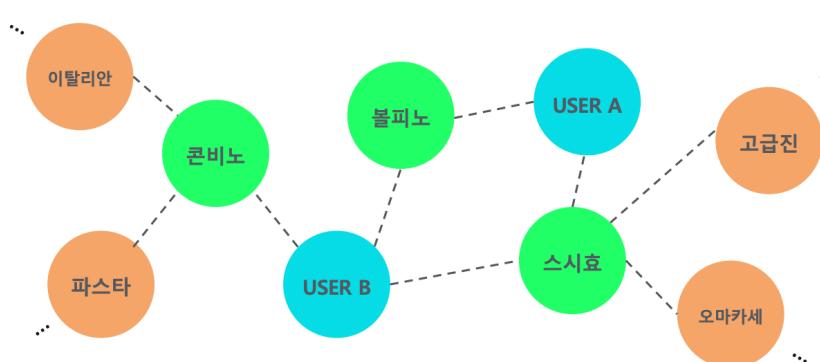


지식 그래프에서 GNN 활용 (2)

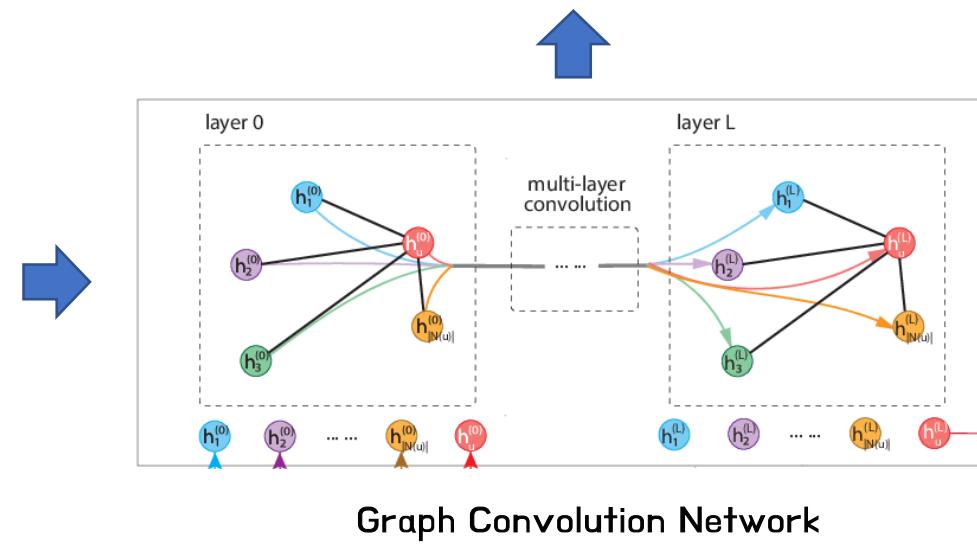
□ 지식 그래프와 GNN을 활용한 장소 추천



유저-장소-메타 지식 그래프



검색과 개인화된 정보를 기반으로
개인화 그래프 추출

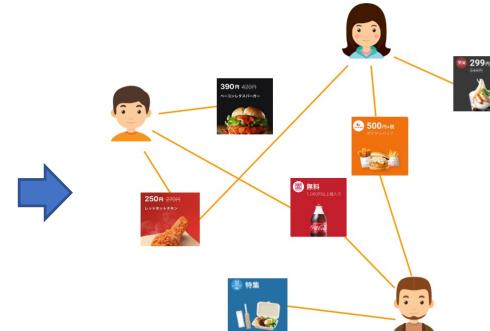
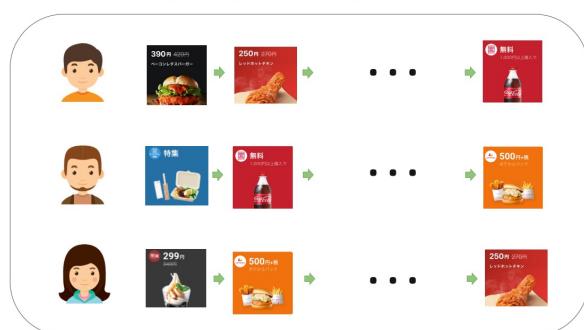


추천 시스템에서 GNN 활용

□ LINE 쿠폰 추천에서 GNN의 활용

N DEVIEW
2020

과거의 쿠폰사용로그



Top Popular 추천 대비 CTR 향상

+ 16%

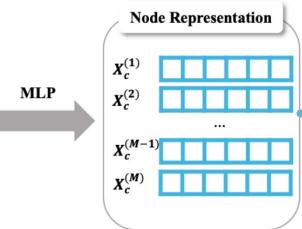
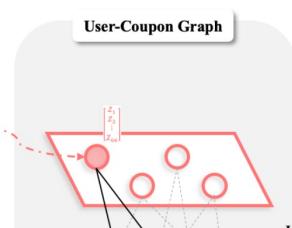
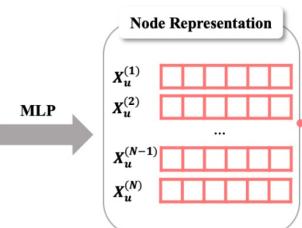
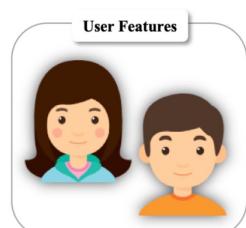


그래프모델

+ 11%



비 그래프모델



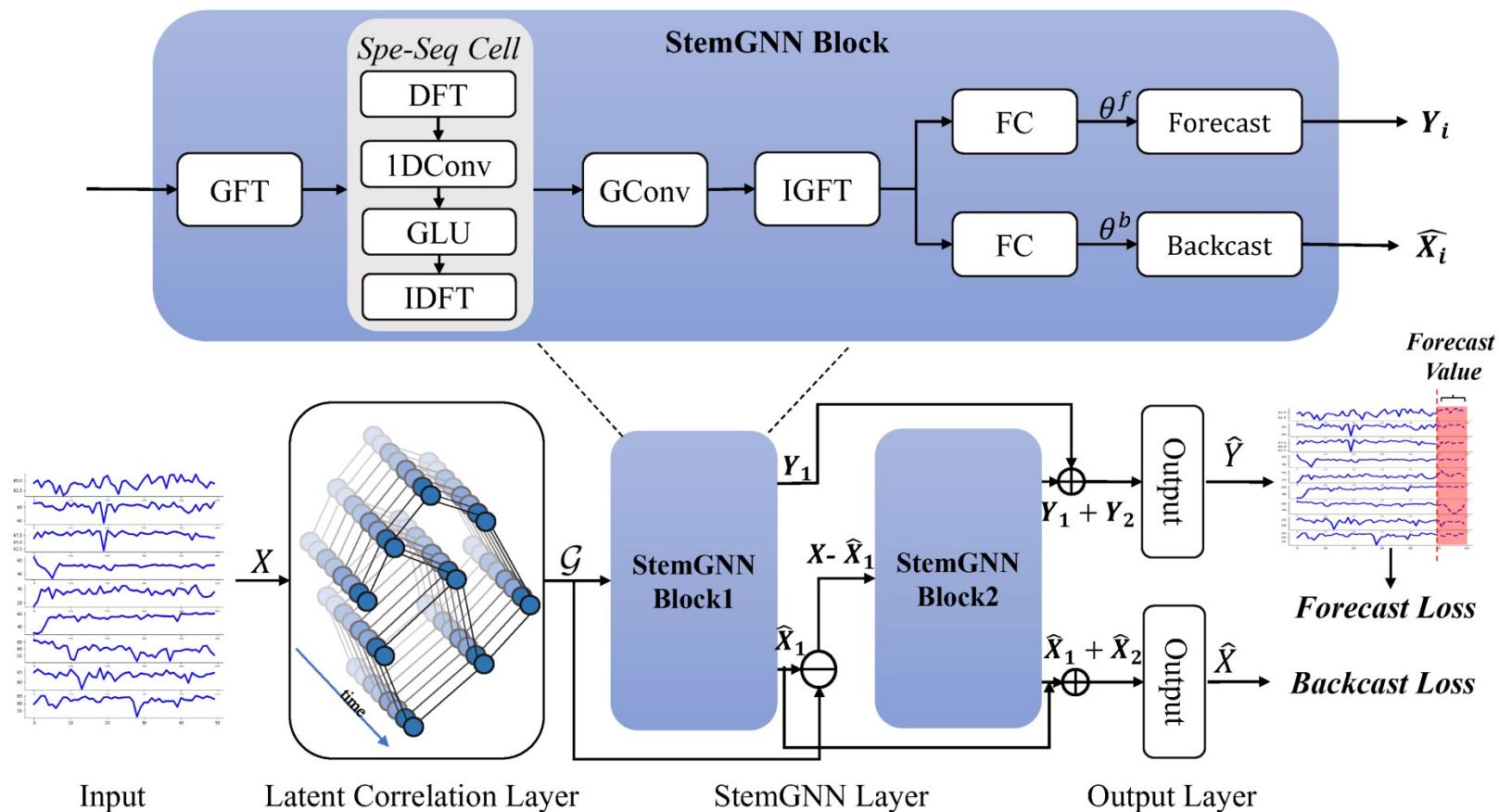
(2) Graph propagation and prediction

(1) Initial Representation via MLP

시계열 데이터의 GNN 활용

□ 시계열 데이터 향후 예측 문제 [NeurIPS'20]

- 시계열 데이터 간의 유사도 학습 & 유사도 그래프를 형성 & 미래값 예측



이미지 검색에 GNN 활용 (1)

□ 의미 기반 이미지 검색이란?

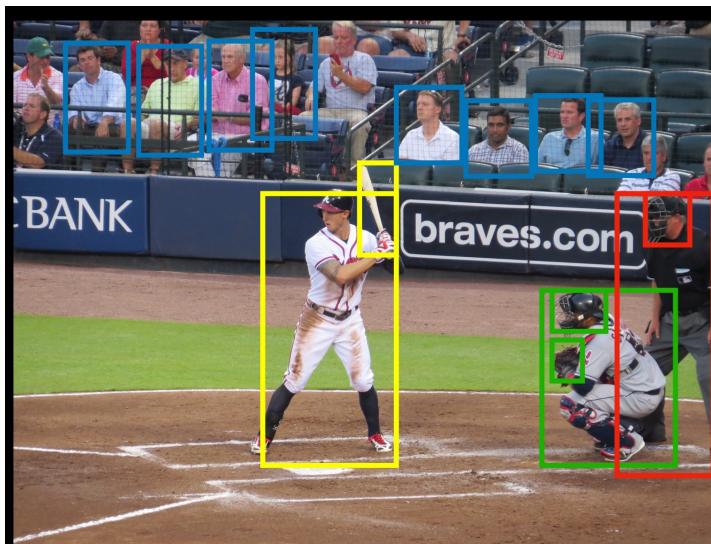
if(kakao)2021



인공지능 모델 관점에서 비슷한 이미지?
"Visually similar"



그렇다면, 좌측 사진과 비슷한 이미지는?
"Semantically similar"



야구 경기의 한 장면

타자와 포수와 심판, 그리고 관중들이 있다.

배트를 들고 서있는 사람, 헤드기어와 글러브를 끼고 앉아있는 사람,

그 뒤에 헤드기어를 쓰고 서있는 사람, 의자에 앉아있는 사람들



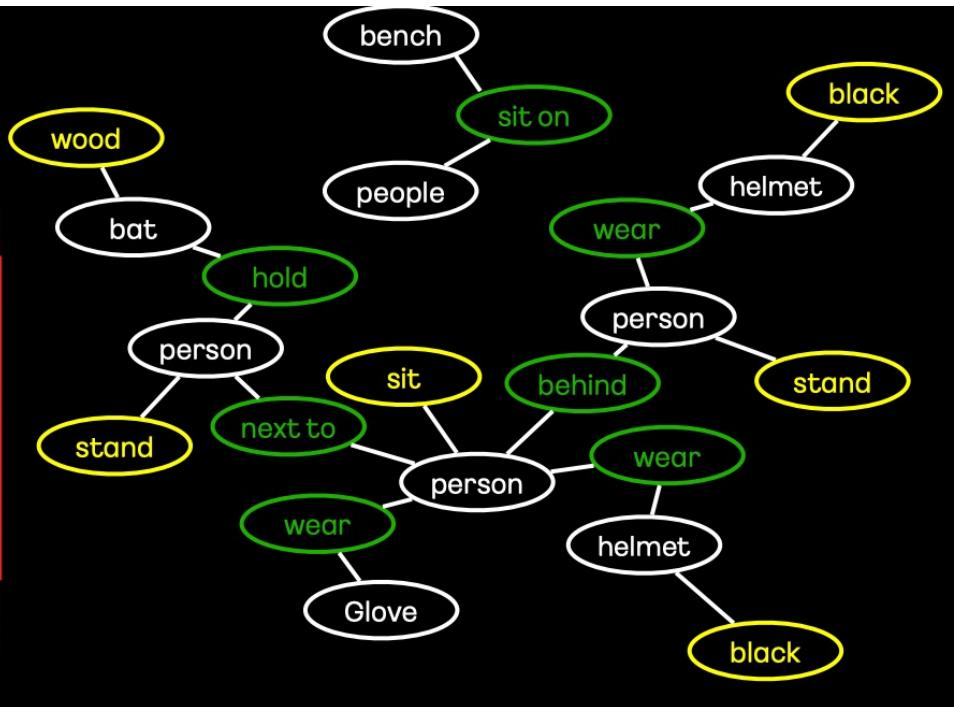
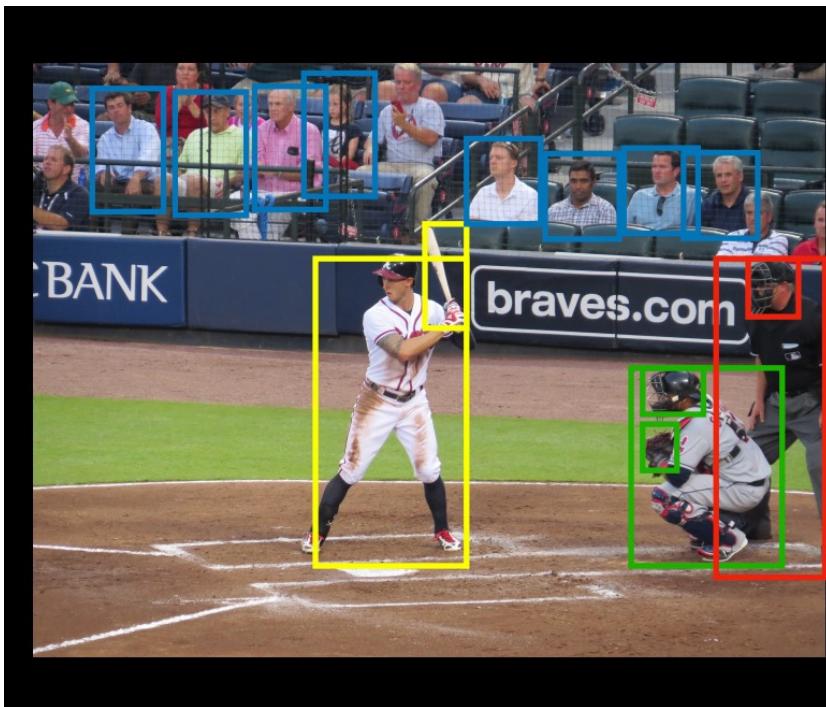
"Semantics of an image"

이미지 검색에 GNN 활용 (2)

□ 장면 그래프 구축 (Scene-graph)

if(kakao)2021

- 이미지 세그멘테이션과 NLP 기법을 활용하여 이미지내 개체간의 관계를
그래프로 표현

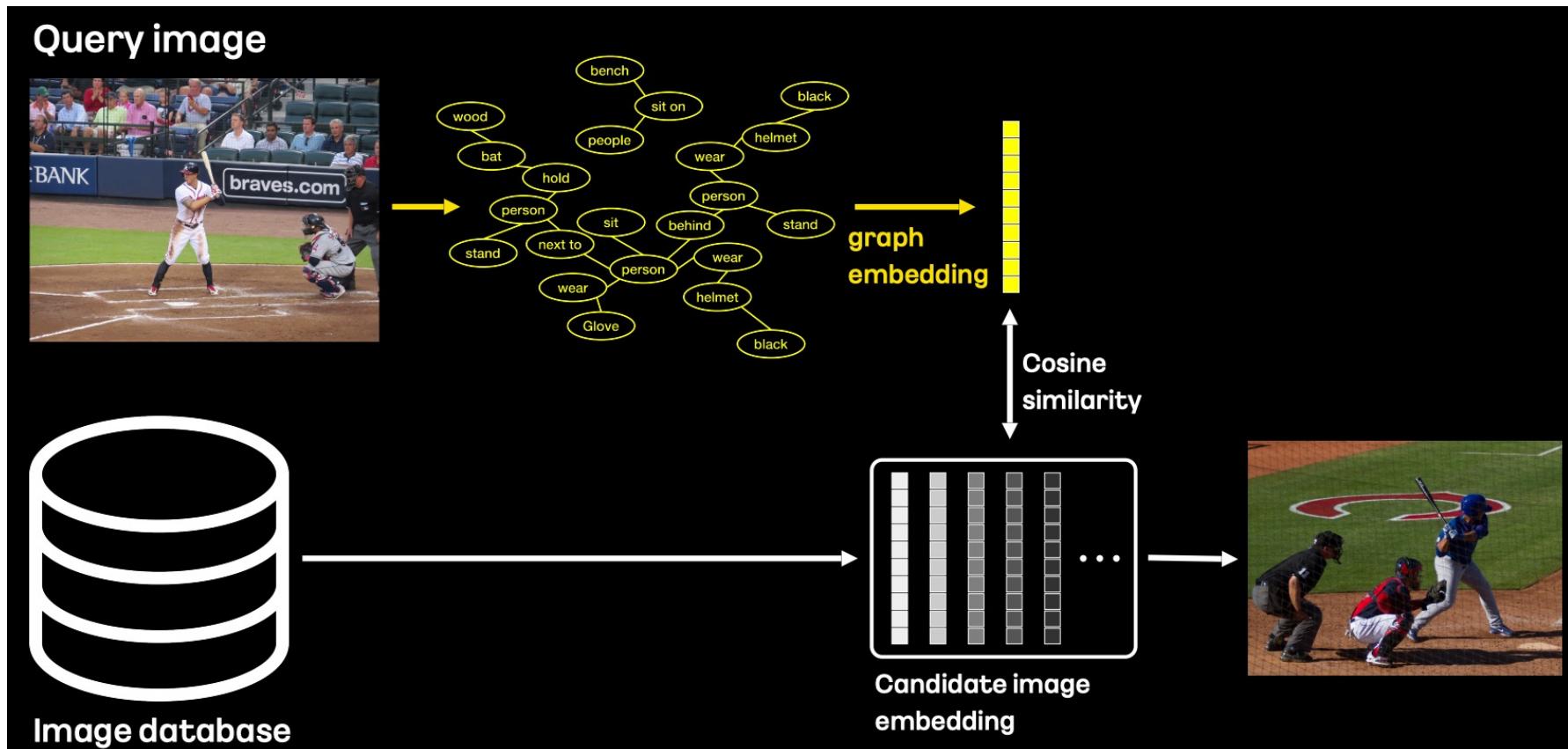


이미지 검색에 GNN 활용 (3)

□ GNN을 통한 그래프 임베딩 학습

if(kakao)2021

- 이미지 => 그래프 => 그래프 임베딩 => 임베딩 비교를 통한 검색

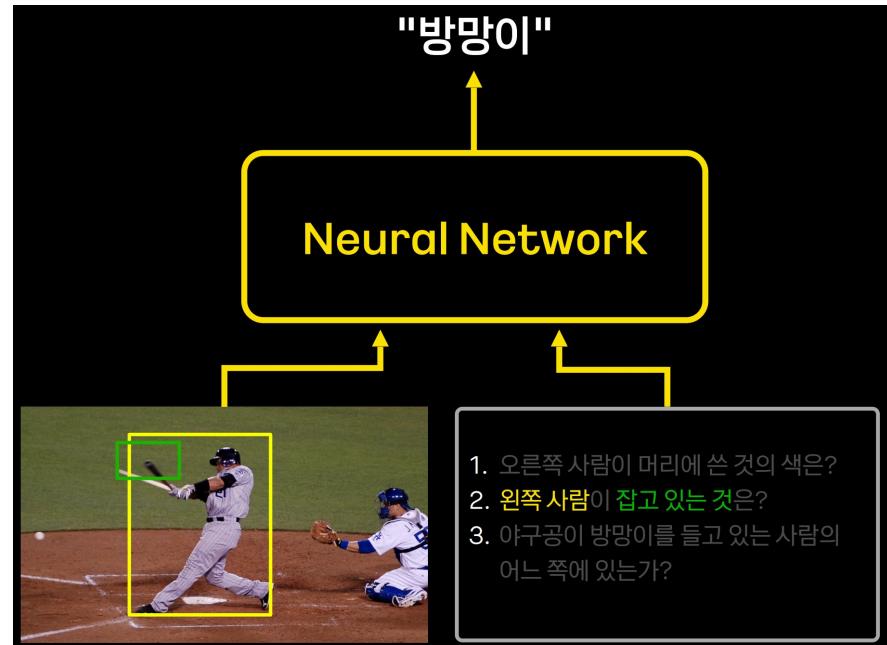
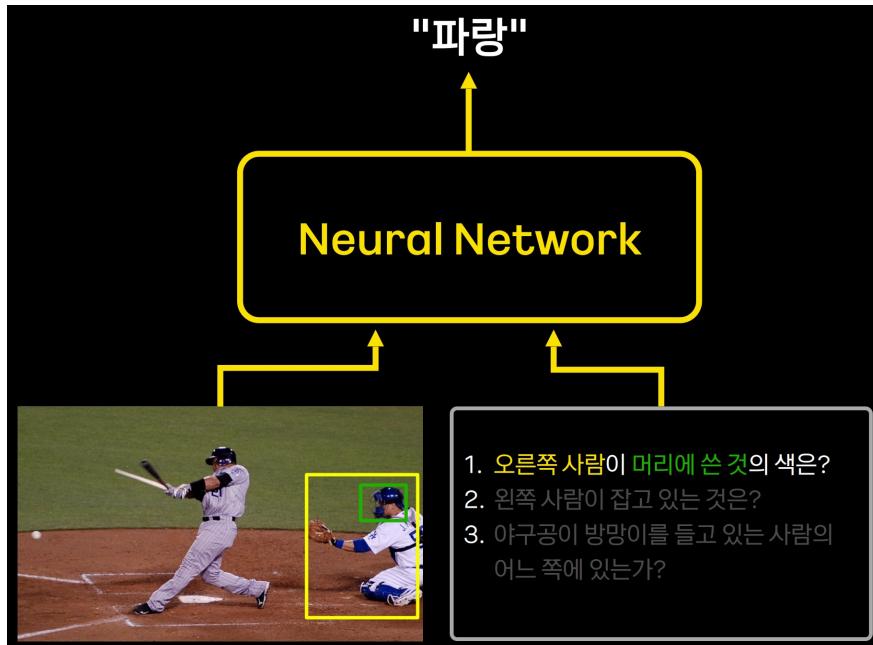


비주얼 QA에 GNN 활용 (1)

□ 비주얼 QA란?

if(kakao)2021

- 이미지와 관련된 질문이 주어지면 이미지에서 관련된 정답을 찾아 대답하는 것

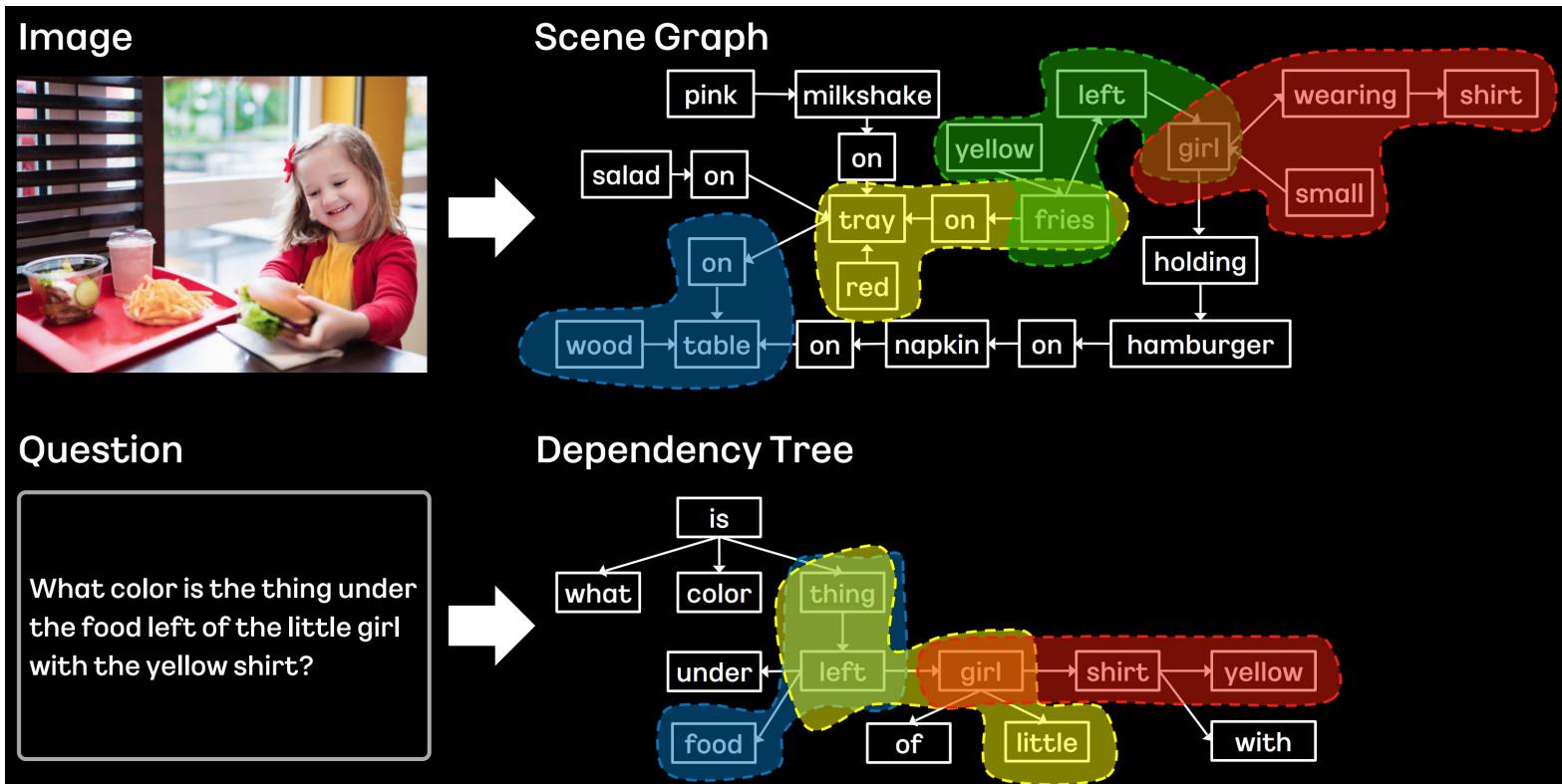


비주얼 QA에 GNN 활용 (2)

□ 장면 그래프를 활용한 비주얼 QA 방법

if(kakao)2021

- 이미지에서 장면 그래프를 추출
- 질문으로부터 지식 그래프를 추출

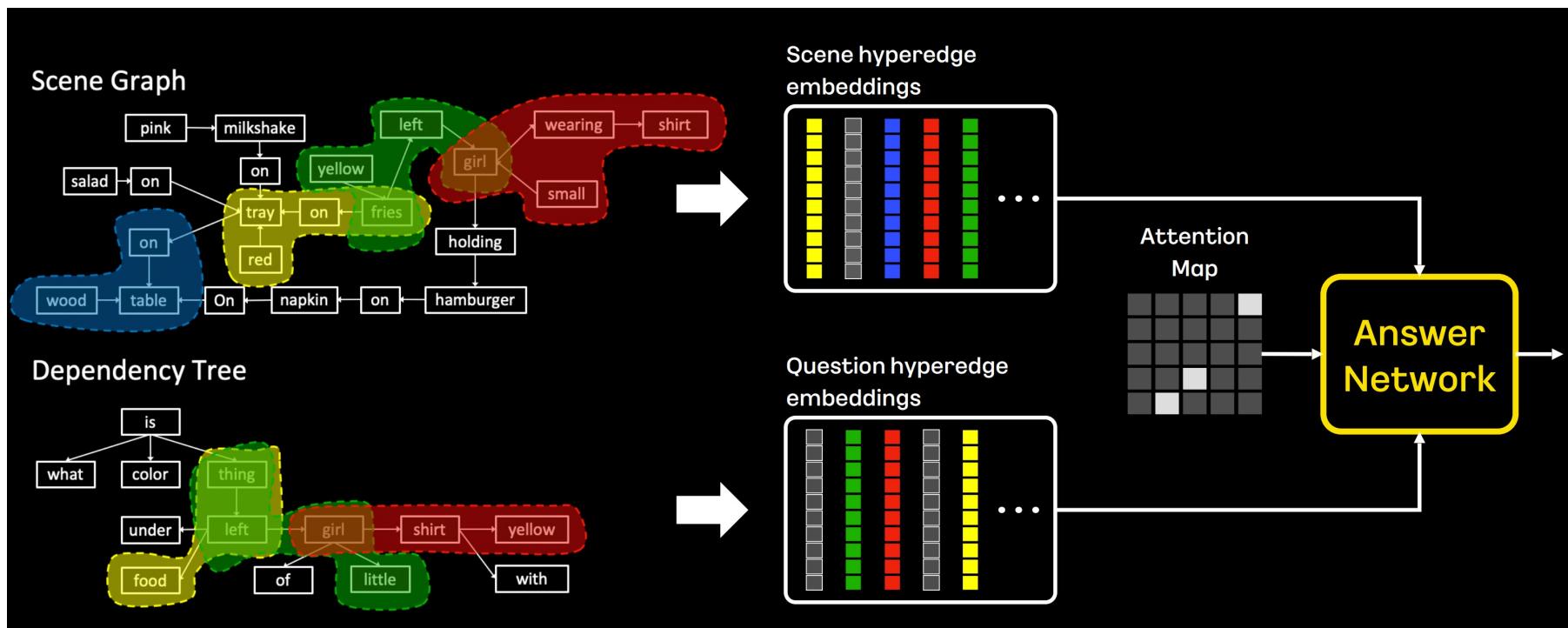


비주얼 QA에 GNN 활용 (3)

□ 장면 그래프를 활용한 비주얼 QA 방법

if(kakao)2021

- 그래프에서 구역을 나누어 부분 임베딩을 구한뒤 관련도를 측정
- 관련도와 임베딩을 기반으로 정답을 추론



GNN의 기타 응용

□ 그 외 다양한 분야에서 GNN이 응용되고 있음

- 컴퓨터 비전, NLP, 강화 학습 분야에서 활용
- 화학 및 생물 분야에서 활용 (약물 발견, 질병 예측 등)
- 교통량 예측 분야에서 활용
- 컴퓨터 네트워크 설계 분야에서 활용
- 프로그래밍 언어 설계 분야에서 활용

□ 농업 분야에서의 GNN 활용은 시작 단계

- 농업 세부 분야에서 데이터가 그래프로 표현된다면 GNN 활용이 가능!
- 단일 개체 보다 개체간 상호작용이 더 많은 정보를 담고 있어 학습에 도움이 될 수 있음

Outline

□ 실세계 그래프 데이터 소개

□ 그래프 뉴럴 네트워크 소개

□ 그래프 뉴럴 네트워크 발전 과정

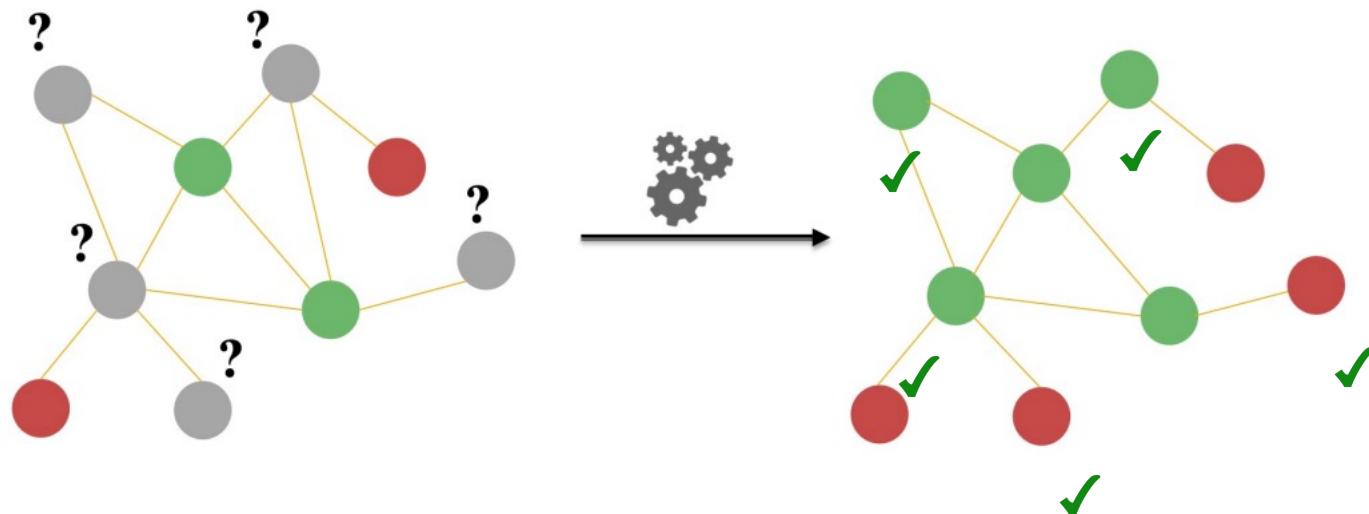
□ 그래프 뉴럴 네트워크 응용

□ 데모: 그래프 정점 분류



그래프 정점 분류

- 정점에 레이블이 일부 주어졌을 때, 레이블이 없는 정점의 레이블을 예측하기
 - Citation network에서 논문의 카테고리를 예측하는데 활용



GNN의 구현 및 데모

□ PyG를 활용하여 구현

- PyG (PyTorch Geometric)은 PyTorch 기반의 GNN 라이브러리



PyG

- 다양한 GNN 모델, 그래프 데이터셋, 그래프 GPU 관련 연산 제공
- 그래프 합성곱인 GCNConv API를 이용하여 그래프 정점 분류 수행

□ 데모를 위한 colab

- 주소: <https://bit.ly/3w9Dw4X>

Conclusion

□ Q1. 우리 주변에 어떠한 그래프 데이터가 있을까?

- Graphs are everywhere!
- 소셜 네트워크, 유저-아이템 네트워크, 문서 네트워크 등등

□ Q2. 그래프 데이터를 어떻게 학습할 수 있을까?

- 그래프 합성곱을 기반으로 그래프 정점 임베딩 및 ML task 종단 학습
- GCN: 그래프 합성곱을 기반으로 하는 기본적인 GNN
- 모델 고도화, 그래프 데이터 다양화, 적용 분야 확장으로 발전 중

□ Q3. 그래프 데이터는 어떻게 활용 될 수 있을까?

- 지식 추론, 추천, 시계열 예측, 이미지 검색 및 QA 등등
- 농업 분야 활용은 시작 단계로 GNN 적용 가능 task에 대한 논의 필요

Thank You and Q&A

□ Contact: jinhongjung@jbnu.ac.kr

□ Web: jinhongjung.github.io

□ Must-read papers on GNN

- <https://github.com/thunlp/GNNPapers>

□ PyG: Graph neural network library for PyTorch

- https://github.com/pyg-team/pytorch_geometric

□ Hands-on tutorials using PyG

- <https://pytorch-geometric.readthedocs.io/en/latest/notes/colabs.html>

Thank You