

# About My Research

Natural Language Processing and Commonsense Reasoning  
for the Next of QnA System

# Contents

1. Current Trend
2. My Research
3. What to do next

# What is PLM?

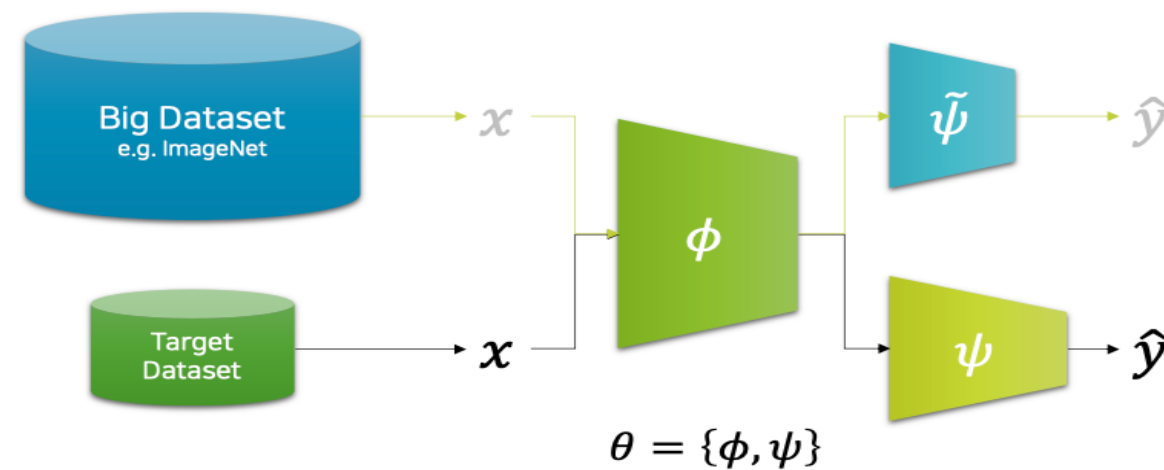
언어모델 Language Model, LM은 문장의 확률 분포를 나타낸 모델

- 문장의 출현 확률을 예측 하거나,
- 이전 단어들이 주어졌을 때 다음 단어의 확률을 예측
- 궁극적인 목표는 우리가 일상 생활에서 사용하는 언어의 문장 분포를 정확하게 모델링 하는 것
  - 특정 분야(domain)의 문장의 분포를 파악하기 위해서 해당 분야의 말뭉치를 수집하기도

# Current Trend

## Transfer Learning

- 큰 데이터셋을 통해 사전학습(pretrained)된 모델을 target task에 fine-tuning



## Self-supervised Learning

- Unlabeled 데이터셋을 지도 학습 scheme에서 학습하자
  - 일부 정보를 통해 나머지 정보를 예측

I love \_\_\_\_\_ go to school .



I love **to** go to school .

- Self-supervised Learning을 통해 좋은 weight parameter의 seed를 얻어, 전이학습(transfer learning)을 통해 한정된 데이터셋에서도 훨씬 더 좋은 성능을 얻자.

# Current Trend

## Era of Transformers

- 2017년, 구글에서 제안

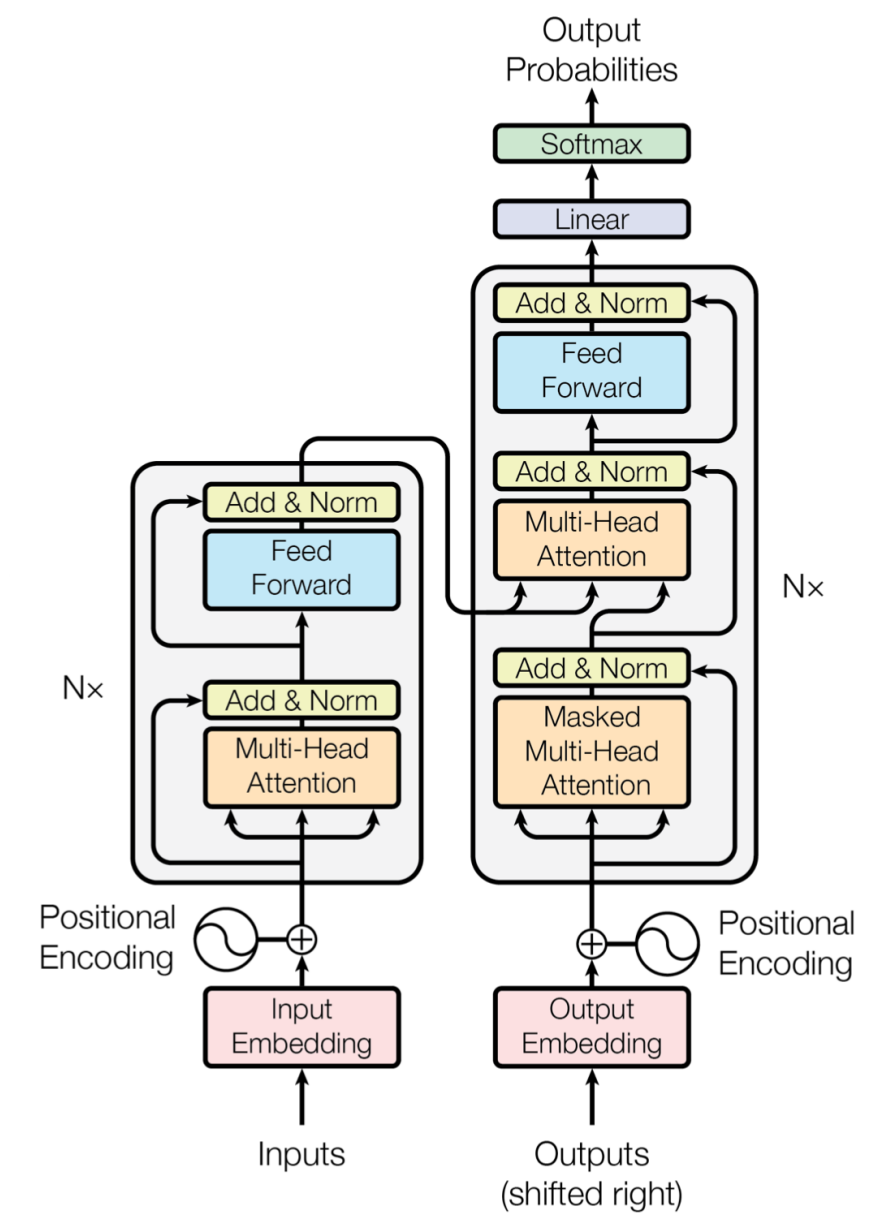
<http://papers.neurips.cc> > paper > 7181-attention-i... PDF

### Attention is All you Need - NIPS papers

A Vaswani 저술 · 52845회 인용 — We propose a new simple network architecture, the Transformer, based solely on **attention** mechanisms, dispensing with recurrence and convolutions entirely.

페이지 11개

- Attention만을 활용하여 아키텍처를 구성
  - 기존 RNN기반의 방식(e.g. LSTM)보다 훨씬 뛰어난 성능

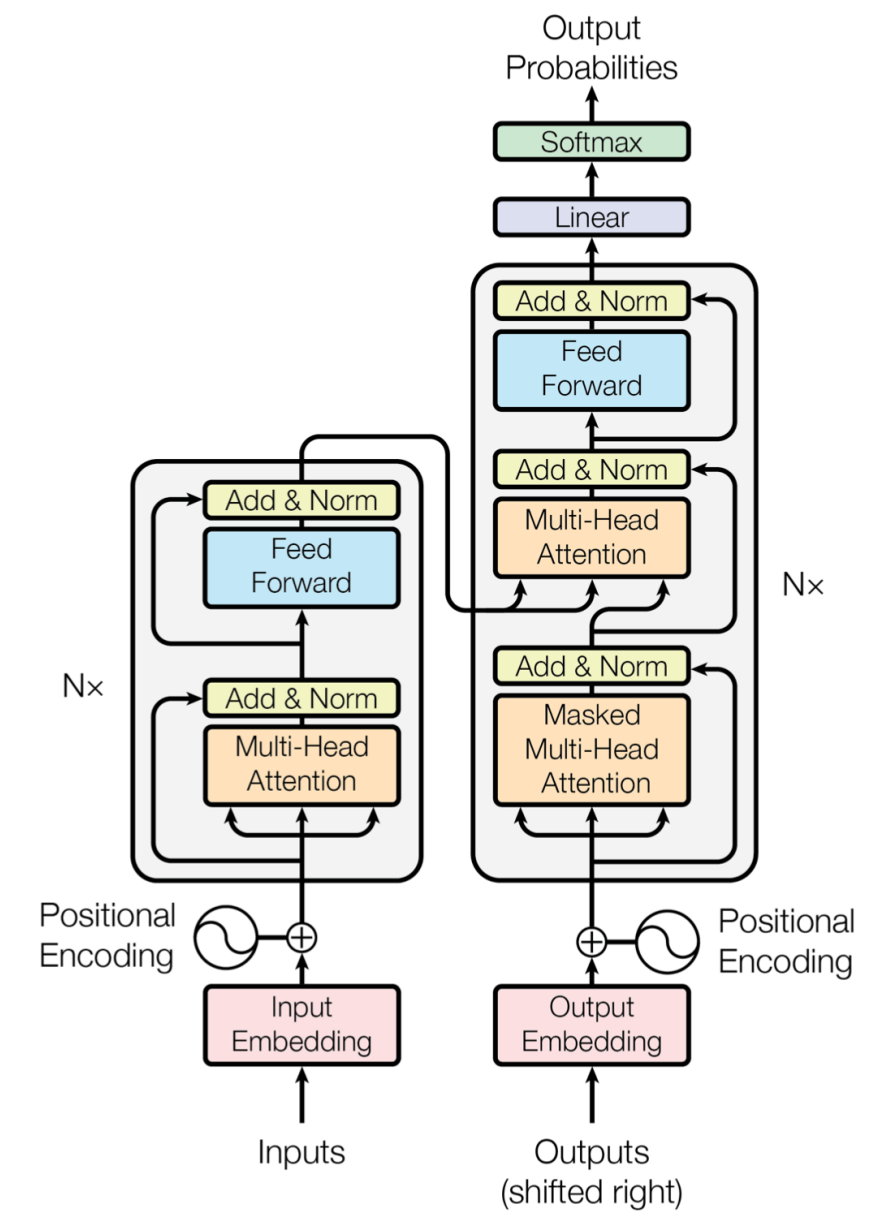


<https://arxiv.org/pdf/1706.03762.pdf>

# Current Trend

## In NLP

- General한 representation을 학습하여, 전이 학습에 활용할 수 있을까?
  - 어떻게 general representation을 학습할 수 있을까?
- Transformer를 back-bone으로 삼아 general representation을 학습하고, 이를 바탕으로 target task에 fine-tuning하자.
- PLM을 통한 성능 향상
  - 1) Feature-based Approach
    - 더 좋은 입력 representation을 갖게 하자.
  - 2) Fine-tuning Approach (+ meta learning...)
    - 더 좋은 weight parameter seed를 갖게 하자.

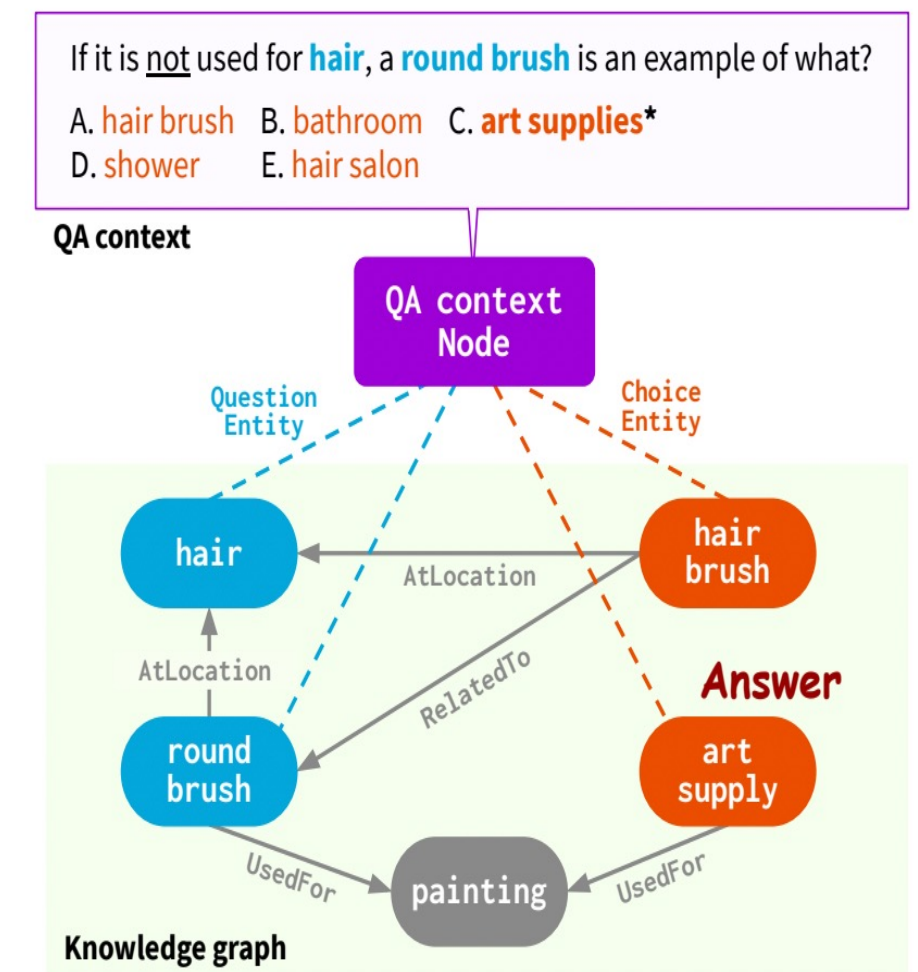


<https://arxiv.org/pdf/1706.03762.pdf>

# Current Trend

## In GNN

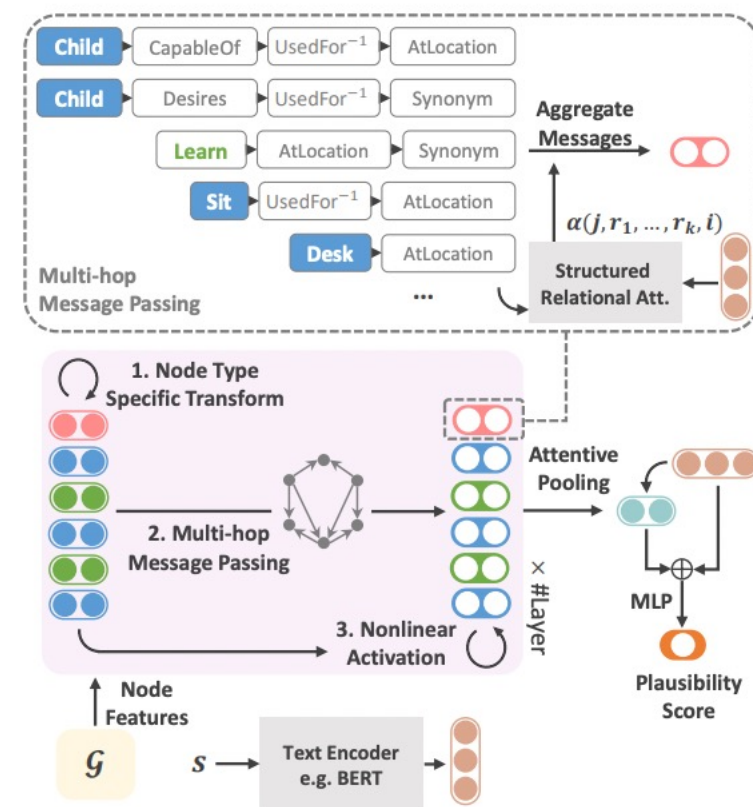
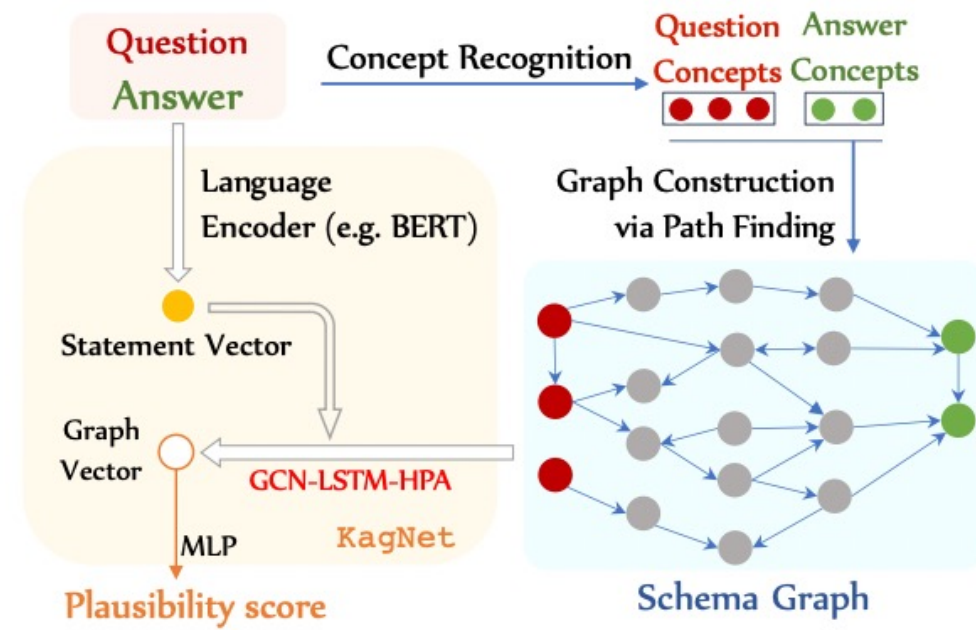
- Recently, pre trained LMs have demonstrated remarkable success in many question answering tasks
- However, while LMS have a broad coverage of knowledge, they do not empirically perform well on **structured reasoning**
- On the other hand, KGs are more suited for **structured reasoning**



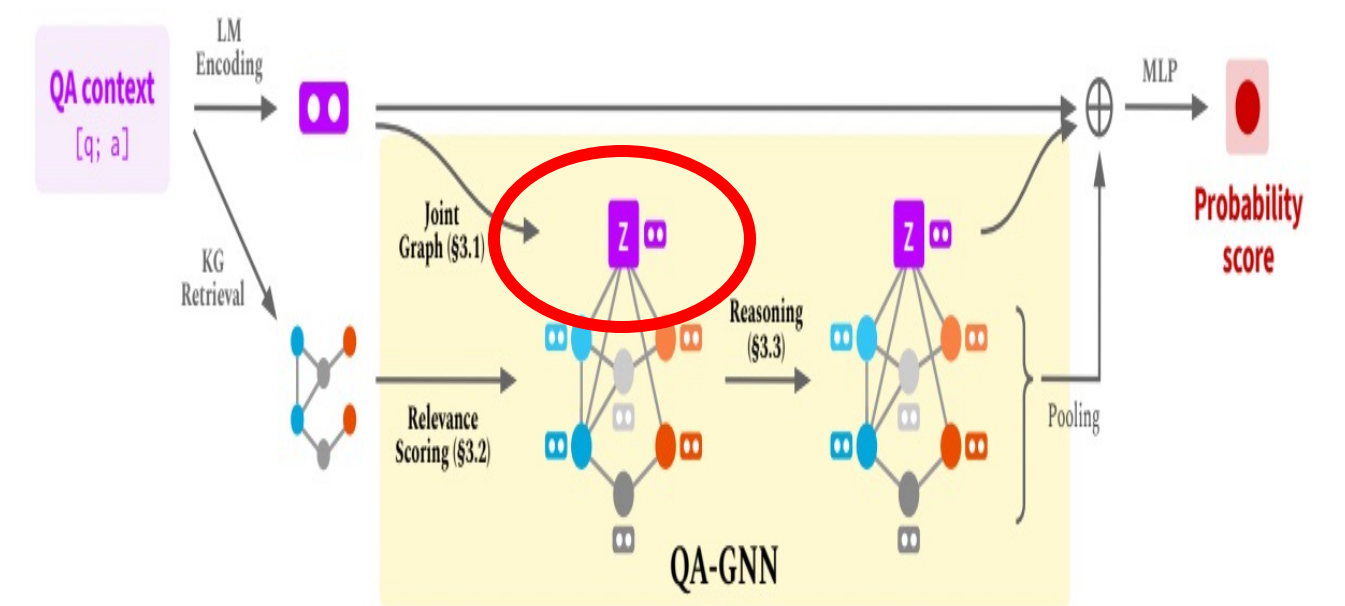
<https://arxiv.org/pdf/2104.06378.pdf>

# Current Trend

## NLP with KG



QA context의 중요성!!!

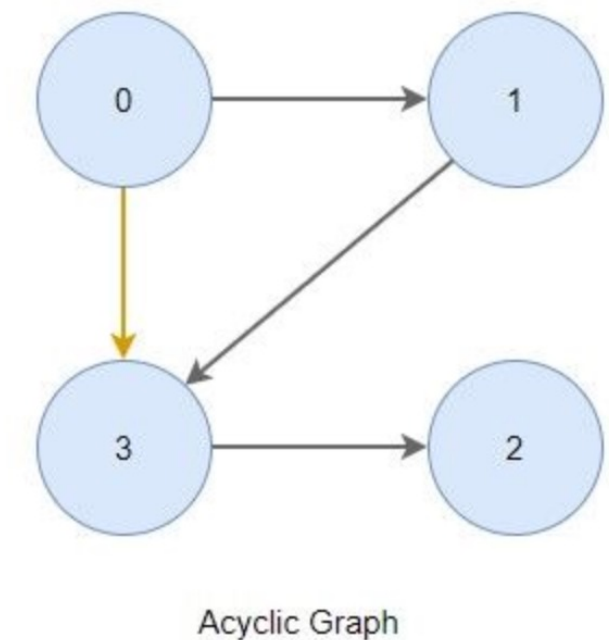
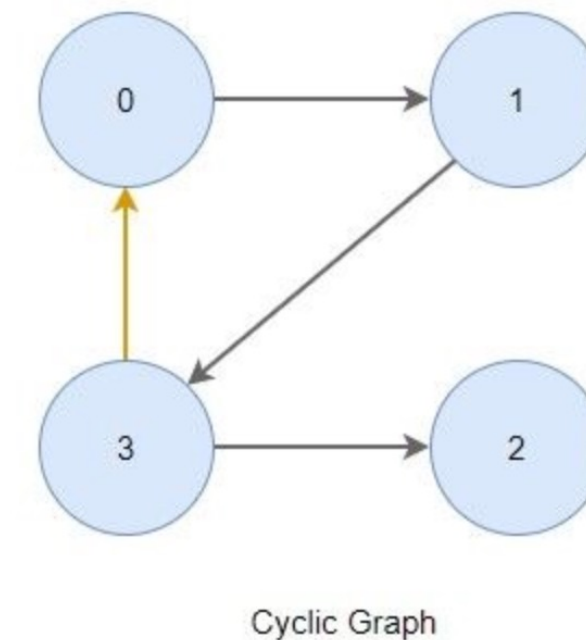




# What is Cyclic Graph

## Definition

- In graph theory, a cycle graph or circular graph is a graph that consists of a single cycle, or in other words, some number of vertices connected in a closed chain.
- 쉽게 말해, 경로 중 시작 정점과 끝 정점이 같은 경우를 순환이라고 하며, 순환(Cycle)을 가지고 있으면 순환 그래프(Cycle Graph)임

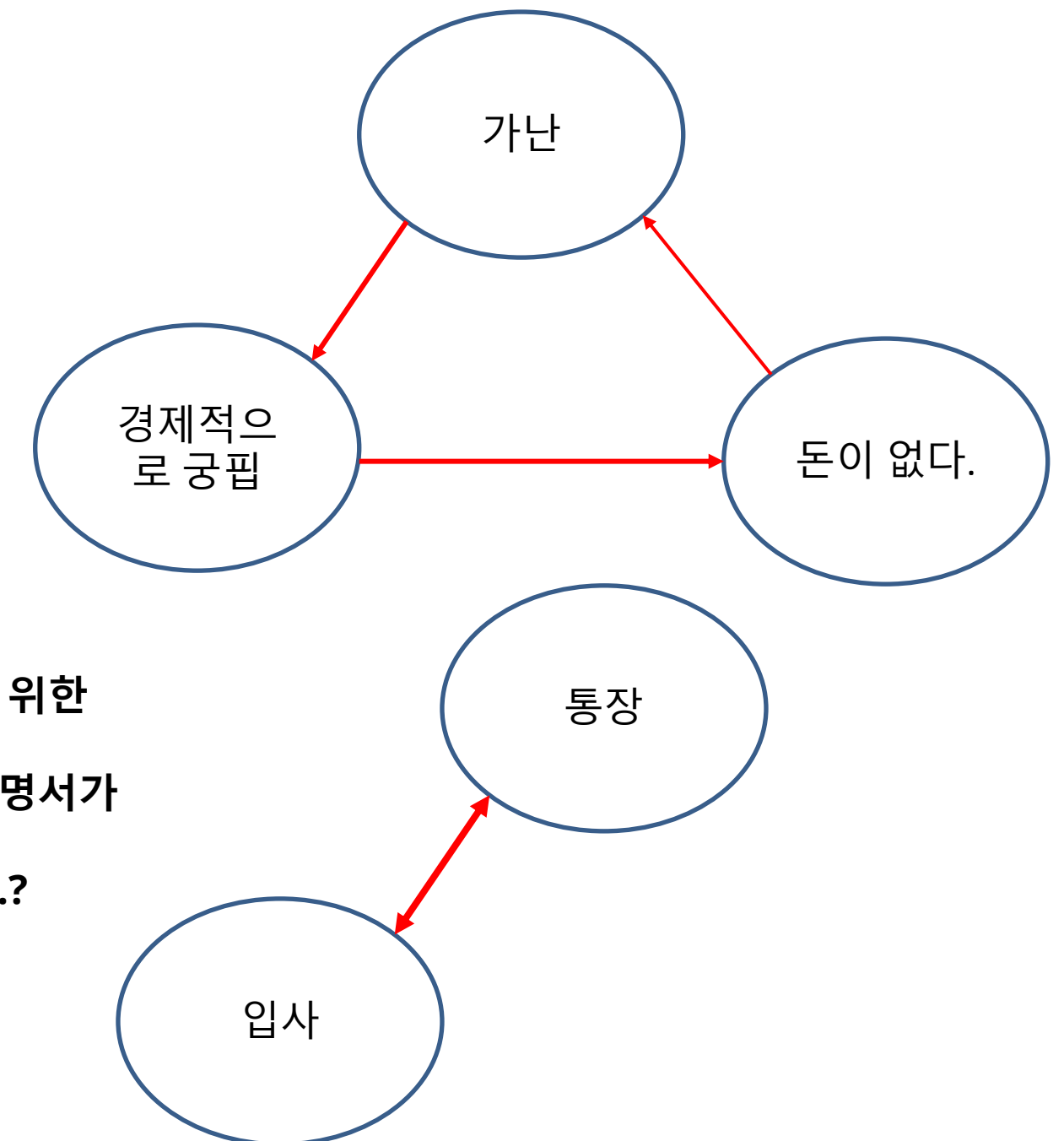


# Do circular graph cause problems in the reasoning process?

## In Human

- 순환 논법(Cyclic Reasoning)이란 어떤 주장을 함에 있어 그 주장의 근거로 그 주장을 사용하는 **오류**. 선결문제의 오류 또는 순환 논리의 오류라고도 하며, 논리적 오류 중에선 비형식적 오류에 해당함.
- 말싸움에서 의미없이 우기는 쪽이나 **모순**된 주장 등에서 자주 보이는 논법이기도 함.
- Ex)
  - 교수: 자네는 가난하네
  - 내담자: 선생님, 그럼 저는 왜 가난한 걸까요?
  - 교수: 그것은 자네가 경제적으로 궁핍하기 때문일세.
  - 내담자: 그럼 선생님, 저는 왜 경제적으로 궁핍한 건가요?
  - 교수: 그건 자네가 돈이 없기 때문일세.
  - 내담자: 왜 제가 돈이 없을까요?
  - 교수: 그야 자네가 가난하기 때문이지!

1. 회사에 입사하기 위해서는 입금을 받기 위한 통장이 필요하다.
2. 통장을 만들기 위해서는 입사했다는 증명서가 필요하다.
3. 입사를 하기 위해서는 통장이 필요하다.?

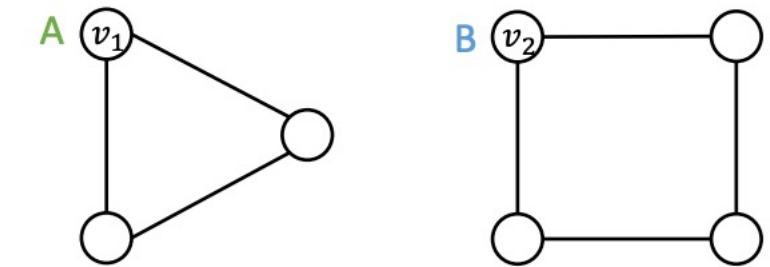


# Do circular graph cause problems in the reasoning process?

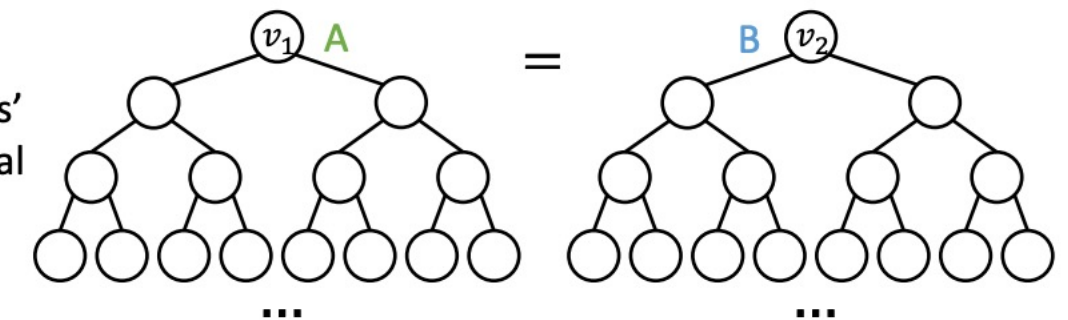
## In AI

- 1. Knowledge Graph에서 추출되는 Subgraph에 Cycle graph가 나오면 인간의 추론과정처럼 문제가 되지 않을까?
- 2. 사이클로 인한 문제가 있음 -> About WL(isomorphic) test, GNN
- 그래프의 동형유무를 판별하는 WL테스트는 그래프가 사이클을 이루고 있을 경우 제대로 판별하지 못함
- 다른 그래프 모양이지만 Message Passing GNN은 두 그래프를 구분할 수 없다. -> 그렇다면 Reasoning에도 문제가 있지 않을까?(because of Using GNN)

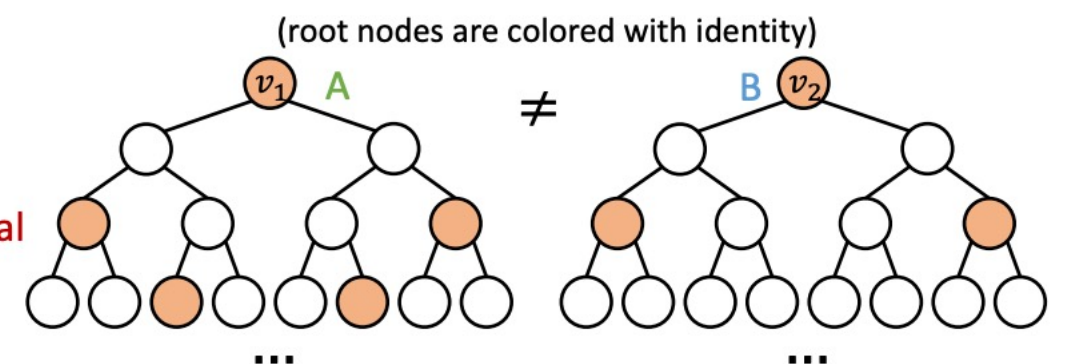
Example input graphs



Existing GNNs' computational graphs



ID-GNNs' computational graphs



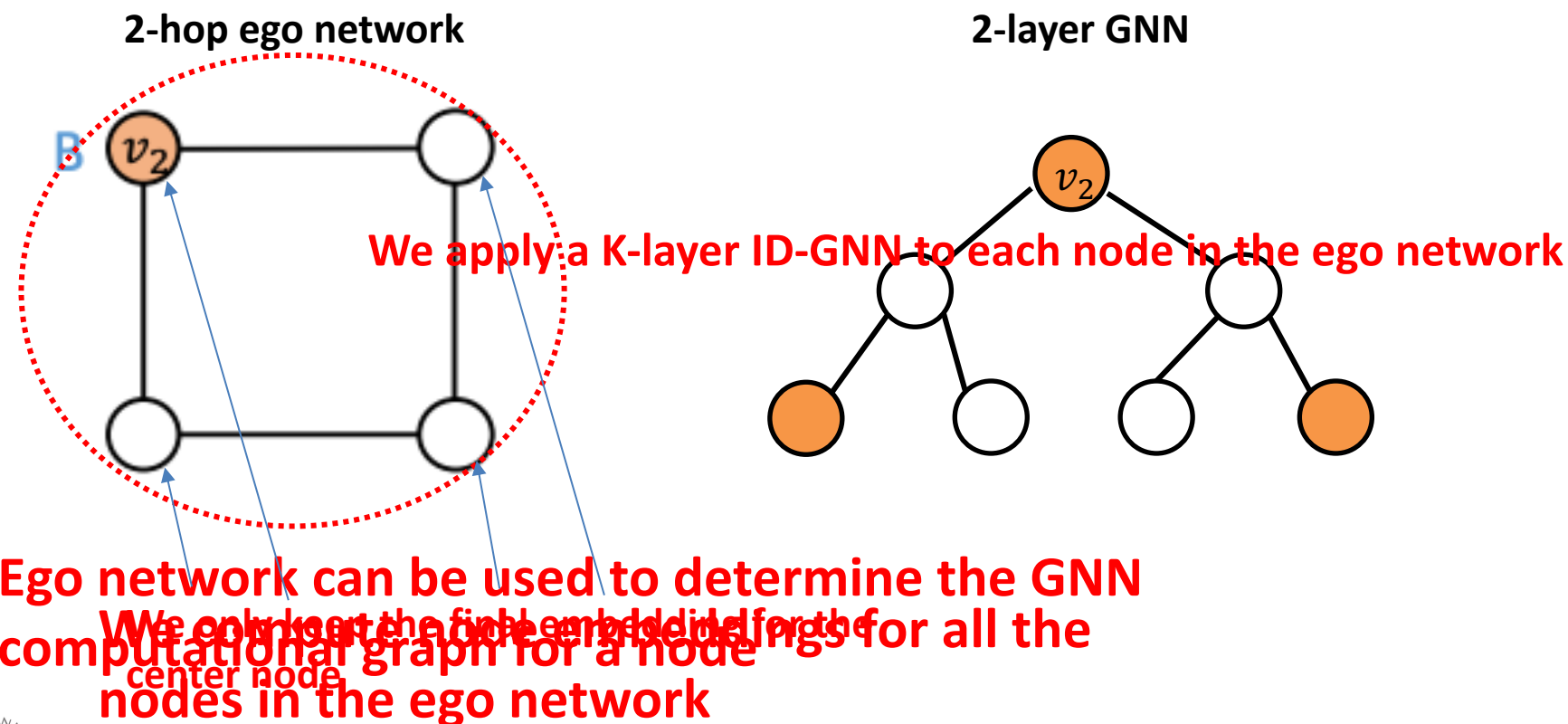
<https://arxiv.org/pdf/2101.10320.pdf>

Rooted node에 대한 identity를 지정하자

# Do circular graph cause problems in the reasoning process?

## ID-GNN Algorithm

We extract the ego network for each node, and initialize node embeddings by raw node features



Other node embeddings are only used to facilitate the computation

### Algorithm 1 ID-GNN embedding computation algorithm

**Input:** Graph  $\mathcal{G}(\mathcal{V}; \mathcal{E})$ , input node features  $\{x_v, \forall v \in \mathcal{V}\}$ ; Number of layers  $K$ ; trainable functions  $\text{MSG}_1^{(k)}(\cdot)$  for nodes with identity coloring,  $\text{MSG}_0^{(k)}(\cdot)$  for the rest of nodes;  $\text{EGO}(v, k)$  extracts the  $K$ -hop ego network centered at node  $v$ , indicator function  $\mathbb{1}[s = v] = 1$  if  $s = v$  else 0

**Output:** Node embeddings  $\mathbf{h}_v$  for all  $v \in \mathcal{V}$

```

1: for  $v \in \mathcal{V}$  do
2:    $\mathcal{G}_v^{(K)} \leftarrow \text{EGO}(v, K)$ ,  $\mathbf{h}_u^{(0)} \leftarrow \mathbf{x}_u, \forall u \in \mathcal{G}_v^{(K)}$ 
3:   for  $k = 1, \dots, K$  do
4:     for  $u \in \mathcal{G}_v^{(K)}$  do
5:        $\mathbf{h}_u^{(k)} \leftarrow \text{AGG}^{(k)}(\{ \text{MSG}_{\mathbb{1}[s=v]}^{(k)}(\mathbf{h}_s^{(k-1)}), s \in \mathcal{N}(u) \}, \mathbf{h}_u^{(k-1)})$ 
6:    $\mathbf{h}_v \leftarrow \mathbf{h}_v^{(K)}$ 
  
```

<https://arxiv.org/pdf/2101.10320.pdf>

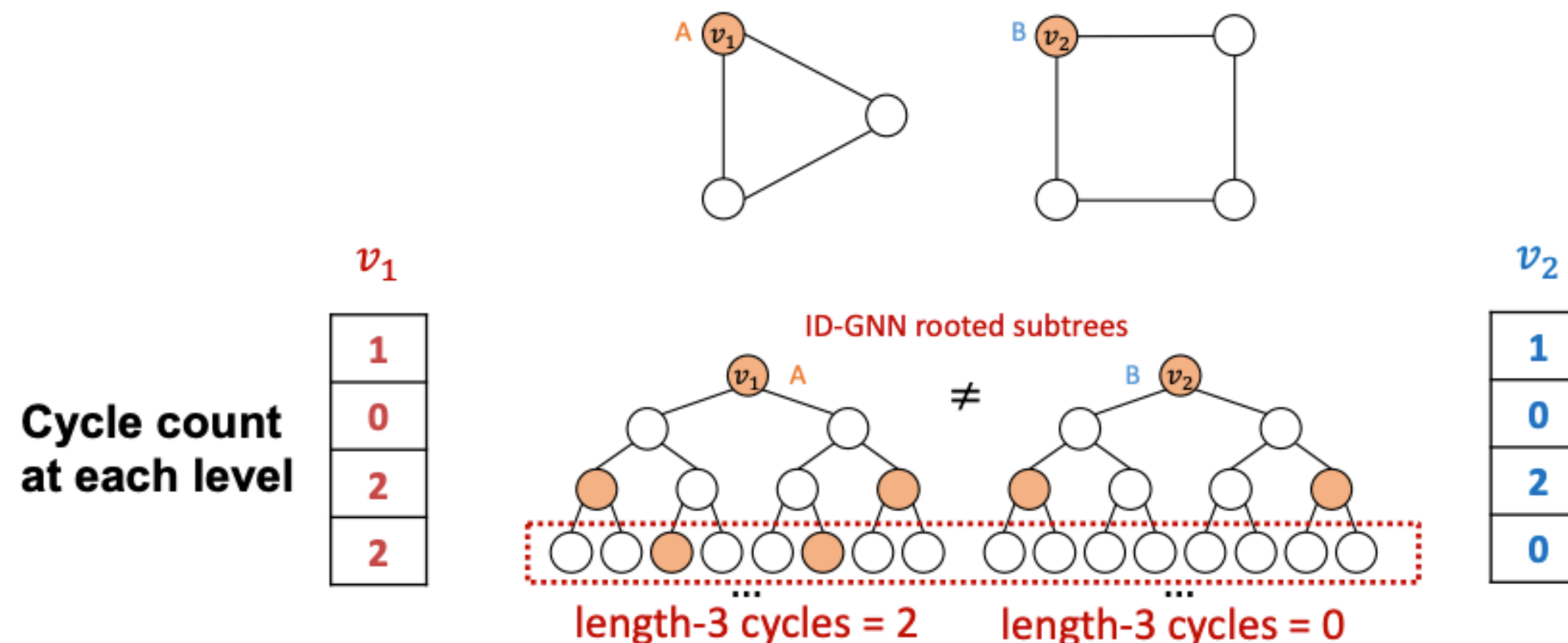
# Do circular graph cause problems in the reasoning process?

## Simplified Version : ID-GNN-Fast

- Based on the intuition, They propose ID-GNN-Fast
  - Use cycle counts in each layer as additional node features
  - Include identity information as an **augmented node feature**  
(no need to do heterogeneous message passing)

Cycle count can be computed efficiently with sparse multiplication via  $x_v^+[k] = \text{Diag}(A^k)[v]$ , where  $A$  is the adjacency matrix.

Ex)  $A_{uv}^2$  specifies #walks of length 2(neighbor of neighbor) between  $u$  and  $v$ .

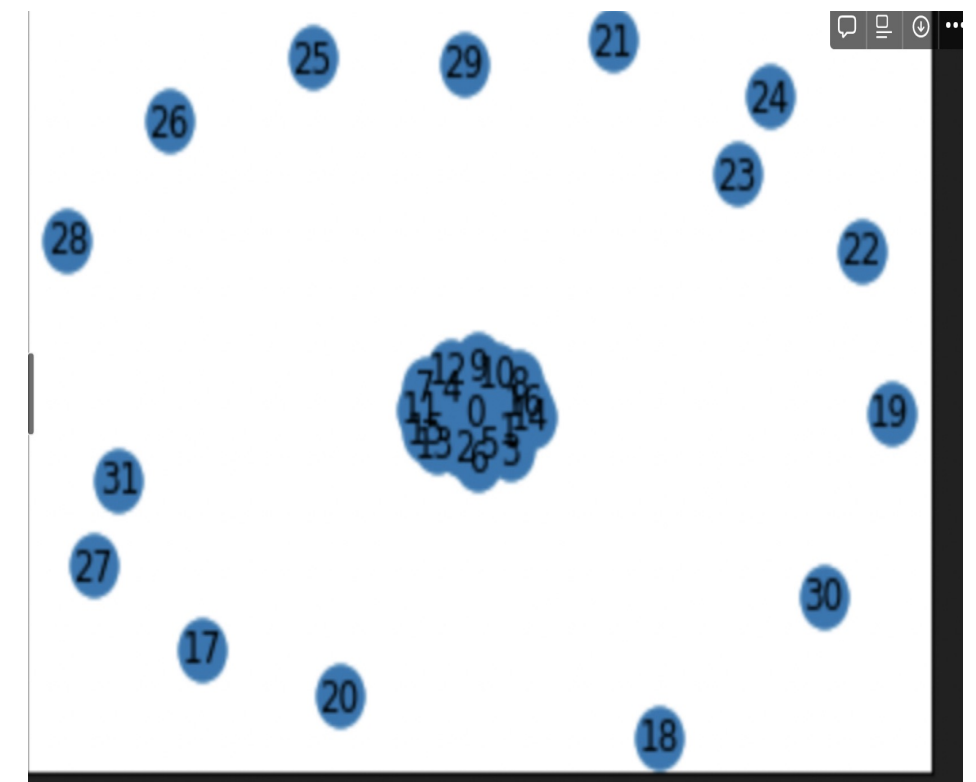




# Do circular graph cause problems in the reasoning process?

## In AI

- 이렇듯 cyclic graph에서 ai도 마찬가지로 reasoning에 문제가 있을거라 생각함.
- 정리하자면, 일반적인 GNN으로는 다른 모양의 subgraph(cycle이 포함된 그래프)일지라도 똑같은 reasoning을 보여줌 -> **node coloring** 사용 -> **cyclic reasoning**을 수행한 것이라 볼 수 있지 않을까?
- 하지만, 이러한 그래프(하나의 사이클로만 이루어진 그래프)는 subgraph 추출시 나오기 힘들 -> Because of Context node
- 물론 극단적으로 생길 수는 있음
- QA-GNN(2021)이 나온 이후로 모든 Commonsense Reasoning시 context node를 사용함
- Context node는 질문 과 답변을 나타내는 문장을 노드로 임베딩 시킨 것으로 모든 question entity(node), answer entity(node)는 context 노드와 양방향 연결로 이루어짐(Subgraph추출시)



# Do circular graph cause problems in the reasoning process?

## In AI

- 노드의 타입은 4가지로 구성됨, context node, question node, answer node, other node
- Other node는 context노드와 직접적으로 연결되지 않음
- 일반적인 GNN은 문제가 있었음 → node coloring을 사용하자...
- 좀 더 효율적이고 기존 GNN의 방식을 바꾼 것이 없나? -> GSC(Graph-Soft-Counter)

---

### Algorithm 1 PyTorch-style code of GSC

---

```
# qa_context: question answer pair context
# adj: edge index with shape 2 x N_edge
# edge_type: edge type with shape 1 x N_edge
# node_type: node type with shape 1 x N_node
edge_emb = edge_encoder(adj, edge_type, node_type)
node_emb = torch.zeros(N_node)
for i_layer in range(num_gsc_layers):
    # propagate from node to edge
    edge_emb += node_emb[adj[1]]
    # aggregate from edge to node
    node_emb = scatter(inputs, adj[0])
graph_score = node_emb[0]
context_score = fc(roberta(qa_context))
qa_score = context_score + graph_score
```

---

# How to develop Cyclic reasoning?

## Cyclic Reasoning

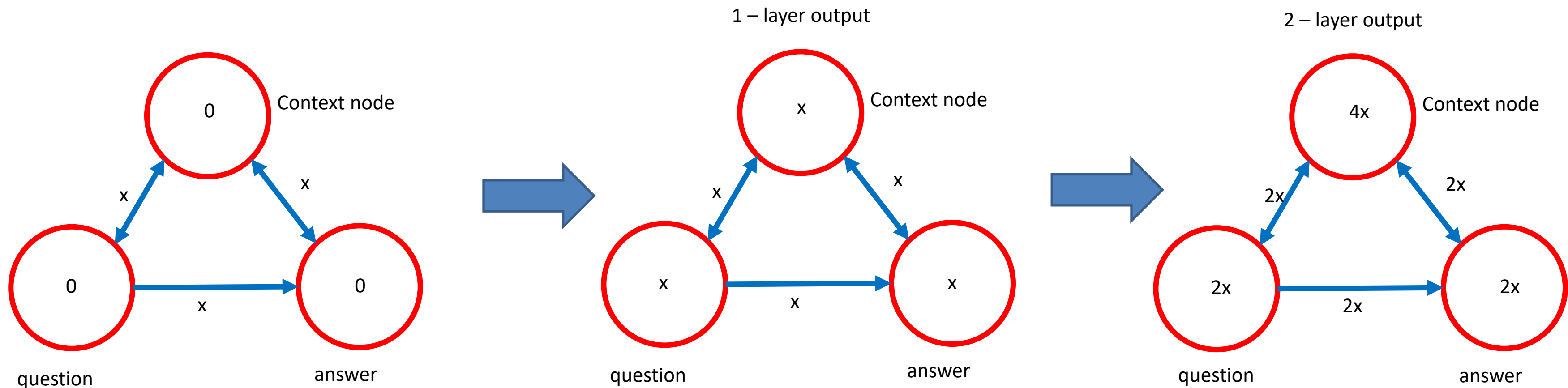
- 그래프에서 cycle은 GNN, WL test에 영향을 미침
- 따라서 그래프를 통한 reasoning에도 문제가 있음
- 기존 GNN방식에서의 문제점이 뭐지? -> **1. ID-GNN(identity node coloring)**
- 기존 GNN방식과 다르게 효율적인 연산 방법을 찾아보자 -> **2. Graph - Soft - Counter**
- GSC와 identity coloring을 같이 사용해서 cyclic reasoning을 만들자.



Context노드는 모든 question node, answer node와 연결되도록 함

## Is it effective for reasoning?

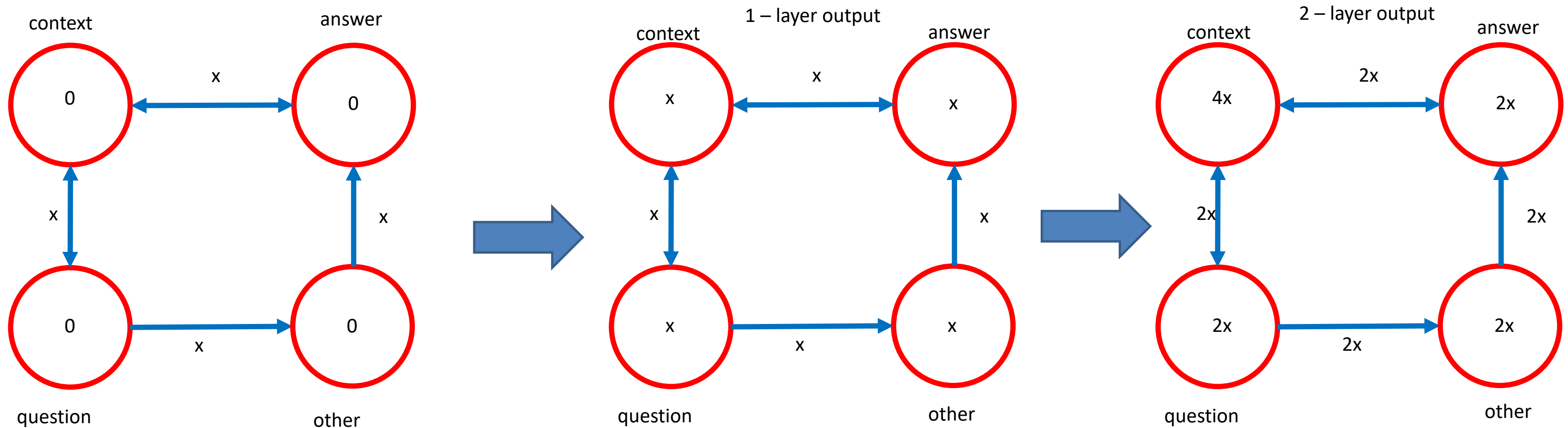
### GSC



Context노드는 모든 question node, answer node와 연결되도록 함

## Is it effective for reasoning?

### GSC



이렇게 subgraph 전체가 하나의 사이클을 이루면 똑같이 해결을 못함

## Is it effective for reasoning?

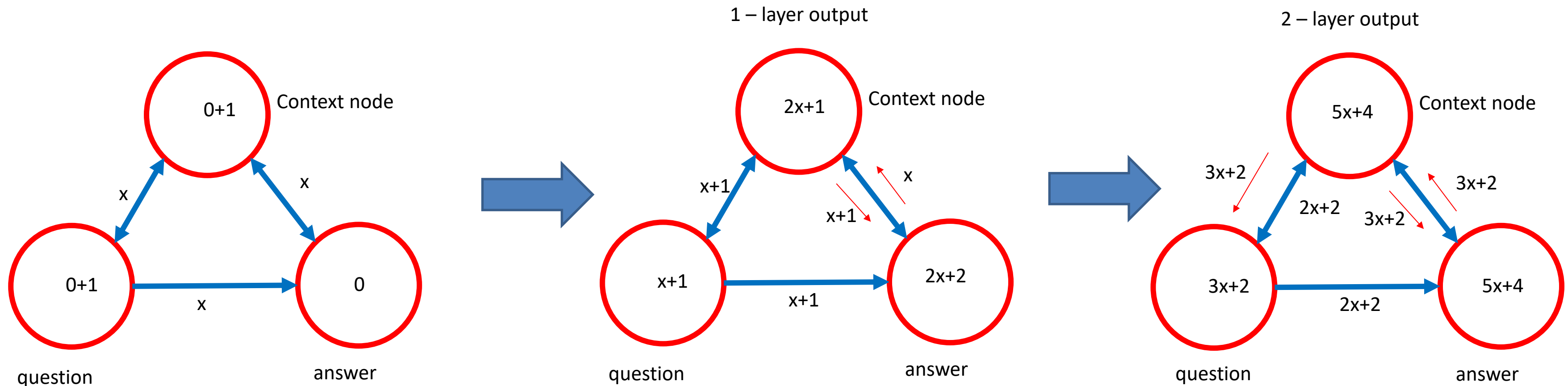
### IDGSC

- 여전히 GSC에도 한계는 있음
- 똑같이 해결 못 함(물론 모든 에지 노드가 다 같은 값일 경우, computational graph는 다름)
- 그렇다면 id-gnn fast의 node coloring feature를 더하면 어떨까? -> 성능 차이 없었음
- 없던 이유를 알아냄 -> 길이 2에 대한 feature를 더했기 때문 -> 길이가 3이상이면 효과가 있음
- 그런데 값이 커지면 추론을 잘한건가? -> 애초에 값이 커진 subgraph가 오답의 subgraph였다면? -> 성능의 차이가 없었던게 당연한건가?

# Is it effective for reasoning?

## IDGSC

노드 초기값에 길이 3의 augmented feature값을 추가



전과 달리 context node값이 달라짐

# Is it effective for reasoning?

## Summary

- 기존 GNN은 다른 형태의 두 그래프의 computational graph가 같아서 reasoning 결과도 같았음
- 이를 해결하기 위해 identity coloring(ID-GNN)이 나옴 + 기존 GNN의 flow와 다른 GSC가 있음
- GSC에 identity coloring을 주입해보자
- 다른 모델에도 identity coloring을 주입해보자

# What to do next

## Contribution & Future work

1. 현재 GNN을 통한 NLP에서 reasoning은 인간의 reasoning과 다름
  - subgraph내에 있는 사이클이 문제가 되는지 안되는지에 대한 정의가 필요
    - Subgraph의 일부분이 cycle이면 output값이 커지므로 해당 답이 될 확률 증가
    - 하지만 정답이 아닌 subgraph에 cycle이 많을 경우는?? -> 오답이 됨
2. WL test, GNN의 문제점을 해결한 Cyclic Reasoning 제시함 -> 비동형인 그래프는 reasoning도 달라야 한다
3. Future 연구에 방향성을 생각해본다면?
  - human의 reasoning을 모방할 수 있을까, Cyclic Reasoning을 더 발전시킬 순 없을까

# Thank you!