

2021 고려대학교 대학혁신지원사업

# KU-Insight Miner AI선배 강의 추천

## Knowledge Graph를 활용한

## R-GCN기반 강의 추천 시스템 개발

고려대학교 디지털정보처 데이터 hub팀

데이터 사이언티스트 이진숙

2021.04.20~

# Chapter 1 : 강의수강 그래프와 GNN적용하기

- AI 학우기반추천에서 얻은 교훈
- Graph Neural Network(GNN)
- Graph Convolutional Network(GCN)
- 어떻게 그래프로 나타낼 것인가? Knowledge Graph

# AI선배 학우기반추천에서 얻은 교훈

평점데이터? 유저기반? 아이템기반? 관계기반!

**“수강이력과 평점데이터만으로는 개인의 선호를 반영하기에 부족하다.”**

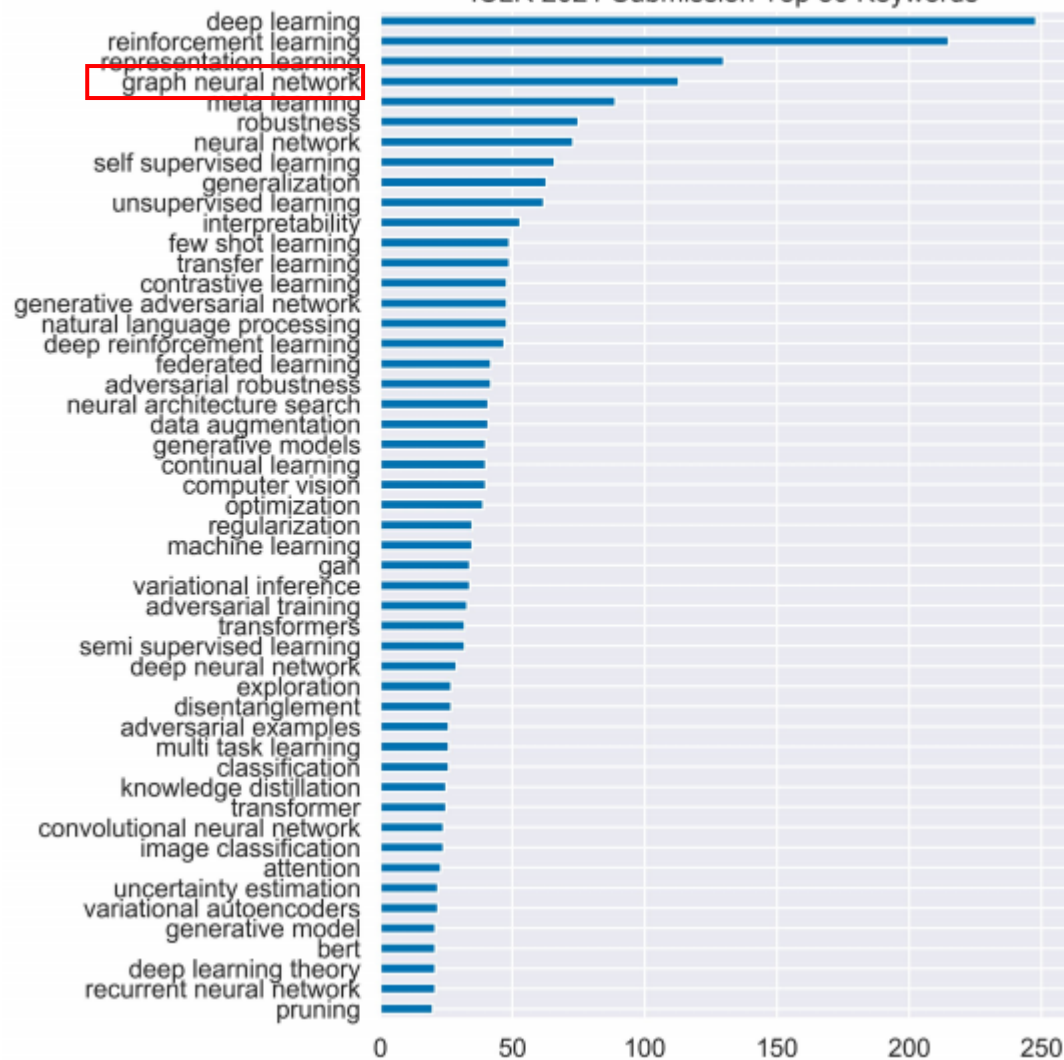
## 데이터에서 얻은 교훈

- 수강한 강의가 선호한 강의라고 보기 어렵다 → 수강신청 실패해서 차순위로 넣은 강의, 전공필수 등
- 평점데이터의 너무 치우친 분포 → 아무리 scaling해도 유저 기반으로 접근하기엔 무리가 있다.
- 갖고 있는 여러 종류의 데이터가 있다 → 강의데이터, 수강데이터, 성적데이터 등
- 새롭게 발견한 데이터들도 그동안 많이 쌓였다 → 희망과목데이터, 블랙보드 로그 데이터 등

## 방법론에서 얻은 교훈

- item based collaborative filtering으로 해결할 수 있지 않을까? → 과목기반추천과 겹칠 가능성
- 관계 기반의 데이터로 보자 → 유클리드 공간에 나타낼 수 없는 데이터 → 그래프 데이터로 해결
- 현존하는 데이터를 적극 활용할 방법 → node, edge feature, link로 이용

## Node와 Edge로 구성된 그래프 구조 데이터를 학습하는 신경망 모델



2008	2009	2010	2013	2015	2016	2017	2018	2019
GNN	GraphESN		GGNN				SSE	
		Spectral CNN		ChebNet		GCN	AGCN	DGCN
	NN4G			DCNN		MPNN	DGC	CGMM
			PATCHY-SAN		GraphSAGE		PGC-DGCNN	GaAN
					GAT		PGC	LGCN
					MoNet			
					DNGR		ARBGA	DGI
					SDNE		DRNE	
					VGAE		NetRA	
					GrammarVAE		Chemical VAE	GraphVAE
							SD-VAE	RGVAE
							DeepGMG	MolGAN
							GraphRNN	NetGAN
				Structural-RNN			GCRN	CGCN
							DCRNN	ST-GCN
								GaAN
								Graph-WaveNet
								ASTGCN

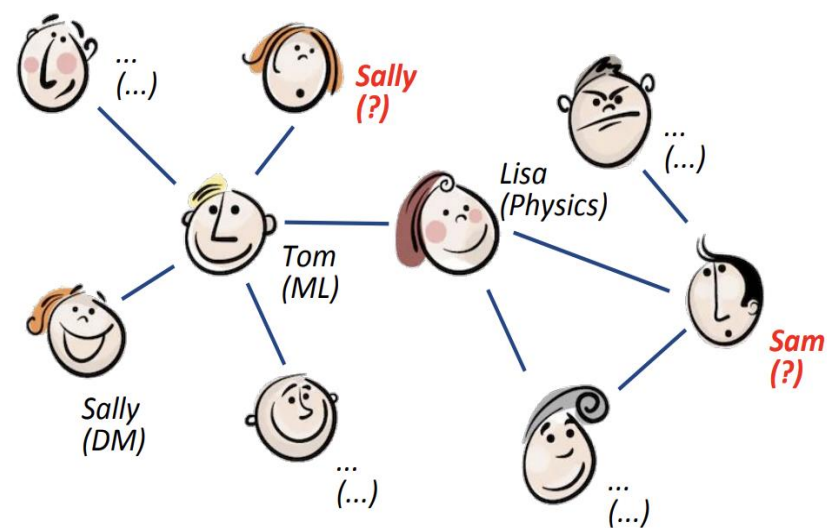
- RecGNNs
- ConvGNNs – Spectral
- ConvGNNs – Spatial
- GAEs – Network Embedding
- GAEs – Graph Generation
- STGNNs

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2020). A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*.  
[https://tootouch.github.io/research/gnn\\_summary/](https://tootouch.github.io/research/gnn_summary/)  
<https://petar-v.com/talks/GNN-Wednesday.pdf>

# 무궁무진한 GNN 활용분야!

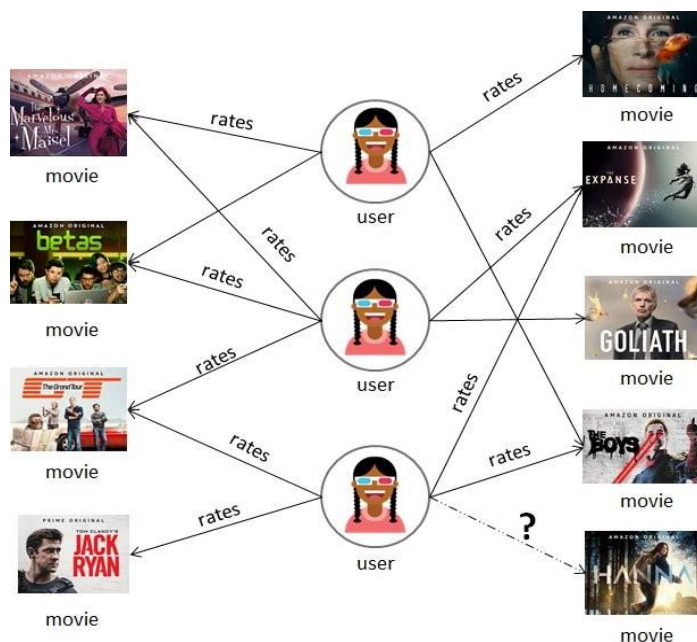
Node, Edge, Graph level task별로 살펴보기

## Node level



Node Classification(semi-supervised)  
(e.g. author's research field prediction)

## Edge level



1) Edge Classification  
(e.g. relationship type)

2) Link Prediction  
(e.g. recommendation)

## Graph level

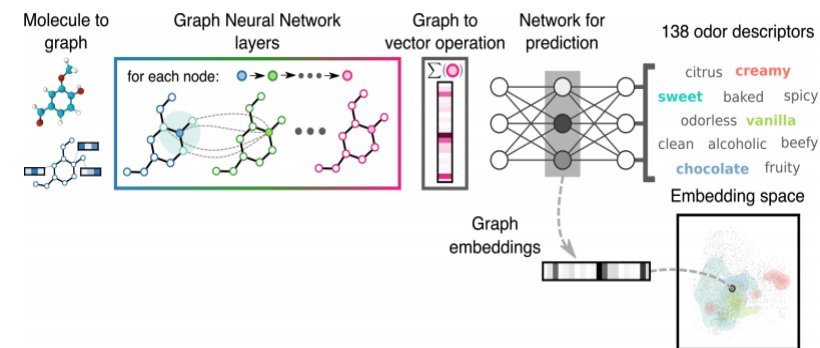
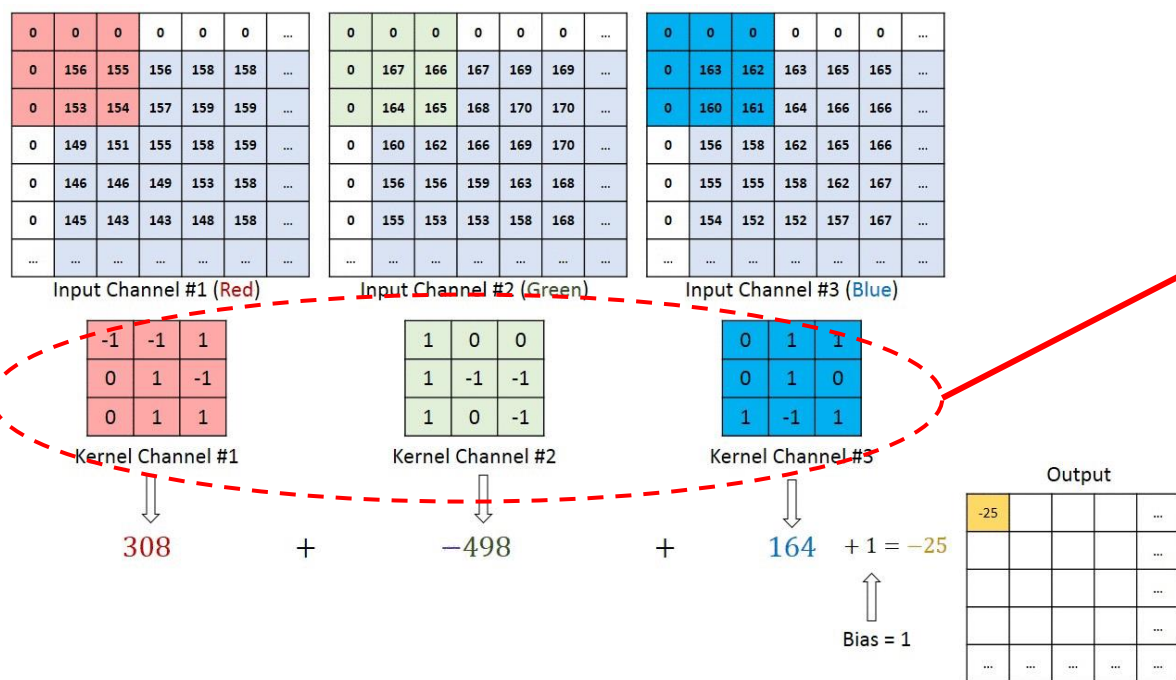


Figure 2: **Model Schematic.** Each molecule is first featured by its constituent atoms, bonds, and connectivities. Each Graph Neural Network (GNN) layer, here represented as different colors, transforms the features from the previous layer. The outputs from the final GNN layer is reduced to a vector, which is then used for predicting odor descriptors via a fully-connected neural network. We retrieve graph embeddings from the penultimate layer of the model. An example of the embedding space representation for four odor descriptors is shown in the bottom right; the colors of the regions in this plot correspond to the colors of odor descriptors in top right.

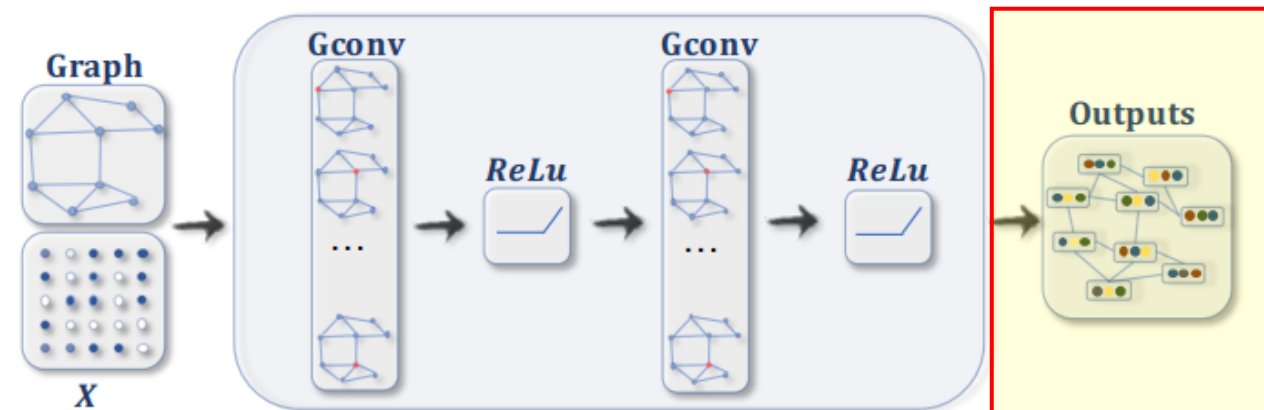
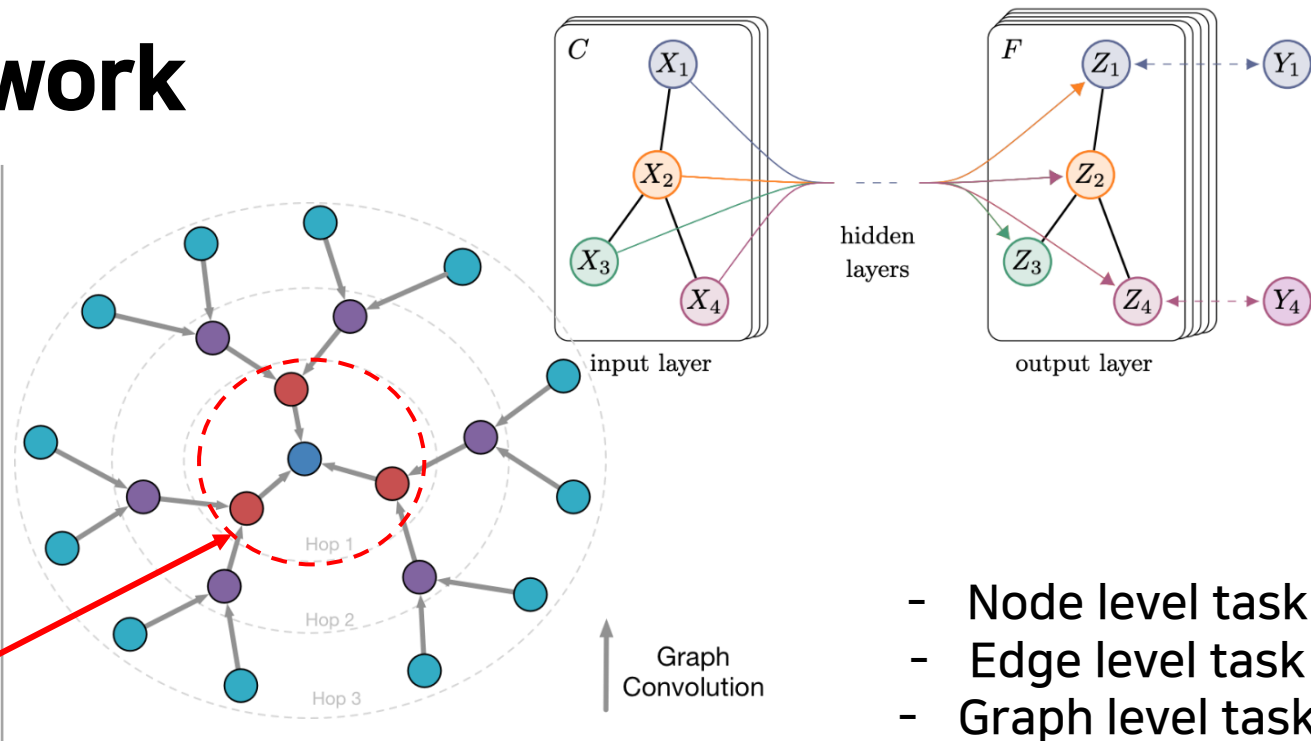
Graph classification  
(e.g. odor prediction)

# Graph Convolutional Network

Graph Neural Network의 기본



Convolutional Neural Network



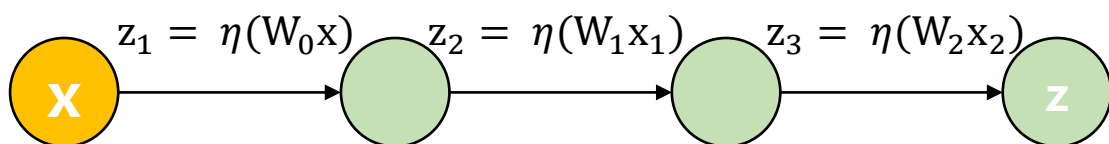
Graph Convolutional Network

# Graph Convolutional Network

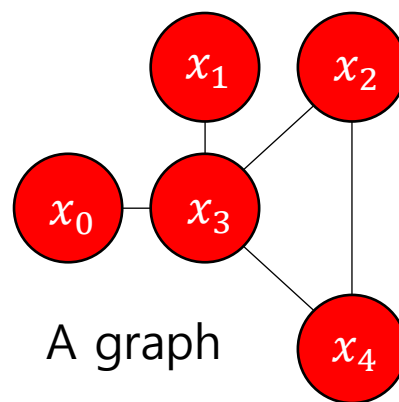
Graph Neural Network의 기본



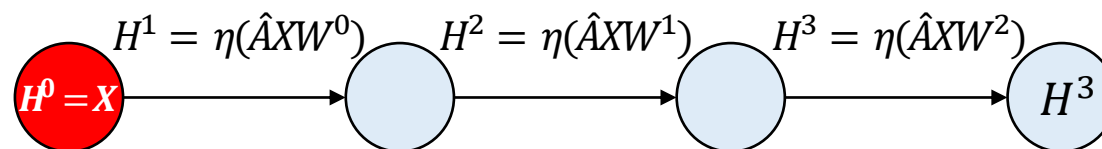
A node



NN with 3 hidden layers



$$\hat{A} = \begin{array}{c|ccccc} & 1 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 \\ \hline 0 & 0 & 1 & 1 & 1 & 1 \end{array}$$



GCN with 3 hidden layers

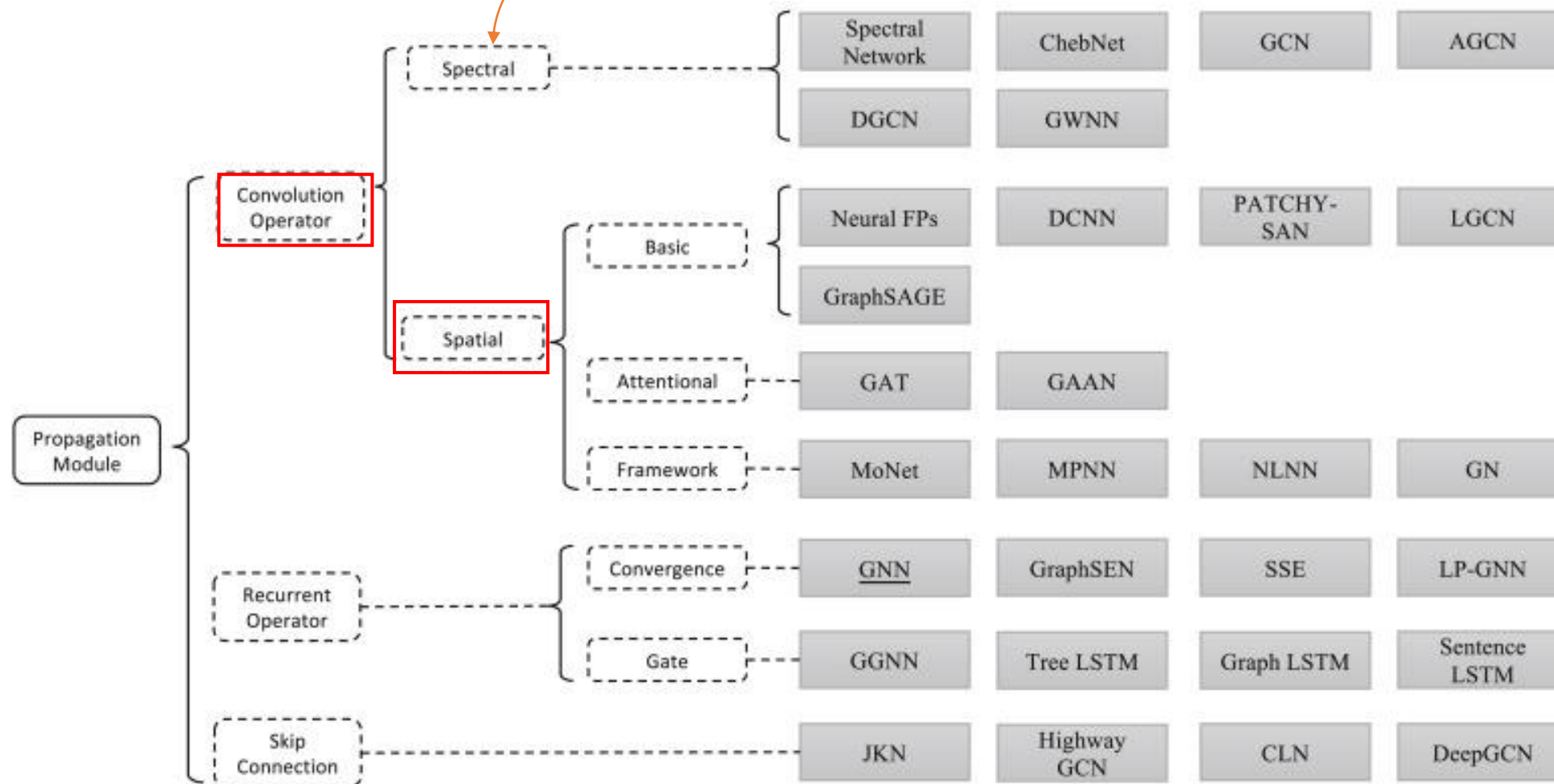


# Graph Neural Network 분류

Graph propagation module

Convolution할 때 Fourier transformation을 사용해서  
Eigenvector decomposition for Laplacian 연산을 하는데 노드의 제곱만큼 연산량이 늘어남

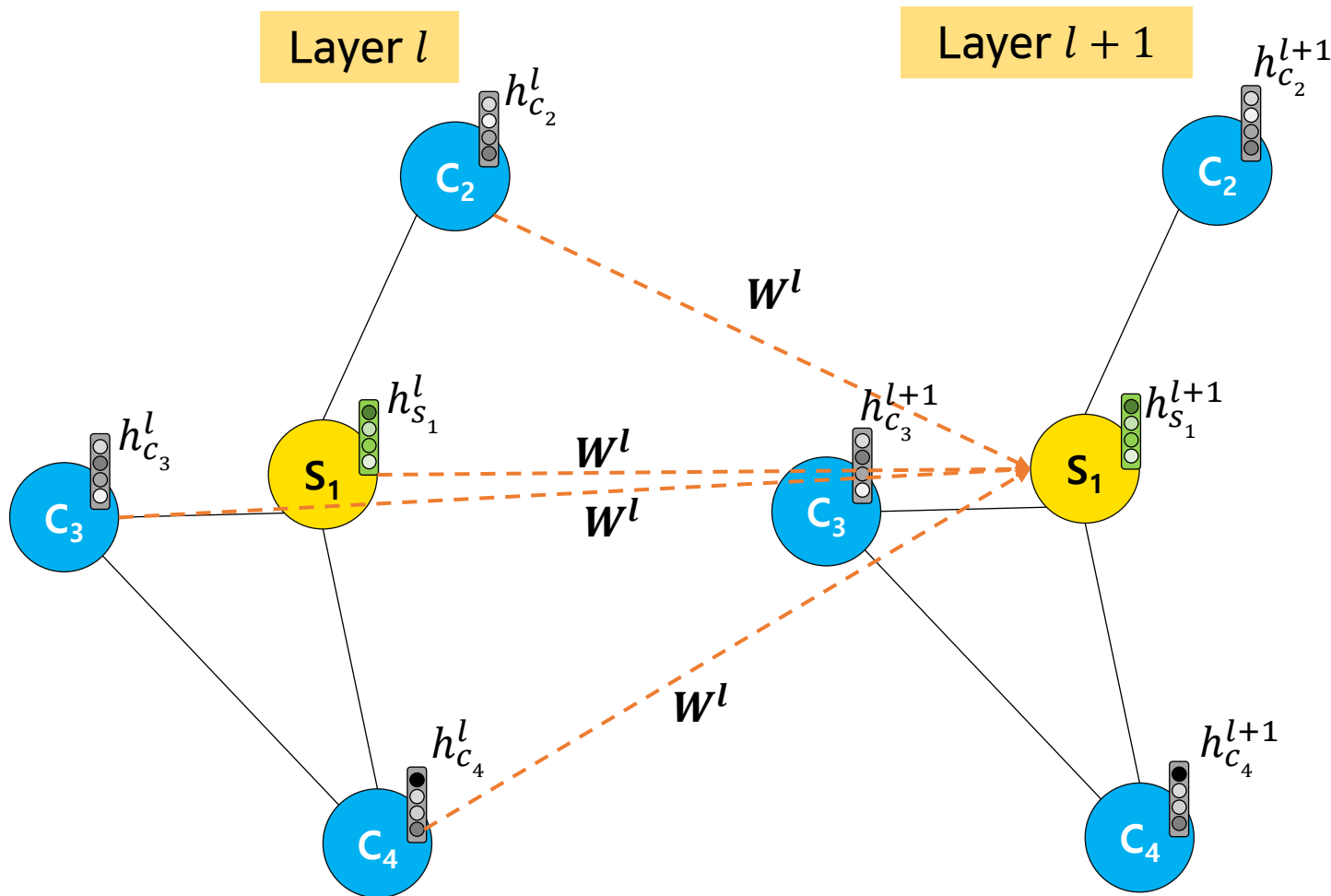
Low efficiency, generality, flexibility ☹️





# Vanilla Graph Convolutional Network(Spatial)

Simplest formulation of spatial GCNs



Matrix representation

$$h^{l+1} = \eta \left( \begin{matrix} n \times n & n \times d \\ D^{-1} A h^l W^l \\ n \times n & d \times d \end{matrix} \right)$$

$n \times d$

Vectorial representation

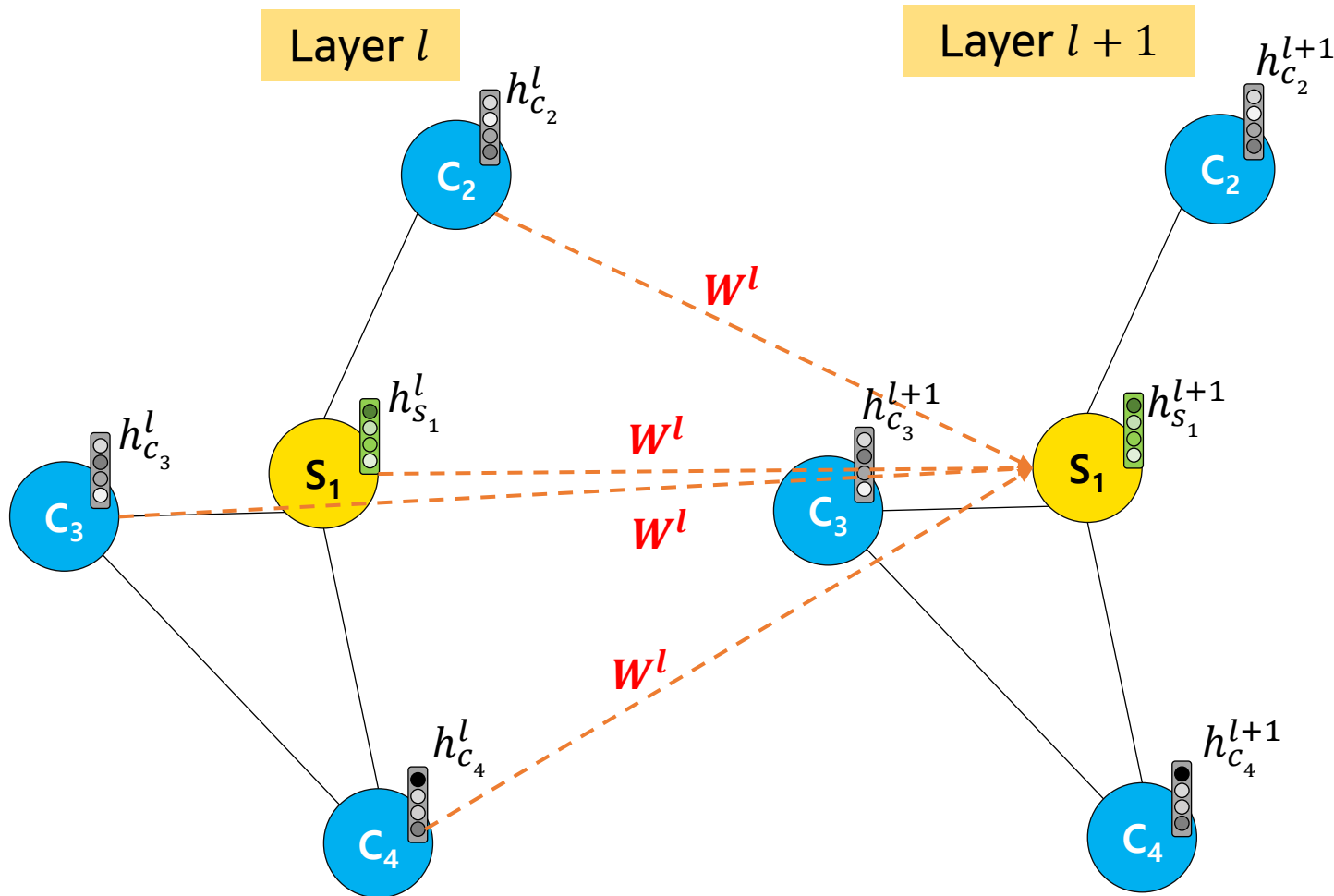
$$h_i^{l+1} = \eta \left( \underbrace{\frac{1}{d_i} \sum_{j \in N(i)} A_{ij}}_{\text{Mean}} \underbrace{W^l}_{d \times d} h_j^l \right)$$

$d \times 1$        $d \times d$        $d \times 1$

- $h^l$ :  $l$  번째 레이어의 hidden state
- $D$ : Degree matrix
- $A$ : Adjacency matrix
- $W^l$ :  $l$  번째 레이어의 가중치
- $\eta(\cdot)$ : non-linear activation function (e.g. ReLU)

# Vanilla Graph Convolutional Network(Spatial)

Simplest formulation of spatial GCNs



## 특징

- 노드 순서, 위치 고려하지 않아도 계산 가능
- 그래프 사이즈, 이웃의 크기에 상관없이 계산 가능
- 이웃의 정보만 고려(Locality)
- Weight sharing (convolution property)
- Limited to isotropic capability (same  $w^l$ )

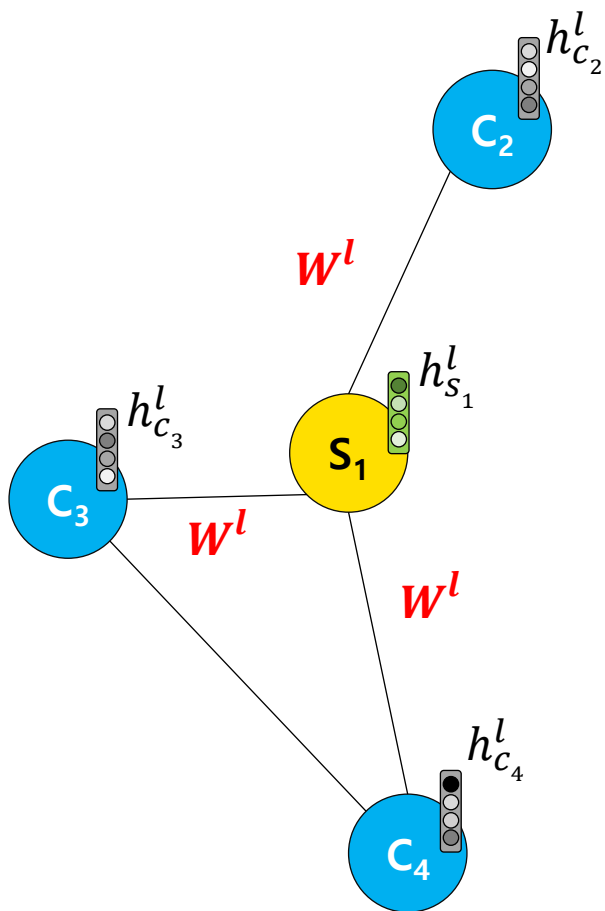
*GNN is basically you change this function...*

$$h_i^{l+1} = \boxed{fGCN}(h_i^l, \{h_j^l : j \rightarrow i\})$$

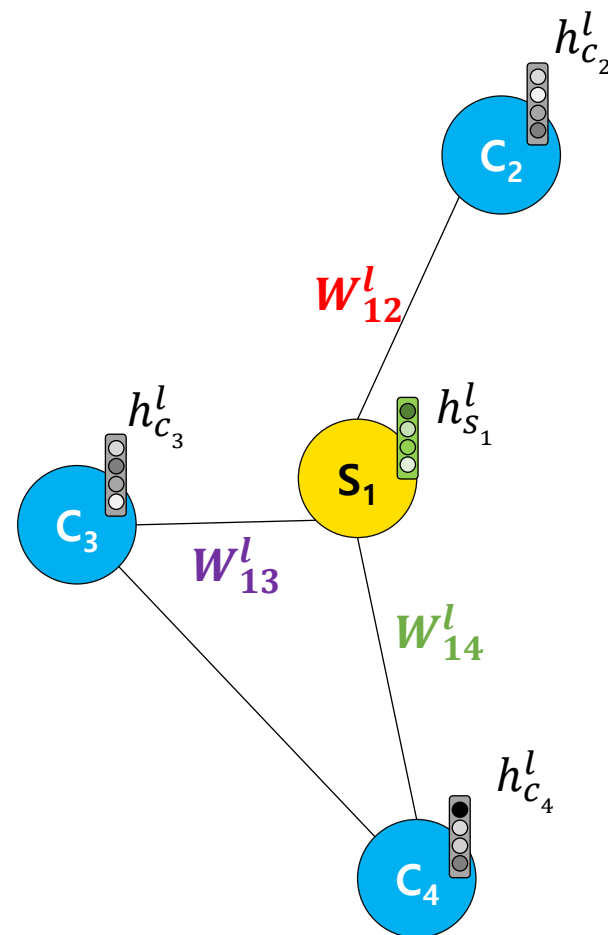
# Anisotropic GCNs

Isotropic vs Anisotropic weights

Isotropic weights

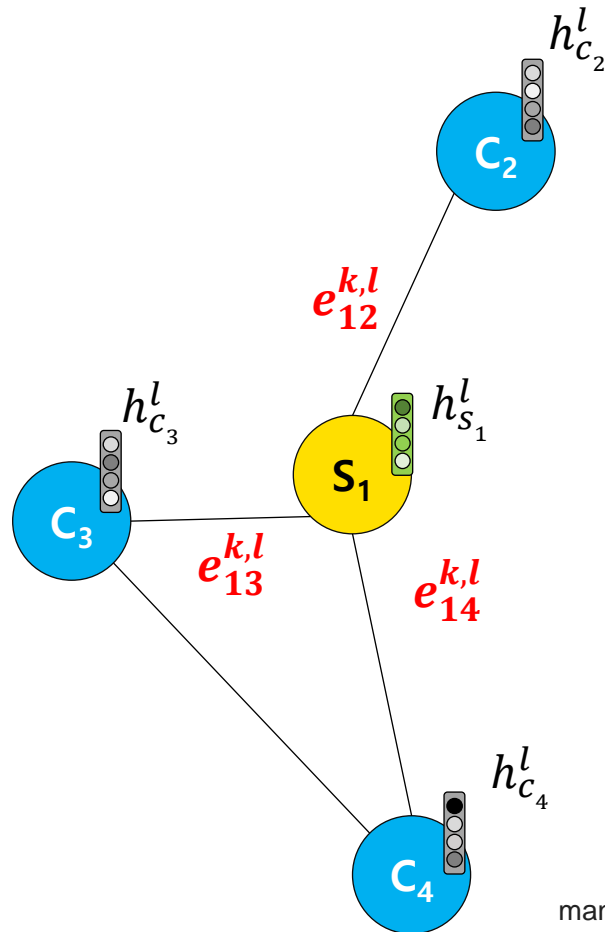


Anisotropic weights



# Anisotropic GCNs

Aggregation하는 방법에 따라 여러가지로 나뉨



## GatedGCNs[2] : Using edge gates

Edge features are explicit  
(important for edge prediction tasks)

Residual connections and batch normalization enhance  
learning speed and generalization

$$h_i^{l+1} = h_i^l + \text{ReLU}(\text{BN}(W_1^l h_i^l + \sum_{j \in N(i)} e_{ij}^l \odot W_2^l h_j^l))$$

$d \times 1$                        $d \times d$   $d \times 1$                        $d \times 1$   $d \times d$   $d \times 1$

$$e_{ij}^l = \frac{\sigma(\hat{e}_{ij}^l)}{\sum_{j' \in N(i)} \sigma(\hat{e}_{ij'}^l) + \varepsilon}$$

$$\hat{e}_{ij}^l = \hat{e}_{ij}^l + \text{ReLU}(\text{BN}(V_1^l h_i^{l-1} + V_2^l h_j^{l-1} + V_3^l \hat{e}_{ij}^{l-1}))$$

$d \times d$                        $d \times d$                        $d \times d$

[1] Monti, F., Boscaini, D., Masci, J., Rodola, E., Svoboda, J., & Bronstein, M. M. (2017). Geometric deep learning on graphs and manifolds using mixture model cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5115-5124).

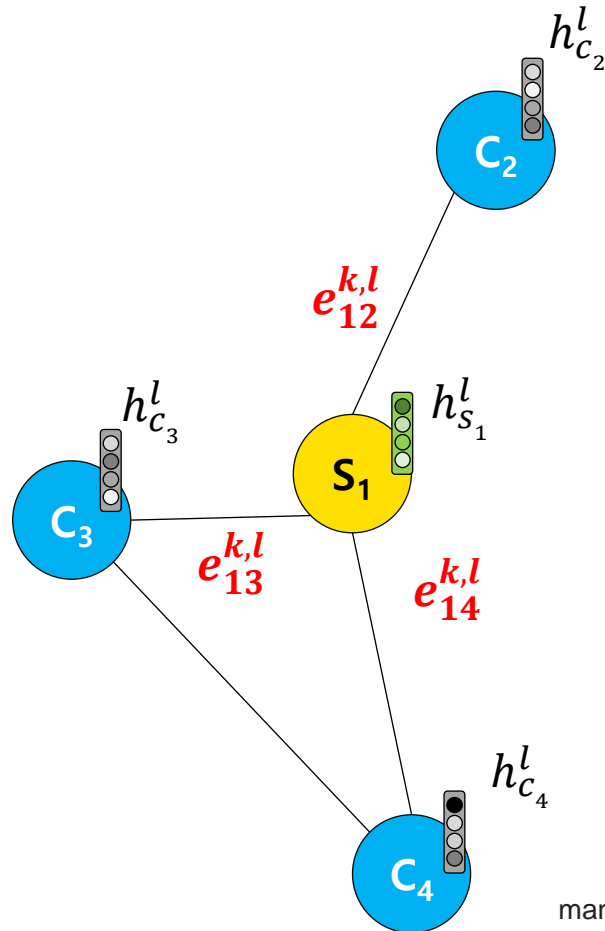
[2] Bresson, X., & Laurent, T. (2017). Residual gated graph convnets. *arXiv preprint arXiv:1711.07553*.

[3] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.

[https://www.youtube.com/watch?v=liv9R6BjxHM&ab\\_channel=AlfredoCanziani](https://www.youtube.com/watch?v=liv9R6BjxHM&ab_channel=AlfredoCanziani)

# Anisotropic GCNs

Aggregation하는 방법에 따라 여러가지로 나뉨



## GAT[3] : Using attention mechanism

Some nodes are more important given by softmax  
중요한 노드들까지도 학습하여 aggregate

$$h_i^{l+1} = \text{Concat}_{k=1}^K \left( \text{ELU} \left( \sum_{j \in N_i} \overset{\text{Scalar}}{e_{ij}^{k,l}} \underset{\frac{d}{K} \times d}{W_1^{k,l}} h_j^l \right) \right)_{d \times 1}$$

$$e_{ij}^{k,l} = \text{Softmax}_{N_i} (\hat{e}_{ij}^{k,l}) = \frac{\exp(\hat{e}_{i,j}^{k,l})}{\sum_{j \in N_i} \exp(\hat{e}_{i,j}^{k,l})}$$

$$\hat{e}_{i,j}^{k,l} = \text{LeakyReLU} \left( \underset{1 \times \frac{2d}{K}}{W_2^{k,l}} \text{Concat} \left( \underset{\frac{2d}{K} \times 1}{W_1^{k,l} h_i^l, W_1^{k,l} h_j^l} \right) \right)$$

[1] Monti, F., Boscaini, D., Masci, J., Rodola, E., Svoboda, J., & Bronstein, M. M. (2017). Geometric deep learning on graphs and manifolds using mixture model cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5115-5124).

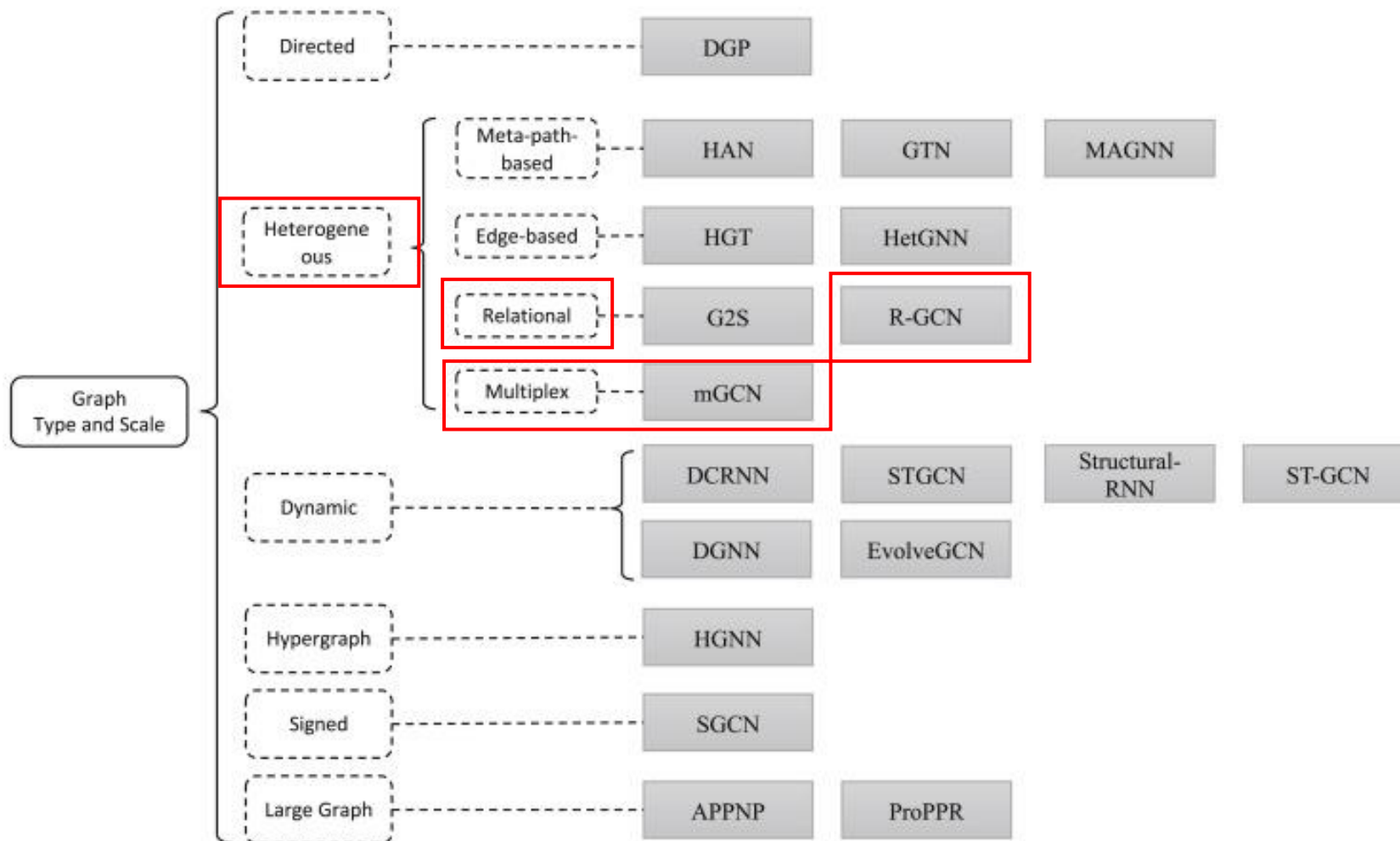
[2] Bresson, X., & Laurent, T. (2017). Residual gated graph convnets. *arXiv preprint arXiv:1711.07553*.

[3] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.

[https://www.youtube.com/watch?v=liv9R6BjxHM&ab\\_channel=AlfredoCanziani](https://www.youtube.com/watch?v=liv9R6BjxHM&ab_channel=AlfredoCanziani)

# Graph Neural Network 분류

Graph type을 무조건 고려해야 한다



# 어떻게 그래프로 나타낼 것인가?

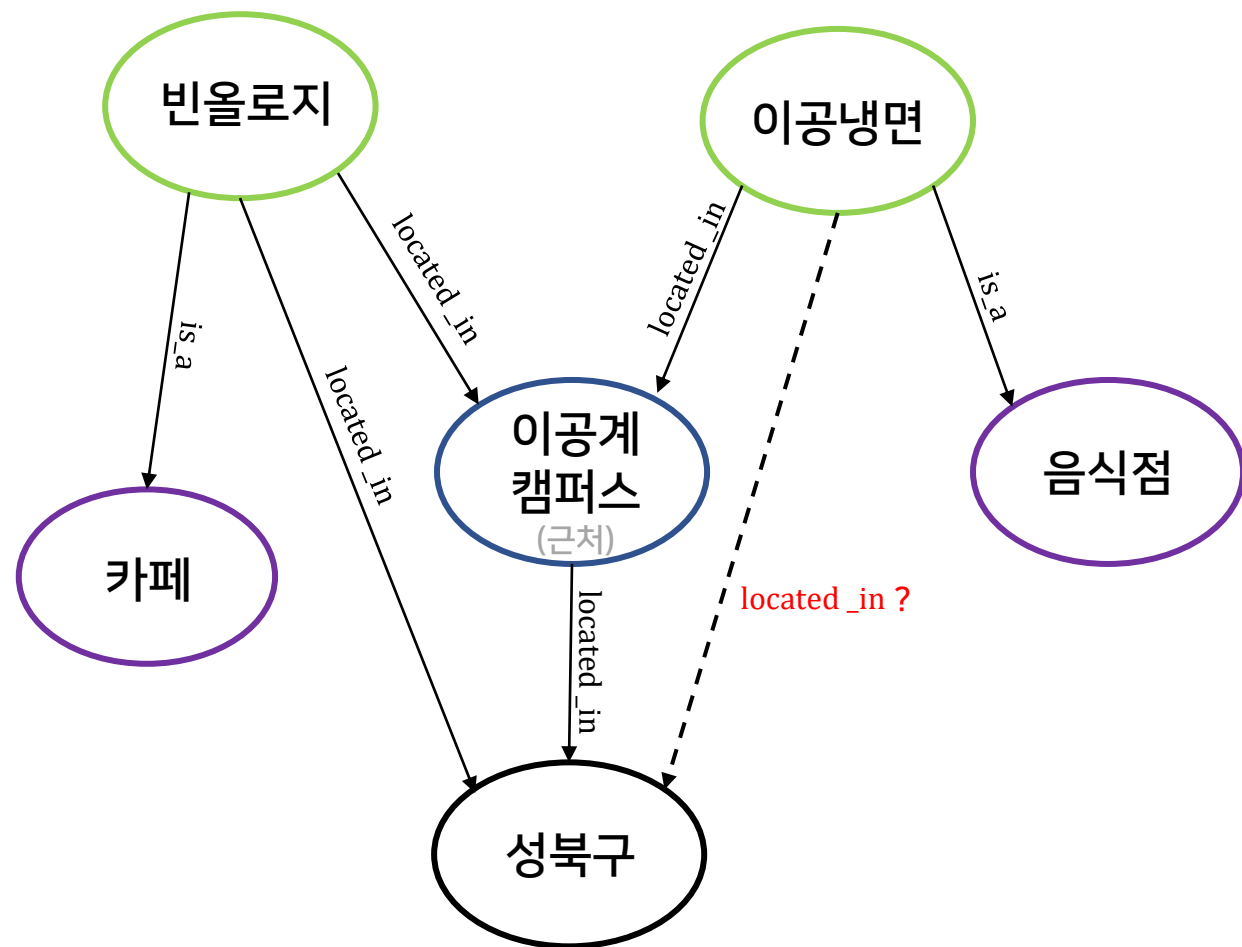
## Knowledge Graph & Knowledge Base

### Knowledge Graph란?

- Knowledge graph(KG) 는 Knowledge base(KB) 를 그래프 형식으로 나타낸 것으로 각각의 entity와 그 사이의 관계를 node와 edge로 나타낸 것
- (*head, relation, tail*) triplet으로 나타냄  
(e.g. (이공냉면, located\_in, 이공계캠퍼스))

### Knowledge Graph Completion

- 대부분의 갖고 있는 데이터는 불완전하기 때문에 Knowledge Graph도 불완전
- GNN으로 추가적인 정보 수집없이 새로운 정보를 예측할 수 있도록 하는 것이 목적 → Link prediction task



예) 고려대 근처 음식점 및 카페 Knowledge Graph



# 어떻게 그래프로 나타낼 것인가?

Knowledge Graph for Course Recommender System

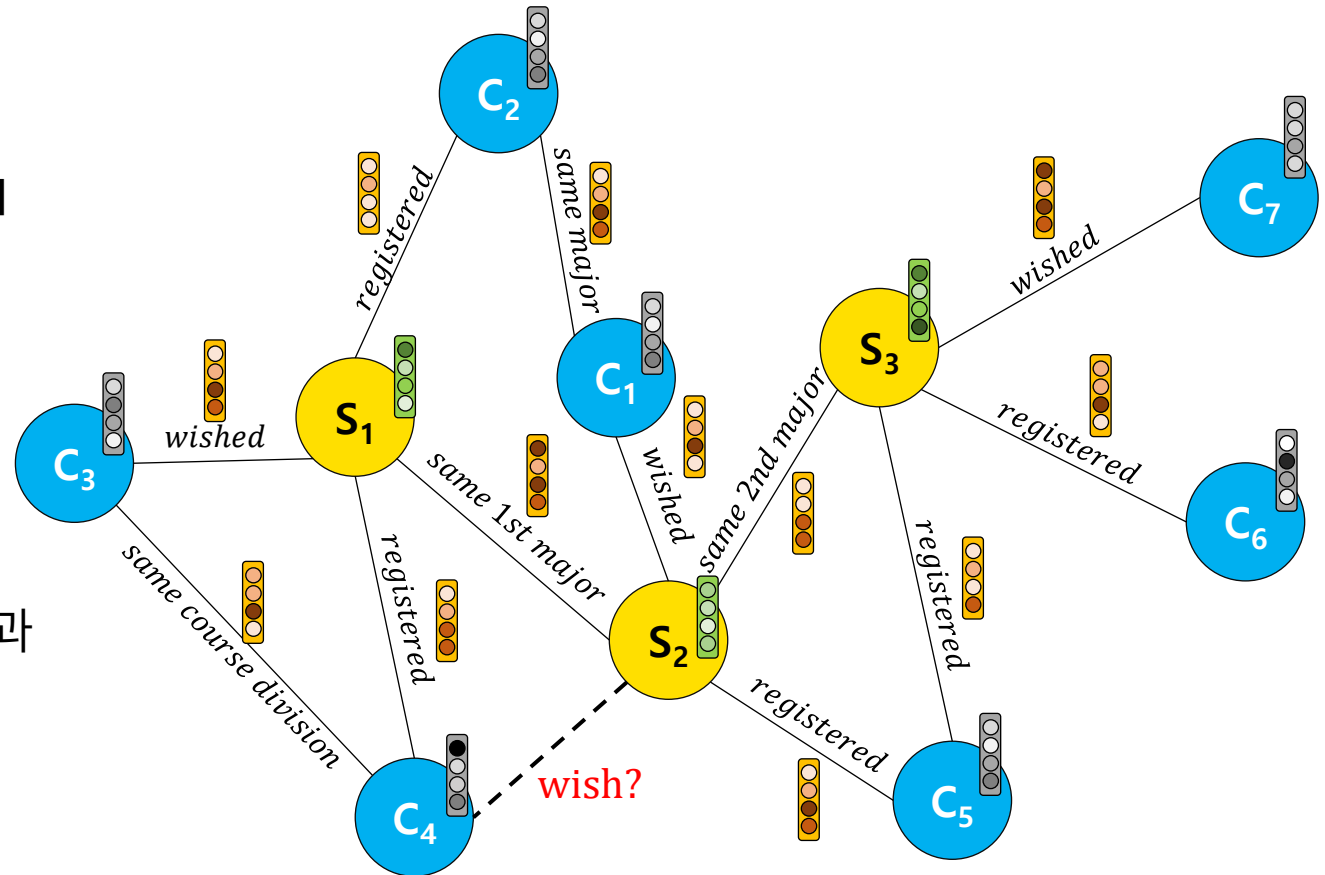
## 강의 수강 Knowledge Graph

- Graph type : Heterogenous, undirected
- Node(vertex) : 학생, 강의
- Edge : Relationships

R1 : 학생-강의 : 희망, 수강

R2 : 학생-학생 : 같은 1전공, 같은 2전공

R3 : 강의-강의 : 같은 이수구분, 같은 개설학과



# 어떻게 그래프로 나타낼 것인가?

Knowledge Graph for Course Recommender System

강의 수강 Knowledge Graph : Relationship이 여러개라면? → This is more like the real world!

- Graph type : **Heterogenous, undirected**
- Node(vertex) : 학생, 강의
- Edge : **Relationships**

R1 : 학생-강의 : 희망

R1-1 : 학생-강의 : 수강

R2 : 학생-학생 : 같은 1전공

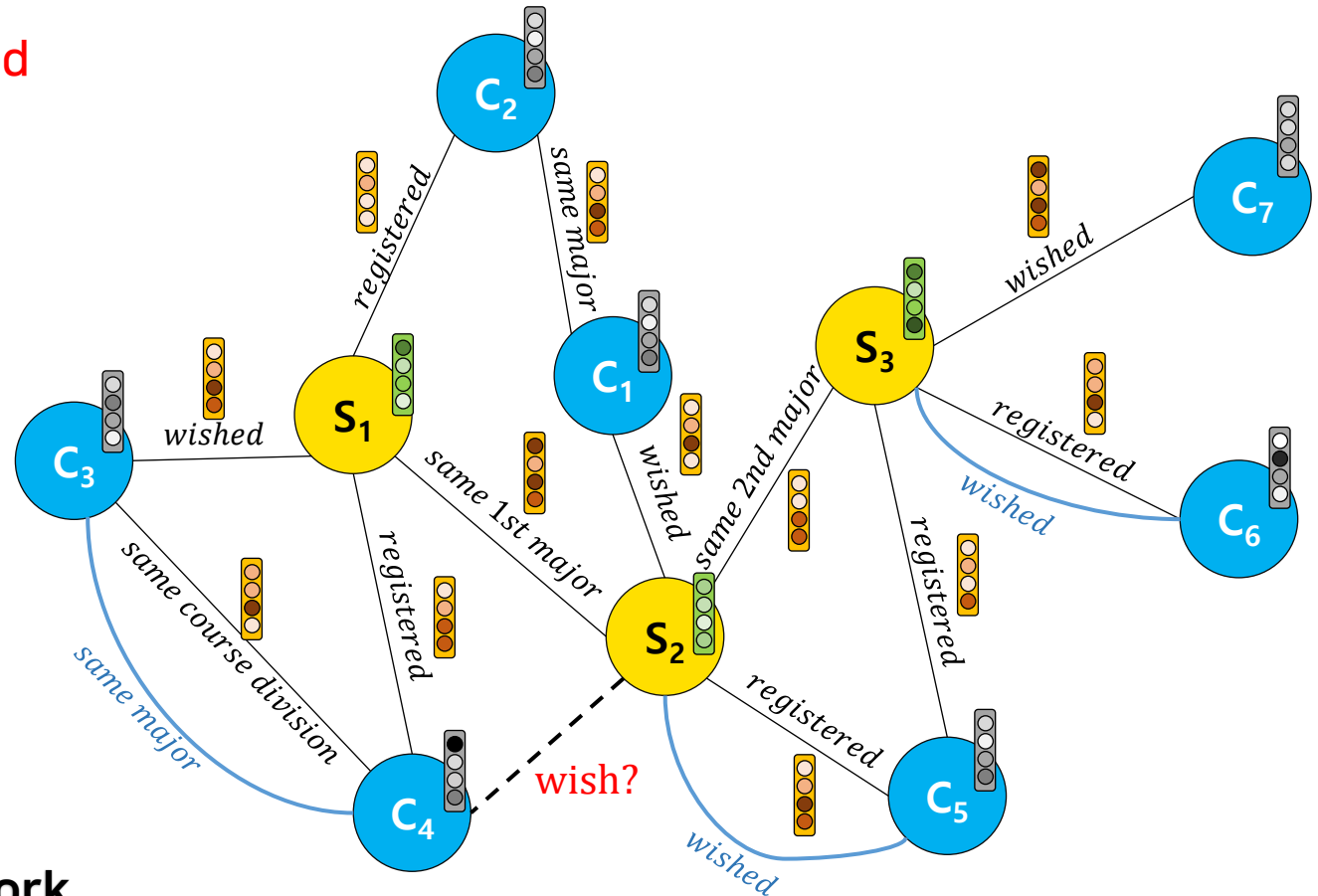
R2-1 : 학생-학생 : 같은 2전공

R3 : 강의-강의 : 같은 이수구분

R3-1 : 강의-강의 : 같은 개설학과

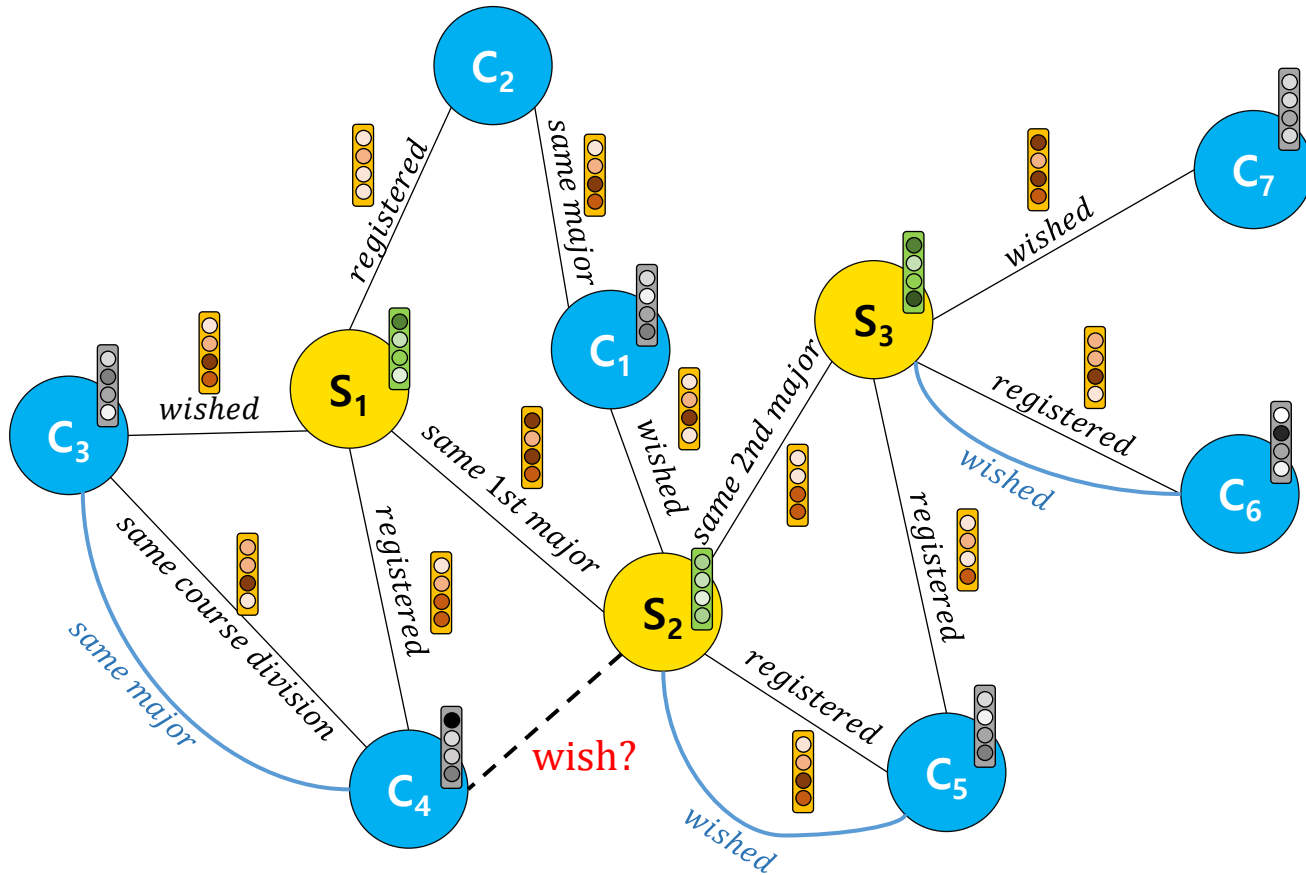


**Multi-Relational Graph Convolutional Network**



# Chapter1 Wrap up!

## Knowledge Graph for Course Recommender System



## Propagation Module

- GNN 중 Convolution propagation으로 접근
- Spectral Convolution 은 높은 비용이 소모되므로 Spatial Convolution 선택
- Isotropic한 모델들은 가중치를 다르게 두지 않기 때문에 Anisotropic 모델 선택(MoNets, GGCN, GAT 등)



## Graph type

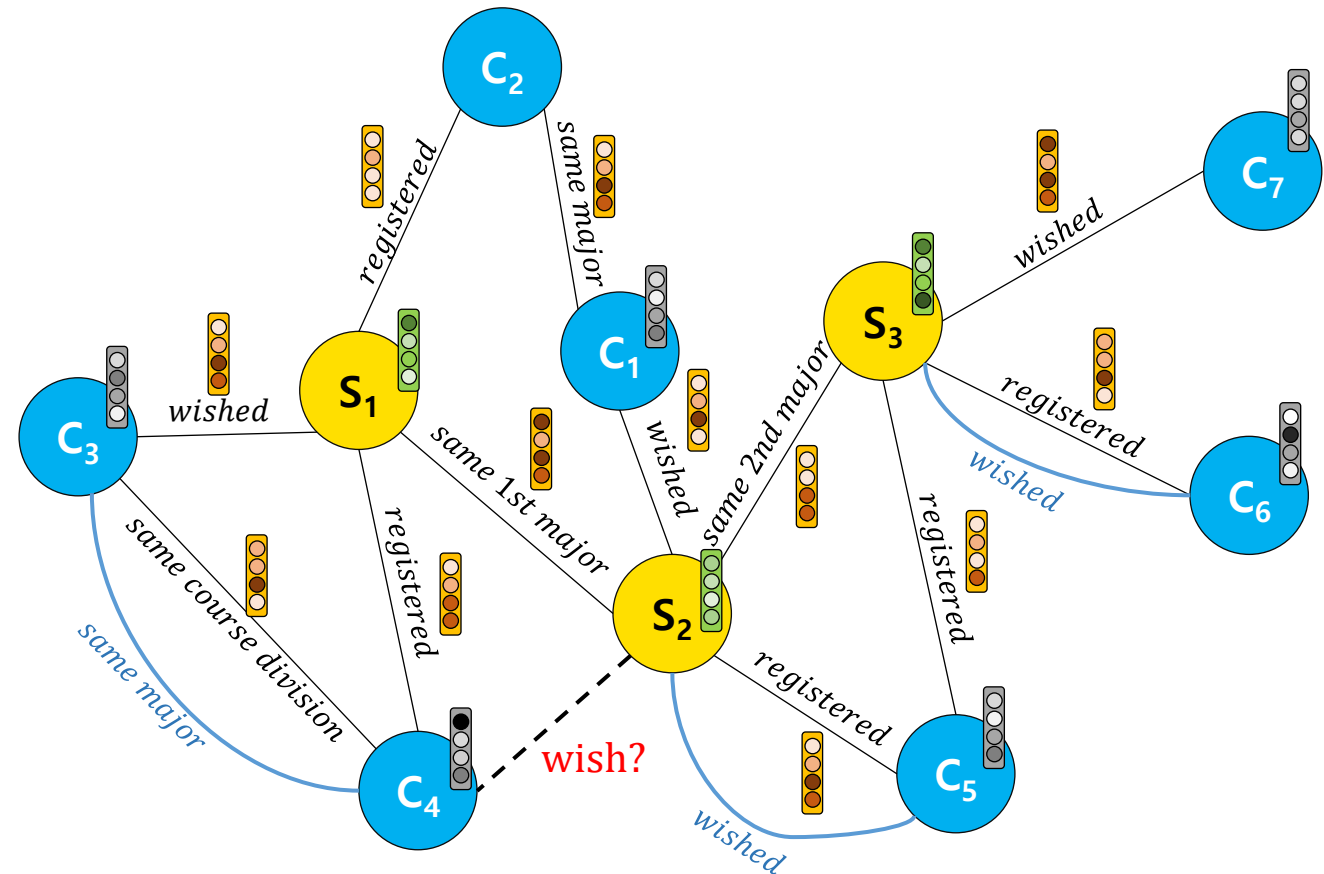
- Graph type : Heterogenous, undirected
- Edge type : Relationships
- 노드의 관계 정보를 알아내고자 하므로 R-GCN선정
- 두 노드 사이의 여러 차원의 관계가 존재할 수 있으므로 multi-relational graph로 접근



## Multi-Relational Graph Convolutional Network

# Chapter 2 : 최종모델선택

## - R-GCN



# R-GCN<sup>[1]</sup>

Relational Graph Convolutional Network

$$h_i^{l+1} = \sigma(\sum_{m \in \mathcal{M}_i} g_m(h_i^l, h_j^l))$$

where

$h_i^l \in \mathbb{R}^{d^l}$  : hidden state of  $l$ th layer

$d^l$  : dimension of  $l$  layer

$\sigma(\cdot)$ : element – wise activation function like ReLU

$\mathcal{M}_i$  : set of incoming messages for node  $v_i$  (often identical to the set of incoming edges)

$g_m(\cdot, \cdot)$ : accumulated and passed through dimension of  $l$  layer

\* note: message specific neural network or simply a linear transformation  $g_m(h_i, h_j) = Wh_j$

$W$  weight matrix like [2] suggested

Graph :  $G = (V, \mathcal{E}, R)$

Node :  $v_i \in V$

Relationship type :  $r \in R$

Edges :  $(v_i, r, v_j) \in \mathcal{E}$

# R-GCN<sup>[1]</sup>

Forward-pass update an entity or node denoted by  $v_i$  in a relational + directed and labeled multi-graph

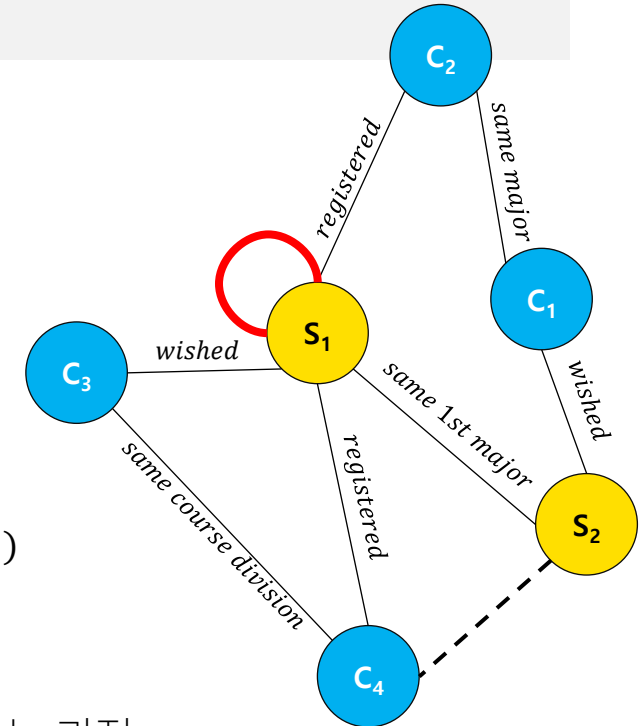
## Propagation model

$$h_i^{l+1} = \sigma(\underbrace{W_0^l h_i^l}_{\text{self-loop}} + \underbrace{\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{c_{i,r}} W_r^l h_j^l}_{\text{Neighbors}})$$

where

$N_i^r$  : set of neighbor indices of node  $i$  under relation  $r \in R$

$c_{i,r}$  : a problem – specific normalization (it can be learned or chosen in advance such as  $c_{i,r} = |N_i|$ )



Neighbor들로부터 전달된 feature vector들을 normalized summation하는 과정

기존의 GCN과는 다르게 relation-specific하게 transformation(엣지의 타입이나 방향성을 고려함)

# R-GCN<sup>[1]</sup>

Two ways of regularization

Rare relationship에서는 overfitting 위험이 있어 Regularization 진행(Reducing parameters)

## (1) Basis decomposition(linear combination of the number of components)

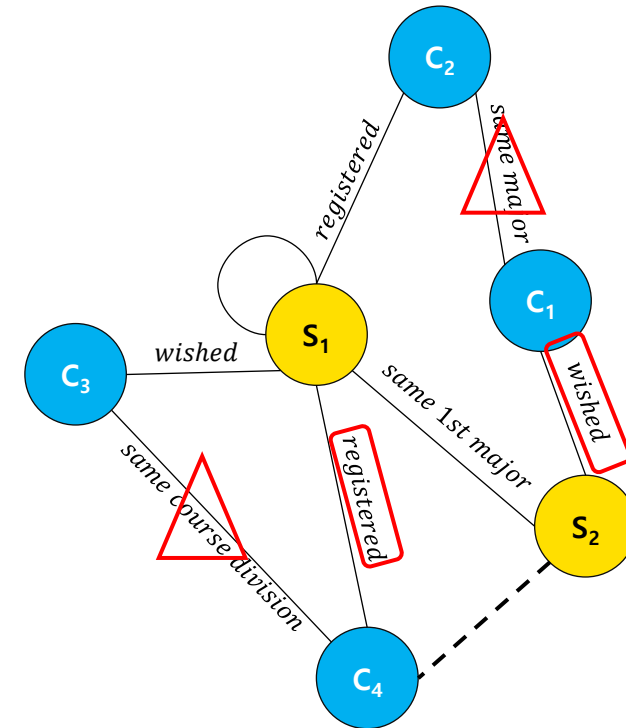
$$W_r^l = \sum_{b=1}^B a_{rb}^l V_b^l$$

$V_b^l \in \mathbb{R}^{d^{l+1} \times d^l}$  with coefficients  $a_{rb}^l$

specifying the numbers of unique  $W$ s that you want to have for the layer  $\rightarrow B$

each of the  $W$ s is calculated by combining those components linearly  
so they learn a coefficient for each of the components

$$h_i^{l+1} = \sigma\left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{c_{i,r}} W_r^l h_j^l + W_0^l h_i^l\right)$$





# R-GCN<sup>[1]</sup>

Two ways of regularization

$$h_i^{l+1} = \sigma\left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{c_{i,r}} W_r^l h_j^l + W_0^l h_i^l\right)$$

Rare relationship에서는 overfitting 위험이 있어 Regularization 진행(Reducing parameters)

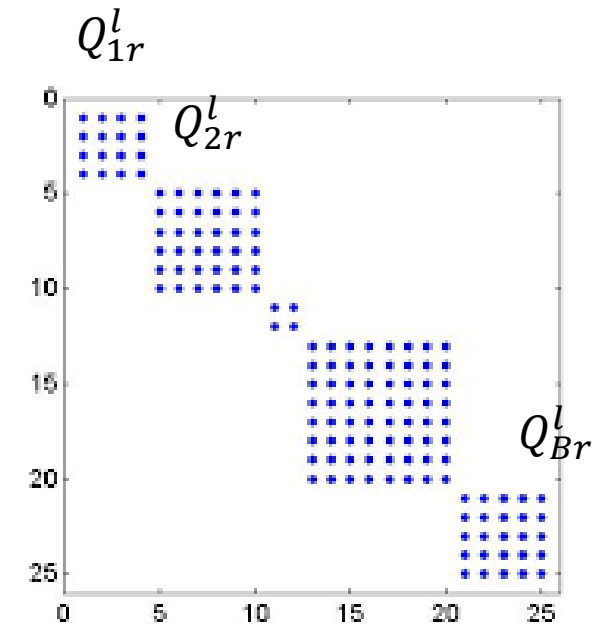
## (2) Block-diagonal decomposition

$$\begin{aligned} W_r^l &= \bigoplus_{b=1}^B Q_{br}^l \\ &= \text{diag}(Q_{1r}^l, \dots, Q_{Br}^l) \end{aligned}$$

$$Q_{br}^l \in \mathbb{R}^{(d^{l+1}/B) \times (d^l/B)}$$

*defined through direct sum over a set of low dimensional matrices*

*Variables are strongly interconnected within a group but don't have much interactions outside of the group*



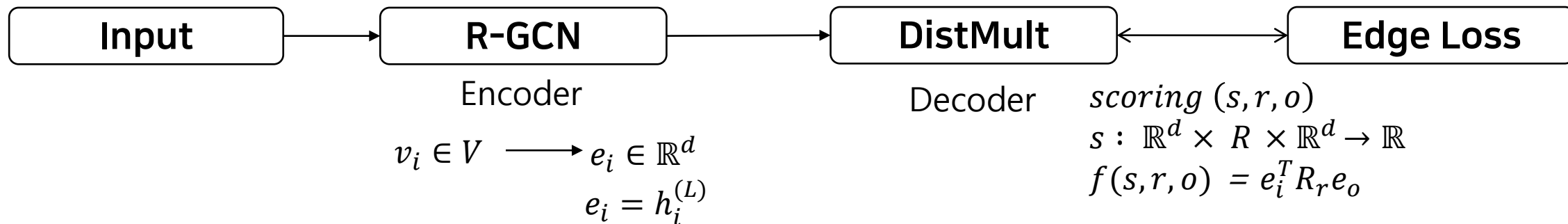
# R-GCN<sup>[1]</sup>

## Link Prediction Task Flow

Graph :  $G = (V, \mathcal{E}, R)$   
 Node :  $v_i \in V$   
 Relationship type :  $r \in R$   
 Edges :  $(v_i, r, v_j) \in \mathcal{E}$

### Link Prediction Task Flow

$\mathcal{E}$  대신에 불완전한 링크  $\hat{\mathcal{E}}$  제공,  $f(s, r, o)$  를 실제  $(s, r, o)$ 에 가깝게 예측



### Training

#### Sampling

Negative Sampling ( $\omega$ )

#### Optimization

Cross entropy

$$\mathcal{L} = -\frac{1}{(1 + \omega)|\hat{\mathcal{E}}|} \sum_{(s, r, o, y) \in \mathcal{T}} y \log(f(s, r, o)) + (1 - y) \log(1 - l(f(s, r, o)))$$

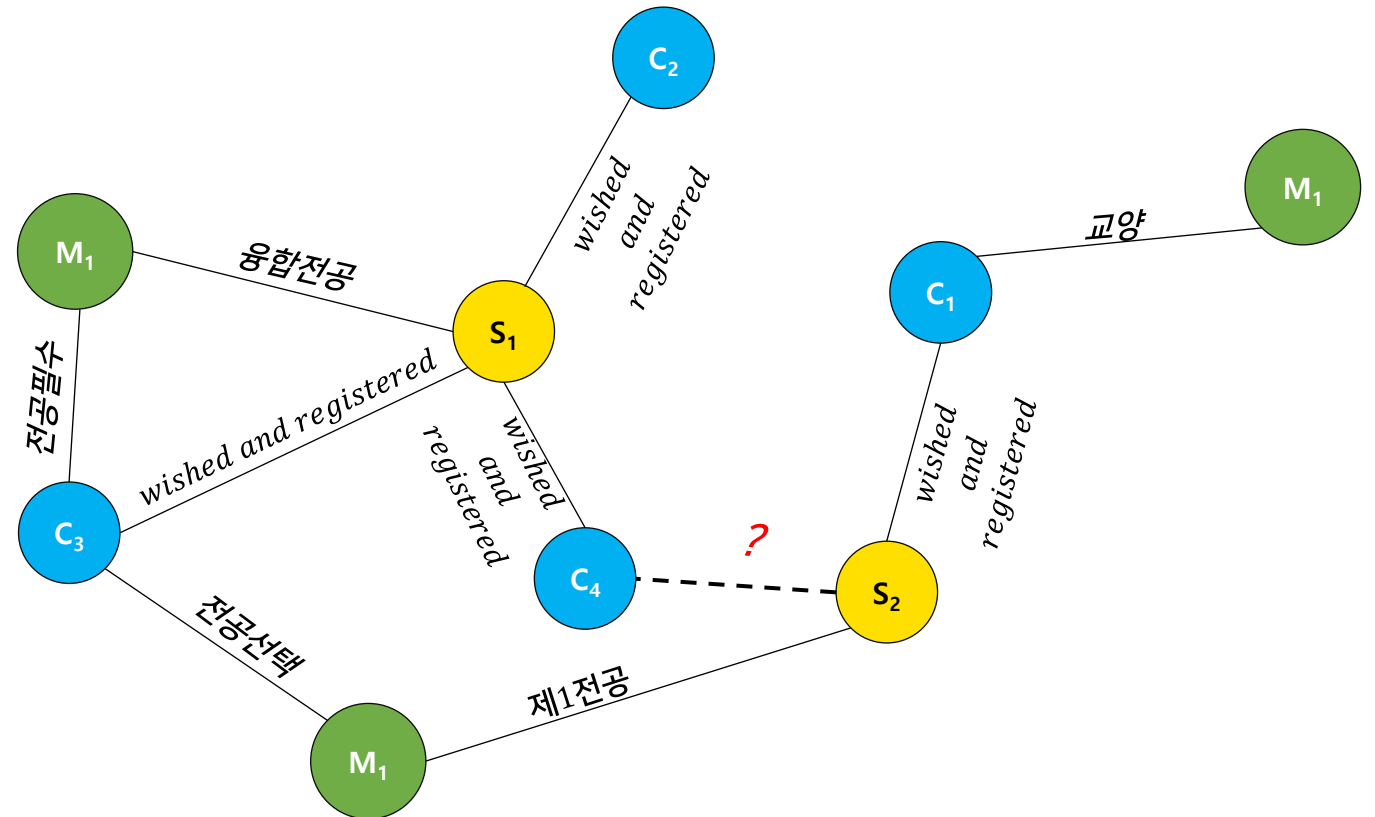
$\mathcal{T}$  : total set of real and corrupted triples

$l$  : logistic sigmoid function

$y$  :  $y = 1$  for positive triples and  $y = 0$  for negative ones

# Chapter 3 : 실제 적용 방법

- Knowledge Graph 정의 및 데이터
- 이용 모델
- 추천 Flow



# Knowledge Graph 정의 및 데이터

14학번~21학번의 수강이력, 관심과목등록, 학적정보, 제1전공, 2전공 커리큘럼

## 이용 데이터

### 1. 관심과목등록 $\cap$ 수강

기준 1) 2014~2021학번 서울캠퍼스학생

기준 2) 2014~2021년도 수강이력

\*주의\* 신입생은 교집합없이 수강이력만 이용

### 2. 제1전공, 제2전공 내역

### 3. 학과, 융합전공, 학생설계전공 커리큘럼

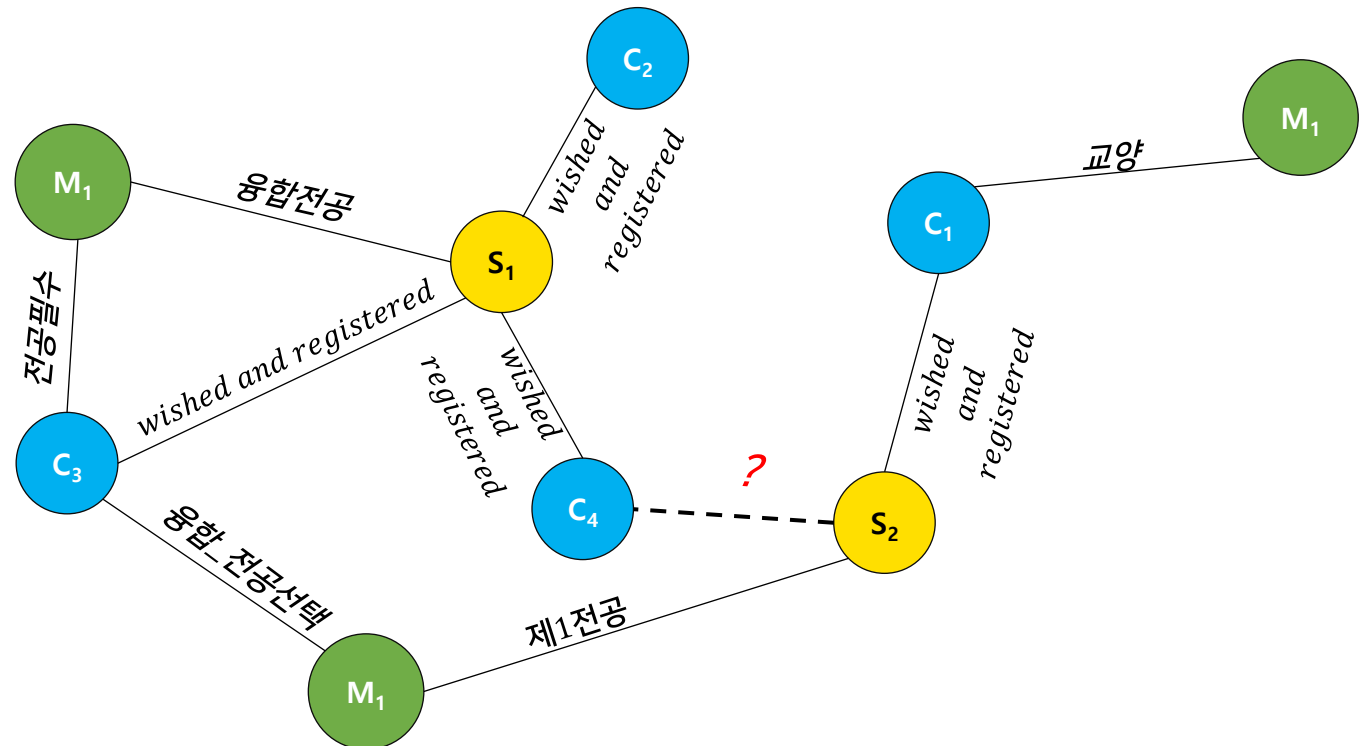
## Knowledge Graph

### 1. Vertex 정의

- 학생
- 강의
- 학과, 제2전공

### 2. Relation 정의

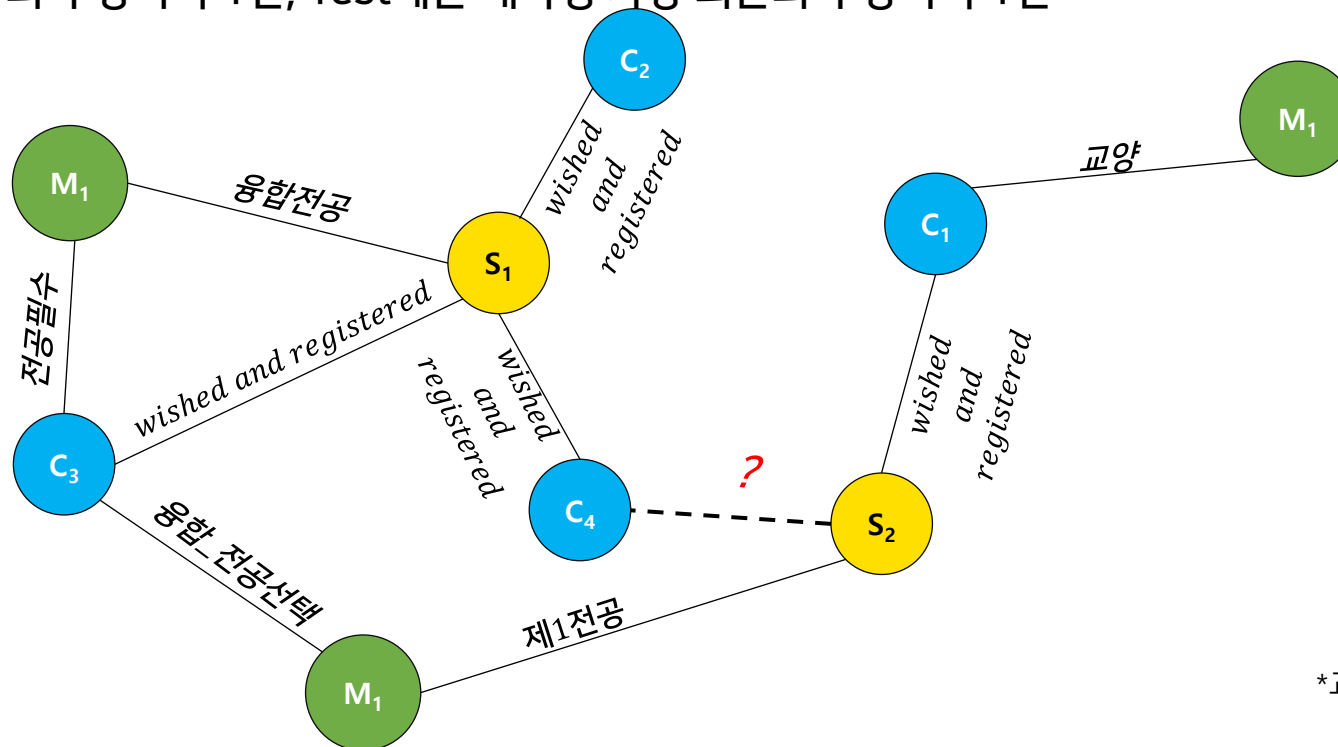
- 희망과목등록 AND 수강
- 전공, 제2전공 이수
- 어떤 이수구분으로 커리큘럼 포함



- Chapter 2에서 제시한 Knowledge Graph처럼 임베딩하고 싶었지만 엣지가 너무 많아 GPU memory out --> Knowledge graph 최대한 간소화

# \* IMPORTANT \* *Train and test set split*

Validation에는 졸업생 가장 최근의 수강이력 1건, Test에는 재학생 가장 최근의 수강이력 1건



\*교양의 경우에 학과는 교양교육원

Train triplets ( $N = 748,751$ )

1. 희망과목등록  $\cap$  수강

기준 1) 2014~2021학번 서울캠퍼스학생  
기준 2) 2014~2021년도 수강이력

2. 제1전공, 제2전공 내역

3. 학과, 이중, 융합, 학생설계 커리큘럼

Valid triplets ( $N = 8,956$ )

희망과목등록  $\cap$  수강 (졸업생)

기준 1) 2014~2021학번 서울캠퍼스학생  
기준 2) 졸업생 가장 최근학기의 이력 랜덤 1건

Test triplets ( $N = 27,625$ )

희망과목등록  $\cap$  수강 (재학생)

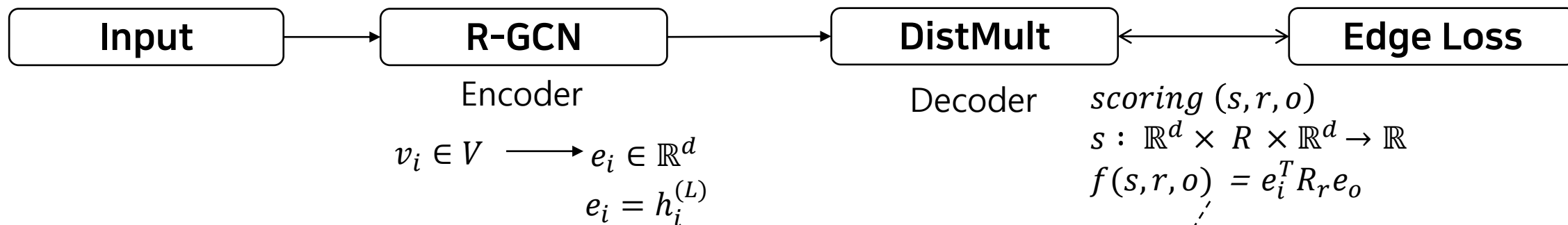
기준 1) 2014~2021학번 서울캠퍼스학생  
기준 2) 재학생 가장 최근학기의 이력 랜덤 1건

# 이용모델: R-GCN

Subject Perturbation과 Object perturbation

## Link Prediction Task Flow

$\varepsilon$  대신에 불완전한 링크  $\hat{\varepsilon}$  제공,  $f(s, r, o)$  를 실제  $(s, r, o)$ 에 가깝게 예측



## DistMult

Relation Embedding의 한 방법으로 가중치의 대각행렬을 이용하여 score계산

Merkel :  $h = (1, 0)^T$   
Germany :  $t = (1, 0)^T$

Is\_leader:  $W_r = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\text{Score} = h^T W_r t$$

$$= \sum (h \odot t) \odot \text{diag}(W_r)$$

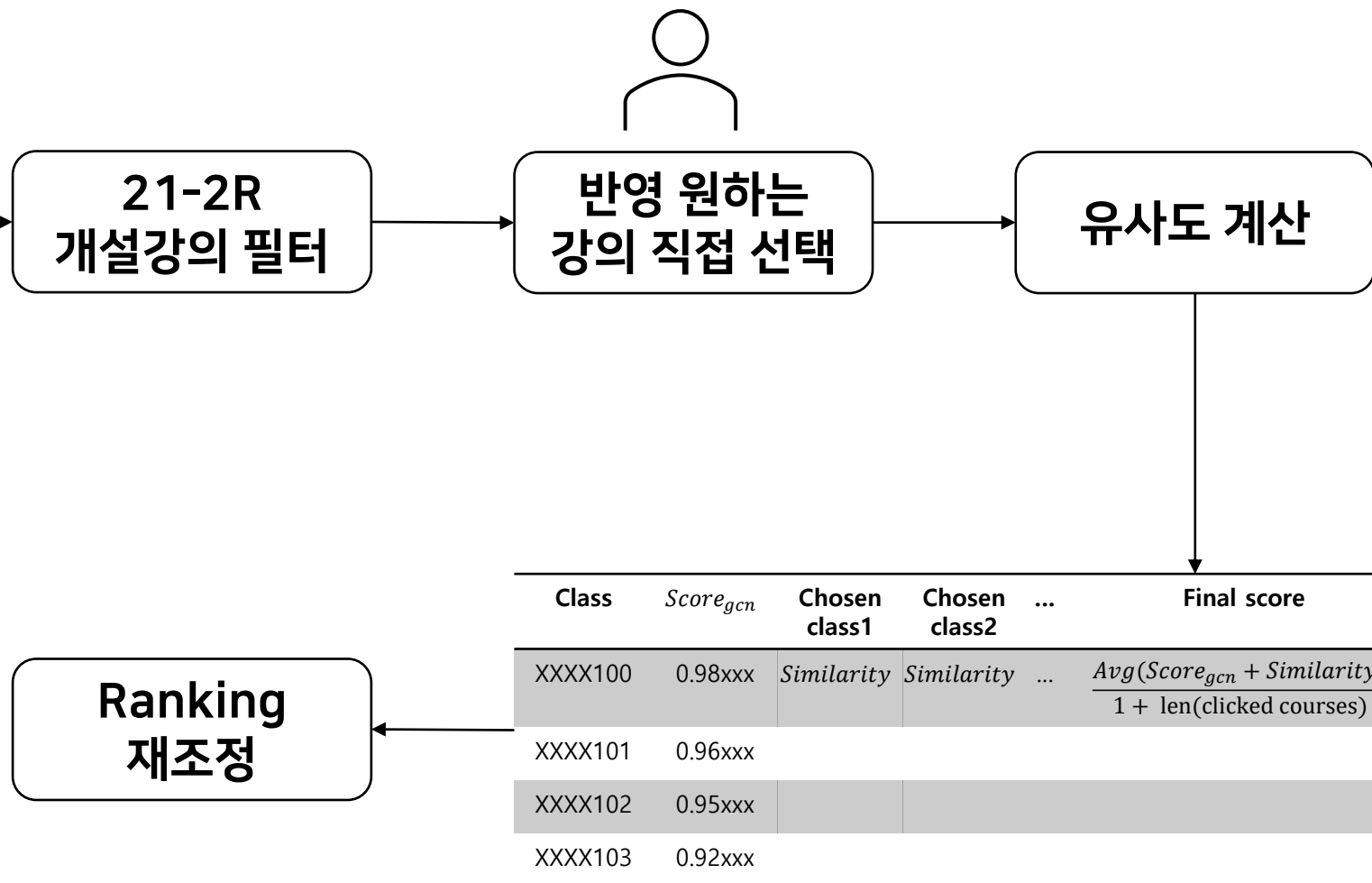
# 추천 Flow

Object perturbation score만 사용

## Recommendation Task Flow

Std_id	Ranking	Class	Score
2021000000	1	XXXX100	0.98xxx
2021000000	2	XXXX101	0.96xxx
2021000000	3	XXXX102	0.95xxx
2021000000	...	XXXX103	0.92xxx

RGCN이 임베딩해준 그래프로부터 산출된  
전공,교양, 교직 등 전체 강의 Score



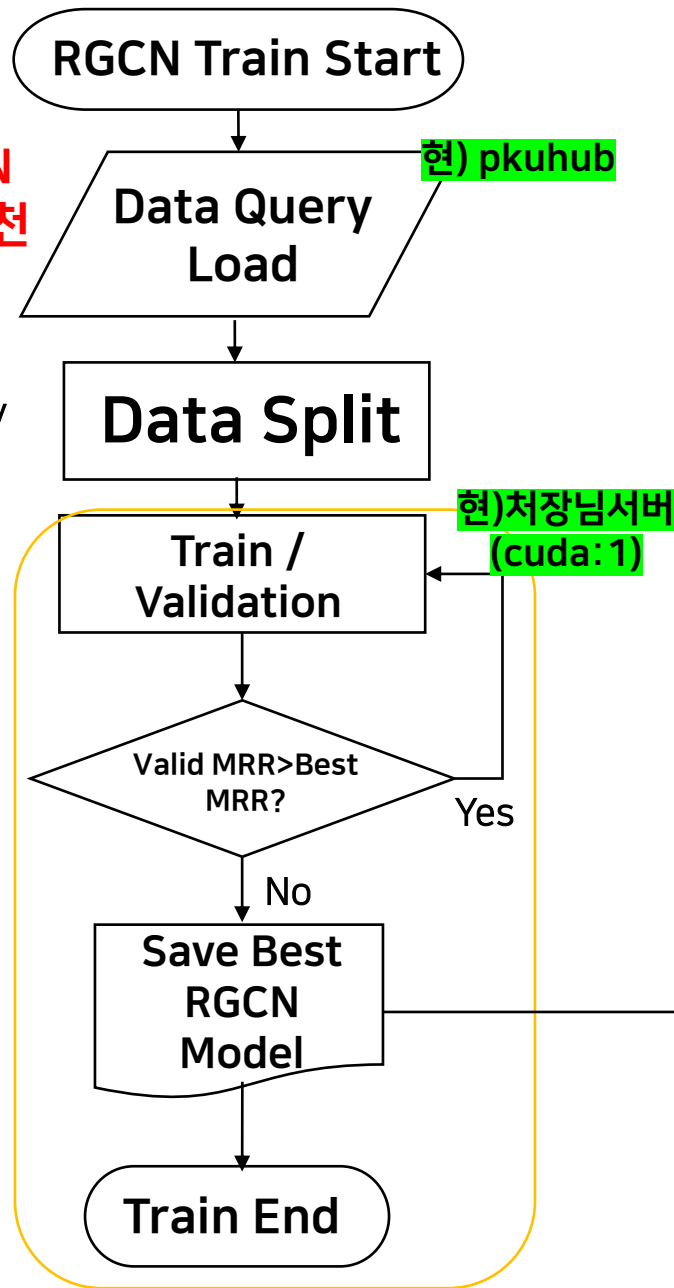


# 추천 Flow (Detailed)1

1

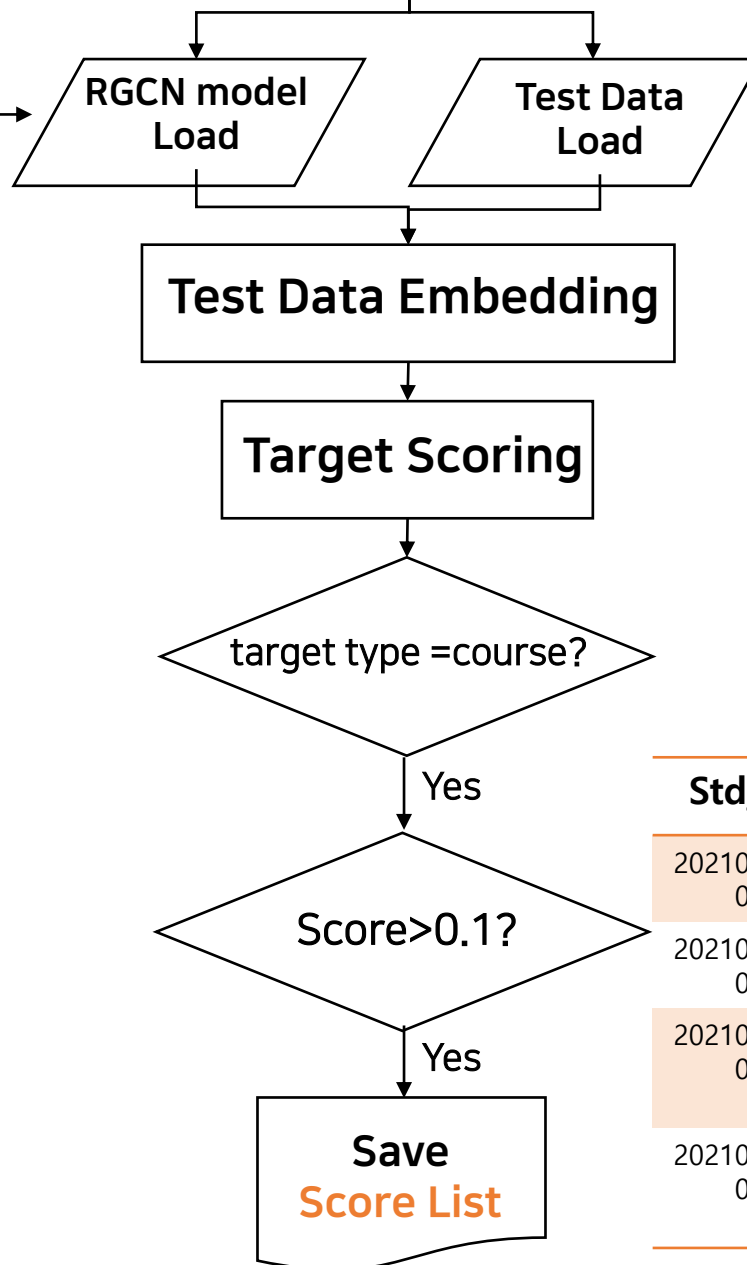
R-GCN  
강의 추천  
학습

python3 utils.py  
python3 model.py  
python3 main.py



RGCN Test Start

(현)처장님서버  
(cuda:1)



2

R-GCN 강의 추천  
결과 추출

python3 save\_result.py

Score List Example

Std_id	Rank	Class	Score
202100000 0	1	XXXX100	0.98xxx
202100000 0	2	XXXX101	0.96xxx
202100000 0	3	XXXX102	0.95xxx
202100000 0	...	XXXX103	0.92xxx

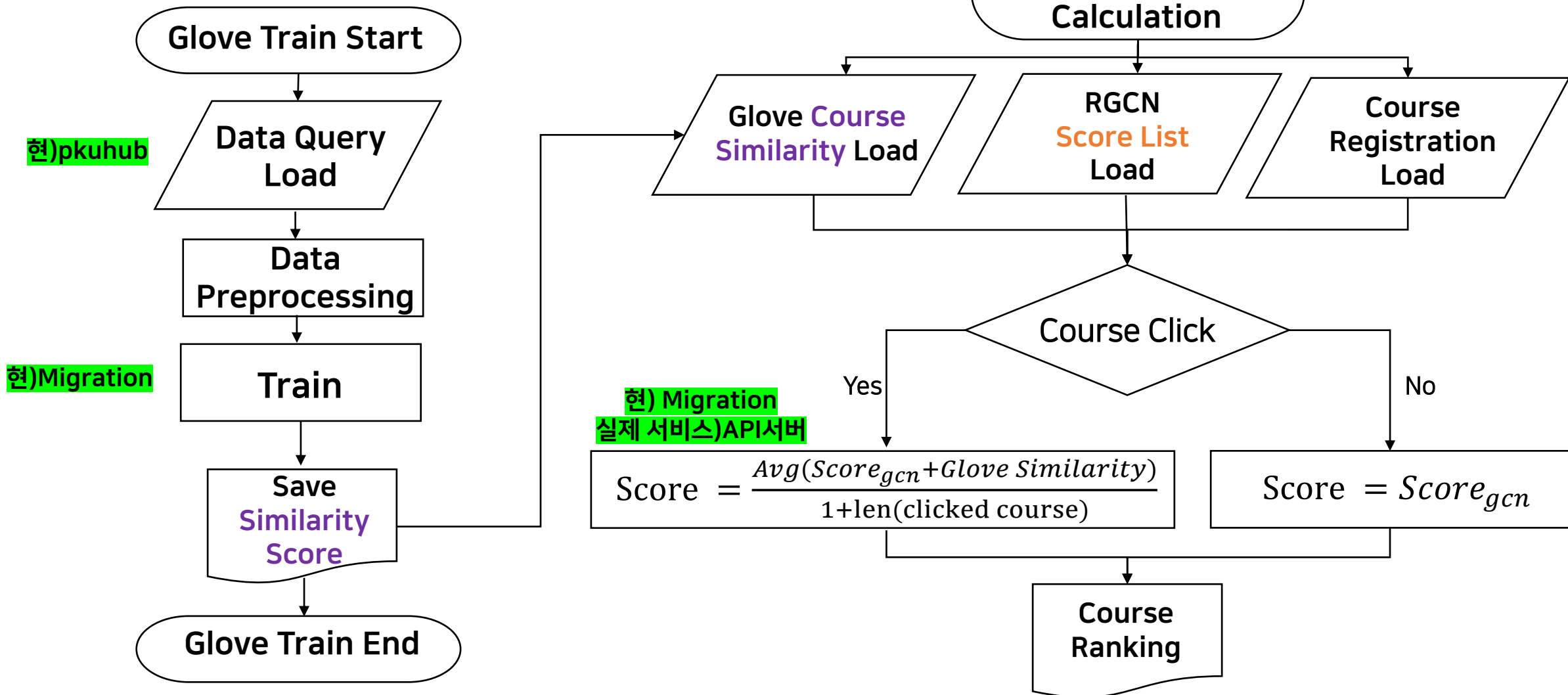
# 추천 Flow (Detailed) 2

3

내용기반강의유사도

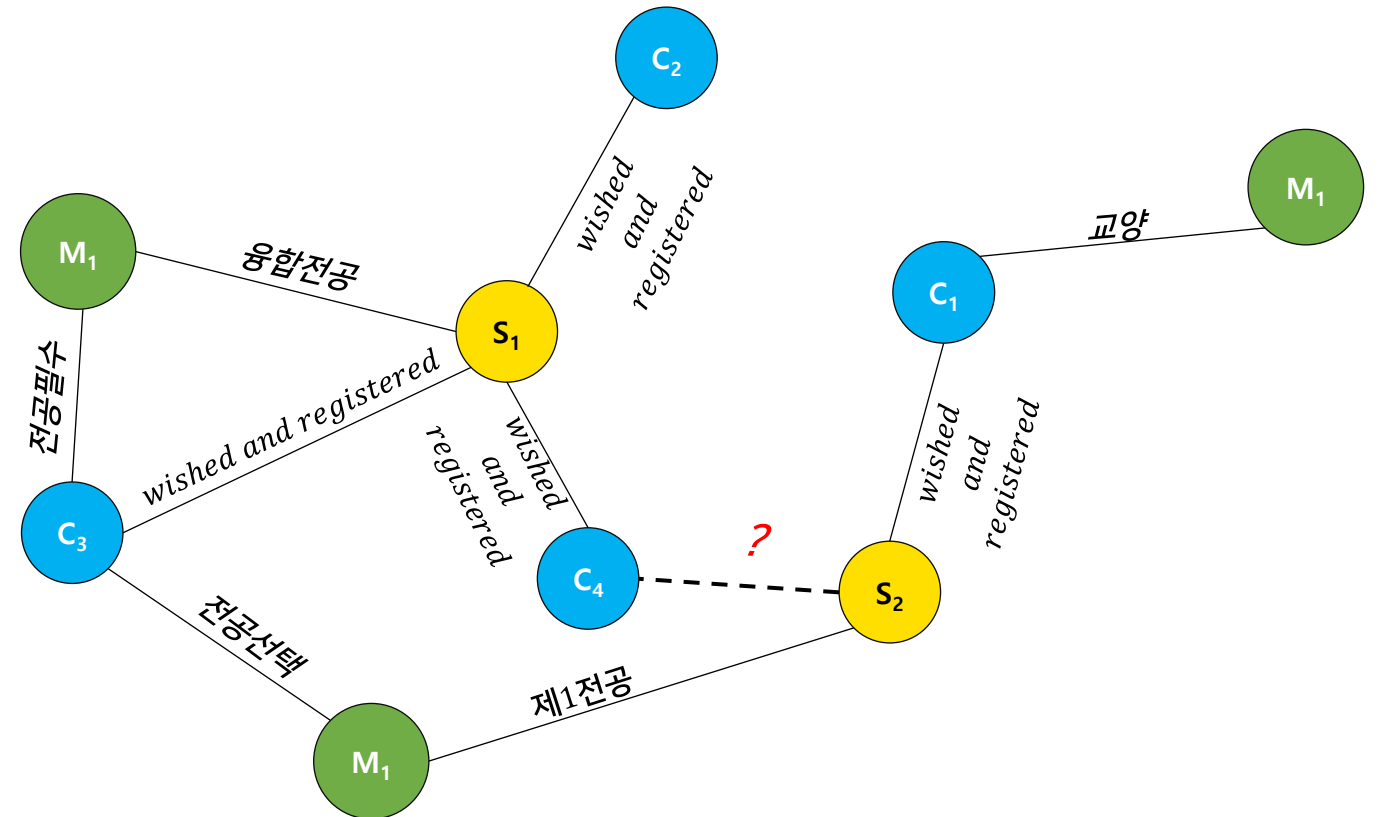
4

실제 서비스 시 동작 파일



# Chapter 3 : 학습&실험결과

- Parameter setting
- Train result
- Sample data test



# Experiment 1. Metric & Hyperparameter setting

Hit@n, Recall@n, MRR

**MRR(Mean Reciprocal Rank)**

$$MRR = \frac{1}{|S|} \sum_{i=1}^{|S|} \frac{1}{rank_i}$$

**Recall**

$$Recall = \frac{1}{|S|} \sum_{i=1}^{|S|} \frac{|courses_i \cap recommended_i|}{|courses_i|}$$

**Hit@1,3,10,21**

N개 안에 포함

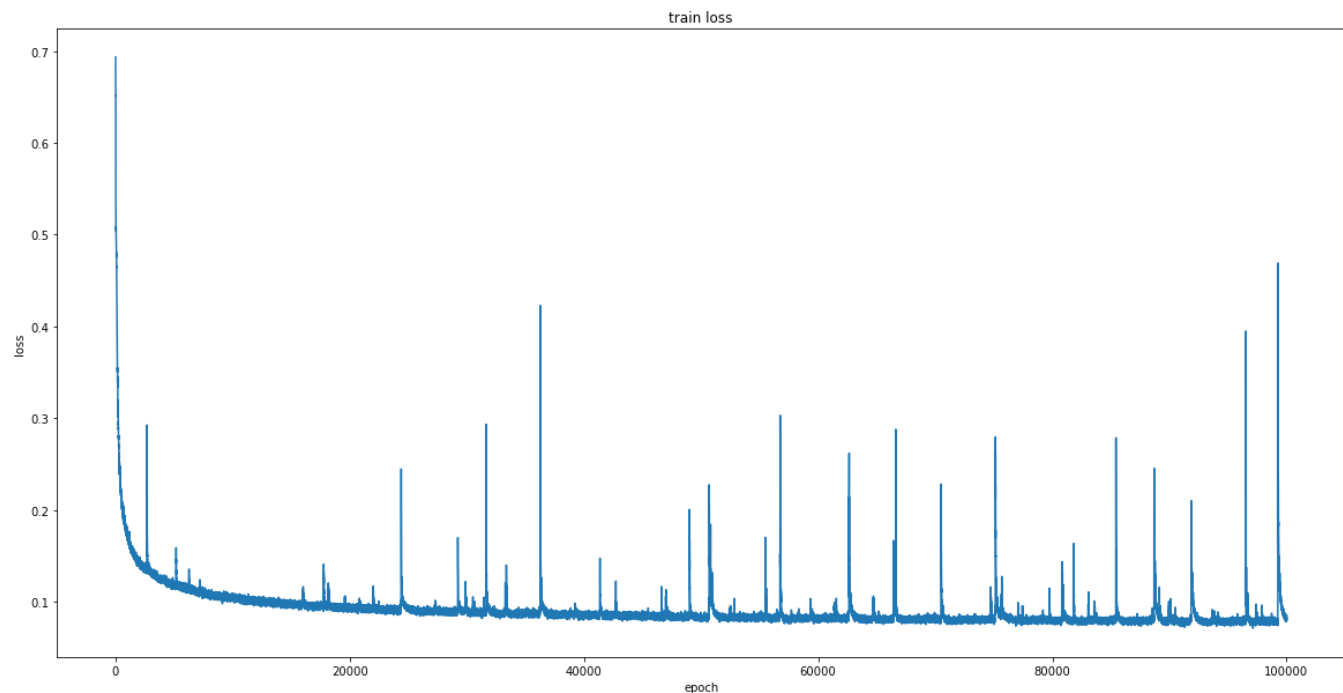
Epoch	Dropout	LR	n-bases	Negative Sample	Composition
~100000	0.2	0.01	4	o	Basic

**Evaluate every 1000**

Loss 대신에 Valid MRR로 Best epoch 설정

# Experiment 1. Result

Hit@n, Recall@n, MRR



num\_entity: 41952  
num\_relation: 24  
num\_train\_triples: 748751  
num\_valid\_triples: 8956  
num\_test\_triples: 27625

학우기반 모델  
**Hit@21(0.43)보다  
0.03 상승!**

## Best Model Description(Evaluated every 1000 epoch)

Epoch	Dropout	LR	n-bases	Train Loss	Valid MRR	Test MRR	Hit@1	Hit@3	Hit@10	Hit@21
95000	0.2	1e-2	4	0.6935	0.095	0.151	0.071	0.145	0.314	0.474

# Test 1: 샘플 학생 수강이력

제1전공 바이오 의공학부, 심화전공, 15학번, 수강할 교양 1과목 남음

1	2015	1R	CHEM151	일반화학및연습 I
1	2015	1R	CHEM153	일반화학실험 I
1	2015	1R	EGRN151	컴퓨터언어및실습
1	2015	1R	GEKS001	1학년 세미나
1	2015	1R	GETE001	사고와표현 I
1	2015	1R	IFLS003	ACADEMIC ENGLISH I(영강)
1	2015	1R	MATH161	미적분학및연습I
1	2015	1R	PHYS151	일반물리학및연습 I (영강)
1	2015	1R	PHYS161	일반물리학실험 I

1	2015	2R	BMED216	해부학(영강)
1	2015	2R	CHEM152	일반화학및연습II
1	2015	2R	CHEM154	일반화학실험II
1	2015	2R	GEKS002	1학년 세미나
1	2015	2R	GETE004	사고와표현II
1	2015	2R	IFLS004	ACADEMIC ENGLISH II(영강)
1	2015	2R	MATH162	미적분학및연습II
1	2015	2R	PHYS152	일반물리학및연습II(영강)
1	2015	2R	PHYS162	일반물리학실험II

1	2016	1R	BMED309	의료영상(영강)
1	2016	1R	CNCE150	정보과학개론
1	2016	1R	GESO121	마음의과학(영강)
1	2016	1R	GEST036	생활원예
1	2016	1R	IFLS125	교양중국어초급(외국어강의)
1	2016	1R	IFLS256	시사일본어(외국어강의)
1	2016	1R	SPGE112	기초음악이론
2	2016	2R	BMED208	생체유체역학
2	2016	2R	BMED218	생리학
2	2016	2R	BMED222	전자기학(영강)
2	2016	2R	BMED226	의공학프로그래밍(영강)
2	2016	2R	GEQR039	데이터로표현하는세상
2	2016	2R	MATH161	미적분학및연습I

2	2020	1R	BMED205	공학수학I(영강)
2	2020	1R	BMED217	회로이론(영강)
2	2020	1R	BMED325	생체데이터과학
2	2020	1R	ECON110	경제학개론
2	2020	1R	GELA077	어떻게영화를읽을것인가
2	2020	1R	IFLS282	학술일본어 I (외국어강의)
3	2020	2R	BMED202	공학수학II(영강)
3	2020	2R	BMED218	생리학
3	2020	2R	BMED312	생체정보계측(영강)
3	2020	2R	BMED323	의공학실험 I (영강)
3	2020	2R	BMED326	의공학실험II(영강)
3	2020	2R	BMED332	생체정보학
3	2020	2R	BMED334	의학물리

# Test 1. 교양 추천 - 21년도 2학기 개설강의 내 Top 30

21학년도 개설 강의 필터링 결과, 반영 원하는 강의 선택 전 결과(score >0.1)

Std_id	cour_cd	cour_nm	score
201525****	GESO119	현대사회와고등교육	0.910874
201525****	KORE155	문학과여성	0.890112
201525****	GEFC091	종교와영화	0.802089
201525****	SPGE108	판소리의이해	0.800478
201525****	KORE151	한국의언어와문학	0.784636
201525****	GEHI087	대한민국헌법사	0.77648
201525****	GEFC099	일본의문화콘텐츠와전통문화	0.747477
201525****	GERM143	독일도시와문학	0.73426
201525****	FRAN133	프랑스문화탐색	0.724651
201525****	JAPN113	교양일본어 I (외국어강의)	0.684292
201525****	GEOG101	도시와국토	0.678964
201525****	JURA150	법학통론	0.638375
201525****	SPGE209	한국음악의세계	0.63623
201525****	PSYC110	심리학의이해	0.443417
201525****	ENGL122	시사영어(영강)	0.426259
201525****	CHEM156	화학수학(영강)	0.424672
201525****	HOEW125	한일교류의역사	0.374022
201525****	GELA138	사진과탈근대사회	0.327954
201525****	GERM142	독일예술과신화	0.291714
201525****	GERM136	독일의대안문화	0.282893
201525****	HOEW137	동아시아사의재조명	0.276103
201525****	HANM110	동양의지혜	0.244505
201525****	KORE159	남북의언어비교	0.243228
201525****	HEED102	부모되기교육	0.231618
201525****	IFLS111	한국어중급Ⅱ	0.203666
201525****	PHEK103	수영	0.17821
201525****	LIBS150	생명과학	0.166742
201525****	GEST146	움직임과건강	0.160997
201525****	PHIL149	비판적사고와논리	0.160359
201525****	GELA140	동서양의사랑시	0.151173



# Test 1. 교양 추천 - 커리큘럼 내 Top 30

2021학년도 커리큘럼 필터링 결과, 반영 원하는 강의 선택 전 결과(score >0.1)

std_id	cour_cd	cour_nm	score	cour_div	year
201525****	KHPM105	보건정책학	0.949	교양	2021
201525****	GEQR050	E=MC제곱	0.929	교양	2021
201525****	KORE129	한국민요와대중가요	0.911	교양	2021
201525****	GESO119	현대사회와고등교육	0.911	교양	2021
201525****	KORE117	비교문학의이해	0.910	교양	2021
201525****	KHPM101	사회와건강	0.899	교양	2021
201525****	KORE155	문학과여성	0.890	교양	2021
201525****	GESO143	중국의대외정책과동아시아국제관계	0.890	교양	2021
201525****	GEST055	21C환경문제이해	0.843	교양	2021
201525****	GEST132	과학기술과창업가정신	0.833	교양	2021
201525****	ENGL101	영미단편강독	0.830	교양	2021
201525****	HANM133	교양한문 I	0.826	교양	2021
201525****	GEFC091	종교와영화	0.802	교양	2021
201525****	SPGE108	판소리의이해	0.800	교양	2021
201525****	KORE151	한국의언어와문학	0.785	교양	2021
201525****	GEHI087	대한민국헌법사	0.776	교양	2021
201525****	GEFC099	일본의문화콘텐츠와전통문화	0.747	교양	2021
201525****	GERM143	독일도시와문학	0.734	교양	2021
201525****	FRAN133	프랑스문화탐색	0.725	교양	2021
201525****	GELA076	영어로시읽기	0.722	교양	2021
201525****	GEFC016	그림을통해본일본고전문학과문화	0.717	교양	2021
201525****	GEST141	기계학습의이해와응용	0.707	교양	2021
201525****	JAPN113	교양일본어 I	0.684	교양	2021
201525****	JAPN113	교양일본어 I (외국어강의)	0.684	교양	2021
201525****	GEST131	인체질환생물학	0.683	교양	2021
201525****	GEOG101	도시와국토	0.679	교양	2021
201525****	GEST090	미생물과건강	0.657	교양	2021
201525****	JURA150	법학통론	0.638	교양	2021
201525****	SPGE209	한국음악의세계	0.636	교양	2021
201525****	POLI104	현대정치와이데올로기	0.587	교양	2021

# Test 1. 전공 추천 - 21년도 2학기 개설강의 내 Top 30

21학년도 2학기 개설 강의 필터링 결과, 반영 원하는 강의 선택 전 결과(score >0.1)

std_id	cour_cd	cour_nm	Cour_div	score
201525****	KHES238	해부생리학및실습(영강)	10	0.978849
201525****	KHPM408	보건의료전략경영론(영강)	10	0.97745
201525****	KHES422	분자세포생물학	10	0.977383
201525****	BSMS424	휴먼마이크로바이옴(영강)	10	0.971937
201525****	BMED326	의공학실험II(영강)	11	0.965896
201525****	JMCO379	미디어산업과융합	10	0.962444
201525****	JMCO272	방송론(영강)	10	0.962432
201525****	JMCO382	엔터테인먼트산업	10	0.961589
201525****	JMCO458	미디어제작워크샵III	10	0.960391
201525****	BMED224	물리화학	10	0.95446
201525****	DISS123	정치학입문(영강)	11	0.952659
201525****	JMCO276	미디어대본작성	10	0.951956
201525****	BSMS214	후생유전학	10	0.940497
201525****	JMCO259	다문화사회와미디어(영강)	10	0.935313
201525****	JMCO452	미디어정책	10	0.929104
201525****	LIBS400	생명과학최신동향세미나	10	0.91777
201525****	BUSS252	마케팅조사론	10	0.912705
201525****	BUSS252	마케팅조사론(영강)	10	0.912705
201525****	BUSS325	Korean Business & Management in the Global Context(영강)	10	0.906686
201525****	HOEW406	한중관계사	10	0.893815
201525****	ECON360	지역도시경제론	10	0.893305
201525****	LIBS457	생물정보학(영강)	10	0.869684
201525****	IMEN302	공급사슬경영(영강)	10	0.862066
201525****	LIBS483	나노생명과학의이해	10	0.855299
201525****	MATE354	미분기하II	10	0.854825
201525****	ECON439	한국거시경제분석	10	0.846445
201525****	BMED212	기초광학(영강)	10	0.84077
201525****	SOCI295	사회통계	11	0.827588
201525****	SOCI295	사회통계(영강)	11	0.827588
201525****	LING471	응용기호학	10	0.81882

# Test 1. 전공 추천 - 커리큘럼 내 Top 30

2021학년도 커리큘럼 필터링 결과, 반영 원하는 강의 선택 전 결과(score >0.1)

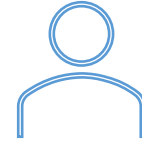
std_id	cour_cd	cour_nm	score	cour_div	year
201525****	CHEM207	분석화학I	0.978895	전공필수	2021
201525****	KHES238	해부생리학및실습	0.978849	전공선택	2021
201525****	KHPM408	보건의료전략경영론	0.97745	전공선택	2021
201525****	KHES422	분자세포생물학	0.977383	전공선택	2021
201525****	BSMS424	휴먼마이크로바이옴	0.971937	전공선택	2021
201525****	JMCO279	애널리틱스프로그래밍	0.969867	전공선택	2021
201525****	BMED326	의공학실험II	0.965896	전공필수	2021
201525****	CHEM231	물리화학	0.965883	전공선택	2021
201525****	LIBT485	임목생명공학	0.964223	전공선택	2021
201525****	KHES232	원자력안전규제	0.964222	전공선택	2021
201525****	KHES225	진단화학및실험	0.963972	전공선택	2021
201525****	JMCO379	미디어산업과융합	0.962444	전공선택	2021
201525****	JMCO272	방송론	0.962432	전공선택	2021
201525****	JMCO382	엔터테인먼트산업	0.961589	전공선택	2021
201525****	JMCO458	미디어제작워크샵III	0.960391	전공선택	2021
201525****	KHES215	환경보건학	0.95918	전공선택	2021
201525****	JMCO374	디지털미디어와민주주의프로젝트	0.958088	전공선택	2021
201525****	KHES211	보건환경독성학	0.957176	전공선택	2021
201525****	LIFS331	유가공학	0.956423	전공선택	2021
201525****	BMED325	생체데이터과학	0.955845	전공선택	2021
201525****	BSMS404	분자질병역학	0.955756	전공선택	2021
201525****	BMED224	물리화학	0.95446	전공선택	2021
201525****	KHES311	분석화학및실험	0.95289	전공선택	2021
201525****	DISS123	정치학입문	0.952659	전공필수	2021
201525****	BSMS203	의생명과학실험I	0.952329	전공선택	2021
201525****	JMCO276	미디어대본작성	0.951956	전공선택	2021
201525****	LIBT391	분자바이러스생명공학	0.950887	전공선택	2021
201525****	KHPM412	지역사회건강증진실습	0.948976	전공선택	2021
201525****	ARCH441	건축설계실무	0.944898	전공필수	2021
201525****	BMED422	바이오인공장기	0.941649	전공선택	2021

# Test 1. 사용자가 원하는 과목 반영 결과

사용자가 원하는 과목 선택 전후 비교

## 과목 선택 전 추천리스트 (RGCN only)

cour_cd	cour_nm	score
GESO119	현대사회와고등교육	0.910874
KORE155	문학과여성	0.890112
GEFC091	종교와영화	0.802089
SPGE108	판소리의이해	0.800478
KORE151	한국의언어와문학	0.784636
GEHI087	대한민국헌법사	0.77648
GEFC099	일본의문화콘텐츠와전통문화	0.747477
GERM143	독일도시와문학	0.73426
FRAN133	프랑스문화탐색	0.724651
JAPN113	교양일본어 I (외국어강의)	0.684292
GEOG101	도시와국토	0.678964
JURA150	법학통론	0.638375
SPGE209	한국음악의세계	0.63623
PSYC110	심리학의이해	0.443417
ENGL122	시사영어(영강)	0.426259
CHEM156	화학수학(영강)	0.424672
HOEW125	한일교류의역사	0.374022
GELA138	사진과탈근대사회	0.327954
GERM142	독일예술과신화	0.291714
GERM136	독일의대안문화	0.282893
HOEW137	동아시아사의재조명	0.276103
HANM110	동양의지혜	0.244505
KORE159	남북의언어비교	0.243228
HEED102	부모되기교육	0.231618
IFLS111	한국어중급II	0.203666
PHEK103	수영	0.17821
LIBS150	생명과학	0.166742
GEST146	움직임과건강	0.160997
PHIL149	비판적사고와논리	0.160359
GELA140	동서양의사랑시	0.151173



AI 선배!

“GESO119 마음의과학(영강)”,  
“SPGE112 기초음악이론”,  
“GEQR039 데이터로표현하는세계”,  
“BMED325 생체데이터과학”,  
“BMED326 의공학실험2(영강)”,  
“BMED332 생체정보학”,  
“BMED334 의학물리”  
과목 반영해주세요

## 과목 선택 후 추천리스트 (RGCN + Glove)

3

내용기반강의유사도

4

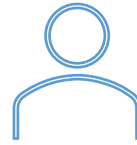
실제 서비스 시 동작 파일

# Test 1. 사용자가 원하는 과목 반영 결과

사용자가 원하는 과목 선택 전후 비교 : 남은 교양 1강의는 아마도 '화학수학(영강)'?

## 과목 선택 전 추천리스트 (RGCN only)

cour_cd	cour_nm	score
GESO119	현대사회와고등교육	0.910874
KORE155	문학과여성	0.890112
GEFC091	종교와영화	0.802089
SPGE108	판소리의이해	0.800478
KORE151	한국의언어와문학	0.784636
GEHI087	대한민국헌법사	0.77648
GEFC099	일본의문화콘텐츠와전통문화	0.747477
GERM143	독일도시와문학	0.73426
FRAN133	프랑스문화탐색	0.724651
JAPN113	교양일본어 I (외국어강의)	0.684292
GEOG101	도시와국토	0.678964
JURA150	법학통론	0.638375
SPGE209	한국음악의세계	0.63623
PSYC110	심리학의이해	0.443417
ENGL122	시사영어(영강)	0.426259
CHEM156	화학수학(영강)	0.424672
HOEW125	한일교류의역사	0.374022
GELA138	사진과탈근대사회	0.327954
GERM142	독일예술과신화	0.291714
GERM136	독일의대안문화	0.282893
HOEW137	동아시아사의재조명	0.276103
HANM110	동양의지혜	0.244505
KORE159	남북의언어비교	0.243228
HEED102	부모되기교육	0.231618
IFLS111	한국어중급II	0.203666
PHEK103	수영	0.17821
LIBS150	생명과과학	0.166742
GEST146	움직임과건강	0.160997
PHIL149	비판적사고와논리	0.160359
GELA140	동서양의사랑시	0.151173



AI 선배!

“GESO121 마음의과학(영강)”,  
 “SPGE112 기초음악이론”,  
 “GEQR039 데이터로표현하는세상”,  
 “BMED325 생체데이터과학”,  
 “BMED326 의공학실험2(영강)”,  
 “BMED332 생체정보학”,  
 “BMED334 의학물리”  
 과목 반영해주세요

## 과목 선택 후 추천리스트 (RGCN + Glove)

cour_cd	cour_nm	final_score
CHEM156	화학수학(영강)	0.547005
GESO119	현대사회와고등교육	0.468287
JURA150	법학통론	0.46542
PSYC110	심리학의이해	0.463409
IFLS307	학술영작문-자연계	0.458688
LIBS150	생명과과학	0.456174
ENGL122	시사영어(영강)	0.419967
JAPN113	교양일본어 I (외국어강의)	0.355499
GEHI087	대한민국헌법사	0.355423
KORE151	한국의언어와문학	0.351955
PHIL149	비판적사고와논리	0.325429
KORE155	문학과여성	0.321151
GEFC099	일본의문화콘텐츠와전통문화	0.300696
FRAN133	프랑스문화탐색	0.281333
SPGE108	판소리의이해	0.28091
GEOG101	도시와국토	0.265615
SOC116	성과사회	0.261987
KORE159	남북의언어비교	0.257
SPGE209	한국음악의세계	0.25684
GERM142	독일예술과신화	0.229467
GERM143	독일도시와문학	0.216234
SOC124	여가와사회	0.210402
GEFC091	종교와영화	0.197427
HOEW137	동아시아사의재조명	0.197183
IFLS111	한국어중급II	0.186112
HANM110	동양의지혜	0.175723



# Test 2. 같은 1전공 다른 2전공 추천교양& 전공 비교

Std A:심리학부 + 정치외교 vs Std B: 심리학부 +생명과학부

## Std A 추천 리스트 (교양)

HOKA155	한국의신분과사회상	0.984608
HOEW139	중국근현대개혁과혁명	0.983874
PHEK126	피트니스앤헬스	0.933635
IFLS308	미디어영어	0.862721
HANM120	생활한자	0.860902
HOEW116	서양기독교사	0.856938
RUSS103	러시아문화의이해	0.847569
PHIL145	불교와현대사회	0.841251
PHEK128	탁구	0.816778
SPAN104	환상문학의세계	0.788622
HOEW138	중국의역사와인물	0.76059
IFLS111	한국어중급II	0.708618
PHIL122	문화의철학적이해	0.699253
COLA402	데이터인문학	0.688647
PHIL120	그리스신화와비극	0.401638
HOKA103	한국전통문화의이해	0.342154
KORE159	남북의언어비교	0.338153
KORE151	한국의언어와문학	0.317408
KORE140	학업한국어작문	0.272574
GELA138	사진과탈근대사회	0.265613
HOEW118	서양근현대사의이해	0.237405
IFLS242	한국어중급말하기	0.235016
HOEW131	서양문화의전통과유산	0.154446
PHEK120	배드민턴	0.153938
GEOG101	도시와국토	0.147955
HOKA102	서울지역의향토문화	0.125531
JURA150	법학통론	0.114065
GEHI087	대한민국헌법사	0.111442
PHIL148	동양철학입문	0.109564
GELA107	한국고전문학과배경사상	0.108411

## Std A 추천 리스트 (전공)

ECON372	경제학과철학적논점들	0.988524
PSYC201	심리학의기초I	0.987947
PAPP362	위험사회와정책	0.987176
JAPN348	일본한문강독	0.986422
POLI378	내전연구(영강)	0.983127
STAT302	시계열분석(영강)	0.982419
KLLE204	현대소설선독	0.981981
POLI220	시민정치론(영강)	0.978334
BUSS435	경영의철학적이해	0.978333
LESE312	자원가치평가론(영강)	0.977356
ARDE164	디자인사	0.976264
STAT243	사회과학을위한데이터과학	0.973812
HOEW412	오리엔트문명사	0.973747
LESE406	국제통상협상론(영강)	0.971278
PHIL020	철학토픽코스 I	0.969047
MATH240	미분방정식및연습	0.967872
MATH240	미분방정식및연습(영강)	0.967872
LESE154	세계화와국가경제(영강)	0.965614
HOKA214	고려후기사	0.95724
STAT204	통계계산프로그래밍	0.956244
PSYC232	임상심리학(영강)	0.946571
ARCH406	건축의장(영강)	0.943267
ECON466	경제정책과프로그램효과와실증분석	0.936106
PSYC331	이상심리학(영강)	0.93402
POLI247	정치제도론(영강)	0.930004
JMCO264	대인커뮤니케이션(영강)	0.916418
RUSS208	중급러시아어회화II(외국어강의)	0.887809
PAPP350	도시계획론	0.883691
LIBT368	식물병제어학	0.875136
NRSG172	건강과국제화	0.854713
POLI425	미국정치론(영강)	0.847158

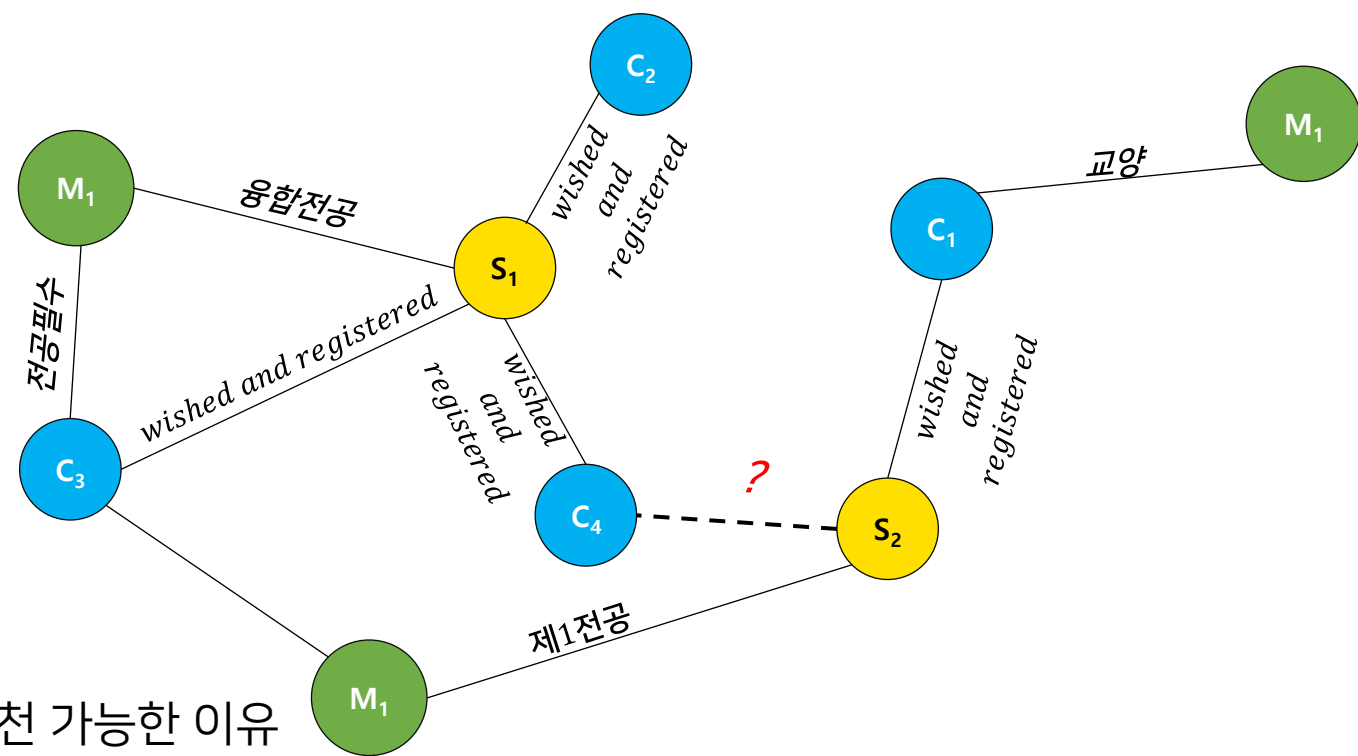
## Std B 추천 리스트 (교양)

JAPN281	영상일본어(외국어강의)	0.952858
HOKA155	한국의신분과사회상	0.924847
KORE155	문학과여성	0.924573
KORE142	근대매체와한국문학	0.915722
HANM120	생활한자	0.8593
LIET152	화학의기초및연습II	0.78654
LIET152	화학의기초및연습II(영강)	0.78654
HOKA115	한국역사의국제적환경	0.705942
GEFC101	언어와문명:세계의언어문명사	0.520037
LIBS154	일반생물학실험II	0.518541
GEST011	뉴튼의시계,하이젠버그의주사위	0.50676
LIET156	화학의기초실험II	0.502494
LIET156	화학의기초실험II(영강)	0.502494
PHEK107	골프	0.438044
HOEW140	새로보는서양의역사	0.393571
PHEK128	탁구	0.382575
GELA142	조선시대문학관의이해	0.291242
GEOG151	지리학입문	0.220453
JAPN253	일본어와일본문화(외국어강의)	0.188521
LIBS152	일반생물학및연습II	0.178454
LIBS152	일반생물학및연습II(영강)	0.178454
SPAN104	환상문학의세계	0.14167
GEST114	건축학과미래사회	0.121573

## Std B 추천 리스트 (전공)

PSYC232	임상심리학(영강)	0.979364
ARDE323	인터페이스디자인	0.977978
HOKA302	한국사적해제	0.968698
KHPM202	역학개론	0.96769
GLKS201	서울의이해	0.953666
PSYC493	사용자경험과심리학(영강)	0.950747
PSYC201	심리학의기초I	0.948371
LESE308	상품선물및옵션거래	0.940641
JAPN348	일본한문강독	0.939053
AMSE329	나노소재(영강)	0.928156
SPAN303	스페인어어원및계통론	0.925007
LIBS466	생물안전성학	0.917438
LIBS450	진화생물학	0.885311
HOKA214	고려후기사	0.879912
PSYC331	이상심리학(영강)	0.869627
BUSS435	경영의철학적이해	0.868307
LIST303	유전학 I	0.830236
LIST202	세포생물학II	0.808756
LIST202	세포생물학II(영강)	0.808756
AMSE414	세라믹재료공정	0.79812
LIST306	생화학II	0.794818
LIST306	생화학II(영강)	0.794818
KORE235	한국어표현론	0.785278
BSMS322	식품학II	0.755606
LIBS316	식물분자생물학	0.74982
JAPN268	일본어회화II(외국어강의)	0.735092
HOKA216	조선후기사	0.717297
LIBS478	염증생물학	0.711338
CHEM457	나노화학(영강)	0.705894
LIBS442	시스템신경과학	0.702306
LIBS442	시스템신경과학(영강)	0.702306
PHYS362	양자역학II	0.686877
LESE302	환경경제학	0.678543
HOEW422	중국사특강II	0.674141
LIBS284	생명과학과법	0.664868
LIST304	유전학II	0.640087
LIBT442	임목분자육종학	0.622479
ISEC303	역공학기법	0.541081

# 결론 및 보완점



- RGCN이 추천해준 강의의 근거 관계 3가지
  - 1) 나의 학과, 제2전공
  - 2) 내가 관심 및 수강한 강의
  - 3) 나의 학과, 제2전공의 커리큘럼 → 전공 강의도 추천 가능한 이유
- 크게 RGCN모델에서 주는 리스트와 특정 강의가 반영된 강의 리스크로 두 단계로 나뉘었기 때문에 수강 여부에 상관없이 클릭 리스트에 반영할 수 있다면 추천 가능
- 교내 데이터 베이스에 아직 2021학년도 2학기 강의 확정이 안되었기 때문에 최종 결과는 바뀔 가능성 있음
- 국문 강의 텍스트 데이터가 없는 경우 Glove모델의 인풋데이터가 없기 때문에 사용자가 클릭한 강의여도 반영이 안됨 → 강의명도 추가하여 최대한 빈 강의가 없도록 조치
- 최적의 Glove 모델을 찾아야 함(2021.7.5 현재 n\_epochs 200보다 300이 안면 타당도가 높음 발견)

2021 고려대학교 대학혁신지원사업

# KU-Insight Miner AI선배 강의 추천

## Knowledge Graph를 활용한

## R-GCN기반 강의 추천 시스템 개발

고려대학교 디지털정보처 데이터 hub팀

데이터 사이언티스트 이진숙

2021.04.20~