Graph Attention Networks

ICLR 2018

1. Intro

GNN

+

Attention

=

GAT

Graph Attention Networks

1. Intro

1 Intro

2 GNN

\+

3 Attention

=

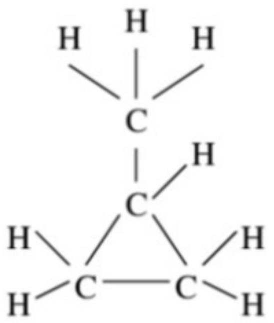
4 GAT

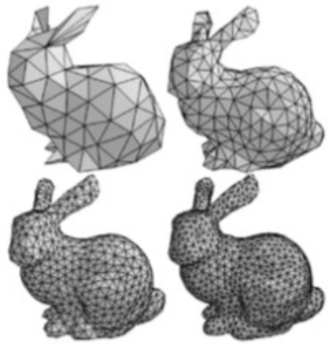
5 Evaluation

2. GNN

Graph Neural Network

그래프 데이터에서의 Neural Network

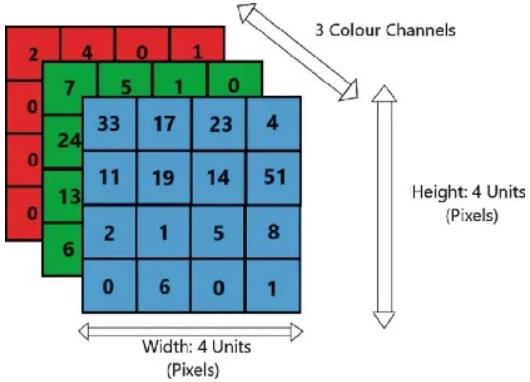




2. GNN

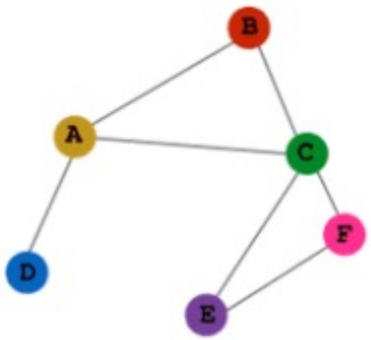
입력 데이터 표현 방법

Neural Network



행렬의 형태로 표현

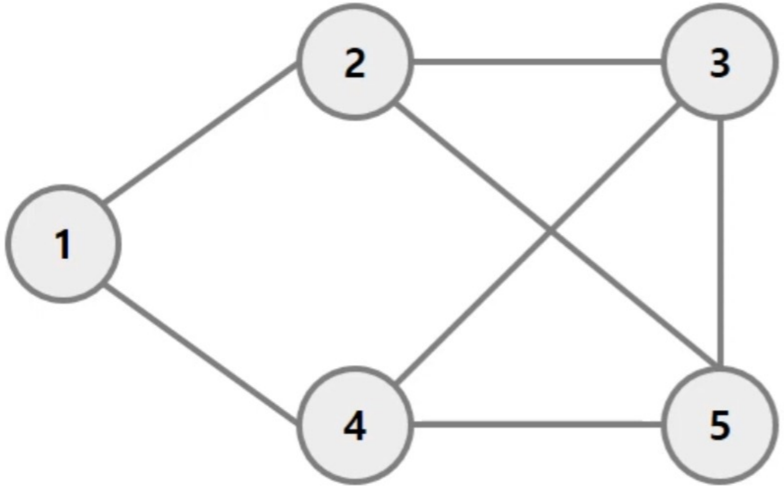
Graph Neural Network



?

2. GNN

입력 데이터 표현 방법 – Adjacency Matrix



[Adjacency Matrix]

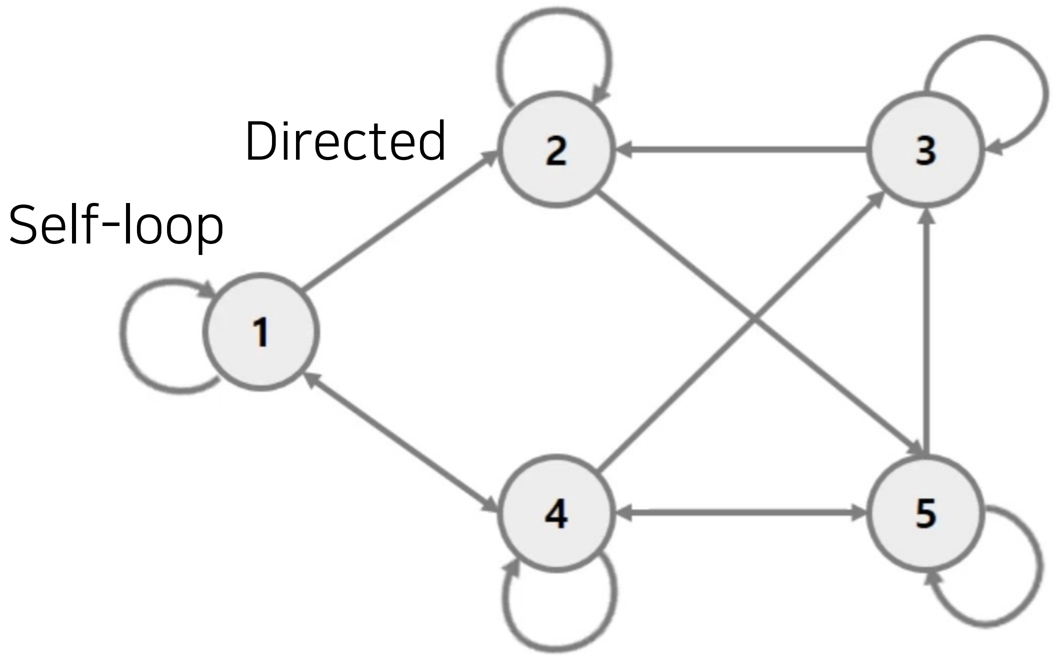
0	1	0	1	0
1	0	1	0	1
0	1	0	1	1
1	0	1	0	1
0	1	1	1	0

N

N

2. GNN

입력 데이터 표현 방법 – Adjacency Matrix



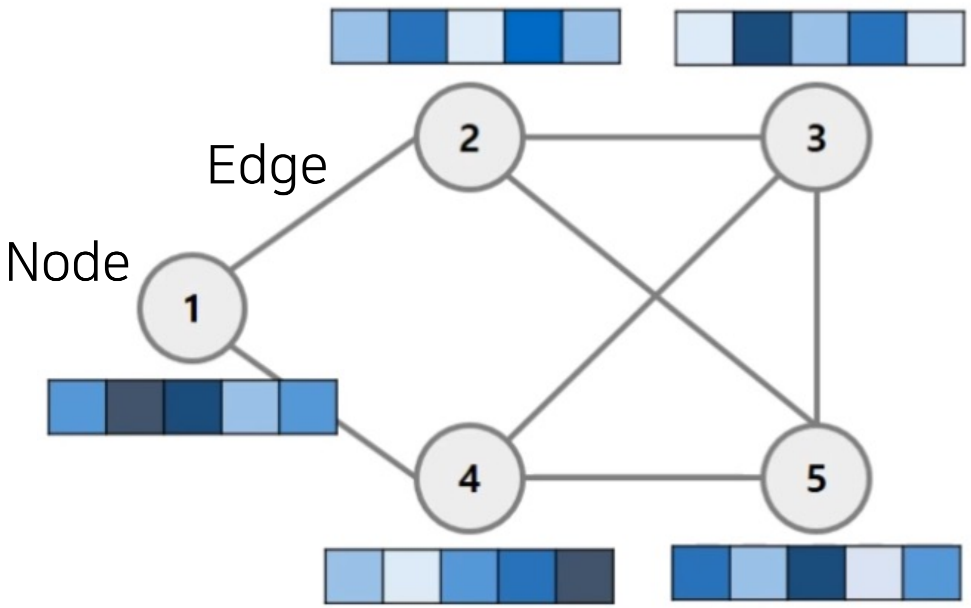
[Adjacency Matrix]

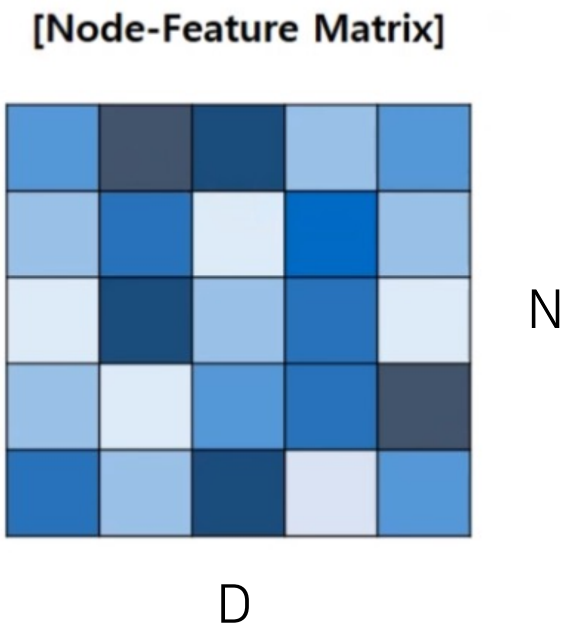
a_{11}	a_{12}	0	a_{14}	0
0	a_{22}	0	0	a_{25}
0	a_{32}	a_{33}	0	0
a_{41}	0	a_{43}	a_{44}	a_{45}
0	0	a_{53}	a_{54}	a_{55}

N

Weight

입력 데이터 표현 방법 – Node Feature

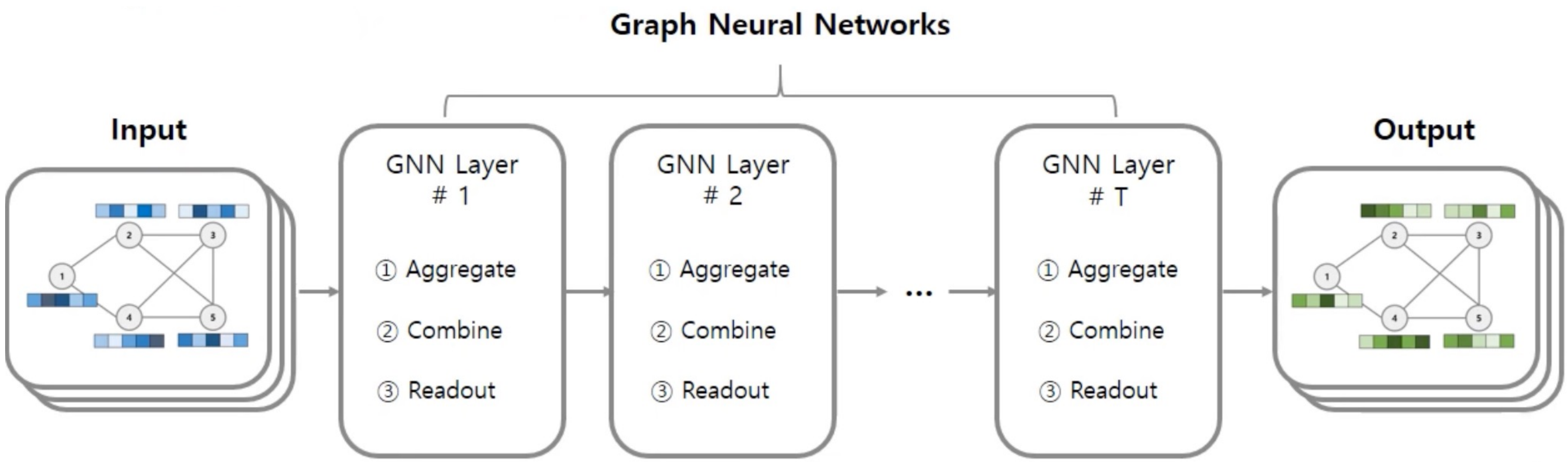




2. GNN

Node Level Task 업데이트 과정

1. Aggregate / Message Passing
2. Combine / Update
3. Readout



3. Attention

Dictionary

(일치하는지, 0,1)

Q: 오렌지

Key (K)	Similarity Sim(K, Q)	Value (V)	Sim(K, Q) * V
레몬	0	새콤한맛	0
오렌지	1	달콤한맛	달콤한맛
아보카도	0	크레파스맛	0

-> 달콤한맛!

3. Attention

Dictionary -> Attention

(유사한 정도의 Softmax)

Q: 귤

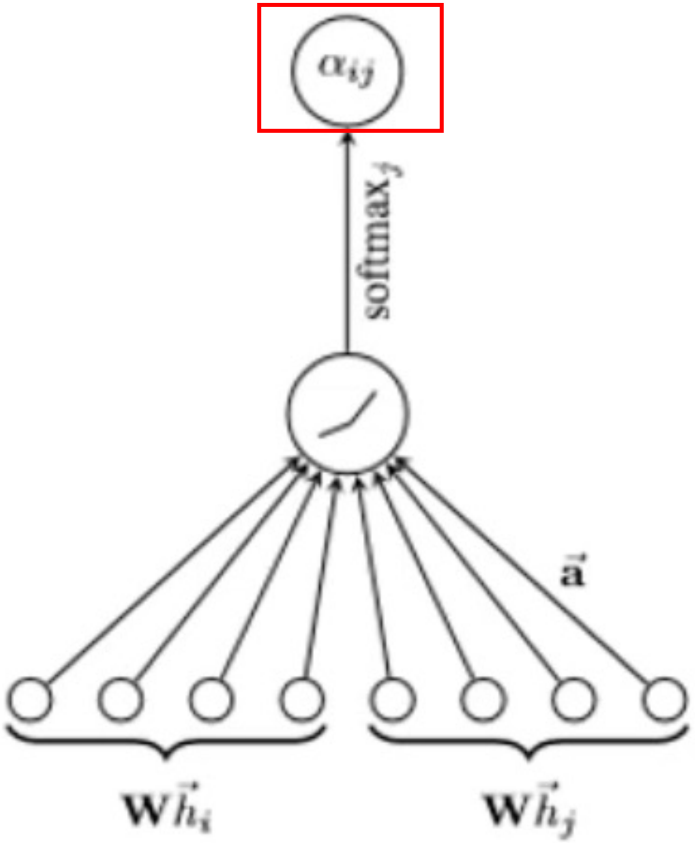
Key (K)	Similarity Sim(K, Q)	Value (V)	Sim(K, Q) * V
레몬	0.35	새콤한맛	0.35 * 새콤한맛
오렌지	0.64	달콤한맛	0.64 * 달콤한맛
아보카도	0.01	크레파스맛	0.01 * 크레파스맛

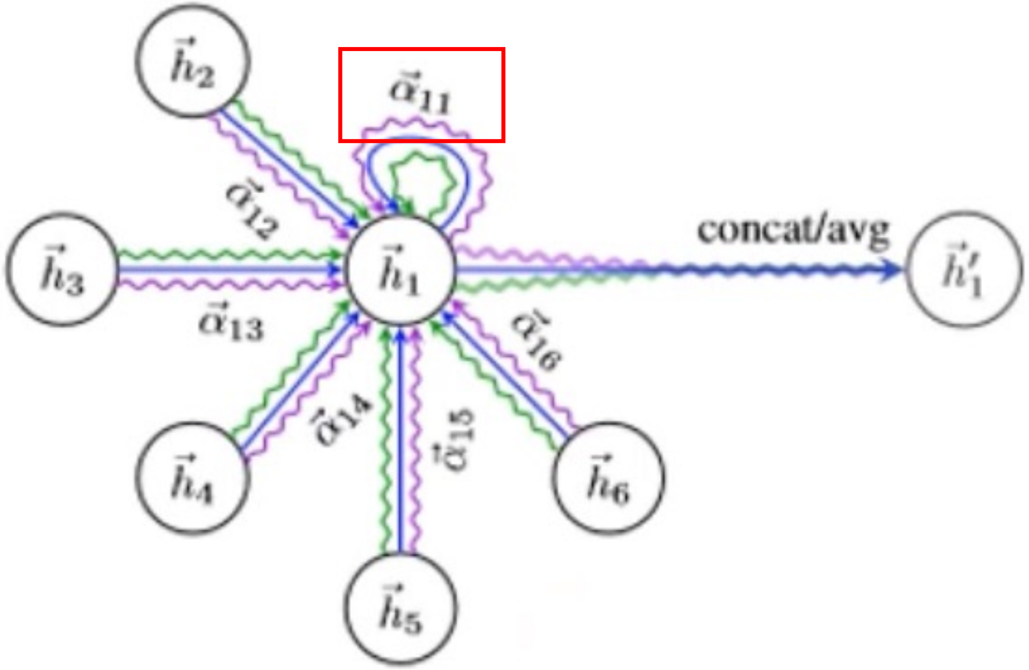
Attention Score

-> 새콤달콤한맛!

4. Graph Attention Network

Overview



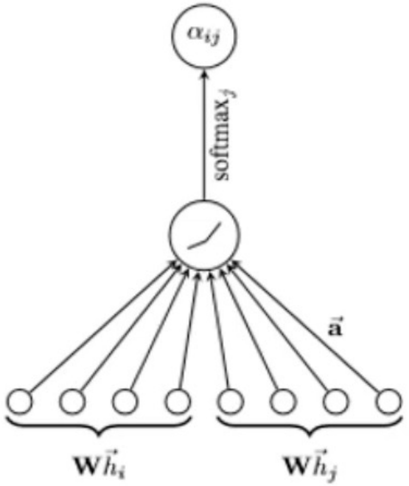


4. Graph Attention Network

Overview

모든 I, J쌍의 Sim 계산

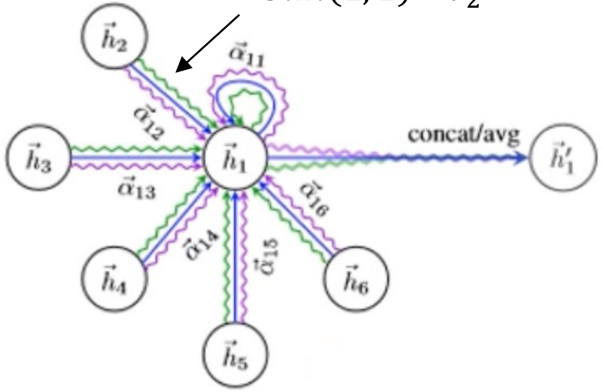
$Sim(I, J)$



노드 I와 노드 J로부터

V_2

$Sim(1, 2) * V_2$



	$Sim(I, J)$	V_2	$Sim(1, 2) * V_2$
Key (K)	Similarity Sim(K, Q)	Value (V)	Sim(K, Q) * V
레몬	0.35	새콤한맛	0.35 * 새콤한맛
오렌지	0.64	달콤한맛	0.64 * 달콤한맛
아보카도	0.01	크레파스맛	0.01 * 크레파스맛

4. Graph Attention Network

Aggregate

유사도 e_{ij} 계산

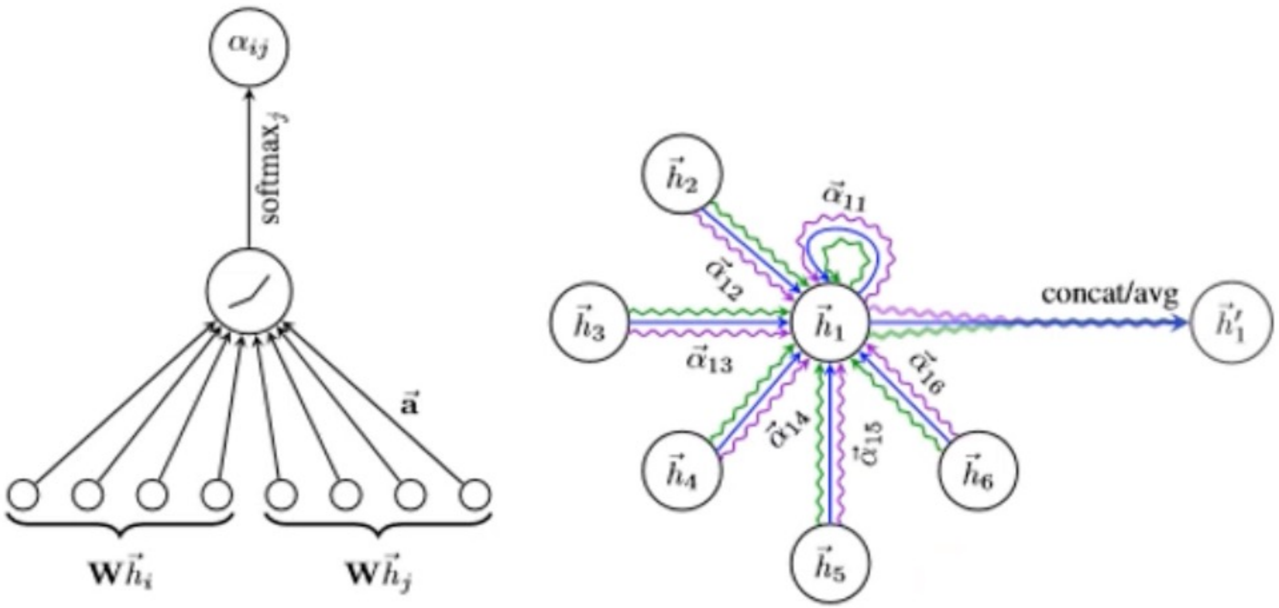
어텐션 스코어 a_{ij} 계산

Combine

이웃들의 정보 $a_{ij}Wh_j$ 집계

Multi-head attention

Readout



4. Graph Attention Network

Aggregate

유사도 e_{ij} 계산

어텐션 스코어 a_{ij} 계산

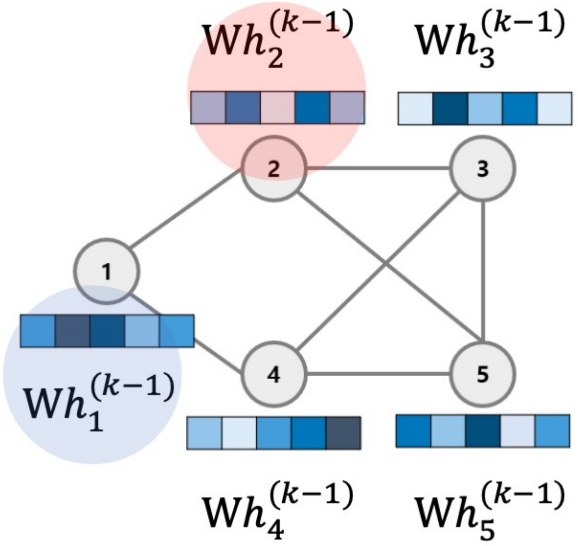
Combine

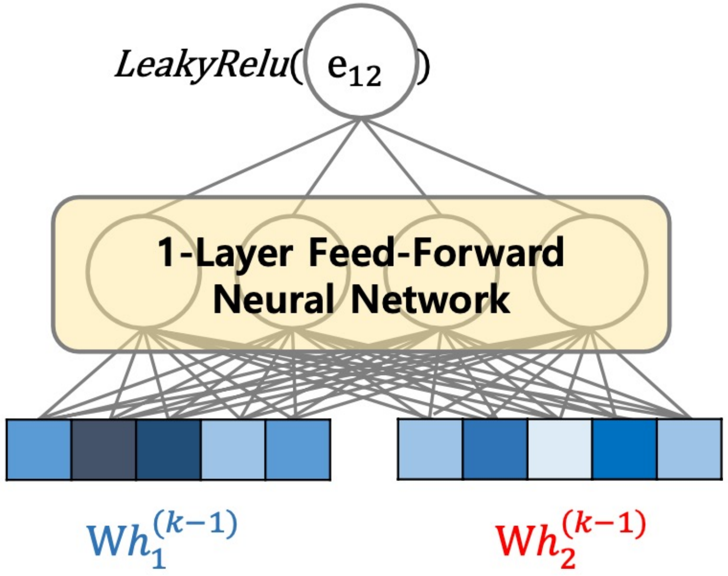
이웃들의 정보 $a_{ij}Wh_j$ 집계

Multi-head attention

Readout

아까의 $Sim(I,J)$ 와 유사





4. Graph Attention Network

Aggregate

유사도 e_{ij} 계산

어텐션 스코어 a_{ij} 계산

Combine

이웃들의 정보 $a_{ij}Wh_j$ 집계

Multi-head attention

Readout

아까의 $Sim(I,J)$ 와 유사

모든 노드들간의 유사도 계산

e_{11}	e_{12}	e_{13}	e_{14}	e_{15}
e_{21}	e_{22}	e_{23}	e_{24}	e_{25}
e_{31}	e_{32}	e_{33}	e_{34}	e_{35}
e_{41}	e_{42}	e_{43}	e_{44}	e_{45}
e_{51}	e_{52}	e_{53}	e_{54}	e_{55}

병렬적으로 계산 가능

3. Attention

Dictionary -> Attention

(유사한 정도의 Softmax)

Q: 귤

Key (K)	Similarity Sim(K, Q)	Value (V)	Sim(K, Q) * V
레몬	0.35	새콤한맛	0.35 * 새콤한맛
오렌지	0.64	달콤한맛	0.64 * 달콤한맛
아보카도	0.01	크레파스맛	0.01 * 크레파스맛

Attention Score

-> 새콤달콤한맛!

4. Graph Attention Network

Aggregate

유사도 e_{ij} 계산

어텐션 스코어 a_{ij} 계산

Combine

이웃들의 정보 $a_{ij}Wh_j$ 집계

Multi-head attention

Readout

모든 노드들간의 유사도

e_{11}	e_{12}	e_{13}	e_{14}	e_{15}
e_{21}	e_{22}	e_{23}	e_{24}	e_{25}
e_{31}	e_{32}	e_{33}	e_{34}	e_{35}
e_{41}	e_{42}	e_{43}	e_{44}	e_{45}
e_{51}	e_{52}	e_{53}	e_{54}	e_{55}

Similarity

병렬적으로 계산 가능

연결된 노드들만 남기기

e_{11}	e_{12}	$-9e^{10}$	e_{14}	$-9e^{10}$
e_{21}	e_{22}	e_{23}	$-9e^{10}$	e_{25}
$-9e^{10}$	e_{32}	e_{33}	e_{34}	e_{35}
e_{41}	$-9e^{10}$	e_{43}	e_{44}	e_{45}
$-9e^{10}$	e_{52}	e_{53}	e_{54}	e_{55}

Masking

구현상의 이유로 마스킹 사용

(각 행은 y번 노드의 모든 이웃)

a_{11}	a_{12}	0	a_{14}	0
a_{21}	a_{22}	a_{23}	0	a_{25}
0	a_{32}	a_{33}	a_{34}	a_{35}
a_{41}	0	a_{43}	a_{44}	a_{45}
0	a_{52}	a_{53}	a_{54}	a_{55}

Softmax

4. Graph Attention Network

Aggregate

유사도 e_{ij} 계산

어텐션 스코어 a_{ij} 계산

Combine

이웃들의 정보 $a_{ij}Wh_j$ 집계

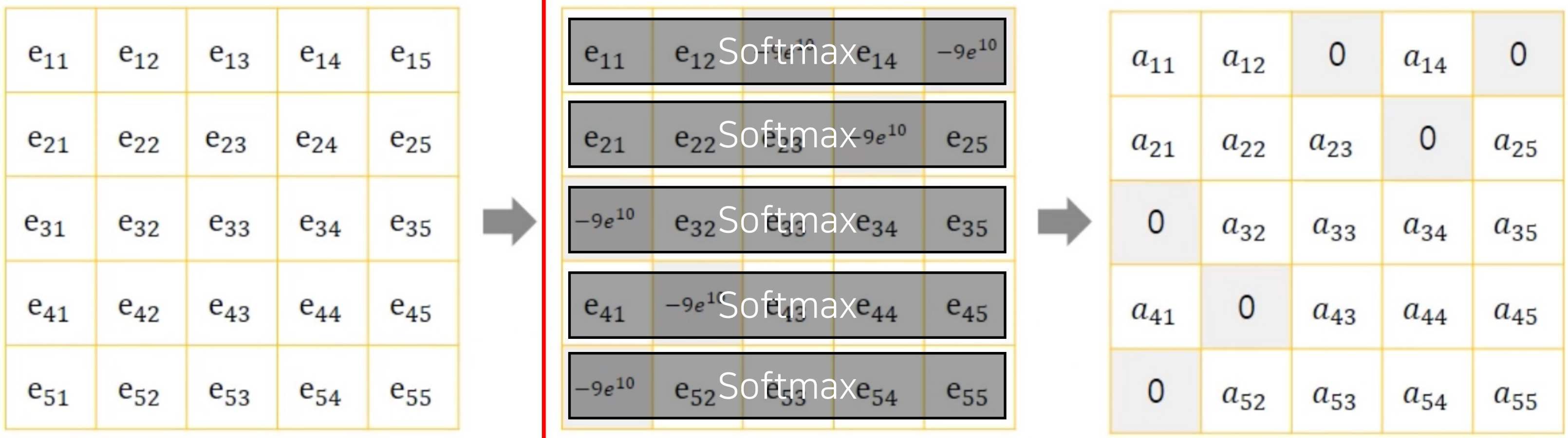
Multi-head attention

Readout

모든 노드들간의 유사도

연결된 노드들만 남기기

Softmax



Similarity

Masking

Softmax

병렬적으로 계산 가능

구현상의 이유로 마스킹 사용

(각 행은 y번 노드의 모든 이웃)

각 행마다, y번 노드 기준
모든 이웃들과의 Attention score

4. Graph Attention Network

Aggregate

유사도 e_{ij} 계산

어텐션 스코어 a_{ij} 계산

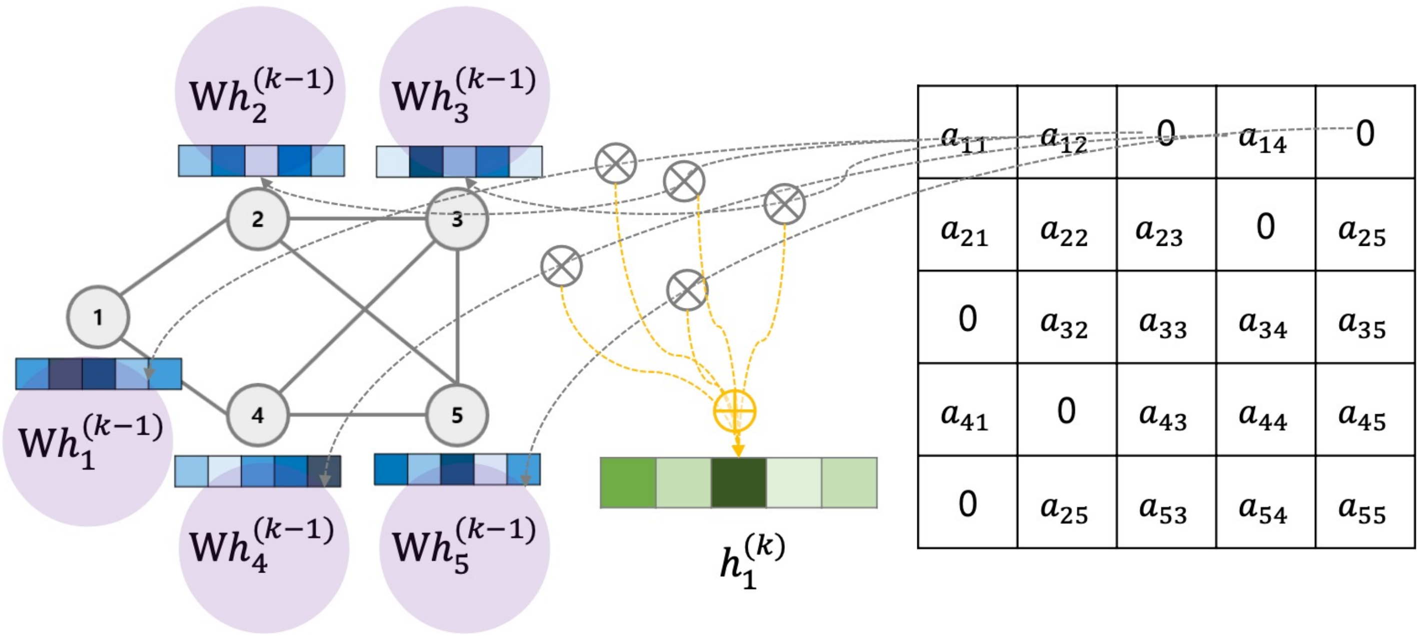
Combine

이웃들의 정보 $a_{ij}Wh_j$ 집계

Multi-head attention

Readout

Key (K)	Score(K, Q)	Value (V)	Score(K, Q) * V
레몬	0.35	새콤한맛	0.35 * 새콤한맛
오렌지	0.64	달콤한맛	0.64 * 달콤한맛
아보카도	0.01	크레파스맛	0.01 * 크레파스맛



4. Graph Attention Network

Aggregate

유사도 e_{ij} 계산

어텐션 스코어 a_{ij} 계산

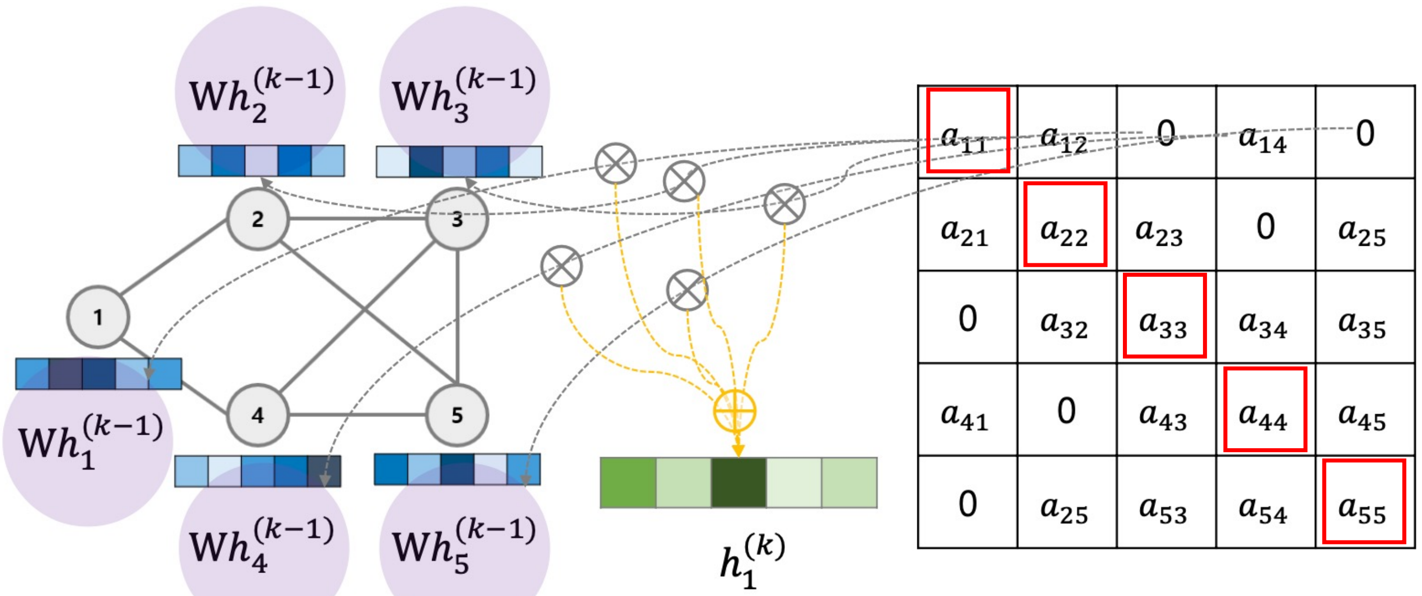
Combine

이웃들의 정보 $a_{ij}Wh_j$ 집계

Multi-head attention

Readout

Key (K)	Score(K, Q)	Value (V)	Score(K, Q) * V
레몬	0.35	새콤한맛	0.35 * 새콤한맛
오렌지	0.64	달콤한맛	0.64 * 달콤한맛
아보카도	0.01	크레파스맛	0.01 * 크레파스맛



모든 노드에 Self-loop 존재

4. Graph Attention Network

Aggregate

유사도 e_{ij} 계산

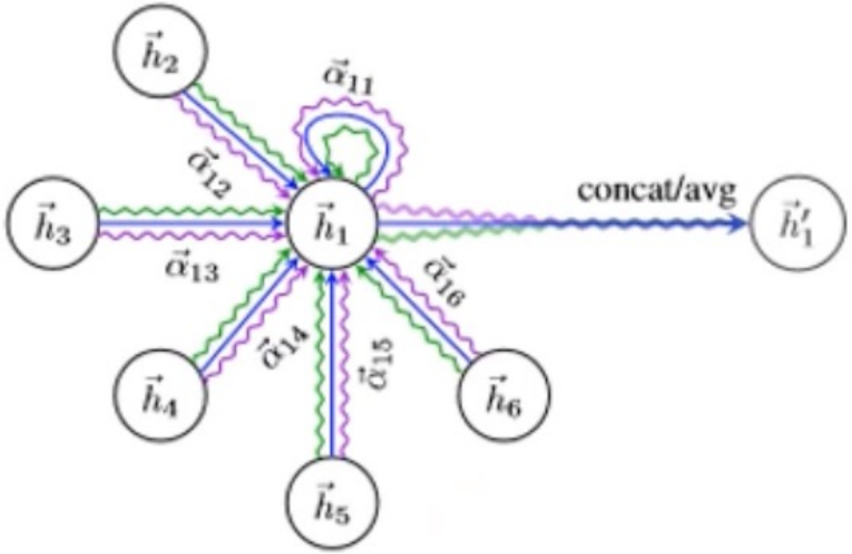
어텐션 스코어 a_{ij} 계산

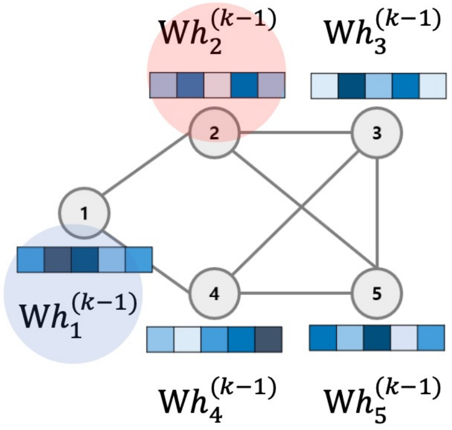
Combine

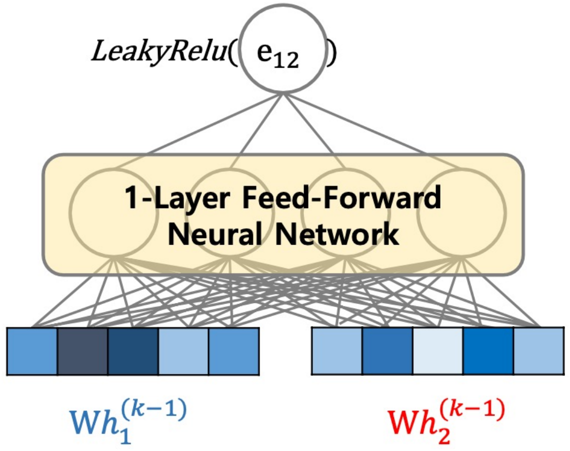
이웃들의 정보 $a_{ij}Wh_j$ 집계

Multi-head attention

Readout







여러 개의 W를 사용
-> 여러 의미에서의 유사도
-> 여러 의미에서의 어텐션

여러 어텐션으로 계산된 값을
평균/concat하여 성능 향상

4. Evaluation

Method	Cora	Citeseer	Pubmed
MLP	55.1%	46.5%	71.4%
ManiReg (Belkin et al., 2006)	59.5%	60.1%	70.7%
SemiEmb (Weston et al., 2012)	59.0%	59.6%	71.7%
LP (Zhu et al., 2003)	68.0%	45.3%	63.0%
DeepWalk (Perozzi et al., 2014)	67.2%	43.2%	65.3%
ICA (Lu & Getoor, 2003)	75.1%	69.1%	73.9%
Planetoid (Yang et al., 2016)	75.7%	64.7%	77.2%
Chebyshev (Defferrard et al., 2016)	81.2%	69.8%	74.4%
GCN (Kipf & Welling, 2017)	81.5%	70.3%	**79.0%**
MoNet (Monti et al., 2016)	$81.7 \pm 0.5\%$	—	$78.8 \pm 0.3\%$
GCN-64*	$81.4 \pm 0.5\%$	$70.9 \pm 0.5\%$	$\mathbf{79.0} \pm 0.3\%$
GAT (ours)	$\mathbf{83.0} \pm 0.7\%$	$\mathbf{72.5} \pm 0.7\%$	$\mathbf{79.0} \pm 0.3\%$

Node Classification Task (Transductive)

출처, 참고

https://arxiv.org/abs/1710.10903

http://dmqm.korea.ac.kr/activity/seminar/296