Densely Connected Graph Convolutional Networks for Graph-to-Sequence Learning

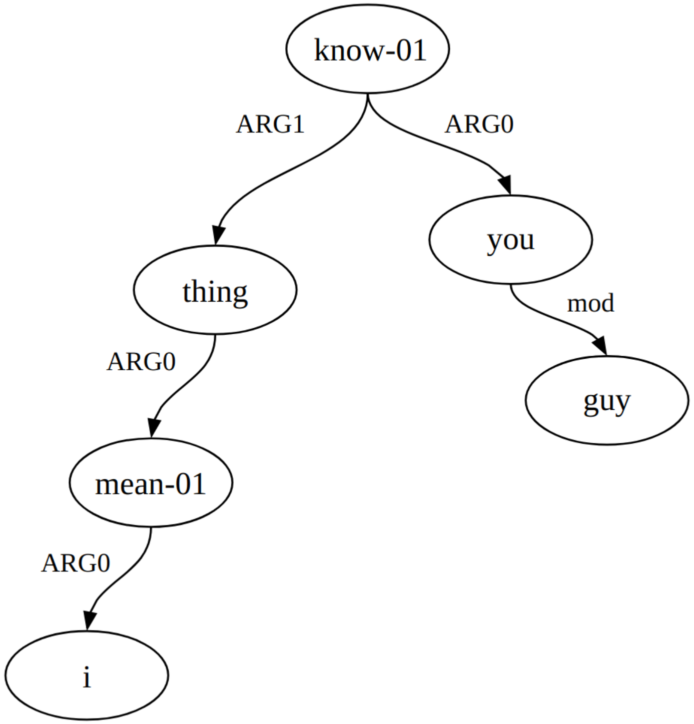
Zhijiang Guo

Joint work with Yan Zhang, Zhiyang Teng, Wei Lu

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

WIAS 浙江西湖高等研究院 WESTLAKE INSTITUTE FOR ADVANCED STUDY

Graph-to-Sequence

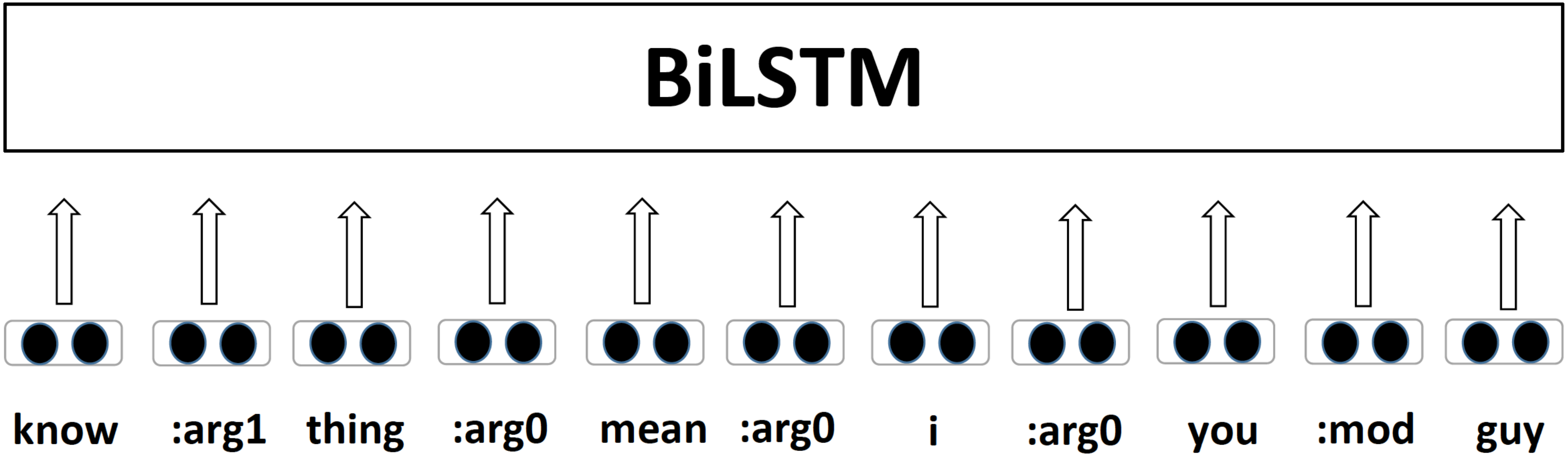


You guys know what I mean

Source: AMR Graph

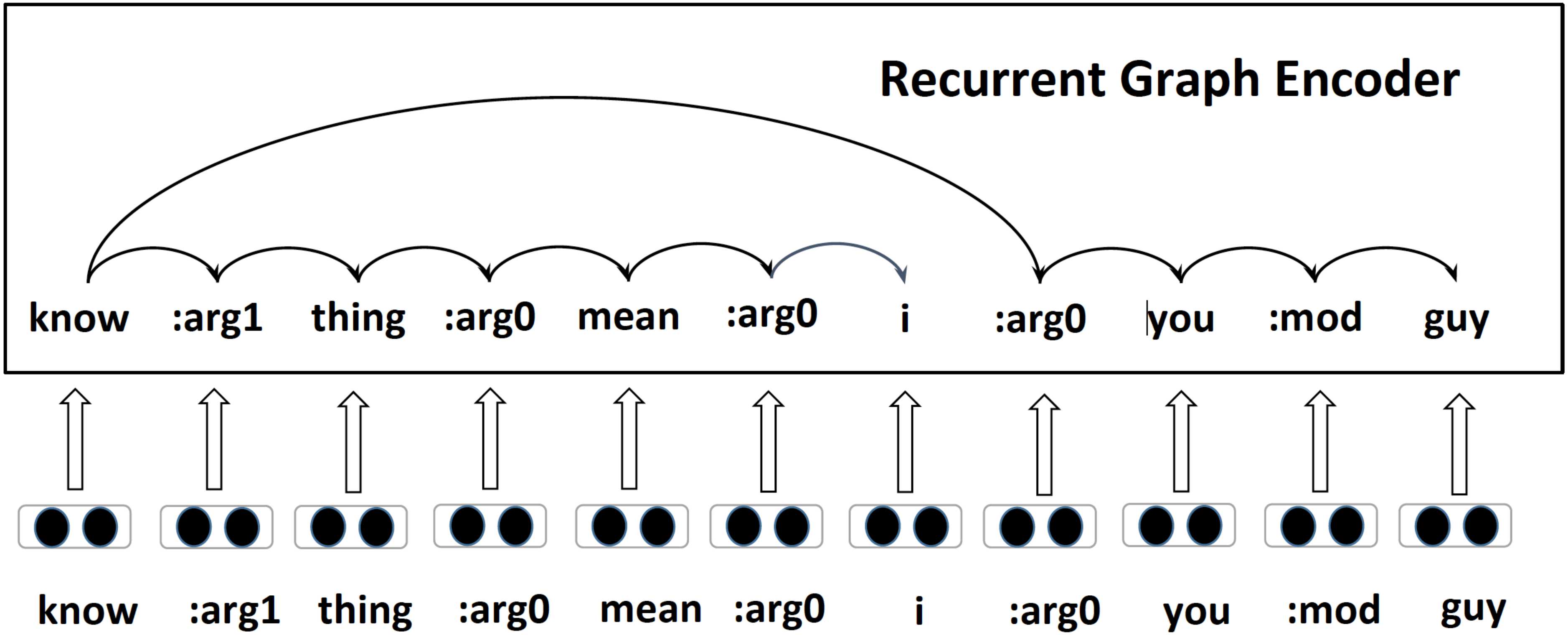
Target: Text Sequence

Previous Work
Sequence Encoder



BiLSTM (Konstas et al., 2017)

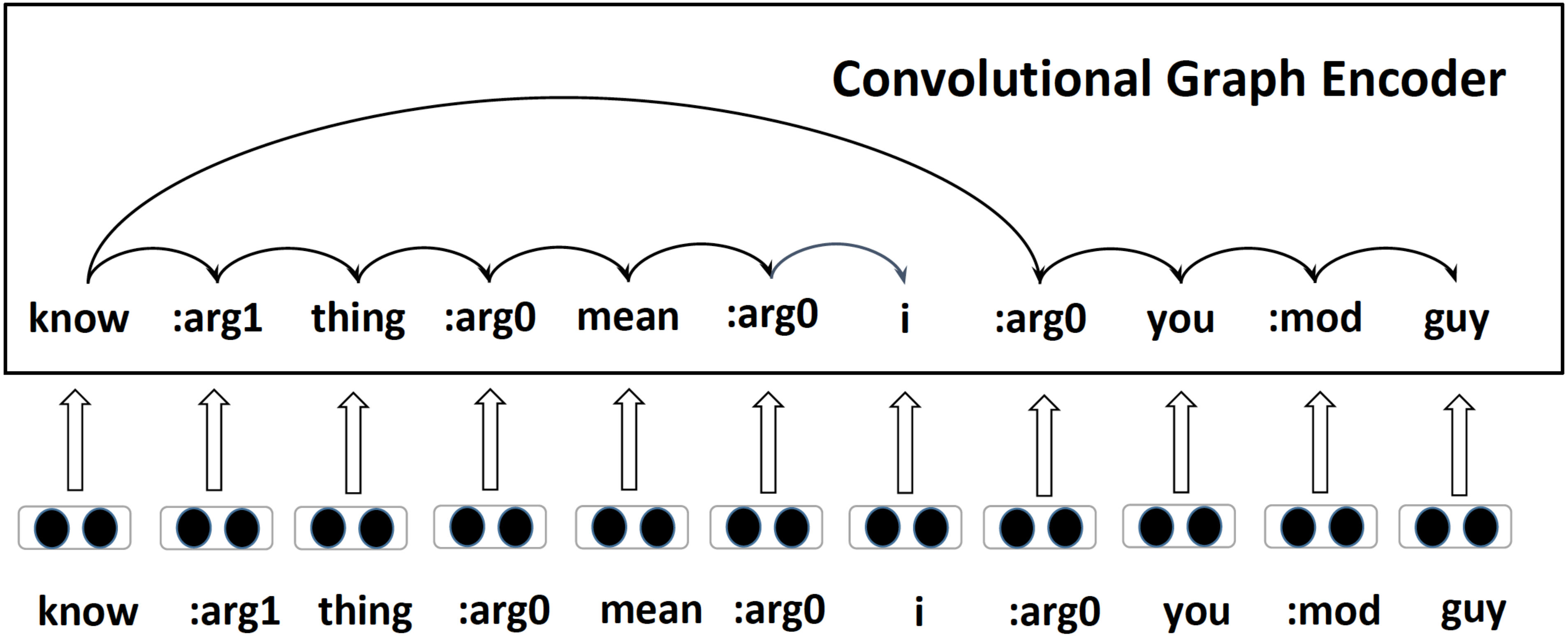
Previous Work
Graph Encoder



Graph State LSTM (Song et al., 2018)

Gated Graph Neural Networks (Beck et al., 2018)

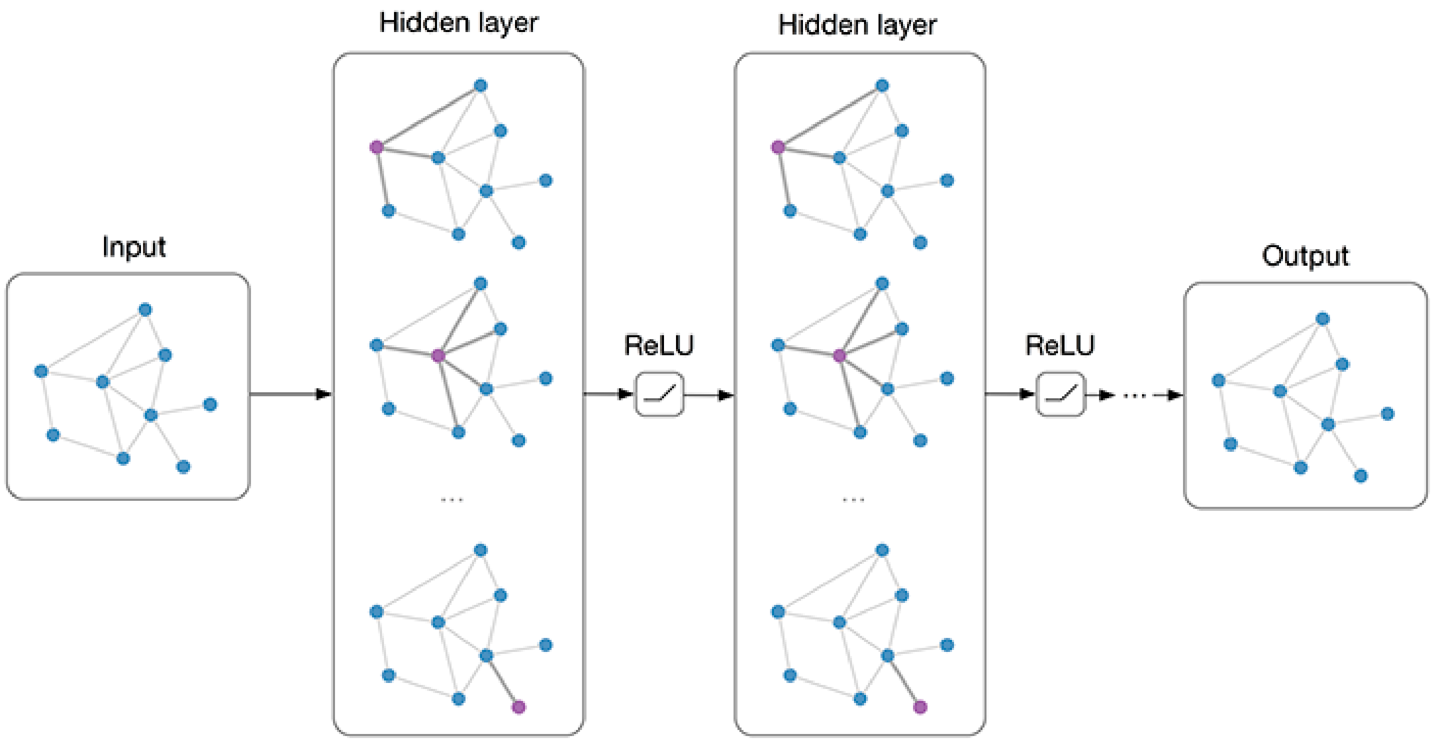
Graph Encoder



GCNs (Kipf and Welling, 2017) ?

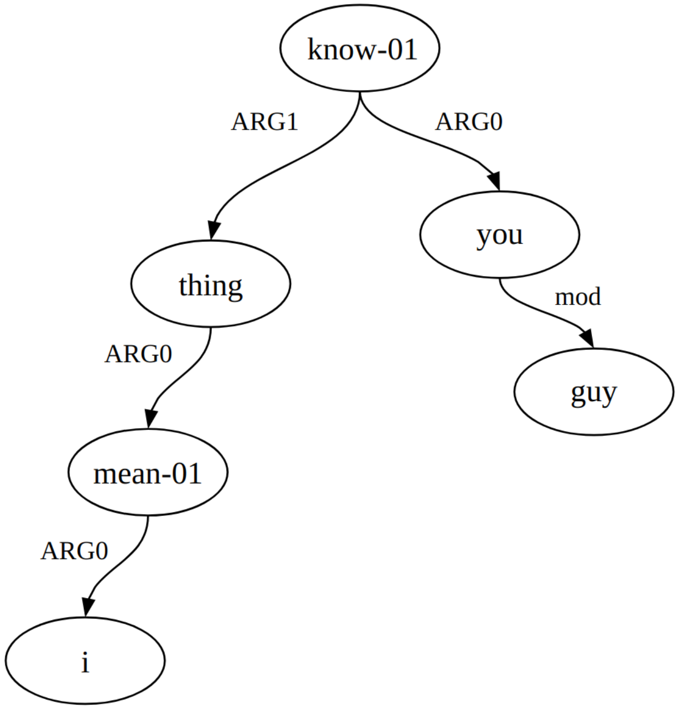
GCNs

GCNs have been successfully applied to NLP tasks (Bastings et al., 2017, Zhang et al., 2018)

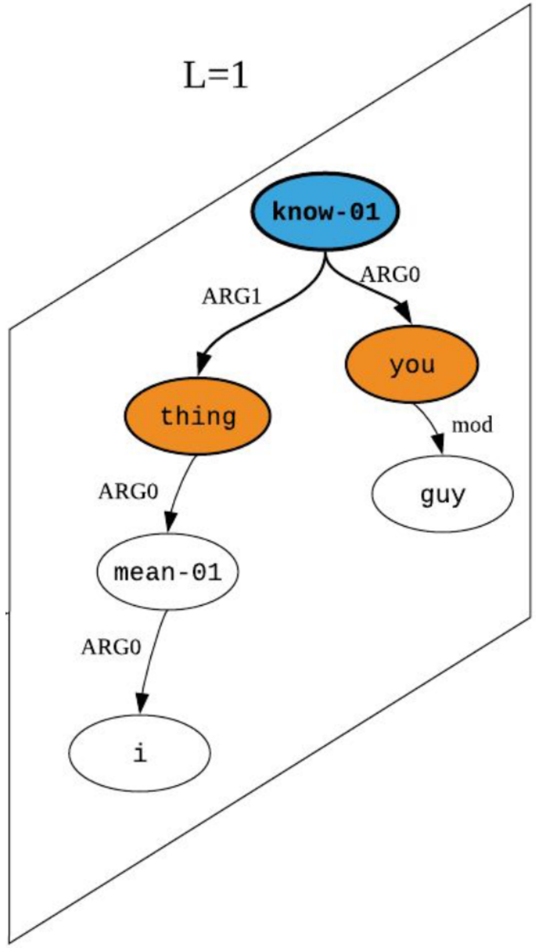


GCNs

First convolutional layer captures **first-order proximity** (immediate neighbors) information

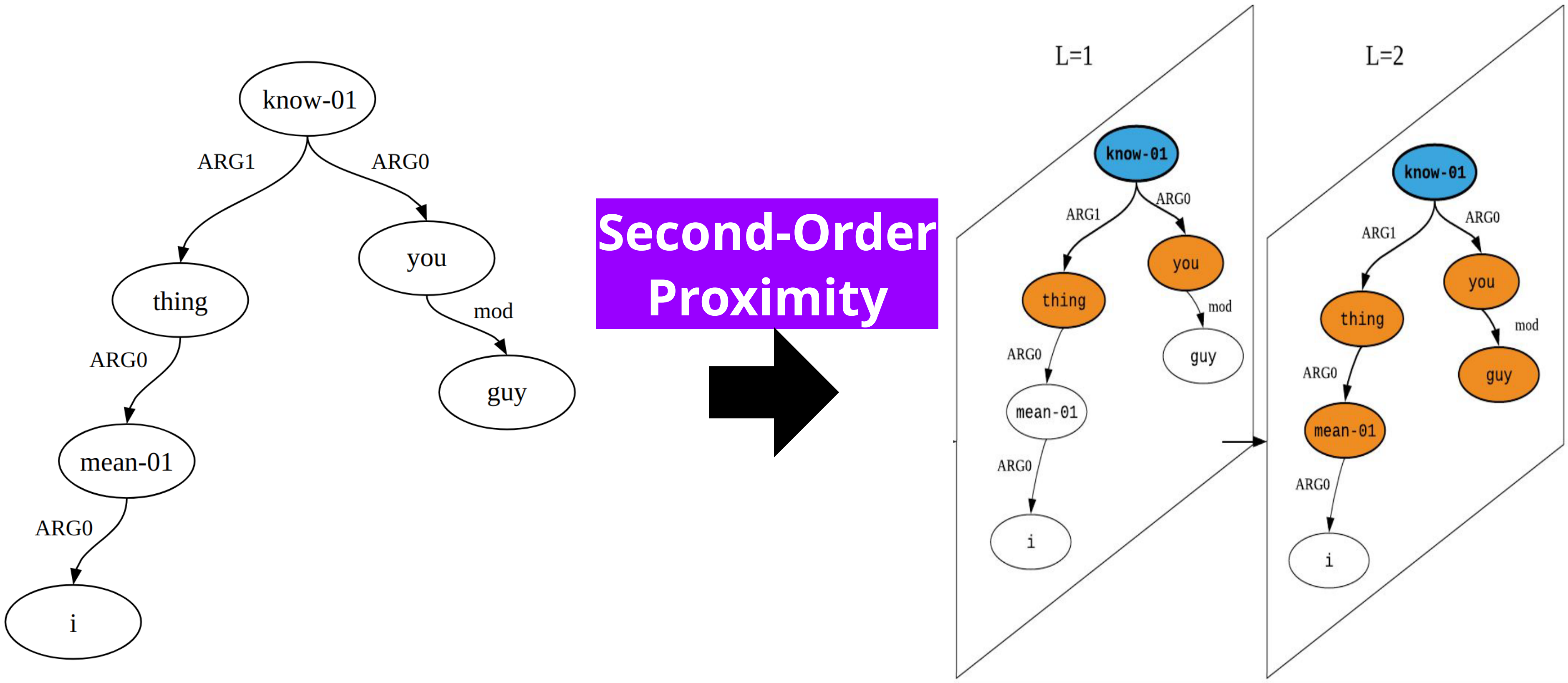


First-Order Proximity



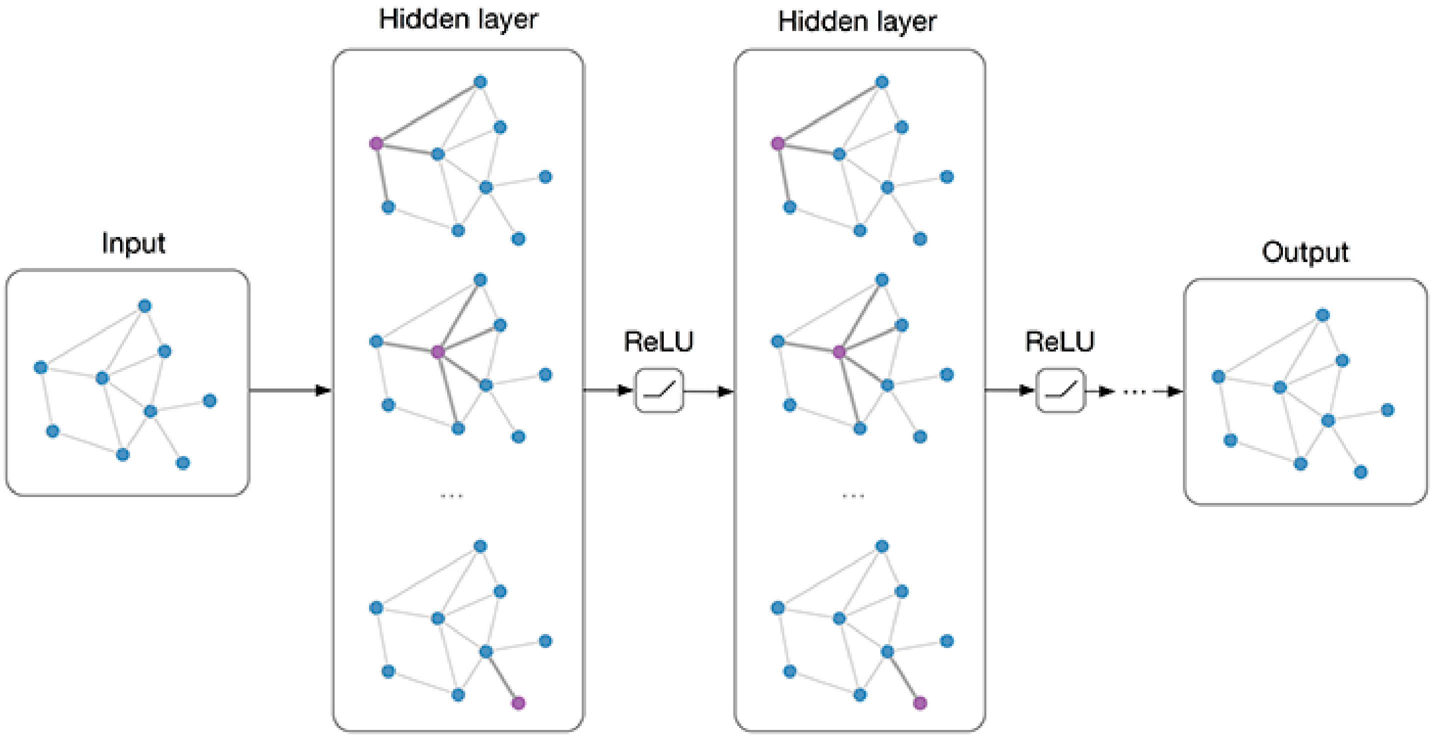
GCNs

Second convolutional layer is able to capture **second-order proximity** information



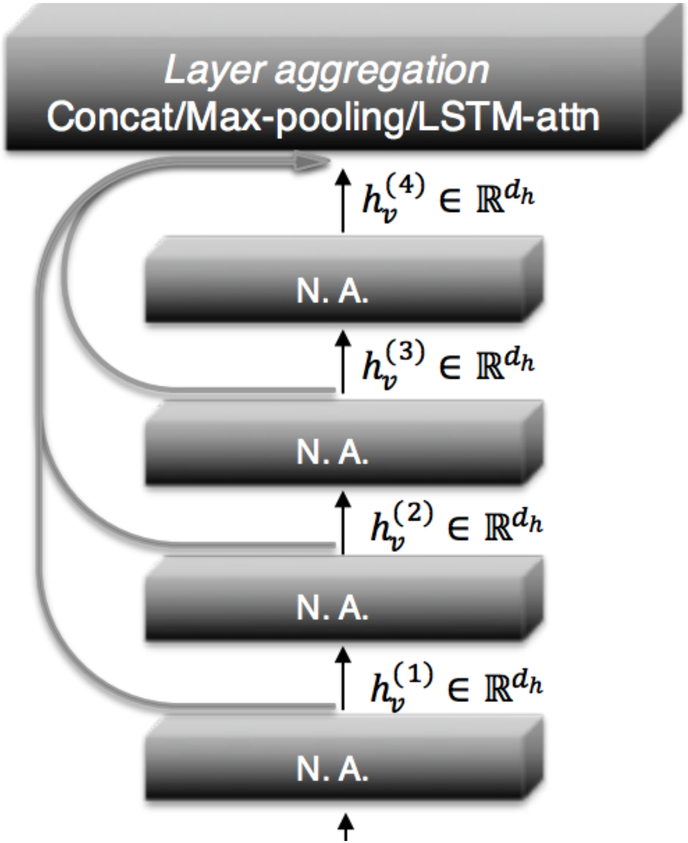
GCNs

Best performance of GCNs is achieved with the relatively shallow **(2-layer)** model (Li et al., 2018)

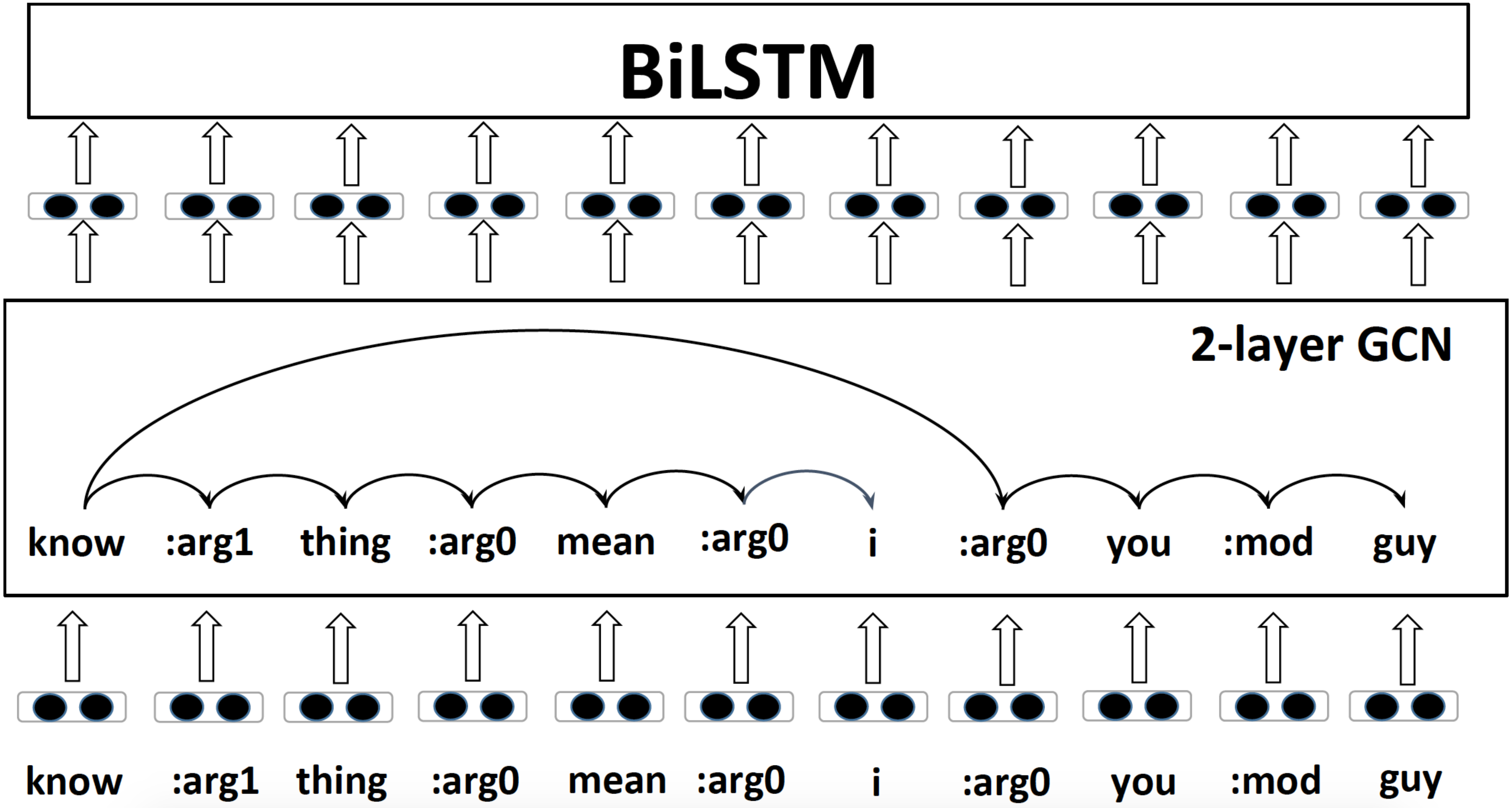


Alternative Solution

LSTM-attention operations are used to aggregate information among layers to capture non-local interactions for larger graphs (Xu et al., 2018)



Alternative Solution



(Damonte and Cohen, 2019)

Better Graph Encoder

Is it possible to build a deep convolutional graph encoder to learn a better graph representation ?

Better Graph Encoder

Is it possible to build a deep convolutional graph encoder to learn a better graph representation ?

Densely Connected Graph Convolutional Networks (DCGCNs)

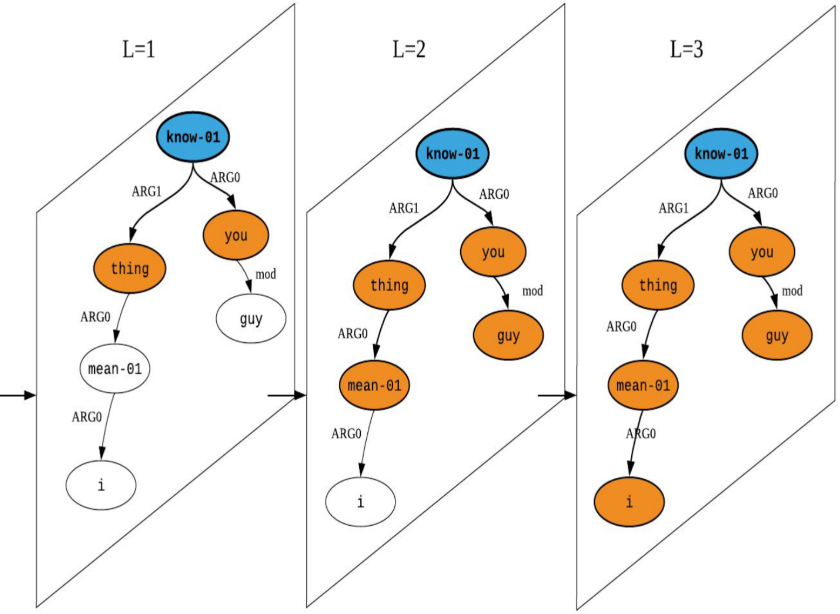
Better Graph Encoder

Is it possible to build a deep convolutional graph encoder to learn a better graph representation ?

Model	**Max GCN Layers**
Jumping Net (Xu et al., 2018)	6
GCN + BiLSTM (Damonte&Cohen, 2019)	2
DCGCNs	**36**

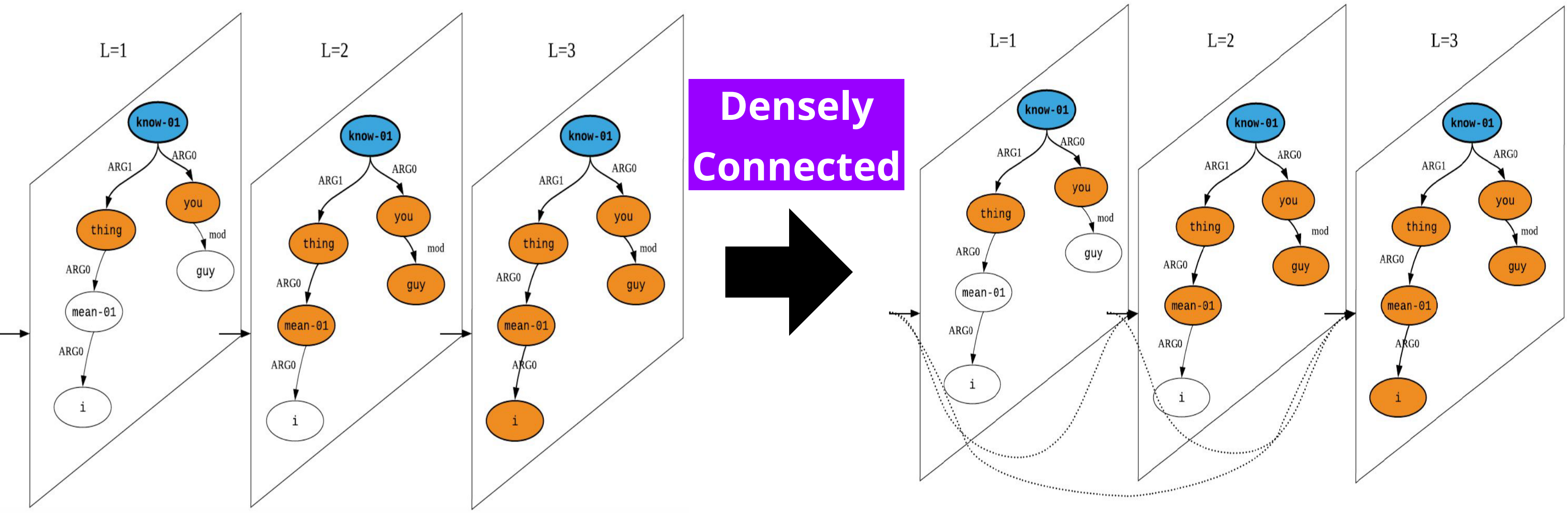
Dense Connectivity

Each layer takes inputs from all preceding layers
(Huang et al., 2017)



Dense Connectivity

Each layer takes inputs from all preceding layers (Huang et al., 2017)

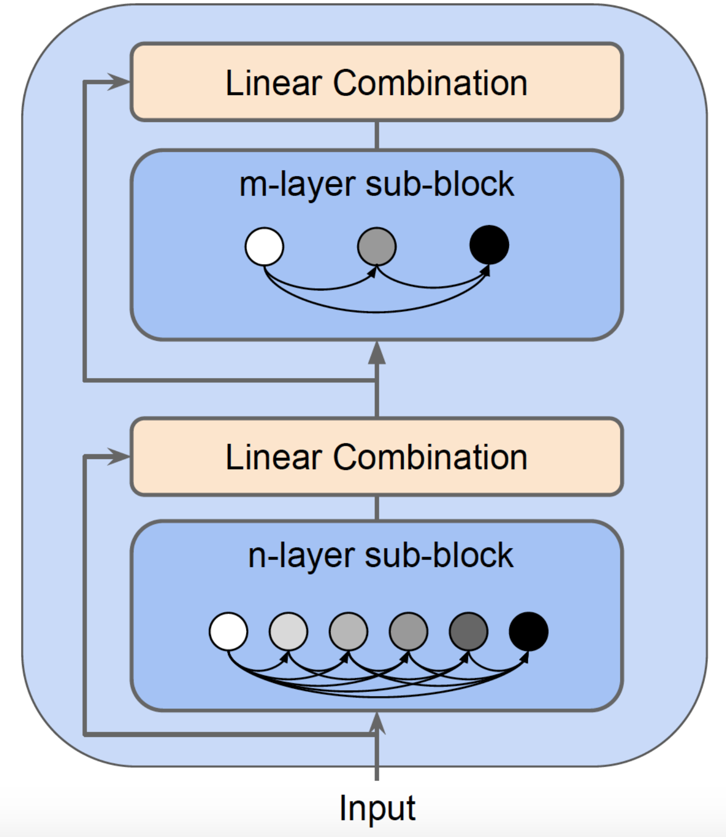


Model

Densely Connected GCNs (DCGCN)

Consists of ***M*** identical Blocks; each has 2 types of components:

- Densely Connected Sub-Block
- Linear Combination Layer

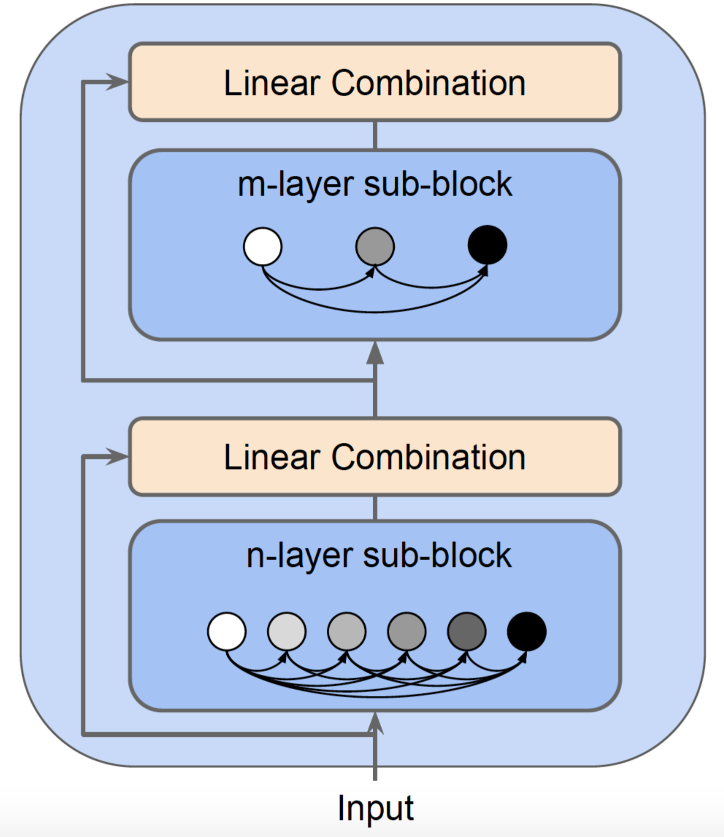


Model

Densely Connected GCNs (DCGCN)

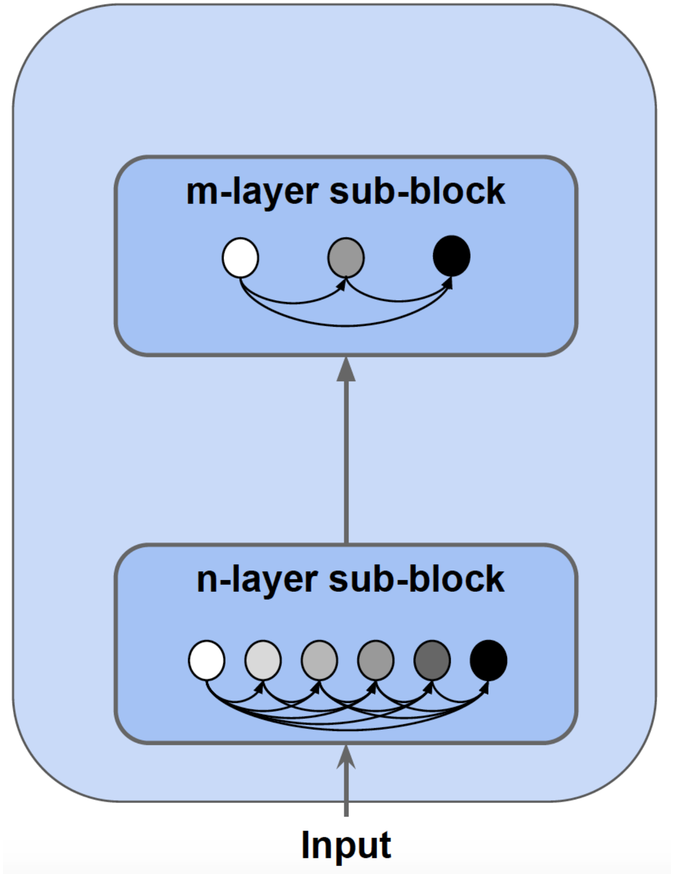
Consists of ***M*** identical Blocks; each has 2 types of components:

- **Densely Connected Sub-Block**
- Linear Combination Layer



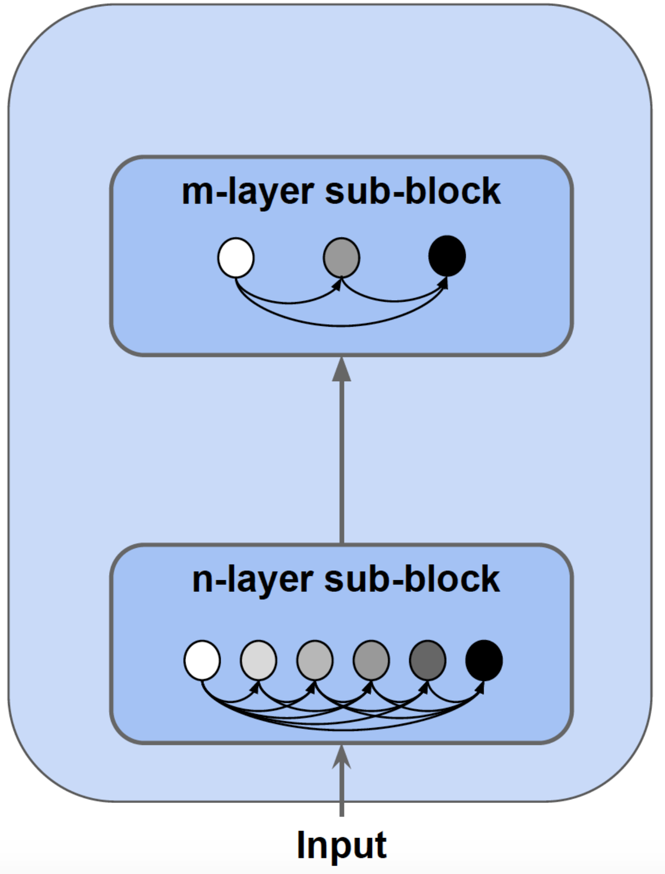
Sub-Block

Each block has **2** sub-blocks. Graph convolutional layers in both sub-blocks are densely connected



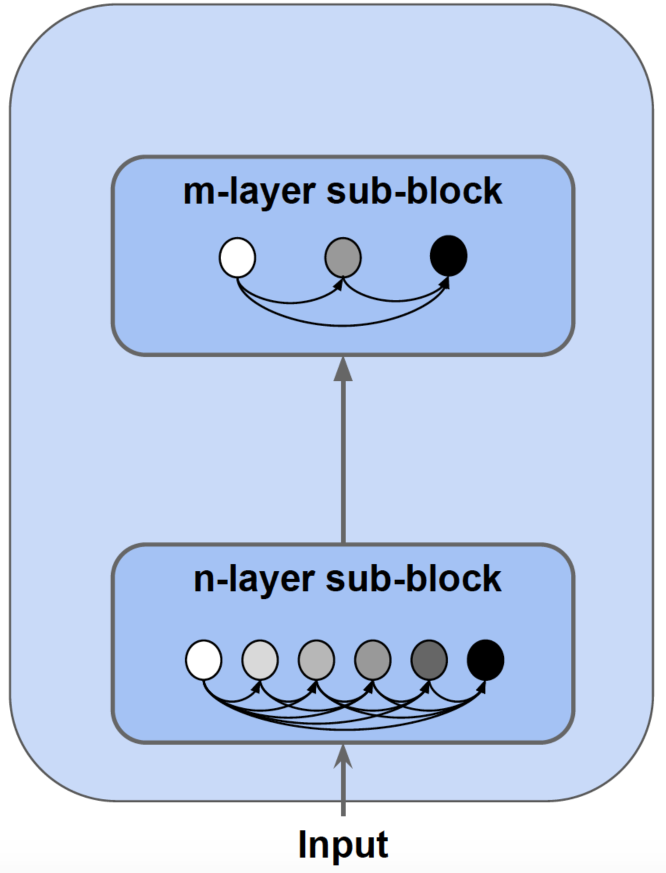
Sub-Block

Intuitively, sub-blocks with different number of layers capture information at **different levels**



Sub-Block

We need to make sure the number of parameters of the deeper DCGCN models is **manageable** and **comparable** to previous methods

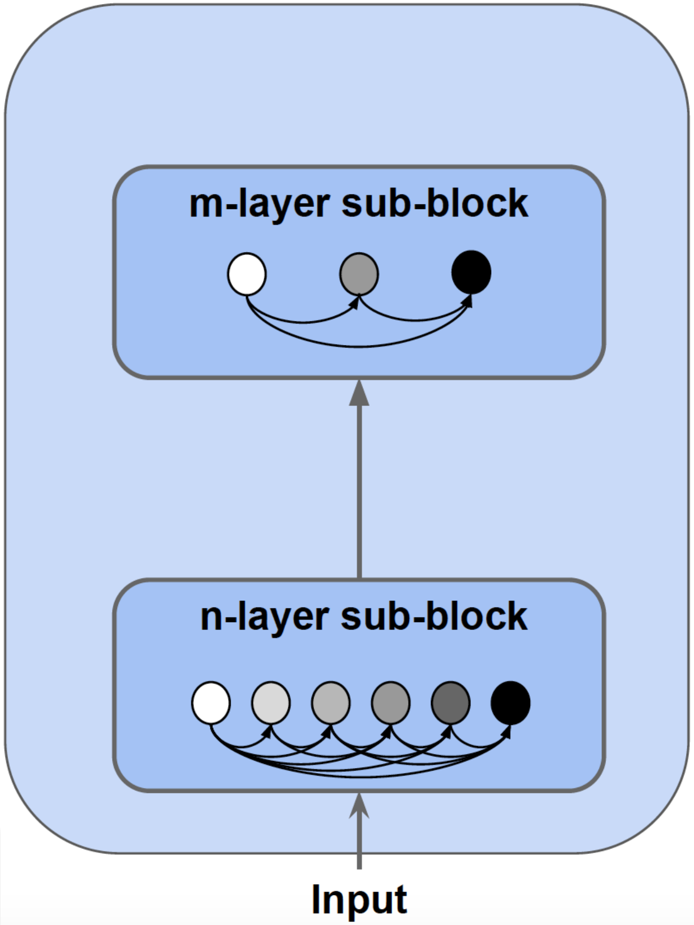


Sub-Block

Output dimension of the sub-block: *D*
(concatenate output from all *L* layers)

Output dimension of each layer:
d = *D* / *L* (proportional to #layers)

Input dimension of the sub-block: *D*
Sub-block layers: *L*

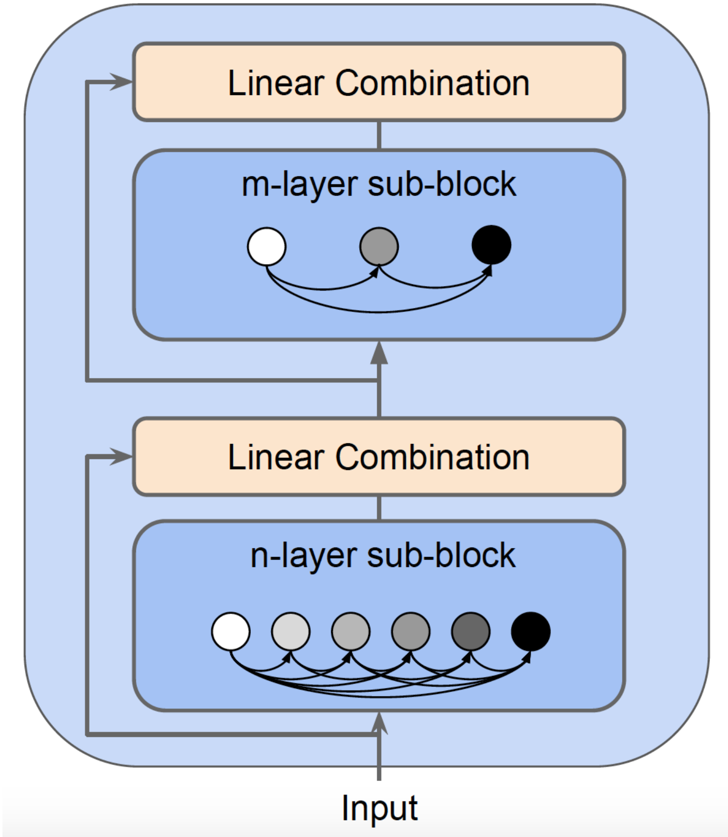


Model

Densely Connected GCNs (DCGCN)

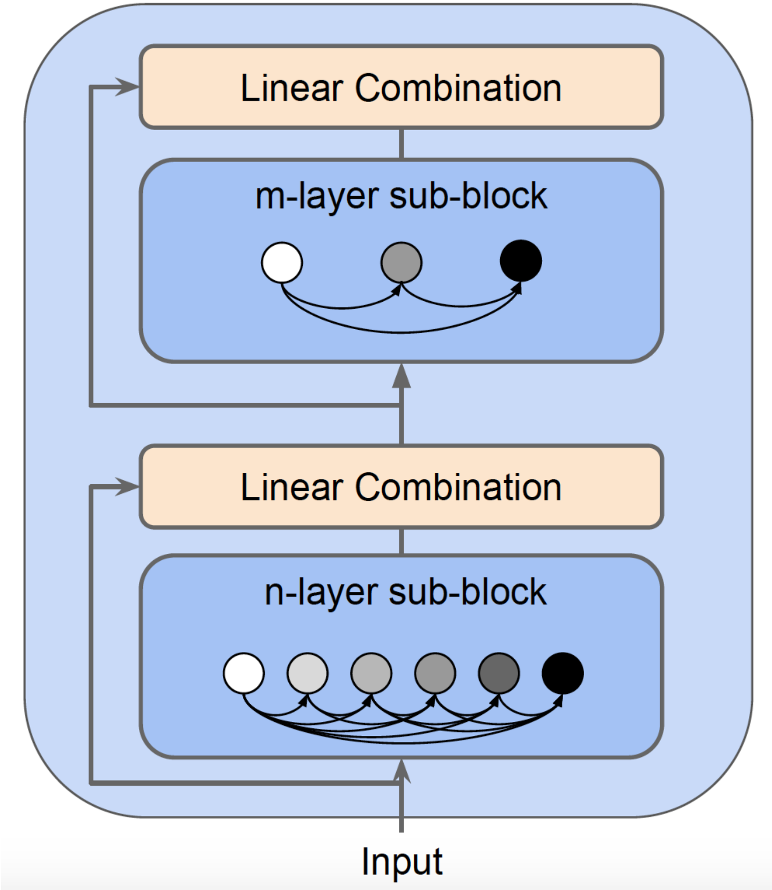
Consists of ***M*** identical Blocks; each has 2 types of components:

- Densely Connected Sub-Block
- **Linear Combination Layer**

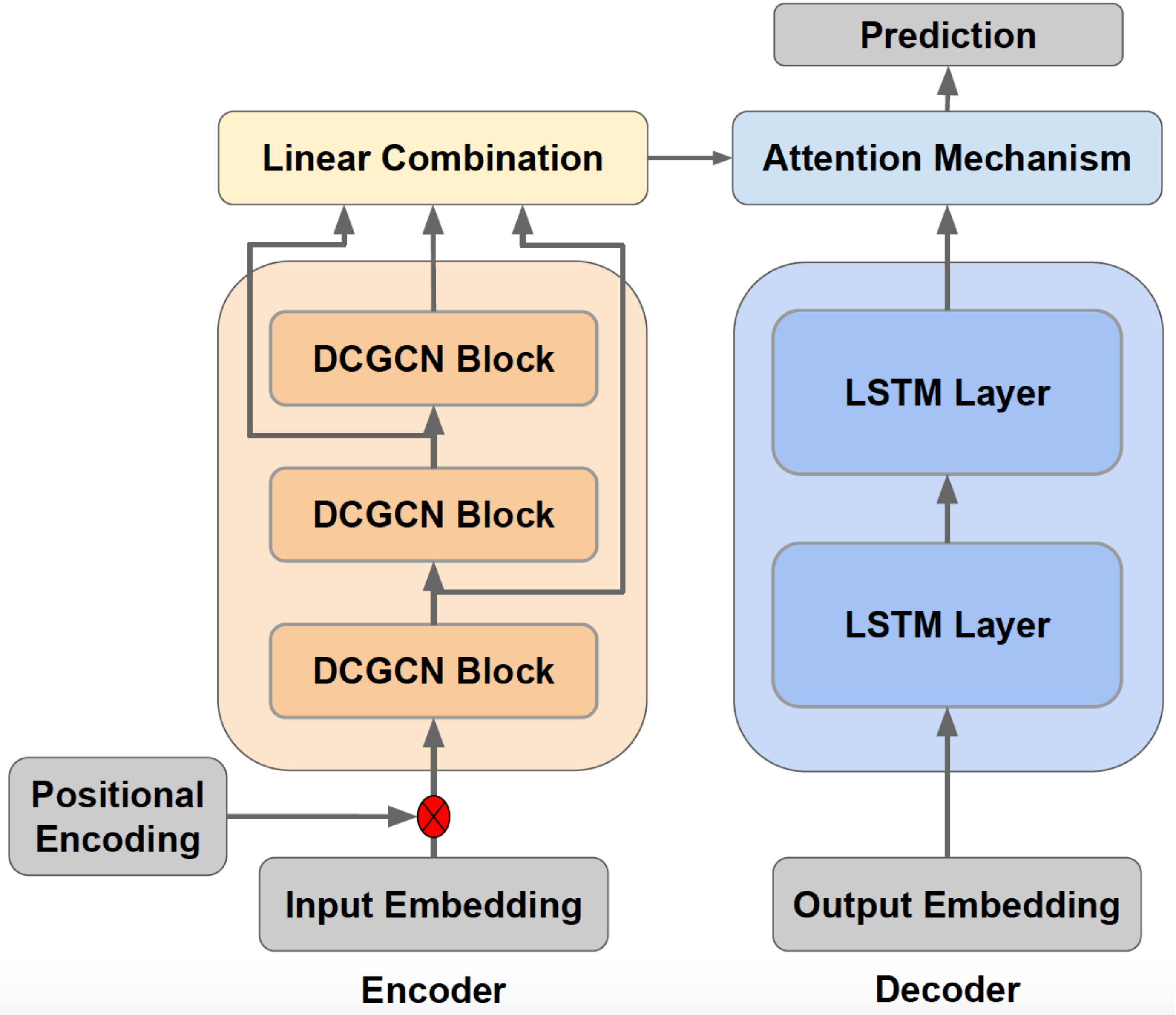


Linear Combination

The layer assigns **different weights** to outputs of **different layers**



Graph-to-Seq Model



Experiments

AMR-to-Text Generation

- AMR 2015 (LDC2015E86)
- AMR 2017 (LDC2017T10)

Syntax-Based Machine Translation

- English-German (WMT16)
- English-Czech (WMT16)

Experiments

AMR-to-Text Generation

- AMR 2015 (LDC2015E86)
- AMR 2017 (LDC2017T10)

Syntax-Based Machine Translation

- English-German (WMT16)
- English-Czech (WMT16)

Data Statistics

Dataset	Train	Dev	Test
AMR 2015	16,833	1,368	1,371
AMR 2017	36,521	1,368	1,371

AMR 2015
Main Results

Model	**Type**	**BLEU**
LSTM (Konstas et al., 2017)	Seq	22.0
GS LSTM (Song et al., 2018)	Graph	23.3
GCNSEQ (Damonte & Cohen, 2019)	Mixed	24.4
DCGCN	Graph	**25.7**

AMR 2015
With External Training Data

Model	**Type**	**Ext**	**BLEU**
LSTM (Konstas et al., 2017)	Seq	0.2M	27.4
GS LSTM (Song et al., 2018)	Graph	0.2M	28.2
DCGCN	Graph	**0.2M**	**31.6**

AMR 2015
With External Training Data

Model	**Type**	**Ext**	**BLEU**
LSTM (Konstas et al., 2017)	Seq	0.2M	27.4
GS LSTM (Song et al., 2018)	Graph	0.2M	28.2
DCGCN	Graph	**0.1M**	**29.0**

AMR 2015
With External Training Data

Model	**Type**	**Ext**	**BLEU**
LSTM (Konstas et al., 2017)	Seq	2M	32.3
LSTM (Konstas et al., 2017)	Seq	20M	33.8
GS LSTM (Song et al., 2018)	Graph	2M	33.6
DCGCN (Single)	Graph	0.3M	33.2
DCGCN (Ensemble)	Graph	**0.3M**	**35.3**

AMR 2017
Single Model Performance

Model	**Type**	**#P**	**B**	**C**
LSTM (Konstas et al., 2017)	Seq	28M	21.7	49.1
GGNNs (Beck et al., 2018)	Graph	28M	23.3	50.4
GCNSEQ (Damonte & Cohen, 2019)	Mixed	N/A	24.5	N/A
DCGCN	Graph	**18M**	**27.6**	**57.3**

#P: Number of Parameters; **B**: BLEU; **C**: CHRF++

AMR 2017
Ensemble Performance

Model	**Type**	**#P**	**B**	**C**
LSTM (Konstas et al., 2017)	Seq	142M	26.6	52.5
GGNNs (Beck et al., 2018)	Graph	141M	27.5	50.4
DCGCN	Graph	**93M**	**30.4**	**59.6**

#P: Number of Parameters; **B**: BLEU; **C**: CHRF++

Experiments

AMR-to-Text Generation

- AMR 2015 (LDC2015E86)
- AMR 2017 (LDC2017T10)

Syntax-Based Machine Translation

- English-German (WMT16)
- English-Czech (WMT16)

Data Statistics

Dataset	**Train**	**Dev**	**Test**
AMR 2015	16,833	1,368	1,371
AMR 2017	36,521	1,368	1,371
En-Cs	181,112	2,656	2,999
En-De	226,822	2,169	2,999

English-German

Model	**Type**	**#P**	**B**	**C**
BoW + GCN (Bastings et al., 2017)	Mixed	N/A	12.2	N/A
CNN + GCN (Bastings et al., 2017)	Mixed	N/A	13.7	N/A
RNN + GCN (Bastings et al., 2017)	Mixed	N/A	16.1	N/A
Seq2Seq (Konstas et al., 2017)	Seq	41.4M	15.5	40.8
GGNNs (Beck et al., 2018)	Graph	41.2M	16.7	42.4
DCGCN	Graph	**29.7M**	**19.0**	**44.1**

#P: Number of Parameters; **B**: BLEU; **C**: CHRF++

English-Czech

Model	**Type**	**#P**	**B**	**C**
BoW + GCN (Bastings et al., 2017)	Mixed	N/A	7.5	N/A
CNN + GCN (Bastings et al., 2017)	Mixed	N/A	8.7	N/A
RNN + GCN (Bastings et al., 2017)	Mixed	N/A	9.6	N/A
Seq2Seq (Konstas et al., 2017)	Seq	39.1M	8.9	33.8
GGNNs (Beck et al., 2018)	Graph	38.8M	9.8	33.3
DCGCN	Graph	**28.3M**	**12.1**	**37.1**

#P: Number of Parameters; **B**: BLEU; **C**: CHRF++

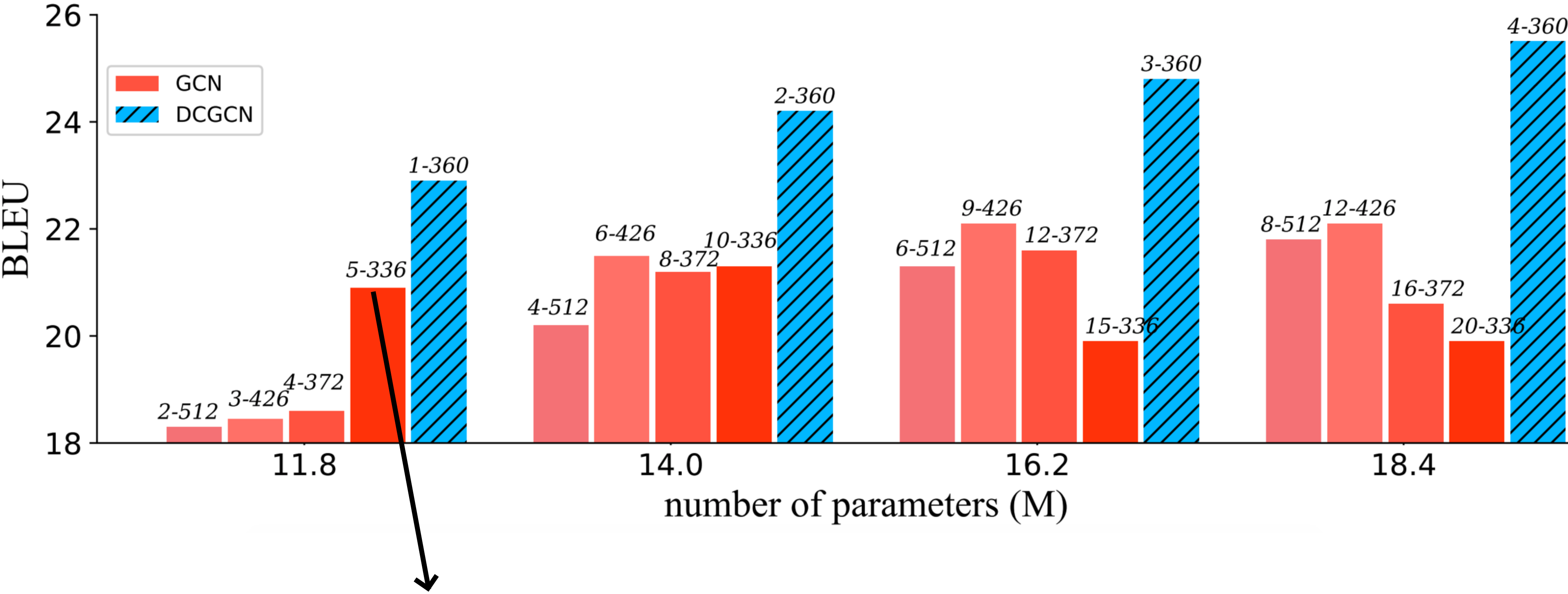
Ablation Test

Model	BLEU
DCGCN	25.5
- Dense Connections	23.2
- Linear Combination (LC)	23.7
- Direction Aggregation	24.6
- Global Node (GN)	24.2

Ablation Test

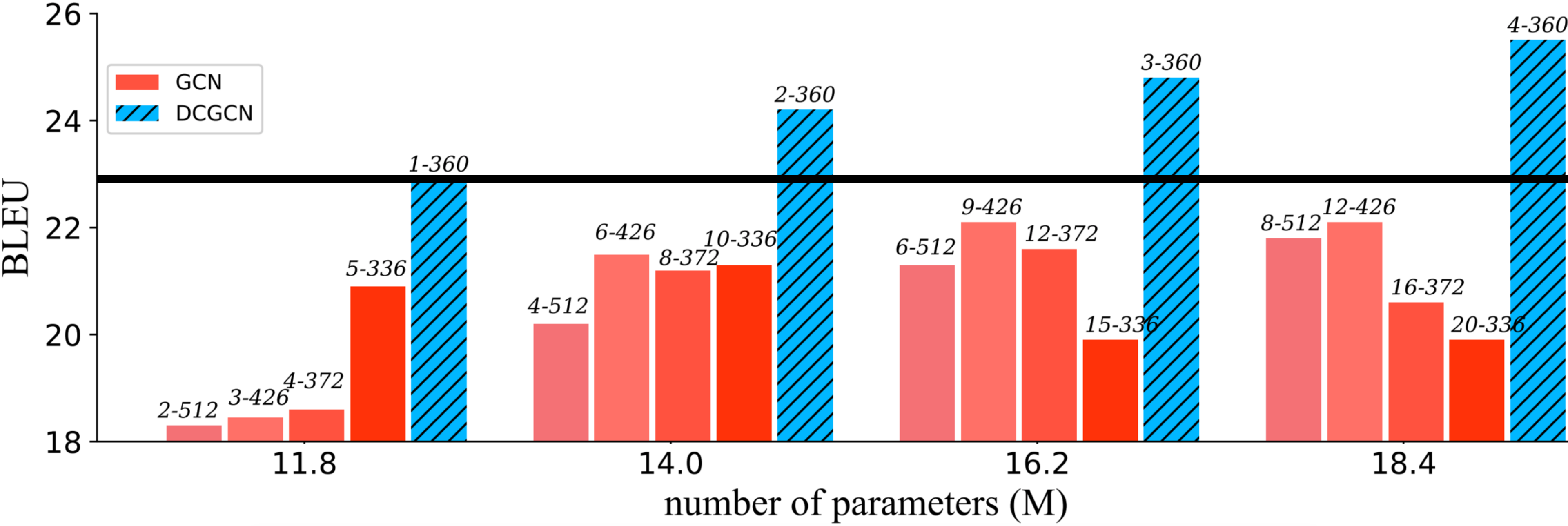
Model	BLEU
DCGCN	25.5
- Dense Connections	23.2
- Linear Combination (LC)	23.7
- Direction Aggregation	24.6
- Global Node (GN)	24.2

Results vs #Parameters



(a 5-layers GCN with the hidden size 336)

Results vs #Parameters



Conclusion

Contribution

A novel GCN model with deeper layers that allows the encoder to better capture the structural information of the graph

Future Work

Explore how other NLP applications can potentially benefit from the proposed approach

Thank You

Code Available

`http://statnlp.org/research/ml`

Graph-to-Sequence

AMR-to-Text Generation

Source: AMR graph

Target: natural language sequence

Syntax-Based Machine Translation (En-De)

Source: English sentence + dependency tree

Target: German sentence

Graph-to-Sequence

AMR-to-Text Generation

Source: AMR graph

Target: natural language sequence

Syntax-Based Machine Translation (En-De)

Source: English sentence + dependency tree

Target: German sentence