GDPNet: Refining Latent Multi-View Graph for Relation Extraction

Fuzhao Xue [1], Aixin Sun[1], Hao Zhang[1,2] ,Eng Siong Chng[1]

[1] *School of Computer Science and Engineering, Nanyang Technological University, Singapore*

[2]*Institute of High Performance Computing, A*STAR, Singapore*

AAAI 2021

What is Relation Extraction (RE)?

Given two entities and a piece of text where the two entities are mentioned in, the task of RE is to predict the semantic relation between the two entities.

- **For sentence-level RE, the context is a single sentence**
- **For dialogue-level RE, the context is a dialogue**

An Example in TACRED Dataset (Sentence-level RE)

Sentence: Edsel Ford, the only child of Henry Ford, died in New York

Entity types: PERSON/PERSON

Relation: per:parents

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

SINGAPORE Agency for Science, Technology and Research

Motivation

When the given text is long, it is challenging to identify indicative words for the relation prediction.

An Example in DialogRE Dataset (Dialogue-level Relation Extraction)

S1: Hey Pheebs.

S2: Hey!

S1: Any sign of your brother?

S2: No, but he's always late.

S1: I thought you only met him once?

S2: Yeah, I did. I think it sounds y'know big sistery, y'know, 'Frank's always late.'

S1: Well relax, he'll be here..

Argument pair	Trigger	Relation type
R1 (Frank, S2)	brother	per:siblings
R2 (S2, Frank)	brother	per:siblings
R3 (S2, Pheebs)	none	per:alternate names
R4 (S1, Pheebs)	none	unanswerab

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

SINGAPORE Agency for Science, Technology and Research

Related Work

RNN-based Relation Extraction
e.g. LSTM, PA-LSTM, SA-LSTM

Graph-based Relation Extraction
e.g. C-GCN, C-AGGCN, LST-AGCN
- *Difference*: 1) Our graph is a latent multi-view graph; 2) We focus on refining this graph.

BERT-based Relation Extraction
e.g. BERT, BERTs, SpanBERT
- *Difference*: 1) our multi-view graph is built on top of token representations by BERT.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Agency for Science, Technology and Research SINGAPORE

Preliminary

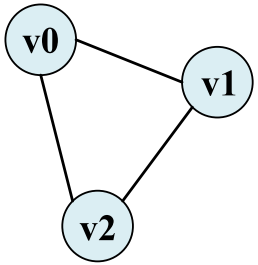
Problem Formulation

- *Input*: a text sequence $X = \{x_1, x_2, \ldots, x_T\}$, and two entities $X_s = \{x_s, x_{s+1}, \ldots, x_{s+m-1}\}$ and $X_o = \{x_o, x_{o+1}, \ldots, x_{o+n-1}\}$.
- *Output*: relation $r \in R$ between X_s and X_o .

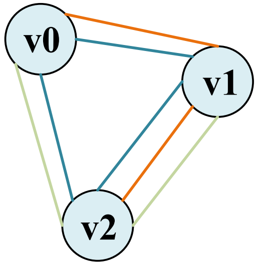
Multi-View Graph

- In a multi-view graph, there exist multiple edges between a pair of nodes, each edge from one view.
- Formally, we represent a multi-view graph as $G = (V, A_1, A_2, \ldots, A_N)$, where V is the set of nodes, A is an adjacent matrix, and N is the number of views.

Homogeneous Graph



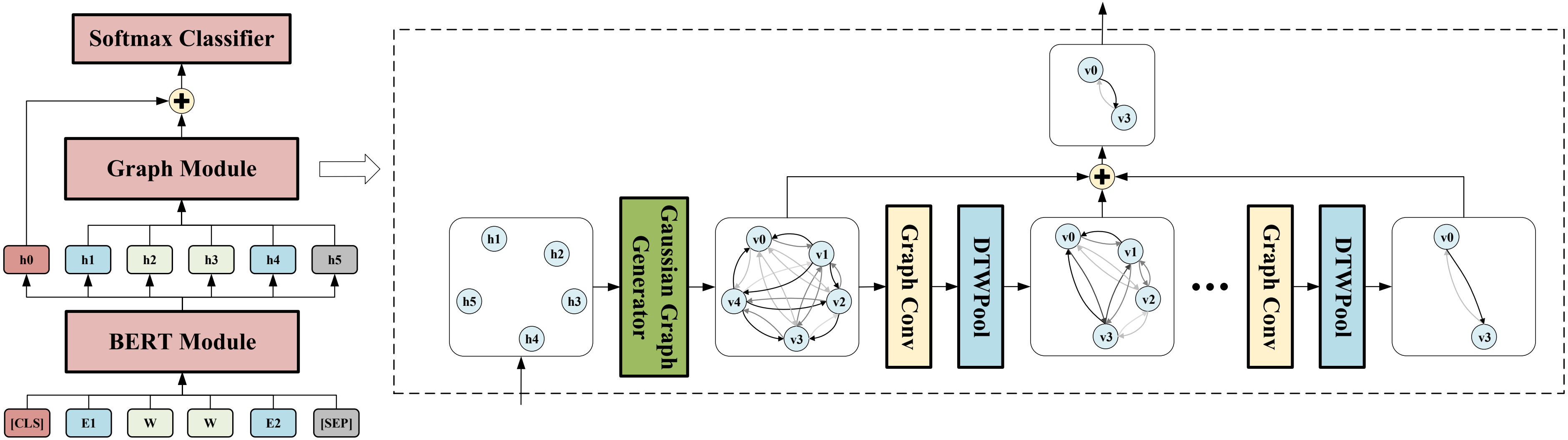
Multi-View Graph



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Agency for Science, Technology and Research SINGAPORE

GDPNet: Overview

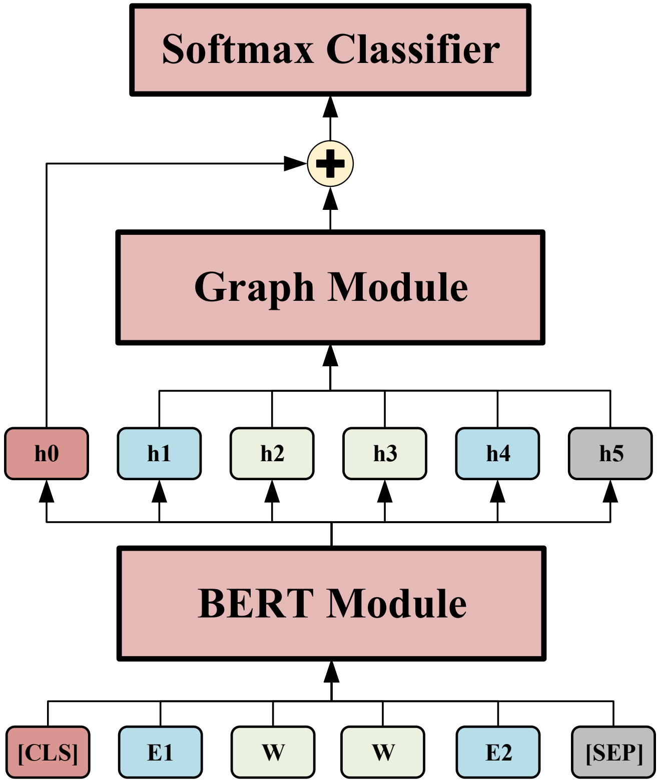


GDPNet: BERT Module

BERT Module

- *Input*: Given a sequence X with T tokens, we map X to a BERT input sequence $X_{input} = \{x_0, x_1, x_2, \dots, x_{T+1}\}$. Here, x_0 denotes the ``[CLS]'' token whi.ch represents the start of sequence $X$, and $x_{T+1}$ is the ``[SEP]'' token which represents the end of the sequence.

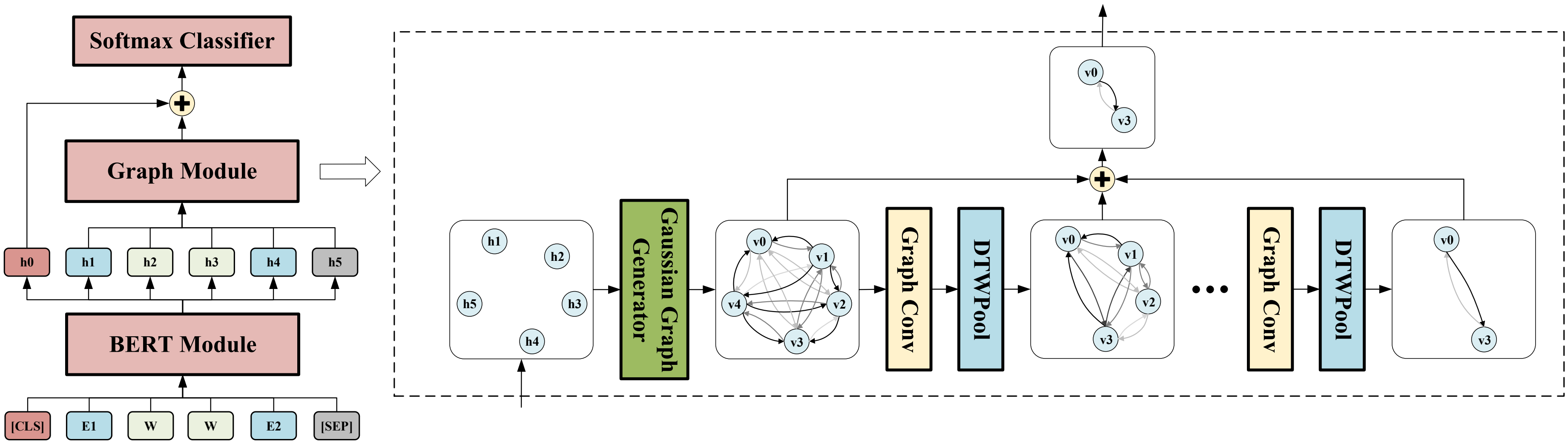
- *Output*: The corresponding token representations from BERT are denoted by $\mathrm{H} = \{h_0, h_1, h_2, \dots, h_{T+1}\}$



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Agency for Science, Technology and Research SINGAPORE

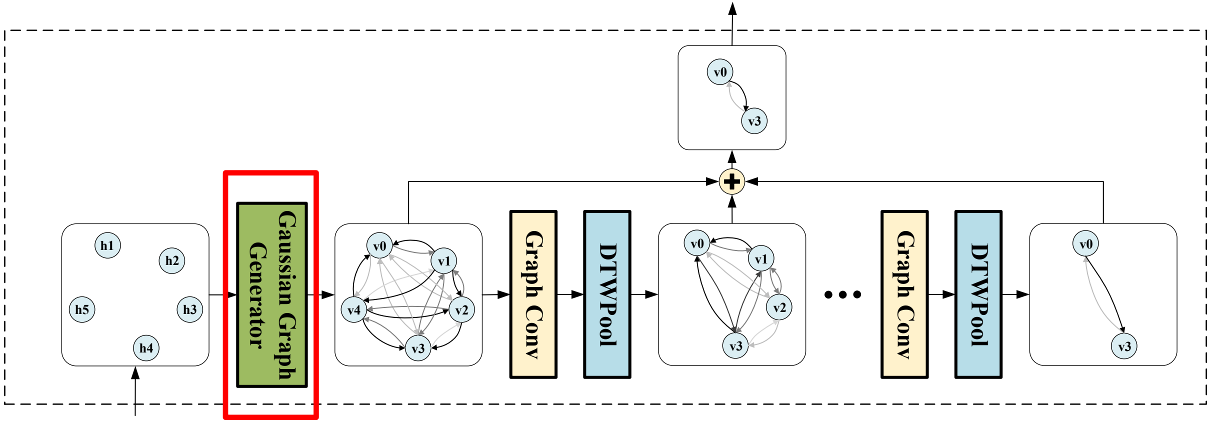
GDPNet: Graph Module



Graph Module

- *Input*: Token representations from BERT $\mathrm{H} = \{h_1, h_2, \ldots, h_{T+1}\}$.
- *Output*: Final task-specific graph V

GDPNet: Graph Module --Gaussian Graph Generator



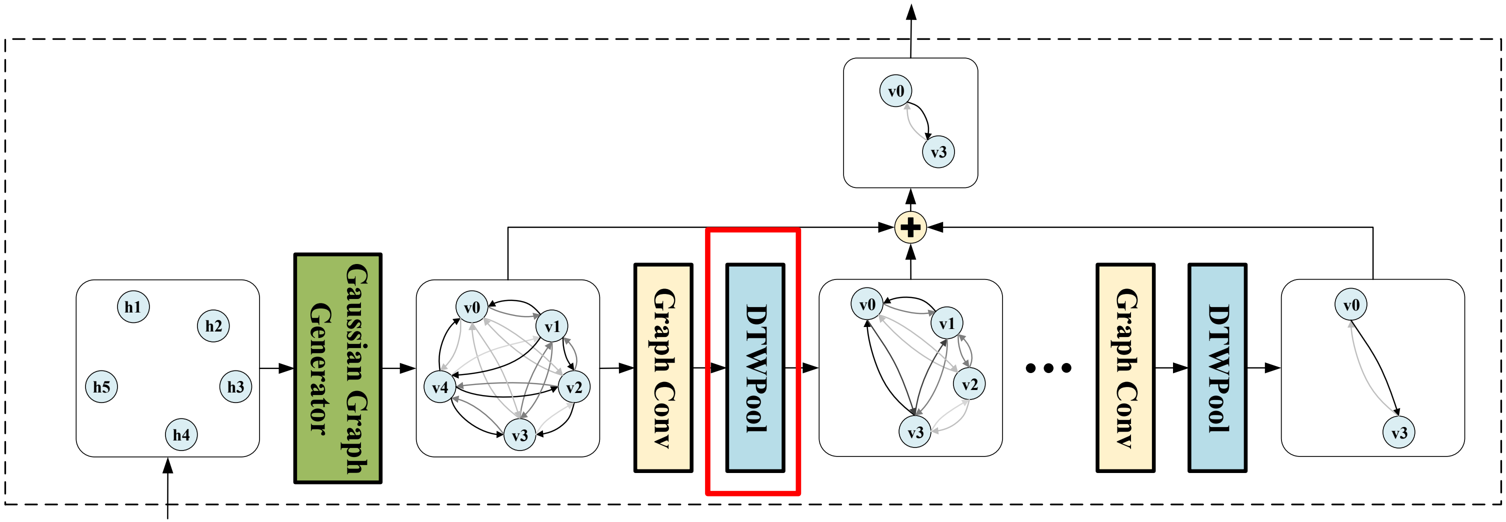
Why: The purpose of the multi-view graph is to capture all possible relations between tokens, so we encourage message propagation between token representations with large semantic differences.

How: Following previous work, we use Gaussian Distribution and KL divergence to model asymmetric relations between two arbitrary tokens.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Agency for Science, Technology and Research SINGAPORE

GDPNet: Graph Module --DTWPool

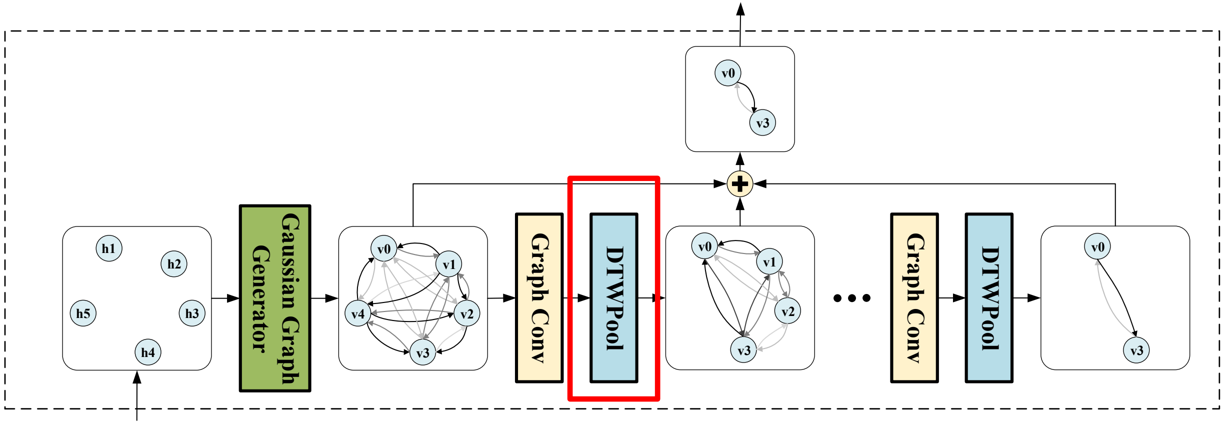


Why do we use Graph Pooling? Graph Pooling is to obtain a task-specific graph from a large graph with sparse informative tokens (nodes).

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Agency for Science, Technology and Research SINGAPORE

GDPNet: Graph Module --DTWPool



Why do we use DTWPool? After initialization, the latent multi-view graph is very large, if the input sequence is long. It is difficult for the RE model to focus on the indicative tokens for relation prediction.

How do we use DTWPool? DTWloss is integrated into our model to preserve the nodes adaptively when we delete a number of nodes in a large graph.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Agency for Science, Technology and Research SINGAPORE

Experiments: Dialogue-level Relation Extraction

- Dataset
  - **DialogRE** is the first human-annotated dialogue-level RE dataset.

- Baseline Models and Experimental Setup
  - **BERTs**: speaker-aware modification of BERT, and it achieves best performance on dialogue-level Relation Extraction.
  - We also include popular baseline models: **CNN, LSTM, BiLSTM and BERT** models.
  - We use the same input format and hyperparameter settings as in BERTs.

Experiments: Dialogue-level Relation Extraction

- Results on DialogRE

Model	Dev		Test	
	F1(σ)	F1c(σ)	F1(σ)	F1c(σ)
CNN (Lawrence et al. 1997)	46.1(0.7)	43.7(0.5)	48.0(1.5)	45.0(1.4)
LSTM (Hochreiter and Schmidhuber 1997)	46.7(1.1)	44.2(0.8)	47.4(0.6)	44.9(0.7)
BiLSTM (Graves and Schmidhuber 2005)	48.1(1.0)	44.3(1.3)	48.6(1.0)	45.0(1.3)
BERT (Devlin et al. 2019)	60.6(1.2)	55.4(0.9)	58.5(2.0)	53.2(1.6)
BERTs (Yu et al. 2020)	63.0(1.5)	57.3(1.2)	61.2(0.9)	55.4(0.9)
GDPNet (ours)	**67.1**(1.0)	**61.5**(0.8)	**64.9**(1.1)	**60.1**(0.9)

GDPNet achieves the best performance on dialogue-level RE.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

SINGAPORE Agency for Science, Technology and Research

Experiments: Dialogue-level Relation Extraction

- Ablation Study

Model	$F1(\sigma)$	$F1c(\sigma)$
GDPNet	**64.9** (1.1)	**60.1** (0.9)
with Homogeneous GGG	63.5 (0.7)	58.4 (0.6)
w/o GGG	62.1 (1.6)	58.1 (1.1)
w/o DTWloss	63.4 (1.4)	58.6 (1.3)
w/o DTWPool	48.9 (1.1)	22.4 (1.0)
w/o GGG & DTWPool	48.2 (1.4)	21.8 (1.0)

Ablation study shows the effectiveness of the two main components in GDPNet, i.e., Gaussian Graph Generator and Dynamic Time Warping Pooling.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

SINGAPORE Agency for Science, Technology and Research

Experiments: Sentence-level Relation Extraction

- Dataset
  - **TACRED** is a widely used large-scale sentence-level relation extraction dataset.
  - **TACRED-Revisit** dataset, released recently, corrects the wrong labels in the development and test sets of TACRED.

- Baseline Models and Experimental Setup
  - **SpanBERT** is the best performing sentence-level Relation Extraction model without incorporating any external knowledge and parser.
  - We also include **RNN- and graph-based models**.
  - We use the same input format and hyperparameter settings as in SpanBERT.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE
Agency for Science, Technology and Research SINGAPORE

Experiments: Sentence-level Relation Extraction

➢ Results on TACRED

Model	TACRED			TACRED-Revisit		
	Pr	Re	$F1$	Pr	Re	$F1$
LSTM (Zhang et al. 2017)	65.7	59.9	62.7	71.5*	69.7*	70.6*
PA-LSTM (Zhang et al. 2017)	65.7	64.5	65.1	74.5*	74.1*	74.3*
C-AGGCN (Guo, Zhang, and Lu 2019)	73.1	60.9	68.2	77.7*	73.4*	75.5*
LST-AGCN (Sun et al. 2020)	-	-	68.8	-	-	-
SpanBERT (Joshi et al. 2020)	70.8	70.9	**70.8**	75.7*	80.7*	78.0*
GDPNet (Our model)	72.0	69.0	70.5	79.4	81.0	**80.2**
KnowBERT (Peters et al. 2019)	71.6	71.4	**71.5**	-	-	79.3

Without external resources, GDPNet achieves comparable performance with the state-of-the-arts on TACRED, and better results on TACRED-Revisit.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Agency for Science, Technology and Research SINGAPORE

Experiments: Sentence-level Relation Extraction

Quantitative Analysis

Type of tokens	DialogRE	TACRED
All tokens	15.6	66.3
Non-repetitive tokens	23.5	67.6
Repetitive tokens	10.0	58.1
Trigger tokens	32.1	-

DialogRE provides manually annotated trigger tokens that are indicative to the relation type. DTWPool selects 32.1% of trigger tokens, given that only 15.6% of tokens are selected among all tokens.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Agency for Science, Technology and Research SINGAPORE

Conclusion

- We propose GDPNet for relation extraction. GDPNet is designed to find indicative words from long sequences (e.g. dialogues) for effective relation extraction.
- GDPNet achieves the best performance on dialogue-level RE and comparable performance on sentence-level RE.
- We show there is a great potential of this mechanism in dealing with long sequences.
- To evaluate the effectiveness of this mechanism on other tasks is part of our future work.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Agency for Science, Technology and Research SINGAPORE

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Agency for Science, Technology and Research
SINGAPORE
CREATING GROWTH, ENHANCING LIVES

Thank You!

Code at: https://github.com/XueFuzhao/GDPNet