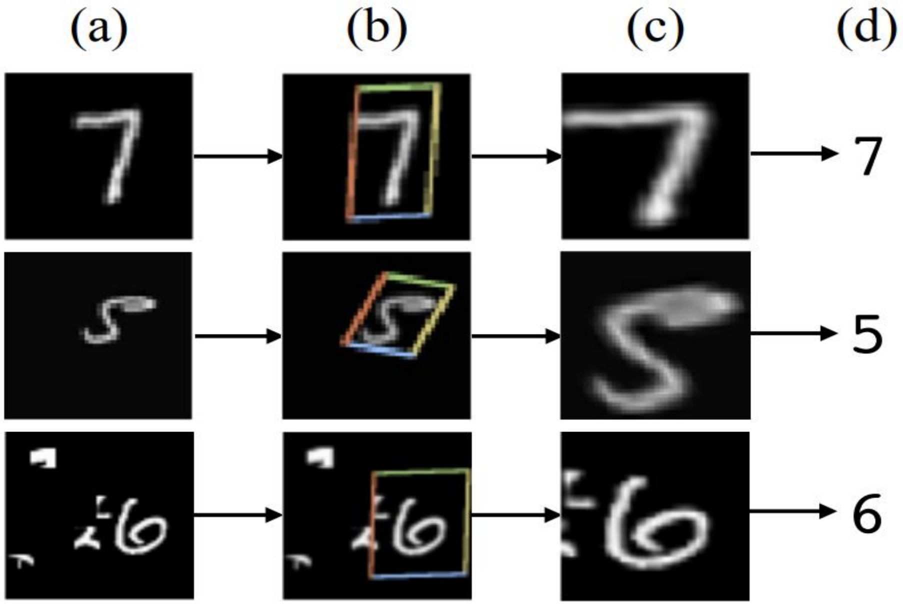
Research

Graph Transformer Networks (NeurIPS 2019)

Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, Hyunwoo J. Kim

Spatial Transformer Networks

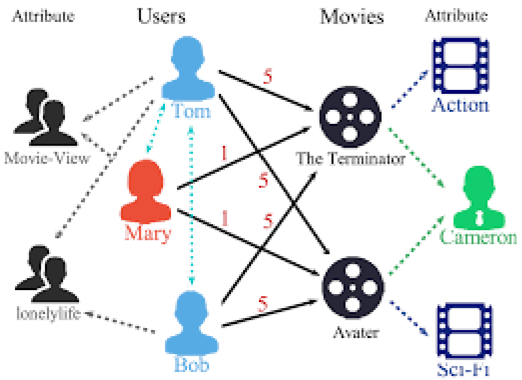


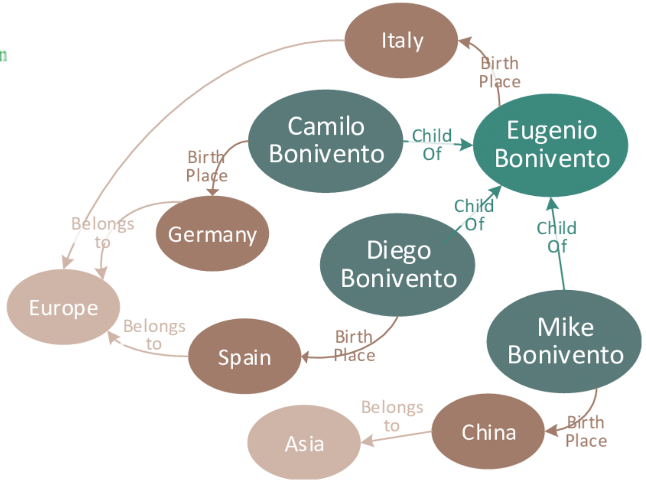
A spatial transformer that crops out and scale-normalizes the appropriate region can simplify the subsequent classification task, and lead to superior classification performance

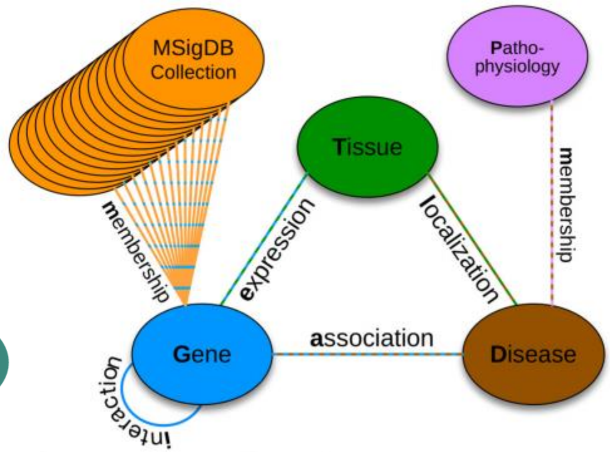
Spatial Transformer Networks, NIPS 2015, M Jaderberg et al.

Heterogeneous Graph

Heterogeneous graph is a graph which contains different types of nodes and edges



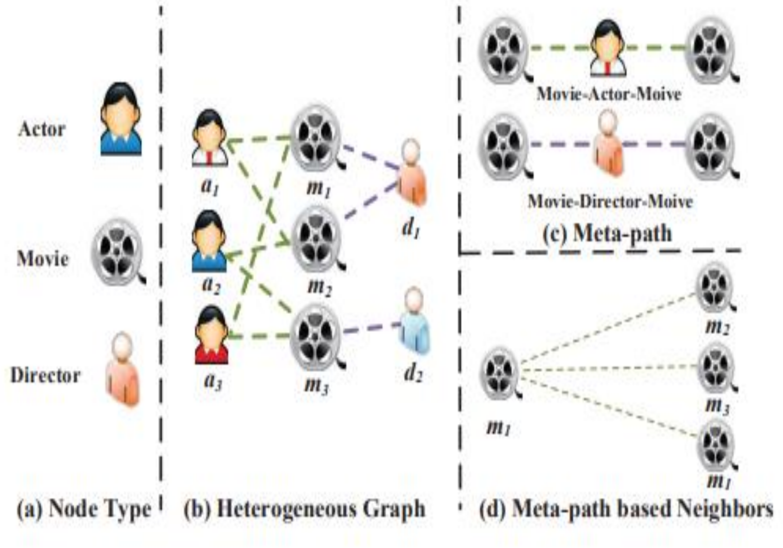




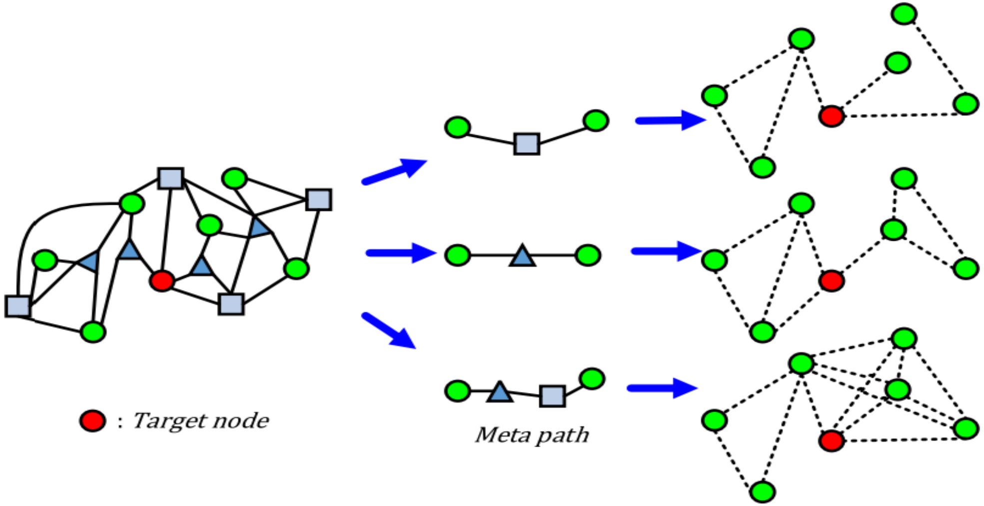
Meta-path

A meta-path is a path consisting of a sequence of relations defined between different object types

Datasets	Examples of meta-path
BlogCatalog	$\text{Blog} \xrightarrow{has} \text{Tag} \xrightarrow{has^{-1}} \text{Blog}$ $\text{Blog} \xrightarrow{written\_by} \text{User} \xrightarrow{written\_by^{-1}} \text{Blog}$ $\text{Blog} \xrightarrow{written\_by} \text{User} \xrightarrow{friend} \text{User} \xrightarrow{written\_by^{-1}} \text{Blog}$
DBLP	$\text{Paper} \xrightarrow{has} \text{Term} \xrightarrow{has^{-1}} \text{Paper}$ $\text{Paper} \xrightarrow{written\_by} \text{Author} \xrightarrow{written\_by^{-1}} \text{Paper}$ $\text{Paper} \xrightarrow{written\_by} \text{Author} \xrightarrow{written\_by^{-1}} \text{Paper} \xrightarrow{has} \text{Term} \xrightarrow{has^{-1}} \text{Paper}$
Chemical Compound	$\text{Compound} \xrightarrow{bind} \text{Gene} \xrightarrow{PPI} \text{Gene} \xrightarrow{bind^{-1}} \text{Compound}$ $\text{Compound} \xrightarrow{treat} \text{Disease} \xrightarrow{cause^{-1}} \text{Gene} \xrightarrow{bind^{-1}} \text{Compound}$ $\text{Compound} \xrightarrow{bind} \text{Gene} \xrightarrow{has} \text{Pathway} \xrightarrow{has^{-1}} \text{Gene} \xrightarrow{bind^{-1}} \text{Compound}$

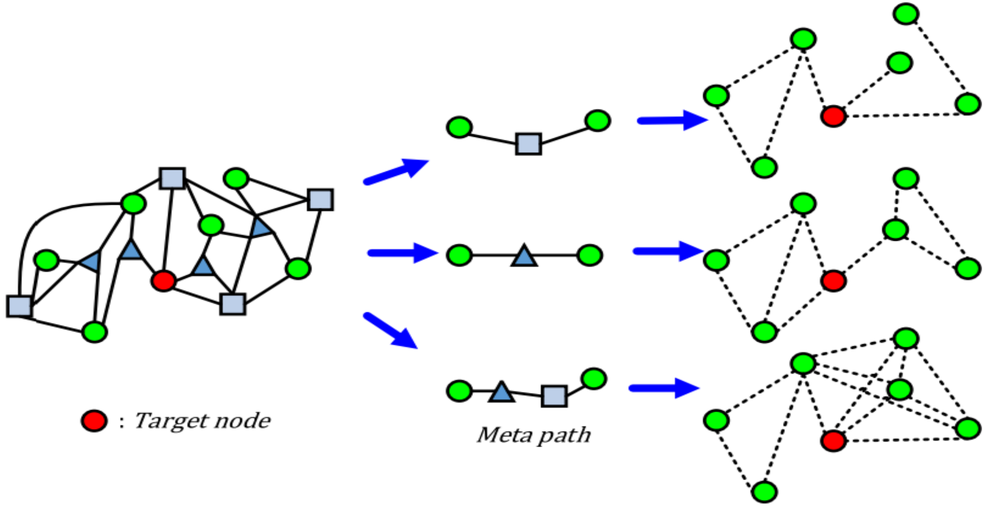


Transform a Graph into new Graphs using Meta-Paths



Previous works about graph neural networks leveraged useful meta-path which selected **manually by domain experts**.

Transform a Graph into new Graphs using Meta-Paths



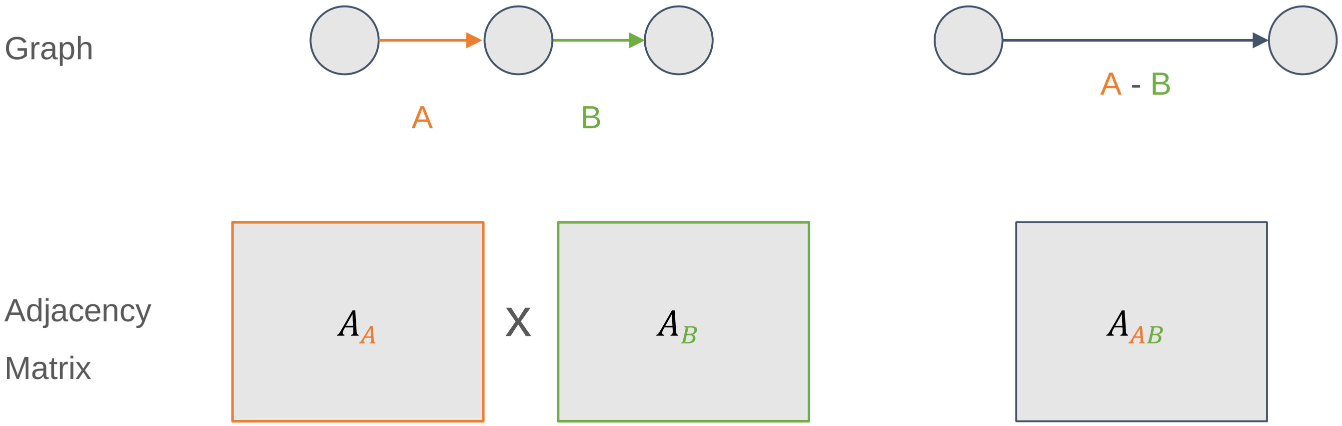
Previous works about graph neural networks leveraged useful meta-path which selected **manually by domain experts**.

Can model **learn to transform** an original graph into a new graph which involves only useful connections for task?

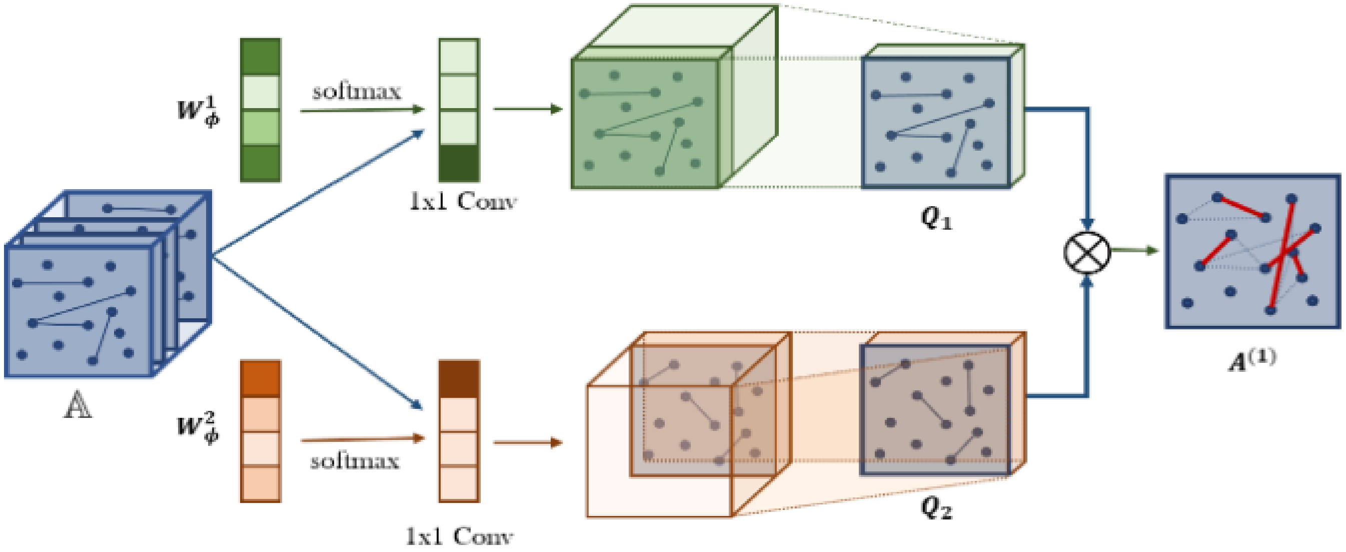
Deep Collective Classification in Heterogeneous Information Networks, WWW 2018, Yizhou Zhang et al.

Multiplication of Adjacency Matrices for Generating Meta-Paths

Edge Type: A,B

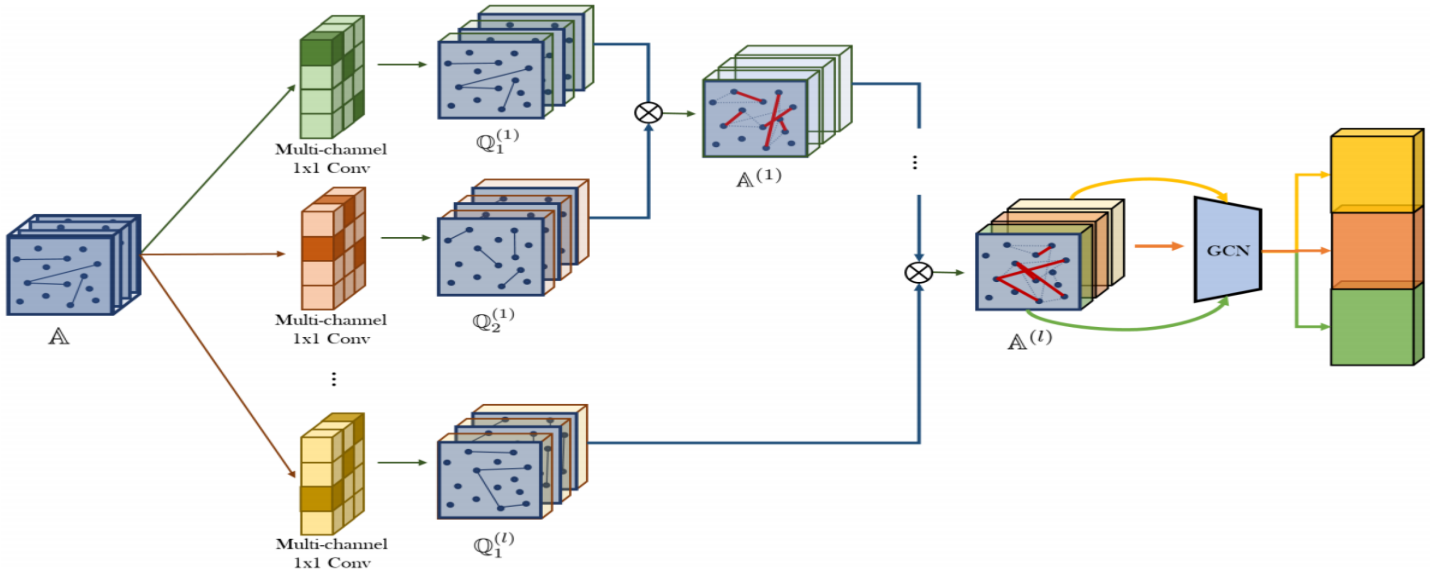


Graph Transformer Layer



Graph Transformer Layer (GTL) softly **selects adjacency matrices** (edge types) from the set of adjacency matrices and **generate** a new meta-path graph via the **matrix multiplication** of two selected adjacency matrices.

Graph Transformer Networks



Graph Transformer Networks (GTNs) learn to **generate a set of new meta-path adjacency matrices** using GT layers and perform graph convolution as in GCNs on the new graph structures.

Q1) Are the new graph structures generated by GTN effective for learning node representation?

Dataset	# Nodes	# Edges	# Edge type	# Features	# Training	# Validation	# Test
DBLP	18405	67946	4	334	800	400	2857
ACM	8994	25922	4	1902	600	300	2125
IMDB	12624	37288	4	1256	300	300	2339

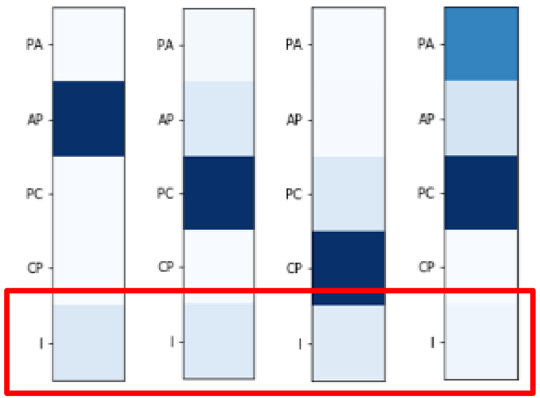
	DeepWalk	metapath2vec	GCN	GAT	HAN	GTN_{-I}	**GTN (proposed)**
DBLP	63.18	85.53	87.30	93.71	92.83	93.91	**94.18**
ACM	67.42	87.61	91.60	92.33	90.96	91.13	**92.68**
IMDB	32.08	35.21	56.89	58.14	56.77	52.33	**60.92**

- Graph Transformer Networks (GTNs) achieves the **highest performance** on all the datasets against all network embedding methods and graph neural network methods
- GTNs performs **better** than HAN which uses the **pre-defined meta paths**.

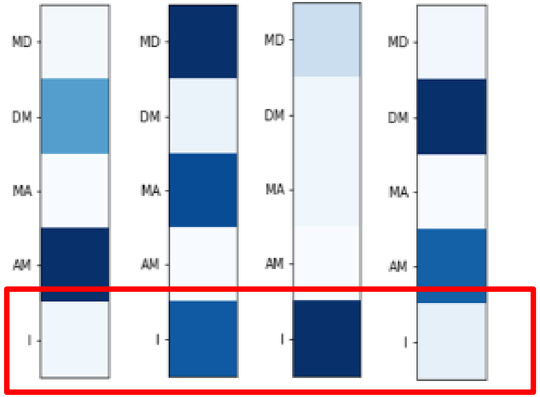
Experiments

Q2) Can GTN adaptively produce a variable length of meta-paths depending on datasets?

The attention score of adjacency matrix (edge type) from each Graph Transformer Layer



(a)
DBLP



(b)
IMDB

- By assigning **higher** attention scores to **the identity matrix (I)**, GTN tries to stick to the shorter meta-paths even in the deeper layer.
- GTN has ability to **adaptively learn most effective meta-path length** depending on the dataset

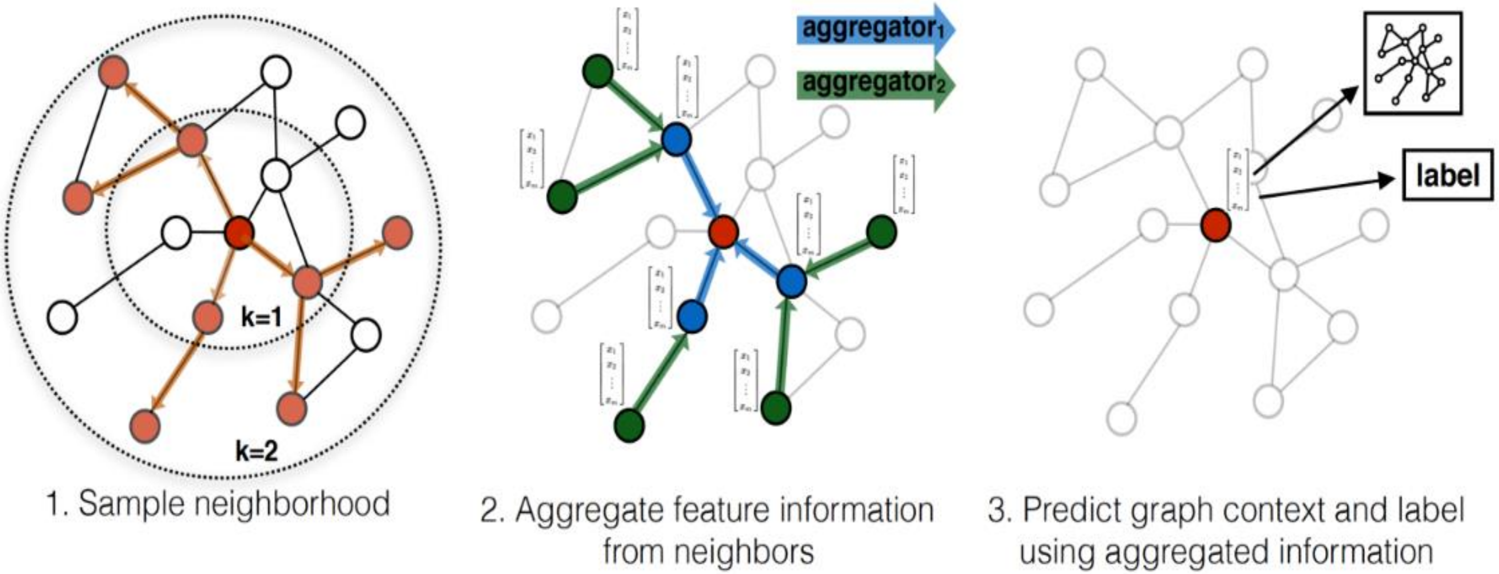
Q3) How can we interpret the importance of each meta-path from the adjacency matrix generated by GTNs?

Comparison with predefined paths and top-ranked meta-paths by GTNs

Dataset	Predefined Meta-path	Meta-path learnt by GTNs Top 3 (between target nodes)	Top 3 (all)
DBLP	APCPA, APA	APCPA, APAPA, APA	APCPC, APCPA, PC
ACM	PAP, PSP	PAP, PSP	PAPA, APA, PAPS
IMDB	MAM, MDM	MDM, MAM, MDMDM	DM, AM, MDM

- The predefined meta-paths by domain knowledge are consistently top-ranked by GTNs as well.
- Our GTNs are capable to learn the importance of meta-paths for tasks

Scalability



Inductive representation learning on large graphs, NIPS 2017, Hamilton Will et al.