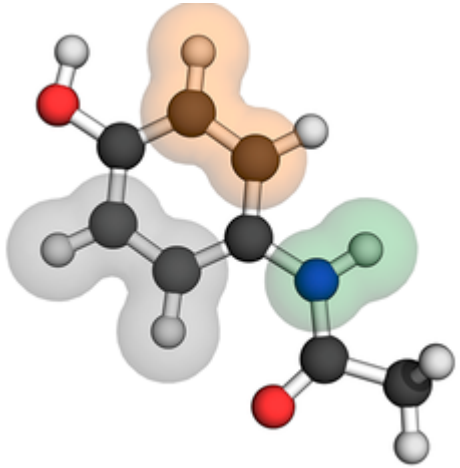
ASGN: an Active Semi-supervised Graph Neural Network for Molecular Property Prediction

Zhongkai Hao, Chengqiang Lu, Zhenya Huang, Hao Wang,
Zheyuan Hu, Qi Liu, Enhong Chen, Cheekong Lee

University of Science and Technology of China

Introduction

- Our task: Molecular property prediction



Input: Molecule

Properties:
U0 (Atomization energy at 0K)
U (Atomization energy at room temperature)
G (Free energy of atomization)
HOMO
LUMO

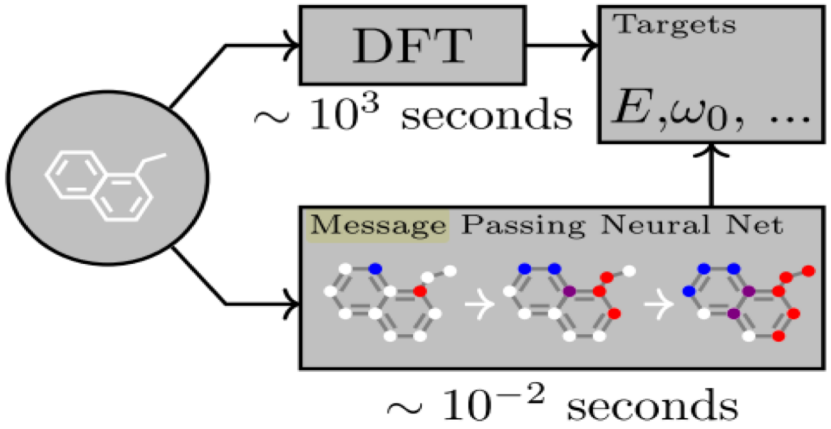
.

Output: Properties

.

- Applications: Drug discovery, material engineering…

Introduction



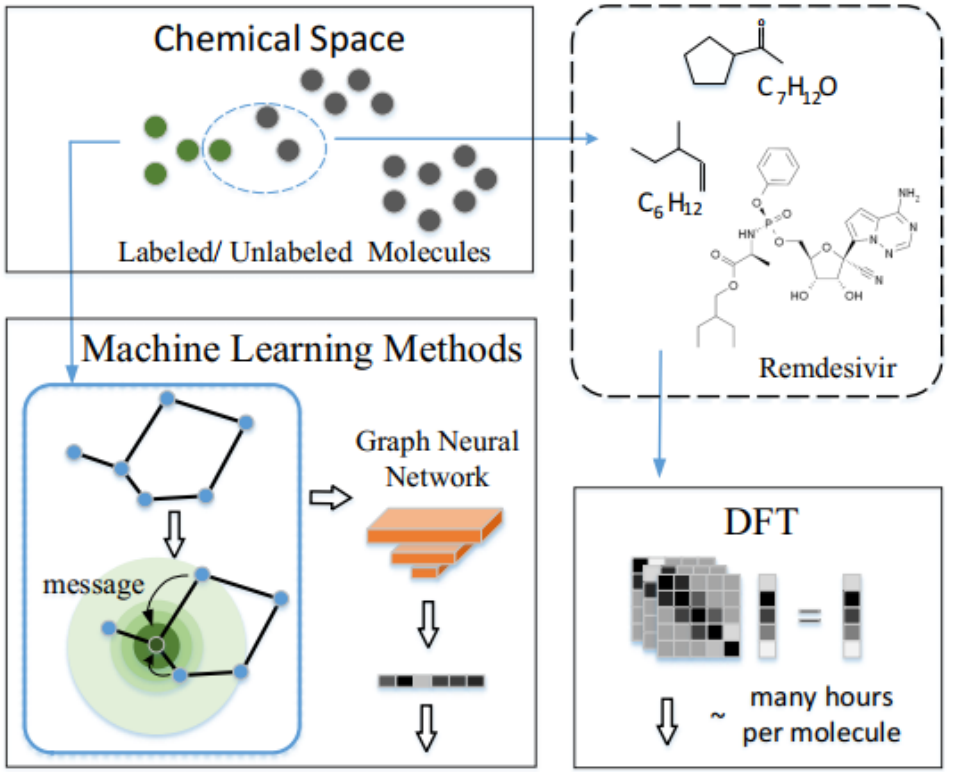
- Measure properties by experiments
- Density Functional Theory
- Modern: Machine learning methods

- A molecule as a graph($G = (V, E)$)
- Pass it to a message passing Graph Neural Networks
- Get the result after 10^{-2} seconds

Introduction

- ML model is data hungry, requires many **labelled** data
- Unlabelled data (molecular graph) is everywhere
- Labelling is expensive
- Our goal: label efficient model

  $f: G \rightarrow R^n$
- Our Solution: **Active semi-supervised learning**



Property	Value
HOMO	-0.38eV
LUMO	0.08eV
U	-196.43eV

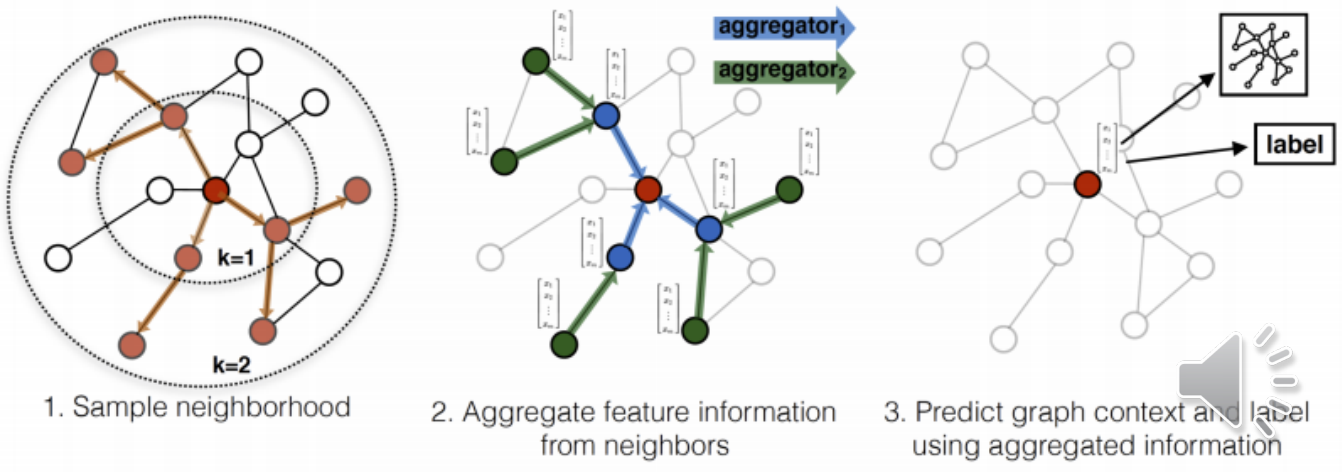
Preliminaries—GNN for molecular property prediction

- Pass message from nodes to nodes
- Aggregate node to get the graph representation

$$z_i^{l+1} = \sigma(\boldsymbol{W}^l \cdot \mathrm{AGG}(z_i^l, \{\boldsymbol{e}(v_i, v_j) : v_j \in \boldsymbol{N}(v_i)\})),$$

$$z_{\mathcal{G}} = \mathrm{Pool}(\{z_i^L : v_i \in \mathcal{V}\}).$$

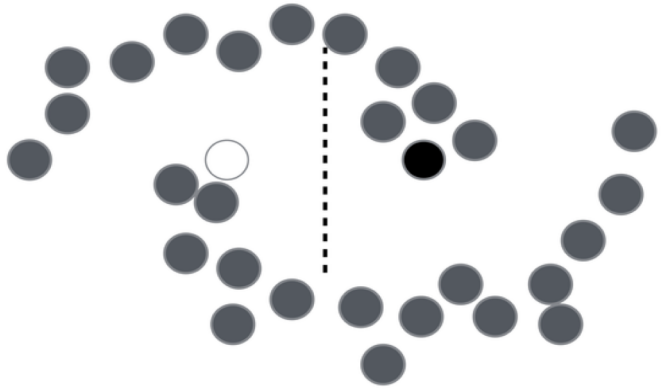
GraphSAGE: A popular MPNN



Related Work—Semi-supervised Learning

- Number of labeled data ≪ unlabeled data
- How can we make use of unlabeled data ?

- Create pseudo labels and predict them!

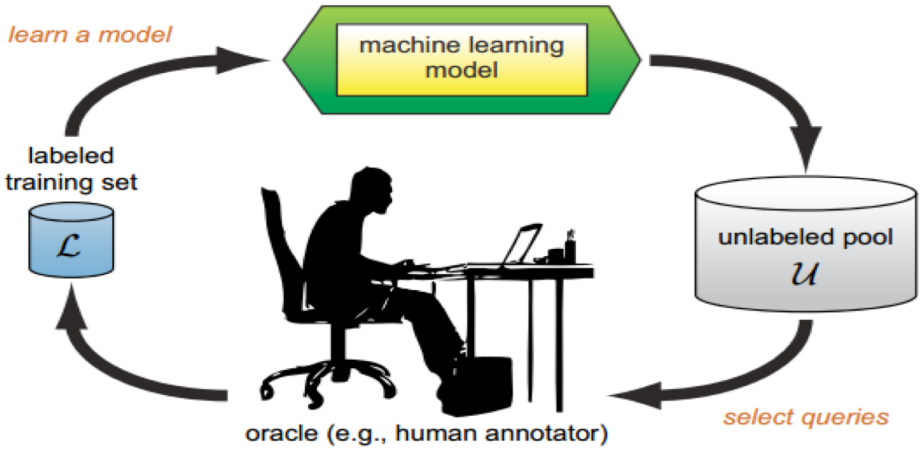


The influence of unlabeled data

Related Work—Active Learning

- Active learning is to improve the value of these labels
- Choose data that is helpful to the model and retrain the model
- Solution: most **representative** and **diversified** subset in the dataset

Framework of active learning.

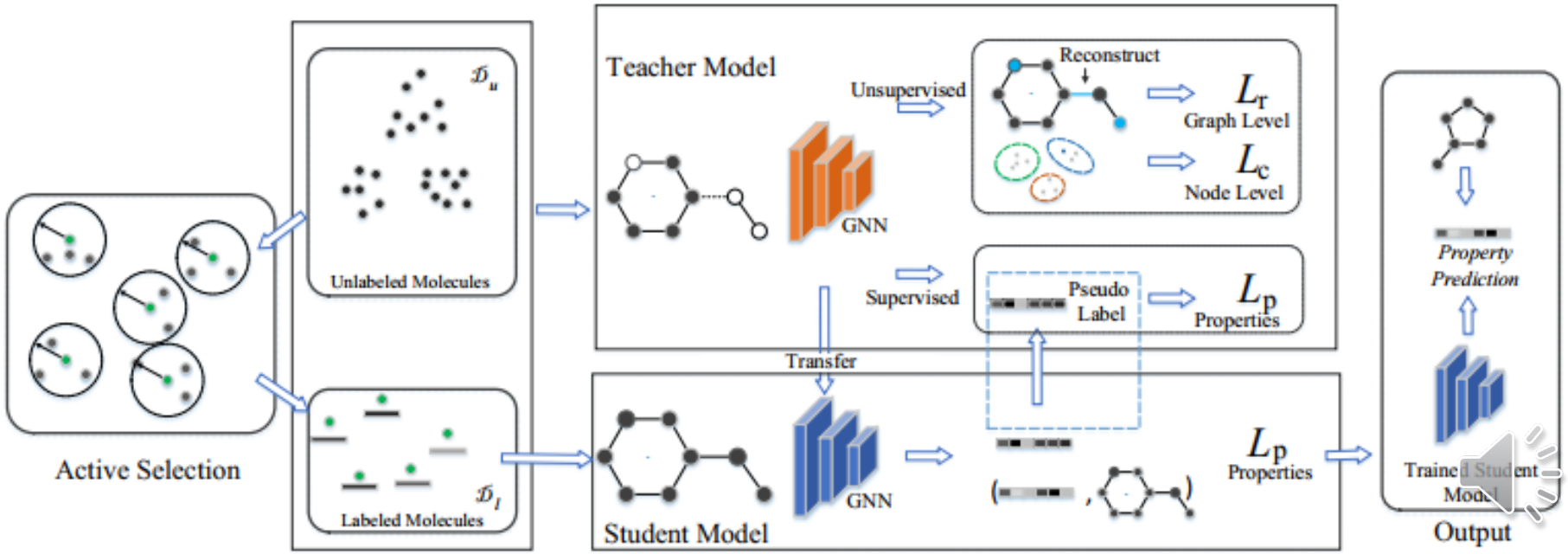


Challenges

- Data structure of molecules is different from traditional images/text/…
- Few works on semi-supervised learning of molecules
- Low training efficiency because of the imbalance data

Model Framework

- Two GNN, a teacher and a student model
- Train the teacher with semi-supervised learning
- Train the student with fully supervised learning for downstream property prediction

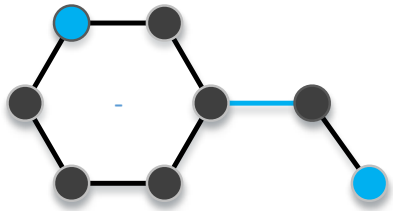


Teacher Model

- Local(node) level pseudo labels—reconstruction
- We believe a good property predictor is able to recover the atom itself from its embedding
- A loss function to reconstruct atom and their distance

$$\mathcal{L}_r = -\mathbb{E}_{v_i \sim \mathcal{V}} \left[\sum_{m=1}^{K_n} f_{im} \log(g_{\theta_n}(z_i)) \right] - \mathbb{E}_{e_{ij} \sim \mathcal{E}} \left[\sum_{m=1}^{K_e} e_{ijm} \log(g_{\theta_e}(z_i, z_j)) \right],$$

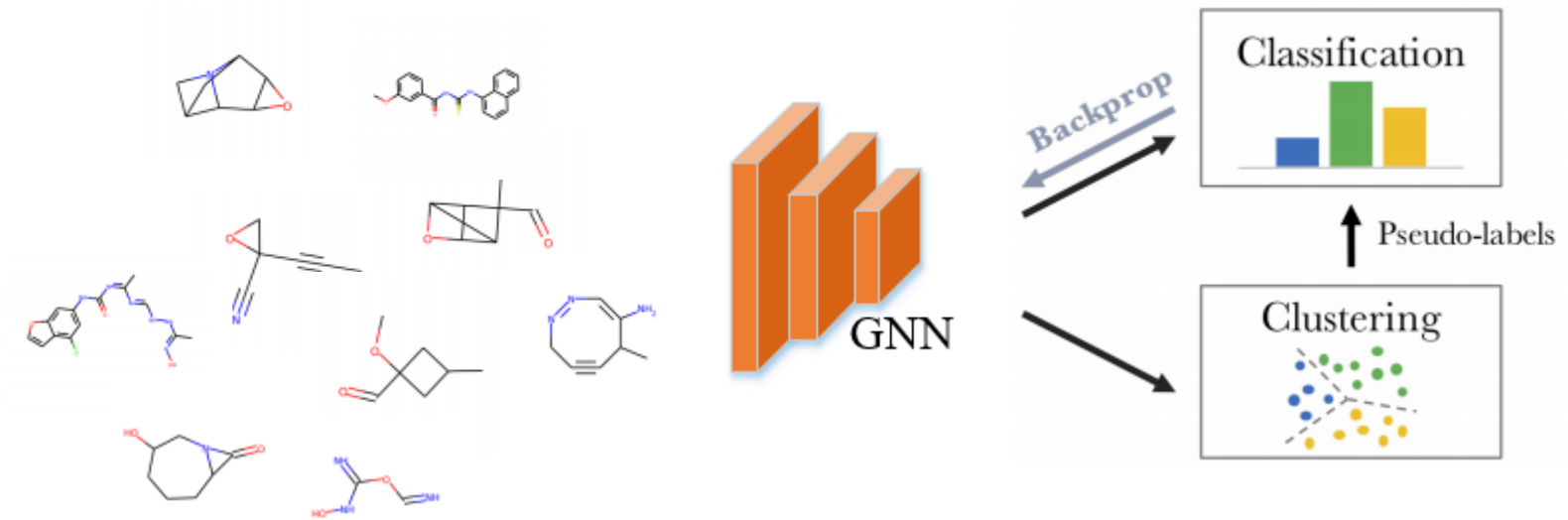
GNN



Sample and reconstruct

Teacher Model

- Global level pseudo labels—clustering loss
- Implicit clustering via optimal transport
- Predict these clusters and repeat iteratively



Teacher model

- Summary of the teacher model
- Add these three loss terms to guide its optimization

(1).property loss

(2).reconstruction loss

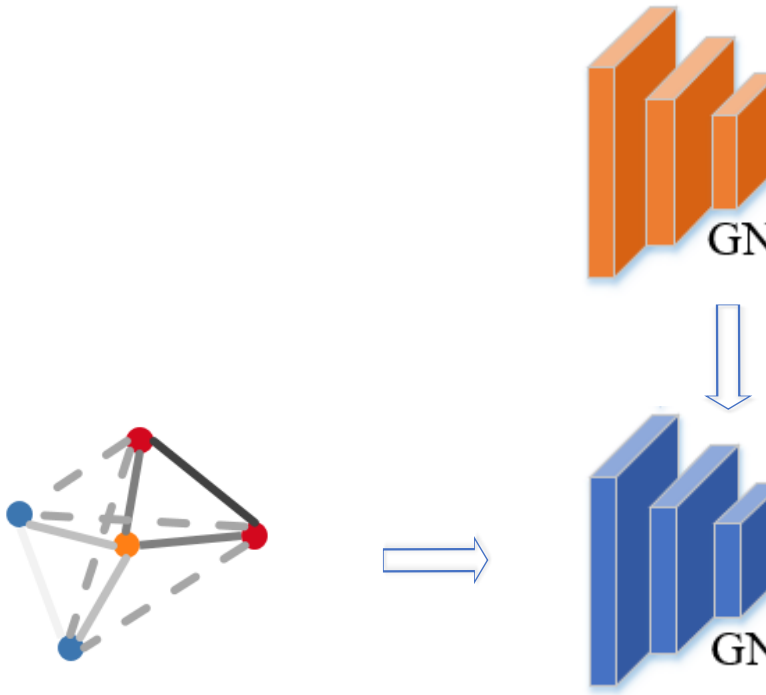
(3).clustering loss

$$\mathcal{L}_t = \sum_{\mathcal{G} \in \mathcal{D}_l} \mathcal{L}_p + \sum_{\mathcal{G} \in \mathcal{D}_u \cup \mathcal{D}_l} \mathcal{L}_r + \sum_{\mathcal{G} \in \mathcal{D}_u \cup \mathcal{D}_l} \mathcal{L}_c .$$

D_l:labeled data D_u:unlabeled data

Student model

- Weight transfer from the teacher model
- Fine tune on property prediction task
- Accelerate convergence and alleviate loss conflict

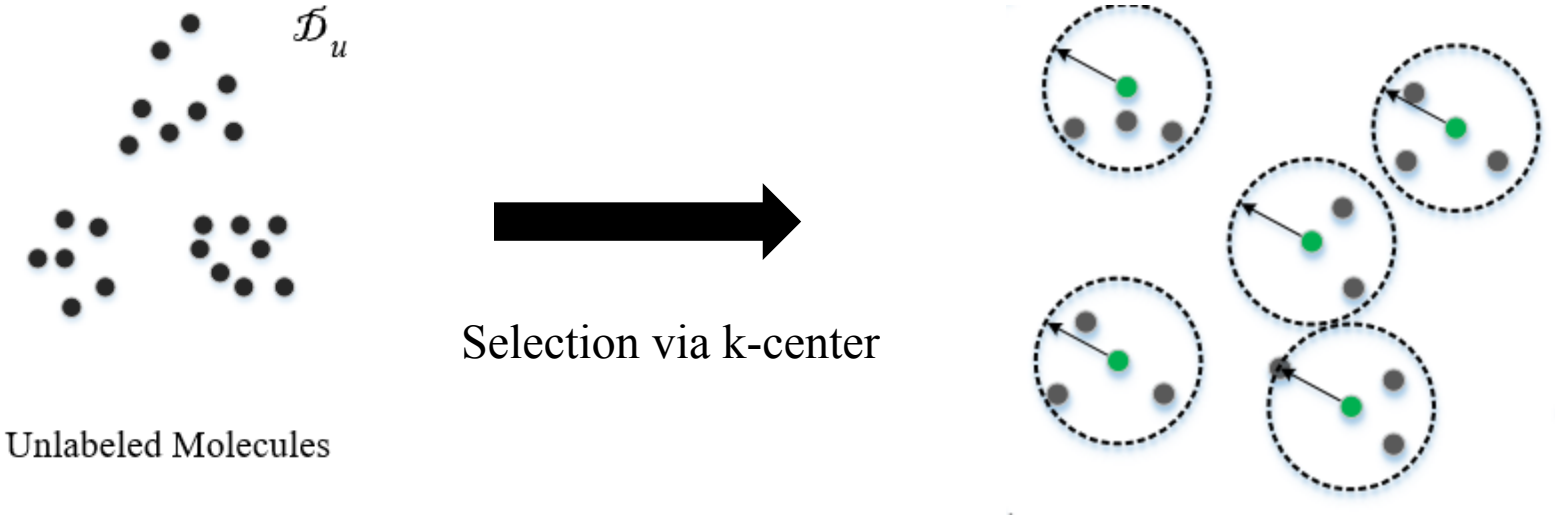


Property	Value
HOMO	-0.38eV
LUMO	0.08eV
U	-196.43eV

Active Data Selection

- Choose most informative data
- K center to choose one molecule from one cluster
- Add them into the labeled dataset

- Repeat the process until label budget is used up



Experiments

- Datasets

(1) QM9: 130,000 molecules, <9 heavy atoms

(2) OPV: 100,000 medium sized molecules

- Properties (All calculated by DFT)

(1) QM9:

U_0	U	G	H	C_v	HOMO	LUMO	gap	ZPVE	R^2	μ	α
eV	eV	eV	eV	Cal/MolK	eV	eV	eV	eV	Bohr^2	Debye	Bohr^3

(2) OPV:

HOMO	LUMO
Hatree	

Experiments

- Effectiveness, compare error on test dataset
- Baselines

(1).Supervised

(2).Mean-teachers

(3).InfoGraph

Experiments

- Results

Properties	U_0	U	G	H	C_v	HOMO	LUMO	gap	ZPVE	R^2	μ	α
Unit	eV	eV	eV	eV	Cal/MolK	eV	eV	eV	eV	Bohr^2	Debye	Bohr^3
Supervised	0.3204	0.2934	0.2948	0.2722	0.2368	0.1632	0.1686	0.2475	0.0007	10.05	0.3201	0.5792
Mean-Teachers	0.3717	0.2730	0.2535	0.2150	0.2036	0.1605	0.1686	0.2394	0.00054	5.22	0.3488	0.5792
InfoGraph	0.1410	0.1702	0.1592	0.1552	0.1965	0.1605	0.1659	0.2421	0.00036	4.92	0.3168	0.5444
ASGN (Ours)	**0.0562**	**0.0594**	**0.0560**	**0.0583**	**0.0984**	**0.1190**	**0.1061**	**0.2012**	**0.00017**	**1.38**	**0.1947**	**0.2818**

Results on QM9

Property	HOMO	LUMO
Unit	Hatree	
Supervised	0.080	0.078
Mean-Teacher	0.078	0.075
InfoGraph	0.077	0.076
ASGN (Ours)	**0.059**	**0.057**

Results on OPV

Experiments

- Efficiency, the label efficiency at a certain error
- Baselines:

(1).Random

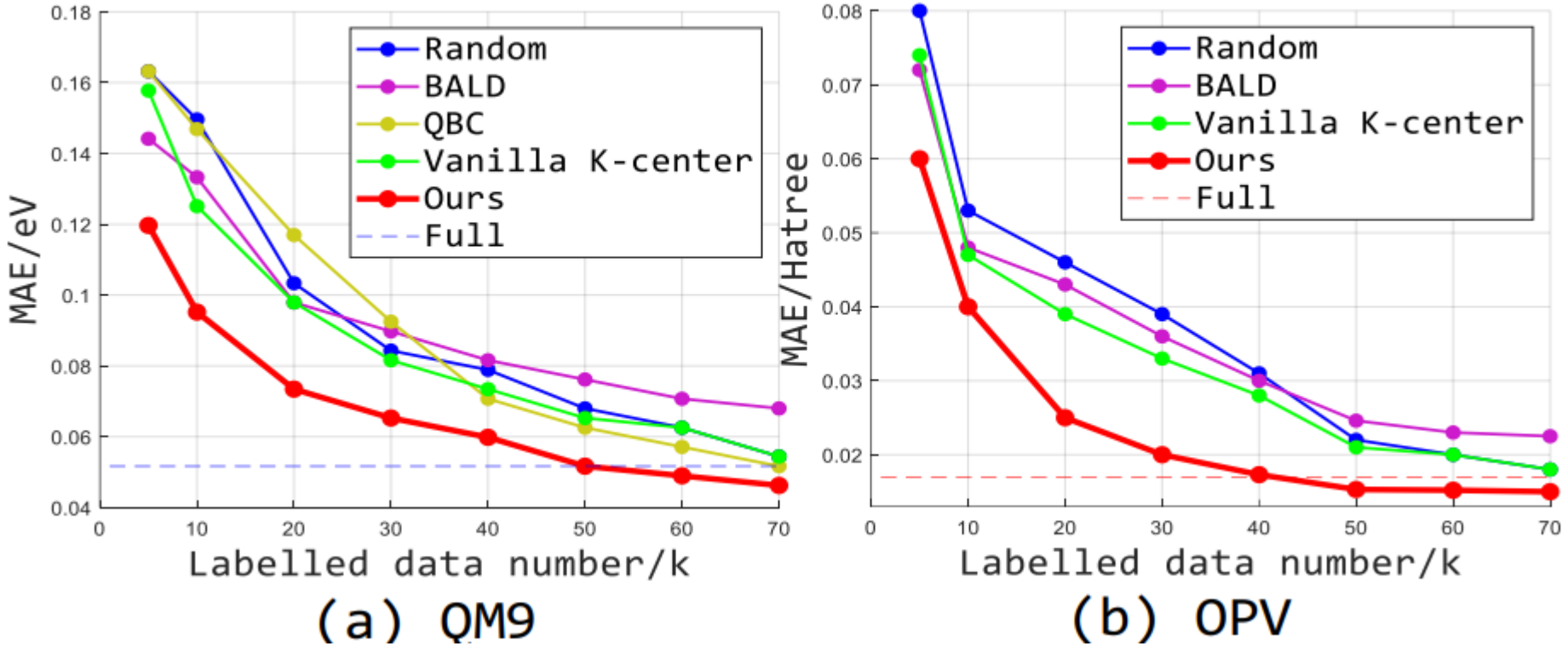
(2).Query by Committee

(3).Deep Bayesian Active Learning

(4).Vanilla K-center

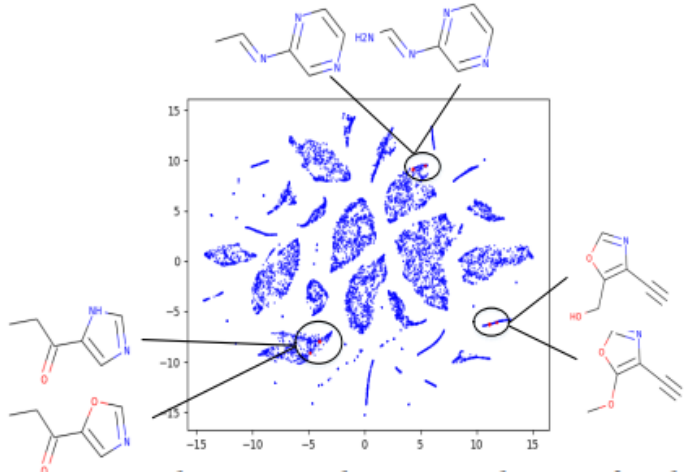
Experiments

- Results



Experiments

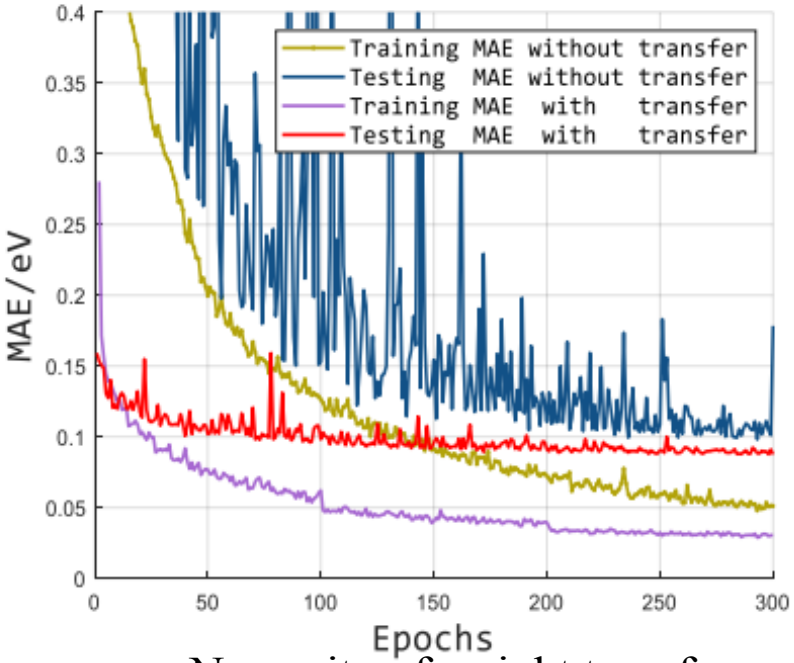
- Ablation Study
- Why using two models (a teacher and a student)
- Why transferring weight from the teacher to the student
- Visualization experiment



Visualization

Name/Dataset	Homo(QM9)			Homo(OPV)		
Unit	eV			Hatree		
Number of data	5k	10k	50k	5k	10k	50k
ASGN-T	0.1668	0.1523	0.0682	0.080	0.053	0.020
ASGN-S	0.1632	0.1252	0.0653	0.076	0.049	0.019
ASGN	**0.1190**	**0.0951**	**0.0517**	**0.060**	**0.039**	**0.015**

Necessity of teacher and student



Necessity of weight transfer

Many thanks!