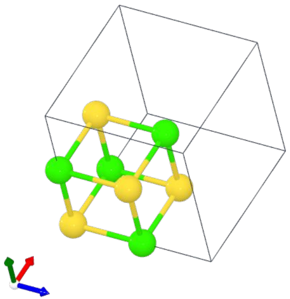
CrysGNN: Distilling pre-trained knowledge to enhance property prediction for crystalline materials.

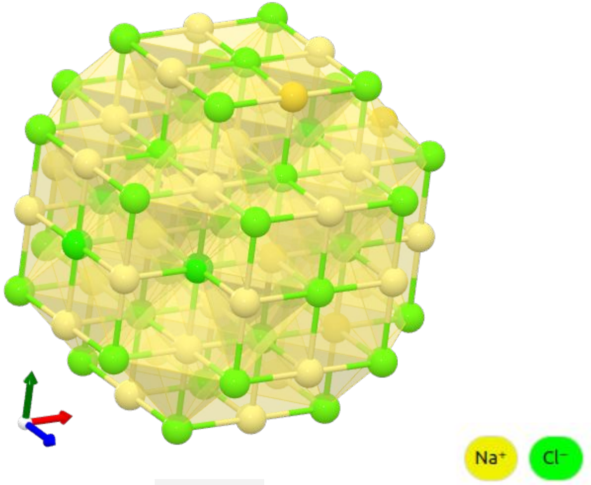
Kishalay Das[1], Bidisha Samanta[1], Pawan Goyal[1], Seung-Cheol Lee[2],
Satadeep Bhattacharjee[2],Niloy Ganguly[1,3]

[1]Indian Institute of Technology Kharagpur, India
[2]Indo Korea Science and Technology Center, Bangalore, India.
[3]L3S, Leibniz University of Hannover, Germany.

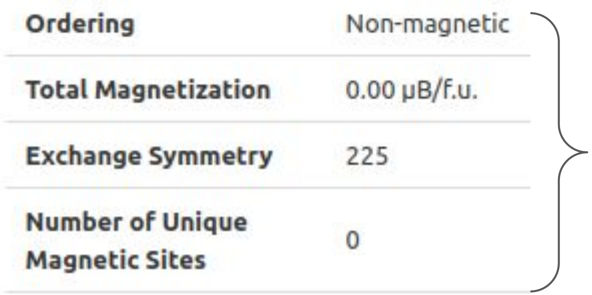
Crystalline Material



Unit Cell

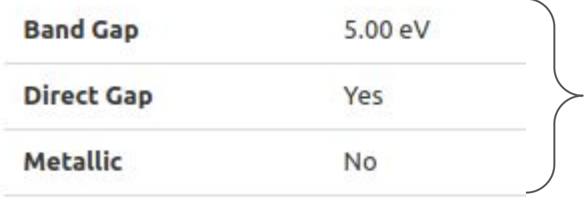


Periodic Structure



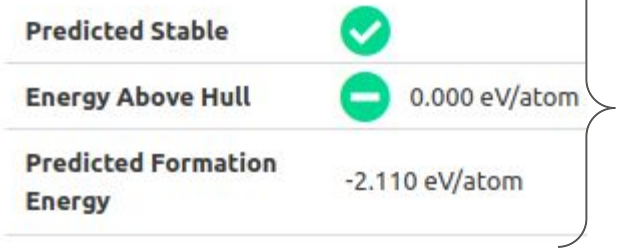
Ordering	Non-magnetic
Total Magnetization	0.00 µB/f.u.
Exchange Symmetry	225
Number of Unique Magnetic Sites	0

Magnetic Structure



Band Gap	5.00 eV
Direct Gap	Yes
Metallic	No

Electronic Structure



Predicted Stable	✓
Energy Above Hull	⊖ 0.000 eV/atom
Predicted Formation Energy	-2.110 eV/atom

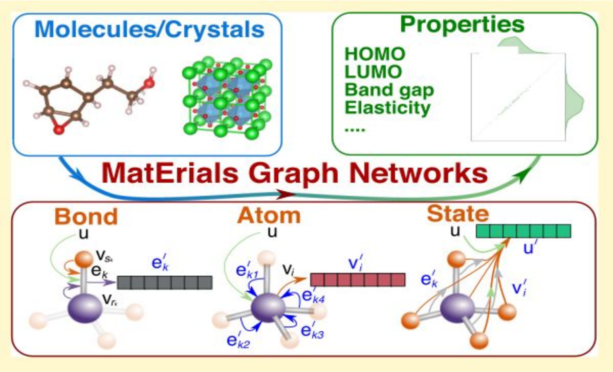
Thermodynamic Stability

Crystal Property Prediction

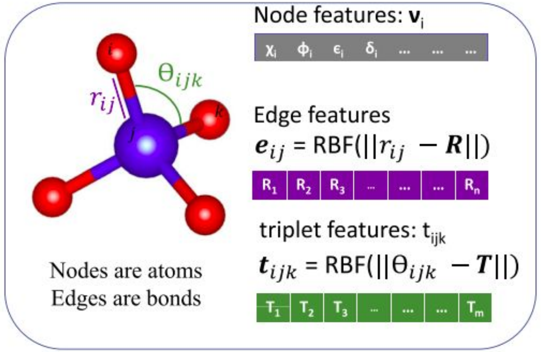
- Given a crystal material's 3D structure, predicting different properties is a challenging and important task in material science.
- **Density functional theory (DFT)** [Orio et al., 2009] : an effective tool to estimate several materials' Properties. But DFT require substantial **computational costs.**
- Recent times, data driven approaches emerged as an effective tool for predicting crystal properties which are **as accurate as DFT,** however, **much faster** than it.
- Majority of the existing approaches, constructs graphs by establishing edges only between nearby atoms and use deep graph neural network to learn crystal structure representation

Limitations of Existing Works

Incorporating specific domain knowledge into a deep encoding module.

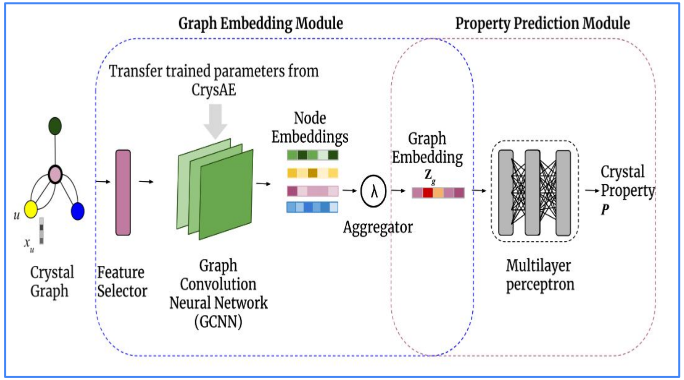


MEGNet/CGCNN/GATGNN



i
j
k
r_{ij}
θ_{ijk}
Node features: $\mathbf{v}_i$
χ_i ϕ_i ϵ_i δ_i ... ...
Edge features
$\boldsymbol{e}_{ij}$ = RBF($||r_{ij} - \boldsymbol{R}||$)
R_1 R_2 R_3 ... ... R_n
triplet features: t_{ijk}
$\boldsymbol{t}_{ijk}$ = RBF($||\theta_{ijk} - \boldsymbol{T}||$)
T_1 T_2 T_3 ... ... T_m
Nodes are atoms
Edges are bonds

ALIGNN



CrysXPP

GNN Based Approaches - *construct graphs by creating edges only between atoms within a **pre-specified distance threshold** (8 Amstrong)*

ALIGNN - *Incorporates **bond angular information** into their encoder module to capture many body interactions between atoms*

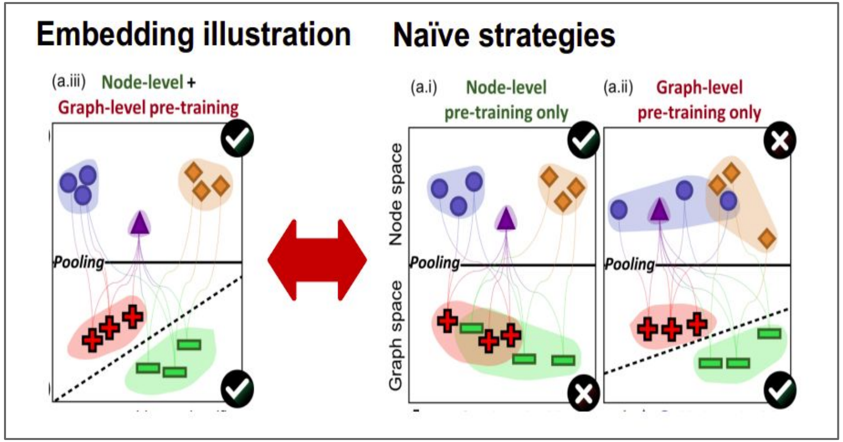
CrysXPP/ ElemNet - *adopt the concept of **transfer learning** to mitigate the data sparsity issue across properties.*

Problem Statement

Can we leverage a **large amount of untagged material structures** to **pretrain a Deep GNN model** which learn the complex hidden features which otherwise are difficult to identify?

Pretraining GNN

- Prior works focus on molecular and biological dataset, which is difficult to extend directly to crystalline material.
- Structural semantics are different between molecules and materials.
- For graph-level pre-training → supervised property prediction using a huge amount of labelled dataset → less effective in material science where property labeled data is extremely scarce.
- Conventional pre-train fine tuning framework limits knowledge transfer capability of the pre-trained model if the downstream task and dataset is different.

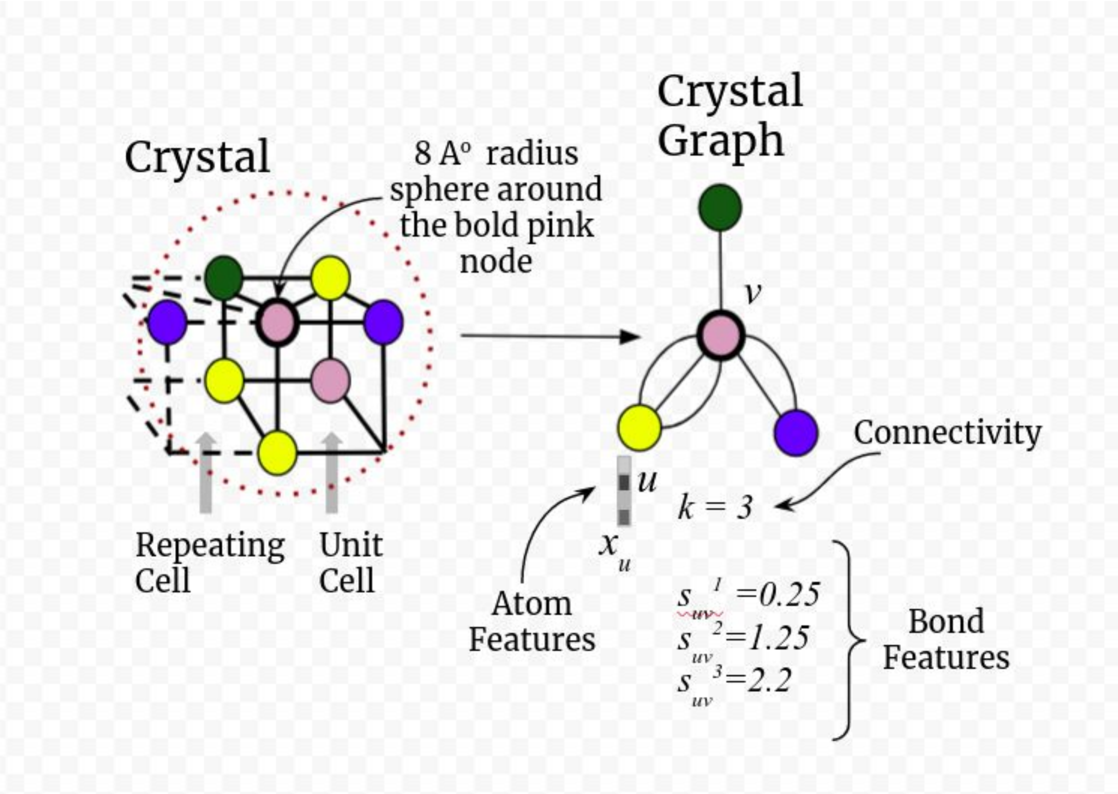

Embedding illustration
Naïve strategies
(a.iii) Node-level + Graph-level pre-training
(a.i) Node-level pre-training only
(a.ii) Graph-level pre-training only
Node space
Graph space
Pooling
Pooling
Pooling

Hu et al. Strategies for Pre-training Graph Neural Networks. (ICLR-2020) : Node-level and graph-leve pre-training on GNNs to capture domain specific knowledge about nodes and edges, in addition to global graph-level knowledge. They perform pretraining on large dataset of chemical and biological dataset.

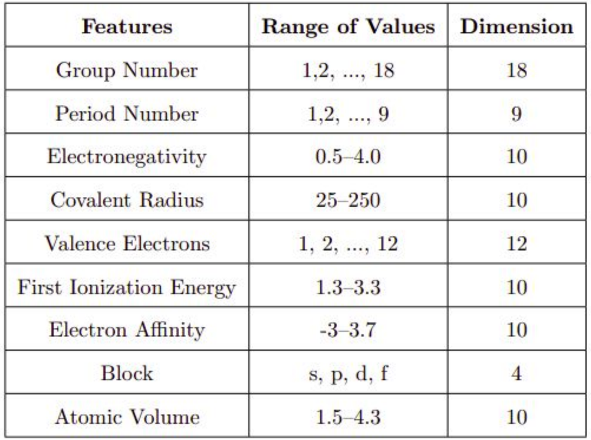
Proposed Methodology

- We developed a **pre-trained GNN model (CrysGNN)** for Crystalline materials, which captures both **local (node level) chemical** and **global (graph level) structural** semantics of crystal graphs.
- We curate a new **large untagged crystal dataset** with **800K crystal graphs** to pretrain CrysGNN.
- We introduce a **self supervised graph pre-training** method which captures (a) connectivity of different atoms, (b) different atomic properties and (c) graph similarity from a large set of unlabeled crystal graph data.
- Subsequently we **distill** important **structural and chemical information** of a crystal from the pre-trained CrysGNN model and pass it to the property predictor.
- **Retrofit** the pre-trained CrysGNN model into any existing state-of-the-art property predictor, to improve their property prediction performance.

Multi-Graph Construction of Crystal

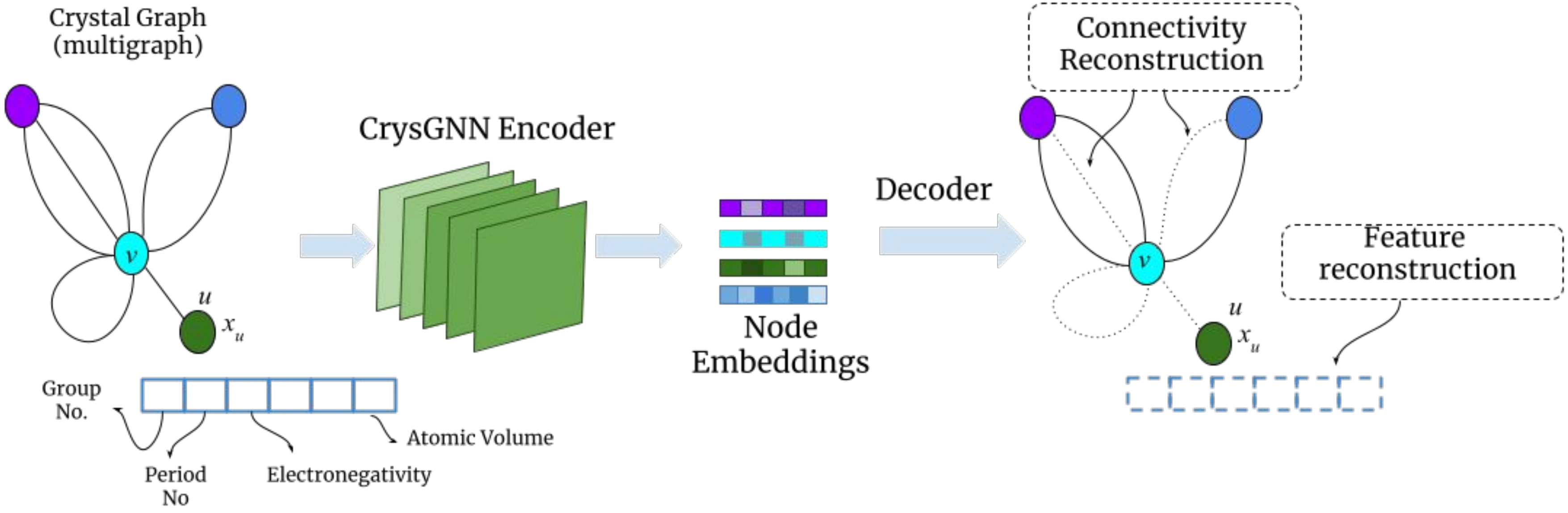


Atom Features



Features	Range of Values	Dimension
Group Number	1,2, ..., 18	18
Period Number	1,2, ..., 9	9
Electronegativity	0.5–4.0	10
Covalent Radius	25–250	10
Valence Electrons	1, 2, ..., 12	12
First Ionization Energy	1.3–3.3	10
Electron Affinity	-3–3.7	10
Block	s, p, d, f	4
Atomic Volume	1.5–4.3	10

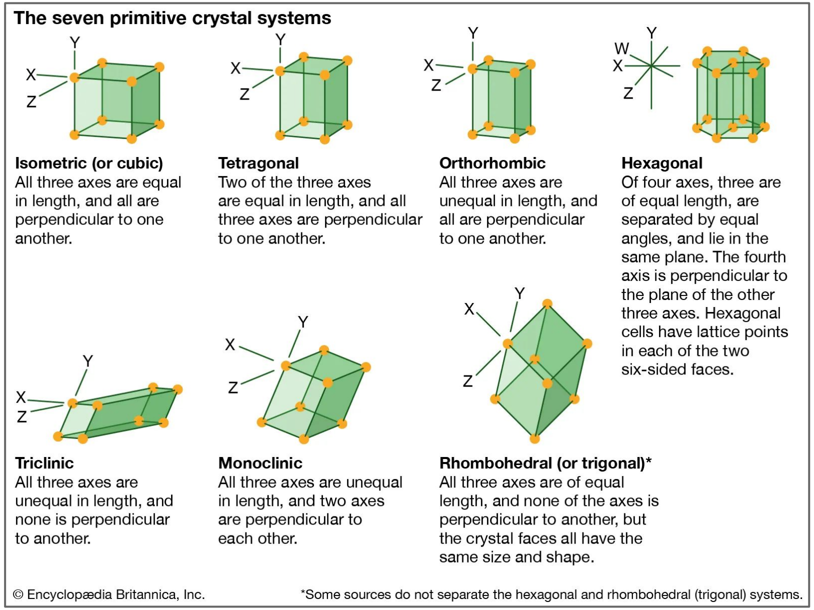
CrysGNN: Node Level Pre-training



CrysGNN: Graph Level Pre-training

- **Space group of Crystal Structure :**
  - Describe the symmetry of a unit cell of the crystal material.
  - Each crystal has a unique space group number.
  - 230 unique space groups

- **Crystal System:**
  - Space group level information can classify a crystal graph into 7 broad groups of crystal systems.



The seven primitive crystal systems

Isometric (or cubic)
All three axes are equal in length, and all are perpendicular to one another.

Tetragonal
Two of the three axes are equal in length, and all three axes are perpendicular to one another.

Orthorhombic
All three axes are unequal in length, and all are perpendicular to one another.

Hexagonal
Of four axes, three are of equal length, are separated by equal angles, and lie in the same plane. The fourth axis is perpendicular to the plane of the other three axes. Hexagonal cells have lattice points in each of the two six-sided faces.

Triclinic
All three axes are unequal in length, and none is perpendicular to another.

Monoclinic
All three axes are unequal in length, and two axes are perpendicular to each other.

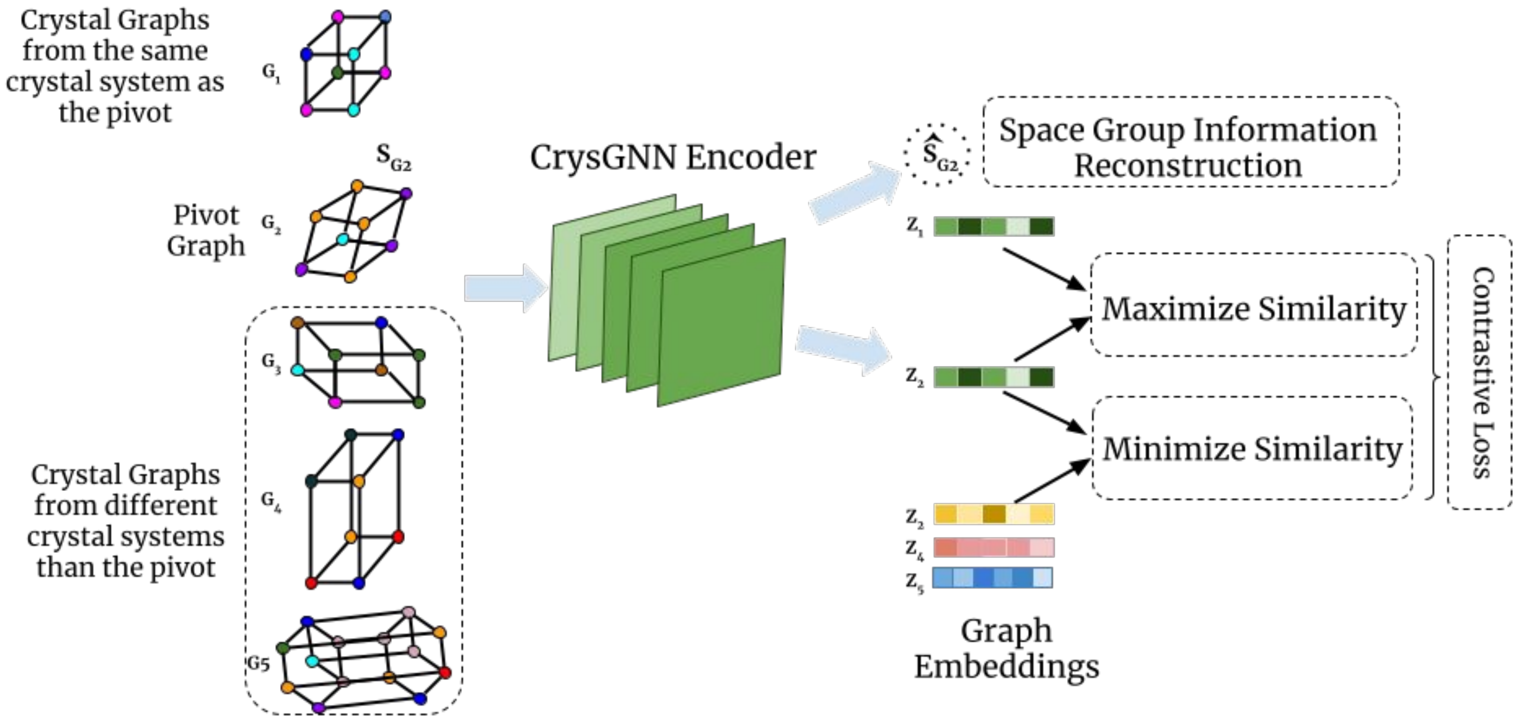
Rhombohedral (or trigonal)*
All three axes are of equal length, and none of the axes is perpendicular to another, but the crystal faces all have the same size and shape.

© Encyclopædia Britannica, Inc.

*Some sources do not separate the hexagonal and rhombohedral (trigonal) systems.

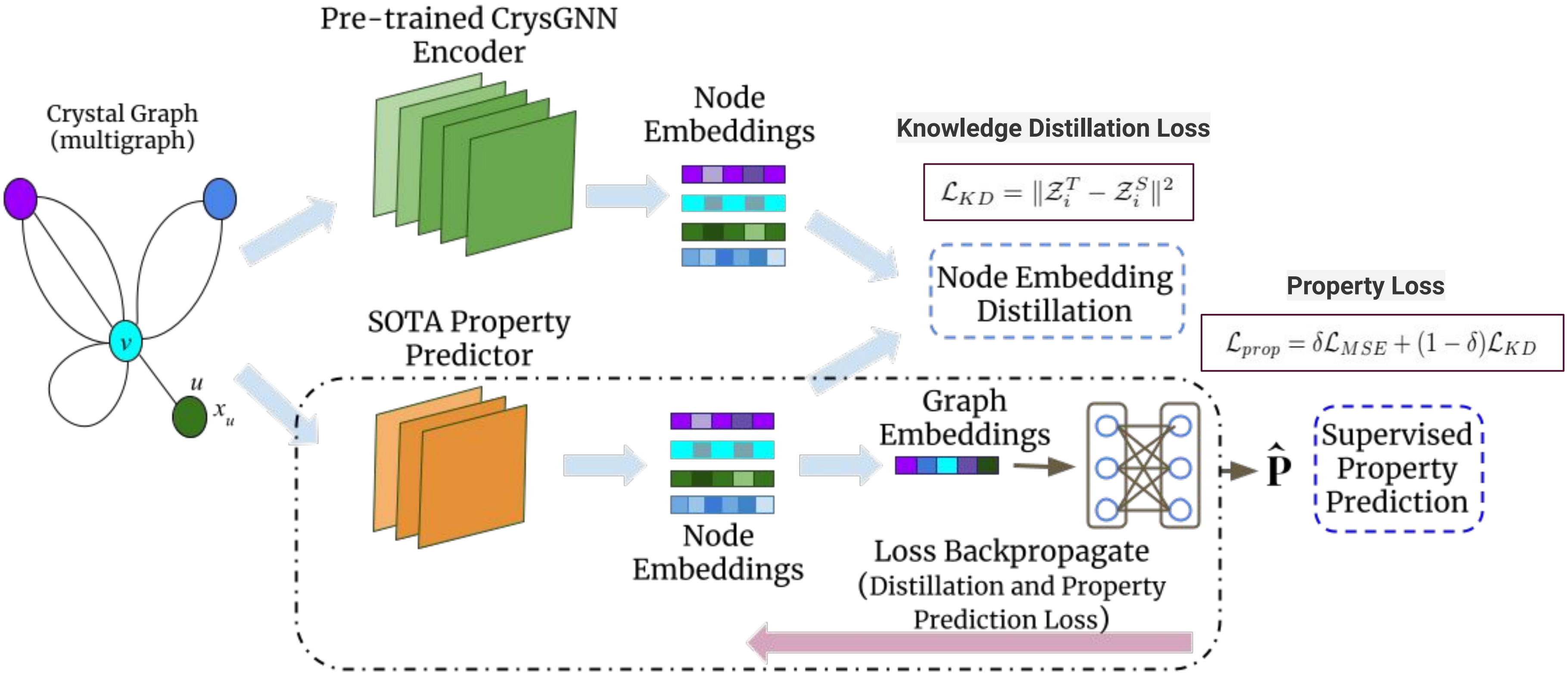
we adopt **supervised** and **contrastive learning** to learn **structural similarities** between graph structures using the **space group and crystal system information** of the materials respectively.

CrysGNN: Graph Level Pre-training



Crystal Graphs from the same crystal system as the pivot
G1
S G2
Pivot Graph
G2
Crystal Graphs from different crystal systems than the pivot
G3
G4
G5
CrysGNN Encoder
Ŝ G2
Space Group Information Reconstruction
Z1
Z2
Maximize Similarity
Minimize Similarity
Contrastive Loss
Z3
Z4
Z5
Graph Embeddings

Distillation and Property Prediction


Crystal Graph (multigraph)
v
u
x_u
Pre-trained CrysGNN Encoder
Node Embeddings
SOTA Property Predictor
Node Embeddings
Graph Embeddings
Node Embedding Distillation
$\hat{\mathbf{P}}$
Supervised Property Prediction
Loss Backpropagate (Distillation and Property Prediction Loss)

Knowledge Distillation Loss

$$\mathcal{L}_{KD} = \|\mathcal{Z}_i^T - \mathcal{Z}_i^S\|^2$$

Property Loss

$$\mathcal{L}_{prop} = \delta\mathcal{L}_{MSE} + (1-\delta)\mathcal{L}_{KD}$$

Dataset Details

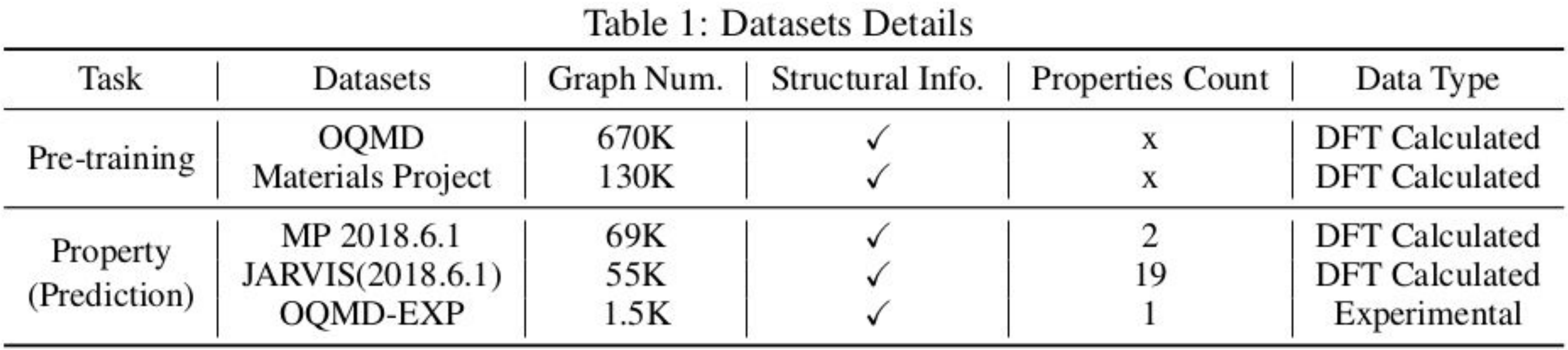
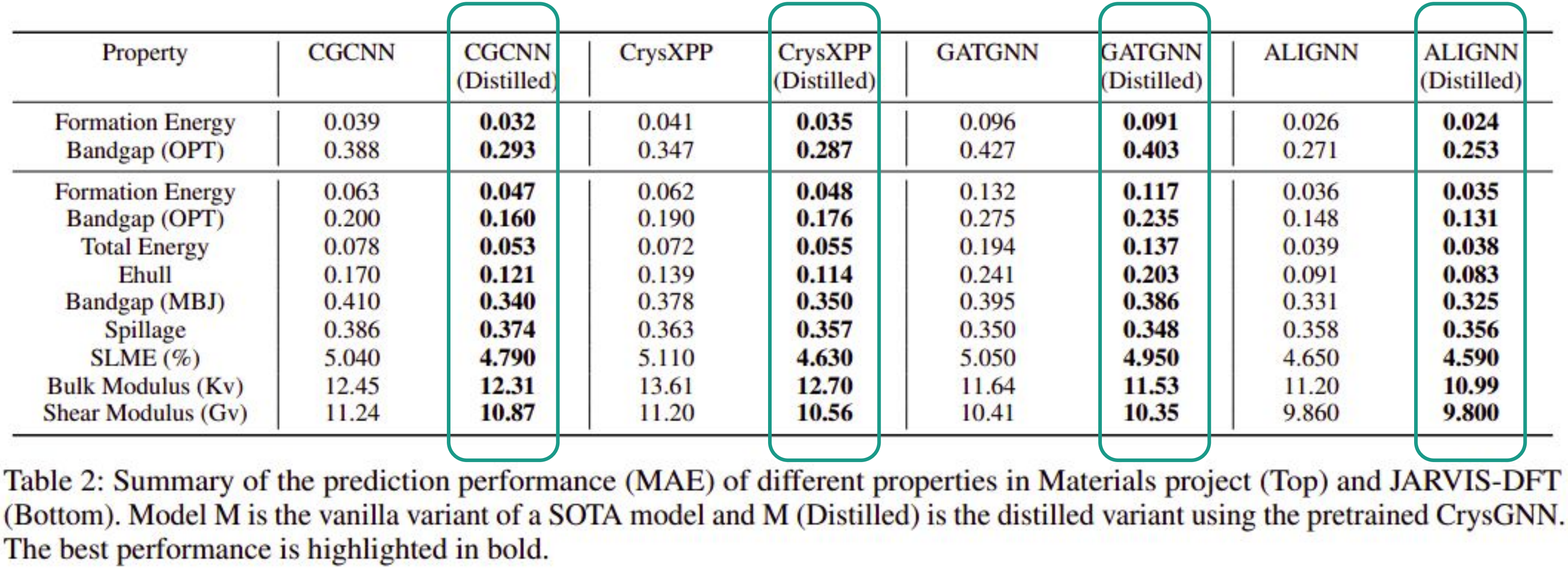


Table 1: Datasets Details

Task	Datasets	Graph Num.	Structural Info.	Properties Count	Data Type
Pre-training	OQMD	670K	✓	x	DFT Calculated
	Materials Project	130K	✓	x	DFT Calculated
Property (Prediction)	MP 2018.6.1	69K	✓	2	DFT Calculated
	JARVIS(2018.6.1)	55K	✓	19	DFT Calculated
	OQMD-EXP	1.5K	✓	1	Experimental

Downstream Task Evaluation



Property	CGCNN	CGCNN (Distilled)	CrysXPP	CrysXPP (Distilled)	GATGNN	GATGNN (Distilled)	ALIGNN	ALIGNN (Distilled)
Formation Energy	0.039	**0.032**	0.041	**0.035**	0.096	**0.091**	0.026	**0.024**
Bandgap (OPT)	0.388	**0.293**	0.347	**0.287**	0.427	**0.403**	0.271	**0.253**
Formation Energy	0.063	**0.047**	0.062	**0.048**	0.132	**0.117**	0.036	**0.035**
Bandgap (OPT)	0.200	**0.160**	0.190	**0.176**	0.275	**0.235**	0.148	**0.131**
Total Energy	0.078	**0.053**	0.072	**0.055**	0.194	**0.137**	0.039	**0.038**
Ehull	0.170	**0.121**	0.139	**0.114**	0.241	**0.203**	0.091	**0.083**
Bandgap (MBJ)	0.410	**0.340**	0.378	**0.350**	0.395	**0.386**	0.331	**0.325**
Spillage	0.386	**0.374**	0.363	**0.357**	0.350	**0.348**	0.358	**0.356**
SLME (%)	5.040	**4.790**	5.110	**4.630**	5.050	**4.950**	4.650	**4.590**
Bulk Modulus (Kv)	12.45	**12.31**	13.61	**12.70**	11.64	**11.53**	11.20	**10.99**
Shear Modulus (Gv)	11.24	**10.87**	11.20	**10.56**	10.41	**10.35**	9.860	**9.800**

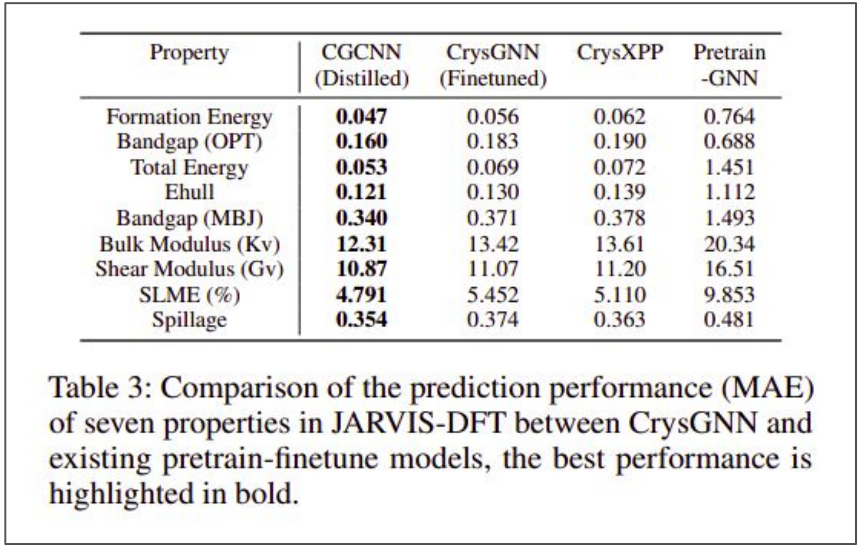
Table 2: Summary of the prediction performance (MAE) of different properties in Materials project (Top) and JARVIS-DFT (Bottom). Model M is the vanilla variant of a SOTA model and M (Distilled) is the distilled variant using the pretrained CrysGNN. The best performance is highlighted in bold.

Downstream Task Evaluation

- Distilled version of any state-of- the-art model outperforms the vanilla model across all the properties.
- Average relative improvement across all properties for ALIGNN (4.19%) and GATGNN (8.02%) is lesser compared to CGCNN (16.20%) and CrysXPP (12.21%).
- **Possible reason :** ALIGNN and GATGNN are more complex models that CrysGNN.
- **Potential Improvement :** Incorporating angle-based information or attention mechanism as a part of pre-training framework may improve further.

Comparison with Existing Pre-trained Models.

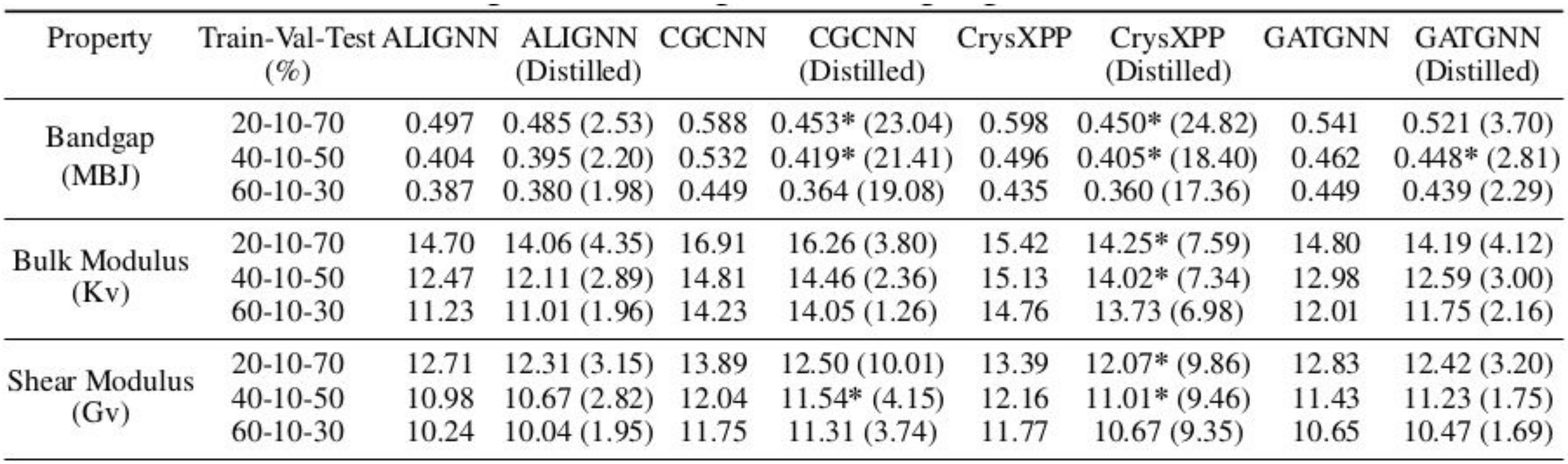
- Demonstrate the effectiveness of the knowledge distillation method vis-a-vis the conventional fine-tuning approaches.
- We finetune CrysGNN and compare with distilled CGCNN, CrysXPP and pretrain GNN by hu et.al.
- Encoding architecture is same for CrysGNN, CGCNN, and CrysXPP (pretrained-finetuned version of CGCNN)
- Distilled CGCNN outperforms finetuned version of CrysGNN and both the baselines



Property	CGCNN (Distilled)	CrysGNN (Finetuned)	CrysXPP	Pretrain -GNN
Formation Energy	**0.047**	0.056	0.062	0.764
Bandgap (OPT)	**0.160**	0.183	0.190	0.688
Total Energy	**0.053**	0.069	0.072	1.451
Ehull	**0.121**	0.130	0.139	1.112
Bandgap (MBJ)	**0.340**	0.371	0.378	1.493
Bulk Modulus (Kv)	**12.31**	13.42	13.61	20.34
Shear Modulus (Gv)	**10.87**	11.07	11.20	16.51
SLME (%)	**4.791**	5.452	5.110	9.853
Spillage	**0.354**	0.374	0.363	0.481

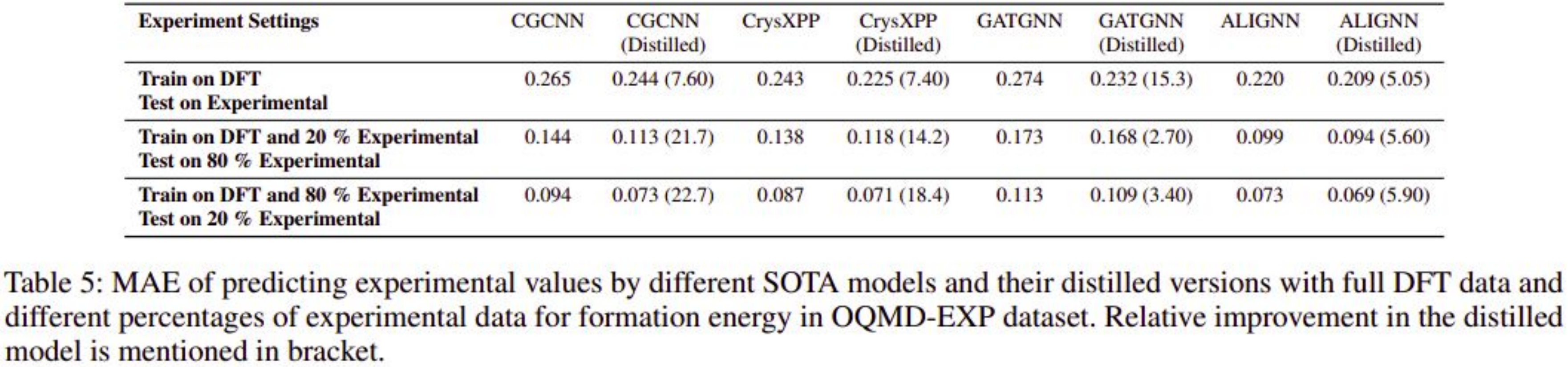
Table 3: Comparison of the prediction performance (MAE) of seven properties in JARVIS-DFT between CrysGNN and existing pretrain-finetune models, the best performance is highlighted in bold.

Effectiveness on sparse training dataset.



Property	Train-Val-Test (%)	ALIGNN	ALIGNN (Distilled)	CGCNN	CGCNN (Distilled)	CrysXPP	CrysXPP (Distilled)	GATGNN	GATGNN (Distilled)
Bandgap (MBJ)	20-10-70	0.497	0.485 (2.53)	0.588	0.453* (23.04)	0.598	0.450* (24.82)	0.541	0.521 (3.70)
	40-10-50	0.404	0.395 (2.20)	0.532	0.419* (21.41)	0.496	0.405* (18.40)	0.462	0.448* (2.81)
	60-10-30	0.387	0.380 (1.98)	0.449	0.364 (19.08)	0.435	0.360 (17.36)	0.449	0.439 (2.29)
Bulk Modulus (Kv)	20-10-70	14.70	14.06 (4.35)	16.91	16.26 (3.80)	15.42	14.25* (7.59)	14.80	14.19 (4.12)
	40-10-50	12.47	12.11 (2.89)	14.81	14.46 (2.36)	15.13	14.02* (7.34)	12.98	12.59 (3.00)
	60-10-30	11.23	11.01 (1.96)	14.23	14.05 (1.26)	14.76	13.73 (6.98)	12.01	11.75 (2.16)
Shear Modulus (Gv)	20-10-70	12.71	12.31 (3.15)	13.89	12.50 (10.01)	13.39	12.07* (9.86)	12.83	12.42 (3.20)
	40-10-50	10.98	10.67 (2.82)	12.04	11.54* (4.15)	12.16	11.01* (9.46)	11.43	11.23 (1.75)
	60-10-30	10.24	10.04 (1.95)	11.75	11.31 (3.74)	11.77	10.67 (9.35)	10.65	10.47 (1.69)

Removal of DFT error bias using experimental data



Experiment Settings	CGCNN	CGCNN (Distilled)	CrysXPP	CrysXPP (Distilled)	GATGNN	GATGNN (Distilled)	ALIGNN	ALIGNN (Distilled)
Train on DFT **Test on Experimental**	0.265	0.244 (7.60)	0.243	0.225 (7.40)	0.274	0.232 (15.3)	0.220	0.209 (5.05)
Train on DFT and 20 % Experimental **Test on 80 % Experimental**	0.144	0.113 (21.7)	0.138	0.118 (14.2)	0.173	0.168 (2.70)	0.099	0.094 (5.60)
Train on DFT and 80 % Experimental **Test on 20 % Experimental**	0.094	0.073 (22.7)	0.087	0.071 (18.4)	0.113	0.109 (3.40)	0.073	0.069 (5.90)

Table 5: MAE of predicting experimental values by different SOTA models and their distilled versions with full DFT data and different percentages of experimental data for formation energy in OQMD-EXP dataset. Relative improvement in the distilled model is mentioned in bracket.

Conclusion

- In this work, we present a novel but simple **pre-trained GNN framework**, CrysGNN, for crystalline materials.
- Captures both **local chemical** and **global structural semantics** of crystal graphs, using node and graph level pre-training respectively
- We curate a new **large untagged crystal dataset** with **800K crystal graphs** to pretrain CrysGNN. We will release the pre-trained model along with the large dataset for the community.
- We **distill important knowledge** from CrysGNN and **inject** it into different state of the art property predictors and **enhance their performance**. We believe this approach can have applications in other domains too.
- Extensive experiments show its superiority over conventional fine-tune models and its inherent ability to remove DFT-induced bias.

Github Repo for CrysGNN : https://github.com/kdmsit/crysgnn
Github Repo for 800K Dataset : https://github.com/kdmsit/crystal_untagged_800K

Thank You for Listening

Any Questions?

Email: kishalaydas@kgpian.iitkgp.ac.in
Complex Network Research Group (CNeRG) : @cnerg