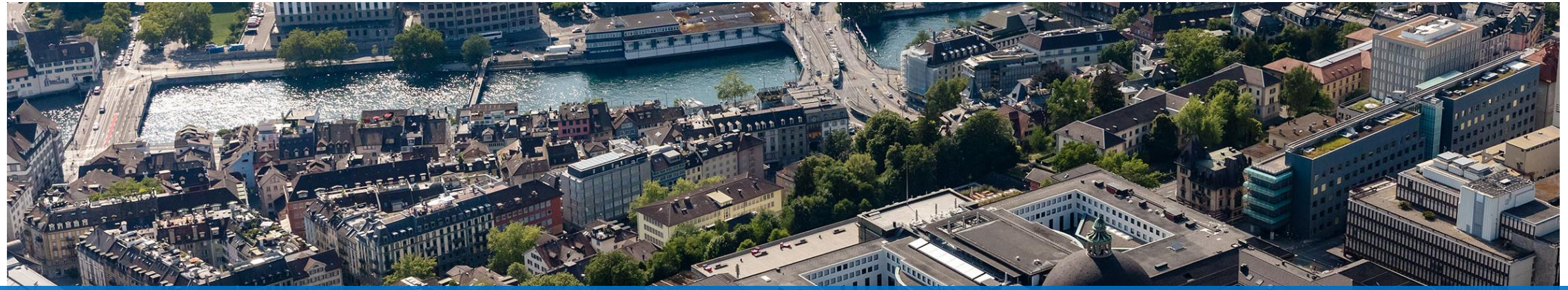
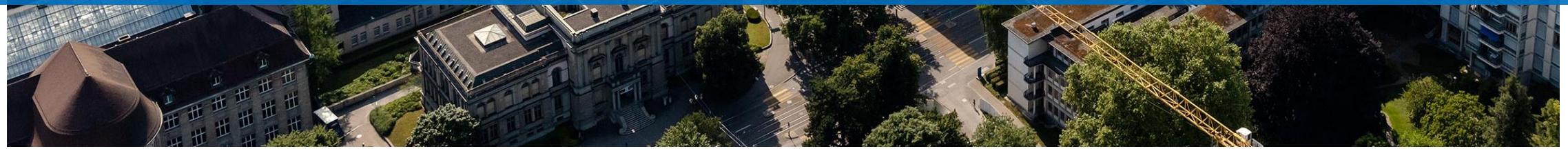
ETH zürich



Design Space for Graph Neural Networks

Jiaxuan You, Rex Ying, Jure Leskovec

Presenter: Jiaqing Xie, Ziheng Chi



ETH zürich

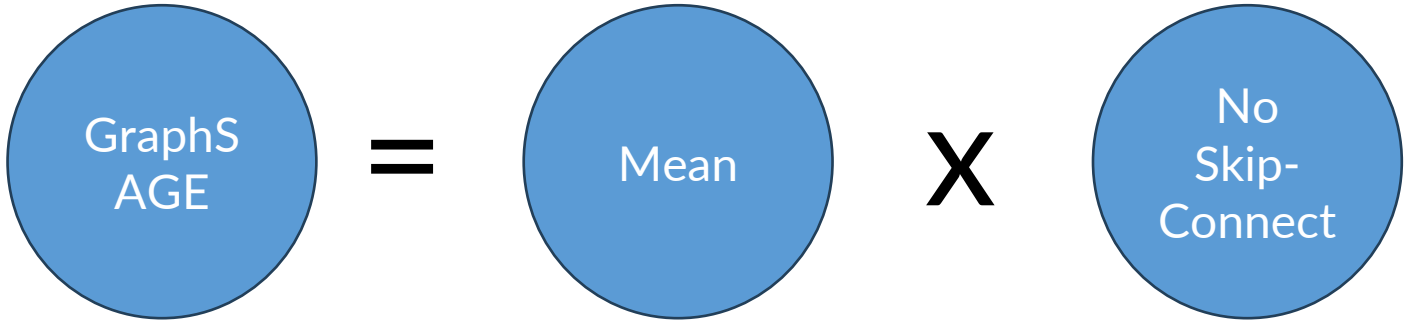
1. Issues

- Lack of General GNN Design
- Lack of Evaluation on New Tasks

Issue 1: Lack of General GNN Design

Example: GraphSAGE

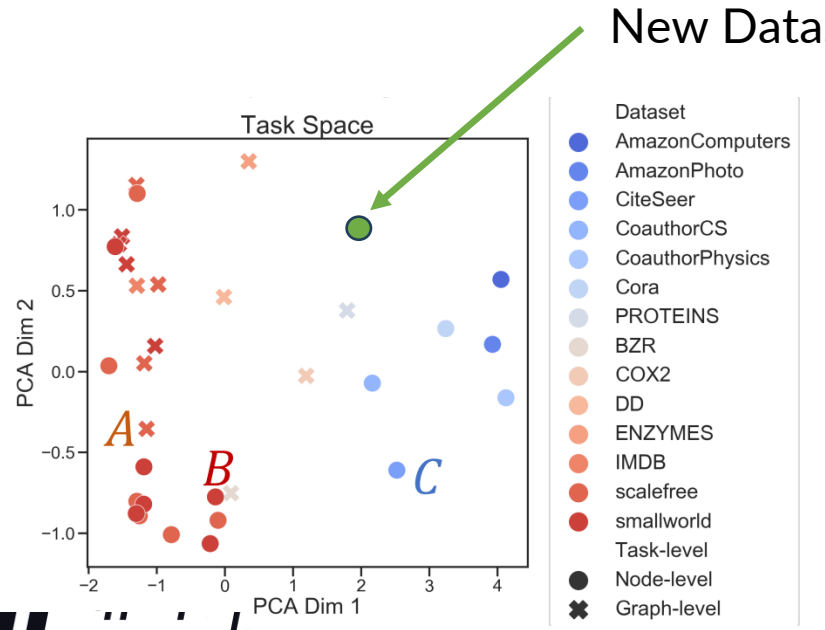
- GraphSAGE: mean / max / LSTM aggregation
- Change aggregation function to summation, no longer GraphSAGE
- Add skip-connection, no longer GraphSAGE
- However, adding summation and skip-connection could help learn some tasks better



Issue 2: Lack of Evaluation on New Tasks

- Evaluate GNN by introducing new tasks
- However new tasks may not resemble existing GNN benchmarks
- Unclear how to design a GNN for new coming tasks

Scenario 1 (Example):



Scenario 2 (Example):

Large Design Space

- GNN-Layers {2 , 4 , 6, 8}
- Aggregation {mean, max, sum}
- Layer Connectivity {skip-cat, skip-sum}
- Batch Size {4 choices}
- Learning Rate {4 choices}
- 4 * 3 * 2 * 4 * 4 = 384 potential models

Exhaustive search to find a SOTA model is not time-efficient.

ETH zürich

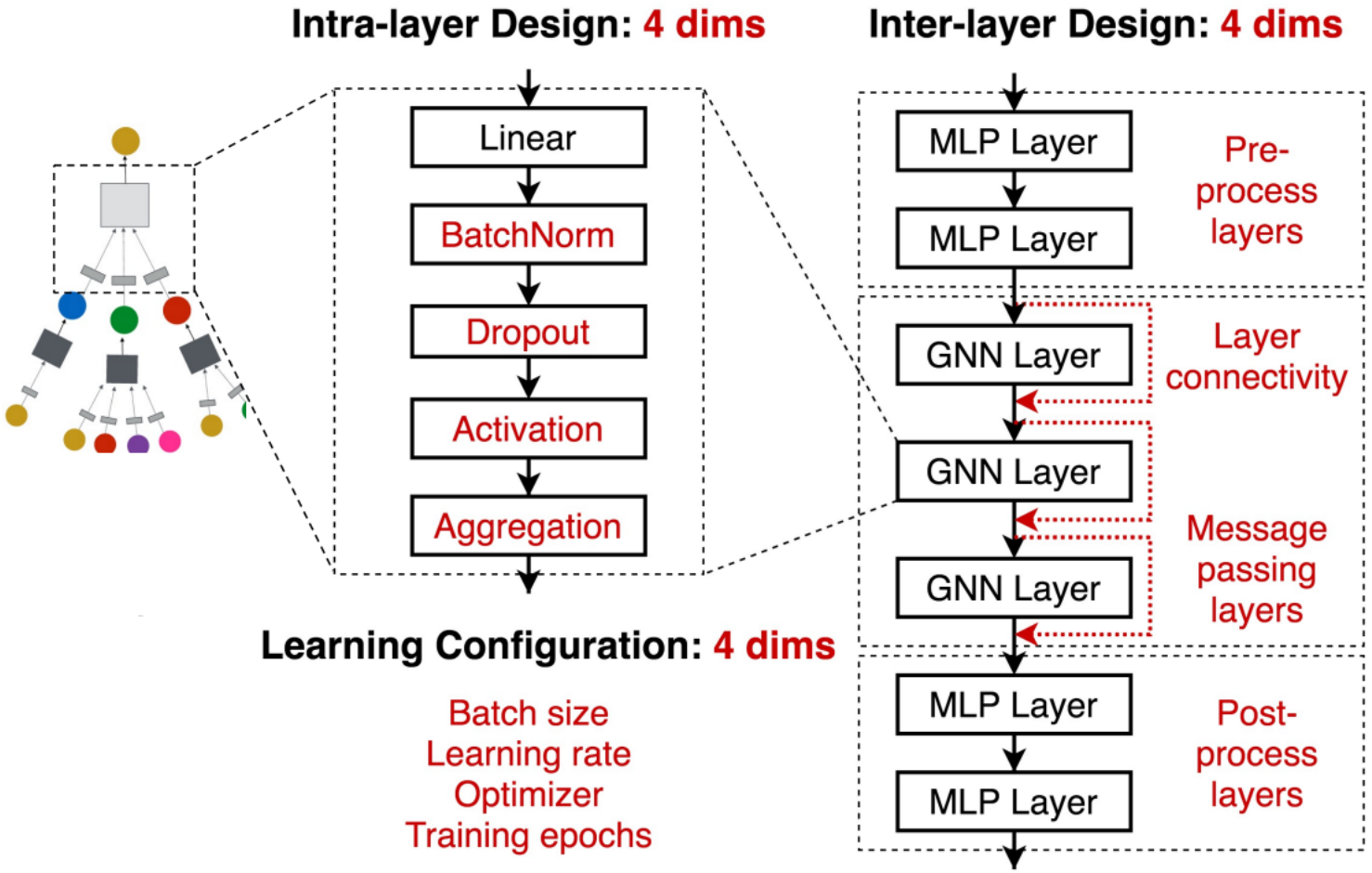
2. Motivations

- Design Space for GNN
- Task Space for GNN

Motivation 1: Design Space for GNN

Main Design Dimensions:
- Intra-layer Design
- Inter-layer Design
- Learning Configuration

315K possible Designs
* Intuition: A condensed search



Motivation 2: Task Space for GNN

It is **difficult** to tell whether GNN is transferable between tasks / datasets:

- Two tasks belong to node classification but result in different SOTA GNN Design

Task Similarity Metric could:

- Transfer GNN design to similar tasks
- Identify new tasks that are dissimilar to all other tasks

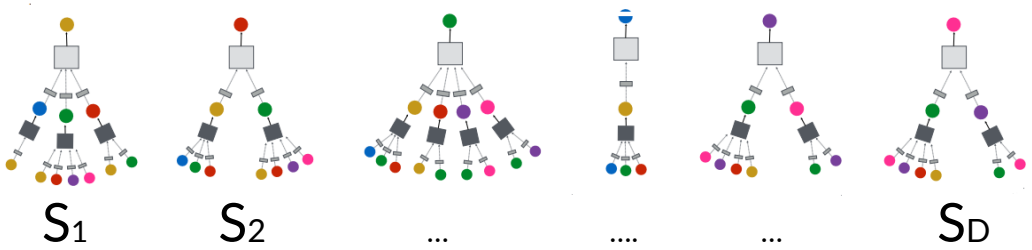
Main Components:

- Selection of anchor models
- Rank distance measurement of the performance of anchor models

Motivation 2: Task Space for GNN

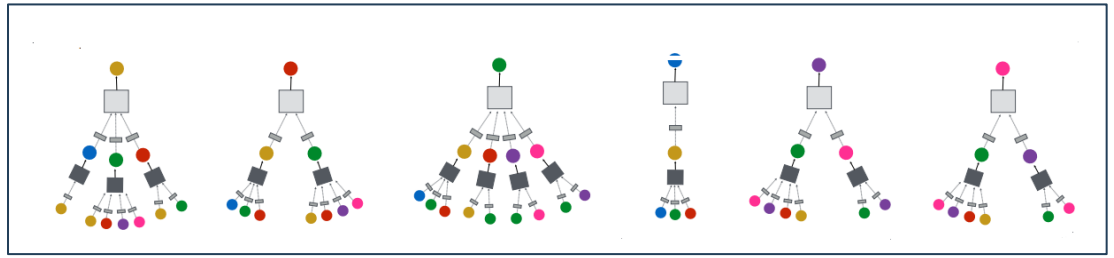
1. Anchor Model: Goal is to find diverse GNN design

- Sample D random GNN candidates from GNN Space : S_1, S_2, …, S_D.



- Fix number of GNN tasks, record each GNN's average performance across tasks.

Tasks
Node
Edge
Graph



D batches of GNN

GNN	Score (average)
S_6	0.86
S_2	0.8
...	...
S_D	0.5
S_1	0.45

} D

- Ranked and sliced into M groups, model with median performance is chosen within each group.

Example D = 110, M = 10

GNN	Rank
S_6	1
S_2	2
...	...
S_1	110

Group	Model ID (Rank)
1	1 – 11
2	12 – 22
...	...
10	100– 110

Group	Anchor Model ID (Rank)
1	6
2	17
...	...
10	105

ETH zürich

Motivation 2: Task Space for GNN

2. Rank Distance Measurement

Kendall rank correlation coefficient between tasks

Task Similarity Metric

	Anchor Model Performance ranking					Similarity to Task A
Task A	M_1	M_2	M_3	M_4	M_5	1.0
Task B	M_1	M_3	M_2	M_4	M_5	0.8
Task C	M_5	M_1	M_4	M_3	M_2	-0.4

M = 12 is enough for comparison
T Tasks lead to a T*T similarity matrix

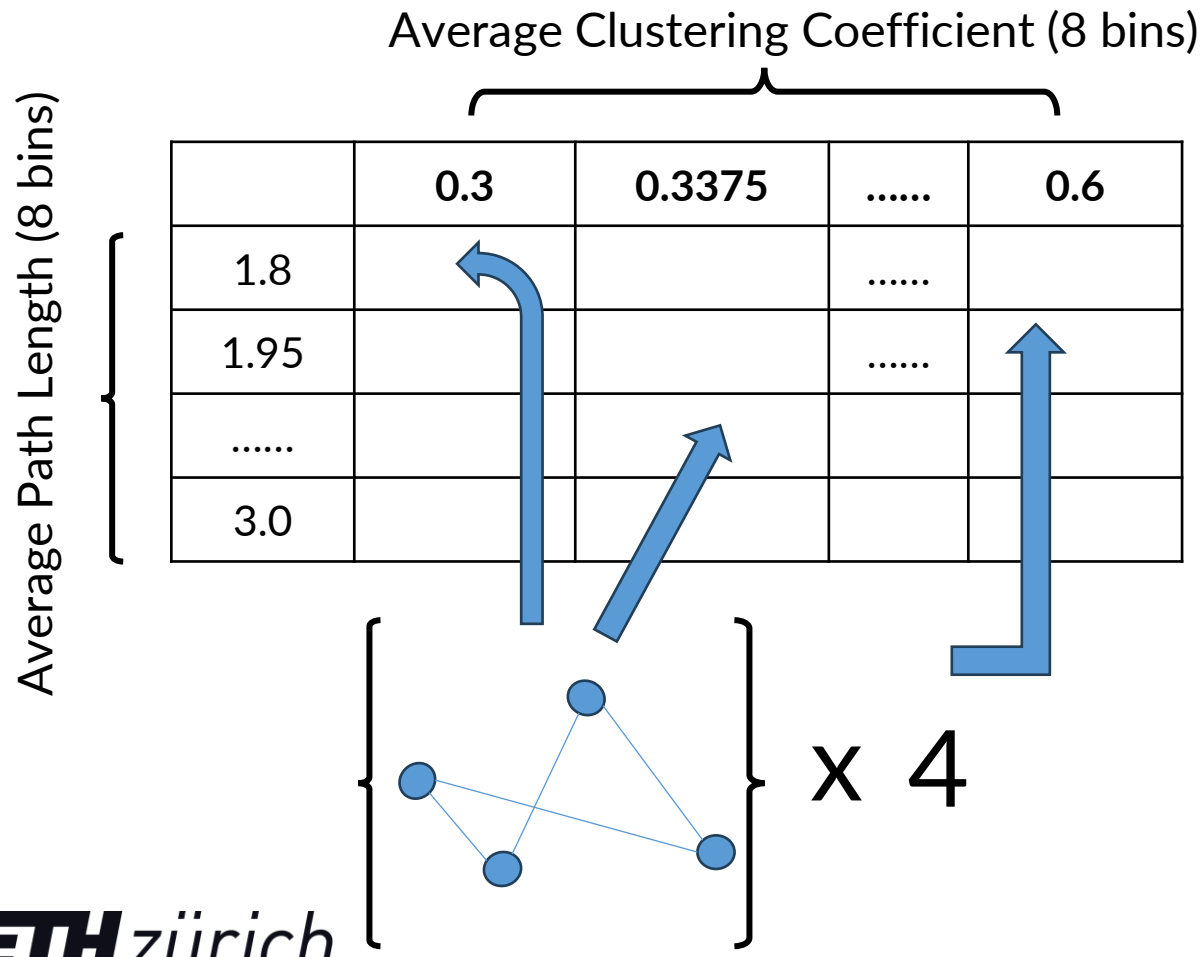
Only care about the ranking instead of the metric of each task.

A node level task might be highly related with a graph level task

Motivation 2: Task Space for GNN

Extended datasets: Synthetic data and Real-World data

Synthetic data: Embed graph statistics



Node-Level Features:
- ☐ Constant features
- ☐ One-hot vectors
- ☐ Node clustering coefficients
- ☐ Node PageRank score

Node-level Labels:
- ☐ Node clustering coefficients
- ☐ Node PageRank score

Graph-level Labels:
- ☐ Average Path Length

Node features predict node labels or graph labels

ETH zürich

3. Experiments

- Design Space Evaluations
- Task Space Evaluations

Evaluation 1: Design dimensions

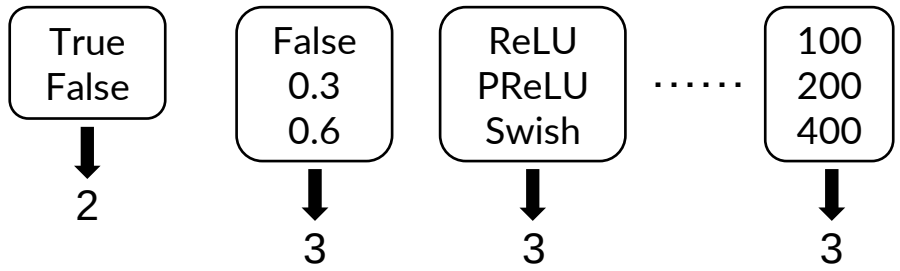
- Setup
  - Previously, total number of task-model pairs: 32 (tasks) × 314,928 (models) ≈ 10,000,000
  - Condensed design space:

32 tasks	96 task-model pairs
Task 1	Task 1 – Model 1
	Task 1 – Model 2
	Task 1 – Model 3
.....	.....
Task 32	Task 32 – Model 94
	Task 32 – Model 95
	Task 32 – Model 96

Task	BatchNorm	Dropout	Activation	.....	Epochs
Task 1	True	0.3	ReLU	.....	200
Task 1	False	0.3	ReLU	.....	200

  - Now, number of task-model pairs to test:

$$96 \times (C_{BatchNorm} + C_{Dropout} + C_{Activation} + \ldots\ldots C_{Epochs}) \approx 3000$$



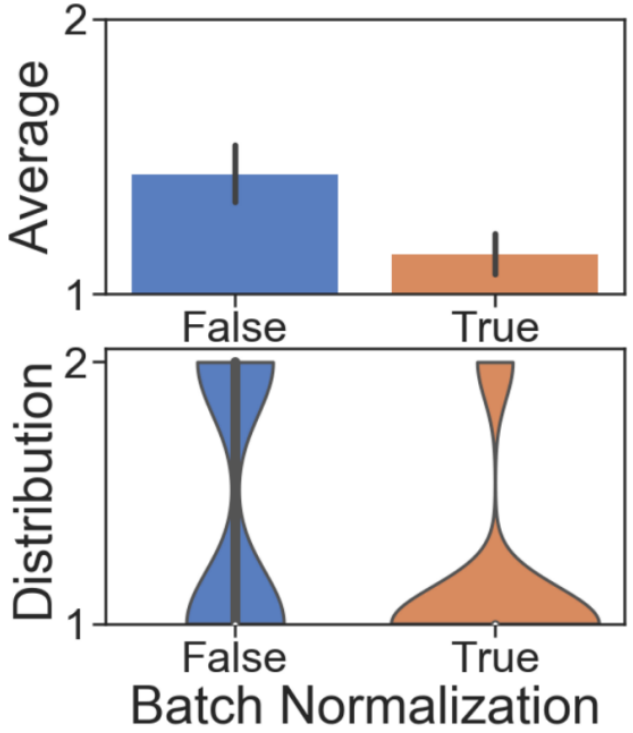
Evaluation 1: Design dimensions

- Results

Rank Design Choices by Performance

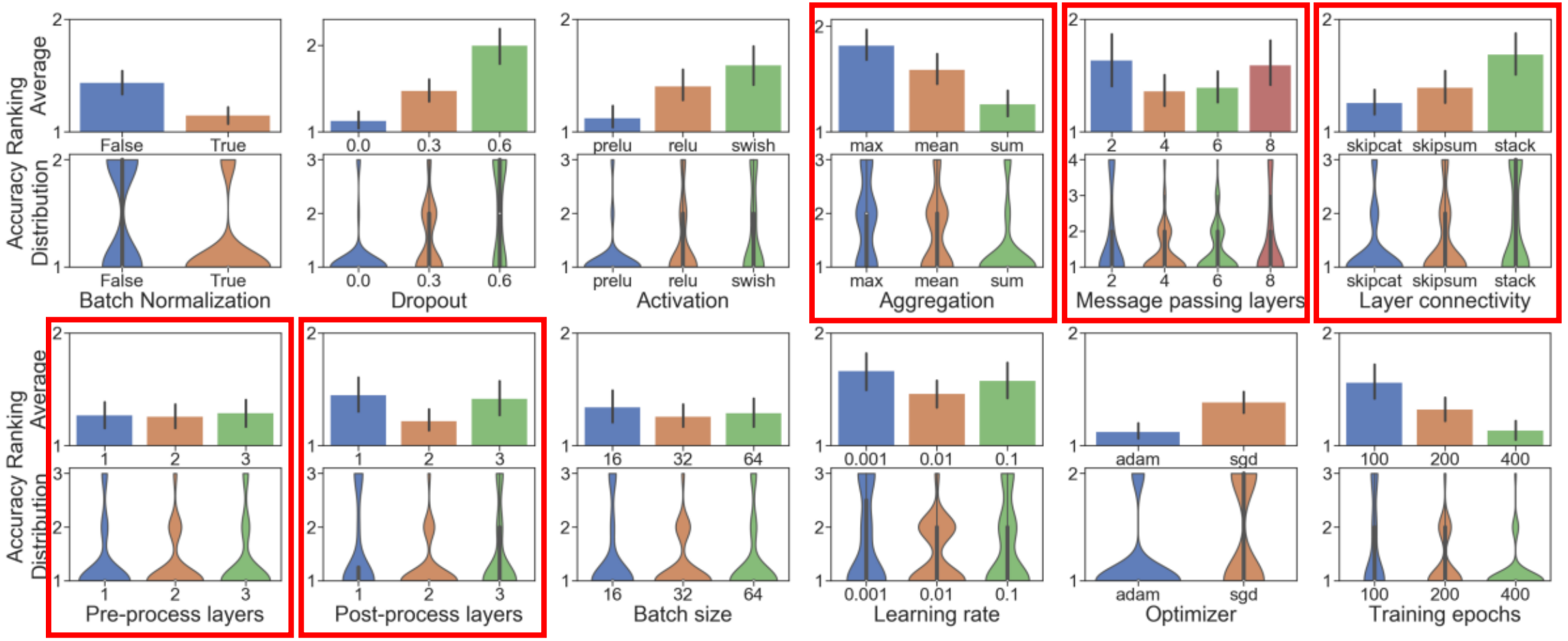
	Experimental Results	
	Val. Accuracy	Design Choice Ranking
Group 1	0.75	1
	0.54	2
Group 2	0.88	1 (a tie)
	0.86	1 (a tie)
	.....	.....
Group 96	0.89	1
	0.36	2

Ranking Analysis



Evaluation 1: Design dimensions

- Results



Evaluation 1: Design dimensions

- Condense the design space
  - Fixed design choices

Activation	BN	Dropout	Batch	LR	Optimizer	Epoch
PReLU	True	False	32	0.01	Adam	400

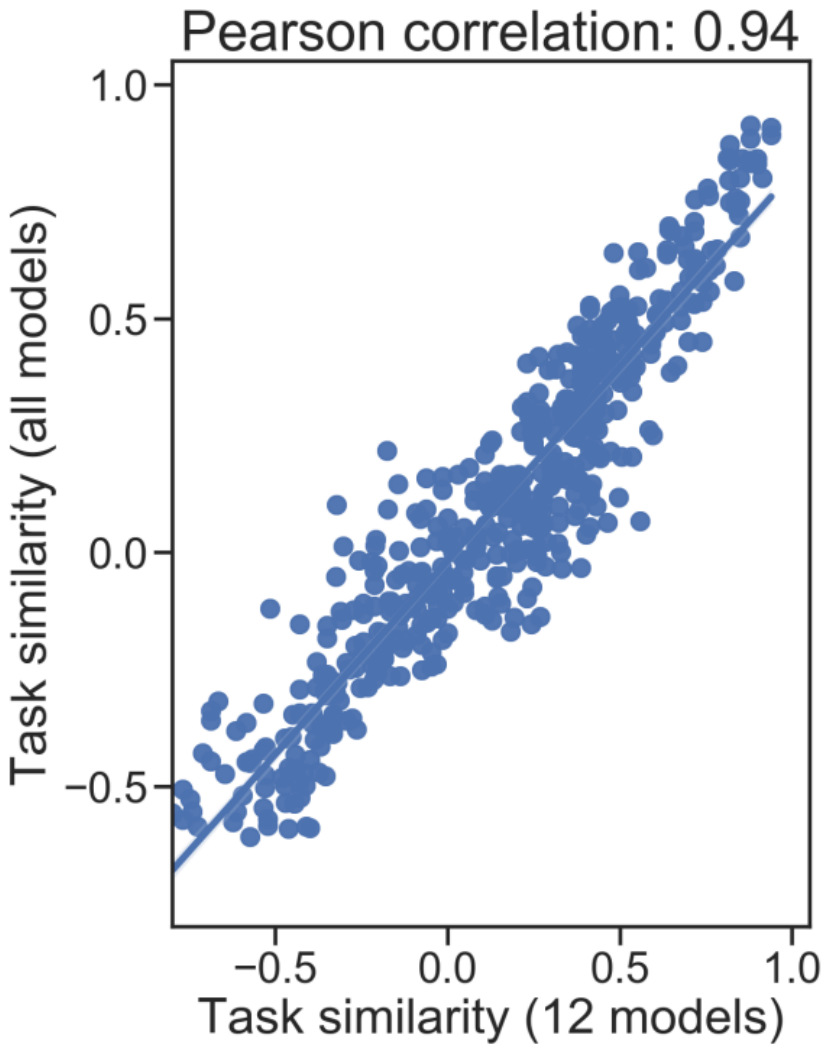
  - Debatable design choices

Aggregation	MP layers	Pre-MP layers	Post-MP layers	Connectivity
Mean, Max, Sum	2,4,6,8	1,2	2,3	Skip-Sum, Skip-Cat
↓	↓	↓	↓	↓
3	4	2	2	2

  - Condensed design space: 3×4×2×2×2 = 96 ≪ 314,928, which allows grid search.

Evaluation 2: Similarity Between 32 Tasks

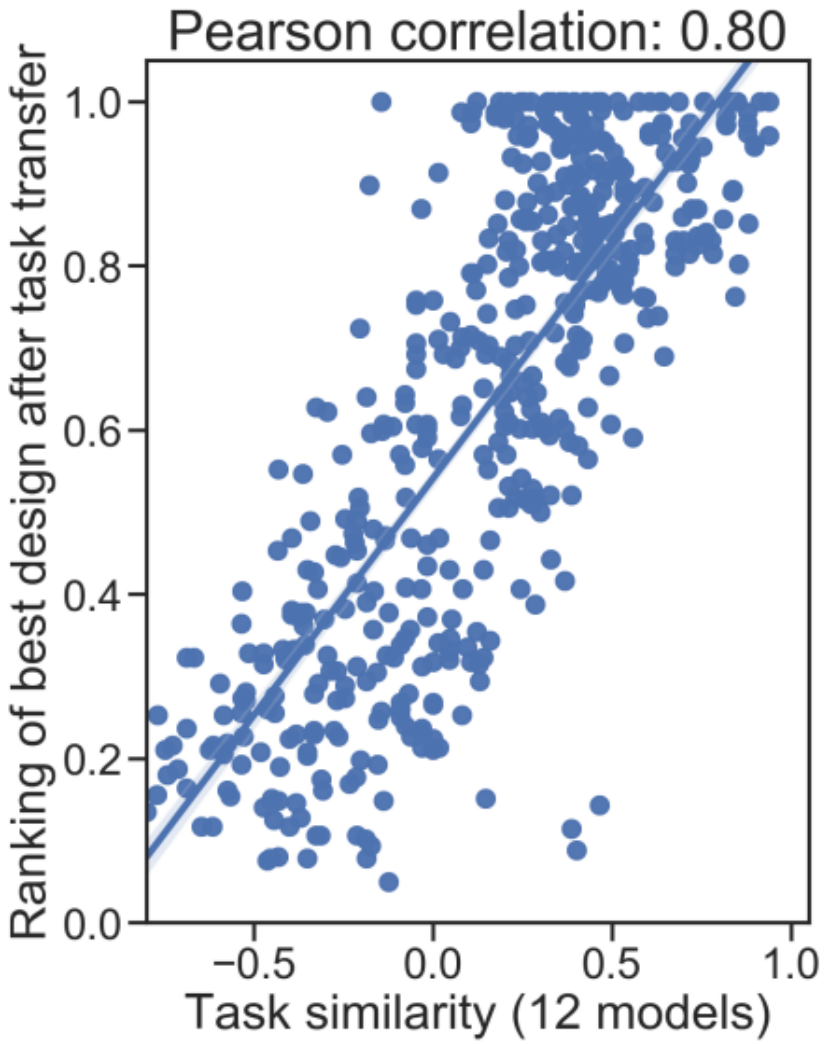
Evaluation 3: Effectiveness of 12 Anchor Models



Notations:

- Each point: A pair of two tasks.
- x-value: Similarity calculated from 12 anchor models.
- y-value: Similarity calculated from 96 anchor models.
- Correlation value: 0.94
  - ➢ Higher → 12 anchors are already representative enough.

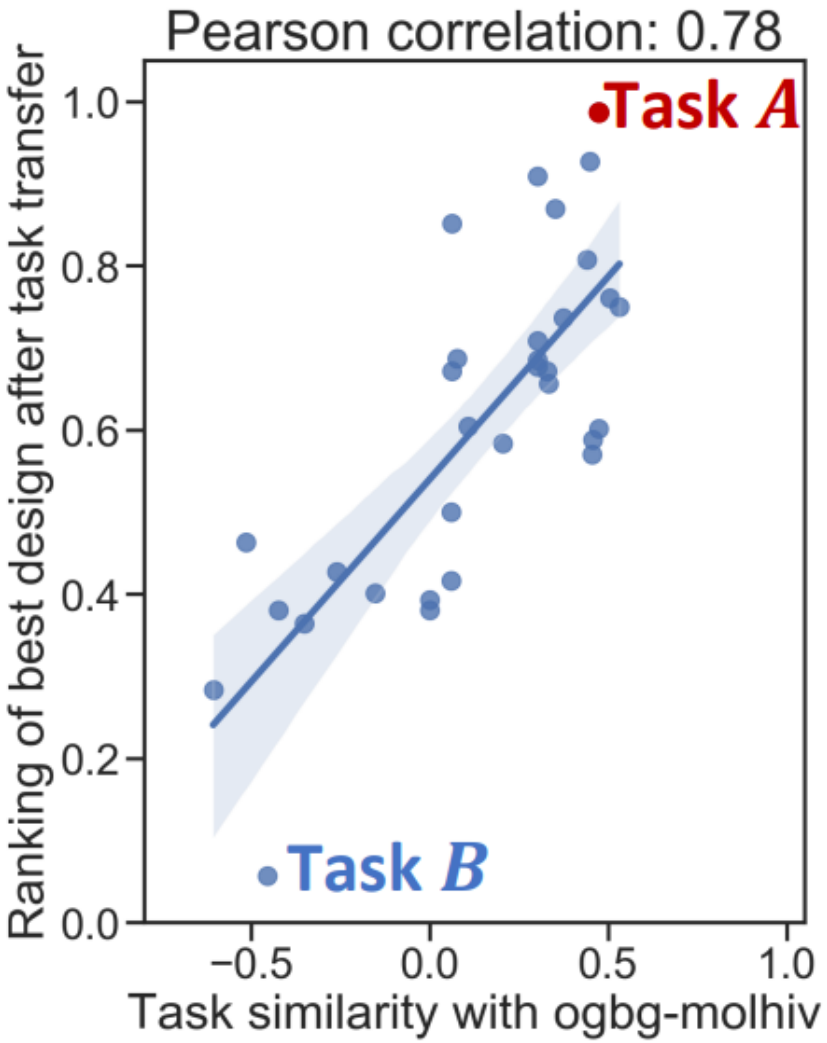
Evaluation 4: Model Transferability



Notations:

- Each point: A pair of two tasks.
- x-value: Similarity of task A and task B.
- y-value: Performance ranking (among the condensed design space) after transferring the best model of task A to task B.
- Correlation value: 0.80
  - ➢ Higher → Similar tasks have similar best models.

Evaluation 5: Application to A New Task



- Each point: One of the 32 tasks.
- x-value: Similarity between the task and the new task.
- y-value: Performance ranking after transferring the best model.

	Task A: graph-scalefree-const-path	Task B: node-CoauthorPhysics	Target task: ogbg-molhiv
Best design in our design space	(1, 8, 3, skipcat, sum)	(1, 4, 2, skipcat, max)	(2, 6, 3, skipcat, add)
Best design's performance	0.865	0.968	***0.792***
Previously reported SOTA	N/A	0.930	**0.771**
Task Similarity with ogbg-molhiv	0.47	-0.61	1.0
Performance after transfer to ogbg-molhiv	***0.785***	0.736	N/A

Any questions?

ETH zürich

Supplementary Slides

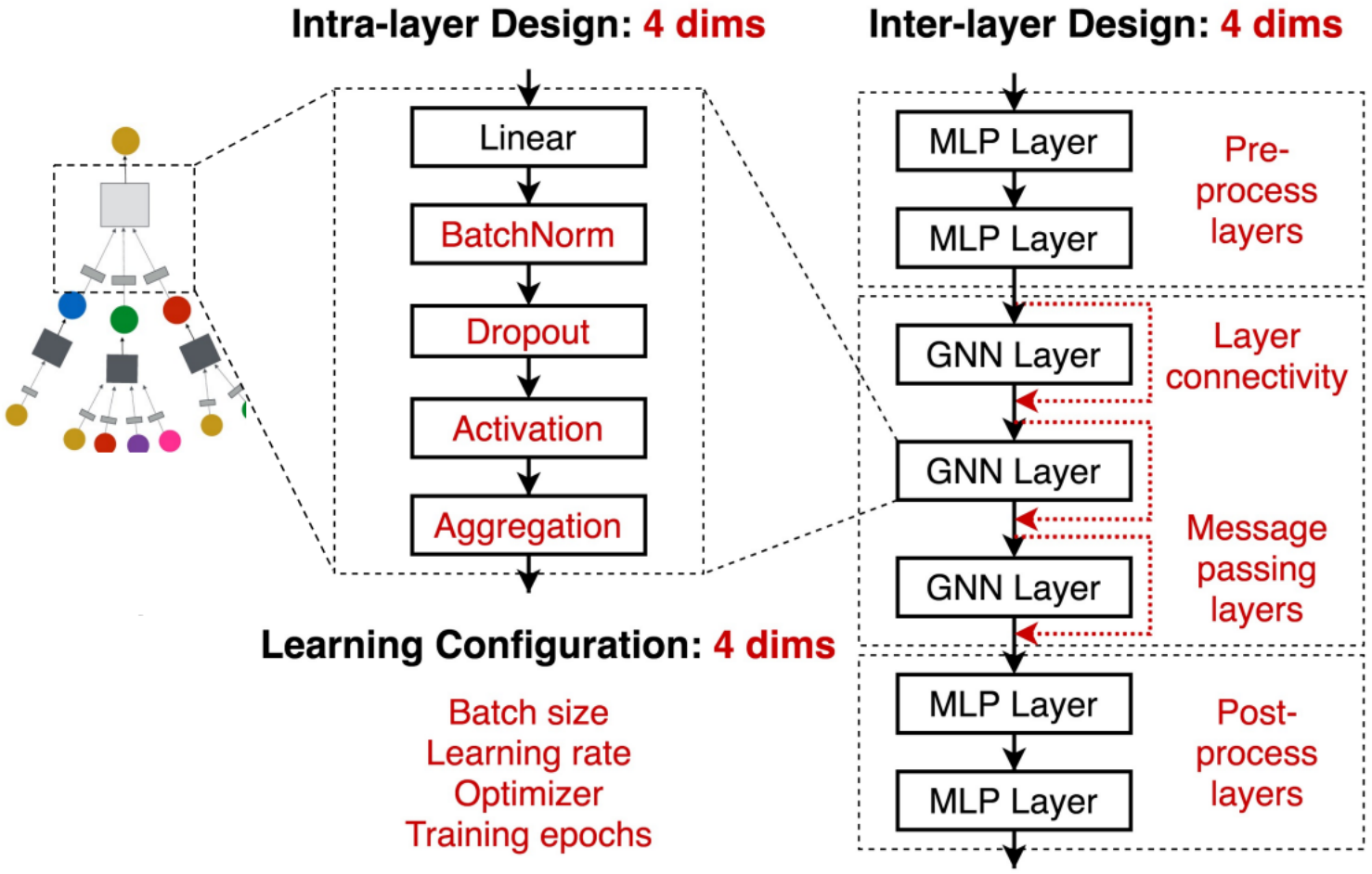
Motivation 1: Design Space for GNN

How 315K Comes from ?

BatchNorm 2 choices
Dropout 3 choices
Activation 3 choices
Aggregation 3 choices
2 * 3 * 3 * 3 = 54

Connectivity 3 choices
Pre-process 3 choices
Message-Passing 4 choices
Post-Process 3 choices
3 * 3 * 4 * 3 = 108

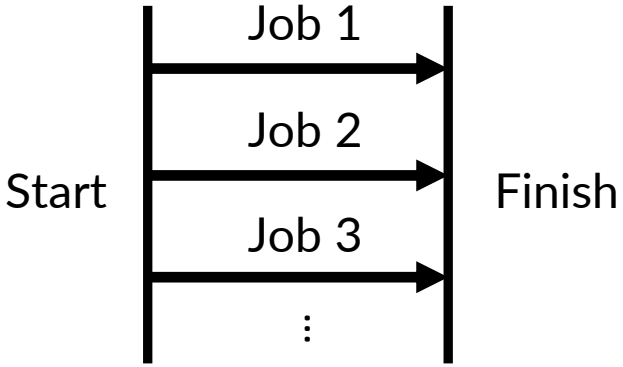
Batch Size 3 choices
Learning Rate 3 choices
Optimizer 2 choices
Training Epochs 3 choices
3 * 3 * 2 * 3 = 54
So together 54 * 54 * 108 = 314928 ≈ 315K



Issue 3: Lack of Software Support on Exploration

Seeking for a Platform where it could perform

- Extensive exploration of design space in parallel
- Auto-generating analyses across seeds and experiments
- Unifying implementation for node, edge, and graph-level tasks

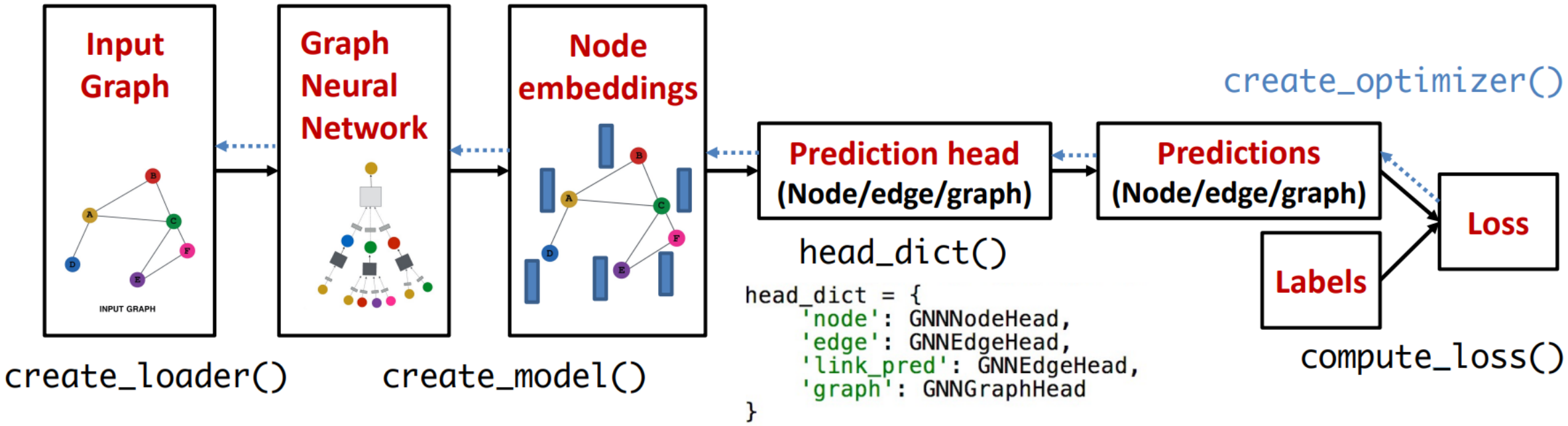


Task Level
Node
Graph
Edge

Additional Task Dimension

Software: GraphGym

Modularized GNN Pipeline



Register your modules and search for best hyper-parameters!

GraphGym: User Case (ID-GNN, You 2021)