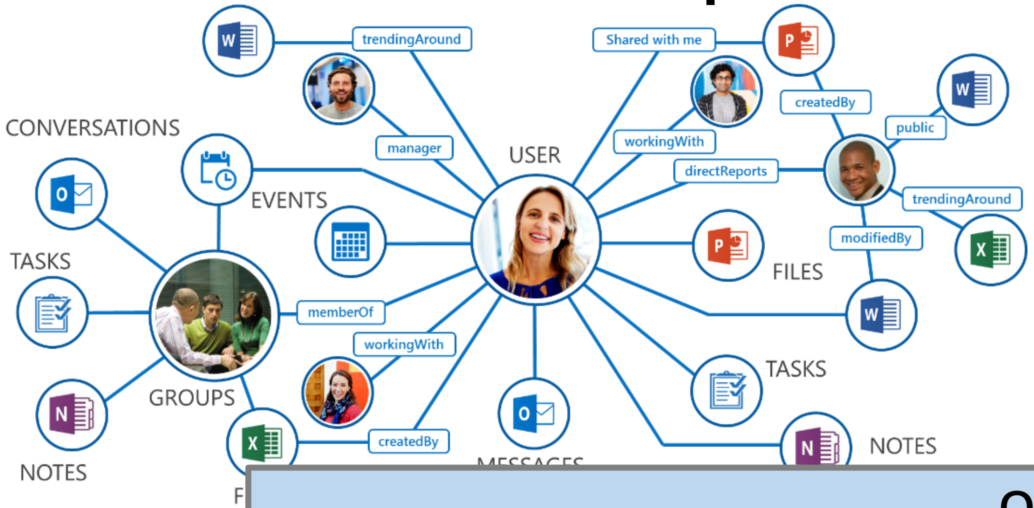
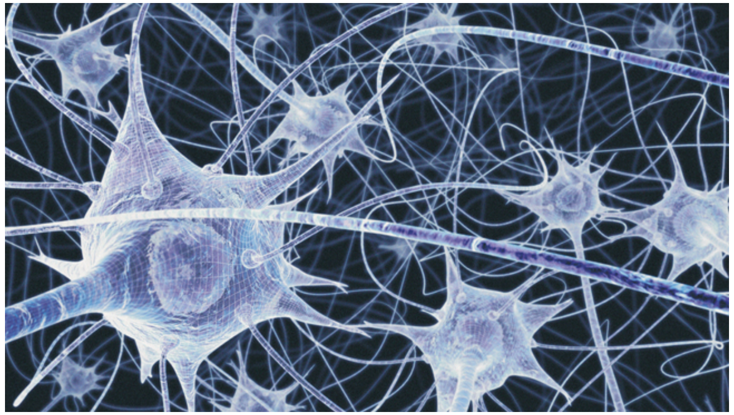
GCC: Graph Contrastive Coding for Graph Neural Network Pre-Training

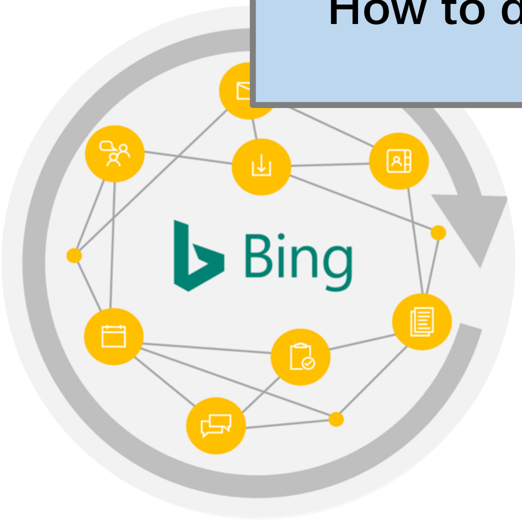
Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang,
Hongxia Yang, Ming Ding, Kuansan Wang, Jie Tang

Real-world Graphs

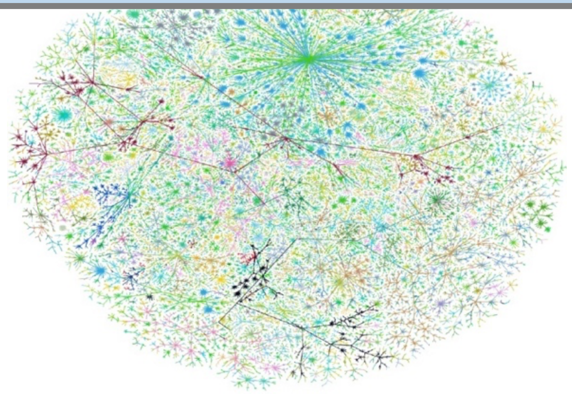




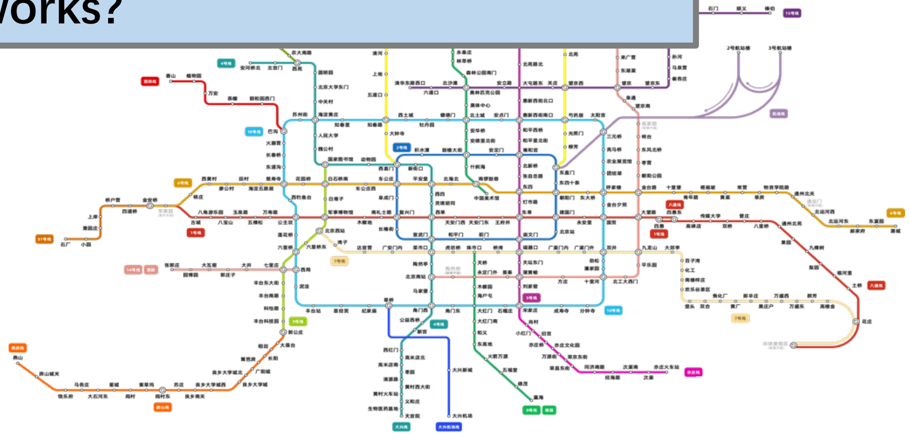
Question:
How to design machine learning models to learn the universal structural patterns across networks?



Knowledge Graph



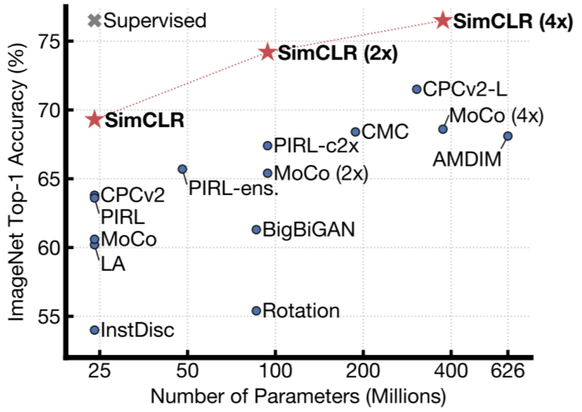
Internet Graph

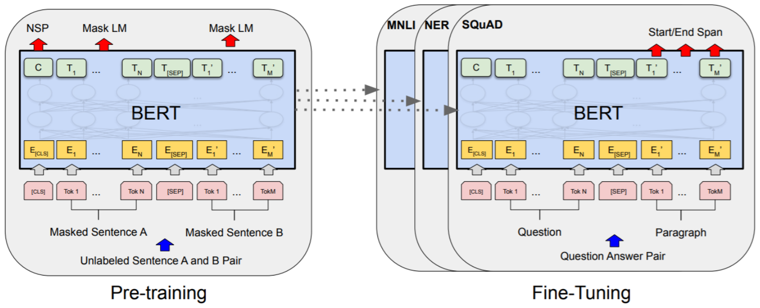


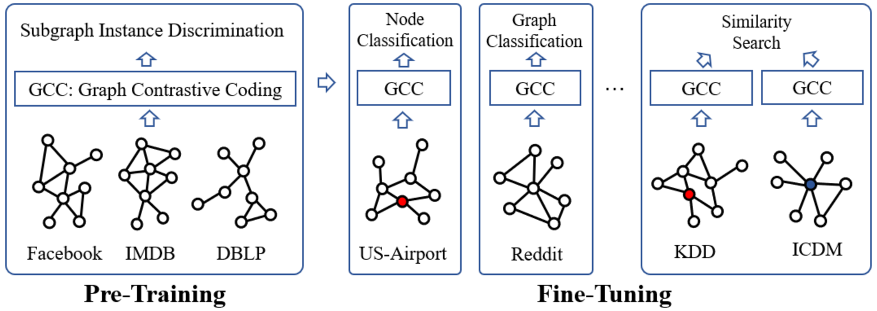
Transportation Graph

figure credit: Web

Pre-training and Fine-tuning







Computer Vision
ResNet
ImageNet

NLP
BERT
Wikipedia + Book corpus

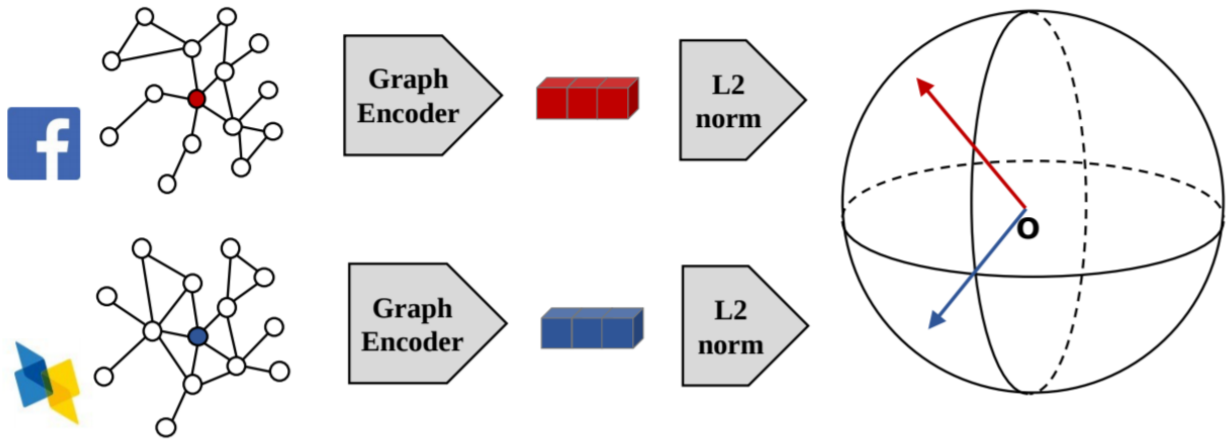
Graph Learning
GCC

Problem

GNN pre-training problem.

The GNN Pre-Training Problem

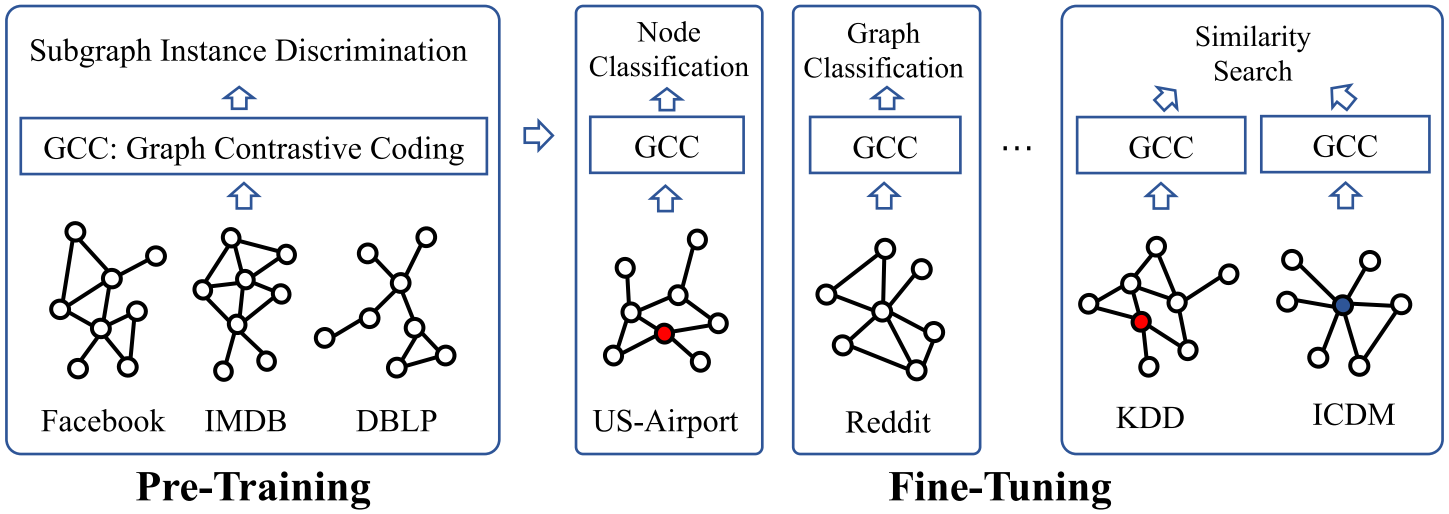
- Problem:
  - Learn a function f that maps a vertex to a low-dimensional vector
  - **Structural similarity**: map vertices with similar local network topologies close in the vector space
  - **Transferability**: compatible with vertices and graphs from various sources, even unseen during training time.



GCC Framework

Graph Contrastive Coding

Graph Contrastive Coding (GCC)



Hypothesis:
Graph structural patterns are universal and transferable across networks.

GCC Pre-training

- **Pre-training Task: Instance** Discrimination
- **InfoNCE Loss:** output **instance representations** that are capable of capturing the **similarities** between instances

$$\mathcal{L} = -\log \frac{\exp\left(\boldsymbol{q}^\top \boldsymbol{k}_+/\tau\right)}{\sum_{i=0}^{K}\exp\left(\boldsymbol{q}^\top \boldsymbol{k}_i/\tau\right)}$$

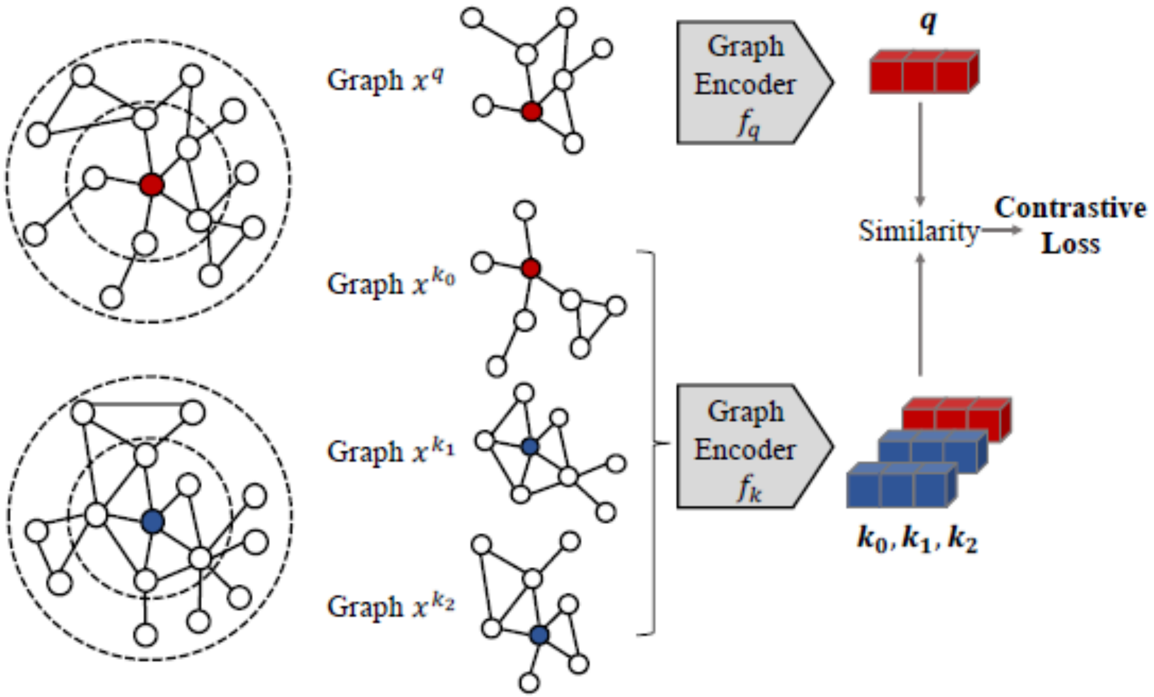
- query instance x^q
- query $\boldsymbol{q}$ (embedding of x^q), i.e., $\boldsymbol{q} = f(x^q)$
- dictionary of keys $\{\boldsymbol{k}_0, \boldsymbol{k}_1, \cdots, \boldsymbol{k}_K\}$
- key $\boldsymbol{k} = f(x^k)$

- Contrastive learning for graphs?
  - **Q1**: How to define **instances** in graphs?
  - **Q2**: How to define **(dis) similar instance** pairs?
  - **Q3**: What are the proper **encoders**?

GCC Pre-training

- **Q1**: How to define **instances** in graphs?
- **Q2**: How to define **(dis) similar instance**?
- **Q3**: What are the proper **encoders**?

$$\mathcal{L} = -\log \frac{\exp\left(\boldsymbol{q}^\top \boldsymbol{k}_+ / \tau\right)}{\sum_{i=0}^{K} \exp\left(\boldsymbol{q}^\top \boldsymbol{k}_i / \tau\right)}$$



GCC Pre-training: Learning Algorithms

- Optimizing Contrastive Loss
  - Encoded query $\boldsymbol{q}$
  - $K+1$ encoded keys $\{\boldsymbol{k}_0, \cdots, \boldsymbol{k}_K\}$

$$\mathcal{L} = -\log \frac{\exp\left(\boldsymbol{q}^\top \boldsymbol{k}_+ / \tau\right)}{\sum_{i=0}^{K} \exp\left(\boldsymbol{q}^\top \boldsymbol{k}_i / \tau\right)}$$

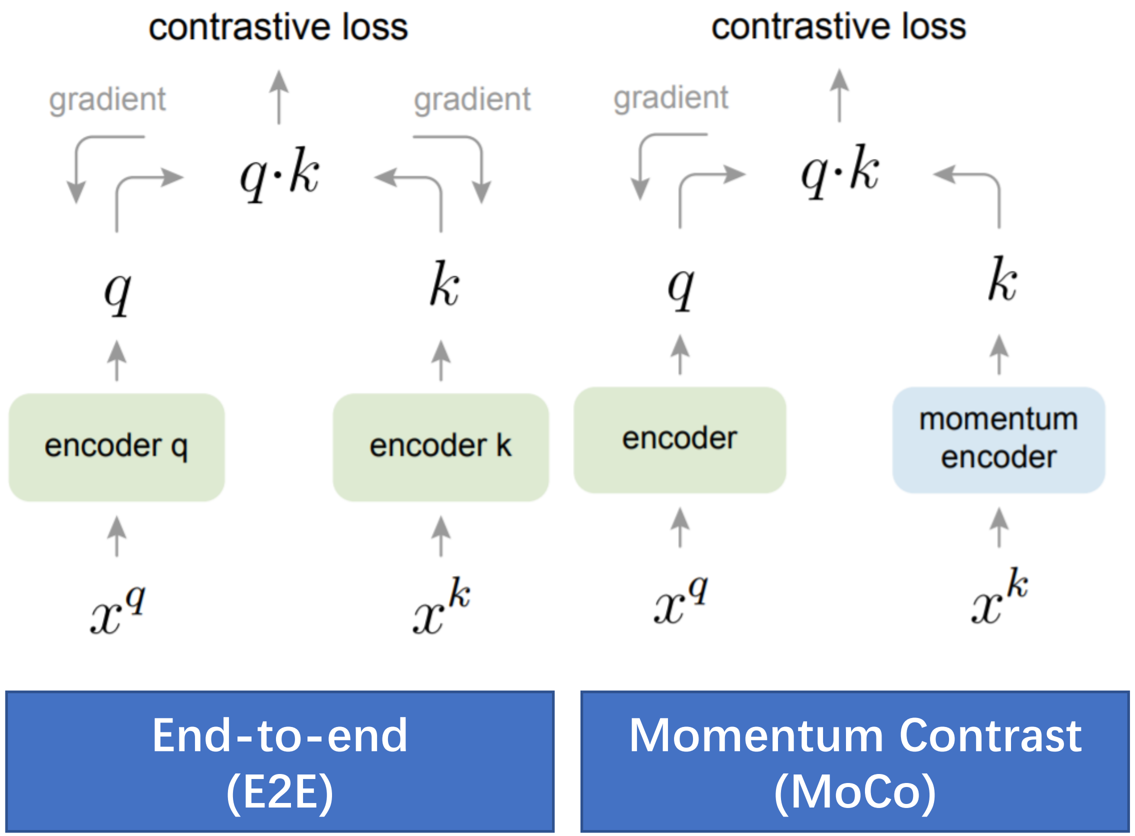
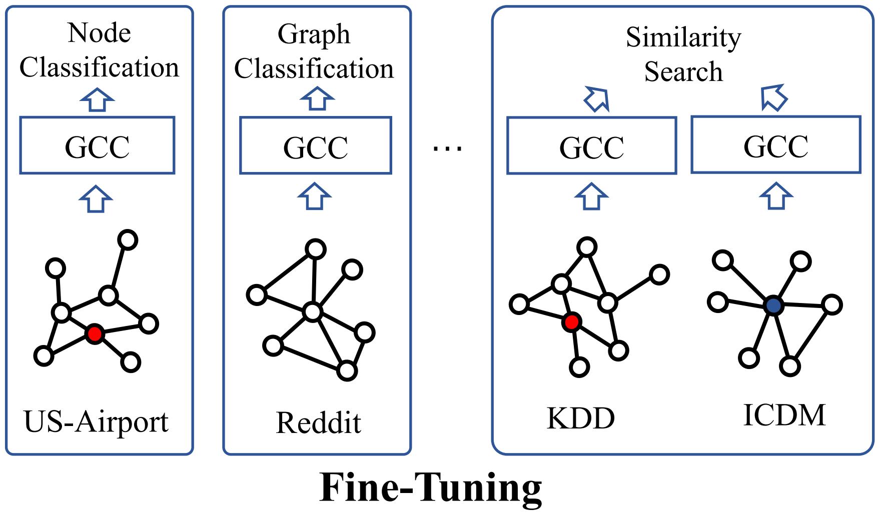
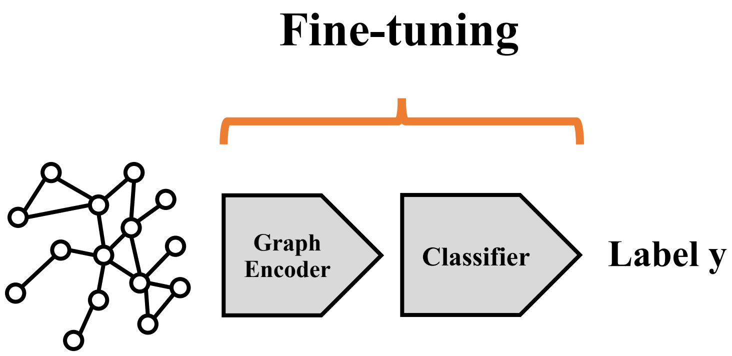


figure credit:
Momentum Contrast for Unsupervised Visual Representation Learning
arxiv.org/abs/1911.05722

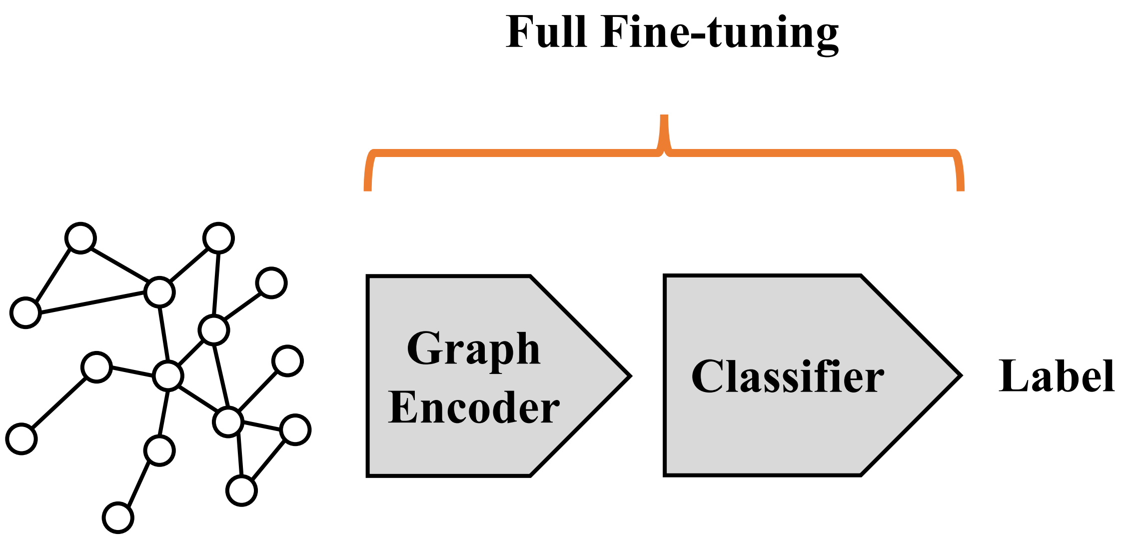
GCC Fine-tuning



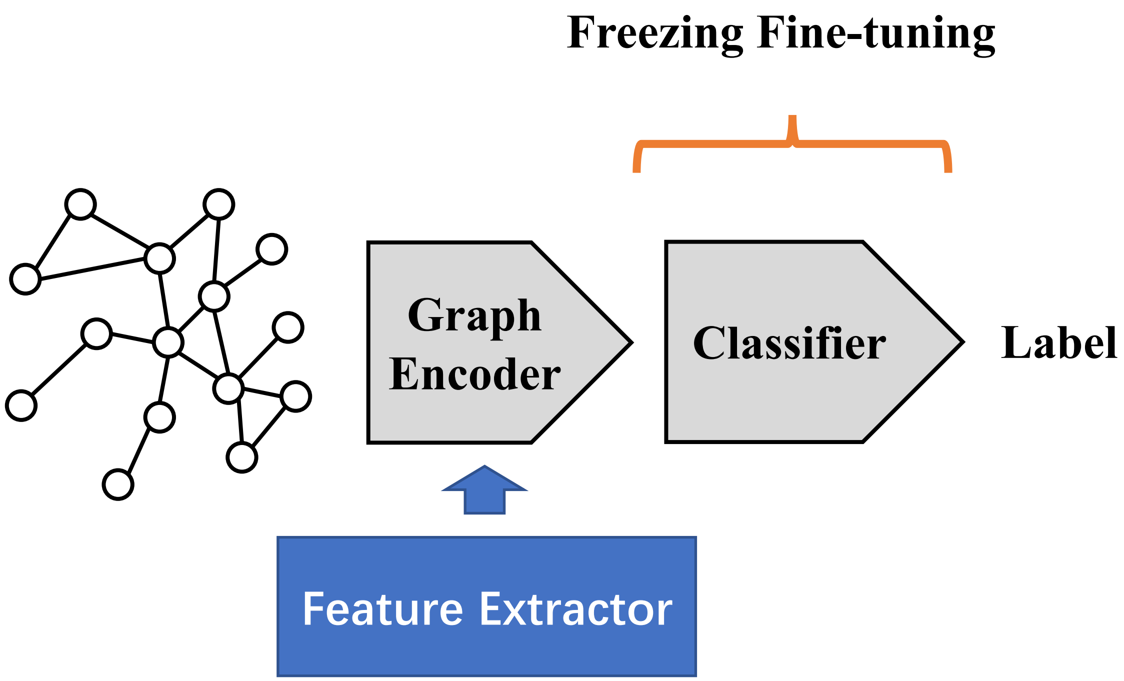


GCC Fine-tuning: Full v.s. Freezing

Full fine-tuning



Freezing fine-tuning



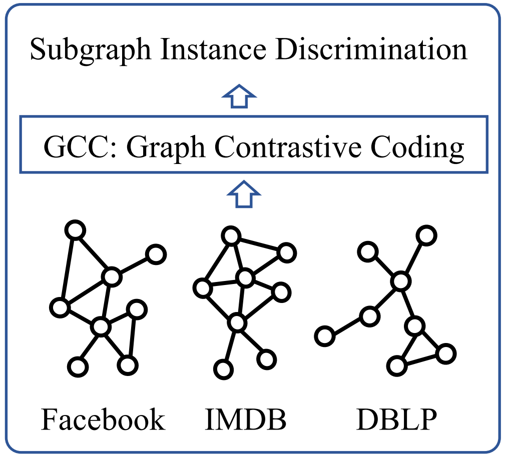
Experiments

GCC Pre-Training / Fine-tuning

- Six real-world information networks for pre-training.

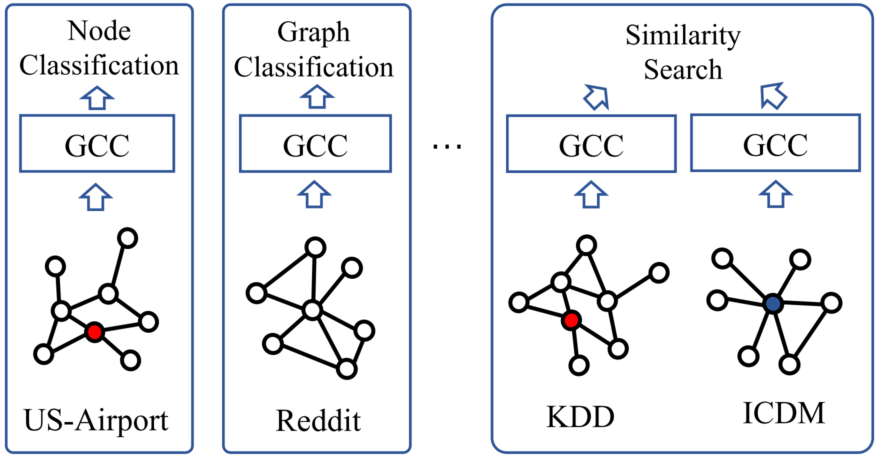
Table 1: Datasets for pre-training, sorted by number of vertices.

Dataset	Academia	DBLP (SNAP)	DBLP (NetRep)	IMDB	Facebook	LiveJournal
$\|V\|$	137,969	317,080	540,486	896,305	3,097,165	4,843,953
$\|E\|$	739,384	2,099,732	30,491,458	7,564,894	47,334,788	85,691,368



Pre-Training

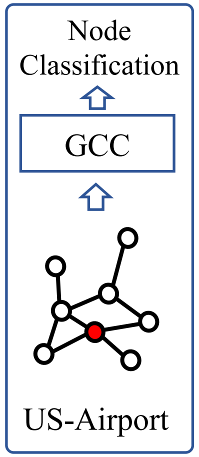
- Fine-tuning Tasks:
  - Node classification
  - Graph classification
  - Top-k Similarity search



Fine-Tuning

Task 1: Node Classification

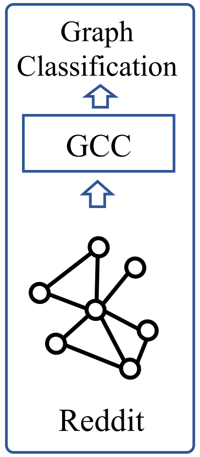
- Setup
  - US-Airport
  - AMiner academic graph



Datasets	US-Airport	H-index
$\|V\|$	1,190	5,000
$\|E\|$	13,599	44,020
ProNE	62.3	69.1
GraphWave	60.2	70.3
Struc2vec	**66.2**	> 1 Day
GCC (E2E, freeze)	64.8	**78.3**
GCC (MoCo, freeze)	65.6	75.2
GCC (rand, full)	64.2	76.9
GCC (E2E, full)	**68.3**	80.5
GCC (MoCo, full)	67.2	**80.6**

Task 2: Graph Classification

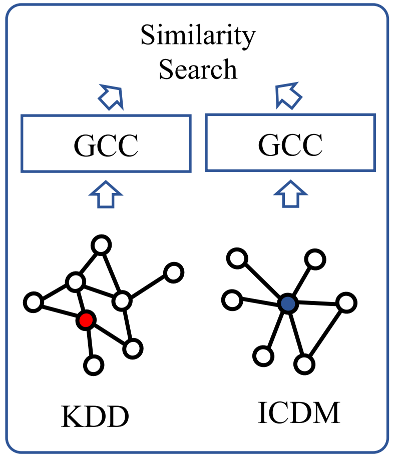
- Setup
  - COLLAB, RDT-B, RDT-M, & IMDB-B, IMDB-M



Datasets	IMDB-B	IMDB-M	COLLAB	RDT-B	RDT-M
# graphs	1,000	1,500	5,000	2,000	5,000
# classes	2	3	3	2	5
Avg. # nodes	19.8	13.0	74.5	429.6	508.5
DGK	67.0	44.6	73.1	78.0	41.3
graph2vec	71.1	50.4	–	75.8	47.9
InfoGraph	**73.0**	**49.7**	–	82.5	53.5
GCC (E2E, freeze)	71.7	49.3	74.7	87.5	52.6
GCC (MoCo, freeze)	72.0	49.4	**78.9**	**89.8**	**53.7**
DGCNN	70.0	47.8	73.7	–	–
GIN	**75.6**	**51.5**	80.2	**89.4**	**54.5**
GCC (rand, full)	**75.6**	50.9	79.4	87.8	52.1
GCC (E2E, full)	70.8	48.5	79.0	86.4	47.4
GCC (MoCo, full)	73.8	50.3	**81.1**	87.6	53.0

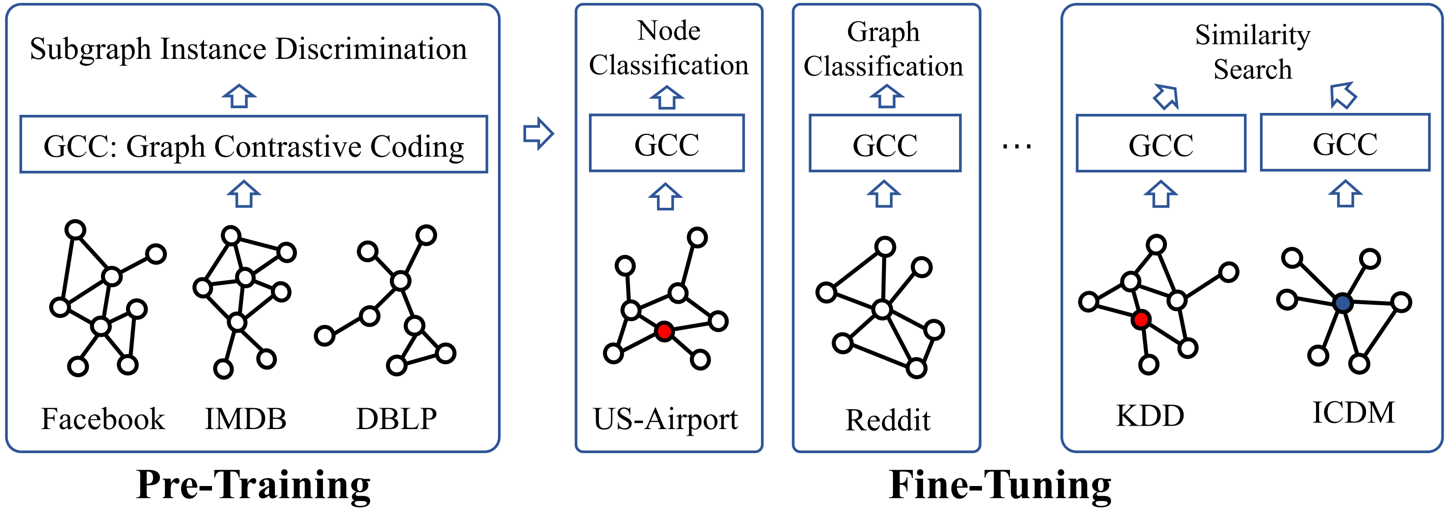
Task 3: Top-k Similarity Search

- Setup
  - AMiner academic graph



	KDD-ICDM		SIGIR-CIKM		SIGMOD-ICDE	
$\|V\|$	2,867	2,607	2,851	3,548	2,616	2,559
$\|E\|$	7,637	4,774	6,354	7,076	8,304	6,668
# groud truth		697		874		898
k	20	40	20	40	20	40
Random	0.0198	0.0566	0.0223	0.0447	0.0221	0.0521
RolX	0.0779	0.1288	0.0548	0.0984	0.0776	0.1309
Panther++	0.0892	0.1558	**0.0782**	0.1185	0.0921	0.1320
GraphWave	0.0846	**0.1693**	0.0549	0.0995	**0.0947**	**0.1470**
GCC (E2E)	**0.1047**	0.1564	0.0549	**0.1247**	0.0835	0.1336
GCC (MoCo)	0.0904	0.1521	0.0652	0.1178	0.0846	0.1425

Conclusion



- We study the pre-training of GNN with the goal of characterizing and transferring structural representations in social and information networks.
- We present Graph Contrastive Coding, which is a graph-based contrastive learning framework to pre-train GNN.
- The pre-trained GNN achieves competitive performance to its supervised trained-from-scratch counterparts in 3 graph learning tasks on 10 graph datasets.

Thanks.

Q&A

https://github.com/THUDM/GCC

Find us at KDD 2020

https://github.com/THUDM/GCC