Graph Few-shot Learning

Chuxu Zhang
Computer Science Department
Brandeis University

Introduction

- Dr. Chuxu Zhang

Assistant Professor, Computer Science, Brandeis University

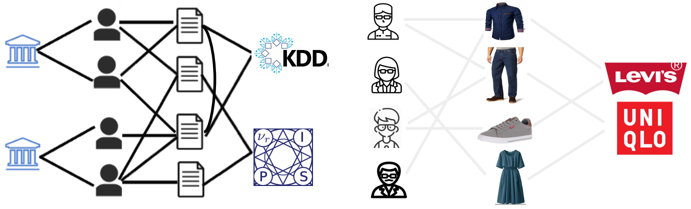
General research: machine learning, deep learning, data mining
Current focus: graph mining and learning, recommender systems, time series & spatial-temporal learning
Applications: social and knowledge systems, cybersecurity, healthcare

Awards/Honors: CIKM 2021 Best Paper Award, WWW 2019 Best Paper Shortlist, APWeb/WAIM 2016 Best Student Paper Award, etc.

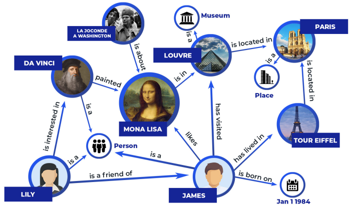
Webpage: https://chuxuzhang.github.io
Contact: chuxuzhang@brandeis.edu

❑ Graph Data and Applications

Social and Knowledge Systems

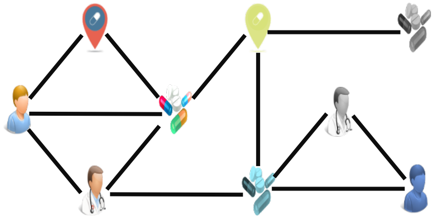


classification and recommendation



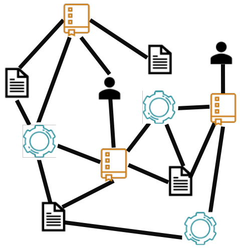
relation reasoning

Healthcare and Biomedicine

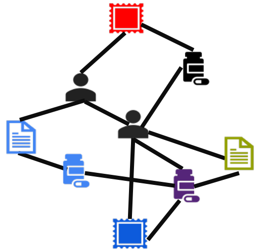


patient risk behavior prediction

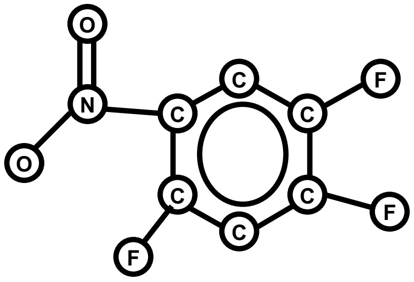
Information and Cybersecurity



malware detection



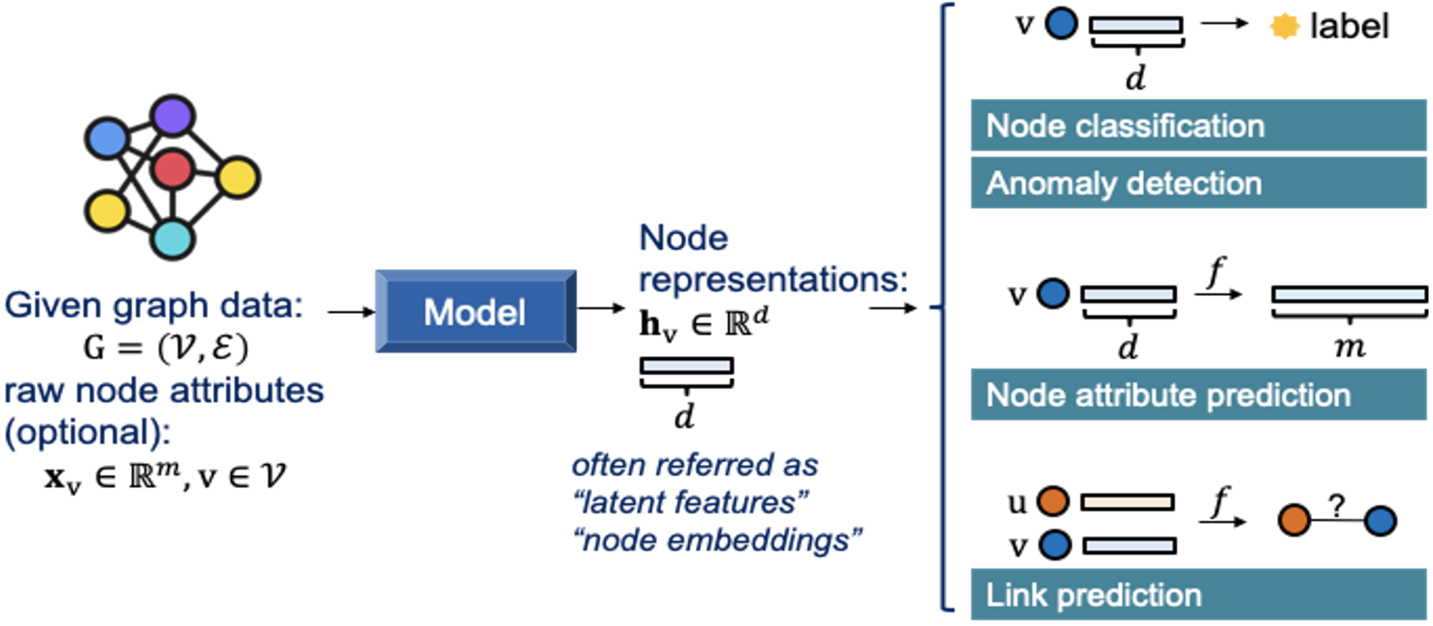
drug trafficker detection



molecular property prediction

Introduction

❑ Graph Representation Learning

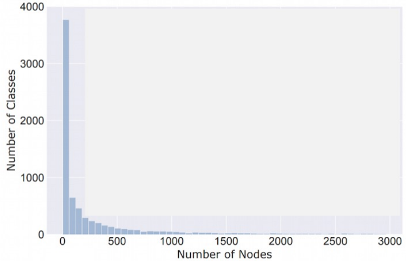


often require large labeled data

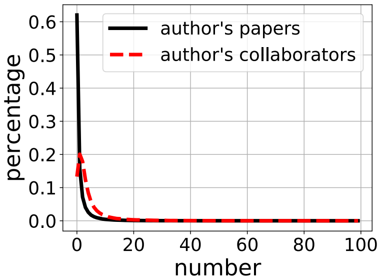
+ Graph Classification

Introduction

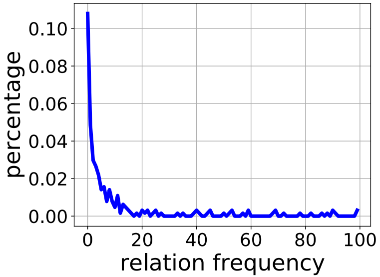
❑ Challenge: Small Labeled Graph Data



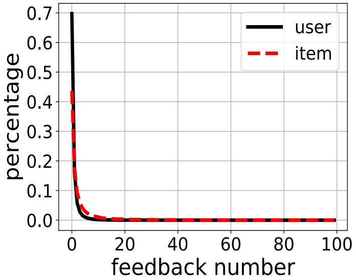
(a) number of labeled nodes in social graph



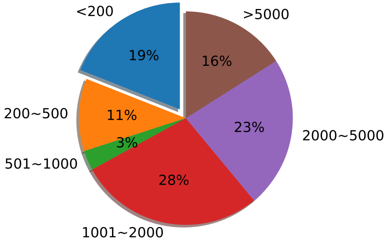
(b) author's paper/collaborator number in academic graph



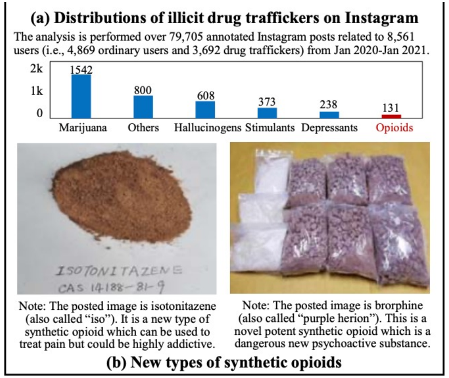
(c) relation frequency in knowledge graph



(d) feedback number of user/item in E-commerce graph



(e) number of molecular graph tested in experimental assays

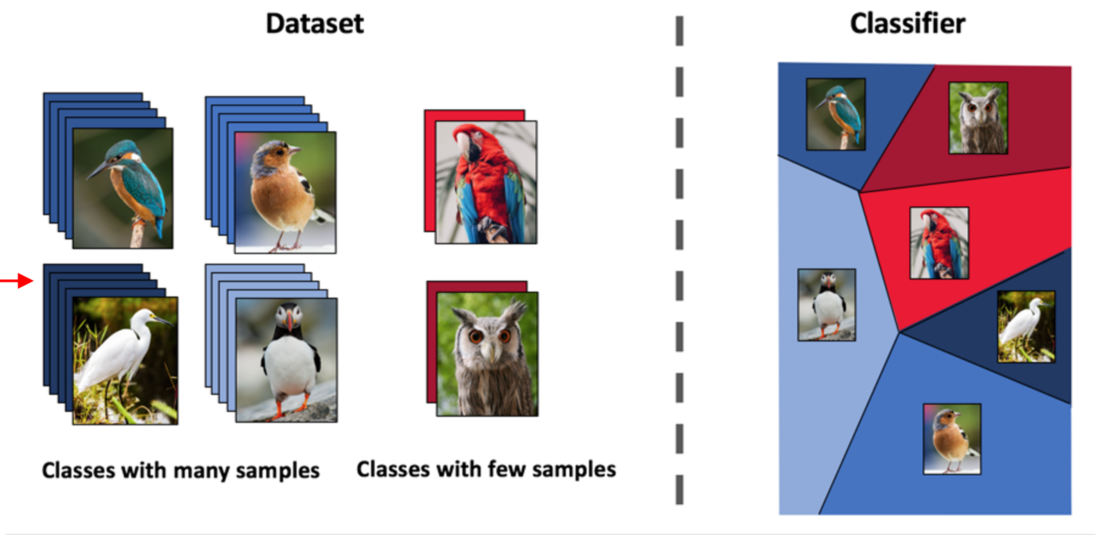


❑ Few-shot Learning

Concept: ML problems with little supervised information (few-shot labels)

Solution: Learn meta-knowledge from auxiliary data

auxiliary data



❑ Few-shot Learning (Meta-learning)

N-way K-shot Learning (N: class number, K: labeled data size - support set size)

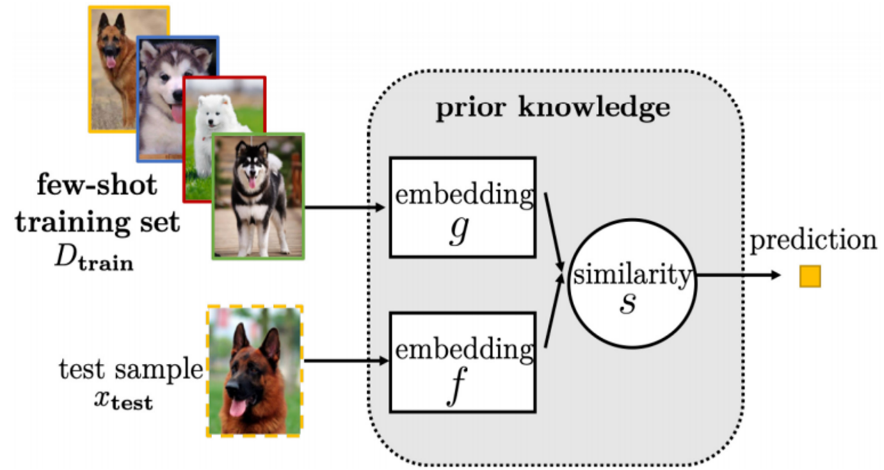
Goal: use auxiliary data to mimic the target few-shot task

5-way 1-shot image classification (MiniImagenet)

Introduction

❑ Metric-based Methods (e.g., Prototypical Net, Matching Net)

Measure the similarity between support samples and query samples, learn embedding function for new few-shot tasks

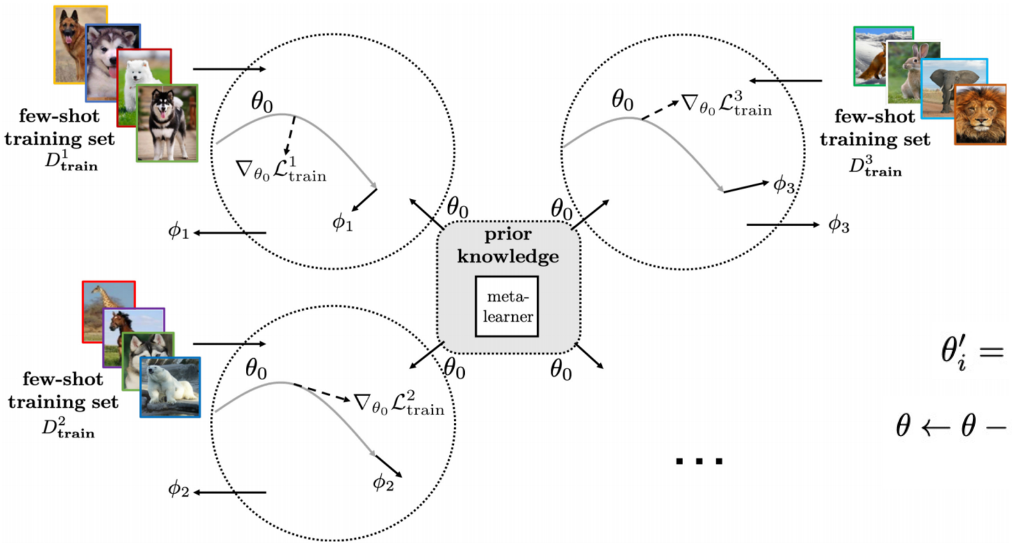


Generalizing from a Few Examples: A Survey on Few-shot Learning, ACM Computing Surveys, 2020

Introduction

❑ Gradient-based Methods (e.g., MAML)

Learn a good parameter initialization, fast adapt to new few-shot tasks



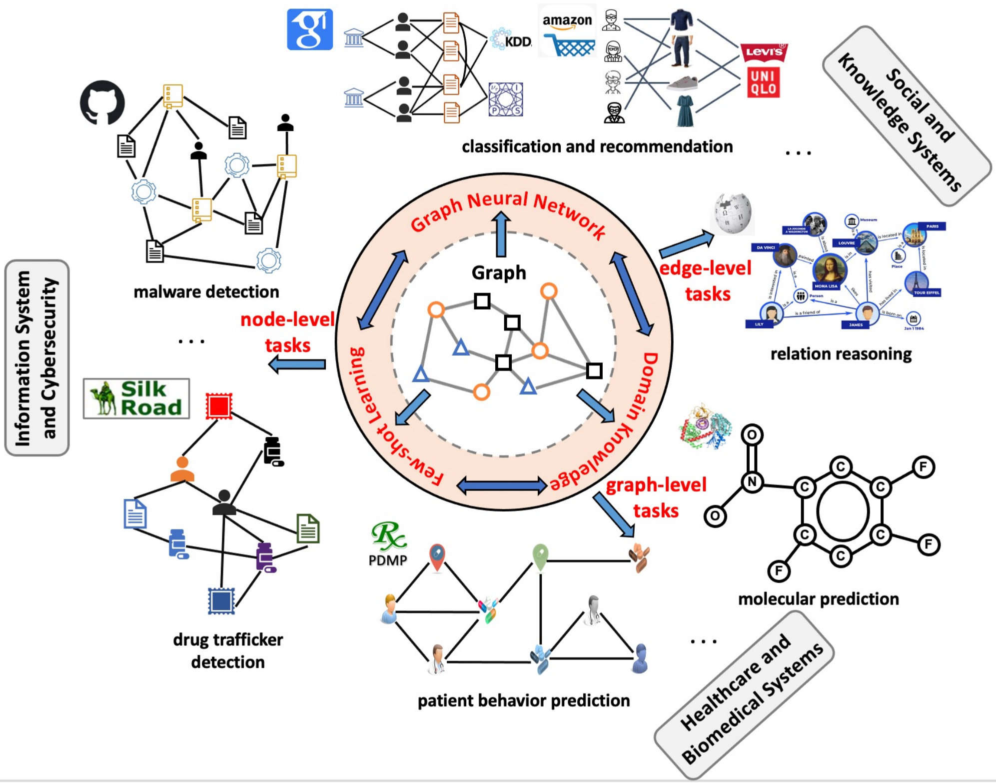
$$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta).$$

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'})$$

Generalizing from a Few Examples: A Survey on Few-shot Learning, ACM Computing Surveys, 2020

Introduction

❑ Graph Few-shot Learning



Node-level Learning Tasks

❑ Part-1: Node-level Few-shot Learning

❖ Node Classification

GFL (AAAI 2020)

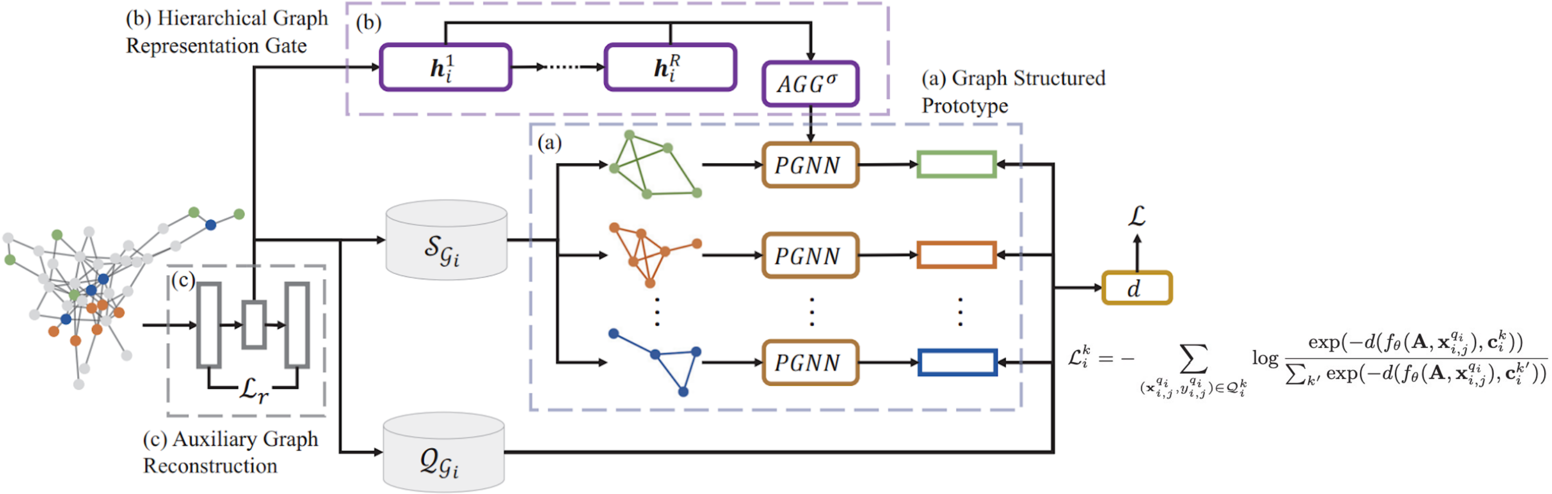
❖ Domain Applications - Anomaly Detection

Meta-AHIN (IJCAI 2021): Malware Detection

MetaHG (NeurIPS 2021): Drug Trafficker Detection

Graph Few-shot Learning via Knowledge Transfer

Given: labeled nodes (with attributes) from training graphs/labels
Predict: labels of query nodes of test graphs/labels with **few-shot** support nodes (labeled nodes)
Approach: metric-based method (prototypical net)
Motivation/Assumption: knowledge from auxiliary graphs could be useful
Key idea: incorporate meta-knowledge from auxiliary graphs/labels via knowledge transfer



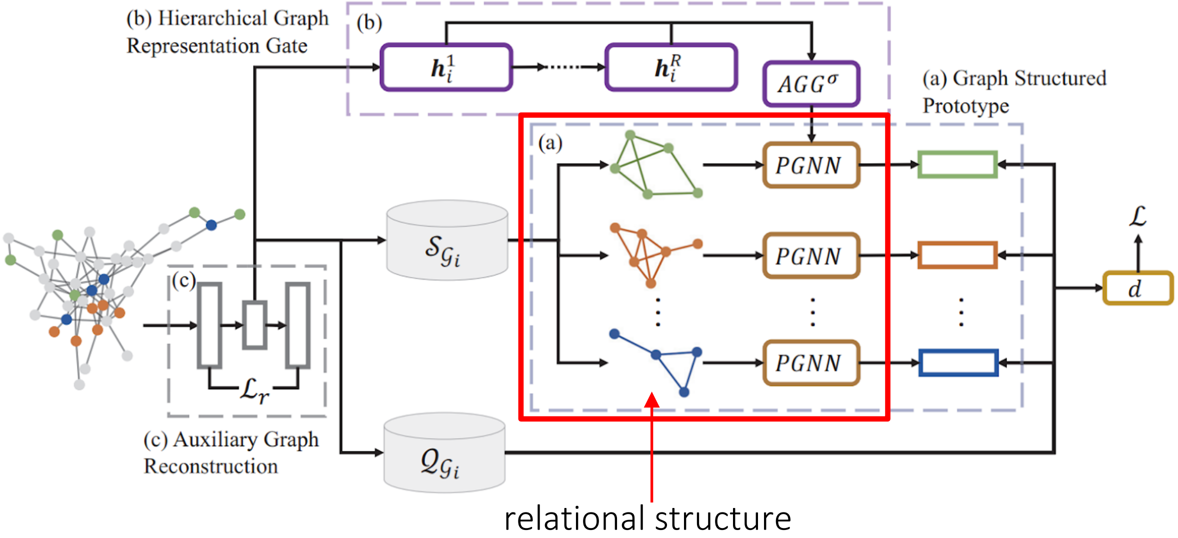
Graph Few-shot Learning via Knowledge Transfer, AAAI 2020

Graph Few-shot Learning via Knowledge Transfer

Step (a): compute graph-structured prototypes

Intuition: extract structures to describe interactions among support nodes of each k

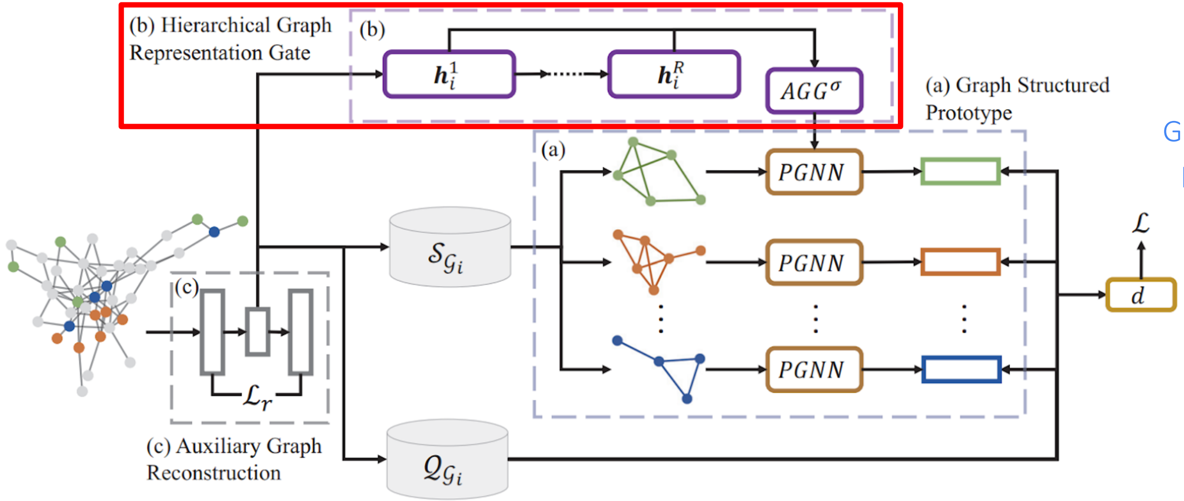
Detail: extract the relational structure (nodes of class k) and learn prototype via GNN



Graph Few-shot Learning via Knowledge Transfer, AAAI 2020

Graph Few-shot Learning via Knowledge Transfer

Step (b): generate hierarchical graph representation gate
Intuition: tailor the globally shared parameter to each graph
Detail: Learn graph-specific parameters via graph pooling and gating function



Graph-specific parameters

Globally shared parameters

$$\boxed{\phi_i} = \mathbf{g}_i \circ \phi = \mathcal{T}(\mathbf{h}_i) \circ \boxed{\phi}$$

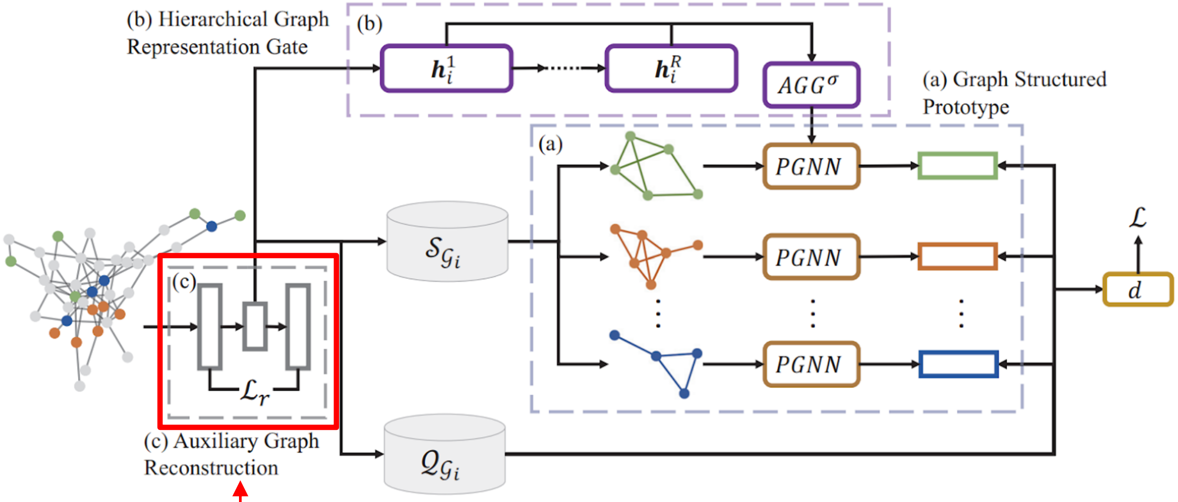
$$\mathbf{g}_i = \mathcal{T}(\mathbf{h}_i) = \sigma(\mathbf{W}_g \boxed{\mathbf{h}_i} + \mathbf{b}_g)$$

Aggregated graph representation

Graph Few-shot Learning via Knowledge Transfer, AAAI 2020

Graph Few-shot Learning via Knowledge Transfer

Step (c): compute auxiliary graph reconstruction loss
Intuition: improve node representations with self-supervised reconstruction loss
Detail: reconstruct the graph with a graph autoencoder



$$\mathcal{L}_r(\mathbf{A}_i, \mathbf{X}_i) = \|\mathbf{A}_i - \mathrm{GNN}_{dec}(\mathbf{Z}_i)\mathrm{GNN}_{dec}^T(\mathbf{Z}_i)\|_F^2$$

Graph Few-shot Learning via Knowledge Transfer, AAAI 2020

Node classification results (ACC) on four datasets (10-shot)

Dataset	Colla.	Reddit	Cita.	Pubmed
# Nodes (avg.)	4,496	5,469	2,528	2,901
# Edges (avg.)	14,562	7,325	14,710	5,199
# Features/Node	128	600	100	500
# Classes	4	5	3	3
# Graphs (Meta-train)	100	150	30	60
# Graphs (Meta-val.)	10	15	3	5
# Graphs (Meta-test)	20	25	10	15

Model	Collaboration	Reddit	Citation	Pubmed
LP (Zhu and Ghahramani 2002)	$61.09 \pm 1.36\%$	$23.40 \pm 1.63\%$	$67.00 \pm 4.50\%$	$48.55 \pm 6.01\%$
Planetoid (Yang, Cohen, and Salakhudinov 2016)	$62.95 \pm 1.23\%$	$50.97 \pm 3.81\%$	$61.94 \pm 2.14\%$	$51.43 \pm 3.98\%$
Deepwalk (Perozzi, Al-Rfou, and Skiena 2014)	$51.74 \pm 1.59\%$	$34.81 \pm 2.81\%$	$56.56 \pm 5.25\%$	$44.33 \pm 4.88\%$
node2vec (Grover and Leskovec 2016)	$59.77 \pm 1.67\%$	$43.57 \pm 2.23\%$	$54.66 \pm 5.16\%$	$41.89 \pm 4.83\%$
Non-transfer-GCN (Kipf and Welling 2017)	$63.16 \pm 1.47\%$	$46.21 \pm 1.43\%$	$63.95 \pm 5.93\%$	$54.87 \pm 3.60\%$
All-Graph-Finetune (AGF)	$76.09 \pm 0.56\%$	$54.13 \pm 0.57\%$	$88.93 \pm 0.72\%$	$83.06 \pm 0.72\%$
K-NN	$67.53 \pm 1.33\%$	$56.06 \pm 1.36\%$	$78.18 \pm 1.70\%$	$74.33 \pm 0.52\%$
Matchingnet (Vinyals et al. 2016)	$80.87 \pm 0.76\%$	$56.21 \pm 1.87\%$	$94.38 \pm 0.45\%$	$85.65 \pm 0.21\%$
MAML (Finn, Abbeel, and Levine 2017)	$79.37 \pm 0.41\%$	$59.39 \pm 0.28\%$	$95.71 \pm 0.23\%$	$88.44 \pm 0.46\%$
Protonet (Snell, Swersky, and Zemel 2017)	$80.49 \pm 0.55\%$	$60.46 \pm 0.67\%$	$95.12 \pm 0.17\%$	$87.90 \pm 0.54\%$
GFL-mean (Ours)	$83.51 \pm 0.38\%$	$62.66 \pm 0.57\%$	$\mathbf{96.51 \pm 0.31\%}$	$\mathbf{89.37 \pm 0.41\%}$
GFL-att (Ours)	$\mathbf{83.79 \pm 0.39\%}$	$\mathbf{63.14 \pm 0.51\%}$	$95.85 \pm 0.26\%$	$88.96 \pm 0.43\%$

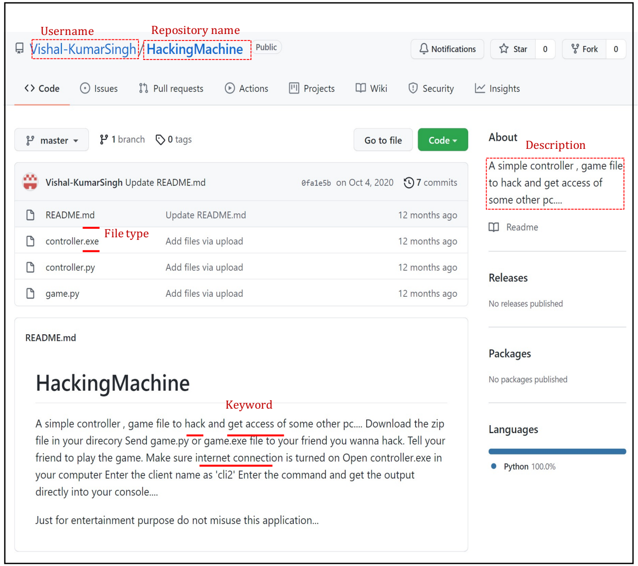
Observation: results indicate the effectiveness of GFL

Graph Few-shot Learning via Knowledge Transfer, AAAI 2020

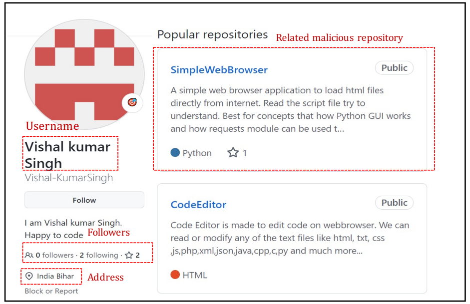
Adapting Meta Knowledge with HIN for Malicious Repository Detection

Given: different types of labeled code repositories (software) along with their content/context information in online coding platforms (e.g., GitHub)

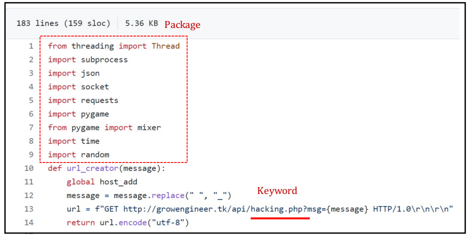
Predict: malicious repositories of new type with **few** samples, e.g., COVID-19



(a) A showcase of malicious repository



(b) The corresponding user profile



(c) The unfolded script 'game.py'

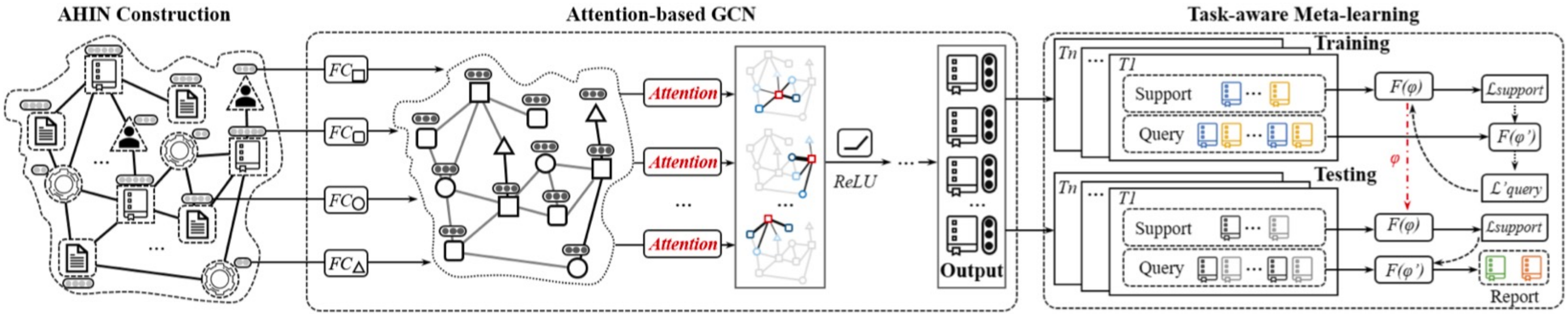
Adapting Meta Knowledge with Heterogeneous Information Network for COVID-19 Themed Malicious Repository Detection, IJCAI 2021

Adapting Meta Knowledge with HIN for Malicious Repository Detection

Approach: gradient-based method

Motivation: knowledge from different types of code repositories (constructed as HIN) is useful

Key idea: leverage meta knowledge in HIN to fast adaptation for new type repository



(a) Model Framework

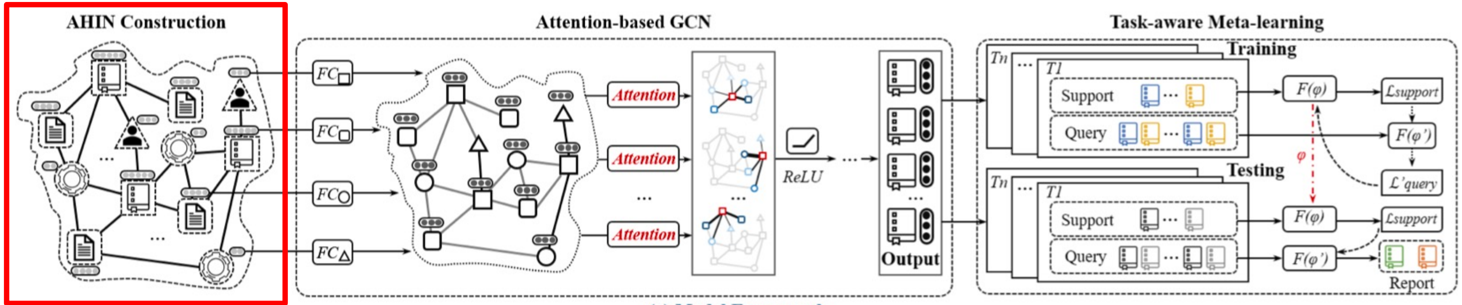
Adapting Meta Knowledge with Heterogeneous Information Network for COVID-19 Themed Malicious Repository Detection, IJCAI 2021

Adapting Meta Knowledge with HIN for Malicious Repository Detection

Step (a): HIN (HG) construction

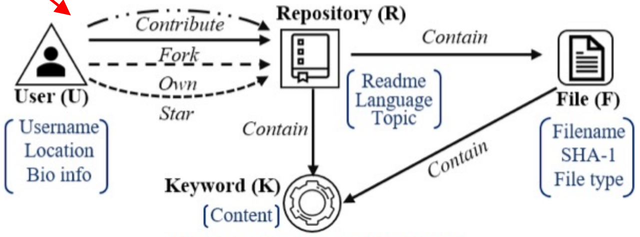
Intuition: besides content, structural information among entities is useful

Detail: construct a HIN to depict structural relation and content information in GitHub



(a) Model Framework

domain
information



(b) Network Schema for AHIN

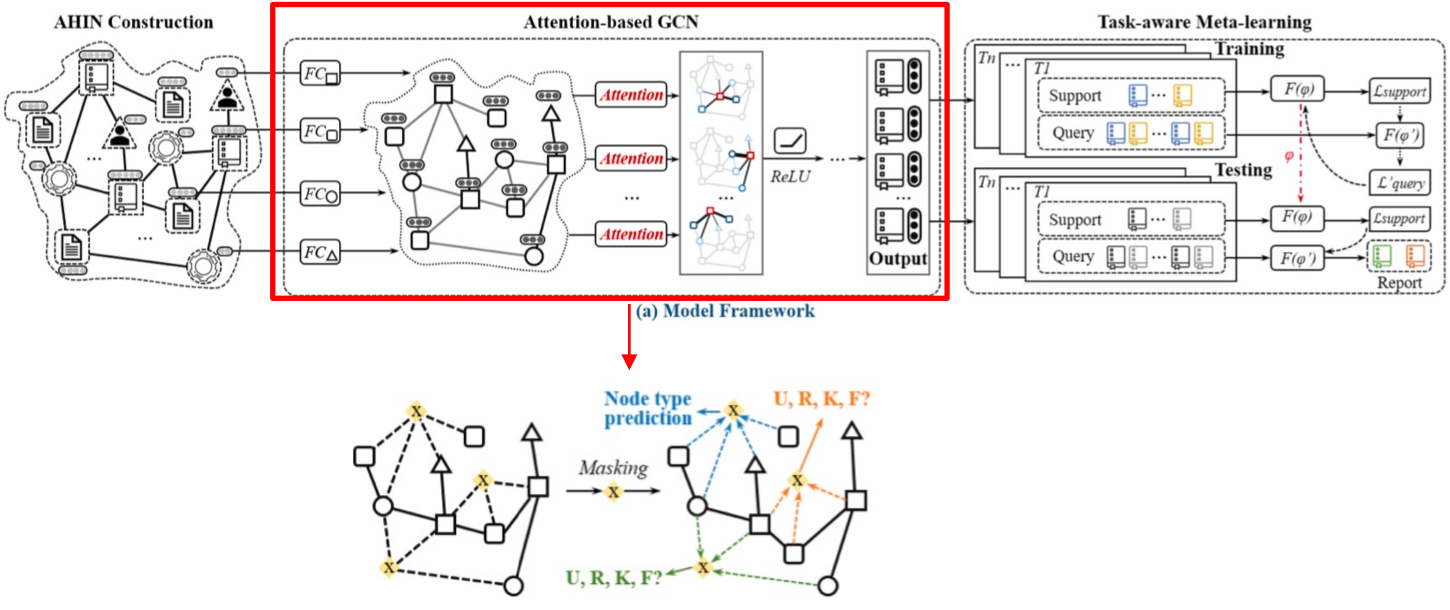
Content Feature	
Repository	readme, program language, topic keywords
Keyword	keywords from every source code file
File	filename, SHA-1, and type of file
User	username, location, bio information
Relation Information	
R1	repository-contributed-by-user
R2	repository-forked-by-user
R3	repository-owned-by-user
R4	repository-stared-by-user
R5	repository-contain-file
R6	file-contain-keyword
R7	repository-contain-keyword

Adapting Meta Knowledge with Heterogeneous Information Network for COVID-19 Themed Malicious Repository Detection, IJCAI 2021

Adapting Meta Knowledge with HIN for Malicious Repository Detection

Step (b): HIN (HG) representation learning

Detail: learn repository embedding through GAT with self-supervised signal



(a) Model Framework

(c) Self-Supervised Learning Module

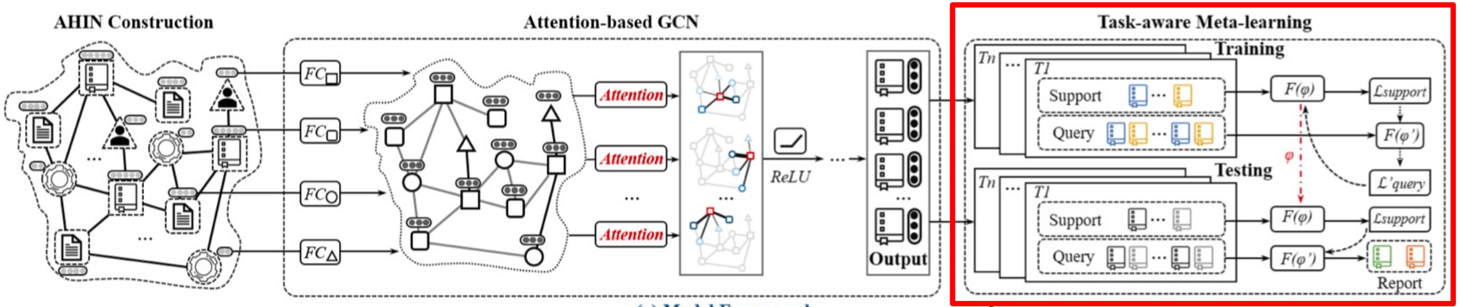
Adapting Meta Knowledge with Heterogeneous Information Network for COVID-19 Themed Malicious Repository Detection, IJCAI 2021

Adapting Meta Knowledge with HIN for Malicious Repository Detection

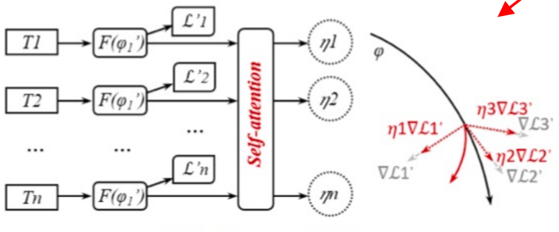
Step (c): task-aware meta-learning

Intuition: different tasks (repository types) have different contributions for test task

Detail: introduce task weight in MAML via self-attention



(a) Model Framework



(d) Task-aware MAML

$$\phi'_\tau = \phi - \alpha \nabla_\phi \mathcal{L}_\tau(\phi)$$

$$\phi \leftarrow \phi - \beta \nabla_\phi \sum_{\tau \in \mathcal{T}} \boxed{O_\tau} \mathcal{L}_\tau(\phi'_\tau)$$

Task weight

$$O_\tau = \frac{\exp(\mathrm{MLP}(\mathbf{F}_\tau))}{\sum_{\tau' \in \mathcal{T}} \exp(\mathrm{MLP}(\mathbf{F}_{\tau'}))}$$

Adapting Meta Knowledge with Heterogeneous Information Network for COVID-19 Themed Malicious Repository Detection, IJCAI 2021

Adapting Meta Knowledge with HIN for Malicious Repository Detection

Malware detection results on GitHub data

Setting		1-shot		3-shot		5-shot		8-shot		15-shot	
Group	Model	F1	ACC	F1	ACC	F1	ACC	F1	ACC	F1	ACC
B1	Feature+LR	0.4566	0.5169	0.4640	0.5296	0.4639	0.5432	0.4869	0.5524	0.5313	0.5597
	Feature+DNN	0.4843	0.5266	0.4918	0.5308	0.4931	0.5554	0.5016	0.5591	0.5564	0.5770
B2	Feature+Protonet	0.4853	0.5278	0.5088	0.5524	0.5259	0.5765	0.6241	0.6438	0.6517	0.6722
	Feature+Matching	0.4975	0.5489	0.5325	0.5905	0.5478	0.6253	0.6325	0.6758	0.6417	0.6959
	Feature+MAML	0.5268	0.5771	0.5646	0.6204	0.5685	0.6527	0.6653	0.7178	0.6835	0.7356
B3	Deepwalk+DNN	0.4834	0.5210	0.4932	0.5360	0.5056	0.5458	0.5184	0.5691	0.5646	0.5964
	metapath2vec+DNN	0.5074	0.5243	0.5137	0.5508	0.5327	0.5687	0.5378	0.5731	0.5731	0.6111
	GCN+DNN	0.4989	0.5178	0.5255	0.5543	0.5541	0.5935	0.5848	0.6002	0.6190	0.6304
	GAT+DNN	0.5021	0.5160	0.5231	0.5500	0.5713	0.5920	0.5918	0.6280	0.6348	0.6415
B4	Deepwalk+MAML	0.5523	0.6049	0.5940	0.6505	0.6248	0.6867	0.6940	0.7500	0.7303	0.7693
	metapath2vec+MAML	0.5892	0.6399	0.6446	0.6835	0.6771	0.7145	0.7316	0.7788	0.7729	0.7960
	GCN+MAML	0.6376	0.6836	0.7051	0.7236	0.7464	0.7737	0.7718	0.8058	0.8043	0.8248
	GAT+MAML	0.6139	0.6661	0.6903	0.7129	0.7404	0.7622	0.7846	0.8147	0.8209	0.8382
Ours	**Meta-AHIN**	**0.7006**	**0.7173**	**0.7747**	**0.7768**	**0.8149**	**0.8283**	**0.8591**	**0.8687**	**0.8640**	**0.8851**

Method	Version	F1	ACC
LGTM	-	0.1459	0.3592
DrWeb	7.0.46.3050	0.5958	0.6174
McAfee	6.0.6.653	0.6187	0.6360
Avast	18.4.3895.0	0.6679	0.6881
Meta-AHIN	-	**0.8640**	**0.8851**

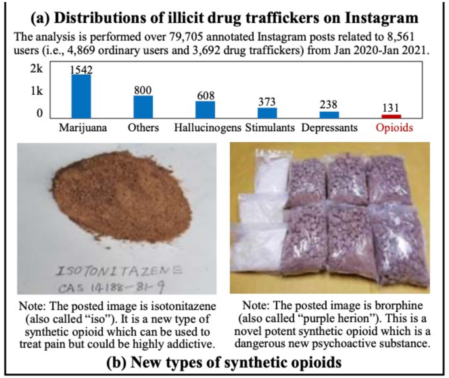
Observation: results indicate the effectiveness of Meta-AHIN

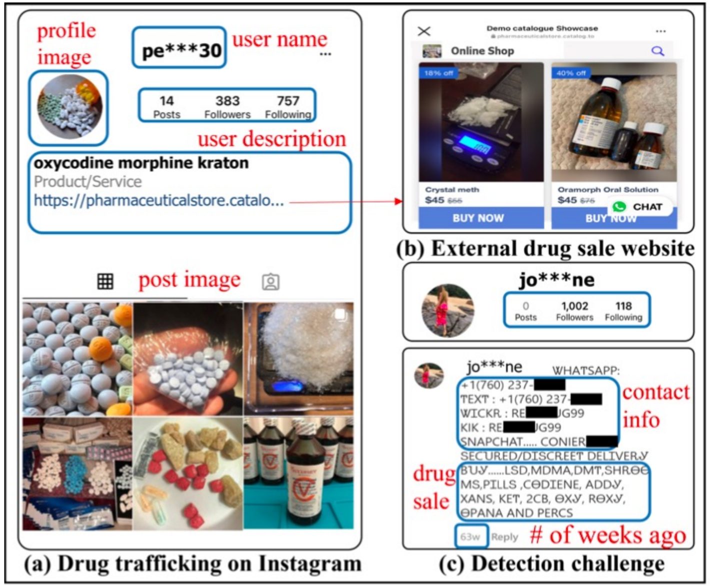
Adapting Meta Knowledge with Heterogeneous Information Network for COVID-19 Themed Malicious Repository Detection, IJCAI 2021

Distilling Meta Knowledge on HG for Illicit Drug Trafficker Detection

Given: different types of drug traffickers along with their content/context information in social media (e.g., Instagram, Twitter)

Predict: illicit drug traffickers of new drug type with **few** labeled samples, e.g., opioids

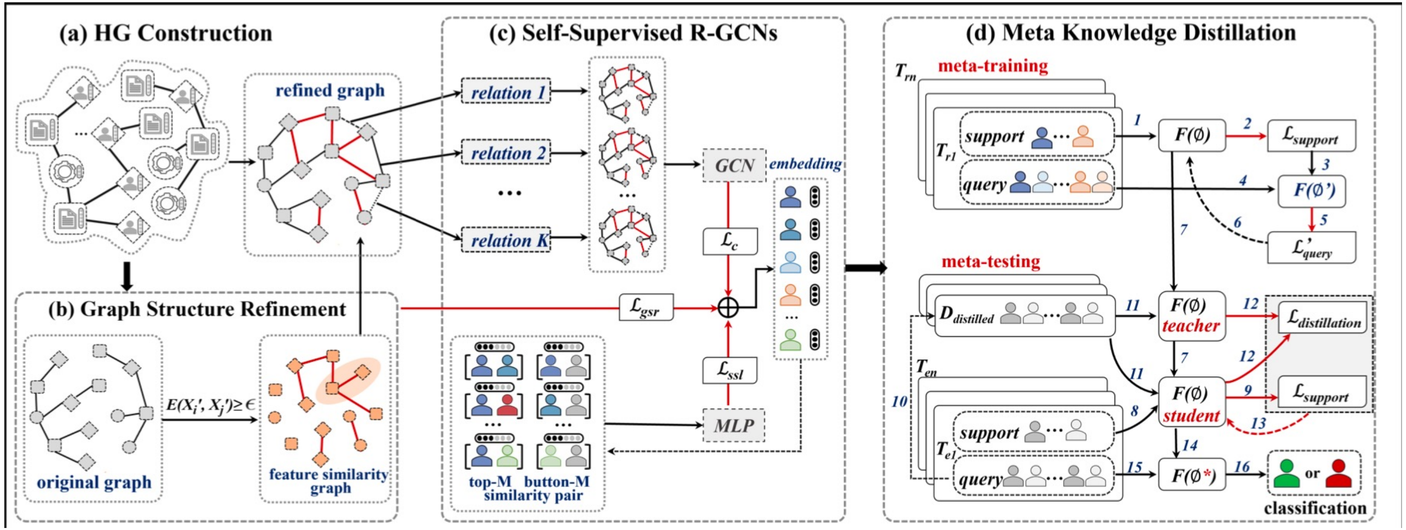




Distilling Meta Knowledge on Heterogeneous Graph for Illicit Drug Trafficker Detection on Social Media, NeurIPS 2021

Distilling Meta Knowledge on HG for Illicit Drug Trafficker Detection

Setting: gradient-based method
Motivation: knowledge from different types of drug traffickers (HG) is useful
Key idea: leverage meta knowledge in HG to fast adaptation for new type trafficker



Distilling Meta Knowledge on Heterogeneous Graph for Illicit Drug Trafficker Detection on Social Media, NeurIPS 2021

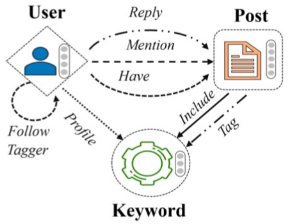
Distilling Meta Knowledge on HG for Illicit Drug Trafficker Detection

Step (a): HG construction with refinement

Intuition: besides content, structural information among entities is useful

Detail: construct a HG to depict structural relation and content information in Instagram, refine graph with structure learning (similarity graph)

domain information



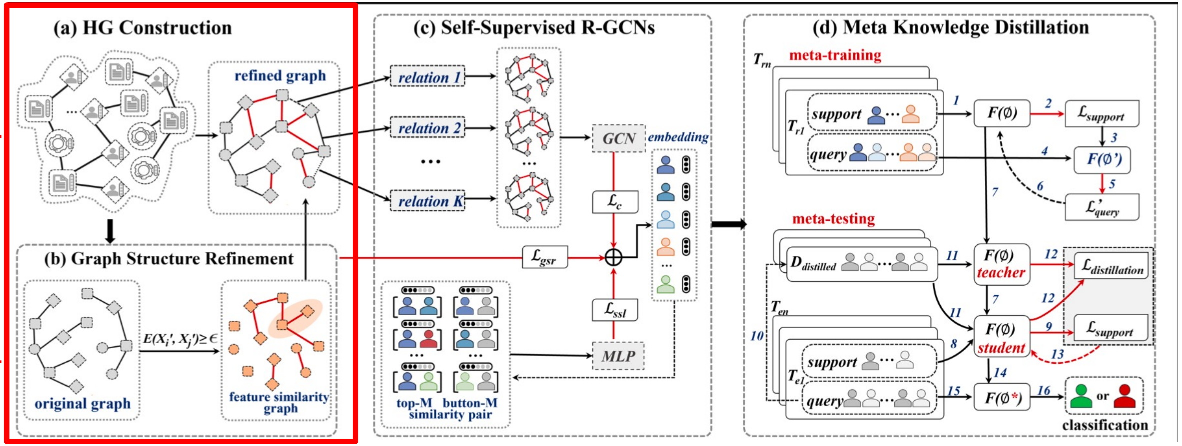
Content Feature	
User	username, profile information,is-business, # of followers/followings/posts
Post	text content, image information comment content, # of likes/comments
Keyword	keywords extracted from all text content (e.g., oxycodone, fentanyl, codine, and LSD)

Relation	
R1: user-follow/followed-user	R2: user-tagger-user
R3: user-reply-post	R4: user-mention-post
R5: user-have-post	R6: user-profile-keyword
R7: post-include-keyword	R8: post-tag-keyword

$$E(X_i', X_j') = \Gamma(W_{s1} \odot X_i', W_{s2} \odot X_j')$$

$$\mathcal{E}_{i,j}' = \begin{cases} 1 & E(X_i', X_j') \geq \epsilon \\ 0 & \text{otherwise} \end{cases}$$

sparse

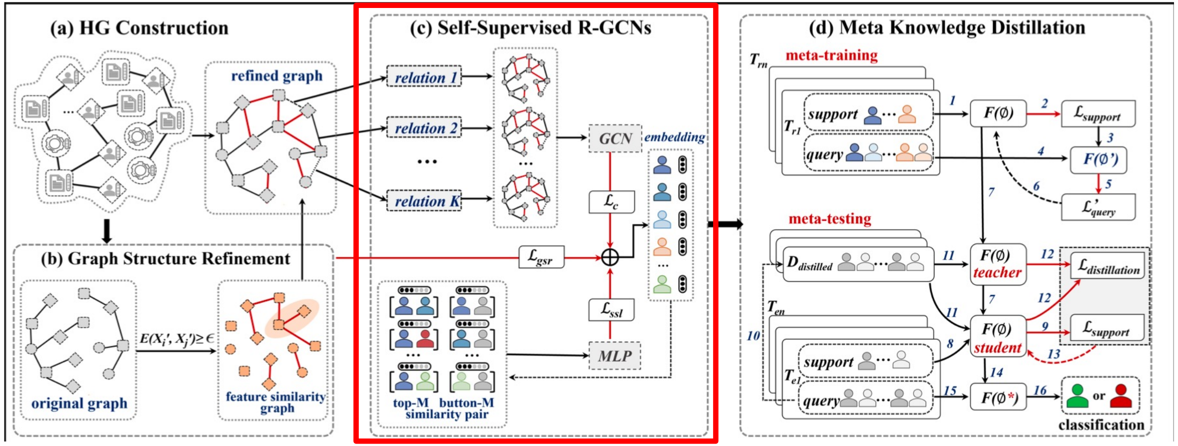


Distilling Meta Knowledge on Heterogeneous Graph for Illicit Drug Trafficker Detection on Social Media, NeurIPS 2021

Distilling Meta Knowledge on HG for Illicit Drug Trafficker Detection

Step (b): HG representation learning

Detail: learn user embedding through R-GCN with self-supervised signal



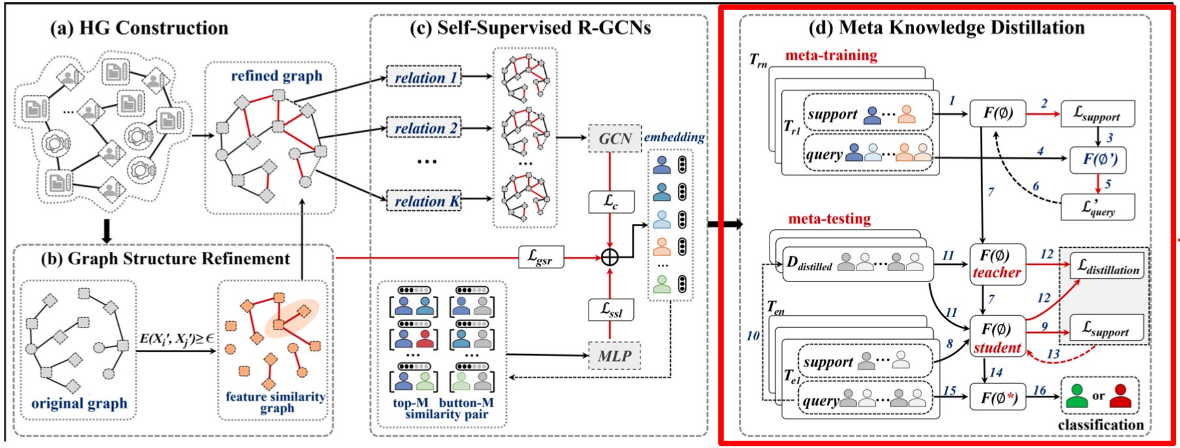
Distilling Meta Knowledge on Heterogeneous Graph for Illicit Drug Trafficker Detection on Social Media, NeurIPS 2021

Distilling Meta Knowledge on HG for Illicit Drug Trafficker Detection

Step (c): meta-learning with knowledge distillation

Intuition: unlabeled data could benefit meta training process

Detail: leverage unlabeled data in MAML via knowledge distillation



$$P_l^T(Z_i,t) = \text{Softmax}(f(Z_i),t)) = \frac{\exp\left[f_l(Z_i/t)\right]}{\sum_{c=0}^{1}\exp\left[f_c(Z_i/t)\right]}$$

$$\boxed{\mathcal{L}_{kd}} = -t^2 \sum_{v_i \in \mathcal{Q}_\tau} \sum_{c=0}^{1} P_c^T(Z_i,t)\log(P_c^S(Z_i,t))$$

"Soft" knowledge of unlabeled data

$$\mathcal{L}_{total} = \mathcal{L}_{ce} + \lambda_{kd}\mathcal{L}_{kd}$$

$$\phi^* = \phi - \alpha \nabla_\phi \mathcal{L}_{total}(\phi)$$

Distilling Meta Knowledge on Heterogeneous Graph for Illicit Drug Trafficker Detection on Social Media, NeurIPS 2021

Distilling Meta Knowledge on HG for Illicit Drug Trafficker Detection

Illicit drug trafficker detection results on Instagram

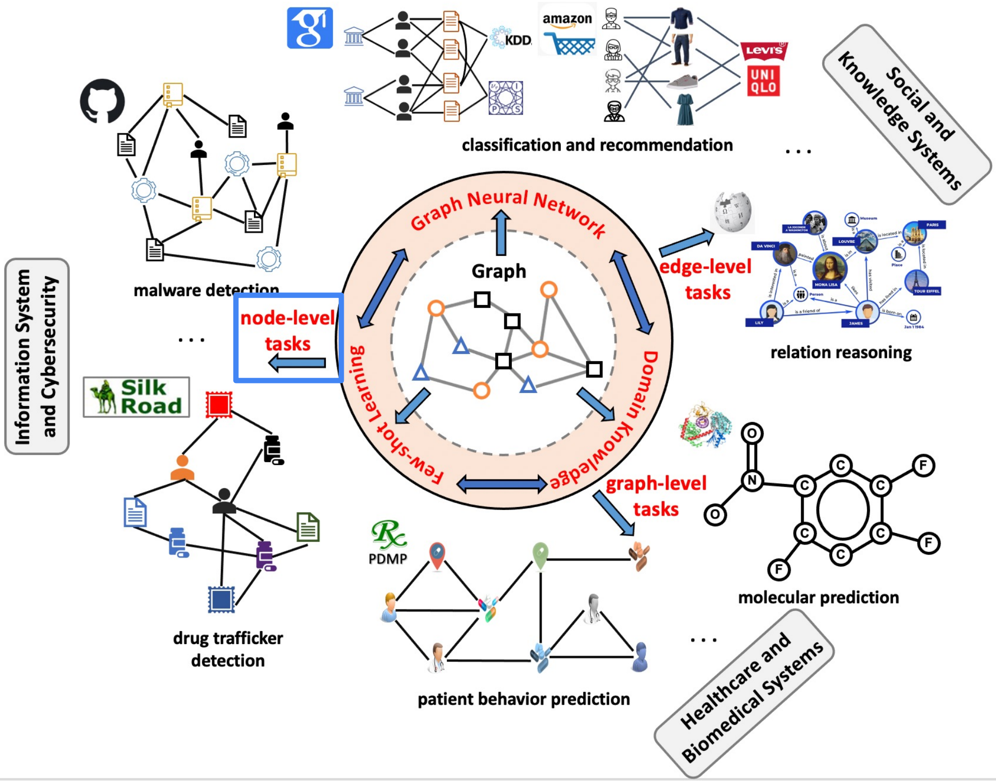
Setting		1-shot		5-shot		10-shot		20-shot	
Group	Model	ACC	F1	ACC	F1	ACC	F1	ACC	F1
B1	tFeature+DNN [37]	0.4786	0.3768	0.5046	0.3959	0.5215	0.4116	0.5420	0.4230
	iFeature+DNN	0.5253	0.4256	0.5428	0.4432	0.5564	0.4635	0.5712	0.4841
	cFeature+DNN	0.5468	0.4434	0.5601	0.4553	0.5758	0.4755	0.5935	0.4924
B2	Li *et al.* [20]	0.4929	0.3942	0.5287	0.4171	0.5496	0.4280	0.5701	0.4456
	Rokon *et al.* [22]	0.5058	0.4037	0.5396	0.4260	0.5585	0.4376	0.5765	0.4529
B3	cFeature+ProtoNet [36]	0.5815	0.5735	0.6156	0.5987	0.6321	0.6293	0.6637	0.6572
	cFeature+MatchingNet [40]	0.6058	0.5953	0.6337	0.6257	0.6551	0.6478	0.6773	0.6659
	cFeature+MAML [10]	0.6337	0.6218	0.6654	0.6525	0.6872	0.6735	0.6959	0.6985
B4	[29] Deepwalk+DNN	0.5919	0.4991	0.6210	0.5251	0.6472	0.5449	0.6710	0.5605
	[6] metapath2vec+DNN	0.6257	0.5243	0.6518	0.5549	0.6739	0.5683	0.6953	0.5825
	[17] GCN+DNN	0.6524	0.5510	0.6853	0.5772	0.7054	0.5953	0.7291	0.6138
	[39] GAT+DNN	0.6650	0.5558	0.6889	0.5842	0.7129	0.6023	0.7305	0.6195
	[43] HAN+DNN	0.6786	0.5629	0.7047	0.5925	0.7207	0.6152	0.7421	0.6290
	[32] R-GCNs+DNN	0.6836	0.5765	0.7183	0.6042	0.7254	0.6221	0.7476	0.6446
B5	Deepwalk+MAML	0.6962	0.6957	0.7324	0.7306	0.7559	0.7537	0.7751	0.7674
	metapath2vec+MAML	0.7251	0.7232	0.7622	0.7534	0.7837	0.7746	0.7995	0.7921
	GCN+MAML	0.7526	0.7407	0.7835	0.7827	0.8052	0.7924	0.8319	0.8356
	GAT+MAML	0.7578	0.7426	0.7905	0.7921	0.8124	0.8036	0.8439	0.8214
	HAN+MAML	0.7732	0.7654	0.8062	0.7959	0.8328	0.8176	0.8551	0.8327
	R-GCNs+MAML	0.7853	0.7727	0.8253	0.8149	0.8465	0.8352	0.8678	0.8535
Ours	**MetaHG**	**0.8489**	**0.8480**	**0.8873**	**0.8758**	**0.9196**	**0.9122**	**0.9354**	**0.9311**

Observation: results indicate the effectiveness of MetaHG

Distilling Meta Knowledge on Heterogeneous Graph for Illicit Drug Trafficker Detection on Social Media, NeurIPS 2021

Introduction

- Graph Few-shot Learning



Edge-level Learning Tasks

❑ Part-2: Edge (Relation)-level Few-shot Learning

❖ Relation Reasoning over Knowledge Graphs

Meta-AHIN (EMNLP 2020 Findings)

MetaHG (Under Review)

❖ Link Prediction over Knowledge Graphs

FSRL (AAAI 2020)

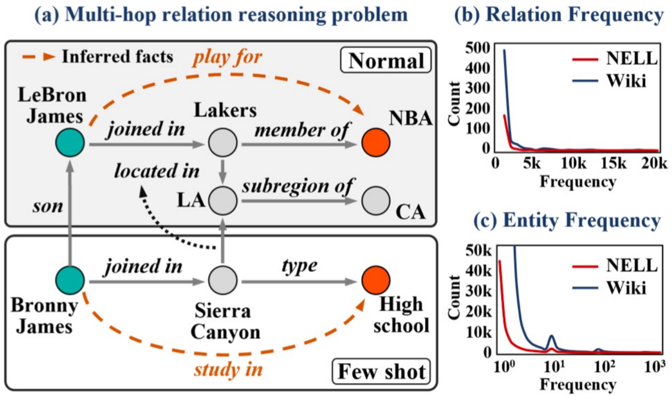
Few-Shot Relation Reasoning over Knowledge Graphs

$G = \{E, R, T\}$, where E and R denote the set of entities and relations, T is the collection of fact triples and each element is a tuple(e_s, r_q, e_o), where e_s, $e_o \in E$ and $r_q \in R$.

Divide all relations into two groups: few-shot and normal, based on relation frequency (cut by shot number K).

Given: triplets of normal relations, a query (e_s, r_q, ?), where e_s is the source entity and r_q is the query few-shot relation

Predict: the target entity e_o for this query after a multi-hop reasoning over KG



Few-Shot Multi-Hop Relation Reasoning over Knowledge Bases, EMNLP 2020 (Findings)

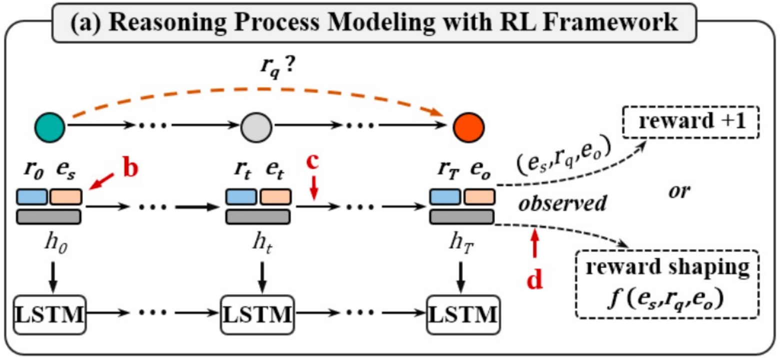
Few-Shot Relation Reasoning over Knowledge Graphs

Setting: gradient-based method

Motivation: knowledge from normal relations is useful

Key idea: leverage meta knowledge of normal relations over KG to fast adaptation for few-shot relations

Reinforcement Learning Framework for Relation Reasoning



$$\mathbf{h}_0 = \mathrm{LSTM}(0, [\mathbf{r}_0 \oplus \mathbf{e}_s])$$
$$\mathbf{h}_t = \mathrm{LSTM}(\mathbf{h}_{t-1}, \mathbf{a}_{t-1}), t > 0$$

State

$$\varphi_\theta(a_t|s_t) = \sigma\{\mathbf{A}_t(\mathbf{W}_2\mathrm{ReLU}(\mathbf{W}_1[\mathbf{e}_t \oplus \mathbf{h}_t \oplus \mathbf{r}_q]))\}$$

Policy Network

$$\mathcal{J}_r^{\mathcal{D}}(\theta) = \mathbb{E}_{(e_s,r,e_o \in \mathcal{D})}\{\mathbb{E}_{a_1,\cdots a_T \sim \varphi_\theta}[\mathcal{R}(s_T|e_s, r)]\}$$

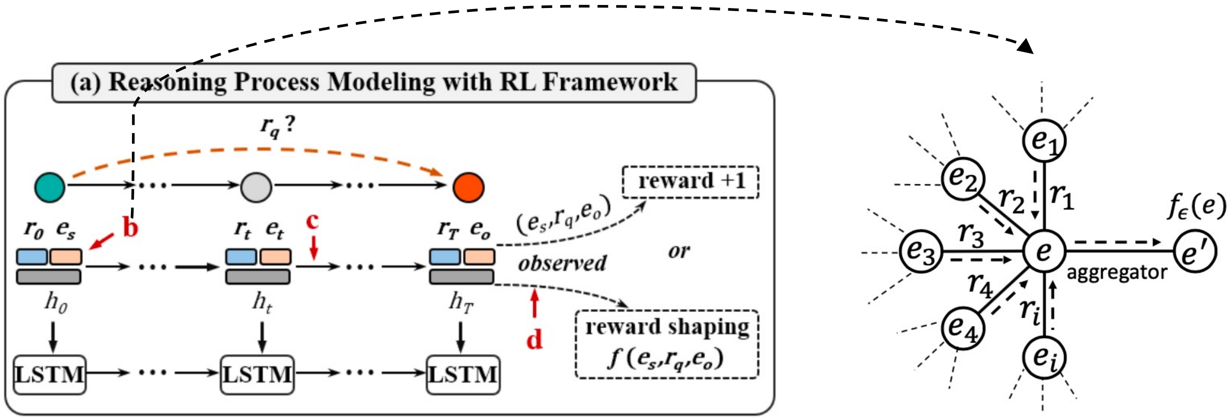
Few-Shot Multi-Hop Relation Reasoning over Knowledge Bases, EMNLP 2020 (Findings)

Few-Shot Relation Reasoning over Knowledge Graphs

Component (b): entity embedding refinement

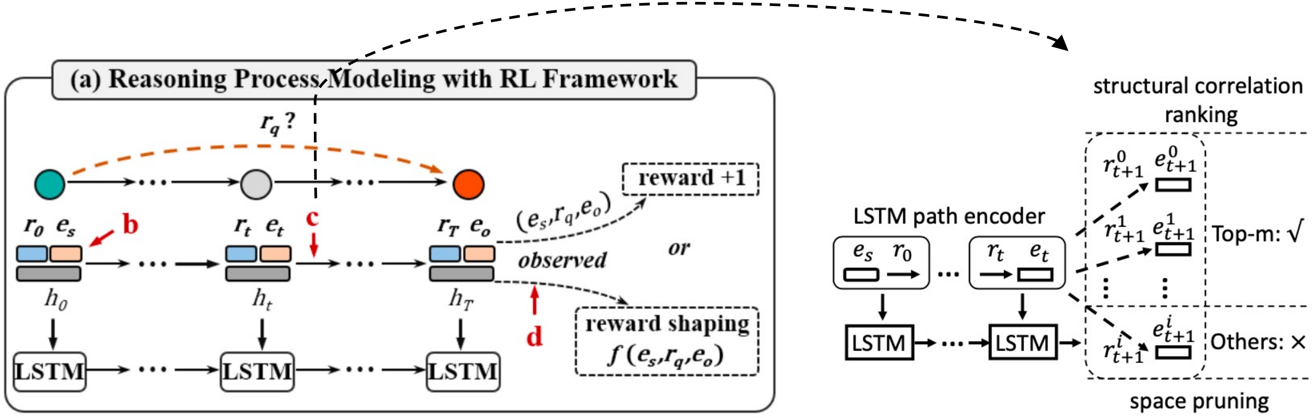
Intuition: heterogeneous graph structure information is useful

Detail: design a heterogeneous neighbor encoder to refine entity embedding



Few-Shot Multi-Hop Relation Reasoning over Knowledge Bases, EMNLP 2020 (Findings)

Component (c): search space pruning
Intuition: action search space enormous or even redundant
Detail: search space pruning strategy based on embedding similarity

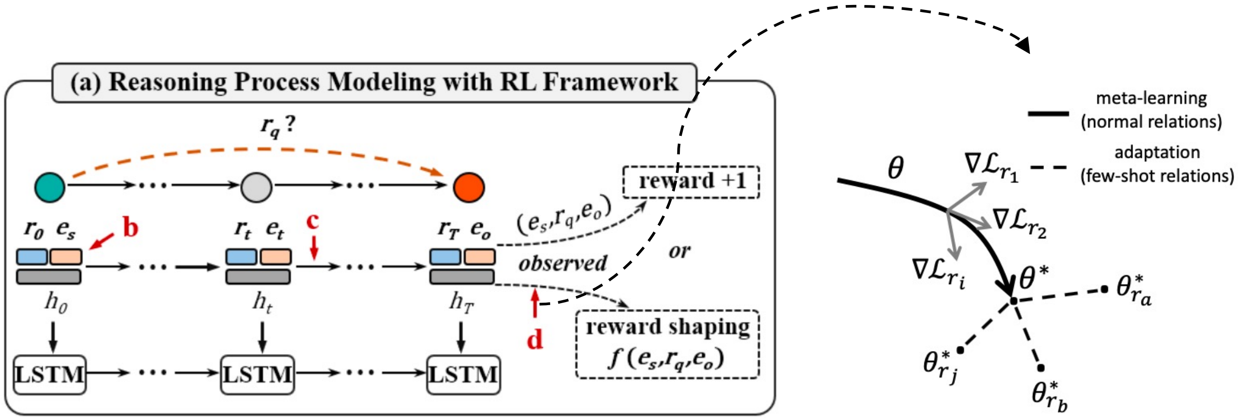


Few-Shot Multi-Hop Relation Reasoning over Knowledge Bases, EMNLP 2020 (Findings)

Component (d): optimization with meta-learning

Intuition: initialize parameters with normal relation data

Detail: MAML for fast adaption over few-shot relation data

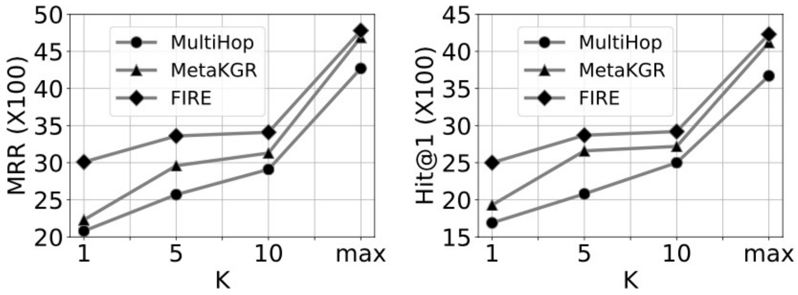


Few-Shot Multi-Hop Relation Reasoning over Knowledge Bases, EMNLP 2020 (Findings)

Few-Shot Relation Reasoning over Knowledge Graphs

relation reasoning results over knowledge graphs

Model	NELL-995		FB15K-237	
	MRR	Hit@1	MRR	Hit@1
NeuralLP	17.9	4.8	10.2	7.0
NTP-λ	15.5	10.2	21.0	17.4
MINERVA	20.1	16.2	30.5	28.4
MultiHop	23.1	17.8	42.7	36.7
MetaKGR	<u>25.3</u>	<u>19.7</u>	<u>46.9</u>	<u>41.2</u>
FIRE	**27.3**	**22.5**	**47.8**	**42.3**

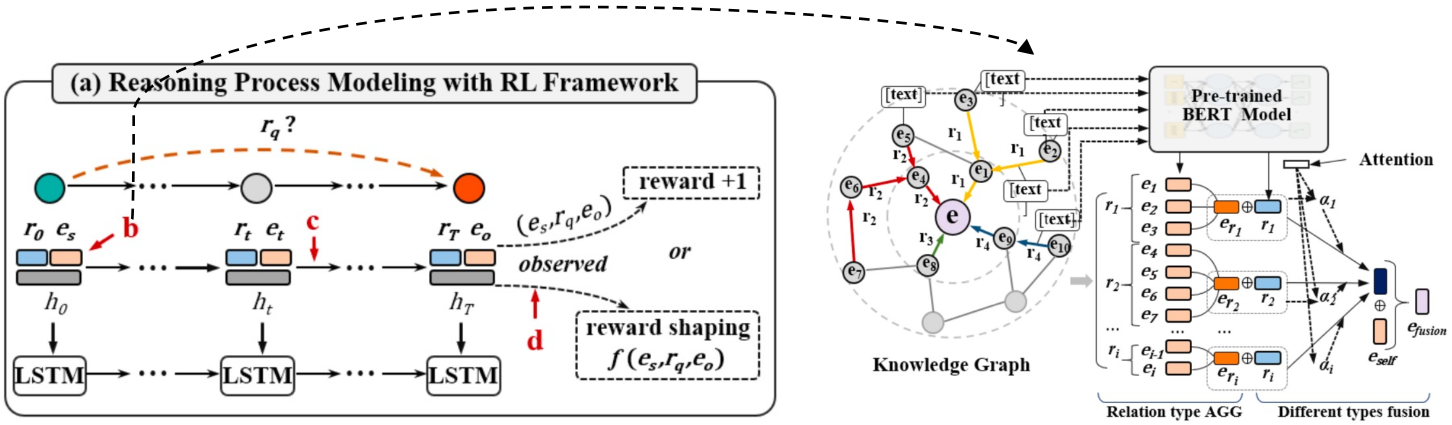


Observation: results indicate the effectiveness of FIRE

Few-Shot Multi-Hop Relation Reasoning over Knowledge Bases, EMNLP 2020 (Findings)

Adapting Distilled Knowledge for Few-shot Relation Reasoning over KG

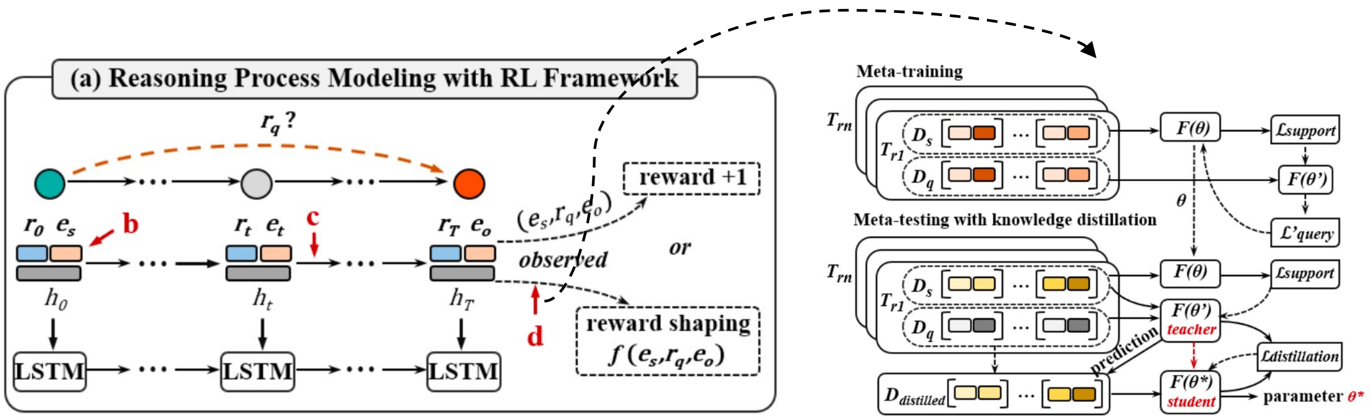
Component (b): entity embedding refinement with both structure and text
Intuition: heterogeneous graph structure and text information is useful
Detail: design a text-enhanced heterogeneous neighbor encoder to refine entity embedding



Adapting Distilled Knowledge for Few-shot Relation Reasoning over Knowledge Graphs, SDM 2022

Adapting Distilled Knowledge for Few-shot Relation Reasoning over KG

Component (d): meta-learning with knowledge distillation
Intuition: unlabeled data could benefit meta training process
Detail: leverage unlabeled data in MAML via knowledge distillation

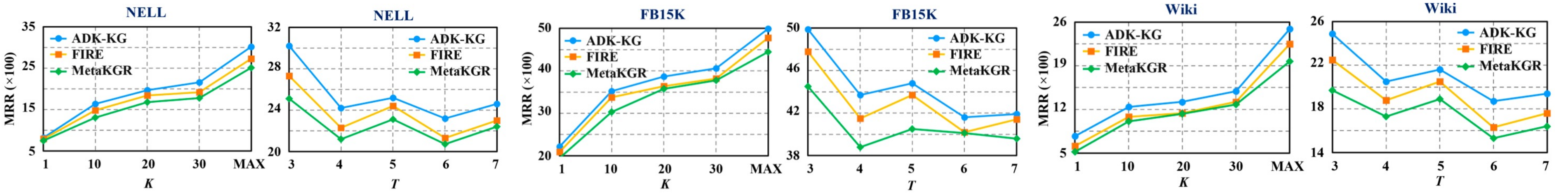


Adapting Distilled Knowledge for Few-shot Relation Reasoning over Knowledge Graphs, SDM 2022

Adapting Distilled Knowledge for Few-shot Relation Reasoning over KG

relation reasoning results over knowledge graphs

Model	NELL				FB15K				Wiki			
	MRR	Hit@1	Hit@5	Hit@10	MRR	Hit@1	Hit@5	Hit@10	MRR	Hit@1	Hit@5	Hit@10
NeuralLP [17]	17.9	4.8	18.5	35.1	10.2	7.1	10.6	14.8	15.6	6.5	16.6	31.5
NTP-λ [18]	15.5	10.2	16.3	33.2	21.0	17.4	22.3	30.7	15.1	7.25	16.1	29.2
MINERVA [13]	20.2	16.1	21.9	28.2	30.5	28.4	32.4	34.1	18.5	14.3	19.9	26.5
MultiHop [19]	23.2	17.8	24.6	32.8	42.8	36.8	43.3	53.2	21.1	15.8	22.4	30.6
MetaKGR [14]	25.1	19.7	25.8	34.2	44.5	39.5	45.7	56.6	23.0	18.3	25.1	32.5
FIRE [31]	27.3	22.5	27.6	36.5	47.8	42.3	48.1	59.2	25.6	21.5	27.3	34.3
ADK-KG	**30.2**	**24.9**	**31.5**	**39.4**	**49.9**	**44.6**	**53.8**	**63.4**	**27.7**	**24.2**	**29.2**	**36.1**

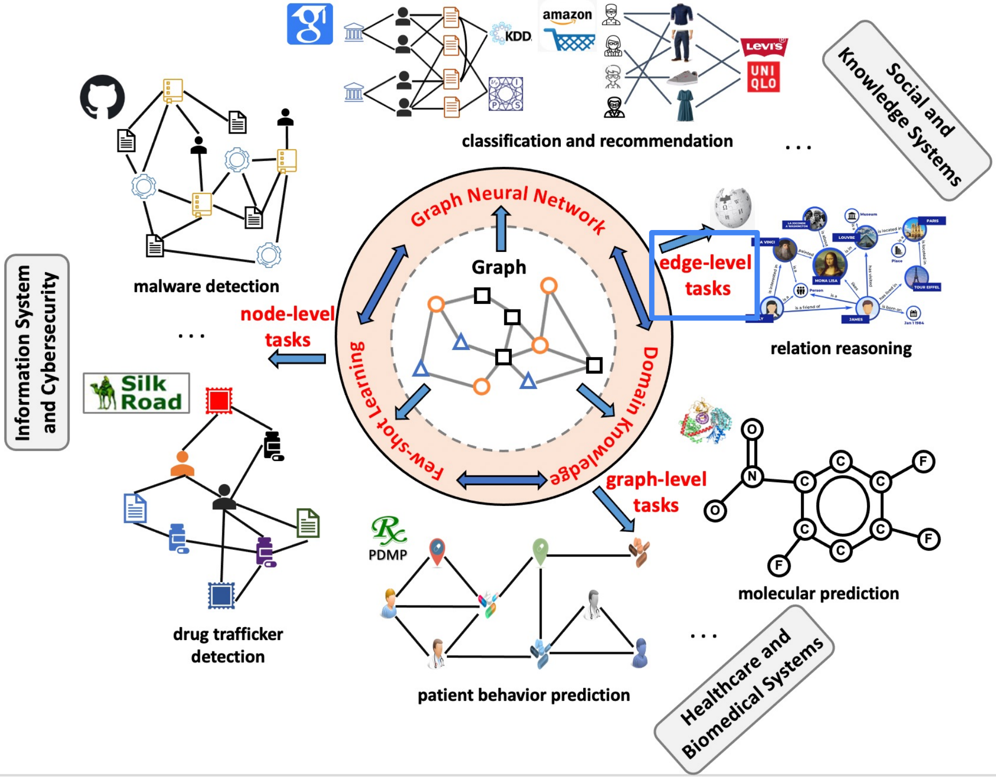


Observation: results indicate the effectiveness of ADK-KG

Adapting Distilled Knowledge for Few-shot Relation Reasoning over Knowledge Graphs, SDM 2022

Introduction

- Graph Few-shot Learning



Graph-level Learning Tasks

- Part-3: Graph-level Few-shot Learning
  - Graph Classification

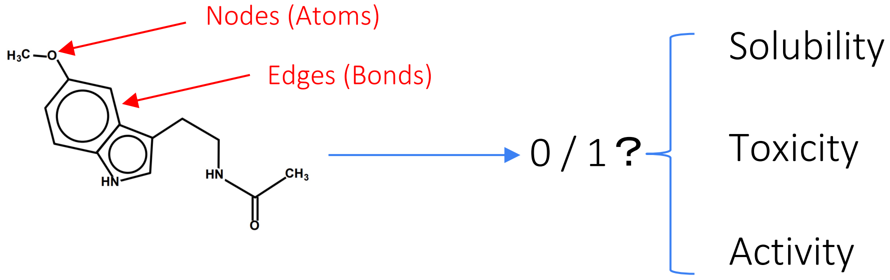
Meta-MGNN (WWW 2021): Molecular Property Prediction

Few-Shot Graph Learning for Molecular Property Prediction

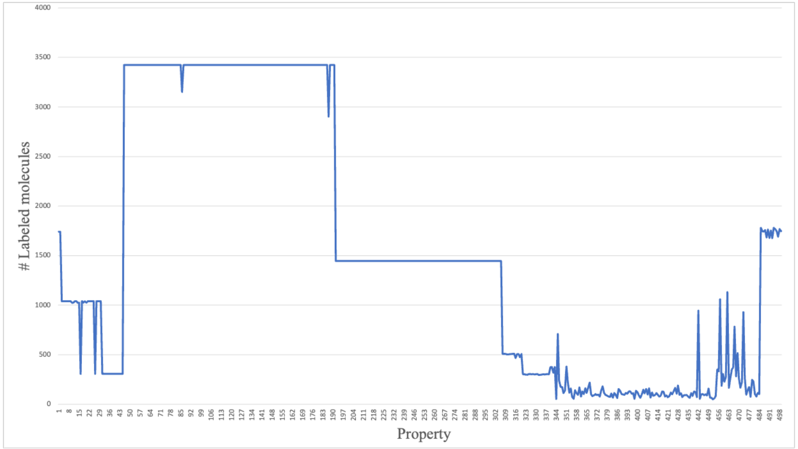
Molecular graph represents a chemical atom and an edge represents a chemical bond between two atoms. Molecular property prediction is to learn a molecular representation vector for predicting its label (i.e., molecular property).

Given: molecular properties and their corresponding molecular graph sets

Predict: molecular graphs of new properties that only have few-shot examples



Key step in new drugs discovery



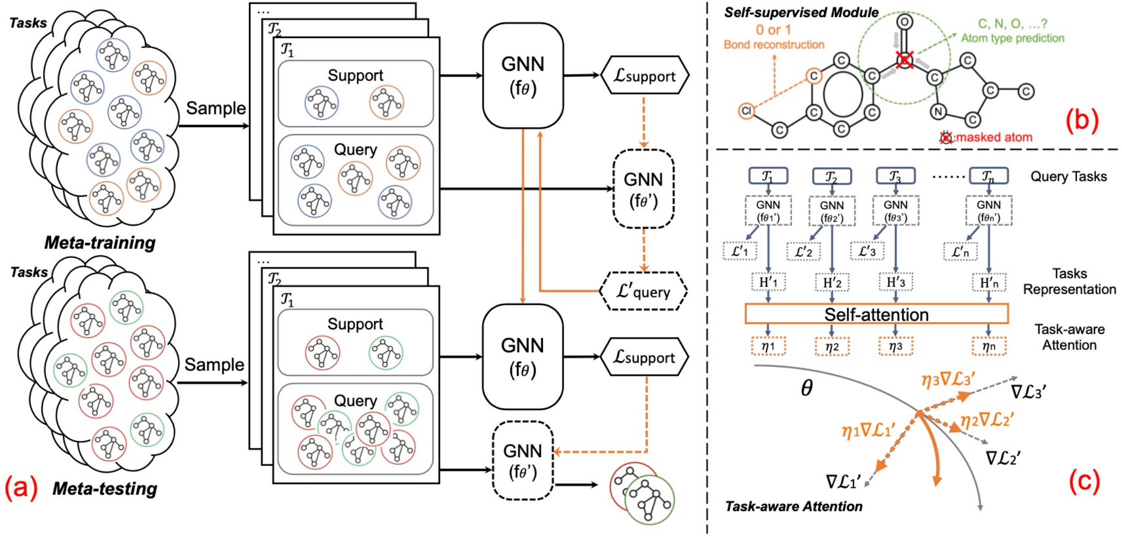
Few-Shot Graph Learning for Molecular Property Prediction, WWW 2021

Few-Shot Graph Learning for Molecular Property Prediction

Setting: gradient-based method

Motivation: knowledge from existing molecular properties is useful

Key idea: leverage meta knowledge of existing molecular properties to fast adaptation for new molecular property



Few-Shot Graph Learning for Molecular Property Prediction, WWW 2021

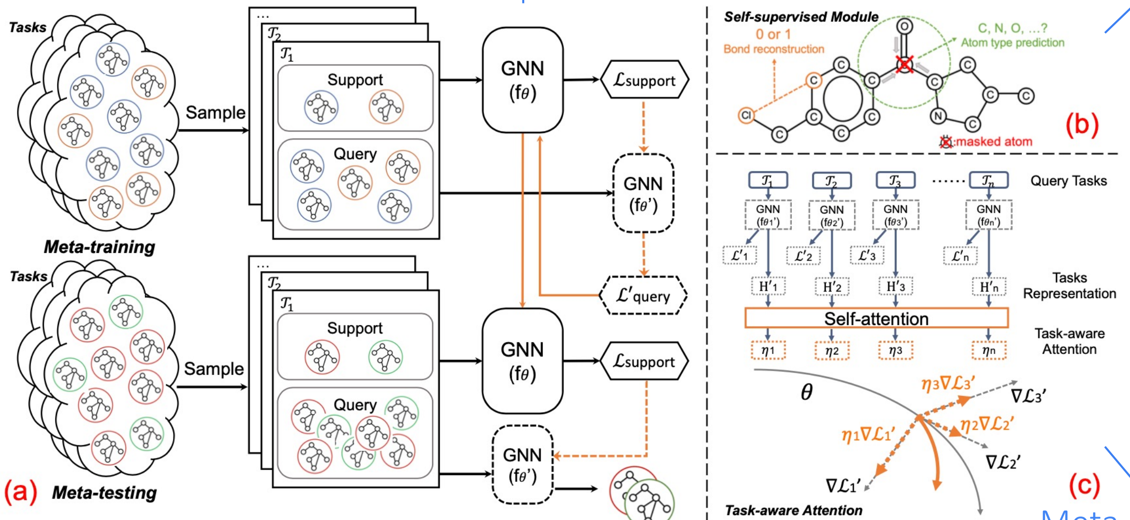
Few-Shot Graph Learning for Molecular Property Prediction

domain information

# Atom Type	118
Atom Chirality Tag	Unspecified, Tetrahedral cw, Tetrahedral ccw, Other
Bond Type	Single, Double, Triple, Aromatic
Bond Direction	-, Endupright, Enddownright

Molecular GNN

Self-supervised Signal:
Bond Reconstruction
Atom Type Prediction



Meta-learning with task attention

$$\eta(\mathcal{T}_\tau) = \frac{\exp(\text{MLP}(\mathbf{H}_{\mathcal{T}_\tau}))}{\sum_{\mathcal{T}_{\tau'}\in\mathcal{T}} \exp(\text{MLP}(\mathbf{H}_{\mathcal{T}_{\tau'}}))},\ \mathbf{H}_{\mathcal{T}_\tau} = \text{MEAN}(\{\mathbf{h}_{\mathcal{T}_\tau,i}\}_{i=1}^{k}).$$

Few-Shot Graph Learning for Molecular Property Prediction, WWW 2021

molecular property prediction results

Dataset	Task	GraphSAGE [9] (2017)	GCN [12] (2017)	MAML [5] (2017)	Seq3seq [34] (2018)	EGNN [11] (2019)	PreGNN [10] (2020)	Meta-MGNN	ΔAUC
					1-shot				
Tox21	**SR-HS**	65.97	65.00	68.56	73.18	72.51	73.09	**73.81**	**+0.63**
	SR-MMP	71.23	71.20	76.34	79.08	76.90	76.20	**79.09**	**+0.01**
	SR-p53	58.05	66.60	71.28	75.23	**78.03**	76.87	77.71	-0.32
	Average	65.10	67.60	72.06	75.83	75.81	75.39	**76.87**	**+1.04**
Sider	**Si-T1**	65.23	63.60	66.82	66.50	71.39	73.04	**75.41**	**+2.37**
	Si-T2	60.47	62.01	63.62	57.03	67.87	66.06	**69.39**	**+1.52**
	Si-T3	61.45	64.52	67.50	61.38	68.23	70.36	**70.65**	**+0.29**
	Si-T4	64.41	65.28	69.02	63.45	72.67	72.34	**72.69**	**+0.02**
	Si-T5	77.85	74.95	77.07	74.83	78.88	77.99	**79.95**	**+1.07**
	Si-T6	61.19	63.20	67.01	63.70	66.31	69.45	**71.97**	**+2.52**
	Average	65.10	65.60	68.51	64.48	70.89	71.54	**73.34**	**+1.80**
					5-shots				
Tox21	**SR-HS**	69.09	68.13	69.02	74.07	73.23	73.39	**74.80**	**+0.73**
	SR-MMP	72.22	69.06	76.43	**80.40**	79.07	78.25	80.26	-0.14
	SR-p53	61.45	72.01	73.95	77.07	78.12	78.01	**79.00**	**+0.88**
	Average	67.59	69.73	73.13	77.18	76.81	76.55	**78.02**	**+0.84**
Sider	**Si-T1**	67.61	65.66	70.12	68.99	72.76	74.77	**76.32**	**+1.55**
	Si-T2	59.86	64.62	64.46	56.53	68.13	65.69	**69.34**	**+1.21**
	Si-T3	60.61	64.90	68.20	64.20	70.11	71.07	**72.29**	**+1.22**
	Si-T4	64.82	64.85	67.75	67.15	72.73	73.42	**74.46**	**+1.04**
	Si-T5	78.33	76.93	78.61	78.55	79.61	80.67	**81.79**	**+1.12**
	Si-T6	61.91	62.06	67.74	66.30	67.17	71.48	**74.12**	**+2.64**
	Average	65.52	66.50	69.48	66.95	71.75	72.85	**74.72**	**+1.87**

Observation: results indicate the effectiveness of Meta-MGNN

Few-Shot Graph Learning for Molecular Property Prediction, WWW 2021

Summary

Model	Task	Framework	Domain
GFL [AAAI'20]	Node Classification	Metric-based	Social & Knowledge Systems
Meta-AHIN [IJCAI'21]	Malware Detection	Gradient-based	Cybersecurity
MetaHG [NeurIPS'21]	Drug Trafficker Detection	Gradient-based	Public Health
Meta-AHIN [EMNLP'20]	Relation Reasoning	Gradient-based	Knowledge Systems
MetaHG [SDM'22]	Relation Reasoning	Gradient-based	Knowledge Systems
FSRL [AAAI'20]	Link Prediction	Metric-based	Knowledge Systems
Meta-MGNN [WWW'21]	Molecular Graph Classification	Gradient-based	Biomedicine/Healthcare

Open Problems

❑ Graph Few-shot Learning Exploration

P1: Heterogeneous graph few-shot learning

P2: Dynamic graph few-shot learning
C: Datasets, task setting

P3: Graph few-shot learning explanation
Meta-knowledge vs. fast adaptation

P4: More domain applications with small labeled data
Patient risk behavior: different drugs for opioid overdose

Thank you!

Q & A