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, data-efficient AI, explainable AI
Applications: healthcare, cybersecurity, social and information networks, recommendation

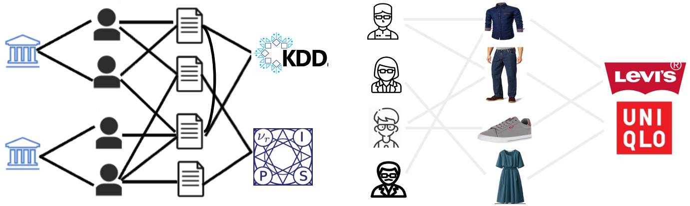
Awards: CIKM 2021 Best Paper Award, WWW 2019 Best Paper Candidate, APWeb/WAIM 2016 Best Student Paper Award, etc.

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

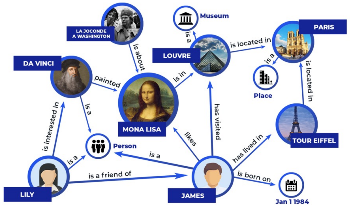
Introduction

❑ Graph Data and Applications

Social and Information Networks

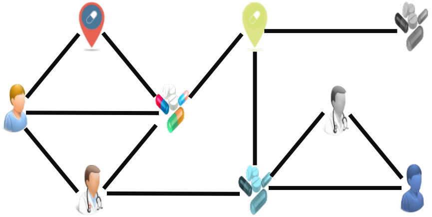


classification and recommendation



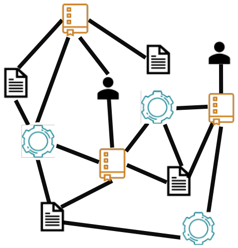
relation reasoning

Healthcare and Biomedicine

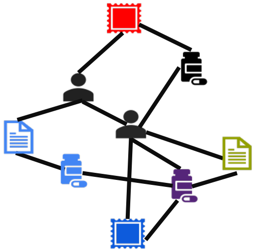


patient overdose behavior prediction

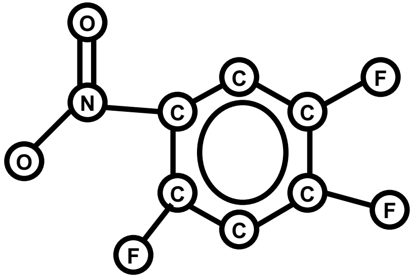
Cybersecurity and Public Health



malware detection



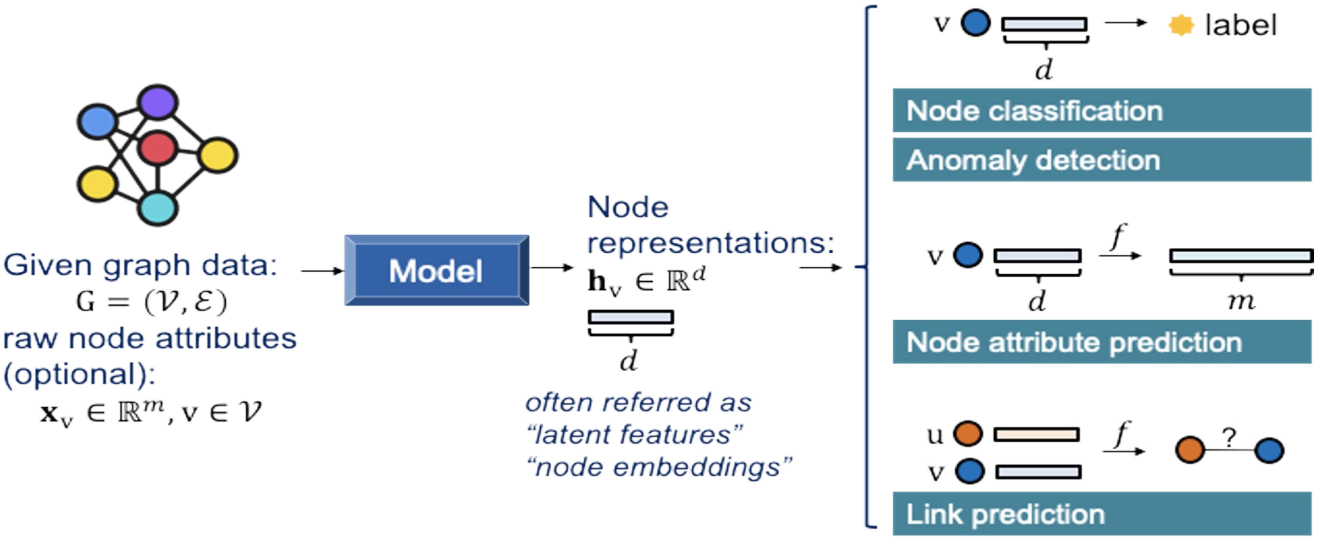
drug trafficker detection



molecular property prediction

Introduction

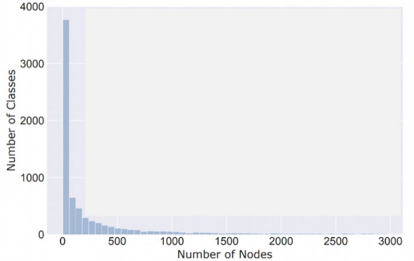
❑ Graph Representation Learning



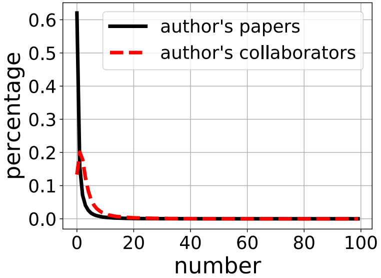
+ Graph Classification

Supervised GNNs often require sufficient labeled data for a specific task.

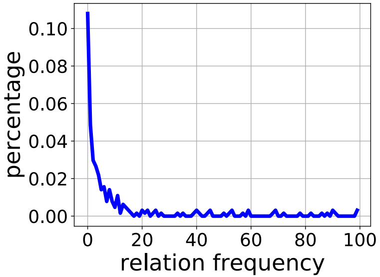
❑ Challenge: Small Labeled Data on Graphs



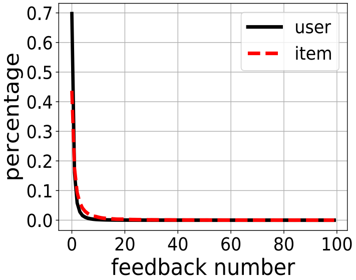
(a) number of labeled nodes in social graph



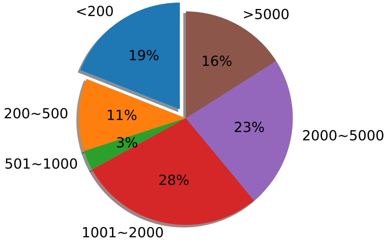
(b) author's paper/collaborator number in academic graph



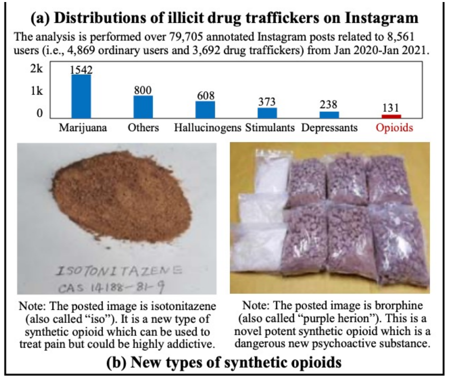
(c) relation frequency in knowledge graph



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



(e) number of molecular graph tested in experimental assays



Introduction

❑ Few-shot Learning

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

Goal: use auxiliary training data to mimic the target few-shot task, learn meta-knowledge from auxiliary data

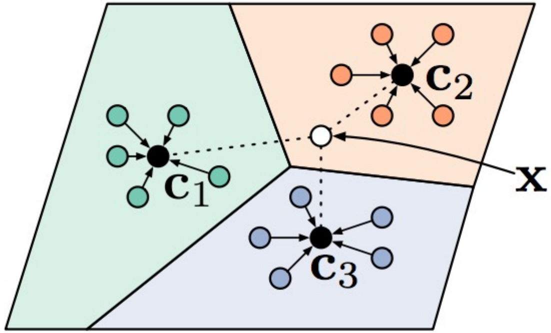
Setting: N-way K-shot Learning (N: class number, K: labeled data size - support set size)

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



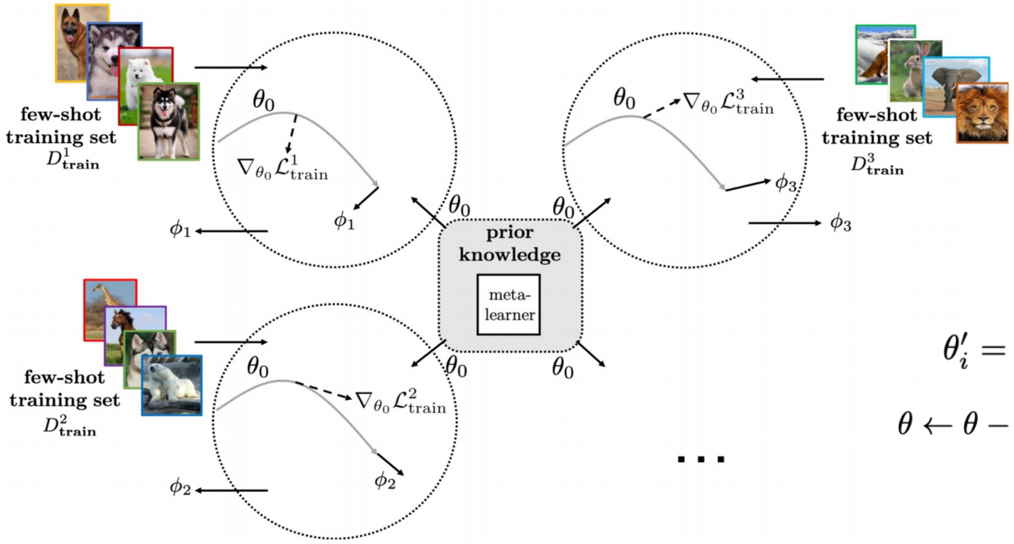
$$\mathbf{c}_k = \frac{1}{|S_k|} \sum_{(\mathbf{x}_i, y_i) \in S_k} f_\phi(\mathbf{x}_i)$$

$$p_\phi(y = k \mid \mathbf{x}) = \frac{\exp(-d(f_\phi(\mathbf{x}), \mathbf{c}_k))}{\sum_{k'} \exp(-d(f_\phi(\mathbf{x}), \mathbf{c}_{k'}))}$$

Prototypical Networks for Few-shot Learning, NeurIPS 2017

❑ Gradient-based Methods (e.g., MAML)

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



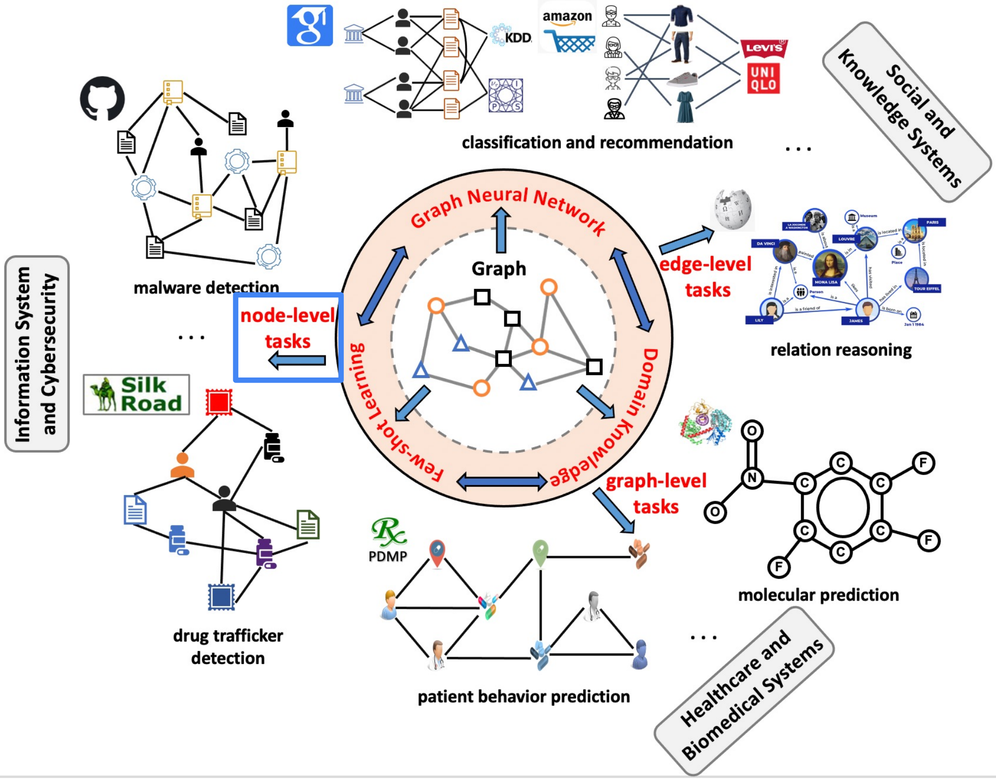
$$\theta_i' = \theta - \alpha\nabla_\theta\mathcal{L}_{\mathcal{T}_i}(f_\theta).$$

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

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 Application - Anomaly Detection

MetaHG (NeurIPS 2021): Illicit Drug Trafficker Detection

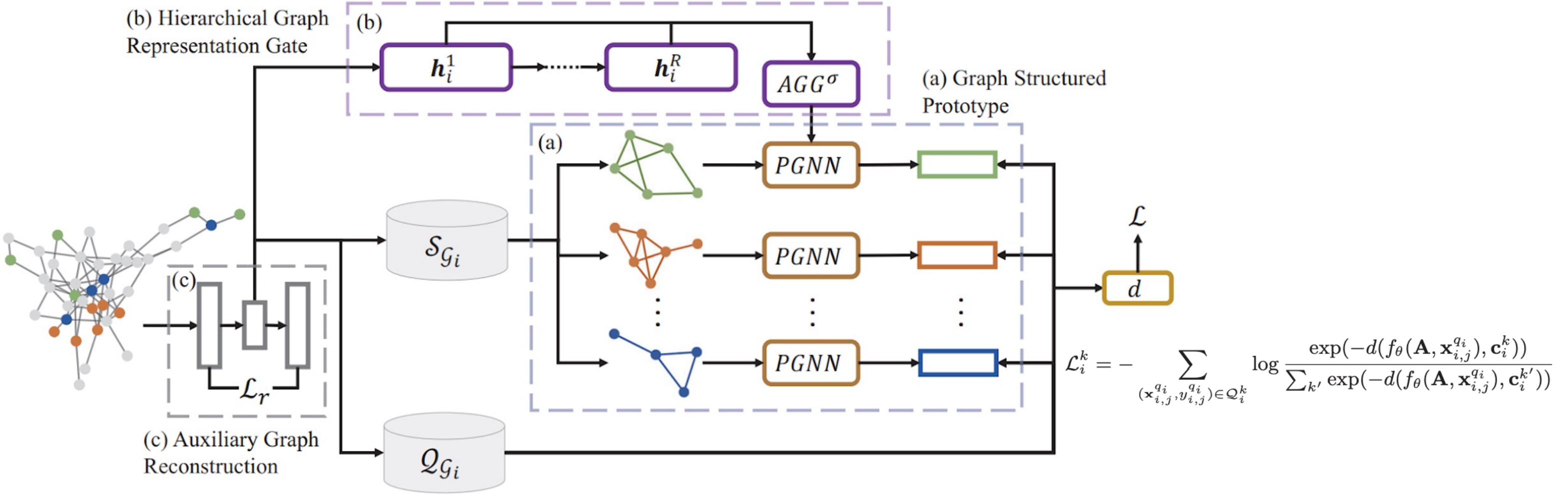
Graph Few-shot Learning via Knowledge Transfer

Given: labeled nodes (with attributes) from training classes

Predict: labels of query nodes of test classes with **few-shot** support set (labeled nodes)

Approach: metric-based method (prototypical net)

Key idea: incorporate meta-knowledge from auxiliary graphs/classes via knowledge transfer



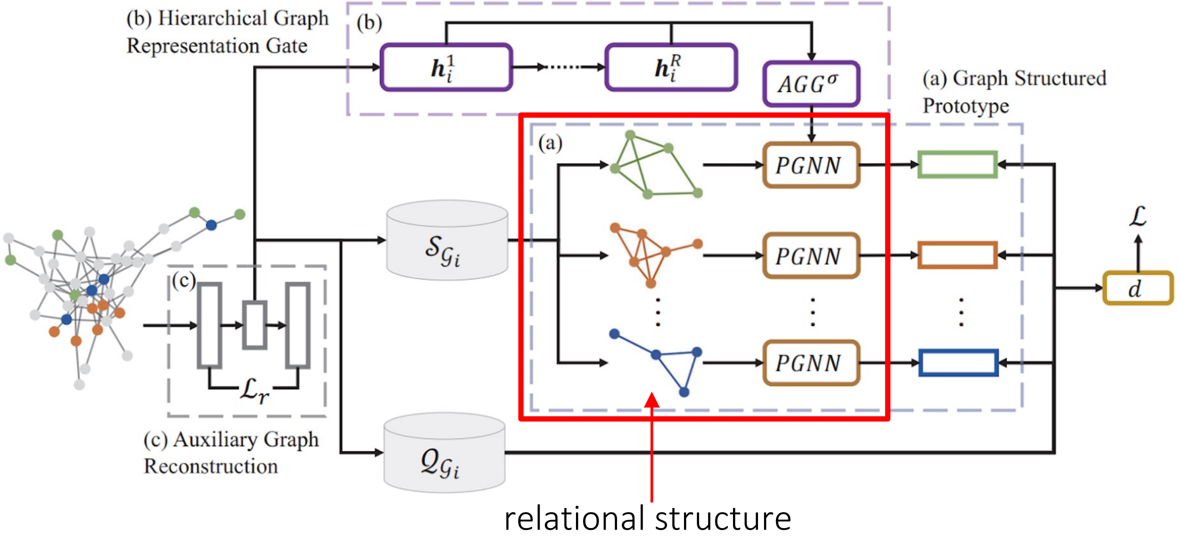
Graph Few-shot Learning via Knowledge Transfer, AAAI 2020

Graph Few-shot Learning via Knowledge Transfer

Step (a): compute graph-structured prototypes

Intuition: extract structure to describe interactions among support nodes of each class

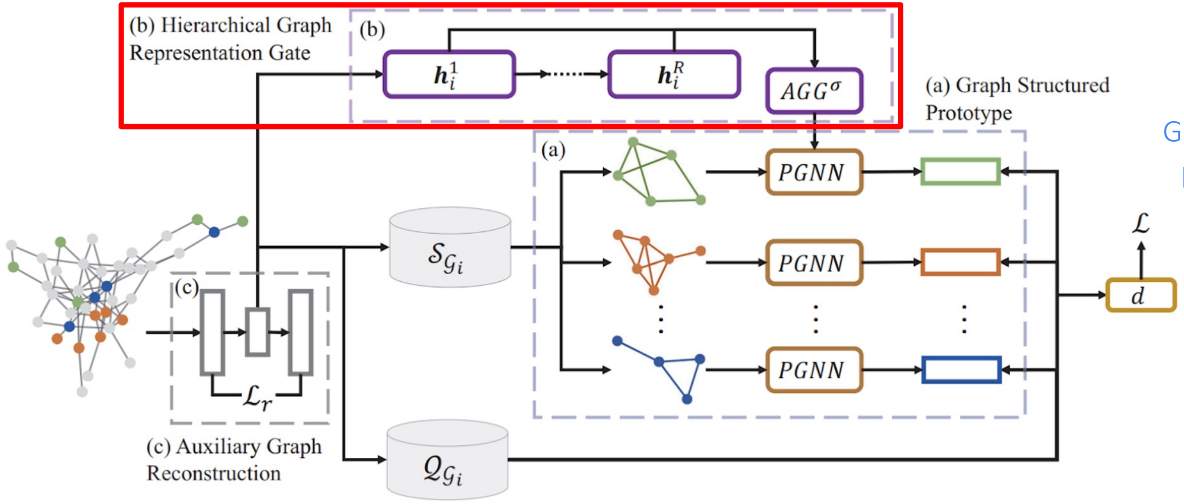
Detail: extract the relational structure (nodes of class k) and learn prototype via GNN



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

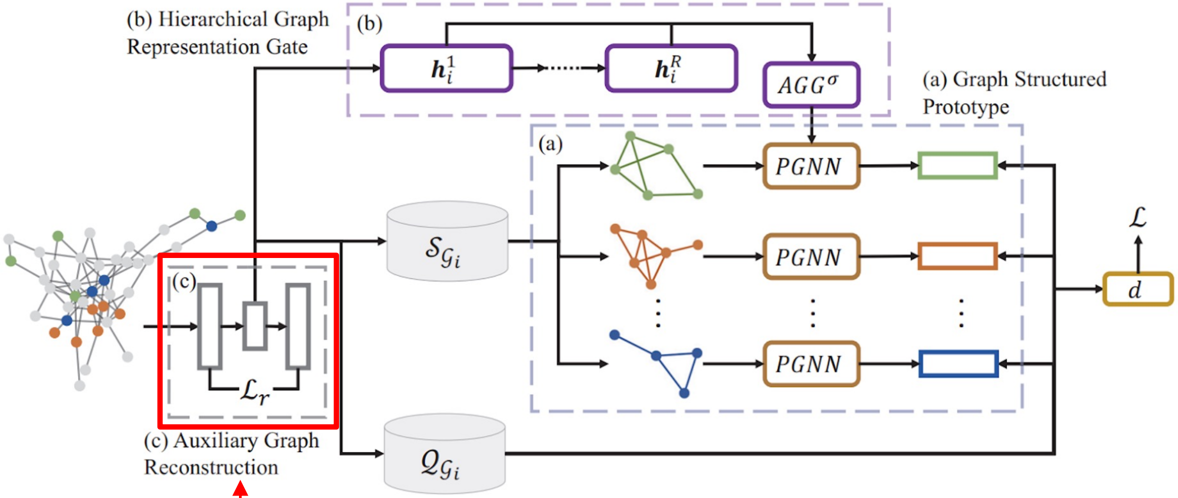
Graph Few-shot Learning via Knowledge Transfer, AAAI 2020

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

Graph Few-shot Learning via Knowledge Transfer

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

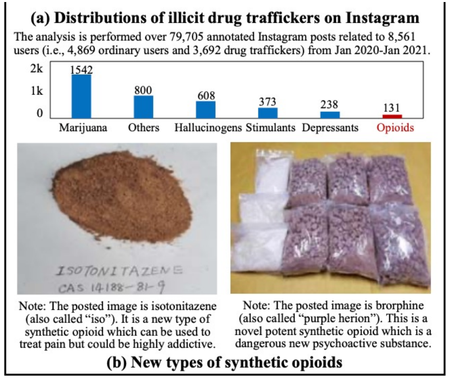
Code: https://shorturl.at/jquCS

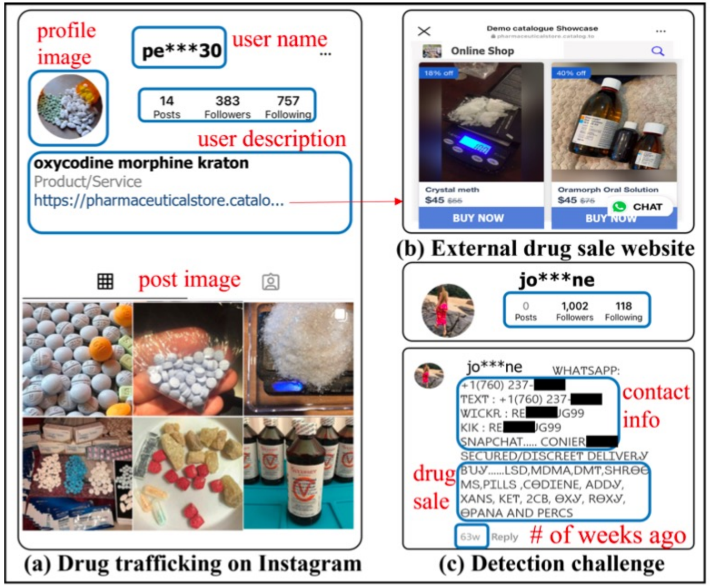
Graph Few-shot Learning via Knowledge Transfer, AAAI 2020

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

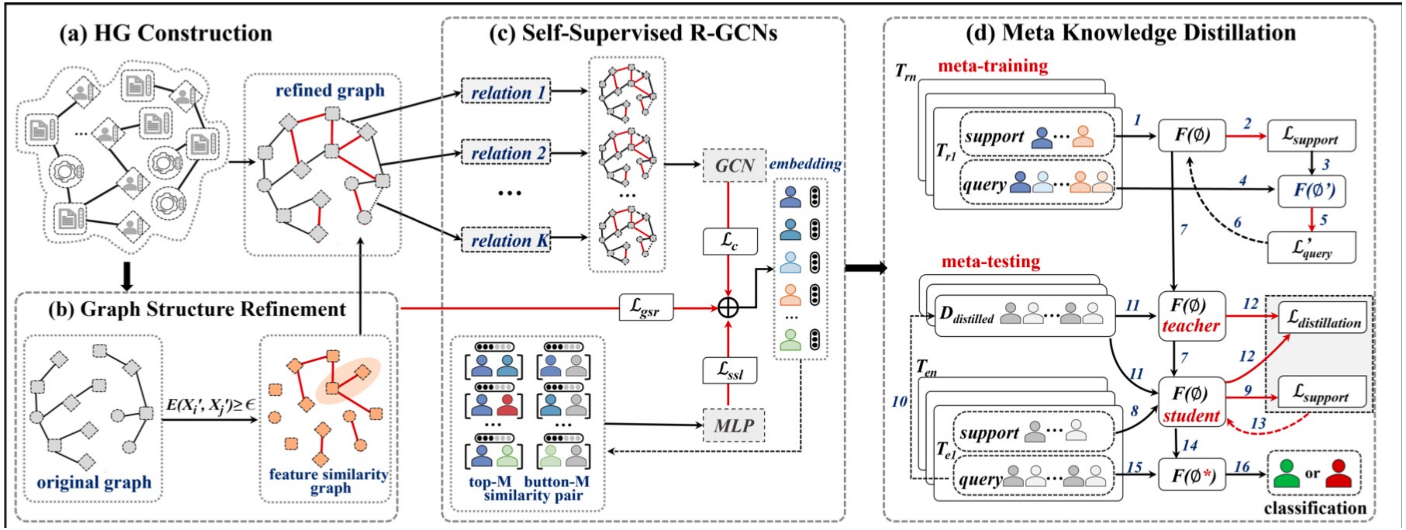




Distilling Meta Knowledge on Heterogeneous Graph for Illicit Drug Trafficker Detection on Social Media, NeurIPS 2021

Distilling Meta Knowledge on HG for Illicit Drug Trafficker Detection

Setting: gradient-based method (MAML)
Key idea: leverage meta knowledge from existing drug traffickers (on graph) to fast adaptation for new typed trafficker



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

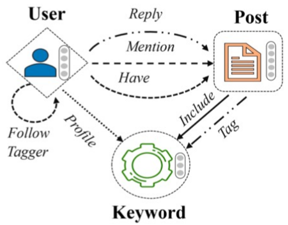
Distilling Meta Knowledge on HG for Illicit Drug Trafficker Detection

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

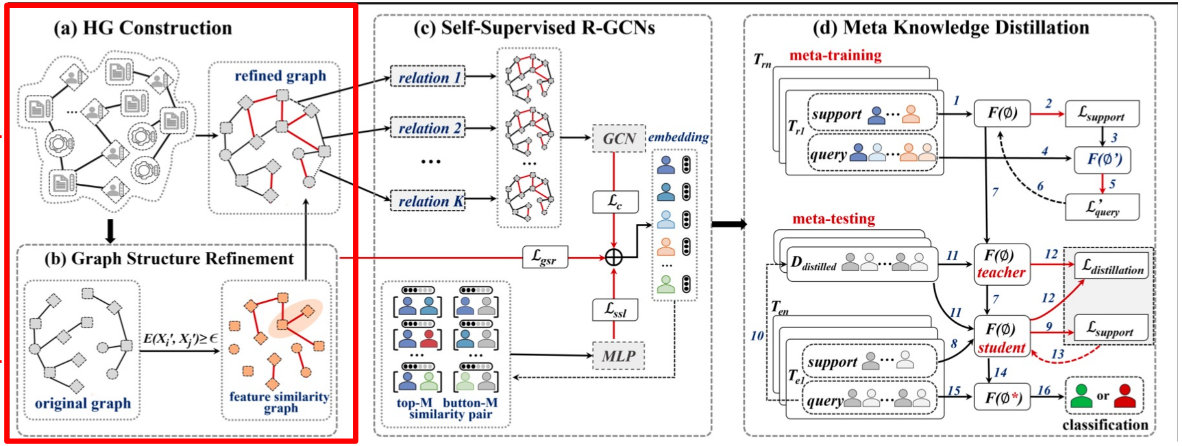


Content Feature	
User	username, profile information,is-business, # of followers/followings/posts
Post	text content, image information comment content, # of likes/comments
Keyword	keywords extracted from all text content (e.g., oxycodone, fentanyl, codine, and LSD)
Relation	
R1: user-follow/followed-user	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

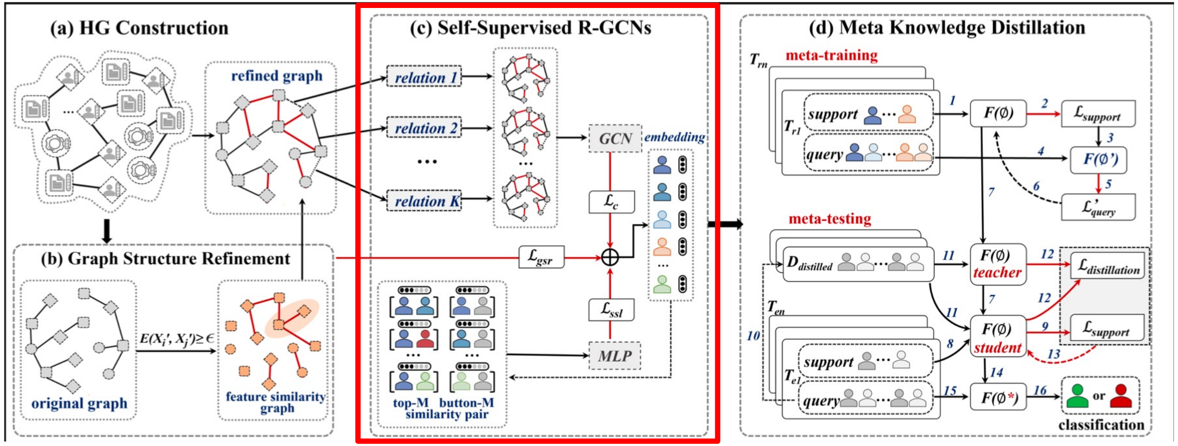


Distilling Meta Knowledge on Heterogeneous Graph for Illicit Drug Trafficker Detection on Social Media, NeurIPS 2021

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



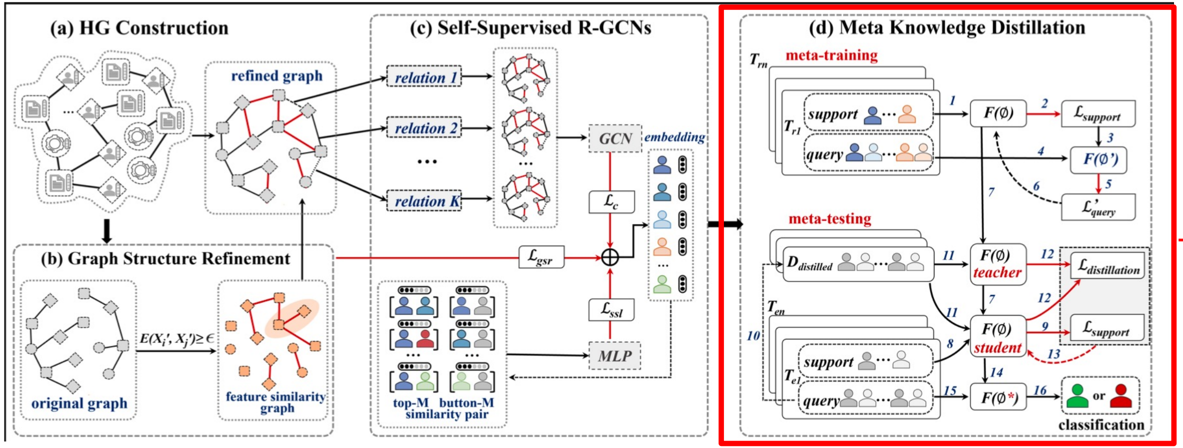
Distilling Meta Knowledge on Heterogeneous Graph for Illicit Drug Trafficker Detection on Social Media, NeurIPS 2021

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



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

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

"Soft" knowledge of unlabeled data

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

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**

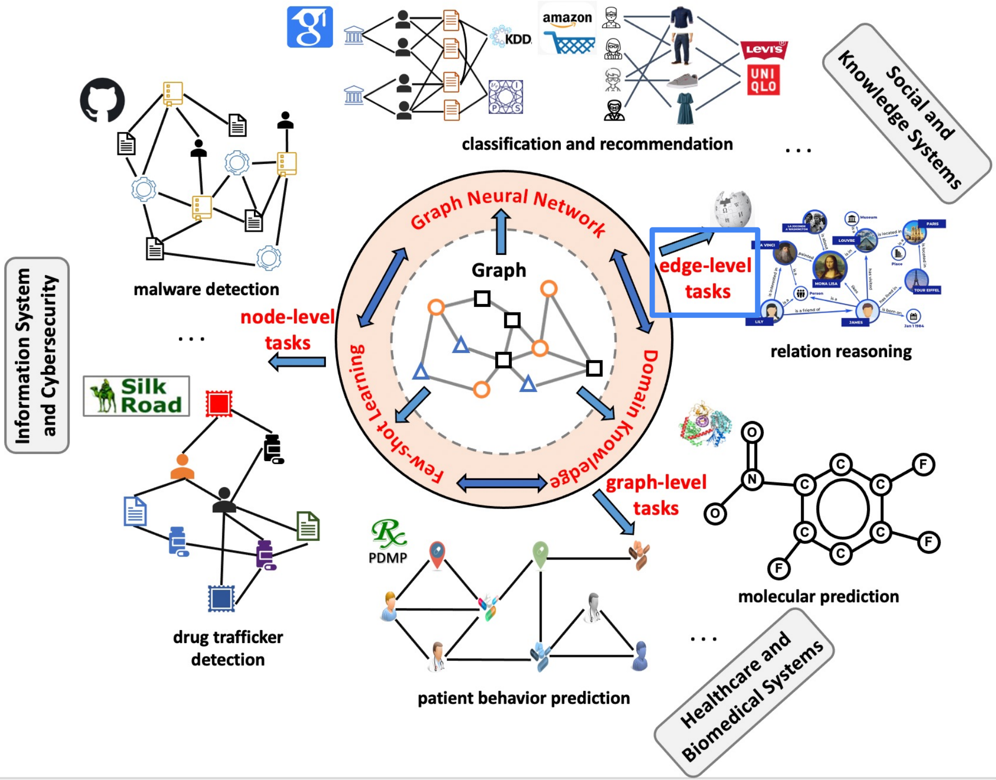
Observation: results indicate the effectiveness of MetaHG

Code: https://shorturl.at/rJV28

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

❖ Link Prediction over Knowledge Graphs

FSRL (AAAI 2020)

❖ Relation Reasoning over Knowledge Graphs

Meta-AHIN (EMNLP 2020 Findings)

Few-Shot Knowledge Graph Completion

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 ∈ E and r_q ∈ R.

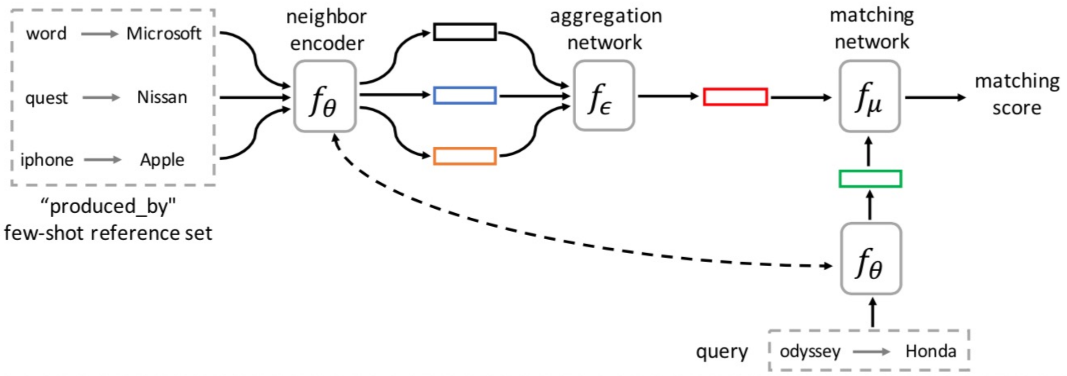
Given: support set (triplet set) and a query (e_s, r_q, ?), where e_s is the source entity and r_q is the query few-shot relation (support set size = k), relation type: class in few-shot learning

Predict: the target entity e_o for this query

Approach: metric-based method (matching network)

Motivation: knowledge from existing relations is useful

Training	
Task #1 (CountryCapital)	
Support	(China, CountryCapital, Beijing)
Query	(France, CountryCapital, Paris)
Task #2 (CEOof)	
Support	(Satya Nadella, CEOof, Microsoft)
Query	(Jack Dorsey, CEOof, Twitter)
Testing	
Task #1 (OfficialLanguage)	
Support	(Japan, OfficialLanguage, Japanese)
Query	(Spain, OfficialLanguage, Spanish)

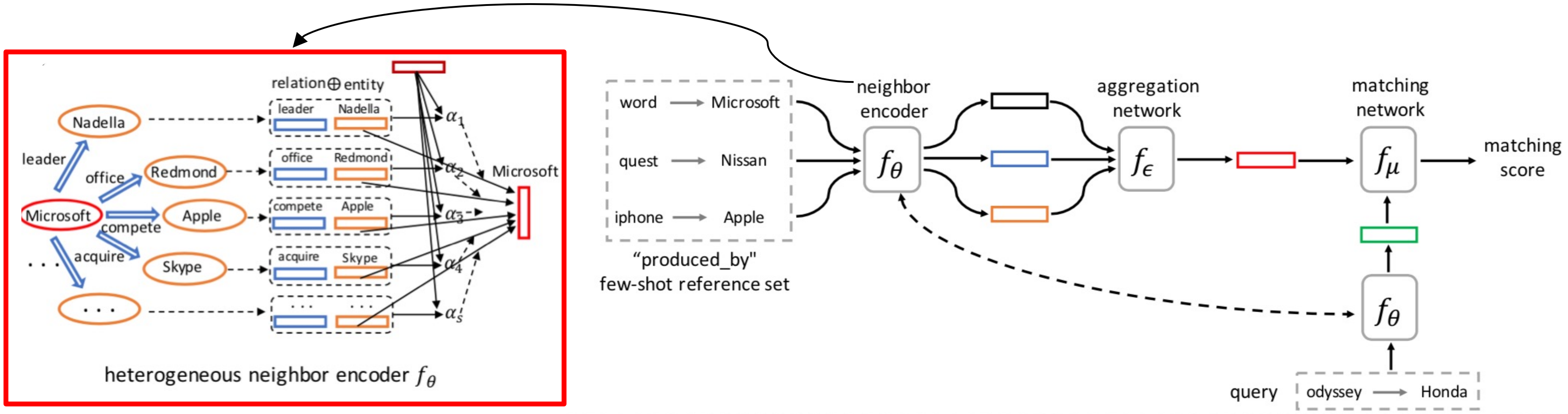


Few-Shot Knowledge Graph Completion, AAAI 2020

Few-Shot Knowledge Graph Completion

Step (a): learn entity embedding (entity pair embedding)

Detail: heterogeneous neighbor encoder (attention network)

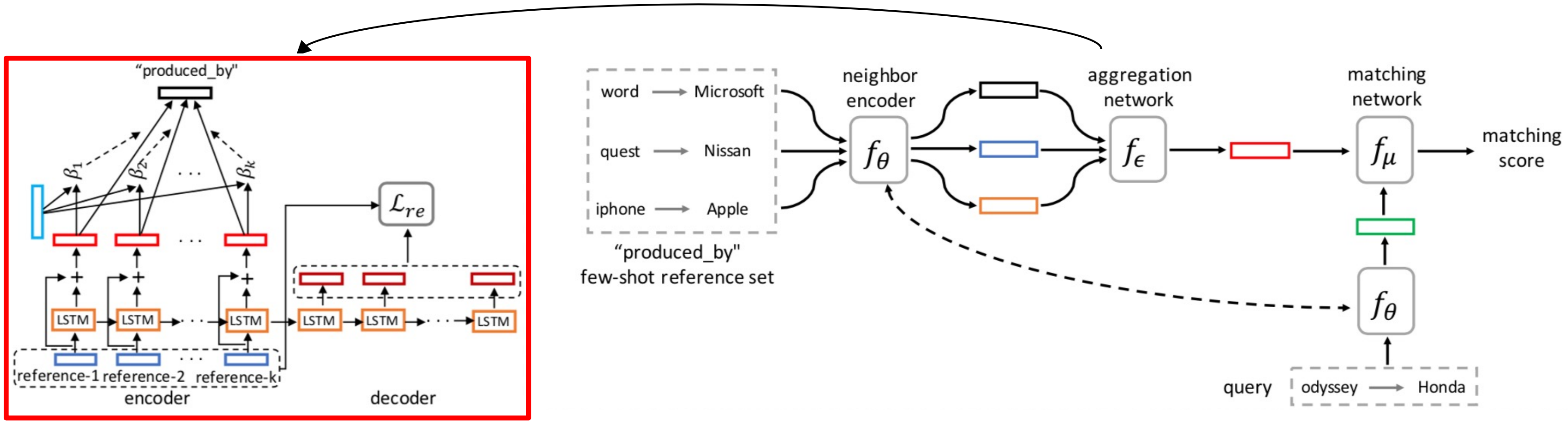


Few-Shot Knowledge Graph Completion, AAAI 2020

Few-Shot Knowledge Graph Completion

Step (b): learn support set embedding

Detail: support set aggregation network (LSTM aggregator: GraphSAGE)

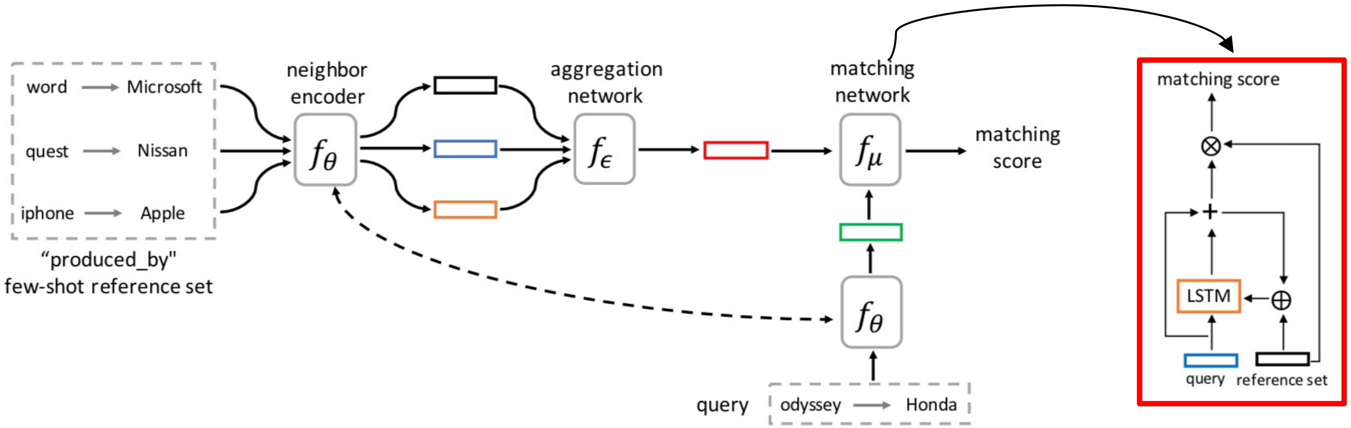


Few-Shot Knowledge Graph Completion, AAAI 2020

Few-Shot Knowledge Graph Completion

Step (c): matching between support set and query set

Detail: matching network



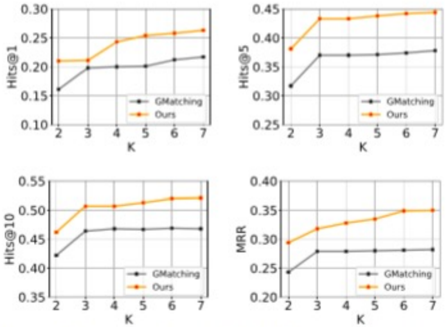
$$\ell_\theta = max(0, \gamma + score_\theta^- - score_\theta^+)$$

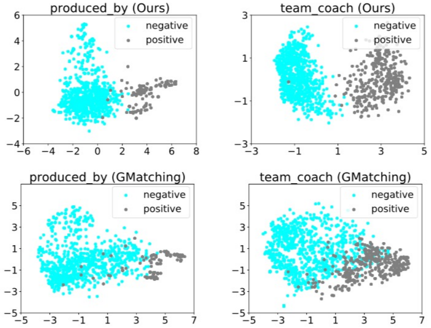
Few-Shot Knowledge Graph Completion, AAAI 2020

Few-Shot Knowledge Graph Completion

Few-shot link prediction on knowledge graphs

	Data: NELL				Data: Wiki			
Model	Hits@1	Hits@5	Hits@10	MRR	Hits@1	Hits@5	Hits@10	MRR
RESCAL	.069/.141	.160/.313	.204/.383	.119/.223	.259/.057	.297/.090	.309/.126	.279/.081
TransE	.056/.119	.112/.256	.189/.320	.104/.193	.186/.069	.352/.134	.431/.176	.273/.111
DistMult	.066/.164	.123/.306	.178/.375	.109/.231	.271/.069	.419/.156	.459/.195	.339/.112
ComplEx	.049/.129	.092/.223	.112/.273	.079/.185	.226/.085	.315/.117	.397/.145	.282/.106
GMatching (MaxP)	.244/.198	.418/.370	.524/.464	.331/.279	.313/.095	.402/.235	.468/.324	.346/.171
GMatching (MeanP)	.257/.186	.455/.360	.542/.453	.341/.267	.290/.128	407/.274	.484/.350	.352/.203
GMatching (Max)	.179/.152	.391/.335	.476/.445	.273/.241	.279/.135	.396/.284	.477/.374	.342/.214
FSRL (Ours)	**.345/.211**	**.502/.433**	**.570/.507**	**.421/.318**	**.338/.155**	**.430/.327**	**.486/.406**	**.390/.241**





Observation: results indicate the effectiveness of FSRL

Code: https://shorturl.at/otAI4

Few-Shot Knowledge Graph Completion, AAAI 2020

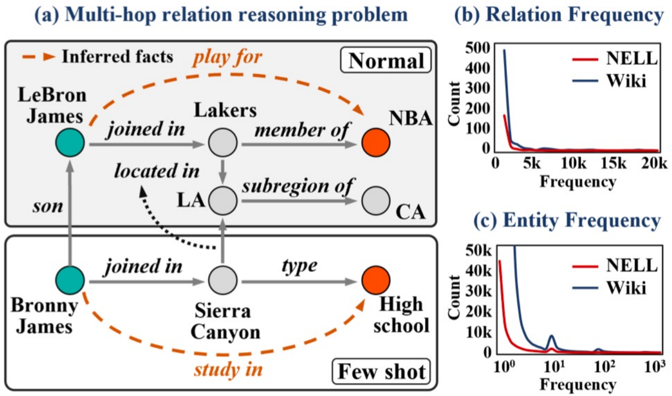
Few-Shot Relation Reasoning over Knowledge Graphs

$G = \{E, R, T\}$, 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)

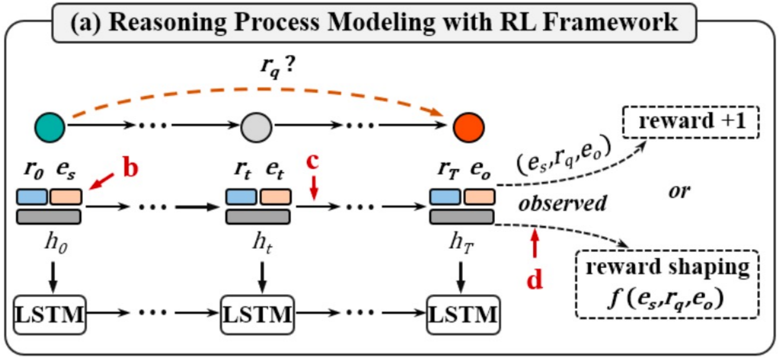
Few-Shot Relation Reasoning over Knowledge Graphs

Setting: gradient-based method (MAML)

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

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

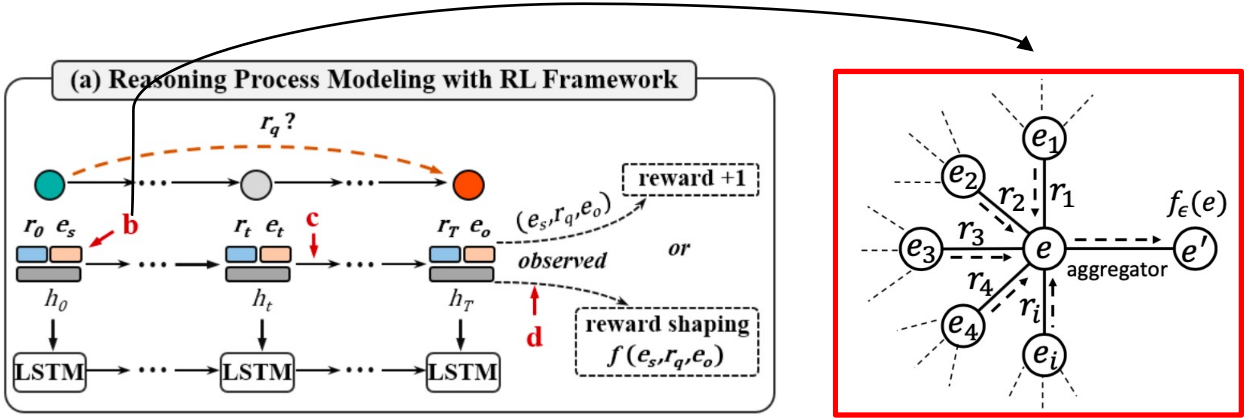
Few-Shot Multi-Hop Relation Reasoning over Knowledge Bases, EMNLP 2020 (Findings)

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



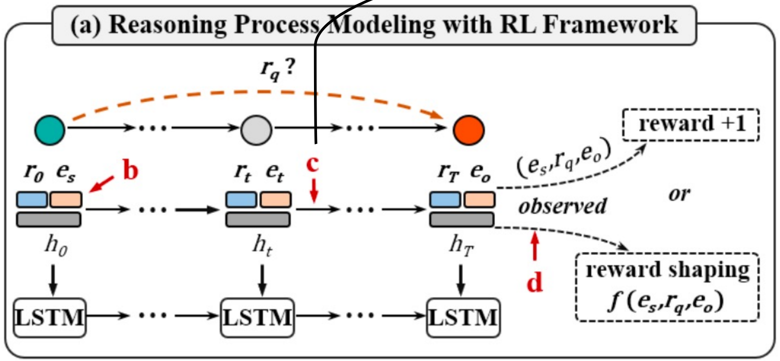
Few-Shot Multi-Hop Relation Reasoning over Knowledge Bases, EMNLP 2020 (Findings)

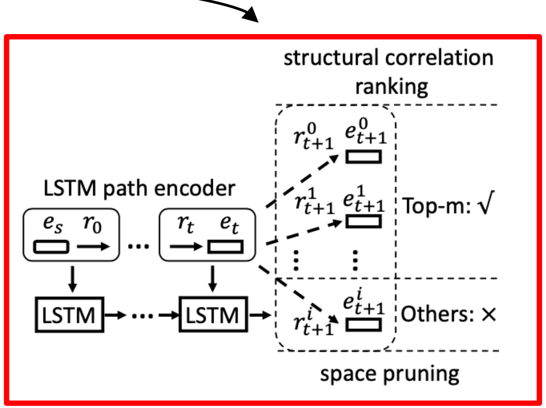
Few-Shot Relation Reasoning over Knowledge Graphs

Component (c): search space pruning

Intuition: action search space enormous or even redundant

Detail: search space pruning strategy based on embedding similarity

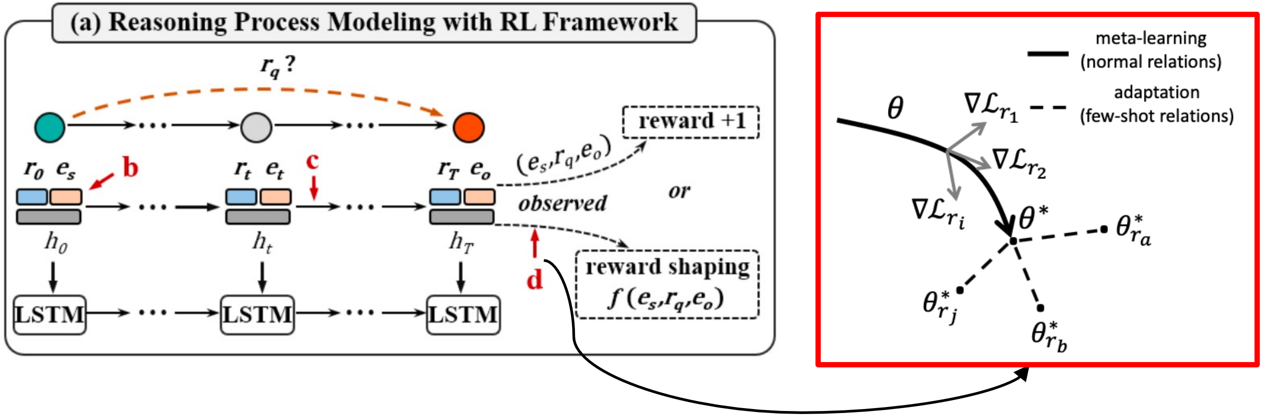




Few-Shot Multi-Hop Relation Reasoning over Knowledge Bases, EMNLP 2020 (Findings)

Few-Shot Relation Reasoning over Knowledge Graphs

Component (d): optimization with meta-learning
Intuition: initialize parameters with normal relation data
Detail: MAML for fast adaption over few-shot relation data

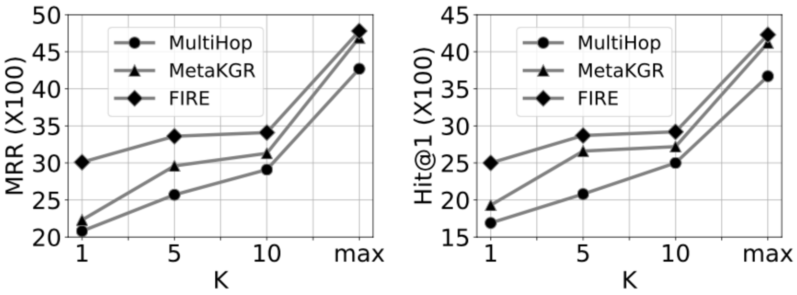


Few-Shot Multi-Hop Relation Reasoning over Knowledge Bases, EMNLP 2020 (Findings)

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	25.3	19.7	46.9	41.2
FIRE	**27.3**	**22.5**	**47.8**	**42.3**

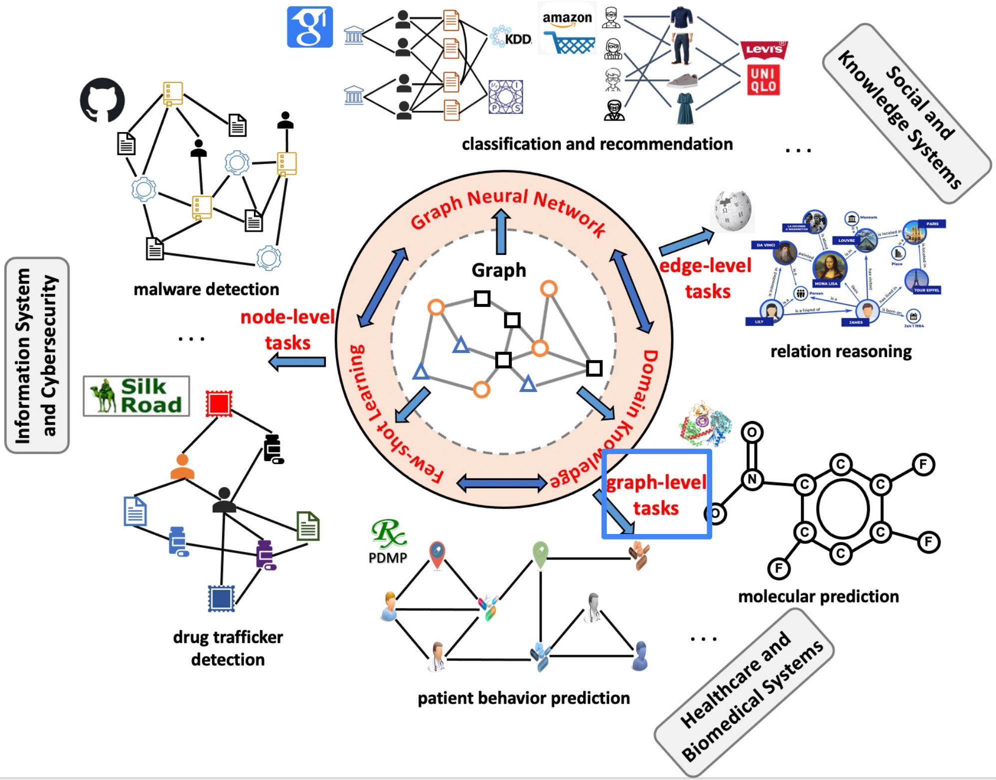


Observation: results indicate the effectiveness of FIRE

Code: https://shorturl.at/suwB6

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

❑ Graph Few-shot Learning



Graph-level Learning Tasks

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

- Graph Classification Application

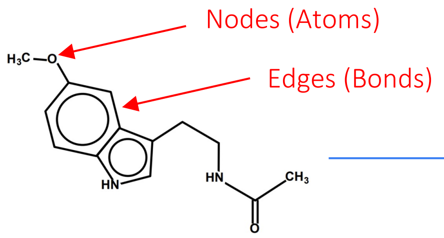
Meta-MGNN (WWW 2021): Few-shot Molecular Property Prediction

Few-Shot Graph Learning for Molecular Property Prediction

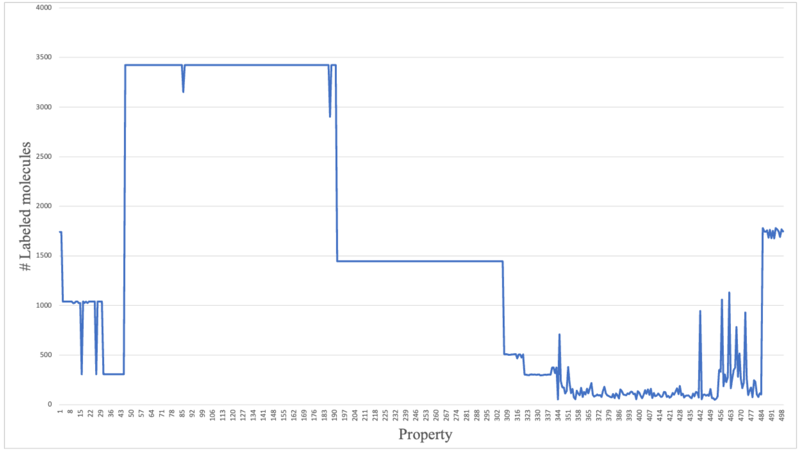
Molecular graph represents a chemical atom and an edge represents a chemical bond between two atoms. Molecular property prediction is to learn a molecular representation vector for predicting its label (i.e., molecular property).

Given: molecular properties and their corresponding molecular graph sets

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



→ 0 / 1 ? { Solubility, Toxicity, Activity }

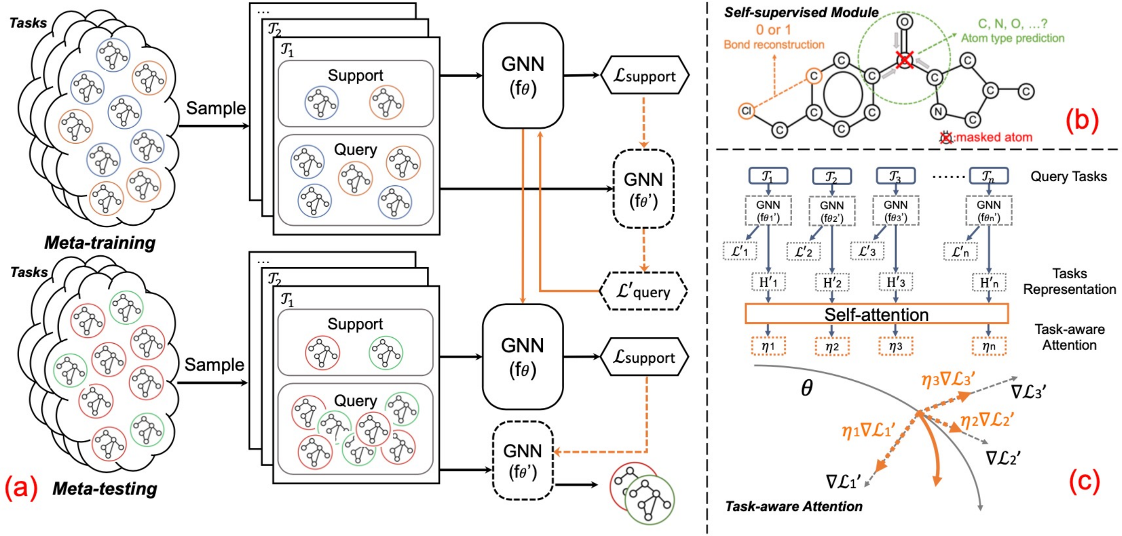


A key step in new drugs discovery

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

Few-Shot Graph Learning for Molecular Property Prediction

Setting: gradient-based method (MAML)
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

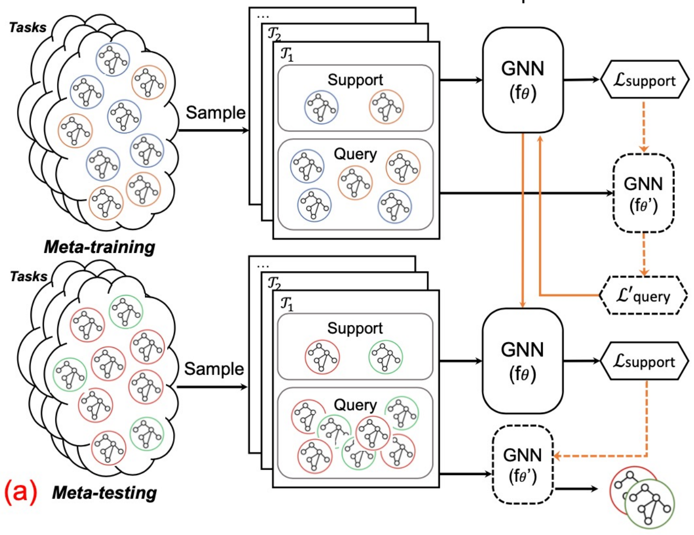
Few-Shot Graph Learning for Molecular Property Prediction

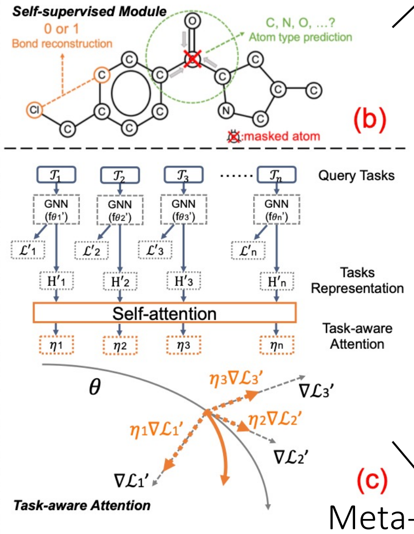
domain information

# 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





(c)

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_\tau \sim p(\mathcal{T})} \eta(\mathcal{T}_\tau) \cdot \mathcal{L}'_{\mathcal{T}_\tau}(\theta'_\tau).$$

Meta-learning with task attention

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

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

Few-Shot Graph Learning for Molecular Property Prediction

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

Code: https://shorturl.at/stxAP

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

Summary

Model	Task	Framework	Domain
GFL [AAAI'20]	Node Classification	Metric-based	Social & Information Networks
MetaHG [NeurIPS'21]	Drug Trafficker Detection	Gradient-based	Public Health
FSRL [AAAI'20]	Link Prediction	Metric-based	Knowledge Graphs
Meta-AHIN [EMNLP'20]	Relation Reasoning	Gradient-based	Knowledge Graphs
Meta-MGNN [WWW'21]	Molecular Property Prediction	Gradient-based	Biomedicine/Healthcare

Open Problems

❑ Graph Few-shot Learning Exploration

P1: Graph model for graph few-shot learning: heterogeneous graphs, dynamic graphs

P2: Graph few-shot learning explanation
Meta training vs fast adaptation

P3: Broader applications with small labeled data
Drug overdose prediction of patients
Molecular generation

Few-shot Learning on Graphs, IJCAI 2022
https://arxiv.org/pdf/2203.09308.pdf

Thank you!

Q & A