Relational Self-Supervised Learning on Graphs

Namkyeong Lee
KAIST ISysE
Daejeon, Republic of Korea
namkyeong96@kaist.ac.kr

Dongmin Hyun
POSTECH PIAI
Pohang, Republic of Korea
dm.hyun@postech.ac.kr

Junseok Lee
KAIST ISysE
Daejeon, Republic of Korea
junseoklee@kaist.ac.kr

Chanyoung Park∗
KAIST ISysE & AI
Daejeon, Republic of Korea
cy.park@kaist.ac.kr

2022.11.15 • ChongQing

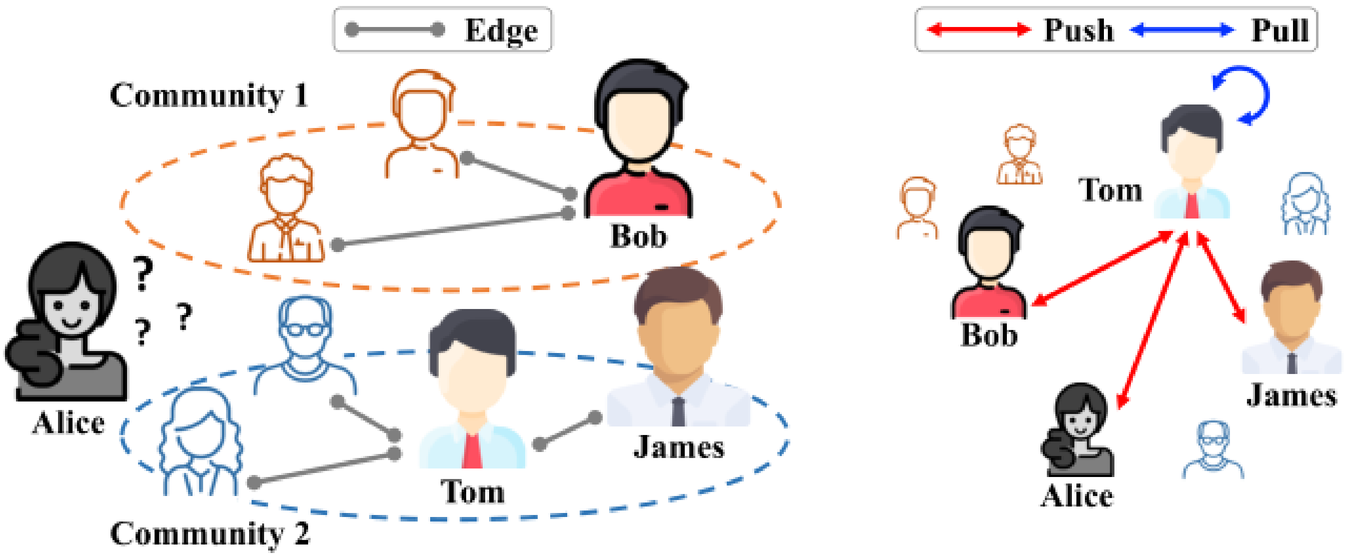
—— CIKM 2022

gesis
Leibniz-Institut
für Sozialwissenschaften

Forschungszentrum · Research Center L3S

Reported by JiaWei Cheng

Introduction



(a) Graph-Structured Data (b) Recent GRL methods

Figure 1: Recent GRL methods cannot fully benefit from the relational information of given graph-structured data.

Overview

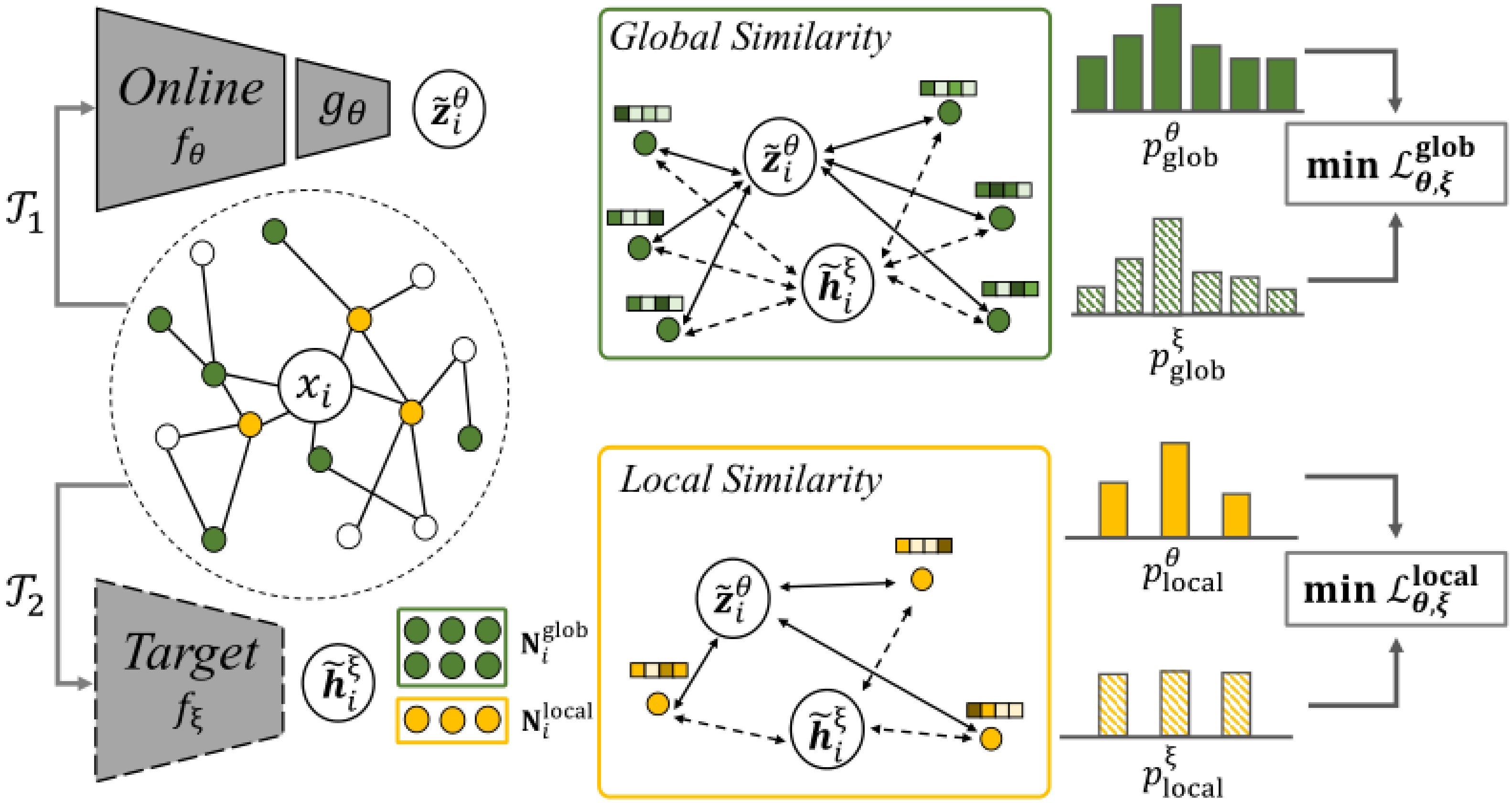


Figure 3: Overall model architecture of RGRL.

Method

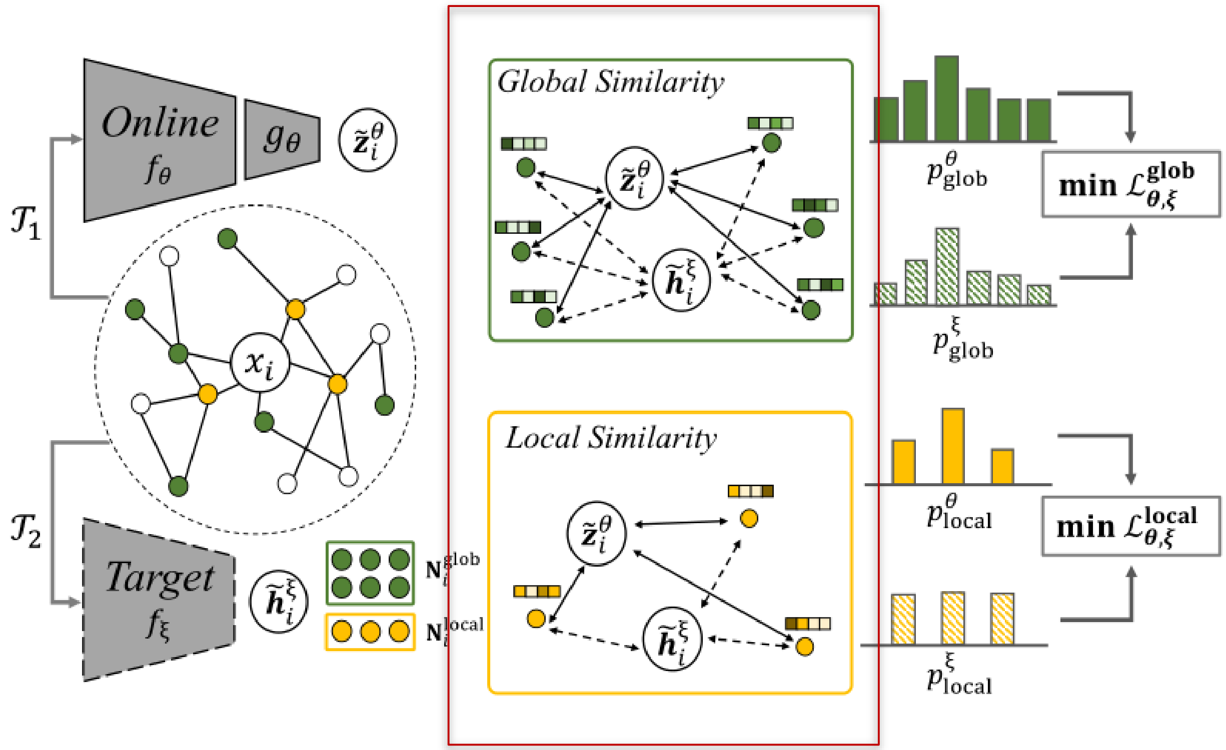


Figure 3: Overall model architecture of RGRL.

$$p_i^\xi(j) = \frac{\exp(\mathrm{sim}(\tilde{\mathbf{h}}_i^\xi, \tilde{\mathbf{h}}_j^\xi)/\tau_\xi)}{\sum_{k \in \mathbf{N}_i} \exp(\mathrm{sim}(\tilde{\mathbf{h}}_i^\xi, \tilde{\mathbf{h}}_k^\xi)/\tau_\xi)}, \forall v_j \in \mathbf{N}_i \tag{1}$$

$$p_i^\theta(j) = \frac{\exp(\mathrm{sim}(\tilde{\mathbf{z}}_i^\theta, \tilde{\mathbf{h}}_j^\xi)/\tau_\theta)}{\sum_{k \in \mathbf{N}_i} \exp(\mathrm{sim}(\tilde{\mathbf{z}}_i^\theta, \tilde{\mathbf{h}}_k^\xi)/\tau_\theta)}, \forall v_j \in \mathbf{N}_i \tag{2}$$

Method

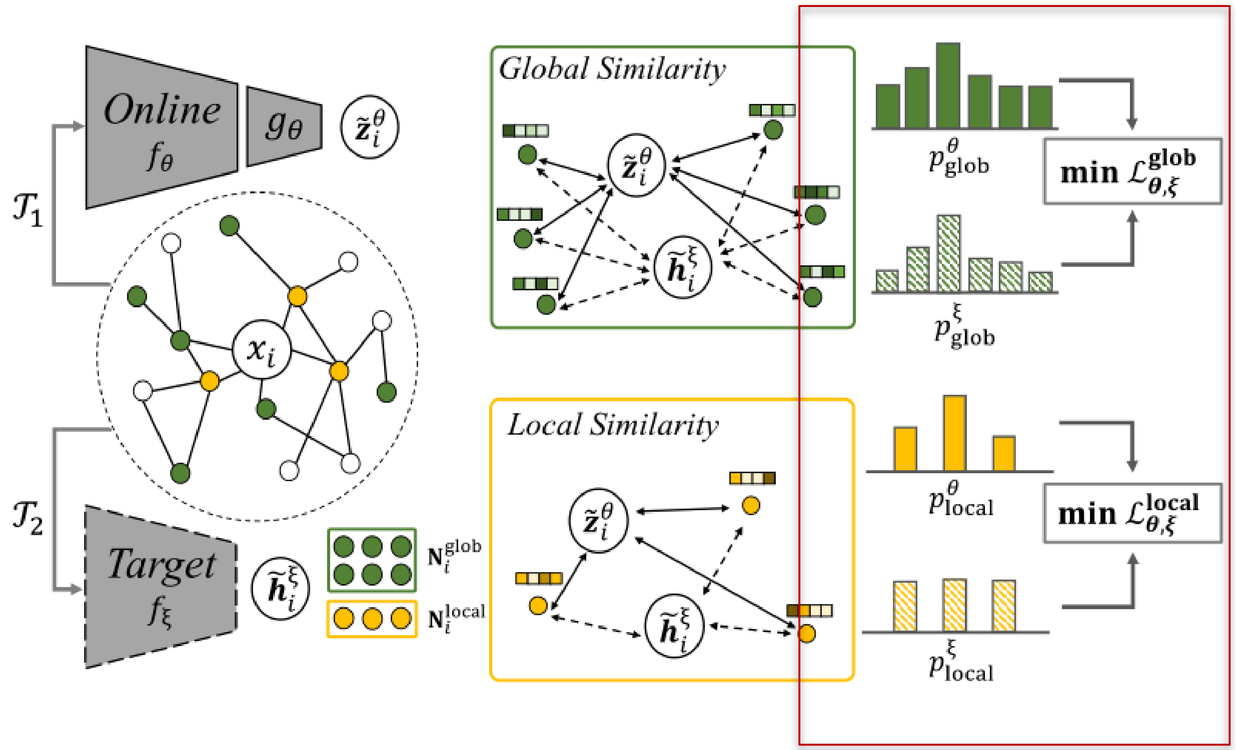


Figure 3: Overall model architecture of RGRL.

$$\mathcal{L}_{\theta,\xi} = \sum_{v_i \in \mathcal{V}} KL(p_i^\theta \,||\, p_i^\xi). \quad (3)$$

$$\mathcal{L}_{\theta,\xi} = \mathcal{L}^{glob}_{\theta,\xi} + \lambda \cdot \mathcal{L}^{local}_{\theta,\xi},$$

Method

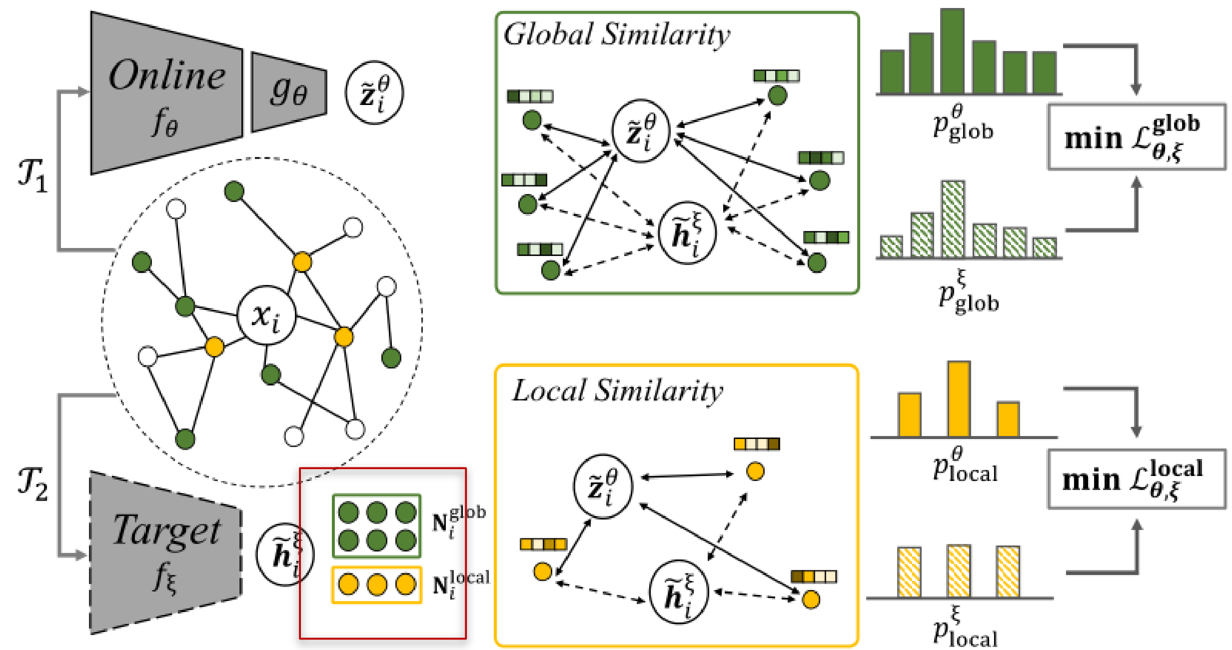


Figure 3: Overall model architecture of RGRL.

$$w_j = \alpha^{\log(\deg_j + 1)} + \beta \tag{4}$$

$$p_{sample}(j) = \frac{w_j}{\sum_{v_k \in \mathcal{V}} w_k}, \forall v_j \in \mathcal{V} \tag{5}$$

$$\mathbf{S} = \sum_{k=0}^{\infty} t(1-t)^k \mathbf{T}^k \tag{6}$$

where $t \in (0, 1)$ is the teleport probability,

$\mathbf{T} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$,

Experiments

Table 1: Comparison on computational complexity

Model	Complexity
GRACE	$4C_{\text{encoder}}(M+N) + 4C_{\text{projection}}N + C_{\text{GRACE}}(N^2)$
BGRL	$6C_{\text{encoder}}(M+N) + 4C_{\text{prediction}}N + C_{\text{BGRL}}(N)$
RGRL	$6C_{\text{encoder}}(M+N) + 4C_{\text{prediction}}N + C_{\text{RGRL}}(NK)$

Experiments

Table 2: Statistics for datasets used for experiments.

Dataset	Type	# Nodes	# Edges	# Features	# Cls.
WikiCS [2]	reference	11,701	216,123	300	10
Amazon Computers [3]	co-purchase	13,752	245,861	767	10
Amazon Photo [3]	co-purchase	7,650	119,081	745	8
Coauthor CS [3]	co-author	18,333	81,894	6,805	15
Coauthor Physics [3]	co-author	34,493	247,962	8,415	5
Cora [4]	citation	2,708	5,429	1,433	7
Citeseer [4]	citation	3,327	4,732	3,703	6
Pubmed [4]	citation	19,717	44,338	500	3
Cora Full [5]	citation	19,793	65,311	8,710	70
ogbn-arXiv [6]	citation	169,343	1,166,243	128	40
Reddit [7]	community	231,443	11,606,919	602	41
PPI (24 Graphs) [8]	interaction	56,944	818,716	50	121
IMDB [9]	co-actor co-director	3,550	66,428 13,788	2,000	3
DBLP [9]	co-author co-paper co-term	7,907	144,738 90,145 57,137,515	2,000	4

Experiments

Table 3: Performance on node classification tasks (OOM: Out of Memory on 24GB RTX3090).

	WikiCS	Computers	Photo	Co.CS	Co.Physics
GCN	77.19 (0.12)	86.51 (0.54)	92.42 (0.22)	93.03 (0.31)	95.65 (0.16)
Feats.	71.98 (0.00)	73.81 (0.00)	78.53 (0.00)	90.37 (0.00)	93.58 (0.00)
n2v	71.79 (0.05)	84.39 (0.08)	89.67 (0.12)	85.08 (0.03)	91.19 (0.04)
DW	74.35 (0.06)	85.68 (0.06)	89.44 (0.11)	84.61 (0.22)	91.77 (0.15)
DW+Feats.	77.21 (0.03)	86.28 (0.07)	90.05 (0.08)	87.70 (0.04)	94.90 (0.09)
DGI	75.35 (0.14)	83.95 (0.47)	91.61 (0.22)	92.15 (0.63)	94.51 (0.52)
GMI	74.85 (0.08)	82.21 (0.31)	90.68 (0.17)	OOM	OOM
MVGRL	77.52 (0.08)	87.52 (0.11)	91.74 (0.07)	92.11 (0.12)	95.33 (0.03)
GRACE	78.25 (0.65)	88.15 (0.43)	92.52 (0.32)	92.60 (0.11)	OOM
GCA	78.30 (0.62)	88.49 (0.51)	92.99 (0.27)	92.76 (0.16)	OOM
CCA-SSG	77.88 (0.41)	87.01 (0.41)	92.59 (0.25)	92.77 (0.17)	95.16 (0.10)
BGRL	79.60 (0.60)	89.23 (0.34)	93.06 (0.30)	92.90 (0.15)	95.43 (0.09)
RGRL	**80.29 (0.72)**	**89.70 (0.44)**	**93.43 (0.31)**	**92.94 (0.13)**	**95.46 (0.10)**

T-test

Experiments

Table 4: Performance on transductive node classification on other datasets (Accuracy), and inductive node classification on Reddit and PPI datasets (Micro-F1).

	Transductive						Inductive	
	Cora	Cite-seer	Pub-med	Cora Full	ogbn-arXiv Valid	ogbn-arXiv Test	Reddit	PPI
GRACE	83.38 (0.95)	70.79 (0.83)	83.96 (0.29)	64.19 (0.36)	OOM	OOM	94.84 (0.03)	67.12 (0.05)
GCA	82.79 (1.01)	70.70 (0.91)	84.19 (0.32)	64.34 (0.42)	OOM	OOM	94.85 (0.06)	66.72 (0.08)
CCA-SSG	83.01 (0.66)	70.35 (1.23)	84.81 (0.22)	64.09 (0.37)	59.43 (0.05)	58.50 (0.08)	94.89 (0.02)	66.09 (0.01)
BGRL	82.82 (0.86)	69.06 (0.80)	**86.16 (0.19)**	63.94 (0.39)	70.66 (0.06)	69.61 (0.09)	94.90 (0.04)	68.89 (0.08)
RGRL	**83.98 (0.78)**	**71.29 (0.87)**	85.33 (0.20)	**64.62 (0.39)**	**72.34 (0.09)**	**71.49 (0.08)**	**95.04 (0.03)**	**69.28 (0.06)**

Experiments

Table 5: Performance on link prediction with random and hard negative edges.

		Computers		Photo		Co. CS		Co. Physics	
		AUC	AP	AUC	AP	AUC	AP	AUC	AP
Random Neg.	GRACE	0.939	0.939	0.962	0.960	0.970	0.970	OOM	OOM
	GCA	0.954	0.954	0.965	0.960	**0.971**	**0.970**	OOM	OOM
	CCA-SSG	0.961	0.959	0.973	0.970	0.949	0.950	0.943	0.936
	BGRL	0.964	0.961	0.978	0.976	0.952	0.948	0.952	0.947
	RGRL	**0.974**	**0.972**	**0.983**	**0.981**	0.967	0.968	**0.964**	**0.963**
Hard Neg.	GRACE	0.933	0.933	0.939	0.929	0.870	0.868	OOM	OOM
	GCA	0.938	0.929	0.948	0.939	0.874	0.869	OOM	OOM
	CCA-SSG	0.954	0.952	0.947	0.943	0.847	0.835	0.871	0.856
	BGRL	0.959	0.956	0.959	0.956	0.845	0.832	0.903	0.892
	RGRL	**0.969**	**0.968**	**0.967**	**0.964**	**0.878**	**0.881**	**0.923**	**0.919**

Experiments

Table 6: Performance on multiplex network.

Dataset	IMDB		DBLP	
Metric	Macro-F1	Micro-F1	Macro-F1	Micro-F1
HAN	0.599	0.607	0.716	0.708
DMGI	0.648	0.648	0.771	0.766
$DMGI_{attn}$	0.602	0.606	0.778	0.770
HDMI	0.650	**0.658**	0.820	0.811
BGRL	0.631	0.634	0.819	0.807
RGRL	**0.653**	**0.658**	**0.830**	**0.818**

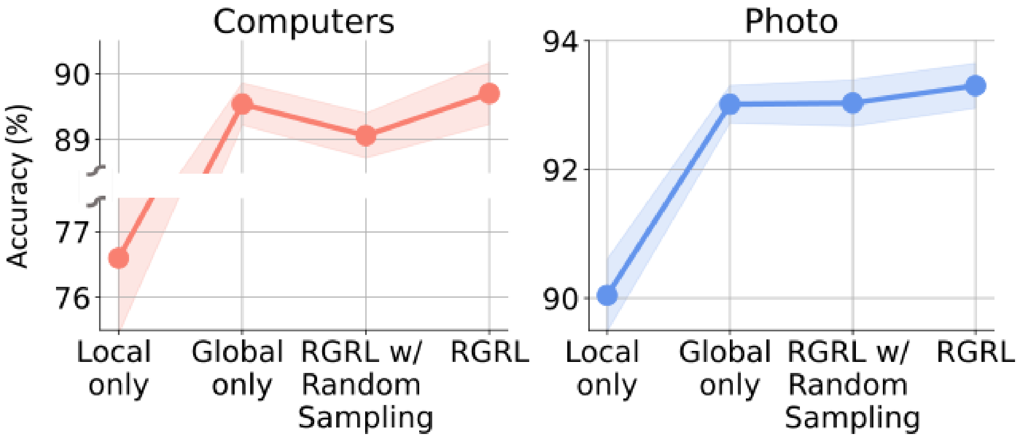


Figure 6: Ablation studies.

Experiments

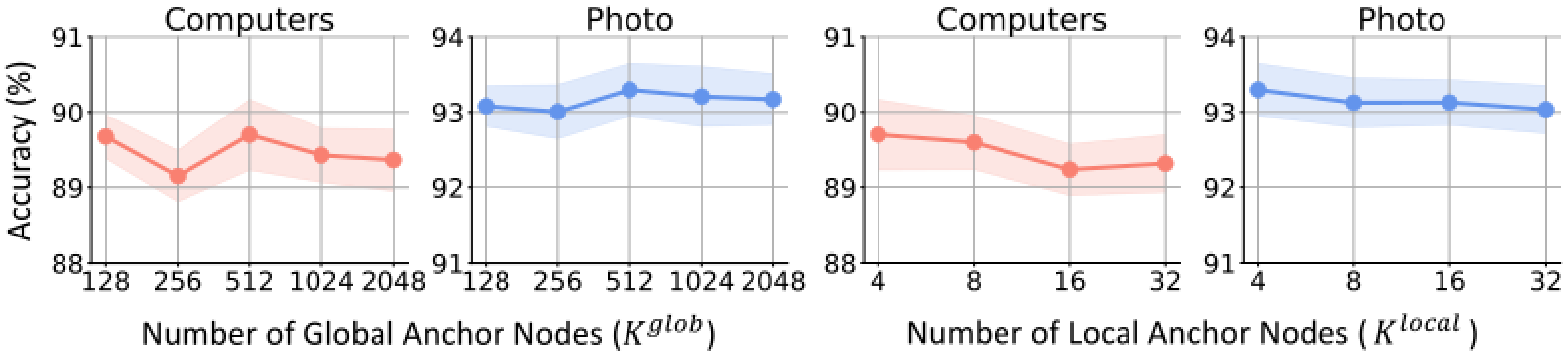


Figure 8: Sensitivity analysis on number of global anchor nodes (Left) and local anchor nodes (Right).

Thanks!