Task-Adaptive Few-shot Node Classification

Song Wang
University of Virginia
sw3wv@virginia.edu

Kaize Ding
Arizona State University
kding9@asu.edu

Chuxu Zhang
Brandeis University
chuxuzhang@brandeis.edu

Jundong Li
University of Virginia
jundong@virginia.edu

Chen Chen
University of Virginia
zrh6du@virginia.edu

2022.10.22 • ChongQing

—— KDD 2022

https://github.com/SongW-SW/TENT

Leibniz-Institut für Sozialwissenschaften

Reported by JiaWei Cheng

Introduction

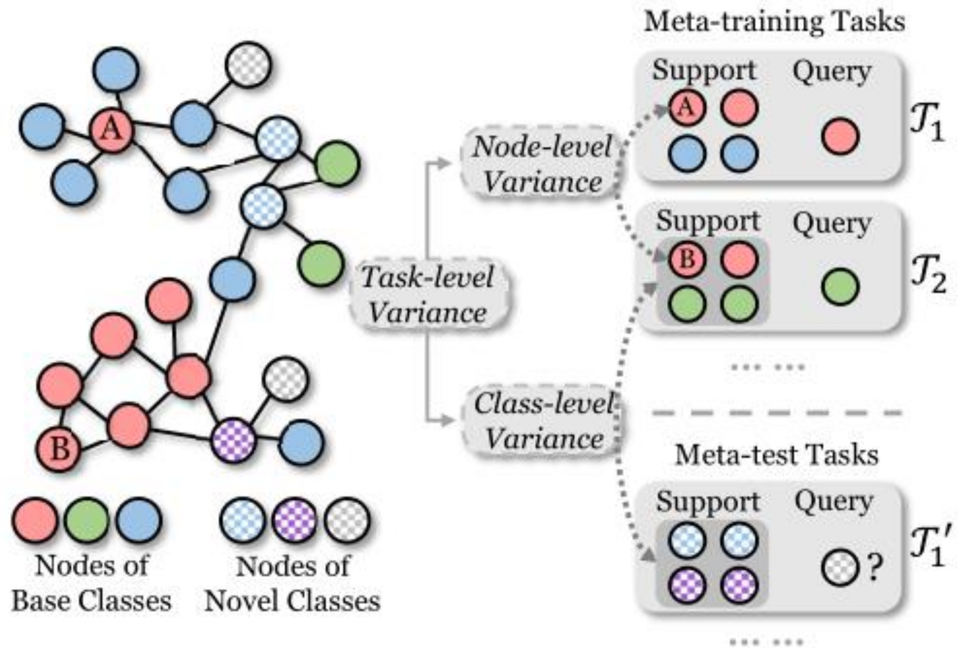


Figure 1: Issues of task variance of existing few-shot node classification frameworks.

Node-level variance represents the differences of node features and local structures of nodes across different meta-tasks.

Class-level variance denotes the difference in class distributions among meta-tasks.

Overview

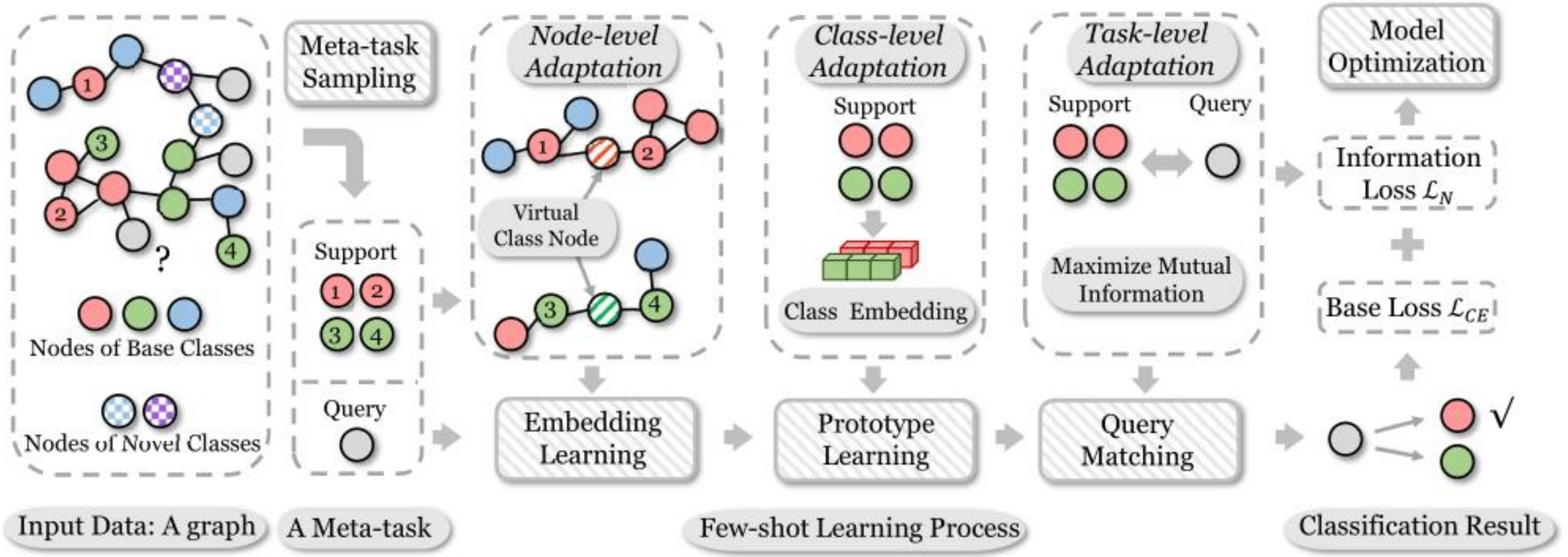
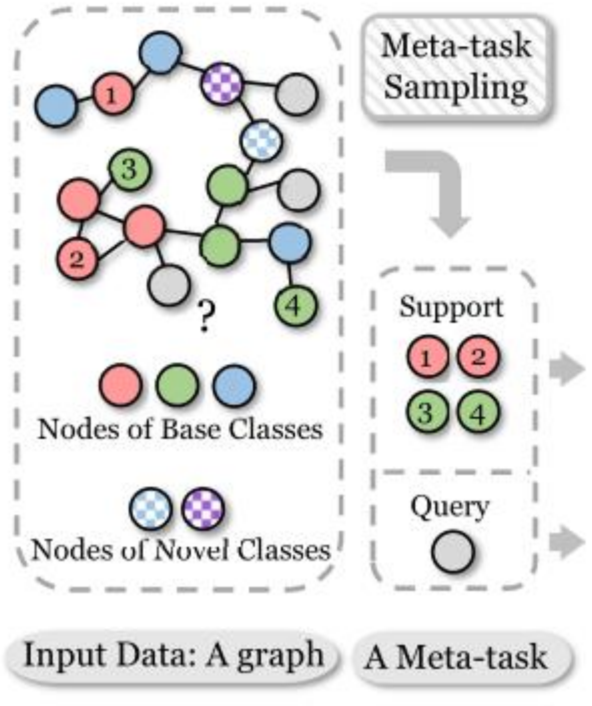


Figure 2: An illustration of the overall process of TENT. We first sample a meta-task from the given graph. Then we construct subgraphs for node-level adaptions and utilize node embeddings in each class for class-level adaptations. We further maximize the mutual information between the support set and the query set during query matching for task-level adaptations.

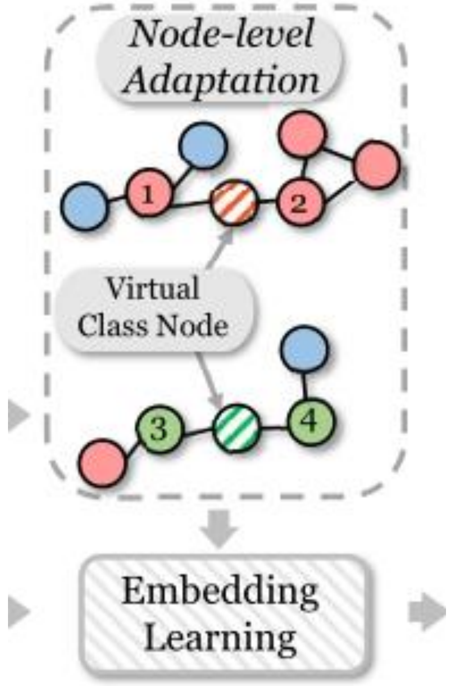
Method



$$\begin{aligned} \mathcal{S}_t &= \{(v_1, y_1), (v_2, y_2), \ldots, (v_{N\times K}, y_{N\times K})\}, \\ \mathcal{Q}_t &= \{(q_1, y'_1), (q_2, y'_2), \ldots, (q_Q, y'_Q)\}, \\ \mathcal{T}_t &= \{\mathcal{S}_t, \mathcal{Q}_t\}, \end{aligned} \tag{1}$$

Method

Node-level Adaptation

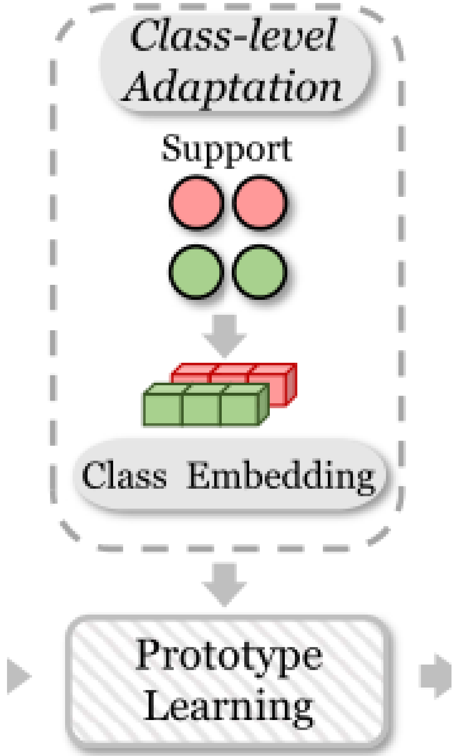


$$\mathbf{H} = \mathrm{GNN}_{\phi}(\mathcal{V}, \mathcal{E}, \mathbf{X}), \quad (2)$$

$$\mathbf{h}_{c_i} = \mathrm{MEAN}(\mathbf{h}_v | v \in \mathcal{S}_i), \quad (3)$$

Method

Class-level Adaptation



$$\alpha_i = \mathrm{MLP}_\alpha\left(\mathrm{MEAN}\left(\{\mathbf{h}_v | v \in \mathcal{S}_i\}\right)\right), \tag{4}$$

$$\beta_i = \mathrm{MLP}_\beta\left(\mathrm{MEAN}\left(\{\mathbf{h}_v | v \in \mathcal{S}_i\}\right)\right), \tag{5}$$

$$\theta_i = (\alpha_i + \mathbf{1}) \circ \theta + \beta_i, \tag{6}$$

$$\mathbf{s}_i = \mathrm{Centroid}\left(\mathrm{GNN}_{\theta_i}(\mathcal{V}_i, \mathcal{E}_i, \mathbf{X}_i)\right), \tag{7}$$

$$\alpha_q = \mathrm{MLP}_\alpha\left(\mathrm{MEAN}\left(\{\mathbf{h}_v | v \in \mathcal{S}\}\right)\right), \tag{8}$$

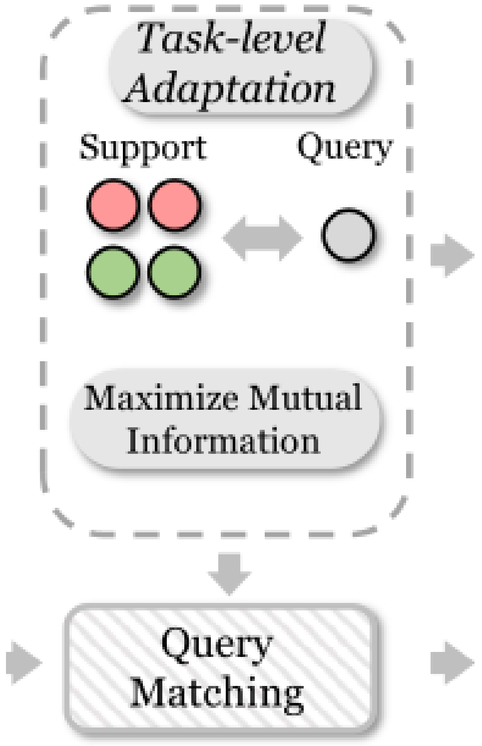
$$\beta_q = \mathrm{MLP}_\beta\left(\mathrm{MEAN}\left(\{\mathbf{h}_v | v \in \mathcal{S}\}\right)\right), \tag{9}$$

$$\theta_q = (\alpha_q + \mathbf{1}) \circ \theta + \beta_q, \tag{10}$$

$$\mathbf{q}_i = \mathrm{Centroid}\left(\mathrm{GNN}_{\theta_q}(\mathcal{V}_i^q, \mathcal{E}_i^q, \mathbf{X}_i^q)\right), \tag{11}$$

Method

Task-level Adaptation



$$\max_{\tilde{\theta}} I(\mathrm{Q};\mathrm{S}) = \max_{\tilde{\theta}} \sum_{i=1}^{Q}\sum_{j=1}^{N} p(q_i, s_j; \tilde{\theta}) \log \frac{p(q_i|s_j;\tilde{\theta})}{p(q_i;\tilde{\theta})}, \qquad (12)$$

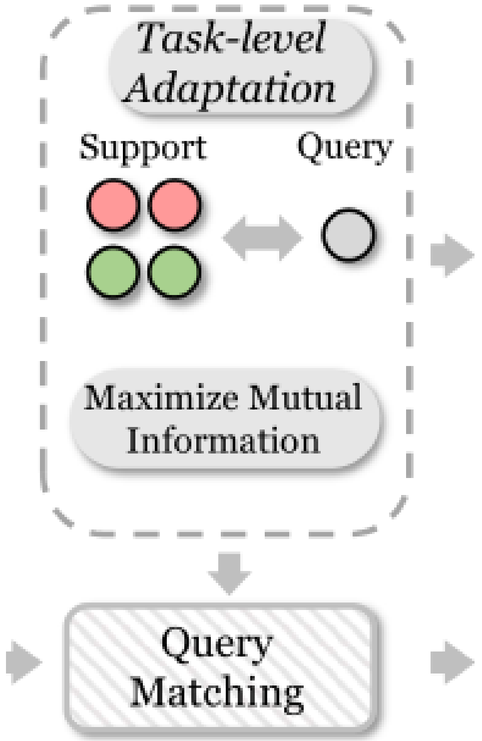
$$I(\mathrm{Q};\mathrm{S}) = \sum_{i=1}^{Q}\sum_{j=1}^{N} p(q_i|s_j;\tilde{\theta}) p(s_j;\tilde{\theta}) \log \frac{p(q_i|s_j;\tilde{\theta})}{p(q_i;\tilde{\theta})}. \qquad (13)$$

$$I(\mathrm{Q};\mathrm{S}) = \frac{1}{N}\sum_{i=1}^{Q}\sum_{j=1}^{N} p(q_i|s_j;\tilde{\theta}) \log \frac{p(s_j|q_i;\tilde{\theta})}{p(s_j;\tilde{\theta})}$$
$$= \frac{1}{N}\sum_{i=1}^{Q}\sum_{j=1}^{N} p(q_i|s_j;\tilde{\theta}) \left(\log(p(s_j|q_i;\tilde{\theta})) - \log\left(\frac{1}{N}\right)\right). \qquad (14)$$

$$I(\mathrm{Q};\mathrm{S}) = \frac{1}{N}\sum_{i=1}^{Q}\sum_{j=1}^{N} \mathbb{1}(q_i \in s_j) \left(\log(p(s_j|q_i;\tilde{\theta})) - \log\left(\frac{1}{N}\right)\right). \qquad (15)$$

Method

Task-level Adaptation



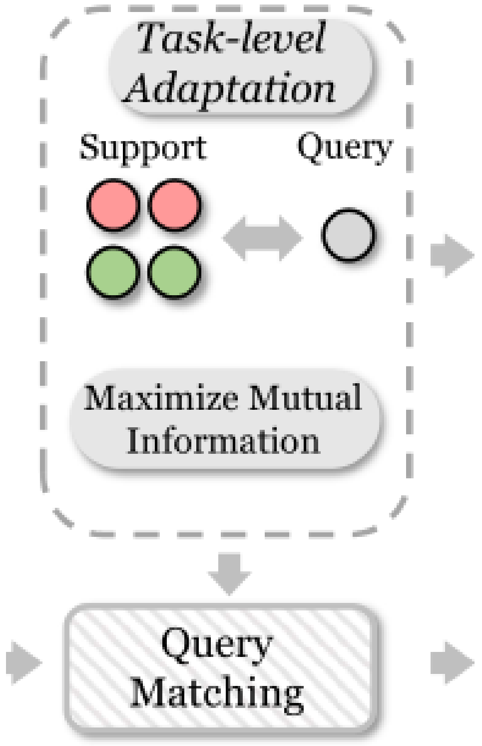
$$\sum_{i=1}^{Q}\sum_{j=1}^{N}\mathbb{1}(q_i \in s_j)\log(p(s_j|q_i;\widetilde{\theta})) = \sum_{i=1}^{Q}\log(p(s_i'|q_i;\widetilde{\theta})), \quad (16)$$

$$I(\mathrm{Q};\mathrm{S}) = \sum_{i=1}^{Q}\log(p(s_i'|q_i;\widetilde{\theta})). \quad (17)$$

$$p(s_i'|q_i;\widetilde{\theta}) = \frac{\exp\left(-(\mathbf{q}_i - \mathbf{s}_i')^2/\tau_i'\right)}{\sum_{j=1}^{N}\exp\left(-(\mathbf{q}_i - \mathbf{s}_j)^2/\tau_j\right)}, \quad (18)$$

Method

Task-level Adaptation



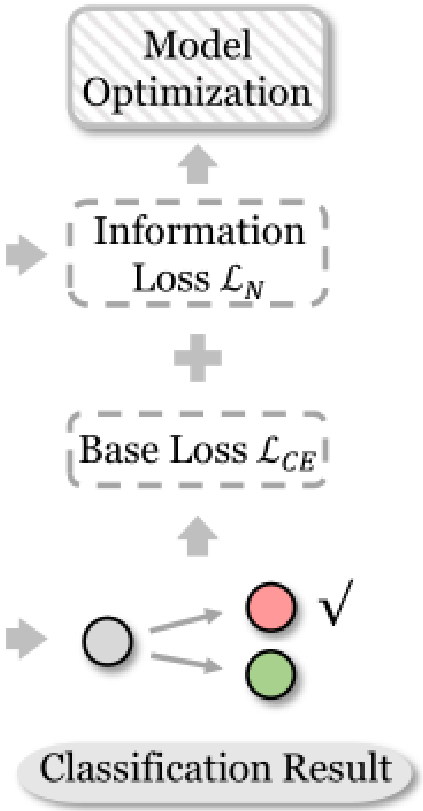
$$\max_{\widetilde{\theta}} I(Q;S) = \min_{\widetilde{\theta}} \sum_{i=1}^{Q} -\log \frac{\exp(\mathbf{q}_i \cdot \mathbf{s}_i'/\tau_i')}{\sum_{j=1}^{N} \exp(\mathbf{q}_i \cdot \mathbf{s}_j/\tau_j)}. \tag{19}$$

$$\tau_i = \frac{N \sum_k^K \|\mathbf{s}_i^k - \mathbf{s}_i\|_2}{\sum_j^N \sum_k^K \|\mathbf{s}_j^k - \mathbf{s}_j\|_2}, \tag{20}$$

$$\mathcal{L}_N = -\sum_{i=1}^{Q} \log \frac{\exp(\mathbf{q}_i \cdot \mathbf{s}_i'/\tau_i')}{\sum_{j=1}^{N} \exp(\mathbf{q}_i \cdot \mathbf{s}_j/\tau_j)}. \tag{21}$$

Method

Few-shot Node Classification



$$\mathbf{p}_i = \text{Softmax}\left(\text{MLP}(\mathbf{h}_i)\right), \tag{22}$$

$$\mathcal{L}_{CE} = -\sum_{i=1}^{Q}\sum_{j=1}^{|C_b|} y_{i,j} \log p_{i,j}, \tag{23}$$

$$\mathcal{L} = \mathcal{L}_N + \gamma \mathcal{L}_{CE}, \tag{24}$$

Experiments

Table 1: Statistics of four node classification datasets.

Dataset	# Nodes	# Edges	# Features	Class Split
Amazon-E	42,318	43,556	8,669	90/37/40
DBLP	40,672	288,270	7,202	80/27/30
Cora-full	19,793	65,311	8,710	25/20/25
OGBN-arxiv	169,343	1,166,243	128	15/5/20

Experiments

Table 2: The overall few-shot node classification results (accuracy in %) of various models under different few-shot settings.

Dataset	DBLP				Amazon-E			
Setting	5-way 3-shot	5-way 5-shot	10-way 3-shot	10-way 5-shot	5-way 3-shot	5-way 5-shot	10-way 3-shot	10-way 5-shot
PN [28]	41.51 ± 3.60	46.17 ± 3.55	28.98 ± 3.87	36.71 ± 3.35	56.80 ± 3.60	62.53 ± 2.80	44.26 ± 2.64	48.20 ± 3.89
MAML [7]	43.06 ± 2.92	49.93 ± 2.57	34.63 ± 3.91	38.44 ± 3.25	56.03 ± 2.11	63.40 ± 3.33	40.80 ± 2.75	47.06 ± 3.15
GCN [15]	62.87 ± 1.44	70.51 ± 1.37	47.22 ± 2.97	53.95 ± 2.49	55.33 ± 1.23	62.96 ± 2.61	45.18 ± 2.61	50.89 ± 2.95
G-Meta [12]	73.49 ± 2.82	78.56 ± 2.86	60.77 ± 3.03	66.26 ± 3.47	64.56 ± 4.23	68.36 ± 4.10	59.75 ± 4.90	63.02 ± 4.11
GPN [6]	76.42 ± 3.11	80.85 ± 3.68	63.14 ± 2.25	69.55 ± 2.56	65.16 ± 3.17	71.89 ± 3.94	62.52 ± 3.12	63.98 ± 2.04
RALE [18]	75.38 ± 4.94	79.85 ± 4.69	62.81 ± 3.48	67.61 ± 3.99	69.55 ± 4.24	74.97 ± 4.66	63.27 ± 3.31	64.85 ± 3.04
TENT	**79.04 ± 3.14**	**82.84 ± 3.97**	**65.47 ± 4.21**	**72.38 ± 4.14**	**75.76 ± 3.63**	**79.38 ± 4.98**	**67.59 ± 4.16**	**69.77 ± 3.76**

Dataset	Cora-full				OGBN-arxiv			
Setting	5-way 3-shot	5-way 5-shot	10-way 3-shot	10-way 5-shot	5-way 3-shot	5-way 5-shot	10-way 3-shot	10-way 5-shot
PN [28]	42.62 ± 3.78	56.66 ± 2.91	35.95 ± 3.95	38.69 ± 3.09	37.99 ± 3.98	49.71 ± 4.20	31.44 ± 3.00	35.79 ± 3.63
MAML [7]	47.10 ± 4.32	54.89 ± 3.09	30.68 ± 3.08	42.22 ± 2.76	41.83 ± 2.54	42.14 ± 3.86	33.15 ± 2.92	36.82 ± 3.03
GCN [15]	49.05 ± 2.04	58.03 ± 3.50	34.27 ± 3.98	39.85 ± 3.50	44.80 ± 2.56	47.29 ± 3.58	35.80 ± 2.21	37.78 ± 2.90
G-Meta [12]	57.93 ± 3.79	60.30 ± 2.93	45.67 ± 3.35	47.76 ± 3.25	47.66 ± 3.27	49.81 ± 4.01	35.93 ± 3.04	40.13 ± 4.35
GPN [6]	58.38 ± 3.49	63.82 ± 2.93	41.65 ± 2.20	45.63 ± 3.17	49.16 ± 3.43	53.06 ± 3.13	37.28 ± 3.99	43.33 ± 3.27
RALE [18]	62.83 ± 3.12	65.93 ± 3.24	48.05 ± 3.09	51.67 ± 3.21	53.90 ± 3.45	56.99 ± 4.43	37.60 ± 4.12	41.42 ± 3.03
TENT	**64.80 ± 4.10**	**69.24 ± 4.49**	**51.73 ± 4.34**	**56.00 ± 3.53**	**55.62 ± 3.13**	**62.96 ± 3.74**	**41.13 ± 4.26**	**44.73 ± 3.42**

Experiments

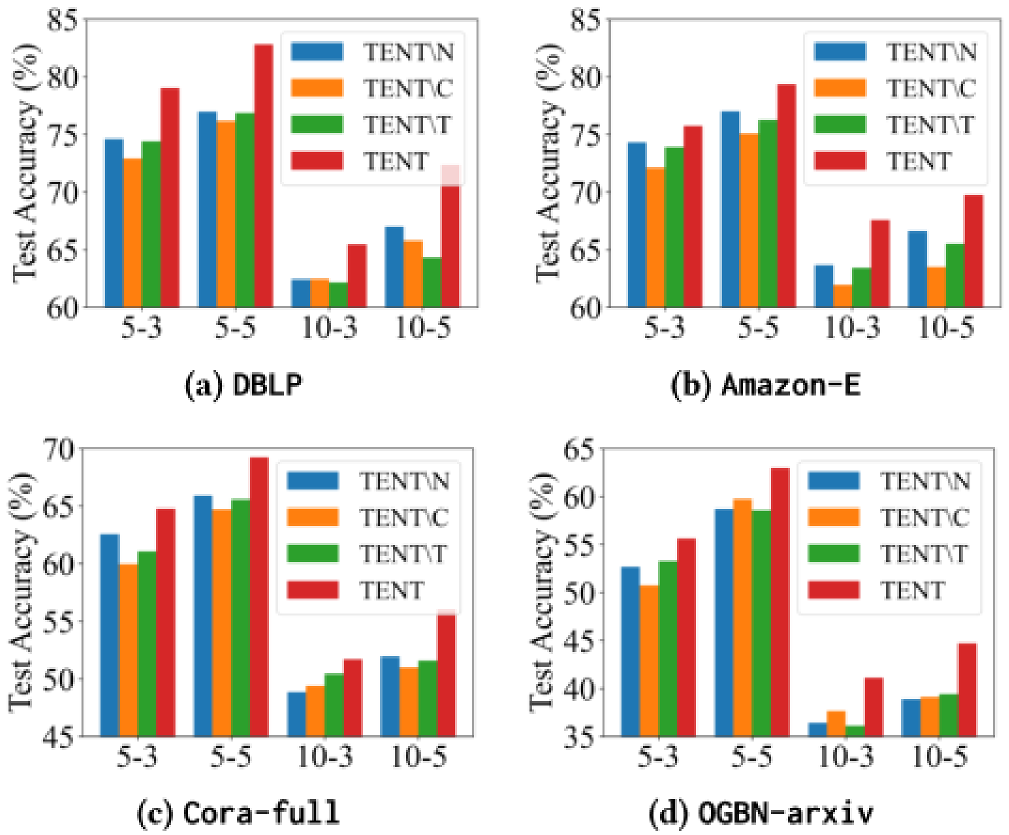


Figure 3: Ablation study on our framework in the N-way K-shot setting.

Experiments

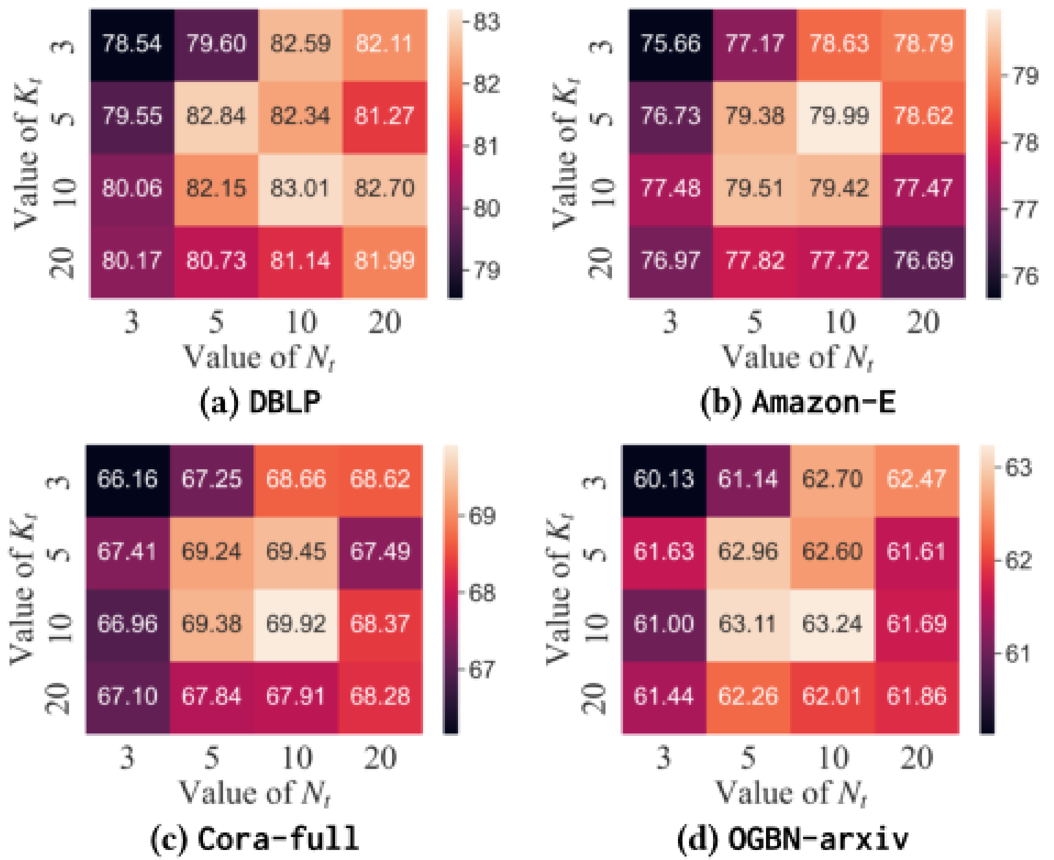


Figure 4: Results of TENT with different N_t and K_t.

Experiments

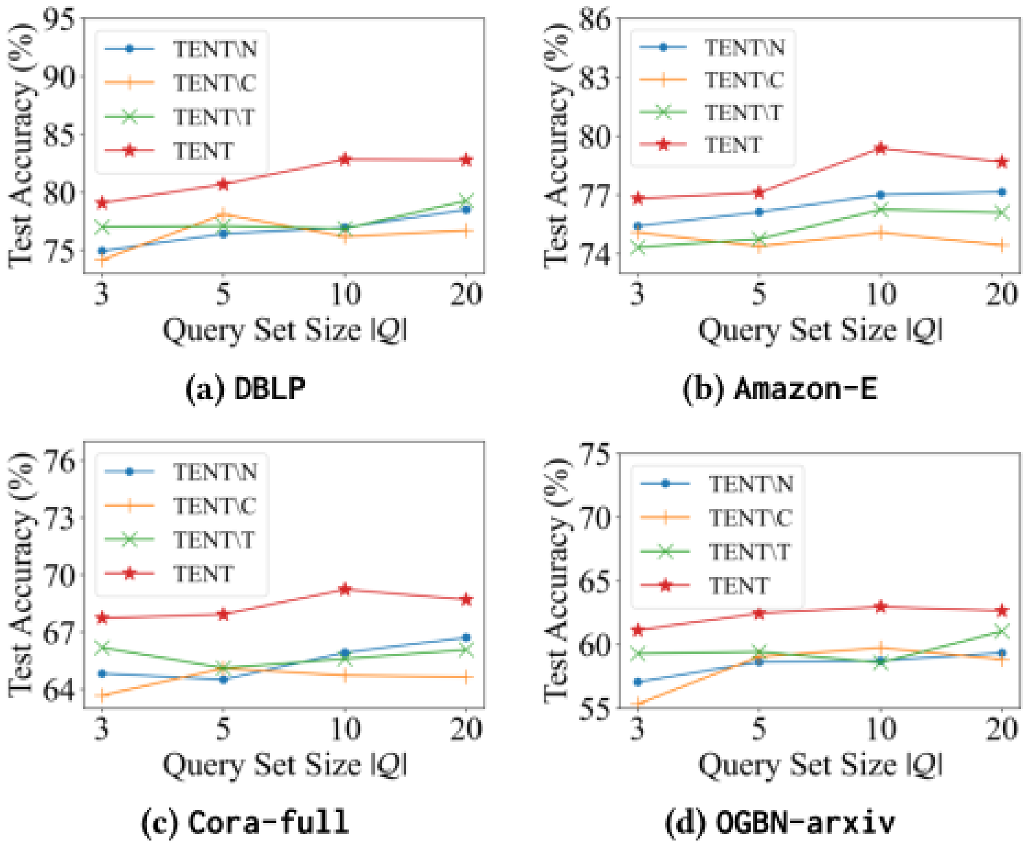


Figure 5: Results of TENT with different values of $|Q|$.

Thanks!