NEURAL INFORMATION PROCESSING SYSTEMS

Similarity-Navigated Conformal Prediction for Graph Neural Networks

Jianqing Song, Jianguo Huang, Wenyu Jiang, Baoming Zhang, Shuangjie Li, Chongjun Wang

Introduction

Conformal Prediction (CP) methods provide a theoretical guarantee for node classification tasks, ensuring that the conformal prediction set contains the ground-truth label with a desired probability (e.g., 95%).

We summarize our contributions as follows:

- Empirically explain that nodes with the same label play a critical role.
- Propose a novel algorithm, namely SNAPS.
- Theoretically show the marginal coverage properties and the validity.
- Extensive experimental results demonstrate the effectiveness.

Preliminary

Theorem 1 *(Vovk et al., 2005) Let calibration data and a test instance, i.e.,* $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n \cup \{(\boldsymbol{x}_{n+1}, y_{n+1})\}$ *be exchangeable. For any non-conformity score function* $s : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ *and any significance level* $\alpha \in (0,1)$*, define the* $1-\alpha$ *quantile of scores as* $\hat{q} := \mathrm{Quantile}\left(\frac{\lceil(1-\alpha)(n+1)\rceil}{n}; \{s(\boldsymbol{x}_i, y_i)\}_{i=1}^n\right)$ *and prediction sets as* $\mathcal{C}_\alpha(\boldsymbol{x}_{n+1}) = \{y | s(\boldsymbol{x}_{n+1}, y) \leq \hat{q}\}$*. We have*

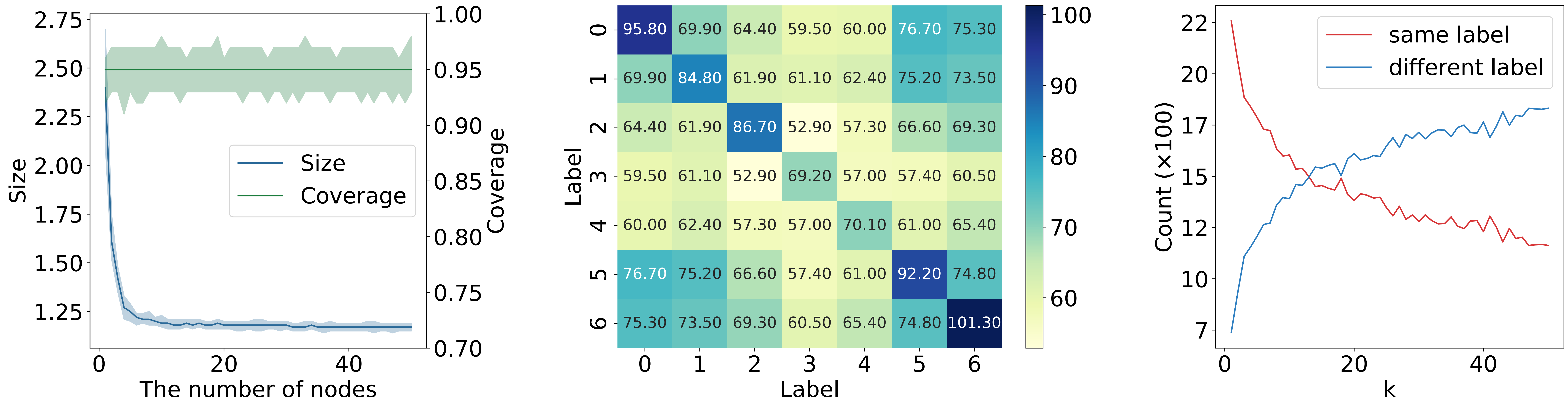
$$1 - \alpha \leq \mathbb{P}[y_{n+1} \in \mathcal{C}_\alpha(\boldsymbol{x}_{n+1})] < 1 - \alpha + \frac{1}{n+1}.$$

Adaptive Prediction Sets. (Romano et al., 2020) Given a data pair $(\boldsymbol{x}, y)$ and a predicted probability estimator $\pi(\boldsymbol{x})_y$ for $(\boldsymbol{x}, y)$, where $\pi(\boldsymbol{x})_y$ is the predicted probability for class y, the non-conformity scores can be computed by:

$$s(\boldsymbol{x}, y) = \sum_{i=1}^{|\mathcal{Y}|} \pi(\boldsymbol{x})_i \mathbb{I}[\pi(\boldsymbol{x})_i > \pi(\boldsymbol{x})_y] + \xi \cdot \pi(\boldsymbol{x})_y,$$

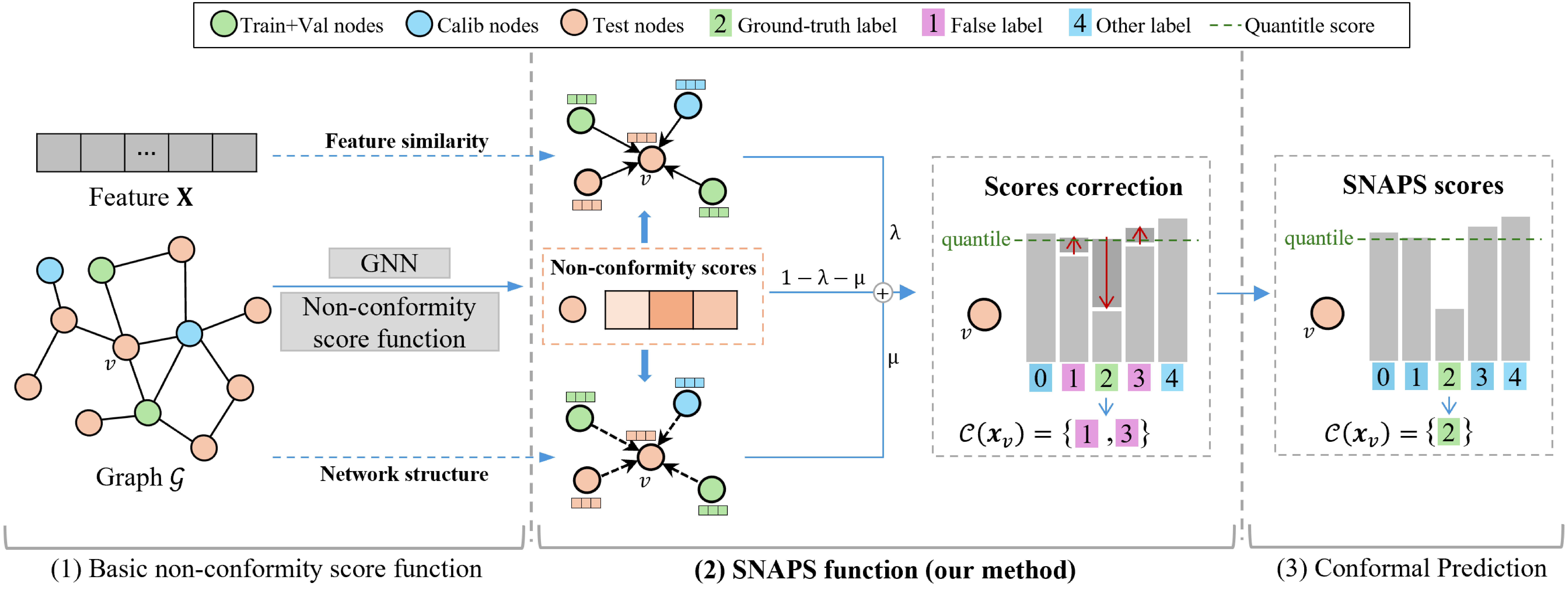
where $\xi \in [0,1]$ is a uniformly distributed random variable. Then, the prediction set is constructed as $\mathcal{C}(\boldsymbol{x}) = \{y | s(\boldsymbol{x}, y) \leq \hat{q}\}$.

Motivation



Empirically show that nodes with the same label as the ego node may play a critical role in the non-conformity scores of the ego node. Specifically, using the scores of nodes with the same label to correct the scores of the ego node could reduce the average size of prediction sets.

Framework



Train+Val nodes
Calib nodes
Test nodes
2 Ground-truth label
1 False label
4 Other label
Quantile score
Feature similarity
Feature X
GNN
Non-conformity score function
Graph $\mathcal{G}$
Network structure
Non-conformity scores
λ
$1-\lambda-\mu$
μ
Scores correction
quantile
0 1 2 3 4
$\mathcal{C}(x_v) = \{1, 3\}$
SNAPS scores
quantile
0 1 2 3 4
$\mathcal{C}(x_v) = \{2\}$
(1) Basic non-conformity score function
(2) SNAPS function (our method)
(3) Conformal Prediction

Method

Feature similarity graph construction

$$\mathrm{Sim}(i,j) = \frac{\boldsymbol{x}_i^\top \boldsymbol{x}_j}{\|\boldsymbol{x}_i\|_2 \cdot \|\boldsymbol{x}_j\|_2}$$

$$\boldsymbol{A}_s(i,j) = \mathrm{Sim}(i,j) \text{ and } \boldsymbol{D}_s(i,i) = \textstyle\sum_j \boldsymbol{A}_s(i,j).$$

Similarity-Navigated Adaptive Prediction Sets (SNAPS)

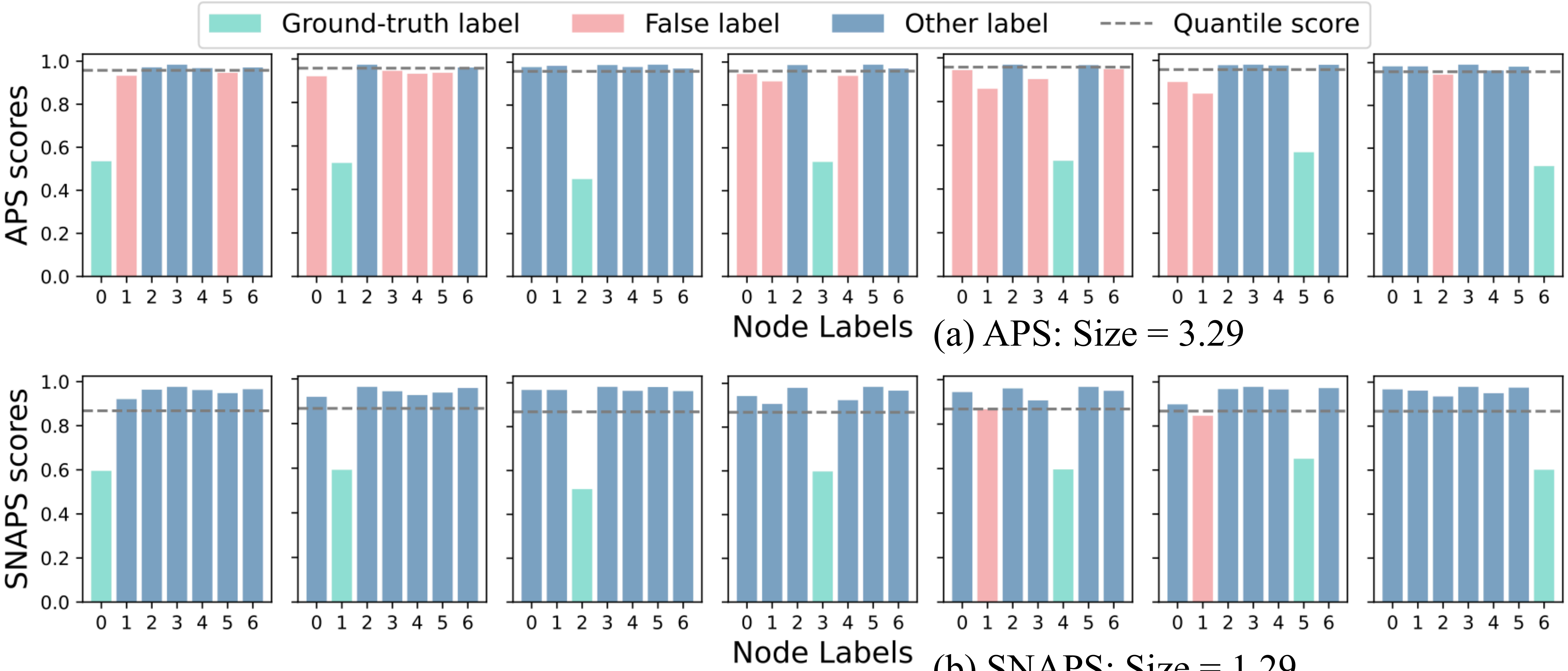
$$\hat{s}(\boldsymbol{x}_i, y) = (1-\lambda-\mu)s(\boldsymbol{x}_i, y) + \frac{\lambda}{\boldsymbol{D}_s(i,i)} \sum_{j=1}^{M} \boldsymbol{A}_s(i,j) s(\boldsymbol{x}_j, y) + \frac{\mu}{|\mathcal{N}_i|} \sum_{v_j \in \mathcal{N}_i} s(\boldsymbol{x}_j, y)$$

Experiments (1)

Table 1: Results of Coverage, Size and SH on different datasets. For SNAPS we use the APS score as the basic score. We report the average calculated from 10 GCN runs with each run of 100 conformal splits at a significance level $\alpha = 0.05$. **Bold** numbers indicate optimal performance.

Datasets	Coverage				Size↓				SH%↑			
	APS	RAPS	DAPS	SNAPS	APS	RAPS	DAPS	SNAPS	APS	RAPS	DAPS	SNAPS
CoraML	0.950	0.950	0.950	0.950	2.42	2.21	1.92	**1.68**	44.89	22.19	52.16	**56.30**
PubMed	0.950	0.950	0.950	0.950	1.79	1.77	1.76	**1.62**	33.67	30.83	35.25	**42.95**
CiteSeer	0.950	0.950	0.950	0.950	2.34	2.36	1.94	**1.84**	50.41	38.99	**59.75**	59.08
CoraFull	0.950	0.950	0.950	0.950	17.54	10.72	11.81	**9.80**	10.23	2.13	**8.67**	5.76
CS	0.950	0.950	0.950	0.950	1.91	1.20	1.22	**1.08**	66.17	78.34	79.80	**87.92**
Physics	0.950	0.950	0.950	0.950	1.28	1.07	1.08	**1.04**	76.74	88.89	88.40	**91.21**
Computers	0.950	0.950	0.950	0.950	3.95	2.89	2.13	**1.98**	27.67	15.85	43.03	**45.48**
Photo	0.951	0.950	0.950	0.951	1.89	1.64	1.41	**1.31**	54.31	56.63	74.57	**78.51**
Arxiv	0.950	0.950	0.949	0.950	4.30	3.62	3.73	**3.62**	22.55	14.52	19.19	**23.53**
Products	0.950	0.951	0.950	0.950	14.92	13.67	10.91	**7.68**	15.51	11.51	19.29	**22.38**
Average	0.950	0.950	0.950	0.950	5.23	4.12	3.79	**3.17**	40.22	36.00	48.01	**52.31**

Experiments (2)



(a) APS: Size = 3.29

(b) SNAPS: Size = 1.29

Experiments (3)

Table 4: Results on Imagenet. The median-of-means is reported over 10 different trials. **Bold** numbers indicate optimal performance.

	Accuracy		APS/SNAPS					
			$\alpha = 0.1$			$\alpha = 0.05$		
Model	Top1	Top5	Coverage	Size ↓	SSCV ↓	Coverage	Size↓	SSCV ↓
ResNeXt101	79.32	94.58	0.899/0.900	19.64/**4.08**	0.088/**0.059**	0.950/0.950	45.80/**14.41**	0.047/**0.033**
ResNet101	77.36	93.53	0.900/0.900	10.82/**3.62**	**0.075**/0.078	0.950/0.950	22.90/**9.83**	0.039/**0.029**
DenseNet161	77.19	93.56	0.900/0.900	12.04/**3.80**	0.077/**0.067**	0.951/0.950	27.99/**10.66**	0.039/**0.026**
ViT	81.02	95.33	0.899/0.899	10.50/**2.33**	**0.087**/0.133	0.949/0.950	31.12/**10.47**	0.042/**0.040**
CLIP	60.53	86.15	0.899/0.899	17.46/**10.32**	0.047/**0.032**	0.950/0.949	34.93/**24.53**	0.027/**0.017**
Average	-	-	0.899/0.900	14.09/**4.83**	0.075/**0.074**	0.950/0.950	32.55/**13.98**	0.039/**0.029**

Conclusion

- Propose a general algorithm, namely SNAPS.
- Present theoretical analyses to certify the effectiveness of this method.
- Extensive experiments demonstrate the effectiveness of SNAPS.
- Extend SNAPS to image classification.

Limitations

- Many classification tasks require inductive learning.
- Graph construction based on feature similarity is both computationally inefficient and lacking accuracy.
- Many heterophilous networks are prevalent in practice.

Thank you!