Generalized Few-Shot Node Classification

Presenter: Zhe Xu

Email: zhexu3@illinois.edu

Zhe Xu (UIUC)

Kaize Ding (ASU)

Yu-Xiong Wang (UIUC)

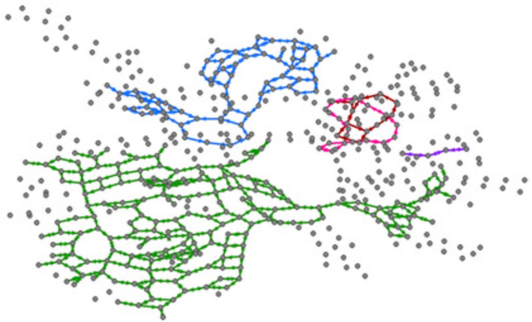
Huan Liu (ASU)

Hanghang Tong (UIUC)

Roadmap

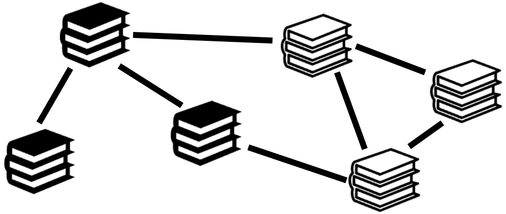
- Motivation
- Proposed Model – STAGER
- Imbalanced Episodic Training
- Experiments
- Conclusion

Graph data is ubiquitous



Traffic graph [1]

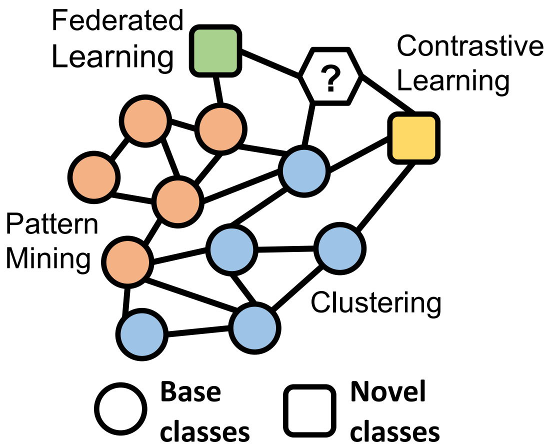
Social graph [2]



Citation graph [3]

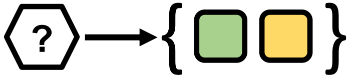
[1] Li, Daqing, et al. "Percolation transition in dynamical traffic network with evolving critical bottlenecks." Proceedings of the National Academy of Sciences 112.3 (2015): 669-672.
[2] R. Zafarani and H. Liu, (2009). Social Computing Data Repository at ASU [http://socialcomputing.asu.edu]. Tempe, AZ: Arizona State University, School of Computing, Informatics and Decision Systems Engineering.
[3] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. ArnetMiner: Extraction and Mining of Academic Social Networks. KDD. 2008

(Generalized) few-shot node classification

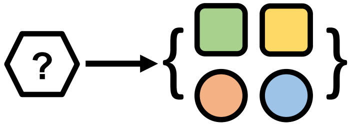


- **shot:** # of labeled samples per class
- **Base classes:** classes with many shots
- **Novel classes:** classes with few shots

- **Existing few-shot node classification researches [1-3]**



- **This work: a more realistic setting**



[1] Zhou, Fan, et al. "Meta-gnn: On few-shot node classification in graph meta-learning." CIKM. 2019.
[2] Ding, Kaize, et al. "Graph prototypical networks for few-shot learning on attributed networks." CIKM. 2020.
[3] Huang, Kexin, and Marinka Zitnik. "Graph meta learning via local subgraphs." NeurIPS. 2020.

Underestimated Novel Classes

$$P(y_i|\mathcal{G}, v_i) = \sum_{C \in \{novel, base\}} \overbrace{P(y_i|C_i, \mathcal{G}, v_i)}^{\text{fine-grained classification given priori}} \underbrace{P(C_i|\mathcal{G}, v_i)}_{\text{A rough classification priori}}$$

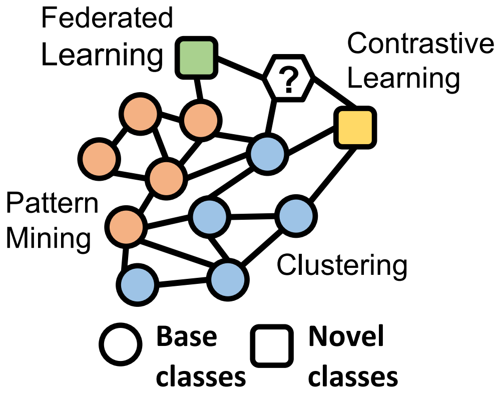
- A given graph $\mathcal{G} = \{\mathbf{A}, \mathbf{X}\}$,
- A rough classification C_i, i.e., novel or base classes

- **If $C_i = base$:** standard node classification
- **If $C_i = novel$:** few-shot node classification
- **If C_i is unknown:**

$$\mathbb{E}[P(C_i = base|G, v_i)] \gg \mathbb{E}[P(C_i = novel|G, v_i)]$$

Dataset	**Method**	$b \to b$	$b \to n$	$n \to b$	$n \to n$
Amazon Clothing	APPNP	100.0	0.0	69.6	30.4
	MetaGNN	100.0	0.0	61.2	38.8
Cora-Full	APPNP	99.2	0.8	66.8	33.2
	MetaGNN	99.0	1.0	72.4	27.6

TABLE I: Test nodes (%) classified from base classes to base classes (i.e., $b \to b$), from base classes to novel classes (i.e., $b \to n$), from novel classes to base classes (i.e., $n \to b$), and from novel classes to novel classes (i.e., $n \to n$), respectively.



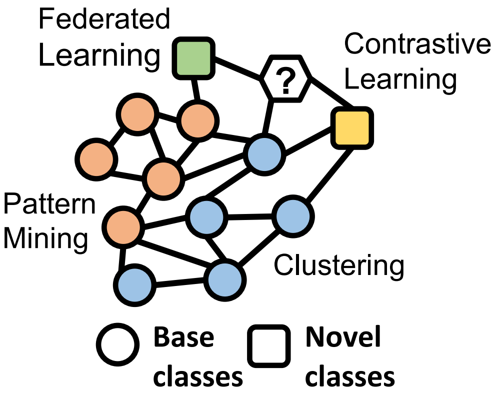
Roadmap

- Motivation ✅
- Proposed Model – STAGER
- Imbalanced Episodic Training
- Experiments
- Conclusion

Epistemic Uncertainty

- Epistemic uncertainty: how well the model fits the data and **is reducible as the size of training data increases** [1-2]

- $P(C_i = base|G, v_i)$ or $P(C_i = novel|G, v_i)$?
- **Number of shots ↔ epistemic uncertainty**



- Measure epistemic uncertainty by a **ranked** classification output from a preliminary classifier g_1 :

$$\tilde{\mathbf{Z}} = \mathrm{rank}(\mathrm{softmax}(g_1(\mathbf{A}, \mathbf{X}, \phi_1)))$$

Implementation detail: dropout variational inference on ϕ_1 [2-3]

The selection of g_1: show in a case study

[1] Alaa, Ahmed, and Mihaela Van Der Schaar. "Discriminative jackknife: Quantifying uncertainty in deep learning via higher-order influence functions." ICML. 2020.
[2] Kendall, Alex, and Yarin Gal. "What uncertainties do we need in bayesian deep learning for computer vision?." NIPS. 2017
[3] Gal, Yarin, and Zoubin Ghahramani. "Bayesian convolutional neural networks with Bernoulli approximate variational inference." arXiv preprint arXiv:1506.02158 (2015).

Proposed Model – STAGER (1)

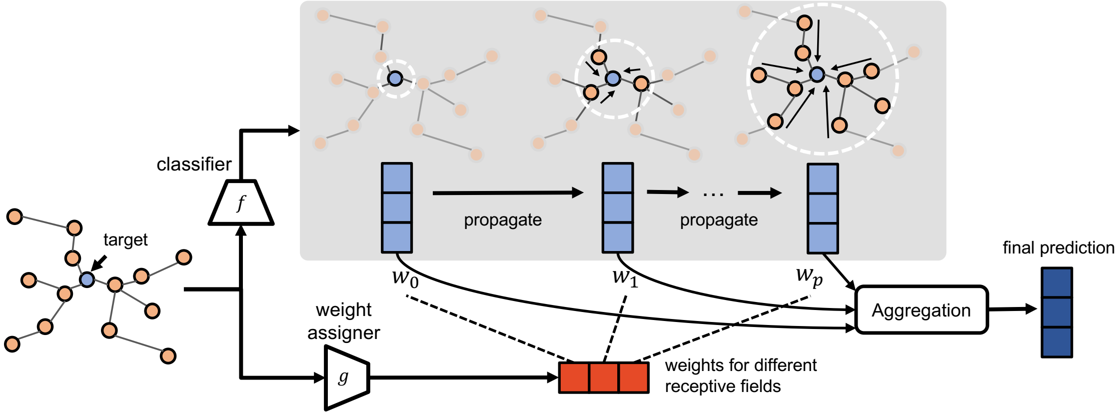
Classifier [1,3]: sum of predictions

$$\mathbf{H}^{(0)} = \mathrm{MLP}(\mathbf{X}, \theta)$$

$$\mathbf{H}^{(j+1)} = \widetilde{\mathbf{A}}\mathbf{H}^{(j)}$$

$$\mathbf{Z} = \mathrm{softmax}\Big(\sum_{j=0}^{p} \big(\underbrace{\mathbf{W}[:,j]\mathbf{1}^{T}}\big) \odot \mathbf{H}^{(j)}\Big)$$

Node-specific weight assignments



- For the **base** classes, **local** classifier.
- For the **novel** classes, **long-range** propagation [1-2]

[1] Gasteiger, Johannes, Aleksandar Bojchevski, and Stephan Günnemann. "Predict then Propagate: Graph Neural Networks meet Personalized PageRank." ICLR. 2019.
[2] Yao, Huaxiu, et al. "Graph few-shot learning via knowledge transfer." AAAI. 2020.
[3] Chien, Eli, et al. "Adaptive Universal Generalized PageRank Graph Neural Network." ICLR. 2021.

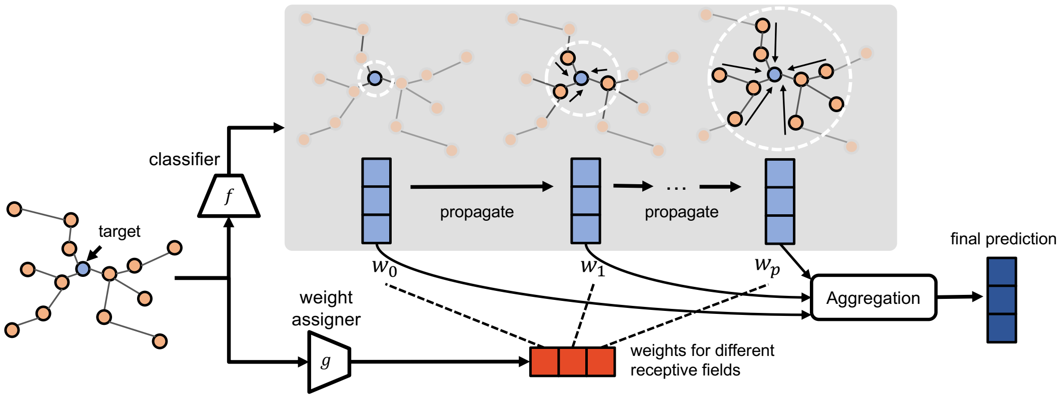
Proposed Model – STAGER (2)

Classifier:

$$\mathbf{H}^{(0)} = \mathrm{MLP}(\mathbf{X}, \theta)$$

$$\mathbf{H}^{(j+1)} = \widetilde{\mathbf{A}}\mathbf{H}^{(j)}$$

$$\mathbf{Z} = \mathrm{softmax}\left(\sum_{j=0}^{p}(\mathbf{W}[:,j]\mathbf{1}^{T}) \odot \mathbf{H}^{(j)}\right)$$



uncertainty (shot)-aware weight assigner:

$$\mathbf{W} = \mathrm{MLP}(\tilde{\mathbf{Z}}, \phi_2)$$

$$\tilde{\mathbf{Z}} = \mathrm{rank}(\mathrm{softmax}(g_1(\mathbf{A}, \mathbf{X}, \phi_1)))$$

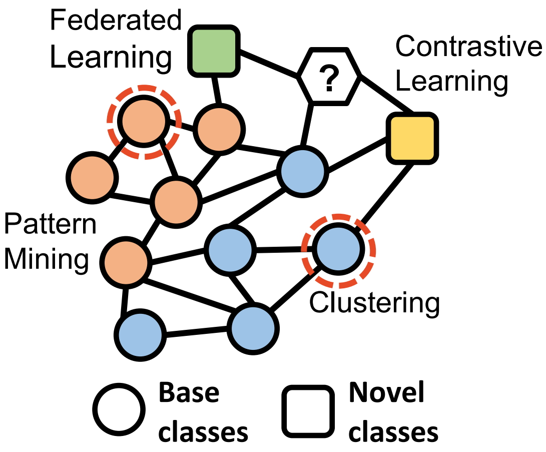
Roadmap

- Motivation ☑
- Proposed Model – STAGER ☑
- Imbalanced Episodic Training
- Experiments
- Conclusion

Meta-learning-based training paradigm

- Existing episodic training [1-5]: **mimic the few-shot scenarios using labeled base samples**
- a few-shot episode: $\mathcal{S}_i = \{v_i, \ldots, v_{N\times K}\}, \mathcal{Q}_i = \{v_i', \ldots, v_{N\times I}'\}$

- N: # of novel classes
- K: shots of the novel classes



$$\phi^* = \arg\min_{\phi} \mathbb{E}_{v_i \in \mathcal{Q}} \mathcal{L}_{cla}(z(\mathcal{G}, \theta^*, \phi, v_i), y_i)$$

Meta-learner ϕ learns from the learning of the learner

$$\theta^* = \arg\min_{\theta} \mathbb{E}_{v_j \in \mathcal{S}} \mathcal{L}_{cla}\big(z\big(\mathcal{G}, \theta, \phi, v_j\big), y_j\big)$$

Learner θ learns to converge

[1] Zhou, Fan, et al. "Meta-gnn: On few-shot node classification in graph meta-learning." CIKM. 2019.
[2] Ding, Kaize, et al. "Graph prototypical networks for few-shot learning on attributed networks." CIKM. 2020.
[3] Huang, Kexin, and Marinka Zitnik. "Graph meta learning via local subgraphs." NeurIPS. 2020.
[4] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In NIPS, 2017.
[5] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In NIPS, 2016.

Imbalanced episodic training

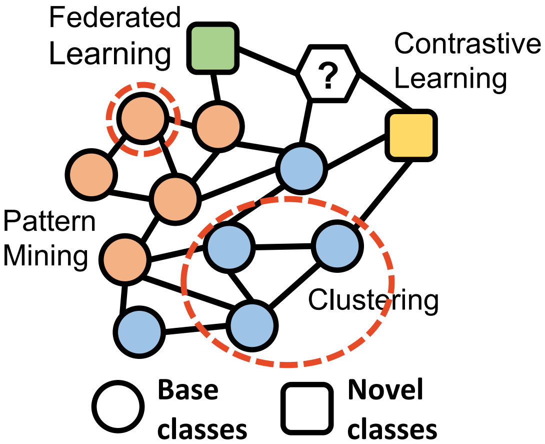
- Goal: Learning to **distinguish few-shot and many-shot samples**
- Key idea: **mimic the few-shot vs. many-shot scenarios**

$$\mathcal{S}_i = \{v_1, \dots, v_{N\times K}\}, \mathcal{Q}_i = \{v'_1, \dots, v'_{N\times I}\}$$

$$\downarrow$$

$$\mathcal{S}_i = \{\underbrace{v_1, \dots, v_{N\times K}}_{\text{pseudo-novel classes}}, \underbrace{v_{N\times K+1}, \dots, v_{N\times K+M\times L}}_{\text{pseudo-base classes}}\},$$

$$\mathcal{Q}_i = \{\overbrace{v'_1, \dots, v'_{N\times I}}, \overbrace{v'_{N\times I+1}, \dots, v'_{(N+M)\times I}}\}$$



- N: # of **pseudo-novel** classes
- K: shots of **pseudo-novel** classes
- M: # of **pseudo-base** classes
- L: shots of **pseudo-base** classes

[1] Zhou, Fan, et al. "Meta-gnn: On few-shot node classification in graph meta-learning." CIKM. 2019.
[2] Ding, Kaize, et al. "Graph prototypical networks for few-shot learning on attributed networks." CIKM. 2020.
[3] Huang, Kexin, and Marinka Zitnik. "Graph meta learning via local subgraphs." NeurIPS. 2020.
[4] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In NIPS, 2017.
[5] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In NIPS, 2016.

Roadmap

- Motivation ☑
- Proposed Model – STAGER ☑
- Imbalanced Episodic Training ☑
- Experiments
- Conclusion

Experimental settings

- Datasets

- Training nodes:
  - Novel: N-way K-shot
  - Base: many-way many-shot
- Test nodes: fixed for every class

Dataset	Nodes	Edges	Features	Labels
Amazon Clothing	24919	91680	9034	77
Amazon Electronics	42318	43556	8669	167
Aminer	40672	288270	7202	137
Cora-Full	18800	62685	8710	56

- Baselines:
  - Classic neural node classifier: APPNP, GPRGNN
  - Few-shot node classifier: Meta-GNN, GPN, G-Meta
  - Imbalanced node classifier: GraphSMOTE

- Metric: accuracy in 10 runs

Main results on the effectiveness

			standard		few-shot			imbalanced		
Dataset	**Setting**	**Class**	**APPNP**	**GPRGNN**	**MetaGNN**	**GPN**	**G-META**	**G-SMOTE**	**STAGER**	**STAGER-I**
Amazon Clothing	5w1s	Base	67.4±1.6	64.7±0.5	64.0±0.5	46.1±3.3	48.7±2.7	67.4±1.6	**69.3±1.1**	67.3±0.4
		Novel	31.4±0.9	31.5±5.4	28.3±0.4	36.1±5.1	39.2±2.9	31.4±0.9	32.4±2.0	**41.3±1.0**
		All	48.5±1.0	47.3±3.0	45.4±0.4	40.9±2.9	43.7±2.2	48.5±1.0	50.0±1.0	**53.7±0.5**
	5w3s	Base	70.5±0.9	69.7±1.0	66.1±1.2	62.9±1.9	63.3±1.7	69.4±0.7	**72.3±1.5**	68.4±1.0
		Novel	48.6±2.5	50.1±3.8	40.4±0.8	46.1±6.7	47.6±6.3	45.6±2.3	53.9±2.1	**66.0±2.4**
		All	59.1±1.2	59.4±2.1	52.6±1.0	54.1±3.4	55.1±3.5	57.0±1.4	62.7±1.1	**67.2±1.1**
	10w1s	Base	73.3±0.3	70.7±1.2	67.6±0.5	42.7±2.4	48.2±2.1	73.3±0.3	**76.7±1.5**	66.7±0.5
		Novel	45.2±0.6	37.7±3.1	41.5±0.5	39.7±5.7	39.9±4.9	45.2±0.6	43.1±2.7	**59.6±0.6**
		All	58.6±0.3	53.5±1.6	54.0±0.4	40.9±3.5	43.9±2.0	58.6±0.3	59.0±1.6	**63.0±0.5**
	10w3s	Base	69.2±0.6	67.5±1.1	65.6±1.2	59.5±2.7	57.4±1.9	68.1±0.7	**70.9±0.7**	69.3±0.4
		Novel	61.4±0.4	58.0±1.5	53.6±0.2	49.6±7.1	54.1±2.8	53.6±3.8	61.8±1.2	**64.6±0.7**
		All	65.2±0.4	62.5±1.3	59.2±0.2	54.3±3.4	55.6±1.6	60.5±1.9	66.2±0.8	**66.8±0.5**
Amazon Elec.	5w1s	Base	60.1±1.8	58.4±0.9	59.7±0.3	19.1±2.1	22.5±3.1	60.1±1.8	**65.8±2.1**	63.9±1.0
		Novel	7.8±0.8	5.1±1.1	6.4±0.3	16.6±5.4	15.3±6.7	7.8±0.8	8.0±0.7	**19.7±1.6**
		All	27.2±0.4	24.8±0.4	26.2±0.2	17.5±3.8	18.0±5.0	27.2±0.4	29.4±1.4	**36.1±1.1**
	5w3s	Base	64.2±1.8	55.1±0.9	63.0±0.7	43.7±1.6	43.6±2.4	63.0±1.4	**69.1±1.6**	69.0±2.9
		Novel	21.6±1.5	13.3±2.0	23.1±0.2	32.7±4.8	28.1±5.6	12.0±4.0	29.8±2.7	**40.7±2.2**
		All	37.4±1.5	28.8±1.4	37.9±0.3	36.8±3.4	33.9±3.5	30.9±2.5	44.3±1.8	**51.2±2.3**
	10w1s	Base	64.4±1.2	59.7±1.3	53.1±1.6	18.5±1.4	20.8±2.2	64.4±1.2	**69.0±0.9**	61.3±0.8
		Novel	8.0±1.3	5.7±1.1	4.9±0.1	15.3±3.7	15.0±3.7	8.0±1.3	11.3±1.5	**15.4±0.3**
		All	34.4±1.0	31.0±1.1	27.7±0.2	16.8±2.3	17.7±2.0	34.4±1.0	**38.3±1.2**	36.9±0.4
	10w3s	Base	58.6±0.4	55.2±0.9	48.8±0.7	43.8±1.7	46.3±1.6	62.9±0.7	**72.3±1.1**	66.5±0.1
		Novel	22.4±1.1	14.8±1.0	16.5±0.2	27.5±2.9	26.2±3.2	13.8±0.3	20.3±2.4	**38.1±2.3**
		All	39.4±0.5	33.7±0.7	31.6±0.4	35.1±1.4	35.6±1.8	36.8±0.2	44.7±1.5	**51.4±1.1**

More results in the paper

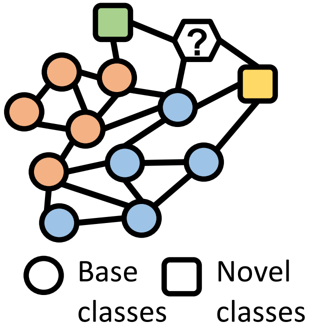
- Compared with **standard node classifiers** (APPNP, GPRGNN), STAGER is competitive.
- **Few-shot node classifiers** (MetaGNN, GPN, G-Meta) cannot work well in the generalized setting.
- **Imbalanced node classifier** G-SMOTE cannot fully use the labeled base classes.
- **Imbalanced episodic training** facilitates **a better tradeoff** between the novel and base classes.

Roadmap

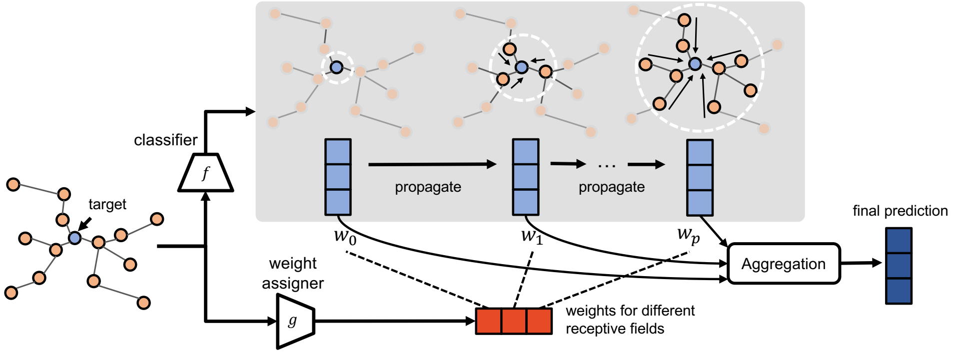
- Motivation ☑
- Proposed Model – STAGER ☑
- Imbalanced Episodic Training ☑
- Experiments ☑
- Conclusion

Conclusion

- **Problem**: generalized few-shot node classification
  - Classify nodes into the *joint set of base & novel classes*



- **Solution**: STAGER + imbalanced episodic training
  - *Uncertainty-aware* node classifier
  - *Few-shot vs. many-shot* scenarios



- **Experiments**
  - *The consistent advantage* over all classes

Title: Generalized Few-Shot Node Classification
Authors: Zhe Xu, Kaize Ding, Yu-Xiong Wang, Huan Liu, Hanghang Tong
Email: zhexu3@Illinois.edu

Thank you!

Q & A