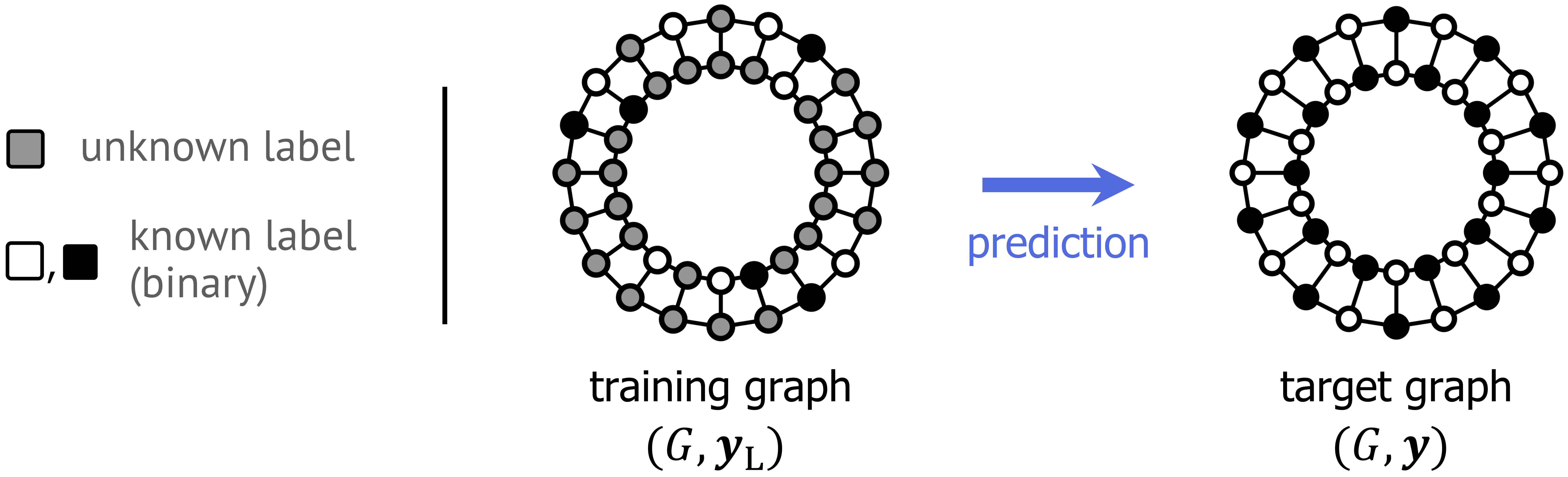
Diffusion Probabilistic Models for Structured Node Classification

Hyosoon Jang[1], Seonghyun Park[1], Sangwoo Mo[2], Sungsoo Ahn[1]

[1]POSTECH [2] University of Michigan

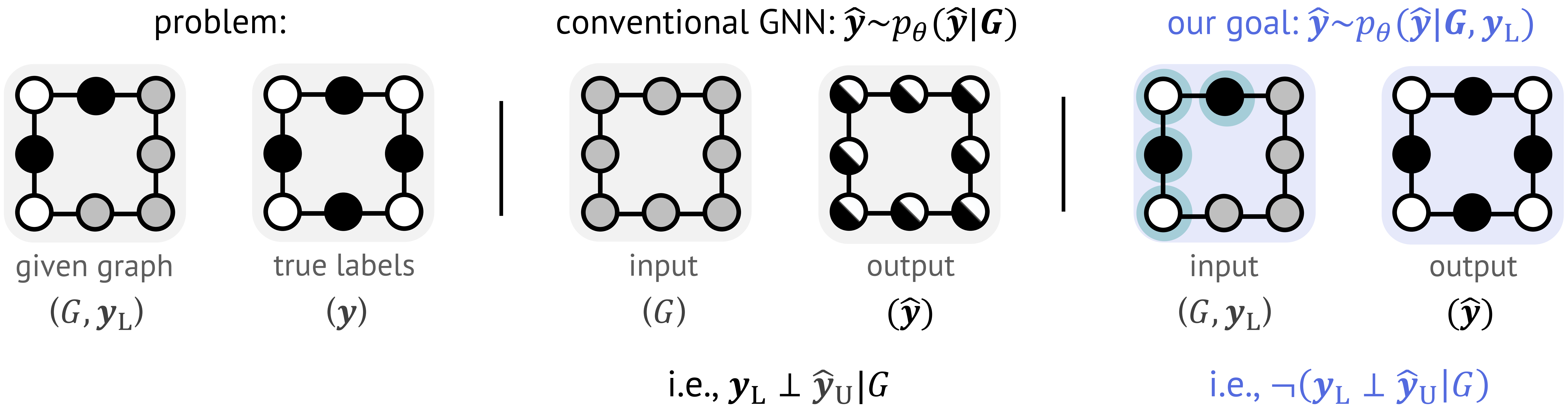
Problem of interest

- Node classification: given a graph G, predict node-wise labels $\boldsymbol{y}$.
    - It typically considers a partially labeled graph that includes known node labels $\boldsymbol{y}_{\mathrm{L}}$ for predicting unknown node labels $\boldsymbol{y}_{\mathrm{U}}$.
    - How to incorporate node-wise label dependencies with the known labels?



How do known labels benefit?

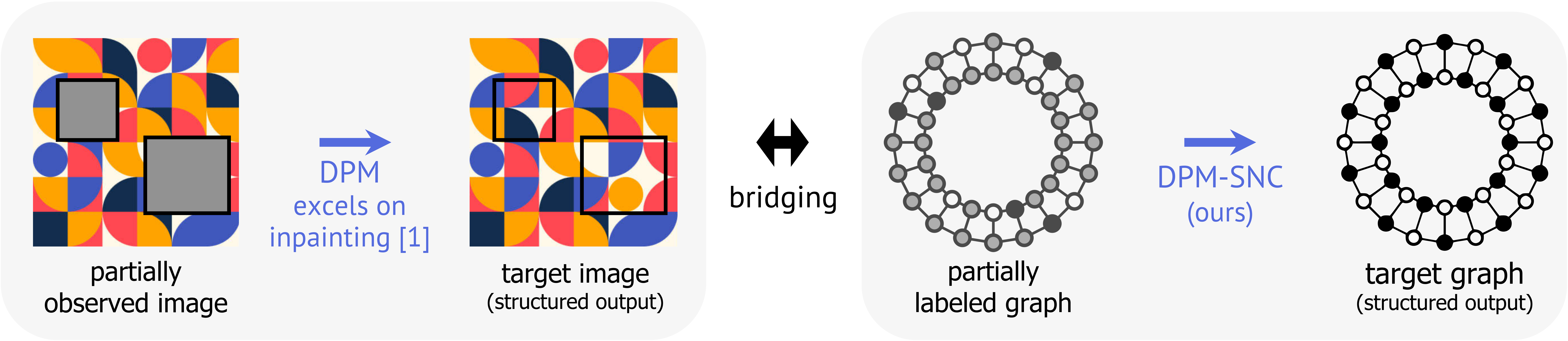
- The outputs of Conventional Graph Neural Network (GNN) ignore the known labels.
- Our goal is structured node classification, which incorporates dependencies with the known labels.



unknown labels $\boldsymbol{y}_\mathrm{U}$, known labels $\boldsymbol{y}_\mathrm{L}$, symbol for independent $\perp$

How to incorporate known labels?

- Key observation: incorporating dependencies with the known label on graph is just like inpainting on image!
  - We investigate diffusion probabilistic models for structured node classification (DPM-SNC)

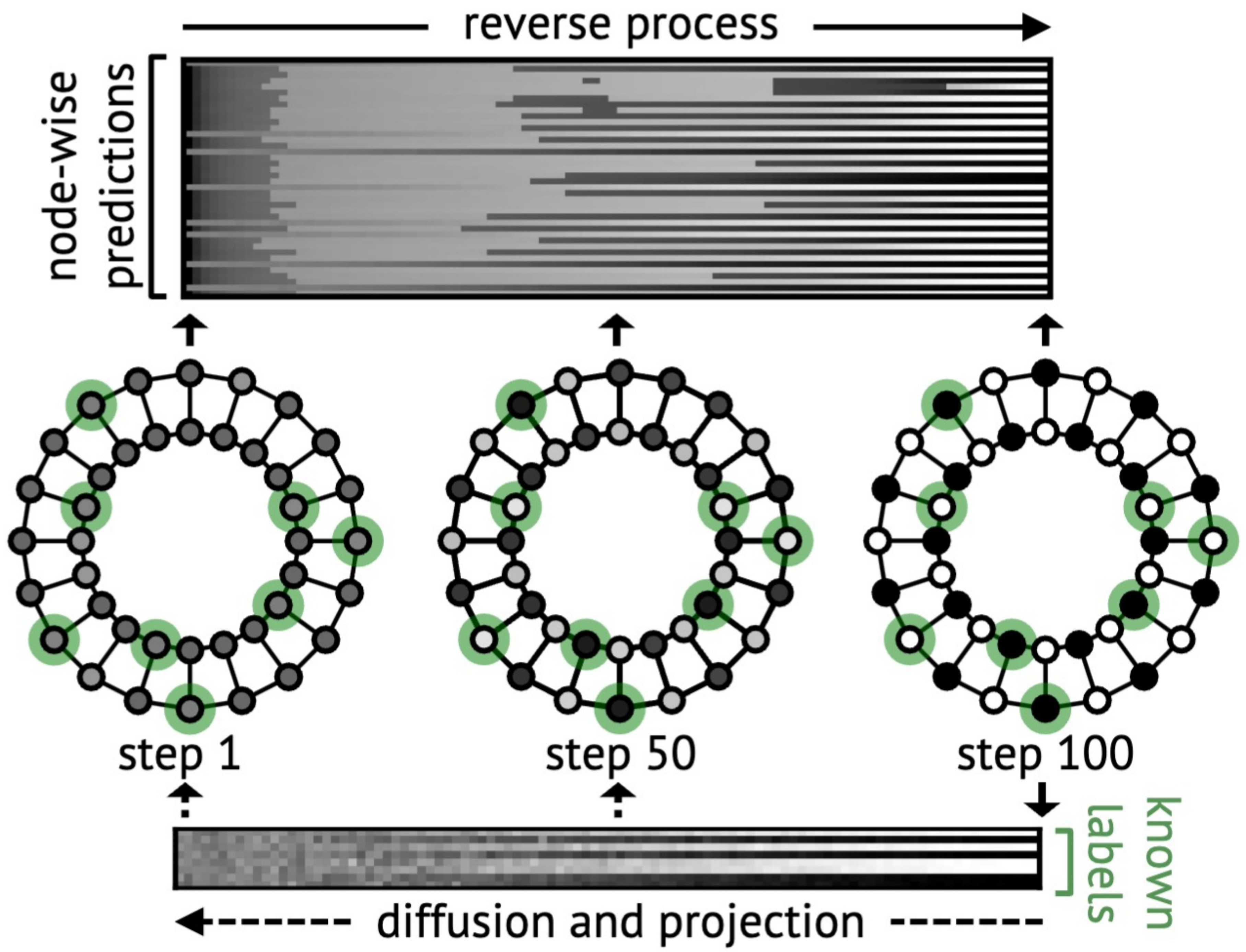


[1] Improving Diffusion Models for Inverse Problems using Manifold Constraints, Chung et al., NeurIPS 2022

DPM-SNC framework

- At a high-level, DPM-SNC consists of:

  - diffusion: injects noise into the node labels
  - reverse diffusion: denoises the noisy labels

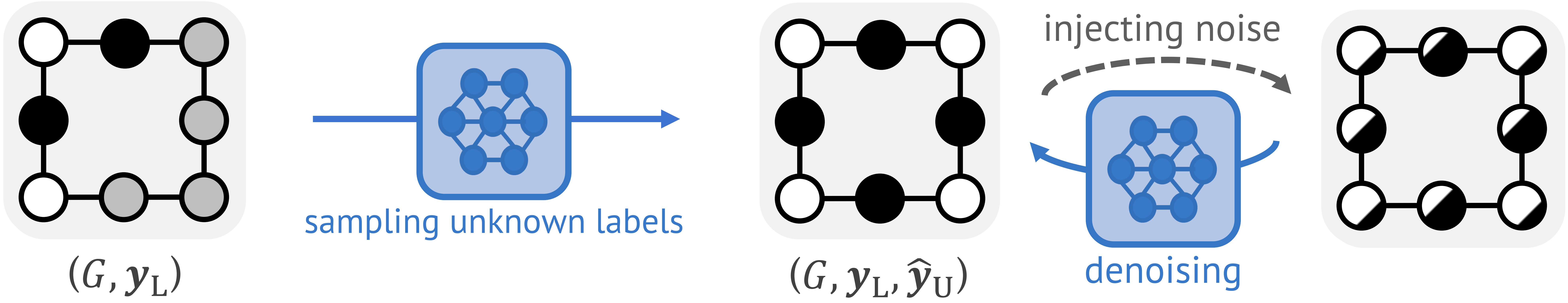


- incorporating known labels via projection [1]

[1] Improving Diffusion Models for Inverse Problems using Manifold Constraints, Chung et al., NeurIPS 2022

Semi-supervised training

- The training of DPM on partially labeled graph is non-trivial.
  - We newly derive a new semi-supervised training algorithm of DPM which alternates two processes:



complete the graph
- conditioned on known labels

update the DPM
- denoising-based training (trivial)

Theoretical analysis

- We also theoretically analyze the expressive power of DPM-SNC.
  - compare with conventional GNNs based on Weisfeiler-Lehman (WL) algorithms [1]

$\underbrace{\text{Conventional GNN* = 1-WL algorithm}}_{[1]} < \underbrace{\text{AGG-WL algorithm = DPM-SNC*}}_{\text{new algorithmic analog}}$

* describes the most expressive condition

[1] How Powerful are Graph Neural Networks?, Xu et al., ICLR 2019

Experiments

- DPM-SNC outperforms on transductive node classification benchmarks.

Table 2: The transductive node classification performance. N-Acc. and Sub-Acc. denote the node-level and subgraph-level accuracy, respectively. **Bold** numbers indicate the best performance among the structured prediction methods using the same GNN.

	Pubmed		Cora		Citeseer		Photo		Computer	
Method	N-Acc.	Sub-Acc.	N-Acc.	Sub-Acc.	N-Acc.	Sub-Acc.	N-Acc.	Sub-Acc.	N-Acc.	Sub-Acc.
LP [42]	69.1±0.0	45.7±0.0	68.1±0.0	46.9±0.0	46.1±0.0	29.8±0.0	81.0±2.0	37.2±1.7	69.9±2.9	15.1±1.1
PTA [43]	80.1±0.2	55.2±0.4	82.9±0.4	62.6±0.8	71.3±0.4	51.4±0.7	91.1±1.5	51.0±1.5	81.6±1.7	26.3±1.0
GCN [3]	79.7±0.3	55.8±0.6	81.4±0.8	59.3±1.1	70.9±0.8	49.8±0.6	91.0±1.2	52.0±1.0	82.4±1.5	27.0±1.5
+LPA [12]	79.6±0.6	53.5±0.9	81.7±0.7	60.3±1.5	71.0±0.6	50.2±1.0	91.3±1.2	52.9±2.0	83.7±1.4	28.5±2.4
+GMNN [7]	82.6±1.0	58.1±1.4	82.6±0.9	61.8±1.3	72.8±0.7	52.0±0.8	91.2±1.2	54.3±1.4	82.0±1.0	28.0±1.6
+G^{3}NN [8]	80.9±0.7	56.9±1.1	82.5±0.4	62.3±0.8	73.9±0.7	53.1±1.0	90.7±1.1	53.0±2.0	82.1±1.2	28.1±2.1
+CLGNN [13]	81.7±0.5	57.8±0.7	81.9±0.5	61.8±0.8	72.0±0.7	51.6±0.9	91.1±1.0	53.4±1.8	83.3±1.2	28.5±1.4
+DPM-SNC (ours)	**83.0**±0.9	**59.2**±1.2	**83.2**±0.5	**63.1**±0.9	**74.4**±0.5	**53.6**±0.6	**92.2**±0.8	**55.3**±2.1	**84.1**±1.3	**29.7**±1.8
GAT [21]	79.1±0.5	55.8±0.5	81.5±0.6	61.3±0.9	71.0±0.8	50.8±1.0	90.8±1.0	50.8±1.9	83.1±1.6	27.8±2.2
+LPA [12]	78.7±1.1	56.0±1.2	81.5±0.9	60.7±0.8	71.3±0.9	50.1±0.9	91.3±0.8	52.7±2.1	**84.4**±1.0	29.4±2.6
+GMNN [7]	79.6±0.8	57.0±0.7	82.3±0.7	62.2±0.8	71.7±0.9	51.4±0.9	91.4±1.0	53.1±1.6	83.3±2.0	29.1±1.8
+G^{3}NN [8]	77.9±0.4	55.9±0.5	82.7±1.3	62.7±1.3	74.0±0.8	53.7±0.5	91.5±0.9	52.6±2.2	83.1±1.7	28.8±2.4
+CLGNN [13]	80.0±0.6	57.5±1.2	81.8±0.6	61.5±0.9	72.1±0.8	52.1±0.8	90.6±0.8	51.9±1.8	82.6±1.2	28.4±1.8
+DPM-SNC (ours)	**81.7**±0.8	**59.0**±1.1	**83.8**±0.7	**63.8**±0.7	**74.3**±0.7	**54.0**±0.9	**92.0**±0.8	**54.0**±2.4	84.2±1.2	**30.0**±2.0
GCNII [44]	82.0±0.8	57.2±1.1	84.0±0.6	63.4±0.8	72.9±0.5	52.1±0.7	91.2±1.2	53.2±1.5	82.5±1.4	26.6±1.3
+DPM-SNC (ours)	**83.8**±0.7	**61.6**±0.9	**85.3**±0.6	**65.8**±0.7	**74.1**±0.5	**54.1**±0.9	**92.8**±1.1	**54.2**±1.2	**84.4**±1.8	**29.2**±1.1

Table 3: The transductive node classification accuracy on heterophilic graphs. **Bold** numbers indicate the best score.

	Empire	Rating
H_2GCN [45]	60.11±0.52	36.47±0.23
CPGNN [46]	63.96±0.62	39.79±0.77
GPR-GNN [47]	64.85±0.27	44.88±0.34
FSGNN [48]	79.92±0.56	52.74±0.83
GloGNN [49]	59.63±0.69	36.89±0.14
FAGCN [50]	65.22±0.56	44.12±0.30
GBK-GNN [51]	74.57±0.47	45.98±0.71
JacobiConv [52]	71.14±0.42	43.55±0.48
GCN [3]	73.69±0.74	48.70±0.63
SAGE [4]	85.74±0.67	53.63±0.39
GAT [21]	80.87±0.30	49.09±0.63
GAT-sep [24]	88.75±0.41	52.70±0.62
GT [53]	86.51±0.73	51.17±0.66
GT-sep [24]	87.32±0.39	52.18±0.80
DPM-SNC (ours)	**89.52**±0.46	**54.66**±0.39

Experiments

- DPM-SNC outperforms on inductive node classification benchmarks.

Table 4: The inductive node classification performance. N-Acc., G-Acc., and F1 denote the node-level accuracy, graph-level accuracy, and micro-F1 score, respectively. **Bold** numbers indicate the best performance among the structured prediction methods using the same GNN.

	Pubmed		Cora		Citeseer		PPI
Method	N-Acc.	G-Acc.	N-Acc.	G-Acc.	N-Acc.	G-Acc.	F1
GCN [3]	$80.25_{\pm 0.42}$	$54.58_{\pm 0.51}$	$83.36_{\pm 0.43}$	$59.67_{\pm 0.51}$	$76.37_{\pm 0.35}$	$49.84_{\pm 0.47}$	$99.15_{\pm 0.03}$
+G^{3}NN [8]	$80.32_{\pm 0.30}$	$53.93_{\pm 0.71}$	$83.60_{\pm 0.25}$	$59.78_{\pm 0.47}$	$76.34_{\pm 0.37}$	$50.76_{\pm 0.47}$	$99.33_{\pm 0.02}$
+CLGNN [13]	$80.22_{\pm 0.45}$	$53.98_{\pm 0.54}$	$83.45_{\pm 0.34}$	$60.24_{\pm 0.38}$	$75.71_{\pm 0.40}$	$50.51_{\pm 0.38}$	$99.22_{\pm 0.04}$
+SPN [9]	$\mathbf{80.78}_{\pm 0.34}$	$54.91_{\pm 0.40}$	$83.85_{\pm 0.60}$	$60.35_{\pm 0.57}$	$76.25_{\pm 0.48}$	$51.02_{\pm 1.06}$	$99.35_{\pm 0.02}$
+DPM-SNC (ours)	$80.58_{\pm 0.41}$	$\mathbf{55.16}_{\pm 0.43}$	$\mathbf{84.09}_{\pm 0.27}$	$\mathbf{60.88}_{\pm 0.36}$	$\mathbf{77.01}_{\pm 0.49}$	$\mathbf{51.44}_{\pm 0.56}$	$\mathbf{99.46}_{\pm 0.02}$
GAT [21]	$80.10_{\pm 0.45}$	$54.38_{\pm 0.54}$	$79.71_{\pm 1.41}$	$56.66_{\pm 1.40}$	$74.91_{\pm 0.22}$	$49.87_{\pm 0.44}$	$99.54_{\pm 0.01}$
+G^{3}NN [8]	$79.88_{\pm 0.62}$	$54.66_{\pm 0.29}$	$81.19_{\pm 0.45}$	$58.68_{\pm 0.38}$	$75.45_{\pm 0.26}$	$50.86_{\pm 0.46}$	$99.56_{\pm 0.01}$
+CLGNN [13]	$80.23_{\pm 0.40}$	$54.51_{\pm 0.36}$	$81.38_{\pm 0.55}$	$58.81_{\pm 0.61}$	$75.45_{\pm 0.36}$	$50.66_{\pm 0.45}$	$99.55_{\pm 0.01}$
+SPN [9]	$79.95_{\pm 0.34}$	$\mathbf{54.82}_{\pm 0.33}$	$81.61_{\pm 0.31}$	$59.17_{\pm 0.31}$	$75.41_{\pm 0.35}$	$51.04_{\pm 0.53}$	$99.46_{\pm 0.02}$
+DPM-SNC (ours)	$\mathbf{80.26}_{\pm 0.37}$	$54.26_{\pm 0.47}$	$\mathbf{81.79}_{\pm 0.46}$	$\mathbf{59.55}_{\pm 0.49}$	$\mathbf{76.46}_{\pm 0.60}$	$\mathbf{52.05}_{\pm 0.71}$	$\mathbf{99.63}_{\pm 0.01}$

Experiments

- DPM-SNC also shows competitive results in graph algorithmic reasoning tasks.

Table 5: Performance on graph algorithmic reasoning tasks. **Bold** numbers indicate the best performance. Same-MSE and Large-MSE indicate the performance on ten, and 15 nodes, respectively.

	Edge copy		Connected components		Shortest path	
Method	Same-MSE	Large-MSE	Same-MSE	Large-MSE	Same-MSE	Large-MSE
Feedforward	0.3016	0.3124	0.1796	0.3460	0.1233	1.4089
Recurrent [54]	0.3015	0.3113	0.1794	0.2766	0.1259	0.1083
Programmatic [55]	0.3053	0.4409	0.2338	3.1381	0.1375	0.1290
Iterative feedforward [56]	0.6163	0.6498	0.4908	1.2064	0.4588	0.7688
IREM [26]	0.0019	**0.0019**	0.1424	0.2171	0.0274	0.0464
DPM-SNC (ours)	**0.0011**	0.0038	**0.0724**	**0.1884**	**0.0138**	**0.0286**