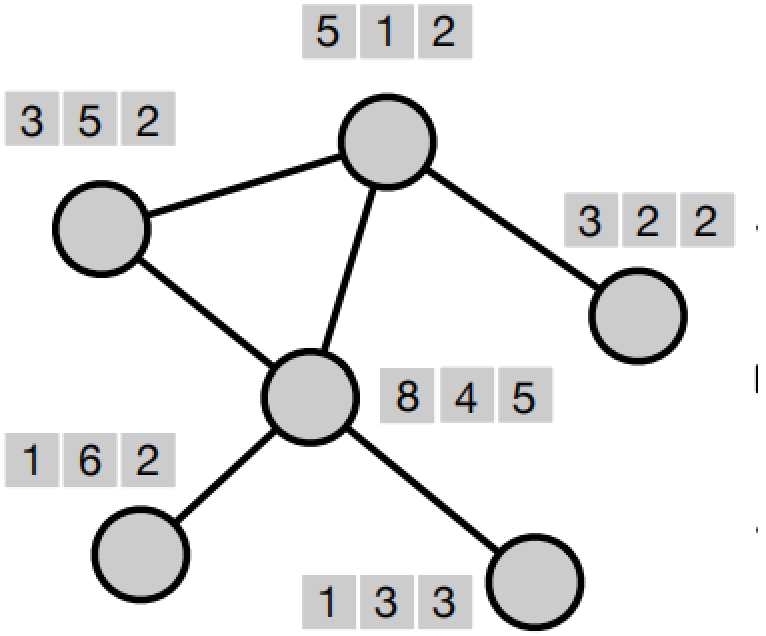
# Graph Random Neural Networks for Semi-Supervised Learning on Graphs

Wenzheng Feng, Jie Zhang, Yuxiao Dong  Yu Han, Huanbo Luan, Qian Xu,Qiang Yang, Evgeny Kharlamov, Jie Tang

NIPS 2020

# Motivation

The assumption of homophily :adjacent nodes tend to have similar features and labels




Graph Data

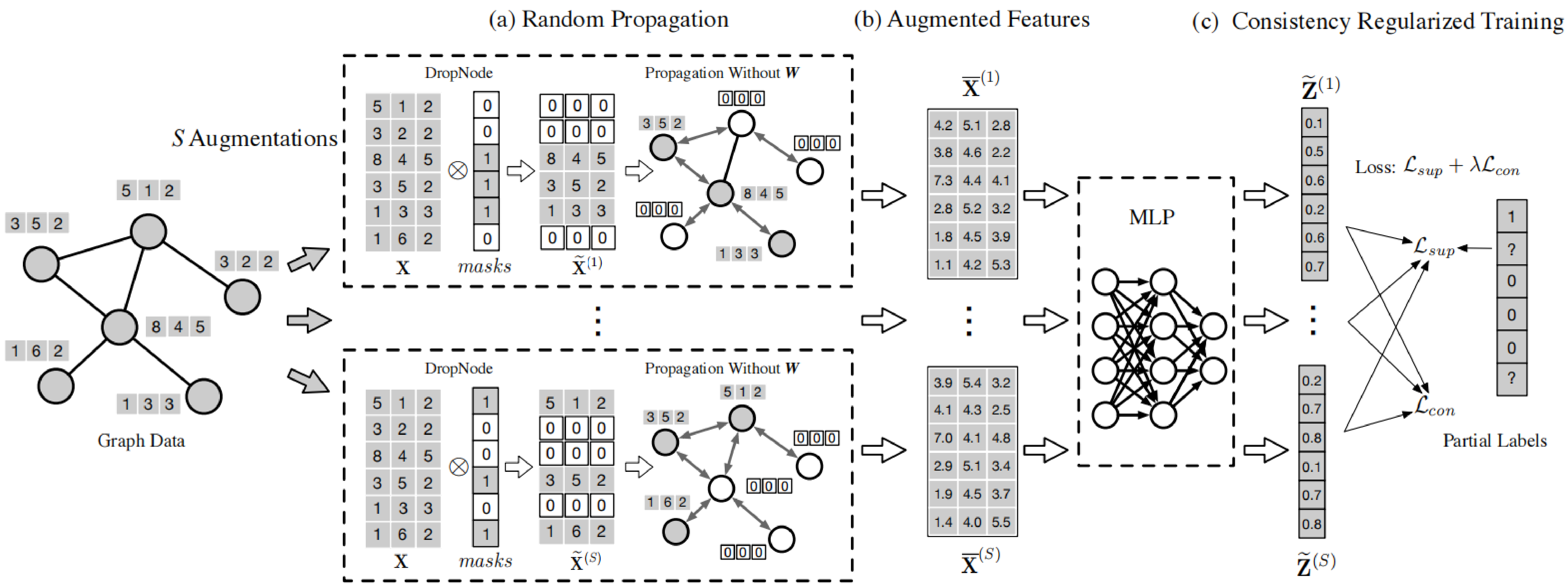
stacking many GNN layers tends to make nodes'features indistinguishable

over-smoothing

Naturally, the deterministic propagation makes each node highly dependentbwith its (multi-hop) neighborhoods, leaving the nodes to be easily misguided by potential data noise and susceptible to adversarial perturbations.

not robust to graph attacks

# Graph Random Neural Network



Random remove node vector : $\widetilde{\mathbf{X}}_i = \epsilon_i \cdot \mathbf{X}_i$ $\quad \epsilon_i \sim Bernoulli(1-\delta)$

Data Augmentation : $\overline{\mathbf{X}} = \overline{\mathbf{A}}\widetilde{\mathbf{X}}$ $\quad \overline{\mathbf{A}} = \sum_{k=0}^{K} \frac{1}{K+1}\hat{\mathbf{A}}^k$

Train model : $\widetilde{\mathbf{Z}}^{(s)} = f_{mlp}(\overline{\mathbf{X}}^{(s)}, \Theta),$

## Supervised Loss:

$$\mathcal{L}_{sup} = -\frac{1}{S}\sum_{s=1}^{S}\sum_{i=0}^{m-1}\mathbf{Y}_i^{\top}\log\widetilde{\mathbf{Z}}_i^{(s)}.$$

## Consistency Regularization Loss:

$$\overline{\mathbf{Z}}_{ij}^{'} = \overline{\mathbf{Z}}_{ij}^{\frac{1}{T}} \Big/ \sum_{c=0}^{C-1}\overline{\mathbf{Z}}_{ic}^{\frac{1}{T}}, (0 \leq j \leq C-1),$$

$$\mathcal{L}_{con} = \frac{1}{S}\sum_{s=1}^{S}\sum_{i=0}^{n-1}\|\overline{\mathbf{Z}}_i^{'} - \widetilde{\mathbf{Z}}_i^{(s)}\|_2^2.$$

**Algorithm 1** GRAND

**Input:**

Adjacency matrix $\hat{\mathbf{A}}$, feature matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$, times of augmentations in each epoch $S$, DropNode/dropout probability $\delta$, learning rate $\eta$, an MLP model: $f_{mlp}(\mathbf{X}, \Theta)$.

**Output:**

Prediction $\mathbf{Z}$.

**while** not convergence **do**
  **for** $s = 1 : S$ **do**
    Pertube the input: $\widetilde{\mathbf{X}}^{(s)} \sim \text{DropNode}(\mathbf{X}, \delta)$.
    Perform propagation: $\overline{\mathbf{X}}^{(s)} = \frac{1}{K+1} \sum_{k=0}^{K} \hat{\mathbf{A}}^k \widetilde{\mathbf{X}}^{(s)}$.
    Predict class distribution using MLP: $\widetilde{\mathbf{Z}}^{(s)} = f_{mlp}(\overline{\mathbf{X}}^{(s)}, \Theta)$
  **end for**
  Compute supervised classification loss $\mathcal{L}_{sup}$ via Eq. 1 and consistency regularization loss via Eq. 3.
  Update the parameters $\Theta$ by gradients descending: $\Theta = \Theta - \eta \nabla_\Theta (\mathcal{L}_{sup} + \lambda \mathcal{L}_{con})$
**end while**
Output prediction $\mathbf{Z}$ via: $\mathbf{Z} = f_{mlp}(\frac{1}{K+1} \sum_{k=0}^{K} \hat{\mathbf{A}}^k \mathbf{X}, \Theta)$.

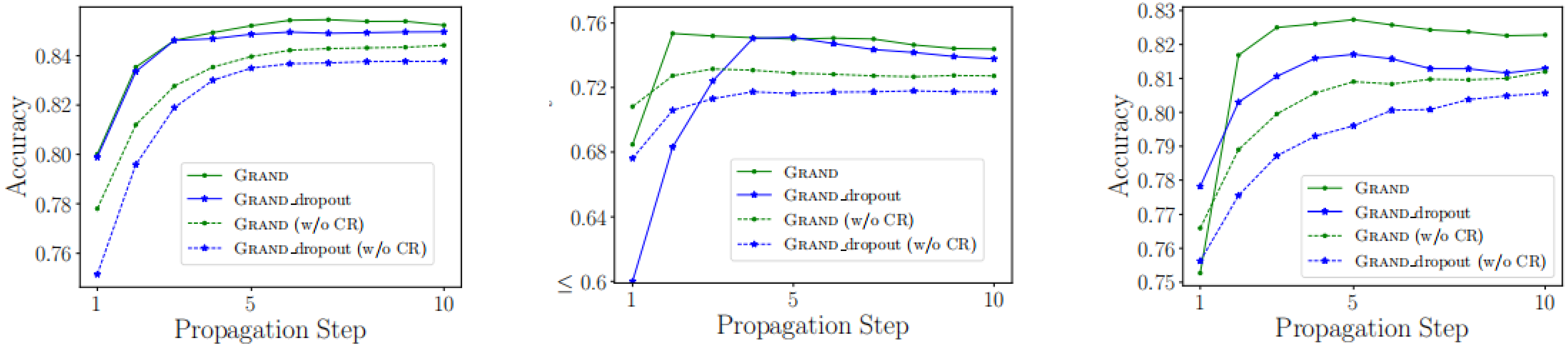
# Experiments

| Method | Cora | Citeseer | Pubmed |
|---|---|---|---|
| GCN [20] | 81.5 | 70.3 | 79.0 |
| GAT [35] | 83.0±0.7 | 72.5±0.7 | 79.0±0.3 |
| APPNP [21] | 83.8±0.3 | 71.6± 0.5 | 79.7 ± 0.3 |
| Graph U-Net [12] | 84.4±0.6 | 73.2±0.5 | 79.6±0.2 |
| SGC [39] | 81.0 ±0.0 | 71.9 ± 0.1 | 78.9 ± 0.0 |
| MixHop [1] | 81.9± 0.4 | 71.4±0.8 | 80.8±0.6 |
| GMNN [31] | 83.7 | 72.9 | 81.8 |
| GraphNAS [13] | 84.2±1.0 | 73.1±0.9 | 79.6±0.4 |
| GraphSAGE [17] | 78.9±0.8 | 67.4±0.7 | 77.8±0.6 |
| FastGCN [7] | 81.4±0.5 | 68.8±0.9 | 77.6±0.5 |
| VBAT [9] | 83.6±0.5 | 74.0±0.6 | 79.9±0.4 |
| $G^3$NN [25] | 82.5±0.2 | 74.4±0.3 | 77.9 ±0.4 |
| GraphMix [36] | 83.9±0.6 | 74.5±0.6 | 81.0±0.6 |
| DropEdge [32] | 82.8 | 72.3 | 79.6 |
| GRAND_dropout | 84.9±0.4 | 75.0±0.3 | 81.7±1.0 |
| GRAND_DropEdge | 84.5±0.3 | 74.4±0.4 | 80.9±0.9 |
| GRAND_GCN | 84.5±0.3 | 74.2±0.3 | 80.0±0.3 |
| GRAND_GAT | 84.3±0.4 | 73.2± 0.4 | 79.2±0.6 |
| GRAND | **85.4±0.4** | **75.4±0.4** | **82.7±0.6** |
| w/o CR | 84.4±0.5 | 73.1±0.6 | 80.9±0.8 |
| w/o mDN | 84.7±0.4 | 74.8±0.4 | 81.0±1.1 |
| w/o sharpening | 84.6±0.4 | 72.2±0.6 | 81.6±0.8 |
| w/o CR & DN | 83.2±0.5 | 70.3±0.6 | 78.5±1.4 |

Table 1: Overall classification accuracy (%).

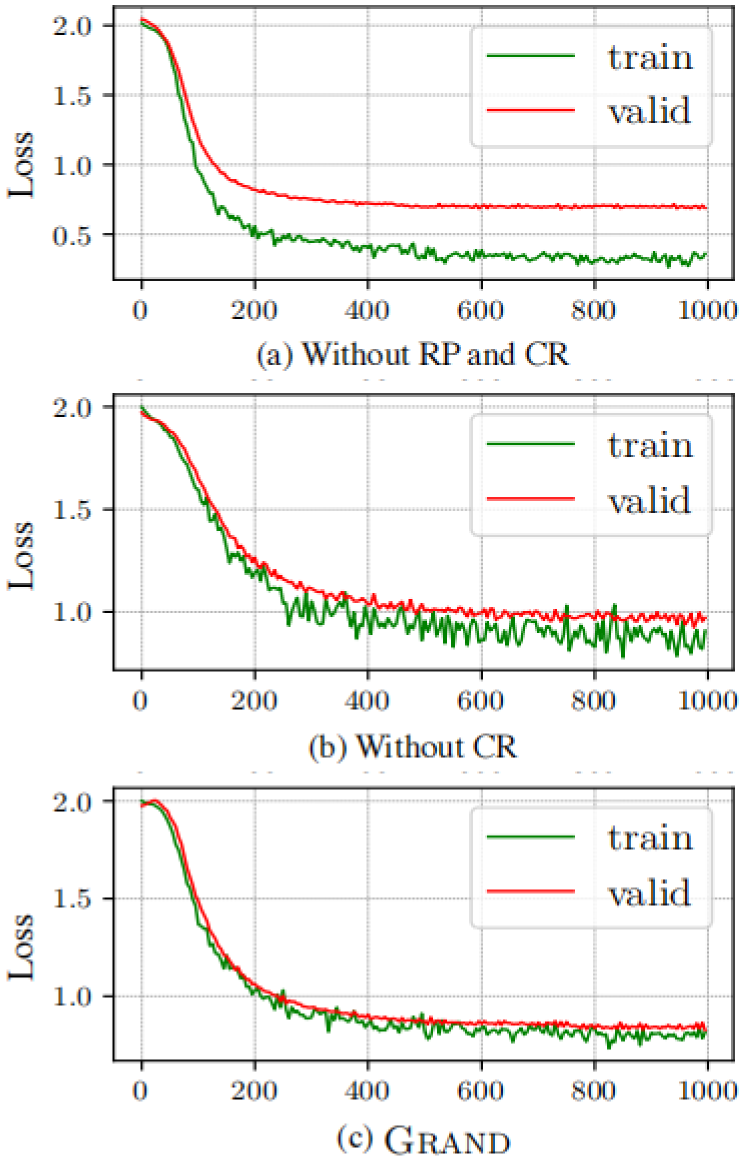
Table 6: Classification Accuracy under different label rates (%).

| Dataset | Cora | | | Citeseer | | | Pubmed | | |
|---|---|---|---|---|---|---|---|---|---|
| Label Rate | 1% | 3% | 5% | 1% | 3% | 5% | 0.1% | 0.3% | 0.5% |
| GCN | 62.8±5.3 | 76.1±1.9 | 79.6±2.1 | 63.4±2.9 | 70.6±1.7 | 72.2±1.1 | 71.5±2.1 | 77.5±1.8 | 80.8±1.5 |
| GAT | 64.3±5.8 | 77.2±2.4 | 80.8±2.1 | 64.4±2.9 | 70.4±1.9 | 72.0±1.3 | 72.0±2.1 | 77.6±1.6 | 80.6±1.2 |
| GRAND | **69.1±4.0** | **79.5±2.2** | **83.0±1.6** | **65.3±3.3** | **72.3±1.8** | **73.8±0.9** | **74.7±3.4** | **81.4±2.1** | **83.8±1.3** |




(a) Cora (b) Citeseer (c) Pubmed

Figure 7: GRAND vs. GRAND_dropout.

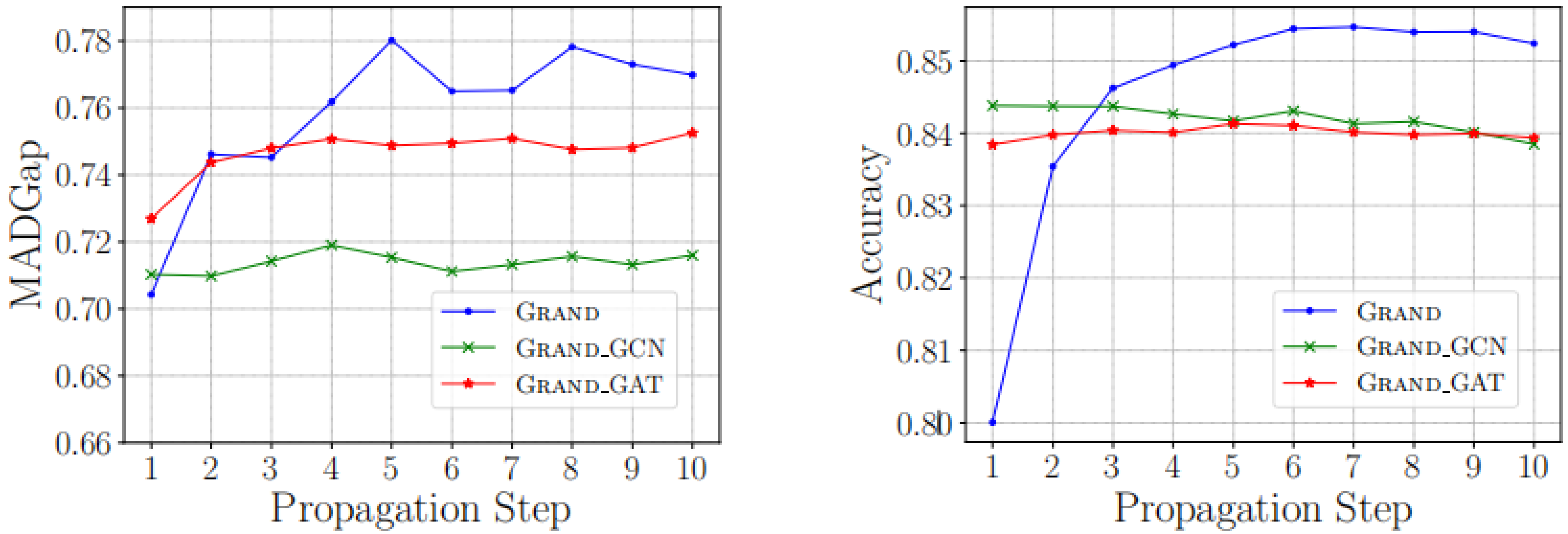



(a) Without RP and CR

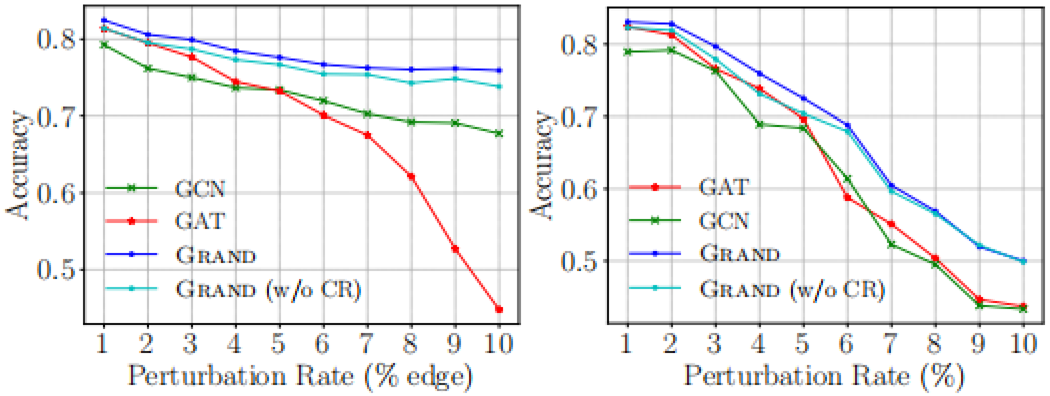
(b) Without CR

(c) GRAND

Figure 2: Generalization on Cora ($x$: epoch).

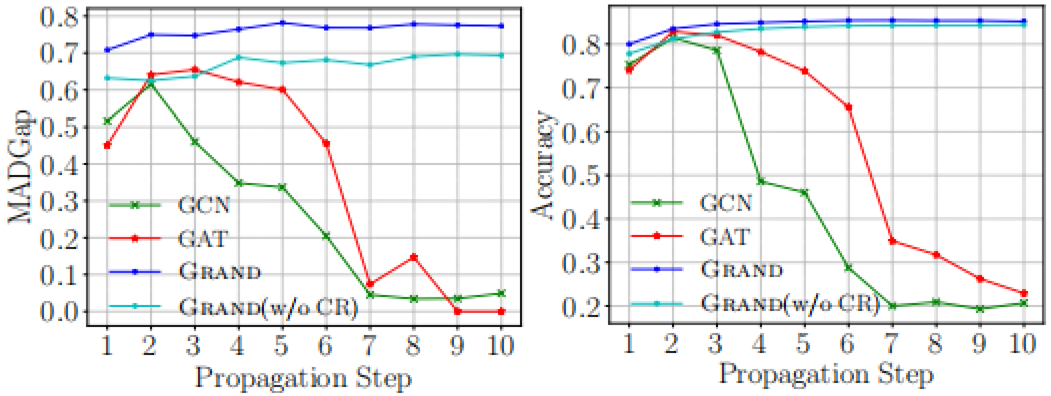



(a) MADGap (b) Classification Results




(a) Random Attack (b) Metattack

Figure 3: Robustness Analysis on Cora.




(a) MADGap (b) Classification Result

Figure 4: Over-Smoothing on Cora