Elastic Graph Neural Networks

Xiaorui Liu
Joint work with Wei Jin (co-first author),
Yao Ma, Yaxin Li, Hua Liu, Yiqi Wang,
Ming Yan & Jiliang Tang

Michigan State University

ICML 2021, July 21st
Data as Graphs

- Social Graphs
- Transportation Graphs
- Brain Graphs
- Web Graphs
- Molecular Graphs
- Gene Graphs
Machine Learning on Graphs

Representation Learning on Graphs

Graph

Features/Representation

Traditional i.i.d. data

Classification

Clustering

......

Ranking
Graph Neural Networks

Message Passing

\[ m_i^{(k+1)} = \sum_{v_j \in N(v_i)} M_k \left( h_i^{(k)}, h_j^{(k)}, e_{ij} \right) \]

Feature Updating

\[ h_i^{(k+1)} = U_k \left( h_i^{(k)}, m_i^{(k+1)} \right) \]

Neural Message Passing for Quantum Chemistry, Justin Gilmer et al, ICML 2017
A Unified View on Message Passing

"Noisy Signal"  

Graph

"Clean Signal"

"Nodes are similar to their neighbors"

\[
\text{arg min}_F \mathcal{L}(F) := \|F - X_{in}\|_F^2 + \mathcal{R}(F, \tilde{L})
\]

Close to the input  
Smoothness prior

A unified view on graph neural networks as graph signal denoising, Yao Ma, Xiaorui Liu et al, 2020
A Unified View on Message Passing

\[ \arg \min_{\mathbf{F}} \mathcal{L}(\mathbf{F}) := \| \mathbf{F} - \mathbf{X}_{\text{in}} \|_F^2 + \mathcal{R}(\mathbf{F}, \tilde{\mathbf{L}}) \]

Close to the input \quad Smoothness prior

**Define Prior \iff Optimization Solver \iff Message Passing**

**Example**

\[ \mathcal{R}(\mathbf{F}, \tilde{\mathbf{L}}) = \lambda \, \text{tr}(\mathbf{F}^\top \tilde{\mathbf{L}} \mathbf{F}) = \lambda \sum_{(v_i, v_j) \in \mathcal{E}} \left\| \frac{\mathbf{F}_i}{\sqrt{d_i + 1}} - \frac{\mathbf{F}_j}{\sqrt{d_j + 1}} \right\|_2^2 \]

- **GCN**
  \[ \mathbf{X}_{\text{out}} = \tilde{\mathbf{A}} \mathbf{X}_{\text{in}} \]

- **PPNP**
  \[ \mathbf{X}_{\text{out}} = \alpha (\mathbf{I} - (1 - \alpha)\tilde{\mathbf{A}})^{-1} \mathbf{X}_{\text{in}} \]

- **APPNP/GCNII**
  \[ \mathbf{X}^{(k+1)} = (1 - \alpha)\tilde{\mathbf{A}} \mathbf{X}^{(k)} + \alpha \mathbf{X}_{\text{in}} \]

A unified view on graph neural networks as graph signal denoising, Yao Ma, Xiaorui Liu et al, 2020
Global Smoothness

\[
\arg \min_F \mathcal{L}(F) := \|F - X_{in}\|_F^2 + \mathcal{R}(F, \tilde{L})
\]

Close to the input \quad Smoothness prior

Example

\[
\mathcal{R}(F, \tilde{L}) = \lambda \text{tr}(F^T\tilde{L}F) = \lambda \sum_{(v_i, v_j) \in E} \left\| \frac{F_i}{\sqrt{d_i} + 1} - \frac{F_j}{\sqrt{d_j} + 1} \right\|_2^2
\]

- **GCN** \quad \( X_{out} = \tilde{A}X_{in} \)
- **PPNP** \quad \( X_{out} = \alpha(I - (1 - \alpha)\tilde{A})^{-1}X_{in} \)
- **APPNP/GCNII** \quad \( X^{(k+1)} = (1 - \alpha)\tilde{A}X^{(k)} + \alpha X_{in} \)

*These MP schemes enforce global smoothness shared across the whole graph*
Local Smoothness

*Can we enhance local smoothness adaptively across different region over the graph?*

Noise graph structure
Local Smoothness

Adversarial graph attack

Graph attack

Feature smoothness

Graph Structure Learning for Robust Graph Neural Networks,
Trend Filtering

Nonparametric regression (univariate)

\[ \hat{\beta} = \arg\min_{\beta \in \mathbb{R}^n} \frac{1}{2} \| y - \beta \|_2^2 + \frac{n^k}{k!} \cdot \lambda \| D^{(k+1)} \beta \|_1 \]

Adapt to the local level of smoothness

\[ D^{(1)} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(n-1) \times n} \]

\[ D^{(k+1)} = D^{(1)} \cdot D^{(k)} \]

\[ L_1 \text{ Trend filtering}, \ S.-J. \ Kim \ et \ al, \ SIAM \ Review, \ 2009 \]

Graph Trend Filtering

GTF

\[
\arg \min_{f \in \mathbb{R}^n} \frac{1}{2} \| f - x \|^2 + \lambda \| \Delta^{(k+1)} f \|_1
\]

Incident matrix \( \Delta \ell = (0, \ldots, -1, \ldots, 1, \ldots, 0) \)

\[
\| \Delta^{(1)} f \|_1 = \sum_{(v_i, v_j) \in E} |f_i - f_j|
\]

\[
\Delta^{(k+1)} = \begin{cases} 
\Delta^\top \Delta^{(k)} = L^{\frac{k+1}{2}} & \in \mathbb{R}^{n \times n} \quad \text{for odd } k \\
\Delta \Delta^{(k)} = \Delta L^{\frac{k}{2}} & \in \mathbb{R}^{m \times n} \quad \text{for even } k
\end{cases}
\]

Trend filtering on graphs, Yu-Xiang Wang et al, JMLR 2016
Graph Trend Filtering

\[ \arg \min_{f \in \mathbb{R}^n} \frac{1}{2} \| f - x \|_2^2 + \lambda \| \Delta^{(k+1)} f \|_1 \]

Local smoothness adaptivity: piecewise behavior

Trend filtering on graphs, Yu-Xiang Wang et al, JMLR 2016
Elastic Graph Signal Estimator

\[
\arg \min_{\mathbf{F}} \mathcal{L}(\mathbf{F}) := \| \mathbf{F} - \mathbf{X}_{\text{in}} \|_F^2 + \mathcal{R}(\mathbf{F}, \tilde{\mathbf{L}})
\]

Close to the input

Smoothness prior

New smoothness prior

\[
\mathcal{R}(\mathbf{F}, \tilde{\mathbf{L}}) = \lambda_1 \| \tilde{\Delta} \mathbf{F} \|_1 + \frac{\lambda_2}{2} \text{tr}(\mathbf{F}^\top \tilde{\mathbf{L}} \mathbf{F})
\]

\[
\tilde{\Delta} = \Delta \mathbf{D}^{-\frac{1}{2}}
\]

\[
\| \tilde{\Delta} \mathbf{F} \|_1 = \sum_{(v_i, v_j) \in \mathcal{E}} \left\| \frac{\mathbf{F}_i}{\sqrt{d_i} + 1} - \frac{\mathbf{F}_j}{\sqrt{d_j} + 1} \right\|_1
\]

\[
\text{tr}(\mathbf{F}^\top \tilde{\mathbf{L}} \mathbf{F}) = \sum_{(v_i, v_j) \in \mathcal{E}} \left\| \frac{\mathbf{F}_i}{\sqrt{d_i} + 1} - \frac{\mathbf{F}_j}{\sqrt{d_j} + 1} \right\|_2^2
\]

Coupling multi-dimensionality

\[
\mathcal{R}(\mathbf{F}, \tilde{\mathbf{L}}) = \lambda_1 \| \tilde{\Delta} \mathbf{F} \|_{21} + \frac{\lambda_2}{2} \text{tr}(\mathbf{F}^\top \tilde{\mathbf{L}} \mathbf{F})
\]

\[
\| \tilde{\Delta} \mathbf{F} \|_{21} = \sum_{(v_i, v_j) \in \mathcal{E}} \left\| \frac{\mathbf{F}_i}{\sqrt{d_i} + 1} - \frac{\mathbf{F}_j}{\sqrt{d_j} + 1} \right\|_2
\]

\[
\text{tr}(\mathbf{F}^\top \tilde{\mathbf{L}} \mathbf{F}) = \sum_{(v_i, v_j) \in \mathcal{E}} \left\| \frac{\mathbf{F}_i}{\sqrt{d_i} + 1} - \frac{\mathbf{F}_j}{\sqrt{d_j} + 1} \right\|_2^2
\]

Data Science and Engineering Lab
Elastic Graph Signal Estimator

Option I

\[
\arg \min_{F \in \mathbb{R}^{n \times d}} \lambda_1 \| \tilde{\Delta} F \|_1 + \frac{\lambda_2}{2} \text{tr}(F^\top \tilde{L} F) + \frac{1}{2} \| F - X_{\text{in}} \|_F^2
\]

\[
\| \tilde{\Delta} F \|_1 = \sum_{(v_i, v_j) \in \mathcal{E}} \left\| \frac{F_i}{\sqrt{d_i} + 1} - \frac{F_j}{\sqrt{d_j} + 1} \right\|_1
\]

Option II

\[
\arg \min_{F \in \mathbb{R}^{n \times d}} \lambda_1 \| \tilde{\Delta} F \|_{21} + \frac{\lambda_2}{2} \text{tr}(F^\top \tilde{L} F) + \frac{1}{2} \| F - X_{\text{in}} \|_F^2
\]

\[
\| \tilde{\Delta} F \|_{21} = \sum_{(v_i, v_j) \in \mathcal{E}} \left\| \frac{F_i}{\sqrt{d_i} + 1} - \frac{F_j}{\sqrt{d_j} + 1} \right\|_2
\]

Define Prior  \iff  Optimization Solver  \iff  Message Passing
Elastic Graph Signal Estimator

\[ \text{arg min}_{F \in \mathbb{R}^{n \times d}} \underbrace{\lambda_1 \|	ilde{\Delta} F\|_{21}}_{g_{21}(\tilde{\Delta} F)} + \underbrace{\frac{\lambda_2}{2} \text{tr}(F^\top L F)}_{f(F)} + \frac{1}{2} \|F - X_{\text{in}}\|_F^2 \]

**Saddle-point reformulation**

\[ \min_{F} \max_{Z} f(F) + \langle \tilde{\Delta} F, Z \rangle - g^*(Z) \quad g^*(Z) := \sup_{X} \langle Z, X \rangle - g(X) \]

**A simple and efficient primal dual solver**

\[
\begin{align*}
F^{k+1} & = F^k - \gamma \nabla f(F^k) - \gamma \tilde{\Delta}^\top Z^k, \\
Z^{k+1} & = \text{prox}_{\beta g^*}(Z^k + \beta \tilde{\Delta} F^{k+1}), \\
F^{k+1} & = F^k - \gamma \nabla f(F^k) - \gamma \tilde{\Delta}^\top Z^{k+1},
\end{align*}
\]
Elastic Message Passing

\[
\begin{align*}
Y^{k+1} &= \gamma X_{in} + (1 - \gamma) \tilde{A} F^k \\
\tilde{F}^{k+1} &= Y^k - \gamma \tilde{\Delta}^\top Z^k \\
\tilde{Z}^{k+1} &= Z^k + \beta \tilde{\Delta} \tilde{F}^{k+1} \\
\{ \\
Z^{k+1} &= \min(\|\tilde{Z}^{k+1}\|_1, \lambda_1) \cdot \text{sign}(\tilde{Z}^{k+1}) \quad \text{(Option I: } \ell_1 \text{ norm)} \\
Z_i^{k+1} &= \min(\|\tilde{Z}_i^{k+1}\|_2, \lambda_1) \cdot \frac{\tilde{Z}_i^{k+1}}{\|\tilde{Z}_i^{k+1}\|_2}, \forall i \in [m] \quad \text{(Option II: } \ell_{21} \text{ norm)} \\
F^{k+1} &= Y^k - \gamma \tilde{\Delta}^\top Z^{k+1}
\}\end{align*}
\]

Figure 1. Elastic Message Passing (EMP). \(F^0 = X_{in}\) and \(Z^0 = 0^{m \times d}\).

Interpretation

- \(\lambda_1 = 0\): standard message passing in Y
  - \(\gamma = \frac{1}{1 + \lambda_2}, \lambda_2 = \frac{1}{\alpha} - 1\): \(F^{k+1} = \alpha X_{in} + (1 - \alpha) \tilde{A} F^k\)
  - \(\gamma = \frac{1}{1 + \lambda_2}, \lambda_2 = +\infty\): \(F^{k+1} = \tilde{A} F^k\)
- \(\lambda_1 > 0\): accumulate \(\tilde{\Delta}^\top Z\) to promote sparsity in \(\tilde{\Delta} F\) and preserve jump edge
Elastic Message Passing

\[
\begin{align*}
\mathbf{Y}^{k+1} &= \gamma \mathbf{X}_\text{in} + (1 - \gamma) \tilde{\mathbf{A}} \mathbf{F}^k \\
\mathbf{F}^{k+1} &= \mathbf{Y}^k - \gamma \tilde{\mathbf{A}}^\top \mathbf{Z}^k \\
\mathbf{Z}^{k+1} &= \mathbf{Z}^k + \beta \tilde{\mathbf{A}} \mathbf{F}^{k+1}
\end{align*}
\]

\[
\begin{cases}
\mathbf{Z}^{k+1} = \min(|\tilde{\mathbf{Z}}^{k+1}|, \lambda_1) \cdot \text{sign}(\tilde{\mathbf{Z}}^{k+1}) & \text{(Option I: } \ell_1 \text{ norm)} \\
\mathbf{Z}_i^{k+1} = \min(\|\tilde{\mathbf{Z}}_i^{k+1}\|_2, \lambda_1) \cdot \frac{\tilde{\mathbf{Z}}_i^{k+1}}{\|\tilde{\mathbf{Z}}_i^{k+1}\|_2}, \forall i \in [m] & \text{(Option II: } \ell_{21} \text{ norm)} \\
\mathbf{F}^{k+1} &= \mathbf{Y}^k - \gamma \tilde{\mathbf{A}}^\top \mathbf{Z}^{k+1}
\end{cases}
\]

Figure 1. Elastic Message Passing (EMP). \( \mathbf{F}^0 = \mathbf{X}_\text{in} \) and \( \mathbf{Z}^0 = \mathbf{0}^{m \times d} \).

**Theorem (Convergence)**

Under the stepsize setting \( \gamma < \frac{2}{1 + \lambda_2 \|\mathbf{\tilde{L}}\|_2} \) and \( \beta \leq \frac{4}{3\gamma \|\tilde{\mathbf{A}} \mathbf{\tilde{A}}^\top\|_2} \), the elastic message passing scheme (EMP) converges to the optimal solution of the elastic graph signal estimator. It is sufficient to choose any \( \gamma < \frac{2}{1 + 2\lambda_2} \) and \( \beta \leq \frac{2}{3\gamma} \) since \( \|\mathbf{\tilde{L}}\|_2 = \|\tilde{\mathbf{A}}^\top \tilde{\mathbf{A}}\|_2 = \|\tilde{\mathbf{A}} \mathbf{\tilde{A}}^\top\|_2 \leq 2 \).

In this work, we fix \( \gamma = \frac{1}{1 + \lambda_2}, \beta = \frac{1}{2\gamma} \).
Elastic GNNs

\[ Y_{pre} = \text{EMP} \left( h_{\theta}(X_{\text{fea}}), K, \lambda_1, \lambda_2 \right) \]

- Follow the decoupled architecture as PPNP but can be used in coupled architecture as well
- EMP is composed by simple and efficient operations, which is friendly to efficient and back-propagation training
- Hyperparameters \( \lambda_1 \) and \( \lambda_2 \) provide better smoothness adaptivity
- Doesn’t require a very large \( K \)
Performance on benchmark datasets

Semi-supervised learning for node classification

*Table 1.* Classification accuracy (%) on benchmark datasets with 10 times random data splits.

<table>
<thead>
<tr>
<th>Model</th>
<th>Cora</th>
<th>CiteSeer</th>
<th>PubMed</th>
<th>CS</th>
<th>Physics</th>
<th>Computers</th>
<th>Photo</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChebNet</td>
<td>76.3 ± 1.5</td>
<td>67.4 ± 1.5</td>
<td>75.0 ± 2.0</td>
<td>91.8 ± 0.4</td>
<td>OOM</td>
<td><strong>81.0 ± 2.0</strong></td>
<td>90.4 ± 1.0</td>
</tr>
<tr>
<td>GCN</td>
<td>79.6 ± 1.1</td>
<td>68.9 ± 1.2</td>
<td>77.6 ± 2.3</td>
<td>91.6 ± 0.6</td>
<td>93.3 ± 0.8</td>
<td>79.8 ± 1.6</td>
<td>90.3 ± 1.2</td>
</tr>
<tr>
<td>GAT</td>
<td>80.1 ± 1.2</td>
<td>68.9 ± 1.8</td>
<td>77.6 ± 2.2</td>
<td>91.1 ± 0.5</td>
<td>93.3 ± 0.7</td>
<td>79.3 ± 2.4</td>
<td>89.6 ± 1.6</td>
</tr>
<tr>
<td>SGC</td>
<td>80.2 ± 1.5</td>
<td>68.9 ± 1.3</td>
<td>75.5 ± 2.9</td>
<td>90.1 ± 1.3</td>
<td>93.1 ± 0.6</td>
<td>73.0 ± 2.0</td>
<td>83.5 ± 2.9</td>
</tr>
<tr>
<td>APPNP</td>
<td>82.2 ± 1.3</td>
<td>70.4 ± 1.2</td>
<td>78.9 ± 2.2</td>
<td><strong>92.5 ± 0.3</strong></td>
<td>93.7 ± 0.7</td>
<td>80.1 ± 2.1</td>
<td>90.8 ± 1.3</td>
</tr>
<tr>
<td>GraphSAGE</td>
<td>79.0 ± 1.1</td>
<td>67.5 ± 2.0</td>
<td>77.6 ± 2.0</td>
<td>91.7 ± 0.5</td>
<td>92.5 ± 0.8</td>
<td>80.7 ± 1.7</td>
<td>90.9 ± 1.0</td>
</tr>
<tr>
<td>ElasticGNN</td>
<td><strong>82.7 ± 1.0</strong></td>
<td><strong>70.9 ± 1.4</strong></td>
<td><strong>79.4 ± 1.8</strong></td>
<td><strong>92.5 ± 0.3</strong></td>
<td><strong>94.2 ± 0.5</strong></td>
<td>80.7 ± 1.8</td>
<td><strong>91.3 ± 1.3</strong></td>
</tr>
</tbody>
</table>

**ElasticGNN:** $L_{21}+L_2$
Performance on benchmark datasets

Better local smoothness adaptivity

*Table 3.* Ratio between average node differences along wrong and correct edges.

<table>
<thead>
<tr>
<th>Model</th>
<th>Cora</th>
<th>CiteSeer</th>
<th>PubMed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell_2$ (APPNP)</td>
<td>1.57</td>
<td>1.35</td>
<td>1.43</td>
</tr>
<tr>
<td>$\ell_{21}+\ell_2$ (ElasticGNN)</td>
<td>2.03</td>
<td>1.94</td>
<td>1.79</td>
</tr>
</tbody>
</table>

Piecewise constant prior

*Table 4.* Sparsity ratio (i.e., $\| (\tilde{\Delta F}_i) \|_2 < 0.1$) in node differences $\tilde{\Delta F}$.

<table>
<thead>
<tr>
<th>Model</th>
<th>Cora</th>
<th>CiteSeer</th>
<th>PubMed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell_2$ (APPNP)</td>
<td>2%</td>
<td>16%</td>
<td>11%</td>
</tr>
<tr>
<td>$\ell_{21}+\ell_2$ (ElasticGNN)</td>
<td>37%</td>
<td>74%</td>
<td>42%</td>
</tr>
</tbody>
</table>
Performance on benchmark datasets

**Impact of K**

![Graph showing the impact of K on test accuracy across different propagation steps.

**Convergence of EMP**

![Graph showing the convergence of the objective value for the problem in Eq. (8) during message passing.

*Figure 2.* Classification accuracy under different propagation steps.

*Figure 3.* Convergence of the objective value for the problem in Eq. (8) during message passing.
## Performance under adversarial attack

### Table 2. Classification accuracy (%) under different perturbation rates of adversarial graph attack.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Ptb Rate</th>
<th>Basic GNN</th>
<th>Elastic GNN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GCN</td>
<td>GAT</td>
</tr>
<tr>
<td>Cora</td>
<td>0%</td>
<td>83.5±0.4</td>
<td>84.0±0.7</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>76.6±0.8</td>
<td>80.4±0.7</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>70.4±1.3</td>
<td>75.6±0.6</td>
</tr>
<tr>
<td></td>
<td>15%</td>
<td>65.1±0.7</td>
<td>69.8±1.3</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>60.0±2.7</td>
<td>59.9±0.6</td>
</tr>
<tr>
<td>Citeseer</td>
<td>0%</td>
<td>72.0±0.6</td>
<td>73.3±0.8</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>70.9±0.6</td>
<td>72.9±0.8</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>67.6±0.9</td>
<td>70.6±0.5</td>
</tr>
<tr>
<td></td>
<td>15%</td>
<td>64.5±1.1</td>
<td>69.0±1.1</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>62.0±3.5</td>
<td>61.0±1.5</td>
</tr>
<tr>
<td>Polblogs</td>
<td>0%</td>
<td>95.7±0.4</td>
<td>95.4±0.2</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>73.1±0.8</td>
<td>83.7±1.5</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>70.7±1.1</td>
<td>76.3±0.9</td>
</tr>
<tr>
<td></td>
<td>15%</td>
<td>65.0±1.9</td>
<td>68.8±1.1</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>51.3±1.2</td>
<td>51.5±1.6</td>
</tr>
<tr>
<td>Pubmed</td>
<td>0%</td>
<td>87.2±0.1</td>
<td>83.7±0.4</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>83.1±0.1</td>
<td>78.0±0.4</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>81.2±0.1</td>
<td>74.9±0.4</td>
</tr>
<tr>
<td></td>
<td>15%</td>
<td>78.7±0.1</td>
<td>71.1±0.5</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>77.4±0.2</td>
<td>68.2±1.0</td>
</tr>
</tbody>
</table>

Basic GNNs < Elastic GNNs
## Performance under adversarial attack

*Table 2. Classification accuracy (%) under different perturbation rates of adversarial graph attack.*

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Pt Rate</th>
<th>Basic GNN</th>
<th></th>
<th>Elastic GNN</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GCN</td>
<td>GAT</td>
<td>$\ell_2$</td>
<td>$\ell_1$</td>
</tr>
<tr>
<td>Cora</td>
<td>0%</td>
<td>83.5±0.4</td>
<td>84.0±0.7</td>
<td>85.8±0.4</td>
<td>85.1±0.5</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>76.6±0.8</td>
<td>80.4±0.7</td>
<td>81.0±1.0</td>
<td>82.3±1.1</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>70.4±1.3</td>
<td>75.6±0.6</td>
<td>76.3±1.5</td>
<td>76.2±1.4</td>
</tr>
<tr>
<td></td>
<td>15%</td>
<td>65.1±0.7</td>
<td>69.8±1.3</td>
<td>72.2±0.9</td>
<td>73.3±1.3</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>60.0±2.7</td>
<td>59.9±0.6</td>
<td>67.7±0.7</td>
<td>63.7±0.9</td>
</tr>
<tr>
<td>Citeseer</td>
<td>0%</td>
<td>72.0±0.6</td>
<td>73.3±0.8</td>
<td>73.6±0.9</td>
<td>73.2±0.6</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>70.9±0.6</td>
<td>72.9±0.8</td>
<td>72.8±0.5</td>
<td>72.8±0.5</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>67.6±0.9</td>
<td>70.6±0.5</td>
<td>70.2±0.6</td>
<td>70.8±0.6</td>
</tr>
<tr>
<td></td>
<td>15%</td>
<td>64.5±1.1</td>
<td>69.0±1.1</td>
<td>70.2±0.6</td>
<td>68.1±1.4</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>62.0±3.5</td>
<td>61.0±1.5</td>
<td>64.9±1.0</td>
<td>64.7±0.8</td>
</tr>
<tr>
<td>Polblogs</td>
<td>0%</td>
<td>95.7±0.4</td>
<td>95.4±0.2</td>
<td>95.4±0.2</td>
<td>95.8±0.3</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>73.1±0.8</td>
<td>83.7±1.5</td>
<td>82.8±0.3</td>
<td>78.7±0.6</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>70.7±1.1</td>
<td>76.3±0.9</td>
<td>73.7±0.3</td>
<td>75.2±0.4</td>
</tr>
<tr>
<td></td>
<td>15%</td>
<td>65.0±1.9</td>
<td>68.8±1.1</td>
<td>68.9±0.9</td>
<td>72.1±0.9</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>51.3±1.2</td>
<td>51.5±1.6</td>
<td>65.5±0.7</td>
<td>68.1±0.6</td>
</tr>
<tr>
<td>Pubmed</td>
<td>0%</td>
<td>87.2±0.1</td>
<td>83.7±0.4</td>
<td>88.1±0.1</td>
<td>86.7±0.1</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>83.1±0.1</td>
<td>78.0±0.4</td>
<td>87.1±0.2</td>
<td>86.2±0.1</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>81.2±0.1</td>
<td>74.9±0.4</td>
<td>86.6±0.1</td>
<td>86.0±0.2</td>
</tr>
<tr>
<td></td>
<td>15%</td>
<td>78.7±0.1</td>
<td>71.1±0.5</td>
<td>85.7±0.2</td>
<td>85.4±0.2</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>77.4±0.2</td>
<td>68.2±1.0</td>
<td>85.8±0.1</td>
<td>85.4±0.1</td>
</tr>
</tbody>
</table>

$L_2 < L_{21}$ in most cases
### Table 2. Classification accuracy (%) under different perturbation rates of adversarial graph attack.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Ptb Rate</th>
<th>Basic GNN</th>
<th>Elastic GNN</th>
<th>(\ell_2)</th>
<th>(\ell_1)</th>
<th>(\ell_{21})</th>
<th>(\ell_1 + \ell_2)</th>
<th>(\ell_{21} + \ell_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GCN</td>
<td>GAT</td>
<td>(\ell_2)</td>
<td>(\ell_1)</td>
<td>(\ell_{21})</td>
<td>(\ell_1 + \ell_2)</td>
<td>(\ell_{21} + \ell_2)</td>
</tr>
<tr>
<td>Cora</td>
<td>0%</td>
<td>83.5±0.4</td>
<td>84.0±0.7</td>
<td>85.8±0.4</td>
<td>85.1±0.5</td>
<td>85.3±0.4</td>
<td>85.8±0.4</td>
<td>85.8±0.4</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>76.6±0.8</td>
<td>80.4±0.7</td>
<td>81.0±1.0</td>
<td>82.3±1.1</td>
<td>81.6±1.1</td>
<td>81.9±1.4</td>
<td>82.2±0.9</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>70.4±1.3</td>
<td>75.6±0.6</td>
<td>76.3±1.5</td>
<td>76.2±1.4</td>
<td>77.9±0.9</td>
<td>78.2±1.6</td>
<td>78.8±1.7</td>
</tr>
<tr>
<td></td>
<td>15%</td>
<td>65.1±0.7</td>
<td>69.8±1.3</td>
<td>72.2±0.9</td>
<td>73.3±1.3</td>
<td>75.7±1.2</td>
<td>76.9±0.9</td>
<td>77.2±1.6</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>60.0±2.7</td>
<td>59.9±0.6</td>
<td>67.7±0.7</td>
<td>63.7±0.9</td>
<td>70.3±1.1</td>
<td>67.2±5.3</td>
<td>70.5±1.3</td>
</tr>
<tr>
<td>Citeseer</td>
<td>0%</td>
<td>72.0±0.6</td>
<td>73.3±0.8</td>
<td>73.6±0.9</td>
<td>73.2±0.6</td>
<td>73.2±0.5</td>
<td>73.6±0.6</td>
<td>73.8±0.6</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>70.9±0.6</td>
<td>72.9±0.8</td>
<td>72.8±0.5</td>
<td>72.8±0.5</td>
<td>72.8±0.5</td>
<td>73.3±0.6</td>
<td>72.9±0.5</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>67.6±0.9</td>
<td>70.6±0.5</td>
<td>70.2±0.6</td>
<td>70.8±0.6</td>
<td>70.7±1.2</td>
<td>72.4±0.9</td>
<td>72.6±0.4</td>
</tr>
<tr>
<td></td>
<td>15%</td>
<td>64.5±1.1</td>
<td>69.0±1.1</td>
<td>70.2±0.6</td>
<td>68.1±1.4</td>
<td>68.2±1.1</td>
<td>71.3±1.5</td>
<td>71.9±0.7</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>62.0±3.5</td>
<td>61.0±1.5</td>
<td>64.9±1.0</td>
<td>64.7±0.8</td>
<td>64.7±0.8</td>
<td>64.7±0.8</td>
<td>64.7±0.8</td>
</tr>
<tr>
<td>Polblogs</td>
<td>0%</td>
<td>95.7±0.4</td>
<td>95.4±0.2</td>
<td>95.4±0.2</td>
<td>95.8±0.3</td>
<td>95.8±0.3</td>
<td>95.8±0.3</td>
<td>95.8±0.3</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>73.1±0.8</td>
<td>83.7±1.5</td>
<td>82.8±0.3</td>
<td>78.7±0.6</td>
<td>78.7±0.7</td>
<td>82.8±0.4</td>
<td>83.0±0.3</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>70.7±1.1</td>
<td>76.3±0.9</td>
<td>73.7±0.3</td>
<td>75.2±0.4</td>
<td>75.3±0.7</td>
<td>81.5±0.2</td>
<td>81.6±0.3</td>
</tr>
<tr>
<td></td>
<td>15%</td>
<td>65.0±1.9</td>
<td>68.8±1.1</td>
<td>68.9±0.9</td>
<td>72.1±0.9</td>
<td>71.5±1.1</td>
<td>77.8±0.9</td>
<td>78.7±0.5</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>51.3±1.2</td>
<td>51.5±1.6</td>
<td>65.5±0.7</td>
<td>68.1±0.6</td>
<td>68.7±0.7</td>
<td>77.4±0.2</td>
<td>77.5±0.2</td>
</tr>
<tr>
<td>Pubmed</td>
<td>0%</td>
<td>87.2±0.1</td>
<td>83.7±0.4</td>
<td>88.1±0.1</td>
<td>86.7±0.1</td>
<td>87.3±0.1</td>
<td>88.1±0.1</td>
<td>88.1±0.1</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>83.1±0.1</td>
<td>78.0±0.4</td>
<td>87.1±0.2</td>
<td>86.2±0.1</td>
<td>87.0±0.1</td>
<td>87.1±0.2</td>
<td>87.1±0.2</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>81.2±0.1</td>
<td>74.9±0.4</td>
<td>86.6±0.1</td>
<td>86.0±0.2</td>
<td>86.9±0.2</td>
<td>86.3±0.1</td>
<td>87.0±0.1</td>
</tr>
<tr>
<td></td>
<td>15%</td>
<td>78.7±0.1</td>
<td>71.1±0.5</td>
<td>85.7±0.2</td>
<td>85.4±0.2</td>
<td>86.4±0.2</td>
<td>85.5±0.1</td>
<td>86.4±0.2</td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>77.4±0.2</td>
<td>68.2±1.0</td>
<td>85.8±0.1</td>
<td>85.4±0.1</td>
<td>86.4±0.1</td>
<td>85.4±0.1</td>
<td>86.4±0.1</td>
</tr>
</tbody>
</table>

\[L_1 + L_2 < L_{21} + L_2\] in most cases
Conclusion

Summary
• Introduce $L_1$ based graph smoothing in the design of GNNs, for the first time
• Derive a novel and general message passing scheme, i.e., EMP
• Develop a family of GNNs, i.e., Elastic GNNs
• Demonstrate better smoothness adaptivity of Elastic GNNs
• Elastic GNNs are intrinsically more robust to adversarial graph attacks and compatible with any other defense strategies

Future directions
• Other node level tasks such as link prediction, community detection, and outlier detection
• Graph level tasks such as graph classification and graph similarity measure
• Higher-order graph difference operators
• EMP as a building block in other GNN architectures

Code: https://github.com/lxiaorui/ElasticGNN
Acknowledgement

Thanks for the funding support from National Science Foundation (NSF), Army Research Office (ARO) and Facebook Faculty Research Award.