Elastic Graph Neural Networks

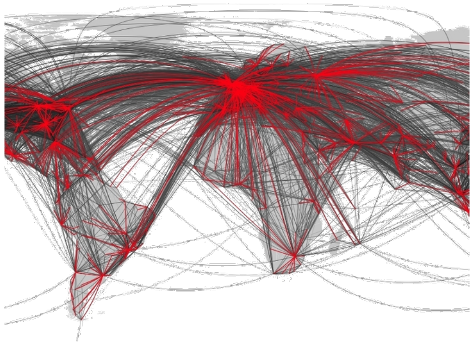
Xiaorui Liu
Joint work with Wei Jin (co-first author),
Yao Ma, Yaxin Li, Hua Liu, Yiqi Wang,
Ming Yan & Jiliang Tang

Michigan State University

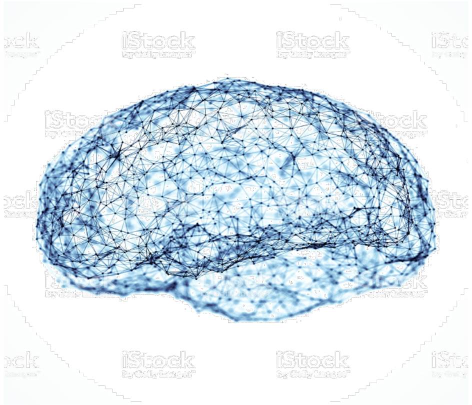
ICML 2021, July 21st

Data as Graphs

Social Graphs

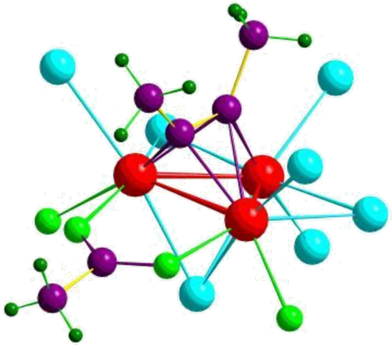


Transportation Graphs

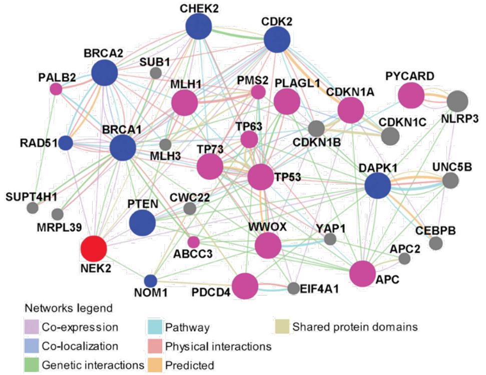


Brain Graphs

Web Graphs

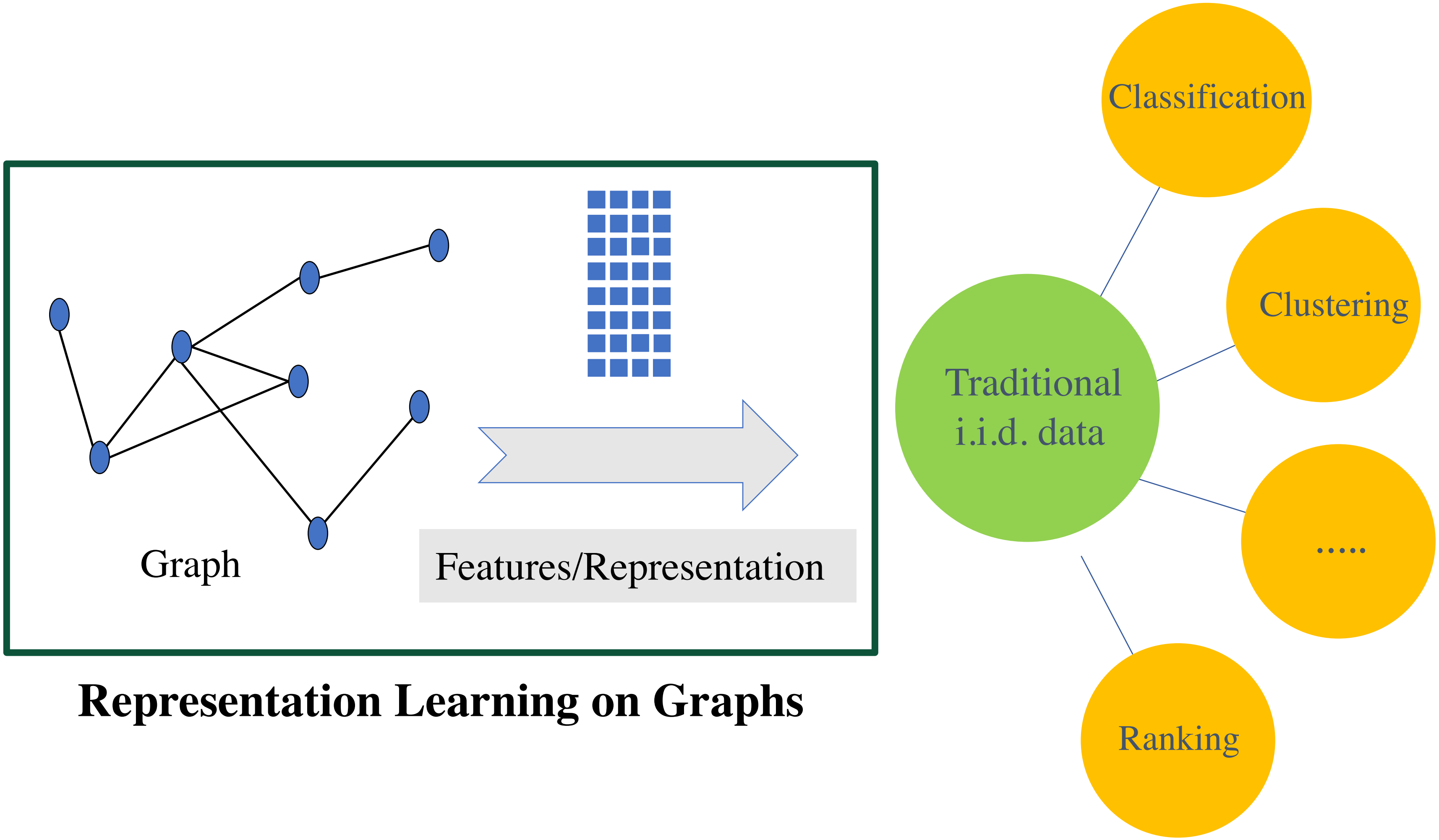


Molecular Graphs



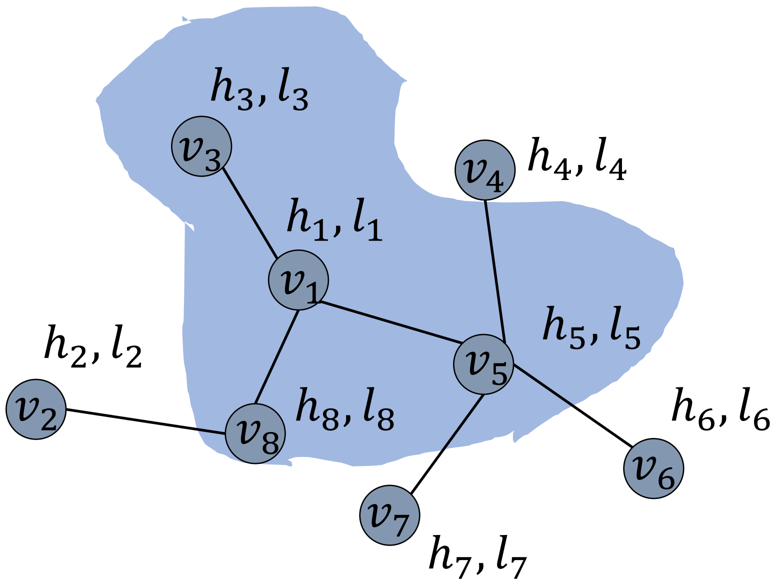
Gene Graphs

Machine Learning on Graphs



Representation Learning on Graphs

Graph Neural Networks



Message Passing

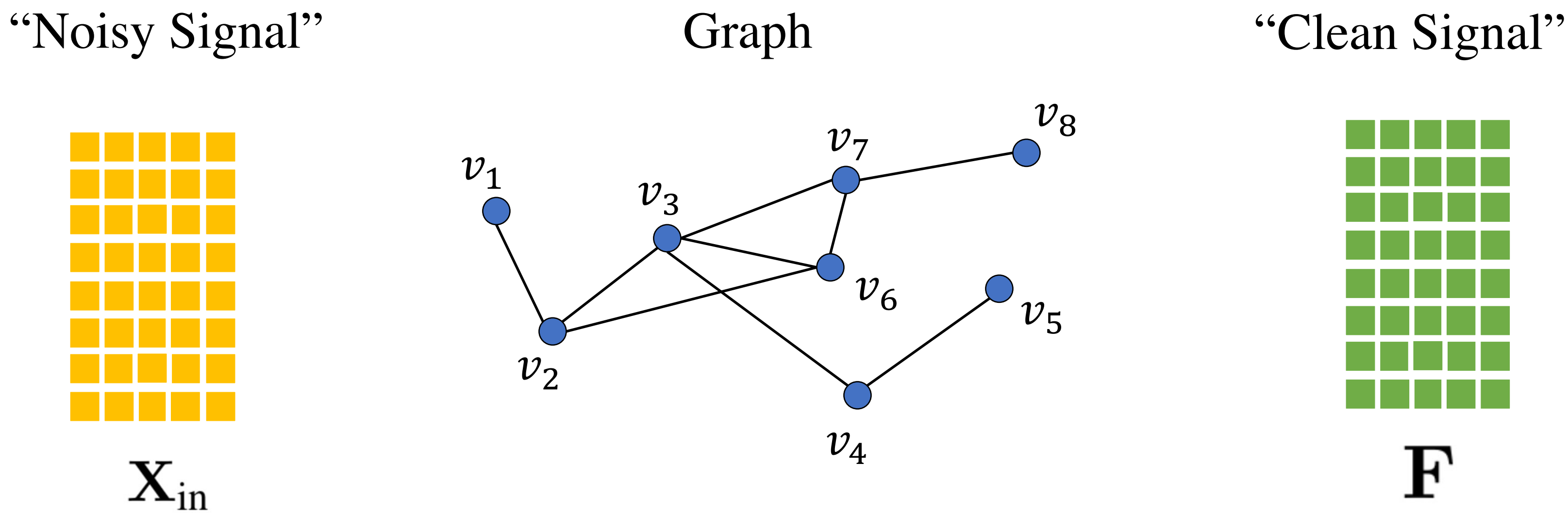
$$m_i^{(k+1)} = \sum_{v_j \in N(v_i)} M_k\left(h_i^{(k)}, h_j^{(k)}, e_{ij}\right)$$

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



"Nodes are similar to their neighbors"

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

Close to the input **Smoothness prior**

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 **Smoothness prior**

Define Prior ⇨ Optimization Solver ⇨ Message Passing

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

- GCN $\mathbf{X}_{\text{out}} = \tilde{\mathbf{A}} \mathbf{X}_{\text{in}}$
- PPNP $\mathbf{X}_{\text{out}} = \alpha \big(\mathbf{I} - (1-\alpha)\tilde{\mathbf{A}}\big)^{-1} \mathbf{X}_{\text{in}}$
- 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

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

Close to the input **Smoothness prior**

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

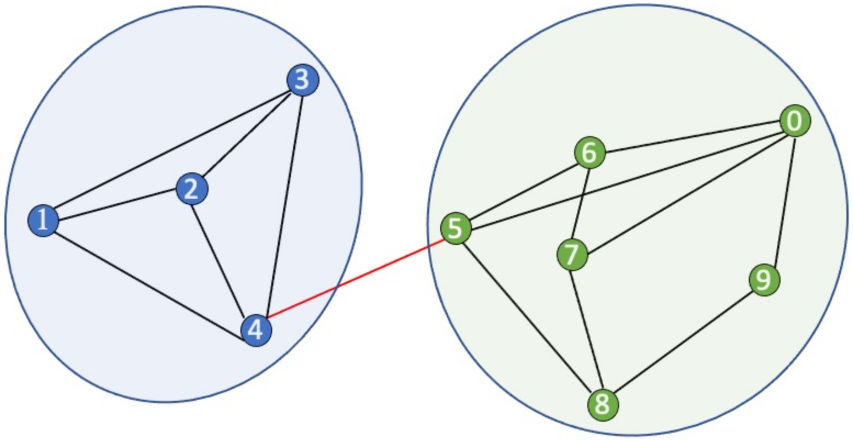
- GCN $$\mathbf{X}_{\text{out}} = \tilde{\mathbf{A}} \mathbf{X}_{\text{in}}$$
- PPNP $$\mathbf{X}_{\text{out}} = \alpha \left(\mathbf{I} - (1 - \alpha) \tilde{\mathbf{A}} \right)^{-1} \mathbf{X}_{\text{in}}$$
- APPNP/GCNII $$\mathbf{X}^{(k+1)} = (1 - \alpha) \tilde{\mathbf{A}} \mathbf{X}^{(k)} + \alpha \mathbf{X}_{\text{in}}$$

These MP schemes enforce global smoothness shared across the whole graph

Local Smoothness

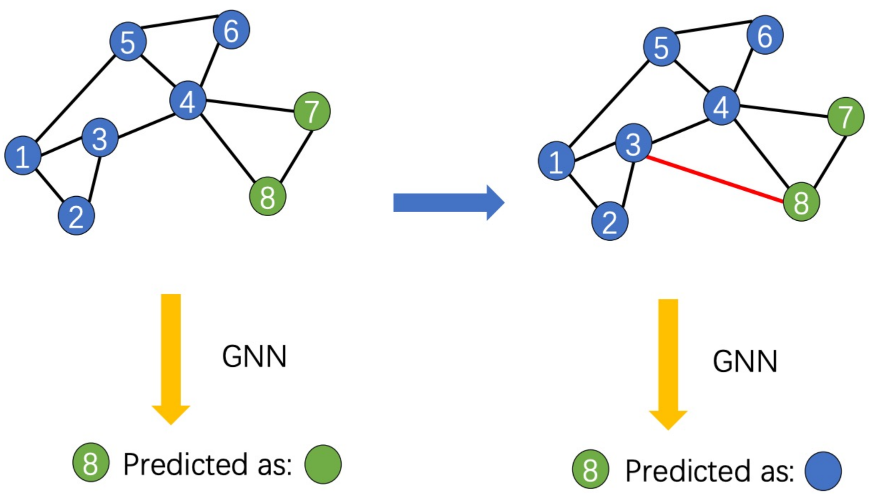
Can we enhance local smoothness adaptively across different region over the graph?

Noise graph structure

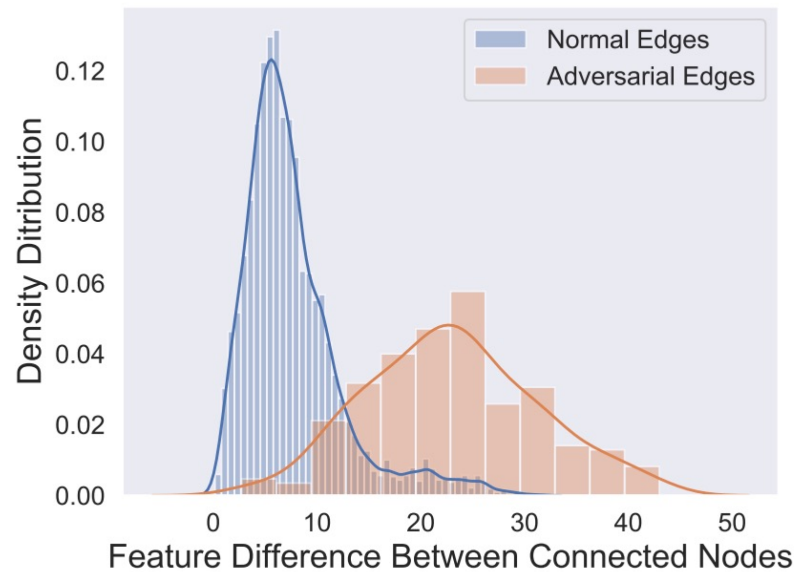


Local Smoothness

Adversarial graph attack



Graph attack



Feature smoothness

Graph Structure Learning for Robust Graph Neural Networks, Wei Jin, Yao Ma, Xiaorui Liu, et al, KDD 2020.

Trend Filtering

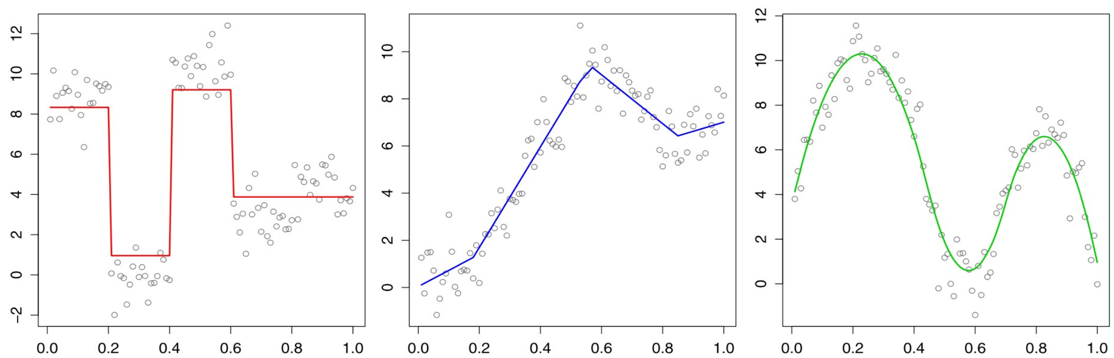
Nonparametric regression (univariate)

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2}\|y - \beta\|_2^2 + \frac{n^k}{k!} \cdot \lambda \|D^{(k+1)}\beta\|_1$$

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

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

Adapt to the local level of smoothness



K = 0 K = 1 K = 2

***L_1* Trend filtering**, S.-J. Kim et al, SIAM Review, 2009

Adaptive piecewise polynomial estimation via trend filtering, Ryan Tibshirani, Annals of Statistics, 2014

Graph Trend Filtering

GTF

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

Incident matrix

$$\Delta_\ell = (0,\ldots,\underbrace{-1}_{i},\ldots,\underbrace{1}_{j},\ldots,0)$$

$$\|\Delta^{(1)}\mathbf{f}\|_1 = \sum_{(v_i,v_j)\in\mathcal{E}} |\mathbf{f}_i - \mathbf{f}_j|$$

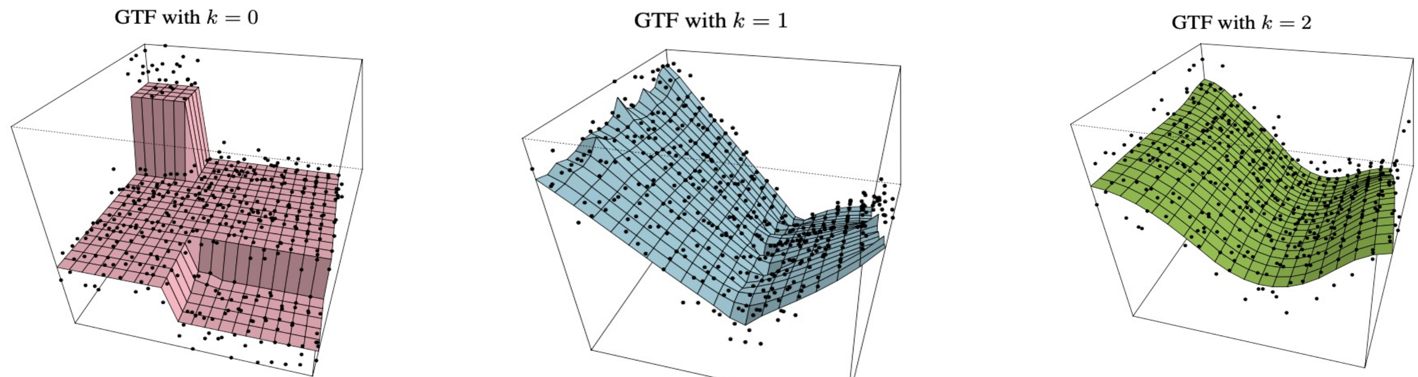
$$\Delta^{(k+1)} = \begin{cases} \Delta^\top\Delta^{(k)} = \mathbf{L}^{\frac{k+1}{2}} \in \mathbb{R}^{n\times n} & \text{for odd k} \\ \Delta\Delta^{(k)} = \Delta\mathbf{L}^{\frac{k}{2}} \in \mathbb{R}^{m\times n} & \text{for even k} \end{cases}$$

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

Graph Trend Filtering

$$\underset{\mathbf{f}\in\mathbb{R}^n}{\arg\min} = \frac{1}{2}\|\mathbf{f}-\mathbf{x}\|_2^2 + \lambda\|\Delta^{(k+1)}\mathbf{f}\|_1$$

Local smoothness adaptivity: piecewise behavior



Penalty: $\sum_{(i,j)\in E} |\beta_i - \beta_j|$

$$\sum_{i=1}^{n} n_i \left|\beta_i - \frac{1}{n_i}\sum_{j:(i,j)\in E}\beta_j\right|$$

$$\sum_{(i,j)\in E}\left|n_i\left(\beta_i - \frac{1}{n_i}\sum_{\ell:(i,\ell)\in E}\beta_\ell\right) - n_j\left(\beta_j - \frac{1}{n_j}\sum_{\ell:(j,\ell)\in E}\beta_\ell\right)\right|$$

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

$$\mathcal{R}(\mathbf{F}, \tilde{\mathbf{L}}) = \lambda_1 \|\tilde{\Delta}\mathbf{F}\|_1 + \frac{\lambda_2}{2} \mathrm{tr}(\mathbf{F}^\top \tilde{\mathbf{L}} \mathbf{F}) \qquad \tilde{\boldsymbol{\Delta}} = \boldsymbol{\Delta}\hat{\mathbf{D}}^{-\frac{1}{2}}$$

$$\|\tilde{\Delta}\mathbf{F}\|_1 = \sum_{(v_i, v_j)\in\mathcal{E}} \left\| \frac{\mathbf{F}_i}{\sqrt{d_i+1}} - \frac{\mathbf{F}_j}{\sqrt{d_j+1}} \right\|_1 \qquad \mathrm{tr}(\mathbf{F}^\top \tilde{\mathbf{L}} \mathbf{F}) = \sum_{(v_i, v_j)\in\mathcal{E}} \left\| \frac{\mathbf{F}_i}{\sqrt{d_i+1}} - \frac{\mathbf{F}_j}{\sqrt{d_j+1}} \right\|_2^2$$

Coupling multi-dimensionality

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

$$\|\tilde{\Delta}\mathbf{F}\|_{21} = \sum_{(v_i, v_j)\in\mathcal{E}} \left\| \frac{\mathbf{F}_i}{\sqrt{d_i+1}} - \frac{\mathbf{F}_j}{\sqrt{d_j+1}} \right\|_2 \qquad \mathrm{tr}(\mathbf{F}^\top \tilde{\mathbf{L}} \mathbf{F}) = \sum_{(v_i, v_j)\in\mathcal{E}} \left\| \frac{\mathbf{F}_i}{\sqrt{d_i+1}} - \frac{\mathbf{F}_j}{\sqrt{d_j+1}} \right\|_2^2$$

Elastic Graph Signal Estimator

Option I

$$\arg\min_{\mathbf{F}\in\mathbb{R}^{n\times d}} \underbrace{\lambda_1\|\tilde{\Delta}\mathbf{F}\|_1}_{g_1(\tilde{\Delta}\mathbf{F})} + \underbrace{\frac{\lambda_2}{2}\mathrm{tr}(\mathbf{F}^\top\tilde{\mathbf{L}}\mathbf{F}) + \frac{1}{2}\|\mathbf{F}-\mathbf{X}_{\mathrm{in}}\|_F^2}_{f(\mathbf{F})}$$

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

Option II

$$\arg\min_{\mathbf{F}\in\mathbb{R}^{n\times d}} \underbrace{\lambda_1\|\tilde{\Delta}\mathbf{F}\|_{21}}_{g_{21}(\tilde{\Delta}\mathbf{F})} + \underbrace{\frac{\lambda_2}{2}\mathrm{tr}(\mathbf{F}^\top\tilde{\mathbf{L}}\mathbf{F}) + \frac{1}{2}\|\mathbf{F}-\mathbf{X}_{\mathrm{in}}\|_F^2}_{f(\mathbf{F})}$$

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

Define Prior ⇨ Optimization Solver ⇨ Message Passing

Elastic Graph Signal Estimator

$$\underset{\mathbf{F}\in\mathbb{R}^{n\times d}}{\arg\min}\ \underbrace{\lambda_1\|\tilde{\Delta}\mathbf{F}\|_{21}}_{g_{21}(\tilde{\Delta}\mathbf{F})} + \underbrace{\frac{\lambda_2}{2}\mathrm{tr}(\mathbf{F}^\top\tilde{\mathbf{L}}\mathbf{F}) + \frac{1}{2}\|\mathbf{F}-\mathbf{X}_{\mathrm{in}}\|_F^2}_{f(\mathbf{F})}$$

Saddle-point reformulation

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

A simple and efficient primal dual solver

$$\begin{cases}\bar{\mathbf{F}}^{k+1} &= \mathbf{F}^k - \gamma\nabla f(\mathbf{F}^k) - \gamma\tilde{\Delta}^\top\mathbf{Z}^k,\\ \mathbf{Z}^{k+1} &= \mathbf{prox}_{\beta g^*}(\mathbf{Z}^k + \beta\tilde{\Delta}\bar{\mathbf{F}}^{k+1}),\\ \mathbf{F}^{k+1} &= \mathbf{F}^k - \gamma\nabla f(\mathbf{F}^k) - \gamma\tilde{\Delta}^\top\mathbf{Z}^{k+1},\end{cases}$$

Elastic Message Passing

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

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

Interpretation

- $\lambda_1 = 0$: standard message passing in Y
  - $\gamma = \frac{1}{1+\lambda_2}$, $\lambda_2 = \frac{1}{\alpha} - 1$: $\mathbf{F}^{k+1} = \alpha \mathbf{X}_{in} + (1-\alpha)\tilde{\mathbf{A}}\mathbf{F}^k$
  - $\gamma = \frac{1}{1+\lambda_2}$, $\lambda_2 = +\infty$: $\mathbf{F}^{k+1} = \tilde{\mathbf{A}}\mathbf{F}^k$
- $\lambda_1 > 0$: accumulate $\tilde{\Delta}^T Z$ to promote sparsity in $\tilde{\Delta} F$ and preserve jump edge

Elastic Message Passing

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

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 $\gamma < \frac{2}{1+\lambda_2\|\tilde{\mathbf{L}}\|_2}$ and $\beta \leq \frac{4}{3\gamma\|\tilde{\Delta}\tilde{\Delta}^\top\|_2}$, 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 $\|\tilde{\mathbf{L}}\|_2 = \|\tilde{\Delta}^\top\tilde{\Delta}\|_2 = \|\tilde{\Delta}\tilde{\Delta}^\top\|_2 \leq 2$.

In this work, we fix $\gamma = \frac{1}{1+\lambda_2}, \beta = \frac{1}{2\gamma}$

Elastic GNNs

$$\mathbf{Y}_{\text{pre}} = \mathbf{EMP}\left(h_\theta(\mathbf{X}_{\text{fea}}), K, \lambda_1, \lambda_2\right)$$

- 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 λ_1 and λ_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.

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

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.

Model	Cora	CiteSeer	PubMed
ℓ_2 (APPNP)	1.57	1.35	1.43
ℓ_{21}+ℓ_2 (ElasticGNN)	2.03	1.94	1.79

Piecewise constant prior

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

Model	Cora	CiteSeer	PubMed
ℓ_2 (APPNP)	2%	16%	11%
ℓ_{21}+ℓ_2 (ElasticGNN)	37%	74%	42%

Performance on benchmark datasets

Impact of K

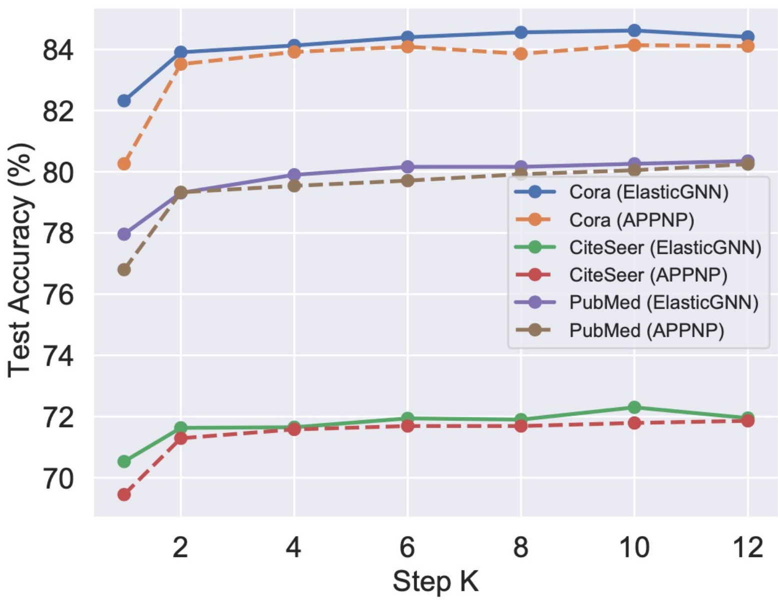


Figure 2. Classification accuracy under different propagation steps.

Convergence of EMP

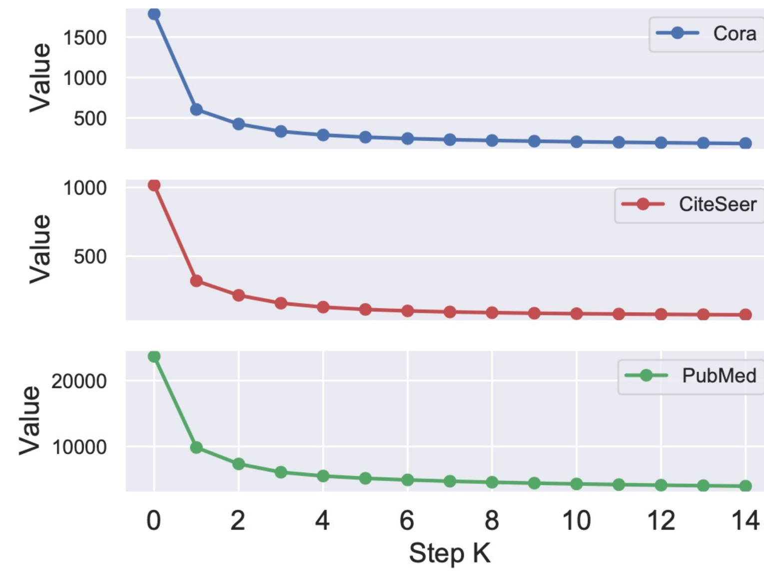


Figure 3. Convergence of the objective value for the problem in Eq. (8) during message passing.

Performance under adversarial attack

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

Dataset	Ptb Rate	Basic GNN GCN	Basic GNN GAT	Elastic GNN ℓ_2	Elastic GNN ℓ_1	Elastic GNN ℓ_{21}	Elastic GNN $\ell_1 + \ell_2$	Elastic GNN $\ell_{21} + \ell_2$
Cora	0%	83.5±0.4	84.0±0.7	**85.8±0.4**	85.1±0.5	85.3±0.4	**85.8±0.4**	**85.8±0.4**
	5%	76.6±0.8	80.4±0.7	81.0±1.0	**82.3±1.1**	81.6±1.1	81.9±1.4	82.2±0.9
	10%	70.4±1.3	75.6±0.6	76.3±1.5	76.2±1.4	77.9±0.9	78.2±1.6	**78.8±1.7**
	15%	65.1±0.7	69.8±1.3	72.2±0.9	73.3±1.3	75.7±1.2	76.9±0.9	**77.2±1.6**
	20%	60.0±2.7	59.9±0.6	67.7±0.7	63.7±0.9	70.3±1.1	67.2±5.3	**70.5±1.3**
Citeseer	0%	72.0±0.6	73.3±0.8	73.6±0.9	73.2±0.6	73.2±0.5	73.6±0.6	**73.8±0.6**
	5%	70.9±0.6	72.9±0.8	72.8±0.5	72.8±0.5	72.8±0.5	**73.3±0.6**	72.9±0.5
	10%	67.6±0.9	70.6±0.5	70.2±0.6	70.8±0.6	70.7±1.2	72.4±0.9	**72.6±0.4**
	15%	64.5±1.1	69.0±1.1	70.2±0.6	68.1±1.4	68.2±1.1	71.3±1.5	**71.9±0.7**
	20%	62.0±3.5	61.0±1.5	**64.9±1.0**	64.7±0.8	64.7±0.8	64.7±0.8	64.7±0.8
Polblogs	0%	95.7±0.4	95.4±0.2	95.4±0.2	**95.8±0.3**	**95.8±0.3**	**95.8±0.3**	**95.8±0.3**
	5%	73.1±0.8	83.7±1.5	82.8±0.3	78.7±0.6	78.7±0.7	82.8±0.4	**83.0±0.3**
	10%	70.7±1.1	76.3±0.9	73.7±0.3	75.2±0.4	75.3±0.7	81.5±0.2	**81.6±0.3**
	15%	65.0±1.9	68.8±1.1	68.9±0.9	72.1±0.9	71.5±1.1	77.8±0.9	**78.7±0.5**
	20%	51.3±1.2	51.5±1.6	65.5±0.7	68.1±0.6	68.7±0.7	77.4±0.2	**77.5±0.2**
Pubmed	0%	87.2±0.1	83.7±0.4	**88.1±0.1**	86.7±0.1	87.3±0.1	**88.1±0.1**	**88.1±0.1**
	5%	83.1±0.1	78.0±0.4	**87.1±0.2**	86.2±0.1	87.0±0.1	**87.1±0.2**	**87.1±0.2**
	10%	81.2±0.1	74.9±0.4	86.6±0.1	86.0±0.2	86.9±0.2	86.3±0.1	**87.0±0.1**
	15%	78.7±0.1	71.1±0.5	85.7±0.2	85.4±0.2	86.4±0.2	85.5±0.1	**86.4±0.2**
	20%	77.4±0.2	68.2±1.0	85.8±0.1	85.4±0.1	86.4±0.1	85.4±0.1	**86.4±0.1**

Basic GNNs < Elastic GNNs

Performance under adversarial attack

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

Dataset	Ptb Rate	Basic GNN		Elastic GNN				
		GCN	GAT	ℓ_2	ℓ_1	ℓ_{21}	$\ell_1 + \ell_2$	$\ell_{21} + \ell_2$
Cora	0%	83.5±0.4	84.0±0.7	**85.8±0.4**	85.1±0.5	85.3±0.4	**85.8±0.4**	**85.8±0.4**
	5%	76.6±0.8	80.4±0.7	81.0±1.0	**82.3±1.1**	81.6±1.1	81.9±1.4	82.2±0.9
	10%	70.4±1.3	75.6±0.6	76.3±1.5	76.2±1.4	77.9±0.9	78.2±1.6	**78.8±1.7**
	15%	65.1±0.7	69.8±1.3	72.2±0.9	73.3±1.3	75.7±1.2	76.9±0.9	**77.2±1.6**
	20%	60.0±2.7	59.9±0.6	67.7±0.7	63.7±0.9	70.3±1.1	67.2±5.3	**70.5±1.3**
Citeseer	0%	72.0±0.6	73.3±0.8	73.6±0.9	73.2±0.6	73.2±0.5	73.6±0.6	**73.8±0.6**
	5%	70.9±0.6	72.9±0.8	72.8±0.5	72.8±0.5	72.8±0.5	**73.3±0.6**	72.9±0.5
	10%	67.6±0.9	70.6±0.5	70.2±0.6	70.8±0.6	70.7±1.2	72.4±0.9	**72.6±0.4**
	15%	64.5±1.1	69.0±1.1	70.2±0.6	68.1±1.4	68.2±1.1	71.3±1.5	**71.9±0.7**
	20%	62.0±3.5	61.0±1.5	**64.9±1.0**	64.7±0.8	64.7±0.8	64.7±0.8	64.7±0.8
Polblogs	0%	95.7±0.4	95.4±0.2	95.4±0.2	**95.8±0.3**	**95.8±0.3**	**95.8±0.3**	**95.8±0.3**
	5%	73.1±0.8	83.7±1.5	82.8±0.3	78.7±0.6	78.7±0.7	82.8±0.4	**83.0±0.3**
	10%	70.7±1.1	76.3±0.9	73.7±0.3	75.2±0.4	75.3±0.7	81.5±0.2	**81.6±0.3**
	15%	65.0±1.9	68.8±1.1	68.9±0.9	72.1±0.9	71.5±1.1	77.8±0.9	**78.7±0.5**
	20%	51.3±1.2	51.5±1.6	65.5±0.7	68.1±0.6	68.7±0.7	77.4±0.2	**77.5±0.2**
Pubmed	0%	87.2±0.1	83.7±0.4	**88.1±0.1**	86.7±0.1	87.3±0.1	**88.1±0.1**	**88.1±0.1**
	5%	83.1±0.1	78.0±0.4	**87.1±0.2**	86.2±0.1	87.0±0.1	**87.1±0.2**	**87.1±0.2**
	10%	81.2±0.1	74.9±0.4	86.6±0.1	86.0±0.2	86.9±0.2	86.3±0.1	**87.0±0.1**
	15%	78.7±0.1	71.1±0.5	85.7±0.2	85.4±0.2	86.4±0.2	85.5±0.1	**86.4±0.2**
	20%	77.4±0.2	68.2±1.0	85.8±0.1	85.4±0.1	86.4±0.1	85.4±0.1	**86.4±0.1**

$L_2 < L_{21}$ in most cases

Performance under adversarial attack

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

Dataset	Ptb Rate	Basic GNN GCN	Basic GNN GAT	Elastic GNN ℓ_2	Elastic GNN ℓ_1	Elastic GNN ℓ_{21}	Elastic GNN $\ell_1 + \ell_2$	Elastic GNN $\ell_{21} + \ell_2$
Cora	0%	83.5±0.4	84.0±0.7	**85.8±0.4**	85.1±0.5	85.3±0.4	**85.8±0.4**	**85.8±0.4**
	5%	76.6±0.8	80.4±0.7	81.0±1.0	**82.3±1.1**	81.6±1.1	81.9±1.4	82.2±0.9
	10%	70.4±1.3	75.6±0.6	76.3±1.5	76.2±1.4	77.9±0.9	78.2±1.6	**78.8±1.7**
	15%	65.1±0.7	69.8±1.3	72.2±0.9	73.3±1.3	75.7±1.2	76.9±0.9	**77.2±1.6**
	20%	60.0±2.7	59.9±0.6	67.7±0.7	63.7±0.9	70.3±1.1	67.2±5.3	**70.5±1.3**
Citeseer	0%	72.0±0.6	73.3±0.8	73.6±0.9	73.2±0.6	73.2±0.5	73.6±0.6	**73.8±0.6**
	5%	70.9±0.6	72.9±0.8	72.8±0.5	72.8±0.5	72.8±0.5	**73.3±0.6**	72.9±0.5
	10%	67.6±0.9	70.6±0.5	70.2±0.6	70.8±0.6	70.7±1.2	72.4±0.9	**72.6±0.4**
	15%	64.5±1.1	69.0±1.1	70.2±0.6	68.1±1.4	68.2±1.1	71.3±1.5	**71.9±0.7**
	20%	62.0±3.5	61.0±1.5	**64.9±1.0**	64.7±0.8	64.7±0.8	64.7±0.8	64.7±0.8
Polblogs	0%	95.7±0.4	95.4±0.2	95.4±0.2	**95.8±0.3**	**95.8±0.3**	**95.8±0.3**	**95.8±0.3**
	5%	73.1±0.8	83.7±1.5	82.8±0.3	78.7±0.6	78.7±0.7	82.8±0.4	**83.0±0.3**
	10%	70.7±1.1	76.3±0.9	73.7±0.3	75.2±0.4	75.3±0.7	81.5±0.2	**81.6±0.3**
	15%	65.0±1.9	68.8±1.1	68.9±0.9	72.1±0.9	71.5±1.1	77.8±0.9	**78.7±0.5**
	20%	51.3±1.2	51.5±1.6	65.5±0.7	68.1±0.6	68.7±0.7	77.4±0.2	**77.5±0.2**
Pubmed	0%	87.2±0.1	83.7±0.4	**88.1±0.1**	86.7±0.1	87.3±0.1	**88.1±0.1**	**88.1±0.1**
	5%	83.1±0.1	78.0±0.4	**87.1±0.2**	86.2±0.1	87.0±0.1	**87.1±0.2**	**87.1±0.2**
	10%	81.2±0.1	74.9±0.4	86.6±0.1	86.0±0.2	86.9±0.2	86.3±0.1	**87.0±0.1**
	15%	78.7±0.1	71.1±0.5	85.7±0.2	85.4±0.2	86.4±0.2	85.5±0.1	**86.4±0.2**
	20%	77.4±0.2	68.2±1.0	85.8±0.1	85.4±0.1	86.4±0.1	85.4±0.1	**86.4±0.1**

$L_1 + L_2 < L_{21} + L_2$ 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 Founding (NSF), Army Research Office (ARO) and Facebook Faculty Research Award.