#5835. Explainable Graph Representation Learning via Graph Pattern Analysis

Xudong Wang[1], Ziheng Sun[1,2], Chris Ding[1], Jicong Fan[1,2]
[1]School of Data Science, CUHK-Shenzhen [2]Shenzhen Research Institute of Big Data
{xudongwang, zihengsun}@link.cuhk.edu.cn, {chrisding, fanjicong}@cuhk.edu.cn

Introduction & Motivation

Explainable AI (XAI) is critical for trustworthy models in domains like healthcare, finance, and transportation. While existing Graph XGL focuses on model/instance levels, **representation-level explainability** remains unexplored.

Fundamental Question

What specific information about a graph is captured in its vector representation g?

Why it matters: Graph patterns carry real-world meaning:
- **Cycles** in molecules indicate chemical properties
- **Cliques** characterize protein complexes
- **Paths** reflect information flow in networks

Limitation of Existing Methods: Graph kernels count patterns but often ignore node features and are high-dimensional. GNNs are powerful but usually lack transparency at the representation level.

Core Contributions & Takeaways

1. **Formal framework** for representation-level XGL. Graph pattern attributions can be inspected via joint learning.
2. **Two novel methods:**
   - **PXGL-EGK**: Explainable ensemble kernels
   - **PXGL-GNN**: Deep pattern-based representations for feature-aware, end-to-end learning
3. **Theoretical guarantees:** Robustness & generalization bounds; Complexity analysis on proposed methods.
4. **Superior performance:** Validated on various benchmarks

Potential Future Work: Extend to dynamic graphs and heterogeneous networks.

The Proposed PXGL-EGK

Given graph $G = (\mathbf{A}, \mathbf{X})$ with adjacency matrix $\mathbf{A}$ and node features $\mathbf{X}$:
Pattern Counting Vector: $\mathbf{h} = \phi(G; \mathcal{P})$ where $h^{(i)}$ counts pattern P_i occurrences
Learnable Ensemble Kernel $\boldsymbol{K}(\boldsymbol{\lambda})$:
Let $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, ..., \lambda_m, ..., \lambda_M]^\top$ be a positive weight parameter vector. The ensemble kernel matrix $\boldsymbol{K}(\boldsymbol{\lambda}) \in \mathbb{R}^{|\mathcal{G}|\times|\mathcal{G}|}$ is defined as the weighted sum of M different kernels $\{K_{\mathcal{P}_1}, K_{\mathcal{P}_2}, \ldots, K_{\mathcal{P}_M}\}$. Given two graphs G_i and G_j in $\mathcal{G}$, the element at the i-th row and j-th column of $\boldsymbol{K}(\boldsymbol{\lambda})$ is given by,

$$K_{ij}(\boldsymbol{\lambda}) = \sum_{m=1}^{M} \lambda_m K_{\mathcal{P}_m}(G_i, G_j)$$

Limitations of Pattern Counting Vector: Ignoring Node Features, Time Complexity; Lacking Implicit Information and Strong Expressiveness.

Loss Function & Optimization

- **PXGL-EGK:**
  Supervised Contrastive Loss:

$$\mathcal{L}_{\mathrm{SCL}}(\boldsymbol{K}(\boldsymbol{\lambda})) = -\sum_{i\neq j} \mathbb{I}_{[\boldsymbol{y}_i=\boldsymbol{y}_j]}(\log K_{ij}(\boldsymbol{\lambda}) - \log[\sum_k \mathbb{I}_{[\boldsymbol{y}_i=\boldsymbol{y}_k, i\neq k]} K_{ik}(\boldsymbol{\lambda}) + \mu \sum_k \mathbb{I}_{[\boldsymbol{y}_i\neq\boldsymbol{y}_k]} K_{ik}(\boldsymbol{\lambda})]),$$ where $\mathbb{I}_{[\cdot]}$ is indicator function.

  Unsupervised KL Divergence:

$$\mathcal{L}_{\mathrm{KL}}(\boldsymbol{K}(\boldsymbol{\lambda})) = \mathbb{KL}(\boldsymbol{K}(\boldsymbol{\lambda}), \boldsymbol{K'}(\boldsymbol{\lambda})) \text{ with } \boldsymbol{K'}_{ij}(\boldsymbol{\lambda}) = \frac{K_{ij}^2(\boldsymbol{\lambda})/r_j}{\sum_{j'} K_{ij'}^2(\boldsymbol{\lambda})/r_{j'}}$$

where $r_j = \sum_j K_{ij}(\boldsymbol{\lambda})$ are soft cluster frequencies.
  Optimization:

$$\boldsymbol{\lambda}^* = \mathrm{argmin}_{\mathbf{1}_M^\top\boldsymbol{\lambda}=1,\ \boldsymbol{\lambda}\geq 0} \quad \mathcal{L}_{\mathrm{ker}}(\boldsymbol{\lambda}),$$

- **PXGL-GNN:**
  Supervised Loss:

$$\mathcal{L}_{\mathrm{CE}}(\boldsymbol{\lambda}, \mathcal{W}) = -\frac{1}{|G|} \sum_{G\in\mathcal{G}} \sum_{c=1}^{C} y_c \log \hat{y}_c$$

  Unsupervised Loss (KL Divergence):

$$\mathcal{L}_{\mathrm{KL}} = \mathrm{KL}(K(\boldsymbol{\lambda}, \mathcal{W}), K'(\boldsymbol{\lambda}, \mathcal{W}))$$

where $K_{ij} = \exp(-\|\mathbf{g}_i - \mathbf{g}_j\|^2/\gamma)$
  Optimization:

$$\boldsymbol{\lambda}^*, \mathcal{W}^* = \underset{\mathcal{W}, \mathbf{1}^\top\boldsymbol{\lambda}=1, \boldsymbol{\lambda}\geq 0}{\arg\min} \mathcal{L}(\boldsymbol{\lambda}, \mathcal{W})$$

The Proposed PXGL-GNN Framework

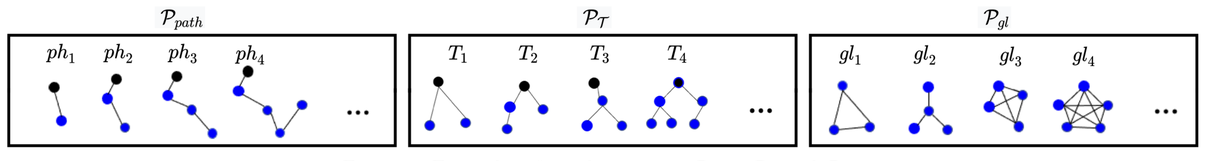


Figure 1: Examples of graph patterns: $\mathcal{P}_{\mathrm{path}}$, $\mathcal{P}_{\mathrm{T}}$ and $\mathcal{P}_{\mathrm{gl}}$

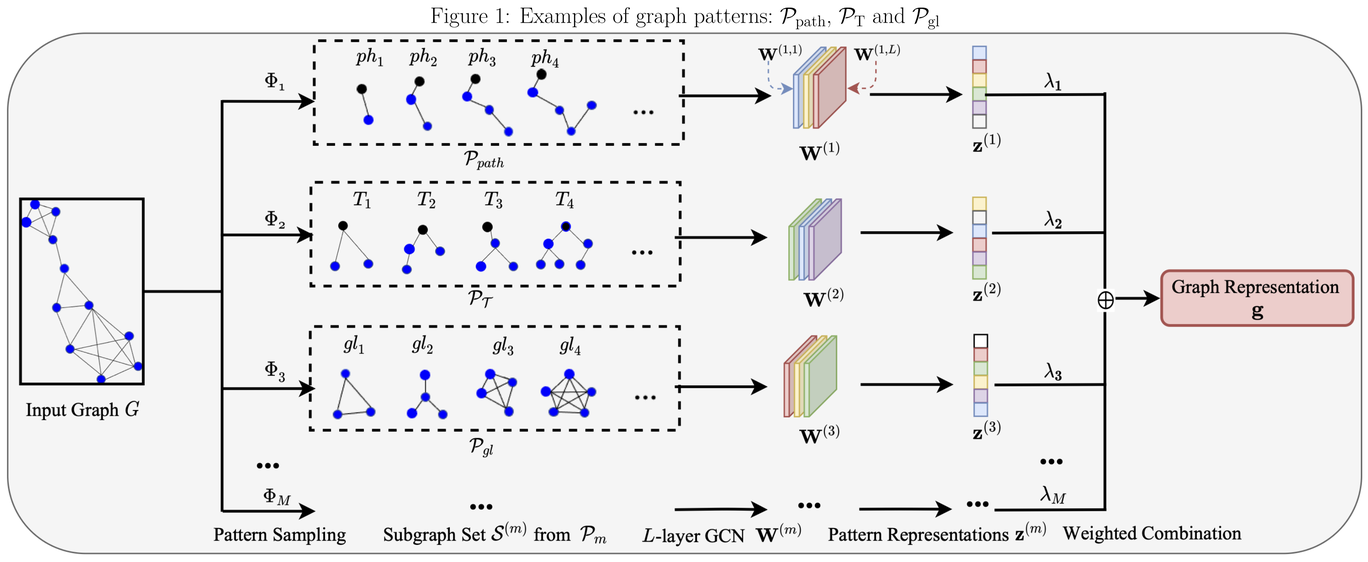


Figure 2: Framework of our proposed PXGL-GNN

Three-Stage Process:

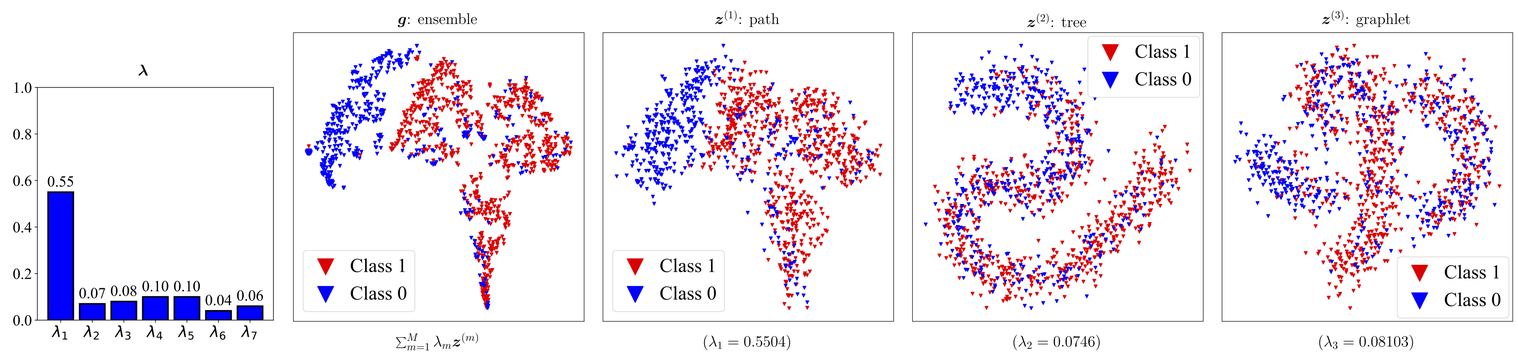
1. **Pattern Sampling:** Extract subgraphs $\mathcal{S}^{(m)}$ matching $\mathcal{P}_m$
2. **Embedding Pattern Representation:** GNN $F(\cdot; \mathcal{W}^{(m)})$ embeds sampled pattern subgraphs to representation $\mathbf{z}^{(m)}$: $\mathbf{g} = \sum_m \lambda_m \mathbf{z}^{(m)}$, where $\boldsymbol{z}^{(m)} = \frac{1}{|\mathcal{S}^{(m)}|} \sum_{S\in\mathcal{S}^{(m)}} F(\boldsymbol{A}_S, \boldsymbol{X}_S; \mathcal{W}^{(m)}), \quad \forall\, m \in [M]$ as the representation for pattern $\mathcal{P}_m$.
3. **Joint Learning of GNN and Pattern Importance Weights:** Refer to the Mathematical Framework.

Theoretical Analysis

Robustness (Theorem 5.1): For perturbations Δ_A, Δ_X, $\|\boldsymbol{A}\|_2 \leq \beta_A$, $\|\boldsymbol{X}\|_F \leq \beta_X$, $\|\boldsymbol{W}^{(m,l)}\|_2 \leq \beta_W$ for all $m \in [M]$ and $l \in [L]$, α is the minimum node degree, then $\|\tilde{\mathbf{g}} - \mathbf{g}\| \leq \frac{\rho^L \beta_W^L}{\sqrt{n}(1+\alpha)^L} \cdot (1 + \beta_A + \|\Delta_A\|_2)^{L-1} \cdot [(1 + \beta_A + 2\|\Delta_A\|_2)\|\Delta_X\|_F + 2L\beta_X(1+\beta_A)\|\Delta_D\|_2]$
Generalization (Theorem 5.2): Proof the detailed estimation error bound η for $|\ell_{\mathrm{CE}}(\boldsymbol{\lambda}_{\mathcal{D}}, \mathcal{W}_{\mathcal{D}}; G) - \ell_{\mathrm{CE}}(\boldsymbol{\lambda}_{\mathcal{D}\setminus i}, \mathcal{W}_{\mathcal{D}\setminus i}; G)| \leq \eta$
Complexity (Sec 5.3): Assume ψ_m is the time complexity of the m-th kernel, the total time complexity of PXGL-EGK is $\mathcal{O}(N^2 \sum_{m=1}^{M} \psi_m)$. Assume F_m is an L-layer GCN with hidden dim d, B is the batch size,
PXGL-GNN (supervised) space complexity: $\mathcal{O}(BMQ(e+nd) + MLd^2 + Cd)$, time complexity: $\mathcal{O}(BMQL(ed + nd^2))$;
PXGL-GNN (unsupervised) space complexity: $\mathcal{O}(BMQ(e+nd) + MLd^2 + Cd + B^2)$, time complexity: $\mathcal{O}(BMQL(ed + nd^2) + B^2)$.

Experiment Results & Pattern Importance Analysis

Example t-SNE visualizations of PXGL-GNN's pattern representations (supervised) for PROTEINS. Paths dominate around 55% importance for graph classification tasks, aligning with the biological importance of protein folding pathways.



The learned explainable pattern weights $\boldsymbol{\lambda}$ of PXGL-GNN (supervised) across all datasets:

Pattern	MUTAG	PROTEINS	DD	NCI1	COLLAB	IMDB-B	REDDIT-B	REDDIT-M5K
paths	0.095 ± 0.014	**0.550 ± 0.070**	0.093 ± 0.012	0.022 ± 0.002	**0.587 ± 0.065**	0.145 ± 0.018	0.131 ± 0.027	0.027 ± 0.003
trees	0.046 ± 0.005	0.074 ± 0.009	0.054 ± 0.006	0.063 ± 0.008	0.105 ± 0.013	0.022 ± 0.003	0.055 ± 0.007	0.025 ± 0.003
graphlets	0.062 ± 0.008	0.081 ± 0.011	0.125 ± 0.015	0.101 ± 0.013	0.063 ± 0.008	0.084 ± 0.011	0.026 ± 0.003	0.054 ± 0.007
cycles	**0.654 ± 0.085**	0.099 ± 0.013	0.094 ± 0.012	0.176 ± 0.022	0.022 ± 0.003	0.123 ± 0.016	0.039 ± 0.005	0.037 ± 0.005
cliques	0.082 ± 0.011	0.098 ± 0.012	**0.572 ± 0.073**	**0.574 ± 0.075**	0.134 ± 0.017	**0.453 ± 0.054**	0.279 ± 0.069	0.256 ± 0.067
wheels	0.026 ± 0.003	0.039 ± 0.005	0.051 ± 0.007	0.012 ± 0.002	0.068 ± 0.009	0.037 ± 0.004	0.036 ± 0.005	0.023 ± 0.003
stars	0.035 ± 0.005	0.056 ± 0.007	0.011 ± 0.002	0.052 ± 0.007	0.021 ± 0.003	0.136 ± 0.017	**0.447 ± 0.006**	**0.578 ± 0.033**

Please see the full results in our paper. In this poster, we only list several representative benchmark methods:
Supervised Graph Classification (ACC↑ %):

Method	MUTAG	PROTEINS	DD	NCI1	COLLAB	IMDB-B	REDDIT-B	REDDIT-M5K
GIN	84.53±2.38	73.38±2.16	76.38±1.58	73.36±1.78	75.83±1.29	72.52±1.62	83.27±1.30	52.48±1.57
SAGNN	93.24±2.51	75.61±2.28	84.12±1.73	81.29±1.22	79.94±1.83	74.53±2.57	89.57±2.13	54.11±1.22
ICL	91.34±2.19	75.44±1.26	82.77±1.42	83.45±1.78	81.45±1.21	73.29±1.46	90.13±1.40	56.21±1.35
PXGL-GNN	**94.87±2.26**	**78.23±2.46**	**86.54±1.95**	**85.78±2.07**	**83.96±1.59**	**77.35±2.32**	**91.84±1.69**	**57.36±2.14**

Unsupervised Graph Clustering Performance (ACC↑ %, NMI↑ %):

Method	Metric	MUTAG	PROTEINS	DD	NCI1	COLLAB	IMDB-B	REDDIT-B	REDDIT-M5K
InfoGraph	ACC	72.9±2.1	71.6±1.9	54.9±3.5	53.5±1.2	59.7±2.0	62.4±1.6	58.2±2.3	59.7±1.9
	NMI	23.6±0.5	23.1±0.3	26.6±0.4	26.3±0.5	31.1±0.8	19.8±0.5	20.6±0.6	28.6±0.6
GraphACL	ACC	74.2±2.3	73.1±2.7	57.2±2.7	52.2±1.3	55.4±1.3	67.9±1.3	59.4±1.4	56.7±2.3
	NMI	34.7±0.7	27.4±0.8	31.2±0.3	26.0±0.7	23.6±0.6	31.5±0.7	21.5±0.6	23.8±0.9
PXGL-GNN	**ACC**	**77.8±2.9**	**74.6±1.9**	**57.6±3.5**	**56.4±1.3**	**61.2±1.4**	**68.6±2.7**	**61.6±1.7**	**60.8±2.3**
	NMI	**35.2±0.6**	**29.2±1.0**	**31.7±0.3**	**32.7±0.8**	**37.2±0.7**	**32.4±1.1**	**22.4±0.9**	**29.5±1.2**