Reconstruction Enhanced Multi-View Contrastive Learning for Anomaly Detection on Attributed Networks

Jiaqiang Zhang[1,2] , **Senzhang Wang**[3] and **Songcan Chen**[1,2*]
[1]College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics
[2]MIIT Key Laboratory of Pattern Analysis and Machine Intelligence
[3]Central South University
{zhangjq, s.chen}@nuaa.edu.cn, szwang@csu.edu.cn

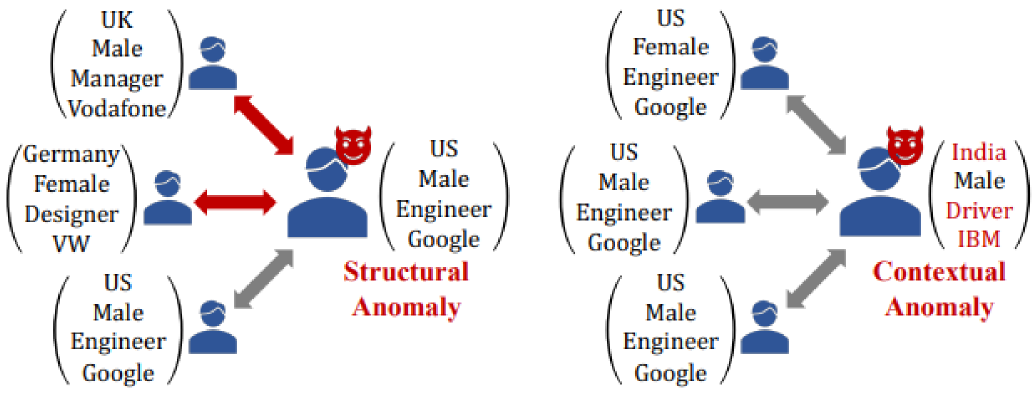
IJCAI 2022

Reported by Nengqiang Xiang

Introduction

Existing approach adopted one-layer GCN to extract the local structure information from the one-hop neighbors, while the global structure information beyond one-hop neighbors cannot be effectively captured and utilized.

This paper proposes a self-supervised learning framework that jointly optimizes a multi-view contrastive learning-based module and an attribute reconstruction-based module to more accurately detect anomalies on attributed networks.



(a) Structural Anomaly (b) Contextual Anomaly

Method

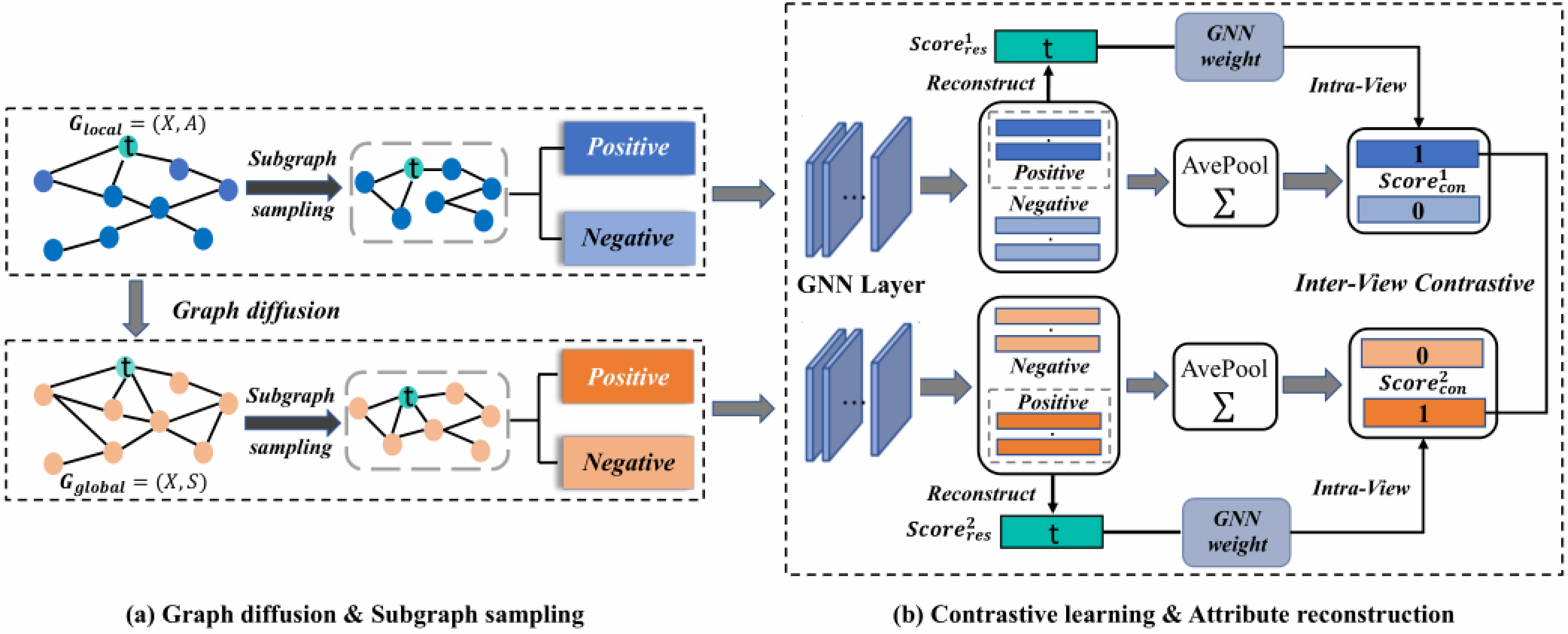
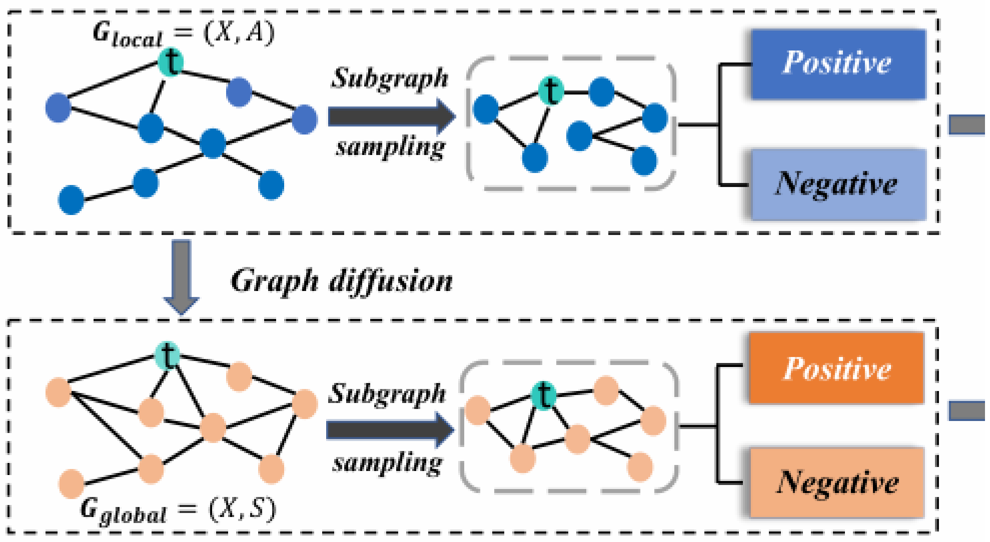


Figure 1: The framework of the proposed model.

Method



(a) Graph diffusion & Subgraph sampling

Symbol Definition:

$\mathcal{G}=(V, \mathrm{E}, \mathbf{X})$: attributed network

$V=(v_1, \cdots v_N)$

$\mathbf{X} \in \mathbb{R}^{N\times F}$: attributed matrix

$\boldsymbol{x}_i \in \mathbb{R}^F$

$\boldsymbol{A} \in \mathbb{R}^{N\times N}$: adjacency matrix

Graph Diffusion：

$$\mathbf{S} = \sum_{k=0}^{\infty} \theta_k \mathbf{T}^k \in \mathbb{R}^{N\times N} \tag{1}$$

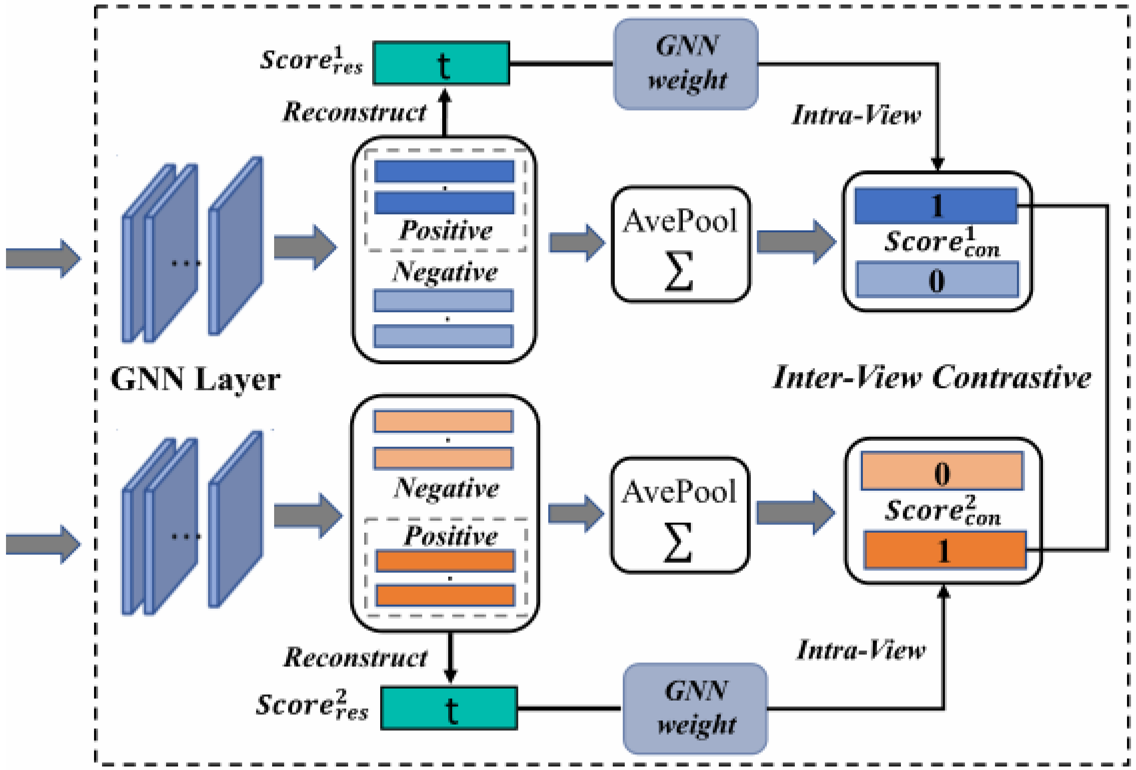
$$\mathbf{T} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2} \quad \theta_k = \alpha(1-\alpha)^k$$

$$\mathbf{S} = \alpha(\mathbf{I} - (1-\alpha)\mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2})^{-1} \tag{2}$$

Subgraph Sampling：

Subgraph Sampling. We adopt random walk with restart (RWR) [Tong *et al.*, 2006] to sample the local subgraph. In both views, a subgraph with the size P is sampled for each node. In particular, we sample nodes and their edges from the original view to form the local-view subgraph, and then select the exact nodes and edges from the other view to form the global-view subgraph. In each view, for node v_i, its subgraph is regarded as its positive pair, and a subgraph corresponding to the other node is regarded as the negative pair.

Method



(b) Contrastive learning & Attribute reconstruction

Intra-View Contrastive Learning:

$$\mathbf{H}_i^l = \phi(\hat{\mathbf{D}}_i^{-1/2}\mathbf{A}_i'\hat{\mathbf{D}}_i^{-1/2}\mathbf{H}_i^{(l-1)}\mathbf{W}^{(l-1)}) \qquad (3)$$

$$\hat{\mathbf{D}}_i \in \mathbb{R}^{N\times N} \qquad \mathbf{A}_i' = \mathbf{A}_i + \mathbf{I}_N$$

$$\mathbf{h}_i^l = \phi(\mathbf{h}_i^{l-1}\mathbf{W}^{l-1}) \qquad (4)$$

$$h_i^0 = \mathbf{x}_i$$

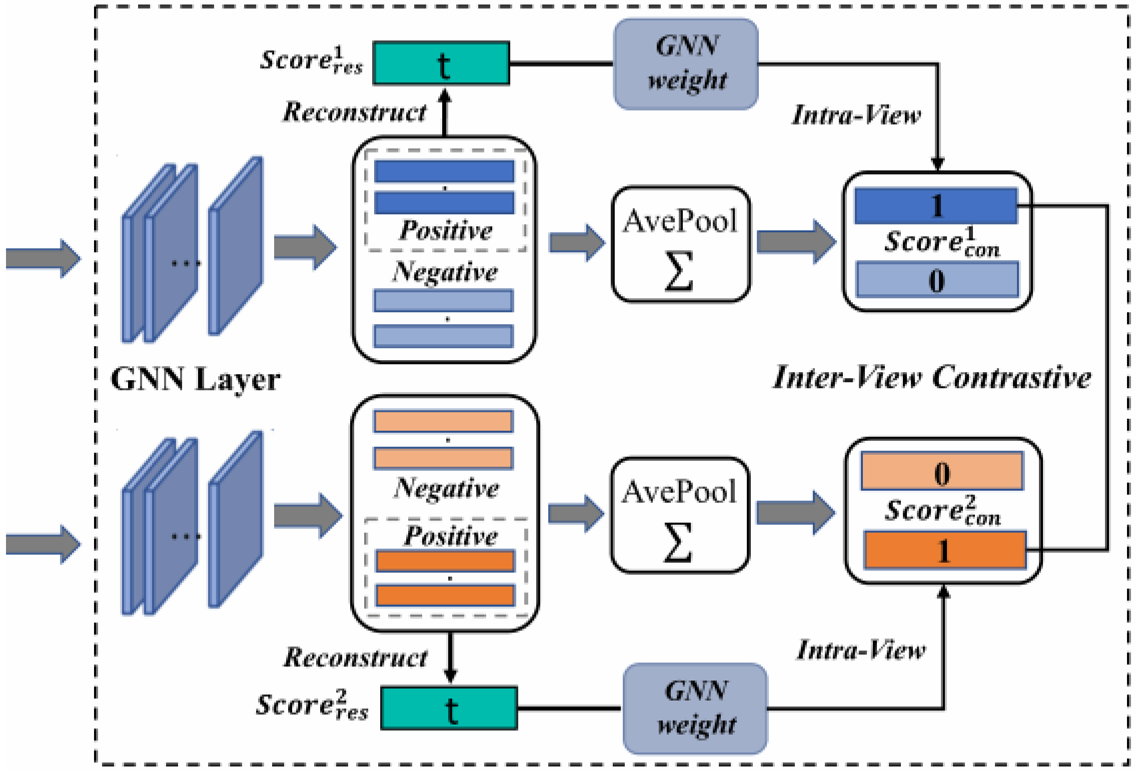
$$\mathbf{e}_i = Readout(\mathbf{H}_i) = \sum_{k=1}^{n_i}\frac{(\mathbf{H}_i)_k}{n_i} \qquad (5)$$

$$s_i = \sigma(\mathbf{h}_i\mathbf{W}_s\mathbf{e}_i^T) \qquad (6)$$

$$\mathcal{L}^1_{intra}(v_i) = -\frac{1}{2}(log(s_i) + log(1-\tilde{s}_i)) \qquad (7)$$

$$\mathcal{L}_{intra} = \frac{1}{2N}\sum_{i=1}^{N}(\mathcal{L}^1_{intra}(v_i) + \mathcal{L}^2_{intra}(v_i)) \qquad (8)$$

Method



(b) Contrastive learning & Attribute reconstruction

Inter-View Contrastive Learning:

$$\mathcal{L}_{inter} = (||\mathbf{s}_1 - \mathbf{s}_2||_F^2) \quad (9)$$

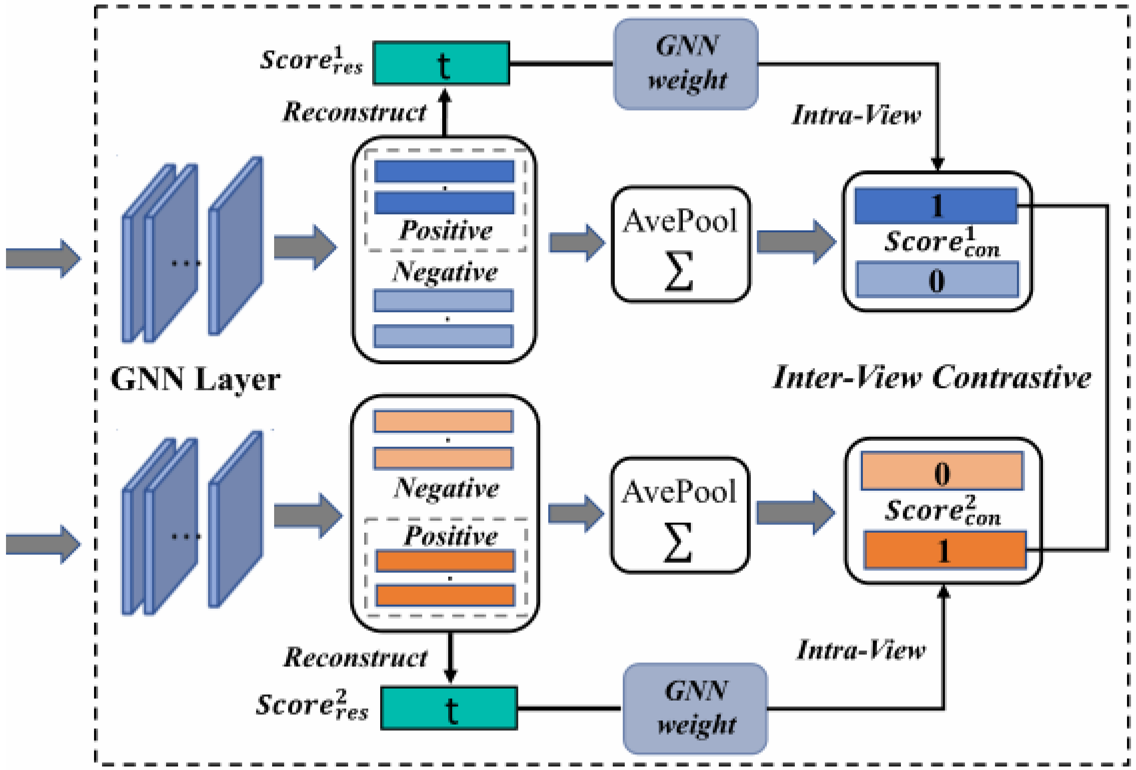
$$\mathcal{L}_{con} = \mathcal{L}_{inter} + \mathcal{L}_{intra} \quad (10)$$

Attribute Reconstruction Based on Neighbors:

$$\mathcal{L}^1_{res}(v_i) = ||g(\mathbf{Z}_i) - x_i||^2 \quad (11)$$

$$\mathcal{L}_{res} = \frac{1}{2N}\sum_{i=1}^{N}(\mathcal{L}^1_{res}(v_i) + \mathcal{L}^2_{res}(v_i)) \quad (12)$$

Method



(b) Contrastive learning & Attribute reconstruction

Anomaly Score Inference:

$$\mathcal{L} = \mathcal{L}_{con} + \gamma \mathcal{L}_{res} \tag{13}$$

$$score_{con}(v_i) = \frac{1}{2}[score^1_{con}(v_i) + score^2_{con}(v_i)] \tag{14}$$

$$score^1_{con}(v_i) = s_i^{1-} - s_i^{1+}$$

$$score^2_{con}(v_i) = s_i^{2-} - s_i^{2+}$$

$$score_{res}(v_i) = \frac{1}{2}\sum_{k=1}^{2} score^k_{res}(v_i) \tag{15}$$

$$score^k_{res}(v_i) = ||g(\mathbf{Z}_i^k) - x_i||_2^2, k = 1, 2 \tag{16}$$

$$score(v_i) = score_{con}(v_i) + \gamma score_{res}(v_i) \tag{17}$$

Experiments

Datasets	Nodes	Edges	Features	Anomalies
BlogCatalog	5196	171743	8189	300
Flickr	7575	239738	12407	450
Cora	2708	5429	1433	150
CiteSeer	3327	4732	3703	150
Pubmed	19717	44338	500	600

Table 1: The statistics of the datasets.

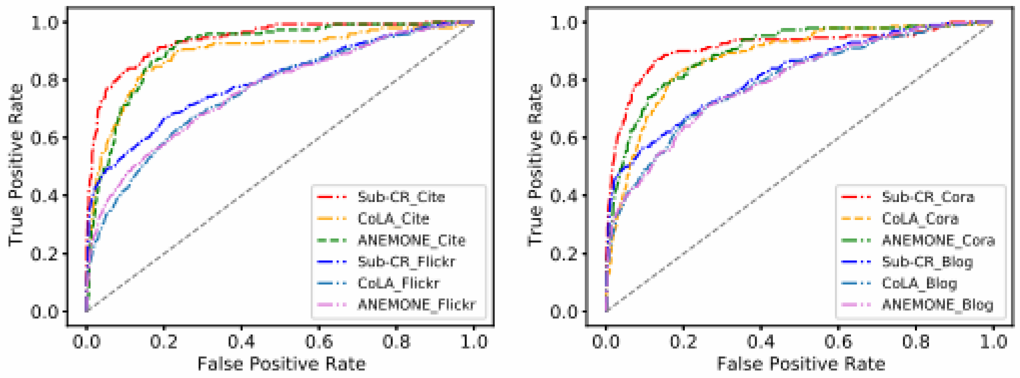


Figure 2: ROC curves on four benchmark datasets.

Experiments

Methods	Blogcatalog	Flickr	Cora	Citeseer	Pubmed
AMEN[2016]	0.6392	0.6573	0.6266	0.6154	0.7713
Radar[2017]	0.7401	0.7399	0.6587	0.6709	0.6233
ANOMALOUS [2018]	0.7237	0.7434	0.5770	0.6307	0.7316
DOMINANT[2019]	0.7468	0.7442	0.8155	0.8251	0.8081
DGI [2019]	0.5827	0.6237	0.7511	0.8293	0.6962
CoLA[2021]	0.7854	0.7513	0.8779	0.8968	0.9512
ANEMONE [2021]	0.8067	0.7637	0.9057	0.9189	0.9548
Sub-CR	**0.8141**	**0.7975**	**0.9132**	**0.9303**	**0.9709**

Table 2: The AUC values comparison on five benchmark datasets.

	BlogCatalog	Flickr	Cora	CiteSeer	Pubmed
Sub-R	0.7943	0.7609	0.9002	0.9017	0.9553
Sub-C	0.7460	0.7434	0.8220	0.7892	0.8006
Sub-weight	0.8083	0.7928	0.9041	0.9275	0.9491
Sub-global	0.8090	0.7923	0.8975	0.9195	0.9625

Table 3: The AUC values of ablation study.

Experiments

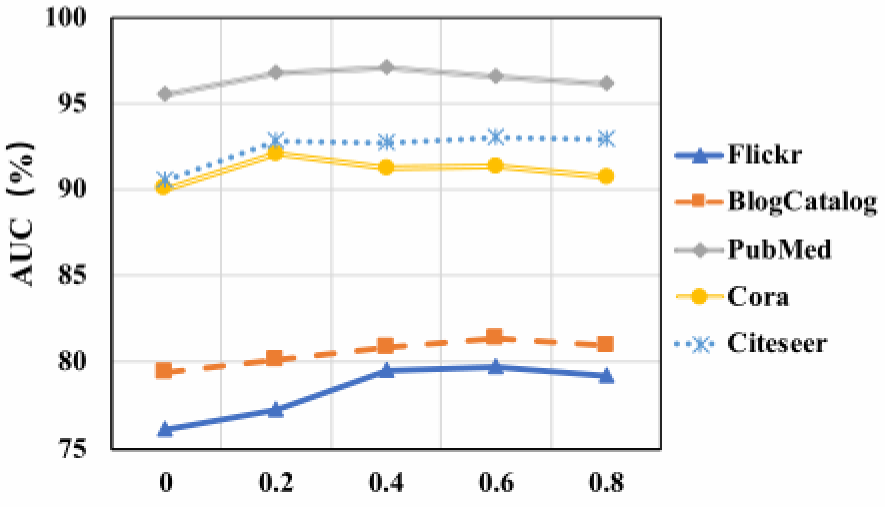


Figure 3: Performance with different γ.

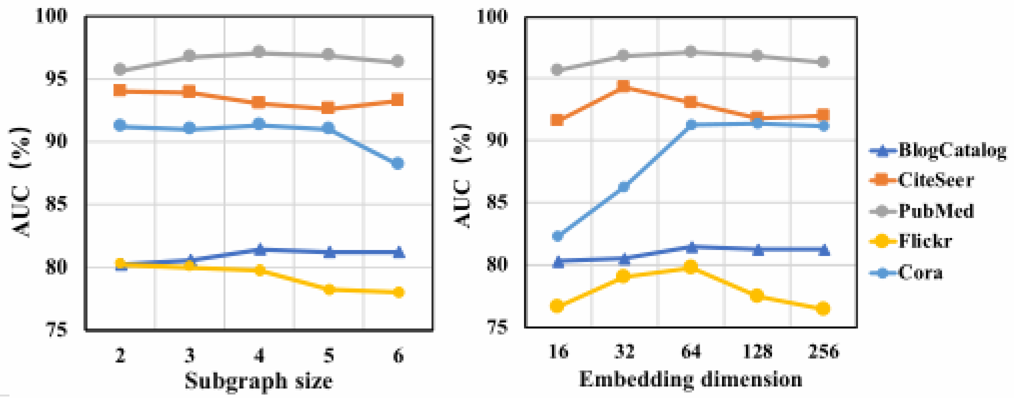


Figure 4: Performance with different subgraph sizes and embedding dimensions.

Thanks