AnomalyDAE: Dual Autoencoder for Anomaly Detection on Attributed Networks

Haoyi Fan [1], Fengbin Zhang [1], Zuoyong Li [2]

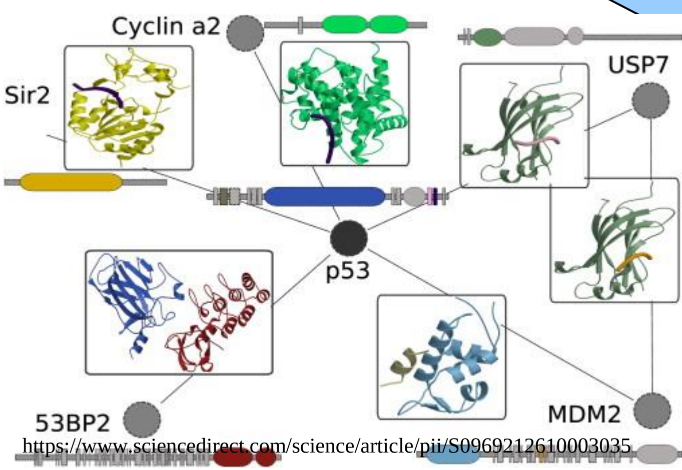
Harbin University of Science and Technology [1]
Minjiang University [2]
isfanhy@hrbust.edu.cn

Background

World Wide Web

Social Network

Networks are ubiquitous !

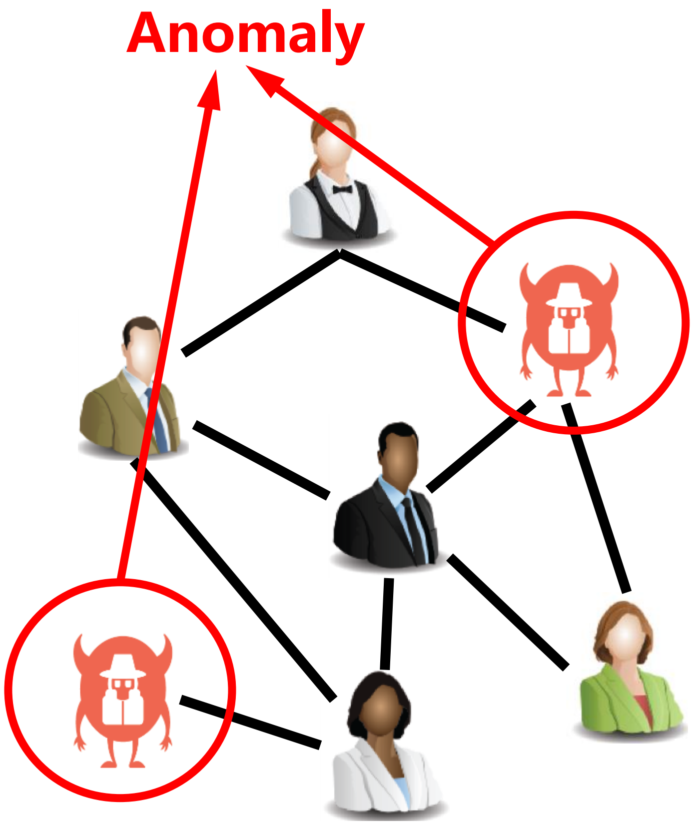


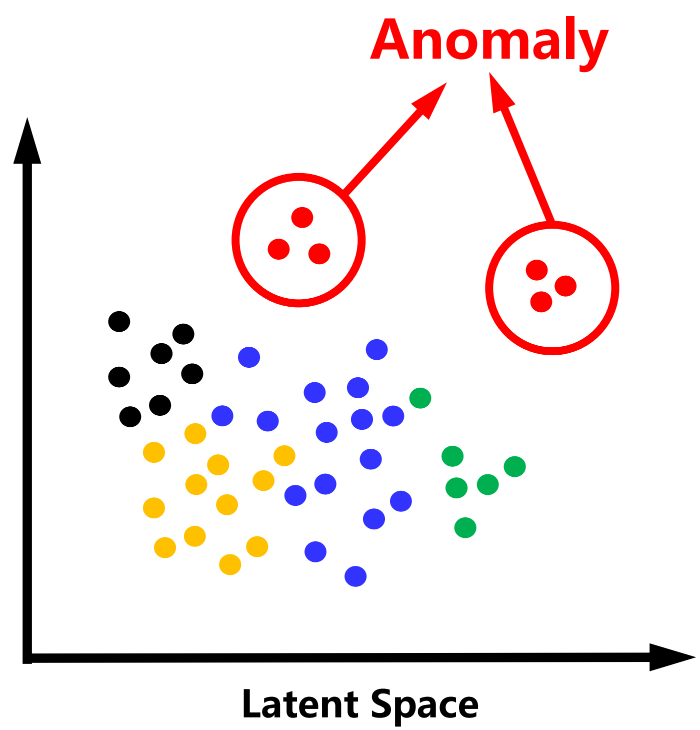
Biology Network

Finance Transaction Network

Background

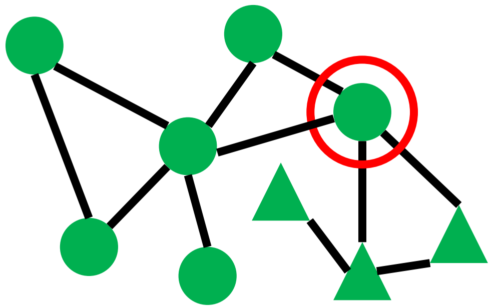
Anomaly Detection on the Attributed Network



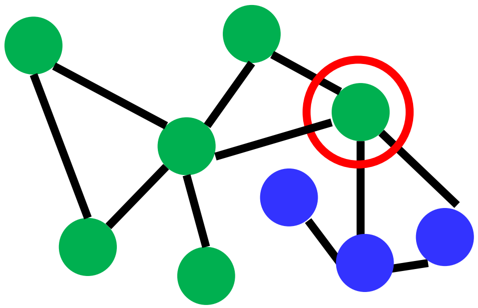


Background

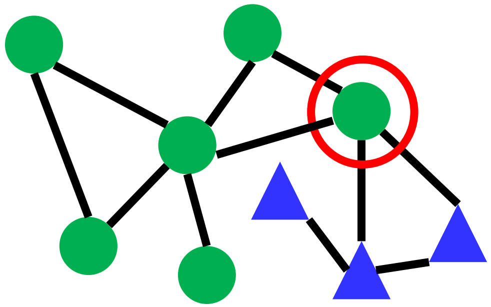
Different types of anomalies



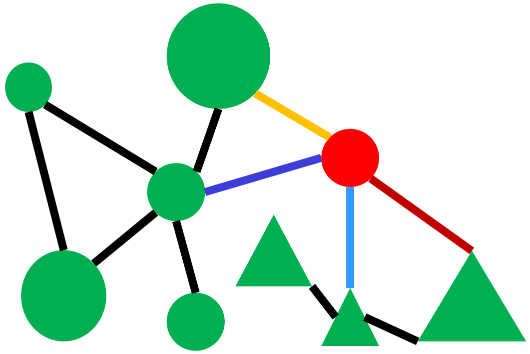
Structure-inconsistent
Attribute-consistent



Structure-consistent
Attribute-inconsistent



Structure-inconsistent
Attribute-inconsistent



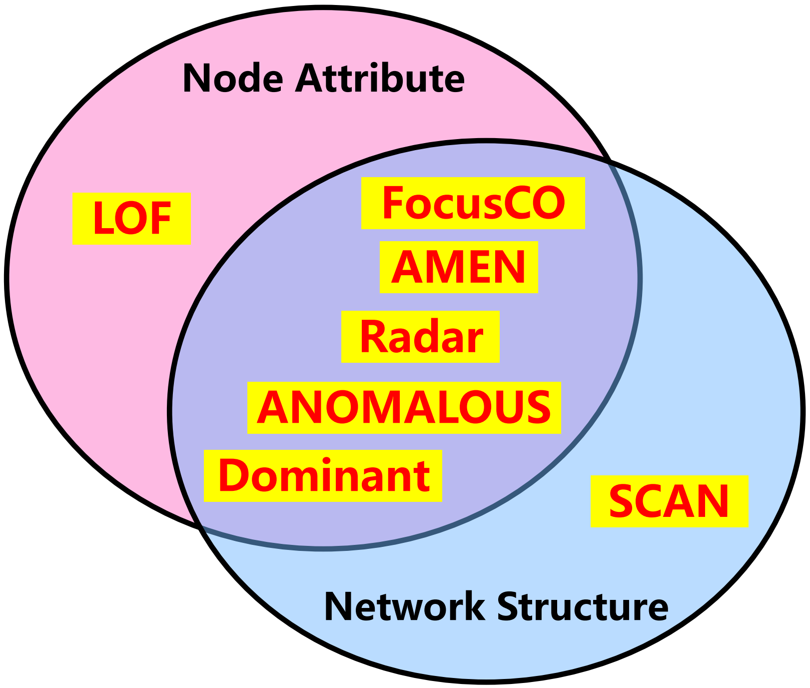
Different neighbors contribute differently for anomaly detection

- **Challenges:**
  - **The cross-modality interactions between the network structure and node attribute**
  - **Neighbor-attention aware anomaly measuring**

Background

Numerous attributed network based anomaly detection methods have been proposed…

LOF Breunig et al. 2000
SCAN Xu et al. 2007
FocusCO Perozzi et al. 2014
AMEN Perozzi et al. 2016
Radar Li et al. 2017
ANOMALOUS Peng et al. 2018
Dominant Ding et al. 2019
…



- **Deep representation learning framework on graph?**
- **The cross-modality interactions between the network structure and node attribute?**

Problem Statement

Problem

Given $\mathcal{G} = \{\boldsymbol{\mathcal{V}}, \boldsymbol{\mathcal{E}}, \mathbf{X}\}$, learn a score function $f: \boldsymbol{\mathcal{V}}_i \mapsto y_i \in \mathbb{R}$, to classify sample x_i based on the threshold λ :

$$y_i = \begin{cases} 1, & if\ f(\boldsymbol{\mathcal{V}}_i) \geq \lambda, \\ 0, & otherwise. \end{cases}$$

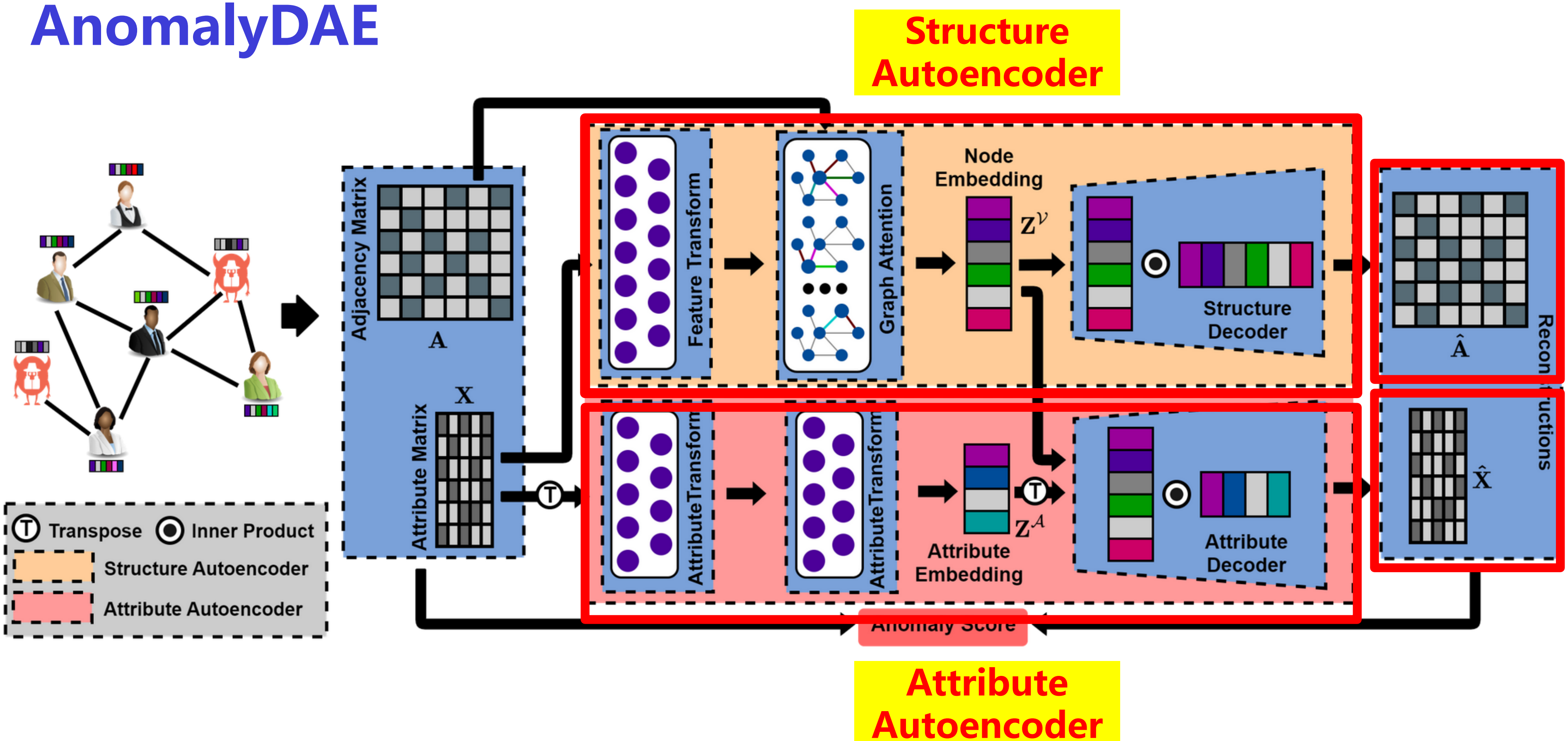
where y_i denotes the label of sample x_i, with 0 being the normal class and 1 the anomalous class.

Notations

$\mathcal{G}$: Attributed network
$\boldsymbol{\mathcal{V}}$: Set of nodes in network.
$\boldsymbol{\mathcal{E}}$: Set of edges in network.
M : Number of nodes.
N : Dimension of attribute.
$\mathbf{A} \in \mathbb{R}^{M\times M}$: Adjacency matrix of a network.
$\mathbf{X} \in \mathbb{R}^{M\times N}$: Attribute matrix of all nodes.

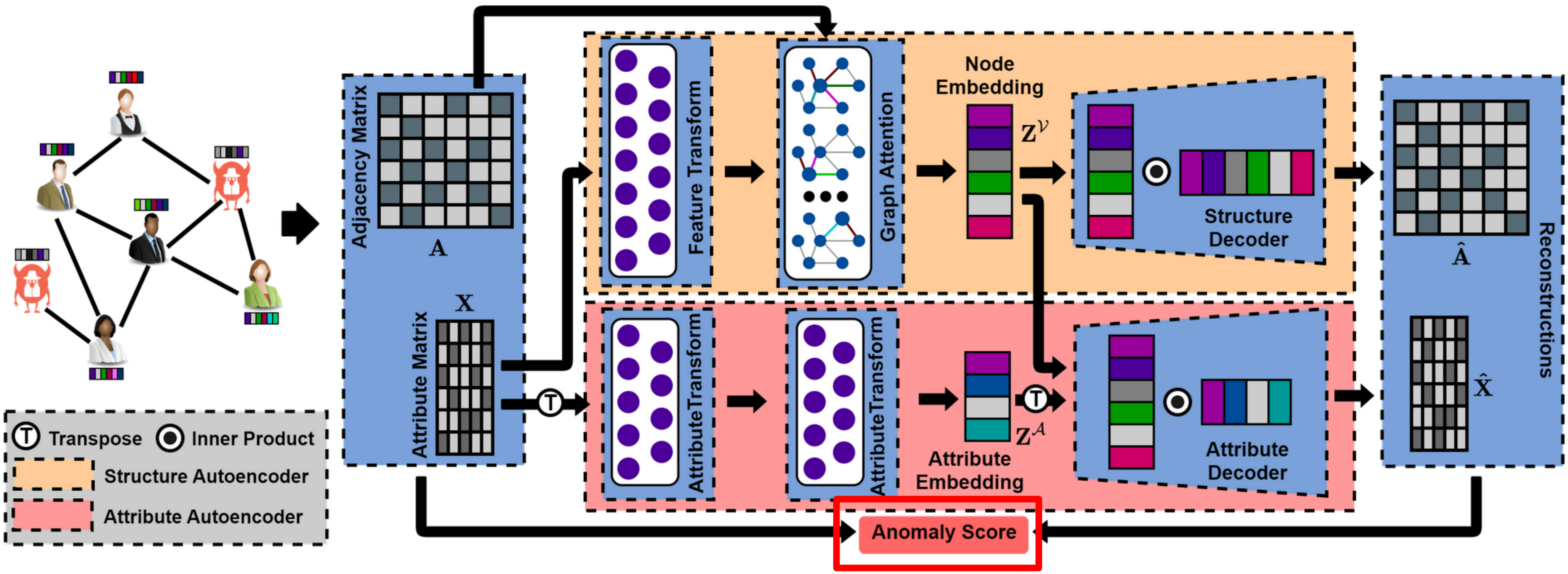
Method

AnomalyDAE



Method

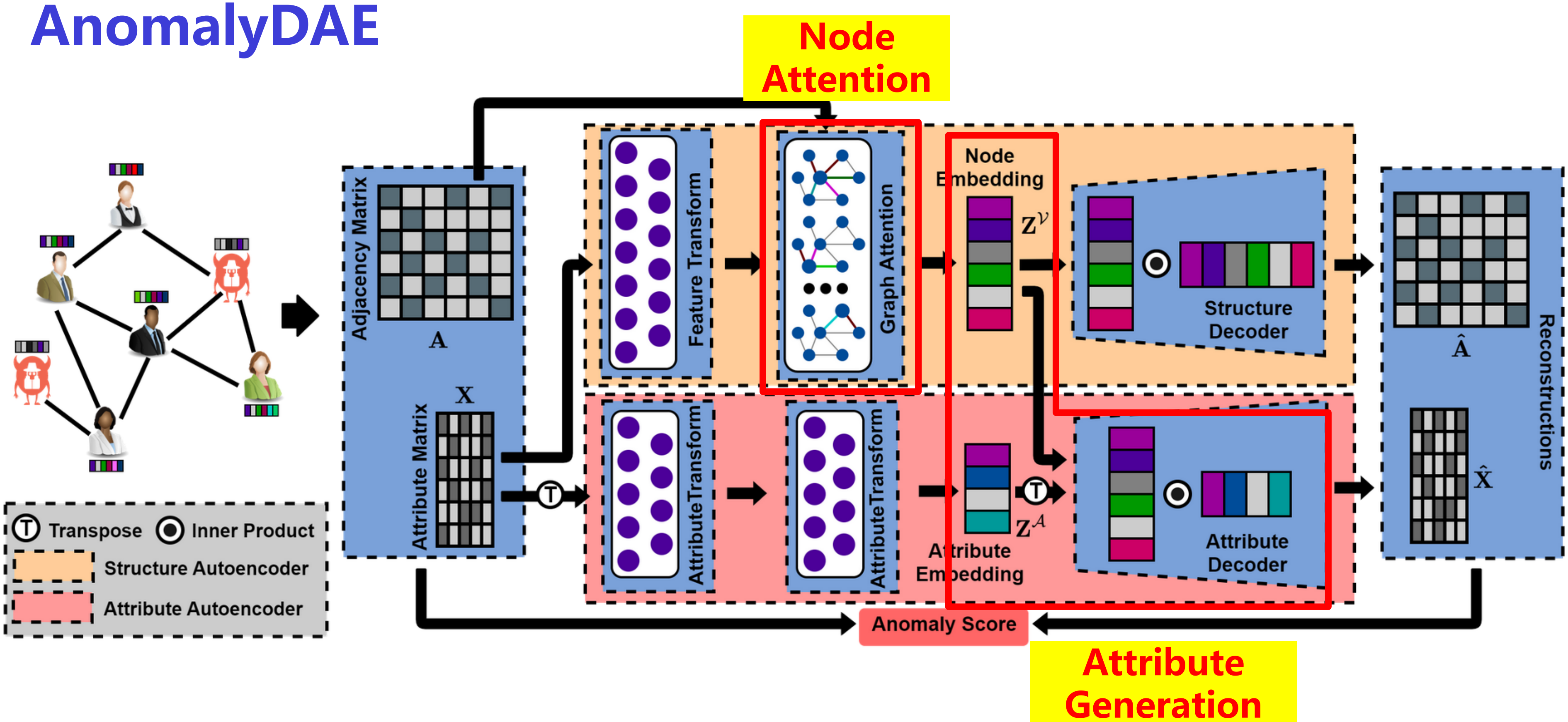
AnomalyDAE



Structure-level and attribute-level anomaly score

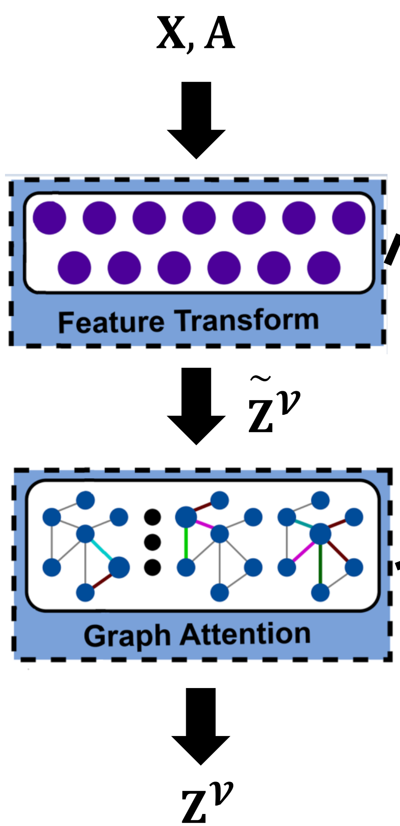
Method

AnomalyDAE



Method

Neighbor-attention Mechanism in Structure Autoencoder



Initial feature transformation:

$$\tilde{\mathbf{Z}}^{\mathcal{V}} = \sigma(\mathbf{X}\mathbf{W}^{\mathcal{V}(1)} + \mathbf{b}^{\mathcal{V}(1)})$$

Importance weights:

$$e_{i,j} = attn\left(\tilde{\mathbf{Z}}_i^{\mathcal{V}}, \tilde{\mathbf{Z}}_j^{\mathcal{V}}\right) = \sigma(\mathrm{a}^{\mathrm{T}} \cdot [\mathbf{W}^{\mathcal{V}(2)}\tilde{\mathbf{Z}}_i^{\mathcal{V}} || \mathbf{W}^{\mathcal{V}(2)}\tilde{\mathbf{Z}}_j^{\mathcal{V}}])$$

Normalization:

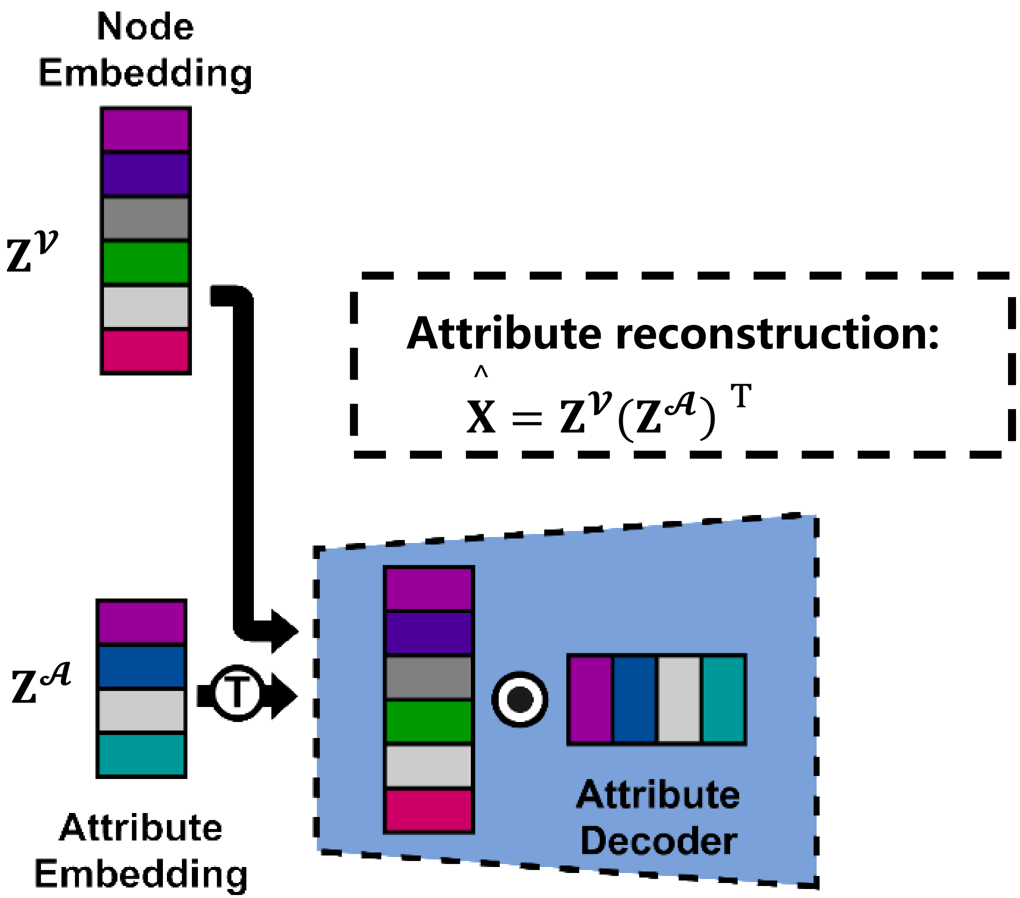
$$\gamma_{i,j} = \frac{\exp(e_{i,j})}{\sum_{k \in \mathcal{N}_i} \exp\left(e_{i,k}\right)}$$

Neighbor-attention aware feature aggregation:

$$\mathbf{Z}_i^{\mathcal{V}} = \sum_{k \in \mathcal{N}_i} \gamma_{i,k} \cdot \tilde{\mathbf{Z}}_k^{\mathcal{V}}$$

Method

Cross-modality Interactions Capturing in Attribute Autoencoder



Method

Loss and Anomaly Score

Loss Function:

$$\mathcal{L}_{rec} = \alpha||(\mathbf{A} - \hat{\mathbf{A}}) \odot \boldsymbol{\theta}||_F^2 + (1-\alpha)||(\mathbf{X} - \hat{\mathbf{X}}) \odot \boldsymbol{\eta}||_F^2$$

$$\boldsymbol{\theta}_{i,j} = \begin{cases} 1 & if\ \mathbf{A}_{i,j} = 0, \\ \theta & otherwise. \end{cases}, \boldsymbol{\eta}_{i,j} = \begin{cases} 1 & if\ \mathbf{X}_{i,j} = 0, \\ \eta & otherwise. \end{cases}$$

Anomaly Score:

$$Score = \alpha\boxed{||(\mathbf{A} - \hat{\mathbf{A}}) \odot \boldsymbol{\theta}||_F^2} + (1-\alpha)\boxed{||(\mathbf{X} - \hat{\mathbf{X}}) \odot \boldsymbol{\eta}||_F^2}$$

Structure-level Anomaly Measure

Attribute-level Anomaly Measure

Method

Loss and Anomaly Score

Loss Function:

$$\mathcal{L}_{rec} = \alpha||(\mathbf{A} - \hat{\mathbf{A}}) \odot \boldsymbol{\theta}||_F^2 + (1-\alpha)||(\mathbf{X} - \hat{\mathbf{X}}) \odot \boldsymbol{\eta}||_F^2$$

$$\boldsymbol{\theta}_{i,j} = \begin{cases} 1 & if\ \mathbf{A}_{i,j} = 0, \\ \theta & otherwise. \end{cases}, \boldsymbol{\eta}_{i,j} = \begin{cases} 1 & if\ \mathbf{X}_{i,j} = 0, \\ \eta & otherwise. \end{cases}$$

Anomaly Score:

$$Score = \alpha||(\mathbf{A} - \hat{\mathbf{A}}) \odot \boldsymbol{\theta}||_F^2 + (1-\alpha)||(\mathbf{X} - \hat{\mathbf{X}}) \odot \boldsymbol{\eta}||_F^2$$

Solution for Problem:

$$y_i = \begin{cases} 1, & if\ f(\boldsymbol{\mathcal{V}}_i) \geq \lambda, \\ 0, & otherwise. \end{cases}$$

$$\lambda = Distribution(Score)$$

Experiment

Datasets

Table 2. Statistics of the used Real-World datasets.

Database	# $\mathcal{V}$	# $\mathcal{E}$	# $\mathcal{A}$	# Anomalies
BlogCatalog	5,196	171,743	8,189	300
Flickr	7,575	239,738	12,047	450
ACM	16,484	71,980	8,337	600

Baselines

LOF Breunig et al. 2000
SCAN Xu et al. 2007
AMEN Perozzi et al. 2016
Radar Li et al. 2017
ANOMALOUS Peng et al. 2018
Dominant Ding et al. 2019
...

Evaluation Metric

AUC (**A**rea **U**nder a receiver operating characteristic **C**urve)

Experiment

Results

Table 3. AUC scores of all methods on three datasets.

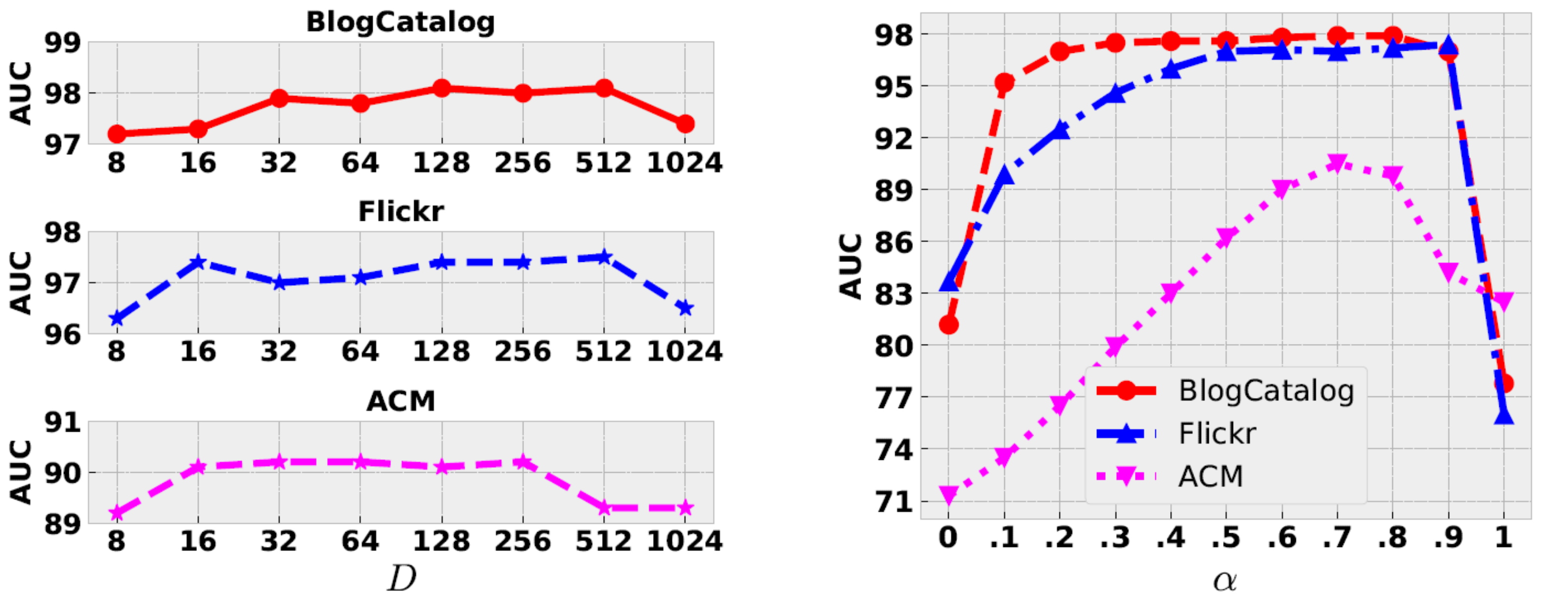
Method	BlogCatalog	Flickr	ACM
LOF [18]	49.15	48.81	47.38
SCAN [19]	27.27	26.86	35.99
AMEN [8]	53.37	60.47	72.62
Radar [12]	71.04	72.86	69.36
Anomalous [13]	72.81	71.59	[illegible]
Dominant [14]	78.13	74.9	74.94
AnomalyDAE	**97.81**	**97.22**	**90.05**

19.68% 22.32% 15.11%

At least 15.11% ~ 22.32% AUC improvement!

Experiment

Results



(a) Embedding dimension (b) Parameter α

Robust and Effective!

Conclusion

- **Traditional machine learning based methods perform poor for feature learning on large graph.**
- **Traditional deep graph model cannot effectively capture the cross-modality interactions between the network structure and node attribute.**
- **We propose a deep joint representation learning framework via a dual autoencoder to capture the complex cross-modality interactions between the network structure and node attribute.**

Reference

- **[LOF]** Breunig, Markus M., et al. "LOF: identifying density-based local outliers." **KDD. 2000**.
- **[SCAN]** Xu, Xiaowei, et al. "Scan: a structural clustering algorithm for networks." **KDD. 2007**.
- **[FocusCO]** Perozzi, Bryan, et al. "Focused clustering and outlier detection in large attributed graphs." **KDD. 2014**.
- **[AMEN]** Perozzi, Bryan, and Leman Akoglu. "Scalable anomaly ranking of attributed neighborhoods." **SIAM, 2016**.
- **[Radar]** Li, Jundong, et al. "Radar: Residual Analysis for Anomaly Detection in Attributed Networks." **IJCAI. 2017**.
- **[GAT]** Veličković, Petar, et al. "Graph attention networks." **ICLR. 2018**.
- **[ANOMALOUS]** Peng, Zhen, et al. "ANOMALOUS: A Joint Modeling Approach for Anomaly Detection on Attributed Networks." **IJCAI. 2018**.
- **[Dominant]** Ding, Kaize, et al. "Deep anomaly detection on attributed networks." **SIAM, 2019**.

Thanks

Thanks for listening!

Contact: isfanhy@hrbust.edu.cn
Home Page: https://haoyfan.github.io/