# Truncated Affinity Maximization: One-class Homophily Modeling for Graph Anomaly Detection

**Hezhe Qiao, Guansong Pang**

Singapore Management University

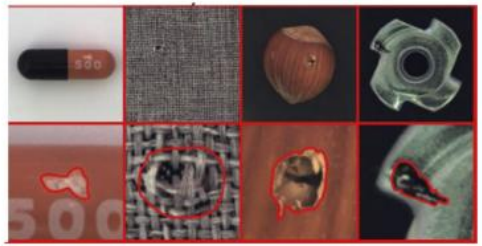
NeurIPS2023@New Orleans, USA

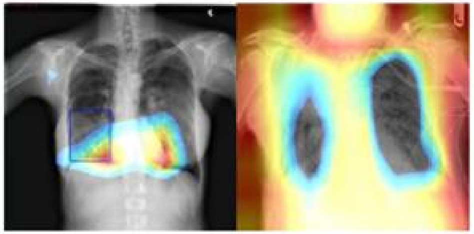
12 Decemeber, 2023

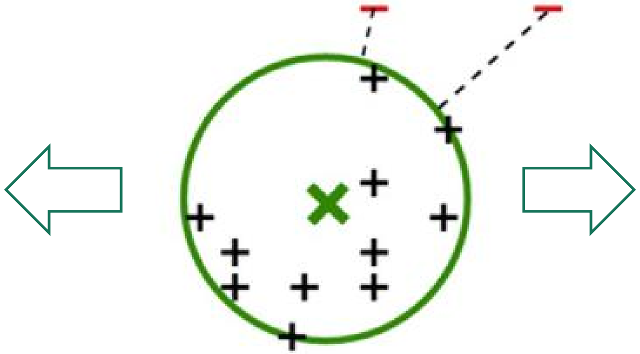
# Background

## What are anomalies?

Anomalies (outliers, novelties): Samples that are signficantly different from the most of the data

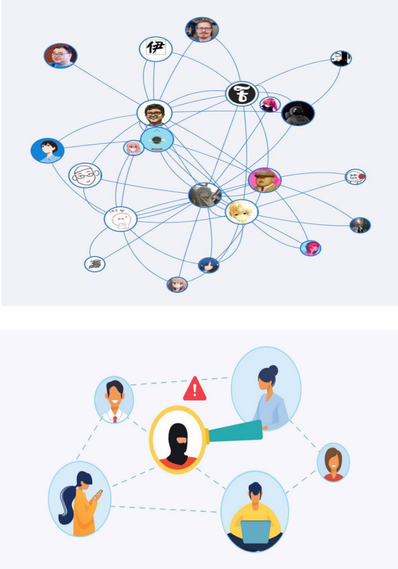




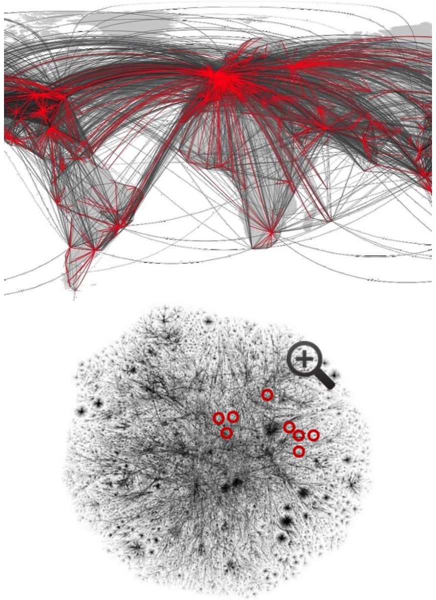


# Background

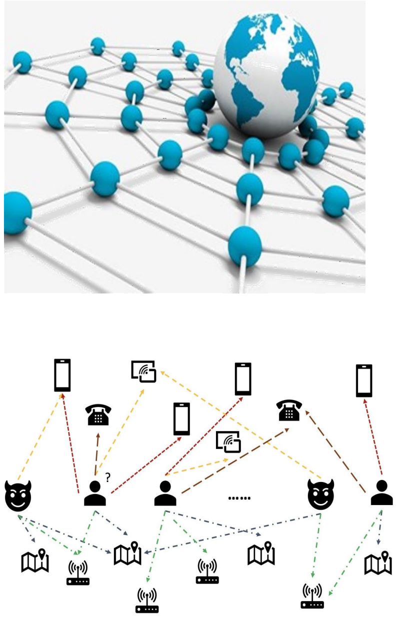
Graph Anomaly Detection (GAD), which aims to identify rare observations in graphs, has attracted more attention nowadays



**Social network**



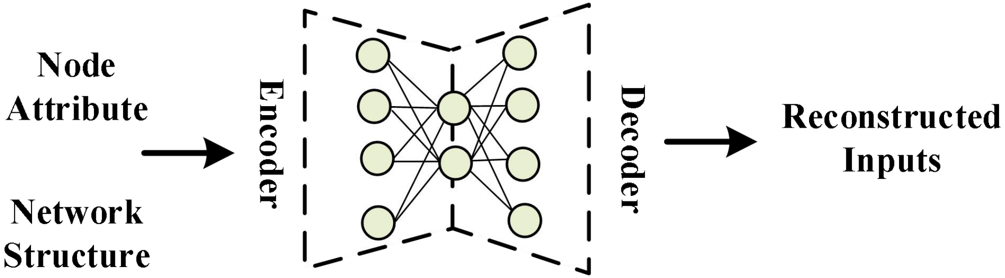
**Bank transaction**



**Web client-server**

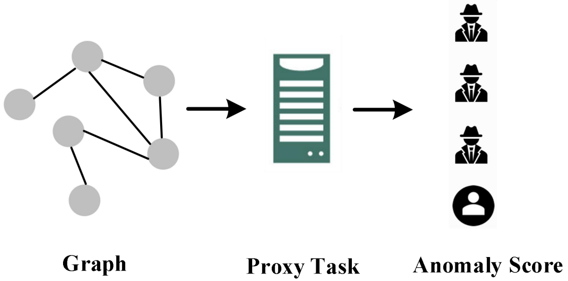
# Related Work

- **Data Reconstruction-based Approach**
  - DOMINATE
  - AnomalyDAE
  - ComGA



- **Self-supervised Approach**
  - CoLA
  - SL-GAD
  - HCM-A

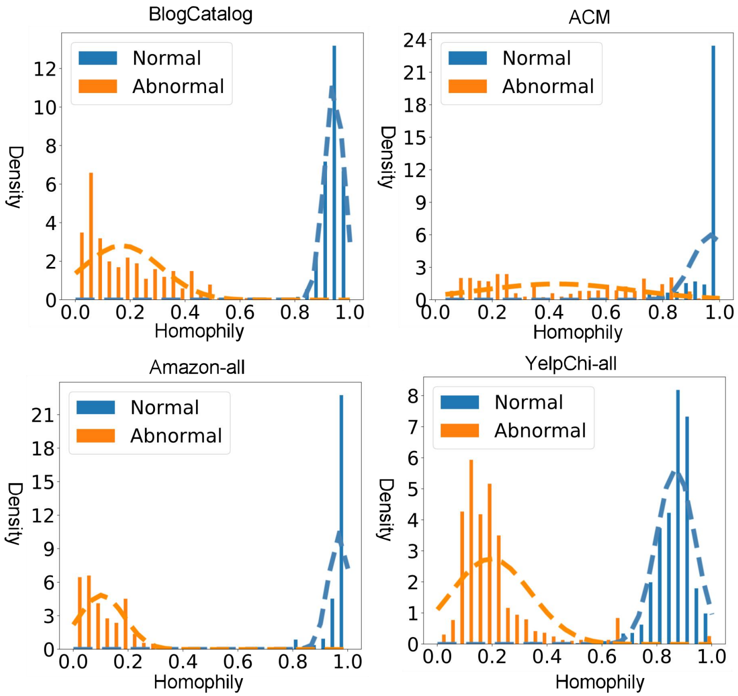
Proxy classification or contrastive learning task



# Motivation(1/2)

## One-class homophily

Normal nodes tend to have strong connection/affinity with each other, while the homophily in abnormal nodes is significantly weaker than normal nodes

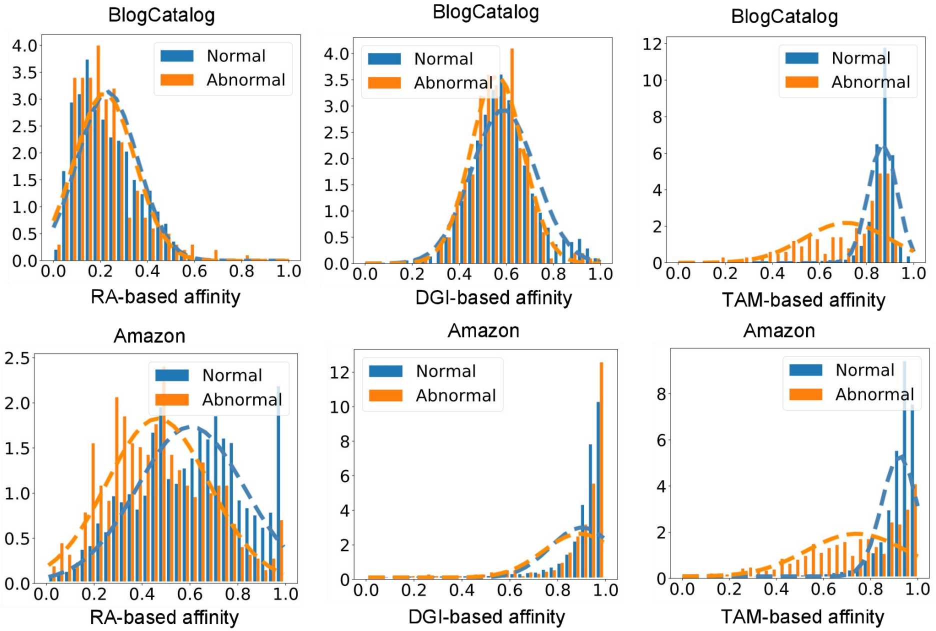


This anomaly-discriminative property is ignored by existing GAD such as data reconstruction, self-supervised training.

# Motivation(2/2)

Optimizing on the original graph structure can be biased by non-homophily edges (i.e., edges connecting normal and abnormal nodes).

- Irrelevant attributes and camouflage
- Normal and abnormal nodes are oversmoothing



# Problem Statement

- Purpose

Given a graph $G=(\mathcal{V},\mathcal{E},\mathbf{X})$ learn a score function $f:\mathcal{V}\mapsto y_i \in \mathbb{R}$

$$y_i = \begin{cases} 1 & f(v_i) \geq \lambda \\ 0 & otherwise \end{cases}$$

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

- Evaluation AUROC, AUPRC

# Local Node Affinity as Anomaly Measure

## One-class homophily

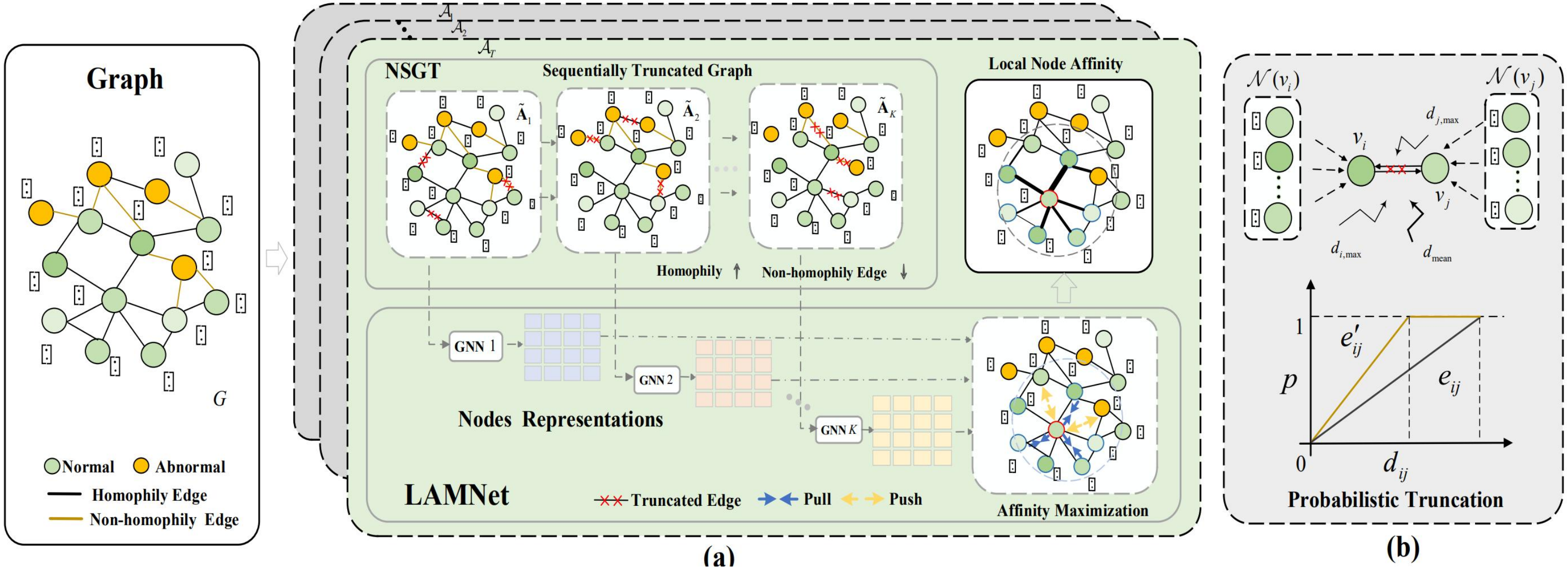
$$f_{TAM}\left(v_i;\Theta,\mathbf{A},\mathbf{X}\right) = -\frac{1}{\left|\mathcal{N}\left(v_i\right)\right|}\sum_{v_j\in\mathcal{N}\left(v_i\right)} \mathrm{sim}\left(\mathbf{h}_i,\mathbf{h}_j\right) \qquad \mathrm{sim}\left(\mathbf{x}_i,\mathbf{x}_j\right)=\frac{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j}{\left\|\mathbf{x}_i\right\|\left\|\mathbf{x}_j\right\|}$$

A new perspective to quantify the **normality/abnormality of nodes**, enabling a much simpler anomaly scoring than existing popular measures such as the **reconstruction error**

$$f\left(v_i\right) = (1-\alpha)\left\|\mathbf{a}_i-\hat{\mathbf{a}}_i\right\|_2+\alpha\left\|\mathbf{x}_i-\hat{\mathbf{x}}_i\right\|_2$$

# Methodolgy

## Our TAM Framework



☐ **Local Affinity Maximization on Graph**

☐ **Normal Structure-preserved Graph Truncation**

# Local Affinity Maximization Networks (LAMNet)

- Obtain the embedding of nodes

$$\mathbf{H}_i^{(l)} = GNN(\mathbf{A}, \mathbf{H}_i^{(l-1)}; \mathbf{W}^{(l-1)}) \qquad \mathbf{H}^{(\ell)} = \phi\left(\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}\mathbf{H}^{(\ell-1)}\mathbf{W}^{(\ell-1)}\right)$$

**Affinity maxmization objective function**

Maximizing the local affinity

Regularization term

$$\min_{\Theta} \sum_{v_i \in \mathcal{V}} \left( \boxed{f_{TAM}(v_i; \Theta, \mathbf{A}, \mathbf{X})} + \boxed{\lambda \frac{1}{\left|\mathcal{V} \setminus \mathcal{N}(v_i)\right|} \sum_{v_k \in \mathcal{V} \setminus \mathcal{N}(v_i)} \mathrm{sim}(\mathbf{h}_i, \mathbf{h}_k)} \right)$$

**The optimization can be largely biased by non-homophily edges**

# Normal Structure-preserved Graph Truncation (NSGT)

NSGT takes a probabilistic approach and performs the graph truncation

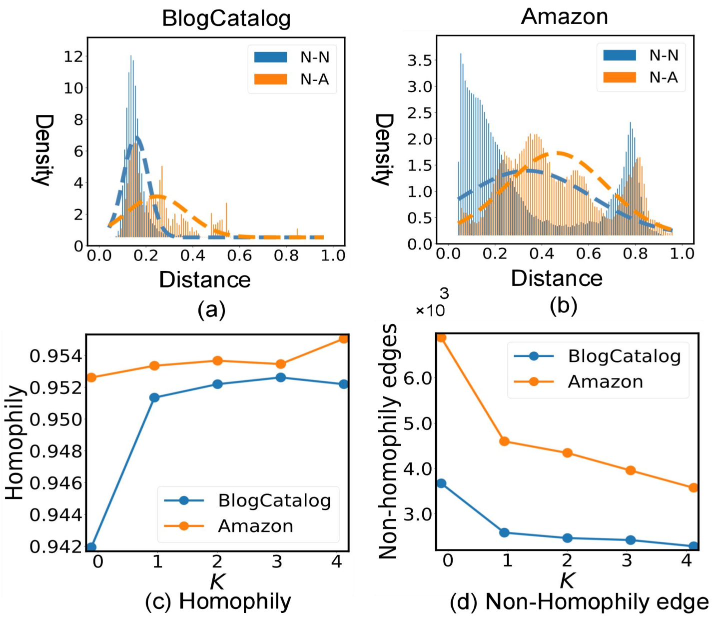
$$r_i = rand(d_{\text{mean}}, d_{\text{i,max}})$$

$$e_{ij} \leftarrow 0 \text{ iff } d_{ij} > r_i \text{ and } d_{ij} > r_j, \forall e_{ij} = 1$$

$$p\left(r_i < d_{ij}\right) = \frac{\max\left(d_{ij} - d_{\text{mean}}, 0\right)}{d_{i,\max} - d_{\text{mean}}}$$

It is considered as a non-homophily edge and removed if and only if the distance between node $v_i$ and $v_j$

$$p\left(\mathcal{E} \setminus e_{ij}\right) = p\left(r_i < d_{ij}\right) p\left(r_j < d_{ij}\right)$$



**(a)** and **(b)** are respectively the Euclidean distance statistic N-N and N-A **(c)** Homophily of normal nodes vs. **(d)** the number of non-homophily edges with increasing truncation iterations/depths

# Training

NSGT performs a sequentially iterative truncation rather than a single-pass truncation.

**How do we know which step is optimal ?**

We train a LAMNet on each of the $K$ sequentially truncated adjacency matrices in $\boldsymbol{A}$, resulting in K LAMNets parameterized by $\{\Theta_1, \Theta_2, \cdots, \Theta_K\}$ for various truncation depths.

$$\mathcal{E}_{i+1} \subset \mathcal{E}_i \qquad \mathcal{A} = \left\{ \widetilde{\mathbf{A}}_1, \widetilde{\mathbf{A}}_2, \cdots, \widetilde{\mathbf{A}}_K \right\} \qquad \mathbf{H}^{(\ell)} = \phi\left( \mathbf{D}^{-\frac{1}{2}} \widetilde{\mathbf{A}}_k \mathbf{D}^{-\frac{1}{2}} \mathbf{H}^{(\ell-1)} \mathbf{W}^{(\ell-1)} \right)$$

We perform NSGT $T$ times independently, resulting in $T$ sets of the truncated adjacency matrix set

$T \times K$ LAMNets

# Inference - Anomaly Scoring

Local node affinity-based anomaly score

$$\text{score}\left(v_i\right) = \frac{1}{T \times K} \sum_{t=1}^{T} \sum_{k=1}^{K} f_{TAM} \boxed{\left(v_i ; \Theta^{*}_{t,k}, \mathbf{A}, \mathbf{X}\right)}$$

Local affinity score calculation based on the original structure

The weaker the local node affinity in the representation under various graph truncation scales, the larger the anomaly score the node $v_i$ has

# Experiments

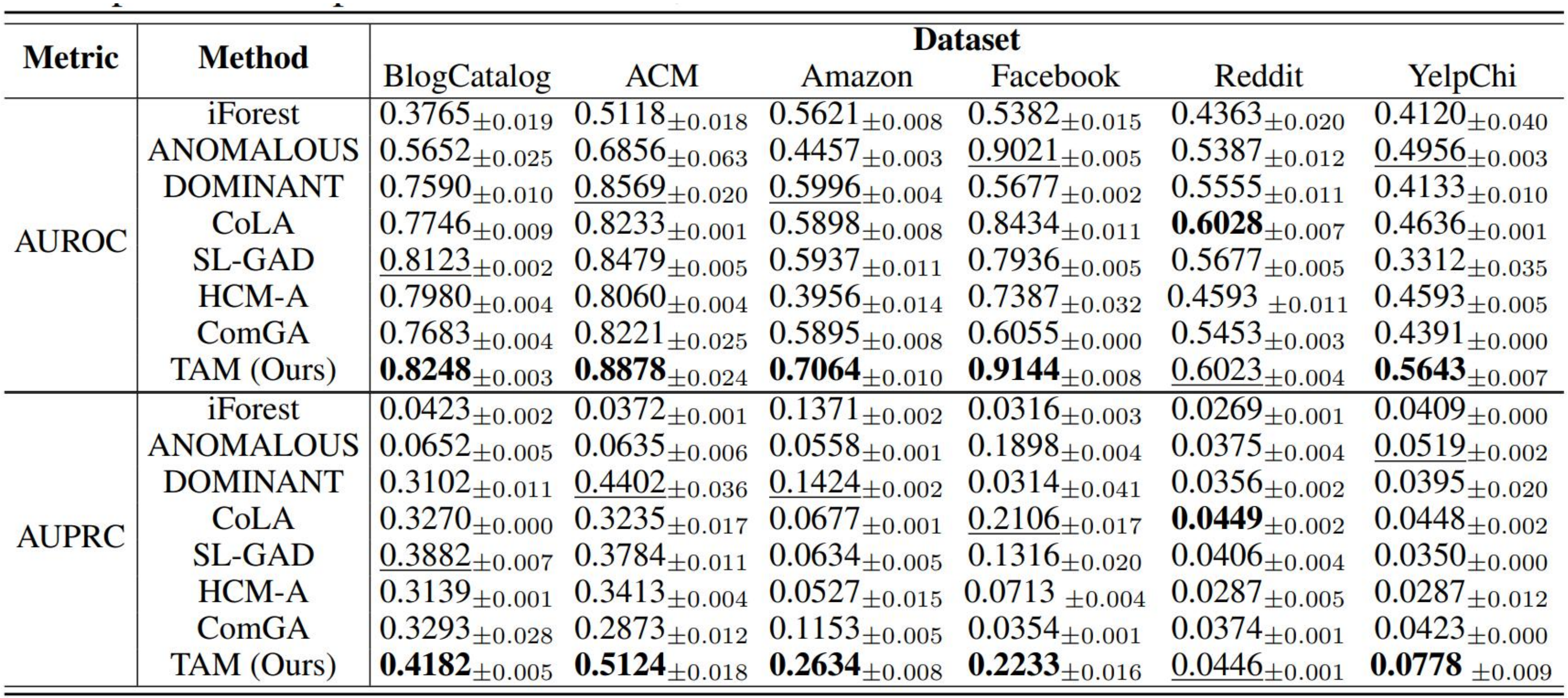
## Dataset

Real-world and large-scale GAD datasets

| Datasets | Type | R\I | Node | Edge | Attribute | Anomaly |
|---|---|---|---|---|---|---|
| BlogCatalog | Soical Network | I | 5,196 | 171,743 | 8,189 | 5.77% |
| ACM | Citation Network | I | 16,484 | 71,980 | 8,337 | 3.67% |
| Amazon | Co-review | R | 11,944 | 175,608 | 25 | 6.66% |
| Facebook | Co-review | R | 4,039 | 88,234 | 576 | 0.67% |
| Reddit | Social Network | R | 10,984 | 175,608 | 64 | 3.33% |
| YelpChi | Transaction Record | R | 45,954 | 49,315 | 32 | 2.65% |
| Amazon-all | Co-review | R | 11,944 | 4,398,392 | 25 | 9.50% |
| YelpChi-all | Transaction Record | R | 45,954 | 3,846,979 | 32 | 14.50% |
| T-Finance | Transaction Record | R | 39,357 | 21,222,543 | 10 | 4.6% |
| OGB-protein | Biology Network | I | 10,984 | 168,016 | 64 | 3.3% |

# Performance Evaluation : Comparing to the state-of-the-art

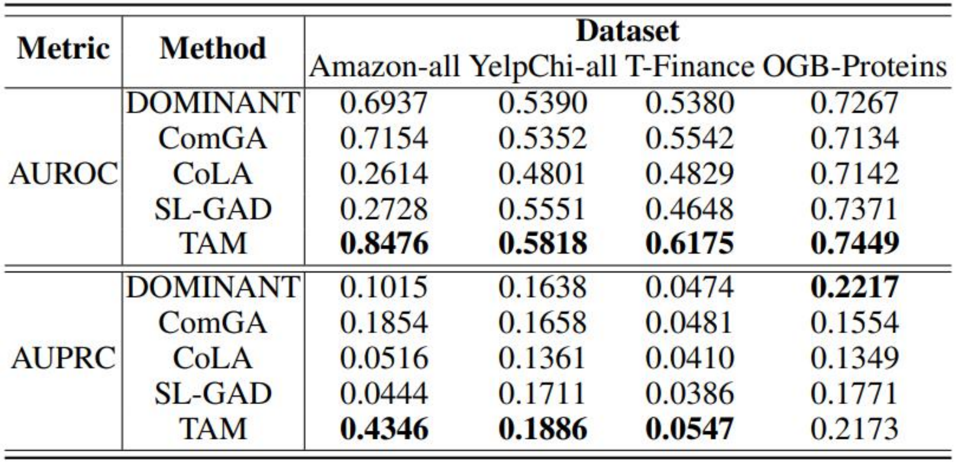
Table 1: AUROC and AUPRC results on six real-world GAD datasets with injected/real anomalies.



| Metric | Method | Dataset | | | | | |
|---|---|---|---|---|---|---|---|
| | | BlogCatalog | ACM | Amazon | Facebook | Reddit | YelpChi |
| AUROC | iForest | 0.3765$_{\pm0.019}$ | 0.5118$_{\pm0.018}$ | 0.5621$_{\pm0.008}$ | 0.5382$_{\pm0.015}$ | 0.4363$_{\pm0.020}$ | 0.4120$_{\pm0.040}$ |
| | ANOMALOUS | 0.5652$_{\pm0.025}$ | 0.6856$_{\pm0.063}$ | 0.4457$_{\pm0.003}$ | 0.9021$_{\pm0.005}$ | 0.5387$_{\pm0.012}$ | 0.4956$_{\pm0.003}$ |
| | DOMINANT | 0.7590$_{\pm0.010}$ | 0.8569$_{\pm0.020}$ | 0.5996$_{\pm0.004}$ | 0.5677$_{\pm0.002}$ | 0.5555$_{\pm0.011}$ | 0.4133$_{\pm0.010}$ |
| | CoLA | 0.7746$_{\pm0.009}$ | 0.8233$_{\pm0.001}$ | 0.5898$_{\pm0.008}$ | 0.8434$_{\pm0.011}$ | **0.6028**$_{\pm0.007}$ | 0.4636$_{\pm0.001}$ |
| | SL-GAD | 0.8123$_{\pm0.002}$ | 0.8479$_{\pm0.005}$ | 0.5937$_{\pm0.011}$ | 0.7936$_{\pm0.005}$ | 0.5677$_{\pm0.005}$ | 0.3312$_{\pm0.035}$ |
| | HCM-A | 0.7980$_{\pm0.004}$ | 0.8060$_{\pm0.004}$ | 0.3956$_{\pm0.014}$ | 0.7387$_{\pm0.032}$ | 0.4593 $_{\pm0.011}$ | 0.4593$_{\pm0.005}$ |
| | ComGA | 0.7683$_{\pm0.004}$ | 0.8221$_{\pm0.025}$ | 0.5895$_{\pm0.008}$ | 0.6055$_{\pm0.000}$ | 0.5453$_{\pm0.003}$ | 0.4391$_{\pm0.000}$ |
| | TAM (Ours) | **0.8248**$_{\pm0.003}$ | **0.8878**$_{\pm0.024}$ | **0.7064**$_{\pm0.010}$ | **0.9144**$_{\pm0.008}$ | 0.6023$_{\pm0.004}$ | **0.5643**$_{\pm0.007}$ |
| AUPRC | iForest | 0.0423$_{\pm0.002}$ | 0.0372$_{\pm0.001}$ | 0.1371$_{\pm0.002}$ | 0.0316$_{\pm0.003}$ | 0.0269$_{\pm0.001}$ | 0.0409$_{\pm0.000}$ |
| | ANOMALOUS | 0.0652$_{\pm0.005}$ | 0.0635$_{\pm0.006}$ | 0.0558$_{\pm0.001}$ | 0.1898$_{\pm0.004}$ | 0.0375$_{\pm0.004}$ | 0.0519$_{\pm0.002}$ |
| | DOMINANT | 0.3102$_{\pm0.011}$ | 0.4402$_{\pm0.036}$ | 0.1424$_{\pm0.002}$ | 0.0314$_{\pm0.041}$ | 0.0356$_{\pm0.002}$ | 0.0395$_{\pm0.020}$ |
| | CoLA | 0.3270$_{\pm0.000}$ | 0.3235$_{\pm0.017}$ | 0.0677$_{\pm0.001}$ | 0.2106$_{\pm0.017}$ | **0.0449**$_{\pm0.002}$ | 0.0448$_{\pm0.002}$ |
| | SL-GAD | 0.3882$_{\pm0.007}$ | 0.3784$_{\pm0.011}$ | 0.0634$_{\pm0.005}$ | 0.1316$_{\pm0.020}$ | 0.0406$_{\pm0.004}$ | 0.0350$_{\pm0.000}$ |
| | HCM-A | 0.3139$_{\pm0.001}$ | 0.3413$_{\pm0.004}$ | 0.0527$_{\pm0.015}$ | 0.0713 $_{\pm0.004}$ | 0.0287$_{\pm0.005}$ | 0.0287$_{\pm0.012}$ |
| | ComGA | 0.3293$_{\pm0.028}$ | 0.2873$_{\pm0.012}$ | 0.1153$_{\pm0.005}$ | 0.0354$_{\pm0.001}$ | 0.0374$_{\pm0.001}$ | 0.0423$_{\pm0.000}$ |
| | TAM (Ours) | **0.4182**$_{\pm0.005}$ | **0.5124**$_{\pm0.018}$ | **0.2634**$_{\pm0.008}$ | **0.2233**$_{\pm0.016}$ | 0.0446$_{\pm0.001}$ | **0.0778** $_{\pm0.009}$ |

# Performance Evaluation: Large-scale graph

Table2 : Results on large-scale graphs



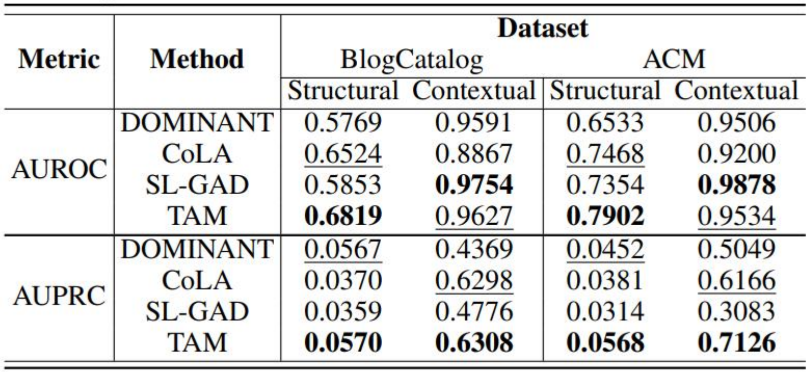
| Metric | Method | Dataset | | | |
|---|---|---|---|---|---|
| | | Amazon-all | YelpChi-all | T-Finance | OGB-Proteins |
| AUROC | DOMINANT | 0.6937 | 0.5390 | 0.5380 | 0.7267 |
| | ComGA | 0.7154 | 0.5352 | 0.5542 | 0.7134 |
| | CoLA | 0.2614 | 0.4801 | 0.4829 | 0.7142 |
| | SL-GAD | 0.2728 | 0.5551 | 0.4648 | 0.7371 |
| | TAM | **0.8476** | **0.5818** | **0.6175** | **0.7449** |
| AUPRC | DOMINANT | 0.1015 | 0.1638 | 0.0474 | **0.2217** |
| | ComGA | 0.1854 | 0.1658 | 0.0481 | 0.1554 |
| | CoLA | 0.0516 | 0.1361 | 0.0410 | 0.1349 |
| | SL-GAD | 0.0444 | 0.1711 | 0.0386 | 0.1771 |
| | TAM | **0.4346** | **0.1886** | **0.0547** | 0.2173 |

TAM can perform consistently well on these large-scale datasets and outperforms four comparing methods including reconstruction-based and contrastive-based methods

It further shows **anomaly-discriminative property** can be well utilized effectively in TAM.

# Performance Evaluation: Structural and contexual anomaly

Table 3: AUROC and AUPRC results of detecting structural and contextual anomalies.



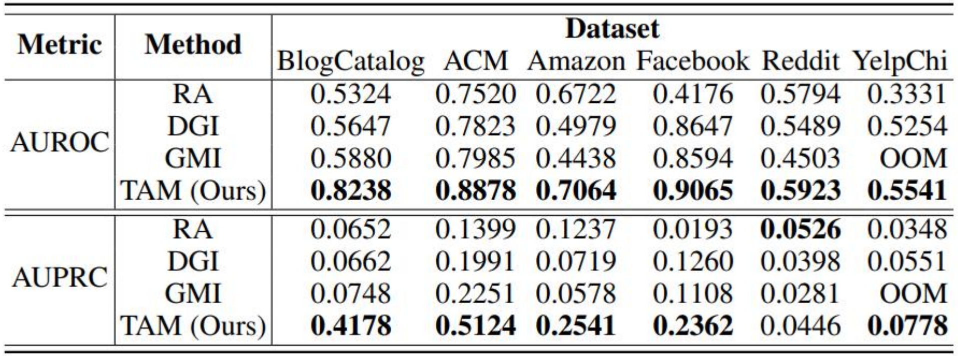
| **Metric** | **Method** | **Dataset** | | | |
|---|---|---|---|---|---|
| | | BlogCatalog | | ACM | |
| | | Structural | Contextual | Structural | Contextual |
| AUROC | DOMINANT | 0.5769 | 0.9591 | 0.6533 | 0.9506 |
| | CoLA | 0.6524 | 0.8867 | 0.7468 | 0.9200 |
| | SL-GAD | 0.5853 | **0.9754** | 0.7354 | **0.9878** |
| | TAM | **0.6819** | 0.9627 | **0.7902** | 0.9534 |
| AUPRC | DOMINANT | 0.0567 | 0.4369 | 0.0452 | 0.5049 |
| | CoLA | 0.0370 | 0.6298 | 0.0381 | 0.6166 |
| | SL-GAD | 0.0359 | 0.4776 | 0.0314 | 0.3083 |
| | TAM | **0.0570** | **0.6308** | **0.0568** | **0.7126** |

Compared to contextual anomalies, it is significantly more challenging to detect structural anomalies, for which TAM outperforms all three methods in both AUROC and AUPRC

# Performance Evaluation: TAM vs. Raw/Generic Node Representation Space

- ☐ Raw Attribute (RA)
- ☐ Deep Graph InfoMax
- ☐ Graph Mutual Informaltion Maximization

Table 4: Using local node affinity on raw attributes (RA) and learned representation spaces.



| Metric | Method | Dataset | | | | | |
|---|---|---|---|---|---|---|---|
| | | BlogCatalog | ACM | Amazon | Facebook | Reddit | YelpChi |
| AUROC | RA | 0.5324 | 0.7520 | 0.6722 | 0.4176 | 0.5794 | 0.3331 |
| | DGI | 0.5647 | 0.7823 | 0.4979 | 0.8647 | 0.5489 | 0.5254 |
| | GMI | 0.5880 | 0.7985 | 0.4438 | 0.8594 | 0.4503 | OOM |
| | TAM (Ours) | **0.8238** | **0.8878** | **0.7064** | **0.9065** | **0.5923** | **0.5541** |
| AUPRC | RA | 0.0652 | 0.1399 | 0.1237 | 0.0193 | **0.0526** | 0.0348 |
| | DGI | 0.0662 | 0.1991 | 0.0719 | 0.1260 | 0.0398 | 0.0551 |
| | GMI | 0.0748 | 0.2251 | 0.0578 | 0.1108 | 0.0281 | OOM |
| | TAM (Ours) | **0.4178** | **0.5124** | **0.2541** | **0.2362** | 0.0446 | **0.0778** |

# Performance Evaluation : Comparing to the alternative approaches

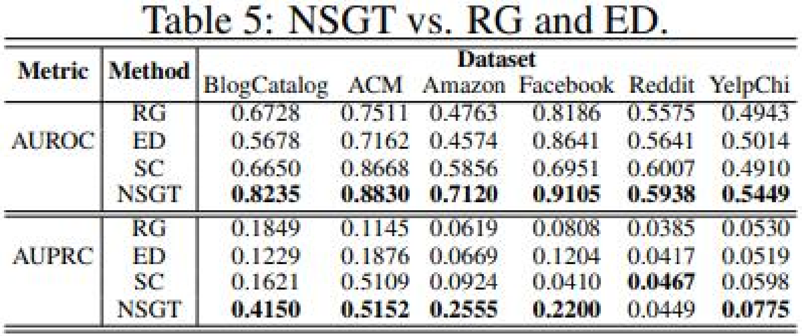


Table 5: NSGT vs. RG and ED.

| Metric | Method | Dataset | | | | | |
|---|---|---|---|---|---|---|---|
| | | BlogCatalog | ACM | Amazon | Facebook | Reddit | YelpChi |
| AUROC | RG | 0.6728 | 0.7511 | 0.4763 | 0.8186 | 0.5575 | 0.4943 |
| | ED | 0.5678 | 0.7162 | 0.4574 | 0.8641 | 0.5641 | 0.5014 |
| | SC | 0.6650 | 0.8668 | 0.5856 | 0.6951 | 0.6007 | 0.4910 |
| | NSGT | **0.8235** | **0.8830** | **0.7120** | **0.9105** | **0.5938** | **0.5449** |
| AUPRC | RG | 0.1849 | 0.1145 | 0.0619 | 0.0808 | 0.0385 | 0.0530 |
| | ED | 0.1229 | 0.1876 | 0.0669 | 0.1204 | 0.0417 | 0.0519 |
| | SC | 0.1621 | 0.5109 | 0.0924 | 0.0410 | **0.0467** | 0.0598 |
| | NSGT | **0.4150** | **0.5152** | **0.2555** | **0.2200** | 0.0449 | **0.0775** |

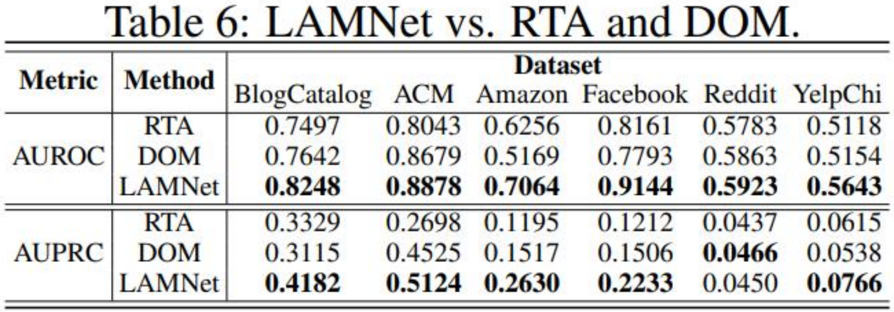


Table 6: LAMNet vs. RTA and DOM.

| Metric | Method | Dataset | | | | | |
|---|---|---|---|---|---|---|---|
| | | BlogCatalog | ACM | Amazon | Facebook | Reddit | YelpChi |
| AUROC | RTA | 0.7497 | 0.8043 | 0.6256 | 0.8161 | 0.5783 | 0.5118 |
| | DOM | 0.7642 | 0.8679 | 0.5169 | 0.7793 | 0.5863 | 0.5154 |
| | LAMNet | **0.8248** | **0.8878** | **0.7064** | **0.9144** | **0.5923** | **0.5643** |
| AUPRC | RTA | 0.3329 | 0.2698 | 0.1195 | 0.1212 | 0.0437 | 0.0615 |
| | DOM | 0.3115 | 0.4525 | 0.1517 | 0.1506 | **0.0466** | 0.0538 |
| | LAMNet | **0.4182** | **0.5124** | **0.2630** | **0.2233** | 0.0450 | **0.0766** |

☐ **Raw Graph (RG)** performs affinity maximization on the original graph structure without any graph truncation

☐ **Drop (ED)** randomly drops some edges (5% edges by default)

☐ **Similarity Cut (SC)** (removing 5% least similar edges).

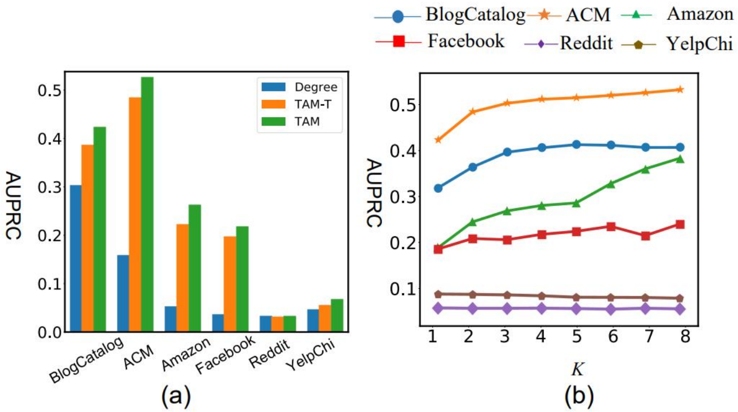
☐ **Raw Truncated Affinity (RTA)** directly calculates the local affinity-based anomaly scores after NSGT

☐ **DOM** performs LAMNet but with our affinity maximization objective replaced by the reconstruction loss used in DOMINANT

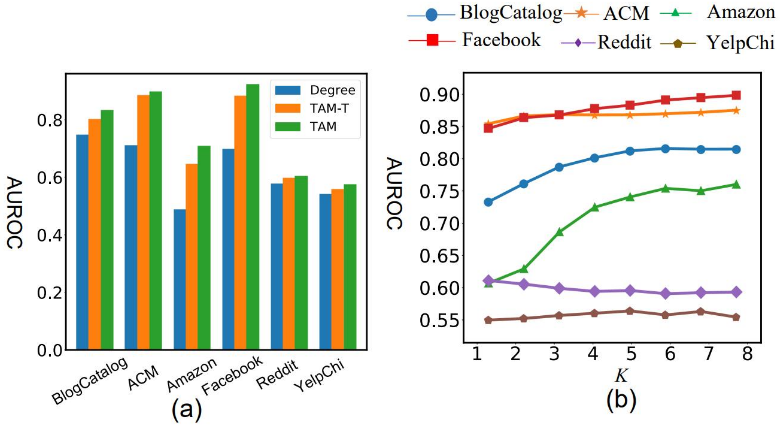
# Performance Evaluation : Comparing different anomaly scoring and truncation

- ☐ **TAM-T** calculates the node affinity on the truncated graph structure rather than the primary graph structure as in TAM
- ☐ **Degree** directly uses results w.r.t. graph truncation depth K. the node degree after our graph truncation as anomaly score

Aggregating the anomaly scores helps largely improve the detection performance



(a) (b)



(a) (b)

# Limitation

## ❖ Primarily Isolated Node

TAM cannot directly handle primarily isolated nodes in a graph, though those isolated nodes are clearly abnormal if they are rare and the other nodes are connected to at least some nodes

## ❖ Memory

TAM  may require a large memory to perform on graphs with a very large node/edge set.

## ❖ Anomaly Dense Block

Datasets with strong heterophily relations/subgraphs of normal nodes

# Conclusions and Future Work

We utilize this property to introduce a novel unsupervised GAD measure, local node affinity, and further introduce a truncated affinity maximization (TAM) approach that end-to-end optimizes the proposed anomaly measure on truncated adjacency matrix with the non-homophily edges eliminated.

## Work to be done . . .

Make it adapt to more large-scale datasets and more complex experimental setting.