

# Alleviating the Inconsistency Problem of Applying Graph Neural Network to Fraud Detection

作者: Yingtong Dou, Zhiwei Liu, Li Sun, Yutong Deng, Hao Peng, Philip S. Yu

Department of Computer Science, University of Illinois at Chicago School of Computer Science,  
Beijing University of Posts and Telecommunications Beijing Advanced Innovation Center for Big  
Data and Brain Computing, Beihang University

## Abstract

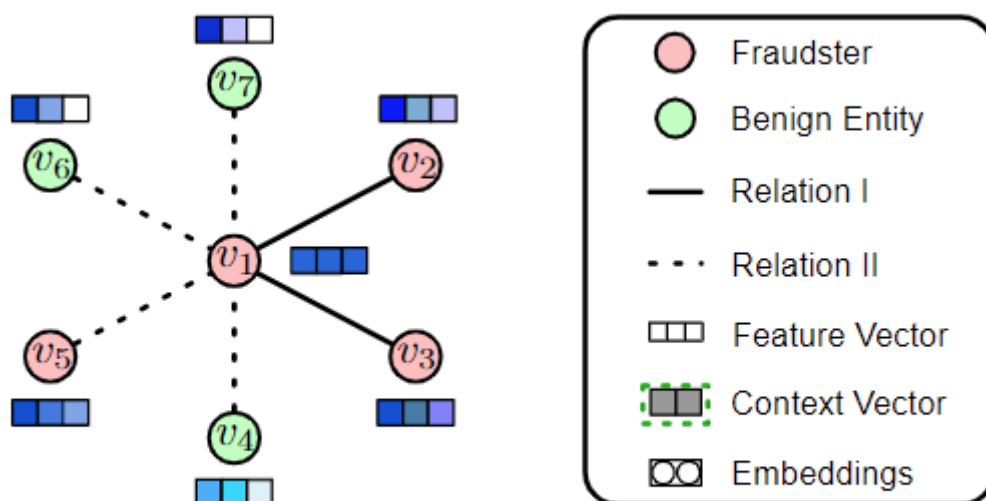
提出一种模型 *Graph - Consis*, 用于处理欺骗则引起的不一致问题

- 对于上下文不一致问题: 结合上下文节点的特征
- 对于特征不一致问题: 设计一个一致性分数来过滤不一致的邻居
- 对于关系不一致问题: 学习关系注意力权重

创新点/贡献/优势:

- 第一篇去解决在GNN模型中的不一致问题
- 从经验上分析了将GNN模型应用于欺诈检测任务的三个不一致问题
- 提出了Graph-Consis来解决三个不一致的问题, 这些问题结合了上下文嵌入, 邻域信息度量和关系注意力

## Introduction



*Left:* Inconsistency Problem

Context inconsistency

聪明的欺诈者可以将自己与伪造的常规实体联系起来, 同时, 欺诈者的数量远少于常规实体。

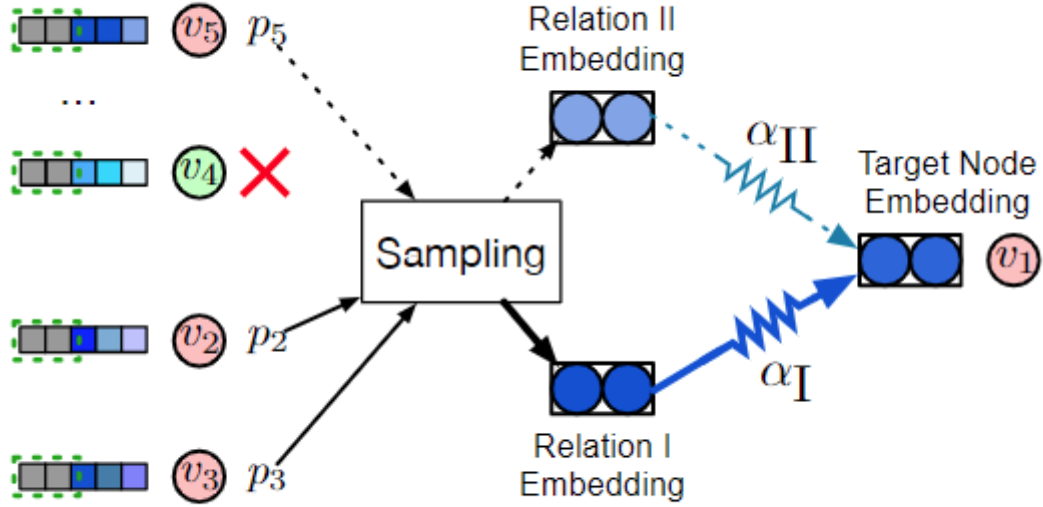
Feature inconsistency

假设有来自同一用户的两条评论，但是关于不同类别的产品，这两条评论具有共同点，因为它们共享同一用户。但是，由于它们与不同的产品相关联，因此它们的评论内容（功能）彼此相距遥远。

### Relation inconsistency

由于实体与多种类型的关系有关，因此平等对待所有关系会导致不一致性问题。例如，同一用户或同一产品可能同时关联了两个评论，这两个评论分别是共同用户关系和共同产品关系。假设一个评论是可疑的，则另一个评论如果通过“公共用户关联”来关联，则应具有较高的可疑性。

## Model



**Right: Proposed GraphConsis Model**

$$\mathbf{h}_v^{(1)} = \{\mathbf{x}_v \| \mathbf{c}_v\} \oplus \text{AGG}^{(1)}(\{\mathbf{x}_{v'} \| \mathbf{c}_{v'} : v' \in \mathcal{N}_v\}), \quad (2)$$

拼接上一个可训练的上下文特征  $c_v$

$$s^{(l)}(u, v) = \exp\left(-\|\mathbf{h}_u^{(l)} - \mathbf{h}_v^{(l)}\|_2^2\right), \quad (3)$$

以差值的  $l^2 - norm$  作为一致性分数

$$p^{(l)}(u; v) = s^{(l)}(u, v) / \sum_{u \in \tilde{\mathcal{N}}_v} s^{(l)}(u, v). \quad (4)$$

计算出每个点被选取的概率

接下来对于每种关系进行计算

$$\alpha_q^{(l)} = \exp\left(\sigma\left(\{\mathbf{h}_q^{(l)} \| \mathbf{t}_{r_q}\} \mathbf{a}^\top\right)\right) / \sum_{q=1}^Q \exp\left(\sigma\left(\{\mathbf{h}_q^{(l)} \| \mathbf{t}_{r_q}\} \mathbf{a}^\top\right)\right), \quad (5)$$

使用  $t_{r_q}$  作为可训练的关系向量，其余部分和GAT相似

$$\text{AGG}^{(l)} \left( \left\{ \mathbf{h}_q^{(l-1)} \right\} \middle|_{q=1}^Q \right) = \sum_{q=1}^Q \alpha_q^{(l)} \mathbf{h}_q^{(l)}, \quad (6)$$

最后就是利用上面求出的注意力系数，对于邻居信息进行聚合

## Experiments

### 数据集

**Table 1: The statistics of different graphs.**

	Graph	#Nodes	#Edges	$\gamma^{(f)}$	$\gamma^{(c)}$
<b>Others</b>	<b>Cora</b>	2,708	5,278	0.72	0.81
	<b>PPI</b>	14,755	225,270	0.48	0.98
	<b>Reddit</b>	232,965	11,606,919	0.70	0.63
<b>Ours</b>	<b>R-U-R</b>	45,954	98,630	0.83	0.90
	<b>R-T-R</b>	45,954	1,147,232	0.79	0.05
	<b>R-S-R</b>	45,954	6,805,486	0.77	0.05
	<b>Yelp-ALL</b>	45,954	7,693,958	0.77	0.07

Yelp数据集包含由Yelp过滤和推荐的酒店和餐厅评论。

R-U-R: it connects reviews posted by the same user

R-S-R: it connects reviews under the same product with the same star rating (1-5 stars)

R-T-R: it connects two reviews under the same product posted in the same month

$\gamma^{(f)}$  表示特征不一致的程度:  $\gamma_r^{(f)} = \sum_{(u,v) \in E_r} \exp(-\|\mathbf{x}_u - \mathbf{x}_v\|_2^2) / |E_r| \cdot d, \quad (8)$

$\gamma^{(c)}$  表示上下文不一致的程度:  $\gamma_r^{(c)} = \sum_{(u,v) \in E_r} (1 - \mathbb{I}(u \sim v)) / |E_r|, \quad (7)$

**Table 2: Experiment results under different training %.**

Method	40%		60%		80%	
	F1	AUC	F1	AUC	F1	AUC
<b>LR</b>	0.4647	<b>0.6140</b>	0.4640	0.6239	0.4644	0.6746
<b>GraphSAGE</b>	0.4956	0.5081	0.5127	0.5165	0.5158	0.5169
<b>FdGars</b>	0.4603	0.5505	0.4600	0.5468	0.4603	0.5470
<b>Player2Vec</b>	0.4608	0.5426	0.4608	0.5697	0.4608	0.5403
<b>GraphConsis</b>	<b>0.5656</b>	0.5911	<b>0.5888</b>	<b>0.6613</b>	<b>0.5776</b>	<b>0.7428</b>

可以看到，在这两个指标上，Graph-Consis在训练数据的80%和60%都优于其他模型，这表明我们可以缓解不一致问题。与其他基于GNN的模型相比，LR在AUC上性能稳定且更好。它表明节点功能很有用，但是GNN中的聚合器在识别欺诈者时会破坏分类器。该观察还证明了不一致问题是至关重要的，在将GNN应用于欺诈检测任务时应予以考虑。

与也学习关系注意的Player2Vec相比，Graph-Consis表现更好。这表明仅使用关系注意不能减轻特征不一致。邻居应该按照我们设计的方法进行过滤和采样。

FdGars直接聚合邻居的信息，而GraphSAGE以相等的概率抽样邻居。两者的性能都比Graph-Consis差，这表明我们的邻居采样技术很有用

## Conclusion

---

在本文中研究了在欺诈检测问题中应用GNN的三个不一致问题。为了解决这些问题，分别设计了三个模块并提出了Graph-Consis。实验结果表明了Graph-Consis的有效性。未来的工作包括为每个关系设计一个自适应采样阈值，以最大程度地提高GNN的接收范围。

调查其他欺诈数据集下的不一致问题是未来研究的另一种途径