Heterogeneous Graph Neural Networks for Malicious Account Detection

Ziqi Liu et al.

Ant Financial Service, Georgia Tech

27th ACM International Conference on Information and Knowledge Management (CIKM'18)
Industry and Case Study Paper

Presenter: Weilin Xu
https://qdata.github.io/deep2Read
Outline

1. Introduction
2. Method
3. Experiments
4. Conclusion
Malicious Account Detection:
To determine if an account is owned by adversary or normal user.
Malicious Account Detection:
To determine if an account is owned by adversary or normal user.

Proposed solution:
Graph Embeddings for Malicious accounts (GEM)
Intuition

Patterns observed from malicious accounts.

- Device aggregation
  Adversary logins to many accounts on one device.
Device Aggregation

**Figure:** Left: Normal; Right: Malicious.
Patterns observed from malicious accounts.

- **Device aggregation**
  Adversary logsins to many accounts on one device.

- **Activity aggregation**
  Adversary’s accounts behave in batches.
Activity Aggregation

**Figure:** Left: Normal; Right: Malicious.
Heterogeneous Graph Construction

- **Vertices**: 1) Account vertices; 2) Device vertices.
- **Edges**: Account is active on Device.

Represented as **adjacency matrix** \( A \in \{0, 1\}^{N,N} \).
- \( A_{i,j} = 1 \): account \( i \) active on device \( j \)
- \( A^{(d)} \): subgraph ignoring edges to non-type-\( d \) devices.

**Features of Vertices**: \( X \in \mathbb{R}^{N,p+|D|+200} \)
- Account vertices only: \( p \) time slots, with activity counts; \( p = 7 \times 24 = 168 \) slots, with activity counts.
- Device vertices only: one hot \( |D| \).
- 6 types of devices.
- Account vertices only: 200 demographics features.
Six device types.

- **Four Hardware ID.**
  - Phone number
  - WiFi MAC address
  - International Mobile Subscriber Identity (IMSI)
  - TID
    - Random number generated with IMSI and IMEI.

- **Two Proprietary Composite Fingerprint**
  - User Machine ID (UMID)
    - Unclear
  - Alipay Device ID (APDID)
    - Consider IMEI, IMSI, CPU, Bluetooth ADDR, ROM.
Six device types.

- **Four Hardware ID.**
  - Phone number
  - WiFi MAC address
  - International Mobile Subscriber Identity (IMSI)
  - TID
    - Random number generated with IMSI and IMEI.

- **Two Proprietary Composite Fingerprint**
  - User Machine ID (UMID) 0.4412 <Secret Weapon>
    - Unclear
  - Alipay Device ID (APDID)
    - Consider IMEI, IMSI, CPU, Bluetooth ADDR, ROM.
Six device types.

- **Four Hardware ID.**
  - Phone number **0.2952**
  - WiFi MAC address
  - International Mobile Subscriber Identity (IMSI)
  - TID
    Random number generated with IMSI and IMEI.

- **Two Proprietary Composite Fingerprint**
  - User Machine ID (UMID) **0.4412** *<Secret Weapon>*
    Unclear
  - Alipay Device ID (APDID)
    Consider IMEI, IMSI, CPU, Bluetooth ADDR, ROM.
Six device types.

- **Four Hardware ID.**
  - Phone number 0.2952
  - WiFi MAC address 0.13
  - International Mobile Subscriber Identity (IMSI)
  - TID
    
    Random number generated with IMSI and IMEI.

- **Two Proprietary Composite Fingerprint**
  - User Machine ID (UMID) 0.4412 <Secret Weapon>
    
    Unclear
  - Alipay Device ID (APDID)
    
    Consider IMEI, IMSI, CPU, Bluetooth ADDR, ROM.
Six device types.

- **Four Hardware ID.**
  - Phone number 0.2952
  - WiFi MAC address 0.13
  - International Mobile Subscriber Identity (IMSI)
  - TID
    Random number generated with IMSI and IMEI.

- **Two Proprietary Composite Fingerprint**
  - User Machine ID (UMID) 0.4412 <Secret Weapon>
    Unclear
  - Alipay Device ID (APDID) 0.0142
    Consider IMEI, IMSI, CPU, Bluetooth ADDR, ROM.
Six device types.

- **Four Hardware ID.**
  - Phone number 0.2952
  - WiFi MAC address 0.13
  - International Mobile Subscriber Identity (IMSI)
  - TID 0.0125
    Random number generated with IMSI and IMEI.

- **Two Proprietary Composite Fingerprint**
  - User Machine ID (UMID) 0.4412 <Secret Weapon> Unclear
  - Alipay Device ID (APDID) 0.0142
    Consider IMEI, IMSI, CPU, Bluetooth ADDR, ROM.
Models

Goal: learn embedding matrix $H$ ($i^{th}$ row is $h_i$ of vertex $i$)

$$H^{(0)} \leftarrow 0$$

for $t = 1, \ldots, T$

$$H^{(t)} \leftarrow \sigma(X \cdot W + \frac{1}{|D|} \sum_{d=1}^{|D|} A^{(d)} \cdot H^{(t-1)} \cdot V_d)$$

**Embeddings** at $t^{th}$ layer: $H^{(t)} \in \mathbb{R}^{N,k}$

**Features:** $X \in \mathbb{R}^{N,p+|D|}$, fed into each layer, ResNet alike.

**Trainable parameters:** $\{V_d\} \in \mathbb{R}^{k \times k}$;

$W \in \mathbb{R}^{P \times k} (P = p + |D|)$, shared among subgraphs.

**Adjacency matrix:** $A \in \{0, 1\}^{N,N}$

**Hyper-parameters:** Embedding size $k$;

$\#$hidden layers $T$ ($\#$hops a vertex needs to look at)
Attention Mechanism

\[ \alpha = [\alpha_1, \ldots, \alpha_{|\mathcal{D}|}]^T \in \mathbb{R}^{|\mathcal{D}|} \]

\[ \text{softmax}(\alpha_d) = \frac{\exp \alpha_d}{\sum_i \exp \alpha_i} \]

\[ H^{(t)} \leftarrow \sigma(X \cdot W + \sum_{d \in \mathcal{D}} \text{softmax}(\alpha_d) \cdot A^{(d)} \cdot H^{(t-1)} \cdot V_d) \]
Logistic Regression Classifier

\[
\min_{W, \{V_d\}, u} \mathbb{L}(W, \{V_d\}, u) = - \sum_{i}^{N_0} \log \sigma(y_i \cdot (u^\top h_i))
\]  

(1)

where \( \sigma = \frac{1}{1+\exp(-x)} \), \( u \in \mathbb{R}^k \)

Expectation Maximization style

- e-step: compute embeddings based on \( W, \{V_d\} \).
- m-step: optimize \( u \), while freezing embeddings.
Datasets

4 consecutive weeks of data from Alipay.
8M vertices, 10M edges
1.7M train labels, 0.2M test labels (number of account vertices?)
374 features

374 features for each vertex:
[Account Only] $p = 7 \times 24 = 168$ slots, with activity counts.
[Device Only] 6 types of devices.
[Account Only] 200 demographics features (yet another secret weapon?)

Train with first 6 days; test with the last day.
4 isolated experiments.
Comparison Methods

Baseline
- Connected Subgraph
- GBDT + Graph
- GBDT + Node2Vec
- Graph Convolutional Network

Variants of this work
- Graph Embeddings for Malicious accounts (GEM)
- GEM-attention
### Result - F-1 Score

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connected Subgraphs</td>
<td>0.5033</td>
<td>0.5567</td>
<td>0.58</td>
<td>0.5421</td>
</tr>
<tr>
<td>GBDT+Graph</td>
<td>0.7423</td>
<td>0.7598</td>
<td>0.7693</td>
<td>0.6639</td>
</tr>
<tr>
<td>GBDT+Node2Vec</td>
<td>0.741</td>
<td>0.7571</td>
<td>0.769</td>
<td>0.6626</td>
</tr>
<tr>
<td>GCN</td>
<td>0.7729</td>
<td>0.7757</td>
<td>0.7957</td>
<td>0.6919</td>
</tr>
<tr>
<td>GEM (Ours)</td>
<td>0.7992</td>
<td>0.8066</td>
<td>0.8191</td>
<td>0.718</td>
</tr>
<tr>
<td>GEM-attention (Ours)</td>
<td><strong>0.8165</strong></td>
<td><strong>0.8133</strong></td>
<td><strong>0.8244</strong></td>
<td><strong>0.7344</strong></td>
</tr>
</tbody>
</table>

**Figure:** F-1 Score
## Result - AUC

<table>
<thead>
<tr>
<th>Method</th>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connected Subgraphs</td>
<td>0.6689</td>
<td>0.6692</td>
<td>0.665</td>
<td>0.6938</td>
</tr>
<tr>
<td>GBDT+Graph</td>
<td>0.8878</td>
<td>0.8835</td>
<td>0.8707</td>
<td>0.8778</td>
</tr>
<tr>
<td>GBDT+Node2Vec</td>
<td>0.8884</td>
<td>0.883</td>
<td>0.8711</td>
<td>0.8773</td>
</tr>
<tr>
<td>GCN</td>
<td>0.8995</td>
<td>0.8932</td>
<td>0.8922</td>
<td>0.881</td>
</tr>
<tr>
<td>GEM (Ours)</td>
<td>0.9159</td>
<td>0.9238</td>
<td>0.9193</td>
<td>0.9082</td>
</tr>
<tr>
<td>GEM-attention (Ours)</td>
<td>0.9364</td>
<td>0.9293</td>
<td>0.9259</td>
<td>0.9155</td>
</tr>
</tbody>
</table>

**Figure:** AUC
98% precision over 89% of rule-based system. Recall unknown.
Precision-Recall Curves on Week 4

[Guess: Recall at 98% precision is about 0.5%.

Figure: Precision-Recall Curves on Week 4.

Ziqi Liu et al. (Alibaba, GeorgiaTech)
Novel graph neural network model for heterogeneous graph.

Exploit two weaknesses of adversary:
  Device aggregation & Activity aggregation.

Detect 10K malicious accounts daily at Alipay.

Future work: beyond adjacency matrix.
Not reproducible.
No open dataset or open source code.
Lack details of secret weapons.

Adaptive adversary.
Fake Hardware ID by hijacking system APIs on rooted devices.
Malicious account can be more active.