计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190600172

# 基于改进边权重的成对马尔可夫随机场模型的社交异常账号检测方法

宋 畅 禹 可 吴晓非
北京邮电大学信息与通信工程学院 北京 100876
(546226118@qq.com)

**摘 要** 社交媒体系统为人们提供了便利的共享、交流和协作平台。人们在享受社交媒体的开放性和便利性时,可能会发生许多恶意行为,例如欺凌、恐怖袭击计划和欺诈信息传播。因此,尽可能准确、及早地发现这些异常活动,以防止灾难和袭击,是非常重要的。近年来,随着在线社交网络(OSN)如 Twitter,Facebook,Google+,LinkedIN 等的成功,丰厚的利益资源使得它们成为了攻击者的目标。社交网络的开放性,使其特别容易受到异常账号攻击的威胁。现有基于图形的最先进分类模型大多使用首先为图的边分配权重,在加权图中迭代地传播节点的信誉分数,并使用最终的后验分数来对节点进行分类的方法。边权重的分配是其中一项重要的任务,此参数将直接影响检测结果的准确度。为此,文中针对社交媒体中异常账号的检测任务,分析了基于社交图全局结构的方法,通过在成对马尔可夫随机场模型中改进边权重的计算方法,使其能够在迭代过程中自适应优化,提出了准确度更高的 GANG+LW,GANG+LOGW 和 GANG+PLOGW 算法。这 3 种算法使用了不同的改进边权重的方法。实验证明,新提出的方法相对于基本的成对马尔可夫随机场模型,可取得更准确的异常账号检测结果,3 种算法中 GANG+PLOGW 得到的结果最好。结果证明,此改进模型在检测社交网络中的异常账号时,能够更有效地解决问题。

**关键词**:社交媒体;异常账号检测;马尔可夫随机场;Sybil 攻击
**中图法分类号** TP181

## Fake Account Detection Method in Online Social Network Based on Improved Edge Weighted Paired Markov Random Field Model

SONG Chang, YU Ke and WU Xiao-fei
School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

**Abstract** Social media systems provide a convenient platform for sharing, communication and collaboration. When people enjoy the openness and convenience of social media, there may be many malicious acts, such as bullying, terrorist attacks and fraudulent information dissemination. Therefore, it is very important to be able to detect these anomalous activities as accurately and early as possible to prevent disasters and attacks. The success of online social networks (OSN) in recent years, such as Twitter, Facebook, Google+, LinkedIn, has made them targests of attacker's goal due to their rich profit resources. The openness of social networks makes them particularly vulnerable to unusual account attacks. Existing classification models mostly use method that first assign weights to the edges of the graph, iteratively propagate the reputation scores of the nodes in the weighted graph, and use the final posterior scores to classify the nodes. One of the important tasks is the settingof edge weight. This parameter will directly affect the accuracy of the test results. Based on the detection task of fake account in social media, this paper analyzed the global structure based on social graph, and improves the algorithm of edge weight in the paired Markov random field model, so that it can adaptively optimize in the iterative process. GANG+LW, GANG+LOGW, and GANG+PLOGW algorithms with higher accuracy were proposed. These three algorithms used three different methods to improve the algorithm of edge weight. Experiments show that the proposed method can obtain more accurate fake account detection results than the basic paired Markov random field model, in which GANG+PLOGW got the best results in the three algorithms. The result proves that this improved model can solve the problem more effectively when detecting fake accounts in social networks.

**Keywords** Social media, Fake account detection, Markov random field, Sybil attack

## 1 引言

社交媒体系统为人们提供了便利的共享、交流和协作平台。人们在享受社交媒体的开放性和便利性时,可能会发生许多恶意行为,例如虚假账号攻击和欺诈信息传播。因此,尽可能准确、及早地发现这些异常活动,以防止恶意行为,非常重要[1]。

到稿日期:2019-06-28 返修日期:2019-09-11 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61601046,61171098);中国 111 基地项目(B08004);欧盟 FP7 IRSES 项目(612212)
This work was supported by the National Natural Science Foundation of China (61601046, 61171098), 111 Project of China (B08004) and EU FP7 IRSES Mobile Cloud Project (612212).
通信作者:禹可(yuke@bupt.edu.cn)

在社交媒体的异常账号检测中，最重要的一类检测是Sybil攻击检测。Sybil攻击是指利用社交网络中的少数节点控制多个虚假身份，从而利用这些身份控制或影响网络的大量正常节点的攻击方式。这种账号最明显的特征就是会发出大量的好友请求，企图与大量的正常用户建立好友关系，为后续发起恶意活动创造条件。Sybil攻击对社交网络的信誉评价体系和用户的信任关系都造成了严重危害。这些攻击破坏了在线社交网络的信任基础，从而导致安全性和隐私侵犯等问题，破坏了网络环境[2]。

本文着重对社交网络中的异常账号检测方法的研究进展进行说明，重点探讨Sybil账号的检测。

现有的异常账号检测方法分为两类：基于局部特征的方法和基于全局结构的方法。基于局部特征的方法不具有强鲁棒性，攻击者通过调整攻击策略容易绕过检测；基于全局结构的方法具有更高的鲁棒性，但其中一些参数的设置不能准确地模拟实际情况，从而降低了算法的检测精度[3-6]。本文着重研究基于社交图全局结构的方法，对重要参数边权重的计算进行改进，并对结果进行分析。

本文第2节介绍了现有Sybil账号检测的相关研究；第3节对改进边权重的Sybil异常检测算法进行详细的描述；第4节对实验数据集和结果进行了详细分析；最后总结全文并展望未来。

## 2 相关研究

现有的异常账号检测方法分为两类：基于局部特征的方法和基于全局结构的方法[7-9]。

1)基于局部特征的方法：旨在通过使用机器学习技术训练的分类器来查找Sybil账号。分析用户活动以提取一些特征，这些特征将作为构建分类器的输入。该类方法利用用户的局部子图结构(如自我网络)、辅助信息(如IP地址、行为和内容)，并将它们与全局用户社交网络结构结合。这些方法的一个关键限制是它们不是对抗稳健的。

2)基于全局结构的方法：将OSN建模为具有分别由节点和链接表示的用户账号和社交关系的图。大多数基于结构的方法可以分为随机游走(Random Walk，RW)和循环置信传播(Loop Belief Propagation，LBP)两类[11]。该类方法利用社交图的全局结构且通常基于直觉，即如果用户与其他欺诈(或普通)用户相关联，则用户可能是欺诈(或正常)的直觉[12]。现有的依靠关联性的方法或者假定对称(即无向)社交链接，过分简化了现实世界OSN中的非对称(即定向)社交图结构，或者仅利用训练数据集中标记的欺诈用户或正常用户(但不是两者)，这限制了它们的检测精度[13]。

现有基于图形的最先进分类方法的重点是分配边权重，如果两个对应节点趋向具有相同标签则边具有大权重，否则具有小权重。大多数现有方法只是为所有边设置一个恒定的权重，难以得到更好的检测效果[14]。

本文基于全局结构的检测方法，提出了一种新的为图的边分配权重的方法，其根据边所连接节点的不同为所有的边设定独特的权重，使检测结果更加准确。

## 3 改进边权重的马尔可夫随机场模型

本文基于Wang等提出的GANG算法[15]进行改进。本节首先对GANG算法进行概述，再详细阐述改进边权重的方法。

### 3.1 定义

假设给予一个有向社交图和一个由标记的sybil节点和良性节点组成的训练数据集，则异常账号检测是预测社交图中每个剩余节点的标签[15]。

在Wang等的工作中，将社交网络构建成有向社交图$G=(V,E)$。其中，一个节点$v\in V$代表一个用户，$|V|$是用户的总数量；有向向$(u,v)\in E$表示$u$在社交图$G$中关注了$v$。每个节点$v\in V$都可能是sybil节点或良性节点。如果边$(v,u)$不存在，称边$(u,v)$为单向边；如果边$(v,u)$存在，称边$(u,v)$为双向边[15]。

定义每个节点的邻居集，用$\Gamma_b(u)$，$\Gamma_i(u)$和$\Gamma_o(u)$表示用户$u$的双向输入、单向输入、单向输出邻居的集合。形式上有$\Gamma_b(u)=\{v|(v,u)\in E \text{ and } (u,v)\in E\}$，$\Gamma_i(u)=\{v|(v,u)\in E \text{ and } (u,v)\notin E\}$，$\Gamma_o(u)=\{v|(u,v)\in E \text{ and } (v,u)\notin E\}$[15]。

### 3.2 基本GANG算法

GANG算法[16]是一种关于有向图的牵连负罪的方法，用于检测OSN中的欺诈用户。它基于一种新型的成对马尔可夫随机场。在基本版本中，给定一个训练数据集，利用LBP来估计每个用户的后验概率分布，并用它来预测用户的标签。但是，基本版本的可扩展性不够，且不能保证收敛，因为它依赖于LBP。因此，GANG的优化版本将其表示为一个简洁的矩阵形式，可以得出收敛的条件。

在GANG中，将二元随机变量与每个用户相关联来建模其标签，$x_u=1$和$x_u=-1$分别表示$u$是sybil节点和良性节点。

GANG算法基于3个直觉认知。

1)邻居$v$是$u$的双向邻居，则$u$倾向于与$v$具有相同标签，因此$u$是欺诈节点的概率可建模为sigmoid函数：

$$\Pr(x_u=1|\bar{x}_{\Gamma_b(u)})=\frac{1}{1+\exp(-\sum_{v\in\Gamma_b(u)}J_{vu}\bar{x}_v)} \tag{1}$$

其中，$J_{vu}$是边$(v,u)$的耦合强度，并且为双向边设置$J_{vu}=J_{uv}$。此外，设置$J_{vu}>0$来为同质性属性建模。在此模型中，如果更多的双向邻居是欺诈性的，那么$u$有较高的欺诈概率。sigmoid函数可以使用定制的pMRF捕捉这些直觉。

2)如果$v$是一个单向传入邻居，那么如果$v$是欺诈性的，$v$不会为$u$的标签提供信息。被欺诈节点链接并不能对判断产生影响，然而，当$v$正常时，$u$也往往是正常的。

$$\Pr(x_u=1|\bar{x}_{\Gamma_i(u)})=\frac{1}{1+\exp(-\frac{1}{2}\sum_{v\in\Gamma_i(u)}J_{vu}(\bar{x}_v-1)} \tag{2}$$

其中，$J_{uv}>0$。在模型中，若已知单向传入邻居是欺诈用户，则不会影响$u$的标签；如果更多的单向传入邻居被认为是正常的，那么$u$就不太可能是欺诈用户。

3)如果$v$是一个单向传出邻居，那么若$v$是一个正常节点，则$v$对$u$的标签没有提供信息。这是因为任何节点都可以链接到OSN中的正常节点。如果$v$是欺诈节点，那么$u$也

往往是欺诈节点，因为普通用户很少关注欺诈节点。

$$\Pr(x_u=1|\bar{x}_{\Gamma_o(u)})=\frac{1}{1+\exp(-\frac{1}{2}\sum_{v\in\Gamma_o(u)}J_{vu}(\bar{x}_v+1)}\tag{3}$$

其中，$J_{uv}>0$。在模型中，如果已知单向传出邻居是正常的，则不会影响 $u$ 的标签；如果更多的单向传出邻居被认为是欺诈节点，那么 $u$ 更可能是欺诈节点。

用一种新颖的定制 pMRF 来捕捉以上直觉。首先，将 $u$ 的标签先验建模为：

$$\Pr(x_u=1)=\frac{1}{1+\exp(-h_u)}\tag{4}$$

其中，$h_u>0$ 和 $h_u<0$ 分别表示根据先验知识 $u$ 倾向于是欺诈或正常的；而 $h_u=0$ 意味着 $u$ 的先验知识不能确定 $u$ 的标签。

假设已经知道 $u$ 的邻居的标签和它的先验知识，将 $u$ 为欺诈节点的概率建模为：

$$\Pr(x_u=1|\bar{x}_{\Gamma(u)}=\frac{1}{1+\exp(-I_b(u)-I_i(u)-I_o(u)-h_u)}\tag{5}$$

其中，双向邻居总影响为：

$$I_b(u)=\sum_{v\in\Gamma_b(u)}J_{vu}\bar{x}_v\tag{6}$$

单向传入邻居总影响为：

$$I_i(u)=\frac{1}{2}\sum_{v\in\Gamma_i(u)}J_{vu}(\bar{x}_v-1)\tag{7}$$

单向传出邻居总影响为：

$$I_o(u)=\frac{1}{2}\sum_{v\in\Gamma_o(u)}J_{uv}(\bar{x}_v+1)\tag{8}$$

$h_u>0$ 是关于 $u$ 的先验知识。

pMRF 为所有 $u\in V$ 建立所有二元随机变量 $x_u$ 的联合概率分布。用 $x_V$ 表示所有二元随机变量的集合。pMRF 如下：

$$H(x_V)=-\frac{1}{2}\sum_{(u,v)\in E_2}J_{uv}x_ux_v-\frac{1}{4}\sum_{(u,v)\in E_1}J_{uv}(x_u-1)(x_v+1)-\frac{1}{2}\sum_{u\in V}h_ux_u\tag{9}$$

$$P_r(x_V)\propto\exp(-H(x_V))\tag{10}$$

其中，$H(x_V)$ 通常被称为能量函数。当 $(u,v)$ 是双向边时，$(u,v)$ 或 $(v,u)$ 单独出现，而不是两者同时出现在 $H(x_V)$ 中。特别地，当观察到 $u$ 的邻居状态时，$u$ 是欺诈节点的条件概率由统一框架给出。如果 $u$ 是欺诈节点或 $v$ 是正常节点，单向边 $(u,v)$ 不影响能量函数和联合概率。

设定节点潜力 $\phi_u(x_u):=\begin{cases}q_u, & \text{if } x_u=1\\ 1-q_u, & \text{if } x_u=-1\end{cases}$；以及边潜力：双向 $\varphi_{uv}(x_u,x_v):=\begin{cases}w_{uv}, & \text{if } x_ux_v=1\\ 1-w_{uv}, & \text{if } x_ux_v=-1\end{cases}$，单向 $\varphi_uv(x_u,x_v):=\begin{cases}w_{uv}, & \text{if } x_u=1 \text{ or } x_v=-1\\ 1-w_{uv}, & \text{otherwise}\end{cases}$。

其中，边权重 $w_{uv}:=(1+\exp\{-J_{uv}\})^{-1}$，被称为同质强度，表示两个节点通过双向边链接时具有相同标签的概率。这里，为所有边设置 $w_{uv}=w>0.5$。

用节点潜力和边潜力可将 $pMRFP_r(x_V)\propto\exp(-H(x_V))$ 构建为 $\Pr(x_r)=\frac{1}{Z}\prod_{v\in V}\phi_v(x_v)\prod_{(u,v)\in E_1\cup E_2}\varphi_{uv}(x_u,x_v)$。其中，$Z=\sum_{x_v}\prod_{v\in V}\phi_v(x_v)\prod_{(u,v)\in E_1\cup E_2}\varphi_{uv}(x_u,x_v)$。

利用上述 pMRF 来检测欺诈节点，得到节点 $u$ 的后验概率分布 $\Pr(x_u)=\sum_{x_V/x_u}\Pr(x_V)$。如果后验结果 $p_u>0.5$，则预测 $u$ 是欺诈节点，否则预测 $u$ 是正常节点。

在 GANG 的基本版本中，使用 LBP[20] 来估计后验概率分布。LBP 在图中的相邻节点之间迭代地传递消息。具体地，在第 $t$ 次迭代中，从 $v$ 发送到 $u$ 的消息是：

$$m_{vu}^{(t)}(x_u)=\sum_{x_v}\phi_v(x_v)\varphi_{vu}(x_u,x_u)\prod_{k\in\Gamma(v)/u}m_{kv}^{(t-1)}(x_v)\tag{11}$$

当两个连续迭代中消息的变化可以忽略不计（如变化的距离小于 $1\times10^{-3}$）或者它达到预定义的最大迭代次数时，LBP 停止。在 LBP 停止后，估计后验概率 $\Pr(x_u)$ 如下：

$$\Pr^{(t)}(x_u)\propto\phi_u(x_u)\prod_{k\in\Gamma(u)}m_{ku}^{(t)}(x_u)\tag{12}$$

$$p_u^{(t)}=\frac{q_u\prod_{k\in\Gamma(u)}m_{ku}^{(t)}}{q_u\prod_{k\in\Gamma(u)}m_{ku}^{(t)}+(1-q_u)\prod_{k\in\Gamma(u)}(1-m_{ku}^{(t)})}\tag{13}$$

其中：

$$m_{ku}^{(t)}=m_{ku}^{(t)}(x_u=1),1-m_{ku}^{(t)}=m_{ku}^{(t)}(x_u=-1)\tag{14}$$

基础模型有两个缺点：1）不可扩展；2）不能保证收敛。因此，文献[7]进行了两步优化，第一步是消除消息保留以增强算法的可扩展性。GANG 的可扩展性不足的主要原因之一是 LBP 在每个边都保留消息。因此，在此优化中，当 $v$ 为 $u$ 准备其消息时，将在其中包含 $u$ 发送给 $v$ 的消息。第二步将其近似为矩阵形式，其中 $\boldsymbol{A}_b$，$\boldsymbol{A}_i$ 和 $\boldsymbol{A}_o$ 表示社交图的双向输入、单向输入和单向输出邻接矩阵。$\boldsymbol{A}_b$，$\boldsymbol{A}_i$ 和 $\boldsymbol{A}_o$ 的第 $u$ 行表示 $u$ 的双向传入、单向传入和单向传出邻居；如果 $u$ 和 $v$ 之间存在双向边，则 $A_{b,uv}=A_{b,vu}=1$，否则 $A_{b,uv}=A_{b,vu}=0$；如果存在从 $u$ 到 $v$ 的单向边，则 $A_{o,uv}=A_{i,vu}=1$。定义第 $t$ 次迭代中所有节点后验的列向量 $\boldsymbol{p}^{(t)}=[p_1^{(t)};p_2^{(t)},\cdots,p_{|V|}^{(t)}]$，所有节点先验的列向量 $\boldsymbol{q}=[q_1,q_1,\cdots,q_{|V|}]$。令 $\hat{\boldsymbol{P}}^{(t)}$ 为由 $|V|$ 个重复列向量 $\hat{\boldsymbol{P}}^{(t)}$ 组成的矩阵 $\hat{\boldsymbol{P}}^{(t)}=[\hat{\boldsymbol{P}}^{(t)},\hat{\boldsymbol{P}}^{(t)},\cdots]$，其中 $\hat{\boldsymbol{P}}^{(t)}$ 为残差向量。

最终得到矩阵形式的结果：

$$\begin{cases}\boldsymbol{A}_i'^{(t-1)}=\boldsymbol{I}(\boldsymbol{A}_i\circ\hat{\boldsymbol{P}}^{(t-1)^{\mathrm{T}}})\\ \boldsymbol{A}_o'^{(t-1)}=\boldsymbol{I}(\boldsymbol{A}_o\circ\hat{\boldsymbol{P}}^{(t-1)^{\mathrm{T}}})\\ \hat{\boldsymbol{p}}^{(t)}=\hat{q}+2\cdot\hat{w}(\boldsymbol{A}_b+\boldsymbol{A}_i'^{(t-1)}+\boldsymbol{A}_o'^{(t-1)})\cdot\hat{p}^{(t-1)}\end{cases}\tag{15}$$

其中，运算符$\circ$表示两个矩阵的元素乘积。

### 3.3　改进边权重的方法

在 GANG 的算法中，使用节点潜力和边潜力构建 pMRF 时，通过边权重 $w_{uv}$ 计算边潜力。$w_{uv}$ 是一个非常重要的参数，直接影响最终结果的准确性。在 GANG 中，双向边潜力为：

$$\varphi_{uv}(x_u,x_v):=\begin{cases}w_{uv}, & \text{if } x_ux_v=1\\ 1-w_{uv}, & \text{if } x_ux_v=-1\end{cases}\tag{16}$$

单向边潜力为：

$$\varphi_{uv}(x_u,x_v):=\begin{cases}w_{uv}, & \text{if } x_u=1 \text{ or } x_v=-1\\ 1-w_{uv}, & \text{otherwise}\end{cases}\tag{17}$$

已有算法中为所有边设置固定值，即 $w_{uv}=w>0.5$，这个固定值不能充分模拟真实情况，限制了算法的性能。本文针对这个问题进行改进，使用新的方法为边分配更合理的数值，从而提升检测算法的准确率。

在原始的 GANG 算法中，最终得到的是每个节点的 0～1 的评分，节点 $u$ 的评分设为 $S_1$，节点 $v$ 的评分设为 $S_2$，则边 $(u,v)$ 的权重值使用 $S_1$ 和 $S_2$ 来计算。改进的边权重按照如下 3 种策略计算，以在迭代中自适应优化。

策略 1 使用归一化线性模型

$$w_{uv}=(s_1-\tau)(s_2-\tau)\times 2+\tau \tag{18}$$

策略 2 使用直接的 Logistic 回归模型

$$w_{uv}=1/(1+\exp(-(s_1-\tau)(s_2-\tau))) \tag{19}$$

策略 3 使用基于概率的 Logistic 回归模型

$$w_{uv}=1/(1+\exp(-(s_1-\tau)(s_2-\tau)\cdot P_T+(s_1-\tau)(s_2-\tau)\cdot(1-P_T))) \tag{20}$$

其中，$\tau$ 是阈值，按照算法中的判定分类，将阈值设定为 0.5；$P_T$ 是前一次迭代中结果的准确率。

在使用 GANG 对原始数据进行第一次处理后，利用第一次得到的结果分别计算每条边的权重 $w_{uv}$，再将得到的边权重代入原来的算法中进行第二次处理，然后使用第二次的结果为所有边分配新的权重，如此构成迭代。

改进边权重的算法如算法 1－算法 3 所示。

**算法 1** LW 算法

Input：社交网络数据集边权重 $w_{uv}$

Output：节点编号及相应后验概率

for 迭代次数 $\tau_1<$设定最大迭代次数 $T_1$

输入 $w_{uv}$，构成邻接矩阵 $\mathbf{A}_b$，$\mathbf{A}_i$，$\mathbf{A}_o$

for 迭代次数 $\tau_2<$设定最大迭代次数 $T_2$

$\mathbf{A}_i'^{(t-1)}=\mathbf{I}(\mathbf{A}_i\circ\hat{\mathbf{P}}^{(t-1)^T})$

$\mathbf{A}_o'^{(t-1)}=\mathbf{I}(\mathbf{A}_o\circ\hat{\mathbf{P}}^{(t-1)^T})$

$\hat{\mathbf{p}}^{(t)}=\hat{q}+2\cdot\hat{w}(\mathbf{A}_b+\mathbf{A}_i'^{(t-1)}+\mathbf{A}_o'^{(t-1)})\cdot\hat{p}^{(t-1)}$

$w_{uv}=(s_1-\tau)(s_2-\tau)\times 2+\tau$

输出节点编号及相应后验概率

End

**算法 2** LOGW 算法

Input：社交网络数据集边权重 $w_{uv}$

Output：节点编号及相应的后验概率

for 迭代次数 $\tau_1<$设定的最大迭代次数 $T_1$

输入 $w_{uv}$，构成邻接矩阵 $\mathbf{A}_b$，$\mathbf{A}_i$，$\mathbf{A}_o$

for 迭代次数 $\tau_2<$设定的最大迭代次数 $T_2$

$\mathbf{A}_i'^{(t-1)}=\mathbf{I}(\mathbf{A}_i\circ\hat{\mathbf{P}}^{(t-1)^T})$

$\mathbf{A}_o'^{(t-1)}=\mathbf{I}(\mathbf{A}_o\circ\hat{\mathbf{P}}^{(t-1)^T})$

$\hat{\mathbf{p}}^{(t)}=\hat{q}+2\cdot\hat{w}(\mathbf{A}_b+\mathbf{A}_i'^{(t-1)}+\mathbf{A}_o'^{(t-1)})\cdot\hat{p}^{(t-1)}$

$w_{uv}=1/(1+\exp(-(s_1-\tau)(s_2-\tau)))$

输出节点编号及相应后验概率

End

**算法 3** PLOGW 算法

Input：社交网络数据集边权重 $w_{uv}$

Output：节点编号及相应的后验概率

for 迭代次数 $\tau_1<$设定的最大迭代次数 $T_1$

输入 $w_{uv}$，构成邻接矩阵 $\mathbf{A}_b$，$\mathbf{A}_i$，$\mathbf{A}_o$

for 迭代次数 $\tau_2<$设定的最大迭代次数 $T_2$

$\mathbf{A}_i'^{(t-1)}=\mathbf{I}(\mathbf{A}_i\cdot\hat{\mathbf{P}}^{(t-1)^T})$

$\mathbf{A}_o'^{(t-1)}=\mathbf{I}(\mathbf{A}_o\circ\hat{\mathbf{P}}^{(t-1)^T})$

$\hat{\mathbf{p}}^{(t)}=\hat{q}+2\cdot\hat{w}(\mathbf{A}_b+\mathbf{A}_i'^{(t-1)}+\mathbf{A}_o'^{(t-1)})\cdot\hat{p}^{(t-1)}$

$w_{uv}=1/(1+\exp(-(s_1-\tau)(s_2-\tau)\cdot P_T+(s_1-\tau)(s_2-\tau)\cdot(1-P_T)))$

输出节点编号及相应的后验概率

End

## 4 实验及结果分析

### 4.1 数据集

本文使用了两个数据集，如表 1 所列。

表 1 测试数据集

Table 1 Test dataset

| 数据集 | 节点数 | 边数 | 平均度 |
|---|---|---|---|
| Pokec[16] | 10000 | 94065 | 19 |
| Twitter1KS-10KN[17] | 11000 | 3260991 | 593 |

文献[16]的真实有向社交图(即 Pokec)数据集来自 SNAP[1)]。从图中提取包含 10000 个节点的连通组件作为良性区域。此外，将 Sybil 区域合成为良性区域的复制品，并在两个区域之间随机、均匀地添加(双向、单向 Sybil 至良性，或/和单向良性至 Sybil)攻击边。

Twitter1KS-10KN 有向社交图数据集包含 1000 个已识别的恶意 Twitter 账号的数据和 10000 个普通 Twitter 账号的数据。

参与实验比较的算法如下：

1)GANG 算法，详情如 3.2 节描述。

2)GANG＋基于归一化线性模型的边权重(GANG＋LW)，在 GANG 算法中使用策略 1。

3)GANG＋基于 Logistic 回归模型的边权重(GANG＋LOGW)，在 GANG 算法中使用策略 2。

4)GANG＋基于概率的 Logistic 回归模型的边权重(GANG＋PLOGW)，在 GANG 算法中使用策略 3。

使用以下指标评估实验结果：

1)TPR(True Positive Rate)，即分类器分类正确的正样本个数占总正样本个数的比例。

2)ACC(Accuracy)，即分类正确的样本个数占所有样本个数的比例。

3)AUC(Area Under the Curve)，即 ROC 曲线下的面积。

### 4.2 结果分析

首先比较改进算法的性能，将 3 种改进策略分别与未改进的算法进行对比。

在合成的 Pokec 数据集上的实验结果如表 2 所列。

表 2 Pokec 数据集上的结果

Table 2 Results on synthesized Pokec dataset

(单位：%)

| | GANG | GANG＋LW | GANG＋LOGW | GANG＋PLOGW |
|---|---|---|---|---|
| *TPR* | 97.1 | 97.3 | 99.6 | 99.7 |
| *ACC* | 98.5 | 98.8 | 99.8 | 99.8 |
| *AUC* | 99.7 | 99.7 | 99.9 | 99.9 |

在 Twitter1KS-10KN 数据集上的实验结果如表 3 所列。

1) http://snap.stanford.edu/data/index.html

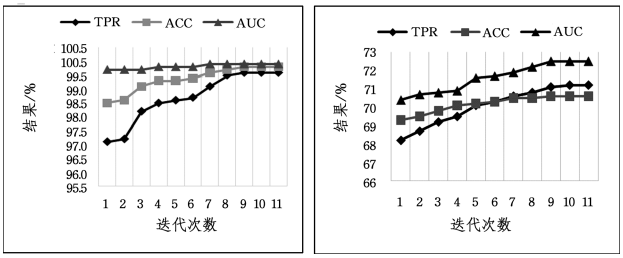
表 3　在 Twitter1KS-10KN 数据集上的结果

Table 3　Results on Twitter1KS-10KN dataset

（单位：%）

| | GANG | GANG+LW | GANG+LOGW | GANG+PLOGW |
|---|---|---|---|---|
| *TPR* | 68.2 | 68.5 | 71.1 | 72.7 |
| *ACC* | 69.3 | 69.1 | 70.6 | 71.3 |
| *AUC* | 70.4 | 70.1 | 72.5 | 72.8 |

其次，分析了迭代次数对改进算法性能的影响。迭代次数对结果的影响如图 1 所示，以策略 2（GANG+LOGW）在两个数据集上的结果为例。




（a）Pokec 数据集　（b）Twitter1KS-10KN 数据集

图 1　迭代次数对结果的影响

Fig. 1　Effect of iterationtimes on results

从图 1 可以看出，在迭代次数达 8～10 时，算法的结果达到最理想状态。所设定的加入边权重的方案在合成数据集和实际数据集上都能较好地提升算法的性能，其中基于概率的 Logistic 回归模型的边权重优化方法的效果最好。

**结束语**　本文研究了基于社交图全局结构的社交网络异常账号检测方法，对其中边权重的参数计算进行了改进，提出了准确度更高的 GANG+LW，GANG+LOGW 和 GANG+PLOGW 算法，并对结果进行分析。实验结果证明，在合成数据集和真实数据集中加入了提出的边权重分配方式后，结果都得到了相应的提升，因此可以证明此边权重分配方式是有效的。而此方法仍然需要有人工标注的数据集进行训练，而此类数据集难以获取且成本较高，因此在 Sybil 账号检测方面，开发有效的无监督算法是未来的工作。

## 参 考 文 献


[1] YU R，QIU H D，WEN Z，et al. A Survey on Social Media Anomaly Detection[J]. ACM SIGKDD Explorations Newsletter，2016，18(1)：1-4.

[2] GAO P，WANG B，GONG N Z，et al. Sybilfuse：Combining local attributes with global structure to perform robust sybil detection[J]. arXiv：1803.06772，2018.

[3] YANG Z，XUE J，YANG X，et al. VoteTrust：Leveraging friend invitation graph to defend against social network sybils[J]. IEEE Transactions on Dependable and Secure Computing，2016，13(4)：488-501.

[4] MISRA S，TAYEEN A S M，XU W. SybilExposer：An effective scheme to detect Sybil communities in online social networks [C]// 2016 IEEE International Conference on Communications (ICC). IEEE，2016：1-6.

[5] ZHENG H，XUE M，LU H，et al. Smoke screener or straight shooter：Detecting elite sybil attacks in user-review social networks[J]. arXiv：1709.06916，2017.

[6] DAVIS C A，VAROL O，FERRARA E，et al. BotOrNot：A system to evaluate social bots[C]// Proceedings of the 25th International Conference Companion on World Wide Web. International World Wide Web Conferences Steering Committee，2016：273-274.

[7] SAVAGE D，ZHANG X Z，YU X H，et al. Anomaly Detection in Online Social Network[J]. Social Network，2014，39：62-70.

[8] YANG Z，WILSON C，WANG X，et al. Uncovering social network sybils in the wild[J]. ACM Transactions on Knowledge Discovery from Data (TKDD)，2014，8(1)：2.

[9] GATTERBAUER W，GÜNNEMANN S，KOUTRA D，et al. Linearized and single-pass belief propagation[J]. Proceedings of the VLDB Endowment，2015，8(5)：581-592.

[10] LIU Y，CHAWLA S. Social media anomaly detection：Challenges and solutions[C]// Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. ACM，2017：817-818.

[11] BIMAL V，BASHIR M A，MARK C，et al. Towards Detecting Anomalous User Behavior in Online Social Networks[M]// Lecture Notes in Computer Science. Berlin：Springer，2014：223-238.

[12] GONG N Z，FRANK M，MITTAL P. SybilBelief：A semi-supervised learning approach for structure-based sybil detection[J]. IEEE Transactions on Information Forensics and Security，2014，9(6)：976-987.

[13] GAO P，GONG N Z，KULKARNI S，et al. Sybilframe：A defense-in-depth framework for structure-based sybil detection [J]. arXiv：1503.02985，2015.

[14] WANG B，ZHANG L，GONG N Z. SybilBlind：Detecting Fake Users in Online Social Networks without Manual Labels[J]. arXiv：1806.04853，2018.

[15] WANG B，JIA J，GONG N Z. Graph-based Security and Privacy Analytics via Collective Classification with Joint Weight Learning and Propagation[J]. arXiv：1812.01661，2018.

[16] WANG B，GONG N Z，FU H. GANG：Detecting fraudulent users in online social networks via guilt-by-association on directed graphs[C]// 2017 IEEE International Conference on Data Mining (ICDM). IEEE，2017：465-474.

[17] WANG B，ZHANG L，GONG N Z. SybilSCAR：Sybil detection in online social networks via local rule based propagation[C]// 2017 IEEE International Conference on Computer Communications. IEEE，2017.

[18] YANG C，HARKREADER R，ZHANG J L，et al. Analyzing Spammers' Social Networks For Fun and Profit：A Case Study of Cyber Criminal Ecosystem on Twitter[C]// Proceedings of the 21st International World Wide Web. New York：ACM，2012.


**SONG Chang**，born in 1995，postgraduate. Her main research interests include online social network analysis and data mining.

**YU Ke**，born in 1977，Ph. D，associate professor. Her main research interests include communication network theory，network data mining，mobile Internet application，machine learning and human-machine intelligence.