applied sciences

MDPI

Article

SybilHP: Sybil Detection in Directed Social Networks with Adaptive Homophily Prediction

Haoyu Lu [1], Daofu Gong [1,*], Zhenyu Li [1], Feng Liu [1,2] and Fenlin Liu [1,*]

[1] Henan Key Laboratory of Cyberspace Situation Awareness, Zhengzhou 450001, China; galuowazhiye@gmail.com (H.L.); zheenyuli@gmail.com (Z.L.); 202012332015274@gs.zzu.edu.cn (F.L.)
[2] School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou 450001, China
* Correspondence: gongdf@aliyun.com (D.G.); liufenlin@vip.sina.com (F.L.)

Abstract: Worries about the increasing number of Sybils in online social networks (OSNs) are amplified by a range of security issues; thus, Sybil detection has become an urgent real-world problem. Lightweight and limited data-friendly, LBP (Loopy Belief Propagation)-based Sybil-detection methods on the social graph are extensively adopted. However, existing LBP-based methods that do not utilize node attributes often assume a global or predefined homophily strength of edges in the social graph, while different user's discrimination and preferences may vary, resulting in local homogeneity differences. Another issue is that the existing message-passing paradigm uses the same edge potential when propagating belief to both sides of a directed edge, which does not agree with the trust interaction in one-way social relationships. To bridge these gaps, we present SybilHP, a Sybil-detection method optimized for directed social networks with adaptive homophily prediction. Specifically, we incorporate an iteratively updated edge homophily estimation into the belief propagation to better adapt to the personal preferences of real-world social network users. Moreover, we endow message passing on edges with directionality by a direction-sensitive potential function design. As a result, SybilHP can better capture the local homophily and direction pattern in real-world social networks. Experiments show that SybilHP works with high detection accuracy on synthesized and real-world social graphs. Compared with various state-of-the-art graph-based methods on a large-scale Twitter dataset, SybilHP substantially outperforms existing methods.

Keywords: social network; sybil detection; semi-supervised learning; belief propagation

Citation: Lu, H.; Gong, D.; Li, Z.; Liu, F.; Liu, F. SybilHP: Sybil Detection in Directed Social Networks with Adaptive Homophily Prediction. *Appl. Sci.* **2023**, *13*, 5341. https://doi.org/10.3390/app13095341

Academic Editor: Giacomo Fiumara

Received: 22 March 2023
Revised: 15 April 2023
Accepted: 18 April 2023
Published: 25 April 2023

1. Introduction

While celebrities and influencers have a huge influence on OSNs, not all of their followers are authentic human beings on the other side of the screen. It was reported that 9–15% of active Twitter users were bots [1,2]. By creating and controlling such bots, or Sybil accounts, malicious adversaries in social networks carry out spamming, phishing scams, referral traffic, and manipulating online public opinion, thereby causing a series of security problems and a crisis of trust.

In order to counter such abuse in social networks, an increasing number of Sybil-detection methods have been proposed. According to the data used, feature-based and graph-based methods are extensively mainstreamed. Feature-based methods train supervised classifiers for detection using diverse information of Sybil and normal users, such as local connections, profiles, IP addresses, and all kinds of behaviors and content features [3–8].

While graph-based methods only make use of the global structure of the social graph, and detection relies on exploiting interrelations among entities (e.g., "friendship" on Facebook or "follow" on Twitter) [9–25], GNN-based methods use both node features and structural characteristics of the OSN data to train graph neural networks (GNNs) for user classification [26–29]. This paper focuses on graph-based detection methods.

Theoretically, an underlying assumption for graph-based Sybil-detection methods is that the benign community and the Sybil community are sparsely connected; therefore, the

connections between nodes follow homophily, i.e., adjacent nodes tend to share the same label. Under this assumption, graph-based methods essentially use the block model [30] to represent a network as a set of blocks and discriminate whether an unknown node belongs to a Sybil community or a benign community. Label propagation algorithms [31,32] are a common class of methods to achieve block segmentation.

Starting from some labeled nodes, label propagation algorithms iteratively propagate the users' influence, trust, or reputation along the social connections between them, until sufficient information for label prediction is obtained. By means of propagation, most graph-based methods can be grouped into random walk (RW)-based [9–15,22] and Loopy Belief Propagation (LBP)-based [16,17,20] methods. However, excluding space and time efficiency, LBP-based methods outperform RW-based methods as they can leverage both labeled benign and Sybil data, and its nonlinear nature endows robustness against label noise [17]. However, existing LBP-based methods suffer from the following problems:

(1) They either assume a global homophily strength for all edges (e.g., GANG [20]) or predefined edge weights as homophily strength (e.g., SybilSCAR-D uses degree-normalized homophily strength [17]), while such an assumption ignores the local homophily difference of edges and, thus, fails to characterize the behavior pattern of nodes. A clear example is that users may have different capabilities to discern another benign user to follow; hence, their following edges should bear different homophily strengths.

(2) They are mostly designed for undirected (symmetric) social graph models, while many real-world platforms, such as Twitter, establish networks by a "follow", "retweet", or "thumbs up", which are asymmetric relationships. Applying these methods directly cannot make full use of edge information. The study in [20] adapts LBP-based methods for directed graphs; however, during two-way message-passing between a directed pair, its edge-potential function acts in the same way, which does not agree with the asymmetric trust relationship.

Our work: In order to overcome the limitations above, we propose a novel Sybil-detection method named SybilHP, a Sybil-detection method optimized for directed social networks with adaptive homophily prediction.

Overall, we use the LBP framework to estimate the posterior probability distribution of nodes for classification or ranking from a labeled subset of the social graph. SybilHP adapts to directed graphs by controlling belief propagation on directed edges with a novel edge-potential-function design, which integrates both the local preference of nodes and the directionality of edges. Specifically, our design involves an adaptive homophily strength estimator for the node's preference that is iteratively updated along with the belief propagation. Moreover, we incorporate a direction-sensitive mechanism into our edge-potential function to better capture the asymmetric interplay between follower and followee. We extensively analyze and evaluate the performance of SybilHP under different conditions, including different parameter settings, attack sparsity, and label noise.

The experiments on synthetic social networks show that the proposed SybilHP has relatively competitive accuracy and robustness. Then, we further evaluate SybilHP and compare it with multiple state-of-the-art methods on a large Twitter dataset. The results demonstrate that the SybilHP performs substantially better than existing methods concerning classification and ranking tasks.

2. Related Work

Both random-walk-based methods and LBP-based methods start from some labeled nodes to predict the unknown labels by semi-supervised learning. The basic idea of random-walk-based methods is that random walks starting from benign nodes tend to reach other benign nodes quickly, while it is difficult for Sybil to reach a benign node in a short walk. From this intuition, SybilGuard [10] and SybilLimit [11] have spawned many works, including SmartWalk [33], SybilIfer [12], SybilRank [9], SybilWalk [19], and Integro [15]. However, note that the training set for these methods should either consist of benign users or Sybil users but not both.

On another front, leveraging both labeled Sybil and benign nodes, LBP-based methods model the joint distribution over each node's label with Markov random fields and then use the LBP algorithm to iteratively estimate posterior probability distributions for unknown labels. Stemming from the seminal SybilBelief [16], SybilSCAR [17] integrates LBP-based and RW-based methods into a unified framework and further simplifies the local update rule for posterior estimation, which largely enhances the detection efficiency of LBP-based methods. SybilFuse [18] incorporates local graph attributes to better estimate the node priors by pre-trained classifiers and then uses LBP to compute the posteriors. Satoshi et al. [24] first showed that existing graph-based Sybil-detection methods can be interpreted in a unified framework of low-pass filtering, and then they proposed SybilHeat.

However, for directed edges, the methods mentioned above either prune the one-way edges and retain those bidirectional ones, or they directly treat all edges as undirected, causing under-use of the original edge information. Worse still, "sparse connection" or "homophily assumption" requires a significant structural gap between Sybil and benign communities. However, such a structural gap in the directed social graph can be particularly obscure because the one-way linkage is much easier to achieve, and Sybils can link to benign users as many as they want. GANG [20] adapted LBP for a directed graph by incorporating the one-way edge scenario into the edge-potential design and further derived a scalable and convergent matrix form. However, the use of a global edge weight still limits its modeling fidelity.

Based on social graph data, recent works have tended to incorporate various side information. For example, SybilHunter [23] provides a hybrid graph-based Sybil-detection approach by aggregating user social behavior patterns. Hosseini et al. [26] first applied graph convolutional networks to social robot detection. The use of GCN makes it possible to perform end-to-end learning using both node attribute information and node structure information. TrustGCN [27] used a "friend request" graph and, by combining social-graph-based defenses and graph neural networks, improved the robustness of adversarial attacks.

BotRGCN [28] and SATAR [29] applied graph convolutional networks on a Twitter follower–followee graph with multi-modal user semantics, properties, and neighborhood information. Improved CGAN [34] manages to extend imbalanced data sets before applying training classifiers to improve the detection accuracy of social bots. RoSGAS [35] is a novel reinforced and self-supervised GNN architecture search framework that adaptively pinpoints the most suitable multi-hop neighborhood and the number of layers in the GNN architecture for social bot detection. However, as the neural network-based models become more complex, the cost of training and deploying the models increases, which limits their scalability and portability.

3. Problem Formulation

3.1. Graph-Based Sybil Detection

Graph-based Sybil detection uses social graph data for detection. We model the social graph as $G(V, E)$, where we take each user as a node $u \in V$, and the directed relationship between users u and v as a directed edge (u, v). For example, "following" and "retweeting" on Twitter or sending friend requests on Facebook can be viewed as forming a directed relationship from one user to another. We call a one-way directed edge unidirectional and a two-way edge bidirectional. Note that we follow the convention of [20] that incoming neighbors, outgoing neighbors, and bidirectional neighbors are separately treated.

Each node in G should be either labeled Sybil or benign, while we only have a part of the whole picture, i.e., a labeled training set T consisting of the labeled Sybil L_s and labeled benign nodes L_b. The goal of the graph-based Sybil detection is to predict those remaining unlabeled nodes with the training set T.

3.2. Sybil Attack and Homophily

Generally, benign and Sybil communities are relatively dense subgraphs of G, and we refer to them as the Sybil region and benign region. Figure 1 shows a Sybil attack on a benign region in a social network, where b_1 and b_2 are compromised nodes attacked by s_1

and s_2. Hopefully, if benign and Sybil regions are sparsely connected, or in other words, if the density of edges between benign and Sybil regions is relatively smaller than the edges among themselves, then such relative sparsity can be partly quantified by the tendency of two linked nodes sharing the same label, i.e., *homophily*. Graph-based Sybil detection exploits homophily to infer the property of unlabeled nodes. However, it is worth noting that an effective Sybil attack can significantly weaken such homophily.

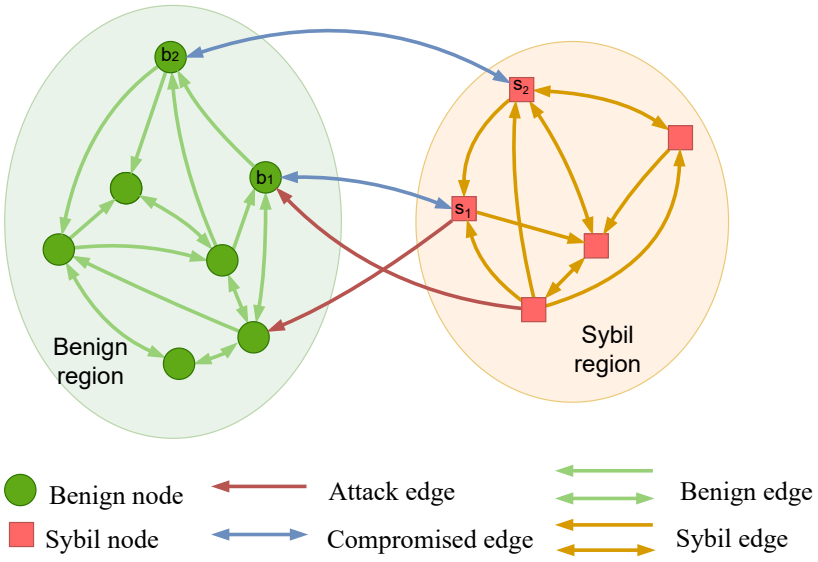

Figure 1. Sybil attack model.

3.3. LBP-Based Sybil Detection

This section first briefly recaps the basic components of the LBP-based method, and then by introducing the existing LBP-based method design, we propose our design motivations.

3.3.1. LBP Framework

As shown in Figure 2, LBP-based methods operate in the following steps. The process starts with the OSN user interaction dataset, where we associate each node $u \in V$ with a binary random variable x_u, whose state could either be -1 or 1, corresponding to benign or Sybil, respectively.

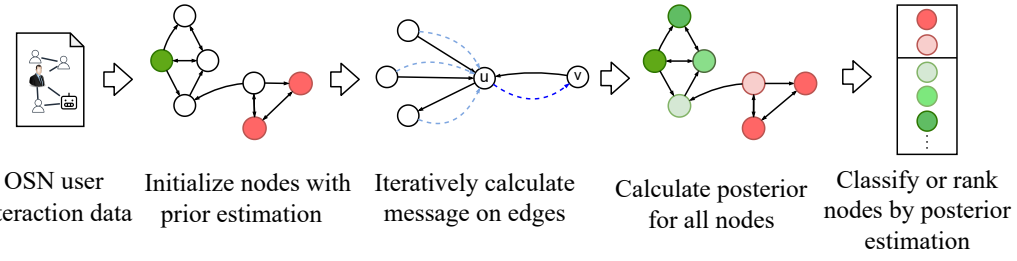

Figure 2. Overview of LBP-based methods.

We then model the joint probability distribution over all binary random variables $\boldsymbol{x_V} = \{x_u\}_{u \in V}$ as a pMRF. Specifically, pMRF factors the joint distribution $P(\boldsymbol{x_V})$ into the multiplication of a series of unary and pairwise potential functions:

$$P(\boldsymbol{x_V}) = \frac{1}{Z} \prod_{u \in V} \phi_u(x_u) \prod_{(u,v) \in E} \varphi_{uv}(x_u, x_v), \tag{1}$$

where $Z = \sum_{\boldsymbol{x_V}} \prod_{u \in V} \phi_u(x_u) \prod_{(u,v) \in E} \varphi_{uv}(x_u, x_v)$ summing over all possible combinations of $\boldsymbol{x_V}$ is the partition function used for probability normalization. The node potential $\phi_u(x_u)$

defined by Equation (2) incorporates the node prior information. Furthermore, the edge potential $\varphi_{uv}(x_u, x_v)$ will be elaborated in the next section.

$$\phi_u(x_u) =: \begin{cases} q_u, & \text{if } x_u = 1 \\ 1 - q_u, & \text{if } x_u = -1, \end{cases} \tag{2}$$

where q_u is the prior probability of node u being a Sybil. Without further prior knowledge, for those nodes labeled with Sybil, we set a soft probability q for them, and $1 - q$ for those labeled with benign. We set the prior probability to be neutral at 0.5 for the unlabeled nodes.

After initializing the prior probability that $x_u = 1$ as q_u for all nodes according to the given labels, we then implement the LBP algorithm on pMRF to estimate the posterior probability of each node x_u being a Sybil, denoted by $p_u = P(x_u = 1 \mid \boldsymbol{x_V})$.

The LBP algorithm can be summarized as the following two steps [36,37], that is, update messages on edges until convergence and then calculate p_u by message aggregation:

(1) For each edge $(u, v) \in E$, a message is sent from u to v in the tth iteration:

$$m_{uv}^{(t)}(x_v) = \sum_{x_u} \phi_u(x_u) \varphi_{uv}(x_u, x_v) \prod_{k \in N(u) \setminus v} m_{ku}^{(t-1)}(x_u). \tag{3}$$

We iteratively apply Equation (3) until the difference between $m_{uv}^{(t)}$ and $m_{uv}^{(t+1)}$ is negligible.

(2) For each node u, the posterior probability distribution of x_u can be estimated from the aggregation of all the converged messages received from its neighbors:

$$p(x_u) = \frac{1}{Z} \phi_u(x_u) \prod_{k \in N(u)} m_{ku}^{(t)}(x_u), \tag{4}$$

where $Z = \sum_{x_u} \prod_{k \in N(u)} m_{ku}^{(t)}(x_u)$ summing over possible x_u for probability normalization.

3.3.2. Existing Potential Function Designs

The LBP-based methods are differentiated mainly according to the edge-potential function $\varphi_{uv}(x_u, x_v)$, which partially reflects the distribution of a neighboring pair x_u and x_v.

SybilBelief [16] designs a potential function that encodes the coupling strength between nodes u and v as follows:

$$\varphi_{uv}(x_u, x_v) =: \begin{cases} w, & \text{if } x_u x_v = 1 \\ 1 - w, & \text{if } x_u x_v = -1. \end{cases} \tag{5}$$

Specifically, when $x_u x_v = 1$ (i.e., x_u and x_v share the same state), $\varphi_{uv}(x_u, x_v)$ takes a presumed homophily strength w ranging from 0.5 to 1, which implies the possibility that x_u coincides with x_v. Similarly, when $x_u x_v = -1$, $\varphi_{uvs.}(x_u, x_v)$ should take the heterogeneity strength between x_u and x_v, that is, $1 - w$.

SybilSCAR [17] follows a similar method to encode the interaction between nodes by neighbor influence. GANG [20], on the other hand, aims to capture the asymmetric relationship by the following design:

$$\varphi_{uv}(x_u, x_v) =: \begin{cases} w, & \text{if } x_u = 1 \text{ or } x_v = -1 \\ 1 - w, & \text{if otherwise,} \end{cases} \tag{6}$$

which is based on the intuition that, in a unidirectional edge, only if the tail is benign or the head is Sybil, then this pair tends to share the same label with homophily strength w. Otherwise, this edge tends to link two nodes with different labels. However, the use of a global homophily strength w still limits its modeling fidelity because different nodes may have different behavior patterns and, therefore, different local homophily with their neighbors.

3.3.3. Our Design Motivation

The motivation behind our potential function design lies in the following considerations.

(1) The message m_{uv} sent from node u to v represents an inference about the state x_v from u's point of view, which partially depends on the directed relationship between node u and v, i.e., following or being followed by. At the same time, the state of the message sender x_u also plays an important role in the message, i.e., $x_u = 1$ or $x_u = -1$. Therefore, we want to portray these cases with a range of initial homophily strengths instead of only one w.

(2) The message m_{uv} sent from u to v should include the discernment of u as a benign user, which can be drawn from u's following preference, and the deception capability as a Sybil, which can be drawn from its followers' statistics. To this end, we set adaptive homophily estimators to capture these local characteristics.

These motivations will guide our potential function design in the next section.

4. Methodology

Based on the motivations mentioned above, in this section, we derive finer modeling to adapt to the directed social graph, highlighting initial homophily strength parameters and adaptive homophily estimators.

4.1. Initial Homophily Strength Parameters

In this subsection, we present how SybilHP profiles the label coordination of a pair of nodes according to different edge types and the states of the message sender.

Note that LBP lets the variables pass messages to exchange their beliefs about each other until the message converges to a consensus [36,37]. Specifically, we take the message passed from u to v as an inference about x_v from u's standpoint [38,39]. However, homophily strength included in a message m_{uv} sent from u to v should differ according to message types and the state of the message sender as shown in Figure 3.

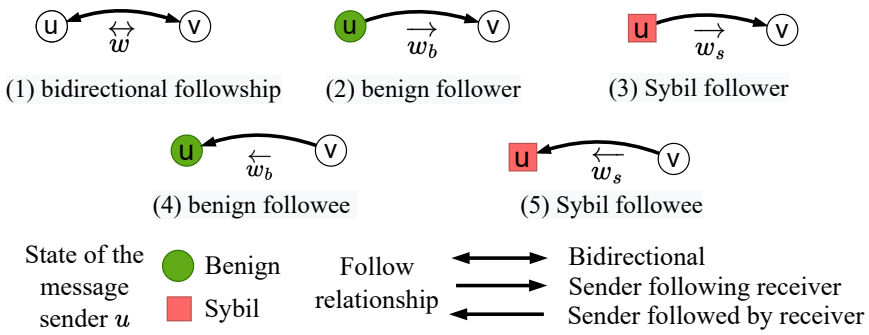

Figure 3. Message m_{uv} differs according to the edge type and the state of the sender u.

Therefore, we design five initial homophily strength parameters for these cases. Our design considerations are as follows.

4.1.1. The Case of Bidirectional Edge

(1) As shown in Figure 3(1), the bidirectional edge (u, v) represents a mutual following relationship, and it naturally implies strong homophily strength between nodes. Furthermore, for this symmetric relationship, we use one parameter $\overleftrightarrow{w}$ representing the initial homophily strength to profile the co-occurrence probability of the pair.

For the unidirectional pair, however, we need more than one parameter to describe such an asymmetric relationship in the message sent from u to v.

4.1.2. The Case When Message Sender u Is a Follower/Tail

(2) As shown in Figure 3(2), given a benign tail u, from u's standpoint, we can assume that v is also benign with high confidence because most human users hold an inherent discernment to follow human users. We use parameter $\overrightarrow{w}_b$ ("b" stands for "benign") to represent such initial homophily strength.

(3) As shown in Figure 3(3), if the tail u is a Sybil, we assume Sybils are group controlled and share a similar "following" pattern [40], that is, the probability that the head v is also Sybil, which we denote as a homophily strength parameter $\overrightarrow{w_s}$ ("s" stands for "Sybil").

4.1.3. The Case When Message Sender u Is a Followee/Head

(4) As shown in Figure 3(4), if head u is benign, from u's standpoint, we estimate that the possibility of the follower v being benign as initial homophily strength $\overleftarrow{w_b}$.

(5) As shown in Figure 3(5), if a Sybil head u is followed by v, we denote the initial homophily strength from u's perspective as $\overleftarrow{w_s}$.

4.2. Adaptive Homophily Estimator

In this subsection, we build adaptive estimators to predict the assortativity [41,42] for each node, i.e., the likelihood that an individual will form connections with other individuals. Here, we do not distinguish between homophily and assortativity, although the former is descriptive, and the latter is predictive. Furthermore, we call the estimators homophily estimators.

Our idea comes from an observation that it is uncommon for Sybils to be actively followed by human users, so the Sybil heads and benign tails tend to play more informative roles compared to benign heads and Sybil tails. To better capture this information, we maintain a pair of homophily estimators to measure a benign user's capability to resist the Sybil attack, and the Sybil's capability to make a benign user compromise. These assortative capabilities can be reflected in the statistics of the neighboring nodes' states. Therefore, for unlabeled nodes, we take their posterior probabilities in the previous iteration as their state; thus, the estimators should be updated in each iteration.

If a benign user u has already followed a certain number of Sybils, then it is safe to say that he/she will do it again. In other words, its discernment depends on the percentage of benign nodes among the nodes it follows as shown on the left of Figure 4. We thus define adaptive homophily estimator $c_{u\_b}^{(t)}$ as follows:

$$c_{u\_b}^{(t)} = \frac{\sum_{v \in N_{\text{out}}(u)} p^{(t)}(x_v = -1)}{|N_{\text{out}}(u)|}, \tag{7}$$

where $N_{\text{out}}(u)$ is the set of outgoing neighbors of u. Furthermore, $p^{(t)}(\cdot)$ is the temporary posterior probability distribution at iteration t, which is calculated by the aggregation of propagated label information at iteration $t-1$:

$$p^{(t)}(x_u) = \frac{1}{Z}\phi_u(x_u) \prod_{k \in N(u)} m_{ku}^{(t-1)}(x_u). \tag{8}$$

Similarly, a Sybil u who has already managed to obtain many benign followers can also entice one more benign follower at a small price. In other words, its deception capability depends on the percentage of benign nodes among its followers as shown on the right of Figure 4. We thus define adaptive homophily estimator $c_{u\_s}^{(t)}$ as follows:

$$c_{u\_s}^{(t)} = \frac{\sum_{v \in N_{\text{in}}(u)} p^{(t)}(x_v = 1)}{|N_{\text{in}}(u)|}, \tag{9}$$

where $N_{\text{in}}(u)$ is the set of incoming neighbors of u. $p^{(t)}(\cdot)$ is the label propagation from iteration $t-1$ defined in Equation (8). Note that we count the bidirectional linked neighbors in both $N_{\text{in}}(u)$ and $N_{\text{out}}(u)$.

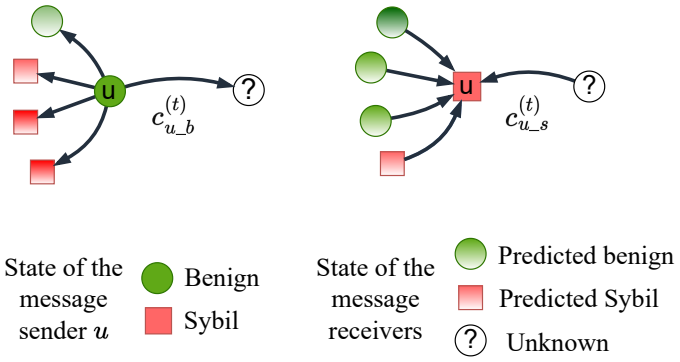

Figure 4. Homophily estimator for the edge incident with a benign tail or a Sybil head.

4.3. Redefine Potential Function

Finally, we integrate the results derived from the previous sections into our edge-potential design. For unidirectional edge (u, v), considering that message passing in the LBP algorithm goes both ways, we make our potential function direction sensitive based on the initial homophily strength parameters and then involve the adaptive homophily estimators to adapt to characteristics of nodes.

Existing work [20,21] assumes that a Sybil's following behavior is unpredictable. So, from a Sybil tail u's standpoint, we cannot gain much effective information about the head's state. Similarly, if a benign user is being followed, chances are slim that one could infer the follower's state. Therefore, in these cases, we only use the initial homophily strength parameters $\overrightarrow{w_s}$ and $\overleftarrow{w_b}$ for a rough estimation of the benign head and the Sybil tail's homophily strength as shown in Figure 5(1,3), respectively.

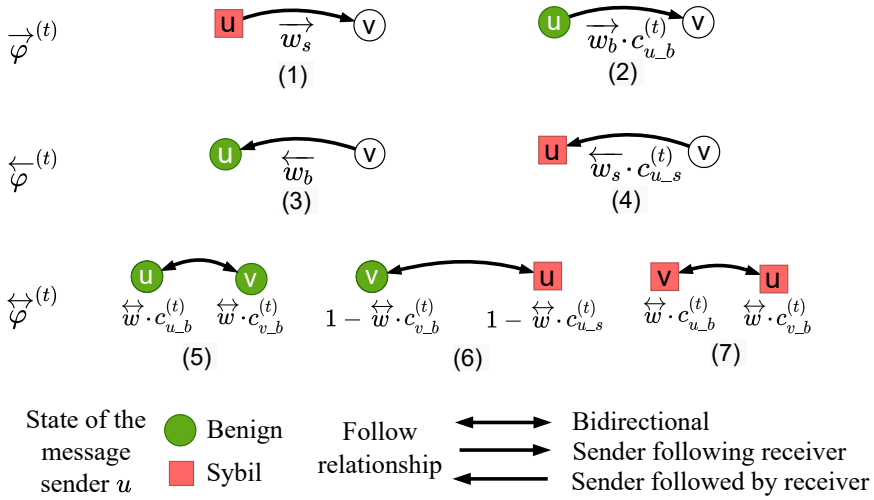

Figure 5. Initial homophily strength parameters and adaptive homophily estimators for potential function.

Dynamic homophily estimators $c^{(t)}_{u\_b}$ and $c^{(t)}_{u\_s}$ can be applied to weaken the homophily-based inference for "dumb" benign nodes who always follow Sybils and enhance the heterogeneity-based inference for "enticing" Sybil nodes who have plenty benign followers. In each iteration, they are updated to further elaborate the initial homophily strength $\overrightarrow{w_b}$ and $\overleftarrow{w_s}$ according to the nodes' preferences as shown in Figure 5(2,4).

Formally, we have unidirectional edge-potential functions as follows. When sending a message from tail u to head v of a unidirectional edge, we have $\overrightarrow{\varphi}^{(t)}(x_u, x_v)$ design:

$$\overrightarrow{\varphi}^{(t)}(x_u, x_v) = \begin{cases} \overrightarrow{w_b} c^{(t)}_{u\_b}, & x_u = -1, x_v = -1 \\ 1 - \overrightarrow{w_b} c^{(t)}_{u\_b}, & x_u = -1, x_v = 1 \\ \overrightarrow{w_s}, & x_u = 1, x_v = 1 \\ 1 - \overrightarrow{w_s}, & x_u = 1, x_v = -1. \end{cases} \tag{10}$$

When sending a message from head u to tail v, $\overleftarrow{\varphi}^{(t)}(x_u, x_v)$ follows:

$$\overleftarrow{\varphi}^{(t)}(x_u, x_v) = \begin{cases} \overleftarrow{w_b}, & x_v = -1, x_u = -1 \\ 1 - \overleftarrow{w_b}, & x_v = -1, x_u = 1 \\ \overleftarrow{w_s} c_{u\_s}^{(t)}, & x_v = 1, x_u = 1 \\ 1 - \overleftarrow{w_s} c_{u\_s}^{(t)}, & x_v = 1, x_u = -1. \end{cases} \tag{11}$$

For bidirectional edges, we adopt homophily strength $\overleftrightarrow{w}$ with the following modification:

$$\overleftrightarrow{\varphi}^{(t)}(x_u, x_v) = \begin{cases} \frac{1}{2}\overleftrightarrow{w}\left(c_{u\_b}^{(t)} + c_{v\_b}^{(t)}\right), & x_u = -1, x_v = -1 \\ 1 - \frac{1}{2}\overleftrightarrow{w}\left(c_{u\_b}^{(t)} + c_{v\_s}^{(t)}\right), & x_u = -1, x_v = 1 \\ \frac{1}{2}\overleftrightarrow{w}\left(c_{u\_s}^{(t)} + c_{v\_s}^{(t)}\right), & x_u = 1, x_v = 1 \\ 1 - \frac{1}{2}\overleftrightarrow{w}\left(c_{u\_s}^{(t)} + c_{v\_b}^{(t)}\right), & x_u = 1, x_v = -1. \end{cases} \tag{12}$$

To sum up, we have the following direction-sensitive edge-potential design:

$$\varphi^{(t)}(x_u, x_v) = \begin{cases} \overleftrightarrow{\varphi}^{(t)}(x_u, x_v), & \text{if } (u, v) \text{ bidirectional} \\ \overrightarrow{\varphi}^{(t)}(x_u, x_v), & \text{if } (u, v) \text{ unidirectional} \\ \overleftarrow{\varphi}^{(t)}(x_u, x_v), & \text{if } (v, u) \text{ unidirectional.} \end{cases} \tag{13}$$

Integrated with the proposed edge-potential function $\varphi^{(t)}(x_u, x_v)$, the pMRF model along with the LBP algorithm forms SybilHP. Given the social graph and a training set, SybilHP returns the posterior probability of nodes being Sybil in graph G for further classification or ranking tasks. Algorithm 1 summarizes the pseudo-code of SybilHP, from which, we can see that the time complexity of SybilHP is $O(iter \cdot |E|)$. As most social networks are often sparse graphs, we have $O(iter \cdot |E|) = O(iter \cdot |N|)$.

Algorithm 1 SybilHP

Require: directed social graph $G = (V, E)$,
the training set $T = (L_s, L_b)$,
the soft probability for labeled Sybil q,
initial homophily strength parameters $\overleftrightarrow{w}$, $\overrightarrow{w_b}$, $\overrightarrow{w_s}$, $\overleftarrow{w_b}$, $\overleftarrow{w_s}$,
and the number of iterations *iter*.
Ensure: posterior probability distribution $p(x_u), \forall u \in V$
for nodes $u \in V$ **do**
initialize q_u according to u,
$q_u = q$ if $u \in L_s$,
$q_u = 1 - q$ if $u \in L_b$,
$q_u = 0.5$ if u is unlabeled,
initialize $p^{(0)}(x_u) = q_u$,
initialize $m_{uv}^{(0)} = 1$ for all $(u, v) \in E$.
end for
for $t \in 1, 2, \ldots, iter$ **do**
for edges $(u, v) \in E$ **do**
compute $m_{vu}^{(t)}(x_u)$ and $m_{uv}^{(t)}(x_v)$ per Equation (3),
update $c_{u\_b}^{(t)}$, $c_{u\_s}^{(t)}$, $c_{v\_b}^{(t)}$ and $c_{v\_s}^{(t)}$ per Equations (7) and (9).
end for
for $u \in V$ **do**
update $p^{(t)}(x_u)$ per Equation (8).
end for
end for
Return $p^{(iter)}(x_u)$ for all u.

5. Experiment

5.1. Experiment Setup

Dataset description: In this section, we first evaluate the influence of various factors including Sybil attack strength, Sybil group scale, and label noise in SybilHP. Since our experiments require social networks with various forms and patterns, we synthesize benign and Sybil regions based on a real-world social graph.

(1) Synthetic-directed Pokec

For the sake of fairness and comparison, we adopted this directed Pokec network [43] from the official repository (https://github.com/binghuiwang/sybildetection (accessed on 2 July 2022)) of for GANG [20], SybilSCAR [17] and SybilBelief [16] for robustness and performance illustration. In particular, we extract a connected component that contains 10,000 nodes and 90,065 edges from Pokec as the benign region, and then we make the Sybil region a replicate of the benign region and add (bidirectional, unidirectional) attack edges between the two regions uniformly at random. If not specified, we add 500 bidirectional edges, 1000 unidirectional Sybil-to-benign attack edges, and 100 unidirectional benign-to-Sybil compromised edges as illustrated in Figure 6. We keep 100 Sybil and 100 benign users as the training set and test on the overall social graph.

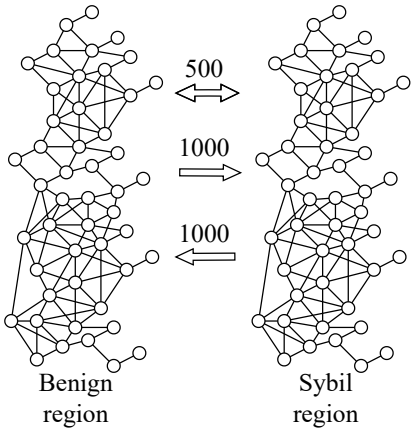

Figure 6. An abridged illustration of the synthesized Pokec dataset.

Furthermore, we compare the detection performance of the proposed method and some state-of-the-art benchmark methods on a large real Twitter dataset.

(2) Real-world Twitter follower–followee graph

By the means of breadth-first search (BFS) graph traversal, we sampled a Twitter follower–followee graph with 269,640 nodes and 6,818,501 edges from [44]. The original data was crawled by Kwak in 2009 [44]. The graph is directed and includes 41,652,230 users and 1,468,364,884 edges. However, only 10,000 Sybils and 10,000,000 benign users are labeled, which can be used as ground truth for training and testing. The remaining nodes are treated as unknown. In other words, there are less than a quarter of labeled nodes in the original dataset.

The ratio of Sybil to benign nodes is even more severely imbalanced at 1:100, which makes the discrimination of benign nodes dominate the performance evaluation results. To address the under-labeling and imbalance of the original dataset, we sampled the original dataset. The BFS starts from these labeled nodes, and we only keep those labeled neighbors until all nodes are reached. Finally, we delete those isolated nodes, and we obtain 91,263 Sybils and 178,377 benign users to form our connected and labeled social graph. We divide 9000 Sybil and 17,000 benign users (about 10%) from them as the labeled training set and test on the overall social graph.

An overview of the datasets is presented in Table 1.

Table 1. Overview of the datasets.

Datasets	# Users	# Edges	# Sybils	# Benign	$\|L_s\|$	$\|L_b\|$
Synthesized Pokec	20,000	182,130	10,000	10,000	100	100
Twitter	269,640	6,818,501	91,263	178,377	9000	17,000

Compared methods: We compared SybilHP with directed graph-based method GANG [20] (including a matrix version and a basic version) and other LBP-based methods SybilSCAR [17] (SybilSCAR-D) and SybilBelief [16]. For undirected graph-based methods, we transformed the directed graph to be undirected by only keeping those bidirectional edges (which is recommended by the original paper). Note that this can cause many nodes isolated and fail to involve in the LBP process.

Parameter setting: For SybilHP, we set the prior probability for labeled Sybil nodes $p = 0.9$, which was also suggested by authors of GANG, SybilSCAR, and SybilBelief; We assigned initial homophily strength parameters $\overleftrightarrow{w} = 0.99$, $\overrightarrow{w_s} = 0.77$, $\overrightarrow{w_b} = 0.97$, $\overleftarrow{w_b} = 0.73$, $\overleftarrow{w_s} = 0.95$, and set the number of LBP iterations $iter = 5$. For GANG, we set $w = 0.51$ adapting to Twitter as suggested by the author. Note that we also adopt the basic version of GANG with an optimized parameter ($w = 0.63$) for our Twitter dataset. We set the parameters of SybilSCAR and SybilBelief in the same way as introduced in [16,17]. An overview of the parameter settings is presented in Table 2.

Table 2. Overview of the parameter settings.

Methods	Parameter Settings
SybilBelief [16]	$w = 0.9, \theta = 0.9$
SybilSCAR [17]	$w = 0.6 (\hat{w} = 0.1), \theta = 0.6$
GANG [20]	$w = 0.51 (\hat{w} = 0.01), \theta = 0.9$
GANG_basic [20]	$w = 0.63 (\hat{w} = 0.13), \theta = 0.9$
SybilHP	$\overleftrightarrow{w} = 0.99, \overrightarrow{w_s} = 0.77, \overrightarrow{w_b} = 0.97,$ $\overleftarrow{w_b} = 0.73, \overleftarrow{w_s} = 0.95, \theta = 0.9$

Evaluation metrics: The following indicators are adopted to evaluate the performance of Bot detection methods:

- Accuracy is the fraction of instances that are correctly classified. It is a simple and intuitive metric, but it can be misleading in cases where the classes are imbalanced.
- Recall is the fraction of positive instances that are correctly classified. It is a measure of how well the model is able to identify positive instances.
- Precision is the fraction of predicted positive instances that are actually positive. It is a measure of how well the model is able to avoid false positives.
- The area under the curve (AUC) is a measure of the overall performance of a classifier. It is calculated by plotting the true positive rate (TPR) against the false positive rate (FPR) for a range of classification thresholds. A higher AUC indicates a better-performing classifier.

We implemented SybilHP in Python 3.8. For the proper comparative experiment, we also ported the original C++ codes (from the authors) of GANG, SybilSCAR, and SybilBelief to Python.

5.2. General Robustness Evaluation

We first briefly evaluate the robustness of SybilHP under different conditions including attack edges density, Sybil community scales, and labeling with noises on the Synthesized Pokec dataset.

Impact of attack edges: We add different numbers of unidirectional attack edges and bidirectional edges (compromised edges) in a ratio of 2:1. Table 3 shows the accuracy decay as the number of attacking edges increases. We found no distinct performance

difference between these homophily-inference-based methods. As the synthesized dataset does not reflect the behavioral preferences of real-world users and Sybil users, the proposed algorithm has only a slight advantage on this dataset, which could be attributed to the additional parameters.

Table 3. The classification accuracy under different numbers of attack edges.

Methods [1]	1K/0.5K	2K/1K	3K/1.5K	4K/2K	5K/2.5K
SybilBelief [16]	**0.975**	0.968	0.936	0.92	0.88
SybilSCAR [17]	0.971	0.965	0.932	**0.923**	**0.881**
GANG [20]	0.973	0.959	0.929	0.92	0.87
SybilHP	**0.975**	**0.969**	**0.938**	0.918	0.878

[1] The notation "1K/0.5K" means the number of added unidirectional attack edges is 1000 and added bidirectional attack edges is 500.

Impact of Sybil community scale: To be more realistic, we split the Sybil region to form several smaller communities to simulate the scenario when multiple Sybil clusters launch attacks on a benign user community, as shown in Figure 7, where (a), (b), and (c) correspond to one, two, and three Sybil clusters attacking the benign region, respectively. Then, the detection performance of SybilHP is examined. Specifically, we partition the Sybil region by obtaining a certain number of neighboring nodes that constitute the community through a BFS traversal starting at a certain point.

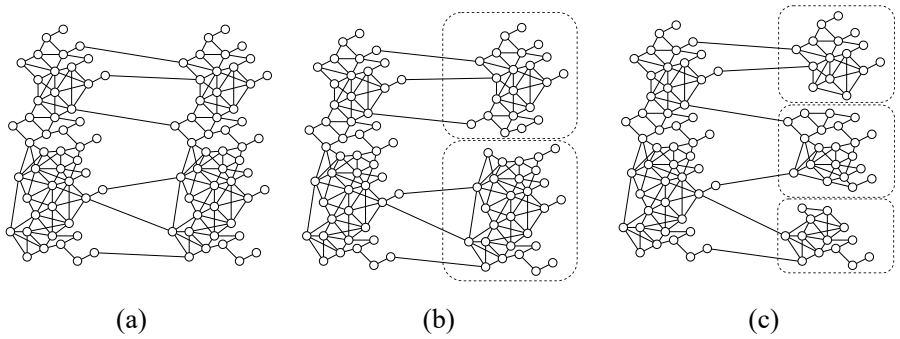

Figure 7. An abridged view of the simulated attack of multiple Sybil clusters.

It can be seen that the total number of Sybil communities increases as the number of individual community nodes decreases, which is similar to the real-world scenario in which Sybil is controlled by different organizations and individuals. We did not change the edges between Sybil and benign regions nor the training set, and the detection performance of SybilHP is shown in Figure 8. It can be seen that the size of the Sybil cluster has almost no impact on the performance of our method. Note, however, that this assumes that nodes in the training set are present in each Sybil cluster.

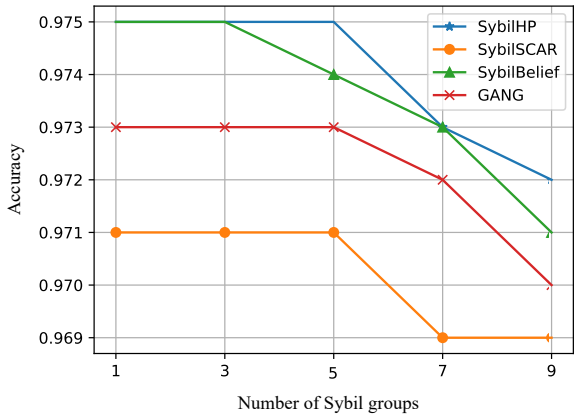

Figure 8. Robustness against attacks from different number/scale of Sybil clusters.

Impact of label noises:In the case of a partially mislabeled training set, LBP-based methods have inherent robustness against label noise. Figure 9 shows the influences of different percentages of false labels on the recall rate for Sybil. We found that SybilBelief and SybilHP showed stronger robustness against label noises compared to SybilSCAR and GANG, which could be due to their nonlinearity.

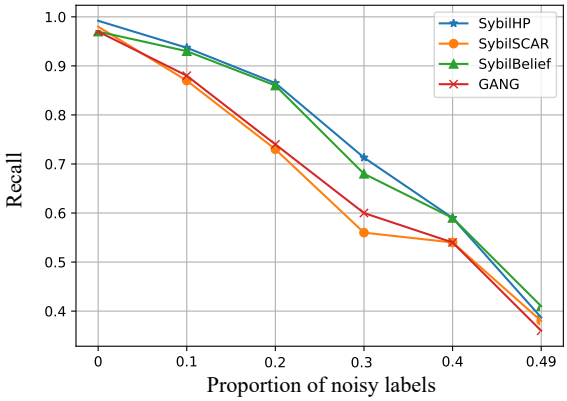

Figure 9. Robustness against noisy labels.

5.3. Comparative Experiments on Real-World Twitter Dataset

In this section, we first focus on the model parameters' adaptation to the real-world Twitter datasets and then give a comparison study with other LBP-based methods.

Model parameter adaptation: The initial homophily strength $\overleftrightarrow{w}, \overrightarrow{w_b}, \overrightarrow{w_s}, \overleftarrow{w_b}, \overleftarrow{w_s}$ for edge potential can be taken as adjustable parameters. We evaluate different configurations of these parameters by variable-controlling on the directed Twitter dataset. Figure 10 shows the variation of detection performance when we vary one of the parameters. We observe that there are points with a good trade-off between precision and recall, and we set them as the parameters for subsequent experiments.

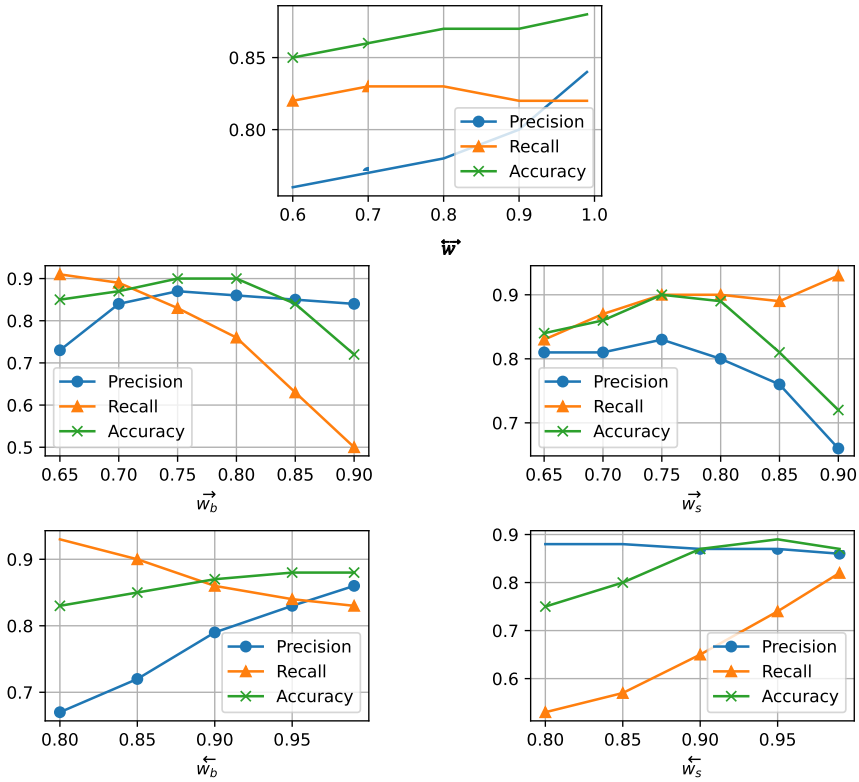

Figure 10. Detection performance under different parameter configurations. Note that the precision, recall, and accuracy are relative and have no reference value since the other parameters are fixed.

Overall classification and ranking performance: Table 4 shows the overall classification performance compared with the other state-of-the-art Sybil-detection methods. SybilHP achieved the highest precision and accuracy and second highest recall. Note that the SybilHP_basic in the table does not incorporate the adaptive homophily estimator mechanism, we see that, with the ablation of the adaptive homophily estimator enhancement, SybilHP still shows superiority over other methods.

Table 4. Classification performance.

Methods	Precision	Recall	Accuracy
SybilBelief [16]	0.873	0.501	0.806
SybilSCAR [17]	0.905	0.508	0.815
GANG_matrix [20]	0.798	0.446	0.74
GANG_basic [20]	0.757	**0.808**	0.847
SybilHP	**0.908**	0.797	**0.904**
SybilHP_basic	0.897	0.764	0.893

As LBP-based detection methods estimate the posterior probability for each node to be Sybil, we can rank the nodes by the posterior probabilities to produce a more thorough performance analysis. We take the Area Under the Receiver Operating Characteristic Curve (AUC) as the evaluation measure for ranking, which can be interpreted as the probability that a randomly sampled Sybil node is ranked higher than a randomly sampled benign node in the testing dataset. Figure 11 shows the overall ranking performance by AUC, and we make the following observations.

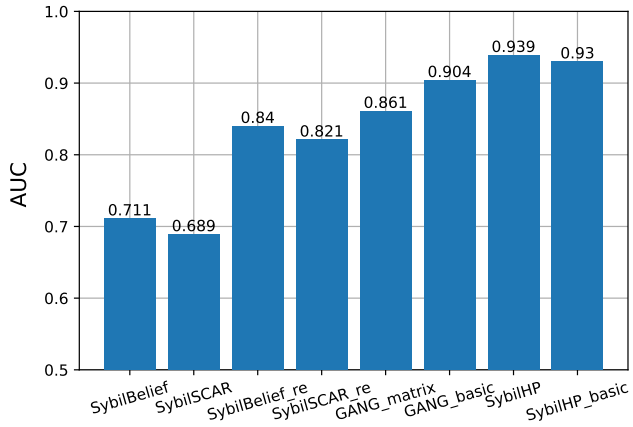

Figure 11. AUCs of compared methods.

First, we found that methods designed for directed social graphs substantially outperformed those methods for undirected graphs. To adapt to methods for undirected graphs, we only kept reciprocal edges in the original social graph (as the original papers suggest), which resulted in some isolated nodes that were unable to be involved in the computation. This is the main reason for the lower AUC performance compared with that reported in the original paper. However, even when we reevaluate these methods after excluding these isolated nodes, their performance is still limited by the loss of directed information, as SybilSCAR_re and SybilBelief_re shown in Figure 11.

Second, we can see that complexity and more refined modeling facilitate the capturing of the characteristics of the data. SybilBelief, the basic version of GANG and SybilHP are more effective than their simplified versions because the computational efficiency can hardly be balanced with accuracy.

Finally, we observed that, without an adaptive homophily estimator mechanism, SybilHP_basic still had decent performance. Furthermore, the performance improvement brought by adaptive homophily estimators is not significant, so the balance of performance and computational cost should be properly considered in practical applications.

Sybil nodes in top-ranked nodes: Since the ranking of nodes can be used as a priority list to do further inspection and verification by system or humans, the accuracy in the top-ranked nodes is important because extra costs for human workers will rule out the majority of nodes. Therefore, we further compare the proportion of Sybil in different fractions of top-80,000 positive-reported nodes. Specifically, we divide top-80,000 nodes (because the dataset only contains around 90,000 Sybils) into 10 intervals and calculate the number of Sybils in each interval.

Figure 12 shows the distribution of Sybils detected in each 10,000 interval. For GANG_matrix, SybilSCAR, and SybilBelief, we can observe a clear drop at the interval 50-60k, while the proposed SybilHP proceeds with its superiority. We speculate that a group of Twitter users with a particular following pattern, "dumb benign followers" or "intriguing Sybils" could have managed to evade these detection methods. However, SybilHP has captured their pattern and discovered them.

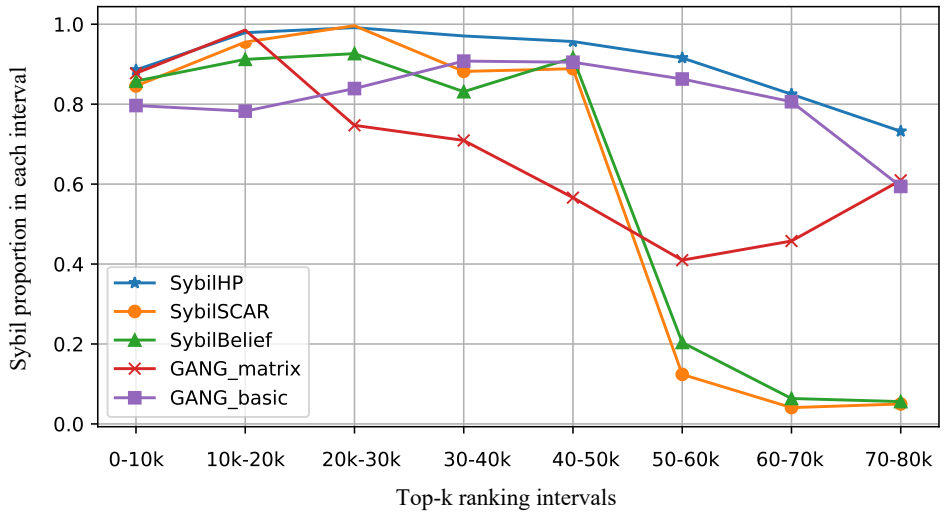

Figure 12. Sybil proportion in each 10,000 interval of top 80,000 positive-reported nodes.

6. Conclusions

In this work, we proposed SybilHP, a Sybil-detection method optimized for directed social networks with adaptive homophily prediction. The proposed algorithm features a novel edge homophily estimator, which is updated iteratively to adapt to the dynamics of homophily between Sybil and benign users in real-world social networks. It also endows message passing on edges with directionality by a direction-sensitive potential function design. We analyzed and compared SybilHP with multiple state-of-the-art graph-based detection methods using a real-world Twitter dataset, and the proposed method achieved superior performance.

7. Discussion of Future Research Directions

In fact, a Sybil-detection system using the proposed method is suitable to be deployed in Twitter, Instagram, TikTok, and many other social media that forms directed social networks. Our future research will focus on a richer range of graph data sources. For example, Sybil detection on a heterogeneous social network can be promising and challenging work. Furthermore, research on time-series homophily also has the potential to shed light on social user behavior analysis and Sybil detection.

Author Contributions: Conceptualization, H.L.; methodology, H.L. and D.G.; software, H.L.; validation, F.L. (Fenlin Liu); formal analysis, Z.L.; investigation, H.L.; resources, F.L. (Feng Liu); data curation, F.L. (Feng Liu); writing—original draft preparation, H.L.; writing—review and editing, Z.L.; visualization, H.L.; supervision, F.L. (Fenlin Liu); project administration, D.G.; funding acquisition, D.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant Nos. 62002387, 61872448, 61772549, and U1804263) and the Science and Technology Research Project of Henan Province, China.

Data Availability Statement: The dataset used in this paper is published on figshare. DOI: 10.6084/m9.figshare.20057300.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Varol, O.; Ferrara, E.; Davis, C.; Menczer, F.; Flammini, A. Online human-bot interactions: Detection, estimation, and characterization. In Proceedings of the International AAAI Conference on Web and Social Media, Montreal, QC, Canada, 15–18 May 2017; Volume 11.
2. Gabielkov, M.; Legout, A. The complete picture of the Twitter social graph. In Proceedings of the 2012 ACM Conference on CoNEXT Student Workshop, Nice, France, 10 December 2012; pp. 19–20.
3. Guo, Q.; Xie, H.; Li, Y.; Ma, W.; Zhang, C. Social Bots Detection via Fusing BERT and Graph Convolutional Networks. *Symmetry* **2022**, *14*, 30. [CrossRef]
4. Yang, Z.; Wilson, C.; Wang, X.; Gao, T.; Zhao, B.Y.; Dai, Y. Uncovering social network sybils in the wild. *ACM Trans. Knowl. Discov. Data* **2014**, *8*, 1–29. [CrossRef]
5. Lee, K.; Caverlee, J.; Webb, S. Uncovering social spammers: Social honeypots+ machine learning. In Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Geneva, Switzerland, 19–23 July 2010; pp. 435–442.
6. Cao, Q.; Yang, X.; Yu, J.; Palow, C. Uncovering large groups of active malicious accounts in online social networks. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, 3–7 November 2014; pp. 477–488.
7. Hu, X.; Tang, J.; Gao, H.; Liu, H. Social spammer detection with sentiment information. In Proceedings of the 2014 IEEE International Conference on Data Mining, Shenzhen, China, 14–17 December 2014; pp. 180–189.
8. Jiang, M.; Cui, P.; Beutel, A.; Faloutsos, C.; Yang, S. Catchsync: Catching synchronized behavior in large directed graphs. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 941–950.
9. Cao, Q.; Sirivianos, M.; Yang, X.; Pregueiro, T. Aiding the detection of fake accounts in large scale social online services. In Proceedings of the Ninth {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 12), San Jose, CA, USA, 25–27 April 2012; pp. 197–210.
10. Yu, H.; Kaminsky, M.; Gibbons, P.B.; Flaxman, A. Sybilguard: Defending against sybil attacks via social networks. In Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, Pisa, Italy, 11–15 September 2006; pp. 267–278.
11. Yu, H.; Gibbons, P.B.; Kaminsky, M.; Xiao, F. Sybillimit: A near-optimal social network defense against sybil attacks. In Proceedings of the 2008 IEEE Symposium on Security and Privacy (SP 2008), Oakland, CA, USA, 18–21 May 2008; pp. 3–17.
12. Danezis, G.; Mittal, P. Sybilinfer: Detecting sybil nodes using social networks. In Proceedings of the NDSS, San Diego, CA, USA, 8–11 February 2009; pp. 1–15.
13. Mohaisen, A.; Hopper, N.; Kim, Y. Keep your friends close: Incorporating trust into social network-based sybil defenses. In Proceedings of the 2011 Proceedings IEEE INFOCOM, Shanghai, China, 10–15 April 2011; pp. 1943–1951.
14. Yang, C.; Harkreader, R.; Zhang, J.; Shin, S.; Gu, G. Analyzing spammers' social networks for fun and profit: A case study of cyber criminal ecosystem on twitter. In Proceedings of the 21st International Conference on World Wide Web, Lyon, France, 16–20 April 2012; pp. 71–80.
15. Boshmaf, Y.; Logothetis, D.; Siganos, G.; Lería, J.; Lorenzo, J.; Ripeanu, M.; Beznosov, K.; Halawa, H. Íntegro: Leveraging victim prediction for robust fake account detection in large scale OSNs. *Comput. Secur.* **2016**, *61*, 142–168. [CrossRef]
16. Gong, N.Z.; Frank, M.; Mittal, P. Sybilbelief: A semi-supervised learning approach for structure-based sybil detection. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 976–987. [CrossRef]
17. Wang, B.; Zhang, L.; Gong, N.Z. SybilSCAR: Sybil detection in online social networks via local rule based propagation. In Proceedings of the IEEE INFOCOM 2017-IEEE Conference on Computer Communications, Atlanta, GA, USA, 1–4 May 2017; pp. 1–9.
18. Gao, P.; Wang, B.; Gong, N.Z.; Kulkarni, S.R.; Thomas, K.; Mittal, P. Sybilfuse: Combining local attributes with global structure to perform robust sybil detection. In Proceedings of the 2018 IEEE Conference on Communications and Network Security (CNS), Beijing, China, 30 May–1 June 2018; pp. 1–9.
19. Jia, J.; Wang, B.; Gong, N.Z. Random walk based fake account detection in online social networks. In Proceedings of the 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Denver, CO, USA, 26–29 June 2017; pp. 273–284.

20. Wang, B.; Gong, N.Z.; Fu, H. GANG: Detecting fraudulent users in online social networks via guilt-by-association on directed graphs. In Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 18–21 November 2017; pp. 465–474.
21. Breuer, A.; Eilat, R.; Weinsberg, U. Friend or faux: Graph-based early detection of fake accounts on social networks. In Proceedings of the Web Conference 2020, Taipei, Taiwan, 20–24 April 2020; pp. 1287–1297.
22. Zhang, X.; Xie, H.; Lui, J.C. Sybil detection in social-activity networks: Modeling, algorithms and evaluations. In Proceedings of the 2018 IEEE 26th International Conference on Network Protocols (ICNP), Cambridge, UK, 25–27 September 2018; pp. 44–54.
23. Mao, J.; Li, X.; Luo, X.; Lin, Q. SybilHunter: Hybrid graph-based sybil detection by aggregating user behaviors. *Neurocomputing* **2022**, *500*, 295–306. [CrossRef]
24. Liu, Y.; Li, Z.; Liang, X.; Liu, Z. Interpreting Graph-based Sybil Detection Methods as Low-pass Filtering. *arXiv* **2022**, arXiv:2206.10835.
25. Li, X.; Lin, Q.; Mao, J. Hybrid graph-based Sybil detection with user behavior patterns. *Procedia Comput. Sci.* **2021**, *187*, 607–612. [CrossRef]
26. Ali Alhosseini, S.; Bin Tareaf, R.; Najafi, P.; Meinel, C. Detect me if you can: Spam bot detection using inductive representation learning. In Proceedings of the Companion Proceedings of The 2019 World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 148–153.
27. Sun, Y.; Yang, Z.; Dai, Y. TrustGCN: Enabling graph convolutional network for robust sybil detection in OSNs. In Proceedings of the 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Virtually, 7–10 December 2020; pp. 1–7.
28. Feng, S.; Wan, H.; Wang, N.; Luo, M. BotRGCN: Twitter bot detection with relational graph convolutional networks. In Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Virtually, 8–11 November 2021; pp. 236–239.
29. Feng, S.; Wan, H.; Wang, N.; Li, J.; Luo, M. Satar: A self-supervised approach to twitter account representation learning and its application in bot detection. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Virtually, 1–5 November 2021; pp. 3808–3817.
30. Scott, J. *Social Network Analysis*; Sage Publications Ltd.: Thousand Oaks, CA, USA, 2000.
31. Zhu, X.; Ghahramani, Z. Learning from labeled and unlabeled data with label propagation. In Proceedings of the 20th International Conference on Machine Learning, Washington, DC, USA, 21–24 August 2003; Volume 20, pp. 936–943.
32. Raghavan, P.; Albert, R.; Kumara, S. Label propagation in social networks. In Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, CA, USA, 12–15 August 2007; pp. 613–622.
33. Liu, Y.; Ji, S.; Mittal, P. Smartwalk: Enhancing social network security via adaptive random walks. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 24–28 October 2016; pp. 492–503.
34. Wu, B.; Liu, L.; Yang, Y.; Zheng, K.; Wang, X. Using improved conditional generative adversarial networks to detect social bots on Twitter. *IEEE Access* **2020**, *8*, 36664–36680. [CrossRef]
35. Yang, Y.; Yang, R.; Li, Y.; Cui, K.; Yang, Z.; Wang, Y.; Xu, J.; Xie, H. RoSGAS: Adaptive Social Bot Detection with Reinforced Self-Supervised GNN Architecture Search. *ACM Trans. Web* 2022, *Accepted*. [CrossRef]
36. Koller, D.; Friedman, N. *Probabilistic Graphical Models: Principles and Techniques*; MIT Press: Cambridge, MA, USA, 2009.
37. Koller, D.; Friedman, N.; Getoor, L.; Taskar, B. Graphical models in a nutshell. *Introd. Stat. Relational Learn.* **2007**, *43*, 125577081.
38. Yedidia, J.S.; Freeman, W.T.; Weiss, Y. Understanding belief propagation and its generalizations. *Explor. Artif. Intell. New Millenn.* **2003**, *8*, 236–239.
39. Jordan, M.I. *An Introduction to Probabilistic Graphical Models*; University of California: Berkeley, CA, USA, 2003.
40. Fu, H.; Xie, X.; Rui, Y.; Gong, N.Z.; Sun, G.; Chen, E. Robust spammer detection in microblogs: Leveraging user carefulness. *ACM Trans. Intell. Syst. Technol.* **2017**, *8*, 1–31. [CrossRef]
41. Newman, M.E.J. Mixing patterns in networks. *Phys. Rev. E* **2003**, *68*, 026126. [CrossRef] [PubMed]
42. Clauset, A.; Shalizi, C.R.; Newman, M.E. Power-law distributions in empirical data. *SIAM Rev.* **2009**, *51*, 661–703. [CrossRef]
43. Leskovec, J.; Krevl, A. SNAP Datasets: Stanford Large Network Dataset Collection. 2014. Available online: http://snap.stanford.edu/data (accessed on 15 April 2023).
44. Kwak, H.; Lee, C.; Park, H.; Moon, S. What is Twitter, a social network or a news media? In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; pp. 591–600.