Sybilアカウント検知手法のグラフ信号処理的解釈とその応用

古谷 諭史[1,2,a)]　芝原 俊樹[1]　秋山 満昭[1]　会田 雅樹[2]

概要：オンラインソーシャルネットワーク (OSN) 上で偽のアカウント（Sybil アカウント）を大量に作成し，スパム・フィッシング URL の拡散や誤情報の流布による世論・市場の操作など様々な悪質な活動を行う Sybil 攻撃への対策は OSN セキュリティにおける重要な課題の一つである．この課題に対して，これまで様々な Sybil 検知手法が提案されてきたが，それらの多くは各手法の性能やノイズ耐性に関して実験的に評価されているのみで，理論的な理解は不十分である．本研究では，既存のグラフベースの Sybil 検知手法はグラフ信号処理の枠組みで統一的に解釈できることを示す．これにより，各手法をフィルタカーネルの特性とシフト行列のスペクトルの 2 つの観点から理論的に比較・解析することが可能となる．さらに，解析に基づいて，我々は新たな Sybil 検知手法 (SybilHeat) を提案する．数値実験を通して，提案手法が高い性能及びノイズ耐性を発揮することを確かめた．本研究はグラフベースの Sybil 検知の理論的基礎を築き，Sybil 検知手法のより良い理解に繋がるものである．

キーワード：オンラインソーシャルネットワーク，Sybil 検知，グラフ信号処理

Graph Signal Processing Interpretation of Sybil Accounts Detection Methods and Its Applications

Satoshi Furutani[1,2,a)]　Toshiki Shibahara[1]　Mitsuaki Akiyama[1]　Masaki Aida[2]

Abstract: Online social networks (OSNs) are threatened by Sybil attacks, in which attackers create a large number of fake accounts (also called Sybils) on OSNs and exploit them for various malicious activities. Therefore, Sybil detection is a fundamental task for OSN security. Although various methods have been proposed recently, theoretical understanding of them is still lacking. In this study, we show that existing graph-based Sybil detection methods can be interpreted in a unified framework of low-pass filtering. This framework enables us to theoretically compare and analyze each method from two perspectives: filter kernel properties and the spectrum of shift matrices. Furthermore, on the basis of the analysis, we propose a novel Sybil detection method called SybilHeat. Numerical experiments demonstrate that SybilHeat performs consistently well on graphs with various structural properties. This study lays a theoretical foundation for graph-based Sybil detection and leads to a better understanding of Sybil detection methods.

Keywords: Online social networks, Sybil detection, graph signal processing

1. はじめに

ソーシャルメディアは，人々の情報発信や交流の主要なプラットフォームとして生活の中に深く浸透し，多くの人々によって日常的に利用されている．しかし，ソーシャルメディアは Sybil 攻撃の脅威に晒されている．Sybil とは，オンライン上での様々な悪質な活動のために作成・運用される偽のボットアカウントであり，最近でも反ワクチンメッセージの流布 [1, 2] や政治的議論の操作 [3, 4] に

[1] NTT 社会情報研究所
NTT Social Informatics Laboratories
[2] 東京都立大学
Tokyo Metropolitan University
[a)] satoshi.furutani.ek@hco.ntt.co.jp

Sybil が悪用されていることが報告されている．このため，Sybil の検知・対策はソーシャルメディアのセキュリティにおける重要な研究課題である．

既存の Sybil 検知手法の多くはオンラインソーシャルネットワーク (OSN) のユーザ間の関係性（グラフ構造）に基づいて検知を行う．攻撃者は Sybil アカウントの作成やそれらの間の繋がりは自由に操作できるが，正規のユーザとの繋がりは操作できない．この結果，OSN において Sybil は Sybil 同士でコミュニティを形成する一方で，正規ユーザとの結び付きは疎になる傾向があるため，OSN の構造を上手く利用することで Sybil と正規ユーザの区別が可能となる [5]．ほとんどのグラフ構造ベースの Sybil 検知手法は，既知の頂点ラベルに基づいて頂点の事前評価値を設定し，グラフ上で評価値を局所的に更新・伝播することで未知の頂点のラベルを推定する．このとき，評価値の更新・伝播アルゴリズムとしてランダムウォークベースの方法と信念伝播法ベースの方法の 2 つがある．ランダムウォークベースの方法では，グラフ上で既知の Sybil または正規ノードからのランダムウォークによって評価値の更新・伝播を行う [6–9]．一方，信念伝播法ベースの方法では，OSN の構造を pairwise Markov Random Field (MRF) としてモデル化して，信念伝播法やその近似手法によって各頂点の周辺分布を計算することで Sybil 検知を行う [10–14]．

上述の通り，これまで様々な Sybil 検知手法が提案されてきたが，ほとんどの場合は各手法の検知性能やノイズ耐性などに関して実験的に比較がなされているのみである．一般に，実験結果は実験に用いるデータの特性や実験条件にしばしば依存するため，検知手法の性能を深く理解するためにはこれらを理論的に比較する必要がある．本研究では，既存のグラフベースの Sybil 検知手法はグラフ信号のローパスフィルタリングの枠組みで統一的に解釈できることを示す．これにより，各手法を「フィルタカーネルの特性」と「シフト行列のスペクトル」の 2 つの観点から理論的に比較・解析することが可能となる．解析を通して，我々は Sybil 検知手法が高い性能を発揮するためには，(i) フィルタカーネルが適切に低周波（高周波）成分を強調（除去）し，(ii) シフト行列の低周波固有ベクトルが高いコミュニティ識別性能を有することが必要であることを明らかにした．これを踏まえて，我々は上記の要件を満たすフィルタカーネルとシフト行列を持つ新しい検知手法 (SybilHeat) を提案する．数値実験を通して，我々の提案手法がグラフのコミュニティ構造の変化に依らず一貫して高い検知性能を示し，ラベルノイズの割合が小さい範囲では高いノイズ耐性を有することを確認した．

2. 準備

2.1 グラフ信号処理

この節ではグラフ信号処理 [15, 16] の基本概念について導入する．自己ループと多重辺のない重みなし無向グラフ $G = (V, E)$ を考える．ここで，$V = \{1, \dots, N\}$ は頂点集合，$E \subset V \times V$ は辺の集合である．グラフ信号はノード上で定義される関数 $x : N \to \mathbb{R}$ であり，N 次元ベクトル $\boldsymbol{x} = (x_1, x_2, \dots, x_N)$ によって表現される．シフト行列 $\boldsymbol{S} = [S_{ij}] \in \mathbb{R}^{N \times N}$ は $(i, j) \notin E$ のときかつそのときに限り $X_{ij} \neq 0$ となる．シフト行列をグラフ信号 $\boldsymbol{x}$ に作用させると，シフトされた信号 $\tilde{\boldsymbol{x}}$ の各要素はその隣接ノードの信号値の線型結合となり．グラフトポロジーに対してグラフ信号をシフトさせることができる．一般的には，シフト行列として隣接行列や Laplacian 行列が用いられる．隣接行列 $\boldsymbol{A} = [A_{ij}] \in \mathbb{R}^{N \times N}$ は $(i, j) \in E$ のとき $A_{ij} = 1$，$(i, j) \notin E$ のとき $A_{ij} = 0$ と定義される実対称行列である．ラプラシアン行列は $\boldsymbol{L} := \boldsymbol{D} - \boldsymbol{A}$ と定義される．ただし，$\boldsymbol{D} := \mathrm{diag}(d_1, d_2, \dots, d_N)$ は次数行列であり，$d_i := \sum_{i=1}^{N} A_{ij}$ はノード i の次数である．

シフト行列 $\boldsymbol{S}$ の固有値 $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ と対応する固有ベクトル $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_N$ に対して，対角行列 $\boldsymbol{\Lambda} := \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ と $\boldsymbol{V} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_N)$ を定義する．このとき，グラフ信号 $\boldsymbol{x}$ に対するグラフフーリエ変換は $\hat{\boldsymbol{x}} = \boldsymbol{V}^{-1}\boldsymbol{x}$ で定義される．また，逆変換は $\boldsymbol{x} = \boldsymbol{V}\hat{\boldsymbol{x}}$ で定義される．入力信号 $\boldsymbol{x}_{\mathrm{in}}$ に対するグラフフィルタリング（グラフ畳み込み）は次式で定義される：

$$\boldsymbol{x}_{\mathrm{out}} = \boldsymbol{V} h(\boldsymbol{\Lambda}) \boldsymbol{V}^{-1} \boldsymbol{x}_{\mathrm{in}} \tag{1}$$

ここで，$h(\boldsymbol{\Lambda}) := \mathrm{diag}(h(\lambda_1), h(\lambda_2), \dots, h(\lambda_N))$ であり，$h(\lambda)$ は区間 $[\lambda_1, \lambda_N]$ 上のフィルタである．古典的な信号処理における畳み込みの定義と同様，頂点領域のグラフ信号はグラフフーリエ変換によって周波数領域の信号に変換され，フィルタ $h(\lambda)$ と掛け合わせて，逆変換によって頂点領域の信号に戻される．これによって，元の信号の特定の周波数成分を強調あるいは除去した信号が出力される．

3. Sybil 検知手法のフィルタリング解釈

無向グラフ $G = (V, E)$ において，既知の Sybil ノード集合を V_s, 既知の正規ノード集合を V_b とする．既存のグラフベースの Sybil 検知手法は，所与の各ノードの事前評価値 $\boldsymbol{q} = (q_1, q_2, \dots, q_N)^\top$ とグラフ G の構造情報をもとに，各ノードのステップ t での評価値 $\boldsymbol{p}^{(t)} = (p_1^{(t)}, p_2^{(t)}, \dots, p_N^{(t)})^\top$ を何らかの更新則 f に従って

$$\boldsymbol{p}^{(t)} = f\left(\boldsymbol{p}^{(t-1)}; \boldsymbol{q}, G\right) \tag{2}$$

を収束するまで逐次的に更新することで評価値を計算し，最終的な評価値 $\boldsymbol{p} = \lim_{t \to \infty} \boldsymbol{p}^{(t)}$ に基づいて未知ノード $i \in V \setminus (V_s \cup V_b)$ が Sybil か否かを判定する手法として理解できる [17]．事前評価値 $\boldsymbol{q}$ や更新則 $f(\cdot)$ がどのようなものかは手法によって異なる．

本節では，式 (2) をのローパスフィルタリングとして定式

化することを考える．すなわち, 式 (2) を $\boldsymbol{p} = \boldsymbol{V} h(\boldsymbol{\Lambda}) \boldsymbol{V}^{-1} \boldsymbol{q}$ という形で表現する．ここで，$\boldsymbol{\Lambda}$ 及び $\boldsymbol{V}$ はそれぞれあるシフト行列 $\boldsymbol{S}$ の固有値の対角行列と固有ベクトルを並べた行列である．$h(\cdot)$ はローパスフィルターカーネルである．この定式化により，既存の Sybil 検知手法にグラフ信号処理的な解釈を与え，異なる手法を同じ観点から理論的に比較することが可能になる．以下では，次の代表的な Sybil 検知手法のローパスフィルタリング解釈について述べる：CIA [6], SybilRank [7], SybilWalk [9], SybilBelief [10], SybilSCAR [14]. なお，ここでは，簡単のため，我々は重みなしの無向グラフを考えるが，我々のアプローチを重みありの無向グラフへ拡張することは容易である．

3.1 CIA

CIA [6] は，既知の Sybil ノードからリスタートありランダムウォークによってノードの悪性スコアを伝播し，Sybil 検知を行う手法であり，$0 < \alpha < 1$ に対して更新式は次式で与えられる：

$$\boldsymbol{p}^{(t)} = \alpha \boldsymbol{A} \boldsymbol{D}^{-1} \boldsymbol{p}^{(t-1)} + (1-\alpha) \boldsymbol{p}^{(0)}, \tag{3}$$

ここで，ノード i の初期スコアは $i \in V_s$ のとき $p_i^{(0)} = 1$ であり，$i \notin V_s$ のとき $p_i^{(0)} = 0$ である．事前評価値 $\boldsymbol{q} = \boldsymbol{p}^{(0)}$ とすると，式 (3) はランダムウォーク Laplacian 行列 $\boldsymbol{\mathcal{L}}_{\mathrm{rw}} := \boldsymbol{I} - \boldsymbol{A}\boldsymbol{D}^{-1}$ の固有値対角行列 $\boldsymbol{\Lambda}$ と固有ベクトル行列 $\boldsymbol{V}$ を用いて

$$\boldsymbol{p} = \boldsymbol{V}(1-\alpha)(\boldsymbol{I} - \alpha(\boldsymbol{I} - \boldsymbol{\Lambda}))^{-1} \boldsymbol{V}^{-1} \boldsymbol{q}. \tag{4}$$

と書くことができる．

3.2 SybilRank

SybilRank [7] は，既知の正規ノードからの打ち切りランダムウォークの滞在確率を求めることによって各ノードの信頼スコアを計算する．これは Sybil ノードと正規ノードの結びつきは疎なので，正規ノードから出発したランダムウォークを有限なステップで打ち切ることによって，ウォーカーは Sybil ノードに到達しにくくなり，結果として滞在確率は正規ノードほど高く，Sybil ノードほど低くなるという仮説に基づく．ノード i の初期スコアは $i \in V_b$ のとき $p_i^{(0)} = 1/|V_b|$，$i \notin V_b$ のとき $p_i^{(0)} = 0$ であり，信頼スコア $\boldsymbol{p}^{(t)}$ は

$$\boldsymbol{p}^{(t)} = \boldsymbol{A}\boldsymbol{D}^{-1} \boldsymbol{p}^{(t-1)} \tag{5}$$

で更新される．SybilRank はこの更新式を $\Gamma = O(\log N)$ ステップで打ち切り，その後で次数のバイアスの影響を取り除くために信頼スコアを次数で正規化することにより計算される (i.e., $p_i = p_i^{(\Gamma)}/d_i$).

$\boldsymbol{p}^{(\Gamma)} = (\boldsymbol{A}\boldsymbol{D}^{-1})^{\Gamma} \boldsymbol{p}^{(0)}$ より，事前評価値 $\boldsymbol{q} = \boldsymbol{p}^{(0)}$ とすると，最終的な信頼スコアは

$$\boldsymbol{p} = \boldsymbol{D}^{-1}(\boldsymbol{I} - \boldsymbol{\mathcal{L}}_{\mathrm{rw}})^{\Gamma} \boldsymbol{q} = \boldsymbol{D}^{-1} \boldsymbol{V} (\boldsymbol{I} - \boldsymbol{\Lambda})^{\Gamma} \boldsymbol{V}^{-1} \boldsymbol{q} \tag{6}$$

のように記述できる．ゆえに，SybilRank は $\boldsymbol{\mathcal{L}}_{\mathrm{rw}}$ によるローパスフィルタリングと次数正規化の組み合わせとして解釈できる．

3.3 SybilWalk

SybilWalk [9] は，元のグラフ G に 2 つのラベルノード（Sybil ラベルノード l_s 及び正規ラベルノード l_b）を追加したグラフ $\widehat{G} = (V \cup \{l_b, l_s\}, \widehat{E})$ 上でのランダムウォークによって各ノードの悪性スコアを計算する．ここで，グラフ $\widehat{G}$ において，l_b と l_s はそれぞれ既知の Sybil ノード及び正規ノードと接続される．ラベルノードの悪性スコアは $p_{l_b} = 0, p_{l_s} = 1$ で与えられ，ノード $i \in V$ の悪性スコアはノード i から出発したランダムウォークが l_b に到達する前に l_s に到達する確率として

$$p_i^{(t)} = \sum_{j=1}^{N} \frac{a_{ij}}{\widehat{d}_i} p_j^{(t-1)} \tag{7}$$

のように計算される．ここで，$\widehat{d}_i$ はグラフ $\widehat{G}$ におけるノード i の次数であり，ノード i が $V_b \cup V_s$ に属する場合は $\widehat{d}_i = d_i + 1$ で，属さない場合は $\widehat{d}_i = d_i$ である．

実は，SybilWalk は l_b と l_s を吸収ノードとする吸収的 Markov 連鎖に等しい．$\widehat{G}$ 上でのランダムウォークにおいて，ユーザノード間での遷移は $\widehat{\boldsymbol{D}}^{-1}\boldsymbol{A} \in \mathbb{R}^{N \times N}$, ユーザノードからラベルノードへの遷移は $\boldsymbol{Q} = (\boldsymbol{q}_b, \boldsymbol{q}_s) \in \mathbb{R}^{N \times 2}$ で与えられる．ここで，列ベクトル $\boldsymbol{q}_b$ の各成分は $i \in V_b$ のとき $q_{bi} = 1/\widehat{d}_i$, $i \notin V_b$ のとき $q_{bi} = 0$ であり，$\boldsymbol{q}_s$ の各成分は $i \in V_s$ のとき $q_{si} = 1/\widehat{d}_i$, $i \notin V_s$ のとき $q_{si} = 0$ である．ゆえに，式 (7) は行列

$$\boldsymbol{\Pi} := \left(\begin{array}{c|c} \widehat{\boldsymbol{D}}^{-1}\boldsymbol{A} & \boldsymbol{Q} \\ \hline \boldsymbol{O} & \boldsymbol{I}_2 \end{array} \right)$$

を用いて $\boldsymbol{p}^{(t)} = \boldsymbol{\Pi} \boldsymbol{p}^{(t-1)}$ と書ける．ここで，$\boldsymbol{I}_2$ は 2×2 の単位行列である．従って，SybilWalk は

$$\begin{aligned} \boldsymbol{p} = \lim_{t \to \infty} \boldsymbol{\Pi}^t \boldsymbol{p}^{(0)} &= \left(\begin{array}{c|c} \boldsymbol{O} & (\boldsymbol{I} - \widehat{\boldsymbol{D}}^{-1}\boldsymbol{A})^{-1}\boldsymbol{Q} \\ \hline \boldsymbol{O} & \boldsymbol{I}_2 \end{array} \right) \begin{pmatrix} \vdots \\ 0 \\ 1 \end{pmatrix} \\ &= (\boldsymbol{I} - \widehat{\boldsymbol{D}}^{-1}\boldsymbol{A})^{-1} \boldsymbol{q}_s = \boldsymbol{V}_{\mathrm{a}} \boldsymbol{\Lambda}_{\mathrm{a}}^{-1} \boldsymbol{V}_{\mathrm{a}}^{-1} \boldsymbol{q}_s \end{aligned} \tag{8}$$

と定式化できる．ここで，$\boldsymbol{\mathcal{L}}_{\mathrm{aug}} := \boldsymbol{I} - \widehat{\boldsymbol{D}}^{-1}\boldsymbol{A}$ は拡張正規化 Laplacian 行列であり，$\boldsymbol{\Lambda}_{\mathrm{a}} = \mathrm{diag}(\lambda_1^{\mathrm{a}}, \ldots, \lambda_N^{\mathrm{a}})$ と $\boldsymbol{V}_{\mathrm{a}} = (\boldsymbol{v}_1^{\mathrm{a}}, \ldots, \boldsymbol{v}_N^{\mathrm{a}})$ はそれぞれ $\boldsymbol{\mathcal{L}}_{\mathrm{aug}}$ の固有値の対角行列と固有ベクトルを並べた行列である．

表 1 代表的な検知手法のローパスフィルタリング解釈のまとめ

Method	Shift matrix $\boldsymbol{S}$	Filter kernel $h(\lambda)$
CIA	$\boldsymbol{\mathcal{L}}_{\mathrm{rw}}$	$(1-\alpha)/(1-\alpha(1-\lambda))$
SybilRank	$\boldsymbol{\mathcal{L}}_{\mathrm{rw}}$	$(1-\lambda)^{\Gamma}$
SybilWalk	$\boldsymbol{\mathcal{L}}_{\mathrm{aug}}$	$1/\lambda$
SybilSCAR	$\boldsymbol{\mathcal{L}}_{\max}$	$1/\lambda$
SybilBelief	$\boldsymbol{H}(r)$	ideal low-pass filter

3.4 SybilBelief

SybilBelief [10] は，社会ネットワーク構造を pairwise MRF としてモデル化し，標準的な信念伝播法 [18] を用いて各ノードが Sybil である確率を近似的に計算する手法である．信念伝播法の更新式は非線形であるため，ローパスフィルタリングとして表現するためには更新式の線形化が必要である．我々は先行研究 [19] において，更新式を固定点の周りで線形化することにより，Bethe-Hessian 行列

$$\boldsymbol{H}(r) := (r^2-1)\boldsymbol{I} + \boldsymbol{D} - r\boldsymbol{A} \tag{9}$$

を用いて，SybilBelief を次式のようにローパスフィルタリングとして再定式化した：

$$\boldsymbol{p} = \boldsymbol{V}_{\mathrm{H}}\, g(\boldsymbol{\Lambda}_{\mathrm{H}})\boldsymbol{V}_{\mathrm{H}}^{-1}\boldsymbol{q} \tag{10}$$

ここで，$\boldsymbol{\Lambda}_{\mathrm{H}} = \mathrm{diag}(\lambda_1^{\mathrm{H}},\ldots,\lambda_N^{\mathrm{H}})$ 及び $\boldsymbol{V}_{\mathrm{H}} = (\boldsymbol{v}_1^{\mathrm{H}},\ldots,\boldsymbol{v}_N^{\mathrm{H}})$ は $\boldsymbol{H}(r)$ の固有値の対角行列と固有ベクトルを並べた行列である．また，$g(\lambda)$ は理想的なローパスフィルタカーネルであり，ある λ' に対して $\lambda < \lambda'$ のとき $g(\lambda) = 1$ if $\lambda < \lambda'$, $\lambda \geq \lambda'$ のとき $g(\lambda) = 0$ となる．

3.5 SybilSCAR

SybilSCAR [14] は SybilBelief のスケーラビリティ及び収束性の問題を回避するため，SybilBelief の更新式において cavity を無視して近似することにより，各ノード i が Sybil である確率 p_i を計算する手法である．SybilSCAR の更新式は

$$\check{\boldsymbol{p}}^{(t)} = 2\check{\boldsymbol{W}}\check{\boldsymbol{p}}^{(t-1)} + \check{\boldsymbol{q}} \tag{11}$$

と表現される．ここで，$\check{\boldsymbol{W}} = (\check{w}_{ij})$ は残差 重み行列で，$(i,j) \notin E$ のとき $\check{w}_{ij} = 0$ である．式 (11) は $\check{\boldsymbol{p}}^{(t)} = \check{\boldsymbol{p}}^{(t-1)}$ となる固定点において $\check{\boldsymbol{p}} = 2\check{\boldsymbol{W}}\check{\boldsymbol{p}} + \check{\boldsymbol{q}} = (\boldsymbol{I} - 2\check{\boldsymbol{W}})^{-1}\check{\boldsymbol{q}}$ と変形できるので，SybilSCAR は

$$\check{\boldsymbol{p}} = \left(\boldsymbol{I} - \frac{1}{d_{\max}}\boldsymbol{A}\right)^{-1}\check{\boldsymbol{q}} = \boldsymbol{V}_{\mathrm{m}}\boldsymbol{\Lambda}_{\mathrm{m}}^{-1}\boldsymbol{V}_{\mathrm{m}}^{-1}\check{\boldsymbol{q}} \tag{12}$$

と定式化できる．ここで，$\boldsymbol{\Lambda}_{\mathrm{m}} = \mathrm{diag}(\lambda_1^{\mathrm{m}},\ldots,\lambda_N^{\mathrm{m}})$ と $\boldsymbol{V}_{\mathrm{m}} = (\boldsymbol{v}_1^{\mathrm{m}},\ldots,\boldsymbol{v}_N^{\mathrm{m}})$ は最大次数正規化 Laplacian 行列 $\boldsymbol{\mathcal{L}}_{\max} := \boldsymbol{I} - \frac{1}{d_{\max}}\boldsymbol{A}$ の固有値の対角行列及び固有ベクトルを並べた行列である．

4. 既存手法の理論的比較

前節では，いくつかの代表的な Sybil 検知手法をローパスフィルタリングとして定式化する方法について説明した．表 1 に示す通り，これらの検知手法の違いは各々のシフト行列とフィルタカーネルの違いに帰着される．ローパスフィルタリングの出力結果は，ローパスフィルタカーネルの振る舞いとシフト行列の選択（すなわち，どのようなフーリエ基底で周波数変換するか）によって異なる．グラフ信号処理の文脈でよく知られているように，Laplacian 行列の小さい（低周波な）固有値に対応する固有ベクトルはグラフの大域的なコミュニティ構造に関する情報を持ち，反対に大きい（高周波な）固有値に対応する固有ベクトルは．noisy な情報を含む [20]．したがって，Sybil 検知手法の性能はローパスフィルタリングがいかに適切に低周波成分を抽出し，高周波成分を取り除くことができるかに依存すると考えられる．本節では，「フィルタカーネルの特性」と「シフト行列のスペクトル」の 2 つの観点から Sybil 検知手法を比較することで，各手法の差異を議論する．また，議論を踏まえて，新しい Sybil 検知手法 (SybilHeat) を提案する．

4.1 フィルタカーネル特性

図 1 は表 1 の 4 種類のフィルタカーネルをプロットしたものである．まず，CIA のフィルタカーネル $h(\lambda) = (1-\alpha)/(1-\alpha(1-\lambda))$ は高周波成分を十分に除去できていないため，出力信号はノイズの多い高周波成分の影響を受ける可能性がある．このため，CIA は検知性能やノイズ耐性が低いことが予想される．次に，SybilRank のフィルタカーネル $h(\lambda) = (1-\lambda)^T$ は中間の範囲の周波数成分は除去できているが，高周波成分は逆に通してしまっている．ゆえに，$\boldsymbol{\mathcal{L}}_{\mathrm{rw}}$ の最大固有値 $\lambda_N (\leq 2)$ が大きい場合，SybilRank の出力は高周波成分の影響を強く受ける可能性がある．後述するように，$\boldsymbol{\mathcal{L}}_{\mathrm{rw}}$ はグラフが sparse なほど最大固有値が大きくなる傾向があるので，SybilRank はスパースなグラフにおいては性能が低くなることが予想される．$h(\lambda) = 1/\lambda$ は低周波成分を強く強調し，高周波成分の寄与が相対的に非常に小さくなっている．$\lambda \to 0$ で $h(\lambda) \to \infty$ となるため，特に $\lambda_1 \simeq 0$ の場合は最小固有値に属する固有ベクトルの影響が支配的になるが，低周波固有値と高周波固有値が十分に分離していれば高い性能を発揮することが期待される．SybilBelief に対応するフィルタカーネルは（定義から）低周波成分を等しく抽出し，高周波成分を完全に除去するものである．このため，低周波固有ベクトルが豊富な情報を持つ限り，SybilBelief は高い検知性能及びノイズ耐性を発揮することが期待される．

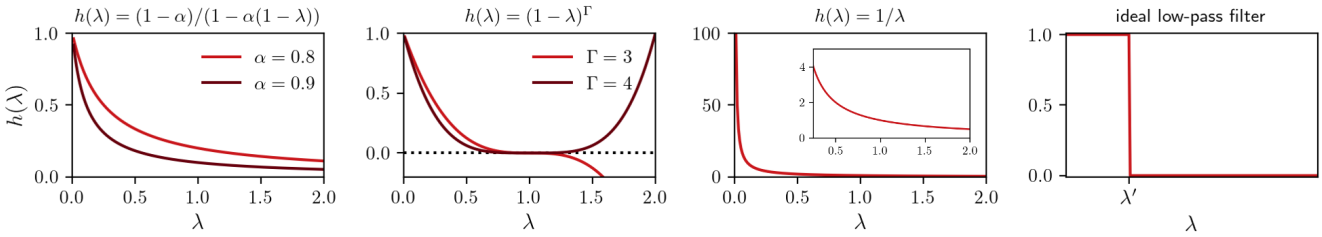

図 1　表 1 中の既存検知手法に対応するフィルタカーネル

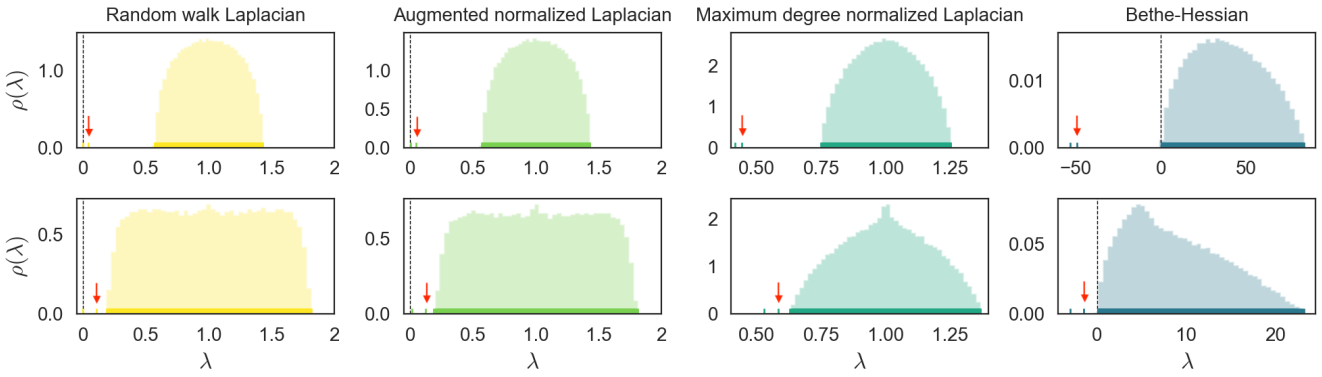

図 2　デンスなモジュラーグラフ（上段）とスパースなモジュラーグラフ（下段）に対する各シフト行列の固有値分布．各図において，矢印は第 2 最小固有値の位置を表す．

4.2　シフト行列のスペクトル

グラフ信号に対するローパスフィルタリングの出力は，周波数（固有値）がどのように分布するか，どのようなフーリエ基底で周波数変換するか（すなわち，どのようなシフト行列を選択するか）によって異なる．本項では，シフト行列のスペクトルの観点から各手法を比較する．

4.2.1　シフト行列の固有値

まず，シフト行列の固有値の観点から検知性能について議論する．一般的な信号処理では周波数スペクトルは等間隔にサンプリングされるが，グラフ信号処理においては，周波数（固有値）の分布は一様ではなく，行列によって異なる．ローパスフィルタリングの良し悪しは，有益な情報を持つ低周波スペクトルをいかに適切に抽出できるかによって決まるため，固有値分布において低周波固有値が高周波固有値のバルクから明確に分離されているほど “良く” ローパスフィルタリングができると予想される．図 2 に Stochastic Block Model (SBM) [21] によって生成されたデンス及びスパースなモジュラーグラフの固有値分布を示す．ただし，Bethe-Hessian $\boldsymbol{H}(r)$ のパラメータ r は文献 [22] と同様 $r = [(\sum_i d_i^2)/(\sum_i d_i) - 1]^{1/2}$ とした．このパラメータ設定では，Bethe-Hessian の informative な固有値は負の値を取り，高周波な固有値は正の値を取るため，両者の峻別が容易になる利点がある．

デンスなグラフでは，各シフト行列の固有値は k 個の小さい（低周波な）固有値がその他の固有値のバルクから明確に孤立することがわかる．一方で，スパースなグラフでは，uninformative な固有値のバルクがべったり広がり，低周波な固有値との峻別が困難となる．このことから，スパースなグラフではローパスフィルタリングによって uninformative なスペクトルの影響を十分に排除することが困難であり，結果としてどの検知手法もデンスなグラフの場合と比較して検知性能が悪化することが予想される．

4.2.2　シフト行列の固有ベクトル

次に，シフト行列の固有ベクトルの観点から検知性能について議論する．Sybil 検知は本質的にグラフの Sybil ノードのコミュニティと正規ノードのコミュニティを特定する問題であるため，各シフト行列の低周波固有ベクトルがグラフのコミュニティ構造に関して豊富な情報を持っているほど良い検知性能を発揮することが期待される．そこで．低周波固有ベクトルの「情報の豊富さ」を測るため，各行列の低周波固有ベクトルによるコミュニティ識別性能を評価する．具体的には，SBM と Degree-Corrected SBM (DCSBM) [23] によって生成された $k = 2$ 個のコミュニティ構造を持つ 2 種類のグラフにおいて，各シフト行列によるスペクトルクラスタリングによってコミュニティを推定し，真のコミュニティと推定コミュニティの正規化相互情報量 (NMI) を計算することでコミュニティ識別性能を評価する．なお，DCSBM は各ノードの次数が任意の次数分布 $p(d)$ に従うように設定可能な SBM の拡張モデルであり，ここでは $p(d) \propto d^{-3}$ とした．

図 3 に SBM 及び DCSBM によって生成されたグラフにおける各シフト行列のコミュニティ識別性能を示す．横軸

はコミュニティ間の結合度であり，c_{in}/N はコミュニティ内のエッジ密度を，c_{out}/N はコミュニティ間のエッジ密度を表し，$(c_{\text{in}}-c_{\text{out}})/2$ が大きくなるほど各コミュニティは明確に分離する．また，破線は識別可能閾値を表し，破線より左側ではどのようなコミュニティ識別アルゴリズムを用いてもコミュニティの識別が理論上不可能であることが示されている [24–26]．図 3 から，SBM グラフにおいては，どのシフト行列も識別可能閾値を境にコミュニティ識別が可能となり，特に $\boldsymbol{H}(r)$ が最も高い識別性能を示すことがわかる．これに対して，DCSBM グラフにおいては，$\boldsymbol{\mathcal{L}}_{\max}$ と $\boldsymbol{H}(r)$ は識別可能閾値を境にコミュニティの識別が可能となる一方，$\boldsymbol{\mathcal{L}}_{\text{rw}}$ と $\boldsymbol{\mathcal{L}}_{\text{aug}}$ は破線の右側の領域においても結合度が小さいときはコミュニティの識別ができないことがわかる．これはコミュニティ数 $k>2$ の場合でも同様である．このことから，スパースで次数不均一性の高いグラフにおいては，SybilBelief や SybilSCAR が高い検知性能を発揮することが予想される．

$\boldsymbol{\mathcal{L}}_{\max}$ や $\boldsymbol{H}(r)$ のコミュニティ識別性能が高い理由は行列の regularization の観点から説明される．先行研究 [27,28] で，通常の正規化 Laplacian 行列を regularized Laplacian 行列 $\boldsymbol{\mathcal{L}}_\tau = \boldsymbol{I}-\boldsymbol{D}_\tau^{-1/2}\boldsymbol{A}\boldsymbol{D}_\tau^{-1/2}$ または $\boldsymbol{\mathcal{L}}_\tau = \boldsymbol{I}-\boldsymbol{D}_\tau^{-1}\boldsymbol{A}$（ただし，$\boldsymbol{D}_\tau = \boldsymbol{D}+\tau\boldsymbol{I}$）に置き換えた Regularized Spectral Clustering (RSC) と呼ばれるアルゴリズムが提案されており，適切に regularization を加える（適切に τ を設定する）ことで RSC のクラスタリング誤差を小さく抑えられることが理論的に示されている [28]．実際，図 3 に示す通り，τ を適切に設定する $(\tau = d_{\text{ave}})$ と，$\boldsymbol{\mathcal{L}}_\tau$ のコミュニティ識別性能は最も性能の高い $\boldsymbol{H}(r)$ に匹敵する．

上記を踏まえると，各シフト行列のコミュニティ識別性能の優劣は次のように説明される．まず，$\boldsymbol{\mathcal{L}}_{\text{rw}}$ は regularization が全く加わっていない状態に相当する．また，対角行列 $\widehat{\boldsymbol{I}}$ を $i\in V_b\cup V_s$ のとき $[\widehat{I}]_{ii}=1$，それ以外のとき $[\widehat{I}]_{ii}=0$ で定義すると，$\boldsymbol{\mathcal{L}}_{\text{aug}} = \boldsymbol{I}-(\boldsymbol{D}+\widehat{\boldsymbol{I}})^{-1}\boldsymbol{A}$ は弱く不均一に regularization が加えられている状態に相当する．一方で，$\boldsymbol{D}_{\text{diff}}$ は $[D_{\text{diff}}]_{ii} := d_{\max}-d_i$ とすると，$\boldsymbol{\mathcal{L}}_{\max} = \boldsymbol{I}-(\boldsymbol{D}+\boldsymbol{D}_{\text{diff}})^{-1}\boldsymbol{A}$ は強く不均一に regularization が加えられている状態に相当する．$\boldsymbol{H}(r)$ は，$\boldsymbol{\mathcal{L}}_\tau = \boldsymbol{I}-\boldsymbol{D}_\tau^{-1}\boldsymbol{A}$ とは次のような関係にある：

$$\boldsymbol{H}(r)\boldsymbol{v}=\lambda\boldsymbol{v} \;\Leftrightarrow\; \boldsymbol{\mathcal{L}}_{r^2-\lambda-1}\boldsymbol{v}=\frac{r-1}{r}\boldsymbol{v}$$

4.3 SybilHeat

前節の解析結果から，Sybil 検知手法が高い性能を発揮するためには，(i) ローパスフィルター $h(\lambda)$ が適切に低周波（高周波）成分を強調（除去）し，(ii) シフト行列 $\boldsymbol{S}$ が高いコミュニティ識別性能を有することが必要である．これを踏まえて，我々はフィルター $h(\lambda)=e^{-s\lambda}$ $(s\geq 0)$，シフト行列 $\boldsymbol{S}=\boldsymbol{\mathcal{L}}_\tau=\boldsymbol{I}-\boldsymbol{D}_\tau^{-1/2}\boldsymbol{A}\boldsymbol{D}_\tau^{-1/2}$ であるような Sybil 検知手法 (SybilHeat) を提案する．$e^{-s\lambda}$ は heat kernel と呼ばれるフィルターカーネルであり，スケーリングパラメータ $s\geq 0$ が大きいほど高周波成分を強く除去する．$\boldsymbol{\mathcal{L}}_\tau$ の固有ベクトルは，上述の通り，Bethe-Hessian と同等の高いコミュニティ識別性能を有する．SybilHeat は所与の事前評価値 $\boldsymbol{q}$ に対して，事後評価値 $\boldsymbol{p}$ を次式で計算する：

$$\boldsymbol{p}=\boldsymbol{V}_\tau e^{-s\boldsymbol{\Lambda}_\tau}\boldsymbol{V}_\tau^{-1}=e^{-s\boldsymbol{\mathcal{L}}_\tau}\boldsymbol{q}=h(\boldsymbol{\mathcal{L}}_\tau)\,\boldsymbol{q} \qquad (13)$$

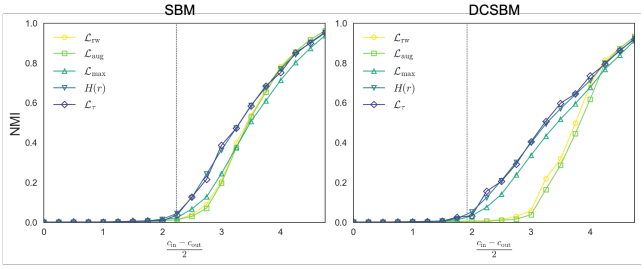

図 3　各シフト行列の低周波固有ベクトルのコミュニティ識別能力

5. 数値実験

前節では，既存の Sybil 検知手法のローパスフィルタリング解釈に基づいて各手法の性能の優劣の理由や各手法が高い性能を発揮する条件について説明し，さらに，その条件を満たすような新たな検知手法として SybilHeat を提案した．本節では，人工グラフでの数値実験を通して，提案手法の検知性能とノイズ耐性を評価する．

本実験では，ノード数 $N=1000$，平均次数 $d_{\text{ave}}=5$ の，SBM 及び DCSBM によって生成された $k=2$ 個の同じサイズのコミュニティを持つ人工グラフを評価に用いる．我々は一方のコミュニティ内のノードを正規ノード，他方を Sybil ノードとした．また，正規ノード及び Sybil ノード全体のうち，ランダムな 10% のノードラベルを既知とした．各手法の実験パラメータは次のように設定した：CIA は $\alpha=0.85$，SybilRank は $\Gamma=\lfloor\log N\rfloor$，SybilSCAR$\theta=0.5$，SybilHeat は $s=8$ とした．

5.1 検知性能

Sybil 検知手法は各ノードに対して Sybil らしさのランキングを計算する [29] ので，我々は検知性能の評価に Area Under the Receiver Operating Characteristic Curve (AUC) を採用する．ある手法の AUC は（ランダムに選ばれた）Sybil ノードが正規ノードよりも高いランクになる確率を意味し，全ての Sybil ノード正規ノードよりも上位にランク付けされている場合，AUC は 1 となる．各ノードを一様にランダムにランク付けした場合，AUC は 0.5 となる．

図 4 に人工グラフでのコミュニティ強度の変化に対する各手法の検知性能を示す．この図から我々は次の結果を観察する．まず，全ての手法はコミュニティ間の結合度

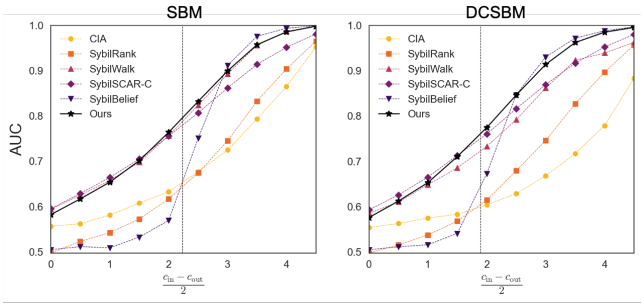

図 4　各検知手法の検知性能

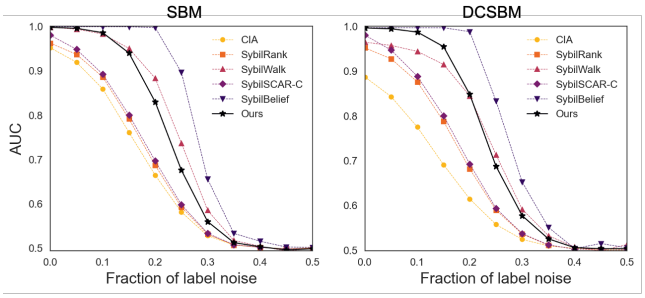

図 5　各検知手法のノイズ耐性

$(c_{\text{in}} - c_{\text{out}})/2$ が小さくなるほど検知性能は悪化する．これは，結合度が小さくなるほど Sybil ノードと正規ノードの間に辺が増え，両者を構造的に峻別するのが困難になるためである．次に，DCSBM グラフにおいて，ランダムウォークベースの検知手法 (CIA, SybilRank, SybilWalk) は SBM グラフの場合と比較して検知性能が悪化する．これは，図 3 の右に示す通り，DCSBM グラフにおいてこれらの手法に対応するシフト行列のコミュニティ識別性能が低いことに起因する．さらに，SybilBelief は，識別可能領域では最も高い性能を示すが，識別不能領域では検知性能が急激に悪化し，他の手法よりも性能が低くなる．言い換えれば，SybilBelief はコミュニティ間の結合が強いグラフでは良好な検知性能を発揮するが，コミュニティ間の結合が弱い場合は十分な性能を発揮できない．これは，SybilBelief が Bethe-Hessian の k 個の小さい固有ベクトル（識別不能領域では情報を持たない）のみに依存して検知を行なっていることに起因すると考えられる．これに対して，提案手法の SybilHeat は他の手法と比較してグラフのコミュニティ構造の変化に影響されず，結合度の変化に対して一貫して高い性能を示す．

5.2　ノイズ耐性

実用上，訓練データには人間のミスによってノイズが含まれる可能性がある [30]．すなわち，いくつかの既知の正規ノードは実際には Sybil ノードであり，いくつかの既知の Sybil ノードは実際には正規ノードである場合がある．本項では，各手法のノイズ耐性を評価するため，既知のノードのうち割合 ε (≤ 0.5) のラベルを逆にしたときの検知性能を評価する．

図 5 に人工グラフにおける各手法のノイズ耐性を示す．まず，4.1 節で議論した通り，SybilBelief に対応するフィルタカーネルは，高周波成分を完全に除去するためノイズに対して非常にロバストである一方，CIA や SybilRank に対応するフィルタカーネルは高周波成分を十分に除去できていないためノイズに対して脆弱であることがわかる．また，SybilWalk と SybilSCAR-C は同じフィルタカーネル $h(\lambda) = 1/\lambda$ であるが，SybilSCAR-C は SybilWalk よりもノイズ耐性が低い．これは，図 2 に示す通り，スパースグラフにおいて SybilSCAR-C のシフト行列 $\boldsymbol{\mathcal{L}}_{\max}$ は全ての固有値が $\lambda = 1$ の周りに凝集する（すなわち，低周波固有値と高周波固有値が接近する）ため，高周波成分の寄与を無視できなくなることに起因すると考えられる．提案手法は SybilBelief に次いでノイズに対してロバストであり，特にラベルノイズが小さい範囲 ($\varepsilon \leq 0.1$) では SybilBelief とほぼ同等である．これは，SybilBelief を除く既存手法に対応するフィルタカーネルと比較して，SybilHeat に対応するフィルタカーネル $h(\lambda) = e^{-s\lambda}$ が高周波成分の寄与を非常に小さく抑えるためである．実用上，訓練データのラベルノイズの割合が 10% 以上になることはあまり多くないと思われるため，提案手法は実データにおいても高い性能を発揮することが期待される．

6.　おわりに

本稿では，既存の代表的なグラフベースの Sybil 検知手法がグラフ信号のローパスフィルタリングの枠組みで統一的に解釈できることを示した．この解釈により，「フィルタカーネルの特性」と「シフト行列のスペクトル」の 2 つの観点から既存の検知手法を理論的に比較・解析を行った．解析を通して，Sybil 検知手法の性能はローパスフィルタリングがいかに適切に低周波成分を抽出し，高周波成分を除去できるかに依存することを明らかにした．言い換えれば，Sybil 検知手法が高い性能を発揮するためには，(i) フィルタカーネル $h(\lambda)$ が適切に低周波（高周波）成分を強調（除去）し，(ii) シフト行列 $\boldsymbol{S}$ の低周波固有ベクトルが高いコミュニティ識別性能を有することが必要である．さらに，上記の 2 つの要件を満たす手法として，フィルタカーネルが heat kernel でシフト行列が regularized Laplacian であるような Sybil 検知手法 (SybilHeat) を提案した．数値実験を通して，我々の提案手法がコミュニティ間の結合度の変化に対して一貫して高い検知性能を示し，ラベルノイズの割合が小さい範囲では SybilBelief に匹敵して高いノイズ耐性を有することを確認した．

参考文献

[1]　Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S. C. and Dredze,

M.: Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate, *American journal of public health*, Vol. 108, No. 10, pp. 1378–1384 (2018).

[2] Allem, J.-P. and Ferrara, E.: Could social bots pose a threat to public health?, *American journal of public health*, Vol. 108, No. 8, p. 1005 (2018).

[3] Bessi, A. and Ferrara, E.: Social bots distort the 2016 US Presidential election online discussion, *First monday*, Vol. 21, No. 11-7 (2016).

[4] Bastos, M. T. and Mercea, D.: The Brexit botnet and user-generated hyperpartisan news, *Social science computer review*, Vol. 37, No. 1, pp. 38–54 (2019).

[5] Alvisi, L., Clement, A., Epasto, A., Lattanzi, S. and Panconesi, A.: Sok: The evolution of sybil defense via social networks, *2013 ieee symposium on security and privacy*, IEEE, pp. 382–396 (2013).

[6] Yang, C., Harkreader, R., Zhang, J., Shin, S. and Gu, G.: Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter, *Proceedings of the 21st international conference on World Wide Web*, pp. 71–80 (2012).

[7] Cao, Q., Sirivianos, M., Yang, X. and Pregueiro, T.: Aiding the detection of fake accounts in large scale social online services, *9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*, pp. 197–210 (2012).

[8] Boshmaf, Y., Logothetis, D., Siganos, G., Lería, J., Lorenzo, J., Ripeanu, M., Beznosov, K. and Halawa, H.: Íntegro: Leveraging victim prediction for robust fake account detection in large scale OSNs, *Computers & Security*, Vol. 61, pp. 142–168 (2016).

[9] Jia, J., Wang, B. and Gong, N. Z.: Random walk based fake account detection in online social networks, *2017 47Th annual IEEE/IFIP international conference on dependable systems and networks (DSN)*, IEEE, pp. 273–284 (2017).

[10] Gong, N. Z., Frank, M. and Mittal, P.: Sybilbelief: A semi-supervised learning approach for structure-based sybil detection, *IEEE Transactions on Information Forensics and Security*, Vol. 9, No. 6, pp. 976–987 (2014).

[11] Fu, H., Xie, X., Rui, Y., Gong, N. Z., Sun, G. and Chen, E.: Robust spammer detection in microblogs: Leveraging user carefulness, *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 8, No. 6, pp. 1–31 (2017).

[12] Gao, P., Wang, B., Gong, N. Z., Kulkarni, S. R., Thomas, K. and Mittal, P.: Sybilfuse: Combining local attributes with global structure to perform robust sybil detection, *2018 IEEE conference on communications and network security (CNS)*, IEEE, pp. 1–9 (2018).

[13] Dorri, A., Abadi, M. and Dadfarnia, M.: Socialbothunter: Botnet detection in twitter-like social networking services using semi-supervised collective classification, *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, IEEE, pp. 496–503 (2018).

[14] Wang, B., Jia, J., Zhang, L. and Gong, N. Z.: Structure-based sybil detection in social networks via local rule-based propagation, *IEEE Transactions on Network Science and Engineering*, Vol. 6, No. 3, pp. 523–537 (2018).

[15] Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A. and Vandergheynst, P.: The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains, *IEEE signal processing magazine*, Vol. 30, No. 3, pp. 83–98 (2013).

[16] Ortega, A., Frossard, P., Kovačević, J., Moura, J. M. and Vandergheynst, P.: Graph signal processing: Overview, challenges, and applications, *Proceedings of the IEEE*, Vol. 106, No. 5, pp. 808–828 (2018).

[17] Wang, B., Jia, J. and Gong, N. Z.: Graph-based security and privacy analytics via collective classification with joint weight learning and propagation, *arXiv preprint arXiv:1812.01661* (2018).

[18] Pearl, J.: *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan kaufmann (1988).

[19] Furutani, S., Shibahara, T., Hato, K., Akiyama, M. and Aida, M.: Sybil Detection as Graph Filtering, *GLOBECOM 2020-2020 IEEE Global Communications Conference*, IEEE, pp. 1–6 (2020).

[20] Ramakrishna, R., Wai, H.-T. and Scaglione, A.: A user guide to low-pass graph signal processing and its applications: Tools and applications, *IEEE Signal Processing Magazine*, Vol. 37, No. 6, pp. 74–85 (2020).

[21] Holland, P. W., Laskey, K. B. and Leinhardt, S.: Stochastic blockmodels: First steps, *Social networks*, Vol. 5, No. 2, pp. 109–137 (1983).

[22] Saade, A., Krzakala, F. and Zdeborová, L.: Spectral clustering of graphs with the bethe hessian, *Advances in Neural Information Processing Systems*, Vol. 27 (2014).

[23] Karrer, B. and Newman, M. E.: Stochastic blockmodels and community structure in networks, *Physical review E*, Vol. 83, No. 1, p. 016107 (2011).

[24] Mossel, E., Neeman, J. and Sly, A.: Stochastic block models and reconstruction, *arXiv preprint arXiv:1202.1499* (2012).

[25] Massoulié, L.: Community detection thresholds and the weak Ramanujan property, *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pp. 694–703 (2014).

[26] Gulikers, L., Lelarge, M. and Massoulié, L.: An impossibility result for reconstruction in the degree-corrected stochastic block model, *The Annals of Applied Probability*, Vol. 28, No. 5, pp. 3002–3027 (2018).

[27] Chaudhuri, K., Chung, F. and Tsiatas, A.: Spectral clustering of graphs with general degrees in the extended planted partition model, *Conference on Learning Theory*, JMLR Workshop and Conference Proceedings, pp. 35–1 (2012).

[28] Qin, T. and Rohe, K.: Regularized spectral clustering under the degree-corrected stochastic blockmodel, *Advances in neural information processing systems*, Vol. 26 (2013).

[29] Viswanath, B., Post, A., Gummadi, K. P. and Mislove, A.: An analysis of social network-based sybil defenses, *ACM SIGCOMM Computer Communication Review*, Vol. 40, No. 4, pp. 363–374 (2010).

[30] Wang, G. A., Mohanlal, M., Wilson, C., Wang, X., Metzger, M., Zheng, H. and Zhao, B. Y.: Social Turing Tests: Crowdsourcing Sybil Detection, *NDSS Symposium 2013*, Internet Society (2013).