Computer Security Symposium 2013
21 - 23 October 2013

SNS の特性を活かした不正アカウント検知手法

津田 侑 †　　遠峰 隆史 †　　井上 大介 †

† 独立行政法人 情報通信研究機構
184-8795　東京都小金井市貫井北町 4-2-1
{tsuda, tomine, dai}@nict.go.jp

あらまし　情報共有やコミュニティ形成を目的とした SNS には実名や性別，所属先などのプライバシーに関する情報の登録を前提とするものもある．そのため，SNS は個人情報やプライバシー情報の集合であると言え，これらが攻撃者によって窃取される危険性がある．情報収集の手段の一つに，攻撃用の不正なアカウントを作成して他者との友達関係の構築を試みる方法がある．不正なアカウントの中には実在の人物を模倣する「なりすましアカウント」と呼ばれるものも存在する．本研究では不正アカウントを用いた攻撃者からの情報収集を未然に防ぐために，まず，不正アカウントの現状を調査する．そして，これらの特徴および SNS やユーザ活動の特性を活かした不正アカウントの検知手法を提案する．

A Method for Detecting Fake Profiles Using the Characteristics of Online Social Networks

Yu Tsuda†　　Takashi Tomine†　　Daisuke Inoue†

†National Institute of Information and Communications Technology.
4-2-1, Nukui-Kitamachi, Koganei, Tokyo, 184-8795, JAPAN
{tsuda, tomine, dai}@nict.go.jp

Abstract　Online social networks (OSNs) aimed at intelligence sharing and community formation promote registering privacy information, for example, real name, gender, affiliation. Therefore, OSNs have huge privacy information, and attackers can collect them. A collection method is creating accounts and trying to become friendship with users. These accounts call "fake profile" since the accounts imitate actual people or accounts. In this research, for protecting privacy information against the attacks, the authors investigate existing state of fake profiles. Then, the authors propose a method for detecting these accounts using the characteristics of the accounts and users' activities on OSNs.

1　はじめに

インターネット上での人間関係の構築を目的としたソーシャル・ネットワーキング・サービス (SNS: Social Networking Service) が普及してきた．SNS では，他のユーザと友達関係を構築したり，特定の事柄に興味を持つユーザ同士でコミュニティを形成することができる．このような特性とユーザ数が増加傾向にあることから，SNS はユーザ間でのコミュニケーションや情報共有が促進されるプラットフォームとなりつつある．

SNS の形体は多種多様である．代表的なも

のとして，短文の投稿でユーザと交流をするTwitter[1]や，ユーザの詳細なプロフィール情報を入力することで友達関係を広げるFacebook[2]がある．その他にも，音楽やエンタテイメントの話題を中心に交流することを目的としたMySpace[3]やビジネスにおける人間関係の構築を目的としたLinkedIn[4]などが挙げられる．

SNSの利用を開始するにあたりユーザはアカウントを作成することになる．このとき，実名や勤務先を原則として登録させるSNSも存在する．このようなSNSでは，実在の人物とユーザアカウントを強く結びつけることができ，日常生活の延長としてインターネット上でも交流を続けられるという利点がある．一方で，SNSにはユーザによって登録された個人のプライバシーに関する情報が集約されるため，これらを狙う攻撃者によって情報を窃取される恐れがあり，多くの既存研究においてもプライバシー情報の漏洩が懸念されている[5, 6, 7]．

SNSからユーザの情報を窃取する攻撃者の手段の一つに，攻撃用の不正なアカウントを作成して他のユーザとの友達関係の構築を試みる方法がある．これは，友達関係を結ぶことにより友達のみに限定公開している情報まで閲覧することができるためであり，攻撃者はユーザが友達以外には非公開にしたいと感じているプライバシーに強く関わる情報を窃取できる．また，不正アカウントの中には既に友達関係を結んでいる友達や著名人といった実在する人物を模倣して作成された「なりすましアカウント」と呼ばれるものも存在する．なりすましアカウントはユーザに誤って友達申請を承認させる手段として，攻撃者にとっては有用であると言える．

本研究では攻撃者からの情報収集を未然に防ぎ，ユーザが安心・安全にSNSを利用できるようにするために，まず，不正アカウントの現状を調査し，そのアカウントの属性の特徴的なものを精査する．そして，これらの特徴およびSNSやユーザ活動の特性を活かした不正アカウントの検知手法についての提案を述べる．

2　不正アカウントによる情報収集

会員登録制のSNSは基本的にはSNS内でのみ他のユーザの情報を閲覧でき，外部に公開されているものは，ユーザによって公開設定されたもののみである．SNS内で閲覧できる情報についても，その公開設定次第では友達やその友達までといった公開範囲を設定できる．すなわち，より詳細に情報を収集するには，標的となるユーザと友達関係を結ぶことが挙げられる．

攻撃者がソーシャルエンジニアリングを用いてSNS上のユーザから情報収集する手法は多岐にわたる．Cutilloらはこれらの手法をプライバシー・完全性・可用性に対する危険性の観点から整理している[8]．プライバシーはユーザ個人の情報の公開制限を自ら設定できることを指し，完全性は第三者によってユーザ個人の情報が改竄されないことを指す．可用性はSNS上での活動が阻害されないことを指す．この中で，プライバシーと完全性を共に脅かすものとして，ユーザになりすますものが挙げられている．ID窃取 (ID theft) は盗み出したIDを用いて不正アクセスすることで本人になりすますことである．プロフィール複製 (Profile cloning) は標的のユーザのプロフィールをそのまま利用した不正なアカウントを作成して活動すること，プロフィール移植 (Profile porting) はプロフィール複製を複数のSNSに渡って行うことを指す．

攻撃者が積極的に情報を収集するソーシャルエンジニアリングに対して，Iraniらは標的が攻撃者に自ら情報を流出するように仕向けるリバースソーシャルエンジニアリングの手法を指摘している[9]．先に挙げた不正アカウントが標的となるユーザから友達申請を受けるためには，不正アカウントの性別や年齢といった属性やプロフィール画像に依存することが挙げられている．さらに，不正アカウントと友達関係を結んだことによって共通の友達が増えることで，不正アカウントへの友達申請の件数が増加することも述べられている．

また，SNS上で不正アカウントを実際に作成してユーザに友達申請を試みる研究もなされている．BilgeらはFacebookを対象として自動的に不正アカウントを作成し友達申請をするシス

テムを構築している [10]．Boshmaf らは SNS 上で活動するボットネットを作成し，それらを活動させることでユーザのプライバシー情報を抽出可能であると述べている [11]．

このように，攻撃者が SNS 上のユーザと友達関係を結ぶことによって個人情報やプライバシー情報を窃取することは十分に起こり得る．攻撃者による情報窃取を防ぐためにはユーザが攻撃者からの友達申請を承認しないように注意喚起する仕組みが有効であると考えられる．また，友達申請を受けた時点で不正アカウントであるかどうかをユーザに提示する必要があるため，不正アカウントが公開設定している名前や性別といった基本的なプロフィール情報のみで判定する必要がある．

一方で，これら不正アカウントへの対処は，Facebook 公認のナビゲーションページである Facebook navi でも述べられているように，最終的にはユーザの判断に任せられている部分もある [12]．本研究ではよりユーザが安心・安全に SNS を利用できるようにするために，まず，次章では Facebook を対象として不正アカウントの現状を調査し，不正アカウントを分析する．第 4 章ではそれを受けて SNS の特性を活かした不正アカウントの検知手法を提案する．

3　不正アカウントの現状の調査

3.1　データ収集

不正アカウントの現状を調査するにあたり，Facebook を対象として不正アカウントの可能性があるものを収集した．収集期間は 2013 年 5 月 23 日から 2013 年 6 月 10 日で，収集方法は Twitter のサイト内で「Facebook」と「なりすまし」を検索ワードとして検索し，掲載されていた URL よりアカウント情報のプロフィールを抽出した．不正アカウントのプロフィール情報は Facebook Graph API[13] を用いて取得した．この Facebook Graph API では，対象となるユーザが公開設定している情報のみ取得できる．表 1 に収集したデータの概要を示す．

次節では，この 51 件の収集データを公開されたプロフィール情報を基に不正アカウントを分類する．

表 1: 収集した不正アカウント

収集期間	2013 年 5 月 23 日 〜 2013 年 6 月 10 日
件数	51 件 男性: 19 件 女性: 32 件
友達の数	最大値: 327 最小値: 0 平均値: 24.0 中央値: 8

3.2　不正アカウントの分析

収集した不正アカウントを公開されている登録名やプロフィール画像といったプロフィール情報に着目して分析する．

登録名と読み方　プロフィール情報のうち，登録名やその読み方は必ず他のユーザに対して公開される．そこで，登録名やその読み方の特徴を調べる．

収集した 51 件の不正アカウントのうち，7 件の登録名とその読み方が妥当ではないものであった．具体的には，漢字表記が「今野 妙子」にも関わらず，カタカナ表記が「ワタナベ トシコ」，ローマ字表記が「Taniguchi Akemi」といったようなものであった．さらに，いずれも女性に付けられることが多い登録名であるにも関わらず，性別は男性として設定されていた．

このようなプロフィール情報は Facebook のような利用規約で実名登録を定めている [14] ために不当なものであると判断し，不正アカウントと判定する一つの要素となると考えられる．

プロフィール画像　登録名と同様にプロフィール画像も必ず他のユーザに対して公開される．妥当な登録名が設定されている不正アカウントのうち，男性用初期画像のものが 12 件，女性用初期画像のものが 1 件であった．また，初期画像とは別の画像が設定されている不正アカウ

表 2: プロフィール画像の分類

	登録名	
	妥当	不当
男性用初期画像	12	1
女性用初期画像	1	-
女性の写真	29	-
女性の集合写真	2	-
キャラクター	-	6
合計	44	7

ントのうち，女性の写真が設定されているものが 29 件，女性の集合写真が設定されているものが 2 件であった．これらのアカウントは全て性別も妥当なものに設定されている．さらに，登録名の読み方が不自然な 7 件の不正アカウントのプロフィール画像は，男性用初期画像が 1 件，アニメやキャラクターの画像が 6 件であった．これらは全て性別が男性と設定されていた．以上のプロフィール画像による分類を表 2 にまとめる．

この他に，初期画像以外のプロフィール画像が設定されているものと類似する画像を Google 画像検索を用いて検索すると，同一の画像を用いた Facebook のアカウントが散見された．このことからも，攻撃者が不正アカウントを作成する際にインターネット上の画像を用いてプロフィール情報を作成している可能性がある．

上記の通り，登録名やプロフィール画像といった公開されたプロフィール情報にも特徴が見られた．以下にその特徴についてまとめる．

- 登録名とその読み方（カタカナ表記・ローマ字表記）の組み合わせが妥当ではない．
- プロフィール画像に特徴があり，どれもインターネット上にある画像を流用していると思われる．
  - 初期画像をそのまま利用．
  - 女性単体・集合写真を利用．
  - アニメやキャラクターの画像を利用．

4 SNS の特性を活かした不正アカウント検知

4.1 ユーザ協力型の不正アカウント検知

前章で述べたように，現状の不正アカウントの公開されたプロフィール情報からも特徴が見られる．これらのプロフィール情のの特徴を用いることで，ユーザが受けた友達申請の中から不正アカウントを検知する．この不正アカウント検知に加えて本研究では，不正アカウントをユーザ間で共有することで正規のユーザが不正アカウントからの友達申請を許可する危険性を低減させることを狙う．

ここでは，Facebook を対象とした不正アカウントの判定と不正アカウントの共有の仕組みを図 1 に示す．まず，ユーザに届いた友達申請の一覧とユーザの友達の一覧を不正なりすまし判定器に送信する．不正なりすまし判定器は不正アカウントの特徴や過去に他のユーザによって登録された不正アカウント情報を基に不正アカウント・なりすまし度を算出し，ユーザに提示する．提示された不正アカウント・なりすまし度を参考にユーザは友達申請の承認と保留を選択でき，さらに任意で不正アカウント DB に登録できる．不正アカウント DB に登録された情報は他のユーザと共有される．

次節より，図 1 中の不正なりすまし判定器における不正アカウントの判定基準についての詳細を述べる．

4.2 不正アカウントの判定

Facebook 上のユーザ個人やその友達関係の情報は，Facebook Graph API を利用することで取得できる．本研究では不正アカウント検知のために，ユーザ個人に関する情報と友達申請に関する情報を用いる．それぞれの情報の一例を表 3 と表 4 に示す．表 4 中の「検索用文字列」にはユーザの登録名（姓・名）やそのカタカナ表記，ローマ字表記が含まれる．

前章で述べた不正アカウントの特徴に加えてこれらの Facebook から取得できる情報から不正アカウントの検知に次の五つの要素を用いる．

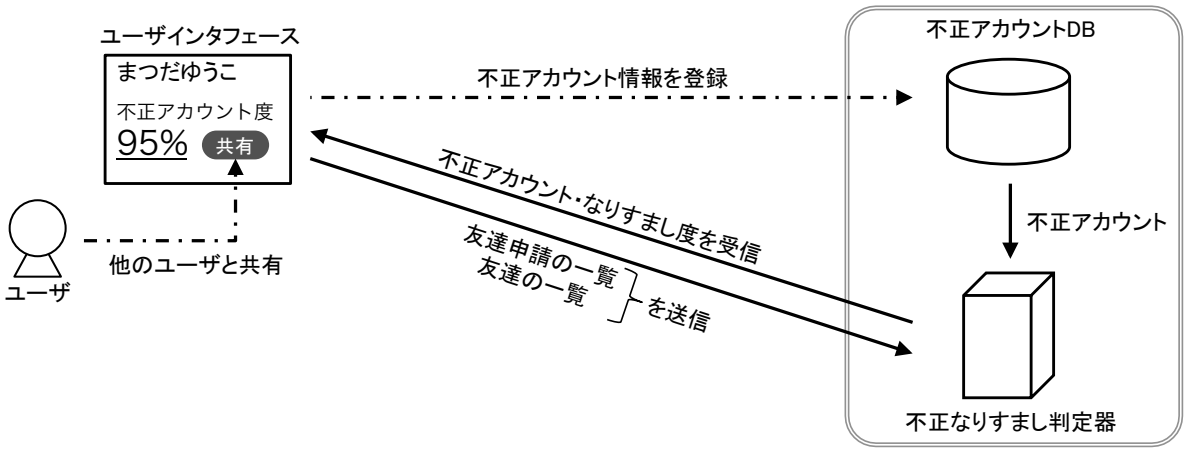

図 1: 不正アカウントの検知の流れ

表 3: 友達申請に関する情報の一例

フィールド	内容
uid_to	友達申請先のユーザ ID
uid_from	友達申請元のユーザ ID
time	友達申請時刻
message	友達申請時のメッセージ

表 4: ユーザに関する情報の一例

フィールド	内容
name	登録名（姓・名）
search_tokens	検索用文字列
pic	プロフィール画像
friend_count	友達の数
mutual_friend_count	共通の友達の数
profile_update_time	最終更新時刻

- 登録名と読み方
- プロフィール画像
- 共通の友達の数
- 友達申請時刻と最終更新時刻
- 不正アカウント情報の共有

以下にその五つの要素の詳細について述べる．

登録名と読み方　Facebook では利用規約によって実名の登録が定められているが，このような登録名が実名であることを前提とした SNS では姓・名の表記およびその読み方・ローマ字表記がユーザによって正しく入力されていると考える．すなわち，その表記や読み方に不整合がある場合，その登録名を持つアカウントは不正である可能性がある．登録されている姓・名とその読み方の組み合わせの整合性は NAIST Japanese Dictionary[15] に登録されてるものと比較して判定する．

プロフィール画像　登録名とは違い，必ずしもプロフィール画像のみでユーザ本人と判別できるものでなくても良い．そのため，正規のユーザは多種多様なプロフィール画像を設定している．一方で，不正アカウントはプロフィール画像を設定せずに初期画像を利用していたり，インターネット上でダウンロードしてきた他者の画像をプロフィールに転用していたりする．そこで，プロフィール画像にこのような設定がされていれば，不正アカウントの可能性があると判定する．また，不正アカウントとして他のユーザに共有された場合は，そのアカウントのプロフィール画像を不正アカウント DB にブラックリストとして登録することで，次回以降の不正アカウント検知に利用する．

共通の友達の数　日常生活において何らかのコミュニティに所属していたり，ある程度の人間関係を築いている人物から友達申請を受ける場合，既に別のユーザがその人物と SNS 上で友達関係を結んでいることが見込まれる．そのため，友達申請を送信してきたユーザと自身の間で共通の友達の数が 0 人の場合，友達申請を送信してきたユーザが不正アカウントである可能性がある．

友達申請時刻と最終更新時刻　攻撃者が不特定多数に友達申請をする場合，不正アカウントを作成後すぐに不特定多数のユーザに友達申請をすることは攻撃にかかる人的・時間的コストを考慮すると効率が良いと考えられる．そのため，不正アカウントの作成時刻（最終更新時刻）と友達申請時刻の間隔は短くなると考えられる．そこで，この間隔が短いものは不正アカウントであると判定する一つの指標とする．

不正アカウント情報の共有　攻撃者が特定の個人を標的にする場合を除き，攻撃者の不正アカウントから友達申請を受けるユーザは複数人存在すると考えられる．このような不正アカウントの情報を他のユーザとも共有することは不正アカウントの検知においても有効であると考えられる．

ここで，筆者が目視で不正アカウントと判定した実例を表 5，表 6 に示す．表中の共通の友達の数は友達申請を受けたユーザと不正アカウントに共通する友達の数のことを指す．申請時刻は不正アカウントからユーザに対して友達申請をした時刻，最終更新時刻は不正アカウントのプロフィールが最後に更新または作成された時刻のことを指す．

不正アカウントの例 (1) では，登録名の漢字表記とカタカナ表記の組み合わせが妥当ではなく，プロフィール画像は初期画像をそのまま使用していた．不正アカウントの例 (2) では，登録名は漢字表記のみで，プロフィール画像はインターネット上で検索できるものが使用されていた．また，それぞれ共通の友達はおらず，友達申請時刻と最終更新時刻の差を見ると比較的短い間隔で友達申請を行っていることがわかる．これらのアカウントは，いづれも本稿執筆時には削除済みになっており，活動期間の短さも不正アカウントの特徴であると言える．

表 5: 不正アカウントの例 (1)

属性	内容
登録名（漢字）	久保田 紗菜
登録名（カタカナ）	タナカ カナコ
登録名（ローマ字）	Kubota Sana
性別	男性
プロフィール画像	男性用初期画像
共通の友達の数	0
申請時刻 – 最終更新時刻	157 秒

表 6: 不正アカウントの例 (2)

属性	内容
登録名（漢字）	徳永 菊恵
登録名（カタカナ）	-
登録名（ローマ字）	-
性別	女性
プロフィール画像	女性の写真
共通の友達の数	0
申請時刻 – 最終更新時刻	56 秒

4.3　なりすましアカウントの判定

不正アカウントの中には，他のユーザのプロフィールを模倣した「なりすましアカウント」が存在する．なりすましアカウントの多くは，既存のユーザの姓・名をそのまま，あるいは一部改変して設定し，そのユーザの友達一覧に登録されているユーザに対して友達申請を行う．このような友達申請を受けたユーザは，友達のプロフィールと似ているために誤って友達申請を承認してしまう恐れがある．

なりすましアカウントによる友達申請の対策として，ユーザ本人がなりすましアカウントを発見し次第，自分の友達に対して注意喚起をすることが有効であると考えられる．一方で，このようななりすましアカウントを SNS 上の全ての

表 7: 文字列の編集距離

文字列 (1)	文字列 (2)	編集距離	正規化
津田侑	津田侑	0	0
津田侑	津田侑子	1	0.25
津田侑	井上大介	4	1.0

ユーザの中から探索することは困難である．そのため，ここでは過去に不正アカウントとして他のユーザが共有したものを対象にしてなりすましアカウントを探索する．

なりすましアカウントはその登録名が自分の登録名と類似している特徴がある．そこで，ユーザの登録名と過去に不正アカウントとして共有されたアカウントの登録名との編集距離を求める．編集距離とは，文字列 S1 から文字列 S2 に変更するまでにかかるコストで，文字の挿入・置換・削除の操作をした回数で決まる．編集距離が小さいほど，二つの登録名は類似しており，その不正アカウントがユーザの「なりすまし」である可能性であるとしてユーザに提示する．

表 7 に二つの文字列の編集距離とそれが正規化されたものの一例を示す．正規化された編集距離 d_{norm} は，編集距離 d，二つの文字列の長さをそれぞれ $str1$，$str2$ としたとき式 (1) で求める．

$$d_{norm} = d \times 1/\max\{str1, str2\} \qquad (1)$$

図 2 にユーザになりすましアカウントを提示する様子を示す．過去に不正アカウントとして報告されたもののうち，編集距離が小さいものをなりすましアカウントの可能性があるとユーザに提示する．ユーザは提示された不正アカウントが自身のなりすましであると判断すると，任意でその情報を他のユーザに共有することができる．

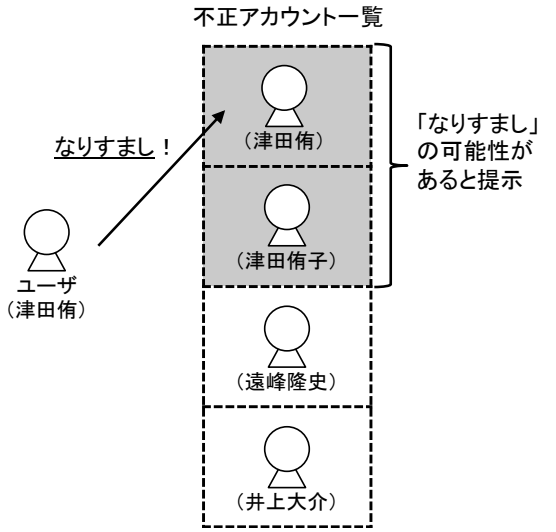

図 2: なりすましアカウントの提示

5 おわりに

本研究では実名登録を前提とした SNS の特性を活かした，不正・なりすましアカウントの検知手法を提案した．提案手法で検知した不正・なりすましアカウントの情報を他のユーザと共有することで，ユーザ同士が互いに不正・なりすましアカウントからの友達申請について警告を発する関係を築くことができる．このような関係を築くことは，攻撃者によるユーザ個人に関するプライバシー情報の窃取を未然に防ぐ効果があると考えられる．

一方で，このような SNS における不正なアカウントの作成手法は日々変化し続けるものであり，また，SNS によって作成手法が異なる可能性もある．そのため，不正アカウントの検知モデルを柔軟に構成できるように設計する必要がある．これにはパターン認識や機械学習のアプローチを用いることで不正アカウントの分類やクラスタリングが考えられる．

今後は，本手法を用いて Facebook と連携した不正・なりすましアカウントの検知システムを実装する．そして，Facebook の既存ユーザにそのシステムを利用してもらうことによって不正・なりすましアカウントの検知性能を評価する．

参考文献

[1] Twitter. http://www.twitter.com/（2013 年 8 月 26 日閲覧）．

[2] Facebook. http://www.facebook.com/（2013 年 8 月 26 日閲覧）.

[3] MySpace. http://myspace.com/（2013 年 8 月 26 日閲覧）.

[4] LinkedIn. http://www.linkedin.com/（2013 年 8 月 26 日閲覧）.

[5] Ralph Gross, Alessandro Acquisti, and H John Heinz Iii. Information Revelation and Privacy in Online Social Networks. In *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pp. 71–80, 2005.

[6] Joshua Fogel and Elham Nehmad. Internet social network communities : Risk taking , trust , and privacy concerns. *Computers in Human Behavior*, Vol. 25, No. 1, pp. 153–160, 2009.

[7] Balachander Krishnamurthy and Craig E Wills. Characterizing Privacy in Online Social Networks. In *Proceedings of the first workshop on Online social networks*, pp. 37–42, 2008.

[8] Leucio Antonio Cutillo, Refik Molva, and Thorsten Strufe. Safebook : A Privacy-Preserving Online Social Network Leveraging on Real-Life Trust. *IEEE Communications Magazine*, No. December, pp. 94–101, 2009.

[9] Danesh Irani, Marco Balduzzi, Davide Balzarotti, Engin Kirda, and Calton Pu. Reverse Social Engineering Attacks in Online Social Networks. In *Proceedings of the 8th International Conference; DIMVA 2011*, pp. 55–74, 2011.

[10] Leyla Bilge, Thorsten Strufe, Davide Balzarotti, Engin Kirda, and Sophia Antipolis. All Your Contacts Are Belong to Us : Automated Identity Theft Attacks on Social Networks. In *Proceedings of the 18th international conference on World wide web*, pp. 551–560, 2009.

[11] Yazan Boshmaf, Ildar Muslukhov, Konstantin Beznosov, and Matei Ripeanu. The Socialbot Network : When Bots Socialize for Fame and Money. In *Proceedings of the 27th Annual Computer Security Applications Conference*, pp. 93–102, 2011.

[12] Facebook navi: なりすまし・アカウントの乗っ取りにご注意！ http://www.facebook.com/fnavigation/posts/682506155111732（2013 年 8 月 26 日閲覧）.

[13] Facebook Graph API. http://developers.facebook.com/docs/reference/api/（2013 年 8 月 26 日閲覧）.

[14] Facebook 利用規約. http://www.facebook.com/legal/terms/（2013 年 8 月 26 日閲覧）.

[15] NAIST Japanese Dictionary. http://naist-jdic.sourceforge.jp/（2013 年 8 月 26 日閲覧）.