

SNS におけるネットワーク成長に基づく  
ユーザプロファイリング手法に関する研究

筑波大学  
図書館情報メディア研究科

2017年3月  
柘植大

# 目次

第 1 章	序論	1
1.1	背景	1
1.2	本研究の目的	2
1.3	本論文の構成	2
第 2 章	関連研究	3
2.1	SNS ユーザの影響力に関する研究	3
2.2	SNS ユーザのプロファイリングに関する研究	3
2.3	SNS ネットワーク特性に関する研究	4
2.4	SNS 上の予測に関する研究	4
2.5	本研究の位置づけ	5
第 3 章	ユーザ分類法と聞き役ユーザの選定法	6
3.1	Reply ネットワークの成長に基づくユーザ分類法	6
3.2	聞き役ユーザ	8
第 4 章	検証と実験	11
4.1	ユーザ分類法の検証	11
4.1.1	検証に用いたデータセット	11
4.1.2	新規ユーザの分類結果	12
4.1.3	分類結果の検証 1 - ユーザの平均 Tweet 数 -	12
4.1.4	分類結果の検証 2 - reply ユーザとの同期性 -	14
4.1.5	分類結果の検証 3 - reply ユーザのネットワーク成長 -	19
4.1.6	分類結果の考察	19
4.2	ネットワーク成長が他ユーザに与える影響力の検証	20
4.2.1	検証に用いたデータセット	20
4.2.2	検証内容	20

---

4.2.3	検証結果 . . . . .	20
4.3	聞き役ユーザの検証 . . . . .	24
4.3.1	評価実験 1 -各特徴量に優れたユーザの影響力- . . . . .	24
4.3.2	評価実験 2 -ネットワーク成長に貢献したユーザの各特徴量- . . . . .	26
第 5 章	考察 . . . . .	28
5.1	ネットワーク成長が他ユーザに与える影響力に関する考察 . . . . .	28
5.2	ネットワーク成長と聞き役ユーザ . . . . .	29
5.2.1	評価実験 1 に関する考察 . . . . .	29
5.2.2	評価実験 2 に関する考察 . . . . .	29
第 6 章	結論 . . . . .	30
	謝辞 . . . . .	31
	参考文献 . . . . .	32

# 目次

3.1	新規ユーザ 100 名の Reply ユーザの増加曲線 . . . . .	6
4.1	新規ユーザの一年間における毎月の各日平均 tweet 数 . . . . .	13
4.2	新規ユーザの Reply ユーザとの同期性 ( $synccrate \geq 0.8$ の場合) . . . . .	17
4.3	新規ユーザの Reply ユーザとの同期性 ( $synccrate \geq 0.3$ の割合) . . . . .	18
4.4	Reply-Reply ユーザの増加曲線 . . . . .	19
4.5	前後 3 ヶ月間における毎月の各日平均 tweet 数 . . . . .	22
4.6	前後 3 ヶ月間におけるネットワーク成長パラメータ $a$ の変動 . . . . .	23
4.7	聞き役ユーザ評価実験 1 の詳細図 . . . . .	24
4.8	被影響ユーザ 400 名の前後 3 ヶ月間における成長パラメータ $a$ の変動 . . . . .	25
4.9	聞き役ユーザ評価実験 2 に用いるスコア . . . . .	26

# 表目次

3.1	関与率で考慮する素性 . . . . .	9
4.1	パラメータ $a$ による新規ユーザの分類結果 . . . . .	12
4.2	各グループにおけるユーザの同期性の割合 . . . . .	16
4.3	被影響ユーザの前後 3 ヶ月間における投稿活動量の比較 . . . . .	21
4.4	提案法による各グループ 10 名の分類結果 . . . . .	25
4.5	影響ユーザ 100 名のスコア上位 10 名・下位 10 名の聞き役ユーザとしての特 徴量 . . . . .	27
5.1	各グループのスコア平均 . . . . .	29

# 第 1 章

## 序論

### 1.1 背景

近年，Twitter<sup>\*1</sup> に代表される SNS(Social Networking Service) が注目を集めている。2015 年 12 月 31 日時点での Twitter 月間アクティブユーザ (MAU) 数は約 3 億人と報告されており，ユーザ数は今後も増加傾向にあるといえる。しかし一方，近年では実際には活動していない非アクティブユーザが増加傾向にあると報告されている。非アクティブユーザとは，アカウントは存在するが，活動を行っていないユーザを意味し，Twitter ユーザの 8 割近くを占めるとのことである。また，Twitter,Inc. による MAU の定義は「一ヶ月中に一度でもログインしたユーザ」であり，ユーザの投稿の有無は考慮されていない。ある調査<sup>\*2</sup> においては，2014 年末時点の Twitter Inc. の MAU 中の 8.5%，約 2500 万人のユーザが投稿活動を行っていないことがわかっている。さらに近年では新規ユーザの離脱率の上昇<sup>\*3</sup>も報告されており，2012 年に利用を開始した新規ユーザの内，2014 年 2 月時点で活動しているユーザは 10.7% とのことである。

以上のような投稿活動の停滞や離脱といった現象について，著者は SNS 上における他ユーザ間とのコミュニケーションが強く影響していると考えている。ここでいうコミュニケーションとは Twitter 上での Reply, Retweet といったアクションを指す。通常の Tweet のみの投稿活動と，Reply のような他者とのコミュニケーション要素を含んだ投稿活動では，ユーザの長期的な投稿活動に差異が生まれると考えられる。具体的には，ユーザ間のコミュニケーションが連鎖的に新たなユーザ間コミュニケーションを発生させ，ユーザのコミュニティの拡大や投稿活動の活性化を誘発していると仮定する。このコミュニケーションの増加やコミュニティ

---

\*1 <https://twitter.com/>

\*2 <http://uk.businessinsider.com/twitter-user-growth-is-worse-than-you-think-2015-2>

\*3 アカウントを作成した直後あるいは短期間の利用の後に，投稿をしなくなる現象を本論文では離脱と称する。離脱したユーザでも，閲覧は継続していることも考えられるので，実際には完全に利用を停止したユーザとは言えない。

の拡大といった現象を，本研究では SNS におけるネットワーク成長と呼ぶ．

## 1.2 本研究の目的

本研究では，SNS におけるネットワーク成長に基づくユーザの分類方法を提案する．その上でネットワーク成長とユーザの投稿活動の関係性について検証を行い，最終的にはユーザの投稿活動の活性化に有益なユーザの検出・推薦を目的としている．以上を検証するために，本研究ではまずユーザの投稿活動における Reply ネットワークの成長に応じてユーザを3つのグループに分類する．そして各グループに属するユーザの投稿活動を長期的に追うことで，Reply ネットワークの有無が投稿活動に与える影響を考察した．

## 1.3 本論文の構成

本論文の構成を以下に示す．まず第2章では，関連研究について概観し，本研究の位置づけを示す．第3章では，Reply ネットワークの成長傾向によるユーザの分類手法と，ネットワーク成長を促す聞き役ユーザの定義について説明する．第4章で実験に用いたデータセットについて述べるとともに，ユーザの分類結果およびグループごとの検証結果，また聞き役ユーザとの関係性について評価し，第5章で考察する．第6章で本論文のまとめと今後の課題について述べる．

## 第 2 章

# 関連研究

### 2.1 SNS ユーザの影響力に関する研究

Twitter 上の特定ユーザの影響力に関する研究は既に数多く行われている。Hannon ら [7] はユーザプロフィールをもとにした content based の推薦手法を提案している。この手法ではユーザの Tweet, Follower, Following, Follower の Tweet, Following の Tweet といったプロフィールをもとにした特徴量を利用して、ユーザの特性を調査し、その結果をユーザ推薦に用いている。Cha ら [6] は、Twitter ユーザの影響力を 3 つの尺度（次数、Retweet された数、Mention<sup>\*1</sup>された数）で評価し、それぞれの評価尺度において抽出されるユーザの特徴についてまとめている。Wang[14] らは Twitter の Mention 機能に着目したユーザの影響力の計測手法を提案している。Wang らは Mention 機能をユーザ推薦手法に取り入れ、この手法において推薦ユーザの Tweet 拡散能力を Social tie model や User influence model といった独自のモデルによって特徴量としている。Tweet の拡散率や拡散する範囲といったものを細かく定義し、それにより「Tweet の拡散相手」として最も最適であるユーザの推薦を実現している。

### 2.2 SNS ユーザのプロファイリングに関する研究

Shaomei ら [16] は「Who Says What to Whom」という観点で、「誰が誰の Tweet を見ているのか」、「誰がどんな Tweet を見ているのか」を調査している。この研究において「誰」の部分は Celebrities, Media, Organizations, Blogs の 4 グループとし、このグループでユーザを分類した際にそれぞれのグループが注視しているグループとその内容についてまとめてい

---

<sup>\*1</sup> Mention とは Reply と同等機能で、Tweet 中の何処かで「@username」と記述すると、指定したユーザのタイムラインに Tweet を送れる仕組みである。Reply が 1:1 の関係性であるのに対し、Mention は 1:N の関係で Tweet を拡散することができる



る。Isabel らの研究 [1] では、実際にオーストリアの Twitter ユーザについて、さまざまな尺度からランク付けを行い、その結果それぞれの尺度において重要となるユーザは異なることを明らかとした。Zhonghua ら [11] は同一の特性を持つ Twitter ユーザのグループ化手法を提案している。ここでは Tweet の内容ベースのグループ化手法と、相互フォロー数や mention 回数に着目したネットワークベースの手法の 2 手法を提案し、それぞれの手法を用いてユーザのグループ化を行なっている。

## 2.3 SNS ネットワーク特性に関する研究

Java らの研究 [8] では、Twitter のネットワーク特性を分析し、人々の Twitter の利用目的について考察している。この研究によって、人々は日常の情報発信・情報収集やコミュニケーションのために Twitter を利用し、またそのコミュニティには複数のタイプがあることが明らかになっている。Yamaguchi らの研究 [19][17] ではユーザの投稿パターンや投稿数の変化といったものに注目し、Twitter の利用開始時期によって投稿パターンが異なることが述べられている。また、投稿活動遷移に着目し、利用継続時間の長短で抽出した 2 つのユーザグループに対して分析を行っている。Kwak ら [9] は Twitter ユーザのランキング方法について、ページランクや Retweet による方法を提案・比較している。ここでは、Retweet Trees という手法を提案し、Twitter ネットワークにおいて Retweet がどれだけ拡散するのかについても論じている。Weng ら [15] は PageRank[10] をトピックモデル用いて改良し、より正確なユーザネットワーク分析を可能としている。ここでは PageRank の遷移確率行列を LDA[5] で抽出されたトピック毎に生成することでトピック志向型の PageRank を提案している。Backstrom ら [2] は Twitter ネットワークを重み付けランダムウォークを取り入れた手法で解析する方法を提案し、これにより SNS ネットワークにおける潜在的なエッジの検出を行なっている。

## 2.4 SNS 上の予測に関する研究

Eytan ら [3][4] は Twitter 上を流れる情報のカスケード<sup>\*2</sup>を追跡することによって、Twitter ユーザの特性および相対的な影響力について考察している。この研究の結論として、巨大なカスケードは以前にも同様な巨大なカスケードを発生させたユーザによって生成される傾向が高いが、一方でそのようなユーザから未来のカスケードを予測することは難しいことがわかっている。Sakaki ら [12] は Twitter のリアルタイム性に注目し、Twitter を監視することによる地震といった特定のイベントの検出アルゴリズムを提案している。この研究ではユーザをセンサーと位置づけ、Follower 間での投稿の関連性をもとにしたイベント検出を試みている。

---

<sup>\*2</sup> 情報の伝搬を意味する。具体的にはユーザ A ユーザ B ユーザ C のように RT や Mention 等によって情報が拡散する流れをカスケードと呼ぶ。

Zhao ら [18] は Twitter 内に発生する質問ツイートを検出する分類機を SVM[13] をもとに構築している。また分類した質問ツイートから Google などの検索エンジンにおける検索クエリの予測が可能であることを示している。

## 2.5 本研究の位置づけ

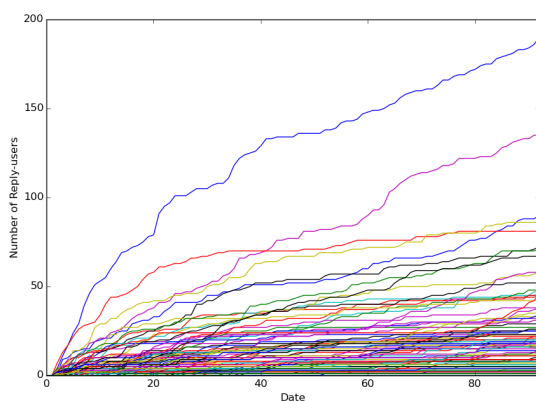
以上のように，SNS や Twitter に関する研究はすでに数多くなされている。本研究も既存研究と被る部分が多々ある。特に投稿活動量や Reply ネットワークの成長に着目している点においては，Yamaguchi らの研究 [19][17] と類似している。一方で本研究では，自身の投稿活動のみならず，それが他ユーザのネットワーク成長に与える影響についてを志向している点に特徴がある。これはすなわち，興味関心の一致といった従来のユーザ推薦で考慮されてきた項目に加えて，そのユーザとつながることによる効果までを考慮することで，よりネットワーク成長に効率的なユーザ推薦を目的としている。

## 第 3 章

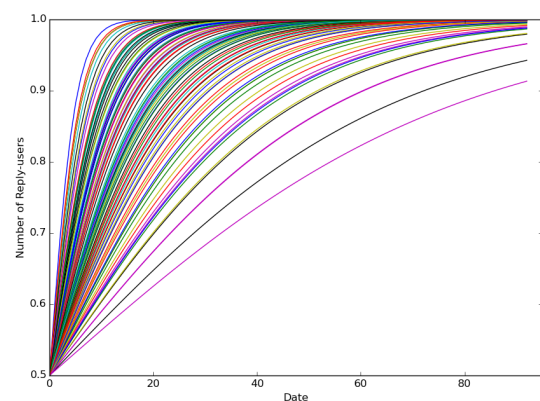
# ユーザ分類法と聞き役ユーザの選定法

本章ではまずユーザの Reply ネットワークの成長に基づく，ユーザ分類法を説明する．次に，ユーザの投稿活動の活性化を目的とした，聞き役ユーザの定義とその選定法について記述する．

### 3.1 Reply ネットワークの成長に基づくユーザ分類法



(a) Reply ユーザ数の累計数



(b) 累計数で正規化した後，シグモイド関数で近似

図 3.1: 新規ユーザ 100 名の Reply ユーザの増加曲線

図 3.1a は Twitter を新たに開始したユーザ計 100 名<sup>\*1</sup>の初期投稿活動（3 ヶ月間）におい

<sup>\*1</sup> 無作為にしたユーザである

て、Reply を行なったユーザの累計数の増加を表す（縦軸が Reply ユーザ数，横軸が日数）。これはユーザ自身が Reply したユーザと，ユーザに Reply を飛ばしたユーザをあわせたものである。最も多いユーザで 193 人の Reply ユーザを獲得しており，一度も Reply が確認できなかったユーザは 6 名となった。また 50% 以上のユーザにおいて Reply ユーザ数は 20 人に満たないことが確認できる。次に Reply ユーザ数の増加傾向に着目すると，多くのユーザの場合において Reply ユーザの増加は活動開始から 2 週間程度で収束している（それ以降は増加していない）ことがわかる。一方では，累計数の多い少ないに限らず，継続的に Reply ユーザを増やし続けているユーザも一部ではあるが確認できる。

本研究では，Reply ユーザ数ではなくその増加傾向に着目する。正確には，Reply ユーザ数の増加の収束速度によってユーザの活動を分類を行う。これは以前に獲得した Reply ネットワークの大きさよりも，将来的にどれだけ多くのユーザとの Reply が期待できるかを判別したいからである。そこでユーザネットワークの成長を，Reply ユーザの増加率と同義とし，Reply ユーザの増加をシグモイド関数で近似する。シグモイド関数は次式で表される。

$$f(x; a) = \frac{1}{1 + e^{-ax}} \quad (3.1)$$

パラメータ  $a$  が大きいほど関数のカーブが急になり，逆に  $a$  が小さいほど変化は緩やかになる。これを Reply ユーザ数の増加と対応付けて考えると，

- $a$  が大: Reply ユーザの増加が初期の投稿活動で収束している
- $a$  が小: Reply ユーザの増加が現在も継続的に行われている

となる。今回は各ユーザ毎に Reply ユーザの総数で正規化した後に，シグモイド関数に近似してパラメータ  $a$  を推定する。近似は  $f(x; a) - y$  の 2 乗誤差の総和が最小となるよう，パラメータ  $a$  を最適化するものとする（ $y$  は実際の正規化した Reply ユーザ数）。これにより，Reply ユーザの増加量は大きいですがすでにネットワークの成長が収束したユーザや，増加は微量であるが継続的に増加しているユーザといった分類が可能となる。

図 3.1a と同様のユーザ 100 名におけるシグモイド関数での近似は図 3.1b となった（色は図 3.1a と対応）。図 3.1b では曲線が線形に近づくほど，Reply ユーザの増加の収束時間が遅いことを意味する（すなわち継続的に Reply ユーザが増え続けているユーザ）。

本研究では，シグモイド関数への近似で推定したパラメータ  $a$  を用いて，ユーザを以下の 3 グループに分類する。

- $a < 0.1$ : 継続的に Reply ユーザを獲得しているユーザ
- $0.1 \leq a < 0.2$ : Reply ユーザの獲得が収束ぎみのユーザ
- $0.2 \leq a$ : Reply ユーザの獲得が収束したユーザ

## 3.2 聞き役ユーザ

聞き役ユーザとは，ユーザの投稿活動を傾聴し，それに対して何らかの反応を示すユーザを指す．新規ユーザに最適な聞き役ユーザを選定し，推薦することにより，本研究の目的であるユーザの投稿活動の活性化につながると考えられる．聞き役ユーザに期待する要素は以下の3つである．

1. 新規ユーザへの関心
2. 新規ユーザへの直接的な働きかけ
3. 新規ユーザの広報

第一に新たなユーザに対して興味関心を持っていることが重要になる．Twitter においては，コミュニケーションを可能とするためには相互 Follow を形成する必要がある．一方的に Follow するだけでは不十分であり，自身に興味を持ち積極的に Follow し返してくれるユーザを期待する．そこで現時点での相互 Follow 率が高いユーザは今後も相互 Follow を形成する確率が高いと仮定し，そのようなユーザを聞き役ユーザの候補と考える．第二に聞き役ユーザから対象ユーザへの直接的なコミュニケーションを期待する．当然であるが他ユーザからの反応は，投稿活動の活性化につながると考えられる．Twitter 上のアクションと対応させると，新規ユーザと聞き役ユーザの間で行われるコミュニケーションは，Reply と Favorites（聞き役ユーザ 新規ユーザの方向）ということになる．3 つめとして聞き役ユーザによる Retweet を介して，聞き役ユーザの Follower 以下に対しても新規ユーザのツイートの伝播を期待する．言い換えると，ネットワークが貧弱な新規ユーザの代わりに，新規ユーザの存在を他者に伝えてもらうということである．これは新規ユーザと聞き役ユーザ間のコミュニケーションとは違うが，他者とのコミュニケーション機会を増加させるという点で重要となる．以上の3つの行動が聞き役ユーザに期待する行動であり，これらの行動を行うユーザが実際に他ユーザに与える影響について調査することが本研究の課題となる．そこでそれぞれの行動毎に特徴量を定義し，聞き役ユーザの選定法を提案する．

相互 Follow 率  $R_{mf}$  は式 (3.2) で与えられる．ユーザをノード，Follow 関係をエッジとした有効ネットワーク  $G = (V, E)$  を考える．このとき  $(u, v)$  は， $u$  が  $v$  を Follow していることを意味し， $|\Gamma_{following}(u)|$  と  $|\Gamma_{mutual}(u)|$  は，それぞれ  $u$  の Following 数， $u$  の相互 Follow 数となる．

$$R_{mf}(u) = \frac{|\Gamma_{mutual}(u)|}{|\Gamma_{following}(u)|} \quad (3.2)$$

$$\Gamma_{following}(u) = \{v \in V | (u, v) \in E\} \quad (3.3)$$

$$\Gamma_{mutual}(u) = \{v \in V | (u, v) \in E, (v, u) \in E\} \quad (3.4)$$

次に新規ユーザに対する直接的働きかけ，すなわち他ユーザへの関与率  $R_g$  について記述する．各ユーザに対する興味関心は一樣であると仮定し，ユーザの単位時間あたり素性数によってこれを求める．関与率で考慮する素性を表 3.1 のとおりである．

表 3.1: 関与率で考慮する素性

新規 Follower 数
新規 Following 数
Favorites 数
Retweet 数
Reply 数

関与率  $R_g$  の定義式を式 (3.5) に示す．新規 Following 数, 新規 Follower 数のエッジの増加数と，Retweet, Reply, Favorites といったコミュニケーションの回数を週単位で収集し，直近 1 ヶ月の増減に応じて値を算出する． $inc$  が 1 週間での各素性の回数であり，例えば  $inc_1(u, Follower)$  は 1 週間前のユーザ  $u$  の新規 Follower 数となる．

$$R_g(u, x) = \frac{\sum_{i=1}^4 R_{gw,i}(u, x)}{4} \quad (3.5)$$

$$R_{gw,i}(u, x) = \begin{cases} 1 & (inc_i(u, x) > \frac{\sum_{j=1}^4 inc_{i+j}(u, x)}{4}) \\ 0 & (inc_i(u, x) \leq \frac{\sum_{j=1}^4 inc_{i+j}(u, x)}{4}) \end{cases} \quad (3.6)$$

$$inc_i(u, x) = \begin{cases} inc_i(u, x) & (inc_i(u, x) > 0) \\ \frac{\sum_{j=1}^3 inc_{i+j}(u, x)}{3} & (inc_i(u, x) \leq 0) \end{cases} \quad (3.7)$$

$x$  は  $x \in \{Follower, Following, Favorites, Retweet, Reply\}$  であり，それぞれの平均値  $R'_g$  (式 (3.8)) を関与率として用いる．

$$R'_g(u) = \frac{\sum_{x \in X} R_g(u, x)}{|X|} \quad (3.8)$$

$$X = \{Follower, Following, Favorites, Retweet, Reply\} \quad (3.9)$$

最後に Tweet の拡散力について詳説する．これについては頻度と影響力の2つを考慮する．まず頻度についてはユーザが発信する Tweet 中の Retweet の割合で求める．すなわち式(3.10)となる．影響力は Retweet がどれだけの範囲に拡散するかであり，頻度に対する重み付けとして考慮する．影響力に適した要素として Follower 数がある．それに応じた重み *Impact* を決定する．*Impact* は式(3.11)となる．

$$Frequency(u) = \frac{|Retweet(u)|}{|Tweet(u)|} \quad (3.10)$$

$$Impact(u) = \log|\Gamma_{follower}(u)| \quad (3.11)$$

頻度 *Frequency* と影響力 *Impact* の積を拡散力に関する特徴量  $R_{diff}$  とする(式(3.12))．

$$R_{diff}(u) = Frequency(u) * Impact(u) \quad (3.12)$$

## 第 4 章

# 検証と実験

本章では初めに 3.1 節で述べたユーザ分類法の検証を行う。分類結果の特徴について考察するとともに、分類の有効性についてまとめる。次にこの分類法で分類した、ネットワーク成長が活発なユーザが他ユーザに与える影響力についての評価を行い、これをまとめる。

### 4.1 ユーザ分類法の検証

#### 4.1.1 検証に用いたデータセット

今回の実験に用いたデータセットは Twitter Search API<sup>\*1</sup> を用いて収集された Tweet である。収集期間は 2012 年 4 月 1 日から 2013 年 6 月 4 日の全 430 日間であり、この期間に日本語で投稿された Tweet が収集されている。今回はこの Tweet データ中において、以下の条件を満たすユーザを用いている。

- 最初の投稿が 2012/05 に見られる
- 2012/05/01 - 2013/06/30 の区間において 80% 以上の日で投稿が確認できる

以上の条件を満たすユーザを検証対象のユーザとして計 100 名用いる。これは対象として新規ユーザ、すなわち Twitter の利用を開始したばかりのユーザを想定しているからである。注意点として、最初の投稿はあくまでも今回用いたデータセット内で確認できる一番最初の投稿である。つまり 2012 年 4 月 1 日以前から Twitter の利用を開始しており、かつ 2012 年 4 月中は投稿がなかったユーザもわずかながら含まれていると考えられる。なお、図 3.1 のグラフに用いたのも同様のユーザである。次に新規ユーザ 100 名の 2012/05/01 - 2012/07/31 における Reply を全て抽出し、その中で区間中に 3 回以上の Reply が確認できるユーザを有効な Reply ユーザとした (計 2187 人 (重複含む))。

---

<sup>\*1</sup> <https://dev.twitter.com/docs/api/1/get/search>



### 4.1.2 新規ユーザの分類結果

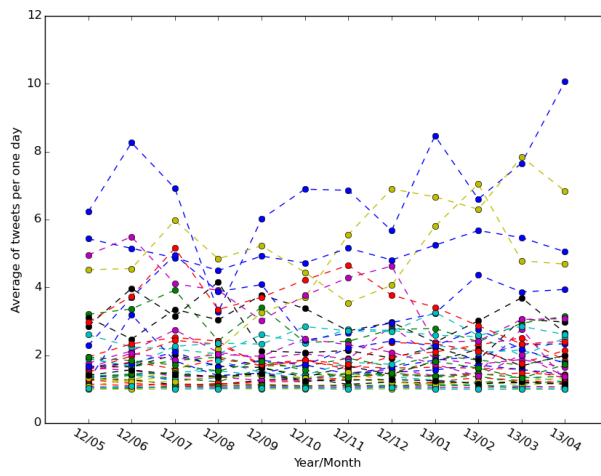
3.1 節で述べた分類法を用いて，起点ユーザを分類した結果を表 4.1 に示す．なお 2012/05/01 - 2012/07/31 の区間において，一度も Reply が確認できなかったユーザ 6 名は除外している．

表 4.1: パラメータ  $a$  による新規ユーザの分類結果

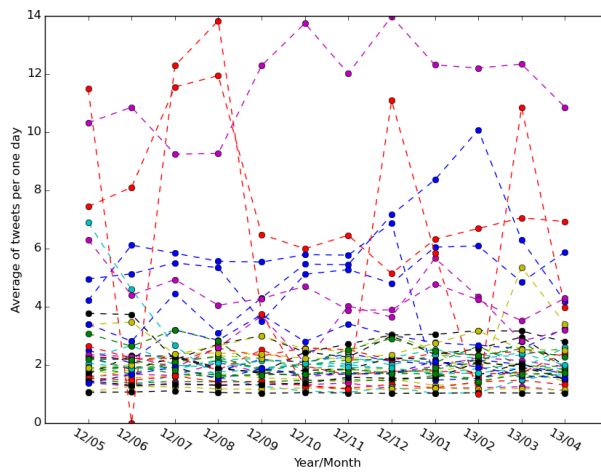
	ユーザ数	最小値	中央値	最大値
$a < 0.1$	39	0.0256	0.0713	0.0995
$0.1 \leq a < 0.2$	37	0.101	0.137	0.195
$0.2 \leq a$	18	0.200	0.256	0.401

### 4.1.3 分類結果の検証 1 - ユーザの平均 Tweet 数 -

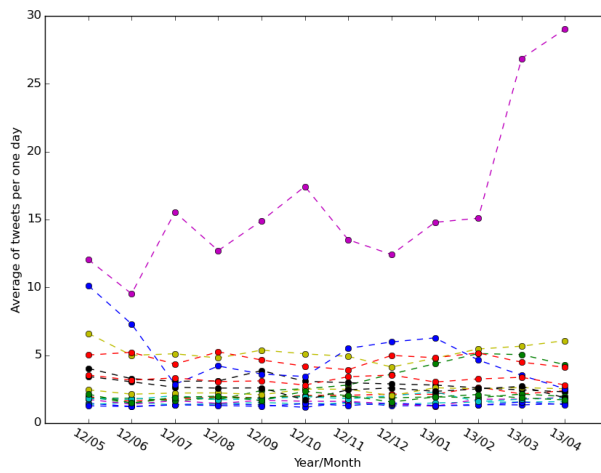
Reply ネットワークの成長がユーザの投稿活動に与える影響を検証するために，月毎の各日の平均 Tweet 数を比較検証する．図 4.1 に各グループごとにまとめた図を示す．図 4.1a に示したユーザ群の投稿活動は図 4.1b と比較すると月毎の分散が小さいことがわかった．つまり継続的に Reply ユーザを獲得しているユーザの投稿活動は，規則的で急激な変化が生じにくいと考えられる．一方で図 4.1b のユーザ群では，平均 Tweet 数の多いユーザは何人が確認できるが，そのいずれのユーザにおいても月毎の投稿数にばらつきがあり，既知の投稿活動からユーザの将来の投稿活動を予測するのは難しいと言える．また 12 年 5 月と 13 年 4 月を比較すると，図 4.1a のユーザ群の方が 13 年 4 月において Tweet 数の向上が見られるユーザが多くいた．以上のことから，Reply ネットワークの成長はユーザの投稿活動の活性化と正の相関があり，また投稿の継続性の向上に寄与することがわかる．Reply ユーザが増え続けているユーザの投稿活動が急激に停滞することは考えにくく，逆に Reply ネットワークの成長が収束したユーザは停滞・離脱の可能性が相対的に高いことになる．



(a)  $a < 0.1$  のユーザ 39 名



(b)  $0.1 \leq a < 0.2$  のユーザ 37 名



(c)  $0.2 \leq a$  のユーザ 18 名

図 4.1: 新規ユーザの一年間における毎月の各日平均 tweet 数

#### 4.1.4 分類結果の検証 2 - reply ユーザとの同期性 -

次に reply ユーザとの投稿活動の同期性について述べる．投稿活動の同期とは，特定ユーザの投稿の異常区間\*2 において，他ユーザも同様に異常が検知されることを意味する．すなわち同期したユーザとは，共通の事柄に対して関心を持っていることになる．本研究では，まず起点となる新規ユーザの投稿活動から異常区間を抽出し，次に reply ネットワークで接続するユーザがその異常区間の周囲において同様に異常区間を迎えているかを評価する．

投稿活動の異常区間検出は以下の手順で行う．ここでは対象ユーザ a の全  $\alpha$  日の投稿活動について正規分布を仮定する（次式  $x_n$  は  $n$  日目における，reply・retweet を含む全 tweet 数を意味する）．

$$x_1, x_2, \dots, x_\alpha \sim n(\mu, \sigma) \quad (4.1)$$

この区間から特定の点  $s$  を定め， $s$  の前後 1 週間をそれぞれ区間  $x, y$  とする．このとき，この 2 区間における投稿活動もそれぞれ正規分布に従うと考えられ，以下のように表される．

$$x_{s-7}, x_{s-6}, \dots, x_{s-1} \sim n_x(\mu_x, \sigma_x) \quad (4.2)$$

$$x_s, x_{s+1}, \dots, x_{s+6} \sim n_y(\mu_y, \sigma_y) \quad (4.3)$$

次に帰無仮説・対立仮説をそれぞれ以下とする．

$$h_0 : \mu_x = \mu_y (= \mu)$$

$$h_1 : \mu_x \neq \mu_y$$

このとき帰無仮説  $h_0$  における尤度関数は，

$$l^0(\mathbf{x}; \mu, \sigma^2, s) = \prod_{n=s-7}^{s+6} \left( \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_n - \mu)^2}{2\sigma^2}\right) \right) \quad (4.4)$$

$\mu$  の推定量， $\hat{\mu}$  は以下．

$$\hat{\mu} = \frac{1}{14} \sum_{n=s-7}^{s+6} x_n \quad (4.5)$$

同様に  $\hat{\sigma}^2$  を求めると

$$\hat{\sigma}^2 = \frac{1}{14} \sum_{n=s-7}^{s+6} (x_n - \hat{\mu})^2 \quad (4.6)$$

\*2 他の区間よりも投稿活動が著しく活発もしくは停滞している区間

対して対立仮説  $h_1$  では  $s$  を前後に  $\mu_x$  と  $\mu_y$  にわかれるため、尤度関数は以下になる。

$$l^1(\mathbf{x}; \mu_x, \mu_y, \sigma_0^2, s) = \prod_{n=s-7}^{s-1} \left( \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(x_n - \mu_x)^2}{2\sigma_0^2}\right) \right) \cdot \prod_{n=s}^{s+6} \left( \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(x_n - \mu_y)^2}{2\sigma_0^2}\right) \right) \quad (4.7)$$

$\mu_x$  と  $\mu_y$  の推定量,  $\hat{\mu}_x, \hat{\mu}_y$  は次式となる ( $\hat{\sigma}_0^2$  は省略)。

$$\hat{\mu}_x = \frac{1}{7} \sum_{n=s-7}^{s-1} x_n, \quad \hat{\mu}_y = \frac{1}{7} \sum_{n=s}^{s+6} x_n \quad (4.8)$$

$l^1, l^2$  の尤度比検定統計量の 2 倍 ( $t$ ) は自由 1 のカイ二乗分布に従う。これより,  $t$  が棄却限界 3.84 を超える点  $s$  を中心する 1 週間を異常区間として抽出する。

$$t = 2 \log\left(\frac{l^1(\mathbf{x}; \hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_0^2, s)}{l^0(\mathbf{x}; \hat{\mu}, \hat{\sigma}^2, s)}\right) \quad (4.9)$$

以上の方法で抽出した異常区間について, reply ユーザがどれだけ対応できているかを同期率で表す。なお, 起点となるユーザの異常区間において, reply 先ユーザも同様に異常区間を迎えている場合を reply ユーザが同期したとみなす。よってユーザ  $u$  に対するユーザ  $v$  の同期率は次式で求まる。

$$\text{syncrate}_{u,v} = \frac{v's \text{ synced count}}{u's \text{ all } s \text{ count}} \quad (4.10)$$

本研究では Reply ユーザのうちどれだけのユーザが同期しているかで, 起点となるユーザの同期性を算出する。すなわち以下の式である。

$$R_{\text{sync}}(u) = \frac{\Gamma_{\text{synced\_reply\_user}}(u)}{\Gamma_{\text{reply\_user}}(u)} \quad (4.11)$$

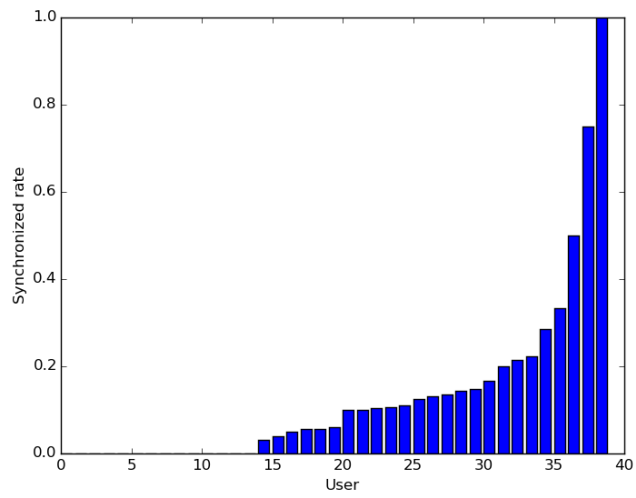
なお以下では, 同期率  $\geq 0.8$  の際の同期性を強い同期性, 同期率  $\geq 0.3$  の際の同期性を弱い同期性と称する。

図 4.2, 4.3 に各グループ全ユーザの同期性を示す。図 4.2 は Reply ユーザとの強い同期性を示す。この図においては Reply ユーザとの同期性が高いユーザが数人存在するが, ほとんどのユーザの同期性は 0.1~0.3 に分布しており, グループごとの差異はなかった。対して同期率  $\geq 30\%$ , すなわち 3 回に 1 回の頻度まで同期率を下げた場合である, 図 4.3 においてはグループごとに違いが確認できる。表 4.2 に各グループにおけるユーザの同期性の割合を示す。 $a < 0.1$  グループの, 弱い同期性における 0.4~0.6 の割合が他 2 グループと比べて高いことが

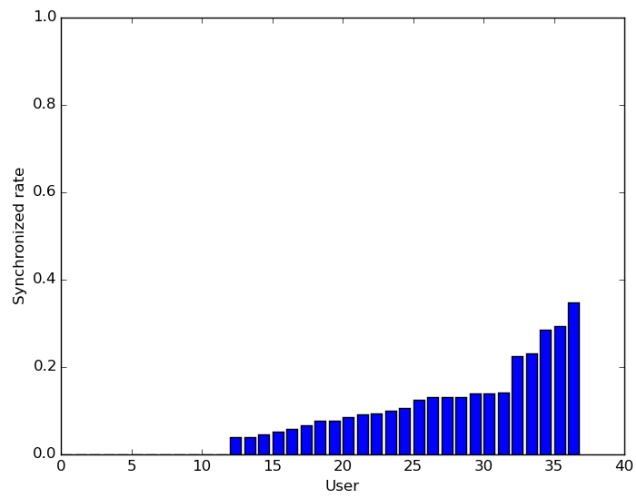
わかる．すなわち  $a < 0.1$  のユーザグループには強い同期性をもつユーザは少ないが，弱い同期性を持つユーザは一定数いるということになる．対してそれ以外グループには同期性を持つユーザはほとんど存在しない．

表 4.2: 各グループにおけるユーザの同期性の割合

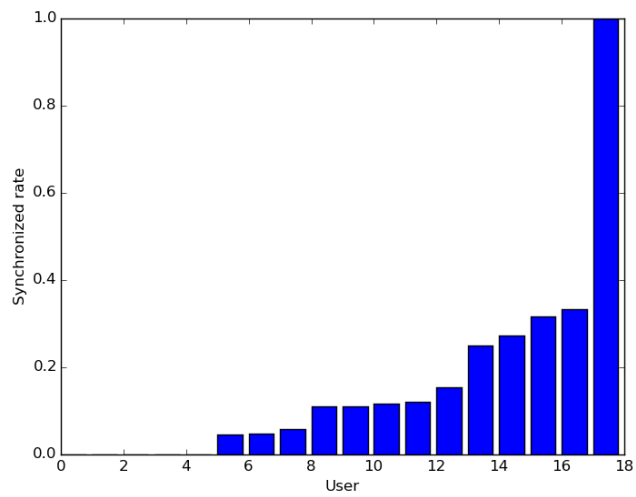
同期性 $R_{sync}$	$a < 0.1$		$0.1 \leq a < 0.2$		$0.2 \leq a$	
	強同期	弱同期	強同期	弱同期	強同期	弱同期
0 ~ 0.2	0.794	0.513	0.864	0.702	0.722	0.722
0.2 ~ 0.4	0.128	0.102	0.135	0.216	0.222	0.166
0.4 ~ 0.6	0.026	0.282	0	0.054	0	0.055
0.6 ~ 0.8	0.026	0.051	0	0.027	0	0
0.8 ~ 1	0.026	0.051	0	0	0.056	0.056



(a)  $a < 0.1$  のユーザ 39 名

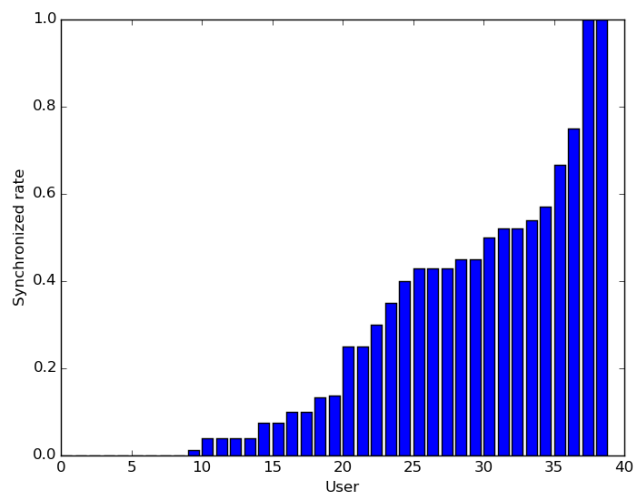


(b)  $0.1 \leq a < 0.2$  のユーザ 37 名

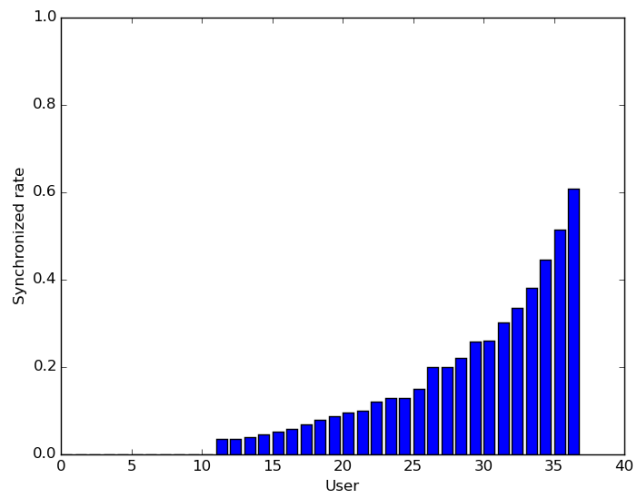


(c)  $0.2 \leq a$  のユーザ 18 名

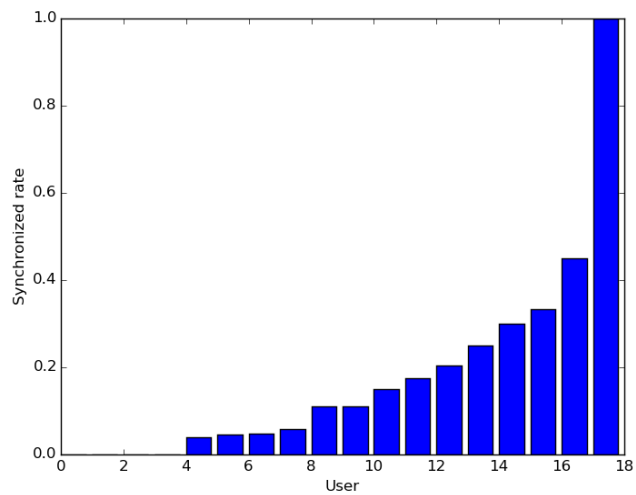
図 4.2: 新規ユーザの Reply ユーザとの同期性 ( $syncrate \geq 0.8$  の場合)



(a)  $a < 0.1$  のユーザ 39 名



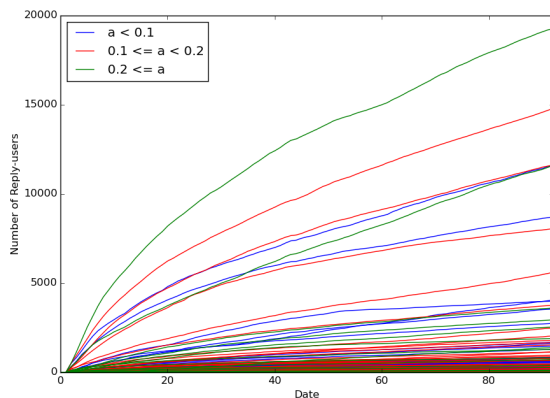
(b)  $0.1 \leq a < 0.2$  のユーザ 37 名



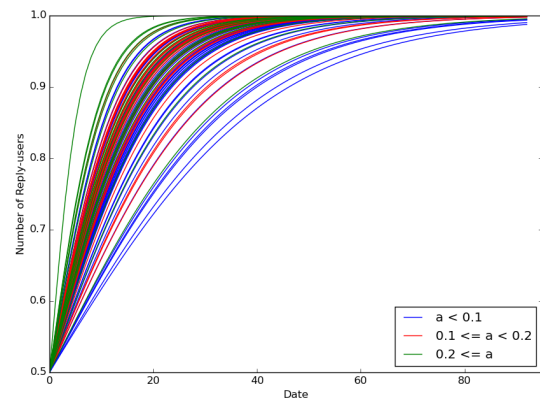
(c)  $0.2 \leq a$  のユーザ 18 名

図 4.3: 新規ユーザの Reply ユーザとの同期性 ( $syncrate \geq 0.3$  の割合)

### 4.1.5 分類結果の検証3 - reply ユーザのネットワーク成長 -



(a) Reply ユーザ数の累計数



(b) 累計数で正規化した後、シグモイド関数で近似

図 4.4: Reply-Reply ユーザの増加曲線

図 4.4a と図 4.4b は、Reply ユーザの Reply ユーザ（以下 Reply-Reply ユーザと呼ぶ）の獲得傾向を示す。ここでは 3 つのユーザ群に属するユーザをそれぞれ赤・青・緑の色で表している。このグラフを比較すると、青 ( $a < 0.1$  のユーザ) は図 4.4a においてはそれほど上位に位置しておらず、むしろ赤・緑の Reply ユーザの方が Reply-Reply ユーザの累計数が多いことがわかる。一方図 4.4b では、シグモイド関数の収束が遅いユーザの多くが青である。すなわち  $a < 0.1$  の Reply ユーザと関与した Reply ユーザは、同様にネットワーク成長が活発であるということになる。

### 4.1.6 分類結果の考察

4.1.3, 4.1.4, 4.1.5 節で述べた検証をもとに、本研究が提案する分類法の特徴をまとめる。まず第一に、本分類法を用いることで継続的な投稿活動が期待できるユーザを発見することが可能であるといえる。具体的には、表 4.1 に示した結果より、ネットワーク成長が盛んなユーザは近い将来の離脱の可能性が低いことがわかった。またそのようなユーザの多くは、4.1.4 節で検証したように、Reply ユーザとの弱い同期性をもつ。すなわち特定のユーザと密な関係を持つのではなく、広くユーザと浅い関係性をもつ特徴がある。このようなユーザを見つけるという点において、ネットワーク成長を示すパラメータ  $a$  を用いることは非常に有効であるといえる。

最後に、Reply ユーザ間のつながりには再帰的な構造があることが 4.1.5 節の結果からわかる。これは、Reply ネットワーク成長が盛んなユーザは、同様にネットワーク成長が盛んな



ユーザと関係を持っているということである。ここから考えられることとして、今回の分類法で分類したユーザを他ユーザに関与させることで、そのユーザのネットワーク成長を促進することができるかもしれない。これについては、以下の4.2節で詳細を検証する。

## 4.2 ネットワーク成長が他ユーザに与える影響力の検証

次に4.1節で分類したユーザを用いて、他ユーザへの影響力について実験を行い検証する。ここでいう影響力とは、ユーザの投稿活動に与える影響である。すなわち、Reply関係を形成することでネットワーク成長や投稿活動に変化があるか、また分類の3グループによって関わったユーザの変化の違いが見られるかを検証する。

### 4.2.1 検証に用いたデータセット

今回は4.1節の分類結果と4.1.1節で述べたデータセットから以下の条件を満たすユーザを検証に用いた。

- 表4.1に示した分類結果から、各グループそれぞれ6名を影響ユーザとして計18名。
- 各影響ユーザと2012/07/01 - 2012/07/31の区間においてReplyが確認できるユーザを被影響ユーザとしてそれぞれ10名(各グループ60名、計180名)。

### 4.2.2 検証内容

被影響ユーザが影響ユーザと初めてReply関係を築いた日を起点として、前後3ヶ月間における投稿活動量、およびネットワーク成長のパラメータ $a$ を算出し、その上で影響ユーザとの関与の前後を比較する。なお投稿活動量については、ユーザの30日毎の各日の平均Tweet数を検証するとともに、前3ヶ月と後3ヶ月における総tweet数を比較した。

### 4.2.3 検証結果

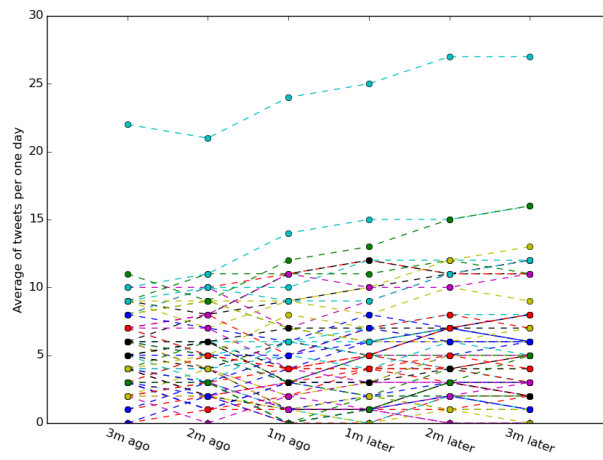
図4.5に各グループの被影響ユーザの前後3ヶ月(計6ヶ月)の毎月各日の平均tweet数を示す。また表4.3は影響ユーザとReply関係を築いた前3ヶ月と後3ヶ月において、どちらのほうが投稿活動量が多かったかを示す。

図4.6はネットワーク成長の変化を表した図となる。まず影響ユーザと関与する前3ヶ月におけるネットワーク成長をもとに本研究が提案している方法で分類し3グループに分ける。その後、関係を持った後3ヶ月間でのネットワーク成長から同様にユーザを分類した場合に、どのようにグループ間で遷移が発生したかを図示している。丸に付された数字はその段階で分類

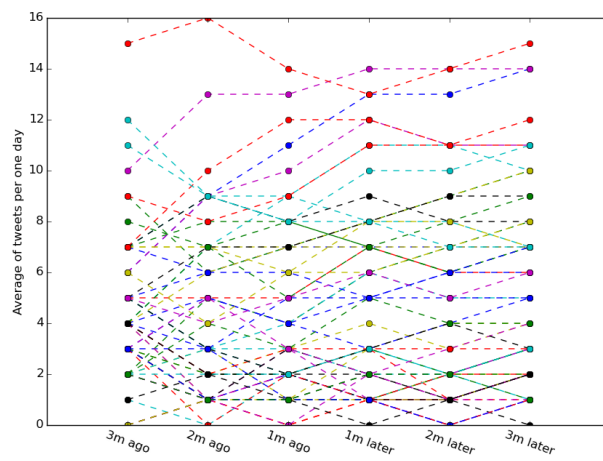
された人数，矢印に付された数字は3ヶ月間で遷移した人数である．

表 4.3: 被影響ユーザの前後3ヶ月間における投稿活動量の比較

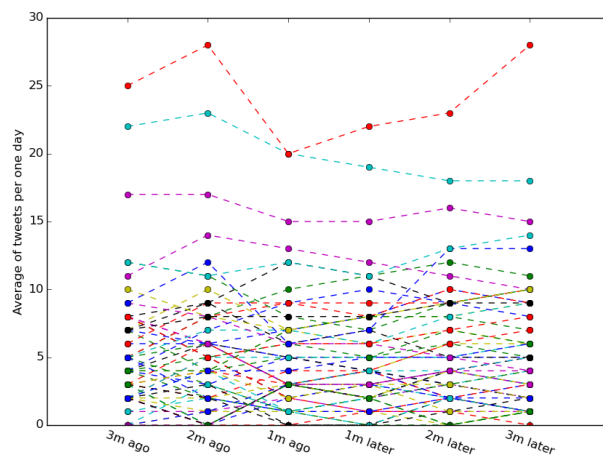
	前3ヶ月 > 後3ヶ月	前3ヶ月 < 後3ヶ月
$a < 0.1$ に対する被影響ユーザ	18	42
$0.1 \leq a < 0.2$ に対する被影響ユーザ	21	39
$0.2 \leq a$ に対する被影響ユーザ	28	32



(a)  $a < 0.1$  に対する被影響ユーザ 60 名

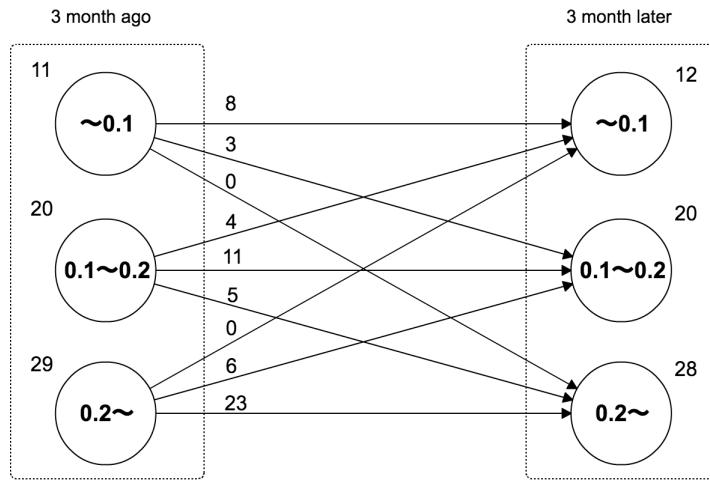


(b)  $0.1 \leq a < 0.2$  に対する被影響ユーザ 60 名

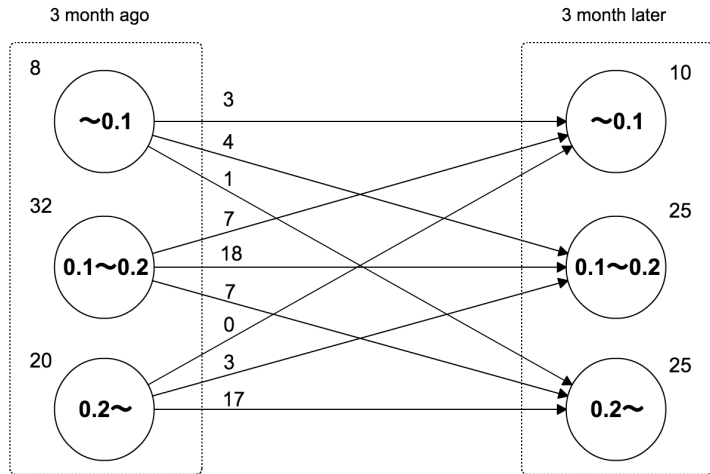


(c)  $0.2 \leq a$  に対する被影響ユーザ 60 名

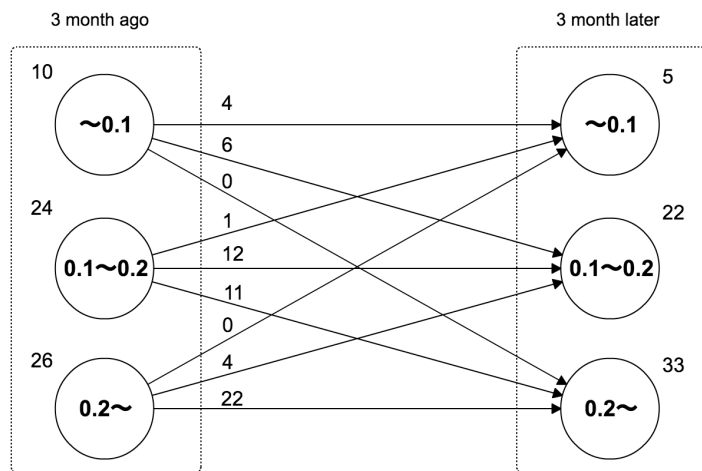
図 4.5: 前後 3 ヶ月間における毎月の各日平均 tweet 数



(a)  $a < 0.1$  に対する被影響ユーザ 60 名



(b)  $0.1 \leq a < 0.2$  に対する被影響ユーザ 60 名



(c)  $0.2 \leq a$  に対する被影響ユーザ 60 名

図 4.6: 前後 3 ヶ月間におけるネットワーク成長パラメータ  $a$  の変動

### 4.3 聞き役ユーザの検証

3.2 節において，ユーザのネットワーク成長を促すユーザに求められる特徴を述べた．本節では上述した検証結果をもとに，実際に他ユーザのネットワーク成長に貢献するユーザが，聞き役ユーザとしての役割を果たしているのか，またはどのような特徴を持つユーザが聞き役ユーザとなるのかを評価する．

#### 4.3.1 評価実験 1 -各特徴量に優れたユーザの影響力-

3.2 節で定義した特徴量， $R_{mf}$ ,  $R'_g$ ,  $R_{diff}$  に優れたユーザを影響ユーザとした場合の，被影響ユーザのネットワーク成長の変化を 4.2.2 節で述べた方法と同様にして評価する．図 4.7 に詳細図を示す．

まずデータセット内 2012/07/01 - 2013/03/31 の区間において Reply が確認できるユーザを無作為に 200 名，候補ユーザとして抽出する．次に候補ユーザ 200 名から，図 4.7 に示した 4 つの影響ユーザグループを作成する（計 40 名）．この影響ユーザと Reply 関係をもっているユーザをそれぞれ 10 名，被影響ユーザとして取得する（計 400 名）．被影響ユーザが影響ユーザと最初に関わった日を境に，前後 3 ヶ月におけるネットワーク成長の変化を図 4.6 と同様にして検証する．

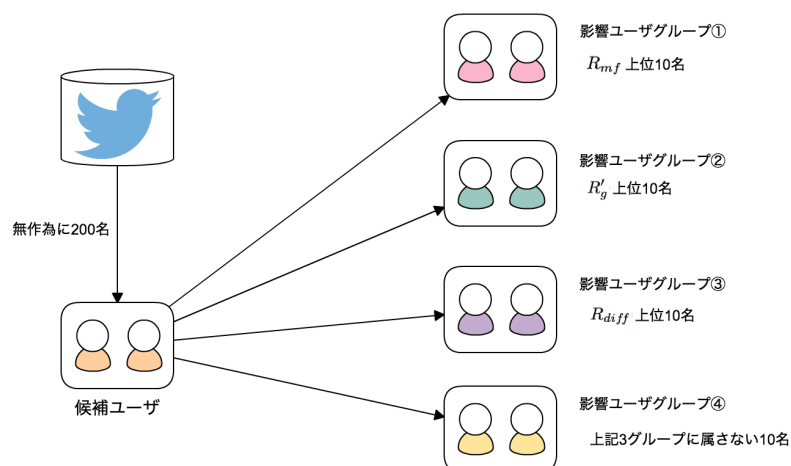


図 4.7: 聞き役ユーザ評価実験 1 の詳細図

表 4.4 に各グループの影響ユーザ 10 名を本研究の提案法で分類した結果を示す．グループ 2 においてネットワーク成長が活発であるユーザが多く見られることがわかる．グループ 2 で用いている  $R'_g$  では素性として時間単位あたりの Reply 数を考慮しているため，Reply ネットワーク成長が盛んなユーザが多く抽出できたと考えられる．それ以外のグループにおいては大

きな違いは確認できなかった。また図 4.8 は、各グループの被影響ユーザが影響ユーザと関与した前後でのネットワーク成長の変動を示している。形式は図 4.6 と同様である。

表 4.4: 提案法による各グループ 10 名の分類結果

	グループ 1	グループ 2	グループ 3	グループ 4
$a < 0.1$	4	6	4	3
$0.1 \leq a < 0.2$	3	3	4	5
$0.2 \leq a$	3	1	2	2

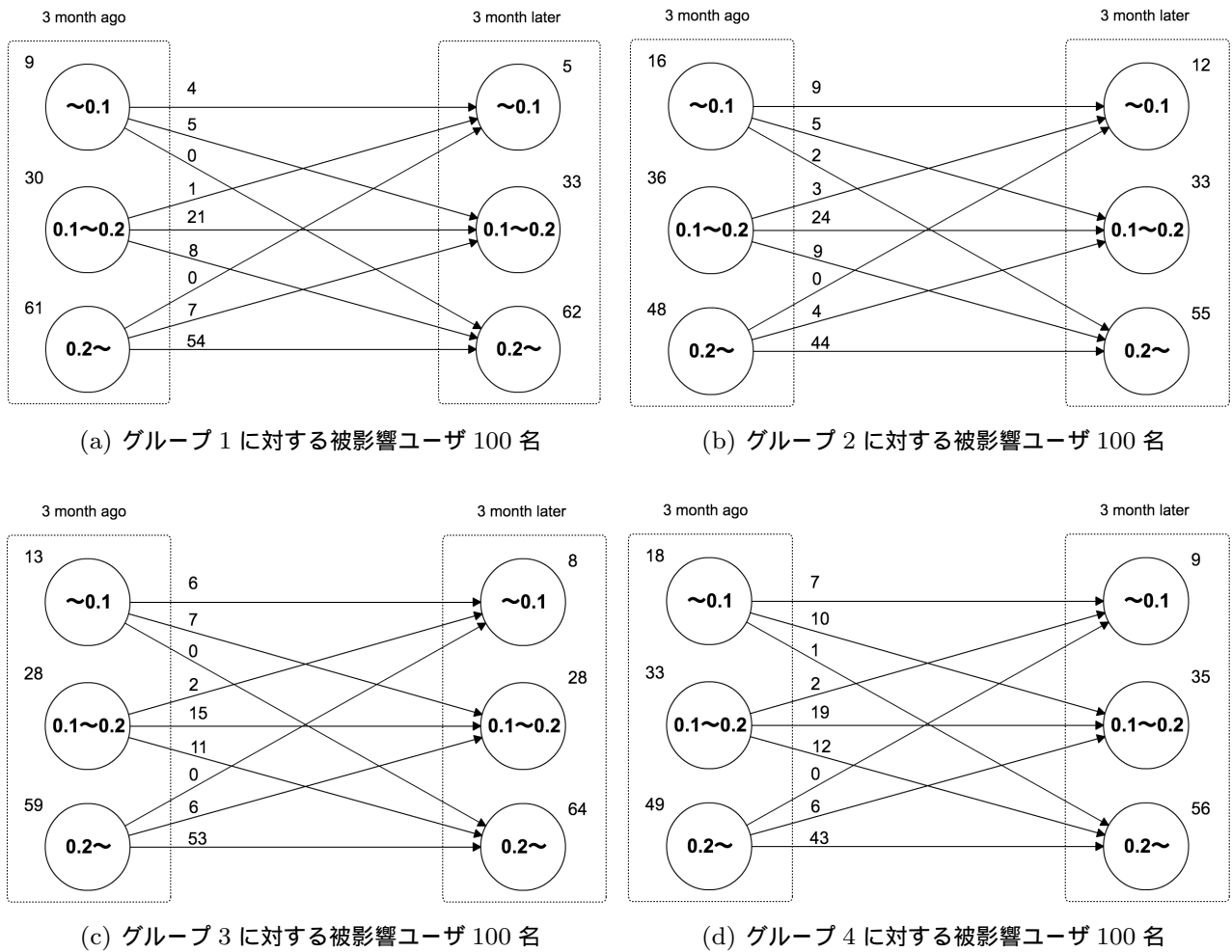


図 4.8: 被影響ユーザ 400 名の前後 3 ヶ月間における成長パラメータ  $a$  の変動

### 4.3.2 評価実験 2 -ネットワーク成長に貢献したユーザの各特徴量-

次に 4.3.1 とは逆に、まずネットワーク成長に貢献したユーザを抽出した上で、そのユーザの聞き役ユーザとしての特徴量がどのようになっていたのかを実験する。方法は以下のとおりである。

まずデータセット内 2012/07/01 - 2013/03/31 の区間において Reply が確認できるユーザ中から、3 ヶ月において  $a < 0.1$  を満たすユーザを 100 名（およびその区間）、影響ユーザとして抽出する。次にこの影響ユーザと Reply 関係をもっているユーザをそれぞれ 10 名、被影響ユーザとして取得する（計 1000 名）。被影響ユーザが影響ユーザと最初に関わった日を境に、前後 3 ヶ月におけるネットワーク成長の変化を図 4.6 と同様にして求める。最後に、図 4.9 のように 3 ヶ月後の被影響ユーザの遷移にスコアを導入し、その総和を影響ユーザのスコアとする。スコアが高いユーザは、被影響ユーザのネットワーク成長に貢献したユーザとなる。そのようなユーザが  $R_{mf}$ ,  $R'_g$ ,  $R_{diff}$  においてどのような特徴を持つか検証する。

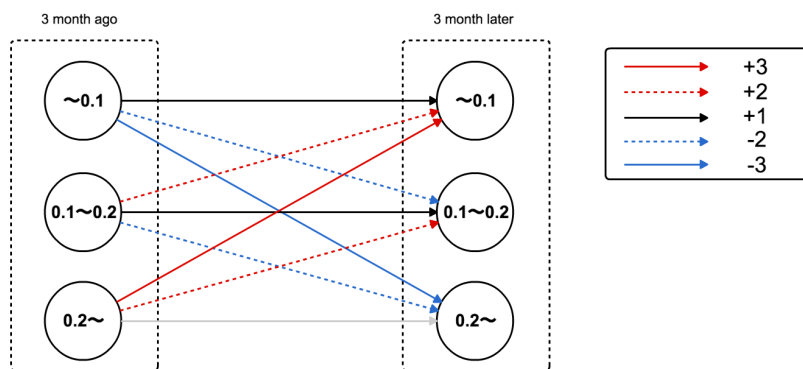


図 4.9: 聞き役ユーザ評価実験 2 に用いるスコア

表 4.5 に影響ユーザ 100 名をスコア付けした結果と各ユーザの特徴量を示す。

表 4.5: 影響ユーザ 100 名のスコア上位 10 名・下位 10 名の聞き役ユーザとしての特徴量

順位	スコア	$R_{mf}$	$R'_g$	$R_{diff}$
1	13	0.74	0.60	0.12
2	12	0.83	0.51	0.06
3	11	0.74	0.62	0.08
3	11	0.89	0.62	0.07
3	11	0.65	0.67	0.12
6	10	0.90	0.52	0.11
6	10	0.83	0.46	0.15
8	9	0.66	0.35	0.08
8	9	0.82	0.55	0.04
8	9	0.74	0.53	0.11
...				
91	-3	0.45	0.56	0.08
91	-3	0.73	0.41	0.07
91	-3	0.44	0.45	0.06
91	-3	0.25	0.63	0.12
95	-4	0.55	0.55	0.11
95	-4	0.33	0.60	0.12
95	-4	0.74	0.46	0.08
98	-5	0.33	0.44	0.05
98	-5	0.27	0.42	0.11
100	-7	0.43	0.45	0.09



## 第 5 章

# 考察

### 5.1 ネットワーク成長が他ユーザに与える影響力に関する考察

表 4.3 より,  $a < 0.1$ ,  $0.1 \leq a < 0.2$  の影響ユーザと Reply 関係を築いた場合, 投稿活動が活発となったユーザが多いことがわかる. またそのようなユーザは関係を構築後に活動量が増えるというわけではなく, 影響ユーザと関わる前 (今回の検証の場合 1 ヶ月程度前) から活動量が増え始め, その中で, 影響ユーザと新たに関係をもっているユーザが多い. 図 4.5a, 4.5b において, 1 ヶ月前から tweet 数が伸びているユーザが多いことも, それを裏付けている. 対して図 4.5c では, そのような傾向は確認できなかった.

図 4.6 は被影響ユーザの Reply ネットワーク成長の遷移を表している. 各図に共通して言えることとして, 一度 Reply ネットワークの成長が停滞したユーザ (図中 0.2 ~ に属するユーザ) については再度ネットワーク成長が活発になる可能性は低い. 加えて, 基本的にはネットワーク成長は時間経過とともに停滞していくため, 方向としては右下にユーザが遷移していく場合が多いことがわかる. そして表 4.3 の結果の大きな要因となったのは,  $0.1 \leq a < 0.2$  のグループに属するユーザの遷移と考えられる. 図 4.6a, 4.6b においては, 停滞するユーザと活発になるユーザがほぼ同じだけ存在する. 一方で図 4.6c では右上に遷移するユーザがほとんどいない.

以上のことから, ネットワーク成長が活発なユーザと関係を築くことは, 他ユーザにとって有益であるということが確認できた. 特に図 4.6 中の  $0.1 \leq a < 0.2$  グループに属するユーザのような, 投稿活動の過渡期を迎えているユーザの活動を後押しする上では, そのようなユーザの存在が重要になる.

## 5.2 ネットワーク成長と聞き役ユーザ

### 5.2.1 評価実験 1 に関する考察

図 4.8 中の 4 つの図についても，5.1 節で述べた特徴が共通していえる．各図を比較してみると，やはり  $0.1 \leq a < 0.2$  のグループに属するユーザの遷移に違いが生じている．図 4.6 のように上へ遷移するユーザはそれほど存在しないものの，下に遷移するユーザ数が図 4.8a, 4.8b と比べて図 4.8c, 4.8d で多くなっている．図 4.9 に示したスコアを各グループのユーザに適応したものを表 5.1 に示す．グループ 4 をベースラインとした場合，グループ 1 とグループ 2 においてスコアが高い結果となった．特にグループ 1 については，有意水準 5% で有意差が認められた．

表 5.1: 各グループのスコア平均

	スコア平均
グループ 1	3.32
グループ 2	2.74
グループ 3	2.12
グループ 4	2.37

### 5.2.2 評価実験 2 に関する考察

表 4.5 からは，スコアの高いユーザ，すなわち他ユーザに対して影響力を持つユーザほど  $R_{mf}$  の値が大きい傾向がわかる．一方で  $R'_g$  の値に上位と下位ともに有意差はなかった．これは 5.2.1 節で述べた結果と異なる．

## 第6章

# 結論

本論文では第一にユーザの Reply ネットワークの成長に着目し，そこからユーザの投稿活動の傾向を分析することを試みた．結果として，Reply ネットワークの成長は将来的な投稿活動の活性化につながり，かつ安定的な活動に寄与することがわかった．逆に停滞や離脱といった投稿活動の変化を予測する上では，Reply ネットワーク成長の収束が手がかりになるといえる．また Reply ネットワークの成長には再帰的な構造があり，他ユーザとの相互的なネットワーク成長が見込めることがわかった．すなわち継続的な Reply ユーザの獲得においては，継続的な Reply ユーザの獲得を実現しているユーザと関係を築くことが重要といえる．

ネットワーク成長が他ユーザに与える影響を評価するため，ユーザとの Reply 関係を起点とした前後のネットワーク成長の比較実験を行なった．その結果からネットワーク成長が盛んなユーザとの関係の有無によって，ユーザの投稿活動に変化が生じることを確認できた．また他ユーザに良い影響を与えると考えられる聞き役ユーザの特徴として，関心・関与・拡散の3特徴量を定義し，その効果について検証した．その結果として関心・関与の特徴量に優れたユーザには他ユーザのネットワーク成長を促す効果があり，特に他ユーザに対して高い関心を持っているユーザにそれが顕著なことがわかった．そのようなユーザを予め選定しユーザとの関わりを調査した結果からも，同様のことが言える．以上から本研究で提案したユーザ分類法および聞き役選定法の有効性を示せたといえる．

今後は本研究で得られた結果をもとに，Reply ネットワークの成長から，自身や他のユーザの成長を推定するモデルを検証したい．また本研究を考慮することによってユーザのネットワーク成長を志向したユーザ推薦を実現したいと考えている．

# 謝辞

第一に指導教員の佐藤哲司教授に心からの感謝の意を表します。先生には3年間熱心にご指導いただき、研究に限らず様々なことを学ばさせていただきました。また本研究を進めるにあたり、常に適切なアドバイスを行ってくださった副指導教員の松林麻実子先生、本当にありがとうございました。そして佐藤研究室の先輩である、山本修平さん、伏見卓恭さん、輪島幸治さん、同期である清野悠希さん、および後輩である大塚泰岳さん、神澤義人さん、高田叶子さん、武田悠佑さん、野崎祐里さんには研究室内外においてお世話になりました。佐藤研究室とともに切磋琢磨できたことに感謝しています。

## 参考文献

- [1] Isabel Anger and Christian Kittl. Measuring influence on twitter. In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, i-KNOW '11, pp. 31:1–31:4, New York, NY, USA, 2011. ACM.
- [2] Lars Backstrom and Jure Leskovec. Supervised random walks: Predicting and recommending links in social networks. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pp. 635–644, New York, NY, USA, 2011. ACM.
- [3] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. Everyone's an influencer: Quantifying influence on twitter. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pp. 65–74, New York, NY, USA, 2011. ACM.
- [4] Eytan Bakshy, Jake M Hofman, Winter A Mason, and Duncan J Watts. Identifying 'influencers' on twitter. In *Fourth ACM International Conference on Web Search and Data Mining (WSDM)*, 2011.
- [5] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, Vol. 3, pp. 993–1022, March 2003.
- [6] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna P. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- [7] John Hannon, Mike Bennett, and Barry Smyth. Recommending twitter users to follow using content and collaborative filtering approaches. In *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys '10, pp. 199–206, New York, NY, USA, 2010. ACM.
- [8] Akshay Java, Xiaodan Song, Tim Finin, and Belle Tseng. Why we twitter: Understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*,

- WebKDD/SNA-KDD '07, pp. 56–65, New York, NY, USA, 2007. ACM.
- [9] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pp. 591–600, New York, NY, USA, 2010. ACM.
- [10] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web, 1999.
- [11] Zhonghua Qu and Yang Liu. Interactive group suggesting for twitter. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, HLT '11, pp. 519–523, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.
- [12] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pp. 851–860, New York, NY, USA, 2010. ACM.
- [13] Vladimir N. Vapnik. *Statistical learning theory*. Wiley, 1 edition, September 1998.
- [14] Beidou Wang, Can Wang, Jiajun Bu, Chun Chen, Wei Vivian Zhang, Deng Cai, and Xiaofei He. Whom to mention: Expand the diffusion of tweets by @ recommendation on micro-blogging systems. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pp. 1331–1340, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.
- [15] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. Twitterrank: Finding topic-sensitive influential twitterers. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pp. 261–270, New York, NY, USA, 2010. ACM.
- [16] Shaomei Wu, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. Who says what to whom on twitter. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pp. 705–714, New York, NY, USA, 2011. ACM.
- [17] Yutaro Yamaguchi, Shuhei Yamamoto, and Tetsuji Satoh. Behavior analysis methods for twitter users based on transitions in posting activities. *IJWIS*, Vol. 10, No. 4, pp. 363–377, 2014.
- [18] Zhe Zhao and Qiaozhu Mei. Questions about questions: An empirical analysis of information needs on twitter. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pp. 1545–1556, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.
- [19] 山口裕太郎, 山本修平, 佐藤哲司. 投稿活動遷移に着目したマイクロブログユーザプロファ

イリングに関する一検討. *DEIM Forum 2014 B2-4*, 2014.