ソーシャルネットワークを活性化する『聞き役』ユーザの発見

Kikiyaku: Special Role Users Activating Community on Twitter

柘植 大 (Hiroshi TSUGE) [1]・佐藤 哲司 (Tetsuji SATOH) [2]・
伏見 卓恭 (Takayasu FUSHIMI) [3]・神門 典子 (Noriko KANDO) [4]
[1]筑波大学図書館情報メディア研究科 博士前期課程 ・ [2]筑波大学図書館情報メディア系 教授[i]
[3]東京工科大学コンピュータサイエンス学部 助教 ・ [4]国立情報学研究所 情報社会相関研究系 教授

［Abstract］
In this paper, focusing on the reply function and the retweet function as active communication tools in the Twitter service, we define a Kikiyaku user who positively tries to engage with other users by using these functions, and attempt to clarify the Kikiyaku users' role played in activation of communities. A Kikiyaku user is interested in other users and actively responds tweets of others by directly connecting to them or informs tweets of others to the members of belonging communities by retweeting the tweets. Based on an assumption that Kikiyaku users will also be involved in new users aggressively who started using Twitter, we aim to clarify whether Kikiyaku user's involvements can change the posting activities of new users. In our approach, we quantify changes of posting activities of each user based on the changes of the slope parameter of a sigmoid curve fitting cumulative number of reply users. Furthermore, we define three indices such as mutual following rate, average participation rate, and tweet diffusion power as characteristics of Kikiyaku users. In our experiments using large tweet data sets, we confirm whether these indices can characterize Kikiyaku users: two indices, mutual following rate and average participation rate, would represent characteristics of Kikiyaku users, however tweet diffusion power does not work well in our experiments.

［キーワード］
ソーシャルネットワーク、コミュニティ成長、ユーザプロファイル、聞き役

1. はじめに

近年、ツイッター[ii]に代表される SNS(Social Networking Service) が急速に普及し、コミュニティ内の情報インフラとして定着してきている。ツイッターの月間アクティブユーザ (MAU) 数は日々増加傾向にあり、すでに約 3 億人を超えているが、近年では実際には活動していない非アクティブユーザも増加傾向にあることが報告されている[1]。非アクティブユーザとは、アカウントは存在するが投稿活動を行っていないユーザであり、2014 年末時点で MAU 中の 8.5%にあたる約 2500 万人のユーザが投稿活動を行っていないとされている。

コミュニケーションツールとしてのツイッターでは、短文を投稿するツイート機能以外にも様々な投稿機能が提供されている。フォロー機能は、他者が投稿したツイートを自身のタイムライン上に表示するための機能であり、あるときフォローしたユーザのツイートは、それ以降、常に自身のタイムラインに表示される。このため、自身にとって情報源となるニュースサイトや興味・関心の高い有名人をフォローすることが一般的に行われているが、一度フォローすれば自動的にツイートが表示されることから、受動的なコミュニケーション手段であると言える。一方、リプライ機能は、投稿されたツイートに対して、その投稿者宛に返信する機能である。また、リツイート機能は、タイムラインに表示されたツイートを、自身をフォローしているユーザであるフォロワーのタイムラインに表示する機能である。このように、リプライ機能とリツイート機能は、ユーザが積極的に「リプライ」あるいは「リツイート」を行う必要があることから、能動的なコミュニケーション手段であると言える。

本論文では、能動的なコミュニケーション手段であるリプライ機能とリツイート機能に着目する。これらの機能を使って積極的に他者と関わろうとする姿勢を持つユーザを聞き役ユーザと称し、聞き役ユーザの存在がコミュニティの活性化、すなわち、リプライによって形成されるコミュニティの成長に果たす役割を明らかにする。

---

[i] satoh@ce.slis.tsukuba.ac.jp
[ii] https://twitter.com/

聞き役ユーザとは、他者への関心が高く積極的にリプライをすることで、そのユーザと直接的に繋がる、あるいは、他者のツイートをリツイートすることで、そのユーザの存在を既存コミュニティに知らせる役割を持つユーザである。特に、本論文では、ツイッターの利用を開始した新規ユーザに対して聞き役ユーザが関わることで、新規ユーザのその後の投稿活動に影響が出るかを実践的に解明することを目指す。

本論文の構成を以下に示す。まず第2章では、SNSにおけるユーザの影響力とネットワーク特性に関する関連研究を概観し、本研究の位置づけを明らかにする。第3章では、本論文における聞き役ユーザの特徴を定義し、ネットワークの成長傾向から聞き役ユーザを発見する手法を提案する。第4章では、著者らが収集した大量のツイートアーカイブを用いて、新規ユーザの成長に対する聞き役ユーザの役割を実験的に明らかにする。最後に、第5章で提案手法の有効性と限界、今後の課題について言及する。

２．関連研究

2.1 SNS ユーザの影響力に関する研究

ツイッター上の特定ユーザの影響力に関する研究はすでに数多く行われている。Hannon ら[2]はユーザプロファイルをもとにしたコンテンツベースの推薦手法を提案している。この手法ではユーザのツイート内容、フォロワー、フォロウィー、フォロワーのツイート内容、フォロウィーのツイート内容などの特徴量に基づくプロファイルを利用して、ユーザの特性を調査し、その結果をユーザ推薦に応用している。Cha ら[3]は、ツイッターユーザの影響力を 3 つの尺度（フォロワー数、被リツイート数、被メンション[iii]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 ら[4]は、ツイッターのメンション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。ユーザのツイート拡散能力を Social tie model や User influence mode といった独自のモデルによって定式化し、ユーザ推薦の精度向上を目指している。ツイート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「ツイートの拡散相手」として最も適したユーザの推薦を実現している。

2.2 SNS ネットワーク特性に関する研究

Java ら[5]の研究では、ツイッターのネットワーク特性を分析し、人々のツイッターの利用目的について考察している。ユーザは日常の情報発信・情報収集やコミュニケーションのためにツイッターを利用し、また属する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た。Yamaguchi らの研究[6][7]では、ユーザの投稿パターンや投稿数の変化に着目し、ツイッターの利用開始時期によって投稿パターンが異なることを明らかにしている。さらに、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[8]はツイッターユーザのランキング手法について、PageRank やリツイートによる方法を提案・比較している。Retweet Trees という手法を提案、ツイッターネットワークにおいてリツイートがどれだけ拡散するのかについても論じている。Weng ら[9]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にした。この手法では、PageRank の遷移確率行列を LDA[11]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[12]は、ツイッターネットワークに対して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、SNS ネットワークにおける潜在的なエッジの検出を行っている。

以上で述べたように、SNS やツイッターに関する研究はすでに数多くなされている。本研究では、Wang らの研究[4]と同様にリプライ（メンション）機能に着目したユーザの影響力の定量化方法を提案する。また、SNS ネットワーク特性にも着目し、特に投稿活動量やリプライネットワークの成長を考慮に入れている点では Yamaguchi らの研究[6][7]と類似していると言える。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を志向している点に特徴がある。

３．聞き役ユーザの特徴と発見手法の提案

3.1 聞き役ユーザとは

本論文では、ソーシャルネットワークの形成／成長に聞き役となるユーザの存在が大きいとの仮説を立てる。「聞き役」あるいは聞き役ユーザとは、他ユーザの投稿を積極的に傾聴し関係を構築することに前向きな姿勢を

[iii] メンションはリプライと同等の機能であり、ツイート本文中に「@user_name」を記述すると、user_name のタイムラインにツイートを送ることができる。リプライが一人のユーザ向けであるのに対して、@を付けてユーザを指定することで複数名にツイートを拡散することができる。

持つユーザとする。現実世界においても、「おせっかい者」と呼ばれる世話焼きが他者を仲介する場面が存在するが、ネットワーク上のソーシャルコミュニティにおいても、このような「聞き役」となるユーザの存在がコミュニティの拡大あるいは活性化に有用であろうという仮説である。特に、コミュニティに馴染んでいない新規ユーザは、発言（投稿）すること自体に障壁があり、そのような場合に、発言に耳を傾けて更なる発言を誘導してくれる聞き役の存在は大きいであろう。

以下、本論文では、ソーシャルコミュニティとしてツイッター環境を前提として、新規ユーザに対して聞き役となるユーザに限定して議論を進める。聞き役ユーザを中心としたユーザ間の関係性を図-1 にモデル化する。

（1）新規ユーザへの直接的な関与

ツイッターでは、未知な新規ユーザのツイートであっても検索等の機能を使って閲読することができる。この時、ツイートを単に閲読するだけでなく、そのツイートを投稿したユーザに関心を持って、直接的な返信であるリプライを返す、あるいは、そのユーザをフォローする。特にリプライは、返信のためのメッセージ（本文）を考えて投稿しなければならないので、新規ユーザに対してその手間を超える関心を持ったと見なすことができる。リプライを返すことで聞き役ユーザが高い関心を持っていることが新規ユーザに伝わり、両者の間に密なコミュニケーション関係、すなわち相互フォローの関係が構築されることが期待される。

（2）新規ユーザの広報

新規ユーザのツイートを聞き役ユーザがリツイートすることで、聞き役ユーザをフォローしているユーザ集合、つまり聞き役ユーザのフォロワー集合に伝播する。各フォロワーから見ると、聞き役ユーザを介して新規ユーザの存在を知ることになり、フォロワー内に上記（1）の条件を満たすユーザが存在すれば、新規ユーザに対してリプライあるいはフォローを行う。こうして、聞き役ユーザのリツイートをきっかけとして、自身のフォロワーと新規ユーザとの間に相互フォロー関係が構築されることが期待される。

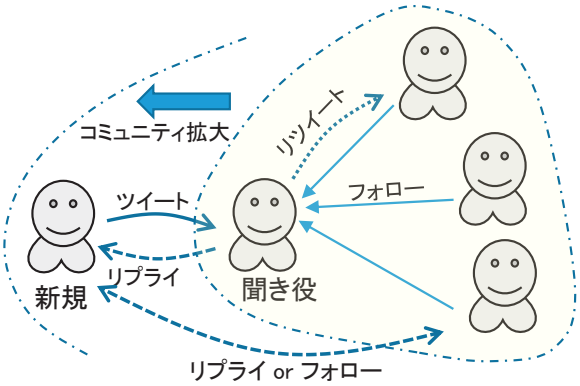


図-1 聞き役ユーザのモデル

ツイッターの基本機能であるフォローは、フォロー先のユーザ（フォロウィー）の了解無しに一方的にツイートが流れてくる関係を形成する。このため新規ユーザは、図-1 における聞き役ユーザあるいはそのフォロワーをフォローすることができるが、それだけでは非対称な関係性しか形成されない。コミュニティの一員として新規メンバーが組み込まれるためには、相互にフォロー関係にある、あるいはお互いにリプライをしあうなど対称な関係を築く必要があり、この対称な関係を構築するのに上述の聞き役ユーザが重要な役目を果たすと考える。本論文では、過去に他のユーザと対称な関係を形成しているユーザは、将来的にも（新規ユーザに対しても）対称な関係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とする。

図-1 において、新規ユーザが相互フォローを形成するユーザとして、聞き役ユーザ以外にも聞き役ユーザのフォロワーが考えられる。この場合、聞き役ユーザは直接的には新規ユーザと繋がってはいないが、自身のフォロワーと新規ユーザとが対称な関係を築くことに貢献したと考え、この場合に聞き役ユーザが重要な役割を果たしたとする。

3.2 リプライユーザ数に着目したユーザプロファイル

本論文では、他者と相互フォローを形成している、あるいは、相互にリプライを投稿しあう関係の形成に寄与するユーザを聞き役ユーザという。そこで、各ユーザのプロファイル（特徴量）として、リプライユーザ数に着目する。これは、リプライ関係にあるユーザ数を表す指標であり、特定のユーザに対して複数回リプライを返してもユーザ数としては1とする。これは、特定のユーザと極めて親しい関係を持つユーザ同士が頻繁にリプライをしてもコミュニティの拡大には寄与しないと考えられることから、リプライ関係にあるユニークユーザ数を採用するのが適当であるとしたからである。

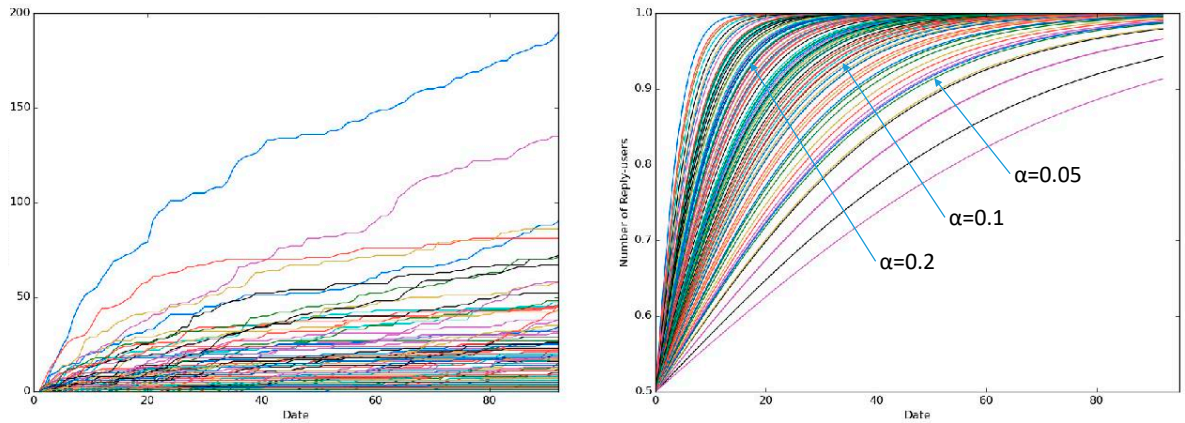
また、フォロー関係の構築もコミュニティの形成には重要な指標であると考えられるが、フォロー機能の特性から、一旦フォローしたユーザとは、積極的にフォロー関係を解消する操作を行わない限りフォローが継続される。特にフォロー先のユーザ（フォロウィー）が投稿活動を停止すると、自身のタイムラインに流れてくるツイートもなくなることから、フォローをしたことさえも忘れてしまうことも発生する。このためフォローは、他のユーザと関係構築をする際の指標としては有用となるが、フォローユーザ数の多寡は、ユーザの現時点での活動状態を表す指標として有用とは言えない。

以上のことから、本論文では、各時点におけるユーザのコミュニティ拡大・形成に寄与する特徴量として、リプライユーザ数を採用することとした。ツイッターを開始した直後のユーザからランダムに抽出した100名に対して、その後の3ヶ月間(90日)にリプライ記事を投稿したユーザ数の累積値を図-2に示す。図-2(a)の縦軸は、各ユーザがリプライしたユーザ数とリプライを受けたユーザ数を合算した異なりユーザ数(リプライユーザ数)を表している。3ヶ月経過後に最も多く獲得したユーザのリプライユーザ数は193名であり、最後までリプライユーザ数が0であったユーザは6名であった。リプライユーザ数の増加傾向に着目すると、多くのユーザはツイッター開始直後の2週間程度でリプライユーザ数を増やしているが、その後は増加傾向が緩やかになり飽和する傾向を示している。図-2(a)は、3ヶ月経過後も増加傾向にあるユーザが一定程度存在することを示しているが、コミュニティが無限に拡がることは考えにくいことから、ユーザ毎に定まる時間経過後に飽和すると考えるのが自然である。

そこで、飽和関数として知られているシグモイド関数を用いて、各ユーザのリプライユーザ数の累積値から各ユーザの成長特性をモデル化する。シグモイド関数とは、次式で示すように時刻tの増加に伴って、パラメータαが大きいほど急速に立ち上がって早期に飽和する関数である。最後までリプライユーザ数が0であった6名を除いた各ユーザの成長特性を表すパラメータαの算出は、$(f(t,\alpha)-y(t))$の二乗誤差の総和が最小となるように最適化する。ここで$y(t)$は、時刻tにおいてユーザが獲得したリプライユーザ数を正規化した値である。

$$f(t,\alpha)=\frac{1}{1+e^{-\alpha t}}$$

図-2(a)に示した100名のユーザから、最後までリプライユーザ数が0であったユーザ6名を除いた各ユーザの成長曲線を図-2(b)に示す。この図において、早期に立ち上がって飽和しているユーザほどαの値は大きく、$\alpha \geqq 0.2$のユーザは、十分早期に飽和し時間が経過しても新規リプライユーザを獲得できないユーザと見なすことができる。一方、$\alpha < 0.1$のユーザは、リプライユーザの獲得が緩やかであり、成長が飽和していないユーザであると見なすことができる。



(a) 実測した成長曲線　　(b) シグモイド関数で近似した成長曲線

図-2 新規ユーザ100名のリプライユーザ数の累積値

3.3 聞き役ユーザの発見手法

3.1 節で定義したように、新規ユーザに対する聞き役ユーザの役割は、（1）新規ユーザへの直接的な働きかけ、および（2）新規ユーザの広報の二つである。これらの役割を定量的に評価することを目的に、以下の 3 種類の指標を定義する。

i）相互フォロー率$R_{mf}(u)$

ツイッターを利用しているユーザ間は、フォロー関係によって比較的安定したネットワークを形成する。そこで、ユーザをノード、フォローをエッジとした有向ネットワーク$\mathrm{G}=(\mathrm{V},\mathrm{E})$を考える。この時、$(\mathrm{u},\mathrm{v})$はユーザuがユーザvをフォローしていることを表し、uが（一方的に）フォローしているユーザ数を$|\Gamma_{following}(u)|$、uが双方向にフォローしているユーザ数を$|\Gamma_{mutual}(u)|$とすると、相互フォロー率$R_{mf}(u)$は以下の式で与えられる。

$$R_{mf}(u)=\frac{|\Gamma_{mutual}(u)|}{|\Gamma_{following}(u)|}\leqq 1$$

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

ii）平均関与率 $R_g(\mathrm{u})$

時間とともに変化する新規ユーザに対する直接的働きかけを評価するために、単位時間あたりの関与を複数の特徴量で計測した上で平均化する。評価に用いた特徴量を表-1 に示す。これらの指標を$x_j(1\leqq j\leqq 5)$として、全区間kにおける平均関与率$R_g(\mathrm{u})$を以下の式で算出する。

$$R_g(u)=\frac{1}{5}\sum_{j=1}^{5}\left(\frac{1}{k}\sum_{i=1}^{k}R'_{g,i}\,(u,x_j)\right)$$

ここで$R'_{g,i}(u,x_j)$は、単位区間$\mathrm{i}(1\leqq \mathrm{i}\leqq \mathrm{k})$における特徴量$x_j$の値が全区間の平均値よりも大きい場合は 1、小さいか等しい場合は 0 とする二値を取る関数とする。したがって、平均関与率が正の値となるユーザuは、表 1 に示す 5 種類の特徴量が平均してプラス、すなわち増加傾向にあることを示している。これらの指標は、いずれも他ユーザとの関わりの程度を表す特徴量であることから、平均関与率が正の高い値を示すユーザは、より積極的に他者と関わる傾向が強いことを示す。

表-1 平均関与率の算出に用いた特徴量

j	変数名	説明
1	$N_{follower}$	自分をフォローしているユーザ数（フォローされているユーザ数）
2	$N_{following}$	自分がフォローしているフォロー先のユーザ数
3	$N_{favorites}$	お気に入り（フェイバリット）として登録しているユーザ数
4	$N_{retweet}$	指定された単位時間の中でリツイートをした回数
5	N_{reply}	指定された単位時間の中でリプライをした回数

iii）ツイートの拡散力$R_{diff}(u)$

聞き役ユーザのモデル（図 1）に示した聞き役となるユーザは、自身のフォローユーザからなるコミュニティに対して、新規ユーザのツイートをリツイートすることでツイートを拡散する。このため、拡散力は、リツイートする頻度とフォロワー数によって決まる。そこで、ユーザuのツイート拡散力$R_{diff}(u)$を以下の式で算出する。

$$R_{diff}(u) = \frac{N_{retweet}(u)}{N_{tweet}(u)} \times \log(|\Gamma_{follower}(u)|)$$

ここで、ユーザuが新規ユーザのツイートをリツイートする確率は、uがこれまでに投稿した全ツイートの内でリツイートが占める割合、すなわち、$N_{retweet}(u)/N_{tweet}(u)$で近似できるとした。すなわち、ユーザは、新規ユーザであるかの区別はせずに、自身のタイムラインの中で一定の比率でリツイートをするとした。また、フォロワー数を対数指標としたのは、一般にフォロワー数の分布はべき乗則に従うこと、有名人などの大量のフォロワーとなるユーザは、図-1 で示したリツイートを閲覧してリプライやフォローなどの直接的な関与に発展する可能性は低いと考えたからである。

4．実験と評価

4.1 実験方法

実運用されているツイッターから収集したデータに、前章の i)から iii）で示した評価式を適用し、聞き役ユーザとして高い特性を持つユーザに関与することで、新規ユーザのツイッター投稿が長期化するかを実験的に明らかにする。

提案手法の評価には、2011 年 11 月から収集を開始した日本国内で投稿されたツイート[13]をデータセットとして使用する。ツイートの収集には、ツイッターの Search API[iv]を使用し、言語に″ja″（日本語）と、日本全域をカバーする位置情報[v]とを検索条件として指定することで、日本国内から投稿された日本語のツイートを収集している。

データセット内から 2012 年 7 月 1 日から 2013 年 3 月 31 日の期間において、リプライが確認できるユーザを無作為に 200 名抽出し、図-3 に示すように聞き役ユーザの候補とする。この候補ユーザの中から前章の提案法で示した相互フォロー率$R_{mf}(u)$、平均関与率R_g(u)、ツイート拡散力$R_{diff}(u)$の高いユーザ各上位 10 名と、それ以外から無作為に抽出した 10 名を抽出し影響ユーザとする。この段階で影響ユーザは、各影響グループ①から④にそれぞれに 10 名、計 40 名が抽出されている。

次に、この影響ユーザとリプライ関係にあるユーザをそれぞれ 10 名、計 400 名を被影響ユーザとして抽出する。抽出した 400 名の被影響ユーザが影響ユーザ/聞き役ユーザと関わることで、自身のネットワーク成長に現れる変化を評価する。

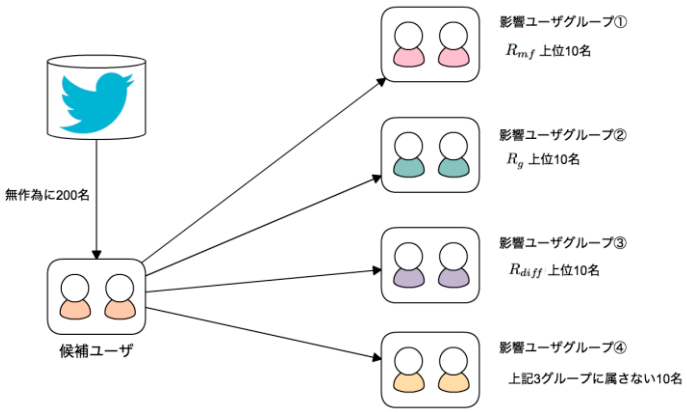


図-3 聞き役ユーザの評価を行うユーザの抽出

被影響ユーザのネットワーク成長の変化は以下の手順で評価する。

(1) 各被影響ユーザが聞き役ユーザからリプライを受けた日を基準日とする。
(2) 基準日から 3 ヶ月前の時点でのユーザプロファイル（パラメータα_P）を 3.2 節に述べた手法を用いて決定する。同様にして、基準日から 3 ヶ月後のパラメータα_Lを決定する。
(3) パラメータ（α_P、α_L）の値はユーザの成長が飽和する早さを表していることから、

iv http://search.twitter.com/search.json
v 兵庫県西脇市を中心とする半径 2,000km 圏内

$\alpha_P < \alpha_L$ ・・・・・・・・・ 時間の経過とともに成長が飽和する傾向が高まる。
$\alpha_P = \alpha_L$ ・・・・・・・・・ 時間の経過によらず成長に変化はない。
$\alpha_P > \alpha_L$ ・・・・・・・・・ 時間の経過とともに飽和していた成長が新たな成長へと遷移している。

以上により、聞き役ユーザと関係を持つことで被影響ユーザの成長に変化が生じるかを評価する。評価に際しては、3.2 節で示したユーザプロファイルのパラメータ α の値でユーザを 3 グループに分割し、グループ間を遷移したユーザ数の多寡から聞き役ユーザの影響力を測定する。ユーザの分割は、図 2-(b)の結果を参考にユーザの成長段階から飽和の程度に応じて表-2 に示す 3 グループとした。特に、成長が止まったユーザ（飽和：$\alpha_P \geqq 0.2$）が、聞き役ユーザと関係を持つことで成長段階（$\alpha_P < 0.1$）に遷移した場合は、聞き役ユーザは被影響ユーザのネットワーク形成に大きな影響を与えたと言える。

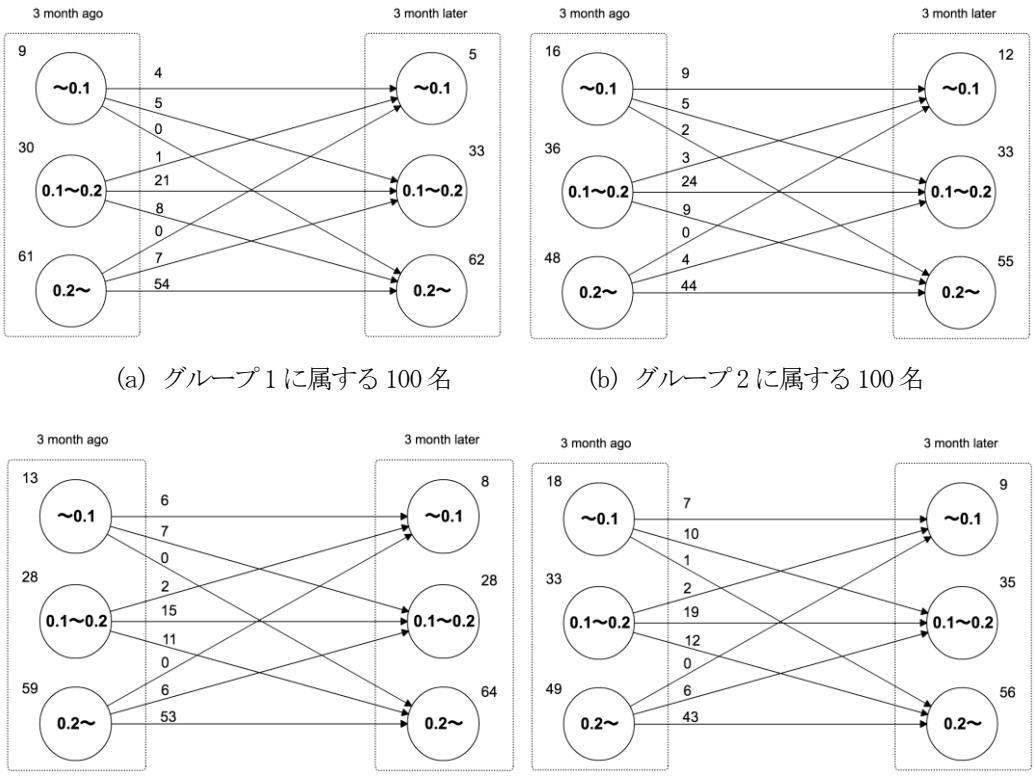
表-2 ユーザの成長

パラメータα_Pとα_Lの値	成長段階
$\alpha < 0.1$	大（飽和していない）
$0.1 \leqq \alpha < 0.2$	中
$0.2 \leqq \alpha$	小（既に飽和している）

4.2 実験結果

聞き役ユーザと関係を持った被影響ユーザを対象に、時間経過に伴う成長の変化を図-4 に示す。この図の(a)から(d)に属する各 100 名は、図-3 に示した方法で区分した被影響ユーザである。各図の四角で囲った枠線内の丸印の中の数値は、表-2 に示した成長を表すパラメータ値であり、各枠内には 3 つのクラスがある。また、丸印の左肩に書いてある数値は、各 100 名の内で当該クラスに属するユーザ数を示している。

左側の四角い枠は、聞き役ユーザと関係を持つ 3 ヶ月前を示しており、右側の枠は 3 ヶ月後、すなわち、左側の枠から右側の枠まで 6 ヶ月の時間経過があり、その期間中にクラス間を遷移したユーザ数が矢印線の上に表記されている。



(a) グループ 1 に属する 100 名　　(b) グループ 2 に属する 100 名

(c) グループ 3 に属する 100 名　　(d) グループ 4 に属する 100 名

図-4 聞き役ユーザの影響評価

図-4 の(a)から(d)のいずれにおいても、3 ヶ月前に成長が飽和していたユーザ（$\alpha_p \geqq 0.2$)の多くは、時間経過後も成長が飽和しているクラス（$\alpha_L \geqq 0.2$)に留まり、成長傾向に変化は見られない。この傾向はランダムに選ばれた聞き役ユーザの場合（図-4（d））でも同様であり、49 名中の 43 名が飽和したままの状態を持続している。一方で、成長が飽和していないクラスに属するユーザの遷移傾向は、図-4 の各図において大きく異なっている。この違いは、聞き役ユーザとして、図-4(a)は相互フォロー率$R_{mf}(u)$の高いユーザを、図-4(b)は平均関与率 $R_g(\mathrm{u})$ の高いユーザを、図-4(c)はツイート拡散力$R_{diff}(u)$ の高いユーザを設定していることに起因している。

そこで、聞き役ユーザの特性の違いが、ユーザの成長に与える影響を評価するために、図-4 の結果を図-5 に示す凡例にしたがって数値化する。期間 6 ヶ月が経過した後に、ユーザが 3 段階に分類されたクラスを変わらずに遷移したら＋１点を加算、成長が飽和していない方向に 1 段階遷移したら＋２点、2 段階遷移したら＋３点を加算することとし、成長が飽和する方向に遷移したらマイナスのスコアを与えるとした。グループ毎に 100 名の被影響ユーザのスコアを平均した値を表-3 に示す。また、t-検定を行った結果、グループ 1 とグループ 4 との間には、有意傾向があること（両側検定で $\mathrm{p} < .10$）が確認できた。

以上の結果から、聞き役ユーザとして最も影響力（成長寄与）があるのはグループ 1、すなわち、相互フォロー率の高いユーザであることが明らかとなった。相互フォロー率の高いユーザは、新規ユーザが自分をフォローしてきた時に、フォロー返しを行う可能性が高いことから、新規ユーザと安定した相互フォロー関係を構築することが期待される。また、平均関与率ならびにツイート拡散力の高いユーザグループも、ランダムに選択したユーザよりも高い影響力を示すことが明らかとなった。ツイート拡散という行為は、被影響ユーザとは直接関与していないが、被影響ユーザのネットワーク拡大に少なからず効果があることが明らかとなったことは、大きな発見であると言える。

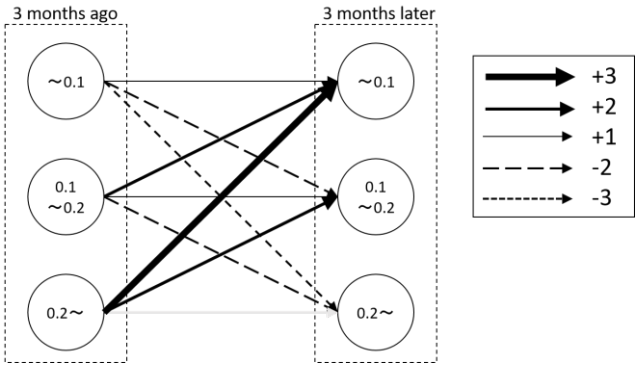


図-5 ユーザの成長を定量化する凡例

表-3 聞き役ユーザの特性によるユーザの成長傾向

グループ	聞き役ユーザの特性	影響力
1	相互フォロー率の高いユーザ	0.69
2	平均関与率の高いユーザ	0.57
3	ツイート拡散力の高いユーザ	0.54
4	ランダムに選択したユーザ	0.38

4.3 考察

上述のグループ分けした実験／評価によって相互フォロー率、平均関与率、ツイート拡散力の順で聞き役ユーザとしての効果が大きいことが明らかとなった。しかし、実際には一人の聞き役ユーザが上記 3 つの指標を持っている。そこで、一人一人の聞き役ユーザに着目して、3 つの指標の組み合わせを調査する。

まず、前述と同じデータセットから以下の 2 条件を満たすユーザ 100 名を聞き役ユーザとして抽出する。

・ 2012 年 7 月 1 日から 2013 年 3 月 31 日の区間に 1 回以上のリプライを行っている。
・ 上記のリプライから 3 ヶ月経過後にシグモイドパラメータが $\alpha < 0.1$を満たしている。

抽出した聞き役ユーザとリプライ関係があるユーザをそれぞれ 10 名ずつ、被影響ユーザとして抽出する。

被影響ユーザが最初に聞き役ユーザと関わった日を基準に、前後 3 ヶ月におけるネットワーク成長の変化を図-4 と同様にして求める。その上で、図-5 に示した凡例にしたがって被影響ユーザのスコアを算出し、聞き役ユ

ーザ毎に対応する被影響ユーザのスコアを合算して、聞き役ユーザのスコアとする。

100 名の聞き役ユーザを上記のスコアが降順となるように並べた際の、上位 10 名と下位 10 名の結果を表-4 に示す。この結果から、上位 10 名の相互フォロー率$R_{mf}(u)$の値は、下位 10 名の値より大きいことが分かる。例外的に、順位 91 と 95 の聞き役ユーザの相互フォロー率は高い値を示しているが、両名の平均関与率$R_g(\mathrm{u})$は相対的に低い値となっている。また、上位の中で相互フォロー率が最低の値 0.65 となっている聞き役ユーザの平均関与率は高い値となっている。このことから、相互フォロー率と平均関与率は、ネットワーク成長に相補的に機能していることが伺える。一方、ツイート拡散力は、上位と下位で大きな傾向の差異も見られず、値自体も小さいことが明らかとなった。

表-4 聞き役ユーザの上位 10 名と下位 10 名の特徴量

聞き役ユーザ上位 10 名					聞き役ユーザ下位 10 名				
順位	スコア	$R_{mf}(u)$	$R_g(\mathrm{u})$	$R_{diff}(u)$	順位	スコア	$R_{mf}(u)$	$R_g(\mathrm{u})$	$R_{diff}(u)$
1	13	0.74	0.60	0.12	91	−3	0.45	0.56	0.08
2	12	0.83	0.51	0.06	91	−3	**0.73**	**0.41**	0.07
3	11	0.74	0.62	0.08	91	−3	0.44	0.45	0.06
3	11	0.89	0.62	0.07	91	−3	0.25	0.63	0.12
3	11	**0.65**	**0.67**	0.12	95	−4	0.55	0.55	0.11
6	10	0.9	0.52	0.11	95	−4	0.33	0.60	0.12
6	10	0.83	0.46	0.15	95	−4	**0.74**	**0.46**	0.08
8	9	0.66	0.35	0.08	98	−5	0.33	0.44	0.05
8	9	0.82	0.55	0.04	98	−5	0.27	0.42	0.11
8	9	0.74	0.53	0.11	100	−7	0.43	0.45	0.09

5. まとめ

本論文では、ソーシャルネットワークの形成・成長に聞き役となるユーザの存在が大きいとの仮説を立て、聞き役ユーザが満たすべき要件を定量的な評価によって検証した。検証はソーシャルネットワークとして広く普及しているツイッターを用いて行い、リプライ関係にあるユーザ数の累積値をシグモイド関数で近似し、シグモイドパラメータの値で新規ユーザのソーシャルネットワークの成長をモデル化した。

聞き役ユーザの特徴として、相互フォロー率、平均関与率、ツイート拡散力の 3 つの指標を定義した。大量に収集したツイートデータを用いた実験を行い、これら 3 つの指標が聞き役ユーザを特徴付けることができることを明らかにしている。特に、新規ユーザと直接的に関与すると期待される相互フォロー率の高い聞き役ユーザが、新規ユーザのネットワーク成長に大きな影響力を持つことが明らかとなった。

しかしながら、各聞き役ユーザの特徴量に着目した実験では、ツイート拡散力の効果を十分に検証できていないので、ツイート拡散力の定義および実験方法については今後の課題である。

謝辞

本研究は JSPS 科研費 JP16H02904 ならびに NII 戦略研究公募型共同研究の助成を受けたものです。ここに記して謝意を示します。

［参考文献］

[1] http://uk.businessinsider.com/twitter-user-growth-is-worse-than-you-think-2015-2
[2] 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, 2010.
[3] 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), pp. 10-17, 2011.
[4] 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, 2013.
[5] 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, 2007.
[6] 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.
[7] Yutaro Yamaguchi, Shuhei Yamamoto, and Tetsuji Satoh. Behavior analysis methods for twitter users based on transitions in posting activities. International Journal of Web Information Systems, Vol. 10, No. 4, pp. 363-377, 2014.
[8] 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, 2010.
[9] 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, 2010.
[10] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999.
[11] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. J. Mach. Learn. Res., Vol. 3, pp. 993-1022, 2003.
[12] 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, 2011.
[13] 山口裕太郎, 水沼友宏, 山本修平, 島田諭, 池内淳, 佐藤哲司. マイクロブログにおける投稿活動に着目したユーザプロファイリング. 情報社会学会, 第 5 回知識共有コミュニティワークショップ論文集, pp.1-10, 2012.

(2017 年 7 月 1 日受理)