

# ネットワーク分析による Twitter ユーザの フォロー形成に関する一考察

小出 明弘<sup>1,a)</sup> 齊藤 和巳<sup>1</sup> 風間 一洋<sup>2</sup> 鳥海 不二夫<sup>3</sup>

受付日 2012年11月3日, 再受付日 2013年1月25日,  
採録日 2013年3月27日

**概要:** 本稿では, Twitter のフォローネットワークを分析することにより, ユーザのフォローがどのような目的で行われているのか議論する. まず, フォローネットワークの特徴を把握するため, ネットワーク内の高次数ノードに着目し, ブログの読者関係とレビューサイトのお気に入り関係を表したそれぞれのネットワーク構造の特徴と比較する. その結果, ブログやレビューサイトでは, 比較的小規模な高コリンクグループが得られたのに対し, フォローネットワークでは, 強い双方向関係により構築された大規模な高コリンクグループと, 双方向関係がほとんど見られない複数の小規模な低コリンクグループが存在することが分かった. さらに, 高次数ノードのツイート集合を分析し, これらのグループは同じようなツイートをしているにもかかわらず, フォロワとの関係に大きな違いが見られることが分かった.

**キーワード:** ソーシャルネットワーク分析, Twitter, 次数, デンドログラム, コリンク

## Study of Twitter's Follow Mechanism Based on Network Analysis

AKIHIRO KOIDE<sup>1,a)</sup> KAZUMI SAITO<sup>1</sup> KAZUHIRO KAZAMA<sup>2</sup> FUJIO TORIUMI<sup>3</sup>

Received: November 3, 2012, Revised: January 25, 2013,  
Accepted: March 27, 2013

**Abstract:** In this paper, we explored Twitter's follow mechanism through a network analysis. In order to characterize the salient structure of Twitter's follow network, we first empirically compared it with those of reader and favorite networks from blog and review sites by focusing on their high degree nodes. From this experiment, we observed a relatively large high co-link group whose nodes are mutually connected to each other and some small low co-link groups whose nodes are not mutually connected to each other in Twitter's network. On the other hand, most groups are relatively small high co-link groups such as discussion groups in blog and review sites. Moreover, by analyzing messages tweeted by these group's users, we found that these groups much differ in relation with followers although these groups resemble in content of tweets.

**Keywords:** social network analysis, Twitter, degree, dendrogram, co-link

### 1. はじめに

マイクロブログサービスの1つである Twitter<sup>\*1</sup>は, 急

速にユーザ数を伸ばしており, 著名人や企業の宣伝活動や災害時の情報伝達の手段として利用されるなど, 新たなメディアとして注目されている [1].

Twitter の特徴として, Facebook<sup>\*2</sup>や mixi<sup>\*3</sup>などの代表的な SNS では, 相手からの承認がなければ投稿を閲覧できないが, Twitter では, 自分の興味のある話題を提供するユーザをフォローするだけで, フォローしたユーザのツイートを閲覧できる. さらに, ユーザはほかのユーザのツイートをリツイートと呼ばれる機能でフォロワへ伝達した

<sup>1</sup> 静岡県立大学大学院経営情報イノベーション研究科  
Graduate School of Management and Information of Innovation, University of Shizuoka, Shizuoka 422-8526, Japan

<sup>2</sup> 和歌山大学システム工学部  
Faculty of Systems Engineering, Wakayama University, Wakayama 640-8441, Japan

<sup>3</sup> 東京大学大学院工学系研究科  
Graduate School of Engineering, the University of Tokyo, Bunkyo, Tokyo 113-8654, Japan

<sup>a)</sup> a.koide1988@gmail.com

<sup>\*1</sup> <http://twitter.com>

<sup>\*2</sup> <http://www.facebook.com/>

<sup>\*3</sup> <http://mixi.jp/>

り、'@" の後ろにユーザ名を明記してツイートすることで、そのユーザを Mention したりすることができる。これらの機能から、Twitter は情報の伝達から対話まで多岐にわたって利用されている。

ユーザがフォロー関係を形成する理由に関する先駆けの研究として、Kwak ら [10] は、ユーザのフォロー関係をネットワークとして表現し、フォロー分布はべき則に従わない、短い有効直径、ユーザ間の低い相互関係など、SNS などのソーシャルネットワークとは異なる構造を有していることを明らかにしている。

本研究でもフォローネットワークを対象とし、フォロー関係の形成について考察する。特に、ネットワーク内で中心的な役割を果たす高次数ユーザに着目して、高次数ユーザ間のフォロー、被フォローの類似関係における特徴を、フォローネットワークと同様にユーザの関係が読み手と書き手によって表されるブログやレビューサイトのネットワークと比較することにより分析する。その際、階層的クラスタリングで用いられるデンドログラムを利用して可視化することにより、高次数ユーザ群の類似関係をマクロな視点で明らかにする。さらに、分析結果を基に高次数ユーザとそれを取り巻くフォロワとの関係を、Mention による対話、リツイートによる情報伝達の2つのツイート行動に基づいて分析する手法を提案し、高次数ユーザの特徴を詳細に分析する。

このような、ソーシャルメディア上の人間の行動によって構築されるネットワークの構造を明らかにすることは、ネットワーク成長の将来予測や、ネットワーク上の情報拡散現象と密接に関わっており、バイラルマーケティングや、デマ情報の抑制への貢献が期待されている [2], [3]。

本稿では、2章で関連研究について述べ、3章で使用したデータセットならびに、基本的な統計量について述べ、さらに次数と様々な指標との関係について述べる。4章では、ユーザ間のフォロー、被フォローの類似性を可視化し、ネットワーク間の構造の違いについて述べる。5章では、特に Twitter に焦点をあて、ユーザのフォロー形成についてより詳細に分析し、6章で本稿を結ぶ。

## 2. 関連研究

ソーシャルネットワークの分析は近年さかんに行われており、SNS やブログ、マイクロブログなど多岐にわたる。SNS 内に構築されるネットワーク構造の分析として、Adamic らは大学内の比較的小規模な SNS のデータを用いて、ユーザの属性とネットワーク構造を組み合わせた分析を行っている [4]。内田らは、ネットワーク指標の観点から、SNS のネットワーク構造は代表的な人工ネットワークモデルでは正確に表現できないことを示した [5]。日本国内の代表的な SNS サービスである mixi に関する研究としては、主にクラスタリングの観点から分析した湯田らの研

究 [6] やコミュニティの観点から分析した松尾らの研究 [7] がある。また、鳥海らは、大量の小規模 SNS を対象にしたネットワーク分析を行い、大規模 SNS である mixi と比較している [8]。ブログを対象にした分析として、風間らは、Ameba ブログの読者ネットワークと mixi ネットワークの構造を比較している [9]。マイクロブログを対象に分析した研究として、Kwak らは、フォローネットワークやトレンドトピックなどを対象に調査し、既存のソーシャルネットワークとは異なる構造を有していることを発見した [10]。Huberman らは、ある期間内に2回以上リプライしたユーザで作成した友人ネットワークは、フォローネットワークのごく一部であったと報告している [11]。Cha らは、フォローネットワーク内の中心的なユーザといくつかのトピックに関する影響力を分析している [12]。

先行研究における Twitter のフォローネットワーク分析では、ネットワーク全体の特徴に着目し、基本的なネットワーク指標を利用して分析していた。本稿では、ネットワーク内で中心的な役割を果たす高次数ユーザに着目して、高次数ユーザ間のフォロー、被フォローの類似関係を分析する。さらに、分析結果を基に高次数ユーザとそれを取り巻くフォロワの間の情報のやりとりに着目して、高次数ユーザの特徴を詳細に分析する。

## 3. 使用データとネットワークの性質

本章では、本分析で使用するデータの詳細と、ネットワークの基本的な統計量を示す。また、次数と各指標との関係について述べる。

今後の共通の設定として、ネットワーク  $G = (V, E)$  を定義し、ノード集合を  $V = \{v_1, v_2, \dots, v_N\}$ 、リンク集合  $E$  を  $V \times V$  の部分集合とする。なお、 $N$  はネットワークの全ノード数を表す。任意のノード  $v_i \in V$  に対し、 $v_i$  からのリンクを有するノード集合 (子ノード集合) を  $A(v_i) = \{v_j; (v_i, v_j) \in E\}$  とし、 $v_i$  へのリンクを有するノード集合 (親ノード集合) を  $B(v_i) = \{v_k; (v_k, v_i) \in E\}$  とする。

### 3.1 データセット

#### 3.1.1 Twitter のフォローネットワーク

1つ目は、Twitter 検索で収集された日本語で書かれたツイートを公開しているユーザのフォローネットワークである。ここで、ユーザ X からユーザ Y へのリンクは、ユーザ X がユーザ Y をフォローしたときに生成される。本稿では、2011年1月31日時点において累積200回以上ツイートしているユーザのフォロー関係から、ユーザをノード、フォロー関係をリンクとした有向ネットワークを作成した。ネットワークのノード数は1,079,986ノード、リンク数は157,371,341本である (以下 TwitterNW)。

さらに、発言内容の判定に用いたツイートデータは、2011

年3月7日 0:00:00 から11日 12:00:00 までの期間に、ネットワーク内に存在する 1,079,986 ノードのツイートを収集したもので、この期間内の総ツイートは 57,625,889 ツイートであった。

### 3.1.2 ブログサイトの読者ネットワーク

2つ目は、日本のブログサービスサイト“Ameba”<sup>\*4</sup>の読者ネットワークである [9] (以下, amebloNW)。“Ameba”の各ブログには読者のブログへのハイパーリンクが張ってある読者リストがある。ここで、ブログ X からブログ Y への読者リンクは、ブログ X がブログ Y をお気に入りブログとして登録したときに生成される。このネットワークのノード数は 56,604 ノード、リンク数は 1,071,080 本である。

### 3.1.3 レビューサイトのお気に入りネットワーク

3つ目は、化粧品の口コミサイト“@cosme”<sup>\*5</sup>から 2009 年 12 月に取得したお気に入りネットワークである (以下, cosmeNW)。サイトに登録しているあるユーザ X がユーザ Y をお気に入り登録すると、ユーザ X からユーザ Y に対してリンクが張られる。このネットワークはランダムに選択したユーザから 10 段までたどって収集したもので、ノード数は 45,024 ノードでリンク数は 546,930 本である。

## 3.2 基本統計量の分析

表 1 に各ネットワークの基本統計量を示す。ここで、リンク密度  $L$  は以下の式で定義される。

$$L = \frac{|E|}{N(N-1)}. \quad (1)$$

また、ネットワーク全体に対する相互関係を表す双方向リンク率  $cr$  は以下の式で定義される [14]。

$$cr = \frac{1}{|E|} \sum_{v_i \in V} |A(v_i) \cap B(v_i)|. \quad (2)$$

なお、この指標は mixi ネットワークのように、リンクがすべて双方向となるネットワークでは 1 となる。

各ネットワークに共通する特徴として、最大出次数に比べ最大入次数が高く、入・出次数には強い正の相関がある。さらに、リンク密度は TwitterNW を筆頭に非常に低く、疎なネットワークである。双方向リンク率は、TwitterNW が最も高く、全体の約 6 割である。amebloNW もほぼ変わらないが、cosmeNW はやや低い値となる。

## 3.3 度数に着目した分析

本節では度数に着目し、入次数と出次数、度数とネットワーク指標の関係を分析する。度数とネットワーク指標の関係を明らかにするために、上位  $k\%$  のノード集合を対象に、 $k$  の値を増やしたときの各指標の変化を分析する。なお、異なるネットワークを比較できるように  $k$  は割合とした。

<sup>\*4</sup> <http://ameblo.jp/>

<sup>\*5</sup> <http://www.cosme.net/>

表 1 各ネットワークの基本統計量

Table 1 Basic Statistics of each network.

	TwitterNW	amebloNW	cosmeNW
平均次数	145.7	13.0	7.80
最大出次数	97,938	3,607	881
最大入次数	175,616	6,291	1,292
入出次数相関	0.74	0.61	0.52
リンク密度	0.000270	0.000347	0.000459
双方向リンク率	0.60	0.54	0.44

### 3.3.1 出次数と入次数の関係

入次数と出次数の関係を、より詳細に分析する。出次数上位  $k\%$  以内のノード集合を  $R^+(k)$ 、入次数上位  $k\%$  以内のノード集合を  $R^-(k)$  とする。ここで、 $R^+(k)$  と  $R^-(k)$  との類似度を求めることにより、度数間関係を明らかにする。類似度は、以下に示す Jaccard 係数を用いる。

$$J(R^+(k), R^-(k)) = \frac{|R^+(k) \cap R^-(k)|}{|R^+(k) \cup R^-(k)|}. \quad (3)$$

本稿では、 $k = 1, 2, \dots, 100$  と設定する。 $k = 100$  となる時、類似度は 1 となる。

出次数集合と入次数集合との Jaccard 係数を図 1 に示す。横軸は対象となる出次数のランキング、縦軸は出次数上位  $k\%$  のノード集合と、入次数上位  $k\%$  のノード集合との Jaccard 係数を表している。また、入次数と出次数の関係がネットワーク固有の性質であることを明らかにするために、順位をランダムに入れ替えた場合とも比較する。

まず、ネットワークごとにランダムデータと比較すると、各ネットワークともにランダムデータより値が高いことから、どのネットワークに関しても出次数上位のノードは入次数においても上位である傾向が強い。また、各ネットワークでの類似度の増減を比較すると、TwitterNW と amebloNW は上位 1% の段階から類似度が高く、その後やや横ばいになりながら緩やかに 1 に近づく。一方、cosmeNW は 2 つのネットワークと比較するとやや低い類似度から右肩上がりに増加している。これらの結果から、各ネットワークで差はあるものの、出・入次数の高いノードは入・出次数も高い傾向にあることが分かる。

### 3.3.2 度数とランキング指標

度数は、ネットワーク中の隣接構造に基づいてランキングする最も基本的な指標である。そのほかに、ネットワーク内でのノードの影響力を測る手法として、PageRank アルゴリズムや HITS アルゴリズムなどがある。本項では、度数と各ランキング指標との関係について述べる。

PageRank は、Web ページの重要度を測るアルゴリズムである [16]。重要な Web ページからリンクされている Web ページは重要な Web ページであると仮定し、そのような Web ページには高い PageRank スコアを与える。

HITS アルゴリズムは、あるトピックに関するオーソリティと、関連オーソリティにリンクしているハブを求



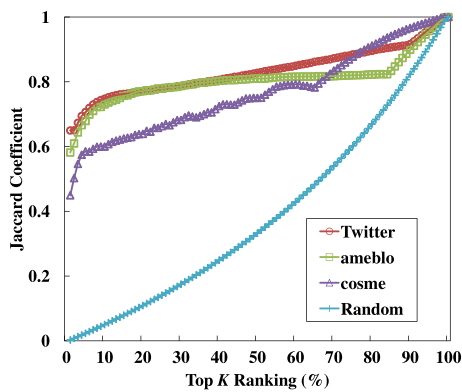


図 1 出次数集合と入次数集合の Jaccard 係数

Fig. 1 Jaccard coefficient between set of outdegree nodes and indegree nodes.

めるアルゴリズムである [17]. 各ノードの Hub 度 (望ましい Authority ノードにリンクを張っている度合い) と, Authority 度 (望ましい Hub ノードからリンクを張られている度合い) を反復計算により算出する.

前項で用いた出次数と入次数のノードランキングに加え, 各ノードの PageRank スコア, Hub 度, Authority 度のランキングを求める. 続いて, 次数とランキング指標との類似関係を分析するため, 前項で用いた Jaccard 係数を利用し, 高入・出次数ノードから見た各ランキング指標との関係を求める. なお, PageRank における大域ジャンプ確率は  $\alpha = 0.15$  とした.

TwitterNW での, 次数とランキング指標の Jaccard 係数を図 2 に示す. ここで, OutD は出次数, InD は入次数, OutPR は出 PageRank, InPR は入 PageRank, Hub は Hub 度, そして Authority は Authority 度をそれぞれ表している. また, 各ランキング指標は, 入・出次数と類似傾向が高い方のみをプロットした. なお, TwitterNW の分析結果だけを示しているが, どのネットワークでの分析でもほぼ同様の傾向が見られる.

次数と各ランキング指標は, 特に上位 1% のノード集合で類似傾向が高いランキング指標が多く, その後やや減少し, 最終的に緩やかに 1 に近づいていく. したがって, 特に次数が高いノードはランキング指標とも強い関係がある. また, ランダムデータと比較すると, 次数の上位のノード集合と各ランキングの上位ノード集合は類似する傾向が強い.

### 3.4 次数とコリンク構造

本節では, ネットワーク内の各ノードの相互関係を分析するため, ネットワーク全体ではなく個々のノードに対してコリンク率を求める [15].

ノード  $v_i$  に対し, 入次数に対する双方向リンク数により得られる入コリンク率  $IC(v_i)$ , 出次数に対する双方向リンク数により得られる出コリンク率  $OC(v_i)$  を以下のように

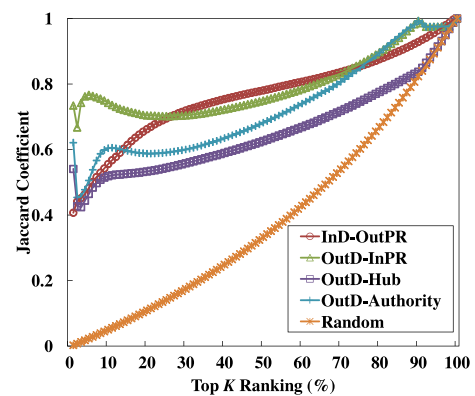


図 2 TwitterNW の次数とランキング指標の Jaccard 係数

Fig. 2 Jaccard coefficient between degree and ranking index on TwitterNW.

定義する.

$$IC(v_i) = \frac{|A(v_i) \cap B(v_i)|}{|A(v_i)|} \quad (4)$$

$$OC(v_i) = \frac{|A(v_i) \cap B(v_i)|}{|B(v_i)|} \quad (5)$$

なお,  $|A(v_i)|, |B(v_i)|$  が 0 になるとき, 本数式は定義されない. さらに, 式 (4) と (5) の調和平均をとることにより, コリンク率として以下の評価式を得る.

$$C(v_i) = \frac{2}{\frac{1}{IC(v_i)} + \frac{1}{OC(v_i)}} = \frac{2 \times |A(v_i)| \cap |B(v_i)|}{|A(v_i)| \cup |B(v_i)|} \quad (6)$$

コリンク率は, 次数に対する相互結合を示す指標で, 各ノードの親ノード集合と子ノード集合が一致するときに 1 となり, まったく一致しないときに 0 となる.

各ネットワークの出次数の順位によるコリンク率の変化を図 3 に, 入次数の順位によるコリンク率の変化を図 4 にそれぞれ示す. 横軸は評価対象となる入次数上位のノードの割合, 縦軸は対象となるノード集合のコリンク率を平均した平均コリンク率を表す. また, 比較のために, 各ネットワークの入次数と出次数を変えずに, ランダムにリンクを張り直した場合の値もプロットした [13].

図 3 では, 特に TwitterNW において次数上位のノードの平均コリンク率がきわめて高いことが分かる. また, 最終的な平均コリンク率も, TwitterNW がその他のネットワークと比較してやや高い. また, 各ネットワークをランダムに張り直した場合と比較すると, 各ネットワークともに, ランダムに張り直した場合より平均コリンク率が明らかに高い. 特に, Twitter ネットワークには, 高次数の場合に双方向性を高める何らかのメカニズムがあると考えられる.

一方図 4 では, TwitterNW の平均コリンク率は高次数ノードでも他のネットワークに比べ高くはなく, むしろ amebloNW の方が高い. また, 平均コリンク率はなだらかに減少する. これらの結果から, 共通の特徴として高次数ノードは平均コリンク率が高くなる傾向があることが分かる.

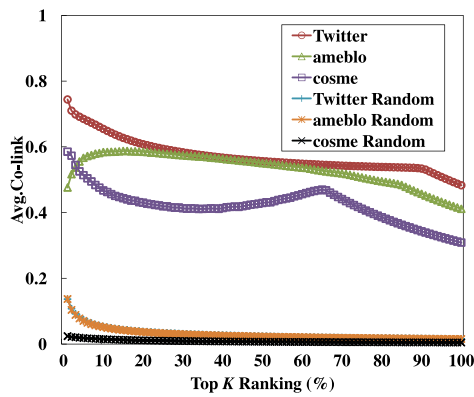


図 3 出次数の順位によるコリンク率の変化

Fig. 3 Change of co-link ratio focus on outdegree ranking.

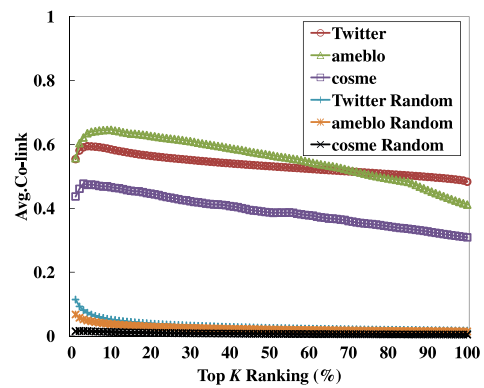


図 4 入次数の順位によるコリンク率の変化

Fig. 4 Change of co-link ratio focus on indegree ranking.

#### 4. 高次数ノード間のリンク集合の類似性

前章では、入・出次数の関係や、次数とランキング指標・コリンク率の高さとの関係について分析した。本章では各ネットワークの高次数部分に限定して、高出次数ノード間の子ノード集合、高入次数ノード間の親ノード集合の類似度をそれぞれ算出し、ノード間の類似関係を最小全域木を用いてデンドログラム可視化する。

##### 4.1 類似度の定義と可視化法

ネットワーク上の高次数ノード間の類似関係を分析するために、親・子ノード集合を用いて、2種類のノード間類似度を定義する。

あるノード  $v_i$  に対し、 $v_j \in A(v_i)$  であるときに1、そうでないときに0とし、ノード  $v_i$  の出次数ベクトル  $\mathbf{a}_i$  を定義すると、出次数に対する隣接行列  $\mathbf{A}$  は以下のように定義される。

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_N \end{bmatrix}^T \quad (7)$$

同様に、入次数に対する隣接行列  $\mathbf{B}$  は以下のように定義される。

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_N \end{bmatrix}^T \quad (8)$$

次に、隣接行列を利用し類似度を算出する。本稿では、類似度としてコサイン類似度を採用した。任意のノード  $v_i$  と  $v_j$  に対するコサイン類似度は、以下のように定義される。

$$\rho(v_i, v_j) = \frac{\sum_{n=1}^N a_{i,n} \cdot a_{j,n}}{\sqrt{\sum_{n=1}^N a_{i,n}^2} \cdot \sqrt{\sum_{n=1}^N a_{j,n}^2}} \quad (9)$$

$$\rho(v_i, v_j) = \frac{\sum_{n=1}^N b_{i,n} \cdot b_{j,n}}{\sqrt{\sum_{n=1}^N b_{i,n}^2} \cdot \sqrt{\sum_{n=1}^N b_{j,n}^2}} \quad (10)$$

上述の処理により、ノード間の類似度を算出できる。なお、コサイン類似度を利用する理由は、多重のリンクを許容するようなネットワークに対し本手法を容易に拡張する

ことができるからである。

さらに、ノード間の類似度を以下の式に基づいて距離に変換し、 $N \times N$  の距離行列  $\mathbf{D}$  を作成する。

$$d(v_i, v_j) = \sqrt{(1 - \rho(v_i, v_j))} \quad (11)$$

そして、ノード間に距離行列に基づいた重み付きエッジを張ることにより、重み付き完全グラフ  $G_C = (V, E)$  を構築し、このグラフから最小全域木  $T$  を作成する。木の作成には以下に示すクラスカル法 [18] を用いた。

- (i) 初期化：グラフ  $G_C$  のリンク集合  $S, T = \emptyset$ , クラスタ  $K$  を定義する。各ノードは単一のクラスタに所属する。
- (ii) 選択：集合  $S$  から距離が最小のリンク  $(v_i, v_j)$  を選択し、 $S = S \cup (v_i, v_j)$  とする。
- (iii) 比較： $v_j \notin K(v_i)$  であれば、 $T$  にリンク  $(v_i, v_j)$  を追加し、 $K(v_i) = K(v_i) \cup K(v_j)$  とする。
- (iv) 反復： $S = \emptyset$  となるまで (ii), (iii) を繰り返す。

この最小全域木  $T$  は、ノード全体の構造を簡潔に表現したものと考えることができる。ここで、作成した木を可視化するため、デンドログラムを利用する。デンドログラムは、基本的なクラスタリング手法の1つである階層的クラスタリングで用いられる。階層的クラスタリングは、1個の要素からなる  $N$  個のクラスタを初期状態とし、最も距離の近いクラスタ間を併合することを1つのクラスタになるまで繰り返す。そして完成したデンドログラムを適当な閾値で階層を分割することでクラスタを求める。最小全域木を利用しデンドログラムを作成することは、階層的クラスタリングにおいてクラスタ間を最短距離法を用いて併合することと同様である。

##### 4.2 可視化結果

各ネットワークの高次数ノードの類似関係をデンドログラムで可視化した結果を図 5, 図 6 に示す。

縦軸はクラスタが併合された際の距離を表す。本分析では、2つのクラスタが併合される際にクラスタの要素数が大きい方を左に配置するように可視化しているため、横方

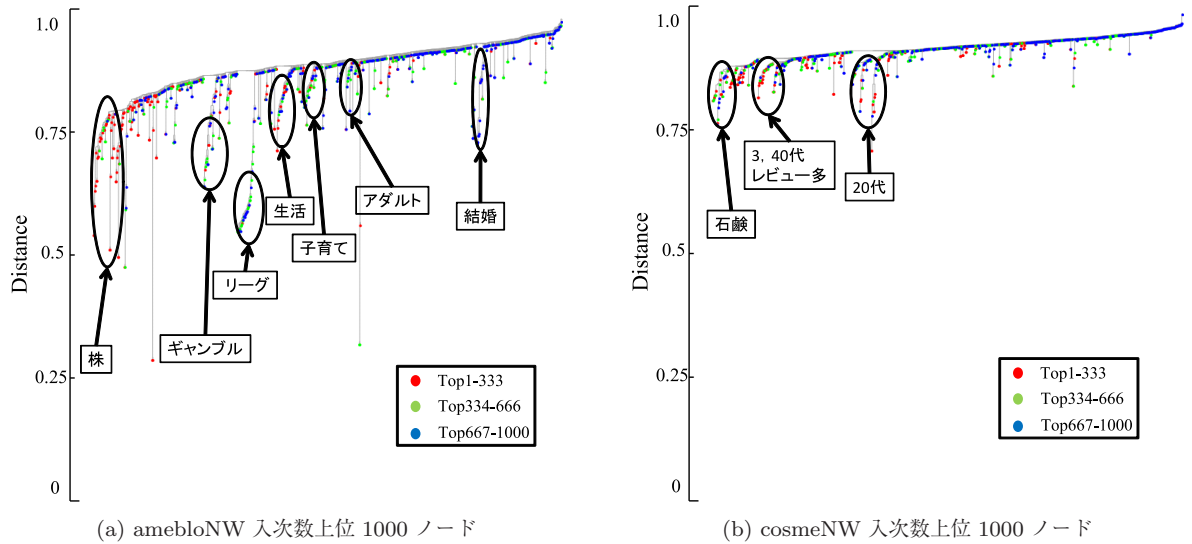


図 5 amebloNW・cosmeNW 上位 1000 ノードのデンドログラム  
 Fig. 5 Dendrogram of top1000 nodes on amebloNW and cosmeNW.

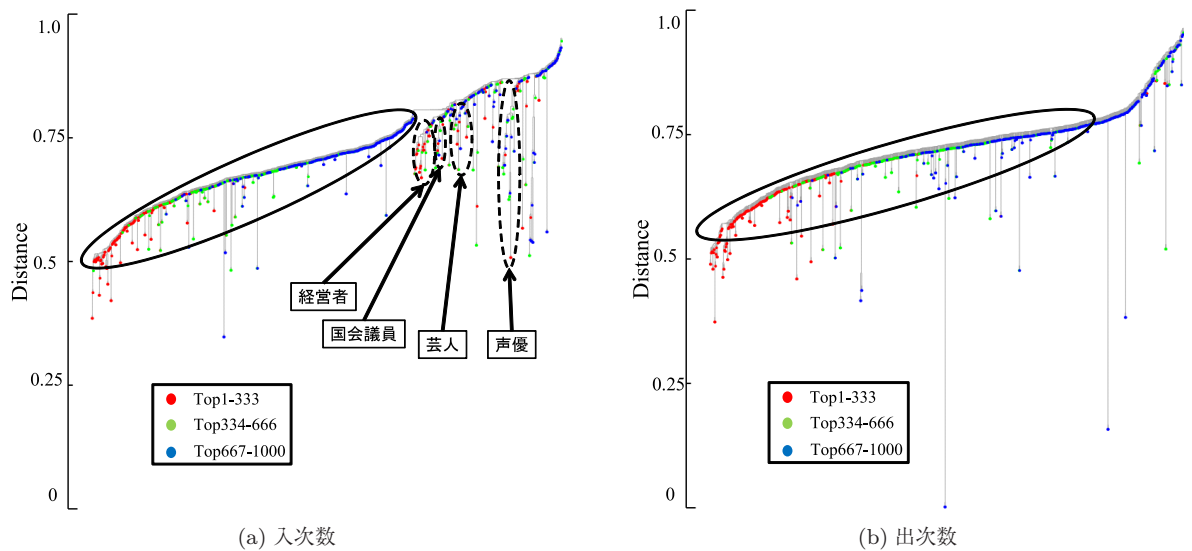


図 6 TwitterNW 上位 1000 ノードのデンドログラム  
 Fig. 6 Dendrogram of top1000 nodes on TwitterNW.

向にはクラスタが併合される順序を表している。なお、本分析では次数上位 1000 ノードを対象に可視化した。また、amebloNW, cosmeNW に関しては、入次数, 出次数ともにほぼ同様の結果が得られたため、出次数の場合の可視化結果を省略した。各ノードは次数上位 1-333 のノードを赤, 上位 334-666 ノードを緑, 上位 667-1000 ノードを青で配色している。

まず、図 5(a) の amebloNW と (b) の cosmeNW には、小規模なノード集合からなる垂れ下がった房のような構造がいくつか見られる。ただし、真下に伸びた線の先にプロットされたノードは、類似した 2 つのノードを表しているが、このような場合は同じユーザが作った別のアカウントであることが多く、本稿では無視する。この小規模ノード集合はほぼ同じ親ノード集合からリンクを張られているノード集合である。また、先ほども述べたように出次数

でもほぼ同様の結果が得られていることから、ほぼ同じ子ノード集合にリンクを張っており、強い双方向性があることが分かる。本稿では、このような双方向性が強いノード集合を高リンクグループ（実線で囲まれたノード群）と呼ぶ。高リンクグループは、amebloNW ではサッカー関係やギャンブル関係, cosmeNW では同じ年齢層や商品のタイプなどの、同じ興味や関心を持つノードによって構成されていた。

図 6 (a) の Twitter の入次数上位 1000 ノードの可視化結果では、図の右側に図 5 でも得られた小規模ノード集合からなる房のような構造がいくつか見られるが、ノード間距離に近い傾向がある図の左側に 1 つの大規模なノード集合が見られる。図 6 (b) では、小規模ノード集合が見られないが、全体的にノード間距離に近い傾向がある。

また、作成した最小全域木のノード間の距離を図 7 に示

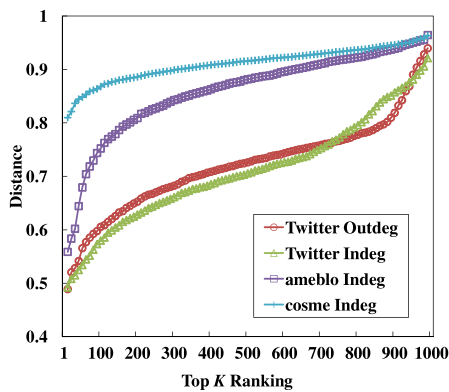


図 7 最小全域木のノード間距離

Fig. 7 Distance between nodes on minimum spanning trees.

す。横軸は、ノード間の距離が短いものから並べたときのランキング、縦軸にそのノード間の距離をプロットした。この結果からも、TwitterNW はその他のネットワークと比較してノード間の距離が全体的に近いことが分かる。

TwitterNW のノード集合について調査すると、図 6(a) の大規模なノード集合は高リンクグループであり、ほとんどが図 6(b) の大規模なノード集合に含まれていた。また、図 6(a) で得られた小規模ノード集合に属するノードは、高出次数上位 1000 ノードには現れない。つまり、ほとんどが 1 方向のリンクである。本稿では、このようなノード集合を低リンクグループ（破線で囲まれたノード群）と呼ぶ。

低リンクグループに属するノードを調査すると、大企業の経営者や国会議員、芸能人など知名度の高いユーザであった。高リンクグループに属するノードを調査すると、中小企業診断士、オークション運営者、旅館の公式アカウントなど多種多様で、低リンクグループのような意味的関連性は見られなかった。したがって、このグループはトピックではなく、他の何らかの特徴を共有していると推測される。

## 5. Twitter における高次数ノードの特性

前章までの分析により、TwitterNW 固有の構造として単一の大規模な高リンクグループと複数の小規模な低リンクグループが検出された。本章では、フォロネットワークにおいて、これらの構造が持つ特徴がどのように作られたかを明らかにする。まず、高次数ノードの詳細なリンク構造を分析する。さらに、静的なフォロネットワークに対し、動的な性質を表すものとしてツイート集合を利用し、ネットワーク構造とツイート行動の関係について分析する。

### 5.1 リンク構造

まず、高リンクグループと低リンクグループのリンク構造を明らかにするために、高入次数上位 1000 ノード

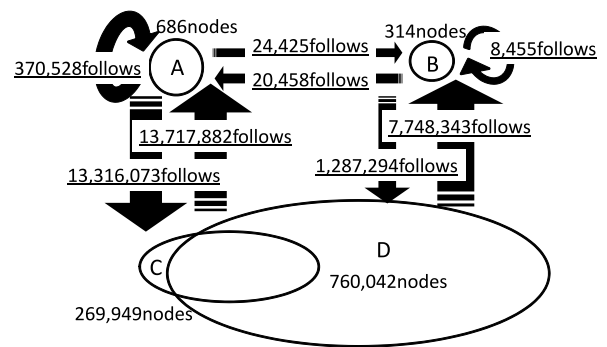


図 8 入次数上位 1000 ノードのリンク構造

Fig. 8 Link structure of top1000 indegree nodes.

集合とその親ノード集合を、次の 4 グループ A, B, C, D に分割する。なお C と D は重なりを持つことに注意する。

- (i) グループ A: 高リンクグループに属するノード集合
- (ii) グループ B: 複数の低リンクグループに属しているノード集合
- (iii) グループ C: グループ A の親ノードかつ、グループ A, B に属さないノード集合
- (iv) グループ D: グループ B の親ノードかつ、グループ A, B に属さないノード集合

グループ A, B, C, D に属するノード集合は、TwitterNW 全体の約 75% であった。

図 8 に、グループ内・グループ間のリンク構造と親ノード集合の重なりを示す。グループ A のノード数は 686 ノードで、グループ B の 314 ノードの約 2 倍だが、その親ノードであるグループ C のノード数はグループ D の約 35% であった。したがって、グループ B は非常に多様なノードからリンクが張られていた。さらに、グループ C のノードの約 90% は、グループ D に含まれていた。

リンク数を見ると、グループ A とグループ C 間のリンク数は、ノード数が大きく異なるにもかかわらずほとんど同じで、双方向リンク率も 0.75 と高かった。さらに、グループ A 内の大部分のノード間と双方向接続しており、双方向リンク率は 0.90 であった。これに対し、グループ B とグループ D 間はノード数、リンク数ともに大きな差があり、双方向リンク率も 0.21 と低かった。また、グループ B 内のノード間のリンク数は非常に少ないが、双方向リンク率は 0.46 とグループ B とグループ D 間に比べるとやや高かった。

### 5.2 URL 付きツイート

高リンクグループと低リンクグループのような違いが生じる 1 つの仮説として、これらのグループのツイート行動自体のパターンが異なっていることが考えられる。そこで、本節では高次数ノードのツイートに含まれる URL について分析する。

情報の出典を示すために付与される URL は、情報の信



頼性やツイートのタイプを分類するための代表的な特徴量の1つであり、各グループの特徴を分析するうえでも重要であると考えられる [19]. 分析では、全ノード、グループ A、グループ B を対象に、グループごとにどの程度 URL が付与されているのかを分析した.

平均の URL 付きツイートの割合を見ると、グループ A が最も高く全ツイートの 37% に URL が付与されていた. また、グループ B の URL 付きツイートの割合は 34% であり、グループ A とそれほど変わらなかった. 一方、全ノードを対象とした場合には 12% であることから、グループ A、B の URL 付きツイートの割合は、全体の約 3 倍と非常に高かった.

グループ A、B に属するノード集合の URL をランダムサンプリングして、その内容を調べると、グループ A では商品紹介や、時事ニュース、そのノードが運営する Web サイトやブログへのリンクなどが多かった. グループ B でも、グループ A とほぼ同様のタイプの URL が多く見られた.

すなわち、高コリンクグループと低コリンクグループは、他の情報の紹介というツイート行動やその情報の種類に差はなく、片方のグループでスパムや宣伝などの行為が顕著に行われているなどの違いによるものでもないことが分かる.

### 5.3 対話構造

本節では、高次数ノードを対話の観点から分析する. Twitter では、ユーザは Mention 機能を使って親しいユーザと繰り返しメッセージのやりとりをしている. そこで、Mention ツイートを対話と見なし、グループ内・グループ間の Mention 機能による対話数を分析した. Mention 数が多い箇所は、関心事が共通だったり、親交が深かったりするなどの理由が考えられる. なお、片方向のやりとりだけでは対話が成立しているとは考えにくいので、Huberman らの研究 [11] のように、1 回以上相互にやりとりした場合を対象に、グループ A、B、C、D に含まれるノード群の対話構造を作成した.

図 9 に、抽出した対話構造を示す. まず、図 8 の結果と異なり、グループ A とグループ B の結び付きが弱かった. また、ほとんどがグループ C、D からグループ A、B に対する対話である一方で、グループ A からグループ C への対話数は、グループ B からグループ D への対話数の約 2 倍であった. またグループ C のノード数は D の 36% にすぎないが、対話に参加したノード数は D の 58% と高くなった. さらに、グループ A、グループ B それぞれの内部でも対話は見られなかった. つまり、他と比べて A、C 間で特に活発に対話していると推測される.

### 5.4 情報伝達構造

次に、グループ内・グループ間のリツイート機能による

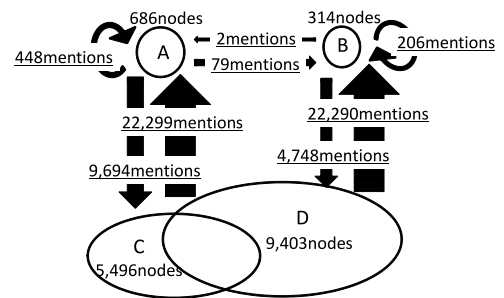


図 9 入次数上位 1000 ノードの対話構造  
Fig. 9 Conversation structure of top1000 indegree nodes.

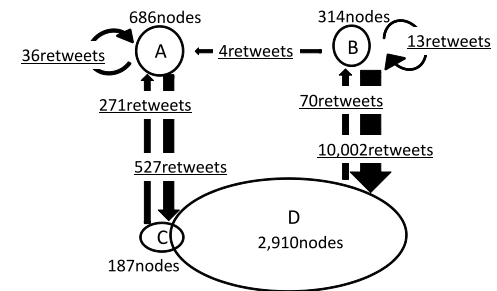


図 10 入次数上位 1000 ノードの情報伝達構造  
Fig. 10 Information transmission structure of top1000 indegree nodes.

情報伝達を分析する. Twitter では、ユーザはリツイート機能を使ってフォローに有用なツイートや多くのユーザに伝えたい情報を伝達しており、リツイートされるツイートが多いほど、発言の内容が高く評価されているとともに、影響力が強いと考えられる. そこで、リツイートを情報伝達と見なし、グループ内・グループ間の情報伝達構造を分析する.

なお、収集期間内で 1 度だけ大量にリツイートされたツイートをしたユーザが存在する可能性がある. このようなユーザは普段から影響力があるとは考えにくいので、本稿では期間内で 2 回以上リツイートされた場合を対象に、グループ A、B、C、D に含まれるノード群の情報伝達構造を抽出した.

図 10 に抽出した情報伝達構造を示す. 情報を伝達する際は、フォローしているユーザのツイートを受け取り、そのツイートをフォローへ伝達することから、フォロー関係とは逆向きのリンクとなる. グループ A から B でほとんどの情報伝達が行われており、他では非常に少なかった. さらに、グループ D で情報伝達するノード数が非常に多く、グループ C で情報伝達するノード数は D の場合の約 6% にすぎない. また、グループ B はグループ D からの情報伝達はほとんどなく、一方的であった.

### 5.5 考察

これまでの分析結果を考慮し、フォローネットワークで単一の高コリンクグループ (グループ A)、と複数の低コ



リンクグループ (グループ B) が得られた背景について考察する。

まず、低コリンクグループは、トピックごとに複数のグループに分割されており、加えて実世界での著名人が多く含まれていた。Twitter では、ユーザは著名人を一方的にフォローして、著名人自身が発言する近況や最新の情報を得ることができることから、知名度の高い著名人になるほど多くのユーザからフォローされやすく、またリツイートによる情報伝達も顕著である。ただし、著名人は Twitter をフォローへの一方的な情報発信のためだけに利用することが多く、フォローとの対話はあまり行わない傾向がある。これは、フォローの双方向性の低さ、フォローとの対話数の少なさで実証されている。このようなツイート内容の制限と交流の少なさが、低コリンクグループをトピックごとに分割している原因だと考えられる。

次に、高コリンクグループは、全体を通じて共通したトピックは見られない巨大な集合であるとともに、多種多様なユーザによって構築されていた。低コリンクグループとは異なり、これらのユーザは実世界における著名人ではないことがほとんどであったが、その被フォロー数の高さから、少なくとも Twitter 上では有名人であると推測できる。ただし、リツイートによる情報伝達はほとんど行われていないことから、低コリンクグループに比べるとツイートの独自性・重要性はフォローにそれほど評価されていないと考えられる。反面、フォローとの対話は活発であることから、これらのユーザが多くのフォローを獲得できたのは、Twitter 上でのフォロー関係形成のための活動と対話の活発さの結果であると考えられる。このような背景から生まれたグループ内のユーザの結び付きの強さが、最小全域木を用いてデンドログラム可視化した際、全体のトピックが明確でないにもかかわらず巨大なグループを構築した理由であると考えられる。

## 6. おわりに

本稿では、Twitter のフォローネットワークの特徴を把握するため、ネットワーク内の高次数ノードに着目し、ブログとレビューサイトの読者関係とお気に入り関係を表したそれぞれのネットワーク構造の特徴と比較した。その結果、フォローネットワークでは、同じ高次数ノードであっても、強い双方向関係により構築された単一の大規模高コリンクグループと、双方向性がほとんど見られない低コリンクグループが存在することを確認した。さらに、高次数ノードのツイート集合を URL 付きツイート・情報伝達・対話の観点から分析し、これらのグループは、URL 付きツイートの割合はほとんど同じであるが、フォローとの関わり方に大きな違いが見られることが分かった。

今後は、フォローネットワークのダイナミクスを考慮した分析により、ネットワークの生成過程をより詳細に分析

する。また、今回の知見が Twitter 内でのネットワークの成長や、情報の拡散とどのように関係しているのかをより詳細に分析する。さらに、既存のモデルを応用し、Twitter 上でのネットワーク成長や情報拡散をより緻密に表現できるようなモデルを構築し、その有効性を評価する予定である。

謝辞 最後に、本研究を行うにあたりデータ収集にご協力いただいた、Twitter 検索を運営する兼山元太氏に感謝する。また、本研究は NTT との共同研究および、科研費 (23500312) の助成を受けた。

## 参考文献

- [1] 総務省：平成 23 年度情報通信白書 (2011), 入手先 <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h23/pdf/>).
- [2] Leskovec, J., Adamic, L.A. and Huberman, B.A.: The Dynamics of Viral Marketing, *Proc. 7th ACM Conference on Electronic Commerce*, pp.228-237 (2006).
- [3] 白井高士, 榎 剛史, 鳥海不二夫, 篠田孝祐, 風間一洋, 野田五十樹, 沼尾正行, 栗原 聡: Twitter におけるデマツイートの拡散モデルの構築とデマ拡散防止モデルの推定, 第 26 回人工知能学会全国大会 (2012).
- [4] Adamic, L.A. and Adar, E.: Friends and Neighbors on the Web, *Social Networks*, Vol.25, No.3, pp.211-230 (2003).
- [5] 内田 誠, 白山 晋: SNS のネットワーク構造の分析とモデル推定, 情報処理学会論文誌, Vol.47, No.9, pp.2840-2849 (2006).
- [6] 湯田聡夫, 小野直亮, 藤原義久: ソーシャル・ネットワーキング・サービスにおける人的ネットワークの構造, 情報処理学会論文誌, Vol.47, No.3, pp.865-874 (2006).
- [7] 松尾 豊, 安田 雪: SNS における関係形成原理—mixi のデータ分析, 人工知能学会論文誌, Vol.22, No.5, pp.531-541 (2007).
- [8] 鳥海不二夫, 山本仁志, 諏訪博彦, 岡田 勇, 和泉 潔, 橋本康弘: 大量 SNS サイトの比較分析, 人工知能学会論文誌, Vol.25, No.1, pp.78-89 (2010).
- [9] 風間一洋, 斉藤和巳, 山田武士, 松林達史: ブログサービスにおける読者ネットワークの分析, 第 2 回ネットワークが創発する知能研究会研究報告要旨集, pp.72-79 (2006).
- [10] Kwak, H., Lee, C., Park, H. and Moon, S.: What is Twitter, a Social Network or a News Media?, *Proc. 19th International Conference on World Wide Web*, pp.591-600, ACM (2010).
- [11] Huberman, B.A., Romero, D.M. and Wu, F.: Social Networks that Matter: Twitter under the Microscope, *First Monday*, Vol.14, No.1 (2009).
- [12] Cha, M., Haddadi, H., Benevenuto, F. and Gummadi, K.P.: Measuring User Influence in Twitter: The Million Follower Fallacy, *Proc. 4th International Conference on Weblogs and Social Media* (2010).
- [13] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U.: Network Motifs: Simple Building Blocks of Complex Networks, *Science*, Vol.298, No.5594, pp.824-827 (2002).
- [14] Ohara, K., Saito, K., Kimura, M. and Motoda, H.: Effect of In/Out-Degree Correlation on Influence Degree of Two Contrasting Information Diffusion Models, *Proc. 2012 International Conference on Social Computing, Behavioral Modeling, and Prediction*, pp.131-138 (2012).

- [15] 小出明弘, 齊藤和巳, 鳥海不二夫, 風間一洋: コリンク構造に着目した多重グラフの特性分析, 第3回ソーシャルコンピューティングシンポジウム (2012).
- [16] Brin, S. and Page, L.: The Anatomy of a Large Scale Hypertextual Web Search Engine, *Proc. 19th International Conference on World Wide Web*, pp.107-117 (1998).
- [17] Kleinberg, J.: Authoritative Sources in a Hyperlinked Environment, *Proc. 9th ACM-SIAM Symposium on Discrete Algorithms*, pp.668-677 (1998).
- [18] Kruskal, B.: On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem, *Proc. American Mathematical Society*, Vol.7, No.1, pp.48-50 (1956).
- [19] Castillo, C., Mendoza, M. and Poblete, B.: Information Credibility on Twitter, *Proc. 20th International Conference on World Wide Web*, pp.675-684 (2011).



鳥海 不二夫

2004年東京工業大学大学院理工学研究科機械制御システム工学専攻博士課程修了。現在、東京大学大学院工学研究科准教授。感性工学, エージェントベースシミュレーション, 人工市場等の研究に従事。電子情報通信学会, 人工知能学会, 日本社会情報学会各会員。博士 (工学)。



小出 明弘

1988年生。2011年静岡県立大学経営情報学部経営情報学科卒業。2013年同大学大学院修士課程修了。同年ヤフー株式会社に入社。複雑ネットワークの研究に関心を持つ。



齊藤 和巳 (正会員)

静岡県立大学経営情報学部教授。1985年慶應義塾大学理工学部数理科学科数学専攻卒業, 1988年東京大学博士 (工学)。複雑ネットワークの研究に従事。電子情報通信学会, 人工知能学会, 日本神経回路学会, 日本応用数理学会,

日本行動計量学会各会員。



風間 一洋 (正会員)

1988年京都大学大学院工学研究科精密工学専攻修士課程修了。同年日本電信電話 (株) 入社。2005年京都大学大学院情報学研究科システム科学専攻博士課程修了。2012年和歌山大学システム工学部教授, 現在に至る。Web

情報検索, Webマイニングの研究に従事。博士 (情報学)。人工知能学会, 日本ソフトウェア科学会, 日本データベース学会, ACM各会員。