

## 〈研究論文〉

## SNS トренд可視化ツールを用いた話題の特性に関する考察

飯 尾 淳

Study on the Characteristics of Topics  
using the SNS Trend Visualization Tool

Jun Iio

## Abstract

We have developed TWtrends, a tool for visualizing the topics from Twitter, to analyze how the topics discussed on social networking services are structured. TWtrends can construct a map of the day's topics from Twitter trends, allowing us to see briefly what topics were being discussed that day. In this study, we noticed a marked difference in the results, and revealed that the topics also have their own characteristics. Based on the results of this study, we expect to be able to automatically classify the characteristics of the day's topics.

## KeyWords

TWtrends, Twitter, topic map, social network services, trends characteristics

## 目 次

- I はじめに
- II 本研究の背景
- III 問題設定と分析方法
- IV 事例による分析
- V 時系列的な考察
- VI おわりに

## I はじめに

インターネット上で人々がさまざまな情報を発信, それにより緩やかな人間のつながりを構築するツールであるソーシャル・ネットワーキング・サービス (Social Networking Services, SNS) が普及

して久しい (飯尾, 2022). 古くは MySpace や Mixi に始まり, 昨今では Facebook, Twitter, Instagram など, 多くの SNS が多数の利用者から支持されている. それらの利用が浸透するにつれ, ある種の「社会を映す鏡」としての考察 (飯尾, 2020) が数多くなされるようになった. とくに研究対象としての Twitter は社会科学研究を志す者にとって魅力的な対象と捉えられているようで, Twitter を対象とした研究は世界中で数多く報告されている (Kwak et al., 2010; Castillo et al., 2011; Weller et al., 2014).

SNS を対象とした研究は, いくつかの類型に分けられる. 初期の研究例に多くみられたものは, SNS のユーザーに注目したもので, ネットワーク分析を行う研究である. SNS の特徴として, ユーザー間にいくつかの関連性が定められるというものがある. 例えば, 友だち (Friendship) という関

係性を定義して、つながりのある範囲内でのみ情報を流通させるというものである。Twitterではフォロー、フォロワーといった関係性、および、それらを利用した鍵アカウントと呼ばれる「関係性のあるアカウントでのみ情報を流通させる仕組み」がある。

ネットワーク分析は、それらの関係性を利用して人間関係の特性を明らかにしようというものである（濱岡, 2012; 小出ら, 2013）。また、さらに踏み込んで、情報発信の特性、とくにノイジーマイノリティと呼ばれる、少数、特定のアカウントから話題の多くが発信されるというようなこと、あるいは、ソーシャルメディアで一般的にみられる情報拡散のパターンを明らかにした研究もある（鳥海ら, 2022）。

また、発信される情報の内容に着目した分析事例も多い。その背景には自然言語処理、テキストマイニングの技術が発達し、ツイートされる内容を簡単に分析できるようになったというものがある（Arthika, 2020; Karami et al, 2022）。日本語のテキスト分析には形態素解析が欠かせないが、最近では非常に高性能な形態素解析機も提案されており（松田 2020）、それらはプログラミング・ライブラリとして提供されているため、気軽に自前の処理プログラムに組み込んで利用できる。

さらに、特定のテーマに対して Twitter でどのような意見交換や合意形成がなされているのかといった研究事例も多い。政治や選挙に対する話題形成や当落予測に関する研究（Tumasjan et al., 2011; Jungherr, 2016）や、健康情報に関するもの（Alotaibi et al., 2020）、あるいは、災害時の情報交換に関する研究（宮部ら, 2011; 山本ら, 2012; Bruns and Liang, 2012）などが目立つ。そのほか、関連するものとしては、情報の真偽性判定に関するものやデマの分散といったもの（柿本ら, 2019; 櫛部・伊藤, 2021; 草竹ら, 2021）も行われている。

これらの研究を踏まえ、本研究では、Twitter で語られる話題の特性に特徴があるのかどうかを、事例により明らかにすることを試みた。本論文は、第Ⅱ章で研究の背景を記述、第Ⅲ章では問題設定

と分析の方法について述べる。さらに、第Ⅳ章で特定の事例について分析した結果について報告し、第Ⅴ章では時系列的な観点からの比較を試みる。最後に、第Ⅵ章でまとめと今後の展望について述べる。

## Ⅱ 本研究の背景

筆者らの研究グループでは、Twitter のトレンド機能に着目し、Twitter のトレンドを可視化するシステム TWtrends を開発、2019 年から運用<sup>1)</sup>してきた（Iio, 2019a）。Twitter のトレンドとは、Twitter 社が提供する機能の一つであり、Twitter でいま注目されているトピックを表すキーワードのことである。

TWtrends は、Twitter が提供する API を用いてトレンドの情報を 20 分毎に収集し、さらにそのキーワードに関連するツイートを取得したうえで、そのトレンドに関する共起ネットワークを作成する機能を備える（図 1 左）。また、1 日のトレンドを収集した後に、トピックマップを作成する機能も持つ（図 1 右）。Twitter トレンド抽出のアルゴリズムは公開されていないが、トレンドの特性として、あるトピック（話題）に対していくつものトレンドが抽出される現象が確認されている。トピックマップは、それらのトレンドの関連性をコサイン類似度で計算し、一定の閾値<sup>2)</sup>を超えるものを関連するトレンドとしてまとめた結果を示すものである。

トピックマップに並ぶ楕円はそれぞれのトレンドを示し、それらをノードとして閾値を超えた関連性を持つものがエッジで結ばれる。関連する話題が形成する一つの塊を、我々は「話題クラスタ」<sup>3)</sup>と呼んでいる。

1) <https://tw.t.iiojun.com/>

2) 当初は閾値として 0.75 としていたが、その後のチューニングにより、現在は閾値として 0.5 を用いている。

3) 話題クラスタが持つ色合いに特別な意味はなく、クラスタをわかりやすく示すために異なる色を割り当てている。

## 統一教会の会見

取得日: 2022年09月22日

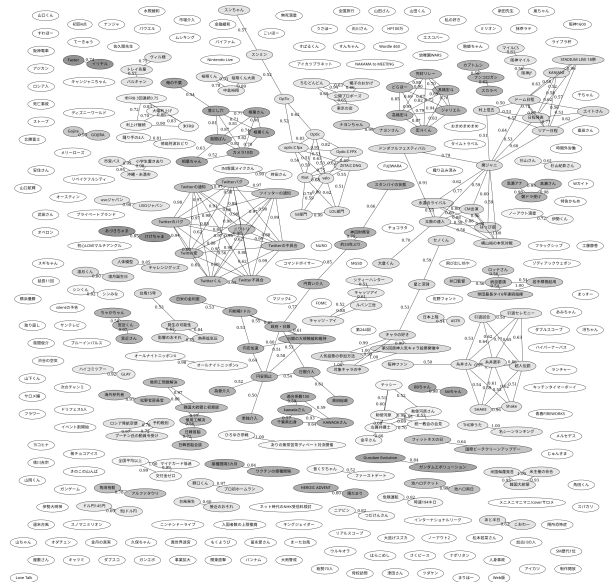
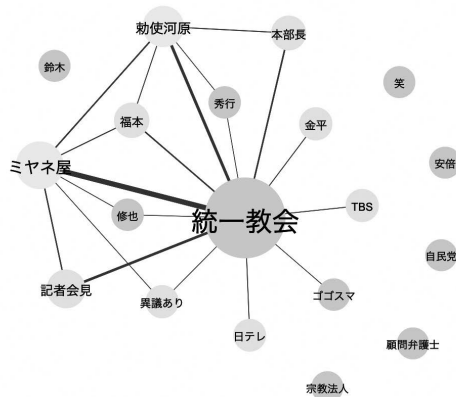


図1 TWtrendsが提供する情報の例：共起ネットワーク（左）とトピックマップ（右）

これまでの運用経験から、トピックマップに示される話題クラスタには、いくつかの類型があることがわかっている（飯尾, 2019b; 金田・飯尾, 2020）。特徴的なものとしては、診断系と呼ばれるものがある。診断系とは、「〇〇診断」などと題されることが多いオンラインゲームの一種であり、特定のテーマに関連して診断結果のような情報<sup>4)</sup>を提示するサービスである。

診断系の話題クラスタは、話題クラスタを構成する全てのノードがエッジで結ばれる密結合のクラスタになる特徴を持つ。診断系サービスにはその結果をツイートする機能があり、あるテーマの診断ゲームが流行するといくつかのトレンドが生成される。それらのトレンドに関連するツイートは診断系サービスが生成するものであり、多くの場合、診断系サービスが提供するキーワードが並べられるものとなっている（図2）。これらのキーワードはさほどバリエーションが多いわけではなく、その結果、関連するトレンドのコサイン類似度を求めると、それぞれが高い値を示す。したがっ

4) その多くは適当にキーワードを生成しているだけであり、実際に何らかの理論に基づいて診断系サービスを提供しているわけではない。

て、診断系の話題クラスタは、全てのノードが密結合されるクラスタとなる特徴を有することになるのである。

Jun Iio  
@iiojun

RTしてくれた人にやります。

- ・呼び方は：
- ・第一印象：
- ・得意そうな教科：
- ・触れ合いたい？：
- ・撫でたい？：
- ・頭の良さ(%)：
- ・して欲しいこと：
- ・最後に一言：

#shindanmaker



図2 診断ゲームによるツイートの例

### Ⅲ 問題設定と分析方法

前章で、診断系の話題クラスタが密結合になる理由を示した。では、他のタイプの話題クラスタではどうだろうか。金田・飯尾 (2020) は、話題クラスタのタイプとして、話題の中心となるトレンドが存在するもの (パターン A)、話題の対象となるイベントが進行するにつれて時系列的に話題が緩やかに変遷していくもの (パターン B)、および先に述べた密結合となる、主に診断系によって形成されるクラスタ (パターン C) と、3 種類の類型化を考察した。ただし、パターン A とパターン B については多分に主観的な議論に留まっている。

本研究ではその考察をさらに推し進め、とくにパターン A とパターン B の違いについて、より客観的、定量的にその違いを示すことにした。分析の手法は以下の通りである。

1. パターン A およびパターン B と考えられる特徴的な話題クラスタを抽出
2. それぞれについて、凝集性クラスタリング (Müllner, 2011) によりクラスタリングを実施
3. 各クラスタの重心を求め、それらの類似度を比較することにより、パターンによる特徴の違いを示す

なお、分析には Python 3.10.5 を使い、Scikit-learn が提供する StandardScaler による正規化と AgglomerativeClustering による凝集性クラスタリングを適用した。

### Ⅳ 事例による分析

本章では、代表的な事例を二つ取り上げ、前章で示した手法によりその違いを示す。

図 3 に、パターン A とパターン B のクラスタが発生しているトピックマップの比較を示す。図 3 左は 2022 年 7 月 2 日のトピックマップで、中央に密結合の大きなクラスタが発生している。このクラスタは、au の通信障害に関するものである。この週末は、長時間にわたり au の通信障害が発生

し社会問題になっていた。

右はその 1 週間後、7 月 9 日のトピックマップである。中央に、多数のノードから構成される大きなクラスタが横たわっている。こちらは左と異なり、緩やかに多くのノードが紐づけられたクラスタである。その話題は、前日に発生した安倍元首相暗殺事件に関するものである。本事件に関連するさまざまな話題が緩やかにつながっているグラフとなっている。

左のクラスタは 54 個のノードから構成され、右のクラスタは 40 個のノードから構成されている。規模は若干左が大きいとはいえ、どちらも通常ではあまり現れない大きさのクラスタである。クラスタの大きさから、いずれの事件も Twitter ユーザーへの影響が大きなものだったことが窺える。

分析の基となるデータは 7 月 2 日および 7 月 9 日のデータである。両日の単語空間<sup>5)</sup>から、対象とする話題クラスタに関連するキーワードのみで構成される部分空間を抽出する。その部分空間において、対象とするクラスタを構成するノードが構成するベクトル群を Pandas の DataFrame で表現、それらに対して Scikit-learn が提供する AgglomerativeClustering を用いて凝集性クラスタリングを行った。なお、クラスタリングを実施する際には、StandardScaler によって正規化してからクラスタリングした。

クラスタリングにおいては三つのクラスタの存在を仮定した。それぞれのグラフをクラスタリングした後、三つに分割されたクラスタ 1～クラスタ 3 までの、それぞれのクラスタの重心<sup>6)</sup>を求め、それぞれの重心間のコサイン類似度を求めた。

クラスタリングの結果を表 1 および表 2 に示す。上から順に、クラスタ番号の 1～3 に対応する。

さらに、それぞれのクラスタにおいて重心の位置を求め、重心間のコサイン類似度を計算した結

5) ここで対象としている単語空間とは、全てのトピックを構成する共起ネットワーク図に登場するキーワードを全て収集し、トレンド毎に出現頻度を成分とする行ベクトルを構成したものを縦に並べたものである。

6) 重心の計算対象は正規化する前のデータとした。



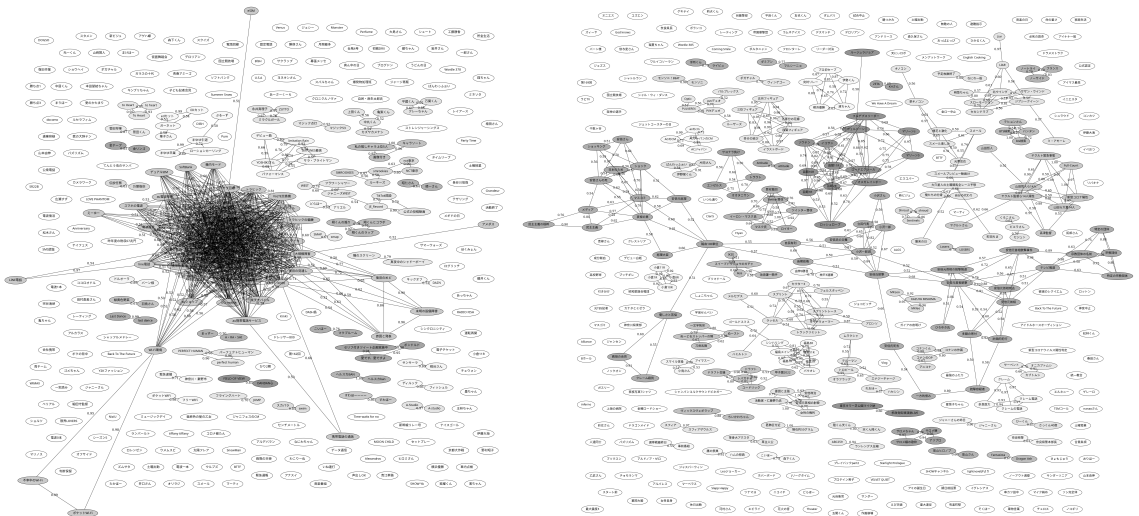


図3 特徴的なクラスターの例：パターン A（左）およびパターン B（右）

表1 携帯電話の障害に関する話題クラスター（7月2日）のクラスタリング結果

KDDI, au と UQ, 復旧のめど, 原因と発表, トラヒックの輻輳, VoLTE 交換機, 復旧の見通し, トラヒック, 未明の設備障害, 携帯電話の通話, 音声・データ, えーゆー, 復旧の目処, UQ mobile, au 通信障害, 通信サービス, 早期復旧, アメダス, au 携帯電話サービス
サブ回線, SoftBank, 楽天モバイル, ポケット Wi-Fi, 不幸中の Wi-Fi, デジチケ, Wi-Fi 環境, eSIM
電波障害, au 回線, au ショップ, au 圏外, au 障害, デュアル SIM, 電波ゼロ, au 電波障害, UQ モバイル, 通信障害, line 電話, au ユーザー, 携帯の電波, au 復活, au くん, スマホの電波, au のせい, 仕事の電話, au の障害, LINE 電話, 機内モード, au さん, アンテナ, 大規模障害, 電波マーク, au 電話, au 復旧

表2 安倍晋三元首相銃撃事件に関する話題クラスター（7月9日）のクラスタリング結果

ひろゆき氏, 宗教団体, 安倍元首相死去, 立民代表, 安倍氏銃撃, 小沢一郎, 宗教団体の名前, 民主主義の根幹, テレビ報道, 小沢さん, 特定の団体, 一方的恨み, 小沢一郎氏, 安倍元首相の銃撃報道, 特定の宗教団体, 銃撃容疑者, 安倍氏死去, 計画的犯行, 安倍元首相, 安倍元首相銃撃事件, 多額の寄付, 安倍元首相銃撃, 安倍氏の災難
民主主義, メディア, ショック, 哀悼の意, 総理大臣, 安倍元総理, 長期政権, 自民有利, 安倍さんの死, ショッキング, 日本のため, マスコミ, 輸血 100 単位, 安倍さん
教授の会見, クレーム殺到, 優しさと苦悩

表3 各クラスターの重心に関するコサイン類似度（左：7月2日, 右7月9日）

	#1	#2	#3
#1	1.000	0.732	0.850
#2		1.000	0.816
#3			1.000

	#1	#2	#3
#1	1.000	0.360	0.142
#2		1.000	0.221
#3			1.000

果を表3に示す。左が7月2日に発生した携帯電話の障害に関する話題クラスターのクラスタリング結果に対するものであり、右が7月9日に発生した安倍晋三元首相銃撃事件に関する話題クラスターのクラスタリング結果に対するものである。

左の結果をみると、明らかに、各クラスター間の

類似度が高く、うまく分割されていないことがわかる。つまり、もともと密結合だったものを無理やり分割してみたものの、それぞれの重心間の類似度は依然として高いままであり、うまく分割できていなかった、という状況と考えられる。

一方の右の結果をみると、対角線は「同じクラ

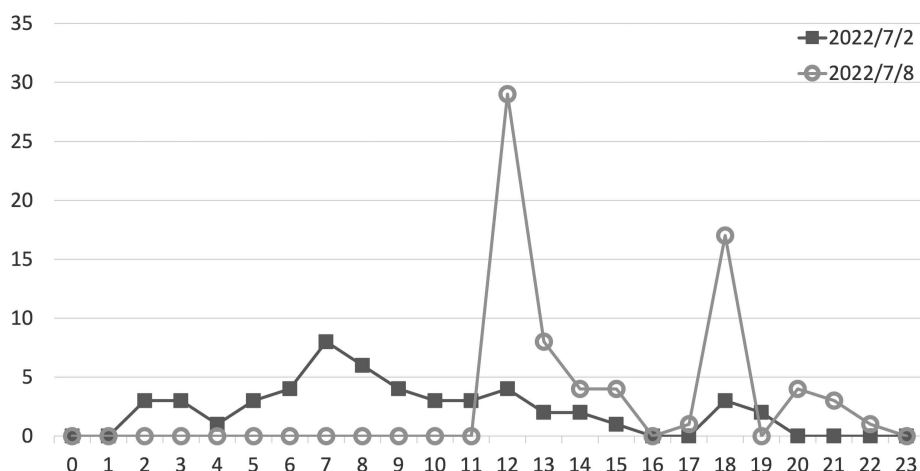


図4 二つの事件に関するトレンドの発生状況（時刻別集計結果）

スタ間の距離」であり「1.0 (=同じもの)」なのは先ほどと同様であるが、それ以外のところでは0.14～0.36と、非常に低い数値になっている。これはすなわち、三つに分類したサブクラスタがきれいに分離できていることを意味する。

## V 時系列的な考察

二つの事件の違い、それらを反映した話題クラスタのパターンに関する相違点をより明確にするために、時系列的な観点での分析を追加で試みた。なお、以降の考察においては、時系列がよりわかりやすい事例として、安倍晋三元首相銃撃事件に関しては、7月9日のものではなく、事件発生当日である7月8日のものを使用した。

図4は、7月2日と7月8日における最大の話題クラスタに関して、それらを構成するトレンドがいつ発生したのかを、時間毎に集計して折れ線グラフで表現したものである。横軸に0時から23時までの時間軸を並べ、縦軸にはその時間帯に収集した関連するトレンドの個数を示している。四角いノードを線でつないだもので示す7月2日のものをみると、通信障害に関するトレンドが日がなほり続いていたことがわかる。

一方、丸いノードをつないだ折れ線で示す7月8日のグラフは、とても特徴的な状況を示してい

る。このグラフは、12時と18時に大きなピークを持つ。安倍晋三元首相が選挙の応援演説中に銃撃されたのは2022年7月8日午前11時半ごろとされている（佐々木, 2022）。事件が報道されるやいなや、Twitterでは大きな話題となったことがこのグラフから窺える。TWtrends システムは20分おきにトレンド情報を収集しているため、11時40分の収集タイミングではまだトレンドとして現れておらず、次の12時のタイミングで本件に関するトレンドを収集できたとの事情もあり、30分程度のタイムラグが存在<sup>7)</sup>する。

次のピークは18時に現れている。銃撃された安倍氏は「心肺停止状態で奈良県立医科大付属病院に運ばれたが、同日午後5時3分に死亡が確認された」（佐々木, 2022）ため、そのニュースが報道されるとTwitterではその件に関する多数のツイートが投稿された。18時に出現したピークは、それらの結果として生じたトレンド群によるものである。

さて、このようなイベント発生状況の違いが、先に示したパターンAとパターンBの違いに現れ

7) TWtrendsの収集タイミングによるタイムラグの他、そもそもTwitterがトレンドとして認識するまでのタイムラグも考慮すべきである。

ていると考えられる。すなわち、今回パターン A として示した通信障害に関しては、通信障害という大きな社会問題に関して状況があまり変化しないまま、しかし Twitter ユーザーへの影響は大きいものが残り続けていたため、「通信障害」というテーマを核とする一つの大きな話題クラスタが出来上がる結果となった。

一方のパターン B として示した元首相銃撃事件に関しては、「元首相銃撃事件」という話題の源泉がありながらも、時系列を追って、複数のイベントが続けて起きた、あるいは、核となる話題が変遷していったと考えられる。その結果、凝集性クラスタリングを行うとサブクラスタにきれいに分類でき、また、時系列的な観点からも一つのテーマのもとでの話題の変遷を確認できたといえる。

## VI おわりに

本研究では、Twitter のトレンドを可視化するツールである TWtrends が示すトピックマップに現れた特徴的な話題クラスタに着目し、先行研究においてパターン A およびパターン B として分類された話題クラスタの形状に関して客観的かつ定量的な分析評価を試みた。

中心的なトピックを持ち比較的密結合となりがちなパターン A に対し、一定の共通話題のもとで時間と共に話題が変遷していくことにより緩いつながりとなるパターン B がどのように異なるのかを、凝集性クラスタリングを適用した結果の各クラスタに関する類似性を計算すること、および、時系列的な集計結果を比較することにより、明示した。

今回示した分析方法の他にも、重要なキーワードを抽出し、それらに関して Jaccard 係数を計算することで、関連性や類似度を計算する方法も考えられよう。今回示した方法だけでは十分な説明がつかないケースも考えられるため、他の指標を用いて、より客観的な話題クラスタの特徴、パターンの違いを示す方法を考案すること、さらに、今回対象としたケース以外でも同様の説明ができることなどを示し、より妥当な手法の提案が、今後

の課題として残されている。

また、それぞれの話題クラスタを代表させるキーワードを如何にして抽出するか、話題クラスタの自動ラベリング問題も、今後、検討が必要な課題の一つである。話題クラスタの生成については、現在、閾値を超えるか否かという単純な方法で行っているが、この方法は、全く異なる話題を一つにまとめてしまうケースと、本来は一つのクラスタとすべきところ、二つ以上のクラスタに分離してしまうなどの問題も含んでいる。そこで、自動ラベリングに関連して、ラベル付けのアルゴリズムを確立したうえで、機械学習を用いたより精緻な話題クラスタの判定方法も、今後、考えていくべきであろう。

## 参考文献

- 飯尾淳 (2022) デジタル世界の拡大と新たな適応・進化、その課題、中央評論, 319, 49-60.
- 飯尾淳 (2020) SNS は世論の代弁者たりえるか？オピニオン、Chuo Online, YOMIURI ONLINE (読売新聞).
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010) What is Twitter, a social network or a news media?. *In Proceedings of the 19th international conference on World wide web*, pp. 591-600.
- Castillo, C., Mendoza, M., & Poblete, B. (2011) Information credibility on twitter. *In Proceedings of the 20th international conference on World wide web*, pp. 675-684)
- Weller, K., Bruns, A., Burgess, J., Mahrt, M., & Puschmann, C. (2014) Twitter and society: An introduction. *Twitter and society [Digital Formations, Volume 89]*, xxix-xxxviii.
- 濱岡豊 (2012) Twitter におけるコミュニケーションの社会ネットワーク分析、第 74 回全国大会講演論文集 2012 (1), 521-522.
- 小出明弘, 齊藤和巳, 風間一洋, 鳥海不二夫 (2013) 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、情報処理学会論文誌 数理モデル化と応用 (TOM) 6 (2), 164-173.
- 鳥海不二夫, 吉田光男, 榊剛史 (2022) ソーシャルメディア上の情報拡散とその理解、電子情報通信学会誌, 105 (4), 326-332
- Arthika, S., Charles, J., Kulathunga, K. M. S., & Lekamge, L. S. (2020) Text Mining Approaches for Social Media Censoring: A Systematic Literature Review, *International Conference on Advances in*

- Computing and Technology (ICACT-2020) Proceedings*, pp. 29-31.
- Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020) Twitter and research: A systematic literature review through text mining. *IEEE Access*, 8, 67698-67717.
- 松田寛 (2020) GiNZA - Universal Dependencies による実用的日本語解析, 自然言語処理, 27 (3), 695-701.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welp, I. M. (2011) Election forecasts with Twitter: How 140 characters reflect the political landscape. *Social science computer review*, 29 (4), 402-418.
- Jungherr, A. (2016) Twitter use in election campaigns: A systematic literature review. *Journal of information technology & politics*, 13 (1), 72-91.
- Alotaibi, S., Mehmood, R., Katib, I., Rana, O., & Albeshri, A. (2020) Sehaa: A big data analytics tool for health-care symptoms and diseases detection using Twitter, Apache Spark, and Machine Learning. *Applied Sciences*, 10 (4), 1398.
- 宮部真衣, 荒牧英治, 三浦麻子 (2011) 東日本大震災における Twitter の利用傾向の分析. 研究報告電子化知財の財産・社会基盤 (EIP), 2011 (17), 1-7.
- 山本雅人, 小笠原寛弥, 鈴木育男, 古川正志 (2012) 東日本大震災時の Twitter における情報伝播ネットワーク. 情報処理, 53 (11), 1184-1191.
- Bruns, A., & Liang, Y. E. (2012) Tools and methods for capturing Twitter data during natural disasters. *First Monday*.
- 柿本大輔, 宮部真衣, 荒牧英治, 吉野孝 (2019) RumorFinder: 流言拡散防止のための情報確認行動促進システム, 和歌山大学災害科学教育研究センター研究報告 3 16-23.
- 櫛部健汰, 伊藤恵 (2021) 観光地に関する風評被害を軽減するための SNS を活用したデマ情報判別システムの提案, 日本ソフトウェア科学会第 37 回大会
- 草竹大暉, 平林真衣, 吉野孝 (2021) Twitter において流言訂正情報を発信・拡散するユーザーの特徴分析, 第 83 回全国大会講演論文集 2021 (1), 161-162.
- Iio, J. (2019a) TWtrends — A Visualization System on Topic Maps Extracted from Twitter Trends, *IADIS International Journal on WWW/Internet*, 17 (2), 104-118.
- 飯尾淳 (2019b) Twitter トレンドの分析 — SNS では何の話題で盛り上がるのか? —, 2019 年度社会情報学会 (SSI) 学会大会, pp. 232-235.
- 金田康之介, 飯尾淳 (2020) Twitter トレンドの可視化による傾向の分析, 第 44 回サイバーワールド研究会, CW2019-09.
- Müllner, D. (2011) Modern hierarchical, agglomerative clustering algorithms. *arXiv preprint arXiv:1109.2378*.
- 佐々木洋 (2022) 事件がわかる 安倍晋三元首相銃撃事件, 毎日新聞, 2022/8/3 10:30 (最終更新 9/8 14:18), <https://mainichi.jp/articles/20220802/osg/00m/040/001000d>