

# グラフの分極度によるマスメディアとTwitterユーザの意見の偏りの分析

藤兼 由生<sup>†</sup> 風間 一洋<sup>††</sup> 吉田 光男<sup>†††</sup> 土方 嘉徳<sup>††††</sup>

<sup>†</sup> 和歌山大学大学院システム工学研究科 〒640-8510 和歌山県和歌山市栄谷 930

<sup>††</sup> 和歌山大学システム工学部 〒640-8510 和歌山県和歌山市栄谷 930

<sup>†††</sup> 豊橋技術科学大学情報・知能工学系 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

<sup>††††</sup> 関西学院大学商学部 〒662-8501 兵庫県西宮市上ヶ原一番町 1-155

E-mail: †s216241@wakayama-u.ac.jp

あらまし 近年、フェイクニュースや釣り記事の報道や、政府の報道機関への介入や利害関係者への忖度などによる報道の抑制などの情報発信側の問題に、フィルターバブルやエコーチェンバー現象などの情報受信側の問題が組み合わさることで、ソーシャルメディアのユーザの意見の分極化が促進されると共に、あるトピックに関する情報を公平な立場から評価することが困難となっている。そこで本稿では、メディアの報道とソーシャルメディアの意見の偏りを分析し、それらを公平に理解することを目標としたメディア・ユーザ意見分析システムのプロトタイプを作成し、その有効性について検証する。まず、あるトピックに関してメディアが発信したニュース記事と、Twitter上でそのニュース記事を言及したツイートから、メディア・ユーザ関係グラフを構築する。さらに、メディアの被共閲覧関係から求めたメディアグラフ、ユーザの共閲覧関係から求めたユーザグラフを作成し、それぞれのグラフのクラスタ構造と分極度から主義主張の偏りを分析する。

キーワード Twitter, ニュース, 分極化, Random Walk Controversy, ネットワーク分析

## 1 はじめに

Twitter や Facebook などのソーシャルメディアの普及より、マスメディアがソーシャルメディアを利用して報道記事を発信したり、逆にソーシャルメディア上のユーザの意見を報道するなど、情報発信者である報道機関と情報受信者であるソーシャルメディアユーザが密に相互作用しあうようになった [1].

情報の発信側の観点からは、Web やソーシャルメディアを利用した情報発信が容易になり、実世界の経済活動や政権アピールなどに活用されるようになった結果、情報の信頼性を損なう行為が増加するようになった。例えば、利益獲得を目的として事実とは異なるフェイクニュースや釣り記事が多く生み出されている。特にソーシャルメディアでは、そういった記事ほど早く広範囲に拡散する傾向がある [2]. また、アメリカの報道機関の事例と異なり、日本国内では、政府の報道機関への介入やクロスオーナーシップによるマスメディア集中、記者クラブにおける談合、報道機関の利害関係者への忖度などにより、たとえ重要な問題や事実であっても、都合が悪い情報は報道されなくなる事例が見られる [3]. これにより、フェイクニュースや釣り記事のように意図的でなくとも、ユーザが多様な情報を知る機会が失われている。

情報の受信側の観点からは、ユーザの情報入手は容易になったが、情報の洪水に悩まされるようになり、この量的な問題の解決のために協調フィルタリングなどの情報推薦やユーザの行動履歴に基づいたパーソナライズが導入された結果、情報獲得の多様性が損なわれるようになった。例えば、広告などの収益最大化のためにユーザが確実に好む一部の情報だけを提示して

アクセス数の増大を図ったことで、ユーザが見ることができる情報を制限してしまうフィルターバブル [4] の問題が生じた。さらに、ソーシャルメディアの隣接ユーザと興味や信念、環境などが類似する同類性が高いソーシャルグラフは、情報推薦やパーソナライズと同様に伝搬する情報を制限する機能を持つだけでなく、類似した隣接ユーザとの情報交換を繰り返させることによって、特定の偏った情報や信念が増幅されるエコーチェンバー現象を引き起こした [5]. このような状況を踏まえて、情報受信者が公平な観点から意思決定するためには、ある事象に関する異なる視点からの報道・議論の全体構造を俯瞰し、中立的な視点から多様なある情報を理解することが必要となる。

本稿では、メディアの報道とソーシャルメディアの意見の偏りを分析し、それらを公平に理解することを目標としたメディア・ユーザ意見分析システムのプロトタイプを作成し、その有効性について検証する。まず、特にマスメディアが報道したニュースがソーシャルメディア上で拡散・言及される現象に着目し、あるトピックに関してメディアが発信したニュース記事と、Twitter上でそのニュース記事を言及したツイートから、メディア・ユーザ関係グラフを構築する。さらに、メディアの被共閲覧関係から求めたメディアグラフ、ユーザの共閲覧関係から求めたユーザグラフを作成し、それぞれのグラフの基本統計量、分極度、クラスタ構造などから主義主張の偏りを分析する。

## 2 関連研究

マスメディアが報道するニュース記事には、それぞれの報道機関の方針が反映されており、その報道傾向を分析する研究が存在する。張らはニュース記事のテキストに着目し、ニュー

ス記事の書き方を「印象」という評価指標で分析することで、ニュースサイトの報道傾向を視覚的に比較・分析する手法を提案した [6]。また、田中らは、報道媒体である新聞の論調が公共事業の実施、公共事業の経済効果に対して批判的である傾向が国民世論やその世論形成に影響を及ぼすとして、大手新聞社の公共事業に対する批判的な報道の時系列的な変遷を分析した [7]。

一方で、Twitter ユーザも自身の主義や嗜好に合致した報道機関のニュースを選んで読み、さらに広く拡散する傾向がある。Garimella らは、ソーシャルメディア上での政治に関するエコーチェンバー現象を分析し、Twitter ユーザは自分の意見に同意する政治的意見にかなりの程度さらされていることを明らかにした [8]。大前らは、Twitter 上で政治的分極化が起きたトピックについてツイートを収集し、リツイートと引用リツイートの関係から作成したユーザグラフを用いて革新派と保守派といった政治的分極性を分析した [9]。また Conover らは、アメリカの議会選挙期間中のツイートデータを収集し、特定のハッシュタグが含まれるツイートをメンションしたユーザ、リツイートしたユーザの関係からそれぞれグラフを作成し、それらに含まれるコミュニティに着目することで、各議会に対するユーザの支持傾向を分析した [10]。

さらに、マスメディアの報道したニュース記事内容と Twitter ユーザが発信したツイート内容に着目した分析も行われている。Sen らは、インドの経済施策に対し、マスメディア、議会、ソーシャルメディアの立場の人々がどのような意見を主張しているかを、報道記事、ツイート、議会が行政活動に提起した質問のような主張が反映されているテキストから LDA で抽出したトピックごとに分析した [11]。

以上のようにマスメディアの報道傾向に関する分析や Twitter ユーザの意見の分極性に関する分析は多く行われているが、報道機関またはユーザだけに注目した対立構造の分析がほとんどであった。本稿では、報道機関の姿勢が明確または一律的に統制される海外と異なり、政府の介入や忖度により報道機関の対立構造が時々刻々と変化し、ユーザもそれに影響されると仮定して、報道機関やユーザのグループを対応づけながら、双方の影響を考慮した意見の分極性を分析する。

### 3 メディアとユーザの意見の分析手法

#### 3.1 メディア・ユーザ意見分析システム

意見の対立分析は主に政治関連のトピックに関して行われることが多かった。しかし、意見の対立の発生は必ずしも政治関連に限らないことから、その他の事件や災害、イベントについてもメディアの報道がソーシャルメディアに与える影響を迅速に分析できることが望ましい。さらに、最近のフェイクニュースや釣り記事による報道や、政府の報道機関への介入や利害関係者への忖度などによる報道の抑制は、ソーシャルメディアのユーザの意見のさらなる分極化を招くと共に、中立的な視点から報道や議論の全体を理解することが困難になっている。

そこで、任意のトピックに対してメディアとユーザの意見の

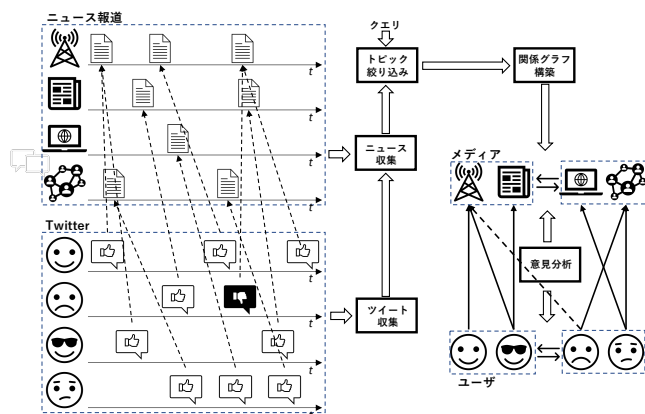


図1 システムの基本アーキテクチャ

偏りや対立を報道・発言内容と関係性から分析するメディア・ユーザ意見分析システムのプロトタイプを作成した。本システムは図1に示すような基本アーキテクチャを持ち、以下の手順で分析する。

- (1) ニュースとツイートデータの収集
- (2) 特定トピックのニュースとツイートへの絞り込み
- (3) メディア・ユーザ関係グラフの構築
- (4) メディアとユーザの意見の分析

#### 3.2 ニュースとツイートデータの収集

まず、報道されたニュースがソーシャルメディア上で拡散・言及される現象に着目し、事前に複数のメディア（情報発信者）のニュースと、それに対してソーシャルメディアユーザ（情報受信者）が反応したツイートを収集する。具体的には、まず観測対象とするメディアへのリンクを含むツイート・リツイートを継続的に収集し、さらにリンク先のニュース記事を抽出する。さらに、収集したデータに関係や属性などを解析した後に、全文検索可能なデータベースに登録する。このデータベースには Elasticsearch<sup>1</sup>を用いる。

以上の処理は、分析前におこなう。

#### 3.3 特定トピックのニュースとツイートへの絞り込み

分析時には、関連記事間のキーワード共有性 [12] を用いて、収集したデータを特定のトピックに絞り込む。キーワード共有性とは、現在注目している記事がある場合に、その関連記事が同一のキーワードを共有している性質のことであり、文書の内容の類似性を用いるよりも適切に関連記事を抽出できることが知られている。

まず、分析対象とするトピックの内容を表すキーワードを手で決定し、それらで構成されるクエリを用いてニュース記事を全文検索して、該当トピックに関するニュース記事集合を抽出する。次に、各ニュース記事の URL を用いてツイートを検索し、検索されたニュース記事集合をツイート・リツイートしたツイート集合を抽出する。この結果、特定トピックに関するニュース記事集合とツイート集合を得る。

1 : <https://www.elastic.co/products/elasticsearch>

なお、適切なトピックに絞り込むためにはキーフレーズの選択が重要であり、さまざまな自動選択手法が提案されているが [12]、現時点では人手で選択している。この理由は、トピックによっては関連記事であっても、さらに別のサブトピックに分かれていることがあり、そのような場合にクエリを人手で修正してトピック粒度やトピック分離を制御できるからである。

### 3.4 メディア・ユーザ関係グラフの構築

次に、特定トピックのデータからメディアとユーザの関係グラフを構築する。

ここで、メディアは情報発信者である報道機関などを表し、主に Web 上で情報を公開しているマスメディアを対象とする。メディア識別子として、ニュース記事を提供するサーバの URL のホスト部を用いるが、ホストの FQDN の先頭が“www”だった場合には削除する。これに伴い、基本的には同一ドメインで公開しているニュース記事をグループ化して扱えるようになる。また、メディア識別子から報道機関を容易に推測できるようになるが、NHK など、同一組織であっても報道内容によって提供するサーバを使い分けている場合に、それらを区別できる。

ユーザは Twitter ユーザを表し、メディアの報道した記事に何らかの反応をしたユーザを対象とする。ユーザ識別子として、Twitter のユーザ ID を用いる。これに伴い、同一ユーザのツイートをグループ化して扱えるようになる。ユーザの識別が容易になる。ただし、そのトピックで過半数を占めるツイート回数が 1 回だけのユーザは、一時的な興味で言及しただけであり、トピックに関する実質的な議論には参加していないものとみなして除去した。これにより、ユーザのクラスタは小さくなるが、クラスタ間の関係がより明瞭となる。

メディア識別子を用いてグループ化したニュース記事集合をメディアノード、ユーザ識別子を用いてグループ化したツイート集合をユーザノードとして、メディアノードに含まれるいずれかのニュース記事を、ユーザノードに含まれるいずれかのツイートがリンクしていた場合に、メディアノードとユーザノードの間にエッジを張ることで、2 部グラフを構築して分析対象とする。これをメディア・ユーザ関係グラフと呼ぶ。

なお、本システムではメディア・ユーザ関係グラフのグラフ構造を分析するだけでなく、メディアノードにはニュース記事集合のテキスト群が、ユーザノードにはツイート集合のテキスト群が対応し、それぞれ報道時刻やツイート時刻の時間情報を持つことから、さらにテキスト内容、時系列特性の分析が可能である。

### 3.5 メディアとユーザの意見の分析

将来的にはグラフ・テキスト・時系列のような複数の側面から分析する予定だが、本稿では、提案するアーキテクチャの有用性の確認と、メディアやユーザの意見の偏りや分断が観測できるかを検証するために、構築したメディア・ユーザ関係グラフに対して、基本的なグラフ構造の分析を試みる。

まず、メディアとユーザの関係を表すメディア・ユーザ関係グラフから、メディアの被共閲覧関係に基づいてメディア間の

関係を表すメディアグラフと、ユーザの共閲覧関係に基づいてユーザ間の関係を表すユーザグラフを作成する。この時、メディアの被共閲覧関係またはユーザの共閲覧関係の強さを、メディアを閲覧したユーザ集合またはユーザに閲覧されたメディア集合の類似度として Jaccard 係数を用いて定量化し、その値の分布から決定した閾値を越える場合にエッジを作成する。なお既存のグラフ分析をそのまま用いるために、どちらのグラフも単純・無向グラフとして表現する。

メディアグラフとユーザグラフから最大連結成分を抽出し、次の 4 種類の分析手法を適用して、エッジで結ばれたメディアやユーザは同類性が高い、つまり意見が類似していると仮定して、メディアとユーザの構造や分極の度合いを分析する。最大連結成分に限定する理由は、孤立ノードになるような複数のメディアを見ないユーザや、あまり見られないメディアを除くと同時に、既存の単一連結成分を前提としたグラフ分析手法を適用できるようにするためである。

#### 3.5.1 グラフの基本統計量による分析方法

メディアグラフとユーザグラフがどのような構造的特徴を持つかを調べるために、最大連結成分のノード数、エッジ数、平均次数、最大次数、クラスタ係数、密度、平均最短距離を用いて分析する。

#### 3.5.2 グラフの分極度による分析方法

あるトピックに関してメディアとユーザそれぞれに主義主張の偏りが存在するとしたら、それらの度合いがメディアグラフ、ユーザグラフの分極度に現れると仮定して、それぞれのグラフの分極度を **RWC** (Random Walk Controversy) [13] で定量化し、グラフの特性を分析する。RWC は今まで提案されてきた他の分極度の指標よりも分離性が良いという利点を持ち、グラフ構造だけから計算できる。Garimella らが提案した RWC はグラフの分極度を定量化する手法であり、(1) グラフ構築 (2) グラフ分割 (3) 分極度の測定、という手順で求める。なお、METIS アルゴリズム [14] を用いて、エッジ切断数が最小となるように、グラフを 2 分割する。

分極度  $RWC$  は、分割された 2 つのノード群  $X, Y$  に対し式 (1) で計算する。

$$RWC = P_{XX}P_{YY} - P_{YX}P_{XY} \quad (1)$$

ここで、ノード群  $A$  からランダムウォークを開始してノード群  $B$  で終了する条件付き確率を  $P_{AB}$  とする。 $RWC$  は理論的には -1 から 1 の間の値を取り、1 に近づくほどグラフの分極度が大きいことを示す。ただし、Garimella らの論文の実験結果を見ると、トピックによって多少異なるが、実際の  $RWC$  の値の分布はより狭い範囲に集中し、0 付近が一番大きい傾向がある。

さらに、METIS アルゴリズムでどのように分割されるかを示すために、メディアグラフとユーザグラフの構造を Gephi の Force Atlas2 アルゴリズムを用いて可視化し、異なる色で着色する。なお、ユーザグラフはメディアグラフよりもノード数・エッジ数共に大幅に多いために、ノードが隠されないようにエッジは描画しない。ただし、RWC では、グラフ構造において意見の対立が起こっていると仮定して 2 分割するために、特

に分極度が小さく、明確な対立が存在しない場合には、可視化結果からわかる分割が直感に反するものになることに注意が必要である。

### 3.5.3 メディア集合の特性分析方法

一般的に、メディアグラフはノード数、エッジ数共に少なく、基本的に公平性を保つという報道倫理に従うことと、ニュース記事の共有も行われることから、明確または複雑なクラスタ構造を持たない傾向がある。

そこで、RWCが比較的高いトピックに対して、メディアグラフをMETISアルゴリズムで分割して得られる2つのメディア集合に、どのようなメディアが含まれ、そのメディアのニュースの被言及数を調べることで、それぞれのメディア群の報道姿勢の違いが反映されるかを分析する。

### 3.5.4 ユーザクラスタの特性分析方法

一般的に、ユーザグラフはノード数、エッジ数共に非常に大きくなる傾向があることからグラフ構造は明確または複雑になり、RWCで用いるMETISアルゴリズムによる2分割結果と直感的に整合しない場合も多い。

そこで、ユーザグラフをLouvain法[15]を用いて、より多くのクラスタに分割する。Blondelらが提案したLouvain法は、Modularityに基づくグラフ分割手法の一つであり、局所最適化と同一コミュニティに属する頂点の集約により、元となったNewmanらのアルゴリズムよりも大規模なグラフ構造を高速に処理でき、高いModularityが得られる。

ユーザグラフがどのようなコミュニティ構造を持つかを明確に示すために、GephiのForce Atlas2アルゴリズムでノードをレイアウトした結果を、クラスタ別に着色して可視化して分析する。なお、ノード数、エッジ数共に膨大なので、レイアウトされたノードや構造が隠されないようにエッジは描画しない。

さらに、ユーザクラスタの特性を、ユーザが言及したメディアとその言及数を用いて分析する。

## 4 分 析

### 4.1 データセット

Twitter APIを用いて2018年5月1日から2018年5月31日までの1ヶ月間の間に拡散(リツイート)されたツイートのうち、ツイート内容にニュース記事URLが含まれるツイート326,906件を言及ツイートデータセットとして利用した。

またニュースデータセットとして、統合型メタ検索エンジンCeek.jp News<sup>2</sup>で取り扱われているニュース記事のうち、言及ツイートデータセット内に含まれるツイートにURLが含まれるニュース記事137,466件を利用した。

### 4.2 トピックの選定

はじめに分析対象とするトピックを選定した。本来は分析対象としては意見の偏りが存在するトピックを選定することが望ましいが、そのようなトピックの発見手法をまだ確立していないために、本稿では、Twitter上で多く言及されているニュー

スの内容から、評価対象とするトピックを選定した。実際には、まず2018年5月の1ヶ月間におけるTwitter上の言及回数の上位10件のニュース記事内容から、以下のニューストピックと、クエリとして用いるキーワードを人手で選定した。

#### a) 東京オリンピック

2020年東京オリンピック・パラリンピックに向けて、大会組織委員会が公表したボランティアの募集要項案の宿泊費などが自己負担となっていることに批判があったことなどを受けて、ボランティアの在り方などを有識者が検討する初会合が開かれたことに関する報道である。

#### b) セクハラ罪

財務省の福田淳一前事務次官のセクハラ問題を巡り、麻生太郎副総理兼財務相が「『セクハラ罪』という罪はない」と繰り返し発言したことに関する報道である。

#### c) 西城秀樹

歌手の西城秀樹さんが亡くなった訃報に関する報道である。

#### d) 加計学園

加計学園の獣医学部新設をめぐる問題に関する報道。加計学園が52年間どこの大学にも認められていなかった獣医学部を新設する「国家戦略特区」の事業者に選定された。その際に加計孝太郎理事長が安倍首相の長年の友であったため特別の便宜を疑われた問題に関する報道である。

#### e) ウイングアークフォーラム

最新テクノロジーやデータを活用する企業が一堂に会し、先進的な取り組みを共有するカンファレンス「ウイングアークフォーラム2017」に関する報道である。

#### f) 日大アメフト部

日本大学と関西学院大学の定期戦で、日大の選手が関学大の選手に危険なタックルをして負傷させた問題。しかし、日大の部員の証言から実際には日大の監督やコーチが危険なタックルを強要していたことが発覚したことに関する報道である。

### 4.3 グラフの基本統計量による分析

選定した6トピックのメディアグラフ、ユーザグラフを作成し、それらのグラフ構造の特性をグラフの基本統計量を用いて分析した。表1、表2にメディアグラフとユーザグラフの基本統計量を示す。

メディアグラフに関しては、ノード数に対して「セクハラ罪」、「加計学園」、「日大アメフト部」のエッジ数、平均次数、最大次数が特に大きい。これは、これらのトピックでは、ユーザが複数のメディアを参照してツイートしている傾向があることを示しており、トピックとしても多くの人の話題になりやすい内容だと考えられる。また、クラスタ係数はどれも0.4以上と高い値を示すが、これはメディアグラフがメディアの被共閲覧関係に基づいて構築されていることが一つの要因である。ただし、「セクハラ罪」、「ウイングアークフォーラム」の密度はかなり低い。これによりクラスタ内部が密であっても、全体的に見ると疎である傾向があることがわかる。また、平均最短距離は「東京オリンピック」、「ウイングアークフォーラム」の場合が長く、これはユーザが見るメディアに偏りがあることを示し

<sup>2</sup>: <http://news.ceek.jp/>

ている。

ユーザグラフに関しては、ノード数に対して「セクハラ罪」, 「加計学園」, 「日大アメフト部」のエッジ数, 平均次数, 最大次数が特に大きい傾向は同じである。なお、ノード数がメディアグラフより多いものの現実のユーザを考えると少ない理由は、単一のメディアのニュースしか見ていないユーザが大部分であり、そのようなユーザはトピックに対し一時的に興味を示しているに過ぎないと仮定して取り除き、さらに提案手法で Jaccard 係数による足切りと最大連結成分に限定したことによって除去されるからである。また、クラスタ係数と密度の差がより顕著になる理由は、ユーザグラフの方がより明確なクラスタ構造を持つからだと考えられる。

#### 4.4 グラフの分極度による分析

3.5.2章で述べた RWC を用いて、メディアグラフとユーザグラフの分極状態を分析した。6 トピックのメディアグラフとユーザグラフの RWC を表3に示す。なお、すでに述べたように実際には RWC の値は-1 と 1 の間の狭い範囲に分布するが、Garimella らの実験結果で 0.3~0.4 以上の場合に分極が起きていたことから、本稿では 0.3 を分極の有無の目安とした [13]。

メディアグラフに関しては、「東京オリンピック」と「ウイングアークフォーラム」の RWC が高く、「西城秀樹」, 「加計学園」, 「日大アメフト部」の RWC は低い。

ユーザグラフに関しては、RWC の値はメディアグラフよりも高くなる傾向があり、「ウイングアークフォーラム」の RWC はさらに高いが、「東京オリンピック」の RWC は比較的低い。

そこで、RWC がメディアグラフとユーザグラフ共に高いトピックとして「ウイングアークフォーラム」を、共に低いトピックとして「加計学園」を選び、そのメディアグラフ・ユーザグラフの可視化結果をそれぞれ図2, 図3に示す。可視化結果は METIS アルゴリズムの分割結果に応じて、2色に彩色した。

「ウイングアークフォーラム」の場合は、RWC が示すようにメディアグラフ、ユーザグラフ共に明確に2つのクラスタに分かれていることがわかる。それに対して、「加計学園」はメディアグラフに関しては明確なクラスタ構造が見られず、ユーザグラフにおいては複数のクラスタがゆるやかにつながっている構造であり、明確に2極化しているとは言い難いことがわかる。

#### 4.5 メディア集合の特性分析

次に、メディアグラフを METIS アルゴリズムで分割した時の、2つのメディア集合の特性の違いについて分析する。「ウイングアークフォーラム」と「加計学園」の各メディア集合の被言及数上位5件のメディアとその被言及数を表4と表5に示す。

「ウイングアークフォーラム」の場合は、メディア集合1では logmi.jp や pc.watch.impress.co.jp のような IT 系のメディアが、メディア集合2では sankei.com や www3.nhk.or.jp のような新聞社やテレビ局などのメディアが含まれていた。「加計学園」の場合は、どちらにも新聞社やテレビ局などのメディアが含まれていて、明確な差は見られなかった。つまり、分極度

が高い場合には、ノード集合を構成するメディアの属性が異なることがわかる。

#### 4.6 ユーザクラスタの特性分析

次に、ユーザグラフを3.5.4章で述べたように Louvain 法で複数のクラスタに分割した時の、ユーザクラスタの特性の違いについて分析する。

まず、「ウイングアークフォーラム」と「加計学園」のクラスタの可視化結果を図4(a)と図4(b)に示す。この結果から、RWC が高い「ウイングアークフォーラム」は分極しているユーザ集合とクラスタがほぼ対応しているが、RWC が低い「加計学園」では必ずしもうまく対応していないことがわかる。

次にユーザクラスタの特性を、ユーザがツイートしたメディア集合に着目して分析する。「ウイングアークフォーラム」は5個、「加計学園」は4個のクラスタに分割されたが、METIS アルゴリズムで分割した2つのユーザ集合に対応するクラスタから、それぞれ最大のクラスタを選択して、分析に使用した。

「ウイングアークフォーラム」と「加計学園」の各ユーザクラスタが言及したメディアとメディアごとの言及数の上位5件を表6と表7に示す。

「ウイングアークフォーラム」の場合は、クラスタ1では logmi.jp や pc.watch.impress.co.jp のような IT 系のメディアが、クラスタ2では sankei.com や www3.nhk.or.jp のような新聞社やテレビ局などのメディアが含まれていた。「加計学園」の場合は、どちらのクラスタも新聞社やテレビ局などのメディアが含まれていて、明確な差は見られなかった。つまり、メディア集合と同様に、ユーザクラスタも分極度が高い場合には、ユーザがツイートしたメディア集合の属性が異なることがわかる。さらに、トピックによってメディアグラフとユーザグラフの構造やクラスタリング結果がかなり異なるものの、メディア集合とユーザクラスタの間に相互依存する関係が存在し、その違いがトピックの性質を表している可能性がある。

#### 4.7 考 察

本稿では、Twitter 上の言及数上位のニュースの内容から評価対象とするトピックを選定して評価したが、トピックによってメディアグラフやユーザグラフの構造や特性が大きく異なる事が明らかになった。

例えば、エッジ数や次数が大きかった「セクハラ罪」, 「加計学園」, 「日大アメフト部」は、不祥事とその隠蔽に関するトピックであることから多くのユーザに注目されていたことがわかるが、メディアグラフ、ユーザグラフ共に RWC の値は低い。この理由として、メディア・ユーザ共に批判的な論調の報道やツイートが大部分であったから、つまり分極化はソーシャルメディア上ではなく、実社会とソーシャルメディアの間で発生していたからだと考えられる。

これに対して、「ウイングアークフォーラム」は RWC の値は高く、実際にメディア集合やユーザクラスタの特性が異なっているだけでなく、それぞれの間に対応関係が見られた。つまり、ユーザの分極化だけでなく、メディアの分極化も起こるこ

| トピック名        | ノード数 | エッジ数 | 平均次数   | 最大次数 | クラスタ係数 | 密度    | 平均最短距離 |
|--------------|------|------|--------|------|--------|-------|--------|
| 東京オリンピック     | 20   | 47   | 4.700  | 10   | 0.557  | 0.247 | 3.142  |
| セクハラ罪        | 99   | 503  | 10.162 | 34   | 0.473  | 0.104 | 2.674  |
| 西城秀樹         | 26   | 70   | 5.385  | 15   | 0.440  | 0.215 | 2.083  |
| ウイングアークフォーラム | 56   | 118  | 4.214  | 15   | 0.422  | 0.077 | 2.892  |
| 加計学園         | 50   | 484  | 19.360 | 42   | 0.737  | 0.395 | 1.674  |
| 日大アメフト部      | 51   | 425  | 16.667 | 32   | 0.748  | 0.333 | 1.816  |

表 1 メディアグラフの基本統計量

| トピック名        | ノード数  | エッジ数    | 平均次数    | 最大次数 | クラスタ係数 | 密度    | 平均最短距離 |
|--------------|-------|---------|---------|------|--------|-------|--------|
| 東京オリンピック     | 89    | 1,528   | 34.337  | 59   | 0.865  | 0.390 | 1.830  |
| セクハラ罪        | 1,482 | 291,461 | 393.334 | 881  | 0.716  | 0.266 | 1.839  |
| 西城秀樹         | 144   | 1,985   | 27.569  | 73   | 0.697  | 0.193 | 2.121  |
| ウイングアークフォーラム | 184   | 2,088   | 22.696  | 61   | 0.756  | 0.124 | 2.742  |
| 加計学園         | 1,712 | 489,063 | 571.335 | 1145 | 0.705  | 0.334 | 1.676  |
| 日大アメフト部      | 1,178 | 192,244 | 326.390 | 722  | 0.709  | 0.277 | 1.740  |

表 2 ユーザグラフの基本統計量

|              | メディアグラフ |      |       |        | ユーザグラフ |         |      |       |
|--------------|---------|------|-------|--------|--------|---------|------|-------|
|              | ノード数    | エッジ数 | 閾値    | RWC    | ノード数   | エッジ数    | 閾値   | RWC   |
| 東京オリンピック     | 20      | 47   | 0.1   | 0.493  | 89     | 1,528   | 0.3  | 0.197 |
| セクハラ罪        | 99      | 503  | 0.03  | 0.220  | 1,482  | 291,461 | 0.25 | 0.150 |
| 西城秀樹         | 26      | 70   | 0.04  | -0.216 | 144    | 1,985   | 0.25 | 0.257 |
| ウイングアークフォーラム | 56      | 118  | 0.03  | 0.320  | 184    | 2,088   | 0.25 | 0.795 |
| 加計学園         | 50      | 484  | 0.004 | -0.038 | 1,712  | 489,063 | 0.25 | 0.150 |
| 日大アメフト部      | 51      | 425  | 0.01  | -0.012 | 1,178  | 192,244 | 0.2  | 0.173 |

表 3 メディアグラフ・ユーザグラフの RWC

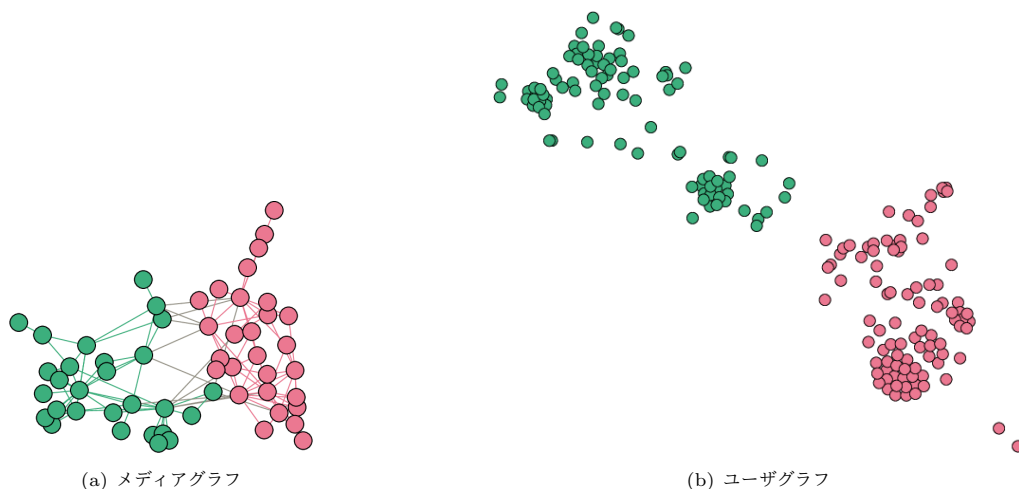


図 2 「ウイングアークフォーラム」の分極化の可視化結果

と、それらの間に何らかの関係が存在する可能性があることがわかる。

ただし、「東京オリンピック」の場合は、メディアグラフの RWC は高く、ユーザグラフの RWC は低かった。本稿ではページ数の関係から詳細を割愛するが、この理由は該当期間に「東京オリンピック」について日常的に報道しているメディアは NHK だけであり、それに関してしかツイートしなかったユーザが大部分であったことと、その他のメディアは東京オリンピックに関する批判的な報道を少数報道していたからであった。このように、メディア側とユーザ側で異なる傾向を示すこともあ

り、メディアとユーザの両方の側面から互いに関係付けて分析することの重要性が明らかになったと言える。

なお、「ウイングアークフォーラム」ではエッジ数や次数が大きくはないことからわかるように、トピックの内容は会議に関するもので意見の対立を産みにくいことから、この分極は新聞社・テレビ系のマスメディアと IT 系のオンラインニュースという異なるメディア属性と、メディア属性に関するユーザの嗜好性から生まれたと推測できる。しかし、本研究で注目する分析対象は、報道姿勢の対立や、フェイクニュースや釣り記事などの誤った情報、政府の介入や利害関係者への忖度によるメディ



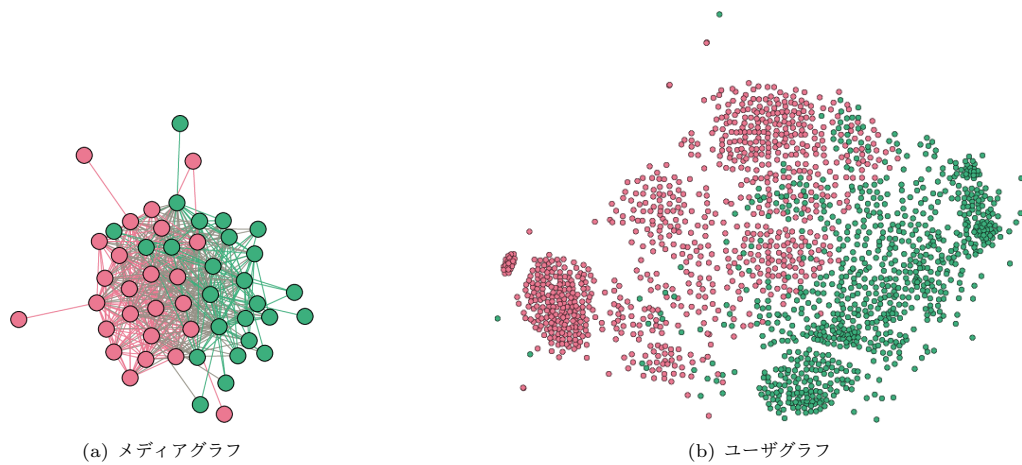


図3 「加計学園」の分極化の可視化結果

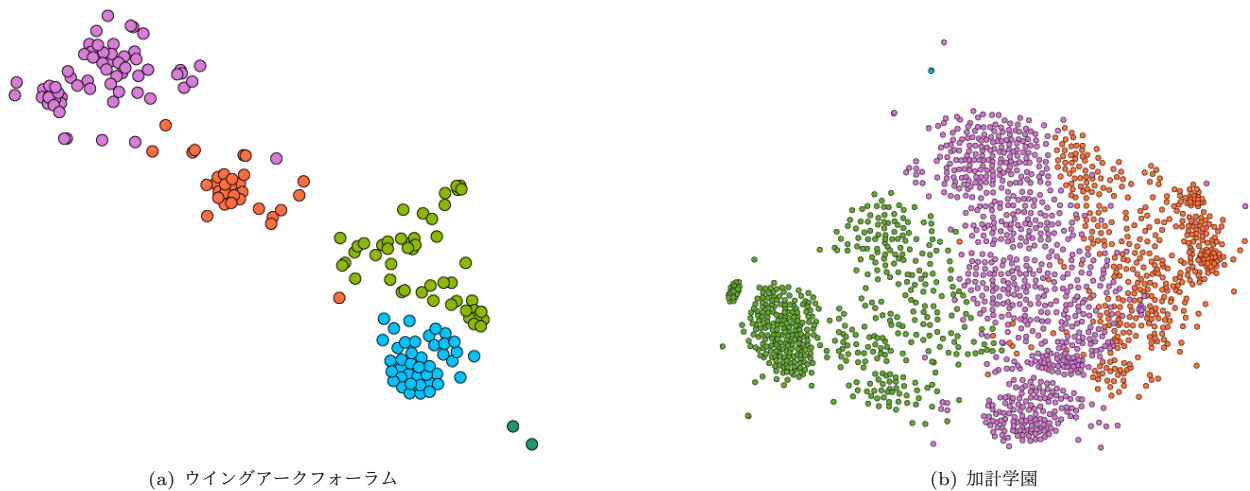


図4 ユーザグラフのクラスターの可視化結果

(a) メディア集合 1

| メディア名             | 識別子                          | 被言及数 |
|-------------------|------------------------------|------|
| ログミー Biz          | logmi.jp                     | 367  |
| PC Watch          | pc.watch.impress.co.jp       | 155  |
| 朝日新聞              | asahi.com                    | 150  |
| ASCII.jp          | ascii.jp                     | 96   |
| AKIBA PC Hotline! | akiba-pc.watch.impress.co.jp | 86   |

(b) メディア集合 2

| メディア名  | 識別子                | 被言及数 |
|--------|--------------------|------|
| 産経新聞   | sankei.com         | 216  |
| NHK    | www3.nhk.or.jp     | 139  |
| 日本経済新聞 | nikkei.com         | 79   |
| 時事通信社  | jiji.com           | 61   |
| スプートニク | jp.sputniknews.com | 43   |

表4 「ウイングアークフォーラム」のメディア集合

(a) メディア集合 1

| メディア名           | 識別子               | 被言及数 |
|-----------------|-------------------|------|
| 毎日新聞            | mainichi.jp       | 1025 |
| 東京新聞            | tokyo-np.co.jp    | 347  |
| 日本経済新聞          | nikkei.com        | 173  |
| ハーバー・ビジネス・オンライン | hbol.jp           | 95   |
| ハフポスト           | huffingtonpost.jp | 95   |

(b) メディア集合 2

| メディア名  | 識別子               | 被言及数 |
|--------|-------------------|------|
| 朝日新聞   | asahi.com         | 2895 |
| 産経新聞   | sankei.com        | 1889 |
| NHK    | www3.nhk.or.jp    | 1481 |
| リテラ    | lite-ra.com       | 883  |
| 日刊ゲンダイ | nikkan-gendai.com | 795  |

表5 「加計学園」のメディア集合

アの分極化と、それに伴うユーザの意見や議論の分極化への影響であるが、今回対象とした言及数上位のニュースは多くの人によって受け入れられやすい内容であることから、必ずしも適切ではないことがわかった。これは、Garimella らが分極化分析対象とするハッシュタグを選ぶ時に頻度に頼らずに人手で選

別していた理由と同じであると考えられる [13]. しかし、日々投稿される膨大なニュースとツイートの中から分極化の発生を迅速に発見できるようにするためには、分析対象を絞り込むことができる、統計値から簡単に算出できるような指標が必要である。

(a) クラスタ 1

| メディア名             | 識別子                          | 言及数 |
|-------------------|------------------------------|-----|
| ASCII.jp          | ascii.jp                     | 90  |
| AKIBA PC Hotline! | akiba-pc.watch.impress.co.jp | 73  |
| ITmedia           | itmedia.co.jp                | 40  |
| ログミー Biz          | logmi.jp                     | 35  |
| PC Watch          | pc.watch.impress.co.jp       | 30  |

(b) クラスタ 2

| メディア名  | 識別子                | 言及数 |
|--------|--------------------|-----|
| NHK    | www3.nhk.or.jp     | 59  |
| 時事通信社  | jiji.com           | 37  |
| 日本経済新聞 | nikkei.com         | 30  |
| スポーツニク | jp.sputniknews.com | 20  |
| 東京新聞   | tokyo-np.co.jp     | 11  |

表 6 「ウイングアークフォーラム」のユーザクラスタがツイートしたメディア

(a) クラスタ 1

| メディア名 | 識別子            | 言及数  |
|-------|----------------|------|
| 産経新聞  | sankei.com     | 1103 |
| 朝日新聞  | asahi.com      | 324  |
| 夕刊フジ  | zakzak.co.jp   | 233  |
| NHK   | www3.nhk.or.jp | 107  |
| 毎日新聞  | mainichi.jp    | 99   |

(b) クラスタ 2

| メディア名  | 識別子               | 言及数 |
|--------|-------------------|-----|
| 朝日新聞   | asahi.com         | 768 |
| リテラ    | lite-ra.com       | 670 |
| 日刊ゲンダイ | nikkan-gendai.com | 661 |
| NHK    | www3.nhk.or.jp    | 302 |
| 毎日新聞   | mainichi.jp       | 266 |

表 7 「加計学園」のユーザクラスタがツイートしたメディア

## 5 おわりに

本稿では、あるトピックに関するニュース記事とそのニュース記事を言及したツイートからメディア・ユーザ関係グラフを構築し、それから得られるメディアグラフとユーザグラフの分極度とクラスタ構造を分析することで、提案システムで任意のトピックにおける意見の偏りや対立を分析できることを確認した。ただし、今回は言及回数上位のトピックを分析したが、RWCの値が高い場合でもそれはメディアの特性の違いによるものであったことから、メディアの報道姿勢に相違や対立が発生しているトピックを発見する手法が必要であることがわかった。

今後の予定として、まず、分極化が発生しているようなニュースやトピックを、RWCよりも簡単に、例えば簡単な統計量で絞り込むことができる手法を検討する。また、今回はメディア・ユーザ関係グラフからメディアグラフとユーザグラフを別々に作成してから分析したが、2部グラフをそのまま処理し、現実のコミュニティ構造を何らかの形で反映できるような分極度を

検討する。さらに、今回のプロトタイプでは扱わなかったテキスト・時系列情報を用いて、クラスタ間の差異を明確に示す手法を追加する予定である。

## 謝 辞

本研究は JSPS 科研費 19H04421 の助成を受けた。

## 文 献

- [1] 東洋経済. ネット炎上「少数の意見」が社会の声に見える訳, 2021. <https://toyokeizai.net/articles/-/409471>.
- [2] Philip Ball. 'News' spreads faster and more widely when it's false. *Nature*, 03 2018.
- [3] 田中皓介, 藤井聡. 記者へのインタビュー調査に基づく公共事業を巡る報道バイアス生成要因の分析. *Policy and Practice*, Vol. 3, No. 2, 2017.
- [4] Eli Pariser. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. Penguin Books, 2011.
- [5] 笹原和俊. フェイクニュースを科学する: 拡散するデマ、陰謀論、プロパガンダのしくみ. 株式会社化学同人, 2019.
- [6] 張建偉, 河合由起子, 熊本忠彦, 白石優旗, 田中克己. 多様な印象に基づくニュースサイト報道傾向分析システム. 知能と情報, Vol. 25, No. 1, pp. 568-582, 2013.
- [7] 田中皓介, 神田佑亮, 藤井聡. 公共政策に関する大手新聞社報道についての時系列分析. 土木学会論文集 D3 (土木計画学), Vol. 69, No. 5, pp. I.373-I.379, 2013.
- [8] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the 2018 World Wide Web Conference*, p. 913-922, 2018.
- [9] 大前拓巳, 豊田正史, 張翔. マイクロブログにおけるユーザ間交流の分断および支持政党の偏りに基づく政治的分極化の分析. 第4回計算社会科学ワークショップ (CSSJ2020), 神戸, 2020.
- [10] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. Political polarization on twitter. In *Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2011.
- [11] Anirban Sen, Debanjan Ghatak, Kapil Kumar, Gurjeet Khanuja, Deepak Bansal, Mehak Gupta, Kumari Rekha, Saloni Bhogale, Priyamvada Trivedi, and Aaditeshwar Seth. Studying the discourse on economic policies in india using mass media, social media, and the parliamentary question hour data. In *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS '19*, p. 234-247, New York, NY, USA, 2019. Association for Computing Machinery.
- [12] 大倉俊平, 小野真吾. 関連記事判定のためのニュース記事キーワード抽出. 言語処理学会 第 24 回年次大会, pp. 1251-1254, 2018.
- [13] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Quantifying controversy on social media. *Transactions on Social Computing*, Vol. 1, No. 1, 2018.
- [14] George Karypis and Vipin Kumar. METIS - unstructured graph partitioning and sparse matrix ordering system, version 2.0. Technical report, August 1995.
- [15] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008, No. 10, p. P10008, October 2008.