

# 感情分析を用いたショック前後における会話ネットワーク上の コミュニティ変化の分析

Analysis of community change on conversation network before and after shock using emotion analysis

與島 仙太郎\*<sup>1</sup> 大知 正直\*<sup>1</sup> 浅谷 公威\*<sup>1</sup> 森 純一郎\*<sup>1</sup> 坂田 一郎\*<sup>1</sup>  
Sentaro Yojima Masanao Ochi Kimitaka Asatani Junichiro Mori Ichiro Sakata

東京大学工学部システム創成学科\*<sup>1</sup>  
Faculty of Engineering, The University of Tokyo

東京大学\*<sup>1</sup>  
The University of Tokyo

With the spread of social media, the movements of information became interactive. As a result, social media plays an important role in spreading information and forming opinions. However, sufficient research has not been done on the formation of such opinions and the process of emotional change. In this research, by using Twitter data of the 2016 US presidential election, how the size and density of the community divided by emotional degree polarity regarding a specific topic changes before and after the presidential election, and the polarity of the positive negative. The purpose of this study is to clarify how the conflict structure between communities with social change will change. We obtained the result that the confrontation structure of the positive community and the negative community will not talk about almost all fields analyzed in the experiment. This suggests that it is causing an Echo chamber.

## 1. はじめに

近年のソーシャルメディアの発達により、情報の発信は双方向化されている。従来は発信の手段や権力を持つメディアが一方向的に情報を発信していた状態であったが、ソーシャルメディアが普及することによって、不特定多数の一般大衆も情報の発信やリアルタイムな意見の発信を行うことが可能になった。このように、ソーシャルメディアにより情報発信が双方向化されたために、ソーシャルメディアは現在の社会において情報拡散や意見形成の面において重要な役割を果たすようになってきている。選挙活動でのソーシャルメディアの活用などは、その顕著な例である。重要度が高まるソーシャルメディアはしかし、その活用について問題点も存在する。デマや過激派の出現などによってレジリエンスが損なわれてしまう場合があるためである。実際にソーシャルメディア上での Echo chamber(意見の偏り)や, Silent Majority(無言の大多数)に関しては存在が確認されている [Goldie 2014][Mustafaraj 2011]。ソーシャルメディア上の意見などの分析の手法の一つとして、感情分析が広く用いられている。

本研究では、2016 アメリカ大統領選挙に関する 10/27-11/15 までの Twitter データを使用して、大統領選挙の前後において、特定話題に関する感情度極性で分けたコミュニティを分析する。コミュニティの規模や密度の変化に注目しながら、ポジティブネガティブの極性を持つコミュニティの間における会話量などを分析することにより、対立構造がどのように変化するかを分析する。それにより、世論調査には表しなかった意見の偏りや感情の実態の把握、及びアメリカの今後の政治に大きく関わるような対立構造の間でのコミュニケーションの変化について把握することを目的とする。

図 1.a ポジティブなユーザとネガティブなユーザのネットワーク

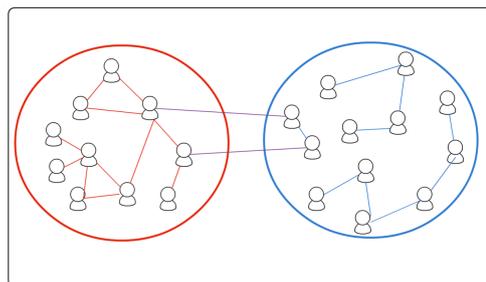


図 1.b ユーザがポジティブに偏っている場合のネットワーク

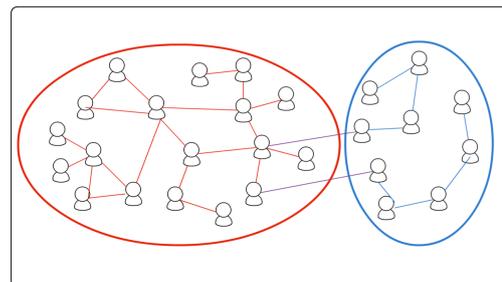
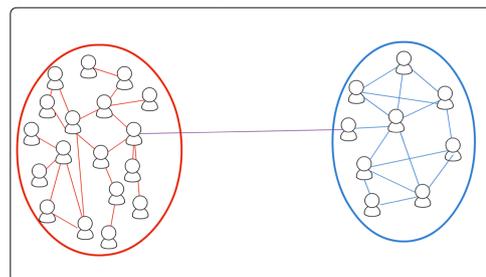


図 1.c 対立するコミュニティの間の会話が少ない場合のネットワーク



## 2. 手法

本研究における分析で使用する手法は以下の通りである。

- データの収集と前処理の実施
- ツイートに対する感情値の付与
- コミュニティの分割
- ネットワークの構築

それぞれについて内容を記述する。

### 2.1 データの収集と前処理

本研究では、大統領選の前後におけるネットワーク構造の変化について感情分析を用いて分析する。リアルタイム性が高く、感情の分析を行いやすいメディアとして、データの収集には Twitter のデータを使用した。

データの収集においては、“trump”、“hillary”、“election 2016”、“us president”、“republican”、“democratic”、“tax”、“job”、“immigration”、“foreign policy”、“trade policy”、“refugee”、“climate change”、“gun”、“abortion”、“childcare”を Twitter API の検索クエリとしてデータを収集し、重複のデータを削除した後、大統領選挙と関連のない投稿を多く含むものとして“tax”、“job”、“refugee”、“climate change”、“gun”、“abortion”、“childcare”のクエリを検索から除外した。

表 1. 使用したデータセット

呼称	期間	データ量
Before	10/27-11/6	869435
After	11/10-11/15	1626309

### 2.2 ツイートに対する感情値の付与

ツイートに対する感情値の付与に関しては、短文に関する感情値の割り当てに関して存在する vader という手法を用いて、ツイートに対して顔文字や感嘆符、スラングなども考慮した形で割り当てを行なった。

ツイートに対する感情値の付与は vader という手法を用いた上で、ポジティブ  $s_{pos}$  ネガティブ  $s_{neg}$  ニュートラル  $s_{neu}$  の合計値が 1 となるように付与[3]し、以下の式を用いて投稿の感情度を 0-1 に正規化した。

$$s_p = s_{pos} * 1 + s_{neu} * 0.5 + s_{neg} * 0 \quad (1)$$

ここで、 $s_p$  は投稿の感情度である。これにより、投稿は 0-1 の感情度を持つこととなり、それが 0.5 より大きいかどうかによってポジティブであるかネガティブであるかを判断する。

### 2.3 コミュニティの分割

本研究ではコミュニティの分割を二段階で行う。まず一つ目が、特定の語 (trump など) に言及しているかどうかである。あらかじめ設定した語に言及している投稿群を切り出し、その後、その投稿を行ったユーザの感情度によって、ポジティブなユーザ群とネガティブなユーザ群の二つに分割を行う。ユーザのポジティブネガティブの判定は、ユーザが特定の語に言及した書く投稿の感情度の平均として扱い、それが 0.5 より大きいユーザをポジティブユーザ、0.5 以下のユーザをネガティブユーザと判定した。

### 2.4 ネットワークの構築

ネットワークの構築は、特定の語に言及している投稿群について切り出した状態で行った。本研究のネットワークにおけるノードは投稿者で、エッジを @ つきの投稿として扱っている無向グラフである。また、エッジに対して属性を付与した。付与した属性は三つであり、ポジティブなユーザ内でのエッジを pos-pos のエッジ、ネガティブユーザ内でのエッジを neg-neg のエッジ、そしてポジティブユーザとネガティブユーザの間でのエッジを interactive なエッジとした。

## 3. 実験

### 3.1 分析した分野とそのクエリ

本研究では大統領選について、予め話題性のある分野を分析分野として、その語に言及しているかによって分割を行っている。今回設定した分野とそのクエリについては以下の通りである。

表 2. 分析対象の分野とクエリ

分野	クエリ
大統領と大統領候補	Hillary,Trump,Obama
選挙関連	republican,democratic,election
人種問題	white,black
外交	UK,immigrants,Russia

### 3.2 ランダムネットワークとの比の算出

本研究では、大統領選の前後において、ポジティブ内/ネガティブ内/ポジティブとネガティブにおける会話の量の変化を分析することで、大統領選の前後においてどのようなコミュニケーションの変化が起こったかを分析する。しかし、大統領選の前後においてネットワークに関するノードやエッジの数は変化しているため、純粋な数によって比較することができない。そこで本研究では、所与のノードとエッジに対してランダムにネットワークを貼った場合との比を大統領選の前後それぞれについて算出し、その値を用いてコミュニケーションの変化を評価した。

## 4. 結果と考察

### 4.1 コミュニティ規模の変化

コミュニティ規模とその密度に関する変化は、コミュニティの規模変化を表す、ポジティブコミュニティやネガティブコミュニティの占める割合とそれぞれのコミュニティ内での会話の活性度を表す密度を用いて分析を行った。結果として各分野に対しては規模の変化は大きく逆転するものから全く割合を変えないものまで幅広く存在したが、密度に関しては小さくなる傾向が強かった。規模を変更するものについては、トランプに対するポジティブユーザや白人に対する投稿数の増加など、結果が出ることによって自分の立場を発言しやすくするようなものも存在していた。

例としてトランプに関するネットワーク密度の表を以下  $s$  に記す。どちらの感情極性のもも密度が下がっており、会話の活発さは失われているとわかる。

表 3. トランプに言及しているネットワークのコミュニティ内での密度

時期	Pos	Neg
Before	0.19	0.36
After	0.08	0.06

#### 4.2 対立構造におけるコミュニケーション量の変化分析

対立構造についてのコミュニケーション量を、大統領選の前後で比較するために、同数のエッジがランダムに貼られた場合の期待エッジ数との比を取ることで比較した。

表 4. トランプに言及しているコミュニティ内やコミュニティ間の会話量における、ランダムにネットワークを貼った場合との比

時期	Pos	Neg	Interactive
Before	0.70	2.57	0.89
After	1.35	1.31	0.65

例えば上のトランプにおけるランダムネットワークとの比を見てみると、ポジティブなものの割合が増加していることから、これはトランプの勝利に合わせてポジティブな中でコミュニケーションが増加していることの表れである可能性が示唆される。またネガティブなものの半減は、悲嘆の際などには会話が活発に起こるわけではなく少数で悲嘆にくれている、という可能性が示唆される。そして、トランプに限らずほぼ全ての分野において小さくなっていったのが、ポジティブなコミュニティとネガティブなコミュニティの間における会話である、インタラクティブなもの比率である。

インタラクティブなもの比率が減少しているということは、感情極性の分かれるポジティブとネガティブのコミュニティに関して、対立構造の間でのコミュニケーションが減り、距離が遠ざかっていると言える。

ほとんどの分野については、このインタラクティブなものが増加しているものが存在した。それが、イギリスに関するものや移民に関するものなどであった。

表 4. イギリスに言及しているコミュニティ内やコミュニティ間の会話量における、ランダムにネットワークを貼った場合との比

時期	Pos	Neg	Interactive
Before	1.79	1.51	0.27
After	1.57	1.40	0.47

表 5. 移民に言及しているコミュニティ内やコミュニティ間の会話量における、ランダムにネットワークを貼った場合との比

時期	Pos	Neg	Interactive
Before	1.37	1.47	0.57
After	1.22	1.34	0.72

イギリスや移民に関するデータにおいては、インタラクティブなコミュニケーションがむしろおおきくなっているが、

それはイギリス人などの英語話者であるが当事者でない者による議論が活発化されているからであると考えられる。

## 5. 結論

本研究では、大統領選の前後について大きく二点に分けて分析を行った。一点目は極性を持つコミュニティの規模や密度などの問題、もう一点が極性を持つコミュニティ間での対立構造の変化についての分析である。本研究の結論として、大統領選前後では、ほぼ全ての分野においてポジティブの感情を持つコミュニティとネガティブの感情を持つコミュニティの間のコミュニケーション比は小さくなり、対立構造の間の分断が深まっていることが示された。対立構造の分断の深まらない少数の例外については、イギリスに関するユーザ群や移民に関するユーザ軍など、外交関係などで当事者でないユーザが多く含まれることが推測される分野に関してであった。コミュニティの規模については、結果がでなければ発言しにくいトピックなどに関して特に大きな変化が見られた。コミュニティ規模の大きな変化の一因には、選挙の前においては発言をすることのなかった Silent Majority が存在していた可能性がある。

対立するコミュニティの分断によっては引き起こされる問題としては、ソーシャルメディア上で自らと同じような意見としか触れなくなり意見の偏りが起こるという Echo chamber の問題が発生することが考えられる。

以上のように、本研究では、大規模なショックとしてアメリカ大統領選挙を取り上げ、その前後におけるコミュニティ規模・密度の変化と、コミュニティ間の対立構造の深まりについて分析を行い、Twitter の感情分析を用いることで意見の偏りなどの状況を示唆する結果を得た。

今後の展望としては、本研究においてはネットワーク構造の変化について大きな変化を見たが、細かい分布の濃度やさらに小さいクラスターの変化に関する分析を行っていないので、より進んだネットワーク分析を行いその変化を観測する必要がある。

## 参考文献

- [Goldie 2014] Goldie, David, et al: "Using bibliometric and social media analyses to explore the "echo chamber" hypothesis." Educational Policy 28.2 (2014): 281-305.
- [Mustafaraj 2011] Mustafaraj, Eni, et al: "Vocal minority versus silent majority: Discovering the opi- onions of the long tail." Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on. IEEE, 2011.
- [Gillbert 2014] Gilbert, CJ Hutto Eric: "VADER: A Parsimonious Rule-based Model for Senti- ment Analysis of Social Media Text." Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) <http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf>. 2014.