

Twitter におけるリツイートによる情報伝播とユーザー特性の分析

川村健二[†] 奈良光紀[‡]
[§]岩手大学大学院 総合科学研究科

1 はじめに

近年, SNS, 特にも Twitter の利用者は爆発的に増加しており, そこで多くの情報が投稿され, 拡散・共有されている. Twitter に関して数多くの研究が行われている. 山下ら [1] はリツイート数の時系列変化に関する研究を行い, 芋づる式にリツイート数が増加することを示唆した. フォロワー数とフォローワー数の対称性・非対称性に関しては, 小出ら [2] と木島ら [3] の研究がある.

本研究では Twitter において, 大きく拡散されたツイートのリツイート数の時系列変化と, ユーザー情報を用いたネットワーク構造の分析を行った. ツイートの収集やユーザー情報の取得は TwitterAPI を使用した.

2 リツイート数の時系列変化

2.1 ツイートの取得

Twitter 上で拡散されているリツイート数が多いツイートを事前に選定・取得した. SearchMethods を用いて, ツイートのテキスト本文の検索をかけることで 72 時間分の全件収集を行った. ただし, 非公開アカウントのツイートは収集できないため今回の調査対象からは除いた. 収集したツイートはウケを狙ったネタ, 告知や時事速報などのニュース, 企業などが広報を目的として行うキャンペーン, 上記 3 つに分類されない呟きの 4 種類に分類して, それぞれ 5 ツイートずつ計 20 ツイートを選定した. なお, リツイート数が多いツイートは TwTimez[4] を参考に選定した.

2.2 リツイート数の時系列変化とリツイート率

図 1 はネタ, ニュース, キャンペーンの 3 つのカテゴリのリツイート数がどのように変化していくかを 5 分ごとに集計したものである. 青線はリツイート数を示している. 黄線はリツイートしたユーザー全員のフォロワー数の合計であり, 元のツイートを何人が受け取ったかを示している. ネタツイートでは 48 時間程度でリツイート数が増加し, その後徐々に減少している. 早朝から昼ごろにかけてがリツイートされやすく, これは既存の調査結果 [5] と同様の傾向が示されている. 告知のツイ

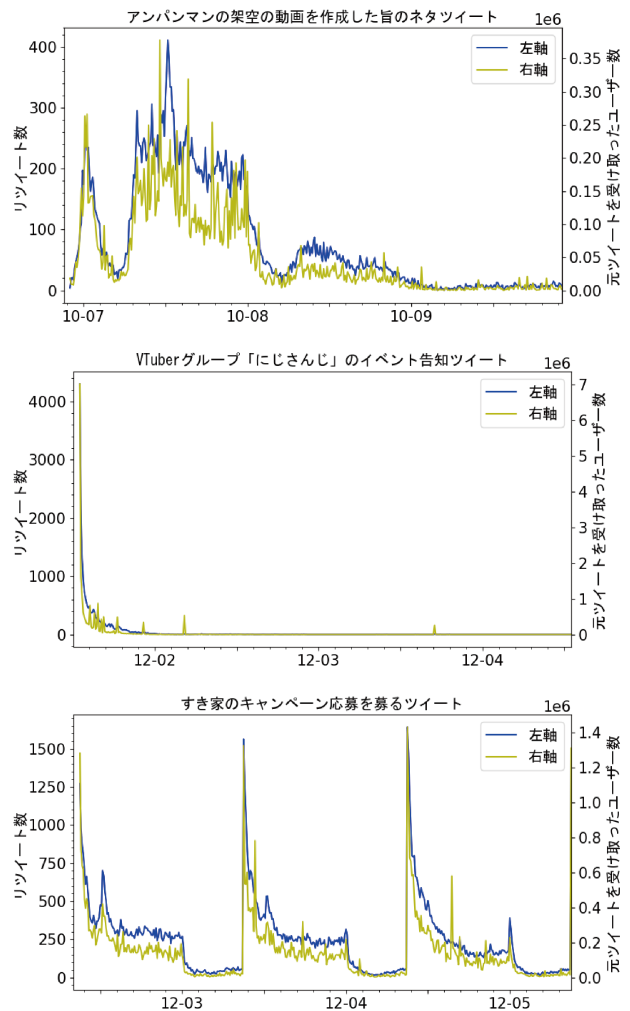


図 1 5 分当たりのリツイート数の時間変化

イトはツイートされた直後が一番多くリツイートされ, 半日以内の比較的短い時間で急速にリツイート数は減少する. キャンペーンツイートは定期的なツイートしたユーザー自身がリツイートを行うことによって, 周期的にリツイート数が増減するパターンをみせている. これは企業などが広く拡散させることを目的に行っていることが考えられる. キャンペーンツイート以外は 72 時間以内にリツイートされにくくなり終息する傾向がみられた.

図 2 は 72 時間経過時点でのリツイート率のヒストグラムである.

$$\text{リツイート率} = \frac{\text{リツイート数の合計}}{\text{元のツイートを受け取った人数の合計}}$$

Information Propagation and Characteristics of Users by Retweeting on Twitter

[†] Kenji KAWAMURA

[‡] Mitsunori NARA

[§] Graduate School of Arts and Science, Iwate University

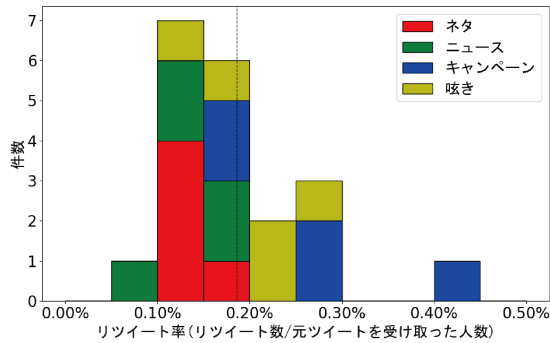


図2 リツイート率のヒストグラム

と定義した. リツイート率は 0.1~0.3% 程度の範囲に収まっている. 内容によって大きくリツイート率が変わらないことがわかった.

3 フォロー関係に基づくネットワーク構造

3.1 ランダムサンプリングによるユーザー情報の収集

ユーザー情報はランダムウォークサンプリングを用いて, UserObject として収集した. 600 人のユーザー情報を集め, 相互フォロー率などを計算し散布図 (図 3) を作成した. 本研究では,

$$\text{相互フォロー率} = \frac{\text{相互フォロー数}}{\text{フォロー数}}$$

と定義した. 相互フォロー数は whotwi[6] を参照し, 相互フォロー数が測定不能だったユーザーに関しては灰色の点で図 3 にプロットした.

3.2 ネットワーク構造の対称性と非対称性

図 3 においてフォロー数=フォロワー数の直線付近にユーザーが集中していることが見てとれる. 実際にこのようなユーザーは相互フォロー率が高く, 対称性を意識して利用しているということが考えられる. その一方でインフルエンサーや主にフォローのみを行うユーザーなど相互フォロー率が低く, 非対称なネットワークを構成しているユーザーも存在している. この結果から, Twitter のネットワーク構造は対称性のあるネットワークと非対称性のあるネットワークが重なり合ったような構造になっていると考えられる.

3.3 ユーザー情報に基づくクラスタリング

図 4 (左) はフォロー数・フォロワー数・相互フォロー数の 3 つの特徴量に基づき, scikit-learn の AgglomerativeClustering で階層型クラスタリングを行った結果である. 2 つにクラスタリングした場合, デンドログラムの左半分が図 4 (右) のオレンジの点となり, 右半分が青の点となる. 左半分側のユーザーは対称性を意識しているクラスターであり, 右半分側のユーザーは対称性を意識していないクラスターに対応している. 相互フォロー率を考慮することで, ネットワークの対称性を考慮したクラスタリングが行うことができた.

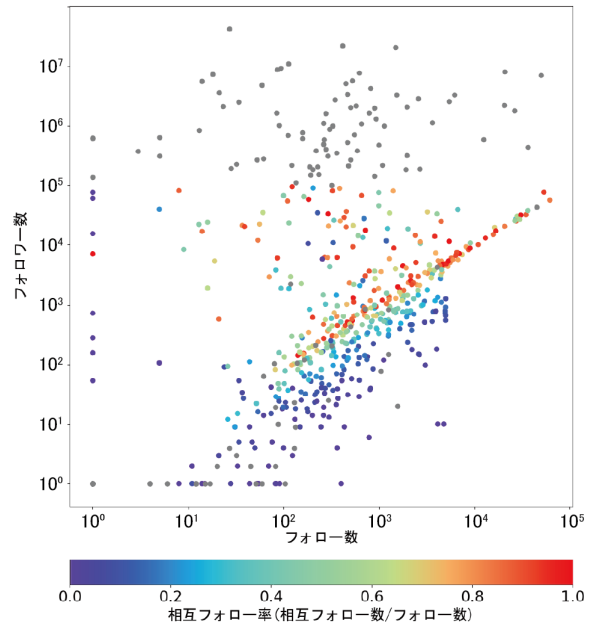


図3 フォロー数・フォロワー数・相互フォロー率

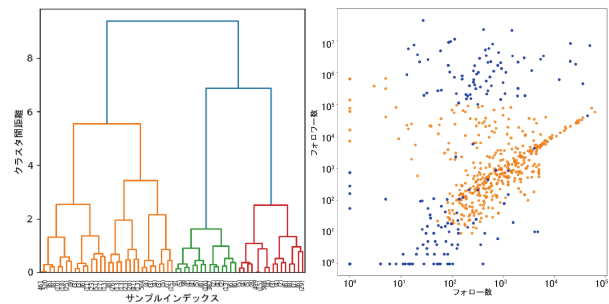


図4 デンドログラムとクラスタリング結果

4 おわりに

本研究ではリツイート数の時系列変化を計測した. その結果, ツイート内容により特有の情報伝播のパターンがあることが分かった. また, ユーザー情報を用いたネットワーク構造の分析を行い, フォロー数・フォロワー数・相互フォロー率に基づくクラスタリングを行い, 各クラスタの特徴づけを行った. 今後の課題としては, リツイート数の推移とネットワーク構造の関係性を明らかにすることが挙げられる.

参考文献

- [1] 山下玲子, 三浦麻子. おもしろツイートはいかに広まったか: 事例研究による「じわる」プロセスの解明. メディア・情報・コミュニケーション研究. 2018 年
- [2] 小出明弘, 斎藤和巳, 風間一洋, 鳥海不二夫. 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. 2013 年
- [3] 木島裕希, 高橋真吾. 入出次数相関と相互リンク率を考慮した Twitter を表現した成長ネットワークモデル. SICPE 第 8 回社会システム部会研究会. 2015 年
- [4] "TwTimez". <http://www.twtimez.net/>
- [5] "データからみるTwitter ユーザー実態まとめ". <https://service.aainc.co.jp/product/echoes/voices/0014>
- [6] "whotwi". <https://ja.whotwi.com/>