

Twitter 上の情報拡散の分析とその利用

富山県立大学工学部電子・情報工学科
1515010 大谷和樹

指導教員：奥原浩之

1 はじめに

近年、Twitter や Instagram, Facebook などのソーシャルネットワークサービス (SNS) が発展し、情報の広がる速度が爆発的に加速している。そこで、SNS の中でも Twitter に着目し、Twitter 社が公開している API を利用して、ある出来事に関係するツイートをキーワードを設定して集め、集めたツイートを利用して情報の拡散の仕方を調べる。その後、その拡散の仕方を利用する何らかの方法を検討する。

2 先行研究

2.1 うわさの伝播モデル

蜷川らの研究 [1] では、構成要素の間にランダムに張り巡らされたネットワーク上を情報が伝播するモデルを提案し、最初に一つの構成要素に情報を与えた場合に、情報が集団内でどの程度ひろがるのかを計算機シミュレーションを用いた確率的な手法によって調べている。そこで示されている伝播モデルは以下のようなものである。

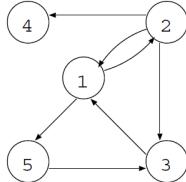


図1. 頂点の数 $N = 5$ の場合のうわさの伝播モデルの有向グラフによる表現

図1は、頂点の数 $N = 5$ の場合のうわさの伝播モデルの例を示す。この例では頂点1にとって頂点2および5頂点はうわさを話す関係にあることを示している。うわさの伝播モデルは離散時間で変化するものとし、ステップが進むにつれ次のように変化する。(1)ステップ0で、無作為に選んだ頂点(たとえば頂点 i とする)の状態を1とし、それ以外の頂点の状態を0とする。これは、人物 i がうわさの発生源となることを表している。(2)次のステップで、頂点 i から隣接しているすべての頂点、すなわち頂点 i から出るすべての弧の終点の状態を1にする。これは人物 i が自分の知人にうわさを話すことに相当する。(3)以降のステップでは、状態1の頂点から隣接するすべての頂点の状態を1にするという動作を繰り返す。これはうわさを聞いた知人がさらにそれぞれの知人にうわさを伝える行動に相当する。このように、うわさの伝播モデルは各頂点の状態が離散時間で変化するループのない有向グラフとして定義される。ステップ t における頂点 j の状態を $aj(t)$ で表し、ステップ t における状態が1の頂点の数を伝播数 $D(t)$ とよび、式(1)で定義する。

$$\begin{aligned} D(t) &= \sum_{i=1}^n \theta(a_i(t)), \\ \theta(x) &= \begin{cases} 0 & x \leq 0 \\ 1 & x > 0. \end{cases} \end{aligned} \quad (1)$$

うわさの伝播モデルではいったん状態が1となった頂点は状態0に戻ることはないので、明らかに伝播数 $D(t)$ はステップ t に対して非減少関数となる。また、全頂点数 N が有限であることから、伝播数 $D(t)$ はあるステップ Ts 以降は一定値をとる。このとき、ステップ Ts で系は定常状態になったと見なし、その時点でシミュレーションを停止するものとする。

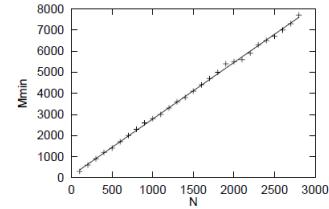


図2. 頂点数 N と非拡散相から拡散相へと転移する最小の弧の数 M_{min} の関係

その後シミュレーションとして、頂点数 N を 100, 200, …, 2800 と 100 ごとに変化させ、それぞれの N について弧の数 M を 100 ごとに変化させた場合に、非拡散相から拡散相に転移する最小の弧の数 M_{min} を求めた。

図 2 から頂点数 N と拡散相に転移する最小の弧の数 M_{min} は式 (2) に示すように比例関係にあることが分かる.

$$M_{min} \simeq \alpha N + \beta \quad (2)$$

式(2)の比例定数 α は 1 頂点あたりの平均出次数にほかならない。最小 2 乗法によると $\alpha = 2.678$, $\beta = 102.381$ となった。この結果より、1 頂点当たりの平均の弧の数は全頂点数 N に依存せず、ほぼ 3 であることがわかった。この結果は、集団を構成する人物が平均して 3 人にうわさを伝えると、そのうわさはほぼ全員に伝わることを意味している。

2.2 ソーシャルメディア上の大規模情報拡散に関する俯瞰的可視化趨向の提案

榎らの研究 [2] では、Twiiter 上の情報拡散を人間が直観的に理解可能な形で可視化する手法を提案している。

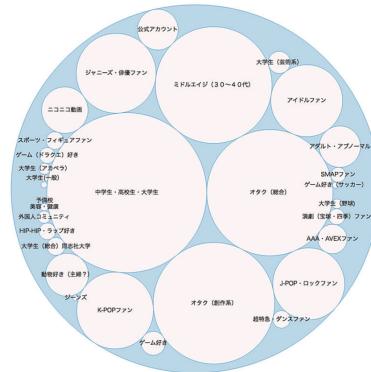


図 3. 第 1 階層の俯瞰図

図3は、第1階層を表した俯瞰図である。図3中の円が一つのコミュニティを表す。また円のサイズは、コミュニティに含まれるユーザ数により相対的に決定される。コミュニティに属するユーザ数が多いほどサイズが大きくなる。また丸ごとにラベルが付与されている。

図4は、第5階層を表した俯瞰図である。第1階層で分かれていたコミュニティの中でもさらに細分化されたコミュニティがあることがわかる。榎らは、Twitter俯瞰図が階層化されているメリットは、全体を俯瞰したければ第1階層の俯瞰図を見て、よ

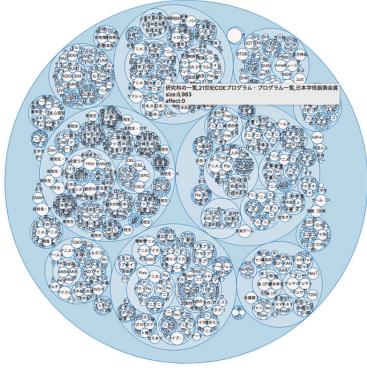


図 4. 第 5 階層の俯瞰図

り細部を見たければ第 2 階層、第 3 階層と下の層の俯瞰図を見るというように、全体的な俯瞰とより細部の俯瞰が両方実現できることであるとしている。

ここで行われているクラスタリングには、社会ネットワークのクラスタリングに用いられる Modularity-Optimization の手法を採用する、Modularity Q とはクラスタの結合度合いを表す指標であり、式 (3) で表される、これが高ければ高いほどネットワーク全体のよくクラスタリングされていると考えられる。ここでは、既存研究で使われていた Louvain 法を用いる。式 (3) において、 ei, j は「コミュニティ i, j に属するノード同士が繋がるリンク数の全リンク数に占める割合」、 $a2i$ は ei, i の期待値を表す。

$$Q = \sum_i (e_{i,i} - a2i^2) \quad (3)$$

このクラスタリングにより得られるクラスタをコミュニティとして抽出する。

2.3 Twitter ネットワークにおけるデマ拡散とデマ拡散防止モデルの推定

白井らの研究 [3] では、デマ情報とそのデマの訂正情報を病気とみなし、感染症疾患の伝染モデル (SIR モデル) を拡張してデマ情報・訂正情報の拡散モデルを構築する。次に、特に情報の拡散スピードが速いといわれている Twitter に焦点を当て、ユーザーが投稿した文章 (ツイート) の情報を収集し、実際に拡散したデマ情報の拡散の様子を調べ、デマ情報・訂正情報拡散モデルによるシミュレーションとの比較を行っている。

3 進捗状況 3.1 TwitterAPI

今回、Twitter 社が提供している API を利用してツイートを集めることを検討している。TwitterAPI を利用するためには Developer 用のカウントを作って申請を行わなくてはいけなかつたので、参考サイト [4] を見ながら申請を進めた。申請が完了すると自身の APP を作ることができるようになるので、作成する。作成が終わると、図 5 のような「Consumer API Keys」、「アクセストークン情報」が入手できる。これを利用することでユーザーが APP から API を通じて Twitter アカウントにアクセスできるようになる。

3.2 ツイートの取得

API が利用できるようになったので、実際にツイートを取得できるのか試してみた。参考サイト [5] の情報をを利用して、今回は Python の Tweepy というライブラリを使うことにした。このライブラリを使うことによって、Python のプログラムでアカウントを操作できる。ツイートを投稿する場合は、

```
>>> api.update_status(status='Hello!')
```

と入力する。実際に投稿されたツイートが図 6 である。

次に、キーワードでツイートを検索するには、

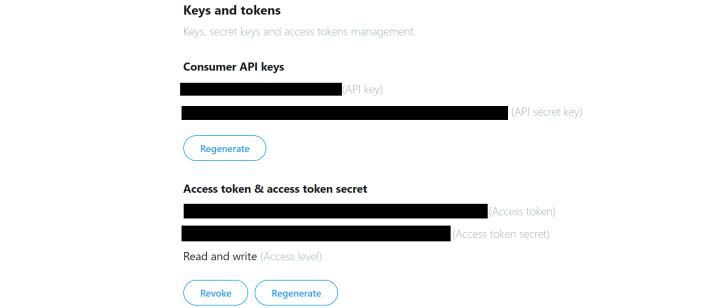


図 5. 入手した Consumer API Keys とアクセストークン情報



図 6. Tweepy での投稿結果

```
>>> search_result = api.search(q='Arashi')
>>> for result in search_result:
...     print result.text
```

と入力する。実際の検索結果が図 7 である。



図 7. キーワード検索結果

4 おわりに

今回は情報拡散についての論文、特に Twitter の情報を利用している論文の紹介と、実際に TwitterAPI を利用してキーワードでツイートを検索する方法を実際に試した結果を述べた。今後は、まだ関連研究についての情報が足らないので、論文の収集を行っていく。また、今回使用したプログラムでは検索結果のツイートは 10 件しか表示できていないため、もっと多数のツイートを検索するプログラムの作成と検索結果を別のファイルに保存する方式を考えていく。

参考文献

- [1] 蟹川繁, 津田伸生, 服部進実ほか. うわさの伝播モデル. 情報処理学会論文誌, Vol. 41, No. 2, pp. 517–520, 2000.
- [2] 楠剛史, 鳥海不二夫, 大知正直. ソーシャルメディア上の大規模情報拡散に関する俯瞰的可視化手法の提案. 人工知能学会全国大会論文集 一般社団法人 人工知能学会, pp. 2D4OS1a01–2D4OS1a01. 一般社団法人 人工知能学会, 2019.
- [3] 白井嵩士, 楠剛史, 鳥海不二夫, 篠田孝祐, 風間一洋, 野田五十樹, 沼尾正行, 栗原聰ほか. Twitter ネットワークにおけるデマ拡散とデマ拡散防止モデルの推定. SIG-DOCMAS, No. B102, 2012.
- [4] <https://qiita.com/kngsym2018/items/2524d21455aac111cdee>
- [5] <https://blog.statsbeginner.net/entry/2015/10/21/131717>.