

October 31, 2019

大谷 和樹

富山県立大学 情報基盤工学講座

1. はじめに
2. 先行研究
3. 進捗状況
4. おわりに

October 31, 2019

背景

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. この状況は, 有用な情報の拡散だけでなく, 間違った情報 (デマ情報) の拡散を手助けする要因にもなっている.

目的

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.

蛭川らの研究では，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．

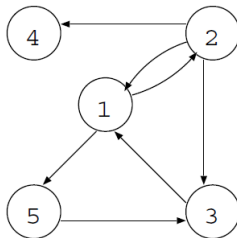


Figure 1. うわさの伝播モデルの有向グラフによる表現

図 1 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．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．

ステップの変化

- (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 $a_j(t)$ 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.

伝播数

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

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 T_s 以降は一定値をとる.
このとき、ステップ T_s 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.

Twitter 上の情報
拡散の分析とその
利用

はじめに

先行研究

進捗状況

おわりに

頂点数 N を 100, 200, \dots , 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。

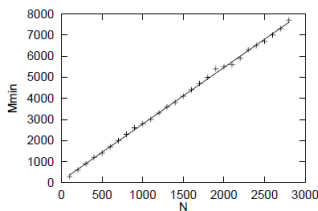


Figure 2. 頂点数 N と非拡散相から拡散相へと転移する最小の弧の数 M_{min} の関係

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.

N と M_{min} の関係

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

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 頂点あ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 人にうわさを伝えと, そのうわさはほぼ全員に伝わることを意味している.

本研究では、Twitter 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。

[illegible]

図3は、第1階層を表した俯瞰図である。図2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。

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趣向の提案

9/15

Twitter 上の情報
拡散の分析とその
利用

はじめに
先行研究
進捗状況
おわりに

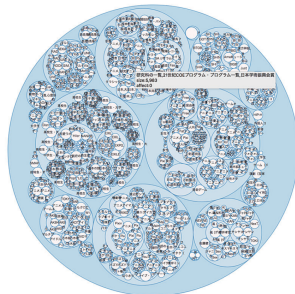


Figure 4. 第 5 階層の俯瞰図

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。

本研究は、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。

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趣向の提案

10/15

Twitter 上の情報
拡散の分析とその
利用

はじめに

先行研究

進捗状況

おわりに

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が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、 $e_{i,j}$ 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」、 a_{2i} は $e_{i,i}$ の期待値を表す。

Modularity Q: クラスタの結合度合いを表す指標

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

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。

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

11/15

Twitter 上の情報
拡散の分析とその
利用

はじめに

先行研究

進捗状況

おわりに

白井らの研究では、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。

SIR モデル

ある集団に属する人を、まだ病気にかかっていない人（S : Suseptible），病気にかかった人（I : Infectious），病気が治って免疫を得た人（R : Recovered）の 3 種類の分類し、時間の経過による S, I, R の人数の変化を表した式。

今回、Twitter 社が提供している API を利用してツイートを集めてく
ることを検討している。TwitterAPI を利用するためには Developer 用の
アカウントを作って申請を行わなくてはならなかった。なので、調べな
がら申請を進めた。

TwitterAPI

TwitterAPI とは、ツイートやタイムラインの取得、リツイートやいいね
といった Twitter のサービスを、公式のウェブサイトを経由せず
に直接利用できるサービスのこと。

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。

Keys and tokens

Keys, secret keys and access tokens management.

Consumer API keys

[Redacted] (API key)

[Redacted] (API secret key)

Regenerate

Access token & access token secret

[Redacted] (Access token)

[Redacted] (Access token secret)

Read and write (Access level)

Revoke

Regenerate

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

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．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．キーワードでツイートを検索した結果が以下の図 6 である．

```
In [8]: search_result = api.search(q='東京')
for result in search_result:
    print (result.text)
```

まだ選挙案が通って無いのにこの報道はすんねわ(´・ω・´) <https://t.co/5422w0d91A>
 高防FCサイトも閉じないんですけどー笑
 RT @mizus_: あんさんぶるスターズ あんスタ 生プロマイド 生プロ

【東】スライ、鹿、弓弦、紀晴、竜人

【東】同僚美智、涉ノ満太郎 泰生、美智

報道での空想を希望しております。また、総てのお取引を優先させて頂きます。検索からも気
 味にお買がけください。http://
 RT @liveda: そして、コササ
 嵐のサカサカ紀晴等
 で紀晴等の一人居まった道場に、総てを下げはじめた奴ら
 ははは、お送りさん！コササで——ずん！！！！！ <https://t.co/gd01vR8AT>
 この間の地盤よりひどいぞ！
 くらっ！お送りさん！
 もう突出ストップして欲しいなm(´_`u´) 3ty
 あと何位増えるのか
 RT @news1: 掲載: 12月号) 嵐の表紙画像公開 18年初の表紙を完全再現
<https://t.co/ao2h997u6>

藤野 E 18
<https://t.co/ao2h997u6>
 RT @news1: フラガ狂のオリコン。嵐さんと並んで欲しい！ありがとうございます <http://t.co/7b02d9k9m>
 RT @liveda: そして、コササ
 嵐のサカサカ紀晴等
 で紀晴等の一人居まった道場に、総てを下げはじめた奴ら
 ははは、お送りさん！コササで——ずん！！！！！ <https://t.co/gd01vR8AT>
 RT @pan_718: 嵐 新 写真集の発売するとは
 RT @abanynewsinfo: 11/7, 14 VS嵐にジャニーズVS嵐
 11/7, 14 VS嵐にジャニーズVS嵐

Figure 6. キーワード検索結果

まとめ

- ① 卒業研究に取り掛かり始めた．今回はほぼ関連論文の紹介になったが，これから時間を見つけて進めていきたいと思う．

今後の課題

- 1 関連研究の論文を探して読む
- 2 ツイートを集めるプログラムの作成 (複数のキーワードで検索，取得するツイート数を増やす，取得したテキストデータを別のファイルに保存など)
- 3 研究の最終目的を定める