

# 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

## Estimation of False Rumor Diffusion Model and Prevention Model of False Rumor Diffusion on Twitter

白井 嵩士<sup>1\*</sup> 榊 剛史<sup>2</sup> 鳥海 不二夫<sup>3</sup> 篠田 孝祐<sup>4</sup>  
風間 一洋<sup>5</sup> 野田 五十樹<sup>6</sup> 沼尾 正行<sup>7</sup> 栗原 聡<sup>7,8</sup>

Takashi Shirai<sup>1</sup>, Takeshi Sakaki<sup>2</sup>, Fujio Toriumi<sup>3</sup>, Kosuke Shinoda<sup>4</sup>,  
Kazuhiro Kazama<sup>5</sup>, Itsuki Noda<sup>6</sup>, Masayuki Numao<sup>7</sup> and Satoshi Kurihara<sup>7,8</sup>

<sup>1</sup> 大阪大学大学院 情報科学研究科

<sup>1</sup> Graduate School of Information Science and Technology, Osaka University

<sup>2</sup> 東京大学

<sup>2</sup> Tokyo University

<sup>3</sup> 名古屋大学

<sup>3</sup> Nagoya University

<sup>4</sup> 理化学研究所

<sup>4</sup> RIKEN

<sup>5</sup> NTT 未来ネット研究所

<sup>5</sup> Network Innovation Laboratories

<sup>6</sup> 産業技術総合研究所

<sup>6</sup> National Institute of Advanced Industrial Science and Technology

<sup>7</sup> 大阪大学 産業科学研究科

<sup>7</sup> The Institute of Scientific and Industrial Research, Osaka University

<sup>8</sup> 科学技術振興機構

<sup>8</sup> JST CREST

**Abstract:** Twitter is a famous social networking service and has received attention recently. Twitter user have increased rapidly, and many users exchange information. When 2011 Tohoku earthquake and tsunami happened, people were able to obtain information from social networking service. Though Twitter played the important role, one of the problem of Twitter, a false rumor diffusion, was pointed out. In this research, we focus on a false rumor diffusion. We propose a information diffusion model based on SIR model, and discuss how to prevent a false rumor diffusion.

## 1 はじめに

現在インターネットは、テレビや新聞といった既存のメディアと並んで積極的に情報収集のツールとして使用されている [2]。さらにインターネットの発展に伴い、コミュニケーション機能を高めたソーシャル・ネットワークキング・サービス (SNS) と呼ばれるサービスが近年登場し、話題を呼んでいる。このようなサービス

では知人や友人らと繋がりを持つことにより社会ネットワークを構築し、情報の共有などを行うことができるのが特徴で、Facebook<sup>1</sup>や Twitter<sup>2</sup>といったサービスがよく知られている。一般の個人であっても簡単に情報を投稿することができ、さらに簡単に情報を収集できることから、こうした SNS を様々な形で活用する動きがある。例えば 2011 年 3 月 11 日に発生した東日本大震災では、携帯電話が通じない中で、Twitter が

\*連絡先：大阪大学産業科学研究科沼尾研究室  
〒 567-0047 大阪府茨木市美穂ヶ丘 8-1  
E-mail: shirai@ai.sanken.osaka-u.ac.jp

<sup>1</sup><http://www.facebook.com/>

<sup>2</sup><http://twitter.com/>

安否確認などの情報を伝達、収集するツールとして機能したことは記憶に新しい [9]. 以上のように, SNS は人々の生活の中で欠かせないものとなりつつある.

しかしながら, 各種 SNS のサービスが積極的に活用される中で, いくつかの問題点も指摘されている. そのひとつとして挙げられるのが, デマ情報 (本論文ではデマを「根拠が無く, 後に誤りを指摘する内容の情報が発表された情報」と定義する) の拡散である. SNS はユーザーが多いことなどから, 情報が急速に広範囲に拡散する. これはデマ情報に関しても同様であり, しばしばデマ情報が拡散してしまい, 社会問題になることがある. 東日本大震災時に緊急の情報を発信する手段として活躍した Twitter であるが, 大震災直後の混乱した状況とも相まって, 非常に多くのデマ情報が広がってしまうこととなった. 被災地への支援に関連する情報や, 非常事態に身を守る方法に関連する情報である場合があり, デマであると気づかずに, 情報を拡散してしまうユーザーが見受けられた.

このような背景において, 本研究では SNS にて拡散するデマ情報の拡散の様子のモデル化と, 拡散を収束させるための方策の検討を行う. まず, 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う. また, 訂正情報の拡散の速度に着目し, 早く訂正情報を拡散させることによるデマ拡散収束方法の検討を行う.

## 2 関連研究

近年, SNS, 特に Twitter を対象とした研究が盛んに行われるようになった.

Twitter を対象とした研究のひとつに, Twitter のユーザーによって構築されるネットワークの特徴を分析した研究がある. Huberman らはフォロワーとツイート投稿数の関係を調べるだけでなく, Twitter 上で 2 回以上直接やり取りを行った (通常ツイートはフォロワー全体に発信されるが, 「@ユーザー名」の識別子をつけることで, 特定のユーザーにのみツイートを発信できる) 相手ユーザーを知り合いであるとみなし, 知り合いの数とフォロワーの関係や, 知り合いの数とツイート投稿数の関係も調べており, ユーザーのフォロワーの多さが知り合いの多さにつながるとは限らないと結論付けている [5]. Akioka らは日本の Twitter ユーザーを対象に, フォロー数やフォロワー数の分布を調べた

[1]. また, Akioka らはこれに加えて Twitter ユーザーの増加の原因と, Twitter のサービスの拡張や他メディアでの Twitter の露出との関連を調べている. Kawak らは Twitter ネットワークとツイート, リツイートの関係について調べている [7]. Kawak らは他メディアのネットワークと比較し, Twitter ネットワークが他の社会ネットワークとは異なる特徴を持つことを示した. 鳥海らは東日本大震災前後に投稿されたツイートから, 直接やり取りを行ったユーザーのネットワークを作成し, 震災の前後でネットワークがどのように変化したかを調べた [10].

情報拡散に関する研究も行われている. Weng らは情報拡散に寄与する Twitter ユーザーの識別方法として, 従来の重要度を測るアルゴリズムである PageRank を拡張した, Twitter ネットワークにてしばしば見られる同類性を考慮した TwitterRank を提案している [12]. Balshy らは口コミなどの商業的観点から URL を含むツイートの拡散に着目し, 拡散の起点となったユーザーの特徴やツイートに含まれる URL のリンク先の内容と, 情報が拡散する範囲の関係などについて調べている [3].

Twitter 上を流れる情報の信頼性についても研究が行われている. Castillo らは, 投稿されたツイートの長さや主題, ツイートを投稿した各種ユーザーの特徴から, ツイートの内容の信頼度を算出する手法を提案している [4]. 梅島らはデマツイートに見られる傾向を把握するため, 東日本大震災時に多くのユーザーにリツイートされたツイートを分析し, デマツイートが持つ特徴を調べている [11].

このように Twitter を対象とした研究が数多く行われている. しかしながらデマツイートの拡散に着目すると, Castillo らや梅島らによれば完全にデマかどうかを判別することができないことに加え, 情報拡散の研究は行われているものの, 情報の収束を対象とした研究は行われていない.

## 3 デマ情報・訂正情報拡散モデル

### 3.1 SIR モデル

伝染病が広まっていく様子を記述する数理モデルの研究は, 古くから行われてきた. SIR モデルは Kermack らによって提案された数理モデルのひとつである [6][8].

SIR モデルではある集団に属する人を, まだ病気にかかっていない人 ( $S$ : Susceptible), 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 種類の分類する. このモデルでは,  $S$  の人は  $I$  の人と接触することにより, 感染率  $\rho_{(S \rightarrow I)}$  に従って病気に感染し,  $I$  に変化する.  $I$  の人は治癒速

度  $\rho_{(I \rightarrow R)}$  に従い、病気を治癒し免疫を獲得し、 $R$  に変化する。時間  $t$  における  $S$  の人数を  $S(t)$ 、 $I$  の人数を  $I(t)$ 、 $R$  の人数を  $R(t)$  とすると、時間の経過による  $S$ 、 $I$ 、 $R$  の人数の変化は式 1 で表される。

$$\begin{cases} \frac{dS(t)}{dt} = -\rho_{(S \rightarrow I)} I(t) S(t) \\ \frac{dI(t)}{dt} = \rho_{(S \rightarrow I)} I(t) S(t) - \rho_{(I \rightarrow R)} I(t) \\ \frac{dR(t)}{dt} = \rho_{(I \rightarrow R)} I(t) \end{cases} \quad (1)$$

集団の全人数  $N$  は、 $N = S(t) + I(t) + R(t)$  であり、常に一定である。このように、SIR モデルにて想定される状況においては、 $S$  の人は  $I$  の人数に比例して病気に感染していく。また、 $S$  は各個人が感染のリスクにさらされるため、感染速度は  $S$  の人数にも比例する。一方で  $I$  から  $R$  に変化する治癒については、周囲の人の感染状態には依存しない。このため、 $I$  の人数のみが  $R$  の増加率に関係する。

## 3.2 病気と情報の伝播の違い

病気が伝染するのと同様に、興味を引くデマ情報も人々の間で広まっていく。このため、デマ情報は SIR モデルにおける病気であるとみなすことができるしかし、デマ情報と病気にはいくつかの違いがあり、単純に「デマ情報 = 病気」とするだけでは不十分であると考えられる。

### R になる条件

人が病気を治す場合は薬の服用や手術、そして自己治癒能力による治癒（死亡も含む）といった方法がある。すなわち病気は時間経過とともに自然に収束する。一方でデマ情報の拡散の場合、情報の内容にもよるが、デマ情報を見た時点ではそれが誤りであるということに気付かない。真実を知っている一部の人が訂正情報を話して、ようやくデマとして認識される。このように病気とは異なり、 $I$  の人は時間経過とともに  $R$  に変化するとは無い。

### 訂正情報の拡散

デマ情報が拡散した場合、 $I$  の人は訂正情報を得ることで  $R$  に変化する。さらに、一部の人はその訂正情報を、他者に広めようとするであろう.. これは言い換えれば、デマ情報だけでなく訂正情報も SIR モデルにおける病気であるとみなすことができ、デマ情報と訂正情報の両方が拡散する様子を記述する必要がある。また、このことから  $S$  から  $I$  を経ずに  $R$  に変化する人も存在する。

## S, I, R 以外の状態

SIR モデルでは  $S$  の人が  $I$  の人に接触した場合、感染率  $\rho_{(S \rightarrow I)}$  によって  $S$  から  $I$  への状態変化が生じた。 $I$  に変化しない場合は  $S$  のままである。一方でデマ情報の場合、デマの内容を知っているだけの場合と、さらに広めた場合に分けられると考えられる。これは後に訂正情報が拡散した場合、デマが拡散していた事実をそもそも知らなかった人 ( $S$  の人)、デマが拡散していた事実を知っていたが広めなかった人 ( $I_{get}$  とする)、デマを広めてしまった人 ( $I$  の人) では、再びデマ情報を取得した場合や、新しく訂正情報を得た場合の振る舞いが異なると考えられるためである。デマの事実を知らなければ、訂正情報を得たとしても、それが重要なものであると思わないかもしれない。このため、 $S$ 、 $I$ 、 $R$  だけでなく、情報を知っているだけの状態  $I_{get}$ 、 $R_{get}$  があり、それぞれの状態で、他の状態に変化する感染率が異なると考えられる。

### 3.2.1 デマ情報・訂正情報拡散モデルの構築

以上の違いを踏まえたうえで、SNS におけるデマ情報・訂正情報の拡散のモデル化を行う。SNS における情報拡散モデルでは、上述したようにデマと真実の 2 つの情報、すなわちデマ情報と訂正情報の 2 種類の情報が拡散する様子を記述する。ユーザーのデマに対する感染状態を、次のように定める。

- $S$  : デマ情報、訂正情報の両方を見たことがない状態。
- $I_{get}$  : デマ情報のみを見たことがある状態。訂正情報はまだ見ていない。
- $I$  : デマ情報を投稿した状態。訂正情報はまだ見ていない。
- $R_{get}$  : 訂正情報を見たことがある状態。
- $R$  : 訂正情報を投稿した状態。

訂正情報をすでに見た  $R_{get}$  や  $R$  のユーザーは、 $S$ 、 $I_{get}$ 、 $I$  に変化することは無いものとする。各感染状態における感染率を  $\rho_{(S \rightarrow I)}$  ( $S$  のユーザーがデマ情報を見たときに、 $I$  となる確率)、 $\rho_{(I_{get} \rightarrow I)}$ 、 $\rho_{(S \rightarrow R)}$ 、 $\rho_{(I_{get} \rightarrow R)}$ 、 $\rho_{(I \rightarrow R)}$ 、 $\rho_{(R_{get} \rightarrow R)}$  とする。全ユーザー数を  $N$ 、全ユーザーの友人（自身が投稿した文章を読むことができる人、Twitter ではフォロワーに相当）の平均人数を  $F$  とする。また、ある時間  $t$  における  $S$ 、 $I_{get}$ 、 $I$ 、 $R_{get}$ 、 $R$  の人数を  $S(t)$ 、 $I_{get}(t)$ 、 $I(t)$ 、 $R_{get}(t)$ 、 $R(t)$  とする。さらに、ユーザーが投稿した情報を、全ての友人が見たと仮定する。SNS でのデマ拡散の場合も、デマ情報の感染速度は SIR モデル同様  $S$  の人数に比例する。しかし SNS では、あるユーザーが投稿した文章は、基本的にそのユーザーの友人しか見ることはない。このた

め、デマの感染速度は、各ユーザーの  $I$  以外の状態である友人の数に比例する。時間  $t$  の時点での、ユーザー 1 人あたりの  $S$  である友人の人数の期待値は  $\frac{FS(t)}{N}$  となる。  $I_{get}$  である友人の人数の期待値も同様に  $\frac{FI_{get}(t)}{N}$  である。以上から、訂正情報がまだ一度も投稿されていない状況では、感染状態は式 2 のように記述できる。

$$\left\{ \begin{array}{l} \frac{dS(t)}{dt} = -\frac{F}{N}I(t)S(t) \\ \frac{dI_{get}(t)}{dt} = (1 - \rho_{(S \rightarrow I)})\frac{F}{N}I(t)S(t) \\ \quad - \rho_{(I_{get} \rightarrow I)}\frac{F}{N}I_{get}(t)I(t) \\ \frac{dI(t)}{dt} = \rho_{(S \rightarrow I)}\frac{F}{N}I(t)S(t) \\ \quad + \rho_{(I_{get} \rightarrow I)}\frac{F}{N}I_{get}(t)I(t) \end{array} \right. \quad (2)$$

同様に、デマ情報と訂正情報の両方が拡散している場合、感染状態は式 3 のように記述できる。

$$\left\{ \begin{array}{l} \frac{dS(t)}{dt} = -\frac{F}{N}I(t)S(t) - \frac{F}{N}R(t)S(t) \\ \frac{dI_{get}(t)}{dt} = (1 - \rho_{(S \rightarrow I)})\frac{F}{N}I(t)S(t) \\ \quad - \rho_{(I_{get} \rightarrow I)}\frac{F}{N}I_{get}(t)I(t) \\ \quad - \frac{F}{N}I_{get}(t)R(t) \\ \frac{dI(t)}{dt} = \rho_{(S \rightarrow I)}\frac{F}{N}I(t)S(t) \\ \quad + \rho_{(I_{get} \rightarrow I)}\frac{F}{N}I_{get}(t)I(t) \\ \quad - \frac{F}{N}I(t)R(t) \\ \frac{dR_{get}(t)}{dt} = (1 - \rho_{(S \rightarrow R)})\frac{F}{N}R(t)S(t) \\ \quad + (1 - \rho_{(I_{get} \rightarrow R)})\frac{F}{N}I_{get}(t)R(t) \\ \quad + (1 - \rho_{(I \rightarrow R)})\frac{F}{N}I(t)R(t) \\ \quad - \rho_{(R_{get} \rightarrow R)}\frac{F}{N}R_{get}(t)R(t) \\ \frac{dR(t)}{dt} = \rho_{(S \rightarrow R)}\frac{F}{N}R(t)S(t) \\ \quad + \rho_{(I_{get} \rightarrow R)}\frac{F}{N}I_{get}(t)R(t) \\ \quad + \rho_{(I \rightarrow R)}\frac{F}{N}I(t)R(t) \\ \quad + \rho_{(R_{get} \rightarrow R)}\frac{F}{N}R_{get}(t)R(t) \end{array} \right. \quad (3)$$

### 3.2.2 情報拡散モデルのシミュレーション

式 3 に示したデマ情報・訂正情報拡散モデルを用いてシミュレーションを行った。シミュレーションの設定を表 2 に示す。ノードは SNS におけるユーザーに相当し、リンク数はユーザーのお気に入りユーザーの数 (Twitter におけるフォロー数) に相当する。ノード  $i$

表 1: シミュレーション手順

Step1: 表 2 のネットワークを読み込む。

Step2: シミュレーション実行ステップ  $t=1$  のとき、無作為に 1 つのノードを選択し、感染状態を  $I$  に変更する。

Step3:  $t=11$  のとき、無作為に 1 つのノードを選択し、感染状態を  $R$  に変更する。

Step4:  $t=35$  のとき、シミュレーションを終了する。

表 2: シミュレーション: ネットワークのパラメータ

ノード数	50,000
リンク数 (次数) の期待値	上限 = 30.00 下限 = 10 パレート指数 = 0.5
リンクされやすさ	上限 = 15.0 下限 = 0.05 パレート指数 = 0.5
感染確率	$\rho_{(S \rightarrow I)} = 0.05$ $\rho_{(I_{get} \rightarrow I)} = 0.05$ $\rho_{(S \rightarrow R)} = 0.1$ $\rho_{(I_{get} \rightarrow R)} = 0.1$ $\rho_{(I \rightarrow R)} = 0.15$ $\rho_{(R_{get} \rightarrow R)} = 0$

がノード  $j$  とつながるかどうかは、ノード  $i$  のリンク数の期待値と、ノード  $j$  のリンクされやすさによって決定される。パレート分布を用いているのは、SNS、特に Twitter のネットワークはスケールフリーネットワークとなるためである [1][12]。シミュレーションの手順は表 1 に示すとおりである。シミュレーションを 100 回行ったときの結果の平均を図 1 に示す。

## 4 ツイートの収集と分析

### 4.1 ツイート収集手順

デマ情報・訂正情報拡散モデルによって実際に SNS にて発生したデマ情報の拡散の様子を再現できるかどうかを確かめた。本研究では SNS のうち、特に拡散速度が速いといわれている Twitter に焦点を当てた。鳥海らの東日本大震災前後における Twitter ネットワークの変化に関する研究 [10] にて収集されたツイートの

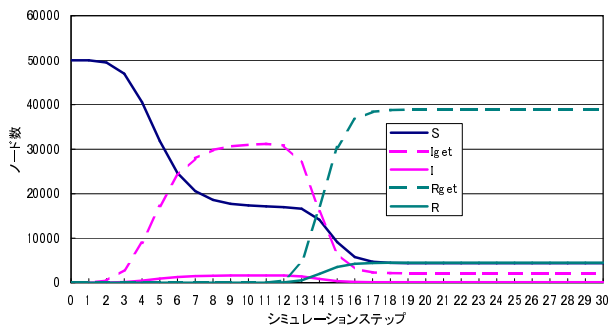


図 1: シミュレーション：ノードの感染状態

うち、2011年3月11日～2011年3月24日の間に投稿されたツイートを対象とし、その中から、あるデマに言及しているデマツイートおよび訂正ツイートを抽出し、分析を行った。なお、ツイートだけでなく、2011年1月30日におけるユーザーのフォローの状況（フォローネットワーク）も得た。ただし、フォローネットワークは全てのユーザーを網羅しているわけではない。

デマツイートおよび訂正ツイートの抽出は表3に示す手順で行った。以下に、本論文にて用いる単語の定義を示す。

- 必須キーワード：ある特定のデマに言及するために、必ず含まれているべきキーワード。
- ネガティブキーワード：ある特定のデマについて、デマ情報を拡散させるツイートに含まれるキーワード。
- ポジティブキーワード：ある特定のデマについて、訂正情報を拡散させるツイートに含まれるキーワード。
- 候補ツイートデータ：全ツイートデータのうち、必須キーワードを含む、ある特定のデマに言及していると思われるツイートの集合。
- デマ・訂正ツイートデータ：候補ツイートデータのうち、ネガティブキーワード、ポジティブキーワードを含むツイートの集合。
- デマツイートデータ：デマ・訂正ツイートデータのうち、ポジティブキーワードを含まないツイートの集合。
- 訂正ツイートデータ：デマ・訂正ツイートデータのうち、ポジティブキーワード1つ以上を含むツイートの集合。

## 4.2 コスモ石油に関するデマ

東日本大震災直後、千葉県市原市のコスモ石油の千葉製油所にて火災が発生した。これに関連して「コス

表 3: デマツイートおよび訂正ツイートの抽出手順

Step1：実際に拡散したデマをひとつ選択する。

Step2：ツイートがそのデマに言及しているかどうかを判別する基準として、デマの内容から必須キーワードを設定する。

Step3：必須キーワードでツイートの検索を行い、デマに関する情報を含むと思われるツイート（候補ツイートデータ）を全て抽出する。

Step4：候補ツイートデータから、RT数が多いデマツイートと訂正ツイートを複数選択する。

Step5：選択したデマツイート、訂正ツイートから、ネガティブキーワード、ポジティブキーワードを設定する。

Step6：必須キーワード、ネガティブキーワード、ポジティブキーワード全てを用いて検索を行い、候補ツイートデータから、デマ・訂正ツイートデータを抽出する。

Step7：必須キーワード、ポジティブキーワードを用いて検索を行い、デマ・訂正ツイートデータから、マツイートデータと訂正ツイートデータを分離する。

モ石油の爆発により有害物質が雲などに付着し、雨などといっしょに降る」といった内容のチェーンメールが出回り、一部のユーザーがこのメールを転載の形でTwitterに投稿したため、多数のユーザーにこの情報が拡散した。3月12日にコスモ石油の公式Webページにおいて、「火災が発生したタンクに貯蔵されていたものはLP（液化石油）ガスであり、人体へ及ぼす影響は非常に少ない」と発表された。

今回、このコスモ石油に関するデマを対象に、デマに言及していると思われるデマツイートおよび訂正ツイートを抽出した。抽出に用いた各種キーワードを表4に示す。このキーワードにて検索を行うことにより得られたデマツイートは9,652件、訂正ツイートは25,883件であった。デマツイートデータおよび訂正ツイートデータの、15分刻みの時間当たりのツイート数を図2に示す。また、コスモ石油に関するデマのデマツイートまたは訂正ツイートを投稿した全ユーザーを対象とした、各時刻における感染状態別の人数を図3に示す。

表 4: コスモ石油に関するデマ:ネガティブキーワードおよびポジティブキーワード

必須キーワード	コスモ石油, 有害物質
ネガティブキーワード	傘, カップ, レインコート
ポジティブキーワード	デマ, ガセ, 嘘, 誤, 偽, 否定, 無害, チェーンメール, チェンメ, 事実 (A-1), ない (A-2), ありません (A-2)

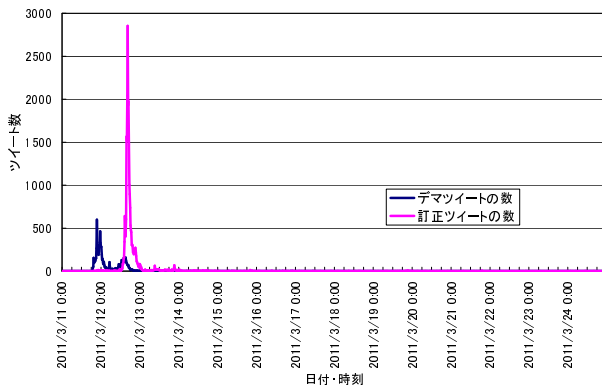


図 2: コスモ石油に関するデマ:デマツイート数および訂正ツイート数

### 4.3 ツイートの分析

シミュレーションにて各種パラメータの設定を行うため, 取得したツイートの特徴の分析を行った.

#### デマツイート・訂正ツイート投稿者の割合

まず, 感染率  $\rho_{(I \rightarrow R)}$  は, デマツイートを投稿したユーザーに対する, デマツイートおよび訂正ツイートの両方を投稿したユーザーの割合から決定できると考えた. このため, 取得した各種ツイートデータから, 各ユーザーのデマに対する感染状態を調べた. 各ユーザーの感染状態の解析は, 表 5 に示す手順で行った.

表 6 に, コスモ石油に関するデマにおける, デマツイートを投稿したユーザーの数 ( $N_I$ ), 訂正ツイートを投稿したユーザーの数 ( $N_R$ ), デマツイートおよび訂正ツイートの両方を投稿したユーザーの数 ( $N_{(I \rightarrow R)}$ ), デマツイート投稿ユーザーに対するデマツイート・訂正ツイート投稿者の割合 ( $N_{(I \rightarrow R)}/N_I$ ) を示す.

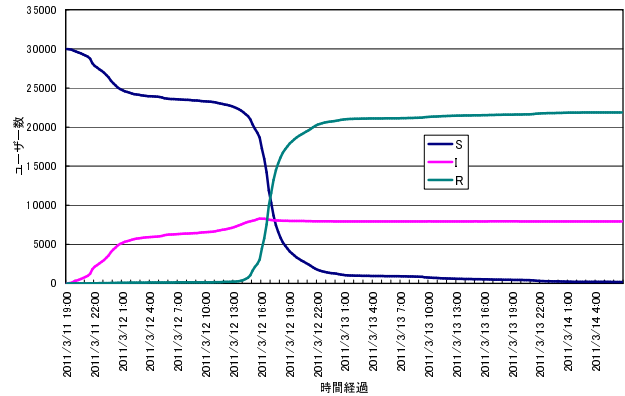


図 3: コスモ石油に関するデマ: ユーザーの感染状態

表 5: 実データにおけるユーザーの感染状態の解析

Step1: デマツイートデータ, 訂正ツイートデータから, デマツイートを投稿したユーザーおよび訂正ツイートを投稿したユーザーのリストを作成する.

Step2: デマツイートを投稿したユーザーのリストと訂正ツイートを投稿したユーザーのリストを照会し, 重複したユーザーのリストを作成する.

#### デマツイート・訂正ツイートのリツイートのされやすさ

次に,  $\rho_{(I \rightarrow R)}$  以外の感染率の設定に関連し, デマツイートおよび訂正ツイートの, リツイートのされやすさを調べた. リツイートのされやすさの解析は, 表 7 に示す手順で行った. デマツイートおよび訂正ツイートを投稿したユーザー  $i$  のフォロワーのうち, デマツイートを投稿したフォロワーの人数, 訂正ツイートを投稿したフォロワーの人数を収集する. 全てのフォロワーがユーザー  $i$  のツイートをリツイートしたとは限らないが, デマツイートを投稿したフォロワーの人数と訂正ツイートを投稿したフォロワーの人数の比率を算出することにより, デマツイートのリツイートのされやすさ, 訂正ツイートのリツイートのされやすさの比較が可能である.

各デマのデマツイートおよび訂正ツイートの両方を投稿したユーザーのうち, 500 人を無作為に抽出し, フォロワーの感染状態について調べた. ただし先述したように, フォローネットワークデータは全てのユーザーのデータを網羅しているわけではない. このため, コスモ石油に関するデマでは, 500 人中 361 人のフォロワーの感染状態を収集することができた. 表 8 に, 収集できたユーザーの人数, および収集できたユーザーのフォロワーのうち感染状態が  $I$  となったフォロワー

表 6: デマツイートおよび訂正ツイート投稿ユーザーの数

$N_I$ (人)	$N_R$ (人)	$N_{(I \rightarrow R)}$ (人)	$N_{(I \rightarrow R)}/N_I$
9,275	17,080	1,364	0.147

表 7: デマツイートおよび訂正ツイートのリツイートされやすさの取得

Step1: 表 5 の手順に従い, デマツイート, デマ訂正ツイートの両方を投稿したユーザーのリストを作成する.

Step2: フォロワーネットワークデータから, 両ツイート投稿ユーザーをフォローしているユーザーのリスト (フォロワーリスト) を作成する.

Step3: 両ツイート投稿ユーザーのフォロワーリストと, デマツイートデータ, 訂正ツイートデータを照らし合わせ, デマツイートを投稿したことがあるフォロワー ( $I$  のフォロワー) または訂正ツイートを投稿したことがあるフォロワー ( $R$  のフォロワー) の数を取得する.

の平均人数 (平均  $F_I$ ), 感染状態が  $R$  となったフォロワーの平均人数 (平均  $F_R$ ),  $I$  となったフォロワーと  $R$  となったフォロワーの人数の比率 (平均  $F_R/\text{平均 } F_I$ ) を示す.

## 他メディアからの情報の流入

さらに, 他のメディアからの情報の流入についても調べた. Twitter ネットワークにおける情報の拡散は他ユーザーのツイートを再投稿するリツイートが主流であるが, テレビや新聞からの情報や他 SNS で流れている情報を投稿するということもある. あるデマツイートを投稿したユーザー  $i$  に着目したとき, ユーザー  $i$  がフォローしているユーザー全てがデマツイートを投稿していない場合, ユーザー  $i$  は Twitter 以外からデマ情報を得たと推測できる. 訂正ツイートのみを投稿したユーザーについても同様である. さらに, これらのユーザーのツイート内容を調べ, リツイートを投稿しているユーザーは Twitter から情報を得たものとして除外した.

コスモ石油に関するデマについては, 500 人中 381 人のデマツイートのみ投稿したユーザー, 500 人中 360 人の訂正ツイートのみ投稿したユーザーの, フォロワーユーザーの感染状態を収集した. このうち 9 人のデマ

表 8: デマツイートおよび訂正ツイート投稿ユーザーの数

平均 $F_I$ (人)	平均 $F_R$ (人)	平均 $F_R/\text{平均 } F_I$
2.64	5.39	2.04

ツイートのみ投稿したユーザーと, 4 人の訂正ツイートのみ投稿したユーザーが, 外部からの情報を得たと思われる結果を得た. 全体に対するこれらの割合はそれぞれ 2.36%, 1.11% であった.

## 5 シミュレーションによる検証

### 5.1 情報拡散モデルと実データの比較

Twitter にて実際に拡散したデマの拡散の様子を把握したため, この拡散の様子がデマ情報・訂正情報拡散モデルにて再現できるかどうかを調べた. ただし, 今回は情報拡散モデルから  $I_{get}$ ,  $R_{get}$  の概念を取り除いたうえでシミュレーションを行った. この理由としては, コスモ石油に関するデマの場合, 図 2 のようにデマツイート, 訂正ツイートともに 1 度だけ拡散していることが確認でき, このような場合では複数のデマツイートや訂正ツイートを見るということはまれであり, 例えば  $I_{get}$  から  $I$  への変化はとても少ないと考えられる. デマ情報・訂正情報拡散モデルのシミュレーションでも, 各ノードのデマ情報を受け取った回数の分布は図 4 の通りであった. シミュレーションにて用いたネットワークは実際のネットワークとは異なる部分があるにせよ, 57.6% のノードがデマ情報を受け取った回数が 1 回以下であること, 複数回受け取った場合でも, 現実の Twitter ではその全てを見たは限らないことなどから, シミュレーションにおいても複数のデマ情報を見たノードは少ないといえる. さらにツイート of データからはツイートを見ただけのユーザーの数を正確に把握できないということもあり, 今回は, 情報を見ただけという状態  $I_{get}$ ,  $R_{get}$  は考慮せずにシミュレーションを行うこととした.

シミュレーションの手順は表 1 と同じである. シミュレーションにて用いたネットワークは, 情報拡散モデルのシミュレーションにて用いたものと同様のネットワークであるが, 感染確率は実データの分析結果を反映し, 表 9 のように設定した. シミュレーションを 100 回行ったときの結果の平均を図 5 に示す. 実データの  $S$ ,  $I$ ,  $R$  のユーザー数と, シミュレーションの  $S$ ,  $I$ ,  $R$  のノード数の時間経過による推移は, おおむね一致していると考えられる. よって, デマ情報・訂正情報拡散モデルにより, デマツイートや訂正ツイートの実際の拡散の様子が説明できたと考えられる.

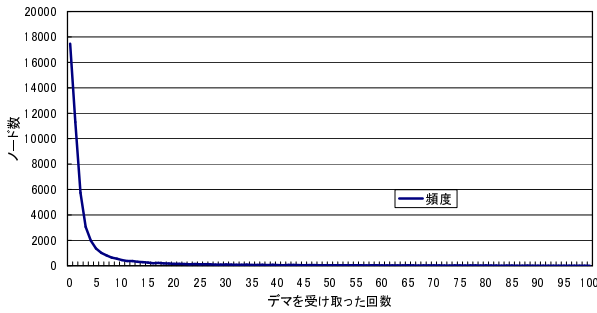


図 4: シミュレーション: デマ情報を受け取った回数の分布

表 9: シミュレーション: 実データを反映した感染率の設定

感染確率	$\rho(S \rightarrow I) = 0.05$
	$\rho(I_{get} \rightarrow I) = 0$
	$\rho(S \rightarrow R) = 2.04 \times \rho(S \rightarrow I)$
	$\rho(I_{get} \rightarrow R) = 0$
	$\rho(I \rightarrow R) = 0.147$
	$\rho(R_{get} \rightarrow R) = 0$

## 5.2 デマ拡散の早期収束に向けた検討

デマ情報・訂正情報拡散モデルを用いて、デマの拡散を早期に収束させるための方法の検討を行った。今回は、訂正情報の拡散によりデマが収束することから、より早く訂正情報を拡散させる方法に着目した。

訂正ツイートを最初に投稿した、すなわち訂正ツイート拡散の起点となったユーザーに焦点を当て、あるユーザーに訂正ツイートを投稿するように依頼するという仮定の状況におけるシミュレーションを行った。起点となるユーザーの選択の方法として以下に示す方法を用意した。

- 起点ユーザー選択ルール A: デマツイート拡散の起点となったユーザーに、訂正ツイート拡散の起点になってもらう。
- 起点ユーザー選択ルール B: ユーザー全員のうち、ネットワークのハブとなっているユーザー 1 人に、拡散の起点になってもらう。
- 起点ユーザー選択ルール C: デマツイートを投稿したユーザーのうち、もっともフォロワーが多いユーザーに、拡散の起点になってもらう。

シミュレーションでは、表 1 の Step:3 での  $R$  となるノードを選択するルールを、上記の起点ユーザー選択ルールに沿って変更した。ただし、ルール B ではリンクされている数が 2000 以上のノード (50000 ノード中

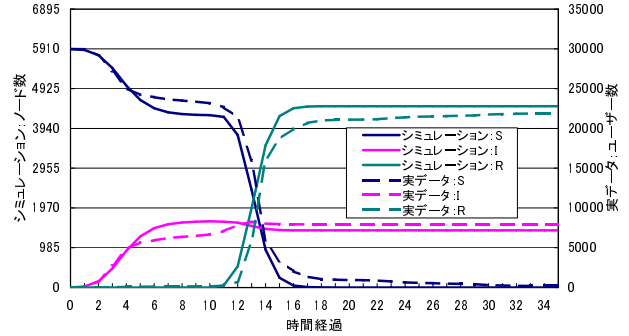


図 5: シミュレーションと実データの感染状態別人数の比較

524 ノード) をハブノードとみなした。各起点ユーザー選択ルールを用いたシミュレーションを 100 回ずつ行い、 $R$  のノードの増加をみた。表 1 Step:3 の通常ルールの場合も含めた結果を図 6 に示す。

選択ルール A は、 $I$  から  $R$  への感染率  $\rho(I_{get} \rightarrow R)$  が最も高いことから、デマツイートの拡散と同様の経路で訂正ツイートを拡散させることでの早期の訂正ツイートの拡散を狙ったものであったが、図 6 より、通常ルールの場合と差はみられなかった。これにより、Twitter のようなフリースケールネットワークでは、起点を無作為に選択する場合、起点に関わらず同じように情報が拡散することが確認できたといえる。一方で選択ルール B や選択ルール C は、通常ルールや選択ルール A よりも拡散スピードが速くなっていることがわかる。これは、訂正ツイートをハブのユーザーが投稿することにより、一般ユーザーよりも早く訂正ツイートの拡散が行われることを示している。ただし、選択ルール B と選択ルール C では若干選択ルール C が優位であったものの、はっきりと優位であるとはいえなかった。しかしながら、あるユーザーに訂正ツイートを投稿するように依頼するという状況を想定した場合、デマツイートが拡散したという事実を知っているユーザーであれば、依頼しやすいという状況も生じると思われる。

## 6 まとめ

SNS は情報を発信したり収集したりするツールとして近年急速に発展を遂げた。特に Twitter はリアルタイム性、速報性が高いという性質があり、情報が短時間で広まるという現象がみられた。一方でデマ情報も同様に拡散することがあり、問題となっていた。

本研究では SNS でのデマ情報の拡散の様子をモデル化した。さらに、投稿された内容がデマであるかを完璧に判別するのは不可能であることから、デマが拡散



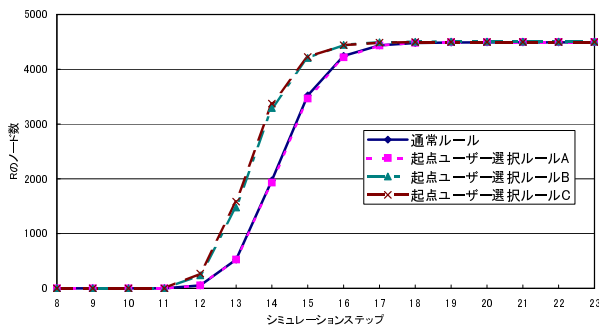


図 6: デマ拡散早期収束シミュレーション:R ノード数

してしまうことを念頭に置き、早急にデマの拡散を収束させる方法についての検討を行った。

まず、SIR モデルの拡張を行い、デマ情報と訂正情報の両方が伝播するモデルを構築した。そして、実際に Twitter ネットワーク上を拡散したツイートを調べ、これから得たパラメータを設定してシミュレーションを行い、この情報拡散モデルが実際のツイートの拡散を再現できることを確かめた。

デマ拡散を収束させる方法についての検討は、デマ情報の拡散を収束させるには訂正情報の拡散速度を高めればよいのではないかと考えて行なった。あるユーザーが起点となって訂正ツイートが拡散していくという場合において、このユーザーの選び方により、通常よりも早く訂正ツイートが広まることが確認できた。実際行うには様々な問題があると思われるが、デマの訂正を行う公式発表に加え、そのデマを知っておりかつフォロワーの多いハブのユーザーに対して訂正ツイートの投稿を依頼できるならば、より早く訂正情報を広めることができると考えられる。

今後の課題としては、複数回デマ情報や訂正情報の拡散が起こる場合の、拡散を重ね合わせたモデルを構築していくことが挙げられる。また、社会学の知見を導入し、情報を見ただけの人についても調べていく必要がある。さらに、Twitter 以外の SNS にて拡散した情報についても、拡散の様子を再現できるかどうかを確かめていく予定である。

## 謝辞

本研究を遂行するに当たり、Twitter 検索のログデータを提供して頂いたクックパッド株式会社の兼山元太氏に感謝致します。

## 参考文献

- [1] Sayaka Akioka, Norikazu Kato, Yoichi Muraoka, Hayato Yamana : Cross-media Impact on Twitter in Japan, *Proceedings of the 2nd international workshop on Search and mining user-generated contents*, pp.111-118, 2010.
- [2] Yasuyoshi Aosaki, Taro Sugihara, Katsuhiko Umemoto: Examining the Trend toward a Service Economy in Information Media through Changes to Technology: Influence of Twitter on Media Companies, *Proceedings of Technology Management for Global Economic Growth (PICMET) '10*, pp.1-5, 2010.
- [3] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, Duncan J. Watts : Everyone's an Influencer: Quantifying Influence on Twitter, *Proceedings of the fourth ACM international conference on Web search and data mining*, pp.65-74, 2011.
- [4] Carlos Castillo, Marcelo Mendoza, Barbara Poblete : Information Credibility on Twitter, *Proceedings of the 20th international conference on World wide web*, pp.675-684. 2011.
- [5] Bernardo A. Huberman, Daniel M. Romero, Fang Wu : Social networks that matter: Twitter under the microscope, *First Monday*, Vol.14, No.1, 2009.
- [6] W. O. Kermack, A. G. McKendrick : A Contribution to the Mathematical Theory of Epidemics, *Proceedings of the Royal Society 115A*, pp.700-721, 1927.
- [7] Haewoon Kwak, Changhyun Lee, Hosung Park, Sue Moon : What is Twitter, a Social Network or a News Media?, *Proceedings of the 19th international conference on World wide web*, pp.591-600, 2010.
- [8] 増田直紀, 今野紀雄 : 複雑ネットワークの科学, 産業図書, 2005.
- [9] 総務省:平成 23 年度情報通信白書, <http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h23/pdf/index.html>, 2011.
- [10] 鳥海不二夫, 篠田孝祐, 栗原聡, 榎剛史, 風間一洋, 野田五十樹 : 震災がもたらしたソーシャルメディアの変化, *JWEIN11*, pp.41-46, 2011.
- [11] 梅島彩奈, 宮部真衣, 荒牧英治, 灘本明代: 災害時 Twitter におけるデマとデマ訂正 RT の傾向, *情報処理学会研究報告*, Vol.2011-IFAT-103, No.4, pp.1-6, 2011.
- [12] Jianshu Weng, Ee-Peng Lim, Jing Jiang, Qi He: TwitterRank: Finding Topic-sensitive Influential Twitterers, *Proceedings of the third ACM international conference on Web search and data mining*, pp.261-270, 2010.
- [13] 吉田光男, 松本明日香 : ソーシャルメディアの政治的活用—活用事例と分析事例から—, *人工知能学会誌*, 27 巻 1 号, pp.35-42, 2012.