

Twitterにおけるデマツイートの拡散モデルの構築と デマ拡散防止モデルの推定

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

白井 嵩士*¹ 榊 剛史*² 鳥海 不二夫*² 篠田 孝祐*³ 風間 一洋*⁴ 野田 五十樹*⁵
Takashi Shirai Takeshi Sakaki Fujio Toriumi Kosuke Shinoda Kazuhiro Kazama Itsuki Noda
沼尾 正行*⁶ 栗原 聡*⁶
Masayuki Numao Satoshi Kurihara

*¹大阪大学大学院 情報科学研究科 Graduate School of Information Science and Technology, Osaka University
*²東京大学 Tokyo University
*³理化学研究所 RIKEN

*⁴NTT 未来ねっと研究所 NTT Network Innovation Laboratories
*⁵産業技術総合研究所 National Institute of Advanced Industrial Science and Technology

*⁶大阪大学 産業科学研究所
The Institute of Scientific and Industrial Research, Osaka University

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

近年の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・ネットワーキング・サービス (SNS) と呼ばれるサービスが登場し、話題を呼んでいる。SNS 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特徴を持ち、Facebook*¹やTwitter*²といったサービスがよく知られている。一般の個人が簡単に情報を投稿したり収集することができ、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 [5]。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、デマ情報が急速かつ広範囲に広がってしまうという問題が指摘された。例えば、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、デマであると気づかずに、情報を広めっ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。Kawak 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 [4]。Kawak 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 [6]。

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 [8]。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 [1]。

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 [2]。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

連絡先: 栗原 聡, 大阪大学産業科学研究所, 〒 567-0047
大阪府茨木市美穂ヶ丘 8-1, Tel(Fax):06-6879-8427, E-mail:kurihara@sanken.osaka-u.ac.jp

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

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

トが持つ特徴を調べている [7].

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 [3].

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。

3.2 病気の伝染と情報の拡散

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。

まず, 病気の伝染は薬の服用や手術, そして自己治癒能力による治癒といった方法により, 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, 訂正情報を得ない限り, ようやくデマとして認識されることはない。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 に変化することは無い。

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。つまり, 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。

さらに, デマ情報拡散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。デマの事実を知らなければ, 訂正情報を得たととしても, それが重要なものであると思わず, 再投稿を行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, それぞれの状態では, 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。

まず, ユーザーのデマに対する感染状態を, 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が, 訂正情報はまだ見ていない状態), I (デマ情報を投稿したが, 訂正情報はまだ見ていない状態), 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) とする。訂正情報をすでに見た R_{get} や R のユーザーは, S, I_{get}, I に変化する事は無いものとする。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にはない。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, 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 1 のように記述できる。

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

4. ツイートの収集と分析

4.1 デマツイートの収集

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, デマ情報・訂正情報拡散モデルによって, 実際に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。

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 [6] にて収集されたツイートのうち, 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。

表 1: コスモ石油に関するデマ:デマキーワードおよび訂正キーワード

デマキーワード	傘, カッパ, レインコート
訂正キーワード	デマ, ガセ, 嘘, 誤, 偽, 否定, 無害, チェーンメール, チェンメ, 事実 and ない, 事実 and ありません

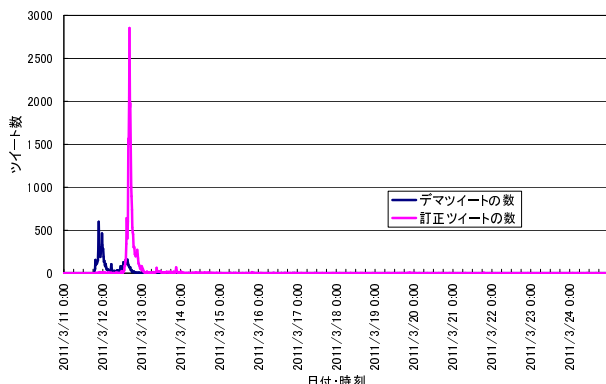


図 1: コスモ石油に関するデマ:デマツイート数および訂正ツイート数

今回は「コス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」といった内容のデマを対象とし、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、キーワードの検索により抽出した。抽出したツイートは、「コスモ石油」と「有害物質」の両キーワードを含み、かつ表 1 に示す、デマツイートまたは訂正ツイートのいずれかを含むツイートである。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 1 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 2 に示す。

4.2 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。

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。デマツイートを投稿したユーザー 9,275 人のうち、訂正ツイートを投稿したユーザーは 1,364 であったため、訂正率は 14.7% であった。

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。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し、人数の比率を算出した。この結果、訂正ツイートはデマツイートの 2.04 倍リツイートされやすいことがわかった。

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

表 2: 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	$\rho_{(S \rightarrow I)} = 0.05$
	$\rho_{(I_{get} \rightarrow I)} = 0$
	$\rho_{(S \rightarrow R)} = 2.04 \times \rho_{(S \rightarrow I)}$
	$\rho_{(I_{get} \rightarrow R)} = 0$
	$\rho_{(I \rightarrow R)} = 0.147$
	$\rho_{(R_{get} \rightarrow R)} = 0$

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。この結果、外部から得た情報を投稿したと思われるユーザーは、デマツイートの投稿ユーザーのうち 2.36%、訂正ツイートの投稿ユーザーのうち 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 1 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。本検証の前にデマ情報・訂正情報拡散モデルを用いた予備実験のシミュレーションを行ったが、予備実験では 57.6% のノードがデマ情報を受け取った回数が 1 回以下であったことから、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。さらにツイートの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。

シミュレーションにて用いたネットワークは、ノード数 50,000、ノードの次数の分布はパレート分布に従うものとし、感染確率は実データの分析結果を反映し、表 2 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 2 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、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。

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

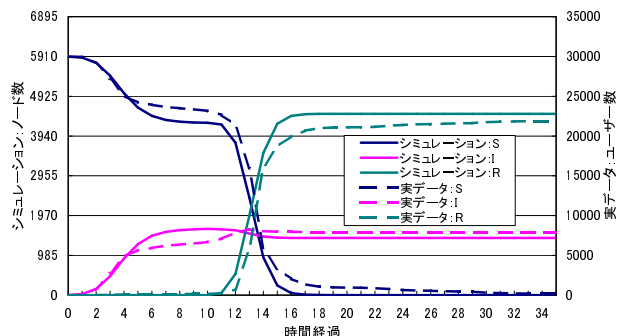


図 2: シミュレーションと実データの感染状態別人数の比較

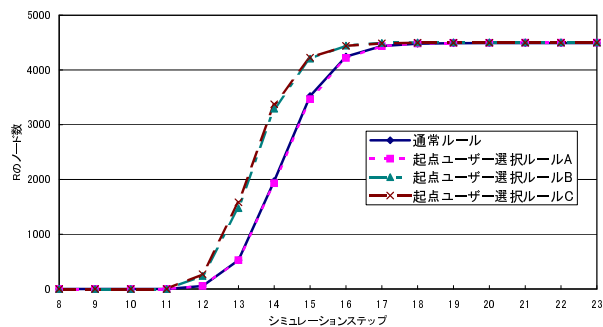


図 3: デマ拡散早期収束シミュレーション:R ノード数

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。

- 通常ルール：全ユーザーのうち、ランダムで 1 人が訂正ツイート拡散の起点となる。
- 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、最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。

このうち、ルール B ではリンクされている数が 2,000 以上のノード (50,000 ノード中 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。通常ルールの場合も含めた結果を図 3 に示す。

選択ルール A の結果は、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を示している。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

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、デマが拡散することを前提に、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、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スピードを高めればよい

という考えに基づいて行った。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。

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。

参考文献

- [1] 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.
- [2] Carlos Castillo, Marcelo Mendoza, Barbara Poblete : Information Credibility on Twitter, *Proceedings of the 20th international conference on World wide web*, pp.675-684, 2011.
- [3] W. O. Kermack, A. G. McKendrick : A Contribution to the Mathematical Theory of Epidemics, *Proceedings of the Royal Society 115A*, pp.700-721, 1927.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] Jianshu Weng, Ee-Peng Lim, Jing Jiang, Qi He : Twitter-Rank: Finding Topic-sensitive Influential Twitterers, *Proceedings of the third ACM international conference on Web search and data mining*, pp.261-270, 2010.