

SNSにおける投稿件数推移分析のための情報拡散モデル

中島 圭佑[†] 南川 雅人^{††} 塩田 茂雄^{††}

[†] 千葉大学工学部都市環境システム学科, 〒 263-8522 千葉市稲毛区弥生町 1-33

^{††} 千葉大学大学院融合理工学府地球環境科学専攻, 〒 263-8522 千葉市稲毛区弥生町 1-33

E-mail: †{afwa5455,adda2062}@chiba-u.jp, ††shioda@faculty.chiba-u.jp

あらまし 実社会で人々の関心事が発生すると, SNSに関連する書き込みが多数投稿され, ピークを迎えたのち沈静化する様子が観察される. 本研究では, この現象は実社会での情報拡散過程が SNS 上に現れた結果であるとして, 感染症の数理モデルとして使われる SIR モデルを用いて投稿数の時間推移を説明する. また, モデルに基づいて, SNS の投稿数の時間推移を解析的に評価する手法を提案する. 数値実験により, 投稿数の時間推移は情報発信ノードの次数などネットワーク構造に大きく依存すること, 隣接するノード同士の相関性に関する仮定が解析結果に影響し, 相関が強いとする仮定のもとで解析結果とシミュレーション結果が良く一致すること等を明らかにする.

キーワード 情報拡散, SNS, SIR モデル, ノード相関,

Information Diffusion Model for Analyzing Temporal Variation of Contents Posted to SNSs

Keisuke NAKAJIMA[†], Masato MINAMIKAWA^{††}, and Shigeo SHIODA^{††}

[†] Faculty of Engineering, Chiba University, 1-33 Yayoi, Inage, Chiba, 263-8522 Japan

^{††} Graduate School of Science and Engineering, Chiba University, 1-33 Yayoi, Inage, Chiba, 263-8522 Japan

E-mail: †{afwa5455,adda2062}@chiba-u.jp, ††shioda@faculty.chiba-u.jp

Abstract When an impressive event occurs in the real world, several contents related to the event are posted to the SNSs; the popularity of the contents rapidly grows and, after a peak, it usually fades over time. In this paper, we explain this temporal variation of contents in the SNSs by using the SIR model, which is often used as a mathematical model of infectious diseases. We find that the temporal variation of the number of contents posted to the SNSs largely depends on the network structure, for example, the degree of the source node of the content. We also find that the simulation results are very consistent with the analytical results obtained by the assumption that the states of neighbor nodes are not independent but strongly correlated.

Key words Information Diffusion, SNS, SIR model, Node Correlation

1. ま え が き

Twitter や Facebook などのソーシャルネットワーキングサービス (Social Networking Service: SNS) が情報発信, 情報収集, コミュニケーションのための日常的なツールとして普及して久しい. SNS の登場により, 対面的なコミュニケーションが主流の時代の想像を超える量のかつ多様な情報が SNS を介して急速に拡散し, 我々の生活に大きな影響を及ぼす時代となっている. このため, SNS 上で得られる情報や SNS で観察される様々な現象を解析し, そこから有為な知見を得るための試みが多数なされている [1]~[13].

実社会で人々の関心事が発生すると, 関連する書き込みが SNS に投稿され, 単位時間あたりの投稿数がピークを迎えた後, 書き込みは次第に減少して, やがて沈静化する現象 (投稿件数のスパイク現象) が起きる. 本研究では, これを実社会で

情報が拡散して, やがて広く知られるところとなり, それ以上の情報伝達 (SNS への投稿) が不要になるまでの過程, つまり実社会における情報拡散過程が SNS 上に表出した現象と解釈する. 情報の拡散は感染症の流行と類似していることから, 本研究では, この SNS への投稿件数の時間変化を, 感染症の数理モデルとして使われている SIR モデル [14], [15] により説明する. その際に, 投稿件数の時間変化が, SNS のネットワーク構造 (ネットワークトポロジ) にどのように依存するかを調べるため, 隣接行列をパラメタとして含むモデルを採用する. さらに, モデルをシミュレーションで評価するだけでなく, SNS の投稿数の時間推移を解析的に評価する手法を提示する.

シミュレーション実験により, SNS の投稿数のスパイク現象は用いたモデルで再現できること, スパイクの形状は情報発信ノードの次数に依存することなどを示す. また, 隣接関係にある二つのノードの状態の相関性をどのように扱うかが解析結果

に大きく影響し、Facebook や Twitter の場合は、相関性が最も強い場合を仮定すると（強相関近似）、シミュレーション結果と大変良く一致する解析結果が得られることを明らかにする。

以下、2. で既存の研究例を説明し、3. において利用するモデルの概要と、モデルを解析的に分析するための手法を説明する。4. で数値例を説明し、5. で結論と今後の課題を述べる。

2. 既存研究

SIS モデルとは、健康状態（Susceptible：状態 S）のノードが、感染状態（Infected：状態 I）の隣接ノードから確率的に感染して状態 I に移行し、また病気が自然治癒して状態 S に戻るということを繰り返すモデルである。SIR モデルでは、病気が治癒すると免疫獲得状態（Recovered：状態 R）に移行し、再び感染することはない。SIS や SIR モデル等を用いた感染症の研究例として、ネットワークの構造を陽に考慮せずに、平均場近似を用いて現象論的なアプローチを試みるもの [16]、次数分布の情報を用いるもの [17], [18]、さらには隣接行列を用いてネットワークの構造を直接モデルの中に取り入れるもの [19]～[23] などがある [15]。

この SIS モデルや SIR モデルを用いて、SNS 上の事象を分析を行った研究の例として [1]～[5] がある。Leskovec ら [1] は Blog サイト間の関連性が生成されるメカニズムを SIS モデルを用いて分析している（被感染者は Blog サイト）。Okada ら [3] は、本研究と同様、SNS の投稿数のスパイク（パースト的な書き込み）は情報拡散の帰結であるとして、（ネットワーク構造を考慮しない）巨視的な SIR モデルにより投稿数のスパイクの生成メカニズムを分析している。Cheng ら [5] は、SIR モデルに免疫獲得状態から感染状態へ移行するメカニズムを取り入れることで、投稿数のスパイクが繰り返し生じる事象を説明できるとしている。

感染症や SNS 上の事象の分析に使われている SIS や SIR 以外の数理モデルの例として、IC（Independent Cascade）モデルや LT（Linear Threshold）モデルがある。IC モデルは、離散時間モデルを採用しており、ある時刻で情報を受け取ったノード i は、次の時刻で各隣接ノードに情報を転送する。ノード i の隣接ノード j は確率 p_{ij} で情報を受け取ることができる（受け取りに失敗することがある）。LT モデルも離散時間モデルを採用している。LT モデルでは、ネットワークを構成しているリンクにそれぞれ重みが付けられており、アクティブな（情報を持つ）隣接ノードからのリンクの重みがしきい値を超えた場合、次の時点でそのノードはアクティブになる。IC モデルや LT モデルを用いた研究例として [10]～[13] がある。

その他の数理モデルを用いた研究例として [6]～[9] などがある。例えば、Matsubara ら [7] は、Blog の投稿数の多様なスパイクの形状を説明するために、ブログへの書き込みが更なる書き込みの誘発要因となること、また誘発の強さが書き込まれてからの時間とともに（power-law 的に）減衰すること等の特徴とするモデルを提案している。

3. 情報拡散モデル

3.1 モデルの説明

有向ネットワークにおいて、1 つないしは複数のノードを起点として、情報が有向リンクを経由してネットワーク内の各ノードに拡散していく現象について考察する。ノードは

状態 0: 情報を知らない

状態 1: 情報は知っているが、未拡散（隣接ノードに未転送）

状態 2: 情報を知り、拡散済み（隣接ノードに転送済み）

の 3 状態のいずれかを取る。状態 0 のノードは（隣接ノードが拡散した）情報を受信することで状態 1 に移行する。状態 1 に移行したノードは、指数分布に従う時間経過後に全ての隣接ノードへの同時情報転送を行い、状態 2 に移行する。状態 2 への移行後は、状態遷移は生じない（状態 2 に留まる；以降、情報転送は行わない）。

上記のモデルは、隣接行列と各ノードの状態 1 から状態 2 への遷移率をパラメータとして有する。以下では、隣接行列を $A = \{a_{ij}\}$ (a_{ij} はノード i からノード j への有向リンクが存在すれば 1、存在しない場合は 0 を取る変数)、ノード i の状態 1 から状態 2 への遷移率を λ_i とする。時刻 t においてノード i がとる状態の番号 (0, 1, or 2) を $Z_i(t)$ で表すこととすると、ノード 1 からノード N までで構成されるグラフの場合、時刻 t でのネットワークの状態は各ノードの状態の組、すなわち

$$(Z_1(t), Z_2(t), \dots, Z_N(t))$$

で記述される。ネットワークの状態遷移は連続時間マルコフ連鎖に従うので、シミュレーションにより情報拡散過程を正確に追うことが可能である。

注 1. 状態 0 を健康状態、状態 1 を感染状態、状態 2 を治癒して免疫を獲得した状態とみなすと、上記のモデルは SIR モデルとほぼ等価になる（SIR モデルでは感染状態のノードに隣接していても、感染するとは限らないが、本モデルでは情報を知るノードに隣接しているノードはいずれ必ず情報を知ることになる、等の差異がある）。

注 2. 隣接行列が対称行列の場合、解析対象は無向ネットワークになる。

3.2 解析手法

本節では、シミュレーションに頼らずに、解析的にモデルを分析する手法について考察する。ネットワークの状態遷移を記述するマルコフ連鎖が既約で正再帰的であれば、いずれ全てのノードが状態 2 に移行し、そこが定常状態となる。すなわち、本モデルの場合、定常状態は自明である。しかし、興味の対象は初期状態から定常状態への移行の過程（過渡特性）にあり、ネットワークの状態数 (3^N) はノード数 N の増加とともに指数関数的に増大するため、この過渡特性を厳密に解析することは不可能である。

ノード i が時刻 t で状態 k にいる確率を $p_i^{(k)}$ で表す。本稿では、 $p_i^{(k)}$ の時間変化に関する微分方程式を近似的に導出し、導出した微分方程式を数値的に解くことで、過渡特性を評価することを試みる。このため、以下の確率変数を導入する。

$$X_i^{(k)}(t) = \begin{cases} 1, & Z_i(t) = k \\ 0, & \text{otherwise} \end{cases}$$

このとき、ノード i が時刻 t で状態 k にいる確率 $p_i^{(k)}$ は以下のように表される。

$$p_i^{(k)} = E[X_i^{(k)}(t)].$$

ノード i の状態 1 への移行は、ノード j がノード i と隣接し（ノード j からノード i への有向リンクがあり）、ノード j が状態 1 にあり、かつノード i が状態 0 にあるときに生じる。また、ノード i の状態 1 から状態 2 への移行は（隣接ノードの状態にかかわらず）遷移率 λ_i で生じ、いったん状態 2 に移行すると、以降、状態 2 に留まり続ける。このことから、

$$\frac{dp_i^{(1)}(t)}{dt} = -\lambda_i E[X_i^{(1)}] + \sum_j a_{ji} \lambda_j E[X_j^{(1)}(t) X_i^{(0)}(t)]$$

$$= -\lambda_i p_i^{(1)}(t) + \sum_j a_{ji} \lambda_j E[X_j^{(1)}(t) X_i^{(0)}(t)],$$

$$\frac{dp_i^{(2)}(t)}{dt} = \lambda_i E[X_i^{(1)}] = \lambda_i p_i^{(1)}(t).$$

ここで、 $X_i^{(0)} + X_i^{(1)} + X_i^{(2)} = 1$ であることに着目し、

$$X_i(t) \stackrel{\text{def}}{=} X_i^{(1)} + X_i^{(2)}, \quad p_i(t) \stackrel{\text{def}}{=} p_i^{(1)} + p_i^{(2)},$$

とおくと ($p_i(t)$ は時刻 t でノード i が情報を知る確率を表す)、以下が得られる。

$$\frac{dp_i(t)}{dt} = \sum_j a_{ji} \lambda_j E[X_j^{(1)}(t) (1 - X_i(t))].$$

さらに、 $a_{ji} = 1$ ならば、 $X_j^{(2)}(t) = 1$ のとき $X_i(t) = 1$ である (ノード j が情報を拡散済ならば、ノード i は必ず情報を受け取っているので、ノード i の状態は 1 か 2 である) ことに注意すると、 $a_{ji} X_j^{(2)}(t) (1 - X_i(t)) = 0$ であり、したがって

$$\begin{aligned} \frac{dp_i(t)}{dt} &= \sum_j a_{ji} \lambda_j E[X_j^{(1)}(t) (1 - X_i(t))] \\ &= \sum_j a_{ji} \lambda_j E[(X_j^{(1)}(t) + X_j^{(2)}(t)) (1 - X_i(t))] \\ &= \sum_j a_{ji} \lambda_j E[X_j(t) (1 - X_i(t))] \\ &= \sum_j a_{ji} \lambda_j (p_j(t) - E[X_j(t) X_i(t)]), \end{aligned} \quad (1)$$

が得られる。

本モデルにおける各ノードの情報転送は、SNS への書き込みの投稿に相当している。時刻 t までに行われた情報転送の回数^(注1)を $L(t)$ とすると、

$$L(t) = \sum_i X_i^{(2)}(t),$$

であるから

$$E[L(t)] = \sum_i E[X_i^{(2)}(t)] = \sum_i p_i^{(2)}(t).$$

ここで、ノード i が情報を受信した時刻を T_i 、情報を受信してから情報転送までにかかった時間を τ_i とすると、

$$\begin{aligned} p_i^{(2)}(t) &= P(T_i + \tau_i \leq t) \\ &= \int_0^t P(T_i \leq t - s) P(s \leq \tau_i < s + ds) \\ &= \int_0^t \lambda_i P(T_i \leq t - s) e^{-\lambda_i s} ds \\ &= \int_0^t \lambda_i p_i(t - s) e^{-\lambda_i s} ds = \int_0^t \lambda_i p_i(s) e^{-\lambda_i(t-s)} ds. \end{aligned}$$

したがって、

$$E[L(t)] = \sum_i \int_0^t \lambda_i p_i(s) e^{-\lambda_i(t-s)} ds. \quad (2)$$

(注1)：隣接ノードには一斉に情報を転送するので、隣接ノードが複数あっても、各ノードの情報転送回数は高々1回とする。

すなわち、(1) を数値的に解いて $p_i(t)$ を求めれば、(2) から投稿数の期待値の時間変化を計算することができる。

3.3 強相関モデル

微分方程式 (1) の右辺に現れる $E[X_i(t) X_j(t)]$ (以下、相関項) は既知ではないので、(1) から $p_i(t)$ を数値的に計算するためには、何らかの仮定をおいて $E[X_i(t) X_j(t)]$ を $p_i(t) = E[X_i(t)]$ と $p_j(t) = E[X_j(t)]$ の関数で表現しておく必要がある。

以前、我々は相関項の上下限を与える以下の結果を用いて、(1) から $p_i(t)$ を数値的に計算する方法を提案した [24].

補題 3.1. $X_i(t)$ と $X_j(t)$ が非負の相関を持つ、すなわち

$$\text{Cov}[X_i(t), X_j(t)] \geq 0,$$

であるならば

$$p_i(t) p_j(t) \leq E[X_i(t) X_j(t)] \leq \min\{p_i(t), p_j(t)\}.$$

Proof. $X_i(t)$ と $X_j(t)$ が非負の相関を持つことから

$$\begin{aligned} \text{Cov}[X_i(t), X_j(t)] &= E[X_i(t) X_j(t)] - E[X_i(t)] E[X_j(t)] \geq 0. \end{aligned}$$

したがって $p_i(t) p_j(t) \leq E[X_i(t) X_j(t)]$ 。また、

$$\begin{aligned} E[X_i(t) X_j(t)] &= P(\{X_i(t) = 1\} \cap \{X_j(t) = 1\}) \\ &\leq P(\{X_i(t) = 1\}) = p_i(t), \end{aligned}$$

および

$$\begin{aligned} E[X_i(t) X_j(t)] &= P(\{X_i(t) = 1\} \cap \{X_j(t) = 1\}) \\ &\leq P(\{X_j(t) = 1\}) = p_j(t), \end{aligned}$$

であることから、 $E[X_i(t) X_j(t)] \leq \min\{p_i(t), p_j(t)\}$ も成り立つ。□

本稿では、(1) において $E[X_i(t) X_j(t)] = \min\{p_i(t), p_j(t)\}$ とする近似を「強相関近似」と呼ぶ。強相関近似を適用すると、(1) は以下ようになる。

$$\begin{aligned} \frac{dp_i(t)}{dt} &= \sum_j a_{ji} \lambda_j (p_j(t) - E[X_j(t) X_i(t)]) \\ &= \sum_j a_{ji} \lambda_j (p_j(t) - \min\{p_i(t), p_j(t)\}) \\ &= \sum_{j: p_j(t) > p_i(t)} a_{ji} \lambda_j (p_j(t) - p_i(t)). \end{aligned}$$

最後の式の和は $p_j(t) > p_i(t)$ を満たすノード j について取っている。すなわち、強相関近似のもとでは、情報は情報保持確率の高いノードから情報保持確率の低いノードへにしか流れないこととなる。なお、強相関近似は情報保持確率の下限を与える (情報拡散は最も遅く進行する) [24].

図1の左側のツリー構造のネットワークの場合、情報は情報発信元から下流のノードに一方方向に流れる。このような場合は強相関近似が厳密に成り立つ。一方、右側の格子状のネットワークの場合、一般にノード A の情報保持確率はノード B やノード C の情報保持確率より高いと考えられるが、情報発信元からノード B やノード C を経由して、ノード A に情報が到達する可能性があるため、強相関近似は厳密には成り立たない。

一方、(1) において $E[X_i(t) X_j(t)] = p_i(t) p_j(t)$ とする近似を「独立近似」と呼ぶ。独立近似は巨視的 (現象論的) に情報拡散過程を解析する研究でしばしば用いられる仮定であり、情報

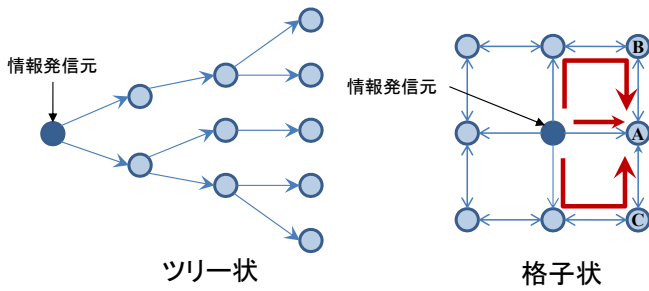


図1 情報拡散経路とネットワークトポロジ

保持確率の上限を与える（情報拡散は最も速く進行する）[24].

独立近似のもとでは、情報保持確率の大小にかかわらず、任意の方向で情報の転送が生じる．このため独立近似のもとでは、例えば図1の左側のツリー構造のネットワークにおいて、情報の逆流（右側から左側への情報の流れ）が生じる．また、図1の右側の格子状のネットワークにおいては、ノードBやノードCを経由して、ノードAに情報が到達する可能性を実際よりも過大に評価する傾向がある．

3.4 平均場近似

強相関近似と独立近似をパラメータ α を用いて線形の式で内挿すると、以下の近似が得られる．

$$E[X_i(t)X_j(t)] = (1 - \alpha)p_i(t)p_j(t) + \alpha \min\{p_i(t), p_j(t)\}.$$

上式で $\alpha = 0$ とすると独立近似に一致し、 $\alpha = 1$ とすると強相関近似に一致する．平均次数を d とし、 $\lambda \stackrel{\text{def}}{=} \sum_{i=1}^N \lambda_i / N$ を定義して、平均場近似を用いると、(1) は以下ようになる．

$$\begin{aligned} \frac{dp(t)}{dt} &= d\lambda(p_j(t) - (1 - \alpha)p(t)^2 - \alpha sp(t)) \\ &= d\lambda(1 - \alpha)p(t)(1 - p(t)). \end{aligned}$$

この微分方程式は以下の解を持つ．

$$p(t) = \frac{p(0)}{p(0) + (1 - p(0))e^{-d\lambda(1-\alpha)t}}. \quad (3)$$

(3) はロジスティック曲線である．つまり、平均場近似を適用すると、情報保持確率の時間変化は個体群成長モデル等で用いられるロジスティック曲線に従う．また、パラメータ α は、事実上、時間 t のスケールパラメータとしての役割を担い、 α が大きくなるほど（強相関極限に近づくほど）時間を緩やかに進ませる作用をもたらす．時刻 t までの投稿総数は以下の式で計算できる．

$$E[L(t)] = \int_0^t \frac{N\lambda p(0)e^{-\lambda(t-s)}}{p(0) + (1 - p(0))e^{-d\lambda(1-\alpha)s}} ds.$$

4. 数値例

4.1 ランダムグラフ

10000 個のノードを持つランダムグラフ（ノード対毎に一定の確率で無向リンクを設定して生成したネットワーク；次数分布はポアソン分布に従う）を生成し、ランダムグラフ上の1つのノードを情報発信元として、提案モデルにより情報拡散を行い、情報拡散の過程で生じる投稿数（情報転送回数）の時間変化を調べる実験を行った．状態1から状態2への遷移率 λ_i は全ノードで1に等しいとした．つまり、各ノードは情報を受け取ってから、平均1の時間で隣接ノードに情報を転送する．

図2は、リンク設定確率が0.001のランダムグラフ（ノード対毎に0.001の確率で無向リンクを設定したランダムグラフ）

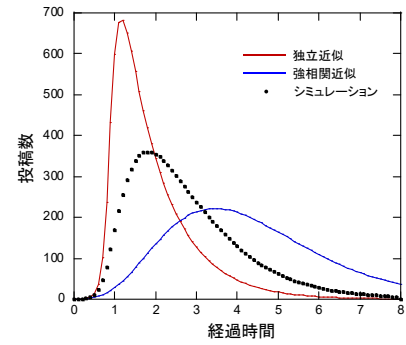


図2 投稿数の時間変化（ランダムグラフ：リンク設定確率 = 0.001）

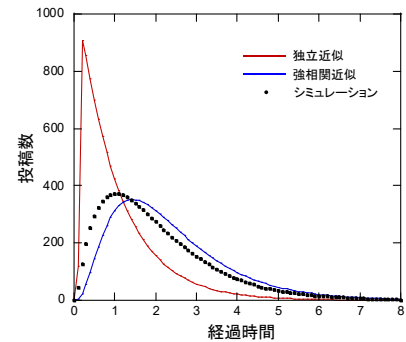


図3 投稿数の時間変化（ランダムグラフ：リンク設定確率 = 0.01）

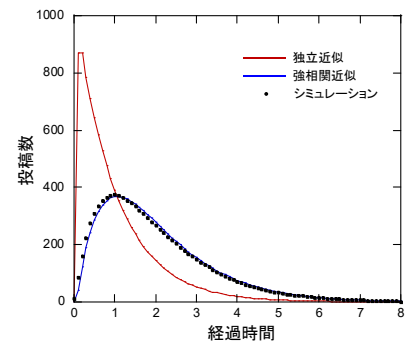


図4 投稿数の時間変化（ランダムグラフ：リンク設定確率 = 0.1）

において、0.1時間（平均情報転送時間の1/10の時間長）あたりの情報転送回数の変化を示したものである．黒丸がシミュレーション結果（5000回のシミュレーションの平均）である．情報発信元のノードの次数は13である（ネットワーク全体の平均次数はおよそ10）．情報拡散の開始より若干の時間の経過後に投稿数の急激な増加が起こり、ピークを迎えたあと次第に投稿数が減少する、という投稿数のスパイクが再現されている．

図2において、青および赤の実線は、それぞれ強相関近似および独立近似を用いて解析的に投稿数を評価した結果を表している．シミュレーション結果は強相関近似と独立近似のおよそ中間に位置している．

図3は、リンク設定確率が0.01のランダムグラフにおいて同様の実験を行い、その結果を表示したものである．情報発信元のノードの次数は105である（ネットワーク全体の平均次数はおよそ100）．同様の傾向が見て取れるが、情報拡散はより速やかに起こり（投稿数のスパイクはより急峻であり）、シミュレーション結果は強相関近似の結果に大きく近づいている．

図4は、リンク設定確率が0.1のランダムグラフにおいて同様の実験を行い、その結果を表示したものである．情報発信元のノードの次数は993である（ネットワーク全体の平均次数はおよそ1000）．シミュレーション結果は強相関近似の結果

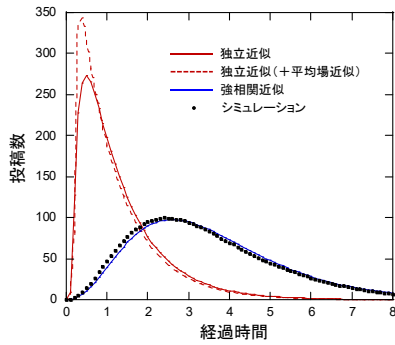


図 5 投稿数の時間変化 (Facebook: 情報発信元次数 = 1)

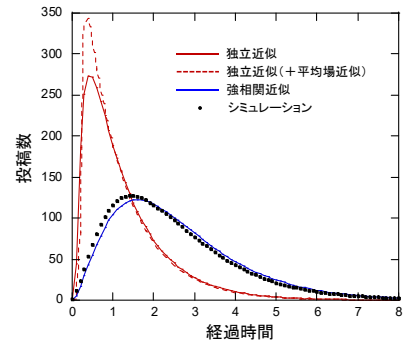


図 7 投稿数の時間変化 (Facebook: 情報発信元次数 = 1045)

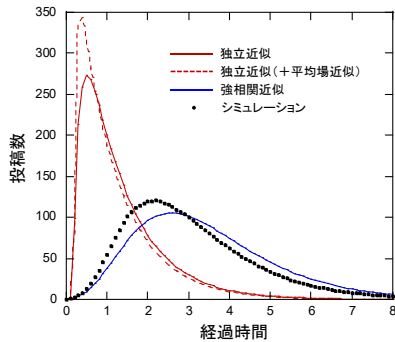


図 6 投稿数の時間変化 (Facebook: 情報発信元次数 = 30)

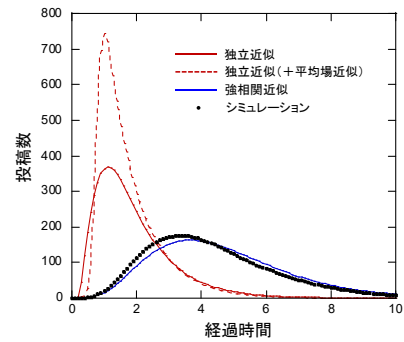


図 8 投稿数の時間変化 (Twitter: 情報発信元次数 = 1)

とほぼ重なる。すなわち、ノード数に対してリンク数の大きいネットワークほど、強相関近似が良く合致するという傾向が確認できる。また、いずれの結果も投稿数の立ち上がりは急峻であるものの、減少は立ち上がりに対して緩やかであり、これは (power-law 的に減衰するという) 実際の現象 [7] と合致しているように見える。

4.2 Facebook

Facebook のトポロジーデータ [25] を用いて、同様の実験を行った。図 5 は、次数が 1 のノードを情報発信元として提案モデルにより情報拡散を行い、情報拡散の過程で生じる投稿数 (情報転送回数) の時間変化を調べたものである。投稿数の時間変化は比較的緩やかであるが、増加後ピークを迎え、やがて減衰するという様子が再現されている。また、Facebook では、強相関近似による解析結果がシミュレーション結果を正確に再現し、独立近似はシミュレーション結果を再現しないことが確認された。3.4 節で説明した手法により、「独立近似+平均場近似」を用いて評価した結果も図に示した。独立近似に平均場近似を適用すると、投稿数のスパイクは最も急峻であるが、シミュレーション結果との差異は最も大きい。

図 6 は次数が 30 のノードを情報発信元として情報拡散を行った結果、図 7 は次数が 1045 のノードを情報発信元として情報拡散を行った結果をそれぞれ示している。どの結果においても、強相関近似により解析的に得た結果は、シミュレーション結果を概ね正しく再現している。これは、Facebook の情報拡散はツリー状のネットワークのように進行していることを意味している。また、情報発信元の次数が大きいほど投稿数のスパイクはより急峻である。つまり、スパイクの形状は情報発信元に依存しており、ネットワーク構造を陽に考慮するモデル化が本質的に重要であることがわかる。

4.3 Twitter

最後に、Twitter のトポロジーデータ [25] を用いて、同様の実験を行った^(注2)。図 8 は、次数が 1 のノードを情報発信元と

して提案モデルにより情報拡散を行い、情報拡散の過程で生じる投稿数 (情報転送回数) の時間変化を調べたものである。やはり、投稿数のスパイクが再現されている。また、強相関近似および独立近似を用いて解析的に投稿数を評価した結果も示した。Twitter においても、強相関近似による解析結果は、正確にシミュレーションを再現する。また、3.4 節で説明した手法により、「独立近似+平均場近似」で解析した結果も図に示した。Twitter は有向ネットワークであるため、全ノードには情報が到達していないが、平均場近似では全ネットワークに情報が到達することを想定するので、投稿数の総数がやや過大に見積もられる。

図 9 は次数が 50 のノードを情報発信元として情報拡散を行った結果、図 10 は次数が 477 のノードを情報発信元として情報拡散を行った結果をそれぞれ示している。やはり、情報発信元の次数が大きいほど投稿数のスパイクはより急峻であり、投稿数のスパイク形状の情報発信元に依存性が確認できる。また、いずれの結果においても、強相関近似により解析的に得た結果は、シミュレーション結果を正しく再現している。これは、Facebook の場合と同様、Twitter においても、情報の拡散はツリー状のネットワークの場合のように (概ね上流から下流に向かって) 進行していることを意味している。

5. 結論

本研究では、SNS への投稿件数のスパイク現象を、ネットワーク構造を陽に取り入れた SIR モデルにより分析した。特に、モデルをシミュレーションで評価するだけでなく、SNS の投稿数の時間推移を解析的に評価する手法について考察した。その結果、SNS の投稿数のスパイク現象はモデル上で再現できること、ネットワーク構造を陽に考慮することが重要であること、強相関近似を用いることで解析的な評価が可能であること

(注2) : ノード数が 8 万、リンク数が 200 万を超える巨大なネットワークであっ

たため、シミュレーションおよび解析はノード数を 1 万に絞った部分ネットワーク上で行った。

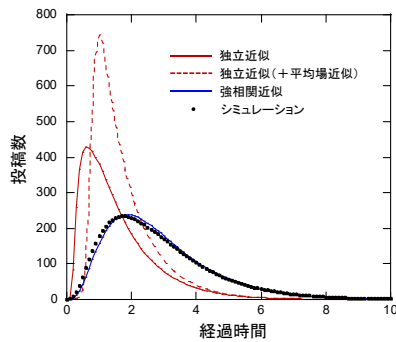


図9 投稿数の時間変化 (Twitter: 情報発信元次数 = 50)

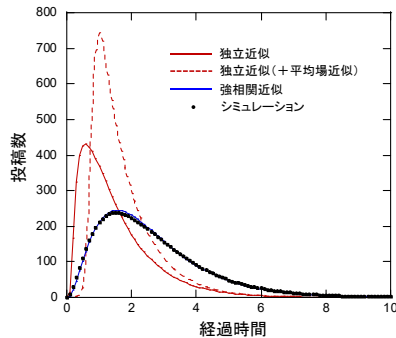


図10 投稿数の時間変化 (Twitter: 情報発信元次数 = 477)

などを明らかにした。

今後は、実際の Tweet 数の時間推移データとのフィッティング、情報は所持しているが SNS に投稿しない状態のモデル化、複数の情報が (相互に影響しながら) 同時に拡散する場合のモデル化等について検討を行いたい。また、投稿数のスパイクが複数個生じる現象を説明するために、拡散の途中で情報の突然変異が生じ、突然変異した情報が新規情報として新たに拡散していく現象をモデルに取り入れる予定である。

文 献

- [1] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst, "Patterns of cascading behavior in large blog graphs," Proceedings of the 2007 SIAM International Conference on Data Mining, 2007.
- [2] M. Cha, F. Benevenuto, Y.Y. Ahn, and K. Gummadi, "Delayed information cascades in flickr: Measurement, analysis, and modeling," Computer Networks, vol.56, no.3, pp.1066–1076, 2012.
- [3] Y. Okada, K. Ikeda, *et al.*, "SIR-extended information diffusion model of false rumor and its prevention strategy for twitter," Journal of Advanced Computational Intelligence and Intelligent Informatics, vol.18, no.4, pp.598–607, 2014.
- [4] C. Bauchhage, F. Hadiji, and K. Kersting, "How viral are viral videos?," 9th International AAAI Conference on Web and Social Media, 2015.
- [5] J. Cheng, L. Adamic, J. Kleinberg, and J. Leskovec, "Do cascades recur?," WWW '16, 2016.
- [6] J. Yang and J. Leskovec, "Modeling information diffusion in implicit networks," Proceedings of the 2010 IEEE International Conference on Data Mining, pp.599–608, 2010.
- [7] Y. Matsubara, Y. Sakurai, B. Prakash, L. Li, and C. Faloutsos, "Maximizing the spread of influence through a social network," KDD '12, pp.6–14, 2012.
- [8] A. Guille and H. Hacid, "A predictive model for the temporal dynamics of information diffusion in online social networks," WWW '12, pp.1145–1152, 2012.
- [9] C. Bauchhage, K. Kersting, and F. Hadiji, "Mathematical models of fads explain the temporal dynamics of internet

memes," 7th International AAAI Conference on Web and Social Media, pp.22–30, 2013.

- [10] D. Kempe, J. Kleinberg, and E. Tardos, "Rise and fall patterns of information diffusion: model and implications," KDD '03, pp.137–146, 2003.
- [11] D. Watts and P. Dodds, "Influentials, networks, and public opinion formation," Journal of Consumer Research, vol.34, no.4, pp.441–458, 2007.
- [12] W. Galuba, K. Aberer, D. Chakraborty, Z. Despotovic, and W. Kellerer, "Outtweeting the twitterers - predicting information cascades in microblogs," WOSN'10, 2010.
- [13] N. Barbieri, F. Bonchi, and G. Manco, "Topic-aware social influence propagation models," Knowledge and Information Systems, vol.37, no.3, pp.555–584, 2013.
- [14] W.O. Kermack and A.G. McKendrick, "A contribution to the mathematical theory of epidemics," Proceedings of the Royal Society A, vol.115, no.772, pp.700–721, 1927.
- [15] R. Pastor-Satorras, C. Castellano, P.V. Mieghem, and A. Vespignani, "Epidemic processes in complex networks," Reviews of Modern Physics, vol.87, pp.926–979, 2015.
- [16] R. Anderson and R.M. May, Infectious Diseases in Humans, Oxford University Press, Oxford, 1992.
- [17] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," Phys. Rev. Lett., vol.86, no.14, pp.3200–3203, 2001.
- [18] M. Boguna and R. Pastor-Satorras, "Epidemic spreading in correlated complex networks," Phys. Rev., vol.E66, p.047104, 2002.
- [19] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," ACM Transactions on Information and System Security, vol.10, no.4, pp.13.1–13.26, 2008.
- [20] P.V. Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," IEEE/ACM Trans. Networking, vol.17, no.1, pp.1–14, 2008.
- [21] K. Sharkey, "Deterministic epidemic models on contact networks: Correlations and unbiological terms," Theor. Popul. Biol., vol.79, no.4, pp.115–129, 2001.
- [22] N. Schwartz and L. Stone, "Exact epidemic analysis for the star topology," Phys. Rev., vol.E87, p.042815, 2013.
- [23] P.V. Mieghem, Performance Analysis of Complex Networks and Systems, Cambridge University Press, Cambridge, England, 2014.
- [24] 南川雅人, 塩田茂雄, "複雑ネットワーク上の情報拡散過程におけるノード相関の影響," 電子情報通信学会 コミュニケーションクオリティ研究会, CQ2017-58, pp.43–58, 2017.
- [25] "Stanford large network dataset collection." <http://snap.stanford.edu/data/>.