

SNS 投稿件数推移分析のための情報拡散モデルと強相関近似解析*

塩田 茂雄* 南川 雅人* 中島 圭佑†

* 千葉大学 大学院融合理工学府 都市環境システムコース

† 千葉大学 工学部 都市環境システム学科

概要: 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 モデルを拡張し、SNS への投稿数の時間推移を解析的に近似評価する手法を提案する。提案手法は、隣接ノード間では情報保持状態にある事象の相関性が高いとする仮定を用いており（強相関近似）、この仮定の妥当性をシミュレーションにより検証する。

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, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 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、感染症の数理モデルとして使われている SIR モデル [14, 15] を拡張し、SNS への投稿件数の時間推移を説明する手法について検討する。提案手法により、情報の拡散が始まってからネットワーク内の情報拡散が終了するまでに、情報拡散を行った単位時間あたりのノード数（単位時間あたりの SNS の投稿数に相当）が次第に増加し、ピークを迎えたあと、次第に減少して落ち着く様子の再現性を確認するとともに、数理モデルに基づいて投稿数の時間推移を解析的に評価する手法についてあわせて検討する。検討の結果、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、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などを明らかにする。

以下、2 章では、提案する数理モデルおよび解析的アプローチの詳細について説明する。また、3 章では、幾つかの数値実験を通して、提案モデルの特徴や解析的アプローチの妥当性について示す。最後に、4 章において結論を述べる。

2. 情報拡散モデル

2.1. 定義

N 個のノードからなる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。ノードは以下のいずれかの状態を取る。

状態 0: 情報を知らない (SIR モデルにおける Susceptible 状態)

状態 1: 情報は知っているが、拡散しない (SIR モデルにおける Infected 状態: 感染力なし)

状態 2: 情報を知っており、将来拡散予定 (SIR モデルにおける Infected 状態: 感染力あり)

状態 3: 情報を知り、拡散済み (SIR モデルにおける Recovered 状態)

時刻 0 では情報発信元のノードは状態 1 にあり、それ以外の全てのノードは状態 0 にある。ノード i は (隣接ノードから) 情報を受信することで、状態 0 から確率 $1 - q_i$ で状態 1 に、確率 q_i で状態 2 に遷移する。状態 2 に遷移後は、平均 $1/\lambda_i$ の指数分布に従う時間経過後に、全ての隣接ノードに同時に情報転送を行い、状態 3 に遷移する。状態 3 に遷移後は、そのまま状態 3 に留まる。状態 0 から状態 1 に遷移したノードは、そのまま状態 1 に留まるが、別の隣接ノードから情報を受信すると確率 q_i で状態 2 に遷移する。

注 1. 状態 1 は、感染はしたものの、他人への感染力がない状態に相当する。通常の SIR モデルは、このような状態を扱わない。しかし、この状態がないと、連結ネットワークでは全てのノードがいずれ感染して周囲にウイルスを拡散させることとなる。これは、最終的に全ユーザが SNS への投稿を行うことと同値であり、投稿を行わないユーザが存在する現実と異なる。このため、本モデルでは、SIR モデルを修正し、感染力のない状態（他のユーザによる書き込みは見たが、自ら投稿を行わない状態）を取り入れることとした。

2.2. 解析的アプローチ

隣接行列を $A = \{a_{ij}\}$ とし、ノード i からノード j への情報転送によりノード j が（状態 0 から）状態 1 に遷移する場合に 0、状態 2 に遷移する場合に 1 を取る確率変数を y_{ij} とする（ $a_{ij} = 0$ のときは $y_{ij} = 0$ とする）。本稿では $Y \stackrel{\text{def}}{=} \{y_{ij}\}_{i,j=1,\dots,N}$ を反応行列と呼ぶ。反応行列の各要素は互いに独立であるとする。時刻 t においてノード i がとる状態の番号を $Z_i(t)$ で表すこととすると、時刻 t でのネットワークの状態は、反応行列 Y および各ノードの状態の組 $(Z_1(t), Z_2(t), \dots, Z_N(t))$ で記述され、 $(Z_1(t), Z_2(t), \dots, Z_N(t))$ の遷移は連続時間マルコフ連鎖に従う。 $(Z_1(t), Z_2(t), \dots, Z_N(t))$ の取り得る状態数は 4^N であるので、 $(Z_1(t), Z_2(t), \dots, Z_N(t))$ の遷移を厳密に追うことは、反応行列 Y の実現値が既知であっても困難である。

本節では、与えられた反応行列 Y の実現値のもとで、ノード i が時刻 t で状態 k にいる確率 $p_i^{(k)}(t|Y)$ の時間変化を記述する微分方程式を導出する。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。

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

ノード i の状態 2 への遷移は、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 2 にあり、ノード i が状態 0 か状態 1 にあり、さらに $y_{ij} = 1$ のときに生じる。また、ノード i の状態 2 から状態 3 への遷移は（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 3 に遷移すると、以降、状態 3 に留まり続ける。このことから、

$$\begin{aligned} \frac{dp_i^{(2)}(t|Y)}{dt} &= -\lambda_i E[X_i^{(2)}(t)|Y] + \sum_j a_{ji} y_{ij} \lambda_j E[X_j^{(2)}(t)(X_i^{(0)}(t) + X_i^{(1)}(t))|Y] \\ &= -\lambda_i p_i^{(2)}(t|Y) + \sum_j a_{ji} y_{ij} \lambda_j E[X_j^{(2)}(t)(X_i^{(0)}(t) + X_i^{(1)}(t))|Y], \end{aligned} \quad (1)$$

$$\frac{dp_i^{(3)}(t|Y)}{dt} = \lambda_i E[X_i^{(3)}(t)|Y] = \lambda_i p_i^{(2)}(t|Y). \quad (2)$$

ここで、

$$X_i(t) \stackrel{\text{def}}{=} X_i^{(2)}(t) + X_i^{(3)}(t), \quad p_i(t|Y) \stackrel{\text{def}}{=} p_i^{(2)}(t|Y) + p_i^{(3)}(t|Y),$$

とおくと、(1) と (2) を足し合わせることで、以下が得られる。

$$\frac{dp_i(t|Y)}{dt} = \sum_j a_{ji} y_{ij} \lambda_j E[X_j^{(2)}(t)(1 - X_i(t))|Y].$$

さらに、 $a_{ji} y_{ji} = 1$ ならば、 $X_j^{(3)}(t) = 1$ のとき $X_i(t) = 1$ である（ $a_{ji} y_{ji} = 1$ ならば、ノード j が情報を拡散済ならば、ノード i は必ず情報を受信しかつ拡散予定か拡散済みなので、ノード i の状態は 2 か 3 である）ことに注意すると、 $a_{ji} y_{ji} X_j^{(3)}(t) (1 - X_i(t)) = 0$ であり、したがって

$$\begin{aligned} \frac{dp_i(t|Y)}{dt} &= \sum_j a_{ji} y_{ji} \lambda_j E[X_j^{(2)}(t) (1 - X_i(t))|Y] = \sum_j a_{ji} y_{ji} \lambda_j E[(X_j^{(2)}(t) + X_j^{(3)}(t)) (1 - X_i(t))|Y] \\ &= \sum_j a_{ji} y_{ji} \lambda_j E[X_j(t) (1 - X_i(t))|Y] = \sum_j a_{ji} y_{ji} \lambda_j (p_j(t|Y) - E[X_j(t)X_i(t)|Y]), \end{aligned} \quad (3)$$

が得られる。(3)に含まれる $E[X_j(t)X_i(t)|Y]$ （以下、相関項）が $p_i(t|Y)$ や $p_j(t|Y)$ で表現できれば、(3)は $\{p_i(t|Y)\}_{i=1}^N$ に関する連立微分方程式となり、数値的に解くことが可能である。 $p_i(t|Y)$ が求めれば、

$$p_i(t) \stackrel{\text{def}}{=} E[X_i(t)] = E[p_i(t|Y)] = E[E[X_i(t)|Y]].$$

により、時刻 t においてノード i が状態 2 か状態 3 にいる確率が求まる。

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¹ (投稿数: 状態 2 から状態 3 に遷移したノード数) を $L(t)$ とすると、

$$L(t) = \sum_i X_i^{(3)}(t),$$

であるから

$$E[L(t)] = \sum_i E[X_i^{(3)}(t)] = \sum_i p_i^{(3)}(t).$$

ここで、ノード i が状態 2 に遷移した時刻を T_i 、状態 2 に遷移してから状態 3 に遷移するまでにかかった時間を τ_i とすると、

$$\begin{aligned} p_i^{(3)}(t) &= P(T_i + \tau_i \leq t) = \int_0^t P(T_i \leq t-s)P(s \leq \tau_i < s+ds) = \int_0^t \lambda_i P(T_i \leq t-s)e^{-\lambda_i s} ds \\ &= \int_0^t \lambda_i p_i(t-s)e^{-\lambda_i s} ds = \int_0^t \lambda_i p_i(s)e^{-\lambda_i(t-s)} ds. \end{aligned}$$

したがって、

$$E[L(t)] = \sum_i \int_0^t \lambda_i p_i(s)e^{-\lambda_i(t-s)} ds. \quad (4)$$

2.3. 強相関近似

実際には $p_i(t|Y)$ が満たす微分方程式

$$\frac{dp_i(t|Y)}{dt} = \sum_j a_{ji} y_{ji} \lambda_j (p_j(t|Y) - E[X_i(t)X_j(t)|Y]).$$

の右辺に現れる相関項 $E[X_i(t)X_j(t)|Y]$ を厳密に知ることはできないため、前節の手順で $L(t)$ を評価するためには、相関項を近似評価する必要がある。相関項の上下限は次の結果により評価できる。

補題 2.1. $X_i(t)$ と $X_j(t)$ が非負の相関を持つ、すなわち

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

であるならば

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

証明. $X_i(t)$ と $X_j(t)$ が非負の相関を持つことから

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

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

$$E[X_i(t)X_j(t)|Y] = P(\{X_i(t) = 1\} \cap \{X_j(t) = 1\}|Y) \leq P(\{X_i(t) = 1\}|Y) = p_i(t|Y),$$

および

$$E[X_i(t)X_j(t)|Y] = P(\{X_i(t) = 1\} \cap \{X_j(t) = 1\}|Y) \leq P(\{X_j(t) = 1\}|Y) = p_j(t|Y),$$

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

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。

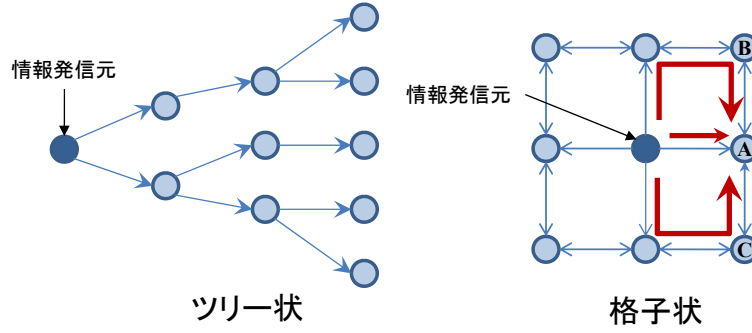


図 1: 情報拡散経路とネットワークトポロジ

本稿では、相関項をその上限式で代用する，すなわち $E[X_j(t)X_i(t) | Y] = \min\{p_j(t | Y), p_i(t | Y)\}$ とする近似を採用する．これを強相関近似と呼ぶ．

強相関近似を適用すると，(3) は以下ようになる．

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

最後の式の和は $p_j(t|Y) > p_i(t|Y)$ 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 ($p_j(t|Y)$) 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，相関項が大きいほど情報拡散はゆっくりと進行するので，強相関近似は各時刻における情報保持確率の下限を与える．

図 2 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，(3) において $E[X_i(t)X_j(t)|Y] = p_i(t|Y)p_j(t|Y)$ 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，各時刻における情報保持確率の上限を与える（情報拡散は最も速く進行する）．

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 4 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．

2.4. 反応行列 Y に関する期待値の取り方

投稿件数の評価のためには， $E[X_i(t)] = E[p_i(t | Y)]$ を求める必要があるが，実際には全ての反応行列 Y の実現値について $p_i(t | Y)$ を求めて平均を取ることは困難である．最も単純なアプローチは

$$\frac{dp_i(t|Y)}{dt} = \sum_j a_{ji}y_{ji}\lambda_j (p_j(t|Y) - E[X_i(t)X_j(t)|Y]),$$

の両辺の期待値をとり，

$$E[y_{ji}p_j(t|Y)] \approx E[y_{ji}]p_j(t), \quad E[y_{ji}E[X_i(t)X_j(t)|Y]] \approx E[y_{ji}]E[X_i(t)X_j(t)],$$

を仮定して，以下の微分方程式

$$\frac{dp_i(t)}{dt} = \sum_j a_{ji}E[y_{ji}]\lambda_j (p_j(t) - E[X_i(t)X_j(t)]) = \sum_j a_{ji}q_{ji}\lambda_j (p_j(t) - E[X_i(t)X_j(t)]), \quad (5)$$

を得て，これに対し，強相関近似を適用する手法である．しかし，次章で示すように，この方法で計算した結果はシミュレーション結果を再現しない．

もう一つの手法は、ランダムに1つ（ないしは複数）選んだ反応行列の実現値について $p_i(t|Y)$ を計算し、その結果を $E[X_i(t)]$ の近似値として用いるものである。やや意外であるが、次章で示すように、反応行列のどの実現値を選んでも結果はおおむね等しく、この手法は充分有効に機能する。

2.5. 平均場近似

強相関近似と独立近似をパラメータ α を用いて線形補間すると、以下の近似が得られる。

$$E[X_i(t)X_j(t)|Y] = (1-\alpha)p_i(t|Y)p_j(t|Y) + \alpha \min\{p_i(t|Y), p_j(t|Y)\}. \quad (6)$$

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。(3)に(6)を代入し、さらに2.4節で示したように先に反応行列の期待値をとって平均場近似を適用すると以下が得られる。

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

ここで、 d は平均（入）次数、 λ は $\lambda \stackrel{\text{def}}{=} \sum_{i=1}^N \lambda_i/N$, $q \stackrel{\text{def}}{=} \sum_{i=1}^N q_i/N$ で定義される変数である。この微分方程式は以下の解を持つ。

$$p(t) = \frac{p(0)}{p(0) + (1-p(0))e^{-dq\lambda(1-\alpha)t}}. \quad (7)$$

(7) 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。

$$E[L(t)] = \int_0^t \frac{N\lambda p(0)e^{-\lambda(t-s)}}{p(0) + (1-p(0))e^{-dq\lambda(1-\alpha)s}} ds. \quad (8)$$

反応行列の各要素が1未満の場合、(情報到達性が損なわれて)情報拡散を行わないノードが出現し、その分、全体の投稿数が減少する。しかし、上記のアプローチではこの効果を取り入れることができないため、特に反応行列の各要素が1未満の場合はシミュレーション結果との一致度は著しく低下する。

3. 数値例

2章で述べた近似解析法の精度を数値的に検証するため、インターネット上に公開されている Facebook と Twitter の部分トポロジーデータ (Facebook: ノード数 4039, リンク数 88234, Twitter: ノード数 81306, リンク数 1768149) [16] を用いて、1つのノードから情報を拡散させるシミュレーション実験を行った。なお、Facebook は無向グラフ、Twitter は有向グラフであると仮定した。シミュレーション開始時点において、情報発信ノードは状態2にセットし、その後、情報発信ノードから情報が拡散していき、それを受けて周囲のノードが次々と情報を拡散していく様子を確認した。

3.1. 投稿数の時間変化: $q = 1$

図2は Facebook のトポロジーデータを用いて、次数30のノードから情報拡散を行った場合の、0.1時間幅内の投稿数（拡散を行ったノード数）の時間推移を示したものである。黒丸のプロットはシミュレーション結果、実線が強相関近似、破線が独立近似による解析結果を示している。パラメタ値として、 $\lambda_i = 1$, $q_i = E[y_{ji} | a_{ji} = 1] = 1$ を仮定した²。 $q_i = 1$ の場合は、情報を受信したノードは必ず情報を転送するので、反応行列の要素は全て1であり、反応行列に関する平均操作の必要はない。シミュレーション結果は5000回の平均を取ったものである。図からわかるように、強相関近似による結果は、シミュレーション結果を大変良く再現している。一方、独立近似による結果は、シミュレーション結果と大きな隔たりがある。図3は Twitter のトポロジーデータを用いて、次数1111のノードから情報拡散を行った場合の結果である。同様の傾向が確認できる。つまり、無向、有向を問わず、強相関近似は非常に有望な解析手段となりうると考えられる。

²時間の単位を例えば「日」ととると、 $\lambda = 1$ は SNS で投稿を見てから自ら書き込みを行うまでの平均時間間隔が1日であることに相当する。

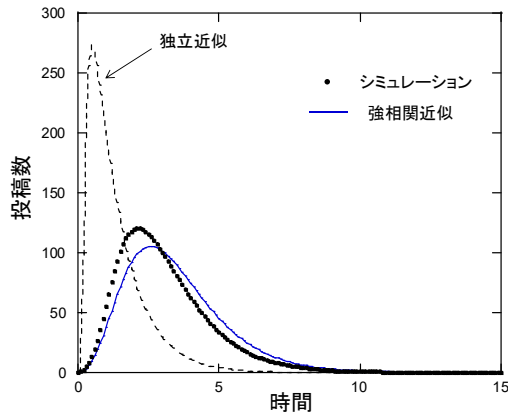


図 2: 投稿数の時間推移 (Facebook: $q = 1$)

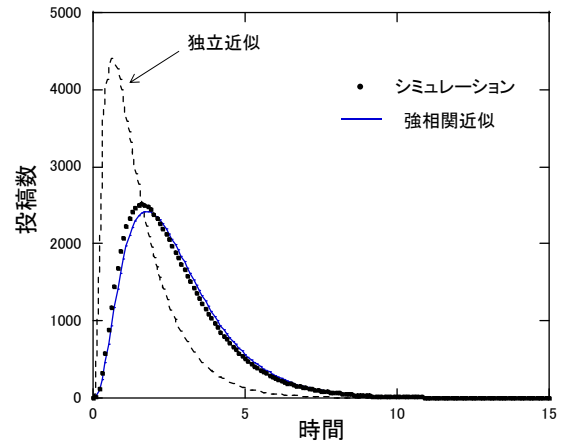


図 3: 投稿数の時間推移 (Twitter: $q = 1$)

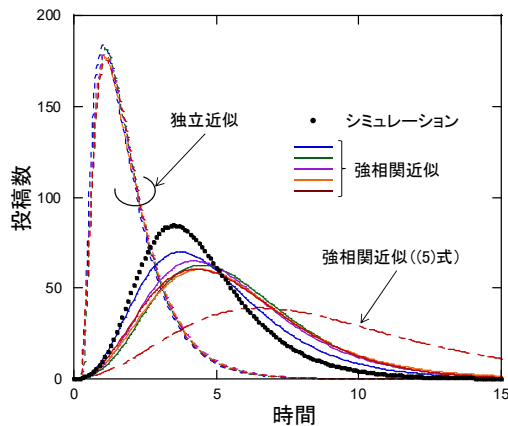


図 4: 投稿数の時間推移 (Facebook: $q = 0.3$)

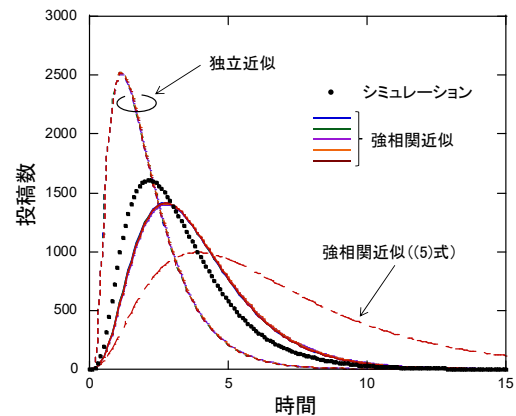


図 5: 投稿数の時間推移 (Twitter: $q = 0.3$)

3.2. 投稿数の時間変化: $q_i = 0.3$

次に, $q_i = 0.3$ として同様の実験を行った. 結果を図 4 (Facebook) および図 5 (Twitter) に示す. 黒丸のプロットはシミュレーション結果, 実線が強相関近似, 破線が独立近似による解析結果を示している. 解析では, ランダムに選んだ 5 通りの反応行列の実現値に対して結果を求めており, 図には各結果をそれぞれ異なる色の線 (5 本の実線および 5 本の破線) で示した. また, 先に反応行列の期待値を取ってから微分方程式を解いた場合 ((5) 式を用いた場合) の結果を, 強相関近似についてのみ一点鎖線で示した. $q_i = 1$ の場合と比べると, 全体的に書き込み数は減少するが, 3 割までは減少しない. これは, 複数の隣接ノードから情報を得ることで, その都度確率 q_i のベルヌーイ試行で状態 2 に遷移するためと考えられる. $q = 1$ の場合と同様, 独立近似よりも強相関近似の方がシミュレーションに近い結果を与える.

加えて, 図からは反応行列の実現値を変えても解析結果にさほど差異は見られないことが確認できる. 特に, Twitter ではほとんど反応行列の実現値依存性は見られない. これは Twitter の方がシミュレーションで用いたネットワークのノード数, リンク数が大きく, ある種の大数の法則により, 十分な平均化が行われた結果ではないかと考えられる. 反応行列の期待値を取ってから微分方程式を解いた ((5) 式を用いた) 場合の結果は, シミュレーションと大きな隔たりがある. その意味で, 反応行列の実現値をランダムに 1 つ選んで計算することは合理的であると考えられる.

3.3. 各サンプルパスの比較

3.1 節および 3.2 節で示したシミュレーション結果は, 5000 回の平均を取ったものである. 3.2 節で述べたように反応行列の実現値依存性は小さいことから, 1 回毎のシミュレーション結果の違いも少ないのではないかと予想される. その点を確認するため, 図 6 (Facebook) および図 7 (Twitter) に乱数の発生系列を変えてシミュレーションを 3 回行い, 各シミュレーションにおける投稿数の時間変化 (3 つのサンプルパス) をそれぞれ図示した. 図からわかるように, サンプルパスの違いは思いのほか大きい. Twitter は約 80000 ノードでのシミュレーション結果であり, その規模のネットワークでも, 大数の法則から期待されるような平均化は必ずしも行われ

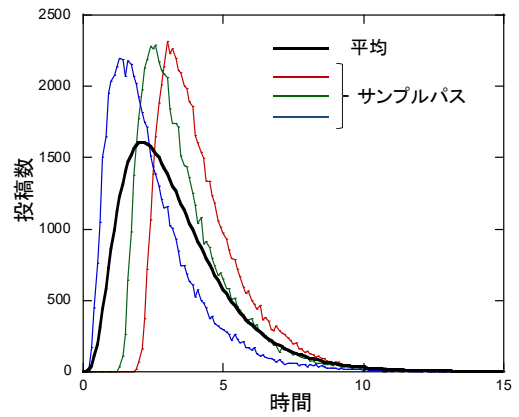
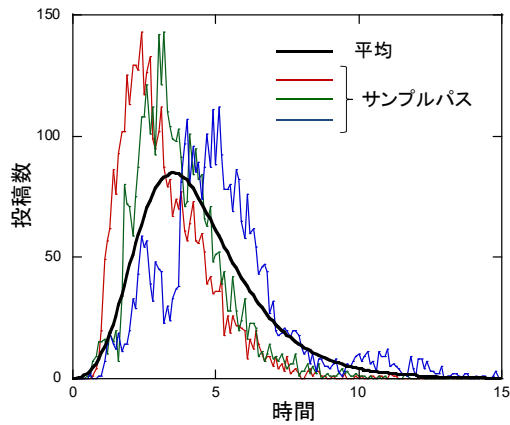


図 6: 投稿数時間推移のサンプルパス (Facebook: $q = 0.3$) 図 7: 投稿数時間推移のサンプルパス (Twitter: $q = 0.3$)

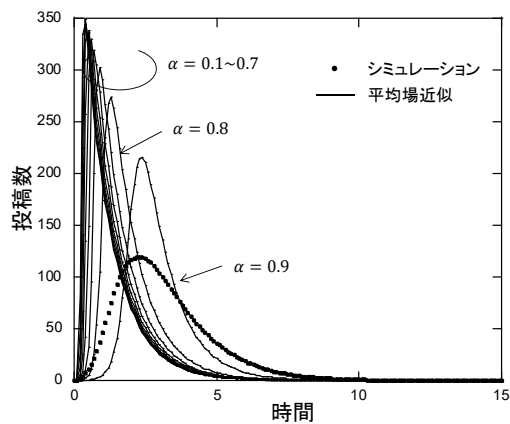


図 8: 平均場近似 (Facebook: $q = 1$)

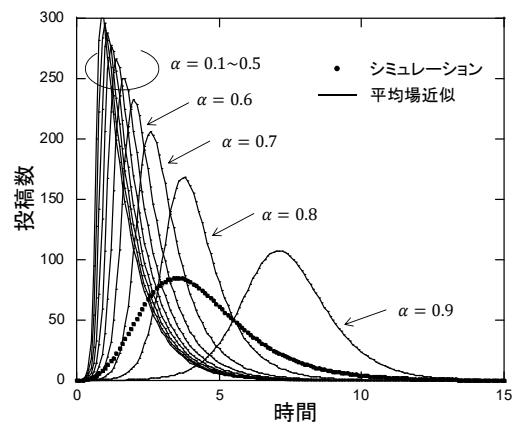


図 9: 平均場近似 (Facebook: $q = 0.3$)

ていない。本モデルにおける投稿数には、一般的に「増加後ピークを迎えて減少に転じる」という時間変化がみられるが、Facebook におけるシミュレーションでは、細かな時間変動が確認でき、複数のピークが見られるサンプルパスも存在している。

なお、サンプルパス毎の差異が大きい原因は、(1) 情報の発信を 1 ノードから行っていること、(2) 情報発信ノードは、シミュレーション開始以降、指数分布に従う時間経過後に周囲のノードに情報を拡散することにあると考えられる。情報発信を複数のノードから行う場合や、情報拡散をシミュレーションの開始と同時に進行する場合には、サンプルパスの差異は図 6 や図 7 より小さくなると想定される。

3.4. 平均場近似

最後に平均場近似 ((8) 式) を適用して得られた投稿数の時間推移の結果を図 9 (Facebook: $q = 1$)、図 10 (Facebook: $q = 0.3$)、図 10 (Twitter: $q = 1$)、図 11 (Twitter: $q = 0.3$) に示す。平均場近似はパラメタ α を含むため、図には α の値を 0 から 0.9 まで 0.1 ずつ変えた結果を示した。パラメタの値をどのように変えてもシミュレーション結果との差異は顕著であり、ネットワーク構造 (トポロジー) の情報を用いない単なる平均場近似では、実際の投稿数の時間推移を再現することはやはり困難である。

4. 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

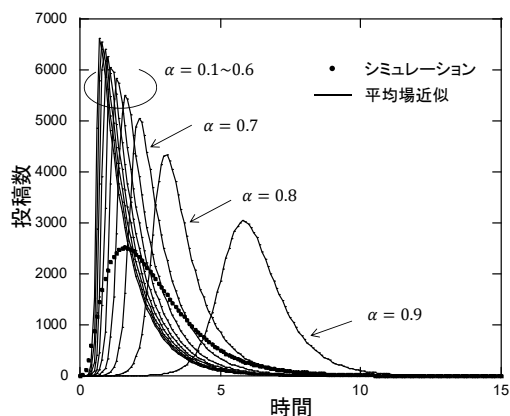


図 10: 平均場近似 (Twitter: $q = 1$)

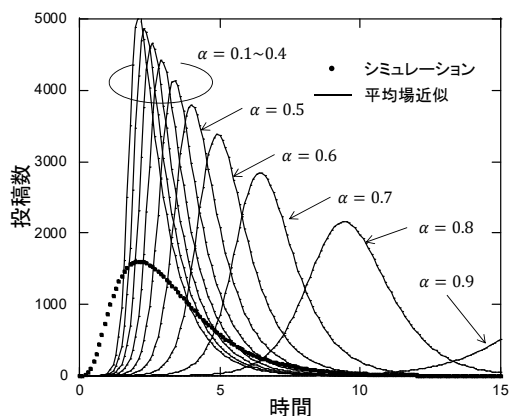


図 11: 平均場近似 (Twitter: $q = 0.3$)

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った。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。

参考文献

- [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. Bauckhage, 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. Bauckhage, 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] "Stanford large network dataset collection." <http://snap.stanford.edu/data/>.