

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

南川 雅人[†] 塩田 茂雄[†]

[†] 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: †adda2062@chiba-u.jp, ††shioda@faculty.chiba-u.jp

あらまし Twitter 等のソーシャルメディアにおける話題や噂の拡がり方は、複雑ネットワーク上での情報拡散現象とみなせる。本研究では複雑ネットワーク上の情報拡散過程を微分方程式によって微視的に表現する数理モデルを用いて、ノード相関（隣接するノードの状態の関係性）が情報拡散速度に及ぼす影響について考察する。その結果、ノード相関が情報拡散過程に大きく影響すること、多くの既存研究で使われる独立モデル（ノード相関がないとするモデル）では情報拡散過程が正しく表現できず、本来、情報が後から伝わるはずのノードに情報が先に伝わる「因果律の破れ」が生じることなどを明らかにする。

キーワード 情報拡散, 複雑ネットワーク, 微分方程式, ノード相関, 因果律

Node Correlation Influence on Information Spreading Processes in Undirected Graphs

Masato MINAMIKAWA[†] and Shigeo SHIODA[†]

[†] Graduate School of Science and Engineering, Chiba University, 1-33 Yayoi, Inage, Chiba, 263-8522 Japan

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

Abstract Topics and rumors spreading in social media such as Twitter can be regarded as information spreading phenomena on complex networks. In this research, we consider the influence of node correlation on information-spreading speed by using mathematical model which microscopically expresses information diffusion process on complex network by differential equation. We find that node correlation greatly affects the information spreading process. We also find that so-called independent models do not accurately express the information spreading processes and the use of independent models may lead to causality breaking.

Key words Information Spreading, Complex Networks, Differential Equation, Node Correlation, Causality

1. ま え が き

1990年代後半からの複雑ネットワーク研究の勃興と急速な進展とともに、現実社会の人間関係や、インターネット、特にオンラインソーシャルネットワーク上での人間関係を複雑ネットワークとしてとらえ、その構造的な特徴を分析する研究が行われている。一方、従来より社会学や経済学分野において、流行、話題、うわさなどの社会現象の広がり方に関する研究が進められていたが、オンラインソーシャルネットワーク等の普及により社会現象の広がり方が変質するに伴い、人間関係を複雑ネットワークとしてモデル化し、数理的に社会現象の広がり方を解析する試みが多数なされている [1]。

本稿では、複雑ネットワーク上での情報の拡散に着目する。情報の拡散は感染症の拡散と類似しており、SIS モデルや SIR モデル [2] などを用いた様々な研究が古くから行われている。SIS モデルや SIR モデルによる研究は、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解析するものが多い [3] が、近年、次数分布の情報を用いるもの [4], [5]、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 [6]~[10] など、ネットワーク構造を考慮できるモデルを用いた手法に研究の関心が移りつつある [1]。

本稿では隣接行列を用いてネットワークの構造を陽に考慮する情報拡散モデルに着目する。各ノードは「情報を持たない」、「情報を持つ」の 2 状態のみを有し、「情報を持たない」から「情報を持つ」への状態遷移のみが生ずる。これは最も単純な情報拡散モデルである。ノード i の状態を X_i で表すこととすると、 X_i の期待値 $E[X_i]$ の時間変化に関する微分方程式が導かれるが、微分方程式には隣接するノードの状態の積の期待値 $E[X_i X_j]$ が含まれる。本稿ではこれを相関項と呼ぶ。多くの既存研究では、この相関項を各ノードの状態の積で近似する独立モデル ($E[X_i X_j] \approx E[X_i] E[X_j]$) が使われているが [1]、本稿ではこの独立モデルの妥当性について詳細に検討する。その結果、相関項が $E[X_i]$ の時間発展（すなわち情報拡散過程）に大きく影響すること、独立モデルでは情報拡散過程が正しく表現できず、本来、情報が後から伝わるはずのノードに情報が先に伝わるといった「因果律の破れ」が生じることなどを明らかにする。

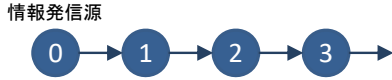


図1 1次元モデル

2. 情報拡散モデル

無向グラフにおいて、1つのノードを起点として隣接ノードへ情報が次々と伝搬していく際の情報伝搬速度について考察する。情報を保持しているノードは、情報を持たない全ての隣接ノードに情報（のコピー）を転送する。ノードが情報を保持してから、（情報を持たない）隣接ノードに情報を転送し終えるまでの時間（情報転送時間）は平均 $1/\lambda$ の指数分布に従う。情報を持たない隣接ノードが複数存在する場合、隣接ノードへの情報転送は同時ではなく個別に行われ、各隣接ノードへの情報転送時間は互いに独立であるとする。いったん情報を保持したノードはそのまま情報を保持し続ける。

時刻 t においてノード i が情報を保持する場合は1、保持しない場合は0をとる確率変数を $X_i(t)$ で表す。ノード i が時刻 t において情報を保持する確率 $p_i(t)$ は以下により表される。

$$p_i(t) = E[X_i(t)].$$

ノード i からノード j への情報転送が行われるのは、ノード i と j が隣接し、ノード i が情報を保持し、かつノード j が情報を保持していない場合に限られる。したがって、 $p_i(t)$ の時間発展は以下の式で記述される。

$$\begin{aligned} \frac{p_i(t)}{dt} &= \lambda \sum_j a_{ij} E[X_j(t)(1 - X_i(t))] \\ &= \lambda \sum_j a_{ij} (p_j(t) - E[X_j(t)X_i(t)]). \end{aligned} \quad (1)$$

ここで a_{ij} は隣接行列 A の要素を表す。つまり

$$a_{ij} = \begin{cases} 1, & \text{ノード } i \text{ と } j \text{ が隣接する} \\ 0, & \text{それ以外} \end{cases}$$

(1) を解析的にもしくは数値的に解くことで $p_i(t)$ が求めれば、 i 番目のノードへの情報到達時間 T_i は以下の式より算出できる。

$$E[T_i] = \int_0^\infty P(T_i > t) dt = \int_0^\infty (1 - p_i(t)) dt.$$

(1) には相関項 $E[X_j(t)X_i(t)]$ が含まれているが、相関項を厳密に求めることは一般に困難であり、(1) を解くためには、相関項について何らかの仮定を置くことが必要になる。

3. 例

3.1 1次元モデル（強相関モデル）

ノードが一列に連なる1次元モデルにおいて、時刻0でノード0のみが情報を保持し、その情報がノード1、2へと伝わっていくケースを考察する（図1）。式(1)は次のように書ける。

$$\begin{aligned} \frac{p_i(t)}{dt} &= \lambda(p_{i-1}(t) - E[X_{i-1}(t)X_i(t)]) \\ &\quad + \lambda(p_{i+1}(t) - E[X_{i+1}(t)X_i(t)]). \end{aligned}$$

図1の1次元モデルでは情報は番号の小さいノードから番号の大きいノードに伝わる。つまり、 $i \leq j$ ならば $\{X_i = 1\} \cap \{X_j = 1\}$ である（ノード i が情報を保持する場合に限り、ノード j も情報を保持する）ことから、

$$\begin{aligned} E[X_{i-1}(t)X_i(t)] &= P(\{X_{i-1}(t) = 1\} \cap \{X_i(t) = 1\}) \\ &= P(\{X_i(t) = 1\}) = p_i(t), \\ E[X_{i+1}(t)X_i(t)] &= P(\{X_{i+1}(t) = 1\} \cap \{X_i(t) = 1\}) \\ &= P(\{X_{i+1}(t) = 1\}) = p_{i+1}(t), \end{aligned}$$

が成り立つ。なお、これは

$$\begin{aligned} E[X_i(t)X_j(t)] &= \min\{E[X_i(t)], E[X_j(t)]\} \\ &= \min\{p_i(t), p_j(t)\}, \end{aligned} \quad (2)$$

を仮定することと同値であるが、(2) は次章で示すように、隣接するノードの相関が最も強い場合に対応することから、以下では(2)を仮定するモデルを「強相関モデル」とも呼ぶこととする。以上より

$$\begin{aligned} \frac{p_i(t)}{dt} &= \lambda(p_{i-1}(t) - p_i(t)) + \lambda(p_{i+1}(t) - p_{i+1}(t)) \\ &= \lambda(p_{i-1}(t) - p_i(t)). \end{aligned} \quad (3)$$

上記の1次元モデルにおいてはノード i への情報到達時間は i 次のアーラン分布に従うはずであるが、次の補題は同じ結論が(3)からも導かれることを示したものである。

補題 3.1. 初期条件

$$p_0(t) = 1, \quad p_i(t) = 0, \quad (i > 0)$$

のもとでは、 i 番目のノードへの情報到達時間 T_i は平均 i/λ の i 次アーラン分布に従う。すなわち、

$$P(T_i \leq t) = p_i(t) = 1 - \sum_{k=0}^{i-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Proof. ノード1が情報を保持する確率 $p_1(t)$ は、(3)より

$$\frac{p_1(t)}{dt} = \lambda(1 - p_1(t)), \quad p_0(t) = 1,$$

を満たす。これを解くと

$$p_1(t) = 1 - e^{-\lambda t},$$

が得られる。またノード i が情報を保持する確率が

$$p_i(t) = 1 - \sum_{k=0}^{i-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t},$$

であるとすると、

$$p_{i+1}(t) = 1 - \sum_{k=0}^i \frac{(\lambda t)^k}{k!} e^{-\lambda t},$$

が成り立つことが、(3)より容易に確かめられる。以上から、数学的帰納法により証明が完結する。□

3.2 独立モデル

本稿で扱う情報拡散モデルは、複雑ネットワーク上で伝染病が拡がっていく過程（epidemic process）を表すモデルの一つに位置付けることができる。伝染病の拡散過程のモデルでは、($i \neq j$ のとき) X_i と X_j が独立であるとするモデルがしばしば用いられる。これを独立モデルと呼ぶこととする。独立モデルでは、式(1)は次のように書かれる。

$$\begin{aligned}\frac{p_i(t)}{dt} &= \lambda \sum_j a_{ij} (p_j(t) - E[X_j(t)]E[X_i(t)]) \\ &= \lambda \sum_j a_{ij} p_j(t) (1 - p_i(t)).\end{aligned}$$

平均場近似 ($X_i(t)$ が i によらず同じ確率的性質を持ち、かつ隣接ノード数 (次数) も全て等しいとする近似) のもとでは、添字 i を省くと

$$\frac{p(t)}{dt} = ap(t) - a(p(t))^2, \quad a \stackrel{\text{def}}{=} \lambda d,$$

となる。ここで d は平均次数である。これを解くと

$$p(t) = \frac{p(0)}{p(0) + (1 - p(0))e^{-at}}. \quad (4)$$

が得られる。(4) はロジスティック曲線である。つまり、独立モデルのもとで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結果となる。

4. 一般の場合

4.1 情報到達時間の上下限

次の補題からわかるように、強相関 (1 次元) モデルは $E[X_i X_j]$ が最大の場合に、また独立モデルは $E[X_i X_j]$ が最小の場合に対応している。

補題 4.1. $X_i(t)$ 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。□

$p_i(t)$ の時間発展を記述する微分方程式 (1) からわかるように、相関項 $E[X_i(t)X_j(t)]$ が大きいほど $p_i(t)$ の増加が抑えられ情報伝送に時間がかかる。従って、強相関モデルは情報到達時間が最大のケース、また独立モデルは情報到達時間が最小のケースに相当しており、一般の場合は両者の中間に位置する。つまり、 $E[X_i(t)X_j(t)] \leq \min\{p_i(t), p_j(t)\}$ 、もしくは $E[X_i(t)X_j(t)] \leq p_i(t), p_j(t)$ を (1) に代入して、数値的に微分方程式を解くことで、情報伝搬時間の上下限が評価できる。

4.2 相関項の時間発展

ノード i と j が隣接しているとき、相関項 $E[X_i(t)X_j(t)]$ の微

分方程式を導いて、相関項を直接数値評価する手法について考察する。 $X_i(t)X_j(t)$ が増加するのは、 $X_i(t) = 0$ かつ $X_j(t) = 1$ で $X_i(t)$ が 0 から 1 に変化するか、 $X_i(t) = 1$ かつ $X_j(t) = 0$ で $X_j(t)$ が 0 から 1 に変化する場合に限られる。また、前者はノード i の隣接ノード (ノード j を含む) のいずれかが情報を保持している場合、後者はノード j の隣接ノード (ノード i を含む) のいずれかが情報を保持している場合になる。したがって、

$$\begin{aligned}\frac{dE[X_i(t)X_j(t)]}{dt} &= \lambda \sum_{k \neq i} a_{ik} E[(1 - X_i(t))X_j(t)X_k(t)] \\ &\quad + \lambda \sum_{k \neq j} a_{jk} E[X_i(t)(1 - X_j(t))X_k(t)].\end{aligned} \quad (5)$$

(5) の右辺は以下のように変形できる。

(5) の右辺

$$\begin{aligned}&= \lambda \sum_{k \neq i} a_{ik} (E[X_j(t)X_k(t)] - E[X_i(t)X_j(t)X_k(t)]) \\ &\quad + \lambda \sum_{k \neq j} a_{jk} (E[X_i(t)X_k(t)] - E[X_i(t)X_j(t)X_k(t)]) \\ &= \lambda (p_i(t) + p_j(t) - 2E[X_i(t)X_j(t)]) \\ &\quad + \lambda \sum_{k \neq i, j} a_{ik} (E[X_j(t)X_k(t)] - E[X_i(t)X_j(t)X_k(t)]) \\ &\quad + \lambda \sum_{k \neq i, j} a_{jk} (E[X_i(t)X_k(t)] - E[X_i(t)X_j(t)X_k(t)]).\end{aligned}$$

ここで、($a_{ij} = a_{ik} = 1$ であるが) ノード j とノード k が隣接しないとき ($a_{jk} = 0$)、

$$\begin{aligned}p_j(t)p_k(t) &\leq E[X_j(t)X_k(t)] \leq \min\{p_j(t), p_k(t)\}, \\ \bar{X}_{jk;i} &\leq E[X_i(t)X_j(t)X_k(t)] \\ &\leq \min\{E[X_i(t)X_j(t)], E[X_i(t)X_k(t)]\}, \\ \bar{X}_{jk;i} &\stackrel{\text{def}}{=} E[X_i(t)X_k(t)]E[X_i(t)X_j(t)]/p_i(t),\end{aligned}$$

が成り立つ。同様に ($a_{ij} = a_{jk} = 1$ であるが) ノード i とノード k が隣接しないとき ($a_{ik} = 0$)、

$$\begin{aligned}p_i(t)p_k(t) &\leq E[X_i(t)X_k(t)] \leq \min\{p_i(t), p_k(t)\}, \\ \bar{X}_{ik;j} &\leq E[X_i(t)X_j(t)X_k(t)] \\ &\leq \min\{E[X_i(t)X_j(t)], E[X_j(t)X_k(t)]\},\end{aligned}$$

が成り立つ。いずれも左辺は相関が最も弱い場合、右辺は相関が最も強い場合に相当する。さらに $a_{ij} = a_{ik} = a_{jk} = 1$ のとき

$$\begin{aligned}\max\{\bar{X}_{ij;k}, \bar{X}_{jk;i}, \bar{X}_{ik;j}\} &\leq E[X_i(t)X_j(t)X_k(t)] \\ &\leq \min\{E[X_i(t)X_j(t)], E[X_i(t)X_k(t)], E[X_j(t)X_k(t)]\},\end{aligned}$$

が成り立つ。以上より、相関が最も強い場合の近似式として次が得られる。

$$\begin{aligned}\frac{dE[X_i(t)X_j(t)]}{dt} &= \lambda (p_i(t) + p_j(t) - 2E[X_i(t)X_j(t)]) \\ &\quad + \lambda \sum_{k \neq i, j} a_{ik} (1 - a_{jk}) (\min\{p_j(t), p_k(t)\} \\ &\quad - \min\{E[X_i(t)X_j(t)], E[X_i(t)X_k(t)]\})\end{aligned}$$

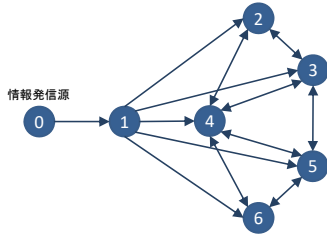


図2 因果律の破れが生じやすいケース

$$\begin{aligned}
& + \lambda \sum_{k \neq i, j} a_{jk}(1 - a_{ik})(\min\{p_i(t), p_k(t)\} \\
& \quad - \min\{E[X_i(t)X_j(t)], E[X_j(t)X_k(t)]\}) \\
& + \lambda \sum_{k \neq i, j} a_{ik}a_{jk}(E[X_j(t)X_k(t)] + E[X_i(t)X_k(t)] \\
& \quad - 2\min\{E[X_i(t)X_j(t)], E[X_i(t)X_k(t)], E[X_j(t)X_k(t)]\}).
\end{aligned} \tag{6}$$

同様に、相関が最も弱い場合の近似式として次が得られる。

$$\begin{aligned}
\frac{dE[X_i(t)X_j(t)]}{dt} & = \lambda(p_i(t) + p_j(t) - 2E[X_i(t)X_j(t)]) \\
& + \lambda \sum_{k \neq i, j} a_{ik}(1 - a_{jk})(p_j(t)p_k(t) - \bar{X}_{jk;i}) \\
& + \lambda \sum_{k \neq i, j} a_{jk}(1 - a_{ik})(p_i(t)p_k(t) - \bar{X}_{ik;j}) \\
& + \lambda \sum_{k \neq i, j} a_{ik}a_{jk}(E[X_j(t)X_k(t)] + E[X_i(t)X_k(t)] \\
& \quad - \max\{\bar{X}_{ij;k}, \bar{X}_{jk;i}, \bar{X}_{ik;j}\}).
\end{aligned} \tag{7}$$

相関が最も強い場合と最も弱い場合を仮定して、微分方程式(5)を数値的に解いた結果はそれぞれ $E[X_i(t)X_j(t)]$ の上下限を与えると考えられるので、それらから $p_i(t)$ および情報到達時間の期待値の上下限が評価できる。

4.3 因果律の破れと対応策

微分方程式(1)を解いて情報拡散現象を評価する際に、相関項 $E[X_{i+1}(t)X_i(t)]$ が適切に定められていないと、本来遅れて情報が伝わるはずのノードに先に情報が到達してしまうという現象が生じる。これを本稿では「因果律の破れ」と呼ぶこととする。

因果律の破れが生じる例として図2を取り上げる。ここではノード0が情報の発信源であり、隣接するノード1にまず情報が伝わり、ノード1の下流に相当するノード2からノード6にはさらに遅れて情報が伝わるので、 $p_1(t) \geq p_i(t)$ ($i = 2, \dots, 6$) が成り立つはずである。しかし、 $\{X_2, X_3, \dots, X_6\}$ 間の独立性を仮定すると、特にノード4のように次数の高いノードは、ノード2, 3, 5, 6を経由して情報が伝わる効果が過大に評価され、ノード1の情報到達確率をノード4の情報到達確率が上回る現象がおきる。これは、正の確率でノード1よりもノード4に早く情報が到達することを意味している。次数の高いノードほど情報到達が早いので、因果律の破れが生じやすい。

因果律の破れは相関項 $E[X_{i+1}(t)X_i(t)]$ の近似誤差に起因する、本来は生じるはずのない現象である。これを防ぐため、本稿では以下の関係が成立することを要請する。なお、情報発信源はノード0とする。

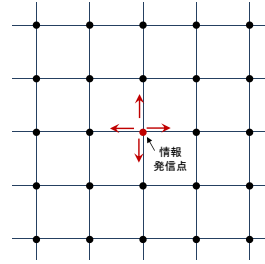


図3 格子状網

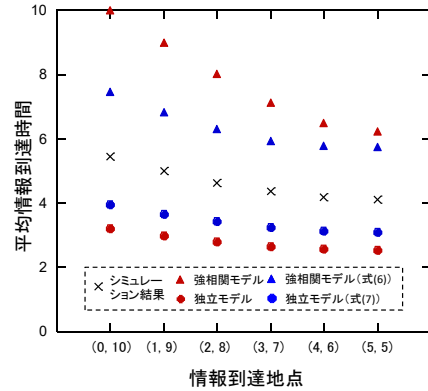


図4 情報転送時間 (格子状網: 10 ホップ離れたノード)

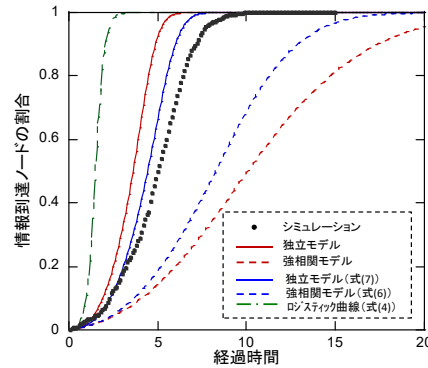


図5 情報既到達ノードの割合 (格子状網)

$$\forall j \neq 0 \quad p_j(t) \leq P\left(\bigcup_{k \in \mathcal{H}(1)} \{X_k(t) = 1\}\right) \leq \sum_{k \in \mathcal{H}(1)} p_k(t). \tag{8}$$

ここで $\mathcal{H}(1)$ はノード0 (発信源) から1ホップで到達するノードの集合である。

5. 数値例

5.1 格子状網

格子状のネットワーク (図3) において、1点から情報を拡散させた際の各ノードへの情報到達時間について、シミュレーションと微分方程式を数値的に解いて得た結果を比較した。

図4は、情報発信点から10ホップ離れた各点までの平均情報到達時間を評価したものである。例えば図4において(2,8)は情報発信地点から x 軸方向に2ホップ、 y 軸方向に8ホップ離れた点に情報が到達する平均時間を表している。1ホップあたりの平均情報転送時間は1とした。×印はシミュレーションにより得られた平均情報転送時間である。平均情報転送時間はおよそホップ数の半分もしくはそれ以下に収まっているが、これは様々な経路に沿った情報転送のうち最短時間で行われるも

表 1 情報発着信ノード (Gnutella)

発信ノード		着信ノード		ノード間ホップ数
ノード番号	次数	ノード番号	次数	
4676	1	5379	1	6
34	2	79	2	4
2470	4	3260	2	4
1071	3	4235	5	4

表 2 情報発着信ノード (Facebook)

発信ノード		着信ノード		ノード間ホップ数
ノード番号	次数	ノード番号	次数	
397	1	1603	1	4
356	4	1730	5	4
11	1	1067	1	3
1262	5	3169	111	3

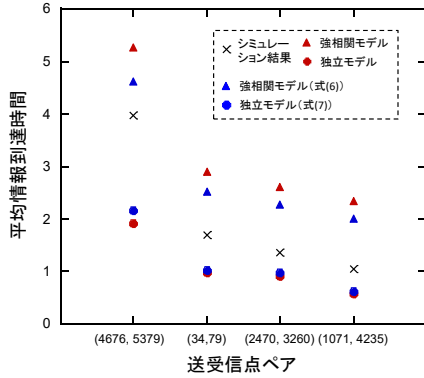


図 6 情報転送時間 (Gnutella)

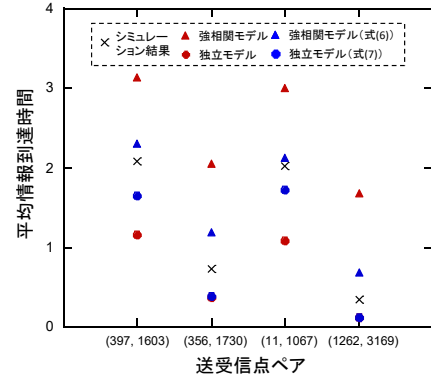


図 8 情報転送時間 (Facebook)

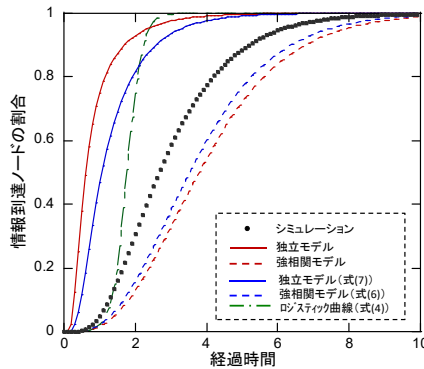


図 7 情報既到達ノードの割合 (発信元: 4676 番, Gnutella)

ので情報転送時間が決まるためと考えられる。

図 4 において、赤丸と赤三角印は、相関項をそれぞれ $E[X_i(t)X_j(t)] = p_i(t)p_j(t)$ (独立モデル) と $E[X_i(t)X_j(t)] = \min\{p_i(t), p_j(t)\}$ (強相関モデル) として微分方程式 (1) を数値的に解いた場合の結果である (因果律を破れを回避する (8) の成立は要請しない)。青丸と青三角印は、相関項をそれぞれ (6) と (7) により計算し、その結果を微分方程式に代入して得た結果を示している ((8) の成立も要請する)。シミュレーションによる情報転送時間は独立モデルと強相関モデルの中間におよそ位置するが、独立モデルと強相関モデルの結果には大きな差があり、それぞれ情報転送時間を過少もしくは過大に評価している。相関項を微分方程式により評価すると、独立モデルと強相関モデルの差は縮まり、シミュレーション結果に対するよりタイトな上下限を与えるようになる。

図 5 は、情報発信点から 20 ホップ以内に存在するノードにおける、情報既到達ノードの割合の時間変化を図示したものである。ホップ数の半分に相当する時間 10 が経過した時点ですべてのノードが情報が到達している。やはり、独立モデルと強相関モデルとの間には乖離がある。またロジスティック曲線 (式 (4)) はシミュレーション結果と大きく異なる。

5.2 Gnutella

次に、Gnutella のネットワークデータ [11] を用いて、ネットワーク上の 1 点から情報を拡散させた際の各ノードへの情報

到達時間について、シミュレーションと微分方程式を数値的に解いて得た結果を比較した。

図 6 に結果を示す。例えば図 6 において (4676, 5379) は 4676 番のノードから情報を発信させたときに、5379 番のノードに情報が伝わるまでの時間を評価した結果を表している。ノードの次数やノード間の距離 (ホップ数) は表 1 にまとめた通りである。4676 番から 5379 番までの情報転送時間は約 4 であるが、両ノードとも次数が 1 であり、4676 番から隣接ノードおよび、5379 番への隣接ノードからの情報転送にそれぞれ平均で 1 の時間がかかることを考慮すると、それ以外の部分での情報転送時間はおよそ 2 であり、その間のホップ数の半分に等しい。その他の発着ノードペア間のホップ数はすべて 4 であるが、それらの情報到達時間は 2 未満である。特に、1071 番から 4235 番への情報転送にかかる時間はおよそ 1 であり、これはホップ数の 1/4 である。全体的に非常に速やかに情報転送が行われている。

微分方程式 (1) を数値的に解いて得た情報転送時間も図 6 に示した。格子状網と同様に、強相関モデルは過大に、独立モデルは過小に情報転送時間を評価する。シミュレーション結果は独立モデルと強相関モデルの結果の間に位置するが、両モデルとシミュレーション結果との差は大きく、(1) における相関項を正しく評価することの重要性が確認できる。格子状網と同様に、相関項を (6) もしくは (7) により計算して (1) に代入する方が、よりタイトな結果を与える。

図 7 は、4676 番のノードが情報発信元の際の、情報既到達ノードの割合の時間変化を図示したものである。Gnutella の直径 (最大ホップ数) は 6 であり、時間 6 が経過した時点で 90% を超えるノードが情報を保持している。やはり、独立モデルと強相関モデルとの間には乖離があり、ロジスティック曲線 (式 (4)) はシミュレーション結果と異なる振る舞いを与える。

5.3 Facebook

Facebook のネットワークデータ [11] を用いて、情報到達時間について評価した結果を図 8 に示す。ノードの次数やノード間の距離 (ホップ数) は表 2 にまとめた通りである。397 番から 1603 番までの情報転送時間は約 2 であるが、両ノードとも次数が 1 であり、397 番から隣接ノードおよび、1063 番への隣接ノードからの情報転送にそれぞれ平均 1 の時間がかかること

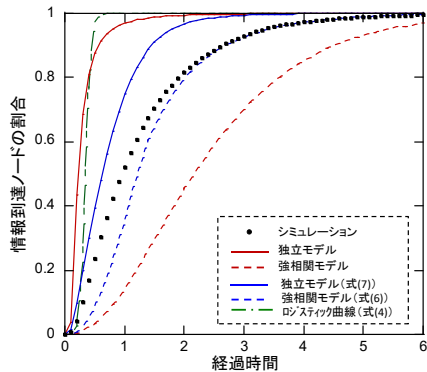


図9 情報既到達ノードの割合 (発信元: 397 番, Facebook)

を考慮すると、それ以外の部分 (2 ホップ分) での情報転送にかかる時間はごくわずかである。11 番から 1067 番までの情報転送にも同様のことが言える。356–1703 間, 1262–3169 間の情報転送にかかる時間は 1 以下である。Facebook は構成する各ノードの次数が大きく、情報拡散は極めてすみやかに行われている。

微分方程式 (1) を数値的に解いて得た情報転送時間も図 8 に示した。格子状網と同様に、強相関モデルは過大に、独立モデルは過小に情報転送時間を評価する結果となる。シミュレーション結果は独立モデルと強相関モデルの結果に間に位置するが、両モデルとシミュレーション結果との差は大きく、(1) における相関項を正しく評価することの重要性が確認できる。格子状網と同様に、相関項を (6) もしくは (7) により計算して (1) に代入する方が、よりタイトな結果を与える。

図 9 は、397 番のノードが情報発信元のとときの、情報既到達ノードの割合の時間変化を明示したものである。Facebook の直径 (最大ホップ数) は 4 であるが、その半分の時間 2 が経過した時点でおよそ 80% のノードが情報を保持している。やはり、独立モデルと強相関モデルとの間には乖離があり、ロジスティック曲線 (式 (eqn:logistic)) はシミュレーション結果と大きく異なる。

5.4 因果律の破れ

最後に、因果律の破れが生じる例を示す。図 10 は Facebook ネットワークにおいて 397 番のノードを発信元として情報拡散させた場合を示している。397 番のノードは次数 1 であり、隣がノード 107 である。ノード 107 は (ノード 397 を含めて) 524 のノードとつながっており、その一つにノード 1 がある。ノード 1 の次数は 3976 である。

図 11 は、ノード 1 とノード 107 の情報保持確率について、独立モデルにより数値的に評価した結果とシミュレーション結果とを比較したものである。シミュレーションでは常にノード 107 の情報保持確率はノード 1 の情報保持確率を上回る。しかし、相関項を $E[X_i(t)X_j(t)] = p_i(t)p_j(t)$ とした独立モデルでは、ノード 1 の情報保持確率がノード 107 の情報保持確率を時間の途中から逆転する。これは、ノード 1 の次数が大きいため、独立モデルのもとでは隣接ノードからの情報転送により情報保持確率が増える効果が過大に評価されるためである。いったんノード 1 の情報保持確率が急増すると、(本来は生じない) ノード 1 からノード 107 に情報が転送される効果によりノード 107 の情報保持確率も急増し、シミュレーション結果と大きく乖離する。同様の現象が Gnutella においても生じる。

6. 結 論

本研究では複雑ネットワーク上の情報拡散過程をネットワーク構造を陽に取り入れるモデルによって分析し、ノード相関が

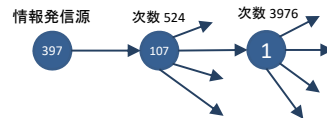


図 10 因果律の破れ (Facebook)

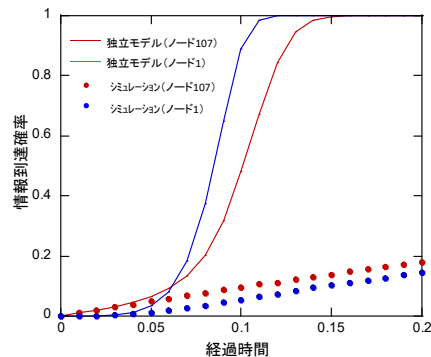


図 11 因果律の破れ (Facebook)

情報拡散過程に大きく影響すること、独立モデルでは情報拡散過程が正しく表現できず、因果律の破れ現象が生じることなどを明らかにした。今後は、情報の拡散と収束の両方を扱えるモデルについて同様の検討を進める。

文 献

- [1] 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.
- [2] 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.
- [3] R. Anderson and R.M. May, *Infectious Diseases in Humans*, Oxford University Press, Oxford, 1992.
- [4] R. Pastor-Satorras and A. Vespignani, “Epidemic spreading in scale-free networks,” *Phys. Rev. Lett.*, vol.86, no.14, pp.3200–3203, 2001.
- [5] M. Boguna and R. Pastor-Satorras, “Epidemic spreading in correlated complex networks,” *Phys. Rev.*, vol.E66, p.047104, 2002.
- [6] 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.
- [7] P.V. Mieghem, J. Omic, and R. Kooij, “Virus spread in networks,” *IEEE/ACM Trans. Networking*, vol.17, no.1, pp.1–14, 2008.
- [8] K. Sharkey, “Deterministic epidemic models on contact networks: Correlations and unbiological terms,” *Theor. Popul. Biol.*, vol.79, no.4, pp.115–129, 2001.
- [9] N. Schwartz and L. Stone, “Exact epidemic analysis for the star topology,” *Phys. Rev.*, vol.E87, p.042815, 2013.
- [10] P.V. Mieghem, *Performance Analysis of Complex Networks and Systems*, Cambridge University Press, Cambridge, England, 2014.
- [11] “Stanford large network dataset collection.” <http://snap.stanford.edu/data/>.