

同質原理に基づくネットワークダイナミクスをもつ情報伝播モデル

An Information Propagation Model with the Network Dynamics based on Homophily

石川 孝

Takashi Ishikawa

日本工業大学

Nippon Institute of Technology

The paper describes an information propagation model with network dynamics based on homophily. In the author's previous work, Twitter information propagation model was proposed to analyze the information resonance phenomenon in social media. However, the proposed model partly modeled the homophily, which states that similarity breeds connection, as the link rewiring mechanism representing users behavior to follow interesting information sources. The purpose of the research is to build a simple and powerful information propagation model only based on homophily, in which the network structure with fast propagation speed emerges in the time evolution. The paper describes the model and the experiment to test its emergent property.

1. はじめに

Twitter などのソーシャルメディアが災害時の情報伝達や直接民主主義の手段として有用であるためには、その情報伝播の速さが重要である。ネットワーク上の情報伝播の速さはネットワーク構造によって変化し、スモールワールドネットワークの情報伝播が速いことが知られている[Castellano 2007][Lü 2011][Doerr 2012]。ソーシャルメディアは人のつながりを作るメディアであると言われており、情報伝播の速いつながりを作ることがその有用性の源である。では、ソーシャルメディアのどのような特性が情報伝播の速いソーシャルネットワークを作るのだろうか？ この疑問に対する本論文の仮説は「同質原理」である。

筆者は昨年、ソーシャルメディアにおけるユーザ行動の特性を明らかにすることを目的として、IC モデルを拡張した情報伝播モデルによってソーシャルメディアにおいて特定のトピックについての情報発信が一時的に急増する情報共鳴現象を解析した[石川 2012a,b]。情報共鳴現象は、オンラインソーシャルネットワーク外のニュースが拡散するだけでなく、情報伝播によってユーザの関心が増えることによって、さらに情報の拡散が増幅されることで共鳴現象が起きていると考えられている。また、ユーザの関心の変化によってソーシャルネットワーク構造が変化することも情報共鳴が起こる要因と考えられる。そのため、ユーザの関心が増えるノードの状態変化と、ユーザのネットワーク構造の変化が互いに影響し合う適応的ネットワーク[Gross 2008]をモデル化の基礎とした。IC モデルに対して、受信メッセージのトピックによるユーザの関心の変化、ユーザの関心に依存する情報発信、および関心のあるトピックの情報源を新たにフォローするリンクつなぎ替え機構を導入した Twitter 情報伝播モデルによるシミュレーションの結果、これらの要因によって情報共鳴現象が生じるというモデルの仮説が確認された。また、ランダムネットワークから情報共鳴が起きやすいソーシャルネットワーク構造が創発することも確認された。しかし、この Twitter 情報伝播モデルには、ソーシャルネットワークダイナミクスの一般的な説明原理である「似た人同士はつながりやすい」という同質原理[McPherson

2001]がリンクつなぎ替え機構にしか反映されていないという欠陥がある。

本研究の目的は、上述の Twitter 情報伝播モデルにおけるソーシャルネットワーク構造の創発を基礎として、情報共鳴に特有と考えられるユーザの関心に依存する情報発信を排除した、情報伝播の速いソーシャルネットワークが創発するような、同質原理だけにに基づく情報伝播モデルを構築することである。目標とするモデルは、ネットワーク上の情報伝播を極力抽象化して、同質原理に基づくネットワークダイナミクスがネットワーク構造の創発の本質的な要因であることを示すための単純化を行う。また、ネットワーク構造の創発と同質原理との因果関係を数学的に証明することを研究の目標とする。このため、伝染病の流行に対するコンタクトプロセス[Liggett 1985]をモデル化の基礎とする。本論文では、上述の目標に対する情報伝播モデルの構築と、数値シミュレーションによるその特性解析の結果について述べる。

本論文は、まず第2章でモデルの数学的基礎となるコンタクトプロセスおよび伝染病の流行に対する SIS モデルとの関係を説明する。つぎに、第3章でネットワークダイナミクスの説明原理である同質原理とその形式化について述べる。そして、第4章で同質原理に基づく情報伝播モデルについて説明し、第5章でその創発的性質を調べる実験について述べる。最後に、本論文の結論と今後の課題を述べる。

2. コンタクトプロセス

まずコンタクトプロセス(contact process)を伝染病の伝播モデルとして直感的に説明する[西尾 2005]。数直線上の各整数点の上に人がいて、健康か病気の2種類の状態のどちらかをとっているとする。病気の人は隣の人の状態とは関係なく一定の割合の時間が経つと自然に健康になる。一方、健康な人は隣の人が病気のときに、隣の病気の人の人数と病気の伝染率に比例して病気になる。これらの状態変化は、連続時間確率過程として定式化すると、次の(1)、(2)のように与えられる。

- (1) 隣の k 人 ($k = 0, 1, 2$) の人が病気で変化しないという条件のもとでは、中央の人が病気になるまでにかかる時間はパラメータ $k \cdot \lambda$ の指数分布とする。これは、微

小時間 dt の間に病気になる確率が、伝染率を λ として、 $k \cdot \lambda dt$ であることを意味する。

- (2) 病気の人が回復するまでにかかる時間は、まわりの人の状態に無関係にパラメータ 1 の指数分布に従う。

数学的には、コンタクトプロセス η_t は $\{0, 1\}^{\mathbb{Z}^d}$ 上の連続時間のマルコフ過程である[今野 2008]。ただし、 \mathbb{Z}^d は d 次元超立方格子である。 η_t は時刻 t での状態の配置を表す。伝染病の伝播モデルとして解釈すると、状態の 0 が健康、1 が病気を表す。数学的な定義は2種類存在し、一つは形式的な生成作用素を用いるもので、もう一つはつぎのグラフ表現によって与えられる。場所 x として、プロセスの状態 $\xi_t = \{x \in \mathbb{Z}^d : \eta_t(x) = 1\} \subset \mathbb{Z}^d$ を考える。 ξ_t は時刻 t で病気である場所の集合を表す。ここで、 ξ_t と η_t は同一視できることに注意する。この ξ_t における時間発展はつぎで決まる。

- (1) もし $x \notin \xi_t$ ならば、 x に隣接する $y \in \xi_t$ である場所の数の λ 倍の推移率で、 $x \in \xi_t$ となる。
 (2) もし $x \in \xi_t$ ならば、推移率 1 で $x \notin \xi_t$ となる。

そして、 $A \subset \mathbb{Z}^d$ から出発するコンタクトプロセスをつぎのように定義する(詳細は[今野 2008]を参照)。

$$\xi_t^A = \{y \in \mathbb{Z}^d : \text{ある } x \in A \text{ に対して、}(x, 0) \text{ から } (y, t) \text{ への経路が存在する}\}.$$

コンタクトプロセスにおいて、すべての場所が 0 である配置は、吸収状態になっている。この性質をプロセスが吸収的であるといい、コンタクトプロセスが吸収的である臨界値 λ_c が存在する。しかし、1 次元の場合ですら λ_c の厳密な値は知られていない。

伝染病の流行についての伝統的なモデルは、均質な集団において人が健康 (S : Susceptible), 病気 (I : Infected), 回復 (R : Recovered, 病気にならない) のいずれかの状態をとり、それぞれの状態にある人の数 $S(t)$, $I(t)$, $R(t)$ の間の関係として流行のダイナミクスを記述する[Dorogovtsev 2010]。その中で SIS モデルは、人が S または I のいずれかの状態をとり、単位時間に状態遷移 $I \rightarrow S$ が回復率 μ で起こり、状態 I の隣人がいるときに状態遷移 $S \rightarrow I$ が伝染率 β で起こると仮定する。このモデルの主パラメータは生産数 $\lambda = \beta / \mu$ である。最初に $I(t)$ が小さい集団において $\lambda < \lambda_c$ であるとき、時間が経つと $I(t)$ が 0 の吸収状態になる流行閾値 λ_c が存在する。逆に $\lambda \geq \lambda_c$ であると、病気は集団に拡がり続ける活性状態になる。SIS モデルのマクロな量 $I(t)$ のダイナミクスは、ミクロな確率過程であるコンタクトプロセスの平均場近似によってつぎの方程式として導出できる[杉峰 2006]。

$$I'(t) = -I(t) + 2d\lambda \{1 - I(t)\} I(t)$$

したがって、コンタクトプロセスはマクロな SIS モデルに対応するミクロなモデルになっている。

コンタクトプロセスや SIS モデルについては、複雑ネットワーク上でのダイナミクスが研究されている[一宮 2008]。基体となるネットワークは、単純グラフ $G = (V, E)$ を考える。 V はノード集合、 E はエッジ集合で、二つのノード間のエッジは高々 1 本、両端が同一のノードであるようなエッジは存在しないとす。ネットワーク構造は、つぎのような隣接行列 $A = (a_{ij})$ で表される。 $V = \{v_1, v_2, \dots, v_N\}$ とし、 v_i と v_j を両端とするエッジを (v_i, v_j) と書くとき、 $(v_i, v_j) \in E$ ならば $a_{ij} = 1$ 、その他は $a_{ij} = 0$ とする。無向グラフでは A は対称行列であり、対角化が可能である。重み付きグラフでは一般に $a_{ij} \in [0, 1]$ である。本論文では、これらの記法を使ってネットワークを表す。

3. 同質原理

「似た人同士はつながりやすい」という同質原理は、様々な種類のソーシャルネットワークのつながりを形成する[McPherson 2001]。この原理の結果は、人々の個人ネットワークが多くの社会地理的特性、行動、対人的特性に関して均質であることである。つまり、人々はふつうは似た人とだけ目立った接触をもつので、人々の性質が局所化する。似た人とだけ相互作用することによって、ネットワーク上の位置の結果として経験するすべてが強化される。同質性 (homophily) は、似た人同士の間の接触が似ていない人同士よりも高い割合で起こるという原理であり、社会的特性での距離がネットワーク上の距離、すなわち二人の人をつなげる情報が伝わる関係性の数に翻訳されることを意味する。

広い意味では、同質性が生じるメカニズムには、選好的同質性 (似た人同士はつながりやすいこと) と誘導的同質性 (つながりによって人が似ること) がある[Kossinets 2009]。この両方の同質原理のメカニズムをネットワークダイナミクスとして形式化すると、ノード間の相互作用の強さ (結合強度) とノード間の類似度が比例することになる (図 1)。相互作用の強さは、病気の伝染や情報の伝播が起こる確率を意味する。また、類似度は、ノードの同値類を定義するノードの状態の関数である。ノードの状態は一般に相互作用によって変化するので、相互作用の強さも時間的に変化することでネットワークダイナミクスが生じる。重みなし (重みがすべて 1 の) グラフで表されるネットワークでは、ネットワーク自体は変化せずに、相互作用の結果として生じるノードの状態変化のダイナミクスをモデル化する。ノードの状態変化によってリンクつなぎ替えが起こる場合は、このダイナミクスを別途にルールで記述することになる[石川 2012a,b]。これに対して重み付きグラフで表されるネットワークでは、相互作用の強さをエッジの重みで表すことによって、ネットワークダイナミクスを隣接行列または結合行列だけを使ってモデル化することができる。そこで本論文では、この隣接行列を使うアプローチによって、同質原理に基づくネットワークダイナミクスをモデル化する。

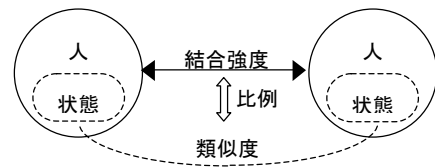


図 1. 同質原理の形式化

4. 同質原理に基づく情報伝播モデル

重み付きグラフで表されるネットワーク上の情報伝播は、コンタクトプロセスと SIS モデルを基礎として、つぎのようにモデル化できる。このモデルを同質性ネットワークモデルと呼ぶ。

- (1) ネットワーク

グラフ $G = (V, E)$ の隣接行列表現を使って、ノード数を N 、 $0 < i, j \leq N$ として、エッジ (v_i, v_j) の重み $w_{ij} \in [0, 1]$ を要素とする結合行列 W によってネットワークを表す。上述の同質原理に基づくネットワークダイナミクスを組み込むため、エッジ重み w_{ij} がノード v_i, v_j 間の類似度と相互作用の強さの両方を表すものと仮定する。このモデルでは、モデルを単純化するため、類似度を定義するためのノードの状態を排除している。なお、 $\sum_i w_{ij}$ がノード v_j の入次数になる。

(2) ノードの状態

ネットワーク上のダイナミクスとしての情報伝播を記述するため、ノード v_i は情報既知を表す値 1 または情報未知を表す 0 のいずれかをとるものとする。あるノード $v_i=1$ に隣接する ($w_{ij}>0$) ノード $v_j=0$ において、 $v_j=1$ への状態の変化が起こったとき、 v_i から v_j に情報伝播が起こったという。

(3) ノードの状態変化

- ノード v_i から v_j への情報伝播は、単位時間あたりに確率 $\beta w_{ij} v_i (1 - v_j)$ で起こるものとする。ここで $\beta \in [0, 1]$ は、伝染率 (infection rate) を表す。
- ノード v_i は、単位時間あたりに確率 μv_i で $v_i=0$ に変化するものとする。ここで μ は回復率 (recovery rate) を表す。

(4) エッジ重みの1次変化

ノード v_i から v_j に情報伝播が起こると、 $r \in [0, 1]$ を学習率として、エッジ (v_i, v_j) の重み w_{ij} が $r w_{ij}$ だけ増加するものとする。ただし、 $(1+r) w_{ij} > 1$ になる場合は、 $w'_{ij}=1$ とする。さらに、 v_j の入次数を一定に保つように、すべての $k \neq i, j$ に対して w_{kj} を $(w'_{ij} - w_{ij}) / (N - 2)$ だけ減少させるものとする。ただし、 $w'_{kj} < 0$ となる場合は、 $w'_{kj}=0$ とし、その差分を他のエッジに誤差が一定値以下になるまで割り振る (詳細は省略)。

(5) エッジ重みの2次変化

基体となるネットワークの機能である情報伝播をより効率的にするため、推移的リンク生成メカニズム [Ishikawa 2010] をつぎのように導入する。推移的リンク生成は、ソーシャルネットワークにおいて新しい友人を作るときに友人の友人を優先して選ぶことを表す。このメカニズムは、同質原理に基づくネットワークダイナミクスによるエッジ重みの2次変化 (非線形効果) であると考えられる [石川 2012a,b]。ノード v_i から v_j に情報伝播が起こったとき、すべての $k \neq i, j$ に対して $\max_k w_{ik} w_{kj} > w_{ij}$ である、すなわち v_k を経由した情報伝播の確率の方が高いとき、(4)の代わりに、

$$w'_{ij} = (1+r) \max_k w_{ik} w_{kj}$$

とする (図2)。ただし、すべてのエッジ重みが $w \in [0, 1]$ となるように (4) と同様に入次数保存のための他のエッジ重みの調整を行う。

上述の情報伝播モデルは、基礎となるコンタクトプロセスと同じように、 $\lambda = \beta / \mu$ によって吸収状態と活性状態の二つの相をもつことが予想される。 λ が閾値 λ_c より大きい活性状態では、ノード数 $n \gg 1$ の範囲に情報が伝播し続け、エッジ重みの変化が持続的に起こる。 $\mu \neq 0$ の場合は、個々のノードの状態は時間的に変化するが、 n の値は確率的に一定の範囲に留まる。このモデルのエッジ重みの2次変化は、情報伝播の速いネットワーク構造が創発する必要条件であって、エッジ重みの1次変化だけではこの創発は起こらない。以下の実験によってこのことを確認する。

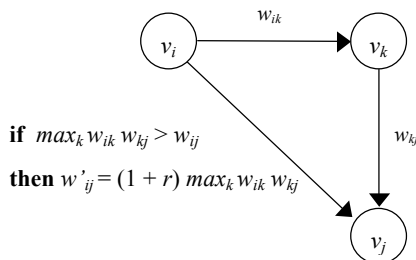


図2. エッジ重みの2次変化

5. 実験

5.1 目的

第4章で定義した同質性ネットワークモデルが、情報伝播の速いネットワーク構造を創発することを確認する。

5.2 方法

1時間ステップごとにすべてのノードで情報の忘却 (I→S) または隣接ノードへの情報の伝播 (S→I) が起こるエージェントベースシミュレーションによって、以下の条件で数値実験を行う。

- 初期ネットワークは、結合行列 W の要素が対角要素 (=0) を除いてすべて等しいとする。ノード数 N は 200 とし、ネットワークの平均入次数は 4 とする。ネットワークダイナミクスによってエッジ重みが変わっても、各ノードの入次数は一定であるが、出次数は変化する。また、時刻 $t=0$ での情報既知のノード数は 1 とする。
- モデルのパラメータは、伝染率 $\beta=0.5$ 、回復率 $\mu=0.5$ 、学習率 $r \in [0, 1]$ とする。 $\lambda = \mu / \beta = 1$ のとき、ほぼ 50% の確率で、約 50% のノードが情報既知となる。
- ネットワークダイナミクスは、同質性ネットワークモデルのルール(1)~(5)を毎時間ステップに実行する。
- 情報伝播の速さは、情報既知のノード数が $N / 3$ に達するまでの時間ステップ数を計測して伝播時間 T とする。ネットワーク構造の変化に伴う情報伝播の速さの変化を調べるため、100 時間ステップごとに情報既知のノードを初期と同じ 1 つにリセットして、 T を 0 から再計測する。ただし、結合行列 W はリセットせずに継続して変化させる。1回の試行において、時刻 $t=50,000$ まで T を 500 回計測して平均する。1つの r の値に対して、この試行を 10 回反復して、 T の総平均と標準偏差を求める。
- エージェントベースシミュレーションのプログラムは、SmallWorldDiffusionSIS.nlogo [Adamic 2008] を基にして上記の拡張を行った。

5.3 結果

学習率 r (increment-rate) を 0 から増やすと、伝播時間の総平均 T (infected-time) は、初期ネットワークよりも一旦減少し、さらに r を増やすと初期ネットワークの伝播時間のレベルに戻る (図3、誤差範囲は標準偏差)。この結果より、 r のある範囲について、初期状態の重みが一般的なネットワークよりも情報伝播の速さが有意に小さいネットワーク構造が創発することが確認された。

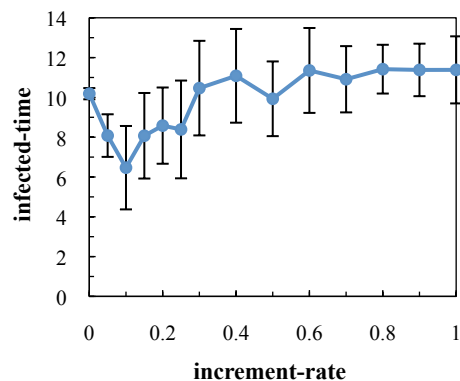


図3. 学習率に対する伝播時間の変化

図3において情報伝播が最も速い学習率 $r=0.1$ に対する結合行列におけるエッジ重みの最大値(max)と標準偏差(std)はその50倍)および伝播時間の総平均(timeはその0.1倍)の典型的な時間的変化を図4に示す。このグラフは、ネットワーク構造の創発が時間ステップ(ticks)の20,000付近で完了し、それ以降は伝播時間がほぼ安定することを示している。また、同質性ネットワークモデルは基本的にコンタクトプロセスであるので、条件は上と同じで β だけを変化させて情報既知ノード数比率の総平均(infected-ratio, 全時間ステップでの情報既知ノード数比率の平均)を調べると、図5(誤差範囲は $n=5$ での土標準偏差)のように $\lambda = \beta/\mu$ によって吸収状態と活性状態の二つの相を分ける閾値 $\lambda_c \approx 0.5$ が存在する。

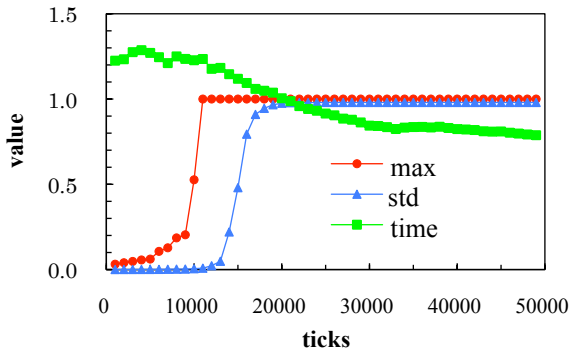


図4. エッジ重みと伝播時間の時間的変化

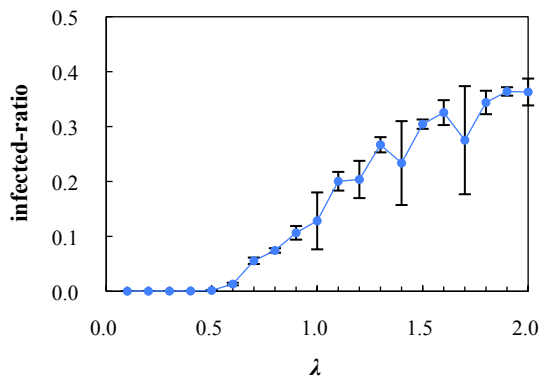


図5. 情報既知ノード数比率の λ による変化

5.4 考察

この同質性ネットワークモデルにおけるエッジ重みの2次変化は、このメカニズムを除いた実験では学習率を変えても伝播時間が変化しないことから、情報伝播の速いネットワーク構造が創発することの必要条件である。

6. おわりに

本論文は、Twitterなどのソーシャルメディアにおける情報伝播について、「似た人同士はつながりやすい」という同質原理に基づくネットワークダイナミクスをもつ同質性ネットワークモデルを提案した。同質原理の効果を表すパラメータを変化させて実験した結果、パラメータのある範囲において情報伝播の速いネットワーク構造が創発することが確認された。しかし、創発するネットワーク構造の特性や情報伝播の拡がり方の時間変化などについてはさらに調査が必要である。また、情報伝播の速いネットワーク構造の創発と同質原理との因果関係を数学的に証明することは今後の課題である。

参考文献

- [Adamic 2008] Adamic L. and Bakshy, E.: Diffusion in a small world (NetLogo). <http://www.ladamic.com/netlearn/>, 2008.
- [Castellano 2007] Castellano, C., Fortunato, S., and Loverto, V.: Statistical physics of social dynamics. *Reviews of Modern Physics*, Vol. 81, No. 2, pp. 591-646, 2009.
- [Doerr 2012] Doerr, B., Fouz, M., and Friedrich, T.: Why Rumors Spread Fast in Social Networks. *Communications of the ACM*, Vol. 55, pp. 70-75, 2012.
- [Dorogovtsev 2010] Dorogovtsev, S. N.: *Lectures on Complex Networks*. Oxford University Press, 2010.
- [Gross 2008] Gross, T. and Blasius, B.: Adaptive coevolutionary networks: a review. *Journal of the Royal Society Interface*, Vol. 5, No. 20, pp. 259-71, 2008.
- [Ishikawa 2010] Ishikawa, T.: The Effect of Transitive Linking on Information Diffusion in Dynamic Acquaintance Networks. *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 2010.
- [Kossinets 2009] Kossinets, G. and Watts, D. J.: Origins of Homophily in an Evolving Social Network. *American Journal of Sociology*, Vol. 115, No. 2, pp. 405-450, 2009.
- [Liggett 1985] Liggett, T. M.: *Interacting Particle Systems*. Springer, 1985.
- [Lü 2011] Lü, L., Chen, D., and Zhou, T.: Small world yields the most effective information spreading. arXiv:1107.0429, 2011.
- [McPherson 2001] McPherson, M., Smith-Lovin, L., and Cook, J. M.: Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, Vol. 27, pp. 415-444, 2001.
- [石川 2012a] 石川 孝: ソーシャルメディアにおける情報共鳴のメカニズム. 2012年度人工知能学会全国大会, 2012.
- [石川 2012b] 石川 孝: ソーシャルメディアにおける情報共鳴のエージェントベースシミュレーション. 日本ソフトウェア科学会 ネットワークが創発する知能研究会 JWEIN12, 2012.
- [一宮 2008] 一宮尚志: 複雑ネットワーク上のコンタクトプロセス. 数理解析研究所講究録, 第 1616 巻, pp. 103-111, 2008.
- [今野 2008] 今野紀雄: 『無限粒子系の科学』. 講談社サイエンスエンティフィク, 2008.
- [杉峰 2006] 杉峰伸明: コンタクトプロセスとその応用. 数理解析研究所講究録, 第 1499 巻, pp. 153-158, 2006.
- [西尾 2005] 西尾眞喜子, 樋口保成: 『確率過程入門』. 培風館, 2005.