

実ネットワークでのゲームモデルのダイナミクスの調査検討

Empirical Study of Dynamics of Game Model using Real Networks

垣貞香奈*¹

Kana Kakisada

薩川大輝*¹

Daiki Satsukawa

斉藤和巳*¹

Kazumi Saito

木村昌弘*²

Masahiro Kimura

元田浩*³

Hiroshi Motoda

*¹静岡県立大学

University of Shizuoka

*²龍谷大学

Ryukoku University

*³大阪大学

Osaka University

Social networks mediate the spread of various information including topics, ideas and innovations, and exploring essential dynamics of the information spread is one of the most central problems in social network analysis. In this paper, we focus on a model of networked coordination games, called the Morris model, and propose a method for analyzing the change of the node states from some initial state over time. More specifically, this method calculates the final influence degree, the size of the state cycle, and the minimum step to reach the state cycle. In our experiments using two kinds of large real undirected networks, we report that although these two networks has somehow different characteristics in terms of their link densities, we obtained quite similar experimental results with respect to the final influence degree, the size of the state cycle, and the minimum step to reach the state cycle.

1. はじめに

アイデアやイノベーションなどが社会ネットワーク上でどのように広がるかのダイナミクスの探求と解明は、現実社会の構造や機能を深く理解することを可能とし、新たなライフスタイルの普及など多様な社会活動や新商品の効果的な宣伝など広報戦略の立案においてきわめて重要と考えられる。ここで社会ネットワークとは、例えばノードを人とし、その友人関係をリンクとして繋いだネットワークを意味する。このようなアイデアなど情報拡散に関する基本的な確率モデルとして独立カスケードモデル (Independent Cascade Model) や線形閾値モデル (linear threshold model) などが広く研究されている [Kempe 03, Gruhl 04, Kimura 07]。ただし、これら確率モデルのダイナミクスは非常に複雑なため、現時点であまり多くの知見が得られていない状況と言える。

一方、前述した線形閾値モデルと類似した考え方に基づき確定的な挙動で規定されるモデルの一つに、Morris モデルと呼ばれるネットワーク上のゲームモデルがある [Morris 00, Kleinberg 07]。Morris モデルにおいては、Contagion 閾値と呼ばれる性質などが研究されている。ここで Contagion とは、無限に多くのノード集合全体に情報が伝わった状態を意味し、ある情報源ノード集合から拡散が開始され Contagion となるためには、Morris モデルにおいて閾値 q と呼ばれるパラメータが 0.5 未満でなければならないことが証明されている [Morris 00, Kleinberg 07]。しかしながら、ある種の構造を持つ有限サイズのネットワークにおいて、閾値パラメータ q とダイナミクスとの関係など未解明の性質も多く、様々な研究課題が残されている。

本稿では、Morris モデルなどの確定的な情報拡散モデルにおいて、ネットワーク構造やモデルのパラメータとダイナミクスの関係を探求するため、基本となる幾つかの分析法を提案する。実世界ネットワークの 2 つのデータセットを用いた実験では、閾値パラメータ q を変化させ、影響度、サイクルのサイズ、および、サイクルまでの最小ステップを評価する。

連絡先: 斉藤和巳, 静岡県立大学経営情報学部, 〒 422-8526
静岡市駿河区谷田 52 番 1 号, 054-264-5436, k-saito@u-shizuoka-ken.ac.jp

2. ネットワーク上のゲームモデル

本節では、無向グラフ $G = (V, E)$ により表現される社会ネットワーク上において、局所インタラクションが定義されたゲームとして規定される Morris モデル [Morris 00, Kleinberg 07] について述べるとともに、そのダイナミクスについて解説する。

2.1 ゲームの定義

ネットワーク上の各ノード $v \in V$ は A (old) または B (new) のどちらかの状態を取るものとする。予め規定する実数 q ($0 \leq q \leq 1$) により、各無向リンク (v, w) において、各ノードは以下の条件に基づき利得 (payoff) が得られるとする。

- v も w も状態 A を選ぶなら、両者は利得 q を得る。
- v も w も状態 B を選ぶなら、両者は利得 $1 - q$ を得る。
- v と w が異なる状態を選ぶなら、両者の利得は 0 となる。

明らかに、各ノードの利得はリンク結合した相手先ノード群の状態のみで決まる。つまり、局所インタラクションのみが定義されたゲームである。

いま、あるノード v に着目し、そのリンク相手先で状態が A のノード数を d_v^A とし、状態が B のノード数を d_v^B とする。ここで、 v のリンク相手先の総数を d_v とすれば、各ノードは A または B のどちらかの状態を取るもので以下が成り立つ。

$$d_v = d_v^A + d_v^B \quad (1)$$

ノード v においては、自らの状態が A ならば利得は qd_v^A であり、一方、状態が B ならば利得は $(1 - q)d_v^B$ となる。したがって、ノード v が状態 B を選ぶのは以下の状況と定義できる。

$$(1 - q)d_v^B \geq qd_v^A \quad (2)$$

この不等式を変形すれば以下を得る。

$$d_v^B \geq qd_v \quad (3)$$

よって、ノード v が状態 B を選ぶのは、そのリンク相手先の q 割が状態 B を選んでいるときとなる。すなわち、このゲームによりシンプルな閾値モデルが定義される。以下では、このモデルを Morris モデルと呼び、パラメータ q を閾値と呼ぶ。

2.2 ゲームモデルの挙動

時刻 $j = 0$ でのネットワークにおける各ノードの初期状態として、状態が B のノード集合を S とし、残りのノード集合 $V \setminus S$ の状態は A とする。ここで、ノード集合 S を初期集合と呼ぶ。本稿では、式 (3) で定義した局所インタラクションで規定される Morris モデルにより、初期集合 S から時刻の経過とともに、どのように各ノードの状態が変化していくかを考える。

いま、与えられた閾値 q と初期集合 S に対して、時刻 j で状態が B となっているノード集合を $h_q^j(S)$ で表す。以下では簡略化のため、 $h_q^j(S)$ を単に状態とも呼ぶ。典型的には、状態 B を新たなアイデアやイノベーションを受け入れた状態 (new) と考え、状態 A をまだ受け入れてない状態 (old) と考えたとき、状態 $h_q^j(S)$ のダイナミクスを分析することは、比較的少数のノードから構成される初期集合 S から、式 (3) で定義した局所インタラクションに基づき、どのようにアイデアなどが広がっていくかを調べることに対応する。

3. 分析方法

ネットワークのノード集合 V が有限なら、任意の初期集合 S に対して、状態 $h_q^j(S)$ のとり得る範囲も有限集合に限定される。よって、ある正のステップ j と c が存在し、 $h_q^j(S) = h_q^{j+c}(S)$ となり、Morris モデルの挙動は確定的なため、状態はサイクルする。以下では、サイクルの最小ステップを以下で定義し、サイクルのサイズと呼ぶ。

$$c_q(S) = \min_{c>0} \{c : \exists j, h_q^j(S) = h_q^{j+c}(S)\} \quad (4)$$

ただし、あるステップ j で $h_q^j(S) = \emptyset$ となる場合、 $c_q(S) = 0$ と定義し、そうでない場合との便宜上の区別を設ける。また、サイクルに至るまでの最小ステップを以下で定義する。

$$j_q(S) = \min_{j>0} \{j : \exists c, h_q^j(S) = h_q^{j+c}(S)\} \quad (5)$$

サイクルに至れば、任意の時刻での状態のサイズ $|h_q^j(S)|$ を求めることができる。よって、サイクルにおける平均状態サイズを以下で定義し、初期ノード集合 S に対する影響度 (influence degree) と呼ぶ。

$$\sigma_q(S) = \frac{1}{c_q(S)} \sum_{j=j_q(S)}^{c_q(S)-1} |h_q^j(S)| \quad (6)$$

ここで、 $h_q^j(S) = \emptyset$ の場合には、 $\sigma_q(S) = 0$ と自然に定義できる。明らかに、ネットワークの構造や閾値 q の設定により、影響度、サイクルのサイズ、および、サイクルまでの最小ステップは変わりうる。

いま、 $\Gamma(v)$ をノード v のリンク相手先ノード集合とすれば、Morris モデルの動作は以下のアルゴリズムで記述できる。

1. ネットワーク構造、初期集合 S 、および閾値 q を入力として与え、時刻を $j = 0$ に設定する。
2. $h_q^{j+1}(S) = \emptyset$ に設定し、各ノード $v \in V$ に対して、 $|h_q^j(S) \cap \Gamma(v)| / |\Gamma(v)| \geq q$ ならば $h_q^{j+1}(S) = h_q^j(S) \cup \{v\}$ とする。
3. $h_q^{j+1}(S) = \emptyset$, $h_q^{j+1}(S) = V$ 、または、ある $k \leq j$ が存在して $h_q^k(S) = h_q^{j+1}(S)$ ならば終了、さもなければ $j = j + 1$ として、2. へ戻る。

ここで、このアルゴリズムで順次求まる $h_q^j(S)$ を保持しておくことで、上述した影響度、サイクルのサイズ、および、サイクルまでの最小ステップは容易に求めることができる。

4. 評価実験結果

評価実験においては、実社会ネットワークの顕著な特徴を多くもつ大規模ネットワークの利用が望ましいと考えられる。本論文では、そのような実世界ネットワークの2つのデータセットを用いた実験結果を報告する。

4.1 データセットと実験設定

ある種の情報は、トラックバックを通してあるブログ著者から別のブログ著者へと伝播しうると考えられるので、ブログのトラックバックネットワークを用いて評価実験を行った。トラックバックネットワークデータは、「goo ブログ」(<http://blog.goo.ne.jp/usertheme/>) の「JR 福知山線脱線事故」というテーマからトラックバックを10段辿ることにより、2005年5月に収集した。収集したネットワークは、12,047ノードと53,315リンクをもつ連結有向グラフであるが、トラックバック作成には、基本として相互承認が必要なため、すべてが双方向リンクを持つようにリンクの追加を行い無向グラフを構築した。その結果、各ノードの次数の総和は79,920となった。本ネットワークは、たいていの大規模な実ネットワークと同様、次数分布も所謂べき乗則に従っていた。以降、本ネットワークデータをブログデータセットと呼ぶ。

もう一つのデータセットとして、「ウィキペディア」内の「人名一覧」から導かれる人物ネットワークを用いて評価実験を行った。具体的には「人名一覧」に登場する人物において、ウィキペディア内の記事中に6回以上共起した2人の人物をリンクすることから得られる無向グラフの最大連結成分を抽出した。以降、本ネットワークデータをウィキペディアデータセットと呼ぶ。ここに、ノード数は9,481であり、各ノードの次数の総和は245,044であった。

実験設定としては、初期集合が一つのノードからなる単純なケースを扱った。明らかに、このような初期集合として、ネットワーク上の任意のノードを指定できる。よって、各ノード $v \in V$ に対して、状態 $h_q^j(\{v\})$ を考えて、閾値 q の設定を変化させつつ、影響度、サイクルのサイズ、および、サイクルまでの最小ステップに関する評価を行った。

4.2 影響度

図1には、式(6)に基づく、ブログデータセットでの影響度の評価結果を示す。ここで、横軸は影響度による各ノードのランクであり、縦軸は各ノードの影響度を示している。図より、閾値 q が0.15と0.2の間で影響度曲線が比較的大きく変化していることが分かる。特に、ブログデータセットでのノード数は12,047であるので、閾値が0.15より小さければ、単一ノードからなる初期集合のうち数百個のものが、全体に影響を及ぼしていることが分かる。なお、閾値 q は小さいほど、一般に影響度は大きくなることに注意。

図2には、ウィキペディアデータセットでの影響度の評価結果を示す。図より、ブログデータセットでの影響度の評価結果と比較して、閾値 q が0.15のとき、ある程度影響度は小さくなっているものの、全体の傾向としては、かなり類似した結果となっていることが分かる。すなわち、ブログデータセットでの各ノードの平均次数は6.63で、ウィキペディアデータセットでの平均次数は25.85であり、ある程度リンク密度には違いがあるものの、同程度の閾値 q で類似した結果となっていることが分かる。

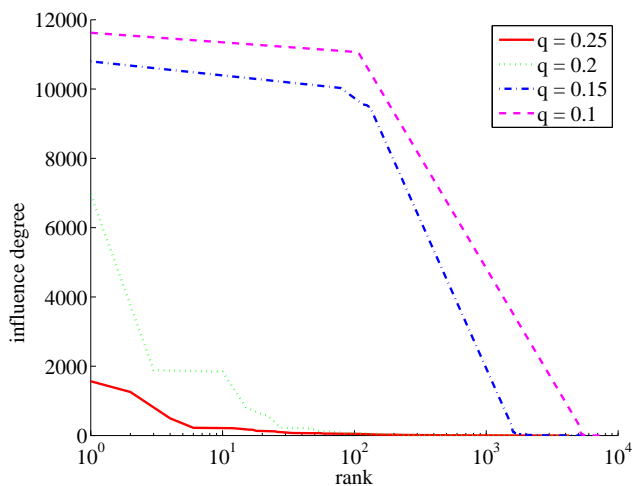


図 1: プログデータでの影響度

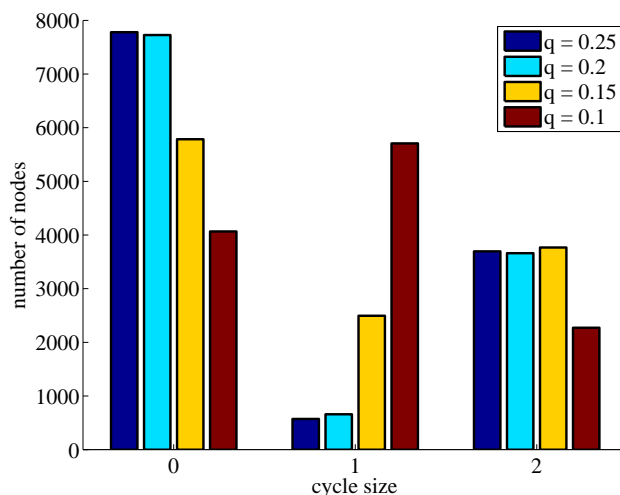


図 3: プログデータでのサイクルのサイズ

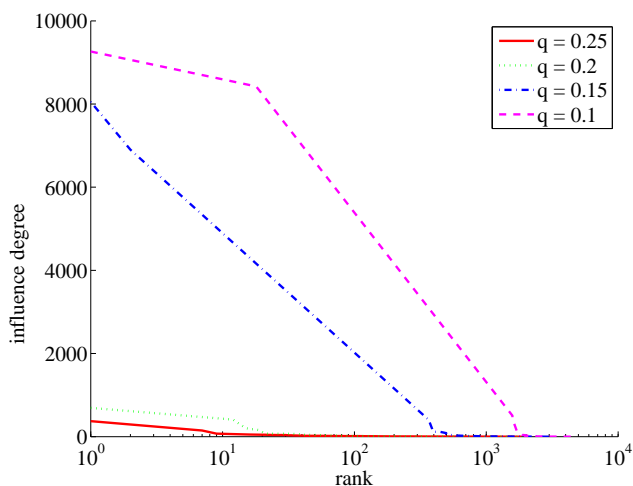


図 2: ウィキペディアデータでの影響度

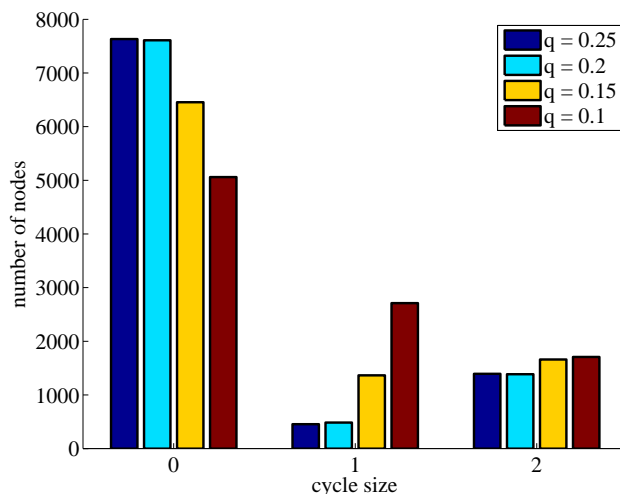


図 4: ウィキペディアデータでのサイクルのサイズ

4.3 サイクルのサイズ

図 3 には、式 (4) に基づく、プログラムデータセットでのサイクルのサイズの評価結果を示す。ここで、横軸はサイクルのサイズであり、縦軸はサイズ毎のノード数を示している。図より、サイクルのサイズは比較的小さく高々 2 であったことが分かる。また、閾値 q が 0.2 より大きければ、単一ノードからなる初期集合の殆どで $h_q^j(S) = \emptyset$ となるため、サイクルのサイズは 0 となっている。一方、閾値 q が 0.1 のときには、多くの初期集合において $h_q^j(S)$ が全ノード集合 V に近付くため、サイクルのサイズが 1 のケースが多くなっていると考えられる。

図 4 には、ウィキペディアデータセットでのサイクルのサイズの評価結果を示す。この評価においても、プログラムデータセットでの影響度の評価結果と比較して、閾値 q が 0.1 のとき、サイズ 2 のサイクルの結果が減少してないものの、全体の傾向としては、かなり類似した結果となっていることが分かる。すなわち、前述したように、2 つのデータセットでは、ある程度リンク密度には違いがあるものの、同程度の閾値 q で類似した結果となっていることが分かる。

4.4 サイクルまでの最小ステップ

図 5 には、式 (5) に基づく、プログラムデータセットでのサイクルまでの最小ステップの評価結果を示す。ここで、横軸はサイクルまでの最小ステップによる各ノードのランクであり、縦軸は各ノードでの実際のステップを示している。図より、閾値 q が 0.15 のとき、比較的多くの初期集合において、サイクルまで長いステップを必要としていることが分かる。詳細には、閾値 q が 0.2 より大きければ、単一ノードからなる初期集合の殆どで $h_q^j(S) = \emptyset$ となるが、そこまで短いステップで到達していると考えられる。一方、閾値 q が 0.1 のときには、多くの初期集合において $h_q^j(S)$ が全ノード集合 V に近付くが、その場合でも短いステップで到達していると考えられる。

図 6 には、ウィキペディアデータセットでのサイクルまでの最小ステップの評価結果を示す。ここでもまた、プログラムデータセットでの影響度の評価結果と比較して、閾値 q が 0.2 のとき、サイクルまでの最小ステップが大きい結果が無くなっているものの、全体の傾向としては、かなり類似した結果となっていることが分かる。すなわち、2 つのデータセットでは、同程度の閾値 q で類似したダイナミクスを示していることが示唆される。

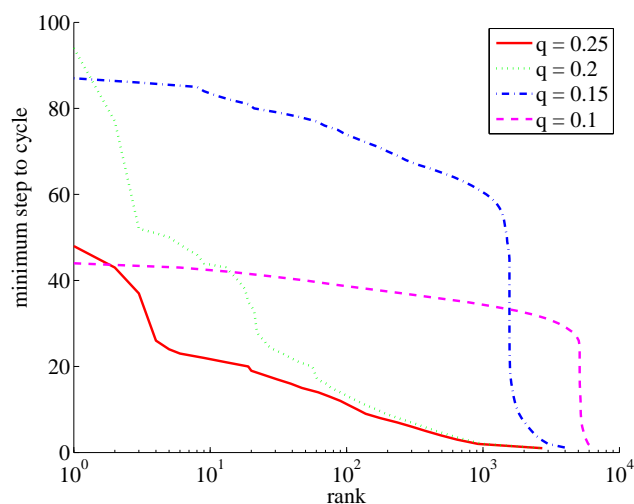


図 5: ブログデータでのサイクルまでの最小ステップ

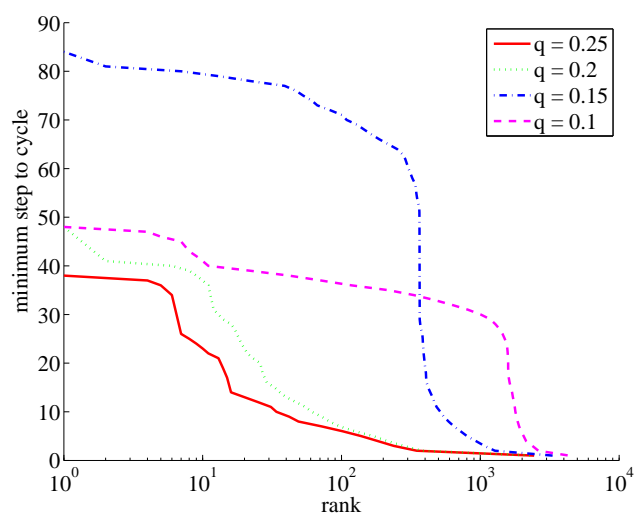


図 6: ウィキペディアデータでのサイクルまでの最小ステップ

5. おわりに

本稿では, Morris モデルなどの確定的な情報拡散モデルにおいて, ネットワーク構造やモデルのパラメータとダイナミクスとの関係を探求するため, 基本となる幾つかの分析法を提案した. 実世界ネットワークの 2 つのデータセットを用いた実験では, 閾値パラメータ q を変化させ, 影響度, サイクルのサイズ, および, サイクルまでの最小ステップに関する評価を行った. 評価結果より, これら 2 つのデータセットでは, ある程度リンク密度に違いがあるものの, 影響度, サイクルのサイズ, および, サイクルまでの最小ステップのすべてで, 同程度の閾値 q で類似した結果となっていることが分かる. このことより, ある種のネットワーク構造に対して, Morris モデルのダイナミクスは不変であることが示唆される. 今後の検討課題として, このような結果をもたらす基本メカニズムのさらなる探求とともに, 多様な構造のネットワークを用いた実証実験などをさらに進める予定である.

参考文献

- [Gruhl 04] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins, Information diffusion through blogspace, Proceedings of the 7th International World Wide Web Conference, pp. 107–117 (2004).
- [Kempe 03] D. Kempe, J. Kleinberg, and E. Tardos: Maximizing the spread of influence through a social network, Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146 (2003).
- [Kimura 07] M. Kimura, K. Saito, and R. Nakano: Extracting influential nodes for information diffusion on a social network. Proceedings of the 22nd AAAI Conference on Artificial Intelligence, pp. 1371–1376 (2007).
- [Kleinberg 07] J. Kleinberg, Cascading Behavior in Networks: Algorithmic and Economic Issues, In Algorithmic Game Theory (N. Nisan, T. Roughgarden, E. Tardos, V. Vazirani, eds.), Cambridge University Press (2007).
- [Morris 00] S. Morris, Contagion, Review of Economic Studies, 67:57–78 (2000).