

情報拡散系列からの効率的なホットスパン検出

Efficient Detection of Hot Span in Information Diffusion Sequences

大原剛三*1
Kouzou Ohara

斉藤和巳*2
Kazumi Saito

木村昌弘*3
Masahiro Kimura

元田浩*4
Hiroshi Motoda

*1 青山学院大学
Aoyama Gakuin University

*2 静岡県立大学
University of Shizuoka

*3 龍谷大学
Ryukoku University

*4 大阪大学
Osaka University

We addressed the problem of detecting the change in behavior of information diffusion from a small amount of observation data, where the behavior changes were assumed to be effectively reflected in changes in the diffusion parameter value. The problem is to detect where in time and how long this change persisted and how big this change is. We solved this problem by searching the change pattern that maximizes the likelihood of generating the observed diffusion sequences. The naive learning algorithm has to iteratively update the pattern boundaries, each requiring optimization of diffusion parameters by the EM algorithm, and is very inefficient. We devised a very efficient search algorithm using the derivative of likelihood which avoids parameter value optimization during the search. The results tested using real world network structures confirmed that the algorithm can efficiently identify the correct change pattern. We further compared our algorithm with the naive method that finds the best combination of change boundaries by an exhaustive search through a set of randomly selected boundary candidates, and showed that the proposed algorithm far outperforms the naive method both in terms of accuracy and computation time.

1. はじめに

今日、Web 空間上には様々な大規模ソーシャルネットワークが展開され、その中でのコミュニケーションが我々の日常生活における行動に大きな影響を与えつつある [Newman 03, Gruhl 04, Domingos 05, Leskovec 06]。そのようなソーシャルネットワークに関しては、ネットワーク構造に基づきノードを特徴づける中心性指標と呼ばれる幾つかの指標が提案されている [Wasserman 94, Bonacichi 87, Katz 53]。中心性指標はソーシャルネットワーク上の情報拡散において重要な役割を果たすノードの特定に利用できるが、可能な限り多くのノードに情報を伝搬させることのできる影響度の高いノード集合を見つける影響最大化問題 [Kempe 03, Kimura 10] などでは、ネットワーク構造だけに基ついた従来の中心性指標では不十分であり、情報拡散メカニズムを陽に用いることが重要となる [Kimura 10]。

一般には、そのメカニズムは確率モデルによってモデル化される。最も代表的で基本的なものは独立カスケード (IC: Independent Cascade) モデル [Goldenberg 01, Kempe 03] と線形閾値 (LT: linear threshold) モデル [Watts 02, Watts 07] であり、我々はさらに非同期時間遅れを考慮したそれらの拡張を提案している [Saito 09]。実際、これらのモデルを用いて同定した影響度の高いノードやリンクは中心性指標により同定されるものとは大幅に異なる。その一方で、これらのモデルを利用する為には、IC モデルにおける拡散確率や LT モデルにおけるノード重みなどの拡散パラメータを事前に指定する必要があるが、それらの真の値を知ることは現実的には不可能であるという新たな問題が生じる。そのため、我々は限られた情報拡散観測データからモデル (のパラメータ) を学習する手法を提案してきた [Saito 09, Saito 10]。

本研究はそれら一連の研究と同じ流れのものであるが、情

報拡散の異なった側面を取り扱う。具体的には、先に挙げた情報拡散モデルは一般には固定であると仮定され、隣接ノード間の相互作用のみを考慮しているのに対し、本研究では我々の行動が近傍の人からだけでなく、他の外的要因からも影響を受けていることに着目し、そのような外的要因により生じる情報拡散モデルの変化を限られた観測データから検出する問題を考える。そのような変化を正確に同定することは、特定の期間に生じ、情報拡散に影響を与えた特別な事象 (外的要因) を限られた観測データから推定する上で極めて重要となる。

実際には、情報拡散過程に変化をもたらす要因は多数存在し得るが、それらすべてをモデルに反映することはできないため、本研究では拡散パラメータの未知の変化としてそれらをモデル化し、その変化がいつ生じ、どのくらい持続し、そしてどの程度の大きさだったのかを検出する。問題の簡単化のために、ここでは扱う情報拡散モデルを非同期時間遅れ付き独立カスケード (AsIC: Asynchronous time delay Independent Cascade) モデル [Saito 09] に限定し、拡散確率の変化としては、ある時点で大きな値に変化し、ある期間その値を持続した後、元の値に戻り、その後は一定のままとなるような矩形的変化のみを考える。以下では、拡散確率が高い期間を“ホットスパン”、その他の期間を“ノーマルスパン”と呼ぶ。したがって、ここでの問題は観測データからホットスパンを検出し、ホットスパンとノーマルスパンにおける拡散確率を推定することとなる。そのために本稿では、文献 [Saito 09] と同じ最適化アルゴリズムを用いる。この最適化アルゴリズムでは、拡散モデルが観測データ系列を生成する尤度を最大化するようにモデルのパラメータ値を反復的に更新する。ホットスパン検出では、パラメータ最適化の反復処理に加え、ホットスパンの探索が必要となるため、さらに複雑な問題となる。もっともナイーブな方法は、ホットスパンの候補を反復的に更新し、各候補に対してパラメータ値を最適化することであるが、候補が多い場合には極めて非効率的となる。これに対し本稿では、尤度関数の拡散確率に関する一階導関数を利用し、候補ごとのパラメータ最適化を回避する非常に効率的で、かつ一般的なアルゴリズムを提案

連絡先: 大原剛三, 青山学院大学理工学部情報テクノロジー学科, 〒 252-5258 相模原市中央区淵野辺 5-10-1, ohara@it.aoyama.ac.jp

する．さらに，実世界ソーシャルネットワークを用い，ナイーブ法と提案アルゴリズムを実験的に比較し，提案アルゴリズムのほうがより効率的，かつより正確に拡散確率を推定し，ホットスパンを検出できることを示す．

2. 情報拡散モデル

AsIC モデルは，時間遅れを考慮していなかった従来の IC モデルに対して，実世界での情報伝搬に見られる非同同期時間遅れの概念を導入したものである．以下に AsIC モデルの形式的な定義について述べる．なお，本稿では議論の簡単化のために遅延時間は指数分布に従うものとするが，べき乗分布など他の分布も同様に用いることが可能である．

以下では，ソーシャルネットワークを有向グラフ $G = (V, E)$ で表す．ここで， V と $E (\subset V \times V)$ はそれぞれ全ノードの集合と全リンクの集合を表す．また，すべてのノード $v \in V$ について， v を始点とするリンクの終点ノードの集合を $F(v) = \{u \in V; (v, u) \in E\}$ ， v を終点とするリンクの始点ノードの集合を $B(v) = \{u \in V; (u, v) \in E\}$ と表す．このとき，ソーシャルネットワーク G 上の情報拡散において，あるノードが対象情報の影響を受けた場合，そのノードはアクティブであると呼び，各ノードはアクティブと非アクティブという二つの状態のいずれか一方の状態を取るものとする．ここで，ノードの状態は非アクティブからアクティブに変わり得るが，その逆の変化は生じないと仮定する．

AsIC モデルは，各リンク $(u, v) \in E$ に対して $0 < p_{u,v} < 1$ という値を取る拡散確率 $p_{u,v}$ と， $r_{u,v} > 0$ という値を取る遅延パラメータ $r_{u,v}$ という 2 つのパラメータをもつ．このとき，AsIC モデルにおける情報拡散過程は与えられた初期アクティブノード（情報源ノード）を起点として連続時間 t の下で次のように展開される．あるノード u が時刻 t でアクティブとなったとき， u はその時点で非アクティブであるノード $v \in F(u)$ をアクティブにする機会を 1 度だけ与えられる．このとき，遅延時間 δ が $r_{u,v}$ をパラメータとする指数分布に従い決められる．そして， u は時刻 $t + \delta$ までに v がアクティブになっていなければ v をアクティブにすることを試み，その試行は確率 $p_{u,v}$ で成功する． u の試行が成功した場合， v は時刻 $t + \delta$ にアクティブとなる．ノードをアクティブにする試行がそれ以上実行できなくなった時点で，この情報拡散過程は終了する．

3. 問題設定

本節では，本稿が取り扱う“ホットスパン検出問題”を形式的に定義する．本研究では，情報拡散過程で何らかの変化が生じ，その変化が内包されたあるトピックについての拡散系列を観察するものと仮定する．このとき，ホットスパン検出問題では，その変化がいつ生じ，どのくらいの長さ持続し，そしてどの程度の大きさであったのかを検出することを目的とする．ここで，ある 1 つのトピックに関する拡散系列に注目していること，そして同じトピックを話題にする際の人々の行動は極めて類似したものになると仮定できること [Saito 09, Saito 10] から，本稿では AsIC モデルのパラメータ $p_{u,v}$ と $r_{u,v}$ はリンク (u, v) には依存せず一定の値を取るという制約をおく．すなわち， $p_{u,v} = p$ ， $r_{u,v} = r$ ($\forall (u, v) \in E$) とする．

次に，ホットスパン検出問題を数学的に定式化する．いま， $[T_1, T_2]$ をあるトピックの拡散過程におけるホットスパン， p_1 ， p_2 をノーマルスパン，ホットスパンそれぞれにおける AsIC モデルの拡散確率であるとする．ここで， $p_1 < p_2$ である．一方，AsIC モデルの下でのあるトピックの拡散結果はアクティ

ブなノードとそのノードがアクティブとなった時刻のペアの集合 $\{(u, t_u), (v, t_v), \dots\}$ により表現される．このとき，期間 $[0, T_1]$ については p_1 ，期間 $[T_1, T_2]$ については p_2 ，そして期間 (T_2, ∞) については p_1 という拡散確率をもつ AsIC モデルによって生成された拡散結果 D を考える．なお，遅延パラメータは全期間 $[0, \infty)$ において同じ値 r を取るものとする．このようにして得られる拡散結果 D を“ホットスパンを伴う拡散結果”と呼ぶ．このとき，ホットスパン検出問題は観測されたホットスパンを伴う拡散結果から，時区間 $[T_1, T_2]$ を検出し，拡散確率 p_1 ， p_2 を推定する問題と定義できる．

4. ホットスパン検出法

以下では， M 本の独立な情報拡散系列からなる観測データ集合 $\{D_m; m = 1, \dots, M\}$ を考える．ここで，各 D_m は m 番目の拡散系列におけるアクティブなノードとそのノードがアクティブになった時刻のペアの集合であり， $D_m = \{(u, t_{m,u}), (v, t_{m,v}), \dots\}$ と定義する．また，各 D_m に対して観測開始時刻 $\phi_m = \min\{t_{m,v}; (v, t_{m,v}) \in D_m\}$ ，および観測最終時刻 $\Phi_m = \max\{t_{m,v}; (v, t_{m,v}) \in D_m\}$ を定義する．ここで， Φ_m は最後に観測されたノードがアクティブになった時刻と必ずしも一致する必要はない．このとき，観測データを $\mathcal{D}_M = \{(D_m, \Phi_m); m = 1, \dots, M\}$ と表記する．そして，すべての $t \in [\phi_m, \Phi_m]$ に対して， $C_m(t) = \{v; (v, t_{m,v}) \in D_m, t_{m,v} < t\}$ とする．すなわち， $C_m(t)$ は m 番目の拡散系列において時刻 t より前にアクティブとなったノードの集合となる．以下では，便宜上， C_m は m 番目の拡散系列におけるすべてのアクティブなノードの集合を表すものとする．

4.1 パラメータ学習

ホットスパンが存在しない場合，AsIC モデルのパラメータ値 p ， r を推定する為の対数尤度関数は次式で与えられる [Saito 09]：

$$\begin{aligned} \mathcal{L}(\mathcal{D}_M; p, r) &= \sum_{m=1}^M \mathcal{L}((D_m, \Phi_m); p, r) \\ &= \sum_{m=1}^M \sum_{v \in C_m} \left(\log h_{m,v} + \sum_{w \in F(v) \setminus C_m} \log g_{m,v,w} \right). \end{aligned} \quad (1)$$

ここで， $h_{m,v}$ は m 番目の拡散系列において $t_{m,v} > 0$ であるノード $v \in D_m$ が時刻 $t_{m,v}$ にアクティブとなる確率密度であり， $g_{m,v,w}$ はリンク $(v, w) \in E$ が存在するときにノード w がノード v によってアクティブにされない確率である． p と r の値は式 (1) を EM アルゴリズムを用いて最大化することにより安定的に求めることができる [Saito 09]．

次に，ホットスパンが存在する場合を考える．ホットスパン $S = [T_1, T_2]$ に対しては，以下のようなパラメータ切替が起こるものとする．

$$p = \begin{cases} p_1 & \text{if } v \in \mathcal{N}_m(S), \\ p_2 & \text{if } v \in \mathcal{H}_m(S). \end{cases}$$

ここで， $\mathcal{N}_m(S)$ と $\mathcal{H}_m(S)$ は m 番目の拡散系列におけるノーマルスパンとホットスパンそれぞれでアクティブとなったノードの集合である．このパラメータ変化を考慮して式 (1) を適切に修正することにより拡張した目的関数 $\mathcal{L}(\mathcal{D}_M; p_1, p_2, r, S)$ を定義することができる．明らかに，十分な量の観測データがあれば， $\mathcal{L}(\mathcal{D}_M; p_1, p_2, r, S)$ は S を真の区間 $S^* = [T_1^*, T_2^*]$ に設定することにより最大化される．ゆえに，ホットスパン検出問題では， $\mathcal{L}(\mathcal{D}_M; \hat{p}_1, \hat{p}_2, \hat{r}, S)$ を最大化する $S = \hat{S}$ を見つけ

ることが目的となる．ここで， \hat{p}_1 ， \hat{p}_2 ，および \hat{r} は与えられた S に対する最尤推定量を表す．そのような \hat{S} を求めるためには，妥当な候補区間の集合 S を用意する必要がある． S を用意する 1 つの方法は，すべての観測点（あるノードがアクティブとなった時刻）のすべての組合せを取ることである．すなわち， $\mathcal{T} = \{t_1, \dots, t_N\}$ を \mathcal{D}_M における観測点の集合としたとき， $S = \{S = [t_1, t_2] : t_1 < t_2, t_1 \in \mathcal{T}, t_2 \in \mathcal{T}\}$ とする．

4.2 ナイーブ法

S から \hat{S} を求める最もナイーブな方法は，2 重の反復処理を内包する．まず内側の反復処理では，目的関数 $\mathcal{L}(\mathcal{D}_M; p_1, p_2, r, S)$ を EM アルゴリズムを用いて最大化し，各候補区間 S に対する最尤推定量 \hat{p}_1 ， \hat{p}_2 ， \hat{r} を求める．次に外側の反復処理にて， $\mathcal{L}(\mathcal{D}_M; \hat{p}_1, \hat{p}_2, \hat{r}, S)$ を最大化する \hat{S} を選択する．このナイーブ法は，観測点数 N が大きいときには極めて非効率的であるため，その計算コストを許容可能なものにするためには， \mathcal{T} から K 個の観測点を選び，候補区間数を制限する必要がある．すなわち， $\mathcal{T}_K = \{t_1, \dots, t_K\}$ としたとき， S の代わりに $S_K = \{S = [t_1, t_2] : t_1 < t_2, t_1 \in \mathcal{T}_K, t_2 \in \mathcal{T}_K\}$ を用いる．ここで， $|S_K| = K(K-1)/2$ であり， K が大きくなるとやはり $|S_K|$ は非常に大きな値となることに注意されたい．

4.3 提案法

前述のナイーブ法は，計算コストを犠牲にして K を大きくすれば十分な精度でホットスパンを検出できるが，計算コストを削減するために K を小さくすれば検出精度は低下する．本節では，この問題を軽減し，観測データ \mathcal{D}_M から効率的，かつ安定的にホットスパンを検出する新たな手法を提案する．提案法では，まず元の目的関数である式 (1) に基づき最尤推定量 \hat{p} ， \hat{r} を求める．このとき，各観測点におけるノードに対して拡散確率 p に関する目的関数 $\mathcal{L}(\mathcal{D}_M; \hat{p}, \hat{r})$ の一階導関数を考えると，式 (1) における一様なパラメータ設定と局所最適性により以下の関係が成り立つことに着目する．

$$\frac{\partial \mathcal{L}(\mathcal{D}_M; \hat{p}, \hat{r})}{\partial p} = \sum_{(u,v) \in E} \frac{\partial \mathcal{L}(\mathcal{D}_M; \hat{p}, \hat{r})}{\partial p_{u,v}} = 0. \quad (2)$$

ここで， $p_{u,v}$ はノード u からノード v への拡散確率である．そして，区間 $S = [T_1, T_2]$ に対して以下の部分和を考える．

$$G(S) = \sum_{m=1}^M \sum_{(u,v) \in E, u \in \mathcal{H}_m(S)} \frac{\partial \mathcal{L}(\mathcal{D}_m, \Phi_m; \hat{p}, \hat{r})}{\partial p_{u,v}}. \quad (3)$$

明らかに，ここでの問題設定においては $S \approx S^*$ であれば $G(S)$ は十分大きな値となり， $p_2 > \hat{p} > p_1$ となる．ゆえに， $G(S)$ を最大化する \hat{S} を探すことによりホットスパン S^* を推定することが可能となる．この方法の利点は，紙面の都合上，式の導出は割愛するが， $\mathcal{T} = \{t_1, \dots, t_N\}$ とし， $i < j$ であれば $t_i < t_j$ としたとき，各候補区間 S を検査するために必要な $G(S)$ の計算において， $G([t_i, t_j])$ を用いて $G([t_i, t_{j+1}])$ を漸進的に計算できることである．この計算コストは，ナイーブ法におけるパラメータ最適化のコストよりもはるかに小さいため，候補区間集合 S を構成する際に可能な観測点すべての組合せを利用することができる．以下に，提案法をまとめる．

1. EM アルゴリズムを用いて $\mathcal{L}(\mathcal{D}_M; p, r)$ を最大化する．
2. 候補時点集合 \mathcal{T} と候補区間集合 S を生成する．
3. $G(S)$ を最大化するホットスパン \hat{S} を検出し，出力する．
4. EM アルゴリズムを用いて $\mathcal{L}(\mathcal{D}_M; p_1, p_2, r, \hat{S})$ を最大化し，得られた最尤推定量 \hat{p}_1 ， \hat{p}_2 ， \hat{r} を出力する．

ここで，提案法では最大化問題を解く EM アルゴリズムを 2 回しか用いないことに注意されたい．

5. 評価実験

5.1 実験設定

本研究では，複数の実世界ソーシャルネットワークを用い，提案法がホットスパン，およびホットスパンとノーマルスパンにおける拡散確率の両方をどれだけ正確に推定できるかをその効率とともに実験的に評価した．比較対象としては，前述のナイーブ法を用い，そのパラメータ K に関しては，5，10，および 20 を用いた．なお，本実験では $M = 1$ の場合，すなわち単一の拡散系列からの学習という最も難しいケースを対象とした．紙面の都合上，それらの実験結果のうち，文献 [Kimura 09] で用いられた日本語ブログにおけるトラックバックネットワークを対象にしたもののみを示す．このネットワークは 12,047 個のノードと 79,920 本の有向リンクをもつ双方向連結ネットワークであり，以下ではブログネットワークと呼ぶ．

本実験では，各ソーシャルネットワークに対して，ホットスパンを伴う拡散系列を AsIC モデルを用いたシミュレーションで生成して利用した．ノーマルスパンにおける拡散確率 p_1 は，文献 [Kempe 03] に従いネットワークの平均出次数 d に対して $1/d$ よりも小さな値となるように設定し，ホットスパンにおける拡散確率 p_2 は p_1 の 3 倍とした．ブログネットワークにおける p_1 と p_2 の値は，それぞれ 0.1 と 0.3 である．遅延パラメータは $r = 1$ とした*1．また，ホットスパンは予備実験の結果に基づき $[T_1 = 10, T_2 = 20]$ とした．これらのパラメータ設定の下，初期情報源ノードを無作為に選択し，異なる 5 本の情報拡散系列を生成した．以下に示す実験結果は，それらを用いた 5 回試行の平均である．なお，いずれの試行も Intel Core 2 Duo 3GHz プロセッサをもつ計算機上で実行した．

5.2 実験結果

まず，図 1(a) にホットスパン \hat{S} の推定精度を絶対誤差 $\mathcal{E}_s = |\hat{T}_1 - T_1| + |\hat{T}_2 - T_2|$ を用いて示す．この結果から，提案法がナイーブ法の推定精度をはるかに上回る良好な推定精度を達成していることがわかる．ナイーブ法に関しては，予想通り K が大きくなるにつれて誤差 \mathcal{E}_s が減少しているが，最良の場合 ($K = 20$) でさえ，その平均誤差は提案法より約 10 倍も大きくなっている．次に，図 1(b) に p_1 と p_2 の推定精度を相対誤差 $\mathcal{E}_p = |\hat{p}_1 - p_1|/p_1 + |\hat{p}_2 - p_2|/p_2$ で示す．同様に，ナイーブ法の相対平均誤差は K が大きくなるにつれて減少しているが，最良の場合 ($K = 20$) でさえ，その値は提案法の約 3 倍となっている．最後に，図 1(c) に各手法の計算時間を示す．この結果から，提案法がナイーブ法に比べてはるかに高速であることがわかる．この計算時間の違いは，EM アルゴリズムの実行回数の差によるものと考えられる．提案法は前述のように 2 度しか EM アルゴリズムを実行しないが，ナイーブ法では全候補区間 $S \in S_K$ に対して EM アルゴリズムを実行する必要があり，その回数は $|S_K| = K(K-1)/2$ 回となる．実際に，ナイーブ法の $K = 5$ に対する計算時間は提案法の 5 倍程度であり，これはナイーブ法の EM アルゴリズム実行回数 $|S_K| = 10$ と提案法の 5 倍となっていることと符合する．この関係は， $K = 10$ ， $K = 20$ の場合についても概ね成り立つ．これらの結果は，たとえナイーブ法が K を十分大きな値に設定して高い推定精度を得たとしても，その代わりに許容しがたい計算時間が必要となることを意味している．

*1 r の変化は拡散結果の時間軸スケールを変化させるだけであり，評価実験においては本質的ではない．

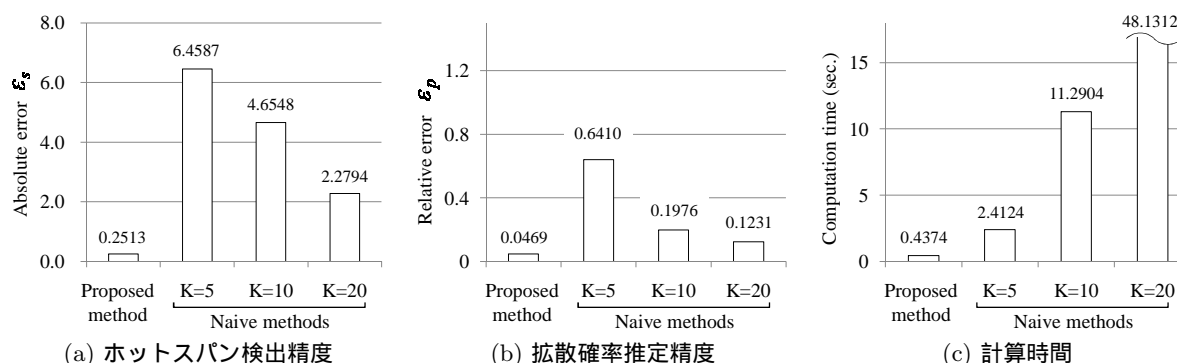


図 1: ブログネットワークにおける提案法とナイーブ法の性能比較

6. おわりに

本論文では、社会ネットワーク上の情報拡散が非同期時間遅れ付き独立カスケードモデルに従うという仮定の下、限られた数の観測データから情報拡散モデルにおける変化を検出する問題を検討した。具体的には、拡散パラメータ（拡散確率）が他の期間と比較して相対的に高い値となる期間をホットスパン、それ以外の期間をノーマルスパンとし、観測データからホットスパンを検出し、両スパンにおける拡散パラメータを推定する手法を提案した。提案法では、情報拡散モデルが観測データを生成する尤度を与える尤度関数の拡散確率に関する一階導関数を利用し、ホットスパンの候補区間ごとのパラメータ最適化を回避することを可能としている。実社会ソーシャルネットワークからシミュレーションにより生成した情報拡散系列を用いた評価実験では、無作為に選択した観測点から構成される候補区間ごとにモデルのパラメータを最適化するナイーブ法と比較して、精度と効率の両方の点において提案法がはるかに優れていることを確認した。なお、本稿では非常に単純化した問題設定を仮定し、紙面の都合上、詳細な議論は割愛したが、提案法はより複雑な拡散パラメータの変化にも適用可能である。今後の課題としては、同様のアプローチの下で時空間的なホットスパンを検出することが挙げられる。

謝辞

本研究は、科学研究費補助金基盤研究 (C) (No. 23500194) の補助を受けた。

参考文献

- [Bonacichi 87] Bonacichi, P.: Power and centrality: A family of measures, *Amer. J. Sociol.*, Vol. 92, pp. 1170–1182 (1987)
- [Domingos 05] Domingos, P.: Mining social networks for viral marketing, *IEEE Intell. Syst.*, Vol. 20, pp. 80–82 (2005)
- [Goldenberg 01] Goldenberg, J., Libai, B., and Muller, E.: Talk of the network: A complex systems look at the underlying process of word-of-mouth, *Market. Lett.*, Vol. 12, pp. 211–223 (2001)
- [Gruhl 04] Gruhl, D., Guha, R., Liben-Nowell, D., and Tomkins, A.: Information diffusion through blogspace, *SIGKDD Expl.*, Vol. 6, pp. 43–52 (2004)
- [Katz 53] Katz, L.: A new status index derived from sociometric analysis, *Sociometry*, Vol. 18, pp. 39–43 (1953)
- [Kempe 03] Kempe, D., Kleinberg, J., and Tardos, E.: Maximizing the spread of influence through a social network, in *KDD 2003*, pp. 137–146 (2003)
- [Kimura 09] Kimura, M., Saito, K., and Motoda, H.: Blocking links to minimize contamination spread in a social network, *ACM Trans. Knowl. Discov. Data*, Vol. 3, pp. 9:1–9:23 (2009)
- [Kimura 10] Kimura, M., Saito, K., Nakano, R., and Motoda, H.: Extracting Influential Nodes on a Social Network for Information Diffusion, *Data Min. Knowl. Disc.*, Vol. 20, pp. 70–97 (2010)
- [Leskovec 06] Leskovec, J., Adamic, L. A., and Huberman, B. A.: The dynamics of viral marketing, in *EC'06*, pp. 228–237 (2006)
- [Newman 03] Newman, M. E. J.: The structure and function of complex networks, *SIAM Rev.*, Vol. 45, pp. 167–256 (2003)
- [Saito 09] Saito, K., Kimura, M., Ohara, K., and Motoda, H.: Learning Continuous-Time Information Diffusion Model for Social Behavioral Data Analysis, in *ACML 2009*, pp. 322–337 (2009)
- [Saito 10] Saito, K., Kimura, M., Ohara, K., and Motoda, H.: Selecting information diffusion models over social networks for behavioral analysis, in *ECML PKDD 2010*, pp. 180–195 (2010)
- [Wasserman 94] Wasserman, S. and Faust, K.: *Social network analysis*, Cambridge Univ. Press, Cambridge, UK (1994)
- [Watts 02] Watts, D. J.: A simple model of global cascades on random networks, *PNAS*, Vol. 99, pp. 5766–5771 (2002)
- [Watts 07] Watts, D. J. and Dodds, P. S.: Influence, networks, and public opinion formation, *J. Consum. Res.*, Vol. 34, pp. 441–458 (2007)