

# 社会ネットワーク上の強影響度ノード同定のための リサンプリングに基づく予測シミュレーション法の提案

Resampling-based Predictive Simulation for Identifying Influential Nodes over Social Network

大原剛三\*1 斉藤和巳\*2 木村昌弘\*3 元田浩\*4  
Kouzou Ohara Kazumi Saito Masahiro Kimura 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 address a problem of efficiently estimating the influence of a node in information diffusion over a social network. Since the information diffusion is a stochastic process, the influence degree of a node is quantified by the expectation, which is usually obtained by very time consuming many runs of simulation. Our contribution is that we proposed a framework for predictive simulation based on the leave- $N$ -out cross validation technique that estimates the approximation error of the influence degree of each node without knowing the true influence degree. We experimentally evaluate the proposed method using the two real world networks, and show that it can serve as a good measure to solve the problem with far fewer runs of simulation ensuring the accuracy if  $N$  is appropriately chosen.

## 1. はじめに

今日、Facebook や Twitter などのソーシャルメディアの爆発的な普及により、インターネット上に巨大な社会ネットワークが形成されつつある。一旦、ソーシャルメディアに投稿された情報は、ときには複数のメディアを跨ぎつつ、それらの上に展開されている社会ネットワークを通して急速、かつ広範囲に拡散され、非常に多くの人々に共有され得る。そのような情報、およびその拡散現象は、我々の日常における意思決定にも多大な影響を与えることから、近年、社会学のみならず計算機科学も含めた多様な分野において盛んに研究されており [Kleinberg 08]、巨大ネットワーク上を伝播する膨大な情報とその影響を効率的に分析することが急務になりつつある。

そのような社会ネットワーク上の情報拡散に関する研究の中で、特に重要と考えられるのは、情報の拡散を最大化するノード集合を同定する影響最大化問題 [Kempe 03] に関するものである。この問題は、いわゆるロコマーケティングなどに応用され得るものであり、情報拡散を最小化するために情報の伝播を阻止すべきリンク集合を求める問題 [Kimura 09] やネットワークの外部から新たな情報を投入する際に、その拡散を最大化するような目標ノード集合を同定する問題 [Saito 13] など、いずれもこの問題の亜種とみなせる。

一方、これらの問題を解くためには、ノードの影響度を推定し、それに基づいてノードを順位づけることが必要となる。ここで、情報拡散プロセスは確率過程であるため、ノード  $v$  の影響度は、 $v$  を情報源とした情報拡散プロセス終了時において、その情報を受け取ったノード数の期待値として定義される。この考えの下、Kempe らは多数の情報拡散シミュレーション結果の平均を用いてノードの影響度を近似し、貪欲探索に基づき影響最大化問題の解となる最適なノード集合を選択する手法を提案している [Kempe 03]。しかし、影響度を精度よく近似するためには膨大な回数のシミュレーション試行が必要となり、多大な計算時間を要する。そのため、効率を改善するための様々な手法 [Kimura 10, Leskovec 07, Goyal 11] が提案され

ているが、試行回数を本質的に削減するには至っていない。

以上のような背景の下、本研究では効率的にノードの影響度を推定するために、限られたシミュレーション試行回数から得られた影響度の近似値の真の影響度に対する近似誤差を精度よく推定する手法を提案する。具体的には、独立試行されたシミュレーションの結果を、ある特定の確率分布から独立に生成されたサンプルとみなし、leave- $N$ -out 法の考えに基づいたサンプリングの枠組みを導入することで、真の影響度を仮定することなく近似誤差を推定する。そして、サンプリングサイズに応じた 2 つの手法を提案し、実ネットワークを用いた評価実験を通して、平均絶対誤差、および二乗平均平方誤差の両方の観点から、数十から数百回程度の試行回数でも提案法が高い精度で影響度の高い強影響度ノードに対する近似誤差を推定可能なことを示す。

## 2. 情報拡散モデル

本研究では、有向グラフ  $G = (V, E)$  により表現される社会ネットワーク上の情報拡散を考える。ここで、 $V, E$  ( $C \ V \times V$ ) はそれぞれネットワーク中のノードとリンクの集合である。以下、情報拡散プロセスにおいて情報の影響を受けた (情報を受け取った) ノードを**アクティブ**、その他のノードを**非アクティブ**と呼び、ノードの状態は非アクティブからアクティブの一方方向にのみ遷移するものと仮定する。また、ノード  $u$  からノード  $v$  へのリンク  $(u, v)$  に対して、 $u$  を  $v$  の**親ノード**、 $v$  を  $u$  の**子ノード**と呼び、任意のノード  $v \in V$  に対して  $\Gamma(v)$  はネットワーク  $G$  における  $v$  のすべての親ノードの集合とする。本稿では、ノード  $v$  がその親ノード  $\Gamma(v)$  のみから直接影響を受ける (情報を受け取る) 確率的カスケードモデルに着目し、以下では簡単化のために、もっとも基本的な離散時間モデルである**独立カスケード (IC: Independent cascade) モデル** [Kempe 03] に議論を絞る。ただし、本稿での議論はより一般的な連続時間モデル [Saito 09] などにも容易に拡張できる。

IC モデルでは、ノード  $u$  から  $v$  に情報が伝わる確率として、過去の拡散過程に依存しない定数である**拡散確率**  $p_{u,v}$  ( $0 \leq p_{u,v} \leq 1$ ) が各リンク  $(u, v)$  にパラメータとして事前に与えられる。この前提の下、IC モデルにおける情報拡散過程は、時

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

刻 0 においてアクティブであるノードの集合  $V_0 \subset V$  が与えられたとき、次のように進行する。ノード  $u$  が時刻  $t$  にてアクティブになった場合、 $u$  はその時点で非アクティブな子ノード  $v$  をアクティブにする機会を 1 度だけ与えられ、その試行は確率  $p_{u,v}$  で成功する。その試行が成功した場合、 $v$  は時刻  $t+1$  でアクティブとなる。 $v$  の複数の親ノードが時刻  $t$  に同時にアクティブとなった場合、それらの親ノードは任意の順序で  $v$  をアクティブにすることを試みるが、いずれの試行も時刻  $t$  で実行される。一方、親ノード  $u$  はその試行の成否に関わらず、それ以降、 $v$  をアクティブにすることを試みることはできない。この情報拡散過程は、いずれの非アクティブノードに対してもアクティブにする試行が実行できなくなった時点で終了する。

ここで、 $A(V_0)$  を拡散過程終了時点でのアクティブノード数とする。このとき、 $A(V_0)$  は確率変数であり、以下ではその期待値を  $\langle A(V_0) \rangle$  と表し、この値をノード集合  $V_0$  の影響度とする。本稿では、特に集合  $V_0$  が単一ノード  $v_0$  のみからなる場合を考え、 $A(\{v_0\})$ ,  $\langle A(\{v_0\}) \rangle$  をそれぞれ単に  $A(v_0)$ ,  $\langle A(v_0) \rangle$  と記述する。本研究では、任意のノード  $v \in V$  に対する  $\langle A(v) \rangle$  を、複数の情報拡散シミュレーション結果の平均により近似することを考える。いま、 $S = \{1, \dots, |S|\}$  なる集合  $S$  を考え、ノード  $v$  を初期アクティブノードとした情報拡散シミュレーションを  $|S|$  回試行するものとし、各試行後のアクティブノード数を  $A_s(v)$  とする ( $s \in S$ )。このとき、任意の  $v \in V$  に対して、 $\bar{A}_S(v)$  を次のように定義する。

$$\bar{A}_S(v) = \frac{1}{|S|} \sum_{s \in S} A_s(v) \quad (1)$$

$|S|$  が十分に大きければ、 $\bar{A}_S(v)$  はノード  $v \in V$  の IC モデルにおける影響度  $\langle A(v) \rangle$  のよい近似となる。

### 3. 予測シミュレーションによる影響度推定

式 (1) によりノード  $v \in V$  の影響度  $\langle A(v) \rangle$  を近似する場合、 $|S|$  をどう設定すればよいか問題となる。実際には、 $|S|$  の大きさが十分であるかどうかは、その結果得られた  $\bar{A}_S(v)$  が真の影響度  $\langle A(v) \rangle$  にどれだけ近いかにより判断することになる。そのような近似誤差の指標としては、一般に平均絶対誤差 (MAE: Mean Absolute Error) と二乗平均平方根誤差 (RMSE: Root Mean Squared Error) の 2 つが多用される。具体的には、MAE と RMSE に基づいたノード  $v$  の影響度の近似誤差はそれぞれ以下のように定義できる。

$$D_S^{(MAE)}(v) = \langle |A(v) - \bar{A}_S(v)| \rangle. \quad (2)$$

$$D_S^{(RMSE)}(v) = \sqrt{\langle (A(v) - \bar{A}_S(v))^2 \rangle}. \quad (3)$$

一方、上記の近似誤差はいずれもその定義に真の影響度  $\langle A(v) \rangle$  を含むため、試行回数  $|S|$  の削減を目的とする本研究では直接用いることはできない。そのため本研究では、 $\langle A(v) \rangle$  に対する  $\bar{A}_S(v)$  の近似誤差を真の影響度  $\langle A(v) \rangle$  を仮定せずに、限られた数のシミュレーション結果  $\{A_s(v) | s \in S\}$  のみから推定することを考える。近似誤差が一定精度で推定可能であれば、盲目的に  $|S|$  に大きな値を設定することなく、その誤差が十分小さくなった時点で試行を止めることが可能となる。

そのために本稿では、 $|S|$  回のシミュレーションは独立に試行されることから、その実行結果の集合  $\{A_s(v) | s \in S\}$  は、平均  $\mu(v)$ 、標準偏差  $\sigma(v)$  であるようなある確率分布に従って独立に生成されたサンプル集合とみなせることに着目し、leave-N-out 交差検証に基づく予測シミュレーション法を提案する。

いま、 $N < |S|$  である正の整数  $N$  に対して、 $\mathcal{B} \subset 2^S$  を要素数が  $N$  となる  $S$  の部分集合の族であるとする。すなわち、任意の  $B \in \mathcal{B}$  に対して  $|B| = N$  が成り立つ。このとき、 $|S|$  回の試行から経験的に得られるノード  $v$  の影響度  $\bar{A}_S(v)$  と真の影響度  $\langle A(v) \rangle$  の近似誤差を以下の式により推定する。

$$\bar{D}_S(v; N) = \sqrt{\langle (\bar{A}_S(v) - \bar{A}_{S \setminus B}(v))^2 \rangle_{B \in \mathcal{B}}} \quad (4)$$

紙面の都合上、詳細な式の導出は割愛するが、この式は次のように書き換えることが可能である。

$$\bar{D}_S(v; N) = C_S(N) \bar{\sigma}_S(v) \quad (5)$$

ここで、 $C_S(N)$  は次のように定義される  $N$  の関数である。

$$C_S(N) = \sqrt{\frac{N}{(|S|-1)(|S|-N)}} \quad (6)$$

また、 $\bar{\sigma}_S(v)$  は、 $|S|$  回のシミュレーション結果  $\{A_s(v) | s \in S\}$  から経験的に導かれる標準偏差であり、次のように定義できる。

$$\bar{\sigma}_S(v) = \sqrt{\frac{1}{|S|} \sum_{s \in S} (A_s(v) - \bar{A}_S(v))^2} \quad (7)$$

すなわち、ここでの提案法は、 $\{A_s(v) | s \in S\}$  から  $B \in \mathcal{B}$  に対応するサイズ  $N$  の部分集合をサンプリングし、残りの結果から得られる影響度  $\bar{A}_{S \setminus B}(v)$  と全結果から得られる影響度  $\bar{A}_S(v)$  の  $B \in \mathcal{B}$  に対する RMSE を計算し、それにより真の近似誤差を推定するものである。

ここで、 $N$  をどう設定するかが問題となる。本稿では、2 つの極端な場合を考える。1 つは、 $N = 1$  ( $\bar{D}_S(v; 1)$ ) の場合であり、このとき  $C_S(1) = 1/(|S|-1)$  となる。これは、Leave-One-Out 交差検証と同じ枠組みとなるため、以下では  $N = 1$  とした提案法を *LOO 法* と呼ぶ。もう一方は、 $N = |S|/2$  ( $\bar{D}_S(v; |S|/2)$ ) の場合であり、このとき  $C_S(|S|/2) = \sqrt{1/(|S|-1)}$  となる。これは、Leave-Half-Out 交差検証と同じ枠組みとなるため、以下では  $N = |S|/2$  とした提案法を *LHO 法* と呼ぶ。

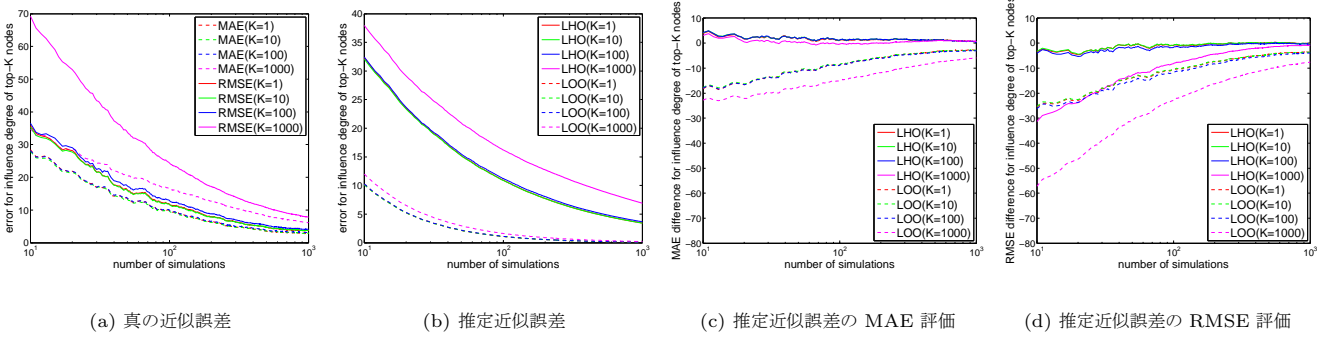
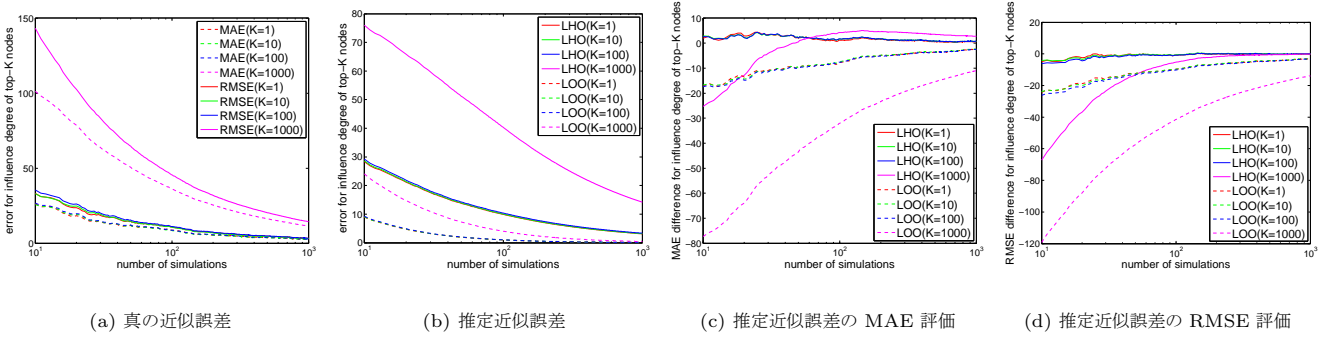
## 4. 評価実験

### 4.1 実験設定

前節で提案した 2 つの手法を、2 つの実ネットワークを用いて実験的に評価した。実験で用いた実ネットワークのうち 1 つは、化粧品に関する口コミサイトである“@cosme”<sup>\*1</sup> から抽出したお気に入りネットワークである。本実験では、2009 年 12 月の時点でランダムに選択したユーザを起点にお気に入りリンクを 10 ステップまで辿ることで生成したノード数 45,024、リンク数 351,299 の有向ネットワークを用いた。以下、コスメネットワークと呼ぶ。もう 1 つは、エンロン E メールデータセット [Klimt 04] に対して、メールの送信者、もしくは受信者をノードとし、各メールの送信者と受信者間にリンクを生成したノード数 19,603、リンク数 210,950 の有向ネットワークである。以下、エンロンネットワークと呼ぶ。

本実験では、上記のネットワークに対して、ボンドパレーション法 [Kimura 10] による独立した 100 万回の情報拡散シミュレーションにおいてノード  $v$  がアクティブにしたノード数の平均値を真の影響度  $\langle A(v) \rangle$  として用いた。ボンドパ

\*1 <http://www.cosme.net/>


 図 1: コスメネットワークにおける影響度上位  $K$  ノードに対する実験結果

 図 2: エンロンネットワークにおける影響度上位  $K$  ノードに対する実験結果

コレクション法の詳細は参考文献を参照されたい. なお, IC モデルの拡散確率  $p_{u,v}$  に関しては, 平均出次数  $\bar{d}$  に基づいた一様な値  $p = 1/\bar{d}$  を用いた. 具体的には, コスメネットワークでは  $p = 0.13$ , エンロンネットワークでは  $p = 0.09$  とした.

一方, LOO 法と LHO 法の試行回数は  $|S| = 1,000$  とし, 各ノード  $v \in V$  について, 各手法により推定された近似誤差  $\bar{D}_{S(s)}(v; N)$  を MAE で評価した真の近似誤差  $D_{S(s)}^{(MAE)}(v)$ , および RMSE で評価した真の近似誤差  $D_{S(s)}^{(RMSE)}(v)$  と比較した. ここで,  $S(s) = \{1, \dots, s\}$  である. また, ここでの興味の対象は影響度の大きい強影響度ノードであるため, 真の影響度における上位  $K$  ノードに対する近似誤差のみを考え, 具体的には  $K = 10^0, 10^1, 10^2, 10^3$  の 4 通りについて評価した. 近似誤差の計算に関しては,  $|S| = 1,000$  回の試行を  $M = 100$  回繰り返し, その平均を用いた. いま,  $D_k(s)$  と  $\bar{D}_k(s)$  をそれぞれ上位  $k$  番目のノードに対して最初の  $s$  回試行から得られた真の近似誤差と提案法による推定近似誤差とすると, MAE を用いた場合の  $D_k(s)$  は,

$$D_{S(s)}^{(MAE)}(v_k) = \frac{1}{M} \sum_{m=1}^M | \langle A(v_k) \rangle - \bar{A}_{S(s),m}(v_k) | \quad (8)$$

となり, RMSE を用いた場合の  $D_k(s)$  は,

$$D_{S(s)}^{(RMSE)}(v_k) = \sqrt{\frac{1}{M} \sum_{m=1}^M (\langle A(v_k) \rangle - \bar{A}_{S(s),m}(v_k))^2} \quad (9)$$

となる.  $\bar{A}_{S(s),m}(v_k)$  は, ノード  $v_k$  に対して  $s$  回のシミュレーションから得られた影響度  $\bar{A}_{\{1, \dots, s\}}(v_k)$  を表す. 他方, 推定近似誤差  $\bar{D}_k(s)$  に関しては,  $\bar{D}_k(s) = M^{-1} \sum_{m=1}^M \bar{D}_{S(s),m}(v_k; N)$

であり,  $\bar{D}_{S(s),m}(v_k; N)$  は同様にノード  $v_k$  に対して  $s$  回のシミュレーションから得られた推定近似誤差  $\bar{D}_{S(s)}(v_k; N) = \bar{D}_{\{1, \dots, s\}}(v_k; N)$  を表す. ここで, LOO 法の場合は  $N = 1$ , LHO 法の場合は  $N = s/2$  となる. 実際には, 推定近似誤差は真の近似誤差の上限となることが望ましいため, 以下のように定義される評価基準を用いて両手法を評価する.

$$\delta_{D_K}(s) = \frac{1}{K} \sum_{k=1}^K \bar{D}_k(s) - \frac{1}{K} \sum_{k=1}^K D_k(s). \quad (10)$$

明らかに,  $\delta_{D_K}(s)$  の値が 0 に近づくほど, その手法はより良好な性能をもつと言える. 加えて, 真の近似誤差の上限を精度よく与えるためには, この値は小さい正の値を取るべきであることに注意されたい.

## 4.2 結果と考察

図 1, 2 に各ネットワークに対する実験結果を示す. 図 1(a), 2(a) は, 各ネットワークに対する真の近似誤差の平均 ( $K^{-1} \sum_{k=1}^K D_k(s)$ ) を示しており, これらの図からは, 1) 試行回数  $s$  が増えるにつれて  $K$  の値に関わらず近似誤差が小さくなる, 2)  $K = 10^0, 10^1, 10^2$  に対する近似誤差は互いに区別できない程度であるが,  $K = 10^3$  に関しては他の場合とは明確に区別できるほど常に大きな値となっている, という 2 つの傾向が MAE, RMSE に関係なく読み取れる. 前者の傾向は, ノードの推定影響度が  $s$  が大きくなるにつれて真の影響度に収束していることを意味し, 後者は順位の低いノードの推定影響度は順位の高いノードの推定影響度と比べてより大きな分散をもち, 大きな  $K$  に対しては近似誤差を推定することはより難しくなることを意味している. また, これらの図からは, RMSE に基づく近似誤差が常に MAE に基づくものより大きいことがわかる. これは, Jensen

の不等式から理論的に導かれる  $D_S^{(MAE)}(v) \leq D_S^{(RMSE)}(v)$  という関係と一致するものである。

同様に、提案法を用いて推定した近似誤差の平均をそれぞれ示している図 1(b), 2(b) においても類似した傾向が見られる。一方、図 1(a) と 1(b), および図 2(a) と 2(b) を比べると、 $K$  の値に関わらず、いずれのネットワークにおいても LHO 法に比べて LOO 法が近似誤差を過小評価する傾向にあることがわかる。この傾向は、図 1(c), 1(d), 2(c), 2(d) において定量的に示されている。図 1(c), 2(c) は MAE を用いた場合の  $\delta_{D_K}(s)$  を、図 1(d), 2(d) は RMSE を用いた場合の  $\delta_{D_K}(s)$  を示している。図 1(c), 2(c) において、LOO 法は常に負の値を取っており、これは推定された近似誤差が真の近似誤差よりも小さいことを意味している。特に、最初の数百回ほどの試行に関しては、LOO 法による推定近似誤差は真の近似誤差よりもずっと小さな値となっている。対照的に、エンロンネットワークに対する  $K = 10^3$  の場合を除いて、LHO 法による結果は正の値から 0 に漸近しており、最初の数十回の試行後には十分 0 に近い値に収束していることが分かる。

ここで、エンロンネットワークに対する  $K = 10^3$  に関しては、いずれの手法の推定精度も他の場合に比べて悪くなっている点に注意されたい。これは、エンロンネットワークの上位 1,000 ノード近辺には、コスメネットワークと比べて影響度の分散が大きなノードが多く存在しているためである。実際、上位 1,000 番目のノードの真の影響度とその標準偏差は、エンロンネットワークが 5,750.461 と 1,209.592 であるのに対し、コスメネットワークでは 11,697.17 と 624.5759 であった。これが、エンロンネットワークにおける上位 1,000 ノードに対する近似誤差の推定をより難しくしていると言える。

図 1(d), 2(d) においても、上記と同様に、真の近似誤差を LHO 法が LOO 法よりも精度よく推定できていることがわかる。実際のところ、図 1(d), 2(d) では、 $K = 10^0, 10^1, 10^2$  に対して LHO 法は最初の数十回の試行後において、図 1(c), 2(c) における MAE による真の近似誤差との比較結果と比べて、RMSE による真の近似誤差をより正確、かつより安定的に推定する傾向にある。これは、提案法が式 (5) に示すように RMSE に基づいているためであると考えられる。しかし、試行回数が 100 未満の場合、LHO 法でさえも  $K = 10^3$  に対する RMSE による真の近似誤差を過小評価しており、これは、大きな  $K$  に対して本推定問題がより難しくなることを改めて示しているものと言える。一方、前述のように  $D_S^{(RMSE)}(v)$  は  $D_S^{(MAE)}(v)$  よりも常に大きな値をとるため、図 1(d), 2(d) における  $\delta_{D_K}(s)$  の値は、図 1(c), 2(c) における値より、推定法に関わらず相対的に小さくなっている。

以上の結果から、限られたシミュレーション回数の下で計算された影響度の近似誤差を推定するという点においては、LHO 法のほうが LOO 法より正確な推定が可能であると言える。実際のところ、少なくとも本実験で用いたネットワーク、および  $K$  と  $M$  の値に関しては、LHO 法による推定近似誤差は、MAE による真の近似誤差に対する上限となり、RMSE による真の近似誤差に対しては、幾分過小評価気味ではあるもののよい近似となっている。これらの結果は、サンプリングサイズ  $N$  に関して、 $N = 1$  (LOO) は小さ過ぎであり、 $N = |S|/2$  (LHO) がより望ましい値であることを示していると言える。

## 5. おわりに

本稿では、社会ネットワーク上の情報拡散を対象に、各ノードの影響度を効率よく推定することを目的とし、そのために限

られた回数のシミュレーション結果から推定した影響度と真の影響度間の近似誤差を精度よく推定する手法を提案した。提案法は leave- $N$ -out 交差検証の考えに基づき、シミュレーション結果の集合から一定サイズの部分集合をサンプリングすることで、真の影響度を仮定することなく近似誤差を推定することを可能としている。実ネットワークを用いた評価実験の結果は、必ずしも他のネットワークに対して一般化され得るものではないが、提案法のうちサンプリングサイズを全シミュレーション回数の半分とした LHO 法による推定結果が真の近似誤差のよい指標になり得ることを示唆している。提案手法の枠組みは情報拡散モデルに限らず、他の確率モデルを用いた予測シミュレーションに広く適用可能なものである。今後は、多様なネットワークにおける影響度分析、および他の領域における推定問題への適用を通し、本予測シミュレーション法のより詳細な評価をする必要がある。

## 謝辞

本研究は科学研究費補助金若手研究 (B) (No. 23700181) の補助を受けた。

## 参考文献

- [Goyal 11] Goyal, A., Lu, W., and Lakshmanan, L.: Celf++: optimizing the greedy algorithm for influence maximization in social networks, in *Proc. of WWW 2011*, pp. 47–48 (2011)
- [Kempe 03] Kempe, D., Kleinberg, J., and Tardos, E.: Maximizing the spread of influence through a social network, in *Proc. of 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 Transactions on Knowledge Discovery from 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)
- [Kleinberg 08] Kleinberg, J.: The convergence of social and technological networks, *Communications of ACM*, Vol. 51, No. 11, pp. 66–72 (2008)
- [Klimt 04] Klimt, B. and Yang, Y.: The Enron corpus: A new dataset for email classification research, in *Proc. of ECML'04*, pp. 217–226 (2004)
- [Leskovec 07] Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., and Glance, N.: Cost-effective outbreak detection in networks, in *Proc. of KDD 2007*, pp. 420–429 (2007)
- [Saito 09] Saito, K., Kimura, M., Ohara, K., and Motoda, H.: Learning Continuous-Time Information Diffusion Model for Social Behavioral Data Analysis, in *Proc. of ACML 2009*, pp. 322–337 (2009)
- [Saito 13] Saito, K., Kimura, M., ohara, K., and Motoda, H.: Which Targets to Contact First to Maximize Influence over Social Network, in *Proc. of SBP 2013*, pp. 359–367 (2013)