

情報拡散モデルにおける入出次数相関と期待影響度の関係について

Relationship between In/Out-Degree Correlation and Influence Degree in Information Diffusion Models

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

How the information diffuses over a large social network depends on both the model employed to simulate the diffusion and the network structure over which the information diffuses. We analyzed both theoretically and empirically how the two contrasting most fundamental diffusion models, Independent Cascade (IC) and Linear Threshold (LT) behave differently or similarly over different network structures. We devised two rewiring structures, one preserving in/out-degree correlation and the other changing in/out-degree correlation while both preserving their in/out-degree distributions, and analyzed how co-link rate and in/out-degree correlation affect the influence degree of each diffusion model using two real world networks, each as the base network on which rewiring is imposed. The results of the theoretical analysis qualitatively explain the empirical results, and the findings help deepen the understanding of complex diffusion phenomena.

1. はじめに

近年、Facebook や Twitter などのソーシャルメディアの急速な普及により、大規模な社会ネットワークがインターネット上に構築され、情報を拡散させる重要な媒体となっている。そのため、情報拡散という観点から社会ネットワークを分析する研究が数多く報告されている [Newman 02, Richardson 02, Leskovec 06, Watts 07, Bakshy 11, Romero 11]。それらの中で最も多く研究されている問題の 1 つに影響度最大化問題がある。これは、情報を効果的に拡散することができるという意味で影響度の高い一定数のノードの組合せを見つけ出す問題であり、これまでに確率に基づく情報拡散モデルを用いてこの問題を解くアルゴリズムが多数提案されている [Kempe 03, Chen 09, Kimura 10, Chen 10]。実際には、そのような確率モデルを用いて情報拡散現象を分析・調査するには、モデル間の振る舞いの違いを理解することや、解析対象となる特定のタスクに対して最も適切なモデルを選択する為の効果的な方法などが必要となる。

既存研究で用いられる基本的な確率モデルとしては、独立カスケード (IC: Independent Cascade) モデルと線形閾値 (LT: Linear Threshold) モデルという 2 つの対照的なモデルがある [Kempe 03]。IC モデルは情報送信者主導のモデルであり、ネットワーク上の情報送信ノードがある確率で隣接ノードに独立に情報を伝える。これに対して LT モデルは情報受信者主導のモデルであり、情報受信ノードは対象となる情報を受け入れた隣接ノードの数がある閾値以上となったときに限り、その情報を受け取る。なお、LT モデルにおける各ノードの閾値は確率変数として扱われる。本研究では、これら IC モデルと LT モデルが互いにどのように異なった、もしくは類似した振る舞いをするのかを構造が異なる多様な社会ネットワークを対象に分析する。

具体的には、各モデルに基づいて得られるノードの影響度をネットワーク構造という観点から比較する。ここで、確率に基づくある情報拡散モデルの下でのノード v のもつ影響度は、

v を起点とした情報拡散過程が終了した時点でアクティブな状態になっているノード数の期待値として定義される。ノードがアクティブであるとは、そのノードが拡散対象となる情報の影響を受けた状態のことを指し、そうでないノードは非アクティブであるとする。本稿では、はじめに IC モデルと LT モデルのスケールフリーネットワーク上での特性を理論的に分析し、1) ノードの入出次数相関が減少するにつれ IC モデルの影響度も減少するが LT モデルの影響度は変化しないこと、2) 双方向リンクの割合が減少するにつれ、両モデルとも影響度が増加するが、IC モデルのほうがその増加率ははるかに低い、という 2 つの性質を示す。そして、これらの性質を実験的に検証する。そのために、基準となるスケールフリーネットワークに 2 つのリンク張り替え手法を適用し、入出次数相関と双方向リンク率が異なる一連のネットワークを系統的に生成し、それらの上での各モデルによる影響度の変化を調べる。いずれのリンク張り替え手法も入出次数分布は保存するが、一方は入出次数相関を保存したまま双方向リンク率だけを減少させ、もう一方は入出次数相関と双方向リンク率をともに減少させる。本稿では、2 つの実スケールフリーネットワークを基準とした実験的検証の結果を示し、前述の 2 つの性質が実際に成立することを示す。

2. 情報拡散モデル

本節では、IC モデルと LT モデルの形式的な定義を述べる [Kempe 03, Kimura 10]。以下、 V を全ノード集合、 $E (c V \times V)$ を全リンク集合とする有向ネットワーク $G = (V, E)$ を用いて社会ネットワークを表現するものとする。また、両モデルとも、情報拡散過程は初期アクティブノードを起点に離散時間 $t \geq 0$ で進行し、ノードの状態は非アクティブからアクティブに変化するが、その逆には変化しないものとする。

まず IC モデルについて述べる。IC モデルでは、各リンク (u, v) はパラメータとして拡散確率 $p_{u,v}$ ($0 < p_{u,v} < 1$) をもつ。そして、ノード u が時刻 t にてアクティブになったとすると、 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 はその試行が成功するかどうかに関わらず、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。

一方、LT モデルでは各リンク (u, v) はパラメータとして重み $q_{u,v}$ (> 0) をもち、その重みは $\sum_{u \in B(v)} q_{u,v} \leq 1$ という関係を満たす。ここで、 $B(v) = \{u \in V; (u, v) \in E\}$ はノード v の親ノードの集合を表す。LT モデルでは、まずすべてのノード $v \in V$ に対して、区間 $[0, 1]$ から一様ランダムに閾値 θ_v を選択し、割り当てる。そして、時刻 t で非アクティブであるノード v は、その時点でアクティブである親ノードとの間のリンクのもつ重みの総和が閾値 θ_v 以上となった場合、すなわち $\sum_{u \in B(v)} q_{u,v} \geq \theta_v$ が満たされた場合に、親ノードの影響を受け、時刻 $t+1$ にアクティブとなる。ここで、 $B_v(v)$ は v の親ノードのうち時刻 t の時点でアクティブであるものの集合を表す。この情報拡散過程は、いずれの非アクティブノードもそれ以上アクティブになることができなくなった時点で終了する。

3. 局所影響度の分析

ここでは、IC モデルと LT モデルの影響度を理論的に分析する。まず、ノード u が直接アクティブにする子ノード数の期待値である u の局所影響度 $\sigma_L(u)$ について考える。IC モデルにおける局所影響度 $\sigma_L^{IC}(u)$ は、 $\sigma_L^{IC}(u) = \sum_{v \in F(u)} p_{u,v}$ により与えられる。ここで、 $F(u)$ は u の子ノードの集合 $\{v \in V; (u, v) \in E\}$ を表す。一方、LT モデルに関しては、 u から出る各リンクに対する重み $q_{u,v}$ を閾値 θ_v が区間 $[0, q_{u,v}]$ から選ばれる確率とみなすことができることから、その局所影響度 $\sigma_L^{LT}(u)$ は、 $\sigma_L^{LT}(u) = \sum_{v \in F(u)} q_{u,v}$ となる。次に、上記の局所影響度に対して、全ノードに対する平均局所影響度 $\bar{\sigma}_L(G)$ を計算する。LT モデルに関しては、すべてのノード $v \in V$ に対して $\sum_{u \in B(v)} q_{u,v} = 1$ という条件を仮定した場合、以下のように $\bar{\sigma}_L^{LT}(G) = 1$ となる。

$$\begin{aligned} \bar{\sigma}_L^{LT}(G) &= \frac{1}{|V|} \sum_{u \in V} \sigma_L^{LT}(u) = \frac{1}{|V|} \sum_{u \in V} \sum_{v \in F(u)} q_{u,v} = \frac{1}{|V|} \sum_{(u,v) \in E} q_{u,v} \\ &= \frac{1}{|V|} \sum_{v \in V} \sum_{u \in B(v)} q_{u,v} = 1 \end{aligned}$$

IC モデルに関しては、文献 [Kempe 03] などの多くの既存研究と同様に、すべてのリンク $(u, v) \in E$ が一様な拡散確率 $p_{u,v} = p$ をもつと仮定するならば、その平均局所影響度 $\bar{\sigma}_L^{IC}(G)$ を次のように計算することができる。

$$\bar{\sigma}_L^{IC}(G) = \frac{1}{|V|} \sum_{u \in V} \sigma_L^{IC}(u) = \frac{1}{|V|} \sum_{u \in V} \sum_{v \in F(u)} p_{u,v} = \frac{1}{|V|} \sum_{(u,v) \in E} p = \frac{|E|}{|V|} p$$

ここで、 $\frac{|E|}{|V|}$ は平均次数 $d = \frac{1}{|V|} \sum_{u \in V} |B(u)| = \frac{1}{|V|} \sum_{u \in V} |F(u)| = \frac{|E|}{|V|}$ となる。したがって、前述の様な拡散確率 p を平均次数の逆数 $\frac{1}{d} = \frac{|V|}{|E|}$ とすることにより、 $\bar{\sigma}_L^{IC}(G) = 1$ を導くことができる。この結果から、IC モデルと LT モデルは平均局所影響度という観点からは等価だといえる。以下では、この結果を踏まえ前述の仮定のもとで影響度を評価するものとする。このとき、IC モデルにおけるノード u の局所影響度は $\sigma_L^{IC}(u) = \sum_{v \in F(u)} p_{u,v} = p|F(u)|$ となることに注意されたい。

ここまでは、ノード $u \in V$ がすでにアクティブになっているという条件下でその局所影響度について議論してきたが、情報拡散が実際には連鎖的なものであることを考慮すると、ノード u がその親ノードによってアクティブにされる確率 $r(u)$ を考える必要

がある。そのため、次に連鎖局所影響度 $\sigma_{CL}(u) = r(u)\sigma_L(u)$ を考える。ここでは、最も単純な場合として、確率 $r(u)$ をノード u がノード集合 V から一様ランダムに選択されたあるノードによって次の時刻にアクティブにされる確率と考える。IC モデルに関しては、その確率 $r^{IC}(u)$ は $r^{IC}(u) = \frac{1}{|V|} \sum_{s \in B(u)} P_{s,u} = \frac{|B(u)|}{|V|}$ となり、LT モデルに関しては、その確率 $r^{LT}(u)$ は $r^{LT}(u) = \frac{1}{|V|} \sum_{s \in B(u)} Q_{s,u} = \frac{1}{|V|}$ となる。したがって、IC モデル、LT モデルそれぞれに対する平均連鎖局所影響度 $\bar{\sigma}_{CL}$ は次のように求められる。平均局所影響度とは違い両者はもはや等価ではない。

$$\bar{\sigma}_{CL}^{IC}(G) = \frac{1}{|V|} \sum_{u \in V} r^{IC}(u)\sigma_L^{IC}(u) = \frac{p^2}{|V|^2} \sum_{u \in V} |B(u)||F(u)| \quad (1)$$

$$\bar{\sigma}_{CL}^{LT}(G) = \frac{1}{|V|} \sum_{u \in V} r^{LT}(u)\sigma_L^{LT}(u) = \frac{1}{|V|^2} \sum_{u \in V} \sigma_L^{LT}(u) = \frac{1}{|V|} \quad (2)$$

一方、ネットワーク G の入出次数相関係数は次式のように定義できる。

$$dc_{I/O}(G) = \frac{\frac{1}{|V|} \sum_{u \in V} |B(u)||F(u)| - d^2}{\sqrt{\frac{1}{|V|} \sum_{u \in V} |B(u)|^2 - d^2} \sqrt{\frac{1}{|V|} \sum_{u \in V} |F(u)|^2 - d^2}} \quad (3)$$

ここで、式 (3) の分母は入出次数分布の標準偏差によって定まることに注意されたい。この事実と、式 (1) から、入出次数分布の標準偏差が一定である場合、IC モデルに対する平均連鎖局所影響度は入出次数相関 $dc_{I/O}(G)$ に依存することがわかる。すなわち、入出次数相関が低くなれば、IC モデルの影響度も同様に減少することが予想できる。これに対して、LT モデルに関しては、式 (2) から、その平均連鎖局所影響度が入出次数相関 $dc_{I/O}(G)$ の影響を受けないことが分かる。

次に、影響度を変化させる重要な要因として、ネットワーク G における双方向リンク率 $cr(G)$ について考える。ここで、 $cr(G) = \frac{1}{|V|} \sum_{u \in V} |B(u) \cap F(u)|$ と定義できる。明らかに、双方向ネットワーク G に関しては、任意の $u \in V$ に関して $B(u) = F(u)$ であることから $cr(G) = 1$ となる。いま、ノード v について、 $v \in B(u) \cap F(u)$ を仮定すると、 v がノード u をアクティブにすることに成功した場合、それ以降、逆向きのリンク (u, v) によってアクティブなノードが増えることはなく、逆に u が v をアクティブにすることに成功した場合、やはりリンク (v, u) はそれ以降、アクティブなノードを増やすことに貢献できなくなる。このことから、IC モデル、LT モデルともに双方向リンク率 $cr(G)$ が減少すれば、影響度が増加することが推測できる。

しかしながら、実際には $cr(G)$ の減少により影響度がどの程度増加するかはそれぞれのモデルで大きく異なり得る。この点を検討する為に、双方向リンク率が 1 に近いネットワークにおいて、ある大きな次数 $D = |F(v)| = |B(v)|$ をもつノード v の k 個の親ノードがアクティブになった場合を考える。このとき、このネットワークの入出次数相関は同様に 1 に近い値となることは明らかである。そして、 v がアクティブとなる確率は、IC モデルに関しては平均次数を d とすると $1 - (1 - 1/d)^k$ 、LT モデルに関しては k/D となる。IC モデルに関しては、その確率は D には依存せず、 k の値が小さくてもある程度以上であれば大きな値となる。したがって、 v がアクティブになったことによって無駄になる逆向きリンクの数 k はそれほど大きな値にならないことが予想できる。これに対して、LT モデルに関しては、 D が大きな値の場合、 k も同様に大きな値を取らなければ v がアクティブとならず、その結果、 v がアクティブになったときには無駄になる逆向きリンクが相当数生じることが予想できる。このことから、双方向リンク率の変化による IC モデルの影響度の変化は、LT モデルの影響度の変化よりもずっと小さくなると考えられる。

4. 実験的検証

前述の理論解析に基づく IC・LT モデルそれぞれの影響度に関する性質を実験的に確認した。実験では、複数の人工ネットワーク、実ネットワークに対して後述する 2 つのリンク張り替え手法を適用したものをを用いたが、ここでは紙面の都合上、それらのうち 2 つの実ネットワークに対する結果のみを示す*1。

4.1 リンク張り替え手法

本実験で用いた 2 つのリンク張り替え手法について説明する。いずれも入出次数分布を保存する為、元のネットワークがスケールフリーである場合、リンク張り替え後のネットワークもまたスケールフリーとなる。1 つ目の手法は、文献 [Melo 02] 中のランダムネットワーク生成法と等価なものであり、与えられたネットワーク G の各ノードの入出次数を保存したままリンクを張り替える。本実験では、一様ランダムに選択した 2 本のリンク $e = (u, v)$, $e' = (u', v')$ の終点ノード v と v' をリンク張り替え確率 α で入れ替えることで、この手法を実現した。以下、この手法を DCP (in/out-Degree Correlation Preserved) 法と呼び、DCP 法により得られたネットワークを $dcp_\alpha(G)$ と表記する。ここで、 α が大きいほど $cr(G)$ は小さくなる。したがって、DCP 法を用いることで、双方向リンク率が IC・LT モデルの影響度にどのように影響を与えるかを調べることができる。

2 つ目のリンク張り替え手法は入出次数相関 $dc_{I/O}(G)$ も変化させるものであり、 $dc_{I/O}(G)$ の変化により IC モデルの影響度が変化することを確認することができる。本実験では、一様ランダムに選択した 2 つのノード v と v' に対して、 v の入リンク集合 $E_I(v)$ と v' の入リンク集合 $E_I(v')$ を確率 α で入れ替えることによって、この手法を実現した。この手法は入リンク集合単位でリンクを張り替えるため、入出次数分布は変化しないことに注意されたい。以下、この手法を DCU (in/out-Degree Correlation Unpreserved) 法と呼び、DCU 法により得られたネットワークを $dcu_\alpha(G)$ と表記する。ここで、リンク張り替え確率 α が大きくなるほど、 $dc_{I/O}(G)$ は小さくなる。

4.2 データセットとネットワーク構造

本稿で実験結果を示す実ネットワークのうち 1 つは、エンロン E メールデータセット [Klimt 04] に対して、双方向に送受信記録がある E メールアドレスに対応するノード間に双方向リンクを張り、最大強連結成分を抽出したノード数 4,254、リンク数 44,314 の双方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。もう 1 つは、化粧品に関する口コミサイトである “@cosme” *2 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 のネットワークを用いた。以下、コスメネットワークと呼ぶ。

上記のネットワークに対して、本実験では $dcp_\alpha(G)$, および $dcu_\alpha(G)$ 中の各ノード v に対する IC モデル、および LT モデルの下での影響度 $\sigma(v)$ を、 α を 0.0 から 1.0 まで 0.1 刻みで変化させて調べた。ここで、 $dcp_{0.0}(G) = dcu_{0.0}(G) = G$ であることに注意されたい。影響度 $\sigma(v)$ に関しては、ボンダパーコレーション法 [Kimura 10] による独立した 10,000 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求めた。また、3. 章での議論に従い、IC モデルの拡散確率 p_{uv} に関しては平均次数 d に基づいたような値

$p = 1/d$ を設定した。具体的には、エンロンネットワークでは $p = 0.10$ 、コスメネットワークでは $p = 0.13$ を設定した。同様に、LT モデルに関しては、すべてのノード $v \in V$ に対して $\sum_{u \in B(v)} q_{u,v} = 1$ となるようにリンクの重みを設定した。

4.3 実験結果

図 1(a), 2(a) にエンロンネットワーク、コスメネットワークそれぞれに DCP 法、および DCU 法を適用した場合の入出次数相関 $dc_{I/O}(G)$ と双方向リンク率 $cr(G)$ の変化を示す。この結果から、両手法ともに意図した通りに機能していることが分かる。具体的には、リンク張り替え確率 α が大きくなるにつれて $cr(G)$ はいずれの手法でもほぼ同様に減少している。これに対し $dc_{I/O}(G)$ は DCP 法では変化していないが、DCU 法では $cr(G)$ と同様に減少している。

次に、エンロンネットワークに関して、DCP 法によるリンク張り替えがノードの最大・平均影響度をどのように変化させるかを図 1(b) に示す。予想通り、LT モデルに関してはリンク張り替え確率が大きくなり、双方向リンク率 $cr(G)$ が小さくなるにつれて、最大影響度、平均影響度ともに増加している。IC モデルに関しては、いずれも増加していないように見えるが、実際には $cr(G)$ が急激に減少する α の値が 0.0 から 0.6 の範囲で若干の増加が確認された。この結果は、3. 章における理論解析の結果を定量的に裏付けるものである。なお、図 2(b) に示すように、コスメネットワークに関しても同様の傾向が確認でき、本稿では割愛した他のネットワークに対する結果においても同様の傾向を確認している。

次に、図 1(c), 2(c) に、DCU 法によるリンク張り替えが IC・LT モデルの最大・平均影響度をどのように変化させるかを示す。DCU 法では α が増加するに従って、入出次数相関 $dc_{I/O}(G)$ と双方向リンク率 $cr(G)$ の両方の値が減少する。これは、IC モデルに関して、その影響度を変化させる 2 つの競合する要因が存在することを意味するが、実験結果からは $dc_{I/O}(G)$ の効果の方が強く、いずれのネットワークにおいても IC モデルにおける影響度が減少していることがわかる。この結果も、 $cr(G)$ の変化により IC モデルが受ける影響が小さいことを定量的に示していると言える。他方、LT モデルにおける影響度は図 1(b), 2(b) と同様に増加しており、このことから LT モデルにおける影響度は $cr(G)$ のみに依存していることが分かる。同様の傾向は他のネットワークでも観測されており、これらの結果もまた、3. 章の理論解析結果を定量的に裏付けるものであると言える。

5. まとめ

本研究では、社会ネットワーク上の情報拡散現象に対する IC モデル、および LT モデルという 2 つの代表的な確率モデルを対象に、ネットワーク構造の変化によるその振る舞いの変化を理論的、実験的に分析した。具体的には、1) IC モデルは入出次数相関に敏感に反応し、その影響度は入出次数相関と正の相関をもつ一方、LT モデルにおける影響度は入出次数相関の影響を受けない、および 2) IC モデル、LT モデルいずれの下でも影響度は双方向リンク率とは負の相関をもち、IC モデルが受ける影響のほうがかはるかに小さい、という 2 つの性質を理論解析を通して定性的に導いた。そして、入出次数相関を保存したまま双方向リンク率を変化させる DCP 法、および入出次数相関と双方向リンク率の両方を変化させる DCU 法という 2 つのリンク張り替え手法を実際の社会ネットワークに適用した実験を通して、これらの性質が実際に成立することを検証した。これらの知見は、今後、社会ネットワーク上の複雑な情報拡散現象をより深く理解するために有用であると思われる。

*1 実際には、BA モデル [Barabási 99]、CNN モデル [Vázquez 03] によって生成した人工ネットワーク、および実データから抽出した 4 つの大規模ネットワークについても同様の実験を行った。

*2 <http://www.cosme.net/>

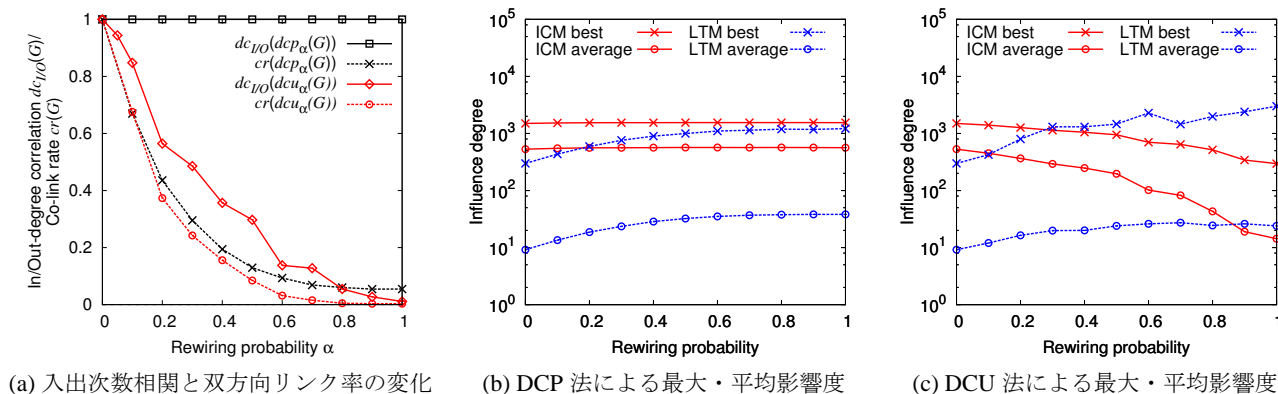


図 1: エンロンネットワークに関する実験結果

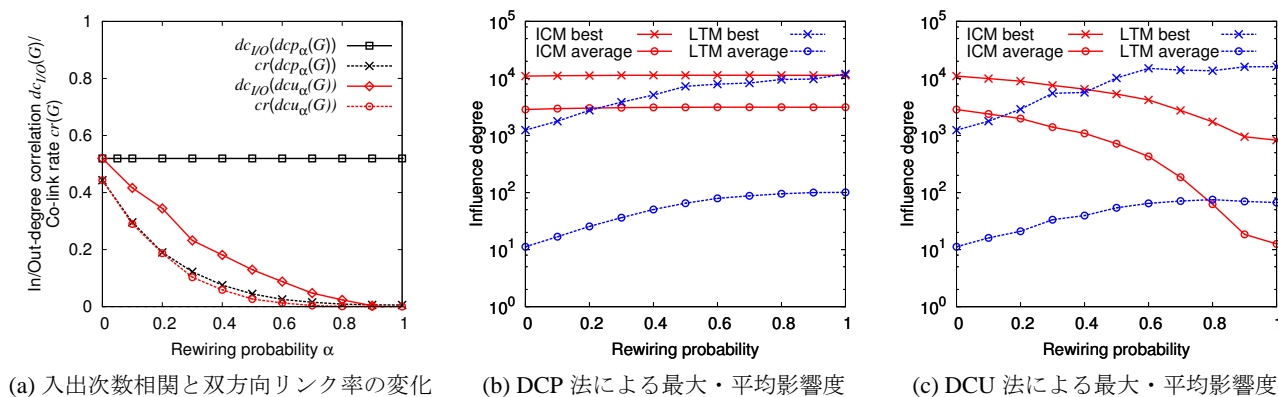


図 2: コスメネットワークに関する実験結果

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

- [Bakshy 11] Bakshy, E., Hofman, J., Mason, W., and Watts, D.: Everyone’s an Influencer: Quantifying Influences on Twitter, in *Proc. of WSDM2011*, pp. 65–74 (2011)
- [Barabási 99] Barabási, A.-L. and Albert, R.: Emergence of scaling in random networks, *Science*, Vol. 286, pp. 509–512 (1999)
- [Chen 09] Chen, W., Wang, Y., and Yang, S.: Efficient influence maximization in social networks, in *Proc. of KDD 2009*, pp. 199–208 (2009)
- [Chen 10] Chen, W., Wang, C., and Wang, Y.: Scalable Influence Maximization for Prevalent Viral Marketing in Large-scale Social Networks, in *Proc. of KDD 2010*, pp. 1029–1038 (2010)
- [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 10] Kimura, M., Saito, K., Nakano, R., and Motoda, H.: Extracting influential nodes on a social network for information diffusion, *Data Mining and Knowledge Discovery*, Vol. 20, pp. 70–97 (2010)
- [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 06] Leskovec, J., Adamic, L. A., and Huberman, B. A.: The dynamics of viral marketing, in *Proc. of EC’06*, pp. 228–237 (2006)
- [Melo 02] Melo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U.: Network Motifs: Simple Building Blocks of Complex Networks, *Science*, Vol. 298, pp. 824–827 (2002)
- [Newman 02] Newman, M. E. J., Forrest, S., and Balthrop, J.: Email networks and the spread of computer viruses, *Physical Review E*, Vol. 66, p. 035101 (2002)
- [Richardson 02] Richardson, M. and Domingos, P.: Mining knowledge-sharing sites for viral marketing, in *Proc. of KDD-2002*, pp. 61–70 (2002)
- [Romero 11] Romero, D., Meeder, B., and Kleinberg, J.: Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter, in *Proc. of WWW2011*, pp. 695–704 (2011)
- [Vázquez 03] Vázquez, A.: Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations, *Physical Review*, Vol. 67, No. 5, p. 056104 (2003)
- [Watts 07] Watts, D. J. and Dodds, P. S.: Influence, networks, and public opinion formation, *Journal of Consumer Research*, Vol. 34, pp. 441–458 (2007)