

# 多重有向ネットワーク成長モデルの特性解析

## Characteristic Analysis of Multiple Directed Network Growth Models

加藤翔子 \*<sup>1</sup>

Shoko Kato

斉藤和巳 \*<sup>1</sup>

Kazumi Saito

\*<sup>1</sup>静岡県立大学大学院

University of Shizuoka

For constructing network growth model which have high multiplicity directed links, we extended Barabasi-Albert model and Connecting Nearest Neighbor model for multiple directed graphs, and proposed link multiplicity preferential attachment in our study. In this paper, we compare the expected influence of information diffusion of these models and real networks, using Independent Cascade model and Linear Threshold model.

### 1. はじめに

複雑ネットワークの成長モデルには、スモールワールドモデル [1] やバラバシ=アルバートモデル [2] など、多くのモデルが提案されているが、その多くが単純無向グラフを生成するモデルである。また、多重有向グラフへ向けたアプローチとしては、クラスタ係数の計算手法の提案 [3] などが先行研究として挙げられるが、グラフ自体を生成するモデリングに関する研究は見られない。従って、現実世界のネットワークの構造を表すことが可能な、リンクの向きや多重度を考慮したモデルの提案が待たれる。

高い多重度を持つ有向リンクを生成するネットワーク成長モデルの構築に向け、先行論文 [4, 5] では、BA モデルと CNN モデルを多重有向に拡張し、リンク多重度優先選択確率と呼ぶリンク生成確率を新たに付加したモデルを提案した。本稿では、これらのモデルと現実世界のネットワークについて、情報拡散した場合の期待影響度を、独立カスケード (IC:Independent Cascade) モデル [6]、線形閾値 (LT:Linear Threshold) モデル [7] の 2 つを用いて比較する。

本論文の構成は以下となる。まず、本研究に用いた情報拡散モデルとネットワーク成長モデルについて説明する。次に、評価に用いたデータについて説明するとともに、実験による結果を報告する。最後に、本論文のまとめについて述べる。

### 2. 情報拡散モデル

ノード集合  $V$  とリンク集合  $E(\subset V \times V)$  からなる多重有向グラフを  $G = (V, E)$  とする。ノードが情報を保持している状態をアクティブ、そうでない状態を非アクティブと呼ぶ。ノードは非アクティブからアクティブへ一度だけ状態が変わることができ、アクティブから非アクティブへ状態が変わることはないものとする。今、初期値としてアクティブノード集合  $S$  が与えられたとし、その他のノードはすべて非アクティブとする。 $(u, v) \in E$  のとき、ノード  $u$  はノード  $v$  の親ノードと呼び、ノード  $v$  はノード  $u$  の子ノードと呼ぶ。各ノード  $v \in V$  に対して、 $v$  の子ノード集合  $C(v)$  と親ノード集合  $B(v)$  を以下のように定義する。

$$C(v) = \{w \in V; (v, w) \in E\}, B(v) = \{u \in V; (u, v) \in E\}$$

#### 2.1 独立カスケード (IC) モデル

IC モデル [6] は感染症の広がり方を示す基本的な確率モデルである。このモデルでは、各リンク  $(u, v)$  に対して前もって実数値  $\kappa_{u,v}$  ( $0 < \kappa_{u,v} < 1$ ) を割り当てる。ここで、 $\kappa_{u,v}$  をリンク  $(u, v)$  における拡散確率と呼ぶ。IC モデルでの拡散過程は離散時間  $t \geq 0$  で展開され、情報源ノードから以下の方法によって広がっていく。ノード  $u$  が時刻  $t$  でアクティブになったとき、ノード  $u$  には現在非アクティブの子ノード  $v$  に対して一度だけアクティブにさせるチャンスが与えられ、それは拡散確率  $\kappa_{u,v}$  で成功する。成功したら、ノード  $v$  は時刻  $t+1$  でアクティブになる。もし、ノード  $v$  の複数の親ノードが時刻  $t$  でアクティブになった場合には、任意の順番で拡散試行が行われるとする。この拡散過程は、新たにアクティブになるノードがなくなったときに終了する。

#### 2.2 線形閾値 (LT) モデル

LT モデル [7] はイノベーションの広がり方を表す基本的な確率モデルである。このモデルでは、すべてのノード  $v \in V$  に対して、 $\sum_{u \in B(v)} \omega_{u,v} \leq 1$  となるように前もってリンク  $(u, v)$  に重み  $\omega_{u,v}$  ( $> 0$ ) を割り当てる。LT モデルでの拡散過程は、初期アクティブ集合  $S$  が与えられた上でランダムルールに従って行われる。まず、全てのノード  $v \in V$  に対して、閾値  $\theta_v$  が区間  $[0, 1]$  から一様ランダムに選ばれる。時刻  $t$  で非アクティブノード  $v$  は各親ノード  $u \in B(v)$  から  $\omega_{u,v}$  の影響を受ける。もし、ノード  $v$  のアクティブな親ノードから受けた重みの合計が  $\theta_v$  以上になった場合、ノード  $v$  は時刻  $t+1$  でアクティブになる。この拡散過程は、新たにアクティブになるノードがなくなったときに終了する。

#### 2.3 情報拡散期待影響度

ネットワーク  $G = (V, E)$  において、IC モデルの拡散確率ベクトルを  $\Theta = (p_{v,w})_{(v,w) \in E}$  とする。初期アクティブノード集合  $S$  に対して、IC モデルの拡散過程終了後のアクティブノード数を  $\varphi(S; \Theta)$  とする。IC モデルの拡散過程は確率過程であるので、 $\varphi(S; \Theta)$  は確率変数となる。よって、 $\varphi(S; \Theta)$  の期待値を  $\sigma(S; \Theta)$  とし、 $\sigma(S; \Theta)$  を、拡散確率ベクトル  $\Theta$  の IC モデルにおけるノード集合  $S$  の“影響度”と定義する。特に、 $S = \{v\}$  のとき、 $\sigma(S; \Theta)$  を  $\sigma(v; \Theta)$  と表記し、 $\sigma(v; \Theta)$  を拡散確率ベクトル  $\Theta$  の IC モデルにおけるノード  $v$  の“影

連絡先: 加藤翔子, 静岡県立大学大学院経営情報イノベーション研究科, 〒422-8526 静岡県静岡市駿河区谷田 52-1, 054-264-5008, j13103@u-shizuoka-ken.ac.jp

響度”と呼ぶ。ここに、影響度の高いノードは、IC モデルに基づく情報拡散において有力なノードである。

LT モデルについても、割り当てられる重みは  $0 < \omega_{u,v} < 1$  であるので、拡散確率の一種と解釈することができる。そのため、期待影響度についても IC モデルと同様に算出する。

### 3. ネットワーク成長モデル

多重グラフとは、同一ノードペア間に複数回リンクが生成されることを許容するグラフであり、多重有向グラフ  $G = (V, E)$  において、 $E$  は多重集合として定式化される。ノード  $v \in V$  が始点となるリンクの本数をノード出次数  $d_v^+$  とし、 $v$  が終点となるリンクの本数をノード入次数  $d_v^-$  と呼ぶ。また、リンク  $e = (u, v) \in E$  の多重度をリンク次数と呼び  $d_e$  と表記する。

また、3.2 において、ノード  $v$  の隣接ノード集合を  $V_{c_v}$  と表記する。連結したノード  $u$  と  $v$  があるとき、 $v$  と  $x \in V_{c_u}$  の間には潜在リンク  $f = (v, x)$  が存在し、 $v$  を一端とする潜在リンク集合を  $F_v$  と定義する。ここで、このモデルは多重を許すため、実リンク  $e = (v, x)$  と潜在リンク  $f = (v, x)$  は同時に存在して良いとする。

#### 3.1 拡張 BA 法

本稿では、BA モデル [2] で提案されたのと同様に、確率  $P(v) \propto d_v^+$  や  $P(v) \propto d_v^-$  でリンクの始点や終点となるノードを選ぶノード次数優先選択を考える。そして、リンク多重度優先選択と呼ぶ、確率  $P(e) \propto d_e$  でリンク  $e$  の多重度を一つ増やすメカニズムを新たに提案し導入する。

以下では、ノード次数優先選択とリンク多重度優先選択を混合させたネットワーク生成モデルを提案する。まず、新ノード生成確率を  $\alpha$  とする。ここで、ノード数  $M$  とリンク数  $N$  程度のネットワーク生成では、 $\alpha = M/N$  に設定する。ただし、この新ノードが始点となるか終点となるかは  $1/2$  の確率で決まるとする。次に、新ノードは生成されずリンクのみ追加されるケースにおいて、ノード次数優先選択とリンク多重度優先選択のどちらを選ぶかの確率を  $\beta$  とする。リンク数  $N$  のネットワーク生成アルゴリズムは以下となる。

- S1.  $V = \{u, v\}$ ,  $V = \{(u, v)\}$ ,  $G = (V, E)$  で初期ネットワークを設定する。
- S2. 確率  $\alpha$  で新ノード生成 S3., さもないければ、リンク追加 S4. のいずれかを  $(N - 1)$  回繰り返す。
- S3.  $V$  に新ノード  $w$  を加え、優先選択確率  $P(v)$  で  $v$  を選び始点か終点かで  $E$  に  $(w, v)$  か  $(v, w)$  を加える。
- S4. 確率  $\beta$  でリンク多重度優先選択でリンク追加 S5., さもないければ、ノード次数優先選択リンク追加 S6. を実行する。
- S5. リンク多重度優先選択確率  $P(e)$  でリンク  $e$  を選び  $E$  に  $e$  を加え、 $e$  の多重度を一つ増やす。
- S6. ノード次数優先選択確率  $P(v)$  で始点  $u$  と終点  $v$  を選び  $E$  に  $(u, v)$  を加える。

#### 3.2 拡張 CNN 法

CNN モデル [8] で提案されたのと同様に、ノード  $v$  の潜在リンク集合  $F_v$  から潜在リンク  $f = (v, x) \in F_v$  をランダムに一つ選んで実リンクにする潜在リンク選択を考える。そして、3.1 で提案したリンク優先選択のパラメータをこちらにも導入する。

以下では、潜在リンク選択とリンク優先選択を混合させたネットワーク生成モデルを提案する。3.1 同様、新ノード生成確率を  $\alpha$  とし、ノード数  $M$  とリンク数  $N$  程度のネットワーク生成では、 $\alpha = M/N$  に設定する。この新ノードが始点となるか終点となるかは  $1/2$  の確率で決まる。次に、新ノードは生成されずリンクのみ追加されるケースにおいて、潜在リンク選択とリンク優先選択のどちらを選ぶかの確率を  $\gamma$  とする。リンク数  $N$  のネットワーク生成アルゴリズムは以下となる。

- S1.  $V = \{u, v, x\}$ ,  $E = \{e = (u, v)\}$ ,  $G = (V, E)$  で初期ネットワークを設定する。
- S2. 確率  $\alpha$  で新ノード生成 S3., さもないければ、リンク追加 S4. のいずれかを  $(N - 1)$  回繰り返す。
- S3.  $V$  に新ノード  $w$  を加え、ノード優先選択確率  $P(v)$  で  $v$  を選び始点か終点かで  $E$  に  $(w, v)$  か  $(v, w)$  を加える。
- S4. 確率  $\gamma$  でリンク優先選択でリンク追加 S5., さもないければ、潜在リンク選択でリンク追加 S6. を実行する。
- S5. リンク優先選択確率  $P(e)$  でリンク  $e$  を選び  $E$  に  $e$  を加え、 $e$  の多重度を一つ増やす。
- S6. ノード優先選択確率  $P(v)$  で  $v$  を選び、始点となるか終点となるかを  $1/2$  の確率で決める。次に、 $F_v$  から  $f = (v, x) \in F_v$  をランダムに一つ選んで実リンクにし、 $E$  に  $(v, x)$  を加える。

### 4. 評価実験

評価に用いたデータと実験結果を述べる。

#### 4.1 データ概要

本稿で取り扱うデータは、Twitter の Favorite, Retweet の 2 機能を多重有向グラフ化したデータである。いずれもユーザーをノード、「Favorite する・される」「Retweet する・される」の関係をリンクとしてネットワークを構築した。

Favorite ネットワークのデータ取得期間は 2011 年 5 月 1 日から 2012 年 2 月 12 日までであり、189,717 ノードと 33,456,690 多重リンクをもつ有向ネットワークである。このデータは Favotter<sup>\*1</sup> の「今日の人気」ページから取得したものである。

Retweet ネットワークのデータ取得期間は 2011 年 3 月 7 日 0 時から同月 11 日 12 時までであり、776,572 ノードと 2,181,151 多重リンクをもつ有向ネットワークである。このデータは鳥海ら [9] のデータを加工したものであり、転載元の発言にコメントをつけるなどして自分の発言として投稿する「非公式リツイート」は除外している。

#### 4.2 実験結果

Favorite ネットワークや Retweet ネットワークと同規模の人工ネットワークを、3. にて説明した拡張 BA 法、拡張 CNN 法を用いて生成し、入次数や出次数と平均期待影響度について比較した結果を以下に示す。

Favotter ネットワークとの比較において、現実ネットワークを赤の丸、拡張 BA 法のうち  $\beta = 0.0$  とした結果を青の四角形、 $\beta = 0.7$  とした結果を緑の三角形、拡張 CNN 法のうち  $\gamma = 0.0$  とした結果を黄色のバツ印、 $\gamma = 0.2$  とした結果をマゼンダ色の菱形でそれぞれプロットした。

\*1 <http://favotter.net/>

また、Retweet ネットワークとの比較においては、現実ネットワークを赤の丸、拡張 BA 法のうち  $\beta = 0.0$  とした結果を青の四角形、 $\beta = 0.6$  とした結果を緑の三角形、拡張 CNN 法のうち  $\gamma = 0.0$  とした結果を黄色のバツ印、 $\gamma = 0.2$  とした結果をマゼンダ色の菱形でそれぞれプロットした。これらはいずれも入次数分布が現実ネットワークと最も近似するパラメーターである。

図 1, 2, 3, 4 では、2. にて説明した IC モデルと LT モデルに基づく期待影響度を、同じ入次数を持つノード群で平均した結果を比較している。横軸はノードの入次数、縦軸は平均期待影響度を表す。

図 1, 2 より、IC モデルにおいて、入次数との相関を見た場合、 $\beta = 0.0$  とした拡張 BA モデルが現実ネットワークと最も近似することが見て取れる。

また、図 3, 4 より、LT モデルにおいては、いずれの拡張成長モデルについても現実ネットワークと近似した期待影響度を持つことがわかる。

図 5, 6, 7, 8 では、2. にて説明した IC モデルと LT モデルに基づく期待影響度を、同じ出次数を持つノード群で平均した結果を比較している。横軸はノードの出次数、縦軸は平均期待影響度を表す。

図 5, 6 より、IC モデルにおいて、拡張 BA モデルよりも拡張 CNN モデルの方が平均期待影響度が低く、現実ネットワークのものと近似することがわかる。

また、図 7, 8 より、LT モデルにおいても、IC モデルほど顕著であるとは言えないが、拡張 CNN モデルの方が拡張 BA モデルより平均期待影響度が低く、現実ネットワークのものと近似することがわかる。

また、いずれのプロットも、IC モデルにおいては、ある一定の入次数や出次数を超えると期待影響度は一定となるが、LT モデルにおいては、IC モデルほど期待影響度が安定せず、バラつきがあることがわかる。

## 5. まとめ

高い多重度を持つ有向リンクを生成するネットワーク成長モデルの構築に向け、拡張 BA モデルと拡張 CNN モデルについて、情報拡散した場合の期待影響度を、IC モデルと LT モデルを用いて分析した。

その結果、入次数との相関では、IC モデルについては  $\beta = 0.0$  とした拡張 BA 法が最も再現性が高く、LT モデルについてはいずれの成長モデルも現実ネットワークと近似した期待影響度を持つことを確認し、出次数との相関では、IC モデル、LT モデルいずれについても、CNN 法を拡張した成長モデルが現実ネットワークと近似した期待影響度を持つことを確認した。また、IC モデルの期待影響度はある次数を超えると一定となり、LT モデルの期待影響度は IC モデルより安定しないことも確認した。

今後は、他の指標と期待影響度の相関関係の分析や、ノード次数優先選択、リンク多重度優先選択、潜在リンク選択の 3 つを混合させた成長モデルの提案などに着手していく予定である。

## 謝辞

本研究は、科学研究費補助金基盤研究 (C)(No. 23500312) の補助を受けた。

## 参考文献

- [1] D. J. Watts, S. H. Strogatz, “Collective dynamics of ‘small-world’ networks”, *Nature*, Vol. 393, pp. 440-442, 1998.
- [2] Barabasi A-L, Albert R, “Emergence of Scaling in Random Networks”, *Science* 286(5439):509-512, 1999.
- [3] 鈴木智也, 「情報伝達に基づいた有向重み付き複雑ネットワーク解析」, 情報処理学会論文誌 数理モデル化と応用 2(1), 70-78, 2009.
- [4] 加藤翔子, 斉藤和巳, 「リンク優先選択を付与した嗜好ネットワーク成長モデルの提案」, 第 9 回ネットワーク生態学シンポジウム, 2012.
- [5] 加藤翔子, 斉藤和巳, 「高多重度有向ネットワーク成長モデルの提案」, 第 5 回データ工学と情報マネジメントに関するフォーラム, 2013.
- [6] Kempe, D, Kleinberg, J. and Tardos, E., “Maximizing the spread of influence through a social network”, *Proc. KDD '03*, pp. 137-146, 2003.
- [7] D.J.Watts, “A simple model of global cascades on random networks”, *Proc. National Academy of Science, USA*, col.99, pp.5766-5771, 2002.
- [8] Alexei Vazquez, “Growing networks with local rules: preferential attachment, clustering hierarchy and degree correlations”, *Physics Review E* 67, 2003.
- [9] 鳥海不二夫, 篠田孝祐, 栗原聡, 榊剛史, 風間一洋, 野田五十樹, 「震災がもたらしたソーシャルメディアの変化」, ネットワークが創発する知能研究会, 2011.

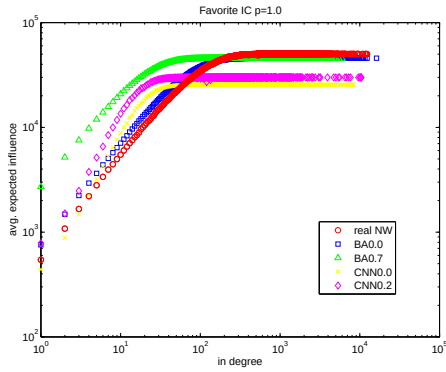


図 1: Favorite 入次数毎 IC モデル

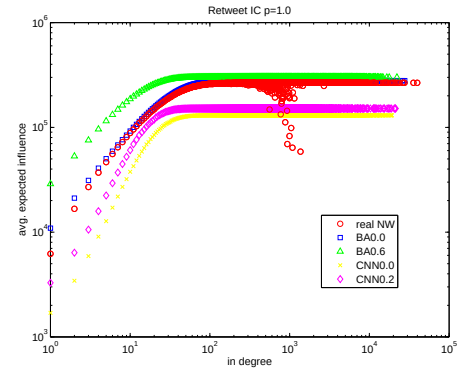


図 2: Retweet 入次数毎 IC モデル

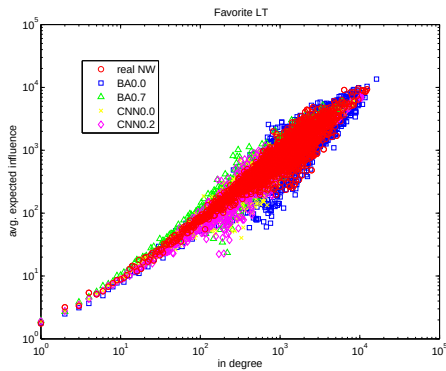


図 3: Favorite 入次数毎 LT モデル

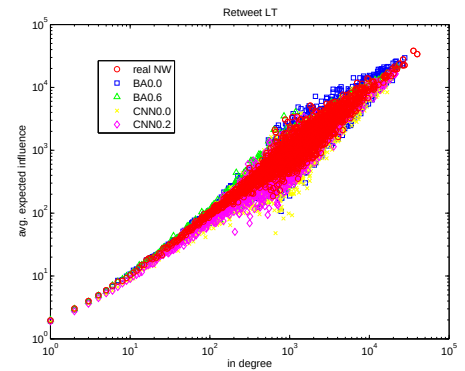


図 4: Retweet 入次数毎 LT モデル

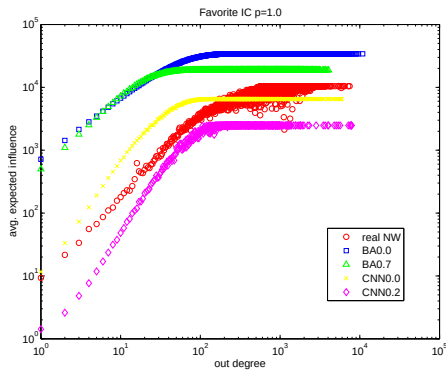


図 5: Favorite 出次数毎 IC モデル

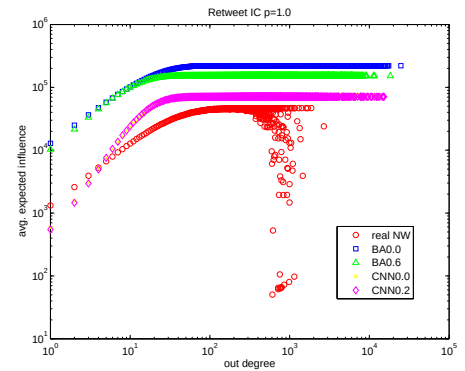


図 6: Retweet 出次数毎 IC モデル

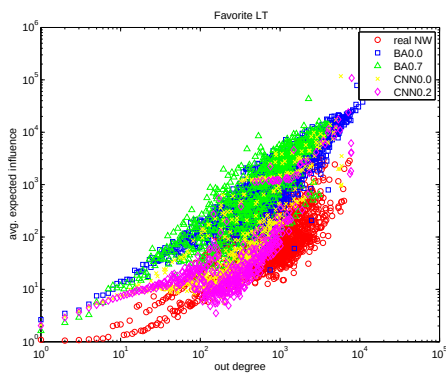


図 7: Favorite 出次数毎 LT モデル

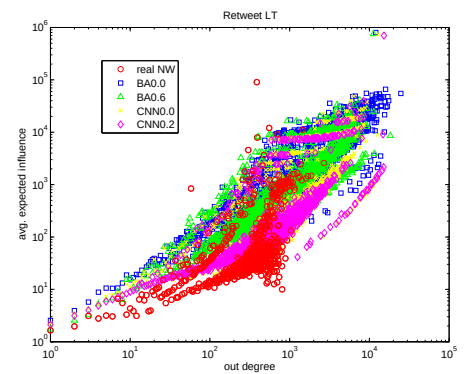


図 8: Retweet 出次数毎 LT モデル