

高多重度有向ネットワーク成長モデルの提案

加藤 翔子[†] 齊藤 和巳[†]

[†] 静岡県立大学 経営情報学部 〒422-8526 静岡県静岡市駿河区谷田 52-1

E-mail: [†]{b09032,k-saito}@u-shizuoka-ken.ac.jp

あらまし 高い多重度の有向リンクを生成することが可能なネットワークの成長モデルの構築へ向け, CNN(Connecting Nearest Neighbor) モデルを多重有向グラフへ応用したもの(多重有向 CNN モデル)を考える. さらに, リンク多重度優先選択と呼ぶリンク生成方法を用いる確率を, 多重有向 CNN モデルと混合させたモデルを新たに提案し, 先行論文で提案した, BA(Barabasi-Albert) モデルを多重有向グラフに拡張したモデルと混合させたものも対象とし, 現実ネットワークとその構造を比較した. 現実ネットワークについては, 高多重度リンクが出現するデータとして, Twitter の Favorite と Retweet を多重有向グラフ化したものを用いた. 評価実験の結果, リンク多重度優先選択のパラメータを付加することで, 提案 BA モデルについてはリンク次数と入次数, 提案 CNN モデルについては平均多重度分布とジニ係数分布が, それぞれ類似することを示した.

キーワード ネットワーク成長モデル, 多重有向グラフ

Shoko KATO[†] and Kazumi SAITO[†]

[†] School of Administration and Informatics, University of Shizuoka

52-1 Yada, Suruga-ku, Shizuoka, 422-8526 Japan

E-mail: [†]{b09032,k-saito}@u-shizuoka-ken.ac.jp

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.

有向リンクの多重度が高い数値を取りうるネットワークの成長モデルの構築へ向け, 先だって発表した論文[4]では, バラバシ=アルバートモデルを多重有向グラフに拡張したもの(多重有向 BA モデル)と, それにリンク多重度優先選択と呼ぶリンク生成方法を用いる確率を新たに付加したモデル(提案 BA モデル)を考え, 現実ネットワークと比較した. 本論文では, CNN(Connecting Nearest Neighbor)モデル[5]を多重有向グラフへ応用したもの(多重有向 CNN モデル)を考え, さらに, 前述のリンク多重度優先選択を多重有向 CNN モデルと混合させたモデル(提案 CNN モデル)についても考える. 評価実験では, 高多重度リンクが出現する現実ネットワークとし

て, Twitter の Favorite と Retweet を多重有向グラフ化したものを用い, [4]で現実ネットワークにより類似すると示した提案 BA モデルとも比較する. その結果, リンク多重度優先選択のパラメータを付加することで, 提案 BA モデルについてはリンク次数と入次数, 提案 CNN モデルについては平均多重度分布とジニ係数分布について, 多重有向グラフに拡張したモデルより Favorite に近い構造や動向を実現できると示した.

BA モデルや CNN モデルは, どのノード間にリンクを張るか, 即ちノードの選択方法に注目したアルゴリズムであるが, 我々が提案するリンク多重度優先選択は, どのリンクの多重度を増やすか, 即ちリンクの選択方法に注目しているため, 十分に新規性を有する研究であると考え.

本論文の構成は以下となる. まず, 本研究に用いた3つのモデルのアルゴリズムについて説明する. 次に, 評価に用いたデータの概要を述べるとともに, 評価実験による結果と考察を報告する. 最後に, 本論文のまとめについて述べる.

2. モデル

ノード集合 V とリンク集合 E からなる多重有向グラフを $G = (V, E)$ とする. 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, E は

多重集合として定式化される．以下では，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，ノード v の隣接ノード集合を $A(v)$ と表記する．リンク $e = (u, v) \in E$ に対して， $x \in A(v), x \neq u$ のとき $f = (u, x)$ または $f = (x, u)$ のリンクを考える．同様に， $x \in A(u), x \neq v$ のとき $f = (v, x)$ または $f = (x, v)$ のリンクを考える．このようにして，リンク e から導かれるリンク f の集合をリンク e の潜在リンク集合 $F(e)$ と呼ぶ．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．

2.1 提案 BA モデル

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 α 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 β 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 α で新ノード生成 S3.，さもなければ，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 β でリンク多重度優先選択でリンク追加 S5.，さもなければ，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．

本稿では便宜上，この方法で生成するモデルを提案 BA モデルと呼ぶ．

2.2 CNN モデル

CNN モデル [5] で提案されたのと同様に，ノード v の潜在リンク集合 $F(e)$ から潜在リンク $f = (v, x) \in F(e)$ 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，2.1 で提案したリンク多重度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．2.1 同様，新ノード生成確率を α 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多重度優先選択のどちらを選ぶかの確率を γ 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 α で新ノード生成 S3.，さもなければ，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，ノード次数優先選択確率 $P(v)$ 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 γ でリンク多重度優先選択でリンク追加 S5.，さもなければ，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. リンク多重度優先選択確率 $P(e)$ で e を選び， $x \in A(v), x \neq u$ または $x \in A(u), x \neq v$ であるような x をノード次数優先選択確率 $P(x)$ で選び，始点と終点のどちらとなるかを $1/2$ の確率で決める． x と u または v を端とする潜在リンク f を実リンクにし， E に加える．

本稿では便宜上，この生成法を提案 CNN 法と呼び， $\gamma = 0$ とした特別なものを多重有向 CNN 法と呼ぶ．

3. 評価実験

評価に用いたデータについての概要，および評価実験の結果とそれに基づく考察を述べる．

3.1 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^(注1) 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 [6] 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．

(注1): <http://favotter.net/>

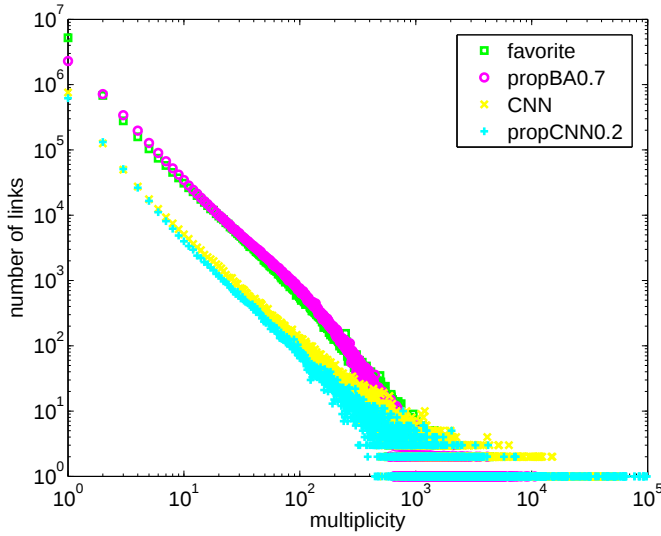


図 1 Favorite リンク次数分布

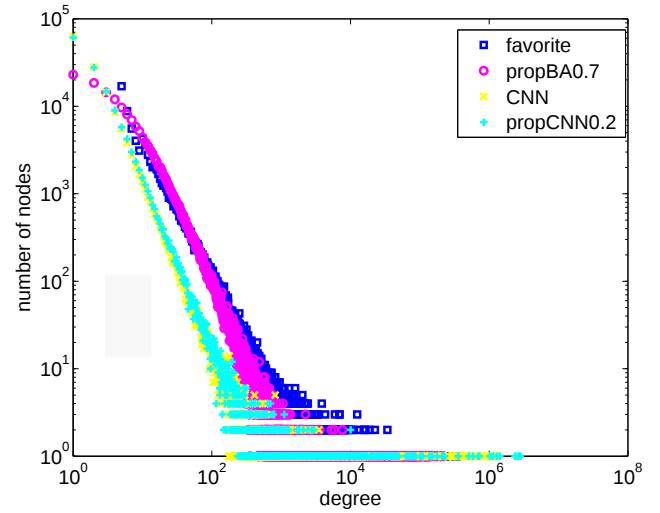


図 2 Favorite 入次数分布

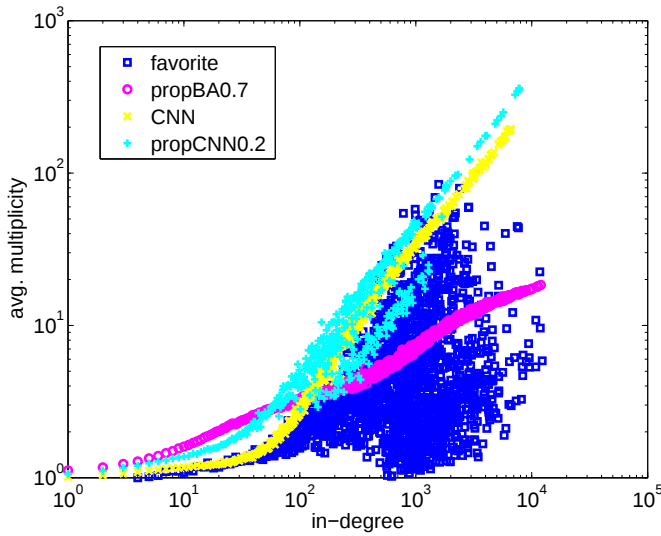


図 3 Favorite 入隣接数毎平均多重度分布

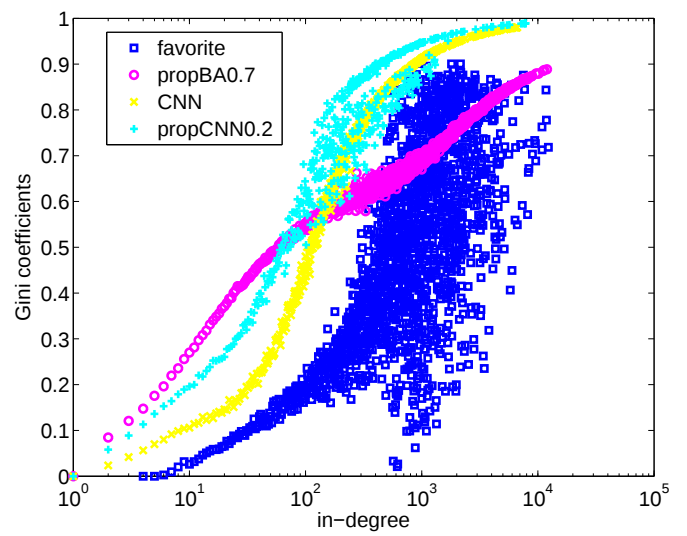


図 4 Favorite 入隣接数毎ジニ係数分布

3.2 実験結果

図 1, 2, 3, 4, 5, 6, 7, 8 は, 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多重有向 CNN モデルと提案 CNN モデルを用いて生成し, その構造を提案 BA モデルと共に比較した結果である. Favorite ネットワークのリンク次数分布は緑の□, 入次数分布, 入隣接数毎平均多重度分布, 入隣接数毎平均ジニ係数分布は青の□でプロットし, 提案 BA モデルを用いたものはマゼンダ色の○, 多重有向 CNN モデルを用いたものは黄色色の×, 提案 CNN モデルを用いたものはシアン色の+でプロットした.

リンク多重度優先選択を用いる確率 β と γ は, Favorite との比較では $\beta = 0.7, \gamma = 0.2$, Retweet との比較では $\beta = 0.6, \gamma = 0.2$ で設定した. それぞれ, $0 < \beta < 1, 0 < \gamma < 1$ の範囲で実験し, β については, 最もリンク次数分布が現実ネットワークと近似するものを目視で推定し設定値とした. γ については, 値が高くなるに連れ平均多重度分布や平均ジニ係数分布が上にシフトするため, 平均多重度分布や平均ジニ係数分布にバラつきが出るもののうち最も低い値とした.

Favorite ネットワークとの比較において, 図 3, 4 より, 平均多重度分布とジニ係数分布ともに, 提案 CNN モデルは他のモデルよりバラつきが顕著であり, Favorite に動向が類似しているが, 全体的に現実より高い値をとっている. これは, 提案 CNN モデルにおけるノード次数やリンク次数の取りうる値が Favorite より広いことが影響すると考えられ, 図 1, 2 のリンク次数分布と入次数分布から推察することができる.

Retweet ネットワークとの比較において, 図 7, 8 より, 平均多重度分布とジニ係数分布ともに, 多重有向 CNN モデルと提案 CNN モデルは, 提案 BA モデルよりバラつきが顕著であり, Retweet に動向が類似するが, 多重有向 CNN モデルは全体的に現実より低い値をとり, 提案 CNN モデルは高隣接数において高い値をとる. これは, 提案 CNN モデルにおけるノード次数やリンク次数の取りうる値が Retweet より広いことが影響すると考えられ, 図 5, 6 のリンク次数分布と入次数分布から推察することができる.

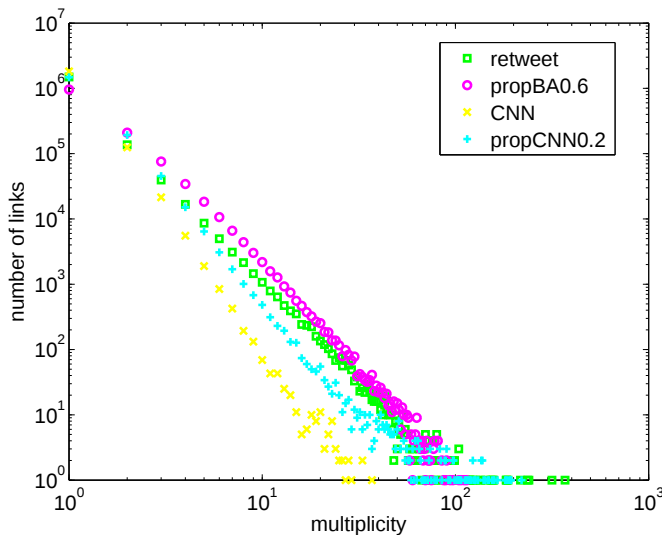


図 5 Retweet リンク次数分布

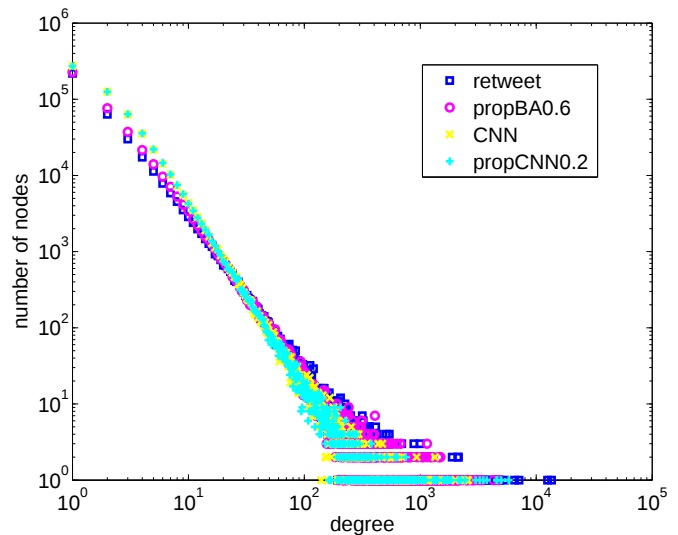


図 6 Retweet 入次数分布

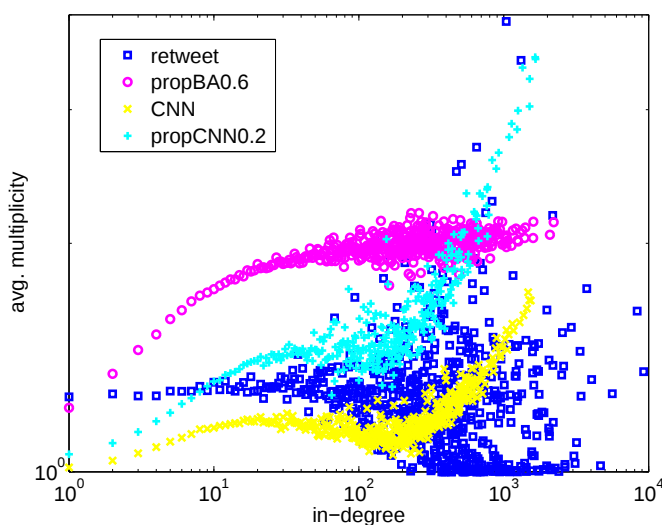


図 7 Retweet 入隣接数毎平均多重度分布

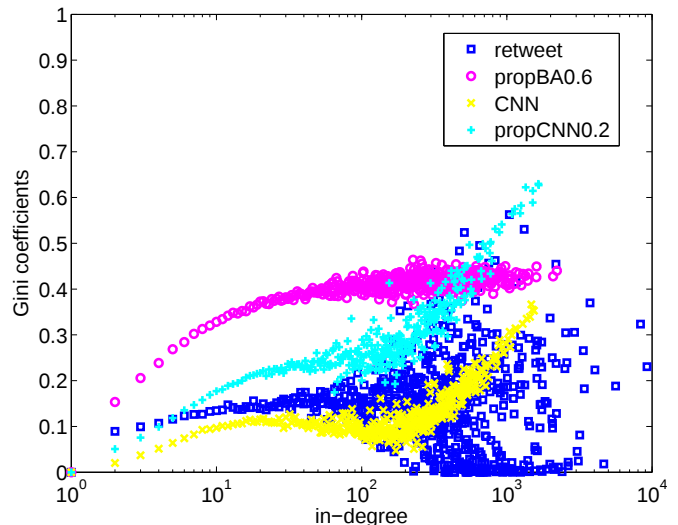


図 8 Retweet 入隣接数毎ジニ係数分布

4. おわりに

高次数な有向リンクを有するネットワークの成長モデルの構築へ向け、提案 BA モデル，多重有向 CNN モデル，提案 CNN モデルの 3 つのモデルについて，現実ネットワークとその構造を比較した．現実ネットワークについては，高多重度リンクが出現するデータとして，Twitter の Favorite と Retweet を多重有向グラフ化したものを用いた．

評価実験により，リンク多重度優先選択のパラメータを付加することで，どちらのモデルも多重有向グラフに拡張したモデルより Favorite に近い構造や動向を実現できると示したが，提案 BA モデルについてはリンク次数と入次数，提案 CNN モデルについては平均多重度分布とジニ係数分布が優れた結果となり，課題が残る分析を互いが補う形となっている．

従って，今後の課題としては，BA モデル，CNN モデル，リンク多重度優先選択の 3 つを融合させたモデルについて検討する予定である．

謝辞 本研究は，科学基礎研究費補助金基盤研究 (C)(No.22500133) の補助を受けた．

文 献

- [1] D. J. Watts, S. H. Strogatz, “Collective dynamics of ‘small-world’ networks”, Nature, Vol. 393, pp. 440-442, 1998.
- [2] A.-L. Barabasi, R. Albert, “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, 「リンク多重度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] A. Vazquez, “Growing networks with local rules: preferential attachment, clustering hierarchy and degree correlations”, Physics Review E 67, 2003.
- [6] 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 (JWEIN'11), 2011.