入出次数相関と相互リンク率を考慮した Twitter を表現した成長ネットワークモデル

○木島裕希　高橋真吾（早稲田大学大学院）

Growing Network Model for Twitter, representing Follow-Follower Correlation and Mutual Link

* Y. Kijima and S. Takahashi (University of Waseda)

概要－　Social Networking Service（以下，SNS）が人々のコミュニケーションのツールとして一般的に用いられている．なかでも Twitter は利用者も非常に多い SNS である．Twitter の特徴は情報の拡散性と即時性にある．その一方で，特徴の弊害として信用性の低い情報が広まりやすい．このような情報伝播の研究において Twitter を表したモデルの特徴は，スケールフリー性にある．しかしながら，入出次数相関と相互リンク率も特徴として挙げられている．そこで本研究では，Twitter のフォロー形成に関する実データを基に，成長ネットワークモデルの提案を行う．そして提案モデルにより得られた結果と考察について述べる．

キーワード: ネットワークモデル，Twitter，複雑ネットワークの科学

1.　序論

1.1　研究背景

現代社会では人と人とのコミュニケーションのツールとして Social Networking Service（以下，SNS）と呼ばれるものが，よく用いられている．Facebook や Mixi, Twitter など，様々なツールが存在している．なかでも Twitter は，利用者も非常に多い SNS である．Twitter が，その他の SNS と異なる点は，特定の場合を除き，相互認証を必要としない SNS であるというところにある．この特性を持つことから，人とつながりを持つことの心理的障壁が小さく，人々がつながりやすい SNS であると言える．

SNS に関する研究は，大きく分けると二つに分類される．一つは，人と人とのつながりを表現しているネットワークがどのように構築されているのか，といった内容の研究．二つ目は，ネットワーク上で情報や病気などが，どのように拡散，伝播していくのかを分析する研究である．一つ目の，「ネットワークの構築」の分野では，日本においては主に Mixi について研究されている．三井ら [7]や宮崎 [9]が，現実の Mixi のネットワークに近い成長ネットワークモデルを提案している．二つ目の「ネットワーク上での伝播」の分野は，湯浅ら [11]がある．湯浅らは感染症流行予測を行う為に，様々なネットワークモデルに対して，感染症のモデルとして有名な SIR モデルを適用し，それぞれのネットワークモデルにおける感染症流行の差異を，ネットワークの統計的指標である平均頂点間距離，クラスタ係数などの観点から分析を行っている．

しかしながら，Twitter に関する研究は「ネットワーク上での伝播」が主であり，「ネットワークの構築」に関する研究は少ない．例えば，池田ら [1]は Twitter の利用者が受け取る情報に，デマが含まれているとし，デマ情報を受け取った後に訂正情報を受け取るといった状態変化を，感染症に見立て，SIR モデルを適用し，どのようにデマの情報が拡がり，訂正されていくのか，といった研究を行っている．この時用いているネットワークモデルは，スケールフリー性を持ったネットワークに固定するため，パレート分布によりモデルを作成している．

1.2　研究目的

本研究では，小出ら [4]による Twitter のフォロー形成に関する実証データ（Table 1）を元にし，それを再現するような成長ネットワークアルゴリズムの提案を行い，構築されたネットワークの特徴の分析を行う．モデルの妥当性評価としても実証データの統計的指標を用いて行う．それらの中でも Twitter の特徴として挙げることの出来る，入出次数相関と相互リンク率を中心に妥当性をとる．この二点を選択した理由として，Twitter が特定の場合を除き，認証を必要とせず，また，相互リンクが確立されている訳ではないことが挙げられる．そういった点があるにも関わらず，入出次数相関は強く，双方向リンク率も比較的高い値をとるのが特徴となっている．

1.3　研究方法

本研究では，モデルの構築にあたって，BA モデルや CNN モデルのような，「成長型ネットワークモデル」を用いる．ネットワークが成長していく過程を，ノードが能動的にネットワークに加わり，リン

クを形成していく様子として表現する．構築したネットワークの評価は統計的指標を用いる．

2. 先行研究

本章では先行研究について説明する．はじめに複雑ネットワークの科学の分野である，ネットワークモデルについて説明する．次に，Twitter の実データの分析により得た，フォロー形成に関する考察を行っている実証研究[4]を説明する．最後に，Twitter における情報伝播の研究[1]に関して説明する．

2.1 ネットワークモデルの研究

ネットワークモデルはグラフ理論により表現され，その研究は古くからなされている[6]．ランダム・グラフに始まり，今では CNN モデル[16]のように，現実に存在するネットワークと近似されたネットワークモデルも存在する．これらネットワークモデルは，コンピュータのつながりを表現していたり，人と人とのつながりを表現していたりする．ネットワークモデルにおける研究の，中心的ともいえる３つの特徴は「スモールワールド性」と「スケールフリー性」，そして「クラスター性」である．以下では，それら特性をふまえたネットワークモデルについて，本研究に最も関連深い，BA モデルと CNN モデルについて詳細を述べる．

2.1.1 BA モデル

BA モデル[13]は，Barabasi と Albert により提案されたネットワークモデルであり，スケールフリー性を持ったネットワークが構築される．その特徴は，成長と優先的選択にある．成長とは，ネットワークの構築をする際，ノードとエッジをステップ毎に加えていく動作のことを示している．優先的選択とは，加わったエッジが接続する先である既存のノードの持つエッジ数に応じて選択確率が変化し，決定される様子のことを示している．

アルゴリズムは以下のように表される：

1. ノード数n_0 $(n_0 \geq 1)$の完全グラフK_{n_0}を作成する．
2. 新たなノードをネットワークに加える．この時，新たなノードはエッジをe $(e \leq n_0)$を持った状態で加わる．
3. 新たなノードが持っているエッジを，既存のノードに接続する．この時，既存のノードを選択する確率は，既存のノードが既に持っているエッジ数に応じて決定される．その際，ループや二重辺とならないようにする．
4. ノード数が，指定された最大ノード数に達していなければ２へ戻る．最大数に達していれば終了する．

BA モデルにおける，優先的選択のノードの選択確率はノードの次数（持っているエッジの数）に比例している．すなわち，次数が高いノードは選択されやすく，次数が小さいノードは選択されにくくなる．この優先的選択により，ステップを重ねる毎に次数の差が開き，選択確率に差が生まれる．結果として構築したネットワークにスケールフリー性が生まれる．

2.1.2 CNN モデル

Connecting Nearest-Neighbor Model[16]は，Vazquez により提案されたネットワークモデルである．その特徴は，クラスター性を高めるために，三角構造を形成していくことにある．

アルゴリズムは以下のように表される：

1. 指定した最大ノード数になるまで確率的に２，３を繰り返す．
2. 確率$1-u$で新しいノードをネットワークに加える．この時，既存のノード一つをランダムに選択し，接続する．そして，新しく加わったノードと，ランダムに選択した既存のノードが接続しているノードの間に，潜在エッジを作る．
3. 確率uで潜在エッジ一つをランダムに選択し，エッジにする．

ここで潜在エッジとは，実際にはエッジが存在していないが，今後エッジとして作成する可能性があるエッジであることを示している．

CNN モデルにおいて，BA モデルと同様に成長はあるものの，優先的選択は無い．しかしながら，BA モデル同様にスケールフリー性（次数分布のベキ則）は生まれる．また，BA モデルでは表現が出来ていない，クラスター性についての表現が可能となっている．クラスター性は SNS に対する研究としては非常に重要な特性である．なぜならば，社会における人と人とのつながりは，コミュニティと呼ばれる近しい属性を持った人々の集団によっても形成されるからである．また，論文においては次数相関についての分析もされている．

2.2 Twitter の分析研究

2.2.1 Twitter の分析

Twitter を詳細に分析している研究として Kwak et al.[15]がある．Kwak et al. は Twitter を利用している 417 万の人々の，1.47 億リンクについてフォロー関係の分析を行なっている．その結果として既存の SNS と Twitter とは異なるネットワーク構造であることを述べている．

フォローについて注目すべきは，次数が 20 の付近と 2,000 の付近である．フォローに関してこの次数付近のノードが多い理由として筆者たちは以下のよ

うに考察を述べている. 次数が 20 の付近での変化は, Twitter を初めて間もない人達に, Twitter 側がおすすめとして紹介する人達が 20 人となっているからである. 検索などの手間なくフォローをすることが出来るため, 20 人程度までは簡単にフォローすることが可能となり, 相補累積分布にも現れていると考えられる. また, 次数 2,000 付近での変化は Twitter にかつて存在した次数制限にあると考えられている. 2009 年以前までは人々がフォローできる数の上限として 2,000 人までというものが設けられていた. フォロワの分布の特徴も指摘されている. フォロワは次数 100,000 の付近まではベキ則に従っており, その冪指数は 2.276 となっている（実在するネットワークのほとんどは 2 から 3 以内の冪指数と言われている）. しかしながら次数 100,000 以降はベキ則の分布よりも多い部分にフォロワを多く持つユーザが多い事を示している. これは他の社会的ネットワークには現れない特徴である.

Akioka et al. [12]は, 日本において Twitter とその他のメディアが与える影響を定量的に分析している. そのなかで筆者らは, フォローとフォロワの次数がベキ則に従うこと, そしてフォローとフォロワに相関が存在することを示している.

筆者達が収集したデータは, ノード数 50,000 ノード, リンク数 64,800,000 本である. Fig. 1 は横軸がフォロー数, 縦軸がノード数の両対数グラフとなっている. Fig. 2 は横軸がフォロワ数, 縦軸がノード数の両対数グラフとなっている. これら二つの Fig. に共通しているのはベキ則が表れていることである. しかし Fig. 1 においてはフォロー数が 20 および 2,000 の付近でベキ則から, 乖離したノード数が存在することがフォロワ数の分布とは異なる. このフォロー数 20 と 2,000 付近でのベキ則からの乖離は Kwak et al. も同様の分析をしているように, おすすめとして表示される 20 人の影響と, 次数制限によるものと考えられる.

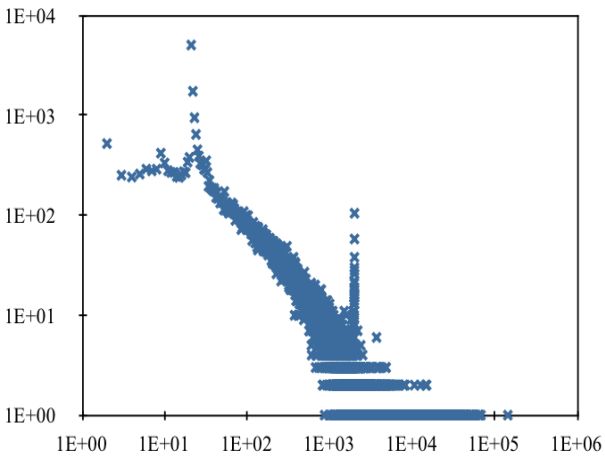

Fig. 1: フォロー数の分布（横軸：出次数，縦軸：ノード数）[12)]

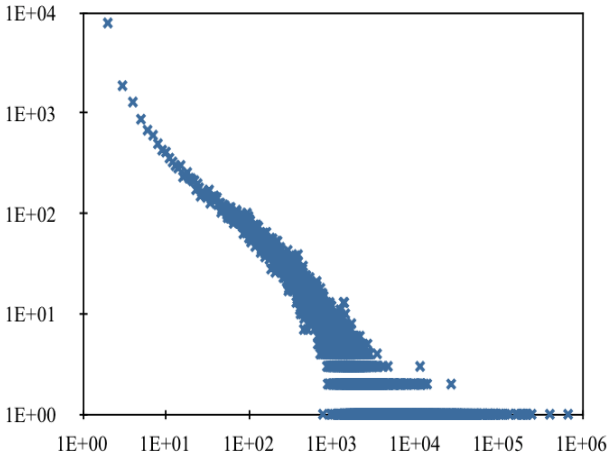

Fig. 2: フォロワ数の分布（横軸：入次数，縦軸：ノード数）[12)]

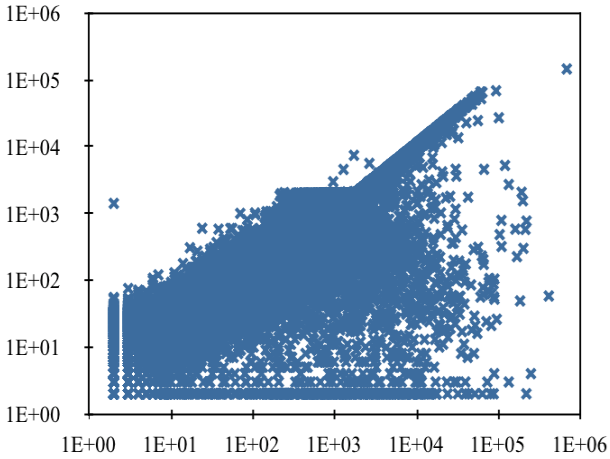

Fig. 3: フォロー・フォロワの散布 Fig.（横軸：入次数，縦軸：出次数）[12)]

Fig. 3 は横軸がフォロワ数，縦軸がフォロー数の両対数グラフとなっている．一般的にフォロー数が多いユーザは，フォロワ数も多いといった線形なプロットが存在していることがわかる．また，フォロワ数とフォロー数が同じ部分の付近でばらついていること，フォロワ数が多い人は，ほとんどがフォロワ数以下だけフォローしていることもわかる．このように，フォロー数とフォロワ数の間には相関があることが確認できる．

2.2.2　フォロー形成に関して

小出ら [4)]による，フォロー形成に関して実証データを基に考察している研究がある．小出らは Twitter のフォローネットワークの高次ノードを分析することにより，その他のネットワークとどのように違うかを分析している．分析した結果，Twitter のフォローネットワークの特徴として，強い双方向関係により構築された大規模な高コリンクグループと，双方向関係がほとんど見られない複数の小規模な低コリンクグループが存在することを示している．筆者達

が使用したデータはノード数 1,079,986 ノード, リンク数 157,371,341 本である.

Table 1: Twitter の基本統計量 [4)]

平均次数	145.7
最大出次数	97,938
最大入次数	175,616
リンク密度	0.000270
入出次数相関	0.74
双方向リンク率	0.60

Table 1 はデータの基本統計量を分析した結果である．Twitter においては，有名人などが非常に多くフォローを集めるネットワークである．しかしながら，有名人はフォローを返さないため，最大入次数と最大出次数の間に開きがある．また，必ずしも双方向性が確立されていないにも関わらず，双方向リンク率は 0.60 と高い値になっている．これは日本の Twitter において特徴的な点であると指摘されている．

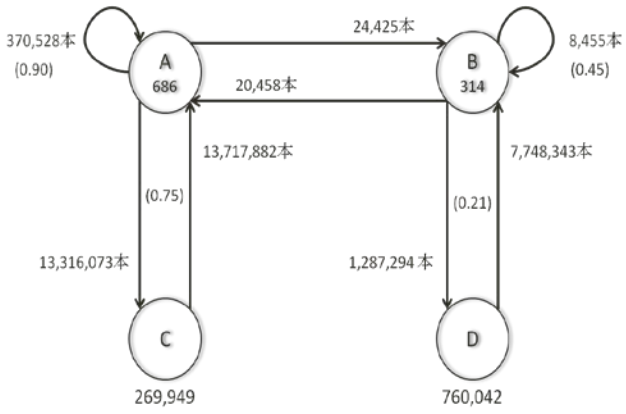

Fig. 4: 入次数上位 1000 ノードのリンク構造 [4)]

Fig. 4 は入次数上位 1000 ノードのフォロー構造をまとめた図である．図におけるノードはネットワークにおけるノード集合を表しており，中に書かれている英字がそのノード集合の名称を，中もしくは下側に書かれている数字が，ノード集合に含まれるノード数を表している．A は「高コリンクグループに属するノード集合」, B は「複数の低コリンクグループに属するノード集合」，C は「A の親ノードかつ，A，B に属さないノード集合」，D は「B の親ノードかつ，A，B に属さないノード集合」を表している．ここで親ノードとはフォローしているノードを，子ノードはフォローされているノードのことを表す．

N は全ノード数，ノード集合を$V = \{v_1, v_2, \dots, v_N\}$, リンク集合 E を$V \times V$の部分集合とする．また，任意のノード$v_i \in V$に対し，$v_i$からのリンクを有するノード集合（子ノード集合）を$A(v_i) = \{v_j; (v_i, v_j) \in E\}$とし，$v_i$へのリンクを有するノード集合（親ノード集合）を$B(v_i) = \{v_k; (v_i, v_k) \in E\}$とする．この時，ネットワーク全体に対する相互関係を表す双方向リンク率は以下の式で表される：

$$cr = \frac{1}{|E|} \sum_{v_i \in V} |A(v_i) \cap B(v_i)|$$

Fig. 4 のノード集合間に存在する矢印はリンクを表しており，隣の数字がそのリンクの本数を表している．同じノード集合から出て入るリンクは，そのノード集合内でのリンクを表す．リンク本数近くに書かれているカッコで囲まれた数字は，そのノード集合間での双方向リンク率を表している．なお，ノード集合 C と D の関係は，D が C を部分的に包括しており，C のノード集合の約 90%は D に含まれる形である．ノード集合 A と C の間では，ノード数が大きく違うにもかかわらず，リンク数がほとんど同じで，双方向リンク率も 0.75 と高い．ノード集合 A 内は大部分が双方向リンクしており，双方向リンク率は 0.90 と高い．これに対し，ノード集合 B と D の間にはノード数，リンク数共に大きな差があり，双方向リンク率も 0.21 と低い．また，ノード集合 B 内のノード間のリンク数は非常に少ないが，双方向リンク率は 0.46 と B，D 間よりは高い値となっている．ノード集合 A，B 間の双方向リンク率，また，それらの親ノード集合の双方向リンク率には触れられていない.

2.3 Twitter 上の情報伝播に関する研究

Twitter の研究で重要な点は情報伝播にもある．Twitter はその情報の拡散性と即時性に特徴を持つ．しかしその特徴の悪影響として，デマ情報が拡散しやすい．これらデマの拡散から，訂正情報によるデマの収束まで扱っている研究として白井ら [5)]や池田ら [1)]がある．白井らは，感染症流行の拡がりを数理的に扱った SIR モデルをデマ情報の拡散に適用している．すなわちデマ情報をまだ受け取っていない人を S(Suseptible)，デマ情報を受け取った人を I(Infectious), 訂正情報を受け取った人と R(Recoverd) としている．シミュレーションで用いているネットワークはスケールフリー性を表現するため，パレート分布を用いたネットワークとしている．

シミュレーションを実際に起きたデマ拡散と比較し，デマ拡散の状況を近似している．その上で，訂正情報が効率的に拡散していくように 3 つのシナリオを用意し，比較検討している．結果として，ネットワークのハブとなっているユーザ，またはデマツイートをしたユーザの内，最もフォロワが多いユーザに訂正情報を流してもらうことが，収束に向けての効果が高いとしている．

Table 2: スケールフリーネットワーク [1)]

ノード数	50,000
リンク数（次数）の期待値	上限 = 3,000 下限 = 10 パレート指数 = 0.5
リンクのされやすさ	上限 = 15.0 下限 = 0.05 パレート指数 = 0.5

以上のように, Twtitter におけるスケールフリー性は表現しているものの，リンク形成による入出次数相関や，双方向リンク率を考慮できていないため，本研究ではこれらの点を考慮した成長ネットワークモデルを提案する．

3. 提案モデル

3.1 モデル概要

本研究では，小出ら [4)]の研究における実証データと考察を基に，ネットワークが成長していくモデルを提案する．モデルはノードが２つの完全グラフから始まる。ノードの最大数は 100,000 とし，ノードが 100,000 存在する状態で，更にノードを追加した場合にシミュレーションは終了する. CNN モデルのように，確率的にリンクの生成とノードの生成を繰り返す．また，BA モデルの優先的選択のように，リンクを生成する際，ノードの属性を確率的に選択し，属性の中からリンクする先をランダムに選ぶ．なお，シミュレーションの１試行につき，１つのネットワークが生成される．

3.1.1 ノードの属性

本研究では小出ら [4)]の研究を基にノードに属性を設ける．Twitter では，ユーザは著名人を一方的にフォローして，著名人（Fig. 2.4 のノード集合 B）自身が発言する近況や最新の情報を得る．そして得た情報をリツイートすることにより，情報を伝達していくという特徴がある．そのため，著名人は多くのユーザからフォローされやすい．しかしながら，著名人自体はフォローをあまりせず，対話もあまり行わない，一方的な情報発信のみである．このことが影響し，トピックごとに複数のグループに分割されている．これらはフォローの双方向性の低さ，フォロワとの対話数の少なさから実証されている．このような実社会で有名であり，Twitter 上での対話が少ない属性を持つ人々を，本研究では「有名人」として定義する．

次に，実世界においては著名ではないが，そのフォロワ数の多さから少なくとも Twitter 上では著名であると推測できる人々について述べる（Fig. 2.4 のノード集合 A）. 全体として共通の話題は無いものの，多種多様なユーザが巨大な集団を作っている．リツイートによる情報伝達が少ないことから，ツイート自体の独自性・重要性は評価されていない．しかしながら，フォロワとの対話は活発である．この対話によりフォローを多く獲得できていると考えられている．このような実社会では有名でないにもかかわらず，Twitter 上での対話が多い属性を持つ人々を，本研究では「活動人」として定義する．

その他，一般的なユーザを「一般人」として定義する．

以上のように，３つの属性を定義し，ネットワークに新たなノードを生成する際に，確率的に有名人であるか，活動人であるか，一般人であるかの属性を割り当てることとする．また，リンクを生成する際には，フォロー先を選択する時に，確率的に属性を選択し，その属性の中からノードをランダムに選択する．

3.1.2 ノード生成フロー

ノードは確率 *createnode* で生成する．ノードは「*nodestatus*」というノードの属性を表すパラメータを持った状態で生成される．*nodestatus* は 1, 2, 3 のいずれかを確率的に割り当てる．1 は有名人，2 は活動人，3 は一般人を表している．ノードに属性を割り当てた後に，同じ属性をもつノードが既存のネットワークに存在していれば，その属性のノードと相互リンクを行う．既存のネットワークに同じ属性を持つノードがなければ，一般人と相互リンクを行うこととする．最初に相互リンクを行う理由は，ネットワークにおいて知人が存在し，その人とは相互にフォローするという仮定をもとにしている．

- ノード生成アルゴリズム：

1. 新しいノードを生成する．
2. 生成したノードに属性を割り当てる．割り当ては確率的に行なう．
   確率 *famous* で有名人に，*semifamous* で活動人に，それ以外は一般人に割り当てる．
3. 割り当てた属性と同じ属性のノード一つをランダムに選択し，相互フォローを行なう．同じ属性のノードが存在しなければ，一般人の属性のノード一つをランダムに選択し，相互フォローを行ない，フローを終了する．

3.1.3 リンク生成フロー

リンクは確率1 – *createnode*で生成する．リンク生成のフローは，フォロー行動をするノードの決定から行なう．フォロー行動を行うノードは属性に関係なくランダムに選択する．フォロー行動を行うノードを決定した後に，どの属性にフォローを行うかを決定する．決定した属性のノードが存在すれば，それらの中からランダムに一つのノードを選択する．

決定した属性のノードが存在しなければ，一般人の属性からランダムに一つのノードを選択しフォローを行う．フォロー先を決定した後に，相互フォローを行なうか否かを確率で選択し，フローを終了する．リンク生成のフローでフォロー行動のみを考えるのは，ユーザであるノードの能動的な行動のみをモデル化するためである．また，フォロー先の属性を確率的に選択するのは，有名人や活動人が，一般人よりはフォローされやすい属性を持っており，Twitter においてユーザを選択する際に用いられるのが検索機能であるからと仮定をもとにしている．また，相互フォローを行なうか否かの決定は，フォローとフォロワの属性により確率的に決定される．

- リンク生成アルゴリズム：

1. フォロー行動を行うノードを，既存のネットワークからランダムに選択する．
2. フォロー先の属性を確率的に選択する．
   確率$(1 - simplechoice) * famouschoice$で有名人の属性に，
   確率$(1 - simplechoice) * semifamouschoice$で活動人の属性に，
   確率$simplechoice$で一般人の属性を選択する．
3. 選択した属性のノード一つをランダムに選択し，フォロー先とする．もし選択した属性のノードが一つもなければ，一般人の属性からランダムに選択する．また，選択した属性のノード全てをフォローしていた場合も，一般人の属性からランダムに選択する．
4. 属性間の相互リンク確率パラメータ *colinkrate* により，確率的に相互リンクを行なうか否かを決定し，終了する．

3.1.4 パラメータ設定

本実験においては小出ら[4]の実証データを基に，パラメータを設定している．これらはシミュレーションにおいて任意に変化させることが可能であり，ネットワークの特性もこれらパラメータにより変化する．Table 3 は実験で用いるパラメータである．

maxnode は 100,000 に設定し，*createnode* は 0.01 に設定を行なう．有名人生成率 *famous* と活動人生成率 *semifamous* は，作成するネットワークの規模から逆算し，0.00314 と 0.00686 に設定する．Fig. 4 においてノード集合 C および D から A にはられているリンク数は 13,316,073 本，B にはられているリンク数は 7,748,343 本，合計で 21,466,225 本である．すなわち，有名人に対しては 36%，活動人に対しては.64%，C と D から A, B に対してリンクが張られている．これをこのまま確率にした場合，フォロー行動を行なうノードが，能動的には一般人を選択しなくなる．そこで，一般人選択確率を設定し，それを差し引いた部分に有名人，活動人に対するリンク割合を適用しパラメータとしている．

Table 3: シミュレーションで用いるパラメータ

パラメータ	値
maxnode	100,000
createnode	0.01
famous	0.00314
semifamous	0.00686
famouschoice	0.324
semifamouschoice	0.576
simplechoice	0.100
colinkrate	各属性に依存

次の Table 4 は各属性間の相互フォロー決定確率である．*colinkrate* のあとに続く数字はどの属性間の相互リンクであるかを表している．有名人と活動人の間の双方向リンク率（*colinkrate12*），および一般人同士の双方向リンク率（*colinkrate33*）は実証データでは示されていない．そのため，有名人と活動人の間は有名人と一般人の間の双方向リンク率（*colinkrate13*）を用いる．

Table 4: 相互リンク率パラメータ

パラメータ	値
colinkrate11	0.45
colinkrate12	0.21
colinkrate13	0.21
colinkrate22	0.90
colinkrate23	0.75
colinkrate33	0.50

4. シミュレーション実験

4.1 シミュレーション結果と考察

小出ら[4]の研究を基に設定した，Table 3 および Table 4 のパラメータを用いて実験を行った結果と，小出らの実証データとの比較を行う．以下 Table 5 はシミュレーション 100 試行の平均と標準偏差である．

Table 5: 出力結果と実証データの比較

	平均	実証データ
平均次数	155.707	145.7
最大出次数	15,453.900	97,938
最大入次数	22,198.000	175,616
リンク密度	0.01104	0.00027
入出次数相関	0.9013	0.74
双方向リンク率	0.7182	0.60

実証データの規模はノード数約 1,000,000 ノードであり，本実験の規模はノード数 100,000 ノードである．そのため，最大出次数，最大入次数，リンク密度は実証データよりも小さい値となっている．また最大出次数，最大入次数の標準偏差が大きい．これは有名人や活動人が確率的にネットワークに加わるため，そのタイミングによってフォローされる数が異なることが影響していると考えられる．また，入出次数相関の標準偏差の値が，双方向リンク率に比べて大きい理由も同じ理由であると考えられる．入出次数相関と双方向リンク率は実証データよりも高い値が出力されている．これは，初期設定における *colinkrate13*, すなわち有名人と一般人の相互フォロー決定確率が大きいことが影響していると考える．

次に示す Fig. 5 Fig. 6 は次数分布と散布図である．なお，示す図は 100 試行中の 1 試行を取り出したものである．その他の 99 試行も同様の図が得られているため，1 試行を取り出して比較しても十分であるとみなす．

Fig. 5 は 1 試行の次数分布である．横軸が次数，縦軸がノード数となっており，四角のプロットが出次数，菱型のプロットが入次数となっている．出次数，入次数ともに次数 100 付近からベキ則が現れているのがわかる．これは Fig. 1 や Fig. 2 と比較すると，出次数分布はピークとしてベキ則から乖離する部分以外は表現できている．出次数分布のピークが表現されていないのは，提案モデルにおいてノードが次数に依存した規則を設けていないためである．入次数分布は実験結果のほうが低次のノード数が少ない結果となってしまっている．提案モデルにおいて低次の入次数が多くなる要因は，相互フォローするためのリフォローが多いことが考えられる．

Fig. 6 は 1 試行の散布図である．三角のプロットが有名人，丸のプロットが活動人，バツのプロットが一般人を表している．Fig. 3 と同様な線形に伸びているプロット，低次数における密集具合を表現できているが，入次数が大きく出次数が小さい範囲におけるプロットが表現されていない．本研究では各属性間における相互リンク率を導入している．そのため，個人個人の相互リンク率を表現しているわけではないため，ばらつきが小さい．それが影響し，特にフォロワ数に対してフォロー数が少ないノードがあまり存在しない. これは, 有名人の特徴である，フォロワ数のみ増加していき，フォロー数が少ないといったノードが少ないことが考えられる．また，フォロワ数が低次の部分でフォロー数だけ多い人も少ない。

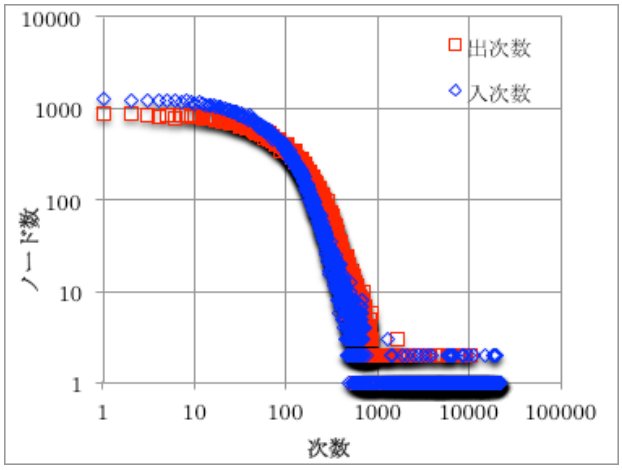

Fig. 5: 1 試行の次数分布

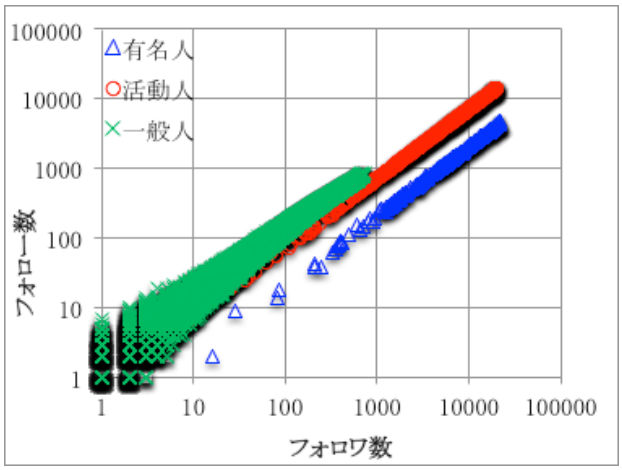

Fig. 6: 1 試行の散布図

Table 5 での出力結果と実データの比較の考察でも述べたように，本シミュレーションでは入出次数相関, 双方向リンク率共に高い値が出力されている．また，散布図も同様の傾向を得ているものの，フォロワ数のみ増加しているプロットの表現ができていない. 実際の Twitter における有名人のフォロー数を見ると，フォロワ数と非常に大きく乖離しているにも関わらず，本シミュレーションで，双方向リンク率により設定した有名人と一般人の相互リンク率 *colinkrate13* は 0.21 と高い値となっている. そこで，ネットワークの生成に大きく関わるパラメータ *createnode*（ノード生成率）の感度分析とキャリブレーションの結果を示し，有名人と一般人との間の相

互リンク率 *colinkrate13* の感度分析及びキャリブレーションを行う．そして実証データとして分析されていない *simplechoice*（一般人選択率）の感度分析を行なう．

4.2　キャリブレーションと感度分析

4.2.1　createnode

はじめに，ネットワーク成長の方向性を決定する *createnode*（ノード生成率）について感度分析とキャリブレーションを行った結果を示す．*createnode* とは 1 回モデルフローを行なう際，どのぐらいの確率でノード生成フローに移行するかを表すパラメータである．期待値の計算を行うと，$createnode = 0.01$ とは，モデルフローを 100 回行なうと，1 回はノード生成フローに移行し，ノード生成を行うということである. その他の 99 回はリンク生成フローに移行する．つまり，1 つのノードがネットワークに加わる間に, 期待値的には最低でも 99 本の新たなリンクが生成されることとなる. 最低でも 99 本というのは, 相互フォローを行う可能性があるため，実際には 99 本以上のリンクが生成されるからである．結果として，*createnode* のパラメータは平均次数に大きく関わるパラメータとなる．以下の Fig. 7 から Fig. 9 が感度分析の結果となる．すべての図で横軸は *createnode* である．

ノード生成率が高くなると平均次数は小さくなり，双方向リンク率は高くなる．しかしながら，入出次数相関に大きな影響は見られない．平均次数が小さくなるのは先も述べたように，期待値の計算を行えば導くことが可能である．双方向リンク率に影響を与えるものは，リンク生成フローに入った際に，どの属性へフォローを行うかを決定するアルゴリズムであると考えられる．有名人にフォローを行えば，双方向リンク率は小さくなる．また，活動人にフォローを行っても双方向リンク率は 0.75 であるため，この値より大きくなることは無い．すなわち，リンク生成数が増えれば，有名人や一般人といった相互リンク率が小さい属性にフォローする数も増加するため，結果として双方向リンク率が小さくなっていると考えられる．入出次数相関に大きな影響はみられないが，ばらつきが大きい結果となっている．この，ばらつきが生まれる要因は，有名人及び活動人のノードがネットワークに生成されるタイミングによるものだと考えられる．

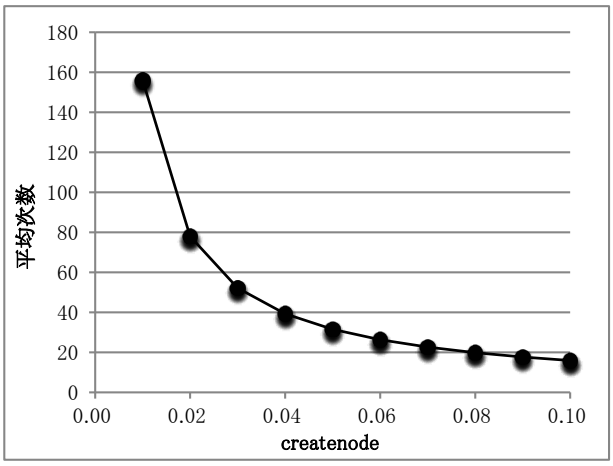

Fig. 7: 平均次数の変化

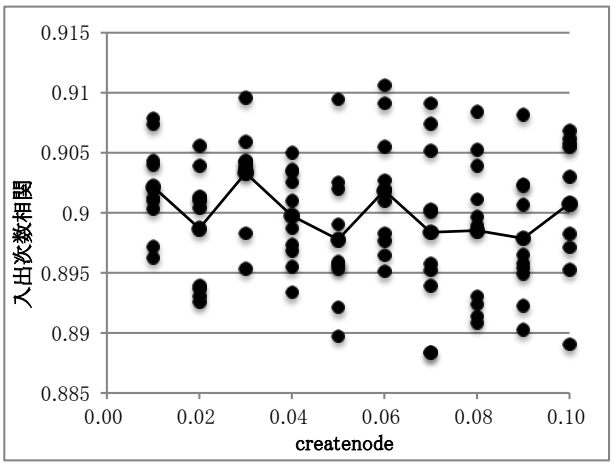

Fig. 8: 入出次数相関の変化

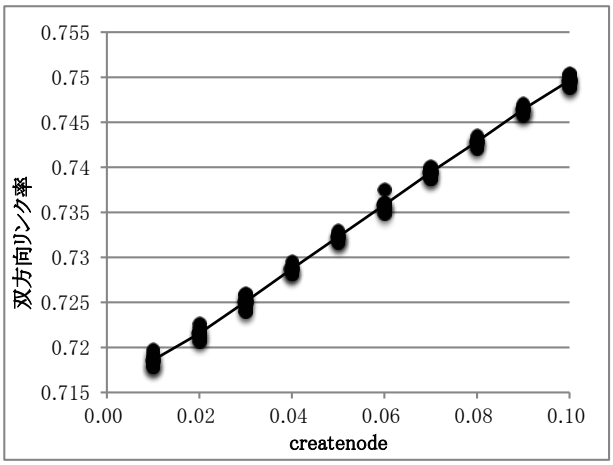

Fig. 9: 双方向リンク率の変化

4.2.2　colinkrate13

次に，相互フォロー決定確率 *colinkrate*13 について感度分析，及びキャリブレーションを行なう．*colinkrate*13 とは有名人と一般人の間の相互フォローを行なうか否かの確率となる．実証データでは双方向リンク率 0.21 となっている．しかしながら実際の Twitter で有名人を検索しそのフォロー数とフォ

ロワ数を調べると，その数に大きな乖離があることがわかる．つまり，双方向リンク率をそのまま相互リンク率として適用することが困難である．また有名人のフォロー先を見てもほとんどが同じ有名人である．Table 6 は meyoo[17]の調査結果である．

Table 6: 有名人のフォロー数，フォロワ数 [17]

有名人	フォロー数	フォロワ数
有吉弘行	190	3,954,102
松本人志	70	2,096,023
吉高由里子	4	1,856,363
ベッキー♪♯	1	1,609,213
宮迫	235	1,599,611
三村マサカズ	166	1,543,069
田村淳	418	1,395,091
宮川大輔	257	1,283,524
徳井義実	152	1,194,003
堀江貴文	416	1,166,577
バカリズム	310	876,283

以下の Fig. 10 から Fig. 12 がその結果となる．すべての図で横軸は *colinkrate13* である．感度分析を行なう範囲は小出らの実証データである双方向リンク率 0.21 より小さい値の範囲で行う．

どの統計指標も *colinkrate13* が小さくなると，小さくなっていることがわかる．有名人が一般人からフォローされた際，および一般人に有名人がフォローした際にリフォローする確率が小さくなっているため，リンク数が増えずに平均次数は小さくなり，リフォローをしなくなっているため双方向リンク率も小さくなっている．入次数と出次数の差も大きくなるため，入出次数相関も小さくなる．また，入出次数相関においては，ばらつきも大きくなっているように見られる，これは先も考察したように，有名人や活動人の生成タイミングが影響していると考えられる．実データに一番近い$colinkrate13 = 0.00$の試行から 1 試行を取り出して，入出次数分布と散布図を示す．

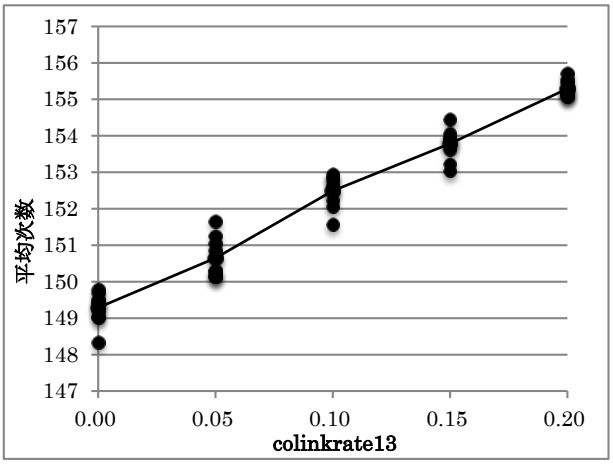

Fig. 10: 平均次数の変化

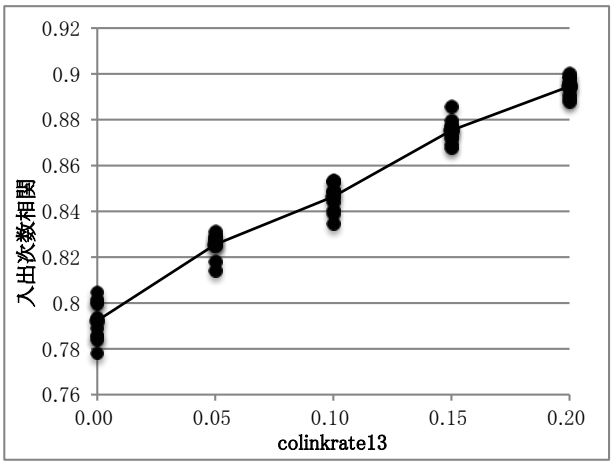

Fig. 11: 入出次数相関の変化

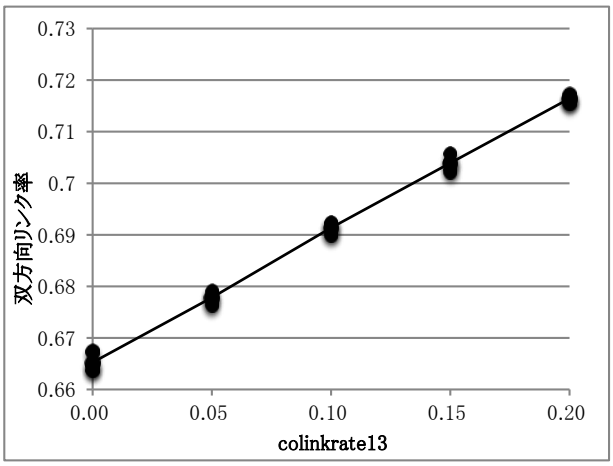

Fig. 12: 双方向リンク率の変化

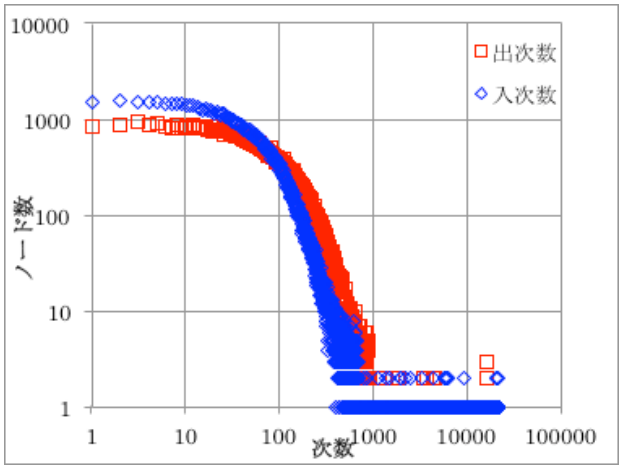

Fig. 13: colinkrate13 = 0.00 の次数分布

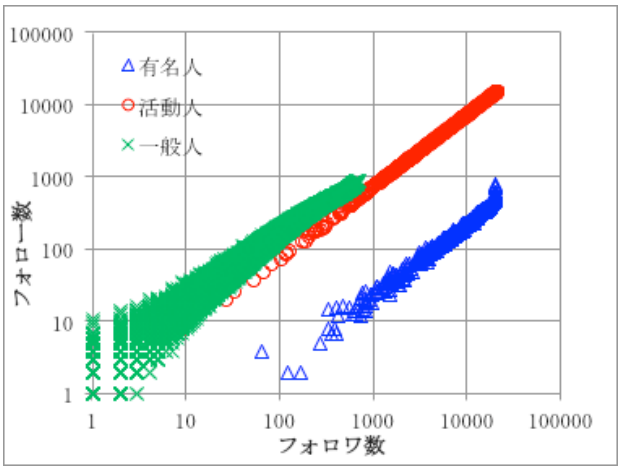

Fig. 14: colinkrate13 = 0.00 の散布図

入出次数分布は大きな変化は見られないが，散布図においては，有名人のフォロー数が減少していることが分かる．Fig. 3 の実際のデータと比較するとまだフォロー数が多い傾向があるといえる．これは，シミュレーション中のステップを経ると，それだけ有名人がフォローを行う可能性があり，それが有名人のフォロー数の増加をさせている．本アルゴリズムには有名人が能動的に行うフォローに対して制約を設けていないため，フォロワ数のみ増加させたい場合は，制約を設ける必要があると考える．

以上の感度分析から，有名人と一般人の相互リンク率は 0 と同定する。

4.2.3　一般人選択率

次に，一般人選択率（*simplechoice*）について感度分析を行う．一般人のフォロー形成に関しての考察は行われていない．しかしながら，*colinkrate*13 の感度分析とキャリブレーションの結果から，一般人選択確率の変化による統計指標の変化は大きいと考えられる．以下に，*simplechoice* を変化させたときの統計指標の変化を示す．

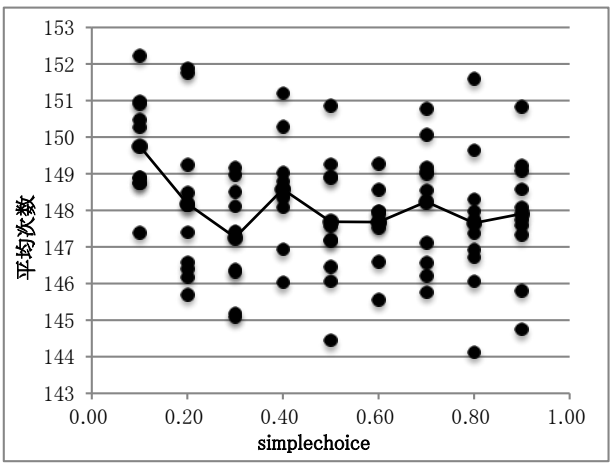

Fig. 15: 平均次数の変化

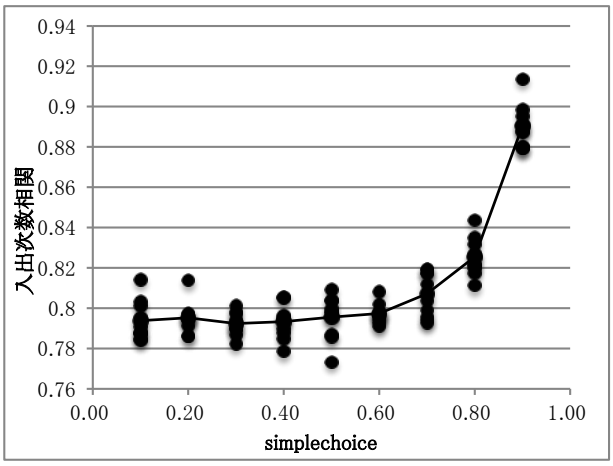

Fig. 16: 入出次数相関の変化

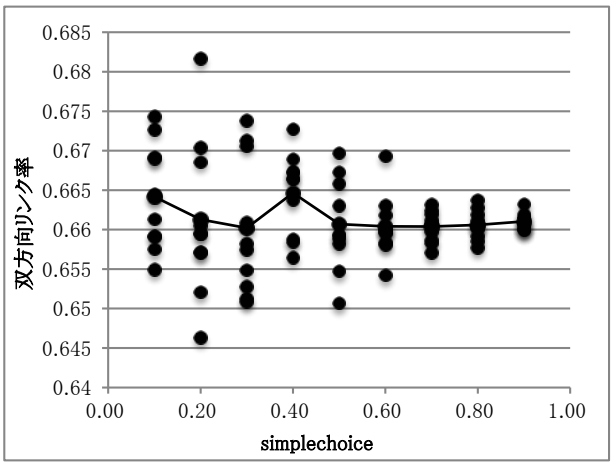

Fig. 17: 双方向リンク率の変化

平均次数はばらつきが大きく，一般人選択率を増加させると平均的にはわずかではあるが減少傾向にある．これは一般人選択確率を増加させることにより，活動人へのフォロー数が減るため，相互フォローを行う回数が減り，平均次数を減少させていると考える．入出次数相関は 0.60 を境に，急激に強い相関となっている．これは，一般人選択率が増加したことにより，フォローをされた際の相互フォローという行動ではなく，フォローをしたそのユーザからすでにフォローをされている可能性が大きくなる．これが結果として入出次数相関を強めていると考える．双方向リンク率は一般人選択率を増加させるとばらつきが小さくなっているのが分かる．これは，一般人選択率が高くなることにより，一般人同士の相互リンク率を適用する回数が多くなり，ばらつきを抑えていると考える．以上より，一般人選択率は入出次数相関, 双方向リンク率の両方に影響を与え，特に 0.60 より大きな値をとると，入出次数相関が強まるという結果を得た．

これらの感度分析の結果から，各指標の実証データからの差分を計算し和をとった値を評価値とする．

各統計指標の重みは同等としている．Fig. 5.19 がその結果であり，評価値から一般人選択率は 0.3 と同定した．

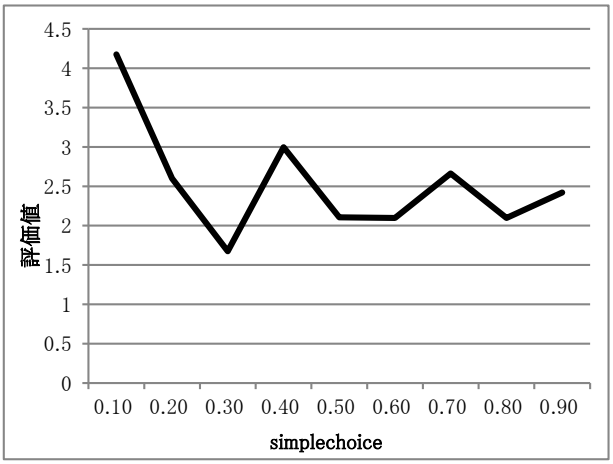

Fig. 18: 評価値の変化

4.3 最終的なパラメータ

結果から導かれた最終的なパラメータを記載する．

Table 7: 最終的なパラメータ

パラメータ	値
maxnode	100,000
createnode	0.01
famous	0.00314
semifamous	0.00686
famouschoice	0.252
semifamouschoice	0.448
simplechoice	0.3
colink11	0.45
colink12	0.21
colink13	0.00
colink22	0.90
colink23	0.75
colink33	0.50

Table 8: 最終的な出力結果（100 試行平均）

	平均	実データ
平均次数	147.2611	145.7
入出次数相関	0.7924	0.74
双方向リンク率	0.6602	0.60

平均次数，入出次数相関，双方向リンク率ともにわずかに高い値が得られた．ノード生成率は平均次数に大きな影響を与え，次数分布の形状にも大きな変化を与えることが分かった．有名人と一般人の相互リンク率は，次数分布に大きな影響を与えることはないが，各評価指標，および散布図に大きな影響を与えることが分かった．一般人選択率は平均次数に影響は与えないが，双方向リンク率のばらつきに影響を与え，入出次数相関には強い影響があることが分かった．

5. 結論

5.1 まとめ

本研究では，Twitter を対象とした実証データを基に，入出次数相関と双方向リンク率を表現するような成長ネットワークモデルの提案を行った．結果，入出次数相関と双方向リンク率はわずかに高い値をとるものの、Twitter を再現するようなネットワークモデルを得ることが出来た．また，ノード生成率，一般人と有名人の相互リンク率，一般人選択確率の感度分析を行なうことで，実証データとは異なった相互リンク率，また深く分析されていなかった一般人同士のフォロー関係について考察を加えた．

5.2 今後の課題

今後の課題として，ユーザ一人一人のリフォロー決定率，一般的な人たちのフォロー関係形成についての分析が望まれる．本研究においてユーザ一人一人のリフォロー決定率を導入していない．散布図をみると，各属性の傾向が実データと近似できるものの，ばらつきを表現できていないことが分かる．ユーザ一人一人のリフォロー決定の構造が明らかとなれば，入出次数相関，双方向リンク率を更に実証データに近似することが可能だと考えられる．また，本研究においてフォロー行動をする人，そしてその人がフォローする先というのは，フォロー先の属性を確率的に決定し，その属性内からはランダムに決定している．しかし，実際の Twitter において，フォロー行動をする人は，既にフォローしている人のフォロー・フォロワ，または自分のフォロワのフォロー・フォロワから探してフォローする可能性がある．このような三角構造についての関係性は，Cui et al.[14]によって分析がなされているが，Twitter 以外の SNS においての分析となっており，三角構造が形成される要因が明らかとされていない．この三角構造の構造原理が明らかとなれば，成長ネットワークモデルにその原理を導入することにより，Twitter の特徴を表現することが可能となる．これら特徴を表現できれば，構築したネットワーク上での情報伝播シミュレーションの分析などの精度を上げることも可能であると考えられる．

参考文献

1) 池田 圭佑, 岡田 佳之, 榊 剛史, 鳥海 不二夫, 篠田 孝祐, 風間 一洋, 野田 五十樹,

諏訪 博彦, 栗原 聡 : マルチエージェント型拡張 SIR モデルを用いた情報拡散シミュレーションの評価，情報処理学会研究報告，(2014)
2) 石井 : マイクロブログ Twitter における日本人利用者の特徴, 筑波大学研究報告書, (2011)
3) 風間 一洋 : Twitter における情報伝播，人工知能学会誌，**27**-1，35/42（2012）
4) 小出 明弘，斉藤 和巳，風間 一洋，鳥海 不二夫 : ネットワーク分析による Twitter ユーザのフォロー形成に関する一考察, 情報処理学会論文誌，**6**-2，164/173（2013）
5) 白井 嵩士，榊 剛史，鳥海 不二夫，篠田 孝祐，風間 一洋，野田 五十樹，沼尾 正行，栗原 聡 : Twitter ネットワークにおけるデマ拡散とデマ拡散防止モデルの推定, 人工知能学会研究会資料，SIG-DOCMAS-B102-6，(2012)
6) 増田 直紀，今野 紀雄 : 複雑ネットワークの科学，産業図書株式会社，(2005)
7) 三井 一平，内田 誠，白山 晋 : コミュニティ構造を有するネットワーク成長モデル, 情報処理学会研究報告，ICS，17/24，(2006)
8) 三井 一平，内田 誠，白山 晋 : ネットワーク上のようなコミュニティとクラスター構造の関係性について，情報処理学会研究報告，ICS，7/14（2006）
9) 宮崎 大樹 : SNS 上でのユーザ間のコミュニケーションに起因するネットワーク成長モデルの提案, 早稲田大学大学院修士論文, (2009)
10) 山本 雅人，小笠原 寛弥，鈴木 育男，古川 正志 : 東日本大震災時の Twitter における情報伝播ネットワーク，情報処理，**53**-11，1184/1191（2012）
11) 湯浅，白山 : 感染症流行予測におけるシミュレーションパラメータの影響分析, 情報処理学会研究報告，(2010)
12) Akioka. S，Kato. N，Muraoka. Y，Yamana. H : Cross-media Impact on Twitter in Japan, International Workshop on Search and Mining User-Generated Contents，111/118，(2010)
13) Barabasi. AL，Albert. R : Emergence of scaling in random networks，Science，509/512（1999）
14) Cui. A.X，Zhang. Z.K，Tang. M，Hui. P.M，Fu.Y ：Emergence of Scale-Free Close-Knit Friendship Structure in Online Social Networks, PLoS ONE，(2012)
15) Kwak. H，Lee. C，Park. H，Moon. S : What is Twitter, a Social Network or a News Media?, International Conference on World Wide Web, 591/600（2010）
16) Vazquez. A : Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations，Physical Review E **67**.5，(2003)
17) meyou, http://meyou.jp/ranking/follower_talent, (2015.01.06)