

# ノードに内在する関係発生能力に注目した ネットワーク生成モデル

時田 遼嗣<sup>1</sup> 武藤 敦子<sup>1</sup> 森山 甲一<sup>1</sup> 犬塚 信博<sup>1</sup>

概要：人や組織の持つ関係性をネットワークとして捉え、複雑ネットワークの特徴を考える分野として社会ネットワーク分析がある。実際の社会ネットワークは、特徴的な性質を持つことが分かっており、その性質を満たすネットワークモデルが研究されてきた。WS や BA のような代表的なモデルの多くは、個々のノードは匿名的に扱われ、構造的特徴のみに基づいてモデル化されてきた。しかし、ノードは固有のネットワーク生成の能力を有すると考えられ、これを分析することも有用である。そこで本論文では、ノード毎に内在する属性に基づいたネットワークモデルの提案を目的とする。その値をパラメータとしてネットワーク群を作成し、現実のネットワークが持つ性質を考えることでネットワークモデルを考察する。

## 1. はじめに

1960年代以降、社会学の分野において社会ネットワーク分析の研究が行われてきた。社会ネットワーク分析とは行為者の属性ではなくその関係性に着目して現象を捉えようとする方法論のことである。つまり、人や組織の持つ関係性をネットワークとして捉え、中心となる行為者や構築される派閥がどのようなものか解明する分野である。研究の対象とするネットワークとしては、人間、企業、産業、国家などの間の関係性で考えられることが多い [1]。

このような現実に存在する社会ネットワークは多様であり、巨大で複雑な構造を有しているが、一定の共通する性質を持つことがわかっている。また社会ネットワークは常に一定の状態を維持するとは限らず、時間の経過によってそのつながりを常に変化させていくものが多い。この変化を数学的なモデルにし、実際のネットワークの性質を有するモデルの研究として代表的なものが二つある。1998年に Watts と Strogatz らによって発表されたネットワークモデル (WS モデル) は極めて単純なアルゴリズムから現実のネットワークが持つ性質を満たすネットワークを生成し注目を集めた [2]。これを契機に、社会ネットワークの性質をもったネットワーク生成モデルへの関心が高まっていった。1999年には Barabasi と Albert らによって WS モデルとは異なる複雑ネットワークが持つ性質を満たすネットワークモデル (BA モデル) が発表された [3]。以後も様々な研究が進む中、近年では SNS の発展から SNS 上の友人関係に注目したネットワークの分析などを中心に、再び注

目を集めている [4]。

このような複雑ネットワークについては、中心性のような静的なものから、構造的な特徴を分析することでネットワーク全体の性質を捉える研究がある [5]。従来のネットワークモデルの多くは、こういった静的な情報から構造的な特徴を指標として変化するものとなっている。しかし、近年ネットワーク分析の中で、ネットワークの変化という動的なものから、構造を変化させる原因となるノードの特徴を分析する研究が行われており、様々な視点から複雑ネットワークの特徴や変化を分析する研究が進んでいる [6]。

従来のネットワークモデルは、ノードの特徴を考えない匿名な対象として、ネットワーク全体の構造を見るモデルが多かった。しかし、関係発生要因を分析した菅田らの研究から、ノードは固有のネットワーク生成の能力を有すると考えられ、これを分析することは有用であると考えられる。そこで本研究では、各ノードには固有の関係構築の性質が内在しているという仮定の下、個々のノードに注目したネットワークモデルを提案し、そのモデルを複雑ネットワークが持つ特徴量をもとに評価を行う。

## 2. 複雑ネットワークの性質と特徴量

モデルの比較を行う上で必要な、複雑ネットワークの性質とネットワーク特徴量について述べる。

スモールワールド性 [7], [8], [9]

ネットワーク理論においては任意のノードからノードへ、少数のノードを経由して辿り着けるネットワークのことをスモールワールド性を持つネットワークとなる。特徴量の点では、平均経路長が短く、クラスター

<sup>1</sup> 名古屋工業大学大学院工学研究科

係数 [8] が高い。

平均経路長  $L$  は次のように定義される。グラフ  $G = (V, E)$  ( $V$ : 頂点集合  $E$ : 辺集合 ( $V \times V$ )) において頂点  $i, j \in V$  の最短距離を  $L_{i,j}$  とする。この時、グラフ  $G$  の  $L$  は

$$L = \frac{1}{|V|P_2} \sum_{i \neq j \in |V|} L_{i,j}$$

となる。Watts と Strogatz らは、このスモールワールド性を満たすネットワークの生成モデルを提案している。

パラメータ  $p$  を 0 とおけば、格子となり、1 とおけばランダムグラフとなる。 $p$  を 0.1 前後とすることで、格子とランダムグラフを併せ持ったような性質のグラフが生成される。

### クラスタ性 [8]

ネットワークの内輪の繋がりやの尺度をクラスタ係数といい、グループ内のコミュニケーションの密さとグループ外とのコミュニケーションの薄さがネットワークの構造に現れる。クラスタ性を測る指標としてクラスタ係数がある。クラスタ係数はグラフの性質によって算出方法が変わるが、ここでは本研究で取り扱う重み無し無向グラフのクラスタ係数の算出法について述べる。

グラフ  $G = (V, E)$  全体のクラスタ係数  $C$  を求めるために、まずは各頂点  $v_i \in V$  ごとのクラスタ係数  $C_{v_i}$  を定義する。頂点  $v_i$  が端点となる辺集合を  $E(v_i)$  とすると、 $v_i$  の次数は  $|E(v_i)|$  となる。 $v_i$  の隣接頂点から任意の 2 頂点のペアを作った際、隣接しているペアの数を  $N$  とすると、

$$C_{v_i} = \frac{N}{|E(v_i)|C_2}$$

であり、グラフ全体のクラスタ係数  $C$  は

$$C = \frac{1}{|V|} \sum_{i=1}^{|V|} C_{v_i}$$

### スケールフリー性 [8]

スケールフリー性とは、一部のノードが多数のノードと隣接しているが大多数のノードはごく少数のノードとのみ隣接しているという性質である。例えば、友人関係ネットワークでは、クラスの人気者のようなごく少数の学生が多数の友人を持っているが、ほとんどの学生は少数の友人しか持たないということになる。

ある頂点の次数を  $k$  とする。次数  $k$  の確率分布  $P(k)$  は

$$P(k) = ak^{-\gamma}$$

に従うとされる ( $a$ : 定数)。スケールフリーネットワークの生成については、Barabasi と Albert らによって、

新規ノードの結合の際に、次数の大きいノードを優先的に選択する、という制約を持たせるモデルの提案がされており、スモールワールド性も満たしている。

## 3. 関係発生要因と指標

### 3.1 関係構築中心性

関係構築中心性 [6] はネットワーク変化から、ノード毎に存在すると考えられる、関係構築の性質を測る指標である。ノード固有である関係構築の要因は二つ考慮されている。ネットワークの変化において新しく関係が発生した 2 頂点  $X$  と  $Y$  に対して、発生前にこれらの両方と関係を持っていた頂点  $Z$  が存在していた場合、 $Z$  を関係発生要因とし、キューピッドと定義した。 $Z$  が存在しなかった場合、 $X$  と  $Y$  を関係発生要因とし、パイオニアと定義されている。

図 1 に例を示す。6 と 7 の間に発生した新たな関係に着目すると 6 と 7 には共通のノード 5 と 8 が存在する。そのため 6 と 7 の間に関係が発生したのは 5 と 8 が要因と考え、この 2 つがキューピッドとなる。次に、1 と 4 の間に発生した関係に着目すると、1 と 4 には共通のノードが存在しない。そのため、この関係は自分たちの力で発生させたと考え、この二人がパイオニアとなる。

### 3.2 キューピッド値とパイオニア値

菅田らはキューピッドとパイオニアに基づく評価値として、キューピッド値 ( $C$  値) とパイオニア値 ( $P$  値) を以下の算出方法で求めている [6]。

#### P 値と C 値の算出方法

- (1) 全ての頂点に  $P$  値と  $C$  値を与える。なお初期値はどちらも 0 とする。
- (2) ある関係が発生した 2 頂点の両方と関係を持っていた頂点の集合を  $F$  とする。 $|F| > 0$  ならば、その時  $F$  の要素であった頂点の  $C$  値に  $1/|F|$  点を加算する。 $|F| = 0$  ならば、関係が発生した 2 頂点の  $P$  値に  $1/2$  点を加算する。
- (3) 2 の操作を全ての発生した関係に対して行った最終的な  $C$  値と  $P$  値を、それぞれの頂点のキューピッド値、パイオニア値と定義する。

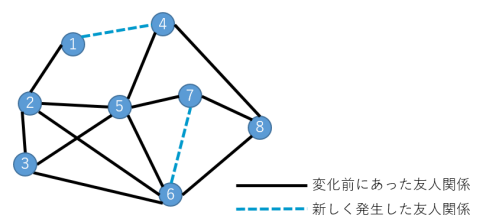


図 1 キューピッドとパイオニア

図1の場合、6,7の間に発生した関係について考えると、5,8が6,7と関係を持っている頂点となるので、5,8にそれぞれ1/2点ずつC値に加算される。1,4の場合は、両方と関係を持つ頂点は存在しないので、1,4のP値に1/2点ずつ加算される。つまり、C値が大きいほど自分の周囲のノード同士をつなげる力が強いことを指し、P値が大きいほど関係を作り出す力が強いことを表している。

#### 4. ノード固有の力に注目したネットワーク生成モデル

3章で、P,C値は、関係発生の要因の指標としていた。そこでP,C値が関係発生の要因となれば、ネットワークの変化がこの能力によって駆動される、つまり、ネットワークの生成が行えるのではないかと考えられる。この能力はノード固有のものであるため、提案するネットワークモデルは、従来のものとは異なる、匿名的ではない個々のノードの特徴に焦点をあてたモデルとなる。

##### 4.1 関係構築能力の定義とモデル

3章で述べたように、算出方法で定義された二値をノードに内在する関係構築能力として、本研究では以下のように定義する。

**P値** 「単位時間あたりに自身と関係を構築するノードの数」

**C値** 「自分と繋がるノードのP値に作用し影響を与える値」

P,C値をこのように定義するのであれば、ネットワーク変化がこの能力によって発生すると考えられる。この考えの下でネットワークを生成する具体的な手順を以下に示す。

##### 提案モデルに基づくネットワーク生成手順

###### 前提条件

- ノード数  $N$ 、辺の数が  $m$  本のランダムネットワークを作成して、それを初期ネットワークとする。
  - 各ノード  $n_i$  ( $i = 1$  から  $N$ ) に、P値  $P_i$  とC値  $C_i$  を定める。
- 各ノード  $n_i$  でそれぞれ、自身と辺が結ばれていないノードをランダムに選ぶ。そのノードを  $n_j$  とする。
  - $n_i$  と  $n_j$  との間に共通の友人がいた場合 (そのノードを  $n_k$  とする)、確率  $P_i \times P_j \times C_k$  で選んだノードとの間に辺を生成する。共通の友人がいない場合は確率  $P_i \times P_j$  で辺を生成する。
  - 2で増加した本数分の辺をランダムに減らす。
  - 1,2,3を繰り返す。

この生成モデルは、各ノードごとに自分と辺の結ばれて

いないノードをランダムに選択する。その選択されたノードとの間に、共通の友人がいなかった場合、お互いのP値の積を確率として辺を生成し、共通の友人がいた場合、それらのC値の積を対象となる二つのノードのP値の積と掛け合わせた値を確率として辺を生成する。これを全ての頂点で行った後、増加した本数分の辺をランダムに減少させる。これを指定した回数分行った時点で終了する。

#### 5. 実験と考察

提案手法で生成されるネットワークが2章で述べたネットワークの性質を持つか確認するため三つの実験を行う。

実験1では、P,C値の影響をノード毎の次数と二値の相関で確認する。実験2は、提案モデルのパラメータであるP,C値が存在するという仮定が正しければ現実に近い値が存在するはずである。そこで、その値を同定するためP,C値を変化させ特徴量を調べる。実験3では、実験2で同定したP,C値で生成したネットワークを従来モデルや現実のSNSネットワークと比較をし、提案モデルの評価を行う。

##### 5.1 実験1: P,C値の影響度実験

P,C値が値に沿った影響を与えるか見るため、二値と次数の相関を分析する。それぞれの影響度を見る際、もう片方の影響は0にするように設定した。ノード数は20前後、次数は5のネットワークで、次数の高さに比例してノードが大きく、ラベルにP,C値を表示するネットワーク図を示し、影響度を確認する。P値は0から1の値で生成し、C値は0.5と100の2つに設定した。

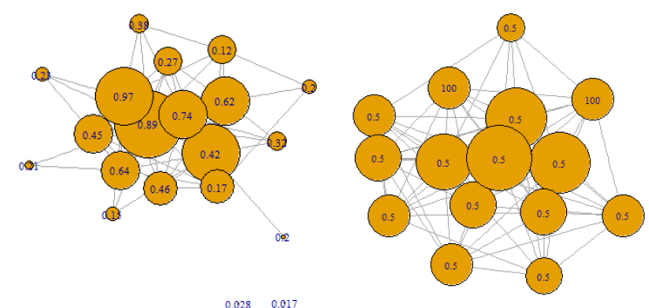


図2 影響度の結果(左P値 右C値)

##### 5.2 実験1の考察

二値の影響を考えたとき、P値が高ければ関係を作る力は強くなり、次数は高くなる。C値の場合は自分の周囲の関係を助け、周囲の次数が高くなる結果になるはずである。結果として、辺の生成は確率であるため多少のブレはあるが、P値は図5.1を見ると、中心に集まったP値の高

いノードは次数が高く、逆も成立している。C値は図5.1からP値ほどはっきりではないが、C値100の2つどちらともつながっているノードは次数が高くなっていることがわかる。

### 5.3 実験2: パラメータの分布変化実験

提案手法で生成するネットワークはパラメータであるP,C値によって異なるネットワークが生成される。そのため、パラメータを変化させネットワークの特徴量の違いからP,C値を同定する。特徴量は2章で述べた複雑ネットワークの特徴を測る指標であるクラスター係数C、平均経路長L、次数分布 $p(k)$ を用いる。パラメータ分布としては正規分布で平均を固定、分散を変化させ比較を行う。二値あるため、どちらかの分布の変化による影響を見る際は残った値は固定して実験を行う。また、ネットワークはノード数150,平均次数を6に統一して行った。

P値の影響を見る場合は、C値の影響をなくするためC値は1で固定した。P値の平均は確率で考え、中央である0.5に設定し、分散は0.01,0.2とし、P値の幅が1を超えないような分散で実験を行う。C値の場合は、P値は0.5で固定した。C値の平均は1に設定し、分散は0.1,0.5で実験を行う。値が0を下回った場合は正の値に変更している。実験結果として、ネットワーク変化1000回に対するクラスター係数C、平均経路長Lと次数分布の図を示す。

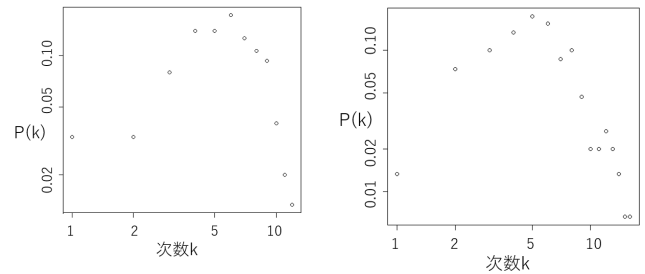


図5 P値 次数分布 (左分散0.01, 右分散0.2)

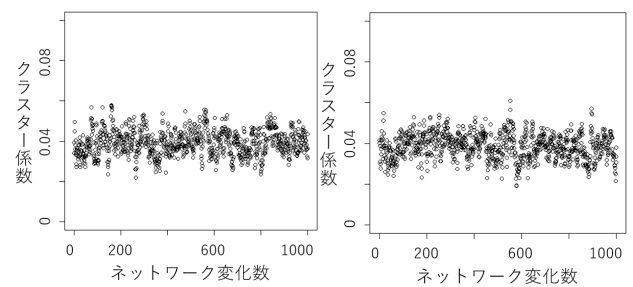


図6 C値 クラスタ係数変化 (左分散0.1, 右分散0.5)

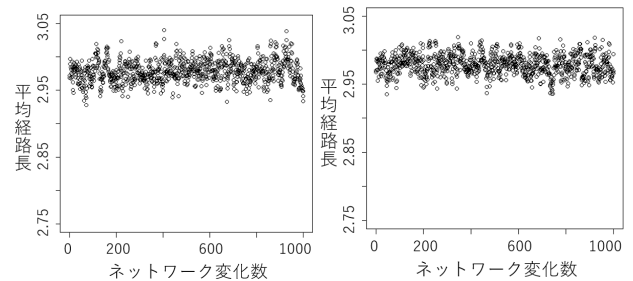


図7 C値 平均経路長変化 (左分散0.1, 右分散0.5)

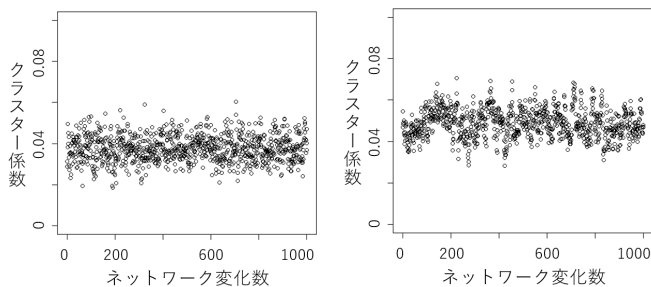


図3 P値 クラスタ係数変化 (左分散0.01, 右分散0.2)

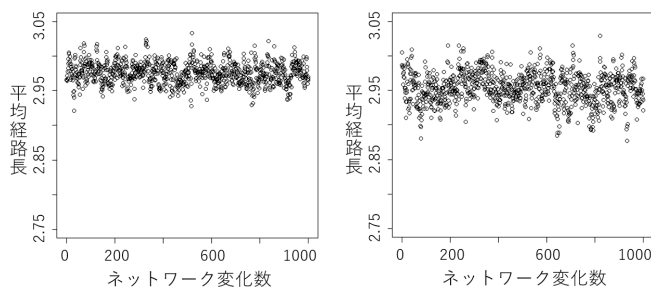


図4 P値 平均経路長変化 (左分散0.01, 右分散0.2)

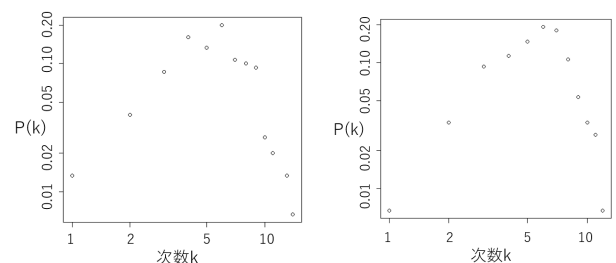


図8 C値 次数分布 (左分散0.1, 右分散0.5)

### 5.4 実験2の考察

P,C値の分散を変え、三つの特徴量を出したが、どちらも大きな差は見られなかった。P値は分散の上昇で辺を引く確率が高いノードが生まれ、わずかにクラスター係数が

高く平均経路長が低いという複雑ネットワークの特徴に近い方向へ推移した。自分と関係のあるノードに作用するC値はクラスター係数に影響すると考えられたが、ネットワーク全体はあまり傾向は見られない結果となった。

次数分布に関しては、どちらも冪分布にはならず、正規分布に近い形となった。このことから、次数分布はP,C値に相関があると考えられる。そのため、分散変化から結果に影響が出たP値を平均0分散0.5の正規分布で負の値を反転させた分布、C値は影響は少ないことから、平均1,分散0.5で、同様の実験を行った。結果は図9-図11のようになり、次数分布はP,C値に相関があるという想定に近い結果が得られた。また、クラスター係数、平均経路長においても初期の値より複雑ネットワークが持つ特徴に近づいていることを確認できた。

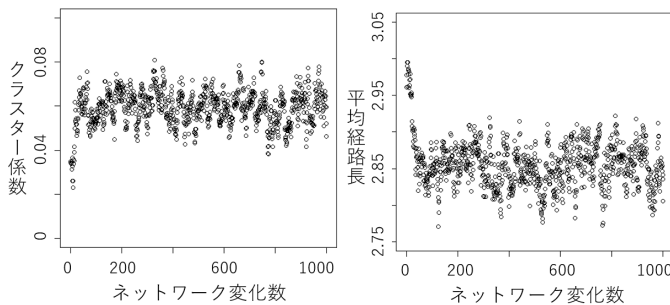


図9 同定値でのクラスター係数

図10 同定値での平均経路長

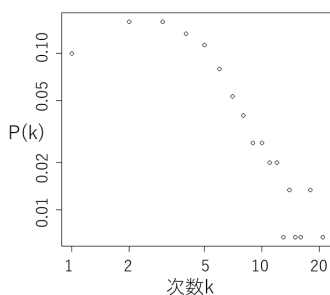


図11 同定値での次数分布

### 5.5 実験3：ネットワークの比較実験

提案手法により生成したネットワークを、従来のネットワーク（ランダムグラフ、WSモデル、BAモデル）とSNSネットワークと比較を行う。条件として、平均次数 $k$ はSNSを基準にして行う。特徴量は実験2と同様にクラスター係数 $C$ 、平均経路長 $L$ 、次数分布 $p(k)$ を用いる。

ランダムグラフ、WSモデル、BAモデル、提案手法は

生成モデルに乱数を用いているため、 $C$ と $L$ には100試行の平均値を使用している。WSモデルの張替え確率 $P$ は $C$ が大きく $L$ が小さい値を示す0.1に設定する。提案手法の $P,C$ 値は前節最後に同定した分布を採用した。実際のSNSネットワークのデータは[11]を参照した。結果は表を用いて $C$ と $L$ の100試行の平均値を示し、次に最大、最小、平均をグラフで示す。次に、次数分布を示し、最後にネットワーク図で比較を行う。

表1  $k=6$ でのクラスター係数 $C$ と平均経路長 $L$

ネットワーク	C	L
ランダムグラフ	0.037	2.99
WSモデル ( $P=0.1$ )	0.318	3.61
BAモデル	0.071	2.81
SNS	0.479	3.28
提案手法	0.064	2.85

表2  $k=10$ でのクラスター係数 $C$ と平均経路長 $L$

ネットワーク	C	L
ランダムグラフ	0.057	2.43
WSモデル ( $P=0.1$ )	0.358	2.75
BAモデル	0.128	2.39
SNS	0.43	2.99
提案手法	0.125	2.36

### 5.6 実験3の考察

比較の結果、次数6,10どちらも似たような傾向の結果が得られた。まずクラスター係数においては、提案手法はSNSやWSモデルが0.3以上の結果となる中、0.1付近と低い結果となった。これは上記でも考察した、 $C$ 値がうまく効いていないことが原因だと考えられる。

平均経路長に関しては、SNSなどに近くスモールワールド性は満たしていると考えられる結果となった。次数分布については、 $k=6$ 場合は、冪分布といえる形を成しているのに対し、 $k=10$ では次数の高い部分に関してはその兆候は見られるが低い値に対しては、スケールフリー性を判断する結果は得られなかった。これは、次数の高さに起因しており、 $P$ 値の低い関係発生能力の低いノードでも、初期ネットワークがランダムネットワークであるため、初期状態の次数が高いと、多数の $C$ 値の影響から高い次数が維持されてしまうことが考えられる。

また、ネットワーク図からわかる通り、提案手法には孤立ノードが存在している。これは減少のフェーズが原因であり、辺の生成にも選択性があるように減少のフェーズにも選択性が必要であることが考えられる。

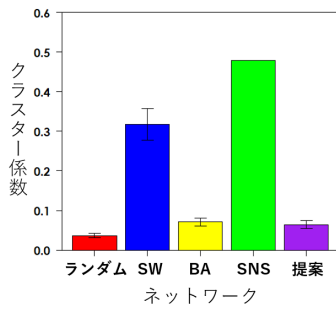


図 12 クラスター係数 k=6

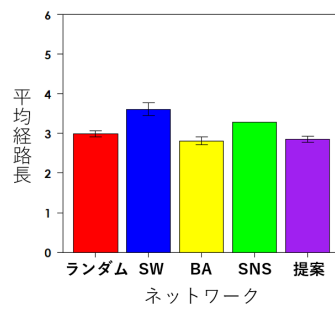


図 13 平均経路長 k=6

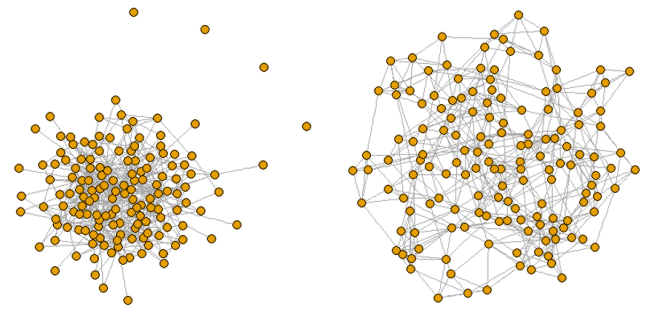


図 18 提案手法

図 19 WS モデル

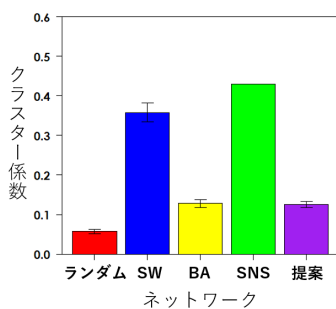


図 14 クラスター係数 k=10

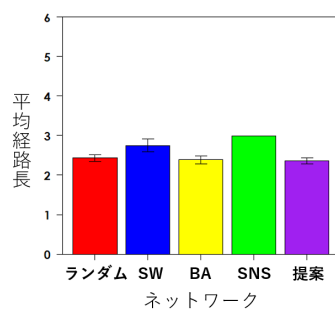


図 15 平均経路長 k=10

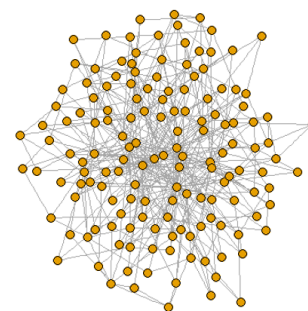


図 20 BA モデル

を持たせた変化を行うことで、より現実のネットワークに近いネットワーク生成が行われると考えられる。

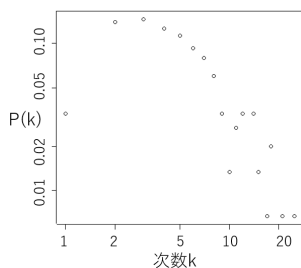


図 16 度数分布 k=6

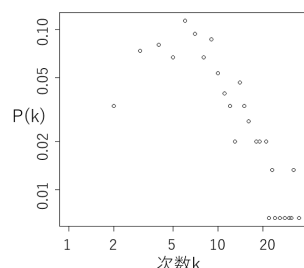


図 17 度数分布 k=10

## 6. おわりに

本研究はノード固有の関係構築の性質が内在しているという仮定で、従来の匿名的なネットワークモデルとは異なる個々のノードに注目したネットワークモデルを提案し、現実のネットワークが持つ特徴を使って従来のモデルと現実の SNS ネットワークを用いて比較を行った。

結果としてスケールフリー性、スモールワールド性を持つ、BA モデルに近いネットワークが生成される結果となった。クラスター性は C 値の同定がうまくいっておらず、影響をあまり与えられなかったことが考えられる。また、次数を上げた際のネットワークの特性や、ネットワーク図から、減少のフェーズが生成のように減少にも選択性

## 参考文献

- [1] 松尾豊. 「社会ネットワーク分析 輪読回 2005」 <http://ymatsuo.com/surveySNA/>
- [2] D.J.Watts and S.H.Strogatz, "Collective dynamics of 'small-world' networks". Nature, 393, 440-442,1998
- [3] Barabasi, A.-L., and Albert, R., Emergence of scaling in random networks, Science 286, 1999, pp. 509-512.
- [4] 内田誠, 白山晋: SNS のネットワーク構造の分析とモデル推定, 情報処理学会論文誌, Vol47, No.9, pp.2840 - 2849 (2006)
- [5] 松島裕, 他: 学生の出席記録を用いた友人ネットワークの分析, 第 24 回人工知能学会全国大会, 1G2-4, 2010
- [6] 菅田貞治, 他: 友人関係ネットワークの変化に対応した関係構築中心性の提案, MPS, (2016)
- [7] D.Watts. SmallWorlds: The Dynamics of Networks between Order and Randomness. Princeton Studies in Complexity. 1999.
- [8] S.Wasserman and K.Faust, "Social network analysis - Methods and applications". Cambridge University Press, 1994
- [9] S.Milgram, "The Small World Problem". Psychology Today, 60-67, 1967
- [10] 鈴木努. 「R で学ぶデータサイエンス 8 ネットワーク分析」 共立出版株式会社. 2009.
- [11] 鳥海不二夫, 他: SNS におけるネットワーク成長モデルの提案, 電子情報通信学会論文誌, Vol. J93-D, No.7, pp. 1135-1143 (2010)