

多目的 GA による複雑ネットワーク生成

近藤 久^{†1} 鴛崎 和宏^{†2,*1}

本論文では、2つのネットワーク特性量を評価関数とした多目的 GA による複雑ネットワークの生成手法を提案する。ネットワーク特性量には、スモールワールド性を特徴付ける平均パス長およびクラスタ係数を用い、スモールワールド性からのアプローチを行う。生成実験の結果、総ノード数に対して限られた辺数の場合にのみ複雑ネットワークの特徴が現れることを確認した。生成されたネットワークと次数分布の等しいランダム有向ネットワークや実ネットワークと比較した。その結果、実ネットワークに多く見られるネットワークモチーフが本手法によって生成されたネットワークにも多数出現すること、および故障と攻撃に対するネットワーク耐性がスケールフリー性を持った実ネットワークと同様の傾向を持つことを確認した。

Generation of Complex Networks by Using Multi-objective Genetic Algorithm

HISASHI KONDO^{†1} and KAZUHIRO TOKIZAKI^{†2,*1}

In this paper, we present a novel generating method of complex networks, generated through optimizing two evaluation functions by using multi-objective genetic algorithm. The one is the average path length of a given network while the other is the clustering coefficient of the network. It has been known that real world networks (complex networks) have considerably short average path length with respect to the number of its vertices, and have higher clustering coefficient than random graphs with a similar number of edges and vertices. Our method generates the networks with these mentioned features. From the experimental results, we found that the generated network would have a degree distribution that approximately follows a power-law (this property is called scale-free) only if the number of edges with respect to the number of vertices in the networks is limited to a value within a certain specific range. Furthermore, we showed that the generated network has many network motifs and tolerance similar to real world networks.

1. はじめに

近年、今まで何の規則性もないと考えられていた WWW のハイパーリンク構造や電力の送電網などの大規模なネットワークに、スモールワールド (*small world*) やスケールフリー (*scale-free*) などに代表される様々な特徴的構造が見つかってきている¹⁾⁻³⁾。そのようなネットワークは複雑ネットワーク (*complex network*) と呼ばれ、様々な研究が行われている⁴⁾⁻⁶⁾。

重要な研究の1つとしてネットワークのモデル化がある。実際の大規模なネットワークをモデル化することにより、ネットワークに特徴的構造が発生するメカニズムの解明やネットワーク上で起こる現象の理解に役立つ。そのため、今までに多くのモデルが提案されている^{2),3),7)-9)}。これらのモデルの多くは無向ネットワークを対象としており、有向ネットワークを扱った研究やモデルは比較的少ない¹⁰⁾⁻¹⁴⁾。しかしながら、WWW のハイパーリンク構造や学術論文の引用関係など多くのネットワークは有向ネットワークとなっている。有向ネットワークにおいてはノード *A* がノード *B* から情報や“もの”(エネルギーや資源, etc.) を得ることが、必ずしも *B* が *A* からそれらを得ることを意味しない。たとえば、上記の例に加えて食物網における捕食被食関係、多くの代謝反応、電子メールや人間関係(ある人 *A* が *B* を友人だと思っけていても、*B* は *A* を友人だと思っけていないことがある)などのネットワークには辺に向きがある。このため辺の向きはネットワークにとって重要な意味を持つ。辺に向きがあると、ネットワーク中の小さな部分グラフ(3ノード間の有向部分グラフ)であるがシステム固有の情報伝達や計算にかかわっていると示唆されているネットワークモチーフの出現¹⁵⁾ や入次数と出次数の相関関係¹⁶⁾、有向ネットワークにおけるパーコレーション¹⁷⁾ など無向ネットワークでは説明できない様々な特徴が現れる。そこで、本論文ではネットワーク特性量に着目した有向ネットワークの生成法を提案する。基礎モデルとして無向ネットワーク用に提案されている多目的 GA を用いた生成法¹⁸⁾ と類似した手法を用いる。生成したネットワークに複雑ネットワークに見られるような特性が現れるかを調べる。そして、他のモデルおよびランダム有向ネットワーク^{15),19)}、実ネットワークとの比

^{†1} 茨城大学工学部知能システム工学科

Department of Intelligent Systems Engineering, Faculty of Engineering, Ibaraki University

^{†2} 茨城大学大学院理工学研究科システム工学専攻

Graduate School of Science and Engineering, Major in Systems Engineering, Ibaraki University

*1 現在、味の素システムテクノ株式会社

Presently with Ajinomoto System Techno Corporation

較を行うことで本手法の有効性を検証する。

以下、2章ではネットワーク特性量と代表的な複雑ネットワークモデルを紹介し、3章ではGAを用いたネットワーク生成手法を述べる。4章でどのような条件を与えたときに複雑ネットワークに見られる特徴を持ったネットワークが生成されるかを実験的に検証し、5章で得られたネットワークをより詳細に解析する。6章はまとめと今後の課題である。

2. ネットワーク特性量と複雑ネットワークモデル

2.1 次数分布

次数 (*degree*) とは、各ノードが持つ辺の本数のことである。各ノードの次数を k としたネットワークに対する次数分布 $P(k)$ はランダムネットワーク (*random network*)²⁰⁾ において次式に示すポアソン分布に従う。ここで、 $\langle k \rangle$ は平均次数である。

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \quad (1)$$

複雑ネットワークではランダムネットワークとは異なり、その次数分布 $P(k)$ は次式のよ
うなべき則 (*power-law*) に従う。

$$P(k) \propto k^{-\gamma} \quad (2)$$

式(2)のべき指数 γ は2~3程度の値を持つことが知られている⁴⁾。この性質は、次数に平均的なサイズがないことからスケールフリーと呼ばれる。

有向ネットワークでは各ノードに入る辺数を入次数 (*in-degree*)、ノードから出ていく辺数を出次数 (*out-degree*) という。WWWのハイパーリンク構造では、入次数、出次数の次数分布がともにべき則に従う傾向にある^{3),6)}。

2.2 平均パス長とクラスタ係数

平均パス長 (*average shortest-path length*) とは、ネットワークにおける任意の2ノード間の最短パスの平均値のことである。複雑ネットワークでは、総ノード数に対して平均パス長が短い傾向にある⁴⁾。ノード i からノード j への最短パスを D_{ij} とすると無向ネットワークに対して、平均パス長 D は次式で表される。ここで、 N は総ノード数である。

$$D = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N D_{ij} \quad (i \neq j) \quad (3)$$

本論文では、有向ネットワークを考えるのでパス長を求める際に辺の向きを考慮する。無向ネットワークではパスがあっても、辺の向きを考慮するとパスが存在しない場合(弱連

結)が存在する。ノード i からノード j へのパスがない場合 $D_{ij} = \infty$ となってしまうネットワーク全体の平均パス長 D を求めることができない。そこで、 D_{ij} の逆数 $1/D_{ij}$ をとることでパスの存在しないノード間の距離を0と見なせる。すべてのノードに対して $1/D_{ij}$ を求め、その平均をとったものが D^{-1} であり、次式のように定義される²²⁾。

$$D^{-1} = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N \left(\frac{1}{D_{ij}} \right) \quad (i \neq j) \quad (4)$$

式(4)から求められる D (調和平均パス長) を本論文での平均パス長 D とする。

クラスタとは、ネットワーク中に現れるノードと辺で作られる三角形のことである。ネットワーク中のノードと辺がどれだけ三角形を形成しているかを示すための指標をクラスタ係数 (*clustering coefficient*)²⁾ と呼ぶ。クラスタがネットワーク中に多ければ多いほどそのネットワークは密であるといえる。たとえば、人をノード、友人関係を辺で表現したネットワークを考える。このネットワーク中にクラスタができるということは、ある人Aに友人Bと友人Cがいたとき、友人Bと友人Cもまた友人どうしであるということであり、クラスタ係数が高ければ親密な友人関係が築かれていることを意味する。複雑ネットワークはクラスタ係数がランダムネットワークに対して非常に高いという特徴を持っている⁴⁾。

本論文では、クラスタ係数を計算する際に辺の向きは考慮せず有向ネットワークを無向ネットワークとして扱う。辺の向きを考慮したクラスタ係数の定義も提案されているが、無向の場合と異なりその一般化は容易ではない²¹⁾。ノードの入次数と出次数を考慮するように無向の場合の定義を拡張することが自然であるが、三角形を閉じる辺が双方向にある場合にどちら向きの辺を用いるのかを考慮しなければならないなどの問題がある。このため有向ネットワークを対象とする場合も辺の向きを無視して無向ネットワークのクラスタ係数の定義を流用することが一般的である。有向な複雑ネットワークの場合に、無向のクラスタ係数の定義を用いても同規模のランダム有向ネットワークと比較してその値は非常に高い。これらの理由から本論文でもクラスタ係数には無向ネットワーク用の定義を用いる。

ノード i に対するクラスタ係数 C_i は次式で求められる。ただし、 k_i はノード i に隣接しているノードの個数(ノード i の次数)、 E_i は k_i 個のノード間に存在する辺の本数(三角形の個数)である。

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \quad (5)$$

この C_i をすべてのノードに対して計算し、その総和を N で割ったものがネットワーク

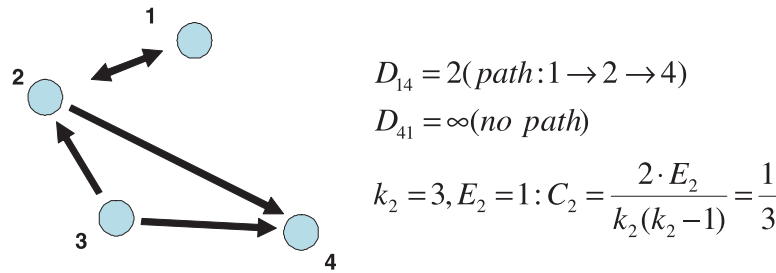


図 1 パス長とクラスタ係数の計算例

Fig. 1 Examples of the shortest-path length and the clustering coefficient.

のクラスタ係数 C であり次式のように定義される。

$$C = \frac{1}{N} \sum_{i=1}^N C_i \tag{6}$$

有向ネットワークに対するパス長とクラスタ係数の計算例を図 1 に示す。例としてノード 1 からノード 4 へのパス長 D_{14} とノード 4 からノード 1 へのパス長 D_{41} を求める。 D_{14} は $1 \rightarrow 2 \rightarrow 4$ という経路が最短であるので 2 となる。それに対して、 D_{41} は存在しないので ∞ である。ノード 2 に対するクラスタ係数 C_2 はノード 2 と隣接しているノードが 3 個で、そのノード間に存在する辺数 E_2 は 1 本であるので式 (5) より $1/3$ となる。

短い平均パス長と高いクラスタ係数から特徴付けられる性質がネットワークのスマールワールド性である。

2.3 ネットワークモチーフ

ネットワークモチーフ (*network motif*) は、ネットワーク中に現れる部分グラフのことである。たとえば、ノード数が 3 つの小さなモチーフでも辺に方向性があるならば 13 種類のモチーフが存在する¹⁵⁾。その一部の例を図 2 に示す。それぞれのモチーフは小さい部分グラフながらもシステム固有の情報伝達や計算にかかわっていると示唆されており、ランダム有向ネットワークにはほとんど出現しない¹⁵⁾。

2.4 複雑ネットワークモデル

複雑ネットワークのモデルは大きく 2 種類に分けられる。1 つは、ノードや辺の総数が生成過程において変化し、ネットワークが成長するモデルである。もう 1 つは、ノードや辺の総数が生成過程において変化せず、辺の張替えのみを行う成長しないモデルである。以下で

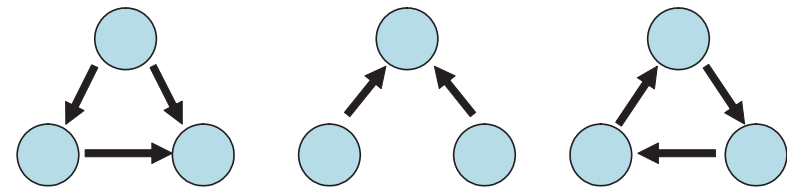


図 2 ネットワークモチーフの例

Fig. 2 Examples of network motifs.

は、成長するネットワークと成長しないネットワークの代表的なモデルをそれぞれ紹介する。

- BA モデル

Barabási-Albert モデル (BA モデル)³⁾ は、優先的選択 (*preferential attachment*) とネットワークの成長 (*growth*) から構成されるモデルである。時間とともにノードがネットワーク中に追加され成長する。その過程において、次数の大きなノードほど新たな辺を集めやすい。富めるものほど富むというダイナミクスを実現しており、その結果として次数分布にベキ則が現れる。このときのベキ指数 γ は $\gamma = 3$ となることが解析的に示されている。BA モデルでは、優先的選択とネットワークの成長によってスケールフリー性が現れることを説明している。

- PRP モデル

Preferential Rewiring Process モデル (PRP モデル)^{7),23)} は、ノードおよび辺の総数が変化しない成長しないモデルである。このモデルの特徴は、はじめに各ノードに適応度を割り振り、その適応度が高くかつ次数の高いノードほど優先的に辺を集めるように辺の張替えを行う。そして次数分布が変化しなくなるまで張替えを行うというモデルである。各ノードの適応度の決め方が重要で、適応度にランダム性があるときにベキ則が現れる。PRP モデルでは、成長しないネットワークにおいて優先的選択と適応度のランダム性からスケールフリー性が現れることを説明している。

3. GA を用いたネットワークの生成法

本章では GA を用いたネットワーク生成手法について述べる。まず、既存研究¹⁸⁾ と本論文手法の違いを説明し、GA を適用するためのネットワークのコーディング方法、個体選択および遺伝的操作 (交叉, 突然変異) について説明する。生成するネットワークは自己ループおよび多重辺を持たない有向ネットワークであり、総ノード数および総辺数は変化し

ない(成長しない).

3.1 既存研究 18) との相違

花田ら¹⁸⁾は平均パス長とクラスタ係数を評価関数とした多目的 GA による無向ネットワークの生成方法を提案した. 本論文手法と同様な点は, ネットワーク生成過程において総ノード数および総辺数が変化しない(成長しない)点および GA を用いて平均パス長とクラスタ係数からアプローチする点である. 以下, 本論文では花田らの手法を HGA 法と呼ぶ.

HGA 法ではノードの配置を TSPLIB から引用した巡回セールスマン問題を用いてネットワーク構築を行っている. 各ノードには座標が割り当てられ, 平均パス長を計算する際にノード間の 2 次元平面上の距離 (Euclidean distance) を考慮している(文献 18)では平均最短移動距離と呼んでいる). ただし, 対象とするネットワークは無向ネットワークであるので辺の向きは考慮していない. それに対して, 本論文手法では各ノードに座標を与えず, ノード間の 2 次元平面上の距離は考えない. ただし, 有向ネットワークを対象としているので平均パス長を計算する際に辺の向きを考慮してネットワークの構築を行っている.

HGA 法と本論文手法との大きな相違点は, 各ノードに座標を割り当てるか否か, 平均パス長を計算する際にノード間の 2 次元平面上の距離を考慮するか否か, および辺が無向か有向かにある.

3.2 ネットワークのコーディング

ネットワークは隣接行列を用いて表現することができる. 総ノード数 N のネットワークに対する隣接行列 A の i 行 j 列の成分を a_{ij} ($1 \leq i, j \leq N$) で表すものとする. ノード i からノード j に向かって辺があるならば $a_{ij} = 1$, さもなくば $a_{ij} = 0$ とする.

ネットワークをコーディングするために, A の各成分 a_{ij} ($i \neq j$) を変換し, 新たに行列 B を生成する. このとき, B の各成分 b_{ij} を式 (7) によって定める.

$$b_{ij} = \begin{cases} 1 & (a_{ij} = 1, a_{ji} = 0) \\ -1 & (a_{ij} = 0, a_{ji} = 1) \\ 2 & (a_{ij} = a_{ji} = 1) \\ 0 & (a_{ij} = a_{ji} = 0) \end{cases} \quad (7)$$

行列 B では, ノード i からノード j に向かって辺が出ていく場合 $b_{ij} = 1$, ノード j からノード i に向かって辺が出ていく場合 $b_{ij} = -1$, そして双方向に辺が存在する場合 $b_{ij} = 2$ と表される. 行列 B の対角成分より上の三角形の部分(点線部分)を 1 行目から順番に 1 列に並べる. このようにすると, $N(N-1)/2$ の長さで有向ネットワークを表現することができる. これ

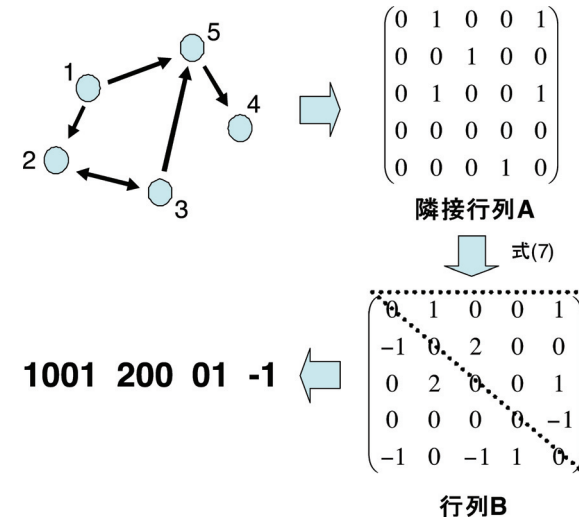


図 3 コーディング例
Fig. 3 An example of chromosome encoding.

を染色体としネットワークのコーディングを行う.

巡回セールスマン問題やロジスティクスネットワーク問題に代表される様々なネットワーク問題に対する複数のコーディング方法が提案されている²⁴⁾. 隣接行列を直接コーディングする方法はネットワークが疎な場合には無駄が多いが, 本論文では定義の簡潔さ, その後の GA 操作の容易さとその計算効率から上記で述べた隣接行列に基づくコーディング方法を用いた.

$N = 5$ のネットワークに対するコーディング例を図 3 に示す. 例では, 与えられたネットワークに対する隣接行列 A を求め, 式 (7) によって新たに行列 B を定めている. 行列 B の対角成分より上の三角形(点線部分)で有向ネットワークが表現できるので, 三角形内の成分を 1 行目から順に 1 列に並べ, 染色体 100120001-1 を得る.

3.3 選 択

親となるペアを選択するためにルーレット選択を用いる. ルーレット選択は個体の評価に応じて選択を行う戦略である. 個体 i に対する評価を F_i とすると, 総個体数 N のとき個体 i の選択確率 P_i は次式で表される.

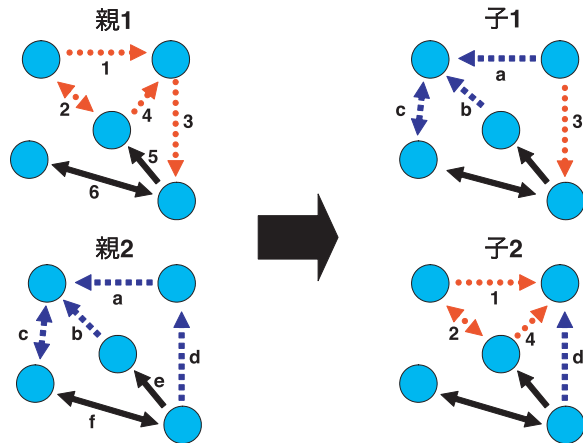


図 4 交叉例
Fig. 4 An example of crossover.

$$P_i = \frac{F_i}{\sum_{j=1}^N F_j} \quad (8)$$

ルーレット選択は評価が高ければ高いほどその個体が選ばれやすいが、評価の低い個体も選ばれる確率が 0 ではないため多様な個体を選ばれる特徴がある。本手法ではエリート保存戦略も用いている。具体的には、現世代で評価値の最も高いエリート個体を次世代の最低評価値の個体と入れ換えることによって行う。

3.4 交 叉

選択によって選ばれた個体に対する交叉の適用方法は以下のとおりとした。

- (1) 親となるネットワークのペアを比較し、辺の向きおよび辺が張られている箇所が等しいものと、それ以外に分ける。
- (2) 辺の向きおよび辺が張られている箇所が等しい辺は固定し、それ以外の辺はランダムに親から引き継がせ子を生成する。ただし、ランダムに引き継がせる際に辺の総数が変化しないように交叉させる本数を決定する。

交叉例を図 4 に示す。左側が親のペア、右側が新たに生まれた子のペアである。実線が共通の辺、点線がそれ以外の辺を表している。例の場合、親 1 と親 2 において辺 5、6 と辺 e、f がそれぞれ共通であるので形質遺伝として子にそのまま引き継がれる。親どうして共通で

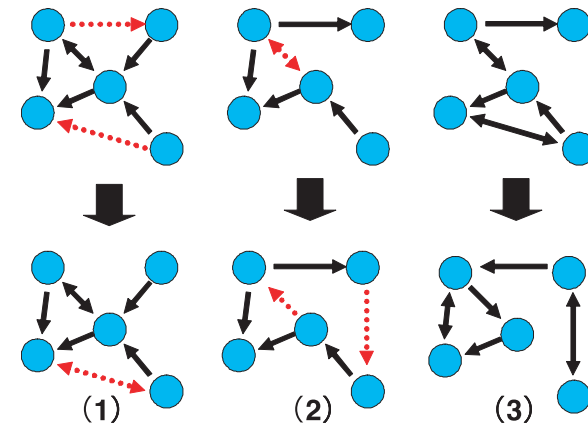


図 5 突然変異例
Fig. 5 Examples of mutation.

ない辺はランダムに親から子に継承するので、子 1 は親 1 から辺 3 を親 2 から辺 a、b、c をそれぞれ引き継いでいる。また、子 2 は親 1 から辺 1、2、4 を親 2 から辺 d をそれぞれ引き継いでいる。

3.5 突 然 変 異

突然変異は解の多様性を維持するために用いられる。本提案手法では、以下の 3 つの方法から適用可能なものをランダムに選び、いずれかを発生させる。

- (1) 片方向の辺を 2 本選び、一方を双方向に、他方を削除する。
- (2) 双方向の辺を選び、それを片方向にする。そして、ランダムに 1 本の辺をネットワークに追加する。
- (3) ランダムに辺の張替えをする。

それぞれの具体例を図 5 に示す。(1) の場合、点線で示す片方向の辺が 2 本選ばれ、それらが 1 本にまとめられている。(2) の場合、点線で示す双方向の辺が選ばれ、それが 2 本に分けられている。(3) の場合では辺をばらばらに張り替えて新たなネットワークを生成している。

4. ネットワーク生成実験

本章では、(1) 平均パス長の最小化 (単一目的 GA)、(2) クラスタ係数の最大化 (単一目

表 1 パラメータ設定

Table 1 Parameters in the experiment 1.

総ノード数	300
総辺数	900
ネットワーク数	100
世代交代回数	2,000
交叉確率	1.0
突然変異確率	0.1

表 2 各ネットワークに対する平均パス長 D およびクラスタ係数 C

Table 2 The average shortest-path length D and the clustering coefficient C for each networks.

	D	C
D の最小化	5.11	0.018
C の最大化	9.66	0.174

的 GA), (3) 平均パス長とクラスタ係数の最適化 (多目的 GA) に対するネットワーク生成を行い, 生成されたネットワークに複雑ネットワークに見られるような特性が現れるのか調べる.

4.1 生成実験 ((1) と (2))

総ノード数は 300 とし, 平均次数 $\langle k \rangle$ が 6 となるように総辺数を 900 と設定した. GA を適用するためのパラメータを表 1 に示す. 初期集団は, ランダム有向ネットワーク 100 個である. 計算機環境は IntelPentium(R)4 (2.80 GHz), メモリ 1GB の PC 上で C 言語を用いて実験を行った. 解を得るまでに約 36 時間を要した. 平均パス長 D の最小化を目的とした場合とクラスタ係数 C の最大化を目的とした場合に生成されたネットワークの D と C を表 2 に示す. ネットワーク生成過程における D と C の平均値の推移をそれぞれ図 6 および図 7 に示す.

図 6 から, D の最小化を目的とすると, D も C も初期ネットワークであるランダム有向ネットワークからほとんど変化が見られなかったことが分かる. これは, 次数分布に変化を与える力がネットワーク生成過程において働かないためである. 生成したネットワークの次数分布が初期ネットワークとほとんど変わらず, ハブと呼ばれるような次数の高いノードが発生しなかった. つまり, ほとんどのノードは平均次数に近い次数を持っており, その状態のまま D を短くするのは難しいことが分かった.

HGA 法では, D のみからアプローチしても短い D と高い C を持つネットワークが生成される. これは, D を計算する際にノード間の 2 次元平面上の距離を考慮しているので, D

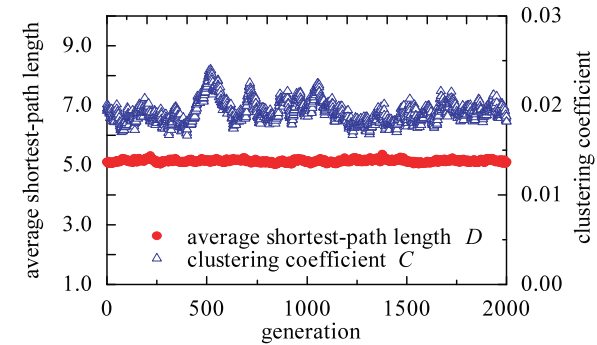


図 6 平均パス長 D とクラスタ係数 C の推移 (D のみ最適化)

Fig. 6 The average shortest-path length D and the clustering coefficient C when only D is optimized.

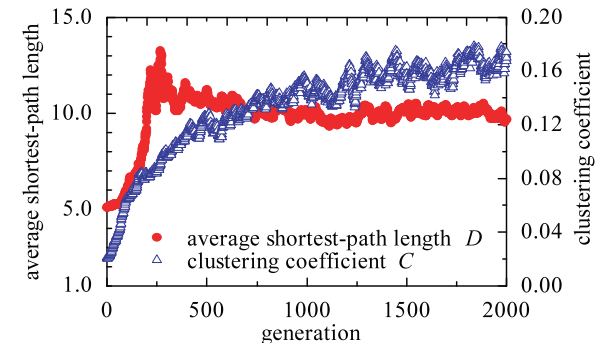


図 7 平均パス長 D とクラスタ係数 C の推移 (C のみ最適化)

Fig. 7 The average shortest-path length D and the clustering coefficient C when only C is optimized.

の最適化が進むと比較的距離の近いノードどうしがつながることによってクラスタが多く発生し, C が高くなるからである.

図 7 から C の最大化を目的とすると C は徐々に高くすることができるが, D が短くならず逆に長くなってしまっていることが分かる. 得られたネットワークは初期ネットワークと比較して C が約 8.6 倍, D が約 1.9 倍であった. これは, D の最小化を目的とした場合

とは異なり、クラスタ係数を高めるために平均次数の 2~3 倍程度の次数を持つ小さなハブのようなノードが発生するが、辺の方向性を考慮していないので、それらのノードを経由する短い有向パスが効果的に生成されず、 D が長くなってしまふからである。ゆえに、 C を高めるアプローチをとるとネットワーク中にハブと考えられるノードが発生するが、 D を短くできないことが分かった。

HGA 法では、 C のみからアプローチすると D には変化が見られないが C が徐々に高くなる傾向にある。 C を高めることによりネットワーク中にハブと考えられるノードが発生するが、2 次元平面上で距離が近いノードどうしがつながらないので D が短くならない。

これらのことから、ネットワーク生成過程において、本手法では辺の向きが、HGA 法ではノード間の距離がネットワークの生成に重要な役割を果たすと考えられる。

4.2 生成実験 2 (3)

実験 1 の結果をふまえ、平均パス長 D とクラスタ係数 C の両方を評価関数とした多目的 GA によりネットワーク生成を行う。 D の最小化 ($1/D$ の最大化) と C の最大化が目的であるので、これらを同時に最適化しなければならない。そこで、パレートランキング法²⁵⁾ を用いてネットワークを評価する。

評価関数を $f_{1/D}(\cdot)$ と $f_C(\cdot)$ の 2 つとして、双方の最大化を目的とする。ネットワーク s と t が与えられたとき、「 $f_{1/D}(s) > f_{1/D}(t)$ かつ $f_C(s) \geq f_C(t)$ 」または「 $f_{1/D}(s) \geq f_{1/D}(t)$ かつ $f_C(s) > f_C(t)$ 」の場合に、 s は t よりも優れているという。つまり、「 s の平均パス長が t の平均パス長よりも厳格に短く、かつ s のクラスタ係数が t のクラスタ係数以上である」または「 s の平均パス長が t の平均パス長以下で、かつ s のクラスタ係数が t のクラスタ係数よりも厳格に大きい」場合に s は t よりも優れている。パレートランキング法では、 n_i 個のネットワークがネットワーク i よりも優れているとき、ネットワーク i のランク R_i は次式で求められる。

$$R_i = 1 + n_i \quad (9)$$

式 (9) を用いてすべてのネットワークにランク付けを行い、この R_i の最小化を目的とする。総ネットワーク数が 7 で縦軸をクラスタ係数 C 、横軸を平均パス長の逆数 $1/D$ とした場合の例を図 8 に示す。各点がネットワークであり、その点に付いている番号がネットワークのランクを表している。このランクを目的関数としてルーレット選択により親となるペアの選択を行う。

多目的 GA を用いて、総ノード数を 100、平均次数を 4 とした場合に生成されたネットワークの例を図 9 に示す。このネットワークは平均パス長 $D = 3.88$ 、クラスタ係数 $C = 0.49$

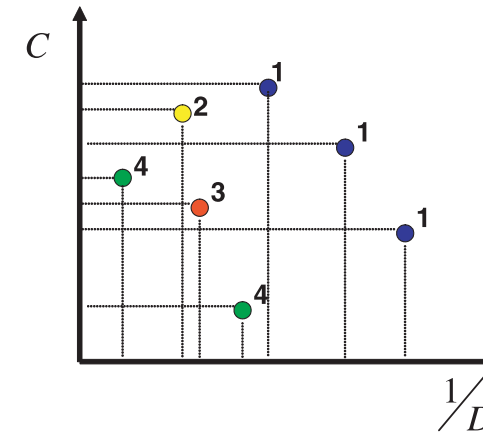


図 8 パレートランキング法の例
Fig. 8 An example of Pareto-ranking method.

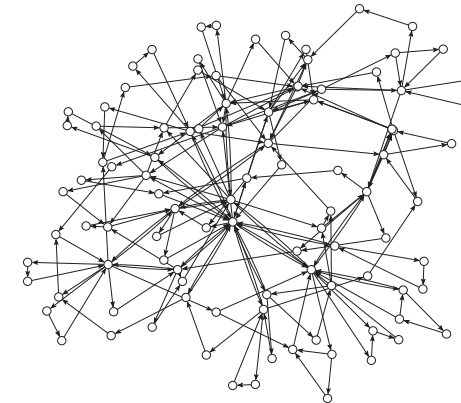


図 9 ネットワークの生成例
Fig. 9 An example of generated network by using multi-objective GA.

であり、短い D と高い C の両方の性質を満たしている。ほとんどのノードは数本の辺しか持たないが、ハブと呼ばれるような多くの辺を持つノードもわずかながら存在することが分かる。以下では、総ノード数を 300 とした場合に生成されたネットワークの特徴をより詳

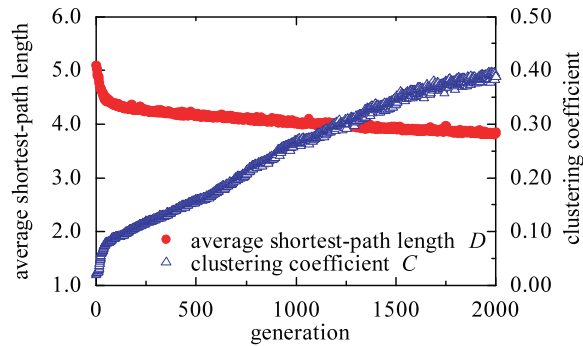


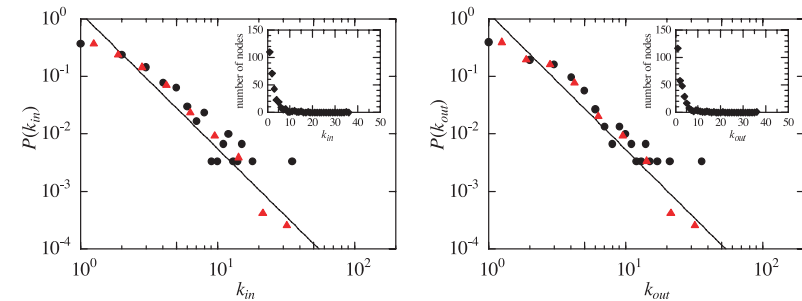
図 10 平均パス長 D とクラスタ係数 C の推移 (多目的 GA 適用時)

Fig. 10 The average shortest-path length D and the clustering coefficient C when multi-objective GA is used.

しく述べる。

多目的 GA を適用するためのパラメータ、総ノード数および総辺数は実験 1 と同様に設定した (表 1)。計算機環境も実験 1 と同様である。解を得るまでに約 72 時間を要した。最終的に得られたネットワークに対する平均パス長 D およびクラスタ係数 C はそれぞれ $D = 3.86$, $C = 0.39$ となった。初期ネットワークと比較して D は約 0.76 倍, C は約 19.8 倍である。

ネットワークの生成過程における D と C の平均値の推移を図 10 に示す。多目的 GA により D と C を同時に最適化することによって、短い D と高い C を持つネットワークが生成されていることが分かる。横軸に入次数 k_{in} , 出次数 k_{out} , 縦軸に入次数分布 $P(k_{in})$, 出次数分布 $P(k_{out})$ とした両対数グラフをそれぞれ図 11 (a), (b) に示す。グラフ内部の小さなグラフは横軸を次数 $k_{in/out}$, 縦軸をノードの出現頻度としたグラフである。赤いマーカは対数スケールで次数の区間長が均一になるように調整し、各区間に含まれる次数の平均値をプロットしたものである。データ数が少ないながらも入, 出次数分布ともにベキ則的傾向を示している。近似直線のベキ指数は $\gamma_{in} = \gamma_{out} = 2.35$ である。この結果より、生成されたネットワークは複雑ネットワークに見られるような短い平均パス長と高いクラスタ係数で特徴付けられるスモールワールド性および次数分布のベキ則で特徴付けられるスケールフリー性に近い性質を持つことが分かる。平均パス長とクラスタ係数の両方からアプローチすることでハブが発生し、そのハブを利用するような有向パスが生成される。その結果とし



(a) 入次数分布

(b) 出次数分布

図 11 $\langle k \rangle = 6$ の場合の入次数分布と出次数分布

Fig. 11 The in- and out-degree distributions ($\langle k \rangle = 6$).

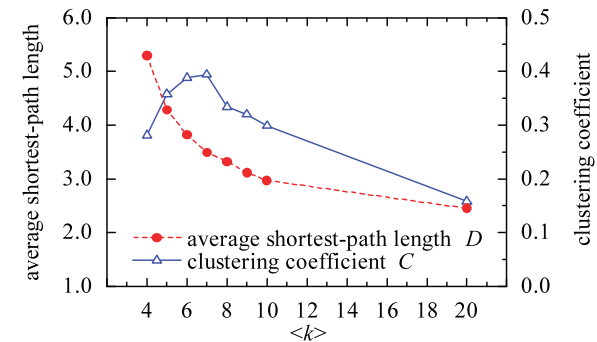


図 12 平均次数 $\langle k \rangle$ に対する平均パス長 D とクラスタ係数 C の推移

Fig. 12 The average shortest-path length D and the clustering coefficient C as a function of mean degree $\langle k \rangle$.

て、入, 出次数分布それぞれにベキ則的傾向が現れたと考えられる。

HGA 法ではノードに座標があり、ノード間の距離が短く、かつクラスタ係数が高い無向ネットワークを構築している (ノード数 101, 辺数 400 の無向ネットワーク)。その結果として比較的高い次数を持つハブとなるノードが生成されたことを報告している¹⁸⁾。ただし、本論文手法とは少なくとも評価関数および辺が無向であることが異なり、得られたネットワークを直接比較することは難しい。

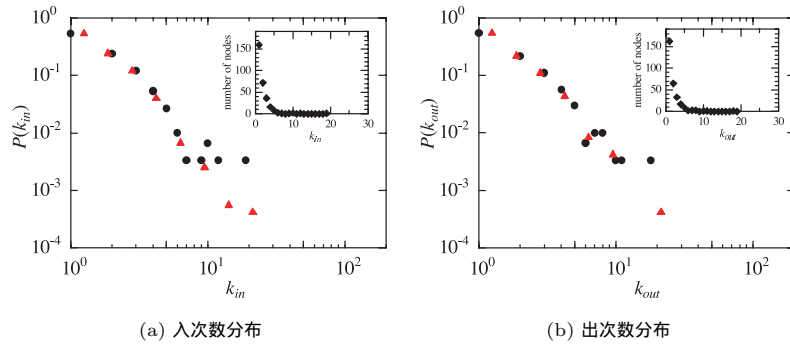


図 13 $\langle k \rangle = 4$ の場合の入次数分布と出次数分布
Fig. 13 The in- and out-degree distributions ($\langle k \rangle = 4$).

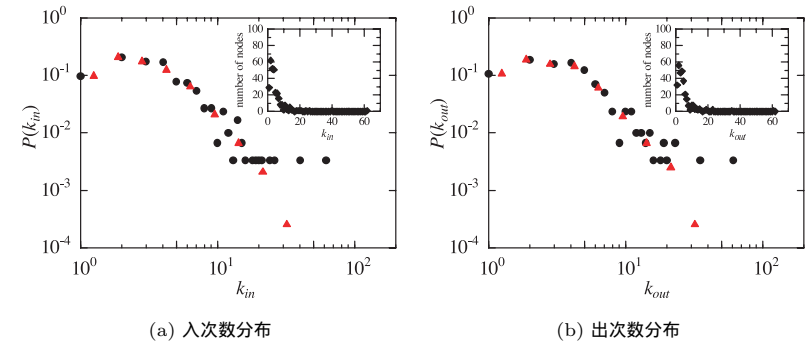


図 15 $\langle k \rangle = 10$ の場合の入次数分布と出次数分布
Fig. 15 The in- and out-degree distributions ($\langle k \rangle = 10$).

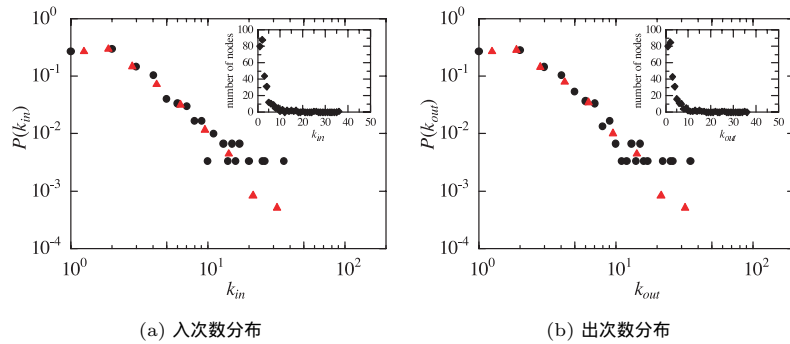


図 14 $\langle k \rangle = 7$ の場合の入次数分布と出次数分布
Fig. 14 The in- and out-degree distributions ($\langle k \rangle = 7$).

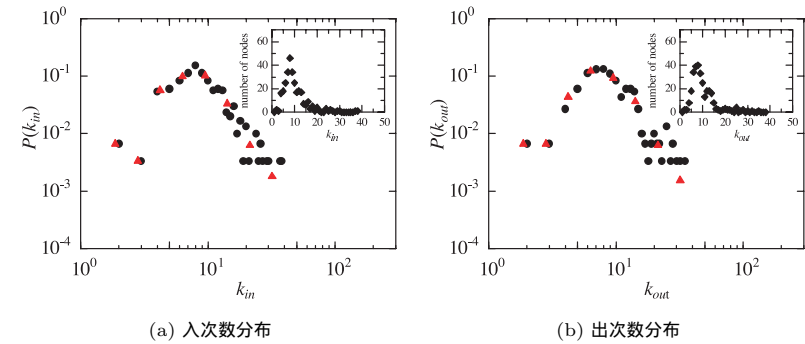


図 16 $\langle k \rangle = 20$ の場合の入次数分布と出次数分布
Fig. 16 The in- and out-degree distributions ($\langle k \rangle = 20$).

4.3 生成実験 3

実験 2 ではネットワーク生成において平均次数 $\langle k \rangle = 6$ とした場合に、複雑ネットワークに見られるような特徴を持ったネットワークが生成された。以下では、 $\langle k \rangle$ を変化させた場合に各特性量にどのような影響があるのか調べる。ノード数は 300 と固定し、生成方法は実験 2 と同じ多目的 GA を用い、 $\langle k \rangle$ のみ変化させる。得られたネットワークの平均パス長 D およびクラスタ係数 C を平均次数別に図 12 に示す。

D の推移を見てみると、 $\langle k \rangle$ に比例して短くなっている。ネットワーク中に辺が増えれば

増えるほどノード間の有向パスが増えるので、 D が短くなるのは明らかである。それに対して、 C は $\langle k \rangle = 7$ でピークを迎え、それ以降減少している。これは各ノードの次数の増加に対して、クラスタ数（三角形の数）の増加が少ないためである。

C のピーク前後でネットワークの次数分布がどのように変化しているのかわかるために $\langle k \rangle = 4, 7, 10, 20$ の場合の、横軸に入、出次数 $k_{in/out}$ 、縦軸に入、出次数分布 $P(k_{in/out})$ とした両対数グラフを図 13、図 14、図 15 および図 16 にそれぞれ示す。グラフ内部の小さなグラフは横軸を次数 $k_{in/out}$ 、縦軸をノードの出現頻度としたグラフである（赤いマ

力は図 11 と同様に区間平均値をプロットしたものである)。C がピークを迎えた $\langle k \rangle = 7$ の度数分布は図 14 に示すように、入、出次数分布がともに直線状に並んでおりベキ則的傾向を示していることが分かる。図 13 では、総ノード数に対して総辺数が少ないので次数の大きなハブは出現しないが、その入、出次数分布はベキ則的傾向を示している。図 15 に示す $\langle k \rangle = 10$ では、次第に両分布にランダム有向ネットワークに見られるような傾向が現れはじめる。さらに、図 16 に示す $\langle k \rangle = 20$ になると両分布が明らかに二項分布（ポアソン分布）に近いものとなっている。図 12 における C がピークを迎える $\langle k \rangle = 7$ までに入、出次数分布にベキ則的傾向が現れ、それ以上 $\langle k \rangle$ を増やすと二項分布的傾向が現れる。ゆえに、限られた辺数でスモールワールド性からアプローチすることで、入、出次数分布にベキ則的傾向が現れることが分かった。

5. ネットワークの解析

本章では多目的 GA により得られたネットワークについてより詳しく調べるために、4.3 節で生成したクラスタ係数 C が最大となった平均次数 $\langle k \rangle = 7$ のネットワークを用いて解析を行う。以下では、このネットワークを GAN と呼ぶ。

5.1 ランダム有向ネットワークとの比較

多目的 GA により生成したネットワーク（GAN）と次数分布の等しいランダム有向ネットワーク（RN）を文献 19) の手法により 100 個生成し、ネットワーク特性量にどのような違いが現れるのか検証する。

GAN と RN の平均パス長 D およびクラスタ係数 C を求めると表 3 のようになった。GAN と RN において D はほとんど変わらないが、 C に明らかな違いが現れている。このことから次数分布からのアプローチでは、必ずしもスモールワールド性が現れるとは限らないということが分かる。

次に GAN と RN に発生するネットワークモチーフについて調べる。対象とするモチーフは 3 ノードからなるものに限定した。その結果を表 4 および図 17 に示す。この結果から、RN にはあまり見られないネットワークモチーフが GAN には多く含まれていることが分かった。

5.2 ランダム故障とピンポイント攻撃への耐性

ネットワークのノードに対して、ランダム故障とピンポイント攻撃を行った場合の特性量の変化を調べる。特性量には最大連結成分に対する平均パス長 D 、ネットワークの最大連結成分の相対サイズ S および分断された成分の平均サイズ $\langle s \rangle$ （最大連結成分以外）を用い

表 3 特性量の比較
Table 3 The comparison of RN and GAN.

	D	C
RN	3.52	0.09
GAN	3.49	0.40

表 4 モチーフの発生割合
Table 4 The rate of network motifs.

	A	B	C	D
RN	14.2	5.9	7.7	14.3
GAN	58	11	49	54

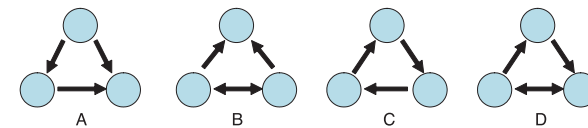


図 17 得られた特徴的モチーフ

Fig.17 Network motifs in the generated network.

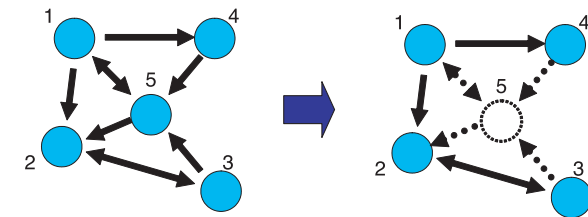


図 18 ノード破壊の概略図

Fig.18 An example of destruction node.

る。ノードの破壊は選ばれたノードの機能が失われることを考え、ノードに接続されている辺を取り除くことで行う。その例を図 18 に示す。この例の場合、ノード番号 5 が破壊され、それにもとない辺が失われている様子を表している。実線が機能しているノードと辺、破線が破壊されたノードと辺である。

ランダム故障とはネットワークの中からランダムに 1 つのノードを選びそのノードを破壊することである。それに対しピンポイント攻撃とはネットワークの中から最も次数の高い

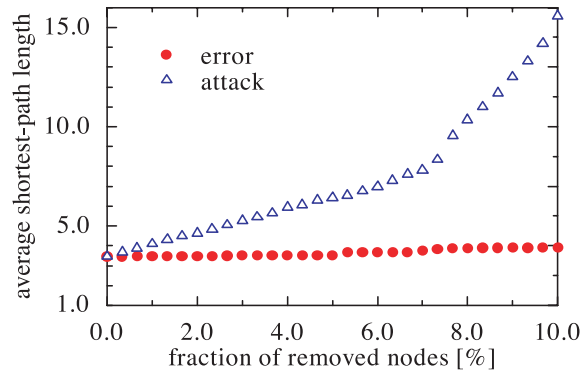
図 19 最大連結成分に対する D の推移

Fig. 19 The average shortest-path length D for the largest cluster S as a function of the fraction of removed nodes.

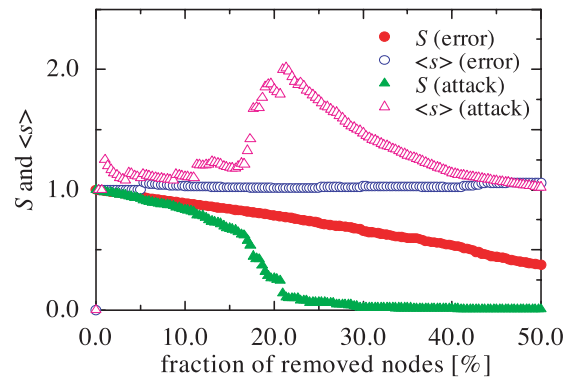
図 20 S および $\langle s \rangle$ の推移

Fig. 20 The relative size of largest cluster S and the average size of the isolated clusters $\langle s \rangle$ as a function of the fraction of removed nodes.

ノードを選択し、そのノードを破壊することである。生成したネットワークに対してノードの破壊を行った場合の各特性量の推移を図 19 および図 20 に示す。図 19 から最大連結成分に対する平均パス長 D はランダム故障に対してほとんど変化がないが、ピンポイント攻撃に対しては急激に長くなっていることが分かる。多くのノードは低い次数を持っているのでランダム故障を起こしてもさほど D は変化しない。しかし、わずかしが存在しない次数

の高いノードがピンポイント攻撃されることで、ハブが破壊され各ノード間の移動が効率良くできなくなってしまうためにこのような現象が見られる。

図 20 より、ランダム故障に対しては S がノードの破壊率に比例して減少するが、ピンポイント攻撃に対してはネットワークが一気に分断し、破壊率が 20% を過ぎた時点でほぼ 0 となり完全に分断されてしまうことが分かる。 $\langle s \rangle$ の推移を見ると、ランダム故障に対してはほぼ 1 である。しかし、ピンポイント攻撃に対しては大きな値をとり、 S が 0 に近づく破壊率 20% 過ぎにピークをむかえ、その後はバラバラになったネットワークを攻撃するだけになるため 1 に近づいていく。

本実験の結果は文献 5) のスケールフリー性を持った実ネットワークに対する実験結果と同様にランダム故障には強いがピンポイント攻撃には弱いという傾向が見られ、本手法で得られたネットワークがこれらの特性においては実ネットワークと同様の傾向を持っていることが分かる。

5.3 他の研究との比較

本節では他の有向スケールフリーネットワークモデルと本論文手法との比較について述べる。以下では本論文のネットワーク生成手法を GAN 法と呼ぶ。

GAN 法と同様に成長しない有向ネットワークを対象とした他の研究としては閾値モデル¹⁰⁾を拡張した中村らの手法¹¹⁾がある。閾値モデルは各ノードがある分布関数に従って決まる重みを持ち、2つのノードの重みの和が閾値を超えたときに限り、この2ノード間に辺が張られる成長しない無向モデルである。文献 11) では各ノードに2つの重み(辺の張りやすさと張られやすさ)を持つように拡張することによって入、出次数分布ともにベキ則に従う有向ネットワークを生成している。ただし、入、出次数分布のベキ指数はともに2を超えない。GAN 法で得られたネットワークはデータ数が少ないながらも近似直線のベキ指数はともに2.35程度となっている(4.2節参照)。実際のWWWのハイパーリンク構造は両方のベキ指数が2以上となることが知られている。

成長する有向ネットワークのモデルとしては、WWWの成長を模倣したモデルがある¹²⁾。このモデルでは確率 p で新たなノードがネットワークに加わる。ネットワーク中に存在する古いノードの入次数に依存した確率でターゲットとなるノードが選択され新たに加わったノードから有向辺が張られる。これはWebページの作者が自身のページから多くのリンクを集めているページ(人気のあるページ)へリンクを張ることに対応している。確率 $1-p$ で新たなリンクがネットワーク中にすでに存在する2つのノード間に張られる。リンクを出すノードとターゲットとなるノードの選択は、それぞれ出次数と入次数に依存した確率で選択

される。これは既存の Web ページから多くのリンクを集めているページへ新たなリンクを張ることに対応し、リンクを多く張っているページが更新されやすいことを意味する。このモデルはパラメータを適切に設定することによって入、出次数分布が実際の WWW のデータに合ったベキ則に従うことが示されている。しかしながら実際の WWW よりもクラスタ係数が低い。このクラスタ係数の低さを解決するために Ding ら¹³⁾ は辺に距離 (*Euclidean distance*) を導入した。

Grabowski ら¹⁴⁾ は、あるノードの出次数はそのノードの現在の出次数に比例して増加するとし、ランダムに選択されたノードへ最大出次数に依存した確率で新たなリンクを張っている。これは新しいリンクが Web ページに加わることで、あるいは友人の多い人がさらに友人を増やすことに対応する。新たなノードの追加は出次数と入次数の和の最大値に依存した確率でなされる。これは WWW において Web ページの追加は出次数 (他のページへのリンク数)、入次数 (他のページから張られているリンク数、人気) がともに大きいページでなされやすいことに対応する。ノードには内部重みを持たせ、ノード間のリンクの張替えはノード間の内部重みの差に基づいて行われる。これは各ノードがネットワーク全体の知識を持つのではなく局所的な知識のみを持つことに対応している (Web ページの作者は WWW のすべての知識を持っているわけではない)。このモデルでは入、出次数分布がともにベキ則に従い、クラスタ係数が実際の WWW と同様に高くなることが示されている。

上記で述べた手法はすべて、次数に基づく確率やノードに重みを持たせるなどの方法によって有向スケールフリーネットワークを生成している (GAN 法と同様にクラスタ係数は辺の向きを考慮していない)。それに対して GAN 法では、平均パス長とクラスタ係数を評価関数 (探索方向) とした多目的 GA により、短い平均パス長および高いクラスタ係数を持つネットワークを生成している。ネットワーク生成過程においてノードの次数分布がベキ則に従うように意図的に設計した機構が存在しないにもかかわらず、総ノード数に対して限られた辺数の範囲において、生成されたネットワークの入、出次数分布はともにベキ則的傾向を示す。この性質は GAN 法の重要な特徴の 1 つであると考えられる。

6. まとめと今後の課題

多目的 GA を用いてスモールワールド性からアプローチすることで、平均パス長が短く、かつクラスタ係数が高いスモールワールド性と、次数分布のベキ則によるスケールフリー的な性質を持つ有向ネットワークを生成することができた。ネットワーク生成過程において、クラスタ係数を高めることで入、出次数分布に変化が生じ、その結果ハブが生成され、か

つ平均パス長を短くすることにより、大多数のノードが少数のハブを利用するように辺の向きが変化する。その結果、入、出次数分布にベキ則的傾向が現れると考えられる。ただし、総ノード数に対して限られた辺数のときにのみ、このような現象が見られた。

生成したネットワークと次数分布の等しいランダム有向ネットワークとの比較から、次数分布がベキ則的になっていてもクラスタ係数が高いとは限らないこと、および特徴的なネットワークモチーフが発生するとは限らないことを確認した。つまり、次数分布のみからのアプローチでは不十分であるといえる。ノードの故障と攻撃に対する特性量の変化を調べることで、これらの特性においてはスケールフリー性を持つ実ネットワークと同様の傾向を持つことが分かった。

本論文では、計算機の都合上ネットワークの総ノード数を 300 としたが、今後は総ノード数を増やした場合の各特性量の変化を調べる必要がある。その際には文献 24) などのコーディング方法や文献 26) の手法が参考となる。生成されたネットワークと実ネットワークの有向辺のつながり方にどのような違いがあるのか、および特徴的なネットワークモチーフが形成される過程を調べることも重要な課題である。

謝辞 本論文に対し有益なご助言をいただいた査読者の方々に感謝いたします。

参考文献

- 1) Barabási, A.-L.: *Linked: The New Science of Networks*, Perseus Books (2002).
- 2) Watts, D.J. and Strogatz, S.H.: Collective dynamics of 'small-world' networks, *Nature*, Vol.393, pp.440-442 (1998).
- 3) Barabási, A.-L. and Albert, R.: Emergence of Scaling in Random Networks, *Science*, Vol.286, pp.509-512 (1999).
- 4) Albert, R. and Barabási, A.-L.: Statistical mechanics of complex networks, *Rev. Mod. Phys.*, Vol.74, No.1, pp.47-97 (2002).
- 5) Albert, R., Jeong, H. and Barabási, A.-L.: Error and attack tolerance of complex networks, *Nature*, Vol.406, pp.378-382 (2002).
- 6) Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. and Wiener, J.: Graph structure in the Web, *Computer Networks*, Vol.33, pp.309-320 (2000).
- 7) Ohkubo, J. and Horiguchi, T.: Complex Networks by Non-growing Model with Preferential Rewiring Process, *J. Phys. Soc. Jpn.*, Vol.74, No.4, pp.1334-1340 (2005).
- 8) Caldarelli, G., Capocci, A., De Los Rios, P. and Muñoz, M.A.: Scale-Free Networks from Varying Vertex Intrinsic Fitness, *Phys. Rev. Lett.*, Vol.89, No.25, p.258702

- (2002).
- 9) Bollobás, B., Borgs, C., Chayes J. and Riordan, O.: Directed scale-free graphs, *Proc. 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '03)*, Philadelphia, PA, USA, pp.132–139 (2003).
 - 10) Masuda, N., Miwa, H. and Konno, N.: Analysis of scale-free networks based on a threshold graph with intrinsic vertex weights, *Phys. Rev. E*, Vol.70, No.3, p.036124 (2004).
 - 11) 中村和弘, 塩田茂雄: ノードが二つの内部重みを持つ有向ネットワークのスケールフリー性, 信学技法, 情報ネットワーク (IN), Vol.107, No.37, pp.13–18 (2007).
 - 12) Krapivsky, P.L. and Redner, S.: A statistical physics perspective on Web growth, *Computer Networks*, Vol.39, pp.261–276 (2002).
 - 13) Dinç, D. and Bingol, H.: The effect of euclidean distance in directed scale-free network generation, *Proc. 22nd International Symposium on Computer and Information Sciences (ISCIS '07)*, Ankara, Turkey, pp.1–5 (2007).
 - 14) Grabowski, A. and Kosiński, R.: Evolving Directed Network with Intrinsic Variables and Local Rules: a Simple Model of WWW networks, *Acta Physica Polonica B*, Vol.38, No.5, pp.1785–1793 (2007).
 - 15) Milo, 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).
 - 16) Newman, M.E.J.: Mixing patterns in networks, *Phys. Rev. E*, Vol.67, No.2, p.026126 (2003).
 - 17) Schwartz, N., Cohen, R., ben-Avraham, D., Barabási, A.-L. and Havlin, S.: Percolation in directed scale-free networks, *Phys. Rev. E*, Vol.66, No.1, p.015104 (2002).
 - 18) 花田良子, 佐藤史隆, 廣安知之, 三木光範, 鈴木泰博: 遺伝的アルゴリズムによるネットワーク特性量に着目したネットワーク設計法, コンピュータソフトウェア, Vol.24, No.1, pp.91–100 (2007).
 - 19) Newman, M.E.J., Strogatz, S.H. and Watts, D.J.: Random graphs with arbitrary degree distributions and their applications, *Phys. Rev. E*, Vol.64, No.2, p.026118 (2001).
 - 20) Bollobás, B.: *Random Graphs, 2nd ed.*, Cambridge University Press (2001).
 - 21) Caldarelli, G.: *Scale-Free Networks – Complex webs in nature and technology*, Oxford University Press (2007).
 - 22) Latora, V. and Marchiori, M.: Efficient Behavior of Small-World Networks, *Phys. Rev. Lett.*, Vol.87, No.19, p.198701 (2001).
 - 23) Ohkubo, J., Yasuda, M. and Tanaka, K.: Replica Analysis of Preferential Urn Model, *J. Phys. Soc. Jpn.*, Vol.75, No.7, p.074802 (2006).
 - 24) 玄 光男, 林 林: ネットワークモデルと多目的 GA, 共立出版 (2008).
 - 25) Goldberg, D.E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley (1989).
 - 26) Kitano, H.: Designing neural networks using genetic algorithms with graph generation system, *Complex Systems*, Vol.4, pp.461–476 (1990).

(平成 21 年 1 月 21 日受付)

(平成 21 年 6 月 4 日採録)



近藤 久 (正会員)

平成 6 年北海道大学大学院工学研究科情報工学専攻博士後期課程修了。博士 (工学)。同年茨城大学工学部システム工学科助手。平成 13 年同講師。平成 17 年同工学部知能システム工学科講師。強化学習, 複雑ネットワーク等の研究に従事。人工知能学会, 日本ソフトウェア科学会, EATCS, IEEE-CS 各会員。



鴫崎 和宏

昭和 60 年生。平成 21 年茨城大学大学院理工学研究科博士前期課程システム工学専攻修了。同年味の素システムテクノ (株) 入社。在学中は, 遺伝的アルゴリズムを用いたスケジューリング問題の最適化や複雑ネットワークの生成に関する研究に従事。