

# ネットワーク特性量を用いた 多目的GAによる複雑ネットワーク生成

鵜崎 和宏<sup>†</sup> 近藤 久<sup>††</sup>

<sup>†</sup> 茨城大学大学院理工学研究科システム工学専攻 <sup>††</sup> 茨城大学工学部知能システム工学科  
〒 316-8511 茨城県日立市中成沢町 4-12-1  
E-mail:07nm924t@hcs.ibaraki.ac.jp<sup>†</sup> H.Kondo@mx.ibaraki.ac.jp<sup>††</sup>

## 和文抄録

本研究では、ネットワーク特性量を用いて多目的GAにより複雑ネットワークの生成を行なう。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提案する。そして、生成されたネットワークをランダム有向ネットワークや実ネットワークと比較することで本手法の有効性を調べる。この実験の結果、多目的GAによりスモールワールド性からアプローチすることで、生成したネットワークにはランダム有向ネットワークには見られないモチーフが多く含まれること、スケールフリー性を持つ実ネットワークに見られる特性が類似することを確認した。

# Generation of Complex Networks by Using Multi-objective Genetic Algorithm

Kazuhiro TOKIZAKI<sup>†</sup> Hisasi KONDO<sup>††</sup>

<sup>†</sup> Graduate School of Science and Engineering, Ibaraki University

<sup>††</sup> Department of Intelligent System Engineering, Ibaraki University

4-12-1 Naka-Narusawa, Hitachi 316-8511, Japan

E-mail:07nm924t@hcs.ibaraki.ac.jp<sup>†</sup> H.Kondo@mx.ibaraki.ac.jp<sup>††</sup>

## Abstract

We present a method of generating complex networks by using multi-objective genetic algorithm for the optimization of two evaluation functions. The first evaluation function is the mean geodesic (i.e, shortest) distance between vertex pairs in a network, the second is the clustering coefficient, which measures the fraction of triples that have their third edge filled in to complete the triangle. In real networks, it is known that the former is comparatively short, and the latter tends to be considerably higher compared with the random graph with a similar number of vertices and edges. We conduct some experiments to see whether the generated network has properties of real networks, and show that our method is effective.

## 1 はじめに

近年、今まで何の規則性も無いと考えられていたWebのハイパーリンク構造や電力網などの大規模なネットワークに様々な特徴的構造が見つかってきている<sup>1)</sup>。その様なネットワークは複雑ネットワーク (*complex networks*) と呼ばれ、様々な研究が行なわれている。

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<sup>2) 3)</sup>。これらのモデルは無向ネットワークを考えたものであり有向

ネットワークを考えたものは少ない。しかしながら、Webのハイパーリンク構造や学术论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。それらの有向ネットワークにはランダム有向ネットワークにはほとんど見られないネットワークモチーフ (*network motifs*) が多く含まれている<sup>4)</sup>。そこで、本研究ではネットワークの特性量に着目した、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手法<sup>5)</sup>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のか調べ、ランダム有向ネットワーク<sup>4, 6)</sup>および実ネットワークとの比較を行なうことで本手法の有効性を検証する。

## 2 複雑ネットワークの特性量

複雑ネットワークに見られる特性量には、主に以下のようなものがある。

### 2.1 次数分布

次数とは、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確率密度関数  $P(k)$  はランダムネットワークにおいて次式に示すポアソン分布に従う。ただし、 $\langle k \rangle$  は平均次数である。

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

しかしながら、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、その確率密度関数  $P(k)$  は次式のようなべき則に従う。ただし、 $\gamma$  はべき指数である。

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

式(2)の  $\gamma$  は 2~3 程度の値を持つことが知られている<sup>7)</sup>。

有向ネットワークでは各ノードに入る辺数を入次数、ノードから出て行く辺数を出次数という。Webのハイパーリンク構造では、入次数、出次数の次数分布が共にべき則に従う傾向にある<sup>3, 8)</sup>。

### 2.2 平均パス長

平均パス長 (*mean geodesic distance*) とは、ネットワークにおける任意の 2 ノード間の最短到達距離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<sup>7)</sup>。

本研究では、有向ネットワークを考えるのでパス長を求める際にノードの向きを考慮する。このとき、ノード  $i$  からノード  $j$  への最短距離を  $D_{ij}$  とする。しかし、ノード  $i$  からノード  $j$  へのパスが無い場合  $D_{ij} = \infty$  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  $\frac{1}{D_{ij}}$  をとることでパスの存在しないノード間の距離を 0 とみなせる。全てのノードに対して  $\frac{1}{D_{ij}}$  を求め、その平均を取ったものが  $D$  であり、次式のように表せる。

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

式(3)から求められる  $D$  (調和平均パス長) を平均パス長  $D$  とする。

### 2.3 クラスタ係数

クラスタとは、ネットワーク中に現れる三角形のことである。このクラスタがネットワーク中にどれだけあるか示すための指標をクラスタ係数 (*clustering coefficient*) 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言える。例えば、人間をノード、友好関係を辺で表現した人間関係のネットワークを考えるとクラスタができるということは、Aの友人Bの友人CもまたAの友人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人間関係が築かれていることを意味する。複雑ネットワークで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<sup>7)</sup>。

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  $i$  が持つ辺数、 $E_i$  はノード  $i$  のクラスタ数である。

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \quad (4)$$

この  $C_i$  を全てのノードに対して計算し、総ノード数  $N$  で割ったものがネットワークのクラスタ係数  $C$  であり次式のように定義される。

$$C = \frac{1}{N} \sum_{i=1}^N C_i \quad (5)$$

## 2.4 ネットワークモチーフ

ネットワークモチーフは、ネットワーク中に現れる部分グラフのことである。例えば、頂点数が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。その例を Fig.1 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見られない<sup>4)</sup>。

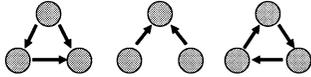


Fig. 1 ネットワークモチーフの例

## 3 ネットワークの生成法

本研究では、2節で述べた特性量のうちクラスタ係数と平均パス長を用い、遺伝的アルゴリズム (*Genetic Algorithms: GA*) を適用することによりネットワークを生成する。本節では、GA を適用するためのネットワークのコーディング方法および GA の適用方法について説明する。また、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を制約条件とする。

### 3.1 ネットワークのコーディング

本研究で設計するネットワークは、単純有向ネットワークであるので隣接行列を用いて表現することができる。そこで、隣接行列を用いてネットワークのコーディングを行う。

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。このとき、 $A$  の各成分  $a_{ij}$  を式 (6) により置換する。

$$a_{ij} = \begin{cases} -1 & (a_{ij} = 0, a_{ji} = 1) \\ 2 & (a_{ij} = a_{ji} = 1) \end{cases} \quad (6)$$

そして、置換後の行列の対角成分より上の三角形の部分をも1行目から順番に1列に並べていく。すると、 $\frac{N(N-1)}{2}$  の長さで有向ネットワークを表現することができる。本研究では、これを染色体としネットワークのコーディングを行う。 $N = 5$  の場

合のコーディング例を Fig.2 に示す。この例の場合、染色体は 100120001-1 となる。

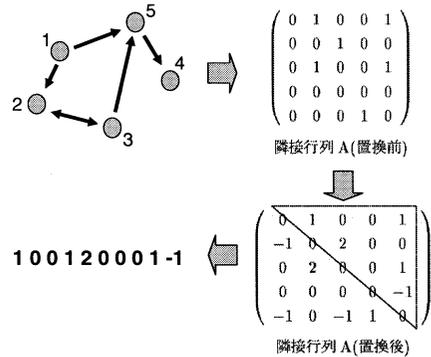


Fig. 2 コーディング例

### 3.2 選択

親となるペアを選択するために、本研究では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なう戦略である。個体  $i$  に対する評価を  $F_i$  とすると、総個体数  $N$  とき個体  $i$  の選択確率  $P_i$  は次式で表される。

$$P_i = \frac{F_i}{\sum_{j=1}^N F_j} \quad (7)$$

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。

### 3.3 交叉

選択によって選ばれた親の特性を残した子を生成するために、共通の辺は親の形質遺伝として固定し、それ以外の辺は親のペア同士でランダムに張り替え、子に継承させることで交叉を行なう。ただし、総辺数を制約とするので、総辺数が変わらないように張替えを行う。その例を Fig.3 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。親同士で共通でない辺は親同士の張替えによって子に継承するので、子 1 は親 1 から辺 3 を親 2 から辺 a, b, c をそれぞれ

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。

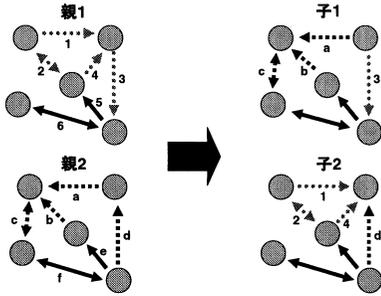


Fig. 3 交叉例

### 3.4 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。本研究では、以下の3つの方法から適用可能なものをランダムにいずれかを発生させる。

- (i) 片方向の辺を2本選び、一方を双方向に、他方を削除する。
- (ii) 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。
- (iii) ランダムに辺の張替えをする。

それぞれの具体例を Fig.4 に示す。(i) の場合、点線で示す片方向の辺が2本選ばれそれが1本にまとめられている。(ii) 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。(iii) の場合では辺をばらばらに張替えて新たなネットワークを生成している。

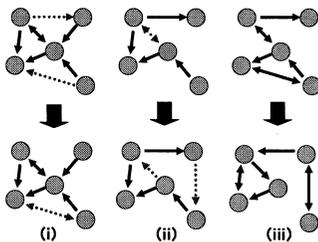


Fig. 4 突然変異例

## 4 GA を用いたネットワーク生成

平均パス長とクラスタ係数をそれぞれ評価関数とし、平均パス長の最小化およびクラスタ係数の最大化をそれぞれ目的として、ネットワークの生成を行なう。最終的に得られた集団の中から評価の最も高いものを1つ選ぶ。そして、そのネットワークに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が現れるのか調べる。

### 4.1 実験結果 1

総ノード数は300とし、平均次数が6となるように総辺数を設定した。GAを適用するためのパラメータを Table.1 示す。総ノード数と総辺数を制約としたランダム有向ネットワークを100個生成し、初期集団とした。

Table 1 パラメータ設定

総ノード数	300
総辺数	900
ネットワーク数	100
世代交代回数	1000
交叉確率	1.0
突然変異確率	0.1

平均パス長の最小化を目的とした場合、およびクラスタ係数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平均パス長  $D$  およびクラスタ係数  $C$  を Table.2 に示す。また、生成過程における平均パス長  $D$  とクラスタ係数  $C$  の平均値の推移をそれぞれ Fig.5 および Fig.6 に示す。

Table 2 単一目的 GA 適用時の  $D$  および  $C$

	平均パス長 $D$	クラスタ係数 $C$
$D$ の最小化	5.44	0.018
$C$ の最大化	10.32	0.061

この結果から、平均パス長  $D$  の最小化を目的とすると、平均パス長  $D$  もクラスタ係数  $C$  も初期ネットワークであるランダム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わかる。これは、ネットワーク全体を連結にしようとする力が働かないためである。つまり、初期状態で連結に

なっている小さな連結成分の中でそれぞれ平均パス長  $D$  の最適化を行ってしまう。その結果、小さな連結成分での平均パス長  $D$  の最小化は簡単なのですぐに収束してしまう。すると、ネットワークは全体として連結にならない。そのためクラスタ係数  $C$  も高くなり複雑ネットワークが持つような短い平均パス長  $D$  と高いクラスタ係数  $C$  を実現できなかったと考えられる。

クラスタ係数  $C$  の最大化を目的とするとクラスタ係数  $C$  は徐々に高くすることができるが、平均パス長  $D$  が短くならず逆に長くなってしまっている。これは、平均パス長  $D$  の最小化を目的とした場合とは異なりネットワークが大きな連結成分になり始める。しかしながら、辺の方向性をまったく考えていないので連結ではあるが各ノード間に効率的なパスが生成されず平均パス長  $D$  が長くなってしまふ。ゆえに、クラスタ係数  $C$  は高くなるが平均パス長  $D$  も長くなってしまふと考えられる。

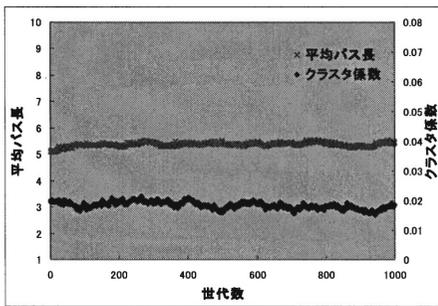


Fig. 5  $D$  と  $C$  の推移 ( $D$  のみ最適化)

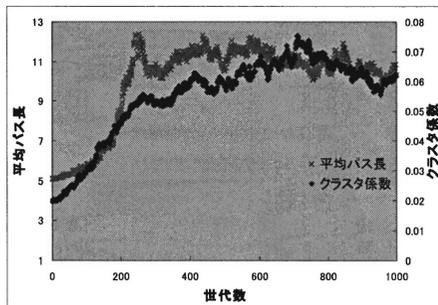


Fig. 6  $D$  と  $C$  の推移 ( $C$  のみ最適化)

この実験から、単純にネットワークの平均パス長を短くするアプローチおよびクラスタ係数を高くするアプローチでは、複雑ネットワークに見られる特性が得られないことがわかった。

## 5 多目的GAを用いたネットワーク生成

実験1の結果を踏まえ、平均パス長  $D$  とクラスタ係数  $C$  の両方を評価関数として多目的GAを適用することでネットワークの生成を行なう。

### 5.1 ネットワークの評価方法

平均パス長  $D$  の最小化とクラスタ係数  $C$  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を用いてネットワークを評価する。パレートランキング法では、ネットワーク  $i$  に対して  $n_i$  個のネットワークが優れているとき、ネットワーク  $i$  のランク  $R_i$  は次式で求められる。

$$R_i = 1 + n_i \quad (8)$$

式(8)を用いて全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 Fig.7 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。

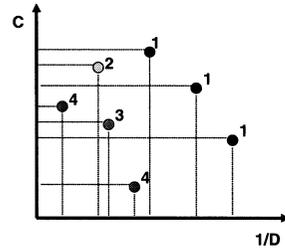


Fig. 7 パレートランキング法の例

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なう。

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 Fig.8 に示す。このネットワークは平均パス長  $D = 3.73$ 、クラスタ係数  $C = 0.62$  であり、短い  $D$  と大きな  $C$  の両方の性質を満たしている。ベキ指

数はそれぞれ  $\gamma_{in} = 1.66, \gamma_{out} = 1.82$  である。5.2 では 4.1 の実験と同様に総ノード数を 300 とした場合に生成されたネットワークの特徴を述べる。

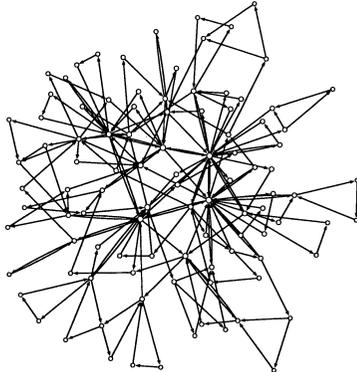


Fig. 8 ネットワークの生成例

## 5.2 実験結果 2

多目的 GA を適用するためのパラメータを Table.3 に示す。初期集団は総ノード数と総辺数を制約としたランダム有向ネットワークとし、総ノード数および総辺数は実験 1 と同様に設定した。

Table 3 パラメータ設定

総ノード数	300
総辺数	900
ネットワーク数	100
世代交代回数	2000
交叉確率	1.0
突然変異確率	0.1

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  $D = 3.80, C = 0.39$  となった。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 Fig.9 に示す。

Fig.9 より、多目的 GA を適用し  $D$  と  $C$  の両方を同時に最適化することで効率的に短い  $D$  と高い  $C$  をもつネットワークが生成されている。

累積分布関数を用いて生成されたネットワークの度数分布を調べると Fig.10 および Fig.11 に示す

ようにデータ数が少ないながらも入次数、出次数共にベキ則に近いものが得られた。それぞれのベキ指数は  $\gamma_{in} = 1.97, \gamma_{out} = 1.99$  である。

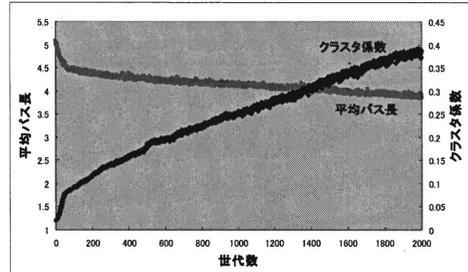


Fig. 9  $D$  と  $C$  の推移 (多目的 GA 適用時)

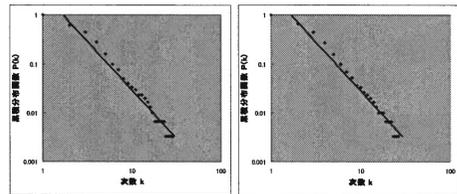


Fig. 10 入次数分布 Fig. 11 出次数分布

この結果より、生成されたネットワークに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度数分布のベキ則で特徴付けられるスケールフリー性をみることができた。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、各ノード間に効率的なパスを生成しながら、ネットワークを大きくする力が働くことになる。すると、ネットワーク全体が連結になり、限られた辺数の中で効率的なパスを実現するために必要なハブが現れる。その結果として、ネットワークにスケールフリー性が発生したと考えられる。

次節では、ランダム有向ネットワークおよび実ネットワークと得られたネットワークの特性を比較することで生成したネットワークをより詳しく解析する。

## 6 ネットワークの解析

### 6.1 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 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。

まず、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 Table.4 のようになった。

Table 4 特性量の比較

	平均パス長 $D$	クラスタ係数 $C$
GAN	3.80	0.39
RN	3.91	0.056

Table.4 より、GAN と RN において平均パス長  $D$  はほとんど変わらないが、クラスタ係数  $C$  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からでは、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わかる。

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 点からなるものに限定した。その結果を Table.5 および Fig.12 に示す。

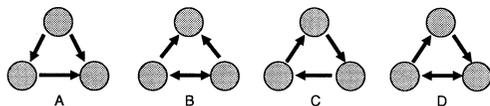


Fig. 12 得られた特徴的モチーフ

Table 5 モチーフの発生割合

	A	B	C	D
GAN	79	10	75	34
RN	16.6	4.6	9	14.9

この結果から RN にはほとんど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わかった。

### 6.2 ランダム故障とピンポイント攻撃に対する比較

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起こした場合の特性量の変化を調べる。特性量には、最大連結成分の平均パス長  $D$ 、ネットワークの最大連結成分  $S$  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なう。その例を Fig.13 に示す。この例の場合、ノード 5 が破壊されそれに伴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。

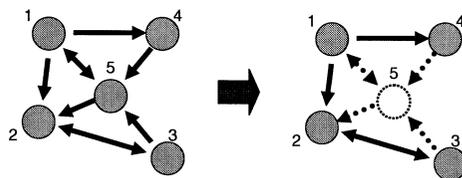


Fig. 13 ノード破壊の概略図

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なった場合の各特性量の推移を Fig.14 および Fig.15 に示す。

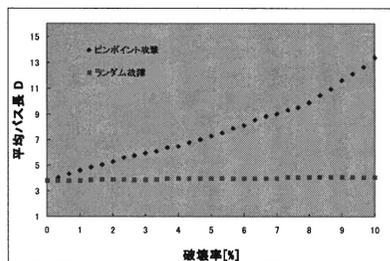


Fig. 14 最大連結成分に対する  $D$  の推移

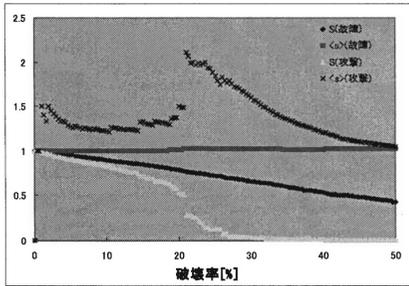


Fig. 15  $S$  および  $\langle s \rangle$  の推移

Fig.14 に示すランダム故障とピンポイント攻撃を行なった場合の最大連結成分に対する平均パス長  $D$  の推移からランダム故障に対して  $D$  はほとんど変わらないが、ピンポイント攻撃に対しては急激に高くなっていることがわかる。これは、次数の高いノードが攻撃されることでハブが破壊され効率よく各ノード間の移動ができなくなっているからであると考えられる。

Fig.15 より最大連結成分  $S$  の推移からピンポイント攻撃に対してネットワークが一気に分断していることがわかる。それに対してランダム故障の場合は、急激にネットワークが分断されることはない。分断された成分の平均サイズ  $\langle s \rangle$  からその様子が良くわかる。

本実験の結果は文献<sup>10)</sup>の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実ネットワークの特性を持っていることがわかる。

## 7 まとめと今後の課題

多目的 GA を用いることで、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性を持つ有向ネットワークを生成することができた。また、生成過程においてネットワーク全体を連結しながら、限られた辺数で効率的なパスを生成しようとするためにハブが出現し、その結果次数分布にベキ則が現れたと考えられる。生成したネットワークには、ランダム有向ネットワークにはほとんど見られない特徴的なネットワークモチーフが多く含まれていた。さらに、ノードの故障と攻撃に対

する特性量の変化を調べると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わかった。以上のことから本手法が有効であると考えられる。

今後は、総ノード数を増やした場合の各特性量の変化、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必要がある。

## 参考文献

- 1) A.-L Barabási. Linked: The New Science of Networks. Perseus books (2002).
- 2) D.J.Watts, S.H.Strogatz. Collective dynamics of 'small-world' networks. Nature, 393, 440-442 (1998).
- 3) A.-L Barabási, R.Albert. Emergence of scaling in random networks. Science, 286, 509-512 (1999).
- 4) R.Milo, S. Shen-Orr, S Itzkovitz, N.Kashtan, D.Chklovskii, U.Alon. Network motifs:Simple building blocks of complex networks. Science, 298, 824-827 (2002).
- 5) 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, Vol24, No.1, 91-100 (2007).
- 6) M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. Physics Reviews E, 64, 026118, 1-17 (2001).
- 7) Réka Albert, A.-L Barabási. Statistical mechanics of complex networks. Reviews of Modern Physics, 74, 47-97 (2002).
- 8) Andrei Broder, Ravi Kumar Ravi, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, Janet Wiener. Graph structure in the Web. Computer Networks, 33, 309-320 (2000).
- 9) D. E. Goldberg: Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley (1989).
- 10) Réka Albert, Hawoong Jeong, A-L Barabási. Error and attack tolerance of complex networks. Nature, 406, 378-382 (2002).