

GA によるモチーフを考慮したネットワーク生成

鶴崎 和宏[†]近藤 久^{††}[†]茨城大学大学院理工学研究科システム工学専攻 ^{††}茨城大学工学部知能システム工学科

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 WWW 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共通の特徴が見つかってきている [1]. その様なネットワークは複雑ネットワークと呼ばれ、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、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]. 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、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。それらの有向ネットワークにはランダム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が多く含まれている [5]. そこで、本研究ではネットワーク特性量に着目し、有向ネットワークのモデル化を行う。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 [4] と類似した手法を用いる。生成されたネットワークとランダム有向ネットワーク [5][6] との比較を行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徴的構造

複雑ネットワークに共通して見られる特徴的構造には以下のようなものがある。

2.1 度数分布

度数とは、各ノードが持つエッジの本数のことである。有向ネットワークでは各ノードに入るエッジ数を入次数、ノードから出て行くエッジ数を出次数という。ランダムネットワークではその度数分布がポアソン分布に従うが、複雑ネットワークでは入次数、出次数が共にべき則に従う [3].

2.2 平均パス長

平均パス長とは任意の 2 ノード間の最短到達距離の平均値のことである。本研究では、パス長を求める際にノードの向きを考慮し、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。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、その平均は次式のように表せる。

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

本研究では、式 (1) 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。複雑ネットワークでは、ノード総数に対して平均パス長が短い傾向にある。

2.3 クラスタ係数

クラスタ係数はネットワークがどれだけ密であるかの指標の 1 つである。3 つのノードがエッジを介してつながっているものをクラスタと呼ぶ。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 が持つエッジ数、 E_i はノード i のクラスタ数である。

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

この C_i を全てのノードに対して計算し、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なる。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高い。

$$C = \frac{1}{n} \sum_{i=1}^n C_i \quad (3)$$

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。頂点数が 3 つの小さなモチーフでもエッジに方向性があるならば 13 種類のモチーフが存在する。その例を図 1 に示す。それぞれのネットワーク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 [5].

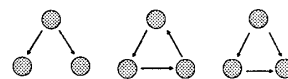


図 1: ネットワークモチーフの例

3 ネットワークの設計方法

本研究では、遺伝的アルゴリズム (GA) を用いてネットワークの設計を行う。GA におけるネットワークのコーディング、交叉および突然変異方法を以下に示す。また、ネットワークのノード数、エッジ数を制約とする。

3.1 ネットワークのコーディング方法

本研究で設計するネットワークは、自己ループおよび多重辺のない有向ネットワークであるのでノードとエッジの関係を隣接行列を用いて表現できる。そこで、隣接行列を利用してコーディングを行う。ノード数 N

Generation of Networks Considering Motifs by Using GA.

[†] Kazuhiro TOKIZAKI (07nm924t@hcs.ibaraki.ac.jp)

^{††} Hisashi KONDO (H.Kondo@mx.ibaraki.ac.jp)

Graduate School of Science and Engineering, Ibaraki University ([†])

Department of Intelligent System Engineering, Ibaraki University (^{††})

4-12-1 Nakanarusawa, Hitachi, Ibaraki 316-8511, Japan

のネットワークに対する、隣接行列 A の要素 a_{ij} を次式により置換する。 $N = 5$ の場合の例を図 2 に示す。

$$a_{ij} = \begin{cases} -1 & (a_{ij} = 0, a_{ji} = 1) \\ 2 & (a_{ij} = a_{ji} = 1) \end{cases} \quad (4)$$



図 2: 置換例

本研究では、置換後の行列の対角成分より上の三角形の部分を選択し染色体としてコーディングを行う。

3.2 交叉方法

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、共通のエッジは形質遺伝として固定し、それ以外のエッジをランダムにどちらかの親から継承するものとする。ただし、エッジの制約を満たすように張替えを行う。

3.3 突然変異

突然変異は、自己ネットワーク内で片方向のエッジを 2 つ選び、一方を双方向に、他方を削除すること、およびランダムにエッジの張替えを行う。

4 シミュレーション (1)

本節では、2 節で述べた特徴的構造である D および C をそれぞれ単一評価関数とした場合のネットワークの形成について調べる。ネットワークのノード数を 100、エッジ数を 250 として集団数を 800 とした。また、GA を交叉確率 100[%]、突然変異確率 1[%]、世代交代回数 1000 回として適用した。そして、最終的に得られたネットワーク集団の中から評価の最も高いものを抽出する。そのネットワークが複雑ネットワークが持つような特徴的構造を持つか調べる。

4.1 シミュレーション結果

D を目的関数とした場合、 $D = 4.12, C = 0.042$ となった。また、 C を目的関数とした場合、 $D = 5.14, C = 0.65$ となった。

この結果から、 D のみを評価関数とすると D はある程度小さくすることができるが、 C がまったく高くならなかった。 D を算出する際に孤立したノードを無視するため、ネットワークが大きくならず小さいネットワークの中での D を最適化してしまう。そのため C が高くないと考えられる。

また、 C のみを評価関数とすると C はかなり高くすることができるが、 D を小さくすることができなかった。これは、単純に方向性を考えない三角形をネットワーク中に多く作ろうとしてしまうため、 D がうまく最適化されないと考えられる。

5 シミュレーション (2)

本節では、シミュレーション (1) の結果を踏まえ、 D と C の両方を評価関数として多目的 GA を用いて最適化を行う。ネットワークの制約および GA のパラメータはシミュレーション (1) と等しいものとした。最終的に得られたネットワーク集団の中からパレート最適なものをつランダムに抽出する。

5.1 シミュレーション結果

最終的に得られたネットワークの D および C はそれぞれ、 $D_{ga} = 3.38, C_{ga} = 0.39$ となった。

この結果より、 D と C の両方を同時に最適化することで複雑ネットワークに見られるような短い D と高い C の両方を実現することができた。

以下では、このネットワークをより詳しく調べるために各ノードの入次数および出次数の等しいランダム有向ネットワーク (RN) を生成し、多目的 GA によるネットワーク (GAN) との比較を行う。

RN の D および C はそれぞれ $D_{rn} = 3.44, C_{rn} = 0.12$ となった。また、GAN の次数分布を図 3 および図 4 に、得られた特徴的モチーフを図 5 に、GAN と RN に対するモチーフの発生個数を表 1 に示す。

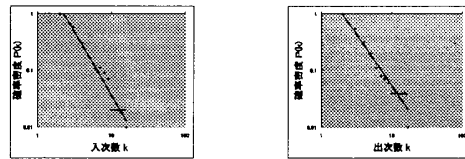


図 3: 次数分布 (入次数) 図 4: 次数分布 (出次数)

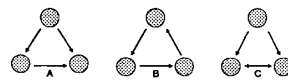


図 5: 得られた特徴的モチーフ

表 1: モチーフの発生割合

	A	B	C
GAN	11	9	11
RN	1.4	1.6	3.7

6 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有向ネットワークを生成できた。その結果、ネットワークが連結になる傾向にあり、ランダム有向ネットワークにはほとんど見られないモチーフが多く含まれている。また、GAN と RN でクラスタ係数に明らかな違いが見られ、より現実の複雑ネットワークに近いものであると考えられる。そして、ノード数が少ないながらも次数分布は入次数、出次数共にベキ則に近いものになっている。

今後は、ノード数を増やした場合の各特性量の変化、および特徴的なモチーフが生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] 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).
- [5] 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).
- [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).