

遺伝的アルゴリズムによる問題サイズの大きな複雑ネットワーク問題の検討

Discussion about solve bigger dimension of the Complex Networks by Genetic Algorithms

佐藤 史隆

Fumitaka SATO

Abstract: Since, Complex networks have gotten a lot of attention. In this report, we solve complex networks problems as combinatorial optimization problems by Genetic Algorithms. In this paper, we discussed about evaluation method for bigger dimension of the problem. It has found that it takes a shorter time using Dijkstra's Algorithm. Then, we solve bigger dimension of the problem using Dijkstra. From these results, It has found that good results which using Single Population GA with Minimal Generation Gap.

1 はじめに

近年、複雑ネットワークに非常に注目が集まっている。複雑ネットワークは、タンパク質の相互作用、インターネットやワールド・ワイド・ウェブ、エイズ(感染のメカニズム)、生態系や細胞の生化学(自然の安定性)、ガンや精神病(遺伝子ネットワーク)など、いたるところに存在し、重要性が高い。しかしながら、複雑ネットワークの構造や特性については分かっていないことが多い¹⁾。複雑ネットワークの構成を知ることは、システムの全体的な特性や挙動を理解するための第一歩となり、非常に重要である。本研究では最適化手法の1つである遺伝的アルゴリズム(Genetic Algorithms: GA)を用いて、離散組み合わせ最適化問題として複雑ネットワークの構成を解くことを目的とする。

本稿では、これまで計算時間の多くかかっていた評価計算方法について検討を行った。また、ノード数を増やし、問題サイズを大きくした場合の実験を行った結果について検討を行う。

2 複雑ネットワーク

複雑ネットワークとは、多数の要素からなるひとまとまりの集団(系)で、各要素が他の要素と絶えず相互作用する結果、全体として見れば、部分の総和以上の独自の振る舞いを示すネットワークの総称である²⁾。

2.1 ネットワークの概要

ネットワークは、Fig. 1 に示すように、中継点である各点のノード(Node)と、それらをつなぎ合わせることで、つまりリンク(Link)させることにより構成される。また、ノード間の距離であるパス長(Path Length)は、あるノードから他のノードに届くまで、いくつのノードを経由していくか、渡っていくリンクの数により表される。

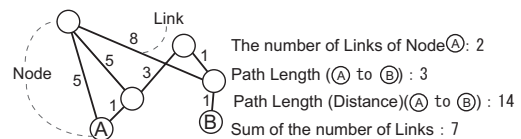


Fig. 1 The example of network composition

本研究では、より実問題に近い問題を想定した、各ノード間の距離を考慮する問題モデルについて扱う。なお、この問題モデルにおいては、平均最短パス長ではなく、より少ない移動距離で到達できるような最短移動距離となるネットワーク構成の構築を目的としている。

2.2 問題モデル

本研究では、各ノード間の距離を考慮する問題モデルを扱い、平均最短移動距離 D を最小化することを目的とする。式(1)に平均最短移動距離 D の定義式を示す。式中、ノード数を n 、また、ノード i からノード j までの最短移動距離を D_{ij} と表す。また、本問題ではノード数、総リンク数を固定することを制約条件とする。

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

本問題モデルでは、任意のノードを基準としたとき、他のノードに到達するために、より少ない移動距離で到達できるような効率の良いネットワーク構成の構築を目的とする。ネットワーク構成の総数は、ノード数を n 、総リンク数を m とするとき、すべてのノード間の組み合わせが、 $n(n-1)/2$ 通りあるため、 $n(n-1)/2 C_m$ となる。よって本問題モデルは、 $n(n-1)/2 C_m$ の中から、最適な組み合わせを選ぶ離散組み合わせ最適化問題となる。

3 遺伝的アルゴリズム

遺伝的アルゴリズム(Genetic Algorithms: GA)は生物の進化のメカニズムを模倣した最適化手法である³⁾

自然界では、環境に適合できない生物は死滅していくが、環境に適合した個体は生き残り、子孫を増やしていく。GAは、この自然界のメカニズムをモデル化し、与えられた環境に最もよく適合したもの、すなわち目的関数に対して最適値を与えるような解を計算機上で求めようとする最適化手法である。遺伝的オペレータ（選択、交叉、突然変異）を繰り返すことにより、最適化を行う。

3.1 コーディング

GAによる複雑ネットワーク問題の最適化を行うためには、まず各ノードのリンクの関係を染色体にコーディングする必要がある。本研究では、各ノード間におけるリンクの関係を、つながっていない場合に0、つながっている場合に1という{0, 1}のビットからなる遺伝子によって個体を構成するコーディング方法を用いる。ノード数を n とすると、ノード間の関係は $n(n-1)/2$ 通りあるため、この方法を用いると遺伝子長は $n(n-1)/2$ となる。

ノード数6、総リンク数5の問題において、このコーディング法を用いた場合の例を Fig. 2 に示す。Fig. 2の左に示すような構成のネットワークから、各ノード間のリンクの関係表を作成することができる。作成した関係表より遺伝子型を作成する。

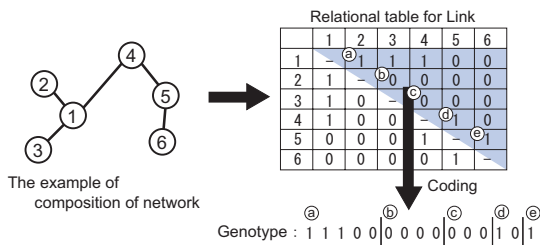


Fig. 2 Coding

4 評価方法の検討

2.2節の式(1)に示したように、本問題モデルでは評価において、全てのノード間の最短経路を求める必要がある。また、ノード数が増加すると、飛躍的に探索空間が増加する。そのため、本問題モデルにおいて、評価にかかる計算時間は非常に大きく、より大きな問題サイズの問題を解くためには評価方法、つまり最短経路を求めるアルゴリズムについて検討を行う必要がある。最短経路を求めるアルゴリズムとして、以下のようなものが挙げられる。

- ダイクストラ法 (Dijkstra 法)
- ウォーシャル・フロイド法 (Warshall-Floyd 法)
- A*探索

本研究では、疎なグラフ¹に対して有効であるとされ、かつ全てのノード間の最短経路を求めるのに適している

¹ノード数 N の時に $N(N-1)/2$ よりリンク数が少ないグラフ

とされるダイクストラ法の実装を行い、これまで用いていた全探索の手法との性能比較を行う。

4.1 ダイクストラ法

ダイクストラ法 (Dijkstra 法) は特定の点からの最短経路・最短距離を求める方法である。以下にダイクストラ法のアルゴリズムを示す。

1. 始点につながっているそれぞれの節点までの始点-節点間の距離を比べ、最小の距離の節点に印をつけて次の点を確定。
2. 印をつけた節点からつながっているそれぞれの節点までの距離を求め、この時点で計算されている (印のついていない) 節点の距離の中で最小の値をもつ節点に印をつけてさらに次の点を確定。
3. 2を全ての節点に印がつくまで繰り返す。

Fig. 3に示すネットワークに対して、始点を1とした時にダイクストラ法を適用した例を Fig. 4に示す。Fig. 4では、ノードを表す丸の下に四角に、各stepの時点で求まっている始点からの最小距離を示している。四角の中でのXの表示は、その点までのルートがその時点において見つかっておらず、最大値が入っている状態を示す。

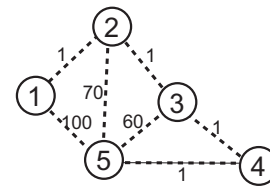


Fig. 3 example of Network

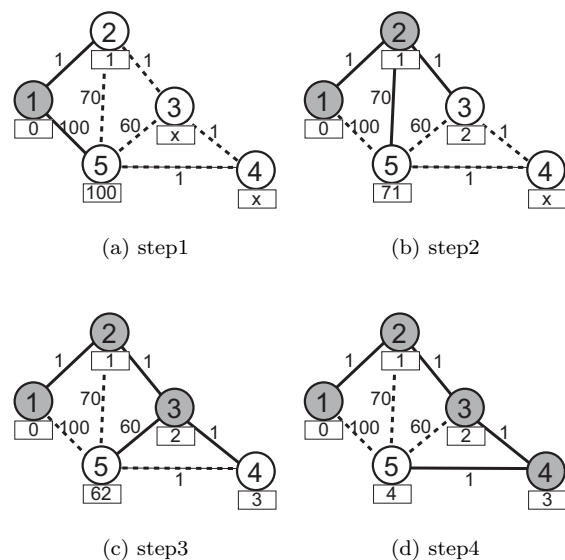


Fig. 4 ダイクストラ法の解法の例

4.2 ヒープ木

ヒープ木とは、完全2分木の各節に必ず親と子の大小関係が成立するようにキーを設定したのが、ヒープである。Fig. 5 にヒープ木の構成例を示す。

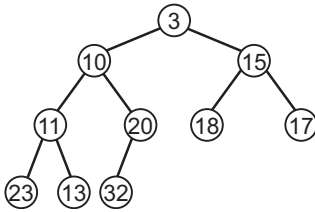


Fig. 5 ヒープの例

このヒープを利用することにより、ダイクストラ法を高速に処理することが可能となる。

4.3 ダイクストラ法における性能比較

これまで用いて全探索の手法、ダイクストラ法、ヒープ木を用いたダイクストラ法の性能比較を行う。対象問題として、TSPLIB に掲載されている bayg29, eil101, pr1002 を使用した。各問題のノード数、及び設定したリンク数を Table 1 に示す。なお、実験には、知的システムデザイン研究室の xenia クラスタを使用した。

Table 1 対象問題のノード数、および設定したリンク数

問題	ノード数	リンク数
bayg29	29	100
eil101	101	500
pr1002	1002	5000

各対象問題における実験結果をそれぞれ Table 2, Table 3, Table 4 に示す。なお、bayg29 における実験結果は、30 種類の計算時間の合計値を示し、eil101, pr1002 は 1 種類の計算時間を示している。

Table 2 bayg29 における各種法の計算時間

手法	計算時間 [s]
全探索	0.1560
ダイクストラ法	0.0160
ダイクストラ法 (ヒープ)	0.0780

Table 3 eil101 における各種法の計算時間

手法	計算時間 [s]
全探索	0.5470
ダイクストラ法	0.0150
ダイクストラ法 (ヒープ)	0.0470

Table 4 pr1002 における各種法の計算時間

手法	計算時間 [s]
全探索	4495.6260
ダイクストラ法	8.9060
ダイクストラ法 (ヒープ)	4.4370

Table 2, Table 3 より、ノード数が少ない場合において、ダイクストラ法が最も優れた性能を示した。Table 4 より、ノード数が多い時、ヒープを用いたダイクストラ法において最も優れた性能を示した。これは、ノード数が少ない時には、最短経路を求める時間よりもヒープ木の構成に時間がかかってしまっているためと考えられる。以上の事から、これまで用いていた探索法より、ダイクストラ法が優れている事が確認された。

5 数値実験

4 節で述べたダイクストラ法を用いることにより、これまでより問題サイズの大きいノード数 101 である eil101 を複雑ネットワーク問題として扱う。分散 GA に世代交代モデル sGA を適用した場合 (以降、場合により DGA+sGA と称す)、単一母集団 GA に世代交代モデル sGA を適用した場合 (以降、場合により SPGA+sGA と称す)、および単一母集団 GA に世代交代モデル MGG を適用した場合 (以降、場合により SPGA+MGG と称す) の性能比較実験を行う。

5.1 パラメータ

実験に用いたパラメータを Table 5 に示す。終了条件は評価計算回数が 1.5×10^7 を超えたときであり、試行回数は 30 とした。

5.2 実験結果

各対象問題における探索終了時 (最大評価計算回数) の評価値 (平均最短移動距離) を Fig. 6 に示す。Fig. 7 に、個体数 400 における DGA+sGA, SPGA+sGA および SPGA+MGG の解探索履歴を示す。Fig. 8 に、個体数 400・サブ母集団数 20・移住間隔 5・世代交代モデル sGA を適用した場合、および個体数 400・サブ母集団数 1・世代交代モデル MGG を適用した場合の解探索終了時のネットワークの構成図を示す。ここに示す構成図は各試行で得られた結果の代表的な一例である。

Fig. 6 より、世代交代モデルに sGA を適用した場合 (DGA+sGA, および SPGA+sGA)、サブ母集団数が多いほど解探索性能が優れていることが確認できる。一方、世代交代モデルに MGG を適用した場合 (SPGA+MGG)、最も優れた解探索性能を示している。また、個体数が 400 の場合に優れた解探索性能を示している。これは、個体数増加による多様性向上の影響であ

Table 5 パラメータ

パラメータ	DGA+sGA	SPGA+sGA	SPGA+MGG
個体数	200,400		
染色体長	$(ノード数) \times ((ノード数) - 1) / 2$		
交叉回数 (MGG)	-	-	25
交叉率	0.5		
サブ母集団数	5,10,20	1	1
突然変異率	1 / (染色体長)		
選択手法	トーナメント選択	トーナメント選択	ランキングルーレット選択
トーナメントサイズ	4	4	-
移住率	0.5	-	-
移住間隔	5,10,20	-	-
試行回数	30		

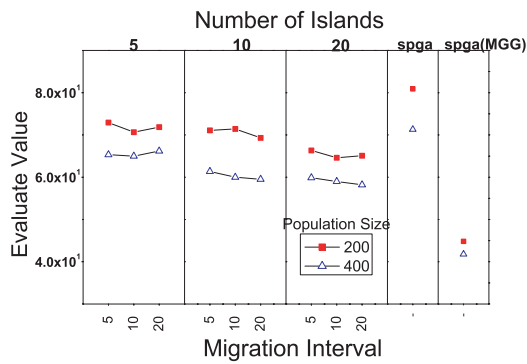
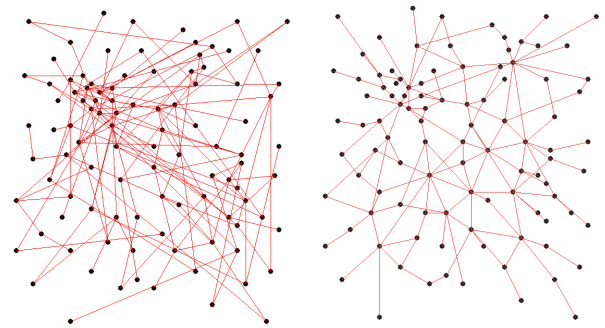


Fig. 6 Result of the Evaluation Value



(a) DGA+sGA

(b) SPGA+MGG

Fig. 8 Network composition

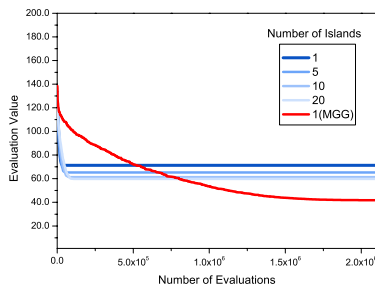


Fig. 7 History of the Evaluation Value

ると考えられる。Fig. 7より、MGGを適用した場合には、sGAと比較して、多様性の維持が図れたことにより、早熟収束せず、良好な解探索性能を示した。

6 まとめ

本稿では、より大きな問題サイズの複雑ネットワーク問題を解くために、計算時間のかかる評価方法について検討を行った。実験結果より、ダイクストラ法を用いることで、評価時間を削減することが可能であることが確

認できた。

また、次にダイクストラ法を用いて、これまでより問題サイズを大きくした場合の実験を行った。実験結果より、分散GAの分散効果による多様性の維持は効果がなく、世代交代モデルにMGGを用いた場合は多様性の維持が図れ、良好な結果を示した。

今後は、コーディング、交叉方法、初期個体の発生方法の検討を課題とする。

参考文献

- 1) A.-L. Barabási, *Linked: The New Science of Network Perseus Books*, NHK 出版, 2002.
- 2) Nicolis, G. and Prigogine, I, *Exploring complexity*, R. Piper GmbH and Co. KG Verlag, 1989.
- 3) D.E. Goldberg, *Genetic Algorithms in Search Optimization and Machine Learning*, Addison-Wesley, 1989.