

進化的ネットワークにおける探索アルゴリズムの提案

緒方 司 (Tsukasa Ogata) * 小野 廣隆 (Hirotaka Ono) †
 定兼 邦彦 (Kunihiko Sadakane) † 山下 雅史 (Masafumi Yamashita) †

概要

グラフにおいて起点頂点より辺で繋がった頂点をたどり、できるだけ少ない頂点を經由して目標頂点に到達したいという問題がある。小規模なグラフではダイクストラ法等のアルゴリズムにより最短経路を求めることができるが、大規模なグラフでは全頂点を訪問することになりうるアルゴリズムは、必要な記憶量が膨大になるため実質上適用不可能である。そこで限定された局所的な情報だけを用いて目標頂点に到達する手法が必要となる。本論文ではスモールワールドネットワーク [6], スケールフリーネットワーク [1] という特定の構造を持つグラフを対象とする。

本論文の目的は、これら2つのネットワーク上で局所情報と記憶 (メモリ) のみを用いた、起点頂点から目標頂点に到達する探索アルゴリズムを提案することである。本論文では k -タブーと l -先読みという遷移と記憶に関する2つの能力を、池田らによる隣接する頂点の次数情報を用いるランダムウォーク [4] と組み合わせた探索アルゴリズムを提案する。そして、提案手法の能力を計算機シミュレーションによって検証する。

1 はじめに

多くの自然界, 人工物, 社会におけるシステムは、頂点と頂点同士を繋ぐ辺からなるグラフによって表現可能である。例えば WWW では頂点を Web ページ, 辺をハイパーリンクとすることにより、グラフとして表すことができる。グラフにおいて、起点頂点

より辺で繋がった頂点をたどり、できるだけ少ない頂点を經由して目標頂点に到達したいという問題がある。これは、情報をなるべく速く目標まで伝えたい、物質の輸送にかかるコストを小さくしたい、なるべく小さいコストで目標まで移動したい等の現実的な要求に対応する。全情報が分かる小規模なグラフでは、ダイクストラ法などにより最短経路を求めることができる。しかし大規模なグラフにおいては、全情報を得ることは困難であり、また仮に全情報が得られたとしても、全頂点を訪問することになりうるアルゴリズムは記憶量が膨大になるため実質上適用不可能である。そこで限られた局所的な情報だけを用いて目標頂点に到達する手法が必要となる。本論文では近年注目を集めている、スモールワールドネットワーク [6], スケールフリーネットワーク [1] という、実在する特定の構造を持つグラフを対象とする。

スモールワールドネットワークとは頂点がクラスタ状に集まっているのにも関わらず、頂点間の距離が短いという特徴を持つグラフである。またスケールフリーネットワークとは、頂点の次数の分布がべき乗則に従う特徴をもつグラフである。

本論文の目的は、これら2つのネットワーク上で局所情報と記憶 (メモリ) のみを用いた、起点頂点から目標頂点に到達する探索アルゴリズムを提案することである。過去の研究ではこの目的のために、一つ前の頂点に戻らない no-back walk や三角形のサイクルを避ける no-triangle-loop walk 等の記憶に関する能力と、隣接頂点に目標頂点が含まれているなら決定的に移動するという遷移に関する能力を、特定の遷移確率を持つランダムウォークに付加した様々な探索手法が提案されている [3, 5, 7]。本論文では記憶と遷移に関する2つの能力と、ランダムウォークの遷移確率をより汎用的に扱える新しい手法を提案する。このために、 k -タブーと l -先読みを考える。 k -

*九州大学大学院システム情報科学府, Department of Computer Science and Communication Engineering, Graduate School of Information Science and Electrical Engineering, Kyushu University

†九州大学大学院システム情報科学研究院, Department of Computer Science and Communication Engineering, Graduate School of Information Science and Electrical Engineering, Kyushu University

タブーとは、直前に通過してきた k 個の頂点を記憶しておき、それらの頂点には遷移しないという手法である。 l -先読みとは、距離 l 以内にある頂点を探索し、目標頂点が含まれているなら必ず目標頂点に決定的に移動する能力である。この2つの手法を、池田らによる隣接する頂点の次数情報を用いるランダムウォーク [4] と組み合わせた探索アルゴリズムを本論文では提案する。そして、スモールワールドネットワークのモデルである WS モデル [6]、スケールフリーネットワークのモデルである BA モデル [1]、比較の対象としてランダムネットワークのモデルである ER モデル [2] において、提案手法の能力を計算機シミュレーションによって検証する。

2 準備

2.1 SW ネットワークと SF ネットワーク

2.1.1 スモールワールドネットワーク

スモールワールドとはもともと、社会心理学で考えられていた、人の繋がり大きさは直観的なイメージよりもはるかに小さいという概念のことをいう。長い間、社会心理学での研究対象であったスモールワールドであるが、1998年に Watts がグラフにおける L と C という2つの特徴量を用いて定式化して以来 [6]、グラフの分野でも注目されるようになった。 L と C の定義は以下の通りである。

- L : 平均経路長 (characteristic path length)
グラフにおけるすべての頂点の組についての距離の平均。
- C : クラスタ係数 (clustering coefficient)
ある頂点 v に対して隣接する頂点数を v_i とする。この v_i 個の頂点間に存在する辺数を $\binom{v_i}{2}$ で割ったものを求め、全頂点の平均をとったものである。

頂点数と辺数が一定のグラフにおいて、通常 L と C は対応しており、 L が大きいなら C も大きく、 L が小さいなら C も小さいという傾向がある。しかし、Watts の定式化によると、スモールワールドネットワークとは $L \approx L_{rand}$ にも関わらず $C \gg C_{rand}$ であるグラフのことを言う。なお、 L_{rand} , C_{rand} とはランダムグラフの L, C のことで、共に小さな値である。つまりスモールワールドネットワークとは、頂点間の平均距離が短いにも関わらず、頂点がクラス

タ状に集まっているグラフのことをいう。この定式化により自然界、人工物、社会における様々なネットワークがスモールワールドネットワークであることが分かっている。

2.1.2 スケールフリーネットワーク

スケールフリーネットワークは1999年に Barabási によって発見された、頂点 v の次数 $\deg(v) = k$ の分布がべき分布 (power law)

$$P(k) \sim k^{-\gamma}$$

に従うネットワークのことをいう [1]。 γ は定数で通常 $2 < \gamma < 3$ である。つまりスケールフリーネットワークでは、 k の値が小さい頂点が多く存在する一方で、極端に k の大きい頂点も少ないながらも存在している。Barabási はこのような特徴を持つネットワークが、以下のような単純で自然な生成規則によって生まれることを明らかにした。

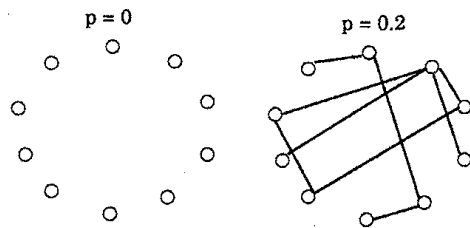
- Growth :
新しい頂点が追加されながら、時間と共にネットワークが成長していく。
- Preferential Attachment :
次数の高い頂点ほど、新しい辺が張られやすい。(rich-get-richer)

例えば WWW において、新しく Web ページを作成したときには、すでに存在している有名なページにリンクを張ることが多く、また論文の引用においても、新しく論文を書くときにはすでに知られている有名な論文を引用することが多いと考えられる。このように時間の経過に従って頂点が増加していき、次数の高い頂点がさらに辺を獲得して次数が高くなりやすいネットワークは、スケールフリーネットワークとなる。

3 ネットワークのモデル

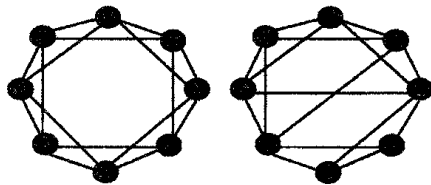
3.0.3 ER モデル

ER モデルは Erdős と Rényi によって提案されたランダムネットワークのモデルである [2]。ER モデルでは N 個の孤立点を初期値に与える。そして、任意の2頂点間に等しい確率 p で辺を張る。 N が十分に大きいとき ER モデルでは、頂点の次数分布がポアソン分布に従うことが知られている。

図 1: ER モデル, $N = 10, p = 0, 0.2$

3.0.4 WS モデル

WS モデルは Watts と Strogatz によって提案された, スモールワールドネットワークのモデルである [6]. WS モデルはまず, 近傍の k 頂点 ($k = 2l, l = 1, 2, \dots$) にのみ辺を張った正則な束グラフ (regular lattice graph) を生成する. 次に, 確率 p で辺の繋ぎかえを行う. 辺の張りかえとは, ランダムに選んだ頂点に辺を張りかえる操作である. $p = 0$ のときは辺の張りかえを全く行わず, 元の束グラフのままである. $p = 1$ の時はすべての辺を張りかえることになり, グラフはランダムグラフとなる. $0 \leq p \leq 1$ において, 近傍を繋ぐ辺により C の値が大きいままで, 張りかえられた辺がショートカットの役割をすることにより, L の値が急激に小さいスモールワールドとなる.

図 2: 束グラフ, $n = 8, k = 4$, WS モデル $n = 8, k = 4, p = 0.1$

3.0.5 BA モデル

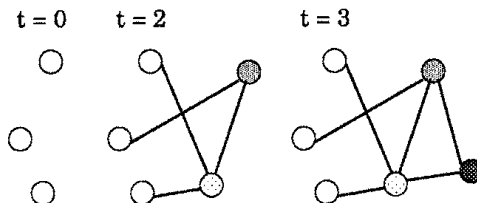
BA モデルは Barabási と Albert により提案されたスケールフリーネットワークのモデルである [1].

- Growth: 初期状態として, 小さな m_0 個の頂点を与える. 各時間ステップごとに 1 つの新頂点と新頂点から既存の頂点へ $m (\leq m_0)$ 本の辺をグラフに加える.
- Preferential attachment: 新頂点から既存の頂点に辺を張るとき, 既存の頂点において頂点 i が次数 k_i を持つならば, 辺の張られる既存の頂点は以下に示す確率 Π で選ばれる. このとき, 多

重辺が存在しないようにする.

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

本論文で考える BA モデルでは, $t = 0$ のとき既存の頂点に辺の張られる確率が 0 となるのを避けるため, 初期状態として m_0 個からなる完全グラフを与える.

図 3: BA モデル, $m_0 = 3, m = 2, t = 0, 2, 3$

4 頂点探索手法

4.1 過去の研究

グラフの全情報を用いた探索手法として, ダイクストラ法等が知られている. グラフが小規模ならばこれらの手法を用いて最短経路を求めることができるが, 実在する大規模ネットワークでは, 記憶量の問題から全情報を用いる実質適用不可能となる. そこで局所情報のみを用いるランダムウォークのような手法が必要となる. 記憶と遷移に関する能力をスタンダードランダムウォークを付加した no-back walk [7] や self-avoiding walk [3, 7] 等の手法がすでに提案されている. 過去の提案手法の大部分が, 記憶と遷移に関する能力を特定の遷移確率を持つランダムウォークに付加したものである. そこで本論文では記憶と遷移に関する能力, それにランダムウォークの遷移確率をより汎用的に扱える新しい手法を提案する.

4.2 提案手法

4.2.1 (k, l, β) - ランダムウォーク

以下の 2 つの能力を考える. この 2 つの能力は局所情報のみを用いると考えられ, k と l が小さい定数であるなら, 任意の頂点探索手法に組み合わせて使用することが可能である.

- k -タブー: グラフ上を移動する際に, 直前に訪問した k 個の頂点をタブーリストの中に記憶しておき, タブーリスト内の頂点には移動しないようにする手法である. ただし, 隣接する頂点

がすべてタブーリストに含まれているなら、タブーリスト内の頂点に遷移する。

- l -先読み: 現在訪問している頂点から距離 l 以内にある頂点を探索し、もし目標頂点が含まれているならば、目標頂点に決定的に遷移する手法である。

本論文では上記した2つの手法を β -シンプルランダムウォーク [4] と組み合わせ、 (k, l, β) -ランダムウォークを提案する。

4.2.2 β -シンプルランダムウォーク

β -シンプルランダムウォークでは、局所情報として隣接頂点の次数情報を使用する。任意の n 個の頂点を持つ有限無向グラフ $G = (V, E)$ において、隣接する頂点の次数情報を用いて、遷移確率 $(p_{u,v}^{(\beta)})_{u,v \in V}$ を以下のように定義する [4]。

$$p_{u,v}^{(\beta)} = \frac{\deg^{-\beta}(u)}{\sum_{w \in N(u)} \deg^{-\beta}(w)}$$

ここで $N(u)$ は頂点 u に隣接する頂点集合である。 β -シンプルランダムウォークでは、 β の値が大きいほど次数の低い頂点に、 β の値が小さいほど次数の高い頂点に遷移しやすくなり、 $\beta = 0$ のときは隣接頂点に等確率で遷移するスタンダードランダムウォークである。このランダムウォークでは、 $\beta = 0.5$ のとき任意の有限無向グラフに対して Hitting Time が $O(n^2)$ 、Cover Time が $O(n^2 \log n)$ であることが証明されている [4]。

提案手法と過去に提案された手法との比較を表 1 に示す。提案手法はそのパラメータを調整することにより、過去に提案された手法と同様のアルゴリズムを実現することができる。

表 1: 提案手法と過去の手法との比較

過去の手法	(k, l, β)
no-back-walk [7]	$(1, 1, 0)$
no-triangle-walk [7]	$(2, 1, 0)$
no-quadrangle-walk [7]	$(3, 1, 0)$
self-avoiding-walk [3, 7]	$(\infty, 1, 0)$
preferentially self-avoiding-walk [7]	$(\infty, 1, -1)$
maximum connectivity first [5]	$(0, 1, -\infty)$

5 計算機実験

5.1 パラメータの設定

$(k, l, \beta) = (\{0, 1, 2, \infty\}, \{0, 1, 2\}, \{-1, -0.5, 0, 0.5, 1\})$ の組み合わせに対して、 (k, l, β) -ランダムウォーク

により実験を行う。ランダムに選んだ起点頂点を一つ固定し、起点頂点から目標頂点に到達するまでに通過した辺の数 (ステップ数) の 50 回平均を計測する。目標頂点は起点頂点以外のすべての頂点とする。起点頂点からすべての目標頂点に対して平均ステップ数を求め、その平均値を測定する。以下、平均ステップ数の平均値を単にステップ数と呼ぶ。対象とするグラフは 3 節で述べた ER モデル、WS モデル、BA モデルによるものである。また、各モデルにおいてグラフのサイズをなるべく等しくするため、頂点数 n のとき辺数が $2n$ となるように各パラメータの設定を行った。各モデルで用いた実際のパラメータは以下の通りである。

- ER モデル: 頂点数 n 、辺数が $2n$ となるように、 $p = \frac{4}{n-1}$ 。
- WS モデル: 頂点数 n 、近傍への辺数 4。また、辺の張りかえる確率 p は、グラフがスモールワールドネットワークとなる $p = 0.1$ 。
- BA モデル: 初期頂点数 2、時間ステップごとに加えられる辺数 2、時間ステップ数 $t = n$ 。

実験では $n = \{25, 30, 35, 40, 45, 50\}$ のときに 30 個のグラフの平均を取った。

6 実験結果と考察

6.0.1 全体的な傾向

$(k, l, \beta) = (1, 0, \{-1, -0.5, 0, 0.5, 1\})$ のときの実験結果の例を図 4~図 6 に示す。横軸は頂点数、縦軸はステップ数である。この実験結果の例ではいずれのパラメータにおいても、ステップ数は頂点数に対してほぼ線形で増加していた。今回得られた実験結果では、他のパラメータの場合においても同様であった。

$n = 50$ のときの $\beta = \{-1, -0.5, 0, 0.5, 1\}$ の実験結果を表 2~表 6 に示す。ステップ数は各モデルとも $l = 0$ 、つまり先読みを使用しないときは $\beta > 0$ の方がステップ数が少なく、 $l > 0$ 、つまり先読みを使用すると $\beta < 0$ の方がステップ数が少ない。これは今回採用したグラフ、パラメータに対して全体的にみられる傾向である。

6.1 パラメータによるステップ数の変化

実験結果に対してパラメータ k, l, β がステップ数に与える影響について考察する。

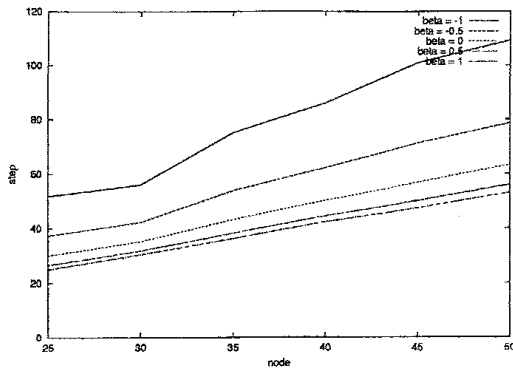


図 4: ER : $(k, l, \beta) = (1, 0, \{-1, -0.5, 0, 0.5, 1\})$

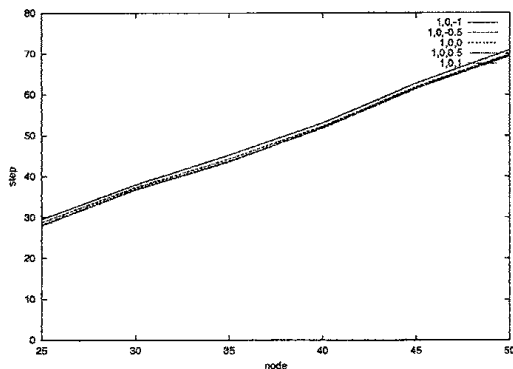


図 5: WS : $(k, l, \beta) = (1, 0, \{-1, -0.5, 0, 0.5, 1\})$

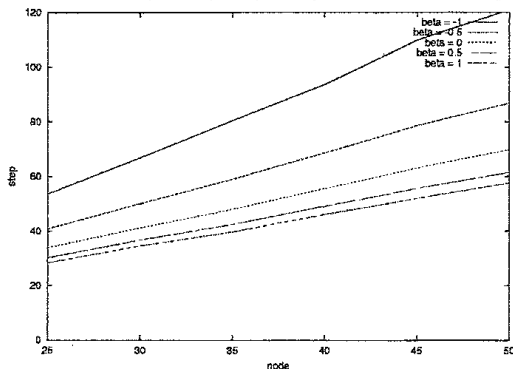


図 6: BA : $(k, l, \beta) = (1, 0, \{-1, -0.5, 0, 0.5, 1\})$

6.1.1 β によるステップ数の変化

- $k = 0, l = 0$

このとき提案手法は β -シンプルランダムウォークとなる。これは一般の有限グラフにおいて、 $\beta = 0.5$ のときに Hitting Time が最小となる β -シンプルランダムウォークの特徴がそのままになっていると思われる。また ER モデル、BA モ

表 2: $n = 50, \beta = -1$

	$l = 0$	$l = 1$	$l = 2$
$\beta = -1$	ER WS BA	ER WS BA	ER WS BA
$k = 0$	188 109 186	30 57 14	7 28 3
$k = 1$	112 71 121	18 35 11	5 17 3
$k = 2$	109 57 111	18 28 11	5 15 3
$k = \infty$	34 36 35	12 20 8	5 12 3

表 3: $n = 50, \beta = -0.5$

	$l = 0$	$l = 1$	$l = 2$
$\beta = -0.5$	ER WS BA	ER WS BA	ER WS BA
$k = 0$	118 107 116	26 60 15	7 29 4
$k = 1$	79 69 87	17 36 12	5 17 3
$k = 2$	79 55 84	17 28 11	5 15 3
$k = \infty$	33 37 35	12 21 9	5 12 4

デルでは β によってステップ数が大きく異なっているのに対し、WS モデルでは β によるステップ数の変化はあまりみられない。これは WS モデルが、正則グラフに近い特徴を持っているためと考えられる。WS モデルにおいてグラフは、束グラフと呼ばれる正則なグラフから確率 p で辺を張りかえることにより作られる。今回の実験ではスモールワールドネットワークとなるように $p = 0.1$ としたために、総辺数の 10% 程度しか辺が張りかえられておらず、結果として得られたグラフは比較的正則グラフに近い特徴を持っていると考えられる。これを確かめるために $n = 50$ における次数の分散を計測した。実際の次数の分散を表 7 に示す。WS モデルは次数の分散が他のモデルよりも小さいことが分かる。正則グラフにおいて β -シンプルランダムウォークの遷移確率は、すべての頂点の次数が等しいために β によらず一定である。

- $k = \{1, 2, \infty\}, l = 0$

ステップ数は $\beta < 0$ よりも $\beta > 0$ の方が小さい。これは提案手法の基にした β -シンプルランダムウォークの特徴がある程度保存されていることを示している。 $k = 0$ から $k = 1$ としたときに

表 4: $n = 50, \beta = 0$

	$l = 0$	$l = 1$	$l = 2$		$l = 0$	$l = 1$	$l = 2$
	ER WS BA	ER WS BA	ER WS BA		ER WS BA	ER WS BA	ER WS BA
$\beta = 0$				$\beta = 1$			
$k = 0$	91 107 94	26 60 17	7 29 4	$k = 0$	93 111 113	35 64 33	10 32 6
$k = 1$	63 69 70	18 36 12	6 17 4	$k = 1$	53 70 58	20 37 15	6 18 4
$k = 2$	63 55 69	17 28 13	5 15 4	$k = 2$	53 54 57	19 29 15	6 16 4
$k = \infty$	33 37 36	13 21 10	5 12 4	$k = \infty$	34 37 40	14 21 12	6 13 4

表 6: $n = 50, \beta = 1$ 表 5: $n = 50, \beta = 0.5$

	$l = 0$	$l = 1$	$l = 2$
	ER WS BA	ER WS BA	ER WS BA
$\beta = 0.5$			
$k = 0$	85 108 95	28 61 23	8 30 5
$k = 1$	56 69 61	19 37 14	6 18 4
$k = 2$	56 54 61	18 29 14	5 15 4
$k = \infty$	33 37 37	13 21 11	5 12 4

は、どのモデルにおいてもステップ数にかなりの改善がみられる。 $k = 1$ のときは直前に訪問していた頂点がタブーリストに挿入されるため、次の遷移先を決定する際に後戻りをしなくなるからである。

一方、 $k = 1$ から $k = 2$ としたときは ER モデル、BA モデルに関してはステップ数の改善はほとんどみらず、WS モデルに関してのみある程度のステップ数の改善がみられる。これは WS モデルのクラスタ係数 C が高いためであると考えられる。すなわち、 C が小さいグラフにおいては、2 ステップ前の頂点は既に現在訪れている頂点とは無関係の位置にあるのに対して、 C が大きいグラフでは 2 ステップ前の頂点も現在訪れている頂点に隣接している可能性が高く、タブーの効果があられやすいためである。実際の C の値を 7 に示す。 $k = \infty$ としたとき、ステップ数は β のによらずほぼ同じ値となっている。このことは、大きな k の値を用いたランダムウォークは、実際にはタブーリストに存在しない頂点への決定的な探索アルゴリズムと変わらない振舞をしていることを示唆するものである。

- $k = \{1, 2, \infty\}$, $l > 0$

ステップ数の大幅な改善がみられる。ステップ数は依然として β に依存しているが、 $l = 0$ の場合と比較するとその差は遥かに小さくなっている。先読みの効果は BA モデルが一番大きく、次に ER モデル、最後に WS モデルの順である。これは、平均パス長 L の大きさに関係していると考えられる。 L が小さいグラフほど先読みのときに、より多くの頂点を探索するからである。実際の L の値を表 7 に示す。

またステップ数は $\beta > 0$ よりも $\beta < 0$ の方がわずかながら小さい。これは次数の高い頂点の方が先読みによって、平均的にはより多くの頂点を探索できるので目標頂点に到達しやすいためだと考えられる。 k によるステップ数の変化の傾向は、 $l = 0$ のときとほぼ同じである。

表 7: $n = 50$: 次数の分散, C, L

	ER	WS	BA
次数の分散	180	18	561
C	0.08	0.39	0.19
L	2.85	3.96	2.66

6.1.2 k によるステップ数の変化

k が小さいとき、ステップ数は β の影響を受けるが、 $k = \infty$ とすると β の影響がほとんどみられない。そこで、 k をどの程度大きくすれば β の影響を受けなくなるのか調べるために、 k によるステップ数の変化を測定した。頂点数 50, $l = 0$ のときの BA モデルにおける実験結果を図 7 に示す。横軸がタブーリストのサイズ k 、縦軸がステップ数である $k = 0$ から $k = 1$ にしたときにステップ数が大きく減少し、 $k \geq 2$ では $l = 0$ では $k = 30$, $l = 1$ では $k = 25$ ぐ

らいまでステップ数が緩やかに減少していき、それより大きい k においてステップ数は減少しない。

他のモデルにおいても同様の結果が得られた。実際にステップ数が減少しなくなったときの k の値を表8に示す。タブーリストのサイズが大きくなるにつれ、提案手法はタブーリストに存在しない頂点への決定的な遷移を行う探索アルゴリズムに近付くために、 β によらずステップ数が等しくなったものと考えられる。

表 8: タブーリストのサイズ k

	ER	WS	BA
$l = 0$	40	25	30
$l = 1$	30	20	25

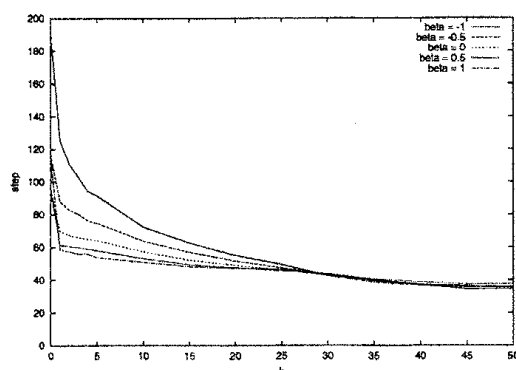


図 7: BA モデル: $(l, \beta) = (0, \{-1, -0.5, 0, 0.5, 1\})$

7 おわりに

従来の局所情報に基づく探索アルゴリズムよりも、記憶と遷移に関する2つの能力とランダムウォークの遷移確率を汎用的に扱える手法を提案した。記憶に関する k -タブーと遷移に関する l -先読みという2つの能力を考え、その2つの能力を β -シンプルランダムウォークに付加することにより (k, l, β) -ランダムウォークという局所情報に基づく探索アルゴリズムを提案した。また、ランダム、スモールワールド、スケールフリーの各特徴をもつグラフを生成し、計算機シミュレーションによりその性能の評価を行った。実験結果よりタブーはクラスタ係数の大きいグラフに、先読みは平均経路長の小さいグラフにより有効であり、タブーリストのサイズが十分に大きければ遷移確率によらず、ステップ数はほぼ等しくなることを示した。

今後は提案手法の実在するネットワーク上における能力の評価、それに理論的な解析を行う予定である。

参考文献

- [1] A.-L.Barabási, R.Albert, and H.Jeong, Mean-field theory for scale-free random networks, Physica A, Vol.272, pp.173-187, 1999.
- [2] P.Erdős, and P.Rényi, On the Evolution of random graphs, Publ.Math.Inst.Hung.Acad.Sci., Vol.5, pp.17-61, 1960.
- [3] C. P. Herrero, Self-avoiding walks on scale-free networks, Physical Review E, Vol 71, pp.1-8, 2005.
- [4] S.Ikeda, I.Kubo, and M.Yamashita, Reducing the Hitting and the Cover Times of Random Walks on Finite Graphs by Local Topological Information, Proceedings of The 2003 International Conference on VLSI, pp.203-207, 2003.
- [5] B.I.Kim, C.N.Yoon, S.K.Han, and H.Jeong, Path finding strategies in scale-free networks, Physical Review E, Vol.65, pp.1-4, 2002.
- [6] D.J.Watts, and S.H.Strogatz, Collective dynamics of small-world networks, Nature, Vol.393, pp.440-442, 1998.
- [7] S.Yang, Exploring complex networks by walking on them, Physical Review E, Vol 71, pp.1-5, 2005.