

Web コミュニティ探索のための量子アルゴリズムの提案と検証

大音 真由美[†] 中 條 拓 伯^{††} 城 和 貴[†]

web 利用の増加にともなって Internet に接続するノード数が急激に増加し、web コンテンツの効率的な選択の重要性が高まっている。現在、web コンテンツの分析および分類には web コミュニティの概念が広く用いられているが、膨大な数の web ページのために単純なキーワード検索では web コミュニティの発見が困難である。web コミュニティの探索は、巨大なデータ群から少数の解を効率的に取り出す点において、量子コンピューティングの持つ無限並列性を利用するのに適している。我々の量子コンピュータ研究の目的は、将来の量子コンピュータに対するキラアプリケーションの提示にある。本論文ではまず、既存のデータベース検索の量子アルゴリズムを用いて web コミュニティ探索アルゴリズムを提案する。次に、web コミュニティの特徴を検討し、実際の web データを用いて実験し、提案したアルゴリズムの検証を行う。最後に、我々がすでに提案している量子コンピュータの命令セットを用いて前述のアルゴリズムのコーディング例を与える。

A Quantum Algorithm to Search for Web Communities and Inspection

MAYUMI OTO,[†] HIRONORI NAKAJO^{††} and KAZUKI JOE[†]

With increase of web pages, the importance of effective choices of web contents rises quickly. Currently, the concept of web community is used for analysis of web contents and classification in general, but the search for web community is quite difficult just with a simple keyword scheme because of the huge number of web pages. The search for web community is suitable for infinite parallelism that quantum computing potentially provides. In this paper, we aim at proposing a killer application for quantum computers. First, we apply an existing quantum algorithm to a conventional search method to propose a web community search algorithm and describe the features of web community. Next, we experiment with the quantum algorithm for database search using actual web data and evaluate the result. In addition, we give an example of program code for the algorithm with the quantum instruction set which we have proposed and show the effectiveness of the algorithm.

1. はじめに

近年の Internet の爆発的成長にともない、web の利用は今や日常である。同時に、情報過多にある Internet の中で、web コンテンツは大規模分散化の一途をたどる。したがって、web コンテンツの有効な流通・提供・利用がますます困難になり、効率的な選択の重要性が高まっている。web コンテンツの流通状況を分析および分類するための指針には、web コミュニティの概念が広く用いられている¹⁾。web コミュニティとは、web ページのオーナーが共通の興味を持つ他の web ページをリンクしてできる web ページとリンクとの集合体である。それらは様々な大きさや形態を持ち、一般的な数学的モデルは提案されていない。

したがって、現在主流のキーワード検索では web コミュニティの発見が困難であり、別のアプローチによる web コミュニティの探索方法が hot spot (人々の関心を集める研究分野)として各方面で研究されている。

web コミュニティの分類には、web コミュニティを web ページの集合として数値的にとらえる方法と、トポロジーなど図形的にとらえる方法とがある。

前者の方法は、web ページ間のリンクの接続情報を表現した隣接行列の固有ベクトルを利用して、hub と authority の価値を各 web ページの重みとして表現する手法である^{2),3)}。authority は、ある topic を扱う web ページの集合における良質な ページで、hub はそれらへリンクしている良質な ページである。また、hub と authority とは互いに補強する関係にある⁴⁾。後者の方法は、web コミュニティのリンクトポロジー

[†] 奈良女子大学

Nara Women's University

^{††} 東京農工大学

Tokyo University of Agriculture & Technology

ある topic に属する多くのページからリンクされている情報源
ある topic に属する多くのページへのリンク集

の解析および視覚化^{5),6)} や, web コミュニティを巨大なデータ群と見なし web mining やグラフ解析を行う手法である^{7)~9)} .

現在 Internet に接続されている web ページの総数は現在のスーパーコンピュータをもってしても解析可能とは想定されないほど多いため, これらはマクロまたはミクロ視点のどちらかに焦点を当てた研究である. ところが隣接行列に対する他の解析手法を用いると, 双方の視点での分析が可能になる. しかし, ノイマン型コンピューティングでは数千万次元に達すると予想される隣接行列を用いて, マクロおよびミクロの視点から解析できる効率的なアルゴリズムの設計はきわめて困難であると予想される.

量子コンピューティングは, データ群の大きさの増加にともなって計算時間が指数的に増大するアルゴリズムにおいて, ノイマン型コンピューティングよりも適していることが知られており, その代表的な例が素因数分解¹⁰⁾, およびデータベース検索アルゴリズム¹¹⁾ である.

web コミュニティの探索および解析においては, 任意のページのリンク先だけでなく, リンク元を知ることも重要な情報源となる. しかし, ノイマン型コンピューティングの枠内では, リンク元を検索するには膨大な時間が必要となる. 量子コンピューティングの利点は, リンク先を探索するのと同じ時間でリンク元も検索できることにある.

我々は, 量子コンピューティング・システムの構築を目標としている. そのために, まず文献 12) において, 量子コンピュータ・アーキテクチャならびに量子命令セットを提案し, 既存の量子アルゴリズムの実際のコーディング例を示した. 次に, 文献 13) において web コミュニティ探索のアルゴリズムを提案し, 小規模な実験を行い, その有効性を確認した. 我々の研究の目的は, 量子コンピュータが実現したときのキラアプリケーションの提示にあるため, 本論文では文献 13) のアルゴリズムに実際の web ページの集合を与え, 意味のある web コミュニティが抽出できることを示す.

まず web コミュニティの定義を与え, その特徴を述べる. 次に, 上記のアルゴリズムの拡張および実際の web データを用いて実験を行い, 結果の解析および検証を行う. そのために, まず任意の web ページに関連した web コミュニティ探索(以下, WCS)アルゴリズムを提案する. これは, 既知の手法と既存の量子アルゴリズムとを組み合わせたものである. さらに, web データを隣接行列に変換し, web コミュニティを探索

するアルゴリズムの実験を行い, その有効性を検証する. そのアルゴリズムの過程においてデータベース探索の量子アルゴリズム¹⁴⁾ を用いる. また, 量子コンピュータを実行するための量子命令セット¹²⁾ を用いて量子アルゴリズム部分のコーディング例を与える.

以下, 2 章では web コミュニティの特徴について述べる. 3 章では, WCS アルゴリズムを提案する. 4 章では, その実験結果の解析を行う. 5 章では, 3 章で述べたアルゴリズム中に用いる既存の量子アルゴリズムについて述べる. また, 我々がすでに提案している量子コンピュータの命令セットについて概説し, 前述の量子アルゴリズムのコーディングを行う. 6 章では, 結論と今後の課題を述べる.

2. web コミュニティの定義と特徴

本章では, web コミュニティと隣接行列ならびにグラフとの関連について述べる.

web ページの集合は, ネットワーク構造で矛盾なく記述できるため, その解析を行うには隣接行列を用いることとする. 隣接行列とはノード (web ページ) 間のエッジ (リンク) の接続情報を表現した行列である. 本論文では, web ページ p_i から p_j へリンクしているとき, p_j をリンク先, p_i をリンク元と表現する. またこのとき, web ページの総数を N とすると, 隣接行列 A は $N \times N$ の正方行列となり, その行列要素は $A_{ij} = \delta_{ij}$ ($i, j = 1, \dots, N$: p_j から p_i へのリンクがあれば 1, なければ 0) である.

以下, web ページの集合の規模によって, web コミュニティとリンク数および連結度の関連を示す.

web コミュニティは Internet に接続された web ページの部分集合であり, 隣接行列でそれらの web ページ番号が連続するように表現したとき, 対角成分上に部分正方行列を形成する.

本論文では, 辺連結度を連結度と表現する. 辺連結度とは, グラフ G において G が k - 辺切断を持つような最小の k をいう. k - 辺切断は k 個の辺からなる G の辺切断である. web ページの集合の連結度が最大のときはすべての web ページが完全グラフを形成し, 最小のときはどのページとも接続しない孤立化した状態にある. これは隣接行列において, すべての行列成分が 1, および 0 のときに相当する.

図 1 は隣接行列の i 行 i 列における行列要素と連結度との関連を示す. A の領域は, i 行目が所属する web コミュニティ内部で, その領域内での 1 の数は web コミュニティ内部でのリンク数であり, これが連結度となる. したがって B の領域は web コミュニティ

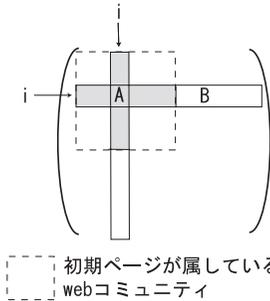


図 1 隣接行列とリンク数および連結度との関連

Fig. 1 Relation between adjacency matrix and both of the number of links and connectivity.

の外部であり， A と B での 1 の数の和がリンク数となる。

本論文では，各 web ページが連結してできる集合体を web の小集団と定義する．小集団は hub や authority，およびその他のページから構成される．また，各小集団が連結し，大集団を形成するものとする．

隣接行列においては，行列要素の 1 の密度より集団の規模が推測できる．一般に小集団になるほど，1 の密度が高くなるからである．

hub や authority を含む web 小集団は，その定義からリンク数は多く連結度も高い．これに対して，大集団の web ページは，リンク数は少なく連結度も低い⁷⁾．このように，各集団の web ページはそれぞれ特徴を持っている．リンク数と連結度から見た hub や authority を含む web 小集団とその他のページとの違いをまとめたものを表 1 に示す．

次に，グラフ理論の観点から見た web コミュニティの特徴について述べる．web ページを十分に持つ小集団は web コミュニティの核となる完全 2 部グラフを含むという前提のもとでコミュニティを探索した研究があり⁹⁾，大集団では，Bow-Tie のような形になるという研究報告がある⁷⁾．Bow-Tie の結び目の部分では web ページの相互接続が進んでいるため，リンクをたどることでのどのページにも到達できる．これがコアページとなる．起点ページは，コアページへのリンクを提供し，到達点ページはコアページからのリンク先となっている．web は大まかには，これら 3 つの部分から成り立っている．web コミュニティはそれらの中間にあたる集団であり¹⁵⁾，それ自身も階層構造となっている場合がある．上記の研究をもとにした web のモデルを図 2 に示す．

$C1$ から $C7$ はそれぞれ web コミュニティで，Bow-Tie の構成部分である．また，web ページ数と密度とを十分に持つ web コミュニティは完全 2 部グラフを

表 1 web ページの規模とリンク数および連結度との比較
Table 1 Comparison with the scale of web pages and both of links and connectivity.

web コミュニティの規模	リンク (リンク元) 数	連結度
hub, authority を含む小集団	多	高
その他の小集団	少	低
大集団	少	低

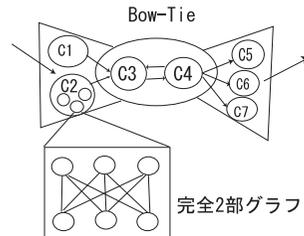


図 2 web コミュニティのモデル
Fig. 2 A model of web community.

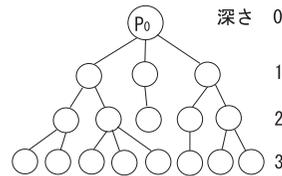


図 3 探索木の構造
Fig. 3 A tree structure.

少なくとも 1 つは持っている。

3. WCS (Web Community Search) アルゴリズム

本章では，web コミュニティを隣接行列を用いて抽出する WCS アルゴリズムを提案する．WCS アルゴリズムは，web コミュニティの大きさに対して，パラメータとして与える深さ上限値が小さい場合は，一部の web コミュニティを出力して終了する．深さ上限値が十分大きい場合は，web コミュニティ中のすべての web ページを出力して終了する．まず，web コミュニティ探索のアルゴリズムで用いたパラメータについて述べる．

- 初期ページ

本論文では，ある web ページに関連した web コミュニティを探索するために，アルゴリズムの最初に入力する web ページを以下，初期ページと呼ぶ．

- 深さ上限値

本アルゴリズムで探索対象とする web ページの集合は，図 3 のように初期ページ p_0 を頂点とする探索木を形成する．しかし多くの場合，この探索はすべての web ページを探索しつくすまで終了しない．そこで，

木構造の探索に深さ上限値を設定し、適切ところで探索を終了する。

- リンク数しきい値

多数のリンクまたはリンク元を持つ web ページは他のコミュニティとも強く連結していると考えられ、これらは探索の次候補の数が多いうえに、初期ページとは異なったコミュニティを探索する可能性が高い。この topic drift¹⁶⁾ を避けるために、リンク数しきい値を導入する。これは、ある一定数以上のリンクまたはリンク元を持つ web ページを対象とした探索を続けないためのパラメータである。具体的には、リンク数しきい値は隣接行列のある 1 行や 1 列における 1 の上限数となる。

- Visit Counter

一度発見された web ページのリンク先を二度と探索しないために、各 web ページに対応する visit counter をそれぞれ設定し、探索で発見するたびに web ページの visit counter に 1 を加算する。また、1 以上の visit counter を持つ web ページが発見されたら、そのページのリンク先およびリンク元を探索しない。したがって、visit counter の数が大きいほど連結度が高くなる。

以下、WCS アルゴリズムを示す。

- (1) web ページの集合を表した $N \times N$ の隣接行列 A を読み込み、初期ページ p 、深さ上限値、およびリンク数しきい値などのパラメータを入力する。
- (2) 入力された初期ページ p の行列要素を p_0 と入れ替え、探索するページ番号の変数 i に 0 を代入する。
- (3) データベース検索の量子アルゴリズム¹⁴⁾ を用いて、隣接行列の i 番目の行要素 $\sum_{j=0}^N A_{ij}$ 、および列要素 $\sum_{j=0}^N A_{ji}$ の中から 1 である j を探索する。 j がなければ、次の i を代入する。
- (4) 見つけたすべてのページの中で visit counter が 1、かつ 1 の数がリンク数しきい値を超えていないページのみを p_j とし、それらの web ページの行列要素を未交換の最も上位の行列要素と交換する。
- (5) 見つけたすべてのページの visit counter に 1 を加える。同じページが複数回探索されたら、重複して加算する。
- (6) p_j の各ページの深さを表す変数を p_i の深さの変数に 1 を加えたものとする。重複は不可とする。それが深さ上限値に達するか、行列の対角要素より後に 1 がなければ (7) へ進む。その他

の場合は、 i に j のページ番号を代入して (3) へ戻る。

- (7) 隣接行列および各 web ページの深さを出力して終了する。

連結度が高い小集団があるとき、それを構成する web ページがリンク数しきい値によって探索の対象から破棄される可能性がある。しかし、対象ページの対角成分より後の行列要素 (図 1 における B) をリンク数しきい値の対象とすることで、この懸念は解消される。

4. 実 験

本章では、WCS アルゴリズムの実験結果を示す。ただし、本実験では前章 (3) で量子アルゴリズムを用いていない。

本実験では、108,123 個の web ページの集合を入力データとして使用する。データは、複数のシード ノードを基点とする前後 2 リンク以内の web ページをすべてリスト化したものである。シード ノードとはある topic の代表的な web ページであり、本実験では国内のコンピュータやメディアに関する企業を topic として選び、以下の 6 個のシード ノードを用意している。

www.ana.co.jp (全日空)

www.compaq.co.jp (コンパック)

www.gnavi.co.jp (ぐるなび)

www.honda.co.jp (本田技研工業)

www.asahi.co.jp (朝日放送)

www.adobe.co.jp (アドビシステムズ社)

ただし、同じドメインを持つ URL のリンクはたどっていない。データのフォーマットは、

```
+SourceURL  $S_1$ 
  DestinationURL  $D_1$ 
  DestinationURL  $D_2$ 
  ...
```

である。

複数の web コミュニティを含むデータを実験対象とすることで、初期ページとは関連のない web コミュニティに属する web ページが雑音 (無作為なデータ) の役割を果たす。また、web コミュニティの属性を決定するうえでシード ノードは重要である。シード ノードは web コミュニティの中心的存在であるため、シード ノードから 4 個先までのリンクをたどるより、前後 2 リンクをたどるほうが妥当である。

図 4 は、WCS アルゴリズム適用前の 108,123 × 108,123 の隣接行列の 1 の要素だけを 1 ピクセルのドットで表現し、全体を 0.03% に縮小したもので、グ

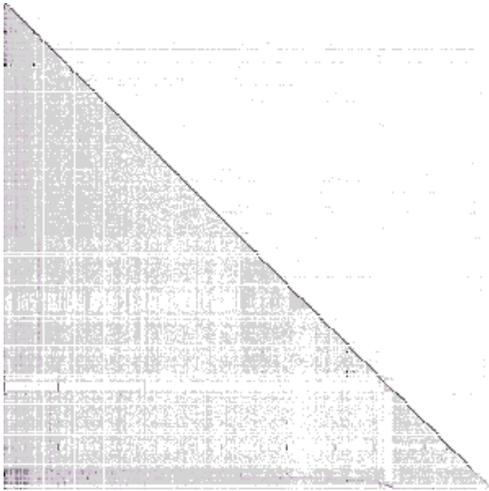


図 4 WCS アルゴリズム適用前の隣接行列

Fig. 4 Adjacency matrix before applying the WCS algorithm.

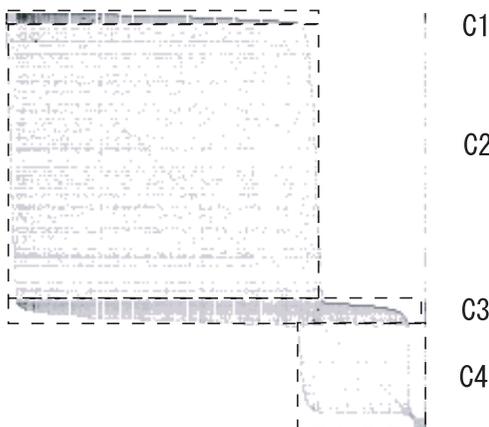


図 5 WCS アルゴリズム適用後の隣接行列

Fig. 5 Adjacency matrix after applying the WCS algorithm.

レスケールで 1 の密度を表現している。データのフォーマットに従って上位の URL から重複分を除いて順に番号化したため、適用前の隣接行列は、対角要素周辺に点が集合した三角行列となっており、その内部でも少数のリンクが存在している。WCS アルゴリズム（量子アルゴリズムを使わない版）の実装は、対象となる隣接行列が $10 万 \times 10 万$ のオーダーであるため、疎行列を利用した。探索に要した計算時間は、1.5 GHz Pentium4、メモリ 1 GB、Linux の PC で約 7 時間であった。

図 5 は WCS アルゴリズム適用後の隣接行列である。本実験でのパラメータはそれぞれ、リンク数しきい値を 1000、深さ上限値を 20、初期ページを

www1e.mesh.ne.jp/osaka-water（大阪市水道局）としている。

まず結果をデータで見ると、最も上位にある集団 C1 では www.tv-osaka.co.jp（大阪テレビ）などのメディア関係、次に、www.osaka-cci.go.jp（大阪商工会議所）などの経済関係、そして、www.osaka-u.ac.jp（大阪大学）などの学術的な団体と、大阪に関係する組織の URL が上から 200 位ほどでは集中している。しかしそれ以下では全国のメディア関係や、transit.yahoo.co.jp（Yahoo!路線情報）などの URL も混ざり、さらに下では、www.gnavi.co.jp/KANSAI（関西版のグルメ情報）や、www.tv-asahi.co.jp/ryoban（朝日テレビの料理番組）などの topic drift が見られる。それから先 C2 まで、大阪に関係する URL は徐々に減少している。また、その下の C3 ではリンクまたはリンク元を多数持つ URL が集合し、C4 のコミュニティでは初期ページとは関連がない URL が無秩序に並んでいる。したがって、本実験では C1 でのキーワードを「大阪」とし、その内部では細分化がなされており、上位にあるほど初期ページとの関連があるといえる。ただし、初期ページと内容的に直接関連がある URL ではない。それは、同じドメインを持つ URL のリンクはたどらないデータ取得方法によるためである。

次に結果を図形的に見ると、大きく 4 個の集合体に分類できる。まず C1 は、適度なリンク数を持ち初期ページに関連したページが比較的多い中心的な集団といえる。正方形に近い集団 C2 は密度が薄いことから、C1 と関連するがリンク数が少なく連結度も低い。次が、リンク数が他の集団より多く、リンク数しきい値によって除外されたページの集団 C3 である。最も下位にあるのが、初期ページとは関連がなく、リンク数も少ないページの集団 C4 である。

リンク数しきい値のみを 100 に変更して、実験した結果が図 6 である。C1 が小さくなり、初期ページと関連のないページが上位で減少している。しかし、この値を小さくしすぎると、初期ページと関連のない web ページが減少すると同様に関連のある web ページも減少するため、必ずしも良い結果が得られるというわけではない。

以上のことから、WCS アルゴリズムは初期ページに関連したコミュニティを最も上位に出力することが確認された。また、各 web ページの深さも出力されるので、初期ページからの「距離」が分かる。

5. 量子アルゴリズム化とコーディング例

本章では、WCS アルゴリズムの中で適用する既存

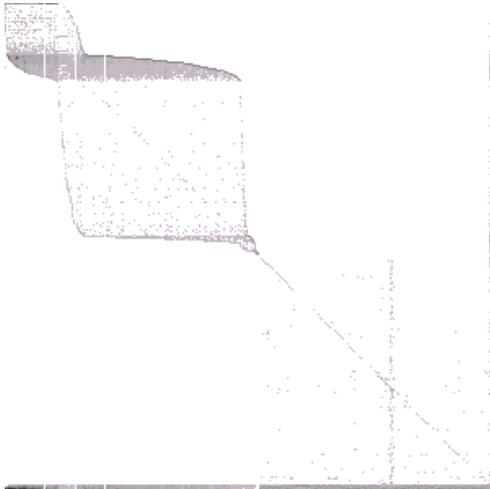


図 6 リンク数しきい値 100 でのアルゴリズム適用後

Fig. 6 Matrix after applying the WCS algorithm with threshold 100.

の量子アルゴリズムの手順，WCS アルゴリズムとの関連，および前述の量子命令セットでのコーディング例について述べる．

5.1 データベース探索アルゴリズムの手順とコーディング

Turing 機械の枠組みでは，隣接行列において，あるページのすべてのリンク先を探す効率的なアルゴリズムはまだ提案されていない．したがって，量子コンピューティングの無限並列性を利用したデータベース探索を行う量子アルゴリズムが WCS アルゴリズムに適用可能である．

3 章の WCS アルゴリズムのステップ (3) で N 行 N 列の隣接行列における任意の行や列からすべての 1 を探索するとき，すでに提案されている量子アルゴリズムを用いる．以下，WCS アルゴリズムのステップ (3) における量子アルゴリズム化が適用できる部分のアルゴリズムを示す．ただし，データ数 N に比べて解となる 1 の数は十分に少ないものとする．

データベース検索の量子アルゴリズム¹⁴⁾ は，解の数が不明の N 個の要素の中から条件 $T[i] = x$ を満たす状態の振幅だけを増幅し，観測を行って任意の解を 1 個取り出す．それを繰り返し，複数解を得る．この量子アルゴリズムは， N 個の行列要素から 1 を検索するステップにおいて用いる．

- (1) $m = 1, \lambda = 6/5$ に初期設定する．
- (2) $k(0 \leq k \leq m - 1)$ をランダムに選ぶ．

- (3) 文献 11) の量子アルゴリズムを k 回繰り返す．
- (4) レジスタを観測して i を得る．
- (5) $T[i] = x$ ならば終了する． $T[i] \neq x$ ならば， m を $\min(\lambda m, \sqrt{N})$ にセットして (2) へ戻る．

次に，我々がすでに提案している量子命令セット¹³⁾ を用いてこのアルゴリズムのコーディング例を示す．ただし紙面の都合上，コーディングすべては示していない．

Store 1, $N-R_m$

Store 6/5, $N-R_\lambda$

L7 { k をランダムに選ぶ }

Store k , $N-R_k$

{ 文献 11) を k 回繰り返す }

{ もし観測した値 i が $T(i) = x$ を満たせば i を表示して終了する．

そうでなければ m を $\min(\lambda m, \sqrt{N})$ にセットして L7 へ戻る }

Stop

次に，文献 11) の量子アルゴリズムについて述べる．これは， N 個の要素から与えられた条件を満たすただ 1 つの解を， $O(\sqrt{N})$ のステップで $1/2$ の確率で探し出す．ノイマン型アルゴリズムでは $O(N/2)$ のステップが必要である．またデータベースには，データが保存されている項目 (item) と，それに対応する表 (table) とが存在する．項目を変数 i ($i = 0, \dots, N$) で表すとき，表は $T[i]$ ($i = 0, \dots, N$) と表される．隣接行列では，ある 1 行または 1 列の要素が表で，行列の番号が項目に相当する．この量子アルゴリズムでは， N 個の要素を S_1 から S_N までラベルづけした量子レジスタを重ねあわせ状態にして，与えられた条件を満たす状態 S_v の振幅だけを増幅する．

- (i) 量子レジスタ R1 に 0 から $N - 1$ までの重ねあわせ状態を入力する．
- (ii) 以下，(a) (b) を $O(\sqrt{N})$ 回繰り返す．

- (a) 任意の状態 S に対して，条件付き位相回転を行う．

$C(S) = 1$ のとき，位相を π 回転する．

$C(S) = 0$ のときは，何もしない．

- (b) 拡散変換 D を適用する．ただし， D は次のように定義される．

$$D_{ij} = 2/N, \text{ if } i \neq j$$

$$D_{ii} = -1 + 2/N$$

- (iii) R1 を観測する．

上記のステップ (ii) (a) において，次の量子オラクルを用いている．

¹⁴⁾ m は， k の上限を設定する変数でステップ 5 で更新される． λ は， $1 < \lambda \leq 4/3$ を満たす任意の数でよい．

$$|S\rangle \rightarrow (-1)^{C(S)}|S\rangle$$

このデータベースは、量子状態で問合せ可能な量子オラクルである必要があり、量子メモリ上に実現されているか量子回路のような形で外部的に与えられているものとする。

以下、この量子アルゴリズム¹¹⁾のコーディング例を記述する。

```

Read N
Load N,N-RN
{ 2n ≥ N を満たす最小の n を計算する }
Store n, N-Rn
L1 QSetLength Q-R1,N-Rn
QExchange Q-R1,I-Reg
Read M
CPhase M, N-RM
QRP Q-R1,N-RM
Read D
CPhase D, N-RD
{ 以下 L2 まで O(√N) 回繰り返す }
QRPS “C(Q-R1)=1”, Q-R1, π
L2 QRP Q-R1,N-RD
QObserve Q-R1, N-Rk
; Q-R1 を観測
{ もし C(N-Rk) = 1 でなければ L1 へ戻る }
Stop

```

5.2 量子命令セット

前節で行ったコーディングで使用した量子命令の仕様についての概略を述べる。詳細は文献 13) を参照されたい。

(1) QSetLength

QSetLength は、この命令の実行後に実行される量子命令のオペランドで指定された量子レジスタ内で使用する qubit の個数 (以後、量子レジスタの長さ) を指定する命令である。つまり、量子レジスタの先頭にある qubit の個数を設定する命令である。

(2) QExchange

QExchange は、量子レジスタを初期化するための命令である。オペランドで指定された量子レジスタに対して、あらかじめその量子レジスタに QSetLength で設定された qubit の個数分だけ、量子初期化レジスタファイルにメモリ領域を確保する。次に、量子レジスタと量子初期化レジスタとのレジスタ番号を交換する。ただし、同じ量子レジスタに対して 2 回以上初期

化命令が実行されるときは、2 回目以降は量子初期化レジスタ内でメモリ領域の確保とレジスタ番号の交換を行う。

(3) QRP

QRP (Quantum Rotate Phase) は、量子レジスタで使用される全 qubit を位相回転するための命令である。qubit のパラメータを、 $Q(\delta_0, \theta_0, \gamma_0)$ とし、それに適用するユニタリー行列のパラメータを、 $U(\delta, \theta, \gamma)$ とする。このとき、ユニタリー行列の適用後の qubit の状態は、 $Q'(\delta_0 + \delta, \theta_0 + \theta, \gamma_0 + \gamma)$ のように表現される。

(4) QRPS

QRPS (Quantum Rotate Phase Selectively) は、条件付きの QRP 命令である。つまり、条件 $T(Q-R_1)=1$ を満たす状態のみを更新する $Q-R_1$ の中の qubit を、選択的に位相回転する命令である。また、 $Q-R_1$ には項目が保存しており、 $T(Q-R_1)$ は $Q-R_1$ の項目に対応する表の値を出力する関数である。

(5) QObservation

量子レジスタの観測結果をノイマン型レジスタに保存する。

5.3 計算量の比較

WCS アルゴリズムで必要な計算量のオーダーは、ステップ (3) における計算量に依存する。それは web ページ数が N のとき、ノイマン型コンピューティングでは $O(N) \times$ (探索した行列の数) であり、量子コンピューティングでは $O(\sqrt{N}) \times$ (解の数) \times (探索した行列の数) である。(探索した行列の数) とは $N/(\text{解の数})$ であり、(解の数) とはある 1 行または 1 列における 1 の数である。

したがって最も計算量が大きくなるのは、各ページがすべて異なるページに 1 個ずつリンクして、1 本の鎖状になっている場合である。このとき (探索した行列の数) は N で、(解の数) は 1 である。すなわち、量子コンピューティングの計算量は $O(N\sqrt{N})$ で、ノイマン型コンピューティングでは $O(N^2)$ となる。また上記の計算量より、量子コンピューティングが WCS アルゴリズムに対して有効なのは、(解の数) が \sqrt{N} に満たないときである。

6. 結論および今後の課題

本論文では、まず隣接行列および web モデルを用いて、数値的および図形的に web コミュニティの特徴を述べた。次に、既存の量子アルゴリズムを用いて web コミュニティの効率的な探索アルゴリズムを提案した。web ページの集合を隣接行列で表すと、それは疎行列

コメント行は; で始まる。

となることが多いので、データベース探索アルゴリズムが効率良く適用される。また、実際の web データを用いて実験を行い、その結果が前述の web コミュニティの特徴を示していることを確認し、WCS アルゴリズムの有効性を検証した。

本アルゴリズムにキーワード探索エンジンでの話題の絞り込み、テキスト、タグおよび URL の解析などの手法を用いれば、より効率良く web コミュニティ探索の精度も上がると予想される。しかし、量子コンピューティングを用いた web コミュニティ探索アルゴリズムはほかに例がないため、将来の量子コンピュータに対するエラー・アプリケーションの提示を本論文での目的として、まず隣接行列のみを用いたグラフ理論からのアプローチを行った。そのため、ある程度の topic drift は避けられないことが判明したので、個人のリンク集などを避ける工夫が別途必要である。

今回用いた web データはリンクの関係によって取得したものである。これに対して、web ページの内容によって取得したデータを用いれば別の興味深い研究結果が期待できるかもしれない。

本実験は、量子アルゴリズム部分のシミュレーションを含んでいない。したがって、量子コンピュータ・エミュレータを開発し、量子アルゴリズムを含めたシミュレーションおよびその評価が急務の課題である。

謝辞 東京大学生産技術研究所の喜連川優教授には web データを提供していただき、また本学理学部情報科学科の植村垂以さんにはプログラムの実装を行っていただきました。感謝いたします。

参 考 文 献

- 1) Flake, G., Lawrence, S. and Giles, C.: Efficient Identification of Web Communities, *6th International Conference on Knowledge Discovery and Data Mining*, pp.150–160 (2000).
- 2) Brin, S. and Page, L.: The Anatomy of a Large-Scale Hypertextual Web Search Engine, *7th International World Wide Web Conference*, pp.107–117 (1998).
- 3) Davison, D., Gerasoulis, A., Kleisouris, K., Lu, Y., Seo, H., Wang, W. and Wu, B.: DiscoWeb: Applying Link Analysis to Web Search, *8th International World Wide Web Conference*, pp.148–149 (1999).
- 4) Kleinberg, J.: Authoritative sources in a hyperlinked environment, *9th ACM-SIAM Symposium on Discrete Algorithms*, pp.668–677 (1998).
- 5) Barabási, A., Albert, R. and Jeong, H.: Scale-free characteristics of random networks: The

topology of the World Wide Web, *Physica A*, Vol.281, pp.69–77 (2000).

- 6) Gibson, D., Kleinberg, J. and Raghavan, P.: Inferring web communities from link topology, *9th ACM Conference on Hypertext and Hypermedia*, pp.225–234 (1998).
- 7) Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. and Wiener, J.: Graph structure in the web, *9th International World Wide Web Conference*, pp.247–256 (2000).
- 8) Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A. and Upfal, E.: The Web as a Graph, *19th ACM Symposium on Principles of Database Systems*, pp.1–10 (2000).
- 9) Kumar, R., Raghavan, P., Rajagopalan, S. and Tomkins, A.: Trawling the web for emerging cyber-communities, *8th International World Wide Web Conference*, pp.1481–1493 (1999).
- 10) Shor, P.: Algorithms for Quantum Computation: Discrete Logarithms and Factoring, *Proc. 35th Annual Symposium on Foundations of Computer Science*, pp.124–134 (1994).
- 11) Grover, L.: A fast quantum mechanical algorithm for database search, *Proc. 28th Annual ACM Symposium on the Theory of Computing*, pp.212–219 (1996).
- 12) Oto, M., Nakajo, H. and Joe, K.: A Possible Instruction Set for Quantum Computer Architectures, *2001 International Conference on Parallel and Distributed Processing Techniques and Applications*, Vol.3, pp.1221–1227 (2001).
- 13) 大音真由美, 中條拓伯, 高田司郎, 城和貴: 量子コンピュータ実現に向けた量子命令セットについて, 情報処理学会論文誌: 数理モデル化と応用, Vol.43, pp.19–28 (2002).
- 14) Boyer, M., Brassard, G., Høyer, P. and Tapp, A.: Tight bounds on quantum searching, *Proc. 4th Workshop on Physics and Computation*, pp.36–43 (1996).
- 15) 村田剛志: ハイパーリンクのグラフ構造に基づく Web コミュニティの洗練, 人工知能学会論文誌, Vol.17, No.3, pp.322–329 (2002).
- 16) Chakrabarti, S., Dom, B., Kumar, R., Raghavan, P., Rajagopalan, S., Tomkins, A., Gibson, D. and Kleinberg, J.: Mining the Web's Link Structure, *IEEE Computer*, Vol.32, No.8, pp.60–67 (1999).

(平成 14 年 11 月 6 日受付)

(平成 15 年 1 月 10 日採録)



大音真由美 (学生会員)

1998年奈良女子大学理学部物理学
 学科卒業。2000年同大学大学院人
 間文化研究科博士前期課程修了。現
 在同大学院博士後期課程在学中。量
 子コンピュータ・アーキテクチャの

研究に興味を持つ。



中條 拓伯 (正会員)

1961年生まれ。1985年神戸大学
 工学部電気工学科卒業。1987年神
 戸大学大学院工学研究科電子工学専
 攻修了。1989年神戸大学工学部シ
 ステム工学科 (後に情報知能工学科)

助手を経て、現在、東京農工大学工学部情報コミュニケー
 ション工学科助教授。工学博士。1998年より1年間 Illi-
 nois 大学 Urbana-Champaign 校 Center for Super-
 computing Research and Development (CSRD) に
 て Visiting Research Assistant Professor 。プロセッ
 サアーキテクチャ、分散共有メモリ、クラスタコン
 ピューティングに関する研究に従事。電子情報通信学
 会、IEEE CS 各会員。



城 和貴 (正会員)

大阪大学理学部数学科卒業。日
 本 DEC, ATR 視聴覚研究所 (日本
 DEC より出向) (株)クボタ・コン
 ピュータ事業推進室で勤務。1993年
 奈良先端科学技術大学院大学情報科

学研究科博士前期課程入学、1996年同研究科後期課
 程修了、同年同研究科助手。1997年和歌山大学シ
 ステム工学部情報通信システム学科講師、1998年同学
 科助教授。1999年より奈良女子大学理学部情報科学
 科教授。工学博士。画像処理、文字認識、ニューラル
 ネットワーク、並列計算機アーキテクチャ、自動並列
 化コンパイラ、並列計算機の解析モデル、視覚化等の
 研究に従事。IEEE、ACM 各会員。