ウェブグラフ—その性質と利用

宇野　裕之

1. はじめに

ウェブはその誕生以来，人々の予想を超える急速な発展を遂げ，われわれの日常生活や社会にさまざまな恩恵をもたらしている．それと同時に，数多くの新しい研究分野を創出した．

ウェブに関する研究分野も拡大を続けて多岐に渡る．その中心は実用的，応用的な分野であるが，そのような分野の消長は激しい．その一方で，ウェブの応用的な利用を支える基礎的，理論的な研究が存在する．その一つが，ウェブのリンク構造をグラフとしてとらえるウェブグラフに関連する分野である．ウェブグラフは，ウェブ上で動作する検索エンジンやクローラなどのウェブアルゴリズム設計のための，最も基本的なモデルとなっている．この分野を広く世に説いたのはKleinbergであり[1]，いまや‘リンク解析’や‘ウェブグラフ’などの用語とともに，重要な位置を築いている．

そこで本稿では，この分野のいくつかの基本的あるいは理論的な成果と，とくにウェブ構造マイニングとよばれるトピックを取り上げる．全体を通してそれらを紹介するとともに，それらを通じて，ウェブの基礎モデルであるウェブグラフに関連する研究において，情報科学や数学の基礎理論，あるいはオペレーションズ・リサーチの手法が，どのように用いられているかという視点を交えて説明する．

2. 基礎的な研究

2.1 ウェブグラフ

ウェブ上に存在するウェブページを点，ウェブページ間に張られたハイパーリンクを有向辺とする有向グラフを**ウェブグラフ（webgraph）**[2]という（図1）．ウェブグラフは，ウェブ上で動作する多くのアプリケーションやウェブアルゴリズムなど設計のための基礎的なモデルであり，極めて重要な概念である（本稿中，ウェブグラフの点をページあるいはノード，辺をリンクとも呼び，これらを区別せず用いる）．

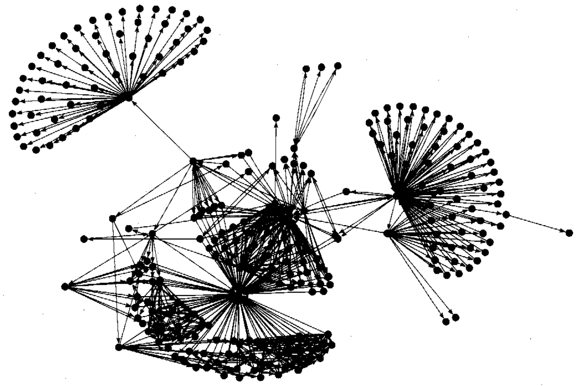

図1　ウェブグラフの一部（2003年8月のデータ）

ウェブグラフに関するもっとも素朴な疑問の一つはその大きさ，すなわちウェブにはいったいいくつのページ（やリンク）が存在するかであろう．その数を知るには，存在するページを実際に確認するしか方法がないが，世界人口よりもはるかに速いスピードで生成死滅を繰り返すページとリンクの「現在」の情報を獲得することは現実的には不可能であり，残念ながらこの問いに対する答えは誰にも分からない．最多のページ数を確認しているのは，ここ数年常にGoogleで，2006年現在，少なくとも約200億ページが存在するようだが，これも下界の一つでしかなく，実際にはこれより多いと考えられている．

このような規模のウェブグラフで動作するアルゴリズムを設計する際に，常に問題となるのはそのスケーラビリティである（当然だが，ウェブグラフのデータは通常の個人用途のPCのメモリに一度には格納でき

うの　ゆうし
大阪府立大学 理学系研究科
〒599-8531 堺市中区学園町1-1

[1] Jon M. Kleinbergは，このことをはじめとする情報科学の基礎理論における多大な功績により，2006年の国際数学者会議（ICM）において，Nevanlinna賞を授与された．
[2] ウェブグラフの英語表記には，まだ統一的なものがなく，webgraph, web graph, web-graph，それぞれの先頭文字を大文字にしたもの，WebGraph, Web-Graphなどの記法が乱立している．

ない）．たとえアルゴリズムが単純であっても，それがウェブ規模で動作する実装が伴わなければ，実用上は無意味である．また，観察や実験のために手にしたウェブデータは，その時点ですでに最新ではなく，しかもウェブの一部でしかないことにも注意する必要がある．

2.2 ウェブグラフの性質

ウェブグラフの性質の中で最も重要なものの一つは，点の入次数（無向グラフの場合は単に次数）の分布がべき乗則に従うという**スケールフリー性**[3]である[2][20]．ただし**べき乗則分布**（power-law distribution）とは，ある確率変数 X（この場合，点の入次数）の値が k である確率が，$P[X=k]\propto 1/k^{\gamma}\ (\gamma\geq 1)$ で表される確率分布であり，γ をべき指数という．このとき，この分布に従う入次数分布を，横軸を入次数，縦軸をその頻度とする両対数軸上に描くと，傾きが $-\gamma$ の直線になる（図2）．ウェブグラフがスケールフリー性をもつことは，極めて多くのリンクを集めるページが少なからず存在することを意味している．航空路線や論文共著関係など，現実社会で自然発生的に形成されるネットワークの多くがスケールフリー性を持つことが観察されており，これらを総称して**スケールフリー・ネットワーク**と呼ぶ．ウェブグラフは，その代表的なものの１つである．

ウェブグラフが持つ別の性質に**スモールワールド性**がある[1][20][30]．これは，あるページからリンクをたどって他の任意のページへ到達するために経由するページ数が，グラフの規模と比較して非常に小さいことを主張するもので，ウェブグラフの直径（あるいは平均点間距離）が小さいことを意味する（n 点のグラフで $O(\log n)$ 程度）．この事実は，古典的には心理学の分野の研究成果として「6次の隔たり（six degrees of separation）」というフレーズで知られている[30]．

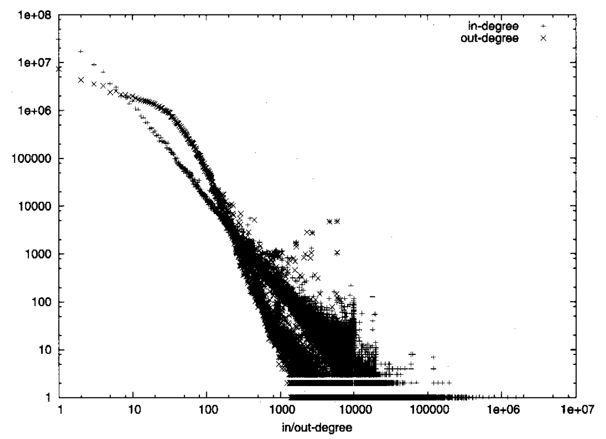

図2　ウェブグラフの入次数（および出次数）の分布がべき乗則に従う様子（2003年8月のデータ，約1億点）

[3] スケールフリー性とは，本来は**自己相似性**を意味する．

その他にも，ウェブグラフの蝶ネクタイ構造や，（図1にも見られるように）密な部分グラフを持つクラスター性など，興味深い性質が数多く知られている[6][20]．上記のような性質のいくつかを共通に持つネットワークは，複雑ネットワークやソーシャル・ネットワークなどの名称で総称されることもある．これらは社会学，生物学，心理学，物理学，数学，情報科学などさまざまな分野で観察，研究され，近年それらの成果を横断的にまとめる学術雑誌や成書も刊行されている[7][24]．

2.3 ウェブグラフのモデル

ランダムな性質を持つグラフの代表的なモデルとして，Erdős と Rényi によるランダムグラフ（ER モデル）があるが，前節で述べた性質の多くは，そのような古典的なモデルでは説明できない[4]．このような性質を持つネットワークは，いったいどのようなメカニズムで自然生成されるのかという疑問とともに，これらを説明するグラフモデルが，主として物理学や数学の分野で多数提案された．

初めにスケールフリー性の説明に成功したのが，Barabási と Albert[2]によるモデル（**BA モデル**）である．彼らは，この種のネットワークが持つ共通の性質として，モデルに次の２つの要素を組み込んだ：(i) ネットワークは常に新しい点が追加されることで成長し続ける，(ii)新しい点が既存の点に対して接続する枝は，次数の高い点に対してより高い確率で接続される（優先的選択（preferential attachment））．

このアイデアにもとづいて実際にネットワークを生成する方法は，具体的には次のようになる．まず，1. 時刻 $t=0$ では，n_0 点（$v_1,\cdots,v_{n_0}$, n_0 は比較的小さな値）からなる完全グラフである；2. 各時刻 $t=i$ で，新しい1点 v_{n_0+i} が加えられ，その点から既存のネットワークの点に対して $m\,(\leq n_0)$ 本のリンクが張られる．このとき時刻 $t=T$ で，新しい点 v_{n_0+T} から既存の各点 v_i に対してリンクが張られる確率 p_i は，$p_i=d_i/\sum_{j<T}d_j$ で定義される．ただし，d_i は点 v_i の次数である．したがってこのネットワークは，時刻 $t=T$

では$\binom{n_0}{2}+mT$本のリンクを持つことになる．この確率p_iは，各点の次数d_iに比例した値であり，これにより，次数が高い点はより多くのリンクを集めやすいという優先的選択の仕組みを実現している．

このメカニズムで生成されたネットワークがスケールフリー性を持つことは，時間と点の次数を連続化した近似的なBAモデルで，次数を時間の関数として記述する微分方程式によって次のように説明される．時刻tに新しい点vがm本の枝とともに加わるとき，点v_iの次数d_iに関して，$\frac{\partial d_i}{\partial t}=m\cdot p_i\cong\frac{d_i}{2t}$という関係が近似的に成り立つ．これを，点$v_i$がネットワークに加えられた時刻$t_i$での初期条件$d_i(t_i)=m$とあわせて解くと，$d_i(t)=m\sqrt{t/t_i}$が得られる．ここで，新しい点は時刻0から等間隔で加えられるとすると，時刻tで点v_iの次数d_iがkより小さい確率$P[d_i(t)<k]$は，

$$P[d_i(t)<k]=P\left[t_i>\frac{m^2t}{k^2}\right]$$
$$=1-P\left[t_i\leq\frac{m^2t}{k^2}\right]=1-\frac{m^2t}{k^2(n_0+t)}$$

となる．したがって，確率密度$P(k)$は

$$P(k)=\frac{\partial P[d_i<k]}{\partial k}=\frac{2m^2}{k^3}\propto k^{-3}$$

となり，べき指数$\gamma=3$が導かれる．また，異なるp_iを定義することで，3とは異なるべき指数に従うスケールフリー・ネットワークが得られることも知られている[2]．この他に近年の進展としては，例えば[9]がある．

一方，WattsとStrogatz[30]は，スモールワールド性を説明するモデルを提案した（**WSモデル**）．これは，比較的大きな直径を持つ規則的につくられたネットワークにおいて，枝のいくつかをランダムに付け替えることで「近道」を生成し，ある範囲の付け替え確率に対して，小さい直径を実現している．この方向の最近の成果には，例えば[5]などがある．

複雑ネットワーク全般で観察される性質を持つグラフモデルの研究は[26]に詳しいが，近年は減少傾向にあるようである．しかしながらウェブグラフでは，ウェブが持つドメインとウェブ全体の2層構造を適切に表現するモデルが欠けていると考えられている[3]．ウェブの全データの取得が困難な状況では，設計されたウェブアルゴリズムの動作確認や検証のためにも，ウェブの実態を反映したモデルの整備が必要である．

2.4 ウェブグラフの探索

ウェブ検索エンジンは，世界中のウェブページを新しく発見したり，一度発見したページの更新を定期的にチェックしなければならない．そのためには，ウェブグラフを効率的に探索することが必要となる[8]．このとき，大域的な構造が分かっているグラフに対する探索とは異なり，局所的な構造しか分からない未知のグラフに対して，そのトポロジーを知るための探索を行うことになる．また，構造が判明しているグラフであっても，実装時のメモリサイズの制約から局所的な情報だけで探索を行うことが想定される．

これらを目的とする探索手法としては，しばしばグラフ上のランダムウォークが用いられ，ウェブグラフを探索する立場で興味があるのはその被覆時間である．任意のグラフにおける通常のランダムウォークの被覆時間については，その上下界がそれぞれ$O(n^3)$, $\Omega(n\log n)$であることが知られている[13][14]．またIkedaら[16]は，隣接点の次数まで考慮した推移確率を構成することで，被覆時間の上界を$O(n^2\log n)$に改善できることを示している．さらに，われわれが対象とするウェブグラフは，前節で見たような性質を持っている．そこでCooperら[10]は，そのような性質を仮定したグラフ上のランダムウォークを考察している．

また，ランダムウォークとは異なるアプローチとしてKurumidaら[23]は，局所的な情報だけで決定的な探索を行うforest searchという枠組みを提案している．

3. 応用的な利用法

3.1 ページのランキング

Googleをはじめとする多くの代表的な検索エンジンは，検索結果を表示する際にその結果を重要度（ランク）の高いものから順に表示する仕組みをもっており，このランクづけの良し悪しが検索システムの性能を決定づける重要な要素になっている．Googleの設立者であるPageとBrinは，このときページの内容を不当に編集することで恣意的な操作が可能となるようなランクではなく，できる限り客観的な基準が必要であると考え[4]，次のようなアイデアを考案した[27]：「ランクが高いページとは，より多くのしかも

[4] ランクの客観性あるいは公平性が重要さを増しているのは，それが商業目的で利用される機会が増大していることにもよる．

ランクの高いページからの厳選されたリンクを受けているページである」(Googleでは，その重要度をページランク (PageRank™) と呼んでいる)．

彼らはこの定義を，ウェブグラフの n 点を状態とし，グラフの隣接行列をもとに定義される行列 P を推移確率行列とするマルコフ連鎖としてモデル化し，ページのランクをその定常分布と考えた．具体的には，A をウェブグラフの隣接行列（すなわち，$a_{ij}=1 \Longleftrightarrow$ 点 i から j へリンクがある）とするとき，P を

$$p_{ij}=\alpha d_i^{-1}a_{ij}+(1-\alpha)n^{-1} \tag{1}$$

を成分とする $n\times n$ 行列で定義する．ただし，d_i は点 i の出次数である．また，$\alpha\in(0,1]$ は減衰率 (damping factor) と呼ばれる係数であり，式(1)の右辺は，閲覧者は確率 α でページ内のリンクを等確率でたどり，確率 $1-\alpha$ でリンクとは無関係に任意のページに等確率でジャンプすることを表している[5]．右辺第2項は，リンク先のないページ (dangling page) を解消し，マルコフ連鎖を既約にする役割も持つ．

このとき，ページランク $\boldsymbol{r}_\alpha$ は P を推移確率行列とするマルコフ連鎖の定常分布であると考え，したがって P^T の固有値を計算することで求められる．例えば，図3のウェブグラフに対する各ページのランクは，純粋にリンク構造だけから計算され，同図に示すとおりとなる．ただし，減衰率 $\alpha=0.85$ としている．多くのリンクを集めるページAと，Aからの唯一のリンクを受けるページBのランクが高いことが分かる．

このように，ページランク計算のアイデアは一見単純だが，ウェブ規模の巨大な行列に対して実際に固有値を計算することは容易ではない．このアイデアで実際に実用的なランクが得られることだけでなく，現実のウェブに対する実装まで行ったことに，PageとBrinの研究の価値がある．大規模かつ疎な行列でのページランク計算の高速化には，さまざまな工夫が必要となる[11]．

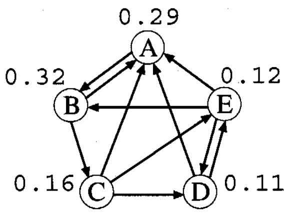

図3 小さなウェブグラフと計算された各ページのランク

[5] ウェブページの閲覧者（サーファー）の典型的な閲覧行動をモデル化していると考え，**ランダムサーファーモデル**と呼ばれる．

また，減衰率 α を変化させることで，異なるランクが得られる．これに関して，さまざまな基準で最良のランクを導く α を得るための研究は多いが，現在でもPageとBrinが示した $\alpha=0.85$ がマジックナンバーとして用いられている．さらに，式(1)右辺の第2項に重みを乗じて p_{ij} を調整し，個人の嗜好に応じたランク（パーソナライズド・ランキング）を得る研究もあり[15]，その成果はGoogleの β 版の機能として実現されている．

3.2 ウェブ構造マイニング

ウェブを巨大なデータベースとみなして，主として単純な照合で情報を得るのが検索であるのに対して，それだけでは発見できない埋もれた二次的な情報を発見することをウェブマイニングとよぶ．その中でも，ウェブのリンク構造に注目して（すなわちウェブグラフ上で）行うマイニングを，とくにウェブ構造マイニングと呼ぶ．構造マイニングは，ウェブ上で特定の話題に興味を持つ隠れたコミュニティの発見などを目指し，その成果は社会現象の解明や，構造にもとづく検索エンジンの効率化などへの利用が期待できる．

それでは，ウェブ上でコミュニティを構成するページは，ウェブグラフ上ではどのような構造を形成するだろうか．Kleinberg[18]はこの問いに対して，**ハブ-オーソリティ・モデル** (hub-authority model) を提案した．例えば，アメリカ大リーグ (MLB) に興味を持つコミュニティがあるとする．その中には，MLBや各球団の公式サイト，あるいは熱狂的なファンやファンに一元的に情報を提供するページが存在するであろう．このとき，各公式サイトはファンのページから多くのリンクを集め，逆にファンのページはそのような公式サイトに数多くのリンクを張っていることが想像される．すなわち，1つのコミュニティは，その話題に関して権威的なページ集合（オーソリティ）とポータル的なページ集合（ハブ）を部集合とする（単に部分グラフとしての）有向2部グラフを含むというものである．さらに，コミュニティの核となるページ集合は，有向2部クリークを構成しているという仮説も考え得る[19][22]．

一般に，コミュニティを構成するページは互いに緊密に，あるいは規則的にリンクを張ると想像され，ウェブグラフの中で絶対的に，あるいは他の部分と比べて相対的に密な部分グラフや，固有の特徴的な部分グラフを構成していると考えられる（その様子が図1でも観察できる)．そのような部分グラフの候補として

は，2部グラフや2部クリークの他に，例えばクリークも考えられるであろう[6].

したがって，これらの仮定のもとでウェブ構造マイニングは，ウェブグラフからそのような部分グラフを抽出することで達成されることになる．ここでは，既存の研究のアプローチを，(1)種ページ集合の伸展（seed set expansion）と，(2)網羅的な列挙（exhaustive enumeration）に分類して説明する．

4. 種ページ集合の伸展

このアプローチは，ユーザが興味を持っている話題を内容に持つページ（種ページ）があらかじめ指定されていて，それに関連する内容のページを探し出してコミュニティとするものである．このアイデアの原型は，やはりKleinberg[18]が2部グラフの発見を試みたHITS（hyperlink-induced topic search）と呼ばれるアルゴリズムだが，ここではFlakeら[12]が異なる枠組みで実現したものを説明する．

その基本的なアイデアは，ある話題に関する種ページをもとに，それを含む点集合で最小カットを定義するものを発見するというものである（図4左）．これにより，グラフの他の部分とは少ないリンクで結ばれている相対的に密な部分を発見することができる．彼らはまず理想的な状況として，種ページsを含むコミュニティC（すなわちウェブグラフの点集合Vの部分集合）を，$C=\{v\in V \mid \deg_C(v)\geq\deg_{V-C}(v), \forall v\in C, s\in C\}$と定義した．この点集合$C$はある条件のもとで，$s$をソース，$C$に属さない点（例えば代表的なポータルサイトなど）をシンクtとするs-t最小カットで求められる点集合と一致することがいえる．

いま，s-t最小カットはs-t最大フロー問題を解く

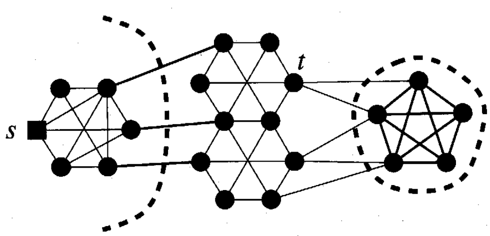

図4　ウェブのコミュニティとそれを抽出する2つのアプローチ．種ページ集合の伸展（左）と網羅的な列挙（右）

[6] ここでのクリークや後述の孤立クリークなど，無向グラフで定義されるウェブグラフで考える際には，辺を無向と見なすなど，ウェブグラフを無向グラフとして扱う必要がある．

ことで求めることができ，最大フロー問題にはそれを解く多項式時間のアルゴリズムが存在する．しかし，たとえ高速な最大フローアルゴリズムがあったとしても，ウェブ規模の入力に対しては，現実的な時間で解を求めることが困難となる．また，わずか一つの話題に関するコミュニティを特定するためだけに，ウェブグラフ全体の情報が必要となる．

そこで彼らは，近似的な最小カットをヒューリスティックに求める手続きを提案している．そのアイデアの概略は次のようなものである：種ページ集合を含むコミュニティを構成するページ集合は，種ページから数リンクの範囲にあると考え，同じ話題の種ページをいくつか準備し（種ページ集合），そこからリンクを実際にたどる（ウェブをクロールする）ことによって，可能性のあるページだけを収集し，さらにグラフに仮想的なシンク，ソースや適切な辺重みを与えて，小規模なネットワークを構成する．そしてその範囲で最大フロー問題を解いた上で，その結果を用いて種ページ集合を更新する．更新された種ページ集合に対して，同じ手続きを数回繰り返す，というものである．

このアイデアの長所は，ウェブグラフ全体の情報を必要とせず，リンク情報をもとに，コミュニティを確定させるために必要なウェブグラフの部分的な情報だけを取得すればよい点であり，それにより実用的な時間でコミュニティを発見することができる．その一方で，あらかじめコミュニティの話題とそれを構成するページを指定する必要があり，水面下で流行の兆しがあるような未知のコミュニティを発見するという目的には適していない．

5. 網羅的な列挙

このアプローチでは，グラフに存在する密なあるいは特徴的な部分構造を直接見つけ出す（図4右）．このとき，発見された部分構造は必ずしも意味のあるコミュニティであるとは限らないが，少なくともコミュニティの候補を網羅的に列挙して提示するという立場である．

組合せ的な対象の効率的な列挙に関する理論的な研究には多くの成果があるが[21][28]，実際の列挙の際に問題となるのは，列挙対象（出力）の潜在的な個数である．既存の効率的な列挙法の多くは，出力サイズに関しては多項式時間であるが，一般に，グラフの中の極大クリークや極大2部クリークは指数個存在する可能性があり（例えば[25]），そのとき列挙に要する

時間は，入力サイズに関しては，指数時間を要することが避けられない．そしてそのようなアルゴリズムは，ウェブ規模のグラフに対しては実用的でなくなる．

Kumar ら[22]の Trawling という方法は，ウェブグラフの部分データ（約2億点規模）に対して，コミュニティの核としての有向2部クリークを網羅的に列挙している．彼らは上記の潜在的な問題に対処するため，列挙の対象とする2部クリーク $K_{i,j}$ を，$i=3\sim6$, $j=3\sim9$ の範囲に限定し，グラフに前処理を行った上で列挙するというヒューリスティックな方法で，実用時間での列挙を達成し，約10万のコミュニティを発見している．

列挙の対象となる部分構造が指数個存在する可能性に対処する方法として Ito ら[17]は，クリークの定義に制限を加えた k-孤立クリークという概念を導入し，入力サイズに関して多項式時間の効率的な列挙アルゴリズムを提案している．彼らの導入した k-孤立クリークとは，グラフ G のクリーク C で $|E(C, G-C)| < k|C|$ を満たすものとして定義される．とくに1-孤立クリーク（単に孤立クリーク）に対しては，(入力)線形時間の列挙アルゴリズムとなる．線形時間のアルゴリズムはウェブ規模のグラフに対しても十分実用的で，約1億点規模のウェブグラフに対しても孤立クリークの列挙は可能となり，その結果，実際にコミュニティの抽出にも成功している[29]．

この方向の今後の可能なアプローチとしては，実際にウェブグラフに頻出する特徴的な構造を発見，同定し，それを列挙が容易な（出現個数がある程度制限されるような）構造として定式化し，実際にそれを列挙するアルゴリズムを構築することである．

6．おわりに

本稿では，ウェブ研究の一分野として，ウェブグラフとそれに関する理論的な研究やその手法のいくつかを紹介した．しかし，幅広い分野に渡るウェブ研究の成果は，著者の力量では到底カバーできるものではなく，また著者の興味の面からも偏った解説になっているかも知れない．それでも，ウェブの中でも比較的理論的な研究に，オペレーションズ・リサーチや情報科学の基礎的な手法がどのように取り込まれているか，その一端が分かっていただければ幸いである．

ウェブの進化のスピードは急速で，今後ウェブの世界で起こることや，どのような理論や技術が生まれるかを予想することは難しい．しかしながら，そのような変化を少しでも感じとっていれば，ページランク技術のように，われわれの手法が活躍できる場があるに違いないと思う．

謝辞　本稿の執筆の機会を与えてくださった，文教大学の根本俊男氏に感謝いたします．

参考文献

[1] R. Albert, H. Jeong and A.-L. Barabâsi. Diameter of the World-Wide Web. *Nature* **401**, 130-131, 1999.

[2] A.-L. Barabâsi and R. Albert. Emergence of scaling in random networks. *Science* **286**, 509-512, 1999.

[3] K. Bharat, B. Cheng, M. Henzinger and M. Rühl. Who links to whom: mining linkage between web sites. *Proc. 5th IEEE ICDM*, 51-58, 2001.

[4] B. Bollobás. *Random Graph* (*2nd edition*). Academic Press, 2001.

[5] B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica* **24**, 5-34, 2004.

[6] A. Z. Broder, S. R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins and J. L. Wiener. Graph structure in the web. *Computer Networks* **33**, 309-320, 2000.

[7] S. Bornholdt, H. G. Schuster (Eds.). *Handbook of Graphs and Networks: From the Genome to the Internet*. Wiley-VHC, 2003.

[8] C. Cooper and A. Frieze. Crawling on web graphs. *Proc. 34th ACM STOC*, 419-427, 2002.

[9] C. Cooper and A. Frieze. A general model of web graphs. *Random Struct. and Algor*. **22**, 311-355, 2003.

[10] C. Cooper and A. Frieze. The cover time of the preferential attachment graph. *Proc. 16th SODA*, 961-970, 2005.

[11] G. M. Del Corso, A. Gullí and F. Romani. Fast PageRank computation via a sparse linear system. *Internet Math*. **2**, 251-273, 2006.

[12] G. W. Flake, S. Lawrence and C. L. Giles. Efficient identification of web communities. *Proc. 6th ACM SIGKDD*, 150-160, 2000.

[13] U. Feige. A tight upper bound on the cover time for random walks on graphs. *Random Struct. and Algor*. **6**, 51-54, 1995.

[14] U. Feige. A tight lower bound on the cover time for random walks on graphs. *Random Struct. and Algor*. **6**, 433-438, 1995.

[15] D. Fogaras and B. Rácz. Towards scaling fully personalized PageRank. *Proc. 3rd WAW*, 105-117, 2004.

[16] S. Ikeda, I. Kubo, N. Okumoto and M. Yamashita. Impact of local topological information on random walks on finite graphs. *Proc. 30th ICALP*, 1054-1067, 2003.

[17] H. Ito, K. Iwama and T. Osumi. Linear-time enumeration of isolated cliques. *Proc. 13th ESA*, 119-130, 2005.

[18] J. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM* **46**, 604-632, 1997.

[19] J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan and A. S. Tomkins. The Web as a graph: measurements, models, and methods. *Proc. 5th COCOON*, 1-17, 1999.

[20] J. Kleinberg and S. Lawrence. The structure of the Web. *Science* **294**, 1894-1895, 2001.

[21] 久保幹雄, 田村明久, 松井知己 (編). 応用数理計画ハンドブック, 14. 4 節 列挙問題. 朝倉書店, 2002.

[22] R. Kumar, P. Raghavan, S. Rajagopalan and A. Tomkins. Trawling the Web for emerging cyber-communities. *Proc. 8th WWW Conf.*, 403-416, 1999.

[23] Y. Kurumida, H. Ono, K. Sadakane and M. Yamashita. Forest search: A paradigm for faster exploration of scale-free networks. Manuscript, 2006.

[24] 増田直紀, 今野紀雄. 複雑ネットワークの科学. 産業図書, 2005.

[25] J. W. Moon and L. Moser. On cliques in graphs. *Israel J. of Math.* **3**, 23-28, 1965.

[26] M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.* **45**, 167-256, 2003.

[27] L. Page, S. Brin, R. Motwani and T. Winograd. The PageRank citation ranking: Bring order to the Web. Technical Report of the Stanford Digital Library Technologies Project, 1998.

[28] 宇野毅明. 効率的な列挙アルゴリズムの構築と利用 (3). 人工知能学会誌 **18**, 586−591, 2003.

[29] Y. Uno, Y. Ota, A. Uemichi and M. Umano. Mining communities and detecting link farms in the Web by isolated cliques. *Proc. 2nd ICKEDS*, 179-187, 2006.

[30] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature* **393**, 440-442, 1998.