

## 地域ウェブ空間の特徴抽出

石村 享久<sup>†</sup> 申 吉 浩<sup>††</sup> 上 林 憲 行<sup>†††</sup>

グラフ理論的手法による WWW の全体の構造の解明が進展しているのに対して、地域の制約下における WWW の部分グラフのネットワーク構造の実証研究は事例に乏しい。本論文では、特定の地域に所属するサイトとそれらの間のリンクからなる WWW の部分グラフを地域ウェブ空間と定義し、その特徴を実証的に提示することを目的とする。分析における計測の指標には、内部到達確率の分布、サイトの次数（リンク保持数）の分布、ハブ度とオーソリティ度によるサイトの「価値」の分布の3つを用いる。内部到達確率の分布の評価方法では、従来の大きな連結成分が1つしか存在しないと仮定した定義を、新たに大きな連結成分が複数存在する場合にも適用可能な定義へと拡張した指標を提案する。また、サイトの価値の評価では、従来のネットワークの規模に依存しない相対的な値を与える指標に対し、新たに絶対的な価値を与える絶対オーソリティ度・絶対ハブ度を提案し、地域ウェブ空間の分析に用いる。実際の調査では、16地域の内部到達確率分布の評価と、その中からさらに同程度規模の8地域を選んで次数分布およびハブ度・オーソリティ度の分布の評価を行った。その結果、地域ウェブ空間では、WWW全体に比べて内部到達確率が低い値に分布していること、次数分布およびハブ度分布では偏向が小さく内部求心性が弱いことが判明した。

### Feature Extraction of Local Web Space Based on Three Aspects

YUKIHISA ISHIMURA,<sup>†</sup> KILHO SHIN<sup>††</sup> and NORIYUKI KAMIBAYASHI<sup>†††</sup>

While many researches focus on the graph-theoretic nature of WWW, only little is known about that of the local-web-space, a subgraph of WWW belonging to a particular geographical area. This paper reveals some nature of the local-web-space. For this purpose, we examine the distributions of the reachability (the portion of the sites in the entire network that are reachable from a site), the degree (the number of links that enter and leave a site) and of the authority/hub weight. All these distributions when applied to WWW are known to indicate that sites in WWW are tightly connected to each other.

Particularly, the paper includes the following. (i) The known method to measure the reachability is extended so that it is applicable to the case where more than one big connected components exist in a network. (ii) The authority/hub weight is extended so that they are proportional to the size of a network. We select 16 areas in Japan to examine the distribution of the reachability and 8 to examine the distributions of the degree and the authority/hub weight. The evaluation indicates that sites in a local web-space loosely connect to each other in comparison with those in WWW.

#### 1. はじめに

WWWはすでに社会に浸透し生活に必要なインフラとなっており、成長を続けているが、その全体像はいまだ明らかではない<sup>1)</sup>。

一方、ネットワーク中におけるノードの次数（保持リンク数）とノードの出現頻度とがべき乗法則に従う、

所謂、スケールフリーネットワーク（scale free network）の事例が社会や自然界に存在する多様なネットワークにおいて続々と発見されたことで、スケールフリーネットワークの研究は燎原の火のごとき勢いで広まっている。WWWがスケールフリーネットワークであることも1999年に報告され<sup>2)</sup>、WWWをグラフ理論的手法により解析しようとする研究の呼び水となった。しかしながら、特定の地域において地理的・社会的・人的な制約下で成長するウェブネットワークに着眼した先行研究事例はわずかである<sup>3)-5)</sup>。

本論文では、特定の地域に属するウェブサイトとそれらウェブサイト間のリンクから構成されるウェブネットワークとして定義される地域ウェブ空間に対して、

<sup>†</sup> 東京工科大学大学院  
Graduate School, Tokyo University of Technology

<sup>††</sup> 東京大学先端科学技術研究センター  
Research Center for Advanced Science and Technology,  
The University of Tokyo

<sup>†††</sup> 東京工科大学  
Tokyo University of Technology

グラフ理論的な指標を適用し、実際に地域ウェブ空間のデータを調査し報告する。

具体的には、Yahoo!JAPAN ディレクトリサービスの地域カテゴリに登録されているサイトを用いて構成される地域ウェブ空間に対し、以下の3つの指標を用いる（各指標の正確な定義に関しては、2章以降を参照されたい）。

- (1) 内部到達確率分布：ネットワーク内のサイトから他のサイトへの到達確率の分布を調べる。Broderら<sup>6)</sup>は、到達可能確率の特性から、ネットワーク内のサイトをSCC, IN, OUT, TENDRILLS, DISCONNECTのカテゴリに分類すると、その構成比が内部到達確率分布の特性を表す指標となることを示した。本論文では、Broderら<sup>6)</sup>の指標の定義を、地域ウェブ空間に適用できるよう拡張している。
- (2) 次数分布：スケールフリーネットワークにおいて、サイトの出現確率と次数との間の関係式におけるべき指数を指標とする。べき指数は負の値となるが、その絶対値が大きいほど、特定の少数のサイトがリンクを独占していることを表す。
- (3) オーソリティ・ハブ度分布：ネットワーク中のリンクの分布の特性から、サイトの重要性を測る指標である。よりハブ度の高いサイトからリンクを集めているサイトのオーソリティ度を高く、また、よりオーソリティ度の高いサイトへリンクを出しているサイトのハブ度を高く定義する。

今回の実証調査では、これらの指標を地域ウェブ空間に適用するのに先立って、以下の2点で指標に拡張を加えた。

まず、Broderら<sup>6)</sup>の指標は、WWWにおける最大の連結成分が全サイトの90%を占めるという事実を前提としている。しかしながら、地域ウェブ空間では大きな連結成分が複数存在する可能性を否定できないため、指標の定義の拡張が必要であると考え、連結成分を特定する方法に改良を加えた。

第2に、従来のオーソリティ度およびハブ度は、単一のネットワークの中に存在するサイトの重要性を比較するための指標であり、異なるネットワーク中の2つのサイトの重要性を比較するには不適当である。たとえば、明らかに異なる重要性を有するサイトがそれぞれのネットワークにおいて同じオーソリティ度・ハブ度を有していたり、逆に、同等の重要性を持つと思われるサイトが異なるオーソリティ度・ハブ度を有して

いたりするケースがありうる。そこで、従来の指標を補完する目的で、同規模の異なるネットワークの間でサイトの絶対的な重要性を比較するための指標（絶対オーソリティ度および絶対ハブ度）を新たに定義した。

本論文は以下のように構成される。2章は本研究における地域ウェブ空間の定義と抽出方法の妥当性を説明し、3章では内部到達確率分布、4章では次数分布のべき指数、5章ではオーソリティ度・ハブ度の各指標に対して、指標に関する説明と、地域ウェブ空間への適用結果を報告する。

## 2. 地域ウェブ空間の定義

地域ウェブ空間は、特定の地域に「所属する」サイトと、これらのサイトを連結するリンクとから構成されるネットワークと定義する。しかし、「地域への所属」の定義はそれほど明確に与えられてはいない。たとえば、サイトの物理的な所在地が、地域への所属を表していないことは明らかである。

本研究の興味は、「地域に所属するサイト」が「地域に所属するサイト」に対して、どのようにリンクを結び、ネットワークを成長させていくかという点にある。したがって、サイトの地域への所属性は、サイトの運用者の意識とサイトの見え方の2点から判断されるべきであると考えられる。すなわち、サイトの運用者が、自らのサイトが特定の地域に所属していることを意識しながら、同一の地域に所属していると判断したサイトに対して張るリンクこそが、地域の特性に制約されたリンクであるとする。本研究では「地域に所属する」サイトを以下の2条件を満足するものと定義する。

- (1) サイトの運用者が地域に所属しているという意識を有していること。
- (2) サイトのコンテンツが地域性を有していること。

上記の2条件を満足するサイトを選択することができれば、それらのサイトを結びリンクは、地域の特性の制約下に生成されたと考えて差し支えないであろう。

以下では「地域に所属する」サイトを実際を選択する方法を考察する。

まず、サイト運用者の所属意識を指標とした選択方法について述べると、検索サイトにおけるカテゴリによる分類を用いる手法が有効である。たとえば、Yahoo!JAPANの場合では、サイトの運用者が自分のサイトの特徴を最もよく現していると思うカテゴリを申請し、Yahoo!JAPANは申請に基づいてサイトを登録するという方法がとられる。すなわち、サイト運用者の意識がこれほど明確に示されることはなく、また、Yahoo!JAPANのような大手の検索サイトであれば、

大きな母集団を期待できることから、選択手法としては有効であると考えられる。

コンテンツの地域性に関しては、以下の2つの理由から有効な選択法を得ることは難しいと考えられる。第1に、地域性の有無の判断に適切な基準がないこと、第2に、多数のサイトに対してコンテンツの内容を効率良く調査する方法を得ることが困難なことである。コンテンツを人が読んで地域性の有無を逐一判断することは不可能である。したがって、判定は機械的に判別できる特徴によらなければならないが、これは判定基準の選択の幅を著しく狭め、判定結果の信憑性を失わせる。

以上の考察から、本研究では、地域に所属するサイトの選択にあたって、大手検索サイト（具体的にはYahoo!JAPAN）における特定の地域のカテゴリを利用することとする。

本研究で採用したサイトの選択方法の精確性に関しては、次の懸念が存在する。

- サイト運用者が複数のカテゴリ候補を申請した場合、検索サイト側で恣意的に最適と思われるものを登録する。このため、サイト運用者が地域への所属を意識していたとしても、地域に所属するサイトとして調査の俎上にのぼらない危険性がある。
- コンテンツが地域性を有しているにもかかわらず、サイト運用者が地域カテゴリに申請しない場合、このようなサイトが排除される危険性がある。

論理的には、これらの危険性に起因して、サンプルに偏りがでる恐れがある。

しかしながら、上記の事由により排除されるサイトは、検索サイト、あるいは、サイト運用者側の視点からみて、コンテンツが有する地域性よりも、他の要素の方がサイトの性格付けとして支配的であると判断されたわけであるから、他者の判断に依拠しているという弱点はあるとしても、地域カテゴリにより選択されたサンプルは地域に所属するサイトの母集団を十分に代表していると考えられる。

### 3. 内部到達確率の分布に基づく地域ウェブ空間の特徴

#### 3.1 評価指標の定義

サイトを頂点、リンクを辺とすると、WWWは有向グラフである。サイトAからリンクをたどってサイトBに到達できるとき、サイトAからサイトBへは到達可能であるという。さらに、ネットワーク中のサイトAを特定し、同じネットワークからランダムに選んだサイトがサイトAから到達可能である確

率を、サイトAの内部到達確率であると定義する。

Broderらの研究<sup>6)</sup>では、互いに到達可能なサイトの集合をSCCと呼び、SCCを基軸とした5つのカテゴリにサイトを分類し、その構成比率をもってWWW空間の到達確率分布の特性を表す指標としている。また、WWWにおける最大の連結成分は他の連結成分に対して圧倒的に大きいという特徴を報告している。SCC(Strongly connected component)

WWWの最大の相互連結成分

IN SCCに属さずSCCのサイトに到達可能なサイトの集合

OUT SCCに属さずSCCのサイトから到達可能なサイトの集合

TENDRILS SCC, IN, OUT以外のサイトのうち、リンクの方向を無視すればSCCに到達可能なサイトの集合

DISCONNECT リンクの方向を無視してもSCCに到達不可能なサイトの集合

上記のカテゴリの構成比率と到達可能確率との関係を概説する。SCC, IN, OUT, TENDRILS, DISCONNECTが全ネットワークに占める割合をそれぞれ $x\%$ ,  $y\%$ ,  $z\%$ ,  $u\%$ ,  $(100 - x - y - z - u)\%$ と表すと、以下が成り立つ。

- SCC中のサイトから到達可能なサイトの割合は $x + z\%$ である。
- IN中のサイトから到達可能なサイトの割合は $x + z\%$ から $x + y + z\%$ の範囲に分布する。
- OUT中のサイトから到達可能なサイトの割合は $0\%$ から $z\%$ の範囲に分布する。
- TENDRILS中のサイトから到達可能なサイトの割合は $0\%$ から $z + u\%$ の範囲に分布する。
- DISCONNECT中のサイトから到達可能なサイトの割合は $0\%$ から $(100 - x - y - z - u)\%$ の範囲に分布する。

上記の考察に基づいて到達確率確率の分布を模式的に表すと図1のようになる。

Broderら<sup>6)</sup>は、WWWにおいて、

$$\text{SCC} : \text{IN} : \text{OUT} : (\text{TENDRILS} + \text{DISCONNECT}) \\ \approx 1 : 1 : 1 : 1$$

が成立することを報告している。

本論文では、地域ウェブサイトはこの指標を適用する際に、地域ウェブ空間において大きな連結成分が複数存在する可能性を否定できない点に配慮した。つまり、最大の連結成分が第2位以下の連結成分に比較して圧倒的に大きいという事実は観察によってWWWで確かめられたものであり、WWWの部分グラフで

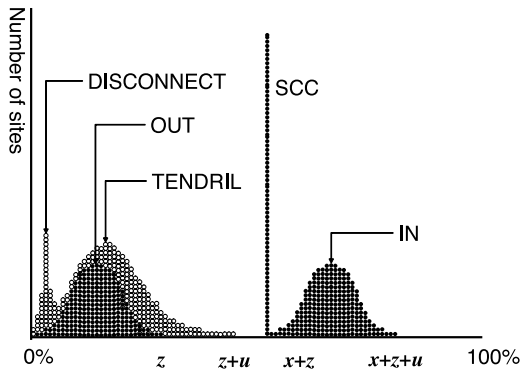


図 1 到達可能確率の分布

Fig. 1 Distribution of reachability probability.

ある地域ウェブ空間でも同じ事実が成立することは保証されないからである。

本論文では、Broder らの手法<sup>6)</sup>を拡張して、複数の大きな連結成分が存在する可能性を斟酌して、カテゴリを以下の流れで再帰的に定義する。

- (ステップ 1) ネットワーク中に含まれる最大の相互連結成分を選び、SCC とする。リンクの向きを問わず SCC と結ばれているサイトの総数が所定の閾値（本論文では 20）を下回る場合、あるいは、SCC が存在しない場合、ネットワークのすべてのサイトを DISCONNECT に分類し、分類を終了する。
- (ステップ 2) SCC に属さず、かつ、SCC のサイトに到達可能なサイトの集合を IN とする。
- (ステップ 3) SCC に属さず、かつ、SCC のサイトから到達可能なサイトの集合を OUT とする。
- (ステップ 4) SCC, IN, OUT のいずれにも属さず、かつ、向きを無視してリンクによって SCC につながるサイトの集合を TENDRILS とする。
- (ステップ 5) ネットワークから、SCC, IN, OUT, TENDRILS に属するサイトと、これらのサイトを始点あるいは終端とするリンクとをすべて取り除いた、新たなネットワークを入力としてステップ 1 に戻り、次の SCC, IN, OUT, TENDRILS を計算する。

上記の方法は、SCC, IN, OUT, TENDRILS を、必ずしも 1 組とは限らずに出力する点に特徴があり、したがって、大きな連結成分が複数存在するケースをカバーしている。ただし、実際に本論文において 16 地域の地域ウェブ空間に適用した事例では、第 2 位以降の連結成分のサイズは閾値より小さく（ほとんどが孤立サイト）、結果として指標として抽出される SCC, IN, OUT, TENDRILS は 1 組のみであった。

表 1 各地域ウェブ空間のカテゴリ構成  
Table 1 Categories in local web spaces.

サイト数	SCC	IN	OUT	TNDL	DCNT	
八王子	460	3.7%	13.7%	8.5%	12.6%	61.5%
相模原	533	9.9%	11.3%	4.7%	6.8%	67.4%
町田市	361	5.3%	14.4%	5.0%	14.7%	60.7%
日野市	99	22.5%	11.2%	18.0%	0%	48.3%
藤沢市	487	7.0%	20.3%	9.2%	8.6%	54.8%
足立区	486	5.6%	9.7%	8.0%	3.9%	72.8%
長野市	518	3.1%	10.2%	1.0%	17.2%	62.3%
盛岡市	361	8.9%	10.8%	8.6%	5.0%	66.8%
松本市	359	4.5%	17.5%	4.2%	23.7%	50.1%
宇都宮市	506	3.4%	12.1%	9.9%	15.1%	59.7%
岡山市	862	6.4%	12.9%	12.8%	16.1%	51.9%
倉敷市	413	11.4%	7.7%	27.8%	7.8%	45.3%
津山市	100	15.0%	29.0%	12.0%	2.0%	42.0%
山形市	260	34.2%	10.0%	25.8%	5.4%	24.6%
広島市	1313	4.6%	13.6%	9.1%	11.8%	60.9%
仙台市	1390	17.2%	14.7%	19.1%	6.9%	42.1%
WWW 全体		27.7%	21.3%	21.2%	21.5%	8.2%

TNDL は TENDRILS, DCNT は DISCONNECT の略

### 3.2 地域ウェブ空間における調査結果

16 地域の地域ウェブ空間を対象として、カテゴリの構成比率を測定した（表 1）。なお、比較のために載せた表中の WWW 全体は Border らによって示された 2000 年の結果である。

前節の定義は複数の SCC の存在を含むが、実際にはすべての地域で第 2 位以下の連結成分のサイズは閾値以下であったため、表 1 に示した調査結果は 1 組のカテゴリのみを含むものとなっている。具体的には、DISCONNECT に分類された最も大きな（無向グラフとしての）連結成分であっても 20 サイトを超えることはなく、おおかた 10 サイトを下回る数となる。さらには、DISCONNECT のほとんどは、地域ウェブ空間内のほかのサイトとリンクを持たないまま、孤立している。これら孤立サイトに関して、岡山県内の 3 市で調査したところ、孤立サイトの 50% 弱のサイトはサイト訪問者への宣伝や連絡を目的としており、同じ地域ウェブ空間内へのリンクに限らず、リンクタグそのものを含んでいないことが分かった。

調査結果から、地域ウェブ空間に関する以下の特徴が抽出された。

- SCC の割合に関しては、10% 以下が 11 地域、10% 超 15% 以下が 2 地域、15% 超 20% 以下が 1 地域と、WWW 全体の比率 24%<sup>6)</sup> に比較して低くなる傾向を示している。逆に、山形市のみは例外で、34.2% と WWW 全体での数値を超えている。
- IN や OUT の比率に関しては、IN 側が大きい調査結果（藤沢市、長野市等）、OUT が大きい調査結果（倉敷市等）、IN と OUT がつりあっている調査結果（足立区、盛岡市、岡山市等）が混在し、地域ウェブ空間に共通する特徴は見出されない。
- DISCONNECT の比率において、地域ウェブ空間の特徴が最も顕著に現れる。WWW 全体にお

ける DISCONNECT の比率は 8.2%<sup>6)</sup> ときわめて小さいのに対し、山形以外の地域ウェブ空間では全体の 50% 近くを占めるにいたっている。例外である山形市でも 24.6% であり、最も高い値を示した足立区では 72.8% にも達している。

#### 4. 次数分布のべき指数に基づく地域ウェブ空間の特徴

##### 4.1 評価指数の定義

WWW はスケールフリーネットワークであることが知られている<sup>2)</sup>。すなわち、サイトを特定して、リンクによりそのサイトを直接参照しているサイトの数を入次数  $d_i$ 、逆に、そのサイトがリンクにより直接参照しているサイトの数を出次数  $d_o$  とすると、入次数  $d_i$  (出次数  $d_o$ ) のサイトの存在確率  $p_i(d_i)$  ( $p_o(d_o)$ ) は、べき指数  $k_i$  ( $k_o$ ) に対して、 $p_i \propto d_i^{-k_i}$  ( $p_o \propto d_o^{-k_o}$ ) と表すことができる。ただし、次数 0 の孤立サイトは無視する (0 のべき乗は 0 なので)。

ネットワークがスケールフリーであるという性質は、特定のサイトに比較的多くのリンクが集中する不平等なネットワークであることを意味し、集中の傾向が強いほどべき指数の絶対値は大きくなることから、べき指数を集中度を表す指標と見なすことができる。

##### 4.2 地域ウェブ空間における調査結果

ほぼ同等のサイズを有する 8 つの地域ウェブ空間 (八王子市、相模原市、藤沢市、足立区、長野市、盛岡市、松本市、宇都宮市) がスケールフリーネットワークであることを確認し、次いで、入次数分布および出次数分布のべき指数の値を調査した。スケールフリー性および次数分布のべき指数を調べる場合には孤立サイトを除外することを前述したが、地域ウェブ空間では DISCONNECT を構成するサイトのほとんどは孤立サイトであることが調査の結果分かっているので、地域ウェブ空間でのスケールフリー性および次数分布のべき指数の評価は SCC を含む (無向グラフとしての) 連結成分に対する評価に近似する。

調査の結果、図 2 が示すように、調査対象とした地域ウェブ空間における次数分布はべき乗分布に従う。

また、地域ウェブ空間における次数分布のべき指数を、表 2 に示す。調査対象とした地域ウェブ空間では、Kumar ら<sup>7)</sup> の報告による WWW 全体に対するべき指数 (入次数に対して  $-2.10$ 、出次数に対して  $-2.38$ ) に比較して、その絶対値が小さく、リンクの集中に関して二極化が進行していないことが分かる。

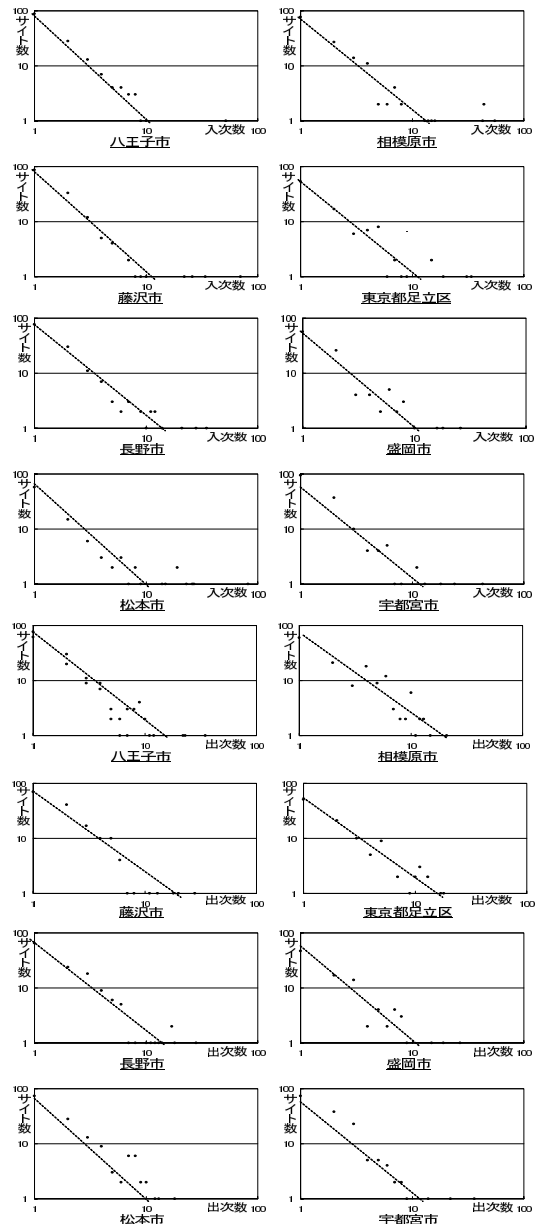


図 2 地域ウェブ空間の入次数および出次数の分布

Fig. 2 Degree distribution for local web spaces.

#### 5. ハブ度とオーソリティ度の分布に基づく地域ウェブ空間の特徴

##### 5.1 評価指標の定義

ネットワークにおけるノードの「価値」の分布を示す指標として、ハブ度とオーソリティ度がある<sup>8)</sup>。サイトのオーソリティ度は、そのサイトを直接参照しているサイトのハブ度の総和が大きいほど大きい値をとり、逆に、ハブ度は、そのサイトが直接参照してい

表 2 地域ウェブ空間の度数分布のべき指数

Table 2 Exponent of degree distribution for local web spaces.

	入次数のべき指数	出次数のべき指数
八王子市	-1.26	-1.19
相模原市	-1.02	-1.41
藤沢市	-1.09	-1.53
足立区	-1.08	-1.35
長野市	-1.27	-1.46
盛岡市	-1.27	-1.23
松本市	-1.07	-1.57
宇都宮市	-1.29	-1.40

る先のサイトのオーソリティ度の総和が大きいほど大きい値をとるように再帰的に定義される。直感的にいえば、オーソリティ度は他のサイトに参照されるべき「内容価値」を表し、ハブ度はそのサイトの参照先の内容価値である「参照価値」を表す指標である。

オーソリティ度およびハブ度は以下の要件を満足するように定義され、この2点で入次数および出次数と相違する。

ネットワークのサイズによらない正規化 一般に、次数はネットワークのサイズに依存する値をとる。実際、 $n$  個のノードから構成されるネットワークにおける次数の上限は  $n-1$  である。したがって、次数を調べる目的がネットワークの極化（少数のノードにリンクが集中すること）の度合いを調べる点にある場合、次数そのままでは指標としては使えない。ネットワークのサイズに依存しないように正規化が必要がある。正規化の1つの方法としては、次数の上限で次数を割るものがあり、 $Cd(v) = \frac{d(v)}{n-1}$  によって与えられる指標を、次数センタと呼ぶ<sup>9)</sup>。オーソリティ度およびハブ度も、ネットワークのサイズに依存しないように正規化された指標である。

リンクの重み付け 次数センタは正規化の要件を満足するが、リンクに重み付けを行わず平等に扱う点で不満が残る。たとえば、出次数の計算では、高い内容価値を持つサイトへのリンクも低い内容価値を持つサイトへのリンクもともに1単位として勘定する。ところが、実際には、少数の内容価値の高いサイトに対してリンクを張るサイトの方が、多数の内容価値の低いサイトにリンクを張るサイトより、高い参照価値を有しているかもしれない。入次数に関しても同様で、高い参照価値を持つサイトからのリンクもその他のサイトからのリンクも同じ1単位で勘定する方法では、忠実に内容価値の評価ができない危険がある。オーソリ

ティ度およびハブ度は、互いに互いをリンクの重みとする再帰的定義により、より忠実な内容価値および参照価値とを表現する。

オーソリティ度とハブ度は、上記2つの要件を満足するように、具体的には以下のように定義される。

サイト  $i$  のオーソリティ度およびハブ度をそれぞれ  $x_i$  および  $y_i$  とし、ネットワークのサイズの影響を排除するために  $\sum_{i=1}^n x_i^2 = 1$  および  $\sum_{i=1}^n y_i^2 = 1$  の条件を設け、正規化する。

ネットワークの隣接行列  $A = [a_{ij}]$  を

$$a_{ij} = \begin{cases} 1 & \text{サイト } i \text{ がサイト } j \text{ を参照している} \\ 0 & \text{サイト } i \text{ がサイト } j \text{ を参照していない} \end{cases}$$

により定める。

サイト  $i$  の参照機能は、そのサイトが参照しているサイトのオーソリティ度の和、すなわち、 $\sum a_{ij}x_j$  で測ることができ、サイト  $i$  の内容価値は、 $\sum a_{ji}y_j$  で測ることができる。したがって、

$$y_i \propto \sum_{j=1}^n a_{ij}x_j, \quad x_i \propto \sum_{j=1}^n a_{ji}y_j \quad (1)$$

の関係が成り立ち、 $\vec{x} = (x_1, \dots, x_n)^T$  および  $\vec{y} = (y_1, \dots, y_n)^T$  は、行列  $A^T A$  および行列  $AA^T$  の固有ベクトルとなる。特に、各行列の固有ベクトルのうち（絶対値が）最大の固有値を持つ固有ベクトルを、オーソリティ度およびハブ度と定義する。

サイト  $j$  が持つ参照機能は  $\sum a_{jk}x_k$  で評価されるので、サイト  $i$  の情報価値は、固有値  $\lambda$  に対して、

$$\sum_{j=1}^n a_{ji} \left( \sum_{k=1}^n a_{jk}x_k \right) = \lambda x_i \quad (2)$$

で測られる。すなわち、各サイトの内容価値の自乗和を最大にする固有ベクトルがオーソリティ度である。

オーソリティ度およびハブ度の絶対値はつねに定数 ( $=1$ ) 以下であり、その分布を調べることにより、ネットワークのサイズに依存することなくネットワークの極化の状態を調べることができる。

その一方で、重み付けの要件を満足したうえで、ノードの絶対的な価値を測る必要がある場合がある。たとえば、完全グラフでは、すべてのサイトのオーソリティ度とハブ度は一律に  $\frac{1}{\sqrt{n}}$  となるが、一様に疎なネット

$\vec{x}^T A^T A \vec{x}$  を  $|\vec{x}| = 1$  の条件のもとで最大にする  $\vec{x}$  をオーソリティ度と定義することもできる。

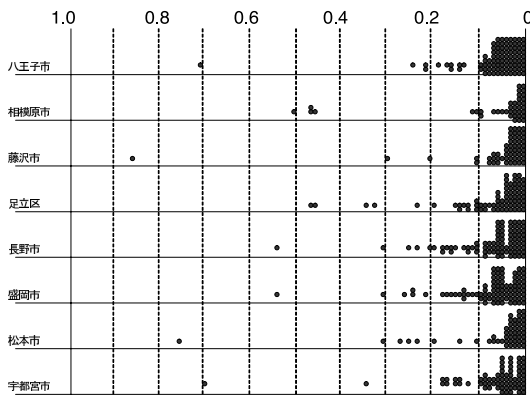


図 3 オorsoリティ度の分布  
Fig. 3 Distribution of authority weight.

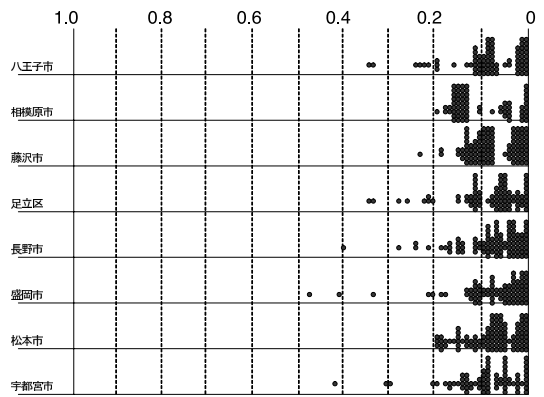


図 4 ハブ度の分布  
Fig. 4 Distribution of hub weight.

ワーク（リング等）でも同じ現象が起きる。すなわち，オorsoリティ度とハブ度の分布から分かることは，内容価値と参照機能の極化の状態のみであり，密な結合で実現されているか，疎な結合で実現されているかについては情報を得ることができない。

本論文では，この問題を解決するために，絶対オorsoリティ度および絶対ハブ度を定義する。オorsoリティ（ハブ）度と絶対オorsoリティ（ハブ）度との関係は，次数センタと次数の関係との関係に等しいと考えれば，理解しやすいかもしれない。

ネットワーク全体の内容価値の総和はネットワークを構成するサイト数  $n$  に比例するという自然な仮定を受け入れると，絶対オorsoリティ度の総和はサイト数に比例すると考えることができる。一方，ノードの絶対ハブ度は，ノードから他のノードを終端として張られるリンクの重み付け和であり，重みには終端ノードのオorsoリティ度を用いる。すなわち，本論文では，絶対オorsoリティ度と絶対ハブ度を以下のように定義する。

サイト  $i$  の絶対オorsoリティ度  $x_i^*$  を

$$x_i^* = \frac{n}{\sum_{i=1}^n x_i} x_i \tag{3}$$

と定め，サイト  $i$  の絶対ハブ度  $y_i^*$  を

$$y_i^* = \sum_{j=1}^n a_{ij} x_j^* \tag{4}$$

により定める

## 5.2 地域ウェブ空間における調査結果

### 5.2.1 ハブ度とオorsoリティ度の分布

地域ウェブ空間の各サイトのオorsoリティ度とハブ度の分布を，次数分布の場合と同じ8つの地域ウェブ空間で調べた（図 3 および図 4）。

表 3 絶対ハブ度上位ランキング  
Table 3 Absolute hub weight rank.

	八王子	相模原	藤沢	足立	長野	盛岡	松本	宇都宮
1 位	86.6	133.6	101.9	81.6	94.7	71.8	98.6	126.9
2 位	84.9	116.6	84.1	77.9	66.6	61.0	97.0	92.0
3 位	61.1	115.1	83.7	62.8	57.9	50.0	95.8	89.4
4 位	58.6	113.6	73.5	58.3	56.9	32.1	93.3	62.2
5 位	57.5	108.0	71.1	51.3	50.8	31.1	93.0	57.3
1	460	533	487	486	518	361	359	506
2	177	174	220	132	194	120	179	200

1, 地域ウェブ空間の規模 2, 最大連結成分の規模

オorsoリティ度の分布では，群を抜いた高いオorsoリティ度を持つサイトが存在する地域ウェブ空間（藤沢市，松本市等）もあれば，最も高いオorsoリティ度であってもそれほど突出していない均質な地域ウェブ空間（足立区，相模原市等）もあり，特定の傾向は見られない。一方，ハブ度に関しては，調査対象の8つの地域で一様に均質な分布をとることが分かる。

### 5.2.2 絶対ハブ度の比較

前述のように，ハブ度に関しては，今回調査を行った地域ウェブ空間すべてで均質な分布がみられる。ハブ度の分布が等しいという事実は，調査対象とした地域ウェブ空間の間でサイトが有する絶対的な参照価値に差があるか否かについては何も示唆しないので，地域ウェブ空間においてハブ度が上位5位に入るサイトに対する絶対ハブ度を調べることで，地域ウェブ空間の間で絶対的な参照価値に差があるか否かを調べた（表 3）。結果，絶対ハブ度には地域間での大きな差がないことが分かる。

また，絶対ハブ度の上限はネットワークのサイト数であり，地域ウェブ空間のすべてにつながる場合に上限値をとる。表 3 に示される値をこの上限値と比較すると，地域ウェブ空間におけるハブの強さが一様に弱

いこともあわせて分かる．

## 6. まとめと今後の研究

### 6.1 まとめ

本論文では、16 地域を対象に、(1) 内部到達確率の分布、(2) 次数分布、(3) オースリティ度・ハブ度の 3 つの指標の値を調査し、地域ウェブ空間の特徴として以下の結果を得た．

- 内部到達確率は低いレベルに分布する．特に、DISCONNECT が占める割合 (40~60%) は、WWW 全体の 8.2% に比較して格段に大きい．
- 地域ウェブ空間の次数分布はべき乗分布に従うべき指数の絶対値は入次数で 1.3 以下、出次数で 1.6 以下と、WWW 全体における値 2.10 および 2.38 と比較して、おしなべて低い．
- オースリティ度の調査では共通した特徴はみられない．一方、ハブ度の分布は均質であり、絶対ハブ度の値も共通に小さい．

### 6.2 今後の研究

内部到達確率分布、次数分布、オースリティ度とハブ度の各指標は、雑駁にいえば、ネットワークを構成するサイト間の結び付きを測る指標であり、地域における物理的および社会的な結び付きがウェブ空間にどのように影響を及ぼすであろうかという関心から取り上げたものである．

著者らは、地域ウェブ空間の特徴を明らかにするうえで、上記の視点に加えて、以下の視点が有効であると考へ、研究に着手している．

時系列におけるネットワーク構成の変化 生物において新しい細胞が生まれては死んでいくのと同様に、一見成長のみを追求しているかにみえる WWW にも代謝が存在する．地域ウェブ空間の各種指標の時系列的な変化をトレースすることは、地域ウェブ空間の代謝原理を解明するうえで、重要な手がかりを提供するものと考えられる．加えて、本論文で取り上げたような静的な (static) な指標のトレースのみではなく、アクセス解析等の動的な手法の導入は有効であると予想できる．

隣接地域間の地域ウェブ空間における結び付き 隣接する地域間の相互作用は、地域ウェブ空間の成長に重要な影響を与えている可能性があり、相互作用を測る指標に基づく調査は有意義である．たとえば、2 つのネットワーク間で存在可能な最大リンク数に対する現存リンク数の比率として定義されるリンク密度は、隣接地域における地域ウェブ空間の間の相互作用を測るうえで有効である．

## 参考文献

- 1) Lawrence, S. and Giles, C.: Searching the World Wide Web, *Science*, Vol.280, No.5360, pp.98-100 (1998).
- 2) Barabasi, A.L. and Albert, R.: Emergence of scaling in random networks, *Science*, Vol.286, pp.509-512 (1999).
- 3) Ishimura, Y., Shin, K. and Kamibayashi, N.: Graph-theoretic Nature of Local Web Spaces, *The 8th International Workshop on Network-Based Information Systems*, pp.22-26 (2005).
- 4) 横山拓哉：地域情報空間における WEB サイト群の静的構造の解明，山形大学理工学研究所修士論文，pp.1-38 (2004).
- 5) 上林憲行：インターネット空間を包含した地域社会デザイン：WEBLOGY と地域情報空間，*情報メディア*，Vol.39, No.3, pp.11-18 (2001).
- 6) Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A. and Wiener, J.: Graph Structure in The Web, *The 9th International World Wide Web Conference* (2000).
- 7) Kumar, R., Raghavan, P., SridharRajagopalan, D., Sivakumar, Tomkins, A. and Upfal, E.: The Web as a graph, *Proc. 19th ACM Symposium on Principles of Database Systems*, pp.1-10 (2000).
- 8) Kleinberg, J.M.: Authoritative Sources in a Hyperlinked Environment, *Proc. ACM-SIAM Symposium on Discrete Algorithms*, pp.668-677 (1998).
- 9) 金光 淳：社会ネットワーク分析の基礎，勁草書房 (2003).

(平成 17 年 5 月 31 日受付)

(平成 18 年 1 月 6 日採録)



石村 享久

2004 年東京工科大学メディア学部メディア学科卒業．東京工科大学大学院バイオ・情報メディア研究科メディアサイエンス専攻修士課程在学．World Wide Web や WWW の部分グラフの成長原理に関する研究に従事．



**申 吉浩**

1985年東京大学大学院理学系研究科修士課程修了。1990年富士ゼロックス(株)入社。2000年(株)アクセスチケットシステムズ取締役, 2004年東京大学先端科学技術研究センター協力研究員, 2006年同センター特認助手。セキュリティ技術, 著作権保護技術の研究に従事。

**上林 憲行**

1980年慶應義塾大学大学院理工学研究科博士課程修了。広島大学工学部助手, 富士ゼロックス(株)入社後, 主幹研究員(Research Fellow), 研究所所長等を歴任。メディアアーキテクチャ, 知識サービス, デジタルコンテンツ流通サービス等の研究開発に従事。2000年山形大学工学部情報科学科教授を経て2003年より現職。人工知能学会理事, 情報処理学会理事, 情報処理学会情報メディア研究会幹事・主査, 論文誌・会誌編集委員, インタラクシオン 98.99のプログラムおよび大会委員長を歴任。現在, 情報処理学会ネットワーク生態学研究G主査。サービス工学(ソーシャルキャピタルサービス, ケアコミュニケーションサービス等), ネットワーク生態工学(WEBLOGY, ソーシャルネットワーキング), 情報臨床工学に関心を持つ。