第　1　章
はじめに

1.1　研究の背景と目的

今日，急速に普及しているインターネットにおいて，その主要なサービスの一つとなっているのが WWW（World-Wide Web）である。WWW は，インターネット上で文字だけでなく音声や画像，動画などの情報もやりとりできるサービスであり，さまざまな情報を入手，発信することが誰にもでき，インターネットにおいて主要なサービスの一つとなっている．こうして、WWW 技術によって構築されたネットワークの上で、インターネットのユーザー達が、情報や知識を提供したり、あるいはお互いに利用しながら、情報や知識の流通が行われている。しかし、WWW ネットワークの範囲は広く、ネットワーク上でのこれらの資源の分布も均一ではない。そこで有効な情報の検索、さらには情報をいかに有効に流通させることできるかどうかが一つの大きな課題となっている。

WWW の特徴は，さまざまな情報を関連づけて次から次へと自由に世界中のネットワークに簡単にアクセスすることができるハイパーテキスト構造にある。その中で、特に重要な役割を果たしているのはハイパーリンク（以下リンクと呼ぶ）である。現在、リンク構造に注目した研究として、WWW 情報空間の弱い構造化[4]やアクセスの高速化[3]などの研究がある。しかし、これらの研究はいずれもネットワークへのアクセスを支援することを目的とするもので、ネットワークそのものについての議論を視野に入れていない。

本研究では、リンクの構造を解析することによって、すべて同じように見えていた

さまざまなネットワークを分析し、ネットワーク上の情報の分布や、ネットワークの形態と情報や知識の流通経路との関連性について検討することが目的である。

1.2　論文の構成

本研究は、ネットワークにおけるリンクの構成状況とページの価値との関連性に注目点を置いたもので、Web ページにおいて参照される側の普段見えないリンク（Hub リンクと呼ぶ）構造を解析する手法の提案と、ネットワークのリンク構造によるネットワーク特性の分析が特色となる。論文の構成として、研究の準備、研究方法と仮説の提案、実験３つの部分からなっている。

第１章では、研究の準備としてネットワーク社会やインターネット技術などに関する研究の現状と研究で使われる情報処理技術やプログラミング言語を紹介する。

第２章では、ネットワーク分析の仮説や、研究手法について説明をする。

第３章では、プログラムを作成する際に考案したネットワーク分析方法を紹介し、アルゴリズムの構成と動作について解説をする。

第４章と第 5 章では、人工的ネットワークを使ったシミュレーション実験とインターネット（WWW）を使った Web 実験に関して、それぞれの結果の比較とネットワーク形態の分析によって、先の仮説と提案の手法の有効性について検証する。

最後に、研究結果のまとめと課題について述べる。

第　2　章
現状の技術と研究に関する仮説

2.1　WWW とネットワーク社会

最近、地球規模の情報通信ネットワークの構築と WWW の技術の普及によって、インターネットのユーザーが爆発的に増えている。このおかげで、世界中にいるインターネットのユーザー達が、ブラウザの上で URL（Universal Resource Locator）を指定するだけで、世界のどこにあるホームページでも簡単にパソコン画面でみられる。一方、情報の利用をはじめ、情報の検索や流通、コミュニケーション、さらに会議やビジネスまでさまざまなことが WWW 上で行われている。

2.1.1　WWW

WWW はヨーロッパの原子核研究所（CERN）で開発されたソフトである。情報をハイパーテキスト形式で表した分散データベースシステムで、インターネット上の情報を統一的に得ることが出来る。その仕組みを説明するために、WWW の各構成要素の役割を次に示す。

- Web の動作

  WWW の主要な要素は、通信回線ネットワークで接続された複数台のホストコンピューターシステムである。モデム、ブラウザソフト、通信ソフトを備えたコンピューター（主にパーソナルコンピューター）が、WWW の入り口である。

Web ドキュメント（HTML ファイルとも呼ばれる）はホスト（サーバー）上にある。各ホストは固有のアドレスを持ち、これによってインターネット上のほかのコンピューターはこのコンピューターを探すことが出来る。また、Web ドキュメントの中にリンクが埋め込まれている。リンクはインターネットの通信プロトコルを使って、あるホストコンピューター上のドキュメントを別のホストコンピューター上のドキュメントにリンクしている。

- Web サイトにおけるページの構成

  Web サイトは、一つ以上のページを相互にリンクし、一つのパッケージにまとめたものである。Web サイトは一般的な情報から専門的な内容までを階層化しただけの単純な構造もあるし、サイト中でページ間にランダムにリンクが張られるような複雑なものもある。

- HTML

  HTML（Hyper Text Markup Language）は Web ドキュメントに埋め込まれている命令のセットである。ブラウザはこれらの命令セットを読み込み、文字や絵を画面に表示することが出来る。

- リンク

  リンクは Web ドキュメントに埋め込まれた HTML コマンドで、Web 上でドキュメント内を移動したり、あるドキュメントから別のドキュメントに移動したりするために使われる。

- URL

  URL は Web ドキュメントの場所を示すもので、ドキュメントはインターネット上のどのコンピューターにあるかをブラウザに指示する。

2.1.2　ネットワークとネットワーク社会

現在、ネットワークという言葉の意味が多様化している。大きく分類すると、物理的体系としてのネットワークと社会的体系としてのネットワークに分けることができる。ここで、本研究の基礎となるネットワーク（本研究の場合、情報通信ネットワークのことを指す）とこれによって生まれるネットワーク社会について説明する。

- ネットワーク

もともとネットワークという言葉は、物理的システムとして相互に密接な関連のある網状組織のことを指していた。情報通信ネットワークや物流システムネットワークなどがその例である。しかし、最近その言葉の意味が拡張され、物理的システムから社会現象までを分析するキーワードにしようとの意図が込まれるようになった。社会システムとしてのネットワークとは、「ネットワーク内の主体間の相互制御関係が、主としては情報や知識の通有を通じての説得・誘導によって行われるような社会システム」である[13]。こうしたいろんなネットワークの概念に基づき、ネットワークを「任意加入によって構成員が決まり、加入者が相互に影響を及ぼし合うグループ」と定義することができる[13]。

- ネットワーク社会

特に、最近インターネット（特にWWW）を利用することによって、コミュニケーション、情報流通、ショッピング、電子投票、物の売買、研究、教育さまざまな社会活動が、直接的な対面交渉を要求しないネットワークを通じて行われるようになった。私たちが生きている現実の社会とは別に、インターネット上で、もう一つの社会システムが形成されつつある。ネットワーク社会に関する研究は、現在、情報通信テクノロジーとしてのネットワークの発達によって、場所と時間の制約を超えて進むコミュニケーションや自律的なコミュニティの形成とそれらの連携の動きに基礎を置いた研究が行われている[13]。

本研究は、ネットワーク社会の主役である小集団ネットワーク（ネットワーク全体

を構成する部分的ネットワーク、ここでは、WWWネットワークの一部を指す）を研究対象とし、それを分析することによって、小集団ネットワークにおける知識や情報の分布、および知識や情報の流通経路などネットワークの状況や特徴を知るための一つのアプローチである。最終的に、ネットワーク社会全体における情報や知識の移動経路と分布特徴の解明を図りたい。

2.2 インターネットエージェントと JAVA

本研究で使われる技術とプログラミング言語を簡単に紹介する。プログラムやアルゴリズムに関する詳細な説明は、次の章（プログラムのアーキテクチャのところ）で説明する。

2.2.1 インターネットエージェント

ここでいうエージェントとは、人間から権利を委任されたパーソナルなソフトウェアアシスタントである。言い換えると、他の人があなたにしてあげられる何かを実行して、人間関係をシミュレートするコンピュータープログラムである。インターネットエージェントとは、インターネットを活動場として、動きまわしながら、人間から委任された仕事をこなすプログラムのことである。代表的なインターネットエージェントとして、以下のようなものがあげられる[14]。

- ウェブロボット(Web robot)、スパイダー(Spider)、ワンダラー(Wanderer)
- ウェブ取引エージェント(Web commerce agent)
- ワームとウイルス(Worm & Virus)
- ＭＵＤエージェントとチャッターボット(MUD agent & Chatterbot)

今回の研究にとって、Web 上の HTML ファイルとリンク情報は特に重要と考えた（2.3 節参照)。しかし、膨大なデータを集める仕事は人間にとって単純でありながら、非常に時間がかかる。その機械的な仕事を人間の代わりにやらせるためにインターネットエージェントを使う。

2.2.2 JAVA

本研究は、ネットワークを分析することが目的なので、実験用ツール（プログラム）作成用プログラミング言語として Java 言語を選んだ。Java を用いる利点は、マルチスレッドサポート、強力なネットワーク機能の API やマルチプラットフォーム対応などが挙げられる。Java 言語は WWW への適用を目的として開発された言語で、以下のような特徴を持っている[12]。

- コンパイラ＋インタープリタ型

  Java 言語では、ソースコードをコンパイルしてバイトコードに変換して、クラスファイルという特別なファイルに格納する。そして、このクラスファイルに格納されたバイトコードをインタープリットして、プログラムを実行する。

- コンピューター（プラットフォーム）に非依存

  コンピューターのハードウエア的な違いは仮想マシンと呼ばれるソフトに吸収されるので、プログラムは OS などが異なる別のネットワーク上のコンピューターにおいても動作をする。

- ネットワーク対応

  インターネット上で動く Telnet や Ftp、Web などほとんどのサーバーへの接続用 API を備えている。

- オブジェクト指向

  オブジェクト指向概念を完全サポートすることで、プログラムの開発効率が高くなっている。

- 安全で信頼性高い

  ファイルシステムへのアクセス制限やメモリ自動管理などによって、安全性が保証されている。

- マルチスレッド

  複数の動作をスレッドと呼ばれるプログラムに同時に分割し、分散処理手法をプログラムのレベルで実現可能である。

本研究では Java 言語を用いて、ページ情報とリンクを収集するプログラム（以下 Web ロボットと呼ぶ）をつくり、インターネット（WWW）上で走らせ、多量の情報を自動的収集させる。

2.3 ページの価値に関する仮説

本節では、今回の研究で使われる重要な概念と言葉の定義について説明する。現在、WWW 上にある膨大な量のページは、さまざまな人々によって作られているため、ページの表現形式は当然異なり、その内容もそれぞれ違う。Web ページの多様性が Web ページをみる人にとっては、大きな楽しみとなっている。一方、ページ価値を評価しようとすると大きな難問に直面する。なぜなら、言語処理的な解析でページ内容や価値を評価しようとすると膨大な計算時間を必要とするのみならず、評価基準にも多様なものが存在するからである。そこで、リンクの多さがページ内容に関する価値の高さを示していると仮定して、Web ページの直接的な内容の解析は行わず、Web ページのリンクに注目する。

リンクはコンピューターによってサポートされる一定の関係を使って、ネットワーク（WWW）内の各ノード（ページ）を結び付ける。これによって、エージェントがネットワーク上を迅速かつ容易に移動することができるという利点をもつ。Web ページ間の関係によって、リンクを以下のように分類する[2]。

- Authority Link

  Authority Link とは、あるページからほかのページへのリンクで、普通にいう Web ページのリンクのことである。（以下 Aut リンクと呼ぶ）

- Hub Link

  あるページがほかのページから張られているリンク。（以下 Hub リンクと呼ぶ）

しかし、上のリンク定義は抽象的な概念であり、絶対的なものではない。つまり、ある Web ページに埋め込まれているリンクはあのページにとって、Aut リンクとなると同時に、リンク先のページにとっては Hub リンクともなる。（図 2.3.1）

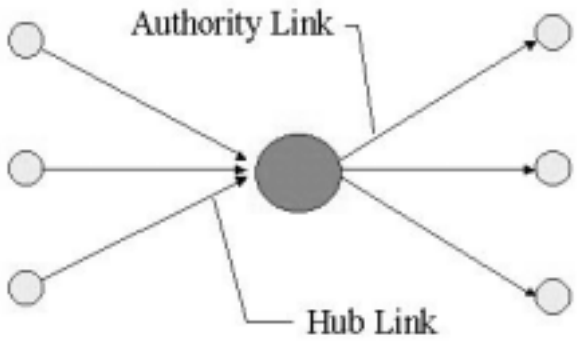


図 2.3.1　Link

Web ページの Link は、人間が自分の価値観と努力を基に作られたものなので、その人の価値観とその人が持っている情報・知識を（何らかの形で）含んでいると考えられる。ここで、リンクの定義から、Web ページのリンク価値を次のように定義する。ページのリンク価値とは、ページから出ている Aut リンクの数とこのページが他のページに張られている Hub リンクの数で特徴付けられるとする。実際、面白い、有用だと思われる Web ページは沢山のページにリンクされている。また、あるページが沢山のページへリンクしているのであれば、そのページが便利なページだと思われる。したがって、ページのリンク価値をページの価値の一部であると考える。

ページのリンク価値 V は次の式で表す。

$$V_i = \{Hub(L_h), Aut(L_a)\}$$

Hub（Ln）：Hub Link の数、Aut（La）：Authority Link の数

ページのリンク価値を前述のように定義すると、以下の仮説が立てられる。

- 仮説１：

  ネットワーク世界においての人間の価値観（趣味・嗜好を含む）は Web 世界のリンク構造によって伝播している。

- 仮説２：

  人間が自分の価値観（趣味・嗜好を含む）に合うリンクを Web ページに追加することは、ネットワーク（WEB）社会における知識の流通につながる。

2.4　人工的ネットワーク

動的に結合されたネットワークシステムは、生物生態系のモデルや伝染病の伝播、ニューラルネットワーク、映画俳優の共演関係、及び多くの自己組織システムなどの研究に使われている[1]。通常のネットワークモデルの場合、接続トポロジーは完全に規則的（レギュラー）か、あるいは完全にランダムであるかに大別される。しかし、多くの社会的、物理的なネットワークは完全なレギュラーでもなく、完全なランダムでもない、一種の中間的なものだと考えられる。

ここで、実際のネットワークを分析する際に、比較の対象として、実世界のネットワークをモデルにした人工的ネットワークを以下のように作る。また、次章で説明する Web ロボットとネットワーク分析ツールの動作にも人工的に生成したネットワークが用いられる。

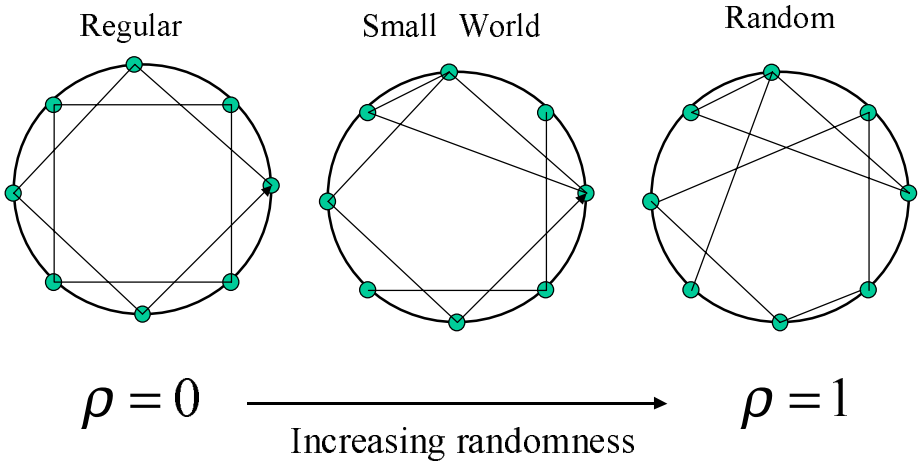


図 2.4.1　人工ネットワーク

人工的ネットワークを作る時に、二つのパラメータを用いる。リンクのランダム性を決定するパラメータとリンクの密度（以下リンクの結合率と呼ぶ）を決定するパラメータである。リンクのランダム性とは、あるノードから出ているリンクの方向を決めるパラメータ。図のように、ランダム性ρが０の時、ネットワーク内のリンクが規則正しく並んでいる、ρが１に近づくにつれ、ネットワーク内のリンクの並ぶ方がランダムになる。リンクの密度とは、ネットワーク内のリンク数とノード数の比例値のことであり、リンクの結合率とも呼ばれる。

人工的ネットワークの生成手順：

１．　ネットワーク内のノードの数を決める。
２．　ネットワーク内のリンクのランダム性（パラメータ１）を決める。
３．　ネットワーク内のリンクの結合率（パラメータ２）を決める。
４．　ネットワークを生成する。

二つのパラメータを調節するで、規則正しい（$\rho=0$）、無秩序（$\rho=1$）とさまざまな中間的な領域（$0<\rho<1$）のネットワークを作り出すことができる。こうして生成された人工的ネットワークは実際のネットワークの結合形態を説明するために使われる。

第　3　章
アルゴリズム

本章では、研究で使われる一連のツール（プログラム）の開発をする際に考案したアルゴリズム、個々のプログラム構成およびシステムとして動作するプログラムの全体の動きについて説明をする。

3.1　ネットワーク分析の手順

前章で述べた仮設を踏まえてここでは、リンクの価値に注目し、ネットワーク内のページおよびリンクの構造解析を利用したネットワーク分析手法を以下のように提案する。

手順：

ページ＝HTML ファイル及びページのリンクデータ（Aut リンク）を Web ロボットを用いて収集する。
得られた Aut リンクデータとページデータから、別に提案するアルゴリズム（第 3 章）を用いて、リンク構造を計算する。Hub リンクデータを得る。
Aut リンクデータと Hub リンクデータを使って、すべてのページの価値を計算する。
さらに、ネットワーク内のリンクの結合率やネットワークの開放度（後述）を計算することによって、ネットワークの特徴を数値化して、分析を行う。
計算結果のグラフを専用ツールに表示させる。ネットワークの特徴とネットワークに

おける情報の分布状況について、定性的に分析を行う。

また、人工的に作られたネットワークを実際のネットワークと比較をして、両者の相似性や、関連性についても検討を加える。

3.2 アルゴリズム

WWW ネットワークの分析をするには、まずネットワークの構造（ノードとパス）を知る必要がある。しかし、ネットワークにおけるある一定の領域内のすべてのノード（ページ）情報とパス（リンク）情報を収集することは人間にとって大変な作業である。ここで、機械的な作業を人間の代わりに Web ロボットにやらせる方法をとった。Web ロボットの動作を以下に示す（図 3.2.1）。

- Web ロボットの動作：

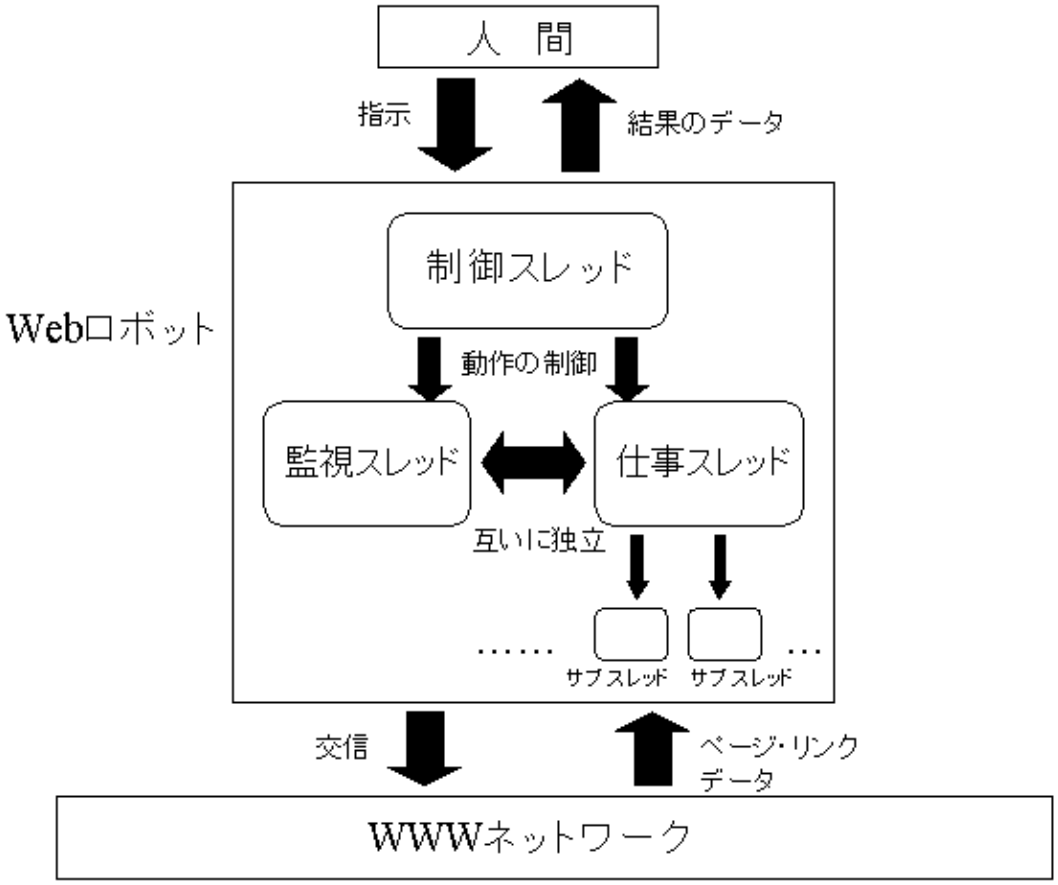
① Web ロボットは、最初のスタートポイントのドキュメント（HTML ファイル）を読み込み、ドキュメントにあるリンクデータを抽出し、自分のインデックスリストに記憶する。

② リストから順番にデータを取り出し、リンク先のドキュメントを読み込む。さらに読み込んだドキュメントを解析しながら、リンクデータを自分のインデックスリストに追加する。

③ 新しいドキュメントを閲覧するたびにインデックスに追加していき、問い合わせは再実行される。

④ Web ロボットはユーザーが設定した条件を満たすまで、或はタイムリミットまで②と③の処理を繰り返す。

こうして、Web ロボットはスタートポイントを中心にリンクを辿って、幅優先探索方式で次々へと探索の輪を広げていく。また、プログラムを設計する際に、ロボットの工作効率を上げるために、Java 言語のマルチスレッド機能を利用して、ネットワークの通信スピードやデータの数と量に応じて、ロボットが自律的に自分自身をコントロールできるようにした。ロボットは、通信スピードの許す限り、最大 10 個の代理（サブスレッド）を呼び出して、処理を分散させることができる。



- リンク情報の抽出

Web ロボットがネットワーク上でリンクデータとページデータ（URL）を収集する際に、リンク情報をデータベースに追加および次の行動を決定するために、Web ページ（HTML ファイル）からリンク情報を抽出する必要がある。

ページのリンク情報を抽出するアルゴリズム：

① リンクデータを格納する LinkStock を作る。

② ページを読込む。読込んだページを解析する。

③ 読込んだページからリンクを抽出し、LinkStock に入れる。

④ このページの解析が終わったら、LinkStock から Link をひとつ取り出す。

⑤ LinkStock が空なるまで、②～④を繰り返す。

- リンク構造解析

Web ロボットが集めてきたリンクデータを解析し、一定のネットワーク内のページとページの間の全体リンク構造を解明する。

リンク構造解析のアルゴリズム：

① 任意の StartNode（PageNumber 以下 SN と略す）を決める。

② 全リンクペア LIST から NS の Aut リンクを取り出し。第 1 階層（1st nearest neighbor）のリンク構造を判明する。

③ 第一階層のページを WorkingList に記憶させる。（重複ページを除く）

④ 全リンクペア LIST から WorkingList に入っているページの Aut リンクを順番に探し出し、第 2 階層（2nd nearest neighbor）のリンク構造を判明する。第 2 階層のページを WorkingList に記憶させる。（重複ページを除く）

⑤ WorkingList が空なるまで、②～④を繰り返す。

- Hub リンクの判明と計算

複数の WEB ページのリンク構造が分かれば、どのページがどのリンクを持っているか、あるいはどのページがどこのページにリンクされているかのが分かる。しかし、各ページに対して、Hub　Link は Authority Link とは違って、そのページの HTML ファイルを解析してもそれを知ることは出来ない。目に見えない Hub リンクの構造を知るため、以下のようなアルゴリズムを考えた（図 3.2.2）。ページの集合を P、L とし、Authority Link 集合を A = {(a , d) , (b , e) , (c , e) , (c , f) , (d , b)}、Hub Link 集合を H = {(a , d) , (b , e) , (c , e) , (c , f) , (d , b)} とする。

Hub リンク計算のアルゴリズム：

① すべてのページの Authority　Link 構造を解析して、Authority Link 集合 A を作成する。

② 任意の２つのページ a , d に対して、a の Authority Link (a , d)は、d にとっ

て Hub Link ともなる。ある Authority Link ペアから Hub Link ペアをつくる。

③ すべての Authority Link ペアがなくなるまで、②を繰り返す。

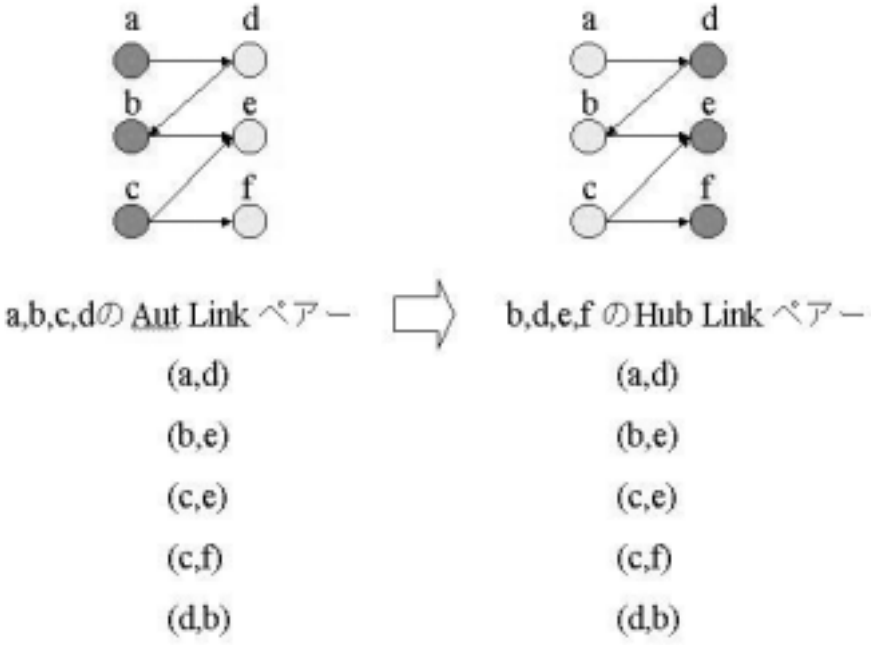


図 3.2.2　Hub リンクペアの作成

第　4　章
実　験

4.1　実験の全体の流れ

図 4.1.1 は、実験の全体の流れを表している。

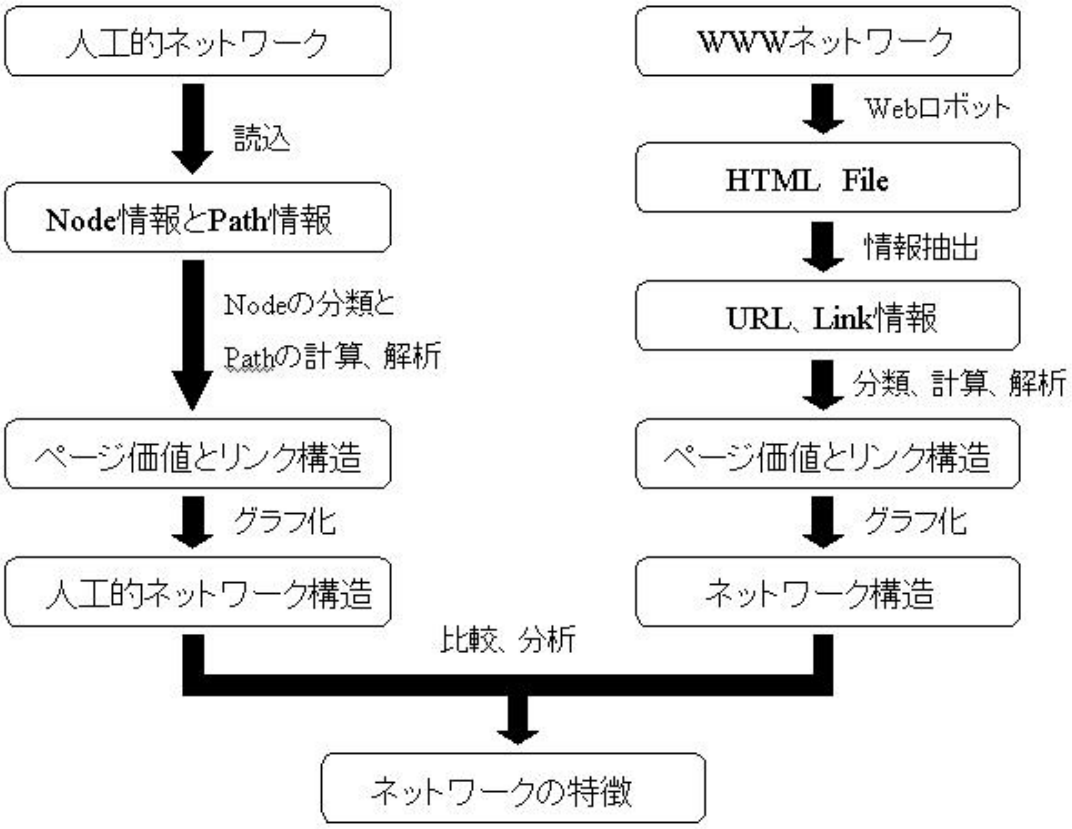


図 4.1.1　実験の流れ

4.2　実験の準備

4.2.1　実験機材

ソフトウエア：

実験のために作成した分析システムツール。Java 言語で書いたプログラム（10 個）と補助プログラムからなる。

ハードウエア：

日立 FLORA370（PentiumⅡ 400）：主に計算能力を要求しないデータの収集、グラフの表示などに使う。

DELL（PentiumⅡ Xeon 400 * 2）：演算スピードと膨大なメモリを必要とするリンクの構造解析、ページの価値計算などに使う。CPU が 2 枚積んでいるので、特にマルチスレッド方式のプログラムの実行に力を発揮する。

ネットワークへの接続形式：

LAN

4.2.2 条件の設定

実験は人工的ネットワークと実際の WWW ネットワークを対象に、複数回に分けて、行われるものである。それぞれの条件設定を以下で示す。

1. 人工的ネットワーク：
   - 実験回数：15 回
   - ノード数：1000
   - リンクのランダム性：０、0.5、１の３段階
   - リンクの結合率：５、15、25、35、100 の５段階
   - 一回実験の所要平均時間：６時間

2. 実際の WWW ネットワーク：
   - 実験回数：12 回
   - ノード数：不定
   - スタートポイント（Web ロボットの探索がはじめる時のページ）：12 ページ
   - 一回実験の所要平均時間：20 時間

4.3　実験

今回の実験を通じて以下のようなデータを得られた。データ（一部）を以下に示す。他のデータは付録 7（実験データの部分）に掲示する。

4.3.1　人工的ネットワーク

表 4.3.1 と 4.3.2 は、12 回の人工的ネットワークを使った実験の結果をまとめたものである。表の第一行目は実験の回数を表している。

実験S:	1	2	3	4	5	6	7
Radom	0	0	0	0	0	0.5	0.5
Link結合率	5	15	25	35	100	5	15
Pages	1000	1000	1000	1000	1000	1000	1000
A-Links	5254	15090	26369	35158	50000	5649	15136
最大A-Links	99	100	100	100	50	100	100
最大H-Links	10	23	36	43	50	14	27
最少A-Links	1	1	1	1	50	1	1
最少H-Links	1	7	18	25	50	0	6
A-Links結合率	5.254	15.09	26.369	35.158	50	5.649	15.136
H-Links	5254	15090	26369	35158	50000	5649	15136
H-Links結合率	5.254	15.09	26.369	35.158	50	5.649	15.136
H-Links内接率	100%	100%	100%	100%	100%	100%	100%

表 4.3.1　人工ネットワーク（1－7）

8	9	10	11	12	10	11	12	平均
0.5	0.5	0.5	1	1	1	1	1	
25	35	100	5	15	25	35	100	
1000	1000	1000	1000	1000	1000	1000	1000	1000
25929	33914	50000	5119	15803	24804	35119	50000	28885.92
100	100	50	97	100	100	100	50	87.25
45	53	57	12	32	43	63	71	42.5
1	1	50	1	1	1	1	50	9.909091
12	18	43	0	4	11	20	28	18.08333
25.929	33.914	50	5.119	15.803	24.804	35.119	50	28.88592
25929	33914	50000	5119	15803	24804	35119	50000	28885.92
25.929	33.914	50	5.119	15.803	24.804	35.119	50	28.88592
100%	100%	100%	100%	100%	100%	100%	100%	100%

4.3.2　人工ネットワーク（8－12）

Link 結合率：　リンクの密度を表す量。
A-Links：　Authority Link の総数。
H-Links：　Hub Link の総数。
最大 A-Link：　最大 A-Link 数を持つページの A－Link 数。
最大 H-Link：　最大 H-Link 数を持つページの A－Link 数。
最少 A-Link：　最小 A-Link 数を持つページの A－Link 数。
最少 H-Link：　最小 H-Link 数を持つページの A－Link 数。
A-Link 結合率：　ネットワークの平均 Aut リンク数。
H-Link 結合率：　ネットワークの平均 Hub リンク数。
H-Link 内接率：　ネットワークの開放度を表す。

図 4.3.1～4.3.4 は、人工的ネットワーク分析の結果をグラフ化したものである。X 軸はネットワーク距離を、Y 軸はリンクの数を表している。また、濃い黒線は Aut リンクを、薄い黒線は Hub リンクのことを表す。(図 4.3.1：ランダム性 R=0、結合率 K=0.15、4.3.2： R=100、K=0.35、4.3.3：R=100、K=0.05、4.3.2： R=100、K=0.15)

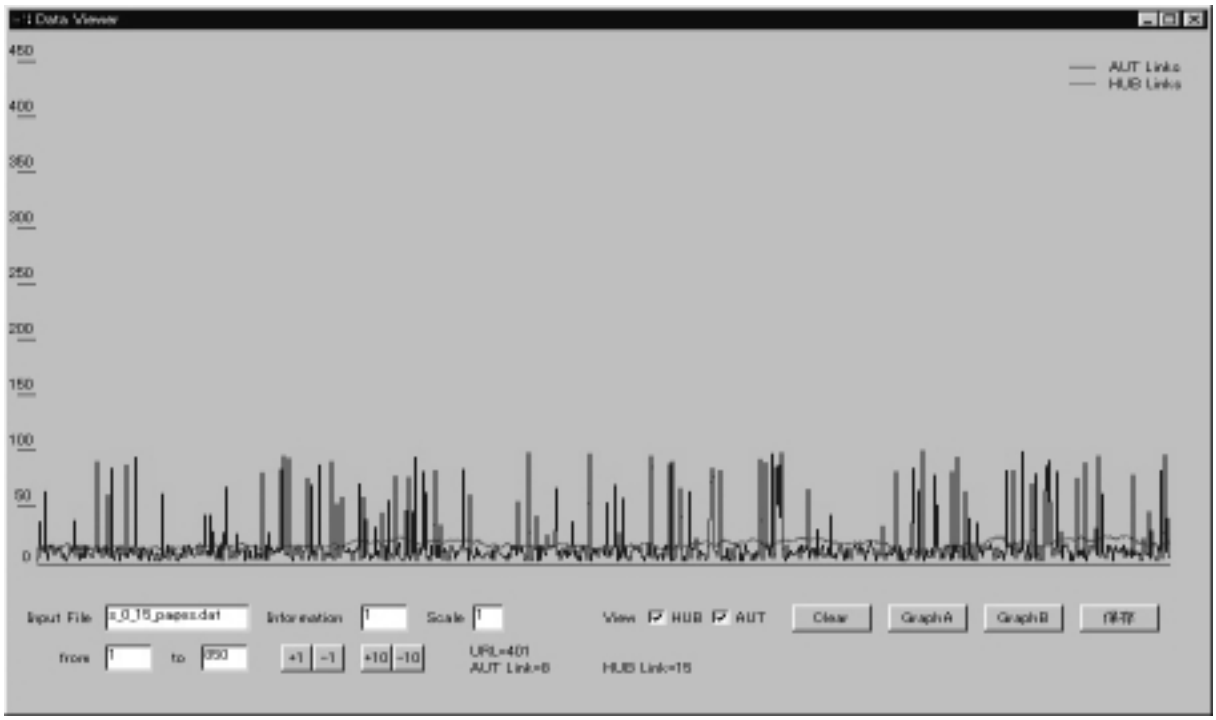


図 4.3.1　人工ネットワークグラフ 1

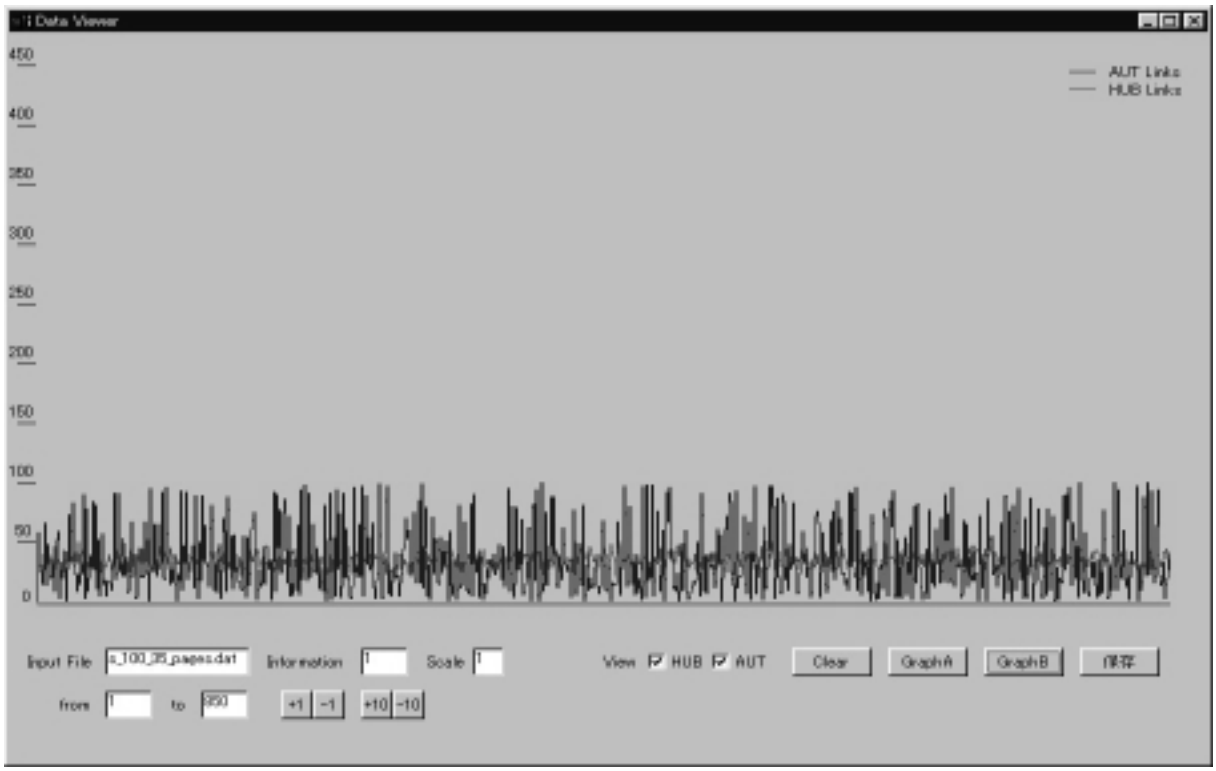


図 4.3.2　人工ネットワークグラフ 2

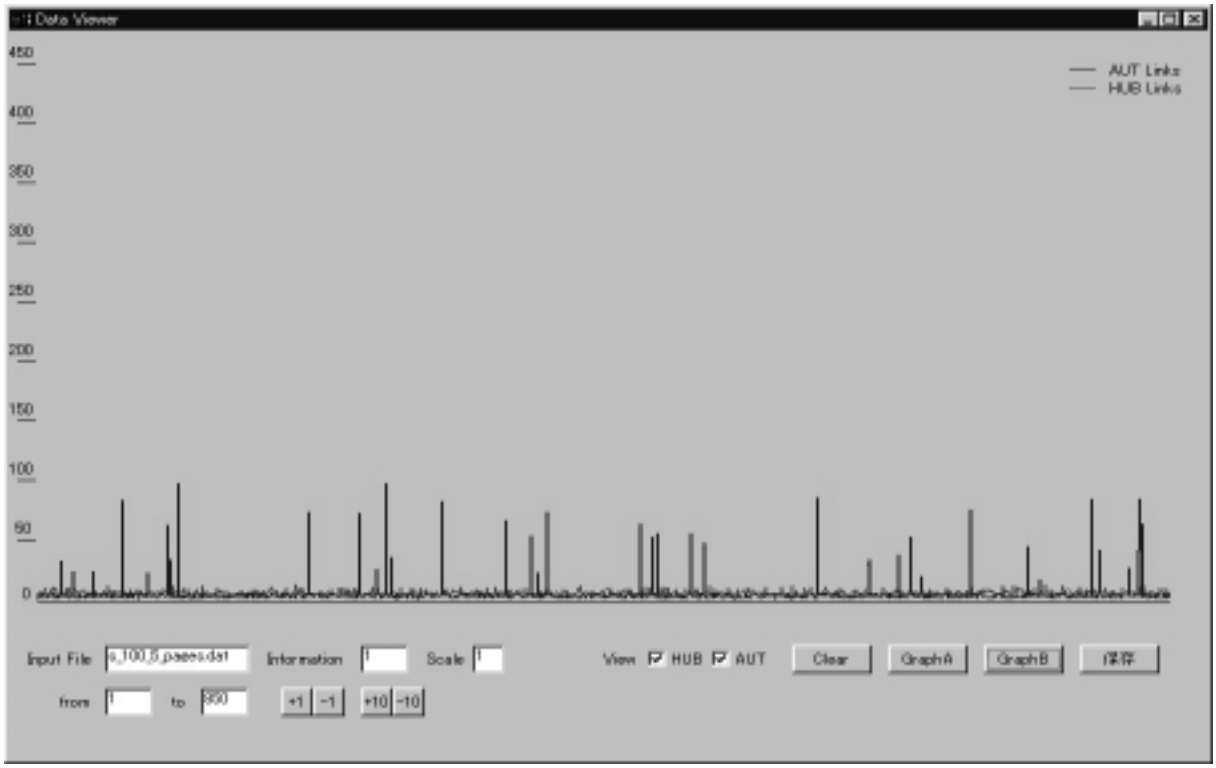


図 4.3.3　人工ネットワークグラフ 3

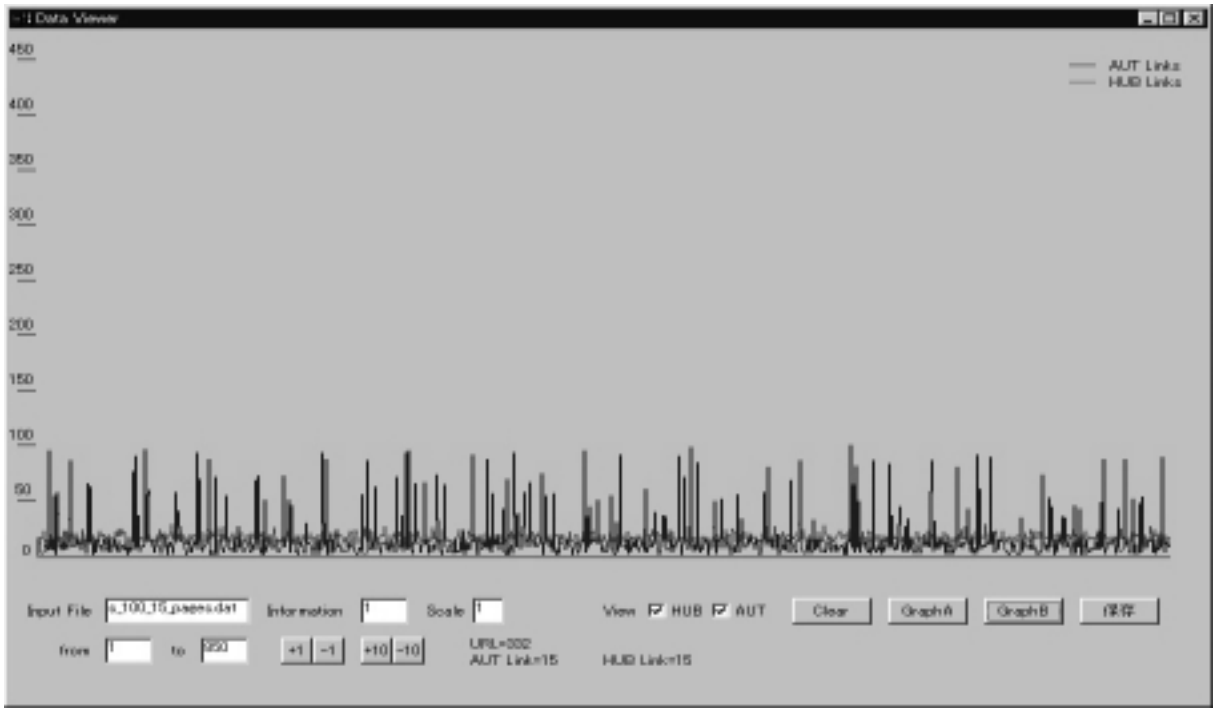


図 4.3.4　人工ネットワークグラフ 4

4.3.2　実際の WWW ネットワーク

WWW ネットワークの実験結果をまとめ、ネットワークのグラフ（一部）を以下で示す。

実験:	1	2	3	4	5	6	7	8	9	10	11	12	
Start Poin	Hp	Huang	Jaist	Java	Ks	Test	Yy	Tkd	Huang *	Jaist *	Shino	Sut	平均
Pages	819	1460	3544	860	4514	2949	609	1161	3384	4349	3790	1416	2404.583
A-Links	14159	7800	21536	21387	27886	27416	13441	17119	19799	82335	24504	36091	26121.08
最大A-Lin	721	161	203	160	197	936	1683	405	829	417	1115	1731	713.1667
最大H-Lin	157	486	1036	215	1258	93	78	141	658	552	1126	188	497.3333
最少A-Lin	1	1	1	1	1	1	1	1	1	1	1	1	1
最少H-Lin	0	0	0	0	0	0	0	0	0	0	0	0	0
A-Links結合率	17.29	5.34	6.08	24.87	6.17	9.3	22.07	14.75	5.85	18.93	6.47	25.49	13.55083
H-Links	3404	3495	9964	6544	12978	6857	2474	3586	7592	24641	10968	8234	8394.75
H-Links結合率	4.16	2.39	2.81	7.61	2.88	2.33	4.06	3.09	2.24	5.67	2.89	5.81	3.828333
H-Links内接率	0.24	0.45	0.46	0.31	0.47	0.25	0.18	0.21	0.38	0.3	0.45	0.23	0.3275
										* :With Other com			

表 4.3.3　実際のネットワーク（WWW）

Link 結合率：　リンクの密度を表す量。
A-Links：　Authority Link の総数。
H-Links：　Hub Link の総数。
最大 A-Link：　最大 A-Link 数を持つページの A－Link 数。
最大 H-Link：　最大 H-Link 数を持つページの A－Link 数。
最少 A-Link：　最小 A-Link 数を持つページの A－Link 数。
最少 H-Link：　最小 H-Link 数を持つページの A－Link 数。
A-Link 結合率：　ネットワークの平均 Aut リンク数。
H-Link 結合率：　ネットワークの平均 Hub リンク数。
H-Link 内接率：　ネットワークの開放度を表す。

図 4.3.5～4.3.8 は、実際のネットワーク分析の結果をグラフ化したものである。X 軸はネットワーク距離を、Y 軸はリンクの数を表している。また、深い黒線は Aut リンクを、浅い黒線は Hub リンクのことを表す。（図 4.3.5：StartPoint=huang、4.3.6：SP=java、4.3.7：SP=test、4.3.8：SP=tkd）

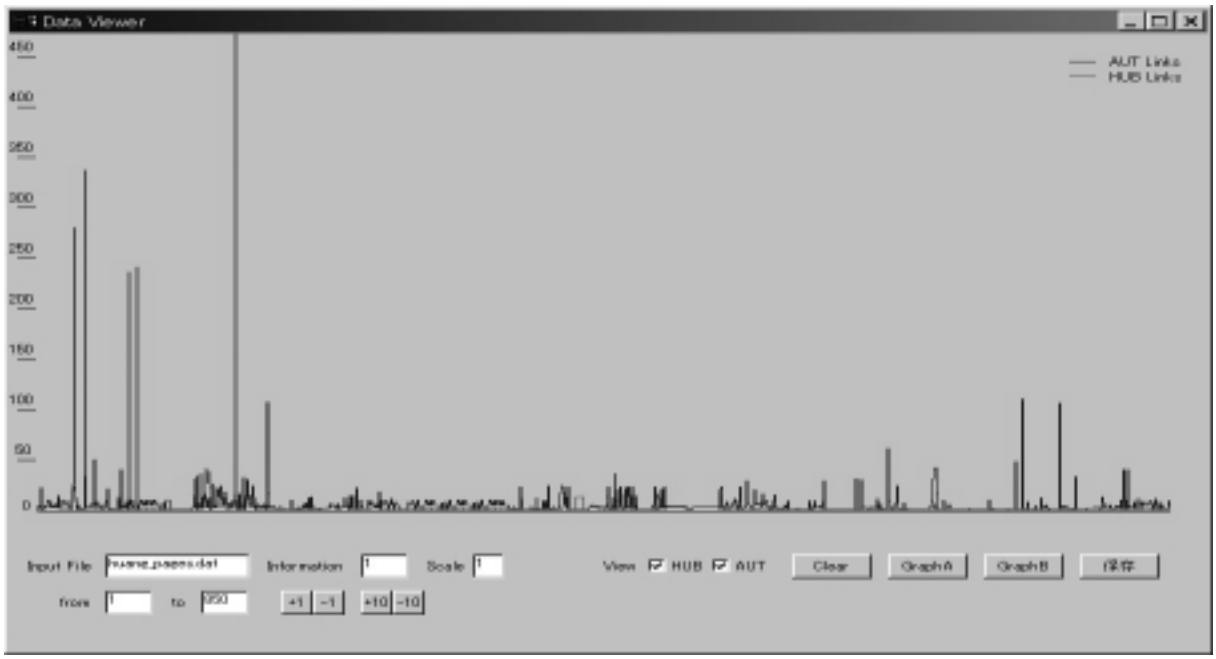


図 4.3.5　ネットワーク（WWW）グラフ 1

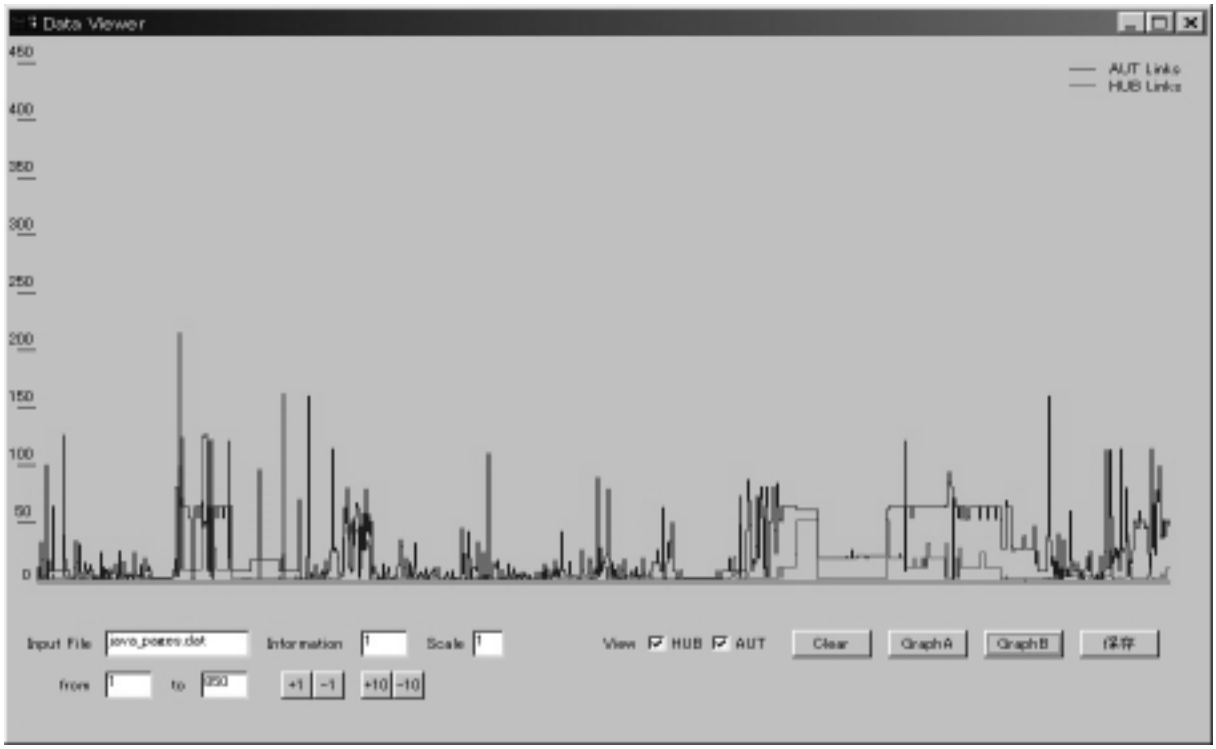


図 4.3.6　ネットワーク（WWW）グラフ 2

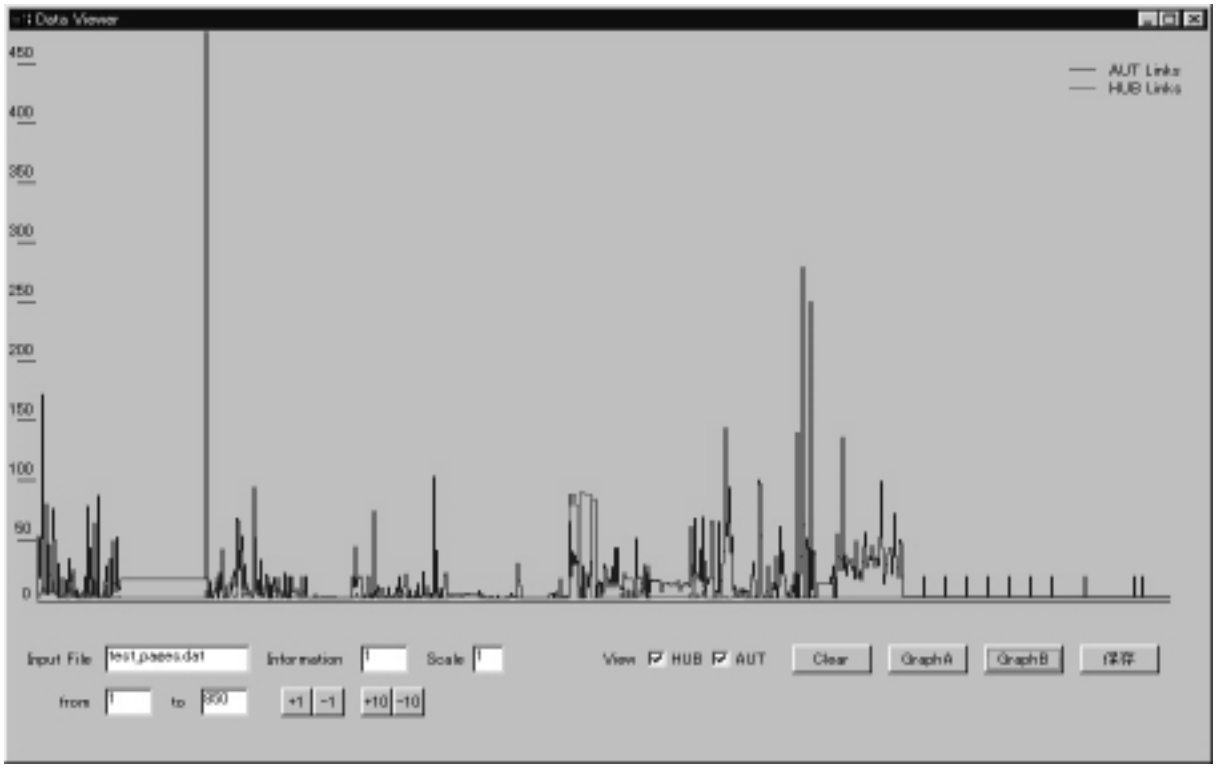


図 4.3.7　ネットワーク（WWW）グラフ 3

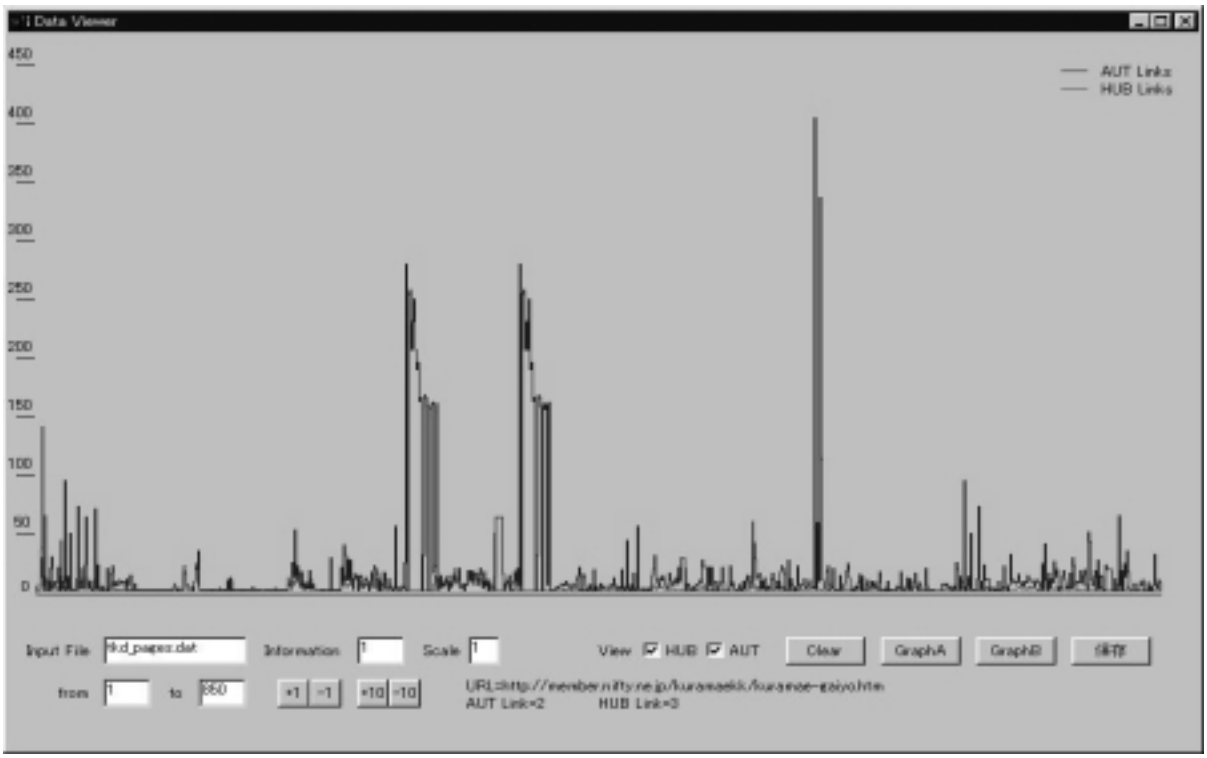


図 4.3.8　ネットワーク（WWW）グラフ 4

4.4 結果

ここで、今回の実験でみられたネットワークの特徴を以下に示す。

- 個々のページにおいて、Aut リンク価値と Hub リンクの価値は同一ではない。

個々のページに対して、そのページの Aut リンク（あるいは Hub リンク）価値が幾ら高くても、そのページの Hub リンク（あるいは Aut リンク）価値も同時に高いとは限らない。

図 4.4.1－4.4.3 のようなネットワーク(huang)の場合、Start-Point は huang で、グラフの X 軸はネットワーク距離を、Y 軸はリンクの価値を表している。図 4.4.1 は Aut リンク価値と Hub リンク価値を一枚のグラフで表し、図 4.4.2 と 4.4.3 はそれぞれに Aut リンク価値と Hub リンク価値だけを表している。また、深い黒線は Aut リンクの価値を、浅い黒線は Hub リンクの価値を表している。このグラフでは、Hub リンクのピークと Aut リンクのピークがそれぞれに、はっきりに分かれている。Hub リンク価値の高いページは、その Aut リンクの価値が低く、Hub リンク価値の低いページは、その Aut リンクの価値が逆に高いという現象が多く見られている。（表 4.4.1）

ページ	Hub 価値最大	Aut 価値最大	その他 Page1	その他 Page2
Aut リンク数	16	829	337	208
Hub リンク数	658	2	33	1

表 4.4.1：図 4.4.1 における Aut リンク価値と Hub リンク価値との比較

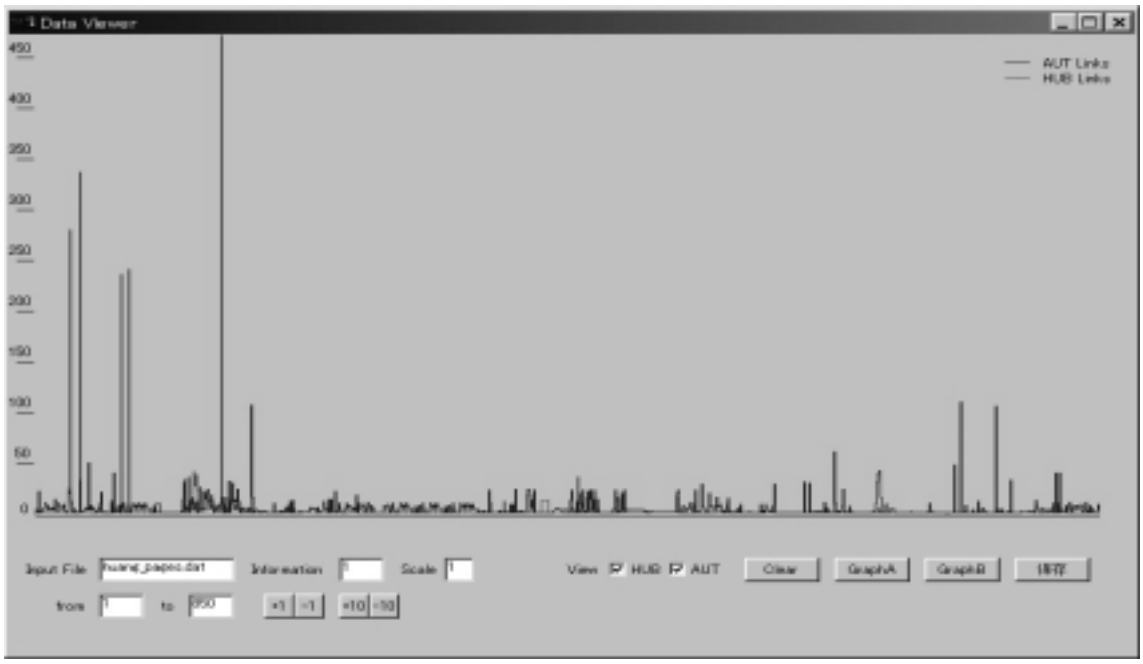


図 4.4.1　リンク・ネットワーク構造グラフ（huang） 1

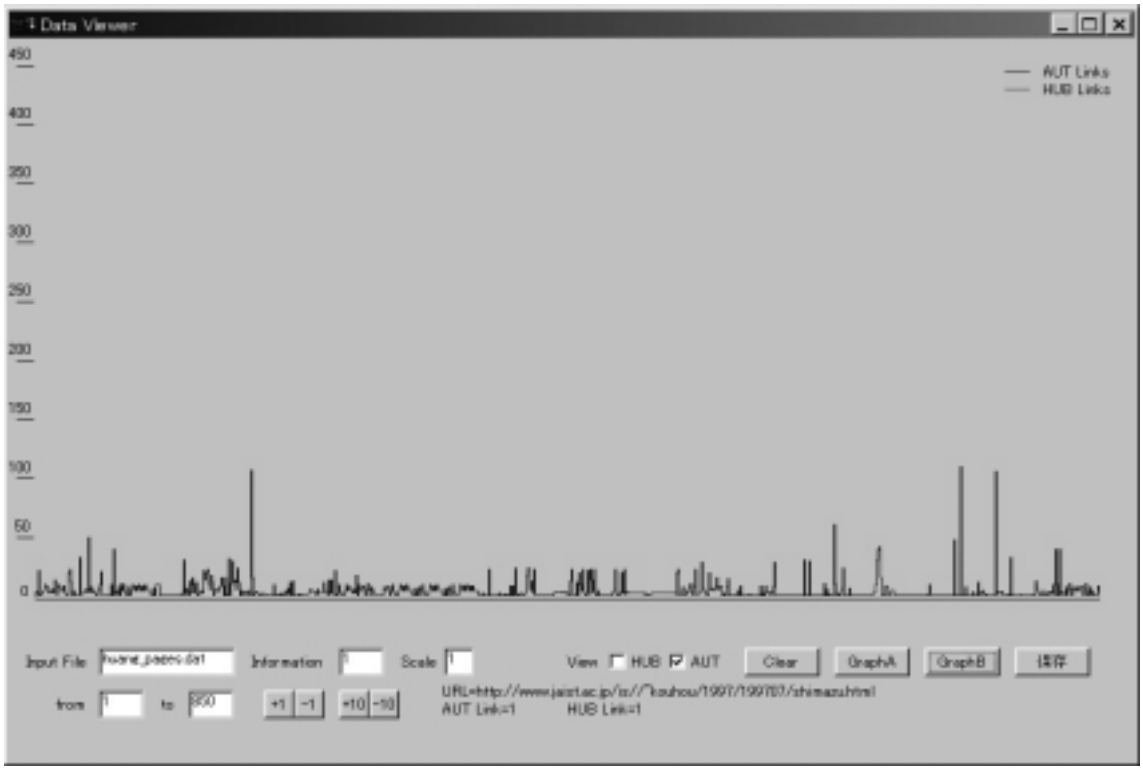


図 4.4.2　リンク・ネットワーク構造グラフ（huang） 2

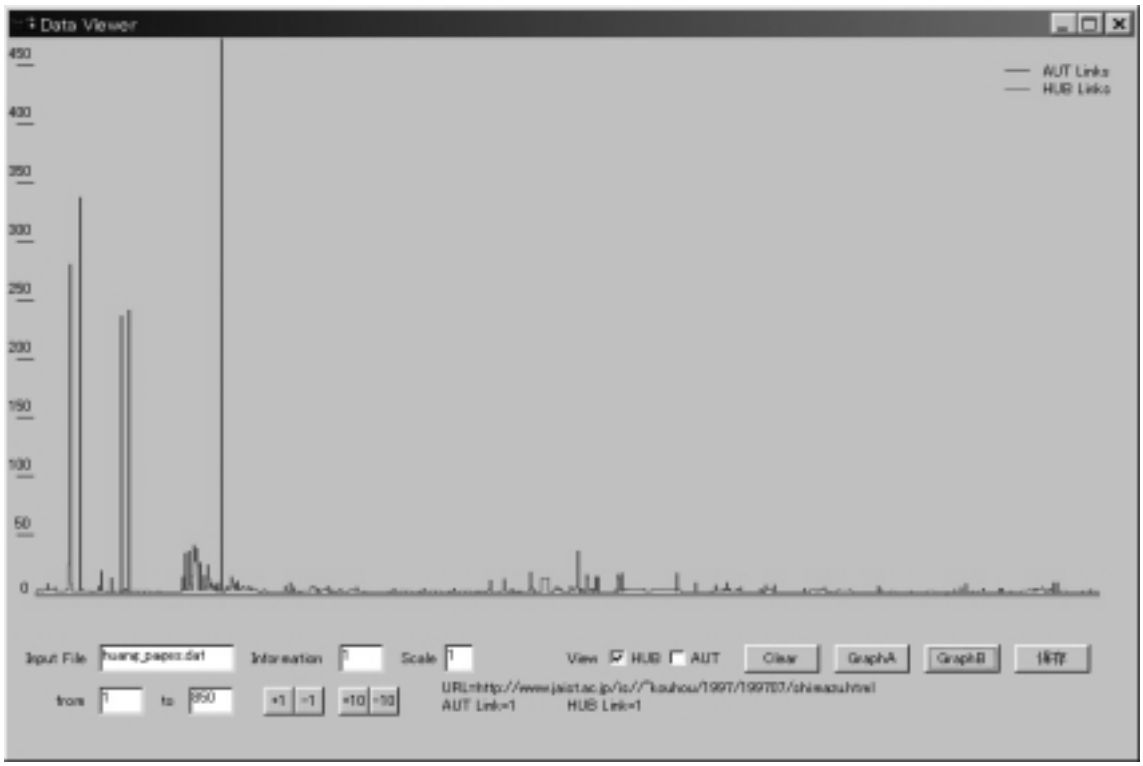


図 4.4.3　リンク・ネットワーク構造グラフ（huang） 3

- Aut リンク価値の高いページは、その Hub 価値も相対的に高い。

前述のように、各ページにおいては、Aut リンク価値と Hub リンク価値は同じではないことが分かった。しかし、ネットワーク全体を見る時に、Hub 価値と Aut リンクに相関傾向がある程度みられる。いくつかの例外を除いて、Aut 価値でソートしたグラフを見ると、多くの場合、Hub リンク価値の高いページが Aut リンク価値の高い区域に相対的に集中する傾向がある。それは、WEB ページのリンク価値が WEB ページの価値ともなることを示していると考えられる。

図 4.4.4、4.4.5、4.4.6 はそれぞれに Aut リンク価値でソートしたグラフである。X 軸はリンク価値の順番を表し、Y 軸はリンク価値を表している。

いずれのグラフでも、グラフの前半（Aut リンク価値の高い（グラフの左側）区域）に Hub リンク価値の高いページが集中する傾向が見られる。

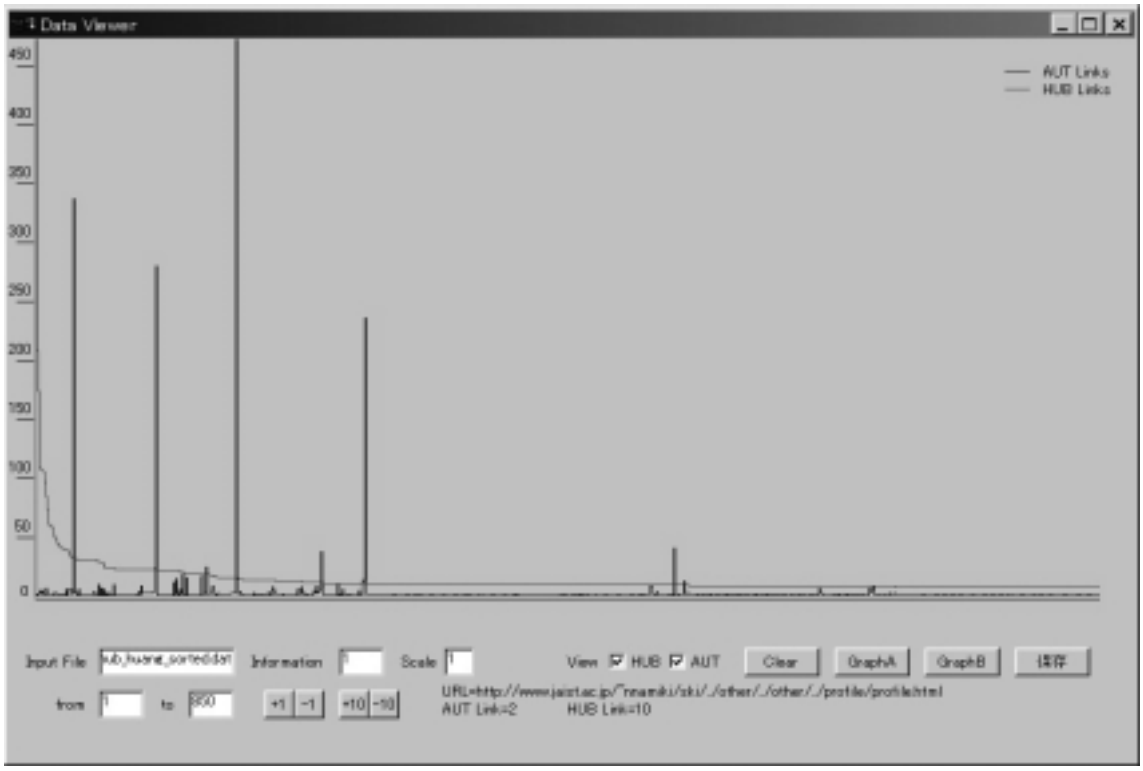


図 4.4.4　SP：Huang

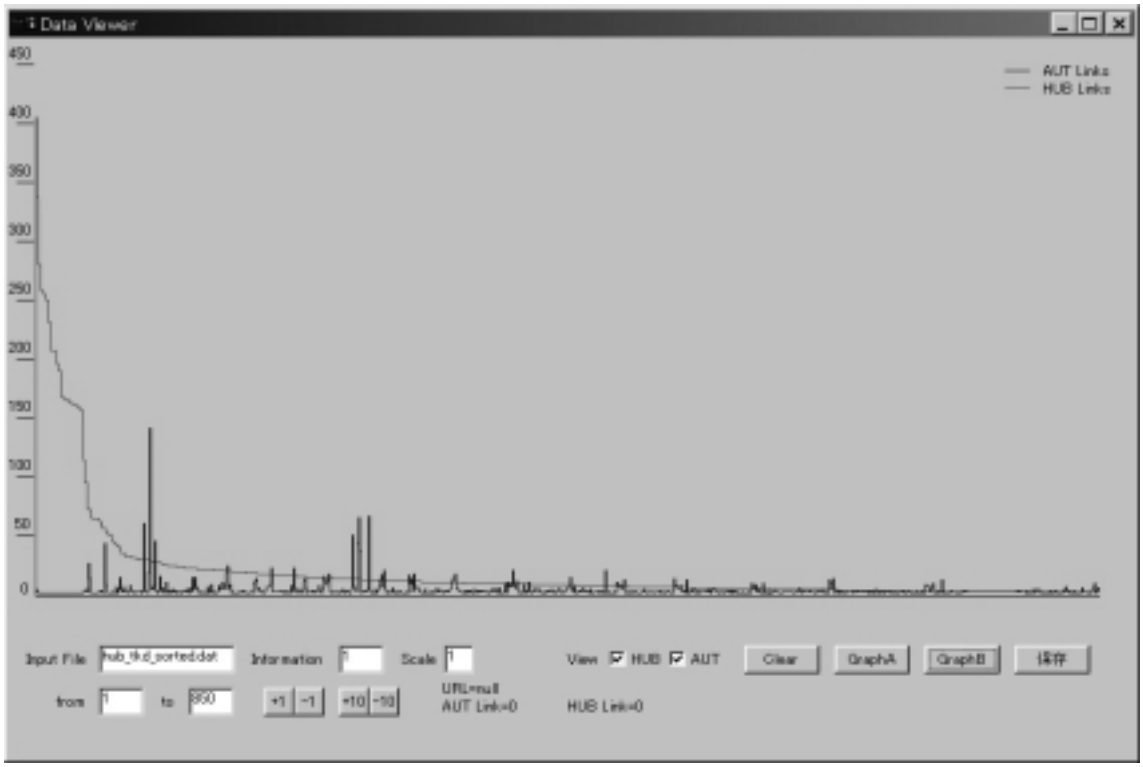


図 4.4.5　SP：東工大

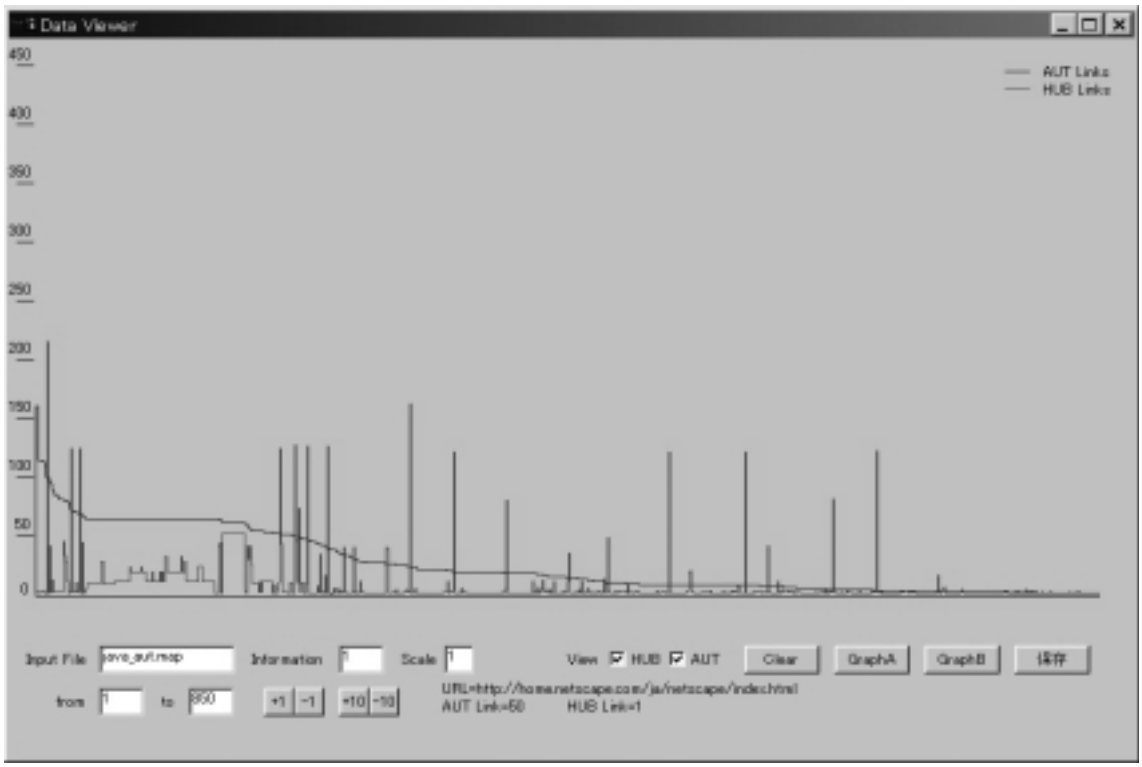


図 4.4.6　SP：　Java 言語のページ

- リンク価値分布について

各ネットワーク区域において、その区域の平均リンク価値（Aut リンク、Hub リンク両方含む）は大きく違い、分布パターンも異なっていることがわかった。図 4.4.7 と 4.4.8 の場合、前者の平均リンク価値は後者より明らかに高いことが分かる。それは、ネットワークにおける知識と情報の分布はランダムではなく、人の嗜好や趣味など主観的要素と集団や組織の構造など物理的要素と非常に関係していることを示していると考えられる。

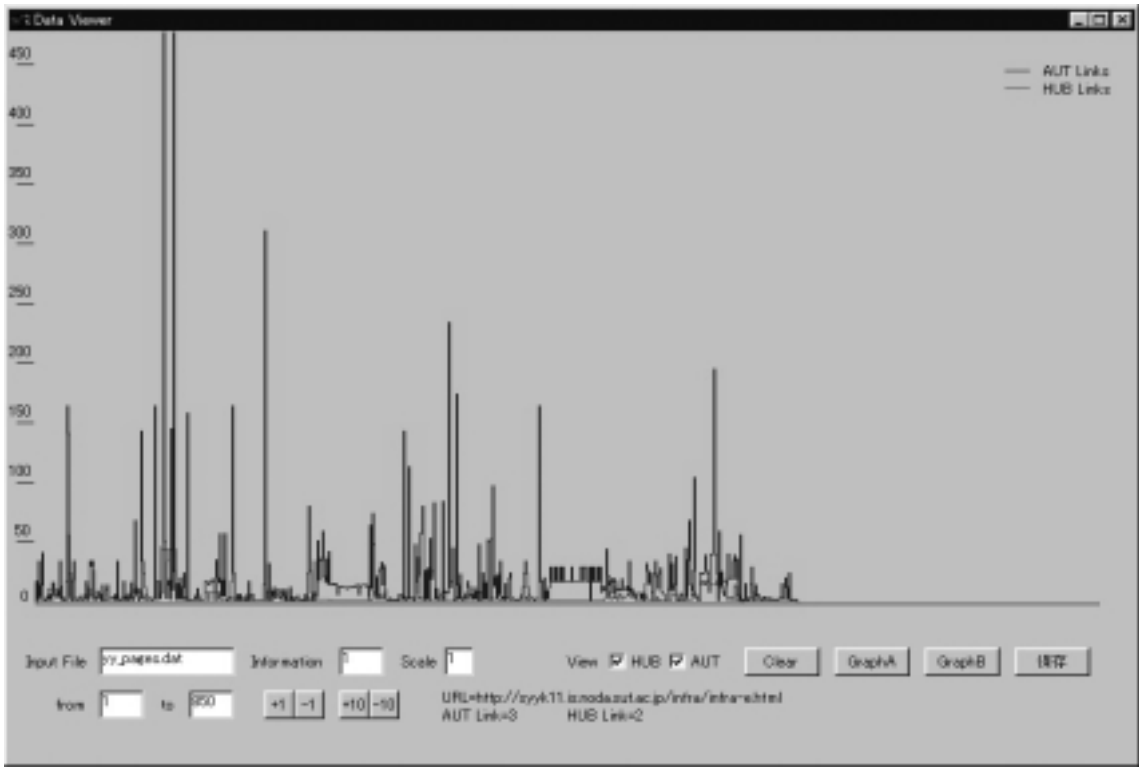


図 4.4.7　平均リンク価値の高いネットワーク

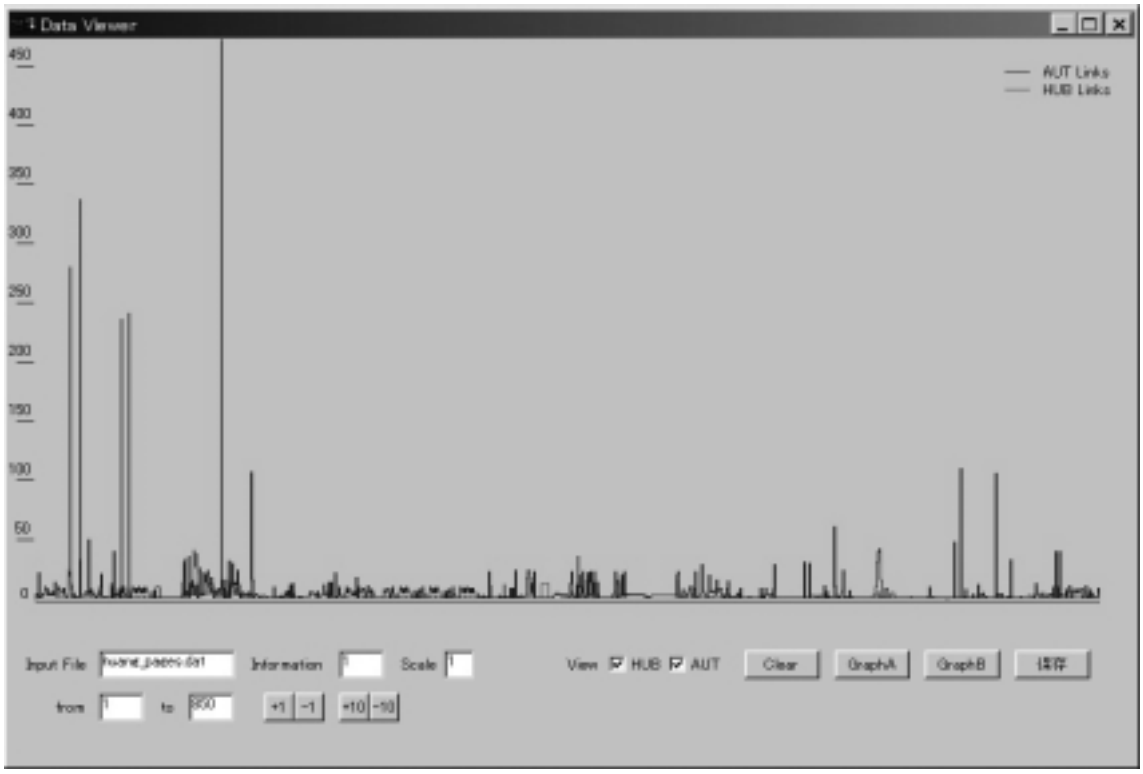


図 4.4.8　平均リンク価値の低いネットワーク

- 明確な目的を持って作ったページはそうでないページよりページのリンク価値が高い。

図 4.4.9 はある人の個人サイトをスタートポイントとしたネットワークグラフで、4.4.10 は Java 言語という共通な話題を持つサイトをスタートポイントとしたネットワークグラフである。両者を比較してみると、前者のリンク（Hub リンクと Aut リンク）価値の平均値が全般的に低く、それに対して、後者の Aut リンク価値と Hub 価値が両方とも高く、リンク価値の平均値も前者より何倍も高いことが分かった。明らかに、後者が、知識と情報の流通を活発に行われると思われる。このことは、前述の仮説のひとつ、つまり、“ネットワーク社会における知識や情報の流通と伝播はネットワーク及びリンクによって行われている”の根拠ともなると考えられる。表 4.4.2 の場合、二つのネットワークの平均リンク価値はおよそ 3 倍も違っている。

ネットワーク	平均 Hub リンク価値	平均 Aut リンク価値
Shino（個人サイト）	2.89	6.47
Java（Java 言語サイト）	7.61	24.87

表 4.4.2　Aut リンクと Hub リンク平均価値の比較

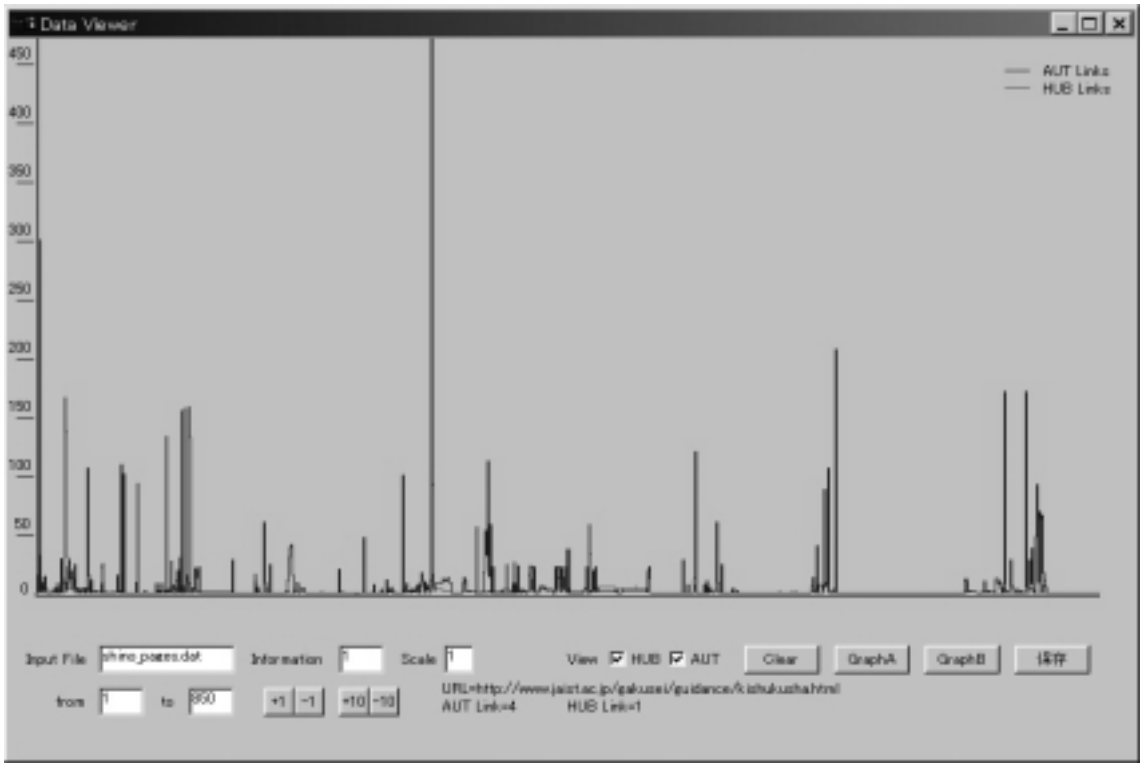


図 4.4.9　SP : shino

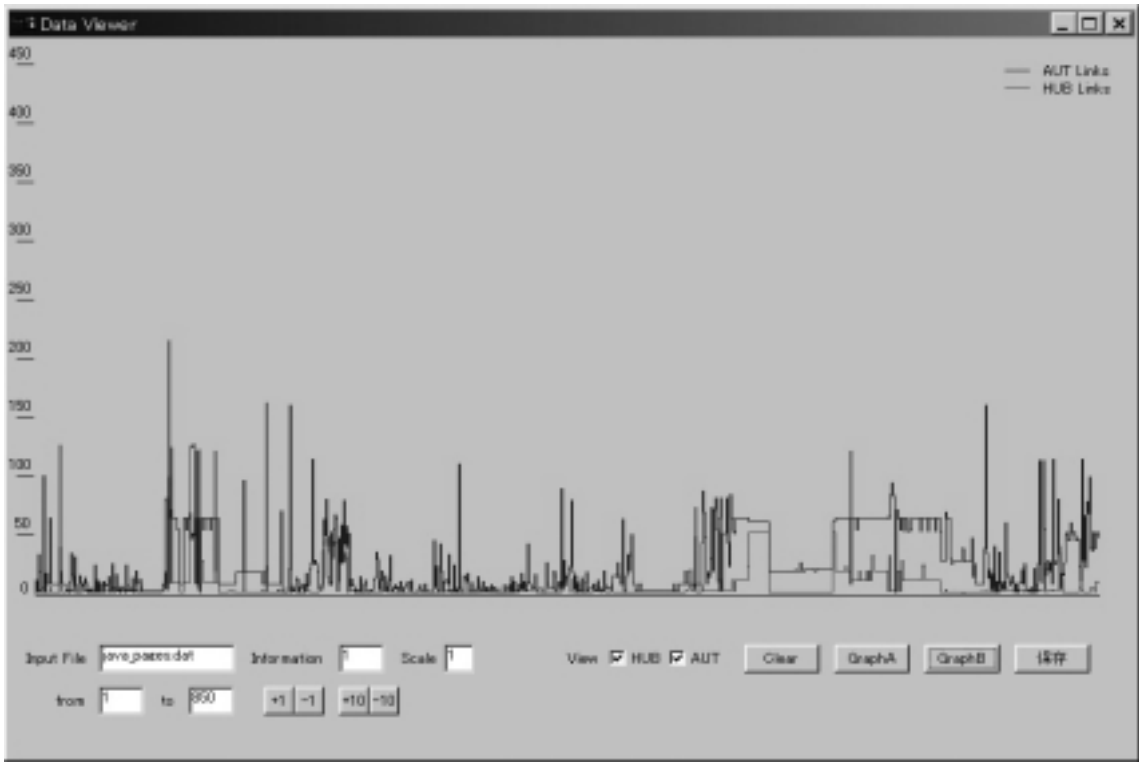


図 4.4.10　SP : Java

第　5　章
考察と課題

5.1　考察

● ネットワークの開放度と開放型ネットワーク

ネットワーク開放度の定義：

ネットワーク開放度とは、あるネットワークにおいて、そのネットワークの中にあるノードから出ているリンクの中に、外部ネットワークのノードへのリンクの割合のことである。（図 5.1.1）

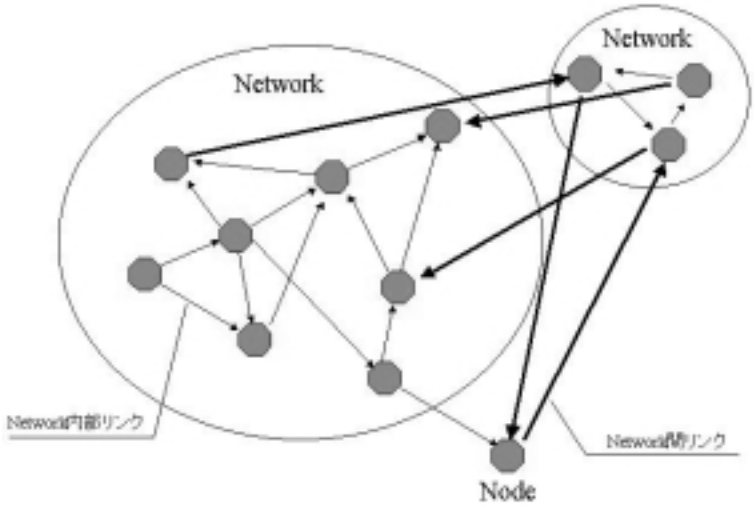


図 5.1.1

あるネット a のネットワーク開放度 Ka を以下の式で定義する。

$$K_a = \frac{\sum N[A(u)]}{\sum N[A(s)]} = \frac{\sum Aut}{\sum Hub}$$

ここで、N はノード、u はネットワーク a の中にあるノードを指向しているリンクで、s はネットワーク a の外にあるノードを指向しているリンクで、A(u)と A(s)はそれぞれリンクの数を表している。

ネットワークの開放度が高ければ高いほど、外の世界（ネットワーク）との連結経路が増え、ネットワーク距離も近くなり、情報や知識の交流も行いやすいと考えられるから、ネットワークの開放度はネットワーク構造を評価する時に一つ重要な参考値となる。

Network	Hp	Huang	Jaist	Java	Ks	Test	Yy	Tkd	SUT
開放度	0.76	0.56	0.54	0.69	0.53	0.75	0.82	0.79	0.77

表 5.1.1：ネットワーク別の開放度の比較

実験に用いたケースで、開放度を計算するとネットワーク YY や TKD、SUT、HP などの値が高く、これらのネットワークは開放的ネットワークであることを示唆(しさ)される。

- 人工ネットワークとの比較

二つのパラメータ（リンクのランダム性とリンクの結合率）を変えて生成されたさまざまな人工ネットワークと実際のネットワークとのグラフを比較する。その結果、図に示すように、リンク価値の分布パターンという点から、実際のネットワークと非常に似た人工ネットワークが存在していることが分かった。（図 5.1.2、図 5.1.3）

つまり、この二つのパラメータ（ランダム R と結合率 L）が実際のネットワークの形態を決定する要素との間に関連があると考えられる。しかし、具体的にどんな関連があるかについては、今後の課題として残されている。

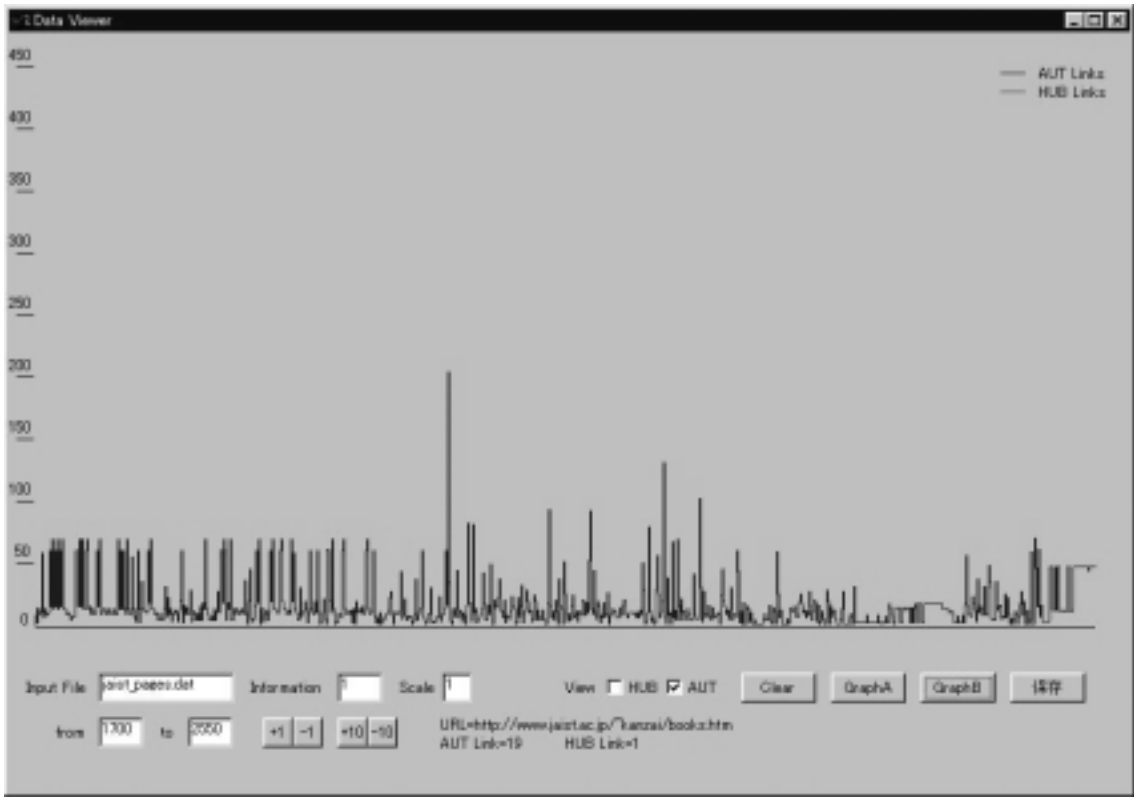


図 5.1.2　実際のネットワーク（SP：Jaist）

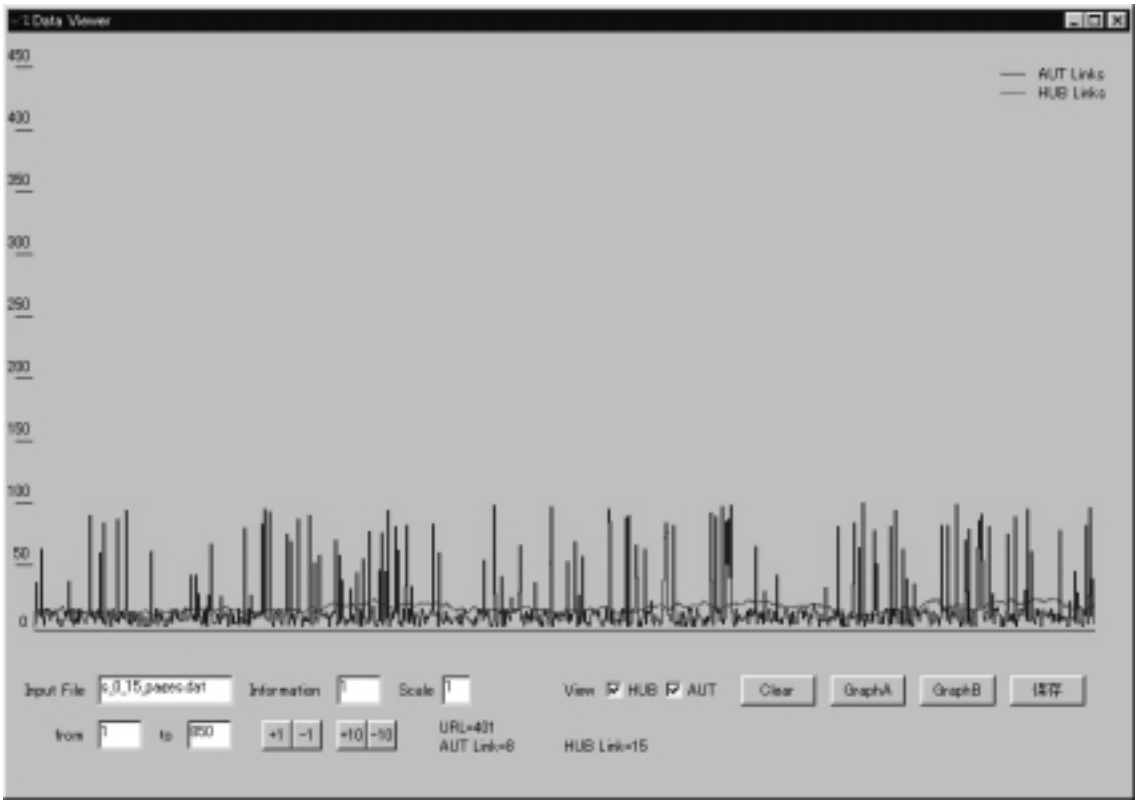


図 5.1.3　人工ネットワーク

● リンク価値の高いページの分布状況

実験では、リンクのランダム性とリンク結合率の二つのパラメータを操作することで、人工ネットワークを作った。前述のように、リンク結合率はネットワーク内のリンクの密度を決めるパラメータで、ネットワークの形態を決定する重要な数値である。今回は、リンクの結合率において、一様分布をとった。（図 5.1.4）

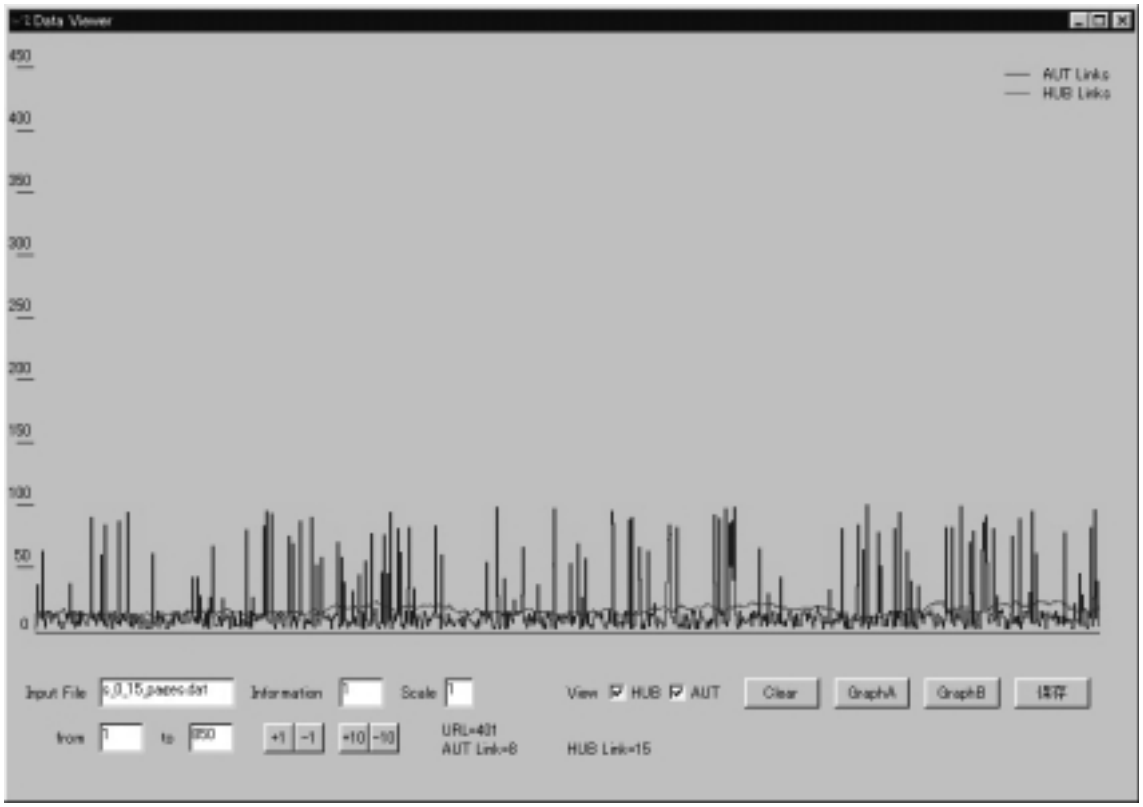


図 5.1.4　人工ネットワークにおけるリンクの結合率（一様分布）

しかし、実際のネットワークを分析したところ、ページ Hub リンク価値が極端に高いページがあることが分かった。それらのページ数はネットワーク内ページの総数の 0.1％にも満たないが、その Hub リンク価値がネットワークの平均 Hub リンク価値よりはるかに高い。（表 5.1.2）

Start Point	平均 Hub 価値	最大 Hub 価値	倍率（最大/平均）
HP	17.29	721	41.7
YY	22.07	1683	76.26
TEST	9.3	936	100.65
HUANG	5.85	829	141.71

表 5.1.2　リンク価値の比較

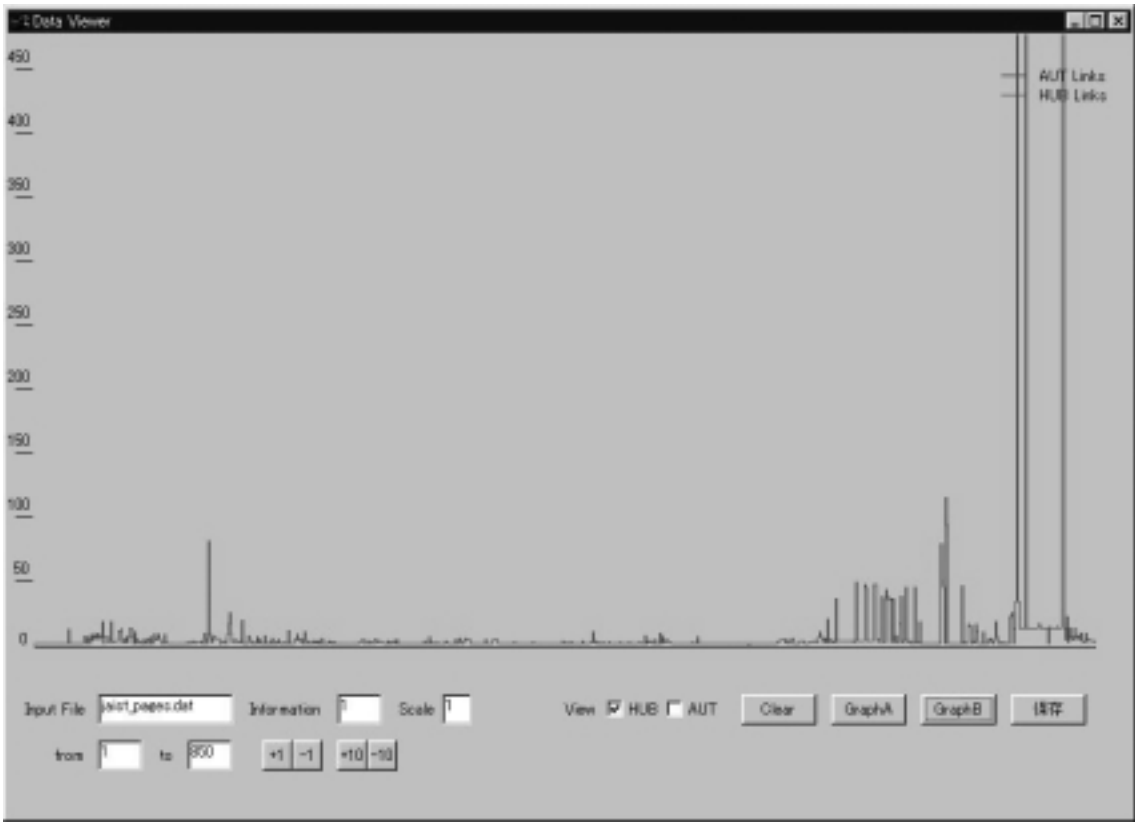


図 5.1.5　Jaist ネットワークにおける Hub リンク価値の分布様子

図 5.1.5 のように、あるページに複数のページからのリンクが集中している現象はリンクの結合率を一様分布で決める人工ネットワークには発生していない。この現象はポアソン分布に相当すると考えられる。ポアソン分布は多くの観察回数のうちほんの

少数の場合しか発生しないような出来事の場合によく当てはまる分布である。故に、Hub リンクが集中するネットワークに近い人工ネットワークを作成する際に、リンクの結合率をポアソン分布で決めるような手法が考えられる。これは今後の課題の一つとして、研究を続けていく予定である。

● 実際への応用

ここでは、本研究で提案したネットワーク分析手法とそれに基づいて、開発されたツールの実際への応用について検討する。

(1) ページ価値の計算：

リンク価値計算により、各ページの価値を数量化することができるという点からページのリンク価値はページの分類をする時に一つの参考値となる。

(2) ページ・ユーザーのグループ化：

趣味、嗜好でネットワーク内のページ・ユーザーをリンク構造によってグループ化することによって、新しいコミュニティの形成を支援する。

(3) 検索結果のランキング：

ページ価値とリンク価値によって、ページ現有の検索エージェントを使って、得た膨大な量の結果を自動的にリコメンデーションすることができる。

(4) ツールの転用：

本研究に使われた Web ロボットや分析プログラムなどのツールはインターネットやネットワークのリンク構造を対象とするユーザーやコミュニティのグループ化などに関する他の課題に使うことができる。

● まとめ

本研究では、対象のネットワークに対して、リンク構造の特徴についての分析を行った。種々の結果をまとめてみると、以下のような結論が得られた。ネットワークの形状はネットワーク内の個体（Webページ）の特徴の表れであり、主・客観的な要素によって、ネットワークの特徴が決められると考えられる。ここでいう主・客観的な要素とは、人の嗜好、趣味、知識、情報及び人の所属団体、グループの構造やインターネットへの接続状況などのことである。われわれが得た実験結果は、ネットワークにおける情報・知識の分布及びネットワーク形態がこれらの要因と密接に関係する可能性を示した。さらなる定量的な分析をすることによって、その具体な関係を発見できるものと期待している。

人工ネットワークと実際のネットワークとを分析することから得られた知見として、ネットワーク内リンクのランダム性、リンクの結合率、ネットワークの開放度、平均 AUT/平均 HUB、四つの評価項目を基準にネットワークの形態を主に以下の４タイプに分類できることが分かった。

① Hub 価値、開放型ネットワーク。

Hub リンクの平均値が高く、ネットワークの開放度の値も高いネットワーク。実用的なページが多く含まれ、ページとページの間にもリンクが積極的に張られている。実用性の高いネットワーク。今回の実験の中で、Java（Java 言語に関するページからなるネットワーク）がそれの典型的な例である。

② Aut 価値、開放型ネットワーク。

Aut リンクの平均値が高く、ネットワークの開放度の値も高いネットワーク。リンク集の多いページが多く含まれ、ページとページの間にもリンクが積極的に張られている。便利性が高いネットワーク。今回の実験の中で、Hp（ホームページを作ることを趣味とページからなるネットワーク）がそれの典型的な例である。

③ Hub 価値、閉鎖型ネットワーク。

Hub リンクの平均値が高いが、ネットワークの開放度の値が低いネットワーク。

実用的なページが多く含まれてるが、ネットワーク外のページへのリンクが相対的にすくない。企業や組織団体の内部ネットワークなどが、それの例である。

④ Aut 価値、閉鎖型ネットワーク。

Aut リンクの平均値が高いが、ネットワークの開放度が低いネットワーク。ネットワーク内部ではリンクが多く張られているが、ネットワーク外のページへのリンクが相対的にすくない。Jaist の知識科学研究科の公式サイトがそれの例である。

5.2　課題

今後の課題として、以下のようにまとめる。

- 一回の実験における平均実験周期（約 20 時間）が長いため、実験で分析したネットワークの数が相対的に少ない。しかし、現実のネットワークの範囲は広く、もっとさまざまな形態のネットワークが存在することが考えられる。

- ページデータとリンクデータを収集する際に、それらのデータをテキストファイルに保存する方式をとっていた。この方法では、プログラムの構成がシンプルになり、プログラムの作成にかかる時間が少なくなるというメリットがあるが、テキストデータの計算処理時間がかかり、ハードディスクとメモリの容量が無駄になるというデメリットもある。解決策として、データベース方式を使うことが考えられる。

- 本研究では、対象のネットワークに対して、リンクの結合形態に関する分析を行った。今後、さらに、ネットワークにおける情報や知識の分布とネットワークを構成する主客観的要素との関係をさらに定量的に分析することで、両者の間の関係を明らかにして行きたい。

謝　辞

本研究を進めるにあたり、終始的確な指導と助言を頂いた北陸先端科学技術大学院大学　知識科学研究科 林 幸雄助教授に心より感謝いたします。

また、論文を書くにあたり、御指導・御助言を頂いた北陸先端科学技術大学院大学知識科学研究科 橋本 敬助教授に深く感謝の意を表します。

実験データを収集する際に、多くの協力と研究内容について様々な質問や指摘をくださった研究室の後輩の皆さまに御礼を述べさせて頂きます。

参考文献

［1］ Duncan J. Watts & Steven H. Strogatz, “Collective dynamics of ‘small-world’ networks,” nature Vol. 393, No. 6684.

［2］ S. Chakrabarti, B. E. Dom, D. Gibson, R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, “Spectral filtering for resource discovery,” ACM SIGIR workshop on Hypertext Information Retrieval on the Web, Melbourne, Australia (1998).

［3］ Krishna Bharat, Andrei Broder, Monika Henzinger, Puneet Kumar, Suresh Venkatasubramanian, “The Connectivity Server: fast access to linkage information on the Web”, Computer Network and ISDN System 30 P469-477 (1998).

［4］ 福島伸一　伊庭斎志　石塚満, “WWW 情報空間のリンク構造を用いた弱い構造化”, 信学技報　AI98-93, KBSE98-63 (1999-03).

［5］ 藤本和則　松澤和光, “インターネット上の記述文から確率知識を構成する一手法”,情報学シンポジウム（1998-01)。

［6］ 山名早人, “インターネット広域分散協調サーチロボット”, Computer Today No.87（1998-09 ）.

［7］ 高橋範泰　山下剛史, “知人のネットワークの概念に基づいた情報共有機構”, 信学技報 OFS98-20, AI98-29(1998-07).

［8］ 吉田仙　亀井剛次　服部文夫, “インターネットにおけるコミュニティ形成支援”, 信学技報 OFS98-21, AI98-30(1998-07).

［9］ カール・シャピロ　ハル R・バリアン, “ネットワーク経済の法則”, ㈱IDC コミュニケーションズ, (1999) .

［10］Roert Lafore, “Java で学ぶアルゴリズムとデータ構造”, SOFTBANK.

［11］Jim Farley, “Java 分散コンピューティング”, オライリー・ジャパン。

［12］Elliotte Rusty Harold, “JAVA ネットワークプログラミング”, オライリー・ジャパン。

［13］林敏彦　大村英昭, “文明としてのネットワーク”, ㈱NTT データシステム科学研究所。

［14］Fah-Chun Cheong , “インターネットエージェント”, ㈱インプレス。