

研究成果報告書

事業名（補助金名） : 基盤的研究開発育成事業（若手研究補助金）
研究開発テーマ名 : 実ネットワークにおける網象因果の解析
研究代表者名 : 河内佑美【北海道大学大学院情報科学研究所／博士後期課程】

1. 背景・目的

グラフ理論的視点からネットワークの要素をノード、要素間の関係をリンクとみなすと、巨大で複雑に見えるが故にランダムであると考えられてきた実世界のネットワーク構造の一部について実は、ある秩序や法則を持った構造が存在するという事実が分かってきた。例えば、スケールフリーネットワーク*と呼ばれる構造が、インターネット、World Wide Web (WWW)、細胞の新陳代謝、たんぱく質の相互作用、映画俳優の共演関係、航空路網など共通に発見されている。また、人間関係、電力網、線虫の神経回路網などに見られるのがスマートワールドネットワーク**と呼ばれる構造である。これらの発見により、我々の身近に存在する様々な領域のネットワークにおける普遍性が垣間見られる。

本研究では、個々のミクロな相互作用から生まれる相互関係によって創発されるマクロな普遍的性質や現象をネットワークダイナミクス上の網象と位置づける。地球上のあらゆるもの現象を万象、天気に関する現象を気象、海洋に関する現象を海象、個人を取り巻く社会現象を環象と呼ぶように、申請者が命名した、ネットワークにおける現象や振る舞いのことである。また、網象における現象の因果関係を網象因果と呼ぶ。本研究は動的環境の中にある実世界のネットワークについて、具体的にはインターネットやWorld Wide Web (WWW) 上でユーザビリティの反映されるシステムを対象とし、網象因果の解析によって普遍的性質を探ることを目的とする。将来的には得られた知見により、ソフトウェア工学的側面からの情報システム開発を目指す。

*スケールフリーネットワーク：少数のリンクを持つノードと、膨大なリンクを持つノードが共に存在するため、リンク数を表現する代表的なスケールが存在しないネットワーク。各ノードが持つリンク数と、特定のリンク数を持つノードの割合の関係がべき分布となる。膨大なリンクを持つノードはハブと呼ばれ、ネットワークにおいて重要な役割を果たす。

**スマートワールドネットワーク：多くの要素がその全構成要素数に比べて、非常に少ない数の要素を介してつながるネットワーク。相互リンクが密であるコミュニティと、コミュニティ間のある程度のリンクの存在によってこのネットワークの特性が現れる。

2. 内容・方法

本研究では電子商取引システム、特にインターネットオークションに焦点を当てる。近年インターネットなどのネットワークを利用した電子商取引サービスが一般消費者に普及し急激な成長を見せている。大規模な情報、データが存在する中でそれらの関係性に注目し特徴を抽出することで、サービスを利用するユーザのビヘイビア特性を導き出す。また、ユーザの意志の働くミクロなユーザ間の相互関係によってマクロ的に普遍的な性質や特徴がネットワーク構造の中に出現するのではないかと考えている。そこでオークションユーザをノード、商品取引を行った出品者と落札者の関係を出品者から落札者への有向リンクと考えネットワークを構成する（図1）。このネットワーク構造を解析することによって、1)ユーザのビヘイビア特徴の抽出や2)ネットワークの時間的成長過程を探る。

まず、特定のオークションサイトで公開されているユーザ間の取引データを、クローリングシステムを構築することで取得する。クローリングシステムとは、指定 URL の HTML データをダウンロードし、テキスト解析・抽出を行うことで必要なデータをデータベースへ蓄積させるものである（図2）。本研究での取得情報は、ある商品の出品者ユーザ ID、落札者ユーザ ID、取引終了時間である。ユーザのビヘイビア特徴抽出のための解析手法として、HITS アルゴリズム[1]を適用する。さらに、ネットワークの成長過程には、各ノードのリンク数分布の時系列的変化を解析する。成長過程については、実データの解析だけではなく、既存のネットワーク成長モデル[2]を用いてシミュレーションを行い実データとの比較によって、得られた結果の検証を行う。以下に HITS アルゴリズムの詳細と、本研究で用いたネットワーク成長モデルについて示す。

2.1 HITS アルゴリズム

有向リンクを通じて各ノードの情報、つまり評価値がネットワーク全体へ伝播され再帰的に各ノードの評価値が計算されるというものである。本研究では、オークションのネットワークを扱うので、出品、落札の区別に関係するリンクの方向に注目すべきである。つまり各ノードに対し、そのノードから出るリンクとノードへ入るリンクの両者を考慮した解析手法が必要であるので、HITS アルゴリズムを用いて解析を行う。

【アルゴリズム】

ネットワークの隣接行列を A とする。このとき Web ページ i から j へのハイパーリンクが存在するとき $A_{i,j} = 1$ であり、存在しないとき $A_{i,j} = 0$ である。また全ページに対し、 a をオーソリティ度のベクトル、 h をハブ度のベクトルとする。各ページ i は以下に示す式(1)、(2)によってハブ度とオーソリティ度が与えられる。ここで、 $i \rightarrow j$ はページ i から j へのハイパーリンクを示す。

$$a_i = \sum_{j \rightarrow i} h_j \quad (1)$$

$$h_i = \sum_{i \rightarrow j} a_j \quad (2)$$

さらに、以下にアルゴリズムの手順を示す。

$$a = [1, 1, 1, \dots, 1] \quad h = [1, 1, 1, \dots, 1] \quad (\text{初期値})$$

Do {

$$a = A^T h; \quad h = Aa;$$

```

 $a = a / ||a||;$   $h = h / ||h||;$  (正規化)

} while ( $a$  と  $h$  が固定点  $a^*$  と  $h^*$  に収束していないとき)

 $a^* = a;$   $h^* = h;$ 

Return  $a^*, h^*$ 

```

a^* と h^* は $A^T A$ と AA^T の固有ベクトルに対応する。またハブ度とオーソリティ度はこのアルゴリズムによって a^*, h^* の安定状態へ収束することが証明されている。ハブ度の高いページは多くのオーソリティ度の高いページへリンクを張り、またオーソリティ度の高いページは多くのハブ度の高いページからリンクを張られていることになる。

2.2 ネットワーク成長モデル

スケールフリーネットワークのベキ乗則を説明した最初の成長モデルが Albert-László Barabási によって提案された。これは、各時間ステップにおいて、新しいノードがネットワークに加わり、優先的選択によって既に存在するノードへリンクを張るというものである。優先的選択とはノード i の持つリンク数 k_i に応じて確率 $\pi(k_i) = k_i / \sum k_j$ を求め、その確率に従ってリンクを張るノードを決定する選択のことである。ノードの付加のみではなく、新しいリンクの付加やつなぎ替えを考慮した成長モデル[2]も提案されている。本研究では実データの性質と照らし合わせ、汎用性の高い後者のモデルを適用する。以下に詳細を示す。

- m 本の新しいリンクが確率 p で加えられる。

新しいリンクの始点となるノードはランダムに選ばれ、また終点はノードの選択確率 $\pi(k_i)$ に従って選ばれる（図 3-(i)）。

- m 本のリンクが確率 q でつなぎ替えられる。

ノード i とそれにつながるリンク l_{ij} がランダムに選択され、まずこのリンクが除かれ、次にノード i から選択確率 $\pi(k_{j'})$ に従って新しいリンクが張られる（図 3-(ii)）。

- 新しいノードが確率 $1-p-q$ で付け加えられる。

新しいノードには m 本の新しいリンクが確率 $\pi(k_i)$ に従ってネットワークに既に存在するノード i につながれる（図 3-(iii)）。

上記のプロセスはステップごとに行われるが、[2] の実験結果からスケールフリー構造が形成されるパラメータ p, q の値が求められている。本研究では、その結果の中から $p=0.65, q=0.0, m=1$ を設定する。この値は、上記モデルにおいてリンクのつなぎ替えを行わないパターンであり、生成されるネットワークのノード数とリンク数が実データとほぼ同様になるものである。

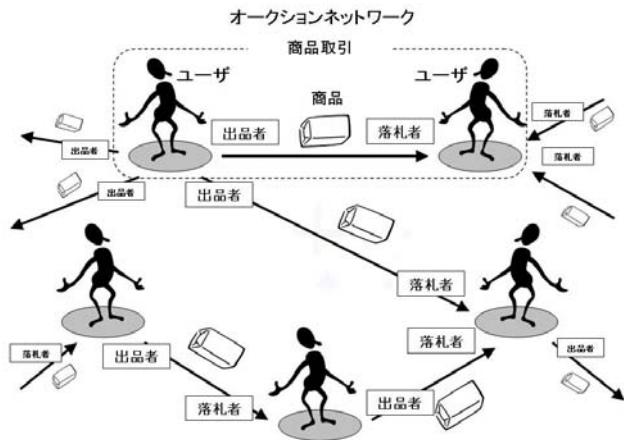


図 1. オークションネットワーク

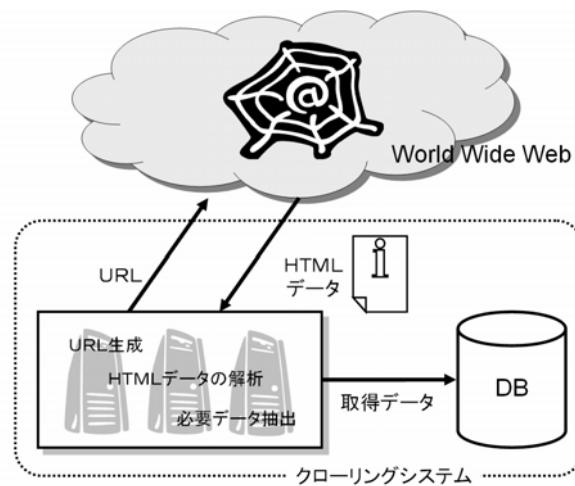


図 2. システム図

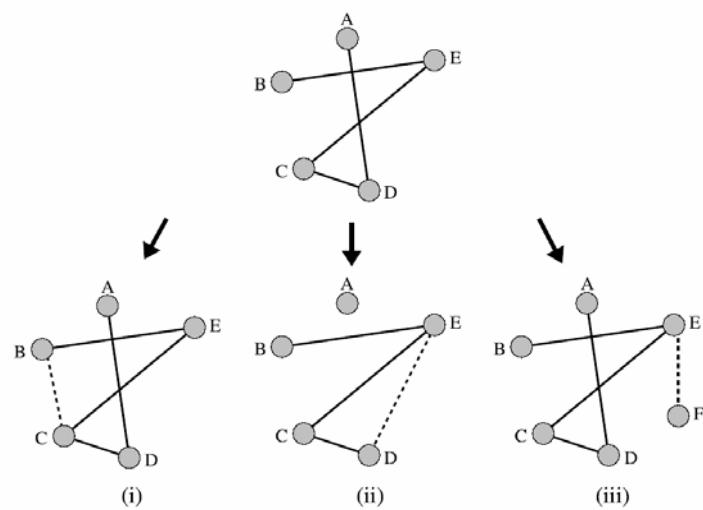


図 3. ネットワーク成長モデルのパターン

3. 結果・成果

3.1 取得データについて

本研究では、Yahoo!オークション (<http://auctions.yahoo.co.jp/>) サイトから以下の取引公開データを取得した。図 4 はあるユーザを中心とした 2 次のネットワークを可視化したものである。

- ユーザ（ノード）数：1,201,371
- 商品取引（リンク）数：3,416,070
- 取引終了時刻（リンク毎）

オークションサイトに公開されている全てのデータを取得するのは時間的、計算機リソース的にも困難であるが、上記のデータ数による解析結果は全体の振る舞いに反映できるものと考える。

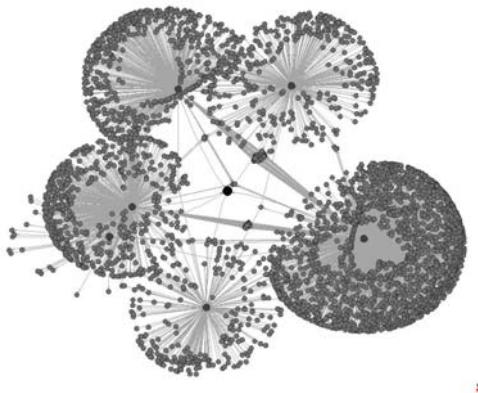


図 4. ネットワークの部分構造

上述したように、オークションネットワークの解析をユーザのビハイビア特徴の抽出とネットワークの時間的成長過程に絞って以下に述べる。

3.2 ユーザのビハイビア特徴の抽出について

取得したオークションのネットワークについて HITS アルゴリズムを用いて各ユーザのハブ度とオーソリティ度を計算する。本研究ではリンクの向きからハブ度を「出品度」、オーソリティ度を「落札度」と呼ぶことにする。各ユーザの出品度に対する出品回数を図 5 に、また各ユーザの落札度に対する落札回数を図 6 に示す。

まず、図 5 について、出品回数がほぼ同数であっても出品度の値に差があることが分かる。ここで、出品回数に対し、出品度の高いユーザ群に注目する。具体的には図 5 において $(\text{出品度} \times 6000 - 100 < \text{出品回数} < \text{出品度} \times 6000 + 100 \wedge \text{出品度} > 0.01)$ の範囲に当てはまる黒丸で示したユーザ群である。次に、各ユーザに対し上記のユーザ群と取引を行った回数の割合が全取引回数の 2 割以上であるユーザは図 6 の黒丸で示したユーザとなることが分かった。これらのユーザは落札度がほぼ同じユーザの中でも落札回数が低いユーザである。

つまり、図 5 と図 6 それぞれに示すユーザ群同士に強い関係性があるといえる。さらに、これらのユーザは図 5 においては出品専門であり、図 6 においては落札専門である傾向が高いことが分かっている。また、取得したユーザの中にはオークションストア（図 5 の白抜き四角）が含まれており、出品回数は多いが必ずしも出品度が高いわけではない。これらのことから、出品を主に行うユーザの中でも、落札を専門に行うユーザから商品を多く落札されているユーザが存在していることが分かる。以

上より、いわゆる“ブランド”ユーザとそれらのユーザと多くの取引を行なながらも全体としての取引回数は同じ落札度を持つユーザに比べて小さく、出品ユーザをある程度慎重に吟味しているのではないかと思われる“ロイヤリティ”ユーザの関係が出現した。

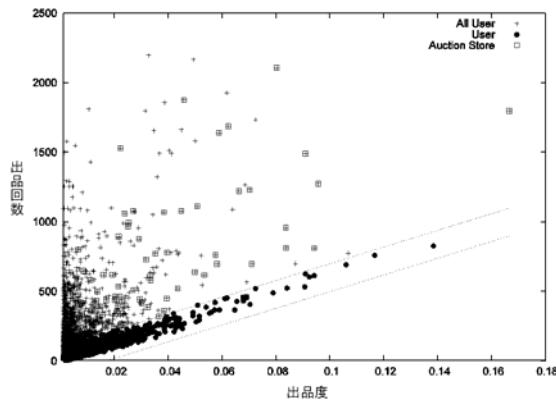


図 5. 出品度一出品回数の分布

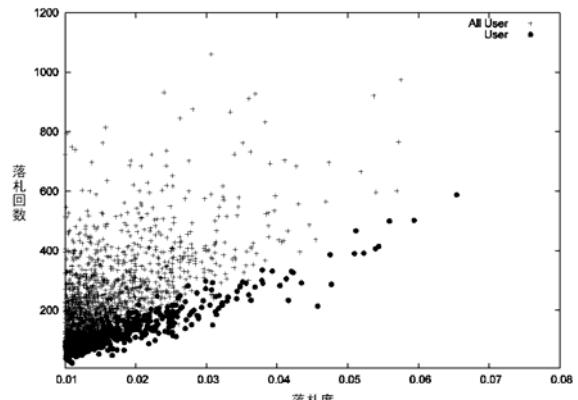


図 6. 落札度一落札回数の分布

3.3 ネットワークの時間的成長過程について

実世界では取引終了時刻間の間隔はそれぞれ異なるが、本研究においては全取得データを時間順に並べ、一つの取引を一ステップと定義して解析を行う。オークションネットワークの構造が時系列的にどのように成長・変化していくのかを調べた。

3.3.1 次数分布変化

まず、ネットワーク全体の構造の時間遷移について調べた。各ノードの次数（出品回数（=出次数）、落札回数（=入次数））とその次数を持つノード数の分布は、図 7 に出品リンクの分布、図 8 に落札リンクの分布を示す。それぞれの図において、1,000,000 ステップ目、2,000,000 ステップ目、3,000,000 ステップ目での次数分布を示している。横軸は次数、縦軸はノード数であり、両対軸とも log スケール化している。

ステップが増すごとに、ノード数やリンク数が増加するので、グラフ全体が上へシフトする。このとき、出品、落札それぞれのグラフの形は同じであるので、ネットワークは同じ構造を維持しながら成長していると分かる。特に、出品リンク（出次数）分布（図 7）はベキ乗則に従っている。一方、落札リンク（入次数）分布（図 8）は完全にベキ乗則に従っているわけではないが、変局点（次数 10~30）の前後で直線的なグラフになっている。

落札リンク分布が出品リンク分布のように一直線のグラフにはならない理由として、躊躇や警戒心、抵抗感なしに知人でないユーザから何かを買うという行為にはある閾値があるのではないかと考えられる。つまり、インターネットオークションにおいて落札の経験があまりないユーザにとっては、匿名性などの不確かさが躊躇や抵抗感へと導いている。経験を積むことによって、その躊躇、抵抗感をもたらす閾値を超えてオークションに落札者として参加するのではないかと考えられる。一方で、出品については金銭的リスクが低いので、落札のような警戒心や抵抗感が少なく好きなように参加できることから、図 7 のように出品リンクの分布がスケールフリーになると考えられる。

以上より、ユーザのビヘイビアや心理的状況がネットワーク構造という定量的に計測可能なものに現れることが分かった。

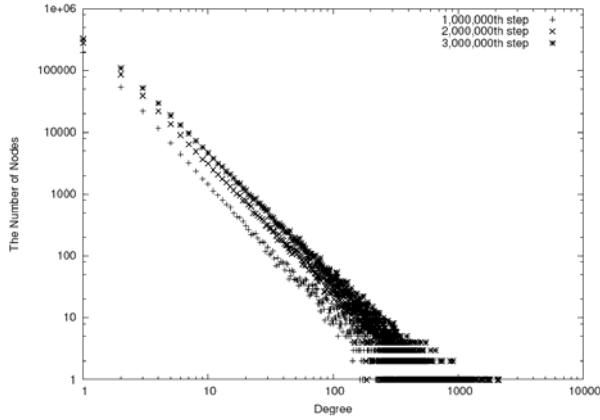


図 7. 出品リンクの次数分布

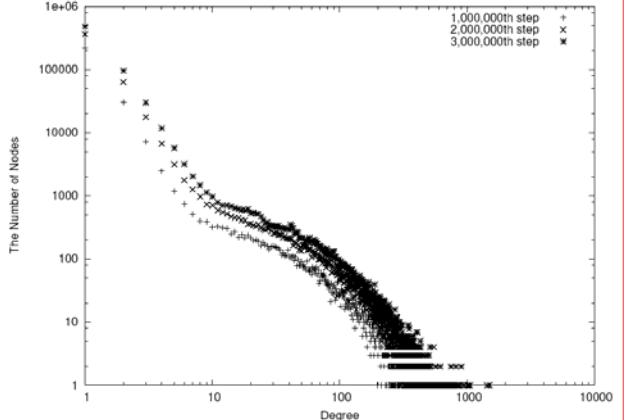


図 8. 落札リンクの次数分布

3.3.2 時系列リンク付加

インターネットオークションにおいて一つの商品取引が行われるとき、商品の値段と同様に、取引相手のオークション経験レベルや信頼性についても取引成立に影響を及ぼす。そこで、各ユーザが行ってきた取引の数とネットワークの構造を時系列的に解析することで、多くの要素を含む経済活動のダイナミクスを説明することが可能になるかもしれない。それぞれの取引がどのようないい間で成立してオークションネットワークが成長していくのかを、ノード間のリンク付加過程を解析することで検証する。特に、取引が成立したユーザのそれまでの出品、落札回数を調べ、リンク付加過程に一定のパターンや法則を見つけ出す。

解析の結果として、図 9 に取引が成立した出品者と落札者とのそれまでの取引回数（=出品回数+落札回数）を示す。横軸は各ノードの次数（取引回数）、縦軸はノード数を表し、両対軸とも log スケール化してある。このグラフは、取得した全時系列データのうち、任意の連続した 100,000 ステップを抽出し、解析したものである。また、どのステップ区間であっても同様の結果を示すことが分かっている。図 9 より、出品ユーザ、落札ユーザ共に同じ次数分布のパターンとなることが分かる。これらのパターンは、インターネットオークションにおいて、商品の値段だけではない基準によって取引相手として選ばれたユーザの分布特徴であるといえる。

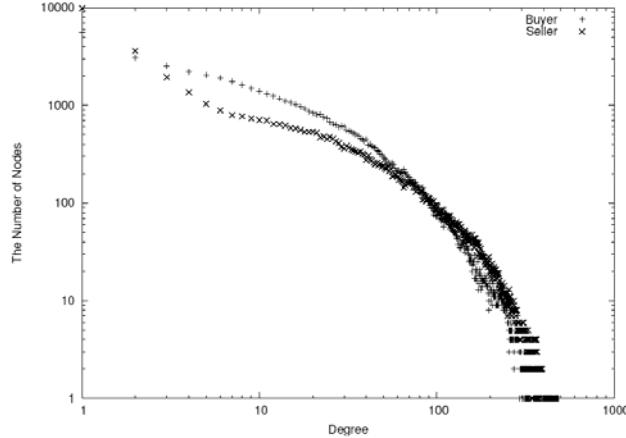


図 9. 出品ユーザと落札ユーザの取引成立時までに行った取扱回数分布

次に、図 9 によって出現したパターンの詳細な解析を行う。2 章で記述した既存のネットワーク成長モデルを用いてシミュレーションを行い、上述と同様のデータをプロットし比較をする。モデルではリンクに方向を定義していないが、本研究ではモデルのリンクやノードの付加プロセスにおいて始点ノードから終点ノードに方向を定義する。そこで、まずシミュレーションによって生成された任意のステップ後のネットワークの出次数、入次数の分布を図 10 に示す。始点ノードはランダムに選ばれるので出次数分布は図のようなカーブになる一方、終点ノードは優先的選択確率によって選ばれるので入次数分布は図のように直線となり、ネットワーク全体としてスケールフリー構造を形成していることが分かる。実データでも解析したように、時系列的にリンクがどのような次数を持つノードに付加されているかをプロットしたものが図 11 である。

図 7、8 から分かるように、オークションネットワークは全体としてほぼスケールフリー構造を形成している。よって同様にスケールフリー構造を形成する成長モデルと照らし合わせると、成長過程でのリンク付加については、実データでの時系列的な出品ユーザと落札ユーザの商品取引時の次数分布図 9 は、シミュレーションでの終点ノードの分布図 11 にほぼ一致する。つまり、オークションネットワークで、出品ユーザも落札ユーザも優先的選択確率に類似した確率で選ばれていることになる。出品回数の多いユーザはより多く、落札回数の多いユーザはより多く今後も取引を行うと予測される。

以上のように、ネットワークの構造や要素間の関係性を調べることで、ユーザのビヘイビアに関する特徴やパターンを抽出することが可能であることが分かった。個人のビヘイビアは個人単体だけで決定されるとは限らず、要素の関係性によっても特徴付けられるのではないかと考えられる。

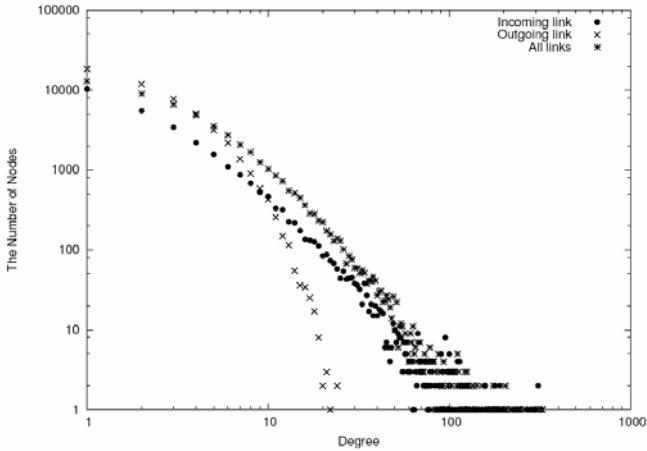


図 10. ネットワーク全体の次数分布

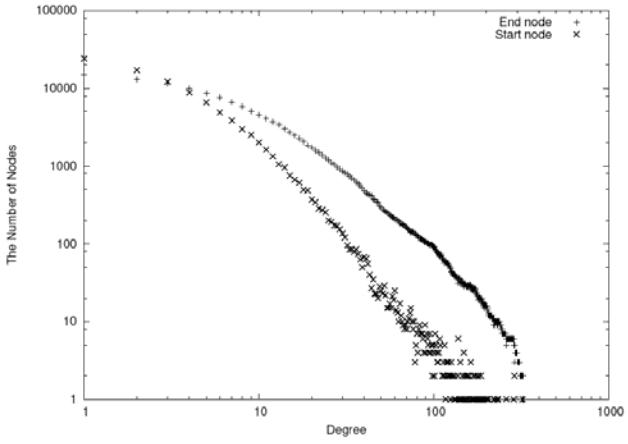


図 11. 選択されたノードの次数分布

4. 今後の展望

ノードとリンクによって構成されるネットワーク構造や構造遷移について、特徴的な性質やパターンが実システム上に出現することが分かった。今後はこれらの特徴をどのように工学的応用へ導くかが問題となってくる。特に人間のビヘイビアを主眼にしたシステムでは、ノードとリンクのみの構造だけでは情報不足である。一方で得られる全ての情報を取り入れてしまっては複雑になり過ぎ、裏側に隠れる重要な性質や特徴を見逃してしまう恐れもある。どのような情報を取り入れ、どのような手法で解析を行うのかを吟味し、工学的応用へ役立つ研究を進めなければならない。

そこで、本研究ではオークションネットワークのリンクにラベルという情報を付加したリンク解析によってユーザの特徴をさらに検証する予定である。リンクのラベルとは、取引における商品のカテゴリのことであり、例えば、コンピュータ、音楽、本、ファッショなど、24種類のカテゴリがサイト側で提示されている。これらの情報をユーザ間の関係性に付加して、関係性やつながりの構造のみからだけではなく、特定の意味や情報を持たせることでユーザのパーソナライゼーションなど工学的応用への道が開けるのではないかと考えている。

【参考文献】

- [1] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604-632, 1999.
- [2] Réka Albert and Albert-László Barabási, Topology of evolving networks: Local events and universality, *Physical Review Letters* 85, 5234 (2000).