

# ソーシャル・ネットワーキング・サービスにおける 人的ネットワークの構造

湯田 聴夫<sup>†,††</sup> 小野 直亮<sup>†</sup> 藤原 義久<sup>†</sup>

ソーシャル・ネットワーキング・サービス (SNS: Social Networking Services) という Web 上でのサービスが世界中で急激な拡大を続けている。SNS においては、先行研究では把握が困難な、相互認証された友人関係という人的ネットワークが大規模に顕在化している。本報告では、日本最大規模の SNS である mixi (ミクシィ) の 2005 年 2 月 15 日時点における 36 万ノード・190 万リンクの人的ネットワークを分析した。基本的なネットワーク解析結果として、次数のゆるやかなスケールフリー性、高い凝集性が確認された。構造を視認するために粗視化を行った。近年開発された解析アルゴリズムを適用し部分的なつながり (link) が密になっている高密度集団 (community) を抽出した。解析結果から独特な内部構造が視覚的に確認された。高密度集団は規模別に大中小の 3 種類に大別された。詳細に高密度集団の人数のばらつきを調べると、集団の人数に全体のトレンドとして Zipf 則が見いだされた。そして 100 人から 300 人ほどの高密度集団が分離されず、100 人以下から、300 人以上の集団へとサイズがスキップするという現象を発見した。発見された内部構造は、既存のモデルでは説明できない独特のものであることを確認した。新しいコミュニケーション・インフラへの進化も期待できる SNS、その研究の端緒を報告する。

## Structural Analysis of Human Network in Social Networking Services

KIKUO YUTA,<sup>†,††</sup> NAOAKI ONO<sup>†</sup> and YOSHI FUJIWARA<sup>†</sup>

Social Networking Services (SNS) have recently prevailed all over the World Wide Web. People grow up connections by making a tie to another who acknowledges as being a friend. Such a giant network of people, with each link being a mutually acknowledged friendship, has not ever been under previous investigation. This work analyzed the largest SNS in Japan, called *mixi*, comprised of 360,000 nodes and 1.9 million links as of February 15, 2005. Our analysis shows scale-free distribution of degree in its tail, and high cliquishness. To observe structure by coarse-graining, we employed a community (highly intra-connected group) extracting method developed by other researchers. As a result, we uncovered three classes of communities according to size. Overall rank-size plot shows Zipf's law for community-size distribution. Nevertheless, we found the existence of a skip in size, which implies absence of community between 100 and 300 in the number of people. This structure cannot be explained by models such as preferential attachment nor connecting nearest neighbor. We report some results of our analysis in anticipation to future advent and development of SNS as an innovative human communication infrastructure.

### 1. はじめに

人が織りなす組織や社会のネットワークを考えるとき、その構造において、人はネットワークの結節点 (ノード) であり、人と人をつなぐ関係性が紐帯 (リンク) となる。リンクの定義と抽出のやり方によって、人

的ネットワークをどのような側面から探求するのかが決まる。先行研究では、学術論文の共著者ネットワーク研究<sup>1)</sup> や、俳優の共演ネットワーク研究<sup>2)</sup> などにおいて、所属ネットワーク (Affiliation Network) を縮約したネットワークで研究されている。ここでは、論文や映画という特定のイベントへ参加している人同士の間で、全員にリンクを張るという操作が行われている。参加者の規模が大きくなれば、直接本人は知らない人も含まれる可能性があるうえ、研究や創作活動など、特定の目的で参加した集団により構成されていることになる。

社会に普遍的に存在する「知人」や「友人」という

<sup>†</sup> ATR ネットワーク情報学研究所  
Network Informatics Laboratories, ATR

<sup>††</sup> 京都大学大学院情報学研究所  
Graduate School of Informatics, Kyoto University  
現在、科学技術新興機構 ERATO 金子複雑系生命プロジェクト  
Presently with Complex Systems Biology Project, ERATO, JST

関係性は、共同研究で論文を書くような厳選された交流関係よりは、おだやかで多様である。職場の同僚や学校の同窓生、趣味や居住地の近さなど、さまざまな背景を有しており、関係性自体が一元的ではなく多元的なつながりである。多様な人と人との関係についての研究は、社会学、とくに数理社会学において、質問紙やインタビューによる調査によって多くの蓄積がある。金光<sup>3)</sup>が示すように近年では社会関係資本の測定へ向けて新しい地平が拓かれてきている。Crossら<sup>4)</sup>は社会ネットワークにおける潜在的な力として、だれが何を知っているか?(who knows what)を質問して集団の相互認識のネットワークへアプローチし、情報伝達、支援関係、問題解決のネットワークなどの調査法を整理している。しかしながら、調査コストもかかるために規模が小さく、ネットワークとしての研究は、“社会”というよりは“組織”というサイズである。

現時点での、社会学におけるネットワーク分析は、「社会的なアクターからなるネットワークの社会的な分析」という意味であり、「社会規模」のネットワーク分析ではない。大規模な社会調査もあるが、各人の友人数などのノードの属性は調査可能だが、交友関係としてのリンクを同定し、編み目状に結び付けたネットワークそのものを大規模に調査しているわけではない。

社会的な規模で、「知人」「友人」といった人のつながりの総体としてのネットワークを求めるとき、電子メール<sup>5),6)</sup>のデータを、社会ネットワークとして解析する研究がある。電子メールのトランザクションからなるネットワークの研究では、送受信者間でリンクが張られる。大学<sup>5)</sup>や企業<sup>6)</sup>という限定した範囲内での交流を調査することで迷惑メールやデータ入手、プライバシーの問題を回避している。しかしネットワークとして解析できる交流範囲は大学内や企業内に限定されている。また、基本的にメールは一方的に送られた向きを持ったリンクとも指摘でき、送られた側にとって、まったく意味をなさない相手である可能性もある。これは個人ブログのトラックバックなどにもあてはまる指摘であり、すべてにスパムの影響を指摘できる。ゆえに、これらの手法は社会ネットワークとしてのデータ精度を上げる付加的な処理が求められる。

そこで今回注目したのが、Web技術によるコミュニケーション支援を受けた、新しい社会ネットワークである。現在、急激な拡大を続けている SNS とは、自己プロフィールの Web 公開を特別な知識がなくてもできるようにしたうえで、会員相互の出会いやコミュニケーションを促進する仕掛けが盛り込まれたサービスである。SNS とは、2003 年 3 月に開始した米国のフ

レンドスター: Friendster<sup>1)</sup>から始まったサービスで、Web 上での実名公開を厳守したコミュニティ・サイトとして、わずか 3 カ月で 100 万人を集め、そのまま急拡大を続け 2006 年 1 月時点で 2,400 万人に達している。Friendster の音楽版とも呼べる Myspace<sup>2)</sup>は 20 代を中心に爆発的な支持を受け、わずか 22 カ月で Friendster を抜き去り、2006 年 1 月時点で 4,500 万人が登録し、群を抜いたアクセス数<sup>3)</sup>で世界最大規模の SNS となっている。ほか、欧州、アジア、各国で SNS は台頭している。日本でも 2004 年の 2 月後半に Mixi: mixi<sup>4)</sup>と、GREE: GREE<sup>5)</sup>がスタートし、当初 GREE の規模の方が大きかったが、2004 年 9 月頃に 10 万規模で逆転した。2005 年 8 月 1 日には mixi が 100 万人を超え、一方 GREE は 20 万人強であった。2006 年 1 月時点で mixi は 250 万人を超えている。ほかにさまざまな SNS が国内で運営されている。

SNS の入会には、すでに入会している会員からの招待が必要であることが特徴の 1 つでもあり、入会時点では招待者のみに友人としてのリンクが張られている<sup>6)</sup>。そこから、共通の趣味や友人の友人などをたどって、さまざまなコミュニケーションを介しながら「自分の友人」として「相互認証したリンク」が張られていく。SNS では、大学内や企業内に限定されないさまざまな属性を持った多様な人的ネットワークが、相互認証に基づき実現されている。これは、質問紙調査でいえば“あなたはこの人と友人ですか?”と、相互に確認を取った数十万、数百万人分の調査結果を意味しており、従来法ではとうてい入手できなかった情報である。

SNS の大規模ネットワーク解析は、Holme らによるスウェーデンの SNS についての社会ネットワーク的な解析がある<sup>7)</sup>。彼らが対象とした SNS は、恋愛相手を見つけることを主眼とした「出逢い系サイト(dating site)」である。すべての SNS 上での接触をネットワークとするとノード数が 3 万弱であるのに対して、友人のネットワークは 1 万 4 千人ほどの規模であり、基本的に各個人は別々に異性を探しているサイトといえる。本稿における「人が織りなす社会のネットワークを分析する」という立場からは、目的も限定

<sup>1)</sup> Friendster: [www.friendster.com](http://www.friendster.com)

<sup>2)</sup> Myspace: [www.myspace.com](http://www.myspace.com)

<sup>3)</sup> [www.alexa.com](http://www.alexa.com) 調査結果: 世界 13 位 (2006/1/10)

<sup>4)</sup> mixi: [www.mixi.jp](http://www.mixi.jp)

<sup>5)</sup> GREE: [www.gree.jp](http://www.gree.jp)

<sup>6)</sup> mixi, GREE は、紹介者が必要。Myspace は不要。

されており、かなり偏ったネットワークである。

mixi は“知り合い系”とも呼ばれており、元々知り合いである人同士が交流する場として、多く活用されている。つまり、実際の社会的な友人・知人関係の人的ネットワークがオンライン上に転写されている傾向があるとも考えられる。そのうえで、さまざまな既存の仲間との交流や、旧友との再会や、新規に仲間や友人を求める活動を、システムが支援している。そこで SNS をさらに分類し、Holme らの出逢い系サイトに対して「コミュニティ・サイト (community site)」として改めて位置づけを明確にする。本稿では、コミュニティ・サイトと位置づけられる SNS として、日本最大の mixi を対象とし、2005 年 2 月 15 日時点、36 万人の人的ネットワークの分析結果を報告する。

いったい、人と人が織りなす社会的なネットワーク構造とは、どのような特徴を有しているのだろうか。本稿の構成は、2 章で基本的なネットワーク指標に基づいて mixi のネットワークを解析する。3 章において Girvan と Newman の研究成果<sup>8),9)</sup> を紹介し、実用的な時間で高密度集団を抽出し、mixi の粗視化結果を示す。4 章では、抽出した高密度集団の分布を精査し、Preferential Attachment モデルと、Connecting Nearest Neighbor モデルを用いて、mixi データとの比較検討を行う。さらに次数の大きいノードを除去した場合の、構造への影響を分析する。5 章では、4 章までで示すユニークな内部構造が生成してくるメカニズムや、SNS 研究の意義について考察する。最後にまとめを行う。

## 2. ネットワークの解析

ネットワークにおけるノード数、つまりユーザ数の推移を図 1 に示した。縦軸がユーザ数、横軸が 2004 年 3 月 1 日からの経過日数となっている。左の図が線形グラフで、右の図は両対数グラフで描画した。ここで mixi の会員数が、冪関数に従って急成長を続けていること、および、その成長速度がいまだに維持されていることが分かった。一般公開は 3 月 3 日からだが、3 月 1 日時点で約 600 人参加していた。一般公開前の 2 月 22 日からプレ・オープンしていたことをヒアリングを通じて確認した。

次にネットワーク全体の特徴を確認した。データは 2005 年 2 月 15 日時点で加入している 363,819 人をノードとし、お互いが友人であると SNS 上で承認し

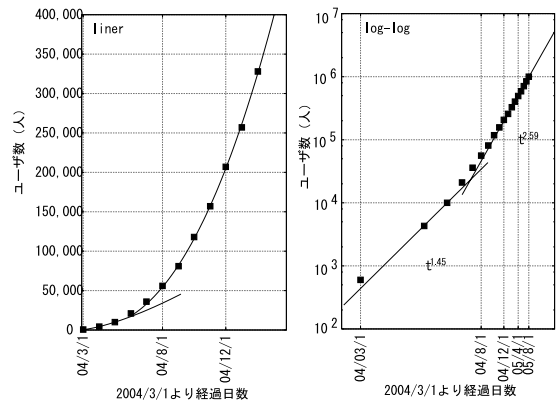


図 1 mixi ユーザ数の増加 (=ノード数の増加)  
Fig.1 Growth in the number of mixi users (nodes).

合った人と人のつながり (mixi では“マイミクシー”と呼ぶ) をリンクとしたネットワークとして扱った。総リンク数は、1,906,878 であった。最大連結成分が含むノード数は 360,802 で、以下の分析における統計量にはほぼ影響がないことから全体のネットワークを対象に分析した。mixi の人的ネットワークは相互認証であるため無向グラフになっている。

### 2.1 次数分布

次数に関して、平均  $\bar{k} = 10.4$ 、標準偏差  $SD = 19.4$ 、最大  $k_{max} = 1,301$ 、また次数が 1 のノード数は 85,846 であった。現在、mixi においては、1,000 人以上のリンクを追加して持つことがシステムのできなくなっている。図 2 は横軸が次数  $k$  に対応し、縦軸が累積確率分布  $P(\geq k)$  に対応する。累積確率分布とは  $k$  以上の次数を持つノードを見いだす確率である。図を確認すると、9 割以上の人々が属するところまでは指数分布的であるが、次数の高い領域にはべき指数が 2.8 程度のスケールフリー性を確認できた。

### 2.2 クラスタ係数

次にクラスタ係数を用い、ネットワークの凝集度合いを確認した。クラスタ係数はネットワークの中でノードが塊になっている度合いである。ノード  $i$  の次数が  $k_i$  だとして、 $k_i$  本のリンク先のノードどうしがすべてつながりあっている場合のリンクの組合せは  $k_i(k_i - 1)/2$  である。クラスタ係数はすべてが実現されていた場合に 1 となり、実際に存在した数を  $E_i$  と

mixi とは mix (交流する) と i (人) を組み合わせた造語  
株式会社ミクシー 2005/8/3 プレスリリース

本研究の学術的な目的のために、mixi の運営会社である株式会社ミクシーにデータ提供を依頼し、秘密保持契約を交わして提供を受けた。データは、個別 ID が特定できないように変換されたノード ID によるリンクデータだけを提供された。ただし、ユーザでもある著者 3 名がどのノードに対応するかだけ、情報の提供を受けた。なお、本研究は、SNS におけるユーザやシステムを何ら個別評価するものではないことを強調しておく。

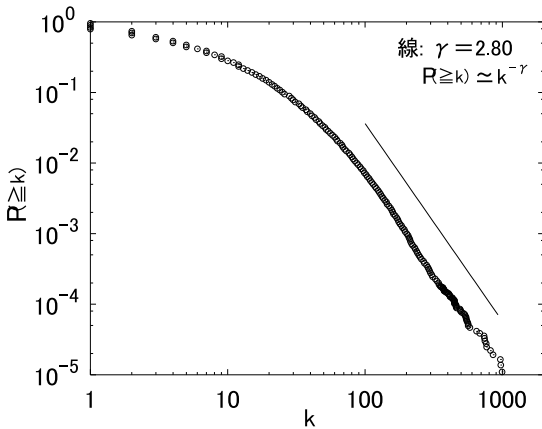


図2 mixi ネットワークにおける次数の累積確率分布  
 Fig. 2 Cumulative distribution of degree in the mixi network.

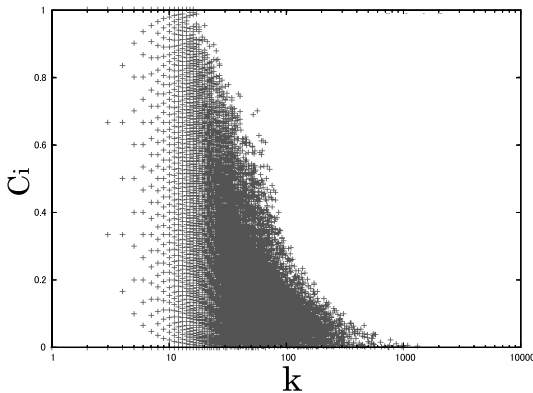


図3 次数に対するクラスタ係数の分布  
 Fig. 3 Scatter plot for degree and clustering coefficient.

すると、ノード  $i$  のクラスタ係数  $C_i$  は、可能な数との比となる。

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \tag{1}$$

ここで、次数に対するクラスタ係数の分布を図3に示した。横軸が次数  $k$  に対応し、縦軸がクラスタ係数  $C_i$  に対応する。次数が100近くなくても0.2近いクラスタ係数を持つノードが確認できるなど総じて高い凝集性が確認できる。各ノードのクラスタ係数を全ノードで平均したネットワーク全体のクラスタ係数は、 $C = 0.328$  となっており、高い凝集性が確認された。

### 3. Girvan と Newman のアルゴリズムによる粗視化

ネットワークはノード数が増えていくと、次数の分布や各種ネットワーク指標の分布のように全体のトレンドは確認できても、ネットワークそのものの構造を

確認することは困難になる。そこで内部にあるリンクの濃い集団の部分を実部構造として抽出し、内部構造の間の関係性に記述しなおすことが重要になる。数理社会学における blocking model などが相当する<sup>3)</sup>。これはブロックとしてまとまりのある集団を抽出し、その集団間の関係性を調べる手法である。社会ネットワーク分析では UCINET を始め各社からツールが提供され、とくに Negopy 分析などが有名であるが、現時点で36万ノードを実用時間で計算できるツールはない。

一方で、近年の物理学者による数多くのネットワーク解析技術の研究により、リンクが密な集団の抽出法が、複数開発されてきた。本報告は mixi のネットワークの特徴を伝えることに焦点を絞るため、36万規模に対応できるため今回用いた Girvan および Newman の方法<sup>8),9)</sup> のみを概略として説明する。

原著では、リンクが密な集団 (community) を抽出する手法をコミュニティ構造抽出法と呼んでいる<sup>9)</sup>。しかし SNS を分析する場合、多くの SNS には、ユーザが自分に近い趣味の人を探したり集ったりするために作ったり参加したりできる“コミュニティ”というサービスがあるため、構造から解析的に抽出するリンクの密な集団を同じ“コミュニティ”と呼称することは混乱を招く恐れがある。また人ではないネットワークであっても抽出アルゴリズムは適応できるが、すべてにコミュニティという用語が馴染むわけではない。そこで、本稿では、Girvan と Newman による手法で抽出されるリンクが密な集団からなる内部構造を GNS (Girvan Newman Substructure) と呼ぶことにする。またその抽出や分析に関しては GNS 抽出および GNS 分析と呼ぶことにする。

#### 3.1 GNS 抽出法

ノード数を  $n$ 、エッジ数を  $m$  としたときに、GNS 抽出アルゴリズムは計算量が  $O((m + n)n)$  で確実に求まる。とくに疎なグラフであれば  $O(n^2)$  の計算量で求まる、これが最大の強みである。Girvan ら<sup>8)</sup> が最初に報告した方法は、媒介中心性 (Betweenness Centrality) を用いていた。より強い媒介者から切断し、切断後に媒介中心性を再計算し同じ処理を繰り返すモデルで、計算量が  $O(m^2n)$  に従い、条件が良い場合で  $O(n^3)$  であった。新しいアルゴリズムは、媒介中心性を用いずに異なるコンセプトで作られているが、計算結果が媒介中心性を用いたものと近似できる

UCINET6=最大 32,767 ノード、実務的には 5,000-10,000 : [www.analytictech.com/ucinet.htm](http://www.analytictech.com/ucinet.htm)  
 An algorithm for extracting community structure

ことを Newman が確認している<sup>9)</sup>。

概要を説明すると、すべてのノードについて「リンク先と合わさって1つの集団になるべきか」を計算していく。その判断基準に、「モジュール性 (modularity)」という指標を導入している。ネットワーク全体を複数の集団に分けた場合に、リンク密度が各集団内で有意に高いかどうかを意味している。具体的な定義として、ネットワーク上から取り出した任意の集団  $i$  および集団  $j$  について集団  $i$  から集団  $j$  へのリンク数が全リンク数に占める割合を  $e_{ij}$  とする。 $e_{ii}$  は集団内部のリンク数の割合を意味する。ここで自集団から自集団を含む全集団へのリンク数の割合を  $a_i = \sum_j e_{ij}$  と表して、彼らはモジュール性  $Q$  を次の式で定義した。

$$Q = \sum_i (e_{ii} - a_i^2) \quad (2)$$

第1項は、そのまま自集団内でのリンク密度を意味し、これだけではネットワーク全体を1つの集団とすると一番高い値になってしまう。そこで第2項は、全体を1つにする場合も含めて、一様ランダムに集団を指定すると第1項と同じ値になり、有意でない場合を打ち消すために設定されている。

ここで最初にすべてのノードを構成要素数が1の“仮集団”と再定義し、個々の“仮集団”を結合しながら更新していく。初期“仮集団”はノードと同数で、ノードのリンクと同じ初期“仮集団”間の任意のペア  $i, j$  に対して、1つの大きな“仮集団”として重合する場合の  $Q$  の変化が計算される。結果として、次の式で  $\Delta Q_{ij}$  を求めていくことができる。

$$\Delta Q_{ij} = e_{ij} + e_{ji} - 2a_i a_j = 2(e_{ij} - a_i a_j) \quad (3)$$

すべてのペアの中で一番  $\Delta Q_{ij}$  値を上げるペアが選ばれて1つの大きな“仮集団”として、比較的少ない計算量で再計算される。このとき、“仮集団”の総数が1つ減る。計算過程で再編される隣接行列上の  $\Delta Q_{ij}$  はステップごとに変わっていき、最終的にどのペアを1つにしても  $Q$  が増加しなくなる時点で終了となる。計算結果としては、同じ“仮集団”に含めると効果の大きい順にリンクのリストが並んでいる。つなげた部分だけを抽出すれば、最終的に残った“集団”のリストを得られる。この各“集団”がそれぞれ GNS である。各 GNS には、他の GNS に含めるよりその GNS に含めた方が全体のモジュール性を高めるノードが含まれている。

最終的な実装は、Clauset ら<sup>10)</sup> により改良されたアルゴリズムを実装した。疎なグラフであれば、

$O(n \log^2 n)$  の計算量で求まる。計算速度として具体的には、Linux 上で、Pentium IV 2.8 GHz、メモリ 1 GB で 36 万ノードで 6 時間となった。

### 3.2 mixi の GNS 描画

mixi のネットワークに対して GNS 分析を行い、リンクの密度が高い集団として GNS が抽出できる。そして、GNS 内のノードが他の GNS 内のノードとリンクしている場合に、GNS どちらのリンクとして再集計した。すると GNS をノードとし GNS 間リンクのある、粗視化されたネットワークが生成できる。そのネットワークを独自に開発した描画ツールで三次元描画した。描画法は、ノード間にリンクがなければ反発し、リンクがあれば引き合うようなモデルを用いた。すると図4の「3D view」に示すような構造が確認された。図中に示す1つ1つの球は GNS を意味しており、大きさは内部に持つノード数の対数に比例している。各球の色は描画上、視認性向上のためランダムに割り振った。注意深く観察すると、中心部に大規模な GNS が3つ確認でき、大規模なものから数多くの小規模な GNS へつながっている。一方、左上には中規模な GNS が確認される。それら中規模の GNS は主に中規模どうして互いに、また大規模 GNS にも接続している。これに対して中規模 GNS と小規模 GNS との接続は著しく少ない。このように複雑な内部構造の様子が視認できる。なお、周囲に散在するとくに小さい GNS は、非連結成分である。紹介を介して広がる mixi ではあるが、途中で退会する人がいるとその人から先が切れてしまい、他に接続がなければ非連結成分となる。視認性を上げるために非連結成分を除去し、薄い円盤上の擬似的な二次元に押し込めた描画の結果を「2D view」に示した。

### 3.3 ネットワーク・モデルによる描画の比較検証

GNS 分析で粗視化して描画した結果、複雑性を有した構造が視認できたが、そもそも GNS 分析や描画法の影響で発生しているのであれば意味がない。そこで、2つの大きな特徴を有するネットワーク生成モデルによって、mixi データと比較検証する。モデルとしては、Preferential Attachment (PA) モデルと Connecting Nearest Neighbor (CNN) モデルの2つを用いた。

PA モデルは、Barabási らによる BA モデル<sup>11)</sup> の、 $m_0 = m = 1$  を初期条件として生成させる 10 万ノードのネットワークとした。ここで  $m_0$  は最初にクリークにするノード数、 $m$  は1ノード追加時の追加リンク数である。つまり、1つのノードから始まり、追加ノードは1本のリンクを持ち、追加時点での既存ノードの

原著では単にグループ (group) と記述<sup>9)</sup>。

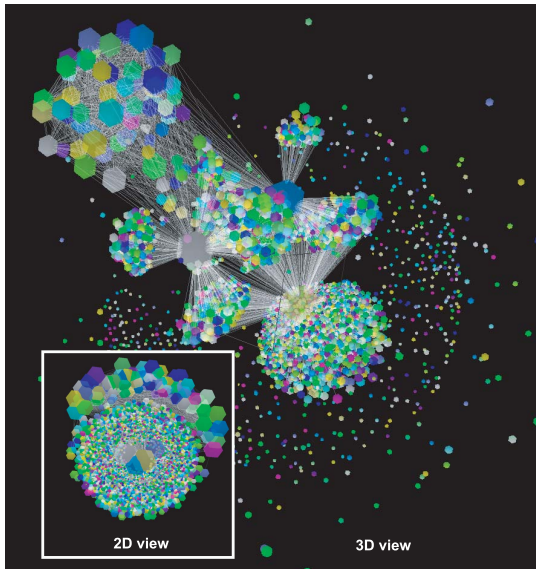


図 4 mixi の GNS 間の結合ネットワークの描画

Fig. 4 3D & 2D views of connections between extracted substructure.

各次数の総次数に対する比を接続確率として、優先的に接続していく。モデルの定義上、ネットワーク内部のリンクには三角形の構造が形成されないため、次数分布におけるスケールフリー性だけを有するツリー構造が生成される。

CNN モデルは、知人を紹介しあう仕組みのモデルで、次数分布のスケールフリー性と同時に、高い凝集性を実現されている。初出は Davidsen らによる DEB モデル<sup>12)</sup>で、ノード数を固定して計算するモデルとなっている。DEB モデルでは、紹介による結合とノードである人の生死のダイナミズムがモデル化されており、定常状態としてネットワークを得る。本研究では、SNS のモデル化も視野にいれるため、後に Vázquez により整理された、ノードが増えていく成長モデルの CNN モデル<sup>13)</sup>を用いた。基本概念は潜在的リンク<sup>13)</sup>という考え方である。新しいノードがネットワークに入るときに、リンク先を友人とすれば、友人の友人との間に潜在的リンクを形成させる。

CNN モデルは、以下の確率過程により、潜在的リンクの形成と実在リンク化を行う。(1) 確率  $1-u$  で新しいノードをランダムに付け加え、そのノードから潜在的リンクも張る。(2) 確率  $u$  で、すでにある潜在的リンクから一様ランダムに選んだ 1 つのリンクを実際のリンクに変更する。本研究では、 $u = 0.3$  で 10 万ノードのネットワークを比較検証用に生成した。

これらのモデルにより生成された、明らかに構造の

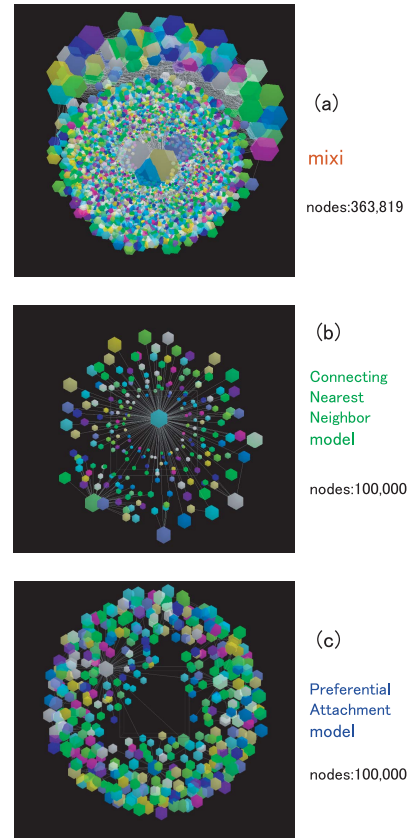


図 5 mixi とモデルの内部構造比較

Fig. 5 Graphical comparison between real data and models.

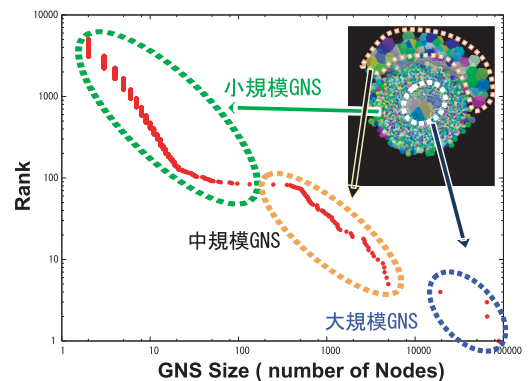


図 6 GNS のランク・サイズ・プロットおよび二次元描画との対応  
Fig. 6 Rank-size plot of GNS with correspondence to 2D view for the mixi data.

異なる 2 種類のネットワークと、mixi における実際の人のネットワークを、GNS 分析と粗視化を通じて比較する。とくに、独自の三次元描画ツールの影響を避けるため、粗視化では、すべて同一パラメータで描画していることを強調したい。結果を、図 5 に示す。描

画の色やサイズは、図4の「2D view」と同様にして、mixiの実データを図5(a)に、CNNモデルを図5(b)に、PAモデルを図5(c)に示した。

その結果、容易に内部構造の大きな違いが確認できた。(a)のmixiでは、中心に大規模GNS、そのまわりに小規模GNSが集まり、中規模GNSは周辺部にかたまっていた。中規模GNSは、間にリンクがあることによって描画上引き寄せられて1カ所に集まって描画されていた。(b)CNNモデルでは、大規模GNSを中心としたスター構造になっておりサイズもさまざまなものが連続的に存在していることが確認できた。(c)PAモデルでは、GNSサイズがどれもほぼ同じであり、中心に位置するような大規模GNSは存在しないことが確認できた。

このように、ノード数が10万以上の大規模なネットワークであっても、GNS分析によって500–5,000ノードほどに粗視化することができることを実際に確認した。そして次数のスケールフリー性や、高い凝集性などのマクロ指標では差が大きくはなくても、著しい内部構造の違いを、粗視化を通じて容易に視認できることを確認した。

#### 4. サイズの分布の検証

4.1 mixiデータのGNSランク・サイズ・プロット  
モデルとのGNS描画の比較を通じて、mixiデータにはGNSサイズの分布に偏差があることが確認された。より定量的に調べるために、GNSのランク・サイズ・プロットを図6に示した。横軸がGNSのサイズ、縦軸はそのサイズの順位を意味している。つまり右下にあるプロットは、多くのノードを有する大規模なGNSで、左上のプロットは、内包するノード数の少ない小規模なGNSである。図4の「2D view」に対応するプロットを図の中に指し示している。その結果、小規模と中規模の間に、サイズが不連続となる独特な構造が発見された。

図7は、不連続であることを確認するために図6と同じプロットを再描画した。サイズを小さい方(図の左上)から大きい方(図の右下)へ向かって確認していくと、サイズが100–300のGNSがほとんど存在していない。それより小さいGNSから、いきなりサイズがスキップして大きくなっている。中規模GNS群の中は、 $x^{-\gamma}$ でガイド線を作図することで、 $\gamma \approx 1$ のZipf則であることが分かる。小規模GNS群の中は、 $\gamma \approx 1.3$ であった。GNSのサイズは基本的にはZipf則に従っているといえるが、このGNSサイズのスキップは、先行研究でも報告のない、きわめてユニークな

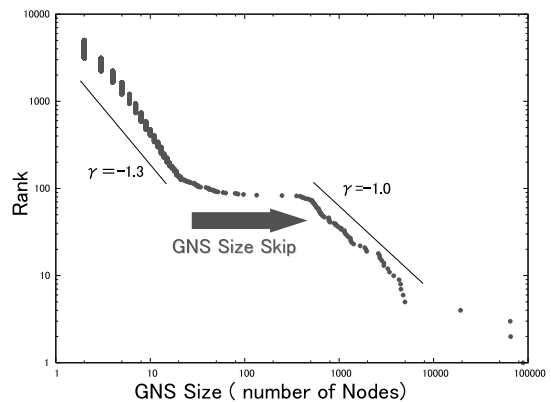


図7 mixiのGNSランク・サイズ分布におけるZipf則とサイズ・スキップ現象

Fig. 7 Rank-size plot of GNS with Zipf's law and size-skip for the mixi data.

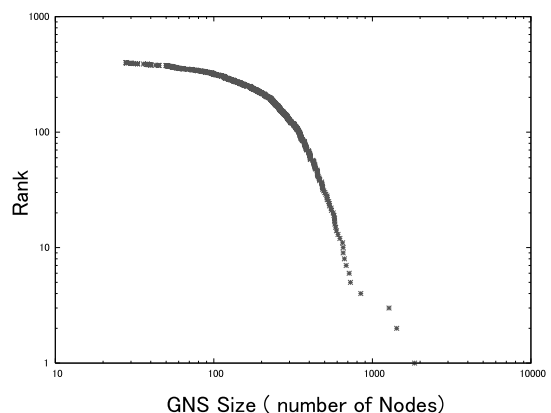


図8 PAモデルのGNS順位サイズ分布

Fig. 8 Rank-size plot of GNS for the PA model.

構造である。

#### 4.2 GNSサイズ・スキップのモデルによる検証

PAモデルとCNNモデルで生成したネットワークを、それぞれGNS分析し、ランク・サイズ・プロットを行った。PAモデルの結果を、図8に示した。横軸はGNSサイズに、縦軸は順位に対応している。このようにPAモデルによるスケールフリー性だけでは、GNSのサイズと順位にZipf則は見られず、むしろ指数的な分布を示していることになる。

次に、CNNモデルで生成したネットワークをGNS分析して、ランク・サイズ・プロットを行った結果を図9に示す。横軸はGNSサイズに、縦軸は順位に対応している。生成するネットワークのノード数に依存せず、Zipf則が確認できた。またノード数が10万以上の規模でネットワークを生成すれば、確率的に1つの実現(realization)であっても、GNS分析の結果にゆらぎがほとんどないことも確認した。CNNモデル

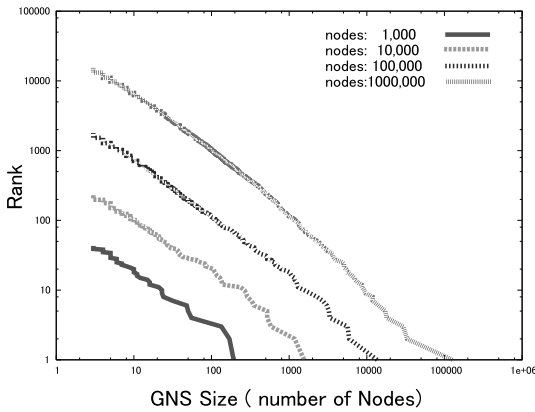


図 9 CNN モデルにおける GNS の順位・サイズ分布 (ネットワークサイズ別)

Fig. 9 Rank-size plot of GNS for the CNN model.

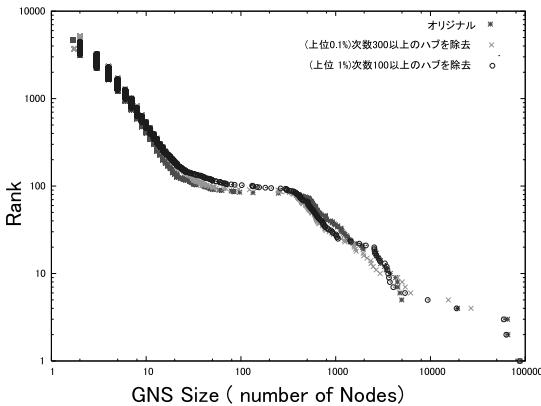


図 10 ハブの除去による内部構造への影響

Fig. 10 Absence of effect due to deletion of hubs in the mixi network.

では、GNS サイズ・スキップは存在しない。

#### 4.3 GNS サイズ・スキップ現象へのハブ除去効果

最後にネットワーク構造に対する、ネットワーク・ハブとも呼ばれる次数の高いノードの影響を検討した。ハブとハブにつながるリンクを除去する操作を行い、残った部分グラフに GNS 分析を行って、その影響を確認した。オリジナルに対して、次数が 300 以上 (0.1%) のノードをすべて除いた部分グラフと、次数が 100 以上 (1%) のノードを除いた部分グラフの、2 種のネットワークを作成した。このとき、ハブが除かれたことにより次数がゼロになってしまうノードも除去した。結果を図 10 に示した。その結果、ハブはわずかに影響するが、Zipf 則の途中に変曲点が入る内部構造である「GNS サイズ・スキップ」は壊れなかった。ハブには影響されず、次数の大きさに強く依存せず、普遍的なノードによるネットワーク構造から、内

部構造が形成されていることが分かる。

## 5. 考 察

本研究で初めて発見された「GNS のサイズ・スキップ現象」は、既存のネットワーク・モデルでは説明されないきわめてユニークな特徴である。GNS のサイズ・スキップの意味とはいったい何なのかを、以下で論考する。基本的な統計量から、スケールフリー性と局所的に高い凝集性が確認され、さらに GNS ランク・サイズ・プロットにおいて基本的に Zipf 則に従うことが確認された。同時に、CNN モデルも「次数分布のスケールフリー性」「高い凝集性」「GNS サイズの Zipf 則」までは満たしていることを確認した。つまり CNN モデルは、GNS サイズ・スキップ以外の特徴は有していることになる。

CNN モデルに対する、mixi という実態との差について考察すると、mixi に付加的に存在するメカニズムとして、「コミュニティ」「検索機能」「日記へのコメント」という機能が浮かび上がってくる。「コミュニティ」という機能は、ネットワークの経路長で距離を測ると遠く離れて散在する複数人がいきなり相互に知り合いとなり、クリーク (clique: 徒党) となることを支援している。「検索機能」は、遠く離れた 2 人を、直接結ばせることを意味している。これらは、CNN モデルにはまったく入ってないメカニズムであるので、今後モデル化すべき課題となる。

「日記へのコメント」を通じた交流のネットワーク的意義は、友人の友人以遠のネットワークの地平線をたぐり寄せていることにある。SNS 内のユーザにとって、直接の友人がネットワークの地平線となっており、その外側は陽炎のようによく分からない。「友人の友人」という地平線の向こう側の人とは、友人の日記へのコメントを介して最も頻繁に出逢うことになる。

Vázquez の CNN モデルでは、新規ノードの追加時にしか潜在リンクを張らないため、この働きは十分に反映されていない。現実的には、友人をたぐり寄せて自分の友人とした時点で、新たにその友人の友人への潜在的リンクが次々に生まれているはずである。CNN モデルは、スケールフリー性と局所的な凝集性をシンプルに実現しているが、SNS のモデルとして成立しているわけではないと指摘できる。

「足跡」という訪問者の履歴を観ることができるユニークな機能もあるが、これは、「相手の存在・働きかけへの気づき」として、上述 3 つの機能すべてを、助長し支援する方向で働いていると考えられる。

次数分布を振り返ると、mixi における友人数が 4 人



以下の人で全体の51%を占め、11人以下では75%に達し、個人では小集団が大勢を占める。紹介者と共通の友人もおもくクラスタ化が進み、実際の観察を合わせても、局所的に少人数でクリーク化している場合が多い。多くの人は1種類の帰属であるように見られたが、少なからず複数の集団に帰属していると見受けられる人がいた。そしてその人の複数の帰属先にいる友人は、その個別の集団内ではクリーク化しているものの、集団間ではリンクがあまりないようだった。

小さな集団に属する1人1人が「たぐったり」「オフ会に参加したり」「検索して再会したり、新たに出逢ったり」しながら、自分の所属する小集団を引き連れて、離れた場所で結合していきクリーク化したら、どうなるであろうか。サイズは、一気に膨らんでしまい、途中のサイズをスキップしてしまう可能性がある。これが、GNSサイズ・スキップのメカニズムとして考えられる。これらの作用はCNNモデルでは1つも考慮されておらず、現実のSNSをモデル化するうえでの課題となる。今後、実データの解析的アプローチを深めると同時に、より実態に適合するSNSネットワーク生成モデルの構築が課題となる。また、国内の他のSNS、海外のSNS、他の社会ネットワーク、さまざまな比較検討を通して、何が起きているのか、そして、何がこれから起きていくのかを明らかにしていくことも課題としたい。

最後に、大規模なSNSをネットワーク的に分析することが、どのような意味を持つのか、その社会的価値について示しておきたい。まず第1に、ネットワークのノード数が増加しても、再帰的な粗視化まで考慮すれば、原理的にネットワーク・サイズに依存せず1枚の画面に「全体」を描画できる。これは「ネットワークの地図」を作れることを意味する。つながりの変化に応じて、全体における粗密も変わるような、動的な地図である。中にいるノードとしての人にとって、自分から織りなされる関係性をひもとくうえでは本質的に意味のある地図となりうる。

また、大規模化の果てに想像できることもある。SNSなどWeb上のコミュニケーションは、よく「リアル(現実)」「ヴァーチャル(仮想)」と分けて議論される。しかし、mixiにおいては、ネットで知り合った友人(マイミク)を、それまでの既存の交友関係と分けて考えないユーザも多く存在するように見受けられる。それには、招待制や実名推奨など、さまざまな理由があると思われるが、従来の「仮想」からは、かなり「現実」に寄ってきたWeb上での交流であるといえるだけの実態がある。

詳細は誌面の都合上割愛するが、著者から、6次の隔たりでミクシィ・ユーザの96%までたどれることを別途検証した。この知見と次数のスケールフリー性を合わせると、日本国内のほとんどの人が6~8次強の友人の友人(FOAF: Friends Of A Friend)ネットワークでたどれることを示唆している。さらに、この地球上、すべての人が、具体的に実在するFOAFをたどって、出逢うことができる可能性を意味している。

SNSには、一度も会ったことのない遠い人であっても、仮想ではなく現実に、自分から親しい人を介してつながっていることを実感できる可能性がある。このようなコミュニケーション・インフラとしてSNSの可能性を眺めるとき、Social Networking Serviceとは、まだ出逢わぬ遠い友人を含めた、FOAFインタフェースであることに改めて気がつく。

## 6. ま と め

本稿は、先行研究にある「出逢い系サイト」ではなく「コミュニティ・サイト」として位置づけられるSNSを分析した、初めての試みであり、さらに、36万人という大規模な交友ネットワークデータの内部構造を調査した最初の事例である。

基礎的なネットワーク解析として、次数のゆるやかなスケールフリー性や、高いクラスタ係数が確認された。内部構造の理解のためにGirvanおよびNewmanにより開発された内部構造(GNS)の抽出法を適用し、マクロ統計量ではとらえられない構造の違いを、粗視化を通じて容易に把握することに成功した。

抽出したGNSは、大規模、中規模、小規模の3種のサイズに大別された。そして定量的な分析の結果、ノード数100~300の小~中規模のGNSがほとんど抽出されない領域があるというGNSサイズ・スキップ現象を発見した。抽出された中規模のGNSにおいては、ノード数(サイズ)とそのサイズの順位にZipf則を確認した。

次に、ネットワーク生成モデルと比較検証を行った。その結果、次数分布がスケールフリーであっても、PAモデルではGNSランク・サイズ・プロットはZipf則にならなかった。CNNモデルにおいてはZipf則が発生したが、GNSサイズ・スキップは発現しなかった。実際のmixiのネットワークは、既存のネットワーク生成モデルでは、説明されないことが確認された。

これがmixiの特徴か、SNSの特徴か、もしくは日本文化の特徴なのかを、今後詳細に分析していく。

謝辞 株式会社ミクシィの代表取締役笠原健治氏、およびシステム担当者の方々には大変お世話になり

ましたことを感謝いたします。本研究は独立行政法人情報通信研究機構の研究委託「人間情報コミュニケーションの研究開発」により実施したものである。

### 参 考 文 献

- 1) Barabási, A.L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A. and Vicsek, T.: Evolution of the social network of scientific collaborations, *Physica A*, Vol.311, pp.590–614 (2002).
- 2) Amaral, L.A.N., Scala, A., Barthélémy, M. and Stanley, H.E.: Classes of small-world networks, *Proc. Natl. Acad. Sci. U.S.A.*, Vol.97, pp.11149–11152 (2000).
- 3) 金光 淳: 社会ネットワーク分析の基礎—社会的関係資本論にむけて, 勁草書房 (2003).
- 4) Cross, R. and Parker, A.: *The hidden power of social networks*, Harvard Business School Press (2004).
- 5) Ebel, H., Mielsch, L.I. and Bornholdt, S.: Scale-free topology of e-mail networks, *Physical Review E*, Vol.66, 035103(R) (2002).
- 6) Tyler, J.R., Wilkinson, D.M. and Huberman, B.A.: Email as spectroscopy: automated discovery of community structure within organizations, *Proc. 1st International Conference on Communities and Technologies*, Kluwer, B.V. (2003).
- 7) Holme, P., Edling, C.R. and Liljeros, F.: Structure and time evolution of an internet dating community, *Social Networks*, Vol.26, pp.155–174 (2004).
- 8) Girvan, M. and Newman, M.E.J.: Community structure in social and biological networks, *Proc. Natl. Acad. Sci. U.S.A.*, Vol.99, No.12, pp.7821–7826 (2002).
- 9) Newman, M.E.J.: Fast algorithm for detecting community structure in networks, *Physical Review E*, Vol.69, 066133 (2004).
- 10) Clauset, A., Newman, M.E.J. and Moore, C.: Finding community structure in very large networks, *Physical Review E*, Vol.70, 066111 (2004).
- 11) Barabási, A.L. and Albert, R.: Emergence of scaling in random networks, *Science*, Vol.286, pp.509–512 (1999).
- 12) Davidsen, J., Ebel, H. and Bornholdt, S.: Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks, *Physical Review Letters*, Vol.88, No.12, 128701 (2002).
- 13) Vázquez, A.: Growing network with local rules: Preferential attachment, clustering hier-

archy and degree correlations, *Physical Review E*, Vol.67, 056104 (2003).

(平成 17 年 5 月 25 日受付)

(平成 18 年 1 月 6 日採録)



湯田 聴夫 (正会員)

昭和 46 年生。平成 9 年東京工業大学大学院生命理工学研究科修士課程修了。同年トーマツ・コンサルティング(株)入社。人事・組織の経営コンサルティング業務に従事。平成

14 年より京都大学大学院情報学研究科博士後期課程在籍。同年 ATR 人間情報科学研究所の研修研究員。平成 17 年より ATR ネットワーク情報学研究所の研究員として、適応進化計算、組織現象の計算機シミュレーション、社会ネットワークの分析研究に従事。組織学会、人工知能学会、経営情報学会、経営行動科学学会、進化経済学会各会員。



小野 直亮

昭和 48 年生。平成 13 年東京大学大学院総合文化研究科博士課程修了。同年京都大学大学院理学研究科吉川研究室ポスドク研究員。平成 14～

17 年 ATR 人間情報科学研究所研究員として、人工生命のモデルによる自己複製システムの創発、自己触媒ネットワークの進化の研究に従事。現在大阪大学大学院情報科学研究科にて大腸菌等の遺伝子発現ネットワークを分析中。物理学会、進化学会、生物物理学会、数理生物学会、International Society of Artificial Life 各会員。



藤原 義久

昭和 39 年生。平成 4 年東京工業大学大学院理工学研究科物理学専攻博士課程修了。日本学術振興会特別研究員、科学技術庁特別研究員等を経て、平成 12～15 年まで、郵政省

通信総合研究所主任研究員。現在、ATR ネットワーク情報学研究所主任研究員。社会経済のネットワーク科学、経済物理学、進化計算等の研究に従事。訳書に『経済における確率的モデルへの招待』(青木正直著、共訳、サイエンス社)等。ヨーロッパ物理学会、システム制御情報学会各会員。