

*MMRC*  
*DISCUSSION PAPER SERIES*

MMRC-J-107

SNS における関係形成原理

—mixi のデータ分析—

独立行政法人 産業技術総合研究所 / スタンフォード大学  
松尾 豊

東京大学大学院 経済学研究科  
安田 雪

2007 年 1 月



東京大学21世紀COE [整備済]  
ものづくり経営研究センター



# SNS における関係形成原理

## —mixi のデータ分析—

### How relations are built within a SNS world -- social network analysis on mixi

独立行政法人 産業技術総合研究所 / スタンフォード大学

松尾 豊<sup>1</sup>

東京大学大学院 経済学研究科

安田 雪<sup>2</sup>

2007 年 1 月

**keywords:** SNS, community, social network analysis

Our purpose here is to (1) investigate the structure of the personal networks developed on *mixi*, a Japanese social networking service (SNS), and (2) to consider the governing mechanism which guides participants of a SNS to form an aggregate network. Our findings are as follows: the

---

<sup>1</sup> Yutaka Matsuo, National Institute of Advanced Science and Technology / Stanford University, y.matsuo@aist.go.jp, <http://ymatsuo.com/>

<sup>2</sup> Yuki Yasuda, University of Tokyo, yasuda@mmrc.e.u-tokyo.ac.jp

clustering coefficient of the network is as high as 0.33 while the characteristic path length is as low as 5.5. A network among central users (over 300 edges) consists of two cliques, which seems to be very fragile. Community-affiliation network suggests there are several easy-entry communities which later lead users to more high-entry, unique-theme communities. The analysis on connectedness within a community reveals the importance of real-world interaction. Lastly, we depict a probable image of the entire ecology on *mixi* among users and communities, which contributes broadly to social systems on the Web.

## 1. はじめに

近年、ソーシャルネットワーキングサービス (SNS) が着目を集めている。SNSとは、WWW上で知人関係を形成し相互交流を行うためのプラットフォームであり、日本では2004年から、GREE, mixi, Orkutなどが、それぞれ独立のサービスとして無料で提供している。中でも、最もユーザ数の多いmixi (ミクシィ) は、ユーザ数が2004年5月に1万人、2004年9月に10万人、2005年8月に100万人に達した<sup>3</sup>。さらに増加のペースは衰えず、2006年10月にはユーザ数600万人を超えている。

SNSが日本で注目を集めた背景として、米国で2003年3月にFriendsterというWeb上のサービスが開始され、瞬く間に万単位の会員を動員したことがある。Friendsterは2006年10月時点で約2900万人のユーザ数を誇る巨大サイトに成長している。現在<sup>4</sup>、最もユーザ数の多いSNSはMySpaceで、アカウント数は1億を超え、世界でも上位10位以内に入るほどページビューが多い<sup>5</sup>。他にもオープンなプラットフォームでブログやコミュニティ機能を提供するLiveJournal (約1億アカウント)、韓国で1500万のアカウント数をもつCyworldなど、SNSは全世界的な広がりを見せている。

ユーザ数の増加もさることながら、最近ではSNSを基盤とし、さまざまなサービスが連携され始めている。例えば、Yahoo! は掲示板やオークションなど各種のサービスを統合する基盤としてSNS (Yahoo! Days) を展開している。mixi では、iTunesなどと連携し、PCで聞いている音楽リストを提供するmixiステーションをリリースし、GREEでも携帯電話との連携を強化している。学術的にもSNSは着目を集めており、数年前から人のネットワークを基盤にした情報検索[沼04]や情報共有[森05]の研究が、国内外を問わず活発に行われている

<sup>3</sup>株式会社ミクシィ 2006年3月2日プレスリリースより。

<sup>4</sup> 2006年10月時点。以下、いずれも同じ。

<sup>5</sup> Alexaによる定期的な視聴率調査による。

る。

SNS では、ユーザはシステム上で、自分の知り合いや友人を登録することで関係を構築していく。自分の日記や写真、本のレビューを知り合いの範囲内で公開することができ、さらにグループやコミュニティと呼ばれる掲示板的なシステムを利用して、特定のトピックに関する情報をやりとりする。そして、ユーザが次々に新しいユーザを招待することでユーザ数が増え続けていく。では、このように連鎖的にユーザ数を増やす SNS の内部ではいったい何が起きているのだろうか？

ユーザはどのような人とどのようなコミュニケーションを行っているのだろうか。SNS では、いかなるネットワークの形成原理が働いており、結果的にどのような関係構造が生成されているのであろうか。こういった問題には、情報系の分野だけでなく、社会学においてもまた経営学・経済学の分野においても高い関心が払われている。現実の人間関係の全体像は、誰にも俯瞰できぬいわば「見えない連鎖」であるが、SNS の関係構造はその見えない連鎖を限定的ではあるが、具現化する可能性があることが、これらの関心を導いていると考えられる。

SNS が生み出す関係構造については、湯田ら[湯田 06]が mixi を対象とした研究で、(i)スケールフリー性と高い凝集性、(ii)抽出したクラスタの大きさにおける Zipf 則とギャップの存在を指摘している。ダイアド（二者関係）の承認関係の集積としてのホールネットワークが持つ構造の特徴を記述しており、きわめて興味深い。現実社会に存在するネットワークが持つスケールフリー性を指摘した研究は膨大にあり、ノードも紐帯も多様なネットワークについてスケールフリー性が論じられている。本研究でも同様に、mixi を対象とした分析を行う。特に、社会ネットワーク分析[安田 01]の視点から、友人関係の相互承認によって形成されたネットワークがどのような特徴を持つか、中心性の高いユーザはどのような関係を構築しているか、また、ユーザの関心の共通性を表すコミュニティがどのようなネットワークを構成し、どのような相互作用を行っているのかなどについて分析する。既存の実運用システムを対象とした研究であるので、データ分析と解釈が主な内容であるが、今後、SNS をはじめとするコミュニティ Web プラットフォーム[大向 06]を構築していく上で、ひとつの重要な知見を提供するものである。

以下、2 章では分析に用いたデータについて述べ、3 章から 5 章では実際の分析結果について述べる。6 章で全体をまとめる議論を行い、7 章で関連研究、8 章で結論を述べる。

## 2. 分析に用いたデータ

本分析では、mixi 上で形成されている関係を対象とする。我々は、2005 年 3 月の時点で、

mixi の運営会社である株式会社イー・マーキュリー（現 株式会社ミクシィ）から学術的な目的でデータの提供を受けた。データは、mixi 上での友人関係、コミュニティの参加に関するものであり、個人情報は一切含まれておらず、個別のユーザ ID も特定できないように加工されている。

mixi の機能について簡単に説明する。mixi は多くの SNS がそうであるように招待制をとっており、mixi ユーザからの招待があってはじめて参加することができる。ユーザ登録後、知人を mixi に招待したり、mixi 上で友人を見つけることによって、紐帯が増加する。この知り合いは「マイミクシィ」（略してマイミク）と呼ばれる。ユーザがマイミクシィの登録をすると、自分のマイミクシィ一覧に相手のニックネームと写真が表示される。マイミクシィの日記やアルバム、レビューが自分のトップページに新着情報として表示され、アグリゲータとしての機能を果たすことになる。紐帯の形成は、具体的には、ユーザ A が別のユーザ B に対して、マイミクシィに追加してもらうように依頼のメッセージを出し、ユーザ B が承認すると、ユーザ A、B 双方に相手がマイミクシィとして登録される。この相互承認をもって作られたマイミクシィの関係を、本論文では友人関係と呼ぶことにする。（マイミクシィと一般的な意味での友人関係は異なるものであるが、それについては 6 章で議論する。）

用いたデータは、ユーザ数が 363,819 人、紐帯総数が 3,813,702 である。ほとんどの紐帯は双方向あるが、108 個の紐帯だけは非対称である<sup>6</sup>。そのため、本分析では方向ありの紐帯が約 381 万本存在するとして議論を進める。（湯田らの分析では、これを約 190 万本の方向なしの紐帯としている。）一人当たりの平均友人数は 10.48 人である。

さらに mixi では、ユーザはコミュニティと呼ばれるグループに所属できる。コミュニティは、ユーザが自由に開設でき、設定したテーマに関心のあるユーザが参加して話題を交わすための場となる。コミュニティに参加するとユーザのトップページに所属するコミュニティが表示され、コミュニティの新しい投稿がユーザのトップページに表示される。コミュニティへの参加は自発的なものであり、誰でも参加可能なもの、管理者の承認が必要なものがある。用いたデータでは、コミュニティの数は 90,795 個であり、コミュニティにひとつでも所属しているユーザ数は 241,423 人（全ユーザの 66.4%）であった。以下、3 章では友人関係のデータ、4 章ではコミュニティのデータ、そして 5 章ではその両方のデータを用いた分析を行う。

---

<sup>6</sup> はっきりした理由は不明である。

### 3. 友人関係の構造

#### 3.1 次数分布, $L$ , $C$

友人関係の紐帯の次数分布を示したものが図 1 である.  $x$ 軸は対数変換した友人数(次数),  $y$ 軸は対数変換した確率(該当ノード数を全ノード数で割ったもの)を示している. 全体の 23.6%にあたる 85851 人が友人を 1 人だけしかもっていない. また友人が 2 人以下の者は全体の 35.7%, 3 人以下の者は全体の 44.3%である. 友人数の最大値は 1301 人であるが, mixi では 2004 年 10 月以降, マイミクシィの数を 1000 人に制限している. mixi の友人関係構造がスケールフリー性を持つことは湯田らも確認しているが[湯田 06], この図からも友人数の分布にスケールフリー性を確認できる. ベキ乗の係数  $\gamma$  は 2.4 程度であった<sup>7</sup>.

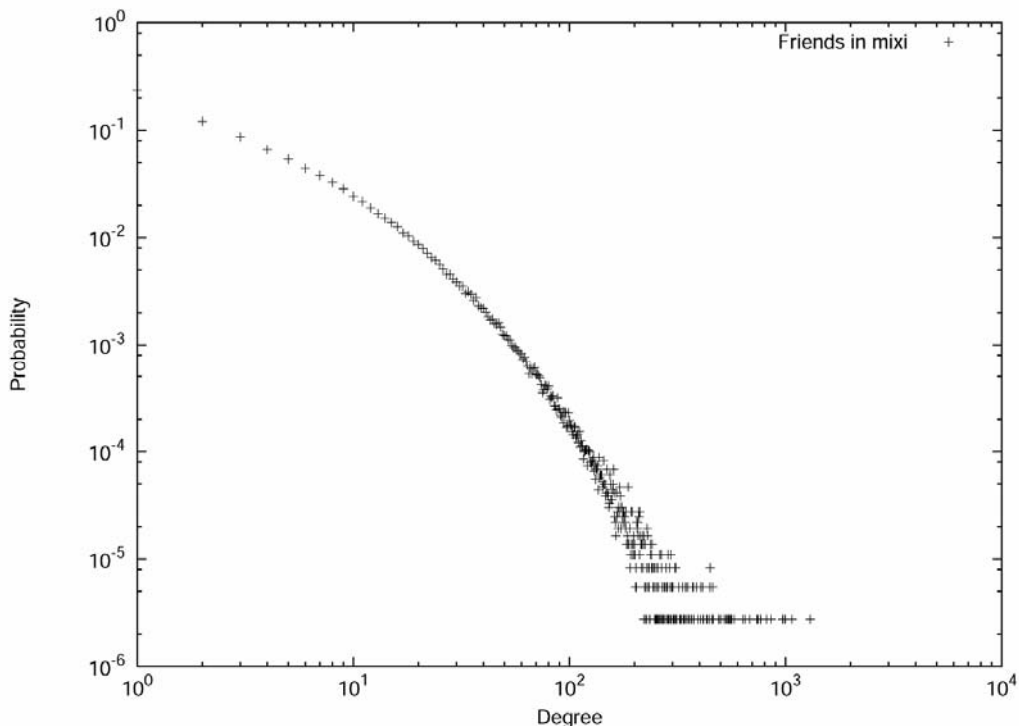


図 1: 次数分布

友人関係のネットワークは, 全体で 1215 の連結コンポネントに分かれており, 最大のものは 360,801 人のユーザから構成される. 2 番目に大きいものでもノード数は 16 であり, ほかのコンポネントはそれ以下である. (ノード数が 2 個のコンポネントが 895 個であった.) これは, ユーザの退会によって孤立したノードであると考えられるが, 全部あわせても全体の 1%以下である. 全体のクラスタ係数  $C$  は 0.328 であり, 最大コンポネントの平均パス長  $L$  は 5.528 であった. この値は, 有名な Milgram の実験における値(5.5)と非常に近い

<sup>7</sup> ただし, グラフのどの部分を用いて回帰するかによって係数は異なる. 湯田らは累積次数分布を用いた上で  $\gamma$  を 2.8 程度としている.

[Milgram67].

用いたデータでは、各ユーザが登録を行った順序関係はおおよそ保存されている。つまり、あるユーザが初期のユーザか新しいユーザかを知ることができる。それを利用して、 $C$ と $L$ の値がユーザ数の増加とともにどのように変化したかを示したのが表 1 である。人数が 10,000 人という行は、初期ユーザ 10,000 人に限った場合のネットワークの諸量を表している。完全にその時点に遡っているわけではなく、初期ユーザ 10,000 人がその 10,000 人内で後から形成した友人関係もあるので、正確な時系列の変化を表しているわけではないが、おおよその概略をつかむことができる。ユーザ数が増えるにしたがって、 $C$ が減少し、 $L$ が増加する。ランダムグラフや Small World では、ノード数  $n$  に対して  $L$  が  $\log(n)$  のオーダーで増加することが知られているが、ここでの  $L$  の増加もこれによく当てはまっており、おおよそ  $L = 0.295 \ln(n) + 1.543$  (決定係数は 0.963) と近似できる。注目すべきは、人数が当初の 2500 人から約 145 倍になってもクラスタ係数は 0.3 を下回らず、一定の局所的凝集性が維持されていることだ。また、同じく平均パス長も、ユーザ数の爆発的増大にもかかわらず、当初の 3.95 からわずか 2 ステップ足らず増えているだけである。局所的凝集性と短いパス長という特徴が、規模が著しく拡大しても保たれている。

表 1: ユーザ数と  $C$ ,  $L$  の変化

人数	$C$	$L$	最大コンポーネント
2,500	0.394	3.951	2,466 人
5,000	0.382	4.082	4,933 人
10,000	0.378	4.279	9,895 人
20,000	0.372	4.454	19,798 人
30,000	0.366	4.524	29,694 人
50,000	0.353	4.634	49,491 人
70,000	0.349	4.729	69,319 人
100,000	0.344	4.847	99,050 人
150,000	0.337	4.998	148,694 人
200,000	0.334	5.124	198,269 人
300,000	0.333	5.335	297,528 人
363,819	0.328	5.528	360,801 人



### 3.2 中心性上位のネットワーク

次に、mixi 内部に多数の友人関係を持つ中心的なユーザのみを抽出し、そのネットワークを調べる。301 人以上の友人を持つユーザはわずか 99 人であり、全体の 0.027% ときわめて少ない。この 99 人は、いわば mixi における著名人であって、立ち上げからのメンバーを含んだネットワークのコアであると予想できる。その友人関係を図示したものが図 2 である。

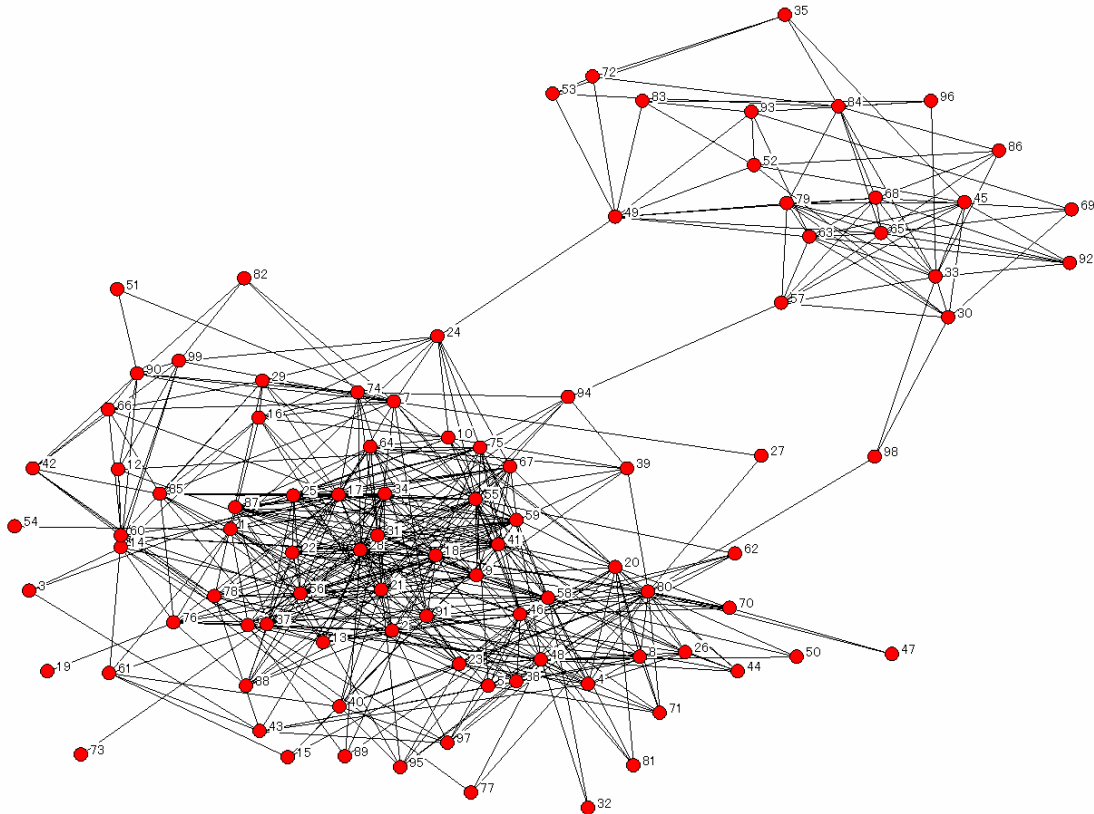


図 2: 300 人以上の友人を持つ中心的なユーザ同士の友人関係

このネットワークでは、内部に大きく 2 つのクリークが存在する。この 2 つのクリークを連結している「橋」の機能を果たすペアが 3 組存在しており、これら 3 つのダイアドが、2 つのクリークを結んでいる。この 99 名が形成するネットワークの  $C$  は 0.382 であり、 $L$  は 2.74 であった。中心的なメンバーの間でもかなり密な関係が形成されていることがわかる。しかし、この凝集性の高さは、2 つのクリークそれぞれの内部での凝集性によるものであり、2 つのクリークの連結は弱く、全体が一つの凝集性を持つグループではないことに留意すべきであろう。

やや範囲を緩めて 201 人以上の友人をもつユーザに限定すると、mixi 内では 365 人、全体の 0.100% である。この 365 人の相互関係のネットワークでは、 $C$  は 0.323 であり  $L$  は 2.877

であった。101人以上の友人を持つユーザは2564人おり、 $C$ は0.233、 $L$ は3.139となる。この範囲でも、大きな2つのクリークの構造は保たれている。理由は定かではないが、mixiには、中心的ユーザが形成する大きな二つのクリーク構造が存在することは興味深い事実である。

### 3.3 コミュニティの構造

友人関係は承認の連鎖として成長していくのに対して、ユーザの関心により成長するのがコミュニティである。多数のユーザの関心を集めたコミュニティは拡大し、また類似した関心を持つユーザは共通のコミュニティに集まる。コミュニティの共通所属は友人関係の契機になりうると同時に、コミュニティ相互の関連性を形成する。

表2: コミュニティのメンバー数上位

	コミュニティ名	人数	中心性*
1	資料になりそうなウェブサイト	10,238人	1
2	Mac ユーザー	8,826人	3
3	まったくわけがわかりません！	7,787人	6
4	Photoshop	7,407人	2
5	クリエイター・デザイナー	7,163人	5
6	iPod User's	7,061人	11
7	面白ネタで笑おう！	7,022人	4
8	美術館・博物館 展示情報	6,824人	13
9	空を見る人	6,581人	7
10	料理作るのが好き	6,467人	15
11	笑える画像	5,828人	12
12	フォント	5,516人	8
13	にゃんこ組	5,342人	38
14	O型	5,337人	27
15	Illustrator	5,309人	9
16	ヴィレッジヴァンガード	5,107人	18
17	名前覚えられません	4,892人	16
18	めんどくさい	4,850人	10
19	カレー大好き	4,849人	22
20	水曜どうでしょう	4,831人	118

\*固有ベクトル中心性による順位

表2は、メンバー数の上位20位までのコミュニティ名及び所属人数を示している。mixi内で最も大きいコミュニティはメンバーが1万人を越え、小さな町の住民程度の規模がある。「資料になりそうなウェブサイト」「Mac ユーザー」といった目的や対象のはっきりしたコミュニティと、「まったくわけがわかりません」「空を見る人」などの漠然とした名前のコミ

コミュニティが上位には混在している.

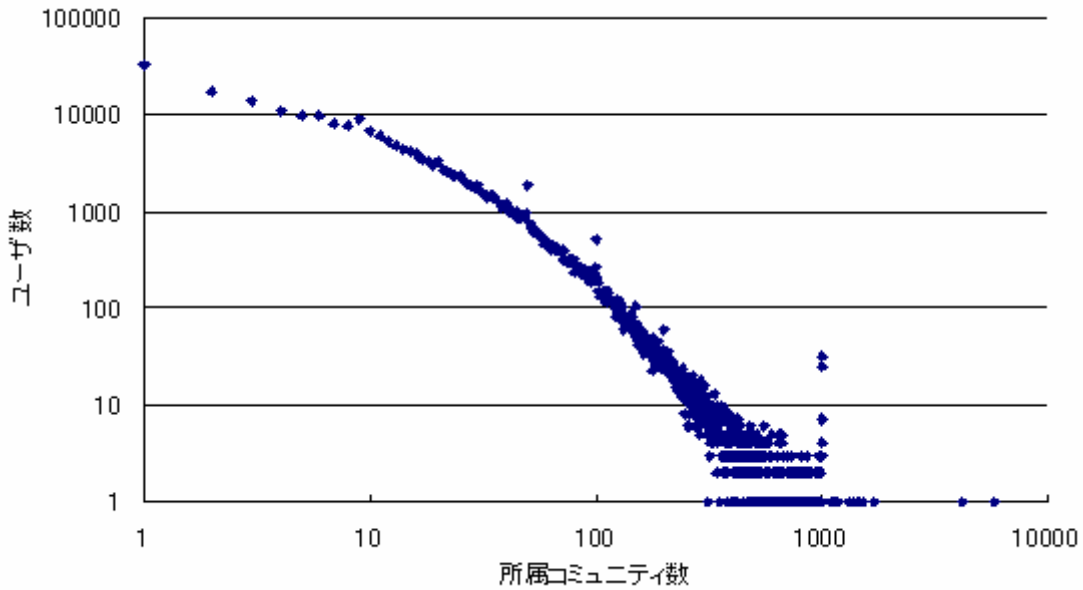


図 3: ユーザの所属コミュニティ数の分布

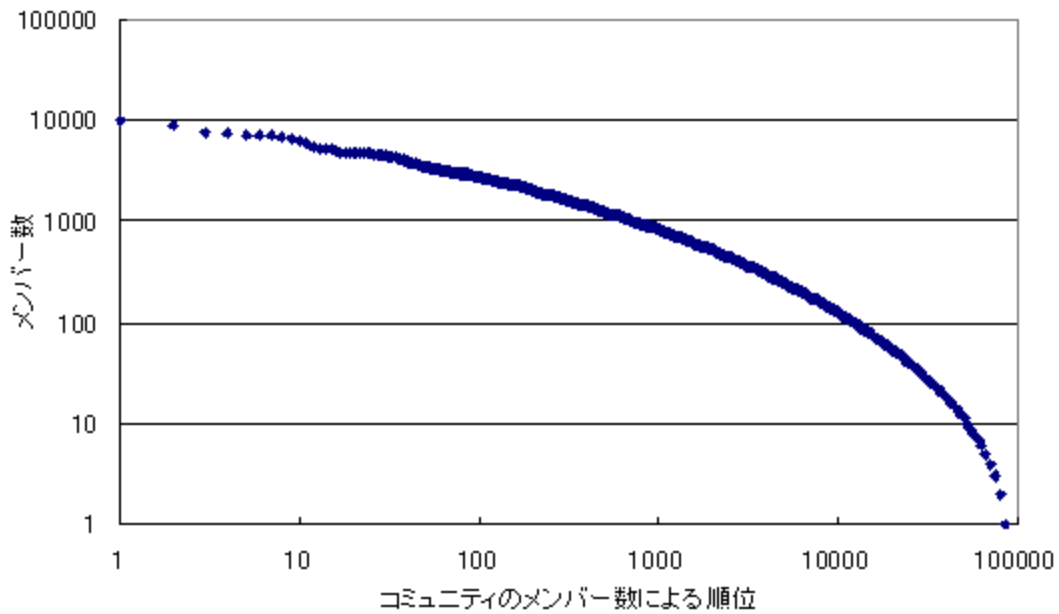


図 4: コミュニティのメンバー数の分布

図 3 は, ユーザがどのくらいの数のコミュニティに所属しているかの分布を示している. 各ユーザの所属コミュニティ数の平均は 24.97 個であり, 所属コミュニティの数が増えるにしたがって人数は少なくなる. 所属コミュニティが数十を超えるあたりからは両対数グラフでほぼ直線状の分布であるが, ところどころ上に飛び出しているプロットが見える. これは

所属コミュニティ数が 50, 100, 200 などのプロットである。特に、所属コミュニティ数が 1000 近辺で大きく上に上がっているが、これは 2004 年 10 月から所属コミュニティの上限を 1000 としているため、所属コミュニティの数が 1000（もしくは 999, 998）のユーザ数が多くなっているためである。

コミュニティのメンバー数の分布を示したものが図 4 である。各コミュニティのメンバー数の平均は 66.40 人であり、所属人数が最大のコミュニティは 10238 人、所属者 1 人だけのコミュニティは 5395 個であった。一部のコミュニティで非常にメンバー数が多く、メンバー数が 1000 を超えるものは全体の 7.9% だけである。大多数のコミュニティはメンバーが数百数十である。しかし、両対数をとったときにこの分布は直線にならないので、Zipf 則ではない。10 人以下のコミュニティは Zipf 則を仮定した場合よりも少なく、すなわちロングテール部分がない。コミュニティという性質上、ある程度の規模までは情報交換しやすく発展しやすいのかもしれない。

分析に用いたデータではコミュニティの時系列の順序関係が保存されており、これを利用してコミュニティの新しさと人数の関係を示したものが図 5 である。ここから、次の 2 点を読み取ることができる。(i)古いコミュニティほどメンバー数が増える傾向があり、その上限は時間によっておおよそ制約されている、(ii)同じ時期のコミュニティでも、メンバー数が非常に多いもの（勝ち組）が存在する。

次にユーザがどのようなコミュニティに所属しているかを使って、コミュニティ間の関連性を見てみよう。対象としたのはメンバー数の上位 200 位までのコミュニティで、メンバー数は最大 10238 人から最小で 2016 人、mixi 全体のコミュニティの延べ人数の 10.6% を構成する。ユーザとコミュニティの所属関係は二分グラフを形成するが、これをコミュニティ間の関係性に縮退させ、ネットワークを抽出する。

つまり、ある二つのコミュニティに共通に所属するユーザが多いほど、その二つのコミュニティの関連性は強いと考える。このネットワークを本論文ではコミュニティ連関ネットワークと呼ぶ。コミュニティ間の関連度は Jaccard 係数を用いて計算する。すなわち、コミュニティ C とコミュニティ D に共通して所属するユーザ数を  $n(C \cap D)$ 、C と D のいずれかに所属するユーザ数を  $n(C \cup D)$  とすると、コミュニティ C と D の関連度は

$$Jaccard(C,D) = n(C \cap D) / n(C \cup D)$$

と計算される。

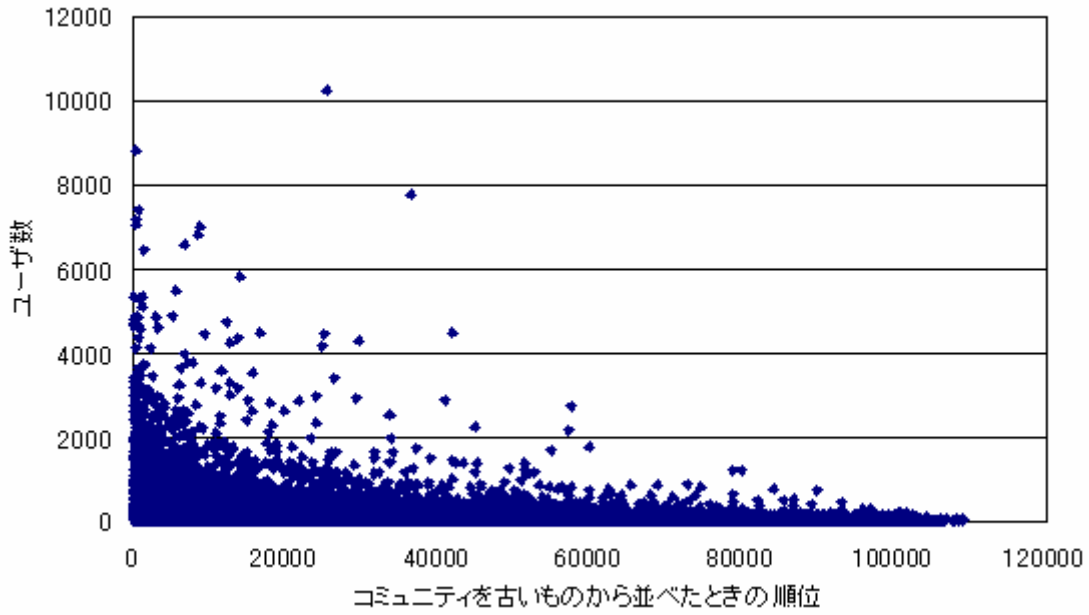


図 5: コミュニティの新しさとメンバー数

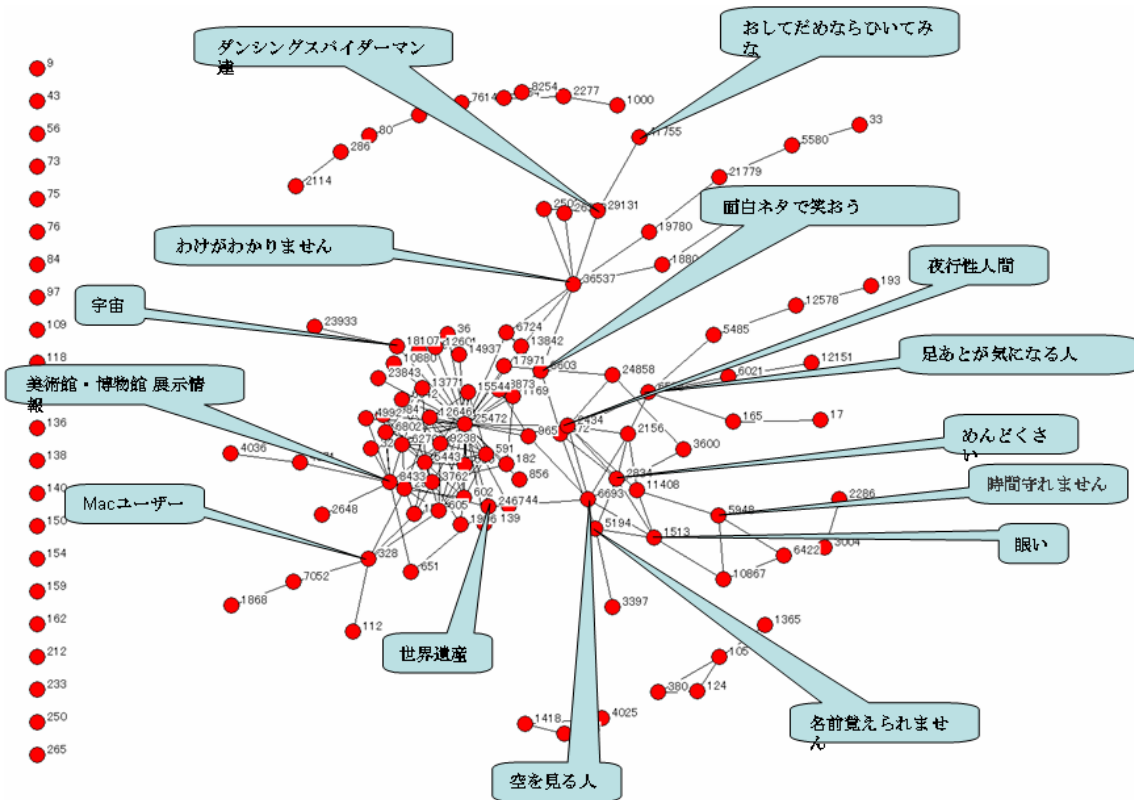


図 6: コミュニティの連関構造

図 6 に、コミュニティの連関ネットワークを示す。Jaccard 係数が 0.2 以上の紐帯のみを表示している。この場合、21 個のコミュニティが孤立点となるが、その他の 179 個のコミュニティは連結したコンポーネントを形成する。図中にいくつかコミュニティ名を挙げているが、「わけがわかりません」から「ダンシングスパイダーマン達」そして「おしてだめならひいてみな」という連関や、「Mac ユーザー」から「Max OS X」そして「PowerBook&iBook」という連関、「恋人いない」から「本当はエロいです。」そして「人生とはエロスで満ちている」など、一般的な話題から次第にテーマが深堀されていく構造を読み取れることができる。

このネットワークでの中心性（固有ベクトル中心性）を表 2 に示している。メンバー数の多いコミュニティは中心性も高い傾向があるが、例えば「にゃんこ組」（中心性 38 位）や「水曜どうでしょう」（中心性 118 位）など、中心性が低くてもメンバー数が上位 20 位までに入っているものもある。連関ネットワークの中心からやや外れたところで、新たにユーザを引き付けつつあるコミュニティである。

さらにこれらのコミュニティのテーマをブロックモデルという手法[安田 01]によりクラスタリングし、ラベル付けしておおまかに分類した結果が図 7 である。8 分割の時点で、共通要素系、共通嗜好系、芸能人系、面白ネタ系、趣味系、TV系、Mac 系、デザイン系の 8 つに分けられる。それぞれがネットワーク図のどのあたりの領域であるかを図 8 に示す。2005 年 3 月の時点でのコミュニティは、このような領域をカバーしながら相互に関連の構造を形作っていたわけである。

1. 共通要素系
  - 酒好き、珈琲党、AB型、B型、関西人
2. 共通嗜好系
  - はてなダイアリー、プログレッシブロック、めがね男子愛好会
3. 芸能人系
  - マシュー南、松本人志、ラーメンズ
4. 面白ネタ系
  - わけわからない、ダンシングスパイダーマン、ゼクシの加藤ローザ
5. 趣味系
  - そうだ京都行こう、GREE、映画愛好会、吉祥寺フリークス、JAZZ、ファミコン
6. TV系
  - たもりクラブ、みうらじゅん、ピタゴラスイッチ
7. Mac系
  - Mac user、ipod user、MAC OSX user
8. デザイン系
  - Web design、Design News、クリエイター&デザイナー

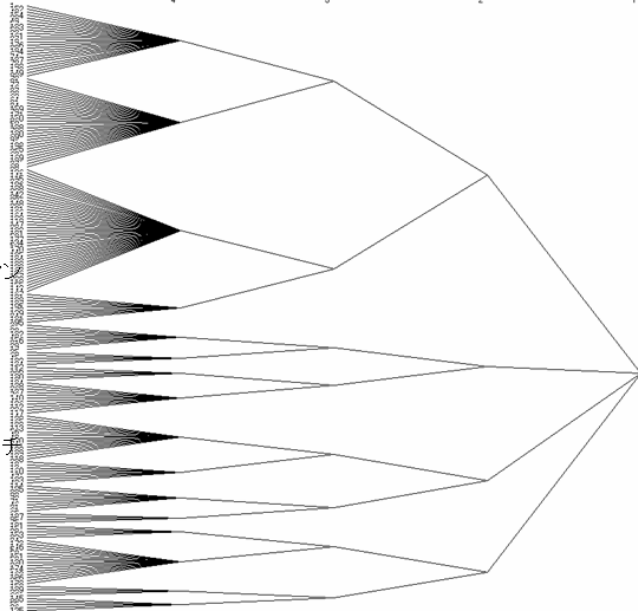


図 7: ブロックモデルによるコミュニティの分類

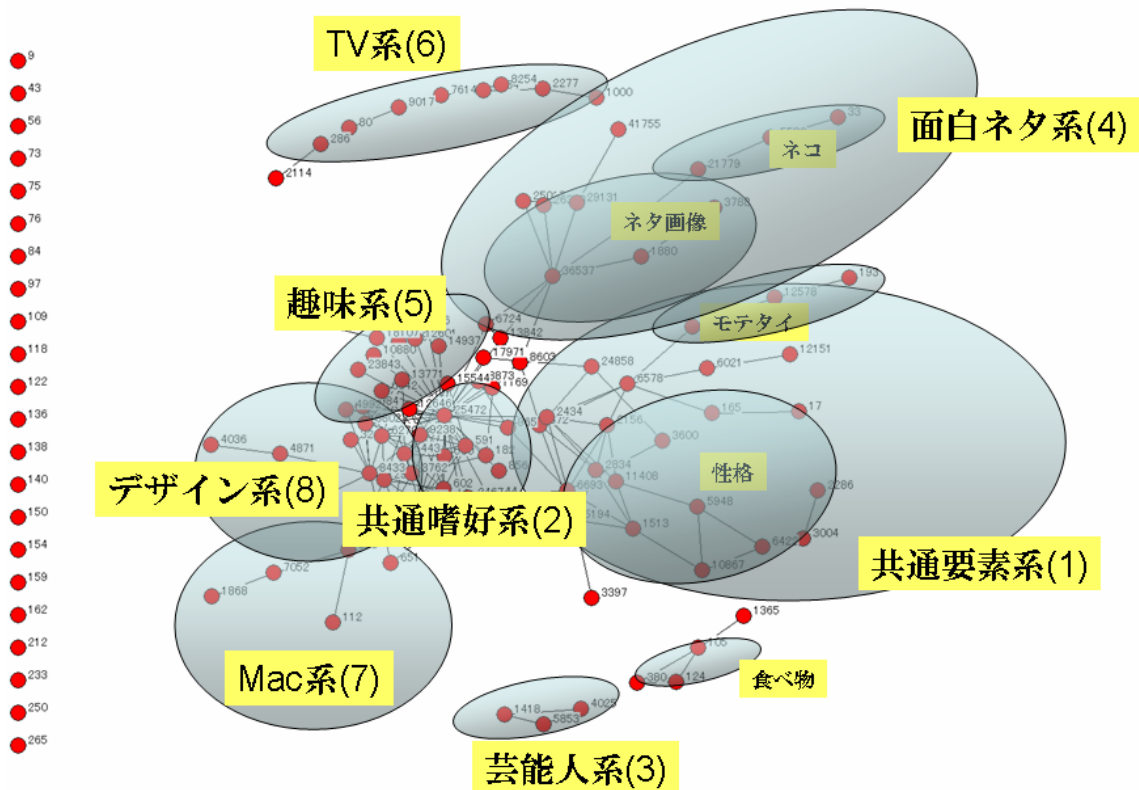


図 8: コミュニティの領域図

#### 4. 友人関係とコミュニティ

mixi では、友人登録とコミュニティへの参加が相互発展的に起こるので、友人関係と共通のコミュニティがあるかどうかはある程度相関している。友人関係のある・なしと、共通のコミュニティのある・なしの分割表を作成したものを表 3 に示す。あまりに全体の数が多いので、友人関係がないー共通のコミュニティがないという負の相関は読み取りにくいですが、友人関係があるー共通のコミュニティがあるという正の相関ははっきりと読み取れる。

表 3: 全ユーザ間の関係における分割表.

	共通コミュあり	なし	計
友人関係あり	$2.04 \times 10^7$	$1.77 \times 10^7$	$3.81 \times 10^7$
なし	$2.80 \times 10^{10}$	$1.30 \times 10^{12}$	$1.32 \times 10^{12}$
計	$2.80 \times 10^{10}$	$1.30 \times 10^{12}$	$1.32 \times 10^{12}$

※ 全ユーザ数  $\times$  (全ユーザ数-1)  $1.32 \times 10^{12}$  である.

では、コミュニティによっては友人関係が多いものとそうでないものがあるのだろうか？  $n$  人から成るコミュニティであれば、最大で  $n(n-1)/2$  組の友人関係が存在し得る。このうち、実際に友人関係が存在する割合を、本論文ではコミュニティの結合性 (CC: community connectedness) と呼ぶことにする。言い換えると、コミュニティに所属するメンバー内でのネットワークの密度である。結合性が 1 であれば、コミュニティ内の全てのメンバーは友人関係であり、0 であればどの 2 人をとっても友人関係でない。

図 9 は、コミュニティのサイズと結合性を調べたものである。コミュニティのサイズが大きくなると一般的に結合性は弱くなる。これは、各ユーザの平均友人数が約 10 人であることから考えると当然であり、結合性はおよそ  $n$  に比例して小さくなる<sup>8</sup>。しかし、同じサイズのコミュニティでも、結合性が大きなものもあれば小さなものもある。図中では、右上に近いものがサイズの割に結合性が高い。

<sup>8</sup> 図 9 の近似曲線は、正確には  $1.098 \log(n) + \log(CC) = 1.005$  であった。



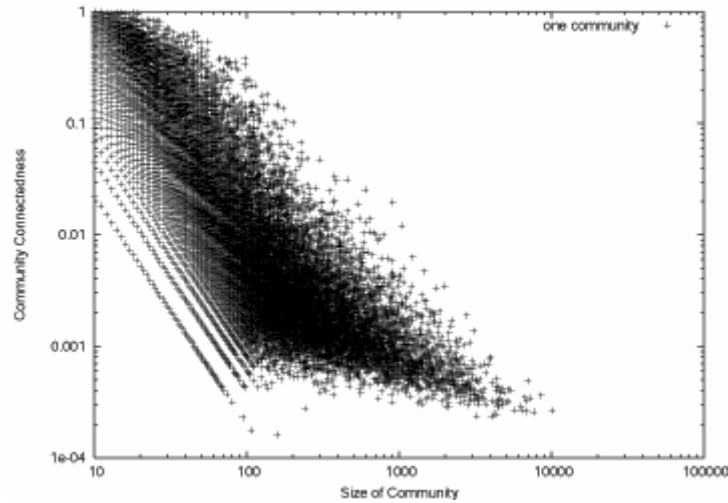


図 9: コミュニティのサイズと内部での結合性

そこで、結合性 $CC$ を $n$ で割った $f=CC/n$ を指標とし、その値の上位 5 位までのコミュニティを示したものが表 4 である<sup>9</sup>。表中の概要を見ると分かるように、これらのコミュニティは何らかの形でリアルな世界でのインタラクションがあるコミュニティであった。6 位以下のものも含め、上位はほぼ完全に何らかの形でオフ会を開催しているコミュニティである。定期的にオフ会を開催し、その感想や次回の予定、その間のコミュニケーションをコミュニティ上で行っているわけである。つまり結合度が高いのは、リアルな空間での関係と mixi 上でのコミュニケーションが融合し、うまくサイクルとなって回っているコミュニティであった。

対照的に、表 5 は指標 $f$ が最も小さいコミュニティである。結合性は上位 5 位と比べて 1000 倍以上の大きな差がある。純粋にユーザの趣味や興味の対象であり、かといって誰しもが共感できるものではなく（しかも興味を持っていること自体で共感を呼ぶほどマニアックでもなく）、友人関係の契機となるのが難しいという印象をもつコミュニティである。

以上見てきたように、コミュニティは、リアルな友人関係を支える結合性の高いコミュニティから、純粋に興味だけでつながっている結合性の弱いコミュニティまで幅広く存在する。もちろん、結合性が高ければよいというものではなく、情報収集という観点からは結合性が低くても十分にその機能を果たしている場合もあるだろうし、弱い紐帯の強さという観点で言えば、結合性が低いコミュニティにこそ貴重な友人関係の可能性があるとと言えるかもしれない。いずれにしても、mixi のコミュニティ機能は、ユーザの多様なニーズに応じてうまく機能しており、結果的に独特の生態系を形作っていると言えるだろう。

表 4: 結合性の強いコミュニティ

<sup>9</sup> コミュニティが比較的小規模であるため、コミュニティ名は示していない。

	コミュニティの概要	人数	結合性
1	Keio SFC のダンス系のコミュニティ	97	0.377
2	鍋と焼肉の集まり	97	0.334
3	割烹のお店の集まり	56	0.539
4	ホームパーティのコミュニティ	84	0.356
5	ある人の家でときどきご飯を食べる会	45	0.656

表 5: 結合性の弱いコミュニティ

	コミュニティの概要	人数	結合性
1	常盤貴子	107	0.000176
2	バイオハザード4 (ゲーム)	94	0.000229
3	カレーうどん	80	0.000316
4	マカロニほうれん荘 (漫画)	158	0.000161
5	鷺沢萌 (作家)	74	0.000370

## 5. 議論

SNS では、友人関係を形成することによるインタラクションの負荷が小さい。社内での職務遂行上の関係や地域での友人関係は、関係構築にも負荷がかかり、関係維持にはさらに継続的な負荷がかかる。そして、関係維持の動機や機会が薄くなったものは淘汰されていく。それに対して mixi では、わずか 2 クリックで関係が形成できるし、ログインしなければ関係維持の負荷も 0 である。また、これまで社会ネットワーク分析で扱われてきた企業の取引関係や産業連関ネットワークと比較して、SNS 上での関係は、経済合理性の影響をほとんど受けていない。企業の取引関係では、取引を行う相手はさまざまな要因を考慮して選ばれ、基本的に企業の利益に寄与しない関係は形成されない。送電網でも神経回路でも、何らかの経済合理性に基づいている。だからこそ、形成されたネットワークには何らかの意味があり、合理性の観点から分析の対象となり得る。したがって、SNS 上でのネットワークは、関係維持の負荷、そして経済合理性という点で、現実のさまざまな社会ネットワークと大きく異なることをまず理解しておく必要がある。

しかし、SNS を日常的に使っているユーザにとって、関係維持の負荷が皆無というわけではない。友人の日記にコメントしたり、友人に自分の所属するコミュニティを見せる（見ら

れる) ことは、多少なりとも社会的な圧力となる。ユーザの所属する集団によっては、友人のメッセージに答えないというのは不評を買うかもしれないし、友人の日記を読んでいないとリアルな世界での友人関係に支障があるかもしれない。したがって、SNS では SNS の世界の中での規範が存在する。それは、自分で友人の数をコントロールしようとしたり、所属するコミュニティを変えたりするなどのユーザの社会的な行為となって現れ、結果的に、他者の存在を意識して抱かれる規範や、各ユーザの目的を反映した社会ネットワークが形成される。この SNS が内部にもつ社会性こそが、SNS のネットワークを他の社会ネットワークと差異づけ、独特の形成原理の要因となるものであろう。

本論文では、mixi 上の「マイミクシィ」を友人関係と呼び替えて議論を進めてきた。しかし、上述の意味では、友人関係とマイミクシィは全く異なるものである。リアルな世界での親しい友人や知り合いを必ずマイミクシィに登録するかというとそうではなく、マイミクシィに登録したからと言って、親しいわけでもない。マイミクシィに登録するのは、自己言及的になるが、マイミクシィに登録してもいいと思うからであって、それは mixi 上の自分のさまざまな行動と関係している。Google で検索するという行為を表す「ググる」という言葉が用いられるように<sup>10</sup>、社会的にも「マイミクに登録する」ということが、ひとつの新しい概念として成立しつつあり、むしろある特定の社会的関係を表す普通名詞として、「マイミク」関係と言うべきなのかもしれない。

本論文では、友人関係とコミュニティに関してさまざまな角度から分析結果を報告した。提供されたデータには個人の属性に関する情報が含まれないなどの制約がある中での分析であり、断定的な知見を述べることは難しいが、本論文の分析結果を総合的に考えると、コミュニティに関して次のような全体像を描くことができるのではないだろうか。

まず表 2 に示したような人気のコミュニティは、誰しもが気軽に参加しやすく共感しやすい、間口の広いコミュニティである。こういったコミュニティは常に多くの人の参加を呼び、時間とともに大きくなる(図 5)。次に、こういった間口の広いコミュニティが、徐々に具体的、マニアックになるコミュニティにつながり、関連コミュニティの領域を形成している(図 6, 図 8)。ユーザは、自分の興味がよりの確に表されたコミュニティを見つけ、そこに遷移していく。コミュニティが特化されるにしたがって人数が少なくなり、数十人程度の小規模コミュニティになると、メンバーの顔が多少は分かり、投稿数・メンバー数とも比較的居心地がよい領域となる(図 4)。時間とともにユーザが増えて規模が大きくなったコミュニティは、さらに特化したコミュニティを生み出す(図 5)。

一方で、友人関係(特にリアルの世界での友人関係)をきっかけにしたコミュニティもあ

<sup>10</sup> Google は、ブランドマネジメントの観点から、この用法に難色を示している。

る。このコミュニティはオフ会が強力な求心力となって、数十人程度の規模まで発展していき、中には数百人の規模に達するものもある（表 4）。人数が増加していくと、オフ会が機能しなくなり求心力を失っていくか、トピック自体に焦点を移した小規模コミュニティに昇華していくことになる（図 9）<sup>11</sup>。

このようなコミュニティの像が正しいかどうかは、今後の mixi やその他の SNS の分析を待たなければならない。しかし、本論文の分析から、SNS におけるユーザやコミュニティの生態系のある側面が垣間見えるのではないだろうか。

## 6. 関連研究

Web 上での社会ネットワークに関する研究は、ここ数年非常に活発に行われており、そのデータも Web のリンク関係から始まり、Email, blog, 名前の共起関係、FOAF データと多様な広がりを見せている[松尾 06]。以下では SNS に特化して関連研究を述べる。

SNS の研究の最も初期のものは、L. Adamic らによるものであろう[Adamic 03]。Stanford 大学の Club Nexus という初期の SNS のデータを用いて、ユーザの属性（学部／大学院や学年、学科、性別、趣味など）と併せたネットワーク分析を行っている。得られたネットワークは 2,469 人で 10,119 本のリンクから成り、 $C$  は 0.17,  $L$  は 4.0 である。表 1 と比較すると、2,500 人のとき  $L$  は同程度であるが  $C$  は mixi の方が高い。日米の違い、時期の違い、データの取り方の違いがあるため断定はできないが、mixi が高い  $C$  の値を保っていることは、友人同士が友人関係になりやすい（もしくは自分の友人を周りから見つけやすい）ということであり、システムの優秀さを示していると考えられる。

Kleinberg を中心とするグループでは、LiveJournal のデータ分析を行っている[Backstrom 06]。特に、グループの機能（mixi でのコミュニティに相当する）に注目し、(i)ユーザがあるグループに入るかどうかの予測モデルを作る、(ii)あるグループが拡大するかどうかの予測をすることを行っている。(i)ではさまざまな属性を用いて決定木によるモデルを構築しているが、特に、自分がそのグループ内に持つ友人が多いほどそのグループに入りやすい、さらにその友人同士が友人関係であるほどそのグループに入りやすいという現象を報告している。本稿でも、コミュニティ内の結合性の幅が非常に広いこと、結合性が高いことがリアルなインタラクションと結びついていることを述べたが、この研究でもグループ内の結合性が重要であるという同様の知見が得られている。(ii)については、ある時点のデータと 4 ヶ

<sup>11</sup> ここで述べた以外にも、mixi では、ユーザの属性を示すためのシンボリックなコミュニティの使い方が多く見られる。コミュニティの写真（画面上でコミュニティ名とともに表示されるもの）の良さといった要素もある。こういったコミュニティは、人気コミュニティであることもあるが、内部でのインタラクションはそもそも重要でなく、友人関係の生成には寄与していないと考えられる。

月後のデータの2つを比較し、グループが拡大するかを予測するには、グループの周辺の人数（つまりグループ内のメンバーと友人関係にあるグループ外の人の数）が重要な属性であり、さらに閉じたトライアドがどのくらい多いかも重要であることを示している。

GoogleではOrkutというSNSのサービスを行っているが、その開発者であるOrkut Buyukkokten本人らはOrkutのデータの分析を行っている[Spertus 05]。この分析では、コミュニティの推薦を対象としており、どのような尺度でコミュニティを推薦すればユーザの受率率が最も高いかを述べている。本論文では、コミュニティの関連度を測るためにJaccard係数という指標を用いたが、この論文では、L1 ノルム、L2 ノルム（コサイン距離）、相互情報量、Salton (*tfidf*的に重み付ける方法)、log-oddsの各指標を比較している。その結果、L2 ノルムがコミュニティの推薦に最も適していることを示した<sup>12</sup>。

mixiについては、本論文でもたびたび触れたように湯田らの研究がある[湯田 05, 湯田 06]。湯田らの研究では、友人関係のネットワークに対して、Girvan-Newman クラスタリングと呼ばれるクラスタリング方法を用いた分析を行っており、クラスタのサイズの分布に興味深いギャップがあると報告している。本論文では、より包括的なネットワーク分析を行っており、湯田らの分析と比較すると、コミュニティの分析により重点が置かれた形になっている。なお、mixi のデータ分析とともにシミュレーションの適用可能性について述べた報告[森 05]もある。

## 7. まとめ

本研究では、mixi のデータを対象とし、友人関係、コミュニティ関係の両面から分析を行った。友人関係については、全体の次数分布やネットワークの諸量の分析、中心性上位のユーザのネットワークの図示を行い、コミュニティについては、全体の統計的な分析、メンバー上位のコミュニティのネットワーク図の図示、そしてコミュニティ内での結合性の分析を行った。

本論文で明らかにしたように、mixi のネットワークは、ユーザ数が36万人の時点でクラスタ係数  $C$  が0.328、平均パス長  $L$  が5.53であった。 $C$ は高く  $L$ は小さく、この値はSNSとしての優秀さ、すなわち友人の局所的連結機能と、効率の高い情報伝播機能を示す値ではないかと思う。また、コミュニティと友人関係は相互の関連を持っており、その両者を考慮することで、ユーザが自分の友人から友人関係を広げていること、結合性の高いコミュニティではオフ会が重要な役割を果たしていることなどを示した。

<sup>12</sup>本論文では、Web上での社会ネットワークに関する他の文献（[Kautz 97, Mika 05]等）でJaccard係数が使われることが多いことからJaccard係数を用いている。

mixi では、コミュニティの検索や人の検索は、(技術的には) 比較的単純な機能に限られている。もちろん、プライバシーを守るという実際的な問題のためであると思うが、この検索機能が限られていることが、逆に、友人関係のつながりによる友人の探索を促し、またコミュニティの連関を生み出し、興味深いネットワークを作り出しているとも見られる。

このように、システムのデザインがユーザの挙動に与える影響は少なくないはずであり、こうした影響についてより一般的な知見を得ることは、SNS をはじめとするコミュニティ Web プラットフォームで今後ますます重要になる課題であろう。今回の分析は、mixi というひとつのシステムのある時点のデータを扱ったものであるが、こういった分析を学術的な知見として積み重ねていくことで、将来的に普遍性の高い有用な知見・発見につながれば著者らの幸いとするところである。

## 参考文献

- [Adamic 03] Adamic, L., Buyukkokten, O., and Adar, E.: A Social Network Caught in the Web, Vol. 8, No. 6 (2003)
- [Backstrom 06] Backstrom, L., Huttenlocher, D., Lan, X., and Kleinberg, J.: Group formation in large social networks: Membership, Growth, and Evolution, in *Proc. SIGKDD'06* (2006)
- [Kautz 97] Kautz, H., Selman, B., and Shah, M.: Referral Web: Combining Social Networks and Collaborative Filtering, *Communications of the ACM*, Vol. 40, No. 3, pp. 63--65 (1997)
- [Matsuo 03] Matsuo, Y., Ohsawa, Y., and Ishizuka, M.: Average-clicks: A New Measure of Distance on the World Wide Web, *Journal of Intelligent Information Systems*, Vol. 20, No. 1, pp. 51--62 (2003)
- [松尾 06] 松尾 豊 : Web2.0 時代の個人とコラボレーション, *情報処理*, Vol. 47, No. 11 (2006)
- [Mika 05] Mika, P.: Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks, {*Journal of Web Semantics*}, Vol. 3, No. 2 (2005)
- [Milgram 67] Milgram, S.: The small-world problem, *Psychology Today*, Vol. 2, pp. 60--67 (1967)
- [Mori 05] Mori, J., Ishizuka, M., Sugiyama, T., and Matsuo, Y.: Real-world Oriented Information Sharing Using Social Networks, in *Proc. ACM GROUP'05* (2005)
- [森 05] 森 祐治, 内田 誠, 柴田 尚樹 : 社会ネットワークサービスを基盤とした情報流通シミュレーションに向けて, *Web が生み出す関係構造と社会ネットワーク分析ワークショップ* (2005)
- [大向 06] 大向 一輝, 松尾 豊, 松村 真宏, 武田 英明 : Community Web プラットフォーム, *人工知能学会論文誌*, Vol. 3, No. 21, pp. 251--256 (2006)
- [Spertus 05] Spertus, E., Sahami, M., and Buyukkokten, O.: Evaluating similarity measures: a

large-scale study in the orkut social network, in *Proc. SIGKDD 2005* (2005)

[安田 01] 安田 雪：実践ネットワーク分析，新曜社 (2001)

[沼 04] 沼 晃介，大向 一輝，濱崎 雅弘，武田 英明：Weblog におけるエゴセントリック検索の提案と実装，第 6 回人工知能学会セマンティックウェブとオントロジー研究会 (2004)

[湯田 05] 湯田 聡夫，藤原 義久：SNS における人のネットワーク構造 その地平線の超え方，Web が生み出す関係構造と社会ネットワーク分析ワークショップ (2005)

[湯田 06] 湯田 聡夫，小野 直亮，藤原 義久：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，情報処理学会論文誌，Vol. 47, No. 3, pp. 865--874 (2006)