

Web2.0の 現在と展望



5. Web2.0時代の個人 とコラボレーション

■ 松尾 豊 産業技術総合研究所 / GBRC 社会ネットワーク研究所 /
スタンフォード大学

Web2.0は、個人の活動の集積として価値あるコンテンツが生成されるという思想が前提となっている。これらの環境の中で、個々のユーザはどのように振る舞い、他者との関係性を構築しているのだろうか？ 本稿では、ユーザの関係性とネットワークという視点から、Web2.0の代表的な例であるブログやSNS、ソーシャルブックマークにおけるユーザのネットワークに関する研究動向を述べる。社会ネットワーク分析や複雑ネットワークという研究分野と関連して多くの研究が行われているが、こういった研究は、Web2.0を理解し今後の展望を考える上で、サービスの提供者にとってもユーザにとっても、重要な知見を提供するものである。

Web2.0と個人

「Web2.0」という言葉が厳密にどう定義されるか定まっているわけではないが、Tim O'ReillyはWeb2.0を特徴付ける7つの概念を挙げている。Webをプラットフォームとして、さまざまなデータの共有とその仕組みの進歩により、結果的にユーザ体験が向上し、集合知が形成される。膨大なブログのエントリが日々生産され、Wikipediaやソーシャルブックマークが日々、その質と量を向上させている。このような「集合知」の形成がマクロな現象だとすると、ミクロな現象、つまり個々のユーザの行為はどのように理解され得るのだろうか？

Web1.0といわれる世界では、各ユーザが各サイト（アプリケーション）と1対1の関係であった。ユーザは、そのサイトを気に入れば、そこに執着したわけであるが、Web2.0では、ほかのユーザの存在がそのユーザに影響を与えるようになった。Wikipediaでもソーシャルブッ

クマークやブログでも、そしてより直接的なソーシャルネットワークサービス（SNS）でも、ユーザは他者の存在を意識的にせよ無意識的にせよ理解し、影響を受けあって、コンテンツを生成している。これは一種のコラボレーションだと考えることができる。Wikipediaで見ず知らずの他人と共同で執筆するのは分かりやすいコラボレーションであるし、またブログに日記を書いてコメントがつく、それがうれしくて、また書こうと思うといった循環のプロセスも1つのコラボレーションである。

このようなコラボレーションを可能にするのは、ユーザ同士の関係性である。それはリアルワールドの友人関係や知り合い関係がWebに持ち込まれたものかもしれないし、Web上でのインタラクションを通じて新たに形成された関係かもしれない。しかし、こうした関係に基づくコミュニケーションがあるからこそ、人は日々、SNSを使ったりブログを書くわけであるし、見知らぬ他者の存在があるからこそソーシャルブックマークやWikipediaは面白いのだろう。結果的に、ユーザ同士の

関係性が、全体として大規模なユーザのネットワークを形成する。ユーザ個人を取り巻くネットワークの存在とその脈動が、Web2.0をユーザサイドから支えているというのが本稿の主題である。

本稿ではユーザの関係性とネットワークに着目する。Web1.0的な文脈では、ユーザの満足度を向上させるには、システムのアルゴリズムやユーザインタフェースの研究が特に重要であった。しかし、Web2.0という仕組みの中では、ユーザを個人として扱うのではなく、他者との関係性を持った個人、ひいてはコミュニティとして扱う必要がある。その研究は必然的に、社会学や現象論としての言語学と接点を持ち始める。本稿でも、社会ネットワーク分析という社会学の分野の知見をベースに話を進めていく。

本稿では、Web2.0時代の個人とコラボレーションと題し、ユーザのネットワークの分析とモデル化に焦点をあてて研究の動向を紹介する。まず、社会ネットワーク分析と複雑ネットワークという研究分野の背景を簡単に説明した後、SNSやブログなどWeb上でのユーザのネットワークを分析する研究を紹介する。さらに、こういったネットワーク上で情報がどう流通するのかについてAmazonやブログサイトでの分析を紹介する。また、SNS上でコミュニティがどう形成されるのか、さらに近年盛んに行われているソーシャルブックマークの分析について述べる。最後に、検索エンジンを用いてさまざまな関係性を抽出する研究を述べ、ユーザの関係性の果たす重要な役割と今後の方向性について述べる。

社会ネットワーク分析と 複雑ネットワーク

本章ではまず、Web2.0をユーザのネットワークという視点から考えるために、基礎となる研究分野について、簡単に概説する。

数年前から、スケールフリーやスモールワールドなどで知られる複雑ネットワーク (complex network) が着目を集めている。ここでいうネットワークとは、ノード (頂点) とエッジ (またはリンク、辺) からなるものであり、たとえば人の関係や都市の交通網などの現実の対象を抽象化したものである。1998年のD. Wattsによるスモールワールドの論文¹⁾をきっかけに、その後、海外ではBarabásiのスケールフリーネットワークを筆頭とするさまざまな研究が活発に行われ、多数の論文がNatureやScienceをはじめとする一流のジャーナルを賑わせた。スモールワールドやスケールフリーの話題は、いくつかの翻訳書に詳しい²⁾が、簡単に説明すると、スモールワールドネットワークは、典型的には小さなクラスタが

少数のリンクでつながれた形をしている (ここでのクラスタとは、ネットワーク中でリンクがたくさん張られた「濃い部分」という意味である)。

スモールワールドは、クラスタ係数 C と平均パス長 L という2つの指標を使って次のように定義される。

- C : 自分の友だち同士が友だちである確率。ネットワーク中のノード v が k_v 個のノードと隣接しているとき、 k_v 個のノード間に存在し得る $k_v C_2 = k_v (k_v - 1) / 2$ 本のエッジに対して、実際に存在するエッジの割合を C_v とする。すべてのノード v について C_v の平均をとったものが C である。
- L : ネットワーク中のすべてのノードの組についての最短パスの長さの平均

スモールワールドは、ランダムなネットワークと比べて C が大きいにもかかわらず L が小さいグラフである。つまり、身近なところでクラスタになっているのに、他の人と短いパスでつながれている。

これと同じような特徴を持つのが、たくさんのリンクを持つ「スーパーノード」が存在するスケールフリーネットワークである。スケールフリーネットワークは、ノードの次数 k (いくつのエッジを持っているか) の分布がべき則 ($P(k) \sim k^{-\gamma}$, γ は定数) に従うというもので、極端に次数の大きいノードが少数存在するが、ほとんどは次数の小さいノードである。

複雑ネットワークに関する研究は主に海外で活発に研究されていたが、国内でも、いくつかの翻訳書が出版されたことで理解が広まり、昨年ごろから急速に活動が盛んになっている。情報処理学会の「ネットワーク生態学研究グループ」、日本ソフトウェア科学会の「ネットワークが創発する知能研究会」などの研究会が立ち上がり、研究の機運が盛り上がっている。

一方で、現実世界に存在するさまざまなネットワークに関する研究は、社会学の分野で古くから行われてきた。社会学では、1930年代から人の関係性を観察しネットワークとして描き分析する手法が提案・洗練されており、社会ネットワーク分析と呼ばれている^{3), 4)}。たとえば、組織内や地域の人の関係性、企業間の関係、産業の連関等、さまざまな対象をネットワークとして捉え、その中でキーとなるプレイヤーは誰か (中心性の分析)、競合にあるのは誰と誰か (構造同値)、誰が効率的にネットワークを張っているのか (structural holes)、どういったグループがあり他のグループとどういう関係を構築しているのか (ブロックモデルなどのクラスタ分析) などの分析手法がある。PageRankはGoogleの基礎的なアルゴリズムとして有名であるが、これは固有ベクトル中心性、

Bonacich 中心性として社会学では以前から知られている概念であった。

ネットワーク分析では、大きく2つのタイプのネットワークデータを扱う(図-1)。1つはノード同士の直接的な関係による隣接行列(adjacent matrix)で表される。つまり、ノード*i*とノード*j*に関係があれば $a_{ij}=1$ 、そうでなければ $a_{ij}=0$ とした行列 $A=\{a_{ij}\}$ で表される。実数値、方向ありなどの拡張ができ、距離が2の関係にあるノード同士を表す行列は A^2 、距離3にあるものは A^3 と簡単に計算することができる。もう1つは、ノードのグループへの所属を表す行列(affiliation matrix)で、ノード*i*がグループ*j*に属していれば $r_{ij}=1$ 、そうでなければ $r_{ij}=0$ とするものである。同一のグループに所属していれば2つのノード間に関係があると解釈することでネットワークとなる。たとえば、企業の取締役は兼務される場合があるが、同じ企業の取締役に所属している関係で人の関係のネットワークを、また同一の人が取締役にいるから企業の関係のネットワークを出すことができる。すなわち、 $R=\{r_{ij}\}$ を転置した行列 R^T を用い、 RR^T でアクタの関係を表す行列が、 $R^T R$ でグループの関係を表す行列となる。こうして得られたネットワークをアフィリエイトネットワークと呼ぶ。

こうした2種類のネットワークを用いると、SNSはもちろん、ブログ、ソーシャルブックマーク、AmazonのユーザやWikipediaのユーザなど、Web2.0のサービス上にかかわるさまざまなユーザをネットワークとして捉えることができる。

ユーザのつながりの分析

本章では、具体的にWeb2.0にかかわるユーザのつながりの研究を見ていこう。

2002年という比較的早い時期に行われたのが、L. AdamicによるSNSの研究である^{5)☆1}。スタンフォードの学生2,470人に対して、Nexusというシステム(SNSの一種)の約1万本の知り合い関係について分析を行った。その結果、ネットワーク上の距離の平均*L*は4.0、クラスタ係数*C*は0.17(ランダムより40倍高い)で、スモールワールドの特徴を備えていることが分かった。この分析では個人のプロフィールとも合わせた分析を行っており、同じ特徴を持つ人がコミュニケーションしやすいなどの結果が報告されている。

日本では、2005年2月時点のデータを用いて、mixiの分析を湯田ら、森ら、安田らが行っている^{☆2}。その

☆1 後にOrkutを作るO. Buyukkoktenも共著者である。
 ☆2 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、社会情報学フェア(2005)。

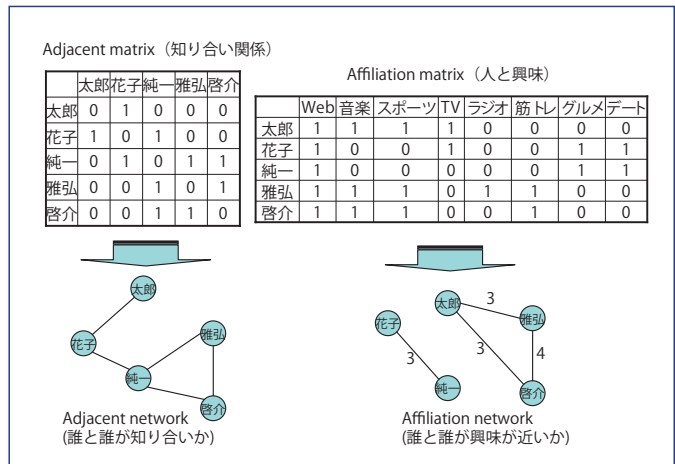


図-1 社会ネットワーク分析

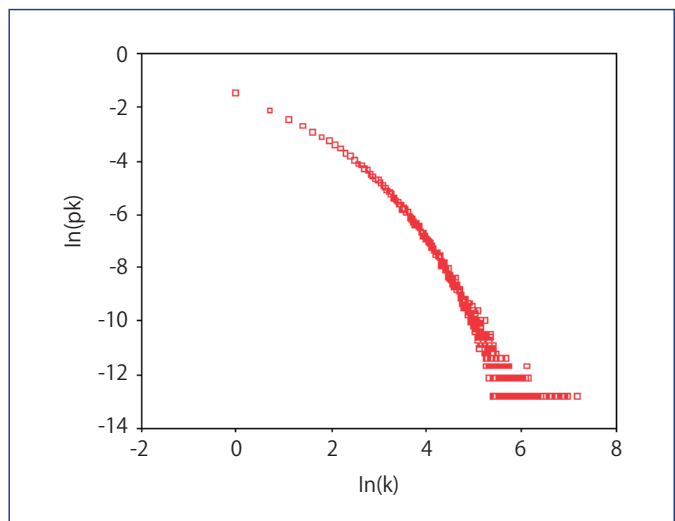


図-2 mixi ネットワークにおける度数分布：横軸は次数、縦軸は全体ユーザにおける割合で、両対数でプロットしている。

後のmixiの成長を考えると早い時期ではあったが、当時の36万ノード、190万リンクについて調査し、知り合いの数が $\gamma=2.80$ のべき分布(図-2)であること、次数平均(マイミクの数の平均)が10.4であり6ホップで96%をカバーする小さな世界であること、クラスタ係数*C*が0.328であり凝集性の高いネットワークであることなどが報告されている。また、湯田らは知り合い関係をGNアルゴリズムという方法でクラスタ化していくと、比較的小規模のクラスタ群と大規模のクラスタ群に二分され、その中間領域が欠けていることを興味深い発見として述べている⁶⁾。mixi上では、自分の周りのクラスタのサイズが徐々に成長していくが、あるときに急激に成長がスキップするわけである。この一般性や含意についてはまだ不明な点も多いが、SNSの何かの性質を示しているものかもしれない。

Web2.0としばしば対立する概念と捉えられているセマンティックWebでも、人の関係性を扱う技術は注目

されている。FOAF (Friend Of A Friend) という人に関する情報を記述する語彙が2003年ごろから整備されており、それが徐々に広まってきている。FOAFでは、自分の名前や連絡先、興味といったプロフィール情報をRDF形式で記述することができ、特に *knows* というプロパティを用いて知り合い情報も記述することができる。LiveJournal や Livedoor ブログ、はてななど、FOAFによる人のメタデータを出力するサイトも多い。自分の知り合いがFOAF文書には記述されているわけであるから、この情報を収集すれば人のネットワークを取り出すことができると考えられる。それを実際にやっているのが、Maryland大のT. Fininらの研究室である。そこで研究開発しているセマンティックWebの検索エンジンSWOOGLEを利用し、FOAFの収集と分析を行っている⁷⁾。2005年の時点で、26,788人の間の15,630個の *knows* の関係を調査した。多くのユーザは孤立しており、小さなクラスタ(842個)に分かれているが、最も大きなクラスタは7,111人であった。次数分布は出次数、入次数ともにべき分布であり、次数が上位であるのは社会的なオーソリティやセマンティックWebの専門家等であると報告している。FOAFファイルは、まだ量的には十分な数があるとはいえないが、その分散性はWeb2.0的であって、今後はユーザのプロファイル管理の仕組みの成長とともに重要性を増してくると考えられる。

さて、ここで紹介した研究は、いずれもWeb上での人のネットワークのスケールフリー性を示すものであった。これが意味するところは何であろうか？ リアルワールドでは、人が1日に使える時間は決まっているので、知り合いの数にも限度があり、知り合い関係はスケールフリーネットワークになり得ない。しかし、Webの場合にはそうではない。Webにおける「知り合い」の基準が人によってまちまちであるのは1つの問題であるが、少なくとも誰かから参考にされている、もしくは情報の流通があるという点でいうと、非常に多くの次数を持つ「スーパーノード」が存在する。一方で、ほとんどの人はそれほど多くのエッジを持たず、その分布はロングテールとなる。しかし、ここで重要なのは、多くの人から支持される一部の人だけではなく、ロングテールの部分の人同士のコミュニケーションを促進する仕組みがSNSやブログにはあるということである。個々のユーザにとって、一部のスーパーノードの質の高い情報も価値があるが、それにも劣らず自分の周りにいる人の日々の雑多な情報も価値がある。Web2.0で議論されるロングテールは、スケールフリーネットワークとは本来は直接のつながりはないが、実は、Webにおけるユーザのネットワークを間に介することで密接に関係している。

情報の伝播モデル：クチコミとブログ

ユーザのネットワークがあるとして、その上で情報はどのように伝播していくのだろうか？ この分析に関する研究をいくつか紹介しよう。

B. Hubermanらは、Amazon.comでの本やDVDの商品の推薦がどのように伝播していくかを分析している⁸⁾。Amazonでは、推薦すると10%値引きされる(さらに推薦したほうにもクレジットが戻る)仕組みがあつて、ユーザには推薦のインセンティブがあるのだが、分析の結果、次のようなことが分かった。

- 2人の間でインタラクションが多くなると、推薦は効かなくなる。
- 推薦を受け入れる確率は、推薦してくれる人の数が増えると急激に増えるがすぐに飽和する。
- 次数の高いスーパーノードがあるが、影響力には限界がある。たくさん推薦する人のことはあまり聞かなくなるからである。
- 推薦の効果は、カテゴリや値段に影響される。

ここで描き出しているのは、ある少数のノードにより全体が影響されるモデルではなく、ネットワークとしてつながれた個々が互いに影響しあいながら、情報が広がっていくモデルである。

また、Richardsonらは、1人のユーザがほかのユーザの購買にどのくらい影響を持つかを数値化し、そのユーザの“network value”を計算する確率モデルを提案している⁹⁾。商品のレビューサイトであるEpinionsのデータを使った結果を示している。この論文が出た2002年当時はまだ、一部のサイトだけで可能な技術であったが、ユーザ間のさまざまな関係が得られる現在では、こういった「ユーザの価値を測定する」手法は適用可能性が広がっている。日本では、化粧品のクチコミサイト@cosmeを対象にした研究が行われている¹⁰⁾。

クチコミといえば、ブログが着目された初期から、ブログによってクチコミマーケティングが可能になるといった話がよく聞かれた。アカデミックの分野でも、2004年ごろからブログの分析は活発に研究されており、研究トピックとして確立し、もはや一段落した感がある。何度かのワークショップの開催を経て、2007年には第1回の国際会議^{☆3}も開催される。2006年からは、情報検索の分野で中心的なTREC(米国NISTが主催)で、ブログのトラックが開始された。ブログの研究としては、

^{☆3} International Conference on Weblogs and Social Media. 2007年3月26～28日、米国コロラド州ボルダー。

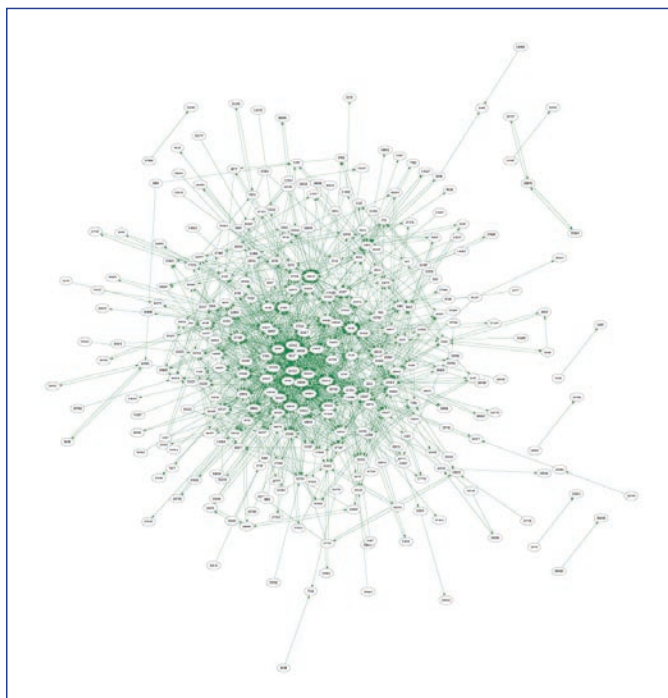


図-3 doblogにおけるユーザのネットワーク：20回以上ブログを訪問しているユーザ同士の関係

たとえば、トピックのバーストの発見、影響力の高いブログのランキングアルゴリズムなどがあり、最近ではブログを書いたユーザの性別や年齢、投稿時の感情などを推測するプロファイリングの研究が盛んである。

ブログが形成するネットワークに関して、E. Adarらのブログ上での情報の伝播の研究が有名である¹¹⁾。ある人が他の人のブログで面白い記事を見たとする。たとえばGiant Microbesというウィルスのぬいぐるみに関する記事であったとしよう。もしこれを読んだユーザがとても気に入れば、自分のブログにも書くかもしれない。それを見た人がまた気に入れば、またブログに書くかもしれない。こうしたブログにおける情報（ここではサイトへのURL）の広まりについて、Adarらは約37,000のブログについて調査を行った。7割以上のエントリが情報の引用元のURLを明示していないが、2つのエントリのテキストの類似度や他へのリンク、時間の情報から、どちらがどちらを参考にしたかを予測するモデルを作った。SVMによる分類で、91%の精度で予測できると述べられている。

また、古川らも同様にブログ上での情報の伝播を、日本のブログホスティングサービスであるdoblogのデータを対象に分析している¹²⁾。図-3は、doblog内でどのユーザがどのユーザのブログを定期的に見ているかの一部を示したものであり、全体として大きなユーザのネットワークを形成する。こうしたユーザ相互の関係が、ブログにおける情報伝播を生み出している。

いずれの研究でも、モデル化の基礎となっているのは、ユーザが情報を得て、それによって他の人に情報を伝播させる力を持つという状況である。Web2.0の世界では、こうした情報の伝播がいたるところで起こっており、ブログや検索エンジン、RSSの規格やRSSリーダーといったそれを効率化する仕組みと技術がこの原動力になっている。こうした情報の伝播の性質が今後の研究でさらに明らかにされれば、ユーザにとってより心地の良い、効率的な情報環境の構築につながっていくと考えられる。

コミュニティの形成

SNSは「コミュニティ」の機能を持っているものが多い。前章まではユーザのネットワークを出して、その濃い部分をクラスターと捉えたわけであるが、知り合い関係でなくとも、「コミュニティ」としてあるトピックを立ててそれに人が集まることが可能である。リアルワールドの関係にとどまらず、Web上で新しい関係を構築するには重要な場所である。

最近の研究では、Live Journalのコミュニティ機能について分析したものがある¹³⁾。ユーザは、自分が入るコミュニティをどう決めているのだろうか？ それを予測するモデルを学習した結果、

- 自分の知り合いの中で、そのコミュニティにすでに入っている数が多ければ、ユーザがそのコミュニティに入る確率が高まる

ことが分かった。これは、ほとんど自明である。面白いのは、

- そのコミュニティにすでに入っている知り合い同士が知り合いであると、そのコミュニティに入る確率が高まる

というものである。自分の知り合いのうち2人があるコミュニティに入っているとすると、その2人が知り合いでない場合より知り合い同士の場合の方が、そのコミュニティに引き込まれやすいわけである。

安田らは、mixiの分析の中で、コミュニティに着目した分析を行っている¹⁴⁾。たとえば、AとBの2つのコミュニティがあつて、それぞれのメンバを U_A 、 U_B とする。このとき、AとBの類似性は、共通するユーザの数を指標化したもの、たとえば、Jaccard係数

$$Jaccard(A,B) = \frac{|U_A \cap U_B|}{|U_A \cup U_B|}$$

により求めることができる。Jaccard係数によってコミュニティのネットワークを作ったのが図-4である。コミュニティ間のつながりが分かり、徐々にマニャックなコミュニティを形成する過程が分かる。ユーザはコミュニティが巨大になってくると、よりマニャックなコミュニ

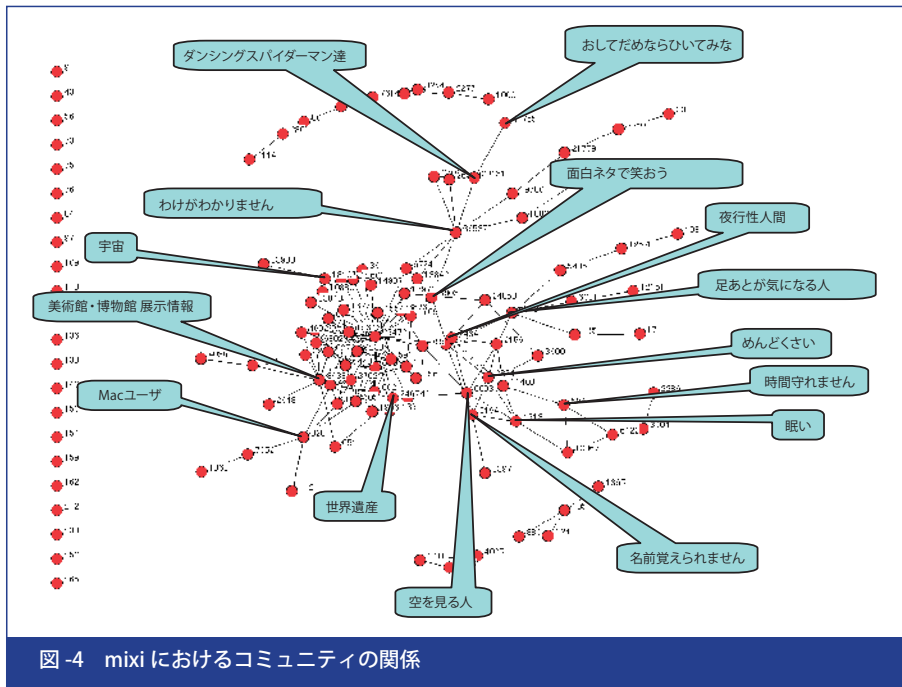


図-4 mixiにおけるコミュニティの関係

ティを作る。結果として、入り口の役割を果たす巨大なコミュニティと、そこから先の徐々にマニアックになる系列コミュニティという生態系が形成されることを示唆している。

Googleの研究者らは、OrkutというSNSにおけるコミュニティの推薦について調べている¹⁵⁾。特定のコミュニティに対して、どういうコミュニティを薦めればユーザは受け入れられるかというものである。上記のJaccard係数と同様、メンバの重なり度の指標(L1, L2ノルム, 相互情報量, IDF, 対数オッズなど)を、400万のコミュニティ推薦、それに対する90万のクリックについて調べ、結果的にL2ノルムが最も良い値であったとしている。こうした研究は類似性尺度の地道な比較であり、派手さはないが、使いやすいコミュニティシステムを作るためには必要不可欠な調査である。こうした積み重ねがWeb2.0におけるユーザ体験の向上につながっている。

ソーシャルブックマークの分析

ソーシャルブックマーク(SB)に関する研究は、まだ始まったばかりであり、2005年に初めての国際ワークショップ^{☆4}が開かれた。そのオーガナイザでもあるS. Golderは、早い時期にdel.icio.usのタグのうち約9万個について分析を行っている¹⁶⁾。インスタンス(ブックマークするWebページ)が増えると、ユーザは新しいタグをつけていくが、その増加率はユーザによって大きく

異なる。つまり、新しいタグを気にせずどんどんつけていく人と、前に使ったタグにこだわって少数のタグを使う人がいる。タグの用途別の種類もいくつかあり、内容を示す普通の意味でのタグのほかに、インスタンスの種類(カテゴリ)を表すもの(articleやblogなど)、所有者を表すもの、質や特徴を表すもの(scaryやfunnyなど)、自己言及的なもの(mystuffなど)、タスクにかかわるもの(toreadなど)に分けられると述べている。

タグは一般的に、3つ組 $\{u, t, i\}$ として記述される。あるタグ t がユーザ u によってインスタンス i に対してつけられたことを表す。インスタンスとは、Webページや写真、動画、論文など、タグをつける対象である。このうち $\{u, t\}$ だけに着目すると、前述のアフィリエーション行列が得られ、アフィリエーションネットワークを作ることができる。つまり、同じようなタグを使っている人のネットワークを作ったり、同じような人からつけられているタグのネットワークを作ることができる。同様に $\{t, i\}$ だけに着目して、タグのネットワーク、インスタンスのネットワークを作ることができる。P. Mikaらはこれをdel.icio.usのデータに対して適用し、タグの関連性を調べるにはユーザの共通性に着目するほうがその精度が良いことを示している¹⁷⁾。丹羽らは、はてなブックマークのデータ約5,800人分を用い、あるタグが1つのインスタンスにどれだけ特徴的に用いられるかというタグとインスタンスの「親和度」を用い、タグの関連性を求めている¹⁸⁾。SBでは、polysemy(多義語)をどのように解消するかが重要なトピックの1つである。最近では、それを自動処理するための研究が行

☆4 Collaborative Web Tagging Workshop (WWW2006).

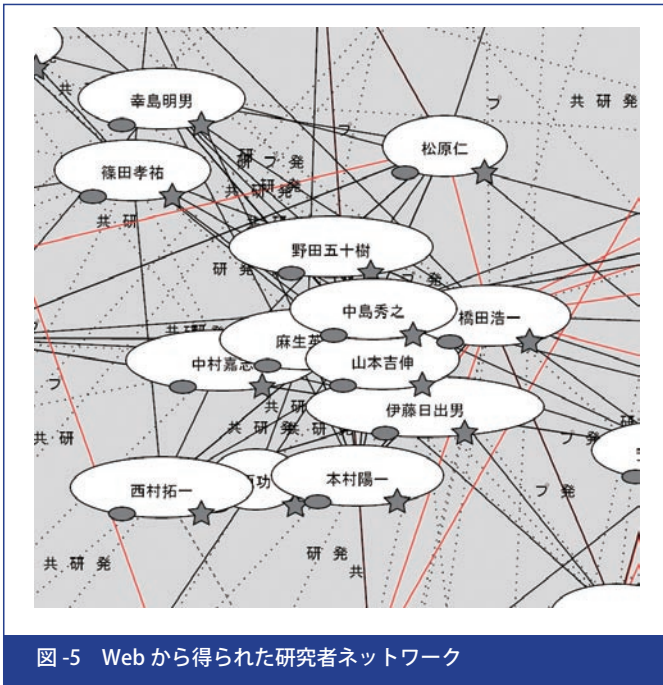


図-5 Webから得られた研究者ネットワーク

われている¹⁹⁾。

SBは、世界をどのように分類するか、その分類がコミュニケーションを通じてどのように共有されるのかという、言語学や人工知能で重要なテーマを含んでいる。世界の分類はある種の知識であり²⁰⁾、簡単な仕組みによって実現されたSBにより、語彙が構築されていく様子を俯瞰できるのは興味深い。言語学者のSaussureはその著書「一般言語学講義」の中で、ラングとパロールという2つの概念を対立させている。パロールが個人の言語実践であり、それが共同体で用いられるようになったものがラングである。SBではまさに、個人にとっての意味がコミュニティで共有されるに至る過程を見ることができる。哲学者のWittgenstein(後期)は、言語は使用によってのみ意味が決まるという言語ゲームの概念を述べたが、その概念もSBの仕組みの上で鮮やかに蘇るのかもしれない。いずれにしても、今までは目に見えなかった言語・社会現象がWebというプラットフォームを通じて可視化され、強化されているということは、言語学や社会学のこれまでの膨大な知識の蓄積に大きな可能性を開いている。

検索エンジン、そして今後のWeb技術

さて、Web2.0のさまざまな現象は、検索エンジンにより適切な情報が探せるようになったという部分に依拠するところが大きい。検索エンジンで探してもらえらるから、Wikipediaには人が来るのであるし、質の高いブログを書く人がいる。検索エンジンは今後ますますインフラ化するだろう。

検索エンジンを1つのモジュールとして用いる研究は、以前からあったが最近では非常に活発に行われている。その中でも、本稿と関係する「人の関係性」、そしてオントロジーの抽出という話題をここでは取り上げよう。検索エンジンに氏名を入れるとその人の情報が出る。2人の氏名を入れると、その2人が共通に含まれるページが出る。これによって、2人のつながりを知ることができる。特に、Web上に情報が顕在する研究者や著名人、企業などは、こういった分析が可能である。筆者らは、検索エンジンを用いて、こういった社会ネットワークを抽出する「ソーシャルネットワークマイニング」の研究を行ってきた²¹⁾。検索エンジンとテキスト処理を用いて、図-5のようなネットワークを抽出することができる。

また、P. Mikaらも、Web上の名前の共起関係やFOAFファイルから社会ネットワークを抽出し図示するFlink^{☆5}というシステムを作っている。

検索エンジンを用いた研究として印象的なのは、P. Turneyらの研究である。TOEFLのシソーラスの同定問題(「次の中から、〜と同義である語を選びなさい」という問題)を、検索エンジンを用いて答えるシステムを作り、ノンネイティブの学生の平均スコアを上回る精度で正解することができると示した²²⁾。情報の量による検索エンジンの「賢さ」を実感することも多いが、実際のテストで(特定の問題に対してではあるが)簡単なアルゴリズムによって人間より高い点が出たわけである。

地名や組織名など各種のエンティティ間の関係を捉えるために検索エンジンを用いる研究も活発である。S. Staabらは、検索エンジンとテキストのパターン分析を用いて、エンティティ間のオントロジーを抽出している²³⁾。実は、Googleの創始者であるS. Brinが、“Extracting patterns and relations from the world wide web”という論文²⁴⁾を1998年に書いているのが非常に示唆的であると私は考えている。その論文では、著者と本のタイトルなどエンティティの関係性を取り出すことの重要性和その手法について、当時の技術からのアプローチが述べられている。Web上の情報処理をユーザを含んでさらに高度化するには、エンティティ間の関係を捉えなければならない。人工知能の技術で古くから扱ってきたのも、関係性を基本とする知識の表現や処理(推論)であった。膨大な記号の世界であるWebをフィールドとして、成熟してきた検索エンジンをインフラとして、エンティティの関係性を捉え処理する技術は、今後の重要な方向性の1つではないだろうか。

以上、本稿では、Web2.0を、ユーザの関係性、ユー

☆5 flink.semanticweb.org

ザのネットワークという視点から捉え、その研究の動向を概説した。世界は確実にネットワーク化している。ここでいうネットワークとは、物理的なネットワークだけでなく、意味や価値といった面でのつながりである。Web2.0を意味や価値というユーザサイドの側面から支えるのは、他者との関係性であり、それが織り成すネットワーク構造であるというのが本稿のメッセージであった。産業界主導に見えるWeb2.0の世界も、アカデミックの分野で数多くの良い研究が行われ、影響を与えている。日本でもこういった研究が、実システムに利用できる本質的な知見を提供することを念頭に、活発に行われるようになることを願っている。

参考文献

- 1) Watts, D. and Strogatz, S. : Collective Dynamics of Small-world Networks, Nature, Vol.393, pp.440-442 (1998).
- 2) Barabási, A. L. : 新ネットワーク思考, NHK 出版 (2002).
- 3) 安田 雪 : 社会ネットワーク分析—何が行為を決定するか—, 新曜社 (1997).
- 4) 安田 雪 : 実践ネットワーク分析, 新曜社 (2001).
- 5) Adamic, L., Buyukkocuten, O. and Adar, E.: A Social Network Caught in the Web, Vol.8, No.6 (2003).
- 6) 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 (Mar. 2006).
- 7) Finin, T., Ding, L. and Zou, L. : Social Networking on the Semantic Web, The Learning Organization (2005).
- 8) Leskovec, J., Adamic, L. A. and Huberman, B. A. : The Dynamics of Viral Marketing (2005).
<http://www.hpl.hp.com/research/idl/papers/viral/viral.pdf>
- 9) Richardson, M. and Domingos, P. : Mining Knowledge-Sharing Sites for Viral Marketing, Proc. SIGKDD'02 (2002).
- 10) 山本 晶 : 発信する顧客は優良顧客か?—サイトの訪問動機とオンライン・ショップの購買履歴データの分析—, 消費者行動研究, 11(1)・(2), pp.35-49 (2005).
- 11) Adar, E. and Adamic, L. A. : Tracking Information Epidemics in Blogspace, Web Intelligence 2005 (2005).
- 12) 古川忠延, 松澤智史, 松尾 豊, 大向一輝, 内山幸樹, 武田正之 : Weblog 間の話題伝播過程に注目した重要トピックの抽出, 第20回人工知能学会全国大会 (2006).
- 13) Backstrom, L., Huttenlocher, D., Lan, X. and Kleinberg, J. : Group Formation Inlarge Social Networks : Membership, Growth, and Evolution, Proc.SIGKDD'06 (2006).
- 14) 安田 雪, 松尾 豊, 濱崎雅弘 : SNS における関係形成原理—No man is an island—, 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).
- 15) Spertus, E., Sahami, M. and Buyukkocuten, O. : Evaluating Similarity Measures : A Large-scale Study in the Orkut Social Network, Proc. SIGKDD 2005 (2005).
- 16) Golder, S. and Huberman, B. A. : The Structure of Collaborative Tagging Systems, Journal of Information Science (2006).
- 17) Mika, P. : Ontologies are Us : A Unified Model of Social Networks and Semantics, Proc. ISWC2005 (2005).
- 18) 丹羽智史, 土肥拓生, 本位田真一 : Folksonomy マイニングに基づく Web ページ推薦システム, 情報処理学会論文誌, Vol.47, No.5 (May 2006).
- 19) Wu, X., Zhang, L. and Yu, Y. : Exploring Social Annotations for the Semantic Web, Proc. WWW2006 (2006).
- 20) 池田晴彦 : 分類という思想, 新潮社 (1992).
- 21) Matsuo, Y., Mori, J., Hamasaki, M., Takeda, H., Nishimura, T., Hasida, K. and Ishizuka, M. : POLYPHONET : An Advanced Social Network Extraction System, Proc. WWW 2006 (2006).
- 22) Turney, P. : Mining the Web for Synonyms : PMI-IR versus LSA on TOEFL, Proc. ECML-2001, pp.491-502 (2001).
- 23) Cimiano, P., Ladwig, G. and Staab, S. : Gimme' The Context : Context-driven Automatic Semantic Annotation with CPANKOW, Proc. WWW 2005 (2005).
- 24) Brin, S. : Extracting Patterns and Relations from the World Wide Web, the International Workshop on the Web and Databases (1998).
(平成 18 年 10 月 2 日受付)

松尾 豊 (正会員)
y.matsuo@aist.go.jp

1997年東京大学工学部電子情報工学科卒業。2002年同大学院博士課程修了。博士(工学)。同年より産業技術総合研究所勤務。2005年よりスタンフォード大学客員研究員。社会ネットワーク研究所研究員。人工知能学会, AAAI 各会員。

