

## パネル討論

[会合] ネットワーク生態学研究会 第3回サマースクール  
 [日時] 平成19年9月5日(水)  
 [会場] 富士教育研修所

## ネットワーク科学の今後

司 会 林 幸雄 (北陸先端科学技術大学院大学 知識科学研究科)

パネリスト 藤原義久 (情報通信研究機構/国際電気通信基礎技術研究所)

友知政樹 (沖縄国際大学 経済学科)

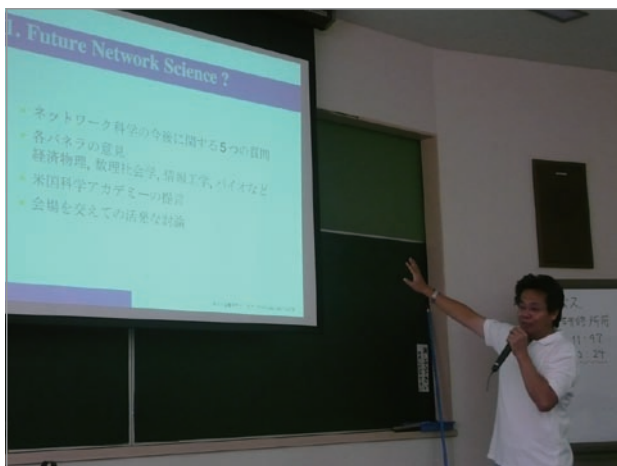
小島一浩 (産業技術総合研究所 知能システム部門)

小野直亮 (大阪大学工学部 情報科学研究科)



## ネットワーク科学が与えた意義

司会: 「ネットワーク科学の今後」についてパネル討論を行いたいと思いますが、まず、その動機付けにもなった2005年に出版された『Network Science』について話したいと思います。これは、米軍の依頼で米国科学アカデミーの下部組織であるNational Research Council (NRC)が、軍とは違うNPOとしての中立な立場で、軍事を含めてネットワーク科学がどのような影響を与えるかをまとめた報告書です。世界中のネットワーク科学に関係する研究者に対するWeb上でのアンケートの回答を元にできています。当然、ネットワーク科学の最先端の研究者であるWatts, Newman, Barabásiらの意見も入っています。その調査結果の提言としてネットワーク科学における7つの重要な研究テーマが挙げられています。まずは、パネラの方々のご意見を聞き、議論が深まった後半にNRCの提言について話したいと思います。



司会: 『Network Science』とは少し異なっていますが、類似した内容でより簡潔にした、以下の5つの質問に対

して、パネラの方々に順に答えていただきたいと思います。

- 1 現在の研究分野と主な研究対象は何か?
- 2 これまでのネットワーク科学が与えた意義や成果は何か?
- 3 今後取り組むべき課題は何か?
- 4 何が本質的に重要で、その理由は何か?
- 5 そのために何をするのがよさそうか?



藤原: 私の興味は人や社会のネットワークにありまして、主に経済を対象に研究しています。経済現象を色々な切り口で見ており、たとえば経済における企業の成長、ゆらぎ、経済ネットワーク、人的ネットワークを扱ってきました。今は、

解析・可視化ツール、コミュニティ分析に興味があります。

これまでのネットワーク科学が与えてきた意義は、1つ目には、社会ネットワーク分析は歴史が長く、さまざまな研究成果が出されており、特に大規模のネットワークの色々な解析技術ができたことです。つまり、統計的にマクロなレベルで捉えるという手法が1つの意義だと思います。2つ目に解析ツールが出てきたことがあります。特に離散数学の研究者の寄与が大きいです。あるマクロな性質を保ったネットワーク、たとえば、一般化ランダムネットワーク<sup>☆1</sup>を統計の帰無仮説<sup>☆2</sup>として

☆1 一般化ランダムネットワーク: 次数分布だけを保存し、それ以外の構造がランダム性を持つようなネットワークのクラス。

☆2 帰無仮説: 統計学の検定で調べる対象となる母集団の性質に関して、最終的に棄却されるべき仮説。

使い、それから出てくる量を実際のデータとどこが違うかを調べる手法が、90年以降出てきました。3番目には、実際のネットワークを見ると、べき乗分布かはともかく、分布のすそ野が非常に長い分布から、帰結が出てきたことです。2つ例を言うと、1つは、クラスタリング係数は、次数がポアソン分布に従う一様ランダムグラフで計算した量とよく比較しますが、次数分布のすそ野が長い同様な分布を持つ一般化ランダムネットワークで評価すると値が百倍ぐらい違います。もう1つは、友達の友達の数を評価するときに、次数分布は非常に重要で、たとえば、社会ネットワーク分析でスノーボールサンプリング<sup>☆3</sup>といったような手法がありますが、そういうとき、次数分布を加味しないと、どちらも実際にサンプリングに対して評価を間違えることになります。

今後取り組むべき課題については、大規模ネットワーク、多元的なネットワーク、動的なネットワークの3つをどうやって解析するかです。大規模ネットワークは可視化の課題などがあります。多元的なネットワークというのは、たとえば人のネットワークは必ずしも1種類のネットワークでつながっているわけではなく、趣味や地域といったさまざまな属性でつながっているネットワークのことです。それをどう捉えるかも課題です。また、動的に変化するネットワークの特徴をどう捉えるかも課題です。

また、研究者ごとに、違う分野の対象のデータを持っていれば、ネットワークの構造だけで分かることと、対象ごとにノードやエッジが持っている属性などさまざまな情報を比較できます。



**友知：** 私の研究分野は、数理社会学と進化経済学で、研究対象は社会ネットワークやその上でのゲーム、さらに、ゲームにおいて戦略が伝搬していくという考え方に基づいた普及現象です。

これまでのネットワーク科学の意義を考える際、問題

を認識して発見する過程において、言葉によってその問題を語らなければならないし、それにより理解する必要があります。その言葉とは、中心性などのネットワークの指標だと思います。指標というものの見方—どういう角度から問題を捉えるか—から、つながり構造の

データを静的に分析したことがこれまでにネットワーク科学が与えた意義ある結果の1つであります。2つ目は、情報フローなどに関する動的ネットワークモデルの提示もまた意義ある結果だと思います。これは、現在、研究が進行しています。

本質的に重要なことは、データの収集方法です。たとえば、世界は本当に狭いのか？という問題を扱ったミルグラムの実験では、データが確かでなかったと思います。今後、大規模社会ネットワークのデータの収集と解析が重要です。その対象として、Webの世界も重要ですが、たとえば、職場などの日常のネットワークも今後の研究課題だと思います。データの収集が本質的に重要ですね。提言としては、大規模ネットワークの収集プロジェクトの立ち上げを行うことがよいと思います。



**小島：** 興味の対象は、かつてニューロを、次に自律分散制御をやって、2000年頃にネットワーク科学を知りました。興味の対象が自律分散制御なので、ネットワークと関連性があるのではないかと思い始め、当時、Peer-to-Peerの研究が盛んに行われ

始めており、ネットワーク上で自律分散させるとは、まさにPeer-to-Peerでの情報のやりとりが焦点になると思いました。それからオーバーレイネットワークや情報検索などを4,5年やっています。自律分散制御の観点から興味の対象は特に、ローカルな通信だけでいかに機能を実現するかです。機能が与えられてそれを設計するにあたって、なるべく、ローカルな情報だけでやりたい、というのが自律分散制御の立場で、常にそれが必要かということ、それが集中制御でよい場合もあります。ただ、私は、ローカルな情報だけで機能を実現する問題に興味があります。

DNAから社会などのサイズやスケールの異なる対象の個性性を無視して抽象化することで、数学的に語ることの意義が大きいですね。しかし、社会学といった現実を対象とすると、抽象化したことで個別の機能を落とすすぎたかどうかを考えねばならないと思います。それが課題の1つ目です。2つ目としては、構造からダイナミクス（ダイナミクスにはネットワーク上での情報フローなどのダイナミクスとネットワーク成長のダイナミクスの2つがあります）にも注目しなければならないことです。

Peer-to-Peerの情報検索をやる際に、ダイナミクスが絡んできますが、情報検索で言うとKleinbergの

☆3 スノーボールサンプリング：1人ないし小集団の人によってリストアップされた知人に、同じ質問調査を次々に雪だるま式に繰り返すようなサンプリング手法。

Navigation 問題<sup>☆4</sup>も絡んできます。結局、どのようなものを目指し、どのような機能をアルゴリズムで実現化したいかによって異なります。このアルゴリズムを評価することがネットワーク科学の本質になります。

提言としては、マイクロ・マクロの評価指標を作成すること、ネットワークの個別性にも注目し、ノードとリンクの意味を考えることです。たとえば、マルチエージェントで社会を扱うときは、社会学の研究者と問題を発見していかなければならないと考えています。



**小野：** バイオインフォマティクスを中心に生物の進化や起源といった、生物の複雑さがいかに創発されたかを、複雑なネットワークとして捉えて解析しています。生物のネットワークといっても色々なレイヤがあります。たとえば、遺伝子同士の相互作用、

多細胞生物の細胞同士のネットワークなどにおいて、生物進化の過程のモデリングや自己組織化の過程をシミュレーションしたり、また進化的な解析を行い、どのように進化の中で複雑さを増してきたかということを中心に行っております。

大規模な可視化や解析によりメタなレベルで統計的な指標により全体を俯瞰し、今まで漠然とたくさんの要素がつながって影響しあっているとしか理解できなかったことを定量的に比較して議論することができる、それによってネットワークがどのようになっているかを理解できたことが、これまでの意義ある成果だと思います。

これからの課題は、ネットワークが持つ機能、そういう構造を持っていると、どういう性質が現れるかに焦点が当てられるのだと思います。構造がどう成長して変化するか、それによりどう機能が変わるか、そこが比較できるようにするには、どこがどの程度変わったかを定量的に理解する、その変化というものが、最終的な生物の分裂なり成長の早さ、適応のしやすさなどに対して、生物系にどれだけ影響を与えているかを比較する方法を見ていく必要があります。

統計的に全体を俯瞰できるようになったことが1つの

成果ですが、ただ、単純な全体の統計量による俯瞰は、個々の性質に対して抽象的すぎます。個々の局所的な構造やノードの性質がどのような機能を持っているかが直接かかわってくるので、そこでのすり合わせを見ていく必要があります。そうすると、単純に全体の統計量というのではなくて、ある分子なら、その要素の周りの構造、その要素の性質と実際の機能を定量的に比較するアプローチが必要です。進化などによってネットワークが時間に従って変化していったり、異なるレイヤの異なる種類のネットワーク間の比較が必要です。

まとめると、ネットワーク自体の変化に興味があります。どのようなモデルからどのようなネットワークができたかが明らかになったことに意義があります。このような方法で、ダイナミクスを表現できるようなアプローチを探らないといけないと思います。単純化しすぎると、現実のネットワークを十分説明できるモデルが出てきにくいと思います。実際のダイナミクスは、ランダムにリンクが張られるのではなく、その背後にある潜在的なインタラクションがあり、それが顕在化することでネットワークとして観測できるのでしょうから、どのようなタイミングで機能が顕在化するかを表す仕組みが必要です。

### ネットワーク構造と機能、サンプリング問題

**司会者：** 皆さんの意見に共通するのは、ネットワーク科学が与えた意義としては、広い範囲の応用で実データを使った分析ができたことと、モデルにより基本原理の解明ができたことですね。今後の課題については、それぞれお考えが少しずつ違うようで、藤原さんは多元的な関係による分析、友知さんはデータ収集、小島さんは構造とダイナミクス、小野さんは構造と機能の関係をそれぞれ重要だとおっしゃっています。小野さんの意見は特に、適応性を強調されたのですね。さて、会場の皆様の意見を聞きたいと思います。

**会場 A：** ネットワーク科学において、機能と構造のどこを特に注視すべきだとお考えですか？ 生物の仕組みを解明することによって、そこで人工物みたいなネットワークにおいて、現状の通信網が持っていない何かいい性質が生物系の中にはあるかもしれない、そういうところを見てみたい、そういうところが重要であるとの認識でよろしいでしょうか？

**小野：** 個々の性質の違いとネットワーク構造の違いというのを、定量的に相関を持って比較できる指標で解析する必要があります。

**会場 B：** 小野さんの立場は、ネットワークによって分子間の相互作用を解明することだと思います。一方で、工学的なものを研究対象としている方々は、ネットワー

☆4 Kleinberg の Navigation 問題：Milgram の手紙転送実験において、参加者は目標人物の限られた情報と自身の友人関係という局所的な情報だけから、なぜネットワークのサイズに比べて小さいステップ数で手紙が到達できるのか？という問題。手紙転送実験は、ネットワークの直径が小さいという「Small World の存在性(事実問題)」と、参加者が目標人物までの効率的な経路を発見するという「Small World の認知性(認知問題)」に分けることができ、Kleinberg の Navigation 問題は、後者に相当する。

クを使って工学的に役立つものを作りたいという立場だと思います。生物のようなロバスト性のあるシステムをネットワーク科学によって解明できるなら素晴らしいと思いますが、小野さんのネットワーク的な理解がその解明に寄与するのでしょうか？

**小野：** 環境が変わったときに生物がその代謝系を操作して、環境に適応していくということが実験的に知られています。代謝系がロバストであるという前提なら、その中でネットワークがどのように変わり、環境に対してどうバランスをとるかが鍵だと思います。その挙動を一般化することで、トラフィックや通信に応用できるかもしれません。

**会場C：** 友知さんに質問なのですが、社会的なネットワークの中で、どのくらいが大規模なのですか？ また、その解析手法は、たとえば、階層によって変わっていくのでしょうか？

**友知：** 対象に応じてさまざまな規模がありますが、1,000ノードぐらいからが大規模でしょう。手法としては、階層性などによって変化してきます。また、ネットワークの一部からネットワーク構造を推定する方法は確立されていません。所得などの大きなデータは、社会学でも昔から使用されてきました。この場合、全数調査でなくサンプル調査によって母集団の平均や分散を推定してきましたが、一歩進めてネットワーク全体を推定できるようになればいいと思います。

**会場D：** サンプルによる統計は、統計物理によると、ある程度の規模がないとべき乗分布で語ることはできないのですが、社会ネットワークでは規模が小さくてもネットワーク科学の知見が使えるのですか？

**友知：** それは何を知りたいかによります。問題の定義によるので、ノード数が1,000だからこういう現象が起こるとはいえません。社会がフラクタル構造であるかもしれません。その場合、全体を知る必要があるのでしょうか？

**司会：** 会場から、あるいはパネラ同士でも結構ですので、これまでの発言に対して何か質問はありませんか。

**小野：** 藤原さんに質問があるのですが、生物のレイヤに興味がありますが、それらレイヤは個々に研究されています。それらをつなぐため、多元的ネットワークの観点から見ると、それらを比較する方法がありますか？

**藤原：** ネットワーク同士は比較が難しいですね。並べて議論する方法はまだありません。その方法を生み出すヒントの1つに伝搬や連鎖に注目することがあります。伝搬させるとき、思ってもみないところで連鎖することがあり、それは、別の関係から伝搬しているということが可能性としてあります。たとえば、企業の取引関係に着目した場合、どうして有力な情報が得られたかが分か

らないときがありますが、実は、別の関係としての株の所有から連鎖している場合があります。

**会場E：** 情報伝搬とは、ネットワークの別の捉え方であると思います。ネットワークから抽出して解析するのではなく、たとえば、ネットワークに刺激を与えた状態で観察する捉え方があると思います。要するに、ネットワークの成長とは違ったダイナミクスのものが解析できるのだと思います。

先ほど、大規模な社会ネットワークのスケールが適当かの話がありました。大規模な社会ネットワークは、コミュニティが多数内包されており、クラスタの中にクラスタが見られる程度の規模が面白いのではないかと思います。クラスタの中のクラスタはフラクタルの話ともつながってくると思います。それに刺激を与えて観察したら個人的に面白い解析ができるのではないかと思います。

**会場F：** 小島さんに質問します。ネットワーク上の情報フローなどのダイナミクスとネットワーク成長のダイナミクスは分離できないと思います。工学的な発想だと、たとえば情報伝搬だと、情報の流れやすいところにリンクが張られることもあるかもしれません。それらを含めた情報伝搬のモデルができればいいと思います。

**小島：** Peer-to-Peerの観点だと、ネットワーク上の情報フローを効率化するために、ネットワーク成長のダイナミクス(Generation)を変化させることを考えます。

～コーヒープレーク～



#### 米国科学アカデミー(NRC)からの提言

**司会：** 我々の考えを大体共有できたと思いますので、ここで、『Network Science』からの提言を、紹介させていただきます。さらに、もう少し広い立場から、ネットワーク科学をどう考えたらよいかということを議論していきたいと思います。

まず、ネットワーク科学は生まれたばかりであり、投資をするには、焦点を絞るのではなく基礎から応用まで

の広い範囲を対象とすべきということです。たとえば、米軍は少なくとも年間一千万ドル（約12億円）の投資をすべきで、さらに、連邦政府機関であるNSF、DOEやNIHなどは別の投資をすべきと提言されています。

また、米軍が想定できる範囲を超えた非常に広い範囲の可能性が期待でき、ネットワーク科学には国家的な取り組みが必要であると明記されています。

特に、情報ネットワークと社会ネットワークの関連が重要であるとの提言がされております。本研究会はまさにそこに注目しているので、我々としても追い風と言えます。

結論として、ネットワーク科学における7つの重要な課題は、以下の図のようになっております。それぞれのどれに注力すべきかは、下図の円グラフのように大きい比率の順で、空間上の伝搬などの構造と機能の関係ダイナミクス、モデリングやツール、要求される機能や特性に適した設計や統合、数学的基盤、異分野間の概念の共通化、より良い実測、頑健性です。

私の印象としましては、7番目の頑健性すなわち故障や攻撃への耐性の課題は、我々の研究会では弱いところ

なので、さらに強化しなければならないと思います。以上、『Network Science』の提言も踏まえて、会場の皆様と議論を深めたいと思います。

### データ収集とツール共有化が重要

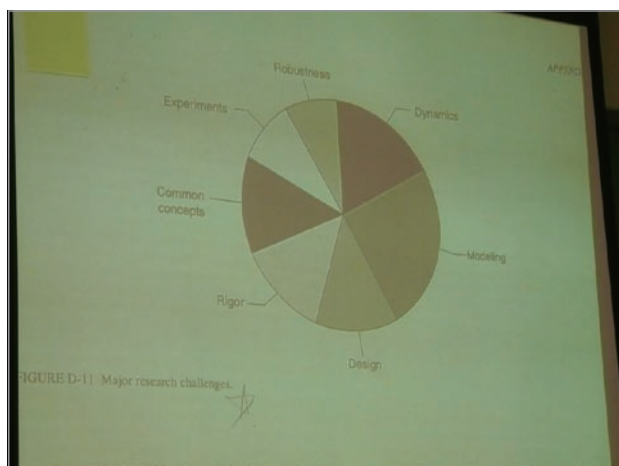
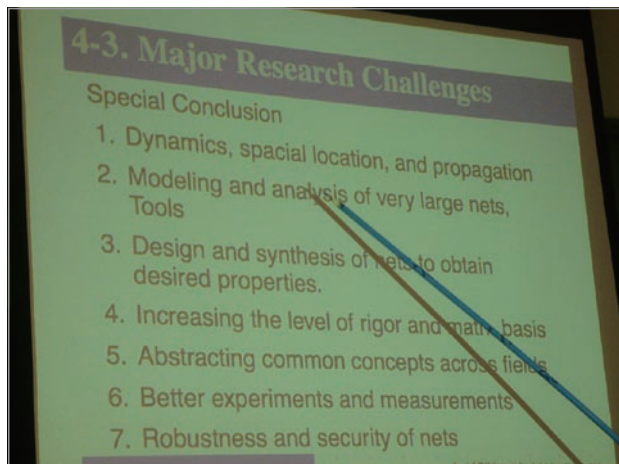
**会場 G:** 友知さんに質問します。まず、社会ネットワークにおいて問題設定をどう捉えればよいでしょうか？ また、問題設定から意味のあるデータを集めなければならないとき、大規模なネットワークが対象のとき、そのデータをどう集めるのがよいと考えられますか？

私が考える課題の1つに、PDFなどの電子化文書を検索してキーワードの共起関係により文献のネットワークを構築した、全世界の知識の構造化があります。これはさまざまな問題解決の手段になるのではないかと思います。ただ、データが膨大なので自動で行う必要があります。それでもものすごい時間がかかります。もしかしたら、スーパーコンピュータを駆使すれば、文献間のネットワークならアーカイブのデータベースがあるので、できるかもしれません。

上記の課題と同様に、社会科学でもデータ収集の自動化が必要になると思います。データベースもなく、アンケート調査や聞き取り調査をするといったことは、人間が海戦術的にやらざるを得ないと思いますが、ある程度の規模ならできるが、1,000人規模になってくると、どうしても自動化が必要になってくると思います。さらに、回収率が高ければデータが集まると思いますので、このための工夫も必要です。このように、体系的に自動的にやる仕組みを、さらに、参加者が進んでデータ収集に協力してくれるシステムを構築しないと大規模な場合では難しいと思います。

**友知:** 問題設定ありき、というのは私もそう思い、また皆さんも同意なさると思います。たとえば、世界中の映画俳優の競演関係を元にしたネットワークには果たして何の意味があるのでしょうか？ その問題設定は、社会的・自然科学的な問題を解くというよりも、スモールワールドのようなネットワークが競演関係を元にしたネットワークにも見られると主張したかっただけなのではないかと思います。

社会学的な問題として、日本のコミュニティが壊れている、昔より暮らしにくくなったという問題をネットワーク科学からのアプローチで見たらどのように解決するかを考える必要があります。このような大規模なデータを扱う必要がある問題設定では、意味あるデータを取ってくるため、多くの人がかかわって議論や収集ができる効率のよい情報収集システムが必要です。Web上で収集できない、たとえば、Webを使えない環境や人がい



Network Science  
by Committee on Network Science for Future Army Applications,  
National Research Council National Academies Press ; 1st edition  
(December 15, 2005) ISBN-10 : 0309100267  
[http://www.nap.edu/catalog.php?record\\_id=11516](http://www.nap.edu/catalog.php?record_id=11516)

るなどのデジタルデバイド問題がある対象などでは、フィールド調査の必要もあり、この場合でも効率のよいデータの収集システムを考えないといけません。

**司会：** Web 自体が、人類の共有物であり、コミュニケーションのツールであり、データ収集の道具でありますね。たとえば、『Network Science』は、Web 上でアンケートを取り集計したものです。さらに、単に集めただけでなく、そのようなデータを共有することがきわめて重要ではないかと思います。また、データを収集しても、分析しなければ意味がありません。そこで、関連してツールについて聞きたいと思います。

**会場 H：** 先ほどからデータの話がいくつか出ていましたが、昨年開催された SIGIR という情報検索の国際学会の中で Yahoo! の CDO (Chief Data Officer) や Microsoft の役員が盛んに言っていたことは、これからはデータである、これまでは科学で色々やってきたけれども、これからはデータ科学だといい、とにかくデータの重要性を強調しておりました。さらに、Watts が Yahoo! に雇用されましたが、Google や Yahoo! といった企業では、社会ネットワーク解析を非常に重視し、社会ネットワークの科学者を大量に雇用して解析しています。彼らがやっていることはデータが非常に大事であるとの考え方でデータ解析を行っております。この傾向から、これからは、データ科学が主流になると思います。また、データの共有化が最も重要であると考えます。

**会場 I：** 企業のデータなどは、クローズになっており、共有化できない問題もあります。たとえば、5月の World Wide Web Conference に出ていたあるチームの発表のデータが、我々が3月まで行っていたモデルで説明できそうだったので、論文に出している Gnuplot のプロットをくださいといいました。しかし、データ提供元である企業との契約上、駄目ですといわれました。プライバシーの問題などもあるのですが、論文発表したデータを反証するために共有化できないというのはそもそも科学なのでしょうか？

**友知：** 私も、科学として問題があると思います。

昔からデータは大切でしたが、今、データがさらに重視されるようになったときさまざまな分野でいわれているのはなぜなのでしょう？ モデルが先行しすぎた反省なのでしょう？

**小島：** データを集める、あるいは保持することについて、サイエンスとビジネスの違いがあって、ビジネスではデータを出さないことで優位性を保ちたいという考えが働きます。だから、データを提供されるはずはないし、そのデータについて研究したいならその企業に就職するしかないと思います。API を公開する動きもありますが、それは限定された動きで、それを自由に使える権利もま

ったくないし、契約上で制限をつけられたり、契約自体を断られたりすることもあります。企業を守るためには、そのデータを守秘しなければならず、仕方のないことです。会場 I の方のケースでも、生データでなくプロットデータですら、企業の方は出したくないという心理が働きます。このようにサイエンスとビジネスとではデータの意味がまったく異なるので、科学者の立場としてデータを提供してくださいとはいえません。科学としてデータを自由に使いたいなら科学者として独自に収集するべきで、どう集め、どう共有化するかを議論しないといけません。

また、問題設定を考えずに単にデータを集めただけでは無意味です。たとえば、地域社会の問題設定なら小規模のデータでよいし、ミルグラムの実験のようなことがやりたいなら大量のデータを集めなければならないと思います。目的からデータ収集の対象や規模を決めて集めなければなりません。このとき、我々がやっている Web マイニングというデータ収集の手法が社会学にそのまま使えるとはいえないですが、社会学の一部として SNS (ソーシャルネットワーキングサービス) という対象に限定すれば、社会一般としては語れないけど、SNS として語るができるデータを集めることができると思います。

**会場 J：** 大規模なデータ収集の手法を議論する前に、私は、なぜ、大規模なデータがそれぞれの分野で軸になっているのか？をお聞きしたいと思います。たとえば、統計的性質が知りたいだけなら、全数調査は必要なく、統計調査としてサンプリングすればよいだけだと思います。このとき、全数サンプリングで得られるメリットは、データを増やしたことにより、統計的な精度が上がるだけだと思います。皆さんの分野のなかで、単に精度を向上させたいだけでなく、何らかの大規模なデータでなければ分からないことはありますか？

私の分野は通信ネットワークなのですが、通信の場合、規模が大きいと、小さいときでは発生しなかったことが起こることがいくつかあります。たとえば、インターネットのルーティングにおいてはネットワークサイズが大きくなると、不安定になることがあります。小さいネットワークと大きいネットワークでは挙動がまったく違うことが最近分かってきました。したがって、対象が大規模なら大規模なデータを扱わねばなりません。パネルの方々にもこういう経験はあるのでしょうか？

**小島：** 私の場合は、分野は Peer-to-Peer で、100 万人でも動くようなプロトコル、さらにネットワーク上で単にブロードキャストするのではなく、いかに自分が欲しい情報にアクセスできるかというのが主題です。なので、必然的に大規模なデータが軸になります。

また、各ノードが持っている情報は小さいけれども、ノード数が大きいから全体のコンテンツ量も大きくなります。ノードごとに持っている情報が異なる中で、いかに自分が欲しい情報にアクセスできるか？というものです。

**友知：** 会場Jの方の質問について、問題設定で対象にしているネットワークがそもそも大規模だから大規模ネットワークデータが必要であると思います。その理由は2つあり、1つ目には、会場Jの方のおっしゃったとおり、ネットワークサイズにより、ルーティングの挙動が変わってくるなど、ネットワーク上のダイナミクスの点があります。また、2つ目は、ネットワーク構造の話なのですが、大規模ネットワークの一部をサンプリングによって抽出し、そこから全構造を推定できるか？という未解決の問題にも帰結すると思います。

数学者に質問なのですが、大規模ランダムネットワークなどを作り、サンプリングを行い、そこから真のデータを推定できるかという研究はありますか？ 具体的には、サンプリングから元の隣接行列を推定することはできるのでしょうか？

**小野：** 極端な例として、それはむしろ、できないという研究があります。100万ノードクラスの普通のランダムネットワークを作り、10%サンプリングした場合、サンプリングの仕方によってスケールフリーに見えてしまうこともあり、大規模な対象は大規模なまま解析する必要があると思います。

**司会：** 推定に関する研究もいろいろあって、もともと一様ならともかく、偏っている場合のサンプリングの問題は難しいということですね。

**藤原：** ツールのことを話す前に、サンプリングの話で補足したいことがあるのですが、ちゃんとした標本調査を設計していないと思います。たとえば、上場企業だけなどのある属性だけを抽出してしまうことになります。そのとき、分布を見るとべき乗分布なのに対数正規分布に見えてしまうことがあります。ですから、標本調査を設計せずに、属性で見てデータを排除してしまうと、先のようなことが起こり得ます。

ネットワーク全体を見てみると、あるノードからの距離が8で最大だとすると、地球儀の上を北極から南極へ行くような形で、最後は距離が8のところに数人がいるような、そういうネットワークになっているはずで、その場合、距離が2のところまでで何らかの指標を選んできると、いくつか間違ってしまうことになります。また、距離が遠い位置に孤立しているノードは届きにくいので、ネットワークからランダムに頂点を選んでサンプリングすることと、リンクを通してサンプリングすることはまったく違います。

**司会：** データ収集の議論を踏まえた上で、ツールのことについて何か意見はありませんか？

**会場L：** 以上の議論からも、収集されたデータは分野によって違ってくると思います。ただ、ネットワーク科学という観点から見ると、さまざまな分野の分析方法やそのツールを、分野を超えて体系化しないと、結局、全体から見た場合、なんの解析をやっているかが分からなくなってくると思います。体系化するにあたってどのような方針でまとめていったらよいと考えられますか？

データを集めても、活かすツールがありません。それを共有化しないといけないのではないのでしょうか？

**藤原：** ツールを体系化することができるかということなのですが、社会ネットワーク分析では、ツールの体系化がかなりされていると思います。少なくとも社会ネットワークではそれらのツールを使えますし、Webの分野だとPageRankに始まり色々な技術が蓄えられていきます。しかし、注意しなければならないのは、対象ごとに使えるツールが違ってくると思います。よって、汎用的に使えるツールがどこまでさまざまな分野で使えるのだろうかという疑問はあります。

**友知：** ツール、すなわち、ものの見方ですが、色々なネットワーク指標があると思います。それはそれで現存するものを使えばよいと思いますし、対象とする問題を解くのに新しい見方が必要なら、新しい指標が出てくると思います。

**小野：** 友知さんのおっしゃったとおり、問題設定に従い、指標を自分で作っていくというようなことがあると思うので、体系化するのは非常に難しいと思います。

既存のツールを並べておいて、それがどこまで使えるかということは、使うときに各人が検討しないと難しいと思います。

**会場L：** 藤原さんが言われたようなアプリケーションのレベルの汎用的なツールでなく、よりメタ的なレベル（たとえば：APIとして）で公開するべきだと思います。そうすると、ある分野の人が別の分野の手法をツールとして実装するときそのツールを組み込むことで、実装を効率的に行うことができます。

さらに他分野の人も手法をツール化することをやれば、全体としてツールを共有化する量が増え、色々な人がより効率的に研究を進められると思います。

**司会：** それは、ソフトウェアのメタデザインが必要で、それに取り組むべきということですね。ツールのことを私も聞きたかったのですが、別の観点から意見はありませんか？

**会場M：** 確かに社会ネットワーク分析では、UCINET, Pajek, NetMinerなどのツールがよく使われているようですし、それらのバージョンもアップして

機能が増えています。個人的には多数の関数や描画機能を含む NetMiner を使っていますが、皆さん自分でツールを作り、個々で一部を公開していると思いますが、私はそれを使ったことはまったくなく、UCINET、Pajek、NetMiner に部分的に作ったツールもあって、これら以外を使いたいとは思いません。ツールの共有化は難しい問題で、皆さんライブラリとして公開されていても使っていないと思いますし、個々人が自分の問題設定でツールを用意していくのがいいと思います。

**司会：** どんどんツールとして公開されていくことが重要ですね。

**会場 M：** そうですね。本に載せられるような有用なツールを発表していけばいいと思います。

**会場 F：** ツールの共有ではなく、いかに大規模なデータ解析ができるツールを作るかという話をしたいと思います。Web を例にとると、Web 上のデータは大規模ですし、それを解析しようとする解析対象も大規模になります。Google では組織内部で色々なツールを作っており、たとえば Google ファイルシステムや多数のコンピュータに簡単に計算を割り振れるといったマップレデュースというシステムを作ったり、それを使った言語を作ったりしています。それにより、たとえば、膨大なリンクの行列を簡単にアクセスできるようにするデータベースを作ったりしています。膨大なデータを独自のツールにより、Google のエンジニアは研究開発対象として扱い、最終的にサービスとして社会に提供しています。膨大なデータのみならず、提供してくれるところがあります。たとえば、ある研究者の方は 100 億から 200 億のページのデータを持っています。そのリンクの情報を出してもいいといっていますが、出されたときに我々は扱えるのか、扱うツールがあるかという問題があります。このようなツール整備の問題は、私の知らないところで進んでいるのでしょうか？ たとえば、Google のようにデータベースからファイルシステムまで整備して扱えるようにする、このような研究をしており、超大規模なデータを扱うことができる人たちにこのような解析の研究が囲い込まれるのではないかと考えています。以上のような超大規模なデータベースやその解析ツールに関して何か意見はないでしょうか？

**会場 L：** 今の話は、効率のよいアルゴリズムの話に帰着してくるのだと思います。今までに作られてきた各分野のアルゴリズムを流用するということはできると思います。ただ、ネットワークに最適なアルゴリズムをとっているかということについてはほとんど研究が行われていません。ですから、独自に効率のよいアルゴリズムを開発して、それを各研究者がなかなか出せないという現状は、結局、共有できていないから独自にアルゴリズム

を開発せねばならず、全体として非効率になっていると思います。

**藤原：** 論文を読むと、Google のこれらのシステムのことが書いてあるのですが、実際、これらは API として実装できますか？

**会場 E：** できます。Google ファイルシステムやマップレデュースは、オープンソースの Java の検索エンジンに実装されているという話です。そこから、取り出してきて使えるようにできます。

大規模な問題に対して、どのように問題を分割して実行するかという問題もあります。そうすると、アルゴリズムも逐次的ではなく、最初から完全に分割して考えなければならないという発想の転換の必要があります。巨大なデータベースの利便性は、妥当な順序に分割して処理できることです。一般的に問題を分割するのは困難ですが、リンクのデータベースがアルファベット順に並んでいたとすると、そうすると妥当に分割して処理すればよいのでそういう意味でやりやすいと思います。API として提供されているというのは、処理した結果であって、我々がやりたいマイニングやネットワーク分析はもっと生データに近いレベルで直接解析するしかないと思います。

**会場 N：** 大規模なデータを扱うとき、それを直感的に見たいということで、ネットワークの可視化を考えますが、ネットワークの可視化はどのように行えばよいのでしょうか？ 個人的にはアニメーションや 3 次元の可視化をやりたかったのですが？

**司会：** 可視化のご専門の方からご意見をどうぞ。

**会場 O：** ネットワークの可視化は、グラフィックボードの性能向上のおかげでより安価かつ容易に行えるようになってきました。ただ、ネットワークによって可視化を考えねばなりません。結局、可視化によって何をしたいかによります。可視化したからといってその意味を考えねば無意味だと思います。重要なことは、可視化ツールを作ったとしても、可視化からより価値あるものを生み出す方法を考えなければなりません。これが見えないため、私としても公開するモチベーションが上がらなかつたりします。本などで可視化のアルゴリズムがあり、その手法自体は難しくなく、可視化で見せることだけ追求すれば 1 カ月ぐらいでツールができると思います。なので、可視化ツールの開発をやってみる価値はあると思います。市販や公開されている現在の可視化ツールは、1,000 ノードぐらいが限度で、さらに大きい物を可視化させようとするは無理です。今のところ、自前でやるしかありません。

**司会：** ツールについての補足ですが、近代科学社から『ネットワーク科学の工具箱』という書籍を編著として出



版します。これには、さまざまなアルゴリズムについて書いてあるので、それらの手法をツール化することに協力していただくとありがたいですね。

話は尽きないと思いますが、そろそろお時間です。『Network Science』の7つの重要なテーマの一部しか取り上げることができませんでしたが、引き続きネットワーク科学の今後を考える取り組みをしていきたいと思えます。

パネルの皆さん、会場の皆さん、どうもありがとうございました。

(平成 19 年 12 月 18 日受付)

林 幸雄 特集 1 参照

藤原義久  
yoshi@yoshi-idea.org

1992 年東京工業大学大学院理工学研究科博士課程修了。理学博士。京都大学基礎物理学研究所研究員等を経て、現在、(株)国際電気通信基礎技術研究所(ATR)主任研究員。(独)情報通信研究機構(NICT)専門研究員を兼務。著書に「パレート・ファームズ」(共著)(日本経済評論社, 2007 年)。

友知政樹  
mtomochi@okiu.ac.jp

1973 年、沖縄県浦添市に生まれる。1998 年、中央大学総合政策学部政策科学科を卒業後、米国へ留学。2002 年、カリフォルニア大学アーバイン校数理行動科学研究所博士課程修了。数理行動科学博士(Ph.D.)。中央大学総合政策学部特任助教授(2003 年度～06 年度)を経て、2007 年より沖縄国際大学経済学部准教授、現在に至る。専門領域は社会ネットワーク分析、進化ゲーム理論、数理社会学など。

小島一浩  
k.kojima@aist.go.jp

2001 年東京工業大学大学院総理工学研究科 知能システム科学専攻博士課程修了。同年独立行政法人産業技術総合研究所入所。現在、同研究所知能システム研究部門分散システムデザイングループ所属。スモールワールドやスケールフリーネットワーク、社会関係資本、P2P システムの研究に従事。博士(工学)。

小野直亮  
nono@ist.eng.osaka-u.ac.jp

大阪大学情報科学研究科 特任准教授。東京大学総合文化研究科修了。ATR にて人工生命モデルを用いた進化ダイナミクスを研究。現在は遺伝子発現情報を元にした代謝反応ネットワークの解析を主なテーマとしている。