



August 2017

vol. **6**

JST ERATO

Kawarabayashi Large Graph Project

NEWSLETTER

Journey と glee、そして Szemerédi と数学

研究総括 河原林 健一

Ken-ichi Kawarabayashi

つい先日、Journeyの特集番組があり、現在の「再ブーム」についての解説もあった。Journeyは、スティーブ・ペリー（現在は脱退している）のハイトーンボイスが耳に残る80年代を代表するロックバンドである。1981年に発売したアルバム「エスケイプ」は1000万枚以上のセールスを記録し、全米チャート1位を獲得した。このアルバムには、「Don't Stop Believin'」「Open Arms」といった彼らの代表曲が収録されている。そして1983年発売のアルバム「フロンティアーズ」も600万枚以上のセールスを記録し全米チャート9週連続2位を獲得している（ちなみに同時期に発売されたマイケル・ジャクソンの「スリラー」は、全米チャート37週連続1位獲得。現在の世界記録である3000万枚超のセールスを記録）。このアルバムには、今年のWBC（ワールドベースボールクラシック）の日本でのテーマ曲である「Separate Ways」も収録されている。

このJourneyが10年ほど前から再ブレイクしている。これは2008年にアメリカのTV番組「glee」の影響が大きい。gleeは、学園物のTVドラマであるのだから、ストーリー以上にパフォーマンスが注目を浴びた。実際gleeは、ミュージカル俳優やミュージシャンを配役に起用し、毎週何曲ものヒット曲のパフォーマンスを行った。そしてそのクオリティの高さは、エンターテインメントの世界で高く評価され、gleeは世界ツアーを行うまでに至っている。そして「Don't Stop Believin'」はgleeの代表曲で、若い世代がJourneyを聞ききっかけとなった。

数学界でも似たようなことが起こっている。SzemerédiのRegularity Lemmaは、ある整数論の結果を導くために、1975年にSzemerédiが証明したグラフ理論の道具である。SzemerédiのLemmaはグラフ理論の分野では非常に高く評価されていたが、2000年くらいまでは、強力ではあるが、グラフ理論の一分野である「極値グラフ理論」の範疇のみに応用可能であると考えられていた。しかしながら2000年以降、整数的組合せ論（Additive Combinatorics）における数多くのインパクトのある結果、離散数学、統計学分野における「Graph Limit (Graphon)」の理論発展、そして理論計算機科学分野の「Property Testing (性質検査)」の発展は、すべてSzemerédiのRegularity Lemmaから始まり、その背景にある「アイデア」と「精神」を引き継いでいる。そしてTao, Gowers, Alonなどの数学界の重鎮達は、Szemerédiの深遠な貢献を大絶賛してきた。

SzemerédiのRegularity Lemmaは、発表当時、彼の見識性と深遠な貢献に数学界が追いつけなかったのだろう。また彼自体、非常に控えめな人格であり、講演も決してうまいわけではない（大半の聴衆は、5分とついていけない。また彼の書いた文字はほぼ解読不能）。しかし2000年以降Szemerédiの先見の明と深い洞察に気づいた数学界はSzemerédiに対して最高級の栄誉を送っている（彼が受賞した国際賞の数々は、ほとんどが2008年以降である）。そして2012年、Szemerédiは「アーベル賞（数学界のノーベル賞）」を受賞した。

以上の例のように、昔の仕事の中には先人の「英知」が詰まってい



国立情報学研究所 情報学プリンシプル研究系 教授 / 所長補佐
ビッグデータ数理国際研究センター センター長

専門分野は離散数学におけるグラフ彩色問題、グラフ構造理論とアルゴリズム、ネットワークフローとパス問題。主な受賞に、日本数学会2015年度日本数学会春季賞 [2015]、第9回日本学術振興会賞、第9回日本学士院学術奨励賞 [2013]、ACM-SIAM Symposium on Discrete Algorithms (SODA13) Best Paper Award [2013]、公益財団法人船井情報科学振興財団 船井学術賞船井哲良特別賞（アルゴリズム的グラフマイナー理論の研究とその応用） [2011] がある。

る。また彼らの先見性は、その仕事をじかに体験しなければ理解できない。ある程度キャリアを積んだ自分は、今後先人の仕事に敬意を払いつつ、先人の「英知」を再ブレイクさせたいと思う。

現在は、研究費などの関係から成果をすぐに求められることが多い。また「Deep Learning」などの分野は数週間、数か月単位での研究サイクルが必要で、瞬間にほとんどの過去の仕事は忘れられてしまう。私のように若くない研究者は、このような分野で生きることにはできないと思うので、今後は今すぐブレイクしなくても、数十年後にブレイクするような研究に取り組みたいと思う（長生きしないと自分の仕事の行く末は見ることができない）。

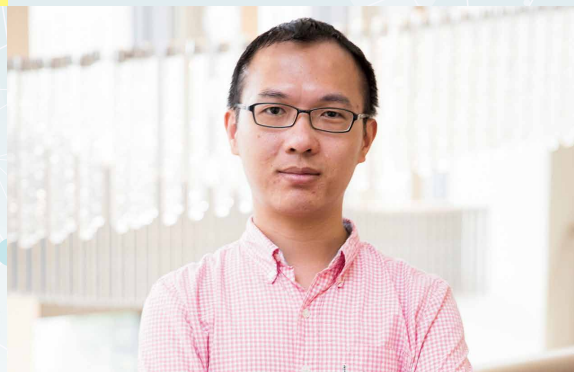
最後に、自分の研究を引用してくれた、また参考にしてくれた研究者にも感謝するべきであろう。Szemerédiは「自分の仕事の評価されたのは、多くの研究者が自分の研究を生かしてくれたからだ。その研究者たちに感謝したい」とアーベル賞の受賞時に述べている。



2017.2.27 成果報告会2017（学術総合センター）

研究紹介

パラメータ化近似 不可能性に対する 新しい証明手法の開発



グラフ・ネットワークにおける理論と最適化グループ

Bingkai Lin

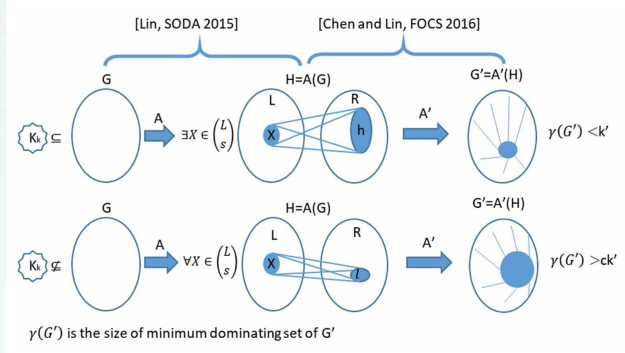
ピンカイ リン

国立情報学研究所 特任研究員

2016年3月、東京大学大学院情報理工学系研究科 博士課程修了。2016年4月より現職。最適化問題のパラメータ化近似不可能性の研究に従事。

計算困難な問題を扱うための理論として、パラメータ化計算量理論と近似アルゴリズム理論は両者ともに重要な役割を果たしています。たとえば、頂点被覆問題という組合せ最適化問題はNP困難と呼ばれるクラスに属しており、グラフの頂点数 n の多項式 $\text{poly}(n)$ に依存する計算時間で解くことは難しいと考えられています。しかし、サイズ k 以下の解が存在するかは $2^k \text{poly}(n)$ 時間で判定できますし、最小サイズの2倍以内の大きさの近似解を $\text{poly}(n)$ 時間で求めることが可能です。パラメータ化近似計算は上の2つのアプローチを合わせたものです。たとえば支配集合問題は、頂点被覆問題とは異なり、どんな関数 f を用いてもサイズ k 以下の解の存在を $f(k) \text{poly}(n)$ 時間で判定できません。また、最小サイズの定数倍以内の解を $\text{poly}(n)$ 時間で求めることもできません。そこで2つのアプローチを合わせ、 $2^k \text{poly}(n)$ 時間で最小サイズ k の2倍以下の支配集合を求められるか、という疑問が自然に湧いてきます。

私の研究では、計算量理論の妥当な仮定のもと、このような計算をするアルゴリズムが存在しないことを示しました (FOCS2016)。これは二部クリークのパラメータ化近似不可能性の証明手法 (SODA2015) の枠組みを拡げることで得られています。最終的には、より多くの最適化問題のパラメータ化近似不可能性を証明するための体系的な手法の構築を目指して、研究に取り組んでいます。



高次元データにおける 確率密度関数の 幾何的な構造の推定



グラフマイニング & WEB&AI グループ

大田 浩史

Hirofumi Ohta

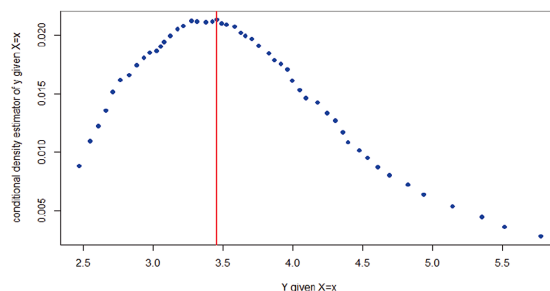
リサーチアシスタント / 東京大学大学院 修士課程

2016年4月から、東京大学大学院経済学研究科 統計学専攻に在学。2016年12月よりRAとして河原林巨大グラフプロジェクトに参加。高次元統計学およびノンパラメトリック推定の研究に従事。

近年、ビッグデータと呼ばれる高次元構造を有するデータが得られることが多くなり、その解析手法の確立が急務となっています。その中でも私は、確率密度関数の幾何的な構造の推定に着目して研究を進めています。確率密度関数の推定の際、カーネル密度推定と呼ばれる方法がよく利用されますが、確率変数の次元の増加に伴ってその推定精度が悪くなることが知られており、この現象は統計学では「次元の呪い」と呼ばれます。

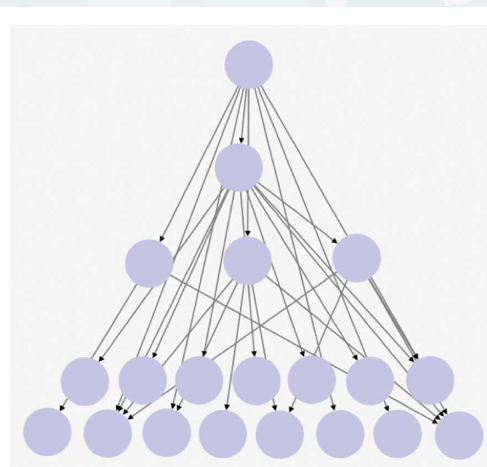
私の研究では、(条件付き) 確率密度関数を、分位点関数を経由してモデリングすることで次元削減を行い、「次元の呪い」を緩和して推定精度を改善する方法を提案しています。また、確率密度関数の幾何的な構造の推定に関して、現在は特に最頻値に着目して研究を進めています。確率密度関数の極大値を与える確率変数の値のことを最頻値といい、統計学において推定対象となる一つの特性値です。また、最頻値は、期待値では捉えられない情報をもっていることが知られています。他にも、確率密度関数にまつわる我々が知りたい幾何的な構造は色々あるため、提案方法を利用したその推定法の研究の発展が期待されます。

条件付き密度関数と条件付き最頻値の推定



人を含めた生物は、分子・細胞・個体・集団・生態系といったあらゆる階層において要素が互いに作用し合う広い意味での社会を築き、生命としての機能を発揮するシステムであると言えます。近年、そのような社会における要素同士の関係性をネットワークとして表現し、その構造を特徴づけることで生命システムを理解する手法が発展したため、生命科学においてネットワークという視点は重要性を増してきています。

私は実際に、アリの巣における個体間の関係性をネットワークとして解析し、行動の数値モデルを構築することで、アリだけでなく、人を含めた動物の社会に共通の構造を明らかにしています。もう一つの研究として取り組んでいるのは、個体や細胞、分子がどのように動くかという問題です。そもそも、ネットワークの要素が作用し合うた

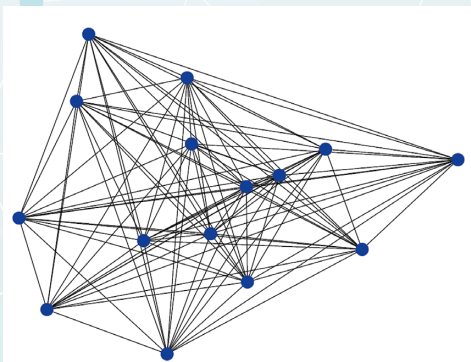


アリのネットワーク

めには、空間の中で要素同士が会うことが必要です。そこで、要素がどのように動き、他の要素と出会うかという問題にも取り組み、最適な遭遇率をもたらす動きのパターンを明らかにしています。これらの研究を通して生命現象を理解し、私たちの実生活や社会に役立つ法則を見つけたいと思っています。

私は幾何学的な「もの」を扱うアルゴリズムの設計に興味をもって研究を行っています。とくに、点集合、直線の集合、およびそれらの間の距離のような幾何的な構成要素に関する問題に取り組んでいます。私の研究目的は、そのような問題を解くためのアルゴリズムを構築する上で有用な幾何的性質を見出すことです。

本プロジェクトで扱う巨大グラフには、有用な幾何情報が大量に含まれています。例として、巨大なアドホックネットワークを用いてメッセージを送る状況を考えます。互いに近いノード間での通信は容易で、メッセージを短い時間で直接送信可能です。一方、遠いノードの間でメッセージを送信する際には、ノード同士でのメッセージの伝達を繰り返す必要がありますが、その回数をできる限り少なくすることが望まれます。また、ネットワーク上の指定された場所に、非常に多くのノードと接続しているハブが配置されており、ハブの間では、高速に通信することが可能です。したがって、遠いノードの間でメッセージを短い時間で送信するためには、如何にしてハブを有効活用



通信ネットワークのイメージ

するか、ということを検討する必要があります。以上の問題では、ノードの位置情報、ノード間の距離、およびノードをつなぐネットワーク上の最短パスなどのような幾何構造が現れ、その解析が重要となります。私は問題に対する効率のよいアルゴリズムの構築に向けて、問題に現れる幾何構造の研究を進めています。

生物の社会を ネットワークの 視点から理解する



複雑ネットワーク・地図グラフグループ

阿部 真人

Masato S. Abe

国立情報学研究所 特任研究員

2015年3月、東京大学大学院総合文化研究科 博士課程修了。
2015年4月より現職。生物ネットワーク、複雑系に関心をもつ。

アルゴリズム構築に 有用な幾何学的 性質を探る



ネットワーク・アルゴリズムグループ

Matias Korman

マティアス コルマン

研究協力者 / 東北大学 助教

2009年9月、東北大学大学院情報科学研究科 博士課程修了。その後、4年間ヨーロッパの大学での研究員を経て、2013年に東北大学に戻り、巨大グラフ内の最短路のアルゴリズムについて研究を続ける。

リサーチアシスタントたちは今… 3

2012年10月のプロジェクト発足以来、現在まで述べ73人（現役16名、OB・OG 57名）もの学生の皆様に Research Assistantまたは Project Assistantとして参加していただいています。OB・OGの中から4人の方にコメントをいただきました。

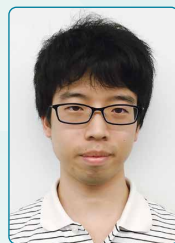


丸茂 直貴

現在：NTTコミュニケーション科学基礎研究所 研究員
旧：グラフ・ネットワークにおける理論と最適化グループ RA
(RA期間：2014年6月～2016年2月)

私は現在NTT研究所で、数理最適化とその機械学習への応用について研究しています。具体的には、スパース最適化問題

等、機械学習で現れる様々な非凸制約付き連続最適化問題を、凸緩和せずに解く手法について研究しています。機械学習分野の国際会議に論文が採択されることを目標として研究していますが、河原林プロジェクトの毎週のセミナーや日々の議論で得た関連分野の幅広い知識・経験は、研究テーマを決める上でも、研究を遂行する上でも直接役立っています。また、ERATO感謝祭の発表者として声をかけていただくなど、業務で本プロジェクト関係者と関わることも多く、ここでの人脈が今も活きています。一流の研究者と議論・交流ができるプロジェクトは、卒業後も研究に携わることになる学生にとって大変ありがたい環境であったと感じます。今後もプロジェクトのOBというアドバンテージを活かし、数理最適化・機械学習分野の研究やその応用に少しでも貢献したいと思っています。



林 孝紀

現在：(株)リクルートホールディングス Indeed エンジニア
旧：複雑ネットワーク・地図グラフグループ RA
(RA期間：2014年5月～2016年2月)

私は現在、世界最大級の求人情報検索サービスを扱うIndeedという会社で、ソフトウェアエンジニアとして働いています。社内

では検索結果の品質向上に取り組むチームに所属し、それぞれの求人がどのような仕事内容に対する募集なのかを分類するシステムの開発に携わっています。仕事内容の分類結果は、ユーザの検索クエリに対する求人情報のマッチングなどに使用されており、適切な分類を行うことは検索品質を高める上で重要です。日々更新される数千万件の求人情報を扱うため、データの性質に合わせて適切なアルゴリズムを選択して高速に処理を行う必要があります。そのような課題に取り組むときに、河原林巨大グラフプロジェクトに参加して、大規模ネットワークに対してデータの性質を利用した高速なアルゴリズムを設計・実装した経験が糧となっています。検索サービスは様々なデータ構造とアルゴリズムの上に成り立っているので、プロジェクトから学んだものを活かして優れたプロダクトを開発していきたいです。



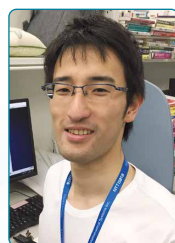
高濱 隆輔

現在：株式会社リクルートライフスタイル
旧：グラフマイニング&WEB&AIグループ RA
(RA期間：2015年5月～2017年2月)

修士課程での2年間、河原林巨大グラフプロジェクトにRAとして在籍していました。大学院を卒業後は株式会社リクルートライフ

スタイルに就職し、データ解析技術を活かした事業の改善などの業務を行っています。

河原林巨大グラフプロジェクトは理論寄りの研究者・学生の方が多く所属されている組織ですが、同時に企業との共同研究にも力を入れており、私自身もヤフー株式会社との共同研究という形で勤務させていただきました。現職では研究とビジネスの境界領域ともいえる業務に携わっていますが、RAでの共同研究の際にビジネス的な知見に基づくアドバイスを頂いた経験が活きていると感じています。今後は、大学院での研究、またRAを通じて多様な研究分野の研究者の方と関わったことで培われた学術的な知識を活用し、それらをビジネスと融合していくことで社会に価値を生み出すことができる人間へ成長していきたいと思っています。



豊岡 祥

現在：株式会社NTTデータ数理システム 研究員
旧：グラフ・ネットワークにおける理論と最適化グループ RA
(RA期間：2015年5月～2017年2月)

当社では、「数理科学とコンピュータ・サイエンスの融合」というテーマを掲げて数理最適化やデータマイニングを活用した事業を中

心に行っております。最近ではやはり機械学習を応用した仕事が多いのですが、周囲には最新の研究に通じている社員が多く、非常に刺激的です。そのような中であって、河原林プロジェクトで毎週のセミナーや成果発表会で色々な分野の方の研究に触れたり、研究会で多くの先生・先輩方と議論した経験は今の自分の血肉となっていると強く感じます。「どのような意義のある研究なのか」「よりよい研究になるためには何が必要か」といった観点は、ビジネスのシーンで研究を活用する上でも非常に大切なことです。

最近では新人研修の一環として、積付け問題 (rectangle packing) という古典的な課題に取り組んでいます。先輩から「世界レベルの結果を」というアツい命令を頂いて、日々アルゴリズムの改良に取り組んでいます。加速度的な進展を見せる数理科学の分野の研究を社会に還元していくべく、幅広い分野にアンテナを張って勉強を続けていきたいと思っています。

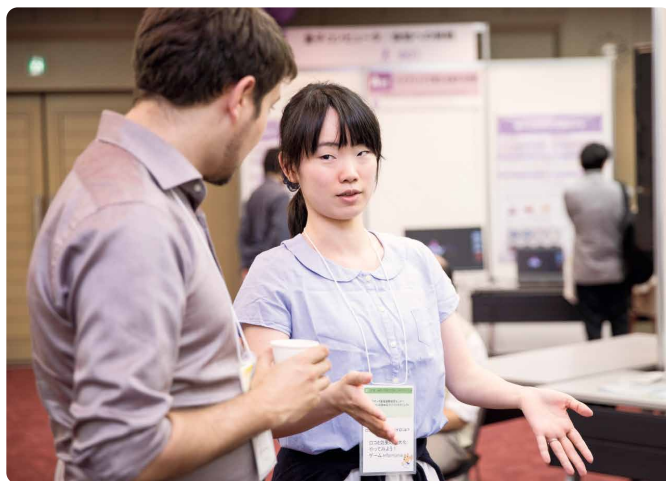
主要な研究成果 2017年3月～2017年8月 国際会議発表論文

論文名	著者	会議名
Portfolio Optimization for Influence Spread	Naoto Ohsaka, Yuichi Yoshida	WWW2017, Perth, Australia, April 3-7
Exact Computation of Influence Spread by Binary Decision Diagram	Takanori Maehara, Hirofumi Suzuki, Masakazu Ishihata	WWW2017, Perth, Australia, April 3-7
Coarsening Massive Influence Networks for Scalable Diffusion Analysis	Naoto Ohsaka, Tomohiro Sonobe, Sumio Fujita, Ken-ichi Kawarabayashi	SIGMOD2017, Chicago, IL, USA, May 14-19
Landmark indexing for Evaluation of Label-Constrained Reachability Queries	Lucien Valstar, George Fletcher, Yuichi Yoshida	SIGMOD2017, Chicago, IL, USA, May 14-19
A Weighted Linear Matroid Parity Algorithm	Satoru Iwata, Yusuke Kobayashi	STOC2017, Montreal, Canada, June 19-23
High Dimensional Consistent Digital Segments	Man-Kwun Chiu, Matias Korman	SoCG2017, Brisbane, Australia, July 4-7
On the Average-Case Complexity of MCSP and its Variants	Shuichi Hirahara, Rahul Santhanam	CCC2017, Riga, Latvia, July 6-9
Online Optimization of Video-Ad Allocation	Hanna Sumita, Yasushi Kawase, Sumio Fujita, Takuro Fukunaga	IJCAI2017, Melbourne, Australia, August 19-25
An Improved Approximation Algorithm for the Subpath Planning Problem and Its Generalization	Hanna Sumita, Yuuma Yonebayashi, Naonori Kakimura, Ken-ichi Kawarabayashi	IJCAI2017, Melbourne, Australia, August 19-25

主要受賞 2017年3月～2017年8月

受賞年月	受賞者	受賞名	主催団体名
2017年4月	吉田 悠一	平成29年度 科学技術分野の文部科学大臣表彰 若手科学者賞	文部科学省
2017年6月	小林 佑輔	Best Paper Award	The 49th ACM Symposium on Theory of Computing (STOC2017)

イベントの様子



2017.6.9～10 国立情報学研究所オープンハウス2017 (学術総合センター)



河原林巨大グラフプロジェクト
Kawarabayashi Large Graph Project

発行：大学共同利用機関法人 情報・システム研究機構 国立情報学研究所 ビッグデータ数理国際研究センター
〒101-8430 東京都千代田区一ツ橋2-1-2 TEL：03-4212-2664
Mail：kk-erato-jimu@nii.ac.jp ホームページ：http://bigdata.nii.ac.jp/
facebook：https://www.facebook.com/kk.erato