様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科研費 KAKENHI

科学研究費助成事業　　研究成果報告書

平成 ２６ 年　 ５ 月 ２７ 日現在

機関番号：１２６０８
研究種目：基盤研究(B)
研究期間：2010～2013
課題番号：２２３０００３
研究課題名（和文）疎グラフ分割問題の研究：統計力学ｖｓ理論計算機科学

研究課題名（英文）Graph bisection problem: approaches from statistical mechanics and theoretical computer science

研究代表者
樺島　祥介（Kabashima, Yoshiyuki）

東京工業大学・総合理工学研究科（研究院）・教授

研究者番号：８０２６０６５２

交付決定額（研究期間全体）：（直接経費）　14,000,000円、（間接経費）　4,200,000円

研究成果の概要（和文）：スピングラス模型など変数間に複雑な依存関係を有する大自由度統計モデルの解析を通じて発展した統計力学の方法が，効率的な近似アルゴリズム，強力な解析手法として情報科学において注目されている．本研究では，統計力学，理論計算機科学の双方において一定の知見が蓄積されている疎なグラフの分割問題に着目し，特に有力な近似解法として知られているスペクトラル法の根拠について，統計力学および理論計算機科学的な方法で分析を行った．また，グラフ分割問題の実問題への応用として単語の語感分類問題を取り上げ，統計力学的な知見にもとづき既存方法の性能改善を行った．

研究成果の概要（英文）：Statistical mechanics has developed various techniques for analyzing large scale and complicated statistical models through research on complex physical systems such as spin glasses. Recently, much attention has been paid for such methodologies as foundations of efficient approximate algorithms and powerful analytical techniques in information sciences. The main objective of this research project is to deepen the understanding of the effectiveness of such methodologies by collaboration of a physicist and a theoretical computer scientist on a concrete problem. The problem that we focused on was graph bisection problems. We particularly examined the origin of the effectiveness of the ``spectral method'', which is known as a dominant approximate solver for the bisection problem. As a practical application of the graph bisection problem, we also improved an exiting method for the extraction of polarity lexicon from word networks utilizing the knowledge of statistical mechanics.

研究分野：情報学

科研費の分科・細目：情報学・情報学基礎

キーワード：キャビティ法　グラフ分割問題　スペクトラル法　ランダム行列　レプリカ法

１．研究開始当初の背景

情報科学における不規則系の統計力学の有用性が注目されるようになって久しい．特に 2000 年以降，統計力学の方法は符号理論，制約充足問題，無線通信理論などにおいて従来研究では得がたい画期的な“成果”を次々と生み出し，もはや従来分野が無視できない存在となっている．これに伴い，国内・国外において組織的な研究プロジェクトが相次いで設定されており，当面この研究潮流が衰えることはないものと予想される．

さて，このように統計力学の方法は情報科学において様々な注目すべき“成果”を挙げてきた一方で，必ずしもそれらがすべて従来分野で受け入れられているわけではない．経験科学である物理の理論として発達した統計力学は「仮説の提示と実験による検証」を基本的な手段としている．一方，情報科学の従来理論は伝統的に数学との結びつきが強く「公理からの演繹」を重視する形式科学の趣が強い．そのため，統計力学の“成果”は情報科学においては"結果"ではなく数学的証明が与えられていない“予想”としか受け取られないことが多いのである．統計力学の方法が浸透するほど，この認識のギャップに対するフラストレーションは高まっており，その解決へ向けた努力の重要性が増している．

２．研究の目的

本研究では，特に，理論計算機科学と統計力学との境界に焦点を当てる．それぞれの方法論のギャップを埋めるためには，共通の問題を共同で議論しながら相互理解を深めることが有効であると考えられる．こうした観点から，本研究では統計力学，理論計算機科学の双方において一定の知見が蓄積されている疎なグラフの分割問題に着目し，両分野の研究者が共同で未解決課題に取り組む．このことを通じて，個別課題に関する新しい知見を得るとともに，統計力学に対してはしばしば“物理的直感”に頼りがちなその有効性の起源に関する明確な根拠を，理論計算機科学に対しては既存研究の延長上にはない新しい発想に基づいたアルゴリズム，解析法の発展をもたらすことを目的とする．

３．研究の方法

研究の端緒として，planted solution 型のグラフ（等）分割問題を取り上げた（図１参照）．2N 個のノードに対し１ノード当たり平均 C 本のエッジが他のノードに張られたグラフが与えられる．等分割問題とは，与えられたグラフに対して切るエッジの本数がなるべく少なくなるように 2N 個のノードを N 個ずつのグループに分割することである．planted solution 型の問題では $i=1,2,\cdots,N$ をグループ A，$i=N+1,N+2,\cdots,2N$ をグループ B に分割することが正解であるとする．与えられるグラフはこの正解に基づいて，同一グループ内のノードは確率 $p=C(1+\Delta)/N$ で，異なるグループ間のノードは確率 $r=C(1-\Delta)/N$ でそれぞれエッジが張られているものとする（$0<\Delta<1$；このようにすると１ノードあたりの平均エッジ数は C になる）．

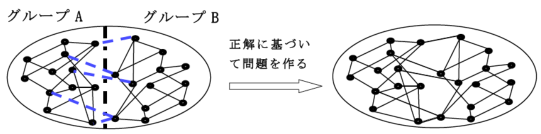

図１：planted solution 型グラフ分割問題

情報科学で特に有用とされる統計力学の方法には i)レプリカ法，ii) キャビティ法，iii)有限サイズスケーリングの３つがある．上述の分割問題に関して我々は，レプリカ法，キャビティ法に基づいた予備的研究から $\Delta<C^{-1/2}$ である場合 $N\to\infty$ の極限ではどのような分割アルゴリズムを用いても正解に関する情報が全く得られないという“予想”を得ている．また，有限の N に関する数値実験のデータについては臨界条件 $\Delta c=C^{-1/2}$ の周辺で N の適当なべき乗を用いてスケーリングしなおせば異なるシステムサイズのデータがすべて同一の曲線上に乗る有限サイズスケーリング仮説の成立を強く示唆する結果も得ている．一方，同一の課題に対して，理論計算機科学でもいくつかの結果が知られており，それらの関係付けが当面の課題となる．

グラフ分割問題の有力な近似解法に，グラフを特徴付ける行列の最大固有ベクトルを利用するスペクトラル法という手法がある．予備的研究から，C が十分大きな場合にはキャビティ法によりスペクトラル法の解析が可能であることが分かった．統計力学の観点からの研究としてはこの結果を $C\sim O(1)$ で各ノードに対して一定の場合，統計的に揺らぐ場合にそれぞれ拡張することが課題となる．加えて，それらの結果，手法の妥当性を理論計算機科学的に検討する．

４．研究成果

（１）第１固有値／固有ベクトルの分析

研究を開始し，議論を進めた結果，グラフ分割問題の分析における中心的課題は，グラフの隣接関係を表現する行列（隣接行列）に関する最大（第１）固有値に対する固有ベクトル（第１固有ベクトル）の形状の評価にある，との認識に至った．そこで，当初予定していた planted solution 型のグラフ分割問題直接ではなく，モデルの本質に焦点を絞りより一般化したランダム疎行列の第１（最大）固有値およびその固有ベクトル（第１固有ベクトル）の詳細な分析に取り組んだ．

まず，次数が一定のランダム疎行列については比較的簡単な考察により，第１固有値と固有ベクトルの形状に関する知見が得られる．一方，次数が分布する場合の解析は一般に複雑化し難しい．そこで，次数のばらつきという特徴は有しながらなるべく単純なモ

デル化を行うという視点から，２種類の次数 c_1 と $(c_1 \le) c_2$ およびそれらを有するノードの割合 p_1，p_2 を与え，それを満たすグラフのアンサンブルを考える．また，グラフ分割の可能性を制御する特徴に対応させて，各リンクに付随する行列要素 J_{ij} は確率 $(1+\Delta)/2$ で+1 を $(1-\Delta)/2$ で-1 の値とるものとする．このような生成法にしたがい得られる行列 $J=(J_{ij})$ について，第１固有値がどのように定まり，第１固有ベクトルの形状がどのように変化するのかについて，$N\to\infty$ の極限において統計力学のレプリカ法およびキャビティ法を用いて調べた．以下，得られた結果のうち主要なものを記す．

① $c_1=c_2=c$ の場合
$\Delta>\Delta c=(c-1)^{-1/2}$ ならば $\Lambda=(c-1)\Delta+1/\Delta$，$\Delta<\Delta c$ ならば $\Lambda=2(c-1)^{1/2}$ である（図２(a)）．第１固有ベクトルは $\Delta>\Delta c$ のとき，$(+1,+1,\cdots,+1)$ と有限の方向余弦 M を持つ（図２(b)）．このことは第１固有ベクトルにもとづいてグラフ分割が可能になることに対応している．つまりスペクトラル法が有効に働く．一方，$\Delta<\Delta c$ ならば第１固有ベクトルはランダムな方向を向く．いずれの場合も，第１固有ベクトルの各成分はすべて同程度の大きさを持つ“広がった”固有ベクトルである（図２(c)）．Δc の周りでは，$M=N^{1/6}g(N^{1/3}(\Delta-\Delta c))$ という有限サイズスケーリング仮説が成立することも実験的に確認した．

② $c_2>c_1$ かつ c_2 のノードが１つの場合
$c_2>2(c_1-1)$ であれば十分に小さな Δ について，$c_2>=c_1(c_1-1)$ であればすべての Δ に対して，①に示した第１固有値は $\Lambda=c_2/(c_2-c+1)^{1/2}$ に置き換えられる（図３(a)参照）．さらに，対応する第１固有ベクトルは O(1)個の成分に大きさが集中した“局在化した”固有ベクトルとなる（図３(b)参照）．局在化した固有ベクトルではほとんどすべての成分がゼロとみなせるほど小さい．つまり，第１固有ベクトルが局在化した形状を有する場合は，その成分の符号のみを手がかりとしてグラフ分割を行うスペクトラル法は有効に働かないことを示唆している．

③ $c_2>c_1$ かつ c_2 のノード数が O(N)の場合
次数 c_2 のノードの割合が多くなるにつれて第１固有値の値は単調に増加し $2(c_2-1)^{1/2}$ に漸近する．ただし，システムサイズ依存性が大きく有限の系については $2(c_2-1)^{1/2}-O((\ln\ln N)^{-2})$ のようにシステムサイズ N に対して非常にゆっくりとした補正が加わるためこのことを数値的に確認するのは非常に困難である（図４(a)）．また，第１固有ベクトルは次数 c_2 のノード数の増加につれて“局在化した形状”→“広がった形状”へ変化する（図４(b)）．

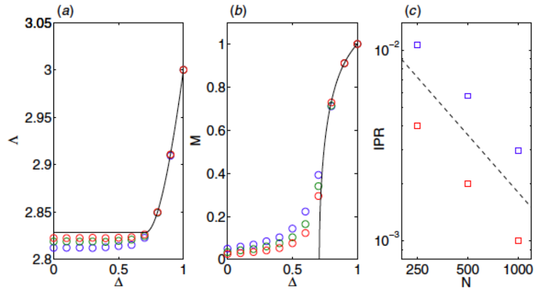

図２：c=3 の場合の(a):第１固有値，(b):(+1,+1,⋯,+1)との方向余弦，(c):第１固有ベクトルの“局在化”の度合い．いずれも論文⑥より転載．

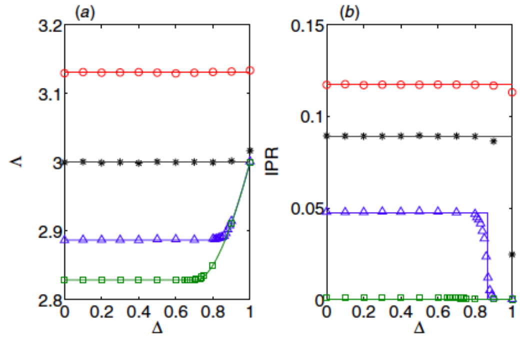

図３：単一次数 $c_1=3$ のグラフの中に１つだけ高い次数 $c_2=4$〜7 のノードを追加した場合．(a)：第１固有値と Δ との関係．(b)：第１固有ベクトルの局在の度合いと Δ との関係．いずれも論文⑥より転載．

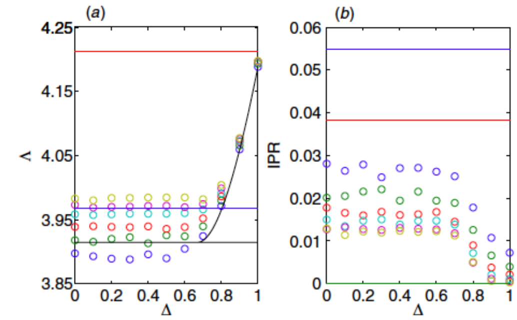

図４：$c_1=3$，$c_2=7$，$p_1=1-p_2=0.9$ の場合に N=1000,2000,4000,8000,16000,32000 と変化させた場合．(a)：第１固有値．下から順に N=1000→N=32000 に対応．(b)：第１固有ベクトルの局在化の度合い．上から順に N=1000→N=32000 に対応．いずれも論文⑥より転載．

以上の結果は，グラフ分割の有力な（近似）解法として知られているスペクトラル法はグラフ内の次数のばらつきにより，その有効性が大きく影響を受けることを意味しており，近似解法の有用性の範囲を定量的に特徴づける指針として重要である．

なお，これらの成果は，不規則系の統計力学の方法にもとづいて得られたものであり，そのままでは数学的な厳密性は担保されない．そこで，理論計算機科学的アプローチにしたがって数学的な厳密化を図った．その結

果，①，②のΔ=1 の場合について厳密な証明を与えることに成功した．多数の要素が複雑に絡んだ対象は組み合わせ論的な複雑さが増大するため，数学的に厳密な方法のみでは見通しよく結果を得ることが難しい場合が多い．我々は，必ずしも厳密性が保証されてはいないが系統的な手続きで結果が得られる統計力学の方法によって“予想”を押さえておき，その上で数学的厳密化を図るという戦略を取った．こうしたアプローチは，大自由度の複雑システムを数学的に分析する際の有効な方法論の雛形になるものと期待される．

（２）高精度な評価表現辞書構築の試み
インターネット環境の急速な普及を背景として，ブログなどに書かれる商品やサービスに対するユーザの評価を自動的に分析し，マーケティングなどに役立てる研究が自然言語処理の分野で活発化している．その際の要素技術に，文中に存在する単語の語感（肯定的あるいは否定的）を分類する評価表現辞書の構築がある．

先行研究（Takamura et al, Proc. ACL’05, pp. 133-140 (2005)）では，i)単語の語感を統計力学のイジングスピンと対応させ，３種類の辞書によって相互作用を導入し ii)good または bad など肯定的／否定的な意味が明らかな数個の単語（種単語）に外部から強制的な極性を与えて適当な温度でのイジングスピンの熱平均を近似的に求める，ことで高精度の語感分類が可能になることを示している．同時に，分類精度が最大になる温度よりも少し低温にするとシステムはいわゆる強磁性体相転移を示し，精度は劇的に劣化することも報告されている．

この語感分類の問題は辞書にもとづいて構成された単語のネットワーク（グラフ）の分割問題にほかならない．本課題では，統計力学の知見にもとづいて，こうした方法の背後にある数理的メカニズムを調べ，それを手法の性能改善および有用性の分析に役立てることを試みた．以下，主要な結果を記す．

① 手法の性能改善
統計力学の平均場理論を用いることにより，相転移は相互作用行列の第１固有ベクトル方向への揺らぎの不安定化により生じることが示される．そこで，先行研究によって提案されている相互作用モデルからその第１固有ベクトルの成分を差し引く方法を提案した．その結果，約８万語の英単語のネットワークに対し，正負の意味を表す種単語が２つの場合に語感を正しく分類する割合が 75.2%から 84.5%に向上するなど，評価表現辞書の精度を大幅に改善することに成功した．対象としている語彙ネットワークは平均次数が１９程度の疎なネットワークであるため，第１固有ベクトルはべき乗法により低計算量で求めることができる．よって，我々の提案した改善法の実装に必要な計算コストは先行研究の方法とほとんど変わらない．

② 有用性の起源の分析
①の結果から，第１固有ベクトルを取り除いた系を考えることにする．ランダムグラフのグラフ分割に関するスペクトラル法の有用性から，グラフを特徴づける結合行列の第２固有ベクトルが分割に関する有用な情報を多く含んでいるものと期待される．しかしながら，実際に観測される固有ベクトルの形状は局在しており分割に有用な情報は少ないことがわかった．

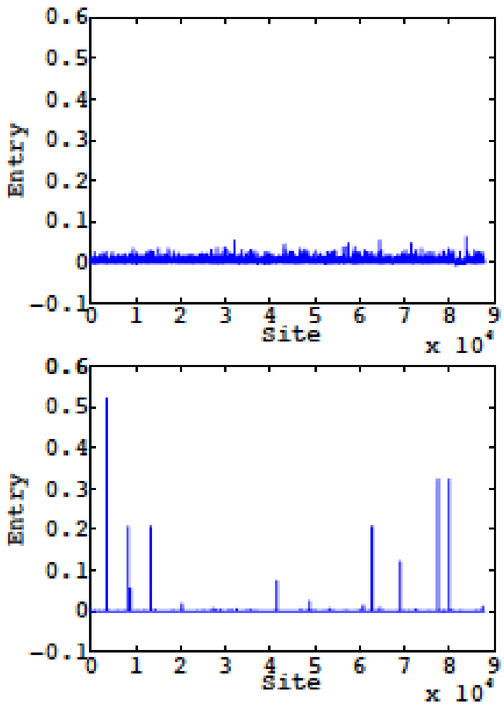

図５：第１固有ベクトルの形状（上）．第２固有ベクトルの形状（下）論文⑤より転載．

そこで，上位１００個の固有ベクトルの形状を更に調べた結果，その多くが局在化していることが判明した．このことは，ランダムグラフに対し知られているグラフ分割の知見が，実際的な問題では必ずしも有用ではないことを示唆している．とはいえ，イジングスピンによる評価表現辞書構築は実験的に有用性が確認されており，何らかのしくみが働いているはずである．このしくみを解明するため，上位１００個の固有ベクトル各々に対して，大きな成分に対応する単語の組み合わせを調べた．例として，第２～第４固有ベクトルに関する結果を以下に記す．

・第２固有ベクトル：
{anniversary, tercenterary, tercentennial, …}
・第３固有ベクトル：
{adieu, adios, arrivederc, …},
・第４固有ベクトル：
{inpour, inpouring, inrush, …}

これらは，それぞれ“…周年”,“さようなら”,“流入”という概念に対応していると思われる．同様に，第５固有ベクトル以下で大きな成分を有する単語の組み合わせは，それぞれ，“計数”,“居住”,“靴”,“植物”,“偵察”,“神”,…といった概念に対応していた．つまり，i)大きな固有値に対する固有ベクトルの多くは各単語の上位概念に対応していること，ii)各概念は直接的には正負の意味に対応していないことがわかる．このことと，高い語感分類の精度が多数の固有ベクトルに無視できない重みが分散される“高温相”で得られるという実験結果を組み合わせると，イジングスピンモデルにもとづいた単語の極性分類が示す高い性能は，種単語に与えられる極性の各単語への（非線形）応答にもとづいているのではないかと考えられる．この“仮説”の妥当性を確かめるために，種単語に与えられる極性からの線形応答のみにもとづいて各単語の極性分類を行った．その結果，改善法ほどではないものの先行研究で得られていた程度の分類性能は得られることが判明し,“仮説”は支持されることとなった．

（３）派生的な成果
グラフ分割問題の分析で発展させた解析技術にもとづいていくつかの派生的な成果も得られた．その中の代表的な成果について簡単に記す．

① 複雑ネットワークの耐性分析への展開
ネットワーク（グラフ）内のノードやリンクを確率的にあるいは選択的に除去すると連結しているクラスタの大きさは減少する．最大クラスタの大きさとネットワークの機能を対応させると，最大クラスタが元のネットワークと同程度の大きさである状況から無視できるくらい小さくなる状況への転移（＝パーコレーション転移）を調べることで，ネットワークの耐性を分析することができる．この問題に第１固有値／固有ベクトルの分析で利用したキャビティ法を応用し，次数相関のあるネットワーク，相互依存型ネットワークなど発展的なネットワークの耐性分析が可能な解析法を開発した．

② MAX-SAT 問題への展開
グラフ分割問題に対する解析手法の応用として，制約可能性問題の解空間の解析を行った．制約可能性問題の中でも，複数の観点から，最も特徴的と言われている XOR-3SAT 問題と CNF-SAT 問題に焦点を合せた．XOR-3SAT 問題では，planted solution model のもとでの，各種伝搬型アルゴリズムの計算限界について明確化した．CNF-SAT 問題では，統計力学的に言われている解空間の構造が，最悪時でも厳密に証明できることを示した．

③ 圧縮センシングへの展開
信号は一般に高次元ベクトルとして表現される．そのなかでも特に典型的に多くの成分がゼロとなる信号は“スパースな信号”とよばれる．近年，スパースな信号に対して，見かけの次元よりも少ない観測値から信号の再構成を行う方法が“圧縮センシング”として注目されている．圧縮センシングにおける信号再構成は，横に長い観測行列による線形変換が与える条件と“スパースである”という拘束条件を組み合わせて線形方程式を解く問題として定式化できる．先行研究では，観測行列はほぼすべての成分がゼロではない密行列を考える場合がほとんどであった．しかしながら，観測や信号復元に要するさまざまなコストを考慮すると，観測行列についてもスパース性を導入することが有利になる場合も多い．スパースな行列は疎結合グラフとして表現できるため，グラフ分割問題の解析で用いた計算手法を応用することができる．こうした観点から，スパースな行列を用いた圧縮センシングの性能評価および信号復元アルゴリズムの開発を行った．

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計２８件）

① S. Watanabe and <u>Y. Kabashima</u>, Cavity-based robustness analysis of interdependent networks: Influences of intranetwork and internetwork degree-degree correlations, Phys. Rev. E 89, 査読有, 012808(1-11), 2014 DOI:10.1103/PhysRevE.89.012808

② <u>O. Watanabe</u>, Message passing algorithms for MLS-3LIN problem, Algorithmica 66, 査読有, 848-868, 2013 http://dx.doi.org/10.1007/s00453-013-9762-7

③ H. Dell, V. Kabanets, D. Van Melkebeek and <u>O. Watanabe</u>, Is Valiant-Varirani's isolation probability improvable?, Computational Complexity 22, 査読有, 345-383, 2013 DOI:10.1007/s00037-013-0059-7

④ N. Nakagawa and H. Yamaguchi, The first eigenvalues of (c,d)-regular graph, IEICE Trans. Information Systems E96-D(3), 査読有, 433-442, 2013 DOI:10.1587/transinf.E96.D433

⑤ T. Goto, <u>Y. Kabashima</u> and H. Takamura, Statistical Mechanical Analysis of Semantic Orientations on Lexical Network, in Proceedings of the 24th International Conference on Compu-

tational Linguistics (COLING2012), 査読有, 977-993, 2012
http://aclweb.org/anthology//C/C12/C12-1060.pdf

⑥ Y. Kabashima and H. Takahasi, First eigenvalue/eigenvector in sparse random symmetric matrices: Influences of degree fluctuation, J. Phys. A: Math. Theor. 45, 査読有, 325001(1-19), 2012
DOI:10.1088/1751-8113/45/32/325001

⑦ D. Lee, Y. Kabashima, K. Takeda, T. Yamada, K. Akabane and K. Uehara, Sparse-Matrix-Based Compressed Sensing for Spectrum Sensing in Flexible Wireless, in Proceedings of the 18th Asia-Pacific Conference on Communications (APCC2012), 査読有, 418-422, 2012
http://www.mlab.t.u-tokyo.ac.jp/attachent/file/256/20120528APCC_final.pdf

⑧ A. Coja-Oghlan, M. Onjo and O. Watanabe, Propagation connectivity of random hypergraphs, The Electronic Journal of Combinatorics 19, 査読有, 1-25, 2012
http://www.combinatorics.org/ojs/index.php/eljc/article/view/v19i1p17

⑨ M. C. Angelini, F. Ricci-Tersenghi and Y. Kabashima, Compressed sensing with sparse, structured matrices, in Proc. Fiftieth Annual Allerton Conference (Allerton House, UIUC, Illinois, USA October 1 - 5, 2012), 査読有, 808–814, 2012
DOI:10.1109/Allerton.2012.6483301

⑩ T. Ando, Y. Kabashima, H. Takahashi, O. Watanabe and M. Yamamoto, Spectral analysis of random sparse matrices, IEICE Trans. Information Systems E94-D(6) , 査読有, 1247-1256, 2011
DOI:10.1587/transfun.E94.A.1247

〔学会発表〕(計１８件)
① Y. Kabashima and S. Watanabe, Cavity-based robustness analysis of complex networks, East Asia Joint Seminars on Statistical Physics, October 22, 2013, Kyoto, Japan.

② Y. Kabashima and S. Watanabe, Cavity approach to complex networks, International Workshop on Phase Transition, Critical Phenomena and Related Topics on Complex Networks, September 9, 2013, Sapporo, Japan.

③ 樺島祥介, ランダム行列とスピングラス理論, 日本物理学会２０１２年秋季大会シンポジウム「ランダム行列の過去・現在・未来」, ２０１２年9月２１日, 横浜国立大学.

④ Y. Kabashima, On the first eigenvalue/eigenvector in sparse random symmetric matrices, School on Large Scale Problems in Machine Learning and Workshop on Common Concepts in Machine Learning and Statistical Physics, August 29, 2012, Trieste, Italy.

⑤ O. Watanabe, Message passing algorithms for MLS-3LIN problem, The 9th Meeting on Analytic Algorithms and Combinatorics (ANALCO'12), January 16, 2012, Kyoto Japan.

⑥ Y. Kabashima, Statistical mechanics approach to the first eigenvalue problem in symmetric random sparse matrices, The 5th International Conference of IMBIC on Mathematical Sciences for Advancement of Science and Technology (MSAST2011), December 18, 2012, Kolkata, India

〔図書〕(計２件)
① 樺島祥介(分担執筆), 森北出版, ランダム行列の数理と科学, 第３章, ２０１４年, ７９～１１２ページ

② O. Watanabe(分担執筆) , World Scientific, Randomness through Computation (H. Zenil ed.), Chapter 24, 2011, pp. 297-308

６．研究組織
(1)研究代表者
樺島　祥介(KABASHIMA, Yoshiyuki)
東京工業大学・大学院総合理工学研究科・教授
研究者番号：８０２６０６５２

(2)研究分担者
渡辺　治(WATANABE, Osamu)
東京工業大学・大学院情報理工学研究科・教授
研究者番号：８０１５８６１７