Final Report for Period: 06/2010 - 05/2011
Submitted on: 08/14/2011
Principal Investigator: Tetali, Prasad .
Award ID: 0701043
Organization: Georgia Tech Research Corp
Submitted By:

Title:
Information Inequalities and Combinatorial Applications

Project Participants

Senior Personnel

Name: Tetali, Prasad
Worked for more than 160 Hours: Yes
Contribution to Project:

Post-doc

Graduate Student

Name: Restrepo, Ricardo
Worked for more than 160 Hours: Yes
Contribution to Project:

Undergraduate Student

Name: Brown, Harrison
Worked for more than 160 Hours: Yes
Contribution to Project:
Worked on a conjecture on graph colorings and made significant progress.

Technician, Programmer

Other Participant

Research Experience for Undergraduates

Organizational Partners

Other Collaborators or Contacts

Mohsen Bayati, Stanford University, Palo Alto, CA
Robert Clayton Jr., Georgia Tech, Atlanta, GA
David Gamarnik, MIT, MA
Satoru Iwata, Kyoto, Japan
Mokshay Madiman, Yale University
Ravi Montenegro, University of Massachusetts, Lowell, MA
Prasad Raghavendra, Georgia Tech, Atlanta, GA
Allan Sly, UC Berkeley, CA
David Steurer, Cornell University

Activities and Findings

Research and Education Activities:

Research Contribution
The research is funded by DMS-0701043 individual grant for the proposal titled, 'Information Inequalities and Combinatorial Applications,' for the period 7/1/07 -- 6/30/2011. The chief contribution during the past year of support is summarized below, which is followed by aspects of mentoring and outreach.

Educational activities:
Besides the student training, mentioned in the next subsection, the PI gave several tutorial-style lectures during the past few years. The most recent ones include a 1-hour lecture (Modified Logarithmic Sobolev Inequalities) at the Newton Institute of Mathematical Sciences, Cambridge, UK and a 2-hour lecture (on Isoperimetric and Functional Inequalities in Discrete Spaces) at the University of Paris, Marne La Vallee, France.
The PI is schedule to deliver a 3-hour tutorial on Markov chains at the Int'l Conference on Machine Learning with Applications in Honolulu, Hawaii, during December 2011.

Findings: (See PDF version submitted by PI at the end of the report)

Summary of Significant Research Findings
While the work proposed in Section 3.2 of the proposal concerning probabilistic analyses of J.M. Pollard's algorithms for the discrete logarithm problem has recently been completely solved by the PI and collaborators, a further open question raised by Pollard has also been addressed and settled in the journal version [6] of the PI's prior work with Ravi Montenegro. In joint work [1] with M. Bayati and D. Gamarnik, settling open problems raised by Aldous, Bollobas, Wormald, and others in establishing the existence of scaling limits for several combinatorial problems in sparse random graph models was one of the main thrusts of this research activity.
Besides the above, two other projects arose from this grant which are being pursued in the next grant: (i) collaboration with M. Madiman on information inequalities in understanding additive number theoretical questions and (ii) collaboration with Satoru Iwata on submodular optimization problems.

A brief description of the results are summarized in the attached pdf file.

Training and Development:

In June 2009, the PI and Fabio Martinelli organized a Working Group in Markov Chains at Georgia Tech, during which 15-20 participants, including several students, actively tackled a set of hard open problems; the workshop resulted in further collaborations, successful resolution of some of the questions and partial progress on others. The PI is currently advising a graduate student (Ricardo Restrepo), and an NSF postdoctoral researcher Kevin Costello. A couple of joint projects with each of these advisees are currently being written up; the main results and a brief summary will be reported in next year's final report for this grant. Supported by the REU part of the current grant, a talented sophomore (Harrison Brown) is working with the PI and Kevin Costello during the summer months of 2010.

During the course of this project, the PI mentored various young researchers. The undergraduate advisees include: Aisha Arroyo, Brian Benson, Antonio Blanca, Harrison Brown, Robert Clayton Jr., and David Hollis. The PI advised graduate students Teena Carroll, Adam Marcus and Ricardo Restrepo during their Ph.D. program. Carroll is now a

tenure-track professor at St. Norbert College, WI. Marcus is completing a Gibbs instructor position at Yale university, and Restrepo will soon start a postdoctoral position at the University of Toronto in Canada.
The PI mentored NSF postdoc Kevin Costello for the past 3 years, and will start mentoring the NSF postdoc Will Perkins. Costello secured a tenure-track position at the University of California, Riverside, CA. The PI is also mentoring (jointly with Ton Dieker in the School of Industrial and Systems Engineering) a postdoc Jinwoo Shin, associated with the Algorithms and Randomness Center (ARC). Since April 2011, the PI is also the Director of ARC. Besides the above, the PI has several papers with other students and postdoctoral students.

Outreach Activities:

In addition, the PI has worked with the undergrad Antonio Blanca during the academic year 2009-2010. Mr. Blanca received the Judge's First Prize in the College of Computing Undergrad Research competition (at Georgia Tech.) in April 2010 for his two projects with the PI. While one of them was more research oriented, the other was to make teaching of algorithms more eff ective in the classroom -- Mr. Blanca implemented several classical algorithms typically taught in an undergraduate course, with visualizations accompanied by synopses of the fundamental ideas and invariants underlying each algorithmic paradigm. The PI and Mr. Blanca tested them successfully in the classroom during Spring 2010 when the PI was teaching CS 3510, a design and analysis undergraduate course in the School of Computer Science at Georgia Tech.

Based on research reported here, the PI gave expository and other invited lectures at the following meetings:

* Aug 1-5, '11 : Probability Summer School in Mambucaba, Brazil

* June 12-17, '11 : IPAM sponsored Combinatorics Reunion workshop, Lake Arrowhead, CA

* March 28-31 ' 11: Discrete Harmonic Analysis, Newton Institute, Cambridge, UK

* January 11-14 '11: Functional Inequalities in Discrete Spaces, University of Paris, Marne La Vallee, France (2-hour tutorial)

* Nov. 29- Dec. 2 '11: Workshop on Combinatorial Counting, Schloss Dagstuhl, Germany

* June 14-15, '10 : Two lectures in the minisymposia on Additive Combinatorics and Probabilistic Combinatorics, SIAM Conference on Discrete Mathematics, Austin, Texas

* May 29, '10: Methods in Discrete Structures Colloquium, Technical University, Berlin.

* May 17-June 12, '10: Six Lectures on Probabilistic Combinatorics, University of Rome 3, Rome, Italy

* January 12, '10: Joint Math and CS Colloquium, Northeastern University

* November 18, '09: Cal Tech Information Sciences & Technology Seminar

* November 12, '09: UCLA Probability Colloquium

* October 5-9, '09: IPAM Workshop on Probabilistic Techniques & Applications

* September 14-17, '09: Probabilistic Methods in Computer Science, Center for Research in Math, Barcelona, Spain

* August 24-28, '09: BIRS Workshop on Probabilistic and Extremal Combinatorics, Ban ff, Canada

During 2009-2011, besides serving as the Editor-in-Chief of the SIAM J on Discrete Mathematics, Tetali edited a special issue of the Journal of Combinatorics, in honor of Joel Spencer, and a special issue of SIAM J on Discrete Math on Constraint Satisfaction Problems and Message Passing Algorithms. Finally, the PI hosts the combinatorics seminar in the School of Mathematics at Georgia Tech.

Journal Publications

M. Madiman and P. Tetali, "Information Inequalities for Joint Distributions, with Interpretations and Applications", IEEE Trans. on Information Theory, p. 2699--27, vol. 56, (2010). Published,

T. Carroll, D. Galvin, P. Tetali, "Matchings and independent sets of a fixed size in regular graphs", Jour. Combin. Theory (Series A), p. 1219--1, vol. 116, (2009). Published,

M. Madiman, A. Marcus and P. Tetali, "Entropy and set cardinality inequalities for partition-determined functions, with applications to sumsets", Random Structures & Algorithms, p. , vol. , (2011). Accepted,

J-H. Kim, R. Montenegro, Y. Peres, P. Tetali, "A Birthday Paradox for Markov chains, with an optimal bound for collision in the Pollard's Rho for Discrete Logarithm", Annals of Applied Probability, p. , vol. 20, (2010). Published,

E. Croot, A. Granville, R. Pemantle, P. Tetali, "Running Time Predictions for Factoring Algorithms", Proc. of ANTS 2008, Springer Lecture Notes in Math, p. , vol. , (2008). Published,

E. Croot, A. Granville, R. Pemantle, P. Tetali, "Sharp Transitions in Making Squares", Annals of Mathematics, p. , vol. , (2011). Accepted,

C. Borgs, J. Chayes and P. Tetali, "Tight Bounds for Mixing of the Swendsen-Wang Algorithm at the Potts Transition Point", Probab. Th. & Rel. Fields, p. , vol. , (2010). Published,

(M. Sammer and P. Tetali, "Concentration on the Discrete Torus using Transportation", Comb. Probab. & Comp., p. 8, vol. 18, (2009). Published,

B. Benson, D. Chakrabarty and P. Tetali, "G-Parking Functions, Acyclic Orientations and Spanning Trees", Discrete Math., p. , vol. 310, (2010). Published,

C. Heitsch and P. Tetali, "Meander Graphs", Proc. of the Int'l Conf. on Formal Power Series and Algebraic Combinatorics, p. , vol. , (2011). Published,

A. Montanari, R. Restrepo, and P. Tetali, "Reconstruction and Clustering in Random Constraint Satisfaction Problems", SIAM J. on Disc. Math., p. , vol. , (2011). Published,

M. Bayati, D. Gamarnik, and P. Tetali, "Existence of limiting constants using a combinatorial approach to an interpolation technique from statistical physics", Annals of Mathematics, p. , vol. , (2010). Submitted,

P. Tetali, J. Vera, E. Vigoda, and L. Yang, "Phase Transition for the Mixing Time of the Glauber Dynamics for Coloring Regular Trees", Ann. of Appl. Probab., p. , vol. , (2011). Accepted,

R. Montenegro and P. Tetali, "Time until intersection of independent walks on Z", J. Number Theory, p. , vol. , (2011). Submitted,

Books or Other One-time Publications

Web/Internet Site

URL(s):

www.math.gatech.edu/~tetali/RESEARCH/pubs.html

Description:

The papers that are listed in this report are posted at the URL and all the papers acknowledge NSF support. In addition, many of the papers are posted on other preprint servers, such as the ArXiv, further enhancing the dissemination of research supported by this grant.

Other Specific Products

Contributions

Contributions within Discipline:

1. Optimal bounds have been obtained for the first time via rigorous mathematical analysis for two classical algorithms, for solving the discrete logarithm problem in cyclic groups, proposed by J.M. Pollard in 1979. Novel Markov chain analyses was the contribution to the discipline.

2. Tight bounds were provided for the stopping time in the so-called square dependence problem, posed by Carl Pomerance at the ICM in 1994. Methods used tools from analytic number theory, probability, and combinatorics. The paper is recently accepted to the Annals of Mathematics.

3. In the context of sparse random graphs and sparse random regular graphs, we proved that the size of a largest independent set in these graphs, normalized by the number of nodes converges to a limit (with high probability), thus resolving an open problem mentioned by several experts: Wormald '99, Aldous-Steele 2003, Bollob\'as-Riordan 2008, as well as Janson-Thomason 2008. Inspired by work from statistical physics, a novel combinatorial interpolation scheme was introduced and used effectively. This has been in review with the Annals of Mathematics.

Contributions to Other Disciplines:

1. Contributed to the understanding of classical number theoretic algorithms of interest in Computational number theory (for the discrete logarithm and factoring problems) and Cryptography.

2. Contributed to the understanding of the role and the power of entropy inequalities in the topics of probability, information theory and additive number theory.

Contributions to Human Resource Development:

Contributions to Resources for Research and Education:

Algorithms simulator developed by the undergraduate advisee Antonio Blanca is being used at Georgia Tech and UC Berkeley in teaching a standard undergraduate algorithms course.

Contributions Beyond Science and Engineering:

Conference Proceedings

Bhatnagar, N;Sly, A;Tetali, P, Reconstruction Threshold for the Hardcore Model, "SEP 01-03, 2010", APPROXIMATION, RANDOMIZATION, AND COMBINATORIAL OPTIMIZATION: ALGORITHMS AND TECHNIQUES, 6302: 434-447 2010

Bayati, M;Gamarnik, D;Tetali, P, Combinatorial Approach to the Interpolation Method and Scaling Limits in Sparse Random Graphs, "JUN 06-08, 2010", STOC 2010: PROCEEDINGS OF THE 2010 ACM SYMPOSIUM ON THEORY OF COMPUTING, : 105-114 2010

Raghavendra, P;Steurer, D;Tetali, P, Approximations for the Isoperimetric and Spectral Profile of Graphs and Related Parameters, "JUN 06-08, 2010", STOC 2010: PROCEEDINGS OF THE 2010 ACM SYMPOSIUM ON THEORY OF COMPUTING, : 631-640 2010

Sammer, M;Tetali, P, Concentration on the Discrete Torus Using Transportation, "MAY 05-09, 2008", COMBINATORICS PROBABILITY & COMPUTING, 18 (5): 835-860 Sp. Iss. SI SEP 2009

Montenegro, R;Tetali, P, How Long Does it Take to Catch a Wild Kangaroo?, "MAY 31-JUN 02, 2009", STOC'09: PROCEEDINGS OF THE 2009 ACM SYMPOSIUM ON THEORY OF COMPUTING, : 553-559 2009

Bayati, M;Gamarnik, D;Katz, D;Nair, C;Tetali, P, Simple Deterministic Approximation Algorithms for Counting Matchings, "JUN 11-13, 2007", STOC 07: PROCEEDINGS OF THE 39TH ANNUAL ACM SYMPOSIUM ON THEORY OF COMPUTING, : 122-127 2007

Madiman, M;Tetali, P, Sandwich bounds for joint entropy, "JUN 24-29, 2007", 2007 IEEE INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY PROCEEDINGS, VOLS 1-7, : 511-515 2007

Croot, E;Granville, A;Pemantle, R;Tetali, P, Running time predictions for factoring algorithms, "MAY 17-22, 2008", ALGORITHMIC NUMBER THEORY, 5011: 1-36 2008

Kim, JH;Montenegro, R;Peres, Y;Tetali, P, A birthday paradox for Markov chains, with an optimal bound for collision in the Pollard Rho algorithm for discrete logarithm, "MAY 17-22, 2008", ALGORITHMIC NUMBER THEORY, 5011: 402-415 2008

Kim, JH;Montenegro, R;Tetali, P, Near optimal bounds for collision in Pollard Rho for discrete log, "OCT 20-23, 2007", 48TH ANNUAL IEEE SYMPOSIUM ON FOUNDATIONS OF COMPUTER SCIENCE, PROCEEDINGS, : 215-223 2007

Categories for which nothing is reported:

Organizational Partners

Any Book

Any Product

Contributions: To Any Human Resource Development

Contributions: To Any Beyond Science and Engineering

Final report on NSF Grant DMS 0701043
Information Inequalities and Combinatorial Applications

1 Research Contribution

The research is funded by DMS-0701043 individual grant for the proposal titled, "Information Inequalities and Combinatorial Applications," for the period 7/1/07 – 5/31/2011. The chief research contribution during *the past two years of support* is summarized below, which is followed by aspects of mentoring and outreach. Most of the relevant publications, acknowledging the NSF support, are available at the PI's website.

1.1 Summary of Significant Research Findings

While the work proposed in Section 3.2 of the proposal concerning probabilistic analyses of J.M. Pollard's algorithms for the discrete logarithm problem has recently been completely solved by the PI and collaborators, a further open question raised by Pollard has also been addressed and settled in the journal version of the PI's prior work with Ravi Montenegro. In joint work with M. Bayati and D. Gamarnik, settling open problems raised by Aldous, Bollobás, Wormald, and others in establishing the existence of scaling limits for several combinatorial problems in sparse random graph models was the main thrust of this past year's research contribution. The PI's continuing grant pursues this line of research to attack recent open problems raised in this context. For a brief description of the results, see below. Some of the other technical contributions are also described shortly after that.

Technical Details of Research Findings:

• Existence of limiting constants using a combinatorial approach to an interpolation technique from statistical physics.
This is joint work with M. Bayati (MSR-NE postdoc, and Stanford University postdoc) and D. Gamarnik (MIT), and appeared in the Proc. of the ACM STOC 2010.

We establish the existence of free energy limits for several sparse random hypergraph models corresponding to certain combinatorial models on Erdös-Rényi graph $\Gamma(N, c/N)$ and random r-regular graph $\Gamma(N, r)$. For a variety of models, including independent sets, MAX-CUT, Coloring and K-SAT, we prove that the free energy both at a positive and zero temperature, appropriately rescaled, converges to a limit as the size of the underlying graph diverges to infinity. In the zero temperature case, this is interpreted as the existence of the scaling limit for the corresponding combinatorial optimization problem. For example, as a special case we prove that the size of a largest independent set in these graphs, normalized by the number of nodes converges to a limit w.h.p., thus resolving an open problem mentioned by several experts: Wormald '99, Aldous-Steele 2003, Bollobás-Riordan 2008, as well as Janson-Thomason 2008. Our approach is based on extending and simplifying the Guerra-Toninelli's interpolation method. Among other applications, this method was used to prove the existence of free energy limits for Viana-Bray and K-SAT models on Erdös-Rényi graphs. The case of zero temperature was treated by taking limits of positive temperature models. We provide instead a simpler combinatorial approach and work with the zero temperature case (optimization) directly both in the case of Erdös-Rényi graph $\Gamma(N, c/N)$ and random regular graph $\Gamma(N, r)$. In addition we establish the large deviations principle for the satisfiability property for constraint satisfaction problems such as Coloring, K-SAT and NAE-K-SAT. For example, let $p(c, q, N)$ and $p(r, q, N)$

denote, respectively, the probability that random graphs $\Gamma(N, c/N)$ and $\Gamma(N, r)$ are properly q-colorable. We prove the existence of limits of $N^{-1} \log p(c, q, N)$ and $N^{-1} \log p(r, q, N)$, as $N \to \infty$.

• **Approximations for the Isoperimetric and Spectral Profile of Graphs and for Restricted Eigenvalues of Diagonally-Dominant Matrices**.
This joint work with Prasad Raghavendra (MSR-New England Postdoc) and David Steurer (Princeton) appeared in the Proc. of the ACM STOC 2010.
Spectral profile (of the Laplacian matrix) of a finite Markov chain or a graph is a somewhat recently introduced refinement of the more classical and fundamental parameter, spectral gap, of the corresponding space. While spectral profile was introduced in the context of studying mixing times of Markov chains (by Goel-Montenegro-Tetali, 2006) as a natural functional analog of the isoperimetric profile, termed *average conductance* by Lovász-Kannan (STOC'99), its origins go back to earlier work in heat kernel estimates on groups and manifolds by various geometers (Barlow-Coulhon-Grigor'yan-Pittet, 2001). Just as the average conductance is characterized by expansion from small sets, spectral profile is characterized by the smallest eigenvalues of certain restrictions of the Laplacian matrix of a graph. Roughly speaking, given a graph G, for each $0 < \delta < 1$, the spectral profile $\Lambda_G(\delta)$ minimizes the Rayleigh quotient (from the variational characterization) of the spectral gap of the Laplacian matrix of G over vectors with support at most δ over a suitable probability measure.
Despite the utility of the spectral profile in the above as well as other settings, the computational complexity of estimating the spectral profile in Markov chains or (equivalently) on graphs remained relatively unexplored, especially compared to the vast literature on estimating the related quantities such as the sparsest cut in a graph.
In this work, we consider a natural semidefinite programming relaxation of the spectral profile of a graph, and prove a logarithmic factor approximation (of order $\log(1/\delta)$) to the spectral profile $\Lambda(\delta)$. Further motivated by applications to sparsification, we extend our techniques in nontrivial ways to provide similar approximation guarantees for the spectral profile of diagonally-dominant matrices, which are also a natural generalization of Laplacian matrices. Our results seem to be the first set of results exploring the computational complexity of spectral profile.
Besides providing tighter estimates on mixing times for various walks (by avoiding the penalty of a slow start), the approach to mixing time in L^2 using the spectral profile facilitated easier derivation of estimates on mixing using other functional approaches – notably the logarithmic Sobolev and Nash inequalities.
Generalizing the standard Cheeger-type inequality relating the spectral gap and conductance (or sparsest cut in the context of a graph), Goel-Montenegro-Tetali (2006) also showed a Cheeger-type inequality relating the spectral profile to the average conductance (or isoperimetric profile). Since one is often interested in knowing when *small sets* expand (as this is sufficient for many applications of expanders), the question of computing or approximating the spectral profile becomes relevant and interesting – with the hope of obtaining a certificate for such small-set expansion. In this work we consider a very natural semidefinite relaxation of the profile and obtain a logarithmic factor approximation.
Inspired by recent applications in optimization over sparse subspaces, we consider a natural extension of the notion of spectral profile to the class of diagonally dominant matrices. Recall that a matrix (over the reals) is diagonally dominant if in every row, the diagonal entry is larger or equal (in absolute value) to the sum of the absolute values of the off-diagonal

entries, and additionally if in some row the inequality is strict. The Laplacian matrix of a graph (being the diagonal matrix of degrees minus the adjacency matrix), with a row and a corresponding column removed, is easily seen to be a special case of such matrices. Our second main contribution is in formulating the spectral profile problem for the diagonally dominant matrices and providing a similar logarithmic factor approximation. This turned out to be a rather nontrivial extension of our first result for graphs.

• **Time until intersection of independent walks on $\mathbb{Z}$.**
This is the journal version of prior joint work with R. Montenegro (University of Massachusetts, Lowell), which had appeared in the Proceedings of ACM STOC (2009). In this work, we develop probabilistic tools for upper and lower bounding the expected time until two independent random walks on $\mathbb{Z}$ intersect each other. This leads to the first sharp analysis of a non-trivial Birthday attack, proving that Pollard's Kangaroo method solves the discrete logarithm problem $g^x = h$ on a cyclic group in expected time $(2 + o(1))\sqrt{b-a}$ for an average $x \in_{uar} [a, b]$. In this final (journal) version, we show that our methods also resolve a conjecture of Pollard's, by showing that the same bound holds when step sizes are generalized from powers of 2 to powers of any fixed n.

• **Entropy and set cardinality inequalities for partition-determined functions**
This is joint work with Mokshay Madiman (Statistics, Yale University) and Adam Marcus (Math, Yale University). Continuation of this work is a component of the PI's continuing grant.
A new notion of partition-determined functions is introduced, and several basic inequalities are developed for the entropy of such functions of independent random variables, as well as for cardinalities of compound sets obtained using these functions. Here a compound set means a set obtained by varying each argument of a function of several variables over a set associated with that argument, where all the sets are subsets of an appropriate algebraic structure so that the function is well defined. On the one hand, the entropy inequalities developed for partition-determined functions imply entropic analogues of general inequalities of Plünnecke-Ruzsa type. On the other hand, the cardinality inequalities developed for compound sets imply several inequalities for sumsets, including for instance a generalization of inequalities proved by Gyarmati, Matolcsi and Ruzsa (2010). We also provide partial progress towards a conjecture of Ruzsa (2007) for sumsets in nonabelian groups. All proofs are elementary and rely on properly developing certain information-theoretic inequalities.

• **Approximating Minimum Linear Ordering Problems**
This is recent joint work with Prof. Satoru Iwata (Kyoto, Japan) and the CS graduate student Pushkar Tripathi (Georgia Tech), and the work has been submitted to the upcoming ACM-SIAM Symposium on Discrete Algorithms (SODA), Kyoto, Japan.
The work addresses a very general Minimum Linear Ordering Problem (MLOP): Given a nonnegative set function f on a finite set V, find a linear ordering on V such that the sum of the function values for all the suffixes is minimized. This problem generalizes well-known problems such as the Minimum Linear Arrangement, Min Sum Set Cover, Minimum Latency Set Cover, and Multiple Intents Ranking. Extending a result of Feige, Lovász, and Tetali (2004) on Min Sum Set Cover, we show that the greedy algorithm provides a factor 4 approximate optimal solution when the cost function f is supermodular. We also present a factor 2 rounding algorithm for MLOP with a monotone submodular cost function, using the convexity of the

Lovász extension. These are among very few constant factor approximation algorithms for NP-hard minimization problems formulated in terms of submodular/supermodular functions. In contrast, when f is a symmetric submodular function, the problem has an information theoretic lower bound of 2 on the approximability.
In addition, the work provides a randomized rounding algorithm for the Min Sum Vertex Cover problem of factor 1.79, improving over the factor 2 algorithm described by Feige, Lovász, and Tetali (2004).