Report from Dagstuhl Seminar 11241

Design and Analysis of Randomized and Approximation Algorithms

Edited by

Martin E. Dyer[1], Uriel Feige[2], Alan M. Frieze[3], and Marek Karpinski[4]

1 **University of Leeds, GB,** m.e.dyer@leeds.ac.uk
2 **Weizmann Institute – Rehovot, IL,** uriel.feige@weizmann.ac.il
3 **Carnegie Mellon University – Pittsburgh, US,** alan@random.math.cmu.edu
4 **Universität Bonn, DE,** marek@cs.uni-bonn.de

Abstract

The Dagstuhl Seminar on "Design and Analysis of Randomized and Approximation Algorithms" (Seminar 11241) was held at Schloss Dagstuhl between June 13–17, 2011. There were 26 regular talks and several informal and open problem session contributions presented during this seminar. Abstracts of the presentations have been put together in this seminar proceedings document together with some links to extended abstracts and full papers.

Seminar 13.–17. June, 2011 – www.dagstuhl.de/11241

1998 ACM Subject Classification C.2 Computer-Communication Networks, E.1 Data Structures, F.2 Analysis of Algorithms and Problem Complexity, F.4 Mathematical Logic and Formal Languages, G.1.2 Approximation, G.1.6 Optimization, G.2 Discrete Mathematics, G.2.2 Graph Theory, G.3 Probability and Statistics

Keywords and phrases Randomized Algorithms, Approximation Algorithms, Approximation Hardness, Optimization Problems, Counting Problems, Streaming Algorithms, Random Graphs, Hypergraphs, Probabilistic Method, Networks, Linear Programs, Semidefinite Programs

Digital Object Identifier 10.4230/DagRep.1.6.24

Edited in cooperation with Mathias Hauptmann

1 Executive Summary

Martin Dyer
Uriel Feige
Alan M. Frieze
Marek Karpinski

License Creative Commons BY-NC-ND 3.0 Unported license
© Martin Dyer, Uriel Feige, Alan M. Frieze, Marek Karpinski

Many, if not most computational tasks that arise in realistic scenarios are computationally difficult, and no efficient algorithms are known that guarantee an exact (or optimal) solution on every input instance. Nevertheless, practical necessity dictates that acceptable solutions be found in a reasonable time. Two basic means for surmounting the intractability barrier are randomized computation, where the answer is optimal with high probability but not with certainty, and approximate computation, where the answer is guaranteed to be within, say, small percentage of optimality. Often, these two notions go hand-in-hand.

Except where otherwise noted, content of this report is licensed under a Creative Commons BY-NC-ND 3.0 Unported license
Design and Analysis of Randomized and Approximation Algorithms, *Dagstuhl Reports*, Vol. 1, Issue 6, pp. 24–53
Editors: Martin E. Dyer, Uriel Feige, Alan M. Frieze, and Marek Karpinski
Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

The seminar was concerned with the newest developments in the design and analysis of randomized and approximation algorithms. The main focus of the workshop was on the following specific topics: randomized approximation algorithms for optimization problems, approximation algorithms for counting problems, methods for proving approximation hardness, as well as various interactions between them. Here, some new broadly applicable techniques have emerged recently for designing efficient approximation algorithms for various optimization and counting problems as well as for proving approximation hardness bounds. This workshop has addressed the above topics and some new fundamental insights and paradigms in this area.

The 26 regular talks and other presentations delivered at this workshop covered a wide body of research in the above areas. The Program of the meeting and Abstracts of all talks are listed in the subsequent sections of this report.

The meeting was hold in a very informal and stimulating atmosphere. Thanks to everyone who made it such an interesting and enjoyable event.

Martin Dyer
Uriel Feige
Alan M. Frieze
Marek Karpinski

Acknowlegement. We thank Annette Beyer and Angelika Mueller-von Brochowski for their continuous support and help in organizing this workshop. Thanks go to Mathias Hauptmann for his help in collecting abstracts of the talks and other related materials for these Proceedings.

2 Table of Contents

Informal Session

Open Problem Session

3 Overview of Talks

3.1 On the usefulness of predicates

Johan Håstad (KTH – Stockholm, SE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Johan Håstad
Joint work of J. Håstad and A. Per

We introduce a notion of usefulness for constraint satisfaction problems. A k-ary predicate P is useful for a k-ary function Q if the following holds. Given a list of k-tuples of literals and a promise that there is an assignment such that P is true on (almost) all of the resulting strings, we can efficiently find an assignment such that when Q is applied to the resulting strings the average is more than the expectation of Q when applied to a random string.

This is an extension of the concept of approximation resistance of standard Max-CSPs in that P is useful for P iff it is not approximation resistant.

A predicate P is useless if it is not useful for any real-valued Q. Among other results we give a simple characterization of uselessness assuming the unique games conjecture: P is useless iff there is a pairwise independent measure supported on the strings accepted by P.

3.2 Counting contingency tables

Alexander Barvinok (University of Michigan – Ann Arbor, US)

License Creative Commons BY-NC-ND 3.0 Unported license
© Alexander Barvinok
Joint work of A. Barvinok and J.A. Hartigan

Let $R = (r_1, \ldots, r_m)$ be a positive integer m-vector and let $C = (c_1, \ldots, c_n)$ be a positive integer n-vector such that

$$r_1 + \ldots + r_m = c_1 + \ldots + c_n = N.$$

We are interested in the number $\#(R, C)$ of $m \times n$ non-negative integer matrices (*contingency tables* with *margins* R and C) with row sums $r_1, \ldots, r_m$ and column sums $c_1, \ldots, c_n$. Namely, we present an efficiently computable asymptotic formula for $\#(R, C)$.

Let us consider the function

$$g(x) = (x + 1) \ln(x + 1) - x \ln x \quad \text{for} \quad x \geq 0.$$

It is easy to see that $g(x)$ is increasing and concave. We extend g to non-negative $m \times n$ matrices X by

$$g(X) = \sum_{i,j} g(x_{ij}) \quad \text{for} \quad X = (x_{ij}).$$

Since g is strictly concave, it attains its maximum on the transportation polytope of all $m \times n$ non-negative matrices X with row sums R and column sums C at a unique point $Z = (z_{ij})$, which we call the *typical matrix*. As is shown in [1], a random non-negative integer matrix with row sums R and column sums C looks more or less like the random matrix of independent geometric random variables with expectation Z. Matrix Z can be efficiently computed by interior point methods.

Let us define a quadratic form $q : \mathbb{R}^{m+n} \longrightarrow \mathbb{R}$ by

$$q(s_1, \ldots, s_m; t_1, \ldots, t_n) = \frac{1}{2} \sum_{i,j} z_{ij} (z_{ij} + 1) (s_i + t_j)^2 .$$

Let $u \in \mathbb{R}^{m+n}$, $u = (1, \ldots, 1; -1, \ldots, -1)$, be vector and let $H = u^\perp$, $H \subset \mathbb{R}^{m+n}$, be the hyperplane that is the orthogonal complement to u. Then the restriction $q|H$ of q onto H is a positive definite quadratic form. We define $\det q|H$ as the determinant of $q|H$. Equivalently, $\det q|H$ is the product of the non-zero eigenvalues of q.

We define the Gaussian probability measure in H with the density proportional to e^{-q}. We consider two random variables $f, h : H \longrightarrow \mathbb{R}$ defined by

$$f(s_1, \ldots, s_m; t_1, \ldots, t_n) = \frac{1}{6} \sum_{i,j} z_{ij} (z_{ij} + 1) (2z_{ij} + 1) (s_i + t_j)^3$$

and

$$h(s_1, \ldots, s_m; t_1, \ldots, t_n) = \frac{1}{24} \sum_{i,j} z_{ij} (z_{ij} + 1) \left(6z_{ij}^2 + 6z_{ij} + 1\right) (s_i + t_j)^4 .$$

We compute

$$\mu = \mathbf{E}\, f^2 \quad \text{and} \quad \nu = \mathbf{E}\, h.$$

We note that computing the expectation of a polynomial with respect to the Gaussian probability measure is a linear algebra problem. In particular, given Z, one can compute μ and ν in $O\left((m+n)^4\right)$ time. To describe the range for which our asymptotic formula is applicable, we need one more definition. Given $0 < \delta < 1$, we say that the margins (R, C) are δ-*smooth* if

$$m \ \geq \ \delta n, \quad n \ \geq \ \delta m \quad \text{and} \quad \delta\tau \ \leq \ z_{ij} \ \leq \ \tau \quad \text{for all} \quad i, j$$

and some

$$\tau \geq \delta,$$

where $Z = (z_{ij})$ is the typical matrix. The following result is proved in [1].

Theorem. Let us fix $0 < \delta < 1$. Let (R, C) be δ-smooth margins. Then, for any $0 < \epsilon \leq 1/2$ the value of

$$\frac{e^{g(Z)} \sqrt{m+n}}{(4\pi)^{(m+n-1)/2} \sqrt{\det q|H}} \exp\left\{-\frac{\mu}{2} + \nu\right\}$$

approximates the number $\#(R, C)$ within relative error ϵ, provided

$$m + n \ \geq \ \left(\frac{1}{\epsilon}\right)^{\gamma(\delta)}$$

for some $\gamma(\delta) > 0$.

If

$$r_1 = \ldots = r_m = r \quad \text{and} \quad c_1 = \ldots = c_n = c$$

then by symmetry we have

$$z_{ij} = \frac{rc}{N} = \frac{r}{n} = \frac{c}{m} \quad \text{for all} \quad i, j$$

and the formula of the above theorem transforms into the asymptotic formula of [3], obtained earlier by Canfield and McKay in the particular case when all the row sums are equal and all the column sums are equal.

References

1 A. Barvinok, *What does a random contingency table look like*, Combinatorics, Probability and Computing 19, 2010, pp. 517–539

2 A. Barvinok and J.A. Hartigan, *An asymptotic formula for the number of non-negative integer matrices with prescribed row and column sums*, Transactions of the American Mathematical Society, to appear, preprint `arXiv:0910.2477` (2009)

3 E.R. Canfield and B.D. McKay, *Asymptotic enumeration of integer matrices with large equal row and column sums*, Combinatorica 30, 2010, pp. 655–680

3.3 Estimating the partition function of the ferromagnetic Ising model on a regular matroid

Leslie Ann Goldberg (University of Liverpool, GB)

License Creative Commons BY-NC-ND 3.0 Unported license
© Leslie Ann Goldberg
Joint work of L.A. Goldberg and M. Jerrum

We investigate the computational difficulty of approximating the partition function of the ferromagnetic Ising model on a regular matroid. Jerrum and Sinclair have shown that there is a fully polynomial randomised approximation scheme (FPRAS) for the class of graphic matroids. On the other hand, the authors have previously shown, subject to a complexity-theoretic assumption, that there is no FPRAS for the class of binary matroids, which is a proper superset of the class of graphic matroids. In order to map out the region where approximation is feasible, we focus on the class of regular matroids, an important class of matroids which properly includes the class of graphic matroids, and is properly included in the class of binary matroids. Using Seymour's decomposition theorem, we give an FPRAS for the class of regular matroids.

References

1 Leslie Ann Goldberg, Mark Jerrum, *A polynomial-time algorithm for estimating the partition function of the ferromagnetic Ising model on a regular matroid*, http://arxiv.org/abs/1010.6231)

3.4 TSP on cubic and subcubic graphs

Leen Stougie (CWI – Amsterdam, NL)

License Creative Commons BY-NC-ND 3.0 Unported license
© Leen Stougie
Joint work of S. Boyd, R. Sitters, S. van der Ster, L. Stougie

We study the Travelling Salesman Problem (TSP) on the metric completion of cubic and subcubic graphs, which is known to be NP-hard. The problem is of interest because of its relation to the famous 4/3 conjecture for metric TSP, which says that the integrality gap, i.e., the worst case ratio between the optimal values of the TSP and its linear programming relaxation, is 4/3. Using polyhedral techniques in an interesting way, we obtain a polynomial-time 4/3-approximation algorithm for this problem on cubic graphs, improving upon Christofides' 3/2-approximation, and upon the $3/2 - 5/389 \approx 1.487$-approximation ratio by Gamarnik, Lewenstein and Svirdenko for the case the graphs are also 3-edge connected. We also prove that, as an upper bound, the 4/3 conjecture is true for this problem on cubic graphs. For subcubic graphs we obtain a polynomial-time 7/5-approximation algorithm and a 7/5 bound on the integrality gap. Just very recently Mömke and Svensson superseded this result by announcing 4/3 bounds for subcubic graphs. However, the techniques we propose here remain interesting and probably more widely applicable.

References

1 Sylvia Boyd, Rene Sitters, Suzanne van der Ster and Leen Stougie, *TSP on Cubic and Subcubic Graphs*, accepted for IPCO 2011

3.5 Approximating Graphic TSP by Matchings

Ola Svensson (KTH – Stockholm, SE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Ola Svensson
Joint work of T. Mömke and O. Svensson

We present a framework for approximating the metric TSP based on a novel use of matchings. Traditionally, matchings have been used to add edges in order to make a given graph Eulerian, whereas our approach also allows for the removal of certain edges leading to a decreased cost.

For the TSP on graphic metrics (graph-TSP), the approach yields a 1.461-approximation algorithm with respect to the Held-Karp lower bound. For graph-TSP restricted to a class of graphs that contains degree three bounded and claw-free graphs, we show that the integrality gap of the Held-Karp relaxation matches the conjectured ratio 4/3. The framework allows for generalizations in a natural way and also leads to a 1.586-approximation algorithm for the traveling salesman path problem on graphic metrics where the start and end vertices are prespecified.

3.6 Connectivity in Discrete Random Processes

Po-Shen Loh (Carnegie Mellon University – Pittsburgh, US)

License Creative Commons BY-NC-ND 3.0 Unported license
© Po-Shen Loh
Joint work of E. Lubetzky and Po-Shen Loh

Half a century ago, a seminal paper of Erdos and Renyi launched the systematic study of random graphs. Since then, this direction of investigation has blossomed into a broad field, and the original model has given rise to many useful variants. Of the properties which have received attention, one of the most fundamental has been that of global connectivity.

Recently, motivated by the practical problem of establishing connectivity in peer- to-peer networks, a natural question of similar flavor arose in the analysis of a natural randomized clustering algorithm. Using methods which originated from physics, but now known to be remarkably useful in the study of random graphs, we establish the asymptotic optimality of this algorithm. We also prove the first rigorous lower bounds on the performance of a closely-related algorithm, extending an approach of Oded Schramm.

3.7 A $(5/3 + \epsilon)$-Approximation for Strip Packing

Lars Prädel (University of Kiel, DE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Lars Prädel

We study strip packing, which is one of the most classical two-dimensional packing problems: Given a collection of rectangles, the problem is to find a feasible orthogonal packing without rotations into a strip of width 1 and minimum height. In this paper we present an approximation algorithm for the strip packing problem with approximation ratio of $5/3 + \epsilon$ for any $\epsilon > 0$. This result significantly narrows the gap between the best known upper bounds of 2 by Schiermeyer and Steinberg and 1.9396 by Harren and van Stee and the lower bound of 3/2.

3.8 Every Hyperfinite Property is Testable

Christian Sohler (TU Dortmund, DE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Christian Sohler

Joint work of I. Newman and Ch. Sohler

A property testing algorithm for a property Π in the bounded degree graph model is an algorithm that, given access to the adjacency list representation of a graph $G = (V, E)$ with maximum degree at most d, accepts G with probability at least 2/3 if G has property Π, and rejects G with probability at least 2/3, if it differs on more than ϵdn edges from every d-degree bounded graph with property Π. A property is *testable*, if for every ϵ, d and n, there is a property testing algorithm $A_{\epsilon,n,d}$ that makes at most $q(\epsilon, d)$ queries to an input graph of n vertices, that is, a non-uniform algorithm that makes a number of queries that is independent of the graph size. A k-disc around a vertex v of a graph $G = (V, E)$ is the subgraph induced by all vertices of distance at most k from v. We show that the structure of a planar graph on large enough number of vertices, n, and with constant maximum degree d, is determined, up to the modification (insertion or deletion) of at most ϵdn edges, by the frequency of k-discs for certain $k = k(\epsilon, d)$ that is independent of the size of the graph. We can replace planar graphs by any hyperfinite class of graphs, which includes, for example, every graph class that does not contain a set of forbidden minors.

We use this result to obtain new results and improve upon existing results in the area of property testing. In particular, we prove that

- graph isomorphism is testable for every class of hyperfinite graphs,
- every graph property is testable for every class of hyperfinite graphs,
- every hyperfinite graph property is testable in the bounded degree graph model,
- A large class of graph parameters is approximable for hyperfinite graphs

Our results also give a partial explanation of the success of motifs in the analysis of complex networks.

References

1 Ilan Newman and Christian Sohler, *Every property of hyperfinite graphs is testable*, STOC 2011, pp. 675–684

3.9 Sublinear Algorithms via Precision Sampling

Alexandr Andoni (Microsoft Research – Mountain View, US)

License Creative Commons BY-NC-ND 3.0 Unported license
© Alexandr Andoni
Joint work of R. Krauthgamer, A. Andoni, K. Onak

Suppose we want to estimate a sum of bounded reals $a_1 + a_2 + \ldots + a_n$ with access to only some "limited information" about the a_i's. A classical setting is where we estimate the entire sum by knowing only a random subset of a_i's. Naturally, there is a trade-off between the size of the subset and the resulting approximation.

Motivated by applications where this tradeoff is not good enough, we introduce Precision Sampling, which is an estimation technique that uses more general kind of "limited information" about the a_i's: Instead of obtaining a subset as above, here we obtain a rough estimate for each a_i, up to various "precision" (approximation). The trade-off is then between the precision of the estimates and the resulting approximation to the total sum. We show that one can obtain a trade-off that is qualitatively better in the precision sampling setting than in the aforementioned (vanilla) sampling setting.

Our resulting tool leads to new sublinear algorithms, including a simplified algorithm for a class of streaming problems, as well as an efficient algorithm for estimating the edit distance.

3.10 Lifting Markov Chains for Faster Mixing

Thomas Hayes (University of New Mexico – Albuquerque, US)

License Creative Commons BY-NC-ND 3.0 Unported license
© Thomas Hayes
Joint work of Th. Hayes and A. Sinclair

Markov Chain Monte Carlo is a powerful tool for sampling from distributions over large sets with combinatorial structure. Generally, the goal is to obtain samples fast, as a function of some parameter that is, say, logarithmic in the space being sampled. In some cases, we know how to sample in say, polynomial time, but really want performance that is a little faster, say $O(n \log(n))$ time or $O(n^2)$. Are there any tools for systematically enhancing the speed of MCMC algorithms?

"Lifting" a given Markov chain produces a new chain, whose ergodic flow projects homomorphically back down to that of the original chain, and hence can be used for sampling the original distribution. We discuss some examples fo which a directed lifting of an undirected original chain gives as much as a quadratic speedup. The main example is a lifting of a tree-structured Markov chain introduced by Jerrum and Sinclair.

3.11 Average-Case Performance of Heuristics for Multi-Dimensional Assignment

Gregory Sorkin (London School of Economics, GB)

License Creative Commons BY-NC-ND 3.0 Unported license
© Gregory Sorkin
Joint work of A. Frieze and G. Sorkin

Beautiful formulas are known for the expected cost of random two-dimensional assignment problems, but in higher dimensions, even the scaling is not known. In 3 dimensions and above, the problem has natural "planar" and "axial" versions, both of which are NP-hard. For 3-dimensional Planar random assignment instances of size n, the cost scales as $\Omega(2/n)$, and a main result of the present paper is the first polynomial-time algorithm that, with high probability, finds a solution of cost $O(n^{-1+\epsilon})$, for arbitrary positive ϵ (or indeed ϵ going slowly to 0). For 3-dimensional Axial assignment, the lower bound is $\Omega(n)$, and we give a new efficient matching-based algorithm that returns a solution with expected cost $O(n \log n)$. Neither algorithm extends to 4 or more dimensions, and finding algorithms with the conjectured scaling for d-dimensional Planar and Axial assignment are open problems.

References

1 A. Frieze and G. Sorkin, *Average-case performance of heuristics for three-dimensional assignment problems*, http://arxiv.org/abs/1008.0390, submitted for publication

3.12 Embedding Spanning Trees in Random Graphs near the Connectivity Threshold

Michael Krivelevich (Tel Aviv University, IL)

License Creative Commons BY-NC-ND 3.0 Unported license
© Michael Krivelevich
Joint work of D. Hefetz, M. Krivelevich, T. Szabo

A disconnected graph G does not contain any spanning trees. Thus, a tree T on n vertices typically does not appear in the binomial random graph $G(n, p)$ before the threshold for connectivity, which is well known to be at $p(n) = \frac{\log(n)}{n}$. We prove that a given tree T on n vertices with bounded maximum degree is contained almost surely in $G(n, p)$ with $p(n) = (1 + \epsilon)\frac{\log(n)}{n}$, provided T belongs to one of the following classes:

(1) T has linearly many leaves

(2) T has a path of linear length all of whose vertices have degree two in T.

3.13 Stochastic Knapsack with Correlations and Cancellation and Application to Non-Martingale Bandit Problems

R. Ravi (Carnegie Mellon University – Pittsburgh, US)

License Creative Commons BY-NC-ND 3.0 Unported license
© R. Ravi
Joint work of A. Gupta, R. Krishnaswamy, M. Molinaro, R. Ravi

In the stochastic knapsack problem, we are given a knapsack with size B, and a set of jobs whose sizes and rewards are drawn from a known probability distribution. However, the only way to know the actual size and reward is to schedule the job—when it completes, we get to know these values. How should we schedule jobs to maximize the expected total reward? We know constant-factor approximations for this problem when we assume that rewards and sizes are independent random variables, and that we cannot prematurely cancel jobs after we schedule them. What can we say when either or both of these assumptions are dropped?

Not only is the stochastic knapsack problem of interest in its own right, but techniques developed for it are applicable to other stochastic packing problems. Indeed, ideas for this problem have been useful for budgeted learning problems, where one is given several arms which evolve in a specified stochastic fashion with each pull, and the goal is to pull the arms a total of B times to maximize the reward obtained. Much recent work on this problem focus on the case when the evolution of the arms follows a martingale, i.e., when the expected reward from the future is the same as the reward at the current state. However, what can we say when the rewards do not form a martingale?

We give constant-factor approximation algorithms for the stochastic knapsack problem with correlations and cancelations, and also for some budgeted learning problems where the martingale condition is not satisfied, using similar ideas. Indeed, we can show that previously proposed linear programming relaxations for these problems have large integrality gaps. We propose new time-indexed LP relaxations; using a decomposition and "shifting" approach, we convert these fractional solutions to distributions over strategies, and then use the LP values and the time ordering information from these strategies to devise a randomized scheduling algorithm. We hope our LP formulation and decomposition methods may provide a new way to address other correlated bandit problems with more general contexts.

References

1 A. Gupta, R. Krishnaswamy, M. Molinaro and R. Ravi, *Approximation Algorithms for Correlated Knapsacks and Non-Martingale Bandits*, paper available as arXiv:1102.3749

3.14 Smoothed Analysis of Multiobject Optimization

Heiko Röglin (University of Bonn, DE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Heiko Röglin
Joint work of R. Beier, T. Brunsch, H. Röglin, S.-H. Teng, B. Vöcking

A well established heuristic approach for solving various multicriteria optimization problems is to enumerate the set of Pareto-optimal solutions. The heuristics following this principle are often successful in practice, even though the number of Pareto-optimal solutions can be exponential in the worst case.

We analyze multiobjective optimization problems in the framework of smoothed analysis, and we prove that the smoothed number of Pareto-optimal solutions in any multiobjective binary optimization problem with a finite number of linear objective functions is polynomial. Moreover, we give polynomial bounds on all finite moments of the number of Pareto-optimal solutions, which yields the first non-trivial concentration bound for this quantity.

References

1 René Beier, Heiko Röglin and Berthold Vöcking, *The Smoothed Number of Pareto Optimal Solutions in Bicriteria Integer Optimization*, IPCO 2007, pp. 53–67
2 Heiko Röglin and Shang-Hua Teng, *Smoothed Analysis of Multiobjective Optimization*, FOCS 2009, pp. 681–690

3 Tobias Brunsch and Heiko Röglin, *Lower Bounds for the Smoothed Number of Pareto Optimal Solutions*, TAMC 2011, pp. 416–427

3.15 k-Means Algorithm Converges

Ravi Kannan (Microsoft Research, IN)

License Creative Commons BY-NC-ND 3.0 Unported license
© Ravi Kannan
Joint work of R. Kannan and A. Kumar

The k-means algorithm is widely used. It is well-recognized that it does not converge to the desirable answer if we start with a bad set of centers. We formalize a simple geometric condition called proximity under which we show it does converge to the desired result. Many known results which assume a stochastic model of input are subsumed by our purely deterministic result.

3.16 Random Geometric Graphs

Tobias Müller (CWI – Amsterdam, NL)

License Creative Commons BY-NC-ND 3.0 Unported license
© Tobias Müller

If we pick points $X_1, ..., X_n$ at random from d-dimensional space (i.i.d. according to some probability measure) and fix a $r > 0$, then we obtain a random geometric graph by joining points by an edge whenever their distance is $< r$.

We give a brief overview of some of the most important results on random geometric graphs and then describe some of my own work on Hamilton cycles, the chromatic number, and the power of two choices in random geometric graphs.

3.17 Hardness of Approximating the Tutte Polynomial of a Binary Matroid

Mark Jerrum (Queen Mary University of London, GB)

License Creative Commons BY-NC-ND 3.0 Unported license
© Mark Jerrum
Joint work of L.A. Goldberg and M. Jerrum

We consider the problem of approximating certain combinatorial polynomials. First, we consider the problem of approximating the Tutte polynomial of a binary matroid with parameters $q \geq 2$ and γ. (Relative to the classical (x, y) parameterisation, $q = (x-1)(y-1)$ and $\gamma = y - 1$.) A graph is a special case of a binary matroid, so earlier work by the authors shows that for $q > 2$ and $\gamma < 0$ there is no FPRAS unless NP = RP, and for $q > 2$ and $\gamma > 0$, the approximation problem is hard for the complexity class $\#\mathrm{RH}\Pi_1$ under approximation-preserving (AP) reducibility. The case $\gamma = 0$ corresponds to the infinite-temperature limit of the Potts model, and is computationally trivial. The situation for $q = 2$ is different. For graphic matroids, the region $\gamma < -2$ is only known to be as hard as approximating perfect matchings in a graph (a problem whose complexity is open), whereas Jerrum and Sinclair have provided an FPRAS for the region $\gamma > 0$. It is known that there is no FPRAS unless NP = RP in the in-between region $-2 \leq \gamma < 0$, apart from at two "special points" where the polynomial can be computed exactly in polynomial time. We show that for binary matroids there is no FPRAS in the region $\gamma < -2$ unless NP = RP. Also, in the region $\gamma > 0$ the approximation problem is hard for the complexity class $\#\mathrm{RH}\Pi_1$ under approximation-preserving (AP) reducibility. Thus, unless there is an FPRAS for all of $\#\mathrm{RH}\Pi_1$, the graphic case differs in approximation complexity from the binary matroid case at $q = 2$. Our result implies that it is computationally difficult to approximate the weight enumerator of a binary linear code, apart from at the special weights for which the problem is exactly solvable in polynomial time. As a consequence, we show that approximating the cycle index polynomial of a permutation group is hard for $\#\mathrm{RH}\Pi_1$ under AP-reducibility, partially resolving a question first posed in 1992.

References

1 L.A. Goldberg and Mark Jerrum, *Approximating the Tutte polynomial of a binary matroid and other related combinatorial polynomials*, URL: http://arxiv.org/abs/1006.5234

3.18 Computational Complexity of the Hamiltonian Cycle Problem in Dense Hypergraphs

Edyta Szymanska (Adam Mickiewicz University – Poznan, PL)

License Creative Commons BY-NC-ND 3.0 Unported license
© Edyta Szymanska
Joint work of M. Karpinski, A. Rucinski, E. Szymanska

We study the computational complexity of deciding the existence of a Hamiltonian Cycle in some dense classes of k-uniform hypergraphs. Those problems turned out to be, along with the hypergraph Perfect Matching problems, exceedingly hard, and there is a renewed algorithmic interest in them. In this paper we design a polynomial time algorithm for the Hamiltonian Cycle problem for k-uniform hypergraphs with density at least $1/2 + \epsilon, \epsilon > 0$. In doing so, we depend on a new method of constructing Hamiltonian cycles from (purely) existential statements which could be of independent interest. On the other hand, we establish NP-completeness of that problem for density at least $1/k - \epsilon$. Our results seem to be the first complexity theoretic results for the Dirac-type dense hypergraph classes.

References

1 M. Karpinski, A. Ruciński, and E. Szymańska, *Computational Complexity of the Hamiltonian Cycle Problem in Dense Hypergraphs*, LATIN 2010, pp. 662–673.

3.19 Distributed Storage Allocation via Fractional Hypergraph Matchings

Andrzej Rucinski (Adam Mickiewicz University – Poznan, PL)

License Creative Commons BY-NC-ND 3.0 Unported license
© Andrzej Rucinski

The following model of distributed storage has been studied in information theory [7, 8, 12]. A file is split into multiple chunks, and then it is replicated redundantly and stored in a distributed storage system with n nodes. The amount of data to be stored in each node i is equal to x_i, where the size of the whole file is normalized to 1. We require that the total amount of data stored does not exceed a given budget T, i.e. $x_1 + \cdots + x_n \leq T$. At the time of retrieval, we attempt to recover the whole file by accessing only the data stored in a randomly chosen subset R of nodes. It is known that there always exists a coding scheme such that we can recover the file whenever the total amount of data accessed is at least the size of this file. Our goal is to find an optimal allocation $(x_1, \cdots, x_n)$ to maximize the probability of successful recovery. In [12], R is taken uniformly at random among all the r-element subsets of $\{x_1, \cdots, x_n\}$. Then the problem can be reformulated as follows: *for a nonnegative sequence* $(x_1, \cdots, x_n)$, *let*

$$\Phi(x_1, \cdots, x_n) = \left| \left\{ S \subseteq [n], |S| = r \textit{ such that } \sum_{i \in S} x_i \geq 1 \right\} \right|.$$

Given integers $n \geq r \geq 1$ *and a real number* $T > 0$, *determine*

$$F(r, n, T) = \max_{\sum x_i = T,\ x_i \geq 0\ \forall i} \Phi(x_1, \cdots, x_n).$$

If the total budget $T \geq n/r$, by setting all x_i equal to $T/n \geq 1/r$, we can recover the original file from any subset of size r. For the case $T < n/r$, the problem of determining $F(n, r, T)$ is equivalent to that of finding the maximum number of edges in an r-uniform hypergraph on n vertices with fractional matching number at most T. Erdős and Gallai [4] determined the integral version of this problem for graphs ($r = 2$). In 1965, Erdős [3] conjectured for r-uniform hypergraphs that the maximum number of edges without a matching of size $s \leq n/r$ is $\max\left\{\binom{rs-1}{r}, \binom{n}{r} - \binom{n-s+1}{r}\right\}$.

For $r = 2$ and $s = T + 1$, where we assume that T is an integer, an easy calculation shows that the above maximum equals the first term if $\frac{2}{5}n \leq T \leq \frac{1}{2}n$, and the corresponding optimal graph is a clique of size, roughly, $2T$. This means that, asymptotically, an optimal allocation is $x_1 = \cdots = x_{2T} = 1/2$ and $x_{2T+1} = \cdots = x_n = 0$. On the other hand, if $T < \frac{2}{5}n$, an optimal allocation is $x_1 = \cdots = x_T = 1$ and $x_{T+1} = \cdots = x_n = 0$.

We now formulate the fractional version of Erdős' Conjecture.
0.4ex]

Conjecture 1: *For all integers* $r \geq 3$ *and a real* s *such that* $0 \leq s \leq r/r$, *if* $\nu^*(H) < s$ *then*

$$|H| \leq \max\left\{ \binom{\lceil rs \rceil - 1}{r}, \binom{n}{r} - \binom{n - \lceil s \rceil + 1}{r} \right\}.$$

In [2] we proved an asymptotic version of Conjecture 1 for $r = 3$ and $r = 4$, and $s \leq \frac{n}{r+1}$. In the proof we used a probabilistic approach based on a special case of an old probability conjecture of Samuels [10]. Samuels' Conjecture says (in a special case we are interested in) that *for all* $\mu \leq \frac{1}{r+1}$*, and all choices of* r *independent random variables* $X_1, \ldots, X_r$ *with common expectation* μ*,*

$$P(X_1 + \ldots + X_r < 1) \geq \left(1 - \frac{\mu}{1-\mu}\right)^r.$$

Samuels proved his conjecture for $r \leq 4$ in [10].

References

1 N. Alon, H. Huang, and B. Sudakov, Nonnegative k-sums, fractional covers, and probability of small deviations, submitted.
2 N. Alon, P. Frankl, H. Huang, V. Rödl, A. Ruciński, and B. Sudakov, Large matchings in uniform hypergraphs and the conjectures of Erdős and Samuels, submitted.
3 P. Erdős, A problem on independent r-tuples, Ann. Univ. Sci. Budapest. Sect. Math. **8** (1965), 93–95
4 P. Erdős, T. Gallai, On maximal paths and circuits of graphs, Acta Math. Acad. Sci. Hungar. **10** (1959), 337–356.
5 P. Frankl, The shifting technique in extremal set theory, in: *Surveys in Combinatorics, Lond. Math. Soc. Lect. Note Ser.*, **123** (1987), 81?-110.
6 P. Frankl, V. Rödl, and A. Ruciński, On the maximum number of edges in a triple system not containing a disjoint family of a given size, submitted.
7 D. Leong, A. G. Dimakis, and T. Ho, Symmetric allocations for distributed storage, *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, (2010).
8 M. Naor and R. M. Roth, Optimal file sharing in distributed networks, *SIAM J. Computing*, **24** (1991), 158–183.
9 V. Rödl and A. Ruciński, Dirac-type questions for hypergraphs – a survey (or more problems for Endre to solve), *Bolyai Soc. Math. Studies 21 (2010), An Irregular Mind (Szemerédi is 70)*, Bolyai Soc. Math. Studies **21** (2010), 561–590.
10 S. M. Samuels, On a Chebyshev-type inequality for sums of independent random variables, *Ann. Math. Statist.* **37** (1966), 248–259.
11 S. M. Samuels, More on a Chebyshev-type inequality for sums of independent random variables, Purdue Stat. Dept. Mimeo. Series no. **155** (1968).
12 M. Sardari, R. Restrepo, F. Fekri, and E. Soljanin, Memory allocation in distributed storage networks, *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, (2010).

3.20 Universally-Truthful Multi-Unit Auctions

Berthold Vöcking (RWTH Aachen, DE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Berthold Vöcking

We present a randomized, polynomial-time approximation scheme for multi-unit auctions. Our mechanism is truthful in the universal sense, i.e., a distribution over deterministically truthful mechanisms. Previously it was only known an approximation scheme that is truthful in expectation which is a weaker notion of truthfulness assuming risk neutral bidders. The existence of a universally truthful approximation scheme was questioned by previous work showing that multi-unit auctions with certain technical restrictions on their output do not admit a polynomial-time, universally truthful mechanism with approximation factor better than two.

Our new mechanism employs VCG payments in a non-standard way. In particular, the deterministic mechanisms underlying our approximation scheme are not maximal-in-range which, on a first view, seems to contradict previous characterizations of VCG-based mechanisms. Although they are not affine maximizers, each of the deterministic mechanisms is composed out of a collection of affine maximizers, one for each bidder. The composite construction ensures that the mechanism's output for a bidder coincides with the output of the affine maximizer for the bidder. This yields a subjective variant of VCG in which payments for different bidders are defined on the basis of possibly different affine maximizers.

3.21 Smoothed Analysis of Partitioning Algorithms for Euclidean Functionals

Markus Bläser (University of Saarland, DE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Markus Bläser

Euclidean optimization problems such as TSP and minimum length matching admit fast partitioning algorithms that compute optimal solutions on almost all of the input points.

We develop a general framework for the application of smoothed analysis to partitioning algorithms for Euclidean optimization problems. Our framework can be used to analyze both the running-time and the approximation ratio of such algorithms. We apply our framework to obtain smoothed analyses of Dyer and Frieze's partitioning algorithm for Euclidean matching, Karp's partitioning scheme for the TSP, a heuristic for Steiner trees, and a heuristic for bounded-degree minimum-length spanning trees.

3.22 Approximating Gale-Berlekamp Games and Related Optimization Problems

Marek Karpinski (University of Bonn, DE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Marek Karpinski
Joint work of M. Karpinski and W. Schudy

We design a linear time approximation scheme for the Gale-Berlekamp Switching Game and generalize it to much wider class of dense fragile minimization and ranking problems including the Nearest Codeword Problem (NCP), Unique Games Problem, constrained form of matrix rigidity, maximum likelihood decoding, correlation clustering with a fixed number of clusters, and the Betweenness Problem in tournaments. As a side effect of our method we obtain also the first optimal under the ETH (exponential time hypothesis) deterministic subexponential algorithm for weighted FAST (feedback arc set tournament) problem with runtime $n^{O(1)} + O^*\left(2^{O(\sqrt{OPT})}\right)$.

Our results depend on a new technique of dealing with small objective functions values of minimization problems and could be of independent interest.

References

1 M. Karpinski and W. Schudy, *Linear Time Approximation Schemes for the Gale-Berlekamp Game and Related Minimization Problems*, Proc. 41st ACM STOC (2009), pp. 313-322.

2 M. Karpinski and W. Schudy, *Approximation Schemes for the Betweenness Problem in Tournaments and Related Ranking Problems*, Proc. 15th Approx-Random (2011), LNCS 6845, Springer, 2011, pp. 277–288 .

3 M. Karpinski and W. Schudy, *Faster Algorithms for Feedback Arc Set Tournament, Kemeny Rank Aggregation and Betweenness Tournament*, Proc. 21st ISAAC (2010), pp. 3-14.

3.23 Vacant Set of a Random Walk on a Random Graph

Alan Frieze (Carnegie Mellon University – Pittsburgh, US)

License Creative Commons BY-NC-ND 3.0 Unported license
© Alan Frieze
Joint work of C. Cooper and A. Frieze

We consider random walks on several classes of graphs and explore the likely structure of the vacant set, i.e. the set of unvisited vertices. Let $\Gamma(t)$ be the subgraph induced by the vacant set of the walk at step t. We show that for random graphs $G_{n,p}$ (above the connectivity threshold) and for random regular graphs G_r, $r \geq 3$, the graph $\Gamma(t)$ undergoes a phase transition in the sense of the well-known Erdős-Renyi phase transition. Thus for $t \leq (1-\epsilon)t^*$, there is a unique giant component, plus components of size $O(\log n)$, and for $t \geq (1+\epsilon)t^*$ all components are of size $O(\log n)$. For $G_{n,p}$ and G_r we give the value of t^*, and the size of $\Gamma(t)$. For G_r, we also give the degree sequence of $\Gamma(t)$, the size of the giant component (if any) of $\Gamma(t)$ and the number of tree components of $\Gamma(t)$ of a given size $k = O(\log n)$. We also show that for random digraphs $D_{n,p}$ above the strong connectivity threshold, there is a similar directed phase transition. Thus for $t \leq (1-\epsilon)t^*$, there is a unique strongly connected giant component, plus strongly connected components of size $O(\log n)$, and for $t \geq (1+\epsilon)t^*$ all strongly connected components are of size $O(\log n)$.

3.24 On Milgram Routing

Alessandro Panconesi (University of Rome "La Sapienza", IT)

License Creative Commons BY-NC-ND 3.0 Unported license
© Alessandro Panconesi
Joint work of S. Lattanzi, A. Panconesi, D. Sivakumar

We demonstrate how a recent model of social networks ("Affiliation Networks", [1]) offers powerful cues in local routing within social networks, a theme made famous by sociologist Milgram's "six degrees of separation" experiments. This model posits the existence of an "interest space" that underlies a social network; we prove that in networks produced by this model, not only do short paths exist among all pairs of nodes but natural local routing algorithms can discover them effectively. Specifically, we show that local routing can discover paths of length $O(\log^2 n)$ to targets chosen uniformly at random, and paths of length $O(1)$ to targets chosen with probability proportional to their degrees. Experiments on the co-authorship graph derived from DBLP data confirm our theoretical results, and shed light into the power of one step of lookahead in routing algorithms for social networks.

References

1 S. Lattanzi, D. Sivakumar, *Affiliation Networks*, In STOC'09, 41 (2009), 427–434

2 S. Lattanzi, A. Panconesi, D. Sivakumar, *Milgram-Routing in Social Networks*, Proc. WWW 2011, 725–734

3.25 The Condensation Transition in Random Hypergraph 2-Coloring

Amin Coja-Oghlan (University of Warwick, GB)

License Creative Commons BY-NC-ND 3.0 Unported license
© Amin Coja-Oghlan
Joint work of A. Coja-Oghlan and L. Zdeborova

For many random constraint satisfaction problems such as random k-SAT or random graph/hypergraph coloring the best current bounds on the thresholds for the existence of solutions are derived via the *first* and the *second moment method*. However, in most cases these simple techniques do not yield matching upper and lower bounds. In effect, for most random CSPs the *precise* threshold for the existence of solutions assignments remains unknown. Examples of this include random k-SAT, random graph k-coloring, or random k-uniform hypergraph 2-coloring ($k \geq 3$). Here we discuss the example of random hypergraph 2-coloring, a case in which the second moment analysis is technically quite simple. We present an approach to improve slightly over the naive second moment argument. But more importantly, we establish the existence of a phase transition below the threshold for the existence of solutions poses a genuine obstacle to the second moment argument. The existence of this so-called *condensation transition* was hypothesized on grounds of non-rigorous statistical mechanics arguments [3].

To define the *random hypergraph 2-coloring* problem, let $V = \{1, \ldots, n\}$ be a (large) set of vertices, and let $H_k(n, m)$ be a random k-uniform hypergraph on V obtained by inserting a random set of m edges (each containing k vertices). A 2-*coloring* of H is a map $\sigma\colon V \to \{0, 1\}$ such that no hyperedge e is monochromatic. We let $m = \lceil r \cdot n \rceil$ for some fixed number r (independent of n). Friedgut's sharp threshold theorem implies that there *exists* a threshold $r_{col} = r_{col}(n)$ such that for any $\varepsilon > 0$ the random hypergraph $H_k(n, m)$ of density $m/n < (1 - \varepsilon)r_{col}$ is 2-colorable with high probability, while in the case $m/n > (1 + \varepsilon)r_{col}$ w.h.p. no 2-coloring exists. The first and the second moment methods can be used to estimate the threshold [1]:

$$r_{second} = 2^{k-1} \ln 2 - (1 - \ln 2)/2 \leq r_{col} \leq r_{first} = 2^{k-1} \ln 2 - \ln 2/2. \tag{1}$$

As observed in [1], for $r > r_{second}$ we have $\mathrm{E}\left[Z^2\right] > \exp(\Omega(n)) \cdot \mathrm{E}\left[Z^2\right]$, i.e., the second moment method fails dramatically. But why? First, it could be the case that the *expectation* $\mathrm{E}[Z]$ is driven up by a tiny fraction of hypergraphs with excessively many 2-colorings, and thus $Z \ll \mathrm{E}[Z]$ w.h.p. Second, it could be that Z is 'close' to $\mathrm{E}[Z]$ with high probability, but without being sufficiently concentrated for the second moment method to apply. The following theorem, which improves the lower bound in (1) by an additive $(1-\ln(2))/2 \approx 0.153$, shows that the second scenario is true.

▶ **Theorem 1.** *There is a sequence $\varepsilon_k \to 0$ such that for*

$$r \leq r_{enhanced} = 2^{k-1} \ln 2 - \ln 2 - \varepsilon_k$$

the random formula $\vec{\Phi}$ is NAE-satisfiable and $\ln Z \sim \ln \mathrm{E}[Z]$ w.h.p.

Even the enhanced second moment argument from Theorem 1 does not give the precise threshold for 2-colorability. The intuitive reason is that for $r > r_{enhanced}$, the expected number $\mathrm{E}[Z]$ is indeed dominated by a tiny fraction of hypergraphs with an abundance of 2-colorings.

▶ **Theorem 2.** *There exist $\varepsilon_k \to 0$, $\delta_k > \varepsilon_k$, and $\zeta_k > 0$ such that the following two statements are true.*

1. *The 2-colorability threshold satisfies $r_{col} > 2^{k-1}\ln 2 - \ln 2 + \delta_k$.*
2. *For any $2^{k-1}\ln 2 - \ln 2 + \varepsilon_k < r < r_k$ we have*

$$\ln Z < \ln \mathrm{E}[Z] - \zeta_k n \qquad \textit{w.h.p.} \tag{2}$$

The first statement shows that indeed $H_k(n,m)$ remains 2-colorable w.h.p. for (at least a small range of) densities $r > r_{enhanced}$. Moreover, the second statement asserts that for densities between $r_{enhanced}$ and the true threshold r_k for 2-colorability, the *expected* number $\mathrm{E}[Z]$ of 2-colorings exceeds the *acutal* number Z by an exponential factor $\exp(\zeta_k n)$ w.h.p. This constrasts with Theorem 1, which shows that below $r_{enhanced}$, Z is of the same exponential order as $\mathrm{E}[Z]$ w.h.p. This so-called *condensation transition* at density $2^{k-1}\ln 2 - \ln 2$ was hypothesized on the basis of non-rigorous statistical mechanics arguments [3].

References

1 D. Achlioptas, C. Moore: Random k-SAT: two moments suffice to cross a sharp threshold. SIAM Journal on Computing **36** (2006) 740–762.
2 E. Friedgut: Sharp thresholds of graph properties, and the k-SAT problem. Journal of the AMS **12** (1999) 1017–1054.
3 F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborova: Gibbs states and the set of solutions of random constraint satisfaction problems. Proc. National Academy of Sciences **104** (2007) 10318–10323.

3.26 Linear Index Coding via Semidefinite Programming

Eden Chlamtac (Tel Aviv University, IL)

License Creative Commons BY-NC-ND 3.0 Unported license
© Eden Chlamtac
Joint work of E. and I. Haviv

In the index coding problem, introduced by Birk and Kol (INFOCOM, 1998), the goal is to transmit n bits to n receivers (one bit to each), where the receivers reside at the nodes of a graph G and have prior access to the bits corresponding to their neighbors in the graph (side information). The objective is to find a code word of minimum length which will allow each receiver to learn their own bit given access to the code word and their side information. When the encoding is linear (this is known as linear index coding), the minimum possible code word length corresponds to a graph parameter known as the minrank of G.

In this talk, we will describe an algorithm which approximates the minrank of a graph in the following sense: when the minrank of the graph is a constant k, the algorithm finds a linear index code of length $O(n^{f(k)})$. For example, for $k = 3$ we have $f(3) \approx 0.2574$. This algorithm exploits a connection between minrank and a semidefinite programming (SDP) relaxation for graph coloring introduced by Karger, Motwani and Sudan.

A result which arises from our analysis, and which may be of independent interest, gives an exact expression for the maximum possible value of the Lovasz theta-function of a graph, as a function of its minrank. This compares two classical upper bounds on the Shannon capacity of a graph.

References

1 E. Chlamtac, I. Haviv, *Linear Index Coding via Semidefinite Programming*, http://arxiv.org/abs/1107.1958

4 Informal Session

4.1 Geometric MAX-CUT Problem

Marek Karpinski (University of Bonn, DE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Marek Karpinski

We raise an open problem on computational status of exact geometric MAX-CUT problem on a real line and the plane. Polynomial time approximation schemes are known for arbitrary metric MAX-CUT problems. The status of the exact geometric MAX-CUT eludes us however completely for both NP-hardness or existence of exact polynomial time algorithms.

4.2 Streaming algorithms for the analysis of massive data sets

Christian Sohler (TU Dortmund, DE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Christian Sohler

Massive data sets occur in many applications of Computer Science. Examples include the WWW, internet traffic logs, and operating system calls. Often data is read sequentially in the form of a data stream, which is too large to be stored in main memory and sometimes even too large to be stored at all. If we want to analyze such massive data sets to, say, build a search engine, detect spreading viruses, or optimize a system's performance, we need special algorithms that use only little memory and process the input sequentially. Such algorithms are called streaming algorithms. In this talk I will give an introduction to streaming algorithms and explain the two major algorithmic concepts used in this area. I will discuss their applications in the development of streaming algorithms for data analysis and close with a discussion of future directions of research.

5 Open Problem Session

5.1 Star Cover Problems

R. Ravi (Carnegie Mellon University – Pittsburgh, US)

License Creative Commons BY-NC-ND 3.0 Unported license
© R. Ravi

Given a set of points S, a distance function d, an integer k and a bound B on maximum load of any facility, the problem is to decide whether we can select k facilities in S and serve other points by these facilities such that the maximum load of any facility is B. The load on a facility f is the sum of distances to the clients it serves. Alternately, since an open facility together with the assigned clients, forms a star and its load is exactly equal to the cost of the star, we can call this the Star Cover Decision Problem (SCDP). We consider two optimization versions of the above stated decision problem:

- Minimum Load Star Cover or MLSC : Given S, d and an integer k, minimize the maximum load on each facility.
- Minimum Cardinality Star Cover or MCSC : Given S, d and a bound B on maximum load, minimize the number of facilities to be opened, such that the load on each facility is at most B.

We can also consider a bicriteria approximation for the decision problem i.e. given S, d, k and B find an (α, β)-approximation, such that at most αk facilities are opened and maximum load on a facility is at most βB.

Known results:

Arkin, Hassin and Levin consider the Minimum Cardinality Star Cover problem, where distance d is a metric and give a $(2\alpha + 1)$-approximation for the problem, where α is the best approximation ratio of the k-median problem.

Even, Garg, Konemann, Ravi and Sinha give a bicriteria approximation of $(4, 4)$ for the case when d is a metric i.e. for given k and B, their algorithm opens at most $4k$ facilities and the completion time is at most $4B$. This is improved to a $(3 + \epsilon, 3 + \epsilon)$ approximation by Arkin, Hassin and Levin.

The star cover decision problem (SCDP) is NP-complete, even when the distance function d is a line metric or a star metric. Furthermore, the problem remains hard even if the facilities to open are specified. The proofs of hardness for line and star metrics are by reductions from 3-PARTITION and MAKESPAN respectively.

The Minimum Cardinality Star Cover or MCSC problem is $\Omega(\log n)$-hard to approximate if the distance function, d, is not a metric, by a reduction from set cover, where $|S| = n$. A similar analysis as that of greedy algorithm for set cover, gives a $\log n$-approximation for the MCSC problem in the general case. There is a 2-approximation for the case when the distance function is a line metric.

For the Minimum Load Star Cover or MLSC problem, the LP relaxation of a natural IP formulation has a large integrality gap. There is a 3-approximation when the distance function is a star metric. In the case of distance function being a line metric, it can be shown that if every point is assigned to either the closest facility to the right or closest facility to the left then the maximum load can be a factor k times worse than the optimum.

Open Problem: Give a nontrivial approximation for the MLSC problem or show a lower bound for the approximation factor. Even the case when the input is a line metric is open.

References

1. E. Arkin, R. Hassin and A. Levin, *Approximations for minimum and min-max vehicle routing problems*, Journal of Algorithms, 59(1), 2006.
2. G. Even, N. Garg, J. Könemann, R. Ravi, A. Sinha, *Covering Graphs using Trees and Stars*. In Proc of the 6th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, 2003.

5.2 Approximating the Euclidean TSP cost in a data stream

Christian Sohler (TU Dortmund, DE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Christian Sohler

Assume you are given access to a stream of points from a discrete space $\{1, \dots, \Delta\}^d$, i.e. the data points arrive sequentially in worst case order. A streaming algorithm is an algorithm that processes the stream of points and uses space polylogarithmic in Δ and the number of points. Is it possible to approximate the cost of the minimum Euclidean TSP problem within a constant smaller than 2 in this model (possibly even upto a factor of $1 + \epsilon$)? It is known that one can approximate the cost of a minimum spanning tree upto a factor of $1 + \epsilon$ [1], so a $(2 + \epsilon)$-approximation of the TSP cost is possible.

References

1 G. Frahling, P. Indyk and Christian Sohler, *Sampling in Dynamic Data Streams and Applications*, Int. J. Comput. Geometry Appl. 18, 1/2, 2008, pp. 3–28,

5.3 Probabilistic Analysis of Local Search for the Max-Cut Problem

Heiko Röglin (University of Bonn, DE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Heiko Röglin

In the max-cut problem we are given an undirected graph $G = (V, E)$ with a weight function $w : E \to \mathbb{R}_{>0}$ and we want to compute a partition of the vertices into two classes such that the total weight of the edges crossing the cut becomes maximal. We consider a simple local search heuristic that starts with an arbitrary cut and improves this cut by moving one vertex from one side of the cut to the other as long as such a local improvement is possible.

We are interested in the number of iterations until a local optimum is reached. In the worst-case this number can only be bounded by $\Theta(\sum_{e \in E} w(e))$. We suspect, however, that the expected number of steps is polynomial in the size of the graph if every edge weight is chosen uniformly at random from the interval $[0, 1]$. So far, we have not been able to prove this conjecture. We even conjecture that also in the framework of smoothed analysis the number of iterations becomes polynomial in the size of the graph and the perturbation parameter.

5.4 Approximation Hardness of TSP

Marek Karpinski (University of Bonn, DE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Marek Karpinski

We discussed approximation hardness results for (metric) TSP as well as asymmetric TSP and raised a question on an existence of direct PCP constructions for those problems for proving stronger approximation hardness results.

5.5 Questions about permantents of nonnegative matrices

Alex Samorodnitsky (The Hebrew University of Jerusalem, IL)

License Creative Commons BY-NC-ND 3.0 Unported license
© Alex Samorodnitsky

We present two conjectures, which, if true, might be useful for approximate counting of some classes of contingency tables (following joint work with Alexander Barvinok).

Conjecture 1: Let $A = (a_{ij})$ be an $n \times n$ stochastic matrix, such that $\sum_{i,j} a_{ij}^2 \leq K$ (think about K as a constant). Then

$$\text{per}(A) \leq n^{K'} \cdot e^{-n},$$

where K' depends only on K.□

Conjecture 2: (This was also independently conjectured by Caputo, Carlen, Lieb and Loss.)

Let $1 \leq p \leq \infty$. The maximum of the permanent of $n \times n$ matrices whose rows are unit vectors in l_p^n attained either at the identity matrix, or at the matrix all of whose entries equal $n^{-1/p}$.

5.6 Understanding the approximability of Graph Balancing

Ola Svensson (KTH – Stockholm, SE)

License Creative Commons BY-NC-ND 3.0 Unported license
© Ola Svensson

One of the most prominent open problems in scheduling theory is to understand whether Lenstra, Shmoys, & Tardos' beautiful 2-approximation algorithm for scheduling jobs on unrelated machines to minimize the makespan can be improved. This problem has been open for more than two decades with little progress. Researchers have therefore started to work on improved algorithms for special cases with the hope to shed light on the more general problem.

It was with this motivation Ebenlendr, Krcal, and Sgall introduced a natural special case, named *Graph Balancing*, defined as follows: given an undirected graph $G = (V, E)$ with weights $w : E \mapsto \mathbb{R}_+$ on the edges, find an orientation of the edges so as to minimize $\max_{v \in V} \sum_{e \in \delta^-(v)} w(e)$ where δ^- denotes the edges directed towards v in the given orientation. In their paper, they give an 1.75-approximation algorithm for Graph Balancing and they show that it is NP-hard to approximate within a factor less than 1.5 (which is the same lower bound as known for the general problem of scheduling on unrelated machines).

n order to obtain their 1.75-approximation algorithm they strengthen the linear program used by Lenstra, Shmoys, & Tardos by adding certain linear inequalities. They then show that the strengthened linear program has an integrality gap of at most 1.75 and they also give instances where this gap is achieved. In order to improve the approximation guarantee further one needs thus an even stronger lower bound. One promising strong lower bound is given by a certain strong linear program known as configuration LP. The configuration LP is believed to be strong but our understanding of it remains rather weak although it has been successfully used to obtain better bounds for the more general restricted assignment problem and for the Santa Claus problem.

The worst case integrality gap instances known for the configuration LP for the restricted assignment problem (and thus Graph Balancing) achieves a gap of 1.5. I strongly believe that this is also the upper bound of the integrality gap and the first step would be to prove this for the Graph Balancing problem. I therefore think progress on the following open problem would be valuable for increasing our understanding of this kind of linear programs and also for developing tools for a better understanding of assignment problems.

Open problem: Show that Graph Balancing has a 1.5-approximation algorithm. In particular, it would be interesting to prove an upper bound of 1.5 on the integrality gap of the configuration LP.

5.7 Random TSP

Alan Frieze (Carnegie Mellon University – Pittsburgh, US)

License Creative Commons BY-NC-ND 3.0 Unported license
© Alan Frieze

Let $C[i,j]$ be an $n \times n$ symmetric matrix where the $C[i,j], i \leq j$ are independent uniform $[0,1]$ random variables. Let F be the minimum weight 2-factor using C as weights. How any cycles $\sigma(F)$ does F have w.h.p.? It is known that w.h.p. $\sigma(F) = O(n/\log n)$. Can this be improved to $o(n/\log n)$? On the face of it, this should be easy. The number of cycles in a random 2-factor on n vertices has $O(\log n)$ cycles. A positive result will simplify algorithms for finding low cost traveling salesman problems with these weights.

References

1 Alan Frieze, *On random symmetric travelling salesman problems*, Mathematics of Operations Research (2004) 878-890.

5.8 Perfect Matchings in k-Uniform Hypergraphs

Andrzej Rucinski (Adam Mickiewicz University – Poznan, PL)

License Creative Commons BY-NC-ND 3.0 Unported license
© Andrzej Rucinski

Let $k \geq 2$ and n be divisible by k. We denote by $m(k,n)$ the minimum m so that for every n-vertex k-uniform hypergraph H, $\delta(H) \geq m$ implies that H has a perfect matching. It is easy to show that $m(2,n) = \lfloor n/2 \rfloor$. It has been conjectured in [6] and again in [3] that $m(k,n) \sim \left(1 - \left(\frac{k-1}{k}\right)^{k-1}\right)\binom{n-1}{k-1}$. The conjecture has been proved for $k = 3$ in [3], [7], and [4], for $k = 4$ in [5], [8], and [2], and for $k = 5$ in [1]. The proof of this last result is based on a lemma which says that $m(k,n) \sim f(k,n)$, where $f(k,n)$ as the minimum m so that $\delta_d(H) \geq m$ implies that H has a *fractional* perfect matching. (Observe that trivially $f(k,n) \leq m(k,n)$.) That lemma reduces the task of finding (asymptotically) $m(k,n)$ to showing a presumambly simpler conjecture that $f(k,n) \sim \left(1 - \left(\frac{k-1}{k}\right)^{k-1}\right)\binom{n-1}{k-1}$.

Problem: Determine, at least asymptotically, $f(k,n)$ and thus $m(k,n)$ for $k \geq 6$.

References

1 N. Alon, P. Frankl, H. Huang, V. Rödl, A. Ruciński, and B. Sudakov, *Large matchings in uniform hypergraphs*, submitted. and the conjectures of Erdős and Samuels

2 P. Frankl, V. Rödl, and A. Ruciński, *On the maximum number of edges in a triple system not containing a disjoint family of a given size*, submitted.
3 H. Hàn, Y. Person, and M. Schacht, *On perfect matchings in uniform hypergraphs with large minimum vertex degree*, SIAM J. Discrete Math. 23(2) (2009), 732–748.
4 I. Khan, Perfect Matching in 3-uniform hypergraphs with large vertex degree, submitted.
5 I. Khan, Perfect matchings in 4-uniform hypergraphs, submitted.
6 D. Kühn and D. Osthus, *Embedding large subgraphs into dense graphs*, Surveys in Combinatorics (editors S. Huczynka, J. Mitchell, C.Roney-Dougal) London Math. Soc. Lecture Notes, Cambridge University Press 365 (2009), 137-167.
7 D. Kühn, D. Osthus, and A. Treglown, *Matchings in 3-uniform hypergraphs*, submitted.
8 A. Lo and K. Markström, *F-factors in hypergraphs via absorption*, submitted

6 Seminar Schedule

Tuesday, June 14th, 2011

09:00–09:10	Opening
Chair:	Marek Karpinski
09:10–09:40	Johan Håstad: *On the Usefulness of Predicates*
09:40–10:10	Alexander Barvinok: *Counting Contingency Tables*
10:10–10:40	Leslie Ann Goldberg: *Estimating the Partition Function of the Ferromagnetic Ising Model on a Regular Matroid*
Chair:	Uriel Feige
11:00–11:30	Leen Stougie: *TSP on Cubic and Subcubic Graphs*
11:30–12:00	Ola Svensson: *Approximating Graphic TSP by Matchings*
Chair:	Martin Dyer
15:00–15:30	Po-Shen Loh: *Connectivity in Discrete Random Processes*
15:30–16:00	Lars Prädel: *A* $(5/3+\epsilon)$*-Approximation for Strip Packing*
Chair:	Alan Frieze
16:30–17:00	Christian Sohler: *Every Hyperfinite Property is Testable*
17:00–17:30	Alexandr Andoni: *Sublinear Algorithms via Precision Sampling*

Wednesday, June 15th, 2011

Chair:	Mark Jerrum
09:00–09:30	Thomas Hayes: *Lifting Markov Chains for Faster Mixing*
09:30–10:00	Gregory Sorkin: *Average-Case Performance of Heuristics for Multi-Dimensional Assignment Problems*
10:00–10:30	Michael Krivelevich: *Embedding Spanning Trees in Random Graphs near the Connectivity Threshold*
Chair:	Michael Paterson
11:00 –11:30	R. Ravi: *Stochastic Knapsack*
11:30 –12:00	Heiko Röglin: *Smoothed Analysis of Multiobject Optimization*
13:30–17:30	Excursion
20:00	Open Problem Session
Chair:	Uriel Feige
Speakers:	R. Ravi, Christian Sohler, Heiko Röglin, Marek Karpinski, Alex Samorodnitsky, Ola Svensson, Alan Frieze, Andrzej Rucinski

Thursday, June 16th, 2011

Chair:	Leslie Ann Goldberg
09:00–09:30	Ravi Kannan: *k-Means Algorithm Converges*
09:30–10:00	Tobias Müller: *Random Geometric Graphs*
10:00–10:30	Mark Jerrum: *Hardness of Approximating the Tutte Polynomial of a Binary Matroid*
Chair:	Alan Frieze
11:00–11:30	Edyta Szymanska: *Computational Complexity of the Hamilton Cycle Problem in Dense Hypergraphs*
11:30–12:00	Andrzej Rucinski: *Distributed Storage Allocation via Fractional Hypergraph Matchings*
Chair:	Michael Krivelevich
15:00–15:30	Berthold Vöcking: *Universally-Truthful Multi-Unit Auctions*
15:30–16:00	Markus Bläser: *Smoothed Analysis of Partitioning Algorithms for Euclidean Functionals*

Friday, June 17th, 2011

Chair:	Johan Håstad
09:00–09:30	Marek Karpinski: *Approximating Gale-Berlekamp Games and Related Optimization Problems*
09:30–10:00	Alan Frieze: *Vacant Set of a Random Walk on a Random Graph*
10:00–10:30	Alessandro Panconesi: *Milgram Routing*
Chair:	Ravi Kannan
11:00–11:30	Amin Coja-Oghlan: *The Condensation Transition in Random Hypergraph 2-Coloring*
11:30–12:00	Eden Chlamtac: *Linear Index Coding via Semidefinite Programming*

Participants

- Alexandr Andoni
Microsoft Research – Mountain View, US
- Alexander Barvinok
University of Michigan – Ann Arbor, US
- Markus Bläser
Universität des Saarlandes, DE
- Eden Chlamtac
Tel Aviv University, IL
- Amin Coja-Oghlan
University of Warwick, GB
- Colin Cooper
King's College – London, GB
- Artur Czumaj
University of Warwick, GB
- Martin Dyer
University of Leeds, GB
- Uriel Feige
Weizmann Inst. – Rehovot, IL
- Alan M. Frieze
Carnegie Mellon University – Pittsburgh, US
- Leslie Ann Goldberg
University of Liverpool, GB
- Johan Håstad
KTH – Stockholm, SE
- Mathias Hauptmann
Universität Bonn, DE
- Thomas Hayes
University of New Mexico – Albuquerque, US
- Mark Jerrum
Queen Mary University – London, GB
- Michael Kaminski
Technion – Haifa, IL
- Ravindran Kannan
Microsoft Research India – Bangalore, IN
- Marek Karpinski
Universität Bonn, DE
- Michael Krivelevich
Tel Aviv University, IL
- Po-Shen Loh
Carnegie Mellon University – Pittsburgh, US
- Tobias Müller
CWI – Amsterdam, NL
- Alessandro Panconesi
University of Rome "La Sapienza", IT
- Michael Paterson
University of Warwick, GB
- Lars Praedel
Universität Kiel, DE
- R. Ravi
Carnegie Mellon University – Pittsburgh, US
- Heiko Röglin
Universität Bonn, DE
- Andrzej Rucinski
Adam Mickiewicz University – Poznan, PL
- Alex Samorodnitsky
The Hebrew University of Jerusalem, IL
- Richard Schmied
Universität Bonn, DE
- Christian Sohler
TU Dortmund, DE
- Gregory B. Sorkin
London School of Economics, GB
- Leen Stougie
CWI – Amsterdam, NL
- Ola Svensson
KTH – Stockholm, SE
- Edyta Szymanska
Adam Mickiewicz University – Poznan, PL
- Claus Viehmann
Universität Bonn, DE
- Berthold Vöcking
RWTH Aachen, DE