

Summary of Papers

Hemanshu Kaul
email: kaul@math.iit.edu

[1] **Global Optima Results for the Kauffman NK Model** (with S.H. Jacobson), *Mathematical Programming*, Volume 106, 2006, 319-338.

The Kauffman NK model is a stochastic combinatorial optimization model that has been used in theoretical biology, physics and management science to model complex systems with interacting components. It could be crudely described as an optimization problem to find a maximum valued (weighted) p -nary vector of length N with the value (weight) of a vector defined in terms of its components' weights and their 'interaction' with K 'neighboring' components. This paper analyzes global optima of the NK model. Most previous papers focused on local optima and simulation based results.

We transform this NP-hard global optimization problem into a stochastic network model that is closely related to two well-studied problems in operations research - project duration in PERT networks and stochastic shortest path problem. This transformation leads to applicable strategies for explicit computation of bounds on the global optima (particularly with K either small or close to N), such as a recursive scheme applicable as a dynamic program and simple stochastic networks that can be processed simultaneously. A general lower bound, which is sharp for $K = 0$, is obtained for the expected value of the global optimum of the NK model in terms of the order statistics of the underlying distribution. We also give a detailed analysis for the expectation and variance of the global optimum when $K = N - 1$ and the underlying distribution is $U(0, 1)$, by converting the analytic problem into a geometric one with estimation of volumes of certain bodies in the N -dimensional hypercube. The lower and upper bounds on the expectation obtained for this case show that there is a wide gap between the values of the local and the global optima. They also indicate that the complexity catastrophe, the tendency of the local optima to collapse towards average behavior, does not arise for the global optima.

[2] **New Global Optima Results for the Kauffman NK Model: Handling Dependency**, (with S.H. Jacobson), *Mathematical Programming*, Special issue on 'Optimization under uncertainty', Volume 108, 2006, 475-494.

This paper generalizes and extends the work from the first paper [1] that focused on the analysis of the (independent) case $K = N - 1$. It presents new global optima results for the NK model by developing tools for handling the dependency between weight functions of different N -vectors due to overlapping weight contributions from their components, when $K \leq N - 2$. Previous papers used Markov chain theory to analyze the cases when $K = 1$, N tends to infinity and the underlying distributions are exponential or negative exponential. The ideas developed here are more combinatorial in nature, independent of specific underlying distributions and especially applicable to K growing with N .

We define and study a dependency graph to handle dependencies among underlying random variables in the NK model. Equitable coloring of the dependency graph is used to bound general order statistics (with dependencies) and consequently, the expected value of the global optima, $\mathbf{E}_{N,K}$, for the NK model. These bounds convert the problem of bounding order statistics of dependent random variables into that of independent random variables while incorporating quantitative information about the mutual dependencies between the underlying random variables. An alternative upper bound on $\mathbf{E}_{N,K}$ using direct arguments is also proposed. These ideas for handling dependence are applied to give a detailed analysis of $\mathbf{E}_{N,K}$ for K close to N ($K = N - \alpha$ and $K = \beta N$, $\alpha \in \mathbf{Z}^+$, $\beta \in (0, 1)$) with sharp bounds when the underlying distribution is normal by

using tools from order statistics theory, and when the underlying distribution is uniform by extending the geometric ideas from the first paper [1]. Finally, for bounded underlying distributions, the global optima is shown to be concentrated around its mean $\mathbf{E}_{N,K}$.

[3] **Extremal Graphs for a Graph Packing Theorem of Sauer and Spencer**, (with A. Kostochka), *Combinatorics, Probability and Computing*, Volume 16, 2007, 409-417.

Let G, H be graphs with maximum degrees $\Delta(G) = \Delta_1, \Delta(H) = \Delta_2$, and orders $n(G), n(H) \leq n$. G and H are said to pack if there exist injective mappings of the vertex sets into $[n]$, such that the images of the edge sets do not intersect. In other words, either G or H is isomorphic to a subgraph of the complement of the other. The concept of graph packing generalizes various extremal graph problems, including problems on existence of fixed subgraphs (such as the Hamiltonian Cycle problem), forbidden subgraphs (Turán-type problems), and equitable coloring. One of the classical results in this area was by Sauer and Spencer (1978) : if $2\Delta_1\Delta_2 < n$ then G and H pack.

We characterize the graphs that achieve equality in the condition on maximum degrees as given in the Sauer-Spencer result for packing of graphs. We show that: If $2\Delta_1\Delta_2 \leq n$, then G and H do not pack if and only if one of G or H is a perfect matching and the other either is $K_{\frac{n}{2}, \frac{n}{2}}$ with $\frac{n}{2}$ odd or contains $K_{\frac{n}{2}+1}$. This gives the Sauer-Spencer theorem as a corollary. This result can be thought of as small step towards the well-known Bollobás-Eldridge conjecture (described in [5] below).

[4] **Analyzing the Performance of Simultaneous Generalized Hill Climbing Algorithms**, (with D.E. Vaughan and S.H. Jacobson), *Computational Optimization and Applications*, Volume 37, 2007, 103-119.

Simultaneous generalized hill climbing (SGHC) algorithms provide a framework for using heuristics to simultaneously address sets of intractable discrete optimization problems where information is shared between the problems during the algorithm execution. A SGHC algorithm probabilistically moves between discrete optimization problems during its execution according to a (problem generation) probability function. Many well-known heuristics (Generalized Hill Climbing (GHC) algorithms), including simulated annealing, threshold accepting and pure local search, can be embedded within the SGHC algorithm framework.

This paper shows that the solutions generated by an SGHC algorithm are a stochastic process that satisfies the Markov property. This allows the problem probability mass functions to be formulated for particular sets of problems based on the long-term behavior of the algorithm. Such results can be used to determine the proportion of iterations that an SGHC algorithm will spend optimizing over each discrete optimization problem. Sufficient conditions that guarantee that the algorithm spends an equal number of iterations in each discrete optimization problem are provided. SGHC algorithms can also be formulated such that the overall performance of the algorithm is independent of the initial discrete optimization problem chosen. Sufficient conditions are obtained guaranteeing that an SGHC algorithm will visit the globally optimal solution for each discrete optimization problem. Lastly, rates of convergence for SGHC algorithms are reported that show that given a rate of convergence for the embedded GHC algorithm, the SGHC algorithm can be designed to preserve this rate.

[5] **On a Graph Packing Conjecture of Bollobás, Eldridge, and Catlin**, (with A. Kostochka and G. Yu), *Combinatorica*, Volume 28, 2008, 469-485.

See the description in [3] above for definition of the graph packing problem. One of the earliest results in this area was by Sauer and Spencer(1978) : if $\Delta_1\Delta_2 < \frac{1}{2}n$ then G and H pack. The main conjecture in the area was made by Bollobás and Eldridge in 1978 as an extension of the Sauer–

Spencer result: if $(\Delta_1 + 1)(\Delta_2 + 1) \leq n + 1$ then G and H pack. If true, this conjecture would be sharp, and would be a considerable extension of the Hajnal-Szemerédi theorem on equitable colorings. The conjecture has only been proved when $\Delta_1 \leq 2$, or $\Delta_1 = 3$ and n is huge.

This paper focuses on proving a result of the form : for a fixed $0 \leq \epsilon \leq 1$, $(\Delta_1 + 1)(\Delta_2 + 1) \leq \frac{n}{2}(1 + \epsilon) + 1 \Rightarrow G$ and H pack. For $\epsilon = 0$, this is essentially the Sauer-Spencer result, while $\epsilon = 1$ gives the Bollobás-Eldridge conjecture. Thus, for any $\epsilon > 0$ this would improve the Sauer-Spencer result. We have proved this result for $\epsilon = 0.2$. That is, we have essentially improved the bound on the product of maximum degrees in the Sauer-Spencer theorem from $(0.5)n$ to $(0.6)n$. This is the current best result towards the Bollobás-Eldridge Conjecture.

[6] Long Local Searches for Large Bipartite Subgraphs, (with D.B. West), SIAM Journal on Discrete Mathematics, Volume 22, 2008, 1138-1144.

We study a basic local-search algorithm for finding a bipartite subgraph with maximum number of edges in a given graph. Starting with an arbitrary vertex partition, move ('flip') a vertex from one partite set to the other if doing so increases the number of edges in the cut. We improve the previous best-known lower bound, $\frac{1}{2}n^{3/2}$, on the maximum number of flips possible in a graph with n vertices to $\frac{2}{25}n^2$. Note that $\frac{1}{4}(n^2)$ is a trivial upper bound. We also prove better upper bounds like $e(G) - \frac{1}{4}d^2$, in terms of d the minimum degree of the graph. We also show that the minimum number of flips needed to reach a global optimum is at most $n/2$, answering a question of Cowen and West.

[7] Distinguishing Chromatic Number of Cartesian Products of Graphs, (with J. Choi and S. Hartke), SIAM Journal on Discrete Mathematics, to appear.

The distinguishing chromatic number of a graph G , $\chi_D(G)$, is the least number of colors needed for a proper coloring of G with the property that the only color-preserving automorphism of G is the identity. That is, we want to give a proper coloring of a graph that breaks all its symmetries, so that the coloring together with the structure of the graph uniquely determines the vertices. This can be thought of as an exact encoding of the vertices using only a proper coloring. It is a common extension of both the chromatic number and the distinguishing number of graphs.

The chromatic number, $\chi(G)$, is an immediate lower bound for $\chi_D(G)$. We show that $\chi_D(G)$ can be surprisingly at most one worse than $\chi(G)$ for G a Cartesian power of any graph. The main theorem is : For every graph G , there exists a constant d_G (explicitly defined) such that for all $d \geq d_G$, $\chi_D(G^d) \leq \chi(G) + 1$, where G^d denotes the Cartesian product of d copies of G . Using our proof techniques, we also find the distinguishing chromatic number for Hypercubes, Hamming graphs (Cartesian products of complete graphs), and Cartesian products of complete multipartite graphs.

[8] Reductions for the Stable Set Problem, (with E.C. Sewell and S.H. Jacobson), submitted for publication.

One approach to finding a maximum stable (independent) set (MSS) (or, equivalently a maximum clique or a minimum vertex cover) in a graph is to try to reduce the size of the problem by transforming the problem into an equivalent problem on a smaller graph. These reductions have been used to study properties of stability critical graphs and facets of the stable set polytope. They have also been used algorithmically in heuristics, polynomial-time algorithms for special classes of graphs, and exact algorithms.

This paper introduces several new reductions for the MSS problem, extends several well known reductions to the maximum weight stable set (MWSS) problem, demonstrates how reductions for

the generalized stable set problem can be used in conjunction with probing to produce powerful new reductions for both the MSS and MWSS problems, and shows how hypergraphs can be used to expand the capabilities of clique projections. The effectiveness of these new reduction techniques are illustrated on a set of challenging MSS problems arising from Steiner Triple Systems.

[9] **Maximum Series-Parallel Subgraph: Approximation Algorithms**, (with G. Calinescu and C.G. Fernandes), submitted for publication.

Consider the NP-hard problem of, given a simple graph G , to find a K_4 -minor-free subgraph (series-parallel subgraph) of G with the maximum number of edges. The algorithm that, given a connected graph G , outputs a spanning tree of G , is a $\frac{1}{2}$ -approximation. Indeed, if n is the number of vertices in G , any spanning tree in G has $n - 1$ edges and any series-parallel graph on n vertices has at most $2n - 3$ edges. We present a $\frac{7}{12}$ -approximation algorithm (current best) for this problem and, constructions and complexity results showing the limits of our approach.

[10] **Forbidden Subgraphs of Unit Disk Graphs**, (with R. Martin), submitted for publication.

A graph $G = (V, E)$ is called a unit disk graph (UDG) if there is a function $f : V(G) \rightarrow \mathbb{R}^2$ such that for $v, w \in V(G)$, $\|f(v) - f(w)\| \leq 1$ if and only if $\{v, w\} \in E(G)$. UDGs are a natural model for wireless and radio networks. UDG is an example of a hereditary property. For every hereditary property \mathcal{H} , there exists a family of graphs $\mathcal{F}(\mathcal{H})$ such that $\mathcal{H} = \bigcap_{H \in \mathcal{F}(\mathcal{H})} \text{Forb}(H)$, where $\text{Forb}(H)$ is the family of graphs with no *induced* copy of H . This together with the fact that recognition of UDGs is NP-hard motivates the study of family of forbidden induced subgraphs of UDG. For example, $K_{1,6}$ was proved to be a forbidden induced subgraph of UDG and this fact was used to give approximation algorithms for various NP-hard problems on UDGs.

We prove that $K_{2,3}$ among other small graphs are forbidden in UDGs. We also give a infinite family of graphs on 14 vertices that are also forbidden. These results are also generalized to Unit Ball graphs in d -dimensions. These results are then applied to an application of Szemerédi's regularity lemma to the edit-distance problem for large UDGs.

[11] **Improved Bound for Guarding Orthogonal Art Galleries with Holes**, (with Y. Jo), preprint.

The original art gallery problem (V.Klee, 1973) asked for the minimum number of guards sufficient to see every point of the interior of an n -vertex simple polygon (Art Gallery). Chvátal (1975) proved that $\lfloor n/3 \rfloor$ guards are always sufficient. If all the edges of the given simple polygon are either horizontal or vertical, then such a polygon is called an orthogonal gallery. Kahn, Klawe and Kleitman (1983) proved that $\lfloor n/4 \rfloor$ guards are sufficient for such a n -vertex gallery.

We study orthogonal gallery with holes, i.e., an orthogonal polygon enclosing some other orthogonal polygons called holes (interior of each hole is empty). In 1982, Shermer conjectured that any orthogonal polygon with n vertices and h holes can be guarded by $\lfloor (n + h)/4 \rfloor$ vertex guards. This conjecture remains open. The best known result shows that $\lfloor (n + 2h)/4 \rfloor$ such guards suffice (O'Rourke 1987). We improve this bound to $\lfloor (n + \frac{5}{3}h)/4 \rfloor$.

[12] **Fall coloring of Graphs**, (with C. Mitillos), preprint.

Fall coloring of graphs is a common extension of two fundamental NP-complete problems - Graph Coloring and Graph Domination. It asks for a partition of the vertices of a graph into sets which are both independent and dominating. The primary question in this topic is the feasibility of such a coloring existing for a given graph. Let $Fall(G)$ be the set of values k such that G is

k -Fall colorable (k is the number of colors allowed). We construct of a family of graphs with pre-determined $Fall(G)$ which shows that $Fall(G)$ can be arbitrarily large with arbitrarily large gaps. We characterize Threshold graphs and Split graphs that are Fall colorable. We also show that the Cartesian product of a k -Fall-colorable graph with a k -colorable graph gives a k -colorable graph.

[13] Counting Local Substructures in Turan-type Problems, (with D. Mubayi), manuscript in preparation.

The fundamental question in extremal combinatorics is to find $ex(n; H)$ the maximum number of edges in an n -vertex graph that does not contain a copy of a given fixed graph H . We are interested in extending this study by counting the number of intersecting copies of H when the number of edges exceeds $ex(n; H)$. In particular, we prove results of the form: If $e(G) \leq ex(n; H) + 1$ then $\exists v \in V(G)$ such v belongs to at least $d(n; H)$ copies of H . We use a new technique utilizing the Hypergraph Removal lemma and the Stability theorem to bound $d_G(n; H)$. For example, we show $d(n; K_3) \geq (1/2 - o(1))n$. We also show similar bounds for $H = K_t$, or more generally H is any fixed edge-critical graph ($\exists e \in E(H)$ such that $\chi(H - e) < \chi(H)$), and generalize this work to hypergraphs like the Fano plane.

[14] Exact Algorithms for the Maximum Independent Set Problem, (with E.C. Sewell and S.H. Jacobson), in preparation.

NP-hard optimization problems cannot be solved in polynomial time unless $P = NP$. To overcome this difficulty, a number of approaches including approximation algorithms and heuristics have been explored. But neither of these approaches guarantee exact solutions. In contrast to these approaches, super-polynomial time algorithms that solve NP-hard problems to optimality have been designed and analyzed. These exact algorithms lead to practical algorithms for moderate instance sizes, in addition to being used as sub-routines in hybrid algorithms with heuristics.

The maximum independent set problem (or, equivalently the maximum clique problem or the minimum vertex cover problem) is a fundamental NP-hard graph optimization problem. The running times of its exact algorithms has improved from $O^*(1.2599^n)$ (Tarjan–Trojanowski, 1977) to $O^*(1.1844^n)$ (Robson, 2001). In this paper, a new and faster algorithm, based partly on reductions from the previous paper [8] of the authors, is proposed.

[15] Lectures on Modern Probabilistic Methods for Discrete Mathematics, (with D.B. West), in preparation.

These lecture notes based on courses offered by me at UIUC and IIT are being prepared for possible publication. These notes cover three areas and their applications that are not well covered in typical graduate courses in Discrete Mathematics: Concentration of Measure, Applications of Shannon Entropy, and Markov Chain Monte Carlo. The lectures present elementary proofs of basic results in these areas with the aim of making these topics accessible to uninitiated students and researchers in CS, ECE, OR, etc., in addition to those in mathematics. The focus is on developing the themes underlying the various methods and illustrating the final results through applications in graph theory, combinatorial optimization and theoretical computer science. Most of the discussions and results would appear for the first time in a book.

[16] A Threshold for Random Geometric Graphs with a Hamiltonian cycle, (with J. Balogh), manuscript.

Naturally evolving massive networks like the social networks of acquaintances (used to model

the spread of diseases, rumors, etc.), the internet, the power grid, airline traffic, etc., share common characteristics like short average distance and high clustering coefficient (average density of sub-graphs induced by neighborhoods). *random geometric graph* (RGG) has been studied as a model for such networks. It is constructed by placing n points randomly (according to uniform distribution or a poisson process) within a metric space, and putting an edge between any pair of points at most distance $r = r(n)$ apart. In addition to its wide range of applications, this model is interesting as it brings together ideas from geometry, probability, and graph theory. As in classical random graphs, the fundamental questions concern the threshold functions for various graph structures. We study RGG on a bounded set with l_∞ -metric and show that the threshold for 2-connectedness is about the same as the threshold for Hamiltonicity for these graphs, i.e., there exists a Hamiltonian cycle as soon as the graph becomes 2-connected.

[17] Multi-Objective Optimization and Sensitivity Analysis For Discrete and Continuous Optimization Problems, (with S.H. Jacobson, G.K. Kao, J.A. Stori and V. Venkat), Technical Report to Austral Engineering and Software, Inc., University of Illinois, Urbana, 2004.

This technical report focusses on providing tools for solving and analyzing various aspects of optimization problems with multiple objective functions. Multi-objective optimization problems frequently arise in real-life applications, but have not been studied in great detail in literature. My contributions to this report include surveying the previous research and proposing algorithms for sensitivity and post-optimality analysis of discrete multi-objective optimization problems as well as for solving mixed discrete-continuous multi-objective optimization problems.

[18] On Queuenumber of Planar Graphs, M.Sc. Thesis (under the guidance of S. Pemmaraju), Dept. of Mathematics, Indian Institute of Technology, Bombay, 1999.

The queuenumber of a graph is the minimum number of queues (FIFO) needed to process its edges (without nesting) over a total order of its vertices. Queue layouts of graphs are dual to stack (LIFO) layouts (better known as book embeddings and the corresponding pagenumber of a graph). Queue layouts have been applied to sorting permutations, parallel process scheduling, matrix computations, and graph drawing.

This thesis studies a long standing open problem in queue layouts of graphs - whether queue number of planar graphs is bounded or not. The queue number of stellation of K_3 is explored through a number of general labelings, using tools like Erdos-Szekeres theorem on long monotone subsequences in arbitrary sequences to analyze and show that these families of labelings give unbounded queue number. (The queuenumber of planar graphs remains unknown to date, 2009.)

[19] An Elementary Construction of \mathbb{R} from \mathbb{Q} , The Mathematica : Annual Journal of the Mathematics Society, St. Stephen's College, Delhi, 1997.

This paper proposes an elementary construction of \mathbb{R} as a completion of \mathbb{Q} . The method taught in high school for approximating the value of an irrational number by upper and lower bounding sequences of rational numbers (essentially the decimal representation) is an illustration of the nested intervals property of reals. Any family of nested intervals of rationals with diameter tending to zero can be said to define a real number. Just as rational Cauchy sequences need not converge in \mathbb{Q} , nested intervals in rationals need not have a non-empty intersection in \mathbb{Q} . This leads to \mathbb{R} being defined as the collection of families of nested intervals in \mathbb{Q} with diameter tending to zero. This construction is shown to follow the underlying theme of the Cantor's construction using Cauchy sequences and Dedekind's construction using Dedekind's cuts, and can be used as a motivation for these constructions.