Acknowledgements

Organising a workshop (however small) is a job that involves a lot of organisation. This workshop is no exception. There have been many people involved and we would like to take this opportunity to thank them.

We owe great thanks to Lindsey Gaunt and Kerrie Walker for their invaluable help in organising accommodation, general conference facilities and for keeping this workshop within budget!

Many thanks to Benjamin Sach for keeping the BAD website up to date, and for acting so swiftly when we asked for changes to be made.

We would like to thank all our speakers for providing their abstracts so promptly, and finally we would like to thank all the delegates attending BAD'08 who will make this event interesting, fun and hopefully will help to bridge some of the gaps between Computer Science, Engineering and Mathematics.

Raphaël Clifford and Paul Sant
(BAD'08 organisers)

Bristol Algorithms Days 2008

Firstly, we would like to welcome you to the 2nd Annual Bristol Algorithms Days workshop. The first workshop (which took place from 19-20th February 2007) was such a success that we decided to come back bigger and BADer (please excuse the pun!).

On a more serious note, the BAD event, and BAD'08 in particular aims to bring together algorithms researchers, mathematicians and engineers from within (and outside) the UK to further research at the boundaries of these disciplines.

During this workshop we hope that you will find time to meet others, discuss some open problems and to foster long term collaborations. We hope that by doing this not only will we be able to undertake important research, but we will hopefully also continue to raise the profile of algorithmic research.

This year we have more talks than ever (10 invited speakers and 10 contributed talks) as well as the introduction of a poster session for PhD students and postdoctoral researchers. We hope that the wide range of topics provides you with an exciting programme and that we can use this "informal" forum to push forward research at the boundaries of Computer Science, Engineering and Mathematics.

Finally, we hope that you enjoy BAD'08 and that we will see you again next year!

Raphaël Clifford and Paul Sant
(BAD '08 organisers)

The Bridging the Gaps Project

Bridging the Gaps is an EPSRC-funded initiative that involves the whole Faculty of Engineering and the Department of Mathematics in the Science Faculty at the University of Bristol. £400k has been awarded to Bristol in a competitive call to set up an internal environment that nurtures interdisciplinary working, with the aim of developing new collaborative research programmes at the interface between mathematical sciences, information and computing technology, and engineering. Under the grant, which will run for three years from 2007, a small team comprising Alan Champneys (Eng Maths), Raphaël Clifford (CS), Yves Tournigny (Maths) and Chris Allen (Aero) supported by project management, solicits proposals from Bristol investigators for seminar programmes, away days, workshops, retreats or any imaginative event aimed at bridging the gaps between disciplines. Four broad themes have been identified, being A. Uncertain Systems, B. Pure Mathematics and Algorithms, C. Media and Materials and D. Computational Science, although any genuinely interdisciplinary collaboration that bridges the gaps will be considered. Involvement of researchers from outside the University is a key part of the programme, and success will be judged by the number of new research outputs (papers, grants, research networks) that result.

Bristol Algorithms Days 2008 is funded and supported by the Bristol Bridging the Gaps project.

For more information, visit http://bridgingthegaps.bristol.ac.uk/

Programme

Monday 3rd March 2008
Stothert Room, Engineers' House

8.30-8.45	Coffee and registration	
Session I		
8.45-9.00	Welcome and introduction	Raphaël Clifford
9.00-9.40	Testing expansion in bounded degree graphs	Artur Czumaj
9.40-10.05	PRIMES is in P	Richard Pinch
10.05-10.30	Peer counting and sampling in overlay networks	Ayalvadi Ganesh
10.30-11.00	Coffee break	
11.00-11.40	Low degree vertices and minimal mess: Large independent sets in random regular graphs	Michele Zito
11.40-12.05	Noise propagation in diffusion tensor imaging	Ida Pu
12.05-1.30	Lunch	
Session II		
1.30-2.10	Matching under preferences: Results old and new	Rob Irving
2.10-2.35	Quantified constraints and containment problems	Barnaby Martin
2.35-3.00	Approximate string matching in subquadratic time	Alexander Tiskin
3.00-3.30	Coffee break	
3.30-4.10	Efficient graph exploration by autonomous agents	Leszek Gasieniec
4.10-4.35	Recursive quantum search and finding the intersection of two sorted lists	Ashley Montanaro
4.35-5.00	On the core and f-nucleolus of flow games	Daniel Paulusma
5.00-5.25	Algorithmic problem solving	Roland Backhouse
Session III		
5.30-6.30	Poster session	

Workshop dinner
The workshop dinner will take place at Goldbrick House, 69 Park Street, Bristol, BS1 5PB at 19.15.

Tuesday 4th March 2008
Stothert Room, Engineers' House

8.30-9.00	Coffee	
Session IV		
9.00-9.40	Overhang bounds	Mike Paterson
9.40-10.20	Storage and retrieval of individual genomes	Veli Makinen
10.20-10.45	A sublinear-time approximation scheme for bin packing	Tugkan Batu
10.45-11.15	Coffee	
11.15-11.55	Multiple random walks in random regular graphs	Colin Cooper
11.55-12.20	Approximating general metric distances between a pattern and a text	Ely Porat
12.20-1.45	Lunch	
Session V		
1.45-2.25	Tolerating faults in parallel computers	Iain Stewart
2.25-3.05	On the stability of dynamic diffusion load balancing	Russell Martin
3.05-3.30	Learning coordinate gradients with multi-task kernels	Yiming Ying
Session VI		
3.30-4.30	Breakout session	
Close		

Testing expansion in bounded degree graphs

Artur Czumaj

University of Warwick

We consider the problem of testing expansion in bounded degree graphs. We focus on the notion of vertex-expansion: an a-expander is a graph $G = (V,E)$ in which every subset U of V of at most $|V|/2$ vertices has a neighborhood of size at least $a|U|$. Our main result is that one can distinguish good expanders from graphs that are far from being weak expanders in time approximately $O(n^{1/2})$. We prove that the property testing algorithm proposed by Goldreich and Ron (2000) with appropriately set parameters accepts every a-expander with probability at least 2/3 and rejects every graph that is ε-far from an a*-expander with probability at least 2/3, where $a^* = O(a^2/(d^2 \log(n/\varepsilon))$, d is the maximum degree of the graphs, and a graph is called ε-far from an a^*-expander if one has to modify (add or delete) at least epsilon dn of its edges to obtain an a^*-expander. The algorithm assumes the bounded-degree graphs model with adjacency list graph representation and its running time is $O(d^2 n^{1/2} \log(n/\varepsilon)/(a^2 \varepsilon^3))$.

Peer counting and sampling in overlay networks

Ayalvadi Ganesh

University of Bristol

We describe a random-walk based technique for sampling a node uniformly at random from a connected graph. We show how it can be combined with the birthday paradox to estimate the number of nodes in the graph. Finally, we describe statistical properties of this estimator and show that it is asymptotically efficient.

(Joint work with Anne-Marie Kermarrec, Laurent Massoulie and Erwan le Merrer.)

Low degree vertices and minimal mess: Large independent sets in random regular graphs

Michele Zito

University of Liverpool

Given a graph $G = (V,E)$, an independent set H is a subset of V which spans no edge. We are interested in finding (by algorithmic means) independent sets of the largest possible cardinality. Let $\alpha(G)$ be the size of the largest independent sets in G divided by $n=|V(G)|$. The problem has a long history. It is one of the first optimization problems whose decision version was shown to be NP-complete. Since then many results have appeared either proving that an optimal structure can be found in polynomial time on special graph classes or showing that particular polynomial-time heuristics return solutions that are not too small for particular classes of graphs or else proving that finding heuristics returning solutions significantly close to the optimal ones is at least as hard as solving the optimization problem exactly.

The maximum independent set problem has been studied thoroughly in several of such random structures. For graphs generated according to the well-known G(n,p) model it has been proven that as long as pn tends to infinity alpha(G(n,p)) n tends to 2 log(np) / log(1/(1-p)) asymptotically almost surely (a.a.s.). For random d-regular graphs, when d is a fixed constant (say 3 or 5 or even 100), the situation is less satisfactory. The best known bounds are reported in the second and third column of Table below for d up to 7. It is quite interesting that for the past 12 years, nobody has been able to improve these bounds (in fact the upper bounds are even older than that).

A.a.s. bounds on α(G) for random d-regular graphs.

d	*l.b*	*u.b*	*α(d)*
3	0.4328	0.4554	0.4348
4	0.3901	0.4163	0.3921
5	0.3566	0.3844	0.3593
6	0.3296	0.3580	0.3330
7	0.3701	0.3357	0.3106

We propose a couple of methods that lead to improvements on the best known lower bounds for all values of d (see fourth column in the Table).

Nick Wormald showed that a simple process (termed "neighbourly" algorithm) that repeatedly picks vertices of minimum positive degree, adds them to the independent set that is being built and then removes all edges at distance at most one from the chosen vertex, builds fairly large independent sets a.a.s. if G is a random d-regular graph. It turns out that, in some cases, it is more convenient to add to the independent set one of the neighbours v of the initially chosen vertex u, rather than u itself. More precisely, v should be chosen if this is guaranteed to create a number of low degree vertices ("sparsification" principle) or if it leads to the removal of very few edges ("minimal mess" principle).

Our solution is simple to analyse: the proof of our results relies on a standard application of the differential equation method. Furthermore, it seems plausible that similar principles may lead to improvements for other optimization problems.

Joint work with William Duckworth (ANU, Canberra)

Noise propagation in diffusion tensor imaging

Ida Pu

Goldsmiths, University of London

Diffusion tensor imaging (DTI) is a useful technique in Magnetic Resonance Imaging to identify tissue macro structures and detect their alteration by various diseases. However, DTI suffers from errors in tensor data due to the low signal-to-noise ratio (SNR) of acquired diffusion-weighted (DW) images. In this talk, we will introduce an analytical algorithm to determine the noise induced errors in DTI. A complete computational chain is established, based on error propagation theory, for the noise propagated from the acquired raw data to various quantities derived from the tensor parameters. We will also demonstrate, with this approach, how the errors in tensor can be predicted in practice.

Matching under preferences: Results old and new

Rob Irving

University of Glasgow

Many algorithmic problems, of both theoretical and practical interest, arise when we wish to form a matching based on the expressed preferences of those involved in the process. Examples of this include the stable marriage and stable roommates problems, the hospitals/residents or college admissions problem, the house allocation problem, and a variety of others.

The history of such problems goes back over half a century, and yet challenging new variants continue to emerge, often from the application domains. This talk presents a survey covering a range of such problems and results, both old and new.

Approximate string matching in subquadratic time

Alexander Tiskin

University of Warwick

Given a text of size n and a pattern of size m, the classical approximate string matching problem asks for all substrings of the text that are within a given edit distance of the pattern. Algorithms for this problem include the standard $O(mn)$ algorithm by Sellers (1980), as well as a $O(mn/\log(n))$ unit-cost RAM algorithm by Wu, Manber and Myers (1986). By extending our method of semi-local string comparison to edit distances, we obtain the first subquadratic approximate string matching algorithm that does not rely on the unit-cost RAM assumption. Our algorithm also solves the more general problem of edit distance matching in time $O(mn\ \mathrm{loglog}(n)/\log(n))$, answering an open problem by Cormode and Muthukrishnan.

Efficient graph exploration by autonomous agents

Leszek A Gasieniec

University of Liverpool

Efficient search in unknown or unmapped environments is one of the fundamental problems in algorithmics. Its applications range from agent navigation in hazardous environment to rigorous exploration (and, e.g., indexing) of data available on the Internet. A large volume of search algorithms has been proposed under various assumptions about the ability of the mobile entity as well as the properties of the explored environment. The environment is very often modeled as two or higher dimensional space ("geometric model"), where the mobile agent can freely traverse (subject to some obstacles); or as a directed or undirected graph ("graph model") where the environment is represented by a finite or infinite graph supported by discrete moves permitted only along its edges.

The design of efficient exploration algorithms in the graph model has been extensively studied under a variety of assumptions, e.g., directed vs undirected graphs, anonymous nodes vs nodes with distinct identities, deterministic vs probabilistic solutions, as well as in the context of different goal functions in mind including optimal time complexity, memory consumption, or use of other computational resources.

We provide a condense summary of core problems in the field complemented by efficient algorithmic solutions. This is followed by presentation of more recent developments and new directions paved in the field.

Recursive quantum search and finding the intersection of two sorted lists

Ashley Montanaro

University of Bristol

Grover's famous quantum search algorithm can find a "marked" element in an unsorted n-element list using only $O(\sqrt{n})$ queries to the list. In this talk, I discuss an extension of Grover's algorithm that allows faster quantum search of databases that have a recursive structure. This is applied to give a quantum algorithm that finds the intersection of two n-element sorted lists in time $O(\sqrt{n})$, which is optimal.

On the core and f-nucleolus of flow games

Daniel Paulusma

University of Durham

A cooperative game is given by a set E of players, and a function v that assigns a positive value $v(A)$ to each subset A of E with $v(\text{empty set}) = 0$. We call a subset A of E a coalition, and we refer to $v(A)$ as the value of coalition A. This value can be interpreted as the gain that the members of A achieve by "cooperating" with each other. The value $v(E)$ is also called the total value of the game.

A flow game (E,v) is defined by a directed graph $G = (V,E)$ with a source $s \in V$, a sink $t \in V$ and positive edge capacities c. Players are represented by arcs, and the value $v(A)$ of each coalition A is equal to the maximum flow in (V,A).

We study the computational complexity of the problem of computing a "fair" allocation of the total value $v(E)$ according to various solution concepts such as the core, the nucleolus and the nucleon.

Firstly, we study simple flow games (in which all edge capacities are equal to one). Using the ellipsoid method,both Deng, Fang, and Sun, and Potters, Reijnierse, and Biswas show that the nucleolus of simple flow games can be computed in polynomial time. Our main result is a new, combinatorial method that is based on eliminating redundant s-t path constraints such that only a polynomial number of these constraints remain. This leads to efficient algorithms for computing the core, nucleolus and nucleon of simple flow games.

Deng, Fang, and Sun also prove that computing the nucleolus for (general) flow games is NP-hard. Our second result generalizes this result. We prove that computing the *f*-nucleolus of flow games is NP-hard for a large class of priority functions f (including the priority functions for the nucleolus, nucleon and per-capita nucleolus).

Algorithmic Problem Solving

Roland Backhouse

University of Nottingham

"Algorithmic Problem Solving" is the title of a 1st-year. 1st-semester module that has been running for six years. Initially an optional module, it grew in popularity and was made compulsory for 1st-year Computer Science students in 2005. It currently has an intake of approximately 150 students per year (for some of whom it remains an optional module).

The module introduces algorithm-design techniques via a collection of carefully selected problems. The problems have been chosen not to be "mathematical" in nature ---they are typically problems that are easily comprehended from short prose statements---- but the goal is to introduce the relevance of (algorithmic) mathematics and mathematical calculation.

This talk will give concrete examples of the problems and how the module is structured so as to introduce algorithm-design principles step-by-step (at first-year level).

Time permitting, the talk will end with a short sketch of a generalisation of the well-known "flashlight problem" (also known as the "U2 problem", the "bridge problem" or the "torch problem"). The original problem involves getting 4 people across a bridge in the shortest time under certain constraints, and is one that I use in the module. The generalised problem involves getting N people across a bridge with capacity C; it seems to be very challenging (at least, I found it so). I have a solution of complexity quadratic in the number of people but I suspect that a linear solution can be constructed. (A linear solution is known when the capacity is 2; no polynomial-time solution has been published so far as I am aware.) I would like to pose the problem as a challenge to the audience. (Needless to say, the generalised problem is not in the syllabus of the module!)

Overhang Bounds

Mike Paterson

University of Warwick

How far can a stack of n identical blocks be made to hang over the edge of a table? The question dates back to at least the middle of the 19th century and the answer to it was widely believed to be of order log n. Recently, we (Paterson and Zwick) constructed n-block stacks with overhangs of order $n^{1/3}$, exponentially better than previously thought possible. The latest news is that we (Paterson, Peres, Thorup, Winkler and Zwick) can show that order $n^{1/3}$ is best possible, resolving the long-standing overhang problem up to a constant factor.

I shall review the construction and describe the upper bound proof, which illustrates how methods founded in algorithmic complexity can be applied to a discrete optimization problem that has puzzled some mathematicians and physicists for more than 150 years.

(Joint work with Yuval Peres, Mikkel Thorup, Peter Winkler and Uri Zwick.)

Storage and retrieval of individual genomes

Veli Mäkinen

University of Helsinki

In near future, biomolecular engineering techniques will reach a state where the sequencing of individual genomes becomes feasible. This progress will create huge expectations for the data analysis domain to reveal new knowledge on the "secrets of life". Quite rudimentary reasons may inhibit such breakthroughs; it may not be feasible to store all the data in a form that would enable anything but most basic data analysis routines to be executed. This paper is devoted into studying ways to store massive sets of complete individual genomes in space-efficient manner so that retrieval of the content as well as queries on the content of the sequences can be provided time-efficiently. We show that although the state-of-the-art full-text self-indexes do not yet provide satisfactory space bounds for this specific task, after carefully engineering those structures it is possible to achieve very attractive results; the new structures are fully able to exploit the fact that the individual genomes are highly similar. We confirm the theoretical foundings by experiments.

Joint work with Jouni Siren and Niko Välimäki.

A sublinear-time approximation scheme for bin packing

Tugkan Batu

London School of Economics

The bin packing problem is defined as follows: given a set of n items with sizes $0<w_1,w_2,\ldots, w_n \leq 1$, find a packing of these items into minimum number of unit-size bins possible.

We present a sublinear-time asymptotic approximation scheme for the bin packing problem; that is, for any $\varepsilon>0$, we present an algorithm A_ε that has sampling access to the input instance and outputs a value k such that $opt \leq k \leq (1+\varepsilon)\, opt+1$, where opt is the cost of an optimal solution. It is clear that uniform sampling by itself will not allow a sublinear-time algorithm in this setting; a small number of items might constitute most of the total weight and uniform samples will not hit them. In this work we use weighted samples, where item i is sampled with probability proportional to its weight: that is, with probability $w_i/\Sigma_i w_i$. In the presence of weighted samples, the approximation algorithm runs in $\tilde{O}(\sqrt{n} \cdot \text{poly}(1/\varepsilon)) + g(1/\varepsilon)$ time, where $g(x)$ is an exponential function of x. When both weighted and uniform sampling are allowed, $\tilde{O}(n^{1/3} \text{poly}(1/\varepsilon)) + g(1/\varepsilon)$ time suffices. In addition to an approximate value to opt, our algorithm can also output a constant-size ``template'' of a packing that can later be used to find a near-optimal packing in linear time.

Joint work with Petra Berenbrink and Christian Sohler.

Multiple random walks in random regular graphs

Colin Cooper

King's College London

It was shown by Cooper and Frieze, for that $r \geq 3$ the cover time C_G of a random r-regular graph G_r is asymptotic to $\theta_r\, n \ln(n)$, where $\theta_r = (r-1)/(r-2)$. This result holds *whp* (with probability tending to one as the number of vertices n tends to infinity).

In this talk we study problems arising from multiple random walks on random regular graphs, and establish the following (*whp*) results.

(i) The time for k independent walks to cover G_r is asymptotic to C_G/k.

(ii) For most starting positions, the expected number of steps before any of the walks meet is $\theta_r n / \binom{k}{2}$.

(iii) If the walks can communicate on meeting at a vertex, we show that (for most starting positions) the expected time for k walks to broadcast a single piece of information is asymptotic to $((2 \ln(k))/k)\, \theta_r\, n$, as k,n tend to infinity.

(iv) We also consider properties of walks where particles interact when they meet at a vertex by coalescing, or by exploding and destroying each other. As an example, the expected extinction time of k explosive particles (k even) tends to $(2 \ln(2))\, \theta_r\, n$ as k tends to infinity.

Approximating general metric distances between a pattern and a text

Ely Porat

Bar-Ilan University

Let $T=t_0 \ldots t_{n-1}$ be a text and $P = p_0 \ldots p_{m-1}$ be a pattern taken from some finite alphabet set Σ, and let *dist* be a metric on Σ. We consider the problem of calculating the sum of distances between the symbols of P, and the symbols of substrings of T of length m for all possible offsets. We present an ε-approximation algorithm for this problem which runs in time $O(1/\varepsilon^2\ n$ polylog$(n,|\Sigma|))$. This algorithm is based on a low distortion embedding of metric spaces into normed spaces (in particular, into ℓ_∞) which is done as a preprocessing stage, and also on a technique of sampling.

Tolerating faults in parallel computers

Iain Stewart

University of Durham

The choice of interconnection network in the construction of a distributed-memory multiprocessor is crucial to the performance of the machine. It is preferable that such interconnection networks: come in parameterized families and have a recursive structure (to aid scalability and programmability); are symmetric (to aid analysis and programmability); have a relatively small diameter (to improve message latency); are highly-connected (to aid fault-tolerance and data transfer); contain embeddings of other such networks (to aid simulation); and so on. Of increasing importance is the capacity of such multiprocessors to tolerate faulty processors and faulty links. In this talk, I shall overview some popular interconnection topologies and their ability to tolerate faults in the context of Hamiltonian circuits and Hamiltonian paths.

On the stability of dynamic diffusion load balancing

Russell Martin

University of Liverpool

We consider the problem of dynamic load balancing in arbitrary (connected) networks on n nodes. Our load generation model is such that during each round, n tasks are generated on arbitrary nodes, and then (possibly after some balancing) one task is deleted from every non-empty node. Notice that this model *fully saturates* the resources of the network in the sense that we generate just as many new tasks per round as the network is able to delete. We show that even in this situation the system is *stable*, in that the total load remains bounded (as a function of n alone) over time. Our proof only requires that the underlying ``communication'' graph be connected. (It of course also works if we generate less than n new tasks per round, but the major contribution of this paper is the fully saturated case.) We further show that the upper bound we obtain is asymptotically tight (up to a moderate multiplicative constant) by demonstrating a corresponding lower bound on the system load for the particular example of a linear array (or path). We also show some simple negative results (i.e., instability) for work-stealing based diffusion-type algorithms in this setting.

(Joint work with Petra Berenbrink and Tom Friedetzky)

Learning coordinate gradients with multi-task kernels

Yiming Ying

University of Bristol

In this talk we propose a multi-task kernel-based formulation for coordinate gradient learning (MGL). In particular, we show that choosing the multi-task kernel appropriately captures the inherent structure of gradient learning. A representer theorem is also provided for his algorithm and appropriate computational optimization algorithms are discussed. Specifically, we show that learning the gradient is equivalent to a learning task in a low dimensional projection space spanned by the input data. Finally, we report some preliminary experiments on simulated data to validate MGL for variable selection as well as determining variable covariation.

(Joint work with Colin Campbell)