D I M A C S

Center for Discrete Mathematics and Theoretical Computer Science

2002-2003 Annual Report

November 2003

DIMACS Center • Rutgers University • 96 Frelinghuysen Road • Piscataway, NJ 08854-8018 • USA
TEL: 732-445-5928 • FAX: 732-445-5932 • EMAIL: `center@dimacs.rutgers.edu`
`http://dimacs.rutgers.edu/`
DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs - Research, Bell Labs, NEC Laboratories America, and Telcordia Technologies with affliates Avaya Labs, HP Labs, IBM Research, and Microsoft Research

Contents

1 Executive Summary

In this report, we summarize the highlights of the accomplishments at DIMACS in the period July 1, 2002 to June 30, 2003, i.e., FY2003.

A Brief History of DIMACS

DIMACS was founded as a National Science Foundation "science and technology center" (STC) through an 11-year award received in 1989. This highly prestigious award – one of only a handful in the country in the mathematical sciences – led us to establish a wide variety of research and education programs that have been widely praised and copied in many parts of the world. The center was founded as a partnership of Rutgers and Princeton Universities (with Rutgers being the lead institution), AT&T Bell Labs, and Bellcore. Other industrial partners have been added as the center has evolved and some 230 researchers, mostly from its partner institutions, are involved with DIMACS as "permanent members" of the center.

DIMACS has an international scope and its programs involve not just the permanent members but also numerous participants from outside the local institutions. In a typical year, the center hosts 30 to 35 research workshops with 1,500 attendees, 100 visiting scientists, 5 postdoctoral fellows, 50 2- and 4-year college faculty, 100 precollege teachers, and 100 high school students, as well as graduate and undergraduate students.

From the beginning, research at DIMACS was unique because the "themes" changed regularly and we developed our signature research activity, the "special year." As it grew, the center became a national and international resource that aimed to catalyze interdisciplinary collaborations and play a leadership role in the rapid development of new directions and subfields of science. Our early special years certainly accomplished this. For instance, DIMACS' 1990-91 Special Year on Complexity Theory of Interactive Computing played a central role in one of the most important developments in theoretical computer science in recent years, the development of ways to reformulate computer "proofs" so that the correctness of a computation can be verified with arbitrarily high confidence by a small number of random spot checks. A critical result leading to these developments, the discovery of connections between interactive proofs and approximation algorithms, was a direct result of the special year. It arose from the interactions between two diverse communities, the cryptography community and the combinatorial optimization community, and came about by applying recent methods of cryptography to the maximum clique problem of combinatorial optimization.

The early special year themes were closely related to discrete mathematics (dm) and theoretical computer science (tcs): Discrete and Computational Geometry; Graph Theory and Algorithms; Combinatorial Optimization. Later programs branched out from dm and tcs, to include programs partnering mathematical scientists with biological scientists, with epidemiologists, with industrial and mechanical engineers, with chemists, with physicists, and with social scientists, to name a few.

As DIMACS evolved, instead of changing our research themes annually as we did at the beginning, we evolved to longer and longer themed programs, "special focus programs," some still going on. Also, over the years, the special focus programs became more complex and reached out more and more to other disciplines. Our "special year" on "Mathematical Support for Molecular Biology" (MSMB), begun in 1994, ended up lasting six years, and was immediately followed by our present special focus on "Computational Molecular Biology." MSMB introduced a large number of computer scientists and mathematicians to the field of computational molecular biology and created important partnerships between mathematical and biological scientists that were groundbreaking and important. These partnerships had remarkable impact. For example, the ability of the t-haplotype, an ultra-selfish gene in mice, to persist at low levels in natural populations has been a subject of debate in the theoretical biology literature for a long time. Ultra-selfish genes are transmitted to offspring more often than would be consistent with Mendel's laws, yet they are harmful to the organism that carries them. A long-standing hypothesis attributed the low level persistence of such ultra-selfish genes to a balance between the loss of the gene due to stochastic effects in small populations and the re-introduction of the gene through migration. In a dramatic collaboration, a DIMACS computer scientist (who was a participant in the MSMB postdoctoral program) and a Princeton molecular biologist, who met through the MSMB tutorial, overturned this 40-year-old hypothesis.

Physicists played a central role in our 1996-98 Special Focus on Discrete Probability, which was also one of the first joint programs we ran with our long-term partner, the Institute for Advanced Study in Princeton. The partnership between computer scientists and physicists has had some important impacts and has continued well beyond that special focus. Some very exciting science came out of this special focus. The general belief in statistical physics is that spontaneous symmetry-breaking is something which, if it occurs at all, occurs when the temperature drops below some unique critical point. This had indeed been proved for the Ising model. During the special focus, a partnership among discrete mathematicians, computer scientists, and statistical physicists led to the solution to the long-open question of whether the same thing holds for the Widom- Rowlinson model, which is to the Ising exactly as vertices are to edges of a graph. Two visitors and a DIMACS member settled this in the negative by constructing a non-regular tree in which, as temperature is lowered, symmetry is broken, then re-established, then broken again. The partnerships between physicists and computer scientists developed during this special focus have played a critical role in the development of a new field of science that can be described as the study of "random discrete systems." Today, this field has matured beyond expectations and is advancing at a tremendous rate.

As time moved on, we also evolved to having more than one themed program each year, up to the point where in FY03 we had eight such programs running simultaneously. The variety of themes and partnerships in these programs has grown tremendously, including, for example, new thrusts in communication security and information privacy, epidemiology, and data mining.

Still, it is important for us to "stick to our roots" and there are quite a few programs at DIMACS devoted to issues in "pure" discrete mathematics and fundamental theoretical computer science. These programs have, over the years, led to significant scientific results. The four color problem, stemming from map coloring, has been one of the motivating problems for research in discrete mathematics, in particular in graph coloring, for over 100 years. Graph coloring has modern applications in areas as diverse as scheduling, traffic phasing, fleet maintenance, frequency assignment, and task assignment, and generalizations of graph coloring form the basis for a joint NSF-supported DIMACS project with international partners in Hungary and the Czech Republic. The original proof that every graph arising from a map could be colored in four or fewer colors was very long, complicated, and tedious and some of the details were not very clear, leading to a prolonged debate about the correctness of the proof. A new proof of this result, obtained by a DIMACS member and two visitors, is simpler and more accessible to complete checking. Moreover, it has been modified to obtain other exciting results about graph coloring.

As DIMACS matured, we started several multi-disciplinary, multi-institutional research projects. Our first two such projects resulted from the 1993-94 Special Year on Massively Parallel Computing. These large projects, supported by DARPA and NSF, involved researchers from DIMACS, WINLAB, AT&T, Bellcore, Georgia Tech, University of Massachusetts, William and Mary, Dartmouth, and two small companies, Ultra Corporation and Cooperating Systems Corporation. The projects, which ended in 2001, developed the first model and simulation of the global Internet with a hundred thousand hosts and routers, running IP packets with real routing protocols at a rate of more than one million packet hops per second. The resulting software is now used by over 40 companies and university research teams. Several of our present multi-disciplinary, multi-institutional projects are closely connected to issues of homeland security.

Industrial partnerships are vitally important to DIMACS. Researchers from our industrial partners play a central role in the life of the center, participating in research projects, organizing special foci and workshops, and participating in educational programs. An early example of research partnerships between university and industrial scientists came during our 1991-92 Special Year on Graph Theory and Algorithms. During that special year, a group of researchers from four different institutions, AT&T Labs, Bellcore, Rice University, and Rutgers, collaborated to solve what at that time was the "world's largest" unsolved Traveling Salesman Problem. The TSP problem they solved arose from the question of how to most efficiently connect 3,038 points for drilling holes in making a circuit board. Among the many ways in which we partner with our industrial colleagues is in shared postdocs. During the 1996-2000 Special Focus on Networks, a researcher from NEC Laboratories and a DIMACS postdoc developed an escrowed identity system to protect privacy in electronic toll collection and related applications. In the first tier of this 2-tiered system, enough information is provided by the user so that an operator can be sure the user is legitimate, but otherwise the user's identity is protected. If more information is required, for example in the case of a crime, the second tier provides it, through involvement of a third party (an escrow agent).

As DIMACS has evolved, we have also expanded our international collaborations. We now run joint activities with ten international partners, including funded research on topics in combinatorics and the connection between computer science and decision making.

From the beginning, we saw our research mission as fundamentally intertwined with our educational mission, and developed programs that integrated research and education from the postdoctoral level to the elementary school level. Our K-12 teacher programs, heavily informed by research at DIMACS on discrete mathematics, quickly received national acclaim and have had major impact on the development of curriculum "standards" in New Jersey and nationally. Literally tens of thousands of teachers nationwide have been impacted by our programs. Our "Reconnect" program has impacted many teaching faculty at 2- and 4-year colleges around the country and our programs for undergraduate research have included a pioneering international "research experiences for undergraduates" program, later mirrored in a unique "research experiences for teachers" program in the DIMACS Connect Institute.

The end of the STC award and its large and stable source of funding required us to raise money for our many projects individually, while challenging us to find ways to support center infrastructure. We now write a large number of grant proposals each year, going to numerous agencies and private sources and numerous programs within those agencies. NSF remains our primary sponsor, but even within NSF, we have obtained funds from numerous programs, including Computer Science; Mathematical Sciences; Social, Behavioral and Economic Sciences; Biological Sciences; Education and Human Resources; and International Programs. Throughout the lifetime of the center, our academic and industrial partners have been strong supporters, especially Rutgers, which has played the central leadership role in DIMACS. Our industrial partners continue to be strong supporters of the center and play a crucial role in the projects and programs of DIMACS, partly through financial support, but mostly through heavy involvement of leading researchers in our center.

Special Focus Activities

The primary scientific activity at DIMACS remains the "special focus," a focused and coordinated collection of activities that span a period of time. The focus concentrates on an area of discrete mathematics/theoretical computer science (broadly speaking) and its applications that is ripe for rapid development, contains problems appropriate for high quality research mathematicians and computer scientists, and, in most cases, has strong potential for connections to other areas of the sciences, the mathematical sciences, and/or industry. This signature way of doing research at DIMACS revolves around a series of tutorials, interdisciplinary research working groups, and workshops (more formal conferences). There are many visitors and we involve postdoctoral fellows, graduate students, and undergraduates from around the world. During the year 2002-2003 we launched our new Special Focus on Computational Geometry and Applications. We held a preliminary workshop in a Special Focus on Communication Security and Information Privacy which goes into high gear in FY04. Our special focus on Computational and Mathematical Epidemiology, launched in Spring 2002, went into high gear.

During FY2003 , we ran the following **Special Focus Programs**:

Next Generation Networks Technologies and Applications (2000-2003)
Computational Molecular Biology (2000-2004)
Mathematics and the Foundations of Computer and Information Science (20 00-2003)
Computational Information Theory and Coding (2001-2004)
Data Analysis and Mining (2001-2004)
Computational and Mathematical Epidemiology (2002-2007)
Computational Geometry and Applications (2002-2005)
Communication Security and Information Privacy (2003-2006)

During 2002-2003, DIMACS held 35 conferences, seminar series, tutorials, workshops, and working group meetings as part of these special focus programs. There were, on average, 63 participants at each of these events, ranging from small numbers at our working group meetings to over 150 at our larger workshops.

We have worked very hard to plan future special foci. In Summer 2003, we launched a new Special Focus on Communication Security and Information Privacy. In FY 2003, we planned two new special foci, one on Computation and the Socio-Economic Sciences scheduled for 2004-2007 and one on Discrete Random Systems scheduled for 2005-2008.

Every special focus activity has led to dramatic scientific achievements – solutions of major open problems, the change of direction of research by a large community, the stimulus of new kinds of collaborations. This shows up dramatically in reports we have solicited on earlier special foci.

Multi-disciplinary, Multi-institutional Research Projects

In addition to special foci, DIMACS has continued or initiated major multi-disciplinary, multi-institutional research projects. In the period covered by this report, five such projects were ongoing, with major external funding for two of them through the intelligence community – an important new source of support for DIMACS' programs. Many of these projects involve research collaborations with international partners. Most of them heavily involve our industrial partners. The projects are:

LINK (software for education and research in discrete mathematics)
Monitoring Message Streams
Mining Multi-lingual Resources using Text Analytics
Bioterrorism Sensor Location
DIMACS-DIMATIA-Renyi Project on Combinatorial Methods
Computer Science and Decision Theory: Applications of Notions of Consensus
Homeland Security (Various Initiatives)

Funding for three of these programs was new in FY2003.

Homeland Security Research and Leadership at Rutgers

Homeland Security research efforts at DIMACS included a variety of projects including modeling social disruptions from bioterrorist attacks, using modern data mining methods to identify patterns underlying emerging diseases or bioterrorist attacks and find new "events" in streams of text messages, and developing methods of privacy-preserving information sharing. These efforts included partnerships with such agencies as the Centers for Disease Control and Prevention, the Secretary of Health and Human Services Smallpox Modeling Task Force, the intelligence community, and state and local health departments.

In addition, DIMACS played a leadership role in the founding of the Rutgers University Homeland Security Initiative (RUHSRI), which aims at coordinating homeland security research at the university, and in launching early RUHSRI projects such as a database of researchers at Rutgers.

International Programs

DIMACS is widely known internationally and hosts hundreds of foreign visitors a year. The center has numerous international partners with whom we have run or are running joint programs, adding this year LAMSADE (the Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision, based at Université Paris IX - Dauphine). During FY03, we applied for NSF funding for a joint project on Computer Science and Decision Theory (Applications of the Notion of Consensus) with LAMSADE, and we were recently informed that this will be funded. This will add to three other grants we have for such international activities. One is for a project on Combinatorial Methods that is joint with DIMATIA (the Center for Discrete Mathematics, Theoretical Informatics, and Applications) in Prague, Czech Republic, and the Renyi Institute (Mathematical Institute) of the Hungarian Academy of Sciences in Budapest. DIMATIA and Renyi are two of the best research in the world in combinatorial mathematics. We also have a grant for a pioneering international Research Experiences for Undergraduates program joint with DIMATIA. Finally, we have a grant for a series of international workshops on strategic information technology, co-sponsored by NSF and the European Union.

Other international partners are in Israel, Denmark, France, Korea, Hong Kong, and Taiwan. We ran our first joint workshop with the National Chiao Tung University in Taiwan in FY03 and continued our partnership with the Hong Kong University of Science and Technology by hosting a jointly-funded postdoctoral fellow.

Graduate Student Programs

Graduate students are a central part of DIMACS' program. Our *local* graduate student program includes small DIMACS awards for travel and software to support the research of local students, small awards for summer and winter break salary to support research projects, and jobs and projects for which graduate students are paid. We supported 23 students this way during summer 2002 and winter 2002-2003. The response to this program has been highly enthusiastic. We are also very pleased at the number of DIMACS researchers, including individuals from our industrial partners, who are participating in this program as research mentors. This program has led to rather impressive scientific payback, as demonstrated by scientific results obtained by the students. DIMACS also supports graduate students as Research Assistants under our special foci and research project grants.

DIMACS sponsors a graduate student organization revolving around discrete mathematics and theoretical computer science and funds a seminar series organized by and for the graduate students. It has also supported a graduate student organization in electronic commerce.

We have also continued our programs of sponsoring DIMACS visits by *non-local* graduate students and continued our long-standing policy of providing a supplemental allocation to each workshop for support of non-local graduate students. The impact of our program on non-local students has been very high and the research of some of our non-local graduate student visitors has also been stimulated and redirected by our programs.

Postdoc Programs

One of DIMACS' goals is to involve outstanding junior researchers in the life of the center. Our postdoc program is designed to accomplish this.

Because we want to train our postdocs to take an interest in educational activities and outreach beyond the mathematical sciences, we ask each postdoc to take part in such activities. Our postdocs meet this requirement in a variety of ways, such as helping with our K-12 and "Reconnect" programs (for 2- and 4-year college teachers) or mentoring in our undergraduate research (REU) program.

We have leveraged our funds for support of postdocs by sharing postdocs with the Institute for Advanced Study and including in our postdoc program postdocs appointed by our industrial partners. We have reached agreement with the Institute for Advanced Study to continue our joint postdoc program for three more years, and have identified resources for our share of the cost, with special help from Rutgers. This year, we also shared postdocs with AT&T Labs and NEC Labs and another with our international partner at Hong Kong University of Science and Technology. Our postdoctoral fellows from past years have been remarkably successful in obtaining positions, both in academics and in industry, and they have also risen rapidly to the top of their profession. A postdoctoral fellowship at DIMACS remains a very highly-sought-after award for junior researchers in discrete mathematics, theoretical computer science, and their applications.

Undergraduate Initiatives

We continued to plan and run our innovative programs in undergraduate education. Our Educational Modules series is aimed at rapid, electronic publication of modules aimed at undergraduates and teachers of undergraduates and in some cases high school teachers and students. This endeavor reached fruition in FY03 with the publication of the first two modules in the series.

The "Reconnect" project is aimed at reconnecting to the mathematical sciences research enterprise faculty whose primary job is teaching undergraduates. As part of this project, we ran a summer 2002 conference aimed at introducing such faculty to current research topics that are relevant to the classroom and at having them develop written materials usable in the classroom. The theme was Voronoi Diagrams (an area of Computational Geometry.) Some of the material developed by participants in our "reconnect" program is being developed in collaboration with editors in our Modules Series. Through an award from NSF, we expanded our "Reconnect" program to "satellite" locations and worked with faculty at those locations to help them develop such workshops. During FY03, we worked with faculty at Salem State College (Massachusetts) and Illinois Institute of Technology, helping them to plan their "Reconnect" programs, and we ran a Salem State satellite conference in June 2003.

Our traditional undergraduate program, the research experiences for undergraduates (REU program, continued in summer 2002 with a larger-than-ever group of students, 26. Through an award from NSF, we have combined our traditional REU program with an international experience, inviting some students from our sister center DIMATIA in Prague, Czech Republic, to join us at DIMACS, and then taking five of our REU students to DIMATIA. There were undergraduates from all over the US and from the Czech Republic involved in our program, including students shared with the Department of Mathematics' VIGRE REU program.

Precollege Programs

Our highly successful precollege programs have continued, with several major new developments. The Leadership Program in Discrete Mathematics for K-8 teachers continued its emphasis on running off-site programs and on developing portable materials that can be used to "franchise" our programs. Funds from the Educational Foundation of America and from state departments of Education have been used to fund this program. Participants in the Leadership Program follow up with workshops in their districts that reached some 5000 teachers in FY03. Our "Crash Course in

Discrete Mathematics" for high school teachers also continued.

Our Young Scholars Program for high school students continued in summer 2002 thanks to a series of gifts from the AT&T Foundation, the American Association of Artificial Intelligence, and the Gannett Foundation.

In FY03, we began to plan a major new initiative to foster interdisciplinary K-12 programs at the interface between the biological and the mathematical sciences. We will begin with a major conference on this topic and experiment with a future summer program for high school teachers.

Integration of Research and Education

We have worked to integrate research and education by including educational components in our special foci, involving postdocs and graduate students in educational programs, and through our undergraduate initiatives such as the modules series and the "Reconnect" project.

The DIMACS Connect Institute (DCI), funded by NSF and the AT&T Foundation, aims at integrating researchers, postdocs, students, and teachers, through a series of research workshops, a high school teacher course, and plenary activities that link these two main groups. There is a research experience for teachers, we work with them to bring their new-found view of research back to their classrooms and to arrange visits to high school classrooms by researchers. We invite teachers and their students back to DIMACS in the Spring for a high school student research conference. In summer 2002, DCI continued its theme of "Graph Theory and its Applications to Problems of Society." The high school student conference had 78 attendees from all over the US.

DIMACS Permanent Members

Since DIMACS started, a key to its strength has been the involvement of researchers from its partner institutions. There are now some 230 scientists who are "permanent" members of DIMACS, many coming from the partner institutions, academic and industrial, and including some 69 from Rutgers. These permanent members serve on DIMACS committees, organize DIMACS special foci and workshops, host visitors, help prepare grant proposals, participate in multi-disciplinary projects, and mentor students and postdocs. These members represent a wide variety of disciplines, not just the mathematical and computer sciences (including operations research and statistics). They include biologists, psychologists, chemists, economists, etc..

Visitor Program

A key to the success of the DIMACS programs is the interaction among a wide variety of people: Permanent members of the center, visitors, postdoctoral fellows, and students. In addition to our workshops, which host many visitors each year, we have a program of short- and long-term visits to DIMACS. These visits are usually centered around a special focus program, but sometimes just support the core research interests of researchers associated with DIMACS. During FY03, we hosted 73 visitors from all over the world in visits of a few days to the entire year. These visits resulted in many fruitful and continuing collaborations. Especially noteworthy are the collaborations initiated between junior and senior researchers and between students and faculty or students and industrial researchers. All in all, between the workshop and visitor programs, we had over 2000 visitors to DIMACS in FY03.

Publication Programs

Dissemination of results, both scientific and educational, is central to the success of DIMACS. A key part of this is our publication program. The centerpiece of this program is the book series. DIMACS volumes are published as handsome, bound books by the American Mathematical Society. Such volumes sometimes reflect the results of a particular workshop, sometimes the themes of a special focus program, and sometimes revolve around educational topics. During FY03, four books were published in the DIMACS series, bringing the series total to 61.

The DIMACS technical report series gives quick, widespread publicity to work done by DIMACS-affiliated researchers, students, postdocs, and visitors. During FY03, we published 61 technical reports.

FY03 saw the publication of the first two modules in the DIMACS Educational Modules series. The series consists of self-contained materials that can be used for one or more class meetings or for independent study, at the undergraduate, community college, or precollege level.

Institutional Membership: Industrial Partners

DIMACS began with two university partners (Rutgers and Princeton) and two industrial partners, AT&T Bell Labs and Bellcore. We now have 10 partners: Rutgers, Princeton, AT&T Labs, Bell Labs (Lucent Technologies), NEC Laboratories America, Telcordia Technologies (formerly Bellcore), Avaya Labs, IBM Research, HP Labs, and Microsoft Research. The first four industrial partners are "full partners." IBM Research and HP Labs were added in FY03. In the current economy, our industrial partners have cut back their financial support of DIMACS. However, the most important thing our institutional members bring to us is the encouragement they give individual scientists at their institutions to be involved in DIMACS activities. This is an especially vital contribution of our industrial partners. This kind of support has always been more important than their financial support and continues at an extremely high level. Industrial scientists are active in all DIMACS programs, organizing workshops, working groups, and special focus programs; serving on committees that meet as often as monthly; helping to prepare grant proposals; and participating in multi-institutional research projects.

Partnerships

A key goal of DIMACS is outreach to other communities, both within and without the mathematical sciences. In the past year, we have made major progress in enhancing and expanding our many partnerships. The highly successful partnership with the Institute for Advanced Study has continued with joint workshops and shared postdocs and plans for shared special focus programs.

Key international partnerships were described earlier.

Partnerships with professional societies and sister centers in the US and Canada continue, with joint programs with the Special Interest Group in Algorithms and Computation Theory (SIGACT) of the Association for Computing Machinery, American Mathematical Society, the Society for Industrial and Applied Mathematics, and such centers as the Institute for Mathematics and its Applications (University of Minnesota) and the Centre de Recherches Mathematiques (University of Montreal).

Prominent partnerships at Rutgers involve the Center for Math, Science, and Computer Education (sharing in our educational programs), the Center for Advanced Information Processing (CAIP), the Wireless Information Network Laboratory (WINLAB), the Center for Molecular Biophysics and Biophysical Chemistry, and the Biomaps Institute (new this year). The Rutgers Center for Operations Research (RUTCOR) is one of the central units in DIMACS.

Programs for Underrepresented Groups

DIMACS has a strong record of involvement of women, African American, and Hispanic scientists as organizers, speakers, visitors, graduate students, and participants in our many research and educational programs.

During FY2003, we continued our collaboration with the Douglass College (Rutgers University) Program for Women in Mathematics, Science, and Engineering.

Several African Americans and Hispanics play prominent roles in DIMACS and the involvement by African Americans in many of our programs such as Young Scholars is strong. A number of prominent African American scholars have agreed to help us recruit African American participants to our programs and in particular to work with us on our Reconnect Program. We have access to extensive mailing lists as a result of our hosting in 1996 of the Second Conference for African American Researchers in the Mathematical Sciences (CAARMS), which resulted in a DIMACS book, *African Americans in Mathematics.* We have followed up on the conference with continuing contacts with African American mathematicians and inform them of the many opportunities to participate in DIMACS activities. We have similar contacts with Hispanic Americans.

We continue to have active involvement with "historically black colleges and universities" (HBCU's) and other minority institutions. We initiated a joint program with HBCU Texas Southern University – as part of our precollege Leadership Program. Faculty from HBCU's Spelman College, Morgan State University, and Howard University as well as from high-minority-enrollment institutions such as University of Houston-Downtown have been involved with our "Reconnect" program. We have worked with Spelman to plan a "Reconnect" event on that campus and are in the process of working with Morgan State on a similar event. We have had extensive connections with University of Puerto Rico, including hosting visitors and REU students.

In the summer of 2002, a large contingent of minority undergraduate students from The Mathematical and Theoretical Biology Institute (MTBI) at Cornell participated in various programs in our Special Focus on Computational and Mathematical Epidemiology, including an international conference and several tutorials. The director of the MTBI summer program said this about the experience: "It is programs like these, programs that seriously attempt to attract and train students in interdisciplinary fields, that will equip minority students with the skills and motivation needed to overcome future challenges. DIMACS is succeeding in cultivating an ethnically diverse and sensitive future scientific workforce through its close collaboration with programs like MTBI."

Changes in Staff

There was quite a bit of turnover in the DIMACS staff this year. Maria Mercado joined us as workshop coordinator (replacing Jennifer Thiemann), Linda Casals as publicity coordinator (replacing Eric Luhrs), and Walter Morris as systems administrator (replacing Pablo Samuel). Dr. Alexander Barg served us as acting part-time associate director. We did a search for a full-time associate director and Dr. Brenda Latka accepted the position in Spring 2003, starting in September 2003 on a one-year trial basis.

Funding

The primary sources of funding for DIMACS are grants and the indirect cost return on those grants provided by Rutgers University. DIMACS received \$5.5M in new external funds in FY03, to go along with some \$3.5M in external funds already in place. We continue to have a very high success rate with our grant proposals.

The remainder of our financial support comes from grants and gifts from industrial partners, workshop registration fees, and Rutgers University. The financial support from our industrial partners continues, but in this difficult economic time for most of them, is down from previous highs. It was never a major source of funds for us. Still, the total value of direct support from our industrial partners was \$101K in FY03. The major contribution of our partners is in the heavy involvement of senior and junior researchers in DIMACS programs. Our four main industrial partners valued this in-kind contribution to DIMACS at \$1.6M in FY03 and it is a main reason that DIMACS is so successful.

In FY03, we continued to receive support from the New Jersey Commission on Science and Technology, which has supported DIMACS since our founding. However, this year marked the last year of NJCST funding. The Commission's \$100K annual contributions have been very helpful in supporting both core and exploratory programs at DIMACS, as "matching" for federal grants, and in supporting the DIMACS infrastructure.

While we keep them relatively low, especially for students, workshop registration fees are a modest source of income and they help pay for some of our workshop staff support.

In spite of difficult economic times, Rutgers has continued to support DIMACS through partial support of one staff person and the director's salary and with space for the center. In addition, in FY03, Rutgers renewed for three years its partial support for the prestigious joint DIMACS postdoctoral program with the Institute for Advanced Study. The indirect cost return from Rutgers is central to the operation of DIMACS. Of great importance is that the "emergency fund" we have built up through this allows us to make commitments for future programs long before we are assured of external funding for them. Without this, it would be very hard to operate.

2 A Brief History of DIMACS

DIMACS was founded as a National Science Foundation "science and technology center" (STC) through an 11-year award received in 1989. This highly prestigious award – one of only a handful in the country in the mathematical sciences – led us to establish a wide variety of research and education programs that have been widely praised and copied in many parts of the world. The center was founded as a partnership of Rutgers and Princeton Universities (with Rutgers being the lead institution), AT&T Bell Labs, and Bellcore. Other industrial partners have been added as the center has evolved and some 230 researchers, mostly from its partner institutions, are involved with DIMACS as "permanent members" of the center.

DIMACS has an international scope and its programs involve not just the permanent members but also numerous participants from outside the local institutions. In a typical year, the center hosts 30 to 35 research workshops with 1,500 attendees, 100 visiting scientists, 5 postdoctoral fellows, 50 2- and 4-year college faculty, 100 precollege teachers, and 100 high school students, as well as graduate and undergraduate students.

From the beginning, research at DIMACS was unique because the "themes" changed regularly and we developed our signature research activity, the "special year." As it grew, the center became a national and international resource that aimed to catalyze interdisciplinary collaborations and play a leadership role in the rapid development of new directions and subfields of science. Our early special years certainly accomplished this. For instance, DIMACS' 1990-91 Special Year on Complexity Theory of Interactive Computing played a central role in one of the most important developments in theoretical computer science in recent years, the development of ways to reformulate computer "proofs" so that the correctness of a computation can be verified with arbitrarily high confidence by a small number of random spot checks. A critical result leading to these developments, the discovery of connections between interactive proofs and approximation algorithms, was a direct result of the special year. It arose from the interactions between two diverse communities, the cryptography community and the combinatorial optimization community, and came about by applying recent methods of cryptography to the maximum clique problem of combinatorial optimization.

The early special year themes were closely related to discrete mathematics (dm) and theoretical computer science (tcs): Discrete and Computational Geometry; Graph Theory and Algorithms; Combinatorial Optimization. Later programs branched out from dm and tcs, to include programs partnering mathematical scientists with biological scientists, with epidemiologists, with industrial and mechanical engineers, with chemists, with physicists, and with social scientists, to name a few.

As DIMACS evolved, instead of changing our research themes annually as we did at the beginning, we evolved to longer and longer themed programs, "special focus programs," some still going on. Also, over the years, the special focus programs became more complex and reached out more and more to other disciplines. Our "special year" on "Mathematical Support for Molecular Biology" (MSMB), begun in 1994, ended up lasting six years, and was immediately followed by our present special focus on "Computational Molecular Biology." MSMB introduced a large number of computer scientists and mathematicians to the field of computational molecular biology and created important partnerships between mathematical and biological scientists that were groundbreaking and important. These partnerships had remarkable impact. For example, the ability of the t-haplotype, an ultra-selfish gene in mice, to persist at low levels in natural populations has been a subject of debate in the theoretical biology literature for a long time. Ultra-selfish genes are transmitted to offspring more often than would be consistent with Mendel's laws, yet they are harmful to the organism that carries them. A long-standing hypothesis attributed the low level persistence of such ultra-selfish genes to a balance between the loss of the gene due to stochastic effects in small populations and the re-introduction of the gene through migration. In a dramatic collaboration, a DIMACS computer scientist (who was a participant in the MSMB postdoctoral program) and a Princeton molecular biologist, who met through the MSMB tutorial, overturned this 40-year-old hypothesis.

Physicists played a central role in our 1996-98 Special Focus on Discrete Probability, which was also one of the first joint programs we ran with our long-term partner, the Institute for Advanced Study in Princeton. The partnership between computer scientists and physicists has had some important impacts and has continued well beyond that special focus. Some very exciting science came out of this special focus. The general belief in statistical physics is that spontaneous symmetry-breaking is something which, if it occurs at all, occurs when the temperature drops below some unique critical point. This had indeed been proved for the Ising model. During the special focus, a partnership among discrete mathematicians, computer scientists, and statistical physicists led to the solution to the long-open question of whether the same thing holds for the Widom- Rowlinson model, which is to the Ising exactly as vertices are to edges of a graph. Two visitors and a DIMACS member settled this in the negative by constructing a non-regular tree in which, as temperature is lowered, symmetry is broken, then re-established, then broken again. The partnerships between physicists and computer scientists developed during this special focus have played a critical role in the development of a new field of science that can be described as the study of "random discrete systems."

As time moved on, we also evolved to having more than one themed program each year, up to the point where in FY03 we had eight such programs running simultaneously. The variety of themes and partnerships in these programs has grown tremendously, including, for example, new thrusts in communication security and information privacy, epidemiology, and data mining.

Still, it is important for us to "stick to our roots" and there are quite a few programs at DIMACS devoted to issues in "pure" discrete mathematics and fundamental theoretical computer science. These programs have, over the years, led to significant scientific results. The four color problem, stemming from map coloring, has been one of the motivating problems for research in discrete mathematics, in particular in graph coloring, for over 100 years. Graph coloring has modern applications in areas as diverse as scheduling, traffic phasing, fleet maintenance, frequency assignment, and task assignment, and generalizations of graph coloring form the basis for a joint NSF-supported DIMACS project with international partners in Hungary and the Czech Republic. The original proof that every graph arising from a map could be colored in four or fewer colors was very long, complicated, and tedious and some of the details were not very clear, leading to a prolonged debate about the correctness of the proof. A new proof of this result, obtained by a DIMACS member and two visitors, is simpler and more accessible to complete checking. Moreover, it has been modified to obtain other exciting results about graph coloring.

As DIMACS matured, we started several multi-disciplinary, multi-institutional research projects. Our first two such projects resulted from the 1993-94 Special Year on Massively Parallel Computing. These large projects, supported by DARPA and NSF, involved researchers from DIMACS, WINLAB, AT&T, Bellcore, Georgia Tech, University of Massachusetts, William and Mary, Dartmouth, and two small companies, Ultra Corporation and Cooperating Systems Corporation. The projects, which ended in 2001, developed the first model and simulation of the global Internet with a hundred thousand hosts and routers, running IP packets with real routing protocols at a rate of more than one million packet hops per second. The resulting software is now used by over 40 companies and university research teams. Several of our present multi-disciplinary, multi-institutional projects are closely connected to issues of homeland security.

Industrial partnerships are vitally important to DIMACS. Researchers from our industrial partners play a central role in the life of the center, participating in research projects, organizing special foci and workshops, and participating in educational programs. An early example of research partnerships between university and industrial scientists came during our 1991-92 Special Year on Graph Theory and Algorithms. During that special year, a group of researchers from four different institutions, AT&T Labs, Bellcore, Rice University, and Rutgers, collaborated to solve what at that time was the "world's largest" unsolved Traveling Salesman Problem. The TSP problem they solved arose from the question of how to most efficiently connect 3,038 points for drilling holes in making a circuit board. Among the many ways in which we partner with our industrial colleagues is in shared postdocs. During the 1996-2000 Special Focus on Networks, a researcher from NEC Laboratories and a DIMACS postdoc developed an escrowed identity system to protect privacy in electronic toll collection and related applications. In the first tier of this 2-tiered system, enough information is provided by the user so that an operator can be sure the user is legitimate, but otherwise the user's identity is protected. If more information is required, for example in the case of a crime, the second tier provides it, through involvement of a third party (an escrow agent).

As DIMACS has evolved, we have also expanded our international collaborations. We now run joint activities with ten international partners, including funded research on topics in combinatorics and the connection between computer science and decision making.

From the beginning, we saw our research mission as fundamentally intertwined with our educational mission, and developed programs that integrated research and education from the postdoctoral level to the elementary school level. Our K-12 teacher programs, heavily informed by research at DIMACS on discrete mathematics, quickly received national acclaim and have had major impact on the development of curriculum "standards" in New Jersey and nationally. Literally tens of thousands of teachers nationwide have been impacted by our programs. Our "Reconnect" program has impacted many teaching faculty at 2- and 4-year colleges around the country and our programs for undergraduate research have included a pioneering international "research experiences for undergraduates" program, later mirrored in a unique "research experiences for teachers" program in the DIMACS Connect Institute.

The end of the STC award and its large and stable source of funding required us to raise money for our many projects individually. We now write a large number of grant proposals each year, going to numerous agencies and private sources and numerous programs within those agencies. NSF remains our primary sponsor, but even within NSF, we have obtained funds from numerous programs, including Computer Science; Mathematical Sciences; Social, Behavioral and Economic Sciences; Biological Sciences; Education and Human Resources; and International Programs. Throughout the lifetime of the center, our academic and industrial partners have been strong supporters, especially Rutgers, which has played the central leadership role in DIMACS. Our industrial partners continue to be strong supporters of the center and play a crucial role in the projects and programs of DIMACS, partly through financial support, but mostly through heavy involvement of leading researchers in our center.

3 Special Focus Programs

The primary scientific activity at DIMACS remains the "special focus," a focused and coordinated collection of activities that span a period of time. The focus concentrates on an area of discrete mathematics/theoretical computer science (broadly speaking) and its applications that is ripe for rapid development, contains problems appropriate for high quality research mathematicians and computer scientists, and, in most cases, has strong potential for connections to other areas of the sciences, the mathematical sciences, and/or industry. This signature way of doing research at DIMACS revolves around a series of tutorials, interdisciplinary research working groups, and workshops (more formal conferences). There are many visitors and we involve postdoctoral fellows, graduate students, and undergraduates from around the world. During the year 2002-2003 we launched our new Special Focus on Computational Geometry and Applications. We held a preliminary workshop in a Special Focus on Communication Security and Information Privacy which goes into high gear in FY04. Our special focus on Computational and Mathematical Epidemiology, launched in Spring 2002, went into high gear.

During FY2003, we ran the following **Special Focus Programs**:

Next Generation Networks Technologies and Applications (2000-2003)
Computational Molecular Biology (2000-2004)
Mathematics and the Foundations of Computer and Information Science (20 00-2003)
Computational Information Theory and Coding (2001-2004)
Data Analysis and Mining (2001-2004)
Computational and Mathematical Epidemiology (2002-2007)
Computational Geometry and Applications (2002-2005)
Communication Security and Information Privacy (2003-2006)

During 2002-2003, DIMACS held 35 conferences, seminar series, tutorials, workshops, and working group meetings as part of these special focus programs. There were, on average, 63 participants at each of these events, ranging from small numbers at our working group meetings to over 150 at our larger workshops.

We have worked very hard to plan future special foci. In Summer 2003, we launched a new Special Focus on Communication Security and Information Privacy. In FY 2003, we planned two new special foci, one on Computation and the Socio-Economic Sciences scheduled for 2004-2007 and one on Discrete Random Systems scheduled for 2005-2008.

Every special focus activity has led to dramatic scientific achievements – solutions of major open problems, the change of direction of research by a large community, the stimulus of new kinds of collaborations. This shows up dramatically in reports we have solicited on earlier special foci.

3.1 2000-2003 Special Focus on Next Generation Networks Technologies and Applications

Overview:

Begun in January 2000, the Special Focus on Next Generation Networks Technologies and Applications is planned to last three and a half years, continuing many of the same themes introduced in the 1996-2000 Special Focus on Networks.

Organizers:

- B. R. Badrinath, Rutgers University
- Sandeep Bhatt, Akamai Technologies

- Joan Feigenbaum (co-chair), Yale University
- S. Muthukrishnan (co-chair), Rutgers University and AT&T Labs

Sponsors and Collaborating Institutions:

The National Science Foundation
The National Security Agency
The New Jersey Commission on Science and Technology
The Office of Naval Research
Rutgers University
Princeton University
AT&T Labs - Research
Bell Labs
NEC Laboratories America
Telcordia Technologies
DIMACS Affiliate Members:
  Avaya Labs
  IBM Research
  Microsoft Research

Theme of Program:

As the computer age has reached the end of the millennium, there has been a major shift, namely, to interconnect computation and global communications as well as share information. This convergence of computing, telecommunications and information is poised to take off even further with novel technology (e.g., cable, wireless and other high speed/bandwidth networking options) and phenomenally successful applications (e.g., web). We are seeing the proliferation of mobile computation units, networking of unprecedented scale, bandwidth demand growing at an exponential rate, and petabytes of distributed shared data. The emerging technological reality offers opportunities that will dramatically transform society.

While the underlying technologies continue to advance at a rapid pace, the fundamental principles governing the design, deployment and use of next-generation networks of unprecedented scale, heterogeneity, and complexity are not entirely clear. How are systems involving the emerging technologies (such as cable or wireless) best designed? This question pervades all levels of networking: at the network level (providing a personal network space, designing QoS for heterogeneous networks such as wired-cable and wired-wireless); at the application level (providing a thin-client view of the web for mobile units, secure electronic commerce, etc.); at the system level (providing interoperability, a homogeneous view of the system in the presence of mobility, etc.); at the protocol level (how to modify TCP for wired-wireless networks, how to multicast in wireless networks, etc.); at the link level (scheduling coding and power level for packet transmissions). New network services raise new challenges: Privacy can be threatened; small faults can cause cascading network failures with severe economic consequences; the sheer number of requests for information can slow the networks and information servers to a crawl; protocols created for previous generation networks may not be appropriate for the scale and complexity of newer technologies; achieving effective interoperability among systems remains extremely difficult; managing distributed systems with mobile software is poorly understood; multiple networks and software systems are being interconnected without a full understanding of the ramifications. This special focus seeks to take advantage of new opportunities and to solve underlying problems by addressing the theoretical foundations on which the emerging environment of commingled communication, information, and computation is based.

The special focus is concentrating on the following:

- Algorithms to achieve cost- and service-effective use of the resources in the emerging environment in virtually all areas, including communications (mobile or stationary, network design, protocols, routing), large scale planning and scheduling, and economic decision making. This will involve expertise from fields such as optimization, approximation, probability and queuing theory, control theory and game theory.

- Models for massive data analysis of multimedia information (e.g., data warehousing and data mining) and traffic and information management in global intelligent networks. Because existing algorithms for data analysis are unlikely to scale, new methods are needed to organize, store, and retrieve data, discover patterns, and use them for planning and decision making. This involves expertise from diverse areas including data structures, databases, data compression and coding, statistics, learning, geometry and visualization.

- Fundamental theoretical work to address problems of reliability, security and privacy. The design of systems that maximize the potential for new technologies while protecting security and privacy requires expertise in fields such as verification, distributed computing, probability, coding theory, number theory, cryptography and security.

This special focus is intended to foster closer working relationships between the algorithms and networking technology communities. With the help of such relationships, we are pursuing the development of the fundamental theory, models, and scalable algorithms that cut through a variety of application domains and will enable us to manage the present technology and design the future technology, from the physical layer to the application layer. Foundational work is required to solve existing problems and to lay the groundwork for the next generation networks technologies and applications, and our special focus is aimed at pinpointing the opportunities for theorists to make a contribution and at stimulating the research needed to take advantage of these opportunities.

Workshops and Working Group Meetings

- **Workshop on Management and Processing of Data Streams**
  - Dates: June 8, 2003
  - Location: San Diego, California
  - Organizers: S. Muthukrishnan, Rutgers University; Divesh Srivastava, AT&T Labs - Research
  - Attendance: 72

  This workshop was part of the Federated Computing Research Conference (FCRC03). Measuring and monitoring complex, dynamic phenomena – atmospheric conditions, stellar patterns, movement of financial markets, traffic evolution in telephone and internet communication infrastructures, usage of world wide web, email and newsgroups – produces stream data, i.e., data that arrives as a series of "observations", often very rapidly. These observations generally have intricate relationships between them, are often grouped in multiple ways into logical events, and are naturally collected via distributed platforms.

  Databases and applications are increasingly required to manipulate stream data. This presents many challenges which are just beginning to be addressed.

  - Fundamental algorithmic models and results for data streams are being studied in the theoretical computer science community.

– Data stream management systems are being designed and prototyped in the database community.
– IP traffic analysis is a powerful application domain for data streams, and is being addressed in the networking community.
– Programming language and software support needed for managing data streams is being determined in the systems community.
– Custom and off-the-shelf hardware support is being developed for data streams in the networking and graphics communities.

FCRC'03 provided a valuable opportunity to engage this large and diverse community of interest in drawing up the specification of data stream management systems, understanding their power and limitations, as well as motivating the applications that will drive the development of data stream systems. This workshop provided a forum to discuss the progress on all aspects of managing stream data. The audience included researchers, practitioners and graduate students interested in data stream management from the database (SIGMOD/PODS) community, as well as from other member FCRC conferences.

Selected Research Results

Signature Computation on Insecure Devices such as Mobile Phones. Visitors Yevgenij Dodis (NYU), Jonathan Katz (Visitor, U-Maryland), Shouhuai Xu (Visitor, UC-Irvine) and Moti Yung (Columbia) studied signature computation on insecure devices such as mobile phones, operating in an environment where the private (signing) key is likely to be exposed. Strong key-insulated signature schemes are one way to mitigate the damage done when this occurs. In the key-insulated model, the secret key stored on an insecure device is refreshed at discrete time periods via interaction with a physically-secure device which stores a "master key". All signing is still done by the insecure device, and the public key remains fixed throughout the lifetime of the protocol. In a strong (t, N)-key-insulated scheme, an adversary who compromises the insecure device and obtains secret keys for up to t periods is unable to forge signatures for any of the remaining $N - t$ periods. Furthermore, the physically-secure device (or an adversary who compromises only this device) is unable to forge signatures for *any* time period. The work of Dodis, Katz and Xu presents constructions of strong key-insulated signature schemes based on a variety of assumptions. It gives an improved construction of a strong (t, N)-signature scheme whose security may be based on the discrete logarithm assumption in the random oracle model. This construction offers faster signing and verification than the generic construction, at the expense of $O(t)$ key update time and key length. These researchers also construct strong $(N - 1, N)$-key-insulated schemes based on any "trapdoor signature scheme"; their resulting construction in fact serves as an identity-based signature scheme as well. This leads to very efficient solutions based on, e.g., the RSA assumption in the random oracle model.

Random Graph Models of Sensor Networks. S. Muthu Muthukrishnan (Rutgers) and Gopal Pandurangan (Visitor, Purdue) studied random graph problems motivated by sensor networks. Sensor networks, a technology likely to have a significant impact on our lives in the future, are modeled by random graphs $G(n, r)$ where the n vertices are represented by random points on the unit square. A pair of vertices at most r apart is connected by an edge. Muthukrishnan and Pandurangan studied three properties of such networks: coverage (how many nodes are needed to cover the entire area), high quality paths (those with a small stretch, i.e., the ratio of the shortest path between the nodes in the graph to the distance in the plane), and disjoint paths (the number of vertex-disjoint paths between a given pair of nodes). As a result of this work, bounds were established on r as a function of n for each of these properties. The bounds show that with each sensor (node) transmitting at power slightly greater than needed to ensure connectivity of the underlying random network, (a) the network will nearly cover the entire area in the plane, (b) shortest paths between any two nodes are high-quality paths, (c) there are many vertex-disjoint paths between any two nodes, all of high quality.

Mix Networks. Markus Jacobsson (Bell Labs and, later, RSA Technologies) proposed and improved techniques for so-called mix networks. These are constructions useful for Internet voting; privacy preserving browsing; and controlled anonymity for payment transactions. The new results allow for more efficient and more versatile mix networks.

Collaboration between Low-Power Mobile Devices. Markus Jacobsson (Bell Labs and, later, RSA Technologies) also worked in the area of low-power mobile devices, such as military sensors, cellular phones, and radio-frequency identification (RFID) tags. The work has involved means for promoting collaboration between selfish devices (who want to maximize their own benefits, but may not care about the benefit of their peers). It has also involved work on providing more efficient computational methods for traversing one-way hash chains, a problem that often comes up in work relating to low-power mobile devices, due to its application for authentication.

3.2 2000-2004 Special Focus on Computational Molecular Biology

Overview: Funded by a substantial three year award from the Alfred P. Sloan Foundation and with a modest three year contribution from the NSF, the Special Focus on Computational Molecular Biology is a highly active focus through its four year duration.

Organizing Committee:

- Martin Farach-Colton, Rutgers University, Chair
- Ron Levy, Rutgers University
- Craig Nevill-Manning, Rutgers University, Assistant Chair
- Wilma Olson, Rutgers University
- Fred Roberts, Rutgers University

Steering Committee:

- Bonnie Berger, MIT
- Helen Berman, Rutgers University
- Douglas Brutlag, Stanford University
- Andrea Califano, First Genetic Trust
- Nick Cozzarelli, UC Berkeley
- Dannie Durand, Carnegie Mellon University
- Jim Fickett, SmithKline Beecham
- Sorin Istrail, Celera Genomics/Applied Biosystems

- Thomas Lengauer, GMD
- Michael Liebman, Roche Bioscience
- Mona Singh, Princeton University
- Martin Vingron, National Cancer Center of Germany
- Michael Waterman, University of Southern California

Sponsors and Collaborating Institutions:

The Alfred P. Sloan Foundation
The National Science Foundation
Celera Genomics
SmithKline Beecham
The New Jersey Commission on Science and Technology
Rutgers University
Princeton University
AT&T Labs - Research
Bell Labs
NEC Laboratories America
Telcordia Technologies
DIMACS Affiliate Members:
    Avaya Labs
    IBM Research
    Microsoft Research

Theme of Program:

The special focus is motivated by major developments at the interface between biology and information science:

First, because of new technologies, we are receiving massive amounts of data and, moreover, the data are arriving in diverse forms and are often unstructured and non-homogeneous. This requires the intervention of experts at handling databases and processing massive amounts of information. Moreover, data are now available at much higher levels of organization than heretofore (for example at the whole genome level rather than at the gene level), which makes possible dramatic new biological understanding if we can find ways to handle the complex algorithmic questions that arise.

Second, new biotechnologies such as gene expression arrays are powerful new experimental tools, but these tools and the development of new ones need to be integrated with fundamental research concerning algorithms. Thus, close collaborations between biological scientists and computer scientists need to be developed and/or strengthened.

Third, new mathematical and computational tools have made it possible to understand biological processes at a much more complex level than before, for example at the level of cellular communities, the immune system, protein structure and design, and processes of learning and memory. This comes at a time when biologists are finding tools to explore how parts of a biological system (e.g., genes or molecules) interact, something that has long been understood to be just as important as understanding the parts themselves but which has been largely inaccessible to experimental or analytical techniques. These new tools make it possible to formulate powerful new mathematical models of biological systems, grounded in the communication and information transmission fundamental to those systems. This calls out for the expertise of mathematical modelers and information scientists who are best able to take advantage of the powerful new mathematical methods while at the same time keeping their models realistic through interactions with biological scientists.

The DIMACS Special Year on Mathematical Support for Molecular Biology commenced in 1994-95 with a very intense level of activity and has continued at a significantly lower level since then. Through that special year, we focused the interests of a large number of computer scientists and mathematicians on the field of molecular biology, created real partnerships between mathematical and biological scientists, and trained outstanding young people to work in computational biology. We feel that the time is right for another highly visible special year thrust, like the one in 1994-95. At the time of our first special year, there was a need to address, review, and coordinate a variety of topics and methods, as well as to introduce many people to the field. Computational biology prior to the first special year was largely focused on a few problems such as alignments, physical mapping, and reconstructing phylogenies. Now, many of the topics emphasized in the first special year have become "standard" topics of research in computational biology and the field has become much broader in its focus. Other topics have begun to be investigated in a much more far-reaching way using methods of computer science, for example problems involving gene finding and motif recognition, protein and RNA folding, protein structure prediction, and linkage analysis. New biotechnologies such as SBH (sequencing by hybridization), optical mapping, and EST (expressed sequence tags) have been brought to the attention of a much larger number of computer scientists so that they can formulate and research the right questions concerning new technologies. But a whole host of new problems that lie at the interface between the computational/mathematical sciences and the biological sciences have arisen, stimulated, as we have noted, by the availability of massive amounts of new data, the integration of new experimental methods with algorithmic methods, and the development of powerful new mathematical tools for modeling ever-more-complex biological systems.

The title of the special focus, "Computational Molecular Biology," is certainly too broad to define its focus. We concentrate on those areas where discrete mathematics (DM) and theoretical computer science (TCS) seem likely to have a major impact. A fundamental premise of the special focus is that some of the most central problems in modern molecular biology are essentially problems involving the combinatorial and algorithmic questions of DM and TCS and it is a basic objective of the focus to create partnerships between biological and mathematical/computer scientists so that some of these central biological questions can be precisely formulated and analyzed. It is our expectation that by bringing together some of the world's leading mathematicians/computer scientists with leading biological scientists, major progress can be made. Moreover, it is our expectation that we can establish and nurture lines of communication and collaboration so that the new methods of information science that arise from the special focus can be quickly adapted to biological applications, and can be readily communicated to the biological community.

Workshops and Special Events

- **Workshop on Computational Methods for SNPs and Haplotype Inference**
  - Dates: November 21 - 22, 2002
  - Location: Rutgers University
  - Organizers: Sorin Istrail, Celera Genomics; Andy Clark, Cornell University; Michael Waterman, University of Southern California and Celera
  - Attendance: 114

  This workshop arose from the fact that The ability to score large numbers of DNA variants (SNPs) in large samples of humans is rapidly accelerating, as is the demand to apply these data to tests of association with diseased states. The problem suffers from excessive dimensionality, so any means of reducing the number of dimensions to the space of genotype classes in a biologically meaningful way would likely be of benefit. Linked SNPs are often statistically associated with one another (in "linkage disequilibrium"), and the number of distinct configurations of multiple tightly linked SNPs in a sample is often far lower than one would expect from independent sampling. These joint configurations, or haplotypes, might be a more biologically meaningful unit, since they represent sets of SNPs that co-occur in a population. Recently there has been much excitement over the idea that such haplotypes occur as blocks across the genome, as these blocks suggest that fewer distinct SNPs need to be scored to capture the information about genotype identity. There is need for formal analysis of this dimension reduction problem, for formal treatment of the hierarchical structure of haplotypes, and for consideration of the utility of these approaches toward meeting the end goal of finding genetic variants associated with complex disease.

- **Workshop on Protein Domains: Identification, Classification and Evolution**
  - Dates: February 27 - 28, 2003
  - Location: Rutgers University
  - Organizers: Stephen Bryant, NIH; Teresa Przytycka, Johns Hopkins University
  - Attendance: 118

  Protein domains may be defined as elementary units of protein structure and evolution, capable, to some extent, of folding and functioning autonomously. Understanding of domain structure, function and evolution is a fundamental step towards understanding of a living organism. Identification of protein domains provides an inside into protein function.

  Current databases provide a vast amount of protein sequence and structure data, and it is likely that members of the majority of protein families have been already observed. Thus we are in the position to ask a question which could not be addressed previously: Can one use this information to identify domain sequences and structures computationally? How can we extract their properties?

  By analyzing the database of known sequences and structures and the properties of conserved domains we hope to get some insight into domain identification, properties and evolution. The amount of the data, it's complexity and biological context open challenges of an interdisciplinary nature. For example, identification of a structural domain requires on one hand sophisticated algorithms that can search for patterns in 3-dimensional data and on the other hand a clear, biologically meaningful and computationally tractable, definition of protein domain.

  The workshop was devoted to computational challenges in this new phase of understanding protein domain organization. The goal of the workshop was to bring together biological and computational/mathematical scientists to discuss the state of art and the open questions focusing on the following aspects of protein domains:

  - Methods for identification of protein domains.
  - Protein domain comparison and classification.
  - Mechanisms of domain evolution.
  - Topological and geometrical properties of protein domains.
  - Relation between sequence and structure conservation.

Leading specialists of the field gave invited presentations. There was a poster session and the organizers invited poster contributions.

- **Working Group on Mathematical and Computational Aspects Related to the Study of The Tree of Life**
  - Dates: March 11 - 14, 2003
  - Location: Rutgers University
  - Organizers: Melvin F. Janowitz, DIMACS; Francois-Joseph Lapointe, Universite de Montreal; F.R. McMorris, Illinois Institute of Technology; Fred Roberts, DIMACS
  - Attendance: 26

  DIMACS has held two working group meetings on “Bioconsensus.” The general goal of these meetings was to investigate the use of consensus techniques in evolutionary biology, and in particular their applications to phylogenetics. The meetings were highly successful and resulted in the publication of volume 61 in the DIMACS book series. There was a strong feeling among the participants in these meetings that there should be a third event, but that it should not focus on the specific topic of Bioconsensus. There was some sentiment for a further meeting on supertrees, and more generally for a working group designed to explore algorithmic and mathematical aspects related to the study of the Tree of Life. There is considerable interest in this type of project in both the Biology and the Mathematics community, as witnessed by the following:

  1. The Tree of Life web project. (www.tolweb.org/tree/phylogeny.html)

  2. The National Science Foundation program (NSF-02-074) whose proposal deadline was May 22, 2002, and whose title is Assembling the Tree of Life.

  3. The lead article by Richard Karp in the May, 2002 AMS Notices ”Mathematical Challenged from Genomics and Molecular Biology”.

  4. The symposium held May 30-June 1, 2002 at the American Museum of Natural History entitled ”Assembling the Tree of Life: Science, Relevance and Challenges”

  5. A symposium on supertrees that was held in conjunction with the annual Evolution Meeting (Chico, CA, June 21-24, 2003). The title of this is: ”Phylogenetic Supertrees: the stage play ”.

  6. Kluwer Academic will be publishing a book in 2003 called *Phylogenetic supertrees: the book.*

  Vast quantities of molecular data are becoming available, and there is a need to provide efficient computer algorithms that will appropriately scale to accommodate the size and quality of the underlying data sets. We recognize that for many reasons viral evolution may have features not necessarily present in the evolution of organisms on a multicellular scale. Since a workshop on the role of evolution in epidemiology is in the planning stage, our intention was to not focus on such organisms.

How does one handle possible errors in the data? What about missing data? Certainly statistics must play a role here, but we need also to study combinatorial methods, consensus methods, some notion of approximate compatibility analysis, generalized pyramids, and some form of cluster analysis. How, for example, should one deal with reticulate evolution? Should errors in data be dealt with by allowing alternate models of evolution, or by allowing reversals and overlap or is there some other technique dependent upon the data? There is a need to develop better techniques for handling microarray data as well. Here new consensus or classification techniques need to be developed. Thought should also be given to the development of appropriate measures of dissimilarity. Are there any techniques other than string matching of interest to the biological community?

Existing methods of supertree construction should be compared with simulations. The NP-completeness of current techniques should be tested, and where appropriate fast approximate algorithms (heuristics) should be developed. Is there a role for compatibility analysis? What about parsimony? There is a need to compare MRP methods with other methods of constructing a supertree. There is a need even to arrive at an acceptable definition of a supertree. How does one deal with massive data sets to derive an accurate Tree of Life?

- **Workshop on Medical Applications in Computational Geometry**
  - Dates: April 2 - 4, 2003
  - Location: Rutgers University
  - Organizers: Danny Chen, University of Notre Dame; Jean-Claude Latombe, Stanford University
  - Attendance: 66

  Computer technology plays an increasingly important role in modern medicine and life sciences. Many medical problems are of a strong geometric nature and may benefit from computational geometry techniques. The DIMACS workshop on Computational Geometry and Medical Applications aimed to provide a forum for researchers working in computational geometry, medicine, and other related areas to get together and exchange ideas, and to promote cross-fertilization and collaborations among these areas. The theme of the workshop was on the exploration of the applicability of computational geometry to medical problems and the new challenges posed by the current medical research and practice to the geometric computing study. Examples of topics discussed included surgical simulation and planning, geometric representation and modeling of medical objects and human-body tissue structures, geometric problems in medical imaging, computational anatomy, registration and matching of medical objects, etc.

- **BioMaPS/DIMACS tutorial: Introduction to Modern Concepts in Biology for Mathematical and Physical Scientists**
  - Dates: June 23 - July 3, 2003
  - Location: Rutgers University
  - Organizers: William Sofer and Paul Ehrlich, Rutgers University
  - Attendance: 40

  This BioMaPS/DIMACS tutorial was a 2-week intensive program divided into three related sections:

  1. molecular biology, genetics, and biotechnology

2. cell biology,
3. presentations of current research

The goal of the tutorial was to introduce participants to topics in molecular and cell biology that are relevant to those who wish to explore or initiate work at the interface among biology, mathematics, computer science, chemistry, and physics. The tutorial was appropriate for graduate students, postdocs, and mid-career scientists who wish to change direction. The first week of the tutorial, entitled "The DNA Revolution," was designed by William Sofer, Professor of Genetics and member of the faculty of the Waksman Institute at Rutgers University, to introduce the fundamentals of modern molecular biology, genetics, and biotechnology to participants who have little background in biology or biochemistry. In addition, the tutorial examined scientific questions that interest present-day biomedical researchers. The emphasis was on problems that can be addressed through quantitative approaches that are currently considered the domain of physics, chemistry, mathematics, or computer science. Lively discussions suitable for postgraduate training of physical scientists and mathematicians occurred in these sessions.

Although the first week was focused on providing participants with a clear understanding of the major issues and accomplishments of modern biology, participants also acquired an essential factual background through the use of an interactive "Internet Textbook." This primer of biological information covers important chemical structures and biochemical processes as well as classical genetics. Building on this foundation are descriptions of several methods and research projects in genetic engineering and bioinformatics. This material equipped participants with the framework for understanding most other relevant technologies. In addition, a virtual laboratory has been designed to offer participants realistic exercises that illustrate the day-to-day activities of modern-day bench scientists. Finally, participants were given the opportunity to develop bioinformatics research projects as part of the "hands-on" approach to training that Professor Sofer believes is most effective.

The second week of this tutorial started with a review of cell biology that included lectures given by Paul Ehrlich, Adjunct Professor of Cell Biology and member of the BioMaPS Institute at Rutgers University. He examined energy generation, cell division including cell cycle control, cell communication including adhesion, and apoptosis as well as the structure and function of the cytoskeleton, membranes, and intracellular compartments. Basic information was supplemented with a summary of open questions published by leaders of each discipline. In addition, major topics were illustrated by an in-depth analysis of recently published biophysical studies that have made a significant contribution to the understanding of biological phenomena, e.g., mathematical models of cell and viral kinetics that explain the emergence of drug-resistant virus despite the use of highly effective drugs. Finally, participants became familiar with important Web sites useful to researchers at the interface of biology with the mathematical and physical sciences.

The second week ended with presentations on current research in fields at the interface among the biological, mathematical, and physical sciences. The first two parts of the tutorial provided the participants with adequate preparation to gain insight and understanding from these presentations. They were given by faculty of DIMACS and the BioMaPS Institute at Rutgers University.

Graduate students and others interested in the interface among the biological, mathematical, and physical sciences typically have extensive training in chemistry, physics, mathematics, statistics, or computer science but are less familiar with the biology-oriented, basic knowledge that modern biologists have acquired. This foundation is indispensable for the identification and appreciation of the main problems in biomedical research. Participation in the BioMaPS Institute/DIMACS tutorial can be seen as a major step in building the foundation.

DIMACS-Celera Genomics/Applied Biosystems Graduate Student Award

Through funds received from Celera Genomics/Applied Biosystems, DIMACS continued its joint DIMACS-Celera Genomics/Applied Biosystems graduate student award program. The awards are given each year of the DIMACS Special Focus on Computational Molecular Biology and are aimed at recognizing outstanding achievements of junior researchers in the field. Winners are invited to present a talk at a DIMACS workshop with expenses paid and they receive a monetary award as well. Nominations are made by letter of recommendation from mentors.

In this third year of the program, the winner in Spring 2003 was Ziv Bar-Joseph of Massachusetts Institute of Technology. At an October 2003 workshop on Data Mining and Bioninformatics, he presented a talk about Structure of Gene Expression.

Selected Research Results

Consensus List Colorings of Graphs and Physical Mapping of DNA. N.V.R. Mahadev (Fitchburg State College) and Fred Roberts (Rutgers) studied a bioconsensus problem relating to graph coloring and investigated the applicability of the ideas to the DNA physical mapping problem. In many applications of graph coloring, one gathers data about the acceptable colors at each vertex. A list coloring is a graph coloring so that the color assigned to each vertex belongs to the list of acceptable colors associated with that vertex. Mahadev and Roberts considered the situation where a list coloring cannot be found. If the data contained in the lists associated with each vertex are made available to individuals associated with the vertices, it is possible that the individuals can modify their lists through trades or exchanges until the group of individuals reaches a set of lists for which a list coloring exists. Mahadev and Roberts described several models under which such a consensus set of lists might be attained. In the physical mapping application, the lists consist of the sets of possible copies of a target DNA molecule from which a given clone was obtained and trades or exchanges correspond to correcting errors in data. They showed that there are graphs and list assignments for which not only is there no list coloring, but almost everyone has to add an acceptable color that was not on their original list in order for a list coloring to exist. In physical mapping, this means that there are situations where essentially every list of copies needs to be expanded.

Immunoglobulin Folds. Boris Galitsky (Visitor), Andrey Dementiev and Sergey Shelepin addressed the contradiction that a rather limited number of residues or classes of amino acids (about 10) determines the fold (Immunoglobulin-like) for the sequence that is about 100 residues long. The Immunoglobulin fold comprises the protein super-families with rather distinguishing sequences with less than 10% identity; their sequence alignment can be accomplished only taking into account the 3D structure. Therefore, it is believed that discovering the additional common features of the sequences is necessary to explain the existence of common folds for these (SCOP) superfamilies. The analysis of pair-wise interconnection between residues of the multiple sequence alignment helped these researchers to reveal the set of mutually correlated positions, inherent to almost every superfamily of protein fold. Hence, the set of constant positions plus the set of variable but mutually correlated ones can serve as a basis of having the common 3D structure for distinguishing protein sequences.

3.3 2000-2003 Special Focus on Mathematics and the Foundations of Computer and Information Science

Overview: The Special Focus on Mathematics and the Foundations of Computer and Information Science began in September 2000 and continued through August 2003, funded by a grant from NSF.

Organizers:

- Michael Fredman, Rutgers University
- Janos Komlos, Rutgers University
- Fred Roberts, Rutgers University

Sponsors and Collaborating Institutions:

The National Science Foundation
The New Jersey Commission on Science and Technology
The Office of Naval Research
Rutgers University
Princeton University
AT&T Labs - Research
Bell Labs
NEC Research Institute
Telcordia Technologies
DIMACS Affiliate Members:
  Avaya Labs
  IBM Research
  Microsoft Research

Theme of Program:

The rapidly changing world of information, communications, and computation is a stimulus for the development of new models and new concepts lying at the foundation of computer and information science and their development will require ever-closer connections with many methods of mathematics. We will investigate the connections between foundational issues in the computer and information sciences and different branches of mathematics in a series of "working" workshops involving both mathematicians and computer scientists. We will de-emphasize formal talks and emphasize discussion time and opportunities to explore new potential collaborations between computer scientists and mathematicians. We will follow up the workshops with opportunities for participants to meet at DIMACS to pursue the research directions identified.

During the first 30 to 40 years of the computer age, the science of computation and communication has been profoundly and broadly influenced by the models, conceptual framework, and methods developed within theoretical computer science (TCS). For example, the theory of languages and automata led to the development of high level programming languages and compilers; the design and implementation of large databases was fundamentally based on theoretical models; the notion of polynomial time tractability which developed within the field of computational complexity provided a fundamental criterion for algorithmic efficiency that is now universally applied; the field of algorithm design and analysis yielded a large library of practical algorithms, as well as basic principles for attacking new computational problems; the study of private secure communication within computational complexity revolutionized the classical field of cryptography, broadening its scope enormously to complex multi-party transactions, and providing fundamental new methods for implementing such transactions.

TCS is inherently a highly mathematical discipline. Its basic philosophy and methodology – formulating and analyzing precise models of computational systems that abstract key aspects of their behavior – has been fundamentally influenced by the methodology of applied mathematics. TCS draws most heavily on discrete mathematics, probability, and logic, but it has growing connections throughout mathematics. The field of TCS initially grew out of logic and the basic notions of computability and complexity. Discrete mathematics soon became a central tool in the design and analysis of algorithms, as well as in the formulation and analysis of general computational models. Probabilistic methods have become increasingly important both to algorithm design, where randomized algorithms and probabilistic analysis have become essential tools, and complexity theory, where the probabilistic viewpoint is fundamental to our understanding of computational models.

Computational geometry has become a major subfield of TCS, drawing heavily on both discrete mathematics and classical geometry. Number theory has played a key role in the design of cryptographic protocols. Computational algebra, algebraic topology, Fourier analysis, and geometry play an increasingly important role in proving lower bounds on computational complexity, as well as in the design of algorithms. Coding theory, long influenced by deep mathematical methods, particularly from algebra and algebraic geometry, presents methods for the robust representation of information and has combined with methods of computational information theory to achieve codes that are close to theoretical limits. The role of these and other areas of mathematics in the development of computer and information science, and in particular in theoretical computer science, is the major theme of this special focus.

At the same time that TCS is increasingly influenced by the methods and results of mathematics, the algorithmic methods of TCS are increasingly important in the analysis of fundamental problems of mathematics. For example, the discovery of algorithms that permit the rapid processing of vectors, polynomials, matrices, and other central mathematical concepts can be a tool in the solution of important mathematical questions. Exploring this direction of the connection between mathematics and TCS is a second theme of this special focus.

Workshops and Working Group Meetings

- **Workshop on Applications of Lattices and Ordered Sets to Computer Science**
  - Dates: July 8 - 10, 2003
  - Location: Rutgers University
  - Organizers: Jonathan Farley, Massachusetts Institute of Technology; Melvin Janowitz, DIMACS / Rutgers University, Jimmie Lawson, Louisiana State University; Michael Mislove, Tulane University
  - Attendance: 43

  Lattices and ordered sets play an important role in many areas of computer science. These range from lattices as models for logics, which are fundamental to understanding computation, to the ordered sets as models for computation, to the role both lattices and ordered sets play in combinatorics, a fundamental aspect of computation. In addition, many applications utilize lattices and ordered sets in fundamental ways. These include such areas as knowledge representation, text categorization and data mining, where order plays a fundamental organizing principle, to the use of lattices and ordered sets to analyze crypto-protocols in security, to inductive logic programming, where ordered sets form basic models. A number of more esoteric structures, such as Kleene algebras and quantales, also feature in recent advances to understanding computation. This workshop will bring researchers from the many areas which rely on lattices and ordered sets together with those doing research in these and related structures, in order to accomplish two goals:

1) to survey the many areas of computation where lattices and ordered sets play a role in order to better understand the problems common across these areas, and

2) to enhance the interactions between researchers in the areas of lattice theory and ordered sets, and those who utilize these structures in modeling computation and in areas of application.

It is expected that researchers in lattice theory and ordered sets will benefit by discovering new and interesting problems where their expertise can be applied, and those who work on applications will have a chance to learn of new advances in these areas that could be applied to problems where lattice theory and ordered sets traditionally have played an important role. In addition to these goals, the intention also is to present talks which are accessible to the broader public that will help those working in related areas to understand the role lattice theory and ordered sets play in theoretical computation and its applications.

The workshop featured several plenary lectures, whose goal was to give a broad background to the workshop, These lectures were complimented by several shorter talks that focused on particular areas of application, as well as on recent advances in lattice theory and ordered sets. The talks by leaders in lattice theory and ordered sets focused on applications to computer science, while the talks by leading researchers in these areas described how techniques from lattice theory and ordered sets have recently led to results and applications to problems in computation.

Selected Research Results

Generalization of Standard Implementation of the Pairing Heaps. Rutgers graduate student Amr Elmsary generalized the standard implementation of the pairing heaps. He introduced a new parameter k. When the node with the minimum value is to be deleted from the heap, the operations to combine the resulting sun-trees into one tree depend on the value of k. When the value of k is equal to 2, the implementation will be equivalent to the standard pairing heaps' implementation. He showed that, for any constant k, this general form achieves the same bounds as the standard implementation. He conducted experimental results showing that, by tuning the value of k, the number of comparisons involved in this operation can be reduced.

Computing Topological Invariants of Semi-Algebraic sets. Computing various topological invariants of semi-algebraic sets in single exponential time is an active area of research. Several algorithms are known for deciding emptiness, computing the number of connected components of semi-algebraic sets in single exponential time etc. Visitors Saugata Basu, Richard Pollack and Marie-Francios Roy developed a new, improved algorithm for computing the Euler-Poincare characteristic (which is the alternating sum of the Betti numbers) of the realization of each realizable sign condition of a family of polynomials restricted to a real variety. The Euler-Poincare characteristic of a finite set of points is just the cardinality of the set, and the algorithm can be viewed as a direct generalization of the algorithm for counting the real roots of a univariate polynomial satisfying different sign conditions on another family of polynomials. A consequence of their result is that the Euler-Poincare characteristic of any locally closed semi-algebraic set can be computed with the same complexity. The best complexity of any previously known single exponential time algorithm for computing the Euler-Poincare characteristic of semi-algebraic sets worked only for a more restricted class of closed semi-algebraic sets and had a higher complexity.

Exact Free Energy Functional for a Driven Diffusive Open Stationary Non-equilibrium System. Joel Lebowitz (Rutgers) obtained the exact probability $\exp[-L\mathcal{F}(\{\rho(x)\})]$ of finding a macroscopic density profile $\rho(x)$ in the stationary non equilibrium state of an open driven diffusive system, when the size of the system $L \to \infty$. $\mathcal{F}$, which plays the role of a non equilibrium free energy, has a very different structure from that found in the purely diffusive case. As there, $\mathcal{F}$ is nonlocal, but the shocks and dynamic phase transitions of the driven system are reflected in non-convexity of $\mathcal{F}$, in discontinuities in its second derivatives, and in non-Gaussian fluctuations in the steady state.

Complexity of Monotone Generation. Endre Boros (Rutgers), Khalid Elbassioni (Visitor), Vladimir Gurvich (Visitor) and Leonid Khachiyan (Rutgers) studied computational complexity of the following class of problems: given a monotone increasing Boolean function defined by a polynomial oracle, generate all its minimal true and maximal false vectors. Examples of problems in this class include the following problems: given a hypergraph, generate all minimal vertex sets which contain at least t hyperedges (Problem I), or, generate all minimal edge-sets whose union contains at least t vertices (Problem II). Two other examples are as follows: given a system of linear inequalities $Ax \geq b$ where A is a nonnegative matrix, generate all minimal feasible vectors (Problem III), or all maximal infeasible vectors (Problem IV). They proved that Problems (I) and (IV) are intractable (NP-complete), while for problems (II) and (III) there exist efficient algorithms, suggested in their paper. This research was a starting point of a study, by the same group of authors, that developed into a coherent theory of complexity of monotone generation.

Quantum Query Complexity of Read-once Boolean Functions Howard Barnum (Visitor) and Michael Saks (Rutgers) studied quantum query complexity. In the query model of boolean computation, an n bit input string x is accessed by an algorithm via queries. In the classical deterministic version of the model (the boolean decision tree model), a single query is an index $i \in \{1, \ldots, n\}$ and the response to a query is the input bit x_i. In the quantum version of the model, a query may be viewed (roughly) as a superposition of deterministic queries i with complex amplitudes α_i, and the response to the query is the corresponding superposition of the bit values x_i. This model is fundamental within quantum computing; both Shor's factoring algorithm and Grover's search algorithm are best formulated within this model. The fundamental question for this model is: what advantage does quantum computation give over classical? One question along these lines has to do with the computation of total Boolean functions. Such a function has domain $\{0,1\}^n$ and has output $\{0,1\}$. Grover's algorithm shows that quantum computation can compute the n-bit OR function in $O(\sqrt{n})$ steps as compared to $\theta(n)$ required for a classical algorithm. An important conjecture (suggested by many researchers) is that quantum computation does not give better than a quadratic speedup for any total boolean function. Barnum and Saks proved this conjecture for the class of *read-once boolean functions* which are functions that can be expressed by a boolean formula in which each variable appears at most one. The proof uses a generalization of a lower bound technique proposed by Ambainis.

3.4 2001-2004 Special Focus on Computational Information Theory and Coding

Overview: Funded by two grants from NSF, the Special Focus on Computational Information Theory and Coding moved into high gear for FY2003.

Organizers:

- Robert Calderbank, (chair), AT&T Labs
- Chris Rose, Rutgers University
- Amin Shokrollahi, Digital Fountain
- Emina Soljanin (co-chair), Bell Labs
- Sergio Verdu, Princeton University

Sponsors and Collaborating Institutions:

The National Science Foundation
The New Jersey Commission on Science and Technology
Rutgers University
Princeton University
AT&T Labs - Research
Bell Labs
NEC Laboratories America
Telcordia Technologies
DIMACS Affiliate Members:
    Avaya Labs
    IBM Research
    Microsoft Research

Theme of Program:

The main purpose of coding theory is the reliable transmission or storage of data. It is common to think about information theory as a science concerned with theoretical limits on rates at which reliable transmission or storage of data is possible. During the course of the 50-year life of information theory, these limits (known as channel capacities) have been computed for various important telecommunications channels based on their underlying physics, which was considered given. Computational Information Theory is concerned with techniques (such as channel coding) for achieving channel capacities. Computational information theory has achieved dramatic scientific breakthroughs in recent years, and codes that come close to theoretical limits have been discovered.

This special focus will explore interconnections among coding theory, theoretical computer science, information theory, and related areas of computer science and mathematics, in order to address some of the challenges for the next 50 years - in particular wireless communication, magnetic/optical storage, the role of signal processing in networks, network information theory, and the creation of quantum information theory.

Today, the development of coding and information theory is closely related to the explosion of information technology, with applications to the Internet and the next generation of networks technologies. The rapid development of a myriad of networked devices for computing and telecommunications presents challenging and exciting new issues for coding and information theory.

In 1948 Shannon developed fundamental limits on the efficiency of communication over noisy channels. The coding theorem asserts that there are block codes with code rates arbitrarily close to channel capacity and probabilities of error arbitrarily close to zero. Fifty years later codes for the Gaussian channel have been discovered that come close to these fundamental limits. Today, theoretical limits on rates at which reliable transmission or storage of data is possible are known for many telecommunications challenges. The same types of challenges face us in today's and tomorrow's new eras of networked, distributed computing and communications.

Coding theory has long used deep mathematical methods. We will explore the use of such mathematical ideas from linear systems theory, automata theory, algebraic geometry, etc. in the understanding of issues at the interface among coding theory, information theory, and theoretical computer science.

There are many connections between coding theory and theoretical computer science. We will explore the connections between coding theory and: upper bounds on the number of random bits used by probabilistic algorithms and the communication complexity of cryptographic protocols; lower bounds on the complexity of approximating numerous combinatorial optimization functions; characterizations of traditional complexity classes such as NP and PSPACE; and the emerging theory of program testing.

Connections between information theory/coding and other parts of science, in particular physics, go back to the need to understand the relations between Shannon's perspective on entropy and the laws of thermodynamics. Today, attention has turned to quantum mechanics and the processing of intact quantum information states, and we will explore such connections.

Coding theory is essential in numerous applied areas such as deep space communication, the theory of wireless channels, and optical/magnetic recording. Such applied areas are posing new and challenging problems for coding theorists and will be a central focus of this special focus.

The development of the "next generation" of networks technologies leads to many new challenges for coding and information theory. Thinking of the network as a channel should help us to understand these challenges.

The time is right to explore in depth the interface among coding theory, theoretical computer science, information theory, and parts of theoretical mathematics. The new and exciting applied problems arising from new information technologies are calling for interdisciplinary approaches. And the theoretical problems have already shown themselves to be ripe for cross-fertilization among techniques and methods.

Workshops and Working Group Meetings

- **Workshop on Signal Processing for Wireless Transmission**
  - Dates: October 7 - 9, 2002
  - Location: Rutgers University
  - Organizer: Sergio Verdu, Princeton University; Jerry Foschini, Bell Labs
  - Attendance: 137

  In contrast to the voiceband telephone channel, the wireless channel suffers from interference from other users and from fading due to destructive addition due to multipath propagation. The aim of this workshop is to explore the ultimate limits that information theory puts on spectral efficiency, as well as the best means of striving toward that efficiency. Multiuser detection and "Dirty Paper Coding" are among the key signal processing countermeasures that promise substantial improvements over existing systems. This workshop investigated such approaches from both a link and network level perspective. The enhancements from multiple antennas were also explored, including means of space-time coding.

- **Workshop on Network Information Theory**
  - Dates: March 17 - 19, 2003
  - Location: Rutgers University
  - Organizers: Piyush Gupta, Gerhard Kramer, Adriaan van Wijngaarden, Bell Labs

- Attendance: 132

The DIMACS Workshop on Network Information Theory focused on the area of efficient and reliable communication in multi-terminal settings. This field has recently attracted renewed attention because of key developments that have spawned a rich set of challenging research problems. Applications such as wireless cellular and LAN data services, ad hoc networks and sensor networks should benefit from these developments. The aim of the workshop was to achieve a better understanding of the underlying information theoretic problems and their solutions.

The workshop topics included, but were not limited to, the following areas:

- Large communication networks: analysis, design, asymptotics
- Multi-terminal capacity/coding: relays, multi-access, random access
- Multi-terminal source coding: distributed sources, multiple descriptions
- Network coding: efficiency and reliability

The workshop consisted of 40-minute invited presentations and ample time for discussion.

- **Workshop on Complexity and Inference**
  - Dates: June 2 - 6, 2003
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizers: Mark Hansen, Bell Labs; Paul Vitanyi, University of Amsterdam; Bin Yu, University of California, Berkeley
  - Attendance: 73

The notion of algorithmic complexity was suggested independently by Kolmogorov, Chaitin, and Solomonoff in the 1960's. Both Kolmogorov and Chaitin introduced the concept as a way to formalize notions of entropy and randomness, building on results from theoretical computer science dealing with partial recursive functions. Independently, Solomonoff defined algorithmic complexity in the pursuit of universal priors for statistical inference. In recent years, Rissanen expanded the applicability of these ideas, employing well-established concepts from information theory to frame his principle of Minimum Description Length (MDL) for statistical inference and model selection.

Each of these lines of research has developed methods for describing data (through coding and compression, or by analogy with some formal computing device); and each of these lines has employed some concept of an efficient representation to guide statistical inference. In this workshop, we explored both the foundational aspects of complexity-based inference as well as applications of these ideas to challenging modeling problems. Participants were drawn from the fields of statistics, information and coding theory, machine learning, and complexity theory. Application areas include biology, information technologies, physics and psychology. The following specific topics were covered by the workshop:

- Kolmogorov complexity and inference
- MDL (MML) theory and applications
- Lossy compression and complexity theory
- Complexity and Bayesian methods
- Individual sequence/on-line prediction and predictive complexity

- Compression methods for clustering
- Machine learning and computational complexity
- Complexity and cognitive science
- Applications

- Princeton-Rutgers Seminar Series in Communications and Information Theory 2002 - 2003
  - Tuesday, October 1, 2002
    Speaker: P. R. Kumar, University of Illinois
    Title: Ad hoc wireless networks: Analysis, protocols, architecture, and towards convergence
    Time: 4:30-5:30pm
    Location: DIMACS Center, Room 431 CoRE Bldg., Rutgers University, Busch Campus, Piscataway, NJ
  - Thursday, October 3, 2002
    Speaker: Jacob Ziv, Professor of Electrical Engineering, Technion–Israel Institute of Technology: President, Israeli Academy of Sciences and Humanities
    Title: Classification with Finite Memory Revisited
    Time: 4:00 pm
    Location: Princeton University, Friend 101
  - Thursday, October 3, 2002
    Speaker: Shlomo Shamai (Shitz), Technion
    Title: On Information Theoretic Aspects of Multi-Cell Wireless Systems
    Time: 4:30-5:30pm
    Location: Princeton University, Friend 004
  - Thursday, November 7, 2002
    Speaker: Michael Luby, Digital Fountain
    Title: LT Codes
    Time: 2:30-3:30pm
    Location: Princeton University, Friend 101
  - Thursday, November 14, 2002
    Speaker: Upamanyu Madhow, University of California, Santa Barbara
    Title: Information-Theoretic Prescriptions for Outdoor Wireless Communication
    Time: 4:30-5:30pm
    Location: DIMACS Center, Room 431 CoRE Bldg., Rutgers University, Busch Campus, Piscataway, NJ
  - Thursday, November 21, 2002
    Speaker: Tsachy Weissman, Hewlett-Packard Labs
    Title: Universal Discrete Denoising: Known Channel
    Time: 4:30-5:30pm
    Location: Princeton University, Friend 101
  - Monday, March 3, 2003
    Speaker: Michael Honig, Northwestern University
    Title: Asymptotic Methods in Wireless Communications
    Time: 4:30-5:30pm
    Location: DIMACS Center, CoRE Bldg, Room 431, Rutgers University, Busch Campus, Piscataway, NJ

- Thursday, March 6, 2003
  Speaker: Babak Hassibi, California Institute of Technology
  Title: Some Asymptotic Results in Wireless Networks
  Time: 4:30-5:30pm
  Location: Princeton University, Friend 006
- Thursday, March 13, 2003
  Speaker: Robert J. McEliece, Professor of Electrical Engineering, California Institute of Technology
  Title: Belief Propagation on Partially Ordered Sets
  Time: 4:30-5:30 pm
  Location: Princeton University, Friend 006
- Friday, March 14, 2003
  Speaker: Venkat Anantharam, University of California, Berkeley
  Title: Message Passing Algorithms for Marginalization
  Time: 4:30-5:30pm
  Location: CoRE Bldg, 1st floor lecture hall, Rutgers University, Busch Campus, Piscataway, NJ
- Monday, April 14, 2003
  Speaker: David N. C. Tse, University of California, Berkeley
  Title: Diversity and Multiplexing: A Tradeoff in Wireless Systems
  Time: 1:30-2:30pm
  Location: DIMACS Center, Room 431 CoRE Bldg., Rutgers University, Busch Campus, Piscataway, NJ
- Thursday, April 24, 2003
  Speaker: Andrea Goldsmith, Stanford University
  Title: Capacity Limits of Wireless Channels with Multiple Antennas: Challenges, Insights, and New Mathematical Methods
  Time: 4:30-5:30pm
  Location: CAIP Seminar Room, Room 601 CoRE Bldg., Rutgers University, Busch Campus, Piscataway, NJ
- Thursday, May 1, 2003
  Speaker: Thomas Richardson, Flarion Technologies
  Title: Principles and Design of Iterative Coding Systems
  Time: 4:30-5:30pm
  Location: Princeton University, Friend 006

Selected Research Results

Efficient Codes for Digital Fingerprinting. Alexander Barg (Bell Labs and DIMACS) and Gregory Kabatiansky (Visitor, IPPI RAN, Moscow) studied efficient constructions of codes for digital fingerprinting. Relying on the ideas of their earlier work (DIMACS report 2001-52) they gave a first construction of codes with the identifiable parent property in which identification of a member of the coalition can be accomplished in time logarithmic in the total number of users of the system. Previously known constructions relied basically upon exhaustive search and thus had exponentially slower identification procedures.

Input-Output of Mutual Information of Multidimensional Channels. Slava Prelov (Visitor, Russian Academy of Sciences) and Sergio Verdu (PM, Princeton) obtained a formula for the second-order expansion of the input-output of mutual information of multidimensional channels as the signal-to-noise ratio goes to zero. While the additive noise is assumed to be Gaussian, very general classes of input and channel distributions were considered. As special cases, these channel models include fading channels, channels with random parameters and channels with almost Gaussian noise. When the channel is unknown at the receiver, the second term in the asymptotic expansion depends not only on the covariance matrix of the input signal but also on the fourth mixed moments of its components. The second-order asymptotics of mutual information finds application in the analysis of the bandwidth-power tradeoff achieved by specific (not necessarily optimum) input signaling in the wideband regime.

Lowering the Bias in Quantum Coin Flipping Protocols. Postdoc Ashwin Nayak and Peter Shor (AT&T Labs) studied bit-commitment-based quantum coin flipping protocols, proving a new bound of 1/16 for the bias of any such protocol. In the same work they analyzed a sequence of multi-round protocols that tries to overcome the drawbacks of previously-proposed protocols in order to lower the bias. By exhibiting an intricate cheating strategy they showed a lower bound of 1/4 on the bias and conjectured that this is the optimal bias for such protocols.

Performance of Expander Codes on a Binary Symmetric Channel. Alexander Barg (Bell Labs and DIMACS) and Gilles Zémor (Visitor, ENST-Paris) studied performance of expander codes on a binary symmetric channel. This research was initiated during a visit of Zémor to DIMACS in 2001 when he, together with Barg, showed that these codes reach channel capacity under iterative decoding, producing the first example of capacity-attaining linear-time decodable codes. This work was extended during a DIMACS visit of Zémor in August 2002 when they introduced another family of expander codes tailored for transmission at rates in the immediate neighborhood of capacity. For this region they estimated the decrease rate (the error exponent) of error probability of decoding for randomized ensembles codes. The resulting estimate gives a substantial improvement of previous results for expander codes and some other explicit code families.

A Fast Algorithm for Generating Pairs of Codes with Prescribed Pairwise Hamming Distances and Coincidences. Vince Grolmusz (Visitor, Eőtvős University, Hungary) developed a fast algorithm for generating pairs of q-ary codes with prescribed pairwise Hamming-distances and coincidences (for a letter $s \in \{0, 1, \ldots, q-1\}$, the number of s-coincidences between codewords a and b is the number of letters s in the same positions both in a and b). The method is a generalization of a method for constructing set-systems with prescribed intersection sizes, where only the case $q = 2$ and $s = 1$ was examined. He also generated codes with prescribed k-wise coincidences and Hamming-distances.

3.5 2001-2004 Special Focus on Data Analysis and Mining

Overview:

Beginning in a preliminary way in Spring 2001, the Special Focus on Data Analysis and Mining is co-sponsored by grants from the National Security Agency and NSF. This focus had a strong program in FY2003.

Organizers:

- Adam Buchsbaum, chair, AT&T Labs
- Pierre Hansen, co-chair, GERAD–University of Montreal
- Diane Lambert, Bell Labs

- David Madigan, Rutgers University
- Mario Szegedy, Rutgers University
- Yehuda Vardi, Rutgers University

Advisory Committee:

- Phipps Arabie, Rutgers University
- Endre Boros, Rutgers University
- J. Douglas Carroll, Rutgers Business
- Melvin Janowitz, Rutgers University
- Buck McMorris, Illinois Institute of Technology
- Daryl Pregibon, AT&T Labs
- Rafail Ostrovsky, Telcordia Technologies
- Larry Shepp, Rutgers University

Sponsors and Collaborating Institutions:

The New Jersey Commission on Science and Technology
The National Science Foundation
The National Security Agency
Rutgers University
Princeton University
AT&T Labs - Research
Bell Labs
NEC Laboratories America
Telcordia Technologies
DIMACS Affiliate Members:
    Avaya Labs
    IBM Research
    Microsoft Research

Theme of Program:

Theoretical and algorithmic approaches to data analysis have played a central role in the development of modern methods for handling data. Now, however, the massive amounts of data gathered in important modern applications ranging from the Internet to credit card fraud detection to astronomy and medicine have dramatically changed the requirements for algorithms and provide ample motivation for a great deal of new theoretical development. We need methods for data analysis and mining that scale to the huge volumes of data that we are getting and can expect to get in such applications. The special focus will be devoted to data analysis and data mining, with emphases on the development of theoretical and algorithmic approaches to the massive data-mining problems that we face today, on the increasingly abstract formulations and models of data mining questions that are being seen in current research, and on the connections between theoretical approaches and practical applications.

The emphasis of this special focus will be on unifying promising approaches to data analysis and data mining that come from many distinct communities of researchers. The topics of interest include methodologies and algorithms for data mining, including clustering, discriminant analysis, enumerative methods, and multidimensional scaling; the increasingly abstract formulations and models of data mining questions using logical methods, conceptual clustering, learning and discovery that are critical in data mining and in particular for automatic, intelligent decision making; and the special problems that arise from applications to such important areas as fraud and intrusion detection, web mining, medical and scientific databases, marketing, and natural language data.

The field of data analysis has a long history with roots in traditional statistical analysis and the development of artificial intelligence. Theoretical analysis of databases has allowed the precise formulation of questions about them and in turn, coupled with a strong effort in algorithms, has led to many powerful techniques for collecting, storing, consolidating, processing, correcting, and retrieving data, for learning, and for finding previously undiscovered patterns.

The emergence of new and powerful data collection technologies has led to the creation of massive amounts of data, often distributed, shared, partially unknown, or having specialized structures. Traditional data analysis tools are incapable of handling the sheer size and complexity of these gigantic data sets. There is a great need to develop new methods and algorithms that can handle these data sets. We need to develop theoretical underpinnings for managing and reasoning about data, and we need new tools for finding patterns or displaying useful summaries of the data.

Among the topics we emphasize in this special focus are the following:

- Massive quantities of disparate data can be stored on a loosely coupled infrastructure such as the World Wide Web, and so we need search and retrieval techniques for distributed, semi- or unstructured "databases."
- Increasingly, the types of data of interest are very different from the strings of text or numerical data toward which much of classical data structures and algorithms are oriented. Thus, we need new theoretical approaches to represent and work with data items of particular, complex, application-dependent form and to integrate data from multiple, heterogeneous sources.
- There is now an increased need for systems that can make intelligent, automatic decisions based on the analysis of data. It is crucial that we understand theoretically the behavior of systems that are permitted to act autonomously, in order to guarantee desired results.
- Most algorithms currently used in data mining do not scale well when applied to very large data sets, often because they rely on random access to the data sets, which scales only while the data sets fit entirely in relatively small main memories.
- Data mining requires dramatic new methods of pattern recognition, including clustering, classification, association, sequence discovery, and visualization.
- Because of the sheer quantity of the data arising in various applications or because of their urgency, it becomes infeasible to store the data in a central database for future access and, therefore, it becomes necessary to make computations involving the data, and decisions about the data (like what to keep), during an initial scan as the data "stream" by.

Workshops and Working Group Meetings

- **Workshop on Visualization and Data Mining**
  - Dates: October 24 - 25, 2002
  - Location: Rutgers University

- Organizers: Emden Gansner, AT&T; Michael Goodrich, Johns Hopkins University; Claudio Silva, AT&T; Roberto Tamassia, Brown University
- Attendance: 35

Visualization focuses on creating insightful visual representations of complex data. In the context of data mining, visualization provides another tool for creating abstraction, and discerning patterns and structure in seemingly undifferentiated data, allowing human perception to aggregate and analyze the data. By its nature, a visual approach can provide a visceral factor which can be compelling as well as merely aesthetic. In certain domains, such as biotechnology, where some data has a physical existence, visualization can provide a connection between this data and related abstract data, and the areas of information and scientific visualization overlap. The increasing analysis of hybrid data, along with the growing size of data sets, will require imaginative new models and techniques for visualization.

This workshop brought together people from the different research communities interested in the interplay of visualization and data mining and exploration, including people in scientific and information visualization, databases, information retrieval, and application areas. The goal was to help establish relationships which can be used to strengthen and foster collaboration across the different areas. The hope was to identify key research issues that need to be addressed, especially those related to visualization of large graphs and non-traditional graph models. The workshop addressed core algorithmic issues as well as aspects of modeling and the use of graphs in data exploration.

- **Workshop on Video Mining**
  - Dates: November 4 - 6, 2002
  - Location: Rutgers University
  - Organizers: Azriel Rosenfeld, Daniel DeMenthon, Dave Doermann, University of Maryland
  - Attendance: 60

Modern computer technology, together with the proliferation of broadcast channels and of video-based surveillance systems, has enabled us to produce vast amounts of both raw and processed video data. The potential uses of this data are many and varied. Monitoring and mining of the content of this already huge, rapidly growing mass of data calls for the development of major computational resources and the development of sophisticated video understanding techniques.

The applications and potential applications of video mining include:

- monitoring of (possibly remote) surveillance cameras in theft protection, fire protection, care of bedridden patients and young children
- automatic recognition of suspicious people in large crowds
- automatic checking to identify passengers entering an airplane, bus, or public building (to verify entry authorization)
- automatic quality control in manufacturing processes
- robotic vision
- retrieval of archived video clips to illustrate a news breaking story
- retrieval of suspicious activities in prerecorded video surveillance sequences

- browsing of DVD and set-top box recordings
- intelligent fast-forward techniques
- detection of multi-lingual scene text to determine origin of broadcast
- video search engines for web browsers
- ranking of video clips by relevance in web query results
- classification of videos into genres for search pruning
- detection of crowd patterns for mob control
- detection of traffic patterns for traffic understanding management

There are some pervasive challenges here. A huge amount of data must be efficiently stored. Once stored, there must exist efficient retrieval algorithms. It must also be possible to retrieve different kinds of data stored in possibly different formats. Some of the applications will require real-time response and action based on an incoming data stream.

The purpose of this workshop was to survey available and potential technologies for video monitoring and mining (and in general methods of fast and efficient content-based analysis of video streams) and to identify promising directions for research in this challenging area. Specific topics covered were the analysis of camera motion and scene activity; temporal segmentation; content-based classification, indexing, and retrieval; representation, browsing, and visualization.

We also investigated related mining problems having to do with audio mining, seismic data mining, and cross-modality mining. Audio mining applications of interest include speech recognition; seismic data mining applications include identification of potential new sources of oil and gas and detection of earthquakes and/or nuclear tests. Cross-modality issues arise for example in problems involving identification from both video and speech.

- **Working Group on Streaming Data Analysis and Mining II**
  - Dates: March 24 - 26, 2003
  - Location: Rutgers University
  - Organizers: Adam Buchsbaum, AT&T Labs and Rajeev Motwani, Stanford University
  - Attendance: 33

This was the second meeting of this working group. The use of the computer in scientific research and as an essential ingredient in commercial systems has led to the proliferation of massive amounts of data. Researchers in a myriad of fields face daunting computational problems in organizing and extracting useful information from these massive data sets. Because of the sheer quantity of the data arising in various applications or because of their urgency, it becomes infeasible to store the data in a central database for future access and, therefore, it becomes necessary to make computations involving the data, and decisions about the data (like what to keep), during an initial scan as the data "stream" by. This working group is concerned with datamining in a "streaming" environment.

Examples of applications requiring immediate decision making based on current information are intrusion detection and fault monitoring. Data must be analyzed as it arrives, not off-line after being stored in a central database, because the problems involved are so urgent from a time-to-react point of view. Some other applications require such quick reactions for theoretical (as well as practical) reasons because of the issues involved in processing and integrating the massive amounts of data generated by a myriad of continuously operating sources. For example, external memory algorithms are motivated by the fact that classical algorithms do not scale when data sets do not fit in main memory. At some point, data sets become so large as to preclude most computations that require more than one scan of the data, as they stream by.

Transactional and time-series applications exemplify current streaming data analysis systems. Transactional applications exploit data recording individual events correlating two or more discrete entities. Examples are phone calls between people (data also studied by the multidimensional scaling working group) and purchases over a credit-card network. One common problem is to maintain behavioral profiles of individual entities (customers, for example). Goals include flagging aberrant transactions, i.e., those not indicated by the models and thus potentially being fraudulent; and detecting paradigm shifts in prevailing trends. In these applications as well as others, analysis of data streams also engenders difficult new problems in data visualization. For example, how is time-critical information best displayed? Can automatic response systems be created to deal with common cases?

Time-series applications exploit sequences of unitary observations taken over time. Examples are reports from sensors monitoring network equipment; inventory levels of parts in a warehouse; positions of objects in successive celestial surveys; and records of prices of commodities, stocks, and bonds. Analyzing and mining time-series data presents many new challenges. What are similar time-series data? How can they be clustered, e.g., to isolate seminal events that cause many simultaneous or near-simultaneous disruptions among the observed elements? How can we find interesting trends?

We exploited recent attention to streaming models of data analysis by the theoretical community as well as recent successes in real-time or near-real-time analysis by practitioners. We brought together an interdisciplinary group of researchers to share their ideas and experiences, in the hopes of initiating new approaches to and motivating others to attack core problems in streaming data analysis.

This working group builds on past DIMACS activities and present DIMACS strengths in various areas related to streaming data analysis and mining. We make use of the DIMACS partnership with AT&T Labs to draw upon the expertise of a variety of researchers at that organization and of the systems that they have developed and we investigate whether these systems might apply to other applications or more general approaches to streaming data analysis. Analysis of data streams is used at AT&T to detect many types of fraudulent behavior; stolen calling cards and impending bill defaults are two examples. Similarly, the JAM system developed at Columbia detects fraudulent activity in financial systems. Gecko, developed by Hume, Daniels, and MacLellan, is a system that audits the largest billing process in AT&T. Using data feeds from passive taps throughout the system, Gecko tracks individual billing records from the network (originating phone call) through settlement (account collection). The main purpose is to detect abnormal processing of records which, if uncorrected, can lead to significant revenue loss. Many open problems remain in such work. For example, how can many heterogeneous data feeds be integrated in one system? How are relevant queries best expressed, giving the user sufficient power while implicitly restraining him/her from incurring unwanted computational overhead? Can domain-specific programming languages be used to provide uniform interfaces to disparate streaming datamining applications?

Critical to the proper functioning of today's massive computer and communications networks is real-time network monitoring. Network monitoring infrastructure creates a sequence of network events and alarms, the correlation of which can facilitate the location of root problems. This is an example of a time-series application. Analysis of data streams is used to detect network faults when they occur, in order to direct timely corrective action. Security trace audits of IP logs are used to detect and react to network attacks, e.g., denial of service attacks. Such an event might trigger some automatic monitoring system, which then prompts intense analysis of recent logs. Among the open problems in this area are the following. How are the results of such monitoring systems best reported and visualized? To what extent can they incur fast and safe automated responses?

The amount of data being collected through scanners at supermarkets and other retail outlets is awesome and marketing research faces the task of making use of these gigantic data sets. Of great interest in marketing is research on "market basket" models and in particular on what items tend to be bought concomitantly. This area of research is moving from off-line settings to more on-line scenarios. For example, performed hourly or even more frequently, such analysis can be used for "just-in-time" provisioning of markets. Recent on-line approaches to such market basket models are discussed in the papers by Ullman and Ganti, Gehrke, and Ramakrishnan.

The DIMACS "Special Year on Massive Data Sets" (1997-1999) featured a workshop on Astrophysics and Algorithms motivated by the huge data sets arising from sky surveys in optical and infrared wavelengths, microwave background anisotropy satellite experiments, helioseismology data, gravitational radiation detection experiments, and results from N-body/hydrodynamical simulations. That special year led to the beginnings of collaborations between computer scientists and astronomers, dealing with the "paradigm shift" in astronomy toward a situation where many researchers spend their time datamining a "digital sky" compiled from a vast array of multi-wavelength sky surveys. Current large scale cosmological simulations generate data of order 100 GB/per simulation. With the advent of higher bandwidth and faster computers, distributed data sets in the petabyte range are being collected. The problem of obtaining information quickly from such databases requires new and improved mathematical methods. Parallel computation and scaling issues are important areas of research. Techniques such as decision trees, vector-space methods, Bayesian and neural nets, and data compression have been utilized.

Financial markets provide time-series data, in particular, the prices of commodities, stocks, and bonds, as functions of time. Spotting and exploiting trends in such data is of great interest, e.g., to trading firms willing to risk capital and to companies that want to hedge against fluctuating international currencies. While DIMACS has not been heavily involved in financial modeling, we have organized a recent workshop on the topic and some of our members are actively involved in the field. Some of the general issues involving streaming data analysis for time series data, such as issues of similarity, clustering, and trend-spotting, present serious challenges in modeling of this kind of data and will provide another motivating application for our working group.

We gathered at DIMACS researchers and developers who work on streaming, for the purpose of exchanging ideas, problems, methods, and conjectures, and initiating joint research activities. We included theoreticians and practitioners from multiple disciplines (e.g., computer science, statistics, astrophysics). Some of the more specific issues addressed by the group have already been mentioned above. Among the broader questions addressed are the following: What is the scope of streaming data analysis? What are the relevant problems in data collection, transmission, processing, and visualization? How do current theoretical models for analyzing massive data streams correlate to methods used in practice? What are the core problems faced by practitioners, and how can theoretical insights be used to attack them? Our primary goals were to build bridges between various communities working on problems in streaming data analysis and mining; to initiate collaborative research; to develop a list of core problems that will motivate future work by the communities; and to produce an infrastructure for sharing results as we go forward.

- **Workshop: Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications**
  - Dates: May 14 - 16, 2003
  - Location: Rutgers University
  - Organizers: Regina Liu, Rutgers University; Yehuda Vardi, Rutgers; Robert Serfling, University of Texas at Dallas; Diane Souvaine, Tufts University
  - Attendance: 92

  Multivariate statistical methodology plays a role of ever increasing importance in real life applications, which typically entail a host of interrelated variables. Simple extensions of univariate statistics to the multivariate setting do not properly capture the higher-dimensional features of multivariate data, nor do they yield geometric solutions because of the absence of a natural order for multidimensional Euclidean space. A more promising approach is the one based on "data depth", which can provide a center-outward ordering of points in Euclidean space of any dimension. Extensive developments in recent years have generated many attractive depth-based tools for multivariate data analysis, with a wide range of applications. The diversity in approaches, emphases, and concepts, however, makes it necessary to seek unified views and perspectives that would guide the further development of the depth-based approach.

  The concept of data depth provides new perspectives to probabilistic as well as computational geometries. In particular, the development of implementable computing algorithms for depth-based statistics has brought about many new challenges in computational geometry. This workshop created a unique environment for multidisciplinary collaboration among computer scientists, theoretical and applied statisticians, and data analysts. It brought together active researchers in these fields to discuss significant open issues, establish perspective on applications, and set directions for further research.

- **Workshop on Management and Processing of Data Streams**
  - Dates: June 8, 2003
  - Location: San Diego, California
  - Organizers: S. Muthukrishnan, Rutgers University; Divesh Srivastava, AT&T Labs - Research
  - Attendance: 72

  This meeting was part of the Federated Computing Research Conference (FCRC'03). Measuring and monitoring complex, dynamic phenomena – atmospheric conditions, stellar patterns, movement of financial markets, traffic evolution in telephone and internet communication infrastructures, usage of world wide web, email and newsgroups – produces stream data, i.e., data that arrives as a series of "observations", often very rapidly. These observations generally have intricate relationships between them, are often grouped in multiple ways into logical events, and are naturally collected via distributed platforms.

  Databases and applications are increasingly required to manipulate stream data. This presents many challenges which are just beginning to be addressed.

  - Fundamental algorithmic models and results for data streams are being studied in the theoretical computer science community.

– Data stream management systems are being designed and prototyped in the database community.
– IP traffic analysis is a powerful application domain for data streams, and is being addressed in the networking community.
– Programming language and software support needed for managing data streams is being determined in the systems community.
– Custom and off-the-shelf hardware support is being developed for data streams in the networking and graphics communities.

FCRC'03 provided a valuable opportunity to engage this large and diverse community of interest in drawing up the specification of data stream management systems, understanding their power and limitations, as well as motivating the applications that will drive the development of data stream systems. This workshop provided a forum to discuss the progress on all aspects of managing stream data. The intended audience includes researchers, practitioners and graduate students interested in data stream management from the database (SIGMOD/PODS) community, as well as from other member FCRC conferences.

- **Workshop on Algorithms for Multidimensional Scaling II**
  – Dates: June 11-12, 2003
  – Location: Doubletree Hotel in Tallahassee, Florida
  – Organizers: J. Douglas Carroll (chairman), Rutgers University; Phipps Arabie, Rutgers University; Larry Hubert, University of Illinois; Michael Trosset, The College of William & Mary; Mike Brusco, Florida State University; Mel Janowitz, DIMACS
  – Attendance: 21

The term "datamining" has come to describe any type of data analysis, with an emphasis on the need to discern new, useful information from the large quantities of data stored in today's databases. Though the term "datamining" is relatively new, the fundamental ideas behind it are not, and the potential approaches to it are grounded in the basic theoretical and algorithmic approaches to data analysis developed over the last several decades. What is new is the challenge of modifying these approaches and developing new ones that can scale to the very large size of the problems that are faced or that apply to new types of data and in new types of applications. Most algorithms currently used in datamining do not scale well when applied to very large data sets, often because they rely on random access to the data sets, which scales only while the data sets fit entirely in relatively small main memories. Multidimensional scaling (MDS) is a case in point. The methodology has its roots in work going back to the first half of the 20th century and its modern roots in work of Shepard, Kruskal, and others. Yet, the traditional algorithms of MDS do not always work efficiently with today's large data sets and the traditional models of MDS are not always the most appropriate for the complex data types that we are seeing today. There has been in recent years a major trend toward the development of totally new and different algorithmic procedures for MDS as well as the modification of traditional procedures; an emphasis on the development of new models of MDS and on the development of new algorithmic procedures for fitting these new models; and a stress on the application of MDS to new types of data and new fields. These new algorithms, new models and new applications are being developed from different points of view and by researchers with different backgrounds and interests. This was the second meeting of our working group on Algorithms for Multidimensional Scaling. In this group, we are involving the developers of new models of, new algorithmic approaches to, and new applications of MDS to share ideas and explore new approaches. We are also including experts in the applications areas and individuals who have not necessarily worked on MDS before but are expert in methodologies such as combinatorial optimization that are increasingly relevant to MDS.

This meeting was held Immediately preceding the Classification Society of North America 2003 meeting, hosted by Mike Brusco at the Doubletree Hotel in Tallahassee, Florida.

Larry Hubert gave an invited talk on MDS. This talk was jointly sponsored by DIMACS and CSNA, and all CSNA attendees were welcome to attend.

One major outcome of this working group meeting is that several working group participants have begun to collaborate on MDS and molecular conformation. The idea is to take actual inner-atomic measurements of molecules, and to use MDS to construct a feasible model of the molecule. The result is that plans have been made to organize a small subgroup of the working group to meet DIMACS. That meeting will bring together chemists who are doing the measurements and mathematicians, computer scientists who are developing the computer algorithms. The working group meeting also led to the attempts to use the second-order sensitivity work discussed there to implement some specialized nonlinear optimization algorithms for the problem of reconstructing molecular structure from bound constrained nuclear magnetic resonance (NMR) data. Group members are also beginning to look more closely at Zhijun Wu's approach to the same sorts of NMR MDS problems, an approach presented at the meeting. Likewise, ideas presented by Suzanne Winsberg both at the meeting and afterward suggested that some of her work may be useful in getting good starting points for optimization, and the group is now investigating this direction, too.

Selected Research Results

Constructing the First Known Data Stream Algorithm for Estimating Dominance of Multiple Signals. Graham Cormode (visitor, University of Warwick) and S. Muthukrishnan (Rutgers) considered streams of multiple signals $(i, a_{i,j})$ where the i's correspond to the domain, the j's index the different signals and $a_{i,j} \geq 0$ gives the value of the j'th signal at point i. They addressed the problem of determining the dominance norms over the multiple signals, in particular the max-dominance norm defined as $\sum_i max_j\{a_{i,j}\}$. Besides finding many applications, such as in estimating the 'worst case influence' of multiple processes, for example in IP traffic analysis or electrical grid monitoring, this norm is a natural measure: it generalizes the notion of union of data streams and may be alternately thought of as estimating the L_1 norm of the upper envelope of multiple signals. Cormode and Muthukrishnan constructed the first known data stream algorithm for estimating max-dominance of multiple signals. The algorithm is simple and implementable; its analysis relies on using properties of stable random distributions with small parameter alpha, which may be a technique of independent interest. They also showed that other dominance norms – min-dominance $\sum_i min_j\{a_{i,j}\}$, count-dominance $(|\{i : a_i > b_i\}|)$ or relative-dominance $\sum_i a_i/max\{1, b_i\}$ – are all impossible to estimate accurately with sublinear space.

Space Lower Bounds for Distance Approximation in the Data Stream Model. Michael Saks (Rutgers) and Xiaodong Sun (graduate student, Rutgers) investigated the problem of approximating the distance between two d-dimensional vectors x and y in the classical data stream model. In this model, the $2d$ coordinates are presented as a "stream" of data in some arbitrary order, where each data item includes the index and value of some coordinate and a bit that identifies the vector (x or y) to which it belongs. The goal is to minimize the amount of memory needed to approximate the distance. For the case of L^p-distance with $p \in [1, 2]$, there were known good approximation algorithms that run in polylogarithmic space in d (here one assumes that each coordinate is an integer with $O(logd)$ bits). Saks and Sun proved that they do not exist for $p > 2$. In particular, they proved a nearly optimal approximation-space tradeoff of approximating L^∞ distance of two vectors. They showed that any randomized algorithm that approximates L^∞ distance of two length d vectors within factor of d^δ requires $\Omega(d^{1-4\delta})$ space. As a consequence they showed that for $p > 2/(1-4\delta)$, any randomized algorithm that approximates L^p-distance of two length d vectors within a factor d^δ requires $\Omega(d^{1-(4p-4)/\delta})$ space. The data stream model has been recognized recently as an important model, because it abstracts the problem of analyzing vast amounts of data "streaming" into a collection point. As for any model of computation, it is of interest to draw a dividing line between things that are efficiently computable in the model and things that are not. Previously other researchers had shown that approximating the distance of two vectors in this model could be done efficiently for standard Euclidean distance and L^1 (Manhattan) distance. These algorithmic techniques did not help for the case of L^∞ (max) distance. The work of Saks and Sun demonstrated that this latter problem is inherently unsolvable efficiently, providing a striking contrast with the other two cases. The lower bound follows from a lower bound on the two-party one-round communication complexity of this problem. This lower bound is proved using a combination of information theory and Fourier analysis.

Proving a Computer-generated Conjecture about the Separator of Fullerenes. It has long been a goal of artificial intelligence to design machines that can think like human beings. A small but quite remarkable research area related to this is concerned with designing machines that can aid in and eventually replace the process of scientific discovery by human beings. The first meeting of the DIMACS working group on Computer Generated Conjectures from Graph-theoretic and Chemical Databases gathered world experts in the use of computers to generate new scientific conjectures. During the working group meeting, a demonstration of the computer program 'Graffiti' led to the computer generation of several conjectures about the chemical structures known as fullerenes. Fullerenes are molecular graphs of carbon isomers. The separator of the fullerene is the difference between the two largest eigenvalues of the graph and is very useful in understanding the structure of these important compounds. 'Graffiti' generated the conjecture that the separator of fullerenes was always at most one, and then the conjecture that the separator is always at most $1 - 3/n$, where n is the number of vertices of the graph. The latter, stronger conjecture ended up being true and was proved during the working group meeting by two of the participants, Gilles Caporossi from Montreal and Dragan Stevanovic from Belgrade. The proof led to insights that later gave rise to the result that the dodecahedron has the largest separator of all fullerenes. It was fascinating to see the computer generate new scientific conjectures that captured the attention of leading researchers and led to new scientific knowledge.

Bounding the Irregularity of a Connected Graph: Proving Conjectures Generated by Computer. The program AutoGraphiX was one of the conjecture-generating programs studied by the DIMACS Working Group on Computer-Generated Conjectures from Graph-theoretic and Chemical Databases. In order to understand the ability of this program to find conjectures in an automated way, help to find further conjectures in an assisted way, and help to prove conjectures (or to conjecture proof strategies) the problem of the irregularity of a graph, defined by Albertson as $\sum_{i,j \in E} ||d_i - d_j||$ where d_j is the degree of vertex j, was studied by visitors Pierre Hansen and Hadrien Melot. An upper bound, as a function of order n and size m,inclusion for the irregularity of a connected graph was found, and it is best possible in the strong sense, i.e., attained by a graph for all n and m compatible with the existence of a connected graph.

Handling Distance Sensitivity Queries on a Weighted Directed Graph. Camil Demetrescu (Visitor, U-Rome) and Mikkel Thorup (AT&T) studied the problem of handling distance sensitivity queries on a weighted directed graph G. A typical example is a query of the form: "What is the shortest distance from vertex x to vertex y in G without using the edge (u, v) ?" This problem is relevant to link failure recovery in routing applications and to weight optimization procedures in network design. Demetrescu and Thorup showed how to preprocess a directed graph with n vertices and nonnegative real-valued edge weights, constructing an oracle that uses $O(n^2 \log n)$ space and answers distance sensitivity queries in $O(\log n)$ time for any vertices x, y and any given edge (u, v). They also showed that the query time can be reduced to $O(1)$ using $O(n^{2.5})$ space.

3.6 2002-2007 Special Focus on Computational and Mathematical Epidemiology

This 5-year program, starting in full in June 2002 (with some winter 2002 events), aimed at creating interdisciplinary partnerships, introducing mathematical scientists to the problems of epidemiology and to researchers working in the field, introducing epidemiologists and biologists to new mathematical and computational methods that might be relevant, and educating students and postdocs interested in the area. Our programs, modeled after the pioneering DIMACS Special Focus on Mathematical Support for Molecular Biology (1994-2000) and its follow-up Special Focus on Computational Molecular Biology (2000-2003), will be organized around mathematical models and methods, with particular emphasis on innovative new approaches based in discrete mathematics and theoretical computer science, combined with a blend of traditional approaches using statistical methods, differential equations, and stochastic process tools. We aim to catalyze partnerships and collaborations among mathematicians, computer scientists, and statisticians on the one hand and biologists, epidemiologists and clinicians on the other.

Organizing Committee:

Mathematics and Computer Science:

- Martin Farach-Colton, Rutgers University
- Mike Fredman, Rutgers University
- S. Muthukrishnan, Rutgers University
- Fred Roberts, Rutgers University

Statistics:

- Don Hoover, Rutgers University
- David Madigan, Rutgers University

Public Health and Epidemiology:

- Marc Lipsitch, Harvard School of Public Health
- Megan Murray, Harvard School of Public Health
- David Ozonoff, Boston University
- Burton Singer, Princeton University
- Daniel Wartenberg, University of Medicine and Dentistry of NJ

Biology:

- Sunetra Gupta, Oxford
- David Krakauer, Santa Fe Institute
- Simon Levin, Princeton University

Advisory Committee:

Ansuman Bagchi (Merck Research), Donald Burke (Johns Hopkins, International Health and Epidemiology), Carlos Castillo-Chavez (Cornell, Biometrics), Andy Dobson (Princeton, Ecology and Evolutionary Biology), Rick Durrett (Cornell, Math), Stephanie Forrest (U. New Mexico, CS), Chris Gilligan (Cambridge, Plant Sciences), John Glasser (CDC), Bryan Grenfell (Cambridge, Zoology), Harry Guess (Merck Research), Hans Heesterbeek (Wageningen, Center for Biometry), Herbert Hethcote (Iowa, Math.), Jennie Hunter-Cevera (University of Maryland Biotechnology Institute), Robert Horsburgh (Boston U., Epidemiology), Valerie Isham (Univ. Coll. London, Stat. Sci.), Sorin Istrail (Celera Genomics), Mel Janowitz (Rutgers, DIMACS), Denise Kirschner (Michigan, Med. Sch., Microbiology and Immunology), Natalia Komarova (Institute for Advanced Study), Joel Lebowitz (Rutgers, Math. and Physics), Ellis McKenzie (NIH), Martin Meltzer (CDC), Denis Mollison (Heriot-Watt, Actuarial Math. and Stat.), Mark Newman (Santa Fe Institute), Martin Nowak (Institute for Advanced Study, Theoretical Biology), Alan Perelson (Los Alamos), Pavel Pevzner (UC San Diego, Computer Science), William Shannon (Wash. U. Sch. of Med., Biostatistics in Medicine), Mona Singh (Princeton, CS), Eduardo Sontag (Rutgers, Math), Pauline van den Driessche (Univ. of Victoria, Math. and Stat.), Robert Vichnevetsky (Rutgers, CS), Shibu Yooseph (Celera Genomics).

Sponsors and Collaborating Institutions:

Office of Naval Research (ONR)
Alfred P. Sloan Foundation
Burroughs-Wellcome Fund
The New Jersey Commission on Science and Technology
The National Science Foundation
The National Security Agency
Rutgers University
Princeton University
AT&T Labs - Research
Bell Labs
NEC Laboratories America
Telcordia Technologies
DIMACS Affiliate Members:
  Avaya Labs
  IBM Research
  Microsoft Research

Theme of the Program:

Epidemic models of infectious diseases go back to Daniel Bernoulli's mathematical analysis of smallpox in 1760 and have been developed extensively since the early 1900s. Hundreds of mathematical models have been published since, exploring the effects of bacterial, parasitic, and viral pathogens on human populations. The results have highlighted and formalized such concepts as the notion of a core population in sexually transmitted diseases and made explicit other concepts such as herd immunity for vaccination policies. Relating to persistent infections, key pathogens that have been studied are: Malaria, Neisseria gonorrheae, M. tuberculosis, HIV, and T. palladum. Important issues such as drug-resistance, rate of spread of infection, epidemic trends, and the effects of treatment and vaccination all have been addressed through mathematical modeling approaches, which with the help of computational tools have provided new insights. Yet, for many infectious diseases, we are far from understanding the mechanisms of disease dynamics. The strength of the modeling process is that it can lend insight and clarification to existing data and theories. Mathematical models provide a unique approach to representing and studying the integrated behavior of complex biological systems and enable us to compare and contrast existing theories of the dynamic interactions in a complex system. The size of modern epidemiological problems and the large data sets that arise call out for the use of powerful computational methods for studying these large models. As pointed out by Levin, Grenfell, Hastings, and Perelson in a 1997 article in Science, "imaginative and efficient computational approaches are essential in dealing with the overwhelming complexity of [such] biological systems." New computational methods are need to deal with the dynamics of multiple interacting strains of viruses though the construction and simulation of dynamic models, the problems of spatial spread of disease through pattern analysis and simulation, and the optimization of drug design through hierarchical and other search methods on adaptive landscapes.

Statistical methods have long been used in mainstream epidemiology largely for the purpose of evaluating the role of chance and confounding associations. Considerable effort is expended by epidemiologists to ferret out sources of systematic error ("bias and confounding") in the observations and to evaluate the role of uncontrollable error (using statistical methods) in producing the results. Interpretation of the results usually depends upon correlative information from the medical and biological sciences. The role of statistical methods in epidemiology is changing due to the large data sets that are arising and this calls for new methods and new approaches, making use of modern information technology for dealing with huge data sets of information on disease patterns.

A smaller but venerable tradition within epidemiology has considered the spread of infectious disease as a dynamical system and applied difference equations and differential equations to that end. But little systematic effort has been made to apply today's powerful computational methods to these dynamical systems models and few computer scientists have been involved in the process. We hope to change this situation. Probabilistic methods, in particular stochastic processes, have also played an important role. However, here again, few computer scientists have been involved in efforts to bring the power of modern computational methods to bear.

A variety of other potentially useful approaches to epidemiological issues have not yet attracted the attention of many in the computer science community nor are the methods made widely available to biological scientists. For example, many fields of science, and in particular molecular biology, have made extensive use of the methods and techniques of discrete mathematics (broadly defined), especially those that exploit the power of modern computational tools. These are guided by the algorithmic and modeling methods of theoretical computer science that make these tools more available than they have been in the past. Yet, these methods remain largely unused in epidemiology. One major development in epidemiology that makes the tools of discrete mathematics and theoretical computer science especially relevant is the use of Geographic Information Systems (GIS). These systems allow analytic approaches to spatial information not used previously. Another development is the availability of large and disparate computerized databases on subjects containing information on many attributes that might be related to disease status.

The role of discrete mathematics and theoretical computer science has also become important with the increasing emphasis in epidemiology of an evolutionary point of view. To fully understand issues such as immune responses of hosts; co-evolution of hosts, parasites, and vectors; drug response; and antibiotic resistance; among others, biologists are increasingly taking approaches that model the impact of mutation, selection, population structure, selective breeding, and genetic drift on the evolution of infectious organisms and their various hosts. Epidemiologists are only beginning to become aware of some of the computer science tools available to analyze these complex problems, such as methods of classification and phylogenetic tree reconstruction grounded in concepts and algorithms of discrete mathematics and theoretical computer science and developed in connection with the explosion in "computational biology," a field in which DIMACS has been a pioneer. Many of the recent methods of phylogenetic tree reconstruction resulted from the DIMACS Special Focus on Mathematical Support for Molecular Biology are described in the DIMACS website in the reports on the accomplishments of the earlier Special Focus. Yet, a great deal more needs to be done.

One important modern topic in theoretical computer science that arose in epidemiology is the theory of group testing, which arose in connection with testing millions of World War II military draftees for syphilis. The idea is to avoid testing each individual and instead to divide them into groups and determine if some individual in the group is positive for the disease, updating the process with groups that test positive. The modern theory of group testing is heavily influenced by combinatorial methods, in particular by the methods of combinatorial designs and coding theory, and many modern algorithmic methods, developed by theoretical computer scientists, are not yet widely known or used in epidemiology.

Mathematical methods of formal logic and ordered algebraic systems have been used to develop the foundations for a theory of measurement with important uses in the physical sciences and, more recently, in the social and biological. While this kind of measurement theory has been applied to data analysis in the social and biological sciences, it is virtually unknown in the epidemiology community (where the term "measurement theory" has other connotations), except to the extent that epidemiological studies use principles, grounded in but sometimes challenged by measurement theory, such as that certain kinds of statistical tests are inappropriate for ordinal data.

New interdisciplinary approaches, involving partnerships among mathematical scientists and biological scientists, epidemiologists, and clinicians, offer promise for making progress on modern epidemiological problems and should take both fields of epidemiology and mathematics/computer science in new and fruitful directions. Mathematical and computational methods seem especially relevant in light of recent modeling approaches to emerging infectious diseases such as the vector-borne diseases from West Nile virus, Eastern equine encephalitis virus and Borrelia burgdorfei (Lyme disease); the spread of "mad cow" disease (transmissible spongiform encephalopathy; and HIV/AIDS. Control measures for these diseases often have counter-intuitive consequences only revealed after sophisticated mathematical analysis.

Similar advances as a result of applications of mathematical and computational modeling have not been as evident in the area of chronic disease epidemiology, although work of considerable promise is being done, for example on modeling of the progression of cancer. In this special focus, we will consider both infectious and non-infectious diseases, and we will explore mathematical and computational approaches to animal and plant diseases as well as to human diseases.

Following the events of September 11, 2001 and thereafter, increasing emphasis on bioterrorist threats have made this special focus especially relevant and also stimulated us to modify the focus accordingly to bring into it a significant emphasis on bioterrorist defense.

Tutorials, Workshops and Working Group Meetings

- **International Conference on Computational and Mathematical Epidemiology**
  - Dates: June 28 - July 2, 2002

- Location: DIMACS Center, Rutgers University
- Organizers: Simon Levin, Princeton University, Fred S. Roberts, Rutgers University
- Attendance: 135

Interest in infectious diseases has increased greatly in recent years as new diseases such as Lyme disease, HIV/AIDS, Hepatitis C, Hantavirus, and West Nile virus have emerged, antibiotic-resistant strains of tuberculosis, pneumonia, and gonorrhea have evolved, and concern has increased over the deliberate introduction of diseases such as anthrax, smallpox or plague by bioterrorists. Mathematical methods have become important tools in analyzing the spread and control of infectious diseases as well as noninfectious ones. Because we believe that partnerships between mathematical scientists (mathematicians, statisticians, computer scientists) and epidemiologists and public health professionals can make important new contributions to the usefulness of such methods, DIMACS launched a five-year "Special Focus on Computational and Mathematical Epidemiology." To "launch" this major program, we held an International Conference on Computational and Mathematical Epidemiology at DIMACS from June 28 to July 2, 2002.

Talks at the conference covered a wide variety of topics relating to modeling of infectious and noninfectious diseases, including:

- mathematical methods for defense against bioterrorist attacks
- vaccination strategies for controlling diseases (deliberately spread or otherwise)
- biosurveillance systems
- analogies between computer viruses and biological viruses
- predictive methodologies for infectious diseases
- spatio-temporal and network modeling of diseases
- emerging infections and biological invasions
- spatially explicit modeling
- evolution of pathogens
- computational biology as a modeling tool in epidemiology
- phylogenetic tree reconstruction methods for rapidly evolving pathogens
- public health modeling
- mathematical modeling of influenza, smallpox, measles, and other diseases
- causal modeling and public health
- challenges from marine diseases
- combinatorial group testing (combinatorial methods for efficiently testing large groups of people for presence of a pathogen)
- ecologic inference

Keynote addresses were given by Rita Colwell, Director of the National Science Foundation, and Nancy Cox, Chief of the Influenza Branch at the Centers for Disease Control.

In addition to the keynote addresses, other very distinguished invited speakers were:

David Banks, FDA
Lora Billings, Montclair State University

Donald Burke, Johns Hopkins University
Carlos Castillo-Chavez, Cornell
Freddy Christiansen, University of Aarhus, Denmark
Christl Donnelly, Imperial College
Ding-zu Du, University of Minnesota
Jonathan Dushoff, Princeton University
Chris Gilligan, University of Cambridge
John Glasser, CDC
Harry Guess, Merck
Herbert Hethcote, University of Iowa
Donald Hoover, Rutgers University
Valerie Isham, University College, London
Edward Kaplan, Yale University
Simon Levin, Princeton University
Marc Lipsitch, Harvard
David Madigan, Rutgers
Ellis McKenzie, NIH
Martin I. Meltzer, CDC
Denis Mollison, Heriot-Watt
Farzad Mostashari, NYC Department of Health
David Ozonoff, Boston University
Marcello Pagano, Harvard University
Alan Perelson, Los Alamos
Joshua Plotkin, Institute for Advanced Study
Fred Roberts, DIMACS, Rutgers University
James Robins, Harvard
Burton Singer, Princeton University
Mike Steel, University of Canterbury, New Zealand
Daniel Wartenberg, UMDNJ-Robert Wood Johnson Medical School

The audience at the conference consisted of mathematical scientists, epidemiologists, and public health professionals. Speakers presented talks that were understandable to non-experts and had a large tutorial/expository component, while at the same time introducing the audience to current issues of importance in research.

One of the major goals of this conference, as well as of the special focus that it launches, was to involve more non-experts in the battle against natural and man-made attacks by infectious diseases. Accordingly, the conference was preceded by a tutorial on Dynamic Models of Infectious Diseases, aimed at introducing those who have not worked on modeling in epidemiological to one important set of tools.

- **Tutorial: Epidemiology for Mathematical Scientists Part I: Introduction to Epidemiological Studies**
  - Dates: August 26 - 30, 2002
  - Location: Rutgers University
  - Organizers: David Ozonoff, Boston University, School of Public Health and Daniel Wartenberg, UMDNJ-Robert Wood Johnson Medical School
  - Attendance: 26

This tutorial introduced participants to the language, concepts and methods of epidemiology, with the objective of allowing mathematicians to communicate with epidemiologists, to understand important open questions and difficult technical problems in epidemiology, and to stimulate their own ideas about how areas of mathematics that have to this point not been applied to epidemiology might be brought to bear. The emphasis was not primarily statistical, although the tutorial discussed statistics to provide some background on how mathematics is currently used in epidemiology.

Special tutorial material was developed specifically to introduce mathematicians to the terminology and concepts used by epidemiologists to investigate patterns of disease occurrence in populations. Emphasized were the types of study designs, measures of association and types of systematic and random error that concern current epidemiological practice. Open questions and practical problems that currently do not have good solutions were also emphasized.

- **Tutorial: Epidemiology for Mathematical Scientists Part II: The Foundations of Molecular Genetics for Non-Biologists**
  - Dates: August 26 - 30, 2002
  - Location: Rutgers University
  - Organizer: William Sofer, Rutgers University
  - Attendance: 32

  The lectures introduced the basics of molecular biology and genetics to those who had only been exposed to classical biology (or who have had no training in biology at all). Topics covered included: polymers, DNA, RNA, proteins, transcription, translation, DNA replication, genetics, genetic engineering, the molecular biology of pathogenesis.

- **Workshop: The Pathogenesis of Infectious Diseases: Host-Pathogen Dynamics**
  - Dates: September 23 - 25, 2002
  - Location: Rutgers University
  - Organizers: Denise Kirschner, University of Michigan and Alan Perelson, Los Alamos
  - Attendance: 53

  Epidemic models of infectious diseases have been developed since the middle of the 1900's. Hundreds of mathematical models have explored the effects of both bacterial and viral pathogens on subgroups of human populations and this workshop will provide an overview. After decades of research focusing on infected patients and experimental animals, most modern research in microbial pathogenesis takes place at the level of cellular and biochemical mechanisms governing host-parasite interaction; however, studies at multiple scales will undoubtedly be needed for a deeper understanding of infectious diseases. For example, linking pathogen-specific information to that regarding the immune system will be critical for understanding the dynamics of most bacterial infections. Components of host-pathogen systems are so numerous and their interactions so complex that intuition alone is insufficient to fully understand the dynamics of the interactions. Mathematical modeling becomes an important experimental tool. Recently, models of host interactions with microbes have begun to appear, including models that explore bacteria-host level interactions. These models have studied, for example, antimicrobial chemotherapy, urinary tract bacterial infections, mycoparasite-immune dynamics, tuberculosis. Models of persistent viral infections, namely HIV-host models, also have a successful recent history. Many of the key results that have shaped our recent understanding of the T-cell and viral dynamics in HIV disease have come from mathematical modeling approaches. Many others have given insight into HIV-immune dynamics as well as disease progression. It is clear that models of within-host events are tied to population trends in diseases as well. A new line of research aims to link information obtained at the host level to predict both prevalence and incidence of disease at the population level. This workshop focussed on this line of research.

- **Working Group Meeting: Adverse Event/Disease Reporting, Surveillance and Analysis (First Meeting)**
  - Dates: October 16 - 18, 2002
  - Location: Rutgers University
  - Organizers: Donald Hoover, and David Madigan, Rutgers University; Henry Rolka, (CDC)
  - Attendance: 52

  Disease or event reporting and surveillance systems represent a primary epidemiological data source for the study of/alert to adverse reactions to medication, emerging diseases, or bioterrorist attacks. These systems synthesize data from millions of reports. The working group brought together pharmacoepidemiologists, statisticians and computer scientists to investigate current major issues confronting adverse event/disease reporting, surveillance, and analysis. We describe the issues for drug reaction reports; those for disease or symptom reports are similar though they raise their own set of issues as well. In the US, the two major reporting/surveillance systems are AERS and VAERS. AERS, the Adverse Event Reporting System (http://www.fda.gov/cder/aers/) is a data base of drug adverse reactions reported by health professionals and others. AERS is administered by the Food and Drug Administration (FDA). The system contains adverse reactions detected and reported after marketing of a drug for a specified time period. AERS contains over two million cases. VAERS, the Vaccine Adverse Event Reporting System (http://www.vaers.org), is a Cooperative Program for Vaccine Safety of the FDA and the Centers for Disease Control and Prevention (CDC). VAERS is a post-marketing safety surveillance program, collecting information about adverse events that occur after the administration of US licensed vaccines. Reports are provided by all concerned individuals: patients, parents, health care providers, pharmacists, and vaccine manufacturers. The VAERS database is publicly available and contains over 100,000 reports. Analyses of AERS and VAERS data must confront several difficulties including adverse event recognition, underreporting, biases, estimation of population exposure, report quality, and, most importantly, no denominator or control group of persons not taking the medication. In many cases it is difficult to discern whether or not a reported adverse reaction was from the medication or instead was a consequence of the underlying conditions that necessitated the medication (see, for example, [Koch-Weser, Sellers, and Zacest (1977), Rawlins (1994), Strom and Tugwell (1990)]). Several methodological issues relating to these reporting mechanisms and others emphasizing disease/symptom reporting will form the agenda for this working group. These include application of computational and statistical methods for early detection of emerging trends; modification of algorithms of streaming data analysis designed to set off early warning alarms; application of data mining methods; development of causal inferential methods in the absence of controls; study of ways to eliminate bias; design of verification methodology. This last issue is especially pressing since large-scale medical record databases that now exist in certain sub-populations (e.g., HMOs, military) can provide a basis both for assessing the quality of AERS and VAERS data and for validating analyses. Still another set of research issues for the working group arises from the use of natural language in reporting systems: Devise effective methods for translating natural language input into formats suitable for statistical analysis (prior work on machine natural language processing and information retrieval is relevant); develop computationally efficient methods to provide automated responses consisting of follow-up questions; develop semi-automatic systems to generate queries based on dynamically changing data, indicating developing epidemiological trends. Relevant to these questions is work in [VanLehn and Niu (to appear)] on interpreting natural language reports based on probabilistic models of context, work in [Langlotz, Shortliffe and Fagan (1986), McConachy, Korb and Zuckerman (1998), Zuckerman, McConachy and Korb (1998)] on communicating uncertain information and summarizing rough trends, and work in [Horvitz and Paek (1999), Horvitz and Paek (2001), Walker (2000)] on decision-theoretic methods for asking followup questions in natural language processing. Earlier work on electronic surveillance reporting from public health reference laboratories is very relevant here (e.g., [Bean and Martin (2001), Bean, Martin and Bradford (1992), Hutwagner, Maloney, Bean, Slutsker and Martin (1997), Mahon, Rohn, Pack and Tauxe (1995)]). Subgroups might be formed to concentrate on drug reactions, emerging diseases, or bioterrorism.

- **Working Group Meeting: Analogies Between Computer Viruses and Immune Systems and Biological Viruses and Immune Systems**
  - Dates: Subgroup Meeting, December 14-16, 2002
  - Location: Rutgers University
  - Organizers: Lora Billings, Montclair State University; Stephanie Forrest, U. New Mexico; Alun Lloyd, Institute for Advanced Study; Ira Schwartz, Naval Research Laboratory
  - Attendance: 4

  This was a working meeting of a small subgroup of a larger working group. The main objective of the working group is to identify, by mathematical analysis, new modeling techniques for describing quantitatively the spread of computer virus attacks and possible defensive strategies. The main themes are the creation of new models to account for the analogies and differences between the spread of computer and biological viruses, identification and metrics of virus spread, and discussions about control and constraints to preserve information flow.

  Current approaches to the detection and control of computer virus spread were discussed including: information flow modeling, operational network modeling, examination of real world virus data, and validation of models against available data. Part of the problem in trying to assess model validity is that data availability is not widespread and incomplete. Network data of virus spread is inherently non-stationary and spatio-temporal in nature. Unlike well-formed physics experiments, a computer virus attack is transient, making control difficult. The new models are meant to aid in the development of early prediction methods, detection strategies, and optimal defensive measures. Of course, all defensive strategies are limited by constraints, whether they are human resources, time, money, or privacy. These factors were also considered during evaluation.

  Computer virus attacks remain a serious threat to the national and international information infrastructure, and may be analyzed through mathematical and computational models. Predicting virus outbreaks is extremely difficult due to the human nature of the attacks, but more importantly, detecting outbreaks early with a low probability of false alarms seems quite difficult. Nonetheless, by developing models based on better data collection, it is possible to characterize essential properties of the attacks from a network topology viewpoint and this is one of the approaches taken by this working group. More resilient networks may be designed to control attacks, while retaining information flow on the networks. Novel methods of control, both static and adaptive, can be designed to slow the spread of an attack, while other means are designed to eradicate the virus from the network. Since the dynamics are based on non-stationary networks, the problem is one of a truly complex spatio-temporal nature. Most of the epidemiological analysis to date on networks has been static, and ignores the time dependence. Given such a complex dynamic, it is clear that mathematical and computer modeling will play a role for very long time and this group is building such models.

- **Working Group Meeting: Spatio-Temporal and Network Modeling of Diseases (First Meeting)**
  - Dates: First Meeting, April 22 - 26, 2003
  - Location: Rutgers University
  - Organizers: Bryan Grenfell, Cambridge; Valerie Isham, University College London; Matthew Keeling, Cambridge; Alun Lloyd, Institute for Advanced Study; Denis Mollison, Heriot-Watt University
  - Attendance: 26

This working group studies issues involved in mathematical modeling of the spread of infectious diseases among human, animal, and plant populations, using models with explicit time and space parameters and models based on networks of individuals. We investigate the use of novel computational methods, emphasizing the use of mathematical models to gain understanding of transmission mechanisms and to determine and evaluate control strategies. The group explores a variety of computer-intensive stochastic spatio-temporal models, including compartment models that divide the population into locally well-mixed populations, continuum models that assume individuals are located at points in a continuous plane, and network models that locate individuals at the vertices of a graph (see, e.g., [Dieckmann, Law and Metz, (eds.) (2000), McGlade, (ed.) (1999)]). We will review connections with current research on interacting particle systems and probabilistic cellular automata (see, e.g., [Levin and Durrett (1997), Mollison (ed.) (1995), Rand Keeling and Wilson (1995)]) and examine statistical modeling issues arising in the collection and analysis of spatial data, such as in the use of Markov chain Monte Carlo methods to fit models to data (see, e.g., [Gilks, Richardson and Spiegelhalter (1995), O'Neill, Balding, Becker, Eerola and Mollison (2000)]). We will develop computational models of spread of disease on networks (e.g., work on STD's and AIDS in [Altmann, M. (1995), Altmann, Wee, Willard, Peterson and Gatewood (1994), Peterson, Willard, Altmann, Gatewood and Davidson (1990)]) and analytical models for epidemics on networks, which are in their infancy (see, e.g., [Altmann (1995), Ball, Mollison, and Scalia-Tomba (1997)]). These models provide major computational challenges in dealing with the complexity of networks [Kretzschmar and Morris (1996), Levin, Grenfell, Hastings and Perelson (1997), Morris and Dean (1994)]. As computers become more powerful, we are able to simulate more and more complex biological interactions. This is especially true of stochastic processes in spatial environments or on large networks. Recently, techniques such as pairwise approximations and moment closure techniques have been used to mimic the effects of space and stochasticity with remarkably good results [Keeling (2000)]. Many of these techniques have been developed in isolation, and essentially none with the aid of computer scientists. One of the aims of the working group will be to connect the many researchers in this field with computer scientists, with the hope of producing a coherent framework. The group will also investigate a suite of models for spatially-extended dynamics that goes across heterogeneous scales, includes periodic forcing (due to seasonal change), involves temporally disturbed environments (due to abrupt changes in the presence of favorable periods for transmission), and deals with changes in host characteristics with age. Among the specific research challenges facing this working group are: synthesizing available epidemiological data and using it to validate and parameterize models; learning how models, particularly stochastic models, scale from small populations to large ones; and applying models to policy questions about endemic and emerging diseases. As noted in the discussion of the Data Mining working group, data about important diseases usually comes in many different forms, from spatially extensive data with few details, such as mortality records or livestock destruction reports, to detailed information from specific places and times. Combining widely varying sources of information is a challenging area requiring both theoretical and computational expertise. Examples of diseases for which large amounts of spatio-temporal data of varying kinds are available include tuberculosis, malaria and HIV, as well as newly emerging or re-emerging diseases illustrated by the recent outbreak of foot and mouth disease in Great Britain, the Four Corners outbreak of Hantavirus, and many other human, animal and plant diseases (see, e.g., [Keeling and Gilligan (2000)]). We explore analytic and simulation techniques for integrating such data from different sources and these should be valuable in assessing the effects of past interventions and thus in refining models used to answer policy questions. We consider theoretical questions about which details are important in scaling models up to large populations and these should be critical in developing workable, large-scale models. The agenda of this working group is sufficiently broad that it will probably make sense to break it into several groups after awhile.

- **DIMACS Working Group on Data Mining and Epidemiology (First Meeting)**
  - Dates: First meeting, May 22 - 23, 2003
  - Location: Rutgers University
  - Organizers: Ilya Muchnik, Rutgers University; S. Muthukrishnan, AT&T Labs - Research and Rutgers University; David Ozonoff, Boston University
  - Attendance: 23

New computational methods are needed to deal with large, complex data sets arising in epidemiology. In this working group, epidemiologists will join with computer scientists and statisticians to explore new uses of data mining methods in epidemiology. Notifiable infectious diseases provide a huge testbed of data for surveillance, for planning and evaluation of intervention programs, and for hypothesis generation in etiologic studies [Levin, Grenfell, Hastings and Perelson (1997)]. Data sets connecting environmental factors and disease can be used in much the same way [Paulu and Ozonoff (1998), Vieira, Webster, Aschengrau and Ozonoff (2001)]. Some examples of large data sets arising in epidemiological studies involve fat-free body mass [Chagnon, Y.C., Borecki, Prusse, Roy, Lacaille, Chagnon, M., Ho-Kim, Rice, Province, Rao and Bouchard (2000)], linkage scans [Province and Single (2000)], the health of Gulf War veterans [Proctor, Heeren, White, Wolfe, Borgos, Davis, Pepper, Clapp, Sutker, Vasterling and Ozonoff (1998)] and alcoholism [Saccone, Kwon, Corbett, Goate, Rochberg, Edenberg, Foroud, Li, T., Begleiter, Reich and Rice (2000)]. A long list of links to large health-related data sets can be found at the website http://www.ehdp.com/vitalnet/datasets.htm. Such data, while often massive in quantity, is uneven in quality and completeness and heterogeneous in nature. This group will build on relevant data mining studies in the epidemiological literature (see e.g., [Brossette, Jones, Sprague, Hardin and Moser (1999), Brossette, Sprague, Hardin, Waites, Jones and Moser (1998), DuMouchel (1999), Forgionne, Gangopadhyay and Adya (2000), Holmes, Durbin and Winston (2000), Openshaw, Turton and MacGill (1999), Pendharkar, Rodger, Yaverbaum, Herman and Benner (1999), Richards, Rayward-Smith, Sonksen, Carey and Weng (2001)]). We will emphasize the development of new algorithmic methods for data mining in epidemiology involving visualization, clustering, and aggregation.

Automatic environmental monitoring and risk evaluation for cancer provides a sample motivation. In the US, there is a rich history of cancer mapping, highlighted by the release of the first US Cancer mortality atlas in 1975, the recent development of the US Cause of Death Atlas, and the National Cancer Institute's data set of about 10 million US cancer cases. Additional data comes from questionnaires based on individual patient and resident information; public registries with cancer incidences aggregated by county; population-based cancer registries aggregated by city and town; birth and death registries; environmental data such as sample databases of water conditions and air quality records; census data such as geographic databases with accurate locations of population; and remotely-sensed data providing information on land use patterns or air pollution distribution. All of these databases have different temporal and spatial assumptions (for example, different frequencies of collection, different spatial resolution (by state, by county, by zip-code, by square kilometer), etc. Cluster analysis offers the promise of pattern extraction from such complex data. Our approach to clustering will start by emphasizing data cleaning tools (see, e.g., [Galhardas, Florescu, Shasha and Simon, E. (2000), Galhardas, Florescu, Shasha and Simon, E. (2000), Lambert, Pinheiro and Sun (1999)] and the website http://www.research.att.com/ tamr/dataquality.html) since so much epidemiological data has problems arising from manual entry, lack of uniform standards for content and formats, data duplication, and measurement errors. We investigate such solutions as duplicate removal, merge purge, and automated detection. Application of traditional clustering algorithms is hindered by the extreme heterogeneity of the data and we discuss new approaches to deal with such heterogeneities. Promising algorithmic methodologies for clustering heterogeneous data are in the papers [Kempner, Mirkin and Muchnik (1997), Kuznetsov and Muchnik (1982), Muchnik and Shvartser (1990)]. We also build on traditional statistical approaches to heterogeneous epidemiological data (see, e.g., [Cox and Piegorsch (1996), Dominici, Parmigiani, Wolpert and Hasselblad (2000), Patil (1991)]). Huge data sets are sometimes best understood by visualizing them. Sheer data sizes require new visualization regimes, which require suitable external memory data structures to reorganize tabular data in secondary storage so that access, usage, and analysis are facilitated (see, e.g., [Abello (1999)]). The working group will be faced with the challenge that developing visualization algorithms becomes harder when data arises from various sources and each source contains only partial information, as is the case, for example, with the cancer monitoring data. In cancer monitoring, we start with data about individuals, regions, etc., and seek to produce aggregative indices, single or multivariate numerical values that measure the risk that an entity in a given group will come down with a certain form of the disease. The groups consist of sets of individuals in different geographic regions, under different environmental conditions (known contaminated drinking water, nearness to environmental risks, etc.), or with different individual risk factors (age, smoking, etc.). While worrying about the "meaningless" statements one can make using such aggregative indices ([Roberts (1994), Roberts (1999)]), we will seek to develop methods for obtaining such indices that are both efficient to compute and useful as predictive tools. We will build on recent approaches to this problem in the bioinformatics context [Hagerty, Muchnik, Kulikowski and Kim (1999)] and in the information technology context [Pennock, Maynard-Reid, Giles and Horvitz (2000), Schapire, Freund, Bartlett and Lee (1998)]. These methods use learning algorithms to develop aggregations using individual "classifiers" based on different sources of data and a "compromiser" to merge the results. This is one of several working groups that benefit from the involvement of researchers from DIMACS' industrial partners who help us modify for epidemiological purposes the methods they have developed for telecommunications/computer network applications - a unique aspect of this project.

- **Working Group on Modeling Social Responses to Bio-terrorism Involving Infec-**

tious Agents

- Dates: First Meeting, May 29 - 30, 2003
- Location: Rutgers University
- Organizers: John Glasser, CDC; Ellis McKenzie, NIH; Fred Roberts. Rutgers University
- Attendance: 34

Models of infectious-disease epidemiology generally assume a fixed social landscape, in which the public consists of passive bystanders and rational actors who comply with health authorities. It is not clear how well this assumption applies to epidemics of any dread disease, much less ones following terrorist attacks, or the extent to which its validity depends on effective communication by authorities and cooperating media. Some analogies (e.g., Three Mile Island, AIDS or West Nile virus) suggest that episodes of mass panic or hysteria would be rare and localized, while actions based on perceived self-interest (e.g., evacuation, queries from the worried-well, antibiotic stockpiling) would be widespread. Acts of spontaneous altruism and mutual aid, as well as criminal opportunism and civil disruption, would also occur. Insofar as changes in social behavior under stress affect the success of medical and public-health interventions, models used to design them would be improved by incorporating relevant social dimensions. The absence of formalized models of social behavior may help to explain the absence of social phenomena in models of infectious diseases or public responses to disease outbreaks. It could be that the inclusion of such phenomena is an elusive goal, however worthy. Nonetheless, this meeting brought social-science experts together with infectious-disease modelers to explore the development of inter-disciplinary methods appropriate to this task.

- **Tutorial: Statistical and other Analytic Health Surveillance Methods**
  - Dates: June 17 - 20, 2003
  - Location: Rutgers University
  - Organizer: David Madigan, Rutgers University; Henry Rolka, CDC; Martin Kulldorff, University of Connecticut
  - Attendance: 71

  Surveillance is a core activity for public health practice. Standard tasks include outbreak detection, post-marketing drug adverse event surveillance, immunization program evaluation and institutional comparisons. Recent concerns about bioterrorism have attracted increased attention to new analytic surveillance methods applied to non-traditional data sources such as over-the-counter drug sales, ambulance dispatches and pre-diagnostic patient encounter summaries. Both the novel methods and new data sources raise significant statistical and computing challenges. The need for innovative statistical methods for surveillance of multiple data streams and the evaluation of utility for new methodologies are especially pressing issues.

  This DIMACS tutorial provided an overview of the challenges and address an array of statistical surveillance tools and techniques. The first two days of the tutorial focused on hands-on use of specific methodologies through case-studies and elementary introductions. The second two days concerned itself more with current research activities and new research challenges.

  The tutorial addressed such topics as:

  - Scan statistics
  - Statistical process control

- Changepoint methods
- Bayesian approaches
- Rule-based surveillance
- False discovery rate and sequential testing
- Text data surveillance

- **Working Group Meeting: The Ecology and Evolution of Influenza and Related Viruses**
  - Dates: June 29 - July 2, 2003
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Jonathan Dushoff, Princeton University, Simon Levin, Princeton University, Joshua Plotkin, IAS/Princeton University
  - Attendance: 25

  Influenza viruses are negative-stranded RNA viruses that cause significant human mortality and morbidity worldwide. The study of influenza epidemiology has profound significance for human health. In addition, influenza viruses offer a remarkable system for evolutionary study. Over the past two decades, the influenza A viral genome has undergone roughly the equivalent of 20 million years of mammalian evolution.

  Influenza viral evolution depends, through frequency-dependent selection, upon influenza's epidemiology - that is, the patterns of host infection. Conversely, annual outbreaks of influenza depend upon the evolution of viral strains. Thus, an understanding of influenza viruses requires a synthetic approach combining both the evolution and the ecology of the pathogen.

  This working group meeting brought together experts who work on influenza from a range of viewpoints. In order to facilitate a synthesis of perspectives, the workshop was organized around half-day sessions, each with one or two short talks followed by active discussion. The five primary topics of inquiry were:

  - The structure and mechanisms of viral antigenic drift
  - Pandemics and interactions between influenza subtypes
  - Disease dynamics and the interpretation of epidemiological data
  - Within-host dynamics and their significance for influenza epidemiology and evolution
  - The implications of historical pandemics for other diseases, particularly SARS

  The workshop also included a grab-bag section to discuss related issues such as variation in viral protein structure and relations between viral genotype and antigenic phenotype. Through short presentations and discussions that include all participants, the workshop improved understanding of influenza viruses on the basis of mathematical modeling combined with molecular, epidemiological, and immunological data.

Selected Research Results

Lattice-theoretic Approaches to Epidemiological Data. One of our epidemiology tutorials introduced epidemiologist David Ozonoff (instructor) to lattice-theorist Alex Pogel (attendee). Their collaboration has blossomed into a full-fledged research effort. Their approach begins with a table giving the attributes of a set of objects. The objects could be individuals and attributes could say if they are sick, if they were exposed to some chemical, and give their age, sex, race, and laboratory values. The lattice diagram of the "closure system" generated by such tabular data provides the analyst with a view that can lead to new insights. Using order-theoretic methods and efficient algorithms Pogel had developed for multiscale analysis of such tabular data and effective visualization of the lattice system, Ozonoff and Pogel studied data giving demographic characteristics of Massachusetts residents who died violent deaths. The relationship of manner of death (suicide, homicide, etc.), means of death (suffocation, poisoning, etc.), age and sex led to a summary of inclusion relations among closed sets generated by the data and a sublattice formed from three of the attributes. From the sublattice one can show that being male and having manner of death "undetermined" implies with 95% confidence that death was from poisoning. This led to hypotheses about why Massachusetts has an "abnormally" high proportion of "undetermined manner" of death classifications, hypotheses later verified through analysis of medical examiner practices. A second data set on features of mothers and low infant birth weight is now being analyzed. The collaboration has already led to new public health hypotheses to explore and a new set of features to be incorporated into the software to make it more useful for public health professionals. The collaboration will also help define the agenda of a DIMACS working group on order-theoretic aspects of epidemiology.

Analogies between Computer Viruses and Biological Viruses. Our Working Group on Analogies between Computer Viruses and Immune Systems and Biological Viruses and Immune Systems has identified new modeling techniques for describing the spread of computer virus attacks and possible defensive strategies. The group has brought together experts from computer science, applied mathematics, physics, and epidemiology and already several fruitful collaborations have been fostered. Guided by the industrial experience of Matthew Williamson (Hewlett Packard), virus propagation patterns were intuitively described and key vulnerabilities were identified. One great concern is the increasing length of the initial occurrence stage, the significant gap between the initial time of attack and the implementation of defensive strategies for a new computer virus. Inspired by ideas discussed by Stephanie Forrest (University of New Mexico) in the working group, Williamson has been developing a "public health" approach that is based on the observation that computers infected by a virus behave differently in one key respect from uninfected computers. Once a virus has infected a machine, it will generally try to connect that machine to as many new computers as possible, as fast as possible, so as to spread itself further. The idea, then, is to limit the rate at which a computer can connect to new computers, where "new" means those that are not on a recent history list. A "throttle" restricts such connections to one a second. In a similar effort, group members Ira Schwartz (Naval Research Labs) and Erik Bollt (Clarkson) are working on a new mathematical method to approximate the topology of a network from a limited number of nodes. The advantage of identifying high traffic regions is that a defensive "ring" strategy will quickly and effectively isolate a new virus. More information can be found on the group's website at http://dimacs.rutgers.edu/Workshops/Analogies.

Models of the Disease Dynamics of Influenza. A research group at Princeton University (DIMACS member Simon Levin, visitors Jonathan Dushoff, Joshua Plotkin, Viggo Andreasen and others) has, for years, been interested in studying the evolution and management of the influenza virus by integrating dynamical models and empirical analyses. Progress towards this goal has benefited greatly from interaction with a diverse group of researchers, ranging from mathematicians to computer scientists to epidemiologists, that have been assembled via DIMACS events, particularly the 2002 International Conference on Computational and Mathematical Epidemiology. The key to the persistence of the influenza virus lies in its continual evolution. Therefore, in order to understand influenza's epidemiological fluctuations and develop strategies for its control, mathematical models of disease transmission must be extended to incorporate evolutionary processes. Aside from these modeling challenges, the study of influenza requires extensive analysis of extant viral sequence data. Formal and informal discussions held at DIMACS workshops, especially with scientists such as N. Cox and A. Perelson, have greatly aided in the conception and integration of these two interrelated research avenues. As a result, the Princeton group has undertaken a successful analysis of the cluster structure in available influenza nucleotides sequences. They have recognized that influenza genotypes do not act as individuals, but as functional clusters or swarms that form the most relevant functional units for modeling. With the sequence analyses in hand, they were able to develop a suite of deterministic and stochastic models for the disease dynamics, incorporating both epidemiological and evolutionary components. These efforts constitute the most extensive work we know integrating the ecology and evolution of multi-strain disease dynamics, and should also point the way to broader development for the theoretical foundations underlying strain interactions in a wide variety of diseases.

Identifying High-risk Patients in a Population of Patients with Known or Suspected Coronary Artery Disease. Sorin Alexe (Graduate student, Rutgers) and Peter Hammer (Rutgers), together with Eugene Blackstone, Hemant Ishwaran , Michael S. Lauer, and Claire E. Pothier Snader (all of Cleveland Clinic Foundation), addressed the problem of distinguishing within a population of patients with known or suspected coronary artery disease, groups at high and at low mortality rates. The study was based on Cleveland Clinic Foundation's dataset of 9454 patients, of whom 312 died during an observation period of 9 years. The Logical Analysis of Data method was adapted to handle the disproportionate size of the two groups of patients, and the inseparable character of this dataset – characteristic to many medical problems. The study has identified a high-risk group of patients representing 1/5 of the population, with a mortality rate 4 times higher than the average, and including 3/4 of the patients who died. The low-risk group identified in the study, representing approximately 4/5 of the population, had a mortality rate 1/3 of the average.

3.7 2002-2005 Special Focus on Computational Geometry and Applications

Overview: Funded by a grant from NSF, the Special Focus on Computational Geometry and Applications commenced in the fall of 2002.

Organizing Committee:

- Steven Fortune (chair), Bell Labs
- Bernard Chazelle, Princeton
- Bill Steiger (local arrangements chair), Rutgers University,

Advisory Committee:

- Pankaj Agarwal, Duke
- Ken Clarkson, Bell Labs
- David Dobkin, Princeton
- Chris Hoffmann, Purdue
- Joe Mitchell, SUNY Stony Brook
- Joe O'Rourke, Smith
- Janos Pach, CUNY

- Ricky Pollack, NYU
- Jarek Rossignac, Georgia Tech
- Jack Snoeyink, UNC

Sponsors and Collaborating Institutions:

The National Science Foundation
The New Jersey Commission on Science and Technology
Rutgers University
Princeton University
AT&T Labs - Research
Bell Labs
NEC Laboratories America
Telcordia Technologies
DIMACS Affiliate Members:
  Avaya Labs
  IBM Research
  Microsoft Research

Theme of the Program:

Computational applications that manipulate geometric models are ubiquitous in science and technology. Efficient algorithms are essential to allow increases in the size, complexity, and completeness of the models. Computational geometry provides efficient algorithms when the models can be described as large collections of relatively simple objects such as points and polygons. Computational geometry has its roots in algorithm analysis and draws heavily from both classical Euclidean geometry and more modern discrete and combinatorial geometry; it includes a strong theoretical foundation of algorithms, data structures, analysis techniques, and conceptual models, as well as an emerging software library.

A few examples illustrate the success of computational geometry methods. First, the solution of partial differential equations is fundamental in many areas, e.g., stress analysis, fluid flow, electromagnetic modeling. A mesh is invariably required to solve the partial differential equation numerically. Delaunay triangulations form the basis of essentially all three-dimensional simplicial mesh generators. Computational geometry has provided a solid theoretical and practical understanding of Delaunay triangulations, including asymptotically efficient algorithms, robust implementations, and connections with other geometric structures. Second, prediction of radio propagation is fundamental to the design of wireless communication networks (a topic emphasized in the special focus on Next Generation Networks Technologies and Applications). Classical geometric optics and the uniform theory of diffraction provide a sophisticated physical model of radio propagation based on ray-tracing. A naive simulation of the model on a large problem instance might require a day or a week of computing time; computational geometry techniques can provide interactive simulations, allowing a system designer to experiment with variations in network parameters. Third, the prospective technology of '3d photocopying' attempts to characterize an object by sampling points on its surface. Computational geometry has provided algorithms that reconstruct a surface model from the sample points, minimizing the number of required sample points and the deviation from the actual object as well as the required computing time.

This special focus on computational geometry and applications follows by 13 years the first DIMACS special year on discrete and computational geometry. During the 13 years, computational geometry has evolved into a mature field with rigorous mathematical foundations. At the same time, there has been a continuing demand for geometric algorithms from affiliated disciplines (graphics, geographic information systems, robotics, VLSI design, computer-aided design and manufacturing, among others). A strong desire has developed within the computational geometry community to improve connections with these and other disciplines. It is timely to have another special focus on computational geometry to exploit these trends, and DIMACS, with its local strength and international connections in the field, is a natural place to have this.

The special focus will deepen the connections between computational geometry and other disciplines. For example, computational geometry and graphics have traditionally been closely tied. However, with changes in technology, e.g., fast cheap PCs and global but relatively slow networks, it may be appropriate to reexamine the algorithmic connections between the two fields. Thus the compression, transmission, and adaptive rendering of meshed models involve combinatorial and topological ideas very much of the flavor of computational geometry; similarly, object-space-based rendering algorithms may become more attractive as the size of models increases. As another example, the solid modeling community has developed in parallel to the computational geometry community. Both communities are concerned with geometric algorithms and have had to deal with many of the same issues, e.g., complexity of representation data structures. The solid modeling community has traditionally been more closely tied to industry and more concerned with curved surfaces; the computational geometry community has been more mathematical and more concerned with linear objects. It is clear that there is considerable potential interaction between the fields, for example on such fundamental conceptual issues as how to describe geometric shapes and how to deal with geometric tolerances. Finally, there are emerging connections with other fields, such as molecular biology (for example through geometric searching).

The special focus will also extend the connections between computational geometry and mathematical) geometry. The connection with discrete and combinatorial geometry is already well-established, but is sufficiently fundamental to develop further; an intriguing prospective connection is with computational real algebraic geometry. In addition, the special focus will address issues of importance within computational geometry, such as the challenging engineering required to implement geometric algorithms.

Workshops and Working Group Meetings

- **Workshop on Geometric Graph Theory**
  - Dates: September 30 - October 4, 2002
  - Location: Rutgers University
  - Organizers: Janos Pach, New York University, Courant Institute of Mathematical Sciences
  - Attendance: 61

  The rapid development of computational geometry in the past two decades presented a powerful new source of inspiration for combinatorial geometry, the theory of geometric arrangements. Many curious extremal problems in recreational mathematics turned out to be crucially important for the analysis of a wide range of geometric algorithms. Perhaps the best example of this is the so-called k-set problem of Erdos, Lovasz, Simmons, and Straus, which is more than thirty years old. Given n points in general position in the plane, what is the maximum number of k-tuples that can be separated from the remaining n-k points by a straight line? After almost twenty years of stagnation and ten years of slow progress, in the past two years Dey and Toth achieved two breakthroughs by substantially improving the best known upper and lower bounds, respectively.

The k-set problem belongs to a newly emerging discipline, geometric graph theory. A geometric graph is a graph drawn in the plane such that its vertices are points in general position and its edges are straight-line segments. The first result on geometric graphs was proved almost seventy years ago by Hopf and Pannwitz: if a geometric graph has no two disjoint edges, then its number of edges cannot exceed its number of vertices. According to Conway's famous Thrackle Conjecture, the same assertion remains true for graphs drawn by (not necessarily straight) continuous arcs with the property that any two of them have at most one point in common. Many recent results in this area are relevant to proximity questions, bounding the number of incidences between points and curves, designing various graph drawing algorithms, etc.

The aim of this workshop was explore the consequences of the new results, and to discuss their extensions and some related questions. We brought together many of the researchers who contributed to the development of this area.

- **Workshop on Computational Geometry**
  - Dates: November 14 - 15, 2002
  - Location: Rutgers University
  - Organizers: Joe Mitchell, SUNY Stony Brook
  - Attendance: 73

This was the twelfth in a series of annual fall workshops on Computational Geometry. This workshop series, founded initially under the sponsorship of the Mathematical Sciences Institute (MSI) at Stony Brook (with funding from the U. S. Army Research Office), has continued during 1996-1999 under the sponsorship of the Center for Geometric Computing, a collaborative center of Brown, Duke, and Johns Hopkins Universities, also funded by the U.S. Army Research Office. In 2000, the workshop returned to the campus of the University at Stony Brook. In 2001, it was held at Polytechnic University in Brooklyn. In 2002, as part of the DIMACS Special Focus on Computational Geometry and Applications, the workshop was hosted and sponsored by DIMACS.

The aim of this workshop was to bring together students and researchers from academia and industry, to stimulate collaboration on problems of common interest arising in geometric computations. Topics covered included, but were not limited to:

  - Algorithmic methods in geometry
  - Geometric data structures
  - Implementation issues
  - Robustness
  - Computer graphics
  - Solid modeling
  - Geographic information systems
  - Applications to computational biology and chemistry
  - Computational metrology
  - Graph drawing
  - Experimental studies
  - Computer vision

- Robotics
- Computer-aided design
- Mesh generation
- Manufacturing applications of geometry
- I/O-scalable geometric algorithms
- Animation of geometric algorithms

Following the tradition of the previous workshops on Computational Geometry, the format of the workshop was informal, extending over 2 days, with several breaks scheduled for discussions. There was also an Open Problem Session in order to promote a free exchange of questions and research challenges.

- **Workshop on Algorithmic Issues in Modeling Motion**
  - Dates: November 18 - 20, 2002
  - Location: Rutgers University
  - Organizers: Pankaj K. Agarwal, Duke University, Leonidas Guibas, Stanford University
  - Attendance: 56

  Motion, like shape, is one of the fundamental modalities to be modeled in order to represent and manipulate the physical world in a computer. As such, motion representations and the algorithms that operate on them are central to all computational disciplines dealing with physical objects: computer graphics, computer vision, robotics, etc. Modeling motion is also crucial for other disciplines dealing with temporally varying data, including mobile networks, temporal data bases, etc. Motion algorithms require computational resources, and frequently sensing and communication resources as well, in order to accomplish their task. Despite the prominent position that motion plays in so many computer disciplines, little has been done to date to provide a clean conceptual framework for representing motion, describing algorithms on moving objects, and analyzing their behavior and performance.

  This workshop brought together people from the different research communities interested in algorithmic issues related to moving objects. The workshop addressed core algorithmic issues as well as aspects of modeling and analyzing motion. The goal was to debate and discuss the issues in representing, processing, reasoning, analyzing, searching, and visualizing moving objects; to identify key research issues that need to be addressed, and to help establish relationships which can be used to strengthen and foster collaboration across the different areas.

- **Workshop on Implementation of Geometric Algorithms**
  - Dates: December 4 - 6, 2002
  - Location: Rutgers University
  - Organizers: Herve Bronnimann, Polytechnic University and Steve Fortune, Bell Labs
  - Attendance: 42

  It is notoriously difficult to implement geometric algorithms. This difficulty arises in part from the conceptual complexity of geometric algorithms, the proliferation of special cases, the dependence of combinatorial decisions on numerical computation, and frequent theoretical focus on worst-case asymptotic behavior.

  This workshop addressed research issues related to the implementation of geometric algorithms. Typical, but not exclusive topics include:

- Numerical issues
- Noisy data and data repair
- Geometric data structures
- Massive geometric data sets
- Algorithm library design
- Algorithm engineering
- Experimental studies

Numerical issues have long been an important concern in the implementation of geometric algorithms. In the last decade the issue has become a central research topic in computational geometry, and a reasonably successful approach based on the use of exact (extended-precision) arithmetic has been developed. However, many significant problems remain—high-level rounding, extension to curved objects, performance—and the practical impact of the research is not yet clear.

Geometric data sets based on physical measurements are inherently noisy. If such geometric data also has combinatorial structure, the geometric and combinatorial information may be inconsistent. To be useful, geometric algorithms must be able to repair such data, that is, in some fashion eliminate inconsistencies. Unfortunately, there is little relevant theory, and current data repair is heuristic at best.

Geometric data structures are known that can represent complex structures in any dimension. However massive data sets in two dimensions, or even modest data sets in high dimension, can require enormous amounts of memory. A challenging research topic is to design algorithms and data structures that are cognizant of the memory hierarchy—cache, main memory, disk—and to provide appropriate implementations.

These are just some of the problems faced by general purpose geometric algorithms libraries. Considerable effort has been expended developing the geometric algorithm library CGAL, which is now reasonably mature. CGAL,just together with the LEDA algorithm library, provide an unparalleled resource for users of geometric algorithms. Further development of algorithm libraries requires attention to many issues—those mentioned above, but also functionality, interface, performance, and support for specific application areas.

We brought together both researchers and practitioners. We hope that practitioners benefited from discussions of the state of the art in research, and that researchers benefited by being exposed to implementation issues of practical importance.

- **Workshop on Medical Applications in Computational Geometry**
  - Dates: April 2 - 4, 2003
  - Location: Rutgers University
  - Organizers: Danny Chen, University of Notre Dame; Jean-Claude Latombe, Stanford University
  - Attendance: 66

Computer technology plays an increasingly important role in modern medicine and life sciences. Many medical problems are of a strong geometric nature and may benefit from computational geometry techniques. The DIMACS workshop on Computational Geometry and Medical Applications aimed to provide a forum for researchers working in computational geometry, medicine, and other related areas to get together and exchange ideas, and to promote cross-fertilization and collaborations among these areas. The theme of the workshop was on the exploration of the applicability of computational geometry to medical problems and the new challenges posed by the current medical research and practice to the geometric computing study. Examples of topics include surgical simulation and planning, geometric representation and modeling of medical objects and human-body tissue structures, geometric problems in medical imaging, computational anatomy, registration and matching of medical objects, etc. The workshop consisted of both invited and solicited contributions.

- **Workshop on Surface Reconstruction**

  - Dates: April 30 - May 2, 2003
  - Location: Rutgers University
  - Organizers: Nina Amenta, University of California - Davis; Fausto Bernardini, IBM - T. J. Watson Research Center
  - Attendance: 55

  Surface reconstruction is the problem of producing a representation of a two-dimensional surface in 3D, give a set of sample points lying on or near the surface. There are a number of interesting systems which collect sample sets and construct surfaces, using laser range scanners, structured light and other techniques. Recent results in computational geometry have focused on algorithms for the surface reconstruction problem and finding ways to guarantee topologically and geometrically correct outputs, given good enough input samples. This workshop encouraged interaction between systems builders and the computational geometry community.

  We surveyed the state of the theory and practice of surface reconstruction, and looked at related problems arising in systems that reconstruct objects, such as the alignment of multiple laser range scans, integrating photometric data, reconstructing dynamic scenes and real-time object acquisition.

- **Workshop on Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications**

  - Dates: May 14 - 16, 2003
  - Location: Rutgers University
  - Organizers: Regina Liu, Rutgers University; Yehuda Vardi, Rutgers; Robert Serfling, University of Texas at Dallas; Diane Souvaine, Tufts University
  - Attendance: 92

  Multivariate statistical methodology plays a role of ever increasing importance in real life applications, which typically entail a host of interrelated variables. Simple extensions of univariate statistics to the multivariate setting do not properly capture the higher-dimensional features of multivariate data, nor do they yield geometric solutions because of the absence of a natural order for multidimensional Euclidean space. A more promising approach is the one based on "data depth", which can provide a center-outward ordering of points in Euclidean space of any dimension. Extensive developments in recent years have generated many attractive depth-based tools for multivariate data analysis, with a wide range of applications. The diversity in approaches, emphases, and concepts, however, makes it necessary to seek unified views and perspectives that would guide the further development of the depth-based approach.

The concept of data depth provides new perspectives to probabilistic as well as computational geometries. In particular, the development of implementable computing algorithms for depth-based statistics has brought about many new challenges in computational geometry. This workshop created a unique environment for multidisciplinary collaboration among computer scientists, theoretical and applied statisticians, and data analysts. It brought together active researchers in these fields to discuss significant open issues, establish perspective on applications, and set directions for further research.

- **Workshop on Geometric Optimization**
  - Dates: May 19 - 21, 2003
  - Location: Rutgers University
  - Organizers: Joe Mitchell, SUNY Stony Brook; Pankaj Agarwal, Duke University
  - Attendance: 62

  Combinatorial optimization typically deals with problems of maximizing or minimizing a function of one or more variables subject to a large number of constraints. In many applications, the underlying optimization problem involves a constant number of variables and a large number of constraints that are induced by a given collection of geometric objects; these problems are referred to as geometric-optimization problems. Typical examples include facility location, low-dimensional clustering, network-design, optimal path-planning, shape-matching, proximity, and statistical-measure problems. In such cases one expects that faster and simpler algorithms can be developed by exploiting the geometric nature of the problem. Much work has been done on geometric-optimization problems during the last twenty-five years. Many elegant and sophisticated techniques have been proposed and successfully applied to a wide range of geometric-optimization problems. Several randomization and approximation techniques have been proposed. In parallel with the effort in the geometric algorithms community, the mathematical programming and combinatorial optimization communities have made numerous fundamental advances in optimization, both in computation and in theory, during the last quarter century. Interior-point methods, polyhedral combinatorics, and semidefinite programming have been developed as powerful mathematical and computational tools for optimization, and some of them have been used for geometric problems.

  This workshop brought together people from different research communities interested in geometric-optimization problems. The goal was to discuss various techniques developed for geometric optimization and their applications, to identify key research issues that need to be addressed, and to help establish relationships which can be used to strengthen and foster collaboration across the different areas.

Selected Research Results

Reconstructing Sets from Interpoint Distances. Paul Lemke (Troy U.), Steven Skiena (SUNY, Stony Brook) and Warren D. Smith (Visitor, DIMACS) studied the problem of reconstructing sets from interpoint distances. The general question addressed by their work was: Which point sets realize a given distance multiset? Interesting cases include the “turnpike problem” where the points lie on a line, the “beltway problem” where the points lie on a loop, and their multidimensional versions. These problems have applications in genetics and crystallography. The research of Lemke, Skiena and Smith gives improved combinatorial bounds for the turnpike and beltway problems. The research further presents a pseudo-polynomial time algorithm as well as a practical $O(2^n n \log n)$-time algorithm that find all solutions to the turnpike problem, and shows that certain other variants of the problem are NP-hard.

3.8 2003-2006 Special Focus on Communication Security and Information Privacy

Overview:

The Special Focus on Communication Security and Information Privacy began in January 2003 with a preliminary workshop and is currently planned to run for three and a half years. During FY2003, we worked extensively on program development for this focus, which got under way in full swing in summer 2003.

Organizing Committee:

- Ronitt Rubinfeld, NEC Research Institute, chair
- Bill Aiello, AT&T Labs, co-chair
- Hugo Krawczyk, Technion
- Avi Rubin, AT&T
- David Wagner, UC Berkeley

Sponsors and Collaborating Institutions:

The National Science Foundation
The New Jersey Commission on Science and Technology
Rutgers University
Princeton University
AT&T Labs - Research
Bell Labs
NEC Laboratories America
Telcordia Technologies
DIMACS Affiliate Members:
  Avaya Labs
  IBM Research
  Microsoft Research

Theme of the Program:

Vitally important aspects of our modern society have become dependent on rapid and secure communication, which is increasingly electronic. The new electronic age offers vast potential for new services and applications, but gives rise to serious new vulnerabilities and security threats. Moreover, many of the most important new applications come at the price of threats to privacy. We plan a 3.5-year "special focus" on Communication Security and Information Privacy, beginning in January 2003, that will explore the new vulnerabilities and threats and new methods for dealing with them.

Within the last decade a tremendous transition has taken place in communications networks. Previously, nearly all communication, whether data, voice or other media, was carried over private networks. Anyone who was not a customer of the network provider was not given physical access to the network. Securing such networks was relatively straightforward. While a great deal of data and media traffic still run over circuit switched or packet switched ATM or Frame Relay private networks, a huge amount and variety of data and media traffic now run over the public Internet, so much so that Internet is now an important national infrastructure whose integrity is vital to the functioning of our economy, culture, and government. The migration of communication services to the Internet is still very much in progress. This migration brings with it new and complex challenges for maintaining communication security.

There are many factors driving the migration to the Internet. One is universal connectivity. The Internet protocol allows users with many different types of local area network technologies (e.g., Ethernet, and 802.11) to be integrated into a single large network. This allows for a type of positive feedback often referred to as the "network effect." The network grows quickly because the number of users, servers, and devices that are already reachable on the Internet make it very valuable to any new IP device. A second factor is unification. Unlike the circuit switched world for which signaling and data/media were carried by two separate networks, signaling and data/media can both be carried over the Internet. For network providers, migrating their services onto an Internet backbone means that they need only deploy, manage, and control a single network, thereby reducing their cost of providing services. Finally, the ultimate promise of the Internet is as a platform for integrating a variety of services such as voice, instant messaging, mobile presence, multimedia, Web and data services. While these are powerful factors driving the migration to IP communications, they have serious security repercussions. Indeed, securing an extremely large, shared services, packet-based IP network with a large number of administrative domains is a much more complex task than securing segregated/circuit switched networks.

Furthermore, through the collection and dissemination of vast amounts of data, the Internet allows users to take advantage of new functionalities that inherently require new notions of security. For example, new issues of privacy for Internet users and applications are arising due to the multitude of data available online. This new electronic reality and the vast potential for interaction between users and computers give rise to new digital applications and services once thought possibly only in the physical tangible world. This, in turn, creates the need for the invention and implementation of new security and cryptographic techniques. Enabling secure electronic commerce and securing digital rights management are some central examples of the new challenges faced in the security area.

Some of the most exciting progress in the fields of communication security and information privacy has come because of the interconnections of practitioners in these fields with researchers developing relevant methods of theoretical computer science and mathematics. This special focus explores these interconnections in order to address some of the fundamental challenges to communication security and information privacy posed by the rapid transition and remarkable growth of new applications in today's communication networks.

The Themes of the Special Focus Include:

- Studying protocol and host vulnerabilities related to Internet communication. Among them are the weakness or total lack of source authentication for the base protocols in the IP suite, lack of admission control mechanisms, vulnerability of hosts to implementation and configuration errors. What is more, protocol and host vulnerabilities can be exploited in tandem to create serious attacks such as distributed denial of service attacks.

- Securing the protocol layer. The special focus analyzes a wide range of security issues related to newer technologies such as wireless access at the lower layers of the protocol stack, or Web services at the higher layer of the protocol stack, including issues dealing with ad-hoc trust establishment, secure roaming between overlay networks, the controlled execution of untrusted code, and peer-to-peer connection in pervasive networking scenarios.

- Seamless data movement vs. privacy and property rights. The power of service providers to automatically log and analyze information on site visitors or customers for collection and dissemination is so great that it must be properly managed or else there is a significant potential for abuse. The special focus examines both violation of property rights and violation of privacy both in the general context and in more specialized applications such as health care data and electronic voting.

- Cryptography and secure protocols. As technology evolves, cryptography faces the task of developing new security models and techniques such as developing a complete suite of solutions that can handle the concurrency and asynchrony of the Internet and obtaining information from multiple data sets while protecting privacy and confidentiality.

Workshops and Working Group Meetings

- **Workshop on Software Security**
  - Dates: January 6 - 7, 2003
  - Location: Rutgers University
  - Organizers: Gary McGraw, Cigital
  - Attendance: 56

  The security of computer systems and networks has become increasingly limited by the quality and security of the software running on these machines. Researchers have estimated that more than half of all vulnerabilities are due to buffer overruns, an embarrassingly elementary class of bugs. All too often systems are hacked by exploiting software bugs. In short, a central and critical aspect of the security problem is a software problem. How can we deal with this?

  The Software Security Workshop explored these issues. The scope of the workshop included security engineering, architecture and implementation risks, security analysis, mobile and malicious code, education and training, and open research issues. In recent years many promising techniques have arisen from connections between computer security, programming languages, and software engineering, and one goal is to bring these communities closer together and crystallize the subfield of software security.

Selected Research Results

Since this special focus did not really get started until summer 2003, with the exception of one workshop held in January 2003, it is premature to list research results.

4 Multi-Disciplinary, Multi-Institutional Projects

During this period, DIMACS had several multi-disciplinary, multi-institutional research projects. Our projects of this kind are scientifically exciting and have taken us into new and exciting research areas with a variety of partner institutions and in a variety of fields.

The participants in these projects, both DIMACS members and outsiders, have repeatedly emphasized that the projects would not have happened without the organizational infrastructure provided by DIMACS. DIMACS brings scientific expertise and experience with large, complex projects that involve a variety of institutions and disciplines, and such projects provide opportunities for students, postdocs, faculty members, and industrial researchers to share in research experiences and learn more about each others' "cultures" and ways of doing science.

Several of these projects involve research collaborations with international parters. Most of them heavily involve our industrial partners.

For projects involving industrial partners, DIMACS provides a neutral site for collaboration. Industrial researchers find it easier to work on a joint high risk precompetitive project under the umbrella of an academic center which enhances the opportunities for interactions.

4.1 The LINK Project

LINK is a software system designed to be a general-purpose environment for experimentation with discrete mathematics. The environment is friendly enough so that non-technical users will be able to visualize problems easily, yet powerful en ough for computer scientists and mathematicians to express complicated computations. The system is designed to support educational pursuits, to be used as a research tool, and to be useful in solving real-life problems. LINK was first funded by an NSF grant to DIMACS for a three year period over 1993 to 1996, and it has also received support from the Office of Naval Research and Los Alamos National Laboratory.

At Los Alamos, LINK was adapted as a foundation for a system to aid the Internal Revenue Service in fraud detection, using graphical clustering techniques to identify similar (possibly fraudulent) returns. The resulting LINK software tool was installed at all IRS service centers for use by tax examiners. Participating scientists in the first stage of LINK were Jonathan Berry (then a DIMACS postdoc, now at Lafayette College), Nathaniel Dean (then at Bell Labs, now at Texas Southern), Mark Goldberg (Rensselaer Polytechnic Institute), Greg Shannon (Spanning Tree Technologies), and Steve Skiena (SUNY Stony Brook).

In FY2003 LINK was a central tool in the Special Focus on Data Analysis and Mining. In planning for that focus, we investigated the use of software to model the process of scientific discovery and in part order to develop and investigate conjectures in graph theory and chemistry. LINK played a role in our receiving an NSF award for a working group to explore this topic and is one of the pieces of software that is playing a major role in it.

4.2 Project on Monitoring Message Streams

Objective of the project:

We aim to improve on existing methods for monitoring huge streams of textualized communications to automatically identify clusters of messages relating to significant "events." We will study retrospective or "supervised" event identification where the system begins with a set of some labeled documents (i.e., documents that have been associated with one or more events) and automatically associates a relevant new message with one or more of these pre-existing events. This task is of great value in forensic activities. In later years, we will also study prospective detection (or "discovery" or "unsupervised learning"), where we decide if a "new event" is present in the flow of messages. Our goal would not be to label the new document clusters, but simply to alert users that a new cluster has been detected. This is considerably more difficult and would be of enormous value as an intelligence tool.

Background:

Our premise is that significant performance improvements in the automatic processing of messages will result from innovations involving more than one component of the process. We have assembled a top-notch, multi-disciplinary team with novel perspectives on each component. This team will conduct coordinated, large-scale, systematic experimentation. This should significantly advance the state-of-the-art and achieve substantial improvements in performance according to accepted measures.

Automatic processing of messages depends on five components: (1) compression of text to meet storage and processing limitations; (2) representation of text in a form amenable to computation and statistical analysis; (3) a matching scheme for computing similarity between documents in terms of the representation chosen; (4) a learning method building on a set of judged examples to determine the key characteristics of a document cluster or "event"; and (5) a fusion scheme that combines methods that are "sufficiently different" to yield improved detection and clustering of documents. Our team comprises experts in each of these areas, including some who have documentably "best" methods in some of these components. Existing methods make use of some or all five components, but don't exploit the full power of the components and/or an understanding of how to apply them to text data. For example, in the TREC-2001 batch filtering evaluation, Lewis used an off-the-shelf support vector machine (SVM) supervised learner, but tuned it for the frequency properties of text data. This combination dominated competing approaches, including two other SVM runs.

Existing methods also aim at fitting into available computational resources without paying attention to upfront data compression. Recent research demonstrates that, for many statistical methods, dramatic reductions in space and time complexity can be realized with only a controllably small impact on the methods' numerical properties, by making a preprocessing pass over the data to reduce its dimension. Unlike feature-extracting dimension reduction, which, done without care, may lead to very bad results, the recent algorithmic techniques, including randomized projections, perform no semantic function themselves, so fit a different paradigm – used, with some success, by Alta Vista, for example. We believe that sophisticated dimension reduction methods in a preprocessing stage followed by the use of sophisticated statistical tools in later stages can be a very powerful approach. Our goal is to identify the best combination of such newer methods through a careful exploration of a variety of tools. We will address issues of effectiveness (how well the task is done) and efficiency (in computational time and space) and use a combination of new or modified algorithms and improved statistical methods built on the algorithmic primitives.

In later years, we want to extend our work to the much harder problem of unsupervised learning, still concentrating on utilizing new methods for the five components defined above. Because the pure supervised learning task is such a challenge for the classifying power of any reasonable method, we will concentrate on "semi-supervised" learning, learning in which human analysts help to focus on features that will be most indicative of change or anomaly and algorithms assess whether incoming documents do or do not deviate "significantly" on those features. New techniques will be needed to represent the data in a way that makes it possible to highlight significant deviation ("abnormality") through an appropriately defined metric and through the use of new clustering algorithms that take into account the analyst-designated features.

Project Researchers:

Endre Boros, Rutgers, RUTCOR
Wen-Hua Ju, Avaya Labs
Paul Kantor, Rutgers, SCILS

David Lewis, Consultant
David Madigan, Rutgers, Statistics
Muthu Muthukrishnan, Rutgers, Computer Science
Rafail Ostrovsky, Telcordia
Fred Roberts, DIMACS and Rutgers Mathematics
Ilya Muchnik, DIMACS
Martin Strauss, AT&T Labs

Other Personnel:
Alex Genkin, DIMACS, Programmer
Vladimir Menkov, Consultant, Programmer
Andrei Anghelescu, Rutgers, Computer Science, Graduate Assistant
Dmitriy Fradkin, Rutgers, Computer Science, Graduate Assistant

Selected Research Results

Improvements in Term Discriminating Power. Combining the representation and matching components in the project's classification of components of automatic message processing, experiments on "term selection" (feature selection specialized to the text case) have shown that it is possible to achieve substantial accuracy in classifying text. Another approach, based on a "voting combination" of several measures of "term effectiveness," did extremely well on a rather peculiar set of TREC2002 tasks called the "topic intersection tasks." The most effective measure of term discriminating power as found by experimentation can be interpreted as a measure that maximizes the Hamming distance between the relevant and the not-relevant documents. The surprising result is that the Hamming measure is relatively robust when moving from the training collection to the test collection.

Nearest Neighbor Classifiers for Text Filtering. One approach to "supervised learning" is to route a message by finding the k most similar training messages (neighbors). We then assign messages to the classes that are most common among their neighbors. This k-nearest neighbor (kNN) heuristic has been studied since the 1950's, but for text messages only since the 1990's. It is moderately effective for text, but is considered inefficient. However, since finding neighbors only needs to be done once, no matter how many classes there are, this may be more efficient than one-classifier-per-topic methods if there are a large number of classes. We worked on fast implementation of kNN. The approach was to store text and classes separately (the is the representation component) and arrange computations to do work proportional only to number of class labels in neighbors, not total number of classes. We developed search engine heuristics, using the in-memory inverted file (the matching component). These retain only high impact terms within each document, or within each inverted list and there is a classification-time selection of inverted lists. Our results showed a great reduction in size of inverted index and time of classification, at a slight additional cost in effectiveness. Effectiveness was slightly below our best methods (Bayesian probit and logistic classifiers). Specifically, our weighted kNN has macro-F1 of 0.49 on a Reuters corpus I dataset, which is within 15% of the best alternatives we tested. The compressed index is 90% smaller than the original index with only a 7-12% loss in effectiveness. Approximate matching is 90-95% faster with only 2-10% loss in effectiveness. Ours were the first large scale experiments on search engine heuristic for neighbor lookup in kNN and the results arose from a very successful partnership between theoreticians and practitioners.

Project Reports

- Project Overview, Paul B. Kantor and Fred S. Roberts, April 2003.
- Optimization of SVM in a Space of Two Parameters: Weak Margin and Intercept, Andrei V. Anghelescu and Ilya B. Muchnik, May 2003.
- Prospective Data Fusion for Batch Filtering, Andrei Anghelescu, Endre Boros, Dmitriy Fradkin, David D. Lewis, Vladimir Menkin, David J. Neu, Kwong Bor Ng., and Paul Kantor, May 2003.
- Feature Selection and Batch Filtering, Endre Boros and David J. Neu, May 2003.
- Sparse Bayesian Classifiers for Text Categorization, Susana Eyheramendy, Alexander Genkin, Wen-Hua Ju, David D. Lewis, and David Madigan, May 2003.
- A Closer Look at Nearest Neighbor Text Classification, David D. Lewis and Vladimir Menkov, May 2003.
- Prospective Data Fusion for Batch Filtering, Andrei Anghelescu, Dmitiry Fradkin, David D. Lewis, Vladimir Menkov, Kwong Bor Bg, and Paul Kantor, February 2003.
- A Closer Look at Nearest Neighbor Classification, David D. Lewis and Vladimir Menkov, January 2003.
- Towards a Formal Framework for Adaptive Filtering, Paul Kantor and David Madigan, December 2002.
- Towards Fast, Effective Nearest Neighbor Text Classification, David D. Lewis, S. Muthukrishnan, R. Ostrovsky, M. Strauss, November 2002.
- Distances from Sketches, Martin Strauss, November 2002.
- Sparse Bayesian Classifiers for Text Categorization - Batch Learning, Alexander Genkin, David D. Lewis, and David Madigan, November 2002.
- Sparse Bayesian Classifiers for Text Categorization - Online Learning, Wen-Hua Ju and David Madigan, November 2002.
- Combinatorial Clustering for Textual Data Representation in Machine Learning Models, Andrei Anghelescu and Ilya Muchnik, November 2002.
- An Experimental Evaluation of a Combinatorial Clustering Model for Textual Data Representation, Andrei Anghelescu and Ilya Muchnik, November 2002.
- A Simple, Coarse Vector Nearest Neighbors Algorithm, S. Muthukrishnan, November 2002.
- Some Statistical Analysis Algorithms on Data Streams, S. Muthukrishnan, November 2002.
- Notes on Term Selection, Endre Boros, November 2002.
- Training a Rocchio classifier for adaptive filtering, Andrei Anghelescu, David D. Lewis, Vladimir Menkov, and Paul Kantor, TREC 2002.
- Feature Selection and Batch Filtering, Endre Boros and David J. Neu, TREC 2002.
- CCR-DIMACS Workshop,

Project Presentations

- Monitoring Message Streams: Retrospective and Prospective Event Detection, Fred Roberts, December 2002.
- Monitoring Message Streams: Future Work, Fred Roberts, December 2002.
- Combinatorial Clustering for Textual Data Representation, Andrei Anghelescu and Ilya Muchnik, December 2002.
- Towards Fast Algorithms for Nearest-Neighbor Searching, Rafail Ostrovsky, December 2002.
- Nearest Neighbor Text Classification, David D. Lewis, S. Muthukrishnan, R. Ostrovsky, and M. Strauss, December 2002.
- Infrastructure: Software and Data, Andrei Anghelescu, Paul Kantor, David D. Lewis, and Vladimir Menkov, December 2002.
- Fusion of Methods, Paul Kantor and Dmitriy Fradkin, December 2002.
- How do we measure success? Paul Kantor, December 2002.
- Towards a Formal Framework for Adaptive Filtering, Paul Kantor and David Madigan, December 2002.
- Bayesian Methods for Text Categorization, Susana Eyheramendy, Alex Genkin, Weh-Hua Ju, David D. Lewis, and David Madigan, November 2002.
- Streaming Data Analysis, S. Muthu Muthukrishnan, November 2002.
- Randomized Projection Methods, Martin Strauss, November 2002.
- Feature Selection, Endre Boros, November 2002.

Selected Project Papers

- A sequential Monte Carlo method for Bayesian analysis of massive datasets, Greg Ridgeway and David Madigan, submitted to JDMKD.
- On Bayesian learning of sparse classifiers, Wen-Hua Ju, David Madigan, and Steven Scott, Working Paper.
- On the Naive Bayes model for text categorization, Susana Ehyeramendy, David D. Lewis, and David Madigan, AI-STATS, 2003.
- On retrieval properties of samples of large collections, David Madigan, Yehuda Vardi, and Ishay Weissman, AI-STATS, 2003.
- String edit distance matching problem with moves, G. Cormode and S. Muthukrishnan, SODA 2002.

- Simple and practical sequence nearest neighbors with block operations, S. Muthukrishnan and S. C. Sahinalp, CPM 2002.

- Efficient algorithms for document retrieval problems, S. Muthukrishnan, SODA 2002.

4.3 Project on Mining Multilingual Resources Using Text Analytics

This project is primarily carried out by our partners at IBM.

To improve the detection of new and significant events or changes within a topic of interest from a multilingual stream of messages, we will develop in this project a novel, automatic, probabilistic system to extract entities, events and relations and use them to populate an ontology (Knowledge Representation) that can be used to identify significant new information. The KR is a unifying framework to represent acquired knowledge independent of the source including sources from different languages; the system will encode both reinforcing and contradicting information by conditioning probabilities on source id and source reliability. The extracted facts will be used to build a probabilistic KR (PKR) that represents the system's view (or knowledge) of the world using probabilistic relational models. This PKR has many uses but we will evaluate our progress in this project by comparing our results to the current state-of- the-art in discovering significant new relations/attribute values in newly received data by building a First Story Detection system, as described below.

By understanding the incoming text beyond simple bag of words models as is commonly done in today's text search and classification systems, we expect to improve the accuracy of discovering new information in a specific subtopic or answering a user's question about some entities or events. Current systems have a high false alarm rate with a low detection rate: incoming messages appear to be new though they do not provide significant new information.

By analyzing documents in multiple languages that describe the same event, called comparable documents, we will have a rich set of facts that may reinforce each other or may be contradictory. The richness of a comparable corpus will require the introduction of a probabilistic model to the KR to handle the information fusion. The PKR will be able to incorporate conditioning the probability of a fact on its source, and other related facts of the same entity or related entities. The PKR will learn its parameters from the comparable corpus.

Knowledge Acquisition using Natural Language Processing: The work will build on a number of technologies developed at IBM Research in the past few years. IBM researchers have built robust statistical extractors in Chinese and English; our project is based on their pioneering work in creating statistical parsers that can be trained for a variety of sources such as newswire, email, and web data. We will also use their state-of-the-art topic detection and tracking systems to home in on a more topic relevant subset of the data. In addition, we will use their recently developed algorithm for identifying comparable documents in a multilingual stream. Comparable documents are those that describe the same fundamental set of facts.

Knowledge Representation: In addition, we will work with an ongoing effort at IBM to build the NETL2 ontology system. NETL2 is a knowledge representation system being developed at Watson Labs by Scott Fahlman. It is based in part on Fahlman's earlier NETL work, but the current version is designed to run well on a standard high-end workstation. NETL2 implements a semantic network in which the nodes represent entities and the links represent statements and relations. It supports reasoning about hierarchical type relationships and can efficiently detect type mismatches: If Fred is a male human, he cannot (in a literal sense) be the mother of some other human or the father of an elephant. He can be placed in a role requiring a human, but not in a role requiring a company or a vehicle. NETL2's hierarchical classification system also extends to relations and actions, and there is an inclusion hierarchy for spatial and temporal regions. So if we acquire from a news story the information that Bill Gates gave a keynote talk in Oslo in May of 2002, this fact would be returned when we ask the system whether any executive of an American software company has been in Norway in 2002.

The NETL2 system serves as the target for our knowledge acquisition work, making the extracted knowledge (and what can be inferred from it) available for other applications. We will in the first year demonstrate the automatic acquisition of the knowledge from running text.

Probabilistic Knowledge Representation: In addition to automatic acquisition, we want to build a probabilistic KR because extracted information may not be correct; in particular, we will have attribution to a source for extracted knowledge; the probability that the fact is correct will depend on the sources used and other attributes in the domain model. We will extend our Conditional Maximum Entropy framework to build probabilistic relational models (a richer model than Bayesian Networks from the Graphical Models family). In the first year, we anticipate that this work will only be a feasibility result, so our results will be an initial proof of concept for learning a PKR.

Data: The data source for the project will be obtained by web crawling on the order of 100 million pages in total in three languages of interest even though our milestones will focus on one language pair, most probably Chinese and English for the first year. We will create a data set of URLs of the crawled pages. This URL dataset will be made publicly available. The URLs can be used to crawl the pages though some will become unavailable as web sites change their content. We may add additional annotations to the URLs such as language ID.

Project Researchers:

Salim Roukos, IBM, PI for IBM Subcontract
Scott Fahlman, IBM

4.4 Project on Bioterrorism Sensor Location

Early warning of the outbreak of a disease or exposure that can lead to a disease is critical in minimizing the impact of that disease. This well-accepted principle in the field of public health is vitally important in the defense against bioterrorism and is a prime factor in the government's plans to develop and place sensors/detectors that might quickly provide warning of a possible bioterrorist attack. Unfortunately, networks of sensors are expensive to purchase and ways to locate them to maximize "coverage" and expedite an alarm in case of an attack are not easy to determine. Approaches that improve upon existing, ad hoc location methods could save countless lives in case of a bioterrorist attack and would undoubtedly save the public a great deal of money in terms of capital and operational costs. This is the motivation for the DIMACS Bioterrorism Sensor Location project.

The project was initiated in FY2002 after the center was contacted by officials at the Defense Threat Reduction Agency (DTRA) and the MITRE Corporation. Plans for the project have developed in consultation with DTRA, MITRE, the Institute for Defense Analyses, the New York City Department of Health, the Department of Homeland Security, and Los Alamos National Laboratories.

This project is aimed at formulating the sensor location problem (SLP) carefully through scenario development and model building and at developing algorithmic approaches for solving it. Working with various government agencies will help to identify relevant modeling parameters and issues. This effort aims to produce new concepts, ideas, and methods that could be of fundamental importance to various defense and homeland security agencies and local health departments. It will produce methods that will be usable by decision makers to choose an appropriate mix of sensors and to decide where to locate them for best protection and early warning. When sensors set off an alarm, public health decision makers need to decide whether or not a bioterrorist attack has taken place, estimate its extent and location, and choose among a variety of possible responses. They can be guided by the pattern of sensor alarms. But the problem of how to interpret this pattern, the Pattern Interpretation Problem (PIP), presents a major theoretical challenge that is intertwined with the SLP. Where the sensors are affects what their messages mean and how the messages will be interpreted affects where the sensors should be placed. This project will formulate the PIP precisely, investigate its implications for finding good solutions to the SLP, and find methods for solving the PIP and attacking the PIP and SLP together. The SLP and PIP are ripe for mathematical modeling and algorithmic analysis using powerful, sophisticated tools of modern information technology. A multi-disciplinary project team is attacking these problems using methods of computer simulation, network design, network analysis, location theory, reliability theory, data mining, fluid dynamic modeling, statistical analysis, source-to-dose modeling, and modeling the spread of threat agents through the atmosphere. It is also building on long experience in siting monitors of air and water quality.

The concepts, algorithms, and approaches developed in this project are aimed at solving a problem of vital importance to society. They will be widely shared with government agencies working on bioterrorist defense. A broader community will be introduced to the problems and methods relevant to the project through a workshop and tutorial, a series of mini-workshops, and a well-publicized seminar series, all open to the public. Students and postdoctoral fellows will be invited to all of these events and, in addition, local students and postdocs will be heavily involved in the research. Lectures on the project will be presented in programs for high school teachers, undergraduate faculty, and undergraduate research students. Because the project is inherently interdisciplinary, it should lead to enhancement of the fledgling lines of communication and collaboration among members of such a large and interdisciplinary team and, ultimately, among others in their disciplines. The project should have impacts well beyond the bioterrorist sensor location problem. The methods developed should be useful in sensor network and pattern detection applications in communications, transportation, civil and chemical engineering, etc.

Project Researchers:

Fred Roberts, DIMACS, Rutgers and Mathematics
Benjamin Avi-Itzhak, RUTCOR
Thomas Boucher, Rutgers Industrial Engineering
Tamra Carpenter, Telcordia Technologies
David Coit, Rutgers Industrial Engineering
Elsayed Elsayed, Rutgers Industrial Engineering
Panos Georgopoulous, UMDNJ-Robert Wood Johnson Medical School
Melvin Janowitz, DIMACS
Paul Kantor, Rutgers Scils
Howard Karloff, AT&T Labs
Jon Kettenring, Telcordia Technologies
Moshe Kress, Center for Military Analyses (Israel)
Paul Lioy, UMDNJ-Robert Wood Johnson Medical School
David Madigan, Rutgers Statistics
S. Muthukrishnan, Rutgers Computer Science
Rafail Ostrovsky, Telcordia Technologies
Michael Rothkopf, RUTCOR
Yehuda Vardi, Rutgers Statistics

Selected Research Results

Using Combinatorial Optimization Approaches to Equipment Placing Problems. In network planning in telecommunications, the "node placement" problem is the problem of determining the locations or nodes at which to install certain types of networking equipment. "Coverage" and cost considerations, of great importance in the SLP, are a major consideration. Researchers at our partner Telcordia Technologies have studied variations of this problem arising from broadband access technologies. There are inherent range limitations that drive placement in these problems. For instance, a customer for DSL service must be within a prescribed distance of an assigned access multiplexer. Thus, an access multiplexer in a DSL network has a "range limit" and can only cover customers within a certain area. The related location problem was formulated as an integer programming problem at Telcordia and Telecordia researchers have developed dynamic programming algorithms to solve two practical variants of it. Methods were tested on both realistic and randomly generated test problems and the methods were implemented as part of Telcordia's "Network Planner" prototype software tool. We expect to modify this work in our project and have begun to think about how to do it. Also relevant is work at Telcordia on emerging transparent optical networks, where there are limitations on how far an optical signal can travel without electrical regeneration. Algorithms for placing regenerating equipment in such networks must explicitly embed the limitations. Such algorithms, using connected dominating sets and "impairment aware routing," have been developed at Telcordia and we have begun to discuss how to modify them for the SLP. Restrictions on types of equipment that can be placed at given nodes are very similar to those we envision on types of sensors that could be placed at given locations, and constraints on how far a signal can travel are related to constraints on area that can be "covered" by a sensor. Work on combining the placement algorithm with demand forecasting should also be relevant to the SLP if we think of various probabilistic versions of "demand" arising from various attack and environmental scenarios. Costs and profit maximization considerations in this work at Telcordia are also relevant to variants of the SLP where costs of sensors come into play. Other relevant methods we shall explore, that were also utilized in the Telcordia work, are general integer programming approaches, approaches to the integer knapsack cover problem, and approaches to local access network expansion problems. This is a good example of cooperation with an industrial partner in the formulation stages of a project and of such a partnership giving us access to recent relevant algorithmic work.

4.5 DIMACS-DIMATIA-Renyi Project on Combinatorial Methods

This project is in collaboration with two international partners, the DIMATIA Center (Center for Discrete Mathematics, Theoretical Informatics, and Applications) in Prague, and the Renyi Institute (Mathematical Sciences Institute) of the Hungarian Academy of Science). These are world-class research centers at the very top in the world in combinatorial mathematics. The project consists of three separate working groups, one on Algebraic and Geometric Methods in Combinatorics, one on Extremal Combinatorics, and one on Graph Colorings and their Generalizations. For more information, see Section 5.1.

4.6 Project on Computer Science and Decision Theory: Applications of Notions of Consensus

This project, in collaboration with our partner LAMSADE (the Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision, based at Université Paris IX - Dauphine), is organized around modern computer science applications of methods developed by decision theorists, in particular methods involving consensus and associated order relations. It is described in detail in Section 5.4.

4.7 Homeland Security Initiatives

DIMACS' research in homeland security encompasses a variety of projects and initiatives in the center and a leadership role for Rutgers University.

Our internal projects involve the application of modern mathematical, statistical, and computer science approaches to detection of bioterrorist attacks, the monitoring of large streams of text messages, and information security and information sharing. Some of our major efforts in this area are described in the sections on our Monitoring Message Streams Project (Section 4.2), our Bioterrorism Sensor Location Project (Section 4.4), and our special foci in Computational and Mathematical Epidemiology (Section 3.6), Communication Security and Information Privacy (Section 3.8), and Data Analysis and Mining (Section 3.5).

Under the epidemiology special focus, several "working groups" have been actively pursuing projects related to homeland security. Our Adverse Events/Disease Reporting, Surveillance and Analysis working group is working closely with the Centers for Disease Control and Prevention and other agencies in developing methods for detecting bioterrorist attacks and emerging disease threats. A new subgroup of that group is looking at health data privacy issues that are critical in doing the related monitoring. Another working group has the long name Mathematical Sciences Methods for the Study of the Deliberate Release of Biological Agents and their Consequences. Its most recent meeting was devoted to modeling the social disruptions that might result from a bioterrorist or other terrorist attack, and in this direction has worked to create partnerships between mathematicians and computer scientists on the one hand and public health specialists, epidemiologistseconomists, psychologists, sociologists, and geographers on the other hand. Through this working group, Fred Roberts was invited to participate in the Department of Health and Human Services Office of Emergency Preparedness smallpox modeling group, as well as numerous meetings on bioterrorism modeling sponsored by NIH and the MITRE Corporation.

Under the epidemiology focus, still another working group is looking at the analogies between computer viruses and immune systems and biological viruses and immune systems, with the goal of finding useful insight both about defense against biological viruses and against computer viruses. Our working group on Modeling Vaccination Strategies has not yet met, but plans are being made for its first meeting in FY04.

Under the special focus on Communication Security and Information Privacy, we are in the process of starting up working groups on Privacy-preserving Data Mining and on Usable Privacy and Security Software, both of which are very relevant to issues of homeland security. We are also in the process of planning for working groups on Intrusion Detection and Network Management and on Mobile Code Security. Our FY03 workshop on software security was very successful.

Under the special focus on Data Analysis and Mining, we have held very successful workshops on video mining (including video surveillance) and data cleaning. Our working group on Streaming Data Analysis and Mining is working on algorithmic methods for handling massive streams of data like those arising in our Monitoring Message Streams project.

DIMACS has played a leadership role in homeland security research at Rutgers. Fred Roberts chairs the Rutgers University Homeland Security Research Initiative (RUHSRI), whose role is to coordinate the Rutgers homeland security research initiatives. During FY2003, we were instrumental in getting RUHSRI started and creating a database of homeland security research at the university, with the help of the Office of Research and Sponsored Programs. We were also active in planning RUHSRI's first major initiatives: proposals for a homeland security research center, a symposium on homeland security research at Rutgers, and a symposium on homeland security research in New Jersey. For more about these activities, see http://dimacs.rutgers.edu/RUHSRI.

5 International Programs

DIMACS is widely known internationally and hosts hundreds of foreign visitors a year. The center has numerous international partners with whom we have run or are running joint programs, adding this year LAMSADE (the Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision, based at Université Paris IX - Dauphine). With these partners, we have funded joint research projects, jointly sponsored workshops, shared postdocs, exchanged visitors and students, and shared undergraduate research programs.

Our international partners are:

- **DIMATIA**: The Center for Discrete Mathematics and Theoretical Informatics and Applications, based at Charles University in Prague, Czech Republic. With DIMATIA, we have several funded projects to be described below. Our first joint project, funded by NSF, led to a major international conference on the Future of Discrete Mathematics, to a series of specialized workshops, and to an important volume in the DIMACS book series. This was followed up by the projects to be described below.
- **Renyi Institute**: The Mathematical Sciences Institute of the Hungarian Academy of Sciences. We have one funded research project with Renyi and DIMATIA. We have previously partnered with The Erdős Institute, also part of the Hungarian Academy of Sciences in Budapest, to run joint workshops.
- **BRICS**: The Center for Basic Research in Computer Science in Aarhus, Denmark. We have discussed a joint workshop, which is still being planned, and hosted a BRICS graduate student on a long-term visit.
- **HKUST**: The Hong Kong University of Science and Technology. We have started a collaboration with the Computer Science Department at HKUST which in 1998 resulted in a joint workshop. We have previously hosted HKUST visitors for long-term visits and in FY03 shared a postdoc.
- **DISCOM**: The Center for Discrete Mathematics and Computer Science at Hebrew University in Jerusalem. We have been discussing joint programs and exchanging visitors and postdocs informally for a number of years. It is likely that the continued interaction will remain relatively informal for now.
- **Con2Mac**: The Center for Combinatorial and Computational Mathematics at Pohang Institute of Technology in Pohang, Korea. Con2Mac was the winner of a very prestigious national award form the Korean National Science Foundation. We have had several visitor exchanges in the past few years and Fred Roberts (Rutgers) participated in and helped to organize the center's inaugural conference in February 2000 in Pohang as well as several followup conferences.
- **National Chiao Tung University**: This university in Taiwan was the site of a shared workshop in FY03.
- **INRIA - Universite de Nice-Sophia Antipolis**: We have hosted several visitors from this partner in France and we have agreed to plan several joint workshops and exchange students.
- **LAMSADE**: The Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision, based at Université Paris IX - Dauphine. LAMSADE is a new partner with which

we have a research project to be described below and with which we have had several visitor exchanges.

- **Centre de Recherches Mathematique (Montreal)**. In FY2003, we initiated plans for our first joint workshop with this major research center in Montreal.

The International Mathematical Sciences Institutes (IMSI) is a group of mathematical sciences research institutes which meet together at international congresses and have agreed to collaborate with each other. DIMACS has become a member of IMSI and Neil Sloane (AT&T Labs) represented DIMACS at the last IMSI meeting.

These international partnerships have led to a variety of active funded projects which we describe in the rest of this section.

5.1 DIMACS-DIMATIA-Renyi Project on Combinatorial Methods

DIMACS has expanded its international partnership with two leading research centers in Europe to launch a joint research project on combinatorial methods, funded by NSF. DIMACS, DIMATIA in Prague, Czech Republic, and the Renyi Institute in the Hungarian Academy of Science, Budapest, Hungary have joined in a tripartite project involving three research working groups:

- Group I: Algebraic and Geometric Methods in Combinatorics (chair Jaroslav Nesetril, DIMATIA)
- Group II: Extremal Combinatorics (chair Gyula Katona, Renyi)
- Group III: Graph Colorings and their Generalizations (chair Fred Roberts, DIMACS)

The groups will each meet three times in the period between December 2002 and September 2005, once in each location.

The first working group will concentrate on algebraic and geometric methods in combinatorics. This working group will concentrate on two broad areas of research: algebraic methods involving the study of homomorphisms of graphs, with special emphasis on problems arising from statistical physics, and problems of combinatorial geometry. The intersection of combinatorics and statistical physics has been an area of great activity over the past few years, fertilized by an exchange not only of techniques but of objectives. This interconnection was featured in a joint DIMACS-DIMATIA workshop on this topic. (See http://dimacs.rutgers.edu/Workshops/Graph-Morph/) Spurred by computing theorists interested in approximation algorithms, statistical physicists and discrete mathematicians have found a wealth of common ground in probabilistic combinatorics. Close connections between percolation and random graphs, between graph homomorphisms and hard-constraint models, and between slow mixing and phase transition, have led to new results and new perspectives. Of special interest is the use of these connections in understanding typical, as opposed to extremal, behavior of combinatorial phenomena such as graph coloring and homomorphisms.

The working group on extremal combinatorics will concentrate on two general topics, extremal graph theory and extremal problems arising from combinatorial search and testing. Extremal graph theory deals with graphs satisfying specified constraints and optimizing some criterion. We plan to investigate jointly a number of specific questions of current research interest in this field, which has played a fundamental role in the history of graph theory. Topics will include generalized matchings, Turan-type extremal problems, problems related to the Kneser conjecture, sequential testing, and combinatorial group testing.

One of the central topics in graph theory is that of graph coloring. Applications of graph theory have led to fascinating generalizations of the notion of graph coloring, with motivation coming from problems of channel assignment in communications, traffic phasing, fleet maintenance, task assignment, and other practical problems. The working group on graph colorings and generalizations will investigate such generalizations of graph coloring as T-colorings, list colorings, L(2,1)-colorings, and set colorings, with an emphasis on the graph coloring concepts that arise from channel assignment problems. As a second area of emphasis, the group will develop algorithms for graph coloring and its various generalizations.

FY2003 Working Group Meetings

- **DIMACS/DIMATIA/Renyi Working Group on Algebraic and Geometric Methods in Combinatorics**
  - First Meeting Dates: December 2 - 6, 2002
  - Location: Samalova chata, Nova Louka, Czech Republic
  - Organizers: Jaroslav Nesetril, DIMATIA, Jan Kratochvil, DIMATIA

  The first meeting of the Working Group on Algebraic and Geometric Methods in Combinatorics was devoted to the topic of Graph Homomorphisms and Geometric Methods in Graph Theory. We explored topics motivated by the intersection of combinatorics and statistical physics, in particular the close connections between percolation and random graphs, graph homomorphisms and hard-constraint models, and between slow mixing and phase transition. We discussed long range order for graph homomorphisms, Gibbs measures that are invariant under automorphisms, optimum partitioning problems, and list homomorphisms. A general lecture on geometric approaches to lower bounds on certain graph parameters was presented, and we explored the topic of geometric graphs.

- **DIMACS/DIMATIA/Renyi Working Group on Extremal Combinatorics**
  - First Meeting Dates: April 10 - 16, 2003
  - Location: Alfred Renyi Institute, Budapest, Hungary
  - Organizer: Gyula Katona, Alfred Renyi Institute

  Participants included a large number of junior researchers from Rutgers, Prague, and Budapest, as well as several senior combinatorialists from the U.S., Canada, Japan, Denmark, Great Britain, Italy, and the Czech Republic. We also used this opportunity to celebrate the 60th birthday of Bela Bollobas, a leading researcher in extremal combinatorics.

Planned FY2004 Working Group Meetings

- **DIMACS/DIMATIA/Renyi Working Group on Graph Colorings and their Generalizations**
  - First Meeting Dates: October 13 - 15, 2003
  - Location: DIMACS Center, Rutgers University, Piscataway, New Jersey
  - Organizers: Fred S. Roberts, DIMACS Stephen Hartke, Rutgers University

We also plan meetings of the working groups on Algebraic and Geometric Methods in Combinatorics and on Extremal Combinatorics during FY2004, the latter at DIMACS.

5.2 European Union – NSF Strategic Information Technology Workshops

In FY2001, we initiated a new program in which DIMACS is playing a central role in joint European Union - NSF cooperation on information technology. This program accelerated in FY 2003. Through a $344K grant from NSF, DIMACS is supporting a series of nine joint EU-US workshops to be held in Europe over a period of three years. The purpose of these workshops is to further cooperation in the area of information technology between the European Union and the United States and it is hoped that the workshops will lead to the identification of topics for future cooperation. These topics all have very promising commercial applications.

DIMACS is collaborating with the European Research Consortium for Informatics and Mathematics (ERCIM), which is handling the EU/NSF joint workshops from the EU end, and with the Strategic Workshop Review Committee (SWRC), a joint EU/US committee responsible for advising NSF and the European Commission on cooperation.

FY2003 Workshops

- Middleware for Mobile Systems, Vienna, Austria, July 1-2, 2002
  - Middleware for Mobile Systems
  - Meeting Dates: July 1-2, 2002
  - Meeting location: "Freihaus", Vienna University of Technology, Wiedner Hauptstasse 8-10, A-1040 Vienna, Austria
  - Organizers: Lus Rodrigues, University of Lisboa, Departamento de Informtica, FCUL, Portugal and Ravi Prakash, Department of Computer Science, University of Texas at Dallas, Richardson, TX, USA
  - Website: http://www.di.fc.ul.pt/~ler/ercimnsf/

We worked with the U.S. scientific community to identify topics for future workshops, and developed three topics for workshops that have been approved by NSF and await approval by the EU.

5.3 International REU program

For the summers of 2002 and 2003, we received NSF funding to bring five students to DIMACS from our sister site DIMATIA at Charles University in Prague, Czech Republic, and to send five of our students to DIMATIA for a three-week research program after their stay at DIMACS as part of the DIMACS International Research Experiences for Undergraduates Program.

For more information about the REU program see Section 6.3.

5.4 DIMACS-LAMSADE Project on Computer Science and Decision Theory: Applications of Notions of Consensus

This project, in collaboration with our partner LAMSADE (the Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision, based at Université Paris IX - Dauphine), is organized around modern computer science applications of methods developed by decision theorists, in particular methods involving consensus and associated order relations. The project seeks to explore the connections between computer science and decision theory, develop new decision-theory-based methodologies relevant to the scope of modern CS problems, and investigate their applications to problems of computer science and also to problems of the social sciences that could benefit from new ideas and techniques. The project will feature exchange visits of graduate students and junior researchers and several workshops that will introduce a broader community to the topics of the collaboration.

The project will emphasize computer science problems that arise in meta-search (combining the outputs of several search engines); collaborative filtering (using learning algorithms to make recommendations for books, movies, etc.); finding centrally located items in large databases, in particular biological ones and ones that arise in homeland security applications; combining ratings of software or hardware; in distributed computing when some processors fail and we seek to reach "agreement" among the remaining processors; finding efficient ways to compute consensus functions; and learning about individuals' preferences as they are revealed gradually in situations of economic cooperation and competition using the Internet. The project will also seek to use the motivating CS applications to improve the methods of decision theory and apply new ideas to decision making problems of the social sciences.

The project will have impact well beyond the small number of people participating in the exchange visits by stimulating a center-to-center exchange that will involve a large number of scientists associated with the two centers in the related scientific discussions and interchanges. Moreover, the impact will be broadened through the workshops that will introduce many of those both associated with and outside the two collaborating centers to this new field. The results should be broadly useful in emerging information technology applications, in areas of economics and political science where methods of decision theory have traditionally been applied, and in new areas of application of decision theory such as to problems of epidemiology and bio-terrorism.

The P.I.'s are Fred Roberts on the DIMACS side and Alexis Tsoukiàs on the LAMSADE side.

In FY03, we planned the project and submitted a proposal to NSF for funding it. We have been informed that it will be funded.

6 Integration of Research and Education

DIMACS believes strongly that research and education should go hand in hand. From the early days of our "science and technology center award," we saw our research mission as fundamentally intertwined with our educational mission, and developed programs that integrated research and education from the postdoctoral level to the elementary school level. Our K-12 teacher programs, heavily informed by research at DIMACS on discrete mathematics, quickly received national acclaim and have had major impact on the development of curriculum "standards" in New Jersey and nationally. Literally tens of thousands of teachers nationwide have been impacted by our programs. Our "Reconnect" program has impacted many teaching faculty at 2- and 4-year colleges around the country and our programs for undergraduate research have included a pioneering international "research experiences for undergraduates" program, later mirrored in a unique "research experiences for teachers" program in the DIMACS Connect Institute. Here, we review the many programs at DIMACS that integrate research and education.

6.1 Postdoctoral Training Program

One of DIMACS' goals is to involve outstanding junior researchers in the life of the center. Our postdoc program is designed to accomplish this. We have leveraged our funds for support of postdocs by sharing postdocs with the Institute for Advanced Study, obtaining grant and industrial support for postdocs, and including in our program postdocs appointed by our industrial partners. This year, we shared postdocs with AT&T Labs, NEC Labs, and the Hong Kong University of Science and Technology, and included in our postdoc program a researcher at NEC Labs. We have instituted regular meetings with our postdocs to discuss the job market. Continuing procedures developed in previous years, each postdoc is assigned a faculty mentor. Moreover, because we want to train our postdocs to take an interest in educational activities and in outreach beyond the mathematical sciences, we ask each postdoc to take part in such activities. Our postdocs meet this requirement in a variety of ways, such as helping with our K-12 and "Reconnect" programs or mentoring in our REU program. Our postdoctoral fellows from past years have been remarkably successful in obtaining positions, both in academics and in industry. They have also rapidly risen to the top of their profession. A postdoctoral fellowship at DIMACS remains a very highly-sought-after award for junior researchers in discrete mathematics, theoretical computer science, and their applications.

A list of the DIMACS 2002-2003 postdoctoral fellows, the program with which they were affiliated, their research interests, and their present affiliations follows:

- Michael R. Capalbo
  DIMACS/Institute for Advanced Study Postdoc at DIMACS
  Ph.D. in Mathematical Sciences, The Johns Hopkins University, 2001
  Last Year: DIMACS/Institute for Advanced Study Postdoc at IAS
  Country of Citizenship: USA
  Research Interests: Explicit construction of combinatorial structures, graph theory, combinatorial optimization

- Amit Chakrabarti
  DIMACS/Institute for Advanced Study
  Ph.D. in Computer Science, Princeton University, 2002

Last Year: Preceptor, Princeton University
Country of Citizenship: India
Research Interests: Complexity theory with an emphasis on lower bounds results in non-uniform models of computation; approximation algorithms, especially for combinatorial optimization problems.

- Graham Cormode
DIMACS Computational Epidemiology Postdoc
Ph.D. in Computer Science, University of Warwick, 2002
Last Year: Teacher and Grader, Case Western Reserve University and University of Warwick
Country of Citizenship: United Kingdom
Research Interests: Algorithms and complexity; streaming algorithms, data mining, sequence comparison.

- Irit Dinur
NEC Research Institute
Ph.D. in Computer Science, Tel-Aviv University, 2001
Last Year: Member, School of Mathematics, Institute for Advanced Study
Country of Citizenship: Israel
Research Interests: cryptography, privacy

- Ofer Melnik
DIMACS/AT&T Labs - Research, Data Analysis and Mining Postdoc
Ph.D. in Theoretical Neuroscience, Brandeis University, 2000
Last Year: Sloan Fellow for Theoretical Neuroscience, Brandeis University
Country of Citizenship: Israel/USA
Research Interests: Data Mining, machine learning, complexity, data analysis

- Kobbi Nissim
DIMACS/NEC Research Institute
Ph.D. in Computer Science, Weizmann Institute of Science, 2001
Last Year: teacher of graduate course, The Open University and Graduate Student, Weizmann Institute of Science
Country of Citizenship: Israel
Research Interests: Efficient and secure multiparty protocols, secure computation of approximations, argument systems for NP, approximability and inapproximability of NP-hard functions, applications of the PCP theorem

- Xuerong Yong
DIMACS/Hong Kong University of Science and Technology Postdoc
Ph.D. in Computer Science, Hong Kong University of Science and Technology, 2002; Ph.D. in Mathematics, Xian Jiaotong University, 1991.
Last Year: Ph.D. student, Hong Kong University of Science and Technology
Country of Citizenship: PR of China
Research Interests: Shannon theory of two-dimensional constrained codes; foundations of computer science

6.2 Graduate Student Programs

Graduate students are a central part of DIMACS' program. Our *local* graduate student program includes small DIMACS awards for travel and software to support the research of local students, small awards for summer and winter break salary to support research projects, and jobs and projects for which graduate students are paid. It was continued during summer 2002 and winter 2002-2003. We supported 23 students this way. The response to this program has been highly enthusiastic. We are also very pleased at the number of DIMACS researchers, including individuals from our industrial partners, who are participating in this program as research mentors. This program has led to rather impressive scientific payback, as demonstrated by scientific results obtained by the students.

DIMACS also supports graduate students as Research Assistants under our special foci and research project grants.

DIMACS sponsors a graduate student organization revolving around discrete mathematics and theoretical computer science and funds a seminar series organized by and for the graduate students. It also supports a graduate student organization in electronic commerce.

We have also continued our programs of sponsoring DIMACS visits by *non-local* graduate students and continued our long-standing policy of providing a supplemental allocation to each workshop for support of non-local graduate students. The impact of our program on non-local students has been very high and the research of some of our non-local graduate student visitors has also been stimulated and redirected by our programs.

6.2.1 Local Graduate Student Program

Scientific results of supported graduate students are included in Sections 3 and 4 and selected publications by supported students are listed below. Here, we provide a list of those graduate students supported, their affiliation, dates of support, and associated DIMACS program.

DIMACS-Supported Graduate Students, 2002 - 2003

Summer 2002

- Gabriela Alexe
  RUTCOR, Rutgers University
  Related to Data Analysis and Mining and Computational Molecular Biology

- Sorin Alexe
  RUTCOR, Rutgers University
  Related to Data Analysis and Mining and Computational Molecular Biology

- Xuhui Ao
  CS, Rutgers University
  Related to Next Generation Networks

- Eva Curry
  Mathematics, Rutgers University
  Related to Mathematics and Foundations of Information and Computer Science

- Navin Goyal
  CS, Rutgers University
  Related to Mathematics and the Foundations of Information and Computer Science, Computational Intractability, and Data Analysis and Mining

- Chan Su Lee
  Elec. & Comp. Eng. (CAIP), Rutgers U.
  Related to Computational Geometry

- Ying Liu
  RUTCOR, Rutgers University
  Related to Computational Geometry and Data Analysis and Mining

- David Nacin
  Mathematics, Rutgers University
  Related to Graph Theory and Algorithms

- Swati Sharma
  Elec. & Comp. Eng., Rutgers University
  Related to Data Analysis and Mining

- Akshaya Kumar Vashist
  CS, Rutgers University
  Related to Computational Molecular Biology

- Anthony Ian Wirth
  CS, Princeton University
  Related to Data Analysis and Mining

Winter 2002-2003

- Diogo Vieira Andrade
  Rutgers University / RUTCOR
  Related to Graph Theory and Algorithms

- Tiberius Bonates
  Rutgers University - RUTCOR
  Related to Combinatorial Optimization and Large Scale Discrete Optimization

- Xiaoling Hou
  RUTCOR, Rutgers University
  Related to Data Analysis and Mining

- Xiaolei Huang
  Rutgers University - CS
  Related to Computational Geometry and Applications

- Marcin Kaminski
  Rutgers University - RUTCOR
  Related to Graph Theory and Boolean Functions

- Tongyin Liu
  RUTCOR, Rutgers University
  Related to: Computational Intractability

- Ying Liu
  RUTCOR, Rutgers University
  Related to Data Analysis and Mining and Computational Geometry

- Jin Ma
  Rutgers University - CS
  Related to Data Analysis and Mining and Computational Molecular Biology

- Marcelo Mydlarz
  CS, Rutgers University
  Related to Next Generation Networks

- Igor Zverovich
  RUTCOR, Rutgers University
  Related to Data Analysis and Mining

Students Supported as RA's on Projects and Special Focus Grants

- Andrei Anghehescu
  CS, Rutgers University
  Monitoring Message Streams Project

- Dmitry Fradkin
  CS, Rutgers University
  Monitoring Message Streams Project

- Jaewok Joo
  Physics, Rutgers University
  Special Focus on Computational and Mathematical Epidemiology

Publications that resulted from DIMACS Support of Graduate Students, 2002 - 2003

- G. Alexe, S. Alexe, P.L. Hammer, B. Vizvari. "Feature selection for logical analysis of microarray data," Third International Conference for the Critical Assessment of Microarray Data Analysis, Nov. 14-15, 2002, Durham, NC, submitted.

- G. Alexe, S. Alexe, D. Axelrod. E. Boros, P.L. Hammer. "Combinatorial analysis of breast cancer data from image cytometry and gene expression microarrays," Journal of Computational Biology, submitted.

- G. Alexe, S. Alexe, M. Reiss, P.L. Hammeran. "Ovarian cancer detection by logical analysis of proteomic data," in preparation.

- G. Alexe, S. Alexe, P.L. Hammer, D. Weissman. "Logical analysis of leukemia microarray data," in preparation.

- S. Alexe, E. Blackstone, P.L. Hammer, H. Ishwaran, M. S. Lauer, C. E. Pothier Snader. "Coronary Risk Prediction by logical analysis of data," Annals of Operations Research, 2002.

- A. Anghelescu, E. Boros, D. Fradkin, D. D. Lewis, V. Menkin, D. J. Neu, K.B. Ng., and P. Kantor, "Prospective data fusion for batch filtering," May 2003, submitted.

- A. Anghelescu, D.D. Lewis, V. Menkov, and P. Kantor, "Training a Rocchio classifier for adaptive filtering," TREC 2002.

- A. V. Anghelescu and I.B. Muchnik, "Optimization of SVM in a space of two parameters: Weak margin and intercept," May 2003, submitted to J. Intelligence Community Research.

- A. Anghelescu and I. Muchnik, "Combinatorial clustering for textual data representation in machine learning models," November 2002, submitted.

- X. Ao and N. Minsky, "A scalable mechanism for multi-enterprise coalition access control," in preparation.

- X. Ao and N. Minsky, "A law-hierarchy based access control model for distributed systems," in preparation.

- T. Bonates, "Solving 0/1 integer linear programs by a pseudo-boolean approach," DIMACS Report, 2003.

- X. Hou, "A data mining problem in stochastic programming," in preparation.

- C. Lee, "Physically accurate haptic rendering of elastic object for a haptic glove," DIMACS Report, 2002.

- T. Liu and Y. Liu, "On generating three-connected planar graphs," DIMACS Technical Report 2002-56, 2002.

- Y. Liu, "Lower bounds of the size of binary space partition of rectangles," in preparation.

- J. Ma, "Clustering analysis on metabolic process research," DIMACS Report, 2003.

- D. Rautenback, I. E. Zverovich, I. I. Zverovich, "Extensions of α-polynomial classes," Australas. J. Combin. **26**, 265-271, 2002.

- I. E. Zverovich, "Extensions of hereditary classes with substitutions," Discrete Appl. Math, to appear.

- I. E. Zverovich, "The domination number of K_p, P_5)-free graphs," Australas. J. Combin. **27**, 95-100, 2003.

- I. E. Zverovich, "Stabex method for extension of α-polynomial hereditary classes," Euro. J. Oper. Res., to appear.

- I. E. Zverovich, "Unique domination and domination perfect graphs," DIMACS Technical Report 2002-43, 2002.

- I. E. Zverovich and I. I. Zverovich, "Ratio of generalized parameters," DIMACS Technical Report 2002-47, 2002.

- I. E. Zverovich and I. I. Zverovich, "An improvement of Gyárfás' bounds on the maximal order of a minimal forbidden induced subgraph for (1, 2)-split graphs," RUTCOR Research Report RRR 37-2002, 2002.

- I. E. Zverovich and V. E. Zverovich, "Basic perfect graphs and their extensions," DIMACS Technical Report 2003-12, 2003.

6.2.2 Celera Genomics/Applied Biosystems Graduate Student Awards 2002-2003

The DIMACS Celera Genomics/Applied Biosystems Graduate Student Award Competition was held for the third time in FY03. The award is given each year of the DIMACS Special Focus on Computational Molecular Biology and is aimed at recognizing outstanding achievements of junior researchers in the field. The winner was Ziv Bar-Joseph, Massachusetts Institute of Technology, Mentor: David Gifford. He gave his talk at the Workshop on Data Mining Techniques in Bioinformatics: October 30 - 31, 2003. His title was "Structure of Gene Expression".

6.2.3 Visiting Graduate Students

The kind of impact our program for visiting graduate students can have is perhaps best illustrated by comments received from Janet Best, a graduate student in Mathematics at Cornell University who attended tutorials and conferences as part of the DIMACS Special Focus on Computational Epidemiology. Her comments summarize the impact of the project this way:

> "I was impressed at the number of strong women role models in the DIMACS programs; providing those examples and giving us opportunities to form/strengthen contacts with these more established women can be crucially important to us women who are still making our way. It was particularly helpful that one of the workshops was co-organized by a strong woman: we need to see that women are leaders and full participants. Further, she was an extremely adept organizer and enormously facile in introducing us to key people in a comfortable manner. There were also a goodly number of women grad students; I found it valuable to get perspective from the UK women.
>
> I personally developed and/or strengthened significant contacts with researchers from UK, Canada, Netherlands, Israel, Australia. I'm now internationally engaged in communication with some of these individuals.
>
> I was exposed to some new (to me) areas and ideas, including engineering, computer science, and data-intensive issues of a sort that I don't often encounter at other mathematical meetings. Examples: questions involving surveillance data and detecting patterns (or deviation from normal patterns), automated cataloging of papers in medical databases, other algorithmic issues. I think the workshops were helpful in getting us to think about how the pure math relates to and can be used in conjunction with computer science.
>
> I thought the talks about vaccination decisions and in particular about smallpox were a great example of creating informed citizens of us. The question of vaccination strategies, in particular with regard to smallpox terrorism, has been front page news for months in this country. These talks really prepared us to appreciate the issues involved and to be able to think critically about and to debate the plans and decisions of the government."

6.3 Research Experiences for Undergraduates (REU) Program

DIMACS has run a Research Experiences for Undergraduates (REU) program since 1993. The program introduces mathematically talented students to research in DIMACS-related fields under the guidance of research mentors, both to discover new results and to provide a basis for making career and graduate school decisions. Students from all over the U.S. come to work one-on-one with DIMACS researchers for eight weeks during the summer and enjoy the stimulating and supportive environment at our center. At the end of the summer, students present their work in front of DIMACS members and visitors and write a short report on their results. For the summers of 2002 and 2003, we also received NSF funding to bring five students to DIMACS from our sister site DIMATIA at Charles University in Prague, Czech Republic and to send five of our students to DIMATIA for a three-week research program after their stay at DIMACS.

A total of 26 undergraduates from around the country plus Prague participated, including students in the VIGRE REU program of the Department of Mathematics at Rutgers, who were given office space at DIMACS and integrated into the DIMACS REU program. Here is a list of the students, their mentors, and their projects.

- Daniel Boylan, Harvey Mudd College, Claremont, CA, worked with Michael Rothkopf, RUTCOR, on Modeling Information Cost in Auctions.

- Jakub Cerny, Charles University, Czech Republic, worked with Janos Komlos, Department of Mathematics, on some open problems related to graph theory.

- Zdenek Dvorak, Charles University, Czech Republic, worked with Janos Komlos, Department of Mathematics, on Non-zero (mod k) dominating sets.

- Inessa Epstein, Rutgers University, New Brunswick, NJ, worked with Peter Hammer, RUTCOR, on Logical Analysis of Data.

- Paul Gross , Rose-Hulman Institute of Technology, Terre Haute, IN, worked with Eric Allender, Computer Science Department, on Extracting a Sequential Algorithm for Solving the Planarity Problem.

- Andrew Hodges, Manchester College, North Manchester, IN, worked with Wilma Olson, Department of Chemistry and Chemical Biology, on the analysis of DNA conformations.

- Steven Jaslar , Rutgers University, New Brunswick, NJ, worked with Vladimir Gurvich, RUTCOR, on generalizations of the Stable Marriage problem.

- Vit Jelinek, Charles University, Czech Republic, worked with Janos Komlos, Department of Mathematics, on open problems related to graph theory.

- Daniel Krasner , University of California at Berkeley, Berkeley, CA, worked with Endre Boros, RUTCOR, on generalizations of the Stable Marriage problem.

- Michael Lesnick, Brown University, Providence, RI, worked with Avy Soffer, Department of Mathematics, on A problem in Quantum Control: Solving an Instance of the Schrodinger Equation.

- Rahul Malhotra, Rutgers University, New Brunswick, NJ, worked with Peter Landweber, Department of Mathematics, and Brian Lins, Graduate Student, Department of Mathematics, on Eigenvalues of Sums of Hermitian Matrices.

- John McClain, Rutgers University, New Brunswick, NJ, worked with Avy Soffer, Department of Mathematics, on Fermis Golden Number for An Interacting Spin System.

- Hoa Nguyen, Rutgers University, New Brunswick, NJ, worked with Kobbi Nissim and Ofer Melnik, DIMACS, on Explaining Data in High Dimensional Space.

- Stanislav Ostrovsky, Rutgers University, New Brunswick, NJ, worked with Carlo Lancellotti, Department of Mathematics, on Synchronization Phenomenon in Dynamical Systems of Nonlinear Coupled Oscillators.

- Pavel Podbrdsky, Charles University, Czech Republic, worked with Janos Komlos, Department of Mathematics, on some open problems related to graph theory.

- Yuki Saka , University of California at Berkeley, Berkeley, CA, worked with Leonid Khachiyan, Department of Computer Science, on Generalized Cuts and Spanning Subgraphs.

- Thomas Schneider, Rose-Hulman Institute of Technology, Terre Haute, IN, worked with David Madigan, Department of Statistics, on Text Categorization Using Bayesian Networks.

- Samuel Stechmann, University of ST. Thomas, St. Paul, MN, worked with Matthew Leingang, Department of Mathematics, on Hilbert's Third Problem in Hyperbolic Geometry.

- Rupert Venzke, University of Pittsburgh, Pittsburgh, PA, worked with Friedrich Knop and Shawn Robinson, Department of Mathematics, on Representation Theory.

- Ursula Whitcher , Swarthmore College, Swarthmore, PA, worked with Rebecca Wright, DIMACS, on Probabilistic Quorum Systems.

Selected Research Accomplishment

Counterexample to a Conjecture about the Stable Marriage Problem. Undergraduates Steven Jaslar (Rutgers) and Daniel Krasner (UC Berkeley) settled a conjecture about the stable marriage problem. Here, there are n men and n women, with each man giving a preference ordering of all the women and each woman giving one of all the men. A matching of men and women is stable if there is no pair (m, w) such that man m prefers woman w to his match and w prefers m to her match. Gale and Shapley found a polynomial-time algorithm that finds a stable matching for any set of preferences. The problem is of serious practical interest, used in matching medical schools with applicants for residencies and job applicants to employers at professional society job fairs. Generalizing the problem, Alkan considered groups of n men, n women, and n cats, forming families of one man, one woman, and one cat. Each man has a preference ordering of all woman-cat pairs; each woman of all man-cat pairs; and each cat of all man-woman pairs. Here (m, w, c) is a breaking triple if m prefers (w, c) to the pair he is with, w prefers (m, c) to the pair she is with, and c prefers (m, w) to the pair it is with. A matching is stable if there are no breaking triples. Alkan showed that a stable matching does not always exist. However, Danilov proved that such a matching always exists in the case of some "acyclic" preferences. He introduced lexicographical-cyclic preferences and conjectured that there was always a stable matching in this case. Jaslar and Krasner found a counterexample to this conjecture. Their examples also show that for all $n > 2$, there is a cyclic preference that does not admit a stable matching. These results were presented at the Prague Midsummer Combinatorics Conference in August 2002.

6.4 The Reconnect Program: Reconnecting Teaching Faculty to the Mathematical Sciences Enterprise

The Reconnect Program exposes faculty teaching undergraduates to the mathematical sciences research enterprise by introducing them to a current research topic relevant to the classroom through a series of lectures by a leading expert and involving them in writing materials useful in the classroom. Participants have the possibility of following up by preparing these materials for publication in the DIMACS Educational Modules Series.

This program, organized primarily around summer conferences, offers the opportunity for junior faculty as well as mid-level and senior faculty to advance to research questions in a new area of the mathematical sciences. Participants also acquire materials and gain ideas for seminar presentations and for undergraduate research projects.

The Reconnect Program is also aimed at reconnecting faculty to the mathematical sciences research enterprise by involving them in a leading research center. There are opportunities to follow up after the conference by getting connected to DIMACS researchers and other DIMACS programs throughout the year.

The Reconnect Program is funded by a four year award from NSF. The program was originally funded by NSF for three years and ran at DIMACS during the summers of 1998, 1999 and 2000. Due to the success of the program and at NSF's recommendation, the program has been redesigned to include a "satellite" component whereby the programs run at DIMACS become the training ground for other college faculty to observe the program and subsequently run a similar program at their own institutions in the succeeding summer. There are currently six institutions that have agreed to run Satellite Reconnect programs: Salem State College and the Illinois Institute of Technology in summer 2003, Lafayette College and St. Mary's College in summer 2004, and Montclair State University and Spelman College in summer 2005. We have begun discussions with Morgan State University to run a Satellite Reconnect Conference during summer 2006. During FY03, we worked extensively with the organizers of the Salem State and Illinois Institute of Technology conferences to help them prepare for their summer 2003 conferences.

The first program for the Satellite Reconnect Program was run in August 2002 on the topic "Voronoi Diagrams - Properties, Algorithms and Applications." The first satellite conference was run in June 2003 on the topic "Some Current Problems in Coding Theory."

FY2003 Conferences

- **Voronoi Diagrams – Properties, Algorithms and Applications**
  - Dates: August 11-17, 2002
  - Location: DIMACS Center, Rutgers
  - Organizers: Rochelle Leibowitz (Wheaton College), Fred Roberts (Rutgers)
  - Attendance: 20

  The Voronoi Diagram of n sites is a subdivision of the plane into n regions, one per site. Each site's region consists of all points in the plane closer to that site than to any of the other $n-1$ sites.

  It is amazing how useful this simply described diagram is. The Voronoi Diagram and its geometric dual, the Delaunay Triangulation, have proved to be two of the most important data structures in Computational Geometry. The Voronoi Diagram and its variants (different metrics, higher dimensions, sites which are segments or polygons instead of points, etc.) have been rediscovered many times in literally dozens of fields, including biology, crystallography, geology, metallurgy, mesh generation, curve and surface reconstruction, meteorology, mathematics, robotics, geography, and marketing. A book-length survey of the types of Voronoi Diagrams and their applications in a variety of fields was written by Okabe, Boots, and Sugihara.

Voronoi diagrams and their applications are an active field of research. In recent years between 3 and 6 of the approximately 40 papers accepted to the annual Symposium on Computational Geometry have been on Voronoi diagrams or their applications.

One of the applications that is the subject of a flurry of recent papers is a group of algorithms for curve and surface reconstruction. The algorithms take a sample of points from a curve in 2-D or a surface in 3-D (perhaps from a medical image or a machine that records 3-D points in space) and construct a representation of the curve or surface sampled.

Another important application of Voronoi diagrams that is still an active area of research is mesh generation. However, not all recent work is applied. Researchers are still exploring fundamental properties of these diagrams and new variations on the diagram.

The conference looked at the properties of Voronoi diagrams that make them so widely applicable, discussed the advantages and drawbacks to the various algorithms for computing them, and looked at a number of specific applications, including the surface reconstruction algorithm discussed above.

Much of the material of the conference covered can be incorporated into undergraduate algorithms courses. The Voronoi diagram (or Delaunay Triangulation) can be constructed via a number of approaches: iterative algorithms, divide-and-conquer, and sweepline methods are among the approaches used. This variety of approaches leads to interesting questions in both theoretical and experimental analysis of algorithms. The topic can also be a large component of a topics course devoted to Computational Geometry.

The properties of the diagram provide an interesting test bed for showing the usefulness of mathematical proofs and give students practice in constructing proofs. The properties are non-trivial and sometimes surprising, but none requires a level of mathematical understanding beyond that gained in high school geometry. Thus they are both challenging and accessible to undergraduate students.

The wide variety of applications and concrete visual representation of the diagram make it a good example for undergraduate courses.

The principal lecturer was Scot Drysdale, Dartmouth College, with a special lecture presented by Chuck Biehl of The Charter School of Wilmington, DE.

- **Some Current Problems in Coding Theory**
  - Dates: June 15-21, 2003
  - Location: Salem State College, Salem, MA
  - Organizers: Maura Murray (Salem State College), Fred Roberts (Rutgers University)
  - Attendance: 20

Whenever information is transmitted across a channel, errors are bound to occur. It is the goal of coding theory to find efficient ways of adding redundancy to the information so that the errors can be detected and even corrected. The first few lectures of this workshop were devoted to a general introduction to the theory of error-correcting codes. The remaining lectures were devoted to a survey of some current topics in coding theory research, including a discussion of several recent papers on the subject.

Because of its immediate real-world applications, coding theory is a topic with great appeal to undergraduates. It can be worked into undergraduate linear algebra, abstract algebra, and number theory courses.

The principal lecturer was Judy Walker of the University of Nebraska.

Selected Research Results

Solution of a Problem in Surface Reconstruction. The Summer 2002 Reconnect Conference at DIMACS featured lectures by Scot Drysdale (Dartmouth) on Voronoi Diagrams (Properties, Algorithms, and Applications). Participant Darren Narayan (Rochester Institute of Technology) was particularly interested in Drysdale's lecture on surface reconstruction. After reading some of the papers suggested by Drysdale, he was led to the web page of Dr. Hugues Hoppe of Microsoft Research and became very intrigued by a problem from graph theory that was posed in the paper. Narayan started working on this problem at the conference and solved it within a week of returning to his school. Surface reconstruction has attracted considerable interest from researchers in the areas of computer graphics, computer aided design, and modeling. A procedure introduced by Eck and Hoppe involves the reconstruction of a tensor product B-spline surface from a set of unorganized points. One of the many important contributions of their paper was the application of a combinatorial optimization method used to minimize distortion during the conversion from a triangle domain to a quadrilateral one. More precisely their process combines pairs of adjacent triangle faces to form quadrilaterals, and it was asked whether or not a complete pairing of the triangles always exists. Narayan used methods from graph theory to establish the affirmative answer.

6.5 DIMACS Connect Institute (DCI)

We ran the third DIMACS Connect Institute (DCI) in summer 2002. The DCI program was funded by NSF and the AT&T Foundation for three years. DCI replaces the DIMACS Research and Education Institute (DREI) which ran for the previous four years. DCI is a five year program, with the first three years supported by an award from NSF and the last two years supported through a combination of funds from the DIMACS partners and corporate supporters, in particular the AT&T Foundation. In DCI we aim to give teachers a true research experience that, in turn, can be reflected in the excitement they bring to their classroom teaching. In the first week of DCI we have a research workshop, an educational program with lectures (called the Classroom Experience), and first research experiences through "scavenger hunts" requiring teachers to get information that is only available in library volumes, on the Internet, on MathSciNet, and through discussions with researchers attending the research workshops associated with DCI. Once the research workshop is over, the teachers remain at DIMACS for two more weeks. Here, they continue into the second week with lectures and then move mid-week on to the Research Experience, where they work with researcher mentors on research projects, development of classroom implementations, and discussion of written materials. We are preparing the teachers for significant in-class follow-up during the subsequent academic year.

During that academic year, teachers are organizing projects in their schools, both in-classroom projects and out-of-classroom projects. We arrange classroom visitations, utilizing teachers to visit other teachers and also Lead Teachers, mentors, and researchers to do some visiting, and we provide written feedback after a visit.

A major innovation that puts more emphasis on student/teacher teaming was our third high school student research conference in May 2003. Seventy-eight high school students from across the United States presented results from original mathematical research during the research conference, "Graph Theory and its Applications to Problems of Society." The outstanding nature of the students' research and the quality of their presentations were truly exceptional and their enthusiasm was infectious. One of the reasons that this conference was so successful was that graph theory is a natural subject for student research. It is easy to understand without a sophisticated mathematical background and easy to come to the frontiers of knowledge.

Our DCI teachers also returned for two weeks in a second summer in 2003. They participated in a second research project, made presentations to the new group of teachers on their in-classroom and out-of-classroom experiences, and worked on their written materials.

For the FY2003 program Linda Lesniak (Drew University) was both overall program director and director of the research program. Barry Tesman (Dickinson College) was the education program director. L. Charles Biehl (teacher, The Charter School of Wilmington, Wilmington, Delaware), Mihaela Nicolescu (teacher, Collegiate School, New York, New York), and Patrick Carney (teacher, The Montclair Kimberley Academy, Montclair, New Jersey) were the lead teachers for the education program. The lead teachers assisted Tesman in running homework sessions, teacher implementation sessions and overall program planning, and assisted the mentors with the research projects. Mentors for the research portion of the program were Nancy Eaton (University of Rhode Island) and Evan Wantland (Warren Wilson College) for the returning teachers from the 2000 and 2001 programs and John Gimbel (University of Alaska at Fairbanks), Dean Hoffman (Auburn University) and Peter Slater (University of Alabama at Hunstville) for the 2002 cohort of teachers.

There were two aspects of the education program. The Classroom Experience consisted of ten total lectures given by Barry Tesman during the first week and a half. During the Research Experience, the teachers engaged in research projects under the guidance of the mentors and lead teachers. A successful part of the education program consisted of teacher presentations based upon a topic introduced by a research program participant. Also, a library of books and materials was set up for the teachers and researchers to use as reference materials and for developing professional plans and presentations. As in previous years, All Institute speakers spent time with the teachers after their presentations to answer questions, etc.

There were 22 teachers participating in the educational component and 41 researchers participating in the research workshop of the summer 2002. During the Research Experience, there were 10 teachers returning from the 2001 program and 3 teachers returning from the 2000 program.

Because of the success of the theme "Graph Theory and its Applications" in our final two years of DREI, we have kept this theme. The topic lends itself beautifully to beginning research experiences for teachers and is ideal for in-classroom and out-of-classroom projects for students of varying abilities. In FY2003, the research workshop at DCI concentrated on the theme "Labelings and Numberings of Graphs."

In FY2003, we also started planning for an international component to DCI. This resulted in a summer 2003 "international" DCI program, including a group of teachers from Romania.

Workshops and Conferences

- **DCI Research Workshop on Labelings and Numberings of Graphs**
  - Dates: July 14-19, 2002
  - Location: DIMACS, Rutgers University
  - Organizers: Peter Slater, University of Alabama, Huntington
  - Attendance: 35

A labeling of a graph is a function f that assigns to each vertex a nonnegative integer and which induces a function h from edges to nonnegative integers, where the value of $h(uv)$ for uv in $E(G)$ depends on $f(u)$ and $f(v)$. For example, $h(uv)$ might be the absolute value of the difference of $f(u)$ and $f(v)$. If f and its induced function h are one-to-one functions, then f is called a numbering. Typical of the graph labeling/numbering problems studied in this workshop are the bandwidth labeling problem, where one seeks to minimize the maximum value of h, with an application to optimal data storage, and the graceful numbering problem with applications to such areas as coding theory, astronomy, and x-ray crystallography. There was a lot of interest in $L(2,1)$-labelings that arise from channel assignment problems.

- **DCI High School Student Research Conference**
  - Dates: May 10, 2003
  - Location: DIMACS, Rutgers University
  - Organizers: Linda Lesniak (Drew University), Fred Roberts (Rutgers University)
  - Attendance: 78

  The conference featured presentations by high school students on their research on topics in graph theory and its applications to the problems of society. The research papers presented at the conference resulted from intensive academic year student research experiences following the summer 2002 programs of research experiences for teachers as part of the DIMACS Connect Institute (DCI). Awards were presented for the research presentations.

Selected Research Results

Modeling Smallpox Response Strategies. In the summer of 2002, Fred Roberts (Rutgers) gave a lecture on "Challenges for Discrete Mathematics and Theoretical Computer Science in Homeland Security" to a group of high school teachers attending DCI. In the lecture, he talked about using mathematical models to help policymakers make decisions about prevention of or response to smallpox attacks and described some simple graph-theoretical models he had been working on. All of the teachers responded enthusiastically, many took the ideas back to their schools, and one, Chuck Biehl of the Charter School of Wilmington (Delaware), encouraged two students to do a project on this topic. The result was a quite-sophisticated computational model, prepared by 12th graders Chintan Hossain and Hiren Patel, designed to compare the effect of different vaccination strategies for smallpox. A beautiful paper prepared by Hossain and Patel summarizes their model and conclusions drawn from it. Hossain and Patel won first place at the New Castle Science Fair for this work and an invitation to a national student project competition. Later, Hossain and Patel made a presentation to the DIMACS workshop on Spatio-temporal and Network Modeling of Diseases in April 2003 and this was one of the highlight presentations of the research workshop.

6.6 K-12 Programs

DIMACS' programs for precollege teachers and students go back to the very first year of the center. Through them, we have had significant influence on the precollege curriculum, the philosophy of mathematics education, and on the teaching and careers of thousands of teachers and students. In this section, we discuss the two main precollege programs of the center, the Leadership Program for K-12 teachers and the Young Scholars Program for high school students, as well as a variety of other programs and products and the impact that they have had.

6.6.1 Leadership Program in Discrete Mathematics

The Leadership Program in Discrete Mathematics (**LP**) has evolved through four phases: Phase I for high school teachers (1989-1991), Phase II for high school and middle school teachers (1992-1994), Phase III for K-8 teachers (1995-2001), and Phase IV for K-8 teachers at satellite locations (2001-present). In this fourth phase, a dissemination phase, mathematicians and computer scientists replicate the Phase III K-8 program at their own institutions.

In Phase III, under a grant from NSF and support from DIMACS and from its STC grant, teachers attended a two-week program during the summer, attended four follow-up days (six were available) during the year, and then returned for a week the following summer. In Phase IV, this program is being replicated at other sites with a grant from the Educational Foundation of America, and grants from state Departments of Education.

The content material for the first summer (two weeks) includes applications of graphs and patterns in numbers and geometry. Participants begin by coloring maps and end by generating spirolaterals and fractals. These topics are presented during morning workshops lasting two and a half hours during which participants are engaged in mathematical activities which illustrate and highlight the key conceptual issues discussed in the workshop.

Study group sessions follow the morning workshop. During this time, the teachers are broken into grade-heterogeneous groups to work on problems related to the morning workshops. These problems will be discussed further during the afternoon consolidation sessions and the following morning, before the next workshop. Once a topic has been fully explored, the teachers at the **LP** work in grade-homogeneous groups exchanging ideas for implementing the material in the classroom and incorporating it into their already full syllabi. Teachers also learn about how they can use various software, including Mathematical Modeling and Microworlds, to enhance their classroom discussions of discrete mathematics and where to find discrete mathematics resources on the Internet.

The **LP** is successful if the teachers return to their schools empowered to become "discrete math leaders." During the program it is emphasized that they have become experts on discrete mathematics among their peers and that they are encouraged and expected to try this material in their classrooms, to introduce it into their curricula, and to give workshops for other teachers in their school or district, telling them about discrete math. Thus the **LP** includes special sessions on problem solving, assessment of mathematical learning, an overview of discrete mathematics (so that they can successfully answer the common question "what *is* discrete math?") and leadership issues. It is fair to say that the **LP** is as leadership-oriented as content-oriented. A typical year's cohort in the K-8 **LP** program gives workshops attended by at least 5000 teachers not participating in the **LP**.

The summer 2002 **LP** involved commuter programs at Auburn University in Alabama, Tufts University in Massachusetts, Wheaton College in Massachusetts, and University of South Dakota.

Summer 2003 programs are planned at Auburn (twice), Tufts, and Valparaiso University in Indiana.

In order to ensure that our off-site programs are comparable in quality to our programs at Rutgers, we have developed high-quality materials which were used at all different sites by a number of different workshop leaders. These materials include detailed instructor's notes, overhead transparencies, handouts, homework problems, and supplementary resource materials. All new instructors are provided with some training in the use of these materials and in the philosophy of the **LP**. Moreover, in order to compensate for the extra time that is available to participants in the resident programs, all of the commuter programs were extended by 2 days, so that they involved 12 days of activities rather than 10.

The materials that have been developed for the **LP** workshops described above are sufficiently detailed and have been sufficiently field-tested and evaluated, that it is possible for other mathematicians to use them as the basis for independently-conducted ("franchised") parallel high-quality institutes, after suitable training provided by the leadership program, and with the assistance of experienced lead teachers.

In FY02, we obtained funding of $150,000 from the Educational Foundation of America. These funds will pay for the salary of the **LP** Assistant Director for two years (beginning January 2002), with the expectation that she will assist four mathematicians and computer scientists each year in bringing the **LP** to their home institutions.

The **LP** is directed by Joseph Rosenstein (Rutgers).

6.6.2 Young Scholars Program in Discrete Mathematics

In its initial eight-year phase, The Young Scholars Program in Discrete Mathematics (**YSP**) received substantial funding from the National Science Foundation, and was awarded four consecutive two-year grants. Because of the success of the program, DIMACS decided that it would attempt to continue the **YSP** in the same format, even though funding from NSF was no longer available for such programs after 1997. The following model was developed and implemented during the period from 1990-1997: The **YSP** is a four-week residential program for 30 mathematically talented students, mostly from New Jersey, entering grades 11-12, with four follow-up sessions during the year. Week-long workshops are given by college and university faculty on topics ranging from number theory to fractals, designed to excite and challenge the students, motivate them to continue their study of mathematics, and consider careers in the mathematical sciences. Rutgers undergraduates and graduate students serve as group leaders and mentors. Maintaining diversity is a priority: each year that it was funded by NSF the program maintained close to a 50/50 balance between females and males, and approximately 20% of the students are minorities.

Since 1996 there has been a significant technology component to the Young Scholars Program. In addition to doing mathematics with computers and graphing calculators, students are involved in a week-long exciting robotics project and competition. Students are expected to master the Control Lab Robotics System (developed at the MIT Media Lab) and, working in teams, use it to solve an open-ended design problem. Each team is provided the same set of Lego materials, including motors and lights, and is able to build a robot that carries a programmable "brick" and employs touch and light sensors that detect its environment and control its motion. The challenge, of course, is to build a robot that will out-perform those constructed by the other teams. In the 2000 Robotics Challenge, for example, each team had "to build an autonomous robotic device that places plastic golf balls in a container within a playing field". This project integrates mathematics, programming, design and construction.

The second phase of the **YSP** began in 1998, with substantial funding from the AT&T Foundation, averaging $40,000 a year for the last five years. Additional support was provided by DIMACS, the American Association for Artificial Intelligence, and by the Gannett Foundation. These grants have covered approximately 1/3 of the costs of the program, with the remainder being paid for by student fees. The cost to the participants in the summer of 2002 was $2,600, a substantial increase from the charges of $700 to $900 when NSF supported the program. The fees are set so that approximately $10,000 in scholarships are available, which enables 8 of the 30 participants to attend at half-price. (Some are actually provided full scholarships.) We are continuing to look for outside funding of the **YSP**.

The **YSP** program is co-directed by Joseph G. Rosenstein (Rutgers) and Michael O'Nan (Rutgers); the late Robert B. Davis also served as co-director.

6.6.3 Additional K-12 Programs and Accomplishments

Over the past eight years, selected teachers in the **LP** have participated in the "Workshops in Your District" outreach program, traveling to many schools to inform their teachers about discrete mathematics. These workshops were initially appropriate for middle school and high school teachers. However, we have now developed workshops for K-8 teachers as well and, as noted above, in a typical year, 5000 teachers attend workshops in our "Workshops in Your District" program.

We also have a newsletter for high school teachers, "In Discrete Math," that has been circulated free once a year to approximately 3500 teachers.

Our LP for high school teachers has been replaced by a series of 3-day "crash courses", introducing those teachers to the excitement of discrete mathematics and its applications and to available resources and materials, held each summer. This course is intended to give high school teachers a taste of what discrete mathematics is all about and an opportunity to learn how to use these contemporary topics in the classrooms. Discrete mathematics makes math concepts come alive for students. It is an excellent tool for improving reasoning and problem-solving skills, and is appropriate for students at all levels and of all abilities. Teachers have found that discrete mathematics offers a way of motivating unmotivated students and challenging honors students at the same time. Discrete mathematics is a rapidly growing and increasingly used area of mathematics with many practical and relevant applications. Because it is grounded in real world problems, discrete mathematics lends itself easily to implementing the recommendations of the NCTM Standards. This program is co-sponsored by the Rutgers Center for Mathematics, Science and Computer Education (CMSCE). The instructors are Joseph G. Rosenstein (Rutgers) and L. Charles Biehl (Dean of Math, Science, and Technology at the Charter School of Wilmington). This program is entirely self-funded, with the cost of participation provided by the teachers or their districts.

DIMACS has launched a new Modules Series that will include educational modules relating to current research that is amenable for discussion in the high school classroom. The Modules Series is discussed in Section 7.3.3.

In earlier years, DIMACS played a central role in the development of the New Jersey curriculum "standards" for mathematics and a curriculum "framework" that is based on those standards. The "standards" say in 30 pages what students at all grade levels are expected to achieve in all areas of mathematics. The "framework" is a massive piece of work, some 680 pages in all, that describes how the standards can be implemented in classrooms and curricula. The DIMACS Associate Director for Education, Joseph Rosenstein (Rutgers), was one of the three principal authors of the "framework" and served as reviewer and chief editor of the entire document. The framework is filled with new and original examples and activities, appropriate for different grade levels, that were originally developed at DIMACS in the process of finding new ways to present discrete mathematics and other fields of mathematics to teachers, and required creative new approaches to engage teachers in problem-solving, the use of mathematics as a tool to make people think, etc. The chapter of the framework on discrete mathematics is the first comprehensive presentation of what topics in discrete mathematics are appropriate for each K-12 grade level. The "framework" and the 1997 DIMACS volume *Discrete Mathematics in the Schools* have had a real impact on mathematics education locally, nationally, and internationally.

Through the various programs mentioned in this section, we have reached tens of thousands of teachers.

Finally, **LP** director Joseph G. Rosenstein and former **LP** Associate Director Valerie DeBellis have received a grant from NSF to develop textbook materials for college courses for pre-service K-8 teachers based on the **LP** materials. "Making Mathematics Engaging: Discrete Math for K-8 Teachers" will bring both the content materials and the instructional strategies of the **LP** to many thousands of prospective teachers, and will also be used in graduate level and professional development activities for many thousands of practicing teachers.

7 Outreach and Knowledge Transfer

7.1 Visitor Program

A key to the success of the DIMACS programs is the interaction among a wide variety of people: Permanent members of the center, visitors, postdoctoral fellows, and graduate students. DIMACS has almost two thousand visitors a year attending our many programs, and the visitors range from senior leaders in the area of a special focus to outstanding young people. We put special emphasis on visitors from fields outside of discrete mathematics and theoretical computer science, on visitors from companies and government labs, and on visitors from smaller institutions where research is not easily accomplished. Visitors come for periods of time ranging from a few days to a few months to a year. Visitors are central to the outreach and knowledge transfer programs of the center. They bring new ideas and new problems to participants in the center's programs. In turn, they bring back new methods, tools, algorithms, and questions to their home institution. Here is a list of DIMACS visitors in 2002-2003, their affiliation, dates of visit, DIMACS program, and DIMACS host.

- James Abello
  AT&T Labs Research
  3/1/2002 - 12/31/2003
  Data Analysis and Mining
  Fred Roberts

- Pankaj Agarwal
  Duke University
  5/18/2003 - 5/25/2003
  Comp. Geometry
  Steve Fortune

- Tanya Alekseyevskaya (Gelfand)
  Rutgers University
  8/31/1998 - 9/1/2004
  General
  I. Gelfand

- Amihood Amir
  Bar-Ilan University
  8/1/2001 - 8/31/2002
  General
  Tomasz Imielinski & Ken Church

- David Assaf
  The Hebrew University
  9/1/2002 - 10/5/2002
  General
  Yuliy Baryshnikov

- Sergio Cabello Justo
  Universiteit Utrecht
  11/10/2002 - 11/23/2002

Comp. Geometry
Bill Steiger

- Tullio Ceccherini-Silberstein
Universita degli Studi del Sannio di Benevento
2/3/2003 - 2/9/2003
General
Lisa Carbone

- Siu-Wing Cheng
Hong Kong University of Science and Technology
11/14/2002 - 11/28/2002
Comp. Geometry
Bill Steiger

- Yves Crama
Ecole d'Administration des Affaires, Universite de Liege
7/17/2002 - 7/31/2002
General
Peter Hammer

- Bhaskar DasGupta
University of Illinois at Chicago
8/10/2002 - 8/17/2002
NGN and CMB
S. Muthukrishnan

- Bhaskar DasGupta
University of Illinois at Chicago
11/10/2002 - 11/17/2002
Comp. Geometry
Bill Steiger

- Bhaskar DasGupta
University of Illinois at Chicago
2/23/2003 - 3/1/2003
Comp. Geometry
Bill Steiger

- Bhaskar DasGupta
University of Illinois at Chicago
6/20/2003 - 6/20/2003
Comp. Geometry
Eduardo Sontag

- Antonis Delingiannakis
University of Maryland
9/2/2002 - 9/20/2002
Data Analysis & Mining
Yannis Kotidis

- Selcuk Demir
Hebrew University of Israel

2/20/2003 - 2/27/2003
General
Lisa Carbone

- Tanka Nath Dhamala
Tribhuvan University
6/4/2003 - 6/30/2003
General
Peter Hammer

- Ioannis Emiris
University of Athens
11/30/2002 - 12/7/2002
Comp. Geometry
Bill Steiger

- Jose Rul de Matos Figueira
University of Coimbra
8/15/2002 - 5/31/2003
General
Fred Roberts

- Matthias Fitzi
ETH Zurich (UC Davis by 9/2002)
11/4/2002- 11/15/2002
General
Juan Garay

- Stephan Foldes
Tampere University of Technology
7/31/2002 - 8/28/2002
General
Peter Hammer

- Christina Fragouli
University of Athens
2/1/2003 - 6/13/2003
Comp. Info Theory & Coding
Emina Soljanin

- Marina Gavrilova
University of Calgary
12/2/2002 - 12/8/2002
Comp. Geometry
Steve Fortune

- Subir Kumar Ghosh
Tata Institute of Fundamental Research
9/16/2002 - 10/6/2002
Comp. Geometry
Bill Steiger

- Alexander Gnedin
  Mathematical Institute
  3/1/2003 - 3/31/2003
  General
  Ed Coffman and Yuliy Baryshnikov

- Anupam Gupta
  Bell Labs (currently a Postdoc)
  10/1/2002- 12/15/2002
  General
  Peter Winkler

- Vladimir Gurvich
  Russian Academy of Science
  8/1/2002 - 8/31/2002
  General
  Endre Boros

- Stuart Haber
  Intertrust - Star Lab
  6/17/2002 - 9/15/2002
  Visiting Research Assoc.
  Fred S. Roberts

- Yuval Ishai
  DIMACS
  9/1/2001 - 8/31/2002
  General
  Fred S. Roberts

- Ravi Janardan
  University of Minnesota
  4/2/2003 - 4/4/2003
  Comp. Geometry
  Bill Steiger

- Michael Joswig
  Institut fuer Mathematik
  12/3/2002 - 12/7/2002
  Comp. Geometry
  Bill Steiger

- Alexander Kelmans
  University of Puerto Rico
  7/30/2002 - 8/31/2002
  General
  Endre Boros and Leonid Khachiyan

- Sang-Baeg Kim
  TNL, Korea Telecom
  10/15/2001- 10/14/2002
  Next Generation Networks
  Tamara Carpenter

- Vladlen Koltun
  University of California at Berkeley
  9/29/2002 - 10/5/2002
  Comp. Geometry
  Steve Fortune
- Francois LaPointe
  Universite de Montreal
  8/24/2002 - 8/26/2002
  General
  Mel Janowitz
- Brenda Latka
  Lafayette College
  9/1/2001 - 8/31/2002
  General
  Fred Roberts
- Angus Macintyre
  University of Edinburgh
  3/17/2003 - 4/15/2003
  General
  Gregory Cherlin
- Benjamin Martin
  Hebrew University of Jerusalem
  2/16/2003 - 2/22/2003
  General
  Lisa Carbone
- Andrew McGregor
  University of Pennsylvania
  6/2/2003- 8/15/2003
  General
  Alexander Barg
- Gopal Pandurangan
  Brown University
  6/17/2002 - 8/16/2002
  Next Generation Networks
  S. Muthukrishnan
- Dmitry Pasechnik
  J.W. Goethe-Universitaet
  10/15/2002 - 10/19/2002
  General
  Leonid Khackiyan
- Benny Pinkas
  Intertrust - Star Labs
  5/29/2002- 8/23/2002
  Visiting Research Assoc.
  Fred S. Roberts

- Sylvain Pion
  Max Planck Institut fur Informatik
  12/2/2002 - 12/6/2002
  Comp. Geometry
  Bill Steiger
- Bartosz Przydatek
  Carnegie Mellon University
  6/16/2003 - 8/15/2003
  Next Generation Networks
  Rafail Ostrovsky
- Lijun Qian
  Bell Labs / Lucent Technologies
  2/1/2003- 3/1/2003
  Visiting Research Assoc.
  Peter Winkler
- Dieter Rautembach
  Lehrstuhl II fr Mathematik
  10/1/2002 - 11/30/2002
  General
  Peter Hammer
- Bruce Reed
  McGill University
  4/3/2003 - 5/3/2003
  Comp. Geometry
  Bill Steiger
- Guenter Rote
  Institut fur Informatik
  9/16/2002 - 10/2/2002
  Comp. Geometry
  Bill Steiger
- Tim Roughgarden
  Cornell University
  7/28/2002 - 8/3/2002
  General
  Anupam Gupta
- Suleyman Cenk Sahinalp
  Case Western Reserve University
  1/5/2003 - 1/11/2003
  Next Generation Networks
  S. Muthukrishnan
- Rahul Santhanam
  University of Chicago
  6/29/2003 - 7/25/2003
  General
  Eric Allender

- Dana Shapira
  Brandeis University
  12/2/2002 - 12/13/2002
  Next Generation Networks
  S. Muthukrishnan
- Micha Sharir
  Tel Aviv University
  5/17/2003 - 5/24/2003
  Comp. Geometry
  Bill Steiger / Pankaj Agarwal (Wksp organizer)
- Iryna Skrypnyk
  University of Jyvaskyla
  3/1/2003 - 4/30/2003
  Data Analysis & Mining
  Tin Kam Ho
- Warren Smith
  NEC Research Institute
  7/1/2002 - 9/30/2002
  Visiting Research Assoc.
  Doron Zeilberger
- Alexander Soifer
  University of Colorado at Colorado Springs
  1/1/2003 - 8/31/2004
  Comp. Geometry
  Saharon Sheleh
- Venkatesh Srinivasan
  Max Planck Institute
  3/15/2003 - 4/15/2003
  General
  Mike Saks
- Ileana Streinu
  Smith College
  9/30/2002 - 10/4/2002
  Comp. Geometry
  Bill Steiger
- Roberto Tamassia
  Brown University
  12/5/2002 - 12/6/2002
  Comp. Geometry
  Bill Steiger
- Monique Teillaud
  INRIA - Sophia Antipolis
  12/2/2002 - 12/6/2002
  Comp. Geometry
  Bill Steiger

- Joel Tropp
  University of Texas
  6/3/2002 - 8/9/2002
  General
  Martin Strauss
- Alexis Tsoukias
  Universite Paris Dauphine
  3/15/2003 - 6/3/2003
  Visiting Research Assoc.
  Fred Roberts
- Victoria Ungureanu
  Rutgers University / MSIS
  9/1/2002 - 5/31/2003
  Next Generation Networks
  Fred Roberts
- Pavel Valtr
  Charles University
  6/22/2003- 6/28/2003
  Partnered
  Fred Roberts
- Emo Welzl
  ETH Zurich
  5/18/2003 - 5/25/2003
  Comp. Geometry
  Steve Fortune
- Rebecca Wright
  Formerly with AT&T Labs Research
  3/18/2002 - 8/15/2002
  AT&T Appt. / Visiting Research Associate
  Fred Roberts
- Ke Yang
  Carnegie Mellon University
  8/19/2002- 8/30/2002
  General
  Juan Garay and Phil MacKenzie
- Gilles Zmor
  E'Cole National Suprieure des Tlcommunications
  8/18/2002- 8/31/2002
  Comp. Info Theory & Coding
  Alexander Barg
- Ming Zhang
  University of Texas MS Anderson Cancer Center
  3/30/2003- 4/12/2003
  Comp. Geometry
  Bill Steiger

- Jian Zhang
  Yale University
  12/2/2002 - 12/13/2002
  Next Generation Networks
  S. Muthukrishnan

- Yunyue Zhu
  NYU - Courant Institute
  7/1/2002- 7/26/2002
  Data Analysis & Mining
  Flip Korn

- Jing Zou
  Princeton University
  6/24/2002- 8/30/2002
  General
  Anna Gilbert

Selected Resulting Publications

Here are some of the publications that resulted from our visitors' stays. Many visitors commented that they are working on publications that are not yet titled. This should be considered just a sampling of the publications that result from our visitors' collaborations initiated at DIMACS.

- D. Assaf and Y. Baryshnikov, "Optimal control of insurance pricing," INSUMA, submitted.

- J. Balogh, J. Csiric, Y. Ishai, E. Kushilevitz, "Private computation using a PEZ dispenser," in preparation.

- Y. Baryshnikov and A. Gnedin, "Exponential-uniform identities," in preparation.

- Y. Baryshnikov abd A. Gnedin, "Objectives in the best-choice problems," in preparation.

- Beimel, Ishai, Kushilevitz, Raymond, "Breaking the $O(n^{1/(2k-1)})$ barrier for information-theoretic private information retrieval," Proc. FOCS 2002, to appear.

- A. Ben-Israel, S. Foldes, "The circumsphere of a simplex," in preparation.

- P. Berman, P. Bertone, B. DasGupta, M. Gerstein, M. Kao, M. Snyder, "Fast optimal genome tiling with applications to microarray design and homology search," 2nd Inter. Workshop on algorithms in Bioinformatics(WABI 2002), LNCS 2452, R. Guigo and D. Gusfield (eds.) Springer Verlag, 419-433, 2002.

- P. Berman, B. DasGupta, M. Kao, "Tight approximability results for test set problems in bioinformatics," submitted.

- P. Berman, B. DasGupta, D. Mubayi, R. Sloan, G. Turan, Y Zhang, "Efficient algorithms for the inverse protein folding problem on 2D and 3D lattices," submitted.

- L. Carbone and B. Martin, "Deformations of lattices in SL_2 over non-Archimedian local fields," in preparation.

- T. Ceccherini-Silberstein, F. Scarabotti, F. Tolli, "Anisotropic expanders I: the Alon-Boppana Theorem," in preparation.

- T. Ceccherini-Silberstein, F. Scarabotti, F. Tolli, "Anisotropic expanders II: edge-colourability," in preparation.

- R. Cramer, S. Fehr, Y. Ishai, E. Kushilevitz, "Efficient multiparty computation over rings," manuscript.

- B. DasGupta, J. Hespanha, E. Sontag, "Honey-pot searching with local sensory information," submitted.

- O. Devillers and S. Pion, "Efficient exact geometric predicates for Delaunay triangulations," 5th Workshop on algorithm Engineering and Experiments (ALENES 03), Baltimore, MD. 2003.

- G. DiCrescenzo and B. Przydatek, "Perfectly secure message transmission with multiple receivers," in preparation.

- S. Foldes and P. L. Hammer, "Submodularity, supermodularity and higher order monotonicities of pseudo-Boolean functions," Math. of Operations Research, submitted.

- S. Foldes and P. L. Hammer, "Closed classes of pseudo-Boolean functions with local character," RUTCOR Research Report RRR 23-2002, 2002.

- C. Fragouli and E. Soljanin, "Decentralized network coding based on subtree decomposition," InfoCom 2004, submitted.

- C. Fragouli, "Connections between convolutional codes and network coding,"in preparation.

- R. Gennaro, Y. Ishai, E. Kushilevitz, T. Rabin, "On 2-round secure multiparty computation," Proc. Of Crypto. 2002.

- S. Greenfield and B. Latka, "Cardinalities of maximal antichains of tournaments," in preparation.

- T.K. Ho and I. Skrypnyk, "Feature selection and training set sampling for ensemble learning on heterogeneous data," DIMACS Technical Report 2003-23, 2003.

- T.K. Ho and I. Skrypnyk, "Clustering heterogeneous data in multiple feature subspaces," Proc. 13th Int. Conf. Pattern Recognition (ICPR-2004), Cambridge, UK, in preparation.

- Y. Ishai and E. Kushilevitz, "Perfect constant-round secure computation via perfect randomizing polynomials," Proc. of ICALP 2002.

- Y. Ishai, E. Kushilevitz, R. Ostrovsky, A. Sahai, "Batch codes and their applications," in preparation.

- A. Kelman and D. Mubayi, "How many disjoint 2–edge paths must a cubic graph have? ," J. of Graph Theory, to appear.

- V. Koltun and M. Sharir, "Curve-sensitive cuttings," in preparation.

- B. Latka, "A classification of antichains of finite tournaments," DIMACS Technical Report 2002-24, 2002.

- B. Latka, "Tournaments that omit N_5 are well-quasi-ordered ," DIMACS Technical Report 2002-31, 2002.

- B. Latka, "Structure theorem for tournaments omitting N_5," J. of Graph Theory **42**, 165-192, 2003.

- B. Latka, "Antichains of bounded size in the class of tournaments," Order, to appear.

- B. Latka, "Maximal antichain of finite tournaments," Proc. 34th Southeastern Conf. Combinatorics, Graph Theory, and Computing,2003, to appear.

- V. Lozin and D. Rautenbach, "On the band-, tree-, and clique-width of graphs with bounded vertex degree," DIMACS Technical Report 2002-54, 2002.

- V. Lozin and D. Rautenbach, "On the boundedness of width parameters for chordal bipartite graphs," in preparation.

- V. Lozin and D. Rautenbach, “Some results on star-free graphs,” in preparation.

- M. Morisio, I. Stamelos, A. Tsoukiás, “Software product and process assessment through profile based evaluation,” J. of Software Engineering and Knowledge Engineering, to appear.

- R. Ostrovsky and B. Przydatek, “On concurrent zero-knowledge,” in preparation.

- M. Ozturk, A. Tsoukiás, Ph. Vincke, “Preference modelling,” in preparation.

- G. Pandurangan, “Random graph problems motivated by sensor networks,” DIMACS Research Report, 2002.

- D. Rautenbach, “A note on linear discrepancy and bandwidth,” DIMACS Technical Report 2002-53, 2002.

- I. Skrypnyk, “Ensemble of learning models at multiple feature subspaces for heterogeneous classification problems,” Licentiate thesis, University of Jyvaskyla, Finland, in preparation.

- A. Tsoukiás, From decision theory to decision aiding methodology,” European J. of Operational Research, to appear.

- A. Tsoukiás, “On the concept of decision aiding process,” in preparation.

- K. Yang, “Non-maleability from un-forgeability,” in preparation.

- G. Zemore, “Concatenated codes: serial and parallel,” in preparation.

Selected Quotes from Visitors

- “During my stay at DIMACS I had the opportunity to talk to people with great experience in my field. I thus got helpful comments by them on my ongoing research, and also suggestions and ideas about my future work. Thus, my visit was extremely helpful to me. I also got the chance to see some of the problems faced in both other areas of databases, and in fields not related at all to my research, such as fraud detection by use of the telephone calls one makes.” Antonios Deligiannakis

- “My stay at DIMACS gave me the opportunity to look at new research directions, and investigate interdisciplinary problems, that involve networking, coding theory, graph theory, algebra and discrete mathematics. This would not have been possible without the support of DIMACS.” Christina Fragouli

- "The research environment at DIMACS was highly supportive of informal and semi-formal exchange of ideas, discussion of new developments in the area and individual research as well. A number of collaborations established and exposure to new directions were extremely useful for my research." Marina Gavrilova

- "I learned of directions in Computer Science which were completely new to me as a probabilist. Many insights I got at DIMACS, in particular, in the field of analysis of algorithms, determine my current work." Alexander Gnedin, University of Utrecht, The Netherlands

- "As a result of my visit, I have started a new research initiative on using geometric methods for problems in medicine/bio-medical engineering/bioinformatics." Ravi Janardan

- "I was surprised to see again Konstantin Rybnikov. We exchanged a few ideas and found out that there is a certain overlap in our current research. This is great because we treated the subject from different angles. We can now think about merging our efforts." Michael Joswig

- "The result of my work was the discovery of a wealth of connections spanning not only several areas of mathematics but also computer science and biology. Understanding these connections will be part of my research agenda for several years." Brenda Latka

- "My stay at DIMACS opened the way to a long term collaboration through LAMSADE and DIMACS aiming to exchange PhD students and senior researchers..." A. Tsoukiás

7.2 Underrepresented Groups

DIMACS has a strong record of involvement of women, African American, and Hispanic scientists as organizers, speakers, visitors, graduate students, and participants in our many research and educational programs.

During FY2003, we continued our collaboration with the Douglass College (Rutgers University) Program for Women in Mathematics, Science, and Engineering.

Several African Americans and Hispanics play prominent roles in DIMACS and the involvement by African Americans in many of our programs such as Young Scholars is strong. A number of prominent African American scholars have agreed to help us recruit African American participants to our programs and in particular to work with us on our Reconnect Program. We have access to extensive mailing lists as a result of our hosting in 1996 of the Second Conference for African American Researchers in the Mathematical Sciences (CAARMS), which resulted in a DIMACS book, *African Americans in Mathematics*. We have followed up on the conference with continuing contacts with African American mathematicians and inform them of the many opportunities to participate in DIMACS activities. We have similar contacts with Hispanic Americans.

We continue to have active involvement with "historically black colleges and universities" (HBCU's) and other minority institutions. We initiated a joint program with HBCU Texas Southern University – as part of our precollege Leadership Program. Faculty from HBCU's Spelman College, Morgan State University, and Howard University as well as from high-minority-enrollment institutions such as University of Houston-Downtown have been involved with our "Reconnect" program. We have worked with Spelman to plan a "Reconnect" event on that campus and are in the process of working with Morgan State on a similar event. We have had extensive connections with University of Puerto Rico, including hosting visitors and REU students.

In the summer of 2002, a large contingent of minority undergraduate students from The Mathematical and Theoretical Biology Institute (MTBI) at Cornell participated in various programs in our Special Focus on Computational and Mathematical Epidemiology, including an international conference and several tutorials. Steve Wirkus, MTBI REU summer director, said this about their experience: "MTBI has strived to prepare a very ethnically and socially diverse group of undergraduates from a variety of academic backgrounds to do quality research in the field of epidemiology. We have always been aware of the critical need for programs that nurture, guide, and strengthen the vast potential that we have in an undergraduate population, especially when this population is from an underrepresented minority group. We had hoped that these tutorials/conferences would continue to support the ideals and goals of MTBI and were very pleased with the results. It really benefited the students to be able to see and interact with so many of the top names in the fields of epidemiology, to see familiar material reinforced, and to be exposed to exciting new ideas.

The positive influence of attending DIMACS events was evident in the content of MTBI student technical reports. There were models of HIV dynamics and issues of epidemics and homeland security also were represented. Moreover, DIMACS succeeded in impressing and reinforcing a positive view of epidemiological research on a group of students who have overcome many challenges to be where they are. It is programs like these, programs that seriously attempt to attract and train students in interdisciplinary fields, that will equip minority students with the skills and motivation needed to overcome future challenges. DIMACS is succeeding in cultivating an ethnically diverse and sensitive future scientific workforce through its close collaboration with programs like MTBI."

7.3 Publications

7.3.1 AMS-DIMACS Volumes

In the last thirteen years DIMACS has published 61 refereed volumes in the series published by the American Mathematical Society (AMS). A volume may be a collection of papers presented at or related to a workshop, papers related to a DIMACS special focus topic, papers on an important educational topoic, or papers on a research topic of interest to the mathematics or computer science communities. The volumes published in FY2003 are:

- Volume Fifty Eight: **Set Theory: The Hajnal Conference, October 15 - 17, 1999**, Editor: Simon Thomas, 2002.
- Volume Fifty Nine: **Data Structures, Near Neighbor Searches, and Methodology: Fifth and Sixth DIMACS Implementation Challenges**, Editors: Michael H. Goldwasser, David S. Johnson and Catherine C. McGeoch, 2002.
- Volume Sixty: **Algorithmic and Quantitative Real Algebraic Geometry**, Editors: Saugata Basu and Laureano Gonzalez-Vega; 2003.
- Volume Sixty One: **Bioconsensus**, Editors: M. F. Janowitz, F. J. Lapointe, F. R. McMorris and F. S. Roberts, 2003.

7.3.2 DIMACS Technical Reports

The DIMACS Technical Report series publishes papers that are due in part to DIMACS activities, such as the work of graduate students, permanent members, visitors, or postdocs. DIMACS Technical Reports appear in the NCSTRL (Networked Computer Science Technical Reports Library) where they can be searched by title, author names, or words in abstracts. 872 Technical Reports have been published since the series started in 1989 and since 1993 the Technical Reports have been available electronically via the DIMACS website. In FY2003, 61 Technical Reports were added to the series. Here is a list of these Technical Reports.

2002 Technical Reports (in FY2003)

- 2002-24 A Classification of Antichains of Finite Tournaments by Brenda J. Latka.
- 2002-25 Strong Key-Insulated Signature Schemes by Yevgeniy Dodis, Jonathan Katz, Shouhuai Xu and Moti Yung.
- 2002-26 An Algorithm for Dualization in Products of Lattices by Khaled M. Elbassioni.
- 2002-27 Extending the Balas-Yu Bounds on the Number of Maximal Independent Sets in Graphs to Hypergraphs and Lattices by Endre Boros, Khaled Elbassioni, Vladimir Gurvich and Leonid Khachiyan.
- 2002-28 The Domination Number of (K_p, P_5)-Free Graphs by I. E. Zverovich.
- 2002-29 Parallel Scheduling Problems in Next Generation Wireless Networks by L. Becchetti, S. Diggavi, S. Leonardi, A. Marchetti-Spaccamela, S. Muthukrishnan, T. Nandagopal and A. Vitaletti.
- 2002-30 On the Edge-Forwarding Indices of Frobenius Graphs by Yan Wang, Xin Gui Fang, and D. Frank Hsu.
- 2002-31 Tournaments that omit N_5 are well-quasi-ordered by Brenda J. Latka.
- 2002-32 Error exponents of expander codes under linear-complexity decoding by Alexander Barg and Gilles Zmor.
- 2002-33 On Convex Polytopes in the Plane "Containing" and "Avoiding" Zero by R. Collado, A. Kelmans and D. Krasner.
- 2002-34 Stable Matchings in Three-Sided Systems with Cyclic Preferences by Endre Boros, Vladimir Gurvich, Steven Jaslar and Daniel Krasner.
- 2002-35 Estimating Dominance Norms of Multiple Data Streams by Graham Cormode and S. Muthukrishnan.
- 2002-36 A class of i.p.p. codes with efficient identification by Alexander Barg and Gregory Kabatiansky
- 2002-37 Reconstructing Sets From Interpoint Distances by Paul Lemke, Steven S. Skiena and Warren D. Smith

- 2002-38 Boundary classes of graphs for the dominating set problem by V.E. Alekseev, D.V. Korobitsyn and V.V. Lozin.

- 2002-39 Report on DIMACS Working Group Meeting: Mathematical Sciences Methods for the Study of Deliberate Releases of Biological Agents and their Consequences by Carlos Castillo-Chavez and Fred S. Roberts.

- 2002-40 Challenges for Discrete Mathematics and Theoretical Computer Science in the Defense Against Bioterrorism by Fred S. Roberts.

- 2002-41 Deferred Assignment Scheduling in Clustered Web Servers by Victoria Ungureanu, Phillip G. Bradford, Michael Katehakis and Benjamin Melamed.

- 2002-42 Using reasoning about dynamic domains and inductive reasoning for automatic processing of claims by Boris Galitsky and Dmitry Vinogradov.

- 2002-43 Unique domination and domination perfect graphs by I. E. Zverovich.

- 2002-44 Bisplit Graphs by Andreas Brandstadt, Peter L. Hammer, Van Bang Le, Vadim V. Lozin.

- 2002-45 Serial and tree-serial dynamic programming with application to fact identification by Vadim Mottl and Ilya Muchnik.

- 2002-46 The PAK Suite: Protocols for Password-Authenticated Key Exchange by Philip MacKenzie.

- 2002-47 Ratio of generalized parameters by I. E. Zverovich and I. I. Zverovich.

- 2002-48 Consensus List Colorings of Graphs and Physical Mapping of DNA by N. V. R. Mahadev and Fred S. Roberts.

- 2002-49 Comprehensive vs. Comprehensible Classifiers in Logical Analysis of Data by Gabriela Alexe, Sorin Alexe, Peter L. Hammer, and Alexander Kogan.

- 2002-50 Spanned Patterns for the Logical Analysis of Data by Gabriela Alexe and Peter L. Hammer.

- 2002-51 Hydrodynamics of Binary Fluid Phase Segregation by Sorin Bastea, Raffaele Esposito, Joel L. Lebowitz, and Rossana Marra.

- 2002-52 Consensus algorithms for the generation of all maximal bicliques by Gabriela Alexe, Sorin Alexe, Yves Crama, Stephan Foldes, Peter L. Hammer and Bruno Simeone.

- 2002-53 A Note on Linear Discrepancy and Bandwidth by Dieter Rautenbach.

- 2002-54 On the band-, tree- and clique-width of graphs with bounded vertex degree by Vadim Lozin and Dieter Rautenbach.

- 2002-55 Local vs. joint measurements for the entanglement of assistance by T. Laustsen, F. Verstraete, and S. J. van Enk.

- 2002-56 On Generating Three-connected Planar Graphs by Tongyin Liu and Yanpei Liu.

- 2002-57 Lower bounds for the size of binary space partition of rectangles by Ying Liu.

- 2002-58 Methods of data fusion in information retrieval: rank vs. score combination by D. Frank Hsu, Jacob Shapiro and Isak Taksa.

2003 Technical Reports (in FY2003)

- 2003-01 Comparison of convex hulls and box hulls by E. Boros, V. Gurvich, Y. Liu.

- 2003-02 On the Notion of Category Size in Multiple Criteria Sorting Models by Vincent Mousseau, Luís C. Dias and José R. Figueira.

- 2003-03 Chordal Bipartite Graphs of Bounded Tree- and Clique-width by V. Lozin and D. Rautenbach.

- 2003-04 Flexible Regulation of Distributed Coalitions by Xuhui Ao and Naftaly Minsky.

- 2003-05 Fast Optimal Genome Tiling with Applications to Microarray Design and Homology Search by Piotr Berman, Paul Bertone, Bhaskar DasGupta, Mark Gerstein, Ming-Yang Kao and Michael Snyder.

- 2003-06 A Data Mining Problem in Stochastic Programming by Andras Prekopa and Xiaoling Hou.

- 2003-07 On a Conjecture of Chvatal by Xiaomin Chen.

- 2003-08 On Bayesian Learning of Sparse Classifiers by Wen-Hua Ju, David Madigan, and Steven L. Scott.

- 2003-09 Uplink Scheduling in CDMA Packet-Data Systems by Krishnan Kumaran, Lijun Qian.

- 2003-10 Perfect Graphs, Kernels, and Cores of Cooperative Games by E. Boros and V. Gurvich.

- 2003-11 Radial Histograms for Spatial Streams by Graham Cormode and S. Muthukrishnan.

- 2003-12 Basic perfect graphs and their extensions by I. E. Zverovich and V. E. Zverovich.

- 2003-13 From the Ground Up: The DIMACS Software Security Workshop by Gary McGraw.

- 2003-14 Tight Approximability Results for Test Set Problems in Bioinformatics by Piotr Berman and Bhaskar DasGupta and Ming-Yang Kao.

- 2003-15 Towards an Efficient Cluster-based E-Commerce Server by Victoria Ungureanu, Michael Katehakis, Benjamin Melamed.

- 2003-16 Class-Dependent Assignment in Cluster-based Servers by Victoria Ungureanu, Phillip G. Bradford, Michael Katehakis, Benjamin Melamed.

- 2003-17 On the Complexity of Some Enumeration Problems for Matroids by Endre Boros, Khaled Elbassioni, Vladimir Gurvich and Leonid Khachiyan.

- 2003-18 IPSOFACTO: A Visual Correlation Tool for Aggregate Network Traffic Data by Flip Korn, S. Muthukrishnan and Yunyue Zhu.

- 2003-19 Checks and Balances: Monitoring Data Quality Problems in Network Traffic Databases by Flip Korn, S. Muthukrishnan and Yunyue Zhu.

- 2003-20 Improved Data Stream Summaries: The Count-Min Sketch and its Applications by Graham Cormode and S. Muthukrishnan.

- 2003-21 From Decision Theory to Decision Aiding Methodology by Alexis Tsoukis.

- 2003-22 Independent domination on $2P_3$-free perfect graphs by I. E. Zverovich.

- 2003-23 Feature Selection and Training Set Sampling for Ensemble Learning on Heterogeneous Data by Iryna Skrypnyk and Tin Kam Ho

- 2003-24 A new kind of graph colorings by Igor Zverovich

- 2003-25 On a problem of Lesniak, Polimeni, and Vanderjagt by Igor Zverovich

- 2003-26 Algorithmic aspects of Reducing Pseudopath Method by Igor Zverovich

7.3.3 DIMACS Educational Modules Series

DIMACS has commenced a modules series: the DIMACS Educational Modules in Discrete Mathematics and Theoretical Computer Science. This series is part of the DIMACS effort to enrich undergraduate education, community college education, and precollege education and to integrate research and education. The modules are self-contained material that can be used for one or more class meetings, or for independent study or outside reports. In FY2003, we published the first modules in the series.

DIMACS modules emphasize topics that are not yet readily available in text material at the module's educational level. Almost every module introduces some aspect of a topic undergoing active research development. For instance, a number of DIMACS modules are being developed in conjunction with the many workshops, tutorials, and other outreach activities at the Center. However, anyone can author a DIMACS Module.

DIMACS modules aim at students at various levels. Some modules are aimed at the junior/senior undergraduate level, for students majoring in math, CS, or other technical fields. Some are for technically oriented first year students. Some are for liberal arts students and provide a taste of what mathematics and computer science are like today. Some are for high school students. Some topics are to appear in both high-end and low-end modules.

DIMACS modules will take advantage of advances in technology. They will all be available electronically, and some may be interactive. We are planning to bundle modules into DIMACS volumes, to be published in the DIMACS book series. This series is published by the American Mathematical Society. AMS is amenable to co-publishing agreements with other publishers such as NCTM and MAA in order to maximize exposure of a volume and it is likely that the volumes of DIMACS modules will be co-published. Modules will be available electronically without charge, and in hard copy at cost.

An Editorial Board includes noted researchers and people with educational and writing experience at the various levels at which modules will be written. The Board has identified many current research topics which seem ripe for modules and has also identified researchers who may be receptive to writing modules, or serving as co-writers or advisors. The Modules Editorial Board is especially eager to arrange for collaborations between researchers, expositors, and teachers.

The modules editorial board has a chief editor, seven associate editors, and other board members. Here is a list of board members:

Chief Editor: Joseph Malkevitch, York College (CUNY)

Associate Editors:
Susanna Epp, DePaul University
Melvin Janowitz, DIMACS, Rutgers University

Rochelle Leibowitz, Wheaton College
Stephen Maurer, Swarthmore
Catherine McGeoch, Amherst College
Joseph Rosenstein, Rutgers University
William Steiger, Rutgers University

Board Members:
Adam Buchsbaum, AT&T Research Labs,
Martin Farach, Rutgers University,
Ronald Graham, UC San Diego,
Philip Hanlon, Michigan,
Robert Hochberg, East Carolina University,
Glenn Hurlbert, Arizona State,
Nancy Maulding, Raritan Valley Community College,
Joseph O'Rourke, Smith,
Warren Page, New York City Technical College, CUNY,
Janet Ray, Seattle Central Community College,
Fred S. Roberts, Rutgers University
Sharon Cutler Ross, DeKalb College,
Robert Sedgewick, Princeton,
Philip Straffin, Beloit,
Barry Tesman, Dickinson,
Douglas West, Illinois,
Peter Winkler, Bell Labs.

As noted earlier, Two modules were made available during 2002-2003. Many additional modules are in preparation.

Modules Published in FY2003

- 03-1 Alexander Kheyfits, Bronx Community College (CUNY)
  Introduction to Clustering Algorithms: Hierarchical Clustering
- 03-2 R. S. (Chuck) Tiberio, Wellesley High School
  Facility Location Problems

8 DIMACS Partnerships and Members

8.1 Industrial Partners

DIMACS has four "full" industrial partners and four "affiliated" partners, two new ones in FY03. The full partners have been involved with us for many years and are truly partners in the center. Senior representatives from these institutions are heavily involved in planning DIMACS activities and play a major organizational role in the center's programs. Moreover, representatives of senior management are extremely active and visible supporters of DIMACS in numerous forums. In the past, DIMACS has received significant support from these institutional partners. They have provided substantial support by appointing DIMACS postdocs, sponsoring workshops, and allowing release time for their researchers to actively participate in DIMACS administrative committees and to organize DIMACS workshops, working groups, and special foci. While the current economic climate has resulted in cutbacks in direct financial support by our full partners, they have otherwise remained extremely involved in the center.

In the last few years, we have attracted four new "affiliated" partners in the center. These affiliated partners are not asked to provide financial support at the level of the full partners. However, they are increasingly active in support of the center, with programmatic advice, involvement in workshops, working groups, and special focus programs.

DIMACS **Key industrial partners** in 2002-2003 were:

AT&T Labs
Bell Labs (Lucent Technologies)
NEC Laboratories America
Telcordia Technologies
Avaya Labs (affiliated partner)
HP Labs (affiliated partner) (new in FY2003)
IBM Research (affiliated partner) (new in FY2003)
Microsoft Research (affiliated partner)

Other companies have also been involved with DIMACS. Several companies have been extremely helpful in planning and carrying out our Special Focus on Computational Molecular Biology. In particular, researchers from Celera Genomics, Roche BioScience, SmithKline Beecham, and IBM are serving on the Advisory Committee for that activity, have organized workshops, and have provided corporate support for our workshops. Also in the past year, we have worked hard to involve pharmaceutical companies in DIMACS molecular biology activities with special emphasis on developing connections with the applied math group at Merck Research. Merck researchers at various levels have been actively involved in DIMACS programs, with the Vice President for Epidemiology actively engaged with our Special Focus on Computational and Mathematical Epidemiology.

Celera Genomics provided $10K to fund a program to recognize outstanding achievements of young people in the field. Winners are invited to present a talk at a DIMACS workshop with expenses paid. (This program has recently been renewed with another gift of $5K for its first year.)

We have had significant industrial involvement in our Special Focus on Next Generation Networks Technologies and Applications, with organizers from AT&T Labs and Akamai Technologies. Our Advisory Committee represents a number of major companies from the information technology field and we have obtained support for this activity from Microsoft and IBM, among other companies. Researchers from Intertrust Star Labs in Princeton were also actively engaged in this special focus in FY2002. Researchers from the new HP Labs office in New Jersey have also been heavily involved.

Our forthcoming special focus in Communication Security and Information Privacy is being organized by researchers from AT&T Labs and NEC Research, with heavy involvement of researchers from HP Labs. Organizers devoted countless hours in FY03 to planning this activity. Microsoft Research funded our first workshop in this focus in FY03.

There has been considerable industrial involvement in the Special Focus on Computational and Mathematical Epidemiology. For example, our work on disease surveillance has involved representatives of companies such as Verispan, Quintiles, SAS, and Harvard Pilgrim. In addition, the project on Bioterrorism Sensor Location, which had its origin in the Epidemiology special focus, was developed in conjunction with the MITRE Corporation and we continue to be involved with MITRE in developing the project.

8.2 Permanent Members of the Center

Since DIMACS started, a key to its strength has been the involvement of researchers from its partner institutions. There are now some 230 scientists who are "permanent" members of DIMACS, many coming from the partner institutions, academic and industrial, and including some 69 from Rutgers. These permanent members serve on DIMACS committees, organize DIMACS special foci and workshops, host visitors, help prepare grant proposals, participate in multi-disciplinary projects, and mentor students and postdocs. These members represent a wide variety of disciplines, not just the mathematical and computer sciences (including operations research and statistics). They include biologists, psychologists, chemists, economists, etc.

Here is a complete list of members.

Permanent Members of DIMACS

- **AT&T Labs - Research**
  - James Abello
  - Bill Aiello
  - David Applegate
  - Matt Blaze
  - Adam L. Buchsbaum
  - Edith Cohen
  - Nick Duffield
  - Peter Fishburn
  - Emden Gansner
  - Anna Gilbert
  - Albert Greenberg
  - David S. Johnson
  - Howard Karloff
  - Jeffrey C. Lagarias
  - Keh-Wei Lih
  - Carsten Lund
  - Michael Merritt
  - S. (Muthu) Muthukrishnan
  - Steven Phillips
  - Madga Procopiuc

- Nick Reingold
- Omer Reingold
- Mauricio G. C. Resende
- Neil J. A. Sloane
- Martin Strauss
- Mikkel Thorup
- Suresh Venkatasubramanian
- Kiem-Phong Vo

- **Avaya Labs**
  - Jon Bentley
  - Mathilde Benveniste
  - Wen-Hua Ju
  - Martin Kappes
  - Mark Karol
  - A.S. Krishnakumar
  - Jim Landwehr, Unit Rep
  - Colin Mallows
  - Ravi Sethi

- **Bell Labs**
  - Matthew Andrews
  - Alexei Ashikhmin
  - Brenda Baker
  - Yuliy Baryshnikov
  - Chandra Chekuri
  - Ken Clarkson
  - Steven Fortune
  - Christopher Fuchs
  - Juan Garay
  - Michael Garey
  - Lov Grover
  - Piyush Gupta
  - Tin Kam Ho
  - Guy Jacobson
  - David Lee
  - Philip MacKenzie
  - Girija Narlikar
  - Arun Netravali

- Marty Reiman
- Iraj Saniee
- F. Bruce Shepherd
- Emina Soljanin
- Alexander Stolyar
- Steven van Enk
- Adriaan Van Wijngaarden
- Qiong Wang
- Gordon Wilfong
- Peter Winkler
- Francis Zane
- Lisa Zhang

- **HP Labs**
  - William Horne
  - Benny Pinkas
  - Tomas Sander

- **Microsoft**
  - Christian Borgs
  - Jennifer Chayes
  - Jeong Han Kim
  - Laci Lovasz
  - David Wilson

- **NEC Laboratories America**
  - Leon Bottou
  - C. William Gear
  - Joe Kilian
  - Dmitry Pavlov
  - Ronitt Rubinfeld

- **Princeton University - Department of Computer Science**
  - Sanjeev Arora
  - Moses Charikar
  - Bernard Chazelle
  - David Dobkin
  - Andrea Lapaugh
  - Nick Pippenger

- Amit Sahai
- Rob Schapire
- Robert Sedgewick
- Jaswinder Singh
- Mona Singh
- Ken Steiglitz
- Robert Tarjan
- Kevin Wayne
- Andrew Yao

- **Princeton University - Department of Ecology and Evolutionary Biology**
  - Simon Levin

- **Princeton University - Department of Mathematics**
  - John Conway
  - Harold Kuhn
  - John F. Nash, Jr.
  - Peter Sarnak
  - Paul Seymour

- **Princeton University - Department of Statistics and Operations Research**
  - Robert J. Vanderbei

- **Rutgers University - Center for Math Science and Computer Education**
  - Gerald Goldin

- **Rutgers University - Department of Chemistry**
  - Ron Levy
  - Wilma Olson

- **Rutgers University - Communication, Information & Library Studies**
  - Paul B. Kantor

- **Rutgers University - Department of Computer Science**
  - Eric Allender
  - B.R. Badrinath
  - Vasek Chvatal
  - Sven Dickinson
  - Martin Farach-Colton

- Michael Fredman
- Michael Grigoriadis
- Haym Hirsh
- Bahman Kalantari
- Leonid Khachiyan
- Dimitris Metaxas
- S. (Muthu) Muthukrishnan
- Craig Nevill-Manning
- William Steiger
- Mario Szegedy
- Endre Szemeredi
- Ann Yasuhara

- **Rutgers University - DIMACS**
  - James Abello
  - Mel Janowitz
  - Brenda J. Latka
  - Fred S. Roberts
- **Rutgers University - Department of Genetics**
  - David Axelrod
- **Rutgers University - Department of Mathematics**
  - Jozsef Beck
  - Richard Bumby
  - Lisa J. Carbone
  - Gregory Cherlin
  - Israel Gelfand
  - Simon Gindikin
  - Stephen Greenfield
  - Andras Hajnal
  - Jeffry Kahn
  - Janos Komlos
  - Joel Lebowitz
  - Richard Lyons
  - Michael O'Nan
  - Fred S. Roberts
  - Joseph G. Rosenstein
  - Michael Saks

  - Saharon Shelah
  - Charles Sims
  - Avraham Soffer
  - Eduardo Sontag
  - Simon Thomas
  - Wolmer Vasconcelos
  - Doron Zeilberger

- **Rutgers University - Department of Psychology**
  - Michael Leyton
  - Zili Liu

- **Rutgers University - Rutgers Center for Operations Research (RUTCOR)**
  - Farid Alizadeh
  - Adi Ben-Israel
  - Endre Boros
  - Jonathan Eckstein
  - Peter. L. Hammer
  - Alan J. Hoffman
  - Alexander Kogan
  - Vadim Lozin
  - Andras Prekopa
  - Michael H. Rothkopf
  - David F. Shanno
  - Albert C. Williams

- **Rutgers University - Department of Statistics**
  - Ram (Ramanathan) Gnanadesikan
  - Richard Gundy
  - Donald Hoover
  - Juan Lin
  - David Madigan
  - Joseph Naus
  - Lawrence Shepp
  - Yehuda Vardi

- **Telcordia Technologies**
  - Tamra Carpenter
  - Giovanni Di Crescenzo

- Martin Eiger
- Stefano Galli
- Jon Kettenring
- K.R. Krishnan
- Arnie Neidhardt
- Rafail Ostrovsky
- Sivaram Rajagopalan
- David Rosenbluth
- David Shallcross

- **Other**
  - Alfred Aho, Columbia University
  - Rajeev Alur, University of Pennsylvania
  - Tewodros Amdeberhan,DeVry Institute of Technology
  - Jonathan Berry, Lafayette College
  - Sandeep Bhatt, Akamai Technologies
  - Enrico Bombieri, Institute for Advanced Study
  - Robert Calderbank, formerly of AT$T Labs - Research
  - Fan Chung, University of California, San Diego (also associated with Telcordia Technologies)
  - Edward G. Coffman, Columbia University
  - Valerie De Bellis, East Carolina University
  - Nathaniel Dean, Rice University
  - Richard DeMillo, Georgia Institute of Technology
  - Kousha Etessami, University of Edinburgh
  - Joan Feigenbaum, Yale University
  - Ronald Graham, University of California, San Diego
  - Stuart Haber, Surety Technologies
  - Rob Hochberg, East Carolina University
  - Frank Hwang, Chiaotung University
  - Markus Jakobsson, RSA Labs
  - Sampath Kannan, University of Pennsylvania
  - Jonathan Katz, University of Maryland
  - Sanjeev Khanna, University of Pennsylvania
  - Natalia Komarova,Institute for Advanced Study
  - Robert Kurshan, Cadence Design Systems
  - Yann LeCun, New York University
  - Linda Lesniak, Drew University
  - Richard Lipton, Georgia Tech

- Tal Malkin, Columbia University
- Stephen B Maurer, Swarthmore College
- Nick Maxemchuk, Columbia University
- Clyde Monma, Formerly of Telcordia Technologies
- Andrew Odlyzko, University of Minnesota
- Andrew Ogielski, Dartmouth College
- Panos Pardalos, University of Florida
- Gabor Pataki, University of North Carolina at Chapel Hill
- David M. Pennock,Overture Services, Inc.
- Carl Pomerance, Dartmouth College
- Avi Rubin, Johns Hopkins University
- Serap Savari, University of Michigan
- Amin Shokrollahi, Digital Fountain
- Peter Shor, Massachusetts Institute of Technology
- Warren Smith, Temple University
- Diane Souvaine, Tufts University
- David Waltz, formerly of NEC
- Ward Whitt, Columbia University
- Avi Wigderson, Institute for Advanced Study
- Andrew Wright, Star Lab - Intertrust Technologies
- Rebecca Wright, Stevens Institute of Technology
- Mihalis Yannakakis, Stanford University

8.3 Honors and Awards of the Members

- **Sanjeev Arora**, DIMACS member from the Princeton University Department of Computer Science, was an **invited speaker at the International Congress of Mathematicians**, Aug 2002 and received the **Distinguished Alumnus Award** from CS division of UC Berkeley, May 2003. Dr. Arora has served as a DIMACS workshop organizer and was previously co-director for Princeton.

- **Moses Charikar**, DIMACS member from the Princeton University Department of Computer Science, received an **Alfred P. Sloan Foundation Fellowship**, 2003 and the **best paper award at the 44th Annual IEEE Symposium on Foundations of Computer Science** (FOCS) 2003 for the paper "On the impossibility of dimension reduction in l_1", with Bo Brinkman (Princeton University). Dr. Charikar has been a workshop organizer.

- **Jonathan Eckstein**, DIMACS member from Rutgers University - RUTCOR, received the **Thomas H. Mott Award for Excellence in Teaching**, awarded by vote of graduating seniors at the Rutgers Business School New Brunswick undergraduate program, May 2003.

- **Joan Feigenbaum**, DIMACS member from Yale University, was the Yale PI on an **NSF Large-ITR award**. Dr. Feigenbaum is is an organizer of the 2004-2006 Special Focus on Computation and the Socio-Economic Sciences and served at Co-Director for the DIMACS Research & Education Institute. Dr. Feigenbaum is a former member of the DIMACS Council and a special focus organizer of the focus on Next Generation Networks.

- **Gerald Goldin** has been awarded a grant and an invitation to King's College, London as **Leverhulme Visiting Professor in Physics** for the academic year 2004-2005. Dr. Goldin is the former Director of the Rutgers University Center for Mathematics, Science, and Computer Education, working closely with DIMACS through CMSCE. He is also a former member of the DIMACS Executive Committee.

- **Stephen J. Greenfield**, DIMACS member from the Rutgers University Department of Mathematics, was invited to give a program of lectures for each of the last three summers to the **New Jersey Governor's School of Engineering and Technology**. Dr. Greenfield is the former Program Director for the DIMACS Research and Education Institute.

- **Natalia Komarova** received the Student Achievement and Advocacy Services first **"Prize for Promise" award**, December 3, 2002, of $ 1000,000 that recognizes young women of exceptional ability, ambition, boldness, brilliance, and dedication within their field of expertise. Dr. Komarova joined Rutgers as of September, 2003. She is on the Advisory Committee for the DIMACS Special Focus on Computational and Mathematical Epidemiology. In announcing the award, Dr. Murray Gell-Mann, a winner of the Nobel Prize in Physics and a founding board member of the Prize for Promise, said: "Her efforts in researching the genetic instability in cancer initiation is a radical departure from traditional oncology research and may play a significant role in understanding the catalysts behind cancer." Dr. Komarova is presently organizing a DIMACS workshop.

- **Jim Landwehr**, DIMACS member from AVAYA, received the 2003 **Outstanding Statistical Application Award** from the American Statistical Association, August 2003, joint with Steve Golowich and Scott Vander Wiel of Bell Labs for a paper published in Technometrics, vol 44 no. 3, 2002, "Interplay Between Physics and Statistics for Modeling Optical Fiber Bandwidth" He received the 2002 **Wilcoxon Award from the American Society for Quality** for the best applications paper to appear in Technometrics, for the same paper as above; awarded in Oct. 2003. Dr. Landwehr is a member of the DIMACS Executive Committee and serves on the DIMACS Council.

- **Simon Levin**, DIMACS member from the Princeton University Department of Ecology and Evolutionary Biology, became a **member of the American Philosophical Society**, May, 2003 and received the **Distinguished Landscape Ecologist Award**, U.S. Regional of the Association International Association for Landscape Ecology, 2003. Dr. Levin is an organizer of the DIMACS International Conference on Computational and Mathematical Epidemiology.

- **Andras Prekopa**, DIMACS member from Rutgers University - RUTCOR, received the 2003 **EURO Gold Medal**, the gold medal of the European Operational research Societies, given away biannually for outstanding research in OR.

- **Mauricio G.C. Resende**, DIMACS member from AT& T Labs - Research, and Panos Pardalos were editors of the book "Handbook of Applied Optimization," Oxford University Press in 2002, one of two books to be named **"Outstanding Professional and Scholarly Titles of 2002"** in the Computer Science category by the Association of American Publishers. It received an Honorable Mention on January 23, 2003.

- **Michael Rothkopf**, DIMACS member from Rutgers University - RUTCOR, was selected as **Fellow of INFORMS**, Fall 2002 and elected **president-elect of INFORMS**, 2003. Dr. Rothkopf is an organizer of the 2004-2006 Special Focus on Computation and the Socio-Economic Sciences and has previously organized a DIMACS workshop and served as an REU program mentor.

- **Rob Schapire**, DIMACS member from the Princeton University Department of Computer Science, with Yoav Freund, was awarded the 2003 **Gödel Prize for outstanding journal article in the area of theoretical computer science** for their paper "A Decision Theoretic Generalization of On-Line Learning and an Application to Boosting," Journal of Computer and System Sciences 55 (1997), pp. 119-139. The Gödel Prize for outstanding papers in the area of theoretical computer science is sponsored jointly by the European Association for Theoretical Computer Science (EATCS) and the Special Interest Group on Algorithms and Computing Theory of the Association of Computing Machinery (ACM-SIGACT). Dr. Schapire has been a workshop organizer at DIMACS.

- **Amin Shokrollahi**, DIMACS member from Digital Fountain, received the **IEEE Information Theory Best Paper Award**, from the IEEE Information Theory Society, July 2003. Dr. Shokrollahi is an organizer of the 2001-2004 Special Focus on Computational Information Theory and Coding.

- **Peter Shor**, DIMACS member from AT& T Labs - Research, was **elected to the National Academy of Sciences** in 2002. Dr. Shor is a former member of the DIMACS Council.

- **Eduardo Sontag**, DIMACS member from the Rutgers University Department of Mathematics, received the **Bode Prize from the IEEE**, December 2003. Dr. Sontag is a DIMACS workshop organizer.

- **Steven J. van Enk**, DIMACS member from Bell Labs, was awarded the **Gilbreth Lectureship**, from the National Academy of Engineering, October 12, 2003.

- **Wolmer Vasconcelos**, DIMACS member from the Rutgers University Department of Mathematics, was **elected to the Brazilian Academy of Sciences** [Academia Brasileira de Ciencias], April 2003.

- **Andrew C. Yao**, DIMACS member from the Princeton University Department of Computer Science, received the **Pan Wen-Yuan Research Award** from the Pan Wen-Yuan Foundation, Taipei, August 2003 and was elected Fellow of the Association for the Advancement of Science, October 2003. Dr. Yao received a Doctor of Science (Honorary Degree), City University of Hong Kong, November 2003. He is a former DIMACS co-director.

8.4 Research, Education, and International Partnerships

A key goal of DIMACS is outreach to other communities, both within and without the mathematical sciences. In the past year, enhancing and expanding our many partnerships has remained a priority of DIMACS.

Center for Math, Science, and Computer Education

A key partner of DIMACS is the Rutgers Center for Math, Science, and Computer Education. With CMSCE, we run the many K-12 educational programs described in Section 6.6, in particular the Leadership Program for precollege teachers and the Young Scholars Program for high school students. Our long-standing partnership with CMSCE has played a crucial role in our success in these programs.

Institute for Advanced Study

The highly successful partnership with the Institute for Advanced Study has continued with joint workshops and shared postdocs. During 2002-03 we shared two postdocs. Michael Capalbo spent 2001-02 at IAS and 2002-03 at Rutgers, while Amit Chakrabarti spent 2002-2003 at IAS. We have agreed to continue the joint postdoc program for at least another three years, with substantial assistance from Rutgers. IAS has been a major partner in our special foci. The Special Focus on Discrete Probability was our first major joint event. A joint event was held at IAS during our Special Year on Large Scale Discrete Optimization. Avi Wigderson, a permanent part-time faculty member at IAS, has been a key organizer of our 1999-00 Special Year on Computational Intractability and arranged a follow-up year at IAS, with continuing activities at DIMACS. Senior IAS researchers have also played an important role in our new Special Focus on Computational and Mathematical Epidemiology by serving on the Organizing Committee, Advisory Committee and workshop/working group organizing committees, and a major special focus workshop is scheduled to be held at IAS. An IAS faculty member is also involved in the planning for the new Special Focus on Communication and the Socio-Economic Sciences, by serving on the special focus Organizing Committee and on a workshop organizing committee.

Partnerships with the Biological Sciences

In Computational Molecular Biology our special focus activities have engaged us in serious collaborations with a number of centers around the country. The departments of Ecology and Evolutionary Biology and of Molecular Biology at Princeton have been involved since the beginning of both of these activities. Professor Laura Landweber of the Molecular Biology Department at Princeton has been one of the organizers of the DNA Computing Special Focus and of a number of its workshops. We have partnered with the Center for Molecular Biophysics and Biophysical Chemistry at Rutgers University in joint sponsorship of a series of workshops over the years. The Director of the Center, Wilma Olson, is one of the organizers of our Special Focus on Computational Molecular Biology. She has also been mentoring in the DIMACS REU program for four summers. We have also partnered in workshops with the Program in Mathematics and Molecular Biology (PMMB) at Florida State University. PMMB has been a sponsor of a workshop in the Computational Molecular Biology Special Focus. The Waksman Institute has partnered with DIMACS on a variety of activities, including playing a crucial leadership role in our 6-year Special "Year" on Mathematical Support for Molecular Biology. In the past year, DIMACS and the Biomaps Institute have partnered in new and important ways, in particular receiving a grant for a tutorial program. Fred Roberts is on the Biomaps Steering Committee and Ron Levy, Biomaps Director, is on the Organizing Committee for the DIMACS Special Focus on Computational Molecular Biology.

Partnerships in Computational and Mathematical Epidemiology

Our Special Focus on Computational and Mathematical Epidemiology has opened up lines of communication with a variety of new partners. Through the efforts of Simon Levin, the Department of Ecology and Evolutionary Biology at Princeton has become a major partner in this endeavor. This focus, with its homeland security emphasis, has also led to direct contacts with federal and state agencies, including cooperation with the federal Department of Health and Human Services, Defense Threat Reduction Agency, the Centers for Disease Control and Prevention, the New Jersey Department of Health and Human Services and Department of Justice, and the New York City Department of Health. Faculty from the Environmental and Occupational Health Sciences Institute (EOHSI) have been heavily involved in planning the special focus.

Partnerships in Homeland Security Research

The DIMACS homeland security research initiatives have involved us with numerous government agencies and other academic units. We have active communication and collaboration with the Centers for Disease Control and Prevention and with the intelligence community. We have strong connections with groups at the MITRE Corporation, the Institute for Defense Analyses (see below), the federal Department of Health and Human Services Office of Emergency Preparedness (through its smallpox modeling task force), the NJ Department of Health and Senior Services, and the New York City Department of Health. We have opened up lines of communication with Los Alamos National Laboratory about joint projects. We have become engaged with numerous new academic partners in these endeavors, and especially through our work with the Rutgers University Homeland Security Research Initiative. Prominent among these are the Environmental and Occupational Health Sciences Institute (EOHSI) and the Center for BioDefense at the UMDNJ Newark campus.

Institute for Defense Analyses
Center for Communications Research, Princeton

The Institute for Defense Analyses - Center for Communications Research (IDA-CCR) in Princeton has partnered with DIMACS in the past in shared postdoc programs. Helene Kulsrud from IDA-CCR served on the steering committee for our Massive Data Sets Special Year, and she organized one of the special year workshops. Victor Miller from IDA served on the organizing committee for our January 2000 workshop on "Unusual Applications of Number Theory." During FY2001, we were approached by IDA-CCR to help them open their new public facility in June 2002 and began planning a tutorial and workshop on quantum computing. After the events of September 11, 2001, we changed the topic to data mining for homeland security and had a rousing opening event the new IDA-CCR facility. The Director of IDA-CCR was also very supportive in helping us formulate our Monitoring Message Streams project that was funded through the intelligence community. We have also opened up lines of communication with the Maryland branch of IDA through our bioterrorism working group and our bioterrorism sensor location project.

Center for Advanced Information Processing, CAIP

We share a building with CAIP and have interacted on a variety of projects over the years. An early collaboration involved the participation of CAIP faculty in the DIMACS Special Year on

Massively Parallel Computing. We share interest in and partnerships with the telecom industry in New Jersey and interact regularly with and through these partnerships. More recently, we have interacted in homeland security initiatives and in plans for projects involving autonomic computing.

Wireless Information Network Laboratory, WINLAB

At Rutgers, our simulations projects have been carried out jointly with faculty at WINLAB, the Wireless Information Network Laboratory, and WINLAB faculty were heavily involved in our research projects on simulations of telecommunications systems and in our Special Focus on Next Generation Networks Technologies and Applications, as members of the Advisory Committee and workshop organizers. A WINLAB professor is a key organizer of our Special Focus on Computational Information Theory and Coding. In addition, several postdocs, graduate students, and undergraduates at DIMACS have been mentored by WINLAB faculty. WINLAB brings to this collaboration extensive practical experience in wireless and mobile computing topics.

Professional Societies and Sister Centers

The Institute for Mathematics and its Applications, IMA, at the University of Minnesota, is one of NSF's mathematical sciences research institutes. We developed, as part of our Special Focus on Next Generation Networks Technologies and Applications, several joint workshops. A September 2002 IMA workshop on bioterrorist defense was a natural follow-up to one at DIMACS in which the director of IMA participated. Further collaboration is under discussion.

At the January 2003 American Mathematical Society national meeting, a number of the research centers in the mathematical sciences held an informational session to introduce the community to programs at these centers. DIMACS shared in the organization of that session with IMA, the Mathematical Sciences Research Institute (MSRI) at Berkeley, the National Institute of Standards and Technology (NIST), the Institute for Pure and Applied Mathematics at UCLA, and other national and international centers.

DIMACS has connections and joint programs with a number of professional societies. We have joined in various projects with SIGACT, the Special Interest Group in Algorithms and Computation Theory of the Association for Computing Machinery. We have sponsored meetings for the NJ Chapter of the Mathematical Association of America and of AMATYC, the American Mathematical Association of Two Year Colleges. AMATYC has also been a strong supporter of our "Reconnect" program. We organized a workshop on Challenges for Theoretical Computer Science with SIAM, the Society for Industrial and Applied Mathematics, and prepared an NSF report on this topic. IEEE, the Institute of Electronic and Electrical Engineers, has also been a collaborator on a workshop.

International Partnerships

DIMACS has also been actively involved in a variety of international partnerships. These partnerships are described in Section 5.4.

9 Administration and Management

9.1 Management of the Center

Management

During 2002-2003 the center management consisted of:

- Director: Fred S. Roberts, RU
- Co-Director: Robert Tarjan, PU
- Associate Director: Mel Janowitz, RU
- Acting Associate Director: Alexander Barg, RU
- Associate Director for Education: Joseph Rosenstein, RU

On December 21, 2002, Alexander Barg, formerly of Bell Labs, joined us as Acting Associate Director, a temporary position. In the Spring of 2003 we began a search for an Associate Director and a one year appointment was accepted by Brenda J. Latka, Lafayette College.

Executive Committee

The center is run by a series of committees. The major decision making body of the center is the Executive Committee. This committee consists of the management plus one representative from each DIMACS unit, the units being the four full industrial members plus Princeton CS, Rutgers CS, Rutgers Math, and Rutgers OR. The Executive Committee meets monthly during the academic year and otherwise as needed. During 2002-2003, the Rutgers members of the Executive Committee were Fred Roberts (Rutgers), Alexander Barg (DIMACS), Lance Fortnow (NEC), Peter Hammer (Rutgers), Mel Janowitz (DIMACS), David Johnson (AT&T), Jon Kettenring (Telcordia), Joseph Rosenstein (Rutgers), Michael Saks (Rutgers), Mario Szegedy (Rutgers), Robert Tarjan (Princeton), and Peter Winkler (Bell Labs). Non voting members represented our affiliated partners: Jennifer Chayes (Microsoft), Jim Landwehr (Avaya), Baruch Schieber (IBM). (Bill Horne (HP) is a new non voting member in FY04.)

DIMACS Council

The center has a large advisory committee of local leaders, called the DIMACS Council. The Council consists of the Executive Committee, one additional representative from each DIMACS unit, and the chairs of all ongoing special foci and DIMACS committees. The Council meets twice a year to discuss general issues concerning the center and the fields of dm/tcs. During 2002-2003, the members of the Council, in addition to those listed above, were Farid Alizadeh (Rutgers), Eric Allender (Rutgers), Endre Boros (Rutgers), Adam Buchsbaum (AT&T), Robert Calderbank (AT&T), Martin Farach-Colton (Rutgers), Steven Fortune (Bell Labs), Michael Fredman (Rutgers), Joe Kilian (NEC), Janos Komlos (Rutgers), S. Muthukrishnan (Rutgers), Charles Sims (Rutgers), Serap Savari (Bell Labs), Peter Shor (AT&T), and David Waltz (NEC).

Other Committees

Other committees are the Postdoctoral Selection Committee, Projects Committee, Education Committee, Graduate Student Selection Committee, Visitor Committee, Computing Equipment Committee, Strategic Planning Committee, and Budget Committee. Each has one representative from each DIMACS unit.

9.2 Advisory Board

The DIMACS Advisory Board met once a year while we had the STC award. During 2002-2003, we had a meeting of a subgroup of the Board and several telephone conference call meetings. Members of the Board are as follows:

- Jennifer Chayes, Microsoft Corp.
- Brenda Dietrich, IBM Research.
- Cynthia Dwork, IBM Almaden Research Center
- Ronald Graham, University of California, San Diego
- Juris Hartmanis, Cornell University
- Richard Karp, University of California, Berkeley and International Computer Science Institute
- Daniel Kleitman, M.I.T.
- Laszlo Lovasz, Microsoft Corp.
- Joseph O'Rourke, Smith College
- Christos Papadimitriou, University of California/Berkeley
- William Pulleyblank, IBM Watson Research Center
- Dana Scott, Carnegie Mellon University
- Eva Tardos, Cornell University

9.3 Staff

Ms. Maria Mercado joined DIMACS in December, 2002 as our new Workshop Coordinator. She replaced Jennifer Thiemann, who left in July. Mr. Walter Morris joined DIMACS in March, 2003 as our new Unit Computing Manager. He replaced Pablo Samuel who left in January. Ms. Linda Casals joined DIMACS in March, 2003 as Publicity Coordinator, replacing Eric Luhrs, who left in January. Other members of the staff are Christine Spassione, Visitor Coordinator; Jessica Herold, Financial Assistant; and Sarah Donnelly, Assistant Director.

10 Financing the Center

The planned end of DIMACS NSF "science and technology center" award in January 2000, with its large and stable source of funding (approximately $2M per year) required us to raise money for our many projects individually, while challenging us to find ways to support center infrastructure.

We now write a large number of grant proposals each year, going to numerous agencies and private sources and numerous programs within those agencies. NSF remains our primary sponsor, but even within NSF, we have obtained funds from numerous programs, including Computer Science; Mathematical Sciences; Social, Behavioral and Economic Sciences; Biological Sciences; Education and Human Resources; and International Programs.

Rutgers-provided return on indirect costs charged in grants is a primary source of support for DIMACS. An "emergency fund" built up from such indirect cost return is crucial in allowing us to make commitments for programs and projects several years in advance, while recruiting program and project organizers to help us raise funds for them.

Throughout the lifetime of the center, our academic and industrial partners have been strong supporters, especially Rutgers, which has played the central leadership role in DIMACS. Our industrial partners continue to be strong supporters of the center and play a crucial role in the projects and programs of DIMACS, partly through financial support, but mostly through heavy involvement of leading researchers in our center.

In FY03, we continued to receive support from the New Jersey Commission on Science and Technology, which has supported DIMACS since our founding. However, this year marked the last year of NJCST funding. The Commission's $100K annual contributions have been very helpful in supporting both core and exploratory programs at DIMACS, as "matching" for federal grants, and in supporting the DIMACS infrastructure.

We also bring in some funds through registration fees for workshops. While we keep them relatively low, especially for students, these fees are a modest source of income and they help pay for some of our workshop staff support.

New Awards:

The following grants are for projects that are newly funded in FY2003. Many of these awards are for multi-year activities. The amount of multi-year awards is noted in total, with the FY2003 appropriation noted as well.

Granting agency/Project Title	Project Period	Total Award	FY03 award
NSF - Monitoring Message Streams	7/1/02-6/30/03	$700,000	$700,000
NSF - Mining Multilingual Resources Using Text Analytics	7/1/03-8/31/04	$693,000	$693,000
NSF - Special Focus on Computational and Mathematical Epidemiology	9/1/02-8/31/07	$2,750,000	$545,387
NSF - DIMACS-DIMATIA-Renyi Collaboration	7/1/02-6/30/05	$155,949	$51,983
NSF - Satellite Reconnect Project	7/15/02-6/30/06	$616,846	$86,955
NSF - DIMACS Special Focus on Computational Information Theory and Coding.	7/1/02-6/30/04	$190,197	$106,448
NSF - DIMACS Exploratory Postdoctoral Program in Computational Epidemiology.	4/15/03-8/31/04	$100,000	$100,000
NSF - Workshop on Information Processing in the Biological Organism	5/1/03-4/30/04	$56,609	$56,609

NSF - Workshop on Software Security	1/1/03-1/31/04	$15,737	$15,737
NSF - Tutorial on Introduction to Modern Concepts in Biology for Mathematical and Physical Scientists.	7/1/02-6/30/03	$25,051	$25,051
NSF - Workshop on Data Depth: Robust Robust Multivariate Analysis, Computational Geometry and Applications.	7/1/02-9/30/03	$18,000	$18,000
NSA - Workshop on Data Depth: Robust Robust Multivariate Analysis, Computational Geometry and Applications and the Workshop on Video Mining.	7/1/02-9/30/03	$20,000	$20,000
The State of New Jersey, Commission on Science and Technology - DIMACS Center Grant	9/1/02-8/31/03	$100,000	$100,000
Univ. of CA - Workshop on Complexity and Inference.	1/7/03-1/6/04	$20,000	$20,000
Burroughs-Wellcome Fund - Summer School Tutorials in Epidemiology for Mathematical Scientists and the Workshop on The Pathogenesis of Infectious Diseases: Host-Pathogen Dynamics.	8/1/02-9/30/02	$35,000	$35,000
ONR - Working Group Meetings on Analogies Between Computer Viruses and Immune Systems and Biological Viruses and Immune Systems.	1/15/02 - 1/14/03	$25,377	$25,377
Microsoft Research - Workshop on Software Security	7/1/02-6/30/03	$5,000	$5,000
Total new awards		$5,526,766	$2,604,547

Continuing Awards:

The following grants were awarded in prior fiscal years. Their research activities continue through FY 2003.

Granting agency/Project Title	Project Period	Total Award	FY03 award
NSF - DIMACS-DIMATIA International REU Program.	5/15/02-4/30/07	$654,305	$109,018
NSF - Working Groups in Data Analysis and Mining.	9/15/01-8/31/04*	$80,000	$29,627
NSA - Working Groups in Data Analysis and Mining.	10/1/01-9/30/02	$17,500	$0
NSF - Planning Workshop on Quantum Communications, Cryptography and Coding.	1/1/02-12/31/02	$24,795	$0
NSF - Working Group on Mathematical Sciences Methods for the Study of Deliberate Releases of Biological Agents and their Consequences.	3/1/02-2/29/04*	$39,988	$0
NSF - U.S.-European Union Joint Conferences	8/15/01-7/31/04	$344,052	$113,229

NSF - DIMACS Special Focus on Mathematics and the Foundations of Computer and Information Science.	9/1/00-8/31/03*	$120,000	$0
NSF - Three Special Focus Programs at DIMACS	9/1/00-8/31/04*	$900,000	$0
NSF - DIMACS Connect Institute	6/1/00-5/31/04*	$766,292	$0
AT&T Foundation - DIMACS Connect Institute	6/1/00-5/31/03	$90,000	$0
Alfred P. Sloan Foundation - DIMACS Special Focus on Computational Molecular Biology.	11/1/99-8/31/04*	$361,778	$100,731
The Alfred P. Sloan Foundation - International Conference on Computational and Mathematical Epidemiology.	2/1/02-9/30/02	$45,000	$0
The State of New Jersey, Commission on Science and Technology - DIMACS Center Grant	9/1/01-8/31/02	$100,000	$0
Total continuing awards		$3,543,710	$352,605

* Includes a one year no-cost extension.

Gifts from DIMACS Partners and Affiliate Members:

The financial support from our industrial partners continues, but in this difficult economic time for most of them, is down from previous highs. It was never a major source of funds for us. Still, the total value of direct support from our industrial partners was $100,781 in FY03. The major contribution of our partners is in the heavy involvement of senior and junior researchers in DIMACS programs. Our four full industrial partners, AT&T Labs, Bell Labs, NEC Laboratories America, and Telcordia Technologies, valued this in-kind contribution to DIMACS at $1.6M in FY03 and it is a main reason that DIMACS is so successful.

A word about the "in-kind" contributions is in order. Several of our industrial partners appointed DIMACS visitors and postdoctoral fellows and paid them directly. The value placed on in-kind contributions includes an estimated $330K for these types of contributions. All of our industrial partner institutions strongly encourage volunteer work by their scinentists and the number of hours spent on DIMACS committee work and organizational activities by their scientists is very large. For example, over 40 permanent members of DIMACS, representing all of the six member organizations, are actively involved in the committees that run the center and its programs, the Executive Committee, Council, Strategic Planning Committee, Budget Committee, Projects Committee, Education Committee, Postdoc Selection Committee, Equipment Committee, Graduate Student Selection Committee, REU Committee. More are involved in an organizational role in our many special focus programs, workshops, and educational projects. Our four full industrial partners have estimated the value of their FY03 in-kind contributions through the participation of their employees in DIMACS organizational activities at $1,160,000.

We have not tried to get similar estimates of "in-kind" contributions from our affiliated partners, though they too are substantial.

11 Future Plans

11.1 Current Special Foci

11.1.1 2000-2004 Special Focus on Computational Molecular Biology

DIMACS-Celera Genomics/Applied Biosystems Graduate Student Award

Funds have been received from Celera Genomics/Applied Biosystems to continue this graduate student award program. We expect to make one new award in FY04.

Workshops, Tutorials, and Working Group Meetings

- **Workshop on Data Mining Techniques in Bioinformatics**
  - October 30 - 31, 2003
  - DIMACS Center, CoRE Building, Rutgers University, Piscataway, NJ
  - Organizers: Mona Singh, Princeton University, and Mark Gerstein, Yale University

- **Short Course: A Field Guide to GenBank and NCBI Molecular Biology Resources**
  - December 10 - 11, 2003
  - DIMACS Center, CoRE Building, Rutgers University
  - Organizers: Paul Ehrlich, BIOMAPS Institute, Mel Janowitz, DIMACS, Tara Matise, Department of Genetics

- **Working Group on New Algorithms for Inferring Molecular Structure from Distance Restraints**
  - January 12 - 16, 2004
  - DIMACS Center, CoRE Building, Rutgers University
  - Organizers: Michael W Trosset, College of William & Mary

- **Workshop on the Interface between Biology and Game Theory**
  - April 5, 2004
  - DIMACS Center, CoRE Building, Rutgers University
  - Organizers: Adam Arkin, Lawrence Berkeley Labs and UC Berkeley, Vijay Vazirani, Georgia Tech, Denise Wolf, Lawrence Berkeley Labs

This working group is a subgroup of the Working Group on Algorithms for Multidimensional Scaling.

11.1.2 2000-2003 Special Focus on Mathematics and the Foundations of Computer and Information Science

Workshops, Tutorials, and Working Group Meetings

- **Workshop on Applications of Lattices and Ordered Sets to Computer Science**
  - July 8 - 10, 2003
  - DIMACS Center, CoRE Building, Rutgers University
  - Organizers: Jonathan Farley, Massachusetts Institute of Technology, Mel Janowitz, DIMACS, Jimmie Lawson, Louisiana State University, Michael Mislove, Tulane University

11.1.3 2001-2004 Special Focus on Computational Information Theory and Coding

Workshops, Tutorials, and Working Group Meetings

- **Workshop on Algebraic Coding Theory and Information Theory**
  - December 15 - 18, 2003
  - DIMACS Center, Rutgers University, Piscataway, NJ
  - Organizers: Alexei Ashikhmin, Bell-Labs, Alexander Barg, DIMACS, Iwan Duursma, University of Illinois

- **Working Group on Optical/Magnetic Recording and Optical Transmission**
  - March 22 - 24, 2004
  - DIMACS Center, CoRE Building, Rutgers University
  - Organizers: Emina Soljanin, Bell Labs, Paul Siegel, University of California - San Diego, Bane Vasic, University of Arizona, Adriaan J. van Wijngaarden, Bell Laboratories

- **Workshop on Information Theory and Symbolic Dynamics**
  - March 25 - 26, 2004
  - DIMACS Center, CoRE Building, Rutgers University
  - Organizers: Emina Soljanin, Bell Labs, Paul Siegel, University of California - San Diego, Bane Vasic, University of Arizona, Adriaan J. van Wijngaarden, Bell Laboratories

- **Princeton-Rutgers Seminar Series in Communications and Information Theory 2003 - 2004**

  This series will continue throughout the year, with meetings at both institutions, and partially supported by Cisco.

11.1.4 2001-2004 Special Focus on Data Analysis and Mining

Workshops, Tutorials, and Working Group Meetings

- **Workshop on Discrete Metric Spaces and their Algorithmic Applications**
  - August 20 - 23, 2003
  - Friend Center, Bowl 006, Princeton University, Princeton, NJ
  - Cosponsored by the Institute for Advanced Study and Princeton University
  - Organizers: Moses Charikar, Princeton University
    Piotr Indyk, MIT
    Nati Linial, Hebrew University
    Jiri Matousek, Charles University
    Yuri Rabinovich, University of Haifa
    Gideon Schechtman, Weizmann Institute

- **Workshop on Data Mining Techniques in Bioinformatics**
  - October 30 - 31, 2003
  - DIMACS Center, CoRE Building, Rutgers University, Piscataway, NJ
  - Organizers: Mona Singh, Princeton University
    Mark Gerstein, Yale University

- **Workshop on Data Quality, Data Cleaning and Treatment of Noisy Data**
  - November 3 - 4, 2003
  - DIMACS Center, CoRE Building, Rutgers University, Piscataway, NJ
  - Organizer: Parni Dasu, AT&T Labs

- **Working Group on New Algorithms for Inferring Molecular Structure from Distance Restraints**
  - January 12 - 16, 2004
  - DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Michael W Trosset, College of William & Mary

  This working group is a subgroup of the DIMACS Working Group on Algorithms for Multidimensional Scaling.

- **Workshop on Geographical Information Systems: Representation Algorithms for Terrain**
  - Date of Workshop: TBA
  - Research Triangle, North Carolina
  - Organizers: Lars Arge, Duke University
    Jack Snoeyink, University of North Carolina, Chapel Hill

- **Workshop on Compact Representation of Data**
  - Dates: TBA
  - DIMACS Center, Rutgers University, Piscataway, NJ
  - Organizer: Moses Charikar, Princeton University

Workshops Under Development:

- **Workshop: Algorithms for Clustering in Data Mining**
  - Dates: TBA (tentatively 2003-2004)
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizers: Pierre Hansen, GERAD - University of Montreal; Moses Charikar, Princeton University

- **Computer-Generated Conjectures from Graph Theoretic and Chemical Databases II (Working Group)**
  - Dates: TBA
  - Location: Centre de Recherches Mathematiques (CRM), Universite de Montreal, Quebec, Canada
  - Organizers: Patrick Fowler, University of Exeter; Pierre Hansen, GERAD - University of Montreal

11.1.5 2002-2007 Special Focus on Computational and Mathematical Epidemiology

Workshops, Tutorials, and Working Group Meetings

- **Working Group: Genetics and Evolution of Pathogens**
  - Dates: November 24 - 25, 2003
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizers: Zhilan Feng, Purdue University

- **Working Group Meeting: Privacy / Confidentiality of Health Data**
  - Date: December 10 - 12, 2003
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Rakesh Agrawal, IBM Almaden; Larry Cox, CDC; Joe Fred Gonzalez, CDC; Harry Guess, University of North Carolina

- **Working Group Meeting: Data Mining and Epidemiology II**
  - Dates: March 18 - 19, 2004
  - Location: DIMACS Center, CoRE Building, Rutgers University

- Organizers: James Abello, Rutgers University; Graham Cormode, Rutgers University; Kenton Morgan, University of Liverpool; David Ozonoff, Boston University

- **Workshop: The Evolution and Control of Drug Resistance**
  - Dates: May 3-4, 2004
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizers: Sally Blower, UCLA Medical School and Tom Lietman, UCSF Medical School

- **Working Group Meeting: Methodologies for Comparing Vaccination Strategies**
  - Dates: May 17 - 21, 2004
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizers: John Glasser, CDC and Herbert Hethcote, University of Iowa

- **Workshop: Genomic Instability in Cancer: Biological and Mathematical Approaches**
  - Dates: June 8 - 9, 2004
  - Location: Institute for Advanced Study, Princeton, New Jersey
  - Organizers: Natalia Komarova, Rutgers University and Institute for Advanced Study, and Arnold Levine, Institute for Advanced Study, and Robert Wood Johnson School of Medicine and Dentistry
  - Cosponsored by the Institute for Advanced Study

Workshops under development:

- **Working Group Meeting: Distributed Computing, Social Networks and Disease Spread Processes (First Meeting, 2003-2004)**
  - Organizers: Vasek Chvatal and Fred Roberts, Rutgers University; Aleksandar Pekec, Duke University

- **Workshop: Models/Methodological Problems of Botanical Epidemiology (2005-2006)**
  - Organizer: Chris Gilligan, Cambridge

- **Workshop: Disease Clusters (2004-2005)**
  - Organizer: Daniel Wartenberg, Robert Wood Johnson Medical School

- **Working Group Meeting: Phylogenetic Trees and Rapidly Evolving Diseases, Meeting I**

- Organizers: Edward Holmes, Oxford University and Mike Steel, University of Canterbury

- **Workshop: The Epidemiology and Evolution of Influenza (2004-2005)**
  - Organizers: Catherine Macken and Alan Perelson, Los Alamos

- **Workshop: Models of Co-Evolution of Hosts and Pathogens (2006-2007)**
  - Organizers: Viggo Andreasen, Roskilde; Andrea Pugliese, Trento

- **Workshop: Genetics and Evolution of Pathogens II**
  - Dates: TBA
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizers: David Krakauer, Santa Fe Institute

- **Working Group Meeting: Order-theoretic Aspects of Epidemiology (First Meeting, 2004-2005)**
  - Organizers: David Ozonoff, Boston University; Melvin Janowitz and Fred Roberts, Rutgers University

- **Workshop: Spatial Epidemiology and Geographical Information Systems (2005-2006)**
  - Organizer: Daniel Wartenberg, Robert Wood Johnson Medical School

- **Working Group Meeting: Predictive Methodologies for Infectious Diseases (First Meeting, 2005-2006)**
  - Organizers: Sally Blower, UCLA Medical School and Christl Donnelly, Imperial College

- **Workshop: Capture-recapture Models in Epidemiology (2005-2006)**
  - Organizer: David Madigan, Rutgers University

- **Workshop: Ecologic Inference (2006-2007)**
  - Organizers: Tom Webster, Boston University and Sylvia Richardson, Imperial School of Medicine

- **Workshop: Combinatorial Group Testing (2006-2007)**
  - Organizers: Ding-zhu Du, University Minnesota; Frank Hwang, Chiatong University

- **Tutorial: Discrete Math and Theoretical Computer Science for Epidemiologists and Biologists (Summer 2004)**
  - Organizers: Alun Lloyd, Institute for Advanced Study; Fred Roberts, Rutgers University; Katherine St. John, CUNY

11.1.6 2002-2005 Special Focus on Computational Geometry and Applications

Workshops, Tutorials, and Working Group Meetings

- **Workshop on Computer Aided Design and Manufacturing**
  - October 7 - 9, 2003
  - DIMACS Center, CoRE Building, Rutgers University
  - Organizers: Deba Dutta, University of Michigan, Ann Arbor Ravi Janardan, University of Minnesota, Minneapolis Michiel Smid, Carleton University

Workshops under development:

- **Workshop on Geographical Information Systems: Representation Algorithms for Terrain**
  - Date of Workshop: TBA
  - Research Triangle, North Carolina
  - Organizers: Lars Arge, Duke University
    Jack Snoeyink, University of North Carolina, Chapel Hill

- **Workshop: Fundamental Geometric Algorithms**
  - Dates: TBA (tentatively 2003-2004)
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizers: Bill Steiger, Rutgers University

- **Algorithm Implementation Challenge: Surface Reconstruction**

11.1.7 2003-2006 Special Focus on Communication Security and Information Privacy

Workshops, Tutorials, and Working Group Meetings

- **Tutorial: Applied Cryptography and Network Security**
  - Dates: August 4 - 7, 2003
  - Location: Rutgers University
  - Organizers: Rebecca Wright, Stevens Institute of Technology

- **Workshop: Large-scale Internet Attacks**
  - Date: September 23 - 24, 2003
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Vern Paxson, ICSI Center for Internet Research

- **Working Group: Privacy / Confidentiality of Health Data**
  - Date: December 10 - 12, 2003

- Location: DIMACS Center, CoRE Building, Rutgers University
- Organizer: Rakesh Agrawal, IBM Almaden; Larry Cox, CDC; Joe Fred Gonzalez (chair), CDC; Harry Guess, University of North Carolina

- **Working Group: Privacy-Preserving Data Mining**
  - Date: March 15 - 17, 2004
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Rebecca Wright, Stevens Institute of Technology; Benny Pinkas, HP Labs; Cynthia Dwork, Microsoft

- **Workshop: Electronic Voting – Theory and Practice**
  - Date: May 27 - 28, 2004
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Markus Jakobsson and Ari Juels, RSA Laboratories

- **Workshop: Security Analysis of Protocols**
  - Date: June 7 - 9, 2004
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: John Mitchell, Stanford and Ran Canetti, IBM Hawthorne

- **Workshop: Usable Privacy and Security Software**
  - Date: July 7 - 8, 2004
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Lorrie Cranor, AT&T; Mark Ackerman, University of Michigan; Fabian Monrose, Johns Hopkins University; Andrew Patrick, NRC Canada; Norman Sadeh, Carnegie Mellon University

- **Working Group: Usable Privacy and Security Software**
  - Date: July 9, 2004
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Lorrie Cranor, AT&T; Mark Ackerman, University of Michigan; Fabian Monrose, Johns Hopkins University; Andrew Patrick, NRC Canada; Norman Sadeh, Carnegie Mellon University

- **Workshop: Cryptography: Theory Meets Practice**
  - Date: October 14 - 15, 2004
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Dan Boneh, Stanford

Workshops under development:

- **Workshop: Intellectual Property Protection**
  - Date: TBA
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Drew Dean, SRI International
- **Workshop: Security of Web Services and E-Commerce**
  - Date: TBA
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Brian LaMacchia, Microsoft
- **Workshop: Mobile and Wireless Security**
  - Date: TBA
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Bill Arbaugh, University of Maryland
- **Working Group: Intrusion Detection and Network Security Management Systems**
  - Date: TBA
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Joe McHugh, CERT, and S. Raj Rajagopalan, Telcordia Technologies
- **Workshop: Security and Trust Issues Associated with Ad-Hoc Computing / Pervasive Networking**
  - Date: TBA
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: S. Raj Rajagopalan, Telcordia Technologies
- **Workshop: Database Security: Query Authorization and Information Inference**
  - Date: TBA
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: H.V. Jagadish, University of Michigan
- **Working Group: Mobile Code Security**
  - Date: TBA
  - Location: DIMACS Center, CoRE Building, Rutgers University
  - Organizer: Dan Wallach, Rice University

11.2 Planned Special Foci

11.2.1 2004-2006 Computation and the Socio-Economic Sciences

Partnerships between mathematicians and social scientists have existed for a long time, but partnerships between computer scientists and social scientists are relatively new. In the past five years, they have begun to sprout and several important new fields of research are already thriving as a result. The topic of game theory and mechanism design has involved many leading computer scientists with economists and, as a result, for the first time many computer scientists are taking serious note of both the problems and the methods of economics. Analysis of the analogies between the growth of social networks and the development of the Internet have led to important insights about both areas and serious interactions between computer scientists and sociologists. Problems of metasearch have engaged computer scientists with political scientists and economists working on voting and social choice. We believe that the time is ripe to build on these early and exciting connections.

Partnerships between computer scientists and biologists are somewhat of a model for what we have in mind. These partnerships have flourished in the past 15 years. They have played a critical role in the human genome project, have led to a major emphasis on information processing in the biological organism, have spawned intriguing new areas of research in computer science such as DNA computing, and have contributed to major developments in tree reconstruction and pattern matching algorithms in computer science. In short, they have had a profound effect on both biology and computer science. There may not be a specific "big science" challenge such as the human genome project in the case of computer science and the social sciences, but the opportunities for real and significant progress both in the computing and social sciences as a result of new interdisciplinary partnerships are upon us. In fact, the interaction between computer science and the social sciences is changing in dramatic ways that can be expected to have lasting impacts on both disciplines. Many applications in computer science and information technology (IT) involve issues and problems that social scientists (economists, political scientists, sociologists, psychologists and others) have addressed for years, issues of preference, utility, decision making, conflict and cooperation, incentives, auctions, bidding, consensus, social choice, and measurement; and the methods social scientists have developed for dealing with these issues and problems form an impressive toolkit. At the same time, with the widespread availability of today's powerful computers and new and exciting data sets, work in economics, political science, sociology and psychology that was only a theoretical possibility a few years ago is becoming a reality. Applying methods of social science to CS/IT problems requires new computational tools and the development of new variants of these methods. Applying computational methods to the solution of modern social science problems requires the development of new data structures, algorithms, and other tools that are in the domain of the computer scientist.

We have begun to see the use of methods developed by social scientists in a variety of IT applications. The requirements associated with these applications place great strain on the social science methods because of the sheer size of the problems addressed, issues involving computational power of agents, limitations on information possessed by players, and the sequential nature of repeated applications. Hence, there is a great need to develop a new generation of methods to satisfy these CS/IT requirements. In turn, these new methods will provide powerful new tools for social scientists. At the same time, great progress is being made on the problems traditionally of interest to social scientists through the use of new methods for finding patterns in data, searching through databases, computing solutions, and testing models.

This Special Focus seeks to develop the new "social-science-based" CS methodologies and to investigate their application to problems of information technology and to problems of the social sciences of fundamental importance to modern society. It also seeks to investigate computer science tools that are especially relevant to emerging problems of the socio-economic sciences. Research into issues of the type we envision requires new interdisciplinary partnerships among computer scientists, mathematicians, operations researchers, economists, experts in business applications, political scientists, psychologists, sociologists, and researchers specializing in the handling of information. We will build on the extensive interdisciplinary experience at DIMACS to bring together such interdisciplinary groups in the course of this special focus. DIMACS special foci on Mathematical Support for Molecular Biology (1994-2000) and Computational Molecular Biology (2000-2004) have played an important role in bringing computer and biological scientists together. This special focus is aimed at doing a similar thing.

Several major research themes span the Special Focus. They include:

- **Computational Tractability/Intractability.** Limits on what can be efficiently computed are very important in social-science-based CS applications and applications of CS methods to the social sciences.
  - Can our understanding of computational complexity help us to build computational tools that will enlarge the set of social science models that can be analyzed and to adapt social-scientific concepts to solve problems of information technology?
  - How can we characterize situations where intractability is a good thing, as for example when we use it to protect privacy or make it difficult to manipulate the outcome of an election?
- **Limitations on Computational Power/Information.** Increasingly, economic decisions have to be made in situations where there are many "actors" each having partial information or where there are limits to computation that prevent actors from obtaining needed information. Limitations include:
  - informational asymmetries
  - computational limits to extracting information that is available (perhaps in large data sets)
  - cognitive limitations on the part of agents.
- **The Impact on New Methodologies of the Sheer Size of CS/IT Applications.** The sheer size of CS/IT applications requires new methods/protocols where efficiency is a central focus and where issues of scalability and robustness with respect to errors in data are prominent.
- **Learning through Repetition.** Issues of learning, taking advantage of repeated games or auctions or through combining repeated inputs, arise in CS/IT applications in fascinating ways. Some tasks may be delegated to software agents and one needs to understand how they should be designed to "learn" user preferences and how to bid and trade.
- **Security, Privacy, Cryptography**. The need for secure communication and privacy sparks the need for cryptographic methods in the analysis and design of voting procedures, games, and decision processes. Such issues arise in areas like electronic commerce and electronic voting, and many will be discussed in detail in the parallel special focus program on "Communication Security and Information Privacy" at DIMACS.
- **Game-theoretic Solution Concepts.** As we consider larger and larger games in modern economic applications, Nash equilibria, solution concepts in cooperative games, and other game-theoretic solution concepts applied in a variety of settings become harder and harder to compute.

- **Markets as Information Sources**. Increasingly individuals and firms are using markets to learn about their competitors' preferences and to aggregate and combine information, for example through the generation of prices. These developments can benefit from methods of distributed computing and combinatorial exchange.

- **Implementation of Auctions**. While dramatic changes in availability of information technology have allowed us to expedite increasingly complex business transactions, for example complex auctions, the ability to implement such transactions often puts a strain on existing computational methods.

- **Designing Markets**. How do we design markets for the exchange of complicated assets like spectrum, stock portfolios, landing rights, and airline routes? The problems involved are different from those of traditional auctions, as there are multiple owners whose assets interact in complex ways.

- **Dynamic Markets**. Ever-more-powerful computers make it increasingly possible to develop sophisticated algorithms that allow actors to adjust prices, bandwidth, etc. in a dynamic way, as well as adjust schedules and allocations in response to changing market conditions or unforeseen events.

- **Computing Utilities**. Utility functions, which have been studied for a long time by economists and psychologists, are increasingly important in considerations of interest to computer scientists. For both groups, there is the need to compute utilities in larger and larger environments and under conditions of distributed information. These call for new models, tools, and algorithms such as the use of polyhedral combinatorial methods which are so central to computer science.

Now would be an ideal time for this special focus. Leading computer scientists are seeking to learn and modify methods of the social sciences and are interacting with social scientists in increasingly sophisticated ways, while social scientists are more and more aware of the kinds of issues that are of central importance in computer science. Attendance at the early meetings organized around this kind of interface both at DIMACS and elsewhere has been large and enthusiastic and we are starting to see very serious research coming out of this interdisciplinary field. This should just be the beginning.

11.2.2 2005-2008 Discrete Random Systems

During the past decade there has been tremendous interplay between discrete mathematics, theoretical computer science, and statistical physics. This program brings together researchers from the various fields to explore topics at their interface.

The focus is on probabilistic algorithms and models which arise in the study of physical systems and combinatorial structures. This broad field requires insights and techniques from the study of phase transitions in statistical physics, probabilistic methods from combinatorics, and randomized algorithms in theoretical computer science.

This area is of increasing commercial, economic, and societal relevance as large-scale real-world systems, from biological epidemics to the world-wide web's structure to efficient highway emergency-repair scheduling, are successfully modeled stochastically.

Strong themes running through these interactions include:

- Phase transitions in (infinite-sized) physical systems such as Ising, Potts, and percolation models

- Rapid or torpid mixing of Markov Chain Monte Carlo (MCMC) processes in finite-sized physical systems
- Approximate counting and approximate (or perfect) uniform sampling of combinatorial structures such as perfect matchings, through MCMC and other techniques
- Randomized algorithms for combinatorial decision, optimization, and enumerated problems
- Random structures such as random graphs and random CNF formulas, including questions of existence, optimization, phase transitions, and connections between phase transitions and computational intractability
- Probabilistic combinatorics including extremal combinatorics
- Technical tools used in the above, including mathematical methods such as martingales, large-deviation inequalities, conductance and related measures, and the differential-equation method, and physics tools such as the replica and cavity methods
- Optimization and analysis for real-world problems and structures that may be modeled by random structures

Or, more compactly:

- Markov-Chain Monte Carlo and other random walks
- Gibbs measures and phase transitions
- Random structures and randomized algorithms
- Probabilistic combinatorics

When DIMACS hosted a workshop on "Statistical physics methods in discrete probability" in 1997, the two communities, roughly of physicists on the one hand and mathematicians and theoretical computer scientists on the other, strove to bridge the gulf between them. Today, the field we are calling Discrete Random Systems has matured beyond expectations.

The physicist's notion of a phase transition in an infinite system can often be related to a computer-science notion of slow mixing, and there are general theorems to convert between the two views. Jerrum and Sinclair's "conductance" bound on mixing rate has been realized to be a version of Cheeger's inequality, and there are a host of other methods for bounding mixing times. A version of the Lovasz Local Lemma, a result in probabilistic combinatorics, has been related to Dobrushin's inequality in physics. A prototypical exact but non-rigorous physical result for a combinatorial problem, random assignment, has been paralleled by rigorous probabilistic methods, and meanwhile new combinatorial problems, notably the SAT phase transition, have been addressed by new non-rigorous results in statistical physics. At least two leading researchers have proposed a sweeping program of rebuilding the great edifice of statistical physics on a solid mathematical foundation.

The field is advancing now at the greatest rate imaginable, and the timing for this program could not be better.

11.3 Multi-Disciplinary, Multi-Institutional Projects

11.3.1 Continuing Projects

The projects described in Section 4 will all continue. The **Monitoring Message Streams Project** will emphasize "semi-supervised" learning where human analysts help to focus on features that will be most indicative of change or anomaly while algorithms assess whether incoming documents do or do not deviate "significantly" on those features.

The **Bioterrorism Sensor Location Project** will explore greedy algorithms and dynamic variants of classic facility location and clustering algorithms.

The joint **DIMACS-DIMATIA-RENYI Project on Combinatorial Methods** moves into its second year, beginning with a working group meeting at DIMACS on generalized graph colorings.

The project on **Computer Science and Decision Theory** will focus on uses of voting methods in metasearch and collaborative filtering. It will also focus on finding centrally located items in large databases, in particular biological ones, and on in distributed computing when some processors fail and we seek to reach "agreement" among the remaining processors

We go into more detail concerning plans for homeland security research in the next subsection.

11.3.2 Homeland Security Research

DIMACS' numerous homeland security research initiatives will continue, and in particular we expect significant continuing activity by the working groups on Adverse Events/Disease Reporting, Surveillance and Analysis, Health Data Privacy, Mathematical Sciences Methods for the Study of the Deliberate Release of Biological Agents and their Consequences, and Analogies between Computer Viruses and Immune Systems and Biological Viruses and Immune Systems. New working groups on Modeling Vaccination Strategies, Privacy-preserving Data Mining, Usable Privacy and Security Software, Intrusion Detection and Network Management and on Mobile Code Security will get started.

DIMACS will continue to play a major role in the Rutgers University Homeland Security Research Initiative (RUSHRI), providing leadership to the effort, coordinating Rutgers' research in this area, and helping Rutgers take a leadership role in the State of NJ in this area. Specifically, DIMACS will take the lead in applying to the Department of Homeland Security to become a DHS Research Center. We will also organize a variety of homeland security symposia both inside Rutgers and with an increasing variety of partners around New Jersey, including Princeton University, UMDNJ, Stevens Institute of Technology, NJ Institute of Technology, Fairleigh Dickinson University, and Monmouth University. We will also work with industrial partners who are increasingly interested in working with the university in this field, and continue to work with various federal, state, and local agencies, such as Fort Monmouth, Picatinny Arsenal, the NJ Technology Council for Homeland Security, the Counter-terrorism Task Force in the NJ Attorney General's office, the Defense Threat Reduction Agency, and the intelligence community.

11.4 Integration of Research and Education

11.4.1 Postdoctoral and Graduate Student Programs

Our postdoctoral and graduate student programs will continue as in the past. Thanks for Rutgers' support, we will have a joint DIMACS-Institute for Advance Study postdoctoral fellow. We will also have postdoctoral fellows in computational and mathematical epidemiology and in biomathematics. We will be seeking to appoint a postdoctoral fellow to work on our Monitoring Message Streams project, funding permitting. Several graduate students will be supported as RA's under that project and a student will also be supported as an RA under our Computational and Mathematical Epidemiology special focus. We will seek additional funding for postdocs and students in various other ways, including continuing to seek to share them with our industrial and international partners.

11.4.2 Research Experiences for Undergraduates Program

We will again partner with the Mathematics Department and run a joint program for about 25 students. The international component of the program will continue with students from our partner DIMATIA in the Czech Republic participating and five US students going to DIMATIA for three weeks.

11.4.3 The Reconnect Program

Our Reconnect Program that aims at "reconnecting" faculty at 2- and 4-year colleges to the mathematical sciences research enterprise continues with two summer 2003 conferences and one June 2004 conference.

FY2004 Conferences

- **Centrality in Graphs with Applications to the Theory of Location of Facilities**
  - Dates: July 16-23, 2003
  - Location: Illinois Institute of Technology
  - Organizers: Fred R. McMorris (Illinois Institute of Technology), Fred Roberts (Rutgers)
- **Internet Algorithms: Toward Scalable Resource Sharing Protocols**
  - Dates: August 10-16, 2003
  - Location: DIMACS Center, Rutgers
  - Organizers: Rochelle Leibowitz (Wheaton College), Fred Roberts (Rutgers)
- **Experimental Algorithmics, with a Focus on Branch and Bound for Discrete Optimization Problems**
  - Dates: June 20-26, 2004
  - Location: Lafayette College
  - Organizers: Jon Berry (Lafayette), Fred Roberts (Rutgers)
  - Attendance: 20

Another Reconnect Conference will be held at DIMACS in Summer 2004 and Satellite Reconnect Conferences are already scheduled for St. Mary's College (California) in Summer 2004 and for Montclair State University and Spelman College in Summer 2005. We have begun discussions with Morgan State University to run a Satellite Reconnect Conference in Summer 2006.

11.4.4 DIMACS Connect Institute

DCI is our program that includes and education program for high school teachers, an international research workshop, partnerships between researchers and teachers, a research experience for the teachers, and a follow-up high school student research conference. The last summer of DCI under

NSF funding and its current model is summer 2003. The research theme for that summer's program: Combinatorial Designs.

Workshops and Conferences

- **DCI Research Workshop on Combinatorial Designs**
  - Dates: July 13-18, 2003
  - Location: DIMACS, Rutgers University
  - Organizers: Robert Jamison (Clemson), Eric Mendelsohn (U. Toronto)
- **DCI High School Student Research Conference**
  - Dates: May 2004
  - Location: DIMACS, Rutgers University
  - Organizers: Linda Lesniak (Drew University), Fred Roberts (Rutgers University)

For summer 2004, we will be changing the theme of DCI as we look to the future. We will be moving from our topics in graph theory and combinatorics to topics at the interface between the mathematical and biological sciences. See Section 11.4.6 for a discussion of a major new initiative along these lines.

11.4.5 K-12 Programs

The Leadership Program in Discrete Mathematics (K-8 teachers) continues with summer programs at satellite locations. Summer 2003 programs were planned at Auburn (twice), Tufts, and Valparaiso University in Indiana.

The Young Scholars Program for high school students also continues with tuition payments by students supplementing the foundation support we have received, which helps to give scholarships for the students.

The "Crash Course in Discrete Mathematics" continues with 3-day summer programs for high school teachers.

At the K-12 level, a major new initiative involves the integration of mathematical and biological sciences. More on this in the next section.

11.4.6 Integrating the Mathematical and Biological Sciences in K-12

DIMACS proposes to explore ideas for integrating the mathematical and biological sciences in grades K-12. We plan to hold a conference to explore methods of doing this at the K-12 level, bringing together those who have tried it, those who have made it work on the undergraduate level, and those who know how to get new programs into the schools. We will also try out some of our ideas in an exploratory two-week summer program for high school mathematics teachers that will introduce them to the topics at the interface of mathematics and biology and give them the opportunity to develop curriculum materials.

Modern biology has changed dramatically in the past two decades. Driven by large scientific endeavors such as the human genome project, it has become very much an information science, closely tied to tools and methods of the mathematical sciences. New algorithms and mathematical models played a central role in sequencing the human genome and continue to play a crucial role as biology develops models of information processing in biological organisms. Increasingly, undergraduate and graduate students are being exposed to this interplay between the mathematical and biological sciences. In K-12, the biology curriculum has made some advances by including such things as genetics and the human genome project, and even some of the mathematics in the Mendelian genetics model. There are also a few isolated efforts to bring biological examples into the mathematics classroom. But K-12 is lagging behind. Current efforts need to be supported and new efforts developed to bring K-12 education up to speed in the integration of mathematics and biology. Students need to be exposed to the excitement of modern biology from both the biological and mathematical point of view. They need to be informed of the new educational and career opportunities that are arising from the interface between these disciplines. Introducing K-12 students to the interface between the biological and mathematical sciences will not only enhance the study of biology, but also the study of mathematics. Students interested in studying biology will realize the importance of understanding modern mathematics. New horizons will be opened for those who might find mathematics interesting, but wonder how it might be useful. There is the potential for all students to study mathematics both longer and more seriously because they are aware early of its importance in applications such as protecting us from bioterrorism, responses to public health crises, and understanding modern diseases.

We plan to take a two-stage approach to explore these ideas. Stage one is a conference in Spring 2004 to explore new curriculum ideas for making the interface between biology and mathematics accessible to K-12 students. We would bring together the following people:

- those who have already explored bringing the mathematical sciences into the K-12 biology classroom or modern biology into the K-12 mathematics classroom;
- those who have integrated the mathematical sciences and biology successfully at the undergraduate level;
- and teachers and administrators who are knowledgeable about the opportunities for and obstacles against introducing the biology/mathematics interface at the K-12 level.

Stage two is an exploratory two-week program in Summer 2004 featuring high school mathematics teachers. The teachers will get an introduction to molecular biology, computational biology, and bioinformatics. They will learn about sequence alignment algorithms, finding the smallest number of mutations of a certain type to switch one sequence into another, algorithms for finding a sequence from its fragments, and other mathematical techniques. They will also learn how mathematical modeling can be applied to the problems of stopping the spread of infectious diseases and defense against bioterrorists attacks. The teachers in the program will produce classroom materials for use in their schools, under the guidance of experienced high school lead teachers and experts in the undergraduate biomath interface, and inform us and each other of school-year classroom implementation experiments. We hope to learn something about which ideas will work and which ideas will not, and use the experience to design future programs that emphasize instructional materials development, teacher preparation and enhancement, as well as research projects on teaching and learning at this important interface.

In the future, we envision experimenting further with the summer program in several ways, depending upon both the outcomes of the conference and the initial exploratory summer program. For example, we hope to be able to experiment with a similar program for biology teachers, and also one in which we partner biology and mathematics teachers, ideally from the same school. The teachers would have the shared experience of participating in the program, learning new material, and solving problems as a team. This would develop relationships that would enable them to continue working together in bringing their new materials and methods to their schools. K-8 programs would come later. We also envision developing satellite activities at partner institutions. Faculty teams from mathematics, math education, computer science, and biology at these satellite locations would run programs modeled on the DIMACS program.

The specific outcome we expect from the conference is a report with new curriculum ideas for making the interface between biology and mathematics accessible to K-12 students. This report will be distributed widely using the extensive DIMACS mailing lists of teachers and supervisors, as well as other mailing lists. The ideas generated by this conference will also inform both the exploratory summer program in stage two and future summer programs. The classroom materials produced by the teachers in the summer program will be experimented with in the classroom and we will then share the materials and the results of the experiments widely.

DIMACS is extremely well positioned to implement this program, which is a natural extension of DIMACS' success in integrating research with education. DIMACS has also been actively involved in the interface between biology and mathematics for over a decade. This includes Special Foci on Mathematical Support for Molecular Biology, DNA Computing, Computational Molecular Biology, and Computational and Mathematical Epidemiology.

DIMACS' programs at the interface between research and education have made us a pioneer in the fields of computational molecular biology, bioinformatics, and computational/mathematical epidemiology. These programs have included tutorials and short courses at the undergraduate, graduate, and postdoctoral levels, including considerable emphasis on the biology/math interface in our "Reconnect" program for 2- and 4-year college faculty and in our "educational modules series." We have experimented with topics in computational biology and computational/mathematical epidemiology in our precollege student and teacher programs and the enthusiastic student and teacher response to these experiments has been one of the stimuli for our developing plans for this new program.

The steering committee for this project includes high school biology and mathematics teachers as well as experts in running K-12 programs in both biology and mathematics.