DIMACS Center
Rutgers University

Special Focus on Computational and Mathematical Epidemiology

Annual Report

May 2005

Participants who spent 160 hours or more:

Fred S. Roberts, Rutgers University, Principal Investigator; involved in all the working groups and all aspects of the project.

Martin Farach-Colton, Rutgers University, Member of Organizing Committee; has been playing a role in the connections of the epidemiology work to issues of bioinformatics/genetics.

David Madigan, Rutgers University, Member of Organizing Committee; Co-Organizer of Working Group on Adverse Event/Disease Reporting, Surveillance and Analysis.

S. Muthukrishnan, Rutgers University, Member of Organizing Committee

Visitors as part of the Special Focus:

> James Abello, Working on a cancer registry modeling project with Dr. Ilya Muchnik and graduate student David Millman and on graph-theoretical models of spread of disease with Michael Capalbo
>
> Michael Capalbo, Working on random graph models of spread of disease with James Abello.

Postdocs as part of the Special Focus:

> Graham Cormode, Working on data mining and epidemiology, primarily with Dr. Muthukrishnan
>
> Patrick DeLeenheer, Working on predator-prey models and chemostat models.
>
> Kah Loon Ng, working on vaccination strategies
>
> Tanya Berger-Wolf, working on kinship relationships in populations
>
> Martin Pal, working on combinatorial optimization and epidemiology

Other Participants:

Sunetra Gupta, Oxford University, Member of Organizing Committee

Donald Hoover, Rutgers University, Member of Organizing Committee; Co-Organizer of Working Group on Adverse Event/Disease Reporting, Surveillance and Analysis.

David Krakauer, Santa Fe Institute, Member of Organizing Committee

Mel Janowitz, DIMACS, Co-Organizer, Workshop and Working Group: Reticulated Evolution; Co-Organizer, Tutorial and Working Group Meeting: Order-theoretic Aspects of Epidemiology

Brenda Latka, DIMACS, Co-Organizer, Working Group: Data De-Identification, Combinatorial Optimization, Graph Theory, and the Stat/OR Interface

Simon Levin, Princeton University, Member of Organizing Committee; Co-Organizer of the DIMACS Epidemiology Minisymposium; Leader of influenza modeling effort; Advisor to PI on all parts of the project.

Marc Lipsitch, Harvard School of Public Health; Member of Organizing Committee

Ilya Muchnik, developing a cancer registry modeling project with Dr. James Abello and graduate student David Millman.

Megan Murray, Harvard School of Public Health; Member of Organizing Committee

David Ozonoff, Boston University; Member of Organizing Committee; Co-Organizer, Tutorial and Working Group Meeting: Order-theoretic Aspects of Epidemiology

Burton Singer, Princeton University, Member of Organizing Committee

Daniel Wartenberg, University of Medicine and Dentistry of New Jersey, Member of Organizing Committee

Working Group Meeting: Phylogenetic Trees and Rapidly Evolving Diseases
September 7 - 8, 2004

Organizers:
Allen Rodrigo, University Of Auckland
Mike Steel, University of Canterbury

Workshop: Reticulated Evolution
September 20 - 21, 2004

Organizers:
Mel Janowitz, DIMACS
Bernard Moret, University of New Mexico
Randy Linder, University of Texas

Working Group Meeting: Reticulated Evolution
September 22, 2004

Organizers:
Mel Janowitz, DIMACS
Bernard Moret, University of New Mexico
Randy Linder, University of Texas

Working Group Meeting: Data De-Identification, Combinatorial Optimization, Graph Theory, and the Stat/OR Interface
November 9 - 10, 2004

Organizers:
Larry Cox, Centers for Disease Control and Prevention

Brenda Latka, DIMACS
Fred Roberts, DIMACS

Working Group Meeting: Adverse Event/Disease Reporting, Surveillance and Analysis
Third Meeting, March 2, 2005

Organizers:
- Donald Hoover, Rutgers University
- David Madigan, Rutgers University
- Henry Rolka, Centers for Disease Control and Prevention

Tutorial and Working Group Meeting: Order-theoretic Aspects of Epidemiology
First Meeting, March 7 - 9, 2005

Organizers:
- David Ozonoff, Boston University
- Melvin Janowitz, Rutgers University
- Fred Roberts, Rutgers University

DIMACS Epidemiology Minisymposium
April 25, 2005

Organizers:
- Simon Levin, Princeton University
- Fred Roberts, DIMACS, Rutgers University

Workshop: Evolutionary Considerations in Vaccine Use
June 27 - 29, 2005

Organizers:
- Troy Day, Queen's University
- Alison Galvani, Yale University
- Abba Gumel, University of Manitoba
- Claudio Struchiner, Oswaldo Cruz Foundation

Graduate students who have undertaken small research projects under support of the project:

German Enciso, RU Mathematics, Winter 04/05
“Systems under positive feedback: Multistability and a reduction theorem”

Pai-Hsi Huang, RU CS, Winter 04/05
“Profile hidden Markov Chains”

David Millman, RU
Working on a cancer registry modeling project with Dr. James Abello and Dr. Ilya Muchnik.

Paul Raff, RU Mathematics, Winter 04/05

> "Work with and extend the work done by Donald Burke and others in their paper 'Toward a Containment Strategy for Smallpox Bioterror: An Individual-Based Computational Approach' and their corresponding computer program."
>
> Liming Wang, RU Mathematics, Winter 04/05
> "Cyclin B-Cdc2/Wee1 system, which underlies the cell cycle, functions as a bistable switch"
>
> Igor Zverovich, RU RUTCOR, Winter 04/05
> "Generalized stable set problem with application to the defense against bioterrorism"

Other Collaborators

Larry Cox, CDC, Co-Organizer, Working Group: Data De-Identification, Combinatorial Optimization, Graph Theory, and the Stat/OR Interface

Troy Day, Queen's U., Co-Organizer, Workshop: Evolutionary Considerations in Vaccine Use

Alison Galvani, Yale U., Co-Organizer, Workshop: Evolutionary Considerations in Vaccine Use

Abba Gumel, U. of Manitoba, Co-Organizer, Workshop: Evolutionary Considerations in Vaccine Use

Randy Linder, University of Texas, Co-Organizer, Workshop and Working Group: Reticulated Evolution

Bernard Moret, University of New Mexico, Co-Organizer, Workshop and Working Group: Reticulated Evolution

Allen Rodrigo, University Of Auckland, Co-Organizer, Working Group Meeting: Phylogenetic Trees and Rapidly Evolving Diseases

Henry Rolka, CDC, Co-Organizer of Working Group on Adverse Event/Disease Reporting, Surveillance and Analysis.

Mike Steel, University of Canterbury, Co-Organizer, Working Group Meeting: Phylogenetic Trees and Rapidly Evolving Diseases

Claudio Struchiner, Oswaldo Cruz Foundation, Co-Organizer, Workshop: Evolutionary Considerations in Vaccine Use

Partner Organizations:

Princeton University: Collaborative Research; Personnel Exchanges. Partner organization of DIMACS. Individuals from the organization participated in the program planning, organization, and the research.

AT&T Labs - Research: Collaborative Research, Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Telcordia Technologies: Collaborative Research; Personnel Exchanges. Partner organization of DIMACS. Individuals from the organization participated in the program planning.

NEC Laboratories America: Collaborative Research; Personnel Exchanges. Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Lucent Technologies, Bell Labs: Collaborative Research; Personnel Exchanges. Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Rutgers, The State University of New Jersey: Collaborative Research

Office of Naval Research: Financial Support

Alfred P. Sloan Foundation: Financial Support

Burroughs-Wellcome Fund: Financial Support

New Jersey Commission on Science and Technology: Financial Support

IBM Research: Collaborative Research; Personnel Exchanges. Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Microsoft Research: Collaborative Research. Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Avaya Labs: Collaborative Research. Partner organization of DIMACS. Individuals from the organization participated in the program planning.

American Statistical Association: Financial Support

Hewlett-Packard Labs: Collaborative Research. Partner organization of DIMACS. Individuals from the organization participated in the program planning.

Centers for Disease Control: Collaborative Research; Personnel Exchanges. Individuals from the organization participated in the program planning and working group/workshop organization.

Activities and Findings

Overview

This is to be a five-year special focus, starting in summer of 2002, following the design of our Center's pioneering special focus on Mathematical Support for Molecular Biology. In this special focus, the center seeks to:

- Develop and strengthen collaborations and partnerships between mathematical scientists (mathematicians, computer scientists, operations researchers, statisticians) and biological scientists (biologists, epidemiologists, clinicians).

- Identify and explore issues in mathematics and computer science that need to be resolved to make progress on important problems in epidemiology.

- Identify and explore methods of mathematical science not yet widely used in studying problems of epidemiology and introduce epidemiologists to them - with an emphasis on methods of discrete mathematics (including discrete probability) and the algorithms, models, and concepts developed in the field of theoretical computer science.

- Introduce outstanding young people from both the mathematical/computer science and biological communities to the issues and problems and challenges of computational and mathematical epidemiology.

- Involve biological and mathematical scientists together to define the agenda and develop the tools of computational and mathematical epidemiology.

The special focus consists of a research program featuring "working groups" concentrating on specific research topics and a program integrating research and education through a series of workshops and tutorials. The 1994-2000 DIMACS Special Focus on Mathematical Support for Molecular Biology played a central role in laying the groundwork for the field of computational molecular biology, led many fledgling concepts and methods grounded in the mathematical sciences to become standard tools in the biological sciences, produced lasting partnerships between biological and mathematical scientists, and introduced many of today's leaders in computational biology to the field and to each other. We are confident that this special focus will do the same.

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 (transmissable 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.

Workshops, Working Group Meetings and Tutorials During This Reporting Period

Working Group Meeting: Phylogenetic Trees and Rapidly Evolving Diseases - First Meeting: DIMACS Symposium on Phylogenetics and Rapidly Evolving Pathogens

Dates: September 7 - 8, 2004
Location: Aotea Centre, Auckland, New Zealand
Organizers: Allen Rodrigo, University Of Auckland and Mike Steel, University of Canterbury
Attendance: 9

This working group builds on phylogenetic methods developed by computational biologists to explore ways in which such methods can be applied and developed to shed new light on the origin, evolution, and likely future development of viruses and other pathogens. Phylogeny is now a central tool for studies into the origin and diversity of viruses such as HIV and dengue fever virus. These and other investigations have provided new insights, such as identifying the possible pattern of transfer of HIV-type viruses between primate species. Phylogenetic techniques have also proved useful in mapping the evolution of different strains of the human influenza A virus, with the goal of predicting which strain is most likely to cause future epidemics, with applications to vaccine development. Many of the phylogenetic techniques in use were originally developed to investigate more traditional and well-behaved evolutionary problems, where historical relationships are typically represented by a binary tree with a small number of species appearing as the leaves (tip vertices). In epidemiology the picture is more complex and this observation underlies the task of this working group. Even if there is a single underlying tree, it may typically have thousands of vertices, and many of these may be of high degree. Furthermore, data may be available not just for the species at the leaves of the tree, but for species distributed at vertices throughout the tree, particularly when the evolution of a virus is studied by serial sampling in patients. This is true for retroviruses which have a very high substitution rate, and whose molecular evolution may be up to 106 times more rapid than eukaryotic or prokaryotic genes. New methods for dealing with these complications will be investigated. To complicate the picture further, it may well be more appropriate to represent the evolution of a virus by a collection of trees, or by a digraph (or network) to recognize the "quasi-species" nature of viruses, such as in the application of split decomposition by Dopazo, Dress, and von Haeseler; we shall pursue this direction of research. Relating population genetics considerations (currently handled by the "coalescent" model) to phylogeny considerations is also potentially useful. However, even here, theory has yet to be developed. For instance, the fact that retroviral evolution occurs within a host means that viral sequences sampled from different hosts must take account of the different dynamics of between-host transmission histories and within-host viral genealogies. This has consequences for the inference of epidemiological parameters based on viral sequences obtained from several hosts, and we will investigate them. Finally, if one wishes to test particular epidemiological hypotheses it would be helpful to have techniques that avoid having to fix attention on one particular tree. This suggests devising fast methods that would average the quantities of interest over all likely trees, weighted by how well they describe the data - a challenge for modern computational tools and our working group.

The first meeting of the working group was held at the INCOB 2004 (*International Conference on Bioinformatics*) in Auckland, New Zealand, 2004. All nine members of the working group attended, and presented talks during the *DIMACS symposium on Phylogenetic trees and rapidly evolving pathogens*, which also included some other associated researchers. The working group also met formally in a closed session during the conference.

Much discussion revolved around the right biological requirements for models. Allen Rodrigo suggested the formulation of better synthetic models (to handle several situations – recombination, serial sampling, migration, selection etc) and the question of how much data would be needed to discriminate these effects. Rod Page was interested in the modeling of gene duplication and horizontal gene transfer in molecular epidemiology. Bruce Rannala saw a need for applying empirical estimators to handle more realistic models that could not be analytically handled. Geoff Chase saw a need for better techniques for testing between models, while Thomas Leitner said it be worthwhile developing techniques to handle transmission, bottleneck effects, recombination and site saturation.

Workshop: Reticulated Evolution (Co-sponsored by the DIMACS Special Focus on Computational Molecular Biology).

Dates: September 20 - 21, 2004
Location: DIMACS Center, CoRE Building, Rutgers University
Organizer: Mel Janowitz, DIMACS; Bernard Moret, University of New Mexico; Randy Linder, University of Texas
Attendance: 52

Species evolution has long been modeled as a branching process that can uniquely be represented by a tree topology. In such a topology, each species can only be linked to its closest ancestor, while interspecies relationships such as species hybridization or lateral gene transfer in bacteria are not allowed. With the advent of phylogenetic analysis at the molecular level, there is increasing evidence that such a model is inadequate. This workshop explored the history and latest status of these new models of "reticulate evolution", and was coupled with a smaller working group meeting designed to explore promising avenues for future research.

Working Group Meeting: Reticulated Evolution (Co-sponsored by the DIMACS Special Focus on Computational Molecular Biology)

Dates: September 22, 2004
Location: DIMACS Center, CoRE Building, Rutgers University
Organizer: Mel Janowitz, DIMACS; Bernard Moret, University of New Mexico; Randy Linder, University of Texas
Attendance: 21

This working group meeting was coupled with a workshop on the same subject. Its goal was to initiate promising avenues of research designed to explore new models of "reticulate evolution" that are biologically meaningful, and computationally feasible.

Working Group Meeting: Data De-Identification, Combinatorial Optimization, Graph Theory, and the Stat/OR Interface (Co-sponsored by the Special Focus on Communication Security and Information Privacy)

Date: November 9 - 10, 2004
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Larry Cox, CDC; Brenda Latka, DIMACS; Fred Roberts, DIMACS
Attendance: 20

In this meeting we explored problems in combinatorial optimization, graph theory, and the interface between statistics and operations research that arise from issues of data privacy and, more specifically, data de-identification.

This was an informal meeting, aimed at involving those with interests in combinatorial optimization, graph theory and the stat/OR interface in working on these problems that have become very important in applications such as health data privacy, government statistical data, and counter-terrorism. The emphasis was on identifying and working on problems of discrete optimization and on identifying and exploring relevant algorithms. No prior knowledge of the application areas was necessary.

Specific problems of interest discussed/examined from the OR perspective include combinatorial structure of the feasible region defined by a partially specified multi-dimensional table or by linked tables; generating extremal points and statistical samples from a feasible region defined by a system of multi-dimensional tabular constraints; and (near)-optimization of (nonlinear) statistical functions over a system of tabular constraints. These problems recently have been approached from the standpoint of the theory of Grobner bases but the intended focus here is on combinatorial and mathematical programming approaches and their computability.

Working Group Meeting: Adverse Event/Disease Reporting, Surveillance and Analysis

Date: Third Meeting, March 2, 2005
Location: Bethesda, Maryland
Organizers: Donald Hoover, and David Madigan, Rutgers University; Henry Rolka, (CDC)
Attendance: 6

A subgroup of participants from the Working Group moderated the session and provided presentations: Howard Burkom (John Hopkins University Applied Physics Laboratory), Gregory F. Cooper (University of Pittsburgh), Martin Kulldorff (Harvard Medical School), David Madigan (Rutgers University) and Henry Rolka (Centers for Disease Control and Prevention).

A variety of analytic approaches have arisen and are in use for performing Bioterrorism (BT) surveillance using social and other public health indicators from various types of data (e.g., pre-diagnostic/chief complaint ambulatory care encounters, nurse call line data, over-the-counter sales, absenteeism, Emergency Department discharge summaries, prescription pharmaceutical sales, 911-emergency calls; etc.). The value of data anomaly investigation and signal detection as technologies in surveillance can be enhanced by more formalized application of data pre-processing methods and applied probabilistic decision science concepts and principles. A full characterization of the usefulness as well as the corresponding development of analytic methods for exploiting opportunistic data are rich areas for research, especially in the context of information system integration. The specific focus of this meeting was on the analytic surveillance component where multiple sources of data are utilized to assess the health-related temporal and geographic status of human health risk.

The purpose of the meeting was to provide (1) an overview of the problem, (2) a view of current practices in operation (2 presentations), (3) evolving application areas of promise (1 presentation) and (4) to elicit feedback and discussion. Ideas and conceptual threads from the discussion were explored and developed into a one-to-five year research agenda for addressing this area.

Tutorial and Working Group Meeting: Order-theoretic Aspects of Epidemiology

Dates: First Meeting, March 7 - 9, 2005
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: David Ozonoff, Boston University; Melvin Janowitz and Fred Roberts, Rutgers University
Attendance: 14

Many practical epidemiological problems involve the comparison of one or more quantities. Most often the quantities are rates or proportions leading to a measure of effect or association, but they may also involve distances, exposure categories, job titles, etc. Often the actual values in question are not important, only whether one value is smaller than or larger than a second, i.e., their order. This working group studied how fundamental order-theoretic concepts of theoretical computer science (TCS) and discrete mathematics (DM) such as semiorders, interval orders, general partial orders, and lattices can be used to improve the results of epidemiological investigations. The participants gave epidemiological concepts a careful definition in the language of partial orders and explored the use of visualization of order-theoretic concepts in epidemiologic studies. The latter involve issues such as how best to visualize a poset through clever presentation of its Hasse diagram - an issue of great interest in the field of TCS known as graph drawing. One application of these ideas arises in the problem of determining cutoffs or boundaries so as to determine exposure categories in epidemiology. This can be modeled by finding n attributes (age, inverse of distance to a pollution source, etc.); for each subject x, finding a number $f_i(x)$ representing a measure of the ith attribute; and saying that x has a higher exposure than y if $f_i(x) > f_i(y)$ for all i. This defines a partial order that is well studied in dimension theory. Finding the *exposure set*, the set of all subjects x whose exposure levels $f_i(x)$ all exceed some threshold, is a common construction in dimension theory. An interesting variant is when only a given percentage of these levels need to exceed threshold. This group sought algorithms for fitting this model to large data sets when the partial order and the exposure set are given, but the attributes and the number of them is not. One promising approach is to use algorithms developed by the DIMACS working group on multidimensional scaling. As another example, point lattices may be regarded as a type of order theoretic lattice. The point lattice construction has found uses in epidemiology through visualizing the relationships of all possible contingency tables to various statistics, effect measures, and cut-off choices and has also been used in statistics. Challenges to this group in extending these ideas include generalizing the concepts to higher-dimensional tables, where there are additional attributes (ordinal, numerical or nominal) besides case status and exposure; and applying lattice-theoretic approaches to measurement error. The point lattices formed by 2 x 2 contingency tables can be represented as *n*-element strings from a 2 letter alphabet *{x,y}*. Measurement errors can be thought of as being caused by a transposition from a substring *xy* to a substring *yx*. Similar transpositions have been studied from a more general viewpoint by lattice theorists as a special case of a Newman commutativity lattice. Many references also mention connections with weak Bruhat orders of Coxeter groups. This group examined how higher dimensional contingency tables relate to what are called multinomial lattices and studied how combinatorial aspects of the Bruhat orders relate to probabilistic questions in epidemiology. Among other things, these considerations gave guidance on how to decide when observed data tables can be explained by chance alone. The group also considered the issue of what kinds of statistical tests are legitimately applied to data where only order matters; this issue is somewhat recognized by epidemiologists, but its order-theoretic subtleties are usually not. Mathematical and computational methods dealing with ordered algebraic systems form the foundations of the modern theory of measurement and can be used to analyze this and important related issues such as what conclusions using scales of measurement are "meaningful." Measurement theory (a term which has a different connotation in epidemiology) does not seem to be known to practicing epidemiologists and this group remedied that, keeping recent applications to software measurement in mind. The group plans to analyze epidemiological studies from a measurement theory point of view. This working group was our most speculative. It built upon a large literature in TCS dealing with order relations, computing them, approximating them, visualizing them, and assigning measures to them, but it did not build upon a large body of work connecting these ideas to epidemiology, mostly upon the view of several active epidemiologists that these ideas are relevant.

The meeting featured a "superquick" tutorial on epidemiology for mathematicians by David Ozonoff and a 1.5 day tutorial on the theory of measurement and its application by Fred Roberts. The latter led to ideas for several multidisciplinary, multi-institutional follow-up research projects. Among other topics discussed were concept lattices in epidemiology, formal concept analysis, and management of quantified semantic taxonomies for biothreat analysis.

DIMACS Epidemiology Mini-symposium
Dates: April 25, 2005
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Simon Levin, Princeton University and Fred Roberts, DIMACS, Rutgers University
Attendance: 18

This one-day mini-symposium marked the midpoint of the Special. We used the opportunity to explore several exciting topics at the interface between mathematics/computer science on the one side and epidemiology/biology on the other side. The keynote talk, given by Don Burke (Johns Hopkins), explored a current research topic ripe for interdisciplinary collaboration: epidemic viruses as computational entities. Four other speakers explored a range of topics, with some emphasis on work that has come out of the special focus: Stephen Hartke (University of Illinois) on the "firefighter problem" (part of his thesis research under Fred Roberts at Rutgers), Jonathan Dushoff (Princeton) on modeling the influenza A virus, Tanya Kostova (Lawrence Livermore National Labs) on modeling homogeneous mixing groups, and James Abello (DIMACS) on visualization experiences with SEER cancer data . We made the program accessible to a mix of mathematicians and computer scientists, biologists, public health people, and epidemiologists and one of our goals was to find ways to enhance the mix of the two types of audiences. We also allowed time for discussion, an overview of upcoming events in the special focus, and ideas for additional future events.

Workshop: Evolutionary Considerations in Vaccine Use
Dates: June 27 - 29, 2005
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Troy Day, Queen's U.; Alison Galvani, Yale U.; Abba Gumel, U. of Manitoba;
Claudio Struchiner, Oswaldo Cruz Foundation
Attendance: 13 (Registration still open for this workshop)

There is a clear need for the development of a predictive framework, based on mathematical modeling and computer simulations, that can be used to help design optimal vaccination strategies. This was the primary objective of the working group organized by John Glasser and Herbert Hethcote that met at DIMACS in May, 2004. One aspect of vaccine use that does not often receive much attention, however, is the evolutionary consequences of these vaccines. For example, what effects might vaccine use have on the evolutionary dynamics of pathogen populations, and how might these evolutionary changes affect the ability of the vaccine to control a certain disease? Additionally, do different vaccination strategies result in different evolutionary outcomes? Given the extensive genetic variability in many pathogens (such as

HIV, influenza A H2N2, malaria and some vaccine-preventable diseases like polio, MMR, Chickenpox, yellow fever, tetanus, pneumococcal disease etc.), evolutionary change in response to vaccination is potentially significant. Our main objective is to examine general evolution-related questions for any disease for which there is a vaccine (or hope for one).

The workshop will focus on the following five main themes:

(i) Modes of Vaccine Action.

(ii) Multiple Levels of Natural Selection.

(iii) Conflicts Between Epidemiology and Evolution.

(iv) Vaccination & Virulence.

(v) Mechanisms of Vaccine Delivery.

The workshop will bring together scientists from diverse backgrounds (mathematicians, epidemiologists, virologists, immunologists, vaccine developers etc.) in order to address the questions raised within the aforementioned themes.

DIMACS Seminar in Quantitative Biology and Epidemiology

Michael Capalbo and James Abello, Special Focus visitors, organized the DIMACS Computational amd Mathematical Epidemiology seminar during the 2004-2005 academic year. In total, there were almost twenty talks given. The audience came from a wide variety of disciplines, including Biology, Mathematics, Medicine, Physics, Chemistry, Computer Science and more, including regular participants from Rutgers, The University of Medicine and Dentistry of New Jersey (UMDNJ), Robert-Wood Johnson Medical School (RWJMS), Princeton and beyond. The program of this series, including titles and speakers is given here. Abstract information can be found at:

http://dimacs.rutgers.edu/SpecialYears/2002_Epid/episeminars.html

Graph-Theoretical Models of the Spread and Control of Disease
Monday, October 11, 2004
Speaker: Fred Roberts, DIMACS

A new clustering coefficient
Monday, October 25, 2004
Speaker: Sara Soffer, Rutgers University

An example of a disease outbreak simulation in an urban environment: TRANSIMS and EpiSims
Monday, November 1, 2004
Speaker: Tanya Berger-Wolf, DIMACS Post Doc

Surveillance Mistakes
Monday, November 8, 2004
Speaker: William E. Halperin, Department of Preventive Medicine and Community Health
University of Medicine and Dentistry

Transmission Probabilities in Disease Spread in Random Power Law Graphs
Monday, November 15, 2004
Speaker: Michael Capalbo, DIMACS

Discussion of the paper "Maximizing the Spread of Influence through a Social Network", by David Kempe, Jon Kleinberg and Eva Tardos
Monday, November 29, 2004
Speaker: Martin Pal, DIMACS

Influences on Breast Cancer Survival via SVM Classification in the SEER Database
Monday, December 6, 2004
Speaker: Ilya Muchnik, DIMACS

Generalized Contingency Tables, Concept Lattices and Graph Partitions
Monday, December 13, 2004
Speaker: Alex Pogel, University of New Mexico

The Impact of Non-Poisson Contact Processes on Virus Spreading
Monday, January 24, 2005
Speaker: Alexei Vazquez, University of Notre Dame (Host: Sara Soffer)

Transmission Probabilities and Disease Spread
Monday, January 31, 2005
Speaker: Michael Capalbo, DIMACS

Modeling Disease Transmission Across Social Networks
Monday, February 7, 2005
Speaker: Stephen Eubank, Virgina Tech

Unbalanced Graph Cuts
Monday, February 21, 2005
Speaker: Martin Pal, DIMACS

What's New in Surveillance?
Monday, February 28, 2005
Speaker: Dona Schneider, Rutgers University

A Survey of Some Results on the Firefighter Problem
Monday, March 21, 2005
Speaker: Kah Loon Ng, DIMACS

Persistent Social Groups
Monday, March 28, 2005
Speaker: Tanya Berger-Wolf, DIMACS

SVM in Analysis of Cross-Sectional Data
Monday, April 4, 2005
Speaker: Dmitriy Fradkin, DIMACS

Gadgets for good: How Computer Researchers can Help Save Lives in Poor Countries
Monday, April 11, 2005

Speaker: Neal Lesh, Harvard University

Choosing a Random Node in a Large-scale Network
Monday, April 18, 2005
Speaker: Jared Saia, University of New Mexico

Computational Aspects of Representing and Analyzing Information, Social & Infrastructure Networks
Monday, May 2, 2005
Speaker: Madhav V. Marathe, Virgina Tech

Exact Methods Applied to Group Sequential and Other Stratified Comparative Poisson Designs
Monday, May 23, 2005
Speaker: Qi Xia, Rutgers University

Findings

Seasonality of Influenza

The underlying cause of seasonal oscillations in influenza incidence remains unclear despite at least 80 years of investigation. These oscillations are presumably due to some mechanism that causes seasonal changes in the effective transmission rate of the virus from person to person (often called seasonal forcing). But laboratory experiments and epidemiological studies have failed to establish whether these transmission changes are due to direct effects of temperature and humidity on transmission, to changes in mixing patterns (e.g., school terms or simply more time spent indoors), or to other factors, such as increased viral production under winter conditions (evidenced by higher febrile reaction to identical influenza exposures). Jonathan Dushoff (Princeton), Joshua B. Plotkin (Harvard), Simon A. Levin (Princeton) and David J. D. Earn (McMaster) collaborated on this problem. For influenza, immunity is not permanent. After recovering from one antigenic variant of the virus, a person is usually at least partially susceptible to new variants within a few years, due to gradual evolution of the virus. They modeled this process of antigenic drift by assuming that the people in their model lose their resistance to the circulating virus and hence the move back to the susceptible class after a few years. Whereas deterministic models treat the number of people in each state as a continuous variable, in reality, individuals are discrete, and the process is more appropriately modeled as a stochastic Markov chain. The resulting demographic stochasticity causes fluctuations in disease incidence to persist. Using these refinements, Dushoff, Plotkin, Levin, and Earn have shown that, in fact, it may be impossible to establish the underlying cause of seasonality in influenza epidemics, because the large observed oscillations in incidence may be caused by undetectably small seasonal changes in the influenza transmission rate that are amplified by the endogenous population dynamics of the host–pathogen system.

Data Mining and Epidemiology

Epidemiology is an observational science that concerns itself with finding and explaining patterns of health and disease in populations, usually of humans, but also populations of animals, insects and plants. Data mining is an active area of research interested in finding algorithms for describing latent patterns in often very large data sets. This DIMACS Working Group on Data Mining and Epidemiology broght together epidemiologists and experts interested in exploring data mining and algorithmic techniques potentially useful in epidemiology and resulted in a DIMACS special volume, edited by James Abello and Graham Cormode, with the objective of documenting and disseminating the results of the collaboration between these two disciplines. The volume includes topics explored by the working group, which emphasized the presentation of specific epidemiological problems and collaborative discussion of ways to use data mining methods, informed by modern tools of computer science, statistics, and mathematics, to gain epidemiological insights. Included papers apply new techniques from theoretical computer science, discrete mathematics, and statistics to epidemiology. Topics include: Regression Methods, Support Vector Machines and Cluster Analysis Disease Dynamics, Epidemiological Studies and the Interpretation of Data, Genetic, Pharmaceutical and Environmental Epidemiology, Elements of Statistical Learning and Models, Association Rules and Lattice/Graph Theoretical Methods. Many of the papers included in the volume give detailed descriptions of specific studies analyzed by methods current in today's epidemiological practice. The studies were selected to represent a range of typical techniques such as exploratory factor analysis, multiple logistic regression, proportional hazards models, generalized additive models and other current methods. Some concentrate on the epidemiological aspects of a study, with explanations of the main statistical ideas. Computer scientists or discrete mathematicians contributed papers that expand upon algorithmic issues and alternate approaches to epidemiological questions. There is a section on open problems from the interface of Epidemiology, Computer Science, Mathematics and Statistics. There are papers that clearly describe fundamental epidemiological problems that can benefit from efficient computational, statistical and mathematical models that can aid in the processing and understanding of combined epidemiological and immunological data. Included are research, expository and survey papers. The list is given below.

Selected Data Mining Concepts, James Abello, Graham Cormode, Dmitriy Fradkin, David Madigan, Ofer Melnik, and Ilya Muchnik

Descriptive Epidemiology: A Brief Introduction, Dona Schneider

*Biostatistical Challenges in Molecular Data Analysis,*William Shannon

Generalized Contingency Tables and Concept Lattices, David Ozonoff, Alex Pogel, and Tim Hannan

*Graph Partitions and Concept Lattices,*James Abello and Alex Pogel

*Influences on Breast Cancer Survival via SVM Classification in the SEER Database,*Jixin Li, Ilya Muchnik, and Dona Schneider

Validation of Epidemiological Models: Chicken Epidemiology in the UK, Dmitriy Fradkin, Ilya Muchnik, Patrick Hermans, and Kenton Morgan

Mining Online Media for Global Disease Outbreak Monitoring, Lynette Hirschman and Laura Damianos

*Random Graphs (and the Spread of Infections in a Social Network),*James Abello and Michael Capalbo

Using Transmission Dynamics Models to Validate Vaccine Efficacy Measures Prior to Conducting HIV Vaccine Efficacy Trials, Kamal Desai, Marie-Claude Boily, Benoıt Masse, and Roy Anderson

*Attempting to Narrow the Integrality Gap for the Firefighter Problem on Trees,*Stephen Hartke

Structure of Social Contact Networks and Their Impact on Epidemics, Stephen Eubank, V. S. Anil Kumar, Madhav Marathe, Aravind Srinivasan, and Nan Wang

Bayesian Modeling for Biosurveillance

In the past year, Gregory Cooper's participation on the DIMACS-organized panel Statistical Issues in Public Health Surveillance for Bioterrorism Using Multiple Data Streams as part of the Adverse Events working group led to useful discussions with the panel participants and organizers about biosurveillance algorithms and new ideas for his project at the University of Pittsburgh. Early, reliable detection of outbreaks of disease, whether natural (e.g., West Nile virus) or bioterrorist-induced (e.g., anthrax and smallpox), is a critical problem today. It is important to detect outbreaks early in order to provide the best possible medical response and treatment, as well as to improve the chances of identifying the source. A primary goal of Cooper's project is to develop new Bayesian models and inference algorithms that then are applied to monitor electronically available healthcare data to achieve early, reliable detection of outbreaks. The scientific challenge of monitoring for outbreaks within an entire population creates major computational challenges in building and applying Bayesian models that are orders of magnitude larger than those developed previously. Cooper's project is applying and extending state-of-the-art probabilistic inference methods to achieve efficient inference. If inference indicates that an outbreak is likely, an alert is raised automatically. Appropriately, however, public health officials are unlikely to blindly trust an outbreak alert, unless there is an explanatory justification. Automated explanation of Bayesian inference is therefore another key project goal. The scientific contributions of Cooper's project involve developing, investigating, and evaluating new modeling and algorithmic techniques that make Bayesian biosurveillance practical for monitoring and diagnosing (in real time) the disease-outbreak status of an entire population. In investigating these issues, Cooper's project is anticipated to make both specific scientific contributions to computer science, statistics, and public health, as well as broader contributions to public safety.

Monitoring Multiple Data Streams to Detect Incidence of Disease

The Working Group on Adverse Event/Disease Reporting, Surveillance, and Analysis has held three substantial meetings since its inception. The first Working Group meeting was held in October 2002 and was attended by about 50 researchers and practitioners. This led to several research collaborations and a spinoff working group. In Summer 2003 they held a week long tutorial on analytic methods for surveillance. This was also attended by about 50 people. In February 2004 they held a meeting focused specifically on methods for multivariate surveillance. The websites for these three events contain extensive details as well as presentation materials. See:

http://dimacs.rutgers.edu/Workshops/AdverseEvent/
http://dimacs.rutgers.edu/Workshops/NewerStat/
http://dimacs.rutgers.edu/Workshops/AdverseEvent2/

One of the outcomes of this working group was a documented group discussion on methods for monitoring multiple data streams to detect incidence of disease. Multiplicity is present in this problem in many forms. We are interested in events, such as release of a toxic agent, which affect the population's health. The variety of possible agents and mode of release mean that an event may be of several types, and it can happen at any time and location. There are also multiple data streams forming the source of information from which we are to deduce whether an event has taken place. When an event occurs, its subsequent effects will depend on various factors that affect dispersal, including population movement and weather conditions (e.g., wind direction). While deterministic models may be available for daily population movement, it is natural to treat weather as a random factor (a more sophisticated approach would be to add observed weather information to the input data). Thus, the question of what one should look for in the data to detect an event is complex. One approach to combining data streams is to monitor individual streams separately, then, when unusual phenomena are seen in one type of data, to look for supporting evidence in others. This raises questions about the nature of the joint effect each possible event type will have on the set of data streams. In assessing the “false alarm rate”, the degree of association between different data streams must also be considered. This working group discussed an alternative approach in which joint modeling of the full set of data streams is undertaken in order to create statistically powerful detectors of event incidence. There are pre-cursors to this in the treatment of multiple endpoints in statistical process control and in the sequential monitoring of clinical trials. In these areas, models are established for the joint behaviors of a number of observed endpoints, including the effect of an “intervention”, i.e., the process becoming out of control in the first example or a successful treatment effect in the latter. It is then a very natural step, and in statistical terms the most efficient option, to define monitoring procedures using the whole set of endpoints together. The challenge in our detection problem is to do something similar while keeping the whole procedure manageable. The working group developed a problem formulation and requirements for any detection algorithm.

A Cost-benefit Analysis of Introducing a Rotavirus Vaccine in Egypt

Rotavirus gastroenteritis is a major cause of mortality and morbidity among children 5 years and younger. In Egypt, approximately 33-44% of all diarrheas among children under 5 years of age are caused by rotavirus. Worldwide, approximately 800,000 deaths are attributable to rotavirus disease each year. The majority of these deaths occur in the developing world. In a study conducted in several developing countries rotavirus was the most frequently detected pathogen in diarrhea episodes during the first year of life. The need for an effective vaccine against the disease has been recognized by the Center for Disease Control and Prevention (CDC), The World Health Organization (WHO), Program for Appropriate Technology in Health (PATH), Pan American Health Organization (PAHO), and The Global Alliance for Vaccines and Immunizations (GAVI). GAVI lists rotavirus vaccine development as one of its top three Priority Research and Development Projects. In light of these developments in rotavirus vaccination, economic evaluation is necessary for countries that are considering using any rotavirus vaccine. In Egypt, approximately 3000 children under five years old die each year from rotavirus infection, while close to 1,709,000 children under five years old become ill with rotavirus each year. The Egyptian Ministry of Health and Population (MOHP) is the main payer of health care. DIMACS visitor Omayra Ortega, University of Iowa, conducted a cost-benefit analysis from the perspective of the MOHP to evaluate the economic impact of introducing a rotavirus vaccine to the current national immunization schedule for the MOHP. Ortega investigated the impact of a rotavirus immunization program in which a rotavirus vaccine, RotaRix, would be administered in 2 doses during the first 6 months of life as a part of a child’s routine immunization. This specific vaccine was chosen because it is the furthest along in vaccine development and the only licensed vaccine in widespread use. The design of the model allows for generalizing to other rotavirus vaccines by changing the number of doses and the efficacy. The impact of an immunization program was measured by the decrease in disease burden and medical costs associated with rotavirus diarrhea over five years. Vaccination would save 2,707 lives, prevent 1,074,799 outpatient’s visits, 413,133 hospitalizations, and cost the MOHP $16,615,942.99 (96,269,475.10 LE) in health expenditures. At the base case level, the Benefit-Cost Ratio is -0.9847:1. The incremental cost-benefit is $17.86 (103.48 LE). Rotavirus vaccine immunization would be beneficial to Egypt both economically and medically.

Evaluating Statistical Methods for Syndromic Surveillance

Special focus participant Michael Stoto (RAND Corporation) collaborated with Ronald D. Fricker, Jr., Arvind Jain, John O. Davies-Cole, Chevelle Glymph, Gebreyesus Kidane, Garret Lum, LaVerne Jones, Kerda Dehan, and Christine Yuan in developing a comprehensive evaluation of statistical methods for syndromic surveillance. They sought (1) to introduce the statistical issues in syndromic surveillance, (2) to describe and illustrate approaches to evaluating syndromic surveillance systems and characterizing their performance, and (3) to evaluate the performance of a couple of specific algorithms through both abstract simulations and simulations based on actual data. Their work, influenced by participation in the special focus, introduced and discussed the statistical concepts and issues in syndromic surveillance, illustrating them with data from an emergency room surveillance system from the District of Columbia. They described methods from the statistical process control literature, including variants on existing multivariate detection algorithms tailored to the syndromic surveillance problem, and compared and contrasted the performance of univariate and multivariate techniques via some abstract simulations. They then compared the new multivariate detection algorithms with commonly used approaches and illustrated the simulation approach to evaluation using simulations based on actual data from seven Washington DC hospital emergency rooms. Their work is a chapter in *Statistical Methods in Counter-Terrorism*, Wilson, A. and D. Olwell (eds.), to appear in the ASA-SIAM series.

Regularization Method for Optimally Switched and Impulsive Systems with Biomedical Applications

As a result meeting in 2002 at a special focus workshop, Erik Verriest (Georgia Tech), Ismael Ortega-Sanchez (CDC), and John Glasser (CDC) began collaborating on the prevention and control of severe and sudden infectious disease epidemics. They consider dynamical systems where the control consists of the choice among a finite number of fixed vector fields. The control is parameterized by the number of switches, m, a ``word" of length m with alphabet A and a sequence of m switching times. They introduced two regularization schemes, replacing the given problem by a smooth control problem, which can be solved by standard numerical optimal control methods. The regularized control is a smooth function, and its quantization to a predetermined set gives the approximation of the switching times and the ``word." Substituting the word in the original problem allows an iterative refinement towards the optimal switching times. Impulsive control determines a timing problem. They considered a fixed affine system where the control takes a singular form. They introduced a regularization scheme, involving a broader class of generalized functions, which allows their use in nonlinear systems. They apply optimal timing problem solutions to the optimal pulse vaccination for the control of epidemics and the optimal scheduling of chemotherapy in cancer. They have submitted two proposals for a Georgia-Tech / CDC collaboration.

A Multi-species Epidemic Model with Spatial Dynamics

Pauline van den Driessche, University of Victoria, and Julien Arino, now at McMaster University, met Richard Jordan**,** Dynamics Technology, Inc. and Mount Holyoke College**,** at a special focus meeting and developed a collaboration that is still ongoing, with the publication of one paper and another one accepted pending revisions. Arino, van den Driessche, and Jordan, together with Jonathan R. Davis Dynamics Technology, Inc.**,** David Hartley, University of Maryland School of Medicine, Joy M. Miller, U.S. Department of Defense formulated a model that describes the spatial propagation of a disease that can be transmitted between multiple species. The spatial component consists, for each species, of a certain number of patches that make up the vertices of a digraph, the arcs of which represent the movement of the various species between the patches. In each of the patches and for each species, a susceptible-exposed-infectious-recovered (SEIR) epidemic model describes the evolution of the disease status of individuals. Also in each patch, there is transmission of the disease from species to species. The group did an analysis of the system, beginning with results on the mobility component. They derived a formula for the computation of the basic reproduction number for s species and n patches, which then determines the global stability properties of the disease free equilibrium. They developed simulations for the spread of a disease in one species and two patches. Arino, Jordan and van den Driessche continue to collaborate.

An Agent-Based Analysis of Smallpox Preparedness and Response Policy

Because conjectural 'thought experiments' can be formalized, refined, and conducted systematically using computers, computational modeling is called for in situations that demand robust quantitative study of phenomena which occur only rarely, or may never occur at all. In light of mounting concerns regarding the threats of bioterrorism in general and smallpox in specific, Benjamin M. Eidelson, Solomon Asch Center for Study of Ethnopolitical Conflict, and Ian Lustick, Political Science Department, University of Pennsylvania and a participant in the working group on Modeling Social Responses to Bio-Terrorism Involving Infectious Agents developed a stochastic agent-based model, VIR-POX, in order to explore the viability of available containment measures as defenses against the spread of this infectious disease. They found the various vaccination and containment programs to be highly interdependent, and ascertained that these policy options vary not only in their mean effects, but also in their subordination to factors of chance or otherwise uncontrollable interference, relationships that themselves fluctuate across ranges of the counterfactual distribution. Broadly speaking, ring vaccination rivaled mass vaccination if a very substantial proportion of smallpox cases could be detected and isolated almost immediately after infection, or if residual herd immunity in the population was relatively high. Pre-attack mass vaccination and post-attack mass vaccination were equivalent in their capacities to eliminate the virus from the population within five months, but the pre-attack strategy did so with significantly fewer deaths in the process. Their results suggest that the debate between ring and mass vaccination approaches may hinge on better understanding residual herd immunity and the feasibility of early detection measures.

Temporal Social Structures

Until recently, questions regarding social structures and social interactions in human or animal populations were answered using statistical tools, summarizing the aggregate information of the big parts of the population. Recently, scientists have started to model the individual social interactions (such as communication, collaborations, sexual contacts, potential disease transmission) as graphs, using either existing detailed data (in email communications or scientific collaborations) or simulations (e.g. EPISIMS simulation). However, the time component is not usually modeled explicitly in the graph of social interactions. Rather, the social graph is static with all the interactions that happened over a period of time present in that graph, sometimes weighted with the frequency of the interaction. However, many questions, especially the ones regarding the stability of social structures, the flow of information or spreading of a disease strongly depend on the temporal aspect of the social interactions. Tanya Berger-Wolf, a post doc at DIMACS in 2004/2005, collaborated with Kah Loon Ng, another DIMACS post doc, Muthu Muthukrishnan, Department of CS, Rutgers, Simon A. Levin, Daniel Rubenstein, Siva Sundaresan, Ilya Fischhoff, Department of EEB, Princeton, and Jared Saia, University of New Mexico on building a consistent mathematical and computational framework that allows investigation of social structures in animal or human populations with the explicit time component. They used equine populations of plain and grevy zebras and the onagers as a case study for their framework. Ng and Berger-Wolf are questioning the general "acceptance" of social networks as being scale-free. They are not entirely convinced that this is true and will proceed with some more work/discussion on identifying a proper framework to model social networks.

Dynamic Data Processing

Often the data collection about a state of an individual and the individual's social interactions is facilitated by various body and remote sensors. The information collected generally has to be transformed from a noisy signal into the abstract state or relationship data. This task becomes especially difficult when the sensor computational, communication and energy power is constrained. Tanya Berger-Wolf, a post doc at DIMACS in 2004/2005, collaborated with Muthu Muthukrishnan, Department of CS, Rutgers, Daniel Rubenstein, Siva Sundaresan, and Ilya Fischhoff, Department of EEB, Princeton to develop general methods for cheap yet accurate dynamic data processing.

Kinship Relationships in Populations

The knowledge of the kinship relationships in a population is often crucial in genetic epidemiology, molecular ecology, population genetics, and conservation biology. The knowledge of familial relationships is needed for many biological applications including the estimation of heritabilities of quantitative characters, studies of mating systems and fitness, and managing populations of endangered species. The more complex the web of relationships, the more generations' data are generally desirable for accurate relationship inference. Tanya Berger-Wolf, a post doc at DIMACS in 2004/2005, has focused on the identification of sibling relationships. Typically, biologists have used parental data to establish sibling groups indirectly through parentage assignments. Reconstructing sibling relationships without parental data is a much more difficult problem, but one that faces many investigators who sample genotype cohorts of offspring rather than parent/offspring groups. In these cases, a reliable method of reconstructing family structure in the population would be extremely useful for studying inheritance patterns, natural selection, breeding biology, and gene flow parameters. Until recently, most sibling relationship reconstruction methods relied on statistical maximum likelihood models. Berger-Wolf and her collaborators, Wanpracha Chaovalitwongse, DIMACS and Department of ISE, Rutgers, Bhaskar DasGupta, DIMACS and Department of CS, University of Illinois at Chicago, Mary Ashley, Department of Biological Sciences, University of Illinois at Chicago, are developing a completely combinatorial approach for reconstruction of sibling relationships from single generation data. Preliminary results show that their approach may provide accurate reconstruction of sibling groups and, unlike statistical methods, it is not sensitive to the number of genetic loci sampled or the allele frequency in any locus. Also, in contrast to the existing statistical methods, it does not require any a priori knowledge about allele frequency, population size, mating system, or family size distributions.

Sensor Placement in Distribution Networks

Tanya Berger-Wolf, a post doc at DIMACS in 2004/2005, collaborated with R. Ravi, Department of CS, Carnegie Melon, William E. Hart, Sandia National Laboratory, and Ari Trachtenberg, Department of ECE, Boston University on the problem of placing sensors in a building or a utility network to monitor the air or water supply. They developed discrete models of the sensor placement problems that aim to either minimize the number of sensors given a time limit on the contamination detection time or minimize contamination detection time given a fixed budget of sensors. They specifically focused on the question of the identification of the source of contamination. They are studying additional aspects and possible objectives of the sensor placement problem.

Unbalanced Graph Cuts as a Model for Limited Vaccination

Ara Hayrapetyan, David Kempe, DIMACS postdoc Martin Pál and Zoya Svitkina collaborated on the problem of isolating the source of a disease from the rest of the population using a limited amount of vaccination. They introduced the Minimum Size Bounded Capacity Cut (MinSBCC) problem, in which there is a given graph with an identified source and the objective is to find a cut minimizing the number of nodes on the source side, subject to the constraint that its capacity not exceed a prescribed bound. In general the MinSBCC problem is NP-complete. They developed an efficient approximation algorithm that either violates the budget constraint or exceeds the optimal number of source-side nodes, but not both. For graphs of bounded treewidth, they showed that the problem with unit weight nodes can be solved optimally in polynomial time, and when the nodes have weights, approximated arbitrarily well. Their work opens directions for future research involving more realistic models of the spread of diseases. The implicit assumption in their node cut approach is that each social contact will always result in an infection. If edges have infection probabilities, for instance, based on the frequency or types of interaction, then the model becomes significantly more complex.

Firefighter Problem: Fire Containment in Grids of Dimension Three and Higher

In the firefighter problem, we seek to control an epidemic working its way through a network by vaccinating a certain number of individuals at each time period or, analogously, seek to control a forest fire by placing a certain number of firefighters at network nodes every time period. The epidemic (fire) spreads each time period to unprotected neighbors (unvaccinated or without firefighters present). While it is known that only two firefighters per timestep are needed in the two dimensional lattice, it had been conjectured by Wang and Moeller that 2d-1 firefighters per timestep are needed to contain a fire outbreak starting at a single vertex in the d-dimensional lattice for d at least 3. Mike Develin and DIMACS graduate student Stephen Hartke (Rutgers, now at University of Illinois) proved this conjecture. They also proved that for each positive integer f, there is some outbreak of fire such that f firefighters per timestep are insufficient to contain the outbreak.

Firefighter Problem on the Two-dimensional Infinite Grid:

Kah Loon Ng, DIMACS post doc and Paul Raff, first year graduate student at Rutgers, are currently working on the firefighter problem on the two dimensional infinite grid. They are mainly interested in the question of whether a fire can be contained after a finite number of moves. It was shown that one firefighter per time step is not sufficient while two per time step is (requiring a minimum of 8 time steps and 18 burnt vertices). A natural question to ask is: What if we have a non-constant number of firefighters per time step? One way of looking at the idea of having a non-constant number of firefighters is to have a periodic sequence of firefighters available for deployment. So in the two cases above, the sequences are 1,1,1,1,.... and 2,2,2,2,... The average number of firefighters in any given period is 1 and 2 for the two sequences above respectively, and so the average of 1 is not sufficient for containment while 2 is. Based on initial empirical results, Ng and Raff conjectured that the cut-off point for the average for successful containment is 1.5. They have successfully implemented a strategy (with codes written in C) to show that any periodic sequence of n ones and (n+1) twos (1,1,1,1,...,2,2,2,...,2) is sufficient for containment. This proves that for any epsilon e>0, 1.5+e is sufficient. Currently, Ng and Raff are working on proving that the average of 1.5 is not enough There are many other possible questions that arise from having a non-constant no. of firefighters. If possible, Ng and Raff would also like to proceed to look at other class of graphs beyond infinite grids, or infinite grids of higher dimension.

Firefighter Problem on Trees and Integer Programming:

DIMACS graduate student Stephen Hartke (now at University of Illinois) studied the integer program of MacGillivrary and Wang that was used to determine the optimal sequence of vaccinations or firefighter responses in the firefighter problem under the special assumption that the network is a tree. Hartke found additional constraints to the integer program that in practice narrow the integrality gap between the integer programming optimal and the optimal of the linear programming relaxation.

Monitoring Multiple Data Streams

S. Muthukrishnan, Rutgers, has been working with Bell Labs researchers on how to analyze data from multiple sensors, in particular, when you have to continuously monitor values from sensors and perform analyses on them. For example, how to continuously monitor what are most frequent observations from many sensors, how to continuously monitor the quantiles of the combined distribution of observed values etc. The key here is that we want to minimize communication as a resource and yet give accuracy guarantees at all times. This needs new algorithms for communication, space and time efficient distributed tracking of the combined distributions. His algorithms will be presented at ACM SIGMOD 2005.

Group EEG Measurements

S. Muthukrishnan, Rutgers, has been working with David Rosenbluth, Telcordia, and Rutgers graduate student, Smriti Bhagat in measuring and analyzing EEG data from multiple subjects simultaneously and trying to form a "group EEG," that is an EEG with significantly more signal in it than any of the individual EEGs. In particular, they have been able to detect stimulus in controlled experiments that are significantly earlier than the response of any individual subject to the stimulus.

An Improved Method of Modeling an Epidemic on a Graph

Alain Franc, Institut National de la Recherche Agronomique (INRA), Département Ecologie des Forêts, Prairies et Milieux Aquatiques, CDA & UMR Biogeco, Bordeaux, was an early participant in this special focus. His involvement, together with encouragement by organizer Simon Levin, Princeton, led to collaboration between Franc, Biometrie INRA, Avignon, on modeling the spread of infectious diseases. Franc and Peyrard started from the already known simple observation that an SIS model is at the same time similar to Levin's equation in metapopulation dynamics and a mean field approximation of a contact process, establishing a direct link between metapopulation dynamics, mathematical epidemiology, and modeling invasive species. In this context, their work bridges two recent trends in mathematical epidemiology connected with the spatial structure of hosts or patches: (i) modeling the spread of an epidemic on a graph, and (ii) pair approximation already implemented on lattices for SIS/SIR models. In this work, they have derived the equation for pair approximation on a graph, knowing the degree distribution only (such an equation existed either for pair approximation on regular lattices, or as mean field only when taking degree distribution on a graph). Second, they have emphasized that pair approximation is a special simplified case of so called Bethe approximation, or cluster variation methods (they exist in physics literature with various wordings). In pair approximation, triplets are always open, i.e. the fact that a triangle in a graph can be closed is not taken into account. They have show how the clustering coefficient, measuring the fraction of closed triangles, can be used to lead to an additional term in the pair approximation at equilibrium. This allows one to compute the Bethe approximation of the spread of an epidemic on a graph taking into account both the degree distribution and the clustering coefficient of the graph.

Exact Methods Applied to Group Sequential and other Stratified Comparative Poisson Designs

Group sequential methods have been commonly applied to clinical trials with continuous or non-rare binary outcome variables, but not until now to comparative Poisson trials (of Poisson or rare binary outcomes). Graduate student Qi Xia, Rutgers University, working with her advisor, Don Hoover, developed a series of exact group sequential methods and applied them to comparative Poisson trials. The methods are based on statistical inference on the relative risk of treated vs. control groups in stratified comparative Poisson studies. Xia showed that application of exact group sequential methods can greatly reduce the comparative Poisson trial length in terms of expected number of events needed to be observed, particularly when the null hypothesis is false. Application of curtailed early stoppage rules (with futility bounds) combined with group sequential methods further reduces the trial length particularly when the null hypothesis is true. Xia developed a set of Splus programs to implement the calculation of test size, power, and expected number of events to be observed for exact group sequential comparative Poisson trials. Xia created examples to demonstrate how to plan a group sequential trial using these Splus programs before a trial and to make dynamic adjustments during a trial. These methods can be expanded to plan an exact stratified group sequential comparative Poisson trial to account for more explanatory variables.

Extending Power and Sample Size Approaches for McNemar's Procedure to General Sign Tests

Current software and textbooks present procedures to estimate power and sample size for sign tests that only apply to settings where positive (i.e., X=1) or negative (i.e., X=-1), but not neutral (i.e., X=0) outcomes occur. However, many studies analyzed by sign tests involve the more general setting where significant amounts of neutral outcomes can occur. Donald Hoover (Rutgers University) applied existing power / sample size approaches, developing software for matched binary responses (McNemar's discordant pairs) to general sign tests with neutral outcomes occurring. These statistical methods are used for testing interventions (such as vaccines) to prevent rare outcomes such as the type that could be of concern for bioterrorism.

Integrating Within-host and Between-host Pathogen Dynamics

Bob Holt’s participation at the DIMACS workshop led him and his collaborator Michael Barfield, both at the Department of Zoology, University of Florida, to work on a new problem, having to do with the integration of within-host and between-host pathogen dynamics. Their work will appear in a DIMACS volume, and a follow-up manuscript in preparation. Briefly, they explored two things. First, for pathogens that are being shed during the course of host infection, transmission can be viewed as a reallocation from potential within-host growth and movement between hosts. This perspective leads to a new formulation for the growth rate of a pathogen population, one that integrates within-host and between-host dynamics. Second, within-host dynamics often have transient phases, and the timing of transmission relative to these transients can have important implications for the overall course of disease progression in a host population.

Unbiased Probability Spaces

The notion of unbiased probability spaces over a set of vertices N is a generalization of the well-known Erdos-Renyi model for random graphs. Making some reasonable simplifying assumptions, it is possible to relate the likelihood that a contagious infection will spread through a social network to the expected size of the largest component in an independent unbiased probability space. DIMACS visitors James Abello and Mike Capalbo have proven a graph theoretical result that “almost” characterizes for which of these spaces it is the case that almost surely a giant component exists. In addition, Abello and Capalbo studied vaccination strategies in social networks, assuming the social network is known and the social network satisfies a certain structural property of being “highly connected.” In particular, Abello and Capalbo developed gives an algorithm for effective vaccinating strategies when there is only a limited amount of vaccine to use during each step.

Blocking Sequences

Motivated by a question formulated by Fred Roberts, DIMACS visitors James Abello and Mike Capalbo have introduced the notion of blocking sequences for seed set S of vertices in a connected graph G. Blocking sequences are a formalization of a sequence of sets that if vaccinated will contain the propagation of an infection that starts on a seed set S in a social network G. Since several variations of this problem are NP-Hard, Abello and Capalbo developed an approximation algorithm for this problem. They also developed another application of blocking sequences. They showed that for d dimensional integer infinite grids, there exist seed sets S and functions f(t) for which there are no blocking sequences whose sizes are upper bounded by f(t). This work is closely related to the firefighter problem studied by Ng and Raff (see above), with whom Abello and Capalbo collaborated while the four were at DIMACS.

Maximum Cliques in Sparse Power Law Graphs with “Large” Clustering Coefficients

There was some hope the Max Clique problem was polynomial computable for sparse power law graphs with reasonable large clustering coefficients. DIMACS visitors James Abello and Mike Capalbo proved that the max clique problem remains NP-Hard even for this class of graphs. It follows from the proof that it cannot even be approximated within a factor of n^{α} with alpha some constant.

Biological Aggregation at the Interface between Theory and Practice

William Day and Mark Wilkinson raised some ideas at the Tree of Life meeting in March 2003 that Day developed into a strategy for understanding evolutionary processes. To understand evolutionary processes better, biologists use aggregation methods to estimate evolutionary relationships; yet properties of the methods are sometimes so imprecisely defined, and their interrelationships so poorly understood, that useful formal results might be difficult to obtain. To address this problem Day developed a strategy for modeling aggregation methods and studying their properties. The approach accommodates impossibility results for aggregating rankings, nonhierarchical classifications, hierarchies, and phylogenies. He plans to formulate other relevant models of biological aggregation and to characterize methods for solving biological problems of agreement and synthesis.

Assessing the Risks to Exotic Trees

In response to The Health Forests Restoration Act of 2003, Title IV (“The President’s Healthy Forests Initiative”), which calls for a comprehensive program to inventory, monitor, characterize, assess, and identify forest stands to identify forest threats. W.D. Smith, USDA Forest Service, was asked to perform a 4 month detail with the USDA Forest Service Forest Health Protection’s (FHP) Forest Health Technology Enterprise Team (FHTET) to develop procedures to produce risk assessment maps for three exotic species in FY2005: *Ips typographus*, *Sirex noctilio*, and *Phytophthora alni*. In addition to addressing these immediate concerns, the procedures developed and documented will provide direction for similar activities within the Agency. Evaluating and mapping the risk and hazard of insect and disease outbreaks has become an essential requirement for Forest Health Protection (FHP). Risk maps have allowed FHP to target invasive species detection efforts for *Phytophthora ramorum* (SOD), and have provided valuable information at the national level as to which forest resources are at the most risk for tree damage and mortality. The continued utility of these outputs though, depends on increasing our capability of producing maps for individual forest pest species of concern. Given improved methods for producing host species and potential pest distribution maps, accumulated over space, these individual species distribution maps could lead to the creation of an improved nation-wide risk and hazard map, and may be used at the regional level to rank forested areas according to risk and management needs. As it is not possible for FHP to collect additional data for the purposes of Smith’s work, existing data that is already collected by various entities must be utilized. These data have often been collected to answer different kinds of questions and with no regard to ecological processes such as the disturbances caused by pest outbreaks. Therefore, innovative techniques for adapting and enhancing available data attributes or data derived from these attributes to answer forest health questions are required. Progress toward producing these forest risk and hazard outputs depends on the development of scientifically credible statistical techniques and methods for integrating and analyzing available data sources. The exposure to data mining techniques Smith acquired as a participant in the DIMACS Working Group on Data Mining and Epidemiology have been invaluable in leading the research component in this effort.

Upper and Lower Bounds in the Weak Sensor Model

In sensor networks, small devices with processing, sensing and communication capabilities called sensor nodes are randomly deployed over an area in order to achieve sensing tasks after self-organizing as a wireless radio network. Sensor networks can be used in almost any application where sensing or measuring physical variables over large areas is needed, e.g. gathering biological data, sensing in catastrophe areas, etc. Unfortunately, since sensor nodes have strong limitations such as memory size, life cycle and range of communication, the initialization and maintenance of such a network is not a trivial task. Only models that reflect all the various constraints of sensor nodes are adequate to study problems in sensor networks. Although sensor networks is a very active research area, most of this work is either empirical or includes unrealistic assumptions regarding the capabilities of the sensor nodes. Hence, there are still many open basic questions. In order to properly address this and other issues in sensor networks, Miguel A. Mosteiro, Martin Farach-Colton, and Rohan Fernandes, all at Rutgers, defined a complete model for sensor nodes in that they call the Weak Sensor Model. Described here are two results they obtained, an upper bound and a lower bound for initialization of a sensor network under the harsh conditions of the Weak Sensor Model. Problems in sensor networks are twofold: geometric and network. Sensor nodes have limited range and are deployed at random over a large area. Therefore, geometric properties like path length, coverage and connectivity need to be understood. Thus, sensor networks are modeled as random geometric graphs in the plane. On the other hand, node limitations, shared radio-communication channels and lack of additional infrastructure at deployment impose limitations on network protocols. Given these difficult conditions, i.e. a group of weak sensor nodes deployed in a geometric random distribution, a natural question is how to organize such a network. Mosteiro, Farach-Colton, and Fernandes developed an optimal-network bootstrapping protocol, i.e. a protocol for network formation under the constraints of a harsh Weak Sensor Model in a geometric random distribution. In this protocol, the optimization criterion is to maximize life cycle subject to the Weak Sensor Model constraints. They observed that the optimal path between any two nodes is the shortest in terms of number of hops and they proved the existence of a hop-optimal subgraph of any random geometric graph modeling a sensor network. Furthermore, we give the first $O(\log^2 n)$ algorithm for constructing the network modeled by such a subgraph within weak sensor conditions. The upper bound described above is the best known for initialization of networks modeled as random geometric graphs. However, tight lower bounds that take advantage of the various constraints of sensor networks were not yet obtained. An efficient solution of various closely related problems such as maximal independent set (MIS), dominating set and leader election, is the ideal building block to dynamically give structure to a sensor network in order to setup a network backbone or to color nodes in order to select different frequency hopping sequences and handle channel contention in this setting. Hence, a lower bound for any of these problems in sensor networks would also be interesting. Any algorithm for initialization of a sensor network, for MIS or for any of the related problems, has to achieve a non-colliding transmission in order to broadcast the topology information. Furthermore, given the unreliability of sensor nodes, no synchronous wake-up of nodes can be assumed. Therefore, a lower bound for the problem of waking up all sensor nodes is also a lower bound for the initialization of the network. The previously demonstrated lower bound left a significant gap with the best-known upper bound. Mosteiro, Farach-Colton, and Fernandes improved that lower bound, closing the gap for the wake up problem in a non-uniform node-distribution setting. They are currently working on extending this result to the uniform case. Farach-Colton is a senior faculty member in the project, Fernandes a project graduate student, and Mosteiro a graduate student in Farach-Colton's group.

Population Dynamics in Spatially Heterogeneous Systems with Drift

Project graduate student Jaewook Joo and Professor Joel L. Lebowitz, both at Rutgers, investigated the time evolution and stationary states of a stochastic spatially discrete population model (contact process) with spatial heterogeneity and imposed drift (wind) in one- and two-dimensions. They considered in particular a situation in which space is divided into two regions: an oasis and a desert (low and high death rates). Depending on the values of the drift and other parameters, the population in the stationary state will be zero, localized, or delocalized. The phase diagram is similar to that obtained by Nelson and coworkers from a deterministic, spatially continuous model of a bacterial population undergoing convection in a heterogeneous medium.

The Computational Epidemiology Research Laboratory

The participation of Tom Jacob and Armin R. Mikler in the early DIMACS workshops and group meetings of this special focus led to several projects at the Computational Epidemiology Research Laboratory (CERL) at the University of North Texas. These projects are described below.

Modeling an Influenza Outbreak

In an effort to prevent an influenza pandemic such as the one witnessed in 1918, which killed as many as 100 million people world wide, disease monitoring and syndromic surveillance methods have been deployed. The methods are designed to identify early cases of influenza and guide the allocation of public health resources to control and contain an outbreak. Nevertheless, the dynamics and progression of influenza in a given population remains elusive and cannot be easily derived. At CERL, faculty and students are currently developing computational models that attempt to reverse engineer influenza outbreaks, thereby extracting the geographic and demographic characteristics that might affect influenza outbreak patterns. The development of a framework for the simulation of influenza outbreaks in multiple regions with different geography, infrastructure, and populations with diverse demographics forms the basis for this effort. Local health officials can generally only observe the combined demand for treatment in the event of an infectious disease outbreak. The tools developed at CERL will facilitate the analysis of an outbreak as the superposition of multiple smaller outbreaks in distinct regions or demographic subgroups. With these tools, epidemiologists and public health officials can engage in a detailed what-if-analysis, thereby experimenting with different vaccination or prevention strategies and optimizing the allocation of public health resources across the region.

Modeling of the Human Papilloma Virus

In the past several years there have been significant improvements in our understanding of cervical cancer. In 2001, the United States health care system spent over $1.5 billion on treatment for cervical dysplasia and an additional $2 billion on screening tests such as pap smears. Human Papilloma Virus (HPV) DNA is found in 99.7% of all cervical cancers. An effective HPV vaccine would have significant impact on HPV infections and cervical disease. Candidate vaccines finished phase 2 testing in the United States and phase 3 trials have begun. Because of the health care costs associated with this virus, it is important to have an effective vaccination strategy in place when this vaccine becomes available in the near future.

Computational models are important tools in determining the transmission dynamics of disease and an efficient and effective vaccination strategy. Jacob and Mikler are creating these models at CERL to aid in our understanding of disease patterns and the probable impact of an intervention or vaccine. Their HPV model stratifies a population into different subgroups based on sexual mixing patterns. They analyze population demographics and census data to extract demographic parameters for their model and mine risk behavior studies in youths to determine the sexual partner exchange rates for a population. Using the HPV model that CERL is developing, Jacob and Mikler hope to be able to offer an effective vaccination solution.

Hidden Markov Models and Bayesian Disease Modeling

Bayesian models are designed to portray the dynamics of diseases in epidemiological sciences. The incidence and prevalence of diseases in a given population, with varied geographic and demographic settings, are analyzed over the temporal domain to build dynamic Bayesian networks. The network illustrates the stochastic dependencies of the demographics on the prevalence of symptoms and their related diseases. An underlying hidden Markov model is designed to inter-link the hidden disease characteristics to the observed prevalence of diseases. A disease outbreak simulator generates synthetic data of epidemic outbreaks in specific populations. While this study involves mining synthetic data generated from the disease outbreak simulator for useful information, it can also be applied to real data from disease studies to uncover previously unknown inferences. High performance computing is a requisite to port the model onto larger domains with finer granularity of results. The Bayesian model shall aid as a predictive framework for analyzing prevalence of diseases in varied geographic and demographic settings.

Outreach Activities

This project is closely intertwined with DIMACS efforts to link mathematics and computer science with biology in the high schools. The project organizers were involved in planning a DIMACS conference on this subject in April 2005 (see http://dimacs.rutgers.edu/Workshops/Biomath/). Also, the project organizers will work closely with the Summer 2005 DIMACS Bio-math Connect Institute (BMCI), which is aimed at introducing high school math/CS and Bio teachers to topics at the interface. This project is informing the BMCI effort and specific topics from the project are being adapted for use in BMCI.

Products:

Books

Abello, J., and Cormode, G. (eds.), *DIMACS Computational Methods in Epidemiology, AMS-DIMACS Volume Series*, American Mathematical Society, 2005, in press.

Papers

Abbas, K., Mikler, A., and Gatti, R., “Temporal analysis of infectious diseases: Influenza,” in *Proceedings of the ACM Symposium on Applied Computing* (SAC '05), Santa Fe, NM, March, 2005.

Abbas, K., Mikler, A., Ramezani, A., and Menezes, S., “Computational epidemiology: Bayesian disease surveillance,” in *Proceedings of the International Conference on Bioinformatics and its Applications* (ICBA'04), Fort Lauderdale, FL, December, 2004.

Abello, J., and Capalbo, M., “Random graphs (and the spread of infections in a social network),” *DIMACS Computational Methods in Epidemiology, AMS-DIMACS Volume Series*, Abello, J., and Cormode, G. (eds.), American Mathematical Society, to appear.

Abello, J., and Capalbo, M., “An approximation algorithm to the modified quarantine problem on expander graphs, to be submitted

Abello, J., and Capalbo, M., “Blocking sequences in infinite grids,” to be submitted.

Abello, J., and Capalbo, M., “Max cliques in sparse power law graphs with large clustering coefficients,” *Internet Mathematics*, to be submitted.

Abello, J., Cormode, G., Fradkin, D., Madigan, D., Melnik, O., and Muchnik, I., “Selected data mining concepts,” *DIMACS Computational Methods in Epidemiology, AMS-DIMACS Volume Series*, Abello, J., and Cormode, G. (eds.), American Mathematical Society, to appear.

Abello, J., Pogel, A., “Graph partitions and concept lattices,” *DIMACS Computational Methods in Epidemiology, AMS-DIMACS Volume Series*, Abello, J., and Cormode, G. (eds.), American Mathematical Society, to appear.

Abello, J., Pogel, A., Miller, L., “Graph partitions and formal concept lattices,” *Journal of Universal Computer Science*, **10** (2004), 934-954.

Abello, J., Pogel, A., Miller, L., “LABFS bigraph decompositions and concept lattices”, *Journal of Universal Computer Science*, **10** (2004).

Angeli, D., De Leenheer, P., and Sontag, E.D., “A small-gain theorem for almost global convergence of monotone systems,” *Systems and Control Letters*, **52** (2004), 407-414.

Angeli, D., De Leenheer, P., and Sontag, E.D., “Monotonicity and convergence in chemical reaction networks,” in *Proceedings of 43rd IEEE Conference on Decision and Control*, Bahamas, 2004.

Arino, J., Davis, J.R., Hartley, D., Jordan, R., Miller, J.M., and Van Den Driessche, P., “A multi-species epidemic model with spatial dynamics,” *Mathematical Medicine and Biology Advance Access*, March 18, 2005, on-line publication.

Berger-Wolf, T.Y., Hart, W.E., and Saia, J., “Discrete sensor placement problems in distribution networks,” *Journal of Mathematical and Computer Modelling*, to appear.

Berger-Wolf, T.Y., DasGupta, B., Chaovalitwongse, W., and Ashley, M.V., “Combinatorial reconstruction of sibling relationships,” Poster, *RECOMB Conference*, Boston, MA, May 14-18, 2005.

Berger-Wolf, T.Y., DasGupta, B., Chaovalitwongse, W., and Ashley, M.V., “Combinatorial reconstruction of sibling relationships,” in *Proceedings of the 6th International Symposium on Computational Biology and Genome Informatics*, Salt Lake City, Utah, July 21 - 26, 2005.

Chaovalitwongse, W.A., Berger-Wolf, T.Y., DasGupta, B., and Ashley, M.V., “Set covering approach for reconstruction of sibling relationships,” *Journal of Optimization Methods and Software*, submitted.

Chen, S., Gaur, A. Muthukrishnan, S., and Rosenbluth, D., “Wireless *in loco* sensor data collection and applications,” *MOBEA II, WWW ‘04*.

Cooper G.F., Dash, D.H., Levander, J.D., Wong, W.K., Hogan, W.R., Wagner M.M., “Bayesian biosurveillance of disease outbreaks,” in *Proceedings of the Conference on Uncertainty in Artifical Intelligence* (2004) 94-103.

Cox, L., Gonzalez, Jr., J. F., and Katzoff, M., “Effects of rounding continuous data using specific rules,” in *Proceedings of the 2004 ASA Joint Statistical Meetings Survey Research Methods Section*, to appear.

Cox, L., Gonzalez, Jr., J. F., and Katzoff, M., “Effects of grouping continuous data on first and second distribution moments,” in *Proceedings of the2004 ASA Joint Statistical Meetings Survey Research Methods Section*, to appear.

Cross, P. C., Lloyd-Smith, J. O., Bowers, J., Hay, C. T., Hofmeyr, M., and Getz, W. M., “Integrating association data and disease dynamics: An illustration using African buffalo in Kruger National Park,” **41** (2004), 879-892.

Day, W. H. E., “Biological aggregation at the interface between theory and practice,” *Proceedings of ASMDA-2005*, submitted.

De Leenheer, P., Angeli, D., and Sontag, E.D., “A tutorial on monotone systems -with an applicationx to chemical reaction networks,” *Proceedings of MTNS2004* (16th International Symposium on Mathematical Theory of Networks and Systems), Leuven, Belgium, July 5-9, 2004.

De Leenheer, P., Angeli, D., and Sontag, E.D., “On predator-prey systems and small gain theorems,” *Mathematical Biosciences and Engineering*, **2** (2005), 25-42.

De Leenheer, P., Angeli, D., and Sontag, E.D., “Monotone chemical reaction networks,” *Journal of Mathematical Chemistry*, to appear.

De Leenheer, P., Angeli, D., and Sontag, E.D., “Crowding effects promote coexistence in the chemostat,” *Journal of Mathematical Analysis and Applications*, submitted.

De Leenheer, P., Levin, S.A., and Sontag, E.D, and Klausmeier, C.A., “Global stability in a chemostat with multiple nutrients,” *Journal of Mathematical Biology*, submitted.

De Leenheer, P., Li,, B., and Smith, H.L., “Competition in the chemostat: some remarks,” *Canadian Applied Mathematics Quarterly*, to appear.

De Leenheer, P., and Malisoff, M., “A small-gain theorem for monotone systems with multi-valued input-state characteristics,” submitted.

De Leenheer, P., and Sontag, E.D., “A note on the monotonicity of matrix Riccati equations,” DIMACS Tech Report 2004-36.

Develin, M., and Hartke, S.G., “Fire containment in grids of dimension three and higher,” submitted.

Dushoff, J., J., Plotkin, B., Levin, S.A., and Earn, D.J.D., “Dynamical resonance can account for seasonality of influenza epidemics,” *PNAS*, **1** (2004), 16915-16916.

Eidelson, B.M., and Lustick, I., “VIR-POX: An agent-based analysis of Smallpox preparedness and response policy,” *Journal of Artificial Societies and Social Simulation*, **7** (2004).

Farach-Colton, M., Fernandes, R.J., and Mosteiro, M.A., “Bootstrapping a hop-optimal network in the weak sensor model,” in preparation.

Farach-Colton, M., Fernandes, R.J., and Mosteiro, M.A, “The wake up problem in the weak sensor model,” in preparation.

Fradkin, D., Muchnik, I., Hermans, P., and Morgan, K., “Validation of epidemiological models: Chicken epidemiology in the UK,” *DIMACS Computational Methods in Epidemiology, AMS-DIMACS Volume Series*, Abello, J., and Cormode, G. (eds.), American Mathematical Society, to appear.

Franc, A., and Peyrard, N., “Cluster variation approximations for a contact proc living on a graph,” *Physics A*, submitted.

Frisén M., "Statistical measures for evaluation of methods for syndromic surveillance," in *Conference Proceedings 2003 National Syndromic Surveillance, Morbidity and Mortality Weekly Report Supplement* (2004).

Frisén, M. and Sonesson, C., "Optimal surveillance," in *Spatial Surveillance for Public Health*, A. Lawson and K. Kleinman (eds), K. Wiley, 2005, 31-57.

Gilbert, A.C., Muthukrishnan, S. and Strauss, M.J., "Improved time bounds for near-optimal sparse fourier representations", DIMACS Tech Report 2004-49.

Hartke, S. G., "Attempting to narrow the integrality gap for the firefighter problem on trees," *DIMACS Computational Methods in Epidemiology, AMS-DIMACS Volume Series*, Abello, J., and Cormode, G. (eds.), American Mathematical Society, to appear.

Hauben, M., Madigan, D., Gerrits, C., and Meyboom, R., "Quantitative methods in drug safety signal detection," *Expert Opinion on Drug Safety*, submitted.

Hayrapetyan, A., Kempe, D., Pál, M., and Svitkina, Z., "Unbalanced graph cuts," *ESA*, (2005), submitted.

Hermans,, P., Fradkin, D., Muchnik, I., and Morgan, K. "Prevalence of wet litter and associated risk factors in broiler ocks in the UK," *Veterinary Review*, (2005), to appear.

Hoover, D.R., "Subject allocation and curtailment for fixed event comparative Poisson trials," *Statistics in Medicine*, **23** (2004), 1229-1245.

Hoover, D.R., "Mental illness and length of hospital stay for Medicaid inpatients infected with HIV," DIMACS Tech Report 2005-10.

Hoover, D.R., "Extending power and sample size approaches for McNemar's procedure to general sign tests," DIMACS Tech Report 2005-16.

Hoover, D.R., "Extending power and sample size approaches developed for McNemar's procedure to general sign tests," *International Statistical Review*, **73** (2005), 103-110.

Hoover, D.R., "Design and analysis of group level studies," in *Handbook of Urban Health*, D. Vlahov and S. Galea (eds.), Verlag-Springer, New York, 2005, in press.

Hoover, D. R., "Subject allocation and curtailment for fixed event comparative Poisson trials," *Statistics in Medicine*, **23** (2004), 1229-1245.

Hoover, D.R., Sambamoorthi, U., Walkup, J., and Crystal, S., "Mental illness and length of inpatient stay for Medicaid recipients with AIDS," *Health Services Research*, **39** (2004), 1319-1340.

Joo, J., and Lebowitz, J. L., "Behavior of susceptible-infected-susceptible epidemics on heterogeneous networks with saturation," *Phys. Rev. E*, **69** (2004), 066105-066113.

Joo, J., and Lebowitz, J. L., "Pair approximation of the stochastic susceptible-infected-susceptible epidemic model on the hypercubic lattice," *Phys. Rev. E*, **70** (2004), 036114-036123.

Joo, J., and Lebowitz, J. L., "Population dynamics in spatially heterogeneous systems with drift: the generalized contact process," in preparation.

Li, J., Muchnik, I., and Schneider, D., “Influences on breast cancer survival via SVM classification in the SEER Database,” *DIMACS Computational Methods in Epidemiology, AMS-DIMACS Volume Series*, Abello, J., and Cormode, G. (eds.), American Mathematical Society, to appear.

Mikler, A., Jacob, R., Gunupudi, V., and Patlolla, P., “Agent-based simulation tools in computational epidemiology,” *Proceedings of the International Conference on Innovative Internet Community Systems* (I2CS '04), Guadalajara,Mexio, June 2004.

Mikler, A., Venkatachalam, S., and Abbas, K., “Modeling infectious diseases using global stochastic cellular automata,” *Journal of Biological Systems*, accepted.

Mitchell, C.S., Hoover, D.R., and Margolick, J.B., “Evaluation of immunologic markers in multiple chemical sensitivity,” in preparation.

Muthukrishnan, S., “Nonuniform sparse approximation with Haar wavelet basis,” DIMACS Tech Report 2004-42.

Muthukrishan, S. and Strauss, M.J., “Approximate histogram and wavelet summaries of streaming data,” DIMACS Tech Report 2004-52.

Ng, K-L., and Raff, P., "Firefighting on a two dimensional infinite grid with a non-constant number of firefighters," in preparation.

Ortega, O., Sanders, J., and Riddle, M., “Cost-benefit analysis of introducing a Rotavirus vaccine in Cairo, Egypt,” in preparation.

Ozonoff, D., “Superfund basic research program: A model for contemporary research programs: Guest Editorial,” *Environmental Health Perspectives*, **111** (2003).

Ozonoff, D., Pogel, A., and Hannan, T., “Generalized contingency tables and concept lattices,” *DIMACS Computational Methods in Epidemiology, AMS-DIMACS Volume Series*, Abello, J., and Cormode, G. (eds.), American Mathematical Society, to appear.

Pilyugin, S. S., Reeves, G. T., and Narang, A., “Stability of mixed microbial cultures: Connecting theory and experiments, Part 1. Unstructured model,” submitted to *Mathematical Biosciences.*

Pilyugin, S. S., Reeves, G. T., and Narang, A., “Stability of mixed microbial cultures: Connecting theory and experiments, Part 2. Structured model,” submitted to *Mathematical Biosciences.*

Roberts, F.S., “Computational and mathematical epidemiology,” *Science, Next Wave* 2004 (also *Science* 303 (2004), 717.

Roberts, F.S., “Decision support algorithms for port-of-entry inspection,” in *Working Together: Research & Development Partnerships in Homeland Security, Proceedings of DHS/IEEE Conference*, Boston, 2005.

Robson, B., “Clinical and pharmacogenomic data mining: 2. A simple method for the combination of information from associations and multivariances to facilitate analysis, decision, and design in clinical research and practice,” *Journal of Proteome Research,* **3** (2004), 697-711.

Roverato, A., “A unified approach to the characterization of equivalence classes of DAGs, chain graphs with no flags and chain graphs,” *Scandinavian Journal of Statistic,* 2005, to appear.

Stoto, M., Fricker, Jr., R.D., Jain, A., Davies-Cole, J.O., Glymph, C., Kidane, G., Lum, G., Jones, L., Dehan, K., and Yuan, C., “Evaluating statistical methods for syndromic surveillance,” in *Statistical Methods in Counter-Terrorism,* Wilson A and Olwell D. (eds.), to be published in the ASA-SIAM series.

Struchiner, C., “Insect gene transformations and DNA vaccines for Malaria,” *Journal of Biological Systems*, accepted.

Venkatachalam, S., and Mikler, A., “Towards computational epidemiology: Using stochastic cellular automata in modeling spread of diseases,” in *Proceedings of the 4th Annual International Conference on Statistics, Mathematics and Related Fields*, Honolulu, HI, January, 2005.

Venkatachalam, S., and Mikler, A., “An infectious disease outbreak simulator based on the cellular automata paradigm,” in *Proceedings of the International Conference on Innovative Internet Community Systems* (I2CS '04), Guadalajara, Mexico, June 2004.

Verriest, E.I., “Regularization method for optimally switched and impulsive systems with biomedical applications,” in *Proceedings of the 42nd IEEE Conference on Decision and Control*, Maui, HI, December 2003, 2156-2161.

Wong, W.K., Cooper, G.F., Dash, D.H., Levander, J.D., Dowling, J., Hogan, W.R., Wagner, M.M., “Bayesian biosurveillance using multiple data streams,” *Morbidity and Mortality Weekly Report*, to appear.

Talks

Berger-Wolf, T.Y., “Combinatorial reconstruction of sibling relationships,” 6th International Symposium on Computational Biology and Genome Informatics, Salt Lake City, Utah, July 21 - 26, 2005.

Berger-Wolf, T.Y., “Who's company? Identifying persistent social groups,” Celebration of Women in Computing, University of Illinois, Urbana-Champaign, April 1-3, 2005.

Berger-Wolf, T.Y., “Reconstructing phylogenetic trees: How good is good enough?” Symposium on Computational Science of Biomolecules: Applications in Medicine and Therapeutics, University of Illinois, Chicago, October 2004.

Berger-Wolf, T.Y., “Reconstructing phylogenetic trees: How good is good enough?” Department of Computer Science, Illinois Institute of Technology, October 2004.

Berger-Wolf, T.Y., “Online consensus of phylogenetic trees,” 4th Workshop on Algorithms in Bioinformatics (WABI), Bergen, Norway, September 2004.

Cormode, G., Garofalakis, M., Muthukrishnan, S., Rastogi, R., “Holistic aggregates in a networked world: Distributed tracking of approximate quantiles,” SIGMOD 2005.

Day, W. H. E., “Biological aggregation at the interface between theory and practice,” ASMDA-2005, Brest, May 2005.

De Leenheer, P., “Monotonicity and convergence in chemical reaction networks,” 43rd IEEE Conference on Decision and Control, Bahamas, 2004.

De Leenheer, P., “A tutorial on monotone systems -with an application to chemical reaction networks,” MTNS2004 (16th international symposium on mathematical theory of networks and systems), Leuven, Belgium, July 5-9, 2004.

De Leenheer, P., “Crowding effects promote coexistence in the chemostat,” Department of Mathematics Colloquium, University of Florida, January 30, 2004.

De Leenheer, P., “Crowding effects promote coexistence in the chemostat,” Department of Mathematics Colloquium, University of Iowa, February 18, 2004.

De Leenheer, P., “Crowding effects promote coexistence in the chemostat,” Department of Mathematics Colloquium, Queen's University, February 23, 2004.

De Leenheer, P., “Growth on 2 nutrients in the chemostat: an application of monotone systems theory,” PACM Dynamical Systems/Nonlinear Science Seminar, Princeton, April 2, 2004.

De Leenheer, P., “An introduction to monotone systems,” Applied Mathematics Seminar, University of Florida, October 26, 2004.

Ozonoff, D., “Challenges of interdisciplinary work,” Superfund Basic Research Program External Advisory Group, NIEHS, North Carolina, April 22, 2004.

Ozonoff, D., “Ask me something easier: a researcher confronts community concerns,” Seminar presentation, Institute for Toxicology, Michigan State University, April 23, 2004.

Ozonoff, D., “Maps, paths and lattices: New tools for old problems,” Distinguished Lecture Series, National Institute for Environmental Health Sciences (NIEHS), Research Triangle Park, NC, April 24, 2003.

Ozonoff, D., “Maps, paths and lattices: New methods for old problems,” Institute of Toxicolgy, Michigan State University, East Lansing, MI, April, 2003.

Ozonoff, D., “Generalized and closed set contingency tables in epidemiology,” Environmental Biostatistics Seminar, Harvard School of Public Health, Boston, MA, May, 2004.

Roberts, F.S., “Homeland security: What can mathematics do?” American Mathematical Society Congressional Briefing, House Office Building, Washington, DC, December 2004.

Roberts, F.S., “Measurement theory and applications,” 6-hour tutorial at DIMACS Working Group on Order-theoretic Aspects of Epidemiology, March 2005.

Roberts, F.S., “On balanced signed graphs and consistent marked graphs,” International Symposium on Graph Colorings and their Generalizations, Budapest, Hungary, April 2005.

Roberts, F.S., "Competition graphs of semiorders," Workshop on Computer Science and Decision Theory, Paris, October 2004 and also at Mathematical Psychology National Meeting, Ann Arbor, Michigan, August 2004.

Main Web Site

http://dimacs.rutgers.edu/SpecialYears/2002_Epid/

Other Specific Products

Web Pages

Working Group Meeting: Phylogenetic Trees and Rapidly Evolving Diseases
http://dimacs.rutgers.edu/Workshops/WGPhylogeneticTrees

Workshop: Reticulated Evolution
http://dimacs.rutgers.edu/Workshops/Reticulated

Working Group: Reticulated Evolution
http://dimacs.rutgers.edu/Workshops/Reticulated_WG

Working Group: Data De-Identification, Combinatorial Optimization, Graph Theory, and the Stat/OR Interface
http://dimacs.rutgers.edu/Workshops/Stat

Working Group Meeting: Adverse Event/Disease Reporting, Surveillance and Analysis
http://dimacs.rutgers.edu/Workshops/AdverseCDC

Tutorial and Working Group Meeting: Order-theoretic Aspects of Epidemiology
http://dimacs.rutgers.edu/Workshops/WGOrder

DIMACS Epidemiology Minisymposium
http://dimacs.rutgers.edu/Workshops/Epimini

Workshop: Evolutionary Considerations in Vaccine Use
http://dimacs.rutgers.edu/Workshops/VaccineUse/

Special Focus Seminar Series 2004-2005
http://dimacs.rutgers.edu/SpecialYears/2002_Epid/episeminars.html

Reports

DIMACS Working Group Report on Data Mining and Epidemiology, March 18-19, 2004
Report Authors: James Abello and Graham Cormode, DIMACS
In preparation.
http://dimacs.rutgers.edu/Workshops/WGDataMining/material/ (password protected)

DIMACS Working Group on Adverse Event/Disease Reporting, Surveillance, and Analysis II, Group discussion on methods for monitoring multiple data streams to detect incidence of disease,

Report Author: David Madigan, Rutgers University
http://dimacs.rutgers.edu/Workshops/AdverseEvent2/

Contributions

Contributions within Discipline

The “discipline” is by definition a combination of disciplines. Many of the results described in the Activities and Findings section of this report illustrate this combination of disciplines, in particular the application of methods of computer science, statistics, and mathematics to problems of epidemiology. Some examples are Dushoff, Plotkin, Levin, and Earn’s work on the seasonality of influenza, Hoover and Xia’s work on developing statistical methods for rare events, Jacob and Mikler’s work at CERL on modeling an influenza outbreak, Stoto, Fricker, Jain, Davies-Cole, Glymph, Kidane, Lum, Jones Dehan, and Yuan’s work on evaluating statitistical methods for syndromic surveillance, and Pogel and Ozonoff’s work on applying lattice theory to epidemiology. Several participants and groups of participants, including Van Den Driessche, Arino, Jordan, Hartley, Miller, Jacob, Mikler, Franc, and Peyrard have created other models of epidemics. Other groups, including Jacob, Mikler, Eidelson, Lustick, Hayrapetyan, Kempe, Pal, Svitkina, Ng, and Raff have studied various vaccination strategies. These are all described in more detail earlier in this report.

Introducing people to this combination of disciplines has been a key goal and a key accomplishment of this project. In addition, the following is a selection of comments we received from participants in our project activities.

“I have participated in three different DIMACS workshops on epidemiology and disease surveillance. I have found them all very stimulating, where new ideas and problems are discussed in a very open and collegial atmosphere. Of special importance is the interaction, interplay and discussions between public health officials and epidemiologists on the one hand and biostatisticians and mathematical statisticians on the other. Neither group has the capacity to solve important methodological public health problems alone, the former lacking the technical skills in statistics and the latter group lacking the deep knowledge of the substantive issues. In addition to the many fruitful interactions at the workshops themselves, I am currently involved in a couple of collaborations that were triggered by first meeting at a DIMACS workshop. I am therefore personally very grateful for the superb work you are doing.” Martin Kulldorff, Harvard Medical School and Harvard Pilgrim Health Care

“I had the opportunity to interact with the activities of the Special Focus on Computational and Mathematical Epidemiology as a participant in the DIMACS Working Group on Methodologies for Comparing Vaccination Strategies. This experience was extremely gratifying to me since I had the opportunity to present my work and receive the feedback from the participants. The work presented at the meeting has recently been accepted at the Journal of Biological Systems. This event also allowed me to establish new collaborative work with many of the participants. With Dr. Abba Gumel, I am now organizing a new workshop on Evolutionary Considerations in Vaccine Use, an offshoot of the activity I had attended in the previous year. This experience is being very fruitful and opening several new opportunities. In addition to the new research projects I am pursuing with the members of the organizing committee, this interaction also allowed for the training of one Brazilian student who will be undertaking her doctoral program under the supervision of Dr. Alison Galvani at Yale University. Dr. Galvani and I are now discussing the possibilities of submitting a research proposal to NIH addressing the impact of the AIDS control program in Brazil that grants free and universal access to HAART therapy for the infected patients. Finally, Dr. Abba Gumel, Dr. Elamin Elbasha and I are now finalizing a manuscript where we discuss the trade off between virus resistance and virus fitness. As an important outcome of this work, we intend to propose new AIDS control programs based on universal drug distribution and vaccination such that virus resistance is kept under control." Claudio Struchiner, Oswaldo Cruz Foundation

"I have attended three Computational and Mathematical EpidemiologySpecial Focus conferences and have found them very useful. As the director of syndromic surveillance operations at the New York City Department of Health and Mental Hygiene, I appreciate the opportunity these conferences give me to step away from the day-to-day activities of the office and listen to and interact with scientists from a wide range of disciplines. The last conference I attended grappled with the thorny problem of multiple data-stream analysis, which we have been exploring in NYC as a way of increasing the specificity of syndromic signals." Rick Heffernan, Bureau of Communicable Disease, New York City Department of Health and Mental Hygiene

"Through the Computational and Mathematical Epidemiology Special Focus, I have been privileged to develop collaborations with David Ozonoff and James Abello. I have also been able to spend time with Mel Janowitz learning about his abstract clustering theory, and this will have an influence on my future work. Over the course of my twice-yearly visits to DIMACS since 2003, I have been exposed to many new ideas in discrete mathematics and also applications of mathematics, especially epidemiology and data mining. This exposure has greatly enriched my work at the Physical Science Laboratory on the NMSU campus, which is primarily focused on the use of concept lattices in data analysis of the output of competitive multi-agent simulations." Alex Pogel, New Mexico State University

"The special year on Epidemiology was very important to my scientific activity. A year ago I started working on Real network. Epidemiology is one of the fields in which I can apply my knowledge and help others. In the past year I learned a lot about possible applications of networks and I got many ideas. I started new collaborations with people from other area of science as computer science and biology. I would never meet those scientists otherwise. I am currently working on some problems with 3 new collaborators that I met at the activity of the special year. I like a lot the workshops in which one can get familiar with a new field. For example, the biological network workshop gave me a good start on important biological terms. I also had the opportunity to invite my collaborator to give a lecture on the spread of disease in social networks. As epidemiology involved many fields of science, meeting scientists from all fields of science is very important and rewarding to all participates." Sara Soffer, Rutgers University.

"The recent meeting on order-theory also spawned a new collaboration with Petko Valchev (U. Montreal), using new lattice diagram visualization and algorithms to apply to the epidemiological work. We have also begun discussions with Dr. T. Webster and A. Ozonoff about Measurement Theory problems in epidemiology (prompted by a presentation by Fred Roberts at the Order-Theoretic Working Group.)." David Ozonoff, Boston University School of Public Health

"My participation in DIMACS provided new connection with people in CDC and WHO. And communications with them thereafter shed new light in my research and deepened the understanding of the problem. I really appreciate the opportunity to participate in the DIMACS meeting (of comparing vaccination strategies working group)." Linda Gao, North Central College, Illinois

"I attended the DIMACS Working Group on Data Mining and Epidemiology. March 18-19, 2004, and presented a paper: The Exploration of Spatial Data Mining (and Mind Mining) to model the risk of Emerald ash borer (EAB) (***Agrilus planipennis***) and its likely spread from current areas of infestation. Participation in the workshop has contributed significantly to the direction of my research." W.D. Smith, USDA Forest Service

"DIMACS has facilitated interactions and collaborations with many people, on a range of issues including influenza genomics, influenza spread, and the spread of antibiotic resistance." Jonathan Dushoff, Princeton University

Contributions To Other Disciplines

Since the "discipline" is inherently multidisciplinary, there is no separate entry in this section.

Contributions Beyond Of Science And Engineering

Several of the outcomes of the special focus have been immediately applicable to real world problems. See, for example, the work of Jacob and Mikler on vaccination clinics in Denton County, Texas, Ortega's cost-benefit analysis of introducing a rotavirus vaccine in Egypt, Struchiner and Galvani's proposal for new AIDS control programs in Brazil, and Smith's work on protecting exotic trees. Additional outcomes are describe by participants as follows:

"The DIMACS conference/event I attended featured time-series analysis (Poisson) and cluster analysis. We have since applied both to asthma studies relating to air toxic compounds and allergens." Steve Miller NJDHSS, Hazardous Site Health Evaluation Program

"The American Association of Poison Control Centers was represented by Bill Watson, Associate Director, Toxicosurviellance, at the DIMACS Tutorial on Statistical and Other Analytical Health Surveillance Methods meeting June 17 to 20, 2003. The information obtained during the presentation, and contacts made there were crucial to understanding the global issues involved in the initiation of the AAPCC Toxicosurveillance program, with receives approximately 6000 new human exposures to potentially toxic substance in a continuous stream daily. It introduced poison centers and the Toxic Exposure Surveillance System as a possible source of useful health surveillance data to the participants, and allowed the AAPCC to initiate SatScan evaluations, as well as more effectively collaborate with CDC surveillance and epidemiology groups." Wm A Watson, Associate Director, Toxicosurveillance, American Association of Poison Control Centers

"In addition to what I've reported in earlier years, , I would add that I have teamed up with Heinbokel and Potash, whom I met at DIMACS, and we are involved in some very interesting modeling work relevant to intelligence problems apart from epidemics of disease. We met at the DIMACS conference." Ian Lustick, University of Pennsylvania

The participation of Tom Jacob and Armin R. Mikler in the early DIMACS workshops and group meetings of this special focus led to several projects at the Computational Epidemiology Research Laboratory (CERL) at the University of North Texas. These projects are described in the section on findings. We should mention in addition their work on mobile vaccination clinics:

> . The recent demand for influenza vaccination has resulted in ad-hoc mass-vaccination clinics held by public health departments throughout the nation. Two such clinics were held in Denton County, Texas on October 15th and 22nd, 2004. These events marked an opportunity to test strategies for clinic setup, crowd control, flow control, and logistics, which had been developed in the context of bioterrorism response. In previous years, most of the influenza vaccinations were conducted at the public health department during regular office hours, throughout the flu season. The lack of adequate supplies of vaccine nationwide, together with the

> specific timeframe during which vaccination could be obtained has resulted in a scenario that resembles that of a mass vaccination during a disaster situation (e.g. smallpox outbreak). Officials from the Denton County Health Department (DCHD), together with students and faculty from CERL, used this opportunity to observe and record problems, delays, and other adverse effects within the clinics. Faculty and students at CERL have been collaborating with DCHD for one year. The primary purpose of this collaboration is the design of outbreak models that facilitate the allocation of public health resources during a public health emergency. In response to what was learnt during the recent vaccination event, Jacob and Mikler believe it is imperative to investigate additional strategies for expedient mass vaccination. At the clinic it was found that the elderly or mobility impaired individuals could not stand in line for hours, which required the use of 'mobile nurses' to attend to individuals who were unable to tolerate the wait. Considering that the recently conducted clinics were restricted to young children, elderly, and immune deficient individuals, waiting lines of several hundred are expected to increase to several thousands in the event of a disaster that triggers mass vaccination. In order to address some of the problems observed, Jacob and Mikler plan to model and analyze the use of mobile distributed vaccination clinics as an alternative vaccination strategy in Denton County.

Contributions To Human Resources Development

Many graduate students, undergraduates, and several postdocs participate in the program. One Rutgers student, Paul Raff (Mathematics), has been heavily involved this year. Other local graduate students and many non-local students were also involved as visitors and workshop/working group attendees. Patrick De Leenheer, a post doc at DIMACS, mentored undergraduate Andrew Hodges on his project "Mutations and treatment of HIV" as part of the 2004 DIMACS REU program. The project has three primary postdocs this year, Tanya Berger-Wolf, Kah Loon Ng, and Martin Pal, as well as participation by visitors James Abello and Michael Capalbo and many visiting postdocs. More senior people were also heavily influenced by the project, being exposed to new directions of research and changing their fields as a result. The impact on the careers of the students and postdocs is illustrated by a few examples.

"My Student Qi Xia finished her Ph.D. thesis this year and will be working as a research biostatistician at Genetech starting this June. This Title is "Exact Methods Applied to Group Sequential and Other Stratified Comparative Poisson Designs." Donald Hoover, Rutgers University

"The DIMACS workshop was very helpful in allowing me and a graduate student to meet other researchers and to discuss state-of-the-art bio-surveillance detection methods. The talks and discussions at the Workshop reinforced the direction we are taking in investigating Bayesian-network-based methods for bio-surveillance. Overall, I found the Workshop to be very useful." Greg Cooper, University of Pittsburgh

"At the time of this conference, I had not yet chosen a topic for my dissertation. At the DIMACS program on Mathematical Modeling of Infectious Diseases, I learned more about which diseases were in need of more research, and which diseases were in the public eye. I chose to work on rotavirus because of a presentation given by Ms. Wolfson, a biostatistician from the WHO. Currently I am working in Cairo, Egypt with the US Naval Medical Research Unit #3 on a cost-benefit analysis of introducing a rotavirus vaccine in Egypt. A cost-effectiveness analysis is in the works. Thank you for offering such wonderful programs at DIMACS!" Omayra Ortega, University of Iowa

"I participated in the workshop on reticulate evolution last summer. As a result, I have a new collaboration with Tandy Warnow and Randy Linder, in an area of research that is somewhat novel for me. It is too early to say much about it, but it stems from the half-baked presentation I gave at the workshop. We have also convinced a math grad student here at Harvard to work on this (Erick Matsen)

and involved Steve Evans (UC Berkeley, stats). As a result of having met Joel Cohen at that DIMACS workshop, I traveled to the Rockefeller University this January and presented some of my research."
John R, Wakeley, Harvard

"I am currently working on a paper with Paul Raff (graduate student at Rutgers). This is regarding the firefighter problem. He got interested in the problem after coming to the talk I gave during the Epi seminar and we have been working together for a while now. Also note that he has applied for the DIMACS summer grad student assistantship award … with me as the recommender and we will be continuing our work on the firefighter problem… We plan to formally start writing up our results in June and submit it by the end of June." Kah Loon Ng

"I have been working on a project concerning the firefighter problem with Kah Loon Ng. Specifically, we are working to achieve tight bounds on firefighters needed to contain a fire on the infinite-dimensional square grid. We are planning on submitting a paper by the end of June. DIMACS as a whole and the work on epidemiology has been a nice breath of fresh air during my first year of grad school, which has mainly been burdened by coursework. It's amazing (and somewhat intimidating at the same time!) to be directly involved in so much work and dialogue going on around the world, centered through DIMACS. Now that the first year at grad school is over, I am very excited about the upcoming time with DIMACS, as I plan to be able to spend more time on my research. The work with Kah Loon will be continuing, and I hope to be involved in more projects with DIMACS. The best thing I like about the work at DIMACS is that it's always known what the application of the work will be - it's always known what the utility of the work is, which is not always the case in pure academic research." Paul Raff, graduate student, Rutgers University

"In one of the discussion groups I presented the question of modeling the importation of and long-distance transport of nursery stock and other commodities that are the major pathway for the introduction of exotic insects and diseases to the U.S. Since the Workshop I have obtained funding to obtain a post-doctorial research associate (Frank Koch, NCSU) to address this problem." W.D. Smith, USDA Forest Service

"Thanks to the Working Group on Disease and Adverse Event Surveillance, I have started work on new statistical methods for post-marketing surveillance for drug safety. I have started collaborating with Manfred Hauben, worldwide head of drug safety for Pfizer. Next year we will jointly supervise Ivan Zorych, a DIMACS postdoc. We have completed one paper (submitted) and have others in the pipeline. I am also supervising a PhD student (Aimin Feng) working in this area. I have been teaching a graduate "data mining" class which uses applications stimulated by the Disease and Adverse Event surveillance Working Group." David Madigan, Rutgers University

In addition, the following graduate students have undertaken small research projects under support of the special focus. Their work is described under their names in the list of project participants.

German Enciso, RU Mathematics, Winter 04/05
"Systems under positive feedback: Multistability and a reduction theorem"

Pai-Hsi Huang, RU CS, Winter 04/05
"Profile hidden Markov Chains"

David Millman, RU
Working on a cancer registry modeling project with Dr. James Abello and Dr. Ilya Muchnik.

Paul Raff, RU Mathematics, Winter 04/05

> "Work with and extend the work done by Donald Burke and others in their paper 'Toward a Containment Strategy for Smallpox Bioterror: An Individual-Based Computational Approach' and their corresponding computer program."
>
> Liming Wang, RU Mathematics, Winter 04/05
> "Cyclin B-Cdc2/Wee1 system, which underlies the cell cycle, functions as a bistable switch"
>
> Igor Zverovich, RU RUTCOR, Winter 04/05
> "Generalized stable set problem with application to the defense against bioterrorism"

Long-Term Special Focus visits have also played a role in people's careers.

> Long-term visitor James Abello has been working on a cancer registry modeling project with Dr. Ilya Muchnik and graduate student David Millman and on random graph models of spread of disease with Michael Capalbo. This has been a completely new line of research for Dr. Abello and he has been involved with a student and a junior researcher in the process.
>
> Long-term visitor Michael Capalbo has been working on random graph models of spread of disease with James Abello. Dr. Cabalbo is a junior researcher (recent Ph.D.) and this long-term visit led to a totally new direction of research for him.

Postdocs as part of the Special Focus:

> 2003/2004
>
> Graham Cormode, working on data mining and epidemiology, primarily with Dr. Muthukrishnan
> Patrick DeLeenheer, working on predator-prey models and chemostat models.
>
> 2004/2005
>
> Kah Loon Ng, working on vaccination strategies
> Tanya Berger-Wolf, working on kinship relationships in populations
> Martin Pal, working on combinatorial optimization and epidemiology

As described in the annual reports for years one and two, mathematician Alex Pogel (New Mexico State University and epidemiologist David Ozonoff (Boston University) began a collaboration through DIMACS that introduced a generalization of the notion of contingency table and they view the concept lattice as a minimal carrier of all the information in this generalized contingency table. Their intention with this work is to help epidemiologists form hypotheses regarding data, particularly in the early stages of an investigation and, most importantly, before any standard low-dimensional statistical tools are applied. Their further development of this idea is leading to new research directions in epidemiology and new methods of data analysis. The first exposition of this research was submitted and accepted for publication in the DIMACS Special Volume on Data Mining and Epidemiology in 2004. This new and fundamentally interdisciplinary collaboration has led to a major change in Pogel's work. It has now led to their receipt of two research grants, first a one-year NIEHS grant, and second a five-year EPA Superfund grant (part of the larger BU grant). Work will continue and they expect the final outcome to be open source software tools available to the epidemiology community, to provide new hypothesis generation abilities. More to the point, this is just one of a variety of such examples that we can point to as truly interdisciplinary, career-changing outcomes of the special focus.