DIMACS Center
Rutgers University

Special Focus on Computational and Mathematical Epidemiology

Annual Report

May 2006

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; Co-Organizer of Working Group on BioSurveillance Data Monitoring and Information Exchange.

S. Muthukrishnan, Rutgers 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.

Visitors as part of the Special Focus:

Michael Bender, working with Martin Farach-Colton on algorithms

Christian Tominski, working with James Abello on models of spread of disease

Kenton Morgan, working with Ilya Muchnik on models of spread of disease

James Abello, working on a cancer registry modeling project with Ilya Muchnik and graduate student David Millman and on graph-theoretical models of spread of disease with Michael Capalbo; Co-Organizer of Tutorial on Data Mining and Epidemiology

Michael Capalbo, working on random graph models of spread of disease with James Abello

Kah Loon Ng, working on vaccination strategies

Nina Fefferman, working on social structure effects on spread of disease

Postdocs as part of the Special Focus:

Ivan Zorych, developing drug monitoring system

Other Participants:

Sunetra Gupta, Oxford University, Member of Organizing Committee

David Krakauer, Santa Fe Institute, Member of Organizing Committee

Simon Levin, Princeton University, Member of Organizing Committee; 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 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

Burton Singer, Princeton University, Member of Organizing Committee

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

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

Workshop: Economic Epidemiology
October 24 - 25, 2005

Organizers:
Dave Smith, NIH
Ramanan Laxminarayan, Resources for the Future

Workshop: The Epidemiology and Evolution of Influenza
January 25 - 27, 2006

Organizers:
Catherine Macken
Alan Perelson, Los Alamos National Labs

Working Group Meeting: Adverse Event/Disease Reporting, Surveillance and Analysis IV
Fourth Meeting, February 10, 2006

Organizers:
Donald Hoover, Rutgers University
David Madigan, Rutgers University
Henry Rolka, CDC

Working Group Meeting: BioSurveillance Data Monitoring and Information Exchange
February 22 - 24, 2006

Organizers:
David Madigan, Rutgers University
Colleen Martin, CDC
Henry Rolka, CDC

Tutorial on Data Mining and Epidemiology
March 23 - 24, 2006

Organizers:
James Abello, DIMACS
Graham Cormode, Bell Laboratories

Workshop: Combinatorial Group Testing
May 17 - 19, 2006

Organizers:
Ding-zhu Du, University of Texas-Dallas
Frank Hwang, Chiatong University

Tutorial: Phylogenetic Trees and Rapidly Evolving Pathogens
June 19 - 20, 2006

Organizer:
Katherine St. John, The City University of New York

Workshop: Phylogenetic Trees and Rapidly Evolving Pathogens
June 21- 22, 2006

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

Working Group Meeting: Phylogenetic Trees and Rapidly Evolving Pathogens II
Second Meeting: June 23, 2006

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

Other Collaborators

Graham Cormode, Bell Laboratories, Co-Organizer of Tutorial on Data Mining and Epidemiology

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

Ding-zhu Du, University of Texas-Dallas, Co-Organizer of Workshop: Combinatorial Group Testing

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

Frank Hwang, Chiatong University, Co-Organizer of Workshop: Combinatorial Group Testing

Ramanan Laxminarayan, Resources for the Future, Co-Organizer, Workshop: Economic Epidemiology

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

Catherine Macken, Los Alamos National Labs, Co-Organizer, Workshop: The Epidemiology and Evolution of Influenza

Colleen Martin, CDC, Co-Organizer of Working Group on BioSurveillance Data Monitoring and Information Exchange

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

Alan Perelson, Los Alamos National Labs, Co-Organizer, Workshop: The Epidemiology and Evolution of Influenza

Allen Rodrigo, University Of Auckland, Co-Organizer, Workshop: Phylogenetic Trees and Rapidly Evolving Pathogens, Co-Organizer of Working Group Phylogenetic Trees and Rapidly Evolving Pathogens

Henry Rolka, CDC, Co-Organizer of Working Group on Adverse Event/Disease Reporting, Surveillance and Analysis; Co-Organizer of Working Group on BioSurveillance Data Monitoring and Information Exchange

Dave Smith, NIH, Co-Organizer, Workshop: Economic Epidemiology

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

Katherine St. John, The City University of New York, Organizer of Tutorial: Phylogenetic Trees and Rapidly Evolving Pathogens

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, organization and 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 a five-year special focus, which started 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 the promise for making progress on modern epidemiological problems and should take both fields of epidemiology and mathematics/computer science in new and fruitful directions. Mathematical and computational methods seem especially relevant in light of recent modeling approaches to emerging infectious diseases such as the vector-borne diseases from West Nile virus, Eastern equine encephalitis virus and Borrelia burgdorfei (Lyme disease); the spread of "mad cow" disease (transmissible spongiform encephalopathy; and HIV/AIDS. Control measures for these diseases often have counter-intuitive consequences only revealed after sophisticated mathematical analysis. Similar advances as a result of applications of mathematical and computational modeling have not been as evident in the area of chronic disease epidemiology, although work of considerable promise is being done, for example on modeling of the progression of cancer. In this special focus, we will consider both infectious and non-infectious diseases, and we will explore mathematical and computational approaches to animal and plant diseases as well as to human diseases.

Workshops, Working Group Meetings and Tutorials During This Reporting Period

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: 48

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. This workshop examined general evolution-related questions for any disease for which there is a vaccine (or hope for one).

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

Workshop: Economic Epidemiology

Dates: October 24 - 25, 2005
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Dave Smith, NIH; and Ramanan Laxminarayan, Resources for the Future
Attendance: 29

The emergence and spread of resistance to antimicrobial agents is a complex interplay between economics, human behavior, and disease ecology. Mathematical models can help make sense of the complexity and vastly improve our understanding of the interplay between these factors. The DIMACS workshop on economic epidemiology is the first of a series of consultations between economists and biologists working on issues of infectious diseases modeling and policy. A special focus of this first workshop was the management of antimicrobial resistance. Economic incentives play an important role in determining antibiotic use, infection control and the evolution of resistance. At first glance, one notes that those who use or prescribe antibiotics have few or no incentives to consider the impact of their decisions on the rest of society. On further reflection, it is evident that this problem of missing incentives extends to institutions such as hospitals and the pharmaceutical industry. Hospitals operating in the vicinity of many other medical care institutions that share patients may have fewer incentives to invest in hospital infection control to manage resistance if the benefits of their actions mainly accrue to other institutions. Drug firms that are involved in the manufacture of antibiotics similarly may fail to consider the impact of their aggressive antibiotic marketing campaigns on cross-resistance with other antibiotics that are being used. Understanding the role of incentives in the evolution of drug resistance, and the implications for the management of resistance formed the agenda of the workshop. The broad purpose of the workshop was to encourage greater application of economic intuition and analytical methods to mathematical models of disease evolution.

Workshop: The Epidemiology and Evolution of Influenza

Dates: January 25 - 27, 2006
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Catherine Macken and Alan Perelson, Los Alamos National Labs
Attendance: 79

This workshop explored the epidemiology and evolution of influenza. The persistence of influenza depends on its ability to evolve so that new strains and subtypes of the virus appear and old ones reappear. This constant evolution means that vaccines need to be updated frequently and that resistance to drug therapies can easily arise. The workshop brought together public health practitioners, immunologists, epidemiologists, evolutionary biologists, mathematicians, statisticians, and computer scientists to explore the evolution and dynamics of influenza. Among the issues explored were the causes and consequences of

patterns of immunological cross-reactivity, and the interactions with drug treatment and vaccination strategies. In addition, the applicability of SIR and agent based models to predict the spread of influenza, and the means of dealing with and planning for an influenza pandemic, were discussed.

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

Dates: February 10, 2006
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Donald Hoover, and David Madigan, Rutgers University; Henry Rolka, CDC
Attendance: 15

This working group is exploring new directions for statistical methods in drug safety. This was the fourth meeting of the group.

Working Group Meeting: BioSurveillance Data Monitoring and Information Exchange

Dates: February 22 - 24, 2006
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: David Madigan, Rutgers University; Colleen Martin and Henry Rolka, CDC
Attendance: 51

Monitoring biosurveillance data in modern public health surveillance systems involves many skill sets, including epidemiology, statistics, and informatics, as well as the necessity to make rapid decisions based on uncertain information. Modern surveillance systems include health-related data that are transmitted on a near real-time or real-time basis. The data is characteristically noisy and/or incomplete. Conclusions from such data invariably have a degree of uncertainty, which is often difficult to assess. Public health officials responsible for monitoring and responding to these data must examine, analyze, report, escalate, and obtain further information in order to make rapid decisions regarding how to react to information from these systems. An added complication involves the use of multiple systems (such as BioSense, ESSENCE, RODS, EARS, and others) to achieve this goal; the consolidation of information across systems and among jurisdictions is an essential and sometimes difficult task.

The intended audience for this workshop included public health officials responsible for monitoring and responding to biosurveillance data. The focus was on issues shared by users of various systems, as opposed to focusing on one system or the other. Users of these systems nationwide have expressed interest in building a peer group for information sharing, communication, and problem solving. Short sessions have been held at the Public Health Information Network (PHIN), National Syndromic Surveillance, and Global Emerging Infection Surveillance (GEIS) Conferences to address these needs and have been well-received by biosurveillance data monitors. The intent of this workshop was to build upon these sessions, to discuss current needs and issues, and to discuss future goals and solutions.
This workshop provided an opportunity for shared dialogue and learning between those practicing monitoring and analysis of biosurveillance data. Workshop goals included the following:

- To increase communication and enhance relationship-building among biosurveillance data monitoring and analytic staff nationwide and at all levels of public health
- To facilitate information exchange regarding biosurveillance monitoring and data analysis, including data anomaly detection methods, monitoring and anomaly characterization, follow-up and communication protocols, experiences and lessons learned, and system and monitor needs
- To frame a biosurveillance data monitors working group

Major topics addressed in this workshop included sharing experiences and lessons learned using various systems, analytical methods, monitoring methods and consolidation of information, and anomaly

characterization and follow-up. Presentations and breakout sessions framed the issues and allowed for ample interaction and networking.

Tutorial on Data Mining and Epidemiology

Dates: March 23 - 24, 2006
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: James Abello, DIMACS; and Graham Cormode, Bell Laboratories
Attendance: 59

Data Mining is now a staple part of Computer Science, and has been applied in a wide variety of different areas. It covers a diverse set of topics from algorithms, statistics and discrete mathematics, with the general goal of identifying patterns in data in order to draw inferences and make predictions. This tutorial brought together experts from Data Mining to introduce the key ideas and techniques from:

- Probability, Decision Trees and Bayesian Statistics
- Machine Learning, Classifiers and Boosting
- Data Stream Analysis and Clustering
- Graph Mining
- Applications to Biology and Epidemiology

The goal was to allow people with little or no knowledge of data mining to understand the basic techniques, and get a flavor of the general methodology and style of results. This tutorial was aimed to be of interest to researchers wishing to work in data mining, and also to researchers from outside computer science who wish to understand these methods in order to apply them. The tutorial included short talks on applications to problems in epidemiology and biology in order to put the general techniques described into perspective.

Workshop: Combinatorial Group Testing

Dates: May 17 - 19, 2006
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Ding-zhu Du, University of Texas-Dallas; and Frank Hwang, Chiatong University
Attendance: Registration is still open for this workshop

This workshop will investigate modern combinatorial and algorithmic methods of group testing, with emphasis on connections to coding theory and combinatorial design. To identify all positive cases in a large population of items, group-testing proceeds by grouping the items into subsets, testing if a subset contains at least one positive item, and then identifying all positive items through iteration of group tests. The theory of group testing arose from the idea (never implemented) of testing millions of World War II military draftees for syphilis and it is very relevant to schemes for large-scale blood testing for viruses such as HIV. Group testing also arises in connection with the mapping of genomes. Here, we have a long list of molecular sequences, form a library of subsequences (clones), and test whether or not a particular sequence (a probe) appears in the library by testing to see in which clones it appears. Because clone libraries can be huge, we do this by pooling the clones into groups. This workshop will investigate both epidemiological and molecular biological applications of group testing.

Tutorial: Phylogenetic Trees and Rapidly Evolving Pathogens

Dates: June 19 - 20, 2006
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Katherine St. John, The City University of New York

Attendance: Registration is still open for this tutorial

Phylogenies, or evolutionary histories, are used throughout biology. In addition to the study of taxonomy, they are used widely to do such things as design drugs, align biomolecular sequences, and to understand rapidly evolving diseases, such as HIV. This tutorial is an introduction to computational phylogenetics and its applications to real-world problems. The topics include standard phylogenetic reconstruction methods and concepts, as well as advanced topics needed to understand the application of phylogeny to rapidly evolving diseases.

The tutorial is aimed at researchers interested in phylogenetics research and their applications. The goal is to understand the standard reconstruction methods and concepts well enough to understand cutting-edge research in the field (i.e. the workshop following the tutorial). The intended audience for this tutorial are those with a graduate background in computer science, discrete mathematics, or biology. No biological or algorithms background is assumed but knowledge of one will be helpful.

The tutorial includes sessions on the necessary biological and algorithmic topics, standard phylogenetic reconstruction methods, and the use of statistical analysis in the field. "Hands-on" laboratory sessions on using these methods and tools are included.

Immediately following this tutorial, DIMACS is running a workshop on rapidly evolving diseases. This tutorial has been coordinated with the workshop to make it possible for the non-specialist to attend and understand most of the talks in the workshop.

Workshop: Phylogenetic Trees and Rapidly Evolving Pathogens

Dates: June 21- 22, 2006
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Allen Rodrigo, University Of Auckland and Mike Steel, University of Canterbury
Attendance: Registration is still open for this workshop

This workshop 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 topics of this workshop. 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 "quasispecies" 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.

Working Group Meeting: Phylogenetic Trees and Rapidly Evolving Pathogens II
Dates: Second Meeting: June 23, 2006
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Allen Rodrigo, University Of Auckland and Mike Steel, University of Canterbury
Attendance: Registration is still open for this working group meeting

This meeting will convene a small group of researchers to explore in depth the topics described in the above workshop and lay out an agenda for future research.

DIMACS Seminar in Quantitative Biology and Epidemiology

Nina Fefferman and James Abello, Special Focus visitors, organized the DIMACS Computational amd Mathematical Epidemiology seminar during the 2005-2006 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/seminars05-06.html

Disease Signatures
Monday, September 19, 2005
Speaker: Nina Fefferman, DIMACS and Tufts University

The Firefighter Problem on d-dimensional Grids and Hartke's Conjecture
Monday, October 3, 2005
Speaker: Michael Capalbo, DIMACS

Termites in the Nation's Service (part 2): More Details than You Wanted
Monday, October 17, 2005
Speaker: Nina Fefferman, DIMACS and Tufts University

GeneralizedFirefighting on the 2 Dimensional Infinite Grid and Centrality Measures in Social Networks
Monday, October 31, 2005
Speaker: Kah Loon Ng, DIMACS

Threshold Phenomena in Simple Epidemics with Recovery
Monday, November 7, 2005

Speaker: Regina Dolgoarshinnykh, Columbia University

Finding and Interpreting Local Models in Analysis of Epidemiological Data
Monday, November 14, 2005
Speaker: Dmitriy Fradkin, Rutgers University

Algorithmic Problems in Epidemiology
Monday, November 28, 2005
Speaker: V.S. Anil Kumar, Virginia Bioinformatics Institute, Virginia Tech

Optimization-Based Data Mining for Epilepsy Research
Monday, December 5, 2005
Speaker: Wanpracha Art Chaovalitwongse, Rutgers University

The Influence of Economic Incentives on Hospital Infection Control: An Application of Economic-Epidemiology
Monday, December 12, 2005
Speaker: Ramanan Laxminarayan, Resources for the Future

Different Scales of BioDefense - Can Societies be Both Safe and Efficient?
Monday, February 13, 2006
Speaker: Nina Fefferman, DIMACS and Tufts University

Graph Maps
Monday, February 27, 2006
Speaker: James Abello, DIMACS and Ask.com

Using Cluster Analysis to Determine the Influence of Epidemiological Features on Medical Status of Lung Cancer Patients
Monday, March 13, 2006
Speaker: Dmitriy Fradkin, Ask.com

Statistical Modeling for Prospective Surveillance: Paradigm, Approach, and Methods
Monday, March 20, 2006
Speaker: Al Ozonoff and Paola Sebastiani, Boston University School of Public Health

A Bayesian Approach to Multistate Life Tables for Use in Social Epidemiology
Monday, March 27, 2006
Speaker: Scott Lynch, Princeton University

Early Detection of Disease Outbreaks
Monday, April 10, 2006
Speaker: Martin Kulldorff, Harvard University

The Simpson-Elsasser-Wolfram (SEW) Framework for Modeling the Living Cell
Monday, April 17, 2006
Speaker: Sung Ji, Rutgers University

Local Likelihood Bayesian Cluster Modeling for Small Area Health Data
Monday, April 24, 2006
Speaker: Andrew Lawson, University of South Carolina

Findings

Impact of Social Structure on Disease Seasonality

Seasonal patterns in infectious diseases have traditionally been incorporated into epidemiological modeling via a variety of mechanisms of external forcing. More recent investigations have examined the possible roles of stochasticity in generating these observed oscillations. DIMACS visitor Nina Fefferman colloborated with Elena Naumova, Tufts Univ. School of Medicine, in using mathematical models and simulations to demonstrate that seasonal patterns can arise solely from the differences in social interactions among etiologically distinct subpopulations. Additionally, these oscillations behave differently depending on the subpopulation into which initial primary pathogen exposure is introduced. Their results demonstrate how societal organization and human behavior can drastically affect the dynamics of infectious disease in a population in a way previously unexplored.

The Influence of Individual Social Behavior on the Centrality of a Network

A number of different measures of centrality have been employed in the study of social networks. These metrics have primarily been employed for the purposes of analyzing 'emerged' (existing) networks. DIMACS post doc Kah Loon Ng and DIMACS visitor Nina Fefferman examined the influence of individual social behavior on the stability of the centrality of a network as it emerges. To do this, they simulated a dynamic directed graph over time, forcing individuals to change their affiliations (outgoing edges) in each computational iteration, with the 'goal' of increasing their own individual node centrality. Individuals were assumed to have knowledge of the centrality measures of their current neighbors, but no other information, and were thus forced to form affiliations at random with others of unknown centrality. Ng and Fefferman examined the influence of these decision processes on the over-all stability of the network's centrality under three different centrality measures: degree centrality, closeness centrality and betweeness centrality. They demonstrated that the behavior of the network's total centrality, over time, depended on the underlying centrality measure employed. Based on these differences in stability within behaviorally homogeneous networks, they then investigated the stability of heterogeneous networks, where all centrality measures are represented, but individuals employ only one on which to base their affiliation choices. Ng and Fefferman examined the stability of these heterogeneous networks, under different starting conditions (e.g. proportion of individuals using degree vs. closeness centrality), to evaluate the effect of behavioral heterogeneity on long-term total network stability. Their work led them to hypothesize that one particular centrality measure should emerge as the most successful strategy for individuals, but which measure achieves that success will depend on the starting conditions of the network. Very few studies have focused on the comparative effects of different centrality measures in dynamic networks. Ng and Fefferman believe that this type of comparison is not only interesting by itself, but lays the groundwork for questions of how centrality-related behavioral choice can influence the evolution of social networks.

The Effect of Social Organization on Disease Threats

While visiting DIMACS, Nina Fefferman, Tufts University began collaborating with Sam Behers, University of Illinois at Urbana-Champaign, to develop mathematical models of social insect species to investigate the impact of different social organizations on the ability of the species to withstand disease threats. Societal structure and social organization shape the types of social interactions expected within a population. These interactions, in turn, are the means by which infectious diseases are spread. The question then arises: Are there ways to structure societies so as to maximize robustness to disease, for example, by minimizing the numbers of infections, or deaths, or economic costs, or the breakdown of

societal infrastructure, or even by minimizing the long-term effect to population growth? These problems (and their potential solutions) occur on a variety of scales: individual, neighborhood, company, local, national, international - each probably leads to a different robustness goal. Fefferman and Behers examined a few models that focus on different scales of disease spread and looked at their society-level implications. The results from these investigations are in preparation for publication and were presented by Fefferman in an invited talk to the Department of Entomology and Center for Infectious Disease Dynamics (CIDD) at Penn State University. After meeting at a DIMACS conference on the economics of epidemiology, Ramanan Laxminarayan, Resources for the Future, began collaborating with Fefferman and Behers on extending this research to investigate the different impacts of disease spread based on different organizational structures of large businesses.

Higher in Utero and Perinatal HIV-Infection Risk in Girls than Boys

Robert J. Biggar, National Cancer Institute, Taha Taha, The Johns Hopkins Bloomberg School of Hygiene and Public Health, Baltimore, Donald R. Hoover, Statistics, Rutgers University, F.Yellin, National Cancer Institute, N. Kumwenda, COM and John Hopkins University, and Robert Lumb Broadhead, COM, analyzed mother-to-child HIV transmission rates by sex and exposure time for babies born to HIV-infected, untreated African women. Data were analyzed from 2 independent studies done in Malawi during the 1990s. Infections were established by polymerase chain reaction on blood samples. Odds ratios (ORs) for transmission were examined by period at risk: in utero (infected in umbilical cord blood), perinatal (infected in 1st postnatal blood Q4 weeks), and postnatal (later postnatal infection). Among 1394 singleton births, girls were more likely to become infected than boys. For in utero transmission, the OR was 1.4 (95% CI: 0.9 to 2.2). For transmission during early life (umbilical cord blood not available) the OR was 2.7 (95% CI: 1.5 to 4.9). However, transmission risks in the perinatal and postnatal infection periods did not differ in boys and girls. Among 303 tested twin-birth pairs, girls were at higher risk than boys for in utero (OR: 2.6; 95% CI: 1.2 to 5.8) and perinatal (OR: 1.9; 95% CI: 1.0 to 3.7) infection. Recognized mother-to-child transmission risk factors did not explain the higher risk of infection in girls. Biggar, et. al. concluded that girls were at higher risk of early (in utero and perinatal) HIV infection than boys. They proposed that minor histocompatibility reactions between maternal lymphocytes and infant Y chromosome derived antigens reduce the risk of HIV transmission in boys. This was quite a complicated collaboration that involved combining data from two studies plus looking at transmission in twins as well as singleton births.

Critical Groups in Dynamic Social Networks

Tanya Berger-Wolf, University of Illinois at Chicago, and Jared Saia, University of New Mexico, addressed the issue of fragility of a network of interactions, such as a social network, in a dynamic setting. They proposed a new mathematical and computational framework that allows analysis of dynamic social networks addressing the time component explicitly. In this framework, they posed the question of finding a critical set of groups at various times whose lack of existence would leave no persistent social structure in the network. Berger-Wolf and Saia formulated this question in terms of a graph optimization problem, proved that it is NP-hard, and provided a polynomial time algorithm for one important special case.

Group Sequential Comparative Poisson Trials For Vaccines And Other Studies

Group sequential methods have been commonly applied to clinical trials with continuous and binary outcomes, but are not well described for comparative Poisson trials (of Poisson or rare binary outcomes) that are commonly used in vaccine studies. Donald R. Hoover and his student Qi Xia developed an exact group sequential procedure for comparative Poisson trials based on exact conditional binomial distributions of a number of events from the treated subjects given the total number of events at each

interim analysis stage. Hoover and Xia showed that application of this exact group sequential procedure can greatly reduce the length of a comparative Poisson trial for vaccines or other interventions in terms of expected number of events needed, particularly when the null hypothesis is false. They also explored the impact of the discreteness of the binomial distribution on the efficiency of the group sequential procedure. Hoover and Xia have developed usable Splus programs to implement this procedure, calculating the test size, power and expected number of events under different conditions.

Pharmacovigilance

A principle concern of pharmacovigilance is the timely detection of adverse drug reactions that are novel by virtue of their clinical nature, severity, and/or frequency. There is understandably an interest in developing data base screening tools to assist human reviewers in identifying associations worthy of further investigation (i.e. signals) embedded within a data base consisting largely of background "noise" containing reports of no substantial public health significance. Data mining algorithms are therefore being developed, tested and/or used by health authorities, pharmaceutical companies and academic researchers. Different data mining approaches in pharmacovigilance include disproportionality analyses, sequential probability ratio tests, correlation analyses, and multivariate regression. Most of the published experience to date has been with so called disproportionality analyses. While the precise operational details of each disproportionality algorithm vary, they all calculate surrogate observed-to-expected ratios in which the reporting experience of each reported Drug-Event Combination (DEC) is compared to the background reporting experience across all/most drugs and events using an independence model. In the appropriate clinical context, DECs that stand out statistically against the background reporting experience may reflect credible signals warranting additional investigation. If there is sufficient correlation between these statistical metrics and novel causal associations, these tools could improve drug safety monitoring. However, many current disproportionality analysis methods have the potential to perform poorly in real-world databases. DIMACS post-doc Ivan Zorych is developing a second generation of data mining algorithms for pharmacovigilance that have the potential to provide sharply improved signal detection performance. These algorithms draw on previous work on large-scale Bayesian logistic regression. Zorych is conducting extensive experiments using the FDA's 'Adverse Events Reporting' database (AERS). Initial results are promising and the experiments are ongoing.

Chemostat Models Applied to a Single Species

Patrick De Leenheer, DIMACS post doc 2003/04, initiated several collaborations during his stay at DIMACS. All of them involved his post-doc advisor, Eduardo Sontag from Rutgers. The first result was obtained from joint work with senior project organizer Simon Levin from Princeton and Christopher Klausmeier from Georgia Tech. They considered chemostat models (a chemostat is a biological reactor in which one or more species compete for one or more nutrients) and studied the dynamics of a single species, limited by two nutrients, assuming that nutrient uptake and growth are decoupled. For a broad class of uptake and growth functions they showed that a nontrivial equilibrium might exist. Moreover, if it exists it is unique and globally stable, generalizing a previous result by Legović and Cruzado. The resulting publication is listed in the Papers section of this report.

Chemostat Models Applied to Multiple Species

A second result on chemostat models followed from collaboration of Patrick De Leenheer with David Angeli from the University of Firenze, Italy, visiting at DIMACS, and Eduardo Sontag from Rutgers. They considered again a chemostat model, but this time an arbitrary number of species was assumed to compete for a single nutrient. Also, here they idealized the conversion of nutrients into new biomass, in contrast to the previously described work. A distinct feature of this model however -one that sets it apart from the traditional chemostat models- is that crowding effects are taken into consideration. The model

can be rewritten as a negative feedback interconnection of two systems that are monotone (as input-output systems). Moreover, these subsystems behave nicely when subject to constant inputs. This allows the use of a particular small-gain theorem that has recently been developed for feedback interconnections of monotone systems. Both global stability and coexistence results have been obtained (the latter holds if crowding effects are large enough). Their coexistence result should be compared to the classical chemostat theory where coexistence does not occur. The resulting publication is listed in the Papers section of this report.

Predator-Prey Systems

Patrick De Leenheer, David Angeli, University of Firenze, Italy, and Eduardo Sontag, Rutgers, obtained an almost global convergence result for Lotka-Volterra systems with predator-prey interactions. Since the early work of Lotka and Volterra, predator-prey systems have continued to attract significant attention. It is well known that these systems may exhibit oscillatory behavior; the best known is the classic Lotka-Volterra predator-prey system. These systems can be written as (negative) feedback systems. The subsystems of the feedback loop are monotone control systems, possessing particular input-output properties. De Leenheer, Angeli, and Sontag used a small-gain theorem, adapted to a context of systems with multiple equilibrium points, to obtain the desired almost global convergence result, which provides sufficient conditions to rule out oscillatory or more complicated behavior that is often observed in predator-prey systems.

Interaction Network Monitoring.

Graham Cormode, DIMACS post doc 2003/04, and Muthu Muthukrishnan, Computer Science, Rutgers, studied the problem of monitoring massive sequences of interactions (person-person or animal-animal, etc.). The challenge of monitoring massive amounts of data generated by interaction networks has led to increased interest in data stream processing. Cormode and Muthukrishnan studied streams of edges in massive interaction multigraphs, defined by (source, destination) pairs. The goal is to compute properties of the underlying graph while using small space (much smaller than the number of participants), and to avoid bias introduced because some edges (interactions) may appear many times, while others are seen only once. They have results for three fundamental problems on multigraph degree sequences: estimating frequency moments of degrees, finding the heavy hitter degrees, and computing range sums of degree values. Space bounds for the summarizing algorithms are significantly smaller than storing complete information. In an experimental study, such summaries are seen to be highly effective, enabling massive multigraph streams to be effectively summarized to answer queries of interest with high accuracy using only a small amount of space.

Sensor Networks

Graham Cormode, Bell Labs, and Martin Farach-Colton, Rohan Fernandes, Miguel Mosteiro and Muthu Muthukrishnan, all from Rutgers, worked on a variety of mathematical optimization problems relating to the placement and organization of sensor networks with particular application to tracking the spread and transmission of disease and other epidemic-like phenomena. They have results bounding the time of communication of events in these networks, and have shown how to compute maximum likelihood estimators for readings with errors between multiple sensors. (Farach-Colton and Muthukrishnan were senior faculty in the project, Fernandes a project graduate student and Mosteiro a graduate student in Farach-Colton's group.) Sensor networks, that is, networks of very inexpensive sensors distributed at random, are likely to become an important tool for epidemiological data gathering, for example when the sensors detect bioactive agents. The cheaper the sensors, the easier they will be to deploy. But cheap sensors have many limitations -- low transmission power, no global positioning systems, low memory. Thus, they don't know where they are when they get thrown out of an airplane,

and they can only talk to a few of their nearby neighbors. Cormode, Farach-Colton, Fernandes, Mosteiro and Muthukrishnan have studied how to find highly connected low-degree subnetworks of sensor networks in sensors randomly distributed in 2-space. They have preliminary results that show that such networks must exist, but their algorithm for constructing them relies on sensors knowing where they are.

The second thrust of their work has been in algorithms for having sensors compute their position from knowledge of the approximate distance to their close neighbors. They have shown that sensor networks, whose distance measurements form distance threshold graphs (that is, you know the distances to nearby neighbors but have no information of distances to distant neighbors) have interesting rigidity properties in two dimensions, which means that the problem of finding the position of each sensor given noisy distance measurements is highly constrained.

Sensor Networks -- Efficient Monitoring

Graham Cormode, Bell Labs, in conjunction with Minos Garofalakis and Rajeev Rastogi of Bell Labs, and S. Muthukrishnan of Rutgers, studied the problem of effectively conducting distributed monitoring tasks within a large sensor network. Emerging large-scale monitoring applications require continuous tracking of complex aggregates and data-distribution summaries over collections of physically distributed streams. Thus, effective solutions have to be simultaneously space efficient (at each remote site), communication efficient (across the underlying communication network), and provide continuous, guaranteed-quality estimates. They proposed novel algorithmic solutions for the problem of continuously tracking complex holistic aggregates in such a distributed-streams setting -- the primary focus is on approximate quantile summaries, but the approach is more broadly applicable and can handle other holistic-aggregate functions (e.g., "heavy-hitters" queries). They presented the first known distributed-tracking schemes for maintaining accurate quantile estimates with provable approximation guarantees, while simultaneously optimizing the storage space at each remote site as well as the communication cost across the network. The algorithms employ a combination of local tracking at remote sites and simple prediction models for local site behavior in order to produce highly communication- and space-efficient solutions. They performed extensive experiments with real and synthetic data to explore the various tradeoffs and understand the role of prediction models in their schemes. The results clearly validated the approach, revealing significant savings over naive solutions as well as the analytical worst-case guarantees.

Cluster Analysis

Graham Cormode, Bell Labs, collaborated with James Abello, Dmitriy Fradkin, David Madigan, Ofer Melnik, and Ilya Muchnik, all of Rutgers, on applying the data mining technique of cluster analysis to epidemiology. The notion of 'clusters' is a very natural one, and occurs frequently in discussions of epidemiology. We hear about 'cancer clusters', areas where the number of reported cancer cases within an area or group of people exceeds the expected amount. Such clusters lead to investigation of possible carcinogens or explanations for greater susceptibility amongst certain groups. Cluster analysis is an unsupervised learning technique that takes large collections of data points and attempts to identify clusters of similar points. More formally, it tries to create clusters to optimize various mathematical properties, such as minimizing the maximum spread of each cluster, or minimizing the sum of the spreads. A variety of algorithms have been proposed to create clusters from a data set, including k-means, hierarchical clustering, and expectation maximization. Cormode developed several new applications of cluster analysis and provided a survey article on this technique, which is listed in Papers.

The Paradoxical Nature of Locating Sensors in Paths and Cycles

The problem of placing sensors or detectors in a network arises in many applications, including epidemiology, homeland security, civil engineering, manufacturing, fault detection in distributed or multiprocessor systems, etc. There are several goals in sensor placement: Rapid and accurate detection of attacks, faults, or contamination, minimizing the cost of sensors used, and identification of the location of an attack or fault or contamination. The network can be represented as an undirected graph. A subset of the vertices identifies the locations of the sensors. The objective is to place the sensors to guarantee source identification. This problem is NP-complete for general networks and surprisingly complex for simple networks such as paths and cycles. David L. Roberts, a graduate student at Georgia Institute of Technology, and Fred S. Roberts, DIMACS, provided the complete solution under one of the formulations given by Bertrand, Charaon, Hudry, and Lobstein when detectors can detect attacks on vertices up to two steps away in the network. In the process of solving this problem, Roberts and Roberts found that the solution is sufficiently counterintuitive that it raises some very interesting paradoxes about sensor location. Roberts and Roberts argue that the complex goals of sensor placement require careful analysis in order to achieve the goals. Since results are sometimes counterintuitive and seemingly paradoxical, sensor placement strategies require the methods of computer science and mathematics.

The Containment Problem for the Spread of Disease

Consider a social or computer network graph G with a set of nodes V and a subset of m nodes of V, each of which is infected with a virus and from which the infection spreads to its non-vaccinated neighbors in one time step. Assume that we have the ability to vaccinate a limited number of nodes, a, during each time step. The goal is to find what nodes to vaccinate at each time step to minimize the total number m of nodes that eventually become infected. This problem is NP hard but James Abello, DIMACS and Ask.com, and Mike Capalbo, DIMACS visitor, devised a tractable approximation algorithm that produces a vaccination strategy that vaccinates $O(\log(V(G))*a$ at each time step, guaranteeing that not more than 3m nodes will eventually become infected. The algorithm is obtained via a linear relaxation of an Integer Programming formulation of the containment problem.

The Firefighter Problem

Graduate student Paul Raff, Rutgers, Mathematics, began a collaboration with DIMACS post doc Kah Loon Ng, after Raff attended a talk by Ng on the firefighter problem. DIMACS visitor James Abello and Rutgers graduate Stephen Hartke, UIUC, have also participated in this research. Raff and Ng investigated a special case of the so-called firefighter problem, which involves the following scenario: in a given graph a finite number of vertices are initially "on fire" or diseased. At each turn t, a number f(t) of firefighters (or vaccines) are able to be positioned at unburned vertices to prevent the fire from spreading there. Once a vertex is on fire or defended it stays that way permanently. The main goal of the firefighter problem is to determine if it is possible to have a scenario where there are vertices that are not defended yet can never catch fire. Raff and Ng's work concentrated on the two-dimensional grid as the underlying graph. It was previously known that having exactly one firefighter per turn would not be sufficient to contain the fire, whereas having exactly two firefighters per turn would. Raff and Ng extended these results and proved that if more than 1.5 firefighters on average are available per turn, then the fire could be contained.

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 2006 (see http://dimacs.rutgers.edu/Workshops/Biomath/). Also, the project organizers are working closely with the Summer 2006 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. Indeed, BMCI's main new theme Summer 2006 will be mathematical and computational epidemiology.

DIMACS will partner with four minority-serving institutions, Clark Atlanta, Howard, Morgan State, and North Carolina A&T, in the development of a new initiative to bring together scientists from the United States and various African countries to collaborate on mathematical modeling of infectious diseases in Africa. DIMACS, in collaboration with the South African Centre for Epidemiological Modeling and Analysis (SACEMA), is planning a 3-day workshop on mathematical modeling and infectious diseases in Africa and later, together with the African Institute for Mathematical Sciences (AIMS), a two week shortcourse enabling those who complete it to participate fully in a following 3-day "capstone" workshop at which shortcourse students and researchers from both the United States and Africa will interact and establish collaborations. The shortcourse is aimed at training junior United States and African participants (graduate students and postdoctoral fellows) in mathematical epidemiology and the control of emerging and re-emerging infectious diseases. The following workshop will be a culmination of the shortcourse, involving active United States and African researchers and aimed at furthering research on the modeling of infectious diseases in Africa. The idea for these activities stemmed from the workshop on Evolutionary Considerations in Vaccine Use, held this year.

Products:

Books

Abello, J., and Cormode, G. (eds.), *Discrete Methods in Epidemiology, AMS-DIMACS Volume Series*, Vol. 70, American Mathematical Society, July 2006.

Cioffi-Revilla, C., *Power Laws in the Social Sciences* (tentative title), Cambridge University Press, under review.

Feng, Z., Dieckmann, U., and Levin, S. (eds.), *Disease Evolution: Models, Concepts, and Data Analyses, AMS-DIMACS Volume Series*, Vol 71, American Mathematical Society, July 2006.

Day, T., Gandon, S., et. al., (eds.) *Evolutionary Considerations in Vaccine Use*, Special Supplement to *Vaccine*, in preparation.

Papers

Abello, J., and Capalbo, M., "Random graphs (and the spread of infections in a social network)," *Discrete Methods in Epidemiology, AMS-DIMACS Volume Series*, Vol. **70**, Abello, J., and Cormode, G. (eds.), American Mathematical Society, July 2006.

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," *Discrete Methods in Epidemiology, AMS-DIMACS Volume Series*, Vol. **70**, Abello, J., and Cormode, G. (eds.), American Mathematical Society, July 2006.

Abello, J., Pogel, A., "Graph partitions and concept lattices," *Discrete Methods in Epidemiology, AMS-DIMACS Volume Series*, Vol. **70**, Abello, J., and Cormode, G. (eds.), American Mathematical Society, July 2006.

Angeli, D., De Leenheer, P., and Sontag, E.D. "On the structural monotonicity of chemical reaction networks," submitted.

Arino, J., Jordan, R., and van den Driessche, P., "Quarantine in a multi-species epidemic model," *Mathematical Biosciences*, to appear.

Berger-Wolf, T.Y., DasGupta, B., Chaovalitwongse, W., and Ashley, M. V., "Combinatorial reconstruction of sibling relationships in absence of parental data," DIMACS Technical Report 2005-27.

Berger-Wolf, T.Y., and Saia, J., "Critical groups in dynamic social networks," DIMACS Technical Report 2005-26.

Berger-Wolf, T.Y., and Saia, J., "A framework for analysis of dynamic social networks," DIMACS Technical Report 2005-28.

Beshers, S., and Fefferman, N.H., "Worker task allocation in social insect colonies: Should the strategy depend on disease presence?," in preparation.

Biggar, R.J., Taha, T.E., Hoover, D.R., Yellin, F., Kumwenda, N. and Broadhead, R., "Higher in utero and perinatal HIV-infection risk in girls than boys," *J. Acquir Immune Defic Synd*r **41** (2006), in press.

Boots, M., Kamo, M., and Sasaki, A., "The implications of spatial structure within populations to the evolution of parasites," *Disease Evolution: Models, Concepts, and Data Analyses, AMS-DIMACS Volume Series,* Vol. **71**, Feng, Z., Dieckmann, U., and Levin, S., (eds.), American Mathematical Society, July 2006.

Charleston, M.A., and Galvani, A.P., "A cophylogenetic perspective on host-pathogen evolution," *Disease Evolution: Models, Concepts, and Data Analyses, AMS-DIMACS Volume Series,* Vol. **71**, Feng, Z., Dieckmann, U., and Levin, S., (eds.), American Mathematical Society, July 2006.

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.

Church, J.D., Jones, D., Flys, T., Hoover, D.R., Marlowe, N., Chen, S., Shi, C., Eshleman, J.R., Guay, L.A., Jackson, J.B., Kumwenda, N., Taha, T.E. and Eshleman, S.H., "Sensitivity of the ViroSeq™ HIV-1 Genotyping System for detection of the K103N resistance mutation in HIV-1 subtypes A, C, and D," *J. Mol. Diagnostics*, (2006), accepted.

Cormode, G., Garofalakis, M., and Muthukrishnan, S., "Holistic aggregates in a networked world: Distributed tracking of approximate quantiles," *Proceedings of ACM SIGMOD International Conference on Management of Data,* 2005.

Day, T., and Gandon, S., "Insights from Price's equation into evolutionary epidemiology," *Disease Evolution: Models, Concepts, and Data Analyses, AMS-DIMACS Volume Series,* Vol. **71**, Feng, Z., Dieckmann, U., and Levin, S., (eds.), American Mathematical Society, July 2006.

De Leenheer, P., Angeli, D., and Sontag, E.D., "Monotone chemical reaction networks," *Journal of Mathematical Chemistry*, March 2006, published electronically.

De Leenheer, P., Angeli, D., and Sontag, E.D., "Crowding effects promote coexistence in the chemostat," *Journal of Mathematical Analysis and Applications*, **319** (2006), 48-60.

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*, **52** (2006), 419-438.

De Leenheer, P., and Pilyugin, S.S., "Feedback-mediated coexistence and oscillations in the chemostat," submitted.

De Leenheer, P., and Angeli, D., "Monotonicity and convergence in chemical reaction networks," *Proceedings of the Joint 44th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC'05)*, Seville, Spain, December 2005.

De Leenheer, P., Mazenc, F., and Malisoff, M., "Stabilizing a periodic solution in the chemostat: A case study in tracking," submitted.

Delmotte, F., Egerstedt, M., and Verriest, E-I., "Pulse control of epidemics by vaccination," in preparation.

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

Eshleman, S.H., Hoover, D.R., Chen, S., et. al., "Nevirapine (NVP) resistance in women with subtype C compared to subtypes A and D HIV after single dose NVP," *J. Infect. Dis.*, **192** (2005), 30-6.

Eshleman, S.H., Hoover, D.R., Chen, S.H., Hudelson, S.H., Guay, L.A., Mwatha, A., Fiscus, S.A., Mmiro, F., Musoke, P., Jackson, J.B., Kumwenda, N., and Taha, T., "Resistance after single dose nevirapine prophylaxis emerges in a high portion of Malawian newborns," *AIDS*, **19** (2005), 2167-69.

Eshleman, S.H., Hoover, D.R., Chen, S., Hudelson, S.E., Guay, L.A., Mwatha, A., Fiscus, S.A., Mmiro, F., Musoke, P., Jackson, J.B., Kumwenda, N., Taha, T., "Nevirapine (NVP) resistance in women with HIV-1 subtype C, compared with subtypes A and D, after the administration of single-dose NVP," *J. Infect. Dis.*, **192** (2005), 30-6.

Eshleman, S.H., Hoover, D.R., Hudleson, S.E., Chen, S., Fiscus, S.A., Piwowar-Manning, E., Jackson, J.B., Kumwneda, N.I., Taha, T.E., "Development of nevirapine resistance in infants is reduced by use of infant-only single-dose nevirapine plus zidovudine postexposure prophylaxis for the prevention of mother-to-child transmission of HIV-1," *J. Infect. Dis.*, **193** (2006), 479-81.

Eshleman, S.H., Church, J.D., Chen, S., Guay, L.A., Mwatha, A., Fiscus, S.A., Mmiro, F., Musoke, Kumwenda, N., Jackson, J.B., Taha, T.E., Hoover, D.R., "Comparison of HIV-1 mother-to-child transmission after single dose nevirapine prophylaxis among African women with subtypes A, C and D," *JAIDS*, (2006), accepted.

Eshleman, S.H., Lie, Y., Hoover, D.R., Chen, S., Hudelson, S.E., Fiscus, S.A., Petropoulos, C.J., Kumwenda, N., Parkin, N., and Taha, T.E., "Association of HIV-1 replication capacity with HIV-1 mother-to-child transmission among antiretroviral drug naïve Malawian women," *J. Infect. Dis.*, (2006), accepted.

Farach-Colton, M., Fernandes, R.J., and Mosteiro, M.A., "Bootstrapping hop-optimal networks in the weak sensor model," in *Proceedings of the 13th Annual European Symposium on Algorithms (ESA), Lecture Notes in Computer Science*, **3669**, Springer-Verlag, (2005), 827-838.

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

Fefferman, N.H., and Naumova, E.N., "Combinatorial decomposition of an outbreak signature," *Mathematical Biosciences*, to appear.

Fefferman, N.H. and Traniello, J.F.A., "Social insects as models in epidemiology: Establishing the foundation for an interdisciplinary approach to disease and sociality," in *Organization of Insect Societies: From Genome to Sociocomplexity*, J. Gadau and J. Fewell (eds.), Harvard University Press, to appear.

Fefferman, N.H., and Starks, P.T.B., "Polistes nest founding behavior: A model for the selective maintenance of alternative behavioral phenotypes," *Special Issue on Polistes Wasps*, to appear.

Fefferman, N.H., Traniello, J.F.A., Rosengaus, R.B., and Calleri, D.V., "Disease prevention and resistance in social insects: Modeling the survival consequences of immunity, hygienic behavior and colony organization using cellular automata," *Behavioral Ecology and Sociobiology*, under revision.

Fefferman, N.H., and Naumova, E.N., "Seasonality in infectious disease incidence: An emergent property of social networks," submitted.

Fefferman, N.H. and Ng, K.L., "An agent-based model of social network evolution under three measures of centrality," in preparation.

Fefferman, N.H., Lofgren, E., and Naumova, E.N., "Theories of observed seasonality in influenza: A review," in preparation.

Feng, Z. and Rong, L., "The influence of anti-viral drug therapy on the evolution of HIV-1 pathogens," *Disease Evolution: Models, Concepts, and Data Analyses, AMS-DIMACS Volume Series,* Vol. **71**, Feng, Z., Dieckmann, U., and Levin, S., (eds.), American Mathematical Society, July 2006.

Fiscus, S.A., Chen, S., Hoover, D.R., Kerkau, M.G., Alabanza, P., Siharath, S., Schmitz, J., Kumwenda, N., Taha, T.E., "Affordable, abbreviated roche monitor assay for quantification of human immunodeficiency virus type 1 RNA in plasma," *J. Clin. Microbiol,*. **43** (2005), 4200-2.

Fradkin, D., Muchnik, I., Hermans, P., and Morgan, K., "Validation of epidemiological models: Chicken epidemiology in the UK," *Discrete Methods in Epidemiology, AMS-DIMACS Volume Series*, Vol. **70**, Abello, J., and Cormode, G. (eds.), American Mathematical Society, July 2006.

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

Getz, W.M. and Lloyd-Smith, J.O., "Basic methods for modeling the invasion and spread of contagious diseases," *Disease Evolution: Models, Concepts, and Data Analyses, AMS-DIMACS Volume Series,* Vol. **71**, Feng, Z., Dieckmann, U., and Levin, S., (eds.), American Mathematical Society, July 2006.

Getz, W.M., Lloyd-Smith, J.O., Cross, P.C., Bar-David, S., Johnson, P.L., Porco, T.C., and Sánchez, M.S., "Modeling the invasion and spread of contagious diseases in heterogeneous populations," *Disease Evolution: Models, Concepts, and Data Analyses, AMS-DIMACS Volume Series,* Vol. **71**, Feng, Z., Dieckmann, U., and Levin, S., (eds.), American Mathematical Society, July 2006.

Hartke, S. G., "Attempting to narrow the integrality gap for the firefighter problem on trees," *Discrete Methods in Epidemiology, AMS-DIMACS Volume Series*, Vol. **70**, Abello, J., and Cormode, G. (eds.), American Mathematical Society, July 2006.

Hauben, M., Madigan, D., Gerrits, C., and van Puijenbroek, E., "The role of data mining in pharmacovigilance," *Expert Opinion in Drug Safety*, **4** (2005), 929-948.

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

Holt, R.D, and Barfield, M., "Within-host pathogen dynamics: Some ecological and evolutionary consequences of transients, dispersal mode, and within-host spatial heterogeneity," *Disease Evolution: Models, Concepts, and Data Analyses, AMS-DIMACS Volume Series,* Vol. **71**, Feng, Z., Dieckmann, U., and Levin, S., (eds.), American Mathematical Society, July 2006.

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.

Huang, P-H., and Pavlovic, V., "An interpretable method for protein homology detection," *International Conference on Intelligent Systems for Molecular Biology (ISMB2006)*, submitted.

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

Kelly, J.K., "Evolutionary and dynamic models of infection with internal host structure," *Disease Evolution: Models, Concepts, and Data Analyses, AMS-DIMACS Volume Series,* Vol. **71**, Feng, Z., Dieckmann, U., and Levin, S., (eds.), American Mathematical Society, July 2006.

Koppelman, W.J., and Adler, F.R., "Do rhinoviruses follow the neutral theory? The role of cross-immunity in maintaining the diversity of the common cold," *Disease Evolution: Models, Concepts, and Data Analyses, AMS-DIMACS Volume Series,* Vol. **71**, Feng, Z., Dieckmann, U., and Levin, S., (eds.), American Mathematical Society, July 2006.

Li, J., Muchnik, I., and Schneider, D., "Influences on breast cancer survival via SVM classification in the SEER Database," *Discrete Methods in Epidemiology, AMS-DIMACS Volume Series*, Vol. **70**, Abello, J., and Cormode, G. (eds.), American Mathematical Society, July 2006.

Lloyd, A.L., and Wodarz, D., "Drug resistance in acute viral infections: Rhinovirus as a case study," *Disease Evolution: Models, Concepts, and Data Analyses, AMS-DIMACS Volume Series,* Vol. **71**, Feng, Z., Dieckmann, U., and Levin, S., (eds.), American Mathematical Society, July 2006.

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

Ng, K.L., and Raff, P., "Fractional firefighting in the two dimensional grid," DIMACS Technical Report 2005-23.

Ng, K.L., and Raff, P., "Firefighting on a two dimensional infinite grid with a non-constant number of firefighters," *Journal of Combinatorial Mathematics and Combinatorial Computing*, submitted.

Ng, K.L., and Raff, P., "A generalization of the firefighter problem in the two dimensional infinite grid," *Discrete Math*, in preparation.

Ozonoff, D., Pogel, A., and Hannan, T., "Generalized contingency tables and concept lattices," *Discrete Methods in Epidemiology, AMS-DIMACS Volume Series*, Vol. **70**, Abello, J., and Cormode, G. (eds.), American Mathematical Society, July 2006.

Roberts, D. L., and Roberts, F. S., "The paradoxical nature of locating sensors in paths and cycles: the case of 2-identifying codes," DIMACS Technical Report 2006-02.

Roberts, D. L., and Roberts, F. S., "The paradoxical nature of locating sensors in paths and cycles: the case of 2-identifying codes," *European Journal of Combinatorics*, submitted.

Rolka, H., Burkom, H., Cooper, G.F., Kulldorff, M., Madigan, D., Wong, W-K., "Issues in applied statistics for public health bioterrorism surveillance using multiple data streams: Research needs," *Statistics in Medicine*, to appear.

Shi, J. and Hoover, D. R., "Exact distribution of the smallest multinomial group when each category has equal probability," in preparation.

Smith, D.L., Boni, M.F., and Laxminarayan, R., "Dynamics and control of antibiotic resistance in structured metapopulations," *Disease Evolution: Models, Concepts, and Data Analyses, AMS-DIMACS Volume Series,* Vol. **71**, Feng, Z., Dieckmann, U., and Levin, S., (eds.), American Mathematical Society, July 2006.

Taha, T.E., Kumwenda, N., Mwakomba, A., Mwenda, R., Kawonga, H., Gaydos, C., Hoover, D.R., Kafalafula, G., "Safety, acceptability, and potential efficacy of a topical penile microbicide wipe," *J. Acquir. Defic. Syndr.*, **39** (2005), 347-353.

Xia Qi, and Hoover, D. R., "The group sequential method applied to comparative poisson trials," in preparation.

Talks

De Leenheer, P., "Bistability and oscillations in the feedback-controlled chemostat," Department of Mathematics and Department of Electrical Engineering, Louisiana State University, February 21, 2006.

De Leenheer, P., “Feedback control in the chemostat,” Department of Mathematics, University of Florida, December 8, 2005.

De Leenheer, P., “Feedback-mediated coexistence and oscillations in the chemostat,” Department of Mathematics, University of North Florida, November 18, 2005.

De Leenheer, P., “Remarks on monotonicity in certain classes of chemical networks,” Conference on Differential and Difference Equations and Applications, Melbourne, Florida, August 1, 2005.

De Leenheer, P., “Remarks on monotonicity in certain classes of chemical networks,” Sixth SIAM Conference on Control and its Applications, New Orleans, Louisiana, July 12, 2005.

De Leenheer, P., “Remarks on monotonicity in certain classes of chemical networks,” Colloquium at Department of Mathematical Analysis, Ghent University, Belgium, May 26, 2005.

De Leenheer, P., “Remarks on monotonicity in certain classes of chemical networks,” Seminar at FUNDP, Namur, Belgium, May 18, 2005.

De Leenheer, P., “Remarks on monotonicity in certain classes of chemical networks,” Colloquium at Department of Mechanical Engineering, Eindhoven University of Technology, The Netherlands, May 17, 2005.

De Leenheer, P., “Remarks on monotonicity in certain classes of chemical networks,” Department of Mathematics, University of Florida, April 19, 2005.

De Leenheer, P. and Angeli, D., “Monotonicity and convergence in chemical reaction networks,” Joint 44th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC'05), Seville, Spain, December 12-15 2005.

Fefferman, N.H., “How would termites prepare for pandemic bird flu and what should we learn from them?” Joint Dept. of Entomology and Center for Infectious Disease Dynamics Seminar, Penn State Univ., 2006.

Fefferman, N.H., “Applications of self-organizing systems to epidemiology,” DIMACS Mixer Series, Rutgers Univ., 2005.

Fefferman, N.H., “Fantastic problems in mathematical ecology,” DIMACS Bio-Math Connect Institute for High School Teachers, Rutgers Univ., 2005.

Mosteiro, M. A., “Bootstrapping a hop-optimal network in the weak sensor model,” Universidad de Buenos Aires, Argentina, Computer Science Department, August 18, 2005.

Ng, K-L. and Fefferman, N., “241 - Comparative effects of different centrality measures in dynamic networks,” XXVI International Sunbelt Social Network Conference, Vancouver, British Columbia, Canada, April 25-30, 2006.

Raff, P., “The firefighter problem in the two-dimensional grid,” Graduate Combinatorics Seminar, Rutgers University, November 2005.

Raff, P., “The firefighter problem in the two-dimensional grid,” Spring School in Combinatorics, Borova Lada, Czech Republic, April 2006.

Roberts, F., "Graph-theoretical problems arising from defending against bioterrorism and controlling the spread of fires," DIMACS/DIMATIA/Renyi Combinatorial Challenges Conference, Piscataway, NJ, April, 2006.

Roberts, F., "Graph-theoretic models of the spread and control of disease and of fighting fires," DIMACS Bio-Math Connect Institute, Piscataway, NJ, July, 2005.

Roberts, F., "Graph-theoretic models of the spread and control of disease and of fighting fires: The Stewart-Dowart Lecture," Trinity College, Hartford, CT, March, 2006.

Verriest, E.I., Delmotte, F., and Egerstedt, M., "Control of epidemics by vaccination," 2005 American Control Conference, Portland, Oregon, June 2005.

Main Web Site

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

Other Specific Products

Web Pages

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

Workshop: Economic Epidemiology
http://dimacs.rutgers.edu/Workshops/EconEpid

Workshop: The Epidemiology and Evolution of Influenza
http://dimacs.rutgers.edu/Workshops/Influenza/

Working Group Meeting: Adverse Event/Disease Reporting, Surveillance and Analysis IV
http://dimacs.rutgers.edu/Workshops/AdverseEvent4

Working Group Meeting: BioSurveillance Data Monitoring and Information Exchange
http://dimacs.rutgers.edu/Workshops/Surveillance/

Tutorial on Data Mining and Epidemiology
http://dimacs.rutgers.edu/Workshops/DataMiningTutorial/

Workshop: Combinatorial Group Testing
http://dimacs.rutgers.edu/Workshops/CGT/

Tutorial: Phylogenetic Trees and Rapidly Evolving Pathogens
http://dimacs.rutgers.edu/Workshops/PhyloTutorial

Workshop: Phylogenetic Trees and Rapidly Evolving Pathogens
http://dimacs.rutgers.edu/Workshops/PhylogeneticTrees

Working Group Meeting: Phylogenetic Trees and Rapidly Evolving Pathogens II
http://dimacs.rutgers.edu/Workshops/WGPhylogeneticTrees2

Special Focus Seminar Series 2005-2006
http://dimacs.rutgers.edu/SpecialYears/2002_Epid/seminars05-06.html

Reports

Abello, J., and Cormode, G., "Report on DIMACS Tutorial on Data Mining and Epidemiology" Available electronically at http://dimacs.rutgers.edu/Workshops/DataMiningTutorial/dmtutorialreport.pdf

Software

In 2004 and 2005, DIMACS fostered a collaboration between David Ozonoff (Boston University) and Alex Pogel (U. of New Mexico). During 2006, Pogel's subcontract with Ozonoff's SuperFund grant has seen much progress made toward building and distributing an open source concept lattice analysis tool for epidemiologists, which will generate contingency tables from concept lattices with a few clicks on lattice nodes. This work stemmed originally from this NSF project. Pogel has hired a team of software engineers, written a specification document, and has a well-defined five-year plan for the completion of the project. The first version of this software is expected to appear at SourceForge.net in mid-2008, and the product will be publicized in the 2009-2010 academic year. Also, in Pogel's work at PSL, various features described in the Ozonoff, Pogel, and Hannan contribution to the AMS-DIMACS Special Volume on Computational Epidemiology ("Generalized Contingency Tables") have been incorporated into Seqer (pronounced "seeker"), the data analysis software developed by Pogel's group over the past three years.

Contributions

Contributions within Discipline

One of the biggest contributions of this project has been the development of a new initiative to bring together scientists from the United States and various African countries to collaborate on mathematical modeling of infectious diseases in Africa. One specific DIMACS workshop was instrumental in leading to the development of the proposed activity in Africa. This was a workshop on ``Evolutionary Aspects of Vaccine Use," organized at DIMACS at Rutgers University in June 2005, and stemming from the work of the DIMACS working group on Methodologies for Comparing Vaccination Strategies. The workshop, organized by Troy Day (Queen's University, Ontario), Alison Galvani (Yale), Abba Gumel (U. of Manitoba), and Claudio Struchiner (Oswaldo Cruz Foundation, Brazil), examined general evolution-related questions for diseases for which there is a vaccine or hope for one. There were 30 scientists participating, from diverse backgrounds, including mathematicians, epidemiologists, virologists, immunologists, and vaccine developers. The special problems of vaccination strategies in Africa that arose in this workshop were one of the primary motivations that led Abba Gumel to propose that DIMACS sponsor a short course and workshop that directly focuses on mathematical modeling of infectious diseases in Africa. DIMACS, in collaboration with the South African Centre for Epidemiological Modeling and Analysis (SACEMA), developed plans to hold a 3-day Workshop on mathematical modeling and infectious diseases in Africa, to be held at the School of Computational and Applied Mathematics at the University of the Witwatersrand, Johannesburg, South Africa, September 26-28, 2006. The workshop will bring together scientists from the United States and various African countries, as well as junior researchers and students. The workshop will provide an agenda for future collaborations between United States and African scientists. It will expose junior United States scientists

and students to the special challenges of modeling the spread of disease in Africa and the opportunities to collaborate with Africans in developing and applying the tools of mathematical modeling to the tremendous health problems caused by such diseases as HIV/AIDS, tuberculosis and malaria, as well as the possible interventions in the case of major new health threats such as pandemic influenza in an underdeveloped region of the world. In addition, DIMACS, SACEMA, and the African Institute for Mathematical Sciences (AIMS), has developed plans for a two week shortcourse on mathematical modeling and infectious diseases in Africa, enabling those who complete it to participate fully in a following 3-day "capstone" workshop at which shortcourse students and researchers from both the United States and Africa will interact and establish collaborations. The shortcourse, to be held at AIMS in Cape Town, South Africa, on June 11-22, 2007, is aimed at training junior United States and African participants (graduate students and postdoctoral fellows) in mathematical epidemiology and the control of emerging and re-emerging infectious diseases. The following workshop, to be held at new conference facilities at The Stellenbosch Institute for Advanced Study (STIAS) June 25-27, will be a culmination of the shortcourse, involving active United States and African researchers and aimed at furthering research on the modeling of infectious diseases in Africa.

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 Ng and Fefferman's work on the effect of social networks on disease spread. Other groups, including Egerstedt, Verriest, and Delmotte 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. Below is a selection of comments from project participants indicating their assessment of the impact of the project:

"This focus has had an important impact on research at the University of Liverpool. We have been awarded two grants to develop collaborative work. The first was from EPSRC the UK Engineering and Physical Research Council. It was a small award, £10000 to enable preliminary work to be carried out on the application of support vector machine (SVM) feature selection to observational epidemiology. The preliminary work was successful and we have recently (2006-9) been awarded £400,000 for a three year collaboration. This collaboration is with Professor Ilya Muchnik [of DIMACS] and originally his graduate student Dmitry Fradkin. The application of SVM to observational epidemiology is novel and promises an alternative to logistic regression. The project will employ 2 postdoctoral students, one computer scientist and 1 veterinarian. The award will enable DIMACS staff to visit the UK and hopefully will result in further collaboration between the groups. The friendly atmosphere at DIMACS and the ease by which it can be reached from Newark on the New Jersey Transit make it a simple trip from Manchester UK, this combined with the relaxed and informal atmosphere make it an ideal "melting pot" for multidisciplinary collaboration. The Director of DIMACS is to be commended for developing and supporting this special focus and serious consideration should be given to extending it. Currently, epidemics of infectious disease in animals and people pose a major threat to global health and prosperity. As part of my visits to DIMACS I have also become involved in other special focus areas e.g. data mining. I look forward to visiting DIMACS 2-3 times a year over the next three years."
Kenton L. Morgan, Professor of Epidemiology, Faculty of Veterinary Science, Leahurst, NESTON

"In addition to affecting the course of my personal interests and undertakings, DIMACS has fostered greater scientific interest in the fields of epidemiological modeling, helping me to co-launch InForMID (the Initiative for the Forecasting and Modeling of Infectious Disease) a new formal project based out of Tufts University School of Medicine. Many of the members of this new Initiative were initially made aware of this undertaking through a connection to DIMACS. (A list of the members who became

interested through contact with DIMACS is given below.) Without the workshops, conferences and seminars DIMACS has hosted, these connections would never have been made."
Nina Fefferman, DIMACS Visitor, Tufts University

InForMID Members due to DIMACS involvement

Shweta Bansal, Univ. of Texas at Austin
Sally Blower, Univ. of California at Los Angeles
Sara Del Valle, Los Alamos National Laboratories
Holly Gaff, Univ. of Maryland School of Med.
Melike Gursoy, BioMAPS/ Rutgers Univ.
Tanya Kostova, Lawrence Livermore National Laboratory
Ramanan Laxminarayan, Resources for the Future
Eduardo Massad, Universidade de São Paulo
Kah Loon Ng, DIMACS/ Rutgers Univ.
Larry Shepp, Rutgers Univ.

"Ever since your workshop, I have tested some of the same ideas ("dominance principles," as per my presentation) in different domains, including a DARPA project in which I was involved last year. The DIMACS interactions were extremely useful and I continue to develop some parts of that project in a forthcoming book, hopefully out in 2006. Tentative title "Power Laws in the Social Sciences." Under review at Cambridge University Press. I will also explore ways in which dominance principles might be applicable in a current project with NSF funding (HSD), on a multi-agent model of social adaptation to environments and change. For example, to my knowledge, dominance principles have not been applied to network analysis, where multivariate situations are common in nonlinear contexts."
Claudio Cioffi-Revilla, Professor of Computational Social Sciences, Director, Center for Social Complexity, Krasnow Institute for Advanced Study, George Mason University

"One of the projects in our new NIH funded superfund program was for pursuing work Alex Pogel and I began under the aegis of this Special Focus. It was approved and will be funded for 5 years."
David Ozonoff, MD, MPH, Professor of Environmental Health, Chair Emeritus, Department of Environmental Health, Boston University

"I attended the Biosurveillance Information Exchange Working Group held on February 22-24, 2006. First of all I wanted to mention upfront that the meeting was excellent. The presenters and attendees were the right group to get together for this group. It was not too long or unfocused as many meetings/conferences have a tendency to be. Secondly, the outcomes of the meeting have already been personally very productive and stimulating. The biosurveillance working group has continued to remain in contact and we are planning on continuing the activities and ideas that we began at the initial DIMACS meeting. Several of us are also taking lead roles as presenters and session leaders for the public health track of the 2006 Syndromic Surveillance Conference pre-conference workshop. Some of the ideas for the track came out of post DIMACS discussions we have had amongst the group. Overall, an excellent group of people, an excellent meeting and excellent facilities."
Marc Paladini, The New York City Department of Health & Mental Hygiene

"I think that our workshop resulted in a number of new research contacts, and we are also now working with the journal 'Vaccine' to publish a special supplement of their journal that highlights aspects of our workshop (Evolutionary considerations in vaccine use). This will help greatly to convey the significance of this kind of research to the broader community of people doing vaccine research. We are still in the initial stages of this so I cannot yet point to a finished product, but the DIMACS workshops was instrumental in making this happen."

Troy Day, Departments of Mathematics & Biology, Jeffery Hall, Queen's University, Kingston, ON, K7L 3N6, Canada

Contributions To Other Disciplines

This "discipline" is inherently multidisciplinary. An example of the type of synergy across disciplines that has been created, Erik I.Verriest, and Ph.D. student Florent Delmotte, School of Electrical and Computer Engineering, College of Engineering, Georgia Institute of Technology, said, "As an engineer specialized in control and optimization, I am now looking for applications in epidemiology (and) one PhD student (F. Delmotte) is actively working on optimal control of delay systems with epidemiology as an example."

Contributions Beyond Science And Engineering

Several of the outcomes of the special focus have been immediately applicable to real world problems. David Madigan reported that the Working Group on Adverse Event/Disease Reporting, Surveillance and Analysis has held a number of meetings that brought together researchers in drug safety with researchers in disease surveillance. This has been moderately successful in bridging the gap between these two areas. For example, Martin Kulldorff (Harvard University), a prominent disease surveillance researcher now has a related program in drug safety. He maintains regular contact with David Madigan, Rutgers and DIMACS, who is using some of Kulldorff's disease surveillance methods in the drug safety context. The Working Group spawned a collaboration between Madigan and Manfred Hauben of Pfizer. This has already resulted in one publication with several more in the pipeline, as well as a pending proposal for funding. Meanwhile on the disease surveillance side, recent Working Group activity has focused on meeting the educational needs of public health practitioners around the country now using syndromic surveillance systems on a daily basis.

Twelve members of the Working Group have composed a letter, "Accelerating Statistical Research in Drug Safey," that they hope to send to the FDA to bring attention to the barriers that still exist to research and development in the area of the application of data mining and statistical algorithms to pharmacovigilance.. This is an example of synergy that has developed between industry and academe.

Additional outcomes are described by participants as follows:

"I attended the recent meeting, "DIMACS Working Group on Adverse Event/Disease Reporting, Surveillance and Analysis IV." This forum was extremely useful as a neutral ground for sharing ideas among competing companies regarding subjects of mutual interest, in the area of drug safety, an area of high importance to the public. One particular result is our work in progress on a letter that we hope will reach the FDA, regarding improved data sources for drug safety research. Beyond this, the forum provided an excellent opportunity to compare and contrast research approaches to pharmacovigilance, and to set directions for future research in this area."
Alan Hochberg, Vice President, Research, ProSanos Corp.

"I greatly enjoyed the one meeting that I was invited to. As head of the signal services unit at Galt Associates, I intend to speak with Dr. Madigan in the near future about some potential collaborations. For me, the program made it apparent that I have a potential academic partner resource with an excellent understanding of safety signaling."
John Clark, MD
Galt Associates

"The DIMACS program has over the years played a role in informing our several presentations/conferences on mathematical modeling of infectious diseases for the intelligence community of the federal government." Alfred D. Steinberg, M.D., Mitre

"My participation in the BIEWG [The Biosurveillance and Information Exchange Working Group] did have a positive impact on my practice of public health. Through the meeting, I built new relationships and rekindled others with epidemiologists doing the same sort of work and struggling with the same sorts of issues. It was (and hopefully will remain) a nice forum to discuss issues of importance to syndromic surveillance monitors. I was also better acquainted with the CDC biosense team of monitors. All of these collaborations were positive developments. I am working with Howard Burkom of Johns Hopkins to put together a poster at the Syndromic Surveillance Conference. The idea was born before our DIMACS meeting, however our collaboration was strengthened there. I would also note that communications between myself and the other biosurveillance monitors across the country that participated has really increased. Although I don't have research or products to point to, I feel it positively impacted my work." Michael A. Coletta, MPH, Enhanced Surveillance Coordinator, Division of Surveillance and Investigation Office of Epidemiology,Virginia Department of Health

Contributions To Human Resources Development

Many graduate students, undergraduates, and several postdocs participate in the program. Local graduate students and many non-local students were also involved as visitors and workshop/working group attendees. The project has two primary postdocs this year, Kah Loon Ng and Ivan Zorych, as well as participation by visitors James Abello and Nina Fefferman 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.

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.

Summer 2005:

Miguel Mosteiro, Rutgers, CS
"Lower and upper bounds for broadcast and MIS in sensor networks"

Paul Raff, Rutgers, Math
"Fire-Fighter Problem"

Winter 05/06:

Suhrid Balakrishnan, Rutgers, CS
"A computational approach to discover gene essentiality"

Rohan Fernandez, Rutgers, CS
"Algorithms and lower bounds"

Pai-His Huang, Rutgers, CS
"Problem of classifying sequenses of variable length"

Miguel Mosteiro, Rutgers, CS

"To reduce remaining gap between the upper and lower bounds"

Paul Raff, Rutgers, Math
"Firefighter problem"

Also the following student was a Graduate Assistant under David Madigan under this project:

Srinivas Maloor, Rutgers
9/1/05-8/31/06

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 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 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. Cabalbo is a junior researcher (recent Ph.D.) and this long-term visit led to a totally new direction of research for him.

"Over the past year my participation in the DIMACS focus on Computational and Mathematical Epidemiology has greatly, positively affected the course of my research. The open and interdisciplinary environment has provided opportunities for me to meet others from a variety of fields I would have otherwise never encountered and form new collaborations with them. The resulting investigations are already, in less than a year, being presented to the greater scientific community at conferences, seminars and journal publications." Nina Fefferman, DIMACS Visitor

The impact on the careers of the students and postdocs is illustrated by a few examples.

"While at DIMACS, under the auspices of the focus on Computational and Mathematical Epidemiology, I have become involved in a collaboration with Dr. Kah Loon Ng, a very talented post-doctoral graph theorist. Together, we have begun investigating the role of individual-level behavioral choice on the stability of social networks. These networks provide the means of transmission of infectious disease, leading to a new level of understanding of different patterns of disease spread in species with different social behaviors. We are just now beginning to use this understanding to evaluate inter-species transmission risks for zoonotic (arising from animals and leading to human exposure) infections for diseases with multiple animal hosts (e.g. avian influenza, west nile virus, etc.). The preliminary results of these studies are being prepared for publication now with an expectation that they will be submitted for review in the next few weeks. They will also be presented by Dr. Ng in a talk entitled "Comparative effects of different centrality measures in dynamic networks" (with an associated published abstract; included below) at the upcoming conference (24^{th} – 30^{th} of April, 2006) of the International Network for Social Network Analysis. We expect to continue this work throughout the coming summer, involving undergraduate students from the DIMACS sponsored Research Experience for Undergraduates and from the Tufts University Summer Scholars program. I am particularly excited that this project will involve young researchers and I am grateful to DIMACS for providing this program."
Nina Fefferman, DIMACS Visitor, Tufts University

"My postdoc under the auspices of this program has started a collaboration with Dr. Daniel Rubenstein, a behavioral ecologist and the Head of the Department of Ecology and Evolutionary Biology at Princeton. Together with Dr. Rubenstein's graduate students Siva Sundaresan and Ilya Fischoff, I am working on

developing computational techniques to analyze the dynamic social interactions of animal populations. The techniques we are developing are applicable to many questions about animal and human populations, including epidemiology. The research project that resulted from the collaboration initially facilitated by DIMACS has just received funding from Microsoft Research (Bioinformatics program)."
Tanya Berger-Wolf, Assistant Professor, Department of Computer Science, University of Illinois at Chicago (DIMACS post-doc 2004/05)

"I don't know exactly which part of DIMACS programs was responsible for it, but the DIMACS postdoc was one of the crucial points that are shaping my career (I am not joking). The collaboration I started with the Princeton biologists while at DIMACS has become the central direction of my research. (The collaboration with Daniel Rubenstein is still very much active.)"
Tanya Berger-Wolf
Department of Computer Science
University of Illinois at Chicago

"The DIMACS meeting was a great opportunity to meet other colleagues. This yielded a collaboration with Troy Day (manuscript in preparation of the evolutionary epidemiology of vaccination) and allowed me to meet with young researchers who are now planning to do a postdoc with me."
Sylvain Gandon, CNRS

Faculty and other researchers have reported that participation in the special focus programs has positively influenced their teaching and led to new interdisciplinary projects with their students. Here are some examples:

"I was a participant in a program several years ago on Mathematical Modeling of Infectious Diseases organized by Herb Hethcote. Following that seminar, I have engaged students in studies using mathematical models as one of their study aims in research projects. In one particular case, I worked with the student to develop a model to assess the effect of HIV on tuberculosis prevention and control in Sub-Saharan Africa. The student performed a population-based survey to collect information on age-specific prevalence of tuberculosis and used this in his mathematical model. The final product of this activity was a publication in Preventive Medicine." Christopher C. Whalen, M.D., M.S. Professor Director, Division of Epidemiology Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine

"Participating in the workshop on Mathematical Epidemiology has had a positive impact on my teaching although not yet on my research. I teach our sophmore/junior level differential equations course and now include a much bigger focus on mathematical modeling. I include a section on the SIR model. Some of my students are doing a course project on HIV/AIDs modeling (using a model developed at COMAP)."
Victor Donnay, Professor of Mathematics, Bryn Mawr College

"The DIMACS conference made me aware of additional surveillance research… In turn, this has influenced the literature review performed by one of the Masters students whom I advise, Ms. Yanna Shen, for her MS project report. I also met Drs. Martin Kulldorf and Henry Rolka at DIMACS, which led to our jointly writing a biosurveillance paper (also with Drs. David Madigan and Howard Burkom) that has been submitted to a journal for review."
Greg Cooper, University of Pittsburgh

"For me, the primary value of the symposia has been educational. I have met a number of new colleagues and been introduced to research methods that I was otherwise completely or partly unaware of.

At this point, I now have a number of approaches to consider in my epidemiologic work (in industry), along with colleagues with whom to potentially work with in the future. I have also given several talks at DIMACS symposia, which I have enjoyed presenting, and which I hope others have enjoyed hearing."
Walter L. Straus, MD MPH, Senior Director, Epidemiologic Research, Merck Research Laboratories

"Investigating the role of social interactions in the dynamics of infectious disease while at DIMACS, I have collaborated with Dr. Elena Naumova (Tufts Univ. School of Medicine), to create models that reveal that demographic heterogeneity in a population can, itself, cause seasonal incidence patterns (as with those observed in influenza). This result has been submitted for publication under the title "The role of social interactions among etiologically distinct subpopulations in the seasonality of infectious disease incidence" and is currently under review.

"While at DIMACS, I have begun work with Dr. Sam Behers at the University of Illinois at Urbana-Champaign, making mathematical models of social insect species to investigate the impact of different social organizations to the ability of the species to withstand disease threats. The results from these investigations are in preparation for publication and were presented (by myself) in an invited talk to the departments of Entomology and Center for Infectious Disease Dynamics (CIDD) at Penn State University entitled "How Would Termites Prepare for Pandemic Bird Flu and What Should We Learn From Them" (abstract included below). After meeting at a DIMACS conference on the economics of epidemiology, Dr. Ramanan Laxminarayan (of Resources for the Future) and I have begun collaborating on extending this research to investigate the different impacts of disease spread based on different organizational structures of large businesses.

"Throughout this past year of my visit at DIMACS, I have presented my ongoing work at the Computational and Mathematical Epidemiology Seminar Series in talks entitled, "Disease Signatures" on Sept. 19th, "Termites in the Nation's Service" on Oct. 17th, and "Different Scales of BioDefense - Can societies be both safe and efficient?" on Feb. 13th. I have also delivered a talk entitled "Selected Problems in Epidemiology" (abstract included below) on the uses of Data Mining in Epidemiology to the DIMACS tutorial on this topic organized by Graham Cormode and James Abello. At this workshop, I was lucky enough to meet Dr. T. Kostova and we have begun serious plans to collaborate in the area of mathematically modeling the adaptive immunological responses in humans to evolving pathogens.

"All of these projects and collaborations were made possible by my involvement with DIMACS over the past nine months. I feel that my individual perspective, even of my own, individual research has been broadened substantially by this opportunity and I am very grateful. I feel lucky to have been a part of this focus in Computational and Mathematical Epidemiology and am looking forward to continuing with these projects and these people."
Nina Fefferman, DIMACS Visitor, Tufts University