DIMACS Center
Rutgers University

Special Focus on Computational and Mathematical Epidemiology

Annual Report

August 2009

Participants who spent 160 hours or more:

Fred S. Roberts, Rutgers University, Principal Investigator
Nina Fefferman, Rutgers University
Paul Kantor, Rutgers University
(participant in supplemental project on Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links)
William Pottenger, Rutgers University
(participant in supplemental project on Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links)
David Lewis (private consultant),
(participant in supplemental project on Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links)
Tong Zhang, Rutgers University
(participant in supplemental project on Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links)
Paul Raff, Rutgers University
(participant in supplemental project on Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links)
Mark Dilsizian, Lehigh University and Rutgers University
(participant in supplemental project on Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links)

Other Participants:

James Abello, Rutgers University
Martin Farach-Colton, Rutgers University
Sunetra Gupta, Oxford University
Donald Hoover, Rutgers University
David Krakauer, Santa Fe Institute
Simon Levin, Princeton University
Marc Lipsitch, Harvard School of Public Health
David Madigan, Columbia University
Ilya Muchnik, developing a cancer registry modeling project with James Abello and graduate student David Millman
Megan Murray, Harvard School of Public Health
S. Muthukrishnan, Rutgers University
David Ozonoff, Boston University
Burton Singer, Princeton University
Daniel Wartenberg, University of Medicine and Dentistry of New Jersey

Working Group Meeting: Spatio-Temporal and Network Modeling of Diseases (Third Meeting)
Dates: October 21 - 25, 2008

Organizers:

Martin Eichner, Tubingen University
Nina Fefferman, DIMACS
Valerie Isham, University College London
Alun Lloyd, North Carolina State University

Denis Mollison, Heriot-Watt University

DIMACS/DyDAn Working Group: Network Models of Biological and Social Contagion
Dates: November 3 - 4, 2008

Organizers:

Lauren Ancel Meyers, University of Texas at Austin
Michelle Girvan, University of Maryland

Workshop: Stochasticity in Population and Disease Dynamics
Dates: December 8 - 10, 2008

Organizers:

Jonathan Dushoff, McMaster University
Todd Parsons, University of Pennsylvania
Joshua Plotkin, University of Pennsylvania

Workshop: Models/Methodological Problems of Botanical Epidemiology
Dates: March 16 - 18, 2009

Organizer:

Chris Gilligan, Cambridge

Tutorial: Statistical De-identification of Confidential Health Data with Application to the HIPAA Privacy Regulations
Dates: April 30 - May 1, 2009

Organizers:

Daniel Barth-Jones, Columbia University
Alina Campan, Northern Kentucky University
Traian Marius Truta, Northern Kentucky University

Workshop: Identifying Genetic Signatures for the Evolution of Complex Phenotypes
Dates: June 11 - 12, 2009

Organizers:

Gyan Bhanot, Rutgers, Cancer Institute of New Jersey and Institute for Advanced Study
Raul Rabadan, Columbia University

Visitors:

Jenny McNulty, University of Montana
Site: Piscataway
Host: Nina Fefferman
Dates: 9/1/2008-8/31/09

Boris Mirkin, Birkbeck College
Site: Piscataway
Host: Fred Roberts
Dates: 9/29/08-10/6/08

Participants in Supplemental Project on Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links:

Fred Roberts, Rutgers
Paul Kantor, Rutgers
David Madigan, Columbia-Rutgers
William Pottenger, Rutgers
Tong Zhang, Rutgers
David Lewis, David Lewis Associates
Vladimir Menkov, Consultant - Programmer
Mark Dilsizian, Lehigh and Rutgers
Paul Raff, Rutgers

Graduate students:

Mark Dilsizian, Lehigh and Rutgers
Paul Raff, Rutgers

Other Collaborators

Daniel Bath-Jones, Columbia University, Co-Organizer, Tutorial: Statistical De-identification of Confidential Health Data with Application to the HIPAA Privacy Regulations

Gyan Bhanot, Rutgers, Cancer Institute of New Jersey and Institute for Advanced Study, Co-Organizer, Workshop: Indentifying Genetic Signatures for the Evolution of Complex Phenotypes

Alina Campan, Northern Kentucky University, Co-Organizer, Tutorial: Statistical De-identification of Confidential Health Data with Application to the HIPAA Privacy Regulations

Graham Cormode, Bell Laboratories and AT&T 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, Queens University, Co-Organizer, Workshop: Evolutionary Considerations in Vaccine Use

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

Jonathan Dushoff, McMaster University, Co-Organizer of Workshop: Stochasticity in Population and Disease Dynamics

Martin Eichner, Tubingen University, Co-Organizer of Working Group Meeting: Spatio-Temporal and Network Modeling of Diseases (Third Meeting)

Nina Fefferman, DIMACS, Co-Organizer of Working Group Meeting: Spatio-Temporal and Network

Modeling of Diseases (Third Meeting)

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

Chris Gilligan, Cambridge, Organizer, Workshop: Models/Methodological Problems of Botanical Epidemiology

Michelle Girvan, University of Maryland, Co-Organizer of DIMACS/DyDAn Working Group: Network Models of Biological and Social Contagion

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

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

Valerie Isham, University College London, Co-Organizer of Working Group Meeting: Spatio-Temporal and Network Modeling of Diseases (Third Meeting)

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

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

Alun Llyod, North Carolina State University, Co-Organizer of Working Group Meeting: Spatio-Temporal and Network Modeling of Diseases (Third Meeting)

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

Denis Mollison, Heriot-Watt University, Co-Organizer of Working Group Meeting: Spatio-Temporal and Network Modeling of Diseases (Third Meeting)

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

Lauren Ancel Meyers, University of Texas at Austin, Co-Organizer of DIMACS/DyDAn Working Group: Network Models of Biological and Social Contagion

Todd Parsons, University of Pennsylvania, Co-Organizer of Workshop: Stochasticity in Population and Disease Dynamics

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

Joshua Plotkin, University of Pennsylvania, Co-Organizer of Workshop: Stochasticity in Population and Disease Dynamics

Raul Rabadan, Columbia University, Co-Organizer, Workshop: Indentifying Genetic Signatures for the Evolution of Complex Phenotypes

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, National Institute of Health (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, Lehman College, Organizer of Tutorial: Phylogenetic Trees and Rapidly Evolving Pathogens

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

Traian Marius Truta, Northern Kentucky University, Co-Organizer, Tutorial: Statistical De-identification of Confidential Health Data with Application to the HIPAA Privacy Regulations

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.

Alcatel-Lucent, 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

This special focus, which started in summer of 2002, is 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. This special focus has been doing 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 needed 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 are considering 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.

A major area of emphasis in the special focus is analysis of health care data. This involves issues of data mining, data sharing, data privacy, etc. A direct result of this emphasis is a new partnership between DIMACS and the Cancer Institute of New Jersey (CINJ). CINJ is a Center of Excellence of UMDNJ-Robert Wood Johnson Medical School and the first and only National Cancer Institute-designated Comprehensive Cancer Center in New Jersey. CINJ has become a full partner of DIMACS and the two institutions have already run joint activities and submitted joint proposals. While the emphasis in this

special focus has been on infectious diseases, we have had a number of cancer-related meetings and working groups. Often times, information collected from clinical, genetic and laboratory tests resides in different locations and due to privacy and security concerns, cannot be easily accessed via the "information superhighway." This often slows the sharing of information and keeps barriers between researchers. By utilizing the latest technology to collect, correlate, analyze and assimilate a full range of biomedical data in a secure manner, researchers will be able to better understand the causes of diseases such as cancer. With such integration, there is also the need to make sure that patient privacy is protected. As advanced technology continues to allow for the simplification and collection of massive amounts of data in the area of health care and research, experts in the field are looking at how to best streamline that process. That is why CINJ and DIMACS, under another grant, recently held a "Biomedical Informatics Summit," bringing together intellectual leaders from a variety of disciplines to help address key challenges in the gathering and analysis of health care information for the benefit of patients.

The tutorials, workshops, and working group meetings that were held during this year are as follows:

Working Group Meeting: Spatio-Temporal and Network Modeling of Diseases (Third Meeting)
Dates: October 21 - 25, 2008
Location: Tubingen, Germany
Organizers: Martin Eichner, Tubingen University; Nina Fefferman, DIMACS; Valerie Isham, University College London; Alun Lloyd, North Carolina State University; Denis Mollison, Heriot-Watt University
Attendance: 43

This working group meeting pulled together the most promising emerging strands of research in spatial and network modeling of epidemics, with particular focus on relating models to data, and on the integration of modeling scales.

Additional foci for this meeting included using data sets to determine age specific differences in susceptibility and infectiousness for airborne infections, using auxiliary information when data is missing, optimal vaccination strategies, epidemic network and intervention modeling, and various control strategies for vector borne illnesses.

Much discussion took place about the top ten future disease threats, and issues of surveillance, vaccination, and growth estimations. The development and use of complex network models was paramount.

Public health and policy issues, including illegal drug use, influenza pandemic potential, and antibiotic resistance completed the agenda.

Working Group: Network Models of Biological and Social Contagion
Dates: November 3 - 4, 2008
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Lauren Ancel Meyers, University of Texas at Austin; and Michelle Girvan, University of Maryland
Attendance: 24

The spread of infectious diseases and the flow of ideas and information through populations both depend on the complex structure of the underlying network of interactions between individuals. Disease ecologists and sociologists have historically studied the dynamics of contagion using models that assume very simple population structures. Recently, however, network modeling has revolutionized both fields

by enabling the rigorous exploration of the relationship between complex individual-level behavior and the higher-level emergence of outbreaks. Statistical tools are used to infer network structure from often limited data. Data driven algorithms are used to generate realistic networks structures, while mathematical approximations predict transmission dynamics that draw from the methods of percolation theory and other areas of statistical physics. The network models are more complex than their mass-action predecessors, but are quite tractable, and often reduce to low-dimensional descriptions and allow for straightforward calculations of the dynamics of contagion.

The workshop focused on these network models and compared them to alternative approaches. Special attention was paid to how to infer the network structure from the sociological and/or epidemiological data the emergence of the structure from simple individual-level behavior, and predicting the dynamics of contagion from simple characterizations of the underlying network.

Participants and speakers concentrated on:

- Measuring and inferring network structure from data;
- Generative models of social and epidemiological networks;
- Modeling the dynamics of biological and social contagion on networks;
- Modeling feedback from contagion dynamics to the network structure;
- Model selection.

Workshop: Stochasticity in Population and Disease Dynamics

Dates: December 8 - 10, 2008
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Jonathan Dushoff, McMaster University; Todd Parsons, University of Pennsylvania; and Joshua Plotkin, University of Pennsylvania
Attendance: 22

Mathematical Models have been a vital tool in epidemiology since the classic paper of Kermack and McKendrick. In spite of almost a century of mathematical analysis of disease dynamics, fundamental questions regarding the expected duration of an epidemic or the likelihood of recurrent outbreaks remain largely unanswered. Stochastic effects – arising both endogenously from the demographics of finite and discrete populations, and from environmental variability can drastically change the behavior of already complex and nonlinear host-disease dynamics, creating intriguing mathematical problems with important consequences for public health.

This workshop brought together experts in mathematical stochastic population dynamics with epidemiologists, who shared knowledge and developed and analyzed new models to bring new analytical and statistical approaches to these questions in disease dynamics. For example, topics included learning from dog rabies, stochasticity in childhood diseases, and tracking the Cuban HIV epidemic.

Workshop: Models/Methodological Problems of Botanical Epidemiology

Dates: March 16 - 18, 2009
Location: DIMACS Center, CoRE Building, Rutgers University
Organizer: Chris Gilligan, Cambridge
Attendance: 25

This workshop included experts from the botanical epidemiology and the genetics communities together with computer scientists and mathematicians interested in modeling using differential equations, discrete systems, and stochastic processes to investigate methodological problems in the spread of disease in plants. Key modeling questions considered included:

- Invasion and persistence of plant disease in spatially-extended and heterogeneous environments, and the consequences for optimizing the spatial and temporal deployment of resistance genes;
- Minimizing the risk of pesticide resistance;
- Prediction of crop loss and yield;
- The effects of changing agricultural processes (GM technology, farm, field size, landscape mosaic) and changing climate on host crop, parasite, vectors and antagonist (biocontrol agents) dynamics;
- The constraints imposed by economic considerations;
- The interplay between population genetics, population dynamics and epidemiology.

Modeling issues confronted were

- The need to confront a complex of models for spatially-extended dynamics that goes across heterogeneous scales from microscopic behavior in soil or a plant, through single infected plants, to a disease patch, multiple patches, whole fields and regions;
- Periodic forcing due to seasonal change;
- Temporally disturbed environments with abrupt changes due to sowing, harvest, and switching on and off of favorable periods for transmission;
- Quenched systems such as host susceptibility changes with host age;
- Scaling from individual to population behavior;
- Development and testing of stochastic models for the evolution of probability distributions within and between replicated epidemics;
- Model reduction, including perturbation and asymptotics.

In addition, data and model testing issues, such as parameter estimation for spatially-explicit and spatially –implicit models, data collection for model testing and parameter estimation, optimization of experimental design for parameter estimation and model discrimination, and analysis of microcosm data to distinguish demographic and environmental stochasticity were covered in the workshop

Applications to tough botanical epidemiology problems such as stem rust, sudden oak death, citrus cankers, and fungicide resistance were considered in these modeling constructs. There was a strong focus on the relationships between climate change and food security and effective models to predict potential problems.

Tutorial: Statistical De-identification of Confidential Health Data with Application to the HIPAA Privacy Regulations

Dates: April 30 - May 1, 2009
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Daniel Barth-Jones, Columbia University; Alina Campan, Northern Kentucky University; and Traian Marius Truta, Northern Kentucky University
Attendance: 32

This tutorial provided researchers, analysts, and managers with an overview of the federal HIPAA Privacy regulations and an introduction to the principles and methods of statistical disclosure limitation that can be used to statistically de-identify health care data to meet the HIPAA privacy regulations.

The focus was on statistical disclosure for various data models (tabular data, microdata, social networks), especially microdata in healthcare databases. The tutorial was designed to be practical and applied, focusing primarily on providing participants with the knowledge needed to understand the statistical de-identification process for healthcare datasets in accordance with the HIPAA privacy rule and to identify confidentiality issues.

As a result of this tutorial, participants will be able to:

- Understand the permissible uses of healthcare data for various purposes under the HIPAA privacy regulations;
- Conceptualize and document data intrusion scenarios;
- Understand the basic principles behind the conduct and documentation of statistical disclosure analyses measuring disclosure risks;
- Understand various disclosure limitation methods;
- Appreciate the associated trade-offs between disclosure risks and statistical information quality.

Workshop: Identifying Genetic Signatures for the Evolution of Complex Phenotypes

Dates: June 11 - 12, 2009
Location: DIMACS Center, CoRE Building, Rutgers University
Organizers: Gyan Bhanot, Rutgers University; Cancer Institute of New Jersey and the Institute for Advanced Study; and Raul Rabadan, Columbia University
Attendance: 80

The focus of this workshop was on methods used to understand the evolution of complex phenotypes in SNP, sequence and gene and protein expression data, and on results obtained by applying such methods to private as well as public datasets. A broad range of topics included the methods to identify selection events in the presence of mutation and recombination, disease and complex phenotype association studies, co-evolution and co-adaptation in host-pathogen systems, correlations between human migration and section, methods to identify evolutionary hotspots, protein evolution, haplotype selection in pathways and signal speciation. Examples of the application of these techniques to uncovering the genes associated with autism, determining the genetic variants for cellular stress response and cancer risk, and understanding master regulators of human malignancy signatures, innate immunity and RNA viruses.

A direct outcome of this special focus was a series of grants for workshops, student training courses, and small research projects emphasizing US and African faculty and students, and stressing topics in computational and mathematical epidemiology. The latest grant helped to establish the *US-African Biomathematics Initiative.* This DIMACS program, joint with the Mathematical Biological Sciences Institute (MBI) at Ohio State University, consisted of three parts during the current year:

A Modeling Data Clinic on Meaningful Modeling of Biological Data

Dates: May 11-19, 2009
Location: Muizenberg, South Africa
Organizers: Steve Bellan, UC Berkeley; Wim Delva, Ghent University; Jonathan Dushoff, McMaster University; Avner Friedman, Ohio State University; Marty Golubitsky, MBI; John Hargrove, SACEMA; Travis Porco, UC San Francisco; Juliet Pulliam, JIH; Fred Roberts, DIMACS, Rutgers University; Brian Williams, WHO retired
Number of participants: 30

Successful model development should never be divorced from data, but sometimes is. Abstract mathematical models, unchallenged by data, do not generally provide the basis required for providing evidence-based advice to policy makers in the fields of biology, epidemiology and medicine. In fact, in some cases they can interfere with understanding and mislead researchers and policy makers. Past institutes and conferences sponsored by DIMACS, in collaboration with African partners, showcased the burgeoning talent of young researchers with an impressive array of oral and poster presentations that highlighted the application of mathematical models to a variety of medical, epidemiological, and biological problems. However, the models were not adequately tested by being fitted to the data. It

became apparent that the next step would be to ensure that such talented computer scientists and mathematicians were engaged in real biological data and questions in a meaningful way. This modeling clinic addressed these problems in an experimental effort at involving participants in analyzing real data sets and fitting models. We hope this effort will be replicated in the future.

Advanced Study Institute and Workshop on Economic Epidemiology
Dates: July 20-31, 2009
Location: Makerere University in Kampala, Uganda
Workshop on Economic Epidemiology
Dates: August 3-5, 12009
Location: Makerere University
Organizers: Nina Fefferman, DIMACS, Rutgers University; Alison Galvani, Yale University; Wayne Getz, UC Berkeley; Abba Gumel, University of Manitoba; Ramanan Laxminarayan, Resources for the Future; Simon Levin, Princeton University; Jan Medlock, Clemson University; Joseph Mugisha, Makeree University; Fred Roberts, DIMACS; David Smith, University of Florida

This two week Institute for US, African, and Canadian graduate students, with a following "capstone" Workshop. This activity followed on the DIMACS workshop on Economic Epidemiology funded through the special focus. Economic epidemiology deals with the mathematical conceptualization of the interplay among economics, community organization, individual human behavior, and disease ecology to improve our understanding of the emergence, persistence, and spread of infectious agents and of optimal strategies and policy to control that spread. Mathematical models of disease spread already exist to allow the examination of the relative efficacy of particular intervention strategies at curtailing disease spread. However, these models frequently assume unmotivated levels of behavioral compliance, making their results difficult to interpret in real-world scenarios. To correctly evaluate health interventions and alternative public policies, models of disease spread must incorporate both group and individual behaviors (which are often the result of economic, and therefore quantifiable, considerations). Incorporating these behaviors entails important and complex mathematical challenges but is necessary in order to understand which of the theoretically efficient policies could result in the most effective real-world disease control.

The field of Economic Epidemiology is really taking off and we can thank the special focus for helping to establish it.

Supplemental project on Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links

During this year, we continued work on the project on "Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links" that was funded by the intelligence community through their KDD program. The work centers around learning patterns from (often sparse and expensive) bodies of analyzed materials. The problems are conceptually quite distinct, although several of our solutions make use of, and extend the power of, the Bayesian Regression software (BXR) that we have developed in earlier intelligence-community-sponsored work. The primary task has been to integrate BXR more fully with Blackbook and in particular tightly couple the two systems' semantic models. New work has extended BXR to deal with new training data and new class definitions in real time and to in general produce a more stream- oriented API for BXR; and we have developed methods for learning patterns that involve higher-order relationships among real or aliased entities and are integrating this new work into Blackbook.

Integration into Blackbook involves data modeling: understanding approaches used by Blackbook, developing conventions for communicating this to BXR, and defining mappings back from BXR to the

Blackbook semantic model. It also involves data formats and communication, system architecture, and data management. Our software is being developed so that it is ready to accept information from Blackbook as an interface to the analyst, and we have made modifications that will greatly simplify the process of computing the data for higher order paths. In connection with this work, we sent one of our graduate students, Mark Dilsizian, to take a course on Semantic Web w/RDF andOWL, given at the University of Texas-Dallas, and run by developers of Blackbook.

We have also worked on a task we call "learning with emergent classes and training data." When information streams are processed over time, both the properties of the data and the classes of interest may change. We are extending our algorithms and software to situations where new training data, and new class distinctions to make are provided incrementally. We are developing online learning algorithms that leverage both changes in stream properties and changes in discrimination sets to improve effectiveness. We have developed and implemented, in BXRtrain, an algorithm for learning from data streams with multiple and emerging classes, while supporting BXRtrain's Bayesian priors.

A final task is concerned with "higher-order path analysis." In addition to knowing how many agents there are at play, and knowing which of them are responsible for which particular products or activities, it is also extremely important to know the links and relationships among these agents. One way to characterize the topology of the relationships in a graph is based on higher-order path analysis techniques. Thus, the strength of connections between entities reflects not only the most direct path, but indirect paths as well. We have developed ways to apply unsupervised higher-order path methods to measure the strength of links between entities/agents and integrated higher-order path statistics into BXR.

Findings

A variety of new results have occurred as a result of the workshops on Computational and Mathematical Epidemiology. Some of the examples follow:

1. **Sequential Decision Making in Epidemiology**

Many problems of epidemiology and public health can be looked at as sequential decision making problems where one decision is made at a time and then later decisions depend upon feedback from earlier ones. Sequential decision making is an old subject with applications in other fields as well, but one that has become increasingly important, with the need for new models and algorithms arising as traditional methods for making decisions sequentially fail to keep pace with problem scale. Making use of methods developed for sequential decision making in the context of container inspection – supported by a synergistic NSF grant – Fred Roberts has applied the sequential decision heuristics to epidemiology and public health. Imagine that we have several potential interventions for a public health crisis. Assume funds limit us to one intervention at a time. Which intervention do we invest in first? On the basis of the outcome of the first intervention, which do we launch next? Interventions are expensive. So are false positive and false negative assessments of the outcome of our interventions. “Cost” is a combination of cost of the intervention and cost of false results. In what order should we launch the interventions in order to minimize total “cost”? We are looking to determine if an epidemic can be controlled. We have a variety of interventions to choose from. In the end, the epidemic is to be classified into one of several categories. In the simplest case, we ask: if the epidemic is in category 0 = controllable, or category 1 = not controllable. An intervention scheme specifies which interventions are to be made based on assessments of previous interventions.

A similar problem arises in medical decision making. A physician is looking to determine if a patient has disease x. The doctor has a variety of tests to choose from. In the end, the patient is to be classified into one of several categories. In the simplest case, we have 0 = “doesn’t have the

disease", 1 = "does have the disease." A testing scheme specifies which tests are to be made based on previous observations.

Roberts formulated these problems as ones of finding a "least cost" Boolean decision function and applied algorithms using simulated annealing developed in the container inspection context to it.

An article describing this is in preparation and the PI gave a talk about this topic at an international workshop on Economic Epidemiology in August 2009.

2. **Infectious Disease and Productivity Tradeoffs**

Epidemiological models have been employed with great success to explore the efficacy of alternative strategies at combating disease outbreaks. Many such models have incorporated an understanding of age-based susceptibility and severity of outcome, trying to determine how best to limit both/either the overall infection within the population and/or the number of serious adverse outcomes. Models built from this perspective frequently recommend the preferential treatment/vaccination of children or elderly, demonstrating how prevention of serious disease within these etiological subgroups can provide both protection within the subgroup itself and indirect protection to the broader population. However it is most frequently the case that these target populations are consumers, rather than providers, of household resources. This distinction implies that, though these results may hold in non-resource limited households, in areas of the globe where continued health of all household members relies on continued provision of resources these models may fail to provide the most effective overall strategies for optimal health outcomes in both target populations and the broader community. DIMACS Visitor Jacques Kibambe Ngoie (U. of Pretoria) and member Nina Fefferman (Rutgers) developed a modified SIR model targeting epidemiological control in resources-limited populations and demonstrated how economic limitation may shift the optimal strategy towards populations at lesser direct risk of serious outcome from infection if they are responsible for providing secondary protection to the health of higher-risk consumers by continuing to produce household resources. Their results demonstrate how household resource limitation can drastically affect the impact of targeted treatment strategies for limiting epidemics.

3. **Community Economics and HIV**

Human immunodeficiency virus (HIV) is one of the most severe and deadly pandemics that this world has ever seen. In some countries, a majority of the HIV infected patients have a way of obtaining treatment so that their virus does not progress into acquired immunodeficiency syndrome (AIDS). However, in many impoverished third world countries monetary resources are very limited and communities are unable to provide treatment for all of the infected population A DIMACS REU student, Immanuel Williams (University of Maryland-Baltimore County), also supported through the Rutgers RISE (Research in Science and Engineering) program aimed at introducing underrepresented groups to research in STEM disciplines, together with mentors Nina Fefferman (Rutgers) and Tami Carpenter (Rutgers), developed a mathematical epidemiological model to simulate an HIV outbreak in a third world country with a community-based economy in which monetary resources that come into the community will be shared by everyone. Examples of such economies might include a small village or an extended family. The epidemiological model includes the interplay between the economic productivity of the community and its ability to provide treatment to the infected population. Within this model there are providers (people who contribute financially to the system) and there are consumers (such as children who consume) and these individuals may be infected or may not. While producers provide immediate

economic benefit to the community, the consumers will provide future economic benefit if they remain sufficiently healthy. Thus, this model is used to gain intuition to potentially inform policy decisions on how to allocate monetary resources between the infected providers and infected consumers. This model is also used to provide a treatment strategy that best assures the stability and long-term survival of the community.

4. **Differential neutralization efficiency of hemagglutinin epitopes, antibody interference, and the design of influenza vaccines**

It is generally assumed that amino acid mutations in the surface protein, hemagglutinin (HA), of influenza viruses allow these viruses to circumvent neutralization by antibodies induced during infection. However, empirical data on circulating influenza viruses show that certain amino acid changes to HA actually increase the efficiency of neutralization of the mutated virus by antibodies raised against the parent virus. Special Focus consultant/organizer Simon Levin and his students Wilfred Ndifon and Ned S. Wingreen demonstrated that this surprising increase in neutralization efficiency after HA mutation could reflect steric interference between antibodies. Specifically, if there is a steric competition for binding to HA by antibodies with different neutralization efficiencies, then a mutation that reduces the binding of antibodies with low neutralization efficiencies could increase overall viral neutralization. Levin, et al. used a mathematical model of virus–antibody interaction to elucidate the conditions under which amino acid mutations to HA could lead to an increase in viral neutralization. Using insights gained from the model, together with genetic and structural data, they predicted that amino acid mutations to epitopes C and E of the HA of influenza A/H3N2 viruses could lead on average to an increase in the neutralization of the mutated viruses. They presented data supporting this prediction and analyzed the implications for the design of more effective vaccines against influenza viruses and other pathogens.

5. **The use of hemagglutination-inhibition for influenza surveillance: Surveillance data are predictive of influenza vaccine effectiveness**

The hemagglutination-inhibition (HI) assay is the main tool used by epidemiologists to quantify antigenic differences between circulating influenza virus strains, with the goal of selecting suitable vaccine strains. However, such quantitative measures of antigenic difference were recently shown to have poor predictive accuracy with respect to influenza vaccine effectiveness (VE) in healthy adults. Simon Levin (Princeton), his student Wilfred Ndifon, and DIMACS Visitor Jonathan Dushoff (McMaster University) re-examined those results using a more rigorous criterion for predictive accuracy – considering only cases when the vaccine (V) and dominant (D) circulating strains are antigenically different – and greater numbers of HI titers. They found that the Archetti–Horsfall measure of antigenic difference, which is based on both the normalized HI titer (NHI) of D relative to antisera raised against V and the NHI of V relative to D, predicts VE very well ($R2 = 0.62$, $p = 4.1\times10{-}3$). In contrast, the predictive accuracies of the NHI of D relative to V alone ($R2 = 0.01$), and two other measures of antigenic difference based on the amino acid sequence of influenza virus hemagglutinin ($R2 = 0.03$ for both measures) are relatively poor. Furthermore, while VE in the elderly is generally high in cases when D and V are antigenically identical (VE = 35%, S.E. = 5%), in other cases VE appears to increase with the antigenic difference between D and V ($R2 = 0.90$, $p = 2.5\times10{-}5$). This paradoxical observation could reflect the confounding effects of prior immunity on estimates of VE in the elderly. Together, the results underscore the need for consistently accurate selection of suitable vaccine strains. Levin, et al. suggest directions for further studies aimed at improving vaccine-strain selection and present a large collection of HI titers that will be useful to such studies.

6. **The BOXER Software**

In the supplemental project "Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links," we built a new software package called BOXER, which can learn binary and multiple classification distinctions from a stream of labeled data. At any point during the learning process, BOXER can issue the current state of its model for use in other calculations. The model itself is a Bayesian logistic regression model, which calculates the probability that an item belongs to a class, based on numerical features characterizing the item. The software maintains a very small "footprint" using a technique called truncated gradients search. This technique ensures that obsolete information is gradually removed from the model. The software has been integrated into a new framework called Blackbook, which has been developed to support the government activities that require the use of data from a wide variety of sources.

Training & Development Section

All of the Workshops have opened doors for researchers and students to work collaboratively together, in a truly integrated fashion.

Two postdoctoral visitors spent the year at DIMACS, partly supported by this grant. Postdoc Debbie Yuster worked on a variety of projects. One of them involved using generating functions in DNA sequence analysis, where she has collaborated with Dr. Gyan Bhanot (Rutgers University and University of Medicine and Dentistry of NJ), on using generating functions to analyze expected frequency of certain motifs in viral DNA. Dr. Bhanot's group is currently relying on Monte Carlo simulation, and Yuster is working to confirm their results rigorously.

Postdoc Paola Vera Licona worked on a variety of projects. Many of these dealt with biological networks and she was especially interested in problems dealing with reverse-engineering for the modeling of such networks, within a polynomial dynamical systems framework.

Steve Bellan, Ph.D student in Environmental Science, Policy and Management at UC Berkeley, describes his growth through participation in DIMACS workshops by saying the following: "This work has led me to think about vector-borne diseases more thoroughly and led to conversations with colleagues on the importance of mosquito demograph in vector-borne disease models. At this workshop and through the additional work I have become increasingly aware of the disconnect between mathematical epidemiology, classical epidemiology, and applied disease control. I am intrigued by the possibility of connecting these three fields through research and teaching."

Outreach Activities

This Special Focus is closely intertwined with our Center's efforts to link mathematics and computer science with biology in the high schools. The DIMACS Bio-Math Connection (BMC) is aimed at introducing high school math, computer science, and biology teachers to topics at the interface. This project is informing the BMC effort and specific topics from the project are being adapted for use in high schools. The materials developed by BMC participants consist of modules that can be flexibly adapted for use in a variety of courses at a variety of grade levels in both biology and mathematics. The project is run by DIMACS in collaboration with the Consortium for Mathematics and its Applications (COMAP) and Colorado State University. More information on BMC is available at http://dimacs.rutgers.edu/BMC//index.html Roughly sixty teachers have been trained to use modules integrating mathematics and biology in computational biology, epidemiology, and ecology. The modules in computational biology and epidemiology will be available this year for widespread nationwide use.

The predecessor to the BMC program led Midge Cozzens and Fred Roberts to edit a book on "Biomath in the Schools" for the DIMACS book series published by the American Mathematical Society. DIMACS postdocs Debbie Yuster and Dan Cranston became involved with the DIMACS Bio-Math program through this book. A high school science teacher, Consuelo Rogers from Hawaii, who had participated in several DIMACS biomath summer programs, wrote an article about her experiences in introducing the biology-math interface into her school. Since Yuster expressed interest in DIMACS educational programs, we suggested she work with Rogers to write up this activity in a way that was publishable. That has now led to a very nice paper entitled "Insights from Math-Science Collaboration at the High School Level." During summer 2004, Charles Mullins, a teacher at the Arkansas School for Mathematics, Sciences, and the Arts, participated in a DIMACS program that emphasized bio-math, and in particular discrete mathematical problems in biology. He returned to school and supervised a research project on the superstring problem, the problem of finding the shortest "superstring" that contains all strings in a set. This problem arises from cutting the DNA molecule into short strands. The project of Mullins' student involved finding out how good an approximation to the optimal superstring can be obtained using a greedy algorithm. Since Cranston expressed interest in DIMACS educational programs, we suggested he work with Mullins to write up this activity. He helped to define the problem and the algorithm for solving it, and to explain the proof that it always obtains an optimal solution for a certain class of strings. He also helped to explain the process used by Mullins to mentor the student in this activity. The result is a paper, "Research at AsMSA based on the DIMACS Biomath program."

Researchers from Howard University and Morgan State University spent time at DIMACS in the summer of 2007 working on "Potential uses of entropy in biosurveillance" through a Department of Homeland Security-sponsored program aimed at faculty and students from minority-serving institutions. Nina Fefferman (DIMACS) and a team of two professors, two graduate students, and two undergraduate students looked into coding theory techniques that can be applied to epidemiology. They hypothesized that the signal-to-noise ratio in reported disease incidence data may increase during the occurrence of disease outbreaks and that these increases may be observable when analyzed using entropy measures. If the hypothesized increases are statistically quantifiable, and if they are found to occur early enough in the progression of disease throughout the affected population, this may prove an invaluable tool for early-warning biosurveillance. The initial results were so promising that this work is now continuing through the newly-founded Department of Homeland Security Center for Excellence for Command, Control, and Interoperability based at DIMACS. It also led one of the undergraduates, Ashley Crump from Howard, to apply to graduate school at Princeton, where she is now going into mathematical/computational epidemiology under the direction of Simon Levin. It led the other undergraduate, Nakeya Williams from Morgan State, to enter a graduate program at North Carolina State. The two graduate students involved, Anthony Obukba from Morgan State and Devroy McFarlane from Howard are actively working in epidemiological modeling and Devroy (and Ashley) were participants in our Economic Epidemiology Advanced Study Institute and Workshop in Uganda in the summer of 2009. This activity has been closely coordinated with and is a clear offshoot of our NSF grant.

A direct outcome of this special focus was a series of grants for workshops, student training courses, and small research projects emphasizing US and African faculty and students, and stressing topics in computational and mathematical epidemiology. The latest grant helped to establish the *US-African Biomathematics Initiative.* This DIMACS program, joint with the Mathematical Biological Sciences Institute (MBI) at Ohio State University, consisted of three parts during the current year: A Modeling Data Clinic on Meaningful Modeling of Biological Data and an Advanced Study Institute and Workshop on Economic Epidemiology. Later activities under this new Initiative will focus on conservation biology and on the topic of genetics and disease control.

Publications

Books

S. Ji, A. Chaovalitwongse, N. Feferman, W. Yoo, and J. E. Perez-Ortin, "Mechanism-based clustering of genome-wide transcript levels: roles of transcription and transcript-degradation rates," *Clustering Problems in Biological Networks*, ed. A. Chaovalitwongse (to appear).

Papers

J. Abello, B. Gaudin, H. J. Schulz, C, Tominski, "Name that cluster: Text vs graphics," IEEE Information Visualization Conference, 2007.

Banerjee, C., L. J. S. Allen, and J. Salazar-Bravo, Models for an arenavirus infection in a rodent population: Consequences of horizontal, vertical and sexual transmission, *Mathematical Biosciences and Engineering*, **5** (2008), 617-645.

S. Bellan, "Reevaluating vector control in the context of age-dependent mosquito mortality: Consequences for the control of chikungunya," PLOS One, August 2009, submitted.

K. Blayneh, J. Mohamed-Awel and Abdul-Aziz Yakubu, Discrete hierarchical competition with reward and cost dispersion, *Journal of Difference Equations and Appl*., **15**, No. 4, 399-414 (2009).

B. Dembele, A. Friedman, and A. -A. Yakubu, "Malaria model with periodic mosquito birth rate," *Mathematical Biosciences*, submitted.

B. Dembele, A. Friedman, and A. -A. Yakubu, "Malaria model on the impact of drug administration protocols," in preparation.

B. Dembele, A. Friedman, and A. -A. Yakubu, Malaria model with periodic mosquito birth and death rates, *Journal of Biological Dynamics*, **3**, Nos. 4, 430-445 (2009).

B. Dembele, A. Friedman, and Abdul-Aziz Yakubu, Mathematical model for optimal use of sulfadoxine pyrimethane as a temporary malaria vaccine, *Bulletin of Mathematical Biology* (revised, March 2009).

P.A. Dreyer, Jr. and F.S. Roberts, Irreversible k-threshold processes: Graph-theoretical threshold models of the spread of disease and of opinion, Discrete Applied Math., **157** (2009), 1615-1627.

N. H. Fefferman and K. L. Ng, "Species-specific behavior affects disease spread throughout and ecosystem," submitted.

Glasser, J.,McCauley, M.M., and Hatchett , R., "Infectious disease physicians advise modelers about aiding in health crises," in preparation.

Ndifon, W., Dushoff, J., and Levin, S., "On the use of hemagglutination-inhibition for influenza surveillance: Surveillance data are predictive of influenza vaccine effectiveness," *Vaccine* 27 (2009) 2447–2452

Ndifon, W., Wingreen, N.S., and Levin, S., "Differential neutralization efficiency of hemagglutinin epitopes, antibody interference, and the design of influenza vaccines," PNAS, 106 (2009), 8701-8706.

F.S. Roberts, “Greedy algorithms in economic epidemiology,” in preparation.

F.S. Roberts, “Meaningful and meaningless statements in epidemiology and public health,” submitted.

Abdul-Aziz Yakubu and Michael Fogarty, Periodic Versus Constant Harvesting of Discretely Reproducing Fish Populations, *Journal of Biological Dynamics*, **3**, Nos. 2-3, 342-356 (2009).

Reports

Talks

Kibambe, J. "Infectious diseases and productivity trade-offs", Graduate School of Business, University of Chicago, March 2009.

Fefferman, N., "The Impact of Household Capital Models on Targeted Epidemiological Control Strategies for Diseases with Age-Based Etiologies", Workshop on Economic Epidemiology, Makerere University, Uganda, August 2009.

Roberts, F.S., “Sometimes it Pays to be Greedy: Greedy Algorithms in Economic Epidemiology,” Workshop on Economic Epidemiology, Makerere University, Uganda, August 2009

Roberts, F.S., “Epidemiology: An Operations Research Approach,” Advanced Study Institute on Economic Epidemiology, Makerere University, Uganda, July 2009

Roberts, F.S., “Meaningless Statements in Epidemiology,” Advanced Study Institute on Economic Epidemiology, Makerere University, Uganda, July 2009

Main Web Site

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

Other Specific Products

Web Pages

http://dimacs.rutgers.edu/Workshops/WGSpatioTemporal3/
Web page for DIMACS Working Group on Spatio-temporal and Network Modeling of Diseases III

http://dimacs.rutgers.edu/Workshops/Contagion/
Web page for DIMACS/DyDAn Working Group on Network Models of Biological and Social Contagion

http://dimacs.rutgers.edu/Workshops/Stochastic/
Web page for DIMACS Workshop on Stochasticity in Population and Disease Dynamics

http://dimacs.rutgers.edu/Workshops/Botanical/
Web page for DIMACS Workshop on Models/Methodological Problems of Botanical Epidemiology

http://dimacs.rutgers.edu/Workshops/HIPAA2009/
Web page for DIMACS Tutorial: Statistical De-identification of Confidential Health Data with Application to the HIPAA Privacy Regulations

http://dimacs.rutgers.edu/Workshops/Phenotypes/
Web page for DIMACS Workshop on Identifying Genetic Signatures for the Evolution of Complex Phenotypes

Contributions

Contributions within Discipline

This Special Focus is by nature multi-disciplinary. It involves applying methods of computer science, statistics, and mathematics to problems in epidemiology. 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 this grant.

One direct result of this grant is a new DIMACS initiative on Climate and Health. This initiative, funded by an "Academic Excellence Award" from Rutgers University, is concerned with the impact of "heat events" and attempts to answer the following questions: What people are most vulnerable to extreme heat events in cities? What evacuation strategies do we use? Where do we take them? There will also be work on the health effects of rolling blackouts during heat events, and how to schedule them to minimize health effects. Another area of emphasis will involve the interaction among pesticide use and climate. Note that the workshop on Botanical Epidemiology in the Activities section included considerable discussion of the effect of climate change on plants/food supply. There will be a study of optimal control of and managing the vector/pathogen populations most effectively given the chemical controls available, while minimizing the health impact to human populations. We will explore how fluctuating environmental conditions can affect the resulting balance of these two concerns and how to tailor models to incorporate the impact of climate conditions on both chemical- and disease-based risks. Finally, in the first stages of this new initiative, there will be work on optimal routing of ambulances and other emergency vehicles during extreme wind or flooding events resulting from climate change. It is likely that this activity will also become a major initiative of the newly-formed University Center for Disaster Prevention and Emergency Response, a partnership of Rutgers University, the University of Medicine and Dentistry of NJ, and the Robert Wood Johnson University Hospital. This partnership, with DIMACS playing a lead role, would never have happened were it not for this NSF grant for computational epidemiology.

DIMACS member Nina Fefferman said "Throughout the past year, I have continued my research into the varying success of organizational strategies of worker allocation under conditions of disease related workforce depletion. The results of these ongoing endeavors have been presented at two DIMACS workshops. Additionally, work to extend their applicability, incorporating economic modeling, is now underway in collaboration with researcher Dr. Ramanan Laxminarayan (of Resources for the Future), whom I met at a DIMACS workshop in 2005. Further possibilities for extensions of this work to examine alternate scenarios of economic impact are just beginning in collaboration with Dr. Martin Meltzer (of the CDC), whom I met at the DIMACS Capstone Workshop in South Africa.

"Due to our mutual involvement with DIMACS, Prof. Dina Fonseca (at the Rutgers Univ. Center for Vector Biology) and I met to discuss our mutual interest in temporal constraints on mosquito reproduction. As an outgrowth of this meeting, we have begun a collaboration to investigate the impact of mosquito habitat selection on invasive spread of a newly introduced mosquito in the state of NJ. Once accomplished, this work will have profound impact on the capability for predictive modeling of mosquito-borne diseases as this species is a known disease vector in its native habitats. This work involved the recruitment of three other researchers, Prof. Rick Lathrop, Prof. Jim Miller, and Prof. Randy Gaugler (of the Rutgers Univ. Depts. of Ecology and Evolution, Oceanography and Marine Science, and the Center for Vector Biology, respectively). Together, the five of us have now submitted a proposal for

funding of this work to the NSF Directorate of Biological Sciences.

“All of these projects and collaborations were made possible by my involvement with DIMACS over the past few years. I feel that my perspective, my own individual research has been broadened substantially by these opportunities 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.”

Sungchul Ji describes his collaboration with Nina Fefferman as follows: “Dr. Fefferman attended one of my seminars given at the DIMACS Center at Rutgers on April 17, 2006 invited by James Abello. Based on the content of my seminar entitled ‘The Simpson-Elsasser-Wolfram (SEW) Framework for Modeling the Living Cell,’ Nina asked me to contribute a paper to a special edition of the journal Ann. Zool. Fennici, for which she is an editor. This led to writing a set of three manuscripts applying the network theory to cell biology. It was through the influence of James Abello, Nina Fefferman and Art Chaovalitwongse whom I met through DIMACS Workshops and Seminars that I was motivated to learn and apply network concepts and theories to molecular and cell biology. In the process, I have been to led to formulate what appears to be a novel theory of biological networks and apply it to modeling the living cell.”

Jonathan Read from the Mathematics Institute at the University of Warwick collaborated with Nina Fefferman (DIMACS) on a variety of projects while visiting DIMACS. The primary focus of their work was expanding the research that Fefferman and former DIMACS postdoc Kah Loon Ng began to examine under what general rules self-organizing dynamic social networks formed similar convergent structures using centrality measures as a metric for similarity. This work continued at the DIMACS funded workshop in Edinburgh. Second, Read collaborated with Fefferman to design a novel presentational encoding strategy for social network graphs which will allow a rapid characterization of the potential for the spread of disease and/or information within the network depending on the individual through which occurs the primary introduction. Third, Read and Fefferman collaborated on models examining the difference in selective pressures acting on pathogen evolution caused by physiological constraints from host-immune evasion. Lastly, Read presented his own research into the social network structure of academia and during his visit collaborated with James Abello (DIMACS) to create a clearer visual representation of his results. Together they were able to construct an image which clearly communicated greater levels of complexity in the network than had been previously possible.

Alina Campan (Northern Kentucky University) and Triaian Marius Truta (North Kentucky University) assisted Daniel Barth-Jones (Columbia) with the tutorial on Statistical De-Identification of Confidential Health Data with Application to HIPAA Privacy Regulations. Here is what Dr. Campan had to say: ”Primarily, the DIMACS Tutorial on Statistical De-identification of Confidential Health Data with Application to the HIPAA Privacy Regulations was a great opportunity to have a longer and insightful meeting with our collaborator, Dr. Daniel Barth Jones. He has a strong industrial experience and his comments, coming from a well seasoned practitioner in the data privacy and statistical de-identification field, helped us crystallize a few ideas about the validity and utility in real-life situations of a new privacy model (constrained k-anonymity) we had introduced very recently, in 2008. We also had the same type of feedback from other practitioners participating to the tutorial. So, in a sense, the tutorial catalyzed the interaction between researchers and practitioners in the data privacy field. The feedback we obtained encouraged us to continue developing the mentioned privacy model. As a result, we will be soon publishing a new paper that reports on our advances; one of the co-authors of this paper is an undergraduate student at NKU. Dr. Truta has also had a few preliminary discussions about a possible collaboration with DaVita Clinical Research, which was represented at the tutorial by Dr. Steven Marshall Wilson.”

Contributions to Other Disciplines

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

Contributions To Human Resources Development

Several graduate students, undergraduates, and postdoctoral researchers have participated in the program.

For example, Jacques Kibambe Ngoie has worked as a post-doctoral student with Nina Fefferman over the past year and they have coauthored a paper on infectious disease and productivity tradeoffs.

Through the Research for Undergraduates (REU) program and its linkage to Rutgers RISE for minority students, Immanuel Williams, worked this past summer jointly with Tami Carpenter and Nina Fefferman, both of DIMACS, applying the ideas from the Fefferman and Ngoie paper to a multigenerational model of local community economics and HIV/AIDS infection. A paper is in preparation.

Two postdoctoral visitors were heavily involved in the special focus in this past year. The work of these postdocs, Debbie Yuster and Paola Vera Licona, is described in the section on Training and Development.

Our supplemental project on Integrating Bayesian Regression into Blackbook, Learning with Emergent Classes and Higher-Order Links has made a substantial contribution to graduate education, providing support during his final year of study for Paul Raff, who successfully defended his doctoral dissertation on automated proof of combinatorial mathematical theorems, on August 14, 2009. In addition, it has opened a substantial new career pathway for Dr. Raff, who applied successfully for a postdoctoral fellowship sponsored by the US government intelligence community.

Contributions to Research and Education

These are summarized in the sections on Human Resource Development, Outreach, and Training and Development earlier in this document.

Contributions Beyond Science And Engineering

The Special Focus on Computational and Mathematical Epidemiology led to direct connections between DIMACS and many agencies concerned with homeland security issues, including the Department of Homeland Security, FEMA, the NJ Office of Homeland Security and Preparedness, the NJ Dept. of Health and Human Services, etc. The PI, Fred Roberts, has been named to NJ Governor Corzine's Health Emergency Preparedness Advisory Committee, HEPAC, which recently has emphasized preparation for the H1N1 virus.

In turn, these connections with homeland security agencies and issues, arguably led to the development of a proposal and subsequent award of a Department of Homeland Security University Center of Excellence for Dynamic Data Analysis (DyDAn) in 2006 and for a DHS University Center of Excellence for Command, Control, and Interoperability for Advanced Data Analysis (CCICADA). Both are based at DIMACS.

One important outcome of this special focus is a new partnership between DIMACS and the Cancer Institute of New Jersey (CINJ). CINJ is a Center of Excellence of UMDNJ-Robert Wood Johnson Medical School and the first and only National Cancer Institute-designated Comprehensive Cancer Center

in New Jersey. CINJ has become a full partner of DIMACS and the two institutions have already run joint activities and submitted joint proposals. Gunaretnam (Guna) Rajagopal, PhD, the executive director of the CINJ Bioinformatics Program, expressed confidence that the partnership will result in new collaborations, “By coming together to identify mutual areas of interest and form program partnerships, we are hopeful that as a collective we will be able to attract funding for this critical work. By taking advantage of the tremendous advances in technology, we have an opportunity to deepen our understanding of cancer and other diseases. We are pleased to partner with DIMACS in leading these efforts.”