Modeling Social Dynamics: Sharing Perspectives across Disciplines October 5 – 6, 2006 Workshop Report[1,2]

[1] The workshop described in this report was supported by the National Science Foundation under Grant No. BCS-0647023. Any opinions, findings or conclusions are those of the authors and do not necessarily reflect the views of the National Science Foundation.
[2] The workshop organizers, Tanzeem Choudhury and James Kitts, would like to give special thanks to Raima Larter, Michelle Chow, Bill Kaghan, Eric Gleave, and Laura Zager, who have helped us extensively in producing this document.

Table of Contents

AGENDA

Thursday, October 5, 2006

07:30am-08:00am	Pick up security badges (at Stafford I)
08:00am-08:30am	Refreshments
08:30am-08:35am	Welcoming Remarks: *Raima Larter*
08:35am-08:45am	Welcoming Remarks: *James Kitts*

SESSION 1: Theoretical Modeling of Social Dynamics

08:45am-09:15am	Plenary Speaker:	*Yaneer Bar-Yam*
09:15am-10:00am	Breakout Sessions	
10:00am-10:15am	Coffee Break	
10:15am-10:45am	Plenary Speaker:	*Peyton Young*
10:45am-11:00am	Participant Introductions	
11:00am-12:00pm	Plenary Discussion	

12:00pm-1:15pm Lunch

SESSION 2: Validating Models of Social Dynamics

1:15pm-1:30pm	Presider:	*Robert Axtell*
1:30pm-2:00pm	Plenary Speaker:	*Bernardo Huberman*
2:00pm-2:30pm	Plenary Speaker:	*Peter Turchin*
2:30pm-3:15pm	Breakout Sessions	
3:15pm-3:45pm	Plenary Discussion	
3:45pm-4:00pm	Break	

WRAP-UP: Recommendations for Promoting and Enhancing Modeling of Social Dynamics

4:00pm-4:10pm	Individual Brainstorming & Feedback
4:10pm-4:30pm	Small Group Brainstorming
4:30pm-5:00pm	Plenary Discussion

6:00pm-7:30pm NSF Hosted "Networking Dinner" at Hotel

Friday, October 6, 2006

08:00am-08:20am	Refreshments
08:20am-08:30am	Welcoming Remarks: *Tanzeem Choudhury*

SESSION 3: Applications: Modeling the Dynamics of Networks

Plenary Panel	Moderator:	*Péter Érdi*
08:30am-08:50am	Speaker:	*Noshir Contractor*
08:50am-09:10am	Speaker:	*Gyorgy Korniss*
09:10am-09:30am	Speaker:	*John Padgett*
09:30am-09:45am	Panel Discussion	
09:45am-10:00am	Coffee Break	
10:00am-10:45am	Breakout Sessions	
10:45am-11:00am	Report on breakout sessions	

SESSION 4: Applications: Modeling Disasters

Plenary Panel	Moderator:	*Carter Butts*
11:00am -11:20am	Speaker:	*Kathleen Carley*
11:20am -11:40am	Speaker:	*Louise Comfort*
11:40am -12:00pm	Speaker:	*Madhav Marathe, Chris Barrett*
12:00pm -12:15pm	Panel Discussion	

12:15pm-1:15pm Lunch

1:15pm-2:00pm Breakout Sessions
2:00pm-2:15pm Plenary Discussion

WRAP-UP: Workshop Synthesis and Conclusion

2:15pm-2:30pm Small Group Discussion
2:30pm-2:45pm Plenary Discussion
2:45pm-3:00pm Workshop Evaluations

3:00pm Workshop Adjourned

INTRODUCTION

The social world that we observe reflects a web of interdependent processes, with macro-level structures of organizations, communities, and societies both emerging from and constraining the micro-level interactions of individuals. Many social phenomena, such as the spread of epidemics, the rise of political insurgency, or the dissolution of organizations, are inherently time varying and depend on interactions between entities within a social system.

Understanding the link between micro-level interactions and macro-level dynamics promises to have profound impact on how human societies, organizations, and nations might be structured and how related policy decisions should be made. Toward this end, an increasing number of scientists are using mathematical and computational models to elucidate theoretical problems in social dynamics, often by applying general theories or methods that are well developed in the natural and physical sciences with a view to gaining insight into the underlying generative processes or the dynamic consequences of social relationships. For example, models from statistical physics have been used to study healthcare organizations, population ecology models have been applied to the study of the evolution of industries, and neural networks have been used to model the origins of religious beliefs. Although these new applications have shown promise, their impact has been limited because well-enforced disciplinary boundaries and conventional scholarly practices constrain the diffusion of ideas and methods across disciplines in the natural and social sciences and engineering.

This meeting was organized to foster intellectual exchange among mathematicians, natural and physical scientists (e.g. biologists, physicists, computer scientists, and engineers), and social and behavioral scientists with a common interest in the development of mathematical and computational methods that can be applied to problems in modeling social dynamics. The meeting served three purposes: (i) to facilitate interdisciplinary conversations among diverse experts and give rise to future collaborative projects of high impact in areas of social dynamics, such as the evolution of social networks, evolutionary and ecological models of social behavior, population dynamics and demography, and organizational dynamics, (ii) to encourage and enable applications of these models to real-world problems, for example, epidemics, distributed failures due to natural disasters and general consequences of environmental change, and finally (iii) to help NSF identify and nurture emerging research agendas as part of the Human and Social Dynamics priority area by providing a forum for researchers to compare and coordinate lines of research,

In line with these goals, the workshop broadly considered three themes:

1. The application of methods and tools from a broad range of disciplines to social dynamics, and the specific challenges in the modeling of social dynamics.
2. The exploration of ways in which interdisciplinary researchers can come together and collaborate fruitfully to extend research in specific disciplines, integrate various lines of research, and advance the state-of-the-art methodologies in modeling human and social dynamics.
3. The application of modeling techniques to real world contexts, and how such applications demonstrate the relevance of modeling human and social dynamics to decision-makers and others outside of the research community

WORKSHOP ORGANIZATION

The workshop consisted of four sessions dealing with different methods for modeling social dynamics and applications of these techniques. These four sessions were:

1. Theoretical modeling approaches for social dynamics
2. Validation of models in social dynamics
3. Applications to theoretical problems, with a focus on network dynamics
4. Applications to real-world problems, with a focus on modeling of disasters

In each session, two or three plenary speakers presented recent work and results related to the topic. In Sessions 1 and 2 the plenary talks were followed by breakout discussion groups. In Sessions 3 and 4, the plenary talks were followed by a panel discussion in which the plenary speakers considered and nominated areas of future research relevant to the session topic. Following each panel discussion, workshop participants responded to the questions through written responses and verbal discussions. In addition, participants brought written answers to "homework" questions relevant to each session (distributed prior to the workshop). These questions served to inform and prepare participants about the session topics, and provided another forum in which participants could voice their ideas and thoughts. At the end of each day, workshop participants discussed ways in which the NSF could help to support and advance the field of modeling social dynamics. This workshop report is a distillation of all material presented and discussed at the workshop, based on recorded discussions and written reports.

SESSION 1: THEORETICAL MODELS FOR SOCIAL DYNAMICS

The first session of the workshop entailed an interdisciplinary discussion on the promises and challenges of modeling social dynamics. There are many examples of applying models from the natural and physical sciences to the social sciences. For example, models from computational neuroscience have been used to investigate the evolution of military alliances, population ecology models have been applied to speciation in manufacturing industries, and methods of sequence analysis from bioinformatics have been applied to the study of residential mobility.[3] In this session two plenary speakers, Yaneer Bar-Yam and Peyton Young, described the application of models from the physical sciences to social science problems. Workshop participants then formed two breakout groups which discussed several theoretical frameworks that could potentially provide insight into research problems in multiple disciplines.

PLENARY TALKS

The first plenary speaker, Yaneer Bar-Yam, described strategies used in modeling complex systems. Such systems – whether physical, biological, or social – often behave in ways that are not immediately inferable from the behaviors of their component parts. Creating useful models of complex systems requires identifying interdependencies between component parts, describing mechanisms of pattern formation, and developing a multi-scale or hierarchical approach.[4] That is, models must simultaneously consider behaviors at a broad range of spatial and temporal resolutions. Bar-Yam's presentation described the use of models developed in ecology and evolutionary biology to account for the spatio-temporal variations in the evolution of social altruism. [5] The presentation and subsequent discussion highlighted the importance of carefully examining the assumptions made during the development of a model when transferring tools and concepts across fields and disciplines. For example, if the structure of contacts influences the population growth at a broader macro level then assuming the global dynamics to be independent of the local interaction structure (e.g. using the standard mean field approximations) can give misleading results. [6],

The second plenary speaker, Peyton Young, described the adaptation of methods from statistical mechanics and stochastic processes to study the dynamics of complex social and economic systems. Such systems often involve large numbers of interacting agents with different information, preferences and rules of behavior. Social scientists have long regarded such dynamic systems as analytically intractable due to their high dimensionality, and have either employed extreme simplifying assumptions (e.g. assuming homogeneous and independent agents) to make models tractable or have used more flexible but less general methods, such as agent-based modeling. Young described an approach based on the

[3] Axelrod, Robert C. and D. Scott Bennet. 1997. *The Complexity of Cooperation*.
Hannan, Michael T. and Glenn Carroll. 2000. *The Demography of Corporations and Industries*.
Stovel Katherine and Marc Bolan. 2004. "Residential Trajectories: The Use of Sequence Analysis in the Study of Residential Mobility." *Sociological Methods and Research* 32: 559-598.

[4] Bar-Yam, Yaneer. 2002. General Features of Complex Systems, *UNESCO Encyclopedia of Life Support Systems*.

[5] Rauch, Erik M. and Yaneer Bar-Yam. 2006. "Long-range interactions and evolutionary stability in a predator-prey system, *Physical Review E 73,* 020903.

[6] Bar-Yam, Yaneer, 1999. "Formalizing the gene centered view of evolution," *Advances in Complex Systems* **2**, pp.277-281.

theory of large deviations in Markov chains that allows the long-run behavior of stochastic systems to be modeled analytically. Young used two examples, Thomas Schelling's classic model of residential segregation[7] and his own research on the distribution of contractual norms in Illinois, to demonstrate the application of this framework[8]. He pointed out that similar analyses can be applied to a very general class of dynamic models that are applicable to social and economic systems but were previously considered intractable. Ongoing challenges in this area include the generalization of the framework to networks with asymmetric influences, the estimation of time it takes to reach equilibrium or steady state, and finding more application scenarios where the predictions of a model can be compared with real-world time-series data.

In the discussion following this presentation, participants inquired about the implications of adapting the Schelling model to consider discrete, bounded neighborhoods rather than egocentric neighborhoods. They also discussed the promises and pitfalls of migrating concepts across disciplines, such as developing an analog to the concept of "temperature" as deployed in the physical and chemical sciences to represent heterogeneity among social actors.

Links to plenary talks slides:
http://seattle.intel-research.net/MSD/talks/Bar-Yam.pdf
http://seattle.intel-research.net/MSD/talks/Young.ppt

BREAKOUT SESSIONS AND GENERAL DISCUSSION

In the breakout sessions and written responses, workshop participants considered both the utility and challenges of adapting mathematical and computational methods and associated theoretical frameworks to address research problems in a wide variety of disciplines including physics and natural sciences, engineering and computer science, mathematics and statistics, and the social sciences. They focused particular attention on the difficulties encountered when applying techniques developed for modeling the dynamics of non-social phenomena to model social systems and their dynamics.

Computational approaches used to develop models for social dynamics

Participants identified a diverse set of computational techniques that are useful for developing and adapting dynamic models across different disciplines.

Differential Equations

Systems of both ordinary and partial differential equations (DEs) are useful tools for building theories in a vast array of disciplines, including social science disciplines such as economics, political science and anthropology. In the social sciences, DEs are very useful for developing models at the aggregate level, but are more limited when considering the interdependent behavior of agents in a population. One area of application of DEs in the social sciences is in non-equilibrium dynamics. Examples of this type of application include models of the rise and fall of states, territorial expansion and the growth of world

---

[7] Thomas Schelling. 1978. *Micromotives and Macrobehavior*, W. W. Norton and Company.

[8] Young, H. Peyton and Mary A. Burke. 2001. "Competition and Custom in Economic Contracts: A Case Study of Illinois Agriculture," *American Economic Review*, 91: 559-573.

regions. Differential equations may also be applicable to modeling of communication channels and the formation of connections, for example, internet communities and geo-coded databases.

Agent-based modeling (ABM)

Agent-based methods are particularly useful for studying the micro-level behavior of individual objects in a system. In an agent-based model, there are many decision-making actors, or agents, dispersed on a social or geographic landscape. Generally it is assumed that there is some degree of heterogeneity among agents, with different preferences, information and degrees of rationality. The interactions between agents may be stochastic according to defined interaction and choice probabilities, and may be subject to exogenous shocks. Agent-based computing is commonly used in a wide spectrum of fields, including biology, economics and sociology, to address a great variety of questions, such as residential segregation and population growth and distribution, as illustrated in the plenary talks. Given the limited applicability of systems dynamics models using differential equations to populations of heterogeneous and interdependent agents, agent-based models may be used in conjunction with such system-level models or with ABMs at different resolutions to serve as a basis for developing multilevel models.

Probabilistic Graphical Models and Random Graph Theory

Probabilistic graphical models combine probability theory and graph theory to efficiently encode dependencies among individuals who collectively form a larger complex system while ensuring that the system is consistent as a whole. A graphical model describes a probability distribution on a set of variables associated to the vertices of a graph, where the edges in the graph encode conditional probabilities. A very common example of a probabilistic graphical model is a Markov Chain, which has been used to model phenomena ranging from social, economic and population genetics processes to traffic flows. While Markov chains have a wide spectrum of applicability, the independence assumptions that they make limit their modeling power. Often more complex dynamics are captured using interacting Markov chains or more complex Bayesian networks or Markov random fields.

Random Graph Theory also lies at the intersection between graph theory and probability theory, and studies the properties of typical random graphs. A random graph is generated by some stochastic process, i.e. a graph is sampled from a probability distribution on a set of graphs. It is possible to have graphical models in which the graph structure itself is random. Random graph models can be used to consider of the evolution of both scale-free and geometric graphs and to identify discriminating metrics for the graph characteristics. Specific applications domains for random graph theory include the propagation, reproduction and evolution of web pages and the replication of genes and operons in biology. An ongoing challenge in the development and application of random graph theory in the social sciences is the identification and computation of important and discriminating metrics of these networks.

General theoretical frameworks that can be applied to a variety of social dynamics problems across multiple areas of study

General theoretical frameworks provide the conceptual underpinnings for a variety of modeling techniques. In this regard, it is important to identify the areas where it is appropriate to apply these different frameworks and the challenges and constraints involved in their application. Some examples of theoretical frameworks discussed in the workshop include:

Game Theory

In its classic form, game theory is a theoretical framework for representing strategic interdependence among rational agents. Originally developed to investigate the card game poker and the associated problem of bluffing, game theory has found applications in a fantastic diversity of fields. Game theorists have studied cartels among manufacturing firms, planned the flight paths of bombers, and modeled temporal patterns in the volume of internet traffic.[9] Work in behavioral game theory has explored and relaxed the strenuous assumptions of rationality in classical game theory, such as common knowledge and maximization of expected utility, using systematic study of strategic choices in laboratory research.[10] Research in non-human animals by evolutionary game theorists has further relaxed traditional assumptions of game theory and has expanded domains of applicability. For example, recent work has examined closed orbit dynamics of mating strategies and coloration of side-blotched lizards.[11] Such general models of strategic interdependence have informed foreign policy choices and suggest implications for the study of inter-group rivalry and conflict, such as that found between Shiites and Sunnis in Iraq, and countless other examples of ethnic conflicts that explode into violence.

One unresolved problem for the application of game theory is the relationship between macro-level (system) and micro-level (agent) equilibrium. The questions of how an economy moves from out-of-equilibrium conditions to equilibrium conditions and whether it is possible to attain macro-level equilibrium without equilibrium being reached at the agent level are particularly important in this area.

Recent research has explored intersections of game theory and social network theory, such as investigating the implications of the structure of social interaction for cooperation in social dilemmas or investigating the implications of strategic interdependence for the evolution of interaction networks. Given problems of analytical tractability for all but the simplest problems in this area, agent-based frameworks are widely used to model dynamics of such systems.

Interacting Particle Systems

The theory of interacting particle systems, well-studied in physics, involves mapping the behavior of individual stochastic actors to the formation and maintenance of larger system domains patterns. Bar-Yam, the first plenary speaker, demonstrated how this framework can be used in his discussion of models for population evolution and growth. Applied studies of opinion dynamics in social structures that provide insights into what are the relevant factors in forming opinions is another example of applications of interacting particle theory. However, there are some challenges in applying this framework in the social sciences. For example, in with interacting particle system models in physics, over time, the system reaches a point in which one state can maintain dominance over another state. In contrast, in modeling democratic political systems in which periodic elections determine the state of the system, it is often very difficult for one party to maintain dominance over another party for an extended period of time. This suggests that there are strong ideological and political lines along which people vote, with a very small proportion of people willing to change their opinion. One way in which to address this issue is to adapt the model by including thresholds where people with opposing opinions are less likely

---

[9] Huberman, Bernardo. 2001. *The Laws of the Web: Patterns in the Ecology of Information.* Cambridge: MIT Press.

[10] Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton University Press.

[11] Kelly R. Zamudio and Barry Sinervo. 2000. "Polygyny, mate-guarding, and posthumous fertilization as alternative male mating strategies." *Proceedings of the National Academy of Science* 97: 14427-14432

to interact with each other. As in the ecological population growth models presented by Bar-Yam, the pre-existing boundaries and environment might impact the final distribution.

Neutrality theory

Population geneticists and population ecologists use models based on neutrality theory to make predictions about variation in species-area relationships as a function of scale, relative species abundance and phylogeny. These models provide insights into speciation, biogeography and biodiversity. Based on the successful adaptation of biological models of population dynamics and epidemics to the social sciences, it is reasonable to posit that neutrality theory may also find applications in the social sciences. Again, it is important to recognize that assumptions underlying the theory in the biological context may not be applicable or appropriate to human societies. Nonetheless, the neutral theory for spatial distributions provides a useful starting point for further exploration.

Biological control networks

Another class of biological models that may be useful in the social sciences deals with the evolution of metabolic and biological control networks. These models are used to study problems such as tracking infections or contaminations back to the source in contexts such as food, drug or water supply chains. These models assume that such contaminations are spread from a single source. Workshop participants suggested that Bayesian methods for diagnosis, which deals with several kinds of unreliability, might be a suitable technique for the application of this framework.

Properties of Models

As discussed in previous sections, importing models from one discipline requires careful consideration of whether or not the assumptions underlying the applications in one disciplinary context are adequate for applications in a different disciplinary context. Making adaptations to underlying assumptions and model properties is often necessary for the successful application of a class of models to a different area of substantive problems.

Assumptions about Independence

An important consideration in the definition of social dynamics models is the assumption of independence between actors. As Bar-Yam discussed during the first plenary talk, standard mean field approximation assumes an average context for time-dependent processes, resulting in a distribution that is independent of boundaries. However, in modeling population dynamics, this assumption does not hold, as the local environmental factors affect the dynamics and final distribution significantly. Therefore, the concept of mean field approximation, while useful in physics, often does not work well in the biologically relevant context of population dynamics. That being said, Bar-Yam pointed out that mean field approximations may still have utility as a preliminary step in modeling complex systems to indicate further directions in model development. More broadly, workshop participants recognized that the usefulness of using statistical methods that make strong independence assumptions and ignore social relations (i.e. dependence) among individuals remains a controversial topic among social scientists.

Underlying Distribution and Continuity of Variables

As with assumptions of independence, assumptions regarding the underlying distribution of variables in model can significantly affect model outcomes. Once again, using the example of population distribution described by Bar-Yam in the first plenary talk, a mean field approximation approach usually assumes the distribution of variables to be Gaussian. However, in some biological context, the spatial distribution of the population is often more accurately described by a power law function rather than a Gaussian or exponential function. Hence, it is extremely important to carefully define and justify assumptions about distributions. Mapping the model to an inappropriate distribution will produce inaccurate predictions. Similarly, workshop participants also emphasized that data or variables being modeled may be continuous or discrete, and assumptions about the continuity of the distributions may impact outcomes of a model.

Interpretation of model parameters

When transferring modeling approaches from one discipline to another, researchers need to remain sensitive to the fact that the interpretation of the meaning of model parameters may shift in the transfer process. For example, there was a discussion about the analogue of "temperature" in social networks following the second plenary talk by Young. In the physical sciences "temperature" is defined as a quantity proportional to the average velocity of the particles in a given system. In models of social interaction, on the other hand, the analogous concept represents random idiosyncrasies of the individual agents. In chemistry, a heterogeneous system of molecules with weak intermolecular interactions and high temperatures will move quickly and mix amongst each other. So too, in a social system, where individuals exhibit a high degree of variability and little interconnectivity, actors, as predicted by Schelling's model, will readily mix. Similarly, the Naming Game (which involves the selection of terms and is discussed in more detail in a later section), is also based on models of interacting particle dynamics in physics. However, in contrast to physics models, the Naming Game assumes a critical threshold value for reaching global consensus. This threshold represents the difference in opinions between neighbors. If this value is too small, there is too much dissimilarity between agents' opinions and global agreement in the system will never be reached.

Time scales and equilibrium

A significant issue in any domain of social dynamics is estimating the waiting time, or the time taken to reach equilibrium. Often there is a disparity between event and real time scales. It may take thousands of generations for some systems to reach equilibrium or for particular properties to emerge. In some situations, such as improving education systems and student performance, solution/equilibrium states need to be reached relatively quickly. If a given solution emerges only over a thousand attempts, it becomes difficult to predict if the solution will be reached in any particular instance. Thus, it is important to consider the appropriate number of iterations required to reach equilibrium or to obtain a reasonable representation of the problem.

Furthermore, while systems may indeed reach equilibrium, they may not necessarily do so in realistic time scales. This was highlighted in the Naming Game model presented in Session 3, where, according to the model, the system converges over a long period of time, but in reality the convergence is never reached. Thus, in developing a model, it is important to consider how process time maps to real time, and how this mapping relates to the problem at hand. In many cases, reaching equilibrium may not be necessary, as the general dynamics may be of more interest than the eventual asymptotic behavior of the model.

Another consideration for time scales in modeling dynamics is the fact that two phenomena may display similar overall dynamics, but may occur over very different time scales. For example, epileptic seizures, earthquakes and stock market crashes are all very sudden catastrophic events which could potentially have similar precursor behaviors that can be represented in models. These models might show similarities and inform each other in their dynamic behavior, however, their time scales could be very different. For example, warning signals such as birds suddenly flying away may occur only minutes before an earthquake, while medically-related signals could occur months or years before an epileptic episode.

Challenges in Modeling Social Dynamics

The fundamental complexity of the dynamics of social phenomena and the complications associated with representing agents who engage in rational, responsive, and realistic interactions present challenges that are significantly more difficult than those faced by modelers of non-social dynamics.

Feedback

In developing models, it is important to account for feedback mechanisms. For example, when attempting to translate fragility modeling into organizational analysis, the need to accommodate the fact that people and organizations can and will change their minds in response to the behavior of the system presents important challenges. Similarly, when using Markov Chains to develop models of transportation systems and traffic flow, it is important to link the responses of social actors (drivers) with non-social physical components such traffic signs, and effectively integrate both of these different elements into the model. In addition, agents often learn from experience and alter their future actions in response to the past model outcomes. These behavioral changes (especially if broadly shared), in turn, may subsequently change the dynamics of the situation. Such feedback loops vastly complicate the construction of models as behavioral changes need to be accounted for and fed back into the analysis of both individual behavior and system performance.

Causality

Dealing with causality in a system is also a significant challenge in modeling social dynamics. Feedback loops, for example, where cause and effect are linked together and cannot be easily separated without affecting overall system dynamics, complicate model development. Modelers should also exercise caution when choosing parameters to model. For instance, in modeling events such as epileptic episodes or earthquakes, one can model observed data and indicators that occur prior to an event itself. However, if the model is based solely on signal correlations at the surface level, rather than the deeper processes that generate those signals, the model may not represent the true causative processes and model predictions are liable to be inaccurate. Nevertheless, it is equally important to balance these considerations against existing limitations in substantive knowledge of underlying generative processes and precursor events. The level of model development is limited by the existing knowledge base in a research community.

Tractability

Participants also raised the issue of model tractability several times during the workshop. Social dynamics models that attempt to account for a large number of interacting agents governed by different behavioral rules and preferences quickly become very complicated. These types of models have a very high dimensionality and have many parameters that need to be specified. Building models often require managing a difficult balance between model complexity and tractability. More complicated models present problems in parameterization and validation. A high degree of complexity can result in a non-transparent model in which it is difficult to determine how parameters should be adjusted to obtain a suitable fit to data. While simpler models like Markov chains are more transparent and exhibit greater analytical and computational tractability and convergence, their quasi-linear form limits the number of interactions that can be considered, thus constraining the modeling power of this approach. An example of working around problems of tractability was presented by Peyton Young, the second plenary speaker in this session, in which he applied the theory of large deviations to Schelling's residential segregation model and the distribution farming contracts in Illinois. Furthermore, for complex probabilistic graphical models there are a variety of well understood approximation methods for tractably learning the parameters and making inferences about the dynamics using likelihood based techniques such as maximum likelihood, pseudo likelihood, variational techniques etc.

Issues related to modeling social dynamics across disciplines

Though different disciplines all share some common scientific values, individual disciplines still tend to have subtly different definitions and goals that influence how they build models and how they communicate about model building. Variations occur both on methodological issues such as expectations about the criteria for evaluating models and theoretical issues such as definitions of concepts. Varying vocabularies in different fields can impact interdisciplinary collaborations as researchers need to adjust their understanding and preconceptions in order to communicate productively and cooperatively with each other. Indeed, workshop participants employed a variety of definitions for the term *model*, including:

1. a formal representation of a theory about a process
2. a formalism that matches a set of input variables with output variables,
3. a representation of a set of constraints, with each definition being a subset of the next.

In attempting cross-disciplinary collaborations, researchers need to be aware of different meanings and connotations between fields and be willing to adjust expectations and ideas about definitions of terms, criteria for evaluating models and uses for models.

Another key consideration when choosing methods is what a particular application requires of the model and whether methods can be extended to make them more powerful. For example, researchers in computer science have worked hard to surmount the limitations and extend the applicability of Markov Chains. Similarly, research conducted in the fields of statistics, computer science, and engineering has focused on distilling information and computing estimates from incomplete network data on nodes and links. While extending the capabilities of tools and methods can be very useful, it is also important to recognize that there are fundamental limits on the extent to which models may be adapted for different systems.

When defining specific theoretical frameworks and approaches, it is also important to remain open-minded and not be limited by the existing repertoire of methods and frameworks. In considering a particular problem, the challenge is to find an appropriate approach, by applying multiple methods to see the problem from many angles. The ultimate challenge is to develop tools for problems that have not been previously solved. In this context, research should be driven by questions and problems, rather than solutions or tools.

Participants noted that that in the fields of population genetics and epidemiology there have been progressive moves from using dynamical models with differential equations that were initially solvable analytically, subsequently required simulation and finally used Markov chain processes that could either be solved or simulated using techniques such as stochastic maximum likelihood methods. Participants discussed whether a similar transition might be made in other fields such as engineering and the social sciences, or whether there is already a synthesis of all of these computational tools being used in an integrated approach.

SUMMARY

The social sciences, in attempting to develop approaches to modeling social dynamics, can borrow and adapt techniques first developed outside the social sciences. This session identified a range of models and techniques from engineering, physics, materials science, chemistry and biology that were potentially useful to the social sciences. Applications for such models include residential and population distributions, language and opinion dynamics, organizational stress, and multi-scale modeling with reference to global epidemic spread.

Models may be grounded in a variety of analytical approaches, including random graph theory, agent-based modeling, data-driven and statistical tools, and differential equations. Furthermore, the development of models can take place in the context of broader theoretical frameworks which again may be applicable in different disciplines. For example, classical Game Theory has been applied to a variety of domains ranging from biology to internet traffic. Other theories such as interacting particle systems and neutrality theory in biology may also be applicable to social contexts.

Several factors should be considered carefully for optimizing the accuracy of models for social dynamics. Assumptions about the independence, underlying distribution and continuity of variables can all affect the results and applicability of a model. When transferring models from other disciplines the underlying assumptions for a model as applied in one field may not be appropriate in the new context. Similarly, the substantive interpretation of the parameters and the time scale of a model require careful consideration. Modeling complex systems of interacting agents who are intelligent and can adapt their behavior presents particular research challenges. Finally, accurately modeling feedback events, accounting for causality and achieving computational and analytical tractability of models are all issues that require much attention and careful thought.

Applying models from other disciplines to social dynamics also raises practical problems of collaboration and research design. In bringing together researchers and methods from different fields, vocabularies, definitions and expectations need to be aligned to ensure productive and effective collaborations. Furthermore, models and techniques may be suitable for different applications. Therefore, identifying what function a model is intended to accomplish should inform the question of which model or technique is the most appropriate choice for achieving the desired function, whether there are valid extensions to existing models that may be applied, or whether the problem at hand requires developing a new model or technique.

SESSION 2: VALIDATING MODELS OF SOCIAL DYNAMICS

In this session, two plenary speakers and the session moderator described and discussed a variety of considerations in validating theoretical models, including methodological goals, criteria for evaluating fit, and model interpretation and use of error. Breakout session discussions focused on specific approaches used to validate theoretical models.

PLENARY TALKS

The session opened with a comprehensive introduction to concepts and issues involved in model validation. The social science modeling community consists of researchers with a heterogeneous mix of disciplinary backgrounds and a wide variety of approaches to validation. Given this diversity, it was important to define and share key unifying concepts and definitions pertinent to the social science modeling community.

The first plenary speaker for this session, Bernardo Huberman, addressed the general issue of the multi-scale nature of social dynamics problems and the challenges involved in ensuring not only that the microscopic description is consistent with the macroscopic dynamics, but also conversely, that macroscopic dynamics resemble the microscopic world. Specifically, he explored the connections between the local actions and the global behavior of large distributed systems. Understanding this relationship is important to a number of domains such as distributed knowledge, economic systems, social organizations and their dynamics. The overall dynamics of these systems is determined by the collective interactions of many autonomous actors operating in domains that often bridge several disciplines. These models display a repertoire of complex and useful behaviors, and may offer new mechanisms for accessing the collective intelligence of social groups. Using several examples, the presentation demonstrated how the ability to use very large data sets from the internet provides observational data on a massive scale, enabling the mapping of interactions and exchanges, which in turn can provide insights into classic social science areas such as social dilemmas (through analysis of internet congestion) and information propagation (through analysis of viral marketing) and the ability to evaluate the theories proposed in these domains.

Questions and discussion following this presentation focused on the future availability of such large data sets to the research community in general. Other participants asked about the potential of mapping interactions through email messages to provide insights into the social dynamics of groups, such as particular research communities. Concern about privacy, especially with respect to the contents of e-mail messages, is a major obstacle to the availability of such data. However, participants emphasized that for analyzing network interactions, access to message content is often not necessary. Hence, if identifiers are stripped from the messages and the content remains unread, there will be fewer impediments to the availability of such data.

The second plenary speaker, Peter Turchin, focused on the validation of dynamic models with time-series data and highlighted two validation techniques. The first technique was hypothesis testing, where validation is performed by comparing two or more alternative hypotheses (each translated into a mathematical model) which purport to explain the behavior of a certain dynamical system. Given that time-series data documents fluctuations of key variables, it is in general possible to decide which hypothesis/model does the best job of accounting for the observed dynamics by contrasting the model predictions with data patterns. Turchin then discussed the use of an alternate type of validation technique, cross-validation, for relatively short and noisy data sets that are typical for biological and social settings. In cross-validation techniques, models are fit on part of the data and then tested for fit using the rest of the data set. Turchin illustrated this approach with two case studies - the dynamics of religious conversion and linkages between oscillating population numbers and sociopolitical instability. Questions following this presentation considered several issues including observational noise in the models (arising from the amount of available data), choosing the endpoint in time for religious conversion and the match to spatial processes of the religious conversion, and the stationary character of population data within noise limits.

http://seattle.intel-research.net/MSD/talks/Turchin.ppt
http://seattle.intel-research.net/MSD/talks/Huberman.ppt

BREAKOUT SESSIONS AND GENERAL DISCUSSION

Following the plenary sessions, the workshop participants divided into breakout groups to discuss various ways of validating models, before reporting back to each other with further discussion. Breakout groups focused on two levels of validation – internal and external.

Internal validation

Before testing models against empirical observations and data, researchers must demonstrate that their models are internally consistent. Participants discussed two mechanisms for evaluating the internal consistency of models. First, there must be some evaluation of the logical and mathematical validity of the model. In cases where models are too complex to be rendered or solved analytically, computer simulation can be a useful tool for internal validation. Most importantly, simulation may provide a means for falsifying a theory or a model, or showing proof of concept. For example, a model indicating that a certain phenomenon cannot occur can be demonstrated as invalid if a computer simulation operating under the assumptions of the model shows that this phenomenon is indeed possible.

A second mechanism for the internal validation of models lies in aligning or docking. The term docking (or alignment) is used to describe the process of comparing two models that are based on the same question or idea, but have been implemented independently and differently. In docking one model with another, researchers observe whether their results overlap or match, and whether one may subsume the other. For example, rendering a model using a different computational approach may produce the same outcome. Docking is a useful mechanism for validating and improving models, as the inability to dock one model to another provides an opportunity to identify flaws in the model. Conversely, the ability of one model to match another suggests that the model is robust. For example, the classic Schelling model of segregation has been successfully re-rendered into many different variations. In contrast, results from the Garbage Can model of organizational decision-making have been more difficult to reproduce.[12] Models do not necessarily have to be docked to be internally valid. Indeed it may be that there is no appropriate model for comparison.

[12] Cohen, Michael D., James G. March, Johan P. Olsen 1972. "A Garbage Can Model of Organizational Choice." *Administrative Science Quarterly*, 17 (1): 1-25.

External/empirical validation

After establishing internal validity, models are tested with empirical data to establish external validity. Models are empirically validated at several different levels depending on their intended use.

Qualitative validation

Qualitative validation involves the comparison of the shapes of distributions, patterns, and trends of overall outcomes predicted by a model with observed patterns of outcomes. One particularly important application of qualitative validation is determining whether a model based on individual microscopic behaviors can produce a macroscopic pattern that matches a system-level phenomenon observed in the real world. In other words, in qualitative validation, researchers assess whether a micro model generates the observed macro pattern.

The replication of an observed empirical pattern by a model can be a powerful approach that does not require the collection or analysis of numerical data. By deriving a qualitative outcome logically from micro-level assumptions, models may help researchers to identify local generative rules and mechanisms that underlie the observed pattern and serve an explanatory role for particular phenomena. The Schelling model of segregation discussed previously provides an excellent example of this form of validation. In this case, simple micro-level assumptions of agents' preferences about their neighbors give rise to a counterintuitive, macro-level outcome - namely, segregated residential populations. In this case the model does not need to account for each individual's residence, but simply to replicate the overall pattern of segregation, despite tolerant attitudes of individuals in their residence choices.

However, while Schelling's simple model is a suggestive illustration, it would not have served this purpose if it had taken into account the many realistic details of empirical housing choices, such as housing prices, crime, highways, facilities and schools. This highlights a limitation of agent based models in the social sciences. As the underlying model becomes more complex (and presumably more realistic), its overall behavior is no longer transparent enough to illuminate the mechanism that underlies the phenomenon being modeled.

Qualitative validation is also useful for evaluating alternative models. An example of this use of qualitative validation is the study of the distribution of farming contracts conducted by Peyton Young in Session 1. Where standard economic theory suggests that there would be a lot of variability in the contracts, Young's modeling of the distribution of farming contracts in Illinois showed distinct clustering in the types of contracts chosen. A qualitative comparison of the model outcomes with the empirical data showed similarities in the overall trends, and thus provided support for Young's alternative to standard economic theory. One caveat to this, however, is that alternative models grounded in different sets of assumptions may generate the same predicted outcome. If the common predicted outcome matches the empirically observed pattern, it is impossible to identify which of the multiple competing mechanisms/models is best without further analysis or discrimination on another criterion, such as parsimony or generality.

Quantitative validation

Many models require more detailed validation against data that go beyond matching outcomes to qualitative empirical observations. Ideally, social science modelers would be able to design experiments to generate the data to test theoretical ideas and mechanisms. However, researchers are often unable to study social systems using experimental methods. In these cases, modelers typically make use of available observational data or data measured in non-experimental settings to assess the validity of models.

Researchers typically use statistical inference to assess the fit of a model to observational data. In some cases, they may first tune the model by using a training data set to define the parameters of the model, and then test the predictive power of the model for its fit to an external test data set. Almost always the ability to generalize to new data is important and especially when the amount of data available is limited. *Model selection* or *model fitting*, is the problem of picking among different mathematical models which all purport to describe the observational data. Selecting a model that has good generalization properties is often done via direct optimization on some measure of goodness of fit or risk of over-fitting on training data. Optimizing the fit of a model with some model-dependant penalty is a common approach to model selection, where complex models or models with many parameters are penalized, as in the Akaike (AIC) or Bayesian Information Criteria (BIC). Cross-validation is another approach used to estimating the ability to generalize to new datasets. In cross validation, the data set is split into distinct training and test sets. The model is specified using the training set and then evaluated on the test set. Turchin used cross-validation to assess fit of two models of population growth and sociopolitical instability in three different historical contexts. Model averaging and structural risk minimization (which penalizes a class of models) are other techniques commonly used for model fitting.

Challenges and considerations

There are several challenges in fitting models to data that are particularly pertinent to the social sciences. First, in contrast to simpler linear regression models, social dynamics models involve accounting for the interactions and mutual influence of rational, adaptive agents. The complexity of such models often makes it difficult to interpret the underlying parameters and to assess the quality of fit to the observational data. Unfortunately, simpler models often do not adequately capture the complexities of real world social phenomena. Consequently, finding the right balance between parsimony and accuracy is difficult.

It is also important to recognize that assessing the fit of a single model to data on its own may not be particularly informative. The value of a given validation metric in isolation may seem promising, but calculating the same metric for a different model may reveal a superior fit to the data. Therefore, as with the qualitative reproduction of empirical patterns by models, when validating models in their fit to data, it is more meaningful to compare the outcome with a single or several explicitly defined alternatives.

There is no universal validation metric, because scholars have a variety of research goals. For example, researchers in network dynamics may consider not just the basic fit to data, but may also include a validation penalty for increased model complexity. In addition, if a model is intended to be used for the design of interventions, it is also important to account for this use in the validation metric, as the errors that may be relatively insignificant in an open-loop context can become amplified in a closed-loop situation.

Researchers may choose either qualitative or quantitative validation to suit their research problems. Sometimes models may not be particularly amenable to quantitative validation or appropriate data may not be available, and the model may still have qualitative utility in explaining a particular phenomenon or producing a theory or metaphor which serves as a catalyst for further discussion of a given issue. Similarly, it may unnecessary for the purposes of the researchers to perform a full quantitative fit of a model to data. For example, in statistical physics, it may be sufficient to recognize that a system exhibits power-law scaling of the form $P(k) = A \times k^{-c}$, and identifying the exponent *c*. However, in the disciplines of engineering or materials science, where the integrity of infrastructure such as bridges relies on accurate quantitative modeling, it is necessary to not only recognize the operation of the power law and identify the exponent c but also to identify the value of the pre-factor A.

Forecasting

Forecasting provides another type of validation. Once a model has been defined and its parameters have been specified, it may then be used to forecast future outcomes. Forecasting involves running the model and recording the output and the forecasts may produce either qualitative or quantitative predictions. The generated predictions may support practical activities such as policy analysis and planning, or subsequently fit to new empirical data as it emerges over time to provide further model validation.

Data Availability and Data Collection

Obtaining data of sufficient quantity and quality is critical to developing validated models that produce credible results. The discussion following the second plenary talk by Turchin highlighted this point. In modeling the historical spread of religions, poor quality data resulted in increased uncertainty around the curve generated by the model.

Participants recognized that large quantities of data are becoming increasingly available because of the digitization of financial and other types of records and electronic traces generated by web-based communities using internet technologies. Having large amounts of high quality data makes it easier to do robust testing of social dynamics models. In fact, increased data availability may help researchers to explore social dynamics phenomena in ways that were previously not possible. During the first plenary talk, Huberman demonstrated how data derived from internet traffic and internet shopping had enabled the testing of complex models of internet congestion (an example of a social dilemma) and viral marketing (an example of information propagation). Though these two topics have attracted much interest from theoretical researchers in the past, previously testing theories in these areas relied primarily on computer simulation or small laboratory experiments. The more recent availability of large amounts of data has enabled the development of high fidelity models to explore and support theories of social dynamics. These developments also promise to foster the ability to build better connections between theory, the laboratory and real world data.

Input and Output Validation

In validating models against empirical data, there is a distinction between input and output validation. Output validation refers to the comparison of the results generated by the model with real-world observations; i.e. how well does the model output predict empirical data? Input validation, on the other hand, requires an evaluation of the realism of the underlying assumptions of the model, including its logical and functional forms parameter values, and initial conditions. In this case, the original assumptions and values are not arbitrarily specified but are validated against empirical observations, often using both quantitative and qualitative data. The extent to which a researcher is interested in either deducing the logical consequences of a set of empirically validated propositions (input validation) or in developing a model that accurately predicts empirical events or trends even if the input assumptions may be unrealistic (output validation) depends on the researcher’s overall goals. Of course, the two approaches to validation may be combined in the same research program.

SUMMARY

The process of validating models includes several complementary layers. The initial layer of internal validation involves establishing that a model is mathematically and analytically self-consistent. Models may also be aligned and docked with other models that are based on the same premise but implemented differently. Successful docking (i.e. obtaining the same result through different implementations) increases confidence in the robustness of a model.

Models that are internally robust and consistent are then subject to external validation which involves validating models against empirical observations. Qualitative model validation involves verifying that a macroscopic pattern generated by a set of microscopic behaviors in the model matches the real-world macro-phenomenon. However, it is important to keep in mind that if two models or two sets of assumptions generate the same empirically observed pattern, qualitative validation does not provide a basis for determining which model is the most accurate representation of the underlying generative mechanism.

Quantitative validation involves fitting models to empirically collected quantitative data sets. A significant challenge in validating and fitting social dynamics models to data is finding a balance between simpler and more transparent models and more complex models that are richer and more accurately capture the details of complex dynamic systems composed of interacting and adaptive agents. Careful definition of validation metrics and improved data collection provides researchers with some additional tools for striking an appropriate balance.

SESSION 3: APPLICATIONS TO THEORETICAL PROBLEMS: MODELING DYNAMICS OF SOCIAL NETWORKS

Researchers use network models to describe a broad diversity of phenomena, including the transmission of disease among living organisms, the formation and dissolution of alliances between firms or between political factions, the dynamics of fads, the spreading of email spam, or the pecking orders among chickens. Given that so many disciplines are making use of network models in some form, this session was aimed at exploring multidisciplinary approaches to modeling networks. Three plenary talks opened the session. A panel discussion followed the plenary talks and examined areas and questions in social network research that remain relatively unexplored.

PLENARY TALKS

The first plenary speaker, Noshir Contractor, described a multi-theoretical multilevel (MTML) model to investigate the dynamics for the creation, maintenance, dissolution and reconstitution of knowledge and social networks[13]. The model paid particular attention to the motivations for forming networks and contextualizing the goals of the community of interest. Contractor described studies that drew on communities and networks in several areas including emergency response networks, public health networks, and online communities from the massively multi-player online role playing game World of Warcraft. His presentation also provided a framework for illustrating how modeling social and knowledge networks enable the growth of cyber-infrastructure, which in turn enables the modeling of networks. The work pulls together and exploits recent advances in:

1) theories about motivations that underlie the dynamics of social network ties
2) cyber infrastructure and web-based technological capabilities to capture relational metadata
3) computational modeling techniques for modeling large-scale multi-agent network systems and
4) exponential random graph modeling techniques for validation

and applies them to the problem of modeling the dynamics of complex networks.

The second plenary speaker, Gyorgy Korniss discussed the "Naming Game", a prototype model for semiotic dynamics, or the emergence of language. Specifically, the model represents how shared communication schemes emerge in a population of autonomous (human or artificial) agents. This phenomenon has been observed in popular web applications where users spontaneously create loose categorization schemes to effectively organize and share large amounts of information. The Naming Game provides a minimalist approach to describe the convergence process toward a shared "vocabulary". The presentation focused on the basic temporal and spatial scaling properties of the Naming Game and considered the effect of the communication topology of the agents on the agreement process. Some of the effects observed are universal to a large class of models of opinion formation where the system eventually reaches global consensus. During the discussion following his talk, Korniss

---

[13] Monge, P. R. & Contractor, N. S. 2003. *Theories of communication networks.* New York: Oxford University Press.

Contractor, Wasserman & Faust. 2006. "Testing multi-theoretical multilevel hypotheses about organizational networks: An analytic framework and empirical example." *Academy of Management Review*, 31, 681-703.

defined the critical variable thresholds in opinion models as a measure of differences in opinions between neighbors. In particular, he noted that if a chosen threshold is too small, global consensus may never occur. However, adding random links may lead to global consensus in spite of a low threshold through the emergence of convergence clusters. A discussion question raised the possibility of the evolution of two distinct regions in a polarized system and the consequent emergence of "bilingual" agents who could communicate in both regions. Korniss acknowledged this as a possibility in the long term but noted that convergence itself may not be observed in realistic time scales and may not be reflected in the dynamic behavior at any given point in time.

The third plenary speaker, John Padgett, described an analysis of Renaissance Florence that adapted a model borrowed from biochemistry. This study focused on questions of invention and emergence of organizations, through continual flow processes, catalysis and transformation of entities to a higher organizational level. Drawing on empirical data collected from Florence over very long periods of time (200 years), the study concluded that the basic biochemical model, while a suitable starting point, was not an appropriate model for invention in Renaissance Florence. The basic framework of catalysis still applies, but inventions are best represented as novel re-mappings across multiple different types of social networks based on the same nodes, rather than a single network that is internally reproducing. In particular, Padgett discussed four multiple-network-mapping mechanisms of organizational invention that he has discovered in his research -- namely, extension and differentiation, fusion and hybridity, transposition and refunctionality, and finally multivocality and robust action. Padgett also used this finding to demonstrate that data can inform model development, as well as its conventional uses for model validation.

Links to plenary talk slides:
http://seattle.intel-research.net/MSD/talks/Contractor.ppt
http://seattle.intel-research.net/MSD/talks/Korniss.pdf
http://seattle.intel-research.net/MSD/talks/Padgett.ppt

PLENARY DISCUSSION, QUESTIONS AND RESPONSES

Following the plenary talks, the three plenary speakers, Contractor, Korniss, and Padgett formed a panel to consider areas of research interest in network modeling. The speakers proposed under explored questions or issues in the field that would serve to illuminate and increase understanding of social network analysis. Workshop participants reacted to these questions and brought up other issues that they felt were pertinent to the field. A general discussion on this area involving all workshop participants rounded out the session.

Theoretical problems in modeling networks

Participants identified three prominent classes of theoretical problems that may be explored through modeling the dynamics of social networks.

The first class of theoretical problems examines social dynamics within fixed networks and considers the diffusion and flow of information through static networks composed of fixed agents with fixed relationships. Examples of such applications include the propagation of rumors or the spread computer viruses.

The second class of theoretical problems examines the dynamics of changing network links and relationships. These problems include the formation and dissolution of connections between fixed agents, e.g. diplomatic relationships between nations or ties between clans formed through marriage. The study of Renaissance Florence described by Padgett in the third plenary presentation reflected that different networks can evolve based on factors such as kinship, economics and politics. Thus, while individual agents remain the same, multiple networks may emerge on the same nodes according to the particular paradigm of the network.

The third class of theoretical problems also looks at the dynamics of network change, but focuses on the emergence and dissolution of nodes, rather than ties, in a network. For example, Contractor's analysis of the content of situation reports (SITREPs) produced following Hurricane Katrina shows the increasing significance of particular players, organizations and concepts following Katrina's landfall in Florida and Louisiana.

These three classes of social networks are not necessarily mutually exclusive. For example, a study of citation networks of patents and publications considers the evolution of both new nodes (papers) and ties (citations to older papers). In this example, the network has some particular characteristics, mainly that ties between existing nodes are never introduced and that there is no dissolution of nodes or edges. Similarly, the model for an emergency response network described by Contractor considers the number and nature of links between agents and the implications for appropriate message dissemination through the network.

Research gaps in theoretical problems

Researchers have explored and developed these three classes of theoretical models to varying levels. Historically, most studies have focused on the dynamics of diffusion on static networks, with many applications including global ordering, epidemic spread, and information diffusion. In studying the dynamics of network change, the emergence and dissolution of links has received the most attention, whereas the emergence and dissolution of nodes has received little attention. Contractor noted that researchers have mostly viewed changing numbers of nodes as a nuisance rather than a substantive problem, and suggested that focusing on the emergence and dissolution of nodes in networks could be particularly fruitful. In addition researchers have focused on the dynamics of single networks and there is very little theoretical or empirical investigation of the "ecology of networks." That is, how do multiple, partially overlapping networks emerge, co-evolve, dissolve, and re-coalesce? Modeling these phenomena is increasingly relevant to help us understand contemporary network forms of organizing. Digital traces of the ecology of networks in virtual worlds provide an outstanding opportunity to empirically test these models.

Such research may be of most value if grounded in highly applied contexts, since dynamics and evolving network structures tend to be very complex and sensitive to the details of the model. Participants also observed that link evolution is often highly context-dependent. However, empirically motivated models can produce simple rules and generic features that reflect the core underlying properties of the network dynamics.

Another research area requiring further exploration is the development of methods for generalizing the properties of a class of networks from a single or few instances. Korniss commented that individual realizations of social network models often display atypical topologies that are incongruent with the ensemble average and therefore may not have relevance for understanding the core properties of a certain class of networks. More broadly, researchers need to consider the question of identifying concrete or quantitative bounds on the predictive power of individual realizations, which will avoid over-fitting of models to specific applications.

Multimodal networks

Contractor identified theorizing about, measuring and modeling multi-modal, or multi-dimensional, networks as another theoretical research area of increasing interest and importance. In multi-modal networks, nodes may include not only people, but a range of other elements including organizations, documents, data sets, analytical tools, or concepts. These elements may all be interlinked. For example, there can be relationships among and between sets of people and databases.

While researchers have been aware of multi-modal networks for some time, limitations on data collection have constrained the development of this area. However, as cyber infrastructure and internet technologies continue to grow, large quantities of data are now becoming available to researchers. This may enable researchers to advance a new generation of theories that need to be developed, tested and applied to the emerging multidimensional networks. This in turn opens up a broad area for further theoretical work, with a need to develop better social theories about how links are forged within such networks and how different types of networks could relate to each other.

As discussed by Padgett, further research opportunities may also lie in considering interconnected networks, in which each network has its own evolutionary logic and dynamics but is interconnected with other networks through shared or interacting nodes. This raises the question of how separate but interconnected networks may influence each other, for example, how a failure in one network can propagate to failure in another. Examples of such networks may be found in transportation, power and communication networks in engineering, and also in political, knowledge, financial and neighborhood networks in the social sciences.

Investigating the content/meaning of nodes and relations

One participant suggested that the research in the evolution of networks should give more attention to the content and meaning of nodes and their ties. This research stream relates to developing formal methods for studying the co-evolution of multiple networks. For example, financial, kinship and social networks are not independent objects, but are coupled structures with important interrelations. Exploring their co-evolution may provide further insight into all of these types of networks.

Encouraging dialog between inductive and deductive modelers

An important issue identified by both the plenary speakers and workshop participants is improving communication between researchers who conduct data-driven, empirical research and experts who build fundamental theories from the first principles. Both models that are deduced from theory and models that are induced from data have much to offer the social sciences, however, these models can be compared and used to inform each other.

Projects and initiatives that connect experts across domain specializations may promote this sort of dialog and bring advances in both computational modeling and social science theory. Participants raised the possibility of a grand challenge problem in social dynamics. This would bring experts into a common forum to compare modeling approaches, tools, and methodologies. Other possible initiatives could include holding conferences targeted at both groups, and/or funding initiatives that would explicitly reward teams composed of modelers and real world empiricists.

Data in social networks

Empirical data plays an integral role for evaluating social dynamics models. However, data is useful not just for validating models, but also for advancing theoretical development. From a contemporary perspective, Contractor's talk illustrated how using data from cyber infrastructure and web-based technology allows computational modeling of network dynamics in large-scale multi-agent systems. From a historical perspective, Padgett's talk illustrated how highly detailed data from Renaissance Florence has provided insights and enabled researchers to develop models of organizational emergence.

Most methods of network analysis assume the researcher has a network census: data on every node and every link in the population of interest. A census is clearly infeasible for most empirical topics outside of classrooms and small experiments. But to move beyond this we need a statistical framework for inference from sampled networks. Methods have been developed to handle networks sampled with egocentric and some link-tracing designs (the latter is based on "adaptive sampling" theory in statistics), but this is an area where more research would be well focused.

Collecting, consolidating, and sharing network data in comparative forms

A major challenge in validating models of social dynamics is accessing data sets that are sufficiently standardized to allow for comparative analysis. While there are many collections of existing data, such data sets tend to be of heterogeneous form and resolution and not developed to be consistent with other data sets. Therefore a key challenge is to find ways to standardize existing data sets in ways that would make them useful to the broader research community. Sometimes it may be necessary to interpolate or make inferences about data between recorded events. Imputing missing observations and fitting models from incomplete data remain significant obstacles, although more recent developments in computer science and engineering may help in this area.

It is also important that social science research does not stagnate in working only with existing data sets. This may lead to highly specialized models that are not suitably general. In this sense, efforts to consolidate existing data sets and standardize data collection methods will help to advance the field of modeling social dynamics and networks.

Another way to address this issue would be to establish broad data collection projects that apply a single data collection protocol over a large number of local networks. Providing large-scale standardized sets that contain millions of data points would enable meta-analysis and better discernment of signal and noise in models. Additionally, developing such comparable data sets would open the door for further advances. The very exercise of compiling the data would present challenges in developing tools such as new database techniques, visualization techniques and online search web-enabled services.

Experimental and micro-level studies of network dynamics

There is a need to integrate experimental techniques into the social network field in order to understand from first principles how and why networks form. At present, while there are good statistical models of overall macro-level structures, inferring the micro-behaviors responsible for generating the macrostructure has been more difficult. Developing and conducting experiments with human subjects promises to illuminate behavioral motivations in networks and lead to better theories of network formation and adaptation.

Deploying sensors for relational data collection

The use of sensors and mobile devices enable automated collection of empirical data about human interactions as well as interactions among non-human agents (i.e. computer networks). These types of data can provide valuable insights into understanding the link between micro-level actions and macro-level behaviors of large networks.

Taking advantage of relational data traces from technology

Related to the problems associated with creating large-scale standardized data sets are the difficulties associated with collecting relational data of comparable density, quality and reliability. However, with continuing growth and development in digital technologies such as phones, mobile devices, computers and the internet, there is an ever-increasing ability to digitally record large amounts of relational data. Examples of sources for such data include communal internet resources such as Wikipedia, online games such as World of Warcraft, the detection of face-to-face interactions through sensors and mobile devices, and online financial transactions. Accumulating this type of digital data opens up huge opportunities for collecting data not only on network structure but also network dynamics. Obviously, there are significant challenges and issues that must be addressed in preserving the privacy of users. Notwithstanding this, the advent of such large amounts of relational data provides researchers with an opportunity to use technology in conjunction with interdisciplinary skills to answer long-standing questions in social dynamics

Data collection from historical contexts

While there is undoubtedly great potential for collecting data on interactions by utilizing new technologies, it is also important to collect data from historical contexts. There has been relatively little work performed in relating events over multiple time scales. The dynamics of events in historical periods such as Renaissance Florence occur over much longer time scales than can be delivered by new technologies. Therefore, there is also a need to collect from past records so as to develop models of historic events and networks.

Other related observations

Expansion of approaches in modeling social networks from other disciplines

While existing network models have been well-developed, it is also important to continue exploring different approaches and to add to the existing repertoire of models and methodologies. Ideas borrowed from different fields may provide new insights into social dynamic processes. Padgett's use of a chemical autocatalytic model as a starting point for studies in Renaissance Florence is an example of applying models and principles and finding meaning for reinterpretation in the social sciences. While Padgett's results reflect that the chemical paradigm is not an appropriate model for the period studied, it still serves as a suitable starting point for inquiry.

Neutrality theory in population genetics and ecology serves as another example. Neutrality theory enables researchers to predict expected distributions over networks over time as a single parameter is tuned. This represents a move away from equilibrium models to more transient dynamics where networks are represented as evolving snapshots in time, thus providing insights into how network structures arise and change. Some of these concepts have already been imported into economics to characterize the dynamics of markets and processes of innovation. Similarly, adopting and exploring neutrality theory in the social sciences may provide insights into the underlying network dynamics.

Another possibility for expanding tools in modeling social networks comes from research performed on networks in engineering. For example, in studying power-station grids, network modelers are centrally concerned with issues of control or intervention and with the inference of internal network behavior from available measurements to facilitate and improve control. One participant suggested that similar approaches may have utility in the social sciences. As such, it may be worthwhile to bring the study of real-time interventions as well as design-time decisions in engineering more prominently into the discussion of social dynamics.

Categorization of relationships in networks

Low-level measurement and categorization of interpersonal relationships in network studies also need to be better defined. Currently, the lack of well-accepted relational taxonomy is an increasingly significant obstacle to progress in the areas of network measurement, error modeling and the standardization of data sets. Further advances in the basic foundational understanding of the building blocks of human relationships, such as what constitutes a social tie or relationship, are required.

Elucidating spatial patterns on finite irregular networks

Michael Macy suggested that another area of social network research that is worthy of attention is determining spatial patterns in irregular networks that better fit observed network topologies. The bulk of simulation work in this area has been performed on regular grids where all nodes have an identical position in the network (e.g. a set of Moore or Von Neumann neighbors). Bringing tools from graph theory to deal with finite networks in which assumptions of regularity are weakened may enable a better understanding of modal domains associated with various characteristics. This would enable the investigation of problems with more realistic network topologies.

Modeling the richness of network edges

In most network graphs presented during the workshop, the richness of nodes was characterized by different attributes, capabilities, or histories. In this light, one participant suggested that hyper-graphs could be developed such that edges, or links, the richness of edges can also be captured. Hyper-graphs can contain subsets of nodes that may overlap with each other. The speaker emphasized that such partitions should not be confused with the phenomenon of clustering on graphs. The development of hyper-edges would enable relations on edges to be richer than merely lines between pairs of nodes.

SUMMARY

Researchers can model and analyze three different classes of network dynamics problems: information flow through static networks, the emergence and dissolution of links in networks, and the emergence and dissolution of network nodes. Models developed to address the first class of problems are relatively well characterized. The second and especially the third classes of models require much further exploration. Other areas of interest in modeling social network dynamics include multi-modal networks, understanding the content of nodes and links and defining quantitative bounds on predictions from single realizations of models.

A key aspect of developing social network models is the use of data. Data sets are useful not only for validation purposes, but also be used for the development of the underlying theory. Useful data sets may be obtained from a diversity of sources, including consolidating existing network data, experimental data, and obtaining relational data through the use of sensor technology or traces of online contacts. There is also great potential in modeling networks with the collection of large, standardized data sets, either through conducting specifically designed studies, or taking advantage of the explosion of relational data arising from modern technologies. In addition to collecting data and modeling networks for recent, present and future events, it may also be useful to develop models for historical periods and events that can also provide valuable insights into micro and macro-social behaviors.

SESSION 4: APPLICATIONS TO SUBSTANTIVE PROBLEMS: MODELING DISASTERS

Disasters represent a particularly challenging class of substantive problems. Environmental disasters, terrorist attacks, sectarian feuds, and outbreaks of disease all involve disruptions of multiple dynamic systems, leading to cascading failures that require coordinated responses among multiple agents who may themselves be debilitated or disconnected. In this session, three plenary speakers discussed the challenges of understanding and modeling disasters and demonstrated progress in this field with empirical cases from their research. A discussion of research areas and questions in modeling disasters that require further development followed the plenary talks.

PLENARY TALKS

The first plenary speaker, Kathleen Carley, described the development of BioWar, a city-level, multi-agent, dynamic-network simulation system that explores the impact of epidemiological events, biological attacks and chemical spills on individuals and the community in different response scenarios. BioWar provides an automated tool for evaluating elements such as response policies, attack severity and detection tools related to biological attacks. In addition, BioWar generates high fidelity virtual data for exploring the potential impact of unprecedented events. The model was designed to facilitate systematic reasoning with respect to the rate and spread of disease, potential response scenarios and to facilitate policy design. Validation problems associated with large quantities of real data led to the development of WIZER, an inference engine for semi-automated validation and tuning of the model. Uncertainties in symptom-based diagnosis of diseases and different behaviors according to symptom severity resulted in the incorporation of a sensitivity analysis and many behavioral sub-models to enable a high degree of flexibility and the ability to model novel diseases.

The discussion following this talk established that symptom-based behavior and quarantine actions have been included in the general BioWar model. Future efforts will attempt to link these behaviors and activities with panic models. At the present time, the model does not include terrorist activity. While BioWar and WIZER are still under development, this simulation provides an example of a high-fidelity, multi-agent simulation model which can be utilized to inform the policy debates that emerge in the course of planning for and responding to disasters.

The second plenary speaker, Louise Comfort, considered the dynamics of rapidly evolving response systems following disasters and the challenges involved in modeling these systems. Within the organizational structure of emergency response systems, “nested sets” of agents at different levels represent a structure that defines specific types of action performed at successive levels of operation. In disaster management, managers need to respond and assess situations quickly. Such models consider how disaster managers select particular actions, what factors enable organizations to adapt and what factors inhibit adaptation under stress.

This talk demonstrated a preliminary model of the dynamics of the information flow immediately before, during, and after extreme events. A comparative analysis of response systems following Hurricane Katrina in Louisiana and Hurricane Rita in Texas in 2005 considered the involvement of emergency response organizations through a range of governance levels. Federal laws and policies were the same in both states. However there was a marked difference in the pattern of collaboration among organizations and jurisdictions in response to the disaster. Thus, modeling interactions within and among the technical and organizational components of the system offers a means to identify the rate and extent of change among organizations involved in coordinated action. This analysis led to the development of a new computer information system, which uses a computational model to aggregate information from different technical and social organizational sources. The information system then communicates the processed information to action agencies that can devise appropriate actions.

The third plenary speaker, Chris Barrett, discussed the need to understand the coupled evolution of human behavior, dense social networks and civil infrastructure in the context of decision support. In these situations, human behaviors, responses to public policies and crisis development processes influence social interactions. These effects must be taken into account in the development of effective strategies. Computer simulation-aided decision support tools that involve the integration of information, situation assessment and course of action analysis provide a practical approach for addressing such issues. The speaker demonstrated the construction of a simulated network-centric system that replicates population activities, event scenarios, and civil infrastructure to create a specific urban environment. This agent-based "Simfrastructure" provides a framework for integrating simulations of civil infrastructure, with applications in policy planning, situational awareness, inter-dependency analysis and economic analysis. Barrett illustrated how the simulation of epidemic spread could be used to support mitigation planning, assess the transmission of an infectious disease through a population, and examine how different interventions affect this process.

Links to plenary talk slides:
http://seattle.intel-research.net/MSD/talks/Carley.ppt
http://seattle.intel-research.net/MSD/talks/Comfort.ppt
http://seattle.intel-research.net/MSD/talks/Barrett.ppt

PLENARY DISCUSSION, QUESTIONS AND RESPONSES

As with Session 3, a panel discussion followed the plenary talks in which the plenary speakers considered and proposed areas of research interest for further advancement of disaster modeling. Workshop participants then discussed these questions.

Considerations and research needs for modeling disasters

There are many applications for disaster models, for example, policy planning, decision support, identifying needs, implementing interventions and disaster relief. Scientists have access to modeling techniques and methods that can be brought together, integrated and applied to disaster situations. Such methods include techniques from agent-based modeling, network modeling, and multimodal modeling. Furthermore, it is possible to adapt models from other fields such as biology to the analysis of disaster and emergency response systems. The use of such fundamental approaches to build tools and large-scale models stimulates further methodological development and innovation, resulting in significant advances in the quality and scope of models.

Large-scale complex models with simulated populations have the capacity not only to inform and assist decision makers and policy setters, but may also be useful resources as test beds for social scientific research. Thus, it is important that advances in the development of these dynamic tools are fed into the social modeling community where they can be leveraged to produce a sustained and long-term return on the significant effort involved in their original conception and development.

Challenges in disaster modeling

There are several challenges and elements to consider when modeling the dynamics of disasters. For example, in epidemiology, medical experts may have divergent views regarding specific symptoms and/or properties of a given disease state, thus complicating the parameterization of models and necessitating sensitivity analyses to determine the margin of uncertainty or error. The behavior of individuals as their condition changes represents another complicating factor. Models and simulations need to account for symptom-based and feedback behaviors. It is also important to obtain an appropriately realistic rendering of the model to ensure that social connections and population structures are accurately simulated and calibrated. For example, a modern-day scenario would be inappropriate for modeling the spread of the 1918 influenza epidemic because social connections and population structures would be different.

It is also important to appropriately identify the component parts of the model for effective analysis. An example of this is defining the unit of analysis that is being modeled, given that there can be many different kinds of agents in interaction. In disaster situations where there are many perhaps loosely-coupled systems in place, it may not be clear how to manage hierarchical organizations and identify the relevant unit of social action. For example, a researcher might want to model the actions of an organization that has been delegated to act on behalf of but acts independently of FEMA. In this case, the modeler has to decide whether a particular action is considered to be that of FEMA, the organization itself, or both. This kind of question is relevant not only to modeling disasters. It represents a deep social science problem that could be observed in many other contexts, and an important consideration in developing models in different fields. In response to this particular question, Comfort suggested that since different individuals perform different tasks entailing different responsibilities for different agents, the most appropriate unit of analysis may be the tasks, or actions that are performed. This definition of tasks as the lowest level unit of analysis allows the model to group actions according to performance level.

Model flexibility

A challenge in modeling disasters is the ability to build flexible models. For models to be meaningful and useful for disaster planning, training or mitigation purposes, they must be realistic, reliable, and credible. The development of such models generally requires large investments of time and money for data collection, computing hardware and human resources. As such, it is fiscally and scientifically preferable to build models that can be applied to different contexts and questions. Models can be flexible in different ways, including the ability to perform more than one type of analysis or allowing for the substitution of different kinds of activity units in the model. For example, the BioWar model has been applied to 6 different cities, 63 diseases and 2 chemical scenarios. Models may also be developed with a modular structure, where standardized modules representing smaller sub-processes are built and validated independently, and then assembled together to produce a full model. A recombination of the same or a subset of such modules may then be useful for modeling a different process.

Another way to develop model flexibility is to apply models developed for one phenomenon to another. For example, while earthquakes, epileptic seizures and stock market crashes are seemingly different phenomena that occur over different time scales, these events may display similarities in their underlying dynamics. Thus, it may be possible to transfer practical lessons, theoretical insights, and useful methodologies from one research context to another. In drawing comparisons between models, however, it is important to consider at what level analysis should take place. For example, should the model be based on symptomatic signals prior to the event or deeper underlying causes and factors? This issue would have to be resolved on a case by case basis depending on the specific problem being considered and the available data.

Workshop participants also identified other important considerations when developing flexible models. In adapting a model to a different problem, it is important to determine when a model may be changed and how much change is appropriate for the model to be both appropriate to the new context and internally consistent. Similarly, it is also important to consider how models can be built to dynamically adjust the unit of analysis and how changing the unit of analysis may influence the data reporting and needs for data validation. Finally, flexible models need to be robust within a reasonable region of parameters while avoiding the hazard of being insensitive to large parameter changes.

Historical and macro-social disasters

A lot of research has focused on the modeling of natural and biological disasters. However, relatively few studies have examined the dynamics of macro-social disasters such as state collapse or genocide. These are well-theorized areas in political science and historical sociology, but there is a general lack of explicit mathematical models to complement the many verbal hypotheses for macro-social disasters. For example, while the collapse of the Soviet Union had a great impact on people and the world in general, the majority of scientists and policy makers did not anticipate this momentous event. In order to better understand such events and learn from history's lessons, there is a need to collect more data and develop models for events and disasters not just from the recent years, but from past decades and centuries as well.

Data

In developing both network structures and simulated populations for disaster modeling, there is a need to collect sufficiently large data sets. Data comes in many forms and is collected in many ways including census data, activity surveys, maps, traffic data, newspapers and situation reports. Collaborations with workers in the specific field of interest can also provide useful data.

An important issue in modeling the dynamics of disasters, as the BioWar model illustrated, is that building models requires large quantities of both real and virtual data. Although virtual data can be formatted to match the real data, the real data is often incomplete, culled from different sources and displays different levels of granularity. Validating and visualizing large scale models containing a lot of data of inconsistent quality presents significant challenges. Data mining and pattern recognition techniques may provide effective methods for validation and new approaches to data summarization making it more amenable for visualization.

Infrastructure

Since disaster modeling is a growing field of research, there is a critical need to develop more detailed methods to facilitate the analysis of different levels of activity in large-scale systems that encompass a whole nation, and global initiatives intended to address major issues such as pollution, climate change, and the spread of infectious diseases. This task will require teams of researchers working at different locations and computational, organizational, and institutional infrastructures that can facilitate data sharing and exchange. It will also require different modes of representing information to the practicing communities, so they may review and provide feedback on research findings.

Cyber-infrastructure

The continuing growth of cyber infrastructure and web-based technologies should be extremely useful for research in modeling disasters. Interactions are increasingly becoming embedded in networks where individuals are able to link not only to other people, but also to documents, data sets, analytic tools, and concepts, which may also all be linked to each other. Cyber infrastructure may provide the technological capability to capture relational metadata from diverse sources, facilitate the feeding of data into models in different locations and disciplines and help to improve the understanding of communities. Cyber infrastructure also has the potential to facilitate connections among researchers and provide modeling communities with collaborative tools. Such tools may include shared data support, shared computing cycles and shared visualization-analytic tools, as well as network recommender systems where information about tools, data sets, documents and resources could be exchanged.

The development of increasingly large scale and complex models calls for greater computing capacity to manage increasingly large amounts of data, extensive simulations and complicated calculations. Similarly, as new and more innovative tools and methodologies continue to be developed, software and hardware requirements will also become significantly more demanding.

Models in practice and public policy

Models for disaster scenarios are usually developed to assist in the planning, training or mitigation of disasters. As such, there is an important feedback process in which the results of the model are communicated to planners and operational staff who are responsible for implementing the action. In using models to support policy and decision making, it is important to consider how the information is presented to the people who make and implement public policy. Researchers need to develop suitable frameworks whereby complex information is presented so that practicing managers can quickly understand a situation while allowing these same managers to drill down to more detailed information if required. This key aspect of implementing disaster models requires an interdisciplinary approach to bring together researchers and substantive field experts to determine appropriate levels of information and communication. Furthermore, different model users are likely to have different levels of understanding of disaster situations as well as different sorts of background knowledge. For example, volunteer firemen in rural areas will have a very different understanding from experienced firemen in urban areas. Therefore, it is important to identify and develop different modes of presentation for different audiences and levels to maximize the effectiveness and applicability of the model.

Developing models is a costly activity. However, it is not as costly as the devastating health and infrastructural consequences of disasters. There is a crucial need for policy makers to recognize modeling as a part of the mitigation process in disaster cycles. Appropriate support will facilitate the development of better models to improve processes, systems and services both in preparation for and in response to disasters. In this regard, government bodies should be encouraged to view the development of models as a preventative investment with a relatively small expenditure having potentially substantial returns through more efficient management systems and better emergency response strategies which save not only resources, but also human lives. Private entities and enterprises may also have a possible role in supporting this kind of work. For example, insurance companies may have interests in helping to minimize losses due to disaster-related damages. Similarly, some computing companies may be altruistically motivated on the basis of public health and safety to provide suitable computing infrastructure for developing models. While recognizing the key and central responsibility of government in responding to and managing disasters, the collaborative potential of investments from private enterprises is worth considering.

SUMMARY

The modeling of disasters brings together many of the themes and techniques discussed throughout this workshop. Realistic real world models and simulations of disasters are almost always large-scale and highly complex. These sorts of models require large amounts of data which may be of variable quality. These characteristics make model validation a very challenging endeavor. Other challenges include dealing with theoretical uncertainties, discrepancies, and disagreements among medical and emergency response professionals, accounting for feedback behaviors as the status of agents change, and appropriately defining the unit of analysis in a given model.

While there are significant ongoing advances in the field of disaster modeling, the plenary speakers and participants identified some ongoing areas of research interest. Building flexible models that can be applied to different contexts or perform different analyses is a potentially valuable area of development. Further developing the technological and computational infrastructures with increased capacities for data management, sharing of tools and methods, and simulations and calculations, promises to help advance the development of disaster modeling. In addition to natural and biological disasters, it is also important to develop models for macro-social disasters. Computational modeling of historical or political events is a research area of great potential interest that remains relatively unexplored.

Finally, it is important to recognize the interplay of modeling disasters with the real world. Given that models are intended to facilitate disaster mitigation and preparation, feeding model outcomes back into the real world context is an integral part of this process. There are important challenges in communicating with agents in the field so that these agents interpret and act on model outputs in an appropriate manner. There is also a need to raise general awareness of the value of developing models for disasters and to encourage both government and private entities to view and invest in this research as part of the disaster mitigation process.

SOLUTION AREAS FOR MODELING SOCIAL DYNAMICS

Workshop participants identified short-term and long-term challenges faced by research programs in social dynamics and also suggested ways that the scientific community may address these challenges. This section summarizes these problems and solutions, concluding with specific recommendations for the NSF.

INTERDISCIPLINARY RESEARCH: HYBRID MINDS AND HYBRID TEAMS

Participants agreed that interdisciplinary perspectives will be an essential part of any solution to problems in modeling social dynamics. They identified two distinct approaches, one based on diverse collaborations of scholars trained in different disciplines ("hybrid teams") and the other emphasizing interdisciplinary training of individual researchers ("hybrid minds").

Interdisciplinary teams draw on the expertise of multiple fields, but also present new challenges. When appraising research grants and projects, it is difficult for substantive experts from different fields to fully appreciate and properly evaluate the merits of interdisciplinary work. Collaborating researchers from different fields also face challenges in finding common ground in conception, terminology and expectations of models.

Researchers who are trained in multiple disciplines may bridge the divide between knowledge bases, combining skills and perspectives from multiple fields. These "hybrid minds" can resolve some of the logistical problems in interdisciplinary teams, but such broadened knowledge can also compromise depth of understanding. A social scientist who dabbles in modeling or a mathematician who builds models with only partial comprehension of the substantive social context are both likely to be limited in their ability to develop complex and enlightening models. Indeed, dialog and collaboration between two researchers with deep expertise in different fields may prove to be more fruitful than two researchers with middling knowledge of both fields.

Participants agreed that both approaches to interdisciplinary research are valuable. There is a need to integrate training in dynamic modeling with exposure to substantive problems in the social sciences, while at the same time there is also great value in promoting conversation and collaboration between experts in different fields.

PROMOTING INTERDISCIPLINARY DIALOG AND COLLABORATION

Bringing together researchers from different fields into productive collaborative relationships poses several significant challenges. Such challenges include resolving differences in terminology, reconciling alternative research goals and criteria for evaluating research, as well as logistical constraints of researchers working in different geographic regions. Addressing these issues will enable efficient and sustained diffusion of theoretical models and tools across disciplines.

Conference/workshop

Participants felt that this NSF workshop was very constructive, and suggested that periodically holding similar meetings would help to form the modeling community, encouraging sustained exchange of ideas and building lasting research relationships across disciplines.

Resources for interdisciplinary research

Participants discussed ways to maintain ongoing links and discussions in the interim between conferences and workshops. Researchers who are new to the field or branching into a novel project area could use a centralized database or clearinghouse to identify other researchers who may be conducting relevant research and who thus may be suitable collaborators. Such a resource could support the growth and integration of the modeling community.

Establishing an ongoing electronic dialog would also help to facilitate the exchange of ideas among researchers. Online discussion spaces could also allow scientists to share beta models or technical reports for peer feedback. It may be useful to establish a research platform for rigorous testing of scientific models of social systems, similar to the NSF-led GENI (Global Environment for Network Innovation) in the computer science field (http://www.geni.net).

Special issues of journals in both the social sciences and physical sciences can identify domains for fruitful interdisciplinary collaboration in modeling social dynamics. For example, recent issues in *American Journal of Sociology, Academy of Management Review*, and *Simulation Modeling Practice and Theory* have promoted these intellectual bridges. Such special issues provide researchers with a tangible publication goal, help bring together different lines of research, and raise the profile of the field in general.

Ongoing interdisciplinary research centers

Interdisciplinary centers provide crucial support for long term collaborations across disciplines. Such centers develop local intellectual communities by offering seminar series, visiting scholars, and ongoing workshops and short courses. Most importantly, they offer sustained research infrastructure, such as data resources, computer platforms, professional programmers, technical staff and administrative assistance to support ongoing research programs.

Investing in sustained support and institutional incentives for interdisciplinary research centers can further advance and nurture the field of social dynamics. Center grants provide tangible incentives for institutions to hire and increase the presence of interdisciplinary researchers. Interdisciplinary training grants similarly promote intellectual ties among faculty as well as training a new generation of researchers. Such sustained support can build mature and productive intellectual communities. For example, the NIH's support of the field of demography over 30 years in the form of center and training grants resulted in a large, active self-sustaining interdisciplinary field.

Grand Challenge Problem

Interdisciplinary collaborations can be particularly productive when considering applications of diverse theoretical approaches to well-defined substantive problems. A tangible 'grand challenge' problem would serve as a focal point for research efforts and innovation, and would provide a context in which to compare different methodologies and tools. Inviting scholars with different disciplinary backgrounds to develop new ways of modeling the problem would open the field for the development of novel and fruitful approaches.

Grand challenge problems can be selected to address important questions of national and societal interest. For example, researchers from a variety of disciplines can tackle problems of disaster planning and mitigation, the social dynamics of network organizational forms, and the propagation of diseases in local and global populations. An important part of defining such a challenge would be to choose a problem for which interventions can be implemented and tested, thus enabling a substantive evaluation with practical implications beyond the scientific community.

EDUCATION

While interdisciplinary collaboration between researchers should be encouraged and supported, there is also a critical need to improve interdisciplinary training through undergraduate, graduate and post-doctoral levels in order to nurture future researchers with skills and appreciation across disciplines. Providing undergraduate curricula in social sciences in tandem with mathematical and programming courses would provide a fundamental basis for further training at the graduate levels. Innovation in this area may be stimulated by ongoing support of interdisciplinary training programs that result in institutional and curriculum change. A few successful examples include the numerous NSF-supported IGERT programs related to modeling social dynamics – such as *Computational Analysis of Social and Organizational Systems* (CASOS) at Carnegie Mellon, *Nonlinear Systems* at Cornell University, *Institutions, Diversity, Emergence, Adaptation, and Structures* (IDEAS) at the University of Michigan, and *IGERT Program in Evolutionary Modeling* (IPEM) at Washington and Washington State universities.

In addition to these long term training programs, short term cross-training programs for graduate students and post-doctoral researchers to learn new skills in different disciplines also help to foster interdisciplinary thinking and build connections between researchers. For example, the NSF funded *International Workshop and Conference on Network Science 2006* brought together over 60 graduate students from the physical, life, and social sciences for a week of intensive tutorials and workshops in the use of network analysis techniques from diverse disciplines.[14] Several interdisciplinary research and training centers – such as CASOS above, the University of Michigan Center for the Study of Complex Systems, and the Santa Fe Institute – also offer such short-term programs.

Funding of training programs, graduate degrees and post-doctoral fellowships that focus on interdisciplinary work and enable students and young PhDs to work in (and possibly transfer to) fields outside that of their formal training would also be valuable initiative. There may also be an opportunity to encourage career development through the sponsoring workshops to enable publication and presentation of results by early career researchers.

DATA COLLECTION

One of the great challenges facing researchers in social dynamics is the disjuncture between basic theoretical models and the data that may be used to specify and validate the models. As models become more complex, researchers have increasing demands for empirical data. Workshop participants discussed promising recent developments in the increasing availability of data on social behavior through digital records such as logs of e-mails, telephone calls, online transactions, GPS location records, or other electronic traces. The availability of such data makes it possible to address larger and more complex problems than have been considered with manually-collected observational data. These rich resources also present significant challenges, however. Issues of personal privacy must be carefully attended, so standards for data collection and processing need to be developed and refined. Even practical issues such as storage, management, and archiving of data become increasingly demanding for data repositories of such large scope. These resources will drive development of new data processing, visualization and search techniques across disciplines. This is an area where developments in cyber infrastructure and web-service technologies may be useful in providing mechanisms to help preserve, archive and make data available to researchers. Of course, such efforts should be executed in tandem with theoretically motivated research programs that guide collection and processing of data.

---

[14] http://vw.indiana.edu/netsci06/

GENERAL DIRECTIONS FOR INTERDISCIPLINARY RESEARCH

While interdisciplinary collaborations are crucial for the advancement of social dynamics research programs, partnerships should not be formed haphazardly. It is important to recognize that some research areas are more naturally aligned and have a greater potential to succeed than others. Problems in which different parties stand to gain by trading knowledge and techniques are more likely to be fruitful, such as where overlapping problem domains between fields provide enough common ground such that working together is both meaningful and useful for all collaborators. For example, studies in the hazards field require input from wide-ranging groups such as seismologists, engineers and sociologists, resulting in the natural formation of interdisciplinary groups. Other potential problem areas that may naturally lend themselves to interdisciplinary collaboration include human impacts of global climate change, the social impact of technology and inference for processes with complex dependence. Thus, encouraging long-term growth in areas where natural synergies and points of contacts are present, focusing on naturally strong links and bringing together researchers with interdependent interests is more likely to be effective and fruitful than assembling a random group of experts from different fields to work together on an arbitrary problem.

While identifying areas of natural interdisciplinary synergy to maximize the effectiveness of research efforts, it is also important to recognize that cutting edge projects involve a certain level of risk. Projects' success becomes less certain with increasing novelty of method, data source and theoretical basis, as well as increasing diversity of disciplines involved and the scope of the problem. Since many project areas in social dynamics involve such risk factors, research in this area is inherently risky. However, truly significant and novel advances in science only come when the boundaries of current knowledge base are challenged, and as such, it is important to recognize the value of supporting high risk, cutting edge projects that will not always succeed. Projects that do succeed are likely to bring great rewards and developmental acceleration for the entire field. Furthermore, such projects may take longer to achieve success and therefore it is also important to sustain support over a longer time-frame. This is not to suggest that poor projects should be supported, but rather, that some high-risk and longer term perspectives should be adopted for continuing growth and advancement of the field.

RECOMMENDATIONS FOR THE NSF

Building on the challenges and solution areas identified by workshop participants, the following recommendations for NSF aim to help promote and facilitate interdisciplinary research and to advance the field of modeling social dynamics.

- ***Support regular conferences and workshops on modeling social dynamics***
  Regular meetings will allow researchers in the field to meet, exchange ideas and establish collaborations in a focused forum, as well as providing an opportunity to define the field and direct its progress.

- ***Provide ongoing support and meso-level infrastructure for interdisciplinary centers***
  In addition to funding individual research projects, the NSF should invest in research centers that facilitate and support ongoing interdisciplinary collaboration in social dynamics, as well as disseminating tools and data to the broader research community. Sustained funding to meet infrastructural, administrative and process-related needs in interdisciplinary centers will enable such centers to focus on fundamental research and also provide incentives and greater recognition for interdisciplinary research at the institutional level. NSF initiatives in cyber

infrastructure should also consider applications to the field of social dynamics and the increasing potential for promoting the exchange of data and ideas.

- ***Define agenda for interdisciplinary research***
  To most effectively conduct interdisciplinary research and establish collaboration, it is important to identify areas and problems that have natural links and points of contact between different disciplines. With a long record of funding research grants, NSF has a rich repository of historic data on different kinds of research projects and collaborations that have been established. NSF could establish focus groups to study this history and determine what kind of collaborations have been successful in the past and also what factors contribute to an effective research partnership. This could then inform and enlighten future collaborative efforts and enable the identification of suitable research projects and areas for interdisciplinary research.

- ***Define a grand challenge problem to be addressed by researchers in the field***
  Given its stature and broad exposure to researchers' agendas, NSF is well-placed to define a grand challenge problem in modeling social dynamics that will bring researchers from different fields together. Furthermore, the NSF has a broad-reaching mandate and well-established history in successfully facilitating such research initiatives, as well as resources that would enable a suitable grand challenge problem to be defined in the context of the current field of knowledge.

- ***Continue supporting teams in areas with established traditions for interdisciplinary exchange***
  Although it is exciting to sponsor novel connections among disparate research areas in the social and physical sciences, it is difficult to identify ex ante which collaborations will foster significant innovations and which will demand much effort with little payoff. Certain highly general theoretical frameworks (such as evolutionary game theory and graph theory) that are very relevant for modeling social dynamics, have been successfully applied and extended by scientists across multiple disciplines. Such common languages help research teams work effectively and also help to identify research problems and tools that will be fruitful for future collaborations.

- ***Support the development of mathematical and computational approaches to modeling the dynamics of social networks***
  Models that help us understand dynamic interdependence among social agents have broad applicability (e.g. understanding diplomatic relationships between nations, coordinating emergency response, etc.). Although there has been some progress in developing models that capture and explain the temporal dynamics and the emergence and dissolution of nodes and links, much more research remains to be done in this area. Supporting interdisciplinary research and intellectual exchange in computational and mathematical models of network evolution could significantly deepen our understanding of social dynamics.

- ***Connect the development of formal theory and modeling tools with the analysis of real world data***
  NSF can serve to bridge the divide between scholars who develop formal theory and researchers who collect and analyze real-world data. Research teams that span this divide will be better positioned to fill the goals outlined in this report. Accordingly, NSF may call for such teams in requests for proposals and may sponsor extended workshops or summer programs that include scholars from both sides of the divide.

- ***Support new tools for automatic collection of longitudinal data, with emphasis on longitudinal network data***

  Most data used in social science research are neither longitudinal nor relational, as such data collection is difficult and expensive using common measurement tools. The paucity of longitudinal network data leads to a widening rift between empirical research and much of the cutting-edge theoretical work in modeling of social dynamics. New technologies for automatic data collection – such as logs of telecommunications or sensor measures of face-to-face interaction – offer opportunities to collect rigorous records of social dynamics extensively and inexpensively.

CONFERENCE PARTICIPANTS

Name	Institution	e-mail
Robert Axtell	Brookings Institution/George Mason	raxtell@brookings.edu
Yaneer Bar-Yam	New England Complex Systems Institute	yaneer@necsi.org
Chris Barrett	Virginia Tech	chris.barrett@vt.edu
Carter Butts	University of California, Irvine	buttsc@uci.edu
Kathleen Carley	Carnegie Mellon University	kathleen.carley@cs.cmu.edu
Tanzeem Choudhury	Intel Research/Dartmouth	tanzeem.choudhury@dartmouth.edu
Louise Comfort	University of Pittsburgh	comfort@gspia.pitt.edu
Noshir Contractor	Northwestern University	nosh@northwestern.edu
Péter Érdi	Kalamazoo College	perdi@kzoo.edu
Irfan Essa	Georgia Tech	irfan@cc.gatech.edu
Bernardo Huberman	Hewlett-Packard Labs	bernardo.huberman@hp.com
Henry Kautz	University of Rochester	henrykautz@yahoo.com
James Kitts	Columbia University	jak2190@columbia.edu
Gyorgy Korniss	Rensselaer Polytechnic Institute	korniss@rpi.edu
Steve Lansing	University of Arizona	lansing@santafe.edu
Raima Larter	National Science Foundation	rlarter@nsf.gov
Michael Macy	Cornell University	mwm14@cornell.edu
Madhav Marathe	Virginia Tech	mmarathe@vbi.vt.edu
Martina Morris	University of Washington	morrism@u.washington.edu
John Padgett	University of Chicago	jpadgett@uchicago.edu
Peter Turchin	University of Connecticut	peter.turchin@uconn.edu
George Verghese	Massachusetts Institute of Technology	verghese@mit.edu
Peyton Young	John Hopkins University	pyoung@brookings.edu

PARTICIPANT INFORMATION

Robert Axtell
raxtell@brookings.edu
http://www.brookings.edu/scholars/raxtell.htm

Economic Studies Program
The Brookings Institution
1775 Massachusetts Ave NW
Washington DC 20036

Rob Axtell has recently joined the faculty of the new Ph.D. Program in Computational Social Science at George Mason University. He is a Nonresident Senior Fellow at the Brookings Institution and an External Professor at the Santa Fe Institute. He holds an interdisciplinary Ph.D. from Carnegie Mellon University. His research involves the intersection of social science and computer science, through so-called agent-based computational models of social phenomena, in which populations of autonomous software agents—each agent representing an individual person—are instantiated. The agents then interact directly with one another according to simple rules of behavior, with patterns and structure emerging at the aggregate level. His book "Growing Artificial Societies: Social Science from the Bottom Up," (MIT Press, 1996; with CD), co-authored with J. Epstein, was an early statement of the potential of multi-agent systems modeling in the social sciences. His research has appeared in academic journals (e.g., "Science," "Proceedings of the National Academy of Sciences," "Economic Journal," "Computational and Mathematical Organization Theory") as well as being reprised in the popular science press (e.g., "Scientific American," "Science News," "New Scientist," "Discover," "Technology Review") and in newspapers and magazines (e.g., "Wall St. Journal," "Los Angeles Times," "Washington Post," "Atlantic Monthly," "New Yorker"). He has recently finished a new book entitled "Artificial Economies of Adaptive Agents: The Multi-Agent Systems Approach to Economics," to be published in 2007.

Yaneer Bar-Yam
yaneer@necsi.org
http://www.necsi.org

New England Complex Systems Institute
24 Mt. Auburn St.
Cambridge, MA 02138

Professor Yaneer Bar-Yam is Founding President of the New England Complex Systems Institute. His research focuses on developing complex systems concepts and applying them to diverse areas of scientific inquiry and to social concerns and problems. In particular, he is interested in fundamental properties of evolution and learning, the evolutionary origins of altruism and other collective behaviors, the relationship between observations at different scales, the relationship of structure and function, information as a physical quantity, and quantitative properties of the complexity of real systems. Applications have been to social, biological and physical systems.

Professor Bar-Yam has been extensively involved in education of complex systems concepts to academic and professional audiences. In particular, he has taught about complex systems around the world (in Australia, Canada, China, Columbia, France, Italy, Japan, Korea, Portugal, Russia and many places in the US). He has given courses for The Centers for Disease Control and Prevention, The Centers for Medicaid and Medicare Services, a wide variety of healthcare organizations, the UN, the World Bank, the US military (Navy, Air Force, Army and Department of Defense) and intelligence community, and engineering contractors (Boeing, Lockheed-Martin, Raytheon, SAIC).

Professor Bar-Yam is chairman of the International Conference on Complex Systems, managing editor of InterJournal, and author of over 150 research papers in professional journals. His book "Dynamics of Complex Systems," published in 1997 by Perseus Press, provides a wide ranging perspective on the field of complex systems. His recent book “Making Things Work: Solving complex problems in a complex world” describes his experience in addressing complexity in military, healthcare, education, third world development, and ethnic violence arenas.

Carter Butts
buttsc@uci.edu
http://www.faculty.uci.edu/profile.cfm?faculty_id=5057

Department of Sociology
University of California, Irvine
SSPA 2145
Irvine, CA 92697-5100

My research involves the application of formal (i.e., mathematical and computational) techniques to theoretical and methodological problems within the areas of social network analysis, mathematical sociology, quantitative methodology, and human judgment and decision making. Currently, my work focuses on: the structure of spatially embedded large-scale interpersonal networks; models for informant accuracy, network inference, and graph comparison; graphical representations of life history data; and models for human behavior in strategic situations. I am also interested in social phenomena related to crisis situations, and am involved in research which seeks to combine social science and information technology to improve crisis response.

Kathleen Carley
kathleen.carley@cs.cmu.edu
http://www.casos.cs.cmu.edu/bios/carley/carley.html

Institute for Software Research International
Carnegie Mellon University
Wean 1323 - CMU, SCS, ISRI
5000 Forbes Ave
Pittsburgh PA 15213

Kathleen M. Carley's research combines cognitive science, social networks and computer science to address complex social and organizational problems. Her specific research areas are dynamic network analysis, computational social and organization theory, adaptation and evolution, text mining, and the impact of telecommunication technologies and policy on communication, information diffusion, disease contagion and response within and among groups particularly in disaster or crisis situations. She and her lab have developed infrastructure tools for analyzing large scale dynamic networks and various multi-agent simulation systems. The infrastructure tools include ORA, a statistical toolkit for analyzing and visualizing multi-dimensional networks. ORA results are organized into reports that meet various needs such as the management report, the mental model report, and the intelligence report. Another tool is AutoMap, a text-mining system for extracting semantic networks from texts and then cross-classifying them using an organizational ontology into the underlying social, knowledge, resource and task networks. Her simulation models meld multi-agent technology with network dynamics and empirical data. Three of the large-scale multi-agent network models she and the CASOS group have developed in the counter-terrorism area are: BioWar a city-scale dynamic-network agent-based model for understanding the spread of disease and illness due to natural epidemics, chemical spills, and weaponized biological attacks; DyNet a model of the change in covert networks, naturally and in response to attacks, under varying levels of information uncertainty; and RTE a model for examining state failure and the escalation of conflict at the city, state, nation, and international as changes occur within and among red, blue, and green forces. Dr. Carley is the director of the center for Computational Analysis of Social and Organizational Systems (CASOS) which has over 25 members, both students and research staff. She is the founding co-editor with Al. Wallace of the journal Computational Organization Theory and has co-edited several books in the computational organizations and dynamic network area.

Tanzeem Choudhury
tanzeem.choudhury@intel.com
tanzeem.choudhury@dartmouth.edu
http://seattle.intel-research.net/people/tanzeem/

Intel Research
1100 NE 45th Street (6th Floor)
Seattle, WA 98105

My research focuses on machine learning and machine sensing of human behavior and group dynamics. I am particularly interested in developing computational models that capture how humans interact with their environment and with each other and how the dynamics of these interactions evolve and change over time. I am currently involved in a project titled '*Creating Dynamic Social Network Models from Sensor Data'*, where we are developing a hybrid method for rigorously observing and modeling structures of social interaction by collecting streaming sensor data (individual's physical location, speech, and motion) along with conventional survey measures.

In April 2008, I will join the faculty of the Computer Science department of Dartmouth.

Louise Comfort
comfort@gspia.pitt.edu
http://www.iisis.pitt.edu

Graduate School of Public and International Affairs
University of Pittsburgh
3221 Wesley W. Posvar Hall
Pittsburgh, PA 15260

Louise K. Comfort has primary research interests in the field of decision making under conditions of uncertainty. She has worked in the field environment of rapid response to extreme events, such as earthquakes, hurricanes and the 2004 Sumatran Earthquake and Tsunami. She is engaged in research to model the interaction among changing conditions in the physical environment, the impact of these changes on the built and social environments, and the capacity of human organizations to assess this risk and respond in innovative, timely ways to reduce the consequences for human communities. Her current research ranges from collaboration with computer scientists in developing a model of early detection of tsunamis to a trial demonstration project in modeling collaborative response to the risk of flooding among twenty-six communities in the Monongahela Valley in the Pittsburgh Metropolitan Region.

Noshir Contractor
nosh@northwestern.edu
http://www.uiuc.edu/ph/www/nosh

Department of Industrial Engineering & Management Science
Department of Communication Studies
Department of Management & Organizations
Northwestern University
Technological Institute
2145 Sheridan Road
Tech C214
Evanston IL 60208-3119

Noshir Contractor is the White Professor of Industrial Engineering & Management Science, Communication Studies, and Management & Organizations at Northwestern University. He is Director of the Science of Networks in Communities (SONIC) Group.. He is investigating factors that lead to the formation, maintenance, and dissolution of dynamically linked social and knowledge networks in communities that include profit, non-profit, government as well as non-government agencies. Specifically, his research team is developing and testing theories and methods of network science to map, understand and enable more effective (i) disaster response networks, (ii) public health networks, (iii) transnational immigrant networks, (iv) massively multiplayer online role-playing game (MMORPG) networks and (v) environmental engineering networks. His research program has been funded continuously for the past decade by major grants from the U.S. National Science Foundation with additional funding from the U.S. National Institutes of Health (NIH), U.S. National Aeronautics and Space Administration (NASA), and the Rockefeller Foundation. Professor Contractor has published or presented over 250 research papers dealing with communication. His book titled *Theories of Communication Networks* (co-authored with Professor Peter Monge and published by Oxford University Press) received the 2003 Book of the Year award from the Organizational Communication Division of the National Communication Association. He is the lead developer of *IKNOW* (Inquiring Knowledge Networks On the Web), and its Cyber infrastructure extension *CI-KNOW*, a network recommender system to enable communities using cyber-infrastructure, as well as *Blanche*, a software environment to simulate the dynamics of social networks.

Péter Érdi
perdi@kzoo.edu
http://cc.kzoo.edu/~perdi/

Center for Complex Systems Studies
Department of Physics
Department of Psychology
Kalamazoo College
1200 Academy Street
Kalamazoo, MI 49006

I have a general interest in applying the dynamical system theory to from the perspective of interacting qualities (molecules, neurons, species, ideas, innovations etc.) A recent topic is 'modeling innovation by a kinetic description of the patent citation system'. My new book, *Complexity Explained* (Springer Verlag), is in press.

Irfan Essa
irfan@cc.gatech.edu
http://www.cc.gatech.edu/~irfan
College of Computing / IIC Division
Georgia Tech
85 5th Street NW
Atlanta, GA 30332-0760

My research group is interested in behavior modeling and activity recognition. Along that thread, one of the areas we are now concentrating on is modeling and analyzing group behaviors. At the meeting, I will be interested in discussions around how we can extend from analysis of behaviors from an individual, to a group, to a larger crowd. Our interests in these topics are from the low-level tracking and sensing of people to higher-level inference and discovery of interesting and sometime abnormal (or anomalous) activities.

Bernardo A. Huberman
Bernardo.huberman@hp.com
http://www.hpl.hp.com/research/idl/people/huberman
Hewlett-Packard Labs
1501 Page Mill Rd. MS 1139
Palo Alto CA 94304

My research for the past five years has concentrated on characterizing collective intelligence and designing mechanisms for harvesting the implicit information that lies within social networks and organizations. While most people agree that a form of social mind can solve problems faster and better than individuals, it is often hard to organize a diverse network of people to do so in efficient fashion. Thus my effort at identifying communities of practice from email patterns and their access to common documents, and characterizing how information flows through such communities. Last but not least we have developed and tested an incentive mechanism that can be used with small groups of people to predict the outcome of uncertain events.

Henry Kautz
henrykautz@yahoo.com
http://www.cs.washington.edu/homes/kautz
Computer Science Department
University of Rochester
PO Box 270226
Rochester NY 14627

My research centers around inferring the behavior, plans, and intentions of people by combining sensor data and commonsense knowledge. A primary application of the work is to assistive technology for persons with cognitive disabilities, in the Assisted Cognition Project. I also participate in a University of Washington Human and Social Dynamics project that uses wearable sensors to infer human interactions. I am particularly interested in ways to learn and reason with models of human behavior that combine logic and probability.

James Kitts
jak2190@columbia.edu
http://www.columbia.edu/~jak2190

Columbia University
Graduate School of Business
704 Uris Hall, 3022 Broadway
New York, NY 10027

My research applies tools of formal demography and ecology to a broad range of social dynamics, including the births and deaths of organizations, flows of members entering, exiting, and migrating between organizations, and the diffusion of information and behavior in networks. I am particularly interested in modeling the co-evolution of behavioral norms and social networks in groups, including dynamics of convergence, factionalism, and polarization in work teams.

Gyorgy Korniss
korniss@rpi.edu
http://www.rpi.edu/~korniss/Research

Department of Physics, Applied Physics, and Astronomy
Rensselaer Polytechnic Institute
110 8th Street
Troy, NY 12180-3590

My current research focuses on dynamics and collective behavior in many-particle or individual-based models. One line of my research focuses on the effects of network topology facilitating the interaction between individuals. Examples include synchronization phenomena in static and dynamic computer networks, optimization in large-scale communication and information networks, and agreement dynamics in sensor networks. Another line of studies focuses on invasion phenomena in spatially explicit individual-based models for ecological invasion, with applications to the spread of an advantageous allele or the introduction of invasive exotic species into a habitat.

Steve Lansing
lansing@santafe.edu
http://www.ic.arizona.edu/~lansing/home.htm

Fall address:
Santa Fe Institute
1399 Hyde Park Road
Santa Fe, N.M. 87501

Spring address:
Dept of Anthropology
University of Arizona
Tucson, AZ 85721-0030

The new field of molecular anthropology has begun to offer fresh perspectives on human prehistory; one of its main foci is to dissect the historical processes that shape broad-scale human genetic patterns. This adds a realistic temporal dimension to our understanding of the social worlds that humans construct for themselves. But this added realism comes at a cost, as we move from simple equilibrium systems to non-equilibrium conditions. We have already outrun the mathematics of classical population genetics, which was developed to analyze equilibrium states. New models designed to make use of the torrent of new molecular data focus on out-of-equilibrium processes, bringing us closer to the finer scales at which most important social and demographic processes actually occur. We are using the last great Neolithic expansion of humans into pristine environments, the colonization of Indonesia and the Pacific, as a laboratory in which to investigate the evolution of social structure and language. The Austronesian expansion has been extensively studied by prehistorians, who use genetic and linguistic markers to reconstruct the history of colonization events. We are a little more ambitious: we wish to understand not only the history of colonization events, but their underlying causes. Using both forward- and backward-in-time simulations, we have begun to test hypotheses about the evolution of social structure and language. For example, what are the effects of different types of kinship systems, or the adoption of irrigated agriculture, on population structure? And what are the respective consequences for relatedness networks of matrilineal or patrilineal descent; patriparallel or matrilateral marriage? How do networks evolve over time, and how are they affected by the physical environment? Our most interesting examples so far involve the development of water temple networks on the island of Bali.

Raima Larter Program Director of Theoretical & Computational Chemistry
rlarter@nsf.gov National Science Foundation

Prior to coming to NSF, my research involved the use of nonlinear dynamics to study chemical and physical processes in chemistry and biology which self-organize into temporal and/or spatial patterns. My group developed computational models for an enzyme-catalyzed oxidation known as the peroxidase-oxidase reaction that occurs in plants and exhibits a wide variety of interesting dynamical phenomena including bistability and both periodic and chaotic oscillations. Using the same sorts of theoretical and computational tools, we also developed models to explore the dynamical basis of the disease of epilepsy. Dynamical diseases are due to malfunctions in timing mechanisms rather than to specific anatomical or biochemical abnormalities. Examples include heart arrhythmias and certain neurological disorders including Parkinson’s disease, as well as epilepsy. The models we developed for temporal lobe epilepsy consist of systems of differential equations that describe chemical kinetic and transport processes involving key molecular components such as glutamate (a neurotransmitter) and calcium ion (a signalling molecule involved in neuronal communication). These processes, when taken together, comprise a complex timing mechanism that may, when malfunctioning, lead to seizures.

Michael Macy Department of Sociology
mwm14@cornell.edu Cornell University
http://people.cornell.edu/pages/mwm14/ 372 Uris Hall
Ithaca NY 14853

Why does our world not degenerate into the world of Mad Max? And why does it sometimes seem as if it may? Social order among interdependent agents can be imposed "from above" by a global policing mechanism or it can emerge "from below" through local interactions among adaptive agents with no centralized coordination. I assume no member of the population has the ability to identify or impose a global solution. But if everyone flies by the seat of their pants, how is social order possible? My current research explores the emergence of norms and conventions in a self-organizing social system. The problem is one that defines the human condition: the overwhelming need for norms that constrain aggressive and mutually destructive behavior is no guarantee that such rules will emerge or be obeyed by anyone except a few "suckers." On the contrary, norms can even make matters worse, by obligating people to engage in behaviors that are individually and collectively harmful. It is not hard to explain why people comply with socially undesirable norms in the face of social pressure, but why would a skeptical population enforce these norms in the first place? My research team uses computational models and laboratory experiments with human subjects to look for elementary principles of interaction in social networks that may yield clues about possible answers.
Recent studies have focused on the importance of network cascades. We found that "small world" networks that are optimal for the spread of information and disease can inhibit the spread of risky or costly collective behaviors characterized by high thresholds of adoption. A new project funded by NSF uses data from the Internet Archive to track the spread of social contagions through on-line networks.

Madhav Marathe
mmarathe@vbi.vt.edu
http://ndssl.vbi.vt.edu/

Virginia Bioinformatics Institute
Department of Computer Science
Virginia Tech
1880 Pratt Drive
Blackburg VA 24061

My interests are in the area of the design of high-performance computational methods for the representation and analysis of large socio-technical and socio-economic systems.

Specific theoretical interests include: (i) formation, evolution, and dynamics of social, information, and infrastructure networks, (ii) co-evolution of social networks, individual behavior, and institutional policies, (iii) efficient, high-performance scalable algorithms and simulations for the analysis and modeling of socio-technical systems, and (iv) complexity science related to human and social dynamics.

The application of the theoretical ideas above to build practical computer based systems are involved. Examples of ongoing work over last ten years include: (i) a highly scalable service oriented information and modeling environment to support interaction based representation and analysis of biological, information, and social systems named Simfrastructure, (ii) an ongoing project funded by the NIH MIDAS program to develop agent based models to represent and analyze the spread of infectious diseases, (iii) the Urban Infrastructure Suite (UIS) of simulations for the Department of Homeland Security as a part of the National Infrastructure Simulation and Analysis Center (NISAC), and (iv) TRANSIMS, a microscopic end-to-end modeling tool used for transportation planning in large urban areas.

Each of these tools has been successfully used in a number of customer defined case studies and exercises related to infrastructure planning, protection, and criticality analysis. They demonstrate the integration of social, behavioral, and engineering methods using simulation-based case studies to support decision making at the federal, state, and local levels; methods for assessing and mitigating vulnerabilities in large critical infrastructures; as well as successful modeling tools for analysis of socio-technical systems spanning very large urban areas. These practical socio-technical modeling tools solve real world problems that consist of approximately approx 10+ million individuals over large urban regions.

Martina Morris
morrism@u.washington.edu
http://faculty.washington.edu/morrism/

Department of Sociology
Department of Statistics
Center for Studies in Demography and Ecology
University of Washington
Box 354312
University of Washington
Seattle WA 98195

Martina Morris is a professor of Sociology and Statistics at the University of Washington. She is the Director of the Center for Studies in Demography and Ecology, the Director of the Sociobehavioral and Prevention Research core of the UW Center for AIDS research (CFAR), and the co-director of the CFAR scientific program on Mathematical Modeling. Her current research focuses on modeling the spread of HIV through partnership networks. Together with colleague Mark Handcock (Statistics, UW), she leads a large NIH-funded interdisciplinary team of researchers at the University of Washington, the UW

Network Modeling Project. The project seeks to integrate network survey design, data collection, statistical models, and dynamic simulations, to ensure a strong connection between theory, methods and data. The survey design and data collection issues are reviewed in her edited volume, Network Epidemiology (OUP, 2004). The statistical modeling and network simulation are based on exponential random graph models (ERGM), and the project has just released the first version of a statistical package, statnet, that implements these models (http://csde.washington.edu/statnet/). The package is written with an interface to the R statistical environment. It is based on an MCMC algorithm that is used for both parameter estimation and network simulation.

John Padgett
jpadgett@midway.uchicago.edu
http://home.uchicago.edu/~jpadgett

Department of Political Science
University of Chicago
5828 S. University Ave.
Chicago, Illinois 60637

For the past fifteen years, I have been constructing from primary archival sources a massive quantitative data set about social-network evolution over two hundred years, 1300-1500, in Renaissance Florence. This unprecedented data set contains information on about 50,000 persons: 10,000+ marriages, 14,000+ loans, 3,000+ business partnerships/firms, 40,000+ tax records, 12,000+ political-office elections, and so forth. Renaissance Florence was the arena for many history-altering organizational and technical inventions, in numerous domains. The project seeks to understand the genesis of many of these economic and political organizational inventions, primarily through tracing empirically and through modeling the catalytic co-evolution of multiple, cross-cutting social networks over time.

Peter Turchin
peter.turchin@uconn.edu
http://www.eeb.uconn.edu/faculty/turchin/Clio.htm

Department of Ecology and Evolutionary Biology
University of Connecticut
75 N. Eagleville Road, U-43
Storrs, CT 06269-3043

Many historical processes are dynamic. Empires rise and fall, populations and economies boom and bust, world religions spread or wither. My current research interests focus on investigating such dynamical processes in history. Because nonlinear dynamical systems are capable of very complex behaviors, explicit mathematical models are a necessary ingredient in any research program for investigating them. We also need data describing how various aspects of the studied systems change with time. Fortunately, much quantitative empirical material on historical systems has been made available over the last couple of decades by workers in the field of cliometrics, and we can confidently expect that this process will continue in the future. The proposed general approach to investigating dynamical systems in history, therefore, is as follows. We start with verbal theories explaining historical change, either proposed by previous theorists, or formulated de novo. The verbal theories are translated into mathematical models, whose predictions can then be rigorously tested with cliometric data. This general approach is more fully developed in my 2003 book Historical Dynamics.

George Verghese
verghese@mit.edu
http://csbi.mit.edu/members/Members/verghese

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Room 38-467
Cambridge, MA 02139

Dynamics on/of networks, and particularly the reflection of network graph structure in the dynamic behavior. Monitoring, inference from limited/local measurements, identification, model reduction. Influence models as tractable representations of coupled stochastic automata. Applications in power systems, synchronization, cascading failure, epidemics, biological networks.

Peyton Young
pyoung@brookings.edu
www.econ.jhu.edu/people/young

Department of Economics
Johns Hopkins University
Baltimore MD 21218-2685

My research is devoted to understanding how social and economic institutions evolve. For this purpose I have developed theoretical techniques for studying the asymptotic behavior very high-dimensional stochastic processes, which is how many economic and social phenomena must be represented. The theory has been applied to such diverse topics as residential sorting and segregation, the evolution of contractual norms, bargaining behavior, and the diffusion of innovations.

EXECUTIVE SUMMARY

Human societies are composed of complex interactions among people, government agencies and other organizations, and the natural environment. Social scientists typically confine their research to a single level of analysis – such as specializing in studying the beliefs and attitudes of individuals, the relationships among organizations, trade and conflict among nations, or evolution of cultures – and there has been relatively little progress in investigating how these processes interrelate. Physical and natural scientists also study phenomena at varying scales of time and space, but their work is held together by shared frameworks grounded in mathematics. This workshop pursued the goal of developing mathematical and computational models to help us understand the dynamics of the social world, partly through adapting approaches and insights from physics, chemistry, biology, ecology, and engineering.

Studying social interactions through mathematical and computational models allows us to better understand our world and society. Insights gained from models of social dynamics can inform government policy planning, implementation, and evaluation, especially for such complex planning tasks as preparing for and responding to natural disasters. However, developing these integrated solutions requires a highly interdisciplinary approach, as mathematical and computational techniques are combined with substantive understanding of social science questions and problems.

This workshop gathered prominent scholars from a great variety of disciplines in the physical and social sciences, providing them with an opportunity to exchange ideas, discuss issues and challenges, build potential collaborations, and gain a greater appreciation of current research efforts in other fields. Over two days, participants discussed methods and techniques that are used to understand social dynamics, theoretical problems to which they can be applied, and ways that they may be used to inform our understanding of real-world problems. They also discussed intellectual and institutional challenges of interdisciplinary research programs, and suggested ways for the National Science Foundation to improve integrative work of this kind. Some specific insights and recommendations raised at the workshop are outlined below.

Examples of research applications

- Applying mathematical and computational models to basic social dynamics such as the emergence of residential segregation, the formation and evolution of friendship and professional networks, the ecology of co-evolving networks, and the spread of opinions among friends and acquaintances
- Management and use of community resources, such as regulating the flow of highway traffic, real estate contracts, health-care resources, and Internet commerce
- Preparing for and mitigating emergencies and disasters such as epidemics, earthquakes and hurricanes
- Effects of human behavior on climate change
- Political transitions, such as the consequences of regime breakdown
- Dynamics of ethnic and religious conflict

Challenges in modeling social dynamics

- Developing models that are complex enough to accurately reflect social processes and yet simple enough to be comprehensible.
- Collecting longitudinal, relational, and multi-scale data applicable to validating models in social dynamics, and developing methods and infrastructure for gathering, analyzing, archiving, and disseminating these data

- Resolving differences in vocabulary, methods and approaches from different fields when engaging in interdisciplinary collaborations
- Promoting interdisciplinary dialog through supporting platforms for collaboration and training
- Identifying suitable areas of interdisciplinary research that will reap further advancements in the field

Solution areas and recommendations for further advancement of social dynamics

- Defining new and potentially fruitful areas for interdisciplinary research partnerships by reviewing the quality and scope of previous research collaborations in modeling social dynamics. In light of the ideas presented at this workshop, participants suggested that mapping the networks of research collaborations in modeling social dynamics may provide key insights for this evaluation.
- Maintaining support for cross-cutting areas – such as Human and Social Dynamics – that sponsor interdisciplinary collaborations on problems of social dynamics, with the support from multiple NSF directorates. Such collaborations may offer scientific breakthroughs but require sustained support over a long period.
- Establishing 'grand challenges' that lead researchers to focus on important topics in social dynamics that demand integrated insights from a variety of disciplines.
- Promoting the diffusion of ideas and methods across disciplines through regular workshops and conferences on modeling social dynamics
- Providing sustained ongoing support for the basic infrastructure in research centers that promote ties across disciplines, especially between the physical and social sciences
- Improving methods for automatic collection, archiving and dissemination of longitudinal data related to social dynamics through continued development of data management and analysis tools and cyber infrastructure
- Promoting collaborations between formal theorists and researchers who analyze empirical data on social dynamics, aiming to bridge the gap between dynamic theory and analysis of longitudinal data
- Pursuing interdisciplinary research on the dynamics of social networks, which present some of the most challenging problems and require input from numerous disciplines, yet have broad applicability across the social sciences.