

Modeling, Simulation and Analysis of Complex Networked Systems

A Program Plan

May 2009

Authors:

James M. Brase

David L. Brown

Lawrence Livermore National Laboratory

May 2009

Front and back covers: Diagrams of the Abstract Syntax Tree (AST) internal representation of a simple C++ program. The firework burst patterns are namespaces and different parts of the internal classes, functions and code within functions. The front cover shows the AST for source code, while the back cover shows a similar AST used to represent the structure of binary executables for analysis.

Source: Dan Quinlan, Center for Applied Scientific Computing, Lawrence Livermore National Laboratory

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

LLNL-TR-412733

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Executive Summary

Many complex systems of importance to the U.S. Department of Energy consist of networks of discrete components. Examples are cyber networks, such as the internet and local area networks over which nearly all DOE scientific, technical and administrative data must travel, the electric power grid, social networks whose behavior can drive energy demand, and biological networks such as genetic regulatory networks and metabolic networks. In spite of the importance of these complex networked systems to all aspects of DOE's operations, the scientific basis for understanding these systems lags seriously behind the strong foundations that exist for the "physically-based" systems usually associated with DOE research programs that focus on such areas as climate modeling, fusion energy, high-energy and nuclear physics, nano-science, combustion, and astrophysics.

DOE has a clear opportunity to develop a similarly strong scientific basis for understanding the structure and dynamics of networked systems by supporting a strong basic research program in this area. Such knowledge will provide a broad basis for, e.g., understanding and quantifying the efficacy of new security approaches for computer networks, improving the design of computer or communication networks to be more robust against failures or attacks, detecting potential catastrophic failure on the power grid and preventing or mitigating its effects, understanding how populations will respond to the availability of new energy sources or changes in energy policy, and detecting subtle vulnerabilities in large software systems to intentional attack.

This white paper outlines plans for an aggressive new research program designed to accelerate the advancement of the scientific basis for complex networked systems of importance to the DOE. It will focus principally on four research areas:

- (1) understanding network structure,
- (2) understanding network dynamics,
- (3) predictive modeling and simulation for complex networked systems, and
- (4) design, situational awareness and control of complex networks.

The program elements consist of a group of Complex Networked Systems Research Institutes (CNSRI), tightly coupled to an associated individual-investigator-based Complex Networked Systems Basic Research (CNSBR) program. The CNSRI's will be principally located at the DOE National Laboratories and are responsible for identifying research priorities, developing and maintaining a networked systems modeling and simulation software infrastructure, operating summer schools, workshops and conferences and coordinating with the CNSBR individual investigators. The CNSBR individual investigator projects will focus on specific challenges for networked systems. Relevancy of CNSBR research to DOE needs will be assured through the strong coupling provided between the CNSBR grants and the CNSRI's.

Modeling, Simulation and Analysis of Complex Networked Systems

Scientists and engineers have long used mathematical and computational models for the analysis and design of physics-based systems, such as those describing the evolution of weather and climate, the behavior of complex physical processes in combustion devices, materials design, weapons systems and energy production. These models can often be described by systems of partial differential equations, and sophisticated methodologies have been developed to translate the physics into mathematical and computational models that can be analyzed and understood to provide predictive understanding of their behavior. However, there are many complex systems of importance to the U.S. Department of Energy (DOE) that are instead described by networks of discrete components. Such systems are not naturally modeled by systems of partial differential equations, and the required mathematics is both fundamentally different from what DOE has historically pursued, and to a large extent is much less well-developed^{1,2,3}.

Examples of networks closely tied to DOE's most critical missions include the electric power grid and the information networks associated with both large-scale "open science" research projects in the Office of Science and with the National Nuclear Security Administration's Stockpile Stewardship program⁴. Understanding and developing a capability to simulate the behavior of the nation's power grid at scale is central to emerging efforts to expand and modernize its operations. The distributed, collaborative teams that perform open science research must be able to reliably and securely share significant amounts of information among institutions and over networks

around the world in order to enable and promote scientific discovery for the DOE and the nation. Assuring the safety and reliability of the nation's nuclear stockpile is fundamentally an information enterprise that depends on assuring the underlying complex information networks. In addition to cyber networks, such as the internet and local area networks, other complex

IMPROVED UNDERSTANDING OF COMPLEX NETWORKED SYSTEMS WILL HELP ANSWER THESE QUESTIONS AND INFORM POLICY

- Can we understand and quantify the efficacy of new security approaches for computer networks?
- Can we improve the design of computer or communication networks to be more robust against partial failures or intentional attacks?
- Can we detect a potential catastrophic failure on the power grid and prevent or mitigate its effects?
- Can we understand how populations will respond to the availability of new energy sources, or to changes in energy policy?
- Can we understand the behavior of large-scale software systems at a level that allows us to detect subtle vulnerabilities to attack?

networks whose improved understanding will have significant impact on federal scientific missions and public policy include social networks, biological networks, networks describing the spread of disease and even networks or graphs that describe the organization and execution of very complex scientific software.

The networks of interest to DOE are often very large, containing hundreds of millions of elements or more, and can range from relatively static in structure (e.g. the power grid), to rapidly changing and evolving (e.g. the internet, or wireless networks). Often, as a practical matter, these networks cannot be studied directly. For example, it is not possible to directly test the effect of controlled shut-down procedures for segments of the power grid, the spread of a computer virus, or the spread of a disease vector under different vaccination options without having potentially drastic economic, safety or health impacts. For infrastructure networks, replicating all or even part of the physical infrastructure may be prohibitively expensive. In both these cases, computational modeling and simulation provides a safe and cost-effective alternative that can help enormously in developing needed understanding.

Practical questions that can be answered about complex networked systems: Mathematical analysis and models of complex networks can be used to computationally explore many practical questions about the performance and behaviors of real systems:

- What are the critical structures and connections in the network? To understand a sudden change in the behavior of the electrical grid, subtle linkages between apparently independent components may be very important.
- What are the important dynamics of the network? Understanding the spreading rate of a human or computer virus requires an understanding of the dynamics of the network components but also of the connections.
- Computational models of complex network models can help to forecast behaviors. For example, in large-scale computer networks small changes in protocol timings may lead to unpredicted large-scale changes in behavior.
- Ultimately these models will form the foundation for control and design optimization of complex network systems. How should power transmission capacity be allocated to optimize real-time grid performance? Where should network flow sensors be placed to minimize bandwidth requirements for real-time activity monitoring?

DOE must continue the development of next-generation complex networked systems that are more secure, less brittle to unexpected events, and more controllable. For these emerging efforts to be successful, it is essential that a firm intellectual foundation be provided for understanding and simulating large-scale networks. Without a strong foundation, DOE will entertain a significant element of risk associated with the lack of essential knowledge for effectively improving network performance and understanding potential modes of failure. DOE has a clear opportunity to exploit its unique set of capabilities in computational and system

science at massive scale towards this unique and timely opportunity to establish a strong mathematical and scientific basis to reduce risk associated with these essential DOE systems and facilities.

Opportunities and Challenges

Many aspects of the DOE mission involve complex networked systems. The Department's most important information, whether for nuclear weapons stewardship or for high-energy physics research, flows in large-scale networks. As plans for new levels of cyberdefense move forward, fundamental understanding of network performance and effects of changes on performance will be needed. The nation's power grid is a complex network. Understanding of its performance and ability to predict the effects of both intentional modifications and unforeseen events is limited^{5,6}. Many of the most important challenges relate to the interactions of multiple networks. For example, in the modern power grid the flow of information interacts strongly with the flow of power. In addition, the interaction of evolving social networks with other network classes is critical to understanding human-generated failure modes. Development of new mathematical and computational approaches, methods, and tools will play important roles in emerging DOE programs in these and other areas.

Foundational mathematical and computational research questions:

Addressing these and other issues involving complex networked systems will require new foundational mathematical and computational research that can answer the following technical questions:

- How can the **structure** of a complex network be reconstructed from necessarily limited observations? There are often critical but unknown interactions among network elements that can create surprising or emergent^{*} behaviors. A fundamental understanding of observability and reconstructability of a network from limited and imperfect observations is needed.
- How do **dynamic processes** on the network evolve? Interactions between network elements on multiple space and time scales can lead to emergent behaviors. Which interactions are important to understanding large-scale behavior? What model representations are best for network dynamics at large scale?
- What are the limits to **prediction** of complex network behavior? Given imperfect system models and limited computational resources can well-understood bounds be placed on behaviors? How can these predictive models be validated in real systems?

* For complex, interconnected systems, exhibited behavior that would not be expected or predicted by inspection of the individual components is often referred to as "emergent".

- Can simulation methods be scaled to performance levels allowing their use in **design and control optimization** for large-scale systems? Ensembles of simulations to explore classes of behavior as design and environmental parameters of the model are varied require efficient and scalable simulation tools.

Research Program Elements

Over the past decade, a significant amount of research has been aimed at understanding complex networks and their structural characteristics^{7,8,9,10}. By analogy with the identification of invariant physical laws for physical systems, the research has focused primarily on the identification of important invariant properties including degree distribution, clustering coefficient distributions, shortest-path distribution and connected components. Understanding the behavior of real complex networks of interest to the DOE requires combining such theoretical studies with the examination of the actual networks themselves, or with accurate computational models of these networks whose structural properties can then be investigated experimentally¹¹. Generating such network models can in itself be an enormous computational task. This is because of the need to potentially represent the networks in terms of very large graphs, and because the models may require generation of extensive synthetic data to provide effective realism. In addition, these networks must be represented at different levels of fidelity. Some studies may require that the fundamental elements of the network be described in great detail, while others may need to exchange fidelity with execution speed. Lower fidelity models that run quickly can be used to support real-time operational or control capabilities for a network, or to make very large statistical studies of network behavior. A full multi-scale approach may also be needed, where some elements are modeled at very high fidelity while coupled to “reduced-order” models that provide a representation of the wider environment.

The ultimate objective of a research program on mathematical modeling, simulation and analysis of complex networked systems is to develop capabilities for understanding, designing and control of those networked systems. To do this requires addressing four major elements: understanding network structure, understanding network dynamics, predictive modeling and simulation of dynamic networks, and situational awareness and control. Some of the emerging research topics in each of these four research areas are described below.

Research area 1: Understanding network structure

To understand network behavior, the underlying topology of the network must be understood. In general the network systems of interest are of sufficient scale and complexity that direct measurement of structure will be impossible. Instead statistical inference and machine learning methods will be used to infer the network structure indirectly.

The fundamental problem is to infer the structure of a complex network from a limited set of noisy observations. The observations are often only indirectly connected to the basic network structures and interconnections. In many cases the network may be spatially distributed (e.g. the electric grid or Internet) and the scale may be very large. Some of the specific mathematical and computational challenges include:

- Machine learning on general graph structures – of particular importance are learning with relational data, learning with limited or no training data, hierarchical structure models, and learning with noisy and missing data¹²;
- Current research in complex network structures has mainly focused on graphs with a single node and link type. To extend the models to more complex systems graphs with multiple types of nodes and multiple classes of relations must be considered. In this case many of the basic approaches to network analysis, such as weighted random walks for similarity metrics, break down and new methods must be developed;
- Development of an understanding of the fundamental limits of network reconstruction will provide an important foundation for network models. Under what conditions of sampling and observation errors can network structures be reconstructed and how should confidence levels be assigned?
- Many of these networks are very large. Methods for analyzing the network structure at varying levels of resolution such as scalable methods for evaluating eigenvalue spectra of large graphs are needed¹³;
- Methods for analyzing measurements from large networks and evaluating competing classes of models and inference methods will be required. Work to scale up statistical data analysis tools on high-performance parallel computer systems to multi-terabyte or petabyte-class data sets will provide needed research tools.

COMPLEX NETWORKED SYSTEMS RESEARCH PROGRAM ELEMENTS

- Understanding network structure
- Understanding network dynamics
- Predictive modeling and simulation
- Situational awareness, design and control of complex networks

Research area 2: Understanding network dynamics

In general the network structure discussed above serves as the substrate for dynamic processes that are the principal objects of interest. The objective is to understand behavior under varying conditions – at steady state and as the network evolves, but particularly large-scale emergent behavior changes due to changes or events in the system or its environment^{14,15}.

- Graph models are used extensively to model network structure. How can these mathematical structures extend to incorporate dynamics? How can they scale as the networks grow?
- Modeling dynamic processes in complex networks will require machine learning methods incorporating multivariate time-varying statistics¹⁶. Understanding the performance of these approaches in noisy, nonstationary environments raises many open mathematical problems. The learning process itself becomes a dynamic system coupled to the system being modeled.
- Emergent behavior is one of the hallmarks of complex network dynamics¹⁷. Capturing the nonlinear interactions that produce these effects is an important component of the network model. Developing stability conditions and understanding phase transitions will be required.
- Understanding the foundations of network dynamics requires providing theoretical bounds on reconstructability of network dynamics and estimates of confidence in model structure and parameters.

Research area 3: Mathematical modeling and simulation

Given a mathematical or statistical model of a complex network, simulations can be developed to realize the model structure and produce dynamic network behaviors¹⁸. Multiple levels of fidelity are needed – at the highest resolution levels the simulations can emulate the bits and instructions of real complex dynamic networked systems while at lower levels sophisticated mathematical models will be required. Accurate and validated simulations of complex dynamic networks will be one of the most important deliverables for DOE programs focused on the applications of complex networks. Areas of research focus include:

- Multi-scale information system simulations that treat parts of the simulated system at very high fidelity through emulation or virtualization approaches while simulating the larger parts at lower fidelity. Multiple time scales will also be important for many types of networks, e.g. the Internet evolves much more quickly than the power grid;
- Highly parallel discrete-event simulations that are the foundation of these simulation tools;
- Performance optimization of large-scale simulations;
- Integration of models for human or social components with complex network simulations;
- Analysis of stability, convergence properties and errors in complex network simulations;
- Confidence measures and validation of simulations with ground truth in real network systems.

Research area 4: Situational awareness, design and control of complex networks

The ultimate goal of modeling and simulation research is to develop methods for understanding and controlling events and anomalies in complex network systems, which will depend upon progress in research areas 1-3 described above. When on-line measurements are used to build models and model predictions are, in turn, used to modify the network operation new dynamics are introduced into this new closed-loop system. The final research area focuses on the specific research required for on-line learning in complex networks and the use of control methods to modify network behavior.

- Real-time adaptive machine learning in networks expands the methods discussed above to work in a real-time on-line environment. The learning methods must adapt to a nonstationary environment and, in many cases, scale to very high speed performance.
- Distributed situational awareness systems will use multiple sensors throughout a physically distributed network to build a model of the network structure and dynamics. Distributing the inputs, the model itself, and the analysis agents that build it is the focus of this area.
- Because of the huge volumes of streaming data often involved in complex networks, the data must be processed in real time, thus requiring analysis methods that maintain minimal state information. Mapping sophisticated learning methods into a streaming environment will be a major challenge.
- Many of the problems that these models and simulations will address involve interaction of the network with some type of adversary. Game-theoretic models of the adversary integrated into a simulation environment are required to understand and model these threats¹⁹.

Program Organizational Structure

The organizational structure of the proposed DOE Office of Science program in complex networked systems is motivated by the need to rapidly develop the fundamental science required to make significant advances in our understanding of these systems. A program that can significantly accelerate the development of theory, modeling and simulation capabilities can have substantial benefits for the DOE as it grapples with challenges associated with real complex networked systems such as the power grid, the internet, and the networks on which the various DOE enterprises depend. In addition to promoting rapid development of a strong scientific foundation for understanding complex networked systems, such a program will be required to develop a substantial base of simulation codes that can be used to provide scientific understanding and discovery capabilities. This requirement strongly supports the involvement of the DOE national laboratories, since collectively they provide substantial experience in the development and maintenance of large simulation code infrastructures that support various DOE mission elements.

An effective program will have two main elements. One is the DOE Complex Networked Systems Research Institute (CNSRI) program, which will include several large research institutes whose charter is to advance theory, modeling, and software infrastructure relevant to large-scale DOE networked systems and which will become long-term focal points for broad national research efforts in complex networked systems. The second component is a Complex Networked Systems Basic Research (CNSBR) Program consisting largely of individual investigator research projects focused on specific challenges for networked systems. To assure the DOE-relevance of research projects supported by the CNSBR program, mechanisms will be provided to strongly couple those projects with the CNSRI's.

The charter of the CNSRI program element is to accelerate the development of the strong scientific foundation for complex networked systems needed to support DOE needs. As part of this charter, the CNSRI's will have the responsibility to identify priority research focus areas in complex networked systems, and will allocate resources to individual PI's who may subcontract to the CNSRI, or who are resident for some period of time at the CNSRI (e.g. academic faculty sabbaticals or summer research opportunities for faculty and students). These activities will be directed by the CNSRI steering committee, consisting of the principal investigators of the CNSRI, representatives from the wider research community, and DOE Office of Science program management.

In addition to their oversight and direction roles, the CNSRI's will be responsible for the design, development and validation of the significant software infrastructure supporting full-scale integrated system models of DOE-relevant complex networked systems. The unique value added by DOE to a research program of this type is the ability to integrate deep research components into large-scale system applications – something that is well-beyond individual investigator driven research. The critical importance to DOE of developing these large-scale system applications supports the

requirement that these Institutes be located at the DOE National Laboratories. Each CNSRI would focus on an end-to-end application area, for example, modeling and simulating the dynamic behavior of the national power distribution grid at full scale. The focus areas for the CNSRI's will demand computational, measurement, and analytic facilities at a National Lab scale.

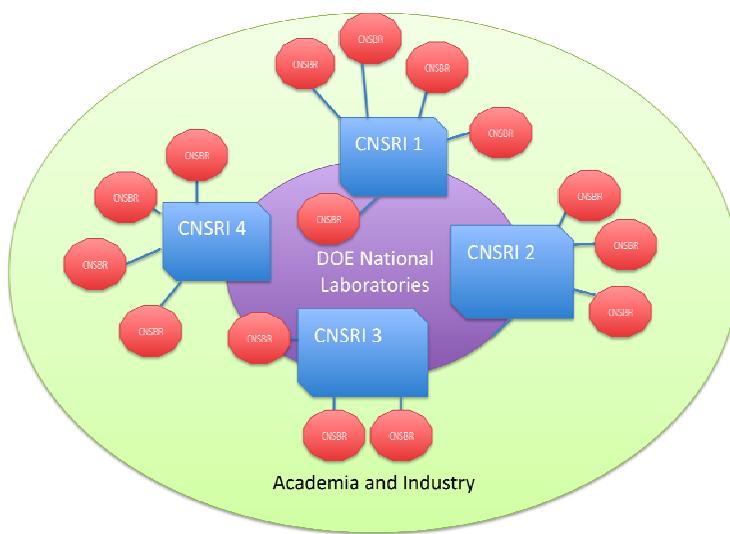


Figure 1: The DOE Complex Networked Systems program consists of two main elements, the Networked Systems Research Institute (CNSRI) program, and the Networked Systems Basic Research (CNSBR) program. The CNSRI's are DOE Lab-directed institutes that coordinate large program sub-areas, while the CNSBR's are individual investigator research efforts based largely in academia and industry. To assure DOE-relevance, each of the CNSBR's is affiliated with one or more of the CNSRI's.

The CNSRI's will also run summer research programs, workshops and conferences that will bring PI's from the CNSBR program, and other federal or international programs, to promote exchange of ideas and to encourage the development of collaborative partnerships.

Complementing the integrative function of the DOE Laboratory-based CNSRI's, the finer-grained CNSBR program is an essential component of the complex network research program. These projects will perform investigations in focused areas of complex network research. Each project in the CNSBR program will be affiliated with one of the CNSRI's whose responsibility it is to integrate the research effort into the full-scale program.

This program structure also provides natural interfaces to other DOE programs in the specific focus areas. For example a CNSRI on Information Network Simulation would provide powerful tools and capabilities to DOE and NNSA cybersecurity and networking programs.

The future of complex network theory research at DOE

DOE has a significant opportunity to develop next-generation complex networked systems that are more secure, less brittle to unexpected events, and more controllable. The establishment of a new and aggressive basic research program in modeling, simulation and analysis of complex networked systems will be an essential enabler of these developments. The resulting strong intellectual foundation will enable DOE to significantly reduce the risk associated with the lack of essential knowledge for effectively improving network performance and understanding potential modes of failure.

With sufficient federal support, a program that couples a group of Complex Networked Systems Research Institutes with a strong program of individual-investigator-based basic research projects can be developed over a very few years that will meet these needs. This program will provide a strong foundation for rapid progress in the development of understanding and simulation capabilities for large-scale complex networks. Research efforts in this program will advance our understanding of network structure, network dynamics, develop predictive modeling and simulation capabilities, and provide tools for the design and control optimization of complex networks. The development of strong foundations in these four areas will be essential for, e.g., understanding and quantifying the efficacy of new security approaches for computer networks, improving the design of computer or communication networks to be more robust against failures or attacks, detecting potential catastrophic failure on the power grid and preventing or mitigating its effects, understanding how populations will respond to the availability of new energy sources or changes in energy policy, and detecting subtle vulnerabilities in large software systems to intentional attack. Furthermore, it will provide a strong intellectual foundation that can lead to future breakthroughs in capabilities to build and operate effective, secure, robust large-scale complex networks in support of the DOE mission.

¹ Hendrickson, B. A., and M. H. Wright, "Mathematical Research Challenges in Optimization of Complex Systems," DOE Workshop Report, December 2006.

² Brown, D. L., et. al., "Applied Mathematics at the U. S. Department of Energy: Past, Present and a View to the Future", Lawrence Livermore National Laboratory Report LLNL-TR-401536, May 2008.

³ Dunlavy, D. M., Hendrickson, B., and Kolda, T. G., "Mathematical Challenges in Cybersecurity", Sandia Report SAND 2009-0805, February 2009.

⁴ Catlett, C. , "A Scientific Research and Development Approach to Cyber Security", <http://www.science.doe.gov/ascr/ProgramDocuments/Docs/CyberSecurityScienceDec2008.pdf>, December 2008.

⁵ Carreras, B. A., Lynch, V.E., Dobson I., and Newman, D. E., "Critical points and transitions in an electric power transmission model for cascading failure blackouts", Chaos **12**, 985-994, (2002).

⁶ Arionos, S., Bonpard, E., Carbone, A., and Xue, F., "Power Grid Vulnerability: A Complex Network Approach", Chaos **19**, 1, March 2009.

⁷ Amaral, L. A. N., and Ottino, J. M., "Complex networks", Eur. Phys. J. B **38**, 147-162 (2004).

⁸ Newman, M. E. J., "The structure and function of complex networks", SIAM Review **45**, 167-256 (2003).

⁹ Albert, R, and Barabasi, A., "Statistical mechanics of complex networks," Reviews of Modern Physics, **74**, 47-97,(2002).

¹⁰ Corogovtsev, A. N., Goltsev, A. V., and Mendes, J.F.F., "Critical phenomena in complex networks," Reviews of Modern Physics, **80**, 1275-1335,(2008).

¹¹ Eubank, S., Kumar, V.S.A., Marathe, M. V., and Srinivasan, A., "Structural and algorithmic aspects of massive social networks," in: Discrete Methods in Epidemiology, Abello J, Cormode, G (ed.). DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **70**, 2006:179-185.

¹² Sen, P., Namata, G., Bilgic, M., Getoor, L., Gallagher, B., and Eliassi-Rad, T., "Collective classification in network data", AI Magazine, **29**(3), 93-106 (2008).

¹³ Aguiar, M. and Bar-Yam, Y., "Spectral analysis and the dynamic response of complex networks", Phys. Rev. E **7**, 016106 (2005).

¹⁴ Strogatz, S., "Exploring complex networks", *nature* 410, **268**, 8 March 2001.

¹⁵ Motter, A., Matias, M., Kurths, J., and Ott, E., "Dynamics on Complex Networks and Applications", *Physica D* **224**, 2006.

¹⁶ Sun, J., Tsourakakis, C., Hoke, E., Faloutsos, C., and Eliassi-Rad, T., "Two Heads Better Than One: Pattern Discovery in Time-evolving Multi-aspect Data", *Data Mining and Knowledge Discovery*, **17**(1), 111-128 (2008).

¹⁷ Sole, R. and Valverde, S., "Information transfer and phase transitions in a model of Internet traffic", *Physica A* **289**, 595, 2001.

¹⁸ Perumalla, K. and Sundaragopalan, S., "High-Fidelity Modeling of Computer Network Worms", Proc 20th Annual Computer Security Applications Conference, 126, 2005.

¹⁹ Huberman, B., and Glance, N., "Evolutionary games and computer simulations", Proc. National Academy of Science USA **90**, 7716, 1993.

