Investigating Complex Network Dynamics through Coarse Graining: Data-Mining and Equation-Free Modeling

Introduction and Goals: From the dreaded phenomenon of rush-hour traffic, composed of thousands of competing vehicles on a tangled infrastructure of highways and roads, to the air we breathe and its mix of chemical species whose binding with our hemoglobin sustains life, complex systems pervade modern life. Thanks to increasing computational power, we have the ability to model the evolution of these systems in higher detail and at greater speed and resolution. Often, as in transportation on roadways and gas-phase molecular reactions, such systems are well described by complex networks whose dynamic behavior arises from interactions among the network's constituent "agents". However, while we desire insight into the long-term macroscopic dynamics of the networks, explicit, accurate equations governing macroscopic, system-level evolution are often unavailable or poorly understood. Instead, detailed information is known only at the microscopic, "agent-based" level. Prof. Ioannis Kevrekidis (under whose supervision this fellowship work will be performed), has pioneered an approach to circumvent this need for macroscopic equations in system-level analysis, known as equation-free (EF) modeling. Particularly exciting was the realization that this development could be linked with dimensionality-reduction techniques to analyze systems that lack a clear macroscopic description. By synthesizing modern data-mining algorithms and the EF method, we will be able to study the coarse dynamics of many currently intractable complex systems, including (as elaborated below) reaction- and neural- networks.

Data-mining algorithms are a key tool in extracting useful information from high-dimensional datasets (e.g. location and capacity of warehouses in a large supply chain network simulation), especially when the low-dimensional representation of the system is unknown. Sometimes the appropriate macroscopic representations are well established: in many chemical reactions, kinetics are governed by species' concentration. However, if the system is, say, a collection of coupled oscillators [1], the correct low-dimensional parameters defining long-term behavior are unclear. Our idea is to turn to numerical dimensionality reduction in these cases (e.g. principal component analysis (PCA), Laplacian eigenmaps, diffusion maps, etc.) to elucidate the significant features of our large, complex network.

Identifying the correct variables for our system-level description is valuable, but even more useful would be to examine these variables' evolution in time. In lieu of explicit equations, we harness the power of EF modeling to analyze our new, coarse system. This framework enables a range of different techniques, including simulation acceleration through coarse projective integration and macroscopic optimization via coarse conjugate gradient methods.

The goal of this project is to bring together data mining algorithms and the EF framework to anlalyze system-level dynamics of complex systems.

Research Outline: Certain dimensionality-reduction methods, such as Laplacian eigenmaps and diffusion maps have already been successfully applied to specific network problems [3]. We would like to not only apply these proven methods to a wider range of network problems, but also expand our investigation into additional algorithms, such as Isomap, nonlinear PCA, and local linear embedding. Besides evaluating their abilities to accurately coarsen a problem, computational efficiency will also be examined. A low-dimensional representation that takes a prohibitive amount of time to produce will be little help in accelerating simulations. Additionally, it has been shown that nonlinear approaches often obscures physical interpretation of the macroscopic variables, as these become nonlinear functions of the original N component system [3]. Thus, we will also search for a method that a.) scales well, such as PCA, and b.) has some discernible relation between the low- and high-dimensional system.

In the course of this project, we will address another important problem in network modeling: developing methods to quantify the similarity of two different networks. Such a metric is a

prerequisite for using many nonlinear dimensionality reducing techniques. Even when working in relatively basic biological systems where information is given in a state vector, the simple Euclidean norm cannot be applied without troublesome weighting [5]. The problem is greatly magnified when each point is itself a network, which would occur, for example, whenever a network's temporal evolution is recorded. Exciting work within the group has showed that, by comparing the spectra of (the adjacency matrices of) networks, one can define a distance measure which, when used in DMAPS, can recover the macroscopic variable governing certain network properties. Developing a more general approach would greatly aid the speed with which a network could be simulated, and will be a secondary aim of this work.

Case Study: Reaction Networks: Reaction pathways in biological systems are incredibly complex, involving various interactions between a wide variety of chemical species, all occurring on a large range of time- and length-scales. Even examining existing models can prove challenging, as the varied scales produce stiff systems of equations. Current attempts to reduce these models, such as asymptotic analysis and lumped-variable approaches [**?**], tend to make certain simplifying assumptions that restrict their applicability. Using data mining techniques, we will be able to develop a low-dimensional description that accounts for behavior across the entire system. Additionally, as the objective of such research is often to influence some sort of control on the reaction, we will apply coarse control techniques currently under development in the group to dictate reaction performance.

Case Study: Neural Network Models: It has been postulated that the rhythmic breathing of mammals is caused by periodically-firing neurons in a specific region of the brainstem known as the pre-Btzinger complex [6]. The neurons individual behavior has a strong experimental basis, while the mechanisms leading to overall dynamics remain unclear. We plan to use our variable-free, equation-free framework to study bifurcations in this system in which neurons fire in a three-, two-, and one-phase rhythm. The coarse-grained variables found by our data-mining will be of special interest. By relating these to physical statistics, we hope to gain a new understanding of the mechanisms leading to oscillatory firing. In this part of the study, there is the possibility of collaborating with Prof. Gero Miesenbck of Oxford, who has developed a method to experimentally investigate neural networks by controlling neurons with light, and Dr. Carlo Laing of Massey University, an expert in computational neuroscience.

Broader Impacts: From the Krebs cycle to Facebook, there are many potential applications for this methodology. A specific example would be modeling food distribution networks. In many countries, hunger is caused not by underproduction of food, but by poor distribution. It is estimated that about forty percent of the food in some countries goes uneaten (the U.S. and India among them) [7]. Gaining a deeper understanding of how this complex network functions would provide insight into areas for improvement. Accurate modeling of the spread of infectious disease is also an urgent task, as illnesses become ever-more antibiotic resistant. A deeper understanding of this phenomenon would inform a governments response to outbreaks, lessening the overall impact.

As engineers, we often search for simple solutions to complicated problems. This sort of thinking allows us to characterize a fluid's flow through a pipe with a single dimensionless number. We hope to empower researchers in a wide range of fields with tools to investigate their specific systems with this level of simplicity.