

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 our modern lives. Often, as in transportation on roadways and gas-phase molecular reactions, they may be characterized as complex networks whose dynamic behavior arises from interactions among the network’s constituent “agents”. Thanks to increasing computational power, scientists have the ability to model the evolution of these systems in higher detail and at greater speed and resolution. However, while researchers desire insight into the long-term macroscopic dynamics of the networks, detailed information is often known only at the microscopic, “agent-based” level. In such cases, explicit, accurate equations governing macroscopic, system-level evolution may be unavailable or poorly understood. Prof. Ioannis Kevrekidis (under whose supervision this fellowship work will be performed), has pioneered a technique to circumvent this need for macroscopic equations in system-level analysis, known as equation-free (EF) modeling. We believe that a synthesis of modern data-mining algorithms with the EF method can be used to study the long-term dynamics of many currently intractable complex systems, including supply chain networks.

Data-mining algorithms are a key tool in extracting useful information from high-dimensional datasets (e.g. positions and velocities of the molecules in a large molecular dynamics simulation), especially when the low-dimensional representation of the system is unknown. Sometimes the appropriate macroscopic representations are understood: 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. Here, we turn to numerical dimensionality reducing techniques (e.g. principal component analysis (PCA), Laplacian eigenmaps, diffusion maps, etc.) to elucidate the significant features of our high-dimensional system.

Having found the correct variables for our system-level description, we would like to examine their 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 minimization via coarse conjugate gradient methods.

The goal of this project is to adapt techniques developed by the data-mining community for use in complex system analysis, by combining them with EF modeling.

Research Outline: Certain dimensionality-reduction methods, such as Laplacian eigenmaps and diffusion maps have already been successfully applied to specific 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 either a.) embeds linearly, such as PCA, or b.) has some discernible relation between the low- and high-dimensional system. Preliminary work in this direction suggests diffusion maps may be able to overcome this barrier [4].

In the course of this project, we will, by necessity, address another important problem in network modeling: developing methods to quantify the distance between two different networks. Such a metric is a prerequisite for using many nonlinear dimensionality reducing techniques. DMAPS

EXAMPLE? 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 (***)Could expand, including adding figures if room permits). 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: Supply Chains: By applying our variable-free, equation-free framework to supply chains, we aim to use coarse minimization to identify configurations of minimum cost and maximum speed of delivery. Additionally, these tool will be used to examine potential bifurcations in system behavior if, for example, receivers suddenly raise demand. By relating the low-dimensional description to model parameters, we may obtain an understanding of the true variables influencing supply chain performance.*** (Clearly I'd need to do background reading on this)

-or-

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.