

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

**Introduction and Goals:** From the disruptive 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 rich mix of chemical species whose binding with our hemoglobin sustains life, complex systems surround us. Thanks to increasing computational power, we have the ability to model the evolution of these systems in fine detail and at high speed and resolution. Often, as in the examples above, the observed dynamic behavior may be interpreted as the result of interactions among the many individual parts, or “agents,” that compose the system, framing the problem’s structure as a complex network. However, while we seek to understand the long-term macroscopic dynamics of the networks, explicit, accurate equations governing this system-level evolution are frequently unavailable or poorly understood. Instead, detailed information is only known 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 compelling was the recent realization that this development could be linked with data-based dimensionality-reduction techniques to analyze systems that lack a clear macroscopic description. By synthesizing modern data-mining algorithms and the EF method, we aim to study the coarse dynamics of currently intractable complex systems; specifically, I hope to extend these abilities to investigate 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 chemical reactions, kinetics are often governed by species’ concentrations. However, in many systems, such as a collection of coupled oscillators [6], the correct low-dimensional descriptors defining long-term behavior are unclear. We intend to turn to numerical dimensionality reduction in these cases (e.g. principal component analysis (PCA), Laplacian eigenmaps, diffusion maps (DMAPS), etc.) to elucidate the significant features of our large, complex networks.

Identifying the correct variables for our system-level description is valuable, but even more illuminating would be to examine these variables’ evolution in time. In lieu of explicit equations, we plan to harness the power of EF modeling to analyze the macroscopic ramifications of the fine-scale system. This framework enables a range of different techniques, including simulation acceleration through coarse projective integration and systems-level optimization and design.

*The ambition of this project is to bring together data mining algorithms and the EF framework to analyze macroscopic dynamics of complex systems, and, in particular, of complex networks.*

**Research Outline:** Certain dimensionality-reduction methods, such as Laplacian eigenmaps and diffusion maps have already been successfully applied to specific network problems. 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. If a low-dimensional representation takes a prohibitive amount of time to produce, it will be of little help in accelerating simulations. Additionally, nonlinear data-mining approaches often obscure physical interpretation of the macroscopic variables, as they become nonlinear functions of the original many-bodied system [3]. Thus, we will also search for a method that: (a) scales well with system size, such as PCA, and (b) provides a physically meaningful coarse description.

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 dimensionality reduction techniques. Even when analyzing relatively basic biological systems, in which information is given in a state vector, the simple Euclidean distance may be quantitatively misleading [1]. This problem of constructing metrics on input data is even more challenging when each data point is itself a network, which would occur, for example, whenever a network’s temporal evolution is recorded. Promising work within the group has showed that, by comparing the spectra of different networks (i.e. the spectra of their adjacency matrices), one can define a distance measure which, when used in DMAPS [2], can recover the macroscopic variables governing certain network properties. Developing a more general approach would significantly accelerate 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 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 [7]. Using data mining techniques, we hope 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 chemical kinetics, we will apply coarse control techniques currently under development in the group to regulate reaction performance.

**Case Study: Neural Networks:** 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-Bötzinger complex. The neurons' individual behavior has been elucidated experimentally, but the mechanisms leading to overall dynamics remain unclear. We plan to use our extended equation-free framework to study bifurcations in this system in which neurons fire in a one-, two-, and three-phase rhythm. The coarse-grained variables found by our data-mining will be of particular interest. By relating these to physical statistics, we hope to gain new insight into the origin of collective oscillatory firing. This may lead to collaborations with Prof. Gero Miesenböck of Oxford, who developed optogenetics, a method to experimentally investigate neural networks by controlling neurons with light, and Dr. Carlo Laing of Massey University, an expert in computational neuroscience [5].

**Broader Impacts:** From the Krebs cycle to Facebook, there are many potential applications for the framework we propose to develop. One specific example is in 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, including the U.S. and India, goes uneaten [4]. Gaining a deeper understanding of how this complex network functions would provide insight into areas for improvement. Accurate modeling of the spread of infectious diseases is also an urgent task, as illnesses become evermore resistant to antibiotics. A deeper understanding of this phenomenon would inform a government's response to outbreaks, lessening their overall impact.

Whether it's characterizing a fluid's flow with a single number or detailing enzyme kinetics with a pair of differential equations, as engineers, we have a history of finding simple solutions to complicated problems. It is this mindset that drives my ambition to empower researchers in a wide range of fields with tools to investigate their specific systems at a practical level of simplicity.

## References

- [1] K. A. Bold, Y. Zou, I. G. Kevrekidis, and M. A. Henson. An equation-free approach to analyzing heterogeneous cell population dynamics. *Journal of mathematical biology*, 55(3):331–52, Sept. 2007.
- [2] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, July 2006.
- [3] A. L. Ferguson, A. Z. Panagiotopoulos, I. G. Kevrekidis, and P. G. Debenedetti. Nonlinear dimensionality reduction in molecular simulation: The diffusion map approach. *Chemical Physics Letters*, 509(1-3):1–11, June 2011.
- [4] D. Gunders. Wasted : How America Is Losing Up to 40 Percent of Its Food from Farm to Fork to Landfill Author. *NRDC*, 2012.
- [5] G. Miesenböck. The optogenetic catechism. *Science (New York, N.Y.)*, 326(5951):395–9, Oct. 2009.
- [6] K. Rajendran and I. G. Kevrekidis. Coarse graining the dynamics of heterogeneous oscillators in networks with spectral gaps. *Physical Review E*, 84(3):036708, Sept. 2011.
- [7] N. Vora and P. Daoutidis. Nonlinear model reduction of chemical reaction systems. *AIChE Journal*, 47(10):2320–2332, Oct. 2001.