From worms to wars, modeling and controlling networked dynamical systems

Megan Morrison

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

J. Nathan Kutz, Chair

Michael Gabbay

Eric Shea-Brown

Program Authorized to Offer Degree:
Applied Mathematics

University of Washington

Abstract

From worms to wars, modeling and controlling networked dynamical systems

Megan Morrison

Chair of the Supervisory Committee:
Robert Bolles and Yasuko Endo Professor J. Nathan Kutz
Applied Mathematics

Networks in nature regularly exhibit dynamics that are difficult to characterize due to their nonlinear nature and use of obscure control signals. These systems are often marked by low-dimensional dynamics, multiple stable fixed points or attractors, and outputs that are generated from nonlocalized network activity. We develop procedures for characterizing nonlinear dynamics in networks with transparent, low-dimensional models and controlling them using bifurcation theory. We apply these techniques to the neural network of *C. elegans*, Hopfield networks, and randomly generated high-dimensional dynamical systems. We show that nonlinear control may be a method by which *C. elegans* regulates its behavior and could be a viable control method in other systems with multiple stable fixed points.

Although much focus rests on the dynamics of nodes in a network, many networks of interest, such as sociopolitical networks, possess edge dynamics in addition to node dynamics. One such network is the international system. We explore dimension reduction techniques and the governing equations for network edge dynamics in addition to node dynamics. We show how final stable states can be predicted from initial network statistics for random matrices under the influence of structural balance dynamics; this analysis is useful for understanding when assortativity in a network, which can occur due to in-group biases, will determine the factionalization that occurs in networks under structural balance dynamics. We further build a matrix dynamical systems model of sociopolitical edge dynamics with low-

conflict and high-conflict stable states. We analyze the edge dynamics in a low-dimensional eigenvalue/eigenvector space and derive bifurcations for state transitions in the eigenvalue space; this is similar to our derivation of state transitions for node dynamics.

Used together, data-driven discovery, dimension reduction, and bifurcation theory can be used to effectively describe, analyze, and control network dynamics. Data-driven techniques allow us to identify the dynamics governing activity in complex networks. Dimension reduction allows us to characterize high-dimensional dynamics with far fewer variables. Bifurcation theory allows us to understand how and why qualitative transitions occur in nonlinear systems. We demonstrate these techniques on several systems including toy models, random networks, the nematode *C. elegans* neural network, and the European international system. We hope that these strategies for building models for network dynamics and evaluating control techniques can be useful in a wider range of networks with nonlinear dynamics.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would like to express my gratitude to Nathan Kutz and Michael Gabbay for advising me during graduate school and helping me grow as a researcher. Nathan has been an incredibly encouraging and supportive advisor through the ups and downs of graduate school and his enthusiasm for research is infectious. Nathan has so many big ideas and is great at getting people excited about the power of applied math. I've learned so much from him and hope that I will be able to be the sort of mentor that he has been to me during this time.

Michael Gabbay has also been very supportive and helpful, thoughtfully going through my work and helping me hash out ideas and improve the way I present findings. He has given me so much great feedback that has improved the way I pursue research ideas and write papers. I have really enjoyed working on applied math applications to social network science with him.

I would also like to thank my other committee members, Eric Shea-Brown and Mehran Mesbahi for their perspectives and helpful comments. Eric Shea-Brown has been a great professor and leader in the mathematical neuroscience community. I have benefited so much from my participating in theoretical neuroscience events at UW.

I am also indebted to the other graduate students and postdocs who helped and mentored me through undergrad and graduate school. I'd like to thank Pedro Maia, Charlie Fieseler, and Charles Delahunt for helping me get started with my initial research projects and for being great people to talk to about mathematical neuroscience ideas. I'd like to thank Kelsey Marcinko for all her help and support and for being a great housemate, classmate, and friend. I'd also like to thank other friends for all their encouragement, in particular Sarah, Janeé, Brooke, Kayley, Sritam, and Cathryn.

Lastly, I would like to thank my wonderful family. I could not have done any of this without the support of my mom and dad and my siblings Elisa, Jenna, and Liam. I am so thankful for them.

DEDICATION

to my mom, who taught me how to do math

Chapter 1

INTRODUCTION

Networks are useful for describing many natural systems in biology, neuroscience, and the social sciences [159]. From the metabolic networks operating at the cellular level, to the neural networks that comprise the brain, to the social networks that determine geopolitical outcomes, natural networks are at work in many areas of our lives. The governing equations and actions required to control dynamics on networks not engineered by humans can be difficult to derive [23, 115]. In this dissertation I discuss methods for characterizing and controlling the dynamics of nodes and edges in networks where the dynamics are nonlinear and possess multiple stable states. The applications that I consider are the neural dynamics and behavioral outputs of *C. elegans*, edge dynamics in the international system and other sociopolitical networks, the dynamics of artificial memory networks, and the dynamics of random networks.

Our first objective is to characterize network dynamical systems with transparent, low dimensional models using data-driven techniques. Many systems produce dynamics on a low-dimensional manifold and we can take advantage of this to describe the dynamics of complex, high dimensional systems with only a few variables [25, 168, 2, 116, 24]. Our second objective is to construct nonlinear control regimes for such systems that take advantage of the systems intrinsic dynamics to shift it between the stable states that exist in the system.

We focus on networks that have independent node ($\mathbf{x}$) and edge ($\mathbf{X}$) dynamics,

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}) + Bu_1(t)$$
$$\frac{d\mathbf{X}}{dt} = g(\mathbf{X}) + \mathbf{C}u_2(t)$$

where $\mathbf{x} \in \mathbb{R}^n$ are the nodes for a network of size n, with nonlinear dynamics $f(\mathbf{x}) \in \mathbb{R}^n$

and control signal $u_1(t)$ applied to the nodes through weights $B \in \mathbb{R}^n$. Edges in a network $\mathbf{X} \in \mathbb{R}^{n \times n}$ can also have nonlinear dynamics $g(\mathbf{X}) \in \mathbb{R}^{n \times n}$ with control signal $u_2(t)$ applied to the system through edge weights $\mathbf{C} \in \mathbb{R}^{n \times n}$. A network may have node dynamics, edge dynamics, or both. We show how the neural network in *C. elegans*, a Hopfield model, and random high dimensional dynamical systems can all be characterized as a nonlinear dynamical system of nodes that can be controlled to shift between fixed points in the system. In other contexts it is the edge states rather than the node states that are of primary concern. Social systems under the influence of structural balance dynamics, such as the international system and militant networks, can be characterized as a nonlinear dynamical system of edges where clusters in the network and conflict levels are a function of edge states. We use matrix dynamical systems and spectral analysis to characterize the dynamics of relationships in social systems and understand how initial network structure influences a network's long-term state. Understanding how structure impacts dynamics is important for assessing stability in sociopolitical networks and risk for system-induced conflict escalation.

In Chapter 2, we observe how nonlinear control could be used by an organism to control its neural dynamics and resulting behavior by constructing a low-dimensional nonlinear control model for the neural dynamics of *C. elegans*. Recent whole-brain calcium imaging recordings of this nematode have demonstrated that the neural activity associated with behavior is dominated by dynamics on a low-dimensional manifold that can be clustered according to behavioral states. Previous models of *C. elegans* dynamics have either been linear models, which cannot support the existence of multiple fixed points in the system, or Markov-switching models, which do not describe how control signals in *C. elegans* neural dynamics can produce switches between stable states. It remains unclear how a network of neurons can produce fast and slow timescale dynamics that control transitions between stable states in a single model. We propose a global, *nonlinear* control model which is minimally parameterized and captures the state transitions described by Markov-switching models with a single dynamical system. The model is fit by reproducing the timeseries of the dominant PCA mode in the calcium imaging data. Long and short time-scale changes in transition

statistics can be characterized via changes in a single parameter in the control model. Some of these macro-scale transitions have experimental correlates to single neuro-modulators that seem to act as biological controls, allowing this model to generate testable hypotheses about the effect of these neuro-modulators on the global dynamics. The theory provides an elegant characterization of control in the neuron population dynamics in *C. elegans*. Moreover, the mathematical structure of the nonlinear control framework provides a paradigm that can be generalized to more complex systems with an arbitrary number of behavioral states.

In Chapter 3, we study how feed-forward control could be used in networks with multiple stable states to induce transitions between stable states. Our method integrates dimensionality reduction, bifurcation theory, and emerging model discovery tools to find low-dimensional subspaces where feed-forward control can be used to manipulate a system to a desired outcome. The method leverages the fact that many high-dimensional networked systems have many fixed points, allowing for the computation of control signals that will move the system between any pair of fixed points. The *sparse identification of nonlinear dynamics* (SINDy) algorithm is used to fit a nonlinear dynamical system to the evolution on the dominant, low-rank subspace. This then allows us to use bifurcation theory to find collections of constant control signals that will produce the desired objective path for a prescribed outcome. Specifically, we can destabilize a given fixed point while making the target fixed point an attractor. The discovered control signals can be easily projected back to the original high-dimensional state and control space. We illustrate our nonlinear control procedure on established bistable, low-dimensional biological systems, showing how control signals are found that generate switches between the fixed points. We then demonstrate our control procedure for high-dimensional systems on random high-dimensional networks and Hopfield memory networks.

Transitioning to sociopolitical networks and the dynamics of network edges, in Chapter 4 we investigate how initial network structure impacts factions formed in networks under structural balance dynamics. The spectrum of the adjacency matrix generated by a stochastic block model with two equal size communities shows detectability transitions in which the community structure becomes manifest when its signal eigenvalue appears outside the main spectral

band. The spectrum also exhibits "sociality" transitions involving the homogeneous structure representing the average tie value. We derive expressions for the eigenvalues associated with the community and homogeneous structure as well as the transition boundaries, all in good agreement with numerical results. Using the stochastically-generated networks as initial conditions for a simple model of structural balance dynamics yields three outcome regimes: two hostile factions that correspond with the initial communities, two hostile factions uncorrelated with those communities, and a single harmonious faction of all nodes. The detectability transition predicts the boundary between the assortative and mixed two-faction states and the sociality transition predicts that between the mixed and harmonious states. Our results may yield insight into the dynamics of cooperation and conflict among actors with distinct social identities.

In Chapter 5 we built upon the matrix dynamical system model of structural balance to include other forces that impact sociopolitical networks and use our model to demonstrate the conditions under which social networks can become destabilized with conflict escalation. Network structure can contribute to the emergence of war in the international system, yet the way in which structure affects dynamics is not well understood. Bifurcation and critical transition theory has been used in other network settings to identify the conditions under which catastrophic transitions will occur in a system such as the destabilization of an ecosystem. While statistical models can find correlations between network features and the emergence of war, dynamical systems models can elucidate why certain network features put a system at risk for destabilization. We model the dynamics that occur in sociopolitical network edges with a bistable dynamical system that contains a low-conflict state and a high-conflict state that corresponds to systemic war. Network dynamics emerge from the interaction between dyadic relationships and structural balance forces. Spectral analysis provides a low-dimensional space in which to analyze the edge dynamics and derive critical transitions to the system. We observe hysteresis in the dynamics and identify structural features that make war more likely, such as dense connectivity, many influential nodes, and community structure. We identify destabilizing ties in the network of great powers leading up to WWI

and connect theories of destabilization in the political science literature to parameters and features in our dynamical systems model. Models for sociopolitical network edge dynamics may be useful in identifying destabilizing edges and features in a network as well as assessing the risk of war.

Finding the dynamics of networked systems in a low-dimensional space allows us to better understand and control them. In this dissertation, we use PCA and eigenvector decomposition to represent high-dimensional dynamics in a low-dimensional subspace and then derive bifurcations for the nonlinear dynamics in the low-dimensional subspace. We show how we can fit noisy high-dimensional network data to a nonlinear dynamical system with control and outline a control regime for similar systems. We highlight how the eigenvalues and eigenvectors of a signed adjacency matrix are the low-dimensional variables that drive edge dynamics in social networks and analyze how community structure impacts the dominant eigenvalues and eigenvectors of a connectivity matrix which in turn determines the stable state to which the network will converge. We hope that this work can be used to better understand complex network dynamics and construct novel control strategies.

Chapter 2

NONLINEAR CONTROL IN THE NEMATODE *C. ELEGANS*

2.1 Introduction

The emergence of large scale neural recordings across model organisms is revolutionizing the potential for the theoretical modeling of how neuron population dynamics is accomplished. With the recent advancements in whole brain imaging technologies for the nematode *C. elegans* [182, 165, 160], the relationship between neural activity and behavioral outcomes can be studied in a holistic fashion. More precisely, *C. elegans* provides a unique opportunity to quantify neuron population dynamics as it has only 302 neurons whose stereotyped electrophysical connectivity map (connectome) is known from serial section electron microscopy [217, 38]. We show that the neuron population dynamics of the *C. elegans* nematode can be characterized by a global nonlinear control model which matches experimental measurements. Moreover, it provides a general mathematical framework that illustrates how nonlinearity can be exploited to produce a global model of neuron population dynamics and how it can be readily applied to more complex model organisms.

Data from *C. elegans* neural recordings show that high-dimensional neuronal activity produces dominant, low-dimensional patterns of activity across the connectome, with interpretable clusters [102, 176, 130, 116, 114, 61]. Previous analysis of behavioral and calcium imaging data can be categorized within three different modeling paradigms, each with their own strengths and weaknesses: Markov models, switching linear dynamical systems, and models with control. An overview is given in Table 2.1.

The most well-established methodology is the Hidden Markov Model (HMM) which has been used for decades [176]. This paradigm simplifies data into clusters, and assumes instantaneous transitions between them. Such models have been used to uncover different

macro-level behaviors in animals that are characterized by different stabilities of individual behaviors, such as the difference between roaming and foraging [6]. While these models capture the differential stability of various behaviors, they are statistical models and do not show how the dynamics in the network generate these transitions.

Recent work extends the HMM paradigm to trajectories in neuron space [128, 127, 129, 44]. This paradigm models the neural network as having distinct linear dynamics within different states, allowing a connection between behavior-level HMMs and neural trajectories at the cost of many more parameters. Mathematically, this is given by $\dot{\mathbf{x}} = \mathbf{A_i x}$, where $\mathbf{x}$ is the state space, the dot represents time differentiation, and i refers to multiple segmented state spaces. However, these models are fundamentally local, and it is unclear whether switching between different states can be biologically achieved.

A recent paradigm has built a global model for calcium imaging dynamics by including a control signal [62]. This is given by the equations $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$ where $\mathbf{x}$ is the state space, the dot represents time differentiation and $\mathbf{u}$ is the control signal. The global matrices $\mathbf{A}$ and $\mathbf{B}$ characterize the intrinsic dynamics, and how actuation forces these dynamics respectively. This paradigm allows the control signals to be studied as independent objects, hypothesizing a separation between the intrinsic dynamics of the network and the mechanisms that cause transitions. However, this work uses a linear framework, which requires that there is only a single fixed point at the origin. From this perspective, all behaviors except one are merely long-lived and do not have their own fixed point.

In contrast to linear models which can only support a single fixed point in the dynamics, nonlinear models offer a more flexible architecture for control, especially in systems like the *C. elegans* where multiple behavioral states appear to be stable. We show that with minimal parametrization, we can construct a global nonlinear model of the underlying *C. elegans* control structure. Our nonlinear control model removes the need for multiple linear models and provides a parsimonious, global control framework parameterized by only a few parameters and consistent with experimental observations. Nonlinear control theory takes the form $\dot{\mathbf{x}} = f(\mathbf{x}) + g(\mathbf{u})$ where $f(\cdot)$ specifies the nonlinear dynamics and $g(\cdot)$ specifies the actuation

on the underlying dynamics. This provides a theoretical framework for circumventing many of the standard limitations inherited from linear control theory. This comes at the expense of provable controllability criteria which can be rigorously stated in linear theory. A fundamental benefit of nonlinear control theory is that one can posit an underlying model with *multiple fixed points* where $f(\mathbf{x}_j) = 0$ and $j = 1, 2, \cdots, N$. In the context of neuron population dynamics and *C. elegans*, these N fixed points correspond to distinct behavioral states, i.e. forward or backward motion. Thus instead of regressing to the matrices $\mathbf{A}$ and $\mathbf{B}$ in constructing a linear model, we instead posit a global model whose features are consistent with experimental observations [102]. We propose a model of the form

$$\boldsymbol{x}' = F(\boldsymbol{x}, \beta) + u(t) \tag{2.1}$$

where $F(\boldsymbol{x}, \beta)$ represents the intrinsic nonlinear dynamics containing multiple stable states; β parameterizes fluctuations in intrinsic dynamics that may occur over long timescales. Fast-timescale control signals $u(t)$ control state location by applying feed-forward control to the intrinsic dynamics.

Nonlinear control has been used to induce and describe transitions between stable attractors in the nonlinear dynamics of other biological networks [170]. In synthetic biology, researchers have created nonlinear, bistable gene regulatory networks in *Escherichia coli* that can be toggled between different states with the use of control signals [72]. Control applied to key nodes can induce a nonlinear system to converge to a desired state rather than an undesired state [43]. Stochasticity is also a mechanism for control and is used by organisms to regulate transcription [105]. Previous work has considered control in bistable systems implemented via feed-forward control pulses [191, 192, 190] and through analysis of saddle points in the system [209]. We hypothesize that the neural network of *C. elegans* uses nonlinear control mechanisms, such as those previously explored, to transition between various stable states and vary its transition probabilities. Feed-forward control signals could take the form of activity in dedicated control neurons, such as sensory neurons, but they may have a distributed representation. We aim to represent the local and non-local neural activity that, holistically,

implements effective control in the nematode, as a single, low-dimensional time series $u(t)$.

Our model has the flexibility to describe *C. elegans* dynamics under a wide variety of internal states and environmental stimulus. Quantitative work on postural analysis of the behaving *C. elegans* has demonstrated there is low-dimensional structure on the level of individual movements and body bends [194, 193]. The statistics of how often these movements happen show the presence of a few discrete clusters [164, 6, 212, 41], or a spectrum of behavioral strategies [69, 88] that are appropriate in different environments and may even be different between individuals [151]. Recent modeling work has used a conceptual or data-driven model of multiple fixed points in the neuron population phase space [39, 176]. However, it remains unclear how statistics of transitions between behaviors can be controlled by global parameters, or how individual trajectories through state space are affected in these cases. Our model is able to reproduce the changes in statistics between the large-scale roaming and dwelling behaviors via changing a single global parameter. In addition, this model reproduces observed short time-scale bursts of reversals interspersed with extremely short-lived forward states.

Our model further could be used to produce testable hypotheses of the effects of neuromodulators on global dynamics. Much work has been done in recent years to extend the understanding of internal *C. elegans* dynamics beyond simple synaptic connections to include additional layers, particularly the slower dynamics of neuromodulators [17, 108]. Specifically, single molecules and simple neuronal circuits have been found to change global statistics related to fundamental behaviors, most clearly the frequency of reversal initiation [65, 6, 18, 126, 141, 41, 88, 212]. Because our model is able to reproduce macro-scale behavioral changes with a single parameter, we hypothesize that there may be a correspondence between some neuromodulators and our model parameters. As we will show, our global nonlinear model is minimally parameterized and provides a parsimonious representation of the neuron population dynamics of the *C. elegans* nematode. These parameters have suggestive connections to experimental work, and some may correspond to one or more neuromodulators. This mathematical framework is general, and can be readily applied to more complex model organisms.

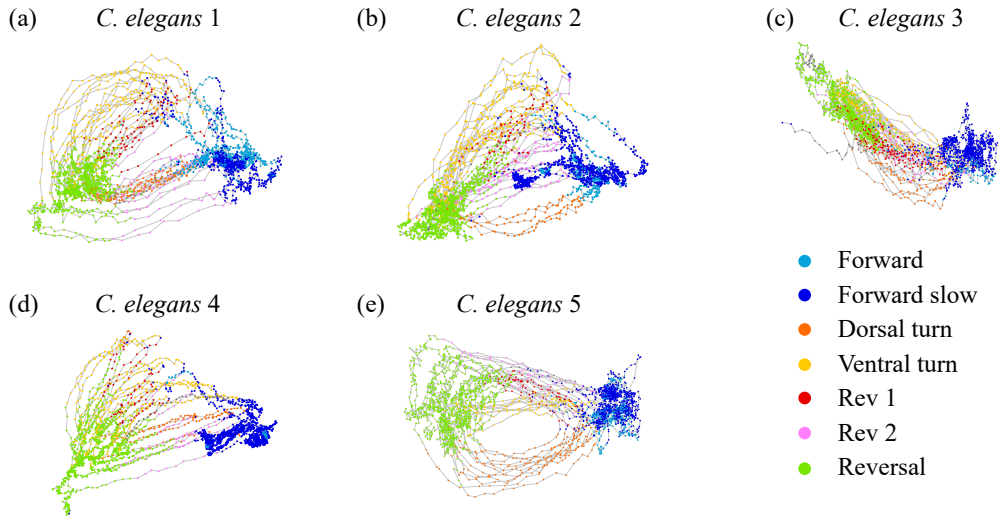

Figure 2.1: *C. elegans* neural activity in the PCA space of the first two modes. Trajectories colored by behavioral state.

2.2 Results

We introduce a nonlinear global model with control for the low dimensional activity of *C. elegans* neuron population dynamics that captures the behavioral dynamics that we aim to model. Any model of this data must satisfy the following requirements: (1) the general structure of the model must support the two fixed points observed in the data, (2) the model must be flexible enough to accommodate the full range of variability observed in *C. elegans*, and (3) the model must be minimally parameterized such that the modulation of only a few parameters can generate this full range of variability. We start by observing the structure of the data and posit a general model whose parameters can be tuned to generate activity that is analogous to the activity observed in the data. We then explore how experimentally observed changes in *C. elegans* behavior can be explained by the modulation of single parameters.

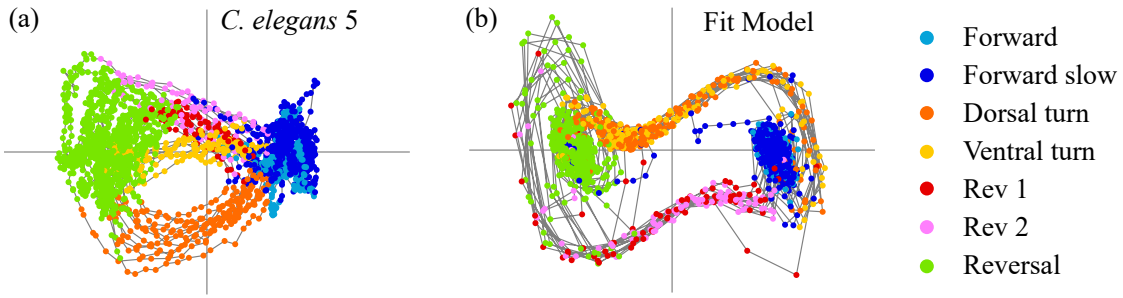

Figure 2.2: (a) PCA activity of *C. elegans* 5. (b) Dynamical systems control model fit to PCA activity of *C. elegans* 5.

2.2.1 Nonlinear global dynamical models for C. elegans

We construct a generalized, low-dimensional representation of the neural activity of *C. elegans* by performing PCA on the activity of neurons in five *C. elegans* from [102]. We then use the first two PCA modes to represent the dynamics linked to behavior (See Methods). Distinct behaviors correspond to different regions of PCA space (Fig 2.1). Forward and reversal behaviors (states 1, 2, and 7) correspond to two distinct stable states in PCA space. Dorsal and ventral turn behaviors (states 3 and 4) correspond to reversal to forward transitions while rev 1 and rev 2 behaviors (state 5 and 6) correspond to forward to reversal transitions. Behavioral states during calcium imaging are determined by [102] and [186]. We fit our general dynamical systems model to the trajectories of the dominant mode $v_1(t)$ which most strongly differentiates the stable states from the transition states.

This low-dimensional representation suggests a feature space for a model decomposition. Specifically, it allows us to build a control model which accurately reproduces the global dynamics with minimal parametrization. The nonlinear parsimonious and global control model takes the form

$$\begin{aligned} x' &= y \\ y' &= f(x, \beta) + \gamma y + u(t), \end{aligned} \tag{2.2}$$

where the nonlinear dynamics is prescribed by the cubic

$$f(x, \beta) = -(x + 1)(x - \beta)(x - 1), \tag{2.3}$$

which has by construction (for $u = 0$) two stable fixed points at $x = \pm 1$ and a single unstable fixed point whose location is determined by the parameter β. Additionally, there is a damping parameter γ and a control input $u(t)$. These relate to the dominant PCA modes directly, where $x(t) = v_1(t)$ and $y(t) = v_2(t)$. Due to the stochastic nature of the observed data, we additionally add stochastic terms and arrive at the system

$$\begin{aligned} dx_t &= y_t dt + \sigma dW_t \\ dy_t &= -(x_t + 1)(x_t - \beta)(x_t - 1)dt + \gamma y_t dt + u(t)dt + \sigma dW_t, \end{aligned} \tag{2.4}$$

where β and γ parameterize the cubic dynamical system, and σ and dW_t characterize the Brownian motion which models the noisy fluctuations observed in experiments. We chose a two dimensional model fit to the first two *C. elegans* PCA modes as this is the minimum number of modes that captures the stable state clusters as well as the variability in the transition trajectories. While a higher dimensional model would capture more of the variance in the neural activity, and a model with more parameters would increase the model fit, we prioritize minimal parameterization. Our objective is to create the lowest dimension model with the fewest number of parameters that is able to represent three features of the *C. elegans* neural activity: (1) the intrinsic stability of the neural activity underlying the forward and reversal behaviors, (2) the variability in transition trajectories, and (3) the destabilization of the stable states under the influence of feed-forward control signals, that is, control of the network's state. A nonlinear control model that is higher-dimensional, or that has more parameters, can be found using the methods outlined in [149].

We fit the model parameters to the low-dimensional *C. elgans* activity by minimizing the error between the dominant PCA trajectory $v_1(t)$ and the trajectory of the corresponding model variable $x(t)$ (see Methods). The labelled, behavioral states timeseries determines

when the dynamical system is in the uncontrolled state $u(t) = 0$ for states 1, 2 and 7 (forward, forward slow, and reversal) or a transition state, $u(t) = u_{3,4}$ (dorsal and ventral turns) or $u(t) = u_{5,6}$ (rev 1 and rev 2 transitions). The dynamical system fitted to *C. elegans* 5 is

$$\begin{aligned} x' &= y + 0.06dW_t \\ y' &= -(x_t + 1)(x_t - 0.11)(x_t - 1)dt - 1.51y_t dt + 0.06dW_t + u(t)dt, \end{aligned} \tag{2.5}$$

with control signal strengths $u_{3,4} = 0.54$ and $u_{5,6} = -0.77$ and timescaling parameter $dt = 0.29$. Figure 2.2 shows a comparison of the *C. elegans* neural activity in PCA space (a) with the dynamical systems model reproducing this activity colored by behavioral state (b). Model parameters for all five *C. elegans* are shown in Table 2.2 and a comparison of the data, model, and errors is shown in Figure 2.3. Across all models, $\beta \approx 0.1$ indicates that the model's reversal state ($x = -1$) is more stable than the forward state ($x = 1$). Because the rev 1 and rev 2 transitions are shorter than the dorsal and ventral turn transitions, stronger control signals are necessary to complete the forward to reversal transition $|u_{5,6}| > |u_{3,4}|$. A weaker forward state stability ($\beta \approx 0.1$) also aids in the forward to reversal transitions.

Figure 2.3 shows the timeseries of the dominant PCA mode ($v_1(t)$) compared with the corresponding model variable ($x(t)$) for each model fit. The error over time ($E(t)$) shows that $E(t) \approx 0$ during the stable states 1, 2, and 7, indicating that the model executes most transitions between the forward and reversal states successfully. The error spikes during state transitions, $|E(t)| > 0$, indicating that the model does not capture the shape of the transitions accurately due to the model's minimal parameterization.

2.2.2 Changes to a single parameter reproduce different long-timescale behaviors of C. elegans

As shown in the methods section, this global model has three fixed points whose stability is determined by the parameter $\beta \in (-1, 1)$. The parameter γ determines the linear growth/decay rate of each fixed point. The parameter σ controls the amount of stochasticity in the system. Figure 2.4(a)-(c) shows the behavior of Eq. (2.4) as a function of β for randomly

generated control signals. For $\beta = 0$, there is a symmetry between the two stable fixed states corresponding to the forward and reversal state, which reproduces the long time-scale distribution of behaviors across individuals. As β approaches unity, the dynamics are skewed in favor of one of the fixed points.

The statistics of reversal length and frequency change drastically across multiple timescales during the life of a *C. elegans*. Our nonlinear control model is able to reproduce three very distinct changes in state distribution and switching frequencies seen in these experimental studies via modulation of a single parameters. The first well-studied change in these dynamics is the switch between dwelling and roaming states [65, 164, 6, 212]. Specifically, the frequency of reversals is much lower in the roaming state, which facilitates the exploration of a larger geographical area. Several neuromodulators [65] and individual neurons [212] have been implicated in this behavioral change; some function of these chemicals or neuron activity levels might directly correspond to the model's control signal onsets and strengths $u_{3,4}$ and $u_{5,6}$.

2.2.3 Distinct behaviors may be controlled by a shared mechanism

Two additional behaviors that are not known to be related can be explained using the same mechanism: spontaneous reversal bouts, and an increase in reversals in an aversive oxygen environment. We use calcium imaging data from [186] to create distributions for *C. elegans* low-dimensional activity during a reversal bout and when in different controlled oxygen states (Fig. 2.4). The reversal bout behaviors, observed in immobilized animals and shown in Figure 2.4(d)-(e), are long-lived behaviors that begin in a reversal state, move into a forward motion state but then fail, and return to a reversal state several times in succession. This can be clearly related to a change in the parameter β, which controls the stability of the fixed points corresponding to forward and backward motion. A known method for experimentally destabilizing the forward state in *C. elegans* is through a modification of their environment. In an environment with a preferred oxygen level of 10%, *C. elegans* tend to have stable forward swimming behavior, Figure 2.4(f)-(h). When the oxygen in their environment is

increases to 21%, they exhibit more transient forward swimming behavior, Figure 2.4(i)-(k), similar to the observed "reversal bouts".

Increasing β, as shown in Figure 2.4(b)-(c), reproduces this unstable forward behavior by retaining the stochastic control signals that would normally transition the system to a forward motion state, but by reducing the stability of that fixed point so that the neural trajectory immediately falls off and returns to a reversal state. We hypothesize that β may also have a biologically correlated neuromodulator or set of neuromodulators and that stabilization of this modulation system would remove the reversal bout phenomenon.

An additional testable prediction is that some subset of neurons correlated with forward motion (e.g. the AVB and RIB pairs) or the ending of reversals (e.g. the SMDD, SMDV, and RIV pairs) may be responsible for stabilizing the forward state and others may be key for initializing the state. Opto-genetic manipulation of the "initiating" neurons without the "stabilizing" neurons should simply produce a failed forward initialization, as seen in the natural reversal bout. Similarly, inhibition of the stabilizing neurons should make forward motion an inaccessible state.

2.3 Discussion

We have produced the first global, nonlinear control model that can capture the dominant features of low-dimensional neural data. Our work demonstrates how the *C. elegans* neural network could control its global dynamics via perturbations to fixed point stability and feed-forward control signals. This model provides a control theory mechanism for switches in stochastic switching models. Our model also extends previous work by explaining incomplete or unsuccessful switching seen in reversal bouts as a change in the stability of the underlying fixed point. This model is minimally parameterized and changes in several parameters can reproduce changes in behavioral distributions akin to that of known neuro-modulators, thus producing a unifying framework for analyzing various changes in distributions of behavior at multiple timescales. In addition, the framework for building this model can be extended to other complex systems with more behavioral states which are defined by fixed points as

discussed in [149].

Several modeling strategies have been used to model *C. elegans* behavioral and neural dynamics, and they can be classified in two ways: direct models of the trajectories in neuron space [127, 62, 129, 128], and abstract Markov models [176]. The former has the advantage of describing neuron-level dynamics at the cost of many parameters, generally hundreds. On the other hand, Markov models do not make specific predictions about neurons or trajectories on the low-dimensional manifold, but generally have a small number of very interpretable parameters. Our model combines the strengths of both approaches, producing a model of dynamics that is both directly connected to neural activity and fits only 6 parameters. It is unclear if these parameters have biological correlates, but the fact that modulating them produces known behavioral outcomes suggests areas for future experimental work.

This modeling strategy has a few limitations. In particular, the entire model was constructed and fit using the first two PCA modes, which only account for 18-23% of the variance in the data. The dominant modes capture the dominant global dynamics; however, there are many secondary structures captured by the later modes such as transient or sparse activity. The fast-timescale signals that control the global dynamics may be captured in the activity of these higher modes. It is almost certainly true that important activity is contained in higher PCA modes, particularly when trying to incorporate more complex behaviors. In addition, it is unclear that PCA modes are the correct basis for producing models whose behaviors have biological correlates. Work regarding an interpretable choice of basis is ongoing, with nonlinear embeddings offering more flexible possibilities [133, 35].

Connected to this issue, the model does not differentiate between ventral and dorsal turns or between transitions rev 1 and rev 2. Transition paths are not clearly separable in the first two PCA modes, even though they are clearly mutually exclusive at the level of muscle activation. A model with more variables would be able to differentiate between these different transition paths from the forward to reversal state and from the reversal to forward state. Extending our framework to incorporate more subtle and complex behaviors is the subject of ongoing work.

The modeling strategy proposed in this paper used polynomials to design fixed points and the transitions between them. Even if the "true" function form is more complex, polynomials can be considered a Taylor expansion approximation of those dynamics. However, no attempt was made to explicitly derive this functional form from neuron-level nonlinearities, or to include information from the known connectome [217]. A derivation from first principles would be an exciting advance and we hope that our model, as one possible macro-scale model, can facilitate this type of theoretical development.

2.4 Methods

We construct a nonlinear control model for *C. elegans* by fitting the parameters of a general dynamical systems model with control to low-dimensional *C. elegans* neural activity. We reduce the dimension of the neural trajectories with PCA and use non-convex optimization to fit the trajectory of the dominant PCA mode to the corresponding dynamical systems model variables.

2.4.1 Dimension reduction

C. elegans have been proposed to have seven different behaviors — forward motion, forward slow, dorsal turn, ventral turn, reversal 1, reversal 2, and sustained reversal [102]. Further references to the forward behavior denote both the forward motion and forward slow states, and references to the reversal behavior denote the sustained reversal state. We used *C. elegans* calcium imaging data collected for [102] to produce the low-dimensional activity shown in Figure 2.1 and the model fits shown in Figures 2.2 and 2.3. We used calcium imaging data collected for [186] to produce the low-dimensional activity and distributions shown in Figure 2.4. We used data from five different *C. elegans* from [102], as their neural patterns expressed activity corresponding to different labeled behaviors. We achieved a low-dimensional representation of the activity by performing *principal component analysis* (PCA) on the time series data and focusing on the activity of the first two PCA modes.

Each of the five *C. elegans* datasets contains calcium imaging from 107-131 neurons. We

performed PCA on the timeseries of each *C. elegans* neural activity, which is calculated as the time derivatives of the normalized Ca^{2+} traces ($\Delta F/F$) [102]. We used the first two PCA modes to construct a dynamical systems model. The first two PCA modes capture 18-23% of the total variance in each dataset meaning that our model only represents the dominant neural activity and is excluding secondary activity. Bleaching causes the calcium imaging signals to dampen over time in each timeseries [102]. Because we are only interested in transitions between behaviors that are known to be faster timescale [100], we correct for this by subtracting a long-timescale moving average from each principal component. Figure 2.5(a) shows the mean centered neural activity of a single *C. elegans*. Figure 2.5(b-d) shows the first two principal components of the neural activity without drift correction while Figure 2.5(e-g) shows the first two principal components with drift correction.

2.4.2 Nonlinear Dynamical Systems Model

Nonlinear dynamical systems are ubiquitous in the engineering, physical and biological sciences for describing many complex phenomena observed in a diverse number of settings. Often, simple qualitative models with polynomial nonlinearities are capable of providing remarkable insight into dynamical behaviors. The nonlinear pendulum, for instance, can be approximated by a Taylor series expansion to characterize the effects of frequency shifts and harmonic generation that is observed in practice. Inspired by well-studied nonlinearities, we consider dynamical systems of the general form

$$\dot{\mathbf{x}} = f(\mathbf{x}, \beta, \gamma) + \mathbf{B}u(t). \tag{2.6}$$

We restrict our focus to polynomial equations with fixed points that can be determined

analytically:

$$\begin{aligned} x' &= y \\ y' &= f(x) + \gamma y + u(t) \\ f(x) &= a\prod_{i=1}^{n}(x - r_i) \end{aligned} \tag{2.7}$$

where $f(x)$ is a polynomial with a leading coefficient a and roots r_i. γ is the damping parameter. This is a second order nonlinear differential equation which can be expressed as $x'' - f(x) - \gamma x' = 0$. If $\gamma = 0$, the system is undamped and the differential equation becomes $x'' = f(x)$ which has an analytical solution. Often however, the solutions are exceedingly complex and it is preferable to take a qualitative approach. We choose a system of this form as the fixed points can be easily placed and assigned a stability type (e.g. saddles, sources, sinks, or centers) through parameter selection. All fixed points lie on the x-axis and are placed and manipulated by varying our polynomial roots r_i, while fixed point stability types are assigned by manipulating γ and a for a given set of roots r_i. Two stable fixed points have been identified in the *C. elegans* dynamics, suggesting a cubic dynamical system. Additional features of the data and how they can be translated into a nonlinear dynamical system are described in Table 2.3.

2.4.3 Model fitting

We fit the parameters ($P = [\beta \;\; \gamma \;\; \sigma \;\; u_{3,4} \;\; u_{5,6} \;\; dt]$) of our general model (Eq. 2.4) to each *C. elegans* low-dimensional activity by minimizing the error $\int |E(t)|dt$ where $E(t) = v_1(t) - x(t)$. Parameter β is the location of the saddle fixed point and controls the relative stability of forward and reversal states, γ is the damping parameter, and σ is the level of stochasticity in the system. Control signal $u(t) = u_{3,4}$ during states 3 and 4 (dorsal and ventral turns) and $u(t) = u_{5,6}$ during states 5 and 6 (rev 1 and rev 2 transitions). The behavioral state timeseries has been determined by [102]. Parameter dt scales the model timesteps so that they

fit the measurement intervals of the calcium imaging data. This is a non-convex optimization problem. We first perform a random search over the parameter space until the model performs most transitions. We then continue optimizing via MATLAB's fminsearch function. The random search resulted in parameters $P = [0.1 \quad -1.5 \quad 0.06 \quad 0.5 \quad -0.7 \quad 0.3]$ which we used as the initial condition for the fminsearch function. We optimized for > 200 interations for each model fit. While this method finds a suitable collection of parameters that execute the transitions observed in the data (Fig. 2.3), it does not guarantee the optimal solution will be found.

2.4.4 Robustness of results to parameter variations

We observe how modifying other system parameters affect the state distribution of the nonlinear system's activity under randomly generated control signals. In Figure 2.6(a) we vary the right fixed point's region of stability by moving the location of the middle fixed point β. We observe the system spends less time at the right fixed point with a smaller stability region. In Figure 2.6(b) we increase the level of Brownian motion (σ) in the system and observe the variability increases in the distributions as a result. In Figure 2.6(c) we observe that increasing the control signal frequency increases the amount of time spend in a transitional state. Figure 2.6(d) shows that increasing the damping strength decreases the distribution variability. Observing these parameter variations holistically, we see that the nonlinear model is able to perform the task of switching between fixed points under a wide range of parameter values which insures the integrity of the system and indicates that *C. elegans* dynamics, if comparable to this model, should be able to operate robustly and stably under a diverse array of environments and internal states.

Paradigm	Parameters	What it models	Implications	Refs
Markov	< 10	Clusters and transitions	Differentiate macro-scale behaviors	[69]
Linear (switching)	>1000	Local linear dynamics in neuron-activity spaces	Connection between behavioral clusters and neural dynamics within them	[128, 127, 129, 44]
Linear (controlled)	>100	Global linear dynamics and control inputs	Disentangle intrinsic dynamics and transition mechanisms	[62]
Nonlinear (controlled)	<10	Global nonlinear dynamics and control inputs	Can model different classes of transitions	This work

Table 2.1: Different modeling paradigms for *C. elegans* with experimental implications.

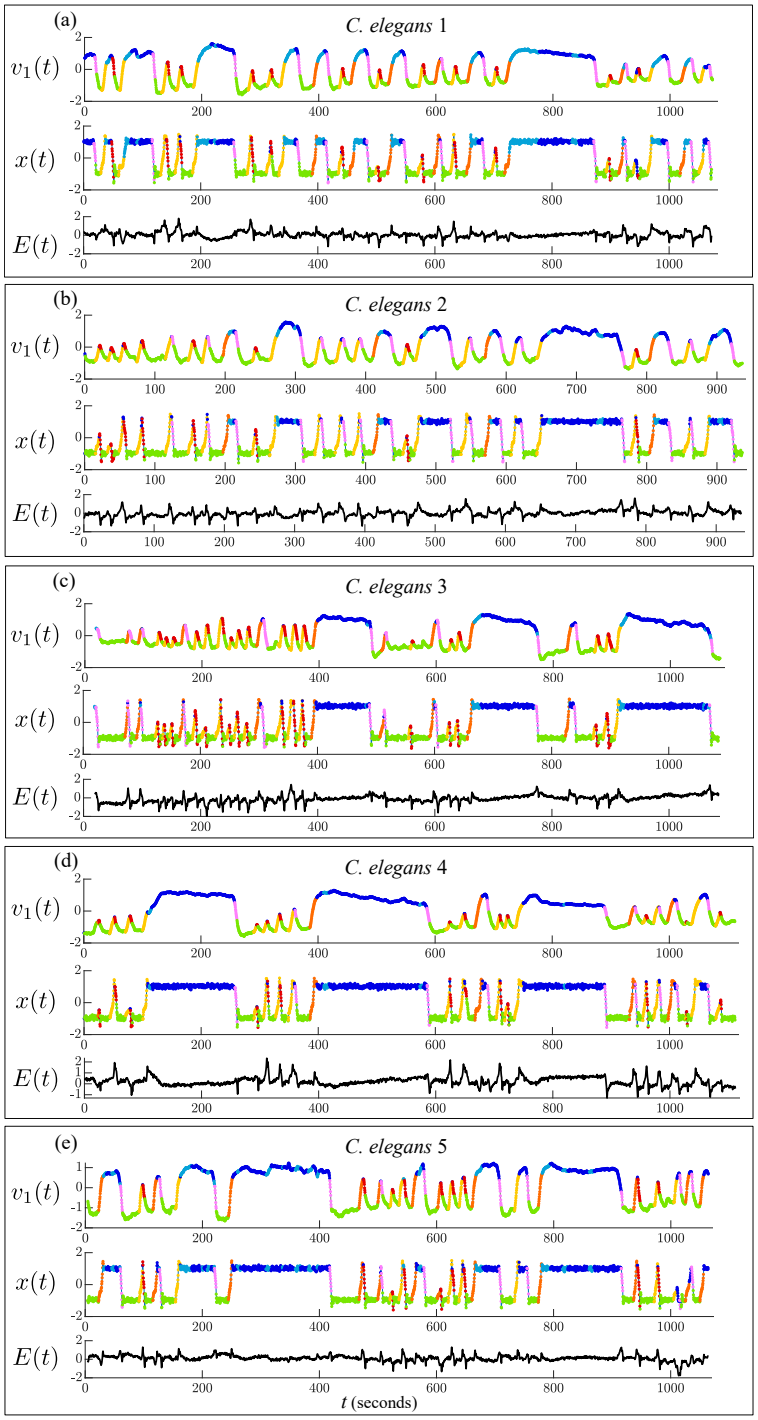

Figure 2.3: Timeseries of dominant mode of *C. elegans* neural activity ($v_1(t)$) and corresponding model variable ($x(t)$). Models are fit to each *C. elegans* by minimizing the error ($E(t)$) between the PCA and model timeseries. Trajectories are colored by behavioral state.

C. elegans	$\boldsymbol{\beta}$	$\boldsymbol{\gamma}$	$\boldsymbol{\sigma}$	$\boldsymbol{u_{3,4}}$	$\boldsymbol{u_{5,6}}$	dt
Id 1	0.1108	−1.5027	0.0574	0.4969	−0.7531	0.2902
Id 2	0.1124	−1.5281	0.0584	0.5059	−0.7159	0.2953
Id 3	0.1107	−1.5321	0.0584	0.5262	−0.7549	0.2951
Id 4	0.1135	−1.4830	0.0591	0.5076	−0.7292	0.3008
Id 5	0.1087	−1.5115	0.0598	0.5350	−0.7731	0.2929

Table 2.2: Parameters for models fit to each *C. elegans* calcium imaging dataset. Data, model, and error timeseries shown in Fig. 2.3.

C. elegans	**Dynamical System**
Two stable fixed points	Globally stable system with two sinks
System functions with variability	System behavior remains qualitatively constant under small parameter perturbations
Trajectories contain stochasticity	System behavior remains qualitatively constant with the addition of noise
Fixed point locations drift	Behavior remains qualitatively constant despite deformations and shifts to the system
Trajectories tend to follow set paths	System path variability set with damping term

Table 2.3: Features exhibited by *C. elegans* neural activity paired with corresponding dynamical system features.

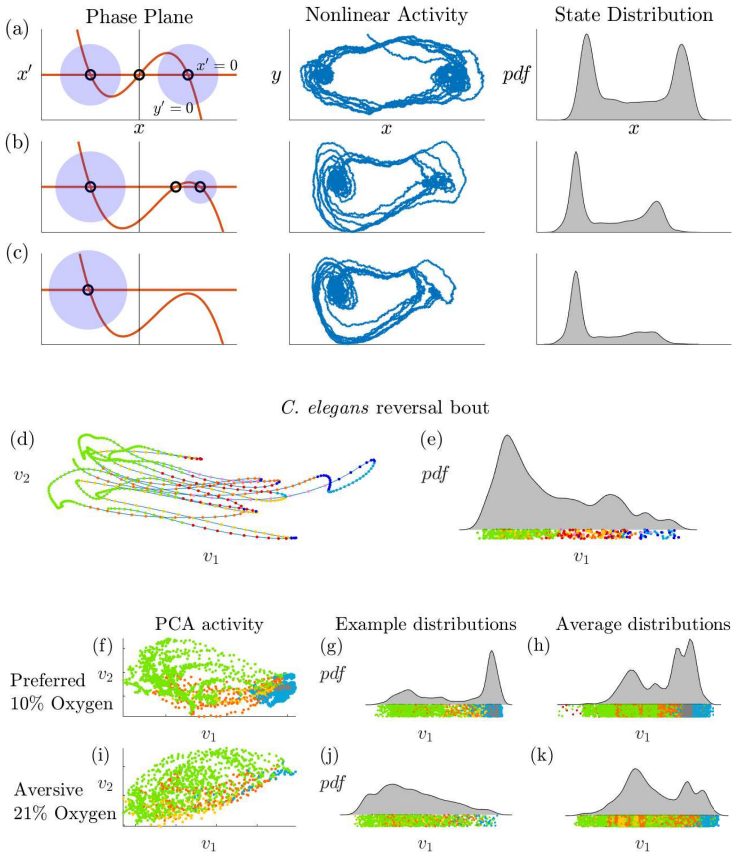

Figure 2.4: (a)-(c) Phase plane, nonlinear stochastic activity, and state distributions of Eq. (2.4) with increasing β values. (a) $\beta = 0$ generates equally stable fixed points. (b) $\beta = 0.6$ generates a less stable fixed point which turns into a slow point as the fixed points merge. (c) $\beta, r_2 \in \mathbb{C}$ and the right fixed point is lost. (d) *C. elegans* PCA trajectory during a reversal bout and (e) the corresponding distribution. The forward fixed point is unstable during this interval. (f)-(h) *C. elegans* activity in a preferred 10% oxygen environment which promotes stability in the forward state compared with (i)-(k) *C. elegans* activity in an aversive 21% oxygen environment which destabilizes the forward state. (f)-(g) PCA activity and distribution of a single *C. elegans* in the preferred oxygen environment compared with the activity of this same *C. elegans* in the aversive oxygen environment (i)-(j). Average distribution for 10 *C. elegans* in the preferred environment (h) compared to the aversive environment (k).

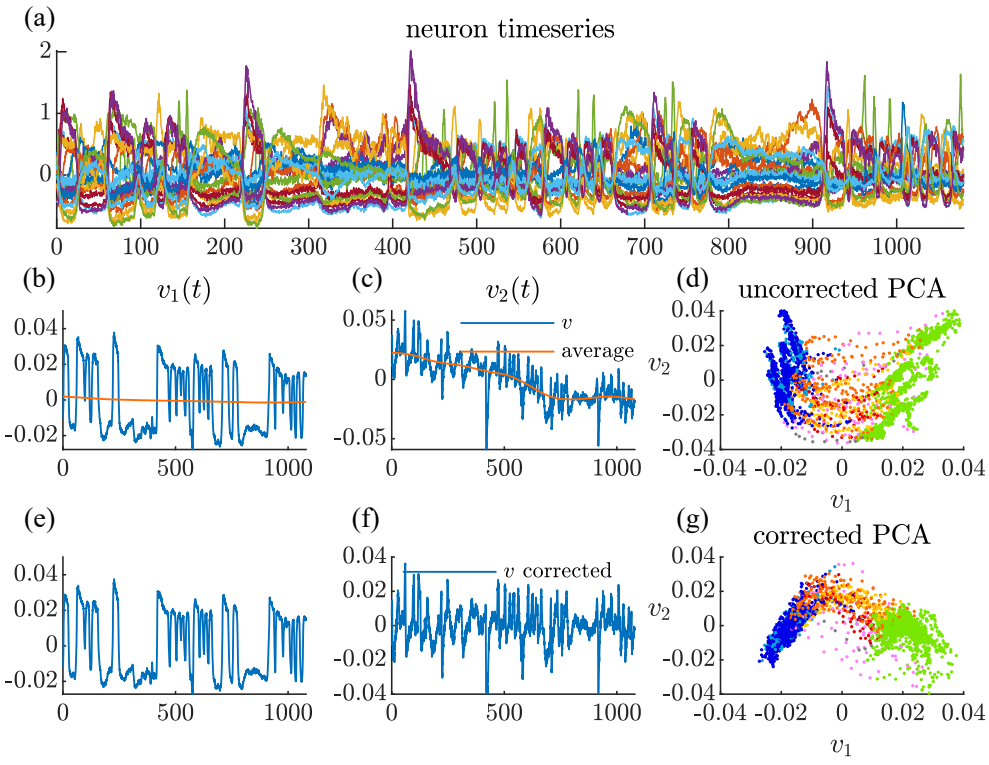

Figure 2.5: Principal component analysis of neural activity. (a) Calcium imaging timeseries mean centered (time in seconds). (b) Timeseries of first principal component with moving average. (c) Timeseries of second principal component with moving average. (d) Neural activity in PCA space using uncorrected PCA. (e) Timeseries of first principal component with moving average subtracted. (f) Timeseries of second principal component with moving average subtracted. (g) Neural activity in PCA space uinsg corrected PCA.

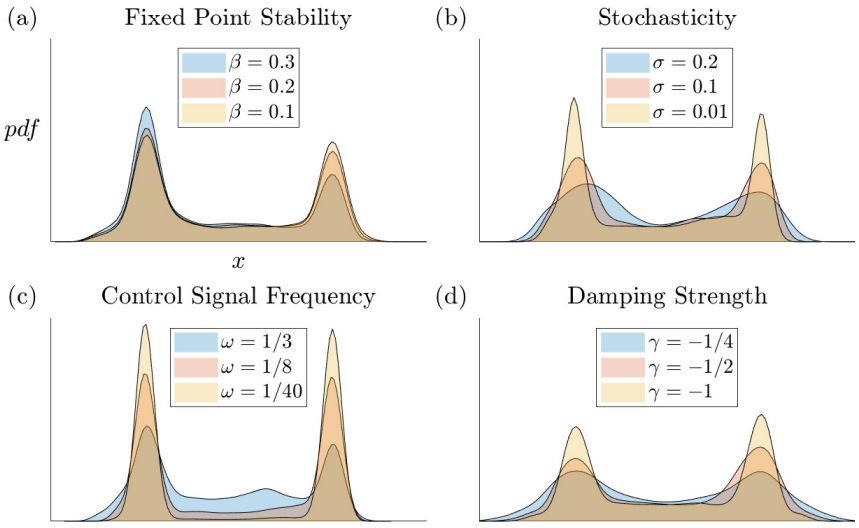

Figure 2.6: State distributions of nonlinear models for various parameter regimes. (a) Fixed point relative locations affects their stability. (b) Increasing levels of Brownian motion (σ) increases the variation about the fixed points. (c) More frequent control signals more evenly distributes the time spent in stable versus transitional states. (d) Stronger damping in the system keeps trajectories close to fixed points.

Chapter 3

NONLINEAR CONTROL OF NETWORKED DYNAMICAL SYSTEMS

3.1 Introduction

[1] Networked dynamical systems are ubiquitous across the engineering, physical, biological and social sciences. They are often characterized by a high-dimensional state space and nonlinearity, making them exceptionally difficult to characterize and control. Indeed, it is typical that the connectivity is so complex that the functionality, control and robustness of the network of interest is impossible to characterize using standard mathematical methods. Moreover, with few exceptions, underlying nonlinearities impair our ability to construct analytically tractable solutions, forcing one to rely on experiments and/or modern high-performance computing to study a given system. Unlike engineered systems that are constructed to be both measurable and controllable, such emergent systems can be difficult to measure and have restricted avenues of control. However, advances over the past decade have revealed a critical observation, that meaningful input/output of signals in high-dimensional networks are often encoded in low-dimensional patterns of dynamic activity [94, 171, 24, 102, 137, 62, 47]. We show that such low-dimensional patterns of activity can be exploited in order to develop principled techniques for a feed-forward architecture for establishing control of high-dimensional, nonlinear networked dynamical systems.

The potential applications of a control framework for networked dynamical systems are extensive. Neuroscience is an especially relevant example where networks of neurons interact to encode and process input stimulus and behavioral responses. Recent observations in a variety of organisms, from the nematode *C. elegans* [113, 102, 62] to insect olfactory processing [94,

[1] ©2021 IEEE. Reprinted, with permission, from [149]

171, 184, 47], shows that the underlying encodings and control are fundamentally low-dimensional. Network models are also common in attempts to understand the formation and retrieval of memories, such as proposed in the Hopfield model where each memory is a fixed point in the high-dimensional, networked dynamical system [87, 150]. Indeed, it is known that the nervous systems carries out an impressive feat of dimensionality reduction when it encodes behavior, collapsing the high-dimensional representation of the stimulus environment into the much lower representations for decision making and motor command. Practical emerging technologies, such as *deep brain stimulation* (DBS), aim to leverage such control protocols to restore patients to their original functional capabilities. Applications extend well beyond neuroscience, with the potential of the method to impact disease modeling [106, 83], social [147] and financial networks, powergrid networks [55], synthetic biology [170], and ecosystems, for instance.

Characterization is only the first step in understanding networked dynamical systems. A principled quantification of the low-dimensional patterns of dynamic activity can help lead to control protocols for manipulating the system into a desired outcome. An extensive body of literature exists on the analysis and control of nonlinear systems [90]. Unlike many linear control models, where controllability and observability can be explicitly computed and guaranteed, nonlinear control remains challenging, especially in networked settings. Nonlinear networked dynamical systems can contain many fixed points, limit cycles and strange attractors, all of which make the development of principled control models difficult. The manifestation of these various phenomenon must be addressed in any practical control paradigm. On the other hand, one can use the existence of such rich dynamical structures to allow the network itself, under suitable manipulation, to evolve to a desired state of behavior. Thus nonlinearity can exploit a much broader class of dynamics and function than linear models.

We integrate methods of dimensionality reduction and data-driven discovery of dynamics to construct principled methods for controlling nonlinear, networked dynamical systems. Specifically, we develop feed-forward control techniques to interpret and regulate the dynamics

of such systems by leveraging dominant, low-dimensional subspaces on which the dynamics evolves. Our mathematical architecture generates a set of actuation signals that, when applied, are able to control the original high-dimensional system. Using bifurcation theory, we find collections of feed-forward control signals that will force convergence to desired objective states, allowing us to move the system from one fixed point of the system to another in a principled manner. Specifically, we can destabilize a given fixed point by making it undergo a saddle node or Hopf bifurcation, while simultaneously making the target fixed point an attractor. This creates a pathway with the feed-forward signals from one fixed point to another. We first demonstrate our nonlinear control procedure on established bistable, low-dimensional biological systems showing how control signals are found that generate switches between attractors. We then show how random high-dimensional networks and Hopfield memory networks can be reliably controlled by discovering low-dimensional subspaces which characterize their dynamic evolution. Our algorithmic procedure is the first of its kind to provide a principled mathematical architecture that simultaneously leverages mode discovery, dimensionality reduction, and bifurcation theory for controlling networked dynamical systems.

The paper is outlined as follows: Sec. II provides a brief overview of the background material necessary for constructing our control framework. Section III provides an analysis of feed-forward control applied to low-dimensional nonlinear dynamical systems. This highlights the basic mathematical architecture that is used in Sec. IV and V for network control applications in both low- and high-dimensional systems respectively. A number of practical applications are considered in Sec. VI, including the Hopfield memory model where we show how control can be used to transition between fixed points, or memories, in the network. The paper is concluded in Sec. VII with a discussion of our results.

3.2 Background

3.2.1 Dimensionality Reduction

It is typically observed that high-dimensional dynamical systems manifest behavior on low-dimensional manifolds [94, 171, 24, 102, 137, 62, 47]. Indeed, low-dimensional structures can be exploited for characterizing pattern forming systems [45] and reduced order models [16]. *Principal component analysis* (PCA) is a linear dimensionality reduction technique based upon the *singular value decomposition* (SVD) that can extract dominant correlated features in high-dimensional data [93, 117, 24], thus producing a coordinate system (subspace) on which our networked dynamics of interest can be projected. Let $\mathbf{X} \in \mathbb{R}^{n \times m}$ represent timeseries data collected from an n dimensional system for m timepoints. The SVD of this matrix produces the matrix decomposition [208, 117, 24]

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^* \tag{3.1}$$

were $\mathbf{U}$ denotes the dominant correlated spatial structures of the n-dimensional system, $\mathbf{S}$ is a diagonal matrix whose singular values characterize an ordered ranking of the correlations, and $\mathbf{V}$ represent the projection of the modes into the temporal dimension. Both $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices with orthonormal columns. A low-dimensional, r-rank system can be optimally approximated in an ℓ_2-sense using the first r columns of each matrix: $\mathbf{X} = \hat{\mathbf{U}}\hat{\mathbf{S}}\hat{\mathbf{V}}^*$ where $\hat{\mathbf{U}} \in \mathbb{R}^{n \times r}$, $\hat{\mathbf{S}} \in \mathbb{R}^{r \times r}$, and $\hat{\mathbf{V}}^* \in \mathbb{R}^{r \times m}$. Such low-rank subspaces are exploited for building reduced order models that approximate the high-dimensional system [16, 24]. It is also exploited in what follows since random networks [Appendix A.3] generically manifest low-dimensional behavior, as shown in Fig. 3.1 where the cumulative variance contained in the initial modes of a randomly generated dynamical system is plotted.

3.2.2 Sparse Identification of Nonlinear Dynamics - SINDy

SINDy is a data-driven approach to find the sparse dynamics driving a dynamical system purely from timeseries data [25]. If $\dot{x} = f(x)$ is the unknown true dynamics of a system from

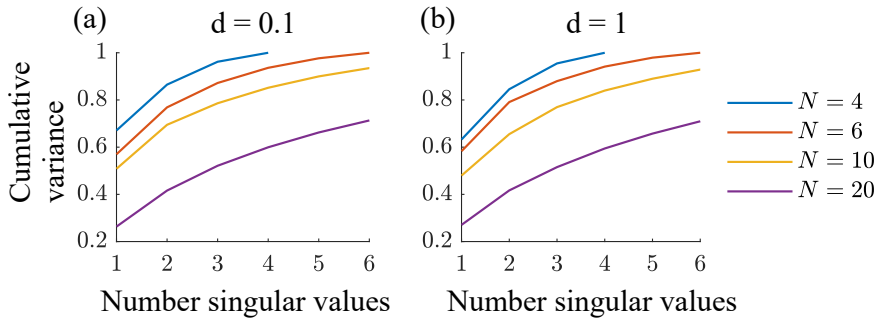

Figure 3.1: Average cumulative variance captured by initial singular values in random dynamical systems of increasing size $N = 4, 6, 10, 20$ [Appendix A.3]. (a) Dynamical systems consisting of a low percentage of possible term combinations, $d = 0.1$ (b) All possible term combinations included in the dynamical system, $d = 1.0$. Variance averaged over 100 trials

which we capture timeseries measurements $\mathbf{X}$, the SINDy algorithm seeks to approximate $f(x)$ by a generalized linear model in a set of candidate basis functions $\theta_k(x)$

$$f(x) \approx \sum_{k=1}^{p} \theta_k(x)\xi_k = \mathbf{\Theta}(x)\mathbf{\Xi}, \tag{3.2}$$

with the fewest non-zero terms in $\mathbf{\Xi}$. It is possible to solve for the relevant terms that are active in the dynamics using sparse regression algorithms that penalizes the number of terms in the dynamics and scales well to large problems. SINDy works well on low-dimensional systems with a high-quality selection of candidate terms in the library [25, 135, 35, 177, 26]. With too little data, or too many library terms (which can result from too many variables), the algorithm can fail to produce a viable model.

3.2.3 Control

Control can be broadly divided into the categories of open-loop and closed-loop [189, 24]. Most advanced control techniques are closed-loop, based on constructing optimal feedback to stabilize a system at a particular set point and are reliant on system measurments [189, 24] [31, 4, 63, 97, 24]. From a dynamical systems perspective, some research considers the global

dynamics that are achievable via feedback control signals by viewing control as parameter manipulations that bring about bifurcations in local dynamics [90]. Strategic perturbations to tunable parameters is a form of control as it affects the stability of the nonlinear system's fixed points. Instead of narrowly considering a single trajectory, this line of research investigates the behavior of ensembles of trajectories. One growing field that uses this approach is synthetic biology [170]; researchers have created synthetic, bistable gene-regulatory networks in *Escherichia coli* that can be triggered to switch between stable states [72]. Strategic perturbations to a select number of nodes can induce dynamics on a network to converge to a desired target state, rescuing it from convergence to an undesirable state [43]. Stochasticity has also been proposed as a useful control mechanism in biological systems. Transcription can be characterized as a bistable dynamical system where stochasticity allows transcription to fluctuate between the two stable states and therefore acts as a mechanism for regulating transcription [105]. Further research found that noise can be exploited to induce desired state transitions in other similarly obscure network dynamical systems [216].

While feedback control methods grant optimal control for systems that can be continuously measured and actuated, many systems of interest are not amenable to constant monitoring or reactive control signals. Furthermore, highly nonlinear systems may contain many stable attractors that do not require feedback control to stabilize, and can be achieved with transient open-loop control signals [146]. Previous work has considered pulse shape and duration to switch between stable states in certain bistable dynamical systems in synthetic biology [191, 192, 190]. While some analyze stable fixed points to understand switching behavior, others have analyzed saddle points to enlighten global decision-making processes in bistable systems [209].

We continue in this dynamical systems view and consider the global and local dynamics that are possible via system bifurcations induced through feed-forward control signals. Previous feed-forward control methods have been limited in that they only consider switching between stable states in certain types of bistable dynamical systems with known dynamics. We provide a general method for switching between any number or type of stable attractor in

networked dynamical systems using open-loop control. This method can be applied to general dynamical systems of various sizes with either known or unknown governing dynamics. Our pulse-based switching procedure takes advantage of the nonlinearities of the system to move between stable attractors as well as to create and stabilize fixed points. Our control method is limited in that only a subset of all possible states and stability levels can be achieved using feed-forward control in an initial condition agnostic fashion. Nonetheless, this type of control may be useful in systems where system measurements are difficult to obtain or feedback mechanisms are impossible.

3.3 Feed-forward control for low-dimensional nonlinear systems

We consider how to generate control signals that will move a system between attractors (fixed points) in a nonlinear dynamical system $\mathbf{x}' = F(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$. We consider systems of the form

$$\frac{dx}{dt} = f(x, y) + u_1(t) \tag{3.3}$$

$$\frac{dy}{dt} = g(x, y) + u_2(t). \tag{3.4}$$

where f and g are the intrinsic dynamics and $u_1(t)$ and $u_2(t)$ are the feed-forward control signals. Systems of this form are controlled by regulating the actuating forces. Many actions taken to control systems can be characterized as feed-forward control signals that do not alter the structure of the underlying system. Adding reactants to a chemical reaction, removing invasive species, performing deep brain stimulation, and taxing individuals can all be characterized as feed-forward control signals applied to nonlinear dynamical systems.

Biological systems may also utilize feed-forward control signals themselves to modulate internal processes. Behavior transitions in the nematode *C. elegans* can be characterized by feed-forward control signals applied to a nonlinear system with two attractor states [146]. While transient feed-forward signals control short-term behavior in this model of *C. elegans*, system parameter modifications affect long-term behaviors, highlighting the different ways biological systems can control their output across multiple timescales. Our control

framework is useful both for developing control strategies for nonlinear systems as well as for understanding how natural systems accomplish endogenous control.

3.3.1 Fixed Point Stability

We use properties of the governing nonlinear system to derive the set of control signals that will move us between stable fixed points by modifying their stability. The Jacobian of the nonlinear system $J(x, y)$ and therefore its trace $T(x, y)$ and determinant $D(x, y)$ are independent of the control signals, meaning that fixed point stability is dependent only on location. These three quantities are given by the following respectively:

$$J(x,y) = \begin{bmatrix} f_x(x,y) & f_y(x,y) \\ g_x(x,y) & g_y(x,y) \end{bmatrix}, \tag{3.5}$$

$$T(x,y) = f_x(x,y) + g_y(x,y), \tag{3.6}$$

$$D(x,y) = f_x(x,y)g_y(x,y) - g_x(x,y)f_y(x,y). \tag{3.7}$$

Given a fixed point (x^*, y^*) occurring at $u_1 = -f(x, y)$ and $u_2 = -g(x, y)$. Fixed points exist in one of four stability regions in the trace-determinant plane. Fixed points in the region $T < 0$ and $D > 0$ are stable sinks, points in the region $T > 0$ and $D > 0$ are unstable sources, while points in the region $D < 0$ are unstable saddles [197]. Note that in the case of a linear system, J is a constant matrix which means that the control signals affect only the location of the fixed point and not its stability. In contrast, feed-forward control signals in a nonlinear system affect both the location and the stability of fixed points; this implies that a much broader range of activity can be generated by feed-forward control signals in nonlinear systems than in linear systems.

Fixed points transition between stability regions along the curves $T(x, y) = 0$ and $D(x, y) = 0$. Transitions between regions can also occur at $D(x, y),\ T(x, y) = \pm\infty$, we therefore define two additional measures: $\hat{D}(x, y) = 1/D(x, y)$ and $\hat{T}(x, y) = 1/T(x, y)$. We thus note that region transitions also occur along the curves $\hat{D}(x, y) = 0$ and $\hat{T}(x, y) = 0$. Saddle-node bifurcations occur along the curves $D(x, y) = 0$ and $\hat{D}(x, y) = 0$ while Hopf

bifurcations occur along the curves $T(x, y) = 0$ and $\hat{T}(x, y) = 0$. For non-smooth functions $F(\mathbf{x})$, such as piecewise functions, there are potentially additional transitions that occur when the function's partial derivatives are undefined. The curves along which the trace and determinant are undefined are additional locations where they may switch signs, meaning these curves must be included as potential transition curves.

By constraining the relation between x and y by $D(x, y) = 0$ and $\hat{D}(x, y) = 0$ we can solve for parameterized curves for the control signals that eliminate stable fixed points in the system by inducing a saddle-node bifurcation. This curve in the control signal space occurs along

$$C^s(t) = (u_1^s(t), u_2^s(t)) \tag{3.8}$$

where $u_1^s(t) = -f(t, y(t))$ and $u_2^s(t) = -g(t, y(t))$. The $y(t)$ in these formulae is the implicit solution to $D(t, y(t)) = 0$ or $\hat{D}(t, y(t)) = 0$. In order for $C^s(t)$ to be a boundary between regions, D must switch signs when crossing the curve. $C^s(t)$ is a boundary curve unless $\frac{\partial D}{\partial \boldsymbol{v}_\perp(t)} = 0$ and $\frac{\partial^2 D}{\partial \boldsymbol{v}_\perp^2(t)} \neq 0$ where $\boldsymbol{v}_\perp(t) = -u_2(t)\hat{\mathbf{i}} + u_1(t)\hat{\mathbf{j}}$ is the direction orthogonal to $C^s(t)$.

Stable fixed points can also be eliminated through Hopf bifurcations which occur along the curve

$$C^h(t) = (u_1^h(t), u_2^h(t)) \tag{3.9}$$

where $u_1^h(t) = -f(t, y(t))$, $u_2^h(t) = -g(t, y(t))$, and $y(t)$ is now the implicit solution to $T(t, y(t)) = 0$ or $\hat{T}(t, y(t)) = 0$. $C^h(t)$ is a boundary curve unless $\frac{\partial T}{\partial \boldsymbol{v}_\perp(t)} = 0$ and $\frac{\partial^2 T}{\partial \boldsymbol{v}_\perp^2(t)} \neq 0$ where $\boldsymbol{v}_\perp(t)$ is the direction orthogonal to $C^h(t)$.

We can use these curves to determine stability regions in the control space for each fixed point in the dynamical system. Let $P = \{p_1, p_2, ..., p_n\}$ be the set of fixed points in the nonlinear system $\mathbf{x}' = F(\mathbf{x})$. Each fixed point is associated with one of four stability regions A, B, C, or D. Let A_k be the set of control signals $u \in \mathbb{R}^2$ such that fixed point p_k is stable under u. Let B_k be the set of control signals $u \in \mathbb{R}^2$ such that fixed point p_k is a source under u. Let C_k be the set of control signals $u \in \mathbb{R}^2$ such that fixed point p_k is a saddle

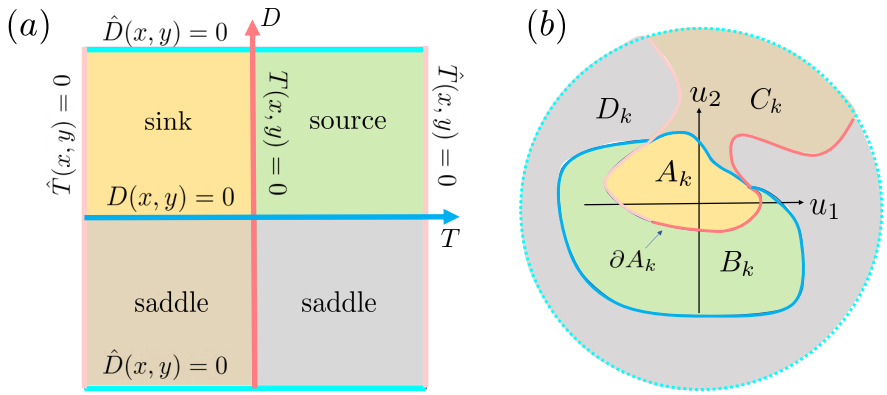

Figure 3.2: Regions of stability and instability for fixed points in a dynamical system. (a) Fixed points are sinks, sources, or saddles, depending on where they lie in the trace-determinant plane. (b) These regions map to stability regions in the control space. Movement across bifurcation curves in the control space correspond to moving between stability regions in the trace-determinant plane by crossing $D = 0$, $T = 0$, $\hat{D} = 0$, or $\hat{T} = 0$.

with $T < 0$ and let D_k be the set of control signals $u \in \mathbb{R}^2$ such that fixed point p_k is a saddle with $T > 0$. A_k, B_k, C_k, and D_k are disjoint regions. Fixed point p_k is destabilized in region $A_k^c = \mathbb{U} \backslash A_k = B_k \bigcup C_k \bigcup D_k$. We denote ∂A_k as the boundary of the stability region A_k of fixed point p_k. Figure 3.2(a) shows the stability regions and boundaries in the trace-determinant plane while Fig 3.2(b) shows the mapping of these stability regions into the control plane.

3.3.2 Controllability Given No Limit Cycles

We can now define necessary and sufficient conditions for moving between fixed points in a system that does not contain limit cycles. The following conditions concern moving between specific fixed points.

Necessary condition to move directly from p_ℓ to p_k:

$$A_\ell^c \bigcap A_k \neq \emptyset \tag{3.10}$$

Sufficient conditions to move directly from p_ℓ to p_k:

$$\bigcap_{i=1,i\neq k}^{n} A_i^c \bigcap A_k \neq \emptyset \tag{3.11}$$

We can extend these conditions to make statements about the reachability of all fixed points in the system.

Necessary condition to move directly from any fixed point in the system to any other fixed point:

$$\forall \ell, k \in 1, ..., n \quad A_\ell^c \bigcap A_k \neq \emptyset \tag{3.12}$$

Sufficient conditions to move directly from any fixed point in the system to any other fixed point:

$$\forall k \in 1, ..., n \bigcap_{i=1,i\neq k}^{n} A_i^c \bigcap A_k \neq \emptyset \tag{3.13}$$

Necessary conditions to move to any point in the system:

$$\forall k \in 1, ..., n \; \exists \ell \in 1, ..., n \; s.t. \; A_\ell^c \bigcap A_k \neq \emptyset \tag{3.14}$$

We can use the control regions to move between fixed points in the system by destabilizing fixed points we want to escape, and stabilizing fixed points we want to achieve. Transitioning between fixed points through saddle-node bifurcations is preferable to transitioning via Hopf bifurcations as Hopf bifurcations can create a stable limit cycle around the source.

While these sets generate collections of control signals that can be used to move between fixed points, it does not reveal the optimal control signals for a transition which depends on additional goals such as energy minimization, transition speed, robustness, or preferred path.

3.3.3 Limit Cycles

Limit cycles are a pervasive feature of many nonlinear systems and so must be considered. They appear in canonical systems such as the Hodgkin-Huxley model [85], FitzHugh-Nagumo

model [64], Brusselator [197], circadian rhythms [145], and extensions to the predator-prey Lotka-Volterra model [86]. We also observed limit cycles and strange attractors appear in our randomly generated nonlinear dynamical systems [Appendix A.3]. Even systems that do not naturally exhibit limit cycles may be perturbed to exhibit them via Hopf bifurcations.

Unfortunately, the location and stability of limit cycles are more difficult to characterize than the location and stability of fixed points in a dynamical system. We can make some statements about where limit cycles do and do not appear by using Dulac's criterion and the Poincaré-Bendixson theorem [197]. We can use Dulac's criterion to solve for regions encapsulating fixed points that cannot contain limit cycles and we can use the Poincare-Bendixson theorem to solve for regions that do contain limit cycles. For example, a globally stable system with a single fixed point that is a source is a compact set and therefore must contain at least one stable limit cycle encapsulating the fixed point. Some systems may contain compact sets enclosing all fixed points. Such sets can be found by eliminating the highest order terms that determine the global stability and then observing the stability of the resulting system as $t \to \infty$. Thus we consider

$$x' \approx P_{n-1}(x, y) + p(x^n, y^n) \tag{3.15}$$

$$y' \approx Q_{n-1}(x, y) + q(x^n, y^n) \tag{3.16}$$

If the global stability of (P_{n-1}, Q_{n-1}) is opposite the global stability of (f, g) and along a Jordan curve surrounding all fixed points $|P_{n-1}(x, y)| > |p(x^n, y^n)|$ and $|Q_{n-1}(x, y)| > |q(x^n, y^n)|$, then there must be a compact set outside of this Jordan curve and therefore a limit cycle.

In some systems multiple limit cycles may encapsulate a fixed point or region. We may find the presence of multiple layers of limit cycles by repeating the process on the approximate lower-order system $P_{n-1}(x, y) = P_{n-2}(x, y) + p(x^{n-1}, y^{n-1})$ and $Q_{n-1}(x, y) = Q_{n-2}(x, y) + q(x^{n-1}, y^{n-1})$. Once again if the lower order terms are larger in absolute value than the higher order terms along a Jordan curve surrounding the fixed points and enclosed by the first Jordan curve, then there must be a second limit cycle surrounding the fixed points.

We can determine the global stability of the system $\mathbf{x}' = F(\mathbf{x})$, that is, the stability as $x, y \to \pm\infty$, by instead considering the stability of the system $\hat{\mathbf{x}}' = G(\hat{\mathbf{x}})$ as $\hat{\mathbf{x}} \to 0$ where $\hat{x} = 1/x$ and $\hat{y} = 1/y$,

$$\frac{d\hat{x}}{dt} = -\hat{x}^2 \left(f\left(\frac{1}{\hat{x}}, \frac{1}{\hat{y}}\right) + u_1(t) \right) \tag{3.17}$$

$$\frac{d\hat{y}}{dt} = -\hat{y}^2 \left(g\left(\frac{1}{\hat{x}}, \frac{1}{\hat{y}}\right) + u_2(t) \right) \tag{3.18}$$

Our original system $\mathbf{x}' = F(\mathbf{x})$ is stable as $x, y \to \pm\infty$ if $G(\hat{\mathbf{x}})$ does not have any stable directions as $\hat{x}, \hat{y} \to 0$ and unstable otherwise.

We can determine that there are no limit cycles surrounding all fixed points by mapping the system to $G(\hat{\mathbf{x}})$ and using the Dulac's criterion to find a region surrounding $\hat{\mathbf{x}} = 0$ that does not contain any limit cycles. Conversely, we can determine if there are limit cycles surrounding a particular fixed point in a system with multiple fixed points by mapping that point to infinity in the system $G(\hat{\mathbf{x}})$ and then finding if there is a limit cycle surrounding all fixed points in $G(\hat{\mathbf{x}})$ using the method outlined above with the Poincare-Bendixson theorem.

3.3.4 Controllability with Limit Cycles

We can use bifurcations that create and eliminate limit cycles to move between limit cycles and fixed points in the system. Andronov-Hopf bifurcations eliminate stable limit cycles by turning the source in the center of the limit cycle into a sink [118, 79]. Homoclinic saddle-node bifurcations, otherwise known as infinite period bifurcations, create a saddle-node bifurcation along the limit cycle which eliminates the cycle [118, 79, 103]. Homoclinic bifurcations occurs when a limit cycle merges with a saddle point creating a homoclinic orbit. Saddle-node bifurcations of periodic orbits merge concentric stable and unstable limit cycles [197].

Feed-forward control may be able to create some of these bifurcations in a given system. We can control the creation and elimination of limit cycles in the dynamical system by using Dulac's criterion and the Poincare-Bendixson theorem to map out control regions that will create and eliminate limit cycles.

3.3.5 Extension to Three-dimensional Systems

This analysis can be extended to create control procedures for three-dimensional systems, $\mathbf{x}' = F(\mathbf{x}) + \mathbf{u}(t)$, where $\mathbf{x}, \mathbf{u} \in \mathbb{R}^3$.

$$\frac{dx}{dt} = f(x, y, z) + u_1(t) \tag{3.19}$$

$$\frac{dy}{dt} = g(x, y, z) + u_2(t) \tag{3.20}$$

$$\frac{dz}{dt} = h(x, y, z) + u_3(t). \tag{3.21}$$

Analogous to the two-dimensional case, we compute the Jacobian of the system $J(x, y, z)$ and fixed point locations $u_1 = -f(x, y, z)$, $u_2 = -g(x, y, z)$, $u_1 = -h(x, y, z)$. We use the linearized system in conjunction with the control signal equations to determine the control surface $(u_1(t, s), u_2(t, s), u_3(t, s))$ along which bifurcations occur. While the stability of fixed points can still be determined in three-dimensional systems, the locations of strange attractors become impossible to determine analytically as Dulac's criterion and the Poincare-Bendixson theorem only apply to two-dimensional systems [197]. Therefore, while we can still stabilize and destabilize fixed points in order to execute transitions, it is unknown where and when attractors will occur in the system, thus potentially compromising the control procedure advocated here. In these circumstances, attractors and their bifurcations are best found experimentally by simulating the dynamics. Using data, Poincaré maps can be found via the SINDy algorithm and from these discovered maps the stability of nonlinear periodic orbits can be determined [19].

3.4 Network control applications

Our feed-forward control procedure can be directly applied to nonlinear systems that consist of two or three variables. While most real world systems realistically involve a large number of variables, many canonical models in chemistry, biology, and the social sciences are formulated using only a few variables. In chemistry the Brusselator describes an autocatalytic chemical reaction with two reagents [167, 166, 197] and the Lokta-Volterra model in ecology describes

predator-prey interactions at the population level between two species [131, 132, 211]. In epidemiology the SIR model describes the spread of an infectious disease by modeling the interactions between susceptible and infected individuals [106, 83] and in neuroscience the FitzHugh–Nagumo model describes spike generation in neurons by measuring membrane voltage and a recovery variable [64, 153]. In the social sciences the classical model of political economy uses capital and labor as variables [178] while Richardson's arms race model describes the interactions between two nations stockpiling nuclear weapons [173, 187]. If the dynamics of a low-dimensional system are unknown, they can be found using the SINDy algorithm [25, 24], allowing us to apply control even to novel systems. We demonstrate the effects of feed-forward control on two variable nonlinear systems using simple chemical reaction models as examples.

3.4.1 Chemical Reaction with Bistability

The following is a minimal example of a chemical reaction with bistability using parameter values $k_1 = 8$, $k_2 = 1$, $k_3 = 1$, and $k_4 = 3/2$ from [219]. We can find control signals that will move the system between stable states by computing stability regions from the control signal bifurcation curves. The governing equations are given by

$$\frac{dx}{dt} = 16y - x^2 - xy - \frac{3}{2}x + u_1 \tag{3.22}$$

$$\frac{dy}{dt} = x^2 - 8y + u_2 \tag{3.23}$$

where the Jacobian, trace and determinant are given by

$$J(x,y) = \begin{bmatrix} -2x - y - \frac{3}{2} & 16 - x \\ 2x & -8 \end{bmatrix} \tag{3.24}$$

$$T(x,y) = -2x - y - \frac{19}{2} \tag{3.25}$$

$$D(x,y) = 2x^2 - 16x + 8y + 12. \tag{3.26}$$

The curve for $T(x, y) = 0$ is

$$u_1^h(t) = -t^2 + 24t + 152 \tag{3.27}$$

$$u_2^h(t) = -t^2 - 16t - 76 \tag{3.28}$$

The curve for $D(x, y) = 0$ is

$$u_1^s(t) = -\frac{1}{4}t^3 + 7t^2 - 32t + 24 \tag{3.29}$$

$$u_2^s(t) = -3t^2 + 16t - 12 \tag{3.30}$$

In the uncontrolled system the chemical reaction has two stable fixed points, Fig. 3.3(a). Figure 3.3(b) shows the stability regions of the fixed points in the control space. Across the blue curve fixed points exhibit saddle-node bifurcations while across the red curve fixed points exhibit Hopf bifurcations. This map shows us that we can eliminate the second fixed point via a saddle-node bifurcation while keeping the first fixed point stabilized and then destabilize the first fixed point via a Hopf bifurcation which creates a stable limit cycle surrounding the first fixed point. The overlay of stability regions shows us the wide variety of states that can be achieved by simply adding or removing certain quantities of reactant.

3.4.2 Brusselator

The Brusselator describes a type of autocatalytic chemical reaction. We take $a = 1$ and $b = 3$ from the general model [197]

$$\frac{dx}{dt} = 1 + x^2y - 4x + u_1(t) \tag{3.31}$$

$$\frac{dy}{dt} = 3x - x^2y + u_2(t), \tag{3.32}$$

and find the bifurcation curve locations.

The Jacobian of the system is

$$J(x, y) = \begin{bmatrix} 2xy - 4 & x^2 \\ 3 - 2xy & -x^2 \end{bmatrix} \tag{3.33}$$

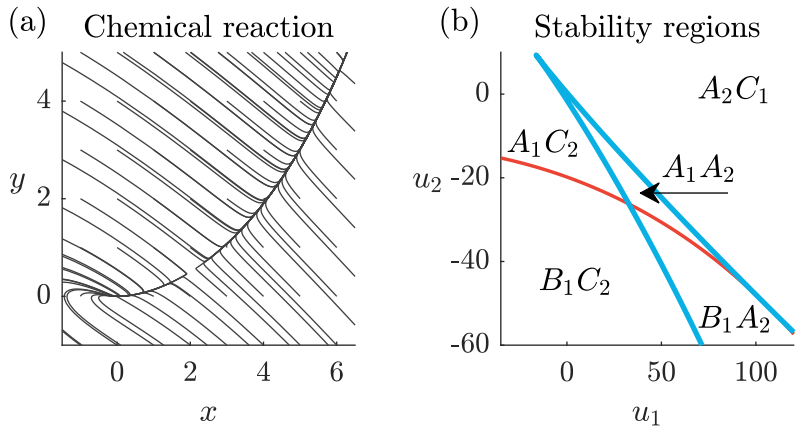

Figure 3.3: Chemical reaction with bistability. (a) In the absence of control signals there are two stable fixed points. (b) Regions of stability and instability for the fixed points in the uncontrolled system. The system has regions with two stable fixed points, one stable fixed point, a stable fixed point and a stable limit cycle, and only a stable limit cycle. Note the stable limit cycle appears surrounding the unstable source.

with trace and determinant given by

$$T(x,y) = 2xy - x^2 - 4 \tag{3.34}$$

$$D(x,y) = x^2. \tag{3.35}$$

We find that $T(x,y) = 0$ along the curve

$$u_1^h(t) = \frac{-t^3}{2} + 2t - 1 \tag{3.36}$$

$$u_2^h(t) = \frac{t^3}{2} - t. \tag{3.37}$$

Note that $\hat{T}(x,y) \to \pm 0$ when $y = c/x^2$ and $x \to \pm 0$ which produces the curve

$$u_1^h(t) = -1 - t \tag{3.38}$$

$$u_2^h(t) = t \tag{3.39}$$

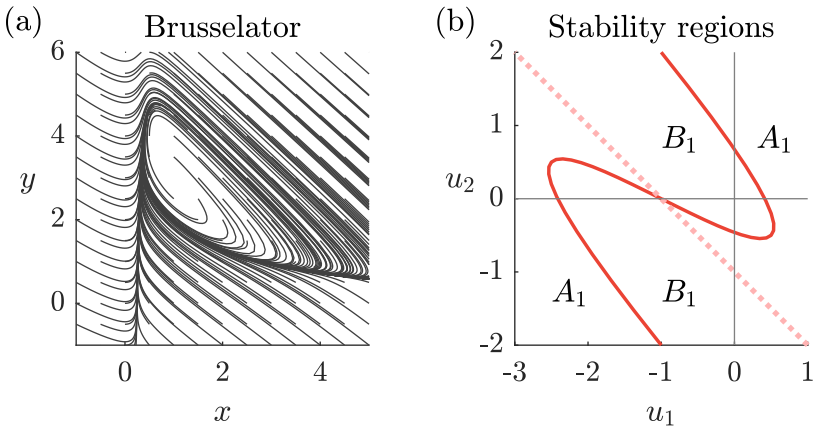

Figure 3.4: Brusselator (a) Brusselator with $u_1, u_2 = 0$. (b) Stability regions in the control space. The system transitions between having a single stable fixed point and an unstable source encapsulated by a stable limit cycle.

The determinant is always greater than zero and therefore does not produce a curve over which there is a sign change. This means that feed-forward control signals that induce saddle-node bifurcations do not exist in this system.

Figure 3.4(a) shows the uncontrolled Brusselator contains a stable limit cycle as also indicated by the control signal stability regions, Fig. 3.4(b). Across the Hopf bifurcation curve $C^h(t)$ the globally stable system transitions between having a single stable fixed point and a single source surrounded by a limit cycle, Fig. 3.4(b). The Brusselator can be compelled to move between a stable state and a limit cycle by adding or removing particular amounts of reactant.

3.5 Feed-forward control for high-dimensional nonlinear systems

So far we have considered control for systems with only two or three variables. While some systems fit this description, most real systems involve a much larger number of variables and interactions. We can use dimensionality reduction in conjunction with data-driven discovery

Algorithm 1: Control for high dimensional systems

Measure system outputs: $\mathbf{X}$

Dimensionality reduction: $\mathbf{USV}^*$

Identify nonlinear system: $\mathbf{x}' = F(\mathbf{x})$

Determine control signals: $\mathbf{x}' = F(\mathbf{x}) + \mathbf{u}_F(t)$

Determine high dimensional signals: $\mathbf{u}_G(t) = \hat{\mathbf{U}}\mathbf{u}_F(t)$

of nonlinear dynamics to extend this control technique to high-dimensional systems. Consider an n-dimensional nonlinear system of the form

$$\mathbf{x}' = G(\mathbf{x}) + \mathbf{u}_G(t) \tag{3.40}$$

where the system variables $\mathbf{x} \in \mathbb{R}^n$ are controlled by the feed-forward control signal $\mathbf{u}_G \in \mathbb{R}^n$ and the form of G may be unknown. We detail a method for finding $\mathbf{u}_G(t)$ from system measurement data, as outlined in Algorithm 1. The steps of the algorithm are detailed as follows:

Measure system outputs

We begin by collecting timeseries data from variables in the high-dimensional system. Data should be collected from many different initial conditions in the entire state space in order to generate an accurate model.

Dimensionality reduction

Next we dimensionality reduce the timeseries data using the SVD. Not all high-dimensional systems have a low-dimensional structure and those that do have a low-dimensional structure may not necessarily have a *linear* low-dimensional structure. SVD is a suitable dimension reduction technique for a given dataset if the first two or three modes capture the majority of the variance in the system. While there is no definite cutoff as to what defines a sufficient amount of variance, issues arise when most of the variance is not allocated to the primary

modes. The low-dimensional trajectories may not satisfy the existence and uniqueness theorem and data that displays many non-unique solutions cannot be adequately characterized by a dynamical system [197]. Even if the trajectories of the primary modes can be approximated well with a dynamical system, the model may still not provide an adequate representation of the original system if later modes capturing high amounts of variance are not considered. The consequence is that the effects of the control signal predicted by the model may diverge from the control signal's true effects on the system.

Identify the nonlinear system using SINDy

Once we have reduced the data to only a few dimensions we can fit a nonlinear dynamical system to the data using the SINDy algorithm [25]. Choosing basis functions can be a difficult task and highly dependent on the system. With polynomials as basis functions, one can capture the local Taylor series approximation for the underlying function if it is smooth. For systems with periodic dynamics or singularities, however, periodic or rational basis functions are a better choice.

If the original system is globally stable the model should also be globally stable. We can determine the stability of the SINDy model using the criteria outlined in Section 3.3.3. In order for all important structures, such as stable attractors, to be represented in the model, samples from the entire region of interest, and samples that converge to different attractors must be included in the regression. The SINDy algorithm can only be expected to generate accurate approximate dynamics within the sample region; dynamics cannot be extrapolated.

Determine feed-forward control signals

Now that we have a low-dimensional nonlinear model for the system's activity, we can generate feed-forward control signals that stabilize fixed points and transition the system between attractors in the SINDy model. The time needed to transition from one fixed point to another is dependent on the amount of stability about the fixed points in the controlled system. Control signals that technically destabilize fixed points yet exist close to bifurcation curves

will generate center manifolds which may take a long time to escape, despite the state being unstable. In contrast, destabilizing control signals that exist far from bifurcation curves will not produce time delays in the system's transition between states yet may in practice require more energy. The time needed for each transition can be determined experimentally from the SINDy model. These control signals must go through a last transformation before they can be applied to the original system.

Determine high dimensional control signals

Finally, we use the feed-forward control signals $\mathbf{u}_F(t)$ discovered for the low-dimensional system to control the original system by projecting the signals back to the original high-dimensional space via the first two SVD modes $\hat{\mathbf{U}}$.

$$\mathbf{u}_G(t) = \hat{\mathbf{U}}\mathbf{u}_F(t) \tag{3.41}$$

$$\mathbf{x}' = G(\mathbf{x}) + \mathbf{u}_G(t). \tag{3.42}$$

Developing control methods in a reduced space has been effective in other settings [24]. Dynamic mode decomposition with control (DMDc) finds spatial-temporal coherent modes with which to construct low-order models that incorporate the effects of control signals [169, 115]. Unsteady wake flows can be effectively described and controlled using POD modes [154]. Structural balance dynamics in social networks and the bifurcations that occur in the system can be analyzed in a low-dimensional eigenspace [137, 147]. Similar to previous methods, we use linear dimension reduction, but unlike previous methods we employ nonlinear, feed-forward control.

3.6 High-dimensional control applications

High-dimensional systems that exhibit low-dimensional dynamics are pervasive in nature. The nematode *C. elegans* has a network of 302 neurons yet exhibits neural activity that exists on a three-dimensional manifold [102, 62]. The low-dimensional encoding of information is a common motif found throughout the neuroscience literature [163, 20, 134, 195] as well as

in the study of artificial neural networks [60, 172, 84, 91, 48, 174]. In the social sciences, social networks and structural balance dynamics both exhibit low-dimensional structures [14, 200, 137, 147]. In biology and epidemiology the dimension of population models can be reduced through mathematical aggregation methods [8]. In molecular biology the dynamics of biological networks exhibiting multiple timescales can be simplified by modeling the system with multiple timescale networks [215] and reduced, hierarchical structured models can describe the dynamics of complex signal transduction networks [42]. We demonstrate our control procedure on several nonlinear high-dimensional systems that exhibit dynamics on a low-dimensional manifolds.

3.6.1 Random High-dimensional Networks

We first illustrate our feed-forward control technique on a randomly generated high-dimensional network dynamical system [Appendix A.3]. We wish to find stable attractors endogenous to the system and then generate control signals that will move the system between stable states. Figure 3.5(a) shows data collected from our high-dimensional random network. The trajectories in the underlying low-rank SVD/PCA space indicates that there are two stable fixed points in the system. Figure 3.5(b) shows the system is indeed low-dimensional as indicated by the rapid decay in singular values in Fig. 3.5(c). The SINDy algorithm finds a model that represents the low-dimensional system and which can be used to derive control signals Fig. 3.5(d-e).

We next use the SINDy model to find the control signal curves along which bifurcations in the system occur, $C^h(t)$ and $C^s(t)$. Figure 3.6 shows the locations of fixed points and control signals in the system when $T(x, y) = 0$. Figure 3.6(a-b) shows the location of a saddle fixed point as it switches types colored by the control signal values that induce these bifurcations. Figure 3.6(c-d) shows these control signal values colored by fixed point locations. The control signals move from stability region C to D as they cross the bifurcation curve.

Figure 3.7 shows the locations of fixed points and control signals in the system when $D(x, y) = 0$. Figure 3.7(a-b) show the locations of the stable fixed points as they go through

saddle-node bifurcations colored by the control signal values that induce these bifurcations while Figure 3.7(c-d) show these control signal values colored by fixed point locations. The stable fixed points start in stability region A and then disappear as they cross the bifurcation curves. $C^s(t)$ defines stability region borders which give us the ability to select control signals that will induce transitions between stable states.

We can now induce transitions between stable states in the system using transient feed-forward control signals selected from the generated stability regions. We first stipulate our objective path or what can be thought of as setpoints over time Fig. 3.8(a) and then select control signals that will induce each transition Fig. 3.8(b). When the control signals are applied to the original system it makes the desired transitions as viewed in the PCA space Fig. 3.8(c-d) as well as the high-dimensional space Fig. 3.8(e-f). The system does not behave exactly as the SINDy model predicts as the model is only an approximation of the system's true dynamics.

3.6.2 Hopfield Networks

Hopfield networks are auto-associative memory networks that converge to stable patterns in the presence of noise and are a model of memory retrieval in the human brain [87, 150]. The Heaviside step function is a key component of the Hopfield model; it is the nonlinearity in the system that allows the Hopfield network to have many fixed points as opposed to one. Because the Heaviside function is a dimension reducing function it makes the dynamics of the Hopfield model much lower dimensional than the network's original size, making it a promising candidate for our control procedure.

We test our control procedure on the dynamics of a Hopfield network. We outline the construction of this network in Appendix A.4. Figure 3.9 shows timeseries data collected from a Hopfield network storing four memories, the trajectories in PCA space, and the SINDy model approximating the system. The rapid decay in singular values shows that the system in PCA space is a good representation of the original data, Fig 3.9(c). Figure 3.10 shows the saddle-node bifurcation curves $C^s(t)$ colored by fixed point locations. These curves indicate

the stability regions for the system's four memories in control space. Figure 3.11 shows the Hopfield network transitioning between fixed points in the manner specified by the objective path. Transient control signals are selected from the stability regions for each transition and applied to the original system through the first two PCA modes. We are able to move directly from any stable fixed point to any other stable fixed point in the system according to the stability regions. Furthermore the system does not exhibit any limit cycles, making it an easy system to control with a feed-forward control signal method.

3.6.3 Systems with Three-dimensional Intrinsic Dynamics

Only some systems can be adequately described using the first two PCA modes. Many systems have higher-dimensional dynamics that require the use of three or more modes. We demonstrate our control procedure on systems with three-dimensional intrinsic dynamics.

Three-dimensional Hopfield

We inflate the intrinsic dimension of our Hopfield network by increasing the number of memories stored [Appendix A.4]. Figure 3.12 shows the dynamics of an intrinsically three-dimensional Hopfield network. We use SINDy to fit a three-dimensional nonlinear dynamical system to the data and control the system along three dimensions, $(u_1(t), u_2(t), u_3(t))$. While control signals can be selected from analytically derived three-dimensional stability regions, it can be easier in practice to find control signals experimentally by systematically perturbing the SINDy model with control signals to find control regions that will destabilize each fixed point.

Three-dimensional Random Dynamical System with a Strange Attractor

We generate a random dynamical system that has intrinsic three-dimensional dynamics and use control signals to move the system between a fixed point and a strange attractor in the system. Figure 3.13(a) shows the SINDy model approximation of the system which captures

the general location and stability of the strange attractor and two stable fixed points. While the location of the fixed point and attractor appear to be adequately captured by the SINDy model, Fig. 3.13(b), the model predicts an oscillation frequency for the strange attractor that is much slower than that observed in the original system, Fig. 3.13(c-d), meaning that while the model may be able to predict whether the system's state is on or off the strange attractor it cannot predict its location along the attractor. Stability regions for the fixed points can be analytically computed; however, the stability region for the strange attractor cannot be analytically computed and therefore must be experimentally determined by measuring the effects of control perturbations to the SINDy model.

3.7 Discussion

Our control procedure shifts nonlinear network dynamical systems between attractors using feed-forward control signals derived from the system equations. This procedure works for high-dimensional systems that exhibit low dimensional dynamics by developing a control procedure for the system in a reduced space and then projecting the derived control signals back to the original space.

Our framework builds off of previous work in which we used a feed-forward control model to demonstrate how the nematode *C. elegans* may switch between short-term behavioral states using externally generated control signals [146]. We observed that the modulation of internal system parameters was another form of control that the nematode could use to implement long-term behavior changes. Our framework also builds off previous pulse-based control methods for switching between stable states in biological dynamical systems [191, 192, 190, 209, 43].

While we have focused on feedforward control, this method could be used in conjunction with feedback control mechanisms by considering feedforward control applied to a system modified by feedback $F_\alpha(\mathbf{x})$, where α is the feedback parameter. Feedfoward control capabilities could be computed and compared for a range of α feedback control values and then α could be selected to render the system most controllable via feedforward control.

While many traditional control strategies used in a wide range of fields can be equivalently viewed as actuating forces applied to nonlinear systems, some control strategies involve modifying internal system parameters or using system feedback, meaning that our control framework does not account for all ways control can be achieved in a network.

In pharmacology [104], certain agonists can be modeled as feed-forward control signals applied to dynamics on a biochemical network while antagonists may be better represented as shifting network parameters. In clinical neuroscience, deep brain stimulation is used as a form of external control to correct motor output in patients with Parkinson's disease and essential tremor as well as to reduce pain in patients with chronic pain [15, 162, 109]. In an alternative setting, deep brain stimulation is used to alter brain structure in patients with treatment-resistant depression and traumatic brain injury [140, 181]. This same medical intervention has the ability to both act as a transient feed-forward control signal to the neural system as well as alter internal structures highlighting how a single control signal applied to a system can have diverse effects.

Network control is also widely implemented in the political and social sciences. Governments desire to optimize economic performance through monetary and fiscal policy [67, 68, 7]. The central bank aims to achieve economic goals by adjusting interest rates, which modify internal economic parameters. Meanwhile, the federal government pursues these same goals through taxation and government spending which are more easily viewed as feed-forward control strategies. States often provide external support to insurgent movements in order to further their own interests [30]; these sponsorships can have a variety of effects on the political network including increasing the power of particular groups or modifying the way groups interact. During an epidemic, public health experts have various ways of suppressing the spread of an infectious disease. They can achieve public health goals by modifying people's social interactions, which changes interaction parameters in infectious disease models, or by introducing external forces such as vaccinating susceptible individuals or quarantining infected ones [106, 83]. Given the limited avenues of control and leverage available to health care professionals, policy makers, and organization leaders, all possible forms of control are

often implemented concurrently. While this has the potential to maximize influence on the system in question, it also makes identifying the effects of multiple control strategies difficult to differentiate.

Sometimes the systemic changes necessary to bring about a desired outcome are either unknown or not possible to implement in a given network while the use of external control presents a clear path to success. We present a feed-forward control strategy as an alternative form of control for nonlinear network systems that are unamenable to feedback control or alterations to system parameters.

One limitation of this control strategy is that the locations and stabilities of limit cycles and strange attractors in the system cannot be analytically determined, particularly for systems with more than two dimensions. This limits our ability to assert control guarantees and means that the stability regions of some attractors must be determined experimentally as they cannot be determined analytically. We aim to extend this work by defining the locations and stabilities of limit cycles and attractors under the influence of feed-forward control by using data-driven discovery methods on Poincare maps as presented in [19].

A practical limitation of this method is that control signals in the original space will most likely be nonsparse if using regular PCA; actuating a majority of variables in many systems is not feasible. This issue may be overcome in some cases by instead performing sparse PCA [223] on the system and then selecting low-dimensional control signals such that $u_1 = 0$ or $u_2 = 0$. While enforcing increasingly sparse constraints on the principal components will reduce the number of variables that require actuation, this comes with a tradeoff — forcing the principal components to be sparse may erode the amount of variability captured by them which in turn makes the model less representative of the true system.

A third limitation of this method is that it can only be applied to high-dimensional systems that can be linearly dimension reduced. Many high-dimensional systems exhibit low-dimensional activity but must be reduced using a nonlinear dimension reduction method. We propose extending feed-forward control to such systems by using autoencoders to perform the nonlinear dimensionality reduction [35] and then transforming the low-dimensional control

signals using the autoencoder. Feed-forward control of nonlinear systems lends itself to many developments including optimization, extensions to higher dimensions, and hybridization with feedback control methods.

To provide a more specific context of the application of network control, we consider some commonly observed inhibitory networks in neuroscience which typically produce winner-take-all dynamics. For instance, experimental evidence from olfaction processing [94, 171, 184, 119] suggests that two key features assumed in our analysis are present in such systems: (i) low-dimensionality and (ii) multiple fixed points. In the odor space discovered from data, odor stimulus plays the critical role of actuation in the system. Thus, application of known odorants in insect olfaction guide the high-dimensional neuronal activity to a fixed point in a low-rank subspace of neural population codes. The modeling framework provided here can be used to understand how such actuations produce the important pathways and trajectories responsible for producing a cognitive recognition, or decision, about the odor applied. Moreover, it highlights the underlying control mechanisms that would need to be enacted by the system to produce a stable decision making process for odor identification.

3.8 Conclusion

Nonlinear network dynamical systems in the physical, engineering, biological and social sciences are typically difficult to characterize and control. We develop a feed-forward control procedure for low-dimensional nonlinear systems by deriving control signals as a function of local bifurcations in the system. We extend this method to high-dimensional systems with unknown dynamics by using dimensionality reduction in conjunction with the model-discovery SINDy algorithm. Our technique is demonstrated on canonical and random network dynamical systems of different dimensions with known and unknown dynamics. We propose this method of control as an alternative to feedback control and as a framework for understanding how actuating forces can be used in nonlinear systems to shift systems between stable states.

To our knowledge, our method provides a principled mathematical architecture that integrates dimensionality reduction, bifurcation theory, and data-driven discovery of dynamics

to construct a feed-forward control model for nonlinear, networked dynamical systems. Our feed-forward control techniques can be interpreted and used to regulate the dynamics of networked dynamical systems by exploiting the dominant, low-dimensional subspaces on which the dynamics evolves. Our mathematical architecture generates a set of actuation signals that, when applied, are able to control the original high-dimensional system. Using bifurcation theory, we find collections of feed-forward control signals that will force convergence to desired objective states, allowing us to move the system from one fixed point of the system to another in a principled manner. Specifically, we can destabilize a given fixed point by making it undergo a saddle node or Hopf bifurcation, while simultaneously making the target fixed point an attractor. This creates a pathway with the feed-forward signals from one fixed point to another.

The potential applications of this control framework are numerous. From neuroscience to powergrids, networked dynamical systems appear in almost every branch of quantitative study. Many modern systems of study are high-dimensional and characterized by nonlinear dynamics, thus they require new mathematical methods to characterize their behavior and exploit their intrinsic dynamics. The methods developed here rely on the simple observation that most high-dimensional dynamical systems manifest low-dimensional dynamics. Thus the low-dimensional subspaces on which the system evolves can be exploited for control, as is done in this manuscript. The number of application areas for which this is true is quite diverse and extensive, including neuroscience [113, 102, 62, 94, 171, 184, 47], powergrids [55], disease modeling [106, 83] and social networks [147].

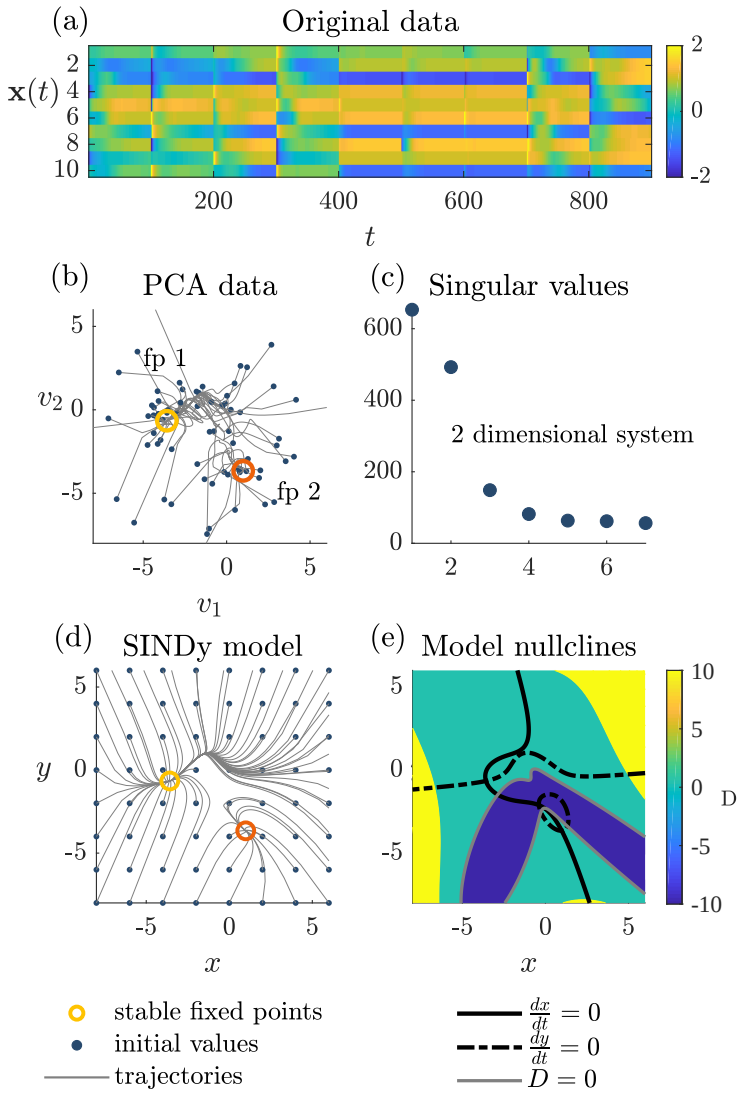

Figure 3.5: Low-dimensional model of a high dimensional random dynamical system. (a) High dimensional system $n = 10$ visualized as a timeseries and (b) shown in PCA space. (c) The first two PCA modes capture a reasonable amount of the data and the SINDy model (d) finds the two stable fixed points. (e) Nullclines and determinant of the SINDy model.

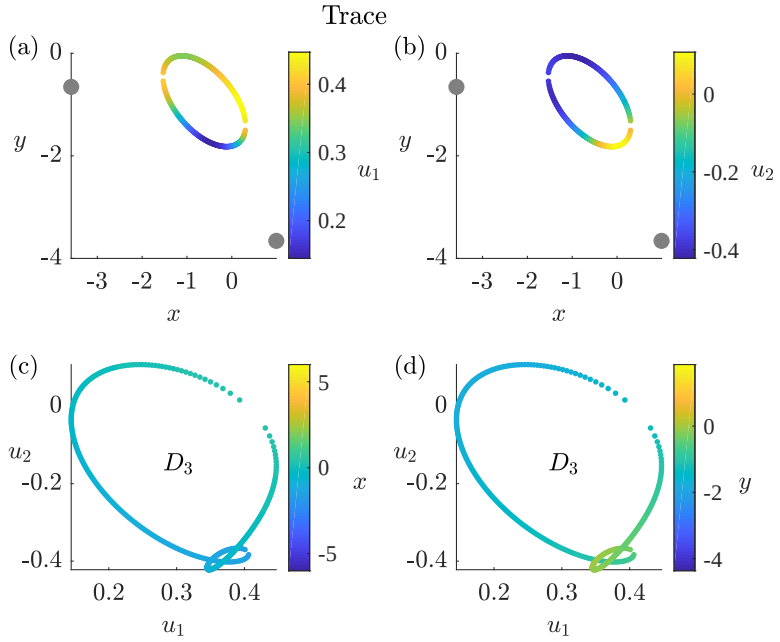

Figure 3.6: Fixed point locations and control signal values along the curve $C^h(t)$ for the low-dimensional model found in Figure 3.5. (a-b) Fixed point locations in the xy-plane for $C^h(t)$ colored by the u_1, u_2 control signal values along $C^h(t)$. (c-d) Control signal locations in the u_1, u_2 plane for $C^h(t)$ colored by (x, y) fixed point location values. The fixed point in this model that transitions across the curve $T = 0$ is the saddle fixed point that sits between the stable fixed points in the uncontrolled model. This saddle fixed point originally is located in region C_3 but transitions to region D_3 across the curve $C^h(t)$.

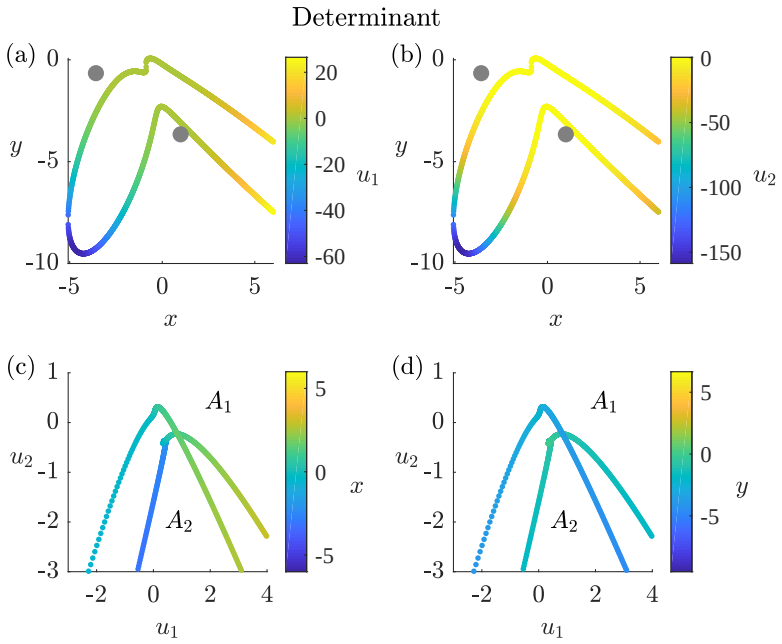

Figure 3.7: Fixed point locations and control signal values along the curve $C^s(t)$ for the low-dimensional model found in Figure 3.5. (a-b) Fixed point locations in the xy-plane for $C^s(t)$ colored by the u_1, u_2 control signal values along $C^s(t)$. (c-d) Control signal locations in the u_1, u_2 plane for $C^s(t)$ colored by (x, y) fixed point location values. The fixed points in this model that transition across the curve $D = 0$ are the stable fixed points in the uncontrolled model. The right stable fixed point goes through a saddle-node bifurcation along the top curve, while the left stable fixed point goes through a saddle-node bifurcation along the bottom curve.

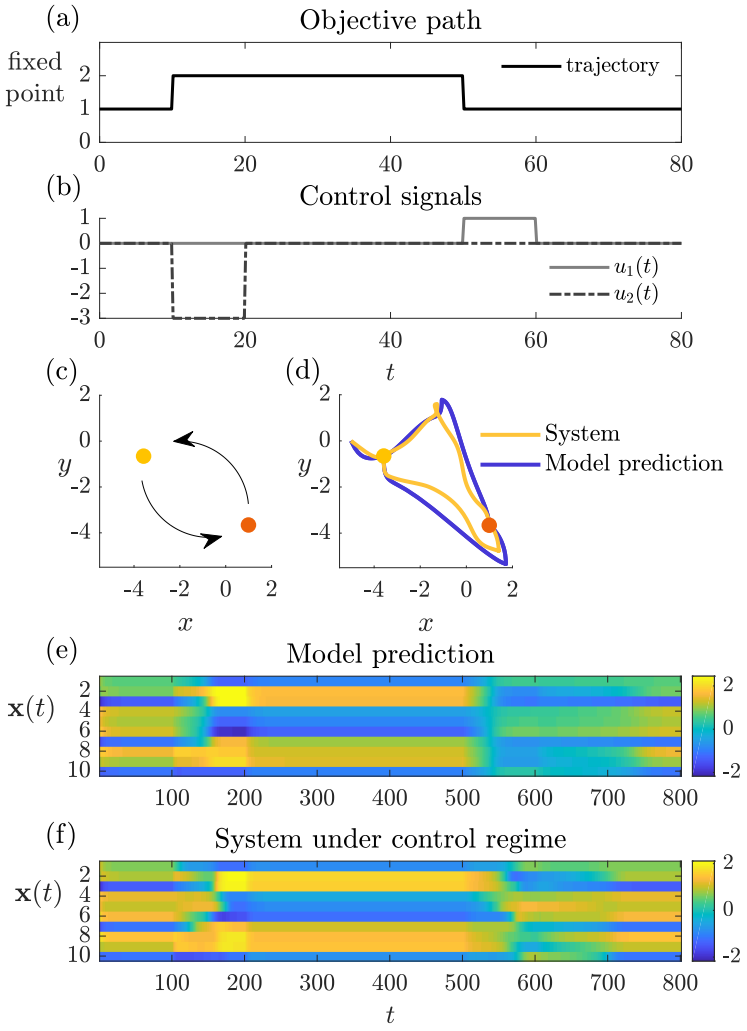

Figure 3.8: Control example for the random system in Figure 3.5. (a) Objective path for the system. (b) Control signals selected using the system's stability region maps to move the system between fixed points. (c) Objective path in PCA space. (d) Predicted and actualized system paths in PCA space. (e-f) Predicted and actualized system activity in the original high-dimensional space.

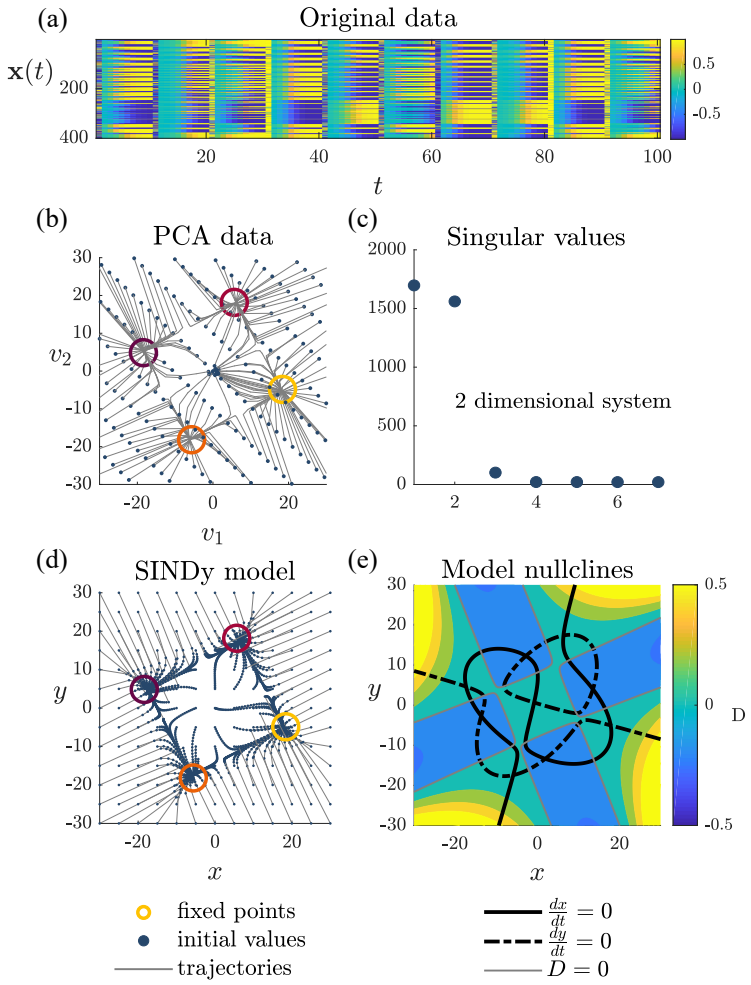

Figure 3.9: Smooth Hopfield network size $n = 400$ with intrinsic 2-dimensional structure. (a) High-dimensional data measured from many initial conditions. (b) Data in PCA space. The network converges to one of 4 "memories". (c) The first 2 modes capture the majority of the variance in the system. (d) SINDy is able to generate a low-dimensional model with similar dynamics that we can use to generate stability regions. (e) Nullclines and determinant of the SINDy model.

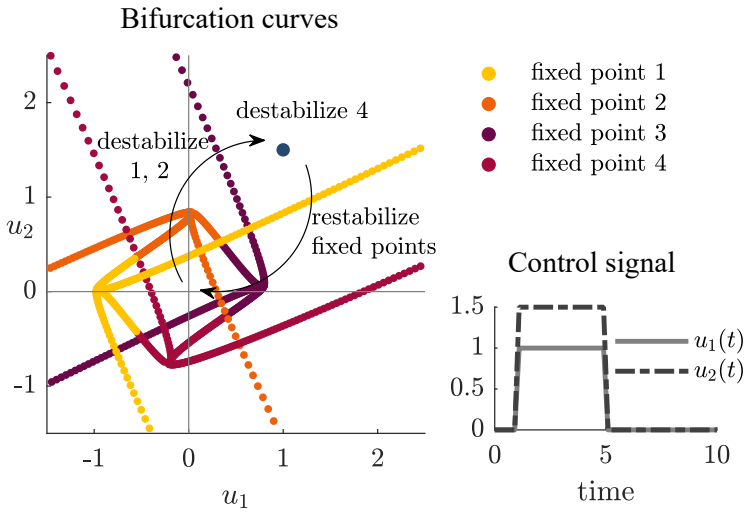

Figure 3.10: Saddle-node bifurcation curves $C^s(t)$ colored by fixed point for the system in Figure 3.9. Moving across the saddle-node bifurcation curve of a stable fixed point either eliminates it from the system or reinstates it.

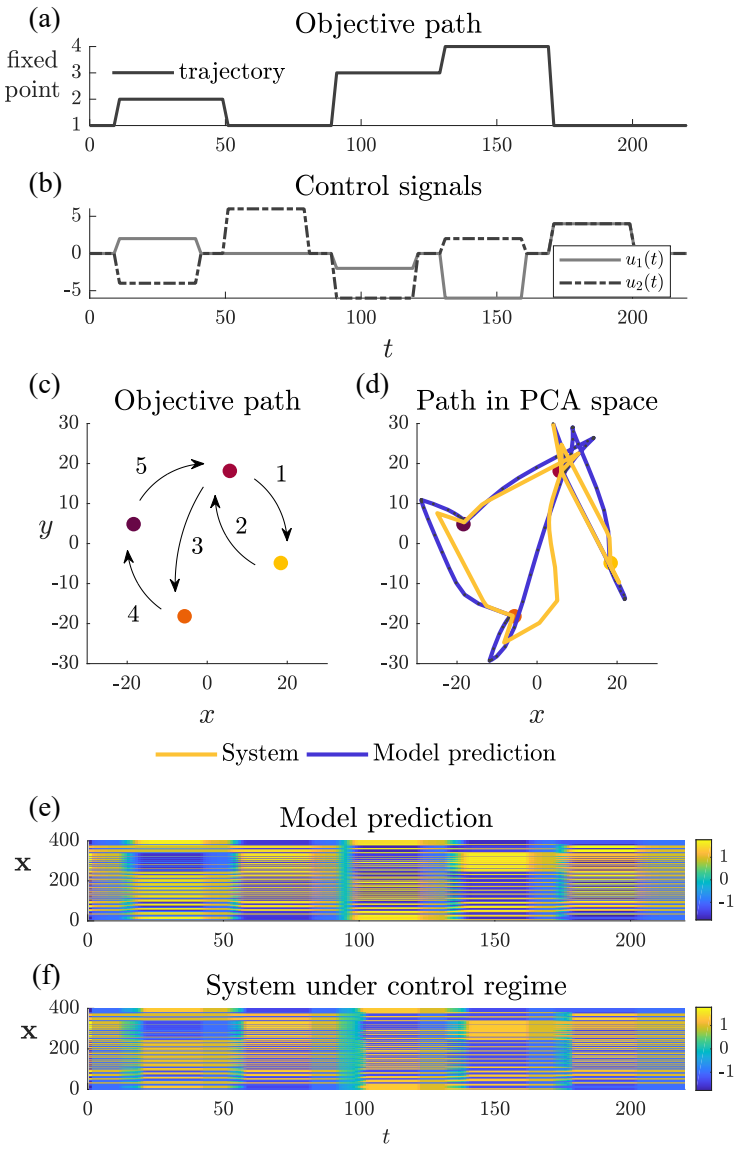

Figure 3.11: Control example for the hopfield system in Figure 3.9. (a) Objective path for the system. (b) Control signals selected using the systems's stability region maps to move the system between fixed points. (c) Objective path in PCA space. (d) Predicted and actualized system paths in PCA space. (e-f) Predicted and actualized system activity in the original high-dimensional space.

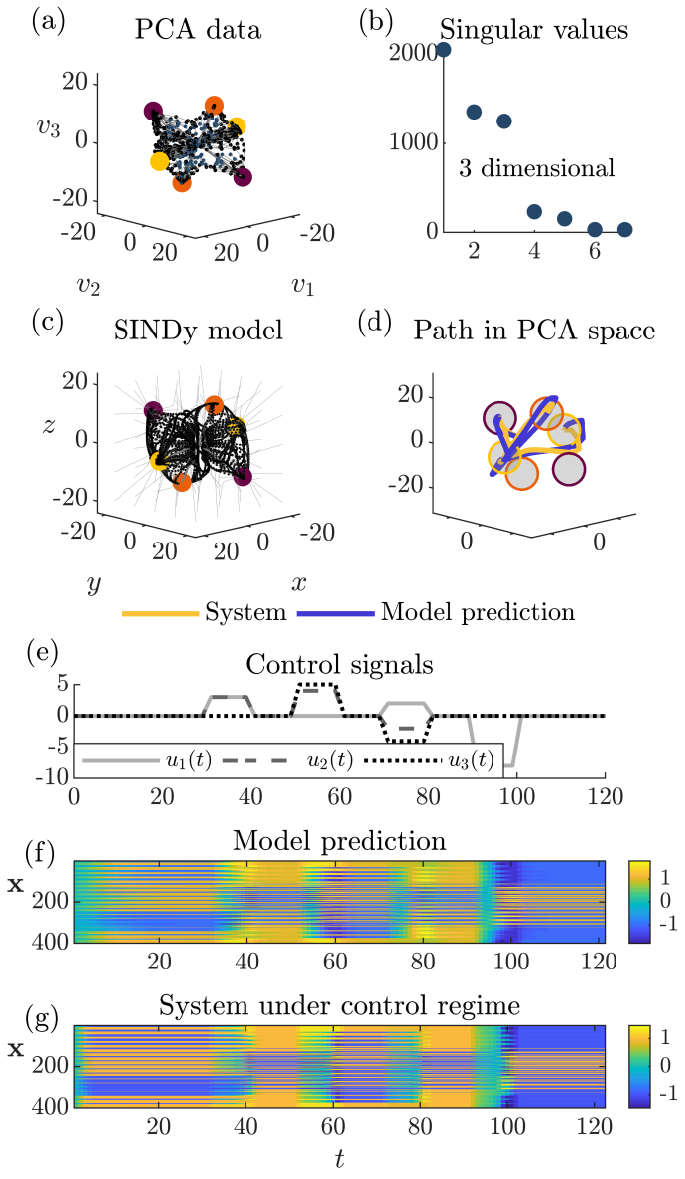

Figure 3.12: Smooth Hopfield network size $n = 400$ with intrinsic 3-dimensional structure. (a) Hopfield data in PCA space. The network converges to one of 6 "memories". (b) The first 3 modes capture the majority of the variance in the system. (c) SINDy is able to capture a low-dimensional model with similar dynamics that we can use to construct a control regime. (d) Predicted and actualized network activity in PCA space. (e) There are now 3 control signals in the control regime as we are using a 3-dimensional system. (f-g) Predicted and actualized network activity in the original space under the specified control regime.

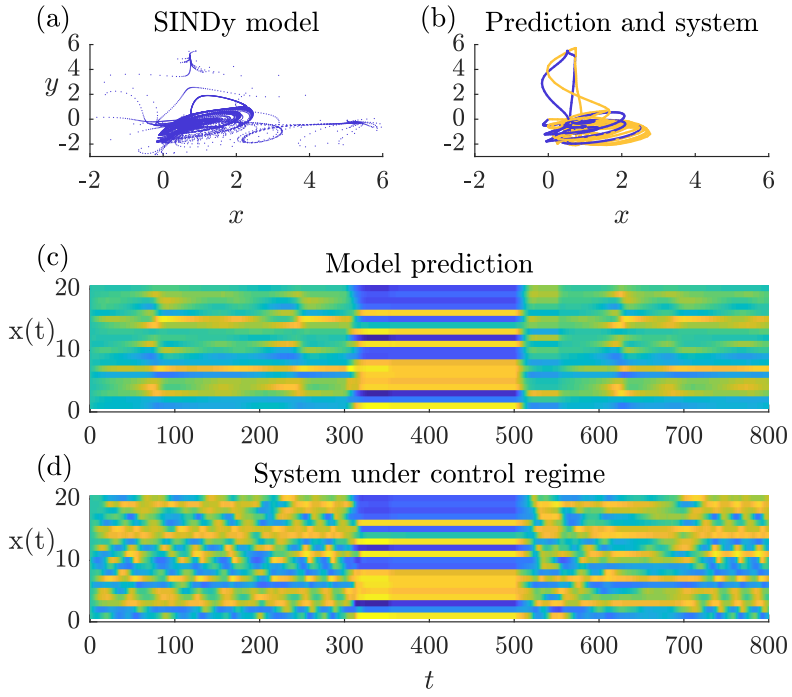

Figure 3.13: Randomly generated high-dimension system $n = 20$ with a strange attractor. (a) SINDy model with 3 variables captures the stable fixed points and strange attractor in the system. (b) Predicted and actualized network activity in PCA space under control. (c) Predicted and actualized network activity in the high-dimensional space. Notice that although the SINDy model captured the location of the strange attractor, it does not adequately capture its frequency as the model predicts a much slower oscillation that the original system actually produces.

Chapter 4

COMMUNITY DETECTABILITY AND STRUCTURAL BALANCE DYNAMICS IN SIGNED NETWORKS

4.1 Introduction

[1] Most research in network science has focused on networks that allow only positive ties. In signed networks, however, ties can take on negative values as well. In social systems, positive ties signify friendly or cooperative relationships between the individual or collective actors represented by the nodes whereas negative ties signify hostile or conflictual relationships between nodes. As examples, signed social networks have been used to represent interpersonal sentiments among students [107], supportive or critical references among opinion makers [22], relationships in online social networks [59], and alliances and military clashes among nations [136, 51].

In this paper, we address community structure in signed networks and its implications for dynamics governed by structural balance, a theory commonly invoked in treatments of signed networks in social systems. In unsigned networks, community structure refers to the presence of clusters within networks characterized by relatively dense intra-cluster ties and sparse inter-cluster ties. A rich set of techniques have been developed to detect communities in unsigned networks [66]. Of particular relevance here, spectral analysis has proven to be a highly valuable tool for probing community structure [157, 36]. For signed networks, the notion of community can be extended to accommodate negative ties by reversing the criteria for positive ties — there should be relatively sparse negative ties within communities and denser ties between them. At the present, however, the literature on community detection in signed networks is itself rather sparse in comparison with unsigned networks [205, 112, 22, 58].

---

[1]

An important phenomenon of community structure in unsigned networks is that of community detectability [46, 152, 222, 74, 220, 1]. Here, community structure can be present — in the sense that the tie generating probabilities in a stochastic block model indeed favor ingroup over outgroup ties — but it is too weak to typically be discerned by analysis of the generated network. For large networks, a phase transition characterizes the passage from undetectable to detectable structure.

We show that detectability transitions also occur in signed networks. We generate our networks using a stochastic block model for two communities in an unweighted and undirected signed network (Sec. 4.2). Examples of simulated networks with community structure that is detectable and undetectable are shown on the left in Fig. 4.1(a) and (b). We describe the transitions observed in the spectra of simulated networks in which outlying eigenvalues corresponding to meaningful signals merge with the main spectral band corresponding to noise (Sec. 4.3). Two sets of spectral transitions are found: one corresponds to the detectability transition involving the two-community structure, while the other affects the ability to observe an overall tendency toward positive or negative tie formation, which we refer to as sociality transitions. Figure. 4.1(c) shows an example of a network with a markedly positive average tie value generated in the regime in which an overall prosocial tendency can be reliably discerned.

We analytically calculate both the key eigenvalues and the transition conditions for large networks. In the main text, we use perturbation analysis to derive expressions for the signal eigenvalues (Sec. 4.4), which are then used to obtain the transition conditions by their equation with the main band edge eigenvalues (Sec. 4.5), these edge eigenvalues being found using random matrix theory (App. A.1). We also present an alternative to our perturbation treatment that derives the signal eigenvalues on the basis of random matrix theory, in keeping with previous treatments of detectability (App. A.2) [152, 222].

The spectral transitions have important implications for the outcomes of structural balance dynamics for networks possessing initial community structure. Structural balance theory, which postulates that triads with one or three negative edges will not endure, can be

implemented as a deterministic, continuous time dynamical system (Sec. 4.6),

$$\frac{dY_{ij}}{dt} = \sum_{k=1}^{N} Y_{ik}Y_{kj}, \tag{4.1}$$

where t is time and N is the number of nodes [111, 137]. This system evolves the connectivity Y_{ij} between nodes i and j as a function of their relationships with mutual neighbors: the product $Y_{ik}Y_{kj}$ increases their connectivity when they share either a common inclination, positive or negative, toward k but decreases it if their inclinations are oppositely signed. This dynamic promotes balanced triads and eradicates unbalanced triads in the network. The model evolves into a fully connected network where either: (1) there are two hostile factions with only positive ties within each and only negative ties between them; or (2) all nodes are positively connected in a single harmonious faction. In either case, the final state is determined by the leading eigenvector of the initial network.

The driving role played by the leading eigenvector of the initial network in the structural balance evolution gives rise to a dynamical manifestation of the detectability transition when the leading eigenvector also carries the community structure signal. For the two-faction outcome, if the leading eigenvector corresponds to the two identity types in the stochastic block model, then the final factions will perfectly align with these identities as shown in Fig. 4.1(a). On the other hand, if the leading eigenvector is merely the edge of the main noise band, as occurs for weak initial structure below the detectability transition, then the composition of the final factions will not align with the identity types as seen in Fig. 4.1(b). An analogous transition to the single-faction outcome is generated by the sociality transition as seen in Fig. 4.1(c). Solutions of the structural balance model starting from networks randomly generated by the stochastic block model do indeed show sharp transitions between behavioral regimes whose boundaries agree with analytical predictions based on the detectability and sociality transitions (Sec. 4.7).

We discuss the potential implications of these results for conflict dynamics among actors with different identity types due to, for instance, ethnicity, religion, or ideology (Sec.4.8). In particular, conflicts such as civil wars may take on a binary nature. If the system starts out

with weak identity-driven structure, then it will not be expected to polarize on the basis of identity. But complete identity polarization results even when initial affinities and animosities between identity types are fairly mild and even though identity itself plays no role in the micro-level conflict dynamics.

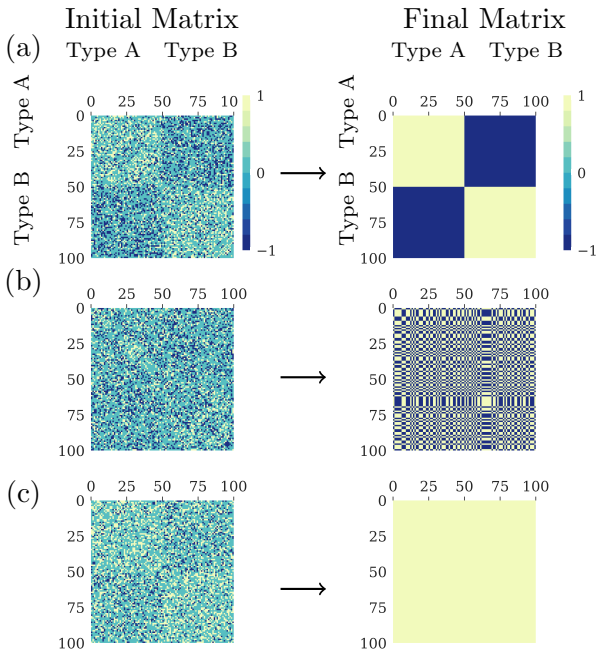

Figure 4.1: Evolution of networks with initial community structure under structural balance dynamics. (a) Moderate initial structuring by group identity leads to a completely connected network consisting of two factions sorted by identity. (b) Weak initial structure leads to two factions of mixed identities. (c) Strong initial positivity in the network leads to a single harmonious faction. Networks represented as adjacency matrices with $\pm 1, 0$ tie values indicated by color. Initial networks generated by stochastic block model, (4.2)–(4.4) with parameters $N = 100$ and $d_{in}, d_{out} = 0.4$ in all networks. $p^+_{out} = 0.3$ and $p^+_{in} = 0.7$ for (a), $p^+_{out} = 0.4$ and $p^+_{in} = 0.5$ for (b), and $p^+_{out} = 0.6$ and $p^+_{in} = 0.8$ for (c). Final networks represent the connectivity signs to which Eq. (4.1) converges (see Sec. 4.6).

4.2 Generating and representing community structure

Communities in an unsigned network are characterized by relatively dense within-community ties and sparse ties between communities. Community detection algorithms seek to discover these communities given an observed network [76, 158, 66]. Stochastic block models, which generate random networks with community structure by setting tie probabilities within and between blocks of nodes, have been used to investigate the behavior of community detection algorithms [159]. In this section, we describe the stochastic block model we use to generate our signed networks, the characterization of community structure via assortativity, and decomposition of the generated networks in terms of the eigenvectors of the average adjacency matrix and a random matrix.

4.2.1 Stochastic block model

Our construction starts with an undirected network of N nodes consisting of two identity groups A and B of equal size $N/2$, where $N \gg 1$. The A group nodes are indexed from 1 to $N/2$ and the B group from $N/2+1$ to N. $\mathbf{A}$ is the signed adjacency matrix where $\mathbf{A}_{ij}$ is the tie value between node i and node j, which can take on values of $\{1, -1, 0\}$ with 0 signifying the absence of a tie. As the network is undirected, the adjacency matrix is symmetric, $\mathbf{A}_{ij} = \mathbf{A}_{ji}$. The probability that a tie, positive or negative, will form between any given ingroup (A with A, B with B) node pair is d_{in}. Similarly, the tie formation probability between outgroup (A with B) node pairs is d_{out}. These tie formation probabilities are equivalent to the expected ingroup and outgroup tie densities and their average yields the expected tie density for the total network, $d = (d_{in} + d_{out})/2$. Given the presence of a tie between ingroup members, the conditional probability that it is positive is p_{in}^{+} and that it is negative is $p_{in}^{-} = 1 - p_{in}^{+}$. Similarly, the positive and negative tie conditional probabilities between outgroup nodes are written p_{out}^{+} and $p_{out}^{-} = 1 - p_{out}^{+}$. For brevity, we refer to p_{in}^{+} and p_{out}^{+} as the ingroup and outgroup *affinities* and p_{in}^{-} and p_{out}^{-} as the in and outgroup *animosities*.

The adjacency matrix can be written in terms of the following block structure:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A_{AA}} & \mathbf{A_{AB}} \\ \mathbf{A_{BA}} & \mathbf{A_{BB}} \end{bmatrix}, \tag{4.2}$$

where each block is a random $N/2 \times N/2$ matrix. The diagonal blocks represent AA or BB ties, whose elements are set using the following probability distribution for the ingroup random variable A_{in}:

$$\mathbb{P}(A_{in} = k) = \begin{cases} d_{in} p_{in}^{+} & , k = 1 \\ d_{in}(1 - p_{in}^{+}) & , k = -1 \\ 1 - d_{in} & , k = 0. \end{cases} \tag{4.3}$$

Since $\mathbf{A}$ is symmetric, there are $(N/2)(N/2 + 1)$ independent, identically distributed ingroup ties.

The off-diagonal blocks, corresponding to AB or BA ties, are transposes of each other resulting in $N^2/4$ *i.i.d.* outgroup ties, which are drawn according to the random variable A_{out}:

$$\mathbb{P}(A_{out} = k) = \begin{cases} d_{out} p_{out}^{+} & , k = 1 \\ d_{out}(1 - p_{out}^{+}) & , k = -1 \\ 1 - d_{out} & , k = 0. \end{cases} \tag{4.4}$$

Equations (4.2)–(4.4) define the stochastic block model used to generate matrices with more or less community structure as seen on the left in Fig. 4.1 and for all the numerically-generated spectra shown in this paper. Note that this model allows nonzero self-ties, unlike in many empirical networks, but this is a standard approximation that facilitates analytical treatment [152, 101, 222]. For large networks, the contribution of the N diagonal elements is negligible in comparison with that of the order N^2 off-diagonal elements. As we note below, the effect of removing self-ties on the average of $\mathbf{A}$ is to shift the eigenvalues by a constant that is independent of N. In addition, the model of structural balance dynamics, Eq. (4.1), allows for self-ties.

Networks that are too sparse become disconnected and the structural balance dynamics of the isolated subgroups will evolve independently of each other as opposed to the holistically evolution we see when the network forms one connected graph. An Erdős-Rényi graph will very likely form a single connected graph if $p > \ln(N)/N$ [57, 1]. Similarly, our stochastic block model matrices with two communities and density probabilities $d_{in} = a/N$ and $d_{out} = b/N$ will have a single giant component if $(a+b)/2 > 1$, and will very likely form a single connected graph if $d_{in} = a\ln(N)/N$, $d_{out} = b\ln(N)/N$ and $(a+b)/2 > 1$ [1].

4.2.2 Assortativity

Assortativity refers to the tendency for nodes of the same type to be more strongly connected than nodes of different types. We extend the standard definition of the assortativity coefficient for discrete node types [156] to our signed network case by calculating separate coefficients for the positive and negative tie networks and then essentially differencing them. We will use the signed network assortativity coefficient to characterize the regimes of the structural balance dynamics in Sec. 4.7.

First, considering the adjacency matrix of positive ties only, we let e_{ij}^{+} denote the fraction of all positive ties that connect a node of type i to one of type j where $i, j \in \{A, B\}$. The assortativity coefficient r^{+} for the network of positive ties, whose adjacency matrix elements are 1 if $A_{ij} > 0$ and zero otherwise, is then

$$r^{+} = \frac{\sum_i e_{ii}^{+} - \sum_i (a_i^{+})^2}{1 - \sum_i (a_i^{+})^2}, \tag{4.5}$$

where $a_i^{+} = \sum_j e_{ij}^{+}$. The assortativity coefficient can range between -1 and 1. A network containing only ingroup (AA or BB) positive ties with no outgroup (AB,BA) ties is completely assortative, $r^{+} = 1$, which in the social network context implies that people only cooperate with members of the same group. A network containing only outgroup ties is completely disassortative, $r^{+} = -1$, implying cooperation across the two groups but not within them. We see this state in the example shown in Fig. 4.2 in which $r^{+} = -1$ when the ingroup affinity $p_{in}^{+} = 0$. As the ingroup affinity increases, r^{+} increases but does not reach one as there are

still outgroup ties due to the nonzero value of the fixed outgroup affinity.

The assortativity for the network of negative ties, r^-, is defined analogously to Eq. (4.5). Whereas positive values of r^+ imply that ingroup relations are more friendly than outgroup relations, assortative mixing in the network of negative ties implies more hostility within groups than between them. Thus, in Fig. 4.2 we see that $r^- > 0$ when there is complete ingroup animosity ($p_{in}^- = 1$ corresponding to $p_{in}^+ = 0$) and $r^- = -1$ when there is no ingroup animosity and so negative ties within groups.

Accordingly, as we want positive values of our overall signed network assortativity coefficient r to signify that ingroup interactions tend to be more amicable than outgroup ones, we average r^+ and $-r^-$, yielding

$$r = \frac{r^+ - r^-}{2}, \tag{4.6}$$

which can take on values between -1 and 1. Figure 4.2 shows that r is negative for low ingroup affinity and positive for high ingroup affinity.

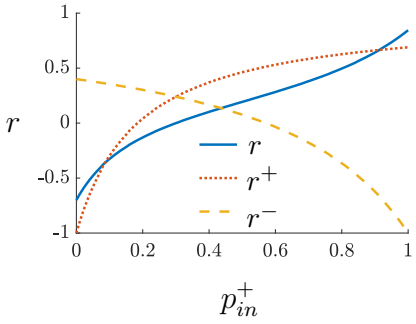

Figure 4.2: Example of assortativity coefficients as a function of ingroup affinity. Values calculated using expected tie numbers. Parameters are $d_{in} = 0.5$, $d_{out} = 0.3$, $p_{out}^- = 0.7$, and $N = 100$.

4.2.3 Adjacency matrix decomposition

This section presents a decomposition of the adjacency matrices generated by the stochastic block model into: (i) a signal component that results from the expected tie values generated by the ingroup and outgroup random variables, A_{in} and A_{out}; and (ii) a noise component due to random deviations from the expected values. This decomposition will form the starting point for our calculation of the network eigenvalues in Sec. 4.4. We write $\mathbf{A}$ as the sum of the average matrix $\langle \mathbf{A} \rangle$ and a random deviation matrix $\mathbf{X}$:

$$\mathbf{A} = \langle \mathbf{A} \rangle + \mathbf{X}. \tag{4.7}$$

Given the block structure of Eq. (4.2), $\langle \mathbf{A} \rangle$ can be written as

$$\langle \mathbf{A} \rangle = \begin{bmatrix} \langle \mathbf{A}_{AA} \rangle & \langle \mathbf{A}_{AB} \rangle \\ \langle \mathbf{A}_{BA} \rangle & \langle \mathbf{A}_{BB} \rangle \end{bmatrix}, \tag{4.8}$$

where each element of $\langle \mathbf{A}_{AA} \rangle$ and $\langle \mathbf{A}_{BB} \rangle$ is equal to $\langle A_{in} \rangle$ and each element of $\langle \mathbf{A}_{AB} \rangle$ and $\langle \mathbf{A}_{BA} \rangle$ is equal to $\langle A_{out} \rangle$. From Eqs. (4.3) and (4.4), we have

$$\langle A_{in} \rangle = d_{in}(2p_{in}^{+} - 1) \tag{4.9}$$

$$\langle A_{out} \rangle = d_{out}(2p_{out}^{+} - 1). \tag{4.10}$$

We define a couple of useful linear combinations of the in and outgroup expected tie values. We denote by μ the average over all the elements in $\langle \mathbf{A} \rangle$,

$$\mu = \frac{\langle A_{in} \rangle + \langle A_{out} \rangle}{2}, \tag{4.11}$$

and we denote by ν the half-difference between the in and outgroup expected tie values,

$$\nu = \frac{\langle A_{in} \rangle - \langle A_{out} \rangle}{2}. \tag{4.12}$$

Both μ and ν range from -1 to 1. Noting that $\langle A_{in} \rangle = \mu + \nu$ and $\langle A_{out} \rangle = \mu - \nu$, these expressions allow us to express $\langle \mathbf{A} \rangle$ as a sum of two outer products,

$$\langle \mathbf{A} \rangle = \mu N \mathbf{u}_H \mathbf{u}_H^T + \nu N \mathbf{u}_C \mathbf{u}_C^T, \tag{4.13}$$

where $\mathbf{u}_H$ and $\mathbf{u}_C$ are orthonormal N-dimensional vectors: $\mathbf{u}_H = \frac{1}{\sqrt{N}}[1, 1, ..., 1]^T$ and $\mathbf{u}_C = \frac{1}{\sqrt{N}}[1, ..., 1, -1, ..., -1]^T$, where the -1's align with the B block node indices. In fact, $\mathbf{u}_H$ and $\mathbf{u}_C$ are readily seen to be the two eigenvectors of $\langle \mathbf{A} \rangle$ with respective eigenvalues μN and νN:

$$\langle \mathbf{A} \rangle \mathbf{u}_H = \mu N \mathbf{u}_H \tag{4.14}$$

$$\langle \mathbf{A} \rangle \mathbf{u}_C = \nu N \mathbf{u}_C. \tag{4.15}$$

The term containing $\mathbf{u}_H$ in Eq. (4.13) generates a homogeneous $N \times N$ matrix whose elements are all equal to μ, the global average tie value. Hence, we refer to $\mathbf{u}_H$ as the *homogeneous* eigenvector. The term containing $\mathbf{u}_C$ generates a matrix whose diagonal block elements are all equal to ν and whose off-diagonal block elements are $-\nu$ and so corresponds to the structure of ingroup and outgroup tie differences. Accordingly, $\mathbf{u}_C$ generates the community structure and we refer to it as the *contrast* eigenvector.

The homogeneous and contrast eigenvectors are *signal* eigenvectors whose ability to be distinguished from the noise generated by $\mathbf{X}$ has important implications for community detectability and structural balance dynamics. From this perspective, μ and ν can be regarded as natural parameters for the signal structure in the network and could be used in place of two of the parameters in the stochastic block model, for instance, the ingroup and outgroup affinities. Doing so is less intuitive from a simulation viewpoint, however.

While $\langle \mathbf{A} \rangle$ is rank 2 in general, if either μ or ν equals zero, $\langle \mathbf{A} \rangle$ becomes a rank 1 matrix composed of either the homogeneous eigenvector or the contrast eigenvector. If $p^+_{out} = 1/2$ then $\langle A_{out} \rangle = 0$ and $\mu = \nu$ and so the two eigenvalues of $\langle \mathbf{A} \rangle$ are degenerate, as is the case when $d_{out} = 0$ and the two blocks are disconnected from each other. On the other hand, taking $d_{in} = 0$ yields a bipartite network between the A and B blocks in which $\mu = \langle A_{out} \rangle / 2 = -\nu$ and so the homogeneous and contrast eigenvalues are equal and opposite. If μ and ν are both zero, $\langle \mathbf{A} \rangle$ vanishes and $\mathbf{A}$ reduces to the noise matrix $\mathbf{X}$.

To remove self-ties from $\langle \mathbf{A} \rangle$, one can subtract $\langle A_{in} \rangle \mathbf{I}$ from Eq. (4.13), where $\mathbf{I}$ is the identity matrix. This shifts the signal eigenvalues by $-\langle A_{in} \rangle = -(\mu + \nu)$.

The noise matrix $\mathbf{X}$ is a symmetric matrix that can be written in the block form,

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_{AA} & \mathbf{X}_{AB} \\ \mathbf{X}_{BA} & \mathbf{X}_{BB} \end{bmatrix}. \tag{4.16}$$

Since $\mathbf{X} = \mathbf{A} - \langle \mathbf{A} \rangle$, the elements of the ingroup blocks $\mathbf{X}_{AA}$ and $\mathbf{X}_{BB}$ can assume values in $\{1 - \langle A_{in} \rangle, -1 - \langle A_{in} \rangle, -\langle A_{in} \rangle\}$ that are distributed according to the random variable X_{in},

$$\mathbb{P}(X_{in} = k) = \begin{cases} d_{in} p_{in}^{+} & , k = 1 - \langle A_{in} \rangle \\ d_{in}(1 - p_{in}^{+}) & , k = -1 - \langle A_{in} \rangle \\ 1 - d_{in} & , k = -\langle A_{in} \rangle. \end{cases} \tag{4.17}$$

Likewise, the entries of the outgroup blocks $\mathbf{X}_{AB} = \mathbf{X}_{BA}^{T}$ are distributed like X_{out},

$$\mathbb{P}(X_{out} = k) = \begin{cases} d_{out} p_{out}^{+} & , k = 1 - \langle A_{out} \rangle \\ d_{out}(1 - p_{out}^{+}) & , k = -1 - \langle A_{out} \rangle \\ 1 - d_{out} & , k = -\langle A_{out} \rangle. \end{cases} \tag{4.18}$$

All the elements of $\mathbf{X}$ have zero mean as $\langle X_{in} \rangle = \langle X_{out} \rangle = 0$. The variances of X_{in} and X_{out} are given by $\sigma_{in}^2 = \langle X_{in}^2 \rangle$ and $\sigma_{out}^2 = \langle X_{out}^2 \rangle$, which are written in terms of the stochastic block model parameters as

$$\sigma_{in}^2 = d_{in} - d_{in}^2 (2p_{in}^{+} - 1)^2 \tag{4.19}$$

$$\sigma_{out}^2 = d_{out} - d_{out}^2 (2p_{out}^{+} - 1)^2. \tag{4.20}$$

These variances will appear as their average,

$$\sigma^2 = \frac{\sigma_{in}^2 + \sigma_{out}^2}{2}, \tag{4.21}$$

in the noise-induced correction to the signal eigenvalues calculated below. The average variance can also be related to the parameters μ and ν as follows,

$$\sigma^2 = \frac{d_{in} + d_{out}}{2} - \frac{d_{in}^2 (2p_{in}^{+} - 1)^2 + d_{out}^2 (2p_{out}^{+} - 1)^2}{2} \tag{4.22}$$

$$= \frac{d_{in} + d_{out}}{2} - \frac{1}{2}(\langle A_{in} \rangle^2 + \langle A_{out} \rangle^2) \tag{4.23}$$

$$= \frac{d_{in} + d_{out}}{2} - \mu^2 - \nu^2, \tag{4.24}$$

where we have used Eqs. (4.9) and (4.10) in the second line and $\langle A_{in}\rangle^2 + \langle A_{out}\rangle^2 = (\mu+\nu)^2 + (\mu-\nu)^2 = 2(\mu^2+\nu^2)$ in the third.

4.3 Signal Eigenvalue Transitions

Spectral analysis has been used to address the number and detectability of communities in unsigned networks by considering the leading eigenvalues that reside outside the (approximately) continuous main spectral band due to its generation as a random graph [36, 152, 222, 179]. For undirected networks, the adjacency matrix is symmetric and hence has a real spectrum. The number of detectable communities is equivalent to the number of positive eigenvalues that lie beyond the main spectral band. Nadakuditi and Newman [152] showed the existence of, and analytically calculated, a detectability transition in which the community structure, as generated by a stochastic block model with two communities, while still present becomes no longer detectable. Under assortative tie formation, this transition occurs once the second eigenvalue of the adjacency matrix, which carries the community structure information, merges with the main spectral band. Using random matrix theory, the authors derived expressions for both leading eigenvalues and the edge of the spectral band, thereby enabling the analytical determination of the transition dependence.

Similar to unsigned networks, Fig. 4.3 illustrates that the spectra of our signed networks consist of a continuous band of eigenvalues originating from $\mathbf{X}$, the variability or noise in the system, and signal eigenvalues originating from $\langle\mathbf{A}\rangle$, the structure in the system. The edges of the main spectral band, $\pm\gamma$, are derived in Appendix A.1,

$$\gamma = 2\sigma\sqrt{N}, \tag{4.25}$$

while the average contrast and homogeneous signal eigenvalues, $\langle\lambda_C\rangle$ and $\langle\lambda_H\rangle$, are derived in Sec. 4.4 and Appendix A.2,

$$\langle\lambda_C\rangle = \nu N + \frac{\sigma^2}{\nu}, \quad |\nu| \geq \frac{\sigma}{\sqrt{N}} \tag{4.26}$$

$$\langle\lambda_H\rangle = \mu N + \frac{\sigma^2}{\mu}, \quad |\mu| \geq \frac{\sigma}{\sqrt{N}}. \tag{4.27}$$

The formulas for $\langle\lambda_C\rangle$ and $\langle\lambda_H\rangle$ consist of their respective eigenvalues from $\langle\mathbf{A}\rangle$ and a correction proportional to the average variance. The conditions $|\nu|, |\mu| \geq \sigma/\sqrt{N}$ imply that $|\langle\lambda_C\rangle|, |\langle\lambda_H\rangle| \geq |\gamma|$ so that (4.26) and (4.27) are only valid when outside of the main spectral band. Note that because of the self-averaging behavior for large N, we can effectively drop the expectation brackets and take $\lambda_C \approx \langle\lambda_C\rangle$ and $\lambda_H \approx \langle\lambda_H\rangle$.

The successive horizontal slices in Fig. 4.3(a) correspond to the eigenvalues of single instances of $\mathbf{A}$ generated by the stochastic block model as the outgroup animosity is increased. There are four points at which the outlying eigenvalues merge with the main band. Considering first the upper right of the plot, the largest eigenvalue, λ_1, is observed to detach from the main band for p_{out}^- greater than about 0.7. The right plot in Fig. 4.3(b), which depicts the signs of components of the first eigenvector $\mathbf{u}_1$, shows that $\mathbf{u}_1$ displays a two block structure for high p_{out}^-. Consequently, in this regime, the leading eigenvector corresponds to a perturbed version of the contrast eigenvector, $\mathbf{u}_C$, of $\langle\mathbf{A}\rangle$. The point at which λ_1 emerges from the main band is then identified with the community detectability transition. Indeed, below this point $\mathbf{u}_1$ loses its block structure and rapidly takes on the appearance of random noise. We observe that in contrast to the analogous case in unsigned networks where the detectability transition involves the second eigenvalue of the adjacency matrix, here the two communities are no longer discernible when the leading eigenvalue merges with the main spectral band. We refer to the transition involving the merging of the contrast eigenvalue with the positive edge of the main band as the *assortative* transition because the communities are preferentially grouped by identity type. With an eye toward its dynamical significance when the contrast eigenvalue is leading, the two final factions produced by the structural balance dynamics are polarized by identity type after λ_1 emerges from the band.

The leading eigenvalue is observed to undergo another transition at $p_{out}^- \approx 0.3$. For lower outgroup animosities, $\mathbf{u}_1$ takes on a single block structure and so can be taken to result from a perturbation of the homogeneous eigenvector $\mathbf{u}_H$. This homogeneous structure disappears from $\mathbf{u}_1$ for p_{out}^- values above the transition. As the homogeneous eigenvalue carries information about the average tie value over all nodes, its emergence from the noise

band can be considered a *sociality* transition. In particular, we refer to transitions that occur on the positive side of the noise band as *prosocial* transitions, in which a pattern of overall positive ties between nodes becomes apparent. The prosocial transition induces a transition in the structural balance dynamics from the two-faction equilibrium (not sorted by identity) to a single harmonious faction consisting of all nodes.

The lower left section of Fig. 4.3(a) shows the intersection of the last and least eigenvalue, λ_N, with the noise band at $p_{out}^- \approx 0.4$. For outgroup animosity values beneath this intersection, the last eigenvector $\mathbf{u}_N$ displays a two block structure as seen on the left plot of Fig. 4.3(b). However, although these blocks align with the A and B identity groups, the warmer outgroup than ingroup relations implies that the blocks are really disassortative "anti"-communities rather than assortative communities (prominent negative eigenvalues are also associated with disassortativity in unsigned networks [158]). Since this *disassortative* transition involves the least eigenvector, it has no significance with respect to the outcomes of the structural balance dynamics.

Finally, the other transition involving λ_N, seen in the upper left of Fig. 4.3(a), represents the emergence of the homogeneous eigenvalue and its corresponding single block structure from the noise band. It is a sociality transition and, in particular, an *antisocial* transition as it occurs on the negative side of the noise band signifying a conflictual relationship among nodes on average. The antisocial transition has no dynamical significance with respect to the structural balance dynamics.

Signal eigenvalues can occur on alternative sides of the spectral band as in Fig. 4.3(a) or on the same side as seen in Fig. 4.4(a). When λ_H and λ_C are on the same side of the spectral band, $p_{out}^- = 1/2$ is the point at which the signal eigenvalues cross one another since $\mu = \nu$. This affects which of the first two eigenvectors carries the community structure, but the structure itself remains apparent (Fig 4.4(b)). However, the signal crossing does affect the balance dynamics, producing a transition between the harmonious and assortative outcomes.

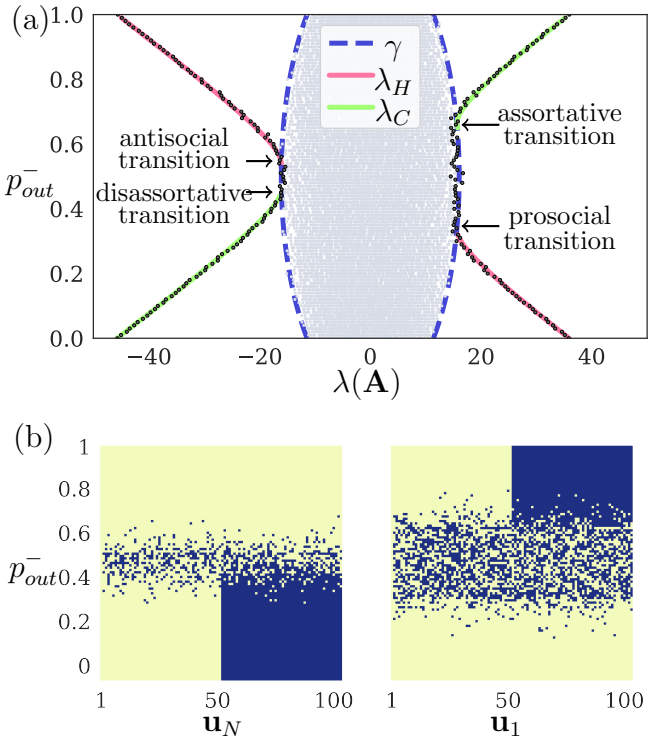

Figure 4.3: (a) Eigenvalue spectrum of $\mathbf{A}$ as a function of outgroup animosity p_{out}^-. The other parameters remain constant: $N = 100$, $p_{in}^+ = 0.4$, $d_{in} = 0.5$, $d_{out} = 0.8$. (b) Signs (yellow positive, blue negative) of the components of the last, $\mathbf{u}_N$, and first, $\mathbf{u}_1$, eigenvectors as functions of p_{out}^-. The theoretical curves for λ_C (solid green), λ_H (solid pink), and γ (dashed blue) are calculated from Eqs. (4.57), (4.58), and (A.15) respectively.

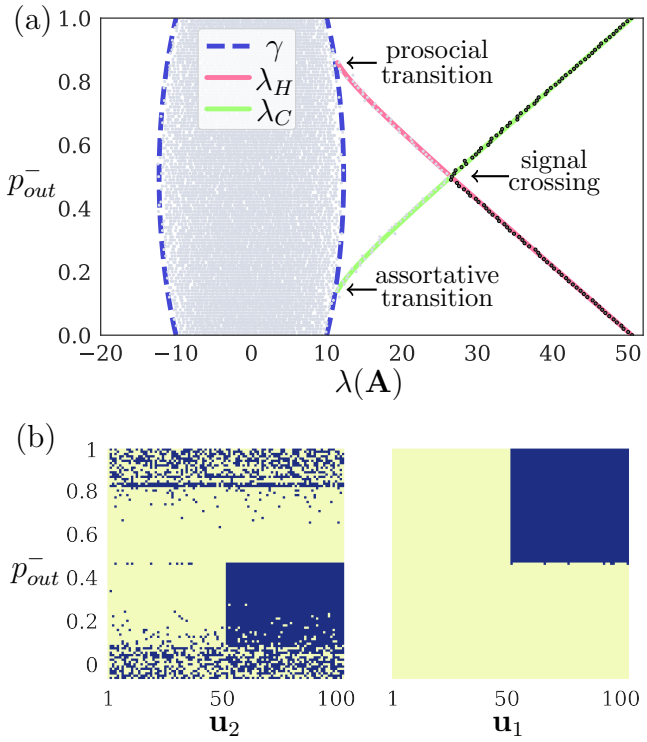

Figure 4.4: (a) Spectrum of $\mathbf{A}$ as a function of p_{out}^- with constant parameters $N = 100$, $d_{in}, d_{out} = 0.5$, and $p_{in}^+ = 1$. (b) Signs (yellow positive, blue negative) of the components of the first and second eigenvectors, $\mathbf{u}_1$ and $\mathbf{u}_2$, as functions of p_{out}^-. The theoretical curves for λ_C (solid green), λ_H (solid pink), and γ (dashed blue) are calculated from Eqs. (4.57), (4.58), and (A.15) respectively.

4.4 Calculation of signal eigenvalues

In this section, we will derive formulas for the signal eigenvalues as a function of our stochastic block model parameters. We employ a perturbation treatment here but present an alternative derivation employing random matrix theory and complex analysis in App. A.2. Equation (4.7), which expresses the adjacency matrix $\mathbf{A}$ as the sum of its expected value $\langle \mathbf{A} \rangle$ and a matrix of random deviations $\mathbf{X}$, will form the starting point of our analysis. We consider $\langle \mathbf{A} \rangle$ as a given deterministic matrix with homogeneous and contrast eigenvalues and eigenvectors, Eqs. (4.14) and (4.15), that is subject to a perturbation from the independent noise matrix $\mathbf{X}$, which induces shifts to the signal eigenvalues and eigenvectors. A perturbation expansion to second order and the statistics of the noise matrix then yield the corrections to the signal eigenvalues. We note that in treating $\mathbf{X}$ as an independent perturbation to $\langle \mathbf{A} \rangle$, we temporarily suspend their linkage via the tie formation probabilities in the stochastic block model.

4.4.1 Perturbation Expansion Setup

We show the perturbation calculation for the case of the contrast eigenvalue. The homogeneous eigenvalue can be obtained in precisely analogous fashion. The eigenvalue equation is

$$(\langle \mathbf{A} \rangle + \mathbf{X})\mathbf{v}_C = \lambda_C \mathbf{v}_C. \tag{4.28}$$

We write the perturbed eigenvector and eigenvalue up to second order as

$$\mathbf{v}_C = \mathbf{u}_C + \mathbf{u}^{(1)} + \mathbf{u}^{(2)} \tag{4.29}$$

$$\lambda_C = \nu N + \lambda^{(1)} + \lambda^{(2)}, \tag{4.30}$$

where $\lambda^{(1)}, \lambda^{(2)}, \mathbf{u}^{(1)}, \mathbf{u}^{(2)}$ are the first and second order perturbations. As the unperturbed eigenvalue is $\mathcal{O}(\nu N)$ we divide Eq. (4.28) by N so that the zeroth order equation is $\mathcal{O}(1)$. Then, using Eqs. (4.29) and (4.30), the eigenvalue equation becomes

$$\begin{aligned} &\left(\frac{\langle \mathbf{A} \rangle}{N} + \frac{\mathbf{X}}{N}\right)(\mathbf{u}_C + \mathbf{u}^{(1)} + \mathbf{u}^{(2)}) \\ =&\left(\nu + \frac{\lambda^{(1)}}{N} + \frac{\lambda^{(2)}}{N}\right)(\mathbf{u}_C + \mathbf{u}^{(1)} + \mathbf{u}^{(2)}). \end{aligned} \tag{4.31}$$

Before embarking upon our perturbation analysis, we specify the appropriate expansion parameter. To do so, we determine the orders of the $\langle \mathbf{A} \rangle$ and $\mathbf{X}$ matrices by evaluating their 2-norms. The 2-norm of our matrix is equivalent to its largest eigenvalue. Consequently, for the unperturbed matrix, $||\langle \mathbf{A} \rangle||_2$ is given by the larger of $N|\nu|$ or $N|\mu|$. Taking ν, μ to be $\mathcal{O}(1)$ therefore implies that $\langle \mathbf{A} \rangle$ is $\mathcal{O}(N)$. In Appendix A.1, using Wigner's semicircle law for the spectral density of a random matrix as well as matrix bounds, we determine that $||\mathbf{X}||_2$ is $\mathcal{O}(\sigma\sqrt{N})$. The ratio of the orders of $\mathbf{X}$ to $\langle \mathbf{A} \rangle$ is $\mathcal{O}(\sigma/\sqrt{N})$, and so successive orders in the perturbation series must diminish by a factor of $\sigma/\sqrt{N}$, which therefore serves as our expansion parameter. At a given N, the perturbation can be made arbitrarily small by letting σ go to zero. But in the large N regime, we need not constrain σ to be small.

Separating Eq. (4.31) out by expansion orders yields

$\mathcal{O}(1)$

$$\frac{\langle \mathbf{A} \rangle}{N}\mathbf{u}_C = \nu \mathbf{u}_C \tag{4.32}$$

$\mathcal{O}(\sigma/\sqrt{N})$

$$\frac{\langle \mathbf{A} \rangle}{N}\mathbf{u}^{(1)} + \frac{\mathbf{X}}{N}\mathbf{u}_C = \nu \mathbf{u}^{(1)} + \frac{\lambda^{(1)}}{N}\mathbf{u}_C \tag{4.33}$$

$\mathcal{O}(\sigma^2/N)$

$$\frac{\langle \mathbf{A} \rangle}{N}\mathbf{u}^{(2)} + \frac{\mathbf{X}}{N}\mathbf{u}^{(1)} = \nu \mathbf{u}^{(2)} + \frac{\lambda^{(1)}}{N}\mathbf{u}^{(1)} + \frac{\lambda^{(2)}}{N}\mathbf{u}_C. \tag{4.34}$$

4.4.2 First Order Treatment

To find the first order eigenvalue perturbation, $\lambda^{(1)}$, we multiply both sides of Eq. (4.33) by $N\mathbf{u}_C^T$. Then using Eq. (4.13) and the orthonormality of $\mathbf{u}_C$ and $\mathbf{u}_H$ gives

$$\nu N \mathbf{u}_C^T \mathbf{u}^{(1)} + \mathbf{u}_C^T \mathbf{X} \mathbf{u}_C = \nu N \mathbf{u}_C^T \mathbf{u}^{(1)} + \lambda^{(1)}. \tag{4.35}$$

Solving for $\lambda^{(1)}$ yields

$$\begin{aligned}\lambda^{(1)} &= \mathbf{u}_C^T \mathbf{X} \mathbf{u}_C \\ &= \frac{1}{N}\left\{ \sum_{i,j=1}^{\frac{N}{2}} X_{ij} + \sum_{i,j=\frac{N}{2}+1}^{N} X_{ij} - 2\sum_{i=1}^{\frac{N}{2}} \sum_{j=\frac{N}{2}+1}^{N} X_{ij} \right\}.\end{aligned} \tag{4.36}$$

The first two terms in the braces above sum ties in the ingroup blocks AA and BB respectively, each tie distributed as X_{in}, and the third term corresponds to the AB and BA outgroup ties, distributed as X_{out}. As $\lambda^{(1)}$ is equal to the sum of zero-mean random ingroup and outgroup variables, its mean therefore vanishes,

$$\langle \lambda^{(1)} \rangle = 0. \tag{4.37}$$

Turning to the variance, each element within the outgroup sum has variance σ_{out}^2, which becomes $4\sigma_{out}^2/N^2$ when the $2/N$ prefactor is included. The contribution to the variance from the $N^2/4$ outgroup variables is therefore σ_{out}^2. Similarly, for the ingroup sums, the symmetry of X implies that, neglecting the diagonal, there are approximately a total of $N^2/4$ independent variables each with variance $4\sigma_{in}^2/N^2$ so that the ingroup variance contribution is σ_{in}^2. Accordingly, the variance of $\lambda^{(1)}$ is

$$\mathrm{Var}(\lambda^{(1)}) = \sigma_{in}^2 + \sigma_{out}^2 = 2\sigma^2. \tag{4.38}$$

To solve for $\mathbf{u}^{(1)}$, we write it as a vector decomposition and solve for the individual components,

$$\mathbf{u}_1^{(1)} = \mathbf{u}_{||}^{(1)} + \mathbf{u}_{\perp}^{(1)}, \tag{4.39}$$

where $\mathbf{u}_{||}^{(1)}$ is the component of $\mathbf{u}^{(1)}$ that is in the $\mathbf{u}_C, \mathbf{u}_H$ plane, and $\mathbf{u}_{\perp}^{(1)}$ is the component orthogonal to that plane. We find $\mathbf{u}_{||}^{(1)}$ and $\mathbf{u}_{\perp}^{(1)}$ by multiplying both sides of the $\mathcal{O}(\sigma/\sqrt{N})$ equation, Eq. (4.33) by $\mathbf{u}_H^T$, the transpose of the homogeneous eigenvector of $\langle \mathbf{A} \rangle$, which gives

$$\mathbf{u}_H^T(\langle \mathbf{A} \rangle \mathbf{u}^{(1)} + \mathbf{X}\mathbf{u}_C) = \mathbf{u}_H^T(\nu N \mathbf{u}^{(1)} + \lambda^{(1)} \mathbf{u}_C), \tag{4.40}$$

which after employing the eigenvector properties becomes

$$\mu N \mathbf{u}_H^T \mathbf{u}^{(1)} + \mathbf{u}_H^T \mathbf{X} \mathbf{u}_C = \nu N \mathbf{u}_H^T \mathbf{u}^{(1)}. \tag{4.41}$$

Rearranging and noting that $\mathbf{u}_H^T \mathbf{u}^{(1)} = \mathbf{u}_H^T \mathbf{u}_{\shortparallel}^{(1)}$, we find

$$\mathbf{u}_H^T \mathbf{u}^{(1)} = \mathbf{u}_H^T \mathbf{u}_{\shortparallel}^{(1)} = \frac{\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C}{\nu N - \mu N}, \tag{4.42}$$

which suggests the following solution for $\mathbf{u}_{\shortparallel}^{(1)}$,

$$\mathbf{u}_{\shortparallel}^{(1)} = \frac{[\mathbf{X}\mathbf{u}_C]_{\shortparallel}}{\nu N - \mu N}. \tag{4.43}$$

Writing $[\mathbf{X}\mathbf{u}_C]_{\shortparallel}$ as a decomposed projection onto $\mathbf{u}_C$ and $\mathbf{u}_H$,

$$[\mathbf{X}\mathbf{u}_C]_{\shortparallel} = (\mathbf{u}_C^T \mathbf{X} \mathbf{u}_C)\mathbf{u}_C + (\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C)\mathbf{u}_H \tag{4.44}$$

allows us to write $\mathbf{u}_{\shortparallel}^{(1)}$ as

$$\mathbf{u}_{\shortparallel}^{(1)} = \frac{(\mathbf{u}_C^T \mathbf{X} \mathbf{u}_C)}{\nu N - \mu N}\mathbf{u}_C + \frac{(\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C)}{\nu N - \mu N}\mathbf{u}_H. \tag{4.45}$$

We now seek to solve for $\mathbf{u}_{\perp}^{(1)}$. Using the decomposition (4.39) in the first order equation, (4.33), gives

$$\langle \mathbf{A} \rangle [\mathbf{u}_{\shortparallel}^{(1)} + \mathbf{u}_{\perp}^{(1)}] + \mathbf{X}\mathbf{u}_C = \nu N [\mathbf{u}_{\shortparallel}^{(1)} + \mathbf{u}_{\perp}^{(1)}] + \lambda^{(1)} \mathbf{u}_C. \tag{4.46}$$

Retaining only the terms that have components orthogonal to the $\mathbf{u}_C, \mathbf{u}_H$ plane and rearranging yields

$$\mathbf{u}_{\perp}^{(1)} = \frac{[\mathbf{X}\mathbf{u}_C]_{\perp}}{\nu N}. \tag{4.47}$$

Using the substitution $[\mathbf{X}\mathbf{u}_C]_{\perp} = \mathbf{X}\mathbf{u}_C - [\mathbf{X}\mathbf{u}_C]_{\shortparallel}$ then gives the solution for the orthogonal component

$$\mathbf{u}_{\perp}^{(1)} = \frac{\mathbf{X}\mathbf{u}_C - [(\mathbf{u}_C^T \mathbf{X} \mathbf{u}_C)\mathbf{u}_C + (\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C)\mathbf{u}_H]}{\nu N}. \tag{4.48}$$

Combining Eqs. (4.45) and (4.48) gives the solution for the first order perturbation to the eigenvector $\mathbf{u}^{(1)}$,

$$\mathbf{u}^{(1)} = \frac{\mathbf{X}\mathbf{u}_C}{\nu N} + \frac{\mu(\mathbf{u}_C^T \mathbf{X} \mathbf{u}_C)}{\nu N(\nu - \mu)}\mathbf{u}_C + \frac{\mu(\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C)}{\nu N(\nu - \mu)}\mathbf{u}_H. \tag{4.49}$$

4.4.3 Second Order Treatment

Having found the first order eigenvalue and eigenvector perturbations, $\lambda^{(1)}$ and $\mathbf{u}^{(1)}$, we can now solve for the second order correction $\lambda^{(2)}$. We multiply both sides of Eq. (4.34) by $\mathbf{u}_C^T$ and then solve to get

$$\lambda^{(2)} = \mathbf{u}_C^T \mathbf{X} \mathbf{u}^{(1)} - \lambda^{(1)} \mathbf{u}_C^T \mathbf{u}^{(1)} \tag{4.50}$$

$$= \frac{\mathbf{u}_C^T \mathbf{X}^2 \mathbf{u}_C}{\nu N} - \frac{(\mathbf{u}_C^T \mathbf{X} \mathbf{u}_C)^2}{\nu N} + \frac{\mu (\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C)^2}{\nu N (\nu - \mu)}, \tag{4.51}$$

where Eqs. (4.36) and (4.49) have been used to obtain the second line.

We seek the expected value $\langle \lambda^{(2)} \rangle$ and consider the righthand terms of Eq. (4.51) in succession. Expanding the expected value of the first term yields

$$\frac{\langle \mathbf{u}_C^T \mathbf{X}^2 \mathbf{u}_C \rangle}{\nu N} = \frac{1}{\nu N} \frac{1}{N} \Bigg\{ \sum_{i,j=1}^{\frac{N}{2}} \langle (\mathbf{X}^2)_{ij} \rangle + \sum_{i,j=\frac{N}{2}+1}^{N} \langle (\mathbf{X}^2)_{ij} \rangle - 2 \sum_{i=1}^{\frac{N}{2}} \sum_{j=\frac{N}{2}+1}^{N} \langle (\mathbf{X}^2)_{ij} \rangle \Bigg\}, \tag{4.52}$$

where $(\mathbf{X}^2)_{ij} = \sum_{k=1}^{N} X_{ik} X_{kj}$. As the elements of $\mathbf{X}$ are independent, the cross-element terms in this sum have vanishing expectation: $\langle X_{ik} X_{kj} \rangle = 0$ for $i \neq j$. When $i = j$, the value of $\langle X_{ik}^2 \rangle$ is either the ingroup or outgroup variance: $\langle X_{ik}^2 \rangle = \sigma_{in}^2$ if $i, k \leq N/2$ or $i, k > N/2$; $\langle X_{ik}^2 \rangle = \sigma_{out}^2$ otherwise. Accordingly, the expectations for the elements of $\mathbf{X}^2$ are given by

$$\langle (\mathbf{X}^2)_{ij} \rangle = \begin{cases} \frac{N}{2}\sigma_{in}^2 + \frac{N}{2}\sigma_{out}^2 \;\; = N\sigma^2, & i = j \\ \\ 0, & i \neq j. \end{cases} \tag{4.53}$$

The above equation reduces the double sums in the first two terms in Eq. (4.52) to single sums over $\langle (\mathbf{X}^2)_{ii} \rangle = N\sigma^2$, which can then be combined. The last term, which contains only off-diagonal elements of $\langle (\mathbf{X}^2) \rangle$, vanishes. The contribution of the first term in Eq. (4.51) to

$\langle\lambda^{(2)}\rangle$ is therefore

$$\frac{\langle \mathbf{u}_C^T \mathbf{X}^2 \mathbf{u}_C \rangle}{\nu N} = \frac{1}{\nu N}\frac{1}{N}\sum_{i=1}^{N} N\sigma^2 \tag{4.54}$$

$$= \frac{\sigma^2}{\nu}. \tag{4.55}$$

We now turn to the second and third terms on the righthand of Eq. (4.51). The numerator of the second term, $(\mathbf{u}_C^T \mathbf{X} \mathbf{u}_C)^2$, involves the square of $\lambda^{(1)}$ by Eq. (4.36). Hence, its expected value is equivalent to the variance of $\lambda^{(1)}$ (which has zero mean), and so goes as σ^2 as shown above. Consequently, the expected value of the second term goes as σ^2/N. The same argument holds for the third term as its numerator depends on $\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C$, which is likewise tantamount to the sum over the random ingroup and outgroup variables. Therefore, in the large-N regime of concern here, the second and third terms, which have a $1/N$ dependence, can be neglected in comparison with the first term, which is independent of N. Accordingly, Eq. (4.55) gives the second order eigenvalue perturbation

$$\langle\lambda^{(2)}\rangle = \frac{\sigma^2}{\nu}. \tag{4.56}$$

Having found that the first order perturbation vanishes on average and given the second order perturbation above, we arrive at the approximate solution for the expected value of the contrast eigenvalue of Eq. (4.28),

$$\langle\lambda_C\rangle = \nu N + \frac{\sigma^2}{\nu}. \tag{4.57}$$

Since the ratio of the second-order correction to the unperturbed eigenvalue goes as σ^2/N, the expansion parameter, given by its square root, is therefore $\mathcal{O}(\sigma/\sqrt{N})$ as stated at the beginning of this calculation.

A similar calculation, this time expanding about the homogeneous eigenvector $\mathbf{u}_H$, gives us the solution for λ_H, which simply involves swapping out ν for μ in the preceding equation,

$$\langle\lambda_H\rangle = \mu N + \frac{\sigma^2}{\mu}. \tag{4.58}$$

The analytical expressions for the signal eigenvalues, (4.57) and (4.58), are plotted in Fig. 4.3(a) for values outside the main spectral band. They are observed to be in good agreement with the outlying eigenvalues of the numerical spectrum. This is the case even though, rather than an average over many generated networks, each horizontal slice represents just one instance, a reflection of the self-averaging behavior of large random networks. These expressions also work well in the non-sparse limit as shown in Fig. 4.5 for the case where the contrast eigenvalue becomes the leading eigenvalue beyond the assortative transition. The predicted λ_C is observed to separate from the spectral edge past a critical density, which decreases with network size, and shows good agreement with the first eigenvalue of the simulated network, particularly for the two larger networks.

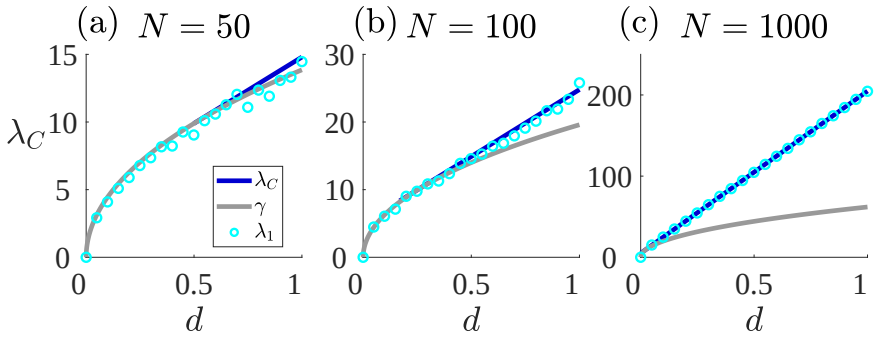

Figure 4.5: Theoretical contrast (λ_C, Eq. (4.57)) and band edge (γ, Eq. (A.15)) eigenvalues, along with the simulated leading eigenvalue (λ_1), as a function of network density. (a) $N = 50$ (b) $N = 100$ (c) $N = 1000$. λ_C is plotted for network density values where it meets or exceeds γ. One leading eigenvalue instance is computed for each d value. $d = d_{in} = d_{out}$, $p^+_{in} = 0.6$, and $p^+_{out} = 0.4$.

For the special case $\mu = \nu$, the same signal eigenvalue expressions still hold but we note how this case affects the derivation. When $\mu \neq \nu$, the last term on the right side of Eq. (4.51) could be neglected above as it is a factor of $1/N$ smaller than the first term. For $\mu = \nu$, however, the third term is singular. Yet, Eq. (4.41) implies that $\mathbf{u}_H^T \mathbf{X} \mathbf{u}_C = 0$ for $\mu = \nu$ (or is

higher order for $\mu - \nu \sim 1/N$) and so the seemingly problematic third term does not arise. For the special cases $\nu = 0$ or $\mu = 0$, $\langle \lambda_C \rangle = 0$ or $\langle \lambda_H \rangle = 0$ respectively and $\langle \mathbf{A} \rangle$ becomes rank 1. Figure 4.7(b) shows example spectra for $\mu = 0$ in which λ_C is the only signal eigenvalue.

The symmetric forms of the expressions for λ_C and λ_H reflect the fact that an orthogonal transformation $\mathbf{K}$ exists that transforms the contrast and homogeneous eigenvectors of $\langle \mathbf{A} \rangle$ into each other, that is, $\mathbf{u}'_H = \mathbf{K}\mathbf{u}_C$ and $\mathbf{u}'_C = \mathbf{K}\mathbf{u}_H$ where the primes denote the transformed system. Specifically, $\mathbf{K}$ is the diagonal matrix $\mathrm{diag}(1, \ldots, 1, -1, \ldots, -1)$ where the negative values start at index $N/2 + 1$. It is its own inverse, $\mathbf{K}^{-1} = \mathbf{K}$. The transformation of the expected adjacency matrix, $\langle \mathbf{A}' \rangle = \mathbf{K}\langle \mathbf{A} \rangle \mathbf{K}$, flips the sign of the off-diagonal blocks so that $\langle A'_{out} \rangle = -\langle A_{out} \rangle$. Therefore, by Eqs. (4.11) and (4.12), $\mu' = \nu$ and $\nu' = \mu$, which swaps the perturbed signal eigenvalues, $\lambda'_H = \lambda_C$ and $\lambda'_C = \lambda_H$.

An alternative calculation of the signal eigenvalues based on random matrix theory and complex analysis is presented in the Appendix A.2. We find the same formulas for the signal eigenvalues (see Eq. (A.44)) as have been derived here.

4.5 Transition Boundaries

In this section, we derive theoretical predictions for the boundaries of the detectability and sociality transitions. As discussed in Sec. 4.3, these transitions occur when the signal eigenvalues merge with the main spectral band. From Eq. (A.15), the edges of the main band of $\mathbf{X}$ are given by $\pm 2\sigma\sqrt{N}$, a formula that is a straightforward adaptation of the band edge of Wigner's semicircle distribution. We consider the community detectability transitions first, that is, those involving the contrast eigenvector $\mathbf{u}_C$, whose eigenvalue is given by Eq. (4.57). The detectability transition will therefore occur when

$$\nu N + \frac{\sigma^2}{\nu} = 2\sigma\sqrt{N}. \tag{4.59}$$

Solving for the critical value ν_* yields

$$\nu_* = \frac{\sigma(\nu_*)}{\sqrt{N}}. \tag{4.60}$$

The notation $\sigma(\nu_*)$ serves as a reminder that ν_* also appears on the righthand side due to the functional dependence of σ given by Eq. (4.24). The community structure is detectable when $|\nu| > \nu_*$. In particular, the assortative transition occurs for $\nu = \nu_*$ and the detectability transition for disassortative structure occurs for $\nu = -\nu_*$.

We observe that the transition condition (4.60) also results by setting the noise power equal to the signal power. Defining the noise power as the projection of $\mathbf{X}^2$ onto $\mathbf{u}_C$, its average, $\langle \mathbf{u}_C^T \mathbf{X}^2 \mathbf{u}_C \rangle$, is found using Eq. (4.55) to be $N\sigma^2$. One could also arrive at this value by considering how much of the total noise variance, $N^2\sigma^2$, is carried on average by each of N randomly chosen orthogonal basis vectors. Equating the signal power to the average noise power, $\nu^2 N^2 = N\sigma^2$, yields (4.60).

The sociality transitions associated with the homogeneous signal occur when the homogeneous eigenvalue given by Eq. (4.58) equals the band edge eigenvalue. This yields a critical value μ_*,

$$\mu_* = \frac{\sigma(\mu_*)}{\sqrt{N}}. \tag{4.61}$$

The prosocial transition occurs for $\mu = \mu_*$ and the antisocial transition occurs for $\mu = -\mu_*$.

We now unpack the transition conditions derived above to express them in alternative ways in parameter space that will further intuitive understanding of the transition behavior and allow for connection with simulation results.

First, we substitute Eq. (4.24) for σ^2 in the detectability condition (4.60) to yield

$$N\nu^2 = \frac{1}{2}(d_{in} + d_{out}) - \mu^2 - \nu^2. \tag{4.62}$$

We point out that $(d_{in} + d_{out})/2$ is simply the overall tie density in the network. For a sparse network, $d_{in}, d_{out} \ll 1$, we can neglect the μ^2 and ν^2 terms on the right-hand side, so that the critical value ν_* is given by

$$\nu_* = \pm\sqrt{\frac{1}{2N}(d_{in} + d_{out})}. \tag{4.63}$$

The positive sign corresponds to the assortative transition and the negative sign corresponds to the disassortative transition. As ν can be regarded as a natural parameter for the community

structure, this structure (assortative or disassortative) becomes easier to detect as the network becomes more sparse since $|\nu_*|$ shifts to smaller values (but care should be taken to distinguish the behavior of ν from that of the affinities and animosities, which can behave oppositely with density as in Eq. (4.72) below). Weaker structure is also more detectable as the size of the network grows, as was already apparent from Eq. (4.60).

We now substitute into Eq. (4.62) the definitions (4.11) and (4.12) for μ and ν and rearrange to obtain

$$\begin{aligned} 0 = \langle A_{out}\rangle^2 - \frac{2N}{N+2}\langle A_{out}\rangle\langle A_{in}\rangle \\ + \langle A_{in}\rangle^2 - \frac{2}{N+2}(d_{in}+d_{out}), \end{aligned} \tag{4.64}$$

which can be solved for the critical value of $\langle A_{out}\rangle$ (omitting the asterisk),

$$\begin{aligned} \langle A_{out}\rangle = \frac{N}{N+2}\langle A_{in}\rangle \\ \pm\sqrt{\frac{2}{N+2}(d_{in}+d_{out}) - \frac{4(N+1)}{(N+2)^2}\langle A_{in}\rangle^2}. \end{aligned} \tag{4.65}$$

To write the detectability transitions completely in terms of the block model probabilities, we substitute Eqs. (4.9) and (4.10) for $\langle A_{in}\rangle$ and $\langle A_{out}\rangle$ into (4.65). Solving for the outgroup animosity and taking the large N limit yields

$$\begin{aligned} p_{out}^- = \frac{1}{2} - \frac{d_{in}}{d_{out}}\left(p_{in}^+ - \frac{1}{2}\right) \\ \pm\frac{1}{d_{out}}\sqrt{\frac{d_{in}+d_{out}-8d_{in}^2(p_{in}^+-\frac{1}{2})^2}{2N}}, \end{aligned} \tag{4.66}$$

where the positive sign corresponds to the assortative transition. Neglecting the second term inside the square root yields the sparse limit, equivalent to Eq. (4.63).

For the sociality transitions, the sparse limit results in the condition

$$\mu_* = \pm\sqrt{\frac{1}{2N}(d_{in}+d_{out})}. \tag{4.67}$$

The positive and negative signs correspond to the prosocial and antisocial transitions respectively. The critical value of the outgroup animosity is given by

$$p_{out}^{-} = \frac{1}{2} + \frac{d_{in}}{d_{out}}\left(p_{in}^{+} - \frac{1}{2}\right) \pm \frac{1}{d_{out}}\sqrt{\frac{d_{in} + d_{out} - 8d_{in}^{2}(p_{in}^{+} - \frac{1}{2})^{2}}{2N}}, \tag{4.68}$$

where now the negative sign is used for the prosocial transition.

4.6 Structural Balance Dynamics

In its simplest incarnation, structural balance theory considers the stability of triads. Triads with all positive ties ("the friend of my friend is my friend") or two negative ties ("the enemy of my enemy is my friend") are considered balanced and so stable. In contrast, a triad with an odd number of negative ties will be unbalanced. For fully connected networks, assuming that all triads must be balanced over time implies that the system achieves either a state of global harmony in which all nodes are positively connected or two hostile camps with positive connections within each camp and negative connections between them [32]. Empirical signed networks in social systems such as international relations, student relationships, and online social networks have been found to be approximately balanced [107, 206, 59], exhibiting a tendency toward partition into two factions.

Although the concept of balance can be extended to arbitrary-length cycles, the triadic notion has motivated the construction of dynamical systems models that evolve the relationship between a pair of nodes as a function of their relationships with their network neighbors [111, 137, 207]. As noted when Eq. (4.1) was introduced, if both members of a dyad have a positive relationship with a third node, that will act towards making the focal dyad's relationship more positive. In contrast, having oppositely signed relationships with the third node will contribute a force pulling the dyad toward a more conflictual relationship.

Unbalanced triads wither away under these dynamics. Building upon Ref. [111], Marvel et al. [137] demonstrated that the model of structural balance dynamics defined by (4.1) almost always achieves a balanced state starting from random initial conditions. Equation (4.1) can be written as a matrix equation,

$$\frac{d\mathbf{Y}}{dt} = \mathbf{Y}^2, \tag{4.69}$$

where $\mathbf{Y}$ is the matrix of signed and continuous connectivity values, Y_{ij}, between node pairs. In support of its empirical relevance, Ref. [137] found that when implemented upon the initial network of several real world systems, this model well predicts the observed final network.

Equation (4.69) is the matrix form of a Riccati equation and has the following closed form solution [137]:

$$\mathbf{Y}(t) = \mathbf{Y}(0)[\mathbf{I} - \mathbf{Y}(0)t]^{-1}. \tag{4.70}$$

The elements Y_{ij} diverge to positive or negative infinity in a finite time t_f and so the solution only holds for $t < t_f$.

For the purposes of analyzing community structure, we convert the connectivity matrix $\mathbf{Y}$ to an adjacency matrix $\mathbf{A}$ with discrete values ± 1 and 0 by taking the sign of the connectivity values so that $A_{ij} = \text{sgn}(Y_{ij})$. The leading eigenvector of the initial connectivity matrix, $\mathbf{Y}(0)$, grows fastest and so dominates the solution as $t \to t_f$. As a result, the final adjacency matrix $\mathbf{A}_f$ corresponding to $\mathbf{Y}$ as $t \to t_f$ converges to the outer product

$$\mathbf{A}_f = \mathbf{u}_1 \mathbf{u}_1^T, \tag{4.71}$$

where $\mathbf{u}_1$ consists of the signs of the leading eigenvector of $\mathbf{Y}(0)$.

The rank 1 structure of $\mathbf{A}_f$ toward which the connectivity matrix converges implies that the final network must partition into either two hostile factions or one harmonious community as consistent with the expectations of structural balance theory [137]. The final network consists of a single harmonious faction if the components of $\mathbf{u}_1$ are of uniform sign, but consists of two hostile factions if $\mathbf{u}_1$ contains both positive and negative values [137]. Note that these results hold only if there is a single dominant eigenvalue and the graph

is connected. When $d_{out} = 0$, the graph is disconnected and the isolated identity blocks will evolve independently of each other and the connectivity matrix will become rank 2 after structural balance dynamics. Although the first two eigenvalues of $\langle \mathbf{A} \rangle$ are equal in the $p^+_{out} = 1/2$ case as well, the network remains connected and, due to stochasticity, one eigenvalue will inevitably be slightly larger in the realized $\mathbf{A}$, causing it to generate the ultimate rank 1 state. The case of $d_{in} = 0$ is also connected and so evolves to rank 1.

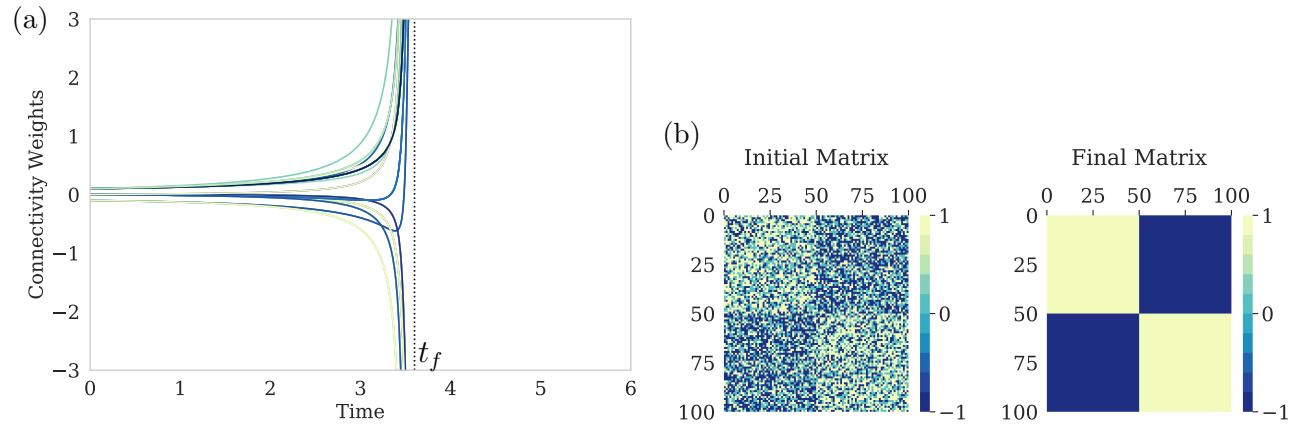

Figure 4.6: Structural balance evolution of a network with community structure. (a) Connectivity network weights $Y_{ij}(t)$ over time evolved by Eq. (4.69) from an initial network generated by a stochastic block model (with values rescaled by $1/N$). (b) Initial and final adjacency matrices. Initial matrix parameters $N = 100$, $d_{in} = 0.7$, $d_{out} = 0.7$, $p^+_{in} = 0.65$, and $p^+_{out} = 0.35$.

We will investigate the evolution of networks with community structure under the structural balance model (4.69). The initial connectivity matrix is taken to be proportional to an initial adjacency matrix $\mathbf{A}_0$ generated using the stochastic block model, in particular $\mathbf{Y}(0) = \mathbf{A}_0/N$. Figure 4.6(a) shows an example of the evolution of network connectivity values over time from a $\mathbf{Y}(0)$ corresponding to the $\mathbf{A}_0$ shown on the left in Fig. 4.6(b). We see that the $Y_{ij} \to \pm\infty$ and the final adjacency matrix $\mathbf{A}_f$ on the right shows the split into two factions with positive ties within each faction and negative ties between factions. It is given by the outer product

(4.71) with $\mathbf{u}_1 = \mathbf{u}_C$ so that the factions correspond with the identity blocks A and B.

4.7 Structural Balance Behavioral Regimes

As the final structure of these dynamical networks is dominated by the initial network's leading eigenvector, we can determine the extent to which networks in our parameter space will become assortative or homogeneous using our transition formulas derived above. First we treat a special case before exploring more general parameters.

4.7.1 Ingroup affinity equals outgroup animosity

We consider the simple case in which the ingroup affinity is set equal to the outgroup animosity, $p_{in}^{+} = p_{out}^{-}$. We also make the simplification $d_{in} = d_{out} = d$. For this case, $\langle A_{in} \rangle = -\langle A_{out} \rangle$ so that $\mu = 0$, i.e., there is no homogeneous signal, and $\nu = -\langle A_{out} \rangle$. Using $\langle A_{out} \rangle = d(1 - 2p_{out}^{-})$ in Eq. (4.63), we solve for the critical outgroup animosity

$$p_{out}^{-} = \frac{1}{2}\left(1 + \sqrt{\frac{1}{dN}}\right). \tag{4.72}$$

Note that we only use the positive sign from (4.63) since it is the assortative transition, not the disassortative one, that involves the leading eigenvector.

Figure 4.7 shows the alignment between the dynamical regimes evolved by the structural balance model and the community structure of the initial network. Panel (a) plots the assortativity r defined by Eq. (4.6) of the final adjacency matrix, $\mathbf{A}_f$ as averaged over 400 initial networks generated by the stochastic block model at each point in the d and p_{out}^{-} parameter space. The yellow region represents the fully assortative outcome where the system evolves into two factions corresponding to the identities defined by the stochastic block model. In the blue region where $r \approx 0$, the two final factions are well mixed by identity. We see that the boundary between these two regions is in good accord with Eq. (4.72). As the network density increases, the assortative regime is observed to grow, extending down to smaller outgroup animosity values. Panels (b) and (c) plot, respectively, the spectrum and first eigenvector of the initial adjacency matrix, $\mathbf{A}_0$ for a constant density value. The transition in

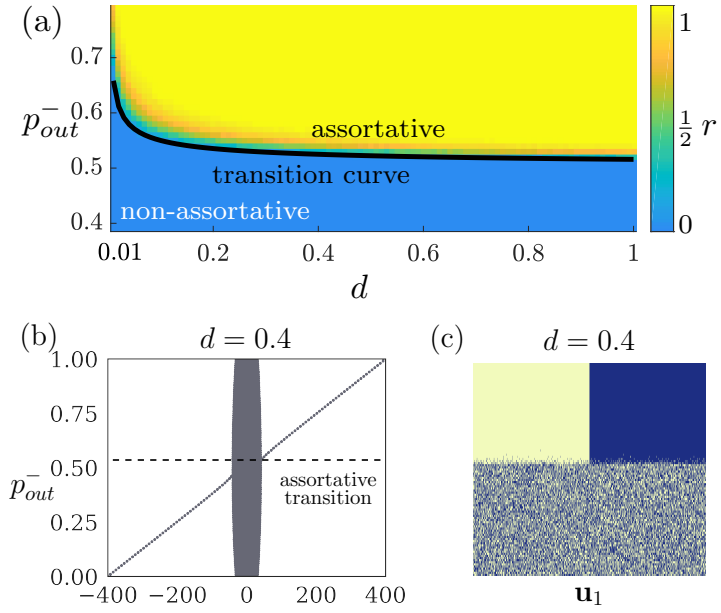

Figure 4.7: (a) Assortativity of final adjacency matrix as a function of d and p_{out}^- for network size $N = 1000$ averaged over 400 simulations. Solid curve is the theoretical transition boundary given by Eq. (4.72). Note that the theoretical curve corresponds to the early part of the transition in which r just begins to rise whereas the more visually distinctive yellow-cyan interface marks the middle of the transition. (b) Initial adjacency matrix spectra for $d = 0.4$ and increasing p_{out}^-. Dashed line indicates theoretical transition point. (c) Component signs for the leading eigenvector $\mathbf{u}_1$ of the initial adjacency matrix. The disassortative transition is not dynamically relevant so only the upper part of the p_{out}^- scale is plotted in (a).

the structural balance dynamics mirrors the behavior of the first eigenvector which undergoes an assortative detectability transition at $p_{out}^{-} = 0.525$. These plots also confirm that the leading eigenvector is never homogeneous and so one does not expect to observe the single faction outcome in this case (being extremely improbable).

4.7.2 General parameter conditions

We now analyze the final states of networks generated using more general parameter conditions. We plot the behavior of the signal transition curves in the two-dimensional parameter space defined by p_{in}^{+} and p_{out}^{-} for fixed values of d_{in}/d_{out}.

As the homogeneous signal was irrelevant in the previous case, we needed only plot the assortativity r. However, the prosocial transition will occur in general and so we must measure the extent to which nodes can be found in one large group. We define the homogeneity h as the fraction of all nodes that can be assigned to a single group by virtue of having a common sign in the leading eigenvector of the adjacency matrix. When all nodes have a common sign, they are positively connected to all other nodes so that $h = 1$, while when nodes are divided into two equal factions, the homogeneity assumes its minimum value, $h = 1/2$.

The top plots of Fig. 4.8 show the assortativity and homogeneity of the final adjacency matrix evolved by the structural balance dynamics in the parameter space defined by the ingroup affinity and outgroup animosity. They can then be linearly combined to effect their joint visualization as shown in the bottom plot. The assortative transition boundary predicted by Eq. (4.66) separates the assortative from non-assortative two-faction states while the prosocial theoretical boundary of Eq. (4.68) separates homogeneous single-faction states from the non-assortative two-faction states. These regimes relate to the initial network spectrum as follows: the blue region is where the homogeneous eigenvalue is both the largest eigenvalue and outside the main band; the cyan region is where the leading eigenvalue is part of the main band; and the yellow is where the contrast eigenvalue is largest and outside the main band.

The horizontal yellow-blue interface observed in Fig. 4.8(c) for larger p_{in}^{+} values corresponds

to the signal crossing transition in which the homogeneous and contrast eigenvectors exchange places, which occurs outside the noise band (see Fig. 4.4). Equating the contrast and homogeneous eigenvalue expressions, (4.57) and (4.58), we find that the transition occurs when $\nu = \mu$, which implies that $\langle A_{out} \rangle = 0$ or equivalently $p_{out}^{-} = 0.5$.

Figure 4.9 shows how the density ratio d_{in}/d_{out} and overall network density $d = (d_{in} + d_{out})/2$ affect the assortative and prosocial transition curves. As d_{in}/d_{out} increases, the transition curves become steeper, implying that denser regions of the connectivity network have more influence on the community structure than sparse regions. Figure 4.10 shows how network size affects the location and shape of the assortative and homogeneous transitions. As N increases, the transitions become sharper and more closely aligned with the theoretical prediction for the critical value of p_{out}^{-}.

4.8 Discussion

In this section, we first make some observations concerning our results on community and, more broadly, network structure, a subject of relevance to both unsigned and signed networks. We then turn to structural balance dynamics, an intrinsically signed network avenue of research. As it is particularly applicable to social systems, we speculate as to connections between our results and the dynamics of conflicts.

Our results can be applied to unsigned networks by taking the ingroup and outgroup animosities to be zero so that $p_{in}^{+} = 1$ and $p_{out}^{+} = 1$. In this case, our expressions for the homogeneous and contrast eigenvalues can be reduced to the sparse-limit forms reported in Refs. [152, 222] for the two leading eigenvalues of the adjacency matrix. These can be obtained by neglecting the μ^2 and ν^2 contributions to σ^2 in (4.24) and then inserting into (4.58) and (4.57) for the homogeneous and contrast eigenvalues respectively. This yields the expressions (outside the noise band),

$$\lambda_H = \frac{N(d_{in} + d_{out})}{2} + 1 \tag{4.73}$$

$$\lambda_C = \frac{N(d_{in} - d_{out})}{2} + \frac{d_{in} + d_{out}}{d_{in} - d_{out}}. \tag{4.74}$$

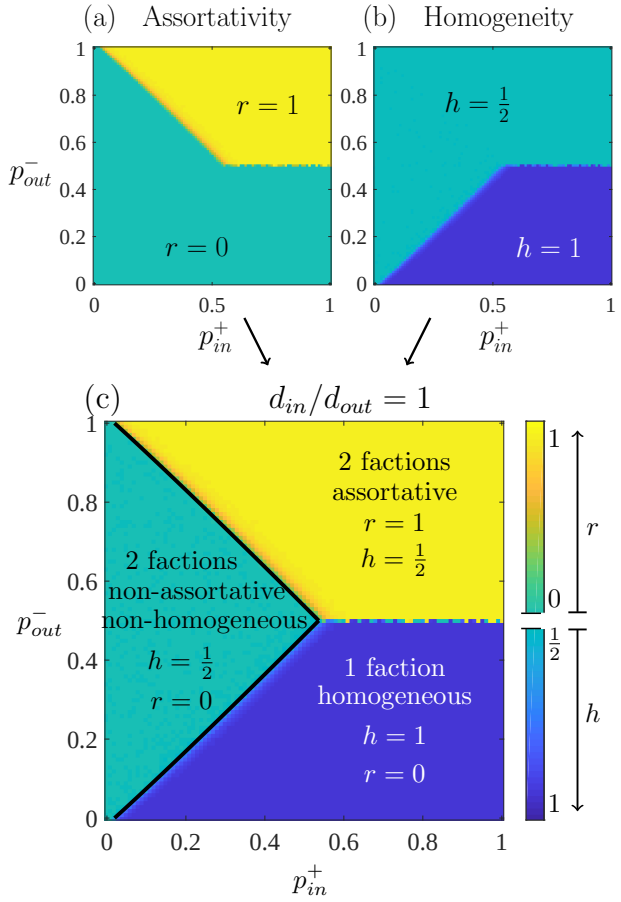

Figure 4.8: (a) Assortativity and (b) homogeneity of final network states evolved by structural balance model (4.69) as a function of ingroup affinity and outgroup animosity. (c) Assortativity and homogeneity are mapped using the measure $z = r - 2h + 1$ to generate the joint heat map. For convenience, two separate color bar scales are shown instead of z. The upper black curve indicates the assortative transition boundary, Eq. (4.66), while the lower black curve indicates the homogeneous transition boundary, Eq. (4.68). Heatmap values generated by averaging over 4 simulations for parameters $d_{in} = d_{out} = 0.45$ and $N = 1000$.

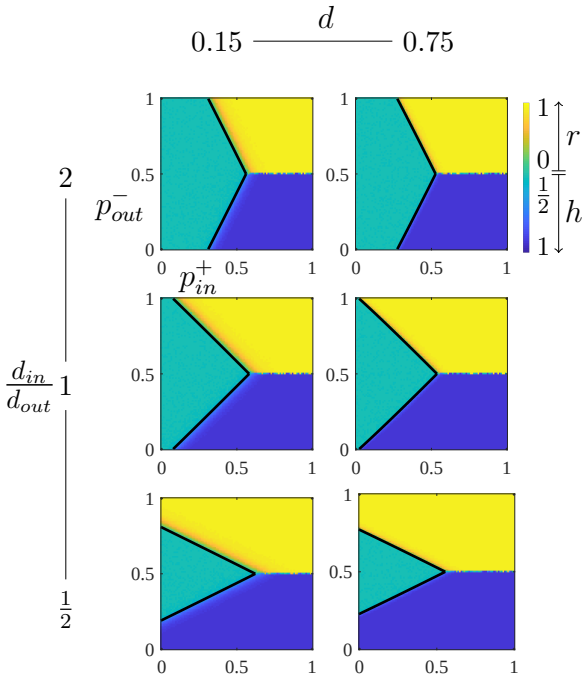

Figure 4.9: Final state heat maps for increasing network density $d = (d_{in} + d_{out})/2$ and decreasing density ratio d_{in}/d_{out}. The assortativity and homogeneity are integrated via the z metric (see Fig. 4.8). Heatmap values generated by averaging over 4 simulations with $N = 1000$.

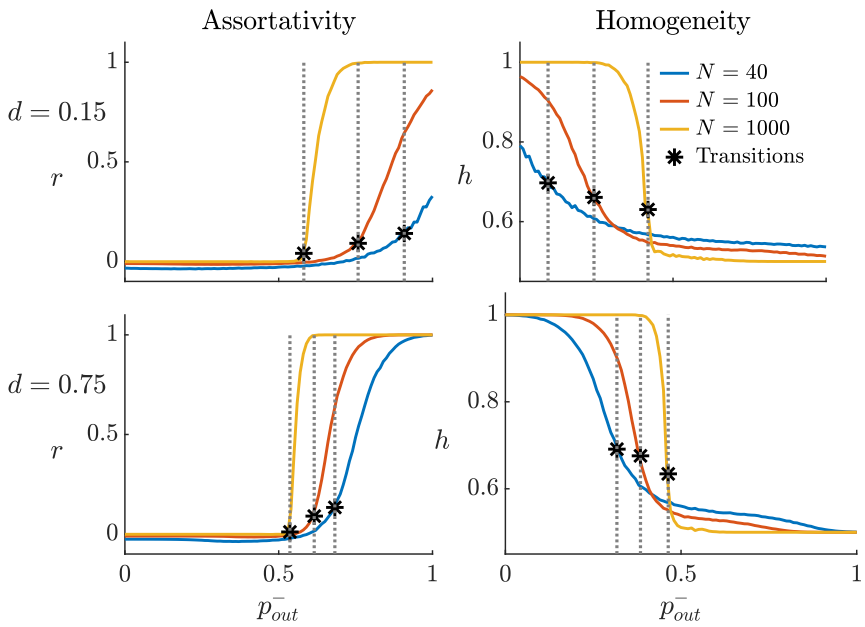

Figure 4.10: Assortativity and homogeneity as a function of p_{out}^{-} for networks of increasing size. Black stars mark the predicted values of p_{out}^{-} at the transition points as given by Eqs. (4.66) and (4.68) for the assortative and prosocial transitions respectively. Parameter values are $p_{in}^{+} = 0.5$, $d = 0.15,\ 0.75$, and $N = 40, 100, 1000$ averaged over 2000, 1000, and 10 trials respectively.

These forms, however, do not manifest the μ, ν interchange symmetry, an essential feature of the signed network case.

An important difference between unsigned and signed networks concerns which of the two leading eigenvectors outside the main spectral band may signify community structure. The above equations can be used to show that we can never observe the contrast eigenvalue as being larger than the homogeneous eigenvalue in the unsigned case. The sign of the difference, $\lambda_H - \lambda_C$, depends on the sign of $N - 2/(d_{in} - d_{out})$. The second term is equal to $1/\nu$, so that we must have $\nu < 1/N$ in order for the contrast eigenvalue to exceed the homogeneous eigenvalue. This condition upon ν in conjunction with Eq. (4.63) for the assortative transition in the sparse limit, which sets the minimum value of ν for λ_C to appear outside the noise band, then necessitates $\mu = (d_{in} + d_{out})/2 < 1/N$. However, this regime is below the threshold, $\mu_* = 1/N$, for the prosocial transition in the sparse limit as obtained from Eq. (4.67). Further, the righthand sides of (4.63) and (4.67) are the same so that the assortative and prosocial transitions occur at the same critical value, $\nu_* = \mu_*$. For unsigned networks, $\nu \leq \mu$ and so if $\mu < \mu_*$, then $\nu < \nu_*$. Therefore, if λ_H is within the noise band then so must λ_C and hence we can never observe $\lambda_C > \lambda_H$.

Consequently, in unsigned networks, assortative community structure is represented by the second eigenvalue of the adjacency matrix (when past the detectability threshold) but not the first. In signed networks, the ordering of μ and ν is not restricted and so the first eigenvalue may signify the community structure while the second, if above the prosocial transition, signifies the homogeneous structure. Relatedly, while the number of outlying eigenvalues is equal to the number of communities in unsigned networks [36], this need not be the case in signed networks. For instance, the case of equal ingroup affinity and outgroup animosity treated in Sec. 4.7 has only one outlying eigenvalue but two communities.

The sociality transitions, which involve the homogeneous eigenvector, bear upon the question of whether a network exhibits a propensity toward positive versus negative tie formation, a question that is unique to signed networks. In unsigned networks, the prosocial transition is present but its significance corresponds to the emergence of a giant connected

component; Eq. (4.67) yields a transition condition of $N(d_{in} + d_{out})/2 = 1$ in accordance with the result for a giant component with two-community structure noted in Sec. 4.2.1. Typically, community structure is taken to connote the existence of multiple communities as it is linked to the community detection problem and the assignment of nodes to communities. The sociality transitions, which are not relevant to the community assignment problem, involve network structure more generally rather than community structure per se. However, the sociality transitions do lend themselves to a sense by which a signed network can be viewed as forming a single community: the existence of a global tendency, irrespective of ingroup and outgroup distinctions, toward the formation of positive versus negative ties. The prosocial transition provides a spectral signature for concluding that the network forms a single community in which relationships are generally friendly or cooperative. Conversely, the antisocial transition provides a signature of a single (anti)community marked by hostility, a Hobbesian state of all against all. The intermediate case, where the homogeneous eigenvalue does not appear outside the noise band, can be taken as marking the absence of a single community with a definitive inclination toward positive or negative tie formation.

Turning to the dynamics, taking networks generated by the stochastic block model and evolving them under the structural balance model entails a dramatic shift in the processes governing network evolution (beyond the change from stochastic to deterministic): a process in which tie formation in a dyad depends only upon its ingroup or outgroup identity is replaced by one where the friendly or hostile relationships among mutual neighbors drives tie value evolution regardless of identity. One rationale for such a shift can be provided by assuming that there is a qualitative change in the nature of the interactions. For instance, hostile relationships characterized by insulting words or gestures may be replaced by physical violence. A second rationale could involve, not a change in the nature of interactions, but a growing awareness that hostile interactions have the potential to become much more prevalent. For example, in a country or region containing two broad identity types, such as ethnicity or ideology, in which individuals or small gangs sporadically clash (either within or across identity lines), a sudden collapse of the central government may lead to a growing sense of

looming systemic violence. In either of these rationales, nodes are motivated to seek and maintain allies so that another node's status as the enemy of an enemy or friend of a friend becomes a crucial determinant in relationship formation and evolution.

Our results may inform debates about the interplay of identity and power in conflicts under anarchy consisting of many actors such as insurgencies, civil wars, and international relations. For ethnic conflict, the shift in models discussed above is supported by the observation that the turn from non-violent to violent conflict represents a qualitative change in dynamics [21]. In the literature on civil wars and ethnic violence, some theories stress mechanisms in which ethnic or religious identity plays an intrinsic role in producing high levels of polarization and violence along identity lines while other theories stress the role of micro-processes of conflict among local actors rather than a pre-existing identity schism [21, 99, 34]. As node identity plays no role in the network evolution, structural balance dynamics is consistent with the latter view. However, the sharp transition to the assortative state shows that the dynamics can lock in initial differences in identity-driven community structure even when they are not large, a behavior consistent with the observation that the polarization and violence in ethnic civil wars often appears to be disproportionate to the initial level of ethnic tension. But the existence of the non-assortative regime implies that identity polarization will not arise when the initial structure is sufficiently weak. Thus, conflict takes on an essentially binary nature in that it is either completely polarized by identity or not at all. Additionally, it has been argued that, contrary to some theories, there is no inherent difference in the dynamics between ideological and ethnic civil wars in terms of their potential for polarization and violence [99]. Our results are consistent with such a claim as it is the initial relationships that matter regardless of whether they are due to similar ethnicity or similar ideology.

4.9 Conclusion

This paper has contributed to two distinct areas of signed network research — community structure and structural balance theory, linking them via the impact of the former upon the latter. We have elucidated the spectrum of unweighted and undirected signed networks

generated by a two-community stochastic block model via two independent methods, perturbation analysis and, in the Appendices, a random matrix theory treatment that extends prior work on unsigned networks [152, 222]. The expected matrix, $\langle \mathbf{A} \rangle$, in the block model can be decomposed into two signals — a homogeneous eigenvector, $\mathbf{u}_H$, related to the expected tie value, μ, over the network and a contrast eigenvector, $\mathbf{u}_C$, related to the half-difference, ν, between the expected ingroup and outgroup tie values and which encodes the community structure. These signal eigenvectors exhibit transitions at the points where they merge with the main spectral band associated with the noise produced by the zero-mean random matrix $\mathbf{X}$. There are four potential transitions corresponding to the intersections of the two signals with the positive and negative edges of the main band. For the contrast eigenvector, these intersections induce the assortative and disassortative transitions respectively and mark changes in community detectability. The homogeneous eigenvector undergoes sociality transitions, prosocial and antisocial, in which emergence from the noise band signifies an overall tendency toward the formation of cooperative or conflictual relationships respectively with other nodes.

We derived analytical expressions for the signal eigenvalues in the presence of the noise by performing a perturbation expansion in which the contributions from the noise $\mathbf{X}$ were treated as small corrections to the eigenvalues of $\langle \mathbf{A} \rangle$. The expressions, (4.26) and (4.27), reveal a second-order correction proportional to the average tie variance and are symmetric under the interchange of μ and ν. The same expressions are derived in the Appendices using random matrix theory along with the formula for the main band edges that is a straightforward modification of Wigner's semicircle law. The transition conditions, (4.60) and (4.61), were determined by equating the signal eigenvalues to the band edge eigenvalues.

We investigated structural balance dynamics in the presence of initial community structure generated by the two-identity stochastic block model. These dynamics completely connect all nodes and allow for three broad regimes of final states: an assortative regime in which two hostile factions emerge that completely align with the two identity blocks; a non-assortative regime in which the two final factions are randomly composed with respect to identity; and a

homogeneous regime consisting of a single harmonious faction with only positive ties. Since the dominant eigenvector of the initial network drives its structural balance dynamics and determines its final state, our spectral analysis allows us to chart the parameter conditions under which each of these states will emerge. The dynamical ascendance of the leading eigenvector implies that the regime boundaries occur where any two of the homogeneous signal, the contrast signal, and the noise band edge exchange places as the first eigenvalue. The assortative transition marks the boundary between the non-assortative and assortative regimes and the prosocial transition divides the non-assortative and homogeneous regimes. The boundary between the homogeneous and assortative regimes represents a reversal in the ranking of the homogeneous and contrast eigenvalues rather than a transition involving the noise band. The theoretically-predicted boundaries were found to agree with the simulation results obtained by solving the structural balance model over many random initial networks.

Finally, we note a few potential directions for future research. As with the unsigned case, spectral analysis of signed networks with community structure could be extended to systems with multiple communities, directed ties, and more realistic network statistics such as non-uniform degree distributions. The structural balance model we used is very simple and, problematically, leads to tie strengths which blow up in finite time. Accordingly, the extent to which the dynamical transitions we have identified persist for more realistic implementations of structural balance dynamics should be explored. More empirical work is also needed to understand the conditions under which real networks can be reasonably modeled by structural balance dynamics or variants thereof.

Chapter 5

MODELING THE DYNAMICS OF CONFLICT AND SYSTEMIC WAR USING NETWORKS

5.1 Introduction

The social sciences attempt to understand the most important forces driving international relations and how conflicts between states lead to widespread wars. Models for interactions between states are dependent both on sociological considerations as well as the mathematical approach taken. We develop a dynamical systems perspective for modeling the network-level state interactions that account for long periods of stability, escalations leading to systemic wars, and transitions back to relative peace. Dynamical systems models of the interaction between states can help identify mechanisms for system destabilization as well as targets for reinforcing stability. These models can be used to determine the risk for minor, localized conflicts to escalate and spread extensively to other nodes in the network. Moreover, they provide a rich set of diagnostic tools for the analysis of escalating conflicts which result in the outbreak of widespread wars.

Dynamical systems models have been useful in understanding and predicting the dynamics of networks in a diversity of fields such as biology, engineering, and sociology [198]. While statistical models are useful in classifying and predicting phenomena, statistical models often fail to provide explicit mechanistic explanations [143, 203, 175]; dynamical systems models, on the other hand, can elucidate the mechanisms responsible for the observed underlying phenomena [209, 146]. In the context of state-level conflicts that lead to widespread war, the aim is to create a mechanistic model that describes how interactions between the various states leads to network level observations of the escalation in tensions that ultimately result in the breakdown of communications and onset of war. Building upon previous models of

social network dynamics, we create a dynamical systems model for the outbreak of war in a network of interacting states. This dynamical systems model leverages ideas and theories about what contributes to the outbreak of war in a dynamical systems framework and builds off of previous dynamical systems models of structural balance by adding stability terms to the system.

5.1.1 Critical transitions between qualitative states and bifurcation theory

Nonlinear dynamics are pervasive in describing phenomena in the biological and social sciences [24, 115]. Using bifurcation theory, the range of a nonlinear system's qualitative behavior can be determined [110, 180]. SIR models describes the spread of infections diseases in a population where critical bifurcations in the system parameters determine whether a disease will spread throughout an entire population [106, 83]. A critical bifurcation in a neuron's voltage dynamics determine when the neuron will transition from a equilibrium voltage to a spiking state [64, 153, 142]. In social networks, critical thresholds determine whether a network under structural balance dynamics will reach a harmonious final state or a factious final state that is random or assortative with respect to initial community structure [137, 147]. Critical transitions appear in the dynamics of many other systems such as ecosystems, climate change, and finance [110, 3, 121, 139]. The dynamics of a system near a critical bifurcation exhibits certain signatures, such as critical slowing down, which allows for estimations of how close the dynamics are from a critical transition leading to a significant qualitative shift in activity [142, 180].

Biological and social systems often exhibit nonlinear dynamics that have multiple stable states between which the system can transition [170, 72, 105, 146, 149]. Stochasticity regulates transcription in bistable gene-regulatory networks, allowing the network to transition between activity states [105]. The nemotode *C. elegans* also has two stable states in its nonlinear neural dynamics which correspond to distinct behaviors [102]. It is hypothesized that transient control signals allow the nemotode to transition between stable states and therefore modify its behavior [146]. Transitions between stable states in artificial networks, such as memories in a

Hopfield neural network, or stable states in random dynamical systems can also be controlled with feed-forward control signals and bifurcation theory [149]. We use bifurcation and control theory to understand the range of dynamics that sociopolitical systems can exhibit as well as how a social network in a low conflict state can transition to a high conflict state under the right conditions.

5.1.2 Dimensionality reduction for network dynamics

High-dimensional dynamical systems can be difficult to model, analyze, and control. Fortunately, high-dimensional dynamics often occur in a low-dimensional subspace, which allows for the construction of reduced order models [16]. Dynamic mode decomposition (DMD) provides a data-driven method for modeling high-dimensional dynamics with an equation-free, low-dimensional model [115]. The Sparse Identification of Nonlinear Dynamics (SINDy) algorithm, identifies the equations of a system's underlying dynamics and works particularly well after a dimension reduction technique, such as PCA, is applied to the data [25, 149]. Low-dimensional spaces are also useful for applying control to a high-dimensional system [24, 98, 149].

When considering network edges, eigenvector decomposition provides a useful, low-dimensional subspace to analyze edge dynamics [137, 147]. In network science, spectral analysis has long been used to understand the dominant structures present in a network such as community structure [158, 37], and can be used to differentiate signal from noise in network edges [152, 147, 218]. We leverage eigenvector decomposition to understand network edge dynamics in a low-dimensional space. We derive the dynamics of the eigenvalues and eigenvectors from the dynamical system of edges and compute bifurcations to the system from the eigenvalue dynamics.

5.1.3 International relations and theories for the causes of war

A significant amount of attention in international relations has been devoted to understanding why states go to war and how they can be averted [95, 122, 214, 188]. Claims about the

primary causes of war vary widely depending on the approach taken — realist, liberal, constructivist, or Maxist — and the level of analysis employed — the examination of within-state features, between-state features, or features of the network as a whole [123, 124, 53]. Theories about the causes of war emphasize a range of different factors including the balance of power, trade, diplomacy, technology, geography, social economics, the desire for expansion, and self-preservation.

From a realist perspective, the international system is anarchic and its dynamics are driven primarily by the self-preserving behavior of states which leads to power balancing dynamics, bandwagoning, structural balance dynamics, and security dilemmas [214, 92, 213, 40]. Systemic wars that arise from the expansionary aspirations of a rising state, such as World War II or the Napoleonic Wars, are heavily driven by the hegemonic ambitions of a single state and are characterized as "wars of mobilization" [144]. Many systemic wars, however, do no arise from state expansionism, and are instead the result of a complex interaction of factors including network structure, security dilemmas, and triggering crises [144, 92]. Wars typically classified as structural wars include World War I [144, 95], the Thirty Years' War [144], and the Peloponnesian Wars [95, 96]. Using game theory, Jervis argues that war can develop among states whose only objective is self-preservation, given the right circumstances [92]. This, unfortunately, indicates that curbing the ambitions of expansionary state is not enough to prevent systemic, structural wars.

5.1.4 Review of predictive models of war

Researchers have modeled anarchic state dynamics using dynamical systems, stochastic models, and game-theory models [92, 143, 144, 204]. Jervis [92] uses game theory models to determine whether countries will cooperate in different circumstances, if arms races will develop among nations, and the maximum level of security that is possible for a state without threatening other states, setting off a cascade that results in a lower level of security. While this perspective is useful for understanding dyadic relationships, it does not account for the entire network structure which often heavily influences tie dynamics.

At the network level, researchers have used statistical features of the international system's network to theorize about the causes of and predict the onset of systemic wars. Midlarsky [143] associates the shift away from an equilibrium in conflict levels with the onset of systemic war. He theorizes the mechanism by which war escalates is the increasing interdependence of conflict and uses conflict equilibrium levels to predict systemic war in Europe. He also finds that the hierarchical structure of the network influences the stability of the international system [144]. Thompson [203], comparing Waltz's analysis of systemic stability and Modelski's long cycle of global leadership, finds that increased mulipolarity in the international system is correlated with more frequent warfare. This indicates that multipolar systems may be inherently more unstable than bipolar or unipolar systems. Thompson [204] also notes the effects that catalysts can play in triggering a system that has a ripe rivalry field. He theorizes that nonlinear interactions in the international system can lead to systemic wars that are akin to systemic accidents that happen in mechanical or engineered systems.

Other researchers have used structural balance to model and predict the onset of war. Antal *et al.* [5] shows how leading up to World War I, ties in the international system evolved to a more balance state, and could have been a force contributing to the war. Marvel *et al.* [137] use a continuous-time model of structural balance to predict the WWII factions using pre-WWII international sytems network data. We incorporate escalatory structural balacing forces into a dynamical systems model of tie dynamics and show how, under the right conditions, this balancing force can destabilize a network at equilibrium.

5.1.5 Previous models of structural balance

Structural balance theory posits that nodes in social systems are driven to form balanced relationships among other nodes [33]. Balanced triads consist of an odd number of positive ties while imbalance triads consist of an even number of positive ties. Evidence for balancing forces have been found in networks of people, states, and wild mammals [107, 50, 89]. In gang networks, inter-gang violence increases among gangs that are a part of imbalanced triads [155]. Marvel *et al.* [137] characterized signed edge weights, $\mathbf{X}(t) \in \mathbb{R}^{n \times n}$, evolving under the

influence of structural balance dynamics with the dynamical system,

$$\frac{d\mathbf{X}}{dt} = \mathbf{X}^2. \tag{5.1}$$

For symmetric matrices, the network is guaranteed to separate into two polarizing factions or a single harmonious faction, determined by the dominant eigenvector of the initial value. For unsymmetric matrices, the system evolves to four factions instead of two [210]. Marvel *et al.*'s [137] model, and a variation of it [207], have been used to predict the factional membership of nodes after a network has been subjected to balancing forces such as the two factions formed in World War I, World War II, and Zachary's karate club.

Gao *et al.* [70] has investigated the origins of balancing forces with respect to nodal interactions, and has considered how to control structural balance edge dynamics using node dynamics [71], with the assumption that the dynamics of nodes and edges are interdependent. Wongkaew *et al.* [221] has considered how to control structural balance dynamics by introducing a 'leader' to the system, while Summers and Shames [199] have evaluated the control abilities of existing nodes in a network by changing ties.

Unfortunately, real networks typically only display structural balance dynamics as characterized in Eq. 5.1 sporadically and transiently, if at all. Healy and Stein [81] found evidence for an increase in balance among the great powers of Europe during the Bismarkian era, and Antal *et al.* [5] showed the great powers increased their structural balance leading up to WWI (1872-1907). However, over longer stretches of time, balance levels in the international system have fluctuated by several measures [52, 29]. Despite the fluctuations, there is still evidence that the international system has remained consistently more balance than null models [107].

5.1.6 Limitations to current models and our contribution

While the model in Eq. (5.1) is able to determine the factions that emerge from a network solely under structural balance dynamics, it is not able to predict the onset of such feed-forward escalation in dynamics or what features could trigger escalation, as the dynamical system does not capture the stable, pre-escalatory state which dominates for long stretches of

time before and after systemic wars. The model also cannot describe de-escalation dynamics, from systemic war back to a low-conflict state.

Previous statistical models have found correlates to the onset of systemic war and posit that certain qualitative interactions are likely to produce conflict escalation. We attempt to ground the predictive statistical features found in previous work in a dynamical systems mechanism for the production of systemic war. We aim to build from the dynamical systems model of conflict escalation in Eq. 5.1 to create a model that describes stability in a peaceful, low-conflict state, transitions to a warring state via balance dynamics, and de-escalation. We create such a model by incorporating stable state terms and limits to the structural balance dynamics. With a dynamical systems model that describes a peaceful state, an escalatory transition to a state of widespread war, and a de-escalatory transition back to a state of peace, we can evaluate the systemic features that move to system closer or farther from escalatory and de-escalatory dynamics. We can also use this dynamical systems model to ground a range of qualitative theories about escalatory risk factors in a mechanistic model.

5.2 Structural balance and stability as a model for international relations

We present a model for signed social network edge dynamics that incorporates three significant influences on the international system — natural dyadic relationships, structural balance dynamics, and short-timescale perturbations. Our model contains a low-conflict equilibrium that corresponds to relative peace and a high-conflict equilibrium that corresponds to systemic war; bifurcations and perturbations to the system can induce the network to transition between the low and high conflict state and analysis of these transitions could be useful for understanding, preventing, and predicting the outbreak of systemic wars.

5.2.1 Modeling tie dynamics in social systems

We built a dynamical system for edges in social networks that represents the interplay of three key forces at work in signed social networks — (1) natural, long-timescale affinity or antipathy between nodes, (2) structural balance dynamics, and (3) short-timescale perturbations that

take the form of transient conflicts or cooperation between nodes. With this dynamical system we aim to model how widespread conflict can escalate through a stable, relatively peaceful network. We show that this conflict spread can be directly related to the global structure of the network which can in turn be characterized and analyzed via spectral analysis. Using spectral analysis, we analytically compute the relationship between the level of edge stability in the network, the strength of balancing forces, and the effect of perturbations. From this we derive stability bifurcations over which conflict will escalate and spread due to the network structure.

Using this model and the derived relationship between stability, balance, and perturbations, we can do the following: (1) compute perturbations that will trigger widespread conflict for a give system, (2) determine optimal stability procedures, (3) determine global network features across multiple networks that put them at risk for network destabilization, and (4) approximate how close the network is to destabilization using critical bifurcation theory.

5.2.2 Social network edge dynamics model

The dynamical system we propose for signed social network edges contains a stability term, a bounded structural balance term, and a perturbation term. The dynamics for edge X_{ij} between nodes i and j in a network containing n nodes is:

$$\frac{dX_{ij}}{dt} = -\beta(X_{ij} - X_{0ij}) + L\tanh\left(\gamma\sum_{k=1}^{n} X_{ik}X_{kj}\right) + X_{Pij}(t) \tag{5.2}$$

where $-\beta(X_{ij} - X_{0ij})$ is the stability term, $L\tanh(\gamma\sum_{k=1}^{n} X_{ik}X_{kj})$ is the bounded structural balance term, and $X_{Pij}(t)$ represents fast-timescale perturbations. X_{0ij} is the edge's natural stable state and β is the strength of the stability term. Without the presence of structural balance or perturbations, $X_{ij} \to X_{0ij}$. The balancing term $L\tanh(\gamma\sum_{k=1}^{n} X_{ik}X_{kj})$ contains parameter γ which represents edge sensitivity to balancing forces. Parameter L and the tanh function limit the maximal force structural balance can exert on the system. The bounded structural balance term approximates unbounded structural balance introduced

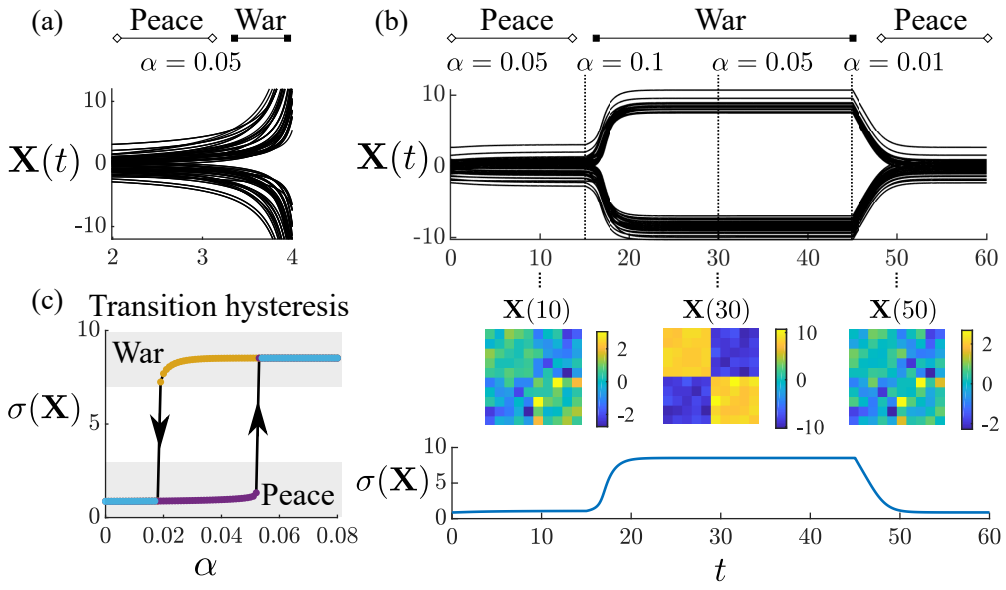

Figure 5.1: (a) Pure structural balance model, Eq. 5.1, has unbound dynamics [137]. (b) Dynamics of structural balance model with stability, (Eq. 5.2). Snapshots of network ties are taken at $t = 10$, 30, and 50. The standard deviation of network ties, $\sigma(\mathbf{X})$, is measured over time. (c) Standard deviation of network ties as a function of the structural balance parameter α as well as the initial state (war or peace).

by Marvel *et al.* [137], $\alpha \sum_{k=1}^{n} X_{ik} X_{kj}$ (Eq. 5.1), where $\alpha = \gamma L$. The structural balance strength, α, is the product of the sensitivity strength and the balancing force upper bound. As $L \to \infty$, $L \tanh(\alpha \sum_{k=1}^{n} X_{ik} X_{kj} / L) \to \alpha \sum_{k=1}^{n} X_{ik} X_{kj}$, the bounded structural balance term approaches the unbounded dynamics. The term $X_{Pij}(t)$ contains transient perturbations to the system caused by fast-timescale events such as a militarized conflict between two states.

The dynamical system implicates that each edge has a natural stable value, X_{0ij} determined by the individual dyadic relationship, and that structural balance forces and transient perturbations produce a new equilibrium value that deviates from X_{0ij}. A number of factors may contribute to determining the naturally stable state such as common ideology or ethnicity,

close economic ties, or mutual security. The structural balance term enhances the dominant eigenvector of the stable solution that exists in the absence of balance dynamics. There is a limit to how much force structural balance dynamics can exert on each edge value thus the need for a bound on the structural balance term.

While the stability term typically brings the system to a low-conflict equilibrium state near X_{0ij}, if structural balance or the transient perturbation to the system are large enough, the edges may no longer have a stable solution near X_{0ij}. After a critical bifurcation in the system parameters, the dynanical system loses stability near X_{0ij} and the system transitions to a factious, high-conflict state. The matrix containing the edges becomes low-rank with the dominant eigenvector of the system determing the sign of all ties. In social networks, this critical bifurcation can be related to the phenomena of feed-forward escalatory conflict that spreads throughout an entire network, requiring each node to participate in the global conflict and take one of two sides. Once the network has stabilized at a high-conflict state, it requires a different bifurcation to bring the system back to a low-conflict state. A return to a low-conflict state can be achieved through a major reduction in the strength of the balancing parameter α.

Figure 5.1(b) shows the dynamics of Eq. 5.2 where the network starts in a peaceful state, transitions to a warring state, and then transitions back to a peaceful state. The structural balance strength, $\alpha = \gamma L$, is modified at times $t = 15$, 30, and 45 by changing the edge sensitivity to balance parameter, γ, and induces changes in the stability of the system, resulting in low/high conflict state transitions. Unlike the pure structural balance system (Eq. 5.1, Fig. 5.1(a)) which is unstable for all α, the model in Eq. 5.2, is stable for small α values, as is evident when $\alpha = 0.05$ (Fig. 5.1(b)). When α is increased to 0.1, the peaceful state destabilizes and the system transitions to the high-conflict, warring state (Fig. 5.1(b)). Once in the high conflict state, lowering α back to 0.05 does not bring the system back to a low-conflict state. α must be decreased even further, to 0.01, to induce the transition back to the peaceful state. This is an example of hysteresis, where a system's future state is determined not only by its parameter values but also by its current state. The standard

deviation of all tie values $\sigma(\mathbf{X})$ is an indicator of the system's current state, with a high standard deviation indicating the high-conflict state. Figure 5.1(c) shows $\sigma(\mathbf{X})$ for various α values and initial conditions. For $\alpha < 0.018$ only the peaceful state is stable and the network will converge to low-conflict for all initial conditions. For $\alpha \in (0.018, 0.053)$, both the peaceful and warring states are stable; the state to which the network converges is dependent on initial conditions. For $\alpha > 0.053$ only the warring state is stable and the network will converge to the high conflict state for all initial conditions. We will analyze these social network edge dynamics in a low-dimensional space and derive the critical bifurcations that determine the network's state.

5.2.3 Stability analysis

Our dynamical system of edges, Eq. 5.2, can be written succinctly as a matrix dynamical system

$$\frac{d\mathbf{X}}{dt} = -\beta(\mathbf{X} - \mathbf{X}_0) + L \tanh\left(\frac{\alpha \mathbf{X}^2}{L}\right) + \mathbf{X}_P(t) \tag{5.3}$$

where $\mathbf{X} \in \mathbb{R}^{n \times n}$ is a symmetric matrix containing all ties X_{ij} between n nodes in the network, $\mathbf{X}_0 \in \mathbb{R}^{n \times n}$ is a matrix of the dyadic stable states X_{0ij}, and $\mathbf{X}_P(t) \in \mathbb{R}^{n \times n}$ is a matrix of the transient perturbations $X_{Pij}(t)$.

We approximate Eq. 5.3 with the analytically tractable system

$$\frac{d\mathbf{X}}{dt} = -\beta(\mathbf{X} - \mathbf{X}_0) + \alpha \mathbf{X}^2 + \mathbf{X}_P(t). \tag{5.4}$$

This system has approximately the same fixed points and bifurcation values as Eq. 5.3 for large L, with the exception of the stable warring state. Approximating the edge dynamics with this equation allows us to perform eigenvector decomposition on the dynamical system. We can analyze the stability of the edge dynamics in the low-dimensional eigenvalue/eigenvector space and derive the system's critical bifurcations in terms of the edge matrices' eigenvalues.

Stable and unstable fixed points

To determine fixed points in the edge dynamics, we re-write Eq. 5.4 to represent the dyadic stable state and the transient pertribation matrix together as a modified transient stable state

$$\frac{d\mathbf{X}}{dt} = -\beta\left(\mathbf{X} - \left(\mathbf{X}_0 + \frac{\mathbf{X}_P(t)}{\beta}\right)\right) + \alpha\mathbf{X}^2 \tag{5.5}$$

where $\tilde{\mathbf{X}}_0(t) = \mathbf{X}_0 + \mathbf{X}_P(t)/\beta$ is the modified transient stable state, simplifying the represetation to

$$\frac{d\mathbf{X}}{dt} = -\beta\left(\mathbf{X} - \tilde{\mathbf{X}}_0(t)\right) + \alpha\mathbf{X}^2. \tag{5.6}$$

We can solve for stable solutions to this system as well as critical bifurcations that destabilize fixed points. Edge dynamics can be expressed in terms of the dynamics of its eigenvalues and eigenvectors which allows for a low-dimensional representation. Let the eigenvector decomposition of $\mathbf{X} = \mathbf{S}\Lambda\mathbf{S}^T$ and the eigenvector decomposition of $\tilde{\mathbf{X}}_0 = \tilde{\mathbf{S}}_0\tilde{\Lambda}_0\tilde{\mathbf{S}}_0^T$. Setting $\frac{d\mathbf{X}}{dt} = 0$ we derive

$$0 = -\beta\left(\mathbf{X} - \tilde{\mathbf{X}}_0(t)\right) + \alpha\mathbf{X}^2 \tag{5.7}$$

$$0 = \alpha\mathbf{X}^2 - \beta\mathbf{X} + \beta\tilde{\mathbf{X}}_0(t) \tag{5.8}$$

$$0 = \alpha\mathbf{S}\Lambda^2\mathbf{S}^T - \beta\mathbf{S}\Lambda\mathbf{S}^T + \beta\tilde{\mathbf{S}}_0\tilde{\Lambda}_0\tilde{\mathbf{S}}_0^T \tag{5.9}$$

$$\mathbf{S}(\alpha\Lambda^2 - \beta\Lambda)\mathbf{S}^T = -\beta\tilde{\mathbf{S}}_0\tilde{\Lambda}_0\tilde{\mathbf{S}}_0^T \tag{5.10}$$

We diagonalize the left hand side by multiplying by $\mathbf{S}^T$ and $\mathbf{S}$.

$$\alpha\Lambda^2 - \beta\Lambda = -\beta\mathbf{S}^T\tilde{\mathbf{S}}_0\tilde{\Lambda}_0\tilde{\mathbf{S}}_0^T\mathbf{S} \tag{5.11}$$

$$\alpha\Lambda^2 - \beta\Lambda = -\beta\mathbf{D}\tilde{\Lambda}_0\mathbf{D}^T \tag{5.12}$$

where $\mathbf{D} = \mathbf{S}^T\tilde{\mathbf{S}}_0$. If the left hand side is diagonal then the right hand side must also be diagonal. Therefore $\mathbf{D}$ must be the identity matrix, $\mathbf{S}^T\tilde{\mathbf{S}}_0 = I$. Therefore $\mathbf{S} = \tilde{\mathbf{S}}_0$. This implies that the eigenvectors of $\mathbf{X}$ will converge to the eigenvectors of $\tilde{\mathbf{X}}_0$. This allows us to diagonalize the matrix dynamical system equation at the fixed points and express the matrix fixed points in terms of its eigenvalues.

$$0 = \alpha\Lambda^2 - \beta\Lambda + \beta\tilde{\Lambda}_0(t) \tag{5.13}$$

$$0 = \alpha\lambda_i^2 - \beta\lambda_i + \beta\tilde{\lambda}_{0i}(t) \tag{5.14}$$

$$\lambda_i = \frac{\beta \pm \sqrt{\beta^2 - 4\alpha\beta\tilde{\lambda}_{0i}(t)}}{2\alpha}, \quad i = 1, ..., n \tag{5.15}$$

where λ_i are the eigenvalue fixed points of the edge matrix. The stable eigenvalues are

$$\lambda_{Si} = \frac{\beta - \sqrt{\beta^2 - 4\alpha\beta\tilde{\lambda}_{0i}(t)}}{2\alpha} \tag{5.16}$$

with the leading stable eigenvalue occuring at

$$\lambda_S = \frac{\beta - \sqrt{\beta^2 - 4\alpha\beta\tilde{\lambda}_{01}(t)}}{2\alpha} \tag{5.17}$$

The other solutions of Eq. 5.15 represent the eigenvalues of the unstable fixed point. The threshold at which the system becomes unstable is

$$\hat{\lambda}_U = \frac{\beta + \sqrt{\beta^2 - 4\alpha\beta\tilde{\lambda}_{01}(t)}}{2\alpha} \tag{5.18}$$

where $\hat{\lambda}_U$ is the stability threshold for Eq. 5.4 and is a lower bound for the true stability threshold, λ_U, of Eq. 5.3. As Eq. 5.3 approaches Eq. 5.4 with an increasingly large bound, L, the approximate estimate becomes exact

$$\lim_{L\to\infty} \lambda_U = \hat{\lambda}_U. \tag{5.19}$$

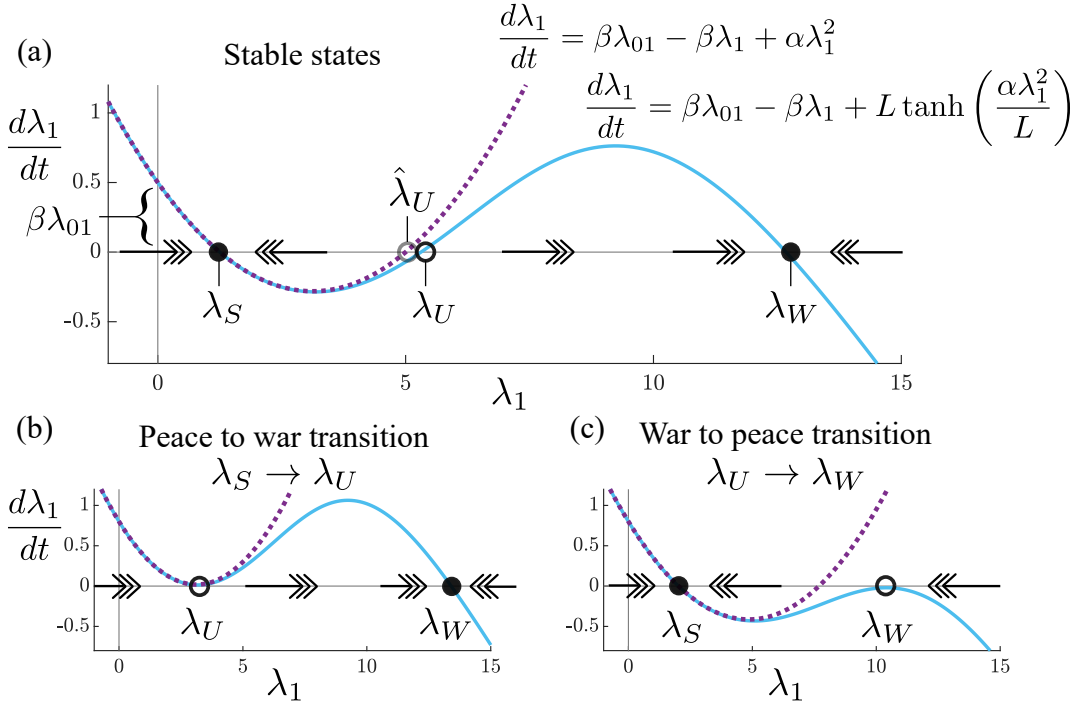

Figure 5.2: (a) Dynamics of the leading eigenvalue as a function of initial conditions. λ_S and λ_W are stable fixed points separated by an unstable fixed point λ_U. (b) A saddle-node bifurcation of the stable and unstable fixed points λ_S and λ_U underlies the peace to war transition. (c) A saddle-node bifurcation of λ_W and λ_U underlies the war to peace transition.

We can infer the dynamics of the edge matrix by analyzing the dynamics of its leading eigenvalue. Figure 5.2(a) shows $d\lambda_1/dt$, as a function of λ_1. λ_1 is stable at λ_S and unstable at $\hat{\lambda}_U$ which approximates the true unstable fixed point λ_U. The approximate and the bounded $d\lambda_1/dt$ deviate significantly after the stability threshold λ_U. The bounds on structural balance generate an additional stable state for λ_1 at a large, highly polarizing value λ_W which we call the warring state. The dynamics of the leading eigenvalue undergoes saddle-node bifurcations during the peace to war transition and the war to peace transition, Fig. 5.2(b,c).

Bifurcation transition from peace to war

Bifurcations in the social network dynamics, Eq. 5.3, induce the system to transition between the low-conflict state where $\lambda_1 = \lambda_S$, and the high-conflict state where $\lambda_1 = \lambda_W$.

The peace-to-war transition occurs when the stable and unstable eigenvalues, λ_S and λ_U, merge in a saddle-node bifurcation and the approximate system no longer has a stable low-conflict state for the leading eigenvalue as shown in Figure 5.2(b). When this bifurcation occurs, the leading eigenvalue converges to the stable high-conflict state λ_W. We can solve for the critical value α^* at which this destabilization occurs by setting $\lambda_S = \hat{\lambda}_U$ which results in

$$\alpha^* = \frac{\beta}{4\tilde{\lambda}_{01}} \tag{5.20}$$

The leading eigenvalue of the system at this stability bifurcation is

$$\lambda_S^* = \frac{\beta}{2\alpha^*}. \tag{5.21}$$

Using this destabilization threshold, we determine that the network will reach equilibrium at the low-conflict state λ_S rather than the high-conflict stable state λ_W if $\lambda_1(\mathbf{X}(t)) < \lambda_U$ and

$$\alpha < \frac{\beta}{4\tilde{\lambda}_{01}} \tag{5.22}$$

$$\tilde{\lambda}_{01} < \frac{\beta}{4\alpha} \tag{5.23}$$

$$\lambda_{\mathbf{1}}\left(\mathbf{X}_0 + \frac{\mathbf{X}_P(t)}{\beta}\right) < \frac{\beta}{4\alpha} \tag{5.24}$$

This indicates that for a given α and β, the critical value for the leading eigenvalue of $\tilde{\mathbf{X}}_0$ is

$$\tilde{\lambda}_{01}^* = \frac{\beta}{4\alpha} \tag{5.25}$$

Beyond this stability bifurcation $\lambda_1 \to \lambda_W$ while the remaining eigenvalues remain small, meaning the edge matrix becomes approximately rank 1 with the dominant eigenvector

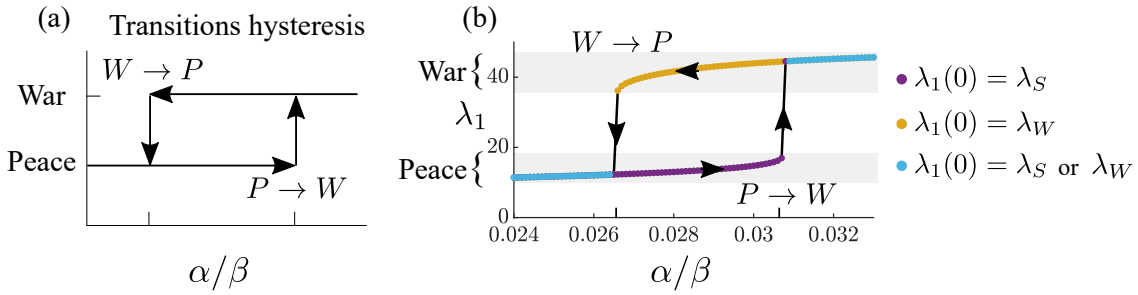

Figure 5.3: Eq. 5.3 displays hysteresis in its parameters. (a) The α/β ratio together with the initial condition determines the final state. The peace to war transition $P \to W$ occurs at a larger threshold than the war to peace transition $W \to P$. (b) Final states of the leading eigenvalue for varying α/β ratios for initial conditions initialized at the peaceful state versus the warring state.

approximating all ties. The size of the dominant eigenvalue is a function of the limit placed on the balance term and goes to infinity without a limiting value on the structural balance term, $\lim_{L\to\infty} \lambda_W = \infty$.

Bifurcation transition from war to peace

Due to the limits on the structural balance term (Eq. 5.3), the leading eigenvalue does not go to infinity as the approximate system Eq. 5.4 would suggest but instead stabilizes at λ_W. In order to transition from a war to a peace state, the system undergoes a different saddle-node bifurcation where $\lambda_U \to \lambda_W$. The warring system will transition back to a peaceful state if

$$\lambda_W < \hat{\lambda}_U \tag{5.26}$$

$$\lambda_W < \frac{\beta + \sqrt{\beta^2 - 4\alpha\beta\tilde{\lambda}_{01}}}{2\alpha} \tag{5.27}$$

$$\lambda_W <\approx \frac{\beta}{\alpha} \tag{5.28}$$

Therefore the critical transition value to move back to a peaceful state from a warring

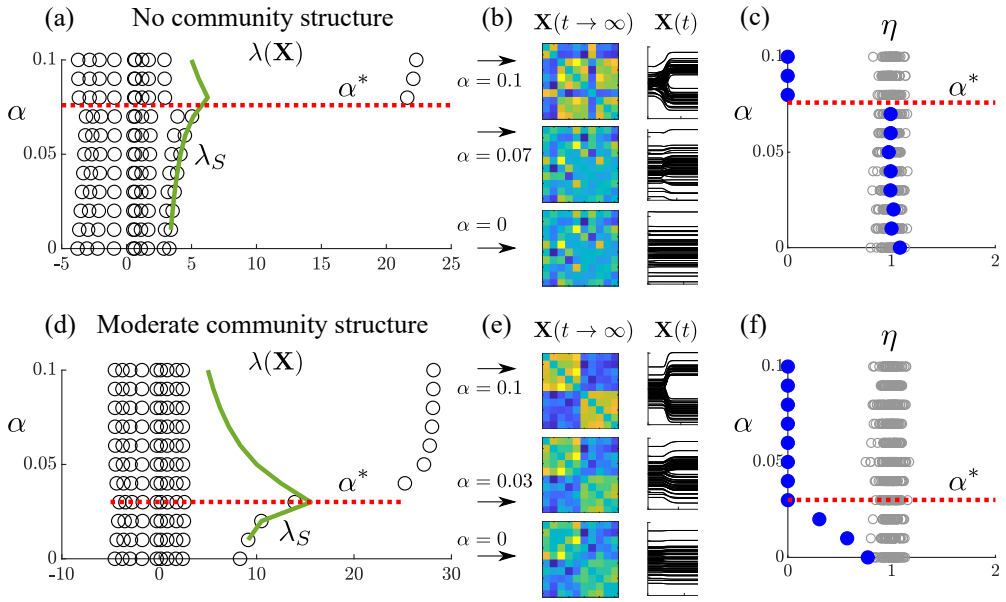

Figure 5.4: Eigenvalues, λ_i, and balance levels, η, of the network at equilibrium for increased balance α. (a) Stable state eigenvalues as a function of α for a random $\mathbf{X}_0$ containing no community structure. The leading eigenvalue agrees with λ_S (Eq. 5.17) before the critical bifurcation α^* (Eq. 5.20) and goes to λ_W after the bifurcation. (b) Networks at equilibrium and edge dynamics for $\alpha = 0.1$, $\alpha = 0.07$, and $\alpha = 0$. (c) Balance levels η as a function of α. (d) Stable state eigenvalues for a network with $\mathbf{X}_0$ containing community structure. (e) Networks at equilibrium and edge dynamics for $\alpha = 0.1$, $\alpha = 0.03$, and $\alpha = 0$. (f) Balance levels for a network with a random $\mathbf{X}_0$ containing community structure.

state is approximately

$$\lambda_W^* \approx \frac{\beta}{\alpha} \tag{5.29}$$

Figure 5.2(c) shows the destabilization of the warring state λ_W when $\lambda_U \to \lambda_W$. Since $\hat{\lambda}_U \leq \lambda_U$, Eq. 5.26 guarentees the system will transition back to a peaceful state.

The war to peace transition happens at about twice the threshold of the peace to war transition $\lambda_W^* = 2\lambda_S^*$. This means that once the system transitions into the war state it takes a significant increase in the stability to structural balance ratio β/α to transition back to a

peaceful state as shown in Figure 5.3. For small α/β ratios, only the peaceful state is stable, for high α/β ratios only the warring state is stable, and for mid-range α/β ratios, both the low and high conflict states are stable, Fig. 5.3(b). In the case when both the low and high conflict states are stable, the equilibrium obtained is dependent on the initial condition. In order to maintain a peaceful state it only requires that $\lambda_U > \frac{\beta}{2\alpha}$, whereas to achieve a peaceful state after a transition to high conflict, it requires that $\lambda_U > \frac{\beta}{\alpha}$ which is twice the value required to maintain the peaceful state. This is an example of hysteresis, where the backward state transition occurs at a different value than the forward state transition. The transitions from peace to war and war to peace are not interchangable and it may take significantly less energy to keep the peaceful state stable than to destabilize the warring fixed point, resuming peace.

5.3 Model behavior and comparison to theories of war

We now observe how the model behaves under parameter manipulation and demonstrate the agreement between our model's computationally and analytically derived results. Figure 5.4 shows network eigenvalues at equilibrium as a function of the balance parameter α, where $\mathbf{X}_0$ either contains community structure (Fig. 5.4(a)) or is completely random (Fig. 5.4(d)). Before the critical bifurcation α^* (Eq. 5.20), the network stabilizes at the low-conflict state and λ_1 matches the analytically computed stable state for the leading eigenvalue λ_S (Eq. 5.17). After α^*, the system transitions to the factious, high-conflict state; this is reflected in the large leading eigenvalue which stabilizes at λ_W and the formula for the leading eigenvalue in the low-conflict state (Eq. 5.17) no longer holds. The network with community structure destabilizes and assumes the warring state at a much smaller α than the random network. This indicates that networks with increased community structure are more easily destabilized.

Balance levels in signed networks can be measured by global balance metric η from Kirkley *et al.* [107] where $\eta = 0$ corresponds to perfect balance and $\eta = 1$ corresponds to an average level of balance compared to null models. The random network does not show a significant level of balance until after the critical bifurcation (Fig. 5.4(c)) at which point the network

transitions to the high-conflict state and becomes perfectly balanced. The network with initial community structure, in contrast, shows significant levels of balance that increase with α even before the critical bifurcation (Fig. 5.4(f)). Figure 5.4 shows that while community structure may lower the threshold for conflict escalation, networks with significant levels of balance can still remain stable at a low-conflict equilibrium if structural balance forces are kept below the bifurcation threshold.

5.3.1 Destabilizing the peaceful state with perturbations

While destabilization of the low-conflict state can be achieved through variation of the balance to stability ratio α/β, destabilization can also occur via a perturbation if $\mathbf{X}_P(t)$ pushes the leading eigenvalue of the system beyond the stability threshold λ_U. We can analyze the types of perturbations that are most likely to destabilize a system for a given $\mathbf{X}_0$. We first consider special cases for the perturbation matrix where the resulting $\tilde{\lambda}_{01}$ can be analytically computed. Beyond these special cases, however, $\tilde{\lambda}_{01}$ is best computed computationally.

Destabilize system with minimum energy perturbation

The perturbation to the low-conflict state that requires the least amount of energy to destabilize is along the direction of the dominant eigenvector. Therefore let the perturbation be $\sigma \mathbf{s}_1 \mathbf{s}_1^T$ where $\mathbf{s}_1$ is the dominant eigenvector of $\mathbf{X}_0$. We solve for the mimimum value of σ that will destabilize the system using the bifurcation threshold Eq. 5.25. The resulting leading eigenvalue is stable at the low-conflict state if

$$\lambda_{\mathbf{1}} \left(\lambda_{01} \mathbf{s}_1 \mathbf{s}_1^T + \frac{\sigma}{\beta} \mathbf{s}_1 \mathbf{s}_1^T \right) < \frac{\beta}{4\alpha} \tag{5.30}$$

Solving for σ we find that the system is destabilized with perturbation σ over the threshold

$$\sigma > \frac{\beta^2}{4\alpha} - \lambda_{01} \tag{5.31}$$

Destabilize system with a low-dimensional perturbation

Perturbations can also take the form of a rank 1 matrix $\mathbf{X}_P(t) = \sigma \mathbf{u}\mathbf{u}^T$ where the factions in the perturbation are contained in the vector $\mathbf{u}$. This type of perturbation will promote the factions contained in $\mathbf{u}$ in the ensuring systemic conflict. With this perturbation our modified transient stable state is

$$\tilde{\mathbf{X}}_0(t) = \mathbf{X}_0 + \frac{\sigma}{\beta}\mathbf{u}\mathbf{u}^T \tag{5.32}$$

The resulting eigenvalues and eigenvectors of the sum of a symmetric matrix and outer product have been studied by Golub [77] and Bunch *et al.* [28] and their derivations can be used, as in the previous example, to compute the minimum σ such that the system destabilizes, $\lambda_1(\tilde{\mathbf{X}}_0(t)) > \beta/(4\alpha)$. This type of perturbation is relevant when all nodes in a network take sides on a particular issue and the resulting low-rank community structure perturbs the network's original structure. Depending on the strength of the perturbation relative to the original structure, the community structure in the perturbation could define some or all of the ensuing conflict.

Destabilize system with a random perturbation

Outside a few special cases, the eigenvalues and eigenvectors of $\tilde{\mathbf{X}}_0(t)$ do not have clean analytical solutions for a generic perturbation $\mathbf{X}_P(t)$. Nevertheless, we can compute $\tilde{\lambda}_{01}$ that will destabilize the system using computational methods and analyze the resulting $\mathbf{X}_P(t)$.

The high-conflict, low-rank structure that appears after destabilizing the low-conflict state will not necessarily contain two equally sized factions. The leading eigenvector could instead contain all nodes or mostly all nodes of the same sign. The interpreation of destabilization in this case would be instead the formation of a tighly-knit alliance structure where, due to structural balance dynamics, nodes are compelled to form positive ties with other nodes due to mutual allies. We can use the eigenvector polarization measure ϕ_1 to determine the extent to which the resulting leading eigenvector contains community structure or a lack of community

structure close to homogeneity [148]. The eigenvector polarization ϕ_i of eigenvector $\mathbf{v}_i$ of the adjacency matrix $\mathbf{X}$ is $\phi_i = \mathbf{v}_i^T \mathbf{M} \mathbf{v}_i$, where $\mathbf{M}$ is the signed modularity matrix [206].

We can use the following optimization schemes to find optimal perturbations $\mathbf{X}_P$ that (1) destabilize the system for any leading mode (Eq. 5.33), (2) destabilize the system with a widespread 2-faction conflict (Eq. 5.34), or (3) destabilize the system with a widespread alliance bloc (Eq. 5.35).

$$\text{Minimum energy:} \quad \underset{\mathbf{X}_P \in A}{\operatorname{argmin}} \left[||\tilde{\lambda}_{01}^* - \tilde{\lambda}_{01}|| \right] \tag{5.33}$$

$$\text{Factional:} \quad \underset{\mathbf{X}_P^F \in A}{\operatorname{argmin}} \left[||\tilde{\lambda}_{01}^* - \tilde{\lambda}_{01}|| + \delta \frac{1}{\phi_1} \right] \tag{5.34}$$

$$\text{Homogeneous:} \quad \underset{\mathbf{X}_P^H \in A}{\operatorname{argmin}} \left[||\tilde{\lambda}_{01}^* - \tilde{\lambda}_{01}|| + \delta \phi_1 \right] \tag{5.35}$$

where $A = \{\mathbf{X} : ||\mathbf{X}|| = \sigma\}$, $\tilde{\lambda}_{01}$ is the leading eigenvalue of $\tilde{\mathbf{X}}_0$, $\tilde{\lambda}_{01}^*$ is the stability threshold (Eq. 5.25), and ϕ_1 is the eigenvector polarization measure of $\tilde{\mathbf{X}}_0$ with weighting δ.

Destabilizing perturbations for pre-WWI network

We use the minimum energy, factional, and homogeneous optimization formulas, Eqs. 5.33, 5.34, and 5.35, to compute minimal perturbations that destabilize the network of great powers leading up to the first world war under the dynamics of Eq. 5.3. World War I is considered a classic example of a systemic war that emerged from the complex dynamics of five European great powers, the United Kingdom, France, Germany, Austria-Hungary, and Russia. We will compare our computationally derived destabilizing perturbations to ties that have been historically identified as critical to the outbreak of the Great War.

We compute the minimum energy perturbations that trigger destabilization, minimal perturbations that also induce widespread conflict, and minimal perturbations that induce a large alliance bloc. $\mathbf{X}_0$ is set to the average edge values over the years 1908-1913 using the militarized interstate disputes (MIDs), alliances, and rivalries data from the Correlates of War dataset and the Handbook of International Rivalries [161, 75, 185, 202].

We set stability and balance parameters to $\beta = 1$ and $\alpha = 0.03$, and normalize the natural stable state matrix, $||\mathbf{X}_0|| = 1$. This results in the stability threshold $\tilde{\lambda}^*_{01} = 8.33$ and minimum perturbation energy $\sigma = 2.3208$ for a perturbation in the direction of the leading eigenvector (Eq. 5.31). Figure 5.5(a) shows the natural stable states, $\mathbf{X}_0$, as well as the eigenvalues of $\mathbf{X}_0$, the modularity matrix $\mathbf{M}$ [158, 206] of the stable state matrix, and the eigenvector polarization measure ϕ [148]. From these spectral measures we see that there is community structure contained in the dominant eigenvector. We analytically compute the optimum perturbation $\mathbf{X}^{op}_P$ with minimum energy $\sigma = 2.3208$ which results in a perturbation with two factions – Germany and Austria-Hungary versus France and Russia, with the United Kingdom loosely aligned with the Franco-Russian bloc.

We solve for alternative, sparse perturbations using the optimization functions delineated above, setting $\sigma = 2.3208$, the energy level for the optimal solution. We compute the error as the amount to which the computationally generated control signal deviates from the stability threshold $E = \tilde{\lambda}^*_{01} - \tilde{\lambda}_{01}$. Figure 5.5(b) shows three sparse perturbations $\mathbf{X}_P$ computed using Eq. 5.33. The resulting destabilized edge system converges to the two factions present during WWI with the United Kingdom, France, and Russia in one faction and Germany and Autria-Hungary in the other faction. Figure 5.5(c) shows sparse perturbations computed using Eq. 5.34. As the leading eigenvector is already factional, the factional solutions are similar to the overall minimum energy solutions in Fig. 5.5(b) and lead to the two factions determined by the system's initial community structure.

Figure 5.5(d) shows the homogenous perturbations $\mathbf{X}^H_P$ computed using Eq. 5.35. The factional structure is much more dominant than other structures in the network resulting in a larger perturbation required to obtain a homogeneous final state, $||\mathbf{X}^H_P|| = 4\sigma$. Ideally, the homogenous perturbations would result in a homogenous struncture among all nodes in the final equilibrium state. With a low level of perturbation energy, however, the system only manages to positively coalesce four of the five countries in a positive alliance bloc. Perturbations $\mathbf{X}^H_{P1}$ and $\mathbf{X}^H_{P3}$ lead to a homogenous coalition against Germany while $\mathbf{X}^H_{P2}$ leads to a homogenous coalition against Austria-Hungary.

We now take a slightly different approach and evaluate which individual ties are most critical in increasing polarization in the network in the year 1913, preceding the outbreak of WWI. Figure 5.6 shows the ties between the great powers and the amount of change measured in the leading eigenvalue λ_1 and the eigenvector polarization metric ϕ_1 for each positive or negative tie perturbation $\delta = \pm 1$. Figure 5.6(b-e) shows that the tie between France and Germany as well as the tie between Germany and Austria-Hungary, were the most crucial in increasing community structure and polarization in the network. The ties between Great Britain and the Continential countries appear to be the least important, implicating that while Great Britain formed an important part of the network, ties with this state were unlikely to create triggering crises. Other significant ties include FRN-AUH, FRN-RUS, GMY-RUS, and AUH-RUS. This finding is to be expected as the German alliance with Austria-Hungary and their disputes with Russia as well as the alliance between Russia and France were all factors that proved integral to the development of World War I [95, 201].

Highly influential perturbations could be computed for other sociopolitical networks to assess the perturbations most likely to destabilize a system and to develop a multilateral approach to stabilizing a network with minimal cost. Figure 5.5 shows that some of the most triggering perturbations to the pre-WWI network are an increase in hostility between Russia and Austria-Hungary and between Germany and France which have been identified as key rivalries in the emergence of WWI. Positive perturbations can also be destabilizing. The reinforcement of the positive alliance between Germany and Austria-Hungary as well as an increased alliance between Russia and France are identified as triggering destabilization. The ties identified as triggering in our computational model correspond to ties that have historically been identified as key to the emergence of WWI. The conflict in Serbia triggered an increase in conflict between Russia and Austria-Hungary due to disputes in this region [204]. As a part of their defense strategy, Germany had a plan to attack France if their was conflict with Russia, known as the Schlieffen Plan, which they carried out after announcing war on Russia [204]. The alliance between Germany and Austria-Hungary as well as the alliance between France and Russia increased tensions in the region as Germany felt threatened by

Russia and France [95, 201]. The destabilizing ties found in this model correspond to WWI destabilization theories from the historical literature.

5.3.2 Critical slowing down near bifurcations

The destabilization perturbations computed in the last section implied knowledge of the dynamical system governing state dynamics and therefore the location of the bifurcation point. In practice, however, the true dynamics are often unknown making it difficult to determine a system's distance to a critical bifurcation. Fortunately, near bifurcations systems exhibit distinctly different behaviors, such as critical slowing down, that can be indicators for nearness to a bifurcation point [180, 110]. Meisel *et al.* [142] found that for saddle-node bifurcations occurring in neural dynamics the recovery rate scales with the square root of the distance from the bifurcation, $r \propto \sqrt{d}$. This scaling law can be derived from the normal form of the saddle-node bifurcation. Our edge dynamical system, Eq. 5.3, exhibits a saddle-node bifurcation in the eigenvalue dynamics when transitioning from the low-conflict to high-conflict state and therefore we expect it to follow this same scaling law.

We measure the standard deviation of the edges after increasingly large perturbations to Eq. 5.3 in the low-conflict state and compare the distance to the bifurcation, $\gamma = \lambda_1 - \lambda_U$, with the decay rate r of the standard deviation. With perturbations that keep the system far from the bifurcation, the network activity decays quickly back to equilibrium (Fig. 5.7(a)) while with perturbations that move the system close to the bifurcation, the network activity exhibits a delay in its decay back to equilibrium (Fig. 5.7(b)). The system moves to the high-conflict state for perturbations past the critical value (Fig. 5.7(c)). Figure 5.7(d) shows that $\log(r)$ scales linearly with $\log(\gamma)$. A linear fit to the data finds that $\log(r) = 0.51 \log(\gamma) - 1.62$. This means that the recovery rate and distance to the bifurcation are approximated related by $r \approx 0.2\sqrt{\gamma}$ which is the same scaling relationship found by Meisel *et al.* [142] for saddle-node bifurcations.

Using the relationship between recovery rates and distance to a bifurcation, we can estimate the distance to a critical bifurcation from decay rates in empirical edge timeseries

data. We estimate the stability of the European great powers network from their edge dynamics leading up to World War I (Fig. 5.8). We use the alliance, rivalry and militarized interstate disputes (MIDs) data from the Correlates of War dataset and the Handbook of International Rivalries [161, 75, 185, 202]. The edges between countries are characterized by long-timescale alliances and rivalries, which can be thought of as natural stable states $\mathbf{X}_0$ (Fig. 5.8(a)), and short-timescale militarized interstate disputes (MIDs) which emerge and decay rapidly and can therefore be thought of as perturbations to the system $\mathbf{X}_P$ (Fig. 5.8(b)). We measure the change in standard deviation in the edges between the years 1873 and 1917 and compute a decay rate after each MID which is used to estimate the distance to the bifurcation (Fig. 5.8(c)). The decay rates, and therefore the estimated distances, show significant variance. Some MIDs showing a significant delay in resolving and therefore a close bifurcation estimate indicating precarious stablity. Other MIDs resolve quickly and therefore estimate that the network is fairly stable. This indicates that while decay rates may provide an estimated range for system stability in empirical networks, other measures may be needed to produce more precise estimates.

5.3.3 Dynamical system destabilization theories of systemic war

We can make several connections between destabilization in our dynamical systems model for sociopolitical dynamics and conflict escalation risk factors identified in the political science literature. These risk factors include nonlinear interactions, ripe rivalry fields, power transitions, and a delay in returning to equilibrium in the number of conflicts [204, 203, 49, 144, 143]. We suggest a dynamical systems mechanism for destabilization cause by these risk factors.

Thompson [204] suggests that an increase in alliance bipolarization increases the likelihood of destabilization in the international system [204]. Polarization directly corresponds to the strength of the leading eigenvalue containing community structure. In terms of the matrix dynamical system, as community structure increases in the dominant eigenvector, the leading eigenvalue separates further from the noise band, and the system becomes destabilized at a

lower balance force threshold as demonstrated in Figure 5.4.

Thompson [204] also suggests that tightly coupled systems with proximate rivalries will experience nonlinear interactions while systems that are not tightly coupled will have only linear interactions. These nonlinear interactions appear in the structural balance term which contains nonlinear interactions only if edges have mutual nodes in common, making these relationships proximate. Edges that do not have nodes in common do not experience a nonlinear interaction in the structural balance term. This keeps such conflicts independent. Nodes that have few mutual allies or rivals will have edge dynamics determined mostly by the individual dyad, uninfluenced by the relationships of other nodes in the network. On the other hand, dense networks will generate many mutual nodes and increase the relative force of the nonlinear interactions determined by the network structure rather than the interactions of the individual dyad.

Network size also effects stability in a similar way. Thompson [203] finds that multipolar system are less stable that bipolar systems which are less stable than unipolar systems. This trend toward instability in systems that contain many influential nodes is reflected in the dynamical system. We can rewrite the nonlinear balance component in Eq. 5.2 as the average edge interaction value ($\mathbf{x}_i^T\mathbf{x}_j/N$) multiplied by N number of nodes, $\mathbf{x}_i^T\mathbf{x}_j = N(\mathbf{x}_i^T\mathbf{x}_j)/N$. If the network increases in size yet the density remains constant then the structural balance term will have more relative weight in the dynamical system. This indicates that increasing the number of nodes in a network and allowing the new nodes to form alliance and rivalries with the existing nodes, keeping the tie density constant, will decrease the stability threshold, making the outbreak of a widespread conflict more likely. This coincides with Thompson's [203] observation that widespread conflicts occur more often in multipolar systems than they do in bipolar systems or unipolar systems.

The decline in power of a global leader and the ascendance of a new leader is also correlated with systemic conflict [204, 203, 49]. Doran and Parsons [49] theorize that changes in power differentials make states less averse to war as their national security is threatened by changes in the status quo. The years leading up to WWI saw a power transition between Great

Britain and Germany (Fig. 5.9(a)) which was a catalyst for the warming ties between Great Britain and France [201, 95]. We can observe the factional alignment of the five great powers over time in the dominant eigenvector (Fig. 5.9(b)). The factions are unstable and maliable during the 1870's and 80's but stabilize in the 1890's with Germany and Austria-Hungary in one faction and France and Russia in the opposite faction. The United Kingdom originally aligns more strongly with the German faction but gradually drifts over to the French side. By taking a weighted sum of the dominant eigenvector we can observe how the balance of power shifted in the network over time (Fig. 5.9(c)). The German coalition loses relative power leading up to WWI which corresponds to historical accounts of Germany increasingly feeling threatened by the other great powers leading up to the start of the war [201, 95]. The lead-up to WWI shows how changes in relative power may cause countries to adjust their alliances and rivalries which can effect the balance of power and network stability.

Midlarsky [143, 144] identifies a delay in returning to an equilibrium in the number of disputes as indicative of instability in the international system. This delay in dispute resolution translates into a slow decay rate in the dynamical system after an MID perturbation, as seen in Figure 5.7. We use the decay rate to estimate distance to the critical bifurcation (Fig. 5.8) similarly to how Midlarsky [143, 144] uses dipute number dynamics to estimate instability in the international system.

Given these mechanisms for destabilization, the formation of new alliances, increased connectivity, the addition of more countries to a system, or fluctuations in relative power could all increase the instability in a system of states and lead to systemic war. While at the dyadic level, forming new connections may not add much conflict to a system or be detrimental for the parties involved, it could inadvertently make a system less stable by allowing for more nonlinear interactions between edges, increasing community structure, or equalizing a power imbalance. Dynamical system models, such as the one introduced here, may be useful in assessing how changes to ties impact not only the countries involved but the entire system; models for system-wide impact could be beneficial in assessing systemic risk and designing de-escalation strategies.

5.4 Discussion

Of key interest in international relations is understanding what factors put a system of states at risk for systemic war as well as what interventionist steps are most effective given instability in a network. We address these questions with a dynamical systems model of interactions between states. The dynamics that occur in sociopolitical networks due to structural balance and dyadic interactions can be modeled with a matrix dynamical system with a stability term and a bounded structural balance term. The matrix dynamical system, Eq. 5.3, has a stable low-conflict state generated from the dyadic stability, and a stable high-conflict state that emerges due to the bounds placed on the structural balance term. This model builds from a previous model of pure structural balance dynamics, Eq. 5.1, that does not contain any stable equilibrium [137].

By approximating the matrix dynamics with an analytically tractable form, Eq. 5.4, we are able to decompose the connectivity matrix $\mathbf{X}$ into eigenvalues and eigenvectors. This dimension reduction technique allows us to reduce the dynamics of hundreds of network edges to the dynamics of a single variable — the leading eigenvalue. We derive bifurcations in terms of the system parameters; these critical bifurcations perturb the system from a peaceful, low-conflict state to a high-conflict state that we categorize as systemic war. A separate bifurcation brings the system from systemic war back to the low-conflict state.

5.4.1 Modeling systemic failures in other fields

In our model we analyze wars that arise from systemic failures in a network not attributable to a single bad actor. Systemic failures are a well-acknowledged feature in other fields such as ecology, finance, engineering, and medicine; the non-localized causes of systemic failures can make them difficult to predict and control [73, 82]. Earth's biosphere has gone through several significant state shifts over the course of time and scientists have found indicators that the climate may be near another critical transition similar to the rapid transitions observed in localized ecological systems [11]. The financial crisis of 2007–2008 emerged from a complex

interaction of many factors and has made regulators more conscious of risk at the systemic level in the financial ecosystem [80]. In 2003, a cascade of failures in the Italian electrical and internet communications network led to an enormous blackout due to vulnerabilities arising from interconnectedness in the system [27]. Systemic failure is a pervasive phenomena in complex, highly connected systems and methods used to measure and anlyze network risks from other fields could prove useful for sociopolitical networks.

5.4.2 Model implications for network stability

Our dynamical systems model for social network dynamics implicates that increasing the size or density of a network decreases network stability. This is similar to findings by May [138] that increasing connectivity in large complex systems decreases the system's stability; in a similar vein, complex ecological networks with many interacting species are at a higher risk of destabilization than simpler networks with fewer interactions among species. Increasing the sociopolitical network's community structure also has a negative impact on stability. The relationship between network structure and stability is also an area of interest in the ecological sciences where scientists have found consistent structures in empirical food webs and assess the risk for perturbations, such as overfishing, to destabilize an ecosystem [12, 196, 54].

Perturbations to ties in the direction of the leading eigenvector of the sociopolitical system have the greatest destabilizing impact and can occur from positive as well as negative ties, meaning that not only new conflicts but also the formation of new alliances can potentially be destabilizing. We estimate the significance of ties between great powers in the WWI network and find that the most influential ties corresponded to key alliances and rivalries that precipitated WWI such as the alliance between Germany and Autria-Hungary as well as the rivalry between Germany and France [201, 95]. When the underlying dynamics of a system are unknown, a system's distance to a critical bifurcation can be estimated by decay rates in the activity after perturbations (Fig. 5.7); we use this method to estimate stability in the great power network timeseries by viewing MIDs as perturbations to the ties (Fig. 5.8). Our relationship between decay rate and distance to the bifurcation, $r = 0.2\sqrt{\gamma}$, has the

same scaling relationship as saddle-node bifurcations in other dynamical systems such as the normal form of the saddle-node bifurcation and transitions to spiking in neurons [142, 180].

5.4.3 Limitations and directions for future research

This model has several limitations. Node power is not taken into account but could be incorporated into the model by weighting ties. There are many other forces that influence tie dynamics beyond the ones included in this model and extensions to represent other key forces may improve the model's validity in particular contexts. One potentially significant force that is not currently represented is power balancing dynamics. Power balancing theorizes that third parties are more likely to ally with the weaker side in a conflict in order to balance the power in a system [81, 122]. Interestingly however, in many cases the opposite occurs, third parties are often more inclines to side with the stronger side in a conflict as this may be in their best interest [81, 213, 40]. A model that includes power balancing would need a parameter that can be tuned to prefer either the weaker or stronger side in a conflict depending on the context.

This model is derived from an anarchic, realist view of the international system. The liberal perspective notes that there are limitations on realist predictive capabilities in the modern era as the realist perspective undervalues the influence of international and non-government organizations and cannot adequately capture the nature of relationships among democracies [53, 183]. This model may better fit a modern setting by taking into consideration these alternative influences. While international organizations may influence the cooperation among states and therefore network structure, the primary sources of power still lies within states, and since many important players in the international system are non-democracies, a realist perspective is still relevant.

5.5 Conclusion

We model the dynamics that occur in sociopolitical networks due to dyadic relationships, structural balance forces, and transient conflicts with a matrix dynamical system that can be

analyzed and understood in terms of the network eigenvalues and eigenvectors. The system has two stable states — low-conflict and high-conflict — and can be perturbed to transition between these two states through parameter manipulation and perturbations. Using spectral analysis we can understand edge dynamics in terms of the dynamics of the eigenvalues and derive analytical solutions for the critical bifurcations that occur in the system. The edge dynamics display hysteresis, where it is easier to transition from the low-conflict to the high-conflict state than it is to transition from the high-conflict back to the low-conflict state.

Using the matrix dynamical system model for edge dynamics, we determine structural features that make systemic war in a network more likely and identify key edges that have the most impact on system stability. By measuring recovery rates after perturbations we can estimate the system's distance from a critical bifurcation and therefore level of stability without knowing the underlying dynamical system parameters. We use our model to analyze the dynamics of the great powers leading up to World War I and find that our model identifies ties that have been implicated as critical to destabilizing Europe in the years leading up to WWI. Risk factors for systemic war found in the international relations literature, such as nonlinear interactions and power transitions, can be interpreted as parameter manipulations, perturbations, and structural changes that destabilize the dynamical system governing the interactions between countries. We hope that models such as this can be used to better evaluate stability in sociopolitical networks and understand the impact that changes to ties in a network have on the dynamics of the system as a whole.

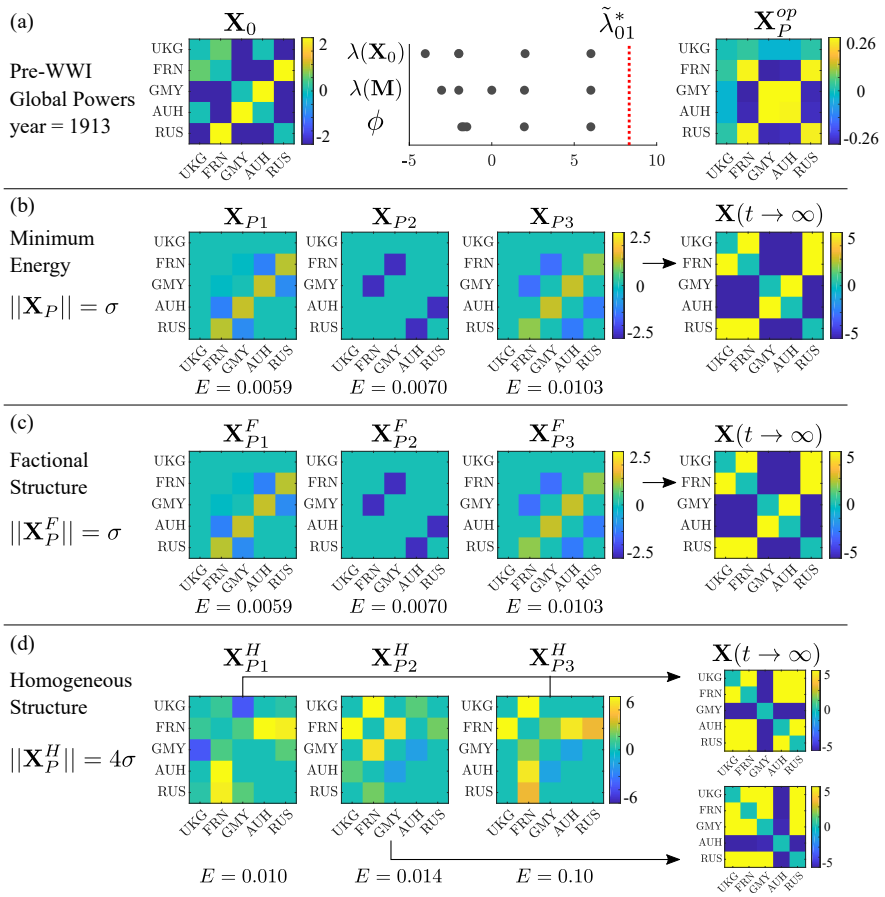

Figure 5.5: Equilibrium pre-WWI network under edge dynamics with perturbations $\mathbf{X}_P$. (a) $\mathbf{X}_0$ is set to the edge weights of the great powers in 1913, immediately preceding WWI. For $\alpha = 0.03$ and $\beta = 1$, the bifurcation threshold is $\tilde{\lambda}^*_{01} = 8.33$. The leading eigenvalue of $\mathbf{X}_0$ is $\lambda_{01} = 6.0125$ resulting in $\sigma = 2.3208$. (b) Perturbations requiring minimal energy (Eq. 5.33), (c) minimal energy while imposing factional structure (Eq. 5.34), and (d) minimal energy while imposing homogenous structure (Eq. 5.35). Error minimized, $E = \tilde{\lambda}^*_{01} - \tilde{\lambda}_{01}$.

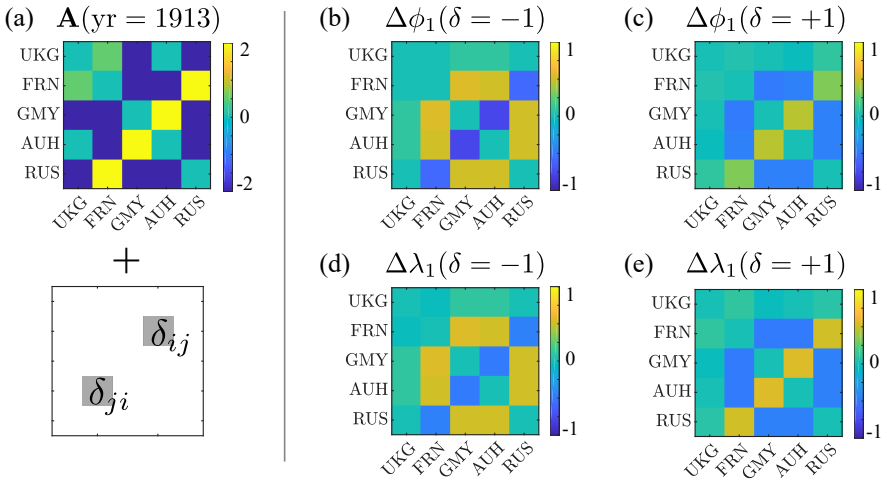

Figure 5.6: (a) Great powers network in 1913. (b) Effects on ϕ_1 for a $\delta = -1$ edge perturbation. (c) Effect on ϕ_1 for a $\delta = +1$ perturbation. (d) Effect on λ_1 for a $\delta = -1$ perturbation. (e) Effect on λ_1 for a $\delta = +1$ perturbation.

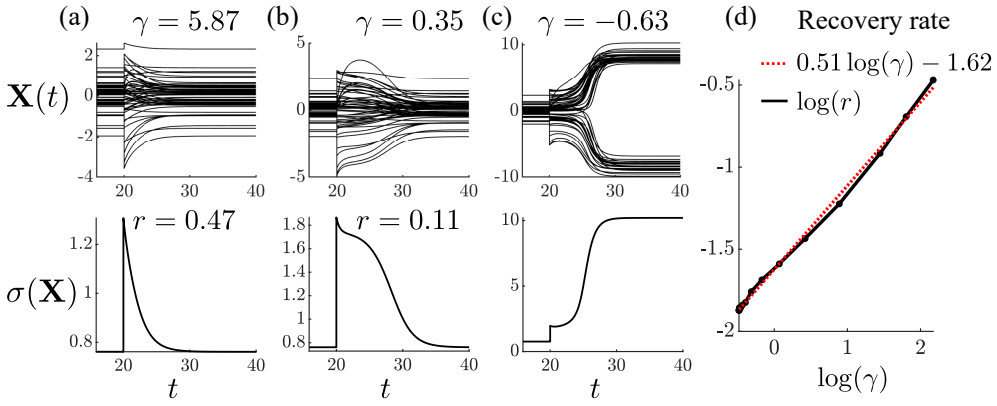

Figure 5.7: Critical slowing down near bifurcation. Recovery rate r is a function of the distance to the bifurcation in the eigenvalue space $\gamma = \lambda_1 - \lambda_U$. (a) Perturbation to the system far from the unstable fixed point, $\gamma = 5.87$, results in a fast recovery rate, $r = 0.47$. (b) Perturbation close to the unstable fixed point, $\gamma = 0.35$, results in a slow recovery rate, $r = 0.11$. (c) Perturbation past the unstable fixed point $\gamma = -0.63$ destabilizes the low-conflict state. (d) The log of the recover rate (r) scales linearly with the log of the distance to the bifurcation (γ), $r \approx 0.2\sqrt{\gamma}$.

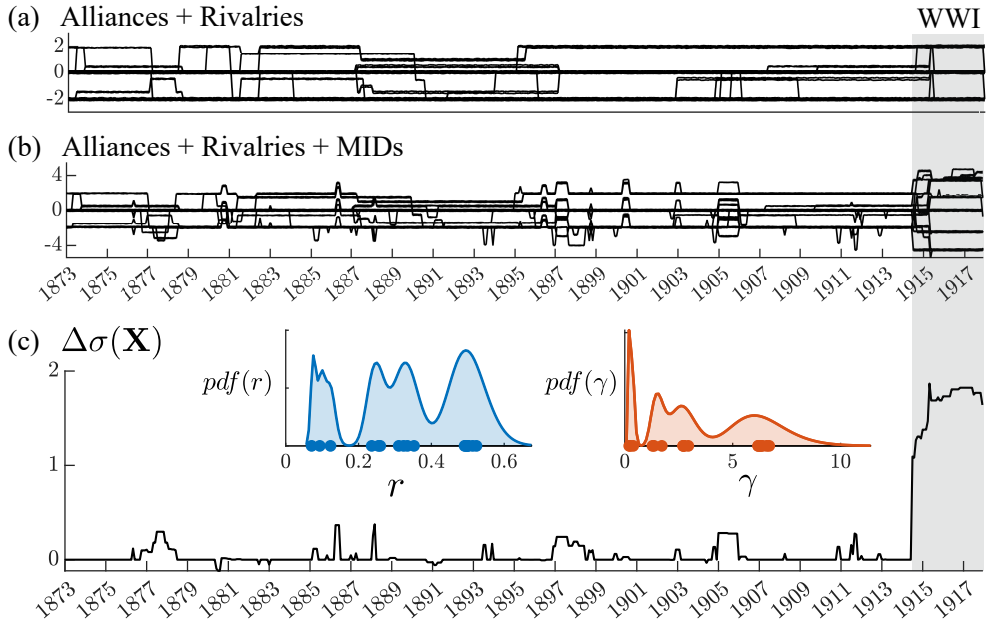

Figure 5.8: (a) Timeseries of great power alliances and rivalries leading up to WWI. (b) Alliances, rivalries, and MIDs. The MIDs act as transient perturbations to the alliance/rivalry network. (c) Change in standard deviation of the edge network over time, $\Delta\sigma(\mathbf{X})$. Decay rates, r, are computed from $\Delta\sigma$, and used to estimate eigenvalue distance to the bifurcation $\gamma = (r/0.2)^2$.

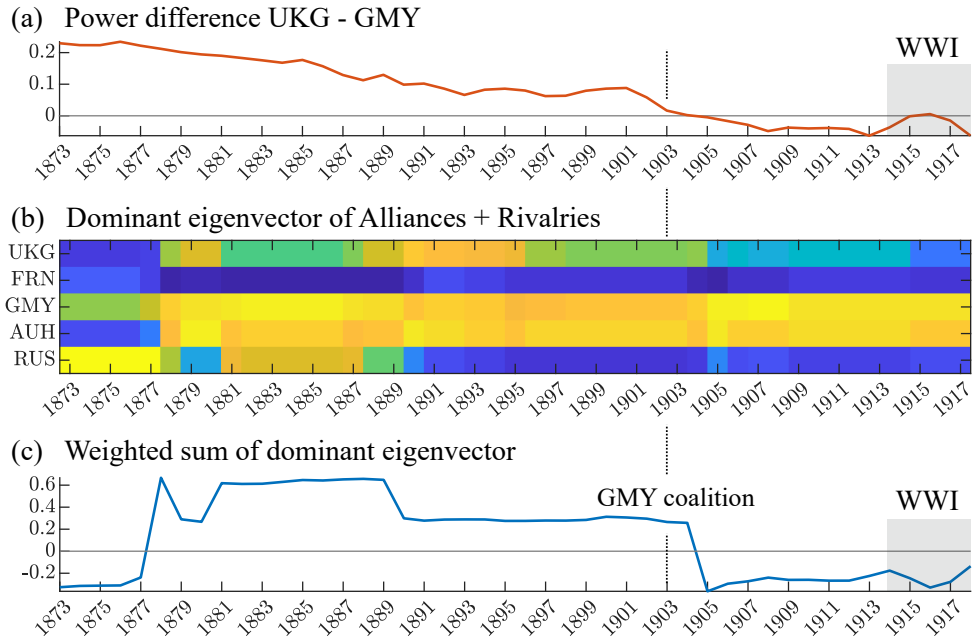

Figure 5.9: (a) Power difference between Great Britain and Germany. (b) Dominant eigenvector of the network of alliances and rivalries. (c) Sum of the dominant eigenvector weighted by relative power.

Chapter 6

CONCLUSION

Network dynamics can be difficult to describe and control particularly when the dynamics are nonlinear, the governing equations are unknown, and control signals play a part in regulating a system's qualitative states. In this dissertation we outline several methods for characterizing nonlinear network dynamics in a low-dimensional space as well as characterizing and deriving control procedures with respect to the system's low-dimensional variables. We model *C. elegans* dynamics using this method and hypothesis that *C. elegans* may make use of nonlinear control to regulate its neural dynamics and ensuing behavioral output. We then proceed to model Hopfield memory network dynamics and random network dynamics in a low-dimensional space and use feed-forward control and bifurcation theory to induce transitions between stable fixed points.

We also model and develop control procedures for network edge dynamics, focusing on structural balancing as well as stabilizing forces that occur in sociopolitical networks. We find that many aspects of a network's structure such as community structure, connectivity density, network size, and nonlinearities affect the dynamics of relationships among nodes. We derive critical bifurcations in these systems that determine when conflict escalation throughout an entire network will occur and what structures in a network are most likely to destabilize a system. We also identify critical thresholds in network statistics that determine whether factions that emerge in a network will be determined from community structure that exists in a network's initial state.

We find that combining techniques from dimension reduction, data-driven discovery of dynamics, and bifurcation theory can result in transparent models that aid in understanding a network's dynamics and determining novel control procedures as well as analysis of control

signals intrinsic to the system. We hope that these techniques can aid in the construction of transparent models and control procedures for nonlinear dynamics in other networks.

Appendix A

A.1 Spectrum of random matrix X

The signal eigenvalues of $\mathbf{A}$ are only visible if they are distinguishable from the noise in the system. In this section we find $\mathcal{O}(||\mathbf{X}||_2)$ as well as γ the spectral edge of the noise matrix $\mathbf{X}$ using random matrix theory.

We start by characterizing the distribution of eigenvalues. The empirical spectral distribution (e.s.d.) of a random matrix $\mathbf{X}$ is defined by

$$\rho(z) = \frac{1}{N}\sum_{i=1}^{n}\delta(z-\omega_i), \tag{A.1}$$

where ω_i are the eigenvalues of $\mathbf{X}$.

Wigner's semicircle distribution Eq. (A.2) defines the eigenvalue density function for a symmetric random matrix size N with i.i.d. entries having variance m^2,

$$\phi(z) = \frac{1}{2\pi N m^2}\sqrt{4Nm^2 - z^2}. \tag{A.2}$$

According to the Wigner limit theorem, the spectra of certain symmetric random matrices converge in distribution to Wigner's semicircle distribution [218, 10], so that

$$\lim_{N\to\infty}\int_{-\infty}^{c}\rho(x)dx = \int_{-\infty}^{c}\phi(x)dx \tag{A.3}$$

Although the semicircle law originally applies only to random symmetric matrices with equal variances for all entries [218, 9, 56], further inquiry has determined that random symmetric block Toeplitz matrices [125, 13] also weakly converge to the semicircle law under certain conditions. $\mathbf{X}$ is a random symmetric block Toeplitz matrix with distribution variances σ_{in}^2 and σ_{out}^2 in the on and off diagonal blocks respectively and therefore has a Wigner semicircle distribution of eigenvalues for some variance parameter m^2 which we have yet to determine.

The edges of the seimicircle enclose the band of eigenvalues in the interval $(-2m\sqrt{N}, 2m\sqrt{N})$. As $N \to \infty$, $\lambda_1(\mathbf{X}) = 2m\sqrt{N}$.

We find the variance parameter by bounding $||\mathbf{X}||_2$ with the Frobenius norm [78],

$$\frac{1}{\sqrt{N}}||\mathbf{X}||_F \leq ||\mathbf{X}||_2 \leq ||\mathbf{X}||_F, \tag{A.4}$$

where

$$||\mathbf{X}||_F = \sqrt{\sum_{j=1}^{\frac{N^2}{2}} |X_{in}|^2 + \sum_{j=1}^{\frac{N^2}{2}} |X_{out}|^2}. \tag{A.5}$$

The ingroup and outgroup sums are distributed as

$$\sum_{j=1}^{\frac{N^2}{2}} |X_{in}|^2 \sim \mathcal{N}\left(\frac{N^2}{2}\sigma_{in}^2, \frac{N^2}{2}Var(|X_{in}|^2)\right) \tag{A.6}$$

$$\sum_{j=1}^{\frac{N^2}{2}} |X_{out}|^2 \sim \mathcal{N}\left(\frac{N^2}{2}\sigma_{out}^2, \frac{N^2}{2}Var(|X_{out}|^2)\right). \tag{A.7}$$

The random variable Z_2 denoting the inside of the square root in (A.5) is therefore distributed as

$$Z_2 = \sum_{j=1}^{\frac{N^2}{2}} |X_{in}|^2 + \sum_{j=1}^{\frac{N^2}{2}} |X_{out}|^2 \sim \mathcal{N}(N^2\sigma^2, N^2\xi^2) \tag{A.8}$$

$$\text{where } \xi^2 = \frac{Var(|X_{in}|^2) + Var(|X_{out}|^2)}{2}. \tag{A.9}$$

The expected value of Z_2 scales with $\sigma^2 N^2$ while the standard deviation only scales with ξN meaning $Z_2 = \mathcal{O}(\sigma^2 N^2)$. Therefore

$$||\mathbf{X}||_F = \mathcal{O}(\sigma N). \tag{A.10}$$

Equation (A.4) then implies that

$$\sigma\sqrt{N} \leq ||\mathbf{X}||_2 \leq \sigma N. \tag{A.11}$$

Therefore $||\mathbf{X}||_2$ must scale with σ as

$$\mathcal{O}(||\mathbf{X}||_2) \sim \sigma. \tag{A.12}$$

This result combined with the Wigner's semicircle distribution scaling implies that

$$\mathcal{O}(||\mathbf{X}||_2) = \sigma\sqrt{N}. \tag{A.13}$$

We continue the argument in order to find the leading eigenvalue of $\mathbf{X}$. We have found that our variance parameter must scale with σ, $m = \mathcal{O}(\sigma)$.

$$m = a\sigma = a\sqrt{\frac{\sigma_{in}^2 + \sigma_{out}^2}{2}} \tag{A.14}$$

for some constant a

We set $\sigma_{in}^2 = \sigma_{out}^2$ which implies that we have a traditional Wigner matrix with i.i.d. entries of variance σ_{in}^2. This means $m = \sigma_{in} = \sigma$ and therefore $a = 1$. Now let $\sigma_{in}^2 \neq \sigma_{out}^2$ while keeping σ^2 constant. m scales only with σ and therefore must be still equal to σ. This means that for all variance values, $m = \sigma$.

This implies that the eigenvalue density function of $\mathbf{X}$ with a diagonal block variance of σ_{in}^2 and an off diagonal block variance of σ_{out}^2 is equivalent to the eigenvalue density function of a random matrix with uniform variance σ^2. This substitution is made in previous derivations of the spectral band of unsigned stochastic block model matrices [152].

Thus, we have determined that the spectra of $\mathbf{X}$ has a Wigner's semicircle distribution with variance parameter σ. Therefore the edge of the spectral band γ of $\mathbf{X}$ is

$$\gamma = 2\sigma\sqrt{N}. \tag{A.15}$$

A.2 Spectra of $\mathbf{A}$ *derived from random matrix theory and complex analysis*

We will now use an alternative method to derive the spectra of $\mathbf{A}$ using random matrix theory and complex analysis. In this argument, we use the eigenvalues of the noise matrix $\mathbf{X}$, whose spectra we have defined in Appendix A.1, to find the eigenvalues of $\mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T$.

We then take these intermediate eigenvalues and use them to find the eigenvalues of $\mathbf{A} = \mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T + \mu N \mathbf{u}_H \mathbf{u}_H^T$.

The expected adjacency matrix $\langle \mathbf{A} \rangle$ is given by Eq. (4.13). We consider the spectrum obtained by adding just the contribution of the contrast eigenvector, $\nu N \mathbf{u}_C \mathbf{u}_C^T$, to the noise matrix which yields the eigenvalue equation

$$(\mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T)\mathbf{v} = z\mathbf{v}. \tag{A.16}$$

We wish to solve for the eigenvalues z, and so use the methods of Ref. [222] to convert Eq. (A.16) into a trace representation Eq. (A.21). We begin by rearranging the terms in Eq. (A.16) to eliminate the eigenvector $\mathbf{v}$,

$$\mathbf{u}_C^T (z - \mathbf{X})^{-1} \mathbf{u}_C = \frac{1}{\nu N}. \tag{A.17}$$

The left hand side of Eq. (A.17) can be written as a sum by performing an eigenvector decomposition on $\mathbf{X}$,

$$\mathbf{u}_C^T \mathbf{S} (z\mathbf{I} - \mathbf{\Lambda})^{-1} \mathbf{S}^T \mathbf{u}_C = \frac{1}{\nu N} \tag{A.18}$$

$$\sum_{i=1}^{N} \frac{(\mathbf{x}_i^T \mathbf{u}_C)^2}{z - \omega_i} = \frac{1}{\nu N}, \tag{A.19}$$

where $\mathbf{S\Lambda S}^T$ is the eigenvector decomposition of $\mathbf{X}$ and $\mathbf{x}_i$ are the eigenvectors of $\mathbf{X}$ (as well as the columns of $\mathbf{S}$). We find that $N(\mathbf{x}_i^T \mathbf{u}_C)^2 \sim \mathcal{X}_1^2$ and therefore $\mathbb{E}[(\mathbf{x}_i^T \mathbf{u}_C)^2] = \frac{1}{N}$ and $Var[(\mathbf{x}_i^T \mathbf{u}_C)^2] = \frac{2}{N^2}$. This allows us to make the approximation $(\mathbf{x}_i^T \mathbf{u}_C)^2 = 1/N$ in Eq. (A.19) giving us

$$\sum_{i=1}^{N} \frac{1/N}{z - \omega_i} = \frac{1}{\nu N} \tag{A.20}$$

$$\frac{1}{N} Tr(z - \mathbf{X})^{-1} = \frac{1}{\nu N}. \tag{A.21}$$

We define $f(z)$ as follows,

$$f(z) = Tr(z - \mathbf{X})^{-1}. \tag{A.22}$$

The values of z that satisfy $f(z) = \frac{1}{\nu}$ are the eigenvalues of the matrix $\mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T$. $f(z)$ has simple poles where $z = \omega_i$, $f(z) \to -\infty$ as $z \nearrow \omega_i$ and $f(z) \to \infty$ as $z \searrow \omega_i$. $f(z)$ is a continuous function within the interval $z \in [\omega_i, \omega_{i-1}]$, therefore for each interval $f(z) = \frac{1}{\nu}$ for some value $z \in [\omega_i, \omega_{i-1}]$. This means the eigenvalues z_i and ω_i are interlaced with the leading eigenvalue $z_1 > \omega_1$. The largest solution to $f(z) = \frac{1}{\nu}$ Eq. (A.22) is the leading eigenvalue of $\mathbf{X} + \mu N \mathbf{u}_C \mathbf{u}_C^T$.

We now can repeat this process to find a formula for both leading eigenvalues by adding the homogeneous signal in addition to the contrast signal back into the noise matrix and solving for the resulting eigenvalues λ [222].

The new eigenvalue equation becomes

$$(\mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T + \mu N \mathbf{u}_H \mathbf{u}_H^T)\mathbf{v} = \lambda \mathbf{v}. \tag{A.23}$$

We solve this equation for the eigenvalues λ_i using the same method used to solve Eq. (A.16) and which is detailed in Ref. [222]. The resulting equation $g(\lambda)$ has a similar form to $f(z)$ but with an additional term,

$$\frac{1/N}{\lambda - z_1} + \sum_{i=2}^{N} \frac{1/N}{\lambda - z_i} = \frac{1}{\mu N} \tag{A.24}$$

$$g(\lambda) = \frac{1}{\lambda - z_1} + \sum_{i=2}^{N} \frac{1}{\lambda - z_i}. \tag{A.25}$$

The values of λ that satisfy $g(\lambda) = \frac{1}{\mu}$ are the eigenvalues of the matrix $\mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T + \mu N \mathbf{u}_H \mathbf{u}_H^T$. Without loss of generality, assume $\nu > \mu$, meaning we have added the largest eigenvalue mode to the noise matrix followed by the second largest eigenvalue mode. When $|\lambda - z_1| >> \frac{1}{N}$, $\frac{1}{\lambda - z_1} = \mathcal{O}(1)$ and $\sum_{i=2}^{N} \frac{1}{\lambda - z_i} = \mathcal{O}(N)$. Because $\frac{1}{\lambda - z_1}$ is the dominant term only when $|\lambda - z_1| = \mathcal{O}(1/N^2)$ and the spectral values z_i that constitute the spectral band of $\mathbf{X} + \nu N \mathbf{u}_C \mathbf{u}_C^T$ are interlaced with the spectrum of $\mathbf{X}$, we may approximate $g(\lambda)$ with our previous function $f(\lambda) = Tr(\lambda - \mathbf{X})^{-1}$ for all λ values away from z_1. $g(\lambda)$ has a singularity at z_1 meaning there is an additional solution to $g(\lambda) = \frac{1}{\mu}$ when $\lambda \approx z_1$.

Therefore the signal eigenvalues, λ_H and λ_C are the largest magnitude solutions to $f(\lambda) = \frac{1}{\nu}$ and $f(\lambda) = \frac{1}{\mu}$. We can find an analytical form for $f(\lambda)$ by taking advantage of the Stieltjes transform representation of $Tr(\mathbf{X} - \lambda)^{-1}$.

The Stieltjes transform $S_\rho(\lambda)$ of density $\rho(t)$ is a function of the complex variable λ and is defined outside the real interval I,

$$S_\rho(\lambda) = \int_I \frac{\rho(t)}{\lambda - t} dt, \quad \lambda \in \mathbb{C} \backslash I. \tag{A.26}$$

The normalized trace of $(\mathbf{X} - \lambda)^{-1}$ is equivalent to the Stieltjes transform of the spectral density of $\mathbf{X}$ [10],

$$\frac{1}{N} Tr(\mathbf{X} - \lambda)^{-1} = S_\rho(\lambda) = \int_I \frac{\rho(x)}{x - \lambda} dx, \tag{A.27}$$

where the e.s.d. has been previously defined, Eq. (A.1). We may substitute the eigenvalue density function $\phi(x)$ for the e.s.d. integrand in the Stieltjes transform since these functions converge in distribution [10],

$$S_\rho(\lambda) = \int_I \frac{\rho(x)}{x - \lambda} dx = \int_I \frac{\phi(x)}{x - \lambda} dx. \tag{A.28}$$

Substituting Eq. (A.2) for $\phi(x)$ yields

$$S_\phi(\lambda) = \int_I \frac{\phi(x)}{x - \lambda} dx = \frac{1}{2\pi N \sigma^2} \int_{-2\sqrt{N}\sigma}^{2\sqrt{N}\sigma} \frac{\sqrt{4N\sigma^2 - x^2}}{x - \lambda} dx. \tag{A.29}$$

We solve this integral using multiple changes of variables. Our argument is adapted from previous work [10]. Letting $x = 2\sqrt{N}\sigma \cos(y)$, the above becomes

$$S_\phi(\lambda) = \frac{1}{\pi} \int_0^{2\pi} \frac{\sin^2(y)}{2\sqrt{N}\sigma \cos(y) - \lambda} dy \tag{A.30}$$

$$= \frac{1}{\pi} \int_0^{2\pi} \frac{(\frac{e^{iy} - e^{-iy}}{2i})^2}{2\sqrt{N}\sigma(\frac{e^{iy} + e^{-iy}}{2}) - \lambda} dy. \tag{A.31}$$

The second change of variables is $\zeta = e^{iy}$ which gives

$$S_\phi(\lambda) = \frac{i}{4\pi\sigma\sqrt{N}} \oint_{|\zeta|=1} \frac{(\zeta^2 - 1)^2}{\zeta^2(\zeta^2 - \frac{\lambda}{\sigma\sqrt{N}}\zeta + 1)} d\zeta. \tag{A.32}$$

$$\text{Let } h(\zeta) = \frac{(\zeta^2 - 1)^2}{\zeta^2(\zeta^2 - \frac{\lambda}{\sigma\sqrt{N}}\zeta + 1)}. \tag{A.33}$$

The function $h(\zeta)$ has three poles,

$$\zeta_0 = 0 \tag{A.34}$$

$$\zeta_1 = \frac{\lambda + \sqrt{\lambda^2 - 4N\sigma^2}}{2\sigma\sqrt{N}} \tag{A.35}$$

$$\zeta_2 = \frac{\lambda - \sqrt{\lambda^2 - 4N\sigma^2}}{2\sigma\sqrt{N}}. \tag{A.36}$$

We must determine which poles are inside the radius $|\zeta| = 1$ and then use the residue theorem to compute the integral. ζ_0 is a pole of order 2 and is inside the contour. Note that $\zeta_1\zeta_2 = 1$, and therefore if $|\zeta_2| \neq |\zeta_1|$ then only one of these poles can be inside the contour. We find, with the argument to follow, that ζ_2 is the pole inside the contour for λ values for which $Im(\lambda) > 0$ and ζ_1 is the pole inside the contour for λ values for which $Im(\lambda) < 0$.

We find that ζ_2 is the enclosed pole for $Im(\lambda) > 0$ by first supposing that $Re(\lambda) > 0$. It follows that $Re(\sqrt{\lambda^2 - 4N\sigma^2}) > 0$ and $Im(\sqrt{\lambda^2 - 4N\sigma^2}) > 0$ and therefore $|\lambda - \sqrt{\lambda^2 - 4N\sigma^2}| < |\lambda + \sqrt{\lambda^2 - 4N\sigma^2}|$ which reveals that $|\zeta_2| < |\zeta_1|$. Let us now suppose that $Re(\lambda) < 0$. It follows that $Re(\sqrt{\lambda^2 - 4N\sigma^2}) < 0$ and $Im(\sqrt{\lambda^2 - 4N\sigma^2}) > 0$, and therefore, by the same argument as above, once again $|\zeta_2| < |\zeta_1|$, meaning that ζ_2 is the enclosed pole.

We use a repetitive argument to show that ζ_1 is the enclosed pole for $Im(\lambda) < 0$. Supposing that $Re(\lambda) < 0$, we find that $Re(\sqrt{\lambda^2 - 4N\sigma^2}) > 0$ and $Im(\sqrt{\lambda^2 - 4N\sigma^2}) > 0$ which informs us that $|\zeta_1| < |\zeta_2|$. Now supposing that $Re(\lambda) > 0$, we find that $Re(\sqrt{\lambda^2 - 4N\sigma^2}) < 0$ and $Im(\sqrt{\lambda^2 - 4N\sigma^2}) > 0$ and therefore $|\zeta_1| < |\zeta_2|$, meaning that ζ_1 is the enclosed pole.

We now know that ζ_1 is inside the contour and ζ_2 is outside for $Im(\lambda) < 0$ while ζ_2 is inside the contour and ζ_2 is outside for $Im(\lambda) > 0$. We use the Residue theorem and the three poles Eqs. (A.34), (A.35), and (A.36) to finish solving Eq. (A.32) by integrating,

$$\oint_{|\zeta|=1} h(\zeta)d\zeta = 2\pi i[Res(h, \zeta_0) + Res(h, \zeta_*)] \tag{A.37}$$

$$\text{where } \zeta_* = \begin{cases} \zeta_1 & , \text{ for } Im(\lambda) < 0 \\ \zeta_2 & , \text{ for } Im(\lambda) > 0. \end{cases} \tag{A.38}$$

Evaluating the residues yields

$$\oint_{|\zeta|=1} h(\zeta)d\zeta = 2\pi i\left[\frac{\lambda}{\sigma\sqrt{N}} \pm \frac{\sqrt{\lambda^2 - 4N\sigma^2}}{\sigma\sqrt{N}}\right]. \tag{A.39}$$

The $\pm$ in the above expression results from the two different residues for $Im(\lambda) > 0$ and $Im(\lambda) < 0$. We now multiply our result by the constant term to finish solving Eq. (A.32),

$$S_\phi(\lambda) = \frac{-\lambda \pm \sqrt{\lambda^2 - 4N\sigma^2}}{2N\sigma^2}. \tag{A.40}$$

Now we have a closed form solution for the Stieltjes transform of the spectral density of $\mathbf{X}$ which gives us an analytical formula for $f(\lambda)$ in the region when λ is outside of the spectral band, $|\lambda| > 2\sigma\sqrt{N} = \gamma$,

$$f(\lambda) = Tr(\lambda - \mathbf{X})^{-1} = \frac{\lambda \pm \sqrt{\lambda^2 - 4N\sigma^2}}{2\sigma^2}, \quad |\lambda| > \gamma. \tag{A.41}$$

We now solve for the eigenvalue λ of $\mathbf{A}$ that corresponds to the contrast signal,

$$f(\lambda) = \frac{1}{\nu} \tag{A.42}$$

$$\frac{\lambda \pm \sqrt{\lambda^2 - 4N\sigma^2}}{2\sigma^2} = \frac{1}{\nu} \tag{A.43}$$

$$\implies \lambda = \nu N + \frac{\sigma^2}{\nu}, \quad |\nu| \geq \frac{\sigma}{\sqrt{N}}. \tag{A.44}$$

If $|\nu| \geq \frac{\sigma}{\sqrt{N}}$ then $|\lambda| \geq |\gamma|$, and the contrast eigenvalue λ is outside the spectral band, otherwise the leading eigenvalue is included in the spectral band. The analogous formula for the eigenvalue corresponding to the homogeneous signal is found by replacing ν by μ in Eq. (A.44). Figure A.1 shows how the intersections between $f(\lambda)$ and $1/\mu$ and $1/\nu$ generate the eigenvalue locations. We have therefore found the same solutions for the signal eigenvalues of $\mathbf{A}$ using random matrix theory and complex analysis as done via perturbation theory in the main text, Eqs. (4.57) and (4.58).

A.3 Random dynamical systems

We generate random dynamical systems with the following dynamics for each node x_i:

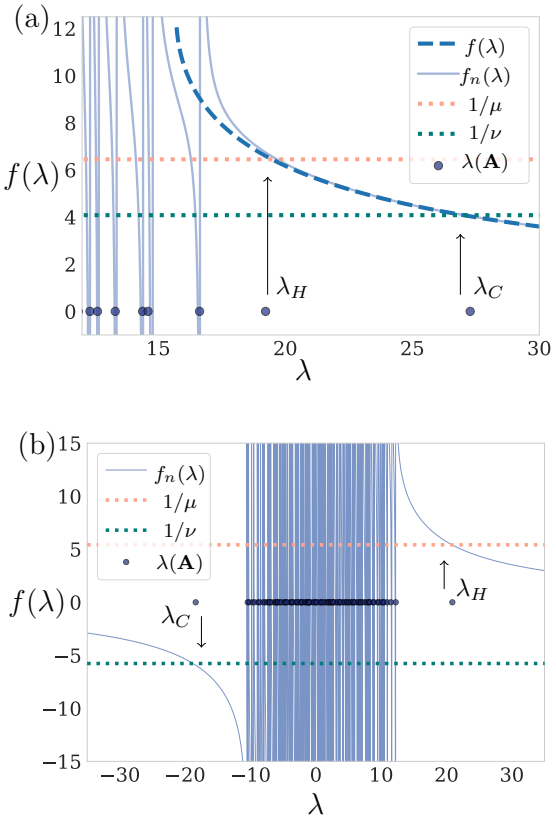

Figure A.1: Intersections of the function $f(\lambda)$ and $1/\mu$ and $1/\nu$ yield eigenvalue locations. (a) $f_n(\lambda)$ is the numerical solution to $f(\lambda)$ and intersects with $1/\mu$ and $1/\nu$ at λ_H and λ_C. $f(\lambda)$ is only defined away from the spectral band; the analytical solution diverges from the numerical approximation upon approaching the spectral edge. (b) The case where λ_H and λ_C are on opposite sides of the spectral band as the signal eigenvalues can be either positive or negative in signed networks.

$$\dot{x_i} = a_i + \sum_{j=1}^{N}(b_{ij}x_j + c_{ij}x_j^2 + d_{ij}x_j^3) + \sum_{j=1}^{N}\sum_{k=1}^{N}(m_{ijk}x_jx_k) - x_i^3 \tag{A.45}$$

where N is the number of nodes and the constants for each term, a, b, c, d, m, are drawn from a sparse normal distribution divided by $\sqrt{N}$ with 10% sparsity. The deterministic term, $-x_i^3$, is added to bias the method to generate systems that do not blow-up in finite time. In order to construct data for the model discovery, we randomly generate a set of initial conditions, find the set of attractors, and then randomly generate additional initial conditions centered about each attractor. Our final set of initial conditions used for the model discovery is selected to have a balanced number of trajectories that converge to each attractor so that all attractors will be adequately represented in the SINDy model. In the 2D random dynamical system we use $N = 10$ while in the 3D random dynamical system we use $N = 20$.

A.4 Hopfield networks

We create a Hopfield model as outlined in previous work [87, 150] using MNIST digits for memories [120]. We use the discrete memory update rule $x(t+1) = \sigma(Wx(t))$ where $\sigma = \tanh(\alpha x)$ and the connectivity matrix, $W = MM^T$, is the memory matrix multiplied by its transpose. We use a tanh function rather than the traditional Heaviside step function in order to make the dynamics smooth and therefore representable by a Taylor series expansion for the SINDy model. Note that $\lim_{\alpha\to\infty}\sigma(x) = 2H(x) - 1$, the traditionally used Heaviside update function. In order to construct data for the model discovery, we generated noisy data points surrounding each memory and then took random samples as initial conditions. For the 2D Hopfield network we used 2 memories, $M \in \mathbb{R}^{400\times 2}$ and $\alpha = 4$. For the 3D Hopfield network we used 3 memories, $M \in \mathbb{R}^{400\times 3}$ and $\alpha = 5$.

BIBLIOGRAPHY

[1] Emmanuel Abbe. Community Detection and Stochastic Block Models: Recent Developments. *Journal of Machine Learning Research*, 18:86, 2018.

[2] Alessandro Alla and J. Nathan Kutz. Nonlinear model order reduction via dynamic mode decomposition. *SIAM Journal on Scientific Computing*, 39(5):B778–B796, January 2017.

[3] R. B. Alley, J. Marotzke, W. D. Nordhaus, J. T. Overpeck, D. M. Peteet, R. A. Pielke, R. T. Pierrehumbert, P. B. Rhines, T. F. Stocker, L. D. Talley, and J. M. Wallace. Abrupt Climate Change. *Science*, 299(5615):2005–2010, March 2003. Publisher: American Association for the Advancement of Science Section: Review.

[4] F. Allgöwer, T. A. Badgwell, J. S. Qin, J. B. Rawlings, and S. J. Wright. Nonlinear predictive control and moving horizon estimation — An introductory overview. In Paul M. Frank, editor, *Advances in Control*, pages 391–449, London, 1999. Springer.

[5] T. Antal, P. L. Krapivsky, and S. Redner. Social balance on networks: The dynamics of friendship and enmity. *Physica D: Nonlinear Phenomena*, 224(1):130–136, December 2006.

[6] Juliette Ben Arous, Sophie Laffont, and Didier Chatenay. Molecular and sensory basis of a food related two-state behavior in C. elegans. *PloS one*, 4(10):e7584, 2009.

[7] Kenneth J. Arrow. *Public investment, the rate of return, and optimal fiscal policy*. Published for Resources for the Future by the Johns Hopkins Press, Baltimore, 1970.

[8] P. Auger, R. Bravo de la Parra, J. C. Poggiale, E. Sánchez, and T. Nguyen-Huu. Aggregation of variables and applications to population dynamics. In Pierre Magal and Shigui Ruan, editors, *Structured Population Models in Biology and Epidemiology*, Lecture Notes in Mathematics, pages 209–263. Springer, Berlin, Heidelberg, 2008.

[9] Z. D. Bai and Y. Q. Yin. Convergence to the Semicircle Law. *The Annals of Probability*, 16(2):863–875, 1988.

[10] Zhidong Bai and Jack W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Springer Series in Statistics. Springer New York, New York, NY, 2010.

[11] Anthony D. Barnosky, Elizabeth A. Hadly, Jordi Bascompte, Eric L. Berlow, James H. Brown, Mikael Fortelius, Wayne M. Getz, John Harte, Alan Hastings, Pablo A. Marquet, Neo D. Martinez, Arne Mooers, Peter Roopnarine, Geerat Vermeij, John W. Williams, Rosemary Gillespie, Justin Kitzes, Charles Marshall, Nicholas Matzke, David P. Mindell, Eloy Revilla, and Adam B. Smith. Approaching a state shift in Earth's biosphere. *Nature*, 486(7401):52–58, June 2012.

[12] Jordi Bascompte. Disentangling the Web of Life. *Science*, 325(5939):416–419, July 2009. Publisher: American Association for the Advancement of Science Section: Perspective.

[13] Riddhipratim Basu, Arup Bose, Shirshendu Ganguly, and Rajat Subhra Hazra. Limiting spectral distribution of block matrices with Toeplitz block structure. *Statistics & Probability Letters*, 82(7):1430–1438, July 2012.

[14] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems*. Neural information processing systems foundation, 2002. ISSN: 10495258.

[15] Alim Louis Benabid. Deep brain stimulation for Parkinson's disease. *Current Opinion in Neurobiology*, 13(6):696–706, December 2003.

[16] Peter Benner, Serkan Gugercin, and Karen Willcox. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review*, 57(4):483–531, 2015. Publisher: SIAM.

[17] Barry Bentley, Robyn Branicky, Christopher L Barnes, Yee Lian Chew, Eviatar Yemini, Edward T Bullmore, Petra E Vértes, and William R Schafer. The multilayer connectome of Caenorhabditis elegans. *PLoS computational biology*, 12(12):e1005283, 2016.

[18] Raja Bhattacharya, Denis Touroutine, Belinda Barbagallo, Jason Climer, Christopher M Lambert, Christopher M Clark, Mark J Alkema, and Michael M Francis. A conserved dopamine-cholecystokinin signaling pathway shapes context–dependent Caenorhabditis elegans behavior. *PLoS genetics*, 10(8):e1004584, 2014.

[19] Jason J. Bramburger and J. Nathan Kutz. Poincar\'e maps for multiscale physics discovery and nonlinear Floquet theory. *arXiv:1908.10958 [nlin]*, August 2019. arXiv: 1908.10958.

[20] K. L. Briggman, H. D. I. Abarbanel, and W. B. Kristan. Optical imaging of neuronal populations during decision-making. *Science*, 307(5711):896–901, February 2005. Publisher: American Association for the Advancement of Science Section: Research Article.

[21] Rogers Brubaker and David D. Laitin. Ethnic and nationalist violence. *Annual Review of Sociology*, 24(1):423–452, 1998.

[22] Jeroen Bruggeman, V. A. Traag, and Justus Uitermark. Detecting communities through network data. *American Sociological Review*, 77(6):1050–1063, 2012.

[23] S.L. Brunton and J.N. Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2019.

[24] Steven L. Brunton and Jose Nathan Kutz. *Data-driven science and engineering: machine learning, dynamical systems, and control*. Cambridge University Press, Cambridge, United Kingdom; New York, NY, 2019.

[25] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, April 2016.

[26] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Sparse identification of nonlinear dynamics with control (SINDYc). *IFAC-PapersOnLine*, 49(18):710–715, January 2016.

[27] Sergey V. Buldyrev, Roni Parshani, Gerald Paul, H. Eugene Stanley, and Shlomo Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464(7291):1025–1028, April 2010.

[28] James R. Bunch, Christopher P. Nielsen, and Danny C. Sorensen. Rank-one modification of the symmetric eigenproblem. *Numerische Mathematik*, 31(1):31–48, March 1978.

[29] Keith Burghardt and Zeev Maoz. Dyadic imbalance in networks. *Journal of Complex Networks*, 8(1):cnaa001, February 2020.

[30] Daniel Byman. *Trends in outside support for insurgent movements*. MR (Series) ; MR-1405. Rand, Santa Monica, CA, 2001.

[31] Eduardo F. Camacho and Carlos Bordons Alba. *Model predictive control*. Advanced Textbooks in Control and Signal Processing. Springer-Verlag, London, 2 edition, 2007.

[32] Dorwin Cartwright and Frank Harary. Structural balance: a generalization of heider's theory. *Psychological Review*, 63(5):277–293, 1956.

[33] Dorwin Cartwright and Frank Harary. Structural balance: a generalization of Heider's theory. *Psychological Review*, 63(5):277–293, September 1956. Publisher: American Psychological Association.

[34] Lars-Erik Cederman and Manuel Vogt. Dynamics and logics of civil war. *Journal of Conflict Resolution*, 61(9):1992–2016, 2017.

[35] Kathleen Champion, Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45):22445–22451, November 2019.

[36] Sanjeev Chauhan, Michelle Girvan, and Edward Ott. Spectral properties of networks with community structure. *Physical Review E*, 80(5):056114, 2009.

[37] Sanjeev Chauhan, Michelle Girvan, and Edward Ott. Spectral properties of networks with community structure. *Physical Review E*, 80(5):056114, November 2009. Publisher: American Physical Society.

[38] B. L. Chen, D. H. Hall, and D. B. Chklovskii. Wiring optimization can relate neuronal structure and function. *Proceedings of the National Academy of Sciences*, 103(12):4723–4728, March 2006.

[39] Xiaowen Chen, Francesco Randi, Andrew M Leifer, and William Bialek. Searching for collective behavior in a small brain. *Physical Review E*, 99(5):052418, 2019.

[40] Thomas J. Christensen and Jack Snyder. Chain Gangs and Passed Bucks: Predicting Alliance Patterns in Multipolarity. *International Organization*, 44(2):137–168, 1990. Publisher: [MIT Press, University of Wisconsin Press, Cambridge University Press, International Organization Foundation].

[41] Matthew A Churgin, Richard J McCloskey, Emily Peters, and Christopher Fang-Yen. Antagonistic serotonergic and octopaminergic neural circuits mediate food-dependent locomotory behavior in Caenorhabditis elegans. *Journal of Neuroscience*, 37(33):7811–7823, 2017.

[42] Holger Conzelmann, Julio Saez-Rodriguez, Thomas Sauter, Boris N. Kholodenko, and Ernst D. Gilles. A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. *BMC Bioinformatics*, 7(1):34, January 2006.

[43] Sean P. Cornelius, William L. Kath, and Adilson E. Motter. Realistic control of network dynamics. *Nature Communications*, 4(1):1942, October 2013.

[44] Antonio C. Costa, Tosif Ahamed, and Greg J. Stephens. Adaptive, locally linear models of complex dynamics. *Proceedings of the National Academy of Sciences*, 116(5):1501–1510, January 2019.

[45] Mark C Cross and Pierre C Hohenberg. Pattern formation outside of equilibrium. *Reviews of modern physics*, 65(3):851, 1993. Publisher: APS.

[46] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Inference and phase transitions in the detection of modules in sparse networks. *Physical Review Letters*, 107(6):065701, 2011.

[47] Charles B Delahunt, Jeffrey A Riffell, and J Nathan Kutz. Biological mechanisms for learning: a computational model of olfactory learning in the Manduca sexta moth, with applications to neural nets. *Frontiers in computational neuroscience*, 12:102, 2018. Publisher: Frontiers.

[48] Emily L Denton, Wojciech Zaremba, Joan Bruna, Yann LeCun, and Rob Fergus. Exploiting linear structure within convolutional networks for efficient evaluation. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 1269–1277. Curran Associates, Inc., 2014.

[49] Charles F. Doran and Wes Parsons. War and the Cycle of Relative Power. *The American Political Science Review*, 74(4):947–965, 1980. Publisher: [American Political Science Association, Cambridge University Press].

[50] Patrick Doreian and Andrej Mrvar. Partitioning signed social networks. *Social Networks*, 31(1):1–11, January 2009.

[51] Patrick Doreian and Andrej Mrvar. Structural balance and signed international relations. *Journal of Social Structure*, 16:1–49, 2015.

[52] Patrick Doreian and Andrej Mrvar. Structural Balance and Signed International Relations. *Journal of Social Structure*, 16, July 2015.

[53] Michael Doyle. Ways of War and Peace: Realism, Liberalism, and Socialism. *Books*, January 1997.

[54] Jennifer A. Dunne, Richard J. Williams, and Neo D. Martinez. Network structure and robustness of marine food webs. *Marine Ecology Progress Series*, 273:291–302, June 2004.

[55] Daniel Dylewsky, Xiu Yang, Alexandre Tartakovsky, and J Nathan Kutz. Engineering structural robustness in power grid networks susceptible to community desynchronization. *Applied Network Science*, 4(1):24, 2019. Publisher: Springer.

[56] Laszlo Erdős. Universality of Wigner random matrices: a survey of recent results. *Russian Mathematical Surveys*, 66(3):507–626, June 2011.

[57] Paul Erdős and Alfr{\'e}d Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5:17–60, 1960.

[58] Pouya Esmailian and Mahdi Jalili. Community detection in signed networks: the role of negative ties in different scales. *Scientific Reports*, 5:14339, 2015.

[59] Giuseppe Facchetti, Giovanni Iacono, and Claudio Altafini. Computing global structural balance in large-scale signed social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 108(52):20953–20958, 2011.

[60] Matthew Farrell, Stefano Recanatesi, Timothy Moore, Guillaume Lajoie, and Eric Shea-Brown. Recurrent neural networks learn robust representations by dynamically balancing compression and expansion. *bioRxiv*, page 564476, December 2019. Publisher: Cold Spring Harbor Laboratory Section: New Results.

[61] C. Fieseler, J. Kunert-Graf, and J. Nathan Kutz. The control structure of the nematode Caenorhabditis elegans: Neuro-sensory integration and proprioceptive feedback. *Journal of Biomechanics*, 74:1–8, June 2018.

[62] Charles Fieseler, Manuel Zimmer, and Nathan Kutz. Unsupervised learning of control signals and their encodings in C. elegans whole-brain recordings. *arXiv:2001.08346 [q-bio]*, January 2020. arXiv: 2001.08346.

[63] Rolf Findeisen, Lars Imsland, Frank Allgower, and Bjarne A. Foss. State and output feedback nonlinear model predictive control: An overview. *European Journal of Control*, 9(2-3):190–206, January 2003.

[64] Richard FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1(6):445–466, July 1961.

[65] Steven W Flavell, Navin Pokala, Evan Z Macosko, Dirk R Albrecht, Johannes Larsch, and Cornelia I Bargmann. Serotonin and the neuropeptide pdf initiate and extend opposing behavioral states in C. elegans. *Cell*, 154(5):1023–1035, 2013.

[66] Santo Fortunato and Darko Hric. Community detection in networks: A user guide. *Physics Reports*, 659:1–44, 2016.

[67] Milton Friedman. The role of monetary policy. *The American Economic Review*, 58(1):1–17, 1968. Publisher: The American Economic Association.

[68] Milton Friedman. *A monetary history of the United States, 1867-1960*. Studies in business cycles ; no. 12. Princeton University Press, Princeton, N.J., 1971.

[69] Thomas Gallagher, Theresa Bjorness, Robert Greene, Young-Jai You, and Leon Avery. The geometry of locomotive behavioral states in C. elegans. *PloS one*, 8(3):e59865, 2013.

[70] Zilin Gao and Yinhe Wang. The structural balance analysis of complex dynamical networks based on nodes' dynamical couplings. *PLOS ONE*, 13(1):e0191941, January 2018.

[71] Zilin Gao, Yinhe Wang, and Lili Zhang. Adaptive control of structural balance for complex dynamical networks based on dynamic coupling of nodes. *International Journal of Modern Physics B*, February 2018. Publisher: World Scientific Publishing Company.

[72] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342, January 2000.

[73] Philip Garnett. Total systemic failure? *Science of The Total Environment*, 626:684–688, June 2018.

[74] Amir Ghasemian, Pan Zhang, Aaron Clauset, Cristopher Moore, and Leto Peel. Detectability thresholds and optimal algorithms for community structure in dynamic networks. *Physical Review X*, 6(3):031005, 2016.

[75] Gibler, Douglas M. *International military alliances, 1648-2008*. CQ Press, 2009.

[76] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[77] Gene H. Golub. Some Modified Matrix Eigenvalue Problems. *SIAM Review*, 15(2):318–334, April 1973. Publisher: Society for Industrial and Applied Mathematics.

[78] Gene H. Golub and Charles F. Van Loan. *Matrix computations*. Johns Hopkins studies in the mathematical sciences. Johns Hopkins University Press, Baltimore, 3rd ed. edition, 1996.

[79] John Guckenheimer and Philip Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, volume 42 of *Applied Mathematical Sciences*. Springer New York, New York, NY, 1983.

[80] Andrew G. Haldane and Robert M. May. Systemic risk in banking ecosystems. *Nature*, 469(7330):351–356, January 2011. Publisher: Nature Publishing Group.

[81] Brian Healy and Arthur Stein. The Balance of Power in International History: Theory and Reality. *The Journal of Conflict Resolution*, 17(1):33–61, 1973. Publisher: Sage Publications, Inc.

[82] Dirk Helbing. Globally networked risks and how to respond. *Nature*, 497(7447):51–59, May 2013.

[83] Herbert W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, January 2000. Publisher: Society for Industrial and Applied Mathematics.

[84] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, July 2006. Publisher: American Association for the Advancement of Science Section: Report.

[85] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500–544, August 1952.

[86] J. Hofbauer and J.W.-H. So. Multiple limit cycles for three dimensional Lotka-Volterra equations. *Applied Mathematics Letters*, 7(6):65–70, November 1994.

[87] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, April 1982.

[88] Ingrid Hums, Julia Riedl, Fanny Mende, Saul Kato, Harris S Kaplan, Richard Latham, Michael Sonntag, Lisa Traunmüller, and Manuel Zimmer. Regulation of two motor patterns enables the gradual adjustment of locomotion strategy in Caenorhabditis elegans. *Elife*, 5:e14116, 2016.

[89] Amiyaal Ilany, Adi Barocas, Lee Koren, Michael Kam, and Eli Geffen. Structural balance in the social networks of a wild mammal. *Animal Behaviour*, 85(6):1397–1405, June 2013.

[90] Alberto Isidori. *Nonlinear control systems design: selected papers from the IFAC Symposium, Capri, Italy, 14-16 June 1989*. IFAC symposia series; 1990, no. 2. Published for the International Federation of Automatic Control by Pergamon Press, Oxford; New York, 1st ed. edition, 1990.

[91] Max Jaderberg, Andrea Vedaldi, and Andrew Zisserman. Speeding up convolutional neural networks with low rank expansions. *arXiv:1405.3866 [cs]*, May 2014. arXiv: 1405.3866.

[92] Robert Jervis. Cooperation Under the Security Dilemma. *World Politics*, 30(2):167–214, 1978. Publisher: [Cambridge University Press, Trustees of Princeton University].

[93] I. T Jolliffe. *Principal component analysis*. Springer series in statistics. Springer, New York, 2nd ed. edition, 2002.

[94] Lauren M Jones, Alfredo Fontanini, Brian F Sadacca, Paul Miller, and Donald B Katz. Natural stimuli evoke dynamic sequences of states in sensory cortical ensembles. *Proceedings of the National Academy of Sciences*, 104(47):18772–18777, 2007. Publisher: National Acad Sciences.

[95] Donald Kagan. *On the origins of war and the preservation of peace*. Doubleday, New York, first edition. edition, 1995.

[96] Donald Kagan. *The Peloponnesian War*. Penguin Books, New York, 2004.

[97] E. Kaiser, J. N. Kutz, and S. L. Brunton. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2219):20180335, November 2018. Publisher: Royal Society.

[98] Eurika Kaiser, J. Nathan Kutz, and Steven L. Brunton. Data-driven discovery of Koopman eigenfunctions for control. *arXiv:1707.01146 [math]*, April 2018. arXiv: 1707.01146.

[99] Stathis N. Kalyvas and Matthew Adam Kocher. Ethnic cleavages and irregular war: Iraq and Vietnam. *Politics & Society*, 35(2):183–223, 2007.

[100] Harris S Kaplan, Oriana Salazar Thula, Niklas Khoss, and Manuel Zimmer. Nested neuronal dynamics orchestrate a behavioral hierarchy across timescales. *Neuron*, 105(3):562–576, 2020.

[101] Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, January 2011.

[102] Saul Kato, Harris S. Kaplan, Tina Schrödel, Susanne Skora, Theodore H. Lindsay, Eviatar Yemini, Shawn Lockery, and Manuel Zimmer. Global brain dynamics embed the motor command sequence of Caenorhabditis elegans. *Cell*, 163(3):656–669, October 2015.

[103] James P. Keener. Infinite period bifurcation and global bifurcation branches. *SIAM Journal on Applied Mathematics*, 41(1):127–144, 1981. Publisher: Society for Industrial and Applied Mathematics.

[104] Terrence P Kenakin. *A pharmacology primer : theory, application and methods*. Elsevier, Amstedam ; Boston, 2004.

[105] Thomas B. Kepler and Timothy C. Elston. Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophysical Journal*, 81(6):3116–3136, December 2001.

[106] William Ogilvy Kermack, A. G. McKendrick, and Gilbert Thomas Walker. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721, August 1927. Publisher: Royal Society.

[107] Alec Kirkley, George T. Cantwell, and M. E. J. Newman. Balance in signed networks. *Physical Review E*, 99(1):012320, January 2019.

[108] Richard Komuniecki, Vera Hapiak, Gareth Harris, and Bruce Bamber. Context-dependent modulation reconfigures interactive sensory-mediated microcircuits in Caenorhabditis elegans. *Current opinion in neurobiology*, 29:17–24, 2014.

[109] Morten L. Kringelbach, Ned Jenkinson, Alexander L. Green, Sarah L. F. Owen, Peter C. Hansen, Piers L. Cornelissen, Ian E. Holliday, John Stein, and Tipu Z. Aziz. Deep brain stimulation for chronic pain investigated with magnetoencephalography. *NeuroReport*, 18(3):223–228, February 2007.

[110] Christian Kuehn. A mathematical framework for critical transitions: Bifurcations, fast–slow systems and stochastic dynamics. *Physica D: Nonlinear Phenomena*, 240(12):1020–1035, June 2011.

[111] Krzysztof Kulakowski, Przemysław Gawroński, and Piotr Gronek. The Heider balance:: A continuous approach. *International Journal of Modern Physics C: Computational Physics & Physical Computation*, 16(5):707–716, 2005.

[112] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. De Luca, and S. Albayrak. Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization. In *Proceedings of the 2010 SIAM International Conference on Data Mining*, Proceedings, pages 559–570. Society for Industrial and Applied Mathematics, April 2010.

[113] James Kunert, Eli Shlizerman, and J Nathan Kutz. Low-dimensional functionality of complex network dynamics: Neurosensory integration in the Caenorhabditis elegans connectome. *Physical Review E*, 89(5):052805, 2014. Publisher: APS.

[114] James M Kunert-Graf, Eli Shlizerman, Andrew Walker, and J Nathan Kutz. Multistability and long-timescale transients encoded by network structure in a model of C. elegans connectome dynamics. *Frontiers in computational neuroscience*, 11:53, 2017.

[115] J Nathan Kutz, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.

[116] J Nathan Kutz, Steven L Brunton, Bingni W Brunton, and Joshua L Proctor. *Dynamic mode decomposition: data-driven modeling of complex systems*, volume 149. SIAM, 2016.

[117] Jose Nathan Kutz. *Data-driven modeling & scientific computation: methods for complex systems & big data*. Oxford University Press, Oxford, 1st ed. edition, 2013.

[118] Yuri A Kuznetsov. *Elements of applied bifurcation theory*, volume 112 of *Applied Mathematical Sciences,*. Springer New York, New York, NY, 2nd ed. edition, 1998.

[119] Gilles Laurent, Mark Stopfer, Rainer W Friedrich, Misha I Rabinovich, Alexander Volkovskii, and Henry DI Abarbanel. Odor Encoding as an Active, Dynamical Process: Experiments, Computation, and Theory. *Annual Review of Neuroscience*, 24(1):263–297, 2001.

[120] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, November 1998. Conference Name: Proceedings of the IEEE.

[121] Timothy M. Lenton, Hermann Held, Elmar Kriegler, Jim W. Hall, Wolfgang Lucht, Stefan Rahmstorf, and Hans Joachim Schellnhuber. Tipping Elements in the Earth's Climate System. *Proceedings of the National Academy of Sciences of the United States of America*, 105(6):1786–1793, 2008. Publisher: National Academy of Sciences.

[122] Jack S. Levy. Theories of General War. *World Politics*, 37(3):344–374, 1985. Publisher: [Cambridge University Press, Trustees of Princeton University].

[123] Jack S. Levy. The Causes of War and the Conditions of Peace. *Annual Review of Political Science*, 1(1):139–165, 1998. _eprint: https://doi.org/10.1146/annurev.polisci.1.1.139.

[124] Jack S. Levy and William R. Thompson. *Causes of War*. John Wiley & Sons, Incorporated, Hoboken, UNITED KINGDOM, 2011.

[125] Yi-Ting Li, Dang-Zheng Liu, and Zheng-Dong Wang. Limit Distributions of Eigenvalues for Random Block Toeplitz and Hankel Matrices. *Journal of Theoretical Probability*, 24(4):1063–1086, December 2011.

[126] Maria A Lim, Jyothsna Chitturi, Valeriya Laskova, Jun Meng, Daniel Findeis, Anne Wiekenberg, Ben Mulcahy, Linjiao Luo, Yan Li, Yangning Lu, et al. Neuroendocrine modulation sustains the C. elegans forward motor state. *Elife*, 5:e19887, 2016.

[127] Scott Linderman and Ryan Adams. Discovering latent network structure in point process data. In *International Conference on Machine Learning*, pages 1413–1421, 2014.

[128] Scott Linderman, Annika Nichols, David Blei, Manuel Zimmer, and Liam Paninski. Hierarchical recurrent state space models reveal discrete and continuous dynamics of neural activity in C. elegans. *bioRxiv*, page 621540, April 2019.

[129] Scott W Linderman, Andrew C Miller, Ryan P Adams, David M Blei, Liam Paninski, and Matthew J Johnson. Recurrent switching linear dynamical systems. *arXiv preprint arXiv:1610.08466*, 2016.

[130] Hexuan Liu, Jimin Kim, and Eli Shlizerman. Functional connectomics from data: Probabilistic graphical models for neuronal network of C. elegans. *arXiv preprint arXiv:1711.00193*, 2017.

[131] Alfred J. Lotka. Analytical note on certain rhythmic relations in organic aystems. *Proceedings of the National Academy of Sciences of the United States of America*, 6(7):410–415, July 1920.

[132] Alfred J. Lotka. *Elements of physical biology*. Williams & Wilkins Company, Baltimore, 1925.

[133] Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9(1):1–10, November 2018.

[134] Christian K. Machens, Ranulfo Romo, and Carlos D. Brody. Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex. *Journal of Neuroscience*, 30(1):350–360, January 2010. Publisher: Society for Neuroscience Section: Articles.

[135] Niall M. Mangan, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Inferring biological networks by sparse identification of nonlinear dynamics. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 2(1):52–63, June 2016.

[136] Zeev Maoz, Lesley G. Terris, Ranan D. Kuperman, and Ilan Talmud. What is the enemy of my enemy? Causes and consequences of imbalanced international relations, 1816 – 2001. *The Journal of Politics*, 69(1):100–115, 2007.

[137] Seth A. Marvel, Jon Kleinberg, Robert D. Kleinberg, and Steven H. Strogatz. Continuous-time model of structural balance. *Proceedings of the National Academy of Sciences*, 108(5):1771–1776, February 2011.

[138] Robert M. May. Will a Large Complex System be Stable? *Nature*, 238(5364):413–414, August 1972. Number: 5364 Publisher: Nature Publishing Group.

[139] Robert M. May, Simon A. Levin, and George Sugihara. Ecology for bankers. *Nature*, 451(7181):893–894, February 2008.

[140] Helen S. Mayberg, Andres M. Lozano, Valerie Voon, Heather E. McNeely, David Seminowicz, Clement Hamani, Jason M. Schwalb, and Sidney H. Kennedy. Deep brain stimulation for treatment-resistant depression. *Neuron*, 45(5):651–660, March 2005.

[141] Richard J McCloskey, Anthony D Fouad, Matthew A Churgin, and Christopher Fang-Yen. Food responsiveness regulates episodic behavioral states in Caenorhabditis elegans. *Journal of neurophysiology*, 117(5):1911–1934, 2017.

[142] Christian Meisel, Andreas Klaus, Christian Kuehn, and Dietmar Plenz. Critical Slowing Down Governs the Transition to Neuron Spiking. *PLOS Computational Biology*, 11(2):e1004097, February 2015.

[143] Manus I. Midlarsky. Preventing Systemic War: Crisis Decision-Making amidst a Structure of Conflict Relationships. *The Journal of Conflict Resolution*, 28(4):563–584, 1984. Publisher: Sage Publications, Inc.

[144] Manus I. Midlarsky. A Hierarchical Equilibrium Theory of Systemic War. *International Studies Quarterly*, 30(1):77–105, 1986. Publisher: [International Studies Association, Wiley].

[145] Atsushi Mochizuki, Bernold Fiedler, Gen Kurosawa, and Daisuke Saito. Dynamics and control at feedback vertex sets. II: A faithful monitor to determine the diversity of molecular activities in regulatory networks. *Journal of Theoretical Biology*, 335:130–146, October 2013.

[146] Megan Morrison, Charles Fieseler, and J. Nathan Kutz. Nonlinear Control in the Nematode C. elegans. *Frontiers in Computational Neuroscience*, 14, 2021. Publisher: Frontiers.

[147] Megan Morrison and Michael Gabbay. Community detectability and structural balance dynamics in signed networks. *Physical Review E*, 102(1):012304, July 2020.

[148] Megan Morrison and Michael Gabbay. The contribution of structure to systemic conflict in signed sociopolitical networks. *(in preparation)*, 2021.

[149] Megan Morrison and J. Nathan Kutz. Nonlinear Control of Networked Dynamical Systems. *IEEE Transactions on Network Science and Engineering*, 8(1):174–189, January 2021. Conference Name: IEEE Transactions on Network Science and Engineering.

[150] Megan Morrison, Pedro D. Maia, and J. Nathan Kutz. Preventing neurodegenerative memory loss in Hopfield neuronal networks using cerebral organoids or external microelectronics. *Computational and Mathematical Methods in Medicine*, 2017:1–13, 2017.

[151] Kyle Moy, Weiyu Li, Huu Phuoc Tran, Valerie Simonis, Evan Story, Christopher Brandon, Jacob Furst, Daniela Raicu, and Hongkyun Kim. Computational methods for tracking, quantitative assessment, and visualization of C. elegans locomotory behavior. *PloS one*, 10(12):e0145870, 2015.

[152] Raj Rao Nadakuditi and M. E. J. Newman. Graph Spectra and the Detectability of Community Structure in Networks. *Physical Review Letters*, 108(18):188701, May 2012.

[153] J. Nagumo, S. Arimoto, and S. Yoshizawa. An active pulse transmission line simulating nerve axon. *Proceedings of the IRE*, 50(10):2061–2070, October 1962. Conference name: Proceedings of the IRE.

[154] Aditya G. Nair, Steven L. Brunton, and Kunihiko Taira. Networked-oscillator-based modeling and control of unsteady wake flows. *Physical Review E*, 97(6):063107, June 2018.

[155] Kiminori Nakamura, George Tita, and David Krackhardt. Violence in the "balance": a structural analysis of how rivals, allies, and third-parties shape inter-gang violence. *Global Crime*, 21(1):3–27, January 2020.

[156] M. E. J. Newman. Mixing patterns in networks. *Physical Review E*, 67(2):026126, February 2003.

[157] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3):036104, 2006.

[158] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E*, 74(3):036104, September 2006.

[159] Mark Newman. *Networks*. Oxford University Press, July 2018. Google-Books-ID: YdZjDwAAQBAJ.

[160] Jeffrey P. Nguyen, Frederick B. Shipley, Ashley N. Linder, George S. Plummer, Mochi Liu, Sagar U. Setru, Joshua W. Shaevitz, and Andrew M. Leifer. Whole-brain calcium imaging with cellular resolution in freely behaving Caenorhabditis elegans. *Proceedings of the National Academy of Sciences*, 113(8):E1074–E1081, February 2016.

[161] Glenn Palmer, Vito D'Orazio, Michael Kenwick, and Matthew Lane. The MID4 dataset, 2002–2010: Procedures, coding rules and description. *Conflict Management and Peace Science*, 32(2):222–242, April 2015. Publisher: SAGE Publications Ltd.

[162] Joel S. Perlmutter and Jonathan W. Mink. Deep brain stimulation. *Annual Review of Neuroscience*, 29(1):229–257, 2006. _eprint: https://doi.org/10.1146/annurev.neuro.29.051605.112824.

[163] David Pfau, Eftychios A Pnevmatikakis, and Liam Paninski. Robust learning of low-dimensional dynamics from large neural ensembles. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2391–2399. Curran Associates, Inc., 2013.

[164] Jonathan T Pierce-Shimomura, Thomas M Morse, and Shawn R Lockery. The fundamental role of pirouettes in Caenorhabditis elegans chemotaxis. *Journal of Neuroscience*, 19(21):9557–9569, 1999.

[165] Robert Prevedel, Young-Gyu Yoon, Maximilian Hoffmann, Nikita Pak, Gordon Wetzstein, Saul Kato, Tina Schrödel, Ramesh Raskar, Manuel Zimmer, Edward S Boyden, and Alipasha Vaziri. Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nature Methods*, 11(7):727–730, July 2014.

[166] I. Prigogine and R. Lefever. Symmetry breaking instabilities in dissipative systems. II. *The Journal of Chemical Physics*, 48(4):1695–1700, February 1968.

[167] I. Prigogine and G. Nicolis. On symmetry-breaking instabilities in dissipative systems. *The Journal of Chemical Physics*, 46(9):3542–3550, May 1967.

[168] Joshua L Proctor, Steven L Brunton, and J Nathan Kutz. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, 2016.

[169] Joshua L. Proctor, Steven L. Brunton, and J. Nathan Kutz. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, January 2016. Publisher: Society for Industrial and Applied Mathematics.

[170] Priscilla E. M. Purnick and Ron Weiss. The second wave of synthetic biology: from modules to systems. *Nature Reviews Molecular Cell Biology*, 10(6):410–422, June 2009.

[171] Misha Rabinovich, Ramon Huerta, and Gilles Laurent. Transient dynamics for neural processing. *Science*, 321(5885):48–50, 2008. Publisher: American Association for the Advancement of Science.

[172] Stefano Recanatesi, Matthew Farrell, Madhu Advani, Timothy Moore, Guillaume Lajoie, and Eric Shea-Brown. Dimensionality compression and expansion in deep neural networks. *arXiv:1906.00443 [cs, stat]*, October 2019. arXiv: 1906.00443.

[173] Lewis F. Richardson. *Arms and insecurity : a mathematical study of the causes and origins of war*. Boxwood Press, Pittsburgh, 1960.

[174] Roberto Rigamonti, Amos Sironi, Vincent Lepetit, and Pascal Fua. Learning separable filters. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2754–2761, 2013.

[175] William M Roberts, Steven B Augustine, Kristy J Lawton, Theodore H Lindsay, Tod R Thiele, Eduardo J Izquierdo, Serge Faumont, Rebecca A Lindsay, Matthew Cale Britton, Navin Pokala, Cornelia I Bargmann, and Shawn R Lockery. A stochastic neuronal model predicts random search behaviors at multiple spatial scales in C. elegans. *eLife*, 5:e12572, January 2016. Publisher: eLife Sciences Publications, Ltd.

[176] William M Roberts, Steven B Augustine, Kristy J Lawton, Theodore H Lindsay, Tod R Thiele, Eduardo J Izquierdo, Serge Faumont, Rebecca A Lindsay, Matthew Cale Britton, Navin Pokala, et al. A stochastic neuronal model predicts random search behaviors at multiple spatial scales in C. elegans. *Elife*, 5, 2016.

[177] Samuel H. Rudy, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, April 2017. Publisher: American Association for the Advancement of Science Section: Research Article.

[178] Paul A. Samuelson. The canonical classical model of political economy. *Journal of Economic Literature*, 16(4):1415–1434, 1978. Publisher: American Economic Association.

[179] Camellia Sarkar and Sarika Jalan. Spectral properties of complex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(10):102101, October 2018.

[180] Marten Scheffer, Stephen R. Carpenter, Timothy M. Lenton, Jordi Bascompte, William Brock, Vasilis Dakos, Johan van de Koppel, Ingrid A. van de Leemput, Simon A. Levin, Egbert H. van Nes, Mercedes Pascual, and John Vandermeer. Anticipating Critical Transitions. *Science*, 338(6105):344–348, 2012. Publisher: American Association for the Advancement of Science.

[181] N. D. Schiff, J. T. Giacino, K. Kalmar, J. D. Victor, K. Baker, M. Gerber, B. Fritz, B. Eisenberg, J. O'Connor, E. J. Kobylarz, S. Farris, A. Machado, C. McCagg, F. Plum, J. J. Fins, and A. R. Rezai. Behavioural improvements with thalamic stimulation after severe traumatic brain injury. *Nature*, 448(7153):600–603, August 2007. Number: 7153 Publisher: Nature Publishing Group.

[182] Tina Schrödel, Robert Prevedel, Karin Aumayr, Manuel Zimmer, and Alipasha Vaziri. Brain-wide 3D imaging of neuronal activity in Caenorhabditis elegans with sculpted light. *Nature Methods*, 10(10):1013–1020, October 2013.

[183] Eric Shiraev and Vladislav Zubok. *International Relations*. Oxford University Press, New York, NY, third edition edition, 2014.

[184] Eli Shlizerman, Jeffrey A Riffell, and J Nathan Kutz. Data-driven inference of network connectivity for modeling the dynamics of neural codes in the insect antennal lobe. *Frontiers in computational neuroscience*, 8:70, 2014. Publisher: Frontiers.

[185] Singer, J. David and Melvin Small. Formal Alliances, 1815-1939. *Journal of Peace Research*, 3:1–31, 1966.

[186] Susanne Skora, Fanny Mende, and Manuel Zimmer. Energy scarcity promotes a brain-wide sleep state modulated by insulin signaling in C. elegans. *Cell Reports*, 22(4):953–966, January 2018. Publisher: Elsevier.

[187] Ron P. Smith. The influence of the richardson arms race model. In Nils Petter Gleditsch, editor, *Lewis Fry Richardson: His Intellectual Legacy and Influence in the Social Sciences*, Pioneers in Arts, Humanities, Science, Engineering, Practice, pages 25–34. Springer International Publishing, Cham, 2020.

[188] Richard Smoke. *War: Controlling Escalation*. Harvard University Press, Cambridge, MA, 1977. Publication Title: War.

[189] Eduardo D. Sontag. *Mathematical control theory: Deterministic finite dimensional systems*. Springer Science & Business Media, November 2013.

[190] Aivar Sootla, Alexandre Mauroy, and Damien Ernst. Optimal control formulation of pulse-based control using Koopman operator. *Automatica*, 91:217–224, May 2018.

[191] Aivar Sootla, Diego Oyarzún, David Angeli, and Guy-Bart Stan. Shaping pulses to control bistable biological systems. In *2015 American Control Conference (ACC)*, pages 3138–3143, July 2015. ISSN: 2378-5861.

[192] Aivar Sootla, Diego Oyarzún, David Angeli, and Guy-Bart Stan. Shaping pulses to control bistable systems: Analysis, computation and counterexamples. *Automatica*, 63:254–264, January 2016.

[193] Greg J Stephens, Matthew Bueno de Mesquita, William S Ryu, and William Bialek. Emergence of long timescales and stereotyped behaviors in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences*, 108(18):7286–7289, 2011.

[194] Greg J Stephens, Bethany Johnson-Kerner, William Bialek, and William S Ryu. Dimensionality and dynamics in the behavior of C. elegans. *PLoS computational biology*, 4(4):e1000028, 2008.

[195] Mark Stopfer, Vivek Jayaraman, and Gilles Laurent. Intensity versus identity coding in an olfactory system. *Neuron*, 39(6):991–1004, September 2003.

[196] D. B. Stouffer, J. Camacho, R. Guimerà, C. A. Ng, and L. A. Nunes Amaral. Quantitative Patterns in the Structure of Model and Empirical Food Webs. *Ecology*, 86(5):1301–1311, 2005. _eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/04-0957.

[197] Stephen Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and Engineering (Studies in Nonlinearity)*. Westview Press, 2001.

[198] Steven H. Strogatz. *Nonlinear dynamics and chaos : with applications to physics, biology, chemistry, and engineering*. Westview Press, a member of the Perseus Books Group, Boulder, CO, second edition edition, 2015.

[199] Tyler H. Summers and Iman Shames. Active influence in dynamical models of structural balance in social networks. *EPL (Europhysics Letters)*, 103(1):18001, July 2013. Publisher: IOP Publishing.

[200] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. LINE: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 1067–1077, Florence, Italy, May 2015. International World Wide Web Conferences Steering Committee.

[201] A. J. P. Taylor. *The struggle for mastery in Europe, 1848-1918.* Oxford,, 1954.

[202] William Thompson and David Dreyer. *Handbook of International Rivalries*. CQ Press, Washington, UNITED STATES, 2011.

[203] William R. Thompson. Polarity, the Long Cycle, and Global Power Warfare. *Journal of Conflict Resolution*, 30(4):587–615, December 1986.

[204] William R. Thompson. A Streetcar Named Sarajevo: Catalysts, Multiple Causation Chains, and Rivalry Structures. *International Studies Quarterly*, 47(3):453–474, September 2003.

[205] V. A. Traag and Jeroen Bruggeman. Community detection in networks with positive and negative links. *Physical Review E*, 80(3):036115, September 2009.

[206] V. A. Traag, P. Doreian, and A. Mrvar. Partitioning signed networks. *arXiv:1803.02082 [physics]*, March 2018. arXiv: 1803.02082.

[207] Vincent Antonio Traag, Paul Van Dooren, and Patrick De Leenheer. Dynamical Models Explaining Social Balance and Evolution of Cooperation. *PLOS ONE*, 8(4):e60063, April 2013.

[208] Lloyd N. Trefethen. *Numerical linear algebra*. Society for Industrial and Applied Mathematics, Philadelphia, 1997.

[209] Laura Trotta, Eric Bullinger, and Rodolphe Sepulchre. Global analysis of dynamical decision-making models through local computation around the hidden saddle. *PLoS ONE*, 7(3):e33110, March 2012.

[210] J. J. P. Veerman. Social Balance and the Bernoulli Equation. *The American Mathematical Monthly*, 125(8):724–732, September 2018. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00029890.2018.1498684.

[211] Vito Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118(2972):558–560, October 1926. Number: 2972 Publisher: Nature Publishing Group.

[212] Tokumitsu Wakabayashi, Izumi Kitagawa, and Ryuzo Shingai. Neurons regulating the duration of forward locomotion in Caenorhabditis elegans. *Neuroscience research*, 50(1):103–111, 2004.

[213] Stephen M. Walt. *The Origins of Alliance*. Cornell University Press, 1990.

[214] Kenneth N. (Kenneth Neal) Waltz. *Theory of international politics*. Addison-Wesley series in political science. Addison-Wesley Pub. Co., Reading, Mass., 1979.

[215] Ruiqi Wang, Tianshou Zhou, Zhujun Jing, and Luonan Chen. Modelling periodic oscillation of biological systems with multiple timescale networks. *IEE Proceedings - Systems Biology*, 1(1):71–84, June 2004. Conference Name: IEE Proceedings - Systems Biology.

[216] Daniel K. Wells, William L. Kath, and Adilson E. Motter. Control of Stochastic and Induced Switching in Biophysical Networks. *Physical Review X*, 5(3):031036, September 2015.

[217] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci*, 314(1165):1–340, 1986.

[218] Eugene P. Wigner. On the Distribution of the Roots of Certain Symmetric Matrices. *Annals of Mathematics*, 67(2):325–327, 1958.

[219] Thomas Wilhelm. The smallest chemical reaction system with bistability. *BMC Systems Biology*, 3(1):90, September 2009.

[220] Mateusz Wilinski, Piero Mazzarisi, Daniele Tantari, and Fabrizio Lillo. Detectability of macroscopic structures in directed asymmetric stochastic block model. *Physical Review E*, 99(4):042310, 2019.

[221] S. Wongkaew, M. Caponigro, K. Kułakowski, and A. Borzì. On the control of the Heider balance model. *The European Physical Journal Special Topics*, 224(17-18):3325–3342, December 2015.

[222] Xiao Zhang, Raj Rao Nadakuditi, and M. E. J. Newman. Spectra of random graphs with community structure and arbitrary degrees. *Physical Review E*, 89(4):042816, April 2014.

[223] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286, June 2006. Publisher: Taylor & Francis.