

©Copyright 2017
Kameron Decker Harris

This Brain Is a Mess: Inference, Random Graphs, and Biophysics to Disentangle Neuronal Networks

Kameron Decker Harris

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

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Eric Shea-Brown, Chair

Stefan Mihalas

Adrienne Fairhall

Program Authorized to Offer Degree:
Applied Mathematics

University of Washington

Abstract

This Brain Is a Mess: Inference, Random Graphs, and Biophysics to Disentangle
Neuronal Networks

Kameron Decker Harris

Chair of the Supervisory Committee:
Professor Eric Shea-Brown
Applied Mathematics

At first glance, the neuronal network seems like a tangled web in many areas throughout the nervous system. Often, our best guess is that such “messy” connections are close to random, while obeying certain statistical constraints, e.g. the number of connections per neuron. However, neuronal wiring is coordinated across larger mesoscopic distances in a way that differentiates between brain layers, areas, and groups of cells. We work across spatial scales in order to understand this hierarchy of order and disorder in brain networks. Ultimately, the goal is to understand how network structure is important for brain function. This leads to:

1. An inference technique which reconstructs mesoscopic brain networks from tracing experiments targeting spatially contiguous groups of neurons.
2. Models of networks which are random, while also having constrained average connectivity and group structure.
3. Comparing simulated and real respiratory rhythms, highlighting the role of inhibitory neurons and connectivity on rhythmogenesis, in particular synchrony and irregularity.

TABLE OF CONTENTS

	Page
Notation	iii
Chapter 1: Connecting network theory, statistics, and physiology	1
1.1 Introduction to networks and graph theory	2
1.2 Finding networks in the real world	4
1.3 Networks built from graph statistics	7
1.4 Biological neural network models	9
1.5 The role of theoretical neuroscience	11
Bibliography	12
Chapter 2: High resolution neural connectivity from incomplete tracing data using nonnegative spline regression	16
2.1 Introduction	17
2.2 Spatial smoothness of mesoscale connectivity	20
2.3 Nonnegative spline regression with incomplete tracing data	21
2.4 Test problem	25
2.5 Finding a voxel-scale connectivity map for mouse cortex	27
2.6 Conclusions	31
Bibliography	32
Chapter 3: Random graph models of communities: spectral gap and its applications	37
3.1 Introduction	37
3.2 Non-backtracking matrix B	42
3.3 Main result	44
3.4 Preliminaries	45
3.5 Proof of Theorem 3	48

3.6	Application: Community detection	73
3.7	Application: Low density parity check or expander codes	81
3.8	Application: Matrix completion	83
	Bibliography	89
Chapter 4:	Different roles for inhibition in the rhythm-generating respiratory network	95
4.1	Introduction	96
4.2	Materials and Methods	98
4.3	Results	110
4.4	Discussion	132
	Bibliography	139
Chapter 5:	Conclusion	149
	Bibliography	150
	List of Figures	154
	List of Tables	160

NOTATION

$\succeq$: entry-wise greater than or equal to, e.g. $W \succeq 0$ means every entry of W is non-negative

$G(V, E)$: graph (directed or undirected) with vertices V and edges E

$\mathbb{N}$: natural numbers $1, 2, 3, \dots$

$\mathbb{R}$: real numbers

$\mathbb{R}_+$: the non-negative real numbers

ACKNOWLEDGMENTS

Thanks to everyone who helped me get through this PhD. When we feel like we are toiling away alone, on something that can seem so specialized, it is our relationships and the teaching we do that make science worth it.

First of all, thank you to my advisor Eric Shea-Brown for being a tireless advocate of me and the rest of his students. You bring a consistently cheerful attitude to your work and life. I am always impressed by your professionalism and drive to create a welcoming and open environment in the lab.

In the Applied Mathematics department, thank you to Keshanie Dissanayake, Lauren Lederer, Derek Franz, Alan Perry, Tony Garcia, and the rest of the staff and faculty. You are the ones that keep everything running!

The support of the rest of the UW computational neuroscience community was essential. In particular, thanks to the other members of our group (past and present): Josh Mendoza, Iris Shi, Doris Voina, Matt Farrell, Ali Weber, Kathleen Champion, Tim Oleskiw, Alex Cayco-Gajic, Guillaume Lajoie, Yu Hu, Braden Brinkman, Hannah Choi, Merav Stern, Gabrielle Gutierrez, and Joel Zylberberg. I appreciate all the help I've received from Adrienne Fairhall and her wonderful group of students. And the many comp neuro happy hours were another joyful release from the frustrations of graduate school life.

Thanks to all of my collaborators as well, especially the committee members who co-advised me on these projects. At the Allen Institute: Stefan Mihalas, Joe Knox, Nile Grad-dis, Nick Cain, David Feng, Lydia Ng, and many others were a great help in getting me going with their data and as the source of many great ideas.

In the Seattle Children's Research Institute: Nino Ramirez, your group has been so

welcoming, and it seems like you guys have the most fun of any lab I know! Thanks to members Tatiana Dashevskiy (sorry for demanding so many *Ns*), Fred Garcia, Aguan Wei, and everyone else.

Thank you to Ioana Dumtriu for your patience teaching me spectral graph theory from the basics, as well as to Marina Meilă. Thank you to my collaborator Gerandy Brito, and congratulations on your baby! I really enjoyed all of our discussions.

Thank you also to the many colleagues and friends I met at conferences and the summer schools in Berkeley, Woods Hole, and Friday Harbor. In particular, Ashok Litwin-Kumar, Haim Sompolinsky, and Larry Abbott, who became my co-authors on a fun project that came from the Woods Hole course.

I also must acknowledge the people who have been with me from the beginning of my foray into research: Chris Danforth and Peter Dodds of UVM. You guys have been such great role models. You were the ones who inspired me to start on this path, and I valued your encouragement highest of all.

The group of friends my (now) wife and I made after moving to Seattle also deserves a huge shout out. In particular, the first friend Aaron Brooks, whom I met in Santa Fe in what feels like another life, thanks for welcoming us into your tribe of so many cool mountain people cross scientists. The massive ski community of Seattle has been another great source of support. There's no better way to refresh the mind and find inspiration than glide silently through the old growth in the early morning, or thread down a glaciated volcano on a fine spring afternoon.

Finally, I must thank my family: my dad, mom, and Jim, along with my grandma, who are always impressed by the scientific stuff although I am not sure you ever really want me to try and explain it. Also, I have the best and most supportive wife, Meira, and I have loved becoming part of your family. When I get frustrated, I only have to think about you and the fourth-graders to get my perspective straight.

DEDICATION

To the glaciers, may you stick around.



Sunrise over the Granite glacier, Adamant Range, British Colombia

Chapter 1

CONNECTING NETWORK THEORY, STATISTICS, AND PHYSIOLOGY

If you look at a piece of brain under a microscope, it will probably look like a mess. Every neuron connects to thousands of others on average, in a way that appears random. If we zoom out, however, hierarchical structure reveals itself. In the cortex, for example, neurons are organized into layers and cell-types, which are furthermore organized into different areas. I study how to measure and model these “messy” networks given incomplete data, both from an inference standpoint and using “toy” random networks whose parameters are easily measureable. I also examine how that structure shapes brain dynamics, in particular synchrony and rhythms.

The brain is a fascinating system which allows us to actively engage with the world through sensation, cognition, and action. Brains are both complicated—intricately composed of many particular parts—and complex—their behavior emerges from the interactions of many neurons connected in a network [Anderson, 1972]. Because of this complexity, reductionism alone will not solve the brain. Instead, we need constructive theories for how assemblies of networked neurons perform computations. These computations rely on network structure.

Neurons are the main information processing units of the brain. They function through rapid fluctuations of their membrane potentials called *spikes*. Spikes are transmitted directionally from the cell body along the *axon*. These potentials are generated and propagated by the action of voltage-sensitive ion channels embedded in the membrane. Neurons receive information via their *dendrites*. Axons and dendrites can be highly branched to

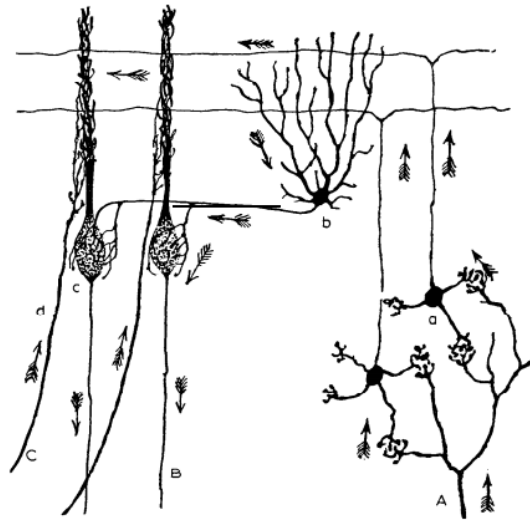


Figure 1.1: An illustration of cerebellar neurons by [Ramón y Cajal \[1906\]](#). Arrows indicate the direction of spike transmission from dendrites to axons. Cajal inferred this correct directionality from the geometry of the neurons alone, before detailed electrophysiological properties of neurons were known. If all of the cerebellar neurons in this area were pictured, we would see a jungle of overlapping axons, dendrites, and cell bodies.

communicate with many thousands of other neurons (Fig. 1.1). When a spike reaches the axon terminals, it causes the release of neurotransmitter molecules which diffuse towards dendrites of receiving neurons. Receptors specialized for that neurotransmitter transduce the signal into a change in electrical potential of the receiving cell, which may elevate (excite) or depress (inhibit) the membrane potential of the downstream neuron. Neurons which communicate this way are *synaptically connected*. Synapses are the most prevalent type of connection in the brain, although there are others like gap junctions (which act like a direct conductance between cells) and coupling through glia.

1.1 Introduction to networks and graph theory

The interconnected web of neurons in the brain forms a *network*. Network is another word for a *graph*, a collection of vertices V and edges E between them. We also use nodes and links to refer to vertices and edges, respectively. Network and graph have different con-

notations, networks being more commonly used in the physics community and graphs among mathematicians. [Hackett \[2011, Section 1.2.1\]](#) puts this nicely:

To model a complex system as a graph is to filter out the functional details of each of its components, and the idiosyncrasies of their interactions with each other, and to focus instead on the underlying structure (topology) as an inert mathematical construct. Although this technique is central also to network theory, the word network, in contrast, usually carries with it connotations of the context in which the overarching system exists, particularly when that system displays any sort of nonlinear dynamics.

In the networks we consider here, nodes usually represent neurons or collections of neurons connected by synaptic links. However, the results may be more generally applicable, and in Chapter 3 we consider abstract graphs which may represent a code, community of people, etc. These neurons and synapses have their own nonlinear dynamics which determine how they generate and transmit spikes. A popular term for any comprehensive whole-brain neural network is *connectome*.

We now introduce some terminology from graph theory which is essential to this research. For a detailed introduction, see [West \[2001\]](#) and [Newman \[2010\]](#). It will be useful to represent the connectivity of the graph with the *adjacency matrix* A , a $|V| \times |V|$ matrix. Entry $A_{ij} = 1$ if node j connects to node i , i.e. $(j, i) \in E$, and $A_{ij} = 0$ otherwise. Sometimes the edges will have weights denoting the strength of a connection, and this will then be the *weight matrix*. The number of neighbors of a node is its *degree*.

If the edges each point one way or the other, the graph is *directed*. In a directed graph, the adjacency matrix is asymmetric, in contrast to a simple graph which has $A = A^T$. We use *in-degree* and *out-degree* for incoming and outgoing edges in a directed graph. A graph is called *d-regular* if every vertex has the same degree d .

In this work, we are generally focused on the relevance of graph theory to dynamical systems which take place on such graphs. Because algebra is one of our most powerful tools for studying dynamical systems, these are the tools we bring to bear on these networks. Many questions about dynamics on graphs, such as the stability of various so-

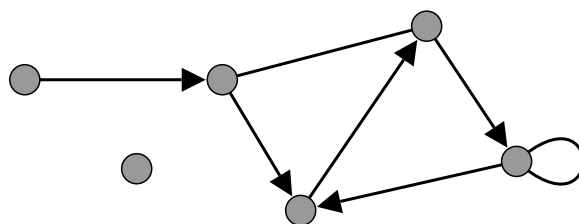


Figure 1.2: A small directed graph. It contains a connected component of 5 nodes and one isolated node. The majority of the edges are directed. There is one undirected edge, shown without arrows. There is also a self-loop.

lutions, are really questions about the eigenvalues and eigenvectors of A and related matrices. However, graph theory is a well-developed area with deep connections to many mathematical fields such as combinatorics (extremal graphs, Ramsey theory, subgraph enumeration), group theory (Cayley graphs, graph symmetries, algebraic constructions of expanders), analysis (Feynman diagrams applied to perturbative expansions), and probability (random graphs). The applications of these various aspects range from computer science, social science, ecology, systems biology and genetics, to physics.

1.2 Finding networks in the real world

Measuring networks is often difficult, especially in the messier biological sciences. If a directed network contains $N = |V|$ vertices, then an experimenter must test all N^2 possible ordered pairs of vertices for the presence of an edge. In neuroscience, there are both anatomical and electrophysiological techniques to measure synaptic connections between single neurons [Sporns, 2010].

The classical electrophysiological technique is to induce spikes in one neuron and see whether it evokes a response in another. Practically, the experimenter must first patch onto each neuron with an electrode. A current is injected into the first neuron which causes it to spike. If the second neuron's membrane potential reliably increases or decreases following the stimulation, possibly evoking a spike, then the first neuron is presy-

naptic to it. The change in the membrane potential, either up or down, indicates whether the presynaptic neuron is excitatory or inhibitory, and the amount of change gives the strength of the synapse.

The anatomical technique measures the fine-scale geometry of neurons, classifies axons and dendrites, and identifies likely connections between them at places such as synaptic boutons. Neuroanatomy began in the nineteenth century with the pioneering work of Santiago Ramón y Cajal, who applied the staining method developed by Camillo Golgi, which sparsely labels neurons. This sparse labeling is incredibly fortuitous, since otherwise all one would see is a tangled mess. He went on to describe many features of the nervous system of animals and humans, proposed that neurons are the building-blocks of brain computation (the neuron hypothesis), proposed the directionality of signal transmission from axons to dendrites, and produced many beautiful illustrations; see Fig. 1.1, reproduced from Ramón y Cajal [1906]. However, staining techniques do not allow for precise identification of synaptic connections and fail to label many neurons. Nowadays, anatomists use three-dimensional electron microscopy (EM) to visualize the finest structure every neurons in a given volume. Image stacks are segmented, usually in a semi-automated manner, to identify, neurons, axons, dendrites, boutons, etc. With these detailed data, we can untangle the mess of neurons.

Each of these approaches can unambiguously identify the presence or absence of synaptic connections at the single neuron level. There have been a number of successes of paired recordings in small circuits, such as central pattern generators in crustaceans, lampreys, and insects [Marder and Calabrese, 1996, Marder and Bucher, 2001]. The technique has also been used in cortical networks to identify the abundances of small motifs by performing simultaneous quadruple recordings [Song et al., 2005]. One famous success of EM reconstructions was the measurement of the first connectome, for the nematode *C. elegans* [White et al., 1986]. However, EM reconstructions have been performed of small circuits, like fly retina, along with small cortical volumes on the order of 10 μm wide [Kleinfeld et al., 2011, Bock et al., 2011, Glickfeld et al., 2013].

The technology to perform network characterization and reconstruction is advancing rapidly. Electrophysiological recordings are inherently limited in the number of neurons that a single neuron can record from due to the inevitable decline of neuron health over time in slices. However, fully-automated cell identification and patching techniques will speed up this process and capture more cells [Suk et al., 2017]. With increased computing storage and power, and improved experimental techniques, the scale of EM datasets that can be collected and processed is growing quickly.

The previous discussion concerned testing for synaptic connections in a network of neurons, which gives a very fine-scale, detailed description: the micro-connectome. However, as the human brain contains roughly 10^9 neurons total, with an average degree of 10^4 , these techniques can only take us so far. Even storing such a matrix, with 10^{13} entries, will be difficult. This roadblock, along with the understanding that the brain is organized into a hierarchy of systems (areas, columns, layers) built up of many neurons, has leads us to consider networks defined at coarser scales, often called the meso- and macro-connectomes [Sporns, 2010].

Current technology is nicely matched to measuring structural connectivity at meso-scale resolutions on the order of 10–100 μm . Large team efforts have performed numerous tracing experiments in fly [Jenett et al., 2012, Peng et al., 2014] and mouse [Oh et al., 2014, Kuan et al., 2015], and they are planned for marmoset [Okano et al., 2015]. In these experiments, a viral or similar tracer is delivered to a specific source site in the brain. This tracer labels the neurons in that area, either highlighting where their axons go (anterograde tracing), or where the axons of neurons presynaptic to the source site neurons come from (retrograde tracing). These experiments are repeated for many source sites across the brain to gather data on brain-connectivity. An example of one of these experiments is shown in Fig. 1.3.

If only a single tracing experiment is performed for each brain region, it is relatively straightforward to build a region-by-region macro-connectome. For an injection delivered precisely to source region j , we can normalize the projection density to region i by the

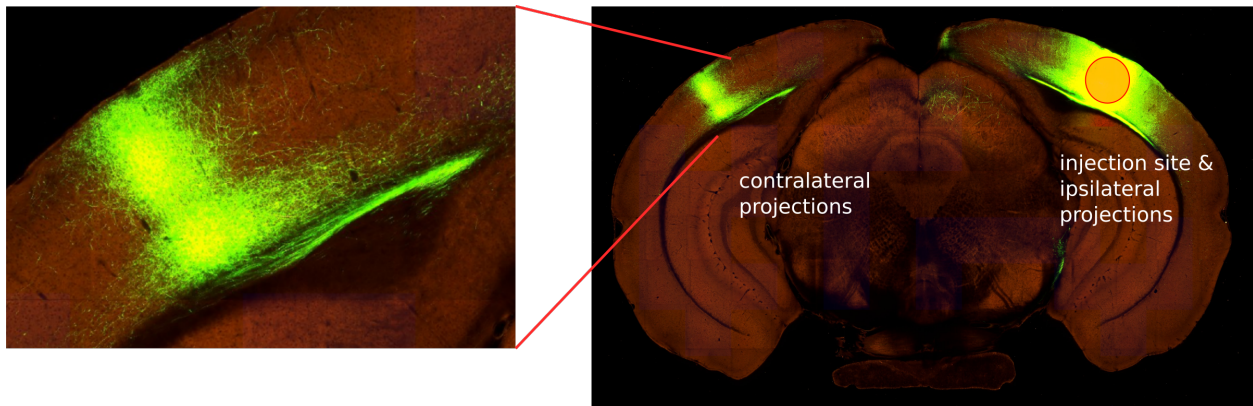


Figure 1.3: Example of meso-scale connectivity data collected via viral tracing. In this experiment, the tracer was injected in the right hemisphere. Axonal projections of neurons whose cell bodies are within the injection site are visible as green fluorescence on both the ipsilateral and contralateral sides. These images show just one of many coronal slices which are stitched together into a 3-D image of fluorescence intensity. These data were collected by the Allen Institute for Brain Science [Oh et al., 2014] and are available online at <http://connectivity.brain-map.org>.

strength of injection and assign this to the weight W_{ij} . However, for situations where there are multiple, possibly inconsistent, injections in the source regions, a regression method is more appropriate [Oh et al., 2014]. Chapter 2 details one method we developed to improve connectome inference and go from a region-specific resolution to a spatially explicit meso-connectome.

1.3 Networks built from graph statistics

Families of random networks are popular ways to model real networks where a full adjacency or weight matrix is not available [Bollobás, 2001, Newman, 2010]. The simplest model, now called the Erdős-Rényi graph, adds edges independently at random with a fixed probability. These graphs have only two parameters, the number of vertices N and the connection probability p . This makes them attractive for modeling networks, e.g. complicated cortical circuits, where it would be impossible to test each connection but easy

to characterize the average probability of a connection. Other classical random graphs include the d -regular random graph and the random graph with $|E|$ edges. These graphs are now well-characterized and nearly everything one might want to know about them has been proven.

In the late 1990s, there was an explosion of interest in graphs that could capture properties of the Internet and other technological and biological networks. These networks often have highly skewed or power-law degree distributions. Classical random graph models do not have this property; the Erdős-Rényi graph, for example, has degrees which are binomially distributed. [Barabási and Albert \[1999\]](#) made the celebrated preferential attachment model to describe how power-law graphs may be generated. Many of these networks also have the small-world property, where the average distance between any two nodes is small while also containing many triangles [[Watts and Strogatz, 1998](#)]. There now exist a multitude of random graph models that contain interesting tunable features like degree distributions, triangle or clique counts, and community structure. The explosion of papers which cover these and other topics have led to the creation of network science as its own discipline [[Newman, 2010](#)].

Community structure in a network refers to a partitioning of nodes into clusters with most connections occurring within each cluster rather than between them. The simplest random network model of community formation, the stochastic block or planted partition model, assigns edges with probability p_{in} within communities and p_{out} between, where $p_{\text{in}} > p_{\text{out}}$ [[Newman, 2010](#)]. Different types of edges, perhaps reflecting neurotransmitter or receptor types with varying dynamics, is another way to form communities. Network communities are very likely how the various neural systems are partitioned into areas or sub-circuits which manage different brain functions. In Chapter 4, we show how dividing a single network into two communities, with predominantly inhibitory connections between communities, can segment a rhythm into two separate phases.

In Chapter 3, we study a generalized version of the d -regular random graph which now has two communities. The vertices V are split into two sets V_1 and V_2 , and edges

are only allowed between the two sets. Any graph where this is possible is called a *bipartite* graph. If the sets are unbalanced, i.e. $|V_1| \neq |V_2|$, then the degrees of vertices in the sets must also be different. Therefore, we call our model the bipartite, biregular random graph. Chapter 3 details various properties of the spectrum of the adjacency matrix for this model. We mention a number of applications of our model for error correcting codes, machine learning, and community detection. While neural networks have certainly evolved to perform similar types of representation and learning of the world, our application is not fundamentally neuroscientific. However, the spectral properties we study are also important for any dynamical system (including a neural one) taking place on such a graph. For instance, they determine whether and how fast coupled oscillators will synchronize [Arenas et al., 2008].

1.4 Biological neural network models

Neural network models began with the binary threshold units of McCulloch and Pitts [1943], who showed that, when wired in the right network, their artificial neurons could compute any logical function. These ideas were extended by Rosenblatt [1958] to form his perceptron, a linear classifier with a training algorithm that formed the basis for artificial neural networks employed to great success [e.g. Silver et al., 2016] in the machine learning boom happening now.

Simultaneously, biophysicists began to isolate and measure ionic currents that create the spikes and more complicated behaviors in real neurons. Hodgkin and Huxley [1952], in their seminal series of papers, used careful voltage clamp experiments to measure the opening and closing kinetics of voltage-dependent sodium and potassium channels in the squid giant axon. Using these data, they created an accurate model of how these channels interact to initiate and propagate spikes. In their careful analysis, the powers which best fit the activation rates of each channel, amazingly, correctly predicted their cooperative binding characteristics many decades before ion channel protein structures were measureable. It is a testament to the power of their modeling approach that it still

is used today, with modifications for the many other types of channels which have since been discovered.

The basic sodium and potassium conductances are adequate to describe spike generation and tonic spiking (limit cycle) behavior. Other channels can lead to different dynamical behavior. For instance, slowly activating or inactivating currents are typically present when neurons fire bursts of spikes, with a quiescent period in between. Rhythmically bursting neurons are common throughout the brain [Buzsaki, 2006], especially in pattern generating circuits [Marder and Bucher, 2001, Grillner, 2006, Grillner and Jessell, 2009, Kiehn, 2011]. The slow timescale causes a periodic cycle of termination or activation of tonic spiking. We use the mathematical techniques of singular perturbation theory, averaging, etc. to study this type of behavior. In general, single neuron dynamics are classified in terms of the bifurcations (normal forms) neurons exhibit across their slow and fast systems [Izhikevich, 2000].

Hodgkin-Huxley type models simplify neuronal dynamics into a set of voltage-dependent currents coupled together with Kirchhoff's circuit laws. In order to model network effects, one must introduce synaptic currents coming from other neurons, which depend on both the mechanisms of neurotransmitter release, diffusion, and signal transmission in the postsynaptic neuron [Destexhe et al., 1994]. Much of the mechanisms for learning depend on neural activity leading to strengthening or weakening of synapses [beginning with Hebb, 1949]. In models of learning and memory, as well as firing rate adaptation, we must capture long- and short-term potentiation of synapses. Furthermore, the actual morphology of axons and dendrites can lead to nonlinear processing effects, leading to the use of spatially extended as opposed to “point” neuron models.

All of this of course ignores many other levels of complexity. Blood flow and oxygenation, interactions between neurons and astrocytes, physical strain on neurons (relevant, for example, when someone experiences a concussion), transcription's regulation of signalling proteins, neuromodulators or drugs—all of these things affect neural dynamics.

In Chapter 4, we study a simple but biophysically realistic model of respiratory rhythm

generation. This process depends on both the intrinsic properties of neurons which lead to rhythmic bursts and the structure of their network, in particular the amount of inhibition and whether those inhibitory cells target their projections to specific cell groups.

1.5 The role of theoretical neuroscience

It should now be clear that there are myriad choices to make when designing a model of any neural system. Even with complete knowledge of a synaptic network and a realistic model for the neurons, it will be hard to predict collective neural behavior. More realistic, and thus more complicated, models involve many more parameters, most of which are unknown in the system of interest. The most important role of the theoretician is in deciding what kind of description is best for the problem at hand.

We prefer a simple description for a more profound reason than needing a model with fewer parameters. Simple models tell us *which ingredients are essential for the effect we care about*. In paring down to the core mechanisms, we gain interpretability. Simplicity reveals how structure gives rise to its function.

Neuroscience is presently undergoing a transition from a data-poor to a data-rich science. Only a few years ago, recording from more than a few neurons in a circuit as very difficult. Now, with multi-electrode arrays and calcium imaging, we can record from hundreds to thousands of neurons in a freely behaving animal. The methods of data collection are constantly changing and pushing the envelopes of detectability.

Because of this constant innovation in experimental methodology, computational scientists are increasingly important for developing algorithms to denoise, sort, and analyze this data. Then, we computational neuroscientists can help bridge from the data to an understanding of the mechanisms of neural computation. This involves comparing the data to models of those theoretical mechanisms in action. Finally, we can propose new mechanisms through a close understanding of the underlying mathematics of neural computation. These theories must exist at both biophysical and more abstract levels.

This is why my research covers all of these aspects: how to measure networks with real

data (Chapter 2), mathematical models for that network structure (Chapter 3), and how network structure gives rise to dynamics, in particular respiratory rhythm generation (Chapter 4). The interplay of these approaches will be part of neuroscience research for the foreseeable future.

Bibliography

P. W. Anderson. More is different. *Science (New York, N.Y.)*, 177(4047):393–396, August 1972. ISSN 0036-8075. doi: 10.1126/science.177.4047.393.

Alex Arenas, Albert Díaz-Guilera, Jurgen Kurths, Yamir Moreno, and Changsong Zhou. Synchronization in complex networks. *Physics Reports*, 469(3):93–153, December 2008. ISSN 03701573. doi: 10.1016/j.physrep.2008.09.002.

Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, October 1999. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.286.5439.509.

Davi D. Bock, Wei-Chung Allen Lee, Aaron M. Kerlin, Mark L. Andermann, Greg Hood, Arthur W. Wetzel, Sergey Yurgenson, Edward R. Soucy, Hyon Suk Kim, and R. Clay Reid. Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, 471(7337):177–182, March 2011. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature09802.

Béla Bollobás. *Random Graphs*. Cambridge University Press, Cambridge, 2001. ISBN 978-0-511-81406-8.

Gyorgy Buzsaki. *Rhythms of the Brain*. Oxford University Press, August 2006. ISBN 978-0-19-804125-2.

Alain Destexhe, Zachary F. Mainen, and Terrence J. Sejnowski. Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism. *Journal of Computational Neuroscience*, 1(3):195–230, August 1994. ISSN 0929-5313, 1573-6873. doi: 10.1007/BF00961734.

Lindsey L. Glickfeld, Mark L. Andermann, Vincent Bonin, and R. Clay Reid. Cortico-cortical projections in mouse visual cortex are functionally target specific. *Nature Neuroscience*, 16(2):219–226, February 2013. ISSN 1097-6256. doi: 10.1038/nn.3300.

Sten Grillner. Biological Pattern Generation: The Cellular and Computational Logic of Networks in Motion. *Neuron*, 52(5):751–766, December 2006. ISSN 0896-6273. doi: 10.1016/j.neuron.2006.11.008.

- Sten Grillner and Thomas M Jessell. Measured motion: Searching for simplicity in spinal locomotor networks. *Current Opinion in Neurobiology*, 19(6):572–586, December 2009. ISSN 0959-4388. doi: 10.1016/j.conb.2009.10.011.
- Adam W. Hackett. *Cascade Dynamics on Complex Networks*. PhD thesis, University of Limerick, 2011.
- Donald O Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Wiley, New York, 1949.
- 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. ISSN 0022-3751.
- Eugene M. Izhikevich. Neural excitability, spiking and bursting. *International Journal of Bifurcation and Chaos*, 10(06):1171–1266, June 2000. ISSN 0218-1274. doi: 10.1142/S0218127400000840.
- Arnim Jenett, Gerald M. Rubin, Teri-T. B. Ngo, David Shepherd, Christine Murphy, Heather Dionne, Barret D. Pfeiffer, Amanda Cavallaro, Donald Hall, Jennifer Jeter, Nirmala Iyer, Dona Fetter, Joanna H. Hausenfluck, Hanchuan Peng, Eric T. Trautman, Robert R. Svirskas, Eugene W. Myers, Zbigniew R. Iwinski, Yoshinori Aso, Gina M. DePasquale, Adrienne Enos, Phuson Hulamm, Shing Chun Benny Lam, Hsing-Hsi Li, Todd R. Lavery, Fuhui Long, Lei Qu, Sean D. Murphy, Konrad Rokicki, Todd Safford, Kshiti Shaw, Julie H. Simpson, Allison Sowell, Susana Tae, Yang Yu, and Christopher T. Zugates. A GAL4-Driver Line Resource for *Drosophila* Neurobiology. *Cell Reports*, 2(4):991–1001, October 2012. ISSN 2211-1247. doi: 10.1016/j.celrep.2012.09.011.
- Ole Kiehn. Development and functional organization of spinal locomotor circuits. *Current Opinion in Neurobiology*, 21(1):100–109, February 2011. ISSN 0959-4388. doi: 10.1016/j.conb.2010.09.004.
- David Kleinfeld, Arjun Bharioke, Pablo Blinder, Davi D. Bock, Kevin L. Briggman, Dmitri B. Chklovskii, Winfried Denk, Moritz Helmstaedter, John P. Kaufhold, Wei-Chung Allen Lee, Hanno S. Meyer, Kristina D. Micheva, Marcel Oberlaender, Steffen Prohaska, R. Clay Reid, Stephen J. Smith, Shinya Takemura, Philbert S. Tsai, and Bert Sakmann. Large-Scale Automated Histology in the Pursuit of Connectomes. *The Journal of Neuroscience*, 31(45):16125–16138, September 2011. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.4077-11.2011.
- Leonard Kuan, Yang Li, Chris Lau, David Feng, Amy Bernard, Susan M. Sunkin, Hongkui Zeng, Chinh Dang, Michael Hawrylycz, and Lydia Ng. Neuroinformatics of the Allen Mouse Brain Connectivity Atlas. *Methods*, 73:4–17, February 2015. ISSN 1046-2023. doi: 10.1016/j.ymeth.2014.12.013.

- Eve Marder and Dirk Bucher. Central pattern generators and the control of rhythmic movements. *Current Biology*, 11(23):R986–R996, November 2001. ISSN 0960-9822. doi: 10.1016/S0960-9822(01)00581-4.
- Eve Marder and Ronald L. Calabrese. Principles of rhythmic motor pattern generation. *Physiological reviews*, 76(3):687–717, 1996.
- Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, December 1943. ISSN 0007-4985, 1522-9602. doi: 10.1007/BF02478259.
- Mark Newman. *Networks: An Introduction*. Oxford University Press, March 2010. ISBN 978-0-19-150070-1.
- Seung Wook Oh, Julie A. Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, Chris Lau, Leonard Kuan, Alex M. Henry, Marty T. Mortrud, Benjamin Ouellette, Thuc Nghi Nguyen, Staci A. Sorensen, Clifford R. Slaughterbeck, Wayne Wakeman, Yang Li, David Feng, Anh Ho, Eric Nicholas, Karla E. Hirokawa, Phillip Bohn, Kevin M. Joines, Hanchuan Peng, Michael J. Hawrylycz, John W. Phillips, John G. Hohmann, Paul Wahnoutka, Charles R. Gerfen, Christof Koch, Amy Bernard, Chinh Dang, Allan R. Jones, and Hongkui Zeng. A mesoscale connectome of the mouse brain. *Nature*, 508(7495):207–214, April 2014. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature13186.
- Hideyuki Okano, Atsushi Miyawaki, and Kiyoto Kasai. Brain/MINDS: Brain-mapping project in Japan. *Phil. Trans. R. Soc. B*, 370(1668):20140310, May 2015. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2014.0310.
- Hanchuan Peng, Jianyong Tang, Hang Xiao, Alessandro Bria, Jianlong Zhou, Victoria Butler, Zhi Zhou, Paloma T. Gonzalez-Bellido, Seung W. Oh, Jichao Chen, Ananya Mitra, Richard W. Tsien, Hongkui Zeng, Giorgio A. Ascoli, Giulio Iannello, Michael Hawrylycz, Eugene Myers, and Fuhui Long. Virtual finger boosts three-dimensional imaging and microsurgery as well as terabyte volume image visualization and analysis. *Nature Communications*, 5, July 2014. doi: 10.1038/ncomms5342.
- Santiago Ramón y Cajal. Nobel Lecture: The Structure and Connexions of Neurons, 1906.
- F. Rosenblatt. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, 65(6):386–408, 1958.
- David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis

- Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, January 2016. ISSN 0028-0836. doi: 10.1038/nature16961.
- Sen Song, Per Jesper Sjöström, Markus Reigl, Sacha Nelson, and Dmitri B. Chklovskii. Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits. *PLOS Biology*, 3(3):e68, March 2005. ISSN 1545-7885. doi: 10.1371/journal.pbio.0030068.
- Olaf Sporns. *Networks of the Brain*. The MIT Press, 1st edition, 2010. ISBN 0-262-01469-6 978-0-262-01469-4.
- Ho-Jun Suk, Ingrid van Welie, Suhasa B. Kodandaramaiah, Brian Allen, Craig R. Forest, and Edward S. Boyden. Closed-Loop Real-Time Imaging Enables Fully Automated Cell-Targeted Patch-Clamp Neural Recording In Vivo. *Neuron*, 95(5):1037–1047.e11, August 2017. ISSN 0896-6273. doi: 10.1016/j.neuron.2017.08.011.
- Duncan J. Watts and Steven H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, June 1998. ISSN 0028-0836. doi: 10.1038/30918.
- Douglas West. *Introduction to Graph Theory*. Prentice Hall, 2 edition, 2001. ISBN 0-13-014400-2.
- J. G. White, E. Southgate, J. N. Thomson, and S. Brenner. The Structure of the Nervous System of the Nematode *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 314(1165):1–340, November 1986. ISSN 0080-4622. doi: 10.1098/rstb.1986.0056.

Chapter 2

HIGH RESOLUTION NEURAL CONNECTIVITY FROM INCOMPLETE TRACING DATA USING NONNEGATIVE SPLINE REGRESSION

Kameron Decker Harris, Stefan Mihalas, Eric Shea-Brown.
In proceedings of NIPS, 2016.

Abstract

Whole-brain neural connectivity data are now available from viral tracing experiments, which reveal the connections between a source injection site and elsewhere in the brain. These hold the promise of revealing spatial patterns of connectivity throughout the mammalian brain. To achieve this goal, we seek to fit a weighted, nonnegative adjacency matrix among 100 μm brain “voxels” using viral tracer data. Despite a multi-year experimental effort, injections provide incomplete coverage, and the number of voxels in our data is orders of magnitude larger than the number of injections, making the problem severely underdetermined. Furthermore, projection data are missing within the injection site because local connections there are not separable from the injection signal.

We use a novel machine-learning algorithm to meet these challenges and develop a spatially explicit, voxel-scale connectivity map of the mouse visual system. Our method combines three features: a matrix completion loss for missing data, a smoothing spline penalty to regularize the problem, and (optionally) a low rank factorization. We demonstrate the consistency of our estimator using synthetic data and then apply it to newly available Allen Mouse Brain Connectivity Atlas data for the visual system. Our algorithm is significantly more predictive than current state of the art approaches which assume regions to be homogeneous. We demonstrate the efficacy of a low rank version on visual cortex data and discuss the possibility of extending this to a whole-brain connectivity matrix at the voxel scale.

2.1 Introduction

Although the study of neural connectivity is over a century old, starting with pioneering neuroscientists who identified the importance of networks for determining brain function, most knowledge of anatomical neural network structure is limited to either detailed description of small subsystems [White et al., 1986, Kleinfeld et al., 2011, Bock et al., 2011, Glickfeld et al., 2013] or to averaged connectivity between larger regions [Felleman and Van Essen, 1991, ?]. We focus our attention on *spatial, structural* connectivity at the *mesoscale*: a coarser scale than that of single neurons or cortical columns but finer than whole brain regions. Thanks to the development of new tracing techniques, image processing algorithms, and high-throughput methods, data at this resolution are now accessible in animals such as the fly [Jenett et al., 2012, Peng et al., 2014] and mouse [Oh et al., 2014, Kuan et al., 2015]. We present a novel regression technique tailored to the challenges of learning spatially refined mesoscale connectivity from neural tracing experiments. We have designed this technique with neural data in mind and will use this language to describe our method, but it is a general technique to assimilate spatial network data or infer smooth kernels of integral equations. Obtaining a spatially-resolved mesoscale connectome will reveal detailed features of connectivity, for example unlocking cell-type specific connectivity and microcircuit organization throughout the brain [Jonas and Kording, 2015].

In mesoscale anterograde tracing experiments, a tracer virus is first injected into the brain. This infects neurons primarily at their cell bodies and dendrites and causes them to express a fluorescent protein in their cytoplasm, including in their axons. Neurons originating in the source injection site are then imaged to reveal their axonal projections throughout the brain. Combining many experiments with different sources then reveals the pathways that connect those sources throughout the brain. This requires combining data across multiple animals, which appears justified at the mesoscale [Oh et al., 2014].

We assume there exists some underlying nonnegative, weighted adjacency matrix

$W \succeq 0$ that is common across animals. Each experiment can be thought of as an injection $\mathbf{x}$, and its projections $\mathbf{y}$, so that $\mathbf{y} \approx W\mathbf{x}$ as in Fig. 2.1A. Uncovering the unknown W from multiple experiments $(\mathbf{x}_i, \mathbf{y}_i)$ for $i = 1, \dots, n_{\text{inj}}$ is then a multivariate regression problem: Each $\mathbf{x}_i$ is an image of the brain which represents the strength of the signal within the injection site. Likewise, every $\mathbf{y}_i$ is an image of the strength of signal elsewhere, which arises due to the axonal projections of neurons with cell bodies in the injection site. The unknown matrix W is a linear operator which takes images of the brain (injections) and returns images of the brain (projections).

In a previous paper, [Oh et al. \[2014\]](#) were able to obtain a 213×213 regional weight matrix using 469 experiments with mice (Fig. 2.1B). They used nonnegative least squares to find the unknown regional weights in an overdetermined regression problem. Our aim is to obtain a much higher-resolution connectivity map on the scale of voxels, and this introduces many more challenges.

First, the number of voxels in the brain is much larger than the number of injection experiments we can expect to perform; for mouse with $100 \mu\text{m}$ voxels this is $O(10^5)$ versus $O(10^3)$ [[Oh et al., 2014](#), [Kuan et al., 2015](#)]. Also, the injections that are performed will inevitably leave gaps in their coverage of the brain. Thus specifying W is underdetermined. Second, there is no way to separately image the injections and projections. In order to construct them, experimenters image the brain once by serial tomography and fluorescence microscopy. The injection sites can be annotated by finding infected cell bodies, but there is no way to disambiguate fluorescence from the cell bodies and dendrites from that of local injections. Projection strength is thus unknown within the injection sites and the neighborhood occupied by dendrites. Third, fitting full-brain voxel-wise connectivity is challenging since the number of elements in W is the square of the number of voxels in the brain. Thus we need compressed representations of W as well as efficient algorithms to perform inference. The paper proceeds as follows.

In Section 2.2, we describe our assumption that the mesoscale connectivity W is smoothly-varying in space, as could be expected from the presence of topographic maps across

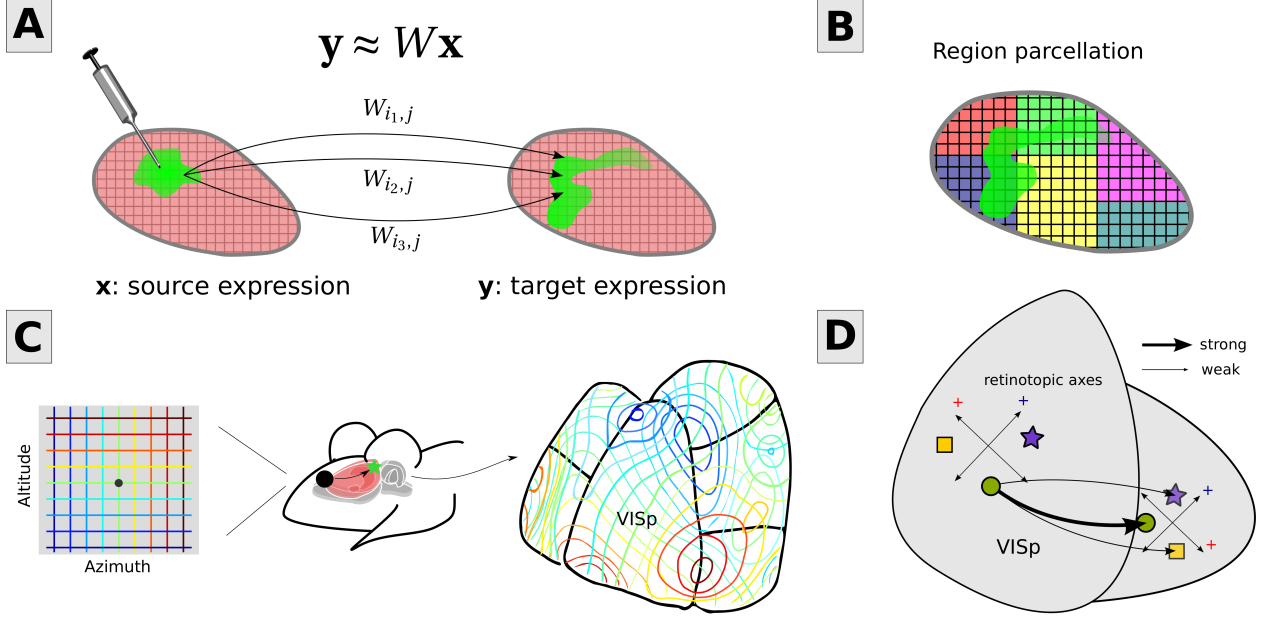


Figure 2.1: **A**, We seek to fit a matrix W which reproduces neural tracing experiments. Each column of W represents the expected signal in target voxels given an injection of one unit into a single source voxel. **B**, In the work of [Oh et al. \[2014\]](#), a regionally homogeneous connectivity matrix was fit using a predefined regional parcellation to constrain the problem. We propose that smoothness of W is a better prior. **C**, The mouse’s visual field can be represented in azimuth/altitude coordinates. This representation is maintained in the retinotopy, a smoothly varying map replicated in many visual areas (e.g. [Garrett et al., 2014](#)). **D**, Assuming locations in VISp (the primary visual area) project most strongly to positions which represent the same retinotopic coordinates in a secondary visual area, then we expect the mapping between upstream and downstream visual areas to be smooth.

much of cortex. Later, we show that using this assumption as a prior yields connectivity maps with improved cross-validation performance.

In Section 2.3, we present an inference algorithm designed to tackle the difficulties of underdetermination, missing data, and size of the unknown W . To deal with the gaps and ill-conditioning, we use smoothness as a regularization on W . We take an agnostic approach, similar to matrix completion [\[Candes and Plan, 2010\]](#), to the missing projection data and use a regression loss function that ignores residuals within the injection site.

Finally, we present a low rank version of the estimator that will allow us to scale to large matrices.

In Section 2.4, we test our method on synthetic data and show that it performs well for sparse data that is consistent with the regression priors. This provides evidence that it is a consistent estimator. We demonstrate the necessity of both the matrix completion and smoothing terms for good reconstruction.

In Section 2.5, we then apply the spline-smoothing method to recently available Allen Institute for Brain Science (Allen Institute) connectivity data from mouse visual cortex [Oh et al., 2014, Kuan et al., 2015]. We find that our method is able to outperform current spatially uniform regional models, with significantly reduced cross-validation errors. We also find that a low rank version is able to achieve approximately $23\times$ compression of the original data, with the optimal solution very close to the full rank optimum. Our method is a superior predictor to the existing regional model for visual system data, and the success of the low rank version suggests that this approach will be able to reveal whole-brain structural connectivity at unprecedented scale.

All of our supplemental material and data processing and optimization code is available for download from:

<https://github.com/kharris/high-res-connectivity-nips-2016>.

2.2 *Spatial smoothness of mesoscale connectivity*

The visual cortex is a collection of relatively large cortical areas in the posterior part of the mammalian brain. Visual stimuli sensed in the retina are relayed through the thalamus into primary visual cortex (VISp), which projects to higher visual areas. We know this partly due to tracing projections between these areas, but also because neurons in the early visual areas respond to visual stimuli in a localized region of the visual field called their *receptive fields* [Hubel and Wiesel, 1962].

An interesting and important feature of visual cortex is the presence of topographic

maps of the visual field called the *retinotopy* [Goodman and Shatz, 1993, Rosa and Tweedale, 2005, Wang and Burkhalter, 2007, Chaplin et al., 2013, Garrett et al., 2014]. Each eye sees a 2-D image of the world, where two coordinates, such as azimuth and altitude, define a point in the visual field (Fig. 2.1C). Retinotopy refers to the fact that cells are organized in cortical space by the position of their receptive fields; nearby cells have similar receptive field positions. Furthermore, these retinotopic maps reoccur in multiple visual areas, albeit with varying orientation and magnification.

Retinotopy in other areas downstream from VISp, which do not receive many projections directly from thalamus, are likely a function of projections from VISp. It is reasonable to assume that areas which code for similar visual locations are most strongly connected. Then, because retinotopy is smoothly varying in cortical space and similar retinotopic coordinates are the most strongly connected between visual areas, the connections between those areas should be smooth in cortical space (Fig. 2.1C and D).

Retinotopy is a specific example of topography, which extends to other sensory systems such as auditory and somatosensory cortex [Udin and Fawcett, 1988]. For this reason, connectivity may be spatially smooth throughout the brain, at least at the mesoscale. This idea can be evaluated via the methods we introduce below: if a smooth model is more predictive of held-out data than another model, then this supports the assumption.

2.3 *Nonnegative spline regression with incomplete tracing data*

We consider the problem of fitting an adjacency operator $\mathcal{W} : T \times S \rightarrow \mathbb{R}_+$ to data arising from n_{inj} injections into a source space S which projects to a target space T . Here S and T are compact subsets of the brain, itself a compact subset of $\mathbb{R}^3$. In this mathematical setting, S and T could be arbitrary sets, but typically $S = T$ for the ipsilateral data we present here.¹ The source S and target T are discretized into n_x and n_y cubic voxels, respectively. The discretization of $\mathcal{W}$ is then an adjacency matrix $W \in \mathbb{R}_+^{n_y \times n_x}$. Mathematically, we

¹Ipsilateral refers to connections within the same cerebral hemisphere. For contralateral (opposite hemisphere) connectivity, S and T are disjoint subsets of the brain corresponding to the two hemispheres.

define the tracing data as a set of pairs $\mathbf{x}_i \in \mathbb{R}_+^{n_x}$ and $\mathbf{y}_i \in \mathbb{R}_+^{n_y}$, the source and target tracer signals at each voxel for experiments $i = 1, \dots, n_{\text{inj}}$. We would like to fit a linear model, a matrix W such that $\mathbf{y}_i \approx W\mathbf{x}_i$. We assume an observation model

$$\mathbf{y}_i = W\mathbf{x}_i + \eta_i$$

with $\eta_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2 I)$ multivariate Gaussian random variables with zero mean and covariance matrix $\sigma^2 I \in \mathbb{R}^{n_y \times n_y}$. The true data are not entirely linear, due to saturation effects of the fluorescence signal, but the linear model provides a tractable way of “credit assignment” of individual source voxels’ contributions to the target signal [Oh et al., 2014].

Finally, we assume that the target projections are unknown within the injection site. In other words, we only know $\mathbf{y}_j$ outside the support of $\mathbf{x}_j$, which we denote $\text{supp } \mathbf{x}_j$, and we wish to only evaluate error for the observable voxels. Let $\Omega \in \mathbb{R}^{n_y \times n_{\text{inj}}}$, where the j th column $\Omega_j = 1 - \mathbb{1}_{\text{supp } \mathbf{x}_j}$, the indicator of the complement of the support. We define the orthogonal projector $P_\Omega : \mathbb{R}^{n_y \times n_{\text{inj}}} \rightarrow \mathbb{R}^{n_y \times n_{\text{inj}}}$ as $P_\Omega(A) = A \circ \Omega$, the entrywise product of A and Ω . This operator zeros elements of A which correspond to the voxels within each experiment’s injection site. The operator P_Ω is similar to what is used in matrix completion [Candes and Plan, 2010], here in the context of regression rather than recovery.

These assumptions lead to a loss function which is the familiar ℓ^2 -loss applied to the projected residuals:

$$\frac{1}{\sigma^2 n_{\text{inj}}} \|P_\Omega(WX - Y)\|_F^2 \quad (2.1)$$

where $Y = [\mathbf{y}_1, \dots, \mathbf{y}_{n_{\text{inj}}}]$ and $X = [\mathbf{x}_1, \dots, \mathbf{x}_{n_{\text{inj}}}]$ are data matrices. Here $\|\cdot\|_F$ is the Frobenius norm, i.e. the ℓ^2 -norm of the matrix as a vector: $\|A\|_F = \|\text{vec}(A)\|_2$, where $\text{vec}(A)$ takes a matrix and converts it to a vector by stacking consecutive columns.

We next construct a regularization penalty. The matrix W represents the spatial discretization of a two-point kernel $\mathcal{W}$. An important assumption for $\mathcal{W}$ is that it is spatially smooth. Function space norms of the derivatives of $\mathcal{W}$, viewed as a real-valued function

on $T \times S$, are a natural way to measure the roughness of this function. For this study, we chose the squared L^2 -norm of the Laplacian

$$\int_{T \times S} |\Delta \mathcal{W}|^2 \, dy dx,$$

which is called the thin plate spline bending energy [Wahba, 1990]. In the discrete setting, this becomes the squared ℓ^2 -norm of a discrete Laplacian applied to W :

$$\|L \operatorname{vec}(W)\|_2^2 = \|L_y W + W L_x^T\|_F^2. \quad (2.2)$$

The operator $L : \mathbb{R}^{n_y n_x} \rightarrow \mathbb{R}^{n_y n_x}$ is the discrete Laplacian operator or second finite difference matrix on $T \times S$. The equality in Eqn. (2.2) results from the fact that the Laplacian on the product space $T \times S$ can be decomposed as $L = L_x \otimes I_{n_y} + I_{n_x} \otimes L_y$ [Lynch et al., 1964]. Using the well-known Kronecker product identity for linear matrix equations

$$(B^T \otimes A) \operatorname{vec}(X) = \operatorname{vec}(Y) \iff AXB = Y \quad (2.3)$$

gives the result in Eqn. (2.2) [Van Loan, 2000], which allows us to efficiently evaluate the Laplacian action. As for boundary conditions, we do not want to impose any particular values at the boundary, so we choose the finite difference matrix corresponding to a homogeneous Neumann (zero derivative) boundary condition.²

Combining the loss and penalty terms, Eqn. (2.1) and (2.2), gives a convex optimization problem for inferring the connectivity:

$$W^* = \arg \min_{W \succeq 0} \|P_\Omega(WX - Y)\|_F^2 + \lambda \frac{n_{\text{inj}}}{n_x} \|L_y W + W L_x^T\|_F^2. \quad (\text{P1})$$

In the final form, we absorb the noise variance σ^2 into the regularization hyperparameter λ and rescale the penalty so that it has the same dependence on the problem size n_x, n_y , and n_{inj} as the loss. We solve the optimization (P1) using the L-BFGS-B projected quasi-Newton method, implemented in C++ [Byrd et al., 1995, Boyd and Vandenberghe, 2004]. The gradient is efficiently computed using matrix algebra.

²It is straightforward to avoid smoothing across region boundaries by imposing Neumann boundary conditions at the boundaries; this is an option in our code available online.

Note that (P1) is a type of nonnegative least squares problem, since we can use Eqn. (2.3) to convert it into

$$w^* = \arg \min_{w \geq 0} \|Aw - y\|_2^2 + \lambda \frac{n_{\text{inj}}}{n_x} \|Lw\|_2^2,$$

where $A = \text{diag}(\text{vec}(\Omega)) (X^T \otimes I_{n_y})$, $y = \text{diag}(\text{vec}(\Omega)) \text{vec}(Y)$, and $w = \text{vec}(W)$. Furthermore, without the nonnegativity constraint the estimator is linear and has an explicit solution. However, the design matrix A will have dimension $(n_y n_{\text{inj}}) \times (n_y n_x)$, with $O(n_y^3 n_{\text{inj}})$ entries if $n_x = O(n_y)$. The dimensionality of the problem prevents us from working directly in the tensor product space. And since the model is a structured matrix regression problem [Argyriou et al., 2009], the usual representer theorems [Wahba, 1990], which reduce the dimensionality of the estimator to effectively the number of data points, do not immediately apply. However, we hope to elucidate the connection to reproducing kernel Hilbert spaces in future work.

2.3.1 Low rank version

The largest object in our problem is the unknown connectivity W , since in the underconstrained setting $n_{\text{inj}} \ll n_x, n_y$. In order to improve the scaling of our problem with the number of voxels, we reformulate it with a compressed version of W :

$$(U^*, V^*) = \arg \min_{U, V \geq 0} \|P_\Omega(UV^T X - Y)\|_F^2 + \lambda \frac{n_{\text{inj}}}{n_x} \|L_y UV^T + UV^T L_x^T\|_F^2. \quad (\text{P2})$$

Here, $U \in \mathbb{R}_+^{n_y \times r}$ and $V \in \mathbb{R}_+^{n_x \times r}$ for some fixed rank r , so that the optimal connectivity $W^* = U^* V^{*T}$ is given in low rank, factored form. Note that we use nonnegative factors rather than constrain $UV^T \succeq 0$, since this is a nonlinear constraint.

This has the advantage of automatically computing a nonnegative matrix factorization (NMF) of W . The NMF is of separate scientific interest, to be pursued in future work, since it decomposes the connectivity into a relatively small number of projection patterns, which has interpretations as a clustering of the connectivity itself.

In going from the full rank problem (P1) to the low rank version (P2), we lose convexity. So the usual optimization methods are not guaranteed to find a global optimum,

and the clustering just mentioned is not unique. However, we have also reduced the size of the unknowns to the potentially much smaller matrices U and V , if $r \ll n_y, n_x$. If $n_x = O(n_y)$, we have only $O(n_y r)$ unknowns instead of $O(n_y^2)$. Evaluating the penalty term still requires computation of $n_y n_x$ terms, but this can be performed without storing them in memory.

We use a simple projected gradient method with Nesterov acceleration in Matlab to find a local optimum for (P2) [Boyd and Vandenberghe, 2004], and will present and compare these results to the solution of (P1) below. As before, computing the gradients is efficient using matrix algebra. This method has been used before for NMF [Lin, 2007].

2.4 Test problem

We next apply our algorithms to a test problem consisting of a one-dimensional “brain,” where the source and target space $S = T = [0, 1]$. The true connectivity kernel corresponds to a Gaussian profile about the diagonal plus a bump:

$$W_{\text{true}}(x, y) = \exp \left\{ - \left(\frac{x - y}{0.4} \right)^2 \right\} + 0.9 \exp \left\{ - \frac{(x - 0.8)^2 + (y - 0.1)^2}{(0.2)^2} \right\}.$$

See the left panel of Fig. 2.2. The input and output spaces were discretized using $n_x = n_y = 200$ points. Injections are delivered at random locations within S , with a width of $0.12 + 0.1\epsilon$ where $\epsilon \sim \text{Uniform}(0, 1)$. The values of $\mathbf{x}$ are set to 1 within the injection region and 0 elsewhere, $\mathbf{y}$ is set to 0 within the injection region, and we take noise level $\sigma = 0.1$. The matrices $L_x = L_y$ are the 5-point finite difference Laplacians for the rectangular lattice.

Example output of (P1) and (P2) is given for 5 injections in Fig. 2.2. Unless stated otherwise, $\lambda = 100$. The injections, depicted as black bars in the bottom of each sub-figure, do not cover the whole space S but do provide good coverage of the bump, otherwise there is no information about that feature. We depict the result of the full rank algorithm (P1) without the matrix completion term P_Ω , the result including P_Ω but without smoothing ($\lambda = 0$), and the result of (P2) with rank $r = 20$. The full rank solution is not shown, but is similar to the low rank one.

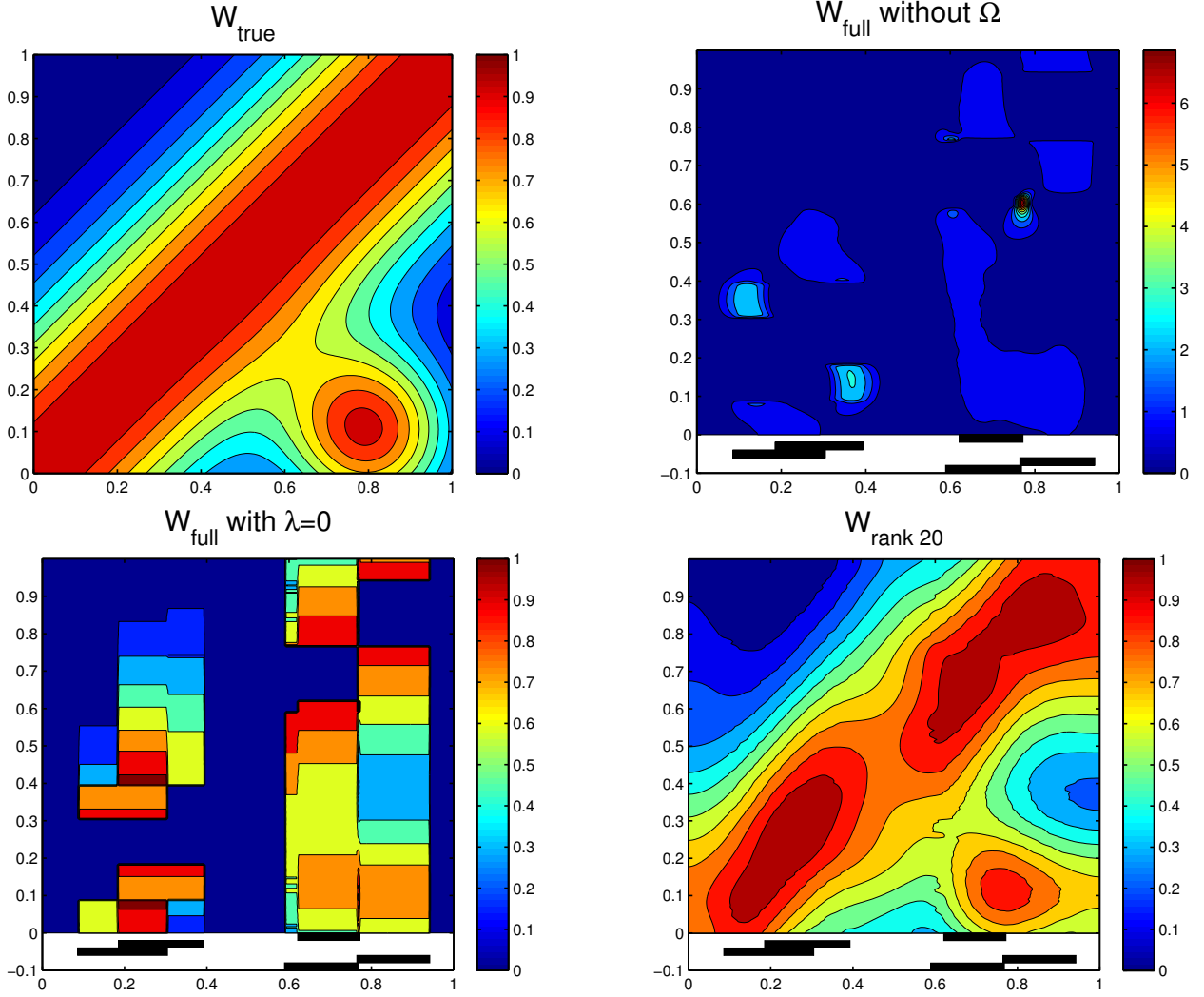


Figure 2.2: Comparison of the true (**Above left**) and inferred connectivity from 5 injections. Unless noted, $\lambda = 100$. **Above right**, we show the what happens when we solve (P1) without the matrix completion term P_Ω . The holes in the projection data cause patchy and incorrect output. Note the colorbar range is $6\times$ that in the other cases. **Below left** is the result with P_Ω but without regularization, solving (P1) for $\lambda = 0$. There, the solution does not interpolate between injections. **Below right** is a rank $r = 20$ result using (P2), which captures the diagonal band and off-diagonal bump that make up W_{true} . In this case, the low rank result has less relative error (9.6%) than the full rank result (11.1%, not shown).

Figure 2.2 shows the necessity of each term within the algorithm. Leaving out the matrix completion P_Ω leads to dramatically biased output since the algorithm uses incorrect values $\mathbf{y}_{\text{supp}(\mathbf{x})} = 0$. If we include P_Ω but neglect the smoothing term by setting $\lambda = 0$, we also get incorrect output: without smoothing, the algorithm cannot fill in the injection site holes nor can it interpolate between injections. However, the low rank result accurately approximates the true connectivity W_{true} , including the diagonal profile and bump, achieving 9.6% relative error measured as $\|W^* - W_{\text{true}}\|_F / \|W_{\text{true}}\|_F$. The full rank version is similar, but in fact has slightly higher 11.1% relative error.

2.5 Finding a voxel-scale connectivity map for mouse cortex

We next apply our method to the latest data from the Allen Institute Mouse Brain Connectivity Atlas, obtained with the API at <http://connectivity.brain-map.org>. Briefly, in each experiment mice were injected with adeno-associated virus expressing a fluorescent protein. The virus infects neurons in the injection site, causing them to produce the protein, which is transported throughout the axonal and dendritic processes. The mouse brains for each experiment were then sliced, imaged, and aligned onto the common coordinates in the Allen Reference Atlas version 3 [Oh et al., 2014, Kuan et al., 2015]. These coordinates divide the brain volume into $100 \mu\text{m} \times 100 \mu\text{m} \times 100 \mu\text{m}$ voxels, with approximately 5×10^5 voxels in the whole brain. The fluorescent pixels in each aligned image were segmented from the background, and we use the fraction of segmented versus total pixels in a voxel to build the vectors $\mathbf{x}$ and $\mathbf{y}$. Since cortical dendrites project locally, the signal outside the injection site is mostly axonal, and so the method reveals anterograde axonal projections from the injection site.

From this dataset, we selected 28 experiments which have 95% of their injection volumes contained within the visual cortex (atlas regions VISal, VISam, VISl, VISp, VISpl, VISpm, VISli, VISpor, VISrl, and VISa) and injection volume less than 0.7 mm^3 . For this study, we present only the results for ipsilateral connectivity, where $S = T$ and $n_x = n_y = 7497$. To compute the smoothing penalty, we used the 7-point finite-difference

Model	Voxel MSE_{rel}	Regional MSE_{rel}
Regional	107% (70%)	48% (6.8%)
Voxel	33% (10%)	16% (2.3%)

Table 2.1: Model performance on Allen Institute Mouse Brain Connectivity Atlas data. Cross-validation errors of the voxel model (P1) and regionally homogeneous models are shown, with training errors in parentheses. The errors are computed in both voxel space and regional space, using the relative mean squared error MSE_{rel} , Eqn. (2.4). In either space, the voxel model shows reduced training and cross-validation errors relative to the regional model.

Laplacian on the cubic voxel lattice.

In order to evaluate the performance of the estimator, we employ nested cross-validation with 5 inner and outer folds. The full rank estimator (P1) was fit for $\lambda = 10^3, 10^4, \dots, 10^{12}$ on the training data. Using the validation data, we then selected the λ_{opt} that minimized the mean square error relative to the average squared norm of the prediction WX and truth Y , evaluating errors outside the injection sites:

$$\text{MSE}_{\text{rel}} = \frac{2\|P_{\Omega}(WX - Y)\|_F^2}{\|P_{\Omega}(WX)\|_F^2 + \|P_{\Omega}(Y)\|_F^2}. \quad (2.4)$$

This choice of normalization prevents experiments with small $\|Y\|$ from dominating the error. This error metric as well as the ℓ^2 -loss adopted in Eqn. (P1) both more heavily weight the experiments with larger signal. After selection of λ_{opt} , the model was refit to the combined training and validation data. In our dataset, $\lambda_{\text{opt}} = 10^5$ was selected for all outer folds. The final errors were computed with the test datasets in each outer fold. For comparison, we also fit a regional model within the cross-validation framework, using nonnegative least squares. To do this, similar to the study by Oh et al. [2014], we constrained the connectivity $W_{kl} = W_{R_i R_j}$ to be constant for all voxels k in region R_i and l in region R_j .

The results are shown in Table 2.1. Errors were computed according to both voxels and regions. For the latter, we integrated the residual over voxels within the regions be-

fore computing the error. The voxel model is more predictive of held-out data than the regional model, reducing the voxel and regional MSE_{rel} by 69% and 67%, respectively. The regional model is designed for inter-region connectivity. To allow an easier comparison with the voxel model, we here include within region connections. We find that the regional model is a poor predictor of voxel scale projections, with over 100% relative voxel error, but it performs okay at the regional scale. The training errors, which reflect goodness of fit, were also reduced significantly with the voxel model. We conclude that the more flexible voxel model is a better estimator for these Allen Institute data, since it improves both the fits to training data as well as cross-validation skill.

The inferred visual connectivity also exhibits a number of features that we expect. There are strong local projections (similar to the diagonal in the test problem, Fig. 2.2) along with spatially organized projections to higher visual areas. See Fig. 2.3, which shows example projections from source voxels within VISp. These are just two of 7497 voxels in the full matrix, and we depict only a 2-D projection of 3-D images. The connectivity exhibits strong local projections, which must be filled in by the smoothing since within the injection sites the projection data are unknown; it is surprising how well the algorithm does at capturing short-range connectivity that is translationally invariant. There are also long-range bumps in the higher visual areas, medial and lateral, which move with the source voxel. This is a result of retinotopic maps between VISp and downstream areas. The supplementary material presents a view of this high-dimensional matrix in movie form, allowing one to see the varying projections as the seed voxel moves. *We encourage the reader to view the supplemental movies, where movement of bumps in downstream regions hints at the underlying retinotopy:* <https://github.com/kharris/high-res-connectivity-nips-2016>.

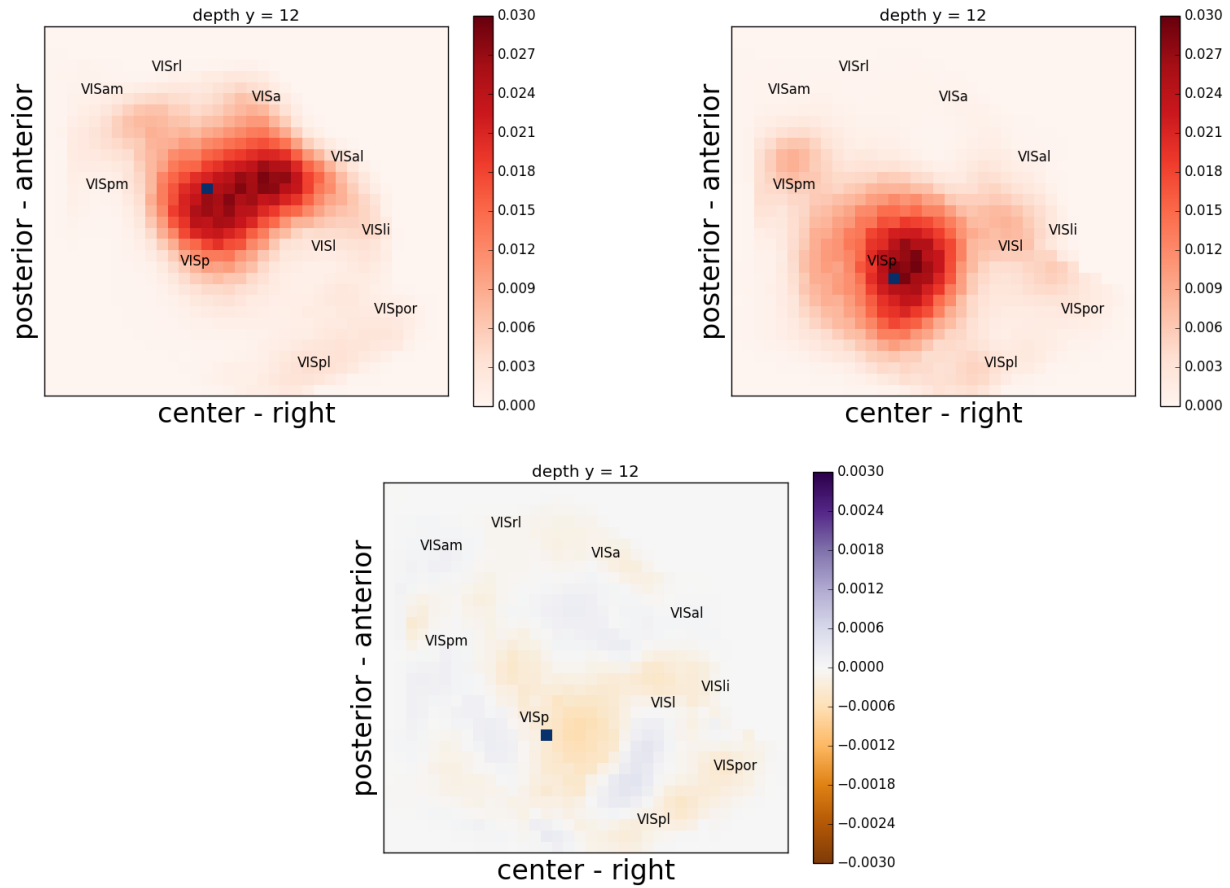


Figure 2.3: Inferred connectivity using all 28 selected injections from visual system data. **Above left**, Projections from a source voxel (blue) located in VISp to all other voxels in the visual areas. The view is integrated over the superior-inferior axis. The connectivity shows strong local connections and weaker connections to higher areas, in particular VISam, VISal, and VISl. Movies of the inferred connectivity (full, low rank, and the low rank residual) for varying source voxel are available in the supplementary material. **Above right**, For a source 800 μm away, the pattern of anterograde projections is similar, but the distal projection centers are shifted, as expected from retinotopy. **Below**, The residuals between the full rank and rank 160 result from solving (P2), for the same source voxel as in the center. The residuals are an order of magnitude less than typical features of the connectivity.

2.5.1 Low rank inference successfully approximates full rank solution for visual system

We next use these visual system data, for which the full rank solution was computed, to test whether the low rank approximation can be applied. This is an important stepping stone to an eventual inference of spatial connectivity for the full brain.

First, we note that the singular value spectrum of the fitted W_{full}^* (now using all 28 injections and $\lambda = 10^5$) is heavily skewed: 95% of the energy can be captured with 21 of 7497 components, and 99% with 67 components. However, this does not directly imply that a nonnegative factorization will perform as well. To test this, we fit a low rank decomposition directly to all 28 visual injection data using (P2) with rank $r = 160$ and $\lambda = 10^5$. The output of the optimization procedure yields U^* and V^* , and we find that the low rank output is very similar to the full result W_{full}^* fit to the same data (see also Fig. 2.3, which visualizes the residuals):

$$\frac{\|U^*V^{*T} - W_{\text{full}}^*\|_F}{\|W_{\text{full}}^*\|_F} = 13\%.$$

This close approximation is despite the fact that the low rank solution achieves a roughly $23\times$ compression of the 7497×7497 matrix.

Assuming similar compressibility for the whole brain, where the number of voxels is 5×10^5 , would mean a rank of approximately 10^4 . This is still a problem in $O(10^9)$ unknowns, but these bring the memory requirements of storing one matrix iterate in double precision from approximately 1.9 TB to 75 GB, which is within reach of commonly available large memory machines.

2.6 Conclusions

We have developed and implemented a new inference algorithm that uses modern machine learning ideas—matrix completion loss, a smoothing penalty, and low rank factorization—to assimilate sparse connectivity data into complete, spatially explicit connectivity maps. We have shown that this method can be applied to the latest Allen Institute data from

multiple visual cortical areas, and that it significantly improves cross-validated predictions over the current state of the art and unveils spatial patterning of connectivity. Finally, we show that a low rank version of the algorithm produces very similar results on these data while compressing the connectivity map, potentially opening the door to the inference of whole brain connectivity from viral tracer data at the voxel scale.

Acknowledgements

We acknowledge the support of the UW NIH Training Grant in Big Data for Neuroscience and Genetics (KDH), Boeing Scholarship (KDH), NSF Grant DMS-1122106 and 1514743 (ESB & KDH), and a Simons Fellowship in Mathematics (ESB). We thank Liam Paninski for helpful insights at the outset of this project. We wish to thank the Allen Institute founders, Paul G. Allen and Jody Allen, for their vision, encouragement, and support. This work was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington.

Bibliography

- Andreas Argyriou, Charles A. Micchelli, and Massimiliano Pontil. When Is There a Representer Theorem? Vector Versus Matrix Regularizers. *J. Mach. Learn. Res.*, 10:2507–2529, December 2009. ISSN 1532-4435.
- Davi D. Bock, Wei-Chung Allen Lee, Aaron M. Kerlin, Mark L. Andermann, Greg Hood, Arthur W. Wetzel, Sergey Yurgenson, Edward R. Soucy, Hyon Suk Kim, and R. Clay Reid. Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, 471(7337):177–182, March 2011. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature09802.
- Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004. ISBN 0-521-83378-7.
- R. Byrd, P. Lu, J. Nocedal, and C. Zhu. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, September 1995. ISSN 1064-8275. doi: 10.1137/0916069.
- E. J. Candes and Y. Plan. Matrix Completion With Noise. *Proceedings of the IEEE*, 98(6): 925–936, June 2010. ISSN 0018-9219. doi: 10.1109/JPROC.2009.2035722.

- Tristan A. Chaplin, Hsin-Hao Yu, and Marcello G.P. Rosa. Representation of the visual field in the primary visual area of the marmoset monkey: Magnification factors, point-image size, and proportionality to retinal ganglion cell density. *Journal of Comparative Neurology*, 521(5):1001–1019, April 2013. ISSN 1096-9861. doi: 10.1002/cne.23215.
- Daniel J. Felleman and D. C. Van Essen. Distributed Hierarchical Processing in the Primate. *Cerebral Cortex*, 1(1):1–47, January 1991. ISSN 1047-3211, 1460-2199. doi: 10.1093/cercor/1.1.1.
- M. E. Garrett, I. Nauhaus, J. H. Marshel, and E. M. Callaway. Topography and Areal Organization of Mouse Visual Cortex. *Journal of Neuroscience*, 34(37):12587–12600, September 2014. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.1124-14.2014.
- Lindsey L. Glickfeld, Mark L. Andermann, Vincent Bonin, and R. Clay Reid. Cortico-cortical projections in mouse visual cortex are functionally target specific. *Nature Neuroscience*, 16(2):219–226, February 2013. ISSN 1097-6256. doi: 10.1038/nn.3300.
- Corey S. Goodman and Carla J. Shatz. Developmental mechanisms that generate precise patterns of neuronal connectivity. *Cell*, 72, Supplement:77–98, January 1993. ISSN 0092-8674. doi: 10.1016/S0092-8674(05)80030-3.
- D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of Physiology*, 160(1):106–154.2, January 1962. ISSN 0022-3751.
- Arnim Jenett, Gerald M. Rubin, Teri-T. B. Ngo, David Shepherd, Christine Murphy, Heather Dionne, Barret D. Pfeiffer, Amanda Cavallaro, Donald Hall, Jennifer Jeter, Nirmala Iyer, Dona Fetter, Joanna H. Hausenfluck, Hanchuan Peng, Eric T. Trautman, Robert R. Svirskas, Eugene W. Myers, Zbigniew R. Iwinski, Yoshinori Aso, Gina M. DePasquale, Adrienne Enos, Phuson Hulamm, Shing Chun Benny Lam, Hsing-Hsi Li, Todd R. Lavery, Fuhui Long, Lei Qu, Sean D. Murphy, Konrad Rokicki, Todd Safford, Kshiti Shaw, Julie H. Simpson, Allison Sowell, Susana Tae, Yang Yu, and Christopher T. Zugates. A GAL4-Driver Line Resource for Drosophila Neurobiology. *Cell Reports*, 2(4):991–1001, October 2012. ISSN 2211-1247. doi: 10.1016/j.celrep.2012.09.011.
- Eric Jonas and Konrad Kording. Automatic discovery of cell types and microcircuitry from neural connectomics. *eLife*, 4:e04250, April 2015. ISSN 2050-084X. doi: 10.7554/eLife.04250.
- David Kleinfeld, Arjun Bharioke, Pablo Blinder, Davi D. Bock, Kevin L. Briggman, Dmitri B. Chklovskii, Winfried Denk, Moritz Helmstaedter, John P. Kaufhold, Wei-Chung Allen Lee, Hanno S. Meyer, Kristina D. Micheva, Marcel Oberlaender, Steffen Prohaska, R. Clay Reid, Stephen J. Smith, Shinya Takemura, Philbert S. Tsai, and Bert Sakmann. Large-Scale Automated Histology in the Pursuit of Connectomes. *The Journal*

- of Neuroscience*, 31(45):16125–16138, September 2011. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.4077-11.2011.
- Leonard Kuan, Yang Li, Chris Lau, David Feng, Amy Bernard, Susan M. Sunkin, Hongkui Zeng, Chinh Dang, Michael Hawrylycz, and Lydia Ng. Neuroinformatics of the Allen Mouse Brain Connectivity Atlas. *Methods*, 73:4–17, February 2015. ISSN 1046-2023. doi: 10.1016/j.ymeth.2014.12.013.
- Chih-Jen Lin. Projected Gradient Methods for Nonnegative Matrix Factorization. *Neural Computation*, 19(10):2756–2779, October 2007. ISSN 0899-7667. doi: 10.1162/neco.2007.19.10.2756.
- R. E. Lynch, J. R. Rice, and D. H. Thomas. Tensor product analysis of partial difference equations. *Bulletin of the American Mathematical Society*, 70(3):378–384, 1964. ISSN 0002-9904, 1936-881X. doi: 10.1090/S0002-9904-1964-11105-8.
- Seung Wook Oh, Julie A. Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, Chris Lau, Leonard Kuan, Alex M. Henry, Marty T. Mortrud, Benjamin Ouellette, Thuc Nghi Nguyen, Staci A. Sorensen, Clifford R. Slaughterbeck, Wayne Wakeman, Yang Li, David Feng, Anh Ho, Eric Nicholas, Karla E. Hirokawa, Phillip Bohn, Kevin M. Joines, Hanchuan Peng, Michael J. Hawrylycz, John W. Phillips, John G. Hohmann, Paul Wohnoutka, Charles R. Gerfen, Christof Koch, Amy Bernard, Chinh Dang, Allan R. Jones, and Hongkui Zeng. A mesoscale connectome of the mouse brain. *Nature*, 508(7495):207–214, April 2014. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature13186.
- Hanchuan Peng, Jianyong Tang, Hang Xiao, Alessandro Bria, Jianlong Zhou, Victoria Butler, Zhi Zhou, Paloma T. Gonzalez-Bellido, Seung W. Oh, Jichao Chen, Ananya Mitra, Richard W. Tsien, Hongkui Zeng, Giorgio A. Ascoli, Giulio Iannello, Michael Hawrylycz, Eugene Myers, and Fuhui Long. Virtual finger boosts three-dimensional imaging and microsurgery as well as terabyte volume image visualization and analysis. *Nature Communications*, 5, July 2014. doi: 10.1038/ncomms5342.
- Marcello G.P Rosa and Rowan Tweeddale. Brain maps, great and small: Lessons from comparative studies of primate visual cortical organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456):665–691, April 2005. ISSN 0962-8436. doi: 10.1098/rstb.2005.1626.
- S B Udin and J W Fawcett. Formation of Topographic Maps. *Annual Review of Neuroscience*, 11(1):289–327, 1988. doi: 10.1146/annurev.ne.11.030188.001445.
- Charles F. Van Loan. The ubiquitous Kronecker product. *Journal of Computational and Applied Mathematics*, 123(1–2):85–100, November 2000. ISSN 0377-0427. doi: 10.1016/S0377-0427(00)00393-9.

- Grace Wahba. *Spline Models for Observational Data*. SIAM, September 1990. ISBN 978-0-89871-244-5.
- Quanxin Wang and Andreas Burkhalter. Area map of mouse visual cortex. *The Journal of Comparative Neurology*, 502(3):339–357, May 2007. ISSN 00219967, 10969861. doi: 10.1002/cne.21286.
- J. G. White, E. Southgate, J. N. Thomson, and S. Brenner. The Structure of the Nervous System of the Nematode *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 314(1165):1–340, November 1986. ISSN 0080-4622. doi: 10.1098/rstb.1986.0056.

INTERLUDE

In the previous Chapter, we introduced a statistical method of to infer the weight matrix of a mesoscopic network. It is essentially a data analysis method, important when the data are incomplete and possibly inconsistent. However, such a spatial regression is inherently dependent on having a number of experiments from which to stitch together the weight matrix. When attempting to model neural dynamics at the microscopic level, the network of synaptic connections is rarely recoverable to anatomists. Sometimes, the best guess is that network of neural connections is random but consistent with certain measureable statistics.

Since the methods of Chapter 2 can yield the mesoscopic connectivity, those weights can be thought of as priors for the connection probability in microscopic random graph model of neural connections. For examples, if W is an $M \times M$ matrix of voxel weights, we could model the microscopic network as M different communities, where the probability of connection between a node in community j and another in community i is a function of W_{ij} . On the other hand, if we zoom out from the micro to macro, the mesoscopic connectivity arises as a spatially-averaged limit of the microscopic connectivity. Zooming in or out on the network, one obtains different effective models at different scales.

In the next Chapter, we present a theoretical study of a type of two-community random graph. This is a simple random graph model with two subsets of vertices who exhibit different connection patterns. We study this as an interesting model in its own right, and find a number of applications in machine learning and computer science.

Chapter 3

RANDOM GRAPH MODELS OF COMMUNITIES: SPECTRAL GAP AND ITS APPLICATIONS

Gerandy Brito, Ioana Dumitriu, Kameron Decker Harris

3.1 Introduction

Random regular graphs, where each vertex has the same degree d , are the best well-known examples of *expanders*: graphs with high connectivity and which exhibit rapid mixing. Expanders are of particular interest in computer science, from sampling and complexity theory to design of error-correcting codes. For an extensive review of their applications, see [Hoory et al. \[2006\]](#). What makes random regular graphs particularly interesting expanders is the fact that they exhibit all three existing types of expansion properties: edge, vertex, and spectral.

The study of regular random graphs took off with the work of [Bender \[1974\]](#), [Bender and Canfield \[1978\]](#), [Bollobás \[1980\]](#), and slightly later [McKay \[1984\]](#) and [Wormald \[1981\]](#). Most often, their expanding properties are described in terms of the existence of the *spectral gap*, which we define below.

Let A be the adjacency matrix of a simple graph, where $A_{ij} = 1$ if i and j are connected and zero otherwise. Denote $\sigma(A) = \{\lambda_1 \geq \lambda_2 \geq \dots\}$ as its spectrum. For a random d -regular graph, $\lambda_1 = \max_i |\lambda_i| = d$, but the “second largest eigenvalue” $\eta = \max(|\lambda_2|, |\lambda_n|)$ is asymptotically almost surely of much smaller order, leading to a “spectral gap.”

The study of the second largest eigenvalue in regular graphs had a first breakthrough

in the Alon-Boppana bound [Alon, 1986], which states that the second largest eigenvalue

$$\eta \geq 2\sqrt{d-1} - \frac{c_d}{\log n}.$$

Graphs for which the Alon-Boppana bound is attained are called *Ramanujan*. Friedman [2003] proved the conjecture of Alon [1986] that almost all d -regular graphs have $\eta \leq 2\sqrt{d-1} + \epsilon$ for any $\epsilon > 0$ with high probability as the number of vertices goes to infinity. Recently, Bordenave et al. [2015] improved that to $\eta \leq 2\sqrt{d-1} + \epsilon_n$ for a sequence $\epsilon_n \rightarrow 0$ as n , the number of vertices, tends to infinity.

In this paper we prove the analog of Friedman and Bordenave's result for bipartite, biregular random graphs. These are graphs for which the vertex set partitions into two independent sets V_1 and V_2 , such that all edges occur between the sets. In addition, all vertices in set V_i have the same degree d_i .

Let $\mathcal{G}(n, m, d_1, d_2)$ be the uniform distribution of simple, bipartite, biregular random graphs. Any $G \sim \mathcal{G}(n, m, d_1, d_2)$ is sampled uniformly from the set of simple bipartite graphs with vertex set $V = V_1 \cup V_2$, with $|V_1| = n$, $|V_2| = m$ and where every vertex in V_i has degree d_i . Note that we must have $nd_1 = md_2 = |E|$. Without any loss of generality, we will assume $n \leq m$ when necessary. Sometimes we will write that G is a (d_1, d_2) -regular graph, when we want to explicitly state the degrees. Let X be the $n \times m$ matrix with entries $X_{ij} = 1$ if and only if there is an edge between vertices $i \in V_1$ and $j \in V_2$. Using the block form of the adjacency matrix

$$A = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix},$$

it is not hard to show that all eigenvalues of A occur in pairs λ and $-\lambda$, where $|\lambda|$ is a singular value of X , along with at least $|n - m|$ zero eigenvalues. For this reason, the second largest eigenvalue is $\lambda_2(A) = -\lambda_{n+m-1}$. Furthermore, just as it was the case with d -regular graphs, the leading (Perron) eigenvalue of A is always $\sqrt{d_1 d_2}$, matched to the left by $-\sqrt{d_1 d_2}$.

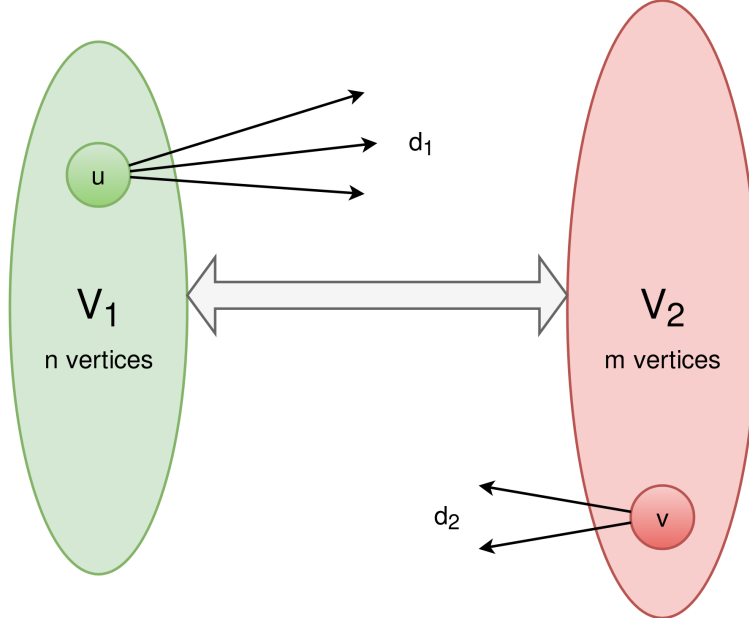


Figure 3.1: The structure of every bipartite, biregular graph. There are $n = |V_1|$ left vertices, $m = |V_2|$ right vertices, each of degree d_1 and d_2 , with the constraint that $nd_1 = md_2$. The distribution $\mathcal{G}(n, m, d_1, d_2)$ is taken uniformly over all such graphs.

Previous work on bipartite, biregular graphs includes the work of [Feng and Li \[1996\]](#) and [Li and Solé \[1996\]](#), who proved the analog of Alon-Boppana bound. For every $\epsilon > 0$,

$$\lambda_2 \geq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} - \epsilon$$

as the number of vertices goes to infinity. This bound also follows immediately from the fact that the second largest eigenvalue cannot be asymptotically smaller than the right limit of the asymptotic support for the eigenvalue distribution, which is $\sqrt{d_1 - 1} + \sqrt{d_2 - 1}$ and was first computed by [Godsil and Mohar \[1988\]](#).

[Marcus et al. \[2013b\]](#) showed that there exist infinite families of (d_1, d_2) -regular bipartite graphs with $\lambda_2 = \sqrt{d_1 - 1} + \sqrt{d_2 - 1}$ by taking repeated lifts of the complete bipartite graph on d_1 left and d_2 right vertices. We note that such graphs are also called Ramanujan. Complete graphs are always Ramanujan but not sparse, whereas d -regular or bipartite (d_1, d_2) -regular graphs are sparse. Our results show that almost every (d_1, d_2) -regular

graph is “almost” Ramanujan.

Beyond the first two eigenvalues, we should mention that [Bordenave and Lelarge \[2010\]](#) studied the limiting spectral distribution of large sparse graphs. They obtained a set of two coupled equations that can be solved for the eigenvalue distribution of any (d_1, d_2) -regular random graph. [\[Dumitriu and Johnson, 2016\]](#) showed that as $d_1, d_2 \rightarrow \infty$ with d_1/d_2 fixed, the limiting spectral distribution converges to a transformed version of the Marčenko-Pastur law. When $d_1 = d_2 = d$, this is equal to the Kesten-McKay distribution [\[McKay, 1981a\]](#), which becomes the semicircular law as $d \rightarrow \infty$ [\[Godsil and Mohar, 1988, Dumitriu and Johnson, 2016\]](#). Notably, [Mizuno and Sato \[2003\]](#) obtained the same results when they calculated the asymptotic distribution of eigenvalues for bipartite, biregular graphs of high girth. However, their results are not applicable to random bipartite biregular graphs as these asymptotically almost surely have low girth [\[Dumitriu and Johnson, 2016\]](#).

Our techniques borrow heavily from the results of [Bordenave \[2015\]](#), who simplified the trace method of [Friedman \[2003\]](#) by counting non-backtracking walks built up of segments with at most one cycle. See also the related work of [Bordenave et al. \[2015\]](#). The combinatorial methods we use to bound the number of such walks are similar to how [Brito et al. \[2015\]](#) counted self-avoiding walks in the context of community recovery in a regular stochastic block model.

Briefly, we now lay out the method of proof that the bipartite, biregular random graph is Ramanujan. The proof outline is given in detail in Section 3.5.1, after some important preliminary terms and definitions given in Section 3.4. The bulk of our work builds to Theorem 3, which is actually a bound on the second eigenvalue of the non-backtracking matrix B , as explained in Section 3.2. The Ramanujan bound on the second eigenvalue of A then follows as Corollary 4.

To find the second eigenvalue of B , we subtract from it a matrix S that is formed from the leading eigenvectors, and examine the spectral norm of the “almost-centered” matrix $\bar{B} = B - S$. We then proceed to use the trace method to bound the spectral norm of the

matrix $\bar{B}^\ell$ by its trace. However, since $\bar{B}$ is not positive definite, this leads us to consider

$$\mathbb{E} \left(\|\bar{B}^\ell\|^{2k} \right) \leq \mathbb{E} \left(\text{Tr} \left((\bar{B}^\ell)(\bar{B}^\ell)^* \right)^k \right) .$$

On the right hand side, the terms in $\bar{B}^\ell$ refer to circuits built up of $2k$ segments, each of length $\ell + 1$ (an entry B_{ef} is a walk on two edges). Because the degrees are bounded, it turns out that, for $\ell = O(\log(n))$, the depth ℓ neighborhoods of every vertex contain at most one cycle—they are “tangle-free.” Thus, we can bound the trace by computing the expectation of the circuits that contribute, along with an upper bound on their multiplicity, taking each segment to be ℓ -tangle-free.

Finally, to demonstrate the usefulness of the spectral gap, we highlight three applications of our bound. In Section 3.6, we show a community detection application. Finding communities in networks is important for the areas of social network, bioinformatics, neuroscience, among others. Random graphs offer tractable models to study when detection and recovery are possible.

We show here how our results lead to community detection in regular stochastic block models with arbitrary numbers of groups, using a very general theorem by Marina Meila (personal communication). Previously, [Newman and Martin \[2014\]](#) studied the spectral density of such models, and the community detection problem of the special case of two groups was previously studied by [Brito et al. \[2015\]](#) and [Barucca \[2017\]](#).

In Section 3.7, we examine the application to linear error correcting codes built from sparse expander graphs. This concept was first introduced by [Gallager \[1962\]](#) who explicitly used random bipartite biregular graphs. These “low density parity check” codes enjoyed a renaissance in the 1990s, when people realized they were well-suited to modern computers [for an overview, see [Richardson and Urbanke, 2003, 2008](#)]. Our result yields an explicit lower bound on the minimum distance of such codes, i.e. the number of errors that can be corrected.

The final application, in Section 3.8, leads to generalized error bounds for matrix completion. Matrix completion is the problem of reconstructing a matrix from observations of

a subset of entries. [Heiman et al. \[2014\]](#) gave an algorithm for reconstruction of a square matrix with low complexity as measured by a norm γ_2 , which is similar to the trace norm (sum of the singular values, also called the nuclear norm or Ky Fan n -norm). The entries which are observed are at the nonzero entries of the adjacency matrix of a bipartite, biregular graph. The error of the reconstruction is bounded above by a factor which is proportional to the ratio of the leading two eigenvalues, so that a graph with larger spectral gap has a smaller generalization error. We extend their results to rectangular graphs, along the way strengthening them by a constant factor of two. The main result of the paper gives an explicit bound in terms of d_1 and d_2 .

3.2 Non-backtracking matrix B

Given $G \sim \mathcal{G}(n, m, d_1, d_2)$, we define the non-backtracking operator B . This operator is a linear endomorphism of $\mathbb{R}^{|\vec{E}|}$, where $\vec{E}$ is the set of oriented edges of G and $|\vec{E}| = 2|E|$. Throughout this paper, we will use $V(H)$, $E(H)$, and $\vec{E}(H)$ to denote the vertices, edges, and oriented or directed edges of a graph, subgraph, or path H . For oriented edges $e = (u, v)$ (here u and v are the starting and ending vertices of e) and $f = (s, t)$, define:

$$B_{ef} = \begin{cases} 1, & \text{if } v = s \text{ and } u \neq t; \\ 0, & \text{otherwise.} \end{cases}$$

We order the elements of $\vec{E}$ as $\{e_1, e_2, \dots, e_{2|E|}\}$, so that the first $|E|$ have end point in the set V_2 . In this way, we can write

$$B = \begin{pmatrix} 0 & M \\ N & 0 \end{pmatrix}.$$

for $|E| \times |E|$ matrices M, N with entries equal to 0 or 1.

We are interested in the spectrum of B . Denote by $\mathbf{1}_\alpha$ the vector with first $|E|$ coordinates equal to 1 and the last $|E|$ equal to $\alpha = \sqrt{d_1 - 1} / \sqrt{d_2 - 1}$. We can check that

$$B\mathbf{1}_\alpha = B^*\mathbf{1}_\alpha = \lambda\mathbf{1}_\alpha$$

for $\lambda = \sqrt{(d_1 - 1)(d_2 - 1)}$. By the Perron-Frobenius Theorem, we conclude that $\lambda_1 = \lambda$ and the associated eigenspace has dimension one. Also, one can check that if λ is an eigenvalue of B with eigenvector $v = (v_1, v_2)$, $v_i \in \mathbb{R}^{|E|}$ then $-\lambda$ is also an eigenvalue with eigenvector $v' = (-v_1, v_2)$. Thus, $\sigma(B) = -\sigma(B)$ and $\lambda_{2|E|} = -\lambda_1$.

3.2.1 Connecting the spectra of A and B

Understanding the spectrum of B turns out to be a challenging question. A useful result in this direction is the following theorem proved by Bass [1992] and Kotani and Sunada [2000]; see also Theorem 3.3 in Angel et al. [2015].

Theorem 1 (Ihara-Bass formula). *Let $G = (V, E)$ be any finite graph and B be its non-backtracking matrix. Then*

$$\det(B - \lambda I) = (\lambda^2 - 1)^{|E| - |V|} \det(D - \lambda A + \lambda^2 I),$$

where D is the diagonal matrix with $D_{vv} = d_v - 1$ and A is the adjacency matrix of G .

From the theorem above we get the following relation between $\sigma(A)$ and $\sigma(B)$

$$\sigma(B) = \{\pm 1\} \cup \{\lambda : D - \lambda A + \lambda^2 I \text{ is not invertible}\}.$$

We use the special structure of G to get a more precise description of $\sigma(B)$. The matrices A and D are equal to:

$$A = \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix}, \quad D = \begin{pmatrix} (d_1 - 1)I_n & 0 \\ 0 & (d_2 - 1)I_m \end{pmatrix},$$

where I_k is the $k \times k$ identity matrix. Let $\lambda \in \sigma(B) \setminus \{-1, 0, 1\}$. Then there exists a nonzero vector v such that

$$(D - \lambda A + \lambda^2 I)v = 0.$$

Writing $v = (v_1, v_2)$ with $v_1 \in \mathbb{R}^n$, $v_2 \in \mathbb{R}^m$, we obtain:

$$Xv_2 = \frac{d_1 - 1 + \lambda^2}{\lambda} v_1, \quad X^*v_1 = \frac{d_2 - 1 + \lambda^2}{\lambda} v_2$$

which imply that

$$\eta^2 = \frac{(d_1 - 1 + \lambda^2)(d_2 - 1 + \lambda^2)}{\lambda^2}$$

is a non zero eigenvalue of both XX^* and X^*X . The above relation gives us the following claim, which relates the spectrum of B and the spectrum of X^*X :

Claim 2. Any $\lambda \in \sigma(B) \setminus \{-1, 0, 1\}$ satisfies:

$$\lambda^4 - (\eta^2 - d_1 - d_2 + 2)\lambda^2 + (d_1 - 1)(d_2 - 1) = 0$$

where η^2 is an nonzero eigenvalue of X^*X or, equivalently, $-\eta$ and η are eigenvalues of A .

3.3 Main result

We spend the bulk of this paper in the proof of the following:

Theorem 3. If B is the non-backtracking matrix of a bipartite, biregular random graph $G \sim \mathcal{G}(n, m, d_1, d_2)$, then its second largest eigenvalue

$$\lambda_2(B) \leq ((d_1 - 1)(d_2 - 1))^{1/4} + \epsilon_n$$

asymptotically almost surely, with $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

Corollary 4 (Spectral gap). If A is the adjacency matrix of a bipartite, biregular random graph $G \sim \mathcal{G}(n, m, d_1, d_2)$, then its second largest eigenvalue

$$\lambda_2(A) \leq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} + \epsilon_n$$

asymptotically almost surely, with $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

Proof. From Claim 2 we obtain that

$$0 = \lambda^4 + (d_1 - 1)(d_2 - 1) - (\eta^2 - d_1 - d_2 + 2)\lambda^2.$$

From Theorem 3, we have that

$$0 \leq 2(d_1 - 1)(d_2 - 1) - (\eta^2 - d_1 - d_2 + 2)(d_1 - 1)^{1/2}(d_2 - 1)^{1/2} + \epsilon.$$

Dividing through by $(d_1 - 1)^{1/2}(d_2 - 1)^{1/2}$ and moving η^2 to the left hand side, we get that

$$\eta^2 \leq (d_1 - 1) + (d_2 - 1) + 2(d_1 - 1)^{1/2}(d_2 - 1)^{1/2} + \epsilon = (\sqrt{d_1 - 1} + \sqrt{d_2 - 1})^2 + \epsilon,$$

and the result follows. $\square$

3.4 Preliminaries

We describe the standard configuration model for constructing such graphs. We then define the “tangle-free” property of random graphs. Since almost all small enough neighborhoods are tangle-free, we only need to count tangle-free paths when we eventually employ the trace method.

3.4.1 The configuration model

The configuration or permutation model is a practical procedure to sample random graphs with a given degree distribution. Let us recall its definition for bipartite biregular graphs. Let $V_1 = \{v_1, v_2, \dots, v_n\}$ and $V_2 = \{w_1, w_2, \dots, w_m\}$ be the vertices of the graph. We define the set of *half edges* out of V_1 to be the collection of ordered pairs

$$E_1 = \{(v_i, j) \text{ for } 1 \leq i \leq n \text{ and } 1 \leq j \leq d_1\}$$

and analogously the set of half edges out of V_2 :

$$E_2 = \{(w_i, j) \text{ for } 1 \leq i \leq m \text{ and } 1 \leq j \leq d_2\},$$

see Figure 3.1. Note that $|E_1| = |E_2| = nd_1 = md_2$. To sample a graph, we choose a random permutation π of $[nd_1]$. We put an edge between v_i and w_j whenever

$$\pi((i - 1)d_1 + s) = (j - 1)d_2 + t$$

for any pair of values $1 \leq s \leq d_1, 1 \leq t \leq d_2$. The graph obtained may not be simple, since multiple half edges may be matched between any pair of vertices. However, conditioning

on a simple graph outcome, the distribution is uniform in the set of all simple bipartite biregular graphs. Furthermore, for fixed d_1, d_2 and $n, m \rightarrow \infty$, the probability of getting a simple graph is bounded away from zero [Bollobás, 2001].

It is often useful to sample the edges one at the time. We call this the *exploration process*. More precisely, we order the set E_1 lexicographically: $(v_i, j) < (v_{i'}, j')$ if $i \leq i'$ and $j \leq j'$. The exploration process reveals π by doing the following:

- A uniform element is chosen from E_2 and it is declared equal to $\pi(1)$.
- A second element is chosen uniformly, now from the set $E_2 \setminus \{\pi(1)\}$ and set equal to $\pi(2)$.
- Once we have determined $\pi(i)$ for $i \leq k$, we set $\pi(k+1)$ equal to a uniform element sampled from the set $E_2 \setminus \{\pi(1), \pi(2), \dots, \pi(k)\}$.

We use the final π to output a graph as we did in the configuration model. The law of these graphs is the same. With the exploration process, we expose first the neighbors of v_1 , then the neighbors of v_2 , etc. This feature will be quite useful in the next subsection.

3.4.2 Tangle-free paths

Sparse random graphs have the important property of being “tree-like” in the neighborhood of a typical vertex. This is the case, also, for bipartite biregular graphs. Formally, consider a vertex $v \in V_1 \cup V_2$. For a natural number ℓ , we define the ball of radius ℓ centered at v to be:

$$B_\ell(v) = \{w \in V_1 \cup V_2 : d_G(v, w) \leq \ell\}$$

where $d_G(\cdot, \cdot)$ denotes the graph distance.

Definition 1. A graph G is ℓ -tangle-free if, for any vertex v , the ball $B_\ell(v)$ contains at most one cycle.

The next lemma says that most bipartite biregular graphs are ℓ -tangle-free up to logarithmic sized neighborhoods.

Lemma 5. *Fix a constant $c < 1/8$ and let $\ell = c \log_d(n)$. Let $G \sim \mathcal{G}(n, m, d_1, d_2)$ be a bipartite, biregular random graph. Then G is ℓ -tangle-free with probability at least $1 - n^{-1/2}$.*

Proof. This is essentially the proof given in [Lubetzky and Sly \[2010\]](#), Lemma 2.1. Fix a vertex v . We will use the exploration process to discover the ball $B_\ell(v)$. To do so, we first explore the neighbors of v , then the neighbors of these vertices, and so on. This breadth-first search reveals all vertices in $B_k(v)$ before any vertices in $B_{j>k}(v)$. Note that, although our bound is for the family $\mathcal{G}(n, m, d_1, d_2)$, the neighborhood sizes are bounded above by those of the d -regular graph with $d = \max(d_1, d_2)$.

Consider the matching of half edges attached to vertices in the ball $B_i(v)$ at depth i (thus revealing vertices at depth $i + 1$). In this process, we match a maximum $m_i \leq d^{i+1}$ pairs of half edges total. Let $\mathcal{F}_{i,k}$ be the filtration generated by matching up to the k th half edge in $B_i(v)$, for $1 \leq k \leq m_i$. Denote by $A_{i,k}$ the event that the k th matching creates a cycle at the current depth. For this to happen, the matched vertex must have appeared among the $k - 1$ vertices already revealed at depth $i + 1$. The number of unmatched half edges is at least $nd - 2d^{i+1}$. We then have that:

$$\mathbb{P}(A_{i,k}) \leq \frac{(k-1)(d-1)}{nd - 2d^{i+1}} \leq \frac{(d-1)m_i}{(1 - 2d^{i+1}n^{-1})nd} \leq \frac{m_i}{n}.$$

So, we can stochastically dominate the sum

$$\sum_{i=1}^{\ell-1} \sum_{k=1}^{m_i} A_{i,k}$$

by $Z \sim \text{Bin}(d^{\ell+1}, n^{-1}d^\ell)$. So the probability that $B_\ell(v)$ is ℓ -tangle-free has the bound:

$$\mathbb{P}(B_\ell(v) \text{ is not } \ell\text{-tangle-free}) = \mathbb{P}\left(\sum_{i=1}^{\ell-1} \sum_{k=1}^{m_i} A_{i,k} > 1\right) \leq \mathbb{P}(Z > 1) = O\left(\frac{d^{4\ell+1}}{n^2}\right) = O\left(n^{-3/2}\right),$$

which follows using that $\ell = c \log_d n$ with $c < 1/8$. The Lemma follows by taking a union bound over all vertices. $\square$

3.5 Proof of Theorem 3

3.5.1 Outline

We are now prepared to explain the main result. To study the second largest eigenvalue of the non-backtracking matrix, we examine the spectral radius of the matrix obtained by subtracting off the dominant eigenspace. We use Lemma 3 in [Bordenave \[2015\]](#) for this:

Lemma 6. *Let T, R be matrices such that $\text{Im}(T) \subset \text{Ker}(R)$, $\text{Im}(T^*) \subset \text{Ker}(R)$. Then all eigenvalues λ of $T + R$ that are not eigenvalues of T satisfy:*

$$|\lambda| \leq \max_{x \in \text{Ker}(T)} \frac{\|(T + R)x\|}{\|x\|}.$$

In the above theorem and throughout the text, $\|\cdot\|$ is the spectral norm for matrices and ℓ^2 -norm for vectors. Recall that the leading eigenvalues of B , in magnitude, are $\lambda_1 = \sqrt{(d_1 - 1)(d_2 - 1)}$ and $\lambda_{2|E|} = -\lambda_1$ with corresponding eigenvectors $\mathbf{1}$ and $\mathbf{1}_{-\alpha}$. Applying the lemma above with $T = \lambda_1^\ell S$ and $R = B^\ell - T$, we get that

$$\lambda_2(B) \leq \max_{\substack{x \in \text{Ker}(T) \\ \|x\| = 1}} \left(\|B^\ell x\| \right)^{1/\ell}, \quad (3.1)$$

where $S = \mathbf{1}_\alpha \mathbf{1}_\alpha^* - \mathbf{1}_{-\alpha} \mathbf{1}_{-\alpha}^*$. It will be important later to have a more precise description of the set $\text{Ker}(T)$. It is not hard to check that

$$\begin{aligned} \text{Ker}(T) &= \{x : \langle x, \mathbf{1}_\alpha \rangle = \langle x, \mathbf{1}_{-\alpha} \rangle = 0\} \\ &= \{(v, w) \in \mathbb{R}^{2|E|} : \langle v, \mathbf{1} \rangle = \langle w, \mathbf{1} \rangle = 0\}. \end{aligned}$$

Above, the vectors v, w and $\mathbf{1}$ are $|E|$ -dimensional, and $\mathbf{1}$ is the vector of all ones.

In order to use Eqn. 3.1, we must bound $\|B^\ell x\|$ for large powers ℓ and $x \in \text{Ker}(T)$. This amounts to counting certain non-backtracking walks. We will use the tangle free property in order to only count ℓ -tangle-free walks. We break up B^ℓ into two parts in Section 3.5.2, an “almost” centered matrix $\bar{B}^\ell$ and the residual $\sum_j R^{\ell,j}$, and we bound each term independently.

To compute these bounds, we need to count the contributions of many different non-backtracking walks. We will use the trace technique, so only circuits which return to the starting vertex will contribute. In Section 3.5.3, we apply a useful result from McKay [1981b] to compute the probability, during the exploration process, of revealing a new edge e given that we have already observed a certain subgraph H . In particular, we find different probabilities depending on whether e shares one of more endpoints in H . We use this to bound the expectation of the product of entries $\bar{B}_{ef}$ along segments ef of a non-backtracking walk. A similar argument appears later, in the proof of Theorem 15, for products of $R_{ef}^{\ell,j}$.

In Section 3.5.4 we cover the combinatorial component of the proof. The total contributions $\|B^\ell x\|$ come from many non-backtracking circuits of different flavors, depending on their number of vertices, edges, cycles, etc. Each circuit is broken up into $2k$ segments of tangle-free walks of length ℓ . We need to compute not only the expectation along the circuit, but also upper-bound the number of circuits of each flavor. we introduce an injective encoding of such circuits that depends on the number of vertices, length of the circuit, and, crucially, the tree excess of the circuit.

Finally, in Section 3.5.5 we put all of these ingredients together and use Markov's inequality to bound each matrix norm with high probability. We find that $\|\bar{B}^\ell\|$ contributes a factor that goes as $((d_1 - 1)(d_2 - 1))^{\ell/4}$, whereas $\|R^{\ell,j}\|$ contributes only a factor of ℓ (up to logarithmic factors in n). Thus, the main contribution to the circuit counts comes from the mean, and, in fact, comes from circuits which are exactly trees traversed forwards and backwards.

Interestingly, that the dominant contributions arise from trees is analogous to what happens when using the trace method on random matrices of independent entries. Our expectation bounds (Section 3.5.3) essentially show that the model $\mathcal{G}(n, m, d_1, d_2)$ adds edges close to independently when exploring small enough neighborhoods. And the combinatorial arguments of Section 3.5.4 show that there are not enough contributions from paths with cycles to compensate for this.

In doing so, however, we are forced to consider tangled paths but which are built up of tangle-free components. This delicate issue was first made clear by [Friedman \[2004\]](#) who introduced the idea of tangles and a “selective trace.” [Bordenave \[2015\]](#), who we follow closely in this part of our analysis, also has a good discussion of these issues and their history. We use the fact that

$$\mathbb{E} \left(\|\bar{B}^\ell\|^{2k} \right) \leq \mathbb{E} \left(\text{Tr} \left((\bar{B}^\ell)(\bar{B}^\ell)^* \right)^k \right), \quad (3.2)$$

and so deal with circuits built up of $2k$ segments which are ℓ -tangle-free. Notice that the first segment comes from $\bar{B}^\ell$, the second from $(\bar{B}^\ell)^*$, etc. Because of this, the directionality of the edges along each segment alternates. See [Figure 3.2](#) for an illustration of a path which contributes for $k = 2$ and $\ell = 2$. Also, while each segment γ_i is ℓ -tangle-free, the overall circuit may be tangled, since later segments can revisit vertices seen before.

3.5.2 Matrix decomposition

For this section, we will assume G is ℓ -tangle-free, which will hold with high probability. Let Γ_{ef}^ℓ be the set of all non-backtracking paths in G of length $\ell + 1$, starting at oriented edge e and ending at f . For a path $\gamma \in \Gamma_{ef}^\ell$, we write $\gamma = (e_1, e_2, \dots, e_{\ell+1})$ where $e_i \in \vec{E}$ for all i , $e_1 = e$ and $e_{\ell+1} = f$. Similarly, define $F_{ef}^\ell \subset \Gamma_{ef}^\ell$ be the set of all non-backtracking, tangle-free paths in G of length $\ell + 1$, starting at oriented edge e and ending at f . Then,

$$(B^\ell)_{ef} = \sum_{\gamma \in \Gamma_{ef}^\ell} \prod_{t=1}^{\ell} B_{e_t e_{t+1}} = \sum_{\gamma \in F_{ef}^\ell} \prod_{t=1}^{\ell} B_{e_t e_{t+1}},$$

where we note the last equality requires G to be ℓ -tangle-free. Denote by $\bar{B}$ the matrix with entries equal to

$$(\bar{B}^\ell)_{ef} = \sum_{\gamma \in F_{ef}^\ell} \prod_{t=1}^{\ell} (B - S)_{e_t e_{t+1}},$$

where

$$S = \begin{pmatrix} 0 & \frac{d_2-1}{n} \mathbf{1}\mathbf{1}^* \\ \frac{d_1-1}{m} \mathbf{1}\mathbf{1}^* & 0 \end{pmatrix}.$$

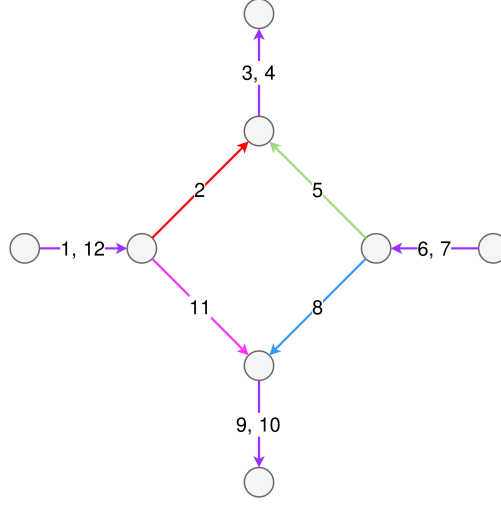


Figure 3.2: An example circuit that contributes to the trace in Eqn. (3.2), for $k = 2$ and $\ell = 2$. Edges are numbered as they occur in the circuit. Each segment $\{\gamma_i\}_{i=1}^4$ is of length $\ell + 1 = 3$ and made up of edges $3(i - 1) + 1$ through $3i$. The last edge of each γ_i is the first edge of γ_{i+1} , and these are shown in purple. Every path γ_i with i even follows the edges backwards due to the matrix transpose. However, this detail turns out not to make any difference since the underlying graph is undirected. Our example has no cycles in each segment for clarity, but, in general, each segment can have up to one cycle, and the overall circuit may be tangled.

Note that $\bar{B}$ is an *almost* centered version of B , and $\text{Ker}(S) = \text{Ker}(T) = \text{span}(\mathbf{1}_\alpha, \mathbf{1}_{-\alpha})$.

The following telescoping sum formula appears in Bordenave [2015]:

$$\prod_{s=1}^{\ell} x_s = \prod_{s=1}^{\ell} y_s + \sum_{j=1}^{\ell} \prod_{s=1}^{j-1} y_s (x_j - y_j) \prod_{t=j+1}^{\ell} x_t.$$

Using this, with $x_s = B_{e_s e_{s+1}}$ and $y_s = \bar{B}_{e_s e_{s+1}}$, we obtain the following relation:

$$(B^\ell)_{ef} = (\bar{B}^\ell)_{ef} + \sum_{\gamma \in F_{ef}^\ell} \sum_{j=1}^{\ell} \prod_{i=1}^{j-1} \bar{B}_{e_i e_{i+1}} S_{e_j e_{j+1}} \prod_{t=j+1}^{\ell} B_{e_t e_{t+1}}. \quad (3.3)$$

This decomposition breaks the elements in F_{ef}^ℓ into two subpaths, also non-backtracking and tangle-free, of length j and $\ell - j$, respectively. To recover the matrices B and $\bar{B}$ by rearranging Eqn. (3.3), we need to also count those tangle-free subpaths that arise from

tangled paths. While breaking a tangle-free path will necessarily give us two new tangle-free subpaths, the converse is not always true. This extra term generates a remainder that we define now.

Let $T_{ef}^{\ell,j} \subset \Gamma_{ef}^\ell$ be the set of non-backtracking paths in $K_{n,m}$ (the complete bipartite graph on n left and m right vertices) of length $\ell + 1$, starting at e and ending at f , such that overall the path is tangled but the first j and last $\ell - j$ edges form tangle-free subpaths of G . Set the remainder

$$R_{ef}^{\ell,j} = \sum_{\gamma \in T_{ef}^\ell} \prod_{i=1}^{j-1} \bar{B}_{e_i e_{i+1}} S_{e_j e_{j+1}} \prod_{i=j+1}^{\ell} B_{e_i e_{i+1}}. \quad (3.4)$$

Adding and subtracting $\sum_{j=1}^{\ell} R_{ef}^{\ell,j}$ to Eqn. (3.3) and rearranging the sums, we obtain

$$B^\ell = \bar{B}^\ell + \sum_{j=1}^{\ell} \bar{B}^j S B^{\ell-j} - \sum_{k=1}^{\ell} R^{\ell,k}. \quad (3.5)$$

Multiplying Eqn. (3.5) on the right by $x \in \text{Ker}(T)$ and using that $B^{\ell-j}x$ is also within $\text{Ker}(T)$, since it is just the space spanned by the leading eigenvectors, we find that the middle term is identically zero. Thus,

$$\|B^\ell x\| \leq \|\bar{B}^\ell x\| + \left\| \sum_{k=1}^{\ell} R^{\ell,k} x \right\| \leq \|\bar{B}^\ell\| + \sum_{k=1}^{\ell} \|R^{\ell,k}\|. \quad (3.6)$$

3.5.3 Expectation bounds

Our goal is to find a bound on the expectation of certain random variables which are products of $\bar{B}_{ef}$ along a cycle. To do this, we will need to bound the probabilities of different subgraphs when exploring G .

The next Lemma follows from McKay [1981b], Theorem 3.5. We use the following notation: Let H be a subgraph of G with vertex set $\{v_1, v_2, \dots, v_k\}$. Let d_i be the degree of v_i in G , so $d_i = d_1$ if v_i is in the set V_1 and d_2 if not. Also, denote by h_i the degree of v_i in H . For natural numbers x and t , we use

$$(x)_t = x(x-1)(x-2) \cdots (x-t+1)$$

to denote the falling factorial.

Lemma 7. *Let $H \subset K_{n,m}$ such that $|E(H)| = o(n)$ and $G \sim \mathcal{G}(n, m, d_1, d_2)$. Then*

$$\mathbb{P}(H \subset G) \leq \frac{\prod_{i=1}^k (d_i)_{h_i}}{(nd_1 - 4d^2)^{|E(H)|}},$$

and

$$\mathbb{P}(H \subset G) \geq \frac{\prod_{i=1}^k (d_i)_{h_i}}{(nd_1 - 1)^{|E(H)|}} \left(\frac{nd_1 - |E(H)| - 5d^2}{nd_1 - c|E(H)| - 5d^2} \right)^{|E(H)|},$$

for some explicit constant $c < 1$, where $d = \max(d_1, d_2)$.

Crucially, we will use Lemma 7 to show that the appearance of edges in random bipartite biregular graphs are weakly correlated, as long as the number of edges is not too big. Computations are similar to those carried out in [Brito et al., 2016].

Lemma 8. *Let $G \sim G(n, m, d_1, d_2)$ and let $H \subset K_{n,m}$ such that $|E(H)| = o(n)$. Let e be an edge not in H , such that H and $H \cup \{e\}$ have the same number of connected components, and let $n_1 = m, n_2 = n$.*

(i) *If e and H share an endpoint of degree d_i in G , it holds that*

$$\mathbb{P}(e \in G | H \subset G) \leq \frac{d_i - 1}{n_i} + O\left(\frac{|E(H)|}{n^2}\right).$$

(ii) *If e has exactly one endpoint in H , and this vertex has degree one in H and degree d_i in G , then*

$$\mathbb{P}(e \in G | H \subset G) = \frac{d_i - 1}{n_i} + O\left(\frac{|E(H)|}{n^2}\right).$$

(iii) *If e has endpoints in both V_1 and V_2 , then we can use the bound in (i) with $d_i = d_1$ or d_2 .*

Proof. Let $e = (u, v)$, where v is the shared endpoint in H . Assume v is degree d_2 in G ; the results for d_1 are analogous. Vertex $u \notin V(H)$ must then have degree d_1 in G . For part (i),

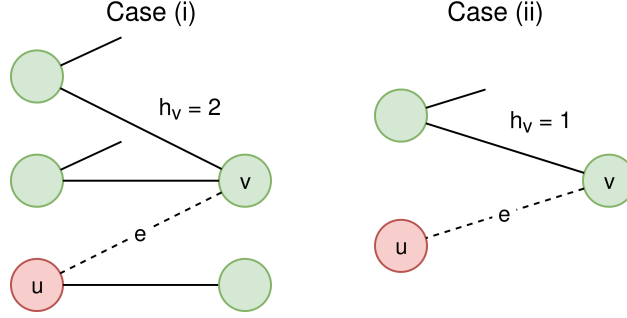


Figure 3.3: The different cases of Lemma 8, which concerns $P(e \in G|H)$: the probability of an edge e existing in the random graph G , given an observed subgraph H , shown in green. The new edge e connects at vertex v . All the green vertices lie in H . When the connecting vertex has degree one in subgraph H , i.e. $h_v = 1$ and case (ii), it induces a two-path. The probability of that edge is closer to the entries of the centering matrix S if $h_v > 1$, case (i).

notice that the graph $H \cup \{e\}$ has at least one vertex v with degree equal to $h_v + 1$. Using the upper bound in Lemma 7, we get

$$\begin{aligned} \mathbb{P}(H \cup \{e\} \subset G) &\leq \frac{d_1(d_2 - h_v)}{nd_1 - 4d^2 - |E(H)|} \mathbb{P}(H \subset G) \\ &\leq \frac{d_2 - 1}{n} \left(1 + O\left(\frac{|E(H)|}{n}\right) \right) \mathbb{P}(H \subset G). \end{aligned}$$

For part (ii), note that $h_v = 1$ and employ the lower bound in Lemma 7 to get equality. $\square$

Remark. Assuming $m \geq n$, and using $\frac{d_1-1}{m} = \frac{d_2-1}{n} + \frac{1}{n} - \frac{1}{m}$, we see that the bound with d_1 is weaker by a factor of $O(\frac{1}{n})$. Finally, $\mathbb{P}(e \in G|H) \leq \frac{d_1}{m}$ for any isolated edge.

Remark (Probability of tree edges). The centering matrix S has entries $(d_2 - 1)/n$ and $(d_1 - 1)/m$. During the exploration process, Lemma 8 states that the probability of adding a new edge differs from the entries of the centering matrix (which are close to but not exactly the expectation) by an order $1/n^2$ correction precisely when adding that edge creates a two-path. See Figure 3.3.

Let $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{2k})$ be a circuit obtained by the concatenation of $2k$ non-backtracking walks of length ℓ . These circuits will appear when we apply the trace method in Section 3.5.5. The circuit γ is traversed in this order: We start at the initial vertex of γ_1 , move

along this path until we meet γ_2 , continue along this path, etc. Denote by $\vec{E}(\gamma)$ the set of oriented edges traversed by γ , and $E(\gamma)$ the same set of edges without orientation. A subpath of γ is just an ordered path of edges traversed as described above. Define

$$X_\gamma = \prod_{ef \in \gamma} (\bar{B}_{ef})^{m_{ef}}, \quad (3.7)$$

where $e, f \in \vec{E}(\gamma)$, and $ef \in \gamma$ means that the oriented path ef is a subpath of γ when traversed as described above, and m_{ef} is the total number of times we traverse ef in γ .

The main result of this section is the following Theorem:

Theorem 9. *Let $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{2k})$ be a circuit obtained by the concatenation of $2k$ non-backtracking walks of length ℓ . If γ visits $K = |E(\gamma)| = o(n)$ different edges and $|V(\gamma) \cap V_2| = r$, then*

$$\mathbb{E}(X_\gamma) \leq \frac{D_\gamma}{m^r n^{K-r}} \left(\frac{K}{n} \right)^\omega (1 + o(1)), \quad (3.8)$$

where $d = \max(d_1, d_2)$,

$$\omega \geq \left\lceil \frac{\sum_{ef \in \gamma} \mathbf{1}\{m_{ef} + m_{f^{-1}e^{-1}} = 1\}}{6d^2} \right\rceil,$$

and $D_\gamma = (d_1 - 1)^r (d_2 - 1)^{K-r}$ if $\omega = 0$, and $D_\gamma = d^K$ otherwise.

The proof of Theorem 9 proceeds like this: We pick an ordering of the set of *undirected* edges visited by γ . With this ordering, we define a filtration, or sequence of nested sigma algebras, $\{\mathcal{F}_t\}_{t=1}^K$ each containing the information of the first t edges in this order. We use the tower of expectation to bound the right hand side of Eqn. (3.8). At each step, a new edge is removed from the filtration and we are able to improve our current bound via Lemma 8. The ordering of the edges is done in a way that allows us to use part (ii) of Lemma 8 a maximal number of times, which will add ω extra factors of $1/n$. These extra factors come from so-called “good” edges (explained shortly) in two-paths which are traversed exactly once.

We start by describing the ordering of the edges. Let $E = \{e_i\}_{i=1}^K$ be a set of undirected edges. A permutation π of $[K]$ can be identified with an ordering in E by taking the first edge to be $e_{\pi(1)}$, the second to be $e_{\pi(2)}$, etc. For a subset $F \subseteq E$ of edges, define

$$N(F) = \{e \in E : e \text{ shares a vertex with some } f \in F\}$$

of *neighbors* of F . Notice that the orientation is not relevant in this definition.

Given an ordering π , we say that the edge $e_{\pi(j)}$ is a *good edge* if there is exactly one value of $i \leq j - 1$ such that $e_{\pi(i)} \in N(e_{\pi(j)})$. In other words, the j th edge is good if it has exactly one neighbor among the previous edges. Thus, when we add a good edge to the graph induced by the previous edges, it must induce a two-path. The following Lemma tells us how many good edges we can guarantee with our ordering π , and gives a recipe to construct that ordering.

Lemma 10. *Let F be some set of undirected edges in a graph with maximal degree d , with $p_2(F)$ equal to the number of two-paths in F . Then there exists an ordering π of the edges with at least $\left\lceil \frac{p_2(F)}{6d^2} \right\rceil$ good edges.*

Proof. We now construct such an ordering by working with three sets of edges E^1 , E^2 , and E^3 . At time $t = 0$, start with the sets $E_0^1 = \emptyset$, $E_0^2 = \emptyset$ and $E_0^3 = F$. For times $t \geq 1$, we build the ordering π iteratively:

1. Choose $e_i, e_j \in E_{t-1}^3$ such that $e_i \in N(e_j)$, i.e. $e_i e_j$ is a two-path.
2. Set $\pi(2t - 1) = i$ and $\pi(2t) = j$.
3. Set $E_t^1 = E_{t-1}^1 \cup \{e_i, e_j\}$, $E_t^2 = E_{t-1}^2 \cup N(\{e_i, e_j\}) \setminus E_t^1$ and $E_t^3 = F \setminus (E_t^1 \cup E_t^2)$.

At all times, E_t^1 , E_t^2 and E_t^3 form a disjoint partition of F . They correspond to: the set of edges already ordered E^1 , the set of the neighbors of the edges already ordered E^2 , and the complement of those two sets E^3 . Also, it is not hard to check that e_{2t} is a good edge for all t . This process will end at step $t = T$ when one of the following exclusive events happen:

- (a) The set $E_T^3 = \emptyset$ and we defined $\pi(t)$ for all $1 \leq t \leq 2T$. We now arbitrarily set $\pi(t)$ for times $2T + 1 \leq t \leq |F|$ using the edges in E_T^2 . By construction, we have at least T good edges. Since at each step we remove from E^3 at most $3d$ edges, and each edge participates in less than $2d$ two-paths, we conclude that

$$T \geq \left\lceil \frac{p_2(F)}{6d^2} \right\rceil.$$

- (b) After T steps, no two edges in E_T^3 form a two-path. By construction, E_T^1 and E_T^3 are disconnected. We use this to continue the ordering for times $2T + 1 \leq t \leq |F|$: Because γ is a connected walk, for each $e_i \in E_t^3$ there exists some $e_j \in E_t^2$ such that $e_i e_j$ or $e_j e_i$ is a subpath of γ . Set $\pi(2t + 1) = j$ and $\pi(2t + 2) = i$. With this choice, e_i is good. We update E_{t+1}^1 , E_{t+1}^2 , and E_{t+1}^3 as before. The process is repeated until time $t = T^*$, when $E_{T^*}^3 = \emptyset$. The number of good edges is again at least

$$T^* \geq \left\lceil \frac{p_2(F)}{6d^2} \right\rceil.$$

□

Remark. The above Lemma gives an algorithm for ordering some set of edges $F \subseteq E$ in a way that guarantees good edges. Suppose this process finishes at time T . The edges in E_T^1 are added in a way that they form a disjoint set of two-paths, and half of these edges are good. We note that we are free to add the edges left in $E \setminus E_T^1$ however we would like.

We will also need the following Lemma:

Lemma 11 (Brito et al. [2015], Lemma 11). Let $X \sim \text{Bernoulli}(q)$ with $q \leq p + r$, where $0 \leq q, p \leq 1$. Then for any integer $m > 1$, the expectation $\mathbb{E}((X - p)^m) \leq p + r$.

Proof. Assume $q < p$. Then,

$$\mathbb{E}(|X - p|^m) \leq (1 - p)^m p + p^m \leq p.$$

The latter inequality follows easily by noting that it is satisfied for $m = 2$, and that $(1 - p)^m p + p^m$ is a decreasing function of m for all $0 \leq p \leq 1$.

If $q > p$, write $q = p + r'$ with $0 < r' < r$. We get:

$$\mathbb{E}(|X - p|^m) \leq (1 - p)^m(p + r') + p^m \leq (1 - p)^m p + p^m + r' \leq p + r,$$

due to similar considerations. $\square$

Proof of Theorem 9. Let $E(\gamma) = \{e_1, e_2, \dots, e_K\}$ be the set of undirected edges visited by γ , ordered by some permutation π of $K = |E(\gamma)|$ elements, and call H_t the graph induced by the set of undirected edges $\{e_1, \dots, e_t\}$. To get an adequate upper bound, we order the edges so that we have a maximal number of good edges with two-path multiplicity one. Let F be the set of undirected edges whose oriented counterparts, in either direction, participate in a two-path which occurs only once in γ . Then,

$$F = \{e \in E(\gamma) : m_{ef} + m_{f^{-1}e^{-1}} = 1 \text{ for some two-path } ef \in \gamma\}.$$

Let $p_2(F)$ denote the number of two-paths in F , then $p_2(F) \geq \sum_{ef \in \gamma} \mathbf{1}\{m_{ef} + m_{f^{-1}e^{-1}} = 1\}$. First, we order the edges in F following the algorithm in Lemma 10, producing an ordering $\tilde{\pi}$. The actual ordering π is set by using $\pi(i) = \tilde{\pi}(i)$ for $1 \leq i \leq |F|$ and arbitrarily extending π to the rest of the edges $E(\gamma) \setminus F$.

Let $\vec{A}_t \subset \vec{E}(\gamma \cup \gamma^{-1})$ be the set of *oriented* edges

$$\vec{A}_t = \{\vec{e}_1, (\vec{e}_1)^{-1}, \dots, \vec{e}_t, (\vec{e}_t)^{-1}\}$$

containing the first t edges with both possible orientations. For simplicity and without any loss of generality, we make the convention that $\vec{e}_t$ goes from set V_1 to V_2 . We use $(\vec{e})^{-1}$ to represent the directed edge with reversed orientation of $\vec{e}$. Define the $\mathcal{F}_t$ as the sigma algebra generated by $\vec{A}_t$, for $t = 1, 2, \dots, K$. Then $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}_K$, so $\{\mathcal{F}_t\}$ is a filtration.

Recall the definition of X_γ from Eqn. (3.7). We define the random variables $\{X_t\}_{2 \leq t \leq K}$ as

$$X_t = \prod_{ef} (\bar{B}_{ef})^{m_{ef}},$$

where the product is over all $e, f \in \vec{A}_t$, where ef is a subpath of γ , and exactly one of $\{e, f\}$ is in the set $\vec{A}_t \setminus \vec{A}_{t-1} = \{\vec{e}_t, (\vec{e}_t)^{-1}\}$. Thus, X_t is the product of all factors in X_γ which involve e_t and earlier edges according to π . Note that the following definition is equivalent:

$$X_t = \prod_{e, f_0, f_1, f_2, f_3} (\bar{B}_{ef_0})^{m_{ef_0}} (\bar{B}_{e^{-1}f_1})^{m_{e^{-1}f_1}} (\bar{B}_{f_2e})^{m_{f_2e}} (\bar{B}_{f_3e^{-1}})^{m_{f_3e^{-1}}}, \quad (3.9)$$

where $e, e^{-1} \in \vec{A}_t \setminus \vec{A}_{t-1}$, and $f_0, f_1, f_2, f_3 \in \vec{A}_{t-1}$. It is not hard to see that $X_\gamma = \prod_{t=2}^K X_t$. Defining $Y_t = \prod_{j=2}^t X_j$ for $2 \leq t \leq K$, we see that $X_\gamma = Y_K$ and $Y_t = X_t Y_{t-1}$ for $2 < t \leq K$. Also, every Y_t is $\mathcal{F}_t$ -measurable. With these ingredients, by the law of total probability we have the following “tower of expectation:”

$$\mathbb{E}(X_\gamma) = \mathbb{E}(\mathbb{E}(Y_K | \mathcal{F}_{K-1})) = \mathbb{E}(\mathbb{E}(Y_{K-1} X_K | \mathcal{F}_{K-1})) = \mathbb{E}(Y_{K-1} \mathbb{E}(X_K | \mathcal{F}_{K-1})). \quad (3.10)$$

We focus on the term $\mathbb{E}(X_K | \mathcal{F}_{K-1})$ in Eqn. 3.10. For oriented edges e, f_0, f_1, f_2, f_3 , we have

$$(B_{ef_0} | f_0) = (B_{e^{-1}f_1} | f_1) = (B_{f_2e} | f_2) = (B_{f_3e^{-1}} | f_3),$$

where we assume that the orientation of the edges is such that these entries are not zero. These equalities say that, under the event that f_0, f_1, f_2, f_3 are edges of the graph, the presence of edge e or e^{-1} attached to any of them is equivalent, making these random variables are identical. Note that this involves the non-backtracking matrix B , not its “centered” version $\bar{B}$. These equivalences allow us to combine the multiplicities in Eqn. 3.9.

Let $\{Z_t\}_{2 \leq t \leq K}$ be independent random variables with distribution

$$Z_t \stackrel{d}{=} (B_{\vec{e}_t f} | f),$$

where $f \in \vec{A}_{t-1}$, $\vec{e}_t \in \vec{A}_t \setminus \vec{A}_{t-1}$, and $\vec{e}_t f$ is a subpath of γ . Since our convention is that $\vec{e}_K$ terminates in set V_2 , we have

$$\mathbb{E}(X_K | \mathcal{F}_{K-1}) = \mathbb{E} \left(\left(Z_K - \frac{d_1 - 1}{m} \right)^a \left(Z_K - \frac{d_2 - 1}{n} \right)^b \right), \quad (3.11)$$

where a is the number of times the new edge $\vec{e}_K$ appears as $B_{f_2 \vec{e}_K}$ and $B_{(\vec{e}_K)^{-1} f_1}$ in X_t , and b is the number of times it appears as $B_{\vec{e}_K f_0}$ and $B_{f_3 (\vec{e}_K)^{-1}}$. That is,

$$a = \sum_{f_1, f_2 \in \vec{A}_{K-1}} m_{(\vec{e}_K)^{-1} f_1} + m_{f_2 \vec{e}_K}$$

and

$$b = \sum_{f_0, f_3 \in \vec{A}_{K-1}} m_{\vec{e}_K f_0} + m_{f_3 (\vec{e}_K)^{-1}}.$$

To evaluate Eqn. (3.11), we have several cases depending on the values of a and b :

Case 1: If $a \geq 2$, then

$$\left(Z_K - \frac{d_1 - 1}{m} \right)^a \left(Z_K - \frac{d_2 - 1}{n} \right)^b \leq \left(Z_K - \frac{d_1 - 1}{m} \right)^2.$$

By Lemmas 11 and 8,

$$\mathbb{E} \left(\left(Z_K - \frac{d_1 - 1}{m} \right)^2 \right) \leq \frac{d_1 - 1}{m} + O \left(\frac{K}{m^2} \right).$$

When $b \geq 2$, the bound is analogous with $(d_2 - 1)/n$ instead.

Case 2: If $a = b = 1$, expand the right hand side of Eqn. (3.11) and use Lemma 8 to get

$$\begin{aligned} \mathbb{E} \left(\left(Z_K - \frac{d_1 - 1}{m} \right) \left(Z_K - \frac{d_2 - 1}{n} \right) \right) &= \mathbb{E} \left(Z_K \left(1 - \frac{d_1 - 1}{m} - \frac{d_2 - 1}{n} \right) + O \left(\frac{1}{nm} \right) \right) \\ &\leq \frac{d_2 - 1}{n} + O \left(\frac{K}{n^2} \right). \end{aligned}$$

Case 3: If $a = 0$ and $b > 1$ or $b = 0$ and $a > 1$, apply Lemma 11 directly to get the same bound as case 1 or 2.

Case 4: If $a = 0$ and $b = 1$, this is the one of the times we may see a bound of second order in n . Suppose that $b = 1$. By Lemma 8 part (ii), for Eqn. (3.11) we then have

$$\mathbb{E} \left(Z_K - \frac{d_2 - 1}{n} \right) \leq \begin{cases} \frac{d_2 - 1}{n} + O \left(\frac{K}{n^2} \right), & e_K \text{ is not a good edge} \\ O \left(\frac{K}{n^2} \right), & e_K \text{ is a good edge} \end{cases}$$

Case 5: If $b = 0$ and $a = 1$, we may also see a bound of second order in n . This is analogous to the previous case, and we get that

$$\mathbb{E} \left(Z_K - \frac{d_1 - 1}{m} \right) \leq \begin{cases} \frac{d_1 - 1}{m} + O \left(\frac{K}{n^2} \right), & e_K \text{ is not a good edge} \\ O \left(\frac{K}{n^2} \right), & e_K \text{ is a good edge} \end{cases}$$

These cases say that adding edge e_K to the graph H_{K-1} results in a different upper bound for $\mathbb{E}(X_K | \mathcal{F}_{K-1})$ depending on: (1) the multiplicity of the two-paths in which it appears, (2) whether it connects to a vertex in V_1 or V_2 when it attaches to H_{K-1} , and (3) whether it is a good edge. Note that cases 1–3 are ambiguous about what bipartite set e_K connects to in the graph H_{K-1} . If both endpoints of e_K connect to edges in H_{K-1} , then it is not a good edge, and Lemma 8 part (iii) says either degree bound is applicable.

Using Eqn. 3.11 and summarizing all of the cases enumerated above, we see that Eqn. (3.10) becomes

$$\begin{aligned} \mathbb{E}(X_\gamma) &= \mathbb{E}(Y_{K-1} \mathbb{E}(X_K | \mathcal{F}_{K-1})) \\ &\leq \begin{cases} \left(\frac{d_1 - 1}{m} + O \left(\frac{K}{n^2} \right) \right) \mathbb{E}(Y_{K-1}), & e_K \text{ connects to } V_1 \text{ or both and } a + b > 1 \\ \left(\frac{d_2 - 1}{n} + O \left(\frac{K}{n^2} \right) \right) \mathbb{E}(Y_{K-1}), & e_K \text{ connects to } V_2 \text{ or both and } a + b > 1 \\ O \left(\frac{K}{n^2} \right) \mathbb{E}(Y_{K-1}), & e_K \text{ is good and } a + b = 1 \\ \left(\frac{d_2}{n} + O \left(\frac{K}{n^2} \right) \right) \mathbb{E}(Y_{K-1}), & e_K \text{ is isolated from } H_{K-1}. \end{cases} \end{aligned} \quad (3.12)$$

Again, we see that if e_K connects two vertices already in H_{K-1} , we are free to choose which bound. We now apply the same argument to $\mathbb{E}(Y_{K-1}) = \mathbb{E}(Y_{K-2} \mathbb{E}(X_{K-1} | \mathcal{F}_{K-2}))$, and continue this process. After a total $K - 2$ times, we get to $\mathbb{E}(Y_2) = \mathbb{E}(X_2) = \mathbb{E}(\mathbb{E}(X_2 | \mathcal{F}_1))$.

The conditional expectation $\mathbb{E}(Y_2|\mathcal{F}_1)$ is just another term like in Eqn. (3.12), and then there is a final term that comes from the first edge. Again, we use the bound that best balances r and Q .

After all of these iterations, we end up with a bound of the form:

$$\begin{aligned}\mathbb{E}(X_\gamma) &\leq \left(\frac{d_1-1}{m}\right)^Q \left(\frac{d_2-1}{n}\right)^{K-Q-\omega} \left(\frac{d_2}{n}\right)^\omega \left(\frac{K}{n}\right)^\omega (1+o(1)) \\ &= \frac{D_\gamma}{m^Q n^{K-Q}} \left(\frac{K}{n}\right)^\omega (1+o(1)).\end{aligned}$$

Here, Q is the number of times the added edge attaches only in set V_1 . The powers of (d_1-1) and (d_2-1) are counted in $D_\gamma = (d_1-1)^Q (d_2-1)^{K-Q}$. The exponent ω counts the number of times we add a good edge with $a+b=1$, which gives an extra factor of K/n in the expectation. But for every good edge guaranteed by our ordering, we must also add an isolated edge first, which gives the weaker factor of d in D_γ for $\omega > 0$.

The order that we add the edges allows us to control Q . First of all, the edges that are ordered in Lemma 10 can be added in such a way that every vertex in V_2 contributes to Q : For a After adding the edges which we guarantee to be good, the process stops at some time $t = T$. For $t > T$, pick some $u \in V_1 \setminus V(H_{t-1})$ and an edge e incident to u . If this edge $e = (u, v)$ connects to a vertex v already in the current $V(H_{t-1})$, then v already has a factor in Q . If $v \notin V(H_{t-1})$, it is a currently isolated edge, and we take the bound that allows us to increase Q . Once every $u \in V_1$ is contained in H_{t-1} , any new edge will connect to just V_1 or both, and we can add one to Q for each $v \in V_2$ that needs it. Once $Q = r$, every vertex has been discovered, and thus any edges left over will connect to both sets V_1 and V_2 . In this case, we use the bound that does not increase Q .

By Lemma 10, we construct an ordering π which guarantees that certain edges are good. These edges must also have $a+b=1$. For a guaranteed good edge $e_{2t} \in F$, all of the edges in $N(\{e_{2t}\}) \setminus E(H_{2t-1})$ appear later. Since e_{2t} occurs in a two-path of multiplicity one in H_{2t} , by construction, then $a+b=1$. We complete the proof with

$$\omega \geq \left\lceil \frac{t(F)}{6d^2} \right\rceil \geq \left\lceil \frac{\sum_{ef \in \gamma} \mathbf{1}\{m_{ef} + m_{f^{-1}e^{-1}} = 1\}}{6d^2} \right\rceil.$$

□

3.5.4 Path counting

This section is devoted to count the number of ways non-backtracking walks can be concatenated to obtain a circuit as in Section 3.5.2. We will follow closely the combinatorial analysis used in [Brito et al., 2016]. In that paper, the authors needed a similar count for self-avoiding walks. We make the necessary adjustments to our current scenario. This is similar to the “cycling times” arguments of Bordenave [2015].

Our goal is to find a reasonable bound for the number of circuits which contribute to the trace bound, Eqn. (3.2) and shown graphically in Figure 3.2. Define $\mathcal{C}_{v,e}^r$ as those circuits which visit exactly $v = |V(\gamma)|$ different vertices, $r = |V(\gamma) \cap V_2|$ of them in the right set, and $e = |E(\gamma)|$ different edges. This is a set of circuits of length $2k\ell$ obtained as the concatenation of $2k$ non-backtracking, tangle-free walks of length ℓ . Note, these are undirected edges in $E(G)$ not directed edges in $\vec{E}(G)$. We denote such a circuit as $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{2k})$, where each γ_j is a length ℓ walk.

To bound $C_{v,e}^r = |\mathcal{C}_{v,e}^r|$, we will first choose set of vertices and order them. The circuits which contribute are indeed directed non-backtracking walks. However, by considering undirected walks along a fixed ordering of vertices, that ordering sets the orientation of the first and thus the rest of the directed edges in γ . Thus, we are counting the directed walks which contribute to Eqn. (3.2). We relabel the vertices as $1, 2, \dots, v$ as they appear in γ . Denote by $\mathcal{T}_\gamma$ the spanning tree of those edges leading to new vertices as induced by the path γ . The enumeration of the vertices tells us how we traverse the circuit and thus defines $\mathcal{T}_\gamma$ uniquely.

We encode each walk γ_j by dividing it into sequences of subpaths of three types, which in our convention *must always occur* as type 1 $\rightarrow$ type 2 $\rightarrow$ type 3, although some may be empty subpaths. Each type of subpath is encoded with a number, and we use the encoding to upper bound the number of such paths that can occur. Given our current

position on the circuit, i.e. the label of the current vertex, and the subtree of $\mathcal{T}_\gamma$ already discovered (over the whole circuit γ not just the current walk γ_j), we define the types and their encodings:

Type 1: These are paths with the property that all of their edges are edges of $\mathcal{T}_\gamma$ and have been traversed already in the circuit. These paths can be encoded by their end vertex. Because this is a path contained in a tree, there is a unique path connecting its initial and final vertex. We use 0 if no old edges occur before the type 2 path, i.e. the path is empty.

Type 2: These are paths with all of their edges in $\mathcal{T}_\gamma$ but which are traversed for the first time in the circuit. We can encode these paths by their length, since they are traversing new edges, and we know in what order the vertices are discovered. We use 0 if the path is empty.

Type 3: These paths are simply a single edge, not belonging to $\mathcal{T}_\gamma$, that connects the end of a path of type 1 or 2 to a vertex that has been already discovered. Given our position on the circuit, we can encode an edge by its final vertex. Again, we use 0 if the path is empty.

Now, we decompose γ_j into an ordered sequence of triples to encode its subpaths:

$$(p_1, q_1, r_1)(p_2, q_2, r_2) \cdots (p_t, q_t, r_t),$$

where each p_i characterizes subpaths of type 1, q_i characterizes subpaths of type 2, and r_i characterizes subpaths of type 3. These subpaths occur in the order given by the triples. We perform this decomposition using the minimal possible number of triples.

Now, p_i and r_i are both numbers in $\{0, 1, \dots, v\}$, since our cycle has v vertices. On the other hand, $q_i \in \{0, 1, \dots, \ell\}$ since it represents the length of a subpath of a non-backtracking walk of length ℓ . Hence, there are $(v+1)^2(\ell+1)$ possible triples. Next, we want to bound how many of these triples occur in γ_j . We will use the following lemma.

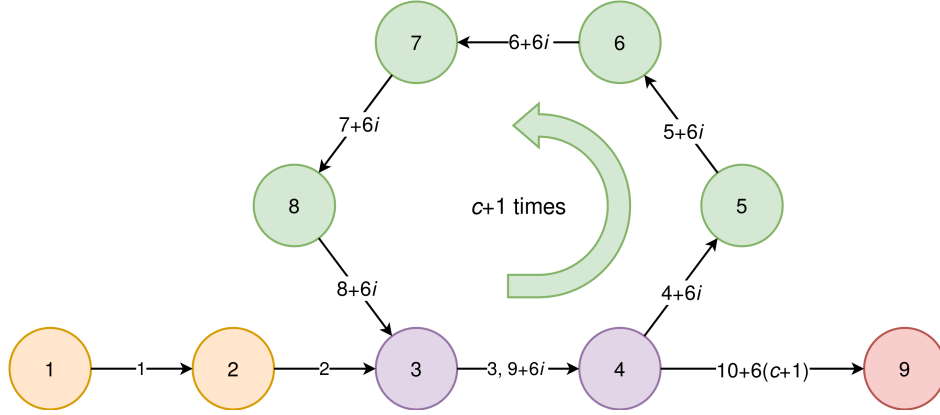


Figure 3.4: Encoding an ℓ -tangle-free walk, in this case the first walk in the circuit γ_1 , when it contains a cycle. The vertices and edges are labeled in the order of their traversal. The segments γ^a , γ^b , and γ^c occur on edges numbered $(1, 2, 3); (4 + 6i, 5 + 6i, 6 + 6i, 7 + 6i, 8 + 6i, 9 + 6i)$ for $i = 0, 1, \dots, c$; and $(10 + 6c)$, respectively. The encoding is $(0, 3, 0) \parallel (0, 4, 3)(4, 0, 0) \parallel (0, 1, 0)$. Suppose $c = 1$. Then $\ell = 22$ and the encoding is of length $3 + (4 + 1 + 1)(c + 1) + 1$, we can back out c to find that the cycle is repeated twice. The encodings become more complicated later in the circuit as vertices see repeat visits.

Lemma 12. Let $(p_1, q_1, r_1)(p_2, q_2, r_2) \cdots (p_t, q_t, r_t)$ be a minimal encoding of a non backtracking walk γ_j , as described above. Then $r_i = 0$ can only occur in the last triple $i = t$.

Proof. We can check this case by case. Assume that for some $i < t$ we have $(p_i, q_i, 0)$, and consider the concatenation with $(p_{i+1}, q_{i+1}, r_{i+1})$. First, notice that both p_{i+1} and q_{i+1} cannot be zero since then we will have $(p_i, q_i, 0)(0, 0, v^*)$ which can be written as (p_i, q_i, v^*) . If $q_i \neq 0$, then we must have $p_{i+1} \neq 0$. Otherwise, we split a path of new edges (type 2), and the decomposition is not minimal. This implies that we visit new edges and move to edges already visited, hence we need to go through a type 3 edge, implying that $r_i \neq 0$. Finally, if $p_i \neq 0$ and $q_i = 0$, then we must have $p_{i+1} = 0$; otherwise, we split a path of old edges (type 1). We also require $q_{i+1} \neq 0$, but $(p_i, 0, 0)(0, q_{i+1}, r_{i+1})$ is the same as (p_i, q_{i+1}, r_{i+1}) , which contradicts the minimality condition. This covers all possibilities and finishes the proof. $\square$

Using the lemma, any encoding of a non-backtracking walk γ_j has at most one triple

with $r_i = 0$. All other triples indicates the traversing of a type 3 edge. We now give a very rough upper bound for how many of such encodings there can be. To do so, we will use the tangle-free property and slightly modify the encoding of the paths with cycles. Consider the two cases:

Case 1: Path γ_j contains no cycle. This implies that we traverse each edge within γ_j once. Thus, we can have at most $\chi = e - v + 1$ many triples with $r_i \neq 0$. This gives a total of at most

$$\left((v+1)^2(\ell+1)\right)^{\chi+1}$$

many ways to encode one of these paths.

Case 2: Path γ_j contains a cycle. Since we are dealing with non-backtracking, tangle-free walks, we enter the cycle once, loop around some number of times, and never come back. We change the encoding of such paths slightly. Let γ_j^a , γ_j^b , and γ_j^c be the segments of the path before, during, and after the cycle. We mark the start of the cycle with $|$ and its end with $\|$. The new encoding of the path is:

$$(p_1^a, q_1^a, r_1^a) \cdots (p_{t^a}^a, q_{t^a}^a, r_{t^a}^a) | (p_1^b, q_1^b, r_1^b) \cdots (p_{t^b}^b, q_{t^b}^b, r_{t^b}^b) \| (p_1^c, q_1^c, r_1^c) \cdots (p_{t^c}^c, q_{t^c}^c, r_{t^c}^c),$$

where we encode the segments separately. Observe that each a subpath is connected and self-avoiding. The above encoding tells us all we need to traverse γ_j , including how many times to loop around the cycle: since the total length is ℓ , we can back out the number of circuits around the cycle from the lengths of γ_j^a , γ_j^b , and γ_j^c . See Figure 3.4. Following the analysis made for Case 1, the subpaths γ_j^a , γ_j^b , γ_j^c are encoded by at most $\chi + 1$ triples, but we also have at most ℓ choices each for our marks $|$ and $\|$. We are left with at most

$$\ell^2 \left((v+1)^2(\ell+1)\right)^{\chi+1}$$

ways to encode any path of this kind.

Together, these two cases mean there are less than $2\ell^2 ((v+1)^2(\ell+1))^{\chi+1}$ such paths.

Now we conclude by encoding the entire path $\gamma = (\gamma_1, \dots, \gamma_{2k})$. We first choose v vertices, r in the set V_2 , and order them, which can occur in $\binom{m}{r} \binom{n}{v-r} \leq m^r n^{v-r}$ different ways. Finally, in the whole path γ we are counting concatenations of $2k$ paths which are ℓ -tangle-free. Therefore, we conclude with the following Lemma:

Lemma 13. *If $C_{v,e}^r = |C_{v,e}^r|$, then this satisfies*

$$C_{v,e}^r \leq m^r n^{v-r} 2^{2k} \ell^{4k} \left((v+1)^2 (\ell+1) \right)^{2k(\chi+1)}. \quad (3.13)$$

3.5.5 Bounds on the norm of $\bar{B}^\ell$ and $R^{\ell,j}$.

Theorem 14. *Let $\ell \leq c \log(n)$ where c is a universal constant. It holds that*

$$\|\bar{B}^\ell\| \leq \log(n)^{15} ((d_1 - 1)(d_2 - 1))^{\ell/4}$$

asymptotically almost surely.

Proof. For any natural number k , we have

$$\mathbb{E} \left(\|\bar{B}^\ell\|^{2k} \right) \leq \mathbb{E} \left(\text{Tr} \left((\bar{B}^\ell)(\bar{B}^\ell)^* \right)^k \right) = \mathbb{E} \left(\sum_{\gamma} \prod_{i=1}^{2k\ell} \bar{B}_{e_i e_{i+1}} \right). \quad (3.14)$$

The sum is taken over the set of all circuits γ of length $2k\ell$, where $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{2k})$ is formed by concatenation of $2k$ tangle-free segments $\gamma_s \in F^\ell$, with the convention $e_1^{s+1} = e_{\ell+1}^s$. Again, refer to Figure 3.2 for clarification.

Recall that two oriented edges e_1 and e_2 form a subpath of γ if we traverse one right after the other. We call such subpath a two-path. With this notation, we define three disjoint sets of circuits:

$\mathcal{C}_1 = \{\gamma : \text{all two-paths in } \gamma \text{ are traversed at least twice, disregarding the orientation}\},$

$\mathcal{C}_2 = \{\gamma : \text{at least one two-path in } \gamma \text{ is traversed exactly once and } v \leq kl + 1\}, \text{ and}$

$\mathcal{C}_3 = \{\gamma : \text{at least one two-path in } \gamma \text{ is traversed exactly once and } v > kl + 1\}.$

The reason for this division is that, by Lemma 8 and Theorem 9, when we have any two-path traversed exactly once, the expectation of the corresponding circuit is smaller. This is precisely because the matrix $\bar{B}$ is nearly centered. Hence, we will see that the leading order terms in Eqn. (3.14) will come from circuits in $\mathcal{C}_1$.

Define the expectations

$$I_j = \mathbb{E} \left(\sum_{\gamma \in \mathcal{C}_j} \prod_{i=1}^{2k\ell} \bar{B}_{e_i e_{i+1}} \right)$$

for $j = 1, 2$ and 3 , so that

$$\mathbb{E} \left(\|\bar{B}^\ell\|^{2k} \right) \leq I_1 + I_2 + I_3. \quad (3.15)$$

We will bound each term on the right hand side above. From Theorem 9 and the path counting bound in Lemma 13, we get that

$$\begin{aligned} I_j &\leq \sum_{\gamma \in \mathcal{C}_j} C_{v,e}^r \frac{D_\gamma}{m^r n^{|E(\gamma)|-r}} \left(\frac{|E(\gamma)|}{n} \right)^{\omega(\gamma)} (1 + o(1)) \\ &\leq \sum_{v,e,r} \sum_{\gamma \in \mathcal{C}_j \cap \mathcal{C}_{v,e}^r} C_{v,e}^r \frac{D_\gamma^*}{m^r n^{e-r}} \left(\frac{e}{n} \right)^{\omega_*} (1 + o(1)) \\ &\leq \sum_{v,e,r} \sum_{\gamma \in \mathcal{C}_j \cap \mathcal{C}_{v,e}^r} n^{v-e} (2\ell)^{4k} \left((v+1)^2 (\ell+1) \right)^{2k(\chi+1)} D_\gamma^* \left(\frac{e}{n} \right)^{\omega_*} (1 + o(1)) \end{aligned} \quad (3.16)$$

where $D_\gamma^* = \max_{\gamma \in \mathcal{C}_{v,e}^r} D_\gamma$, and $\omega_* = \min_{\gamma \in \mathcal{C}_{v,e}^r} \omega(\gamma)$. Since we have employed Theorem 9, we split the circuits into subsets depending on the number of vertices $v = |V(\gamma)|$, number of edges $e = |E(\gamma)|$, and the number of vertices r in set V_2 . The sum over $\gamma \in \mathcal{C}_j$ implicitly sums over v , e , and r . We will use Eqn. 3.16 to bound each I_j .

Bounding I_1

In all cases, each circuit traverses $2k\ell$ two-paths. Hence, for each $\gamma \in \mathcal{C}_1$, where each two-path is repeated twice, we have at most $k\ell$ different two-paths. Furthermore, since each edge can be in multiple two-paths, we have that the total number of different two-paths is greater than or equal to the total number of edges traversed by γ . We then have

that $e \leq k\ell$. Since γ is connected, we have $v \leq k\ell + 1$. Lastly, observe that $\omega = 0$ for any $\gamma \in \mathcal{C}_1$. Thus, on the right hand side of Eqn. (3.16) we get

$$I_1 \leq \sum_{v=\ell+1}^{k\ell+1} \sum_{e=v-1}^{k\ell} n^{v-e} (2\ell)^{4k} \left((v+1)^2 (\ell+1) \right)^{2k(\chi+1)} D_{\gamma}^* (1 + o(1)).$$

The leading order term in n corresponds to $v = e + 1 = k\ell + 1$ and $e = k\ell$. Because any γ is connected, for these values of v and e the undirected graph induced by γ is a tree, which implies that $\chi = 0$. We conclude that

$$I_1 \leq n \left(2\ell(k\ell + 2)^2 (\ell + 1) \right)^{4k} ((d_1 - 1)(d_2 - 1))^{\lceil \frac{k\ell}{2} \rceil} (1 + o(1)). \quad (3.17)$$

Note that this term comes from paths γ that leave some vertex, explore the graph up to a distance $k\ell$, then return along the same path in the opposite direction, thus traversing the undirected path twice. Each length ℓ segment is non-backtracking and tangle-free, but the overall path is backtracking. At the end it must backtrack to reverse direction, as it does at each point where the segments are joined, as shown in Figure 3.2.

Bounding I_2

We turn our attention to I_2 . Because there is at least one two-path traversed exactly once, we have $e \geq v$ for $\gamma \in \mathcal{C}_2$. Eqn. (3.16) becomes

$$I_2 \leq \sum_{v=\ell+1}^{k\ell+1} \sum_{e=v}^{2k\ell} n^{v-e} (2\ell)^{4k} \left((v+1)^2 (\ell+1) \right)^{2k(\chi+1)} D_{\gamma}^*,$$

where we dropped the term $\left(\frac{e}{n}\right)^{\omega_*} \leq 1$. Now the leading order term comes when $e = v = k\ell + 1$, implying $\chi = 1$ and yielding

$$I_2 \leq \left(2\ell(k\ell + 2)^2 (\ell + 1) \right)^{4k} d^{k\ell} (1 + o(1)). \quad (3.18)$$

Bounding I_3

We focus on I_3 last. Notice that cycles in $\mathcal{C}_3$ will visit many vertices, since $v > k\ell + 1$. We first show that, in this case, $\omega(\gamma)$ is also large. Let $v = k\ell + t$. Define $p_2(\gamma)$ as the number

of different two-paths traversed by γ , and $p_2^*(\gamma)$ as the number of two-paths traversed exactly once. We have $p_2(\gamma) \geq e \geq v = k\ell + t$. Also, since γ has length $2k\ell$, we deduce that

$$2(p_2(\gamma) - p_2^*(\gamma)) + p_2^*(\gamma) \leq 2k\ell,$$

which implies that $p_2^*(\gamma) \geq 2t$. Therefore, $\omega_* \geq \frac{2t}{6d^2}$. Eqn. (3.16) then gives

$$I_3 \leq \sum_{v=k\ell+1}^{2k\ell} \sum_{e=v}^{2k\ell} n^{v-e} (2\ell)^{4k} \left((v+1)^2(\ell+1) \right)^{2k(\chi+1)} D_\gamma^* \left(\frac{e}{n} \right)^{\frac{2(v-k\ell)}{6d^2}}.$$

Since $D_\gamma^* \leq d^e$, for the leading order term $v = e$ we have

$$D_\gamma^* \left(\frac{e}{n} \right)^{\frac{2(v-k\ell)}{6d^2}} \leq d^v \left(\frac{v}{n} \right)^{\frac{2(v-k\ell)}{6d^2}} = d^{k\ell} \left(d \left(\frac{v}{n} \right)^{\frac{2}{6d^2}} \right)^{(v-k\ell)} \leq d^{k\ell},$$

since $v \leq 2k\ell = o(n)$, d is constant, and $v - k\ell \geq 1$. Now $v = e$ means that $\chi = 1$, and we get

$$I_3 \leq k\ell \left(2\ell(v+1)^2(\ell+1) \right)^{4k} d^{k\ell} (1 + o(1)). \quad (3.19)$$

Finishing the proof of Theorem 14

Now we compute the leading order contribution to $\mathbb{E} (\|\bar{B}^\ell\|^{2k})$. We consider

$$k = \left\lfloor \frac{\log(n)}{\log(\log(n))} \right\rfloor.$$

with $\ell \leq \frac{1}{8} \log(n)$. We see that I_1 has n in front, as opposed to I_2 and I_3 , and that the other terms are of similar order. Thus, plugging Eqns. (3.17), (3.18) and (3.19) into Eqn. (3.15), we find that only I_1 contributes:

$$\mathbb{E} (\|\bar{B}^\ell\|^{2k}) \leq I_1 \leq n \left(2\ell(k\ell+2)^2(\ell+1) \right)^{4k} ((d_1-1)(d_2-1))^{\lfloor \frac{k\ell+1}{2} \rfloor} (1 + o(1)).$$

We now apply Markov's inequality. With this choice of k and ℓ , we have

$$\left(2\ell(k\ell+2)^2(\ell+1) \right)^{4k} = O(n^{28}).$$

and $\log(n)^{30k} = n^{30}$. Therefore,

$$\begin{aligned} \mathbb{P}(\|\bar{B}^\ell\| > \log(n)^{15} ((d_1 - 1)(d_2 - 1))^{\ell/4}) &\leq \frac{\mathbb{E}(\|\bar{B}^\ell\|^{2k})}{\log(n)^{30k} ((d_1 - 1)(d_2 - 1))^{k\ell/2}} \\ &\leq n^{-29} \left(2\ell(k\ell + 2)^2(\ell + 1)\right)^{4k} (d_1 - 1)(d_2 - 1)(1 + o(1)) \\ &= o(1). \end{aligned}$$

□

Theorem 15. *Let $1 \leq j \leq \ell \leq c \log(n)$ where c is a universal constant. Then*

$$\|R^{\ell,j}\| \leq \log(n)^{16},$$

asymptotically almost surely.

Proof. The proof is analogous to the proof of Theorem 14. We have, for any integer k

$$\mathbb{E}(\|R^{\ell,j}\|^{2k}) \leq \mathbb{E}\left(\text{Tr}\left((R^{\ell,j})(R^{\ell,j})^*\right)^k\right) = \mathbb{E}\left(\sum_{\gamma} \prod_{s=1}^{2k} \prod_{i=1}^{j-1} \bar{B}_{e_i^s e_{i+1}^s} S_{e_j^s e_{j+1}^s} \prod_{i=j+1}^{\ell} B_{e_i^s e_{i+1}^s}\right). \quad (3.20)$$

Now, the sum is over circuits $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{2k})$ of length $2k\ell$ formed from $2k$ elements of $T^{\ell,j}$, $\gamma_s = (e_1^s, e_2^s, \dots, e_{\ell+1}^s)$ for $s \in [2k]$, with the convention $e_1^{s+1} = e_{\ell+1}^s$. Recall the definition of $T^{\ell,j}$ in Section 3.5.2 and the definition of $R^{\ell,j}$ in Eqn. 3.4.

Denote by $\mathcal{D}_{v,e}^r$ as set of circuits that visit exactly v vertices, r of which are in V_2 , and $e + 2k$ different edges. We have to slightly modify the argument of Section 3.5.4 for this case. Here, the extra $2k$ edges are the j th edge in each segment γ_s , which connects the first j edges to the last $\ell - j$. For $\gamma \in \mathcal{D}_{v,e}^r$, each γ_s is divided into two tangle-free, non backtracking walks of length j and $\ell - j$. By encoding each of these paths as in Section 3.5.4, we conclude that there are at most

$$4\ell^4((v+1)^4(\ell+1)^2)^{\chi+1}$$

such γ_s . Concatenating $2k$ many of these gives

$$D_{v,e}^r = |\mathcal{D}_{v,e}^r| \leq n^{v-r} m^r (4\ell^4)^{2k} ((v+1)^4 (\ell+1)^2)^{2k(\chi+1)}. \quad (3.21)$$

We will now obtain an analagous result to Theorem 9. Returning to Eqn. (3.20), notice that for each γ we have terms of the form

$$(\bar{B}_{ef})^{m_{ef}} (B_{ef})^{m'_{ef}},$$

since now two-paths are weighted by entries of both $\bar{B}$ and B . Here, m'_{ef} is the number of times we traverse the oriented two-path ef and get contributions from B . If $m'_{ef} > 0$, we have

$$(\bar{B}_{ef})^{m_{ef}} (B_{ef})^{m'_{ef}} \leq (1 - B_{ef})^{m_{ef}} B_{ef} \leq B_{ef}$$

and the corresponding conditional expectation can be upper bounded by $\frac{d-1}{n}$, by Lemma 8. If $m'_{ef} = 0$, we proceed as in the proof of Theorem 9. After dropping the term $(K/n)^\omega$ terms, which will not be important, we get the corresponding expectation bound

$$\mathbb{E} \left(\|R^{\ell,j}\|^{2k} \right) \leq \sum_{v,e,r} \sum_{\gamma \in \mathcal{D}_{v,e}^r} D_{v,e}^r \frac{(d-1)^e}{m^r n^{e-r}} \left(\frac{d-1}{n} \right)^{2k} (1 + o(1)),$$

where we included an upper bound on the factor arising from the $2k$ entries of the matrix S (see Section 3.5.2). Using the bound in Eqn. (3.21), we obtain

$$\mathbb{E} \left(\|R^{\ell,j}\|^{2k} \right) \leq \sum_{v,e,r} \sum_{\gamma \in \mathcal{D}_{v,e}^r} n^{v-e-2k} (4\ell^4)^{2k} ((v+1)^4 (\ell+1)^2)^{2k(\chi+1)} (d-1)^{2k} (1 + o(1)).$$

Finally, we notice that $1 \leq v \leq 2k\ell$ and $v-1 \leq e \leq 2k\ell$. Furthermore, if $v-e=1$, then γ induces a tree and r takes a unique value. We show that this will lead to a term linear in n . Also, for fixed v and e there are less than n different values of r . Since we have at most $4k^2\ell^2$ pairs, we conclude:

$$\begin{aligned} \mathbb{E} \left(\|R^{\ell,j}\|^{2k} \right) &\leq (4k^2\ell^2) n (4\ell^4)^{2k} ((2k\ell+1)^4 (\ell+1)^2)^{2k} \left(\frac{(d-1)^{\ell+1}}{n} \right)^{2k} (1 + o(1)) \\ &\leq (4k^2\ell^2) n (4\ell^4)^{2k} ((2k\ell+1)^4 (\ell+1)^2)^{2k} (1 + o(1)). \end{aligned}$$

Now to finish the proof. Let

$$k = \left\lfloor \frac{\log(n)}{\log(\log(n))} \right\rfloor,$$

so that now

$$\mathbb{E} \left(\|R^{\ell,j}\|^{2k} \right) \leq n^{29} (1 + o(1)).$$

Then, by Markov's inequality and since $\log(n)^{32k} = n^{32}$,

$$\begin{aligned} \mathbb{P}(\|R^{\ell,j}\| > \log(n)^{16}) &\leq \frac{\mathbb{E}(\|R^{\ell,j}\|^{2k})}{\log(n)^{32k}} \\ &\leq n^{-31} 4k^2 \ell^2 (4\ell^4)^{2k} ((v+1)^4 (\ell+1)^2)^{2k} (1 + o(1)) \\ &= o(1). \end{aligned}$$

□

3.5.6 Proof of the main result, Theorem 3

By Eqns. (3.1) and (3.6),

$$|\lambda_2|^\ell \leq \|\bar{B}^\ell\| + \sum_{k=1}^{\ell} \|R^{\ell,j}\|.$$

Combining Theorem 14 and 15:

$$\begin{aligned} |\lambda_2| &\leq \left(\log(n)^{15} ((d_1 - 1)(d_2 - 1))^{\ell/4} + \ell \log(n)^{16} \right)^{1/\ell} \\ &= ((d_1 - 1)(d_2 - 1))^{1/4} + \epsilon_n. \end{aligned}$$

3.6 Application: Community detection

In many cases, such as online networks, we would like to be able to recover specific communities in those graphs. In the typical setup, a community is a set of vertices that are more densely connected together than to the rest of the graph.

The model we present here is inspired by the planted partition or stochastic block-model (SBM) [Holland et al., 1983]. In the SBM, each vertex belongs to a class or community, and the probability that two vertices are connected is a function of the classes of the

vertices. It is a generalization of the Erdős-Rényi random graph. The classes or blocks in the SBM make it a good model for graphs with community structure, where nodes preferentially connect to other nodes depending on their communities [Newman, 2010].

There are many methods for detecting a community given a graph. For an overview of the topic, see Fortunato [2010]. Spectral clustering is a common method which can be applied to any set of data $\{\zeta_i\}_{i=1}^n$. Given a symmetric and non-negative similarity function S , the similarity is computed for every pair of data points, forming a matrix $A_{ij} = S(\zeta_i, \zeta_j) = S(\zeta_j, \zeta_i) \geq 0$. The spectral clustering technique is to compute the leading eigenvectors of A , or matrices related to it, and use the eigenvectors to cluster the data. In our case, the matrix in question is just the Markov matrix of a graph, defined soon. We will show that we can guarantee the success of the technique if the degrees are large enough.

Our graph model is a regular version of the SBM. We build it on a “frame,” which is a small, weighted graph that defines the community structure present in the larger, random graph. Each class is represented by a vertex in the frame. The edge weights in the frame define the number of edges between classes. What makes our model differ from the SBM is that the connections between classes are described by a regular random graph rather than an Erdős-Rényi random graph. However, the graph itself is not necessarily regular.

A number of authors have studied similar models. Our model is a type of random *lift* of the frame, which is said to *cover* the random graph [e.g. Marcus et al., 2013a, Angel et al., 2015, Bordenave, 2015]. This type of random graph was also studied by Newman and Martin [2014], who called it an equitable random graph, since the community structure is equivalent to an equitable partition. This partition induces a number of symmetries across vertices in each community which are useful when studying the eigenvalues of the graph. Barrett et al. [2017] studied the effect of these symmetries from a group theory standpoint. The work of Barucca [2017] is closest to ours: they consider spectral properties of such graphs and their implications for spectral community clustering. In particular, they show that the spectrum of what we call the “frame” (in their words, the discrete spectrum,

which is deterministic) is contained in that of the random graph. They use the resolvent method (called the cavity method in the physics community) to analyze the continuous part of the spectrum in the limit of large graph size, and argue that community detection is possible when the deterministic frame eigenvalues all lie outside the bulk. However, this analysis assumes that there are no stochastic eigenvalues outside the bulk, which will only hold with high probability if the graph is Ramanujan. Our analysis shows that, if a set of pairwise spectral gaps hold between all communities, then this will be the case.

3.6.1 The frame model

We define the *random regular frame graph* distribution $\mathcal{G}(n, H)$ as a distribution of simple graphs on n vertices parametrized by the “frame” H . The frame $H = (V, E, p, D)$ is a weighted, directed graph. Here, V is the vertex set, $E \subseteq \{(i, j) : i, j \in V\}$ is the directed edge set, the vertex weights are p , and the edge weights are D . Note that we drop the arrows on the edge set in this Section, since it will always be directed. The vertex weight vector $p \in \mathbb{R}^{|V|}$, where $\sum_{i \in V} p_i = 1$, sets the relative sizes of the classes. The edge weights are a matrix of degrees $D \in \mathbb{N}^{|V| \times |V|}$. These assign the number of edges between each class in the random graph: D_{ij} is the number of edges from each vertex in class i to vertices in class j . The degrees must satisfy the balance condition

$$p_i D_{ij} = p_j D_{ji} \quad (3.22)$$

for all $i, j \in V$ where (i, j) or (j, i) are in E . This requires that, for every edge $e \in E$, its reverse orientation also exists in H . We also require that $n_i = np_i \in \mathbb{N}$ for every $i \in V$, so that the number of vertices in each type is integer.

Given the frame H , a random regular frame graph $G \sim \mathcal{G}(n, H)$ is a simple graph on n vertices with n_i vertices in class i . It is chosen uniformly among graphs with the constraint that each vertex in class i makes D_{ij} connections among the vertices in class j . In other words, if $i = j$, we sample that block of the adjacency matrix as the adjacency matrix of a D_{ii} -regular random graph on n_i vertices. For off-diagonal blocks $i \neq j$, these

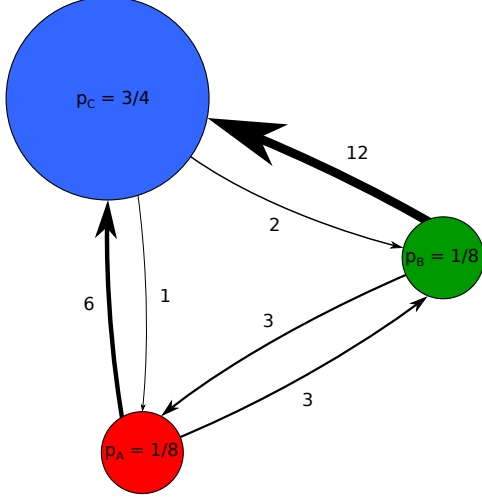
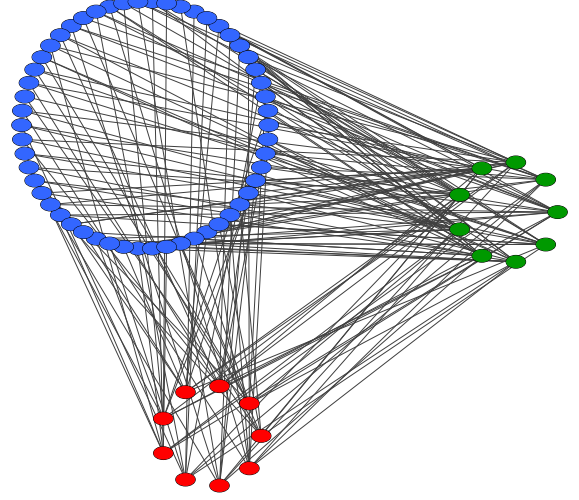
A Frame**B** Random regular frame graph

Figure 3.5: Schematic and realization of a random regular frame graph. **A**, the frame graph. The vertices of the frame (red = A, green = B, blue = C) are weighted according to their proportions p in the random regular frame graph. The edge weights D_{ij} set the between-class vertex degrees in the random regular frame graph. This frame will yield a random tripartite graph. **B**, realization of the graph on 72 vertices. In this instance, there are $1/8 \times 72 = 9$ green and red vertices and $3/4 \times 72 = 54$ blue vertices. Each blue vertex connects to $k_{CA} = 1$ red vertex and $k_{CB} = 2$ green vertices. This is actually a multigraph; with so few vertices, the probability that the configuration model algorithm yields parallel edges is high.

are sampled as bipartite, biregular random graphs $\mathcal{G}(n_i, n_j, D_{ij}, D_{ji})$.

Sampling from $\mathcal{G}(n, H)$ can be performed similar to the configuration model, where each node is assigned as many half-edges as its degree, and these are wired together with a random matching [Newman, 2010]. The detailed balance condition Eqn. (3.22) ensures that this matching is possible. Practically, we often have to generate many candidate matchings before the resulting graph is simple, but the probability of a simple graph is bounded away from zero for fixed D .

An example of a random regular frame graph is the bipartite, biregular random graph. The family $\mathcal{G}(n, m, d_1, d_2)$ is a random regular frame graph $\mathcal{G}(n + m, H)$, where the frame

H is the directed path on two vertices: $V = \{1, 2\}$ and $E = \{(1, 2), (2, 1)\}$. The weights are taken as $p_1 = n/(n + m)$, $p_2 = m/(n + m)$, $D_{12} = d_1$, and $D_{21} = d_2$.

Another example random regular frame graph is shown in Figure 3.5. In this case, the frame H has $V = \{A, B, C\}$ and $E = \{(A, B), (A, C), (B, A), (B, C), (C, A), (C, B)\}$ with weights p and D as shown in Figure 3.5A. We see that this generates a random tripartite graph with regular degrees between vertices in different independent sets, shown in Figure 3.5B.

3.6.2 Markov and related matrices of frame graphs

Now, we define a number of matrices associated with the frame and the sample of the random regular frame graph.

Let G be a simple graph. Define $D_G = \text{diag}(d_G)$, the diagonal matrix of degrees in G . The Markov matrix $P = P(G)$ is defined as

$$P = D_G^{-1}A,$$

where $A = A(G)$ is the adjacency matrix. The Markov matrix is the row-normalized adjacency matrix, and it contains the transition probabilities of a random walker on the graph G . Let $L = D_G^{-1/2}A D_G^{-1/2}$ be the normalized Laplacian. Then P and L have the same eigenvalues, but L is symmetric, since $L_{ij} = \frac{A_{ij}}{\sqrt{d_i d_j}}$.

Suppose $G \sim \mathcal{G}(n, H)$, where the frame $H = (V, E, p, D)$. Another matrix that will be useful is what we call the *Markov matrix of the frame* R , where $R_{ij} = \frac{D_{ij}}{\sum_j D_{ij}}$. Thus, R is a row-normalized D , in the same way that the Markov matrix P is the row-normalized adjacency matrix A . Furthermore, R is invariant under any uniform scaling of the degrees. Because of this equitable partition property of random regular frame graphs, eigenvectors of the frame matrices $D = D(H)$ or $R = R(H)$ lift to eigenvectors of $A = A(G)$ or $P = P(G)$, respectively. Suppose $Dx = \lambda x$, then it is a straightforward exercise to check that $A\tilde{x} = \lambda\tilde{x}$

for the piecewise constant vector

$$\tilde{x} = \begin{bmatrix} \mathbf{1}_{n_1} x_1 \\ \mathbf{1}_{n_2} x_2 \\ \vdots \end{bmatrix}.$$

Using the same procedure, we can lift any eigenpair of R to an eigenpair of P with the same eigenvalue.

Bounds on the eigenvalues of frame graphs in terms of blocks

The following result is due to Marina Meila (2015, personal correspondence).

Proposition 16. *Let G be a random regular frame graph $G(n, H)$, P its Markov matrix, and L the Laplacian with vertices ordered by class in both cases. Let R be the Markov matrix of the frame $H = (V, E, p, D)$, with $|V(H)| = K$ classes. Define the matrices $L^{(kl)}$ as the (k, l) block of L with respect to the clustering of vertices by class. For $l \neq k$, let*

$$M^{(kl)} = \begin{pmatrix} 0 & L^{(kl)} \\ L^{(kl)} & 0 \end{pmatrix} = \begin{pmatrix} 0 & L^{(kl)} \\ L^{(lk)*} & 0 \end{pmatrix}.$$

For $l = k$, let $M^{(kk)} = L^{(kk)}$. Assume that all eigenvalues of D are nonzero and pick a constant C such that

$$\frac{|\lambda_2^{(kl)}|}{\lambda_1^{(kl)}} \leq C < 1$$

for every $k, l = 1, \dots, K$, where $\lambda_1^{(kl)}$ and $\lambda_2^{(kl)}$ are the leading and second eigenvalues of $M^{(kl)}$.

Under these conditions, the eigenvalues of P which are not eigenvalues of R are bounded by

$$C \max_{k=1, \dots, K} \left(R_{kk} + \sum_{l \neq k} \sqrt{R_{kl} R_{lk}} \right) \leq \frac{C}{2} \left(1 + \max_{k=1, \dots, K} \sum_{l=1}^K R_{lk} \right).$$

The spectrum of the Markov matrix $\sigma(P)$ enjoys a simple connection to $\sigma(A)$ when A is the adjacency matrix of a graph drawn from $G(n, m, d_1, d_2)$. In this case, $P = L = \frac{A}{\sqrt{d_1 d_2}}$, so the eigenvalues of P are just the scaled eigenvalues of A . This and the spectral gap for bipartite, biregular random graphs, Corollary 4, lead to the following remark:

Remark. For a random regular frame graph, $M^{(kl)}$ corresponds to the normalized Laplacian of a bipartite biregular graph $G(n_k, n_l, D_{kl}, D_{lk})$. Thus,

$$\frac{|\lambda_2^{(kl)}|}{\lambda_1^{(kl)}} \leq \frac{\sqrt{D_{kl} - 1} + \sqrt{D_{lk} - 1}}{\sqrt{D_{kl} D_{lk}}} + \epsilon.$$

Suppose we are given a frame that fits the conditions of Proposition 16; namely, D cannot have any zero eigenvalues. Then we can uniformly grow the degrees, which leaves R invariant, but allows us to reach an arbitrarily small C . This ensures that the leading K eigenvalues of P are equal to the eigenvalues of R . Note that this actually means that the entire random regular frame graph satisfies a weak Ramanujan property. We now show that this guarantees spectral clustering.

3.6.3 Spectral clustering

Spectral clustering is a popular method of community detection. Because some eigenvectors of P , the Markov matrix of a random regular frame graph, are piecewise constant on classes, we can use them to recover the communities so long as those eigenvectors can be identified. Suppose there are K total classes in our random regular frame graph. Then, given the eigenvectors $x^1, x^2, \dots, x^K$, which are piecewise constant across classes, we can cluster vertices by class. For each vertex $v \in V(G)$, associate the vector $y^v \in \mathbb{R}^K$ where $y_j^v = x_v^j$. Then if $y^v = y^u$ for $u, v \in V(G)$, vertices u and v belong to the same class¹. It is simple to recover these piecewise constant vectors $x^1, x^2, \dots, x^K$ when they are the leading eigenvectors. These facts lead to the following theorem:

Theorem 17 (Spectral clustering guarantee in frame graphs). *Let G be a random regular frame graph $G(n, H)$ and P its Markov matrix. Let R be the Markov matrix of the frame $H = (V, E, p, D)$, with $|V(H)| = K$ classes and $\lambda_1 \geq \dots \geq \lambda_K$ the eigenvalues of R and $|\lambda_K| > 0$.*

¹ In the SBM case, the eigenvectors are not piecewise constant, but they are aligned with the eigenvectors of R and thus highly correlated across vertices in the same class. A more flexible clustering method such as K -means must be applied to the vectors y in that case.

Then we can scale the degrees by some $\kappa \in \mathbb{N}$, $D \rightarrow \kappa D$, so that the vertex classes are recoverable by spectral clustering of the leading K eigenvectors of P .

Remark. The conditions of Theorem 17, while very general, are also weaker than may be expected using more sophisticated methods tailored to the specific frame model. We illustrate this with the following example.

Example: The regular stochastic block model

Brito et al. [2016] and Barucca [2017] studied a regular stochastic block model, which can be seen as a special case of our frame model. Let the frame H be the complete directed graph on two vertices, including self loops, where

$$D = \begin{pmatrix} d_1 & d_2 \\ d_2 & d_1 \end{pmatrix}$$

and $p = (1/2, 1/2)$. Define the regular stochastic block model as $\mathcal{G}(2n, H)$. This is a graph with two classes of equal size, representing two communities of vertices, with within-class degree d_1 and between-class degree d_2 . We assume $d_1 > d_2$, since communities are more strongly connected within. Brito et al. proved the following theorem:

Theorem 18. *If $(d_1 - d_2)^2 > 4(d_1 + d_2 - 1)$, then there is an efficient algorithm for strong recovery, i.e. recovery of the exact communities with high probability as $n \rightarrow \infty$.*

Theorem 18 gives a sharp bound on the degrees for recovery, which we can compare to our spectral clustering results. The eigenvalues of D are $d_1 + d_2$ and $d_1 - d_2$, and the Markov matrix of the frame R has eigenvalues 1 and $(d_1 - d_2)/(d_1 + d_2)$. The diagonal blocks $L^{(11)}$ and $L^{(22)}$ each correspond to the Laplacian matrix of a d_1 -regular random graph on n vertices, whereas the off-diagonal block term $M^{(12)}$ corresponds to the Laplacian of a d_2 -regular bipartite graph on $2n$ vertices. Using our results and the previously known results for regular random graphs [Friedman, 2003, 2004, Bordenave, 2015], we can pick some $C > 2\sqrt{d_2 - 1}/d_2$ since $d_1 > d_2$ and we will eventually take the degrees to

be large. Using Proposition 16, we find that the spurious eigenvalues of P come after the leading 2 eigenvalues if

$$\frac{2\sqrt{d_2-1}}{d_2} < \frac{d_1-d_2}{d_1+d_2},$$

to leading order in the degrees. Rearranging, we obtain the condition

$$(d_1-d_2)^2 > 4(d_2-1) \left(\frac{d_1+d_2}{d_2} \right)^2.$$

Assuming $d_2/d_1 = \beta < 1$ fixed, and taking the limit $d_1, d_2 \rightarrow \infty$, we find that the Brito et al. result becomes

$$d_1 > 4 \frac{1+\beta}{(1-\beta)^2} + o(1),$$

whereas our result becomes

$$d_1 > \frac{4}{\beta} \left(\frac{1+\beta}{1-\beta} \right)^2 + o(1),$$

illustrating that the spectral threshold is a factor of $(1+\beta)/\beta$ weaker.

3.7 Application: Low density parity check or expander codes

Another useful application of random graphs is as expanders, loosely defined as graphs where the neighborhood of a small set of nodes is large. Expander codes, also called low density parity check (LDPC) codes, were first introduced by Gallager in his PhD thesis [Gallager, 1962]. These are a family of linear error correcting codes whose parity-check matrix is encoded in an expander graph. The performance of such codes depends on how good an expander that graph is, which in turn can be shown to depend on the separation of eigenvalues. For a good introduction and overview of the subject, see the book “Modern Coding Theory” by Richardson and Urbanke [2008].

Following Tanner [1981], we construct a code $\mathcal{C}$ from a (d_1, d_2) -regular bipartite graph G on $n+m$ vertices and two smaller linear codes $\mathcal{C}_1$ and $\mathcal{C}_2$ of length d_1 and d_2 , respectively. We write $\mathcal{C}_1 = [d_1, k_1, \delta_1]$ and $\mathcal{C}_2 = [d_2, k_2, \delta_2]$ with the usual convention of length, dimension, and minimum distance. We assume the codes are all binary, using the finite

field $\mathbb{F}_2$ the codeword is $x \in \mathcal{C} \subset \mathbb{F}_2^{|E|}$ where $|E| = nd_1 = md_2$. That is, we associate a bit to each edge in the graph bipartite graph G . Let $(e_i(v))_{i=1}^{d_v}$ represent the set of edges incident to a vertex v in some arbitrary, fixed order. Then the vector $x \in \mathcal{C}$ if and only if the vectors $(x_{e_1(u)}, x_{e_2(u)}, \dots, x_{e_{d_1}(u)})^T \in \mathcal{C}_1$ for all $u \in V_1$ and $(x_{e_1(v)}, x_{e_2(v)}, \dots, x_{e_{d_2}(v)})^T \in \mathcal{C}_2$ for all $v \in V_2$. The final code $\mathcal{C}$ is also linear. With this construction, the code $\mathcal{C}$ has rate at least $k_1/d_1 + k_2/d_2 - 1$ [Tanner, 1981].

Furthermore, Janwa and Lal [2003] proved the following bound on the minimum distance of the resulting code:

Theorem 19. *Suppose $\delta_1 \geq \delta_2 > \eta/2$. Then the code $\mathcal{C}$ has minimum distance*

$$\delta \geq \frac{n}{d_2} \left(\delta_1 \delta_2 - \frac{\eta}{2} (\delta_1 + \delta_2) \right),$$

where η is the second largest eigenvalue of the adjacency matrix of G .

Corollary 20. *Suppose the code $\mathcal{C}$ is constructed from a biregular, bipartite random graph $G \sim \mathcal{G}(n, m, d_1, d_2)$ and the conditions of Theorem 19 hold. Then the minimum distance of $\mathcal{C}$ satisfies*

$$\delta \geq \frac{n}{d_2} \left(\delta_1 \delta_2 - \frac{\sqrt{d_1 - 1} + \sqrt{d_2 - 1}}{2} (\delta_1 + \delta_2) - \epsilon_n \right).$$

We see that these Tanner codes will have maximal distance for smallest η , and used our main result, Corollary 4, to obtain the explicit bound in Corollary 20. By growing the graph, the above shows a way to construct arbitrarily large codes whose minimum distance remains proportional to the code size nd_1 . That is, the relative distance $\delta/(nd_1)$ is bounded away from zero as $n \rightarrow \infty$. However, the above bound will only be useful if it yields a positive result, which depends on the codes $\mathcal{C}_1$ and $\mathcal{C}_2$ as well as the degrees.

Remark. *In general, the performance guarantees on LDPC codes that are obtainable from graph eigenvalues are weaker than those that come from other methods. Although our method does guarantee high distance for some high degree codes, analysis of specific decoding algorithms or a probabilistic expander analyses yield better bounds that work for lower degrees [Richardson and Urbanke, 2008].*

3.7.1 Example: An unbalanced code based on a $(14, 9)$ -regular bipartite graph

We illustrate the applicability of our distance bound with an example. Let $\mathcal{C}_1 = [14, 8, 7]$ and $\mathcal{C}_2 = [9, 4, 6]$. These can be achieved by using a Reed-Salomon code on the common field $\mathbb{F}_q$ for any $q > 14$ [Richardson and Urbanke, 2008]. We take $q = 2^4 = 16$ for inputs that are actually binary, and this means each edge in the graph actually contains 4 bits of information. Employing Corollary 20, the Tanner code $\mathcal{C}$ will have relative minimum distance $\delta/(nd_1) \geq 0.0014$ and rate at least 0.016. Taking $n = 216$ and $m = 336$ gives the code a minimum distance of at least 4.

3.8 Application: Matrix completion

Assume we have some matrix $Y \in \mathbb{R}^{n \times m}$ which has low “complexity.” Perhaps it is low-rank or simple by some other measure. If we observe Y_{ij} for a limited set of entries $(i, j) \in E \subset [n] \times [m]$, then *matrix completion* is any method which constructs a matrix $\hat{Y}$ so that $\|\hat{Y} - Y\|$ is small, or even zero. Matrix completion has attracted significant attention in recent years as a tractable algorithm for making recommendations to users of online systems based on the tastes of other users (a.k.a. the Netflix problem). We can think of it as the matrix version of compressed sensing [Candes and Plan, 2010, Candès and Tao, 2010].

Recently, a number of authors have studied the performance of matrix completion algorithms where the index set E is the edge set of a regular random graph [Heiman et al., 2014, Bhojanapalli and Jain, 2014, Gamarnik et al., 2017]. Heiman et al. [2014] describe a *deterministic* method of matrix completion, where they can give performance guarantees for a fixed observation set E over many input matrices Y . The error of their reconstruction depends on the spectral gap of the graph. We expand upon the result of Heiman et al. [2014], extending it to rectangular matrix and improving their bounds in the process.

3.8.1 Matrix norms as measures of complexity and their relationships

We will employ a number of different matrix and vector norms in this Section. These are all related by the properties of the underlying Banach spaces. The complexity of Y is measured using a particular factorization norm:

$$\gamma_2(Y) = \min_{UV^*=Y} \|U\|_{\ell_2 \rightarrow \ell_\infty^n} \|V\|_{\ell_2 \rightarrow \ell_\infty^m}.$$

The minimum is taken over all possible factorizations of $Y = UV^*$, and the norm $\|X\|_{\ell_2 \rightarrow \ell_\infty^n} = \max_i \sqrt{\sum_j X_{ij}^2}$ returns the largest ℓ_2 norm of a row. So, equivalently,

$$\gamma_2(Y) = \min_{UV^*=Y} \max_{i,j} \|u_i\|_2 \|v_j\|_2,$$

where u_i and v_i are the rows of U and V . See [Linial et al., 2007] for a number of results about the norm γ_2 . In particular, note that

$$\frac{1}{\sqrt{nm}} \|Y\|_{\text{Tr}} \leq \gamma_2(Y) \leq \|Y\|_{\text{Tr}} \quad (3.23)$$

$$\gamma_2(Y) \leq \sqrt{\text{rank}(Y)} \|Y\|_\infty, \quad (3.24)$$

so we see that γ_2 is related to two common complexity measures of matrices, the trace norm (sum of singular values, i.e. the $\ell_2^m \rightarrow \ell_2^n$ nuclear norm) and rank [Candès and Tao, 2010]. Note also the well-known fact that

$$\|Y\|_{\text{Tr}} = \min_{UV^*=Y} \|U\|_F \|V\|_F,$$

where $\|X\|_F = \sqrt{\sum_{ij} X_{ij}^2}$ is the Frobenius norm. We see that the trace norm constrains factors U and F to be small on average via $\|\cdot\|_F$, whereas the norm γ_2 is similar but constrains factors uniformly via $\|\cdot\|_{\ell_2 \rightarrow \ell_\infty^n}$. However, we should note that computing $\gamma_2(Y)$ is more costly than the trace norm, which can be performed with just the singular value decomposition.

3.8.2 Matrix completion generalization bounds

The method of matrix completion that we study, following [Heiman et al. \[2014\]](#), is to return the matrix X which is the solution to:

$$\begin{aligned} & \underset{X}{\text{minimize}} \quad \gamma_2(X) \\ & \text{subject to} \quad X_{ij} = Y_{ij}, \quad (i, j) \in E. \end{aligned} \tag{3.25}$$

[Heiman et al. \[2014\]](#) analyze the performance of the convex program (3.25) for a square matrix Y using an expander argument, assuming that E is the edge set of a d -regular graph with second eigenvalue η . They obtain the following theorem:

Theorem 21 ([Heiman et al. \[2014\]](#)). *Let E be the set of edges of a d -regular graph with second eigenvalue bound η . For every $Y \in \mathbb{R}^{n \times n}$, if $\hat{Y}$ is the output of the optimization problem (3.25), then*

$$\frac{1}{n^2} \|\hat{Y} - Y\|_F^2 \leq c \gamma_2(Y)^2 \frac{\eta}{d},$$

where $c = 8K_G \leq 14.3$ is a universal constant and $\|\cdot\|_F$ is the Frobenius norm.

Considering rectangular matrices, we find a more general theorem which reduces to Theorem 21 if $n = m$ and $d_1 = d_2 = d$, but improved by a factor of two:

Theorem 22. *Let E be the set of edges of a (d_1, d_2) -regular graph with second eigenvalue bound η . For every $Y \in \mathbb{R}^{n \times m}$, if $\hat{Y}$ is the output of the optimization problem (3.25), then*

$$\frac{1}{nm} \|\hat{Y} - Y\|_F^2 \leq c \gamma_2(Y)^2 \frac{\eta}{\sqrt{d_1 d_2}},$$

where $c = 4K_G \leq 7.13$.

Proof. We start by considering a rank-1 sign matrix $S = uv^*$, where $u, v \in \{-1, 1\}^{n \times m}$. Let $S' = \frac{1}{2}(S + J)$, where J is the all-ones matrix, so that S' has the entries of -1 in S replaced by zeros. Then $S' = 1_A 1_B^* + 1_{A^c} 1_{B^c}^*$ for subsets $A \subset V_1 = [n]$ and $B \subset V_2 = [m]$, where

$A = \{i : u_i = 1\}$ and $B = \{j : v_j = 1\}$. Consider the expression

$$\begin{aligned}
 \left| \frac{1}{nm} \sum_{i,j} s_{ij} - \frac{1}{|E|} \sum_{(i,j) \in E} s_{ij} \right| &= \left| \frac{1}{nm} \sum_{i,j} (2s'_{ij} - 1) - \frac{1}{|E|} \sum_{(i,j) \in E} (2s'_{ij} - 1) \right| \\
 &= 2 \left| \frac{1}{nm} \sum_{i,j} s'_{ij} - \frac{1}{|E|} \sum_{(i,j) \in E} s'_{ij} \right| \\
 &= 2 \left| \frac{|A||B| + |A^c||B^c|}{nm} - \frac{E(A, B) + E(A^c, B^c)}{|E|} \right| \\
 &\leq 2 \left| \frac{|A||B|}{nm} - \frac{E(A, B)}{|E|} \right| + 2 \left| \frac{|A^c||B^c|}{nm} - \frac{E(A^c, B^c)}{|E|} \right|.
 \end{aligned}$$

The following is a bipartite version of the expander mixing lemma [De Winter et al., 2012]:

$$\left| \frac{E(A, B)}{|E|} - \frac{|A||B|}{nm} \right| \leq \frac{\eta}{\sqrt{d_1 d_2}} \sqrt{\frac{|A||B|}{nm} \left(1 - \frac{|A|}{n}\right) \left(1 - \frac{|B|}{m}\right)} = \frac{\eta}{\sqrt{d_1 d_2}} \sqrt{\frac{|A||B||A^c||B^c|}{(nm)^2}}.$$

We find that

$$\begin{aligned}
 \left| \frac{1}{nm} \sum_{i,j} s_{ij} - \frac{1}{|E|} \sum_{(i,j) \in E} s_{ij} \right| &\leq \frac{4\eta}{\sqrt{d_1 d_2}} \sqrt{\frac{|A||B||A^c||B^c|}{(nm)^2}} \\
 &= \frac{4\eta}{\sqrt{d_1 d_2}} \sqrt{xy(1-x)(1-y)} \\
 &\leq \frac{\eta}{\sqrt{d_1 d_2}},
 \end{aligned}$$

since $xy(1-x)(1-y)$ attains a maximal value of 2^{-4} for $0 \leq x, y \leq 1$. This improves on Theorem 21 by a factor of 2, because the version of the expander mixing lemma we used allowed us to combine both terms without approximation.

The rest of the proof develops identical to the results of Heiman et al. [2014], which we include for completion. Next, we apply the result to rank-1 sign matrices to any matrix R . Let $R = \sum_i \alpha_i S^i$, where S^i is a rank-1 sign matrix and $\alpha_i \in \mathbb{R}$. For a general matrix R , this might require many rank-1 sign matrices. Define the sign nuclear norm $\nu(R) = \sum_i |\alpha_i|$. Then,

$$\left| \frac{1}{nm} \sum_{i,j} r_{ij} - \frac{1}{|E|} \sum_{(i,j) \in E} r_{ij} \right| \leq \nu(R) \frac{\eta}{\sqrt{d_1 d_2}}.$$

It is a consequence of Grothendieck's inequality, a well-known theorem in functional analysis, that there exists a universal constant $1.5 \leq K_G \leq 1.8$ so that $\gamma_2(X) \leq \nu(X) \leq K_G \gamma_2(X)$ for any real matrix X [Heiman et al., 2014].

Now, let the matrix of residuals $R = (\hat{Y} - Y) \circ (\hat{Y} - Y)$, where $\circ$ is the Hadamard entry-wise product of two matrices, so that $R_{ij} = (\hat{Y}_{ij} - Y_{ij})^2$. Since

$$\frac{1}{|E|} \sum_{(i,j) \in E} r_{ij} = 0,$$

we conclude that

$$\frac{1}{nm} \sum_{i,j} r_{ij} \leq \nu(R) \frac{\eta}{\sqrt{d_1 d_2}} \leq K_G \gamma_2(R) \frac{\eta}{\sqrt{d_1 d_2}}.$$

Furthermore, $\gamma_2(R) \leq \gamma_2(\hat{Y} - Y)^2 \leq (\gamma_2(\hat{Y}) + \gamma_2(Y))^2$. Since $\hat{Y}$ is the output of the algorithm and Y is a feasible solution, $\gamma_2(\hat{Y}) \leq \gamma_2(Y)$. Thus, $\gamma_2(R) \leq 4\gamma_2(Y)^2$ and the proof is finished. $\square$

Remark. If we minimize the trace norm of the solution, which is a more practical method than working with γ_2 , the same bounds hold in terms of $\|Y\|_{\text{Tr}}$. This is because $\gamma_2(Y) \leq \|Y\|_{\text{Tr}}$. We only need to modify the final part of the proof.

3.8.3 Noisy matrix completion bounds

Furthermore, our analysis easily extends to the case where the matrix we observe is corrupted with noise. As mentioned in the above remark, similar results will hold for the trace norm. In the noisy case, we solve the problem

$$\begin{aligned} & \underset{X}{\text{minimize}} \quad \gamma_2(X) \\ & \text{subject to} \quad \frac{1}{|E|} \sum_{(i,j) \in E} (X_{ij} - Z_{ij})^2 \leq \delta^2 \end{aligned} \tag{3.26}$$

and obtain the following theorem:

Theorem 23. Suppose we observe $Z_{ij} = Y_{ij} + \epsilon_{ij}$ with bounded error

$$\frac{1}{|E|} \sum_{(i,j) \in E} \epsilon_{ij}^2 \leq \delta^2.$$

Then solving the optimization problem (3.26) will yield a bound of

$$\frac{1}{nm} \|\hat{Y} - Y\|_F^2 \leq c\gamma_2(Y)^2 \frac{\eta}{\sqrt{d_1 d_2}} + 4\delta^2,$$

where $c = 4K_G \leq 7.13$.

Proof. Denote $\hat{Y}$ the solution to P3.26. It will be useful to introduce the sampling operator $\mathcal{P}_E : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^{n \times m}$, where $(\mathcal{P}_E(X))_{ij} = X_{ij}$ if $(i, j) \in E$ and 0 otherwise. Following Heiman et al. [2014], let $R = (\hat{Y} - Y) \circ (\hat{Y} - Y)$ be the matrix of squared errors, then

$$\begin{aligned} \left| \frac{1}{nm} \|\hat{Y} - Y\|_F^2 - \frac{1}{|E|} \|\mathcal{P}_E(\hat{Y} - Y)\|_F^2 \right| &= \left| \frac{1}{nm} \sum_{i,j} (\hat{Y}_{ij} - Y_{ij})^2 - \frac{1}{|E|} \sum_{(i,j) \in E} (\hat{Y}_{ij} - Y_{ij})^2 \right| \\ &\leq K_G \gamma_2(R) \frac{\eta}{\sqrt{d_1 d_2}}. \end{aligned}$$

However, since Y is a feasible solution to P3.26, we have

$$\gamma_2(\hat{Y}) \leq \gamma_2(Y),$$

so that

$$\gamma_2(R) \leq (\gamma_2(\hat{Y} - Y))^2 \leq (\gamma_2(\hat{Y}) + \gamma_2(Y))^2 \leq 4\gamma_2(Y)^2.$$

By the triangle inequality

$$\|\mathcal{P}_E(\hat{Y} - Y)\|_F \leq \|\mathcal{P}_E(\hat{Y} - Z)\|_F + \|\mathcal{P}_E(Z - Y)\|_F \leq 2\delta\sqrt{|E|}$$

using the bounds on the distance of the solution to the data and on the noise. Because

$$\frac{1}{nm} \|\hat{Y} - Y\|_F^2 \leq \left| \frac{1}{nm} \|\hat{Y} - Y\|_F^2 - \frac{1}{|E|} \|\mathcal{P}_E(\hat{Y} - Y)\|_F^2 \right| + \frac{1}{|E|} \|\mathcal{P}_E(\hat{Y} - Y)\|_F^2$$

we get the final bound

$$\frac{1}{nm} \|\hat{Y} - Y\|_F^2 \leq 4K_G \gamma_2(Y)^2 \frac{\eta}{\sqrt{d_1 d_2}} + 4\delta^2.$$

□

3.8.4 Application of the spectral gap

Theorem 22 provides a bound on the mean squared error of the approximation X . Directly applying Corollary 4, we obtain the following bound on the generalization error of the algorithm using a random biregular, bipartite graph:

Corollary 24. *Let E be sampled from a $G(n, m, d_1, d_2)$ random graph. For every $Y \in \mathbb{R}^{n \times m}$, if $\hat{Y}$ is the output of the optimization problem (3.25), then*

$$\frac{1}{nm} \|\hat{Y} - Y\|_F^2 \leq c \gamma_2(Y)^2 \frac{\sqrt{d_1 - 1} + \sqrt{d_2 - 1} + \epsilon_n}{\sqrt{d_1 d_2}},$$

where $c = 4K_G \leq 7.13$ is a universal constant.

Bibliography

- Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, June 1986. ISSN 0209-9683, 1439-6912. doi: 10.1007/BF02579166.
- Omer Angel, Joel Friedman, and Shlomo Hoory. The non-backtracking spectrum of the universal cover of a graph. *Transactions of the American Mathematical Society*, 367(6):4287–4318, 2015. ISSN 0002-9947, 1088-6850. doi: 10.1090/S0002-9947-2014-06255-7.
- Wayne Barrett, Amanda Francis, and Benjamin Webb. Equitable decompositions of graphs with symmetries. *Linear Algebra and its Applications*, 513(Supplement C):409–434, January 2017. ISSN 0024-3795. doi: 10.1016/j.laa.2016.10.017.
- Paolo Barucca. Spectral partitioning in equitable graphs. *Physical Review E*, 95(6):062310, June 2017. doi: 10.1103/PhysRevE.95.062310.
- Hyman Bass. The Ihara-Selberg zeta function of a tree lattice. *International Journal of Mathematics*, 03(06):717–797, December 1992. ISSN 0129-167X. doi: 10.1142/S0129167X92000357.
- Edward A. Bender. The asymptotic number of non-negative integer matrices with given row and column sums. *Discrete Mathematics*, 10(2):217–223, January 1974. ISSN 0012-365X. doi: 10.1016/0012-365X(74)90118-6.
- Edward A Bender and E. Rodney Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, 24(3):296–307, May 1978. ISSN 0097-3165. doi: 10.1016/0097-3165(78)90059-6.

- Srinadh Bhojanapalli and Prateek Jain. Universal Matrix Completion. In *PMLR*, pages 1881–1889, January 2014.
- Béla Bollobás. A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs. *European Journal of Combinatorics*, 1(4):311–316, December 1980. ISSN 0195-6698. doi: 10.1016/S0195-6698(80)80030-8.
- Béla Bollobás. *Random Graphs*. Cambridge University Press, Cambridge, 2001. ISBN 978-0-511-81406-8.
- C. Bordenave, M. Lelarge, and L. Massoulié. Non-backtracking Spectrum of Random Graphs: Community Detection and Non-regular Ramanujan Graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1347–1357, October 2015. doi: 10.1109/FOCS.2015.86.
- Charles Bordenave. A new proof of Friedman’s second eigenvalue Theorem and its extension to random lifts. *arXiv:1502.04482 [math]*, February 2015.
- Charles Bordenave and Marc Lelarge. Resolvent of large random graphs. *Random Structures & Algorithms*, 37(3):332–352, October 2010. ISSN 10429832. doi: 10.1002/rsa.20313.
- Gerandy Brito, Ioana Dumitriu, Shirshendu Ganguly, Christopher Hoffman, and Linh V. Tran. Recovery and Rigidity in a Regular Stochastic Block Model. *arXiv:1507.00930 [math]*, July 2015.
- Gerandy Brito, Ioana Dumitriu, Shirshendu Ganguly, Christopher Hoffman, and Linh V. Tran. Recovery and Rigidity in a Regular Stochastic Block Model. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 1589–1601, Philadelphia, PA, USA, 2016. Society for Industrial and Applied Mathematics. ISBN 978-1-61197-433-1.
- E. J. Candes and Y. Plan. Matrix Completion With Noise. *Proceedings of the IEEE*, 98(6): 925–936, June 2010. ISSN 0018-9219. doi: 10.1109/JPROC.2009.2035722.
- E.J. Candès and T. Tao. The Power of Convex Relaxation: Near-Optimal Matrix Completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080, May 2010. ISSN 0018-9448. doi: 10.1109/TIT.2010.2044061.
- Stefaan De Winter, Jeroen Schillewaert, and Jacques Verstraete. Large Incidence-free Sets in Geometries. *The Electronic Journal of Combinatorics*, 19(4):P24, August 2012. ISSN 1077-8926.
- Ioana Dumitriu and Tobias Johnson. The Marčenko-Pastur law for sparse random bipartite biregular graphs. *Random Structures & Algorithms*, 48(2):313–340, March 2016. ISSN 1098-2418. doi: 10.1002/rsa.20581.

- Keqin Feng and Wen-Ch'ing Winnie Li. Spectra of Hypergraphs and Applications. *Journal of Number Theory*, 60(1):1–22, September 1996. ISSN 0022-314X. doi: 10.1006/jnth.1996.0109.
- Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3–5):75–174, February 2010. ISSN 0370-1573. doi: 10.1016/j.physrep.2009.11.002.
- Joel Friedman. A Proof of Alon's Second Eigenvalue Conjecture. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, pages 720–724, New York, NY, USA, 2003. ACM. ISBN 978-1-58113-674-6. doi: 10.1145/780542.780646.
- Joel Friedman. A proof of Alon's second eigenvalue conjecture and related problems. *arXiv:cs/0405020*, May 2004.
- R. Gallager. Low-density parity-check codes. *IRE Transactions on Information Theory*, 8(1): 21–28, January 1962. ISSN 0096-1000. doi: 10.1109/TIT.1962.1057683.
- David Gamarnik, Quan Li, and Hongyi Zhang. Matrix Completion from $O(n)$ Samples in Linear Time. *arXiv:1702.02267 [cs, math, stat]*, February 2017.
- C. D. Godsil and B. Mohar. Walk generating functions and spectral measures of infinite graphs. *Linear Algebra and its Applications*, 107(Supplement C):191–206, August 1988. ISSN 0024-3795. doi: 10.1016/0024-3795(88)90245-5.
- Eyal Heiman, Gideon Schechtman, and Adi Shraibman. Deterministic algorithms for matrix completion. *Random Structures & Algorithms*, 45(2):306–317, September 2014. ISSN 1098-2418. doi: 10.1002/rsa.20483.
- Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic block-models: First steps. *Social Networks*, 5(2):109–137, June 1983. ISSN 0378-8733. doi: 10.1016/0378-8733(83)90021-7.
- Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006. ISSN 0273-0979, 1088-9485. doi: 10.1090/S0273-0979-06-01126-8.
- H. Janwa and A. K. Lal. On Tanner Codes: Minimum Distance and Decoding. *Applicable Algebra in Engineering, Communication and Computing*, 13(5):335–347, February 2003. ISSN 0938-1279, 1432-0622. doi: 10.1007/s00200-003-0098-4.
- Motoko Kotani and Toshikazu Sunada. Zeta Functions of Finite Graphs. *Journal of mathematical sciences, the University of Tokyo*, 7(1):7–25, 2000. ISSN 13405705.
- Wen-Ch'ing Winnie Li and Patrick Solé. Spectra of Regular Graphs and Hypergraphs and Orthogonal Polynomials. *European Journal of Combinatorics*, 17(5):461–477, July 1996. ISSN 0195-6698. doi: 10.1006/eujc.1996.0040.

- Nati Linial, Shahar Mendelson, Gideon Schechtman, and Adi Shraibman. Complexity measures of sign matrices. *Combinatorica*, 27(4):439–463, July 2007. ISSN 0209-9683, 1439-6912. doi: 10.1007/s00493-007-2160-5.
- Eyal Lubetzky and Allan Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Mathematical Journal*, 153(3):475–510, June 2010. ISSN 0012-7094, 1547-7398. doi: 10.1215/00127094-2010-029.
- A. Marcus, D. A. Spielman, and N. Srivastava. Interlacing Families I: Bipartite Ramanujan Graphs of All Degrees. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 529–537, October 2013a. doi: 10.1109/FOCS.2013.63.
- Adam Marcus, Daniel A. Spielman, and Nikhil Srivastava. Interlacing Families II: Mixed Characteristic Polynomials and the Kadison-Singer Problem. *arXiv:1306.3969 [math]*, June 2013b.
- Brendan D. McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applications*, 40(Supplement C):203–216, October 1981a. ISSN 0024-3795. doi: 10.1016/0024-3795(81)90150-6.
- Brendan D McKay. Subgraphs of random graphs with specified degrees. *Congressus Numerantium*, 33:213–223, 1981b.
- Brendan D McKay. Asymptotics for 0-1 Matrices with Prescribed Line Sums. *Enumeration and Design*, pages 225–238, 1984.
- Hirobumi Mizuno and Iwao Sato. The semicircle law for semiregular bipartite graphs. *Journal of Combinatorial Theory, Series A*, 101(2):174–190, February 2003. ISSN 0097-3165. doi: 10.1016/S0097-3165(02)00010-9.
- M. E. J. Newman and Travis Martin. Equitable random graphs. *Physical Review E*, 90(5):052824, November 2014. doi: 10.1103/PhysRevE.90.052824.
- Mark Newman. *Networks: An Introduction*. Oxford University Press, March 2010. ISBN 978-0-19-150070-1.
- T. Richardson and R. Urbanke. The renaissance of Gallager’s low-density parity-check codes. *IEEE Communications Magazine*, 41(8):126–131, August 2003. ISSN 0163-6804. doi: 10.1109/MCOM.2003.1222728.
- Tom Richardson and Ruediger Urbanke. *Modern Coding Theory*. Cambridge University Press, New York, NY, USA, 2008. ISBN 978-0-521-85229-6.
- R. Tanner. A recursive approach to low complexity codes. *IEEE Transactions on Information Theory*, 27(5):533–547, September 1981. ISSN 0018-9448. doi: 10.1109/TIT.1981.1056404.

Nicholas C. Wormald. The asymptotic connectivity of labelled regular graphs. *Journal of Combinatorial Theory, Series B*, 31(2):156–167, October 1981. ISSN 0095-8956. doi: 10.1016/S0095-8956(81)80021-4.

INTERLUDE

The previous Chapter presents a model for a random graph with two communities, each of the bipartite sets. The results we develop for the two community case are then extended to state when community detection is possible with an arbitrary number of groups, along with applications in channel coding and machine learning.

We next consider a random graph model of the respiratory network. To model neural dynamics, we introduce a set of biophysical equations for the membrane potentials and channel openings of the various neurons, as well as synaptic currents. We study the effect of changing network connectivity on the resulting oscillatory dynamics. We show how splitting the network into two communities segments the oscillations into two phases, one for each sub-network, which are pushed apart by inhibition. We accompany the model with experiments performed with slices taken from the brainstem respiratory areas. Similarities and differences between the real neural recordings and model simulations lead to new hypotheses that increase our understanding of how the brain generates this important rhythm.

Chapter 4

DIFFERENT ROLES FOR INHIBITION IN THE RHYTHM-GENERATING RESPIRATORY NETWORK

**Kameron Decker Harris, Tatiana Dashevskiy, Joshua Mendoza, Alfredo J. Garcia III,
Jan-Marino Ramirez, and Eric Shea-Brown.
Journal of Neurophysiology 118(4), 2070-2088. 2017.**

Abstract

Unraveling the interplay of excitation and inhibition within rhythm-generating networks remains a fundamental issue in neuroscience. We use a biophysical model to investigate the different roles of local and long-range inhibition in the respiratory network, a key component of which is the pre-Bötzinger complex inspiratory microcircuit. Increasing inhibition within the microcircuit results in a limited number of out-of-phase neurons before rhythmicity and synchrony degenerate. Thus, unstructured local inhibition is destabilizing and cannot support the generation of more than one rhythm. A two-phase rhythm requires restructuring the network into two microcircuits coupled by long-range inhibition in the manner of a half-center. In this context, inhibition leads to greater stability of the two out-of-phase rhythms. We support our computational results with in vitro recordings from mouse pre-Bötzinger complex. Partial excitation block leads to increased rhythmic variability, but this recovers following blockade of inhibition. Our results support the idea that local inhibition in the pre-Bötzinger complex is present to allow for descending control of synchrony or robustness to adverse conditions like hypoxia. We conclude that the balance of inhibition and excitation determines the stability of rhythmogenesis, but with opposite roles within and between areas. These different inhibitory roles may apply to a variety of rhythmic behaviors that emerge in widespread pattern generating circuits of the nervous system.

4.1 Introduction

Rhythmic activity is critical for the generation of behaviors such as locomotion and respiration, as well as apparently non-rhythmic behaviors including olfaction, information processing, encoding, learning and memory [Marder and Bucher, 2001, Buzsaki, 2006, Kopell et al., 2010, Ainsworth et al., 2012, Skinner, 2012, Missaghi et al., 2016]. These rhythms arise from central pattern generators (CPGs), neuronal networks located within the central nervous system that are capable of generating periodic behavior due to their synaptic and intrinsic membrane properties [Marder and Bucher, 2001, Grillner, 2006, Grillner and Jessell, 2009, Kiehn, 2011].

An increasingly important concept is that a given behavior may involve the interaction between several rhythmogenic microcircuits [Anderson et al., 2016, Ramirez et al., 2016]. In the neocortex, multiple rhythms and mechanisms are involved in a variety of cortical processes [Buzsaki, 2006]. In breathing, which consists of the three dominant respiratory phases—inspiration, post-inspiration, and expiration—each phase seems to be generated by its own autonomous, excitatory microcircuit, sub-populations of the overall network which act as rhythm-generating modules [Anderson et al., 2016, Lindsey et al., 2012]. The timing between these excitatory microcircuits is established by inhibitory interactions. In locomotion, each side of the spinal cord contains rhythmogenic microcircuits that are similarly coordinated by inhibitory mechanisms in order to establish left-right alternation [e.g. Kiehn, 2011]. Assembling a behavior by combining different microcircuits may imbue a network with increased flexibility. This strategy could also facilitate the integration and synchronization of one rhythmic behavior with another. Sniffing, olfaction, whisking, and rhythmic activities in hippocampus and locus coeruleus are all rhythmically coupled to the inspiratory rhythm generated in the pre-Bötzinger complex (preBötC) [Sara, 2009, Moore et al., 2013, Ferguson et al., 2015, Ramirez et al., 2016, Huh et al., 2016]. This small microcircuit, located in the ventrolateral medulla, is the essential locus for the generation of breathing [Smith et al., 1991, Tan et al., 2008, Gray et al., 2001, Schwarzscher et al., 2011,

Ramirez et al., 1998].

First discovered a quarter of a century ago, the preBötC is among the best understood microcircuits [Smith et al., 1991]. It continues to generate fictive respiratory rhythm activity when isolated in vitro, reliant on excitatory neurotransmission. Rhythmicity in the preBötC ceases when glutamatergic synaptic mechanisms are blocked, while it persists following the blockade of synaptic inhibition. However, almost 50% of the preBötC neurons are inhibitory [Shao and Feldman, 1997, Winter et al., 2009, Morgado-Valle et al., 2010, Hayes et al., 2012]. Despite the abundance of inhibitory neurons, the majority of neurons in the preBötC are rhythmically active in phase with inspiration. A small group of approximately 9% of neurons in the preBötC are inhibited during inspiration and discharge in phase with expiration [Morgado-Valle et al., 2010, Nieto-Posadas et al., 2014, Carroll et al., 2013]. A recent optogenetic study by Sherman et al. [2015] showed that stimulation of glycinergic inhibitory preBötC neurons can delay or halt a breath, and inhibition of those neurons can increase the magnitude of a breath. This is consistent with pharmacological agonist-antagonist experiments by Janczewski et al. [2013] which found that inhibition can modulate rhythm frequency or trigger apnea but is not essential for rhythm generation. The inhibitory population may thus be an “actuator” that allows descending pathways to control respiration. However, with only a few studies available, the role of these inhibitory preBötC neurons is not well-understood.

These experimental findings raise important questions: What is the role of inhibitory neurons within this microcircuit [Cui et al., 2016]? Why does the preBötC generate primarily one rhythmic phase despite the presence of numerous inhibitory neurons? Our modeling study arrives at the conclusion that this microcircuit can only generate one rhythmic phase. Synaptic inhibition seems to primarily serve to titrate the strength of this single rhythm while creating a small number of apparently anomalous expiratory cells. In order to generate more than one phase, it is necessary to assemble a network where excitatory microcircuits are segmented, via inhibition, into different compartments. Mutually-inhibitory circuits have been proposed for the inspiration–active expiration net-

work [Smith et al., 2013, Molkov et al., 2013, Koizumi et al., 2013, Onimaru et al., 2015] and preBötC–post-inspiratory complex (PiCo) networks [Anderson et al., 2016].

The novelty of our theoretical study lies in two conceptually important findings: A single microcircuit is unable to generate more than one phase based on the currently known network structure, and the generation of different phases necessitates the inhibitory interaction between excitatory microcircuits. Based on these findings we propose that the generation of rhythm and phase arise from separate network-driven processes. In these two processes, inhibition plays fundamentally different roles: local inhibition promotes desynchronization within a microcircuit, while long-range inhibition establishes phase relationships between microcircuits. Consistent with our proposal is the observation that breathing does not depend on the presence of all three phases at any given time. In gasping and some reduced preparations, the respiratory network generates a one-phase rhythm consisting of inspiration only. Under resting conditions, breathing primarily oscillates between inspiration and post-inspiration. This eupneic rhythm also involves a late expiratory phase according to Richter and Smith [2014]. Under high metabolic demand or coughing, another phase is recruited in form of active expiration. This modular organization may be a fundamental property of rhythm generating networks.

4.2 Materials and Methods

4.2.1 preBötC network simulations

We model the preBötC network as a simple directed Erdős-Rényi random graph on $N = 300$ nodes, where edges are added at random with fixed probability. We denote a directed edge from node j to node i as $j \rightarrow i$. The connection probability $p = (k_{\text{avg}}/2)/(N - 1)$ so that the expected total degree, that is the in-degree plus the out-degree, of a node is k_{avg} , which we vary. We prefer to parametrize these networks by degree k_{avg} rather than p , since in this case our results do not depend on N once it is large [Bollobás, 1998].

Each node is of type bursting (B), tonic spiking (TS), or quiescent (Q) with correspond-

ing probabilities 25%, 45%, and 30% [Peña et al., 2004, Del Negro et al., 2005]. Neurons are inhibitory with probability p_I , another parameter, and all projections from an inhibitory neuron are inhibitory. The sets of excitatory and inhibitory nodes are denoted $\mathcal{N}_E$ and $\mathcal{N}_I$. Edges are assigned a maximal conductivity g_E for excitatory connections and g_I for inhibitory connections. In our parameter sweeps, we vary these conductivities over the range 2–5 nS. This matches the postsynaptic potential deflections observed in experiments (typical IPSPs: -1.2 to -1.8 mV, EPSPs: 1.6 to 2.3 mV; data from Aguan Wei).

We use “model 1” from Butera et al. [1999a] as the dynamical equations for bursting, tonic spiking, and quiescent neurons. All parameters, given in Table 4.1, are shared among the dynamical types with the exception of the leak conductance g_L which is adjusted for the desired dynamics (B, TS, Q). Parameter values besides g_L are taken from Park and Rubin [2013], most of which are the same or close to the original values chosen by Butera et al. [1999a]. With the chosen parameters, the bursting neurons fire 6-spike bursts every 2.4 s, and the tonic spikers fire 3.5 spikes per second.

The full system of equations is

$$\begin{aligned}\dot{V} &= -(I_L + I_{Na} + I_K + I_{Na,p} + I_{syn} - I_{app}) / C \\ \dot{h} &= (h_\infty(V) - h) / \tau_h(V) \\ \dot{n} &= (n_\infty(V) - n) / \tau_n(V)\end{aligned}\tag{4.1}$$

with currents calculated as

$$\begin{aligned}I_L &= g_L(V - E_L) \\ I_{Na} &= g_{Na}m_\infty^3(V)(1 - n)(V - E_{Na}) \\ I_K &= g_Kn^4(V - E_K) \\ I_{Na,p} &= g_{Na,p}m_{p,\infty}(V)h(V - E_{Na}),\end{aligned}$$

and the activation and time constants are

$$\begin{aligned}x_\infty(V) &= \frac{1}{1 + \exp((V - \theta_x)/\sigma_x)} \\ \tau_x(V) &= \frac{\bar{\tau}_x}{\cosh((V - \theta_x)/(2\sigma_x))}.\end{aligned}$$

Parameter	Value
C	21 pF
E_{Na}	50 mV
E_{K}	-85 mV
E_{L}	-58 mV
θ_m	-34 mV
θ_n	-29 mV
$\theta_{m,p}$	-40 mV
θ_h	-48 mV
σ_m	-5 mV
σ_n	-4 mV
$\sigma_{m,p}$	-6 mV
σ_h	5 mV
$\bar{\tau}_n$	10 ms
$\bar{\tau}_h$	10,000 ms
g_{K}	11.2 nS
g_{Na}	28 nS
$g_{\text{Na},p}$	1 nS
I_{app}	0 pA
$g_L^{(\text{B})}$	1.0 nS
$g_L^{(\text{TS})}$	0.8 nS
$g_L^{(\text{Q})}$	1.285 nS
$E_{\text{syn,E}}$	0 mV
$E_{\text{syn,I}}$	-70 mV
θ_{syn}	0 mV
σ_{syn}	-3 mV
$\bar{\tau}_{\text{syn}}$	15 ms

Table 4.1: Parameters for the network model are taken from the literature [Butera et al., 1999a, Park and Rubin, 2013]. We modify g_L for quiescent (Q), tonic spiking (TS), and intrinsically bursting (B) cells. The system of equations is simulated in the given units, so that no conversions are necessary. Those parameters below the lower horizontal break are for the synaptic dynamics.

To model network interactions, we model synaptic dynamics with first-order kinetics [Destexhe et al., 1994]. The synaptic current neuron i receives is

$$I_{\text{syn},i} = \sum_{j \in \mathcal{N}_E: j \rightarrow i} g_E s_{ij} (V_i - E_{\text{syn},E}) + \sum_{j \in \mathcal{N}_I: j \rightarrow i} g_I s_{ij} (V_i - E_{\text{syn},I}),$$

where g_E and g_I are the maximal excitatory and inhibitory synapse conductances. The reversal potentials $E_{\text{syn},E}$ and $E_{\text{syn},I}$ for excitatory and inhibitory synapses, shown in Table 4.1, correspond the appropriate values for glutamatergic and glycinergic or GABAergic synapses. The variables s_{ij} represent the open fraction of channels between cells j and i , and they are governed by the differential equations

$$\begin{aligned} \dot{s}_{ij} &= \left((1 - s_{ij}) m_{\infty}^{(ij)}(V_j) - s_{ij} \right) / \tau_{\text{syn}} \\ m_{\infty}^{(ij)}(V_j) &= \frac{1}{1 + \exp((V_j - \theta_{\text{syn}}) / \sigma_{\text{syn}})}. \end{aligned}$$

Excitatory and inhibitory synapses share the parameters $\bar{\tau}_{\text{syn}}$, θ_{syn} , and σ_{syn} (Table 4.1).

Each model run starts from random initial conditions and lasts 100 s of simulation time with 1 ms time resolution. The first 20 s of transient dynamics are removed before postprocessing. Rather than save all state variables during long runs, we record a binary variable for each neuron that indicates whether or not the neuron fires a spike in the given time step. A spike is registered when V surpasses -15 mV for the first time in the previous 6 ms. This spike raster is then stored as a sparse matrix. The simulation code is configurable to output voltage traces or all state variables; these were examined during development to check that the model and spike detection function correctly.

We examine the effects of network connectivity, inhibition, and synaptic strength on the dynamics of our model by varying k_{avg} , p_I , g_E , and g_I . To capture the interactions of these parameters, we sweep through all combinations of parameters in the ranges $k_{\text{avg}} = 1.0, 1.5, \dots, 12.0$; $p_I = 0.00, 0.05, \dots, 1.00$; $g_E = 2, 3, \dots, 5$ nS; and $g_I = 2, 3, \dots, 5$ nS, with 8 repetitions of each combination. The only randomness in the model is randomness present in the graphs and initial conditions, since the dynamics are deterministic. This amounts to 61,824 graph generation, simulation, and postprocessing steps. Network

generation, simulations, and postprocessing were performed with custom code available from the first author at <http://github.com/kharris/prebotc-graph-model>. The code was written in Python and C++, and some analysis was performed with MATLAB. Numerical integration used backwards differentiation formulae in VODE called via `scipy.integrate.ode`, suitable for stiff equation systems. We experimented with the tolerance to be sure it resolves all timescales. We used the Hyak cluster at the University of Washington to conduct parameter sweeps. Each simulated 100 s took less than 3 hours and could be performed on a standard consumer machine.

4.2.2 *Two population network model*

The preBötC is thought to be connected to another microcircuit, alternately the BötC, PiCo, and lateral parafacial group, in a mutually inhibitory manner [Smith et al., 2013, Molkov et al., 2013, Huckstepp et al., 2016, Anderson et al., 2016] which allows them to generate stable two-phase rhythms as in a half-center oscillator [Marder and Bucher, 2001]. We study this case with a two microcircuit model, a where each microcircuit is represented by a different population of cells (Pop. 1 and Pop. 2); we arbitrarily refer to the preBötC as Pop. 1.

We use a two group stochastic block model for the network. The stochastic block model [Holland et al., 1983] is a generalization of the directed Erdős-Rényi random graph, where the connection probability varies depending on the population label of each neuron. Each population has recurrent connections from excitatory to all other cells, with each connection occurring with a fixed probability. As we describe below, we vary probabilities of connections from inhibitory neurons to other neurons in the same population (intra-group) and in the other population (inter-group).

Let N_1 be the number of neurons in Pop. 1 and N_2 be the number of neurons in Pop. 2. We assume $N_1 = N_2 = 300$, so the network has a total of 600 neurons. To generate this network we begin by assigning each neuron to one of the two populations. We then

assign each neuron a type: quiescent, tonic or bursting, using the same method as the single population model. Afterwards, we randomly assign neurons to be inhibitory with probability $p_I = 0.5$ [Shao and Feldman, 1997, Winter et al., 2009, Hayes et al., 2012, Morgado-Valle et al., 2010]; otherwise they are excitatory. We then assign connections to the neurons with probabilities:

$$P^{(I)} = \begin{bmatrix} \frac{k_{\text{intra}}}{N_1-1} & \frac{k_{\text{inter}}}{N_2} \\ \frac{k_{\text{inter}}}{N_1} & \frac{k_{\text{intra}}}{N_2-1} \end{bmatrix}, \quad P^{(E)} = \begin{bmatrix} \frac{3}{N_1-1} & 0 \\ 0 & \frac{3}{N_2-1} \end{bmatrix},$$

where $0 \leq k_{\text{intra}}, k_{\text{inter}} \leq 4$. The matrix entries (i, j) are the probability of a connection between an inhibitory or excitatory neuron in population i to a neuron in population j . This model allows us to tune between a half-center network containing only inter-group inhibition and a network with equal amounts of both intra- and inter-group inhibition.

The matrix $P^{(E)}$ contains the probability of connection for a projecting excitatory neuron. It is diagonal, reflecting the assumption that excitatory neurons only project within the local population, and each excitatory neuron has an average out-degree of 3. The matrix $P^{(I)}$ describes the probability of connection for inhibitory projecting neurons. The variable k_{intra} is the expected number of projections per inhibitory neuron to other neurons within its own population, and k_{inter} is the expected number of projections from an inhibitory neuron to neurons in the other population. We normalize these values in the matrix to ensure that the average in-degree is the sum of the columns and and out-degree is the sum of the rows, both equal to $k_{\text{intra}} + k_{\text{inter}} + 3$. The total inhibitory degrees then depend on the values of k_{intra} and k_{inter} , which affect only the inhibitory connection probabilities. Unless explicitly stated, connections are assigned a fixed conductance of $g_E = g_I = 2.5 \text{ nS}$ for excitatory and inhibitory connections.

We examine the effects of inhibition both within a population and between populations. To do this, we sweep through the parameters $k_{\text{intra}}, k_{\text{inter}} = 0.0, 0.5, \dots, 4.0$ and simulate 8 realizations (i.e., samples from the distribution of random graphs with these parameters) for each parameter pair. This leads to 648 graph generation, simulation,

and post processing steps. As for the single population model, all code is available at <http://github.com/kharris/prebotc-graph-model>.

4.2.3 *Slice experiments*

Brainstem transverse slices were prepared from CD1 mice (P7–12). All experiments were performed with the approval of the Institute of Animal Care and Use Committee of the Seattle Children’s Research Institute. Mice were maintained with rodent diet and water available ad libitum in a vivarium with a 12 h light/dark cycle at 22°C. Thickness of slices containing the preBötC varied between 550–650 μm . Slices were placed into the recording chamber with circulating artificial cerebrospinal fluid (aCSF) containing NaCl 118 mM, KCl 3 mM, CaCl_2 1.5 mM, MgCl 1 mM, NaHCO_3 25 mM, NaH_2PO_4 1 mM, d-glucose 30 mM and equilibrated with 95% O_2 and 5% CO_2 , pH 7.4. We maintained the temperature of the bath at 31°C, with an aCSF circulation rate of 15 mL/min. Rhythmic activity of preBötC was induced by slow up-regulation of KCl concentration from 3 mM to 8 mM in aCSF. The details of the technique are described in Ramirez et al. [1997a] and Anderson et al. [2016].

We recorded extracellular neuronal population activity in the preBötC region with a protocol that first measured the control activity, then activity following application of a partial excitation block, and finally with an additional complete block of inhibition. We used 700 nM DNQX disodium salt, a selective non-NMDA receptor antagonist which blocks glutamatergic ion channels generating fast excitatory synaptic inputs, to effect the partial excitation block. Picrotoxin (PTX), an ionotropic GABA_A receptor antagonist blocking inhibitory chloride-selective channels, was used at 20 or 50 μM to shut down inhibition. Both concentrations of PTX were equally effective at blocking inhibition. DNQX disodium salt and PTX were obtained from Sigma-Aldrich, St. Louis, MO. After application of either drug, we waited 5 min for the drugs to take effect and used at least 10 min of data to measure the resulting rhythm.

In additional experiments, we supplemented the extracellular population-level data with multi-electrode recordings in the contralateral preBötC. Extracellular neural activity from the transverse medullary slice was recorded on a 16 channel commercial linear multi-array electrode (model: Brain Slice Probe, Plexon, Dallas, TX). Each electrode had a recording surface of 15 microns and interelectrode spacing was fixed at 50 microns. Neural signals were amplified and recorded using the Omni-Plex D system (Plexon). Wide-band data was filtered with a Butterworth lowpass filter, 200 Hz cutoff, and spike sorting was performed offline and post-hoc using Offline Sorter v4.1.0 (Plexon). Specifically, individual unit waveforms were detected and sorted using principle component analysis, visualized in a three-dimensional cluster view. Waveforms were detected and sorted using Offline Sorter with manual cluster cutting single electrode-based feature spaces. Care was taken to follow nonstationarities in waveform shapes in assigning spikes to separate units, and auto- and cross-correlation histograms were examined as a check on sorting results [Lewicki, 1998]. All neurons with good isolation were kept for analysis.

We kept only those slices that initially showed robust rhythms, as determined by the experimentalist. We performed a total of 5 multi-electrode experiments and discarded one in which the rhythm went away after application of DNQX and never recovered. We recorded extracellularly from 15 slices and excluded 2 outliers from statistical analysis, because their rhythms slowed considerably more than the others with DNQX. In vitro slice data were analyzed by hand using Axon pClamp (Molecular Devices, Sunnyvale, CA) to extract burst locations and amplitudes, which were exported to a table for analysis using custom Python programs available at <http://github.com/kharris/prebotc-graph-model>.

4.2.4 *Postprocessing*

Because of the large number of simulations needed to explore the parameter space, we can examine only a small fraction of the simulations by eye and must rely on summary

statistics to characterize the dynamics.

Binning and filtering

First, the spike raster data is aggregated into 50 ms bins of spike counts to compress the size of the matrix. We denote the spike raster vector timeseries $\mathbf{x}^{\text{bin}}(t)$. The unbinned spike rasters are then convolved with a Gaussian kernel $k(t) = \left(\sigma\sqrt{2\pi}\right)^{-1} \exp\left(-\frac{1}{2}t^2/\sigma^2\right)$, where $\sigma = 60$ ms, to produce the continuous timeseries $\mathbf{x}^{\text{filt}}(t) = (k * \mathbf{x})(t)$, which is then downsampled to the same time bins. To characterize the overall population output, we compute what we call the integrated trace $x^{\text{int}}(t)$. This is defined as the lowpass-filtered population average, where the population average $\bar{x}(t) = \frac{1}{N} \sum_{i=1}^N x_i(t)$. We use a second-order Butterworth filter with cutoff frequency 4 Hz. The integrated trace is normalized to have units of spikes per second per neuron.

Synchrony statistic

Our principle aim is to quantify how different networks give rise to varying degrees of synchrony across the population of bursting neurons. We choose to characterize the overall synchrony of the population with one statistic [Golomb, 2007, Masuda and Aihara, 2004]

$$\chi = \left(\frac{\langle \bar{x}^{\text{filt}}(t)^2 \rangle_t - \langle \bar{x}^{\text{filt}}(t) \rangle_t^2}{\frac{1}{N} \sum_{i=1}^N [\langle x_i^{\text{filt}}(t)^2 \rangle_t - \langle x_i^{\text{filt}}(t) \rangle_t^2]} \right)^{1/2} \quad (4.2)$$

where the angle brackets $\langle \cdot \rangle_t$ denote averaging over the timeseries and $\bar{x}^{\text{filt}}(t) = \frac{1}{N} \sum_{i=1}^N x_i^{\text{filt}}(t)$. The value of χ is between 0 and 1. With perfect synchrony, $x_i^{\text{filt}}(t) = \bar{x}^{\text{filt}}(t)$ for all i , then we will find $\chi = 1$. With uncorrelated signals $x_i^{\text{filt}}(t)$, then $\chi = 0$. Examples of network activity for different values of χ are shown in Fig. 4.1.

Burst detection and phase analysis

The respiratory rhythm is generated by synchronized bursts of activity in the preBötC. In order to identify these bursts in the integrated traces, we needed a method of peak-detection that identifies large bursts but ignores smaller fluctuations. To do this we identify times t^* in the integrated timeseries $x^{\text{int}}(t)$, where $x^{\text{int}}(t^*)$ is an absolute maxima over a window of size 600 ms (12 time bins to either side of the identified maximum), and its value is above the 75th percentile of the full integrated timeseries. This ensures that the detected bursts are large-amplitude, reliable maxima of the timeseries.

Using the detected burst peak times $t_1^*, t_2^*, \dots, t_{n_{\text{bursts}}}^*$, we can examine the activity of individual neurons triggered on those events, the burst triggered average (BTA). The time between consecutive bursts is irregular, so in order to compute averages over many events, we rescale time into a uniform phase variable $\phi \in [-\pi, \pi]$. A phase $\phi = 0$ happens at the population burst, while $\phi = -\pi \equiv \pi \pmod{2\pi}$ occurs in-between bursts. To define this phase variable, we rescale the half-interval $[(t_n^* - t_{n-1}^*)/2, t_n^*]$ preceding burst n to $[-\pi, 0]$. Similarly, we rescale the other half-interval $[t_n^*, (t_{n+1}^* - t_n^*)/2]$ which follows burst n to $[0, \pi]$. Each rescaling is done using linear interpolation of the binned spike rasters. Let $\Phi(t)$ denote the mapping from time t to the phase. Then the BTA activity of neuron i is

$$x_i^{\text{BTA}}(\phi) = \frac{1}{n_{\text{bursts}}} \sum_{j=1}^{n_{\text{bursts}}} \int_{-(t_j^* - t_{j-1}^*)/2}^{(t_{j+1}^* - t_j^*)/2} x_i^{\text{filt}}(t_j^* + t) \delta(\Phi(t_j^* + t) - \phi) dt, \quad (4.3)$$

where $\delta(\cdot)$ is the Dirac delta measure which ensures that x_i^{filt} is sampled at the correct phase.

The BTAs exhibit two characteristic shapes. The first shape is peaked at a particular value of ϕ ; these neurons are phasic bursters. Of course, most phasic bursters take part in the overall population rhythm and have their BTA maximum near zero. Cells that are in-phase with the population rhythm are *inspiratory*. However, there are some bursters with a BTA peak near π , and we call these out-of-phase cells *expiratory*. The second shape is weakly peaked or flat; these neurons are *tonic*.

We define a complex-valued *phase-locking variable* z_i as the circular average of the BTA normalized by its integral:

$$z_i = \frac{\int_{-\pi}^{\pi} x_i^{\text{BTA}}(\phi) e^{i\phi} d\phi}{\int_{-\pi}^{\pi} x_i^{\text{BTA}}(\phi) d\phi}. \quad (4.4)$$

Normalization allows us to compare cells with different firing rates. The magnitude of phase-locking (peakedness of x_i^{BTA}) is quantified by the magnitude $|z_i|$. We use the argument $\arg(z_i)$ to define the dominant phase of a cell's activity. These phase-locking variables are similar to the order parameters used to study synchrony [Arenas et al., 2008]. We classify cell i as inspiratory, expiratory, tonic, or silent by:

1. Silent: firing rate is less than 0.1 Hz,
2. Inspiratory: $|z_i| > 0.2$ and $|\arg(z_i)| \leq \pi/2$,
3. Expiratory: $|z_i| > 0.2$ and $|\arg(z_i)| > \pi/2$,
4. Tonic: otherwise.

Two population phase analysis

For the two microcircuit model, we are also interested in the phase relationship between the two populations. To study this, we examine the burst-by-burst phase differences between the two populations' integrated traces and extract descriptive statistics of the phase differences. The N_1 neurons in Pop. 1 and N_2 neurons in Pop. 2 define two separate groups that we analyze as in Sections 4.2.4, 4.2.4, and 4.2.4. Note that because of the symmetry of $P^{(E)}$ and $P^{(I)}$, Pop. 1 and Pop. 2 are statistically equivalent. The burst times define two vectors t^{1*} and t^{2*} , where t_j^{i*} is the time for the j th peak in the signal of population $i = 1$ or 2. Pop. 1 is set as the reference signal for phase analysis. We then define a window with respect to the reference as $W_j = [t_j^{i*}, t_{j+1}^{i*}]$, where i is the chosen reference signal. For each peak ℓ in the non-reference signal, which we write as $t_{\ell}^{\bar{i}*}$, we find the reference window W_j so that $t_{\ell}^{\bar{i}*} \in W_j$. In other words, for each peak in the non-reference signal we find

the two peaks it lies between in the reference signal; we say that these peaks delineate the reference window. Once we have the reference window to use for the given peak, we define the phase difference between the two signals as $\theta_i = \frac{t_{j+1}^{i*} - t_\ell^{i*}}{t_{j+1}^{i*} - t_j^{i*}} \in [0, 1]$.

For an accurate description of the overall phase difference between the signals, we use directional statistics [Jammalamadaka and SenGupta, 2001], which account for the fact that $\theta = 0$ and 1 are identified. We can imagine that each phase difference is mapped to a circle, where we can then calculate the average position of those phase differences and how spread out the values are on that circle with respect to that average. To do this, we map the θ_i onto the unit circle using the equation $\zeta_k = e^{2\pi i \theta_k}$. We then take the average of these complex-valued points, $\zeta_{\text{avg}} = \frac{1}{n} \sum_{k=1}^n \zeta_k$.

We next calculate two quantities: the average phase difference $\Phi = \arg(\zeta_{\text{avg}})/(2\pi)$ and the phase order $\Omega = |\zeta_{\text{avg}}|$. The average phase difference Φ is the circular average of the peak-by-peak phase difference between the two signals through time. The phase order Ω tells us how concentrated the phase differences are compared to the average. If $\zeta_k \approx \zeta_{\text{avg}}$ for all k , then $|\zeta_{\text{avg}}| \approx 1$. However, if the values of ζ_k are uniformly spread around the unit circle, we would have a $\zeta_{\text{avg}} \approx 0$, since opposite phases cancel out. Thus, the phase order $0 \leq \Omega \leq 1$, and the closer it is to one, the more reliable the phase difference is between the two rhythms over time.

Irregularity scores

We define the irregularity score of sequence x_j as

$$\text{IRS}(x) = \frac{1}{n_{\text{bursts}}} \sum_{j=1}^{n_{\text{bursts}}} \frac{|x_{j+1} - x_j|}{|x_j|}. \quad (4.5)$$

Here, x_j denotes either the amplitude of the j th detected burst (amplitude irregularity) or the period between bursts j and $j + 1$ (period irregularity). The irregularity score $\text{IRS}(x)$ measures the average relative change in x .

Statistical tests

We analyzed the amplitude, period, amplitude irregularity, and period irregularity using a linear mixed effects model. This model captures the repeated measurement structure inherent in our experimental design. In particular, we model the response (amplitude, period, etc.) $y_{s,d}$ of a slice s to drug d as

$$y_{s,d} = a + a_s + \mu_d + \epsilon_{s,d},$$

where a is a fixed intercept (representing the control level of y), a_s is a zero-mean random effect for each slice, μ_d is a fixed effect for each drug (DNQX or DNQX+PTX), and $\epsilon_{s,d}$ is a zero-mean noise term. We fit this model using the `lmerTest` package in R, and the code and data used for fitting and analysis are provided in the Data Supplement. In the results we report the estimate of the fixed effects (a , μ_d), standard error (SE), degrees of freedom (DF), t value, and p value.

4.3 Results

We developed a network model of the preBötC and used this to examine the impact of connectivity and inhibition. Each cell in the network is governed by membrane currents that can produce square wave bursting via the persistent sodium current $I_{Na,p}$ [Butera et al., 1999a]. We include bursting pacemaker (B), tonic spiking (TS), and quiescent (Q) cell types in realistic proportions. Through simulations, we examine the effects of network connectivity and the presence of inhibitory cells on rhythm generation. To achieve this, we vary three key parameters over their biologically plausible ranges: (1) the fraction of inhibitory cells p_I , (2) the average total degree k_{avg} , i.e. the average total incoming and outgoing connections incident to a neuron, and (3) excitatory and inhibitory maximal synaptic conductances g_E and g_I . The parameter k_{avg} controls the sparsity of synaptic connections present in the network; as k_{avg} increases, the network becomes increasingly connected.

As we detail below, we compute metrics of synchronous bursting within the microcircuit as these network parameters vary. We then generalize the model to two coupled microcircuits and test whether the added network structure can generate multi-phase rhythms. Finally, we also compare these model effects to experiments with preBötC slice preparations, where we use a pharmacological approach to modulate the efficacy of excitatory and inhibitory synapses.

4.3.1 *Inhibition and sparsity weaken the model rhythm*

We first fix a moderate level of network sparsity, so that each cell receives and sends a total of $k_{\text{avg}} = 6$ connections on average, and we also fix the synaptic strengths (g_E and $g_I = 2.0$ nS). In Fig. 4.1, we show the behavior of the network for varying amounts of inhibitory cells p_I .

In Fig. 4.1A, the inhibitory fraction $p_I = 0$, so the network is purely excitatory. In this case it generates a strong, regular rhythm, and the population is highly synchronized. This is clear from both the integrated trace x^{int} , which captures the network average activity and thus the rhythm (defined in Section 4.2.4), and the individual neuron spikes in a raster, which are clearly aligned and periodic across many cells in the microcircuit. To further quantify the levels of synchronized firing, we use the synchrony measure χ , a normalized measure of the individual neuron correlations to the population rhythm, formally defined in Eqn. (4.2). Values of $\chi \approx 1$ reflect a highly-synchronized population, whereas $\chi \approx 0$ means the population is desynchronized. The cells in panel A are visibly synchronized from the raster, and have synchrony $\chi = 0.88$.

We introduce a greater fraction of inhibitory cells $p_I = 0.2$ in panel B. Here, we see more irregularity in the population rhythm as well as reduced burst amplitude and synchrony ($\chi = 0.72$). In panel C, with a still greater fraction of inhibitory cells, $p_I = 0.4$, the network shows further reduced synchrony ($\chi = 0.28$) and a very irregular, weak rhythm. In this case, the “rhythm” is extremely weak, if it even can be said to exist at all,

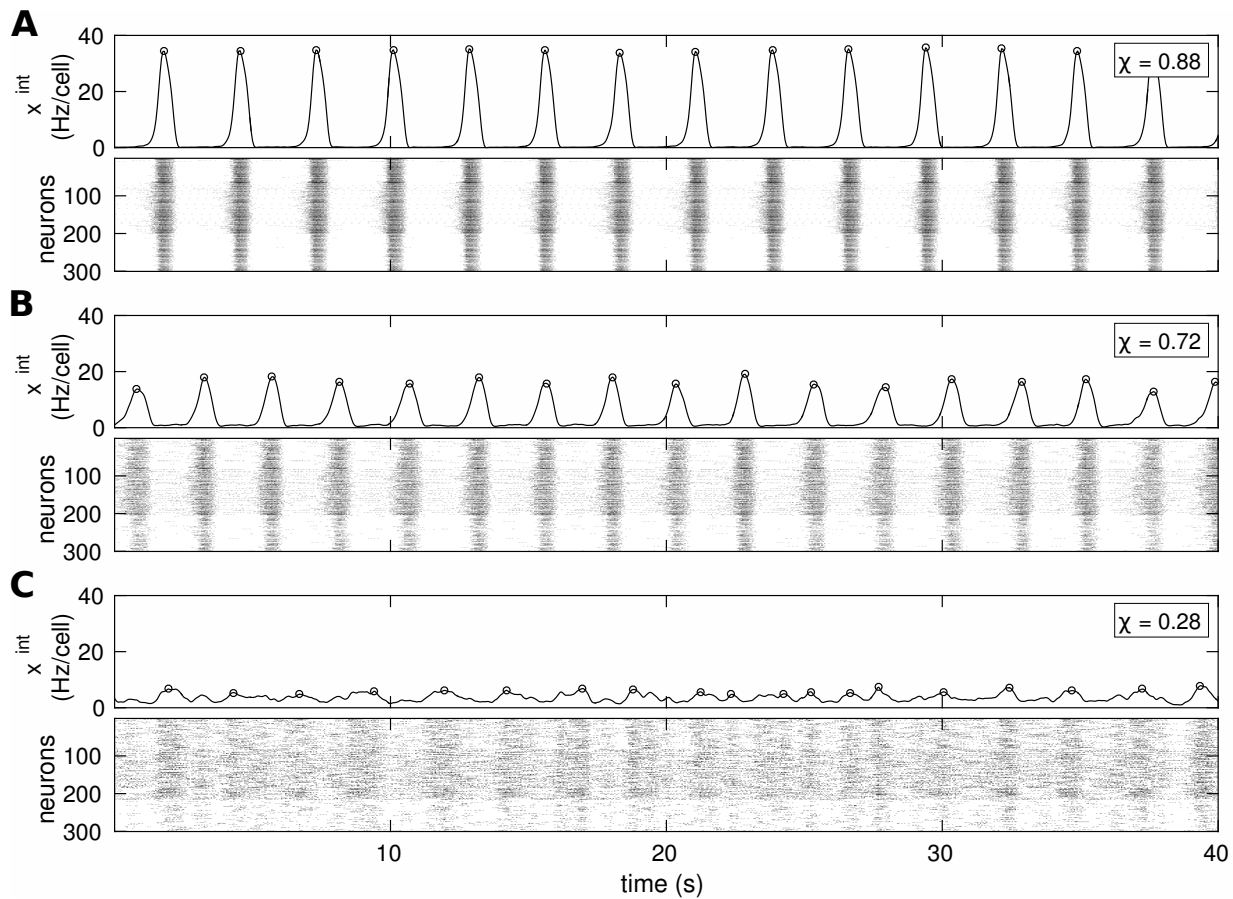


Figure 4.1: **With higher fraction of inhibitory cells, synchrony and burst amplitude decrease, and the integrated timeseries becomes more variable.** Three simulations of the respiratory network model: **A**, $p_I = 0\%$; **B**, $p_I = 20\%$; **C**, $p_I = 40\%$. Above, we show the integrated trace, which is a lowpass-filtered average of the spiking activity of all $N = 300$ neurons in the network. Below, we show the spike raster of individual neuron activity. In all cases, $k_{\text{avg}} = 6$, $g_E = g_I = 2.0$ nS. Detected bursts are marked by open circles on the integrated traces. At lower levels of synchrony, as in part C, what constitutes a burst becomes ambiguous.

and could not drive healthy breathing.

Building on these three examples, we next studied the impact of inhibition on synchrony over a wider range of network connectivity parameters. Here, we vary not only the fraction of inhibitory cells p_I , but also the sparsity via k_{avg} . In Fig. 4.2, we summarize

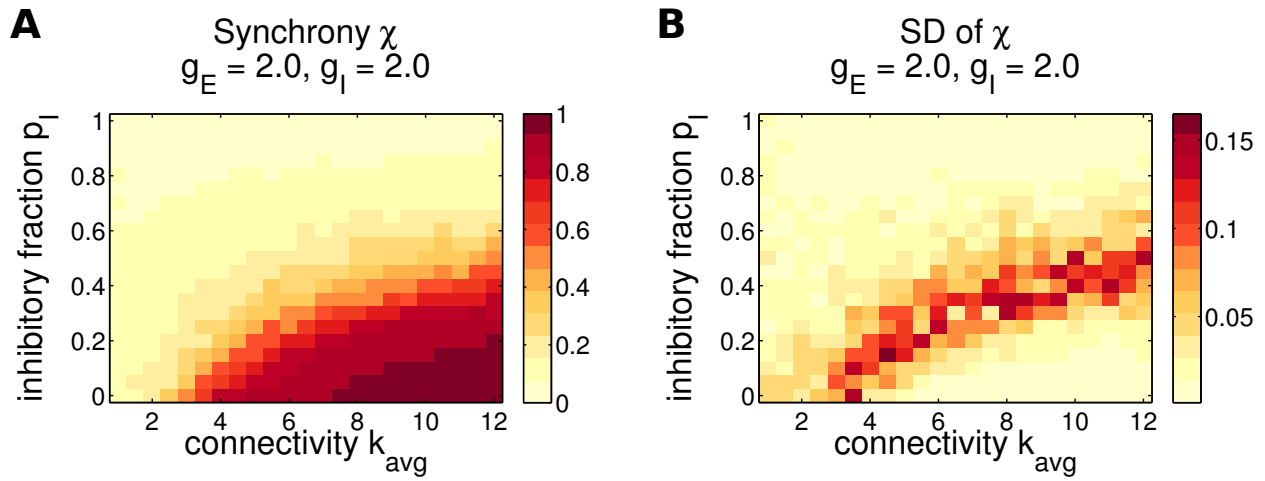


Figure 4.2: **Synchrony decreases with inhibition and sparsity. The highest variability across networks occurs at the synchronization boundary.** **A**, Synchrony parameter χ averaged over 8 network realizations, plotted versus the amount of connections k_{avg} and the fraction of inhibitory neurons p_I . **B**, Standard deviation of χ over network realizations. Higher standard deviation indicates that the synchrony is not reliable for different networks with those parameters. The area of highest standard deviation occurs at the boundary of low and high synchrony, $\chi \approx 0.5$. This is indicative of a phase transition between synchronized and desynchronized states.

the effects of inhibition and sparsity on synchrony by plotting χ as those parameters vary. Each point in the plot is the average χ over 8 network realizations with the corresponding parameters. The main tendency is for higher synchrony with higher k_{avg} , i.e. higher connectivity and less sparsity, and lower synchrony with higher p_I . A similar effect occurs when varying g_E and g_I , where stronger excitation synchronizes and stronger inhibition desynchronizes (shown in Fig. 4.8 for comparison with pharmacological experiments).

Inhibition thus decreases the synchrony within the preBötC microcircuit, which hinders the rhythm. At or above $p_I = 50\%$, the network is desynchronized for all connectivities k_{avg} . With an inhibitory majority, most inputs a neuron receives are desynchronizing, thus no coherent overall rhythm is possible. This is one of our first major results: In a single microcircuit, constructed with *homogeneous* random connectivity and with I_{NaP} -

driven burst dynamics [Butera et al., 1999a], inhibition cannot lead to the creation of a multi-phase rhythm. Inhibition only has the effect of desynchronizing bursting neurons and disabling the overall rhythm.

For any type of random connectivity, there is no single network corresponding to a given inhibitory fraction and sparsity level. Rather, each setting of these parameters defines a probability distribution over a whole family of networks, and we can study rhythm generation on sample realizations. This raises the question of how consistent our findings are from one of these networks to the next. To address this, we next depict the standard deviation of χ across the 8 network realizations, shown in Fig. 4.2B. The standard deviation tells us how much variation in synchrony to expect for different random networks with these parameters, with a higher standard deviation indicating less reliability. The variability in networks is a result of their random generation. The highest standard deviation occurs near the border between synchrony and disorder, where the average $\chi \approx 0.5$ (see panel A). Above this border, almost all networks exhibit low synchrony, and below it networks consistently show the same levels of high synchrony. Near the transition, random variations in the network structure have a larger effect on synchrony. The increase in standard deviation at the boundary between high and low synchrony is indicative of a “phase transition” between synchronized and desynchronized network states [Arenas et al., 2008].

4.3.2 *Inhibition creates an expiratory subpopulation*

In the preBötC, the majority of cells fire in phase with inspiration, but there are also cells that fire during other phases (post-inspiratory or expiratory) along with tonically active cells. A goal of our study is to identify the network and inhibitory effects leading to this variety of cells.

In order to analyze the time during the ongoing population rhythm at which individual model neurons are active, we identify robust peaks in the integrated trace as popula-

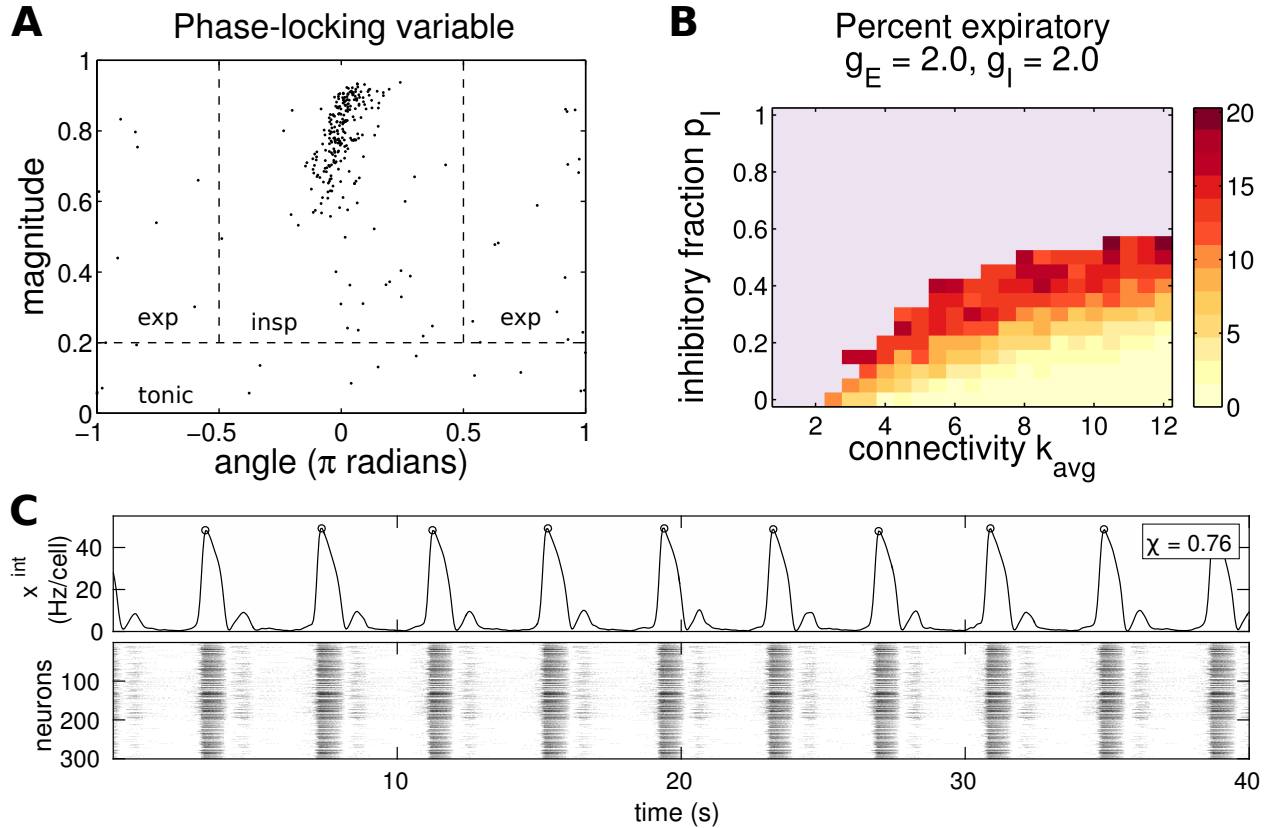


Figure 4.3: Expiratory cells arise from inhibition, but can only occupy a minority without disrupting the inspiratory rhythm. **A**, Neuron phase-locking variables for the simulation in Fig. 4.1B ($k_{avg} = 6, p_I = 20\%$). Each neuron has an associated complex number z_i with $0 \leq |z_i| \leq 1$. The magnitude $|z_i|$ is plotted against angle $\arg z_i$. These are used to define inspiratory, expiratory, and tonic neurons via the labeled regions separated by the dashed lines. **B**, Expiratory (anti-phase with main rhythm) neurons as a function of network parameters k_{avg} and p_I . The fraction of expiratory neurons increases with inhibition or as the connectivity becomes weaker. The blue indicates the absence of any overall rhythm, defined as $\chi < 0.25$. **C**, An example of a simulation with two-phase activity, with $k_{avg} = 6, p_I = 30\%, g_E = 5.0$, and $g_I = 2.0$. A minority of neurons produce a reliable, small bump after every burst. It is aligned near 0.7π , so it is more of a post-inspiratory or pre-expiratory burst. These expiratory cells are rebound bursting after being disinhibited. This is similar to the “handshake” mechanism of Wittmeier et al. [2008]. However, this type of two-phase rhythm is very rare in simulations.

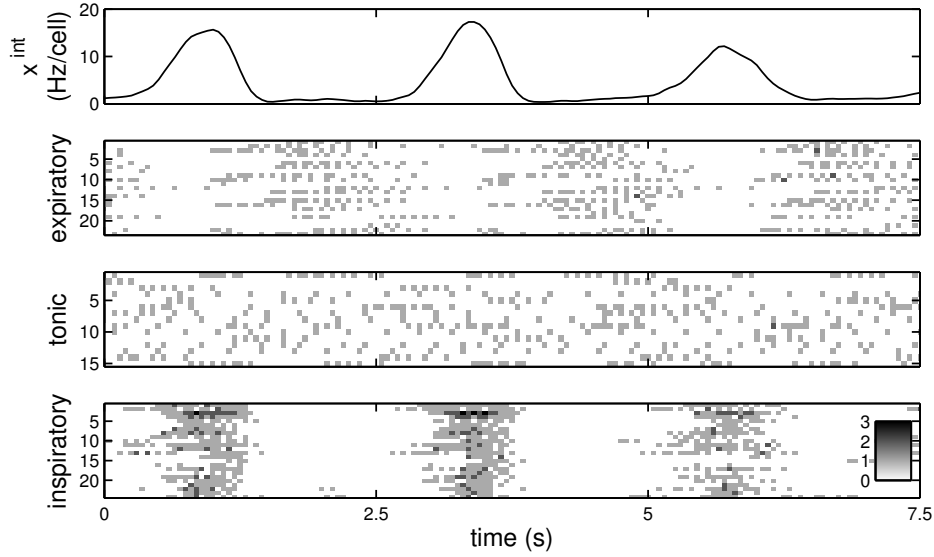


Figure 4.4: **Example rasters of expiratory, tonic, and inspiratory cells.** Expiratory cells exhibit lower firing rates than inspiratory ones, similar to the typical tonic firing observed in slices. As shown, tonic classified cells can be bursting so long as their bursts do not occur reliably at any given phase. The inspiratory cells shown are a random subset. Data are for a representative network with $k_{\text{avg}} = 6$, $p_I = 20\%$ (same as Figs. 4.1B and 4.2B).

tion bursts (see Section 4.2.4 for details). This allows us to map time into a phase variable $\phi \in [-\pi, \pi]$ and study neuron activity triggered on phase. Each peak in the rhythm occurs as the population bursts in synchrony and sets the phase $\phi = 0$. Values of $\phi \approx 0$ correspond to the inspiratory phase, since this corresponds to activity in phase with the overall population rhythm, which for the preBötC is inspiration. A phase near π or $-\pi$ we call expiratory. We examine cells' firing rates as a function of phase, which we call the burst triggered average (BTA, Eqn. 4.3). Using this, we define a phase-locking variable z_i (Eqn. 4.4) for each cell. The magnitude $|z_i|$ reflects how selectively cell i responds to phase, and the angle $\arg(z)$ tells the phase it prefers. This allows us to classify cells as inspiratory, expiratory, tonic, or silent. Fig. 4.3A shows the phase-locking variables z_i for an example simulation with parameters that generate a realistic rhythm ($k_{\text{avg}} = 6$, $p_I = 20\%$, $\chi = 0.716$, with raster and integrated trace in Fig. 4.1B). In this case we see most neurons

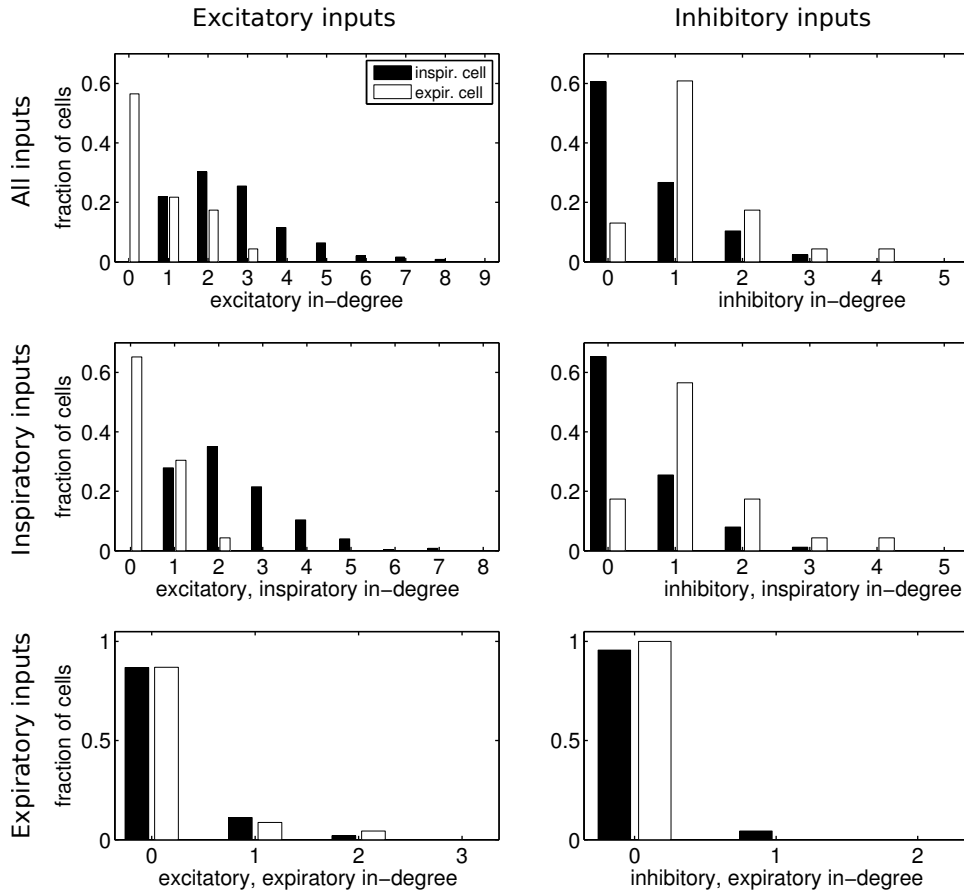


Figure 4.5: **Expiratory cells preferentially receive input from other excitatory, expiratory cells and inhibitory, inspiratory cells.** **B**, The top row shows the distribution of inputs, excitatory on left and inhibitory on right, colored by whether the receiving cell is inspiratory (black bars) or expiratory (white bars). Expiratory cells receive less excitatory and more inhibitory connections than inspiratory cells. The center and bottom rows break down these inputs by the phase of the presynaptic neuron, inspiratory inputs shown in the center and expiratory below. Expiratory cells preferentially receive excitatory input from other expiratory cells (compare middle left and bottom left). Furthermore, inhibitory input to expiratory cells tends to come from inspiratory cells rather than other expiratory cells (middle right and bottom right). Data are for a representative network with $k_{\text{avg}} = 6$, $p_I = 20\%$ (same as Figs. 4.1B and 4.2B). There were 251 inspiratory, 23 expiratory, 15 tonic, and 11 silent cells.

are inspiratory, with a dominant cluster of phase-locking variables centered on $|z| \approx 0.8$ and $\arg(z) \approx 0$. The rest of the cells are distributed approximately uniformly at random in the phase/magnitude cylinder. In this example, the majority of cells are inspiratory, with a smattering of expiratory and tonic cells.

Panel B in Fig. 4.3 shows our main results. For any connectivity level k_{avg} , we find that the number of expiratory neurons increases as the fraction of inhibitory cells p_I increases until the rhythm degrades entirely. Note that there can be a few expiratory neurons even with $p_I = 0$ for $k_{\text{avg}} < 4$. However, at this connectivity each cell has less than 2 incoming connections on average. The expiratory cells in that case are isolated from the rest of the network and have in-degree zero, with their phase only reflecting random initial conditions. Comparing Figs. 4.2A and 4.3B, we see that the number of expiratory neurons grows as synchrony decreases.

Another key finding of panel 4.3B is that there are never more than 20% expiratory cells. This means that, in this kind of unstructured microcircuit, it is not possible to create a two-phase rhythm where the expiratory burst is of similar magnitude to the inspiratory burst. Up to approximately 20% of neurons can be expiratory without destroying the rhythm, defined as maintaining $\chi \geq 0.25$. Fig. 4.3C shows an example of a rhythm with two phases, where the expiratory or post-inspiratory phase recruits only a minority of cells. The expiratory burst in this case is caused by rebound bursting of expiratory cells when they are released from inhibition. However, a two-phase rhythm of this magnitude is rare in our simulations. For example, it does not occur in other network realizations with the same parameters as Fig. 4.3C.

One of our goals is to understand the network mechanisms that give rise to expiratory cells. In Fig. 4.4, we show the firing properties of some example expiratory, tonic, and inspiratory classified cells. Expiratory and tonic cells both fire at lower rates than inspiratory cells, which are active in tight bursts. The modeled expiratory cells thus show tonic active behavior which is suppressed by inhibition, as observed in slice [Shao and Feldman, 1997, Lieske et al., 2000]. Note that some of the tonic cells in Fig. 4.4 are bursting,

just not at a reliable rhythm phase.

Each neuron's phase-locking properties are determined by its intrinsic dynamics and the excitatory and inhibitory synaptic currents it receives during various phases of the rhythm. In the model, we find that expiratory cells receive different synaptic inputs than inspiratory cells. We can see this by plotting their input properties in Fig. 4.5, in this case for a typical simulation in the partially synchronized regime, the same parameters as Fig. 4.1B. Overall, expiratory cells have less excitatory inputs and more inhibitory inputs than inspiratory cells (top panels). We also break down these inputs by the phase of the presynaptic cell. Expiratory cells receive less excitation during the inspiratory phase, and they similarly receive more inhibition during the inspiratory phase (center panels). Given that expiratory cells are the minority, the trends for inputs during the expiratory phase are not as strong (bottom panels). This suggests that expiratory cells emerge from random configurations in the network, which partitions itself into different phases based on the types of interactions in each cell's neighborhood. Excitatory synapses drive the postsynaptic neuron into phase with the presynaptic one, while inhibitory synapses drive neurons out of phase.

As we have shown in the preceding two sections, the presence of inhibition leads to changes in the population rhythm generated in microcircuits: a degradation of the overall population synchrony as well as an increasing presence of expiratory cells. The average degree k_{avg} controls the sparsity of connections in the network, and lower values also lead to less synchrony. Moreover, we have shown that cells become expiratory due to the arrival of inhibition during the the inspiratory phase as well as excitation during the expiratory phase.

4.3.3 *Two population network shows the benefits of half-center inhibition*

In Section 4.3.1 we examined the effect of inhibition on rhythmic spiking in a single microcircuit, as would model, for example, an isolated preBötC [e.g. Ramirez et al., 1997a].

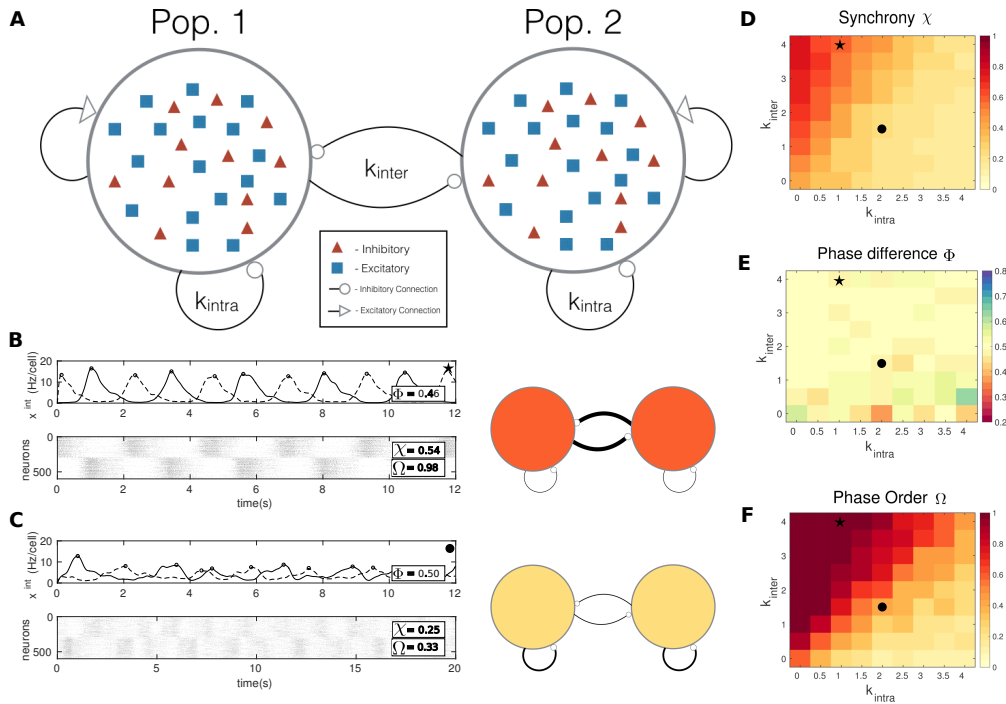


Figure 4.6: **A**, Schematic of the two population network. The average amount of excitatory connections on average are fixed, but we vary the expected intra-group and inter-group inhibition k_{intra} and k_{inter} . **B** and **C**, Two simulations of the network with different parameters: Each simulation also has a schematic on the right demonstrating the differences in inhibitory network strengths. A thicker line indicates more connections, and a darker color indicates a more reliable rhythm. Panel B shows the case $k_{intra} = 1.0$ and $k_{inter} = 4.0$. Panel C depicts $k_{intra} = 2.0$ and $k_{inter} = 1.5$. There, we observe less reliable rhythms, with decreased phase order Ω and decreased synchrony χ , despite approximately equal average phase difference Φ . **D**, Average synchrony over 8 realizations for each k_{inter} and k_{intra} pair. Higher values of χ occur above the diagonal $k_{inter} = k_{intra}$ line. **E**, Average phase difference Φ of rhythmic bursts between the two populations. No clear trends are evident, and the value is close to $\Phi = 0.5$, perfectly out-of-phase, in much of the region. **F**, Average phase order Ω . Higher phase order indicates the relative phase of bursts in Pop. 1 and Pop. 2, i.e. Φ in panel E, are reliable. The phase order appears to be proportional to the synchrony, with the highest values above the diagonal. Star and circle symbols in D–F are the network parameters used to produce the rhythms in panels B and C.

There we saw that increasing inhibition causes the synchrony and rhythmicity of neural spiking to degrade. Here, we extend our analysis to a model of two coupled microcircuits. Each microcircuit, taken separately, is a heterogeneous subnetwork of cells with exactly the same properties and parameterization as for the networks studied above. The two microcircuits are then coupled with mutual inhibition in the manner of a classical half-center pattern generator. We explore the effects of inhibition on the synchrony within each microcircuit, as well as on the phase of the two microcircuits relative to one another.

Figure 4.6A shows a schematic of our network model. As in the previous sections, each microcircuit (a distinct *population* of cells) contains both excitatory and inhibitory neurons. For simplicity, since we want to isolate the effects of inhibitory structure, the excitatory neurons only project locally, that is within the same microcircuit. We vary inhibitory connectivity via the parameters k_{inter} and k_{intra} , the intra-group and inter-group average degrees for inhibitory cells. For example, setting $k_{\text{inter}} = 0$ yields independent populations that do not interact; when $k_{\text{intra}} = 0$ and $k_{\text{inter}} \neq 0$, we have a network version of the classic half-center oscillator, with inhibition purely between the two microcircuits. We will investigate network activity at these two extremes and intermediate levels of connectivity.

Panels B and C in Fig. 4.6 illustrate the role of inhibitory connectivity on rhythmic spiking dynamics in two representative cases. The upper network (see schematic), has weaker inhibition within each population than between the populations, with parameters $k_{\text{intra}} = 1.0$ and $k_{\text{inter}} = 4.0$. The population activity exhibits a strong, regular, and synchronous rhythm with little change in the phase relationship over time. The bottom network has the opposite connectivity: stronger inhibition within each population and weaker inhibition between ($k_{\text{intra}} = 2.0$ and $k_{\text{inter}} = 1.5$). This network demonstrates a weak, sporadic rhythm with a varying phase relationship through time. These suggest that inhibition within microcircuits competes with inhibition between them to determine the strength and phase relationships of rhythms. We now explore this trend across a broad range of connectivity levels.

First, we show how intra- and inter-group inhibition affect the synchrony in the two population model. To quantify this, we compute the synchrony measures for each population separately (χ_1 and χ_2), and report the average $\chi = (\chi_1 + \chi_2) / 2$. Figure 4.6D shows the results. As intra-group inhibition k_{intra} increases, there is a degradation in synchrony. This is consistent with the results from the single population model, where unstructured local inhibition reduces the strength and regularity of the population rhythm. Panel C gives an example of network activity in this regime, and is indicated by a circle in panel D–F. However, as we add inhibitory connections between the two populations by increasing k_{inter} , synchrony recovers: overall, we see stronger synchrony above the diagonal where $k_{\text{inter}} = k_{\text{intra}}$. Panel B, indicated by the star in D–F, illustrates this. Overall, Figure 4.6D suggests that intra-group inhibition destabilizes synchrony, while inter-group inhibition can have the opposite effect.

In order to drive breathing, in which each microcircuit presumably generates a different phase in a motor pattern, the model should produce two rhythms with reliable phase separation. To analyze this, we first compute a measure of the average, over time, of the difference between the phases of each microcircuit, which we call Φ . A value $\Phi = 1$ or 0 indicates that the two rhythms are, on average, in-phase, and $\Phi = 0.5$ indicates the two rhythms are, on average, perfectly out-of-phase (see further details in methods Sections 4.2.4 and 4.2.4). Figure 4.6E shows that $\Phi \approx 0.5$ over the range of inhibitory connectivity. Thus, the two microcircuits appear to be out of phase on average, regardless of connectivity. A glance back at panels B and C reveals that this out-of-phase behavior can arise in different ways: either for two reliable rhythms that are phase-locked, or for two unreliable rhythms that drift broadly with respect to one another over time. To quantify this difference, we use a phase order metric Ω (Section 4.2.4), shown in Fig. 4.6F. Here, $\Omega = 1$ indicates that the phase differences are completely repeatable over time, while $\Omega = 0$ indicates phase differences are completely unreliable, instead being evenly spread over time. In agreement with the two cases illustrated in panels B and C, as we increase the inhibition within microcircuits k_{intra} , phase reliability Ω decreases; conversely,

increasing k_{inter} increases Ω .

These results lead to the important conclusion that it is not a particular number of inhibitory connections in a network that leads to a stable two-phase rhythm, but instead the relative strengths of intra- and inter-group connectivity. For a stable two-phase rhythm, there need to be at least as many inhibitory connections between populations as within populations. The key rhythm metrics, synchrony χ and phase order Ω , demonstrate the same effect, because χ and Ω are strongly correlated. This makes sense because the rhythms are generated through synchronous bursting. Note that an irregularity score for the phase differences would yield similar results as Ω , but we prefer Ω since it takes into account the circular structure of the phase variable. Increasing intra-group inhibition pushes the system to the edge of stability. However, we are able to recover some rhythm stability and phase separation reliability by increasing inter-group inhibition. In summary, we see the same desynchronizing effect of local inhibition as in the single population model, with some benefit to synchronous rhythms possible from inter-group inhibition.

4.3.4 *Partial synchrony of in vitro preBötC rhythms in multi-array recordings*

We now turn to experiments with the preBötC, to test the model predictions about the role of inhibition in such circuits. We recorded from mouse transverse brainstem slices containing the preBötC, keeping only those that initially exhibited robust rhythms. This yielded a collection of 17 recordings of the population rhythm using a large extracellular local field potential (LFP) electrode. Of these, 4 were simultaneously recorded with a linear electrode array to capture the behavior of multiple neurons (16, 29, 33, and 29 cells were isolated in individual experiments). From the multi-array data, we extracted individual spikes and calculated the synchrony metric χ as in the model.

Our experiments reveal that a fully synchronized network such as in Fig. 4.1A is not realistic under our experimental conditions. This is because preBötC slices exhibit signif-

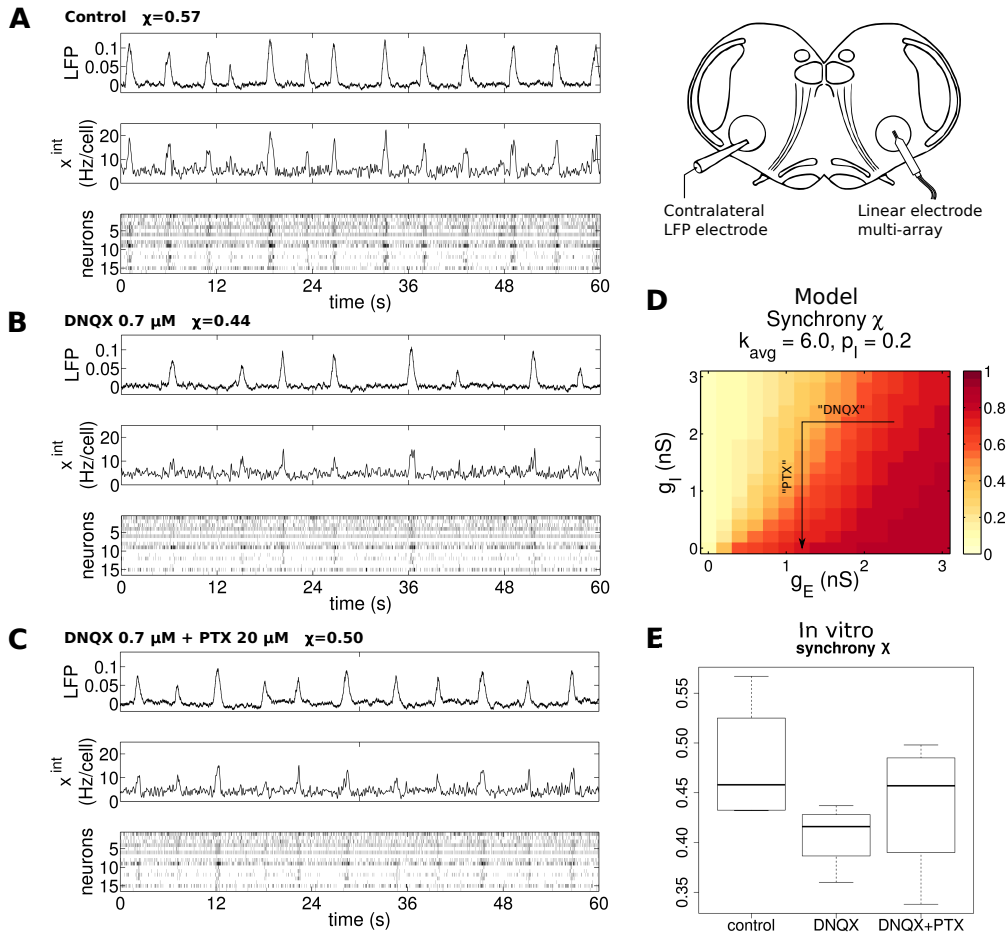


Figure 4.7: In vitro array recordings from transverse slice preparations exhibit partial synchrony. We performed in vitro preBötC slice experiments, where we measured the rhythm in control, partial excitation block (DNQX 0.7 μM), and partial excitation block with full inhibition block (DNQX 0.7 μM + PTX 20 μM). We record the preBötC population activity with a large electrode (LFP, arb. unit) as well as individual neurons in the contralateral area using an array. The average activity of the isolated units is also shown (x^{int} , Hz/cell). **A**, Control conditions show a robust population rhythm with some amplitude and period irregularity. **B**, Partial excitation block using DNQX degrades the population synchrony, with decreased burst amplitude, slower rhythm, and more irregular intervals between bursts. **C**, Blocking inhibition with PTX allows the rhythm to recover toward control conditions. **D**, Synchrony in the model, as a function of excitatory and inhibitory synaptic conductances g_E and g_I , increases with stronger excitation and decreases with stronger inhibition, similar to varying connectivity k_{avg} and inhibitory fraction p_I . Arrows indicate the presumed effects of DNQX and PTX on the model. **E**, Measurements of synchrony from our 4 array recording experiments. Synchrony takes intermediate values in all conditions, decreasing with DNQX and recovering after PTX.

icant cycle-cycle variability [Carroll et al., 2013, Carroll and Ramirez, 2013]. So real networks are somewhere in the intermediate synchrony range. We confirmed this in multi-array in vitro experiments. An example experiment with 16 cells is shown in Fig. 4.7A. We observe that there is significant cycle-to-cycle period and amplitude variability in the rhythm, which is reflected in the partial synchrony of the 16 neurons recorded ($\chi = 0.57$). With $n = 4$ multi-electrode control experiments, we measured an average $\chi = 0.48$ (SD 0.055).

The number of expiratory neurons observed in other experiments is also consistent with the degree of partial synchrony in the model. Multi-array recordings by Carroll et al. [2013] found 5.0% expiratory and 3.9% post-inspiratory cells. Counted together, as we are doing, a realistic percentage of expiratory cells is 9%. Referring to Figs. 4.2A and 4.3B, we see that this occurs near the region where $\chi \approx 0.6$. This value is not far from the experimentally measured average $\chi = 0.48$. However, we did not observe any expiratory cells in our limited set of 4 multi-array experiments, which is expected based on Carroll et al. [2013].

In Fig. 4.7B and C, we also show the behavior of the slice under pharmacological manipulations of the efficacy of excitatory and inhibitory synaptic transmission, shown here for completeness and explored in more detail in Section 4.3.5. Specifically, we use the glutamatergic antagonist DNQX and the GABA and glycine receptor antagonist picrotoxin (PTX) (Section 4.2.3). After recording the control rhythm, we applied DNQX 0.7 μM to partially block excitation and observed the resulting rhythm. After recording in DNQX conditions, we follow with application of picrotoxin (PTX) 20 μM . The dosages are chosen so that DNQX partially blocks excitation [Honore et al., 1988] but does not stop the rhythm, whereas the PTX dosage is high enough to effect near-complete disinhibition [see Fig. 1 in Othman et al., 2012]. We see in Fig. 4.7B that DNQX leads to less synchrony and a visibly degraded, slower rhythm. Moreover, Fig. 4.7C shows that when this inhibition is reduced by adding PTX, the rhythm recovers toward control values of frequency, amplitude, and synchrony.

When varying synaptic conductances in a simulation of the effects of DNQX and PTX, the computational model behaves as one might expect from our earlier results. We generated 8 networks with average degree $k_{\text{avg}} = 6$ and inhibitory fraction $p_I = 20\%$. Then we varied the maximal conductances of excitatory and inhibitory synapses g_E and g_I while keeping the network structure fixed. We show the synchrony χ as a function of g_E and g_I in Fig. 4.7D. Increased g_E leads to enhanced synchrony, while, as expected from the results above, increased g_I desynchronizes the population. Thus, once again we find that excitation synchronizes and inhibition desynchronizes activity within a microcircuit.

Finally, in Fig. 4.7E we summarize the synchrony χ across all 4 multi-array experiments and pharmacological conditions. Clearly, the networks are all partially synchronized. Synchrony χ decreases by about 0.07 (SE 0.02, DF 8, $t=-3.414$, $p=0.009$) with DNQX, with a recovery to near baseline following PTX. These trends are shown in only 3 out of 4 experiments, so we stress that this is marginally significant according to the mixed effects model (see Table 4.2). We next show how proxies for the synchrony which measure regularity of the rhythm can be applied to our larger collection of LFP recordings to further illuminate this trend.

4.3.5 *Excitatory and inhibitory balance modulates rhythm irregularity in vitro and in silico*

In Sections 4.3.1–4.3.3, we use a computational model to show how population rhythms depend on levels of inhibitory connectivity within and between microcircuits. We have demonstrated that in vitro preBötC networks are naturally in a partially synchronized state, Sec. 4.3.4. We now investigate how in vitro preBötC rhythms behave under the modulation of synaptic conductances using pharmacological techniques. To quantify rhythm quality from the integrated LFP signal, available in all 17 of our recordings, we turn to amplitude and period irregularity. These measure the cycle-to-cycle variability of the sequence of burst amplitudes and inter-burst-intervals [Sec. 4.2.4 and Carroll et al., 2013, Carroll and Ramirez, 2013].

Our experiments use the synaptic antagonists DNQX and PTX to pharmacologically modulate the efficacy of excitatory and inhibitory synapses *in vitro*, analogous to lowering g_E and g_I , respectively. This is illustrated with the arrows in Fig. 4.7D. In Fig. 4.8, we also illustrate the behavior of the amplitude and period irregularity scores in the model as g_E and g_I vary. Comparing Figs. 4.8 and 4.7D, it is apparent that both irregularity scores increase in the model as χ decreases. In the 13 experiments where we have only an LFP signal, this suggests that irregularity can stand in as a proxy for neuron synchrony, which we could only measure with multi-cell array recordings.

We plot *in vitro* irregularity across conditions in Fig. 4.8 using box plots. The results of statistical tests using a linear mixed effects model are shown in Table 4.2. To summarize, amplitude irregularity shows no significant trends with the blocking of excitation via DNQX and inhibition via PTX. However, we noted a statistically significant increase (DF=34, $t=5.03$, $p=1.6 \times 10^{-5}$) in period irregularity of about 0.12 (SE 0.02) following application of DNQX and subsequent decrease with PTX to near baseline. The qualitative effect on period irregularity matches trends present in the computational network model.

The model also predicts that there would be a slight decrease in irregularity with initial application of PTX after control, i.e. a variant of the previous protocol without DNQX. We performed limited experiments with varying doses of PTX and found some small decreases in period irregularity which were not significant (data not shown). However, it did appear that the more irregular control slices showed greater decreases in irregularity with application of PTX, as also would be expected from the model results in Fig. 4.8.

With regards to the lack of a trend in amplitude irregularity, we note that the “landscapes” of the amplitude and period irregularity scores produced by the computational model (heat maps in Fig. 4.8) show markedly different regions of high irregularity. In the amplitude irregularity case, the red region of high values is much wider than in the period irregularity case. For amplitude, it is shaped like a plateau rather than the steep slope of period irregularity. This suggests that amplitude irregularity is less sensitive to synaptic modulation, perhaps making trends harder to identify in pharmacological

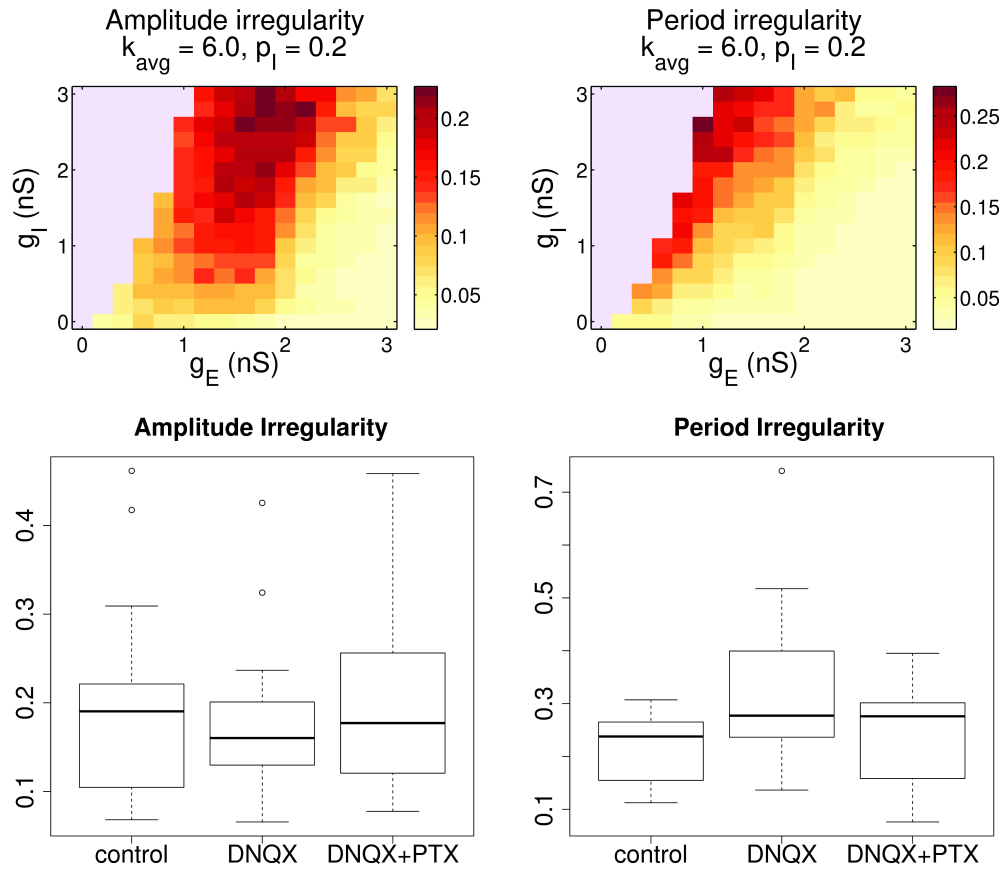


Figure 4.8: **Modulation of inhibition and excitation changes the rhythm in comparable ways for experiments and the model. (Above: model)** The effect of changing conductances g_E and g_I . Burst amplitude and period irregularity decrease with stronger excitation and weaker inhibition. Both of these measures are negatively correlated to the population synchrony, shown in Fig. 4.7D. **(Below: experiments)** This plot summarizes 17 experiments. We extracted bursts from the LFP and measured the amplitude and frequency irregularity of those rhythms. Amplitude irregularity showed no significant trends across conditions. However, period irregularity showed a significant increase from control with DNQX, a decrease from DNQX to DNQX+PTX, and a small increase between control and DNQX+PTX. See Table 4.2 for the full output of the statistical tests.

	Fixed Effect	Std. Error	DF	t value	Pr(> t)
Amplitude Irregularity					
Intercept	0.197894	0.024664	24.63	8.024	2.48×10^{-8}
DNQX	-0.016169	0.018067	34	-0.895	0.377
DNQX+PTX	-0.005418	0.018067	34	-0.3	0.766
Period Irregularity					
Intercept	0.21622	0.02627	30.38	8.23	3.14×10^{-9}
DNQX	0.12076	0.02401	34	5.03	1.57×10^{-5}
DNQX+PTX	0.031	0.02401	34	1.291	0.205
Amplitude Mean (a.u.)					
Intercept	0.070486	0.009401	18.72	7.498	4.76×10^{-7}
DNQX	-0.01134	0.003552	34	-3.192	0.00304
DNQX+PTX	-0.000614	0.003552	34	-0.173	0.86379
Period Mean (s)					
Intercept	4.0594	1.0299	26.11	3.942	0.00054
DNQX	4.6371	0.8105	34	5.721	1.98×10^{-6}
DNQX+PTX	1.9396	0.8105	34	2.393	0.02238
Synchrony χ					
Intercept	0.47875	0.02506	6.497	19.104	5.94×10^{-7}
DNQX	-0.0715	0.02094	8	-3.414	0.00917
DNQX+PTX	-0.04125	0.02094	8	-1.97	0.08439

Table 4.2: Statistical results for in vitro measurements of amplitude irregularity, period irregularity, amplitude, and period. We report the estimated fixed effect for the intercept, DNQX, and DNQX+PTX conditions, as well as standard error (SE), degrees of freedom (DF), t value, and p value for each effect. These data summarize 17 LFP recordings save the synchrony fit, which comes from 4 multielectrode recordings.

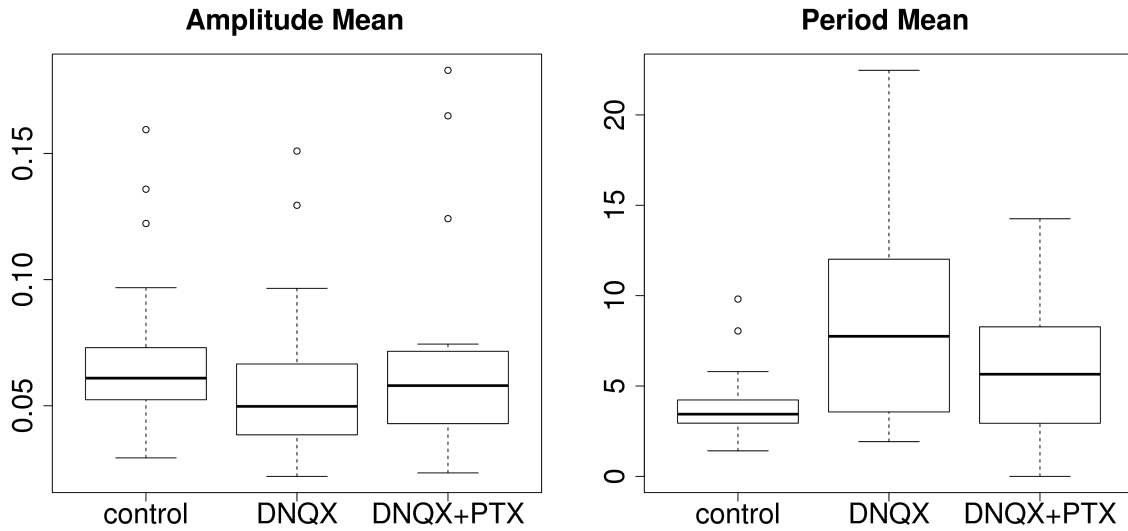


Figure 4.9: The effect of DNQX and PTX on in vitro rhythm amplitude and period, similar to Figure 4.8. Amplitude decreases with DNQX while period increases, with both recovering to near baseline after addition of PTX. See Table 4.2 for the result of statistical tests on this data.

experiments. However, it could also be that bursting in the real preBötC is essentially an “all-or-nothing” phenomenon, with amplitude irregularity a result of noise but not strongly dependent on details of the burst dynamics, in contrast to the model we study. This would make it insensitive to blockers, since once a burst is triggered it is reliable and consistent, similar to the triggering of an action potential. This is interesting in the context of the burstlet hypothesis [Kam et al., 2013].

4.3.6 *In vitro* rhythm slows following excitatory block

Besides variability, we found in experiments that synaptic blockers also significantly change the overall period and amplitude of rhythmic bursts, as shown in Fig. 4.9 and Table 4.2. Mean burst amplitude is decreased by -0.011 units (SE 0.004, DF=34, $t=-3.192$, $p=0.003$) following DNQX and recovers to baseline with application of PTX. This is consistent with the effect of varying g_E and g_I in the model. In experiments, we also see a significant slowing of the rhythm. The burst period increases with DNQX by 4.6 s (SE 0.81, DF=34,

$t=5.72$, $p=2 \times 10^{-6}$) and only partially recovers with application of PTX, remaining 1.9 s (SE 0.81, DF=34, $t=2.39$, $p=0.02$) above baseline. As described above, while our network model qualitatively predicts the experimental trends for period variability and amplitude modulation in the isolated preBötC, it does not reproduce overall changes in burst period.

Simple modifications to the model capture the period slowing with excitatory blockers. Suppose each respiratory cell receives concurrent input from excitatory and inhibitory pools of tonic neurons [Ramirez et al., 1997b]. These cells determine a baseline drive to the preBötC, which we model as a constant current I_{app} . Tonic external conductances g_E^{app} and g_I^{app} have the same effect but complicate our parameter tuning due to modification of the effective leak current. DNQX would then lower the excitatory drive, leading to decreased I_{app} . A negative drive current then slows the amount of time it takes a neuron to integrate to bursting, lowering the neuron's intrinsic burst frequency. PTX, by lessening the influence of the inhibitory tonic pool, causes a net disinhibitory effect on the neuron, restoring I_{app} to near baseline. So far, we have taken $I_{app} = 0$ as the baseline, but these differential effects remain regardless of the baseline tonic current. Mimicking DNQX with $I_{app} = -4$ pA causes the period to approximately double (not shown but tested for $k_{avg} = 6$, $p_I = 0.2$, $g_E = g_I = 3.0$ in control, $g_E = 1.8$ under DNQX).

One consequence of this tonic pool hypothesis is that changing the baseline drive also changes the intrinsic dynamics of neurons. Increased hyperpolarization can cause tonic cells to become bursters, and bursters to become silent in the absence of network effects. However, this can benefit synchrony, since when the large pool of originally tonic cells shift into bursting mode, they can help maintain a strong rhythm despite the reduced excitatory synaptic drive. In a check for a few network structures, we found that the 'main' effects of excitation and inhibition on rhythms persist when we also make these I_{app} changes.

To recap, our experimental results show that the control preBötC networks lie in the partially-synchronized regime. The results also confirm that the relative balance of excitation and inhibition determine the level of synchrony and variability of the rhythm. In

experiments, we also find a strong dependence of rhythm frequency on the amount of inhibition, and we have discussed changes to the model which could explain this effect.

4.4 Discussion

4.4.1 Network structure of respiratory areas

The preBötC contains neurons which are silent, tonic spiking, or periodically bursting pacemakers [Thoby-Brisson and Ramirez, 2001, Peña et al., 2004, Ramirez et al., 2011]. Numerous models are proposed for the preBötC, at the level of single neurons with pacemaker dynamics [Butera et al., 1999a, Best et al., 2005, Rubin et al., 2009b, Toporikova and Butera, 2011, Park and Rubin, 2013] as well as networks of these neurons (Butera et al., 1999b, Best et al., 2005, Purvis et al., 2006, Rubin et al., 2009a; Rubin et al., 2009b, Schwab et al., 2010, Gaiteri and Rubin, 2011, Lal et al., 2011, Rubin et al., 2011, Carroll et al., 2013, Carroll and Ramirez, 2013, Wang et al., 2014). Traditionally, these models have consisted of just the excitatory, essential core of inspiratory neurons. However, Ramirez et al. [1997b] showed that inspiratory cells receive concurrent excitation and inhibition in the inspiratory phase during both in vitro and in vivo recordings from cat preBötC. Furthermore, Morgado-Valle et al. [2010] demonstrated the existence of glycinergic inspiratory pacemakers within preBötC, likely candidates for the inhibitory population presynaptic to those found by Ramirez et al. We have chosen to study the consequences of mixed excitatory and inhibitory cells in this network.

The details of network structure in the preBötC is currently unknown, and molecular markers for rhythmogenic neurons have been found only recently [Wang et al., 2014]. Reklung et al. [2000] recorded from pairs of cells and estimated that 13% (3 of 23 pairs) were synaptically connected. However, the distance between the connected neurons of the 3 pairs is unknown. This, along with the small sample size, makes it difficult to know whether this connectivity is representative for the entire preBötC. Moreover, synaptic transmission was not entirely reliable. Thus, the robustness of these excitatory con-

nections is difficult to assess from those exceedingly difficult paired recordings. [Hartelt et al. \[2008\]](#) imaged the dendrites and axons of neurons in the area and found a network with spatially localized, modular structure similar to a small-world network. They estimated average neuron degrees were between roughly 2 and 6 [[Hartelt et al., 2008](#)]. [Carroll and Ramirez \[2013\]](#) recorded from in vitro slice preparations and argued for roughly 1% connectivity using cross-correlation analysis of 10,778 pairs. The number of cells in the preBötC is estimated to be around 300–600 [[Wang et al., 2014](#), [Winter et al., 2009](#), [Hayes et al., 2012](#), [Feldman et al., 2013](#)], although this differs significantly with the estimate of 3000 neurons by [Morgado-Valle et al. \[2010\]](#). This difference is mainly due to varying functional definitions of what constitutes a preBötC neuron. However, our results should not change much with the network size: because we parametrize the connectivity by the average degree, the in-degree distribution and thus variability of input signal to a given neuron (proportional to $k_{\text{avg}}^{-1/2}$) will not change significantly.

The exact structure of the preBötC network remains debatable, but it appears clear that the connectivity is relatively sparse. Many original models of the isolated preBötC assume a fully-connected network, i.e. a complete graph [[Butera et al., 1999b](#), [Purvis et al., 2006](#), [Rubin et al., 2009b](#)]. [Gaiteri and Rubin \[2011\]](#) studied a variety of different topologies and their effects on the rhythm. Random graphs have recently become more popular [[Schwab et al., 2010](#), [Gaiteri and Rubin, 2011](#), [Lal et al., 2011](#), [Carroll et al., 2013](#), [Carroll and Ramirez, 2013](#), [Wang et al., 2014](#)], however only a few of these studies have looked at sparse random networks with average degree less than 10 [[Carroll et al., 2013](#), [Carroll and Ramirez, 2013](#)]. We believe this sparse regime is relevant to the irregularity observed in vitro [[Carroll et al., 2013](#)].

While a clustered connectivity may be present in the preBötC, where it would have profound effects on rhythm generation [[Gaiteri and Rubin, 2011](#)], direct evidence for this is limited to the study of [Hartelt et al. \[2008\]](#). Furthermore, the preBötC is a bilateral rhythm generator with each side coupled to the other principally by excitatory connections [[Lieske and Ramirez, 2006](#), [Koizumi and Smith, 2008](#)], making the two-population

model perhaps well-suited for the preBötC. There is also evidence for excitatory connections between the expiratory and inspiratory centers [Onimaru et al., 2009, Tan et al., 2010, Huckstepp et al., 2015]. We did try adding a few excitatory projections between the two populations, and in our model only a few projections will make the two centers synchronize. Having predominantly excitatory connections between bilateral preBötC areas could further stabilize the rhythm. However, we have chosen to first model the simpler, sparse but unstructured random connectivity as presented. We leave a full exploration of such effects to future work.

4.4.2 *Rhythm patterning by inhibition*

The neural circuits that drive respiration can generate basic rhythms through excitation alone, yet they also include strong inhibitory connections both within and between different microcircuits. Our aim here is to shine light on the role of this inhibition. Through modeling studies that explored thousands of network configurations, we show that inhibition plays two main roles in excitatory rhythm generators that depend systematically on the structure of the underlying connectivity. Unstructured local inhibition within a single excitatory microcircuit, as for our model of an isolated preBötC, destabilizes rhythmic bursting by preventing the synchronization of excitatory neurons. This is in contrast to the spiking models where inhibition facilitates synchrony and relevant, for example, in the gamma oscillation [Börgers and Kopell, 2003]. Within such a single microcircuit with sparse, random, and homogeneous connectivity, adding inhibitory cells does not create a robust two-phase rhythm (i.e., inspiration *and* expiration). However, such inhibition does explain the presence of expiratory cells as have been observed experimentally [Carroll et al., 2013, Nieto-Posadas et al., 2014]. Our pharmacological experiments in the transverse preBötC slice also support the presence of local inhibition that is destructive to homogeneous synchrony: when we first partially block excitation, and then inhibition, we see that levels of period irregularity first increase and then decrease.

The same qualitative effects of local inhibition persist in a two population inspiratory-expiratory model, suggesting that the synchronizing and desynchronizing roles of excitation and inhibition within a population persist in more complicated systems. Moreover, long-range inhibition *between* excitatory microcircuits both stabilizes rhythms locally (reflected in their synchrony) and enforces reliable phase separation between microcircuits (phase order), reminiscent of the concept of the half-center [Brown, 1911, Stuart and Hultborn, 2008, Sharp et al., 1996] This suggests twin roles for inhibition: Within a single microcircuit, it reduces synchrony and introduces some out-of-phase cells; between populations, it facilitates partitioning of the overall rhythm into different phases. As such, inhibition balances against excitation in a way that depends on the overall connectivity of the network.

How strongly do the twin roles for inhibition play out in biological circuits for breathing? Anatomical studies have suggested substantial inhibition within microcircuits, and recordings have shown some cells with expiratory or post-inspiratory firing within the predominantly inspiratory preBötC [Carroll et al., 2013, Morgado-Valle et al., 2010, Nieto-Posadas et al., 2014]. Intriguingly, our model predicts that the level of local inhibition that is consistent with these observations moves the circuits as a whole toward the boundary between ordered, synchronous and disordered, asynchronous activity. This could be useful for making the network more sensitive to control signals. For instance, descending excitatory inputs that selectively target the inhibitory population could lead to pauses in the rhythm.

This frames two questions: First, what constructive role could such destabilizing inhibition play? Possibly, it could produce a rhythm that has a particular temporal pattern (e.g. ramping) or that could be more flexibly controlled. Second, what role might destabilizing inhibition play in disease states in which rhythms within and between respiratory and other centers degrade?

Physiological studies suggest interesting answers to the first of these questions. Local inhibition within the preBötC has a critical role in shaping the inspiratory pattern

[Janczewski et al., 2013, Sherman et al., 2015], as our modeling study also shows. One of the hallmarks of “eupnea” or normal breathing is an augmenting ramp-like inspiration which is lost when inhibition is blocked in the isolated preBötC [Lieske et al., 2000]. Characterizing the synaptic profile of inspiratory neurons reveals the presence of concurrent inhibition and excitation which likely prevent an effective synchronizing between the excitatory neurons, thereby slowing down the build-up of inspiratory activity. Indeed, we hypothesize that the presence of local, desynchronizing role of inhibition within the preBötC could also explain an ongoing debate in the field of respiration, i.e. why an isolated preBötC can generate a eupnea-like inspiratory activity pattern in the absence of the other phases of respiratory activity [Lieske et al., 2000, Ramirez and Lieske, 2003]. The augmenting inspiratory discharge in the isolated preBötC is very sensitive to the blockade of inhibition. In hypoxia, when synaptic inhibition is suppressed, the desynchronizing effect of local inhibition is lost and the isolated preBötC generates an inspiratory burst that is characterized by a fast rise time reflective of a facilitated synchronization. However, the Butera model we implemented does not exhibit these rise time effects at the single-cell level. Instead, the behavior only becomes evident in the population due to the misalignment of individual neuron bursts, and this overall effect is quite weak (data not shown). It is likely that other currents are important for the individual burst characteristics and that future models including these will provide further evidence for a role for local inhibition in shaping inspiratory bursts.

4.4.3 *Limitations of our study*

There was considerable variability in the control rhythms and the responses to drugs. We believe this is principally due to intrinsic variability of the preBötC network structure across mice, the slicing procedure which damages the network to varying degrees, and the moderate dose of DNQX. The multi-electrode recordings captured between 16 and 33 units. This small sample of cells contributes significant variance to our synchrony

measure χ , and we believe this is why we cannot see a significant effect on synchrony. We placed the electrode array where we could record from many inspiratory cells, however we also found almost as many tonic cells. It is possible that these are cells which are not integrated into the network and therefore could bias χ to lower values. In future work, it would be important to see whether the rhythm also degrades with inhibitory agonists, e.g. muscimol [see Janczewski et al., 2013]. However, agonists introduce a tonic input which is rather different than modifying the synaptic efficacies, thus they will have a different effect than antagonists or optogenetic stimulation.

Our slice experiments showed a slowing down with excitation block and no statistically significant variation in amplitude irregularity, both in contrast to the model. Other membrane currents may explain these salient features of our pharmacological studies. We proposed that tonic populations could drive the change in frequency. However, the CAN current is another likely candidate. Since CAN-dependent pacemakers can rely on accumulation of excitatory synaptic events to initiate bursting [Rubin et al., 2009a, Del Negro et al., 2010], excitatory synaptic block will slow this accumulation, leading to an increase the rhythm period. This mechanism would be similar to the synaptic integrator model of Guerrier et al. [2015], which reproduced the period effects of NBQX (similar to DNQX). As mentioned above, the CAN current is also probably important for generation of augmenting, ramping discharges. Our model excluded CAN for simplicity and because the vast majority of respiratory models use the Butera et al. [1999a] persistent sodium equations. Also, it appears that cadmium-sensitive intrinsic bursting neurons (presumably the same as CAN-dependent) are only a minority of the respiratory neurons in the preBötC [Peña et al., 2004]. Hayes et al. [2008] present evidence that a low-threshold, inactivating K^+ current I_A is present in preBötC neurons and significantly affects rhythmogenesis. They conclude that I_A helps control amplitude and frequency irregularity by preventing or delaying those neurons from responding without massive excitatory input. Beyond irregularity, I_A and I_{CAN} are also important for overall burst shape, duration, inter-spike intervals, burstiness, etc., which are interesting topics for future study. Finally, synaptic

delays can be very important determinants of synchronization strength and phase relationships [Brunel and Hakim, 1999]. Future models will need to investigate how these many currents interact with excitatory and inhibitory synaptic dynamics in rhythm generation.

4.4.4 Conclusions

Our results contribute to a large body of modeling and experimental work in the field. Because local inhibition has a desynchronizing role, the preBötC cannot generate a two-phase rhythm, consistent with lesioning experiments performed by Smith et al. [2007]. Multiarray recordings from more than 900 neurons that indicate less than 9% of the neurons in the preBötC are expiratory [Carroll et al., 2013] also support this finding. Moreover, our modeling study also provides theoretical support for the respiratory network organization recently proposed by Anderson et al. [2016]. They propose that each phase of the respiratory rhythm is generated by its own excitatory microcircuit located in a different region of the ventral respiratory group, the inspiratory phase being generated by the preBötC, post-inspiration by its own complex (the PiCo) [Anderson et al., 2016], and active expiration by the so-called lateral parafacial/retrotrapezoidal group [Janczewski and Feldman, 2006, Onimaru et al., 2009, Huckstepp et al., 2016]. This idea is similar in spirit to the microcircuit models of Smith et al. [2013], Molkov et al. [2013], Koizumi et al. [2013], Onimaru et al. [2015], which contain more areas. However, each of these excitatory microcircuits contains neurons with different anatomical, physiological and modulatory properties, and each is dependent on excitatory synaptic transmission, able to generate rhythmicity in the absence of synaptic inhibition [Ramirez et al., 2016].

Overall, a modular organization of rhythm generating networks has both evolutionary [Ramirez et al., 2016] and functional implications; the latter may explain, for example, why we can hop on one leg without requiring a major network reconfiguration. We hypothesize that the separation of a rhythmic behavior into several excitatory microcircuits

may indeed be dictated by the architecture of these sparsely connected excitatory networks that generate rhythmicity based on excitatory synaptic mechanisms. The addition of local inhibition to each microcircuit adds another layer of complexity to the generation of rhythms which can affect synchrony and controllability. The lessons learned from the respiratory circuit may also apply to networks that generate locomotion or other rhythmic behaviors, where each phase may be composed of separate microcircuits that are interacting with inhibitory connections.

Acknowledgements

We would like to thank Jonathan Rubin, Peter J. Thomas, Juan Restrepo, Bard Ermentrout, and the reviewers for discussions and suggestions. KD Harris was supported by a Boeing fellowship and an NIH Big Data for Genomics and Neurosciences Training Grant. J Mendoza was supported by a postbaccalaureate fellowship from the University of Washington Institute for Neuroengineering. J-M Ramirez, FAJ Garcia III, J Mendoza, and T Dashevskiy were supported by NIH grants R01 HL126523 and P01 HL090554. We also acknowledge support of a NSF Career Award DMS-1056125 to E Shea-Brown and NIH F32 HL121939 to T Dashevskiy. This work was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington.

Bibliography

- Matt Ainsworth, Shane Lee, Mark O. Cunningham, Roger D. Traub, Nancy J. Kopell, and Miles A. Whittington. Rates and rhythms: A synergistic view of frequency and temporal coding in neuronal networks. *Neuron*, 75(4):572–583, August 2012. ISSN 1097-4199. doi: 10.1016/j.neuron.2012.08.004.
- Tatiana M. Anderson, Alfredo J. Garcia, Nathan A. Baertsch, Julia Pollak, Jacob C. Bloom, Aguan D. Wei, Karan G. Rai, and Jan-Marino Ramirez. A novel excitatory network for the control of breathing. *Nature*, 536(7614):76–80, August 2016. ISSN 0028-0836. doi: 10.1038/nature18944.

- Alex Arenas, Albert Díaz-Guilera, Jurgen Kurths, Yamir Moreno, and Changsong Zhou. Synchronization in complex networks. *Physics Reports*, 469(3):93–153, December 2008. ISSN 03701573. doi: 10.1016/j.physrep.2008.09.002.
- Janet Best, Alla Borisyuk, Jonathan Rubin, David Terman, and Martin Wechselberger. The Dynamic Range of Bursting in a Model Respiratory Pacemaker Network. *SIAM Journal on Applied Dynamical Systems*, 4(4):1107–1139, January 2005. ISSN 1536-0040. doi: 10.1137/050625540.
- Béla Bollobás. Random Graphs. In *Modern Graph Theory*, number 184 in Graduate Texts in Mathematics, pages 215–252. Springer New York, 1998. ISBN 978-0-387-98488-9 978-1-4612-0619-4. doi: 10.1007/978-1-4612-0619-4.7.
- Christoph Börgers and Nancy Kopell. Synchronization in networks of excitatory and inhibitory neurons with sparse, random connectivity. *Neural computation*, 15(3):509–538, 2003.
- T. Graham Brown. The Intrinsic Factors in the Act of Progression in the Mammal. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, 84(572):308–319, 1911. ISSN 0950-1193.
- Nicolas Brunel and Vincent Hakim. Fast Global Oscillations in Networks of Integrate-and-Fire Neurons with Low Firing Rates. *Neural Computation*, 11(7):1621–1671, October 1999. ISSN 0899-7667. doi: 10.1162/089976699300016179.
- Robert J. Butera, John Rinzel, and Jeffrey C. Smith. Models of respiratory rhythm generation in the pre-Bötzinger complex. I. Bursting pacemaker neurons. *Journal of neurophysiology*, 82(1):382–397, 1999a.
- Robert J. Butera, John Rinzel, and Jeffrey C. Smith. Models of respiratory rhythm generation in the pre-Bötzinger complex. II. Populations of coupled pacemaker neurons. *Journal of Neurophysiology*, 82(1):398–415, 1999b.
- Gyorgy Buzsaki. *Rhythms of the Brain*. Oxford University Press, August 2006. ISBN 978-0-19-804125-2.
- Michael S. Carroll and Jan-Marino Ramirez. Cycle-by-cycle assembly of respiratory network activity is dynamic and stochastic. *Journal of Neurophysiology*, 109(2):296–305, January 2013. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.00830.2011.
- Michael S. Carroll, Jean-Charles Viemari, and Jan-Marino Ramirez. Patterns of inspiratory phase-dependent activity in the in vitro respiratory network. *Journal of Neurophysiology*, 109(2):285–295, January 2013. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.00619.2012.

- Yan Cui, Kaiwen Kam, David Sherman, Wiktor A. Janczewski, Yu Zheng, and Jack L. Feldman. Defining preBötzinger Complex Rhythm- and Pattern-Generating Neural Microcircuits In Vivo. *Neuron*, 91(3):602–614, August 2016. ISSN 0896-6273. doi: 10.1016/j.neuron.2016.07.003.
- Christopher A. Del Negro, Consuelo Morgado-Valle, John A. Hayes, Devin D. Mackay, Ryland W. Pace, Erin A. Crowder, and Jack L. Feldman. Sodium and Calcium Current-Mediated Pacemaker Neurons and Respiratory Rhythm Generation. *The Journal of Neuroscience*, 25(2):446–453, December 2005. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.2237-04.2005.
- Christopher A. Del Negro, John A. Hayes, Ryland W. Pace, Benjamin R. Brush, Ryoichi Teruyama, and Jack L. Feldman. Synaptically Activated Burst-Generating Conductances Underlie a Group-Pacemaker Mechanism for Respiratory Rhythm Generation in Mammals. *Progress in Brain Research*, 187:111–136, 2010. ISSN 0079-6123. doi: 10.1016/B978-0-444-53613-6.00008-3.
- Alain Destexhe, Zachary F. Mainen, and Terrence J. Sejnowski. Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism. *Journal of Computational Neuroscience*, 1(3):195–230, August 1994. ISSN 0929-5313, 1573-6873. doi: 10.1007/BF00961734.
- Jack L. Feldman, Christopher A. Del Negro, and Paul A. Gray. Understanding the Rhythm of Breathing: So Near, Yet So Far. *Annual Review of Physiology*, 75(1):423–452, February 2013. ISSN 0066-4278, 1545-1585. doi: 10.1146/annurev-physiol-040510-130049.
- K. A. Ferguson, F. Njap, W. Nicola, F. K. Skinner, and S. A. Campbell. Examining the limits of cellular adaptation bursting mechanisms in biologically-based excitatory networks of the hippocampus. *Journal of Computational Neuroscience*, 39(3):289–309, October 2015. ISSN 0929-5313, 1573-6873. doi: 10.1007/s10827-015-0577-1.
- Chris Gaiteri and Jonathan E. Rubin. The Interaction of Intrinsic Dynamics and Network Topology in Determining Network Burst Synchrony. *Frontiers in Computational Neuroscience*, 5, 2011. ISSN 1662-5188. doi: 10.3389/fncom.2011.00010.
- David Golomb. Neuronal synchrony measures. *Scholarpedia*, 2(1):1347, 2007. ISSN 1941-6016. doi: 10.4249/scholarpedia.1347.
- Paul A. Gray, Wiktor A. Janczewski, Nicholas Mellen, Donald R. McCrimmon, and Jack L. Feldman. Normal breathing requires preBötzinger complex neurokinin-1 receptor-expressing neurons. *Nature Neuroscience*, 4(9):927–930, September 2001. ISSN 1097-6256. doi: 10.1038/nn0901-927.

- Sten Grillner. Biological Pattern Generation: The Cellular and Computational Logic of Networks in Motion. *Neuron*, 52(5):751–766, December 2006. ISSN 0896-6273. doi: 10.1016/j.neuron.2006.11.008.
- Sten Grillner and Thomas M Jessell. Measured motion: Searching for simplicity in spinal locomotor networks. *Current Opinion in Neurobiology*, 19(6):572–586, December 2009. ISSN 0959-4388. doi: 10.1016/j.conb.2009.10.011.
- Claire Guerrier, John A. Hayes, Gilles Fortin, and David Holcman. Robust network oscillations during mammalian respiratory rhythm generation driven by synaptic dynamics. *Proceedings of the National Academy of Sciences*, page 201421997, July 2015. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1421997112.
- N. Hartelt, E. Skorova, T. Manzke, M. Suhr, L. Mironova, S. Kügler, and S.L. Mironov. Imaging of respiratory network topology in living brainstem slices. *Molecular and Cellular Neuroscience*, 37(3):425–431, March 2008. ISSN 10447431. doi: 10.1016/j.mcn.2007.10.011.
- John A. Hayes, Jeffrey L. Mendenhall, Benjamin R. Brush, and Christopher A. Del Negro. 4-Aminopyridine-sensitive outward currents in preBötzinger complex neurons influence respiratory rhythm generation in neonatal mice. *The Journal of Physiology*, 586(7):1921–1936, April 2008. ISSN 1469-7793. doi: 10.1113/jphysiol.2008.150946.
- John A. Hayes, Xueying Wang, and Christopher A. Del Negro. Cumulative lesioning of respiratory interneurons disrupts and precludes motor rhythms in vitro. *Proceedings of the National Academy of Sciences*, 109(21):8286–8291, May 2012. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1200912109.
- Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic block-models: First steps. *Social Networks*, 5(2):109–137, June 1983. ISSN 0378-8733. doi: 10.1016/0378-8733(83)90021-7.
- T. Honore, S. N. Davies, J. Drejer, E. J. Fletcher, P. Jacobsen, D. Lodge, and F. E. Nielsen. Quinoxalinediones: Potent competitive non-NMDA glutamate receptor antagonists. *Science*, 241(4866):701–703, August 1988. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.2899909.
- Robert T. R. Huckstepp, Kathryn P. Cardoza, Lauren E. Henderson, and Jack L. Feldman. Role of Parafacial Nuclei in Control of Breathing in Adult Rats. *Journal of Neuroscience*, 35(3):1052–1067, January 2015. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.2953-14.2015.
- Robert TR Huckstepp, Lauren E. Henderson, Kathryn P. Cardoza, and Jack L. Feldman. Interactions between respiratory oscillators in adult rats. *eLife*, 5:e14203, June 2016. ISSN 2050-084X. doi: 10.7554/eLife.14203.

- Carey Y. L. Huh, Bénédicte Amilhon, Katie A. Ferguson, Frédéric Manseau, Susana G. Torres-Platas, John P. Peach, Stephanie Scodras, Naguib Mechawar, Frances K. Skinner, and Sylvain Williams. Excitatory Inputs Determine Phase-Locking Strength and Spike-Timing of CA1 Stratum Oriens/Alveus Parvalbumin and Somatostatin Interneurons during Intrinsically Generated Hippocampal Theta Rhythm. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 36(25):6605–6622, June 2016. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.3951-13.2016.
- S. Rao Jammalamadaka and Ashis SenGupta. *Topics in Circular Statistics*. World Scientific, April 2001. ISBN 978-981-4495-24-0.
- Wiktor A Janczewski and Jack L Feldman. Distinct rhythm generators for inspiration and expiration in the juvenile rat. *The Journal of Physiology*, 570(Pt 2):407–420, January 2006. ISSN 0022-3751. doi: 10.1113/jphysiol.2005.098848.
- Wiktor A. Janczewski, Alexis Tashima, Paul Hsu, Yan Cui, and Jack L. Feldman. Role of Inhibition in Respiratory Pattern Generation. *The Journal of Neuroscience*, 33(13):5454–5465, March 2013. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.1595-12.2013.
- K. Kam, J. W. Worrell, W. A. Janczewski, Y. Cui, and J. L. Feldman. Distinct Inspiratory Rhythm and Pattern Generating Mechanisms in the preBotzinger Complex. *Journal of Neuroscience*, 33(22):9235–9245, May 2013. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.4143-12.2013.
- Ole Kiehn. Development and functional organization of spinal locomotor circuits. *Current Opinion in Neurobiology*, 21(1):100–109, February 2011. ISSN 0959-4388. doi: 10.1016/j.conb.2010.09.004.
- Hidehiko Koizumi and Jeffrey C. Smith. Persistent Na⁺ and K⁺-Dominated Leak Currents Contribute to Respiratory Rhythm Generation in the Pre-Bötzinger Complex In Vitro. *The Journal of Neuroscience*, 28(7):1773–1785, February 2008. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.3916-07.2008.
- Hidehiko Koizumi, Naohiro Koshiya, Justine X. Chia, Fang Cao, Joseph Nugent, Ruli Zhang, and Jeffrey C. Smith. Structural-Functional Properties of Identified Excitatory and Inhibitory Interneurons within Pre-Bötzinger Complex Respiratory Microcircuits. *The Journal of Neuroscience*, 33(7):2994–3009, February 2013. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.4427-12.2013.
- Nancy Kopell, Mark A. Kramer, Paola Malerba, and Miles A. Whittington. Are Different Rhythms Good for Different Functions? *Frontiers in Human Neuroscience*, 4, November 2010. ISSN 1662-5161. doi: 10.3389/fnhum.2010.00187.

- Amit Lal, Yoshitaka Oku, Swen Hülsmann, Yasumasa Okada, Fumikazu Miwakeichi, Shigeharu Kawai, Yoshiyasu Tamura, and Makio Ishiguro. Dual oscillator model of the respiratory neuronal network generating quantal slowing of respiratory rhythm. *Journal of Computational Neuroscience*, 30(2):225–240, April 2011. ISSN 0929-5313, 1573-6873. doi: 10.1007/s10827-010-0249-0.
- M. S. Lewicki. A review of methods for spike sorting: The detection and classification of neural action potentials. *Network (Bristol, England)*, 9(4):R53–78, November 1998. ISSN 0954-898X.
- S. P. Lieske, M. Thoby-Brisson, P. Telgkamp, and J. M. Ramirez. Reconfiguration of the neural network controlling multiple breathing patterns: Eupnea, sighs and gasps. *Nature neuroscience*, 3(6):600–607, 2000.
- Steven P. Lieske and Jan-Marino Ramirez. Pattern-Specific Synaptic Mechanisms in a Multifunctional Network. I. Effects of Alterations in Synapse Strength. *Journal of Neurophysiology*, 95(3):1323–1333, March 2006. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.00505.2004.
- Bruce G. Lindsey, Ilya A. Rybak, and Jeffrey C. Smith. Computational Models and Emergent Properties of Respiratory Neural Networks. *Comprehensive Physiology*, 2(3):1619–1670, July 2012. ISSN 2040-4603. doi: 10.1002/cphy.c110016.
- Eve Marder and Dirk Bucher. Central pattern generators and the control of rhythmic movements. *Current Biology*, 11(23):R986–R996, November 2001. ISSN 0960-9822. doi: 10.1016/S0960-9822(01)00581-4.
- Naoki Masuda and Kazuyuki Aihara. Global and local synchrony of coupled neurons in small-world networks. *Biological Cybernetics*, 90(4):302–309, April 2004. ISSN 0340-1200, 1432-0770. doi: 10.1007/s00422-004-0471-9.
- Kianoush Missaghi, Jean-Patrick Le Gal, Paul A. Gray, and Réjean Dubuc. The neural control of respiration in lampreys. *Respiratory Physiology & Neurobiology*, 234:14–25, August 2016. ISSN 1878-1519. doi: 10.1016/j.resp.2016.08.007.
- Yaroslav I. Molkov, Bartholomew J. Bacak, Thomas E. Dick, and Ilya A. Rybak. Control of breathing by interacting pontine and pulmonary feedback loops. *Frontiers in Neural Circuits*, 7:16, 2013. doi: 10.3389/fncir.2013.00016.
- Jeffrey D. Moore, Martin Deschênes, Takahiro Furuta, Daniel Huber, Matthew C. Smear, Maxime Demers, and David Kleinfeld. Hierarchy of orofacial rhythms revealed through whisking and breathing. *Nature*, 497(7448):205–210, April 2013. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature12076.

- Consuelo Morgado-Valle, Serapio M. Baca, and Jack L. Feldman. Glycinergic Pacemaker Neurons in PreBötzinger Complex of Neonatal Mouse. *Journal of Neuroscience*, 30(10): 3634–3639, March 2010. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.3040-09.2010.
- Andrés Nieto-Posadas, Ernesto Flores-Martínez, Jonathan-Julio Lorea-Hernández, Ana-Julia Rivera-Angulo, Jesús-Esteban Pérez-Ortega, José Bargas, and Fernando Peña-Ortega. Change in network connectivity during fictive-gasping generation in hypoxia: Prevention by a metabolic intermediate. *Frontiers in Physiology*, 5, July 2014. ISSN 1664-042X. doi: 10.3389/fphys.2014.00265.
- Hiroshi Onimaru, Keiko Ikeda, and Kiyoshi Kawakami. Phox2b, RTN/pFRG neurons and respiratory rhythmogenesis. *Respiratory Physiology & Neurobiology*, 168(1–2):13–18, August 2009. ISSN 1569-9048. doi: 10.1016/j.resp.2009.03.007.
- Hiroshi Onimaru, Kayo Tsuzawa, Yoshimi Nakazono, and Wiktor A. Janczewski. Mid-line section of the medulla abolishes inspiratory activity and desynchronizes pre-inspiratory neuron rhythm on both sides of the medulla in newborn rats. *Journal of Neurophysiology*, 113(7):2871–2878, April 2015. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.00554.2014.
- Nidaa A. Othman, Michael Gallacher, Tarek Z. Deeb, Daniel T. Baptista-Hon, David C. Perry, and Tim G. Hales. Influences on blockade by t-butylbicyclo-phosphoro-thionate of GABAA receptor spontaneous gating, agonist activation and desensitization. *The Journal of Physiology*, 590(1):163–178, January 2012. ISSN 1469-7793. doi: 10.1113/jphysiol.2011.213249.
- Choongseok Park and Jonathan E. Rubin. Cooperation of intrinsic bursting and calcium oscillations underlying activity patterns of model pre-Bötzinger complex neurons. *Journal of Computational Neuroscience*, 34(2):345–366, April 2013. ISSN 0929-5313, 1573-6873. doi: 10.1007/s10827-012-0425-5.
- Fernando Peña, Marjorie A. Parkis, Andrew K. Tryba, and Jan-Marino Ramirez. Differential contribution of pacemaker properties to the generation of respiratory rhythms during normoxia and hypoxia. *Neuron*, 43(1):105–117, 2004.
- L. K. Purvis, J. C. Smith, H. Koizumi, and R. J. Butera. Intrinsic Bursters Increase the Robustness of Rhythm Generation in an Excitatory Network. *Journal of Neurophysiology*, 97(2):1515–1526, November 2006. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.00908.2006.
- J. M. Ramirez, U. J. A. Quellmalz, and B. Wilken. Developmental Changes in the Hypoxic Response of the Hypoglossus Respiratory Motor Output In Vitro. *Journal of Neurophysiology*, 78(1):383–392, July 1997a. ISSN 0022-3077, 1522-1598.

- J. M. Ramirez, P. Telgkamp, F. P. Elsen, U. J. Quellmalz, and D. W. Richter. Respiratory rhythm generation in mammals: Synaptic and membrane properties. *Respiration Physiology*, 110(2-3):71–85, November 1997b. ISSN 0034-5687.
- J. M. Ramirez, S. W. Schwarzacher, O. Pierrefiche, B. M. Olivera, and D. W. Richter. Selective lesioning of the cat pre-Bötzinger complex in vivo eliminates breathing but not gasping. *The Journal of Physiology*, 507(3):895–907, March 1998. ISSN 1469-7793. doi: 10.1111/j.1469-7793.1998.895bs.x.
- Jan-Marino Ramirez and Steven P. Lieske. Commentary on the definition of eupnea and gasping. *Respiratory Physiology & Neurobiology*, 139(1):113–119, December 2003. ISSN 1569-9048. doi: 10.1016/S1569-9048(03)00195-2.
- Jan-Marino Ramirez, Henner Koch, Alfredo J. Garcia, Atsushi Doi, and Sebastien Zanella. The role of spiking and bursting pacemakers in the neuronal control of breathing. *Journal of Biological Physics*, 37(3):241–261, June 2011. ISSN 0092-0606, 1573-0689. doi: 10.1007/s10867-011-9214-z.
- Jan-Marino Ramirez, Tatiana Dashevskiy, Ibis Agosto Marlin, and Nathan Baertsch. Microcircuits in respiratory rhythm generation: Commonalities with other rhythm generating networks and evolutionary perspectives. *Current Opinion in Neurobiology*, 41: 53–61, December 2016. ISSN 0959-4388. doi: 10.1016/j.conb.2016.08.003.
- Jens C. Rekling, Xuesi M. Shao, and Jack L. Feldman. Electrical Coupling and Excitatory Synaptic Transmission between Rhythmogenic Respiratory Neurons in the PreBötzinger Complex. *The Journal of Neuroscience*, 20(23):RC113–RC113, January 2000. ISSN 0270-6474, 1529-2401.
- Diethelm W. Richter and Jeffrey C. Smith. Respiratory Rhythm Generation In Vivo. *Physiology*, 29(1):58–71, January 2014. ISSN 1548-9213, 1548-9221. doi: 10.1152/physiol.00035.2013.
- J. E. Rubin, N. A. Shevtsova, G. B. Ermentrout, J. C. Smith, and I. A. Rybak. Multiple Rhythmic States in a Model of the Respiratory Central Pattern Generator. *Journal of Neurophysiology*, 101(4):2146–2165, February 2009a. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.90958.2008.
- Jonathan E. Rubin, John A. Hayes, Jeffrey L. Mendenhall, and Christopher A. Del Negro. Calcium-activated nonspecific cation current and synaptic depression promote network-dependent burst oscillations. *Proceedings of the National Academy of Sciences*, 106(8):2939–2944, 2009b.
- Jonathan E. Rubin, Bartholomew J. Bacak, Yaroslav I. Molkov, Natalia A. Shevtsova, Jeffrey C. Smith, and Ilya A. Rybak. Interacting oscillations in neural control of breathing:

- Modeling and qualitative analysis. *Journal of Computational Neuroscience*, 30(3):607–632, June 2011. ISSN 0929-5313, 1573-6873. doi: 10.1007/s10827-010-0281-0.
- Susan J. Sara. The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews Neuroscience*, 10(3):211–223, March 2009. ISSN 1471-003X. doi: 10.1038/nrn2573.
- David J. Schwab, Robijn F. Bruinsma, Jack L. Feldman, and Alex J. Levine. Rhythmogenic neuronal networks, emergent leaders, and k-cores. *Physical Review E*, 82(5), November 2010. ISSN 1539-3755, 1550-2376. doi: 10.1103/PhysRevE.82.051911.
- Stephan W. Schwarzacher, Udo Rüb, and Thomas Deller. Neuroanatomical characteristics of the human pre-Bötzinger complex and its involvement in neurodegenerative brainstem diseases. *Brain*, 134(1):24–35, January 2011. ISSN 0006-8950, 1460-2156. doi: 10.1093/brain/awq327.
- X. M. Shao and J. L. Feldman. Respiratory rhythm generation and synaptic inhibition of expiratory neurons in pre-Bötzinger complex: Differential roles of glycinergic and GABAergic neural transmission. *Journal of Neurophysiology*, 77(4):1853–1860, April 1997. ISSN 0022-3077.
- A. A. Sharp, F. K. Skinner, and E. Marder. Mechanisms of oscillation in dynamic clamp constructed two-cell half-center circuits. *Journal of Neurophysiology*, 76(2):867–883, August 1996. ISSN 0022-3077, 1522-1598.
- David Sherman, Jason W. Worrell, Yan Cui, and Jack L. Feldman. Optogenetic perturbation of preBotzinger complex inhibitory neurons modulates respiratory pattern. *Nature Neuroscience*, 18(3):408–414, March 2015. ISSN 1097-6256. doi: 10.1038/nn.3938.
- Frances K. Skinner. Cellular-based modeling of oscillatory dynamics in brain networks. *Current Opinion in Neurobiology*, 22(4):660–669, August 2012. ISSN 1873-6882. doi: 10.1016/j.conb.2012.02.001.
- J. C. Smith, H. H. Ellenberger, K. Ballanyi, D. W. Richter, and J. L. Feldman. Pre-Bötzinger complex: A brainstem region that may generate respiratory rhythm in mammals. *Science*, 254(5032):726–729, January 1991. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1683005.
- Jeffrey C. Smith, A. P. L. Abdala, Hidehiko Koizumi, Ilya A. Rybak, and Julian FR Paton. Spatial and functional architecture of the mammalian brain stem respiratory network: A hierarchy of three oscillatory mechanisms. *Journal of neurophysiology*, 98(6):3370–3387, 2007.
- Jeffrey C. Smith, Ana P.L. Abdala, Anke Borgmann, Ilya A. Rybak, and Julian F.R. Paton. Brainstem respiratory networks: Building blocks and microcircuits. *Trends in Neurosciences*, 36(3):152–162, March 2013. ISSN 01662236. doi: 10.1016/j.tins.2012.11.004.

- Douglas G. Stuart and Hans Hultborn. Thomas Graham Brown (1882–1965), Anders Lundberg (1920–), and the neural control of stepping. *Brain Research Reviews*, 59(1): 74–95, November 2008. ISSN 0165-0173. doi: 10.1016/j.brainresrev.2008.06.001.
- Wenbin Tan, Wiktor A. Janczewski, Paul Yang, Xuesi M. Shao, Edward M. Callaway, and Jack L. Feldman. Silencing preBötzinger Complex somatostatin-expressing neurons induces persistent apnea in awake rat. *Nature Neuroscience*, 11(5):538–540, May 2008. ISSN 1097-6256. doi: 10.1038/nn.2104.
- Wenbin Tan, Silvia Pagliardini, Paul Yang, Wiktor A. Janczewski, and Jack L. Feldman. Projections of preBötzinger Complex neurons in adult rats. *The Journal of Comparative Neurology*, 518(10):1862–1878, May 2010. ISSN 00219967, 10969861. doi: 10.1002/cne.22308.
- Muriel Thoby-Brisson and Jan-Marino Ramirez. Identification of Two Types of Inspiratory Pacemaker Neurons in the Isolated Respiratory Neural Network of Mice. *Journal of Neurophysiology*, 86(1):104–112, July 2001. ISSN 0022-3077, 1522-1598.
- Natalia Toporikova and Robert J. Butera. Two types of independent bursting mechanisms in inspiratory neurons: An integrative model. *Journal of Computational Neuroscience*, 30(3):515–528, June 2011. ISSN 0929-5313, 1573-6873. doi: 10.1007/s10827-010-0274-z.
- Xueying Wang, John A. Hayes, Ann L. Revill, Hanbing Song, Andrew Kottick, Nikolas C. Vann, M. Drew LaMar, Maria Cristina D. Picardo, Victoria T. Akins, Gregory D. Funk, and others. Laser ablation of Dbx1 neurons in the pre-Bötzinger complex stops inspiratory rhythm and impairs output in neonatal mice. *eLife*, 3:e03427, 2014.
- Stefan M. Winter, Jens Fresemann, Christian Schnell, Yoshitaka Oku, Johannes Hirrlinger, and Swen Hülsmann. Glycinergic interneurons are functionally integrated into the inspiratory network of mouse medullary slices. *Pflügers Archiv - European Journal of Physiology*, 458(3):459–469, July 2009. ISSN 0031-6768, 1432-2013. doi: 10.1007/s00424-009-0647-1.
- Steffen Wittmeier, Gang Song, James Duffin, and Chi-Sang Poon. Pacemakers handshake synchronization mechanism of mammalian respiratory rhythmogenesis. *Proceedings of the National Academy of Sciences*, 105(46):18000–18005, 2008.

Chapter 5

CONCLUSION

We have presented three distinct but related approaches to studying network structure, with an emphasis on neuroscientific applications.

In Chapter 2, we highlight a statistical inference method which recovers the weighted connection matrix of the whole brain from tracing data. So far, the tracing data that researchers collect for mouse [Oh et al., 2014], rat [Bota and Swanson, 2007], and marmoset [Okano et al., 2015, 2016], have been analyzed at the coarse regional level. The tools we develop are like a magnifying glass on top of this, revealing important bumps and cranies in the fabric of projections. Longer term, next-generation electron microscopy data offer a bridge between single-neuron microscale connectivity and the mesoscale networks we analyze today. In the next 10 years, more connectivity datasets will become available, formats will standardize, and tracing techniques will improve. In all cases, inference methods like the ones we developed are essential for accessing the underlying connectome. And the methods we use to understand connectomes—network science [Sporns, 2010], topographic organization, dynamical models, and hopefully new paradigms—are crucial.

These mesoscopic models can be tied to the microscopic level by viewing them as priors on the average levels of connectivity, i.e. the average degree, of connections between neurons in different spatial positions. This leads naturally to the kind of random graph models studied in Chapter 3. These types of models are already being studied empirically as models of cortical columns [Markram et al., 2015, Reimann et al., 2015, Cain et al., 2016, Reimann et al., 2017, Schuecker et al., 2017] in attempts to understand “canonical” cortical computations [Miller, 2016]. Random graphs have also been successful in pre-

dicting the properties of feedforward neuronal networks [Marr, 1969, Albus, 1971, Babadi and Sompolinsky, 2014, Litwin-Kumar et al., 2017, Cayco-Gajic et al., 2017]. The spectral properties of these models are important for understanding the behavior of dynamical systems on these graphs. For instance, the spectral gap is related to whether a network of oscillators will synchronize. Furthermore, we have seen that random graphs lead to interesting mathematical problems, and that they are much more generally applicable in computer science, machine learning, etc.

In Chapter 4, we model the respiratory rhythm using a biophysical dynamical system on a random network. We show how the structure of that network, both the composition of inhibitory and excitatory cells as well as the average connectivity, drives a transition from a synchronous to an irregular state. This seems promising for explaining the role of the “anomalous” inhibitory cells there. Understanding synchrony is important for pattern generation and other brain rhythms [Buzsaki, 2006, Kopell et al., 2010, Ainsworth et al., 2012]. However, there is also a huge amount of interest in the apparently asynchronous and chaotic state exhibited by cortical networks, which seems to arise from plasticity rules that maintain excitation/inhibition balance [Ginzburg and Sompolinsky, 1994, Sussillo and Abbott, 2009, 2012, Renart et al., 2010, Deco et al., 2014, Markram et al., 2015, among others].

Determining what kinds of computation are possible in these different dynamical states, as well as how they are maintained and modulated through plasticity and learning, remains a crucial area for computational neuroscience. The role of the network is central to it all.

Bibliography

Matt Ainsworth, Shane Lee, Mark O. Cunningham, Roger D. Traub, Nancy J. Kopell, and Miles A. Whittington. Rates and rhythms: A synergistic view of frequency and temporal coding in neuronal networks. *Neuron*, 75(4):572–583, August 2012. ISSN 1097-4199. doi: 10.1016/j.neuron.2012.08.004.

- James S. Albus. A theory of cerebellar function. *Mathematical Biosciences*, 10(1–2):25–61, February 1971. ISSN 0025-5564. doi: 10.1016/0025-5564(71)90051-4.
- Baktash Babadi and Haim Sompolinsky. Sparseness and Expansion in Sensory Representations. *Neuron*, 83(5):1213–1226, September 2014. ISSN 0896-6273. doi: 10.1016/j.neuron.2014.07.035.
- Mihail Bota and Larry W. Swanson. Online workbenches for neural network connections. *The Journal of Comparative Neurology*, 500(5):807–814, February 2007. ISSN 1096-9861. doi: 10.1002/cne.21209.
- Gyorgy Buzsaki. *Rhythms of the Brain*. Oxford University Press, August 2006. ISBN 978-0-19-804125-2.
- Nicholas Cain, Ramakrishnan Iyer, Christof Koch, and Stefan Mihalas. The Computational Properties of a Simplified Cortical Column Model. *PLOS Computational Biology*, 12(9):e1005045, September 2016. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005045.
- N. Alex Cayco-Gajic, Claudia Clopath, and R. Angus Silver. Sparse synaptic connectivity is required for decorrelation and pattern separation in feedforward networks. *Nature Communications*, 8(1):1116, October 2017. ISSN 2041-1723. doi: 10.1038/s41467-017-01109-y.
- Gustavo Deco, Adrián Ponce-Alvarez, Patric Hagmann, Gian Luca Romani, Dante Mantini, and Maurizio Corbetta. How Local Excitation–Inhibition Ratio Impacts the Whole Brain Dynamics. *The Journal of Neuroscience*, 34(23):7886–7898, April 2014. ISSN 0270-6474, 1529-2401. doi: 10.1523/JNEUROSCI.5068-13.2014.
- Iris Ginzburg and Haim Sompolinsky. Theory of correlations in stochastic neural networks. *Physical Review E*, 50(4):3171–3191, October 1994. doi: 10.1103/PhysRevE.50.3171.
- Nancy Kopell, Mark A. Kramer, Paola Malerba, and Miles A. Whittington. Are Different Rhythms Good for Different Functions? *Frontiers in Human Neuroscience*, 4, November 2010. ISSN 1662-5161. doi: 10.3389/fnhum.2010.00187.
- Ashok Litwin-Kumar, Kameron Decker Harris, Richard Axel, Haim Sompolinsky, and L. F. Abbott. Optimal Degrees of Synaptic Connectivity. *Neuron*, 93(5):1153–1164.e7, March 2017. ISSN 0896-6273. doi: 10.1016/j.neuron.2017.01.030.
- Henry Markram, Eilif Muller, Srikanth Ramaswamy, Michael W. Reimann, Marwan Abdellah, Carlos Aguado Sanchez, Anastasia Ailamaki, Lidia Alonso-Nanclares, Nicolas Antille, Selim Arsever, Guy Antoine Atenekeng Kahou, Thomas K. Berger, Ahmet Bilgili, Nenad Buncic, Athanassia Chalimourda, Giuseppe Chindemi, Jean-Denis Courcol, Fabien Delalondre, Vincent Delattre, Shaul Druckmann, Raphael Dumusc,

- James Dynes, Stefan Eilemann, Eyal Gal, Michael Emiel Gevaert, Jean-Pierre Ghobril, Albert Gidon, Joe W. Graham, Anirudh Gupta, Valentin Haenel, Etay Hay, Thomas Heinis, Juan B. Hernando, Michael Hines, Lida Kanari, Daniel Keller, John Kenyon, Georges Khazen, Yihwa Kim, James G. King, Zoltan Kisvarday, Pramod Kumbhar, Sébastien Lasserre, Jean-Vincent Le Bé, Bruno R. C. Magalhães, Angel Merchán-Pérez, Julie Meystre, Benjamin Roy Morrice, Jeffrey Muller, Alberto Muñoz-Céspedes, Shruti Muralidhar, Keerthan Muthurasa, Daniel Nachbaur, Taylor H. Newton, Max Nolte, Aleksandr Ovcharenko, Juan Palacios, Luis Pastor, Rodrigo Perin, Rajnish Ranjan, Imad Riachi, José-Rodrigo Rodríguez, Juan Luis Riquelme, Christian Rössert, Konstantinos Sfyarakis, Ying Shi, Julian C. Shillcock, Gilad Silberberg, Ricardo Silva, Farhan Tauheed, Martin Telefont, Maria Toledo-Rodriguez, Thomas Tränkler, Werner Van Geit, Jafet Villafranca Díaz, Richard Walker, Yun Wang, Stefano M. Zaninetta, Javier DeFelipe, Sean L. Hill, Idan Segev, and Felix Schürmann. Reconstruction and Simulation of Neocortical Microcircuitry. *Cell*, 163(2):456–492, October 2015. ISSN 0092-8674. doi: 10.1016/j.cell.2015.09.029.
- David Marr. A theory of cerebellar cortex. *The Journal of Physiology*, 202(2):437–470.1, June 1969. ISSN 0022-3751.
- Kenneth D Miller. Canonical computations of cerebral cortex. *Current Opinion in Neurobiology*, 37:75–84, April 2016. ISSN 0959-4388. doi: 10.1016/j.conb.2016.01.008.
- Seung Wook Oh, Julie A. Harris, Lydia Ng, Brent Winslow, Nicholas Cain, Stefan Mihalas, Quanxin Wang, Chris Lau, Leonard Kuan, Alex M. Henry, Marty T. Mortrud, Benjamin Ouellette, Thuc Nghi Nguyen, Staci A. Sorensen, Clifford R. Slaughterbeck, Wayne Wakeman, Yang Li, David Feng, Anh Ho, Eric Nicholas, Karla E. Hirokawa, Phillip Bohn, Kevin M. Joines, Hanchuan Peng, Michael J. Hawrylycz, John W. Phillips, John G. Hohmann, Paul Wahnoutka, Charles R. Gerfen, Christof Koch, Amy Bernard, Chinh Dang, Allan R. Jones, and Hongkui Zeng. A mesoscale connectome of the mouse brain. *Nature*, 508(7495):207–214, April 2014. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature13186.
- Hideyuki Okano, Atsushi Miyawaki, and Kiyoto Kasai. Brain/MINDS: Brain-mapping project in Japan. *Phil. Trans. R. Soc. B*, 370(1668):20140310, May 2015. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2014.0310.
- Hideyuki Okano, Erika Sasaki, Tetsuo Yamamori, Atsushi Iriki, Tomomi Shimogori, Yoko Yamaguchi, Kiyoto Kasai, and Atsushi Miyawaki. Brain/MINDS: A Japanese National Brain Project for Marmoset Neuroscience. *Neuron*, 92(3):582–590, November 2016. ISSN 0896-6273. doi: 10.1016/j.neuron.2016.10.018.
- Michael W. Reimann, James G. King, Eilif B. Muller, Srikanth Ramaswamy, and Henry Markram. An algorithm to predict the connectome of neural microcircuits. *Frontiers in*

Computational Neuroscience, 9, October 2015. ISSN 1662-5188. doi: 10.3389/fncom.2015.00120.

Michael W. Reimann, Max Nolte, Martina Scolamiero, Katharine Turner, Rodrigo Perin, Giuseppe Chindemi, Paweł Dłotko, Ran Levi, Kathryn Hess, and Henry Markram. Cliques of Neurons Bound into Cavities Provide a Missing Link between Structure and Function. *Frontiers in Computational Neuroscience*, 11, 2017. ISSN 1662-5188. doi: 10.3389/fncom.2017.00048.

Alfonso Renart, Jaime de la Rocha, Peter Bartho, Liad Hollender, Néstor Parga, Alex Reyes, and Kenneth D. Harris. The Asynchronous State in Cortical Circuits. *Science*, 327 (5965):587–590, January 2010. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1179850.

Jannis Schuecker, Maximilian Schmidt, Sacha J. van Albada, Markus Diesmann, and Moritz Helias. Fundamental Activity Constraints Lead to Specific Interpretations of the Connectome. *PLOS Computational Biology*, 13(2):e1005179, February 2017. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005179.

Olaf Sporns. *Networks of the Brain*. The MIT Press, 1st edition, 2010. ISBN 0-262-01469-6 978-0-262-01469-4.

David Sussillo and L. F. Abbott. Generating Coherent Patterns of Activity from Chaotic Neural Networks. *Neuron*, 63(4):544–557, August 2009. ISSN 0896-6273. doi: 10.1016/j.neuron.2009.07.018.

David Sussillo and L. F. Abbott. Transferring Learning from External to Internal Weights in Echo-State Networks with Sparse Connectivity. *PLOS ONE*, 7(5):e37372, May 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0037372.

LIST OF FIGURES

Figure Number		Page
1.1	An illustration of cerebellar neurons by Ramón y Cajal [1906]. Arrows indicate the direction of spike transmission from dendrites to axons. Cajal inferred this correct directionality from the geometry of the neurons alone, before detailed electrophysiological properties of neurons were known. If all of the cerebellar neurons in this area were pictured, we would see a jungle of overlapping axons, dendrites, and cell bodies.	2
1.2	A small directed graph. It contains a connected component of 5 nodes and one isolated node. The majority of the edges are directed. There is one undirected edge, shown without arrows. There is also a self-loop.	4
1.3	Example of meso-scale connectivity data collected via viral tracing. In this experiment, the tracer was injected in the right hemisphere. Axonal projections of neurons whose cell bodies are within the injection site are visible as green fluorescence on both the ipsilateral and contralateral sides. These images show just one of many coronal slices which are stitched together into a 3-D image of fluorescence intensity. These data were collected by the Allen Institute for Brain Science [Oh et al., 2014] and are available online at http://connectivity.brain-map.org	7
2.1	A , We seek to fit a matrix W which reproduces neural tracing experiments. Each column of W represents the expected signal in target voxels given an injection of one unit into a single source voxel. B , In the work of Oh et al. [2014], a regionally homogeneous connectivity matrix was fit using a predefined regional parcellation to constrain the problem. We propose that smoothness of W is a better prior. C , The mouse’s visual field can be represented in azimuth/altitude coordinates. This representation is maintained in the retinotopy, a smoothly varying map replicated in many visual areas (e.g. [Garrett et al., 2014]). D , Assuming locations in VISp (the primary visual area) project most strongly to positions which represent the same retinotopic coordinates in a secondary visual area, then we expect the mapping between upstream and downstream visual areas to be smooth. . .	19

- 2.2 Comparison of the true (**Above left**) and inferred connectivity from 5 injections. Unless noted, $\lambda = 100$. **Above right**, we show the what happens when we solve (P1) without the matrix completion term P_Ω . The holes in the projection data cause patchy and incorrect output. Note the colorbar range is $6\times$ that in the other cases. **Below left** is the result with P_Ω but without regularization, solving (P1) for $\lambda = 0$. There, the solution does not interpolate between injections. **Below right** is a rank $r = 20$ result using (P2), which captures the diagonal band and off-diagonal bump that make up W_{true} . In this case, the low rank result has less relative error (9.6%) than the full rank result (11.1%, not shown). 26
- 2.3 Inferred connectivity using all 28 selected injections from visual system data. **Above left**, Projections from a source voxel (blue) located in VISp to all other voxels in the visual areas. The view is integrated over the superior-inferior axis. The connectivity shows strong local connections and weaker connections to higher areas, in particular VISam, VISal, and VISl. Movies of the inferred connectivity (full, low rank, and the low rank residual) for varying source voxel are available in the supplementary material. **Above right**, For a source 800 μm away, the pattern of anterograde projections is similar, but the distal projection centers are shifted, as expected from retinotopy. **Below**, The residuals between the full rank and rank 160 result from solving (P2), for the same source voxel as in the center. The residuals are an order of magnitude less than typical features of the connectivity. 30
- 3.1 The structure of every bipartite, biregular graph. There are $n = |V_1|$ left vertices, $m = |V_2|$ right vertices, each of degree d_1 and d_2 , with the constraint that $nd_1 = md_2$. The distribution $\mathcal{G}(n, m, d_1, d_2)$ is taken uniformly over all such graphs. 39
- 3.2 An example circuit that contributes to the trace in Eqn. (3.2), for $k = 2$ and $\ell = 2$. Edges are numbered as they occur in the circuit. Each segment $\{\gamma_i\}_{i=1}^4$ is of length $\ell + 1 = 3$ and made up of edges $3(i - 1) + 1$ through $3i$. The last edge of each γ_i is the first edge of γ_{i+1} , and these are shown in purple. Every path γ_i with i even follows the edges backwards due to the matrix transpose. However, this detail turns out not to make any difference since the underlying graph is undirected. Our example has no cycles in each segment for clarity, but, in general, each segment can have up to one cycle, and the overall circuit may be tangled. 51

- 3.3 The different cases of Lemma 8, which concerns $P(e \in G|H)$: the probability of an edge e existing in the random graph G , given an observed subgraph H , shown in green. The new edge e connects at vertex v . All the green vertices lie in H . When the connecting vertex has degree one in subgraph H , i.e. $h_v = 1$ and case (ii), it induces a two-path. The probability of that edge is closer to the entries of the centering matrix S if $h_v > 1$, case (i). 54
- 3.4 Encoding an ℓ -tangle-free walk, in this case the first walk in the circuit γ_1 , when it contains a cycle. The vertices and edges are labeled in the order of their traversal. The segments γ^a , γ^b , and γ^c occur on edges numbered $(1, 2, 3); (4 + 6i, 5 + 6i, 6 + 6i, 7 + 6i, 8 + 6i, 9 + 6i)$ for $i = 0, 1, \dots, c$; and $(10 + 6c)$, respectively. The encoding is $(0, 3, 0)|(0, 4, 3)(4, 0, 0)|(0, 1, 0)$. Suppose $c = 1$. Then $\ell = 22$ and the encoding is of length $3 + (4 + 1 + 1)(c + 1) + 1$, we can back out c to find that the cycle is repeated twice. The encodings become more complicated later in the circuit as vertices see repeat visits. 65
- 3.5 Schematic and realization of a random regular frame graph. **A**, the frame graph. The vertices of the frame (red = A, green = B, blue = C) are weighted according to their proportions p in the random regular frame graph. The edge weights D_{ij} set the between-class vertex degrees in the random regular frame graph. This frame will yield a random tripartite graph. **B**, realization of the graph on 72 vertices. In this instance, there are $1/8 \times 72 = 9$ green and red vertices and $3/4 \times 72 = 54$ blue vertices. Each blue vertex connects to $k_{CA} = 1$ red vertex and $k_{CB} = 2$ green vertices. This is actually a multigraph; with so few vertices, the probability that the configuration model algorithm yields parallel edges is high. 76
- 4.1 **With higher fraction of inhibitory cells, synchrony and burst amplitude decrease, and the integrated timeseries becomes more variable.** Three simulations of the respiratory network model: **A**, $p_I = 0\%$; **B**, $p_I = 20\%$; **C**, $p_I = 40\%$. Above, we show the integrated trace, which is a lowpass-filtered average of the spiking activity of all $N = 300$ neurons in the network. Below, we show the spike raster of individual neuron activity. In all cases, $k_{\text{avg}} = 6$, $g_E = g_I = 2.0$ nS. Detected bursts are marked by open circles on the integrated traces. At lower levels of synchrony, as in part C, what constitutes a burst becomes ambiguous. 112

- 4.2 Synchrony decreases with inhibition and sparsity. The highest variability across networks occurs at the synchronization boundary.** **A**, Synchrony parameter χ averaged over 8 network realizations, plotted versus the amount of connections k_{avg} and the fraction of inhibitory neurons p_I . **B**, Standard deviation of χ over network realizations. Higher standard deviation indicates that the synchrony is not reliable for different networks with those parameters. The area of highest standard deviation occurs at the boundary of low and high synchrony, $\chi \approx 0.5$. This is indicative of a phase transition between synchronized and desynchronized states. 113
- 4.3 Expiratory cells arise from inhibition, but can only occupy a minority without disrupting the inspiratory rhythm.** **A**, Neuron phase-locking variables for the simulation in Fig. 4.1B ($k_{\text{avg}} = 6$, $p_I = 20\%$). Each neuron has an associated complex number z_i with $0 \leq |z_i| \leq 1$. The magnitude $|z_i|$ is plotted against angle $\arg z_i$. These are used to define inspiratory, expiratory, and tonic neurons via the labeled regions separated by the dashed lines. **B**, Expiratory (anti-phase with main rhythm) neurons as a function of network parameters k_{avg} and p_I . The fraction of expiratory neurons increases with inhibition or as the connectivity becomes weaker. The blue indicates the absence of any overall rhythm, defined as $\chi < 0.25$. **C**, An example of a simulation with two-phase activity, with $k_{\text{avg}} = 6$, $p_I = 30\%$, $g_E = 5.0$, and $g_I = 2.0$. A minority of neurons produce a reliable, small bump after every burst. It is aligned near 0.7π , so it is more of a post-inspiratory or pre-expiratory burst. These expiratory cells are rebound bursting after being disinhibited. This is similar to the “handshake” mechanism of Wittmeier et al. [2008]. However, this type of two-phase rhythm is very rare in simulations. 115
- 4.4 Example rasters of expiratory, tonic, and inspiratory cells.** Expiratory cells exhibit lower firing rates than inspiratory ones, similar to the typical tonic firing observed in slices. As shown, tonic classified cells can be bursting so long as their bursts do not occur reliably at any given phase. The inspiratory cells shown are a random subset. Data are for a representative network with $k_{\text{avg}} = 6$, $p_I = 20\%$ (same as Figs. 4.1B and 4.2B). 116

- 4.5 Expiratory cells preferentially receive input from other excitatory, expiratory cells and inhibitory, inspiratory cells.** **B**, The top row shows the distribution of inputs, excitatory on left and inhibitory on right, colored by whether the receiving cell is inspiratory (black bars) or expiratory (white bars). Expiratory cells receive less excitatory and more inhibitory connections than inspiratory cells. The center and bottom rows breaks down these inputs by the phase of the presynaptic neuron, inspiratory inputs shown in the center and expiratory below. Expiratory cells preferentially receive excitatory input from other expiratory cells (compare middle left and bottom left). Furthermore, inhibitory input to expiratory cells tends to come from inspiratory cells rather than other expiratory cells (middle right and bottom right). Data are for a representative network with $k_{\text{avg}} = 6$, $p_I = 20\%$ (same as Figs. 4.1B and 4.2B). There were 251 inspiratory, 23 expiratory, 15 tonic, and 11 silent cells. 117
- 4.6 A**, Schematic of the two population network. The average amount of excitatory connections on average are fixed, but we vary the expected intra-group and inter-group inhibition k_{intra} and k_{inter} . **B** and **C**, Two simulations of the network with different parameters: Each simulation also has a schematic on the right demonstrating the differences in inhibitory network strengths. A thicker line indicates more connections, and a darker color indicates a more reliable rhythm. Panel B shows the case $k_{\text{intra}} = 1.0$ and $k_{\text{inter}} = 4.0$. Panel C depicts $k_{\text{intra}} = 2.0$ and $k_{\text{inter}} = 1.5$. There, we observe less reliable rhythms, with decreased phase order Ω and decreased synchrony χ , despite approximately equal average phase difference Φ . **D**, Average synchrony over 8 realizations for each k_{inter} and k_{intra} pair. Higher values of χ occur above the diagonal $k_{\text{inter}} = k_{\text{intra}}$ line. **E**, Average phase difference Φ of rhythmic bursts between the two populations. No clear trends are evident, and the value is close to $\Phi = 0.5$, perfectly out-of-phase, in much of the region. **F**, Average phase order Ω . Higher phase order indicates the relative phase of bursts in Pop. 1 and Pop. 2, i.e. Φ in panel E, are reliable. The phase order appears to be proportional to the synchrony, with the highest values above the diagonal. Star and circle symbols in D–F are the network parameters used to produce the rhythms in panels B and C. 120

- 4.7 In vitro array recordings from transverse slice preparations exhibit partial synchrony.** We performed in vitro preBötC slice experiments, where we measured the rhythm in control, partial excitation block (DNQX 0.7 μ M), and partial excitation block with full inhibition block (DNQX 0.7 μ M + PTX 20 μ M). We record the preBötC population activity with a large electrode (LFP, arb. unit) as well as individual neurons in the contralateral area using an array. The average activity of the isolated units is also shown (x^{int} , Hz/cell). **A**, Control conditions show a robust population rhythm with some amplitude and period irregularity. **B**, Partial excitation block using DNQX degrades the population synchrony, with decreased burst amplitude, slower rhythm, and more irregular intervals between bursts. **C**, Blocking inhibition with PTX allows the rhythm to recover toward control conditions. **D**, Synchrony in the model, as a function of excitatory and inhibitory synaptic conductances g_E and g_I , increases with stronger excitation and decreases with stronger inhibition, similar to varying connectivity k_{avg} and inhibitory fraction p_I . Arrows indicate the presumed effects of DNQX and PTX on the model. **E**, Measurements of synchrony from our 4 array recording experiments. Synchrony takes intermediate values in all conditions, decreasing with DNQX and recovering after PTX. 124
- 4.8 Modulation of inhibition and excitation changes the rhythm in comparable ways for experiments and the model. (Above: model)** The effect of changing conductances g_E and g_I . Burst amplitude and period irregularity decrease with stronger excitation and weaker inhibition. Both of these measures are negatively correlated to the population synchrony, shown in Fig. 4.7D. **(Below: experiments)** This plot summarizes 17 experiments. We extracted bursts from the LFP and measured the amplitude and frequency irregularity of those rhythms. Amplitude irregularity showed no significant trends across conditions. However, period irregularity showed a significant increase from control with DNQX, a decrease from DNQX to DNQX+PTX, and a small increase between control and DNQX+PTX. See Table 4.2 for the full output of the statistical tests. 128
- 4.9 The effect of DNQX and PTX on in vitro rhythm amplitude and period,** similar to Figure 4.8. Amplitude decreases with DNQX while period increases, with both recovering to near baseline after addition of PTX. See Table 4.2 for the result of statistical tests on this data. 130

LIST OF TABLES

Table Number	Page
2.1 Model performance on Allen Institute Mouse Brain Connectivity Atlas data. Cross-validation errors of the voxel model (P1) and regionally homogeneous models are shown, with training errors in parentheses. The errors are computed in both voxel space and regional space, using the relative mean squared error MSE_{rel} , Eqn. (2.4). In either space, the voxel model shows reduced training and cross-validation errors relative to the regional model.	28
4.1 Parameters for the network model are taken from the literature [Butera et al., 1999a, Park and Rubin, 2013]. We modify g_L for quiescent (Q), tonic spiking (TS), and intrinsically bursting (B) cells. The system of equations is simulated in the given units, so that no conversions are necessary. Those parameters below the lower horizontal break are for the synaptic dynamics.	100
4.2 Statistical results for in vitro measurements of amplitude irregularity, period irregularity, amplitude, and period. We report the estimated fixed effect for the intercept, DNQX, and DNQX+PTX conditions, as well as standard error (SE), degrees of freedom (DF), t value, and p value for each effect. These data summarize 17 LFP recordings save the synchrony fit, which comes from 4 multielectrode recordings.	129