Robustness of Evolutionary and Glassy Systems

Vaibhav Mohanty

Rudolf Peierls Centre for Theoretical Physics

University of Oxford

This dissertation is submitted for the degree of

Doctor of Philosophy in Theoretical Physics

October 2021

To my parents, Sangeeta and Bidyut

In memory of my grandmother, Manorama (1928-2022)

Acknowledgements

I am first and foremost thankful to my advisor, Ard Louis. Not only has Ard been an inspiration to me as a scientist, but he has really has embodied what it means to be an incredible mentor and a thoughtful advocate for his students for which I am incredibly grateful. He has given me insightful guidance and ample room to grow as a thinker and a scientist. This DPhil could not have happened without you.

Science is meant to be collaborative, and this work would not have been possible without incredible collaborators, including Kamal Dingle, Sam Greenbury, Tasmin Sarkany, Shyam Narayanan, and Yoonsoo Nam. Thank you all!

I am also grateful to the Louis group members, former and current, who along with Ard have created a welcoming environment. In addition to Tasmin and Yoonsoo, I am thankful to Guillermo Valle-Pérez, Chris Mingard, Vuk Radovic, Shuofeng Zhang, and Wilber Lim for fruitful and interesting discussions. I would also like to thank Mehrana Nejad, Shuofeng, and Wilber for being wonderful officemates in the Beecroft Buildling, and am I grateful for the wonderful discussions and conversations that emerged from working alongside each other. I also thank Alex Kelser, Sloan Nietert, and Alex Wei for fruitful discussions.

I am incredibly grateful to my parents, the rest of my family and loved ones, and friends. Their unwavering support throughout this time—and especially during the pandemic—has meant the world to me. The Christ Church and Marshall communities especially have given me the experience of a lifetime!

Lastly, my heart goes out to all who have been affected by the COVID-19 pandemic, especially to those who have lost loved ones. Part of this thesis was inspired by the challenges faced during this ongoing pandemic and the vast uncertainty surrounding it. I hope a better understanding of biological evolution can aid the effort in combating this pandemic and preventing future ones.

This DPhil was supported by a Marshall Scholarship (2019-2021) from the Marshall Aid Commemoration Commission. I am incredibly grateful for the opportunity and community the Marshall offered.

Thesis submitted 2021, viva voce 2022.

Abstract

Systems which take in a sequence-based input to produce a nontrivial output are ubiquitous across science. One key property of such maps is the robustness of the output upon random changes of the inputs. Interestingly, recent work on genotype-phenotype maps in biology has shown that the robustness, defined as the probability that a point mutation to the genotype does not change the phenotypic output, is typically exponentially higher than a null expectation based on a random input-output pairings. This high robustness must arise from correlations in the input-output map. In this thesis, we investigate the origins of these correlations.

We first study spin glasses, defining the inputs as the set of interactions, and the outputs as the ground-state spin configurations. We find that the robustness in this generic model exhibits behaviour that is consistent with the scaling laws observed for the biological genotype-phenotype maps, suggesting that the robustness behaviour has a more universal origin.

Next, we use a maximum-entropy approach to study generic sequence-based input-output maps on Hamming graphs. If we constrain the global robustness, we find that there are two states, as a function of the strength of the constraint. The fragile phase is similar to the null-model based on completely random pairings, and has low robustness. By contrast, for a stronger constraint, there is a rather sudden change to a different correlated behaviour. Interestingly, this very simple model exhibits scaling of the robustness, the distribution of neutral component sizes, and the probability of obtaining a different output upon mutations that are remarkably similar to those found for a series of genotype-phenotype maps. Finding these detailed results in such a simple model suggests that they have near-universal origins in GP maps, and possibly in a wider set of real-world mappings.

Maximally robust neutral networks are then explored from a graph-theoretic perspective, elucidating connections between the robustness bound and a function from number theory. We then discuss trade-offs natural systems face in simultaneous optimisation of robustness and information content and robustness and population neutrality. Upon coarse-graining phenotypes, it has been observed that robustness decreases relative to what one might expect from the underlying finer-grained phenotypes. We explain this non-intuitive behaviour quantitatively. We then extend the notion of robustness to input-output maps with continuous inputs, predicting a novel natural scaling law for robustness in these systems. Recent numerical results from deep neural networks support our theory.

The thesis concludes with a return to spin glasses, now in the context of evolutionary adaptation. We study short-term evolution on glassy fitness landscapes following non-adiabatic perturbation of metastability and characterise how variances in epistatic and external interactions influence timescales and change in fitness.

Originality, Publications, and Presentations

This work has not been submitted to any other university or to any other program at the University of Oxford and is wholly original unless otherwise stated in the thesis.
Datasets used in this thesis:

- Numerical RNA/HP folding data, unless otherwise noted, from the Greenbury-Schaper-Ahnert-Louis (GSAL) dataset in ref. [1] were used with permission from collaborator Sam Greenbury.
- RNA abstract shapes results/figures were used with permission from collaborator Tasmin Sarkany.
- Deep neural network raw data was used with permission from collaborator Yoonsoo Nam.

Notes on specific chapters:

1. Chapter 1 uses the GSAL dataset mentioned above and is otherwise wholly original. Part of the introduction appears in the pre-print [2]:
   Mohanty, V and Louis, A.A. “Robustness and Stability of Spin Glass Ground States to Perturbed Interactions”. (2020). arxiv:2012.05437.
2. Chapter 2 is wholly original. The work appears as the pre-print [2] above.
3. Chapter 3 is wholly original and was presented as follows:
   Mohanty, V and Louis, A.A. “A Phase Transition Between Random (Fragile) and Correlated (Robust) Phases of Input-Output Maps”. (2021). Session E17: Stochastic Thermodynamics of Biological and Artificial Information Processing - I (Focus Session), *American Physical Society March Meeting 2021.* A manuscript is in preparation.
4. Chapter 4 is a collaboration between the author and Sam Greenbury, Shyam Narayanan, Tasmin Sarkany, Kamal Dingle, Sebastian Ahnert, Ard Louis. Theorem 4.2.1 was discovered by the author independently of Greenbury’s [3] PhD thesis. Greenbury and the author are now collaborating together (with others mentioned) to publish those and other results together. The GSAL dataset and Sarkany’s figures are used. All writing and theory/derivations presented here are the author’s own unless stated in the thesis. Parts of Chapter 3 and of Chapter 4 will be presented as follows:
   Mohanty, V and Louis, A.A. “Maximally robust neutral networks in the

correlated phase of input-output maps". (2022). Session A08: Network Theory and Application to Complex Systems I (Focus Session), *American Physical Society March Meeting 2022.* A manuscript is in preparation.

5. Chapter 5 is a collaboration between the author and Yoonsoo Nam. All writing and theory/derivaitions presented here are the author's own unless stated in the thesis, all raw numerical data is Nam's.
6. Chapter 6 is wholly original.
7. Chapter 7 is wholly original.
8. Appendix A is wholly original.
9. Appendix B is wholly original.
10. Appendix C is a proof required by Chapter 4 that is primarily due to collaborator Shyam Narayanan. It is a part of the broader collaboration between the author, Sam Greenbury, Tasmin Sarkany, Kamal Dingle, Sebastian Ahnert, and Ard Louis.
11. Appendix D is wholly original.

Contents

Appendices

1
Introduction and Background

Contents

1.1 Introduction to Genotype-Phenotype (GP) Maps

In biology, a *genotype* is a set of stored biological information. Often, it refers to a sequence such as that of DNA or RNA responsible for coding for a protein or some other regulatory function, but it can also be represented, say, by the weights in a gene regulatory network. A genotype can map onto a *phenotype*, which is a biologically observed output or behaviour, in what is known as a genotype-phenotype (GP) map. Examples include 4 letter RNA sequences and 20 letter protein sequences that can be mapped to their physical folded states, and gene-regulatory networks, which can, for example, be described by Boolean networks [4] where a set of weights represent the gene interaction strengths. The response of such systems to changes in the input sequences have been extensively studied computationally and analytically

[1, 5–19]. For GP maps, an important concept is the set of genotypes (sequences) that map to a particular phenotype, often called the neutral network. These present a number of commonalities across GP maps [1, 15, 20]: the neutral networks are typically highly connected so that they can be traversed by single mutational steps, leading to enhanced evolvability, which is the ability to discover new phenotypes [7, 21]. In some cases, the neutral network is split into smaller component networks which are disconnected due to biophysical constraints [8, 22]). The number of sequences/vertices in the neutral network for different phenotypes can vary over many orders of magnitude, and is typically strongly biased, with a small fraction of the phenotypes taking up the majority of genotypes. Such phenotype bias can strongly affect evolutionary outcomes [11, 14, 23].

It has been shown that many naturally occurring genotype-phenotype maps exhibit a common set of features [1, 15, 20]:

1. **Redundancy**, which refers to the notion that the GP map is many-to-few: multiple genotypes can map to one phenotype.

2. **Phenotype bias**, which indicates that some phenotypes have many more genotypes mapping to them than others.

3. **Neutral correlations** refer to the notion that sequences that are close in terms of Hamming distances, are correlated, that is they often map to similar phenotypes.

A key property of neutral networks for GP maps is the *robustness*, typically defined in the literature as the fraction ρ_p of single-character mutations in a genotype that produce the same phenotype p, averaged over the neutral network of all genotypes that produce that particular phenotype. That is to say, for a biological genotype-phenotype map $f(\mathbf{g})$ which takes in a genotype $\mathbf{g}$ of ℓ characters chosen from an alphabet $K = \{K_0, \ldots, K_{k-1}\}$, the robustness of phenotype p is defined as

$$\rho_p(f) = \frac{1}{|\mathcal{G}_p|\ell(k-1)} \sum_{\mathbf{g}\in\mathcal{G}_p} n_{p,\mathbf{g}} \tag{1.1}$$

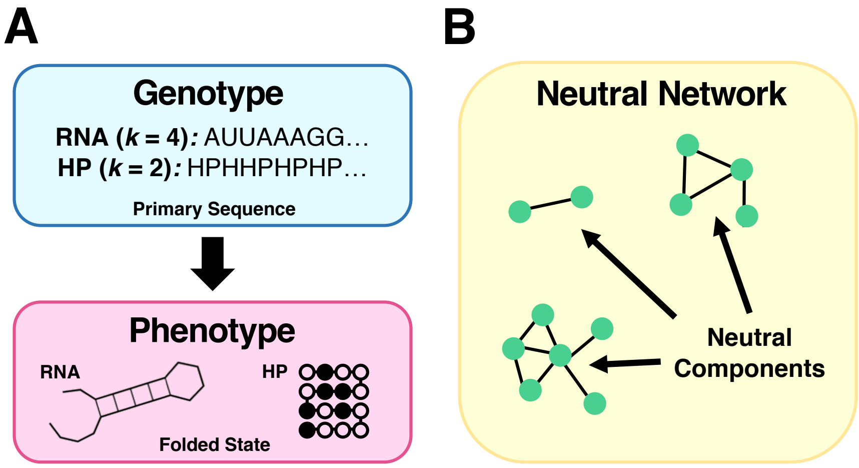


Figure 1.1: Schematic representation of genotype-phenotype maps and neutral networks. **(A)** In RNA secondary structure maps, there are $k = 4$ nucleotides (*A*denine, *C*ytosine, *G*uanine, and *U*racil) to choose from when building a genotype, and in HP protein folding GP maps, there are $k = 2$ classes of amino acids (*H*ydrophobic and *P*olar). Each of these sequences maps onto a 2-dimensional secondary structure phenotype. **(B)** Each structure (more generally, phenotype or output) has multiple inputs which map to it. The neutral network or neutral set corresponds to the set of genotypes which map to a particular phenotype.

where $\mathcal{G}_p$ is the neutral set of all genotypes $\mathbf{g}$ whose output is the phenotype p, and $n_{p,\mathbf{g}}$ is the number of genotypes $\mathbf{g}'$ satisfying $f(\mathbf{g}') = f(\mathbf{g}) = p$ that differ from $\mathbf{g}$ by a Hamming distance of 1. We call $n_{p,\mathbf{g}}$ the number of nearest-neighbours of $\mathbf{g}$ mapping to p. Thus $\rho_p(f) \in [0,1]$ measures the mean probability that a mutation from $\mathbf{g} \in \mathcal{G}_p$ to a neighbouring genotype $\mathbf{g}' \in \mathcal{G}_p$ results in the same phenotype p.

GP maps outside of biology may be called input-output (IO) maps (or may still be called genotype-phenotype maps). Throughout this thesis, we will be using both terms interchangeably. Many natural input-output maps both within and outside of biology are well-studied empirically: e.g. RNA secondary structure folding [1, 8], lattice models of protein folding [1], protein self-assembly [1], genetic programming [19], and gene regulatory networks [18].

In the RNA secondary structure GP map, the input or genotype is a nucleic acid sequence drawn from an alphabet of $k = 4$ characters corresponding to *A*denine, *C*ytosine, *G*uanine, and *U*racil. Programs like ViennaRNA [24] use free energy minimisation to determine a secondary structure which the oligonucleotide would

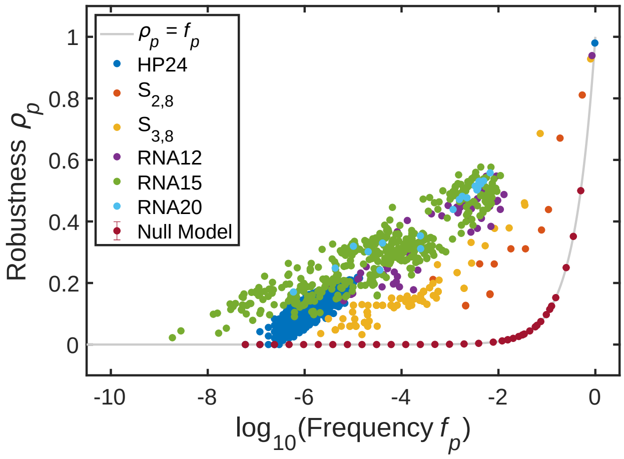


Figure 1.2: Reproduction of Figure 2(a) from ref. [1]. Phenotype robustness ρ_p plotted against the log of the phenotype frequency f_p for many natural genotype-phenotype maps.

adopt; this is the output or phenotype of the GP map. The HP model of protein folding [25] is also a sequence-to-structure GP map. A protein primary sequence (the input or genotype) is modelled as a chain of either **H**ydrophobic or **P**olar amino acids (so $k = 2$ characters in the alphabet: H and P), and the output is a 2-dimensional non-self-intersecting random walk on a lattice corresponding to a predicted secondary structure that that the peptide would adopt. Schematic representations of these GP maps are shown in Figure 1.1A. The length of the RNA or HP input sequence is denoted by a number following the model name. So, "RNA12" is the RNA secondary structure GP map for oligonucleotides of length 12, "HP24" is the HP protein folding GP map for primary sequences of length 24. The Polyomino GP map is a protein quaternary structure GP map [12, 26] in which domino-like square tiles can be rotated and matched up on a 2-dimensional lattice with other tiles based on the types of interfaces expressed on the sides of the squares; these assembled structures are called polyominoes. A polyomino model with N_t tiles and N_c interfaces is denoted by S_{N_t,N_C}.

What all of the input-output maps mentioned above—including both the biological and non-biological models—hold in common is that they each have high levels of mean robustness to changes in their inputs, excepting pathological or adversarial examples (such as the 1-dimensional Edwards-Anderson model in spin glass ground state GP maps discussed in the following chapter). Naively, one might guess that in a large, uncorrelated, and randomly-assigned GP map, the probability that a nearest-neighbouring genotype yields phenotype p is approximately equal to the probabilitity that any genotype drawn at random from the entire input space yields phenotype p. This probability is $|\mathcal{G}_p|/k^\ell$, so $n_{p,\mathbf{g}} \approx \ell(k-1)|\mathcal{G}_p|/k^\ell$, and $\rho_p \approx |\mathcal{G}_p|/k^\ell$. And indeed this is true for completely uncorrelated and randomly-assigned GP maps [1]. However, biologically inspired GP maps such as the RNA secondary structure maps, protein folding maps, gene regulatory networks, genetic programming, and others all exhibit a scaling law for robustness $\rho_p \sim \mathcal{O}(\log|\mathcal{G}_p|)$ and not $\rho_p \sim \mathcal{O}(|\mathcal{G}_p|)$. We reproduce Figure 2(a) from ref. [1] in our Figure 1.2, which shows the observed robustenss values of all of the phenotypes for the RNA12 and RNA15 secondary structure genotype-phenotype maps, the HP24 protein folding map, the $S_{2,8}$ and $S_{3,8}$ Polyomino protein self-assembly maps, and a subset of sampled phenotypes for the RNA20 secondary structure map. Also plotted are the expectation of robustness for the random null model in which phenotypes are randomly assigned to genotypes. We see clearly that the natural genotype-phenotype maps show elevated robustness relative to the null model, indicating the presence of neutral correlations in the genotype-phenotype map: in general, the natural maps tend to follow the $\rho_p \sim \mathcal{O}(\log|\mathcal{G}_p|)$ scaling law while the null model follows the $\rho_p \sim \mathcal{O}(|\mathcal{G}_p|)$, uncorrelated scaling law. Such enhanced robustness in the natural maps is believed to be critical for the evolutionary process because it allows neutral (i.e. fitness cost-free) exploration of the neutral network, which enhances the ability of a population to find new phenotypic variation [6, 7, 10, 15, 21, 27].

1.2 Biological and Graph-Theoretic Connections

In this section, we will provide mathematical definitions of key objects and parameters which appear throughout this thesis in the discussion of genotype-phenotype maps (and input-output maps more broadly) and robustness. We begin by explicitly making connections between biological terminology in the GP map literature and the corresponding graph theoretic definitions. For a graph G, we will refer to its set of vertices as $V(G)$ and its set of edges as $E(G)$. The cardinality of a set A will be denoted by $|A|$.

As mentioned before, we take a *genotype* to be a sequence of length ℓ drawn from an alphabet of k characters which are fixed for a particular system. The standard graph representation of the genotype space involves treating each of the k^ℓ genotypes as vertices in a graph, with two vertices being connected by an edge if and only if the corresponding sequences differ by exactly one character—i.e. the Hamming distance between the two sequences is 1. Each vertex then has degree $\ell(k-1)$, the graph is called a Hamming graph:

Definition 1.2.1. *A* ***Hamming graph*** $H_{\ell,k} \equiv (K_k)^{\square \ell}$ *is the Cartesian product of* ℓ *copies of the complete graph* K_k *(which is a graph of* k *vertices where all vertices are connected to all other vertices).*

The Cartesian product $A \square B$ of two graphs A and B is defined to have a vertex set $V(A \square B) = \{(a,b) \,:\, a \in A \wedge b \in B\}$, and an edge exists between vertices (a,b) and (a',b') for $a, a' \in A$ and $b, b' \in B$ if $a = a'$ and (b,b') is an edge or if (a,a') is an edge and $b = b'$. Therefore, each vertex, which corresponds to a sequence of length ℓ, is connected to every other vertex which has a sequence that differs from the first vertex at only one site. For $k = 2$, the Hamming graph is identical to the ℓ-dimensional hypercube graph Q_ℓ.

We assume that each genotype maps deterministically to a phenotype, and the set of all phenotypes is assumed to be finite. Mathematically, a GP map is therefore a vertex labelling of the Hamming graph, where each vertex label has been taken from a finite set of phenotypes of cardinality N_p. We now consider the set of genotypes which all map to the same phenotype p. The subgraph of $H_{\ell,k}$ *induced* by these vertices contain all of the edges present in $H_{\ell,k}$ which connect 2 genotypes which both map to p. Formally, this is known as a neutral network:

Definition 1.2.2. *The* ***neutral network*** $G^{(p)}$ *belonging to phenotype* p *is the induced subgraph of the Hamming graph* $H_{\ell,k}$ *which contains all of the vertices*

Table 1.1: Evolutionary biology definitions and graph theoretic translations. All genotypes are considered to be sequences of length ℓ drawn from k characters. We refer to a generic phenotype p and the graph $G^{(p)}$ consisting of all vertices/sequences which map to that phenotype.

Evolutionary Biology	**Graph theory**
Set of all genotypes	Hamming graph $H_{\ell,k}$
Neutral network/set $G^{(p)}$	Induced subgraph of $H_{\ell,k}$
Neutral component of a neutral network	Connected component of a graph
Robustness $\rho(G^{(p)})$	*Density $\dfrac{2}{\ell(k-1)}\dfrac{\lvert E(G^{(p)})\rvert}{\lvert V(G^{(p)})\rvert}$

*Up to prefactor.

(genotypes) which map to phenotype p. The induced subgraph contains all of the edges in $H_{\ell,k}$ which connect vertices in $G^{(p)}$. Mathematically, $E(G^{(p)}) = \{\{u, v\} \in H_{\ell,k} \,|\, u \in G^{(p)} \wedge v \in G^{(p)}\}$, where $\{\cdot,\cdot\}$ is an an unordered pair.

In the neutral theory of evolution [28–30], neutral networks which are highly connected play an essential role. The fact that all genotypes within a neutral network map to the same phenotype allow for traversal of a potentially large portion of the Hamming graph without incurring any fitness penalty. Larger neutral networks also tend to share edges with a larger number of other distinct neutral networks, facilitating the discovery of new phenotypes.

A graph can have multiple connected components whose union forms the entire graph, but which individually are not connected to each other; in evolution, a connected component of a particular neutral network is called a *neutral component* (see Figure 1.1B). The number of neutral components as well as the number of sequences comprising the largest neutral component for phenotype neutral networks in RNA, HP, and Polyomino models are plotted against the frequency of those phenotypes in Figure 1.3. We see that the biological GP maps maintain distinct trends from the random null model. The correlated input-output map structure ensures that even for small frequencies, neutral components tend to be large and clustered as opposed to small and broken apart as in the random null model. Neutral networks which have many small neutral components do not really offer the phenotypic discovery advantage described previously.

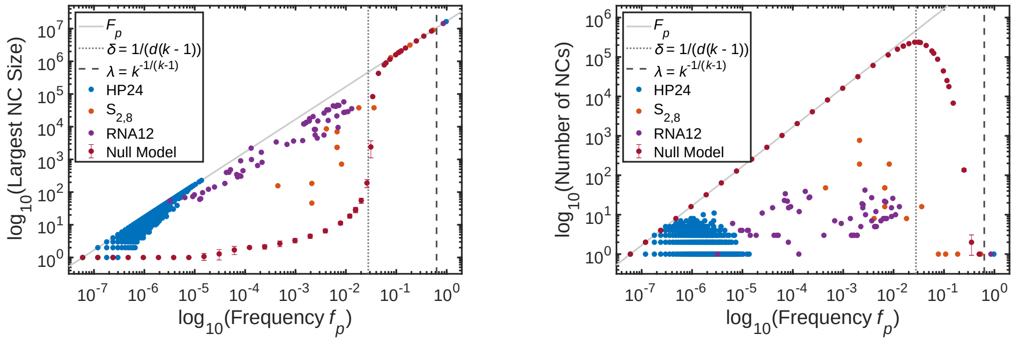


Figure 1.3: Reproduction of Figure 3 from ref. [1]. (Left) Log-log plot of the number of sequences/vertices comprising the largest neutral component for each phenotype in RNA12, HP24, and $S_{2,8}$ polyomino GP maps as well as the expectation from the random null model (for which $k = 4$ and $d = 12$ define the number of characters in the alphabet and the sequence length, respectively) versus the frequency of the phenotype. Notice that for the biological GP maps, the largest neutral component increases with the frequency of the phenotype while for the random null model, the largest neutral component stays fairly small across all frequencies below a threshold frequency $\delta = 1/(d(k-1))$. Above this percolation threshold, a giant component emerges, and this becomes a single component when the frequency of the phenotype is greater than $\lambda = k^{-1/(k-1)}$. (Right) Log-log plot of the number of neutral components for the biological and random null models versus phenotype frequency. The biological models maintain a small number of neutral components regardless of frequency while the null model has neutral networks which are broken into many smaller components which eventually percolate into a giant component at the same threshold δ and a single component at λ. Also shown in both of these plots is the line F_p which corresponds to the largest component or the number of components equaling $F_p = k^d f_p$, the number of sequences which map to output p.

We now define robustness, a quantity which is central to this thesis:

Definition 1.2.3. *The* ***robustness*** $\rho_p \equiv \rho(G^{(p)})$ *of the* p*-th phenotype's neutral network* $G^{(p)}$*, which is an induced subgraph of* $H_{\ell,k}$*, is proportional to the density of the graph:*

$$\rho_p = \frac{2}{\ell(k-1)} \frac{\left|E(G^{(p)})\right|}{|V(G^{(p)})|} \tag{1.2}$$

Due to the identity

$$2\left|E(G^{(p)})\right| = \sum_{v \in V(G^{(p)}} \deg(v), \tag{1.3}$$

we can see that definition 1.2.3 is identical to eq. (1.1). As mentioned before, natural systems tend to exhibit a robustness scaling law of $\rho_p \sim \mathcal{O}(\log|V(G_p)|)$ and not $\rho_p \sim \mathcal{O}(|V(G_p)|)$, which would be expected from a random mapping from input to output [1]. The biological and graph theoretic definitions of the above important parameters are summarised in Table 1.1.

A property that is closely related to the robustness is the transition probability between two phenotypes, defined as:

Definition 1.2.4. *The* ***transition probability*** *that a single point mutation in the genotype leads to a change from phenotype p to phenotype q is given by*

$$\begin{aligned} \phi_{qp} &= \frac{|E(G_p, G_q)|}{\ell(k-1)|V(G_p)|}, \quad p \neq q \\ &= \frac{1}{|V(G_p)|\ell(k-1)} \sum_{\mathbf{g} \in V(G_p)} n_{q,\mathbf{g}}, \end{aligned} \tag{1.4}$$

where G_p and G_q are the neutral networks for two different phenotypes, $E(G_p, G_q)$ is the set of edges present in the Hamming graph which each have one vertex belonging to G_p the other belonging to G_q (i.e. the edges connecting neutral networks G_p and G_q), and $n_{q,\mathbf{g}}$ is the number of neighbours the genotype $\mathbf{g}$ has which map to output q (note that $\mathbf{g}$ itself maps to output p).

Note that $\phi_{qp}|V(G_p)| = \phi_{pq}|V(G_q)|$. We define the diagonal terms $\phi_{pp} \equiv \rho_p$ so that there is an additional prefactor of 2 because $|E(G_p, G_p)| = |E(G_p)|$ only counts the edges within G_p once, but we need the edges counted twice. In both natural systems and random systems [1], the observed scaling law is $\phi_{qp} \sim \mathcal{O}(|V(G_q)|)$; i.e. it appears that the neighbours of output p which are not themselves mapping to p are chosen at random based on their frequency in the map. The linear scaling law is apparent in RNA and HP secondary structure GP maps, as shown in Figure 1.4.

In this thesis, we emphasise the network-theoretic nature of input-output and genotype-phenotype maps. In many cases, thinking about the mapping from input to output graphically and using the properties of Hamming graphs help derive novel biological intuition. The above graph theoretic parameters and definitions will be used throughout the thesis, subject to notational modifications as needed.

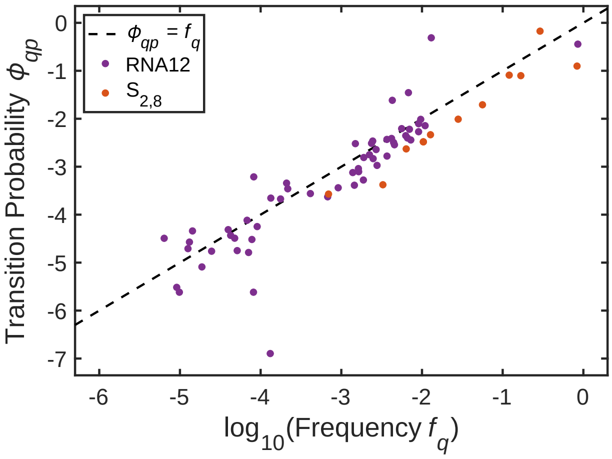


Figure 1.4: Transition probabilities ϕ_{qp} from phenotype p to phenotype q upon a single character mutation of an input sequence, plotted against the frequency of output q. Here, we plot ϕ_{qp} for the RNA12 and HP24 models for the phenotype with the second largest number of genotypes mapping to it. Note that $\phi_{qp} \approx f_q$. These plots include the self-transition $\phi_{pp} = \rho_p$. Data obtained from the authors of ref. [1] (the Greenbury-Schaper-Ahnert-Louis, or GSAL, dataset).

1.3 Thesis Outline

The remainder of the thesis is outlined as follows:

In Chapter 2, we examine spin glass ground states as input-output maps and show that the network theoretic properties observed in natural biological GP maps are also reproduced in the spin glass maps, suggesting a deeper, universal origin. This motivates Chapter 3, in which we propose a maximum entropy model of genotype-phenotype maps and analyze its behaviour. We show that a phase transition emerges between robust and fragile phases of genotype-phenotype map organisation and suggests that natural GP maps belong to one of two important classes.

Chapter 4 then explores maximally robust neutral networks from a graph theoretic perspective. We explore 3 parameters: robustness, population neutrality, and base information content of maximally robust neutral networks known as bricklayer's graphs. We also elucidate for the first time the connection between a number-theoretic fractal function and biological GP maps. Also in Chapter 4 we propose a novel theory of phenotype coarse-graining and discuss why neutral networks themselves tend to stray away from the optimal robustness curve while neutral components in many natural maps actually meet it.

In Chapter 5, we propose for the first time an extension of the GP map notion of robustness to continuous systems, showing how robustness is related to the surface area-to-volume ratio of neutral sets in these input-output maps. We then provide a series of approximations to the shapes of the neutral set (with zero, then one, then two tunable parameters) which elucidate a theoretical power law for the behaviour of robustness with respect to neutral set size. We fit the model to numerical data for deep neural networks and show that it is consistent with the power law form, and that the fitted parameters take on reasonable values.

In Chapter 6, we investigate a new direction of research, inspired by the question of how a population can find a new fitness peak when the landscape changes. In particular, our numerical experiments suggest that the discovery of new metastable states can be thought of as a Poisson process, and that the observed power-law relationship between change in fitness and metastable state discovery time is strengthened or weakened according to the variance in the epistatic interactions in the fitness model.

Chapter 7 recapitulates the findings of the thesis.

2

Robustness of Spin Glass Ground States to Perturbed Interactions

2.1 Introduction

Given the generality of the high robustness observed in biologically inspired GP maps, in this chapter we ask the question whether a similar phenomenon can be found in spin glasses, which have a rich history in statistical and condensed matter physics. They have been intensely studied since the 1970s [31, 32] (and most recently have been recognized with a 2021 Nobel Prize in Physics to Giorgio Parisi—a spin glass pioneer). Spin glasses have led to many important insights in physics and other related disciplines, including computer science [33–35]. More recently, the spin glass Hamiltonian has been used as a phenomenological model for epistatic genotype to fitness landscapes in which different sites (*e.g.* DNA, genes, or amino acids) may couple to each other [36–42]. An important application has been to viruses [37–40]. By taking sequence data over time, inverse statistical physics methods can be employed to "learn" the interactions (or couplings) between different sites. In the context of evolution, therefore, one can interpret the ground state of the spin glass energy landscape as the global fitness peak on an evolutionary fitness landscape. The interactions between spins in such a system can depend on a number of biological or environmental factors [36, 40]. In the context of robustness, the interesting question here for these genotype (sequence) to fitness landscapes is again, what is the robustness of the ground state configuration (the sequence of spins which gives the lowest energy) to mutations in the interactions (the couplings)? In other words, if we perturb an interaction, how likely is the same set of spins still the ground state?

Here, for tractability, we treat a simpler system than those typically used for fitness landscape inference, namely the $\pm J$ spin glasses on random graphs. In particular, we examine special cases, namely the Sherrington-Kirkpatrick model and 1D Edwards-Anderson model. Various network-topological properties are computed for subgraphs of the input space which include the sets of interactions all mapping to the same ground state configuration. We find that such subgraphs obey the same logarithmic scaling law between robustness and phenotype network size as in the analogous biological GP maps described above, suggesting that this high robustness may hold for a much wider set of physical systems.

2.2 Model and Methods

2.2.1 Spin Glass Model

Consider an undirected, unweighted random graph $G(V, E)$ with V vertices and E edges, such as the one in Figure 2.1(a). We place Ising spins on each vertex, and each edge represents a nonzero interaction between spins. A spin configuration $s \in \{\pm 1\}^{|V|}$ can be written as a sequence of $+1$ and -1 values, so it is essentially a binary sequence of length $|V|$. A set of interactions $J \in \{\pm 1\}^{|E|}$ similarly is a sequence of $+1$ and -1 values of length $|E|$. The spin glass Hamiltonian

$$\mathcal{H}_G(s; J) = -\sum_{\{i,j\}\in E} J_{ij} s_i s_j - \sum_{i \in V} h_i s_i \tag{2.1}$$

contains couplings between all spins which are connected by an edge in G. The single-spin, external magnetic field interactions h_i are chosen from a random distribution while maintaining $|h_i| \ll |J_{ij}|$ in order to break the possible degeneracies of the spin glass ground state, yielding only a unique ground state. Details are explained further in the Numerical Methods section.

The input-output map considered in our study is the spin glass ground state optimisation function

$$\Omega_G : \{\pm 1\}^{|E|} \to \{\pm 1\}^{|V|} \tag{2.2}$$

defined for the graph G. For a set of interactions J, $\Omega_G(J)$ outputs the ground state configuration s that minimizes the Hamiltonian eq. (2.1). This ground state is not necessarily unique for spin glasses in general, but we will choose values for the various J_{ij} and h_i parameters in our simulation such that there is in fact a unique ground state. The most common task in spin glass theory is to find s given a particular set of interactions J. In this chapter, we study an inverse problem, namely the relationship between the set of all input sets $\{J\}$ that generate a particular output s.

To efficiently represent this system, we note that any binary sequence of length n can be represented by a n-dimensional undirected hypercube graph $Q_n(U, F)$, with vertices U such that $|U| = 2^n$ and edges F such that $|F| = 2^{n-1}n$. This is accomplished by mapping each binary sequence to a vertex in $Q_n(U, F)$ and placing edges between two vertices if the corresponding sequences have a Hamming distance of 1 between them.

The domain of Ω_G accordingly has a mapping to the $|E|$-dimensional hypercube graph $Q_{|E|}(U, F)$. In general, for graphs G which have sufficiently high connection density to produce geometrical frustration in the spin glass, $\Omega_G(J)$ follows no pattern and is difficult to calculate [43], even more so because of the degeneracy-breaking external random field interactions $\{h_i\}$. But, because of frustration, two sets of interactions $J^{(i)}$ and $J^{(j)}$ mapped to connected vertices often have $\Omega_G(J^{(i)}) = \Omega_G(J^{(j)})$. The vertices corresponding to all J such that $\Omega_G(J) = s$ for some fixed s induce a subgraph $H_s(U_s, F_s)$ of $Q_{|E|}$. It follows that $\bigcup_s U_s = U$. These subgraphs are the equivalent of neutral sets in the GP map literature.

In this chapter, we numerically compute network-topological properties of the induced subgraphs $H_s(U_s, F_s)$ of the hypercube $Q_{|E|}(U, F)$ for a spin glass on a random graph $G(V, E)$. We consider three cases for G: a sparse random graph with $|E| \lesssim \frac{1}{2}\binom{|V|}{2}$, a dense random graph with $|E| \gtrsim \frac{1}{2}\binom{|V|}{2}$, and a complete graph

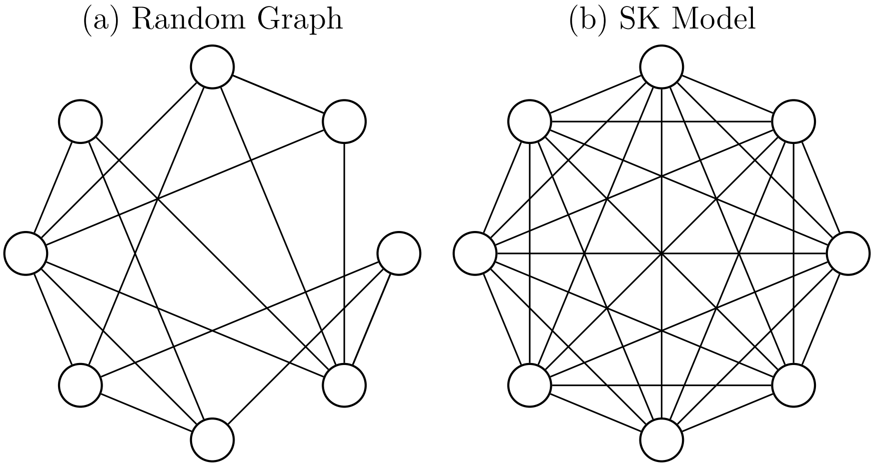


Figure 2.1: $G(V, E)$ for: (a) a dense random graph and (b) the fully connected Sherrington-Kirkpatrick model. Spins are placed on vertices and nonzero interactions are on the edges.

$|E| = \binom{|V|}{2}$. The latter case is known as the Sherrington-Kirkpatrick (SK) model of a spin glass [31]. We also consider the 1-dimensional Edwards-Anderson model [32], for which the relationship between induced subgraph edge count $|F_s|$ and vertex count $|U_s|$ (equivalent to robustness) becomes analytically solvable with knowledge of the degree distribution, which is presented in the Results section.

2.2.2 Definitions of Topological Quantities

The following parameters are computed for the spin glass input-output maps in our study:

Robustness, normalised edge count, or mean degree. This property is exactly the mean robustness defined in eq. (1.1) and Definition 1.2.3. Locally, the neighbour count $n_{p,\mathbf{g}}$ is equivalent to the degree of a particular vertex $v \in H_s(U_s, F_s)$, where H_s is the induced subgraph of $Q_{|E|}$ such that all vertices in H_s correspond to interactions which result in ground state spin configuration s.

The mean degree is related to the number of edges by

$$\sum_{v \in U_s} \deg(v) = 2|F_s|. \tag{2.3}$$

Here, we compute a normalised mean degree or normalised edge count (or equivalently the robustness), simply dividing the above quantity by the size of the subgraph $|U_s|$ and by the length of the input sequence $|E|$:

$$\rho_s \equiv \phi_{ss} \equiv \frac{2|F_s|}{|U_s||E|} \in [0,1]. \tag{2.4}$$

The notation ϕ_{ss} will become clear below when we also treat the transition probability ϕ_{rs} of a vertex leading to a different ground state r. For the remainder of the chapter, we use "normalised edge count" and "robustness" interchangeably. In many "real-world" input-output maps including RNA and protein folding, Boolean threshold networks, and genetic algorithms, it has been observed that $\rho_s \sim \log |U_s|$ or equivalently $|F_s| \sim |U_s| \log |U_s|$. We will test this scaling for the spin-glass system.

Transition probability. Consider two induced subgraphs of $Q_{|E|}(U,F)$ called $H_s(U_s,F_s)$ and $H_r(U_r,F_r)$, with $s \neq r$ so that s and r are two spin different configurations. Let $T_{rs} = T_{sr}$ be the set of edges connecting the induced subgraphs H_s and H_r. Mathematically, $T_{rs} = \{\{u,v\} \in Q_{|E|}(U,F) \,|\, u \in U_s \wedge v \in U_r \wedge u \neq v\}$. We define the *transition probability* of $s \to r$

$$\phi_{rs} \equiv \frac{|T_{rs}|}{|U_s||E|} \in [0,1], \quad s \neq r \tag{2.5}$$

as the probability that a random bit flip in an input sequence corresponding to a vertex in U_s will result in a sequence that is found in U_r. That is to say, by performing a sign flip of one spin coupling in the set of interactions $J \mapsto J'$, ϕ_{rs} gives us the conditional probability $\mathbb{P}(\Omega_G(J') = r \,|\, \Omega_G(J) = s)$. In many "real-world" input-output maps, it has been observed that often $\phi_{rs} \sim |U_r|$, or equivalently $|T_{rs}| \sim |U_r||U_s|$ [1, 11]. We will test this scaling for the spin-glass system.

It is now easy to see that, for the case where $s = r$, the transition probability T_{ss} simply counts the number of edges in the induced subgraph $H_s(U_s,F_s)$ and therefore is equivalent to F_s. However, in the definition of ρ_s in eq. (2.4), we have a necessary factor of 2 which comes from the double counting of edges. This double counting is required for consistency with the notion of a transition probability due to a bit flip—or in this case, staying within the same subgraph despite a bit flip. From the definitions, the normalisation condition $\sum_r \phi_{rs} = 1$ holds, where the sum includes $r = s$.

Rank-size and degree distributions. A rank-size plot, plotted on a log-log scale, may be used to deduce if there is a power law (i.e. generalized Zipf's Law) relationship between rank and number of vertices U_s in the subgraph as is seen for some GP maps. We also compute the distribution of degrees of all the vertices in a given subgraph, for all subgraphs. This provides more information about the modality and skew of the distribution than simply counting the edges or finding the mean of the degree.

Clustering coefficients. The local clustering coefficient defined at a vertex w gives the ratio of all neighbours of w connected to each other to the ratio of all possible pairs of the neighbours of w, the latter of which is $\binom{\deg(w)}{2}$. As such, $C(w)$ calculates the fraction of triangles involving w out of all possible triangles involving w. Given an induced subgraph $H_s(U_s, F_s)$, the local clustering coefficient $C_s(w)$ for a vertex $w \in U_s$ is defined as

$$C_s(w) = \frac{2|\{\{u, v\} \in F_s \mid u, v \in N_s(w) \wedge u \neq v\}|}{\deg(w)(\deg(w) - 1)}, \tag{2.6}$$

where $N_s(w) = \{v \in U_s \mid \{v, w\} \in F_s\}$ is the neighbourhood of w, i.e. the set of all vertices connected to w. An averaged clustering coefficient

$$\overline{C}_s = \frac{1}{|U_s|} \sum_{w \in U_s} C_s(w) \tag{2.7}$$

can also be defined for the entire induced subgraph.

In our system, we note that the hypercube graph $Q_{|E|}(U, F)$ has no triangles to begin with and thus has $C(w) = 0$ for all $w \in U$. It immediately follows that all local clustering coefficients are zero for all induced subgraphs.

Assortativity. A network's *assortativity* is a measure of correlation between the degrees of two connected vertices. Typically, the Pearson correlation coefficient r is used as a quantitative measure of assortativity. For a subgraph $H_s(U_s, F_s)$, this is calculated by finding [8]

$$r_s = \frac{\sum_{v\in U_s} \deg(v)^2\, \overline{\mathrm{nndeg}}(v) - \frac{1}{2|F_s|}\left(\sum_{v\in U_s} \deg(v)^2\right)^2}{\sum_{v\in U_s} \deg(v)^3 - \frac{1}{2|F_s|}\left(\sum_{v\in U_s} \deg(v)^2\right)^2}, \tag{2.8}$$

where $\overline{\mathrm{nndeg}}(v) = \frac{1}{|N_s(v)|}\sum_{u\in N_s(v)} \deg(u)$ is the average degree of vertices in the neighbourhood $N_s(v)$ of vertex $v \in U_s$. Networks with $r > 0$ are said to be assortative, and vertices with higher degree tend to be connected to vertices with higher degree. Accordingly, networks with $r < 0$ are said to be dissortative, and vertices with relatively high degree tend to be connected to vertices with relatively low degree.

Betweenness centrality. The *betweenness centrality* $B_s(v)$ of a vertex $v \in U_s$ for a $H_s(U_s, F_s)$ is defined as

$$B_s(v) = \frac{1}{2}\sum_{u,w\in U_s} \frac{g(u,v,w)}{g(u,w)}, \quad u \neq v \neq w, \tag{2.9}$$

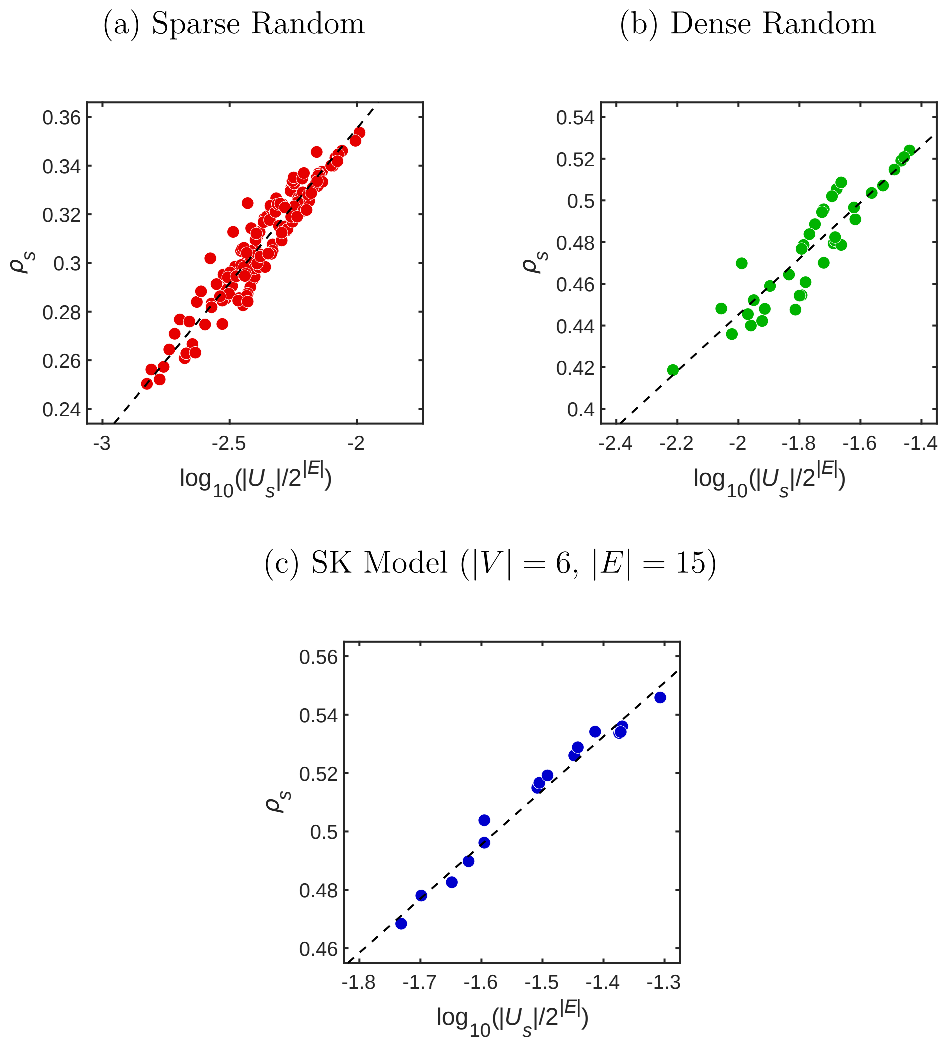
where $g(u,w)$ is the number of shortest paths between u and w and $g(u,v,w)$ is the number of shortest paths between u and w that pass through v. $B_s(v)$ is often plotted against $\deg(v)$.

2.2.3 Numerical Methods

For each instance of a random graph (and for the special cases of the SK model and the 1D Edwards-Anderson model), we fixed random fields h_i drawn from the distribution

$$\mathbb{P}(h_i) = \frac{1}{2}\left[\delta(h_i - \epsilon) + \delta(h_i + \epsilon)\right], \quad \epsilon \ll 1 \tag{2.10}$$

in the spin glass Hamiltonian given in eq. (2.1). Such a distribution enforces $|h_i| = \epsilon \ll 1 = |J_{ij}|$ and drastically raises the probability that ground state degeneracies—due to $\mathbb{Z}_2$ symmetry (i.e. $s_i \mapsto -s_i$ for all i), underlying symmetries of


(a) Sparse Random

(b) Dense Random

(c) SK Model ($|V| = 6$, $|E| = 15$)

Figure 2.2: Plot of normalised edge count ρ_s (equivalent to robustness) versus log of normalised number of subgraph vertices, $\log_{10}(|U_s|/2^{|E|})$ (equivalent to number of vertices in induced subgraph) for a sparse random graph ($|V| = 9$, $|E| = 15$), dense random graph ($|V| = 7$, $|E| = 15$), and SK model ($|V| = 6$, $|E| = 15$). The correlation between ρ_s and $\log_{10}(|U_s|/2^{|E|})$ is **(a)** sparse random graph: Pearson $r = 0.9814$, Spearman $\rho = 0.9817$, **(b)** dense random graph: Pearson $r = 0.9679$, Spearman $\rho = 0.9750$, and **(c)** SK model: Pearson $r = 0.9904$, Spearman $\rho = 0.9915$. Dashed line is the line of best of fit in the linear-log scale.

the random graph $G(V, E)$, and particular configurations of J_{ij}—are broken, leaving only a unique ground state spin configuration s which minimizes the Hamiltonian in eq. (2.1). The ground state for each possible configuration of J_{ij} was computed by exhaustively enumerating over all possible spin configurations $s \in \{-1, +1\}^{|V|}$; we chose $\epsilon = 10^{-4}$ and numerically verified that it was indeed unique for simulation

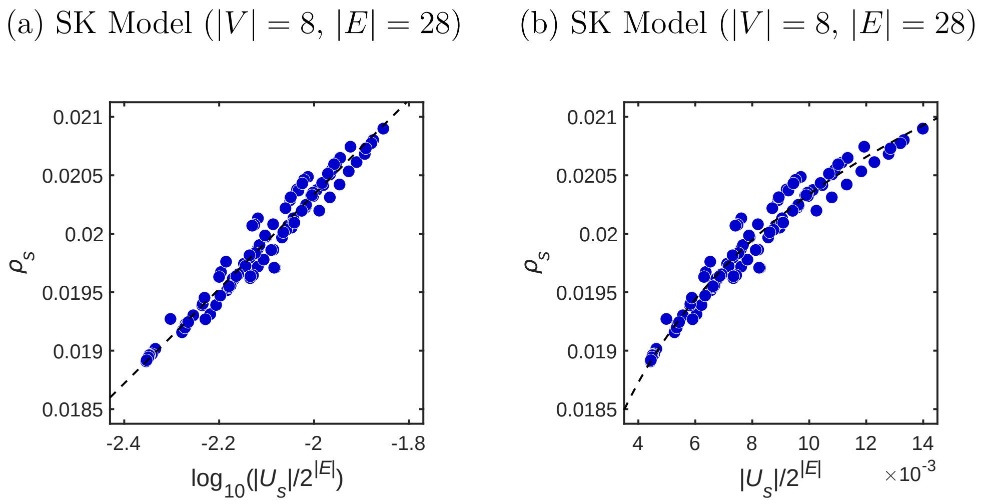
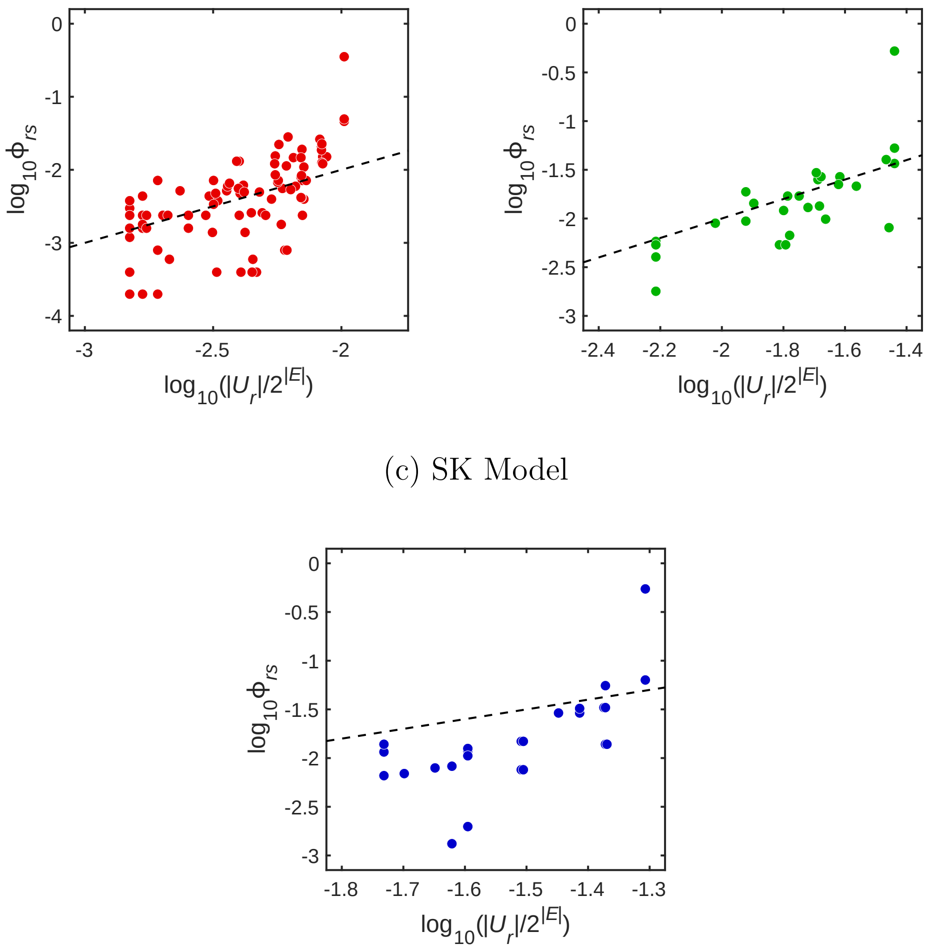

(a) SK Model ($|V| = 8$, $|E| = 28$) (b) SK Model ($|V| = 8$, $|E| = 28$)

Figure 2.3: **(a)** Plot of normalised edge count ρ_s (equivalent to robustness) versus log of normalised number of subgraph vertices $\log_{10}(|U_s|/2^{|E|})$ and **(b)** plot of normalised edge count ρ_s (equivalent to robustness) versus normalised number of subgraph vertices $|U_s|/2^{|E|}$ for SK model ($|V| = 8$, $|E| = 28$). The data shown in both plots are identical; only the scaling of the abscissa is modified to demonstrate clearly that the scaling law for the normalised edge count/robustness is indeed logarithmic in the subgraph vertex count. The dashed line in the left panel is the ordinary least squares best fit line; the same line is log-transformed in the abscissa coordinate to the dashed logarithmic curve in the right panel.

cases we considered. This study was repeated for many instances of sparse and dense random graphs $G(V, E)$ (and various choices of random h_i); numerical results for single representative samples are shown in the following section.

2.3 Results

We now present results for the various topological quantities described in the previous section for the subgraphs $H_s(U_s, F_s)$ (the equivalent of neutral sets) for spin glasses defined on random graphs $G(V, E)$ including: (a) sparse random graphs, (b) dense random graphs, and (c) the Sherrington-Kirkpatrick model. We also present the 1D Edwards-Anderson model as a special case where we can calculate the exact relationship between edge count and induced subgraph vertex count. The SK model simulations typically involve lower $|V|$ due to computational constraints.



(a) Sparse Random

(b) Dense Random

(c) SK Model

Figure 2.4: Log-log plot of *nonzero* transition probability ϕ_{rs} as a function of normalised induced subgraph vertex count $|U_r|/2^{|E|}$ for sparse random graph ($|V| = 9$, $|E| = 15$), dense random graph ($|V| = 7$, $|E| = 15$), and SK model ($|V| = 6$, $|E| = 15$). Correlation coefficients for plot points which do not have $\phi_{rs} = 0$ are **(a)** sparse random graph: Spearman $\rho = 0.6612$, **(b)** dense random graph: Spearman $\rho = 0.8034$, and **(c)** SK model: Spearman $\rho = 0.7049$. Dashed line is given by eq. (2.11).

2.3.1 Subgraph Edge Count or Mean Degree

For all topologies for G, our numerical simulations showed that each induced subgraph $H_s(U_s, F_s)$ has exactly one connected component, regardless of size. In other words, each network of inputs (set of interactions) which map to a particular output (a particular ground state) has only a single component.

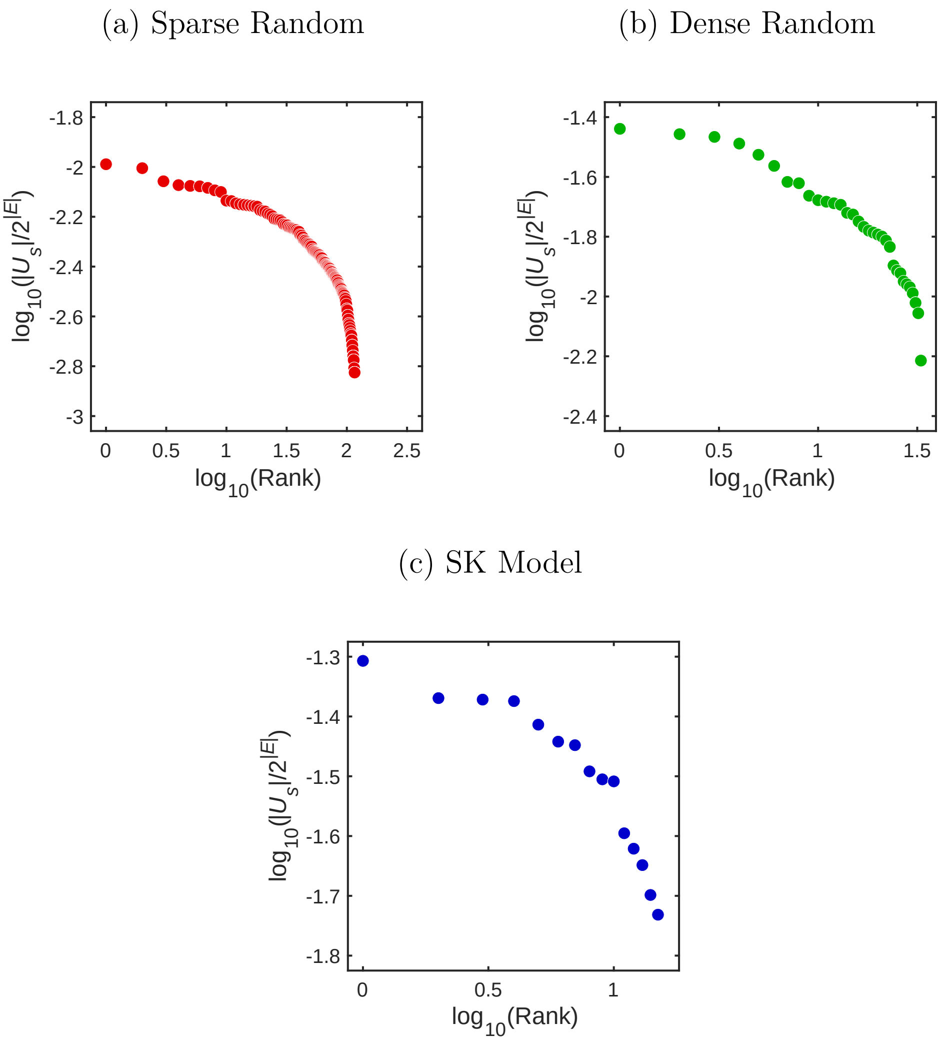
In Figure 2.2, we plot the normalised edge count of induced subgraphs versus the logarithm of the induced subgraph vertex count. Spin glasses on sparse, dense, and the complete graphs all approximately display the $\rho_s \sim \log |U_s|$ relationship which is also seen for the closely related scaling of robustness with neutral set size found for many GP maps. These can be compared with Figure 1.2 to see indeed that the robustness scaling behaviour exhibited in the biological genotype-phenotype maps is also demonstrated here in spin glasses. To further emphasize the logarithmic nature of this scaling, we plot both linear-linear and linear-log scale plots for robustness of a larger SK Model spin glass ($|V| = 8$, $|E| = 28$, so there are 2^{28} unique inputs and $2^8 = 64$ unique outputs) in Figure 2.3. It is clear from especially the linear-axis plot on the right that the scaling is indeed logarithmic. This result indicates that there is a high degree of robustness (or a lack of sensitivity) in the spin glass input-output map to perturbations of the couplings.

The vertices of induced subgraphs of $Q_{|E|}(U, F)$ tend to be located near each other in the hypercube. To compare our result to a random null model, take some spin glass ground state mapping $\Omega_G(J)$, and randomize the input-output pairings while keeping the subgraph vertex counts the same. A perturbation of a single interaction $J \mapsto J'$ will result in a spin configuration $s = \Omega_G(J')$ being selected with probability $|U_s|/2^{|E|}$, regardless of $\Omega_G(J)$. Thus, $\phi_{rs} \approx |U_r|/2^{|E|}$, even for $r = s$. Thus, the scaling behaviour for ρ_s is distinctly different for this random mapping as compared to the mappings observed here, which show high robustness to changes in the input set of interactions.

2.3.2 Transition Probabilities

Transition probabilities are plotted in Figure 2.4 for the largest induced subgraph of each spin glass model. As found for other GP maps [1, 11], ϕ_{rs} is typically much smaller than the ρ_s found in Figure 2.2 or Figure 2.3. A random null model which states that the probability of obtaining r upon a random step is just proportional to $|U_r|$ gives the prediction

$$\phi_{rs} \approx \left(\frac{1 - \rho_s}{2^{|E|} - |U_s|} \right) |U_r| \tag{2.11}$$

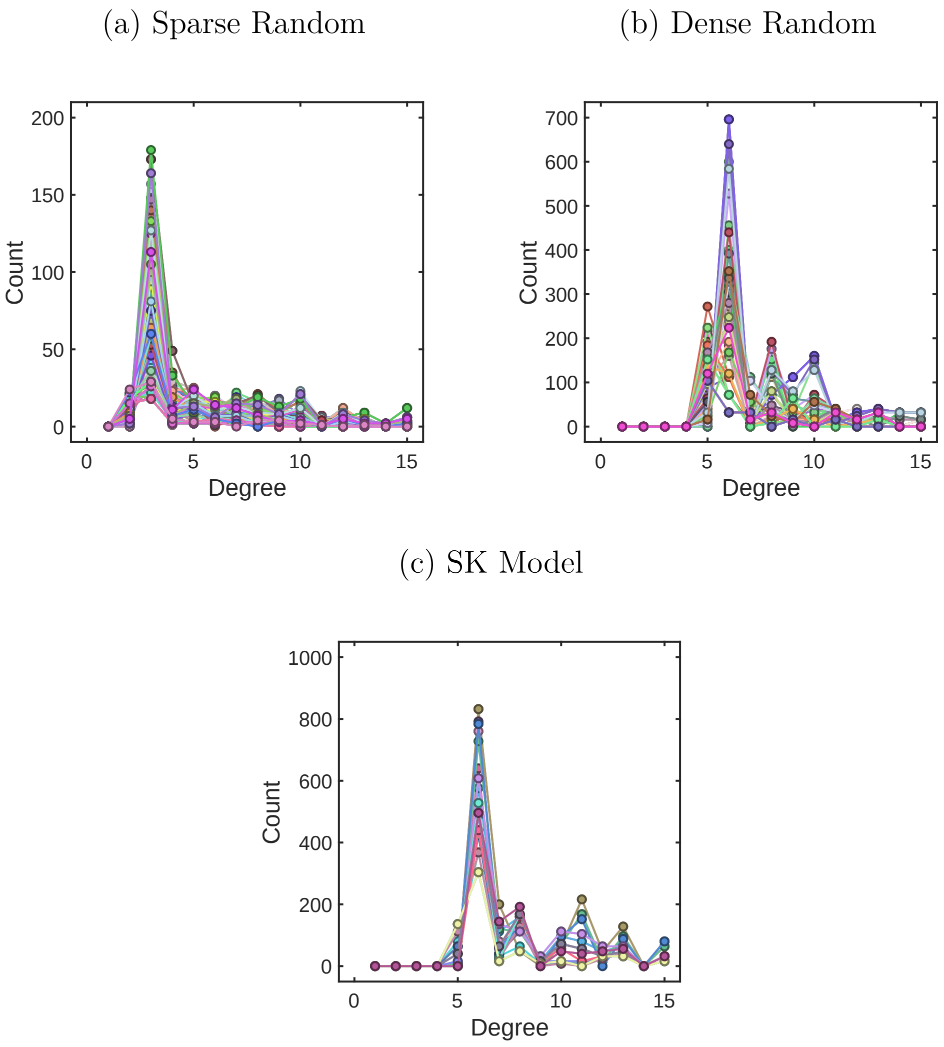

(a) Sparse Random

(b) Dense Random

(c) SK Model

Figure 2.5: Log-log plot of normalised induced subgraph vertex count $|U_s|/2^{|E|}$ (equivalent to the neutral set size) versus the rank of the size for **(a)** sparse random graph ($|V| = 9$, $|E| = 15$), **(b)** dense random graph ($|V| = 7$, $|E| = 15$), and **(c)** SK model ($|V| = 6$, $|E| = 15$).

for $r \neq s$, where the prefactor is a normalisation constant taking into account neutral steps that lead to $r = s$. Overall, as can be seen in Figure 2.4, this predicted curve does a good job, suggesting that vertices of subgraphs $\{U_r\}$ ($r \neq s$) with nonzero transition probability are approximately randomly distributed in the neighbourhoods of all vertices $v \in |U_s|$ with frequency $\approx |U_r|/2^{|E|}$.


(a) Sparse Random

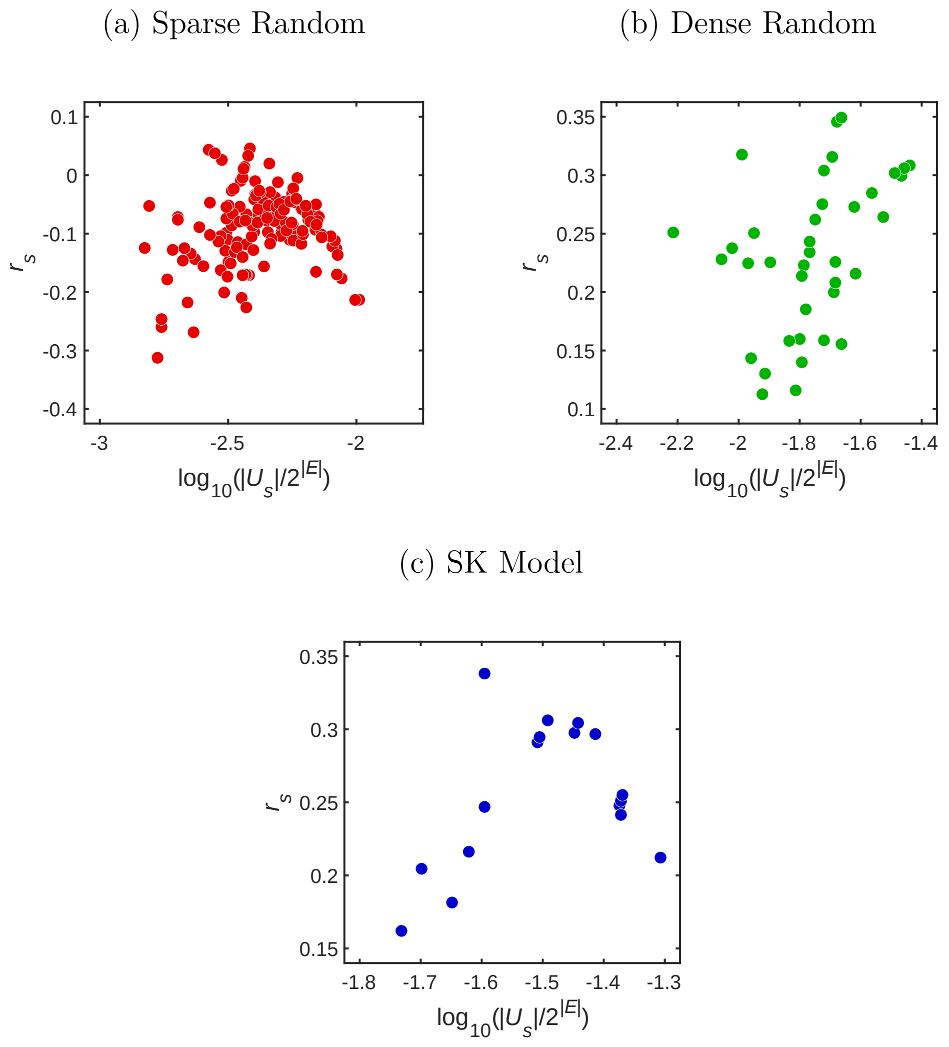
(b) Dense Random

(c) SK Model

Figure 2.6: Degree distributions of all induced subgraphs for **(a)** sparse random graph ($|V| = 9$, $|E| = 15$), **(b)** dense random graph ($|V| = 7$, $|E| = 15$), and **(c)** SK model ($|V| = 6$, $|E| = 15$). The distributions are approximately unimodal.

2.3.3 Size-Rank Distributions

In the GP map literature there has been a lot of interest in phenotype bias, the observation that the neutral set sizes can vary over many orders of magnitude, which can even determine evolutionary outcomes [14, 23] even when natural selection is also at play. In Figure 2.5 we show that such phenotype bias also exists for this spin glass system. The rank plots show a consistent behaviour independent of spin glass graph topology G. An open question is whether or not the distribution of neutral set sizes obeys a Zipf-like power law, applicable to models in which input site ordering


(a) Sparse Random

(b) Dense Random

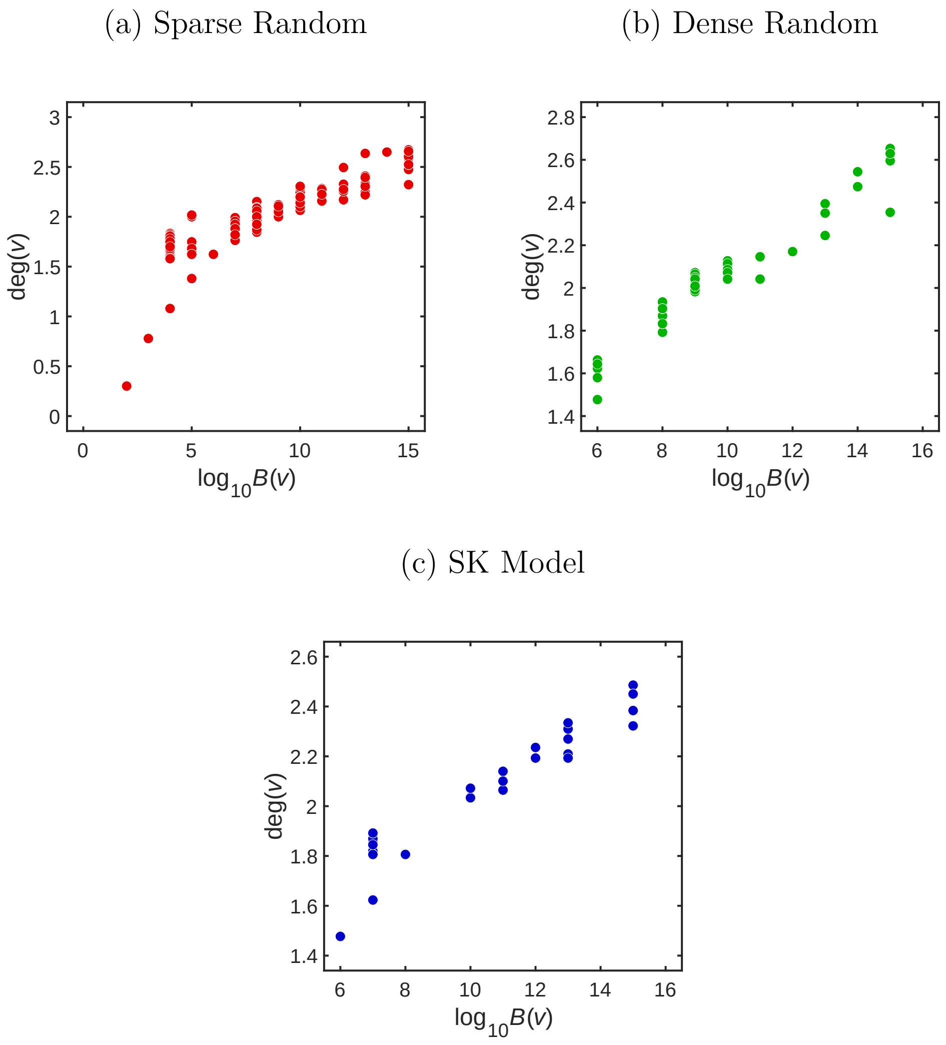
(c) SK Model

Figure 2.7: Plot of assortativity r_s of all induced subgraphs versus log of the normalised induced subgraph probability $|U_s|/2^{|E|}$ for sparse random graph ($|V| = 9$, $|E| = 15$), dense random graph ($|V| = 7$, $|E| = 15$), and SK model ($|V| = 5$, $|E| = 10$). Correlation coefficients are **(a)** sparse random graph: Spearman $\rho = 0.4188$, **(b)** dense random graph: Spearman $\rho = 0.2178$, and **(c)** SK model: Spearman $\rho = 0.4544$.

is strongly constrained (including Boolean neural networks [44, 45]), or a log-normal distribution, which appears in RNA secondary structure GP maps [14, 46]. We think that the current systems are still too small to conclusively answer this question.

2.3.4 Subgraph Degree Distributions and Clustering

In Figure 2.6, we plot the degree distribution of each subgraph for a sparse, dense, and complete graphs. These tend to be unimodal, with very few vertices attaining



(a) Sparse Random

(b) Dense Random

(c) SK Model

Figure 2.8: Plot of vertex degree v versus log of betweenness centrality $B(v)$ for all vertices v in the largest induced subgraph for sparse random graph ($|V| = 9$, $|E| = 15$), dense random graph ($|V| = 7$, $|E| = 15$), and SK model ($|V| = 6$, $|E| = 15$). Correlation coefficients are **(a)** sparse random graph: Spearman $\rho = 0.9919$, **(b)** dense random graph: Spearman $\rho = 0.9615$, and **(c)** SK model: Spearman $\rho = 0.9922$.

or coming close to attaining the maximum possible degree of $|E|$. The peak shifts toward higher degree as the size of the induced subgraph also increases, as is expected. The mean of this degree distribution is of course proportional to the edge count found earlier.

Our simulations also confirm the trivial result that clustering coefficients are always zero for induced subgraphs of hypercubes.

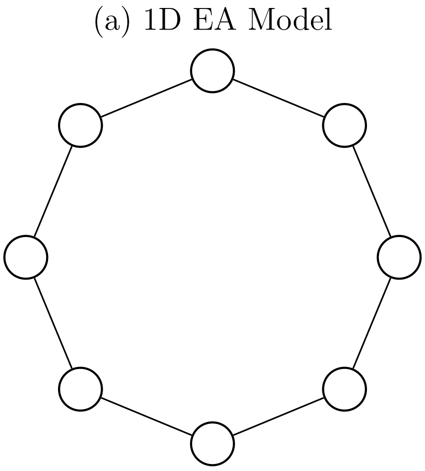
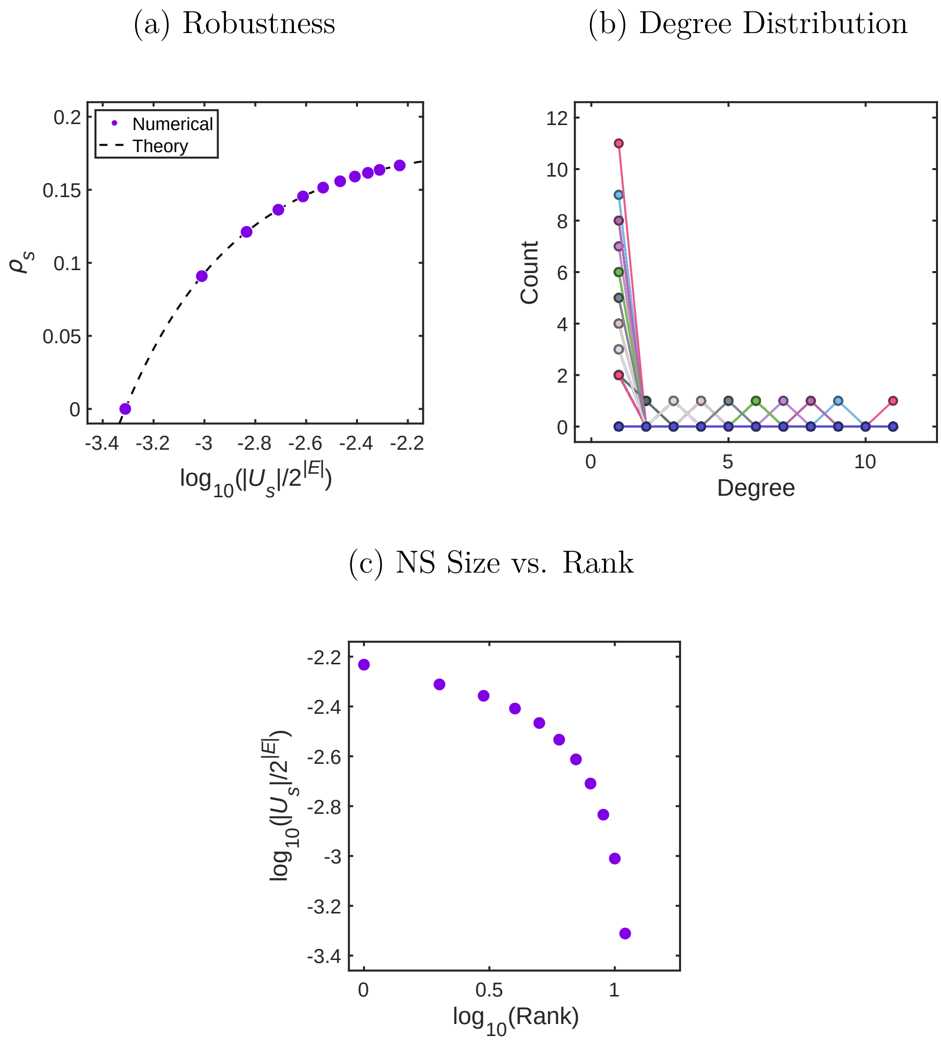


Figure 2.9: Connectivity of the 1D Edwards-Anderson model.

2.3.5 Assortativity and Betweenness Centrality

In Figure 2.7 we show the assortativity, defined in eq. (2.8), of induced subgraphs plotted against the log of induced subgraph vertex count. The values are mainly negative for the sparser graph, and positive for the denser graph and the SK model. The value of the assortativity itself is an indication of the correlation the degree of a vertex and the degree of its neighbours. Positive (negative) r_s of subgraph indicates that the degree of a vertex and the degree of its neighbours tend do be positively (negatively) correlated. In [8], a weak positive correlation between assortativity and network size was found for RNA. It may be that our systems are too small to resolve such trends.

We also plot the betweenness centrality versus degree for the largest induced subgraph in Figure 2.8. It is clear from the positive correlation found in all plots that vertices with higher degree tend to also be more central, i.e. there are more shortest paths travelling through that vertex. Such positive correlation between betweenness centrality and degree has also been observed in induced subgraphs in RNA folding GP maps [8]. These two measures—both being measures of centrality—often have positive correlation in unweighted networks [47].



(a) Robustness

(b) Degree Distribution

(c) NS Size vs. Rank

Figure 2.10: Results for 1D Edwards-Anderson model: **(a)** robustness/normalised edge count ρ_s versus neutral set size $|U_s|$ (the dashed line is the analytical result from eq. (2.12)), **(b)** degree distribution of neutral set vertices, and **(c)** neutral set size versus rank plot on log-log scale.

2.3.6 Special Case: 1D Edwards-Anderson Model

The Edwards-Anderson (EA) model is another special case which deserves individual treatment. The EA model, the original theory of spin glasses [32], is simply a spin glass on a lattice with nearest neighbour interactions only. For the 1D Edwards-Anderson (EA) model with periodic boundary conditions, the topology of which is shown in Figure 2.9, the behaviour of ρ_s is analytically tractable from the degree distribution. The 1D EA model has $|E| = |V|$, so for $|V| > 5$, $G(V, E)$

is sparsely connected for the 1D EA topology. The degree distribution of each subgraph indicates that for a subgraph of size $|U_s|$, there are exactly $|U_s|-1$ vertices with degree 1 and 1 vertex with degree $|U_s|-1$. This means, for the 1D EA model, an induced subgraph $H_s(U_s, F_s)$ is exactly a star graph $S_{|U_s|-1}$, which has $|F_s| = |U_s| - 1$ edges. Thus, it immediately follows that

$$\rho_s = \frac{1}{|U_s||E|} \sum_{v \in U_s} \deg(v) = \frac{2}{|E|} \left(1 - \frac{1}{|U_s|}\right) \tag{2.12}$$

The points in Figure 2.10(a) fall exactly along this curve. Fig. 2.10(b) shows that the degree distribution is quite simple. A size-rank plot is also shown for the 1D EA model in panel (c). For the 2D EA we could not find an analytically tractable ρ_s.

Remark: pathological examples. The 1D EA model serves as a example in which linear-log scaling is not seen for ρ_s, and other topological properties of induced subgraphs may not behave exactly the same way as in the representative figures which describe the overwhelming majority of random graphs found in our simulations. It is not surprising that a small set of pathological examples exists for this input-output map, as they could for any input-output map. However, the major takeaway is that the ground states are obviously robust to changes in the interactions, even for the 1D EA model shown here. The relationship between edge count and subgraph vertex count exhibited in the 1D EA model (and of course the general cases discussed prior) display ρ_s much higher that would be expected from a random mapping Ω_G of inputs to outputs.

2.4 Discussion

In this chapter, we probed the robustness and stability of the ground states of $\pm J$ spin glasses on random graphs to perturbations of the interactions. Through numerical simulation, we calculated the properties of induced subgraphs of the hypercube graph which all map to the same ground state output. Such properties include the relationship between edge count and vertex count, the probability of transition from ground state to another after sign flip of a single interaction, the size-rank distribution of the induced subgraphs, degree distribution, clustering coefficients, assortativity, and betweenness centrality. In addition to sparsely and densely connected random graphs, we also studied the special cases of the Sherrington-Kirkpatrick model and 1D Edwards-Anderson model, the latter of which has an analytically tractable relationship between edge and vertex count.

Our main result is that these ground states have relatively large sets of interactions J that map to them. In other words there is a fair amount of redundancy. Moreover, the ground states are remarkably robust to flipping the interactions. Robustness was found to scale as the log of the size of these (neutral) sets, quite similarly to what has been observed for many different GP maps [1, 12, 13, 18, 19]; a caveat is that our systems are quite small, so that the scaling was not observed over much more than an order of magnitude in frequency. Another interesting result, also seen for GP maps [1, 11] is that, in contrast to the robustness, the transition probabilities, defined as the likelihood of a flip of the spin yielding a different ground state, do scale proportionally to the neutral set size, as one would expect from a random model. The similarity to the GP map behaviour suggests that there may be a more universal argument (based for example on algorithmic information theory [48, 49]) for these scaling properties.

Our spin-glass models are relatively small, because, as is well known, finding the ground state of a spin glass can be typically computationally expensive and difficult in general. Depending on graph topology, finding the ground state of a spin glass can be NP-hard [43]. Knowledge that the robustness of a ground state is large may potentially offer improvements to ground-state finding algorithms by providing a measure of stability of a certain ground states as a function of parameter space. It would also be interesting to check some of our results on significantly larger networks. We find, for example, that all our subgraphs that map to the same ground state are connected by single flips of the J_{ij}s. Will this percolation property hold for larger systems, or will these subgraphs start to fragment?

Mapping epistatic interactions onto spin glass Hamiltonians to derive sequence to fitness maps has been especially important for the study of viral evolution [37–41].

These models typically have continuous J, and so it will be interesting to see if the kind of scaling properties of robustness that we find here for the $\pm J$ spin glasses carry over to these more complex systems. If, as we expect, a concept akin to high robustness persists, then this may have implications for the stability of fitness peaks to environmental changes.

Another potentially interesting future direction of research is to explore these results about robustness in the complementary setting of the sensitivity of Boolean functions [50–53]. It would be interesting to see whether similar high robustness/low sensitivity results can be found in this related context.

3
Phase Transition Between Fragile and Robust Phases of Input-Output Maps

Contents

3.1 Introduction

We have seen in Chapter 2 and in refs. [1, 15, 18–20] that natural genotype-phenotype or input-output maps seem to obey universal scaling laws for robustness and transition probabilities, among other properties. Here, we propose a maximum entropy model with only a single constraint which reproduces the universal scaling laws for many properties without those properties themselves being constrained. Our statistical mechanical model exhibits phase transition-like behaviour, separating a robust and a fragile phase, each of which has distinct scaling laws for robustness and other topological properties of neutral networks. We conduct numerical simulations to explore these phases and the network-topological properties they predict for neutral networks. We then examine examples of analytical biological GP maps, and show that in the large sequence limit they tend towards the robust phase.

3.2 Minimally Constrained Maximum Entropy Model of Input-Output Maps

Consider the sequence-based input-output map

$$\Omega : \{0, \dots, k-1\}^d \to \{0, \dots, q-1\} \tag{3.1}$$

from the space of sequences of length d from an alphabet with k characters to a finite set of discrete outputs with cardinality q. The input space of sequences of length d and alphabet of k characters comprises the vertex set of the Hamming graph $H_{d,k} = (K_k)^{\square d}$, which is the Cartesian product of d copies of the complete graph K_k, and an edge exists between two vertices if the Hamming distance between the two sequences is 1. Let V be the vertex set and E the edge set of $H_{d,k}$. Thus, Ω induces a graph partition of $H_{d,k}$ such that the induced subgraph vertex set $V_n = \{x \in \{0, \dots, k-1\}^d \,|\, \Omega(x) = n\}$. Now, each induced subgraph, which has

$V_n \subseteq V$ and $E_n \subseteq E$, called a *neutral space* in the evolutionary biology literature, contains all vertices which map to the same observable output.

We now define robustness and sensitivity as they are used in evolutionary biology and computer science, respectively.

3.2.1 Preliminary Definitions

Definition 3.2.1. *The* ***(local) robustness*** $\rho(\Omega, x)$ *of an input-output map* Ω *given an input* $x \in \{0, \dots, k-1\}^d$ *is the fraction of characters in* x *which can be changed without changing* $\Omega(x)$*. Equivalently, we can use our graph-theoretic definition. Suppose* $\Omega(x) = n$*. Then, the vertex* $x \in V_n$ *belongs to the induced subgraph* V_n*. Then,*

$$\rho(\Omega, x) = \frac{\deg_{G_n}(x)}{d(k-1)}, \tag{3.2}$$

where $\deg_{G_n}(x)$ *is the degree of* x *in the induced subgraph in* G_n*.*

Definition 3.2.2. *The* ***robustness*** $\rho_n(\Omega)$ *of an output* $n \in \{0, \dots, q-1\}$ *is given by the average of the local robustness over all vertices in the induced subgraph* G_n*:*

$$\rho_n(\Omega) = \frac{1}{|V_n|} \sum_{x \in V_n} \rho(\Omega, x) = \frac{2|E_n|}{d(k-1)|V_n|}. \tag{3.3}$$

Robustness in evolutionary systems is typically calculated for each phenotype, and the scaling relation between the robustness ρ_n and the size of the induced subgraph $|V_n|$ is observed. This is equivalent to understanding the general scaling relationship between the number of edges and number of vertices of the induced subgraph $|V_n|$.

In this chapter, we hypothesize that an aggregate measure of robustness for an input-output map, which we call *global robustness*, is by itself a sufficient constraint to reproduce all of the important topological features of genotype-phenotype maps that have been observed in nature, including the scaling laws for robustness, transition probabilities, number of neutral components per phenotype, size of largest neutral component, etc. The global robustness is a collective measure of the robustness of all the outputs in an input-output map, weighted by the output frequencies:

Definition 3.2.3. *We define the **global robustness** of an input-output map as the frequency-weighted average robustness*

$$\mathbb{E}_n\left[\rho_n\right] \equiv \sum_{n=0}^{q-1} \mathbb{P}(\Omega(x)=n)\rho_n(\Omega) = \sum_{n=0}^{q-1} f_n \rho_n(\Omega), \tag{3.4}$$

where $\rho_n(\Omega)$ is the robustness of the n-th output for the input-output map Ω on a Hamming graph $H_{d,k}$, and $f_n = |V_n|/k^d$ is the frequency of the output appearing, which is identical to the number of vertices in the output neutral network divided by the total number of vertices in the Hamming graph, which is k^d.

We note that the frequency distribution often varies across several orders of magnitude in natural GP maps [48, 54, 55]; as a result, the global robustness value of the global robustness will mostly be determined by the robustness of the largest phenotypes. In Section 3.2.2, by using a maximum entropy approach and constraining *only* the global robustness, we will discuss the phase transition-like behaviour of the statistical mechanical model, most importantly showing that each of these scaling laws are actually heavily dependent on each other.

Connections Between the Sensitivity of Boolean Functions and Robustness

Here, we also point out an important connection between robustness and a mathematical counterpart which appears in computer science, *sensitivity*, which measures the likelihood that flipping a single input bit will alter the output bit of Boolean functions $f : \{0,1\}^d \to \{0,1\}$, that map binary sequences of length d onto a single binary output. In other words, low sensitivity corresponds to high robustness. Huang's [51] famously short solution, published in 2019, to the decades-old Sensitivity Conjecture [50] concerning induced subgraphs of the n-dimensional hypercube graph has recently brought a great deal of attention to the sensitivity analysis of Boolean functions. While scaling laws for sensitivity are typically not measured in the manner done for biological robustness, there are many mathematical equivalencies between the two. In the literature, see e.g., [50–53]), quantities are defined such as local sensitivity, function sensitivity, and average sensitivity, which are similar to biological robustness. In this chapter, we do not explicitly use these definitions from sensitivity analysis, instead opting for the GP map inspired ones, but point out that there is a seemingly understudied connection between robustness in biological systems and sensitivity in Boolean functions. Moreover, the notion of average sensitivity is used to motivate a similar parameter we call "global robustness" our input-output maps, which we believe can function as a master constraint on many different network topological properties of neutral networks.

Boolean functions $f : \{0,1\}^d \to \{0,1\}$ are a special case of the input-output maps discussed above, with $k = 2$ and $q = 2$. Sensitivity is closely related to evolutionary robustness.

Definition 3.2.4. *The* ***(point) sensitivity*** *[56] $s(\Omega, x)$ of a Boolean function (Ω defined above, with $k = 2$ and $q = 2$) at a particular input $x \in \{0,1\}^n$ is the number of elements of x which, when changed, result in $\Omega(x)$ also changing. This is exactly*

$$s(\Omega, x) = d(1 - \rho(\Omega, x)). \tag{3.5}$$

Definition 3.2.5. *The* ***average sensitivity*** *[56] of a Boolean function is the average of the point sensitivity over all inputs*

$$\overline{s}(\Omega) = \frac{1}{2^d} \sum_x s(\Omega, x). \tag{3.6}$$

The sensitivity of a Boolean function commonly studied in computer sciencce is different from average sensitivity, and it is given by $\max_x(s(\Omega, x))$.

From the definitions above, one sees that, for $k = 2$ and $q = 2$, the average sensitivity is proportional to the sum of the robustness values for all outputs weighted by the size of the corresponding induced subgraph:

$$\begin{aligned} \overline{s}(\Omega) &= d\left[1 - \frac{1}{2^d}(\rho_0(\Omega)|V_0| + \rho_1(\Omega)|V_1|)\right] \\ &= d\left(1 - \frac{|E_0| + |E_1|}{d2^{d-1}}\right). \end{aligned} \tag{3.7}$$

Note that average sensitivity is proportional to the fraction of edges which belong to any induced subgraph. Another definition relevant in computer science appears here as well: in the analysis of Boolean functions, the *influence* of the i-th (with $i \in \{1, \dots, d\}$) site in the input sequence is defined as the probability that flipping the bit at the i-th site will change the output. Therefore, the sum of the influences [56] of each site is exactly equal to the average sensitivity as defined in eq. (3.6). In the case of evolution, where outputs need not be Boolean, we could have $q \geq 2$. We could now consider the following equalities, which resemble the sum of influences or average sensitivity:

$$\begin{aligned} \frac{2}{k^d d(k-1)} \left| \bigcup_{n=0}^{q-1} E_n \right| &= \frac{2}{k^d d(k-1)} \sum_{n=0}^{q-1} |E_n| \\ &= \sum_{n=0}^{q-1} \frac{|V_n|}{k^d} \rho_n(\Omega) = \sum_{n=0}^{q-1} \mathbb{P}(\Omega(x) = n) \rho_n(\Omega) = \mathbb{E}_n \left[\rho_n \right]. \end{aligned} \tag{3.8}$$

Here, $\mathbb{P}(\Omega(x) = n)$ is the probability of finding an output $n \in \{0, \dots, q-1\}$ after randomly sampling an input x from a uniform distribution on the space of inputs. In the final step we note that this is exactly equal to the global robustness.

3.2.2 Deriving the Canonical Ensemble with Maximum Entropy

In order to be able to apply some techniques from statistical physics, as well as those from graph theory, we define, for a specific input-output map Ω from eq. (3.1), a cost function (akin to a Hamiltonian in physics) which we propose is minimised in natural input output maps. Denote the negative of the global robustness defined above as $\mathcal{H}(\Omega)$, which we showed to be directly proportional to the total number of edges induced by the map on the Hamming graph:

$$\begin{aligned} \mathcal{H}(\Omega) = -\mathbb{E}_n \left[\rho_n \right] &= -\frac{2}{k^d d(k-1)} \sum_{n=0}^{q-1} |E_n| \\ &= -\frac{2}{k^d d(k-1)} \sum_{\{x,y\} \in E} \delta(\Omega(x), \Omega(y)), \end{aligned} \tag{3.9}$$

where $\delta(\Omega(x), \Omega(y))$ is the Kronecker delta, and the sum is performed over pairs of vertices $\{x, y\}$ connected by each edge in $H_{d,k}$. We use the notational convention $\mathcal{H}(\Omega)$ to refer to the negative global robustness because the negative global robustness now resembles the Hamiltonian of a classical Potts model in statistical physics. This now makes $\Omega(x) \in \{0, \dots, q-1\}$ a classical Potts spin, and the negative global robustness counts all sets of adjacent spins that map to the same value $\Omega(x)$, i.e. the number of edges in the Hamming graph whose two vertices map to the same output. Then the mean value of the negative global robustness $\langle \mathcal{H}(\Omega) \rangle_\Omega$ over some ensemble is then equal to the mean total number of edges over that particular ensemble of maps.

Define

$$f_n = \mathbb{P}(\Omega(x) = n) = \frac{|V_n|}{k^d} \tag{3.10}$$

to be the probability of finding an output $n \in \{0, \dots, q-1\}$. The vector $\mathbf{f} = (f_0, \dots, f_{q-1})$ specifies the sizes of all of the induced subgraphs of $H_{d,k}$. The L^1-norm $\|\mathbf{f}\|_1 = 1$. Let $\mathbb{P}(\Omega \,|\, \mathbf{f})$ be the probability of selecting an input-output map Ω given the constraint that the subgraph sizes are given by $\mathbf{f}$. Then the question arises, what is the least biased distribution over all possible maps Ω? We then follow the classical maximum entropy (MaxEnt) strategy [57] under external constraints. The idea is that the least-biased probability $\mathbb{P}(\Omega|\mathbf{f})$, given a set of constraints, is the one that maximises the entropy under the same constraints. In statistical mechanics, this principle was first proposed by E.T. Jaynes [57], who showed that, for example, the canonical ensemble is obtained when maximising the Shannon information entropy under the constraints of normalisation and a fixed average energy. Here we follow the same strategy. For a fixed set of subgraph sizes given by $\mathbf{f}$, we will maximise the Shannon entropy under the usual constraint of a normalised probability, and we add a constraint on the negative global robustness, which is equivalent to a constraint on the average of the Potts model Hamiltonian. This is written down as follows in terms of Lagrange multipliers:

$$\begin{aligned} S[\mathbb{P}(\Omega \,|\, \mathbf{f})] = &- \sum_{\Omega \,|\, \mathbf{f}} \mathbb{P}(\Omega \,|\, \mathbf{f}) \log \mathbb{P}(\Omega \,|\, \mathbf{f}) \\ &- \Lambda \left(\sum_{\Omega \,|\, \mathbf{f}} \mathbb{P}(\Omega \,|\, \mathbf{f}) - 1 \right) - \frac{1}{T} \left(\sum_{\Omega \,|\, \mathbf{f}} \mathbb{P}(\Omega \,|\, \mathbf{f}) \mathcal{H}(\Omega) - C \right), \end{aligned} \tag{3.11}$$

where C is a constant, and Λ and $1/T$ are Lagrange multipliers. We use T to resemble the "temperature" in statistical physics; implemented as a Lagrange multiplier, T

modulates the strength of the constraint on the negative global robustness. High T corresponds to a weaker global robustness constraint, and as T becomes lower, the constraint becomes stronger. We now maximise S with respect to $\mathbb{P}(\Omega \,|\, \mathbf{f})$, first by taking the functional derivative with respect to $\mathbb{P}(\Omega \,|\, \mathbf{f})$ and setting it to zero:

$$\frac{\delta S}{\delta \mathbb{P}(\Omega \,|\, \mathbf{f})} = -\log \mathbb{P}(\Omega \,|\, \mathbf{f}) - \Lambda - \frac{1}{T}\mathcal{H}(\Omega) = 0, \tag{3.12}$$

so

$$\mathbb{P}(\Omega \,|\, \mathbf{f}) = \exp\left[-\Lambda - \frac{1}{T}\mathcal{H}(\Omega)\right]. \tag{3.13}$$

Using the normalisation condition $\sum_{\Omega \,|\, \mathbf{f}} \mathbb{P}(\Omega \,|\, \mathbf{f}) = 1$, we have that

$$e^{\Lambda} = \sum_{\Omega \,|\, \mathbf{f}} e^{-\frac{\mathcal{H}(\Omega)}{T}}. \tag{3.14}$$

Defining $Z(\mathbf{f}) \equiv e^{\Lambda}$ yields a partition function

$$Z(\mathbf{f}) = \sum_{\Omega \,|\, \mathbf{f}} \prod_{\{x,y\} \in E} \exp\left[\frac{1}{T}\frac{2}{k^d d(k-1)}\delta(\Omega(x), \Omega(y))\right] \tag{3.15}$$

which is the partition function of the classical Potts model for an *a priori* determined $\mathbf{f}$. We have now defined the equivalent of a canonical ensemble on the space of input-output maps with a fixed set of output frequencies $\mathbf{f}$.

Thus, the probability distribution becomes

$$\mathbb{P}(\Omega \,|\, \mathbf{f}) = \frac{1}{Z(\mathbf{f})} \exp\left[\frac{1}{T}\frac{2}{k^d d(k-1)}\delta(\Omega(x), \Omega(y))\right]. \tag{3.16}$$

Robustness can be written as

$$\begin{aligned}\rho_n(\Omega) &= \frac{2|E_n|}{k^d d(k-1) f_n} \\ &= \sum_{\{x,y\} \in E} \frac{2\delta(\Omega(x), \Omega(y))\delta(\Omega(x), n)}{k^d d(k-1) f_n} \\ &\equiv -\frac{1}{f_n}\mathcal{H}_n(\Omega),\end{aligned} \tag{3.17}$$

where $\mathcal{H}_n(\Omega)$ is one of the terms in the expansion of the negative global robustness:

$$\mathcal{H}(\Omega) = \sum_{n=0}^{q-1} \mathcal{H}_n(\Omega). \tag{3.18}$$

3.3 Unconstrained Maximum Entropy (High T Limit)

If MaxEnt is performed under no constraints aside from normalisation, then the entropy-maximising distribution is the uniform distribution, and all states have equal probability:

$$\mathbb{P}(\Omega\,|\,\mathbf{f}) = \frac{1}{Z_0(\mathbf{f})}, \tag{3.19}$$

where

$$Z_0(\mathbf{f}) \equiv \sum_{\Omega\,|\,\mathbf{f}} 1 = \frac{(k^d)!}{\prod_{n=0}^{q-1}(k^d f_n)!}. \tag{3.20}$$

This is equivalent to the $T \to \infty$ limit (i.e. the "high temperature" limit). In the unconstrained MaxEnt scenario (the $T \to \infty$ limit), the robustness of the n-th output is

$$\begin{aligned}\lim_{T\to\infty} \langle \rho_n(\Omega)\rangle &= \frac{2}{k^d d(k-1) f_n}\frac{1}{Z_0(\mathbf{f})}\sum_{\Omega\,|\,\mathbf{f}} |E_n| \\ &= \frac{2}{k^d d(k-1) f_n}\frac{1}{Z_0(\mathbf{f})}\frac{(k^d(1-f_n))!}{\prod_{a\neq n}(k^d f_a)!}\sum_{\Omega\,|\,f_n} |E_n|.\end{aligned} \tag{3.21}$$

In the above calculation, $\sum_{\Omega\,|\,\mathbf{f}} |E_n|$ is the sum of the number of edges in the n-th subgraph taken over all configurations Ω given the output frequencies $\mathbf{f}$. Consider the vertices which do not belong to the n-th subgraph for a particular configuration Ω; these vertices can be swapped or rearranged in exactly $\frac{(k^d(1-f_n))!}{\prod_{a\neq n}(k^d f_a)!}$ ways without changing the number of edges in the n-th subgraph. Therefore,

$$\sum_{\Omega\,|\,\mathbf{f}} |E_n| = \frac{(k^d(1-f_n))!}{\prod_{a\neq n}(k^d f_a)!}\sum_{\Omega\,|\,f_n} |E_n|, \tag{3.22}$$

where on the right hand side we are summing over $\sum_{\Omega\,|\,f_n}$, the unique configurations Ω having only constrained the number of vertices in the n-th subgraph. There are $\binom{k^d}{k^d f_n}$ such configurations. The average number of edges in an induced subgraph of a Hamming graph with $k^d f_n$ vertices is thus given by

$$\binom{k^d}{k^d f_n}^{-1}\sum_{\Omega\,|\,f_n} |E_n|. \tag{3.23}$$

The authors of ref. [58] calculated this average exactly using a combinatorial proof. First, they calculate the number of edges in all possible induced subgraphs of the Hamming graph $H_{d,k}$; then, they manipulate their expression to determine the multiplicity of graphs in which a particular vertex has some fixed degree. Lastly, they use knowledge of how many available neighbours any vertex has in a Hamming graph to simplify their sum. They prove that an induced subgraph of $H_{d,k}$ which has $k^d f_n$ vertices has an average number of edges given by

$$\binom{k^d}{k^d f_n}^{-1} \sum_{\Omega \,|\, f_n} |E_n| = \frac{d(k-1)(k^d f_n)(k^d f_n - 1)}{2(k^d - 1)}. \tag{3.24}$$

Now substituting back into equation (eq. (3.21)), we can simplify to show that

$$\lim_{T\to\infty} \langle \rho_n(\Omega) \rangle = \frac{k^d f_n - 1}{k^d - 1}. \tag{3.25}$$

This result for robustness of the n-th output is obtained when the constraint on the negative global robustness is not included during entropy maximisation (equivalently, $T \to \infty$ in the entropy definition). This ensemble-averaged value is in fact also the *exact* robustness of the random null model, in which inputs are randomly assigned to the outputs. Our result in eq. (3.25) simplifies to the null expectation $\rho_n(\Omega) \approx f_n$ in the $d \to \infty$ ("thermodynamic") and/or $k \to \infty$ limit.

We can now calculate the asymptotic bound on the ensemble-averaged negative global robustness for unconstrained MaxEnt:

$$\begin{aligned} \lim_{T\to\infty} \langle \mathcal{H}(\Omega) \rangle &= \sum_{n=0}^{q-1} \lim_{T\to\infty} \langle \mathcal{H}_n(\Omega) \rangle \\ &= -\sum_{n=0}^{q-1} \lim_{T\to\infty} f_n \langle \rho_n(\Omega) \rangle = -\frac{1}{k^d - 1}\left(k^d \sum_{n=0}^{q} f_n^2 - 1\right). \end{aligned} \tag{3.26}$$

Although we considered the weak constraint (high T limit), it is also interesting to consider more intermediate values of T and the behaviour of the negative global robustness as the constraint becomes stronger. To do so, we must be able to calculate the partition function in eq. (3.15). This problem appears to be analytically intractable, but in Appendix A we consider the special case for Boolean outputs $q = 2$. We perform a cluster (high T) expansion of the partition function and are able to calculate the coefficients of this expansion analytically using combinatorics. This may be useful for better understanding the behaviour of the system's global robustness as the constraint on maximum entropy strengthens.

3.4 Strongly Constrained Maximum Entropy (Low T Limit)

We now consider the case in which the constraint on the maximum entropy is strong; this corresponds to the $T \to 0$ limit, the "low temperature" limit. When the entropy constraint is strong, we know from statistical physics that the log of the partition function (times $-T$) is equal to the minimum of the Hamiltonian, which is the negative global robustness:

$$\lim_{T\to 0}\left[-T\log Z(\mathbf{f})\right] = \min_{\Omega\,|\,\mathbf{f}} \mathcal{H}(\Omega). \tag{3.27}$$

This means that in the strong constraint limit, we are trying to maximise the global robustness (minimise the negative global robustness). This is directly difficult to do because it depends on distribution of the number of vertices in each subgraph corresponding to a different output. The best approximation we can make is to assume that each of the individual subgraphs has its robustness maximised:

$$\min_{\Omega\,|\,\mathbf{f}} \mathcal{H}(\Omega) = \min_{\Omega\,|\,\mathbf{f}} \sum_{n=0}^{q-1} \mathcal{H}_n(\Omega) \le \sum_{n-0}^{q-1} \min_{\Omega\,|\,f_n} \mathcal{H}_n(\Omega), \tag{3.28}$$

where we recall that $\mathcal{H}_n(\Omega)/f_n \equiv \rho_n(\Omega)$. The inequality becomes an equality for some special cases of $\mathbf{f}$. For instance, when

$$\mathbf{f} = k^{-d}\left(1,\ \underbrace{1,\ldots,1}_{k-1\text{ copies}},\ \underbrace{k,\ldots,k}_{k-1\text{ copies}},\underbrace{k^2,\ldots,k^2}_{k-1\text{ copies}},\ldots,\underbrace{k^{d-1},\ldots,k^{d-1}}_{k-1\text{ copies}}\right), \tag{3.29}$$

$\mathcal{H}(\Omega)$ is minimised when each of the $\mathcal{H}_n(\Omega)$ is minimised[1]. Moreover, we find numerically that even outside of these cases, taking the inequality as an approximate

[1] This is the case because we can take $k^{\widetilde{d}}$ vertices (with $\widetilde{d} \in \{0, 1, \ldots, d-1\}$) and construct an induced subgraph of the main Hamming graph $H_{d,k}$ which is itself a small Hamming graph $H_{\widetilde{d},k}$. In this case, each neutral network has maximal possible robustness, which means the global robustness is maximised.

equality works well (see Section 3.5). So, we work in the approximation that the maximum of the global robustness (which is the frequency-weighted average of all of the output robustnesses) is approximately the same as the frequency-weighted sum of the maximal robustnesses for each output on its own.

We discuss the subgraphs of a Hamming graph which maximise robustness thoroughly in Chapter 4, where we use a result from coding theory to prove the exact maximal robustness. The true maximal robustness is given by a complicated expression which is actually a fractal. Instead of working with such an expression, we use an approximation to the exact maximal robustness, which has been proven in ref. [58] as well, though it is easy to see this approximation in the following chapter as well.

An induced subgraph of $H_{d,k}$ which has $k^d f_n$ vertices has number of edges with an upper bound

$$\begin{aligned} |E_n| &\leq \frac{(k-1)(k^d f_n)\log\left(k^d f_n\right)}{2\log k} \\ &= \frac{d(k-1)(k^d f_n)}{2} + \frac{(k-1)(k^d f_n)\log f_n}{2\log k}. \end{aligned} \tag{3.30}$$

Thus, we have that

$$\begin{aligned} \min_{\Omega\,|\,f_n} \mathcal{H}_n(\Omega) &= -\frac{2}{k^d d(k-1)} \max_{\Omega\,|\,f_n} |E_n| \\ &\leq -f_n\left(1+\frac{\log f_n}{d\log k}\right). \end{aligned} \tag{3.31}$$

Immediately, we have

$$\max_{\Omega\,|\,f_n} \rho_n(\Omega) \leq 1 + \frac{\log f_n}{d\log k}, \tag{3.32}$$

from which it follows that

$$\lim_{T\to 0} \langle \rho_n(\Omega)\rangle = -\frac{1}{f_n}\lim_{T\to 0} \langle \mathcal{H}_n(\Omega)\rangle \approx 1 + \frac{\log f_n}{d\log k}, \tag{3.33}$$

which is the exact result given by generalising the Fibonacci model genotype-phenotype map, proposed by Greenbury and Ahnert [13]. In particular, we have shown that in the low temperature limit, the relationship between robustness and induced subgraph size is $\mathcal{O}(\log f_n)$, which is the trend observed in biological systems such as RNA secondary structure maps, protein folding maps such as the HP and Polyomino model, Boolean threshold networks, as well as spin glass input-output maps (as we showed in the previous chapter).

We can also calculate the minimum value of the negative global robustness (and accordingly, the maximum value of the global robustness) for the entire input-output map (i.e. the $T \to 0$ limit):

$$\begin{aligned} \min_{\Omega\,|\,\mathbf{f}} \mathcal{H}(\Omega) &= \lim_{T\to 0} \langle \mathcal{H}(\Omega) \rangle = \sum_{n=0}^{q-1} \lim_{T\to\infty} \langle \mathcal{H}_n(\Omega) \rangle \\ &= -\left(1 + \frac{1}{d \log k} \sum_{n=0}^{q-1} f_n \log f_n \right) . \end{aligned} \tag{3.34}$$

3.5 Numerical Methods: Markov Chain Monte Carlo Simulation of Potts Model

In sections Section 3.3 and Section 3.4, we provided an exact analytical treatment of the maximum entropy model's high T and low T limits, which respectively show that the system reproduces distinct scaling laws that correspond, respectively, to the behaviours seen in the random null model GP map and the logarithmic scaling observed in the biological RNA secondary structure GP map [1, 8], the HP protein folding map [1], the Polyomino protein self-assembly map [1], genetic algorithms [19], Boolean threshold networks [18], and spin glasses (Chapter 2). We are also interested in the temperature-dependent behaviour and would like to explore whether other scaling laws are likely to emerge from the model as well at different simulation temperatures. Temperature here, implemented as a Lagrange multiplier in the Shannon entropy expression, modulates the strength of the constraint on the frequency-weighted average of the robustness (also defined earlier as being proportional to "global robustness").

In the canonical ensemble formulation, the system can be equilibrated at a particular temperature. To understand the equilibrium properties of our model, we

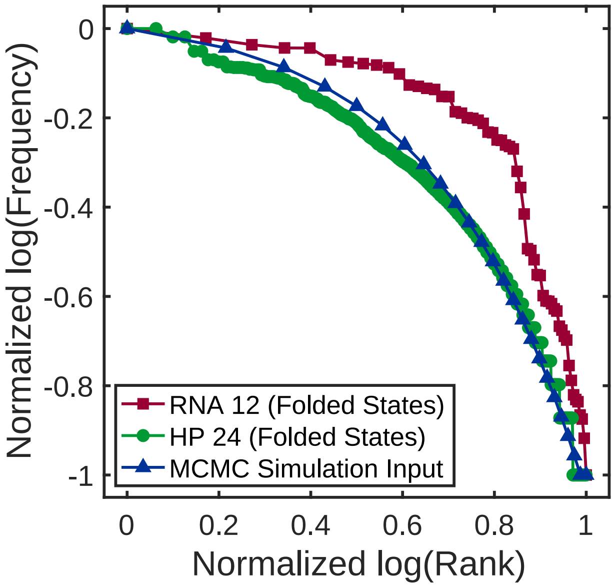


Figure 3.1: Log-log plot of frequency versus rank for the MCMC simulation frequency vector $\mathbf{f}$ plotted alongside the frequency-rank plots for the distributions of the phenotype frequencies for RNA12 and HP24 molecules (with the unfolded state omitted). The RNA secondary structure map is known to have a rank-frequency plot that scales as $f(r) \sim \frac{\log r + a}{r}$, where r is the rank, and a is some constant [46]. Other folding systems are expected to have a Zipf-law like distribution where $\log f(r) \propto \log r$. Here, we choose an exponential distribution for our MCMC simulation in order to uniformly cover the log(frequency) axis with data points, but we do not believe the frequency distribution would affect the qualitative results or the conclusions.

employ a Markov chain Monte Carlo (MCMC) simulation—namely, the Metropolis-Hastings algorithm [59, 60]. This technique allows for estimation of ensemble-averaged observables without sampling exactly over the entire configuration space, instead using a Markov chain to estimate the observables in a time-dependent fashion. In our simulations, we are exploring the configuration space of all possible input-output maps Ω for an *a priori* fixed frequency vector $\mathbf{f}$, as described in the partition function formulation earlier. This system can be mapped to a classical ferromagnetic Potts model on a Hamming graph with fixed frequencies for each Potts state.

3.5.1 Initialisation and Choice of Frequency Distribution

We first initialise the system at a particular configuration. For this to happen, the parameters of the Hamming graph corresponding to alphabet size k and sequence length d are chosen *a priori* and fixed for the entire simulation. We use the combinations $k = 2$ with $d = 24$ and $k = 4$ with $d = 12$, as this will allow for direct comparison with natural GP map robustness, transition probability, and neutral component size/number data for the HP24 (length 24) protein folding map and the RNA12 (length 12) secondary structure map, respectively, which is published in ref. [1]. The distribution of output frequencies $\mathbf{f}$ is also determined *a priori* and fixed for the entirety of the simulation. For both of these simulations we have 25 outputs, whose frequencies are given by an exponential distribution:

$$f_0 = 1 \quad \text{and} \quad f_i = 2^{i-1}, i = \{1, \ldots, 24\}. \tag{3.35}$$

We do not expect the actual choice of distribution to significantly affect the overall trends in robustness and other network topological properties in part because differences observed between biological GP maps in this property do not seem to affect global trends in scaling of the robustness. However, the exponential distribution is chosen for two reasons: (1) this allows for maximum spread of data points along the $\log f$ axes (abscissa) for the robustness, transition probability, and neutral component size and number plots shown in the results section, and (2) as shown in the log-log frequency-rank plot Figure 3.1, the exponential distribution resembles the distribution of *folded* phenotypes/outputs in the HP24 and RNA12 models. Each of these models has one very large phenotype that corresponds to unfolded states, but the overall distribution for the smaller frequencies appears to be well-mimicked by the exponential distribution we have chosen here. RNA folding systems are known to have a frequency-rank relationship estimated to be asymptotically $f(r) \sim \frac{\log r + a}{r}$, where r is the rank, and a is some constant [46]. Zipf laws have been observed for other input-output maps as well; in this case one would have $\log f(r) \propto \log r$. We have chosen an exponential distribution for our MCMC simulation in order to uniformly and maximally cover the log(frequency) axis in order to best understand the the robustness-frequency relationship.

The initialisation of the input-output map configuration itself for the MCMC simulation can be random (which would typically be in the disordered/random/fragile phase) or an ordered/ground state initialisation (which would consist—as explained in the following chapter—approximately of a set of bricklayer's graphs). For $k = 2$ with $d = 24$ and $k = 4$ with $d = 12$ we begin with an ordered simulation in which we simply assign output number 25 to the first 2^{23} sequences, output number 24 to the next 2^{22} sequences, output number 23 to the next 2^{21} sequences, and so on. The ordered start for these large systems is chosen for computational feasibility, as many temperatures can be simulated in parallel without risk of unwanted blockage in local energy[2] minimum.

We perform an additional simulation with a random/disordered initialisation for a smaller system ($k = 2$ with $d = 8$), also with an exponential frequency distribution:

$$f_0 = 1 \quad \text{and} \quad f_i = 2^{i-1}, i = \{1, \ldots, 8\}. \tag{3.36}$$

For this case, we perform simulations at each temperature in a sweep from high temperature to low temperature serially, using the final configuration Ω at a particular temperature as the starting configuration for the next (lower) temperature, in effect performing simulated annealing, which ensures the system does not get stuck in a local energy minimum.

3.5.2 Simulation Details

The theoretical details of the Metropolis-Hastings MCMC simulation we conduct are described in Appendix B. For our finite size simulations, we assume our system is ergodic and mixes sufficiently rapidly that we are able to calculate ensemble observables by time-averaging over those observables for configurations Ω_t (the configuration Ω at time step t) at regular intervals. For our ($k = 2$, $d = 24$) and ($k = 4$, $d = 12$) simulations which have ordered starting configurations, we simulate at temperatures $T^* = 0.01$ to $T^* = 15.01$ at intervals of $\delta T^* = 0.1$, where $T^* = T(k^d d(k-1)/2)$ is a scaled temperature. The system equilibrates at the simulation temperature very quickly, so we used a transient "burn-in" period of 50 Monte Carlo sweeps, each of which consists of $k^d = 2^{24} = 4^{12}$ individual time

[2] Note that energy here refers to the negative global robustness

steps/proposals. During the transient period, observables were not calculated at all. By monitoring system energy (negative global robustness), one could see that the system was equilibrated well before the 50 transient Monte Carlo sweeps. The main simulation then proceeded for 100 Monte Carlo sweeps, totaling $100 * 2^{24} = 100 * 4^{12} = 1,677,721,600$ proposals/time steps. For the ($k = 2$, $d = 8$) simulation with the random/disordered initialisation, we simulate from temperatures $T^* = 0.01$ to $T^* = 5.01$ at intervals of $\delta T^* = 0.5$, which we found to be sufficiently small spacing to ensure proper convergence to the robust phase during the annealing process. The number of burn-in/transient Monte Carlo sweeps was increased to 1000 (for a total number of proposals/time steps of $1000 * 2^8$), and the number of Monte Carlo sweeps in the main simulation during which observables were recorded was increased to 5000 (for a total number of proposals/time steps of $5000 * 2^8$).

After the transient period, observables were calculated at the end of each Monte Carlo sweep (i.e. every k^d proposals) and incorporated into the time-averaged estimates. We measured the system energy E (equivalent to the negative global robustness) as well as the scaled heat capacity C^*, given by

$$C^* = \frac{\partial E}{\partial T^*} = \frac{\langle E^2 \rangle - \langle E \rangle^2}{(T^*)^2}. \tag{3.37}$$

We additionally measured robustness of *each* output, the number of neutral components (i.e. connected components) within each output's induced subgraph/neutral network, the size of the largest neutral component for each output, the transition probability ϕ_{pn} between each pair of outputs. The transition probability ϕ_{pn} that a single-character mutation of a sequence mapping to a vertex in the n-th neutral will result in the output changing to p is mathematically defined as

$$\phi_{pn} = \frac{|E(G_n, G_p)|}{\ell(k-1)|V(G_n)|}, \quad n \neq p, \tag{3.38}$$

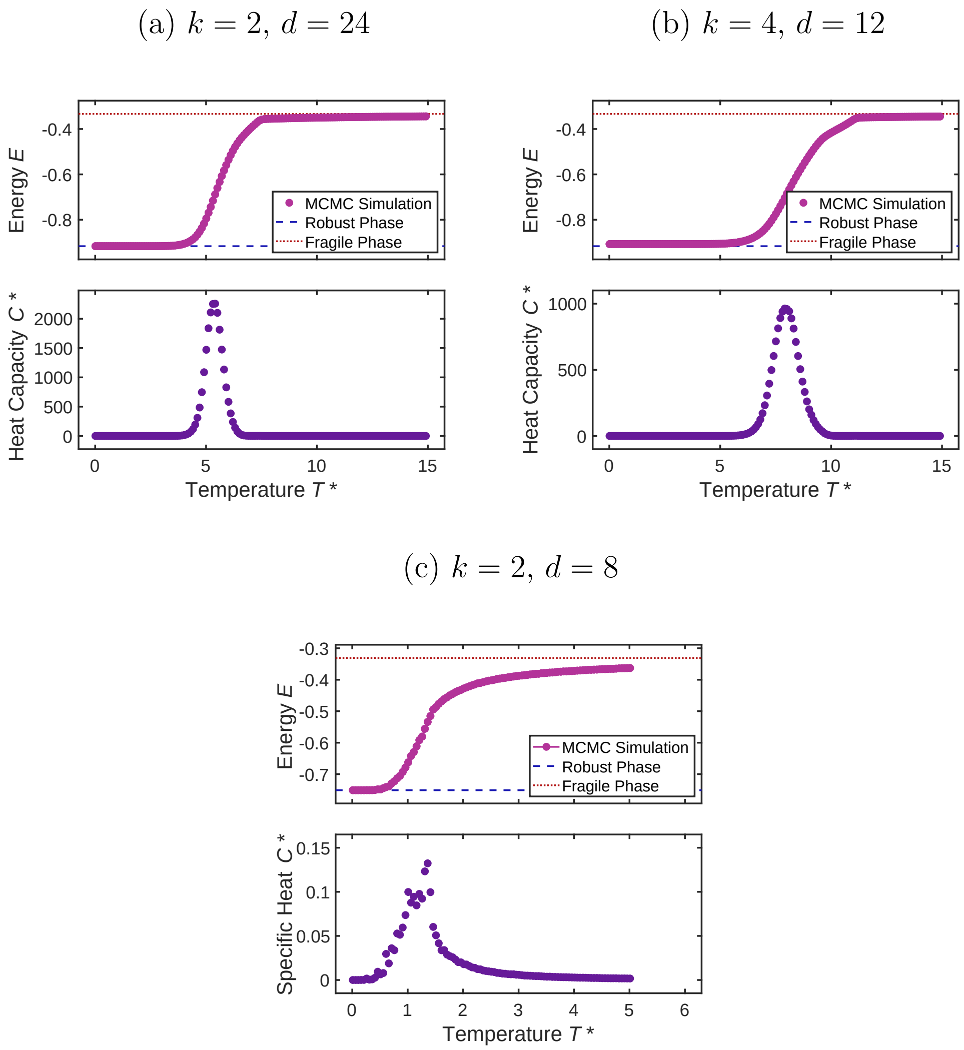
where $E(G_n, G_p)$ is the set of edges connecting the neutral networks/induced subgraphs of outputs n and p. These results are presented in the following section.

3.6 Numerical Results and Evidence of Phase Transition-Like behaviour

Here, we present results of the Markov Chain Monte Carlo (MCMC) simulation for the classical Potts model on a Hamming graph for fixed $\mathbf{f}$. Numerically, we are interested in the the behaviour of the ensemble-averaged energy (negative global robustness), heat capacity, robustness, transition probabilities, number of neutral components, and size of the largest neutral component. These results can then be compared to the biological and computer science GP map data in refs. [1, 15, 18–20].

3.6.1 Negative Global Robustness, Heat Capacity, and Indication of Phase Transition

In Figure 3.2, we show both the behaviour of the energy E (negative global robustness) and heat capacity C. For all three simulations, it is clear that there is phase transition-like behaviour between two distinct energy limits. In the energy plots, we see that in the low temperature regime the system's energy is minimal based on the frequency distribution $\mathbf{f}$ chosen; it does not leave the ordered state (for the ordered initialisation), or it finds an ordered (or at least very low energy) state during the annealing process for the disordered initialisation. Up to corrections to the log-scaling rule mentioned previously that will be addressed in the following chapter rigorously, the data suggest that the "robust phase" is indeed very close to the low temperature limit. Meanwhile, with increasing temperature, there is a steep transition to a higher energy phase that approaches the high temperature "fragile phase" limit we calculated in the high T limit previously. The peak in the heat capacity suggests that in the infinite limit there would be a divergence there, suggesting that a phase transition exists between the robust and fragile phases. These clearly appear to be the only two phases, for this system, and the scaling laws for robustness and other network topological properties that emerge from those two phases are the only ones we expect to see in nature.



(a) $k = 2$, $d = 24$

(b) $k = 4$, $d = 12$

(c) $k = 2$, $d = 8$

Figure 3.2: Energy E (equivalent to the negative global robustness) and (scaled) heat capacity C^* versus (scaled) temperature T^* for the simulated Potts model with fixed frequency vector $\mathbf{f}$, showing phase transition-like behaviour between the robust and fragile phases. Simulations **(a)** and **(b)** were initialised in the robust phase, with each temperature's MCMC carried out in parallel and simulation **(c)** was initialised with a random configuration and each temperature was simulated serially, starting with $T^* = 5.01$.

3.6.2 Robustness

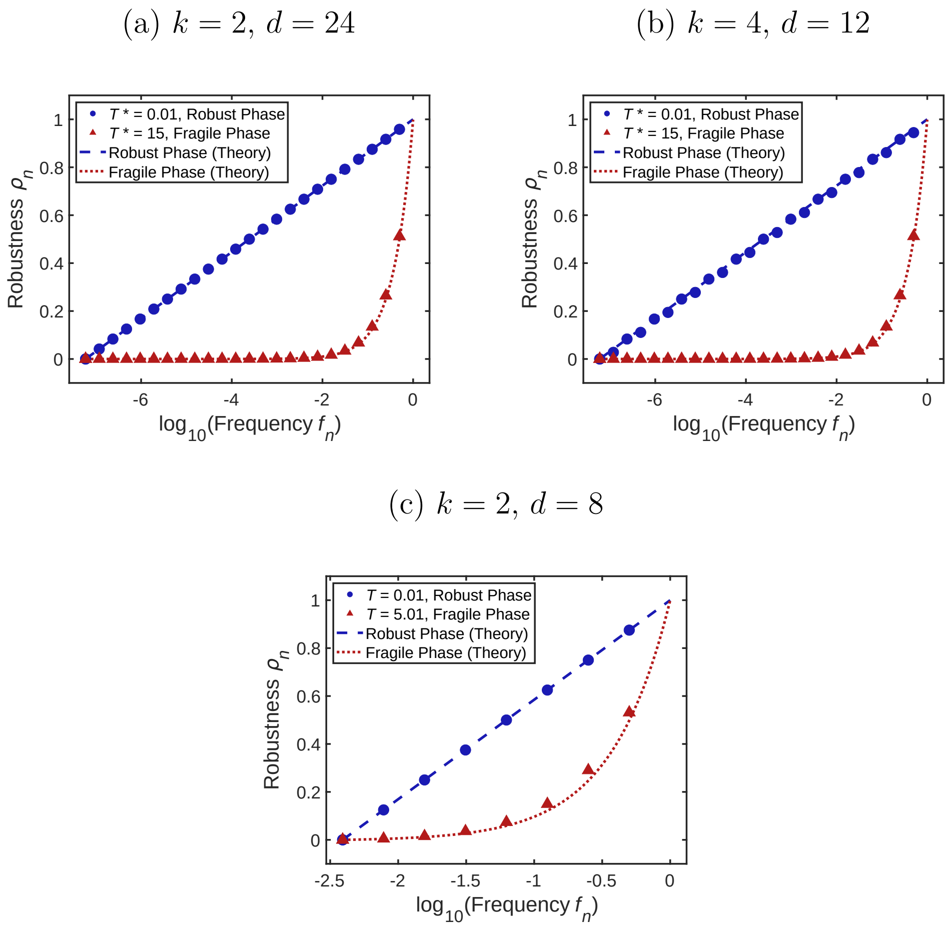
In Figure 3.3, we plot robustness ρ_n of each output versus the logarithm of the frequency f_n. For each simulation size (with ordered intialisation for the first two cases and a disordered initialisation for the third), we find very clearly that in the low temperature limit, the system is in the robust phase and agrees with the theoretical calculation of the robust phase robustness. For the high temperature phase, the system is in the fragile phase and agrees with our low temperature limit calculations. We note that for the $k = 4$, $d = 12$ simulations, alternating outputs (i.e. the odd-indexed outputs) stray a small amount from the logarithmic theory prediction. As mentioned before, this is because the exact maximum robustness line is not simply logarithmic; there is an additional correction factor for outputs whose frequencies are not powers of k. We have neglected that small correction factor for the theory line in this plot; it is discussed rigorously in the following chapter.

Due to the phase transition between the robust and fragile phases, we would expect that the only two scaling laws for robustness that should emerge are the logarithmic (robust) scaling law and the linear (fragile) scaling law. The former corresponds to the naturally observed scaling laws in the biological sequence-to-structure maps, the gene regulatory networks, the genetic algorithms, and the spin glasses, and latter corresponds to the random null model's scaling law discussed in [1]. Our results suggest that in the infinite sequence limit $d \to \infty$, the phase transition becomes exact, and there would be no other scaling laws observed in the model besides these two.

We emphasize once again that the robustnesses of individual outputs have not been constrained in this simulation; only the frequency-weighted average over the robustness—the global robustness—has been constrained. Yet all of the individual robustnesses collectively obey the same scaling law. This supports our hypothesis that only a single constraint on global robustness is sufficient to reproduce the naturally observed robustness behaviour of all the outputs/phenotypes.

3.6.3 Transition Probabilities

In Figure 3.4, we show a log-log plot of the transition probabilities ϕ_{pn} (for a mutation resulting in a change in output/phenotype from n to p) versus the frequency of the target output f_p. The empirical observations in ref. [1] show that $\phi_{pn} \propto f_p$ or even $\phi_{pn} \approx f_p$. Proportionality, but not approximate equality, would emerge if the


(a) $k = 2$, $d = 24$

(b) $k = 4$, $d = 12$

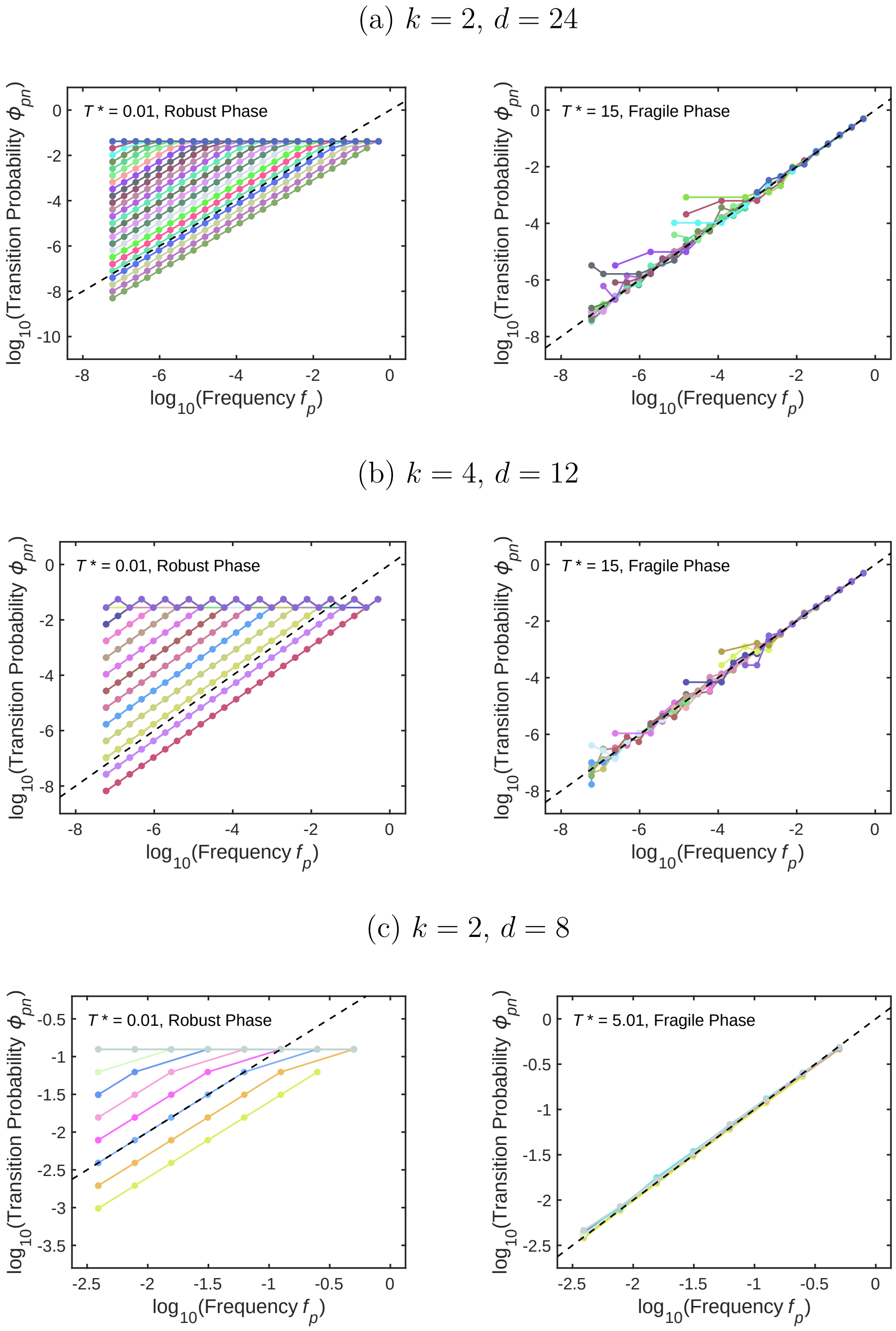
(c) $k = 2$, $d = 8$

Figure 3.3: Robustness ρ_n versus frequency f_n at low ($T^* = 0.01$) and high ($T^* = 15$) scaled temperatures for MCMC simulation. For each Hamming graph geometry, which is the same size as the **(a)** HP24 map, the **(b)** RNA12 map, and **(c)** a smaller map, there is clear evidence that two distinct robustness scaling laws exist on either side of the phase transition. The MCMC results coincide with the theory plots for the robust and fragile scaling laws. Simulations **(a)** and **(b)** were initialised in the robust phase, with each temperature's MCMC carried out in parallel and simulation **(c)** was initialised with a random configuration and each temperature was simulated serially, starting with $T^* = 5.01$.

robustness $\rho_n n$ is sufficiently large that transitions to other outputs that are not the n-th output would be different from the null expectation of $\phi_{pn} \propto f_p$.

We see that, for all of our simulations, in both the robust and fragile phases corresponding to low and high temperatures, respectively, the proportionality $\phi_{pn} \propto f_p$ is maintained. The dotted lines in Figure 3.4 are the line $\phi_{pn} = f_p$. In the high temperature limits (fragile phase), the $\phi_{pn} \approx f_p$ is observed. This is because in the fragile phase, the inputs are effectively randomly assigned to the outputs, as in the high temperature limit the constraint of the negative global robustness on the Shannon entropy becomes very weak. As a result, the neighbour of any given input/vertex is random with probability approximately equal to the frequency of that (neighbour's) output f_p. All of the MCMC ϕ_{pn} values thus converge onto the $\phi_{pn} = f_p$ line.

In the low temperature (robust) phase, the proportionality $\phi_{pn} \propto f_p$ is still maintained for many transitions, after which there is a plateau in the transition probability. The offsets from the dotted line and the plateau are due to two different phenomena: the offsets can be attributed to the fact that the mapping back onto the starting output $\phi_{nn} = \rho_n$ is much higher than would be expected from a random uncorrelated map. In the robust phase, when outputs tend to be clustered (maximally or nearly maximally) near each other due to high correlations, starting outputs with high frequency f_n are substantially more likely to map back onto themselves than any other output. This penalizes many of the ϕ_{pn} values, offsetting them below the diagonal line. Meanwhile, however, there are some outputs p which are highly abundant in the neighborhood of output n (again, beause of the clustering in the robust phase). These transitions are offset above the diagonal line. Lastly, the plateau in ϕ_{pn} occurs as frequency of the target output f_p increases because, for a starting output of frequency f_n, there are at most $k^d f_n$ edges connecting the induced subgraphs of the n-th and p-th outputs. So ϕ_{pn} is bounded above by the starting output's frequency f_n.



(a) $k = 2$, $d = 24$

(b) $k = 4$, $d = 12$

(c) $k = 2$, $d = 8$

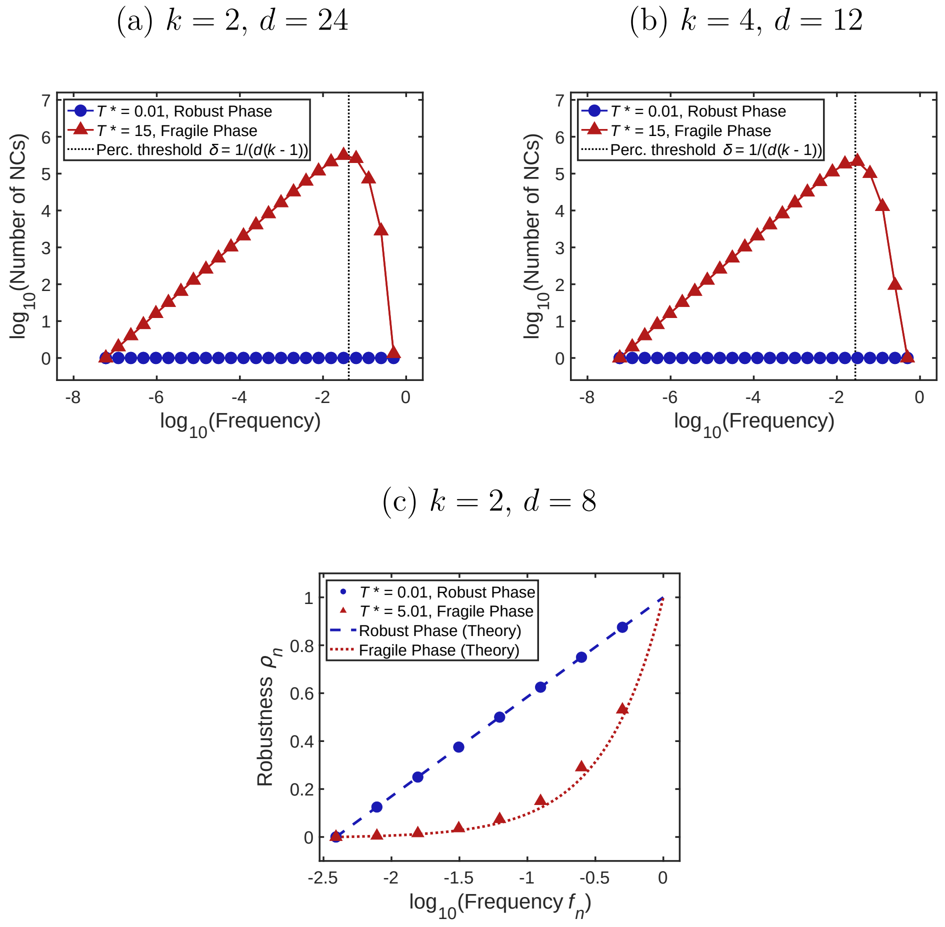
Figure 3.4: Log-log plot of the transition probabilities ϕ_{pn} (indicating a transition $n \to p$) versus frequency f_p of the target output at **(a,b,c left)** low ($T^* = 0.01$) and **(a,b,c right)** high ($T^* = 15$) scaled temperatures for MCMC simulation. Each colour represents a starting output n, and each dot on each line is placed according to the frequency f_p of the target output p. The dotted line indicates the expectation $\phi_{pn} \approx f_p$ that has been observed empirically in [1]. In both the low and high temperature limits, the scaling law holds. Simulations **(a)** and **(b)** were initialised in the robust phase, with each temperature's MCMC carried out in parallel and simulation **(c)** was initialised with a random configuration and each temperature was simulated serially, starting with $T^* = 5.01$.

These graphs show excellent phenomenological agreement with the empirical data from biological GP maps [1] (reproduced in Figure 1.4) as well as the spin glass maps from Chapter 2.

3.6.4 Number of Neutral Components and Size of Largest Neutral Component

Each output's induced subgraph/neutral network may be disconnected into several neutral components (called connected components in graph theory) or may be fully connected. In Figure 3.5, we plot the number of neutral components of each induced subgraph for each output versus the frequency of that output. In the robust phase at low temperature, we see very clearly that the number of neutral components remains at 1; this is because the entire output network tends to cluster tightly in the robust phase as the constraint on maximum entropy is strong. In the fragile, high temperature phase, we find that for the smallest frequencies the number of neutral components is small, and it grows as frequency increase until a percolation threshold [1] δ is reached. Since each vertex/input in the Hamming graph has $d(k-1)$ neighbours, when the frequency of a particular output is approximately $f_n \approx \delta \equiv 1/(d(k-1))$, the expected number of neighbours mapping to the same (n-th) output becomes approximately 1. Beyond this threshold, the probability of finding at least one neighbour mapping to the same output becomes very close to 1, which means that the entire induced subgraph/neutral network for that output is very likely to be (almost) fully connected. This percolation threshold, known as the giant component threshold, therefore shows a rapid drop in the number of neutral components for sufficiently large frequency.

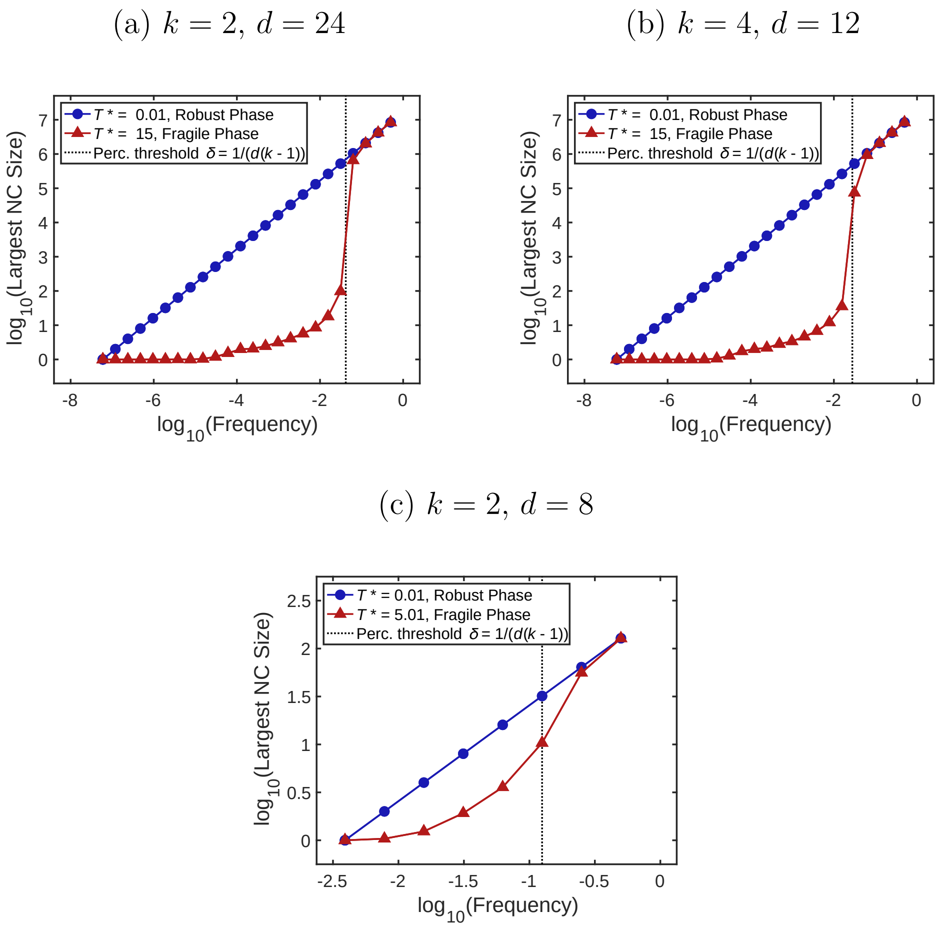
The size of the largest component is plotted in Figure 3.6 versus the frequency of the output. In the robust phase, since there is only one component per output, the number of vertices in the largest component is simply the number of vertices in that induced subgraph. In the fragile phase, the size of the largest component increases, but very slowly, as the probability of finding small connected clusters increases with frequency. However, the clusters are still likely to be small and spread out until the


(a) $k = 2$, $d = 24$

(b) $k = 4$, $d = 12$

(c) $k = 2$, $d = 8$

Figure 3.5: Log-log plot of the number of neutral components versus frequency f_p of each output in both the robust (low temperature) and fragile (high temperature) phases. These results agree with the data for biological GP maps in [1]. A percolation transition occurs [1] at the frequency $\delta = 1/(d(k-1))$, where the largest outputs' induced subgraphs/neutral networks are sufficiently large that they connect into one giant component; the number of neutral components drops to a small number (essentially 1) in the giant component regime.


(a) $k = 2$, $d = 24$

(b) $k = 4$, $d = 12$

(c) $k = 2$, $d = 8$

Figure 3.6: Log-log plot of the largest neutral component size versus frequency f_p of each output in both the robust (low temperature) and fragile (high temperature) phases. These results agree with the data for biological GP maps in [1]. A percolation transition occurs [1] at the frequency $\delta = 1/(d(k-1))$, where the largest outputs' induced subgraphs/neutral networks are sufficiently large that they connect into one giant component.

giant component percolation threshold is reached, beyond which a giant component appears. Now, for sufficiently large frequencies, the largest component is also a giant component, or the entire induced subgraph may even be fully connected.

The neutral component size and number data has excellent agreement with the biological GP map data published in ref. [1]; this model is more idealised and less, noisy of course.

3.7 Global Robustness of Analytical Biological GP Maps

In the previous sections we have shown that a simple constraint on the global (or average) robustness is enough to push a mapping from a fragile phase to a robust phase that is near to maximum robustness. The open question, of course, is do real GP maps have such a constraint? One way of answering this question is to look at some simplified GP maps, and to see if they naturally result in a higher robustness. In this section we consider a recent body of work that has done just that, starting with a paper by Greenbury and Ahnert [13], who looked at a simple picture of constrained and unconstrained parts of a genotype. This work was then followed by additional papers [16, 46] that built in more sophisticated versions of the same constrained/unconstrained idea.

Since these models all show that a very simple biological fact, having constrained and unconstrained parts of a sequence, leads to relatively high robustness, this suggests that natural GP maps will be in the high robustness phase. We now look at two models from this literature, examined analytically by Weiß and Ahnert [16], in more detail.

3.7.1 Maximally Robust behaviour of Weiß-Ahnert Gene-Like Model

Weiß and Ahnert [16] generalise Greenbury and Ahnert's previous Fibonacci model of genotype-phenotype maps [13] (which modeled genes containing variable-length sequences of coding and non-coding DNA by using binary sequences with effective "stop codons" built in) to an input-output map that admits an alphabet of more than 2 characters. This "gene-like model" has one stop codon among the alphabet of k characters. Their input-output map takes in a sequence of d characters from the alphabet, and the output is defined uniquely by the identity of the sequence prior to the stop codon, which by the very nature of the system is mutable. This model closely mimics the idea of coding and non-coding regions in DNA or RNA regions. They also compare this model to a gene-like reference model in which the first stop codon is fixed. The robustness of the gene-like model and the gene-like reference model are exactly solved.

First, we consider the gene-like model with input output map Ω_g. Suppose that a particular output is generated from a (contiguous) coding sequence of length ℓ. Each output that has a coding sequence of length ℓ will have a frequency

$$f(\ell) = k^{-\ell}, \tag{3.39}$$

and the number of outputs which have coding sequences of length ℓ have multiplicity

$$\mu(\ell) = (k-1)^{\ell-1}. \tag{3.40}$$

When $\ell = 0$, that means there is no stop codon present. The authors of [16] assign a label of "undefined" to this output. The undefined output has a frequency $f_u = \left(\frac{k-1}{k}\right)^d$. It can be verified that

$$f_u + \sum_{\ell=1}^{d} f(\ell)\mu(\ell) = 1. \tag{3.41}$$

The robustness of an output with a coding sequence of length ℓ is given by

$$\rho(\ell) = 1 - \frac{\ell}{d}. \tag{3.42}$$

The undefined output has a robustness $\rho_u = \frac{k-2}{k-1}$. We can now write the negative global robustness $\mathcal{H}(\Omega_g)$ as

$$\begin{aligned}\mathcal{H}(\Omega_g) &= -\left(f_u\rho_u + \sum_{\ell=1}^{d}\mu(\ell)f(\ell)\rho(\ell)\right)\\ &= -\left(1 - \frac{k}{d} + \frac{(k-1)^d}{k^{d-1}d} + \frac{(k-2)(k-1)^{d-1}}{k^d}\right).\end{aligned} \tag{3.43}$$

Given the frequency distribution of this model, the lower bound becomes

$$\begin{aligned}\min_{\Omega_g\,|\,\mathbf{f}} \mathcal{H}(\Omega_g) = \lim_{T\to 0}\langle\mathcal{H}(\Omega)\rangle &= -\left(1 + \frac{f_u \log f_u}{d\log k} + \sum_{\ell=1}^{d}\frac{\mu(\ell)f(\ell)\log f(\ell)}{d\log k}\right)\\ &= -\left(1 - \frac{k}{d} + \frac{(k-1)^d}{k^{d-1}d} + \frac{(k-1)^d\log(k-1)}{k^d\log k}\right),\end{aligned} \tag{3.44}$$

and the upper bound on $\langle \mathcal{H}(\Omega_g) \rangle$—i.e. the negative global robustness in the unconstrained maximum entropy limit—is

$$\begin{aligned}\lim_{T\to\infty} \langle \mathcal{H}(\Omega) \rangle &= -\frac{1}{k^d-1}\left(k^d f_u^2 + \sum_{\ell=1}^{d} \mu(\ell) f(\ell)^2 - 1\right) \\ &= -\frac{1}{k^d-1}\left(\frac{(k-1)^{2d}}{k^d} + \frac{k^d-(k-1)^d k^{-d}}{k^2-k+1} - 1\right).\end{aligned} \tag{3.45}$$

It is clear that in the limit of large sequence length d, we see that

$$\lim_{T\to 0} \langle \mathcal{H}(\Omega) \rangle \xrightarrow{d\gg 1} -1\left(1-\frac{k}{d}\right) \xrightarrow{d\to\infty} -1 \tag{3.46}$$

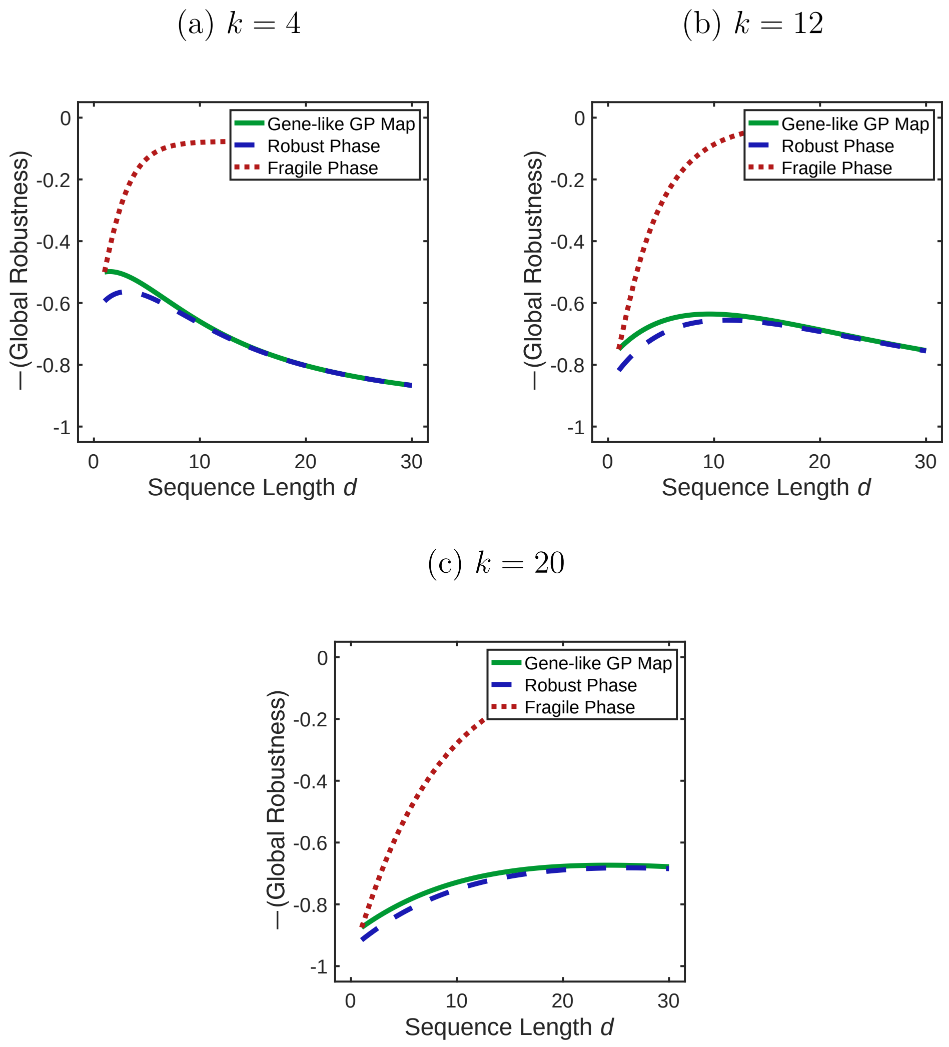
and

$$\lim_{d\to\infty} \mathcal{H}(\Omega_g) \xrightarrow{d\gg 1} -1\left(1-\frac{k}{d}\right) \xrightarrow{d\to\infty} -1. \tag{3.47}$$

Thus, the gene-like model converges to the highest possible edge count (i.e. lowest possible negative global robustness) in the large sequence length limit. In Figure 3.7, we see plots of the lower and upper bounds on $\langle \mathcal{H}(\Omega) \rangle$ and the negative global robustness of the gene-like model $\mathcal{H}(\Omega_g)$.

3.7.2 Maximally Robust behaviour of Modified Weiß-Ahnert Gene-like Model with Fixed Number of Constrained Sites

The authors of ref. [16] also propose a gene-like "reference" model in which the stop codon is not mutable. This means that if the i-th element in the input sequence is (the first) "stop" codon (indicating that the constrained part of the input sequence ends there), it cannot be changed to any other character in the alphabet. That means that for a sequence of length d, there are $d-1$ mutable sites. In their model, they consider cases where the stop codon could be at any of the d sites and also calculate the frequency f_n and robustness ρ_n of outputs for all possible positions of the stop codon. They also calculate the robustness and frequency of the undefined phenotype, when no stop codon is present. In their model, the restriction on not being able to mutate the stop codon results in edges being pruned from the Hamming graph—thus, our analytical calculations on the Hamming graph will not be exactly correct (though the behaviour will be the same in the large sequence length d limit).


(a) $k = 4$

(b) $k = 12$

(c) $k = 20$

Figure 3.7: Global robustness from the gene-like genotype-phenotype map from Weiß and Ahnert [16] as a function of sequence length. (Red) Lower bound (low temperature limit) $\min_{\Omega\,|\,\mathbf{f}} \mathcal{H}(\Omega_g) = \lim_{T\to\infty} \langle\mathcal{H}(\Omega)\rangle$ from eq. (3.44), (Blue) upper bound (high temperature limit) $\lim_{T\to\infty} \langle\mathcal{H}(\Omega)\rangle$ from eq. (3.45), and (Green) negative global robustness $\mathcal{H}(\Omega_g)$ from eq. (3.43) for the gene-like model of genetic input-output maps for (a) $k = 4$, (b) $k = 12$, and (c) $k = 20$.

We present a slightly modified version of the gene-like reference model: here we have a fixed number of coding (i.e. constrained) sites $c = \{0, \ldots, d\}$ out of a total number of sites d. Moreover, indices of those c constrained sites are specified *a priori* (e.g. the first c sites in the input sequence would be a realistic exon-intron pair). The number of constrained sites c is an additional degree of freedom in the model. The frequency of an output $f(c)$ that has c constrained sites is thus given by the number of combinations of characters which can be generated from only the unconstrained sites

$$f(c) = k^{-c}, \tag{3.48}$$

and the robustness of such an output is the number of unconstrained sites divided by the total number of sites:

$$\rho(c) = 1 - \frac{c}{d}. \tag{3.49}$$

The number of outputs which have c constrained sites is given by the multiplicity

$$\mu(c) = k^c. \tag{3.50}$$

Clearly, we have $\mu(c)f(c) = 1$. Using the above relations it is clear that

$$\rho(c) = 1 + \frac{\log f(c)}{d \log k}, \tag{3.51}$$

regardless of the number of constrained sites, so a model that is purely constrained or unconstrained displays the expected linear-log scaling relationship between robustness and frequency. Referring to this input-output map as $\Omega_g(c)$, where c specifies the number of constrained sites, we can write the negative global robustness

$$\mathcal{H}(\Omega_g(c)) = -\mu(c)f(c)\rho(c) = -\left(1 - \frac{c}{d}\right). \tag{3.52}$$

The lower bound on the negative global robustness is given by

$$\begin{aligned} \min_{\Omega_g(c)\,|\,\mathbf{f}} \mathcal{H}(\Omega_g(c)) &= \lim_{T\to 0} \langle \mathcal{H}(\Omega) \rangle \\ &= -\left(1 + \frac{\mu(c)f(c)\log f(c)}{d \log k}\right) \\ &= -\left(1 - \frac{c}{d}\right). \end{aligned} \tag{3.53}$$

Clearly, the total edge fraction is *always* at its maximum value in this model, and is thus the “coldest” it could possibly be. For the sake of completeness, we show that the upper bound—i.e. the negative global robustness in the unconstrained maximum entropy limit—is given by

$$\begin{aligned} \lim_{T\to\infty} \langle \mathcal{H}(\Omega) \rangle &= -\frac{1}{k^d - 1}\left(\mu(c)f(c)^2 - 1\right) \\ &= -\frac{k^{d-c} - 1}{k^d - 1}. \end{aligned} \tag{3.54}$$

3.7.3 Highly Robust behaviour of Weiß-Ahnert RNA-Like Model

The authors of ref. [16] have also developed an analytical model that behaves like secondary structure maps–in particular, the RNA secondary structure map in which oligouncleotide sequences can fold to form stem-loop structures. In their model, one character of the k characters in the alphabet represents a character which can induce binding. The furthest separated binding characters are linked in a matter that would resemble base-pairing in a stem-loop structure. A sequence with an even number of such binding characters would have pairs forming from the outside-in. A sequence with an odd number of such binding characters would have one binding character in the central loop which remains unbound. The input-output map, which we call Ω_r is decided based on mutations of the paired/bound versus unpaired/unbound characters.

For an input sequence of length d, suppose there are i characters which are of the binding character. If i is even, then robustness also depends on the number j of characters intervening between the two innermost characters. The authors have shown that, as a function of i and j, the frequency of outputs is

$$f(i,j) = \frac{(k-1)^{d-i} + j(k-1)^{d-i-1}}{k^d}, \tag{3.55}$$

and while the multiplicity of each output cannot be analytically derived, the multiplicity of the number of outputs given i and j can. This is given by

$$\widetilde{\mu}(i,j) = \binom{d-1-j}{i-1}. \tag{3.56}$$

The undefined output has only one or zero coding letters, and it has frequency

$$f_u = \frac{(k-1)^d + d(k-1)^{d-1}}{k^d}. \tag{3.57}$$

The robustness values of these outputs has also been derived analytically, with the robustness as a function of i and j given by

$$\rho(i,j) = \frac{k-2}{k-2} + \frac{i(k-2)}{d(k-1)} + \frac{jk}{d(k-1)(k-1+j)}, \tag{3.58}$$

and the robustness of the undefined output is

$$\rho_u = \frac{k - 1 + (d(k-2)+1)/(k-1)}{(k-1+d)}. \tag{3.59}$$

There is no closed form expression for the total edge fraction, but it is given by

$$\mathcal{H}(\Omega_r) = -\left(f_u \rho_u + \sum_{i \in I_d} \sum_{j=0}^{d-1} \widetilde{\mu}(i,j) f(i,j) \rho(i,j), \right). \tag{3.60}$$

where $I_d = \{2, 4, \ldots, 2\lfloor d/2 \rfloor\}$. The lower bound is, as usual, given by

$$\begin{aligned} \min_{\Omega_r \,|\, \mathbf{f}} \mathcal{H}(\Omega_r) &= \lim_{T \to 0} \langle \mathcal{H}(\Omega) \rangle \\ &= -\left[1 + f_u \log f_u + \sum_{i \in I_d} \sum_{j=0}^{d-1} \widetilde{\mu}(i,j) f(i,j) \log f(i,j) \right], \end{aligned} \tag{3.61}$$

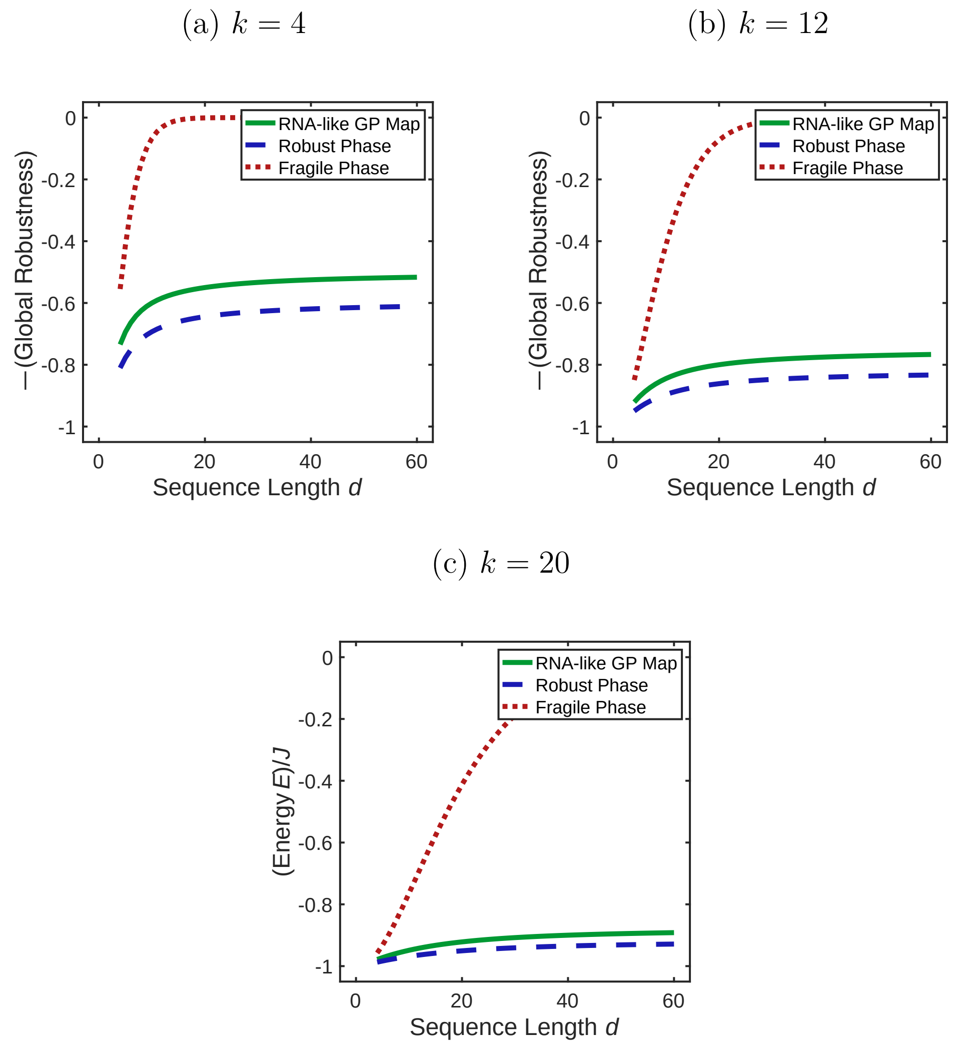
and the upper bound is

$$\lim_{T \to \infty} \langle \mathcal{H}(\Omega) \rangle = -\frac{1}{k^d - 1} \left(\sum_{i \in I_d} \sum_{j=0}^{d-1} \widetilde{\mu}(i,j) f(i,j)^2 - 1 \right). \tag{3.62}$$

For various values of k, we plot in Figure 3.8 the total edge fraction for this model, the lower bound, and the upper (random) bound. It is clear that, as the sequence length increases, the total edge fraction remains near—though does not seem to asymptotically approach as in the gene-like model—the lower bound on the negative global robustness. These systems are not globally maximally robust, but their features [16] resemble those of the robust phase.

3.8 Discussion

In this section, we proposed a statistical physics approach to understanding input-output maps. We hypothesised that constraining a single parameter, the (negative) global robustness of an input-output map would be enough to reproduce many of the network topological properties of natural input-output maps which have been observed in RNA genotype-phenotype maps, protein folding genotype-phenotype maps, and more. We showed that the negative global robustness $\mathcal{H}(\Omega)$ of an input-output map Ω can be written in the form of the classical Potts model, and



(a) $k = 4$

(b) $k = 12$

(c) $k = 20$

Figure 3.8: Global robustness from the RNA-like genotype-phenotype map from Weiß and Ahnert [16] as a function of sequence length. (Red) Lower bound (low temperature limit) $\min_{\Omega\,|\,\mathbf{f}} \mathcal{H}(\Omega_r) = \lim_{T\to\infty} \langle \mathcal{H}(\Omega)\rangle$ from eq. (3.53), (Blue) upper bound (high temperature limit) $\lim_{T\to\infty} \langle \mathcal{H}(\Omega)\rangle$ from eq. (3.54), and (Green) energy $\mathcal{H}(\Omega_r)$ from eq. (3.52) for the gene-like model of genetic input-output maps for (a) $k = 4$, (b) $k = 12$, and (c) $k = 20$.

therefore many of the methods from statistical physics can be used to analyze the distribution over the space of input-output maps.

By working with a maximum entropy model, we showed than in the unconstrained limit, the input-output map is the same as the random null model, and its robustnesses are linear in the frequencies of the outputs. This corresponded to the high temperature limit in the statistical physics analogy. In the limit in which the negative global robustness was strongly constrained (the low temperature limit), we saw that a logarithmic robustness scaling law was recovered. Moreover, through MCMC simulation, we also found that transition probabilities between phenotypes, neutral component sizes, and number of neutral components all obeyed the same scaling laws that had been seen in natural genotype-phenotype maps. We concluded the chapter by studying analytical models of genotype-phenotype maps and calculated their negative global robustnesses in the limit of infinite sequence length. We saw that in the gene-like models, these negative global robustnesses converged to the optimum in the infinite sequence limit. For the RNA-like model, the negative global robustness did not converge to the minimum value but remained very close. The value of examining these simple analytical models is that, despite only using some fairly basic assumptions of how sequences work in practice for genotype-phenotype maps, one ends up with relatively high global robustness and high individual robustnesses for individual phenotypes. Therefore, from this knowledge and from our maximum entropy calculations, we should *expect* natural genotype-phenotypes to exhibit the robust phase. Moving forward, the next step would be to (a) see if we can better elucidate the phase transition behaviour analytically and properly define a thermodynamic limit and (b) investigate at much larger biological scales with much more complex GP maps beyond the simple models which have been studied in the literature and here.

4
Maximally Robust Neutral Networks and Coarse-Grained Phenotypes

Contents

4.1 Introduction

Even though there have been decades of research in the biological community into mutational robustness, an exact upper bound on the robustness for a neutral network of fixed size is not widely known. 1960s work in the field of coding theory (error-correcting codes, code transmission in noisy channels, etc.) proved that, among subgraphs of Hamming graphs, a certain class of graphs maximise, for a fixed number of vertices, the *edge-to-vertex ratio* $|E(G)|/|V(G)|$ of the graph G. As we saw in previous chapters, the edge-to-vertex ratio of a graph is equivalent (up to a scaling factor) to what biologists call mutational robustness. Here, we elucidate the connections between those graph theoretic results and biological robustness by studying maximally robust neutral networks, called *bricklayer's graphs*.

The term "bricklayer's graph" was coined by Reeves et al. [61], who describe them as "the graphs that interpolate pointwise between hypercube graphs of consecutive dimension (the point, line, line and point in the square, square, square and point in the cube, and so on)." In this manner, adding an $(n+1)$-th vertex to a bricklayer's graph which already has n vertices is performed in a manner akin to stacking bricks serially in a (hyper)cube. The coding theorists in the 1960s did not actually use the term "bricklayer's graphs," but Harper [62] and Lindsey [63] worked with the graph-theoretic representation of sequence spaces and their subsets, just as we have done with biological genotype-phenotype maps. Harper and Lindsey discovered (for hypercubes and all Hamming graphs, respectively) that these bricklayer's graphs are optimal in the sense that, given a set of sequences (whose vertex representations are) in a bricklayer's graph, the average single-character error tolerance of the sequences is maximised. This is essentially exactly the same as maximising mutational robustness of a phenotype in a genotype-phenotype mapping, but thus far the mathematical relationships between coding theory results from the 1960s and recent work on mutational robustness have not been established. We make these relationships explicit and dive further into the mathematical descriptions of the bricklayer's graphs.

In particular, we showcase a surprising connection between biological sequence robustness in genotype-phenotype maps and number theory by discussing how the robustness of bricklayer's graphs is related to the sums-of-digits function, which is notably continuous everywhere but differentiable nowhere. The known results from number theory show that the logarithmic scaling law observed in the natural systems discussed in previous chapters is only an asymptotic approximation to the true maximal robustness. Real RNA folding simulations have shown that many individual connected components within neutral networks of folded phenotypes can attain the bricklayer's graph bound exactly, but their phenotypes are comparatively not as robust. We use a property of the sums-of-digits functions that we prove to analytically calculate a lower bound on this deviation from the ideal neutral network robustness. We then show that real biophysical genotype-phenotype maps have phenotypes which are close to this bound, which is what we expected to see from Chapter 3. Many of the phenotypes' neutral networks which deviate from the bound we prove are composed of multiple neutral components. Using a newly derived property of the sums-of-digits function that we prove, we are able to bound the deviation of the robustness of a phenotype comprised of multiple neutral components from its optimal value calculated from the bricklayer's graphs.

After considering the robustness of phenotypes which consist of multiple neutral components, we then consider the coarse-graining of phenotypes. We show how robustness and transition probabilities change when multiple phenotypes are combined together into abstract phenotypes. Our formulas help explain unexpected behaviour observed in recent numerical simulations [64] in RNA phenotype coarse-graining.

We also discuss the maximisation of other topological properties of neutral networks, including base information content and population neutrality. The base information content in a phenotype is defined as the difference between the maximum possible and actual Shannon entropy of a set of sequences which map to a phenotype. The lower the information content, the less informative each position in the sequence is in determining a unique phenotype. It had been suggested by Greenbury [3] that certain graphs simultaneously optimise mutational robustness and base information content. We show, however, that *in general* these two parameters are not simultaneously optimised, and we use the sums-of-digits function to discover a special set of bricklayer's graphs in which simultaneous optimisation does occur.

Robustness and population neutrality are both incredibly important to evolutionary adaptive dynamics. The enhanced, logarithmically scaling robustness allows an evolving population to traverse a substantial portion of the Hamming graph, staying within the robust subnetwork of the current phenotype, without incurring any fitness cost. This aids the discovery of many more routes of escape to other phenotypes of higher fitness than would be possible from a random mapping of genotypes to phenotypes. The population neutrality quantifies a higher-order anisotropy in the equilibrium population distribution at infinite time which, albeit, may not ever be reached in natural systems. In fact, this same population neutrality quantifies the deviation of Haldane's *genetic load* [65] from it's typical value; the deviation results from the fact that the highest fitness phenotype may be coded for by multiple genotypes—i.e. there is neutrality present in the GP map. It has been shown by Van Nimwegen et al. [66] that the population neutrality, despite its importance in the asymptotic evolutionary dynamics of a population, actually only relies on the topological properties of the most-fit neutral network. In particular, the population neutrality is equivalent, up to a scale factor, to the principal eigenvalue of the adjacency matrix of the neutral network. With our collaborators, we prove lower and upper bounds on the population neutrality of bricklayer's graphs; the latter resolves a conjecture by Reeves et al. [61].

It has recently been shown [67] that Hamming spheres, a particular class of graphs, are the subgraphs of Hamming graphs which maximise the population neutrality asymptotically (for genotype-phenotype maps with alphabets with $k = 2$ characters). The Hamming sphere is important in coding theory: when an encrypted message of length ℓ is passed through a channel that may potentially induce errors, a buffer radius is built in such that the message may be decrypted even if up to $\eta \leq \ell$

errors appear in the message. The “sphere-packing bound” or “Hamming bound” gives the “volume” (order, or number of vertices in the graph) of the Hamming sphere which buffers a particular codeword. Given a set of codewords of length ℓ in a k-ary alphabet, the Hamming bound indicates how large the Hamming spheres should be so that every codeword can be recovered successfully from an encrypted message even if errors have occurred. In this chapter, we exactly calculate the robustness of Hamming spheres and show that it is below the bricklayer’s graph maximum. As a result, for $k = 2$, there is mutual exclusivity in the maximisation of robustness and population neutrality.

Having studied the robustness, information, and neutrality properties of individual phenotypes’ neutral networks, and coarse-graining of neutral components and of phenotypes, this chapter presents a new graph-theoretic look at the topological properties of biologically relevant netural networks.

4.2 Bricklayer’s Graphs: Maximally Robust Neutral Networks

First, we define the notion of a bricklayer’s graph. Originally arising in the context of optimal codes in coding theory, it was shown [62, 63] that if the vertices of a bricklayer’s graph (that is an induced subgraph of a Hamming graph) represent the set of k-ary sequences which map to a particular codeword A being transmitted, that set of sequences minimizes the number of single-site mutations which would cause an incorrect transmission of codeword A, averaged over all of the sequences mapping to codeword A. This is akin to maximizing mutational robustness in biology, which measures the number of point mutations which do *not* change the phenotype, averaged over all sequences mapping to that phenotype. The term “bricklayer’s graph” was coined by Reeves et al. [61] because these graphs are constructed by repeatedly adding an adjacent vertex in the Hamming graph, resembling the process of laying bricks; we continue to use it throughout this thesis.

Definition 4.2.1. *A* ***bricklayer's graph*** $G_{n,k}$ *is an induced subgraph of a Hamming graph* $H_{\ell,k}$ *containing* $|V(G_{n,k})| = n$ *vertices* $\{0, 1, \ldots, n-1\}$ *such that* $(i, j) \in E(G_{n,k})$ *if the base-*k *representations of* i *and* j *differ in exactly one digit.*

As we described in the previous chapters, in evolutionary biology one is often concerned with calculating the robustness of a neutral network (induced subgraph of a Hamming graph) because of its important implications for dynamic discovery of phenotypes. We now calculate the exact robustness of bricklayer's graphs and draw a connection to a function with a rich history in number theory.

4.2.1 Exact Robustness/Number of Edges

Theorem 4.2.1. *A bricklayer's graph* $G_{n,k}(V, E)$ *with* n *vertices has* $|E| = S_k(n) = \sum_{i=0}^{n-1} s_k(i)$ *edges, where* $s_k(i)$ *is the sum of all digits in the base-*k *representation of the integer* i*. We will call* $S_k(n)$ *the sums-of-digits function.*

Proof. To see this, let ℓ be the length of the input sequence so $\ell \geq \log_k n$, and let $(x_{\ell-1}(n), \ldots, x_0(n))$ be the vector of integers containing the digits of the base-k representation of the integer n such that $n = \sum_{i=0}^{\ell-1} x_i(n)k^i$. Consider the bricklayer graph $G_{n-1,k}$. When we add one more vertex such that $G_{n-1,k} \mapsto G_{n,k}$, we look at the base-$k$ representation of n. An edge can be added if the base-k representation of n differs from the base-k representation of the neighbouring vertex by exactly one digit. Going through digit by digit, we see that the only allowed flips for the i-th digits are $x_i(n) \mapsto \{0, \ldots, x_i(n) - 1\}$. This set has cardinality $x_i(n)$. Summing this over all digits, we find that the number of edges added to the graph $G_{n-1,k}$ when adding an additional vertex n is the sum of digits of n in the base-k representation $s_k(n)$. Therefore, the total number of edges in $G_{n,k}$ is $S_k(n) = \sum_{i=0}^{n-1} s_k(i)$.[1] □

The asymptotic, logarithmic behaviour of the sums-of-digits function $S_k(n)$ was first given by Bush [68], and an exact analytical form for $k = 2$ was given by Trollope

[1] This theorem was proven by the author independently, without knowledge of Sam Greenbury's derivation in his PhD thesis. The author and Greenbury are, at the time of writing of this thesis, collaborating on a manuscript to publish this result along with the others found in this chapter.

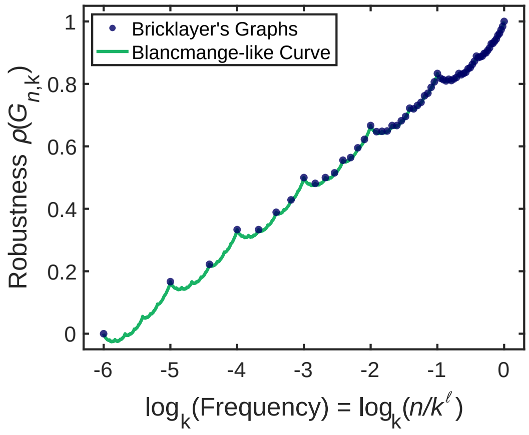


Figure 4.1: The linear-log plot of the bricklayer graphs' robustness $\rho_n(G_{n,k}) = 2S_k(n)/(n\ell(k-1))$ versus frequency (number of vertices n divided by k^ℓ), where $\ell = 6$ and $k = 2$. The "blancmange-like" curve is plotted as well; this is the continuous-everywhere, differentiable-nowhere function that corresponds to the continuous limit of $S_k(n)$.

[69] and later generalised by Delange for all k [70]. The function can be written as:

$$S_k(n) = \frac{n}{2}\left[(k-1)\log_k n - g_k\left(k^{\{\log_k n\}-1}\right)\right], \tag{4.1}$$

where $\{x\}$ is the fractional part of x, and

$$g_k(x) = (k-1)\log_k x + \frac{D_k(x)}{x}, \tag{4.2}$$

where $D_k(x)$ is the *Delange function* (using the modified definition in ref. [71]) given by

$$D_k(x) = \sum_{n=0}^{\infty} \frac{D_{k,0}(k^n x)}{k^n}, \quad D_{k,0}(x) = \int_0^x \mathrm{d}t\,(2k[t] - 2[kt] + k - 1), \tag{4.3}$$

where $[x]$ is the integer part of x. For $k = 2$, the Delange function $D_k(x)$ is the same as the continuous everywhere, differentiable nowhere Takagi function first described in 1903 [72]. The sums-of-digits function has interesting connections to number theory, namely the Riemann zeta function. Delange [70] showed that the Fourier series coefficients c_n of $g_k\left(k^{\{x\}-1}\right)$ (which is periodic in x with a period of one), defined by

$$g_k\left(k^{\{x\}-1}\right) = \sum_{n\in\mathbb{Z}} c_n e^{\imath 2\pi n x} \tag{4.4}$$

are

$$c_n = \int_0^1 \mathrm{d}x\, e^{\imath 2\pi n x} g_k\left(k^{\{x\}-1}\right) = \imath \frac{k-1}{n\pi}\left(1+\frac{\imath 2n\pi}{\log k}\right)^{-1} \zeta\left(\frac{\imath 2n\pi}{\log k}\right), \tag{4.5}$$

where ζ is the Riemann zeta function, and $\imath$ is the unit imaginary number.

4.2.2 Bounds on Bricklayer's Graph Robustness

From the above definitions, we may clearly see that the biological robustness is given by

$$\rho(G_{n,k}) = \frac{2|E(G_{n,k})|}{\ell(k-1)|V(G_{n,k})|} = \frac{\log_k n}{\ell} - \frac{g_k\left(k^{\{\log_k n\}-1}\right)}{\ell(k-1)}, \tag{4.6}$$

which we already see reproduces the conventional scaling law of $\rho = \ell^{-1}\log_k n$, *plus* an additional term. This robustness $\rho(G_{n,k})$ is plotted in Figure 4.1.

Galkin and Galkina [71] have shown that for some base k,

$$A_k \leq \frac{S_k(n)}{n} - \frac{k-1}{2}\log_k n \leq 0, \tag{4.7}$$

where $A_k < 0$ is a constant specified in ref. [71]. Since $|E| = S_k(n)$ for a bricklayer graph, the edge count $|E|$ also follows these bounds. Therefore, the robustness of a bricklayer graph is bounded below and above by

$$\frac{\log_k n}{\ell} + \frac{2A_k}{(k-1)\ell} \leq \rho(G_{n,k}) \leq \frac{\log_k n}{\ell}. \tag{4.8}$$

There is no short formula to calculate A_k, but Galkin and Galkina [71] have found an exact, though fairly involved, algorithm to determine A_k. For $k=2$, $A_2 = \log_4 3 - 1$; for $k=3$, $A_3 = \log_3 2 - 1$ and for $k=4$, $A_4 = (3/2)\log_2 5 - (9/4)$. Moreover, as $k \to \infty$,

$$A_k = -\frac{k}{2}\left[1 - \frac{\log\log k}{\log k} + \mathcal{O}\left(\frac{1}{\log k}\right)\right]. \tag{4.9}$$

An algorithm for exactly calculating A_k for a fixed value of k is given in ref. [71]. By plugging eq. (4.9) into eq. (4.8), we see that, for any fixed k, the robustness of a bricklayer graph is bounded above by the traditional logarithm curve and below by a $\mathcal{O}(1/\ell)$ additive term.

4.2.3 Optimal Upper Bound on Robustness of All Neutral Networks

Finding the subgraph G with a fixed number of vertices of a Hamming graph $H_{\ell,k}$ that maximises the number of edges (and robustness) is equivalent to minimising the "edge boundary" of the subgraph, i.e. minimising the number of edges $\{u, v\}$ which connect a subgraph vertex $u \in V(G)$ to a vertex outside the subgraph $v \in V(H_{\ell,k} \backslash G)$. This is known as the "edge-isoperimetric problem" for the Hamming graph. Harper [62] had showed that bricklayer's graphs attain the maximum bound for the $k = 2$ case, and Graham [73], and Hart [74] calculated the exact value of the bound for $k = 2$, namely $|E(G)| \leq S_2(n)$. Lindsey [63] generalised the work of Harper [62] to prove that bricklayer's graphs attain the maximum bound for all $k \geq 2$ but did not calculate the value of the bound. Using combinatorial methods, Squier et al. [58] provided the value of the tightest-known bound for $k > 2$:

$$|E(G)| \leq \frac{k-1}{2} n \log n. \tag{4.10}$$

From ref. [63], it is known that bricklayer's graphs are maximally robust for all $k \geq 2$—i.e. they (not necessarily uniquely) attain the upper bound for the edge count (and therefore the graph density and robustness of a neutral network/induced subgraph of a Hamming graph). Greenbury [3] argued that bricklayer's graphs provide the maximum robustness, but, to our knowledge, connections to Lindsey's proof [63] have never been made, so the assertion was not rigorous. Below, this thesis now uses the coding theory result of Lindsey [63] to formalise the final piece of the generalisation, as the following theorem generalises the proof by Graham [73] and Hart [74] for all $k \geq 2$, and improves (and optimises) the bound given by Squier et al. [58]:

Theorem 4.2.2. *The number of edges $|E(G)|$ of a subgraph G of a Hamming graph $H_{\ell,k}$ with fixed number of vertices $n = |V(G)|$ is optimally bounded by $|E(G)| \leq S_k(n)$. Accordingly, the biological robustness of any neutral network is optimally bounded by $\rho(G_{n,k}) \leq 2S_k(n)/(n\ell(k-1))$.*

Proof. The statement follows from ref. [63], which proves that bricklayer's graphs attain the maximal edge count among all subgraphs of a Hamming graph of a fixed number of vertices n, and our/Greenbury's Theorem 4.2.1, which shows that bricklayer's graphs have exactly $S_k(n)$ edges. This upper bound is optimal because we can always construct a bricklayer's graph with number of vertices n. The bound on the robustness follows by applying its definition. □

This bound now allows us to rigorously show, for the first time, an interesting property of the sums-of-digits function $S_k(n)$, generalising the proof by Graham [73], who proved the following for $k = 2$, which we now prove for general k:

Theorem 4.2.3. *For k nonnegative integers $\{n_1, n_2, \dots, n_k\}$ obeying $n_1 \leq n_2 \leq \dots \leq n_k$, the following property of the sums-of-digits function holds:*

$$\sum_{i=1}^{k} S_k(n_i) + \sum_{i=1}^{k-1} (k-i) n_i \leq S_k\left(\sum_{i=1}^{k} n_i\right) \tag{4.11}$$

Proof. Let $n = \sum_{i=1}^{k} n_i$, and choose ℓ such that $n_k \leq k^\ell$. We must necessarily have that $n \leq k n_k \leq k^{\ell+1}$. Consider the Hamming graph $H_{\ell+1,k}$. Since $H_{\ell+1,k} = H_{\ell,k} \Box K_k$, we can decompose the edge set of $H_{\ell+1,k}$ into

$$E(H_{\ell+1,k}) = \left(\bigcup_{i=1}^{k} E(H_{\ell,k}^{(i)})\right) \cup \left(\bigcup_{1 \leq i < j \leq k} E(H_{\ell,k}^{(i)}, H_{\ell,k}^{(j)})\right), \tag{4.12}$$

where $H_{\ell,k}^{(i)}$ is the Hamming graph consisting of the base-k representations of the integers whose (arbitrarily) first digit is $i-1$, and $E(G_1, G_2) = \{\{u, v\} \mid u \in E(G_1) \wedge v \in E(G_2)\}$ is the set of edges in $H_{\ell+1,k}$ which join two subgraphs G_1 and G_2. Note that for $1 \leq i \leq k$, we can construct a bricklayer's graph $G_{n_i,k}$ with n_i vertices that is a subgraph of the Hamming graph $H_{\ell,k}^{(i)}$. By Theorem 4.2.1, each bricklayer's graph has size (number of edges) equal to $S_k(n_i)$. Let us assume that each bricklayer graph has been constructed starting on the vertex such that its index's first digit is $i-1$, and all other digits are 0. This ensures that the $\left|E(G_{n_i,k}, G_{n_j,k})\right| = n_i$ for $i < j$ is maximal. The total number of edges in this graph G—i.e. the subgraph G of the Hamming graph $H_{\ell+1,k}$ induced by the vertex set

$V(G) = \bigcup_{i=1}^{k} V(G_{n_i,k})$—is given by the contributions from within the bricklayer's graphs and the connections between them:

$$\begin{aligned} |E(G)| &= \sum_{i=1}^{k} |E(G_{n_i,k})| + \sum_{1\le i<j\le k} \left|E(G_{n_i,k}, G_{n_j,k})\right| = \sum_{i=1}^{k} S_k(n_i) + \sum_{1\le i<j\le k} n_i \\ &= \sum_{i=1}^{k} S_k(n_i) + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} n_i = \sum_{i=1}^{k} S_k(n_i) + \sum_{i=1}^{k-1} (k-i) n_i. \end{aligned} \tag{4.13}$$

However, we also know that G has $|V(G)| = n$ and, by Theorem 4.2.2, size $|E(G)| \le S_k(n)$. Therefore,

$$|E(G)| = \sum_{i=1}^{k} S_k(n_i) + \sum_{i=1}^{k-1} (k-i) n_i \le S_k(n), \tag{4.14}$$

and this completes the proof. □

Theorem 4.2.3 is not only an interesting property of the sums-of-digits function which may be useful in coding theory. In the following section, we show how it can be used to provide a deeper understanding and analytical quantification of how the robustness of phenotypes in real biological systems *deviates* from the actual maximum robustness attained by the bricklayer's graphs.

4.3 Neutral Components of Biological GP Maps Attain the Bricklayer's Graph Bound

Greenbury [3] had shown that the neutral components of RNA secondary structure maps with only G and C nucleotides attain the bricklayer's bound; the present author used the Greenbury-Schaper-Ahnert-Louis (GSAL) dataset [1] to uncover that the RNA-GC secondary structure maps are not the only ones which attain the bricklayer's bound. In fact, the RNA (length 12 and 15) secondary structure maps with all 4 nucleotides (ACUG) as well as the HP protein folding models (length 24, and 5 × 5 lattice) have neutral components that exactly meet the bricklayer's bounds as well.

Recall that in Figure 1.1B, we showed that a phenotype's neutral network can be broken into multiple neutral components of various sizes. In Figure 4.2, the robustness values of each neutral component in the RNA12, RNA15, HP24, and

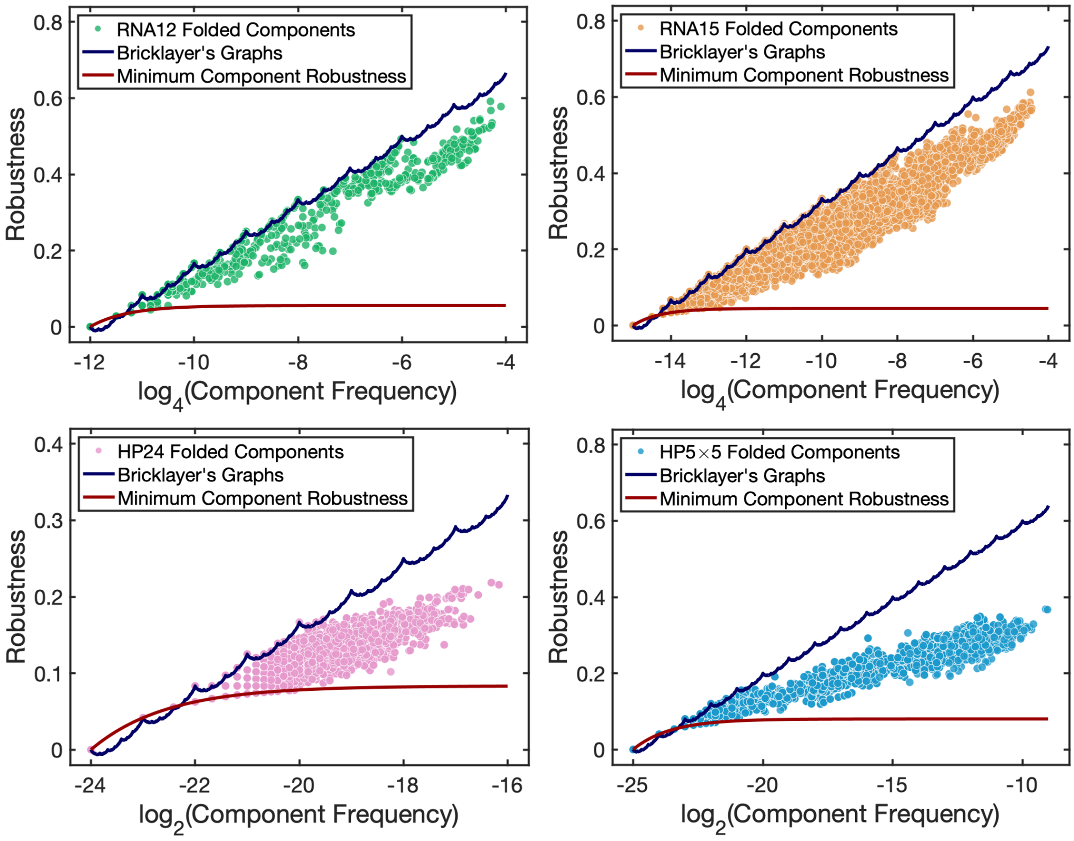


Figure 4.2: RNA and HP protein folding secondary structure GP maps. The robustness of every component of every folded phenotype for both the RNA and HP GP maps (of various lengths) is plotted against the frequency (fraction of vertices $|V(G)|/k^\ell$ in the entire Hamming graph) alongside the bricklayer's graph upper bound (blue line) we have derived here and the minimum robustness (red line) of a neutral component (easily derived from the minimum number of edges in a connected graph). The natural maps all contain neutral components which attain the bricklayer's graph bound (as well as the minimum bound). The unfolded (trivial) phenotype is ommitted from each of these plots. The minimum robustness appears to be larger than the bricklayer's graph line for low frequencies; but, this only happens for non-integer values of the number of vertices. Of course, any graph will have an integer number of vertices; in all of those cases, the bricklayer's graph robustness will be greater than or equal to the minimum robustness.

HP5 × 5 models are plotted against the logarithm of the number of vertices in that neutral component. The plots also show the bricklayer's graph (i.e. maximum possible) robustness for each neutral component size calculated in the previous section as well as the minimum robustness of each neutral component G_i, given in ref. [3] by:

$$\rho_{\min}(G_i) = \frac{2}{\ell(k-1)}\left(1 - \frac{1}{|V(G_i)|}\right). \tag{4.15}$$

The above formula follows from the fact that a neutral component, by definition, is connected, and the minimum number of edges in a connected graph with n vertices is $n-1$ [3]. This minimum value can be attained by many graphs, including the path graph P_n and star graph $K_{1,n}$. In fact, this was the robustness that individual phenotypes/outputs had for the 1D Edwards-Anderson spin glass input-output map studied in Chapter 2.

For all of the systems studied, sufficiently small neutral components do attain the same robustness as bricklayer's graphs over several orders of magnitude of phenotype frequencies (number of vertices). We notice that in the HP model maps, the size of the largest neutral components which still reach the bricklayer's graph line have fewer vertices than the same for RNA secondary structure GP maps. This is likely due to the architecture of the GP maps themselves; it has been shown recently [75] that neutral components are often modular in that they consist of highly packed clusters of vertices which are then connected to other clusters by a smaller set of linking vertices. We speculate that in the HP map there may be higher modularity, leading to less-than-maximally robust neutral components above a lower threshold.

In agreement with the theory, we also see that no neutral components exceed the bricklayer's graph bound. Our natural results highlight the importance of the rigorous calculations of an upper bound. Previous studies have shown these robustness values plotted against the asymptotic logarithmic bound, but these bounds were clearly not as tight as possible; we have now proven the exact maximum robustness that can be attained by any GP map. The constraint is not imposed due to the specific properties of the GP map (biological or not); rather, the robustness is bounded above due the underlying mathematics of the GP map, now understood more directly in the Hamming graph framework.

4.4 Robustness of Phenotypes Deviates From the Bricklayer's Bound

While some of the neutral components of natural maps do achieve the bricklayer's bound (the maximum robustness value), the *phenotype* robustness values of natural biological GP maps tend to deviate from this optimum, as seen clearly in Figure 2(A) in ref. [1], which is reproduced in Chapter 1 as Figure 1.2. The immediate reason, of course, is the fact that a bricklayer's graph is maximally dense in its edge-to-vertex ratio; it must certainly be fully connected. The mere existence of neutral components which are not connected to each other in a particular phenotype/output's neutral network means that the phenotype neutral network definitely cannot be a bricklayer's graph, even if its neutral components are.

While this is straightforward, a discussion of *how much* real phenotypes actually deviate from the bricklayer's bound is lacking in the literature, especially any sort of analytical mathematical treatment. We now use Theorem 4.2.3 to provide a tight bound on the difference in number of edges in a real phenotype consisting of multiple neutral components and the bricklayer's graph with the same number of vertices. First, we consider a neutral network which has n vertices and is split into m neutral components. If each neutral component is maximally robust (as many of the RNA/HP neutral components are), then each robustness would be

$$\rho(G_{n_i,k}) = \frac{2S_k(n_i)}{n_i\ell(k-1)}, \quad 1 \leq i \leq m, \tag{4.16}$$

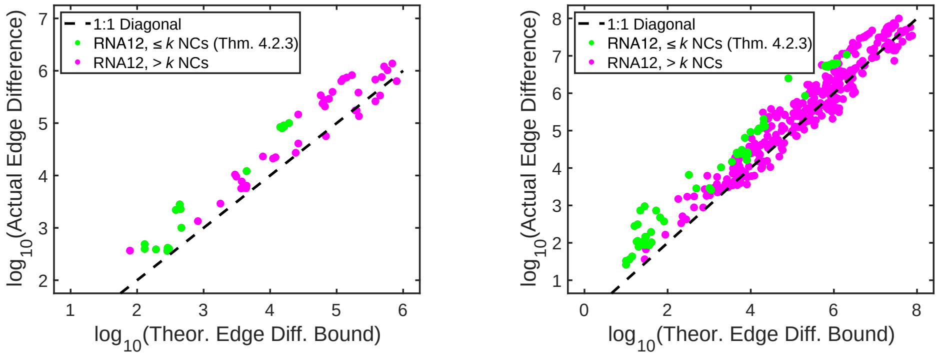
where $n = \sum_{i=1}^{m} n_i$, with n_i being the number of vertices in the i-th neutral component. The result of Theorem 4.2.3,

$$\sum_{i=1}^{k} S_k(n_i) + \sum_{i=1}^{k-1}(k-i)n_i \leq S_k\left(\sum_{i=1}^{k} n_i\right), \tag{4.17}$$

could be rewritten as

$$\left(\rho(G_{n,k}) - \frac{1}{n}\sum_{i=1}^{k} n_i\rho(G_{n_i,k})\right) \geq \frac{2}{n\ell(k-1)}\sum_{i=1}^{k-1}(k-i)n_i. \tag{4.18}$$

We note that $(1/n)\sum_{i=1}^{k} n_i\rho(G_{n_i,k})$ is the robustness of the entire phenotype neutral network consisting of multiple bricklayer's graphs as neutral components; a further generalisation and discussion of this formula is provided in section Section 4.5. This is now a rigorous bound on the difference between the maximum possible robustness of a phenotype and its true robustness, assuming that its neutral components are bricklayer's graphs.



(a) RNA12, $k = 4$, $\ell = 12$

(b) RNA15, $k = 4$, $\ell = 15$

Figure 4.3: Plot of left hand side (ordinate) and right hand side (abscissa) of eq. (4.19) for real phenotype data for RNA12 and RNA15 from the GSAL dataset [1]. Green plot points represent phenotypes with $\leq k = 4$ neutral components; the theoretical bound in eq. (4.19) rigorously holds for these phenotypes. Magenta plot points have > 4 neutral components; despite the fact that the bound should not rigorously hold for such pheontypes, it still does seem to hold for a large number of phenotypes, or at least many plot points lie close to the dashed line.

The assumption that the neutral components are bricklayer's graphs is not necessary, however. Weakening this assumption to assume that each neutral component has an arbitrary topology simply weakens the tightness of the bound. For an arbitrary neutral component A_{n_i} with n_i vertices, $\rho(A_{n_i}) \leq \rho(G_{n_i,k})$, so

$$\frac{n\ell(k-1)}{2}\left(\rho(G_{n,k}) - \frac{1}{n}\sum_{i=1}^{k} n_i \rho(A_{n_i})\right) \geq \sum_{i=1}^{k-1}(k-i)n_i. \qquad (4.19)$$

It is important to note that this inequality has been proven to hold only when the number of neutral components is less than or equal to k. Indeed, it does hold perfectly for the biological RNA neutral component/phenotype robustness data from the GSAL dataset [1]. In Figure 4.3, each plot point represents a phenotype, and the vertical axis coordinate is given by the log of the left hand side of eq. (4.19), which is (the log of) the difference in the optimal number of edges and the actual number of edges for that phenotype. The horizontal axis coordinate is given by the log of the right hand side of eq. (4.19), which is a theoretical bound computed from the frequencies of the neutral components for that phenotype. The left plot in Figure 4.3 shows actual RNA12 phenotype data, and the right plot shows RNA 15 phenotype data. Green plot points have $k = 4$ or fewer neutral components; it is for these phenotypes that the theoretical bound rigorously holds. The biological data support the theory if all green plot points are *above or on* the dashed 1:1 diagonal line, as this would indicate that the inequality is valid; indeed that is the case.

Most of the phenotypes in the RNA12 and RNA15 secondary structure maps do not have $\leq k = 4$ neutral components, however. Also in Figure 4.3, we have plotted, in magenta, the values of the left and right hand side of eq. (4.19) for phenotypes with $> k = 4$. The theoretical bound seems to hold even for most of the cases outside of the range for which it is proven, but it begins to fail for sufficiently large phenotypes.

Here, we discussed the process by which many neutral components are combined into one larger phenotype; these neutral components are not connected to each other (by definition), so the phenotype robustness is simply a frequency weighted average of the component robustnesses which will be necessarily lower than the maximum possible achievable robustness for a single-component phenotype. With this being said, in practice, an evolving population will typically be confined to a component, as double neutral mutations are typically quite rare. Therefore, even though the robustness of the phenotype is lower, the robustness experienced in shorter timescales by the population may often be closer to the bricklayer's graph bound.

In the following section, we will discuss the process of coarse-graining multiple *phenotypes* together into abstract phenotypes; this requires a more generalised approach because different phenotypes typically have nonzero transition probabilities between each other, unlike neutral components.

4.5 Theory of Robustness of Coarse-Grained Outputs/Phenotypes

In discussing neutral network and neutral component topologies for natural systems, one should keep in mind that phenotype definitions are to some extent arbitrary. What constitutes a phenotype for a particular input-output/genotype-phenotype map is determined *a priori.* Giegerich et al. [76] have, for instance, defined a set of new "abstract shape" RNA secondary structure GP maps where the genotype remains an RNA sequence, and the phenotype is a coarse-grained secondary structure that, at increasing levels of coarse-graining, ignores fine details of the stem and loop lengths and nesting. Dingle et al. [23] have shown that the most frequent coarse-grained RNA structures that appear in nature are also strongly biased towards in the genotype-phenotype map as well. In coarse-grained maps such as this, how parameters such as robustness scale becomes determined by the level of coarse-graining as well as by the underlying Hamming graph topologies. As we showed in the previous section, what appears robust at the neutral component level may join with other relatively robust neutral components to form a phenotype with much lower robustness.

Here, we consider the robustness of a phenotype and transition probabilities between phenotypes which have been generated from the union of multiple neutral networks. We refer to this process of merging neutral networks of different phenotypes as "coarse-graining" of phenotypes. We show general, exact formulas which follow from the underlying graph theory. We then examine collaborator Tasmin Sarkany's [64] numerical results on coarse-grained RNA GP maps and point out the surprising change in robustness as phenotype coarse-graining occurs. We conclude by deriving a critical transition probability that would be needed for two coarse-grained phenotypes to maintain "high" robustness when coarse-grained together.

4.5.1 Notational Preliminaries

Consider a GP map whose genotypes consist of sequences of length ℓ drawn from an alphabet of k characters. The genotype space is the Hamming graph $H_{\ell,k}$, and phenotype neutral networks are induced subgraphs of $H_{\ell,k}$. The i-th phenotype's neutral network G_i (assuming $1 \leq i \leq N_p$, where N_p is the total number of phenotypes) is an induced subgraph of $H_{\ell,k}$. Let $V(G)$ denote the vertex set of a graph G, $E(G)$ denote the edge set of G, and $E(G_i, G_j) = E_H(G_j, G_i)$ denote the set of edges induced in graph $H_{\ell,k}$ by union $V(G_i) \cup V(G_j)$ which are neither elements of $E(G_i)$

nor $E(G_j)$, where we have taken both G_i and G_j for $i \neq j$ to be induced subgraphs of $H_{\ell,k}$. Mathematically, $E(G_i, G_j) = \{\{u, v\} \in H_{\ell,k} \,|\, u \in V(G_i) \wedge v \in V(G_j)\}$.

As we have defined before, the robustness of the i-th phenotype is given by

$$\rho_i = \frac{2|E(G_i)|}{\ell(k-1)|V(G_i)|}, \tag{4.20}$$

so it is proportional to the ratio of edges to vertices in the neutral network. The transition probability that a single point mutation in the genotype leads to a change from phenotype i to phenotype j is given by

$$\phi_{ji} = \frac{|E(G_i, G_j)|}{\ell(k-1)|V(G_i)|}, \quad i \neq j. \tag{4.21}$$

Note that $\phi_{ji}|V(G_i)| = \phi_{ij}|V(G_j)|$. We define the diagonal terms $\phi_{ii} \equiv \rho_i$ so that there is an additional prefactor of 2.

4.5.2 Robustness and Transition Probabilities for Coarse-Grained Phenotypes

Robustness of Coarse-Grained Phenotypes

We first compute a general formula for robustness of coarse-grained phenotypes. Let S be the set of phenotype indices that indicate which phenotypes are being coarse-grained into a new neutral network G_S. The vertex set of G_S is the union of all vertices

$$V(G_S) = \bigcup_{s \in S} V(G_s). \tag{4.22}$$

The edge set of G_S includes all edges in each individual neutral network as well as the edges joining the neutral networks:

$$E(G_S) = \left(\bigcup_{s \in S} E(G_s) \right) \cup \left(\bigcup_{(r,s) \in S} E(G_r, G_s) \right), \tag{4.23}$$

where $(r, s) \in S$ denotes an ordered pair of elements of S such that $r \neq s$. It follows that the robustness of coarse-grained phenotype S is

$$\rho_S = \frac{2}{\ell(k-1)} \frac{\sum_{s \in S} |E(G_s)| + \sum_{(r,s) \in S} |E(G_r, G_s)|}{\sum_{s \in S} |V(G_s)|}. \tag{4.24}$$

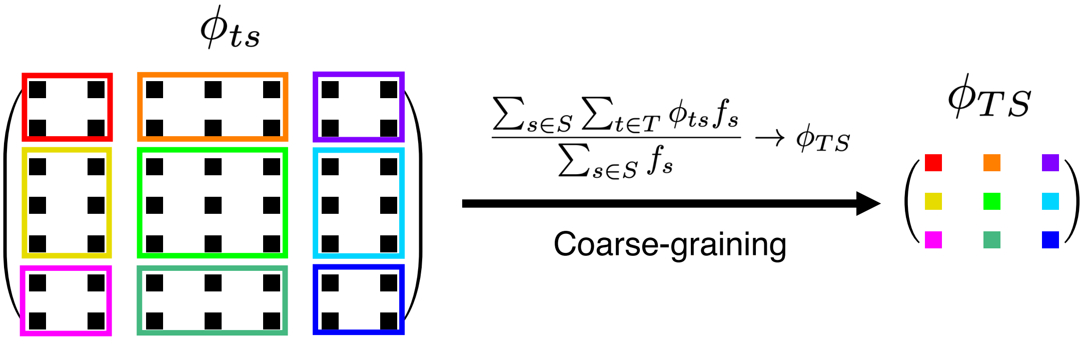

ϕ_{ts}
$\frac{\sum_{s\in S}\sum_{t\in T}\phi_{ts}f_s}{\sum_{s\in S}f_s} \rightarrow \phi_{TS}$
Coarse-graining
ϕ_{TS}

Figure 4.4: Schematic diagram of phenotype coarse-graining on the transition matrix $\phi_{ts} \rightarrow \phi_{TS}$, which includes transition probabilities (off-diagonals) and robustness (diagonals). If two non-overlapping sets of original phenotypes are coarse-grained into two new coarse-grained phenotype T and S, then the transition probability from coarse-grained phenotype S to coarse-grained phenotype T is given by ϕ_{TS}, calculated in eq. (4.27), which involves taking a frequency-weighted sum over the transition probabilities ϕ_{ts} between the original, non-coarse-grained phenotypes which comprise T and S.

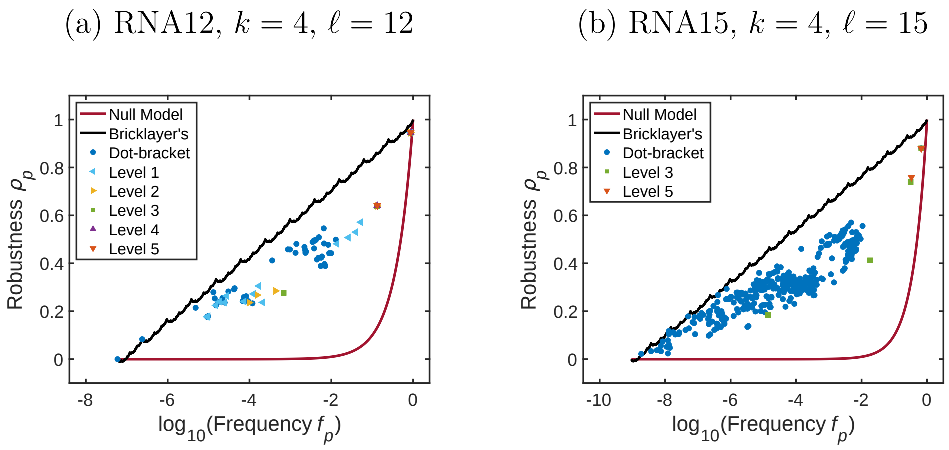
Using eq. (4.20) and the normalised phenotype frequency $f_i = |V(G_i)|/k^\ell$, we can rewrite the coarse-grained robustness in terms of familiar biological parameters

$$\rho_S = \frac{\sum_{s\in S}\rho_s f_s + \sum_{\{r,s\}\in S}\phi_{rs}f_s}{\sum_{s\in S}f_s} = \frac{\sum_{s\in S}\sum_{r\in S}\phi_{rs}f_s}{\sum_{s\in S}f_s}, \tag{4.25}$$

where $\{r, s\} \in S$ denotes an unordered pair of elements of S such that $r \neq s$, and in the last step we have used $\phi_{ss} = \rho_s$. It is easy to check that, if $S = \{1, 2, \ldots, N_p\}$ is the set of *all* phenotypes, then $\sum_{1\leq s\leq N_p}\phi_{rs}f_s = f_r$ (this is easily verified from the definition of ϕ_{rs}), so $\rho_S = 1$ as expected.

Transition Probabilities between Coarse-Grained Phenotypes

We now calculate a general formula for transition probabilities between coarse-grained phenotypes. Let S and T be two non-overlapping sets of phenotype indices that indicate which phenotypes are being coarse-grained into two coarse-grained neutral networks G_S and G_T, respectively. The set of edges $E(G_S, G_T)$ joining G_S and G_T is the union of all sets of edges that adjoin every pair of (non-coarse-grained) phenotypes, where within each pair one element is picked from the constituent phenotypes of S and the other is picked from constituent


(a) RNA12, $k = 4$, $\ell = 12$

(b) RNA15, $k = 4$, $\ell = 15$

Figure 4.5: RNA abstract phenotype robustness plots for various levels of coarse-graining for (a) RNA12 and (b) RNA15 models, primarily due to collaborator Tasmin Sarkany [64]. “Dot-bracket” structures are the standard folded RNA phenotypes which approximately match the RNA Vienna folding results. Level 1 is the first abstracted (coarse-grained) phenotype, including one or more dot-bracket structures based on coarse-grained topology. Level 2 includes phenotypes which are further coarse-grained from Level 1; level 3 includes phenotypes which are even further coarse-grained, etc. In the case of RNA12, Levels 4 and 5 are identical because the Level 4 phenotypes are already coarse-grained as much as possible. Also plotted are the bricklayer’s bound indicating the maximum possible robustness, the null model robustness, and the minimum robustness for a phenotype which contains only one component; this would be the robustness of a star graph, which we provided in eq. (2.12).

phenotypes of T. It follows that

$$E(G_S, G_T) = \bigcup_{s\in S}\bigcup_{t\in T} E(G_s, G_t) \tag{4.26}$$

It now follows that the transition probability ϕ_{TS} from the S-th coarse-grained phenotype to the T-th coarse-grained phenotype (assuming S and T have no overlap) is

$$\phi_{TS} = \frac{|E(G_S, G_T)|}{\ell(k-1)|V(G_S)|} = \frac{1}{\ell(k-1)} \frac{\sum_{s\in S}\sum_{t\in T}|E(G_s, G_t)|}{\sum_{s\in S}|V(G_s)|} = \frac{\sum_{s\in S}\sum_{t\in T}\phi_{ts}f_s}{\sum_{s\in S} f_s}. \tag{4.27}$$

We can now see that from eq. (4.25) and eq. (4.27) the coarse-graining procedure takes on the same functional form, which is represented graphically in Figure 4.4.

As we mentioned previously in this chapter, the process of coarse-graining neutral components into a phenotype neutral network is a specific case of the general process we have derived here, but with all transition probabilities between neutral components $\phi_{ji} = 0$ since neutral components are not connected to each other, by definition. In Figure 4.5, we reproduce results by Sarkany [64] in which RNA secondary structure GP maps had robustness values calculated for various levels of coarse-graining. Coarse-graining was performed using the RNA SHAPES tool [77]. The "dot-bracket" structure is the actual RNA secondary structure phenotype; Level 1 of coarse-graining ignores some details of the dot-bracket structure and combines similar phenotypes into the same abstract phenotype; the Level 2 structures include further coarse-graining, and so forth. There are 5 possible levels of coarse-graining.

Even though the systems in Figure 4.5 are too small to show a large number of Level 4 or 5 coarse-grained phenotypes, the overall trends are visible. The dot-bracket structures appear to be closest to the bricklayer's graph maximum robustness curve. At the highest levels of coarse-graining (Level 4/5), abstract phenotypes are so densely packed with dot-bracket phenotypes that a substantial portion of the Hamming graph sequence space is covered by only a small number of abstract phenotypes. This leads to a percolation-like phenomenon that allows for highly coarse-grained, large-frequency phenotypes having high robustness as would be intuitively expected.

At lower levels of coarse-graining, however, we see an unexpected trend. One may expect that coarse-graining dot-bracket phenotypes together would simply "push" the robustness parallel to the diagonal logarithm line. However, the data show that coarse-grained phenotypes with sufficiently small frequencies deviate from the maximal possible robustness (the bricklayer's graph bound) more than the phenotypes that comprise them. This is because the transition probabilties between these phenotypes being coarse-grained are likely too low to provide adequate increase in robustness after coarse-graining.

4.5.3 Critical Threshold for the Coarse-Graining of Phenotypes with High Robustness

We now consider the example of coarse-graining two phenotypes and ask how much the transition probability between those phenotypes should be in order to keep them along the same diagonal robustness line parallel to the bricklayer's graph

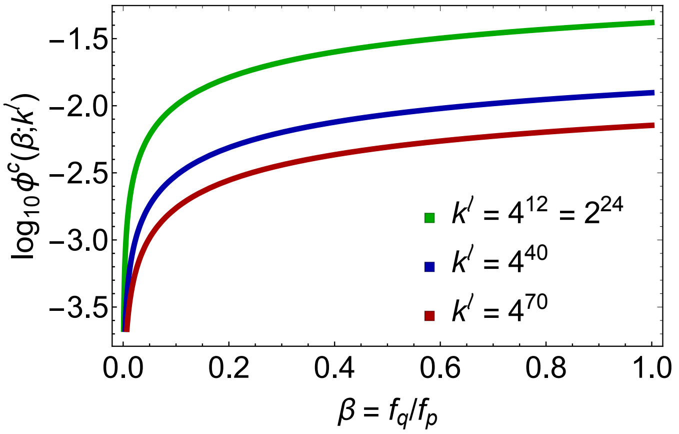


Figure 4.6: Plot of the theoretical critical transition probability $\log_{10}\phi^c(\beta;k^\ell)$ versus the ratio of frequencies $\beta \equiv f_q/f_p$. For $\phi_{qp} < \phi^c(\beta;k^\ell)$, the true coarse-grained robustness would be lower than the prediction.

bound. Recall that a phenotype's neutral network G_i which contains n vertices has at most $|E(G_i)| = S_k(n)$ edges, where $S_k(n)$ is once again the sums-of-digits function. We know that aymptotically $S_k(n) \sim (n/2)\log_k n$, and a reasonable approximation to the maximum robustness is

$$\rho_i \leq \frac{2S_k(n)}{n\ell(k-1)} \leq 1 + \frac{\log_k f_i}{\ell}. \tag{4.28}$$

In the *high-robustness* asymptotic assumption, we take $\rho_i \approx 1 + \ell^{-1}\log_k f_i$. This approximation will be employed below.

For two phenotypes p and q which are being coarse-grained into a new phenotype S, we can use eq. (4.25) to show that

$$\rho_S = \frac{\rho_p f_p + \rho_q f_q + 2\phi_{qp} f_p}{f_p + f_q}. \tag{4.29}$$

Let us assume that the two phenotypes have robustness values which are displaced by the same amount Δ from the (asymptotic) optimal robustness curve:

$$\rho_p = 1 + \frac{\log_k f_p}{\ell} - \Delta, \quad \rho_q = 1 + \frac{\log_k f_q}{\ell} - \Delta. \tag{4.30}$$

This approximation is consistent with empirical observations of robustness trends in many GP maps. Substituting these approximations into eq. (4.29), we have

$$\rho_S \approx 1 - \Delta + \frac{f_p \log_k f_p + f_q \log_k f_q + 2\ell\phi_{qp} f_p}{\ell(f_p + f_q)}. \tag{4.31}$$

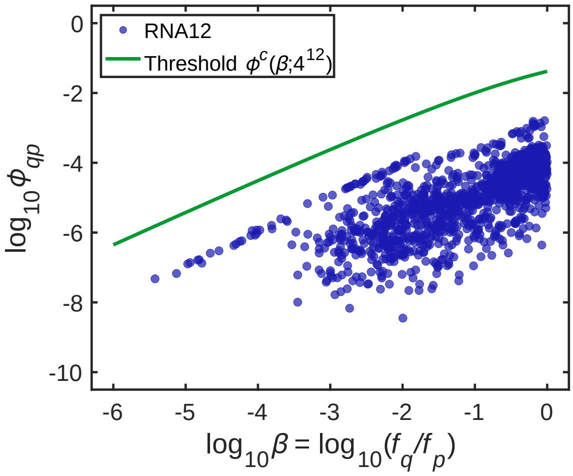


Figure 4.7: Plot of the critical transition probability $\log_{10}\phi^c(\beta;k^\ell)$ versus the ratio of frequencies $\beta \equiv f_q/f_p$ along with the real non-zero ϕ_{qp} transition probabilities for the RNA12 secondary structure map. For $\phi_{qp} < \phi^c(\beta;k^\ell)$, the true coarse-grained robustness is lower than the prediction.

We would intuitively expect two very robust phenotypes which have dense connections to each other (i.e. relatively high values of $\phi_{qp}f_p = \phi_{pq}f_q$) to be approximately collinear with the points $(\log f_p, \rho_p)$ and $(\log f_q, \rho_q)$ on a linear-log plot of robustness versus frequency. That is to say, we expect the robustness of the coarse-grained phenotype S to be

$$\rho_S^* \approx 1 + \frac{\log_k(f_p + f_q)}{\ell} - \Delta. \tag{4.32}$$

We now perform a change of variables to $\beta \equiv f_q/f_p$ and $f_S \equiv f_p + f_q$, so $f_p = f_S/(1+\beta)$ and $f_q = f_S\beta/(1+\beta)$. Without loss of generality, we can choose $f_p \geq f_q$, so $0 < \beta \leq 1$. We now rewrite eq. (4.31) as

$$\begin{aligned}\rho_S &\approx 1 - \Delta + \frac{f_S}{\ell(1+\beta)}\log_k\left(\frac{f_S}{1+\beta}\right) + \frac{f_S\beta}{\ell(1+\beta)}\log_k\left(\frac{f_S\beta}{1+\beta}\right) + \frac{2\phi_{qp}}{1+\beta} \\ &= \left(1 + \frac{\log_k f_S}{\ell} - \Delta\right) - \frac{\log_k(1+\beta)}{\ell} + \frac{\beta\log_k\beta}{\ell(1+\beta)} + \frac{2\phi_{qp}}{1+\beta}.\end{aligned} \tag{4.33}$$

The actual coarse-grained robustness deviates from the prediction in eq. (4.32) by

$$\rho_S - \rho_S^* = -\frac{\log_k(1+\beta)}{\ell} + \frac{\beta\log_k\beta}{\ell(1+\beta)} + \frac{2\phi_{qp}}{1+\beta}. \tag{4.34}$$

We can see that for $\phi_{qp} = 0$ (which implies $\phi_{pq} = 0$) or sufficiently small ϕ_{qp} in general, $\rho_S < \rho_S^*$ because $-\log_k(1+\beta) < 0$ and $\log_k \beta \leq 0$. The transition probability ϕ_{qp} needs to exceed a critical threshold $\phi^c(\beta; k^\ell)$ in order for the true robustness to equal or exceed the prediction. The true coarse-grained robustness is lower than the prediction for $\phi_{qp} < \phi^c(\beta; k^\ell)$, which is given by

$$\phi^c(\beta; k^\ell) = \frac{(1+\beta)\log(1+\beta) - \beta\log\beta}{2\log k^\ell}, \quad 0 < \beta \equiv \frac{f_q}{f_p} \leq 1. \tag{4.35}$$

In Figure 4.6, we plot $\phi^c(\beta; k^\ell)$ versus β for various realistic values of k^ℓ that correspond to the RNA12/HP24, RNA40, and RNA70 GP maps.

We hypothesize that the vast majority of transition probabilities ϕ_{qp} for any substantially large GP map will fall below this threshold. In Figure 4.7, we show that, from the GSAL dataset [1], *none* of the RNA12 transition probabilities between dot-bracket phenotypes fall above our approximate critical threshold.

We have now provided new theoretical intuition regarding the behaviour of robustness during the process of coarse-graining phenotypes, which is a new active area of research in the GP map community [23].

4.6 Base Information Content and Robustness of Bricklayer's Graphs

Information theory provides a useful lens through which to study genotype-phenotype maps. In this section, we explore the information content of two classes of graphs—the maximally robust bricklayer's graphs and the optimally informative Hamming graphs.

First, we introduce the information-theoretic definitions relevant to the study of input-output maps. The Shannon entropy (which we saw in a different context in Chapter 3) of the distribution of a discrete random variable X given by

$$\mathcal{H}(X) = -\sum_{i=1}^{N} \mathbb{P}(X = x_i)\log\mathbb{P}(X = x_i), \tag{4.36}$$

where $\{x_1, \ldots, x_N\}$ are the N possible outcomes in the support of X. It is easy to see that, with no constraints, the Shannon entropy is maximised when X is uniform over its support, so $\mathbb{P}(X = x_i) = 1/N$ for all i; the maximum Shannon entropy is

$$\mathcal{H}_{\max} = \log N. \tag{4.37}$$

Following Adami [78], in the context of genotype-phenotype maps, we know that in the mapping of sequences of length ℓ drawn from an alphabet of k characters, we have a total possibility of k^ℓ sequences, so a phenotype which is mapped onto by all possible sequences contains the maximal Shannon entropy of

$$\mathcal{H}_{\max} = \ell \log k. \tag{4.38}$$

Greenbury [3] notes that a phenotype p with F_p sequences would similarly have maximal entropy $\log F_p$. Following these two authors, we define the information $\mathcal{I}_p$ stored about the genotype sequences which map to phenotype p as the difference

$$\mathcal{I}_p = \mathcal{H}_{\max} - \mathcal{H}_p = \ell \log k - \log F_p. \tag{4.39}$$

In our discussion we will refer to $\mathcal{I}_p$ as the *total information content* in phenotype p.

Adami [78] approximates this information content by assuming that the total entropy $\mathcal{H}_p$ of the set of sequences is equal to the entropies of each base at each position, independently of each other. The sum is carried out over all sequences in the set. We write the Adami-approximated information content—to which we will refer as the *base information content*—as

$$\mathcal{I}'_p = \mathcal{H}_{\max} - \sum_{i=1}^{\ell} \sum_{j=1}^{k} \left(-P_{ij}^{(p)} \log P_{ij}^{(p)} \right), \tag{4.40}$$

where P_{ij} is the probability that the j-th base is found at the i-th position in phenotype p. These probabilities, by definition, are exactly

$$P_{ij}^{(p)} = \frac{(\text{number of occurrences of base } j \text{ at site } i)}{F_p}. \tag{4.41}$$

We now introduce the *relative base information error* between the base information content and the entire information content:

$$\langle \Delta \mathcal{I}_p \rangle = \frac{\mathcal{I}'_p - \mathcal{I}_p}{\mathcal{I}_p}. \tag{4.42}$$

By definition, $-1 \leq \langle \Delta \mathcal{I}_p \rangle \leq 0$ since $0 \leq \mathcal{I}'_p \leq \mathcal{H}_{\max} - \log F_p$ by subadditivity of entropy. This also means that $\mathcal{I}'_p \leq \mathcal{I}_p$.

Take an input-output map which is a labelling of the Hamming graph $H_{\ell,k}$, and suppose A_n is a neutral network with $1 < n \leq k^\ell$ vertices in the map, so it is an induced subgraph of $H_{\ell,k}$. Greenbury [3] made 2 central claims about small Hamming graphs and bricklayer's graphs:

Proposition 4.6.1 (Greenbury [3])**.** *Both robustness and base information content are simultaneously maximised for a Hamming graph $H_{r,q}$, but in general base information content and robustness are not simultaneously optimizable.*

Proposition 4.6.2 (Greenbury [3])**.** *Consider a bricklayer's graph $G_{n,k}$. Suppose one can form a Hamming graph $H_{r,q}$ which has the same number of vertices as $G_{n,k}$. Both $G_{n,k}$ and $H_{r,q}$ will have equal robustness.*

Greenbury provided a single numerical example in ref. [3] to support both propositions. However, neither claim holds fully in general. In this section, we qualify Proposition 4.6.1 with a caveat and prove the modified proposition in general analytically, and disprove Proposition 4.6.2 by counterexample, showing that it does not hold in general but may hold for special cases. We then go on to show that the numerical example provided by Greenbury was indeed a special case, and we analytically derive the entire set of solutions in this class of special cases by using the connections between robustness and the Trollope-Delange formula explored earlier in this chapter. In doing so, we additionally prove an interesting identity of the sums-of-digits function $S_k(n)$ which is novel, to the author's knowledge.

We now rigorously prove the following proposition, for which Greenbury had provided a numerical example in ref. [3]:

Theorem 4.6.1. *The relative base information error is (not necessarily uniquely) optimised by the Hamming graph $H_{r,q}$.*

Proof. Consider a GP map with an input sequence of length ℓ and k characters in the alphabet. Now, let G_p be a phenotype neutral network with q^r vertices, with

$r \leq \ell$ and $q < k$. It is easy to see that the Hamming graph $H_{r,q}$ is an induced subgraph $H_{\ell,k}$. This Hamming graph contains the set of all sequences of some length r that is drawn from some alphabet of size q. So, just like the background Hamming graph $H_{\ell,k}$, all of the possible q bases are present at all sites in equal proportion. Without loss of generality, suppose the sequence set for this phenotype consists of all sequences in which the first r sites contain one of the first q bases, and the remaining $\ell - r$ bases are fixed.

As a concrete example, let us consider the GP map which takes in sequences of length $\ell = 4$ and has an alphabet of $k = 4$. Let us use the base-k integer representations for simplicity, so our sequences are

$$\{0000, 0001, 0002, 0003, 0010, 0011, \ldots, 3332, 3333\}. \tag{4.43}$$

Now, suppose phenotype p is specified when sites 3 and 4 (counting from the left) are equal to 2, and when the first two sites contain either a 0 or a 1. So, this neutral network, $H_{r,q}$ where $q = 2$ and $r = 2$, has a genotype set given by

$$\{0022, 0122, 1022, 1122\}. \tag{4.44}$$

In general, for the Hamming graph neutral network $H_{r,q}$, the probability of finding the j-th base at the i-th site is

$$P_{ij}^{(p)} = \frac{1}{q}, \quad \forall i \in \{1, \ldots, r\}, j \in \{1, \ldots, q\} \tag{4.45}$$

The total information content $\mathcal{I}_p$ is

$$\mathcal{I}_p = \ell \log k - r \log q, \tag{4.46}$$

and the base information content $\mathcal{I}'_p$ is

$$\mathcal{I}'_p = \ell \log k - \sum_{i=1}^{r} \sum_{j=1}^{q} \left(-P_{ij}^{(p)} \log P_{ij}^{(p)} \right) \quad = \ell \log k - r \log q, \tag{4.47}$$

where we point out that the sum over i runs from 1 to r because the neutral mutations can only occur at the first r sites based on our definition (recall that this choice is arbitrary), and the sum over j runs from 1 to q because the neutral mutations can only occur between the first q letters of the alphabet, based on our definition (this choice is arbitrary as well). It is now clear that, for the Hamming graph neutral network $H_{r,q}$, we always have optimisation of the base information content, as $\mathcal{I}_p = \mathcal{I}'_p$, meaning that the base information content and total information content are equal, so the relative base information error is $\langle \Delta\mathcal{I}_p \rangle = 0$. □

It follows from the theorem above that a bricklayer's graph $G_{q^r,k}$ with q^r vertices, with $q \neq k$, does not necessarily optimise the base information content. Of course, for the special case $q = k$, the bricklayer's graph $G_{q^r,k}$ is a Hamming graph $H_{r,q}$, and therefore it *will* optimise base information content. Therefore, the part of Greenbury's proposition that states that base information content is simultaneously maximised for Hamming graphs is shown to be true in general.

Now, we discuss the robustness of Hamming graphs and provide counterexamples to Greenbury's claim that Hamming graphs and bricklayer's graphs provide the same robustness when they can both exist. The total number of edges in the Hamming graph $H_{r,q}$ is given by the standard formula $|E(H_{r,q})| = (q^r r(q-1))/2$. From Definition 1.2.3, the robustness of $H_{r,q}$ is

$$\rho(H_{r,q}) = \frac{r(q-1)}{\ell(k-1)}. \tag{4.48}$$

Meanwhile, the bricklayer's graph $G_{q^r,k}$ which contains the same number of vertices has a robustness of

$$\rho(G_{q^r,k}) = \frac{2}{q^r \ell(k-1)} S_k(q^r). \tag{4.49}$$

In general, since bricklayer's graphs optimise robustness, we must of course have

$$\rho(H_{r,q}) \leq \rho(G_{q^r,k}). \tag{4.50}$$

And as we mentioned, when $q = k$, $\rho(H_{r,q}) = \rho(G_{q^r,k})$. However, if we consider, for example, the case where $q = 3$, $k = 4$, and $r = 3$, we find that the number of edges in the Hamming graph is 288 while the number of edges in the bricklayer's graph with the same number of vertices is 292. Therefore, the robustness of the Hamming graph is lower than that of the bricklayer's graph. We believe the vast majority of cases where $q < k$ (since $q > k$ would not allow for the production of a Hamming graph) will yield a nonzero difference between the edge counts of bricklayer's graphs and Hamming graphs.

However, Greenbury [3] had provided a simple example where these two robustness values are equal. Namely, he based his claim on the fact that when $q = 2$, $r = 2$, and $k = 3$, we find that there are 4 edges in both neutral networks. We show that this special case is actually not an isolated one. In fact, the robustness of a bricklayer's graph $G_{(k-1)^2,k}$ and a Hamming graph $H_{2,k-1}$ contained within a larger Hamming graph $H_{\ell,k}$ for $\ell > 2$ will be equal regardless of the value of k chosen. Greenbury's example was a particular case within this class of special cases. We now show the interesting identity from which this result follows:

Theorem 4.6.2. *The sums-of-digits function $S_k(n)$ possesses a closed-form solution for the special case*

$$S_k\left((k-1)^2\right) = k^3 - 4k^2 + 5k - 2 \tag{4.51}$$

for all k.

Proof. For the $k = 2$ case, this formula is $S_2(1) = 0$, which is trivial. So, we need to consider the $k \geq 3$ cases. Firstly, we note from the Trollope-Delange formula that

$$S_k\left((k-1)^2\right) = \frac{(k-1)^2}{2}\left[(k-1)\log_k(k-1)^2 - g_k(k^{\{\log_k(k-1)^2\}-1})\right]. \tag{4.52}$$

Notice that the fractional part $\{\cdot\}$ can be written using the integer part $[\cdot]$ function instead:

$$\{\log_k(k-1)^2\} = \log_k(k-1)^2 - [2\log_k(k-1)] \tag{4.53}$$

We know that $\log_k(k-1) < 1$, is monotonic, and in the $k \to \infty$ limit, $\log_k(k-1)$ asymptotically approaches 1. Thus, we simply take note of the fact that for $k = 3$ (the lowest k we are considering here), $\log_3 2 \approx 0.631 > 0.5$, so $[2\log_k(k-1)] = 1$ for all integers $k \geq 3$. Now, we can write the argument of g_k as

$$k^{\{\log_k(k-1)^2\}-1} = k^{2\log_k(k-1)-2} = \frac{(k-1)^2}{k^2}. \tag{4.54}$$

Using the definition of g_k, we see

$$\begin{aligned} S_k\left((k-1)^2\right) &= \frac{(k-1)^2}{2}\left((k-1)\log_k(k-1)^2 - (k-1)\log_k\left(\frac{(k-1)^2}{k^2}\right)\right. \\ &\qquad \left. - \frac{k^2}{(k-1)^2} D_k\left(\frac{(k-1)^2}{k^2}\right)\right) \\ &= (k-1)^3 - \frac{k^2}{2} D_k\left(\frac{(k-1)^2}{k^2}\right). \\ &= (k-1)^3 - \frac{k^2}{2}\sum_{n=0}^{\infty} k^{-n} D_{k,0}\left(k^{n-2}(k-1)^2\right). \end{aligned} \tag{4.55}$$

Now, we simplify $D_{k,0}$:

$$\begin{aligned} D_{k,0}\left(k^{n-2}(k-1)^2\right) &= \int_0^{k^{n-2}(k-1)^2} \mathrm{d}t\,(2k[t] - 2[kt] + (k-1)) \\ &= k^{n-2}(k-1)^3 + 2\int_0^{k^{n-2}(k-1)^2} \mathrm{d}t\,(k[t] - [kt]) \end{aligned} \tag{4.56}$$

For the remaining integral, we first consider the cases where $n \geq 2$. The first term integrates over a step-ladder like function, yielding:

$$2k\int_0^{k^{n-2}(k-1)^2} \mathrm{d}t\,[t] = 2k \sum_{i=1}^{k^{n-2}(k-1)^2-1} i = k\left(k^{n-2}(k-1)^2\right)\left(k^{n-2}(k-1)^2 - 1\right). \tag{4.57}$$

For the second term, we perform the substitution $u = kt$, so

$$\begin{aligned} 2\int_0^{k^{n-2}(k-1)^2} \mathrm{d}t\,[kt] &= \frac{2}{k}\int_0^{k^{n-1}(k-1)^2} \mathrm{d}u\,[u] = \frac{2}{k}\sum_{i=1}^{k^{n-1}(k-1)^2-1} i \\ &= \frac{1}{k}\left(k^{n-1}(k-1)^2\right)\left(k^{n-1}(k-1)^2 - 1\right). \end{aligned} \tag{4.58}$$

Subtracting the previous two equations, we now have

$$2\int_0^{k^{n-2}(k-1)^2} \mathrm{d}t\,(k[t] - [kt]) = -k^{n-2}(k-1)^3, \tag{4.59}$$

so

$$D_{k,0}\left(k^{n-2}(k-1)^2\right) = 0, \quad n \geq 2. \tag{4.60}$$

Thus, only the first two terms in the series survive, which gives us

$$D_k\left(k^{n-2}(k-1)^2\right) = D_{k,0}\left(\frac{(k-1)^2}{k^2}\right) + \frac{1}{k} D_{k,0}\left(\frac{(k-1)^2}{k}\right). \tag{4.61}$$

Consider the $n = 0$ term first:

$$\begin{aligned} D_{k,0}\left(\frac{(k-1)^2}{k^2}\right) &= \int_0^{k^{-2}(k-1)^2} \mathrm{d}t \, (2k[t] - 2[kt] + (k-1)) \\ &= \frac{(k-1)^3}{k^2} + 2\int_0^{k^{-2}(k-1)^2} \mathrm{d}t \, (k[t] - [kt]) \, . \end{aligned} \tag{4.62}$$

The first term is

$$2\int_0^{k^{-2}(k-1)^2} \mathrm{d}t \, k[t] = 0 \tag{4.63}$$

because $k^{-2}(k-1)^2 < 1$ at most, and $[t] = 0$ in this range of values. The second term is evaluated using $u = kt$:

$$\begin{aligned} 2\int_0^{k^{-2}(k-1)^2} \mathrm{d}t \, [kt] &= \frac{2}{k}\int_0^{k^{-1}(k-1)^2} \mathrm{d}u \, [u] = \frac{2}{k}\left(\frac{k-2}{k} + \sum_{i=1}^{k-3} i\right) \\ &= \frac{2}{k}\left(\frac{k-2}{k} + \frac{(k-2)(k-3)}{2}\right) \end{aligned} \tag{4.64}$$

since $k-2 < k^{-1}(k-1)^2 < k-1$. So,

$$D_{k,0}\left(\frac{(k-1)^2}{k^2}\right) = \frac{(k-1)^3}{k^2} - \frac{2}{k}\left(\frac{k-2}{k} + \frac{(k-2)(k-3)}{2}\right) \tag{4.65}$$

Now, consider the $n = 1$ term:

$$\begin{aligned} \frac{1}{k}D_{k,0}\left(\frac{(k-1)^2}{k}\right) &= \frac{1}{k}\int_0^{k^{-1}(k-1)^2} \mathrm{d}t \, (2k[t] - 2[kt] + (k-1)) \\ &= \frac{(k-1)^3}{k^2} + \frac{1}{k}\int_0^{k^{-1}(k-1)^2} \mathrm{d}t \, (2k[t] - 2[kt]) \, . \end{aligned} \tag{4.66}$$

The first integral is solved the same way as eq. (4.64), so

$$2\int_0^{k^{-1}(k-1)^2} \mathrm{d}t \, [t] = 2\left(\frac{k-2}{k} + \sum_{i=1}^{k-3} i\right) = \left(\frac{k-2}{k} + \frac{(k-2)(k-3)}{2}\right) \tag{4.67}$$

The second integral is solved again via substitution $u = kt$, so

$$\begin{aligned} \frac{2}{k}\int_0^{k^{-1}(k-1)^2} \mathrm{d}t \, [kt] &= \frac{2}{k^2}\int_0^{(k-1)^2} \mathrm{d}u \, [u] \\ &= \frac{2}{k^2}\sum_{i=1}^{(k-1)^2-1} = \frac{1}{k^2}(k-1)^2\left((k-1)^2 - 1\right) . \end{aligned} \tag{4.68}$$

Thus, the $n = 1$ term can now be simplified to

$$\begin{aligned} &\frac{1}{k}D_{k,0}\left(\frac{(k-1)^2}{k}\right) \\ &= \frac{(k-1)^3}{k^2} + 2\left[\frac{k-2}{k} + \frac{(k-2)(k-3)}{2}\right] - \frac{1}{k^2}(k-1)^2\left((k-1)^2 - 1\right) . \end{aligned} \tag{4.69}$$

Adding the $n = 0$ and $n = 1$ terms, we find that

$$\frac{k^2}{2(k-1)} \sum_{n=0}^{\infty} k^{-n} D_{k,0} \left(k^{n-2}(k-1)^2\right) = \frac{k^2}{2} \frac{2(k-1)^2}{k^2} \quad = (k-1)^2, \tag{4.70}$$

so

$$S_k \left((k-1)^2\right) = (k-1)^3 - (k-1)^2 = k^3 - 4k^2 + 5k - 2. \tag{4.71}$$

This completes the proof. □

We can now show the following:

Corollary 4.6.2.1. *The robustness values of the bricklayer's graph $G_{(k-1)^2,k}$ and the Hamming graph $H_{2,k-1}$ are the same for all k.*

Proof. A Hamming graph $H_{r,q}$ has $q^r r(q-1)/2$ edges. Setting $q = k-1$ and $r = 2$, we see that the Hamming graph $H_{2,k-1}$ has $(k-1)^2(k-2) = k^3 - 4k^2 + 5k - 2$ edges. The bricklayer's graph $G_{(k-1)^2,k}$ has $S_k\left((k-1)^2\right)$ edges, and $S_k\left((k-1)^2\right) = k^3 - 4k^2 + 5k - 2$ by Theorem 4.6.2. □

Thus, we have shown that, in general, bricklayer's graphs and Hamming graphs do not have the same robustness except for some special cases, one class of which we have worked out explicitly. From numerical simulations, we are not aware of any other special cases that exist, but we cannot, at present, rule this out.

4.7 Population Neutrality and Hamming Spheres

For a population evolving on a Hamming graph $H_{\ell,k}$, suppose phenotype p corresponds to the highest-fitness phenotype. In the infinite-time steady-state limit, a very high fraction P of the population will lie in $G^{(p)}$, with some fraction P_s on each node $s \in V(G^{(p)})$ (such that $\sum_{s\in V(G^{(p)})} P_s = P$). Consider

$$\lambda(G^{(p)}) \equiv \sum_{s\in V(G^{(p)})} \frac{P_s \deg_{G^{(p)}}(s)}{P}, \tag{4.72}$$

where $\deg_{G^{(p)}}(s)$ is the number of neighbours of s which are also in $G^{(p)}$. This is an average of the asymptotic population proportion on each node of $G^{(p)}$, weighted by the degree of the node in $G^{(p)}$. Van Nimwegen et al. [66] ask us to consider 3 evolutionary random walks:

1. Consider a "blind ant" random walker which starts on $G^{(p)}$ which, at each time step, picks one of its $\ell(k-1)$ neighbours and steps only if that chosen neighbour is in $G^{(p)}$. Otherwise, it stays at the current node. This random walker spends equal time at all vertices in $G^{(p)}$ [79], and

$$\lambda(G^{(p)}) = \frac{1}{|V(G^{(p)})|} \sum_{s \in V(G^{(p)})} \deg_{G^{(p)}}(s) \equiv \rho(G^{(p)})\ell(k-1), \tag{4.73}$$

where $\rho(G^{(p)})$ is exactly the mutational robustness.

2. Consider a "myopic ant" random walker in which the walker starts on a random node in $G^{(p)}$. At the next time step, the random walker computes all of the neighbouring nodes which are also in $G^{(p)}$ and steps to one at random. In this case, [79]

$$\lambda(G^{(p)}) = \rho(G^{(p)})\ell(k-1) + \frac{\mathrm{Var}_s[\deg_{G^{(p)}}(s)]}{\rho(G^{(p)})\ell(k-1)}, \tag{4.74}$$

where $\mathrm{Var}[\deg_{G^{(p)}}(s)]$ is the variance in the degree distribution of the nodes in $G^{(p)}$.

3. Finally, consider more realistic mutation-selection dynamics which obey the discrete-time version of Eigen's quasispecies evolution model [80] in which every vertex is assigned a fitness. The maximally fit neutral network $G^{(p)}$ is taken to have some fitness (the exact value does not matter in the infinite time limit) which is uniform across the entire network, and in each time step, a population of random walkers (of constant size) is replenished, with individuals randomly drawn based on the distribution of fitnesses from the previous time step. In this limit, Van Nimwegen et al. [66] proved that the population distribution within $G^{(p)}$ does not rely on the mutation rate, the average fitness of the population, or the fitness of the neutral network itself. Rather, the equilibrium population distribution is a static property dependent only on the topology of the neutral network itself. Namely, the fraction of the population $G^{(p)}$ on a particular node is given by the corresponding component of the principal eigenvector of the adjacency matrix. The population neutrality

is equal to the weighted average of the equilibrium population fraction on each node of the neutral network, with each weight equal to the degree of that node.

The result of Van Nimwegen et al. [66] is formalised below (note that we now define "population neutrality" as $\widetilde{\lambda}(G^{(p)}) = \lambda(G^{(p)})/(\ell(k-1))$ instead of $\lambda(G^{(p)})$):

Definition 4.7.1. *The population neutrality $\widetilde{\lambda}(G^{(p)})$ of a neutral network $G^{(p)}$, which is an induced subgraph of $H_{\ell,k}$, is given by $\widetilde{\lambda}(G^{(p)}) = \lambda(G^{(p)})/(\ell(k-1))$, where $\lambda(G^{(p)})$ is the principal eigenvalue of the adjacency matrix of $G^{(p)}$.*

Reeves et al. [61] refer to $\widetilde{\lambda}(G^{(p)})$ as "robustness," deviating from the terminology used most commonly in the genotype-phenotype map literature. We will continue to use "robustness" to refer to one-step mutational robustness and "population neutrality" to refer to the parameter defined above, which is proportional to the principal eigenvalue of the adjacency matrix of a neutral network.

4.7.1 Population Neutrality of Bricklayer's Graphs

In the previous subsections, we discussed the robustness and information content of maximally robust neutral networks, the bricklayer's graphs $G_{n,k}$. We now discuss the principal eigenvalue $\lambda_1(G_{n,k})$ of bricklayer's graphs. Reeves et al. [61] conjectured the following log-bound for the upper bound of the population neutrality:

Conjecture 4.7.1 (Reeves et al. [61])**.** *The principal eigenvalue of a bricklayer's graph is bounded above by*

$$\lambda_1(G_{n,k}) \leq (k-1)\log_k n, \tag{4.75}$$

Our collaborator, Shyam Narayanan, has rigorously proven Conjecture 4.7.1; his proof is reproduced in Appendix C. We also have a rigorous and tight lower bound on the population neutrality of bricklayer's graphs, by Theorem 4.2.1:

Theorem 4.7.1. *The principal eigenvalue $\lambda_1(G_{n,k})$ of a bricklayer's graph $G_{n,k}$ is bounded below by $\lambda_1(G_{n,k}) \geq 2S_k(n)/n$.*

Proof. A well-known identity is that $\lambda_1(G) \geq \overline{d}(G) = 2|E(G)|/|V(G)|$, the average degree of the graph G. For a bricklayer's graph $G_{n,k}$, we have from Theorem 4.2.1 that

$$\frac{2|E(G_{n,k})|}{n} = \frac{2S_k(n)}{n} = (k-1)\log_k n - g_k\left(k^{\{\log_k n\}-1}\right) \leq \lambda_1(G_{n,k}). \tag{4.76}$$

This is the tightest known lower bound on the principal eigenvalue of a bricklayer's graph. Recall from eq. (4.1) and eq. (4.7) that $2A_k \leq g_k(x) \leq 0$. □

With these two results, we can now bound the population neutrality of bricklayer's graphs both above and below:

$$(k-1)\log_k n - g_k\left(k^{\{\log_k n\}-1}\right) \leq \lambda_1(G_{n,k}) \leq (k-1)\log_k n, \tag{4.77}$$

showing that the robustness and population neutrality both obey a logarithmic scaling law $\ell\rho(G_{n,k}) \sim \ell(k-1)\lambda_1(G_{n,k}) \sim (k-1)\log_k n - \mathcal{O}(1)$.

Van Nimwegen et al. [66] described the robustness as a low order approximation to the principal eigenvalue in describing the asymptotic population distribution. But here, it is clear that for bricklayer's graphs, the robustness certainly deserves more credit than that. The strong agreement between ρ and λ_1 for bricklayer's graphs shows that, for evolution on maximally robust neutral networks (or within maximally robust neutral components), the robustness, which has an exact analytical form, would likely provide an excellent approximation to the principal eigenvalue, which has no known analytical solution and can become expensive to compute. By using the bound given by Galkin and Galkina [71], we can guarantee that the population neutrality can be approximated by the robustness to within a margin of at most

$$\frac{\lambda_1(G_{n,k})}{\ell(k-1)} - \rho(G_{n,k}) \leq -\frac{2A_k}{\ell(k-1)} \sim \mathcal{O}\left(\frac{1}{\ell}\right), \tag{4.78}$$

as $k \to \infty$.

4.7.2 Hamming Spheres

Reeves et al. [61] considered whether bricklayer's graphs maximise the population neutrality relative to all possible induced subgraphs of the hypercube/Hamming graph $H_{\ell,2}$ (i.e. only the $k = 2$ case was covered). They found numerically that this is true for small graphs, but for $\ell \geq 19$ this no longer holds in general. There appears to be no existing literature on the types of subgraphs of Hamming graphs $H_{\ell,k}$ for $k > 2$ which maximise the prinicipal eigenvalue, but there have been investigations into the $k = 2$ (hypercube) case aside from the numerical approach of Reeves et al. [61].

Bollobás et al. have shown [67] that the star graph $K_{1,n}$ with $|V(K_{1,n})| = n$ vertices maximises the principal eigenvalue for $n \leq \ell$ for $|V| \geq 105$. The robustness of a star graph is relevant to the input-output map subgraph properties for the ground states of the 1-dimensional Edwards-Anderson $\pm J$ spin glass model as shown in Chapter 2, and it also is the minimum robustness of any fully connected neutral network or neutral component [3]. A star graph with n vertices has $|E(K_{1,n})| = n - 1$ edges, so the robustness is

$$\rho(K_{1,n}) = \frac{2}{\ell(k-1)} \left(1 - \frac{1}{n}\right). \tag{4.79}$$

Bollobás et al. [67] have more recently argued that a type of graph known as a *Hamming sphere* (or *Hamming ball*) maximises the principal eigenvalue (and therefore the population neutrality) for the $k = 2$ case; although Friedman and Tillich [81] have provided specific examples of very large or very small Hamming spheres which are not principal eigenvalue maximisers, they do not make claims about principal eigenvalue maximisation asymptotically. Bollobás et al. [67] show that Hamming spheres of radius $\eta \sim o(d)$ (where d is the underlying dimension of the hypercube $H_{d,2}$) have a principal eigenvalue which is asymptotically within $1 + o(d)$ of the true maximum. Moreover, they also show that for a fixed radius η, these Hamming spheres *exactly* maximise the largest eigenvalue for sufficiently large dimension d.

We now precisely introduce the Hamming sphere. Recall that a Hamming sphere contains the set of (vertices which represent) sequences which have at most a fixed number of errors relative to a fixed "codeword":

Definition 4.7.2. *A **Hamming sphere** $B_{\eta,\ell,k}$ with radius $0 \leq \eta \leq \ell$ is an induced subgraph of a Hamming graph $H_{\ell,k}$ whose vertex set corresponds to the set of all sequences within a Hamming distance at most η from a fixed "codeword" sequence.*

The number of vertices in a Hamming sphere is typically written in series form; we provide an analytical form that has not appeared in the literature, to the author's knowledge. For $0 \leq i \leq \eta$ errors in a sequence, there are $\binom{\ell}{i}$ ways the errors can be chosen, and each of the i erroneous sites can take on one of $k-1$ values (since one of the k values is "correct"). So, the number of sequences with exactly i errors is $\binom{\ell}{i}(k-1)^i$. Summing over allowed values of i (from 0 to η, inclusive), the Hamming sphere's $B_{\eta,\ell,k}$ number of vertices is

$$\begin{aligned} |V(B_{\eta,\ell,k})| &= \sum_{i=0}^{\eta} \binom{\ell}{i}(k-1)^i \\ &= k^\ell - (k-1)^{\eta+1}\binom{\ell}{\eta+1}{}_2F_1(1; 1+\eta-\ell; 2+\eta; 1-k), \end{aligned} \tag{4.80}$$

where ${}_2F_1(a;b;c;z)$ is the ordinary (Gaussian) hypergeometric function.

4.7.3 Exact Robustness of Hamming Spheres

We now calculate the exact number of edges (and from it, the robustness) of the Hamming sphere.

Theorem 4.7.2. *The size (number of edges) of a Hamming sphere $B_{\eta,\ell,k}$ is given by*

$$|E(B_{\eta,\ell,k})| = \frac{k-1}{2}\left\{\ell k^\ell + (k-1)^\eta\binom{\ell}{\eta}\left[\eta - {}_2F_1(1; \eta-\ell; 1+\eta; 1-k)\right]\right\}. \tag{4.81}$$

Proof. Suppose our sequences are of length ℓ and we are using a k-ary alphabet. A sequence with $0 \leq i \leq \eta - 1$ errors, has the maximal number of neighbours $\ell(k-1)$ because any site can be flipped and the number of errors will remain at most η.

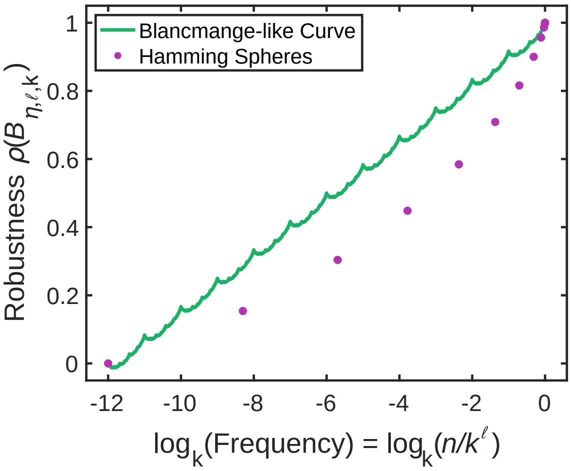


Figure 4.8: Plot of the robustness of a Hamming sphere $B_{\eta,\ell,k}$ as a function of the log fraction of vertices occupied within a larger Hamming graph $H_{\ell,k}$. Here, we use $k = 2$ and $\ell = 12$. For comparison, the bricklayer's graph bound is plotted, illustrating that the Hamming sphere does not meet the bricklayer's graph bound for graphs which do not have maximal (1) or minimal (k^ℓ) vertices.

As explained in the previous subsection, the number of sequences with exactly i errors is $\binom{\ell}{i}(k-1)^i$, so the sum of the degrees of the vertices within the boundary of the Hamming sphere (but not on the boundary) is $\ell(k-1)\sum_{i=0}^{\eta-1}\binom{\ell}{i}(k-1)^i$. On the boundary, there are $\binom{\ell}{\eta}(k-1)^\eta$ sequences which have η errors each. The only allowed flips would have to occur at one of the η erroneous sites, as they could be flipped either to another erroneous character in the alphabet or to the "correct" letter. Thus, the sum of the degrees of the vertices on the border of the Hamming sphere is $\eta(k-1)\binom{\ell}{\eta}(k-1)^\eta$. The sum of degrees is equal to $2|E(B_{\eta,\ell,k})|$, so the total number of edges is

$$|E(B_{\eta,\ell,k})| = \frac{k-1}{2}\left[\ell\sum_{i=0}^{\eta-1}\binom{\ell}{i}(k-1)^i + \eta\binom{\ell}{\eta}(k-1)^\eta\right], \tag{4.82}$$

which can be shown from a table of integrals to evaluate to the expression given in the theorem. □

To our knowledge, this is the first analytical expression of the number of edges of a Hamming sphere.

Now, given the number of vertices and edges of Hamming spheres as a function of radius η, we can write the robustness as a parametric function of η:

$$\rho(B_{\eta,\ell,k}) = \frac{k^\ell + (k-1)^\eta \binom{\ell}{\eta} \left[\frac{\eta}{\ell} - \ell {}_2F_1(1; \eta - \ell; 1 + \eta; 1 - k)\right]}{k^\ell - (k-1)^{\eta+1} \binom{\ell}{\eta+1} {}_2F_1(1; 1 + \eta - \ell; 2 + \eta; 1 - k)}. \tag{4.83}$$

In Figure 4.8, we plot the exact robustness of the Hamming sphere $B_{\eta,\ell,k}$ versus the fraction of vertices it occupies in the encompassing Hamming graph $H_{\ell,k}$. We observed that changing the size of the encompassing Hamming graph (i.e. shifting ℓ) does not change the curve that is traced out by the parameter η. The plot therefore suggests that

$$\rho(B_{\eta,\ell,k}) \sim 1 + \frac{\log_k |V(B_{\eta,\ell,k})|}{\ell} + o\left(\frac{\log_k |V(B_{\eta,\ell,k})|}{\ell}\right), \tag{4.84}$$

meaning the the Hamming sphere robustness asymptotically seems to scale just like that of the bricklayer's graph, but the offset from the logarithmic trajectory $\sim 1 + \frac{\log_k \left|V(B_{\eta,\ell,k})\right|}{\ell}$ is larger than that for the bricklayer's graphs. Of course, the bricklayer's graphs will always have equal or greater robustness than any other graph.

4.8 Discussion

In this chapter, we have used graph theory to investigate maximally robust neutral networks known as bricklayer's graphs. By using results from coding theory and number theoretic results on the sum-of-digits functions, we are able to exactly show the maximal robustness of biological neutral networks and that biophysical GP maps are close to this bound, which is what is expected from Chapter 3. We used the connection to the sums-of-digits function to then provide a theoretical lower bound on how much the robustness of a phenotype which has multiple neutral components would deviate from the maximal bricklayer's bound. We then move from considering the robustness of phenotypes comprised of multiple neutral components to the robustness of coarse-grained phenotypes which consist of other phenotypes which may not necessarily have nonzero transition probabilities between each other. We show how our general theory explains trends found in real RNA coarse-grained phenotypes.

Next, we considered the information content stored in the sequences which comprise a phenotype. In opposition to what has been suggested in the literature, we showed that, in general, the base information content and robustness are not simultaneously optimised. But, we exactly proved that a class of special cases does indeed allow for simultaneous optimisation.

We then discussed the population neutrality of bricklayer's graphs, which is equivalent to the principal eigenvalue of the adjacency matrix of the graph and has been suggested in the literature [66] to be more important than mutational robustness in evolutionary dynamics. The population neutrality had been conjectured by Reeves et al. [61] to be bounded above logarithmically. The author and collaborator Shyam Narayanan prove the lower and upper bounds, thereby resolving the conjecture and placing tight bounds which show that, in fact, the population neutrality and mutational robustness are nearly equal to each other, suggesting that the mutational robustness is not actually a poor substitute for population neutrality but is much more easily computable (as it can be sampled) compared to the principal eigenvalue.

Having worked out these mathematical properties of robustness on graphs, it is interesting to consider what the biological relevance is of these results. Intuitively, one might muse that natural neutral networks ought to be "optimal" from the viewpoints of information content, error tolerance, asymptotic population dynamics, or one-step mutation robustness, or all of the above. In the previous sections, we learned, however, that in constructing neutral networks, nature actually faces a number of trade-offs when dealing with simultaneous optimisation of base information content, population neutrality, and robustness.

We have investigated three classes of graph topologies for neutral networks, each of which is responsible for maximising a different parameter of evolutionary relevance. The bricklayer's graphs maximise robustness exactly, the Hamming graphs optimise base information content, and the Hamming spheres optimise population neutrality (at least for $k = 2$). There is no immediate reason to assume that optimisation of any of these parameters is mutually exclusive; and for some specific cases, it is possible to optimise all of them: for instance, consider the small graph $G_{16,4} = H_{2,4}$.

This optimises robustness, base information content, and population neutrality: robustness is optimised because it is a bricklayer's graph; base information content is optimised because it is also a small Hamming graph; and population neutrality is optimised because, by eq. (4.77), the lower and upper bounds are equal to each other, and the principal eigenvalue of the adjacency matrix thus equals the maximal value. But for large natural systems, these specific examples are highly unlikely to appear. The fact that neutral components for small RNA and HP protein folding systems attain the bricklayer's graph bound suggests that perhaps these natural sequence-to-structure maps "prioritise" robustness over information content and population neutrality. Here—like in Chapter 3—questions to be addressed in the future include, what happens to the robustness for much more complex systems, say, the human genome? Do these principles we have uncovered here still hold?

5

Robustness Scaling Law for Continuous Input-Output Maps

Contents

5.1 Introduction

Thus far, our study of robustness has centered around *discrete* input-output maps of the form

$$\Omega : \{0, \ldots, k-1\}^d \to \{0, \ldots, q-1\}. \tag{5.1}$$

However, some of the systems with discrete inputs that we have studied can alternatively be formulated (and perhaps more realistically) with continuous inputs. For example, Carmago and Louis' [18] study of Boolean threshold networks as gene regulatory networks relies on only Boolean inputs representing the regulatory weights between genes. A more realistic picture would likely allow for activation, noninteraction, supression, and behaviours in between. Our spin glass input-output maps in Chapter 2 were formulated for the $\pm J$ model on random graphs. While the $\pm J$ model is useful (as is its ability to convey sign epistasis in spin glass fitness landscape models), it would be more general to consider models that admit any level of positive, negative, or zero interaction between spins by making J a continuous variable and perhaps imposing some minimum or maximum cutoff.

These examples are inspiring, but perhaps one of the most immediate and important applications is to deep learning. Deep neural networks (DNNs) are graphs with an input layer of nodes that are densely connected to a hidden layer of nodes, which is connected to another hidden layer of nodes, and so on, until the nodes in the last hidden layer are connected to the nodes in output layer. Training a DNN in a supervised learning setting involves updating weights such that some pre-defined loss function dependent on the results of the output nodes is progressively minimised as the DNN "learns" more about the training data by comparing the DNN outputs to the known answers in the training data. The weights are fixed after training, and one can feed in new test data and expect to receive fairly accurate output results.

The parameter space of DNNs is typically very high-dimensional (due to the existence of many weights on edges as well as nodes in the graph), and navigation of the parameter space is guided by a process such as stochastic gradient descent, which uses a loss function to guide the path of the weights. In this regard, stochastic gradient descent is similar to simulated annealing in thermodynamics (funnelling down an energy landscape) or biological evolution involving climbing up a fitness landscape. The energy-loss-fitness analogies motivate the use of methods from physics and evolution to think about deep learning.

Our work on discrete systems in evolution and statistical physics motivates an extension of the robustness framework to continuous input spaces. In this chapter, we define robustness in a continuous space that has been discretised. We then derive the equivalent of the logarithm power law for robustness that is expected from natural discrete systems—in these "continuous input"-"discrete output" maps, we find that the robustness should scale as a *power law*. We then derive a series of Ansätze/approximations which are increasingly more parameterised and which try to capture the salient details of the neutral spaces (the sets of points which map to the same output). Using robustness data from deep neural networks collected by collaborator Yoonsoo Nam [82], we are able to show that deep neural networks indeed do appear to have power law neutral spaces, supporting our prediction.

5.2 Theory

5.2.1 Definition of Robustness

In the genotype-phenotype maps and input-output maps discussed in prior chapters, the number of inputs was finite, and a notion of distance was imposed on sequences via the Hamming distance. Two sequences were connected by an edge in the input space if the Hamming distance between them was 1. Now, we are dealing with continuous inputs, so there can exist a Euclidean metric on this space. However, to be able to define robustness in terms of “nearest-neighbours” as is done in the discrete systems, we first discretise the Euclidean space.

We will consider systems with continuous weights, but discrete outputs, such as is found for classification problems in machine learning. Now, consider such a continuous-to-discrete input-output map:

$$\phi : \mathbb{Z}^n \to \{\mathbf{v} \in \mathbb{Z}^q \,|\, \|\mathbf{v}\|_0 = 1\}, \tag{5.2}$$

on an n-dimensional hypercubic lattice with arbitrary spacing ϵ. Let $\phi_i(\mathbf{w})$ be the indicator function such that $\phi_i(\mathbf{w}) = 1$ if the input vector $\mathbf{w}$ maps to the i-th output and $\phi_j(\mathbf{w}) = 0$ for $j \neq i$. Assume that there are q outputs.

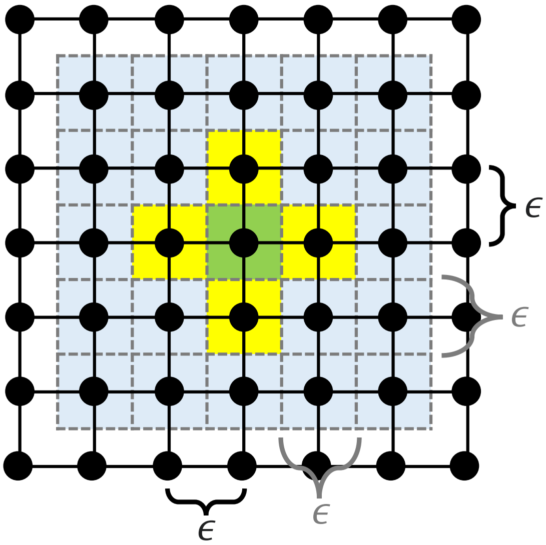


Figure 5.1: Discretisation of function output space. Lattice points are shown in black. (Green box) the n-dimensional box surrounding the lattice point $\mathbf{w}$. (Yellow boxes) the neighbouring lattice points $\mathbf{w}$ and the $(n-1)$-dimensional hypersurfaces (gray dashed lines) shared with with the green box. Robustness is the fraction of neighbours of each box within a neutral set $\mathcal{G}_i$, averaged over all lattice points in the neutral set $\mathcal{G}_i$.

The standard definition of robustness comes from the biological literature in which ϕ_i is calculated on a Hamming graph instead of a hypercubic lattice graph as we have here. Here, the robustness ϕ_i of the i-th output is defined as

$$\rho_i = \frac{1}{2nF_i} \sum_{\mathbf{w}\in\mathcal{G}_i} \sum_{k=1}^{n} \left[\phi_i(\mathbf{w}+\epsilon\hat{\mathbf{e}}_k) + \phi_i(\mathbf{w}-\epsilon\hat{\mathbf{e}}_k)\right], \tag{5.3}$$

where $\mathbf{w}$ is a lattice point to begin with, and by our definition $\mathbf{w}+\epsilon\hat{\mathbf{e}}_k$ and $\mathbf{w}-\epsilon\hat{\mathbf{e}}_k$ map onto other lattice points which are not necessarily found within the neutral set (i-th output function space) $\mathcal{G}_i = \{\mathbf{w} : \phi_j(\mathbf{w}) = \delta_{ij}, \forall j\}$. In eq. (5.3), $F_i = |\mathcal{G}_i|$ is the number of lattice points within $\mathcal{G}_i$. The lattice is given by the black points (and edges between them) in fig. 5.1.

Now we consider the n-dimensional hypercubic dual lattice (gray dotted lines in fig. 5.1). Each lattice point in the primary lattice is contained within a volume ϵ^n in the dual lattice, and each of these n-dimensional cubes shares an $(n-1)$-dimensional hypercubic face of volume ϵ^{n-1} with the boxes of the $2n$ neighbouring lattice points. From the definition of robustness in eq. (5.3), we see that robustness counts the fraction of neighbours (out of $2n$) of a lattice point $\mathbf{w} \in \mathcal{G}_i$ which are also within

$\mathcal{G}_i$, then averages this over the F_i lattice points in $\mathcal{G}_i$. Therefore, $1-\rho_i$ counts the fraction of neighbours (out of $2n$) which are *not* in $\mathcal{G}_i$; this is the same as counting the number of edges connecting a point $\mathbf{w} \in \mathcal{G}_i$ to a point outside of $\mathcal{G}_i$, then dividing by $2n$ and averaging over $\mathcal{G}_i$. This number of edges is equal to the hypersurface area S_i of the neutral set $\mathcal{G}_i$, divided by the $(n-1)$-hypersurface area ϵ^{n-1}. Similarly, the fraction $F_i = V_i/\epsilon^n$, where V_i is the volume occupied by the boxes surrounding the lattice points in $\mathcal{G}_i$. We can now exactly write

$$1-\rho_i = \frac{1}{2n}\frac{S_i/\epsilon^{n-1}}{V_i/\epsilon^n} = \frac{\epsilon}{2n}\frac{S_i}{V_i}, \tag{5.4}$$

or

$$\rho_i = 1 - \frac{\epsilon}{2n}\frac{S_i}{V_i}. \tag{5.5}$$

5.2.2 Limiting Cases

The n-dimensional object with the smallest surface area-to-volume ratio is the n-hypersphere, with volume $V_\circ(r;n) = r^n\pi^{n/2}/\Gamma(1+n/2)$ and surface area $S_\circ(r;n) = nr^{n-1}\pi^{n/2}/\Gamma(1+n/2)$, giving $S_\circ(r;n)/V_\circ(r;n) = n/r$ for radius r. As $r \to \infty$, $S_\circ(r;n)/V_\circ(r;n) \to 0$. For $r \gg \epsilon$ (the dense or "aggregated" limit, i.e. the low surface area-to-volume limit), we can write

$$\rho_i \approx 1 - \frac{\epsilon}{2r} \xrightarrow{r\gg\epsilon} 1. \tag{5.6}$$

In general this is true for any surface area-to-volume ratio that asymptotically approaches 0 for large volume.

In the opposite limit, suppose the volume $V_\circ(r;n)$ is fragmented into 1 (or more) spheres of *diameter* ϵ (as this is the lowest resolution permitted by our discretisation), we find that there are $V_\circ(r;n)/V(\epsilon/2;n)$ individual spheres that have surface area $S_\circ(\epsilon/2;n)$ each. The total surface area to volume ratio is therefore $S_\circ(\epsilon/2;n)/V_\circ(\epsilon/2;n) = 2n/\epsilon$. It now follow that in the "fragmented" limit,

$$\rho_i \approx 1 - \frac{\epsilon}{2n}\frac{2n}{\epsilon} = 0, \tag{5.7}$$

which is the expected behaviour.

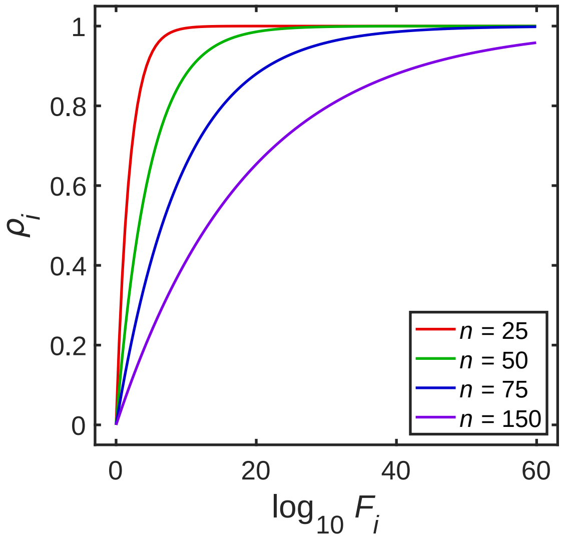


Figure 5.2: The power law upper bound for robustness is plotted against the log of the frequency F_i for various values of n.

Since we have a hypercubic lattice, it is perhaps more rigorous to look at the "fragmented" (i.e. large surface area-to-volume ratio limit) by considering a neutral set/function space that has been fragmented into n-dimensional hypercubes of side length ϵ, as this is the lowest resolution we can reach. The volume of each hypercube is $V_\square(\epsilon; n) = \epsilon^n$ and the surface area is $S_\square(\epsilon; n) = 2n\epsilon^{n-1}$. It follows that the surface area to volume ratio $S_\square(\epsilon; n)/V_\square(\epsilon; n) = S_\circ(\epsilon/2; n)/V_\circ(\epsilon/2; n) = 2n/\epsilon$, and eq. (5.7) is obtained for the robustness once again.

5.3 Power Law Theory for Robustness in Continuous Input Spaces

5.3.1 Upper Bound Derivation

We approximate the i-th output function space $\mathcal{G}_i$ as a prism with lengths $\{d_1\epsilon, \dots, d_n\epsilon\}$, with $d_k \in \mathbb{N}$ for all k. It follows that the (hyper)surface area is

$$S_i(d_1, \dots, d_n) = 2\epsilon^{n-1} \sum_{j=1}^{n} \frac{1}{d_j} \prod_{k=1}^{n} d_k. \tag{5.8}$$

The volume and frequency are given by

$$V_i(d_1, \ldots, d_n) = \epsilon^n \prod_{k=1}^{n} d_k, \quad F_i(d_1, \ldots, d_n) = \prod_{k=1}^{n} d_k \tag{5.9}$$

Thus, the robustness is

$$\rho_i = 1 - \frac{\epsilon}{2n} \frac{S_i(d_1, \ldots, d_n)}{V_i(d_1, \ldots, d_n)} = 1 - \frac{1}{n} \sum_{k=1}^{n} \frac{1}{d_k}. \tag{5.10}$$

The arithmetic mean of a set of positive numbers is always greater than or equal to the geometric mean, so

$$\frac{1}{n} \sum_{k=1}^{n} \frac{1}{d_k} \geq \left(\prod_{k=1}^{n} \frac{1}{d_k} \right)^{1/n} = \left(\prod_{k=1}^{n} d_k \right)^{-1/n} = \frac{1}{F_i^{1/n}}. \tag{5.11}$$

We can now bound the robustness from above as a function of the frequency:

$$\rho_i \leq 1 - \frac{1}{F_i^{1/n}}, \tag{5.12}$$

where equality holds if and only if $d_1 = d_2 = \cdots = d_n$. The limits $F_i \to 1$, $\rho_i \to 0$ and $F_i \to \infty$, $\rho_i \to 1$ hold as expected. The upper bound is plotted for various dimensions (number of inputs) n in fig. 5.2.

Trivially, the lower bound of the robustness is 0 for all F_i.

5.3.2 Effective Dimension Approximation (High and Intermediate Sensitivity Approximation)

We now assume that the effective dimension of the function space is less than n. In this approximation, we assume that our function space is a prism once again, with n_{eff} of the sides being length ϵ, and the remaining $n - n_{\text{eff}}$ being length $b\epsilon$, where b is some integer larger than 1. This would imply that there are n_{eff} axes/directions that are highly sensitive to small changes in the value of the input, since the width in those directions is only ϵ. The remaining $n - n_{\text{eff}}$ directions are of intermediate sensitivity if $b\epsilon \ll \Lambda$ or are not very sensitive at all if $b\epsilon \approx \Lambda$. The inspriration behind this splitting of the space into "stiff" and "sloppy" directions comes from the literature on sloppiness [83], which shows that for systems with any parameters, there are typically a few stiff directions, where small changes in the parameters

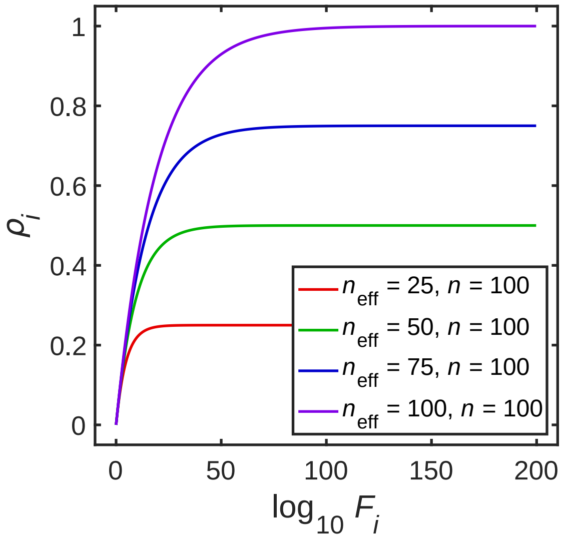


Figure 5.3: The robustness power law is shown for various effective dimensions n_{eff} of the i-th function space in an n-dimensional space.

lead to large changes in the outputs, and many sloppy directions, where changes in the parameters have little effect on the outputs.

The surface area of the neutral space prism is

$$S(b, n_{\text{eff}}) = 2\epsilon^{n-1}\left((n - n_{\text{eff}})b^{n_{\text{eff}}} + n_{\text{eff}}b^{n_{\text{eff}}-1}\right), \tag{5.13}$$

and the volume and frequency are

$$V(b, n_{\text{eff}}) = b^{n_{\text{eff}}}\epsilon^n, \quad F(b, n_{\text{eff}}) = b^{n_{\text{eff}}}. \tag{5.14}$$

Thus, the robustness is

$$\rho_i = 1 - \frac{\epsilon}{2n}\frac{S(a, n_{\text{eff}})}{V(a, n_{\text{eff}})} = \frac{n_{\text{eff}}}{n}\left(1 - \frac{1}{F_i^{1/n_{\text{eff}}}}\right). \tag{5.15}$$

Therefore, $n_{\text{eff}} < n$ places a ceiling on the maximum robustness; this behaviour is shown in fig. 5.3. In the limit $n_{\text{eff}} \to n$, this faithfully reproduces eq. (5.12) discussed previously.

5.3.3 Low and Intermediate Sensitivity Approximation

In this approximation, the prism has lengths larger than the ϵ length scale in all directions. However, m of these directions are taken to be unimportant or "sloppy," meaning that changes along those axes do not affect the output (very much). We assume that our function space is a prism once again, with m of the sides being length $a\epsilon$, and the remaining $n-m$ being length Λ, where $1 \leq a < \Lambda/\epsilon$. The m important directions are of intermediate or high sensitivity if $a\epsilon \ll \Lambda$ or very low sensitivity if $a\epsilon \approx \Lambda$.

The surface area of the neutral space prism is

$$S(a,m) = 2\left(m\Lambda^{n-m}(a\epsilon)^{m-1} + (n-m)\Lambda^{n-m-1}(a\epsilon)^{m}\right), \tag{5.16}$$

and the volume and frequency are

$$V(a,m) = (a\epsilon)^m\Lambda^{n-m}, \quad F(a,n_{\mathrm{eff}}) = a^m\left(\frac{\Lambda}{\epsilon}\right)^{n-m}. \tag{5.17}$$

Thus, the robustness is

$$\begin{aligned}\rho_i &= 1 - \frac{\epsilon}{2n}\frac{S(a,m)}{V(a,m)}\\ &= 1 - \frac{1}{n}\left(\frac{m}{a} + \frac{n-m}{(\Lambda/\epsilon)}\right) = 1 + \frac{\epsilon}{\Lambda}\left(1 - \frac{m}{n}\right) - \frac{m}{n}\frac{(\Lambda/\epsilon)^{\frac{n}{m}-1}}{F^{1/m}}.\end{aligned} \tag{5.18}$$

In the limit $a \to 1$, this approximation produces the same result as the Effective Dimension Approximation in the previous subsection for the case where $b = \Lambda$.

5.3.4 Anisotropic Prism Approximation

The previous two sets of approximations assumed only one free parameter. Here, we maintain the approximation that some axes are much stiffer than others, but we add an additional parameter. In this approximation, we examine the intermediate regime in which all of the dimensions are $\gg \epsilon$, but the function space is anisotropic such that there are m dimensions with length scale $a \gg 1$ that continue to grow when F_i increases, while there are $(n-m)$ dimensions of length b which are *fixed* and do not increase with increasing F_i (i.e. are approximately the same for different function spaces) but maintain $b\epsilon \gg \epsilon$. Fixing $(n-m)$ dimensions of length $b\epsilon$ is unrealistic as F_i becomes small, as we should have $F_i \to 1$. However, for intermediate and larger values of F_i, the derivations here should hold. This case is a generalisation of the previous section.

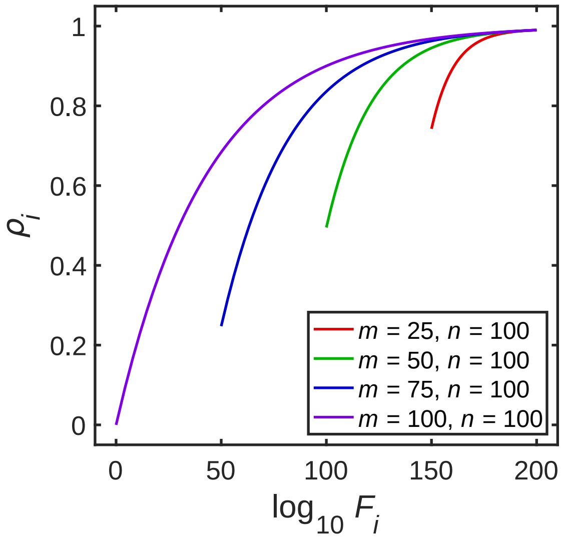


Figure 5.4: The robustness power law is shown for various numbers of important directions m in the low and intermediate sensitivity approximation. The number of bins is taken to be $\Lambda/\epsilon = 100$.

The surface area of the function space, approximated as a prism, should be

$$S(a,b,m) = 2\epsilon^{n-1}\left((n-m)a^m b^{n-m-1} + m a^{m-1} b^{n-m}\right), \tag{5.19}$$

and the volume and frequency are

$$V(a,b,m) = a^m b^{n-m}\epsilon^n, \quad F(a,b,m) = a^m b^{n-m}. \tag{5.20}$$

Thus, the robustness is

$$\rho_i = 1 - \frac{\epsilon}{2n}\frac{S(a,b,m)}{V(a,b,m)} = 1 - \frac{1}{n}\left(\frac{n-m}{b} + \frac{m}{a_i}\right) \tag{5.21}$$

We now discuss how this formulation behaves under various assumptions; using the assumption that b is fixed for different function spaces, we have that

$$\rho_i = 1 - \frac{n-m}{nb} - \frac{m b^{\frac{n}{m}-1}}{n F_i^{1/m}}. \tag{5.22}$$

With the additional assumption that $b \gg a_i \gg 1$ for all neutral spaces, we see that

$$\rho_i \approx 1 - \frac{m b^{\frac{n}{m}-1}}{n F_i^{1/m}} \quad \Rightarrow \quad \log(1-\rho_i) \sim \mathcal{O}(\log F_i) \tag{5.23}$$

Recall that this holds for sufficiently large F_i, so the robustness ρ_i does not appear to go to zero for $F_i \to 1$. In general, for sufficiently large F_i, this theory predicts power law behaviour (with an exponent that can deviate from $1/n$, as in the isotropic case discussed in the previous section) for the robustness as a function of the function space size.

5.3.5 Sampling in Numerical Simulations and Boundary Restriction

In numerical simulations of systems such as deep neural networks, spin glasses, or evolutionary fitness landscapes, sampling must of course be confined to a particular region of parameter space. This is reasonable for practical systems such as the one mentioned.

In our numerical studies, our parameters are all confined to the range, $[-\Lambda/2, \Lambda/2]$. This means that the box containing these sample points is bounded by the cube $[-(\Lambda-\epsilon)/2, (\Lambda-\epsilon)/2]^n$, so the volume contained is Λ^n, and the number of points in the sample space is $(\Lambda/\epsilon)^n$. Now, we suppose that all of the outputs available are confined to this box. This means that we are approximating F_i to be the global absolute frequency; this is not necessarily a valid assumption. However, even in real scenarios, the parameters are often restricted to some range (or at least practically confined to some range); this can be implemented in deep learning, for example, by adding a "regularisation" term to the loss function. As such, we can interpret F_i to be the number of inputs mapping to the i-th output which are *practically* accessible in a simulation (perhaps due to regularisation). The effect of sampling from within a confined region is that neutral sets will not necessarily be confined within the box in the sloppy directions (akin to principal components which are low rank).

We suppose that $f_i = F_i/(\Lambda/\epsilon)^n$ is the fraction of points mapping to the i-th output which are contained within the sample space, and suppose that N samples are taken. In the large N limit, we would expect that the theory in eq. (5.21) would be obtained, which is reprinted as first-order linear equation in $\log(1-\rho_i)$

and $\log f_i$, which makes least squares linear regression very easy:

$$\log(1-\rho_i) = \log\left(\frac{mb^{(n/m)-1}}{n(\Lambda/\epsilon)^{n/m}}\right) - \frac{\log f_i}{m} \tag{5.24}$$

We can fit sampled robustness and frequency data to fit b and m. We would expect that along unimportant directions the neutral spaces would extend out much farther than the sampling boundaries, so the value of b should be close to Λ if our simple approximation holds.

5.4 Numerical Methods

Collaborator Yoonsoo Nam [82] simulated small neural networks in order to sample various points within the discretised parameter space to calculate the robustness of each output space/neutral space. Each of these deep neural networks had 10 nodes in the input layer, 5 input nodes in each hidden layer, and 2 nodes in the output layer. Simulations were performed for 2, 3, and 4 hidden layers. The input space was taken to be restricted to the box $[-5,5]^n$, where n was the number of dimensions determined by the number of hidden layers. The 2 hidden layer system had $n = 67$ dimensions/independent parameters, the 3 hidden layer system had $n = 122$ dimensions, and the 4 hidden layer system had $n = 177$ dimensions. The input space $[-5,5]^n$ was discretised along each orthogonal axis into either 99, 199, or 399 bins.

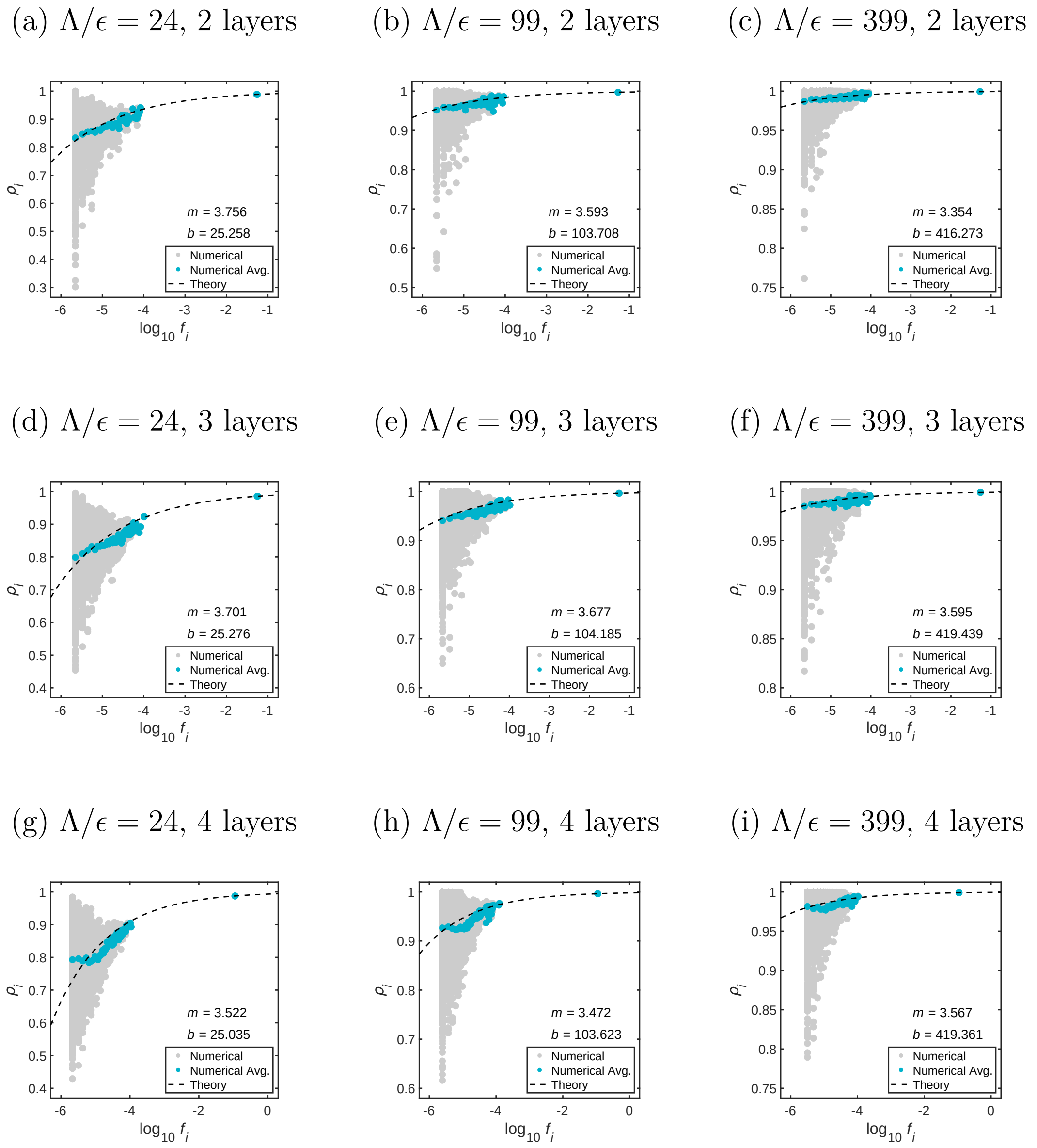
The input-output map was defined as follows: every point/bin in the hypercube lattice was taken to be an input with weights corresponding to that point in Euclidean space. The most probable DNN output from each of the 2^{10} binary inputs was appended to a list and assigned a label signifying a unique output. This means that two IO map outputs (phenotpyes) were the same if each and every one of the 2^{10} outputs mapped to exactly the same DNN outputs (from the neural network itself). Approximately 10^7 samples were taken for each choice of number of hidden layers and number of bins.

The author fit eq. (5.24) to Nam's [82] data using weighted least squares linear regression, using weights $w_i \propto f_i$, having approximated the standard deviations of the $1 - \rho_i$ as scaling as $\sigma_i \sim \mathcal{O}(1/\sqrt{(Nf_i)})$. In the dataset we kept only the outputs/phenotypes which appeared more than once.

5.5 Results and Discussion

In Figure 5.5 and Figure 5.6, we plot the robustness versus frequency as a linear-log plot and a log-log plot, respectively, from the numerical simulations. We additionally plot the Anisotropic Prism Approximation, eq. (5.24), with the parameters fitted to the data. It is clear, especially from the log-log plot, that the data support the power-law prediction introduced in this chapter. We find that for all combinations of architectures (number of hidden layers) and resolution of discretisation (number of bins), the value of b is within approximately 5% of Λ/ϵ. Despite using such a greatly simplified model—using only 2 parameters to specify the number of important/unimportant directions and the length of the neutral space in the unimportant directions—we still see surprisingly clear agreement with the approximation that $b \approx \Lambda/\epsilon$ in the unimportant directions. This would be, as mentioned prior, due to the fact that in the unimportant directions, the neutral space is likely to extend far beyond the sampling boundary imposed; the neutral set would get cut off at the boundary.

The value of m—the number of important directions—appears to stay between 3 and 4, only slightly decreasing as the dimension n increases. This suggests that, at the resolutions of discretisation that we have used in the simulation, we are only able to see approximately 3 or 4 major important directions. Future experiments that may further help elucidate the function of m would involve alterations of the DNN architecture. There is an interesting connection here between the number of "important directions" and the lottery ticket hypothesis by Frankle and Carbin [84] they show that DNNs often only need a small subset of the parameters to represent a solution. This suggests that only a low-dimensional manifold of the data is needed, and it may be that the dimension of “important” directions is related to the dimension of this manifold.



(a) $\Lambda/\epsilon = 24$, 2 layers

(b) $\Lambda/\epsilon = 99$, 2 layers

(c) $\Lambda/\epsilon = 399$, 2 layers

(d) $\Lambda/\epsilon = 24$, 3 layers

(e) $\Lambda/\epsilon = 99$, 3 layers

(f) $\Lambda/\epsilon = 399$, 3 layers

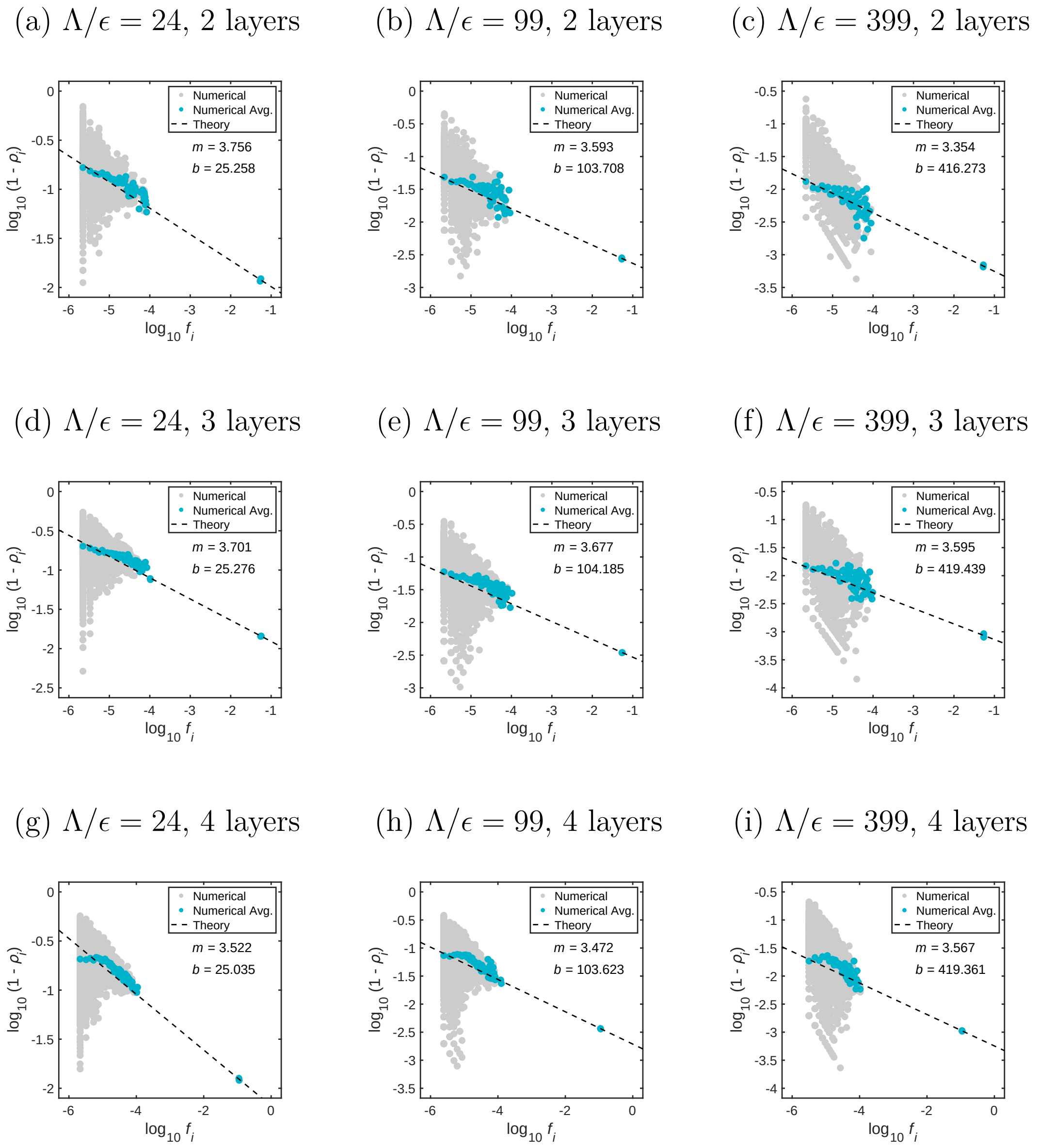
(g) $\Lambda/\epsilon = 24$, 4 layers

(h) $\Lambda/\epsilon = 99$, 4 layers

(i) $\Lambda/\epsilon = 399$, 4 layers

Figure 5.5: Linear-log plot of sampled robustness ρ_i versus frequency f_i for function spaces within deep neural network parameter space. Neural networks each have 10 input nodes, hidden layers (2, 3, or 4) with 5 nodes each, and 2 output nodes. Discretisation of the parameter space is chosen such that $\Lambda/\epsilon = 24, 99,$ or 399. Parameters range between -5 and 5. Theory (dotted) has been fit to the data (gray). Numerical averages (turquoise) have been shown for each unique frequency for convenience.

However, while the DNN input-output map is a particular example of a continuous input-output map we have chosen to study in this chapter, the central focus has been to expand the notion of robustness as it is defined in the biological GP map literature to continuous input-output maps and provide a prediction for the



(a) $\Lambda/\epsilon = 24$, 2 layers

(b) $\Lambda/\epsilon = 99$, 2 layers

(c) $\Lambda/\epsilon = 399$, 2 layers

(d) $\Lambda/\epsilon = 24$, 3 layers

(e) $\Lambda/\epsilon = 99$, 3 layers

(f) $\Lambda/\epsilon = 399$, 3 layers

(g) $\Lambda/\epsilon = 24$, 4 layers

(h) $\Lambda/\epsilon = 99$, 4 layers

(i) $\Lambda/\epsilon = 399$, 4 layers

Figure 5.6: Log-log plot of 1 minus sampled robustness $1 - \rho_i$ versus frequency f_i for function spaces within deep neural network parameter space. Neural networks each have 10 input nodes, hidden layers (2, 3, or 4) with 5 nodes each, and 2 output nodes. Discretisation of the parameter space is chosen such that $\Lambda/\epsilon = 24, 99,$ or 399. Parameters range between -5 and 5. Theory (dotted) has been fit to the data (gray). Numerical averages (turquoise) have been shown for each unique frequency.

naturally observed scaling law for robustness. To this end, the DNN data have shown excellent agreement with the robustness power law predicted by the simple prism-based approximations that have been made regarding neutral space geometry, which are based on the intuition of generic behaviour with a few stiff and many sloppy directions. The next steps will be to apply these theories to other continuous systems, such as the many models used in systems biology.

6
Dynamic Perturbation of Metastable Peaks on Glassy Fitness Landscapes

Contents

6.1 Introduction

Thus far, the discussion has been centered around the *static* properties of genotype-phenotype maps and input-output maps, namely in the context of network-theoretic properties that are relevant to evolution, such as mutational robustness. In this chapter, we investigate some basic questions around evolutionary *dynamics* on glassy landscapes.

Spin glasses have been used for decades to model molecular evolution. By thinking of spin configurations as molecular sequences and system energy as evolutionary fitness, we can draw a direct analogy between the spin glass model and molecular fitness landscapes. The dynamics of such a model then mimics evolution on a complex landscape. More recently, a growing number of researchers have been using spin glass-like models to infer fitness landscapes from sequencing data [37–40]. For the human immunodeficiency virus (HIV), inference of the fitness using the spin glass model has been successful in reproducing experimental replicative fitness measurements [37]. A major advantage of these models is that their inferred parameters offer direct interpretation. The spin-spin interaction coefficients, for example, can be directly interpreted as epistasis between two molecular sites. Moreover, it has been shown that higher-order spin-spin interactions are well-captured in the model despite that fact that the inferred fitness function only contains one-spin and two-spin terms [40].

In late 2019, the RNA virus SARS-CoV-2 began infecting humans, and the resulting COVID-19 pandemic—which continues at the time of writing of this thesis—has caused over 200 million infections and over 4.5 million deaths worldwide. Vaccines against the original human SARS-CoV-2 virus are being disseminated around the world; yet, the virus continues to mutate, and new dominating variants had already begun emerging before wide dissemination of the vaccine. Many of these variants are able to cause breakthrough infections even in vaccinated individuals. When a vaccine is administered in the host population (presumably on a faster timescale than it takes for infection to spread), hosts' immune systems tend to develop antibodies towards specific sites (epitopes) on the viral surface proteins. Immunity is also developed by unvaccinated, infected but recovered individuals as well. The applied immune pressure from the vaccines and recovered individuals effectively alters the shape of the fitness landscape for the virus, and a stable fitness peak in the original strain's fitness landscape may no longer be stable under immune pressure, as certain amino acids at the epitopes may now face negative selection pressure. Viruses may then "escape" to find strains which are stable under the new host immunity environment. It has thereby become important to ask, how does a population that has found a stable fitness peak evolve once that fitness landscape has been perturbed, as happens for example with a vaccine?

One way of approaching this problem using a mathematical framework is to consider a spin glass landscape where we can easily perturb the landscape. Ideally, we would know the exact intrinsic fitness landscape for a particular virus and simulate evolutionary dynamics and immune pressure changes on the landscape. But that problem remains quite hard, and since we are interested in generic effects, in this chapter, we study a schematic spin glass-like fitness landscape in which inferred terms in the model are drawn from a known Gaussian distribution.

Using numerical simulation, we consider a population that has reached a local fitness maximum (which we will interchangeably call a *metastable state*)[1]. We then perturb the stability of that peak by altering the fitness function, which looks like a spin glass Hamiltonian with one-site "external field" terms denoted by h_i for the i-th site and two-site "interaction" terms denoted by J_{ij} for the interaction between sites i and j, and we examine the evolutionary behaviour of the population until it finds another metastable state. In this chapter,

1. We show that the difference in fitness $\Delta F = F_{\text{new}} - F_{\text{initial}}$ between the initial and new peak depends on the variance σ_h^2 of the one-site h terms and the variance σ_J^2 of the two-site J terms in the fitness function. Our data suggest that while F_{new} and F_{initial} both only depend on $\sigma_h^2 + \sigma_J^2/2$, ΔF has a nontrivial and entirely different dependence on the variances. Specifically ΔF, appears to be larger when either of the variances is large and smaller when the two variances are closer in value to each other.

2. We examine the distribution of ΔF over many evolutionary trajectories and show that it is unimodal with a long right tail.

[1] In spin glass physics, from which this language emerges, a "stable" state often refers to a global energy minimum while a "metastable" peak or state refers to a local energy minimum. We adopt this language in the discussion of evolutionary fitness landscapes which are modelled as the *negative* of spin glass energy landscapes. So, *metastable states* in our evolutionary context are referring to local fitness maxima. Metastability of a particular viral sequence indicates that any single change to that sequence will result in a worse/lower fitness.

3. We consider the time interval τ between the fitness peak perturbation and the discovery of the new stable fitness peak and describe its dependence on the variances of the one-site and two-site terms. Unlike ΔF, τ depends nontrivially on the one-site variance σ_h^2 and the two-site variance σ_J^2 terms; we plot the dependencies graphically.

4. Our data strongly show τ is exponentially distributed, suggesting that finding a new fitness peak after perturbation may be akin to a Poisson process.

5. Lastly, we examine the relationship between the change in fitness ΔF and the time taken to discover the new peak τ and find that the correlation between these two variables depends on the variance in the one-site and two-site terms in the fitness function. As the variance in the one-site term increases, the Pearson correlation between $\log \Delta F$ and $\log \tau$ *decreases*, and as the variance in the two-site term increases, the Pearson correlation between $\log \Delta F$ and $\log \tau$ *increases*. The two-site term represents epistasis between sites in the viral genome, and this suggests that increase in the variance in epistasis makes the fitness trajectory between the perturbed initial peak and the newly discovered stable peak more power-law-like (in the sense that the correlation $\log \Delta F$ and $\log \tau$ is linear and closer to 1).

These findings open the door to the possibility of studying evolution on more realistic landscapes, including those derived from SARS-CoV-2 (and other viral) prevalence data, perhaps in the context of trajectories of viral escape after vaccination of a host population.

6.2 Inferring Fitness Landscapes

We consider a viral genome of length L in which every site in the genome takes on one of two possible values (this is a simplification to the binary case from 4 nucleotides or 20 amino acids). Such a simplification can be made, for example, by coding the most prevalent residue at a particular site in the genome as +1 and the clumping the remaining residues together and coding them as -1. In the inference process of viral fitness landscapes, one assumes that the fitness landscape can be modelled as a spin glass Hamiltonian (times -1):

$$F(\mathbf{s}) = \frac{1}{2}\sum_{i\neq j} J_{ij}s_is_j + \sum_i h_is_i, \tag{6.1}$$

where the sequence $\mathbf{s} \in \{-1,+1\}^L$. Here, the J_{ij} and h_i parameters are meant to be inferred from the true prevalence data. The h_i parameter (called the "external field" in spin glass physics) for each site i is related to intrinsic preference for a particular spin state at site i; that is to say, at site i, if site $s_i = 1$ is leads to higher fitness, then we would expect h_i to have large magnitude and be greater than zero. An h_i that is negative with large magnitude would indicate that $s_i = -1$ contributes to higher fitness, and and h_i with small magnitude indicates that it does not really matter whether $s_i = 1$ or $s_i = -1$. The J_{ij} term is an "interaction" between sites i and j and represents the *epistasis* between those two sites. *Positive sign epistasis* ($J_{ij} > 0$) indicates that $s_i = s_j$ will provide a higher fitness contribution, and *negative sign epistasis* ($J_{ij} < 0$) indicates that $s_i = -s_j$ will provide a higher fitness contribution.

It is assumed that the theoretical prevalence $P_{th}(\mathbf{s})$ of each sequence $\mathbf{s}$ ("viral strain") is given by

$$P_{th}(\mathbf{s}) = \frac{e^{F(s)}}{\mathrm{tr}_{\mathbf{s}}\, e^{F(s)}}, \tag{6.2}$$

where $\mathrm{tr}_{\mathbf{s}}$ is the sum over all spin configurations. Various methods can be used to minimise the KL divergence $KL(P_{th}||P_{real})$ between the theoretical prevalence landscape P_{th} and the empirical prevalence landscape P_{real} [37]. One of the most used methods is the adaptive cluster expansion (ACE) method. This method performs the convex optimisation task for gradually larger clusters of spins/sites at a time. After doing so, the inferred fitness landscapes, despite only containing up to two-spin correlation information (via the J_{ij} couplings), tend to agree well at higher orders as well [40]. This has been used to infer Boltzmann machines [85], to solve the inverse Ising problem [86], to infer lattice protein models [87], to infer protein contacts [42], to infer neuronal connections from brain activity

[88], and to infer the fitness landscapes of human immunodeficiency virus (HIV) by using prevalence data of those viruses [37–40].

Fitness landscapes inferred from prevalence data (called prevalence landscapes) do not give the full picture of the true replicative fitness (called the intrinsic fitness landscape) for that particular virus. Incompleteness of prevalence data certainly biases the calculation; moreover, unfit strains will have low or zero prevalence and therefore will not contribute to the inference of low-fitness regions in the landscape. From eq. (6.2) we see that, in a prevalence landscape, the most prevalent viral strain will necessarily be the most fit; therefore, the most prevalent viral strain will be at the global maximum of the prevalence landscape. In the true intrinsic fitness landscape, a viral population will not necessarily have reached the global maximum; the global maximum in the prevalence landscape could actually just be a local maximum in the intrinsic fitness landscape. Thus, there is an important distinction between the prevalence landscape inferred from prevalence data and the intrinsic fitness landscape which requires replication experiments. Nonetheless, Ferguson et al. [89] report excellent correlation between inferred prevalance landscapes and *in vivo* experiments, suggesting that the prevalence landscape may be a fair proxy for the true intrinsic fitness landscape.

In this chapter, we are concerned with the evolutionary behaviour of a viral population that has reached a local fitness peak which is then perturbed. We work with idealised spin glass fitness landscapes which have Gaussian distributed couplings and field terms. This is quite a simplification, but the analytical calculations below cannot at present be done on inferred prevalence landscapes nor on intrinsic replicative fitness landscapes from *in vivo* experiments. We hope that the general principles will still be valid, and that in the future numerical calculations may be extended to empirical landscapes.

6.3 Static Properties of Spin Glass Fitness Landscapes

Let us consider the spin glass fitness function in eq. (6.1) which defines the fitness landscape for our problem:

$$F(\mathbf{s}) = \frac{1}{2}\sum_{i\neq j} J_{ij}s_i s_j + \sum_i h_i s_i. \tag{6.3}$$

Recall that $\mathbf{s}$ is a particular (viral) genomic sequence, the h_i and J_{ij} variables parameterise the landscape with one-site (self) and two-site (interaction) terms, respectively. In our calculations, we make the simplifying assumption that the parameters are all Gaussian, so $J_{ij} \sim \mathcal{N}(0, \sigma_J^2/L)$ (symmetric with $J_{ii} = 0$) and $h_i \sim \mathcal{N}(0, \sigma_h^2)$. We now discuss static properties of the fitness landscape which will be relevant to numerical simulation.

The *density of states* $\eta(F)$ as a function of the fitness F is the fraction of the total number of sequences $\mathbf{s}$ with fitness F, and the *density of metastable states* $\eta_m(F)$ is the fraction of the total number of sequences which have fitness F and are at a local fitness maxima. The density of metastable states has been calculated in the absence of a Gaussian external field h_i and in the presence of a constant external field in refs. [90–93], but there does not appear to be any exact calculation of the density of metastable states in the presence of a one-site external field term h_i drawn from a Gaussian distribution, which is relevant to our case. In the following sections, we follow the derivations of Tanaka and Edwards [93] by working in the annealed approximation, which is when the disorder in the random h_i and J_{ij} variables is averaged over directly. The authors of refs. [90, 91] argue that such averaging is less physical than averaging over the logarithm of any observable variables (e.g. such as the expected number of metastable states which we will calculate later). However, they also found that the quenched and annealed approximations yielded identical results above the glass transition energy. In the evolutionary sense, this means that below a certain fitness value, these two approximations should be identical; though they may not be above a certain critical fitness cutoff. We believe our evolutionary adaptation likely takes beneath the glassy "summit" of the fitness mountain—outside of the glassy regime (but still in the presence of metastability)—so in this section we follow the calculations of Tanaka and Edwards [93] having included the additional complication of a Gaussian external field to the model.

6.3.1 Density of States

For a sequence of length L with fitness F given by eq. (6.3), we first define the *intensive fitness* (fitness normalised by length) $\widetilde{F} = F/L$. The number of states which have fitness equal to $\widetilde{F}$ is denoted as $\Omega(\widetilde{F})$, and the *density of states* $\eta(\widetilde{F})$ at a particular $\widetilde{F}$ is the number of states normalised by the total number of sequences possible (2^L):

$$\eta(\widetilde{F}) = \frac{\Omega(\widetilde{F})}{2^L}. \tag{6.4}$$

First we show that the total density of states is Gaussian (as shown in [36]):

$$\eta(\widetilde{F}) = \frac{1}{2^L}\,\mathrm{tr}_{\mathbf{s}} \left\langle \delta\left(L\widetilde{F} - \sum_{\langle i,j\rangle} J_{ij}s_is_j - \sum_i h_is_i \right) \right\rangle_{J,h}. \tag{6.5}$$

This is difficult to solve exactly, but if we take the Fourier transform of the expression within the average $\langle\cdot\rangle$, then it becomes easily tractable:

$$\begin{aligned} \eta(F) &= \frac{1}{2^L}\,\mathrm{tr}_{\mathbf{s}} \left\langle \int_{-\infty}^{\infty} \frac{\mathrm{d}\theta}{2\pi} \exp\left[-\imath\theta \left(L\widetilde{F} - \sum_{\langle i,j\rangle} J_{ij}s_is_j - \sum_i h_is_i \right) \right] \right\rangle_{J,h} \\ &= \int_{-\infty}^{\infty} \frac{\mathrm{d}\theta}{2\pi} \exp\left[\left(-\imath L\theta\widetilde{F} - \frac{L\theta^2\sigma_J^2}{4} - \frac{L\theta^2\sigma_h^2}{2} \right) \right] = \frac{1}{\sqrt{2\pi\sigma_F^2}} e^{-\frac{F^2}{2\sigma_F^2}}, \end{aligned} \tag{6.6}$$

where the approximation $L - 1 \approx L$ has been used. This is a Gaussian distribution with zero mean and with variance $\sigma_F^2 = L\left(\frac{\sigma_J^2}{2} + \sigma_h^2\right)$. The intensive fitness $\widetilde{F} = F/L$ also obeys a Gaussian distribution with zero mean and with variance $\sigma_{\widetilde{F}}^2 = \frac{1}{L}\left(\frac{\sigma_J^2}{2} + \sigma_h^2\right)$.

If the J_{ij} terms were to disappear, this would physically correspond to the case where there are no epistatic interactions between any two genomic sites s_i and s_j. The fitness landscape in this case has no ruggedness, and there is a single fitness peak which is also the global maximum. The fitness function would only depend on the one-site, external field terms h_i, and it is possible to find the exact fitness of the single global fitness peak fitness peak. As $\sigma_J \to 0$, the J_{ij} terms disappear, and we find

$$\max_{\mathbf{s}} \left\langle \lim_{\sigma_J \to 0} F(\mathbf{s}) \right\rangle_h = \max_{\mathbf{s}} \sum_i h_i s_i = \sum_i \langle |h_i| \rangle_h = L\sqrt{\frac{2}{\pi}} \simeq 0.79L. \tag{6.7}$$

So, the global fitness maximum on this landscape has expected fitness $\simeq 0.79L$. This value would be expected to increase as interactions J_{ij} are included.

6.3.2 Average Number of Metastable States

Finding the average number of local fitness maxima is the same as finding the average number of local fitness minima of the spin glass Hamiltonian

$$-F(\mathbf{s}) = -\frac{1}{2}\sum_i \sum_{j \neq i} J_{ij} s_i s_j - \sum_i h_i s_i = H_{\text{spin glass}}(\mathbf{s}) \tag{6.8}$$

For this reason, we will be referring to the i-th element of the sequence $\mathbf{s}$ as the "spin" s_i, which may be "up" ($s_i = +1$) or "down" ($s_i = -1$). Flipping the i-th spin results in a fitness change ΔF_i given by

$$\frac{-\Delta F_i}{2} = x_i = s_i \left(\sum_j J_{ij} s_j + h_i \right). \tag{6.9}$$

The stability of a spin configuration requires that $x_i > 0$ for all i. It has been shown in [90] that, for the SK model, averaging the observables (annealed approximation) yield the same results as averaging the logarithm of the observable (quenched approximation—the "physical" case) below a certain fitness threshold when off-diagonal elements in replica space vanish. The joint probability density function for the x_i is given by

$$\eta_m(\mathbf{x}) = \frac{1}{2^L} \operatorname{tr}_{\mathbf{s}} \left\langle \prod_i \delta \left(x_i - \sum_j J_{ij} s_i s_j - h_i s_i \right) \right\rangle_{J,h,b}, \tag{6.10}$$

where $\langle \cdot \rangle_{J,h,b}$ is the average over the quenched disorder, and the trace is performed over all spin configurations. Noting that $\langle J_{ij} \rangle_J = 0$, $\eta_m(\mathbf{x})$ is invariant under the gauge transformation $J_{ij} \mapsto J_{ij} s_i s_j$, we can write

$$\eta_m(\mathbf{x}) = \frac{1}{2^L} \operatorname{tr}_{\mathbf{s}} \left\langle \prod_i \delta \left(x_i - \sum_j J_{ij} - h_i s_i \right) \right\rangle_{J,h}. \tag{6.11}$$

We use the Fourier representation of the delta functions to average over the energies

$$\eta_m(\mathbf{x}) = \frac{1}{2^L}\prod_i \int_{-\infty}^{\infty} \frac{\mathrm{d}\phi_i}{2\pi} e^{\imath\phi_i x_i}\, \mathrm{tr}_{s_i} \left\langle \exp\left(-\imath\phi_i \sum_j J_{ij} - \imath\phi_i h_i s_i\right)\right\rangle_{J,h} . \tag{6.12}$$

Performing the J average, we have

$$\begin{aligned} \left\langle \exp\left[-\imath\sum_{i\neq j} J_{ij}\phi_i\right]\right\rangle_J &= \left\langle \exp\left[-\imath\sum_{\langle i,j\rangle} J_{ij}(\phi_i+\phi_j)\right]\right\rangle_J \\ &= \exp\left[-\frac{\sigma_J^2}{2L}\sum_{\langle i,j\rangle}(\phi_i+\phi_j)^2\right], \end{aligned} \tag{6.13}$$

which can be expanded as

$$\exp\left[-\frac{\sigma_J^2}{L}\sum_i \phi_i^2 - \frac{\sigma_J^2}{L}\sum_{\langle i,j\rangle}\phi_i\phi_j\right] \approx \exp\left[-\frac{\sigma_J^2}{2}\sum_i \phi_i^2 - \frac{\sigma_J^2}{2L}\left(\sum_i \phi_i\right)^2\right]. \tag{6.14}$$

This simplifies to a form in which there are no products $\phi_i\phi_j$:

$$\sqrt{\frac{L}{2\pi\sigma_J^2}}\int_{-\infty}^{\infty} \mathrm{d}z \exp\left[-\frac{Lz^2}{2\sigma_J^2}\right]\prod_i \exp\left[-\frac{\sigma_J^2}{2}\phi_i^2 + \imath z\phi_i\right]. \tag{6.15}$$

Above, we have assumed that $L-1\approx L$ for large L, and additionally that the $\mathcal{O}(1)$ term is much smaller than the $\mathcal{O}(L)$ terms in the exponents . Performing the h and b averages, we have

$$\langle \mathrm{tr}_{\mathbf{s}_i} \exp\left[-\imath\phi_i h_i s_i\right]\rangle_h = 2e^{-\frac{\sigma_h^2\phi_i^2}{2}}. \tag{6.16}$$

It now follows that

$$\begin{aligned} \eta_m(\mathbf{x}) &= \sqrt{\frac{L}{2\pi\sigma_J^2}}\int_{-\infty}^{\infty} \mathrm{d}z\, e^{-\frac{Lz^2}{2\sigma_J^2}}\prod_i \int_{-\infty}^{\infty}\frac{\mathrm{d}\phi_i}{2\pi} e^{-\frac{(\sigma_J^2+\sigma_h^2)\phi_i^2}{2}+\imath(x_i+z)\phi_i} \\ &= \sqrt{\frac{L}{2\pi\sigma_J^2}}\int_{-\infty}^{\infty} \mathrm{d}z\, e^{-\frac{Lz^2}{2\sigma_J^2}}\prod_i\left[\frac{1}{\sqrt{2\pi(\sigma_J^2+\sigma_h^2)}}e^{\frac{-(x_i+z)^2}{2(\sigma_J^2+\sigma_h^2)}}\right]. \end{aligned} \tag{6.17}$$

The expected number of metastable states $\langle g_0\rangle$ is given by

$$\langle g_0\rangle = 2^L\int_0^{\infty}\mathrm{d}^L\mathbf{x}\,\eta_m(\mathbf{x}) = \sqrt{\frac{L}{2\pi\sigma_J^2}}\int_{-\infty}^{\infty}\mathrm{d}z\, e^{LS_0(z)}, \tag{6.18}$$

where

$$S_0(z) = -\frac{z^2}{2\sigma_J^2} + \log\operatorname{erfc}\left(\frac{z}{\sqrt{2(\sigma_J^2+\sigma_h^2)}}\right). \tag{6.19}$$

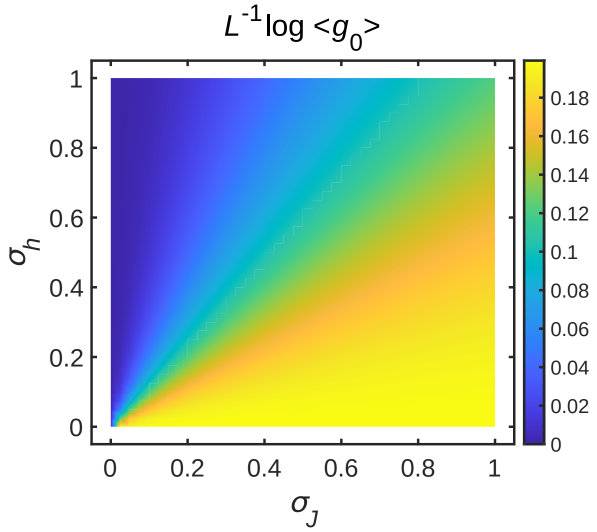


Figure 6.1: Log of $\langle g_0 \rangle$, the expected number of metastable states from theory, as a function of σ_J and σ_h. We can see that when $\sigma_h \gg \sigma_J$, the number of metastable states in the fitness landscape decreases and tends to 1. This is the regime in which the external field terms dominate, and ruggedness of the landscape disappears, leaving only one global fitness peak, indicated by the blue-most regions. Likewise, in the limit $\sigma_J \gg \sigma_h$, the landscape is maximally rugged, and there is high metastable peak density, indicated by the yellow-most regions.

Using Laplace's method for large L, we can approximate

$$\langle g_0 \rangle \approx e^{LS_0(z^*)}, \tag{6.20}$$

where z^* is the solution to the fixed-point equation

$$\frac{\partial S_0(z)}{\partial z} = 0. \tag{6.21}$$

The number of metastable states $\langle g_0 \rangle$ is plotted as a function of σ_J and σ_h in Figure 6.1.

6.3.3 Density of Metastable States

From the Hamiltonian (equivalently the negative fitness) in equation eq. (6.8), we can find the distribution over the fitnesses of the local fitness maxima:

$$\eta_m(F) = \int_0^\infty \mathrm{d}^L\mathbf{x}\,\mathrm{tr}_{\mathbf{s}} \left\langle \frac{\delta\left(F - \sum_{\langle i,j\rangle} J_{ij}s_is_j - \sum_i h_is_i\right)\prod_i \delta\left(x_i - \sum_j J_{ij} - h_is_i\right)}{g_0(\{J,h\})} \right\rangle_{J,h}. \tag{6.22}$$

We make the approximation that $g_0(\{J, h\})$ can be replaced by $\langle g_0 \rangle$ as calculated in the previous section. Defining the number of metastable states as a function of the fitness, $\Omega_m(F) = \eta_m(F) \langle g_0 \rangle$, we have

$$\Omega_m(F) = \int_0^\infty \mathrm{d}^L\mathbf{x}\, \mathrm{tr}_{\mathbf{s}} \left\langle \delta\left(F - \sum_{\langle i,j\rangle} J_{ij}s_is_j - \sum_i h_is_i\right) \prod_i \delta\left(x_i - \sum_j J_{ij} - h_is_i\right)\right\rangle_{J,h} . \tag{6.23}$$

Using the Fourier representations of the delta functions, we find

$$\begin{aligned}\Omega_m(F) = \int_0^\infty \mathrm{d}^L\mathbf{x} \int_{-\infty}^\infty \frac{\mathrm{d}\theta}{2\pi} \int_{-\infty}^\infty \frac{\mathrm{d}^L\boldsymbol{\phi}}{(2\pi)^L} e^{\imath L\theta\widetilde{F} + \imath\phi_i x_i} \\ \times\, \mathrm{tr}_{\mathbf{s}} \left\langle \exp\left[-\imath \sum_{\langle i,j\rangle} J_{ij}(\theta + \phi_i + \phi_j) - \imath \sum_i h_i(\theta + \phi_i)\right]\right\rangle_{J,h} ,\end{aligned} \tag{6.24}$$

which becomes

$$\begin{aligned}= 2^L \int_0^\infty \mathrm{d}^L\mathbf{x} \int_{-\infty}^\infty \frac{\mathrm{d}\theta}{2\pi} \frac{\mathrm{d}^L\boldsymbol{\phi}}{(2\pi)^L} e^{\imath L\theta\widetilde{F} + \imath\phi_i x_i} \\ \times \exp\left[-\frac{\sigma_J^2}{2L} \sum_{\langle i,j\rangle} (\theta + \phi_i + \phi_j)^2 - \frac{\sigma_h^2}{2} \sum_i h_i(\theta + \phi_i)\right] .\end{aligned} \tag{6.25}$$

This can be rewritten as

$$\begin{aligned}= 2^L \int_{-\infty}^\infty \frac{\mathrm{d}z}{\sqrt{2\pi\sigma_J^2/L}} \int_0^\infty \mathrm{d}^L\mathbf{x} \int_{-\infty}^\infty \frac{\mathrm{d}\theta}{2\pi} e^{\imath L\theta\widetilde{F} - \frac{Lz^2}{2\sigma_J^2} - \frac{L\theta^2}{2}\left(\frac{\sigma_J^2}{2} + \sigma_h^2\right)} \\ \times \int_{-\infty}^\infty \frac{\mathrm{d}^L\boldsymbol{\phi}}{(2\pi)^L} \prod_i \exp\left[\imath\phi_i(x_i + z) - \frac{(\sigma_J^2 + \sigma_h^2)\phi_i^2}{2} - \theta\phi_i(\sigma_J^2 + \sigma_h^2)\right] .\end{aligned} \tag{6.26}$$

This expression now becomes

$$\begin{aligned}= 2^L \int_{-\infty}^\infty \frac{\mathrm{d}z}{\sqrt{2\pi\sigma_J^2/L}} \int_{-\infty}^\infty \frac{\mathrm{d}\theta}{2\pi} \int_0^\infty \mathrm{d}^L\mathbf{x}\, e^{\imath L\theta\widetilde{F} - \frac{Lz^2}{2\sigma_J^2} - \frac{L\theta^2}{2}\left(\frac{\sigma_J^2}{2} + \sigma_h^2\right)} \\ \times \prod_i \frac{1}{\sqrt{2\pi(\sigma_J^2 + \sigma_h^2)}} e^{-\frac{(x_i + z + \imath\theta(\sigma_J^2 + \sigma_h^2))^2}{2(\sigma_J^2 + \sigma_h^2)}} ,\end{aligned} \tag{6.27}$$

which finally equals

$$= \int_{-\infty}^\infty \frac{\mathrm{d}z}{\sqrt{2\pi\sigma_J^2/L}} \int_{-\infty}^\infty \frac{\mathrm{d}\theta}{2\pi} e^{\imath L\theta\widetilde{F} - \frac{Lz^2}{2\sigma_J^2} - \frac{L\theta^2}{2}\left(\frac{\sigma_J^2}{2} + \sigma_h^2\right)} \left[\mathrm{erfc}\left(\frac{z + \imath\theta(\sigma_J^2 + \sigma_h^2)}{\sqrt{2(\sigma_J^2 + \sigma_h^2)}}\right)\right]^L . \tag{6.28}$$

Performing the substitution $z \mapsto z - \imath\theta(\sigma_J^2 + \sigma_h^2)$ followed by $\theta \mapsto \imath\theta$, we have that

$$\Omega_m(F) = \int_{-\infty}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi\sigma_J^2/L}} \int_{-\infty}^{\infty} \frac{\mathrm{d}\theta}{2\pi} e^{LS_1(z,\theta)}, \tag{6.29}$$

with—in the limit of large L—the saddle point action

$$S_1(z,\theta) = \widetilde{F}\theta + \frac{\theta^2}{2}\left(\frac{\sigma_J^2}{2} + \sigma_h^2\right) - \frac{(z - \theta(\sigma_J^2 + \sigma_h^2))^2}{2\sigma_J^2} + \log\operatorname{erfc}\left(\frac{z}{\sqrt{2(\sigma_J^2 + \sigma_h^2)}}\right). \tag{6.30}$$

One of the saddle point equations can be analytically solved to yield a solution:

$$\frac{\partial S_1(z,\theta)}{\partial\theta} = 0 \quad \Rightarrow \quad \theta^* = \frac{2(z(\sigma_J^2 + \sigma_h^2) + \widetilde{F}\sigma_J^2)}{\sigma_J^4 + 2\sigma_J^2\sigma_h^2 + 2\sigma_h^4}. \tag{6.31}$$

The saddle point action $S_2(z) = S_1(z,\theta^*)$ is now updated:

$$S_2(z) = \frac{2z(2\widetilde{F} + z)\sigma_h^2 + (2\widetilde{F}^2 + 4\widetilde{F}z + z^2)\sigma_J^2}{2(\sigma_J^4 + 2\sigma_J^2\sigma_h^2 + 2\sigma_h^4)} + \log\operatorname{erfc}\left(\frac{z}{\sqrt{2(\sigma_J^2 + \sigma_h^2)}}\right). \tag{6.32}$$

The number of metastable states as a function of the (intensive) fitness is therefore

$$\Omega_m(F) \approx e^{LS_2(z^*)}. \tag{6.33}$$

where z^* is the solution to the saddle point equation

$$\frac{\partial S_2(z)}{\partial z} = 0. \tag{6.34}$$

In Figure 6.2, we plot the density of states and density of metastable states for various values of α, which is a parameter defined such that:

$$\sigma_J = \sqrt{2}\cos\left(\frac{\pi\alpha}{2}\right), \quad \sigma_h = \sin\left(\frac{\pi\alpha}{2}\right). \tag{6.35}$$

Note that α has been defined such that the density of states defined in eq. (6.6) is independent of α:

$$\frac{1}{\sqrt{2\pi\sigma_F^2}} e^{-\frac{F^2}{2\sigma_F^2}} = \frac{1}{\sqrt{2\pi L}} e^{-\frac{F^2}{2L}}, \tag{6.36}$$

since

$$\sigma_F^2 \equiv L\left(\frac{\sigma_J^2}{2} + \sigma_h^2\right) = L\left[\cos^2\left(\frac{\pi\alpha}{2}\right) + \sin^2\left(\frac{\pi\alpha}{2}\right)\right] = L. \tag{6.37}$$

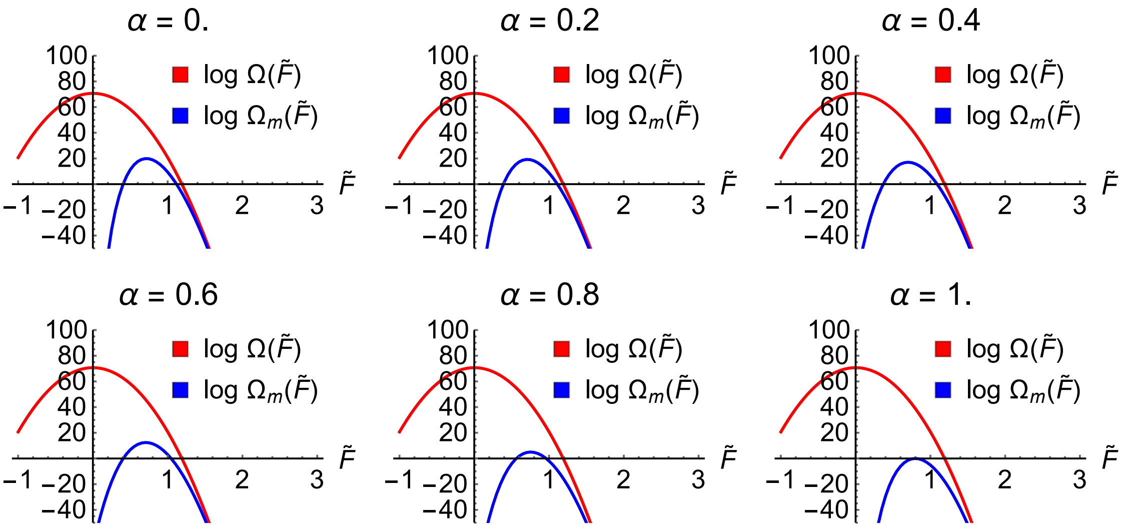


Figure 6.2: Plot of (blue) the log number of total configurations $\log \Omega(\widetilde{F})$ that have an intensive fitness $\widetilde{F}$ (as determined from the density of states in eq. (6.6)) and (red) the log number of metastable states $\Omega_m(\widetilde{F})$ as a function of intensive fitness as determined from the metastable density of states in eq. (6.33). As α, defined in eq. (6.35), approaches 1, we approach the limit where $\sigma_h \to 1$ and $\sigma_J \to 0$, in which case the epistatic/interaction J_{ij} terms do not matter, leaving only a single fitness peak. This is seen for the blue curves, which continue to narrow as α increases, eventually reaching a point where $\Omega(\widetilde{F}) = 1$ (or $\log \Omega(\widetilde{F}) = 0$) only at a single point, with $\log \Omega(\widetilde{F}) < 0$ for all other $\widetilde{F}$. The plots were made with $L = 100$.

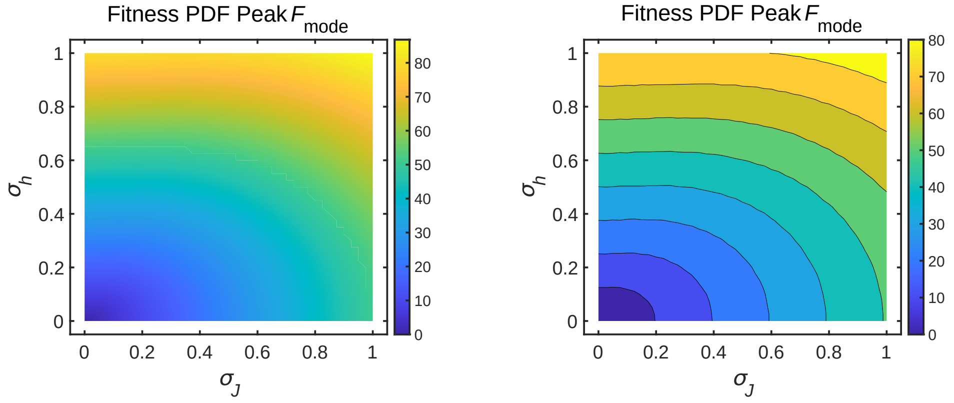


Figure 6.3: Numerically calculated fitness values F_{mode} at the mode of the distribution of metastable states calculated from numerical optimisation of the density of metastable states eq. (6.33). Left and right plots show identical data, but the left plot shows a smooth plot of all the F_{mode} values while the right plot bins the F_{mode} values and colours each bin differently so that the general shape of the contour lines separating the bins are clearly denoted.

From the plots, we see that increasing α (which increases σ_J relative to σ_h) does not change the total number of states, but the distribution of metastable states tends to become narrower. In the limit as $\alpha \to 1$, then we have $\sigma_J \to 0$ and $\sigma_h \to 1$, which is the limit discussed earlier in which the epistasis/interaction terms J_{ij} all vanish, and we are left with one unique fitness peak. Thus, we see that $\log \Omega_m(\widetilde{F})$, the log of the number of metastable states, is equal to 0 at only one point (and all other values are negative); this indicates that there is a single unique fitness peak.

These plots actually break down (due to the annealed approximation) for sufficiently large fitness, but this should not matter for our numerical simulations, as most of our salient numerical data will be collected for fitness values sufficiently low that we have metastability but not breakdown of the annealed approximation.

In Figure 6.3, we plot the position of the peak F_{mode} of the distribution of metastable states. The contours in Figure 6.3 appear to be distorted ellipses. It is important to note that F_{mode} is not dependent only on $\frac{\sigma_J^2}{2} + \sigma_h$ in the way that the density of states $\Omega(\widetilde{F})$ is; this shows that changing α not only narrows the density of metastable states but also shifts the peak of the metatstable state density slightly.

6.4 Evolutionary Dynamics Simulation

During the course of the COVID-19 pandemic, the development and administration of vaccines to SARS-CoV-2 has led to the broad-scale immunisation of the world's population. Although the vaccines have been successful in attenuating the spread of the virus, the emergence of novel variants of SARS-CoV-2 has proven to be an emerging challenge that will warrant the development of new vaccines to combat the pathogen's evolving spike protein structure. Undoubtedly, the administration of the vaccine exerts selection pressure, causing an effective change in the fitness landscape for the virus. We are interested in the evolutionary dynamics of a viral population that has reached a local fitness maximum (a metastable state), only to have the metastability of that local peak perturbed by an external interaction, similar to the way a rapidly disseminated vaccine in a host population would affect the effective fitness landscape of the viral population. In particular, we are interested in how much time it takes for the viral population to find a new metastable peak after the external perturbation and the change in the fitness from one peak to another. We now present a numerical simulation based on methods used in ref. [94], which have also been used by the authors of [40].

Consider a population of $N = 100$ viruses with genome of length L. The genome of the n-th virus is a sequence $\mathbf{s}^{(n)} = (s_0^{(n)}, \dots, s_{L-1}^{(n)})$, where $s_i^{(n)} \in \{\pm 1\}$. The population of viruses evolves on the fitness landscape defined earlier:

$$F(\mathbf{s}^{(n)}) = \frac{1}{2}\sum_{i\neq j} J_{ij} s_i^{(n)} s_j^{(n)} + \sum_i h_i s_i^{(n)}. \tag{6.38}$$

In each generation, each viral sequence n survives to reproduce with probability

$$P_{\text{surv}}(n) = \frac{e^{-\beta(\overline{F}-F_n)}}{1+e^{-\beta(\overline{F}-F_n)}}, \tag{6.39}$$

where F_n is the fitness of the n-th organism, $\overline{F}$ is the mean fitness of the population, and β is a tunable parameter that determines how strong the penalty is for deviating below the mean fitness or how advantageous the reward is for deviating above the mean (we set $\beta = 1$ for our simulations, by the convention in ref. [40]). The next generation is chosen by sampling without replacement from the survivors of the previous generation to keep the population fixed at N. At the beginning of the next time step, each organism can mutate. The site mutation probability per generation is taken to be μ; this is arbitrarily set to 0.1 in our simulations, which is fairly high, but—as argued by Shekhar et al. [40]—each “generation” in the simulation actually represents multiple cycles of replication, and selection operates on a much slower timescale, so the effective number of mutations seen per generation may be modelled appropriately with a higher mutation rate. Our choice of $\mu = 0.1$ per site per generation is high for an RNA virus (which is typically three or four orders of magnitude lower), but we would not expect the qualitative features of the dynamics here to change.

In Figure 6.2, we showed that the density of states is approximately symmetric above and below fitness $F = 0$ (i.e. the density of states is approximately an even function). However, the metastable states, which form a subset of the total

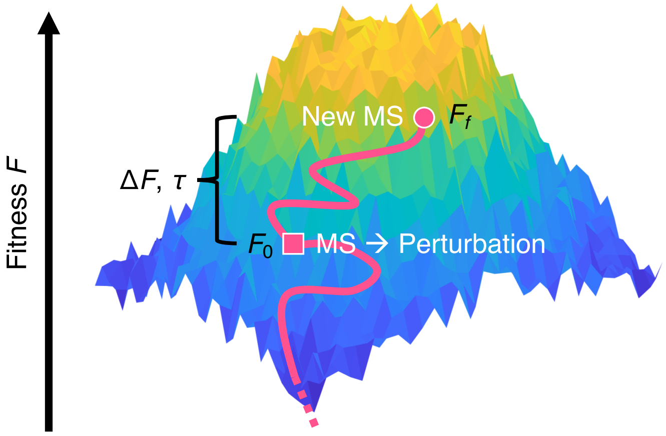


Figure 6.4: Schematic representation of the numerical simulation. On a spin glass fitness landscape, a population evolves until it reaches a metastable state (local fitness maximum). Then, an external perturbation is applied to the h_i magnetic field terms such that the current state is no longer metastable. The population then continues to evolve until it reaches a new metastable state. MS = metastable state, F_0 = fitness of first metastable state encountered, F_f = fitness of metastable state encountered after perturbation, $\Delta F = F_f - F_0$ = the difference in fitness between the two metastable states, τ = the time between the perturbation and the encountering of a new metastable state.

set of states, only tend to occur for positive fitnesses. We therefore initialise our simulations well below the "base" of the mountain of metastable states—i.e. they have fitnesses F approximately equal to or less than zero. When the mutation-selection cycle begins, there is an initial burn-in period during which fitness tends to increase without encountering any metastable states until all at once it reaches a fitness value at which the density of metastable states is no longer trivially small.

Eventually, at least 50% of the population reaches the metastable state. This is the "first" metastable state encountered. We then apply a non-adiabatic perturbation such that the metastable state is destabilised (i.e. the local fitness maximum is no longer a local fitness maximum). This is our schematic model of external immune pressure (such as from a vaccine). Inspired by Shekhar et al. [40], this is done by applying an external field b_i to $\eta = 10\%$ of the sites (representing targeting of epitopes on viral surface proteins), so

$$F(\mathbf{s}^{(n)}) \mapsto F(\mathbf{s}^{(n)}) - \sum_i b_i s_i^{(n)}. \tag{6.40}$$

We specifically choose $b_i = -ah_i$, where $a = 2$ initially so that the external field just switches sign $h_i \mapsto h_i - b_i = -h_i$. For many cases, this perturbation is enough to destabilise the metastable state. However, there are often cases where the local fitness maximum encountered is very sharp; it needs a much stronger perturbation to be destabilised. If our initial perturbation is not substantial enough to destabilise the metastable state, we increment a from 2 to 3 (and later 3 to 4, 4 to 5, and so on, up until a cutoff) until we find some a for which the applied perturbation is strong enough to destabilise the metastable state.

After the perturbation has been applied, the population continues to evolve until a new metastable state is encountered (and at least 50% of the population reaches that state). The process then may repeat, with perturbations being applied each time a new metastable state is found. A schematic for the entire simulation is shown in Figure 6.4.

An effect of the application of the (potentially strong) perturbations is that the total density of states may distort, as the h_i's may no longer follow the original Gaussian distribution. This effect is more likely to be prominent as fitness increases (and local fitness maxima become sharper), so studying the early regime of the fitness trajectory (the trajectory between the first few low-lying metastable states) is most convincing. Here, we focus our attention on the dynamics between the very first metastable state encountered, the first perturbation applied to this metastable state, and then the next most-immediate metastable state encountered. Although we also do look at the overall fitness growth trajectory, this first transition is the only one in which we can guarantee that multiple trials can be conducted with random initialisations with low lying fitness and still ensure that the initial fitness is roughly the same value. At these low metastable peaks, the destabilising perturbation is usually found with the the default value of a, the total density of states is not distorted, and we obtain interesting dynamics with consistent distributions for observables of interest.

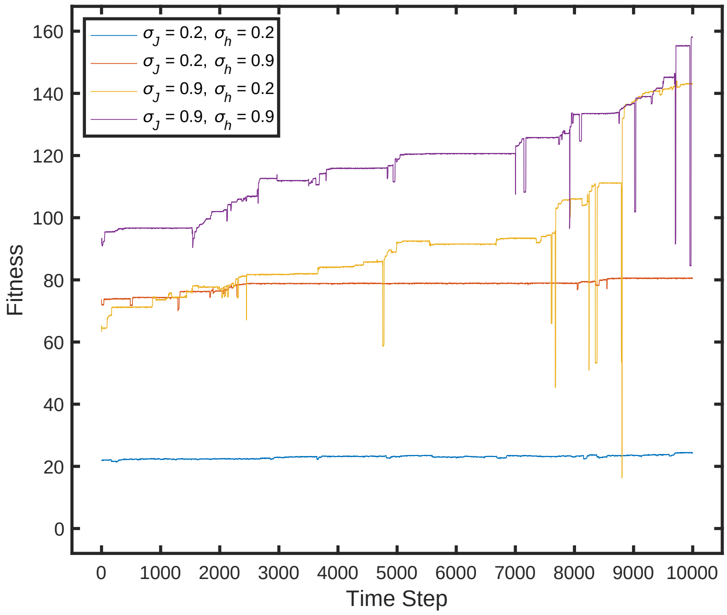


Figure 6.5: For pairs (σ_J, σ_h), we plot sample long term fitness trajectories. The large dips correspond to perturbations at metastable states which had to be increased in magnitude in order to effect a destabilisation of the metastable states. It appears that large σ_J and σ_h resulted in the discovery of deep glassy local maxima which are difficult to perturb.

6.5 Numerical Results

Examples of long term fitness growth trajectories during repeated process of dynamic perturbation of metastable states reached are shown in Figure 6.5. We note that the fitness does tend to exceed what would be expected from the density of metatstable state predictions because the dynamic perturbations that must be applied at higher fitnesses tends to distort the Gaussian distribution of h_i and increases its standard deviation beyond σ_h.

Thus, we focus the remainder of the discussion on the behaviour of the population between the first and second metastable states encountered, since this region appears to be representative of most other inter-metastable state transitions that occur. Firstly, the population is initalised well below the metastable region and evolves up until a metastable state is encountered. The fitness of the initial state encountered after the simulation begins below the metstable region is plotted in the left panel

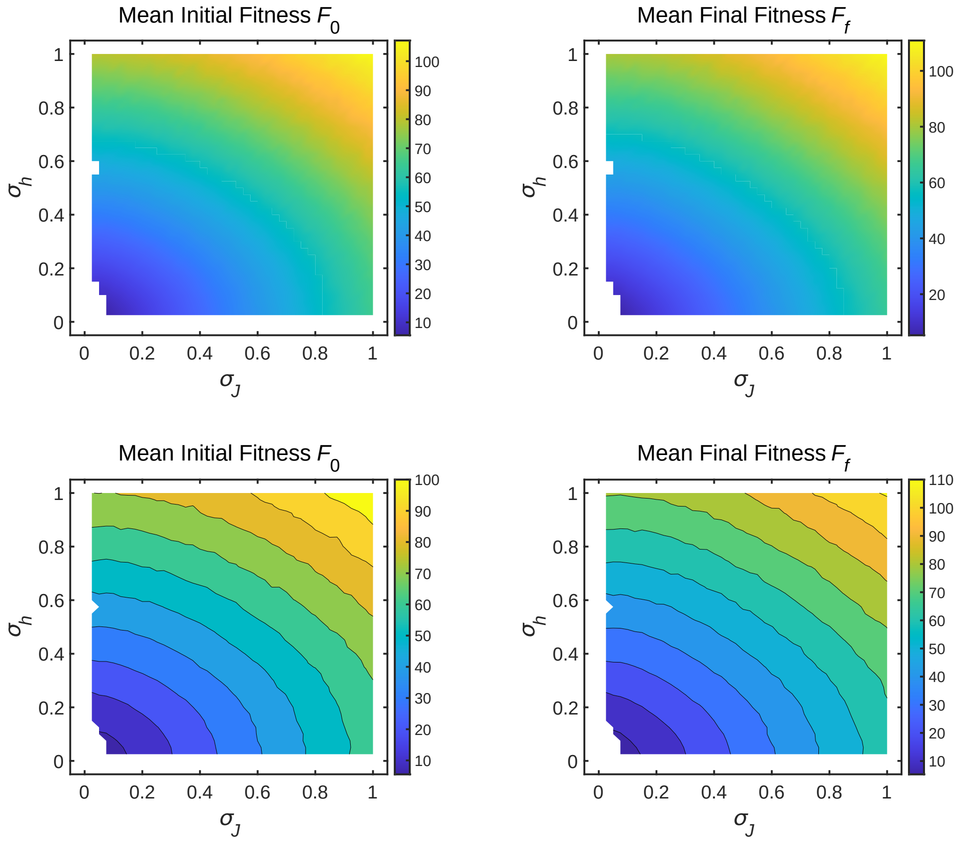


Figure 6.6: Plots of the initial (left; bottom and top) and final (right; bottom and top) value of the fitness at which the simulation begins. Lines indicate contours of constant fitness. This corresponds to (left; bottom and top) the first metastable state that the population finds while evolving from below the metastable state threshold and (right; bottom and top) the next metastable state that is found after the perturbation destabilises the first one. Bottom and top plots show identical data; lines of constant fitness are easier to distinguish on the contour plots on the bottom.

in Figure 6.6. After perturbing the landscape externally, the population then evolves again until a new metastable state is encountered. This fitness is plotted in the right panel of Figure 6.6.

As we would expect, the initial fitness distribution has radial symmetry (with eccentricity since we plot σ_J instead of $\sigma_J/\sqrt{2}$). This means that the epistatic variance σ_J and the onsite variance σ_h contribute approximately equally to the overall fitness growth below the metastable region; this is expected. Similarly, after perturbation of the metastable state and encountering of a new metastable state, the final fitness also has the same symmetry. Thus far in the simulation, σ_J and

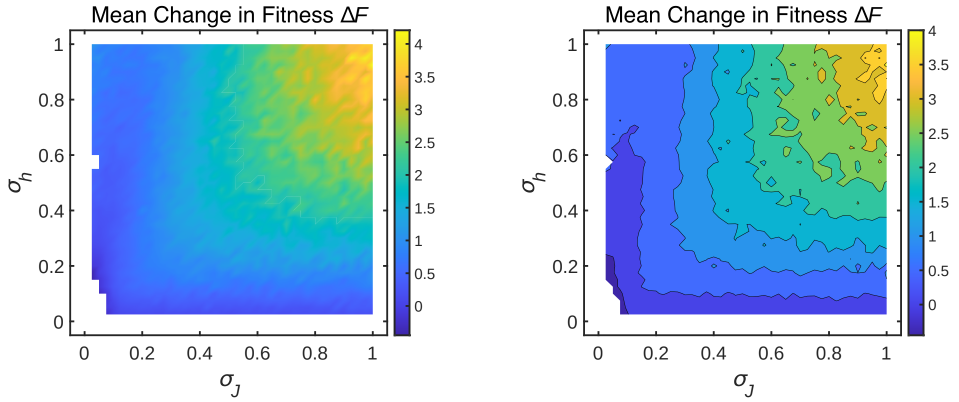


Figure 6.7: Plots of the change in fitness $\Delta F = F_f - F_0$ experienced between the transition between the first and second metastable states encountered. Increasing either σ_h or σ_J seems to have similar effects (up to a scale factor). Left and right plots show identical data; lines of constant fitness are easier to distinguish on the contour plots on the right. Notably, the contour lines do not show the same shapes as fig. 6.6.

σ_h have symmetrically affected the population fitness.

The difference in fitness ΔF between the initial F_0 and final F_f metastable states is plotted in Figure 6.7. The influence of σ_h and σ_J appears to be symmetric (approximately) once again, but it is clear that the radial symmetry in the graph is now broken; the contours of constant fitness no longer look like ellipses. This means that $\frac{\sigma_J^2}{2} + \sigma_h^2$ would not produce the same value of ΔF for all combinations of (σ_J, σ_h). Still, increasing either σ_J and σ_h increases ΔF. The contours now appear to have behaviour similar to $\sigma_J \sigma_h$ = constant.

We now bring to attention the interesting result that the *recovery time* taken for a population to reach another metastable state is *not* symmetrically affected by σ_J and σ_h. Plotted in Figure 6.8, these contours now curve such that there appears to be an optimal value of σ_h for a given σ_J that maximises the recovery time. Moreover, increasing σ_J quickly lowers the recovery time, but increasing σ_h does not do so as quickly. The likely explanation for this is that increasing σ_J relative to σ_h increases the density and overall number of metastable states, as shown previously in Section 6.3.2; this would increase the chance that a metastable state is found sooner.

We now look at the actual distribution of recovery times for various values of σ_J and σ_h in Figure 6.9, taken over several trials. Strikingly, these distributions

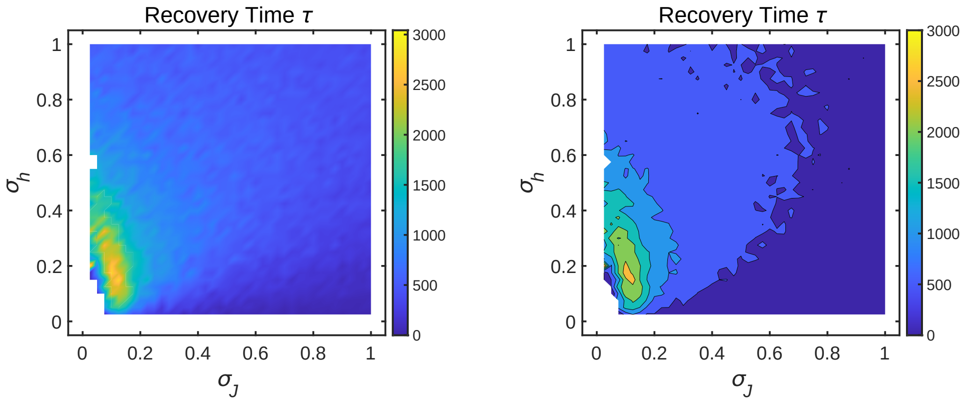


Figure 6.8: Plots of the time τ taken for metastability to be recovered by 50% of the population after perturbation of the first encountered metastable state. There is clear asymmetry in the effects of increasing σ_J versus σ_h. Increasing σ_J tends to create far more metastable states within the same range of fitnesses, so the time taken to discover a new metastable state is much smaller relative to an equivalent increase in σ_h.Left and right plots show identical data; lines of constant fitness are easier to distinguish on the contour plots on the right. Notably, the contour lines do not show the same shapes as Figure 6.6, nor do they match with the shapes in Figure 6.7. Time units are measured in generations.

seem to be very well-fit by exponential probability distributions. This suggests that we can think of metastable state-finding as a Poisson process, since Poisson process time intervals are distributed exponentially.

These distributions differ from the distributions of ΔF, shown in Figure 6.10. The ΔF distributions appear to have a single peak close to zero (but positive). A few values showed changes in $\Delta F < 0$, but these trials were considered anomalous and omitted from these results since we are examining trajectories for positive fitness growth. These distributions are also right-tailed, but the lack of an exponential shape suggests that the relationship between recovery time and change in fitness is not exactly one-to-one, which was already suggested by the differences in the shapes of the contours for these plots.

To better examine the relationship between τ and ΔF, we plot each trial's ΔF against τ in the log-log plot in Figure 6.11. According to the arguments of Good and Desai [95] and Guo and Amir [36], we would expect fitness trajectories to follow a power-law relationship in the strong selection, weak mutation regime (SSWM), meaning that ΔF and τ would also be related by a power law. However, because of

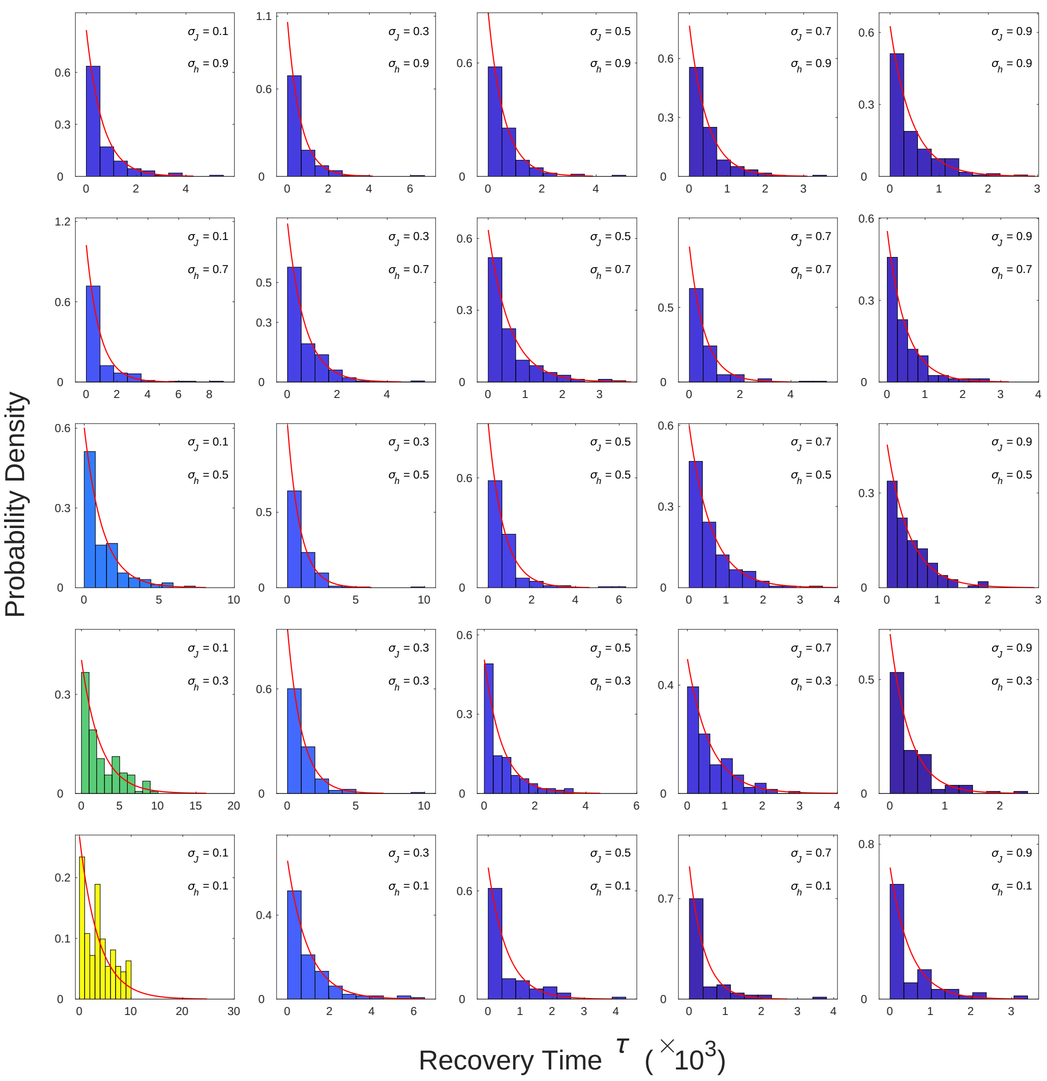


Figure 6.9: Histograms of the times τ taken for metastability to be recovered by 50% of the population after perturbation of the first encountered metastable state. The exponential distribution suggests that finding a metastable state after perturbation of the original metastable state is a Poisson process, or can be approximated as such. Plots for intermediate values of σ_J and σ_h are found in Appendix D.

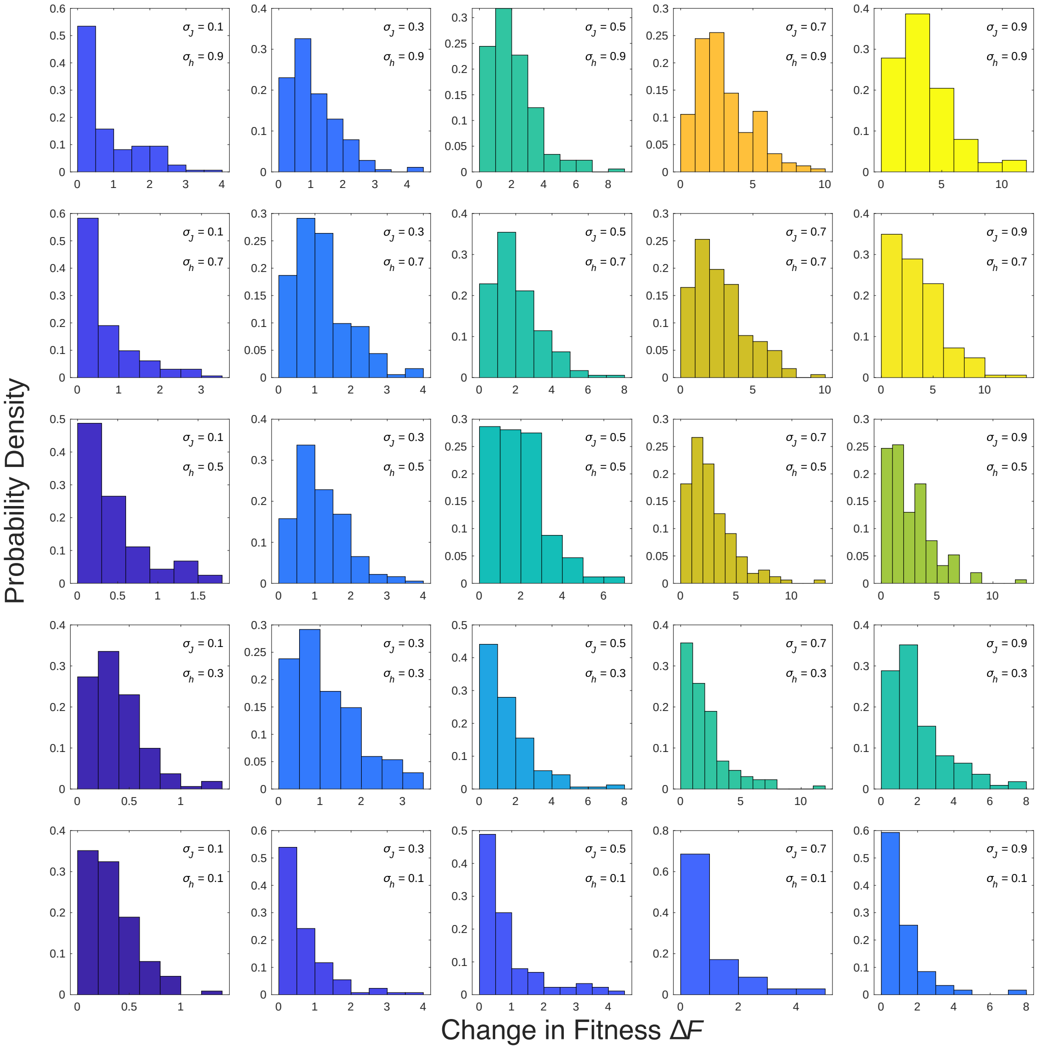


Figure 6.10: Histograms of the changes in fitness $\Delta F = F_f - F_0$ experienced between the transition between the first and second metastable states encountered. The distributions tend to be right skewed and unimodal near, but not at $\Delta F = 0$, in general. Plots for intermediate values of σ_J and σ_h are found in Appendix D.

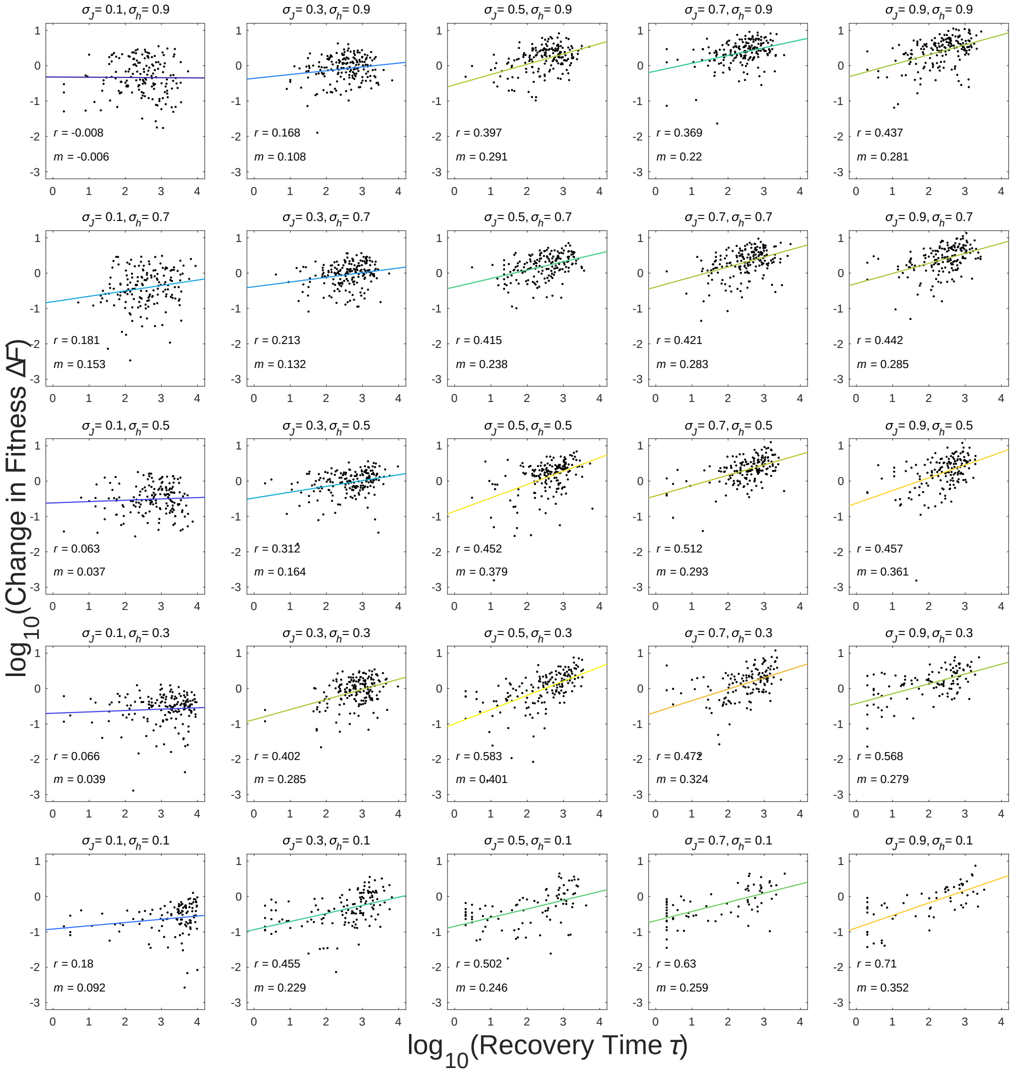


Figure 6.11: Correlations between log change in fitness $\log \Delta F$ and time taken to recover metastability $\log \tau$ following a perturbation appear to be linear, supporting the idea that the fitness trajectory is approximately a power law, as suggested by [95]. In general, the correlation tends to increase as σ_h gets smaller and as σ_J gets larger. Plots for intermediate values of σ_J and σ_h are found in Appendix D.

our numerical procedure and the choice of $\beta = 1$ for the simulation, we are not in the SSWM. Nonetheless, in the proper limits of σ_h and σ_J, we do indeed see a power-law relationship between ΔF and τ, since there are positive linear correlations between $\log \tau$ and $\log \Delta F$ (though this cannot be used to conclude whether the entire trajectory follows a power law). The plot in Figure 6.11 reveals that the correlation is generally positive, but it decreases with *decreasing* σ_J and *increasing* σ_h.

6.6 Discussion

Motivated by the use of inferred spin glass Hamiltonians as fitness functions for viral evolution, we proposed a schematic spin glass fitness landscape with Gaussian interactions J_{ij} with mean zero and variance σ_J and external fields h_i with mean zero and variance σ_h. First, we investigated the density of local fitness maxima, also known as metastable states, by following and expanding the theoretical findings of Tanaka and Edwards [93], who had examined the system without any Gaussian external field. We found the relationship between the two variances σ_J and σ_h is responsible for modulating many of the static properties of the total density of states as well as the density of metastable states in this schematic fitness landscape.

The dissemination of vaccines for COVID-19 and the emergence of viral variants calls for a better understanding between the viral population dynamics and externally applied immune pressure. We utilised our a schematic fitness landscape to understand the behaviour of an evolving viral population which has reached a local fitness peak after the metastability of that peak has been perturbed by an external force. Although the evolutionary dynamics during stochastic tunnelling between local fitness peaks had been studied by Guo and Amir [36] in the strong selection weak mutation regime, a general numerical study incorporating mutation-selection dynamics governed by quasispecies interactions outside of this regime had not been conducted. Inspired by Shekhar et al. [40], by incorporating additional external field terms b_i (which flip the sign of the original external fields h_i), we model the presence of immune pressure, which make certain residues for the virus unfavorable on their epitopes, the sites targeted by the immune system.

We studied studied the post-perturbation dynamics of the evolving viral population up until the point at which it discovers and fixes at a new local fitness maximum. Our focus lay in low-lying local fitness maxima so that we could reproducibly conduct several numerical trials while maintaining initial fitness values.

Our data suggest that while the change in fitness experienced during the transition to a new metastable state after an old one has been destabilised is equally influenced by σ_J and σ_h, the contour surfaces take on a nontrivial shape that is not radially symmetric. This is unexpected, because the fitness values at the initial metastable peak and the final metastable peak both have radially symmetric dependence on σ_J and σ_h.

The time taken for metastability to be recovered τ is, furthermore, asymmetrically affected by the two variances, with σ_J being more responsible for rapidly decreasing τ. The recovery time is exponentially distributed, suggesting that a Poisson process may serve as a good model for metastable state finding. Lastly, the power-law prediction by Good and Desai [95] from the SSWM approximation holds reasonably well in particular limits of σ_J and σ_h, but changes in σ_J and σ_h asymmetrically affect the correlation between $\log \Delta F$ and $\log \tau$. These are fundamental population dynamic insights which may contribute to a better understanding of evolution after a rapid, non-adiabatic fitness-reducing event, like when certain traits become unfavorable due to an external event. We hope that our numerical simulations can eventually be run on intrinsic fitness landscapes constructed from *in vivo* replication experiments or at least on prevalence landscapes inferred from viral prevalence data.

7 Conclusion

Input-output maps which take in a sequential input and produce some nontrivial output are ubiquitous in science. One of their most interesting properties are those of robustness: how likely is it that a perturbation to an input does not change the output?

Here, we have advanced the understanding of robustness in evolutionary and glassy systems in several ways. In Chapter 2, we turned to spin glasses, which are widely used in many fields of science to model densely interacting phenomena, including evolutionary fitness landscapes. We showed that a spin glass input-output map mapping from a set of interactions to a ground state obeys the same scaling laws as previously examined biological and computer science systems. The spin glass model we defined is one of the simplest models of complex interacting components which optimise a cost function. Many natural systems can be viewed this way, so this suggests a wide applicability of observed robustness (and other relevant) scaling. It also further motivates investigation into the more universal origins of robustness.

In Chapter 3, we explored this more universal origin of robustness and other network topological properties of input-output maps. In the literature and in Chapter 2, many of these network topological properties are treated separately from each other, with their scaling laws characterised independently of each other. We have adopted a new perspective, hypothesising only a single constraint on a parameter we call global robustness is needed to not only reproduce robustness scaling laws for each individual output but also for other scaling laws observed for transition probabilities, neutral component sizes, and largest neutral component size for each output. The constraint is placed on a maximum entropy model taken over the probability distribution of all input-output maps for fixed output frequencies. The tunable "temperature" parameter in the model can be swept to show that there exists a phase transition between a robust and fragile phase which respectively reproduce the natural and random GP map scaling laws observed in the literature. This suggests that, if the maximum entropy assumption holds, nature seems to have two options—out of what may otherwise have appeared to be arbitrarily many options—for scaling laws for these network-topological parameters including robustness. Moreover, real biological systems tend to be found in or near the maximally robust phase.

In Chapter 4, we then characterised the robust phase by examining maximally robust neutral networks known as bricklayer's graphs. The exact bound on the maximum possible robustness is made fully rigorous here for the first time, and the connections between biological input-output maps and the sums-of-digits function studied in number theory are elucidated. After proving the generalisation of an inequality regarding the sums-of-digits function, we calculate a realistic tight bound on the deviation of real natural phenotypes from the maximal bricklayer's bound. We then explore information content of neutral networks, proving an early argument that Hamming graphs optimise information content, disproving the claim that robustness of Hamming graphs and bricklayer's graphs generally are the same, and rigorously showing that a class of special, nontrivial cases exist for which robustness does match. The latter is accompanied by the analytical solution to a special case for the sums-of-digits function which is novel, to the author's knowledge. We then discuss population neutrality and show its lower bound for bricklayer's graph, and in conjunction with a collaborator's upper bound, show that it is tightly bounded, and that the error between population neutrality and robustness for a bricklayer's graph is small. We then discuss Hamming spheres, which have been proven to optimise the population neutrality for binary input sequences. We calculate their exact robustness, which is of course lower than that of the bricklayer's graphs. This suggests that nature generally faces optimisation trade-offs between robustness and information content (for neutral network sizes in which both Hamming graphs and bricklayer's graphs can exist) as well as between robustness and population neutrality. We then introduce a theory of phenotype coarse-graining, providing analytical formulas derived from graph theory for how robustness and transition probabilities should change under the coarse-graining process. This suggests an explanation for the recent, unexpected numerical results from a collaborator's RNA secondary structure coarse-graining simulations.

In Chapter 5, we extend the notion of robustness for the first time from input-output maps with discrete inputs to input-output maps with continuous inputs. We show how discretisation of the continuous input space and application of the graph-theoretic definition of robustness generates a novel expected scaling law that is not dependent on the discretisation itself. Unlike in the discrete systems, where the natural scaling law for robustness is logarithmic in frequency, for the continuous input-output maps we predict a power law. We then calculate a series of increasingly parameterised Ansätze for the shape of neutral spaces, providing intuition regarding how informative and uninformative parameter axes affect the geometry of the neutral spaces. We fit our prediction to a collaborator's numerical data from deep neural network simulations in which robustness was sampled; the data support our prediction of power law robustness behaviour.

In Chapter 6, we returned to spin glasses as in Chapter 2; but, motivated by the COVID-19 pandemic, we study a problem in viral evolution, namely evolutionary adaptation at short timescales following the non-adiabatic perturbation of a metastable state. We found that the variances in the epistatis/coupling terms and the onsite/field terms symmetrically impact the fitness trajectory near the beginning of the evolutionary trajectory, but asymetrically affect the timescales at which new metastable states are discovered. Our data show clearly that the distribution of mestastable state recovery times is exponential, suggesting that discovery of a metstable state after an external perturbation has disturbed metastability can be treated as a Poisson process. Lastly, we characterised the correlations between change in fitness and metastable state recovery time, showing that the correlation resulting from a power law relationship decreases with decreasing epistasis and/or with increasing variance in the onsite/field term.

In this thesis, we have advanced the understanding of natural robustness scaling laws via graph-theoretic and maximum entropy approaches and have provided exploratory steps in extending this notion to new continuous frameworks. We have also explored glassy evolution models from the perspective of robustness and in terms of adaptive dynamics. We hope that a deeper understanding of neutrality and mutational robustness of evolutionary systems will provide new avenues for predicting biological evolution, considering how important this has become during the COVID-19 pandemic.

Appendices

A
Cluster Expansion for $q = 2$ (Boolean/Ising Outputs)

For the Hamiltonian given in eq. (3.15), we consider the case where $q = 2$; i.e. the outputs are Boolean, so $\Omega(x) \in \{0, 1\}$. This is relevant to functions and systems encountered in computer science as well as some evolutionary systems like Gavrilets' "holey fitness landscape" [96]. The classical Potts model maps onto an Ising model. We perform the switch to Ising spins $\sigma_x = 2\Omega(x) - 1 \in \{-1, 1\}$. Now, the Hamiltonian can be rewritten as

$$\begin{aligned}\mathcal{H}(\boldsymbol{\sigma}) &= -\frac{2}{k^d d(k-1)} \sum_{\{i,j\}\in E} \frac{\sigma_i \sigma_j + 1}{2} \\ &= -\frac{1}{2} - \frac{1}{k^d d(k-1)} \sum_{\{i,j\}\in E} \sigma_i \sigma_j .\end{aligned} \tag{A.1}$$

Now, note that $\mathbf{f} = (f_0, f_1)$ and $f_0 + f_1 = 1$, so the partition function is entirely defined by the quantity $M = f_1 - f_0$, which is the magnetisation of the Ising model. The partition function is now

$$Z(M) = e^{\frac{1}{2T}} \sum_{\boldsymbol{\sigma}\,|\,M} \prod_{\{i,j\}\in E} e^{\frac{1}{Tk^d d(k-1)}\sigma_i \sigma_j}. \tag{A.2}$$

Making the replacement

$$e^{\frac{1}{Tk^d d(k-1)}\sigma_i \sigma_j} = \gamma\left(1 + \tau \sigma_i \sigma_j\right) \tag{A.3}$$

where

$$\gamma = \cosh\left(\frac{1}{Tk^d d(k-1)}\right), \quad \tau = \tanh\left(\frac{1}{Tk^d d(k-1)}\right). \tag{A.4}$$

The partition function can be rewritten

$$Z(M) = e^{\frac{1}{2T}} \gamma^{\frac{k^d d(k-1)}{2}} \sum_{\boldsymbol{\sigma}\,|\,M} \prod_{\{i,j\}\in E} (1 + \tau\sigma_i\sigma_j). \tag{A.5}$$

Now, τ will serve as the series expansion variable. After expanding the product, the partition function becomes a sum of terms proportional to

$$\sum_{\boldsymbol{\sigma}\,|\,M} \sigma_{i_1}^{a(i_1)} \dots \sigma_{i_r}^{a(i_r)}, \tag{A.6}$$

which is a r-spin correlation function at $T \to \infty$ for fixed magnetisation M. Here the $a(i)$ is the multiplicity of the i-th spin modulo 2. Thus, only the spins with odd multiplicity survive. Suppose there are W spins with odd multiplicities. Switching from Ising spins $\{-1, 1\}$ to bits $\{0, 1\}$ by performing $\sigma_{i_w} = 2b_{i_w} - 1$, we convert this problem into a combinatorial one:

$$\sum_{\boldsymbol{\sigma}\,|\,M} \prod_{w=1}^{W} \sigma_{i_w} = \sum_{\mathbf{b}\,|\,M} \prod_{w=1}^{W} (2b_{i_w} - 1) = (-1)^W \sum_{\mathbf{b}\,|\,M} \left[1 + \sum_{j=1}^{W} (-2)^j \sum_{v\in\mathcal{P}_j(S)} \prod_{\alpha\in v} \alpha\right], \tag{A.7}$$

where $S = \{\sigma_{i_1}, \dots, \sigma_{i_w}\}$ and $\mathcal{P}_j(S) = \{s \in \mathcal{P}(S)\,|\,|s| = j\}$, with $\mathcal{P}(S)$ denoting the power set of S. It now follows that

$$\begin{aligned}
&(-1)^W \sum_{\mathbf{b}\,|\,M} \left[1 + \sum_{j=1}^{W} (-2)^j \sum_{v\in\mathcal{P}_j(S)} \prod_{\alpha\in v} \alpha\right] \\
&= (-1)^W \binom{k^d}{\frac{k^d(M+1)}{2}} + (-1)^W \sum_{j=1}^{W} (-2)^j \sum_{v\in\mathcal{P}_j(S)} \sum_{\mathbf{b}\,|\,M} \prod_{\alpha\in v} \alpha \\
&= (-1)^W \binom{k^d}{\frac{k^d(M+1)}{2}} + (-1)^W \sum_{j=1}^{W} (-2)^j \sum_{v\in\mathcal{P}_j(S)} \left[\binom{k^d}{\frac{k^d(M+1)}{2}} \mathbb{P}(\alpha = 1, \forall\alpha \in v)\right],
\end{aligned} \tag{A.8}$$

where

$$\sum_{\mathbf{b}\,|\,M} \prod_{\alpha\in v} \alpha = \binom{k^d}{\frac{k^d(M+1)}{2}} \mathbb{P}(\alpha = 1, \forall\alpha \in v) \tag{A.9}$$

relates the left hand side to the probabilty $\mathbb{P}(\alpha = 1, \forall\alpha \in v)$ that all $\alpha \in v$ are unity. This is identical to the probability of picking j coins which are heads up,

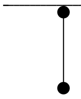
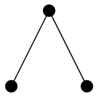
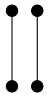
g	$\lambda(g)$	$W(g)$	$\eta(g)$
	1	2	$\frac{k^d d(k-1)}{2}$
	2	2	$\frac{k^d d(k-1)(d(k-1)-1)}{2}$
	2	4	$\frac{k^d d(k-1)((k^d-4)d(k-1)+2)}{8}$

Table A.1: Lowest order graphical terms in the high T expansion.

without replacement, given that $k^d(M+1)/2$ of them (out of k^d) are heads to begin with. This is the probability mass function of the hypergeometric distribution, and the above expression is exactly

$$\begin{aligned}
&\binom{k^d}{\frac{k^d(M+1)}{2}}(-1)^W\left[1+\sum_{j=1}^{W}(-2)^j\sum_{v\in\mathcal{P}_j(S)}\prod_{\ell=0}^{j-1}\frac{\frac{k^d(M+1)}{2}-\ell}{k^d-\ell}\right]\\
&=\binom{k^d}{\frac{k^d(M+1)}{2}}(-1)^W\left[1+\frac{\left(\frac{k^d(M+1)}{2}\right)!}{(k^d)!}\sum_{j=1}^{W}(-2)^j\binom{W}{j}\frac{\left(k^d-j\right)!}{\left(\frac{k^d(M+1)}{2}-j\right)!}\right] \qquad (A.10)\\
&=\binom{k^d}{\frac{k^d(M+1)}{2}}(-1)^W{}_2F_1\left(-\frac{k^d(M+1)}{2};-W;-k^d;2\right),
\end{aligned}$$

where we have used the fact that the summand of $\sum_{v\in\mathcal{P}_j(S)}$ is independent of v, and ${}_2F_1(a;b;c;z)$ is the ordinary (Gaussian) hypergeometric function, and the last step can be verified from the definition of the ordinary hypergeometric function

$$ {}_2F_1(a;b;c;z)=\sum_{n=0}^{\infty}\frac{(a)_n(b)_n}{(c)_n}\frac{z^n}{n!}, \qquad (A.11)$$

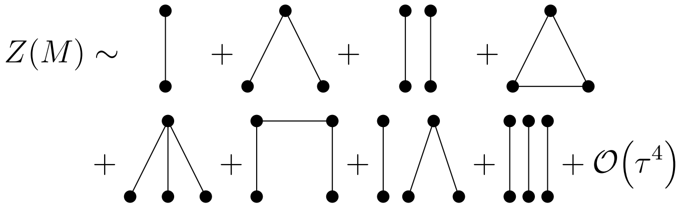
and $(m)_n$ is the Pochhammer symbol

$$(m)_n=m(m+1)\cdots(m+n-1)=\frac{\Gamma(m+n)}{\Gamma(m)}. \qquad (A.12)$$

We now return to the partition function in eq. (A.5). A cluster (high temperature) expansion can be performed in τ, with the expansion terms (i.e. "clusters") each corresponding to *edge-induced* subgraphs of $H_{d,k}$:

$$\begin{aligned}
Z(M)&=e^{\frac{1}{2T}}\gamma^{\frac{k^d d(k-1)}{2}}\binom{k^d}{\frac{k^d(M+1)}{2}}\\
&\times\left[1+\sum_{g\subseteq H_{d,k}}(-1)^{W(g)}\eta(g)\tau^{\lambda(g)}{}_2F_1\left(-\frac{k^d(M+1)}{2};-W(g);-k^d;2\right)\right], \qquad (A.13)
\end{aligned}$$

where g is an edge-induced subgraph of $H_{d,k}$ which has $\lambda(g)$ edges, $W(g)$ vertices with odd degree, and multiplicity $\eta(g)$ in $H_{d,k}$. Note that $(-1)^{W(g)} = 1$ is always true because there are always an even number of vertices with odd degree in any graph. The lowest order terms in τ thus have only one edge, then two edges, and so on. Graphically, the partition function can be represented as



(A.14)

In Table A.1, we present g, $\lambda(g)$, $W(g)$, and $\eta(g)$ for the first few lowest order terms in the expansion. Up to $\mathcal{O}(\tau^2)$, we have

$$Z(M) = e^{\frac{1}{2T}} \gamma^{\frac{k^d d(k-1)}{2}} \binom{k^d}{\frac{k^d(M+1)}{2}} \left[1 + a_1\tau + a_2\tau^2 + \mathcal{O}\left(\tau^3\right)\right], \tag{A.15}$$

where

$$a_1 = \frac{k^d d(k-1)}{2} {}_2F_1\left(-\frac{k^d(M+1)}{2}; -2; -k^d; 2\right) \tag{A.16}$$

and

$$\begin{aligned} a_2 = {} & \frac{k^d d(k-1)(d(k-1)-1)}{2} {}_2F_1\left(-\frac{k^d(M+1)}{2}; -2; -k^d; 2\right) \\ & + \frac{k^d d(k-1)((k^d-4)d(k-1)+2)}{8} {}_2F_1\left(-\frac{k^d(M+1)}{2}; -4; -k^d; 2\right). \end{aligned} \tag{A.17}$$

The log partition function can also be expanded up to $\mathcal{O}(\tau^2)$:

$$\begin{aligned} \log Z(M) = {} & \frac{1}{2T} + \frac{k^d d(k-1)}{2} \log\gamma + \log\binom{k^d}{\frac{k^d(M+1)}{2}} \\ & + a_1\tau + \tau^2\left(a_2 - \frac{a_1^2}{2}\right) + \mathcal{O}\left(\tau^3\right). \end{aligned} \tag{A.18}$$

The expectation of the Hamiltonian can now be evaluated

$$\begin{aligned} \langle \mathcal{H}(\boldsymbol{\sigma}) \rangle = {} & -\frac{\partial \log Z(M)}{\partial (T^{-1})} = -\frac{1+\tau}{2} \\ & - \frac{1}{\gamma^2 k^d d(k-1)}\left[a_1 + \tau\left(a_2 - \frac{a_1^2}{2}\right) + \mathcal{O}\left(\tau^2\right)\right]. \end{aligned} \tag{A.19}$$

The $q = 2$ case occurs, for instance, when dealing with Boolean functions in which $k = 2$ and $q = 2$, or with Gavrilets' holey fitness landscape [97], in which fitness values are either 0 or 1 indicating “inviable” or “viable” (and k will depend on model specifics). With a sufficient number of terms, the high temperature expansion can be used to study the behavior of the partition function as the maximum entropy constraint strengthens.

B
Metropolis-Hastings Algorithm and Markov Chain Proposal Steps

Once the system in Chapter 3 has been initialised, the simulation proceeds at a (fixed) temperature[1] T. In the Metropolis-Hastings algorithm [59, 60], the system proceeds via a Markov chain through a sequence of states Ω_t labelled by step number t. The state Ω_{t+1} is obtained from the previous state Ω_t with some transition probability. This transition probability $g(\Omega_t \to \Omega_{t+1})$ obeys detailed balance:

$$\pi(\Omega_t)g(\Omega_t \to \Omega_{t+1}) = \pi(\Omega_{t+1})g(\Omega_{t+1} \to \Omega_t), \tag{B.1}$$

where $\pi(\Omega_t) \equiv \pi(\Omega_t \,|\, \mathbf{f})$ is the probability of obtaining state Ω_t at equilibrium (i.e. it is the stationary distribution of the Markov process) where the frequency vector $\mathbf{f}$ is specified *a priori* (and omitted from further equations). Given that the simulation is taking place at a fixed temperature T, we know that the equilibrium probability distribution is time-independent and given by the Gibbs measure

$$\pi(\Omega_t) = \pi(\Omega_{t_1}) \equiv \pi(\Omega_t \,|\, \mathbf{f}) = \frac{e^{-\mathcal{H}(\Omega)/T}}{Z(T; \mathbf{f})}. \tag{B.2}$$

[1] We actually use a scaled temperature $T^* = Tk^d d(k-1)/2$, but in the following section we proceed with T for simplicity.

Therefore, the ratio of forwards to backwards transition probabilities only depends on the energy difference between the current (step t) and proposed (step $t+1$) states:

$$\frac{g(\Omega_t \to \Omega_{t+1})}{g(\Omega_{t+1} \to \Omega_t)} = \frac{\pi(\Omega_{t+1})}{\pi(\Omega_t)} = \exp\left[-\frac{1}{T}\left(\mathcal{H}(\Omega_{t+1}) - \mathcal{H}(\Omega_t)\right)\right]. \tag{B.3}$$

Since this does not uniquely determine the transition probability, the Metropolis choice is typically used:

$$g(\Omega_t \to \Omega_{t+1}) = \min\left\{1, e^{-\frac{1}{T}(\mathcal{H}(\Omega_{t+1})-\mathcal{H}(\Omega_t))}\right\} \tag{B.4}$$

In other words, if the change in energy $\mathcal{H}(\Omega_{t+1}) - \mathcal{H}(\Omega_t) < 0$, the proposed new state is accepted at time $t+1$. Otherwise, the transition happens only with probability $e^{-\frac{1}{T}(\mathcal{H}(\Omega_{t+1})-\mathcal{H}(\Omega_t))}$.

For a Metropolis-Hastings algorithm, the exact change in the configuration which is proposed at each time step depends on the particular system. Our system is a classical Potts model with fixed frequency vector $\mathbf{f}$. In order to keep the frequency vector fixed at all times, we set our proposal steps to be a *swap* of two input-output pairs. That is, at step t the system is in state Ω_t; we choose two random sequences/inputs/vertices x and y. Suppose we have input-output pairs $\Omega_t(x) = A$ and $\Omega_t(y) = B$; the proposed configuration Ω_{t+1} would have $\Omega_{t+1}(x) = B$ and $\Omega_{t+1}(y) = A$.

C
Proof of Upper Bound on Population Neutrality of Bricklayer's Graphs

This proof is due to Shyam Narayanan, who is a collaborator on the work introduced in Chapter 4. We reproduce it with his permission here to provide completeness to the arguments presented in Chapter 4.

C.1 Preliminaries

First, we note some basic preliminary inequalities.

Proposition C.1.1. *For any* $x \geq -1$*,* $\log_k(1+x) \leq \frac{x}{\ln k}$*.*

Proof. This is immediate by the facts that $\log_k(1+x) = \frac{\ln(1+x)}{\ln k}$ and $\ln(1+x) \leq x$ by convexity. □

Proposition C.1.2. *For* $x > 1$*,* $\frac{x-1}{\ln x}$ *is a positive and increasing function.*

Proof. Since $f(y) = \ln(1+y)$ is increasing and convex on $y \in (0, \infty)$ and since $f(0) = 0$, this means that $\frac{f(y)}{y} = \frac{\ln(1+y)}{y}$ is a positive and decreasing function on $y > 0$, so $\frac{y}{\ln(1+y)}$ is a positive and increasing function on $y > 0$. The claim follows by setting $y = x - 1$. □

Proposition C.1.3. *Let $n = m \cdot k + r$ for $k \geq 2, m \geq 1$, and $1 \leq r \leq k-1$. Then,*

$$\frac{k-r}{m} + \frac{r}{m+1} \geq \frac{k^2}{n}.$$

Proof. Since $\frac{1}{x}$ is convex, we can use Jensen's inequality to get

$$\begin{aligned} \frac{k-r}{m} + \frac{r}{m+1} &= k\left[\frac{(k-r)/k}{m} + \frac{r/k}{m+1}\right] \\ &\geq k \cdot \frac{1}{m \cdot \frac{k-r}{k} + (m+1) \cdot \frac{r}{k}} = \frac{k}{m + \frac{r}{k}} = \frac{k^2}{n}. \quad \square \end{aligned} \tag{C.1}$$

For a graph $G = (V, E)$ on $n = |V|$ vertices, we let $A(G)$ represent the adjacency matrix of G, i.e., $A(G)_{i,j} = 1$ if $(i, j) \in E$ and $A(G)_{i,j} = 0$ otherwise. Note that the diagonal of the adjacency matrix is all 0's. For any square matrix A, we let $\lambda_{\max}(A)$ represent the maximum eigenvalue of A. In addition, define $A_{n,k}$ to be the adjacency matrix of the bricklayer graph $G_{n,k}$, and define $\lambda_{n,k}$ to be the largest eigenvalue of the adjacency matrix $A_{n,k}$, i.e., $\lambda_{n,k} = \lambda_{\max}(A_{n,k})$.

We also recall that for two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, the *Cartesian product* $G = G_1 \times G_2$ is the graph on the set product of vertices $V_1 \times V_2$, such that (v_1, v_2) and (w_1, w_2) are connected in G if and only if either $v_1 = w_1$ and $(v_2, w_2) \in E_2$, or $v_2 = w_2$ and $(v_1, w_2) \in E_1$. Note that the Hamming graph $H_{d,k} = G_{k^d,k}$ is just the Cartesian product of d copies of the complete graph K_k on k vertices.

Next, we note the following well-known proposition:

Proposition C.1.4. *For any graphs G_1, G_2 with Cartesian product G, $\lambda_{\max}(A(G)) = \lambda_{\max}(A(G_1)) + \lambda_{\max}(A(G_2))$.*

As a direct corollary, since the maximum eigenvalue of the complete graph K_k's adjacency matrix is $k-1$, this implies that the Hamming graph $H_{d,k}$ has maximum eigenvalue $d(k-1) = (k-1)\log_k(k^d)$.

We also note the following simple result:

Proposition C.1.5. *For any integer $m \geq 1$, $G_{mk,k}$ is isomorphic to $G_{m,k} \times K_k$.*

Proof. Note that for $0 \leq s \neq s' \leq mk - 1$, vertices s, s' are connected if and only if s, s' differ in exactly one coordinate in their base k representation. If we write $s = k \cdot \ell + r$ and $s' = k \cdot \ell' + r'$, where $0 \leq \ell, \ell' \leq m$, $0 \leq r, r' \leq m - 1$, then this is equivalent to either $\ell = \ell'$ and $r \neq r'$, or ℓ and ℓ' differ in exactly one coordinate and $r = r'$. Thus, if we represent s by the pair (ℓ, r) and s' by the pair (ℓ', r'), s, s' are connected if and only if (ℓ, r) and (ℓ', r') are connected in the product graph $G_{m,k} \times K_k$. □

By combining Propositions C.1.4 and C.1.5, we have the following corollary:

Corollary C.1.1. $\lambda_{mk,k} = \lambda_{\max}(K_k) + \lambda_{m,k} = (k-1) + \lambda_{m,k}$.

Next, we note the Courant-Fischer formula [98] on maximum eigenvalues of symmetric matrices:

Proposition C.1.6. *(Courant-Fischer) For any (real) symmetric matrix A, the maximum eigenvalue $\lambda_{\max}(A)$ of A satisfies*

$$\lambda_{\max}(A) = \sup_{\|v\|_2=1} v^T A v = \sup_{v \neq 0} \frac{v^T A v}{\|v\|_2^2}.$$

In addition, $v \neq 0$ is an eigenvector with eigenvalue λ if and only if $v^T A v = \lambda \cdot \|v\|_2^2$.

Using the Courant-Fischer formula, we prove another simple result on properties of $\lambda_{n,k}$.

Proposition C.1.7. *If $n \leq n'$, then $\lambda_{n,k} \leq \lambda_{n',k}$. In other words, $\lambda_{n,k}$ is an increasing function of n.*

Proof. Recall that $\lambda_{n,k} = \sup_{\|v\|_2=1} v^T A_{n,k} v$. Then, by setting $w \in \mathbb{R}^{n'}$ to just be v concatenated with $n' - n$ 0's, we have that $w^T A_{n',k} w = v^T A_{n,k} v = \lambda_{n,k}$, so $\lambda_{n',k} = \sup_{\|w\|_2=1} w^T A_{n',k} w \geq \lambda_{n,k}$. □

We also use the Courant-Fischer formula to prove the following:

Proposition C.1.8. *Let A be a symmetric matrix with all nonnegative entries. If v is an eigenvector of A with eigenvalue $\lambda_{\max}(A)$, then so is v', the vector where each coordinate v_i is replaced with its absolute value $|v_i|$.*

Proof. Note that $(v')^T A(v') \geq v^T A v$, since $(v')^T A(v') = \sum_{i,j} |v_i| \cdot |v_j| \cdot A_{ij} \geq \sum_{i,j} v_i \cdot v_j \cdot A_{ij}$. However, v' and v have the same ℓ_2-norm. So, by Proposition C.1.6, $\lambda_{\max}(A) \geq \frac{(v')^T A(v')}{\|v'\|_2^2} \geq \frac{v^T A v}{\|v\|_2^2} = \lambda_{\max}(A)$, so $\frac{(v')^T A(v')}{\|v'\|_2^2} = \lambda_{\max}(A)$. Thus, v' is an eigenvector of A with eigenvalue $\lambda_{\max}(A)$. □

Our final preliminary results relate to graph automorphisms.

Definition C.1.1. *For a graph $G = (V, E)$, where $V = [n]$, a permutation $\pi : [n] \to [n]$ is called an* automorphism *if for all pairs (i, j) for $1 \leq i < j \leq n$, (i, j) is an edge in E if and only if $(\pi(i), \pi(j)) \in E$.*

Proposition C.1.9. *Let π be an automorphism of $G = (V, E)$, where $V = [n]$. Then, if v is an eigenvector of A_G with eigenvalue λ, then $\pi(v)$ is also an eigenvector of A_G with eigenvalue λ, where $\pi(v)$ is defined to have ith coordinate $v_{\pi(i)}$.*

Proof. Suppose that $A_G v = \lambda v$. Then, for all $i \in V$, $\sum_{j:(i,j) \in E} v_j = \lambda \cdot v_i$. Note that by the definition of automorphism, if π is an automorphism, then so is π^{-1}, so $\sum_{j:(\pi^{-1}(i),(\pi^{-1}(j)) \in E} v_j = \lambda \cdot v_i$ for all $i \in V$. Therefore, replacing i with $\pi(i)$ and j with $\pi(j)$, we get that $\sum_{j:(i,j) \in E} v_{\pi(j)} = \lambda \cdot v_{\pi(i)}$ for all i. Therefore, $\pi(v)$ is an eigenvector of A_G with eigenvalue λ. □

As a corollary, we have the following result.

Corollary C.1.2. *Let $\mathcal{G}$ be a subgroup of S_n, the permutation group, such that for all $\sigma \in \mathcal{G}$, σ is an automorphism. Then, A_G, there exists a (nonzero) eigenvector v of eigenvalue $\lambda_{\max}(A_G)$, such that for all i and all $\sigma \in \mathcal{G}$, $v_i = v_{\sigma(i)}$.*

Proof. Let w be some (nonzero) eigenvector of A_G with eigenvalue $\lambda_{\max}(A_G)$. Since all entries of A_G are nonnegative-valued, we can assume without loss of generality that each entry of w is nonnegative, by Proposition C.1.8. Now, let $v = \frac{1}{|\mathcal{G}|} \cdot \sum_{\pi \in \mathcal{G}} \pi(w)$, i.e., v is the average of $\pi(w)$ overall all $\pi \in \mathcal{G}$. Note that by Proposition C.1.9, $\pi(w)$ is an eigenvector of eigenvalue $\lambda_{\max}(A_G)$, so therefore, v, as a linear combination of the vectors $\pi(w)$, is also. In addition, note that since w has all nonnegative entries and is nonzero, this means $\pi(w)$ also has all nonnegative entries and is nonzero for any $\pi \in \mathcal{G}$, so $v = \frac{1}{|\mathcal{G}|} \cdot \sum_{\pi \in \mathcal{G}} \pi(w)$ does as well. Importantly, v is nonzero.

Finally, we need to check that $v_i = v_{\sigma(i)}$ for all $\sigma \in \mathcal{G}$. However, note that

$$v_{\sigma(i)} = \frac{1}{|\mathcal{G}|} \sum_{\pi \in \mathcal{G}} \pi(w)_{\sigma(i)} = \frac{1}{|\mathcal{G}|} \sum_{\pi \in \mathcal{G}} w_{(\pi \circ \sigma)(i)} = \frac{1}{|\mathcal{G}|} \sum_{\pi \in \mathcal{G}} w_{\pi(i)} = v(i),$$

since the set $\{\pi \circ \sigma\}$ over all π counts each permutation in $\mathcal{G}$ exactly once. □

C.2 Population Neutrality of Bricklayer Graphs: Main Theorem

In this section, we prove the following theorem, conjectured by Reeves et al. [61]

Theorem 1. *For all integers $k \geq 2$ and all integers $n \geq 1$, $\lambda_{\max}(A_{n,k}) \leq (k-1) \cdot \log_k n$.*

Because Reeves et al. [61] proved this theorem for the special case $k = 2$, we only focus on $k \geq 3$. We note that our techniques can in fact generalise to the $k = 2$ case as well, though this requires additional computation and significant tweaking of parameters.

If $k < n$ but $k \nmid n$, then we can write $n = k \cdot m + r$ for some $1 \leq r \leq k-1$. Then, note that we can split the bricklayer graph into k layers, based on the last bit of the base k representation of each number in $\{0, 1, \ldots, n-1\}$. The first r layers can be permuted arbitrarily without changing the structure of the graph (i.e., these permutations are automorphisms of $G_{n,k}$), as can the last $k - r$ layers. Based on this, we can permute the ordering of the graph $G_{n,k}$ and draw the adjacency matrix $A_{n,k}$ as in Figure C.1.

The following lemma will be a very important observation for bounding $\lambda_{n,k}$.

Lemma 1. *Suppose that $n = k \cdot m + r$, where $1 \leq r \leq k-1$. Then,*

$$(\lambda_{n,k} - \lambda_{m+1,k} - (r-1)) \cdot (\lambda_{n,k} - \lambda_{m,k} - (k-r-1)) \leq r(k-r). \tag{C.2}$$

$$\begin{array}{r} r\left\{\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right. \\ k-r\left\{\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right. \end{array}
\left[\begin{array}{cccc|ccc}
A_{m+1,k} & I_{m+1} & \cdots & I_{m+1} & J & \cdots & J \\
I_{m+1} & A_{m+1,k} & \cdots & I_{m+1} & J & \cdots & J \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
I_{m+1} & I_{m+1} & \cdots & A_{m+1,k} & J & \cdots & J \\
\hline
J^T & J^T & \cdots & J^T & A_{m,k} & \cdots & I_m \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
J^T & J^T & \cdots & J^T & I_m & \cdots & A_{m,k}
\end{array}\right]$$

$$\underbrace{\hphantom{A_{m+1,k}\;I_{m+1}\;\cdots\;A_{m+1,k}}}_{r}\qquad\underbrace{\hphantom{A_{m,k}\;\cdots\;A_{m,k}}}_{k-r}$$

Figure C.1: The adjacency matrix $A_{km+r,k}$ expressed in terms of $A_{m,k}$ and $A_{m+1,k}$. Here, $I_m \in \mathbb{R}^{m\times m}$ represents the $m \times m$ identity matrix (and similarly for $I_{m+1} \in \mathbb{R}^{(m+1)\times(m+1)}$. Finally, $J \in \mathbb{R}^{(m+1)\times m}$ is the identity $m \times m$ matrix with an extra row of all 0's.

Proof. We use the order of vertices of the bricklayer graph $G_{k\cdot m+r,k}$ to produce an adjacency matrix as in Figure C.1 (where $k \cdot m + r = n$). Then, due to the automorphisms by permuting the first r layers and last $k-r$ layers of the bricklayer graph, by Corollary C.1.2 we can choose an eigenvector u of maximum eigenvalue of the form $u = (v, \ldots, v, w, \ldots, w)$, where there are r copies of $v \in \mathbb{R}^{m+1}$ concatenated together with $k-r$ copies of $w \in \mathbb{R}^m$. Suppose that $\|v\|_2 = \alpha$ and $\|w\|_2 = \beta$. Then,

$$\|u\|_2^2 = r \cdot \alpha^2 + (k-r)\cdot\beta^2, \tag{C.3}$$

and recalling that $J \in \mathbb{R}^{(m+1)\times m}$ is the identity $m \times m$ matrix with an extra row of 0's,

$$\begin{aligned}
u^T A_n u &= r \cdot v^T A_{m+1,k} v + (k-r)\cdot w^T A_{m,k} w \\
&\qquad + r(r-1)\|v\|_2^2 + (k-r)(k-r-1)\|w\|_2^2 + 2r(k-r)\cdot v^T J w \\
&\le r \cdot \lambda_{m+1,k}\cdot\alpha^2 + (k-r)\cdot\lambda_{m,k}\cdot\beta^2 \\
&\qquad + r(r-1)\cdot\alpha^2 + (k-r)(k-r-1)\cdot\beta^2 + 2r(k-r)\cdot\alpha\beta.
\end{aligned} \tag{C.4}$$

The inequality follows by the properties of maximum eigenvalues, definitions of α and β, and the Cauchy-Schwarz inequality which tells us that if $v_{-(m+1)}$ is v with its last coordinate removed, then $v^T J w = \langle v_{-(m+1)}, w\rangle \le \|v_{-(m+1)}\|_2 \cdot \|w\|_2 \le \alpha\cdot\beta$.

If $\beta = 0$, then $u^T A_n u = r \cdot [\lambda_{m+1,k} + (r-1)]\cdot\alpha^2$, whereas $\|u\|_2^2 = r\cdot\alpha^2$. So, $\lambda_{n,k} = \frac{u^T A_n u}{\|u\|_2^2} = \lambda_{m+1,k} + (r-1)$, and thus the left hand side of Equation (C.2) is 0, so the Lemma immediately follows. Otherwise, assume $\beta \neq 0$: by scaling, we can assume without loss of generality that $\beta = 1$. In this case, by Equations (C.3) and (C.4),

$$\begin{aligned}\lambda_{n,k} = \frac{u^T A_n u}{\|u\|_2^2} \leq [r \cdot \alpha^2 + (k-r)] \Big(& r \cdot \lambda_{m+1,k} \cdot \alpha^2 \\ & +(k-r) \cdot \lambda_{m,k} + r(r-1) \cdot \alpha^2 + (k-r)(k-r-1) + 2r(k-r) \cdot \alpha \Big) .\end{aligned} \tag{C.5}$$

We will attempt to maximise the above fraction over α. Indeed, for fixed $\lambda_{n,k}, \lambda_{m,k}, \lambda_{m+1,k}$, if there exists a solution to the above inequality, then

$$\alpha^2 \cdot r \cdot (\lambda_{n,k} - \lambda_{m+1,k} - (r-1)) - 2r(k-r) \cdot \alpha + (k-r) \cdot (\lambda_{n,k} - \lambda_{m,k} - (k-r-1)) \leq 0 \tag{C.6}$$

has a solution in α. In general, for real numbers A, B, C, where $A \geq 0$, $Ax^2 + Bx + C \leq 0$ has a real solution in x if and only if the discriminant condition $B^2 - 4AC \geq 0$ holds. However, by setting u to be r copies of a maximum eigenvector of $A_{m+1,k}$ concatenated with $k-r$ copies of $\mathbf{0} \in \mathbb{R}^m$, we saw that $\frac{u^T A_n u}{\|u\|_2^2} = \lambda_{m+1,k} + (r-1)$, so $\lambda_{n,k} \geq \lambda_{m+1,k} + (r-1)$. So, in (C.6), the quadratic term $r \cdot (\lambda_{n,k} - \lambda_{m+1,k} - (r-1))$ is nonnegative. Now, the discriminant condition is

$$4r^2(k-r)^2 - 4r(k-r)(\lambda_{n,k} - \lambda_{m+1,k} - (r-1)) \cdot (\lambda_{n,k} - \lambda_{m,k} - (k-r-1)) \geq 0,$$

and by dividing by $4r(k-r)$, we get that Equation (C.2) holds. □

Our goal will be to prove, by strong induction, that whenever $k \geq 3$ and n is not a power of k, that $\lambda_{n,k} \leq (k-1)\log_k n - \frac{\alpha_k}{n}$, where α_k is $\frac{1}{6}$ for $k = 3$ and $\frac{1}{2}$ for $k \geq 4$.

First, we consider cases where n is very small, or close to a power of k.

Proposition C.2.1. *For all $k \geq 2$ and any $1 \leq r \leq k-1$, $\lambda_{k+r,k} \leq (k-1) + \frac{c_k \cdot r}{k+r}$, where $c_k = 1.3$ if $2 \leq k \leq 7$ and $c_k = 2.2$ if $k \geq 8$.*

Proof. Set $n = k + r$ and $m = 1$, so $n = km + r$. Then, by Equation (C.2), $(\lambda_{n,k} - \lambda_{2,k} - (r-1)) \cdot (\lambda_{n,k} - \lambda_{1,k} - (k-r-1)) \leq r(k-r)$. Since $\lambda_{1,k} = 0$ and

$\lambda_{2,k} = 1$, this means that $(\lambda_{n,k} - r) \cdot (\lambda_{n,k} - (k - r - 1)) \leq r(k - r)$. However, note that if $k \geq 8$, then

$$\begin{aligned}
&\left((k-1) + \frac{r \cdot c_k}{n} - r\right) \cdot \left((k-1) + \frac{r \cdot c_k}{n} - (k - r - 1)\right) \\
&= \left((k - 1 - r) + \frac{r \cdot c_k}{n}\right) \cdot \left(r + \frac{r \cdot c_k}{n}\right) \\
&\geq r(k-r) - r + \frac{(k-1) \cdot r \cdot c_k}{k+r} \\
&= r(k-r) + r \cdot \left[c_k \cdot \frac{k-1}{k+r} - 1\right] \\
&\geq r(k-r),
\end{aligned}$$

since $2.2 \cdot \frac{k-1}{k+r} \geq 2.2 \cdot \frac{k-1}{2k-1} \geq 1$ if $k \geq 8$. Therefore, we must have $\lambda_{n,k} \leq (k-1) + \frac{r \cdot c_k}{n}$.

Finally, if $2 \leq k \leq 7$, this proposition can be checked numerically. □

Proposition C.2.2. *For all $k \geq 2$, $a \geq 1$, and $1 \leq r \leq k-1$, we have $\lambda_{k^a+r,k} \leq a(k-1) + \frac{c_k \cdot r}{k^a+r}$.*

Proof. We fix k and prove this by induction on a. Proposition C.2.1 covers the base case of $a = 1$. Suppose the statement is true for a and all $1 \leq r \leq k-1$; we will try to prove it for $a+1$ and all $1 \leq r \leq k-1$. Set $n = k^{a+1} + r$ and $m = k^a$. Then, it suffices to show that

$$\begin{aligned}
&\left((a+1)(k-1) + \frac{c_k \cdot r}{n} - a(k-1) - \frac{c_k}{m+1} - (r-1)\right) \\
&\qquad \cdot \left((a+1)(k-1) + \frac{c_k \cdot r}{n} - a(k-1) - (k - r - 1)\right) \geq r(k-r). \quad \text{(C.7)}
\end{aligned}$$

To see why, since $\lambda_{m,k} = a(k-1)$ and $\lambda_{m+1,k} \leq a(k-1) + \frac{c_k}{m+1}$ by our induction hypothesis, we must have that $(a+1)(k-1) + \frac{r \cdot c_k}{n} \geq \lambda_{n,k}$ in order for Equation (C.2) to hold.

We can simplify the left hand side of Equation (C.7) as

$$\begin{aligned}
&\left((k-r)+c_k\cdot\left(\frac{r}{n}-\frac{1}{m+1}\right)\right)\cdot\left(r+\frac{c_k\cdot r}{n}\right)\\
&=r(k-r)+c_k\left(\frac{r(k-r)}{n}+\frac{r^2}{n}-\frac{r}{m+1}+\frac{rc_k}{n}\left(\frac{r}{n}-\frac{1}{m+1}\right)\right)\\
&=r(k-r)+rc_k\left(\frac{k}{n}-\frac{1}{m+1}-\frac{c_k}{n(m+1)}+\frac{rc_k}{n^2}\right)\\
&\geq r(k-r)+rc_k\left(\frac{k}{n}-\frac{1}{m+1}-\frac{c_k}{n(m+1)}+\frac{rc_k}{n(m+1)k}\right)\\
&=r(k-r)+rc_k\left(\frac{k-r-c_k+\frac{rc_k}{k}}{n(m+1)}\right),
\end{aligned}$$

which is at least $r(k-r)$ since $k-r-c_k+\frac{rc_k}{k}=\frac{1}{k}\cdot(k-r)(k-c_k)\geq 0$. □

Proposition C.2.3. *For all $k\geq 2$, and if $a\geq 2$ and $1\leq s\leq k-1$ or if $a=s=1$, then $\lambda_{k^a-s,k}\leq a(k-1)-\frac{s\cdot(k-1)}{k^a-s}$.*

Proof. Again, we fix k and prove this by induction on a. First, note that when $a=1$ and $s=1$, $\lambda_{k-1,k}=k-2=a(k-1)-\frac{k-1}{k^a-1}$. Now, suppose the statement is true for a; we will try to prove it for $a+1$. Set $n=k^{a+1}-s$, $m=k^a-1$, and $r=k-s$. Then, it suffices to show that

$$\begin{aligned}
&\left((a+1)(k-1)-\frac{s(k-1)}{n}-a(k-1)-(k-s-1)\right)\\
&\quad\cdot\left((a+1)(k-1)-\frac{s(k-1)}{n}-a(k-1)+\frac{k-1}{m}-(s-1)\right)\geq s(k-s). \quad \text{(C.8)}
\end{aligned}$$

This is because $s(k-s)=r(k-r)$, $\lambda_{m+1,k}=a(k-1)$, and $\lambda_{m,k}\leq a(k-1)-\frac{k-1}{m}$ by our induction hypothesis, so we must have that $(a+1)(k-1)-\frac{s(k-1)}{n}\geq\lambda_{n,k}$ in order for Equation (C.2) to hold.

We can simplify the left hand side of Equation (C.8) as

$$\begin{aligned}
&=\left(s-\frac{s(k-1)}{n}\right)\cdot\left((k-s)+\frac{k-1}{m}-\frac{s(k-1)}{n}\right)\\
&=s(k-s)+s(k-1)\left(\frac{1}{m}-\frac{s}{n}-\frac{k-s}{n}-\frac{k-1}{mn}+\frac{s(k-1)}{n^2}\right)\\
&=s(k-s)+s(k-1)\left(\frac{s(k-1)}{n^2}-\frac{s-1}{mn}\right)\\
&=s(k-s)+s(k-1)\cdot\frac{s(k-1)m-(s-1)n}{n^2m}.
\end{aligned}$$

Finally, since $m = k^a - 1, n = k^{a+1} - s$, we get that $s(k-1)m - (s-1)n = (k-s)(k^a - s) \geq 0$, so the left hand side is at least $s(k-s)$. This concludes the proof. □

For all positive integers n, k, we define $\mu_{n,k} = (k-1)\log_k n$ and $\nu_{n,k} = (k-1)\log_k n - \frac{\alpha_k}{n}$.We prove the following lemma:

Lemma 2. *For all $n = m \cdot k + r$, where $1 \leq r \leq k-1$ and $m \geq \frac{k-1}{2\alpha_k (\ln k)^2}$, we have that*

$$(\nu_{n,k} - \nu_{m+1,k} - (r-1)) \cdot (\nu_{n,k} - \nu_{m,k} - (k-r-1)) \geq r(k-r). \tag{C.9}$$

Proof. Note that for $n = m \cdot k + r$, $\mu_{n,k} - \mu_{m+1,k} - (r-1) = (k-1)\log_k n - (k-1)\log_k (m+1) - (r-1)$. Noting that $(k-1)\log_k(m+1) = (k-1)\log_k(km+k) - (k-1)$, we have that

$$\begin{aligned} \mu_{n,k} - \mu_{m+1,k} - (r-1) &= (k-1)\log_k \frac{n}{km+k} + (k-r) \\ &= (k-r) - (k-1)\log_k \left(1 + \frac{k-r}{n}\right) \\ &\geq (k-r)\left(1 - \frac{k-1}{n \cdot \ln k}\right). \end{aligned}$$

The final inequality follows by Proposition C.1.1.

Likewise, $\mu_{n,k} - \mu_{m,k} - (k-r-1) = (k-1)\log_k n - (k-1)\log_k m - (k-r-1)$. Noting that $(k-1)\log_k m = (k-1)\log_k(km) - (k-1)$, we have that

$$\begin{aligned} \mu_{n,k} - \mu_{m,k} - (k-r-1) &= (k-1)\log_k \frac{n}{km} + r \\ &= r - (k-1)\log_k \left(1 - \frac{r}{n}\right) \\ &\geq r\left(1 + \frac{k-1}{n \cdot \ln k}\right). \end{aligned}$$

The final inequality again follows by Proposition C.1.1.

Therefore,

$$\nu_{n,k} - \nu_{m+1,k} - (r-1) \geq (k-r) \cdot \underbrace{\left(1 - \frac{k-1}{n \cdot \ln k}\right)}_{A} + \alpha_k \cdot \underbrace{\left(\frac{1}{m+1} - \frac{1}{n}\right)}_{B} \tag{C.10}$$

and

$$\nu_{n,k} - \nu_{m,k} - (k - r - 1) \geq r \cdot \underbrace{\left(1 + \frac{k-1}{n \cdot \ln k}\right)}_{C} + \alpha_k \cdot \underbrace{\left(\frac{1}{m} - \frac{1}{n}\right)}_{D} \tag{C.11}$$

Note that for any $k \geq 2$ and $m \geq k+1$, that $\frac{k-1}{n \ln k} \leq \frac{k-1}{(k+1)\ln k} \leq \frac{1}{2}$. In addition, note that $m, m+1 \leq n$. So, with A, B, C, D as defined in Equations (C.10) and (C.11), we have that $A \geq \frac{1}{2}, C \geq 1$, and $B, D \geq 0$. Therefore,

$$\begin{aligned}
&(\nu_{n,k} - \mu_{m+1,k} - (r-1)) \cdot (\mu_{n,k} - \mu_{m,k} - (k-r-1)) \\
&\geq [(k-r) \cdot A + \alpha_k \cdot B] \cdot [r \cdot C + \alpha_k \cdot D] && \text{By (C.10) and (C.11)} \\
&\geq r(k-r) \cdot AC + \alpha_k \left[r \cdot B \cdot C + (k-r) \cdot A \cdot D\right] && \text{By ignoring } B \cdot D \text{ term} \\
&\geq r(k-r) \cdot AC + \frac{\alpha_k}{2} \left[r \cdot B + (k-r) \cdot D\right] && \text{Since } A, C \geq \frac{1}{2} \\
&\geq r(k-r) \cdot AC + \frac{\alpha_k}{2} \cdot \left(\frac{k(k-1)}{n}\right) && \text{By Proposition C.1.3} \\
&= r(k-r) - r(k-r) \cdot \frac{(k-1)^2}{n^2(\ln k)^2} + \frac{\alpha_k}{2} \cdot \left(\frac{k(k-1)}{n}\right) && \text{By definition of } A, C \\
&= r(k-r) - \frac{k^2}{4} \cdot \frac{(k-1)^2}{n^2(\ln k)^2} + \frac{\alpha_k}{2} \cdot \left(\frac{k(k-1)}{n}\right) && \text{Since } r(k-r) \geq k^2/4.
\end{aligned}$$

Finally, $\frac{k^2}{4} \cdot \frac{(k-1)^2}{n^2(\ln k)^2} \leq \frac{\alpha}{2} \cdot \left(\frac{k(k-1)}{n}\right)$ as long as $n \geq \frac{k(k-1)}{2\alpha_k(\ln k)^2}$, which is true if $m \geq \frac{k-1}{2\alpha_k(\ln k)^2}$. □

We are now ready to state and prove (a slightly stronger version of) our main theorem.

Theorem 2. *For all $k \geq 3$ and $n \geq 1$, $\lambda_{n,k} \leq (k-1)\log_k n$. In addition, if n is not a power of k, then $\lambda_{n,k} \leq (k-1)\log_k n - \frac{\alpha_k}{n}$, where $\alpha_k = \frac{1}{2}$ for $k \geq 4$ and $\alpha_k = \frac{1}{6}$ for $k = 3$.*

Proof. First, note that for $n = k^a$ (including $n = 1 = k^0$), $\lambda_{n,k} = a(k-1) = (k-1) \cdot \log_k n$, as we saw in the preliminary section.

Next, suppose that $2 \leq n \leq k-1$. Then, $\lambda_{n,k} = n-1$, so we just have to show that $n-1 \leq \frac{k-1}{\ln k} \cdot \ln n - \frac{1}{2n}$. This is equivalent to $1 - \frac{1}{n} + \frac{1}{2n^2} \leq \frac{k-1}{\ln k}$ for all $2 \leq n \leq k-1$.

Since $\frac{k-1}{\ln k}$ is an increasing function over $k \geq 2$, it suffices to prove this for $k = n+1$, or $n - 1 \leq \frac{n}{\ln(n+1)} \cdot \ln n - \frac{1}{2n}$. However, $\ln(n+1) = \ln n + \ln\left(1 + \frac{1}{n}\right) \leq \ln n + \frac{1}{n}$, so

$$\frac{n}{\ln(n+1)} \cdot \ln n - \frac{1}{2n} \geq \frac{n \cdot \ln n}{\ln n + \frac{1}{n}} - \frac{1}{2n} \geq n\left(1 - \frac{1}{n \ln n}\right) - \frac{1}{2n} \geq n - \frac{1}{\ln n} - \frac{1}{2} \geq n - 1$$

if $n \geq 8$. If $2 \leq n \leq 7$, it is still true that $n - 1 \leq \frac{n}{\ln(n+1)} \cdot \ln n - \frac{1}{2n}$, which can be checked manually.

We now suppose $n \geq k$. We prove our claim by strong induction on n, splitting into five cases.

Case 1: $\boldsymbol{k|n}$. Then, by Proposition C.1.4, $\lambda_{n,k} = \lambda_{n/k,k} + (k-1) \leq (k-1) + (k-1)\log_k(n/k) = (k-1)\log_k n$. In addition, if n is not a power of k, then $\lambda_{n,k} = \lambda_{n/k,k} + (k-1) \leq (k-1) + (k-1)\log_k(n/k) - \frac{\alpha_k}{n/k} \leq (k-1)\log_k n - \frac{\alpha_k}{n}$.

Case 2: $\boldsymbol{n = k^a + r}$ for some $\boldsymbol{1 \leq r \leq k-1, a \geq 1}$. Then, we know by Proposition C.2.2 that $\lambda_{n,k} \leq a(k-1) + \frac{c_k \cdot r}{n}$. However,

$$\begin{aligned}(k-1)\log_k n &= (k-1)a + \frac{k-1}{\ln k} \cdot \ln\left(\frac{k^a + r}{k^a}\right) = (k-1)a - \frac{k-1}{\ln k} \cdot \ln\left(\frac{k^a}{k^a + r}\right) \\ &\geq (k-1)a + \frac{k-1}{\ln k} \cdot \frac{r}{n}.\end{aligned} \tag{C.12}$$

Therefore,

$$(k-1)\log_k n - \lambda_{n,k} \geq (k-1)a + \frac{k-1}{\ln k} \cdot \frac{r}{n} - \left[a(k-1) + \frac{c_k \cdot r}{n}\right] \geq \frac{r}{n}\left(\frac{k-1}{\ln k} - c_k\right).$$

It is simple to see that $\frac{k-1}{\ln k} \geq 1.8$ for all $k \geq 3$ and $\frac{k-1}{\ln k} \geq 2.7$ for all $k \geq 8$, so by the definition of c_k, $\frac{k-1}{\ln k} - c_k \geq 0.5$. Thus, $(k-1)\log_k n - \lambda_{n,k} \geq 0.5 \cdot \frac{r}{n} \geq \frac{\alpha_k}{n}$.

Case 3: $\boldsymbol{n = k^a - s}$ for some $\boldsymbol{1 \leq s \leq k-1, a \geq 2}$. Then, we know by Proposition C.2.3 that $\lambda_{n,k} \leq a(k-1) - \frac{s \cdot (k-1)}{n}$. However,

$$\begin{aligned}(k-1)\log_k n &= (k-1)a + \frac{k-1}{\ln k} \cdot \ln\left(\frac{k^a - s}{k^a}\right) = (k-1)a - \frac{k-1}{\ln k} \cdot \ln\left(\frac{k^a}{k^a - s}\right) \\ &\geq (k-1)a - \frac{k-1}{\ln k} \cdot \frac{s}{n},\end{aligned} \tag{C.13}$$

Therefore,

$$\begin{aligned}(k-1)\log_k n - \lambda_{n,k} &\geq (k-1)a - \frac{k-1}{\ln k} \cdot \frac{s}{n} - \left[a(k-1) - \frac{s \cdot (k-1)}{n}\right] \\ &= \frac{s}{n}\left[(k-1) - \frac{k-1}{\ln k}\right].\end{aligned} \tag{C.14}$$

For $k \geq 4$, $k-1-\frac{k-1}{\ln k} \geq 0.5$, and for $k=3$, $k-1-\frac{k-1}{\ln k} \geq \frac{1}{6}$. Therefore, $(k-1)\log_k n - \lambda_{n,k} \geq \frac{\alpha_k}{n}$.

Case 4: $\boldsymbol{n = k \cdot m + r}$, where $\boldsymbol{1 \leq r \leq k-1}$ and $\boldsymbol{2 \leq m \leq \frac{k-1}{2\alpha_k(\ln k)^2}}$. If $k=3$, then this means $m \leq \frac{6}{(\ln 3)^2} < 5$, so we just need to check for $n < 18$, which can be done manually. If $4 \leq k \leq 7$, then this means $m \leq \frac{k-1}{2\alpha_k(\ln k)^2} = \frac{k-1}{(\ln k)^2} < 2$, so $2 \leq m \leq \frac{k-1}{(\ln k)^2}$ is impossible, so the statement is vacuously true.

Otherwise, if $k \geq 8$, then $m \leq \frac{k-1}{(\ln k)^2} \leq k-2$, and $\lambda_{n,k} \leq \lambda_{(m+1)k,k} = (k-1) + \lambda_{m+1,k}$, where $m+1 \leq 1 + \frac{k-1}{(\ln k)^2} \leq k-1$. So, $\lambda_{m+1,k} = m$, so $\lambda_{n,k} \leq k-1+m$. However, $(k-1)\log_k n \geq (k-1) + (k-1)\log_k m$. So, $(k-1)\log_k n - \frac{\alpha_k}{n} - \lambda_{n,k} \geq (k-1)\log_k m - m - \frac{1}{2n} \geq (k-1)\log_k m - m - \frac{1}{16}$ since $n \geq k \geq 8$. Our goal is to show that this quantity is nonnegative.

Note that $(k-1)\log_k m - m - \frac{1}{16}$ is concave in m, so to show that $(k-1)\log_k m - m - \frac{1}{16} \geq 0$ for all $2 \leq m \leq \frac{k-1}{(\ln k)^2}$, it suffices to check this for $m=2$ and some $m \geq \frac{k-1}{(\ln k)^2}$. Since $k \geq 8$, we have that $\frac{k-1}{(\ln k)^2} \leq \frac{k}{(\ln 8)^2} \leq 0.5k$. So, we verify that $(k-1)\log_k m - m - \frac{1}{16} \geq 0$ for $m=2$ and $m=k/2$. For $m=2$,

$$(k-1)\log_k m - m - \frac{1}{16} = (k-1)\log_k 2 - 2 - \frac{1}{16} = \ln 2 \cdot \frac{k-1}{\ln k} - \frac{33}{16} \geq \ln 2 \cdot \frac{7}{\ln 8} - \frac{33}{16} \geq 0.$$

For $m = k/2$,

$$\begin{aligned}(k-1)\log_k m - m - \frac{1}{16} &= (k-1) - (k-1)\log_k 2 - \frac{k}{2} - \frac{1}{16} \\ &= \frac{k}{2} - \frac{17}{16} - (k-1)\log_k 2 \geq \frac{k}{2} - \frac{17}{16} - \frac{k-1}{3} \geq 0,\end{aligned} \tag{C.15}$$

since $k \geq 8$ which means $\log_k 2 \leq \frac{1}{3}$ and $\frac{k}{2} - \frac{17}{16} - \frac{k-1}{3} = \frac{k}{6} - \frac{35}{48} \geq 0$. This concludes this case.

Case 5: Remaining Cases. In this case, $n = k \cdot m + r$ for some $m \geq \frac{k-1}{2\alpha_k(\ln k)^2}$ and $1 \leq r \leq k-1$, but n neither equals $k^a + r$ nor $k^a - s$ for any $1 \leq r, s \leq k-1$. So, we know that neither m nor $m+1$ is a power of k, so by our induction hypothesis, $\lambda_{m,k} \leq \nu_{m,k}$ and $\lambda_{m+1,k} \leq \nu_{m+1,k}$. However, we know that by Lemma 2, $(\nu_{n,k} - \nu_{m+1,k} - (r-1)) \cdot (\nu_{n,k} - \nu_{m,k} - (k-r-1)) \geq r(k-r)$, so $(\nu_{n,k} - \lambda_{m+1,k} - (r-1)) \cdot (\nu_{n,k} - \lambda_{m,k} - (k-r-1)) \geq r(k-r)$. Therefore, by Lemma 1, we must have that $\lambda_{n,k} \leq \nu_{n,k} = (k-1)\log_k n - \frac{\alpha_k}{n}$. □

D
Supplementary Evolutionary Simulation Data

In this Appendix, we present additional simulation data for Chapter 6: namely, we replot Figure 6.9, Figure 6.10, and Figure 6.11 with intermediate values of σ_J and σ_h. These are, respectively, histograms for the metastable state recovery time τ, the change in fitness between the first two encountered metastable states ΔF, and the log-log correlations between ΔF and τ.

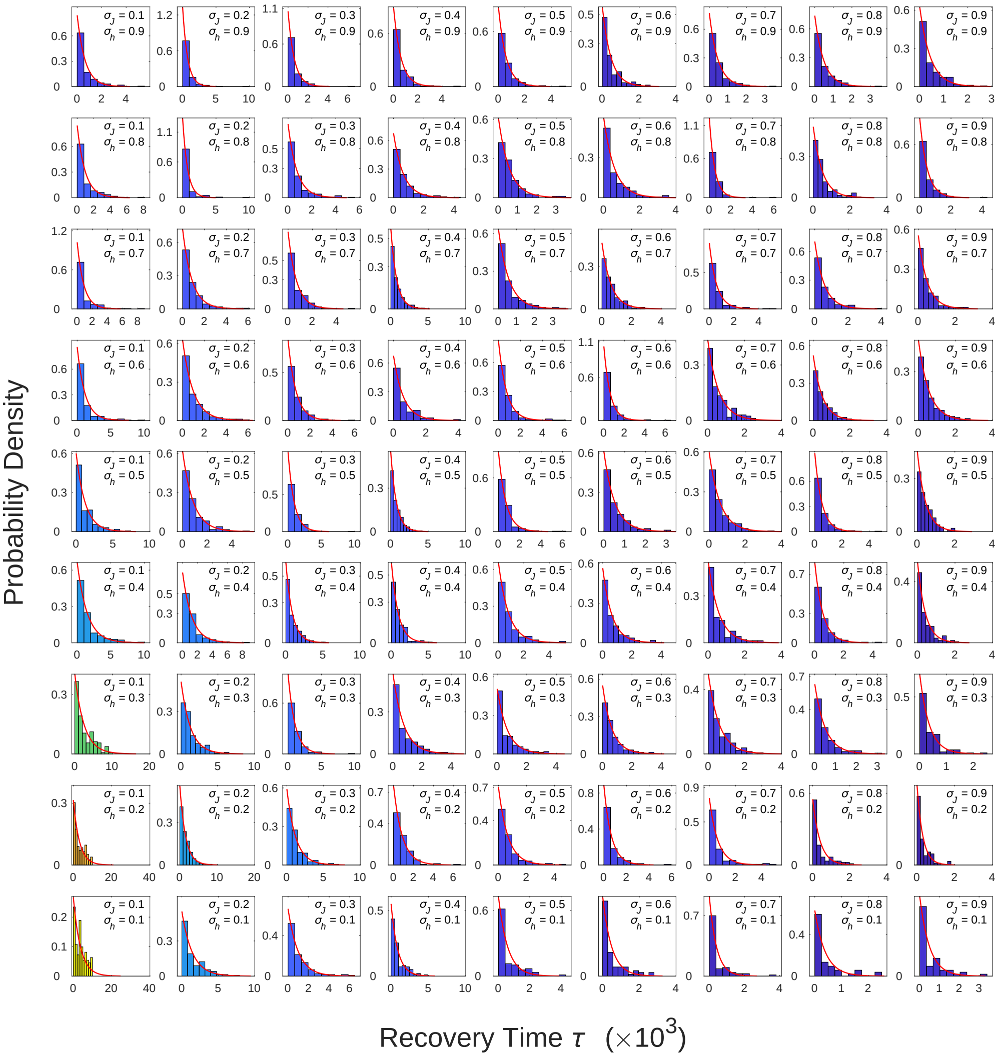


Figure D.1: Histograms of the times τ taken for metastability to be recovered by 50% of the population after perturbation of the first encountered metastable. The exponential distribution suggests that finding a metastable state after perturbation of the original metastable state is a Poisson process, or can be approximated as such.

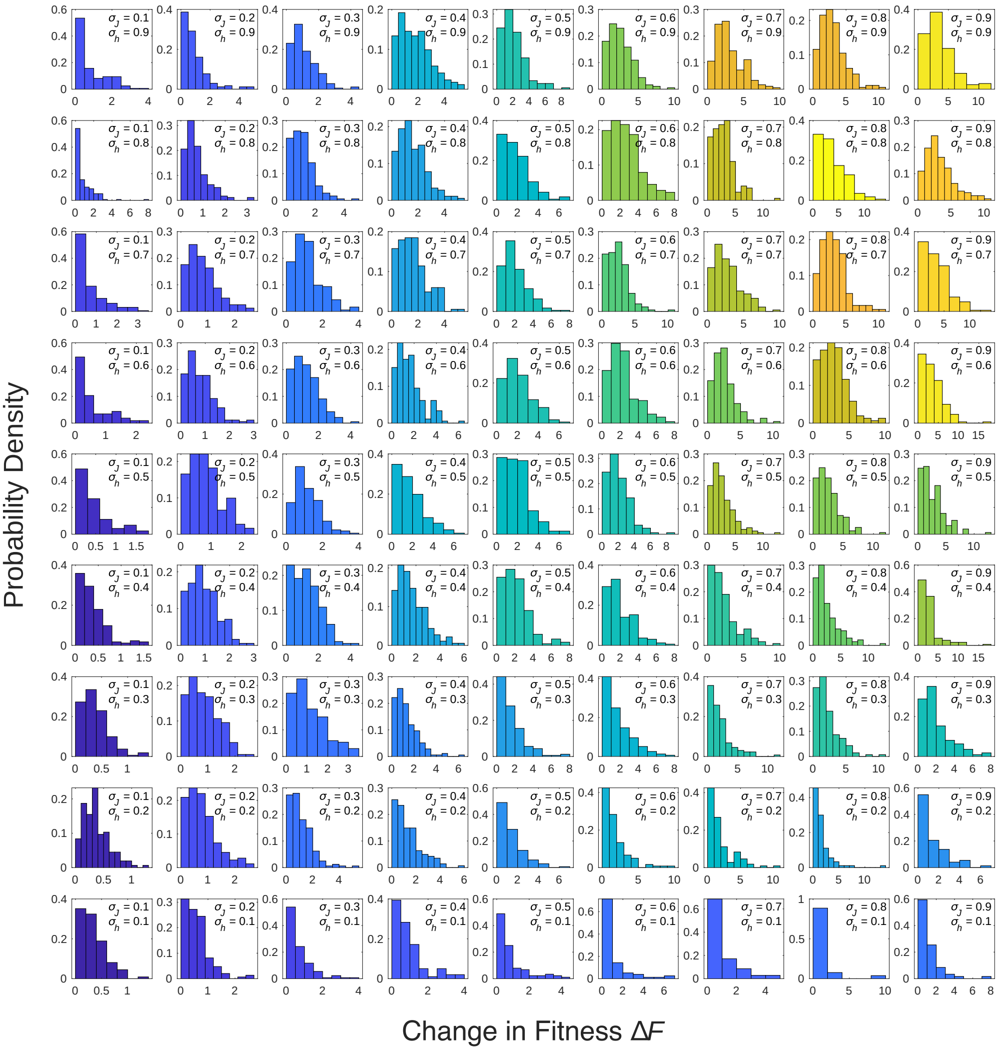


Figure D.2: Histograms of the changes in fitness $\Delta F = F_f - F_0$ experienced between the transition between the first and second metastable states encountered. The distributions tend to be right skewed and unimodal near, but not at $\Delta F = 0$, in general.

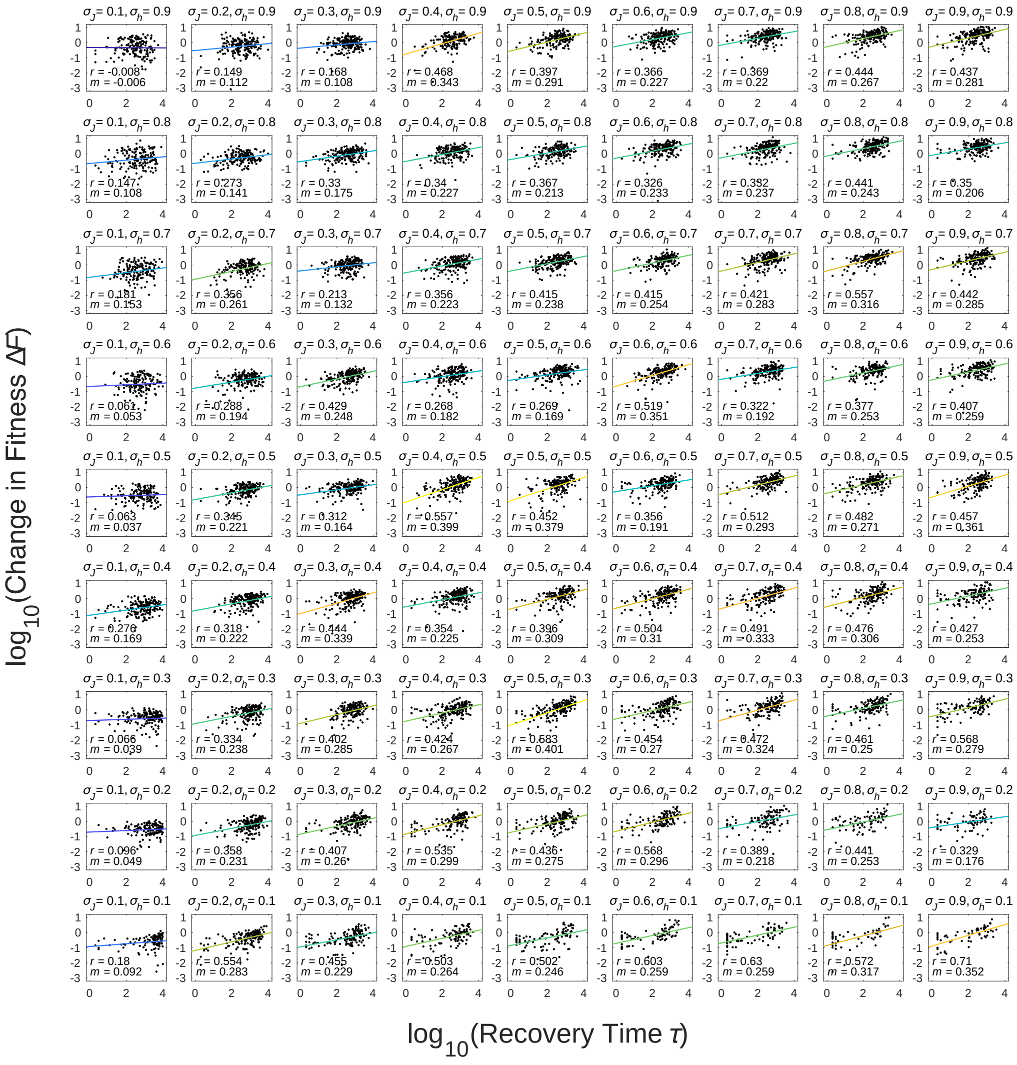


Figure D.3: Correlations between log change in fitness $\log \Delta F$ and time taken to recover metastability $\log \tau$ following a perturbation appear to be linear, supporting the idea that the fitness trajectory is approximately a power law, as suggested by [95]. In general, the correlation tends to increase as σ_h gets smaller and as σ_J gets larger.

References

[1] Sam F. Greenbury et al. “Genetic Correlations Greatly Increase Mutational Robustness and Can Both Reduce and Enhance Evolvability”. en. In: *PLOS Computational Biology* 12.3 (Mar. 2016). Ed. by Richard A Goldstein, e1004773.

[2] Vaibhav Mohanty and Ard A. Louis. “Robustness and Stability of Spin Glass Ground States to Perturbed Interactions”. In: *arXiv:2012.05437 [cond-mat]* (Dec. 2020). arXiv: 2012.05437.

[3] Sam Francis Greenbury. “General properties of genotype-phenotype maps for biological self-assembly”. en. PhD thesis. Univeresity of Cambridge, 2014.

[4] Stuart Kauffman. “Homeostasis and Differentiation in Random Genetic Control Networks”. en. In: *Nature* 224.5215 (Oct. 1969), pp. 177–178.

[5] Andreas Wagner. “Distributed robustness versus redundancy as causes of mutational robustness”. en. In: *BioEssays* 27.2 (Feb. 2005), pp. 176–188.

[6] Andreas Wagner. *Robustness and evolvability in living systems.* eng. 3. print., and 1. paperback print. Princeton studies in complexity. OCLC: 845177181. Princeton, NJ: Princeton Univ. Press, 2007.

[7] Andreas Wagner. “Robustness and evolvability: a paradox resolved”. In: *Proceedings of the Royal Society B: Biological Sciences* 275.1630 (Jan. 2008). Publisher: Royal Society, pp. 91–100.

[8] Jacobo Aguirre et al. “Topological Structure of the Space of Phenotypes: The Case of RNA Neutral Networks”. en. In: *PLoS ONE* 6.10 (Oct. 2011). Ed. by Yamir Moreno, e26324.

[9] Joshua L. Payne and Andreas Wagner. “Constraint and Contingency in Multifunctional Gene Regulatory Circuits”. en. In: *PLoS Computational Biology* 9.6 (June 2013). Ed. by Réka Albert, e1003071.

[10] Joshua L. Payne, Jason H. Moore, and Andreas Wagner. “Robustness, evolvability, and the logic of genetic regulation”. eng. In: *Artificial Life* 20.1 (2014). Number: 1 Publisher: MIT Press, pp. 111–126.

[11] Steffen Schaper and Ard A. Louis. “The Arrival of the Frequent: How Bias in Genotype-Phenotype Maps Can Steer Populations to Local Optima”. en. In: *PLoS ONE* 9.2 (Feb. 2014). Ed. by Suzannah Rutherford, e86635.

[12] Sam F. Greenbury et al. “A tractable genotype–phenotype map modelling the self-assembly of protein quaternary structure”. en. In: *Journal of The Royal Society Interface* 11.95 (June 2014), p. 20140249.

[13] S. F. Greenbury and S. E. Ahnert. “The organization of biological sequences into constrained and unconstrained parts determines fundamental properties of genotype–phenotype maps”. en. In: *Journal of The Royal Society Interface* 12.113 (Dec. 2015), p. 20150724.

[14] Kamaludin Dingle, Steffen Schaper, and Ard A. Louis. “The structure of the genotype–phenotype map strongly constrains the evolution of non-coding RNA”. In: *Interface Focus* 5.6 (Dec. 2015). Publisher: Royal Society, p. 20150053.

[15] S. E. Ahnert. “Structural properties of genotype–phenotype maps”. In: *Journal of The Royal Society Interface* 14.132 (July 2017). Publisher: Royal Society, p. 20170275.

[16] Marcel Weiß and Sebastian E. Ahnert. “Phenotypes can be robust and evolvable if mutations have non-local effects on sequence constraints”. In: *Journal of The Royal Society Interface* 15.138 (Jan. 2018), p. 20170618.

[17] Daniel Nichol et al. “Model genotype–phenotype mappings and the algorithmic structure of evolution”. en. In: *Journal of The Royal Society Interface* 16.160 (Nov. 2019), p. 20190332.

[18] Chico Q. Camargo and Ard A. Louis. “Boolean Threshold Networks as Models of Genotype-Phenotype Maps”. en. In: *Complex Networks XI* (2020). Publisher: Springer, Cham, pp. 143–155.

[19] Ting Hu, Marco Tomassini, and Wolfgang Banzhaf. “A network perspective on genotype–phenotype mapping in genetic programming”. en. In: *Genetic Programming and Evolvable Machines* (Jan. 2020).

[20] Susanna Manrubia et al. “From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics”. en. In: *Physics of Life Reviews* 38 (Sept. 2021), pp. 55–106.

[21] Joshua L. Payne and Andreas Wagner. “The causes of evolvability and their evolution”. en. In: *Nature Reviews Genetics* 20.1 (Jan. 2019), pp. 24–38.

[22] Steffen Schaper, Iain G. Johnston, and Ard A. Louis. “Epistasis can lead to fragmented neutral spaces and contingency in evolution”. en. In: *Proceedings of the Royal Society B: Biological Sciences* 279.1734 (May 2012), pp. 1777–1783.

[23] Kamaludin Dingle et al. “Phenotype bias determines how RNA structures occupy the morphospace of all possible shapes”. en. In: *bioRxiv* (Dec. 2020). Publisher: Cold Spring Harbor Laboratory Section: New Results, p. 2020.12.03.410605.

[24] Ronny Lorenz et al. “ViennaRNA Package 2.0”. In: *Algorithms for Molecular Biology* 6.1 (Nov. 2011), p. 26.

[25] Ken A. Dill. “Theory for the folding and stability of globular proteins”. en. In: *Biochemistry* 24.6 (Mar. 1985), pp. 1501–1509.

[26] Iain G. Johnston et al. “Evolutionary dynamics in a simple model of self-assembly”. In: *Physical Review E* 83.6 (June 2011). Publisher: American Physical Society, p. 066105.

[27] Joshua L. Payne and Andreas Wagner. “The Robustness and Evolvability of Transcription Factor Binding Sites”. en. In: *Science* 343.6173 (Feb. 2014). Publisher: American Association for the Advancement of Science Section: Report, pp. 875–877.

[28] Motoo Kimura. “Stochastic Processes and Distribution of Gene Frequencies Under Natural Selection”. In: (1955).

[29] Motoo Kimura. “Evolutionary Rate at the Molecular Level”. en. In: *Nature* 217.5129 (Feb. 1968), pp. 624–626.

[30] Motoo Kimura. “The neutral theory of molecular evolution: A review of recent evidence”. In: *The Japanese Journal of Genetics* 66.4 (1991), pp. 367–386.

[31] David Sherrington and Scott Kirkpatrick. “Solvable Model of a Spin-Glass”. en. In: *Physical Review Letters* 35.26 (Dec. 1975), pp. 1792–1796.

[32] S F Edwards and P W Anderson. “Theory of spin glasses”. In: *Journal of Physics F: Metal Physics* 5.5 (May 1975), pp. 965–974.

[33] Marc Mezard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond*. World Scientific lecture notes in physics v. 9. OCLC: ocm14929802. Singapore ; New Jersey: World Scientific, 1987.

[34] Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford graduate texts. OCLC: ocn234430714. Oxford ; New York: Oxford University Press, 2009.

[35] Hidetoshi Nishimori. *Statistical physics of spin glasses and information processing: an introduction*. International series of monographs on physics 111. OCLC: ocm47063323. Oxford ; New York: Oxford University Press, 2001.

[36] Yipei Guo, Marija Vucelja, and Ariel Amir. “Stochastic tunneling across fitness valleys can give rise to a logarithmic long-term fitness trajectory”. en. In: *Science Advances* 5.7 (July 2019), eaav3842.

[37] Raymond H. Y. Louie et al. “Fitness landscape of the human immunodeficiency virus envelope protein that is targeted by antibodies”. en. In: *Proceedings of the National Academy of Sciences* 115.4 (Jan. 2018), E564–E573.

[38] Thomas C. Butler et al. “Identification of drug resistance mutations in HIV from constraints on natural evolution”. en. In: *Physical Review E* 93.2 (Feb. 2016), p. 022412.

[39] John P. Barton et al. “Relative rate and location of intra-host HIV evolution to evade cellular immunity are predictable”. en. In: *Nature Communications* 7.1 (Sept. 2016), p. 11660.

[40] Karthik Shekhar et al. “Spin models inferred from patient-derived viral sequence data faithfully describe HIV fitness landscapes”. en. In: *Physical Review E* 88.6 (Dec. 2013), p. 062705.

[41] Thomas A Hopf et al. “Mutation effects predicted from sequence co-variation”. en. In: *Nature Biotechnology* 35.2 (Feb. 2017), pp. 128–135.

[42] Simona Cocco et al. “Inverse statistical physics of protein sequences: a key issues review”. en. In: *Reports on Progress in Physics* 81.3 (Mar. 2018), p. 032601.

[43] F Barahona. “On the computational complexity of Ising spin glass models”. In: *Journal of Physics A: Mathematical and General* 15.10 (Oct. 1982), pp. 3241–3253.

[44] C. Van den Broeck and R. Kawai. “Learning in feedforward Boolean networks”. en. In: *Physical Review A* 42.10 (Nov. 1990), pp. 6210–6218.

[45] Guillermo Valle-Pérez, Chico Q. Camargo, and Ard A. Louis. “Deep learning generalizes because the parameter-function map is biased towards simple functions”. In: *arXiv:1805.08522 [cs, stat]* (Apr. 2019). arXiv: 1805.08522.

[46] Susanna Manrubia and José A. Cuesta. “Distribution of genotype network sizes in sequence-to-structure genotype–phenotype maps”. en. In: *Journal of The Royal Society Interface* 14.129 (Apr. 2017), p. 20160976.

[47] Stuart Oldham et al. “Consistency and differences between centrality measures across distinct classes of networks”. en. In: *PLOS ONE* 14.7 (July 2019). Ed. by Satoru Hayasaka, e0220061.

[48] Kamaludin Dingle, Chico Q. Camargo, and Ard A. Louis. “Input–output maps are strongly biased towards simple outputs”. en. In: *Nature Communications* 9.1 (Dec. 2018), p. 761.

[49] Kamaludin Dingle, Guillermo Valle Pérez, and Ard A. Louis. “Generic predictions of output probability based on complexities of inputs and outputs”. en. In: *Scientific Reports* 10.1 (Dec. 2020), p. 4415.

[50] Noam Nisan. “On the degree of boolean functions as real polynomials”. en. In: *comput complexity* (1994), p. 13.

[51] Huang. “Induced subgraphs of hypercubes and a proof of the Sensitivity Conjecture”. en. In: *Annals of Mathematics* 190.3 (2019), p. 949.

[52] A. Bernasconi. “Sensitivity vs. block sensitivity (an average-case study)”. en. In: *Information Processing Letters* 59.3 (Aug. 1996), pp. 151–157.

[53] Sourav Chakraborty. “On the sensitivity of cyclically-invariant Boolean functions”. In: *Discrete Mathematics & Theoretical Computer Science* 13.4 (Dec. 2011), pp. 51–60.

[54] Peter Schuster et al. “From sequences to shapes and back: a case study in RNA secondary structures”. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 255.1344 (Mar. 1994), pp. 279–284.

[55] W. Grüner et al. “Analysis of RNA sequence structure maps by exhaustive enumeration I. Neutral networks”. en. In: *Monatshefte für Chemie / Chemical Monthly* 127.4 (Apr. 1996), pp. 355–374.

[56] Ehud Friedgut. “Boolean Functions With Low Average Sensitivity Depend On Few Coordinates”. en. In: *COMBINATORICA* 18.1 (Jan. 1998), pp. 27–35.

[57] E. T. Jaynes. “Information Theory and Statistical Mechanics”. In: *Physical Review* 106.4 (May 1957). Publisher: American Physical Society, pp. 620–630.

[58] R. Squier, B. Torrence, and A. Vogt. “The number of edges in a subgraph of a Hamming graph”. en. In: *Applied Mathematics Letters* 14.6 (Aug. 2001), pp. 701–705.

[59] Nicholas Metropolis et al. “Equation of State Calculations by Fast Computing Machines”. In: *The Journal of Chemical Physics* 21.6 (June 1953). Publisher: American Institute of Physics, pp. 1087–1092.

[60] W. K. Hastings. “Monte Carlo Sampling Methods Using Markov Chains and Their Applications”. In: *Biometrika* 57.1 (1970). Publisher: [Oxford University Press, Biometrika Trust], pp. 97–109.

[61] T. Reeves et al. “Eigenvalues of neutral networks: Interpolating between hypercubes”. en. In: *Discrete Mathematics* 339.4 (Apr. 2016), pp. 1283–1290.

[62] L. H. Harper. “Optimal Assignments of Numbers to Vertices”. en. In: *Journal of the Society for Industrial and Applied Mathematics* 12.1 (Mar. 1964), pp. 131–135.

[63] John H. Lindsey. “Assignment of Numbers to Vertices”. In: *The American Mathematical Monthly* 71.5 (1964). Publisher: Mathematical Association of America, pp. 508–516.

[64] Tasmin Sarkany. *Personal communications with Tasmin Sarkany.* 2021.

[65] J. B. S. Haldane. “The Effect of Variation of Fitness”. In: *The American Naturalist* 71.735 (July 1937). Publisher: The University of Chicago Press, pp. 337–349.

[66] Erik van Nimwegen, James P. Crutchfield, and Martijn Huynen. “Neutral Evolution of Mutational Robustness”. In: *Proceedings of the National Academy of Sciences of the United States of America* 96.17 (1999). Publisher: National Academy of Sciences, pp. 9716–9720.

[67] Béla Bollobás, Jonathan Lee, and Shoham Letzter. “Eigenvalues of subgraphs of the cube”. en. In: *European Journal of Combinatorics* 70 (May 2018), pp. 125–148.

[68] L. E. Bush. “An Asymptotic Formula for the Average Sum of the Digits of Integers”. In: *The American Mathematical Monthly* 47.3 (Mar. 1940). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00029890.1940.11990954, pp. 154–156.

[69] J. R. Trollope. “An Explicit Expression for Binary Digital Sums”. In: *Mathematics Magazine* 41.1 (1968). Publisher: Mathematical Association of America, pp. 21–25.

[70] Hubert Delange. “Sur la foncion sommatoire de la foncion « somme des chiffres »”. fr. In: (1975). Medium: text/html,application/pdf Publisher: Fondation L’Enseignement Mathématique.

[71] O. E. Galkin and S. Yu. Galkina. “Global extrema of the Delange function, bounds for digital sums and concave functions”. en. In: *Sbornik: Mathematics* 211.3 (Mar. 2020), pp. 336–372.

[72] Teiji Takagi. “A simple example of the continuous function without derivative,” in: 1 (1903), pp. 176–177.

[73] R. L. Graham. “On Primitive Graphs and Optimal Vertex Assignments”. en. In: *Annals of the New York Academy of Sciences* 175.1 (July 1970), pp. 170–186.

[74] Sergiu Hart. “A note on the edges of the n-cube”. en. In: *Discrete Mathematics* 14.2 (Jan. 1976), pp. 157–163.

[75] Marcel Weiß and Sebastian E. Ahnert. “Neutral components show a hierarchical community structure in the genotype–phenotype map of RNA secondary structure”. en. In: *Journal of The Royal Society Interface* 17.171 (Oct. 2020), p. 20200608.

[76] Robert Giegerich, Björn Voß, and Marc Rehmsmeier. “Abstract shapes of RNA”. In: *Nucleic Acids Research* 32.16 (Aug. 2004), pp. 4843–4851.

[77] Stefan Janssen and Robert Giegerich. “The RNA shapes studio”. In: *Bioinformatics* 31.3 (Feb. 2015), pp. 423–425.

[78] Christoph Adami. “Information theory in molecular biology”. en. In: *Physics of Life Reviews* 1.1 (Apr. 2004), pp. 3–22.

[79] B. D. Hughes. *Random walks and random environments.* Oxford : New York: Clarendon Press ; Oxford University Press, 1995.

[80] Manfred Eigen. “Selforganization of matter and the evolution of biological macromolecules”. en. In: *Naturwissenschaften* 58.10 (Oct. 1971), pp. 465–523.

[81] Joel Friedman and Jean-Pierre Tillich. “Generalized Alon–Boppana Theorems and Error-Correcting Codes”. en. In: *SIAM Journal on Discrete Mathematics* 19.3 (Jan. 2005), pp. 700–718.

[82] Yoonsoo Nam. *Personal communications with Yoonsoo Nam.* 2021.

[83] B. B. Machta et al. “Parameter Space Compression Underlies Emergent Theories and Predictive Models”. en. In: *Science* 342.6158 (Nov. 2013), pp. 604–607.

[84] Jonathan Frankle and Michael Carbin. “The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks”. In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019* (2019).

[85] S. Cocco and R. Monasson. “Adaptive cluster expansion for inferring boltzmann machines with noisy data”. eng. In: *Physical Review Letters* 106.9 (Mar. 2011), p. 090601.

[86] S. Cocco and R. Monasson. “Adaptive Cluster Expansion for the Inverse Ising Problem: Convergence, Algorithm and Tests”. en. In: *Journal of Statistical Physics* 147.2 (Apr. 2012), pp. 252–314.

[87] Hugo Jacquin et al. “Benchmarking Inverse Statistical Approaches for Protein Structure and Design with Exactly Solvable Models”. en. In: *PLOS Computational Biology* 12.5 (May 2016). Ed. by Debora S. Marks, e1004889.

[88] John Barton and Simona Cocco. “Ising models for neural activity inferred via selective cluster expansion: structural and coding properties”. en. In: *Journal of Statistical Mechanics: Theory and Experiment* 2013.03 (Mar. 2013), P03002.

[89] Andrew L. Ferguson et al. “Translating HIV Sequences into Quantitative Fitness Landscapes Predicts Viral Vulnerabilities for Rational Immunogen Design”. en. In: *Immunity* 38.3 (Mar. 2013), pp. 606–617.

[90] A. J. Bray and M. A. Moore. “Metastable states in spin glasses”. en. In: *Journal of Physics C: Solid State Physics* 13.19 (July 1980). Publisher: IOP Publishing, pp. L469–L476.

[91] A J Bray and M A Moore. “Metastable states in spin glasses with short-ranged interactions”. en. In: *Journal of Physics C: Solid State Physics* 14.9 (Mar. 1981), pp. 1313–1327.

[92] D S Dean. “On the metastable states of the zero-temperature SK mode”. en. In: *Journal of Physics A: Mathematical and General* 27.23 (Dec. 1994), pp. L899–L905.

[93] F Tanaka and S F Edwards. “Analytic theory of the ground state properties of a spin glass. I. Ising spin glass”. en. In: *Journal of Physics F: Metal Physics* 10.12 (Dec. 1980), pp. 2769–2778.

[94] C. Amitrano, L. Peliti, and M. Saber. “Population dynamics in a spin-glass model of chemical evolution”. en. In: *Journal of Molecular Evolution* 29.6 (Dec. 1989), pp. 513–525.

[95] Benjamin H. Good and Michael M. Desai. “The Impact of Macroscopic Epistasis on Long-Term Evolutionary Dynamics”. en. In: *Genetics* 199.1 (Jan. 2015), pp. 177–190.

[96] Sergey Gavrilets. *Fitness Landscapes and the Origin of Species.* English. Princeton, N.J: Princeton University Press, July 2004.

[97] Sergey Gavrilets. “Evolution and speciation on holey adaptive landscapes”. en. In: *Trends in Ecology & Evolution* 12.8 (Aug. 1997), pp. 307–312.

[98] Richard Courant and David Hilbert. *Methods of mathematical physics. Vol.1.* eng. Vol. 1. Weinheim: Wiley-VCH, 2009.