SAPIENZA
UNIVERSITÀ DI ROMA

Statistical mechanics for biological applications: Focusing on the immune system

Dipartimento di Scienze di Base e Applicate per l'Ingegneria (SBAI) - sezione di Matematica

Dottorato di Ricerca in Modelli e Metodi Matematici per la Tecnologia e la Società – XXVI Ciclo

Candidate

Lorenzo Asti

ID number 698834

Thesis Advisor

Dr. Adriano Barra

Co-Advisor

Dr. Elena Agliari

A thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Applied Mathematics

November 2013

Thesis defended on 12 February 2014
in front of a Board of Examiners composed by:

Livio Triolo (chairman)
Brunello Tirozzi
Francesco Vaccarino

Statistical mechanics for biological applications: Focusing on the immune system
Ph.D. thesis. Sapienza – University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: lorenzo.asti@sbai.uniroma1.it

To the sixtieth birthday
of a wonderful woman.

Acknowledgments

I would like to thank my advisors Adriano Barra of Dipartimento di Fisica of Sapienza, Università di Roma and Elena Agliari of Dipartimento di Fisica of Università di Parma for providing me a guide through these crucial years of PhD training.

A special thanks is reserved to Andrea Pagnani of Human Genetics Foundation (HuGeF) and Politecnico di Torino who gave me the possibility of spending a very pleasant time in Turin working on a consistent part of this Thesis and learning many things. Thanks also to the entire group in HuGeF.

Thanks to Guido Uguzzoni who shared with me much of the way throughout the PhD training and, with this, several helpful scientific (and many non-scientific!) discussions together with many moments of both satisfaction and frustration. And a lot of fun! It was nice to have such a mate along these years.

Thanks to my mother Elisabetta Affabris of Dipartimento di Biologia, Università di Roma Tre for all the support and for having taught me a lot of biology. And, of course, for much, much more...

I also wish to thank Prof. Daniela Giachetti of Dipartimento di Scienze di Base e Applicate per l'Ingegneria (SBAI), director of my PhD school, for having always been a helpful and kind reference point.

Lastly, as I believe that a lot in my work is the result of the high quality of many courses that I attended to during my university career, I would like here to acknowledge some among the best teachers that I had in Sapienza, Università di Roma: Thank you Massimo Testa, Carlo Di Castro, Omar Benhar, Andrea Cavagna, Giulio D'Agostini and many others. After my short experience in teaching, I know that effective, understandable and enjoyable lessons are not easy to prepare and I believe that they are the fruit of a sincere engagement in many hours of work. Thank you teachers for that.

Contents

Introduction

The emergence in the last decades of a huge amount of data in many fields of biology triggered also an increase of the interest by quantitative disciplines for life sciences. Mathematics, physics and informatics have been providing quantitative models and advanced statistical tools in order to help the understanding of many biological problems. Statistical mechanics is a field that particularly contributed to quantitative biology because of its intrinsic predisposition in dealing with systems of many strongly interacting agents, noise, information processing and statistical inference.

In this Thesis a collection of works at the interphase between statistical mechanics and biology is presented. In particular they are related to biological problems that can be mainly reconducted to the biology of the immune system.

Beyond the unification key given by statistical mechanics of discrete systems and quantitative modeling and analysis of the immune system, the works presented here are quite diversified. The origin of this heterogeneity resides in the intent of using and learning many different techniques during the lapse of time needed for the preparation of the work reviewed in this Thesis. In fact the work presented in Chapter 3 mainly deals with statistical mechanics, networks theory and networks numerical simulations and analysis; Chapter 4 presents a mathematical physics oriented work; Chapter 5 and 6 deal with data analysis and in particular wth clinical data and amino acid sequences data sets, requiring the use of both analytical and numerical techniques.

The Thesis is conceptually organized in two main parts. The first part (Chapters 1 and 2) is dedicated to the review of known results both in statistical mechanics and biology, while in the second part (Chapters 3, 4 and 6) the original works are presented together with briefs insights into the research fields in which they can be embedded.

In particular, in Chapter 1 some of the most relevant models and techniques in statistical mechanics of mean field spin systems are reviewed, starting with the Ising model and then passing to the Sherrington-Kirkpatrik model for spin glasses and to the Hopfield model for attractors neural networks. The *replica method* is presented together with the *stochastic stability method* as a mathematically rigorous alternative to replicas.

Chapter 2 is dedicated to a very schematic overview of the biology of the immune system.

In Chapter 3, Section 3.1 is dedicated to the presentation of a mathematical phenomenological model for the study of the idiotypic network while Section 3.2 serves as a review of the statistical mechanics based models proposed by Elena

Agliari and Adriano Barra as toy models meant to underline the possible role of complex networks within the immune system.

In Chapter 4 the mathematical model of an analogue neural network on a diluted graph is studied. It is shown how the problem can be mapped in a bipartite diluted spin glass. The model is rigorously solved at the replica symmetric level with the use of the *stochastic stability technique* and fluctuations analysis is used to study the spin glass transition of the system. A topological analysis of the network is also performed and different topological regimes are proven to emerge though the tuning of the model parameters.

In Chapter 5 a model for the analysis of clinical records of testing sets of patients is presented. The model is based on a Markov chain over the space of clinical states. The machinery is applied to data concerning the insurgence of Tuberculosis and Non-Tuberculous Infections as side effects in patients treated with Tumor Necrosis Factor inhibitors. The analysis procedure is capable of capturing clinical details of the behaviors of different drugs.

Lastly, Chapter 6 is dedicated to a statistical inference analysis on deep sequencing data of an antibodies repertoire with the purpose of studying the problem of antibodies affinity maturation. A partial antibodies repertoire from a HIV-1 infected donor presenting broadly neutralizing serum is used to infer a probability distribution in the space of sequences that is compared with neutralization power measurements and with the deposited crystallographic structure of a deeply matured antibody. The work is still in progress, but preliminary results are encouraging and are presented here.

The works reviewed in this thesis have been published or are in preparations and are the followings:

[1] Elena Agliari, Lorenzo Asti, Adriano Barra, and Luca Ferrucci. Organization and evolution of synthetic idiotypic networks. *Physical Review E*, 85(5):051909, 2012.

[2] Elena Agliari, Lorenzo Asti, Adriano Barra, Raffaella Burioni, and Guido Uguzzoni. Analogue neural networks on correlated random graphs. *Journal of Physics A: Mathematical and Theoretical*, 45(36):365001, 2012.

[3] Elena Agliari, Lorenzo Asti, Adriano Barra, Rossana Scrivo, Guido Valesini, and Robert S. Wallis. Application of a stochastic modeling to assess the evolution of tuberculous and non-tuberculous mycobacterial infection in patients treated with tumor necrosis factor inhibitors. *PloS one*, 8(1):e55017, 2013.

[4] Lorenzo Asti, Paolo Marcatili, Andrea Pagnani, and Guido Uguzzoni. Multivariate Gaussian Modeling for abs affinity maturation. *In preparation*

Chapter 1

Statistical mechanics of spin systems

The aim of this Chapter is that of introducing the statistical mechanics of the Hopfield model, a paradigm for attractor neural networks whose proposal for an immunological interpretation will be described in Chapter 3. As it is a complex mean field spin model, we will first review the statistical mechanics of the simplest mean field spin model, the Ising model; then we will introduce complex mean field spin models and the *replica technique* for solving them and finally we will review the solution of the mean field Hopfield model in the low and high (at replica symmetric level) storage limit.

1.1 Ordinary spin systems

1.1.1 The paradigm of spin systems: the Ising model

The Ising model was originally introduced as statistical mechanics model to study the ferromagnetic transition in magnetic materials with planar symmetry by Ernst Ising after the suggestion of Wilhem Lenz, Ising's s doctoral thesis supervisor. Ising solved it in the one dimensional case where the model displays no phase transition [5].

As spins are binary variables, after a while, the applicability of the model suddenly appeared to be wider than the pure physical one as large systems of interacting binary variables can be encountered in many field of quantitative science, from computer science [6], to the modeling of biological [7], social or economic systems [8].

The model is introduced by a Hamiltonian of binary variables (spins) $\sigma_i = \pm 1$, $i = 1 \dots N$, that is composed of an interaction (two-bodies) term and an external field (one-body) term:

$$H_{\text{Ising}}(\boldsymbol{\sigma}; J, h) = -J \sum_{(ij)} \sigma_i \sigma_j - h \sum_{i=1}^{N} \sigma_i \,, \tag{1.1}$$

where the sum in the interaction term extends over the connected couples of spins in a cubic lattice structure. The coupling strength J and the external field h are the

parameters of the model and they tune respectively the strength of the interaction and of the external polarization source. The interest for this kind of models resides in studying their statistical mechanics, namely in computing the partition function that is the sum, in the thermodynamic (infinite system) limit, of the Boltzmann weight of all the possible configuration, $\boldsymbol{\sigma}$, of the system,

$$Z = \sum_{\sigma_1=\pm1\ldots\sigma_N=\pm1} e^{-\beta H(\boldsymbol{\sigma})} \equiv \sum_{\{\sigma\}} e^{-\beta H(\boldsymbol{\sigma})} \,, \tag{1.2}$$

where $\beta \equiv 1/K_B T$ is the inverse temperature in unit of the Boltzmann constant K_B that from now on we will set equal to one. From the partition function the thermodynamical behavior of the system is recovered by computing the intensive free energy as a function of the temperature and the other system parameters through the relation

$$f = \lim_{N\to\infty} -\frac{1}{\beta N} \log Z \,. \tag{1.3}$$

Moreover the Boltzmann measure allows to compute any thermodynamical quantity $\mathcal{O}$ as an average over the phase space of its microscopic value. In this chapter will be indicated with the symbol

$$\langle \mathcal{O} \rangle = \sum_{\{\sigma\}} \mathcal{O}(\boldsymbol{\sigma})\, e^{\beta H(\boldsymbol{\sigma})} \tag{1.4}$$

Of course the behavior of the system strongly depends on the dimensionality of the lattice: in 1D the solution is easy to find [9] and the system displays no phase transition; in 2D the Onsager procedure [10] permits to find the exact solution which shows that a second order para/ferromagnetic phase transition takes place; in 3D the exact solution is still not know and perturbative techniques have been developed based on Renormalization Group, one of the most successful ideas in the physics of the twentieth century. In dimension $D > 4$ the critical (near the transition) behavior is known.

The easiest step for the solution of the model is the so called *mean filed* approximation that is equivalent to consider a fully connected model, where every spin is connected to all the others with a constant coupling of order $\mathcal{O}(1/N)$. The Hamiltonian for the mean field ferromagnetic Ising model, also known as *Curie-Weiss model* is so:

$$H_{\mathrm{CW}}(\boldsymbol{\sigma}; J, h) = -\frac{J}{N} \sum_{i<j} \sigma_i \sigma_j - h \sum_{i=1}^{N} \sigma_i \,. \tag{1.5}$$

Let us consider for the moment the case of vanishing external field, $h = 0$, in which the Hamiltonian is invariant under the operation of global spin flip ($\sigma_i \to -\sigma_1, \forall i$). The difficulty in performing the partition sum of interacting systems arises from the two body term $\sim \sum_{\{\sigma\}} e^{\sum_{ij} \sigma_i \sigma_j}$. Nevertheless the partition function for this mean field model can be computed if we introduce the magnetization of a configuration as the fraction of positive spins with respect to the negative ones, namely

$$m(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i \,. \tag{1.6}$$

In this way the partition function can be written as

$$\begin{aligned} Z(\beta, J, h=0) &= \sum_{\{\sigma\}} e^{\frac{\beta J}{N}\sum_{i<j}^{N}\sigma_i\sigma_j} = \sum_{\{\sigma\}} e^{\frac{\beta J}{2N}\sum_{i,j=1}^{N}\sigma_i\sigma_j - \frac{\beta J}{N}\sum_{i=1}^{N}s_i^2} = \\ &= \sum_{\{\sigma\}} e^{\frac{\beta J}{2N}\sum_{i,j=1}^{N}\sigma_i\sigma_j - \beta J} = \sum_{\{\sigma\}} e^{\frac{\beta J N}{2}m^2(\boldsymbol{\sigma})} . \end{aligned} \tag{1.7}$$

It has to be noted that in the last passage a constant (non depending on N) multiplicative contribution to the partition function has been neglected; this happens because terms of this kind give vanishing contributions, in the thermodynamic limit, to the intensive free energy f defined in 1.3. We will implicitly do that in the following along all this Thesis.

Every value of the magnetization corresponds to a different number of configuration, so if we now sum over the magnetization value, instead of the configuration we have to consider its degeneration. For doing so we use the Dirac-δ having the property

$$f(x) = \int_{-\infty}^{+\infty} dy f(y)\delta(x-y) \tag{1.8}$$

and whose Fourier representation reads as

$$\delta(x-y) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} dt e^{i(x-y)t} = \frac{1}{2\pi}\int_{-i\infty}^{+i\infty} dt e^{(x-y)t} . \tag{1.9}$$

Using the Dirac-δ, the degeneration of the magnetization can be taken into account by writing

$$\begin{aligned} Z(\beta, J, h=0) &= \int_{-\infty}^{+\infty} dm \sum_{\{\sigma\}} \delta\big(Nm - \sum_{i=1}^{N}\sigma_i\big)\, e^{\frac{\beta J N}{2}m^2} \\ &= \int_{-\infty}^{+\infty} dm \int_{-i\infty}^{+i\infty} dx \sum_{\{\sigma\}} e^{Nmx - x\sum_i \sigma_i + \frac{\beta J N}{2}m^2} . \end{aligned} \tag{1.10}$$

Now the sum over the spin configurations involves no more than a one-body term and it can easily performed as

$$\sum_{\{\sigma\}} e^{-x\sum_{i=1}^{N}\sigma_i} = \prod_{i=1}^{N}\sum_{\sigma_i=\pm 1} e^{-x\sigma_i} = \prod_{i=1}^{N} 2\cosh x = N 2\cosh x , \tag{1.11}$$

so that now

$$\begin{aligned} Z(\beta, J, h=0) &= \int_{-\infty}^{+\infty} dm \int_{-i\infty}^{+i\infty} dx\, e^{N(mx + \log\cosh x - \log 2 + \frac{\beta J N}{2}m^2)} \equiv \\ &\equiv \int_{-\infty}^{+\infty} dm \int_{-i\infty}^{+i\infty} dx\, e^{-N\beta f(m,x;\beta,J)} , \end{aligned} \tag{1.12}$$

where

$$f(m, x;\, \beta, J) = -\frac{mx}{\beta} - \frac{1}{\beta}\log\cosh x - \frac{J}{2}m^2 - \frac{\log 2}{\beta} . \tag{1.13}$$

As now the exponent in (1.12) is linear in the system size N, the integral can be solved by *saddle point (or Laplace) method*, which implies that, in the thermodynamic limit, the leading contribution to the free energy comes from substituting the integral in (1.12) with the maximum value of the integrand function, so, by neglecting the integral sign and minimizing $f(m,x)$[1]. Following this method, the extremization of $f(m,x)$ in (1.13) gives

$$\begin{cases} x &= -m\beta J \\ m &= -\tanh x \end{cases} \tag{1.15}$$

and hence

$$m = \tanh(\beta J m)\,. \tag{1.16}$$

So the physical free energy $f(\beta, J)$ is obtained by solving this equation and plugging the result in

$$\begin{aligned} f(m;\,\beta,J) &= f(m, x=-m\beta J;\,\beta,J) = \\ &= \frac{J}{2}m^2 - \frac{1}{\beta}\log\cosh(\beta J m) - \frac{1}{\beta}\log 2\,; \end{aligned} \tag{1.17}$$

this is the form of the intensive free energy as a function of the order parameter (in this case magnetization) and the other macroscopic parameters. By writing $f(m;\,\beta,J) = e(m;\,\beta,J) - \frac{1}{\beta}s(m;\,\beta,J)$, with

$$e(m;\,\beta,J) = -\frac{J}{2}m^2\,, \tag{1.18}$$

$$s(m;\,\beta,J) = -\beta J m^2 + \log\cosh(\beta J m) + \log 2\,, \tag{1.19}$$

it is possible to isolate an energetic contribution and an entropic one, where the latter logarithmically counts the number of effectively accessible configurations $\mathcal{N}$ of a system with a determined macroscopic magnetization: $s = \log(\mathcal{N})/N$. Consistently one could check that at infinite temperature ($\beta = 0$) every configuration is equally accessible so that $s = \log 2$ and $\mathcal{N}(\beta = 0) = 2^N$, while at zero temperature ($\beta \to \infty$) the entropy vanishes and only a finite number of configuration is accessible.

Equations such as (1.16) are known in statistical mechanics as *self consistency relations*. Solving it, checking that the solution is stable, in the sense that it minimizes $f(m)$, and that it represents the absolute minimum of the free energy gives the physical value of the order parameter, namely the one that is exponentially (in the system size) the most probable to find by picking at random a configuration of the system with the Boltzmann probability.

Expanding the self consistency equation (1.16) for small m gives

$$m = \beta J m - \frac{\beta J}{3}m^3\,. \tag{1.20}$$

[1] In fact for a regular function $g(x)$ that achieves its maximum value in x^*,

$$\begin{aligned} \log\int_{-\infty}^{+\infty} dx\, e^{Ng(x)} &= \log\int_{-\infty}^{+\infty} dx\, e^{Ng(x^*)+\frac{1}{2}Ng''(x^*)(x-x^+)^2+\dots} = \log\left(e^{Ng(x^*)}\sqrt{\frac{2\pi}{Ng''(x^*)}}\right) = \\ &= Ng(x^*) - \frac{1}{2}\log N + \text{const} + \dots \end{aligned} \tag{1.14}$$

so the computation of leading term in $N >> 1$ does not involve any integration procedure.

By studying the solution of this equation for different values of the temperatures it is possible to observe that two very different thermodynamical regimes are separated by the presence of a critical temperature T_c:

- For $\beta < \beta_c = 1/J$ the only solution is $m = 0$ and it is stable. This correspond to a high temperature ($T > T_c = J$) paramagnetic phase in which the system displays a null magnetization and it is ergodic in the sense that all the configurations with a non-null statistical measure are connected by the (local) operation of a single spin flip. By performing on a finite system a dynamic that is consistent with the Boltzmann measure (Glauber [11], Metropolis-Hastings [12]), independently on the initial condition, after a transient, one should observe the system exploring configurations where the fluctuations of the magnetization are small, namely they vanish with the system size as $\delta m \to \mathcal{O}(1/\sqrt{N}) \to 0$.

  In the paramagnetic phase the free energy is simply

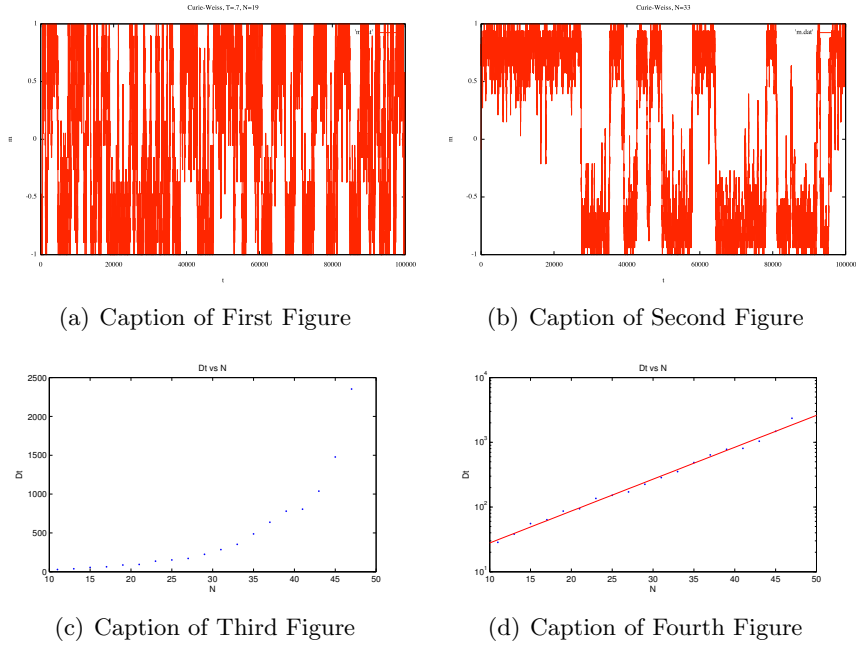
$$f(\beta < \beta_c, J) = -\frac{1}{\beta} \log 2 \,. \tag{1.21}$$

- For $\beta > \beta_c = 1/J$ the $m = 0$ solution is no longer stable and two symmetric stable solution $m = \pm m^*$ appear, each corresponding to the system being polarized in one of the two possible directions, spontaneously breaking the symmetry of the Hamiltonian. As the two solutions are mapped one into the other by the transformation under which the Hamiltonian is invariant (the global spin flip: $m \to -m$), the two solutions give the same value of the free energy. This phase is called ferromagnetic.

  In this phase the system is no longer ergodic. In fact the Boltzmann measure displays two "peaks" in the phase space that are not connected by local moves. In this regime a dynamics at finite size will show the system fluctuating around each of the two symmetric magnetization and jumping between them. Increasing the system size N, fluctuations around one solution vanish as $\delta m \to \mathcal{O}(1/\sqrt{N}) \to 0$ and the time of permanence in one of the two state grows exponentially as $\sim e^N$ so that a very large system becomes trapped in one of the two state and only one "peak" of the Boltzmann measure is explored: the system remains confined to a restricted region of the phase space, being this region smaller as the temperature is lower. In the limiting case of $T = 0$ the system explore only one configuration, namely one of the two ground states $\sigma_i = 1 \,(\text{or}\, \sigma_i = -1), \forall i$. This fact is know as *ergodicity breaking* and the (two, in this case) different sets of available configurations in which the system gets restricted are know in statistical mechanics as *pure states* or *ergodic components*. The result of a dynamical simulation in this condition is depicted in Fig. 1.1.

Equation (1.16) can be solved without approximation by graphical methods and the above discussion still holds.

Of course the saddle point method implies the necessity to sum over all the (two) equivalent saddle points when computing macroscopic observable, so that, also



(a) Caption of First Figure

(b) Caption of Second Figure

(c) Caption of Third Figure

(d) Caption of Fourth Figure

Figure 1.1. In this Figure the result of a simulation of the fully connected Ising model with Metropolis dynamics is show. In the upper panels the temporal behavior of the magnetization is show for a system of $N = 19$ - 1.1(a) - and $N = 33$ - 1.1(b) - spins. In the lower panels the scaling of the average jumping time Dt with the system size N is shown in regular - 1.1(c) - and logarithmic - 1.1(c) - scale. In a system with very large size, only configurations belonging to one of the two magnetization states are would be effectively explored.

in the low temperature phase, $\langle m(\boldsymbol{\sigma})\rangle_{Boltzmann} = 0$ while physically one observes $\langle m(\boldsymbol{\sigma})\rangle \neq 0$. This inconsistency is solved by considering that a physical system where the external field is exactly null does not exist and also in the simulations the choice of a non symmetric the initial condition (or the first fluctuation from a symmetric one) has an effect analogous to a small external field. If we consider the system to be in a non null external field $h > 0$ the free energy and the self consistency equation modify according to

$$f(m;\, \beta, J, h) = \frac{J}{2}m^2 - \frac{1}{\beta}\log\cosh(\beta J m + h) - \frac{1}{\beta}\log 2 \tag{1.22}$$

and

$$m = \tanh(\beta J m + h) \tag{1.23}$$

respectively. The presence of a postive external field explicitly breaks the symmetry of the Hamiltonian so that, in this case, in the low temperature phase, the two saddle points value of the magnetization , $m_1 > m_2$, are no longer equivalent: the state corresponding to the greater one has a lower free energy, $f(m_1) < f(m_2)$, and so is the one that gives the leading contribution to the free energy, while the contribution

of the smaller one vanishes exponentially; in fact

$$\begin{aligned} f(\beta, J, h > 0) &= \lim_{N\to\infty} -\frac{1}{N\beta} \log\left[e^{-N\beta f(m_1)} + e^{-N\beta f(m_1)}\right] = \\ &= \lim_{N\to\infty} f(m_1) - \frac{1}{N\beta} \log\left[1 + e^{-N\beta(f(m_2)-f(m_1))}\right] = \\ &\simeq \lim_{N\to\infty} f(m_1) - \frac{1}{N} e^{-N\beta(f(m_2)-f(m_1))} = f(m_1)\,. \end{aligned} \tag{1.24}$$

The consistent way to have a result with spontaneous symmetry breaking in case of vanishing external field, $h = 0$ is to compute the free energy in the thermodynamic limit at non zero field and then to send its value to zero:

$$f(\beta, J, h = 0) = \lim_{h\to 0} f(\beta, J, h \neq 0) = \lim_{h\to 0} \lim_{N\to\infty} f(\beta, J, h \neq 0, N)\,; \tag{1.25}$$

in this way, for example, in the low temperature phase $\langle m(\boldsymbol{\sigma})\rangle \neq 0$.

The ferromagnetic phase transition in the Ising model is a *second order phase transition*. In fact a discontinuity appear, at the critical point in the susceptibility $\chi = \partial m/\partial h = \partial^2 f/\partial h^2$ that is the second derivative of the free energy with respect to the controll parameter of the system, the external magnetic field h. In chis case the order parameter, m, grows continuously from zero. As we will see in the following, *fist order phase transitions* exist in which the order parameter display a jump at the critical point.

1.1.2 From two to many attractors: the low storage Hopfield model

In Section 1.1.1 we have solved the statistical mechanics of the simplest mean field spin model, the Ising model. In that case the ground state of the Hamiltonian obviously corresponds to the two states of maximum polarization where every spin is positively (negatively) oriented, $\sigma_i = 1\,(-1)$, $\forall i$ and $m(\boldsymbol{\sigma}) = 1\,(-1)$. The above analysis should have convinced the reader that a large dynamical spin system with the equilibrium distribution corresponding to the Boltzmann distribution of the mean field Ising model, at small temperature, starting from a random configuration, eventually falls in the basin of attraction of one of the two ground state and then fluctuates around it. The selection the initial configuration of the dynamics in one of the two basins, assures that, after a transient, one specific ground state is reached by the system and the symmetry is broken in the chosen direction.

An interesting extension of this situation would be achieved by considering a spin system with many (more than two symmetrical) ground states. If we could easily tune the parameters in the Hamiltonian such that arbitrary configurations of our choice are the ground states of the system, this would be a good toy model for an *associative neural network*. In fact this minimum energy configurations could be thought as *stored patterns* and, by appropriately choosing the initial configuration, the system would *remember* one of these patterns.

This scenario is actually realized by the low storage Hopfield model defined by the Hamiltonian

$$H_{\text{Hopfield}}(\boldsymbol{\sigma}; \boldsymbol{\xi}, \boldsymbol{h}) = -\frac{1}{N} \sum_{i<j}^{N} \left(\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu\right) \sigma_i \sigma_j - \sum_{i=1}^{N} \left(\sum_{\mu=1}^{p} h_i^\mu\right) \sigma_i\,, \tag{1.26}$$

where we note $\boldsymbol{\xi} = (\xi_1^1, \dots, \xi_N^1; \dots; \xi_1^p, \dots, \xi_N^p)$, $\boldsymbol{\xi^\mu} = (\xi_1^\mu, \dots, \xi_N^\mu)$, $\boldsymbol{\xi_i} = (\xi_i^1, \dots, \xi_i^p)$, $\boldsymbol{h} = (h_1^1, \dots, h_N^1; \dots; h_1^p, \dots, h_N^p)$, $\boldsymbol{h^\mu} = (h_1^\mu, \dots, h_N^\mu)$, $\boldsymbol{h_i} = (h_i^1, \dots, h_i^p)$.

In this model the usual Ising couplings are defined as a sum of projectors over p vectors in the N-dimensional space,

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu \,, \tag{1.27}$$

and the Ising external fields are simply the sum over p vectors in the same space,

$$h_i = \sum_{\mu=1}^{p} h_i^\mu \,. \tag{1.28}$$

It is straightforward to check that, if the pattern entries $\xi_i^\mu \in \{-1, 1\}$, at zero external field the ground state ground state configuration (the global minimum of the energy function 1.26) corresponds to the cases where the spins are parallel (or antiparallel) to one vector $\boldsymbol{\xi}^\mu$. Of course, depending on the choice of the bit-strings $\boldsymbol{\xi}$, more degenerate ground states could coexist. Moreover the configurations that are parallel to the other patterns correspond metastable states, namely local minimum of the energy. So, by performing a dynamic with an initial condition close to a stored patterns, at low temperature, the complete pattern will be retrieved. We call the states correlated with the p stored memories *pure states*.

Unfortunately, together with the (wanted) pure states, the above procedure causes the presence of unwanted *mixture states* that are correlated with local minima of the energy that correspond to combinations of the stored patterns.

Solution of the low storage Hopfield model with uncorrelated and unbiased random patterns

We now aim to solve the statistical mechanics of the Hopfield model where the pattern entries are drawn from an unbiased distribution in an uncorrelated way:

$$P(\xi_i^\mu) = \frac{1}{2}\delta_{\xi_i^\mu,1} + \frac{1}{2}\delta_{\xi_i^\mu,-1} \,, \tag{1.29}$$

so that

$$\langle \Phi(\boldsymbol{\xi}_i) \rangle_{\boldsymbol{\xi}} = 2^{-p} \sum_{\boldsymbol{\xi}_i \in \{1,-1\}} \Phi(\boldsymbol{\xi}_i) \qquad \text{and} \qquad \langle \xi_i^\mu \xi_i^\nu \rangle = \delta_{\mu\nu} \,. \tag{1.30}$$

Following a procedure analogous to the one used to find the solution of the mean field Ising model in 1.1.1, we can introduce p order parameters known as pattern overlaps

$$m_\mu(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu \sigma_i \tag{1.31}$$

so that the Hamiltonian can be rewritten as a function of the order parameters as

$$H(\boldsymbol{\sigma}; \boldsymbol{h}) = -\frac{1}{2} N \sum_{\mu=1}^{p} m_\mu^2(\boldsymbol{\sigma}) + \frac{1}{2} N p \tag{1.32}$$

and the free energy becomes

$$f(\beta) \;=\; -\frac{1}{N\beta}\log\int_{-\infty}^{+\infty}\prod_{\mu=1}^{p} dm_\mu \sum_{\{\sigma\}} \delta\big(Nm_\mu - \sum_{i=1}^{N}\xi_i^\mu\sigma_i\big)\, e^{\frac{1}{2}\beta N\boldsymbol{m}^2 + p/2}\,. \tag{1.33}$$

Following the same reasoning as in Section 1.1.1, by the use of Fourier representation of the δ function and of the saddle point method, we can state that, in the thermodynamic limit, the physical values taken by the order parameters are the solutions of the p equations:

$$\boldsymbol{m} = \langle \boldsymbol{\xi}\tanh(\beta\,\boldsymbol{m}\cdot\boldsymbol{\xi})\rangle_{\boldsymbol{\xi}} \tag{1.34}$$

that locally extremize the free energy

$$f(\boldsymbol{m}) = \frac{1}{2}\boldsymbol{m}^2 - \frac{1}{\beta}\langle\log\cosh(\beta\,\boldsymbol{m}\cdot\boldsymbol{\xi})\rangle + \frac{\alpha}{2} - \frac{1}{\beta}\log 2\,. \tag{1.35}$$

A solution of the self consistency equation (1.34) is a pure state if it minimizes (and not maximizes) the free energy (1.35). So, in order to discriminate between stable and non-stable solutions, the spectrum of the Hessian matrix

$$\frac{\partial^2 f}{\partial m_\mu \partial m_\nu} = \delta_{\mu\nu} - \beta\langle\xi_\mu\xi_\nu\left[1-\tanh^2(\beta\boldsymbol{\xi}\cdot\boldsymbol{m})\right]\rangle \tag{1.36}$$

has to be studied.

A crucial question regards the stability of pure states and mixture states. In order to answer it an *ansatz* has to be done on the form of the vector $\boldsymbol{m}$. The simplest one focuses on the situation where n patterns are retrieved with the same intensity so that, when $n = 1$, the considered state is pure, while it is a mixed state for $n > 1$. For simplicity let us suppose that the first n patterns are recalled; of course the symmetry of the problem assures that the conclusions hold for cases in which other patterns are retrieved. So symmetric *ansatz* is

$$\boldsymbol{m} = m_n(\overbrace{1,\ldots 1}^{n\text{times}},\overbrace{0,\ldots,0}^{p-n\text{times}}) \tag{1.37}$$

Plugging it into equations (1.34), (1.35) and (1.36) and studying them would lead to the situation resumed in Figure 1.2. As for the mean field Ising model, over the critical temperature $T_c = 1$, for every n, only the solution $m_n = 0$ is stable, so every pattern overlap vanish and the system is in a paramagnetic state. For $T < T_c$, the paramagnetic solution is no longer stable and the pure state where only one pattern is retrieved ($n = 1$) is always stable. The solutions with even n are always unstable. Lowering the temperature, mixture states with odd $n > 1$ start being stable; the higher n the lower is the temperature at which the corresponding solution becomes stable. As the free energy of the pure state is always lower than the mixture states ones (Figure 1.2, right panel), in the thermodynamic limit only the pure states give contribution to the free energy. Nevertheless the mixture states are attractors of a large but finite system; so, in order for a pure state to be retrieved by a dynamical process, the initial condition should be close enough to it.

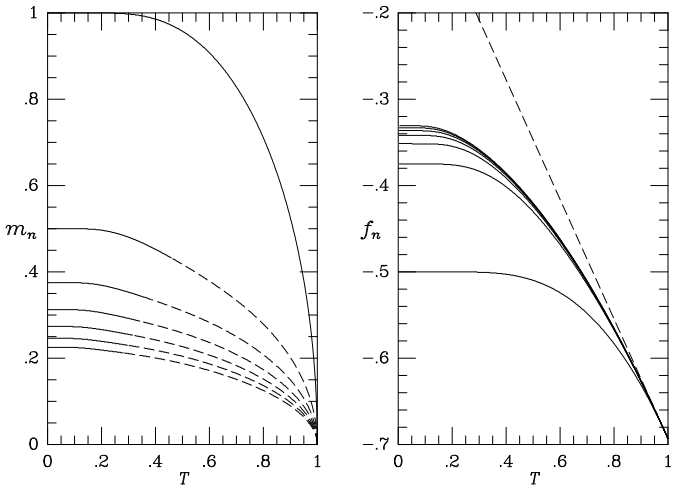


Figure 1.2. Amplitudes m_n of the mixture states of the low storage Hopfield model as a function of the temperature. From top to bottom: $n = 1, 3, 5, 7, 9, 11, 13$. Region where they are stable are drawn with a solid line, while unstable solutions are drown as a dashed line Dashed.
Right picture: corresponding free energies $f(m_n)$. From bottom to top $n = 1, 3, 5, 7, 9, 11, 13$. The free energy of the paramagnetic state is drawn for comparison with a dashed line. Figure from [13].

1.2 The replica trick for disordered spin systems: The Sherrington-Kirkpatrik model

In last Section the Hopfield model was studied in the regime of finite number p of stored patterns. If we consider the condition where the number of patterns are infinite and scale linearly with the system size the procedure outlined above is not useful anymore since the dimension of the integral in (1.33) diverges as fast as the exponent of the integrand function so that the saddle point method can no longer be applied. Nevertheless the system can still be studied but it turns out to be similar to a *glassy* one and more sophisticated computational tools are needed. Among them one of the most (and first) used is the *replica method* [14]. In this Section, with the aim of introducing it, we outline the (replica symmetric) solution of the most famous spin glass model: the Sherrington-Kirkpatrick model.

1.2.1 The quenched disorder in mean field systems and the replica trick

The class of spin models that will be discussed from now on are characterized by the fact that the couplings between spin variables are heterogenous and in particular they can be both positive and negative. The intuitive picture of the consequence of the presence of such disorder is that these systems (in particular regimes) experience *frustration*. This means that the interaction network displays closed loops where the couplings are such that, if we start choosing the spins values in order to minimize the

energetic contribution of the loop, both choices of the last spin value are equivalent and one pair of variables in the loop involving that (*frustrated*) spin will give a positive contribution to the energy. Frustration in the network gives rise to a complex energy (and free energy) landscape with many metastable (locally minimizing the energy) configurations.

Given that thermodynamics is studied for large (infinite in the limit) systems, it is sensible to express the disorder in terms of a probability distribution of parameters. For spin systems this for example would mean that the couplings are not homogeneous as in the Ising model of Section 1.1.1 but are thrown from a distribution $P(J)$.

The hypothesis of *self averaging* of the thermodynamic quantities states that the right procedure to compute the physical value of them, in the thermodynamic limit, is that of performing the sum over configurations at fixed realizations of the disorder and then to average these thermodynamic quantities over the disorder distribution. In particular, indicating with $\mathbb{E}$ the operation of average over the disorder ($\mathbb{E}\, g(J) \equiv \int dJ g(J) P(J)$) , the physical value of the free energy is obtained as

$$\bar{f} \equiv \mathbb{E}\, f = \lim_{N\to\infty} \frac{1}{N} \mathbb{E} \log Z(J) \,, \tag{1.38}$$

where $Z(J)$ is the partition function at finite N computed at fixed disorder. It has to be remarked that the average of the logarithm of the partition is conceptually very different from (and computationally much harder than) the simple average of the partition function. Averaging Z over that disorder would mean to average the disorder together with the dynamic variables, while the couplings are meant to be fixed while the spins evolve.

To perform the average over the disorder of $\log Z$ we introduce the so called *replica trick* which makes use of the limit

$$\mathbb{E} \log Z = \lim_{n\to 0} \frac{\mathbb{E} Z^n - 1}{n} \tag{1.39}$$

or equivalently

$$\mathbb{E} \log Z = \lim_{n\to 0} \frac{1}{n} \log[\mathbb{E}\, Z^n] \,. \tag{1.40}$$

This formal equivalence is not of practical use unless the parameter n is considered as an integer; this corresponds to compute the partition function for n independent copies of the system with the same realization of the disorder and to average their product. The limit is then interpreted as an analytical continuation to real values of n. This is of course mathematically not safe; moreover, as we will see in practical cases, Z^n is usually computed performing the thermodynamic ($N \to \infty$) limit before the $n \to 0$ one. This exchange of the ordering of the limits is usually not justified. Nevertheless in some cases (as for instance for the Sherrington-Kirkpatrick model; see [15]) for the procedure is proven to give the exact results by rigorous arguments

1.2.2 The replica symmetric solution of the Sherrington-Kirkpatrick model

The Sherrington-Kirkpatrick (SK) model is the paradigm for disordered spin systems for the compactness of its definition and the richness of its phenomenology.

It is a mean field spin model defined through the Hamiltonian

$$H_{SK} = -\frac{1}{\sqrt{N}} \sum_{i<j}^{N} J_{ij}\sigma_i\sigma_j \,, \tag{1.41}$$

where the couplings J_{ij} are independently thrown from a standard normal distribution:

$$dP(J_{ij}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{J_{ij}^2}{2}} dJ_{ij} \equiv d\mu(J_{ij}) \quad \text{and} \quad P(J_{ij}, J_{kl}) = P(J_{ij})P(J_{kl}), \quad \forall (i,j) \neq (k,l), \tag{1.42}$$

where, from now on, $d\mu(x) \equiv (2\pi)^{1/2} \exp(-x^2/2)\, dx$ indicates the standard gaussian measure. Here we will outline the solution that was first proposed for this model by David Sherrington and Scott, Kirkpatrick in [14]. This solutions which is known as the *replica symmetric* (RS) solution, is wrong but represents a good approximation for the true solution at temperature slightly below the transition temperature. The right solution of has been proposed by Giorgio Parisi in [16] and the rigorous proof of its correctness takes contributions from the works of Francesco Guerra [17] and Michel Talagrandt [18].

As the contribution of the diagonal terms in (1.41) is negligible in the thermodynamic limit, the definition of the model will be approximated by

$$H_{SK} = -\frac{1}{2\sqrt{N}} \sum_{ij}^{N} J_{ij}\sigma_i\sigma_j \,. \tag{1.43}$$

Using the replica trick in the form of Equation (1.40),

$$\bar{f}(\beta) = \lim_{N\to\infty} -\frac{1}{\beta N} \mathbb{E} \log Z = \lim_{N\to\infty} -\frac{1}{\beta N} \lim_{n\to 0} \log \mathbb{E} Z^n \,. \tag{1.44}$$

Performing the disorder average of the replicated systems involves just gaussian integrations so that

$$\mathbb{E} Z^n = \sum_{\{\sigma\}} \int_{-\infty}^{\infty} \prod_{ij} dJ_{ij} e^{-\frac{J_{ij}^2}{2} + \sum_a^n \frac{\beta}{2\sqrt{N}} J_{ij}\sigma_i^a\sigma_j^a} = \sum_{\sigma} e^{\sum_{ij}^N \frac{\beta^2}{4N} \sum_{ab}^n \sigma_i^a\sigma_i^b\sigma_j^a\sigma_j^b} \,. \tag{1.45}$$

We now define what is, at this level, the order parameter of the system: the overlap matrix

$$q_{ab}(\boldsymbol{\sigma}) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^a \sigma_i^b \tag{1.46}$$

whose entries represent the identity fraction of two replicas of the system. $q_{ab} \in [-1, 1]$: $q_{ab} = 1(-1)$ when the two replicas are in the same (in completely anti-correlated) configurations and $q_{ab} = 0$ when the two replicas are in completely different configurations. Now, calling the integration measures $Dq \equiv \prod_{ab}^n dq_{ab}$ and

$D\Lambda \equiv \prod_{ab}^{n} d\Lambda_{ab}$

$$\begin{aligned}
\mathbb{E}Z^n &= \sum_{\{\sigma\}} e^{\frac{N\beta^2}{4}\sum_{ab}^{n} q_{ab}^2(\boldsymbol{\sigma})} = \sum_{\{\sigma\}} \int_{-\infty}^{\infty} \prod_{ab}^{n} dq_{ab}\delta(Nq_{ab} - \sum_i \sigma_i^a \sigma_i^b) e^{\frac{N\beta^2}{4}\sum_{ab}^{n} q_{ab}^2} = \\
&= \int_{-\infty}^{\infty} Dq \int_{-i\infty}^{i\infty} D\Lambda \sum_{\{\sigma\}} e^{N\sum_{ab}\Lambda_{ab}q_{ab} - \sum_{ab}\sum_i \sigma_i^a \Lambda_{ab}\sigma_i^b + N\frac{\beta^2}{4}q_{ab}^2} = \\
&= \int_{-\infty}^{\infty} Dq \int_{-i\infty}^{i\infty} D\Lambda\, e^{-N\beta f(q,\Lambda)}\,, \qquad (1.47)
\end{aligned}$$

with

$$f(q,\Lambda) = -\frac{1}{\beta}\sum_{ab}\Lambda_{ab}q_{ab} - \frac{1}{\beta}\log\left(\sum_{\{\sigma\}} e^{\sum_{ab}\sigma^a\Lambda_{ab}\sigma^b}\right) - \frac{\beta}{4}\sum_{ab} q_{ab}^2\,, \qquad (1.48)$$

that is symmetric under permutations of replica indexes. The physical value of the intensive free energy is given for formula (1.44). Exchanging the orders of the limits in that fomrula is the only way to compute the integrals over the measures Dq and $D\Lambda$. In that case, taking the thermodynamic limit before sending the number of replicas to zeros allows to use the saddle point method. So the extremization of $f(q,\Lambda)$ over q allow to eliminate the additional variables Λ with

$$\Lambda_{ab} = -\frac{\beta^2}{2}q_{ab}\,, \qquad (1.49)$$

so that the expression of the free energy as a function of the overlap matrix is

$$f(q;\beta) = \frac{\beta}{4}\sum_{ab} q_{ab}^2 - \frac{1}{\beta}\log\left(\sum_{\{\sigma\}} e^{\frac{\beta^2}{2}\sum_{ab}\sigma^a q_{ab}\sigma^b}\right)\,. \qquad (1.50)$$

In last formula it is evident the effect of the replica trick: it has decoupled sites but now replicas are coupled by the overlap matrix. Last expression has to be computed in the saddle point and the limit $n \to 0$ is still to be performed. In order to do this a parametrization of the overlap matrix needs to be made explicit. A good parametrization could permit to decouple replicas.

The simplest *ansatz* for the form of the matrix overlap is the one that respects the symmetry of the $f(q)$, so a matrix whitch has 1 on the diagonal and a parameter q_0 on all the entries out of the diagonal, i. e.

$$q_{ab} = q_0 + (1-q_0)\delta_{ab}\,. \qquad (1.51)$$

This is known as the *replica symmetric (RS) anstatz*. With this parametrization the free energy becomes

$$\begin{aligned}
f(q_0;\beta) &= \lim_{n\to 0}\frac{1}{n}f(q_{RS}) = \lim_{n\to 0}\frac{\beta}{4}nq_0^2 + \frac{\beta}{4}(1-q_0^2) + \\
&\quad -\frac{1}{\beta n}\log\left(\sum_{\sigma} e^{\frac{1}{2}\beta^2 q_0\sum_{ab}^{n}\sigma^a\sigma^b} e^{\frac{\beta^2}{2}(1-q_0)n}\right) = \\
&= -\frac{\beta}{4}(1-q_0)^2 - \lim_{n\to 0}\frac{1}{\beta n}\log\left(\sum_{\sigma} e^{\frac{1}{2}\left(\beta\sqrt{q_0}\sum_a^n \sigma^a\right)^2}\right)\,. \qquad (1.52)
\end{aligned}$$

Rewriting the last term as the result of a gaussian integration

$$\begin{aligned} f(q_0;\beta) &= -\frac{\beta}{4}(1-q_0)^2 - \lim_{n\to 0}\frac{1}{\beta n}\log\sum_{\sigma}\int d\mu(z) e^{z\beta\sqrt{q_0}\sum_a \sigma_a} = \\ &= -\frac{\beta}{4}(1-q_0)^2 - \lim_{n\to 0}\frac{1}{\beta n}\log\int d\mu(z)\prod_a^n\sum_{\sigma_a} e^{z\beta\sqrt{q_0}\sigma_a} = \\ &= -\frac{\beta}{4}(1-q_0)^2 - \lim_{n\to 0}\frac{1}{\beta n}\log\int d\mu(z)\left(2\cosh(z\beta\sqrt{q_0})\right)^n . \end{aligned} \tag{1.53}$$

We can now use the replica trick relation (1.40) in the inverse sense so that we have the final result for the RS free energy of the SK model

$$f(q_0;\beta) = -\frac{\beta}{4}(1-q_0)^2 - \frac{1}{\beta}\int d\mu(z)\log\left(\cosh(z\beta\sqrt{q_0})\right) - \frac{1}{\beta}\log 2 . \tag{1.54}$$

This expression has to be minimized with respect to the parameter q and this procedure gives the self consistency relation

$$q_0 = \int d\mu(z)\tanh^2(z\beta\sqrt{q_0}) . \tag{1.55}$$

By studying equations (1.54) and (1.55) it is possible to verify that for $T > T_C \equiv 1$ only the solution $q = 0$ is stable while for $T < 1$ stable solutions with $q > 0$ appear. The nature of this phase transition appears obscure at the moment so that we now aim to deep the physical meaning of the order parameter.

1.2.3 The physical interpretation of the overlap and the spin glass phase

As we have shown in Section 1.1.1 for the case of the mean field Ising model, in interacting systems ergodicity can break in the low temperature phase in which the phase space can split into ergodic regions where the system remains trapped. Let us indicate states with greek letters. Pure sates, in mean field systems (not at the critical point), at equilibrium, have the *clustering properties* which states that spins are independent inside a generic state α:

$$\langle\sigma_i\sigma_j\rangle_\alpha = \langle\sigma_i\rangle_\alpha\langle\sigma_j\rangle_\alpha . \tag{1.56}$$

In finite dimension systems, the property still holds for spins that are far apart in the space structure, i.e. for $|i-j|\to\infty$. Calling $Z_\alpha = \sum_{\{\sigma\in\alpha\}} e^{-\beta H(\sigma)}$ and $w_\alpha = Z_\alpha/Z$, such that $\sum_\alpha w_\alpha = 1$ and $\sum_\alpha Z_\alpha = Z$, the thermal average can be split into ergodic components as

$$\langle\Phi(\boldsymbol{\sigma})\rangle = \sum_{\{\sigma\}}\Phi(\boldsymbol{\sigma})\, e^{-\beta H(\boldsymbol{\sigma})} = \sum_\alpha w_\alpha\langle\Phi(\boldsymbol{\sigma})\rangle_\alpha . \tag{1.57}$$

As we have defined the overlap between configurations in (1.46), it is possible to investigate the overlap between two states, defined as

$$q_{\alpha\beta} = \frac{1}{N}\sum_i\langle\sigma_i\rangle_\alpha\langle\sigma_i\rangle_\beta . \tag{1.58}$$

The structure of the states overlap gives informations about the structure of the states of the system. For example the *self-overlap* of a state α, $q_{\alpha\alpha}$ encodes the information about the size of the state: $q_{\alpha\alpha} = 1$ when the state correspond to only one configuration and $q_{\alpha\alpha} = 0$ when the state is the largest possible one, the paramagnetic state.

Let us consider the probability distribution of the overlap between two replicas of the system at fixed disorder

$$\begin{aligned} P(q) &= \frac{1}{Z} \sum_{\{\sigma_a \sigma_b\}} e^{-\beta H(\sigma^a) - \beta H(\sigma^b)} \delta(q - q_{ab}) = \\ &= \sum_{\alpha\beta} w_\alpha w_\beta \, \delta(q - q_{\alpha\beta}) \end{aligned} \tag{1.59}$$

From this relation is possible to understand the role played replicas in spin glasses: they serves as probes to explore the structure of states of the system which, as we will see, can be very complex. So last quantity is of uttermost importance for spin glasses and in fact the true order parameter of this kind of systems turns out to be the full average overlap distribution $\mathbb{E}P(q)$. A hint that could justify this claim comes from the definitions of the moments of the average overlap distribution that, by the use of the clustering property, can be easily shown to be

$$q^{(k)} = \int dq \, q^k \, \mathbb{E}P(q) = \frac{1}{N^k} \sum_{i_1 \dots i_k} \mathbb{E}\langle \sigma_{i_1} \dots \sigma_{i_k} \rangle^2 \, . \tag{1.60}$$

In particular the first moment is

$$q^{(1)} = \int dq \, q \, \mathbb{E}P(q) = \frac{1}{N} \sum_i \mathbb{E}\langle \sigma_i \rangle^2 \equiv q_{EA} \, . \tag{1.61}$$

This quantity, also known as the *Edward-Anderson order parameter*, gives a very intuitive information about the presence of an "ordered" phase in spin glasses. In fact as the couplings are randomly distributed around zero, the naïf idea about the broken phase is that spins are frozen around random directions evenly distributed among sites. So in spin glasses the global magnetization $m = 1/N \sum_i \langle \sigma_i \rangle$ vanishes also if spins are frozen in random directions and the local magnetizations $m_i = \sum_i \langle \sigma_i \rangle \neq 0$; while, thanks to the presence of the power two in its definition, q_{EA} is a global quantity that can distinguish between the paramagnetic phase, where $\langle \sigma_i \rangle = 0$, $\forall i$, and the spin glass phase $\langle \sigma_i \rangle \neq 0$ but with different signs depending on the site.

We stress anyway that, even if q_{EA} can give a very effective information on the presence of symmetry breaking, the true order parameter of mean field spin glasses is the full distribution $\mathbb{E}P(q)$. In order to compute it, the replica trick is needed. In fact it can be used in general to compute thermal averages of a generic quantity

$\Phi(\boldsymbol{\sigma})$ as

$$\begin{aligned}\langle\Phi(\boldsymbol{\sigma})\rangle &= \frac{\sum_{\{\sigma\}} e^{\beta H(\boldsymbol{\sigma})}\Phi(\boldsymbol{\sigma})}{\sum_{\{\sigma\}} e^{\beta H(\boldsymbol{\sigma})}} = \lim_{n\to 0}\sum_{\{\sigma\}} e^{\beta H(\boldsymbol{\sigma})}\Phi(\boldsymbol{\sigma})\left[e^{\beta H(\boldsymbol{\sigma})}\right]^{n-1} = \\ &= \lim_{n\to 0}\sum_{\{\sigma\}}\Phi(\boldsymbol{\sigma}^1)\, e^{\beta\sum_a^n H(\boldsymbol{\sigma}^a)} = \\ &= \lim_{n\to 0}\sum_{\{\sigma\}}\frac{1}{n}\sum_a^n\Phi(\boldsymbol{\sigma}^a)\, e^{\beta\sum_b^n H(\boldsymbol{\sigma}^b)}\,. \end{aligned} \tag{1.62}$$

So the average overlap distribution can be computed as

$$\mathbb{E}P(q) = \lim_{n\to 0}\mathbb{E}\sum_{\{\sigma\}}\frac{2}{n(n-1)}\sum_{a>b}\delta(q-q_{ab})\, e^{-\beta\sum_c^n H(\boldsymbol{\sigma}^c)}\,. \tag{1.63}$$

Using the exchange of limits and the saddle point methods exactly as for the computation of the free energy, the result is

$$\mathbb{E}P(q) = \lim_{n\to 0}\frac{2}{n(n-1)}\sum_{a>b}\delta(q-q_{ab}^{SP})\,, \tag{1.64}$$

where q_{ab}^{SP} stands for the overlap matrix computed at the saddle point.

The insertion of the RS *ansatz* give as a result

$$\mathbb{E}P(q)_{RS} = \delta(q-q_0)\,, \tag{1.65}$$

where $q_0 = q_{EA}$ and satisfies (1.55).

For the SK models the replica symmetric solutions is not exact and the the scientific community was suddenly aware of this as, in their paper [14], Sherrington and Kirkpatrick pointed out that, at low temperatures, their solution gives a negative entropy (that is not possible for discrete variables where the entropy is the logarithm of the number of available configurations at temperature T, that is always greater than one and so its log greater than zero).

As already said, the Parisi solution [16] gives the correct structure of the low temperature phase which is much more complex of the RS scenario and it predicts a continuum structure for the $\mathbb{E}P(q)$.

Details about the realization of the symmetry breaking in the SK model are beyond the scope of this Thesis. We stress anyway that the RS solution of disordered systems is a useful first step in their knowledge as it provides informations about the presence of a spine glass phase transition and locate it in the parameter space. For example, by letting the average J_0 and standard deviation J of the coupling distribution be two tunable parameters of the model the RS approximation allows separate a paramagnetic phase, a ferromagnetic phase and a spin glass phase and so to draw the phase diagram in Figure 1.3.

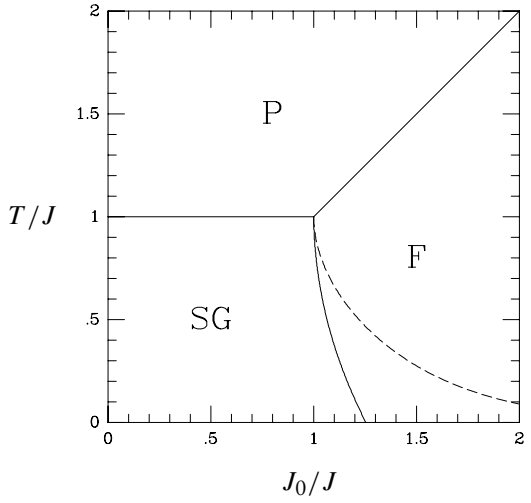


Figure 1.3. Phase diagram of the SK model. Lines separated the paramagnetic (P), the ferromagnetic (F) and the (SG) phases. The dashed line is the so called *Almeida-Thouless (AT) line* that separate the region of the parameter space where the RS *ansatz* gives the exact solution from that in which it represents an approximation. Figure from [13].

1.3 The RS solution of the Sherrington-Kirckpatrick model thorough stochastic stability method

As shown in last Section, the RS solution of the SK model can be obtained through the use of the replica trick. Despite the latter has a historical and conceptual value (as it represented the first technique that was used to successfully tackle the mean field spin glass problem) as already pointed out it is not completely free from mathematically risky passages. For this reason the mathematical physics community has developed techniques that aim to rigorously prove results achieved by the use of the replica trick.

Here we present the *stochastic stability technique*, that was introduced in [19], and we use it to derive the RS solution of the SK model, that in this way, comes rigorously from the RS assumption.

In order to compute the free energy of the SK model, the stochastic stability method suggests to introduce an interpolating Hamiltonian as:

$$H_t(\boldsymbol{\sigma}; \boldsymbol{J}, \boldsymbol{\eta}, a) = -\frac{1}{\sqrt{N}} \sum_{i<j}^{N} J_{ij}\sigma_i\sigma_j - a\sum_i \eta_i\sigma_i \,, \tag{1.66}$$

where, as usual for the SK model J are independently normally distributed and the same holds for the random fields η. The constant a has been introduced as the scale of the random field and a clever choice of its value will be provided in the following.

The interpolating free energy reads off as

$$\begin{aligned} f(\beta,t) &= -\frac{\mathbb{E}}{N\beta}\sum_{\{\sigma\}} e^{-\beta H_t} = \\ &= -\frac{\mathbb{E}}{N\beta}\log\sum_{\{\sigma\}} e^{\sqrt{t}\frac{\beta}{\sqrt{N}}\sum_{i<j}^{N} J_{ij}\sigma_i\sigma_j+\sqrt{1-t}\,a\sum_i^N \eta_i\sigma_i}\,. \end{aligned} \tag{1.67}$$

As can be easily verified, the complete solution of the SK model corresponds to $f(\beta, t=1)$, while $f(\beta, t=0)$ can be easily solved as it is a problem of independent spins in a random field.

Once the interpolating free energy has been introduced, the strategy for the solution of the problem will be to compute its $t=0$ value, its derivative $df(\beta,t)/dt$ and to compute the free energy as

$$f(\beta) = f(\beta, t=1) = f(\beta, t=0) + \int_0^1 \frac{df(\beta,t)}{dt}\,dt\,. \tag{1.68}$$

The computation of the $t=0$ term gives

$$\begin{aligned} f(\beta,t=0) &= -\frac{\mathbb{E}}{\beta N}\log\sum_{\{\sigma\}} e^{a\,\sum_i \eta_i\sigma_i} \\ &= -\frac{\mathbb{E}}{\beta N}\log\left(2\cosh a\eta\right)^N \\ &= -\frac{1}{\beta}\int d\mu(\eta)\log\cosh(a\eta) - \frac{1}{\beta}\log 2\,. \end{aligned} \tag{1.69}$$

Introducing the interpolating partition function $Z_t \equiv \sum_{\{\sigma\}} e^{-\beta H_t}$, the interpolating thermal average operator $\omega_t(O) \equiv \sum_{\{\sigma\}} O(\boldsymbol{\sigma})\, e^{-\beta H_t}$ and the interpolating complete average as $\langle O\rangle_t \equiv \mathbb{E}\,\omega_t(O)$, the streaming derivative of the interpolating free energy is:

$$\begin{aligned} \frac{df}{dt} &= -\frac{\mathbb{E}}{N\beta}\frac{1}{Z_t}\sum_{\{\sigma\}}\left(\frac{\beta}{2\sqrt{Nt}}\sum_{i<j} J_{ij}\sigma_i\sigma_j - \frac{\beta a}{2\sqrt{1-t}}\sum_i \eta_i\sigma_i\right) e^{(\dots)} = \\ &= -\frac{\mathbb{E}}{N\beta}\left(\frac{\beta}{2\sqrt{Nt}}\sum_{i<j} J_{ij}\omega_t(\sigma_i\sigma_j) - \frac{\beta a}{2\sqrt{1-t}}\sum_i \eta_i\omega_t(\sigma_i)\right) = \\ &= -\frac{\mathbb{E}}{N\beta}\left(\frac{\beta}{2\sqrt{Nt}}\sum_{i<j} \partial_{J_{ij}}\omega_t(\sigma_i\sigma_j) - \frac{\beta a}{2\sqrt{1-t}}\sum_i \partial_{\eta_i}\omega_t(\sigma_i)\right) = \\ &= -\frac{\mathbb{E}}{N\beta}\left[\frac{\beta^2}{2N}\sum_{i<j}\left(1-\omega_t^2(\sigma_i\sigma_j)\right) - \frac{\beta^2 a^2}{2}\sum_i\left(1-\omega_t^2(\sigma_i)\right)\right] = \\ &= -\frac{\mathbb{E}}{N\beta}\left[\frac{\beta^2}{4N}\sum_{ij}\left(1-\omega_t^2(\sigma_i\sigma_j)\right) - \frac{\beta^2 a^2}{2}\sum_i\left(1-\omega_t^2(\sigma_i)\right)\right] = \\ &= -\frac{\beta}{4}\left(1-\langle q_{12}^2\rangle\right) - \frac{\beta a^2}{2}\left(1-\langle q_{12}\rangle\right)\,, \end{aligned} \tag{1.70}$$

where in the third passage we have use integration by parts that, for gaussian variables J and regular functions $g(J)$, gives

$$\int d\mu(J)\, Jg(J) = \int d\mu(J)\, \partial_J g(J)\,. \tag{1.71}$$

Introducing the *source of fluctuations* as

$$S(\beta, t) = \frac{\beta}{4}\langle (q_{12} - q_0)^2 \rangle_t \tag{1.72}$$

and choosing

$$a = \sqrt{q_0}\,, \tag{1.73}$$

the streaming derivative becomes

$$\frac{df}{dt} = S(\beta, t) - \frac{\beta}{4}(1 - q_0)^2\ . \tag{1.74}$$

In the above formula, the only dependence on t is in the source of fluctuations $S(\beta, t)$. As already pointed out, the RS approximation corresponds to neglecting fluctuations of the overlap so to fix $S(\beta, t) = 0$, so that

$$\left.\frac{df}{dt}\right|_{RS} = -\frac{\beta}{4}(1 - q_0)^2\ . \tag{1.75}$$

Inserting last result and (1.69), with the choice (1.73), into Equation (1.68) RS free energy of Equation (1.54).

It has to be underlined that the above calculation is rigorous and shorter than the replica trick one.

1.4 A paradigm for neural networks: The high storage Hopfield model

At this point we have all the computational tools to take into account the so called *High storage Hopfield model* in which the number p of patterns is infinite and it scales linearly with the system size so that the *load* can be defined by the finite quantity

$$\epsilon = \lim_{N\to\infty} \frac{p}{N}\,. \tag{1.76}$$

The existence of an infinite number of patterns can be responsible for the destruction of the stable retrieval states [20]. This can be intuitively understood by simple dimensional arguments. In fact the energy of a general Ising system can be written as a function of the fields acting on every spins defined by $w_i(\boldsymbol{\sigma}) = \sum_J^N J_{ij}\sigma_j$, so that $H = -\sum_i^N w_i(\boldsymbol{\sigma})\sigma_i$. Let us separate the contribute of the first pattern

$$w_i(\boldsymbol{\sigma}) = \frac{1}{N}\sum_j^N \xi_i^1\xi_j^1\sigma_j + \frac{1}{N}\sum_j^N\sum_{\mu>1}^p \xi_i^\mu\xi_j^\mu\sigma_j\,. \tag{1.77}$$

Let us consider a configuration close to the first pattern, so to say, $\boldsymbol{\sigma} \simeq \boldsymbol{\xi}^1$. In this case, as the ξ are random binary variables equally distributed between $+1$ and -1,

the same holds for the product $\xi_j^\mu \sigma_j$, $\forall j$, $\forall \mu > 1$. So, by the fact that a sum of M such random variables has null average and variance $\sim M$, $\sum_{\mu>1}^{p} \xi_i^\mu = \mathcal{O}(\sqrt{p})$ and $\sum_j^N \xi_j^\mu \sigma_j = \mathcal{O}(\sqrt{N})$

$$w_i(\boldsymbol{\sigma}) \simeq \xi_i^1 + \eta_i \,. \tag{1.78}$$

where the η_i are a random fields with null average and magnitude of order $\mathcal{O}(\sqrt{p/N})$. So, when p is a finite number, the contribution of the random field vanishes and the retrieved state is stable. Nevertheless, in the high storage limit, the random fields have the same magnitude of the contribution due to the retrieved pattern and they can totally destabilize the retrieval state.

In fact when the number of patterns is infinite the Central Limit Theorem of probability shows that, being the couplings J defined in (1.27) an infinite sum of random variables, if the number of patterns is too high, they can become approximately normally distributed, so presenting a situation similar to that of a *spin glass*. This is a hint about the fact that the solution of the high storage Hopfield model can be obtained by the use of the replica method.

1.4.1 Replica symmetric solution of the high storage Hopfield model

The replica method can be applied in order to compute the free energy of the high storage Hopfield model. Unfortunately, the quenched random couplings are given in terms of the ξs that have discrete probability distribution. This make the computation much harder than the one for the SK model where gaussian quenched averaged allow more tractable computations. As a result the full RSB solution is not known, while only the RS solution and its 1RSB correction [21] have been computed.

Here we report the result of the RS computation. The detailed replica calculation can be found in [20] or in [13]; both versions are very didactical. In the case in which a finite number s of patterns is recalled, the free energy reads off as

$$\begin{aligned} f(\boldsymbol{m}, q, r; \beta, \epsilon) \;=\; & -\frac{1}{\beta}\ln 2 + \frac{1}{2}\epsilon + \frac{1}{2}\sum_{\nu}^{s}(m^\nu)^2 + \\ & + \frac{\epsilon}{2\beta}\left(\ln\left[1-\beta(1-q)\right] - \frac{\beta q}{1-\beta(1-q)}\right) + \frac{\epsilon\beta r}{2}(1-q) + \\ & - \frac{1}{\beta}\int d\mu(z)\, \mathbb{E}_\xi \ln\cosh\left[\beta(\sqrt{\epsilon r}z + \sum_{\nu}^{s} m^\nu \xi^\nu)\right] . \end{aligned} \tag{1.79}$$

Minimization of last formula with respect to the parameters m, q and r gives the self consistency equations

$$m \;=\; \mathbb{E}_\xi\, \xi \int d\mu(z) \tanh\beta[\sqrt{\epsilon r}z + \sum_{\rho}^{2} m^\rho \xi^\rho] \,, \tag{1.80}$$

$$q \;=\; \int d\mu(z) \tanh^2\beta[\sqrt{\epsilon r}z + \sum_{\rho}^{2} m^\rho \xi^\rho] \,, \tag{1.81}$$

$$r \;=\; \frac{q}{[1-\beta(1-q)]^2} \,. \tag{1.82}$$

The saddle point values of the order parameters are

$$m = \mathbb{E}\,\frac{1}{N}\sum_{i=1}^{N}\xi_i\langle\sigma_i\rangle\,, \tag{1.83}$$

$$q = \mathbb{E}\,\frac{1}{N}\sum_{i=1}^{N}\langle\sigma_i\rangle^2\,, \tag{1.84}$$

$$r = \mathbb{E}\,\frac{N}{\epsilon}\sum_{\mu=s+1}^{p}\left(\frac{1}{N}\sum_{i=1}^{N}\xi_i^1\langle\sigma_i\rangle\right)^2\,. \tag{1.85}$$

As usual the values that solve these equations have to be plugged into the free energy (1.79) to obtain its physical value.

The parameter r represents the effect of the non recalled patters; their single amplitude is of order $\mathcal{O}(1/\sqrt{N})$ but their amount is $\sim p$ so that their cumulative effect is finite also in the thermodynamic limit. As can be seen by direct inspections of the self consistency relations, the effect of the non recalled patterns is that of an additional source of (slow) noise.

Relation (1.82) permits to easily eliminate the variable r so that the relevant order parameters are $\boldsymbol{m}$ and q.

In particular in the case in which only one (let us say the first) pattern is recalled (i.e. with the *pure state anstaz*: s=1, $\boldsymbol{m} = (m, 0, \ldots, 0)$), eliminating r the self consistency relations read as

$$m = \int d\mu(z)\tanh\beta\left[\frac{\sqrt{\epsilon q}}{1-\beta(1-q)}z + m\right]\,, \tag{1.86}$$

$$q = \int d\mu(z)\tanh^2\beta\left[\frac{\sqrt{\epsilon q}}{1-\beta(1-q)}z + m\right] \tag{1.87}$$

and they minimize

$$f(m,q;\beta,\epsilon) = -\frac{1}{\beta}\ln 2 + \frac{1}{2}\epsilon + \frac{1}{2}m^2 + \frac{\epsilon}{2\beta}\left(\ln\left[1-\beta(1-q)\right] - \frac{\beta q}{1-\beta(1-q)}\right) + \\ + \frac{\epsilon\beta q(1-q)}{2[1-\beta(1-q)]^2} - \frac{1}{\beta}\int d\mu(z)\ln\cosh\left[\beta(\sqrt{\epsilon r}z + m)\right]\,. \tag{1.88}$$

Numerical solution of the self consistency equations and their stability study permits to draw the phase diagram in Figure 1.4. Four phases can be distinguished:

1. A *paramagnetic phase* with $q = 0$ and $m = 0$.
2. A *spin glass phase* with $q > 0$ and $m = 0$ in which the slow noise disrupt the stability of the retrieval states.
3. A *mixed phase* in which the retrieval state (with $q > 0$ and $m \neq 0$) develops having a dynamical effect, but where the thermodynamics is still dominated by the spin glass state which has the lowest free energy.
4. A *retrieval phase* in which the retrieval state become the absolute minimum of the free energy, so the thermodynamically relevant one.

It is important to note that also at zero noise level ($T = 0$), the retrieval states can be destroyed by the presence of the slow noise due to the extensive number of non recalled patterns; if the ratio between the number of stored pattern and the system size exceed a critical value, only the spin glass state exist. The RS prediction for this value is $\epsilon_c \simeq 1.38$

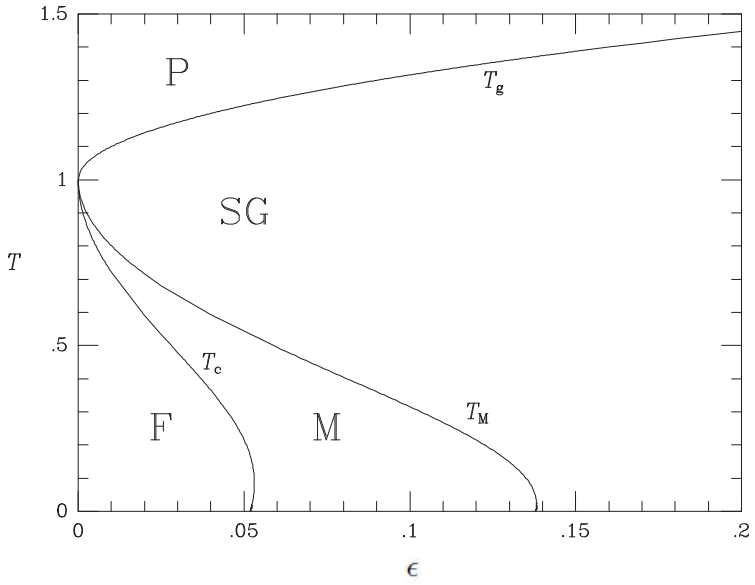


Figure 1.4. Phase diagram of the Hopfield model. P: paramagnetic state; SG: spin glass phase; M: mixed phase; F: retrieval state. Figure from [13].

Chapter 2

Basic elements of immunology

This Chapter is aimed to provide to the reader very basic elements of the biology of the adaptive immune system. It is not at all meant to be exhaustive on the argument but just to qualitatively expose the elements and mechanisms of the immune system that will be taken into account along this Thesis, specially in Chapters 3 and 6. The mechanisms of the immune system are very differentiate and they involve a huge quantity of molecule and cell types whose functional role is often of multiple nature. Despite the compelling (and still relatively poorly understood) complexity of the system here we present a selection of its constituents and mechanisms that are described in a very schematic way.

The immune system has evolved with the scope of fight pathogenic agents and cancer cells in the superior organisms.

When a pathogen enters an individual's body, the hardest task for the immune system is not simply its physical elimination, but instead its identification and separation from the endogenous molecules and cells (the so called *self*); once this task has been accomplished the elimination of the pathogen is a minor issue. In fact the *self/non-self discrimination* is one of the constitutive and more striking features of the immune system. Its specificity is accomplished by an enormous differentiation of the agents and mechanisms of the immune system.

The most coarse grained classification within the immune system is the distinction between *innate* and *adaptive* response. The former is less specific but also less efficient and consists of the first set of strategies that are set up in order to fight pathogens. A classification of the mechanisms of the innate response is beyond the scopes of this Thesis.

Some elements of the adaptive response will be taken into account along the Thesis so a brief review of its main mechanisms will be given in the following of this Chapter.

2.1 The adaptive immune response

The identifying feature of the adaptive response is its ability to adapt the cells populations, biochemical environment and cells states in response to specific

pathogens or loss of homeostasis (e.g. emergence of cancer cells). Different pathogens give rise to different organizations of the adaptive immune system in response to them.

The most important class of cells of the adaptive immune system are B and T lymphocytes. Their synergic action is responsible for the identification of specific pathogens. Once a pathogen cell or molecule is recognized by B and T (helper) lymphocytes its eventual elimination is achieved by other elements of the immune system, like the macrophage, complement molecules, etc. Moreover, specific T lymphocytes (cytotoxic T lymphocytes) can both recognize and selectively eliminate pathogenic agents.

Both B and T cells develop in *clonal* populations. They are divided into groups of cells, *clones*, each being identified by the expression of specific receptors that are able to recognize different external molecules, that are called *antigens*. The difference responsible of the diversity of B cells and T cells receptors happens at the genetic level so that different clones have genomes that diverge in the part coding for the specific receptors.

The action mechanisms of B and T lymphocytes will be sketched in the next Sections.

2.1.1 The B lymphocytes

The effector mechanisms of B cells is the production of *antibodies* (Abs), also known as *immunoglobulins* (Igs). This class of proteins, whose structure and repertoire generation will be depicted in Section 2.1.2, when released out of B cells, can fight pathogens by binding them in such a way to prevent some pathogenic mechanism (*neutralizing Abs*) or by acting as tags for other effector cells of the immune system that eventually destroy bound pathogens.

Igs are made up of a constant domain and a variable domain; the latter is responsible for the antigen binding. B cells expressing Igs with the same variable domain are said to form a clone. Igs are also expressed as membrane proteins though which B lymphocytes recognize antigens; in this case they are named *B cell receptor* (BCR).

Both B and T cells originate from the Hematopoietic Stem Cells (HSC) present in the primary lymphoid tissues, i.e. the bone marrow and fetal liver. HSC eventually gives birth to all circulating cells; among them they can differentiate in Common Lymphoid Progenitors (CLP). At this point the developmental history of B and T cell splits. CLPs can differentiate in the progenitors of T and Natural Killer cells or in the progenitors of the B cells, known as pro-B cells that do not produce Abs.

The next step of development of the B cell is known as pre-B cell. At this level of development, B cells undergo a stochastic genetic rearrangement of the DNA regions coding for the Abs variable regions (i.e. variable region of the light and heavy chains). This mechanism, known as *V(D)J genetic recombination*, will be described in more detail in Section 2.1.2; heavy chain - and subsequently light chain - genetic elements undergoes this genetic recombination (see Section 2.1.2). That first happens in one of the two homologous chromosomes; if the rearranged sequence is productive, a signal inhibits the rearrangement of the second chromosome, otherwise the second chromosome undergoes the rearrangement; if both chromosomes rearrange in a non

productive way the cell dies by apoptosis, otherwise the cell continues its evolution and Igs are produced from the rearrangement of just one of the two inherited alleles; this fact is known as *allelic exclusion*.

At this level, these cells are known as *immature* B cells; after having successfully rearranged their DNA, they produce Igs as membrane receptors, also known as *B cell receptors* (BCRs). Before leaving the primary lymphoid tissues (the bone marrow in adults) a *negative selection* is performed: B cells that produce Abs with high affinity to the *self* molecules presented in the bone marrow undergo apoptosis or *receptor editing* (a new round of genetic rearrangement). Cells that pass this self-immunity check are free to circulate as *mature* B cells and reach the secondary lymphoid tissues. Despite the presence of this gate against self immunity, some self reactive mature B cells can be found also in disease free individuals; a simple cause for this is that not every kind of self molecule is present in primary lymphoid tissues, so that the so called *central tolerance* is not an exhaustive check against self immunity. Additional mechanisms that work outside of the primary lymphoid tissues are responsible for the so called *peripheral tolerance* that makes self reactive lymphocyte become *anergic*, so to say, in a "frozen" state that block their activity.

Mature B cells that never encounter the antigens they are able to bind are called *naïve* B cells. They mainly express Abs of the class IgM and IgD as membrane receptors (see Section 2.1.2). In secondary lymphoid tissues a repertoire of naïve B clones is accumulated. When a B cell happens to encounter an antigen that it recognizes through its membrane Igs, it gets activated. Polymeric antigen (that present the same antigenic determinant repeated many times), such as polysaccharides and lipids, directly activate the bound B cell, while, when a proteic antigen is bound, a check signal by a T helper cell is needed for the B cell to get activated (see Section 2.1.3). For this reason the latter are defined as *T-dependent antigens* while the former are *T-independent*.

A B cell that receives the activation signal starts a proliferation process so that the related clone gets expanded. The majority of the proliferating cells differentiate into *plasma cells* that are able to produce Igs in the soluble form and spread them around the body to fight the pathogen. Many of the plasma cells undergo a *class switching* process by which they change the isotype of the produced Igs so that the most common secrete Igs are of the IgG class in blood and of the IgA class in the mucous tissues.

While producing Abs, B cell mainly remain confined in lymph nodes secreting Abs that circulate through the blood and the lymphatic circulation.

Some of the proliferating B cell differentiate in *memory B cells* that can live for a long period in order to maintain a population that is ready to efficiently react to further expositions to the related antigen.

Abs affinity maturation

In lymphoid follicles, that are present in secondary lymphoid tissues, Follicular Dendridic Cells (FDC) capture antigens bound in immuno complexes with Igs, expose them on their surface and recruit a very limited number of plasma cells (of the order of a few units). Starting from these cells in interaction with FDC, around the dendridic structures, a community of active B cells, called *germinal center*

(GC), is formed. These B cells are characterized by a remarkably high duplication velocity (average duplication time of 6-12 hours) and high mutation rate (1 point like mutation over 10^3 base pairs, so $10^3 - 10^4$ times more frequently than in normal duplication processes) in the region coding for the variable domain of heavy and light Ab chains (*somatic hypermutation*).

As FDC provide a chemical signal that delay apoptosis to the best binders, B cells in the GC undergo an evolutionary process competing for the best affinity to the antigen exposed by the FDC. B cells proliferation and selection seems to happen in rounds in two morphologically different regions of the GC, respectively called dark and light zone due to the different concentrations of cells. As a result of this process called *affinity maturation*, some of the selected B cells exit the GC and produce Igs with an average of 5% of mutation from the original B cell and improved affinity. The dynamics of this process is still under investigation and also mathematical modeling is involved (see for example [22] for experimental observations and [23] for mathematical modeling).

2.1.2 The immunoglobulines

Antibodies are soluble circulating proteins, also named immunoglobulins, expressed on the membrane of the B lymphocytes or secreted by plasma cells. They bind antigens in part of them called *epitopes* or *antigenic determinants*. The Abs effect can be of two kind:

1. They can bind special points on the antigen that are crucial for the effector mechanism of the pathogen and block them. In this case they are called *neutralizing Abs*.

2. They can tag antigen so that it can be eliminated by effector cells of the immune system that eliminate it.

Igs have a symmetric structure (schematically represented in Figure 2.1) with two light chains and two heavy chains that are covalently bound together by stable disulfide bonds. Both light and heavy chains structure is characterized by the presence of the so called *Ig domains* formed by two β sheets kept together by disulfide bonds. Light chain contain two Ig domains while heavy chains are equipped with four or five of them. Light chains can exist in two form named κ and λ in humans.

Both light and heavy chain have a constant and a variable domain respectively indicated by C_L, V_L, C_H and V_H. Constant domain are responsible for the effector mechanism of the Igs and the possibility for them to be exposed on the cell membrane or circulate in a soluble form. The heavy chain constant domain can assume few different structures that determine the isotype of the Ab; there exist nine different isotypes among which IgM and IgD are mainly produced by naïve B cells, IgG are mainly produced by plasma cells and found in blood, IgA are mainly found in mucoid tissues and IgE have a role in the elimination of parasites and in allergies. The roles of the different isotypes will not be deepened here.

Variable domains of both light and heavy chain consist in one Ig domain. They are responsible for the binding of the antigen. Different B clones are identified

by the variable domains of the Igs that they produce. In the variable domains of both light and heavy chains variability is concentrated in three regions identified as *complementary determining regions* (CDRs) that are exposed in the zone where the Ig develop its structural complementarity to the antigen. These regions are separated by more conserved regions called *frameworks* (FWRs). CDRs can be easily identified by inspection of the variability along the amino acid sequence of a variable region as shown in Figure 2.1.

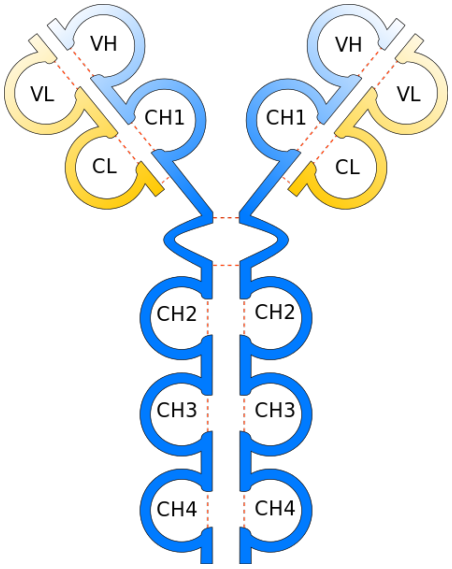


Figure 2.1. Schematic representation of the immunoglobulines structure. Loops represent Ig domains while red dashed lines stand for disulphide bonds.

Genetic V(D)J recombination

Three different genetic *loci* code for the Igs heavy, κ-light and λ-light chains. Each *locus* is localized on a different chromosome.

Pro-B cells display genes coding for Igs in their germinative configuration. For heavy chain this is organized with multiple V (variability) genes followed by multiple D (diversity) genes, multiple J (joining) genes and lastly by the nine C genes coding for the constant regions of the different Ig isotypes. For light chains the germinative configuration start with multiple V genes followed by multiple J genes and by a single C gene different for κ and λ chains. There are no D genes coding for the light chains.

The nomenclature for these genes is organized as follows. Let us consider for example the gene IGHV1-2*02: this indicate the allele 02 of V gene number 2 belonging to the family 1 of the heavy chain.

The number of V, D and J germline genes depends on the particular species. In human the germinative genetic configuration presents 85 IGHV, 20 IGHD and 6 IGHJ genes responsible for the coding of the variable part of the heavy chain, while for the light chain there are 1000 IGκV, only IGλV genes, 5 IGκJ and one IGλJ gene. In addition, V,D and J genes are known to be polymorphic.

When the cell differentiate a genetic recombination takes place during which single V,D, J and C germline genes for the heavy chain and single V, J and C germline genes are randomly selected and maintained in the genome while the rest is cut away. This mechanisms give rise to a *combinatorial diversity*. Moreover between the selected genes some nucleotides can be added or removed. Depending on the mechanisms of insertion they are called P (cut of hairpin loops) or N nucleotides. So, over the combinatorial diversity, a *junctional diversity* contribute to the diversification of the Abs repertoire of naïve B cells.

As described in Section 2.1.1, the last mechanism that contribute to the diversity of circulating Abs is the the affinity maturation of activated B cells.

Resuming, the different mechanisms that contribute to Abs diversity are

1. genetic V(D)J recombination;
    - combinatorial diversity,
    - junctional diversity,
2. negative selection in primary lymphoid tissues that avoids self recognition;
3. positive selection through affinity maturation.

The shape of an individual Igs repertoire is determined by the statistical distribution of the above mechanisms.

2.1.3 The T lymphocytes

T cells mainly differentiate in the thymus. They can be divided in many subgroups. The more coarse grain distinction within T cells can be done between *cytotoxic T lymphocytes* (CTL), *helper T lymphocytes* and *regulatory T lymphocytes* (Treg).

T cells recognize antigens through a membrane protein called *T cell receptor* (TCR). Like B lymphocytes also T cell differentiate into clones that express different TCRs, the membrane protein through which they recognize antigens. As for the Igs, alsoTCR diversity results from a V(D)J genetic recombination process. Differently form B lymphocytes, T cell do not undgergo any affinity maturation process.

While B cell can bind free antigens, T lymphocytes recognize antigens only when they are presented by other cells through the *major histocompatibility complex of class I or II* (MHC-I and MHC-II).

Among the most important activities of T cells is the production of *cytokines*. This is a huge and very complex class of molecules that locally modulate the immune response. Their mechanism of action is very complex, as a single type of cytokine has in general multiple cell targets and multiple effects.

Cytotoxic T lymphocytes

CTLs are also known as $CD8^+$ T lymphocytes as they express the CD8 receptor on their surface. CTLs bind their antigen when it is presented by MHC-I. The binding of MHC-I is mediated by CD8. MHC-I membrane molecules are expressed by all type of cells. When a cell is infected by a pathogen it can expose through the MHC-I an antigen that has been synthesized in the endogenous environment as

an effect of the infection. A CTL that recognize the exposed antigen can bind the infected cell and chemically induce its death. All this mainly happens in the sites of infections.

T helper cells

T helper cells recognize the antigen only when it is exposed through the MHC-II by the so called *antigen presenting cells* (APC). They characterized by the expression of the MHC-II, and the most relevant of them include dendritic cells (DC), macrophage cells and B lymphocytes.

DCs are responsible for the capture of antigens and their transport in the lymphoid follicles. There, they expose antigens to T helpers through the MHC-II. T helper cells express the surface receptor CD4 (for that are also called $CD4^+$ lymphocytes) that is responsible for the contact with MHC-II. In case of recognition the T cell gets activated, starts cytokine productions and the related clone gets expanded because of the cells proliferation.

An activated T helper that recognize the antigen exposed by macrophages elicits the activation of the latter through cytokines secretion promoting the elimination of the pathogen via a cascade of effects.

Activated T helper are also crucial for the activation of a naïve B cell that recognizes an antigen. In fact when this happens, the B lymphocyte process the antigen and expose a part of it on its membrane through the MHC-II. If the exposed peptide is recognized by an active T helper, the latter trigger the activation of the B cell involving proliferation, Abs secretion and class Ab switching. This mainly happens through local production of excitatory cytokines and the communication through the receptors CD40 on the T helper and the CD40L on the B cell.

So to be promoted to plasma cell and start its immune response, a B lymphocyte needs the check by an appropriate T helper cell that confirms the exogenous nature of the antigen. This double check is crucial for the immune system to avoid response to self molecules. Multiple checks are extensively present in several mechanisms of the immune system.

T helper cell can also negatively regulate B cells through the production of inhibitory cytokines. This multiple interactions in the adaptive immune response is responsible for the fine tuning of its response.

Regulatory T cells

Like B cells, T lymphocytes have to pass a negative selection in primary lymphoid tissues where T cells that respond to a sample of the self molecules are directed toward apoptosis. Some of these cells do not die but instead differentiate into *regulatory T cells* (Treg) that can negatively regulate macrophages, DCs (and so the T branch of the response that is activated by those APC) through the production of immunosuppressor cytokines like the Interleukin-10.

As Treg cells inhibit the activity of T helper cells that can activate B clones, Treg activity can be thought to have an inhibitory net effect on B cells.

2.1.4 The idiotypic network

The variable domain of an Ig produced by a B clone that is not expanded is a molecule that can hardly be found in the host body. But when a B clone is expanded, the Igs concentration of its idiotype, both in the soluble or membrane form, increases dramatically. This high concentration of a molecule that has never been so present in the body before could induce the immune system to react to it as to an external antigen. In fact, a part of the Igs variable region of the first expanded clone could be the epitope for other Abs. When located on the variable domain of an Ig this possible epitope is called *paratope*; the latter is so recognized by the *idiotope* of Abs produced by other B clones. Using this mechanism different B clones could form a network of interaction between different idiotypes known as *idiotypic network*. This hypothesis has been originally formulated in the 1970's by Niels Jerne in [24]. The role that was hypothesized for the idiotypic interactions was that of turning off B monoclonal responses after the disappearance of the antigenic stimulus, helping the recovering of the homeostasis regime.

Because of the experimental difficulties in proving the existence and the importance of such idiotypic interactions, the idiotypic network hypothesis did not have too much following after its proposal. Nevertheless, between late 1980's and early 1990's, it received a renewed attention after the theoretical works of Francisco Varela, António Coutinho and John Stewart. This authors recognized the possibilities of a systemic role of the idiotypic interactions. In fact, following the idiotyipic network hypothesis, an antigenic stimulation should not only produce a reaction of the B clones that directly recognize the antigen, but also of anti-antibodies that recognize the directly stimulated ones. So the effect of the antigenic stimulations should spread over the network of interacting B clones and the response should be thought as performed by the whole idiotypic network. In particular the antigenic stimulation should have different effects depending on the connectivity degree of the directly stimulated clone in the idiotypic network.

In a famous paper, [25], Stewart, Varela and Coutinho tested this hypothesis by implementing a mathematical model on an experimentally identified subnetwork of a real idiotypic network of a few tens of B clones. As a result of the model implementation, two situations can be discriminated: when poorly connected clones are stimulated by antigens they are free to extensively respond; on the other hand, the stimulation of a highly connected clone generate a weak expansion of the clone itself and the response is spread over more clones that weakly expand to contrast the production of the directly stimulated Abs.

The authors' interpretation of this theoretical result was that the experimentally validated presence of self directed B clones in healthy individuals could be justified by the regulatory action of the idiotypic network that prevents these B cells to extensively react to the self. Within this framework, some autoimmune diseases could be though as caused by pathological modification on the network structure such that self directed B clones are not adequately connected and the network cannot control their self reaction.

It has to be pointed out that a limitation in [25] was that the very limited size of the considered real interaction network made hard to extend the conceptual conclusions to an entire idiotypic network.

Chapter 3

Statistical mechanics based models for the adaptive response in the immune system

The main subject of this Chapter is the description of the model for the generation of a synthetic idiotypic network that we proposed in [1]. This work will be depicted in Section 3.1.

Furthermore in Section 3.2 we will present a review of the *corpus* of models proposed by Elena Agliari, Adriano Barra and coworkers aimed to underline possible analogies of the immune system interaction networks with some mathematical models inspired by neural networks. The idiotypic network plays a crucial role also in this framework so we will show how our model for the idyotipic network can be embedded in a wider and consistent theoretical apparatus.

3.1 Organization and evolution of synthetic idiotypic networks

In Section 2.1.4 we described how different B clones can interact via the antibody/anti-antibody mechanism originally proposed by Jerne [24], giving rise to an effective imitative interaction network between the clones, the *idiotypic network*. In that Section it was pointed out that, as highlighted later by Varela [25, 26], this network could have a systemic role that permits the existence of non dangerous self reactive B clones in a living organism. In fact Varela's idea is that, if a B clone is highly connected in the idiotypic network, the repression induced by the topological neighbors can be so high that the clone never gets activated, even in presence of a massive concentration of antigen. So, following this reasoning, self directed B clones could exist in a living organism provided that they are highly connected in the idiotypic network. Varela and coworkers enforced this idea by the use of a mathematical model of interacting B clones based on coupled differential equations.

The topology of the idiotypic network plays a crucial role in the repression mechanism proposed by Varela, so a mathematical model for generating realistic synthetic idiotypic networks is of great interest for the mathematical modeling of

the immune system. Agliari and Barra in [27] proposed a very simple mathematical rule for constructing networks that should mimic features of the idiotypic networks. Here and in [1] we propose an upgrade of this rule which goes in the direction of a more realistic description of the biological reality. Basically this modification deals with the fact that the antibody repertoire of an individual is not completely random, as hypothesized in [27] as a first approximation, but instead a certain degree of correlation in the generation of the antibodies needs to be considered. The structure of this correlation is caused in the real world by several complex mechanisms such as genetic recombination, self/non self recognition and maturation of the immune repertoire due to infections. Despite this high complexity, here we propose an ultra simplified mathematical model for the generation of antibodies repertoires and antibodies interactions embedding all the correlation of the repertoire in a unique parameter. The idea is also that, with the passing of time, an individual encounters more and more antigens that have a correlated (not totally random) structure and so, in response, its repertoire becomes more and more correlated, so that the correlation parameter should be thought as a (slowly) evolving quantity of an idiotypic network. We study the effect of the tuning of this correlation parameter on the topology of the network focusing in particular on the differences between the weighted and bare topologies, showing that the mechanism proposed by Varela does not need, in this framework, the existence of an over percolated bare network, a scenario which seems more in agreement with experimental findings.

This Chapter is organized as follows: in Section we introduce a class of weighted graphs meant to describe idiotypic networks. Then, such a model is investigated for its global topology in Section 3.1.2, including degree distribution and clustering features, and as for the distribution of weights associated with its links (3.1.3). Next the ageing of the system by increasing the degree of bias among bit strings; this process induces a progressive dilution of the network and the resulting percolation is analyzed in Section 3.1.4. Finally, in Section 3.1.5, the results are discussed together with possible perspectives

3.1.1 The model for the generation of a correlated idiotypic network

As described in Section 2.1.4, the idiotypic network is a network of imitative effective interactions between different B clones mediated by immunoglobulins. Our aim is here that of constructing a rule to assign an interaction strength to a couple of B clones, given their idiotypicity, i.e. given the sequence of the antibodies that they produce.

Let us consider a set of N_B B clones. Each B clone is identified by its idiotypicity, namely by the sequence of the variable region of the antibody that it expresses. Let us call b_μ a logarithmic measure for the concentration of cells of the μ-th clone with respect to the physiological concentration c_r of a clone in absence of antigenic stimulation:

$$b_\mu = \log(\frac{c_\mu}{c_r}) \, ; \quad \mu = 1, 2, \ldots, N_B \, . \tag{3.1}$$

A possible differential equation for the evolution of the idiotypic network is a Lagevin

equation of the form

$$\frac{db_\mu}{dt} = \sum_{\nu=1}^{N_B} J_{\mu\nu}(b_\nu - b_\mu) + \sum_k J_{\mu k} A_k + \sqrt{\frac{1}{\beta}}\eta_\mu \,, \tag{3.2}$$

where the couplings $J_{\mu\nu}$ in the first term encode the contribution of the idiotypic interactions, the second term mimic the antigenic stimulation by antigens with concentration A_k that can simulate multiple clones though the matrix $J_{\mu k}$ and the last contribution is a usual white noise term.

Here we do not concentrate on the time evolution of the system ruled by the above equation, but instead on a topological study of the idiotypic network $J_{\mu\nu}$. In order to construct an effective interaction network described by interaction couplings $J_{\mu\nu}$, we make the minimal hypothesis that the information about the specificity of the antibodies is encoded in a binary string of length L that we indicate with:

$$\Psi_\mu^i = 0, 1 \,, \quad \text{where} \quad \mu =, 1, \dots N_B \quad \text{and} \quad i = 1, \dots L \,. \tag{3.3}$$

The strings entries are chosen to have biased distribution: we assume that each entry μ of the i^{th} string is extracted randomly according to the distribution

$$P(\Psi_\mu^i = 1) = \frac{(1+a)}{2}, \; P(\Psi_\mu^i = 0) = \frac{(1-a)}{2}, \tag{3.4}$$

with $a \in [-1, +1]$. In this way, when $a = 0$, we recover the previous unbiased model [28] and, in general, the average similarity between a pair of strings can be tuned via a as $\langle \Psi_\mu^i \Psi_\nu^i \rangle \sim (1+a)^2/4$. As we will see, Equation ((3.4)) provides a basic way to bias the repertoire, which allows us to study the direct effects on the network performance; more refined models can of course be obtained.

Given a couple of clones, say μ and ν, the μth entries of the corresponding strings are said to be complementary, iff $\Psi_\mu^i \neq \Psi_\nu^i$. Therefore, the number of complementary entries $\chi_{\mu\nu} \in [0, L]$ can be written as $\chi_{\mu\nu} = \sum_{i=1}^{L} [\Psi_\mu^i (1 - \Psi_\nu^i) + \Psi_\nu^i (1 - \Psi_\mu^i)]$. Of course, $\chi_{\mu\nu}$ strongly depends on the correlation parameter a and, in turn, it directly affects the affinity between μ and ν. In fact, the non-covalent forces acting among antibodies depend on the geometry, on the charge distribution and on hydrophilic-hydrophobic effects which give rise to an attractive (repulsive) interaction for any complementary (non-complementary) match. In principle, once the protein folding problem is solved [29], the whole analysis of this kind of network could be extremely simplified by directly studying the VDJ genes and their reshuffling; however, as this bridge among micro and meso biological scenarios is lacking, we rely on "effective descriptions." In particular, we assume that each complementary/non-complementary entry yields an attractive/repulsive contribution [30, 31]; the ratio between their intensities is denoted by the positive parameter α. Hence, we introduce the measure for the affinity between Ψ_μ and Ψ_ν,

$$f_{a,\alpha,L}(\Psi_\mu, \Psi_\nu | a) \equiv [\alpha \chi_{\mu\nu} - (L - \chi_{\mu\nu})], \tag{3.5}$$

which ranges from $-L$ (when $\Psi_\mu = \Psi_\nu$) to αL (when all entries are complementary, i.e. $\Psi_\mu^i \equiv 1 - \Psi_\nu^i$, $i = 1, ..., L$). Now, when the repulsive contribution prevails, that

is, $f_{a,\alpha,L} \leq 0$, the two antibodies do not see each other and the coupling among the corresponding lymphocytes $J_{\mu\nu}(a, \alpha, L)$ is set equal to 0; conversely, when $f_{a,\alpha,L} > 0$, we take as $J_{\mu\nu}$ the exponential of the affinity. This choice is the simplest trial able to mimic a key-lock mechanism for a sharp pattern recognition. Thus, we have:

$$J_{\mu\nu}(a, \alpha, L) \equiv \Theta(f_{a,\alpha,L}(\Psi_\mu, \Psi_\nu|a)) \exp[f_{a,\alpha,L}(\Psi_\mu, \Psi_\nu|a)], \tag{3.6}$$

where $\Theta(x)$ is the discrete Heaviside function [1] returning 1, if $x > 0$, and 0, if $x \leq 0$. Indeed, the expression in Equation ((3.6)) ensures that the coupling strength among lymphocytes spans several orders of magnitude ($J_{\mu\nu} \in [0, \exp(\alpha L)]$), as expected from experimental results [32].
One could possibly introduce a proper normalization in order to fix an average value for the coupling strength, which in turn fixes a scale for the level of noise ruling the thermodynamics of the system. This procedure is allowed due to the fact that, as the system is a ferromagnet, the average coupling is positive definite and that, for a given size N, couplings display an upper bound. Nonetheless, the normalization is somehow arbitrary and does not qualitatively affect the behavior of the system. As for our current aims, we can neglect this and take Equation ((3.6)) as an effective definition for the coupling strength between node i and node j.

As we are interested in studying the limit of large networks ($N >> 1$), we also stress that N and L are intrinsically connected to each other. This can be easily seen in the case where the match among antibodies had to be perfect for reciprocal recognition; then, in order to reproduce all possible antibodies obtained by the L entries in the bitstrings, the immune system would need $N_B = 2^L$ lymphocytes. Here, having relaxed the hypothesis of perfect match, only a fraction of this quantity needs to be retained to manage the whole repertoire, and we can introduce the following scaling between the number of all possible idiotypically different lymphocytes and the effective size of the repertoire:

$$L = \gamma \log N_B, \tag{3.7}$$

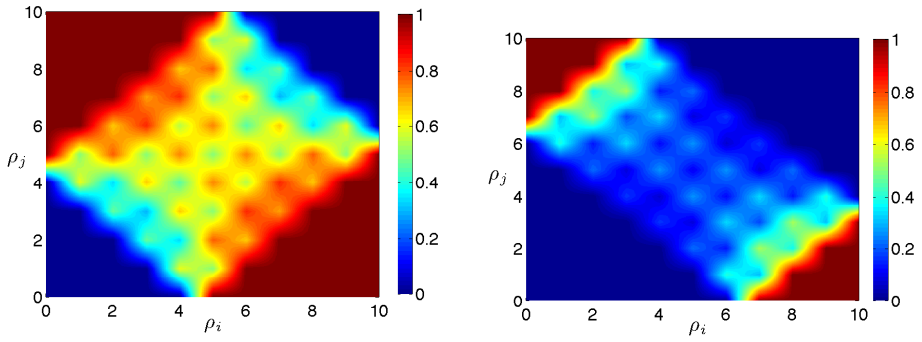
where $\gamma > 0$.

As we will see in more detail in Section 3.2 the the B clones system can be considered in analogy with a ferromagnetic system of *soft* (gaussian) spins. Moreover, in that Section, we will consider bipartite networks made up of B cells and T cells clones and we will see that the ferromagnetic interactions in the B cells party will give rise to interesting emerging phenomena in complete system. Let us now concentrate on the topological analysis of idiotypic networks defined here, following [1].

3.1.2 Global topology

If we forget, for the moment, the weights of the couplings, the N different B clones, interacting pairwise, define a graph $\mathcal{G} = \{V, \Gamma\}$, where V denotes the set of nodes and Γ the set of links. The cardinality of V is given by $|V| = N_B$, that is the

[1] The effect of the theta function is to remove those links whose strength is mathematically different from zero, but, from a practical point of view are so weak that whatever level of noise would clean them as well.



(a) (b)

Figure 3.1. (Color on line) Link probability $P_{\text{link}}(\rho_\mu, \rho_\nu; \alpha, L)$ as a function of ρ_μ and ρ_ν with $L = 10$ and $\alpha = 0.5$ (left) or $\alpha = 1.1$ (right).

total amount of idiotypically different clones. The topological properties of $\mathcal{G}$ are completely determined by the adjacency matrix A defined as $A_{ij} = 1$ if $J_{ij} \neq 0$ and $A_{ij} = 0$ if $J_{ij} = 0$. For instance, the degree of a node i (i.e. coordination number) is given by $z_i = \sum_{j \in V} A_{ij}$. In the following, we provide the main definitions and formula to describe the topology of the emergent graph and later we will deepen its global features.

First, let us introduce the probability that a string Ψ_i displays ρ non null entries; this follows a binomial distribution

$$P(\rho; a, L) = \binom{L}{\rho} \left(\frac{1+a}{2}\right)^{\rho} \left(\frac{1-a}{2}\right)^{L-\rho} . \tag{3.8}$$

Correspondingly, the probability that two strings ξ_i and ξ_j, displaying ρ_i and ρ_j non-null entries respectively, exhibit χ complementary matches is

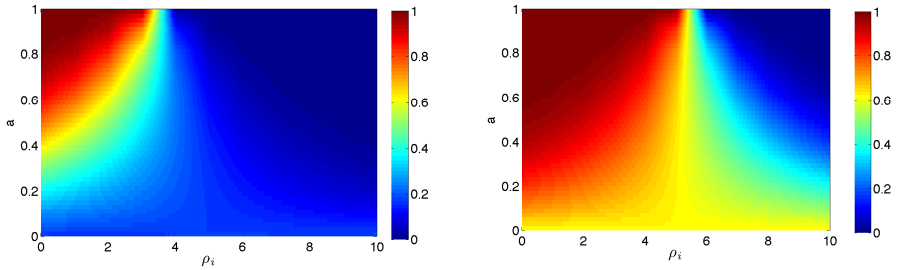
$$P(\chi; \rho_\mu, \rho_\nu) = \binom{L}{\rho_\mu}^{-1} \binom{L}{\rho_\nu}^{-1} \times \frac{L!}{\left(\frac{\rho_\mu - \rho_\nu + \chi}{2}\right)! \left(\frac{\rho_\nu - \rho_\mu + \chi}{2}\right)! \left(\frac{\rho_\mu + \rho_\nu - \chi}{2}\right)! \left(L - \frac{\rho_\mu + \rho_\nu + \chi}{2}\right)!} . \tag{3.9}$$

Now, the link probability $P_{\text{link}}(\rho_\mu, \rho_\nu; \alpha, L)$ can be obtained by summing $P(\chi; \rho_\mu, \rho_\nu)$ over the values of χ compatible with (ρ_μ, ρ_ν) and such that $f_{\alpha,L}(\Psi_\mu, \Psi_\nu) > 0$, namely, recalling Equation (3.9),

$$P_{\text{link}}(\rho_\mu, \rho_\nu; \alpha, L) = \sum_{\chi = \max(|\rho_\mu - \rho_\nu|,\ \lceil \frac{L}{\alpha+1} \rceil)}^{\min(\rho_\mu + \rho_\nu,\ 2L - \rho_\mu - \rho_\nu)} P(\chi; \rho_\mu, \rho_\nu) . \tag{3.10}$$

Figure 3.1 shows an example of $P_{\text{link}}(\rho_\mu, \rho_\nu; \alpha, L)$ for different choices of the parameter α.

By averaging $P_{\text{link}}(\rho_\mu, \rho_\nu; \alpha, L)$ over ρ_i, we get the average link probability for a



(a) (b)

Figure 3.2. (Color on line) Link probability $P_{\text{link}}(\rho_\mu; a, \alpha, L)$ as a function of ρ_μ and a, while α and L are fixed. In the left panel $\alpha = 0.5$ while in the right panel $\alpha = 1.1$; in both cases $L = 10$. It is straightforward to see that on the line $a = 0$ the probability link is independent on ρ_μ. When $a \neq 0$, the link probability changes with varying ρ_μ. Due to symmetry, only the range $a \in [0, 1]$ is shown.

node whose string displays ρ_μ non-null entries, that is,

$$P_{\text{link}}(\rho_\mu; a, \alpha, L) = \sum_{\rho_\nu=0}^{L} P(\rho_\nu; a, L) P_{\text{link}}(\rho_\mu, \rho_\nu; \alpha, L), \tag{3.11}$$

from which the average degree for j reads

$$z(\rho_\mu; a, \alpha, L, N_B) = N_B P_{\text{link}}(\rho_\mu; a, \alpha, L). \tag{3.12}$$

Numerical calculations of $P_{\text{link}}(\rho; a, \alpha, L)$ are shown in Figure 3.2: Notice that a uniform bit distribution within the antibodies (i.e. $a = 0$) corresponds to an unbiased graph, where the average link probability of a node does not depend on the pertaining string.
By further averaging $P_{\text{link}}(\rho_\mu; a, \alpha, L)$ over ρ_μ, we get the average link probability for an arbitrary node of the system

$$\overline{P}_{\text{link}}(a, \alpha, L) = \sum_{\rho_\mu, \rho_\nu=0}^{L,L} P_{\text{link}}(\rho_\mu, \rho_\nu; \alpha, L)\, P(\rho_\mu; a, L)\, P(\rho_\nu; a, L),$$

from which the average coordination number follows as

$$\overline{z}(a, \alpha, L, N_B) = N_B \overline{P}_{\text{link}}(a, \alpha, L). \tag{3.13}$$

Finally, within a mean-field approach, we can use $P_{\text{link}}(\rho; a, \alpha, L)$ to write the degree distribution:

$$P_{\text{degree}}(z|\rho; a, \alpha, L, N_B) = \binom{N_B - 1}{z} \left[P_{\text{link}}(\rho; a, \alpha, L)\right]^z \left[1 - P_{\text{link}}(\rho; a, \alpha, L)\right]^{N_B - 1 - z},$$

and, by further averaging over ρ,

$$\overline{P}_{\text{degree}}(z; a, \alpha, L, N_B) = \sum_{\rho=0}^{L} P_{\text{degree}}(z|\rho; a, \alpha, L, N_B)\, P(\rho; a, L). \tag{3.14}$$

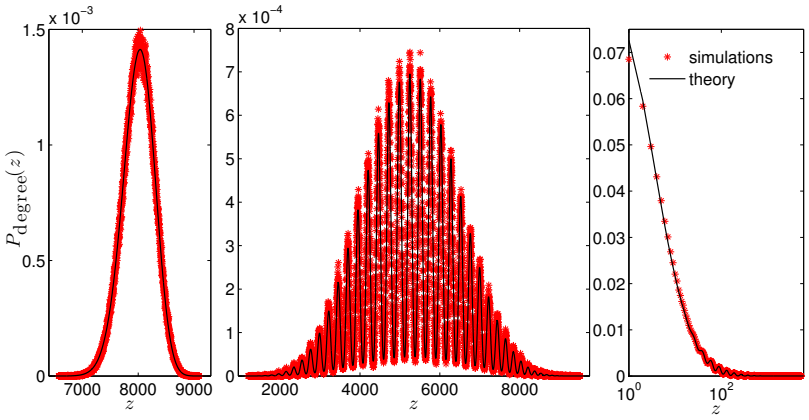


Figure 3.3. (Color on line) Degree distribution for different values of a and fixed $\alpha = 1.2$, $N = 10000$, $\gamma = 10$. From left to right: $a = 0.1$ (uni-modal behavior; the mean degree is about $0.8N$), $a = 0.3$ (multimodal behavior; the mean degree is $0.5N$), and $a = 0.6$ (there is no extensive network; the average degree is $0.03N$).

Multimodal degree distribution

As shown by Equation ((3.14)), $\overline{P}_{\text{degree}}(z; a, \alpha, L, N)$ is the sum of L binomial distributions, each referring to a different "mode" ρ. Therefore, the average degree distribution will show a multimodal behavior as long as two consecutive modes have disjoint supports.

In general we expect that, when a is close to 0, peaks merge since, at low bias, the link probability weakly depends on ρ (see also Figure 3.2), which is uniformity, while, when a is close to 1, most of the modes are off and the network is sparse. Therefore a very multimodal distribution is expected only for intermediate values of a. Indeed, these arguments are corroborated by Figure (3.3) which shows different plots of $\overline{P}_{\text{degree}}(z; a, \alpha, L, N_B)$ for different values of a; also notice that numerical data are finely fitted by the analytical calculation of Equation ((3.14)).

As already pointed out the parameter a can be thought to rise in absolute value with the ageing of an organism. Following this hypothesis the correlations in the structure of the network should increase with time in an individual and so the multimodal structure of the degree distribution should should be thought as an evolving feature of an idiotypic network.

Scaling in thermodynamics limit.

As evidenced in the previous subsection, a crucially affects the topology of the idiotypic network. We therefore investigate in more detail the global connectivity of the system in terms of $\overline{P}_{\text{link}}(a, \alpha, L, N_B)$. To compute this quantity, one would plug Equation (3.9) into Equation (3.10) and (3.13). However, since here we are interested in the large L limit and relative fluctuations in ρ decrease as $1/\sqrt{L}$, we can adopt a mean-field like approach and approximate Equation (3.8) as

$$P(\rho; a, L) \simeq P_{MF}(\rho; a, L) = \delta(\rho - \bar{\rho}), \tag{3.15}$$

with $\bar{\rho} = (1+a)L/2$, so that Equation (3.9) can be restated as

$$P_{MF}(\chi; a, L) = \frac{\binom{(1+a)L/2}{\chi/2}\binom{(1-a)L/2}{\chi/2}}{\binom{L}{(1+a)L/2}}. \tag{3.16}$$

such that

$$\langle \chi \rangle_{MF} = \frac{1-a^2}{2} L \tag{3.17}$$

$$\langle \chi^2 \rangle_{MF} - \langle \chi \rangle^2_{MF} = \left(\frac{1-a^2}{2} \right)^2 L \,. \tag{3.18}$$

So in the limit of large system the distribution of the complementary bit-strings can be approximated, using Central Limit Theorem, by

$$P_{MF}(\chi; a, L) \simeq \mathcal{N}\left(\langle \chi \rangle_{MF} \,, \frac{\langle \chi \rangle_{MF}}{L} \right) \tag{3.19}$$

and in the infinite system limit fluctuations of the number of complementarities vanish.

Moreover, exploiting the parity symmetry for a, we focus on the range $a \in [0, 1]$. Thus, Equation (3.10) can be approximated as

$$\begin{aligned} \overline{P}_{\text{link}}(a, \alpha, L, N) &\approx \\ \overline{P}^{MF}_{\text{link}}(a, \alpha, L, N) &= \frac{\sum_{\chi=\lceil \frac{L}{2(\alpha+1)} \rceil}^{(1-a)L} \binom{\frac{1+a}{2}L}{\chi/2}\binom{\frac{1-a}{2}L}{\chi/2}}{\binom{L}{\frac{1+a}{2}L}}. \end{aligned} \tag{3.20}$$

Now, since we are interested in scaling laws, we can neglect all terms in the sum, but the leading one. For instance, focusing on $a < 1/2$ ($a > -1/2$) and $\alpha > 1$, it is easy to see that this is given by $\chi = L/2$. Then, via Stirling approximation, we get

$$\log \overline{P}_{\text{link}}(a, \alpha, L, N_B) \sim f(a) L, \tag{3.21}$$

where $f(a) = (1+a)\log(1+a) + (1-a)\log(1-a) - (1+2a)/4 \log(1+2a) - (1-2a)/4 \log(1-2a)$ is a symmetrical monotonically decreasing function from $a = 0$ to $a = \pm 1/2$ and roughly plays the role of an (compressed in the interval $[-1/2, +1/2]$) entropy of the bit-strings, such that, as in other biased approaches [33], a can be thought of as the "bit-string magnetization".

Therefore, we can write

$$\overline{P}_{\text{link}}(a, \alpha, L, N) \sim N_B^{\gamma f(a)}. \tag{3.22}$$

In this way, as the degree of bias a is increased, the graph turns from highly connected ($\overline{P}_{\text{link}} = \mathcal{O}(1)$) to diluted ($\overline{P}_{\text{link}} \sim N_B^{-\gamma c}$, with $c = \log(27/32)/2$). Similar results can be found for a different parameter range.

Clustering Coefficient

The local clustering coefficient is defined as the number of triangles stemming from a node μ over the maximum number of triplets centered on μ itself (see, e.g., Refs. [34, 35, 36]). Due to the "anti-transitive" nature of the idyotipic network, triangles, i.e., 3-cycles, are expected to be unlikely, while quadrilaterals, i.e. 4-cycles, are expected to be favored. Roughly speaking, an antibody Ab1 elicits its anti-antibody Ab2, which in turn elicits the anti-anti-antibody Ab3, whose structure should be close to Ab1's, so that a 4-cycle finally develops.

Indeed, this was partially shown in Ref. [28], where an idiotypic network and an analogous Erdös-Rény (ER) graph, namely, a purely random graph exhibiting the same average degree $\bar{z}$ were compared, obtaining that the former displays on average a significantly smaller number of triangles. We now extend those results considering also 4-cycles (devoid of diagonals). Given a node i with z_μ nearest neighbors, the number of expected squares stemming from i, in the case where i belongs to a random ER graph and in the case where it belongs to our idiotypic network, are respectively

$$Q_{ER}(z_\mu) = \binom{z_\mu}{2}(N-1-z_\mu)p^2(1-p), \tag{3.23}$$

$$Q(z_\mu) = \binom{z_\mu}{2}(N-1-z_\mu)p'^2(1-p''), \tag{3.24}$$

where $p = \bar{z}/N$, p' is the probability that in $\mathcal{G}$ a neighbor of i is linked to a node not belonging to the ith neighborhood, and p'' is the probability that two neighbors are linked. The latter is just the clustering coefficient for node i, which, as shown in Ref. [28], is smaller for a graph where links are based on complementarity features, so that $(1-p) \leq (1-p'')$. Moreover, the idiotypic and the ER graphs we are comparing display, by construction, the same coordination number. If we impose this condition to be true also for the average degree of any site j that is linked with i, we get

$$\begin{aligned} \bar{z_\nu} &= 1 + p''(z_\mu - 1) + p'(N - z_\mu - 2) \\ &= 1 + p(z_\mu - 1) + p(N - z_\mu - 2). \end{aligned} \tag{3.25}$$

Therefore, as $p > p''$, we get $p < p'$ and finally $Q(z_\mu) > Q_{ER}(z_\mu)$. More generally, this suggests that in our idiotypic networks, as links are based on complementarity, 4-cycles are motifs while 3-cycles are anti-motifs [28, 37]. Indeed, Figure 3.4 numerically confirms that the number of quadrilaterals [triangles] appearing in our graph is larger [smaller] than the number expected for an analogous ER graph, estimated as $\binom{N}{4}p^4(1-p)^2$, $[\binom{N}{3}p^3]$.

3.1.3 Coupling and weighted degree distributions

In Section 3.1.2 we studied the bare topology of the network, while here we focus on the properties related to the distribution of weights $J_{\mu\nu}$ associated with links (μ, ν). As mentioned above, these features retain a strong biological meaning. For instance, nodes displaying a high weighted degree feel, under normal conditions, a larger (internal) quiescent stimulus [27].

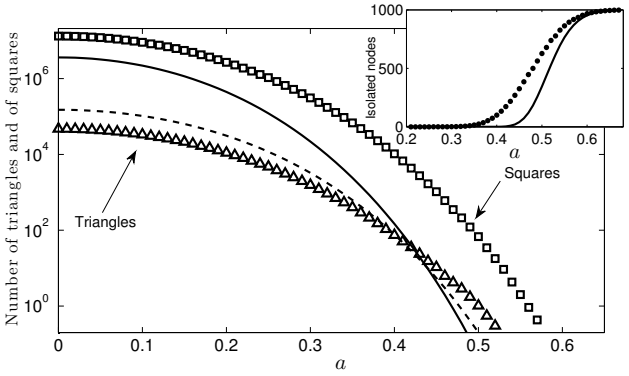


Figure 3.4. Number of triangles (△) and of quadrilaterals (□) averaged over 100 realizations in our idiotypic network as a function of a. Parameters characterizing the network are $N_B = 1000$, $\gamma = 7$ and $\alpha = 0.7$. Curves represent the number of triangles (dashed line) and of squares (solid line) expected for an analogous ER graph. Inset: Number of isolated nodes present in the system; again, the idiotypic network (•) and ER graph (line) are compared. The latter displays a larger number of triangles as long as a giant component can be detected.

Coupling distribution

Given two nodes μ and ν with ρ_μ and ρ_ν non-null entries, respectively, recalling Eqs. (2) and (3), the coupling distribution is

$$P(J_{\mu\nu}|\rho_\mu,\rho_\nu;\alpha,L) = \begin{cases} P(\chi_{\mu\nu} = \frac{\log(J_{\mu\nu})+L}{\alpha+1};\rho_\mu,\rho_\nu,L) & \text{if } J_{\mu\nu} > 1 \\ 0 & \text{if } J_{\mu\nu} \leq 1 \end{cases} , \qquad (3.26)$$

where $J_{\mu\nu}$ can span over $[1, e^{\alpha L}]$.

To obtain the mean coupling probability one should average Equation ((3.26)), so $P(\chi;\rho_\mu,\rho_\nu)$ of Equation (3.9), over the binomial distribution of ρ_μ and ρ_ν in Equation (3.8). To give an analytical estimate of this quantity, as in section 3.1.2, one can use a mean-field like approach, namely Equations (3.16), (3.17) and (3.18), so that, in the limit of large system the number of complementary entries is normally distributed following

$$\chi_{MF} \sim \mathcal{N}\left(\frac{1-a^2}{2}L,\ (\frac{1-a^2}{2})^2 L\right) \qquad (3.27)$$

as the Θ function in the definitions (3.6) of the couplings set to zero couplings that are smaller than zero, we can neglect it for the sake of analytical simplicity in determining the approximate coupling distribution considering $J = \exp[(\alpha+1)\chi - L]$.

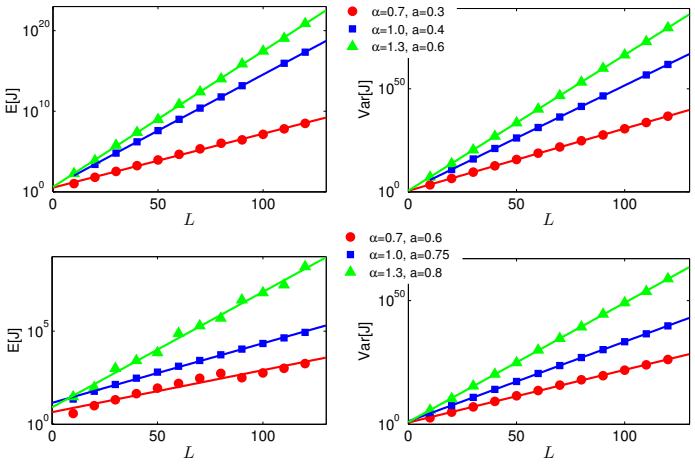


Figure 3.5. (Color on-line) Average value of the coupling strength E[J] and of its variance Var[J], versus L for several choices of parameters, depicted in different colors. We considered $\alpha = 0.7$ ($\circ$), $\alpha = 1.0$ ($\square$) and $\alpha = 1.3$ ($\triangle$) for different values of a pertaining to a connected (upper panels) and disconnected (lower panels) regimes (see the legend). Symbols refer to data obtained from exact numerical calculations, while curves are drawn according to the approximations (3.31) and (3.32).

So in the mean filed approximation the couplings J are log-normally distributed as

$$\langle \log J \rangle_{MF} \simeq [(\alpha+1)\frac{1-a^2}{2} - 1]L\,, \tag{3.28}$$

$$\mathrm{Var}_{MF}[\log J] \simeq (\alpha+1)^2 \left(\frac{1-a^2}{2}\right)^2 L\,, \tag{3.29}$$

$$P_{MF}(J) \simeq \log \mathcal{N}\Big(\langle \log J \rangle_{MF}\,, \sqrt{\mathrm{Var}_{MF}[\log J]}\Big) = \\ = \frac{1}{J\sqrt{2\pi L \frac{1-a^2}{2}(\alpha+1)}} e^{-\frac{1}{2}\frac{[\log J - ((\alpha+1)(1-a^2)/2-1)]^2}{L(\alpha+1)(1-a^2)/2}}\,. \tag{3.30}$$

Thus, by calling $\theta := (\alpha+1)\frac{1-a^2}{2} > 1$, and using the properties of the log-normal distribution and the scaling relation 3.7,

$$\langle J \rangle_{MF} \simeq N_B^{\gamma[\theta^2/2+\theta-1]} \tag{3.31}$$

$$\mathrm{Var}_{MF}[J] \simeq N_B^{2\gamma[\theta^2+\theta-1]}\,. \tag{3.32}$$

Hence, $\langle J \rangle_{MF}$ is expected to scale as a power of the system size and exponentially with a^2. These results have been successfully checked by numerical simulations (see Figs. 3.5 and 3.6).

Weighted connectivity

We now extend the bare degree $z_\mu = \sum_\nu A_{\mu\nu}$ to a *weighted* degree referred to as w_i and defined as

$$w_\mu = \sum_\nu J_{\mu\nu}. \tag{3.33}$$

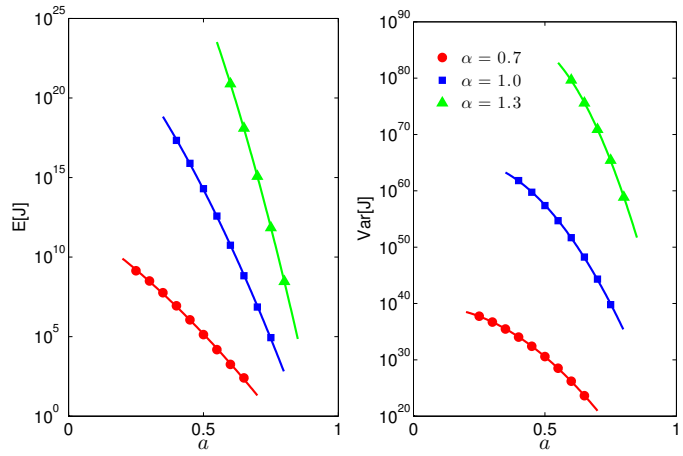


Figure 3.6. (Color on line) Average value of the coupling strength E[J] and of its variance Var[J], versus a for several choices of parameters, depicted in different colors. We considered $\alpha = 0.7$ ($\circ$), $\alpha = 1.0$ ($\square$) and $\alpha = 1.3$ ($\triangle$) and $L = 120$ (see the legend). Symbols refer to data obtained from exact numerical calculations, while curves are drawn according to the approximations (3.31) and (3.32).

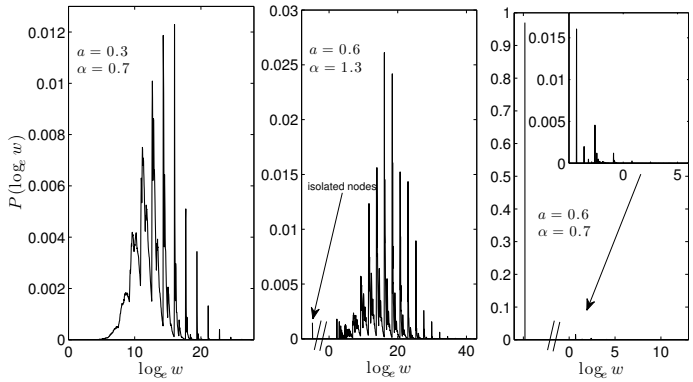


Figure 3.7. Semilogarithmic plot for the distribution $P(w; a, \alpha, L)$ obtained by averaging 100 systems made of $N = 2 \times 10^5$ nodes with $\gamma = 7$; results shown were averaged over 10^3 different realizations. We considered several values of a and α, corresponding to either over- or under- percolated regimes.

As we will see in more detail in the next Section, as the clones are connected by the couplings J the weighted topology is much more significative than the bare one.

In general, w_μ depends on the number of nearest-neighbors z_μ and on the coupling with each of them. So we define the weighted degree probability as follows:

$$P(w_\mu|\rho_\mu, z_\mu; a, \alpha, L) = \sum_{J_{\mu 1}....J_{\mu z_\mu}} P(J_{\mu 1}|\rho_\mu)\ldots P(J_{\mu z_\mu}|\rho_\mu)\ \delta(w_\mu - \sum_{\nu=1}^{z_\mu} J_{\mu\nu})\,. \quad (3.34)$$

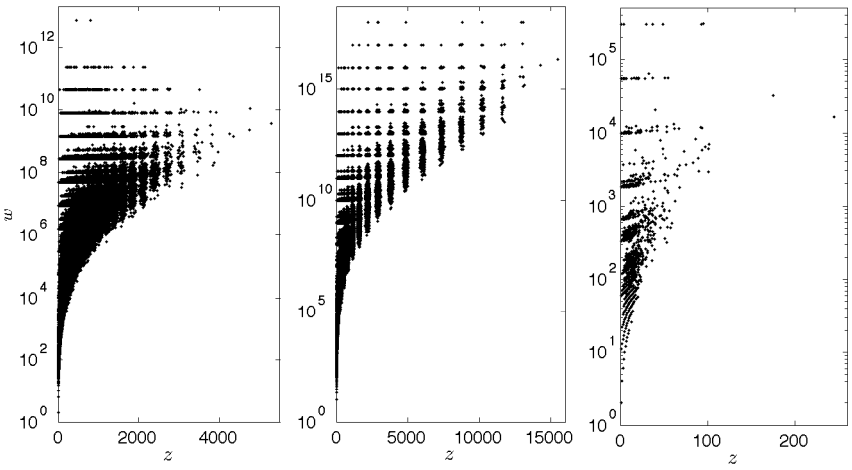


Figure 3.8. Each point shown corresponds to a node of the idiotypic network and its coordinates correspond to its weighted degree w and its bare degree z, respectively. The 100 system simulated is made up of $N = 2 \times 10^5$ nodes and we fixed $\gamma = 7$. We considered the same values of a and α used in Figure 7. Notice that nodes displaying a different degree (and therefore a different ρ), may display the same weighted degree. The reshuffling among weighted and non-weigheted peaks is evident.

Now, averaging over all the possible numbers of nearest-neighbor,

$$P(w_\mu|\rho_\mu; a, \alpha, L) = \sum_{z_\mu} P(z_\mu|\rho_\mu; a, \alpha, L) P(w_\mu|\rho_\mu z_\mu; a, \alpha, L) \tag{3.35}$$

and averaging over nodes we obtain the mean degree distribution

$$P(w; a, \alpha, L) = \sum_{\rho_\mu} P(w_\mu|\rho_\mu; a, \alpha, L) P(\rho_\mu; a, L). \tag{3.36}$$

This is the theoretical description of the empirical (by numerical simulations) distributions in Figure 3.7 which were obtained by numerical simulations of the network. The weighted degree distribution displays a fine structure similarly to $P_{\text{degree}}(z)$ (see Section (3.1.2)), and even a "hyperfine" structure. This is ultimately due to the fact that Equation (3.34), being a sum of terms that can be localized in the w range, can, by itself, display a multimodal distribution. When summing over ρ_i in Equation ((3.36)), several multimodal distributions are superposed, giving rise to the complicated structure shown in Figure 3.8. Otherwise stated, ρ_i may univocally determine a range for the degree z_i [leading to a multimodal $P_{\text{degree}}(z)$], but z_i, in turn, does not univocally determine a range for w_i. The "reshuffling" between bare and weighted degrees can be seen in the scatter plots in Figure 8 and it is mirrored by the non-trivial structure of $P(w)$. As a result, nodes that are lazier in reacting to antigenic stimulation are not necessarily those with a large number of neighbors, but, rather, those with a large weighted degree; the two subsets cannot be trivially mapped into each other.

It is worth emphasizing that, as the weighted degree is a sum of exponential factors, the support of $P(w)$ covers several orders of magnitude, in both the connected and the disconnected regimes. This is consistent with the co-existence of highly

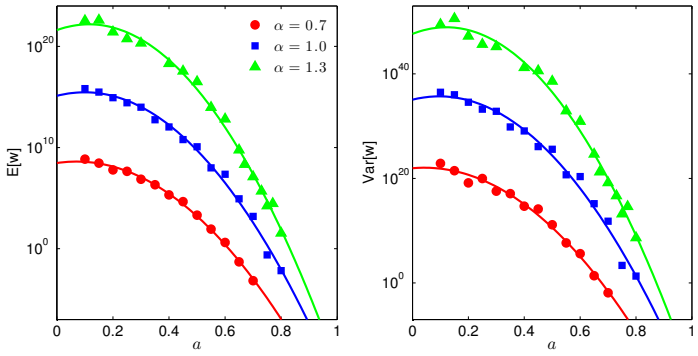


Figure 3.9. (Color on line) Average weighted degree $E[w]$ and its variance $Var[w]$ versus a for a system of $N = 10^4$ sites, $\gamma = 7$ and $\alpha = 0.7$ ($\circ$), $\alpha = 1.0$ ($\square$), and $\alpha = 1.3$ ($\triangle$). Symbols represent data from numerical simulations, while curves are the best fit obtained from Equation ((3.37)).

(poorly) susceptible agents with respect to external stimuli, i.e. nodes with small (large) w. This difference has been attributed to the self-addressed or non-self-addressed attitude of lymphocytes [38, 39], and, interestingly, it also survives in the underpercolated regime.

In particular, we can introduce the relations

$$\begin{aligned} \langle w \rangle &\sim N_B \langle J \rangle , \\ \mathrm{Var}\,[w] &\sim N_B \,\mathrm{Var}[J], \end{aligned} \tag{3.37}$$

which hold as long as the couplings insisting on the same nodes can be approximated as independent. The expressions in Equation ((3.37)) have been used to fit the numerical data in Figure (3.9).

As shown in [40], the envelope of the distribution $P(w)$ can be approximated by a log-normal distribution.

Finally, we stress that the support of $P(w)$ remains spread over several order of magnitude also in the region of the parameter space where the network is underpercolated (see Section 3.1.4): This suggests that the existence of an extensive *Jerne-like* network may not be strictly necessary for *self/non-self* discrimination in a systemic way.

3.1.4 Bias and specificity

As explained before, an increase in the bias parameter a corresponds to a progressive smoothing of the repertoire variability. A growth in a also results in a dilution of the graph itself, eventually leading to percolation phenomena (see Figure (3.11)).

In order to study this process, we compare its features to those pertaining to an ER random graph $\mathcal{G}_{\mathrm{ER}}(N, p)$, where links are drawn independently with probability $p = \bar{z}/N$, in such a way that the average degree is the same for both networks.

Therefore, the evolution of $\mathcal{G}(N, p)$ as p ranges from 0 to 1, eventually leading to a percolation transition (see, e.g., [41, 42]) is compared to the evolution of our graph $\mathcal{G}(a, L, \alpha)$ as a ranges from 0 to 1.

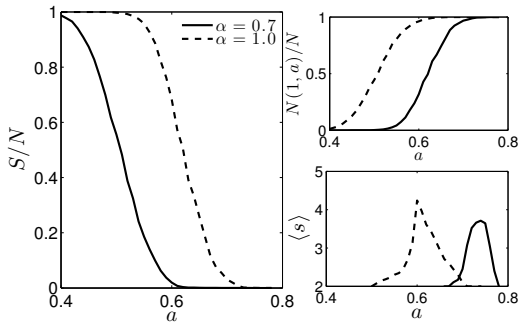


Figure 3.10. These plots describe the evolution of the network as a is tuned while we fix the system size $N = 10000$, $\gamma = 7$, and $\alpha = 0.7$ (solid line) and $\alpha = 1$ (dashed line). Left panel: Average size of the giant component S/N. Right panels: relative number of isolated nodes $N(1, a)/N$ (upper panel) and average size of the non-giants components excluding isolated nodes (lower panel).

Giant component and distribution of cluster size

In order to evaluate the impact of removing ties, we measure the relative size of the largest connected component S as a function of the fraction of links left f .

In general, as a ranges from 0 to 1, S gets continuously smaller: nodes with large ρ are those more likely to remain isolated or to form small clusters. Differently from the ER case, beyond the giant component, clusters typically display a small size. As evidenced in Ref. [34], these features give rise to a rather gentle percolation transition.

In order to clarify this point, we focus on the evolution of the internal organization of clusters by measuring the distribution $N(a, s)$, representing the number of clusters of size s present when the correlation parameter is a.

As shown in Figure (3.10), as a grows, the typical cluster size shrinks from N (a unique giant component) to 1 (there are only isolated nodes). For instance, at $\alpha = 0.7$, when $a = 0.46$ there are a few isolated nodes and a giant component whose typical (normalized) size is around 0.8. On the other hand, when $a = 0.60$ many nodes are isolated, $\sim 39\%$, and the typical size of the remaining (non-giant) clusters is around 4.2. Beyond isolated nodes and the giant component, the statistics of cluster size is rather uniform, suggesting that minor disconnected clusters display small sizes, i.e., $s < 10$.

Emergence of small components

The difference between our idiotypic network and a random graph is emphasized by the analysis of the emergence of motifs around the the percolation threshold. We study the number of isolated k-loops, namely, unicycles of length k, and isolated k-stars, namely, subgraphs made of a central node connected to k nodes with unitary

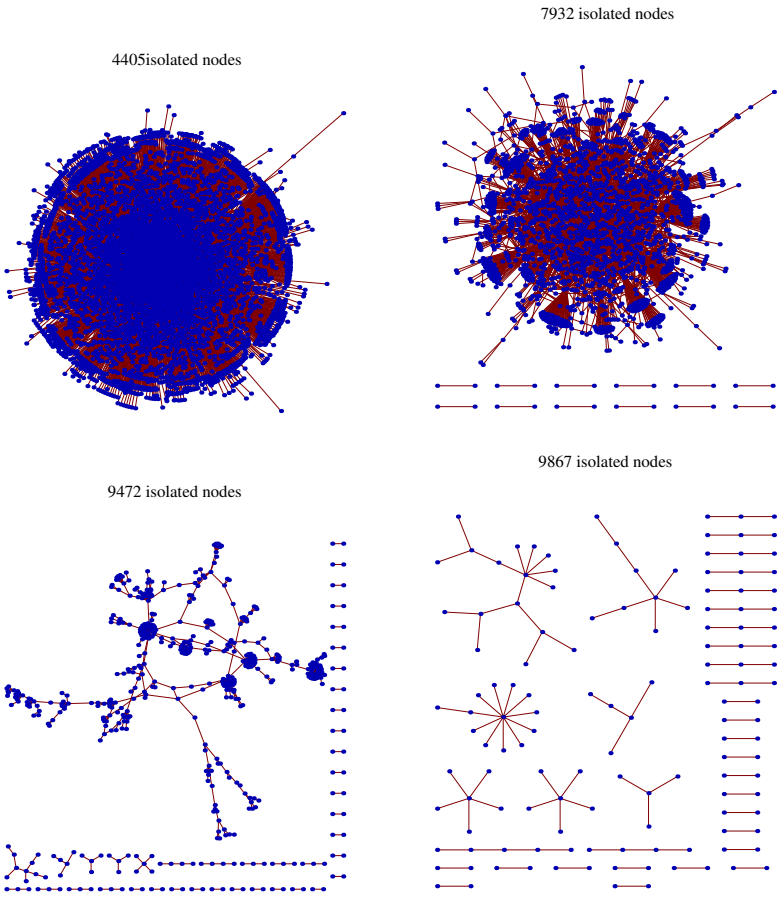


Figure 3.11. (Color on line) Realization of idiotypic networks made up of $N = 10000$ nodes with $\gamma = 3$ and $\alpha = 0.7$. Different values of a have been considered: from left to right $a = 0.50$, 0.55, 0.59, 0.62. Although these plots refer to one single realization, we have checked that the system displays robustness in this sense.

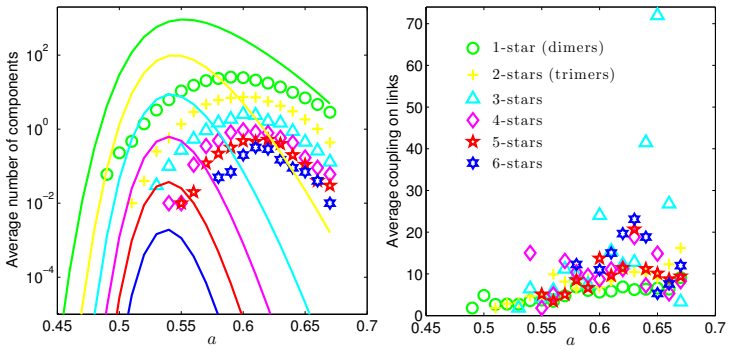


Figure 3.12. Left: Number of k-stars (a dimer is a 1-star and a trimer is a 2-star) averaged over 100 numerical realizations of N=10000 and α=0.7 idiotypic networks, while a (circles) is variated. Solid lines represent the ER prediction for the same quantity at the relative link probability. Right: Mean coupling value within a star.

degree. In an ER graph these quantities follow

$$\langle N_{k\text{-star}} \rangle = p^k (1-p)^{(k+1)N-3k-1} \frac{\prod_{i=0}^{k}(N-i+1)}{(k+1)!}, \tag{3.38}$$

$$\langle N_{k\text{-loop}} \rangle = p^k (1-p)^{k(N-k)} \frac{\prod_{i=0}^{k}(N-i+1)}{k!}, \tag{3.39}$$

which are drawn for comparison in Figure 3.12. In the idiotypic network the above-mentioned components are present in a smaller range of a, and the number of all stars, except the dimers (1-star), is eventually larger than ER predictions. This is due to the bias characterizing strings, as it makes the percolation transition smoother - in such a way that components not belonging to the giant component hardly develop - and induces inhomogeneity among nodes - in such a way that those associated with a value of ρ much lower than its average value are likely to act as (local) hubs. This picture is confirmed by the fact that k-stars with large k have a higher mean coupling (see Figure 3.12, right).

Moreover, in our simulations we never find isolated triangles (3-loop) or quadrangles (4-loop), which are instead present in random graphs.

3.1.5 Conclusions

In the last Sections we studied a class of weighted random networks aimed to describe the mutual interactions between cells inside the B-branch of the immune system. As we will see in the next Sections of this Chapter, the studied interaction network among B clones can be studied in a systemic perspective as a ferromagnetic system on the network isolated (as in Equation (3.2))or in interaction with the T cell clones (as we will se in Section 3.2.3). In this framework, the Burnet clonal expansion theory [43, 44] will appear to be the standard one-body response of the system, while the Jerne theory is ruled by the two-body term, whose coupling encodes the Varela prescription.

The main novelty introduced is the existence of a biased repertoire: The "mean field similarity" among the entries in the bit-strings encoding for antibodies ruled by the tuning of a scalar parameter, a, ranging from 0 (completely random antibodies) to 1 (completely deterministic antibodies). From a biological viewpoint, the extension to a value $a \neq 0$ is a step towards more realistic descriptions of the system, since antibodies are not completely random objects [45] and since during ontogenesis in bone marrow, a part of the strongly self-reacting repertoire is killed [46, 47]. Although more realistic assumptions would imply the use of an a_μ^i (with the price of a highly intractable mathematics), we skipped the fine dependence in favor of a simpler mean field choice $a_\mu^i = a \neq 0$ for all positions in the bit-strings coding for the idiotypicity. This introduction of bias indeed gives rise to striking, at least qualitatively, effects: In a broad range of $a \neq 0$, the degree distribution displays a multi-modular structure which mirrors the history of the system itself. Such fringes in the (weighted) degree distribution were experimentally noted in the pioneering investigations by Varela and coworkers and an immunological interpretation was provided in Refs. [48, 38, 39]. Although based on synthetic data, our analysis extends these findings in some details, suggesting the existence of a hyperfine structure, richer than the one obtained by sampling over the experimentally available subset of the repertoire. Furthermore, the reshuffling among modes of the network raises the question of which should be the physical observable to characterize the reaction attitude of nodes: the bare degree (which is actually the standard one checked in experiments) or, rather, the weighted connectivity which we claim to be the most relevant; Here, we just stress that, while the average degree varies over a linear scale, the weighted degree displays a distribution with a log-normal like envelope, hence spanning several order of magnitude. This spread is robust as it holds also also for regimes of high dilution (corresponding to large a): Interestingly, this implies that mechanisms such as self/non-self discrimination may work despite a truely over-percolated network of B-cells.

A progressive increase in a eventually leads to an under-percolated network, where nodes not belonging to the giant component typically form small-size clusters, which mirrors the cascades of anti-antibodies commonly seen in experiments (see e.g. [49]). Further, the local topology displays squares as motifs and triangles as non-motifs, as expected from a complementary-based network, which is in agreement with experiments on idiotypic networks too.

3.2 Neural networks inspired models for B and T lymphocites interaction network

In a series of recent papers ([27], [47], [50], [1], [51], [52], [53], [54], [40]) Agliari, Barra and coworkers proposed a *corpus* of statistical mechanics inspired mathematical models aimed to describe the adaptive response of the immune system. The intent of this theoretical effort is that of underlining how some general features of the immune system can be thought as emergent behaviors of a complex system of many interacting agents, which in this case are the different B-cell and T-cell clones.

In this Chapter we analyze the theoretical framework of this modeling of the immune system by reviewing the above cited works and we describe in more detail the analysis presented in [1] which finds its place in the general scheme.

The models that we are going to consider have to be thought as coarse grain descriptions of the real biological systems. In order to be mathematically treatable those models are build on rude simplifications and generalizations that could appear unrealistic to the pure biologist eye. Nevertheless the aim of those mathematical models is not that of making precise quantitative prediction on the behavior of the real system; on the contrary they should be considered as *toy models* whose scope is that of underlining how some general concepts of the mathematical modeling of the complex systems could play a role in understanding the immune system behavior.

3.2.1 The minimal model

In the following we describe a model that accounts for the interactions of the different B cells and T cells clones. The model has the aim to underline possible analogies of the immune system with the neural networks.

Let us consider a system made up by N_B different clones of B lymphocytes and N_H different clones of helper and regulatory T lymphocytes. As in Section 3.1Let us call b_μ a logarithmic measure for the concentration of cells of the μ-th clone with respect to the physiological concentration c_r of a clone in absence of antigenic stimulation:

$$b_\mu = \log(\frac{c_\mu}{c_r}) \, ; \quad \mu = 1, 2, \dots, N_B \, . \tag{3.40}$$

We will describe the helper/regulatory T clones as two state systems that are encoded binary variables

$$h_i = \pm 1 \, ; \quad i = 1, 2, \dots, N_H : \tag{3.41}$$

a clone is considered to have +1 (-1) value when it is in "activation" ("rest") state.

As we are interested in the statistical properties of system with a macroscopic number of degrees of freedom, eventually we will consider the thermodynamic limit in which $B, H \to \infty$ with a finite ratio:

$$\alpha = \lim_{H \to \infty} \frac{N_B}{N_H} \, . \tag{3.42}$$

The nature of the interaction between every cell clone is encoded in the parameters $\xi_i^\mu = \pm 1$ which are equal to 1 (-1) if the interaction between the μ-th B clone i-th and the T clone is of the type "helper" ("regulatory"), i.e. if the activation activation of the i-the T cell clone elicits (repress) the expansion of the interacting μ-th B

cell clone. As we are interested in the systemic properties of the system the simple hypothesis is to consider them as random variables with a fair distribution:

$$P(\xi_i^\mu = +1) = P(\xi_i^\mu = -1) = \frac{1}{2}\,, \quad \forall\, i, \mu\,. \tag{3.43}$$

To describe the time evolution of the concentration of a generic B cell clone given the parameters and the state of the T clones, we can use the Langevin equation:

$$\frac{db_\mu}{dt} = -b_\mu + \frac{1}{\sqrt{N_H}} \sum_{i=1}^{N_H} \xi_i^\mu h_i + \sum_k J_{\mu k} A_k + \sqrt{\frac{1}{\beta}}\eta_\mu\,, \tag{3.44}$$

where the first term accounts for the death of b cells (death rate=1), the second and the third term describe the expansion/repression signal coming from the T cells and the expansion signal due to the presence of antigens respectively, while the last term is a white noise:

$$\langle \eta(t) \rangle = 0\,, \quad \left\langle \eta(t)\eta(t') \right\rangle = \delta(t - t')\,, \tag{3.45}$$

and β (the inverse temperature) is a parameter related to the fluctuation of this noise term.

As long as the couplings ξ_i^μ are symmetric, the Langevin dynamics admits a Hamiltonian description

$$\frac{db_\mu}{dt} = -\frac{d}{db_\mu}\tilde{\mathcal{H}}(b, h|J, \xi) + \sqrt{\frac{1}{\beta}}\eta_\mu\,, \tag{3.46}$$

where, by integration over b_μ,

$$\tilde{\mathcal{H}}^{N_B,N_H}(b, h|J, \xi) = \sum_{\mu=1}^{N_B} b_\mu^2 - \frac{1}{\sqrt{H}} \sum_{i,\mu=1}^{N_B,N_H} \xi_i^\mu b_\mu h_i - \sum_{\mu=1}^{N_B} J_{\mu,k} b_\mu A_k \tag{3.47}$$

At the same time, as T clones activity is described by discrete variables their dynamics is chosen to be described by Glauber equation, so that given the state of the B clones, the T clones activity evolve following:

$$P(h_i \to -h_i) = \frac{1}{1 + \exp[2\beta\Delta\tilde{\mathcal{H}}^{B,H}]}\,. \tag{3.48}$$

Is a know result [55] that the equilibrium distribution of the Langevin and Glauber dynamics is the Boltzmann distribution:

$$P(b, h|\beta; \xi, J) = \frac{e^{-\beta\tilde{\mathcal{H}}^{B,H}}}{Z(\beta; \xi, J)} \tag{3.49}$$

where the normalization factor

$$\begin{aligned} Z(\beta; N_B, N_H|\boldsymbol{\xi}) &= \sum_{\{h\}} \int \prod_\mu db_\mu e^{-\beta\tilde{\mathcal{H}}^{N_B,N_H}} = \\ &= \sum_{\{h\}} \int \prod_\mu db_\mu e^{-\frac{\beta}{2}\sum_\mu^{N_B} b_\mu^2 + \frac{\beta}{\sqrt{H}}\sum_{\mu,i}^{N_B,N_H} \xi_i^\mu b_\mu h_i} \end{aligned} \tag{3.50}$$

is the partition function.

Following the principles of statistical mechanics, the thermodynamics of the system is recovered once the free energy is computed as

$$f(\beta;\epsilon) = \lim_{N_B,N_H\to\infty} -\frac{1}{N_H}\,\mathbb{E}\ln Z(\beta,N_B,N_H|\xi)\,, \tag{3.51}$$

where the symbol $\mathbb{E}$ indicates the average over the *quenched* variables ξ_i^μ. We note that as the free energy is the physical quantity that is aimed to compute, multiplicative constants on Z do not affect the result as in the thermodynamic limit they give vanishing contributions. For this reason, from now on, these constants will be neglected.

Let us consider first the case in which there is no antigenic stimulation on the B lymphocytes, i.e. let us equate to zero the term $\sum_k J_{\mu k}A_k$ in Equation (3.44). In the computation of the partition function a first step can be easily performed; that is the integration over b_μ as it just involves gaussian integrals. It gives

$$Z(\beta;N_B,N_H|\boldsymbol{\xi}) = \sum_{\{h\}} \exp\left(\frac{\beta}{2H}\sum_{i,\mu}^{N_B,N_H}\xi_i^\mu\xi_j^\mu h_i h_j\right)\,. \tag{3.52}$$

So the thermodynamics of the system is completely equivalent to that of a Hopfield model in the variables h_i with N_B patterns, namely to a system with a Hamiltonian of the form:

$$\mathcal{H}^H(h|\xi) = -\frac{1}{H}\sum_{ij}^{H}\left(\sum_\mu^B \xi_i^\mu\xi_j^\mu\right)h_i h_j - \sum_i \tilde{J}_{il}h_i A_l\,, \tag{3.53}$$

where the second term is a possible external field due to the presence of an antigen A_l.

Here the role of the load of the network is played by ϵ defined in (3.42), the ratio between the number of B clones and that of T clones. In complete analogy with the high storage Hopfiled model we can define the pattern overlap and the Edward-Anderson overlap as:

$$m_\mu \equiv \frac{1}{N_H}\xi_i^\mu\langle h_i\rangle\,, \tag{3.54}$$

$$q \equiv \frac{1}{N_H}\langle h_i\rangle^2\,. \tag{3.55}$$

In this framework the patterns stored in the system could be interpreted, at the immunological level, as the strategies that the T lymphocytes orchestrate to coordinate the B lymphocytes in response to the presence of an antigen which acts as an external perturbation on the system of interacting clones. In fact the retrieval of a stored pattern, let us say of the pattern $\mu = 1$, would give

$$m_1 = m > 0;\; m_{\mu>1} = 0\,. \tag{3.56}$$

That would give on average a non null net expansion term in the dynamical equation (3.44) only to the 1-th B clone This situation would correspond to an activation

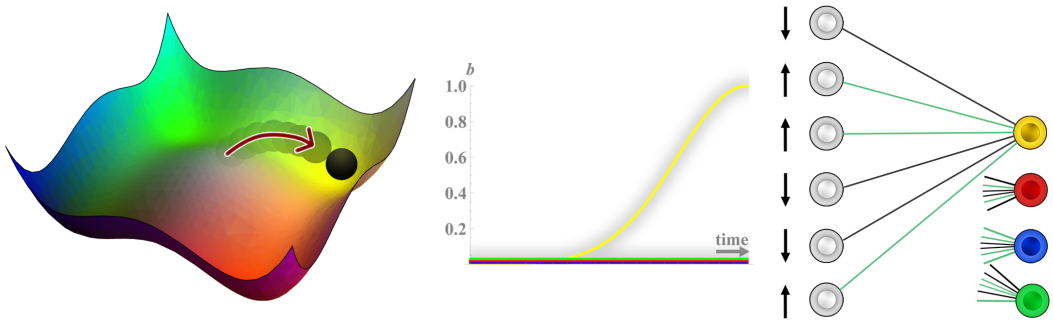


Figure 3.13. Schematic representation of the consequence of retrieval capabilities by the helper network in the bipartite network made up of both T and B clones: In the left panel a free-energy landscape of the helper network, with four minima (each corresponding to retrieval of instruction for a particular B clone) is shown. The black ball represents the state of the system, which is driven into the yellow hole (e.g. due to antigenic stimulation). Consequently, as all the helpers in the bipartite network (right panel) become parallel to the sign of the cytokines linking them to the yellow B clone (green (black) link are meant as positive (negative)). This results in maximal strength conferred on the retrieved clone, which undergoes clonal expansion. The latter is represented in the middle plot, together with the lack of growth by the other clones (not retrieved). Figure from [40].

configuration of the T clones that, on average, gives an expansion signal only to the 1-th B clone, while the others clones feel a null average net contribution by the T cells. A pictorical view of this immunological interpretation of the Hopfield model is provided in Figure 3.13.

As shown in Chapter 1 and depicted in Figure 1.4, the Hopfield model admits stable retrieval states only in a closed region of the parameter space enclosed in the critical line below the spin glass phase. This would be reflected, in the immunological interpretation of the Hopfield model, by the fact that the system would be able to recover a defense strategy only provided that the noise is not too high and that the ratio between the amount of B and T clones does not exceed a critical value.

3.2.2 The role of an extensive antigenic stimulus on the network capabilities

In [50] the authors studied the effect of an extensive antigenic stimulation on the the network model described in Section 3.2.1. We review here the principal results of this work.

Let us consider the effect of the antigenic term on the dynamical Equation (3.44) for the B lymphocytes: let us rewrite it as $\sum_k J_{\mu k} A_k = b_0^\mu$. This has the effect of changing the logarithmic equilibrium concentration of the μ-th B clone from 0 to b_0^μ. Let us consider the extreme case in which

$$b_0^\mu = b_0 \,, \quad \forall\, \mu \,. \tag{3.57}$$

This affects the marginalization with respect to the b_μ as now the partition function

in Equation (3.50) is replaced by

$$\begin{aligned} Z(\beta; N_B, N_H|\boldsymbol{\xi}) &= \sum_{\{h\}} \int \prod_{\mu} db_\mu e^{-\beta \tilde{\mathcal{H}}^{N_B,N_H}} = \\ &= \sum_{\{h\}} \int \prod_{\mu} db_\mu e^{-\frac{\beta}{2}\sum_{\mu}^{B}(b_\mu - b_0)^2 + \frac{\beta}{\sqrt{H}}\sum_{\mu,i}^{N_B,N_H} \xi_i^\mu b_\mu h_i} \end{aligned} \tag{3.58}$$

and integration over b_μ gives now an effective hamiltonian for the h party

$$\mathcal{N}_{\mathcal{H}}{}^{H}(h|\xi) = -\frac{1}{N_H}\sum_{ij}^{H}\left(\sum_{\mu}^{N_B}\xi_i^\mu \xi_j^\mu\right) h_i h_j - \sum_i \tilde{J}_{il} h_i A_l - \Phi \sum_i \eta_i h_i \,, \tag{3.59}$$

where we have called a rescaled activation parameter $\Phi = \sqrt{\alpha} b_0$ and

$$\eta \equiv \frac{1}{\sqrt{N_B}} \sum_{\mu} \xi_i^\mu \tag{3.60}$$

plays now the role of an external random field. In the thermodynamic limit the Central Limit Theorem of probabilities assures that the distribution of this field converges to a Normal distribution:

$$P(\eta) = \mathcal{N}(0,1) \,. \tag{3.61}$$

An extensive activation of the B clones, in this toy model framework, affects the system by generating a new source of noise which is represented by a random field in the Hopfield network of the T helper clones. The intuitive effect of the presence of this additional noise is that of making more difficult for the network to retrieve a strategy. This is confirmed by the analytic solution of the model performed in [50], to which we refer for details. The standard Hopfield model replica symmetric solution (1.88) is modified accordingly to

$$\begin{aligned} f(\epsilon, \beta, \Phi; m, q) &= -\frac{\log 2}{\beta} + \frac{\epsilon}{2}[1 + \frac{\beta\sqrt{q}(1-q)}{[1-\beta(1-q)]^2}] + \frac{1}{2}m^2 + \\ &+ \frac{\epsilon}{2\beta}\Big[\log[1-\beta(1-q)] - \frac{\beta q}{1-\beta(1-q)}\Big] \\ &- \frac{1}{\beta}\int d\mu(\eta)\int d\mu(z)\log\cosh\Big(\beta m + \frac{\beta\sqrt{\epsilon q}}{1-\beta(1-q)}z + \sqrt{\beta}\Phi\eta\Big), \end{aligned} \tag{3.62}$$

where $m_\mu = m(1,0,0,...,0)$ has been fixed.

Extremizing this replica symmetric free energy with respect to m, q, the self-consistent relations

$$\begin{aligned} m &= M(\epsilon, \beta, \gamma, \Phi; m, q) = \\ &= \int d\mu(\eta)\int d\mu(z)\tanh\Big(\beta m + \frac{\beta\sqrt{\epsilon q}}{1-\beta(1-q)}z + \sqrt{\beta}\Phi\eta\Big), \end{aligned} \tag{3.63}$$

$$\begin{aligned} q &= Q(\epsilon, \beta, \gamma, \Phi; m, q) = \\ &= \int d\mu(\eta)\int d\mu(z)\tanh^2\Big(\beta m + \frac{\beta\sqrt{\epsilon q}}{1-\beta(1-q)}z + \sqrt{\beta}\Phi\eta\Big) \end{aligned} \tag{3.64}$$

can be found.

The numerical solution of these equations permits to draw the plots in Figure 3.14 and 3.15. Form those plots it is possible to see clearly how the presence of the random field reduces the retrieval region in the (α, β) plane.

In [50] the authors suggest an immunological analogy with the this mathematical result. In fact it has been observed that an abnormally high concentration of lymphocytes gives rise to autoimmunity. In the model, when the system is out of the retrieval phase, its dynamics becomes very complex and the stored strategies are not recovered: the formal B clones get expanded in a random fashion and it is so possible that they attack the self.

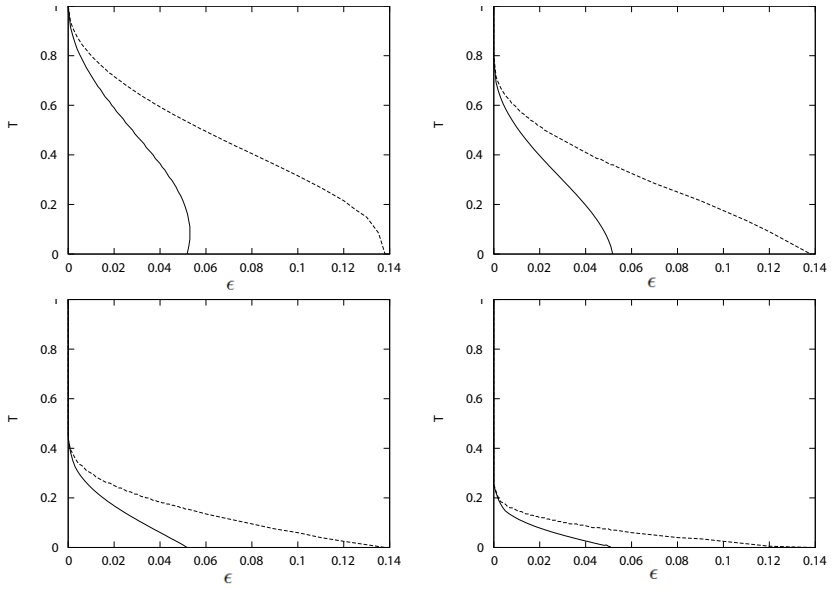


Figure 3.14. The effect of the random field on the Hopfield model phase diagram: the dashed line represents the critical line β_M, which distinguishes among retrieval (in general sense) and spin glass phases, while the continuous line represents the critical line β_C, which confines the pure state phase. Upper panel: $\Phi = 0$ (left) and $\Phi = 0.5$ (right); Lower panel: $\Phi = 1$ (left) and $\Phi = 1.5$ (right). Figures form [50].

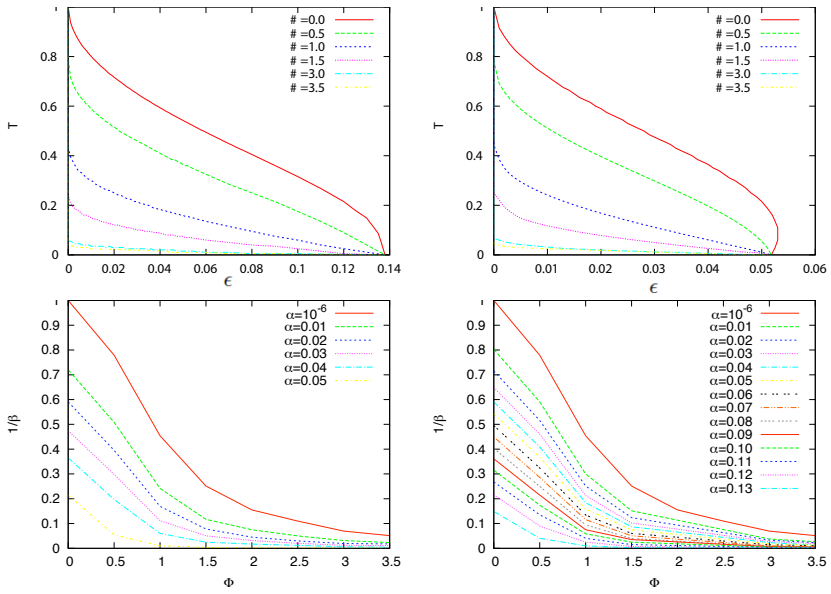


Figure 3.15. Left panels: Phase diagrams given by the critical surface β_M at various α (upper panels) and at various Φ (lower panel). Right panels: Phase diagrams given by the critical surface β_C at various α (upper panels) and at various Φ (lower panel). Figures from [50].

3.2.3 The role of the idiotypic interactions on the network capabilities

In the toy models for the B and T core of the adaptive response reviewed in Sections 3.2.1 and 3.2.2 the B clones do not interact among themselves. The theory of idiotypic network presented in Section 2.1.4, for which we have presented a descriptive model in Section 3.1, states that in fact interactions among B clones can exist. In this Section we review the work presented in [40] in which the minimal model presented in Section 3.2.1 is enriched by the presence of the idiotypic interactions. Following [40], the B-B interactions modify the equation for the time evolution of the logarithmic B clones (3.44) into

$$\frac{db_\mu}{dt} = -\sum_{\nu=1}^{N_B} J_{\mu\nu}(b_\nu - b_\mu) + \frac{1}{\sqrt{N_H}} \sum_{i=1}^{N_H} \xi_i^\mu h_i + \sum_k J_{\mu k} A_k + \sqrt{\frac{1}{\beta}}\eta_\mu \,, \tag{3.65}$$

where now the first term encodes for the idiotypic interactions and the $J_{\mu\nu} \geq 0$ follow the definitions and the structure presented in Section 3.1.

The Hamiltonian for the b and h variables eventually modifies as

$$\tilde{\mathcal{H}}^{N_B,N_H}(b,h|J,\xi) = \sum_{\mu,\nu}^{N_B} \frac{J_{\mu\nu}}{4}(b_\mu - b_\nu)^2 - \frac{1}{\sqrt{H}} \sum_{i,\mu=1}^{N_B,N_H} \xi_i^\mu b_\mu h_i - \sum_{\mu=1}^{N_B} J_{\mu,k} b_\mu A_k \,. \tag{3.66}$$

As long as the analysis is restricted to the case in which almost every B clone has the rest concentration ($b_\mu \simeq 0$) and just a few clones get expanded, a sort of "mean field" approximation, stating that the term $b_\mu b_\nu \simeq 0$, holds so that

$$\sum_{\mu,\nu}^{N_B} \frac{J_{\mu\nu}}{4}(b_\mu - b_\nu)^2 \simeq \frac{1}{2}\sum_{\mu}^{N_B} w_\mu b_\mu^2 \,, \tag{3.67}$$

where the weighted connectivity w_μ follows the definition (3.33) and, as explained in detail in Section 3.1.3, has a distribution with approximately log-normal envelope.

According to the last approximation the partition function of the system can be computed as

$$\begin{aligned} Z(\beta; N_B, N_H|\boldsymbol{\xi}) &= \sum_{\{h\}} \int \prod_\mu db_\mu e^{-\beta\tilde{\mathcal{H}}^{N_B,N_H}} = \\ &= \sum_{\{h\}} \int \prod_\mu db_\mu e^{-\frac{\beta}{2}\sum_\mu^{N_B} w_\mu b_\mu^2 + \frac{\beta}{\sqrt{H}}\sum_{\mu,i}^{N_B,N_H} \xi_i^\mu b_\mu h_i} \,. \end{aligned} \tag{3.68}$$

Gaussian integration over the b variables leads to

$$Z(\beta; N_B, N_H|\boldsymbol{\xi}) = \sum_{\{h\}} \exp\left(\frac{\beta}{2H} \sum_{i,\mu}^{N_B,N_H} \frac{\xi_i^\mu \xi_j^\mu}{w_\mu} h_i h_j\right) \tag{3.69}$$

which can be interpreted as the partition function of a modified Hopfield model with weighted patterns for the spin variables h having Hamiltonian

$$\mathcal{H}^H(h|\xi, w) = -\frac{1}{N_H}\sum_{ij}^{N_H}\left(\sum_\mu^{N_B} \frac{\xi_i^\mu \xi_j^\mu}{w_\mu}\right) h_i h_j - \sum_i \tilde{J}_{il} h_i A_l \,, \tag{3.70}$$

that represents the update of the simple Hopfield Hamiltonian (3.53) in presence of a non negligible idiotypic network. From last expression it is evident that, within the "mean filed" approximation, the net effect of the idiotypic interactions is that of removing the degeneracy of the attractor states of the neural network associated to the effective interactions among T clones: attractors associated to B clones with a high weighted connectivity are less pronounced than those that are related to the less interacting B clones in the idiotypic network. This result opens the possibility of a modification of the Varela interpretation of the idiotypic network effect: the repression of B clones with strong idiotypic interactions not only comes from the direct interaction with the neighbouring B clones, but, as a "second order" effect, also from the fact that they do not receive the elicitation signal from the T cells as the strategy that would maximally expand those highly interacting B clones are suppressed in the hs neural network.

The above discussion is enforced by the RS solution of this weighted Hopfield model. The computation details can be found in [40]. As a result the standard Hopfield model free energy (1.79) is modified as

$$\begin{aligned} f(m,q,r|\beta,\epsilon) \;=\; & \mathbb{E}_w \left\{ -\frac{1}{\beta}\log 2 + \frac{1}{2}\sum_{\nu=1}^{s}\frac{m_\nu^2}{w_\nu} \right. \\ & + \lim_{N_H\to\infty}\frac{1}{2\beta N_H}\sum_{\mu>s}^{N_B}\left[\ln\left(1-\frac{\beta}{w_\mu}(1-q)\right) - \frac{\frac{\beta q}{w_\mu}}{1-\frac{\beta}{w_\mu}(1-q)}\right] \\ & \left. + \frac{\epsilon\beta r}{2}(1-q) - \frac{1}{\beta}\,\mathbb{E}_\xi \int d\mu(z)\,\ln\cosh\left[\beta\Big(\sum_{\nu=1}^{s}\frac{m_\nu}{w_\nu}\xi^\nu + \sqrt{\epsilon r}z\Big)\right]\right\} \end{aligned} \tag{3.71}$$

which admits the following self consistency relations

$$m^\nu \;=\; \mathbb{E}_{\xi,w}\int d\mu(z)\tanh\beta\left[\sqrt{\epsilon r}z + \sum_{\rho}^{s}\frac{m^\rho\xi^\rho}{w_\rho}\right], \tag{3.72}$$

$$q \;=\; \mathbb{E}_{\xi,w}\int d\mu(z)\tanh^2\beta\left[\sqrt{\epsilon r}z + \sum_{\rho}^{s}\frac{m^\rho\xi^\rho}{w_\rho}\right], \tag{3.73}$$

$$r \;=\; \mathbb{E}_w \lim_{N_h\to\infty}\frac{1}{\epsilon N_H}\sum_{\mu>s}^{p}\frac{q}{[w_\mu-\beta(1-q)]^2}, \tag{3.74}$$

that are meant to be compared with the standard Hopfield ones, (1.80)-(1.82).

An approximate solutions for the above equations can be given in the pure state case by approximating Equation (1.82) as

$$r \simeq \lim_{N_h\to\infty}\frac{1}{\epsilon N_H}\sum_{\mu>s}^{p}\frac{q}{[\mathbb{E}_w w_\mu-\beta(1-q)]^2}\,. \tag{3.75}$$

Three extreme cases can be distinguished

1. the recalled clone has weighted connectivity much smaller than the average one; in this case we can approximate $\mathbb{E}_w w \to \infty$, so that r vanishes and the

self consistency equation for m simply reduces to

$$m = \tanh \beta \frac{m}{w} \,, \tag{3.76}$$

so to the low storage case, in which case the pattern is simply recalled at low temperature.

2. the recalled clone has weighted connectivity close to the average one; in this case we can approximate $\mathbb{E}_w w = \mathcal{O}(1)$; in this case the model reduced to the standard high storage Hopfield model.

3. the recalled clone has weighted connectivity much larger than the average one; in this case we can approximate $\mathbb{E}_w w \to 0$, so that the self consistency equations reduce to

$$m = \int d\mu(z) \tanh \beta \left(\frac{m}{w} + \frac{\sqrt{\epsilon q} z}{\beta(1-q)} \right) \,, \tag{3.77}$$

$$q = \int d\mu(z) \tanh^2 \beta \left(\frac{m}{w} + \frac{\sqrt{\epsilon q} z}{\beta(1-q)} \right) \,. \tag{3.78}$$

The effect of a large w is so that of increase the level of noise and eventually destroy the recalling of the strategy related to expansion of the considered B clone.

We stress that, as discussed within the framework of a descriptive model in Section 3.1, probably an overpercolated idiotypic network is not actually required for the mechanism considered here to works as the distribution of the weighted idiotypic connectivity (which is the parameter that enters in the effective T helper network) remains broad even for extremely diluted regimes (far beyond the percolation threshold).

3.2.4 The fully connected assumption

Several unrealistic assumptions have been made in building up the models depicted in Section 3.2 starting from the symmetry of the interactions to the fully connected structure of the interaction network in which every couple of B and T clones are considered to be in interaction. The models have to be considered as very coarse grained description of the biological reality that have the aim of underline the importance of some systemic mechanism.

Nevertheless some of these assumptions can be relaxed, as for example the fully connected topology. The possible effects of the dilution of the network has been investigated in [51] and [52] in the low storage regime ($N_B \sim \log N_H$), while medium ($N_B \sim {N_H}^{\delta}$, $\delta \in [0,1)$) and high ($N_B \sim N_H$) level of load are still under investigation and preliminary results can be found respectively in [53] and [54].

The description of the results of the above cited work is behind the scopes of this Thesis. Nevertheless we outline that a result of the dilution of the interaction between B and T clones, i.e. of the entries of the ξ vectors, is that, in some regimes, the system is allowed to retrieve more than one strategy in parallel. We refer to the articles for details.

Chapter 4

The stochastic stability methods for neural networks

In this Chapter and in [2] we analyze a Hopfield model [56, 33], where pattern entries can be either extracted from a Gaussian distribution or set equal to zero. More precisely, entries are drawn from a normal distribution $\mathcal{N}[0,1]$ with a probability $(1+a)/2$ or set equal to zero with probability $(1-a)/2$, where $a \in [-1,+1]$ is a tunable parameter controlling the degree of dilution of patterns. We focus on the high-storage limit, namely the amount of patterns L is linearly diverging with the system size N, i.e. $L = \alpha N$.

This kind of "analogue" neural networks has been intensively studied on fully connected topologies (see for instance [57, 58, 59, 60]) and further interest in the model lies in its peculiar "soft retrieval" as explained for instance in [61].

Here, we first study the topological properties of the emergent weighted network, then we pass to the thermodynamic properties of the model.

In particular, we calculate analytically the average probability for two arbitrary nodes to be connected and we show that, by properly tuning a, the network spans several topological regimes, from fully connected down to the percolation threshold. Moreover, even if the network is very sparse, it turns out to display a large degree of cliquishness due to the Hebbian rule underlying its couplings. The coupling distribution is also explicitly calculated and shown to be central and with extensive variance, as expected.

From a thermodynamic perspective, using an exact Gaussian mapping, we prove that this model is equivalent to a bipartite diluted spin-glass, whose parties are made up by binary Ising spins and by Gaussian spins, respectively, while interactions among them, if present, are drawn from a standard Gaussian distribution $\mathcal{N}[0,1]$; of course, there are no links within each party. The size of the two parties are respectively N, for the Ising spins, and L, for the Gaussian ones, and the dilution in the Hopfield pattern entries corresponds to standard link removal in this bipartite counterpart.

Then, extending the technique of multiple stochastic stability (developed for fully connected Hopfield model in [60] and for ferromagnetic systems on small-world graphs in [28]) to this case, we solve its thermodynamics at the replica symmetric level. Once introduced suitably order parameters for this theory, we obtain an

explicit expression for the free-energy density that we extremize with respect to them to obtain the self-consistencies that constraint the phase space of the model. As in other works on diluted networks [28, 62, 63], the order parameters are two (one for each party) series of overlaps defined on all the possible subgraphs through which the network can be decomposed. A study of their rescaled and centered fluctuations allows to obtain the critical surface delimiting the ergodic phase from the spin-glass one. The same result is also recovered through small overlap expansion from the self-consistencies; the agreement confirms the existence of a second order phase transition [20, 64].

The Chapter is organized as follows: in Section 4.1 the model is defined with all its related parameters and variables, while in Section 4.2 its topological properties are discussed. Section 4.3 deals with the statistical mechanics analysis while in Section 4.4 fluctuation theory is developed. Section 4.5 is left for a discussion and outlooks.

4.1 The model

Given N Ising spins $\sigma_i = \pm 1$, $i \in (1, ..., N)$, we aim to study a mean-field model whose Hamiltonian has the form

$$\tilde{H} = -\frac{1}{D} \sum_{ij}^{N} J_{ij} \sigma_i \sigma_j \,, \tag{4.1}$$

where the couplings are built in a Hebbian fashion [65][56] as

$$J_{ij} = \sum_{\mu=1}^{L} \xi_i^\mu \xi_j^\mu, \tag{4.2}$$

and D is a denominator whose specific form is discussed in Section 3. In fact, in general, as the coordination number may vary sensibly according to the definition of patterns ξ, in order to ensure a proper linear scaling of the Hamiltonian (4.1) with the volume, D has to be a function of the system size N and of the parameters through which patterns ξ are defined.

We consider the high-storage regime [13], such that, in the thermodynamic limit (i.e. $N \to \infty$), the following scaling for the amount of stored memories (patterns) is assumed

$$\lim_{N\to\infty} \frac{L}{N} = \alpha \in \mathbb{R}^+, \tag{4.3}$$

even though we use the symbol α for the ratio between the number of patterns and the system size also at finite N, bearing in mind that the thermodynamic limit has to be performed eventually.

The quenched entries of the memories ξ_i^μ are Gaussian and diluted, namely they are set to zero with probability $(1-a)/2$, while, with probability $(1+a)/2$, they are drawn from a standard Gaussian distribution:

$$P(\xi_i^\mu) = \left(\frac{1-a}{2}\right) \delta(\xi_i^\mu) \, + \, \left(\frac{1+a}{2}\right) \mathcal{N}_{[0,1]}(\xi_i^\mu) \,. \tag{4.4}$$

The parameter a can in principle be varied in the range $a \in [-1, 1]$, and, in general, small values correspond to highly diluted regimes. As proved in Section 4.2.2, a scaling law for this parameter has to be introduced in order to avoid the topology to become trivial in the thermodynamic limit. Thus, we consider the following scaling

$$a = -1 + \frac{\gamma}{N^\theta} , \tag{4.5}$$

where θ determines the topological regime of the network, while γ plays the role of a fine tuning within it. More precisely, $\gamma \in (0, 2]$ and, of course, for $\gamma = 0$ we get $P(\xi_i^\mu) = \delta(\xi_i^\mu)$, that is, there is no network, so we discuss only the case $\gamma > 0$. Finally, notice that fixing $\theta = 0$ and $\gamma = 2$ yields to $a = 1$, corresponding to the standard analogue Hopfield model [60].

4.2 Topological analysis

4.2.1 Coupling distribution

Let us consider the definition of the coupling strength in Equation (4.2): the probability p that the μ-th term $\xi_i^\mu \xi_j^\mu$ is zero corresponds to the probability that at least one between ξ_i^μ and ξ_j^μ is zero, which is

$$p \equiv \left(\frac{1-a}{2}\right)^2 + 2\left(\frac{1-a}{2}\right)\left(\frac{1+a}{2}\right) = \frac{3-a^2-2a}{4} = 1 - \left(\frac{1+a}{2}\right)^2 , \tag{4.6}$$

while its complement is the probability that a Gaussian number is drawn for both entries, that is $(1-p) = [(1+a)/2]^2$. Thus, the probability that the link connecting i and j has strength J_{ij} can be written as

$$\begin{aligned} P(J_{ij}) &= p^L \delta(J_{ij}) + \sum_{k=1}^{L} p^{L-k}(1-p)^k \binom{L}{k} \; P_k\big(\sum_{\nu}^{k} \xi_i^\nu \xi_j^\nu = J_{ij}\big) = \qquad (4.7) \\ &= p^L \delta(J_{ij}) + \sum_{k=1}^{L} f(k) \; P_k\big(\sum_{\nu}^{k} \xi_i^\nu \xi_j^\nu = J_{ij}\big) = \\ &= p^L \delta(J_{ij}) + \sum_{k=1}^{L} f(k) \; \int_{-\infty}^{+\infty} \frac{dl}{2\pi} \frac{e^{-ilJ_{ij}}}{(1+l^2)^{k/2}} , \qquad (4.8) \end{aligned}$$

where to simplify the notation we defined $f(k) = p^{L-k}(1-p)^k \binom{L}{k}$ and $P_k\big(\sum_\nu^k \xi_i^\nu \xi_j^\nu = J_{ij}\big)$ is the probability that k pairs of Gaussian entries, pairwise multiplied, sum up to J_{ij}, namely

$$\begin{aligned} P_k\Big(\sum_{\nu}^{k} \xi_i^\nu \xi_j^\nu = J_{ij}\Big) &= \int_{-\infty}^{\infty} \prod_{\nu=0}^{k} d\xi_i^\nu d\xi_j^\nu \; P(\xi_i^\nu) P(\xi_j^\nu) \, \delta\Big(\sum_{\nu=0}^{k} \xi_i^\nu \xi_j^\nu - J_{ij}\Big) = \\ = \int_{-\infty}^{+\infty} \frac{dl}{2\pi} \prod_{\nu=0}^{k} d\xi_i^\nu d\xi_j^\nu \, \frac{e^{-\frac{(\xi_i^\nu)^2}{2}}}{\sqrt{2\pi}} \frac{e^{-\frac{(\xi_i^\nu)^2}{2}}}{\sqrt{2\pi}} \, e^{il\left(\xi_i^\nu \xi_j^\nu - J_{ij}\right)} &= \int_{-\infty}^{+\infty} \frac{dl}{2\pi} \frac{e^{-ilJ_{ij}}}{(1+l^2)^{k/2}} . \qquad (4.9) \end{aligned}$$

From Equation (4.8) one can easily specify the characteristic function of the coupling distribution

$$F(l) \equiv \int_{-\infty}^{+\infty} e^{ilJ} P(J) dJ = p^L + \sum_{k=1}^{L} \frac{f(k)}{(1+l^2)^{k/2}} = \frac{1}{2} p^L \left[1 + \left(\frac{1+l^2 p}{p+l^2 p} \right)^L \right] , \quad (4.10)$$

where we dropped the indices i and j, due to the arbitrariness of the couple of nodes considered. From $F(l)$ it is possible to obtain all the momenta by simple differentiation. For instance, first and second moment read respectively as

$$\mathbb{E}[J_{ij}] = (-i) \frac{\partial F(l)}{\partial l} \bigg|_{l=0} = 0, \quad (4.11)$$

$$\mathbb{E}[J_{ij}^2] = (-i)^2 \frac{\partial^2 F(l)}{\partial l^2} \bigg|_{l=0} = L(1-p) = L \left(\frac{1+a}{2} \right)^2 = \frac{\alpha \gamma^2}{4} N^{1-2\theta} . \quad (4.12)$$

Now, for fixed a and α, we expect that J, being a sum of Gaussian variables, is also normally distributed (except the point $J = 0$), at least for large N. Indeed, numerical simulations confirm that the distribution $P(J)$ converges in the thermodynamic limit ($L \to \infty$) to a Gaussian distribution with zero mean and variance given by Equation (4.12) (see Figure 4.1), except for the point $J = 0$ which will be discussed in the following section.

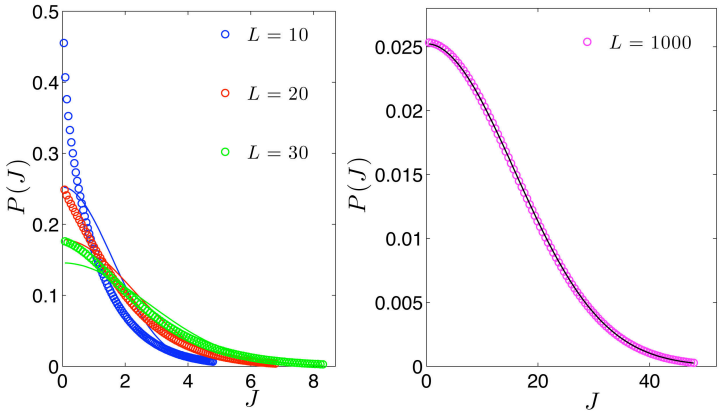


Figure 4.1. Coupling distribution $P(J)$ for $L = 10$ (left panel, blue), $L = 20$ (left panel, red), $L = 30$ (left panel, green) and $L = 1000$ (right panel). Circles represent the coupling distribution calculated according to Eq. 4.7, while the continuous lines represent normal distributions with momenta given by Eq. 4.11 and Eq. 4.12: as the thermodynamic limit is approached, the agreement gets better and better. Notice that, when L grows, the divergence in $J = 0$ becomes weaker. Only the positive values of J are considered due to the symmetry.

4.2.2 Link Probability and topology regimes

Let us consider the bare topology. The quantity of interest is the average link probability P_{link}:

$$P_{\text{link}} = 1 - P(J = 0) . \quad (4.13)$$

Looking at Equation (4.7), in principle $P(J=0)$ has two contributions: one from the delta function and one from the sum over k random numbers, but the latter has a null measure in the limit $L \to \infty$. To show this we consider the second term in Equation (4.7) and we calculate its measure over the interval $J_{ij} \in [-\epsilon, +\epsilon]$, highlighting for clarity the term $k=1$:

$$\int_{-\epsilon}^{\epsilon} dJ_{ij} \sum_{k=1}^{L} f(k)\ P_k(\sum_{\nu}^{k} \xi_i^\nu \xi_j^\nu - J_{ij}) =$$
$$= p^{L-1}\ (1-p) L \int_{-\epsilon}^{\epsilon} P_1(r) dr\ +\ \sum_{k=2}^{L} \binom{L}{k}\ p^{L-k}\ (1-p)^k\ \int_{-\epsilon}^{\epsilon} P_k(r) dr\,. \quad (4.14)$$

In fact, we notice that $P_1(r)$ has a weak divergence in $r=0$ and its integral scale as $\sim \epsilon \log(\epsilon)$, so that the divergence is suppressed by the prefactor in the limit $L \to \infty$, so that the first term in Eq. 4.14 is vanishing. As for $P_{k>1}(r)$, its integral is non-diverging and can be upper bounded [1] to show that the second term is also negligible in the limit $L \to \infty$.

Hence, in the thermodynamic limit, $J_{ij}=0$ only if $\xi_i^\mu \xi_j^\mu = 0$, for any μ, namely

$$P(J_{ij}=0) = p^L = \left(\frac{3-a^2-2a}{4}\right)^L . \quad (4.15)$$

Now, looking at (4.13) and (4.15) in the thermodynamic limit, it is clear that, if we consider a as finite and constant, only two trivial topologies can be realized. In fact, if $a=-1$, P_{link} is zero and the system is fully disconnected, while, with $a > -1$, P_{link} tends to one exponentially fast with the system size, and the graph becomes fully connected.

Nevertheless, with the scaling (4.5),

$$P_{\text{link}} = 1 - \left(1 - \frac{\gamma^2}{4N^{2\theta}}\right)^{\alpha N} \simeq 1 - e^{-\frac{\alpha\gamma^2}{4N^{2\theta-1}}}, \quad (4.16)$$

where the last expression holds for large N and $\gamma \in (0,2]$. Now, by tuning the value of θ, we realize different topological regimes; within each regime the parameter γ acts as a fine tuning. Following a mean-field approach, namely just focusing on the average link probability, we can distinguish:

- $\underline{\theta=0}$: $P_{\text{link}} = 1 - \left(1 - \frac{\gamma^2}{4}\right)^{\alpha N} \to 1$. Fully Connected graph, with average degree equal to the system size ($\bar{z} = N-1$).
  The coupling distribution converges to the Gaussian one with variance $\text{Var}[J] \propto N$, as in the with Sherrington-Kirkpatrick model.

[1] From the the two inequalities $\int_{-\epsilon}^{\epsilon} P_k(r)dr < 2\epsilon P_k(0)$, $\quad P_k(0) < \text{c}\frac{\log(k)}{k}$ (with c a constant), it follows

$$\sum_{k=2}^{L} \binom{L}{k}\ p^{L-k}(1-p)^k\ \int_{-\epsilon}^{\epsilon} P_k(r)dr < 2\epsilon\text{c}, \sum_{k=2}^{L} \binom{L}{k}\ p^{L-k}\ (1-p)^k \frac{\log(k)}{k}$$

which goes to zero in the limit $L \to \infty$.

- $0 < \theta < 1/2$: $P_{\text{link}} = 1 - e^{-\frac{\alpha\gamma^2}{4}N^k} \to 1$ (where $0 < k < 1$). Fully Connected graph, with average degree equal to the system size ($\bar{z} = N - 1$).
  The coupling distribution converges to the Gaussian one with $\text{Var}[J] \propto N^k$.

- $\theta = 1/2$: $P_{\text{link}} \simeq \frac{\alpha\gamma^2}{4} = \text{const}$. The link probability is finite and the average coordination number is linearly diverging with the system size, namely $\bar{z} = \alpha\gamma^2(N-1) = \mathcal{O}(N)$, and $\text{Var}[J] \propto \text{cost}$.

- $1/2 < \theta < 1$: $P_{\text{link}} = 1 - e^{-\frac{\alpha\gamma^2}{4}N^k} \simeq \frac{\alpha\gamma^2 N^k}{4} \to 0$ (where $-1 < k < 0$). Extreme Diluted Graph, characterized by a sublinearly diverging average coordination number, $\bar{z} = \mathcal{O}(N^{1-k})$, and $\text{Var}[J] \propto N^k$.

- $\theta = 1$: $P_{\text{link}} = 1 - e^{-\frac{\alpha\gamma^2}{4N}} \simeq \frac{\alpha\gamma^2}{4N} \to 0$. Finite Coordination Regime with $\bar{z} = \alpha\gamma^2/4$, and $\text{Var}[J] \propto 1/N$.

- $\theta > 1$: $P_{\text{link}} = 1 - e^{-\frac{\alpha\gamma^2}{4}N^k} \simeq \frac{\alpha\gamma^2 N^k}{4} \to 0$ (where $k < -1$). Fully Disconnected Regime with coordination number vanishing for any choice of α and γ. The variance of the coupling distribution is vanishing superlinearly with N.

A contour plot of P_{link} as a function of γ and θ is shown in Figure 4.2.

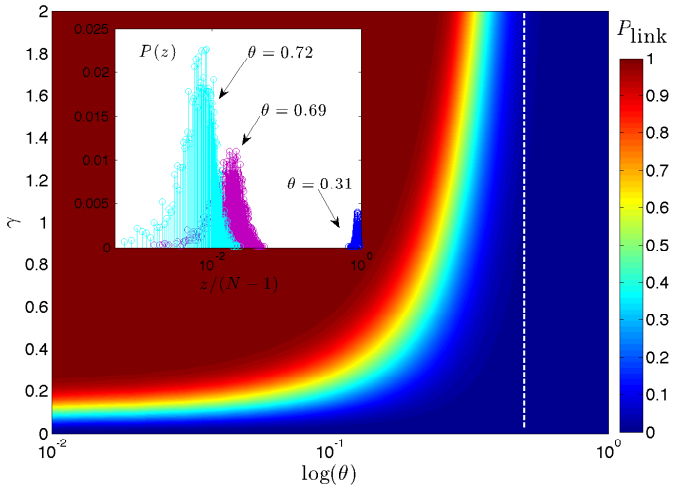


Figure 4.2. The main figure represents the contour plot of P_{link} (see Equation 4.16) as a function of γ and of $\log\theta$. The dashed, vertical line corresponds to $\theta = 1/2$ and demarcates the onset of a disconnected regime. The inset represents the degree distribution $P(z)$ as a function of the normalized number of nearest neighbors; three values of θ are considered as specified. Notice that, as expected, larger values of θ yields to sparser graphs. Both figures refer to systems made up of $N = 6000$ nodes, with $\alpha = 0.05$ and $\gamma = 1$.

4.2.3 Small-world properties

Small-world networks are characterized by two main properties: a small diameter and a large clustering coefficient, namely, the average shortest path length scales

logarithmical (or even slower) with the system size and they contain more cliques than what expected by random chance [66]. The small-world property has been observed in a variety of real networks, including biological and technological ones [67].

First, we checked that, in the over-percolated regime, the structures considered here display a diameter growing logarithmically with N, as typical for random networks [68].

As for the clustering coefficient C, it is basically defined as the likelihood that two neighbors of a node are linked themselves, that is, for the i-th node,

$$c_i = \frac{2E_i}{z_i(z_i-1)}\,, \tag{4.17}$$

where z_i is the number of nearest neighbors of i and E_i is the number of links connecting any couple of neighbors; when E_i equals its upper bound $z_i(z_i-1)/2$, the neighborhood of i is fully connected. The global clustering coefficient then reads as

$$C = \frac{1}{N}\sum_{i}^{N} c_i. \tag{4.18}$$

A clustering coefficient close to 1 means that the graph displays a high "cliquishness", while a value close to 0 means that there are few triangles.

It is easy to see that for the Erdős-Rényi graph, where each link is independently drawn with a probability P, the average clustering coefficient is $C_{\rm ER} = P$. Therefore, for our network, we measure C and we compare it with the average link probability $P_{\rm link}$; results obtained for different choices of θ are shown in Figure 4.3.

First, we notice that, for a given system size N, the behavior of C and of $C_{\rm ER}$, with respect to θ, is markedly different (see the inset): the latter decreases monotonically due to the analogous decrease of the link probability, while the former exhibits two extremal points at a relatively large degrees of dilution. In fact, as long as the networks are highly connected, the disappearance of a few links yields, in both cases, a modest drop in the overall cliquishness. On the other hand, when dilution is significant, the intrinsic structure of the "Hebbian graph" matters: as patterns get sparser and sparser, surviving links are those connecting nodes whose related patterns display matching with non-null entries. In this way, the neighbors of a node are also likely to be connected [37, 28, 34] and the clustering coefficient grows. Finally, at a very large degree of dilution, the system approaches the fully-disconnected regime and the clustering coefficient decreases.

In order to compare more effectively our graph and an analogous ER graph, we also considered the ratio $C/C_{\rm ER}$ (see the main figure). Interestingly, for θ relatively large, as N gets larger this ratio grows confirming that the few links remaining are very effective in maintaining the cliques. This can be understood as follows: to fix ideas let us take $\theta = 1$, so that the average number of non-null entries in a string is $L\gamma/(2N)$ which equals $\gamma\alpha/2$ in the high-storage regime under investigation. For simplicity, let also assume that $\gamma\alpha/2 \approx 1$, and that this holds with vanishing variance for all nodes. Therefore, if the node i has k neighbors, its (local) clustering coefficient is either 0 (if $k \leq 2$) or 1 (if $k > 2$). Hence, the expected local clustering coefficient can be estimated as the probability for a node to display $k > 2$ nearest neighbors,

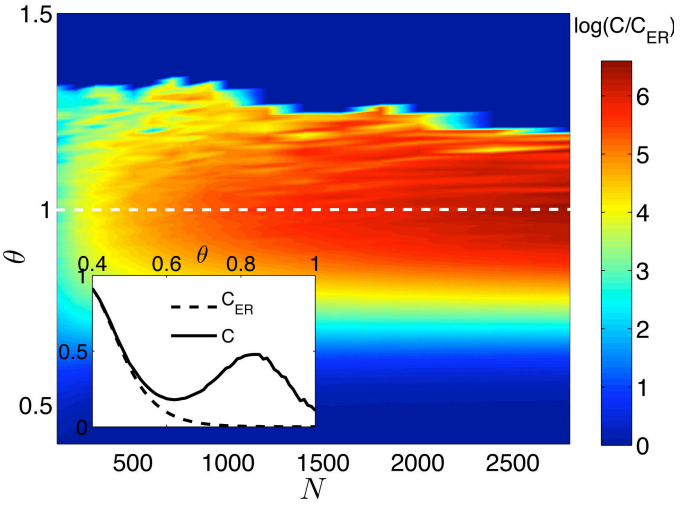


Figure 4.3. Inset: Mean clustering coefficient C (continuous line) for different choices of the parameter θ, while $N = 1600$, $\gamma = 2$ and $\alpha = 0.5$ are kept fixed; the mean has been performed over all nodes making up the graph and over 10^2 realizations. An analogous ER random graph is also considered and the related clustering coefficient C_{ER} (dashed line) is shown for comparison. Notice the qualitative different behaviors of C and C_{ER}. Main figure: contour plot for the logarithm of the ratio C/C_{ER}, as a function of N and θ. Notice that, although for θ close to 1 both graphs are sparse, $C \gg C_{\mathrm{ER}}$. On the other hand, for $\theta > 1$, both coefficient converge to zero, in the thermodynamic limit; the non-null values appearing in the figure are due to finite-size effects.

namely $1 - (1-q)^{N-1} - Nq(1-q)^{N-2}$, where $q = \gamma/(2N)$ is the probability that the pattern of an arbitrary node $j \neq i$ has the non-null entry matching with the one of ξ_i. With some algebra we get $c_i \approx 1 - e^{-\gamma/2}(1 + \gamma/2)$, which remains finite also in the thermodynamic limit, in agreement with results from simulations. For $\theta > 1$, $q = \gamma N^{1-\theta}/2$ and c_i converges to zero.

4.3 The statistical mechanics analysis

In this section we study the thermodynamic properties of the system introduced: At first we show its equivalence to a bipartite spin-glass and figure out the order parameters of the theory, then we define an interpolating free-energy which generalizes the multiple stochastic stability developed in [60]; this technique allows to obtain the replica-symmetric solution in form of a simple sum rule. As a last step, we extremize the free energy finding self-consistencies for the order parameters, whose critical behavior is also addressed.

4.3.1 The equivalent diluted bipartite spin-glass

As we deal with a structure whose average coordination number may range in $[0, N]$, from a statistical mechanics perspective, we aim to define the normalization constant D for the Hamiltonian in Equation (4.1), in such a way that its average (which defines the extensive energy of the system and is denoted symbolically with

the brackets) is linearly diverging with the system size, namely $\langle \tilde{H} \rangle \propto N$.

By a direct calculation, it is possible to show that this condition is fulfilled by

$$D = N^{1-\theta}, \tag{4.19}$$

so that, using the explicit definition for the couplings, we can write

$$\tilde{H} = -\frac{1}{N^{1-\theta}} \sum_{i<j}^{N} \sum_{\mu}^{L} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j = -\frac{1}{N^{1-\theta}} \sum_{i<j}^{N} J_{ij} \sigma_i \sigma_j \,, \tag{4.20}$$

For a single realization of the disorder encoded in the memories, the partition function reads off as:

$$\tilde{Z}_{N,L}(\beta;\xi) = \sum_{\{\sigma\}} \exp\left\{ \frac{\beta}{2N^{1-\theta}} \sum_{i,j}^{N} \sum_{\mu}^{L} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j \right\} . \tag{4.21}$$

Note that, as usual in the Hopfield model, the diagonal term gives an extensive contribution to the partition function. In the above expression we neglected this diagonal term, directly by adding it as a term $\frac{\alpha\beta}{2}(\frac{1+a}{2}) = \frac{\alpha\beta\gamma}{4N^\theta}$ to the final expression of the free energy [59] (see Equation (4.46)).

Now, we can introduce another party made up of L *soft* spins $\{z_\mu\}$, namely i.i.d. variables with an intrinsic standard Gaussian distribution $\mathcal{N}[0,1]$, that interact only with the original party of binary spins $\{\sigma_i\}$ via the couplings $\{\xi_i^\mu\}$; the related partition function is

$$Z_{N,L}(\beta;\xi) = \sum_{\{\sigma\}} \int \prod_{\mu} d\mu(z_\mu) \exp\left\{ \sqrt{\frac{\beta}{N^{1-\theta}}} \sum_{i}^{N} \sum_{\mu}^{L} \xi_i^\mu \sigma_i z_\mu \right\} , \tag{4.22}$$

with $d\mu(z_\mu)$ standard Gaussian measure for all the z_μ. By applying Gaussian integrations as usual [28], it is easy to see that $\tilde{Z}_{N,L}(\beta;\xi)$ and $Z_{N,L}(\beta;\xi)$ are thermodynamically equivalent. The advantage of the expression (4.22) is that it is linear with respect to the memories ξ_i^μ, so that the bare topology is simply that of a bipartite random graph with link probability $p_{\text{link}} = (1+a)/2$, like in [37].

Taken O as a generic observable, depending on the spin configurations $\{\sigma, z\}$, we define the Boltzmann state $\omega_\beta(O)$ at a given value of (fast) noise β as

$$\omega_\beta(O) = Z_{N,L}(\beta;\xi)^{-1} \sum_{\{\sigma\}} \int \prod_{\mu} d\mu(z_\mu) O(\{\sigma, z\}) e^{\sqrt{\frac{\beta}{N^{1-\theta}}} \sum_i^N \sum_\mu^L \xi_i^\mu \sigma_i z_\mu}, \tag{4.23}$$

and we introduce a product space on several replicas of the system as $\Omega_s = \omega_1 \bigotimes \omega_2 \bigotimes ... \bigotimes \omega_s$ [60].

For a generic function of the memories $F(\xi)$, the quenched average will be defined by the symbol $\mathbb{E}$ and performed in two steps: first we fix the number l of links between the two parties and we perform the average over the Gaussian distribution of the memories:

$$\mathbb{E}_\xi^{(l)}[F(\xi)] \equiv \int_{-\infty}^{+\infty} \prod_{(i,\mu)=1}^{l} \frac{d\xi_i^\mu}{\sqrt{2\pi}} e^{-\frac{(\xi_i^\mu)^2}{2}} F(\xi_i^\mu) = \int F(\xi)\, d\mu_l(\xi) \equiv f(l)\,; \tag{4.24}$$

then, we perform the average over the binomial distribution for the number of links,

$$\mathbb{E}_l[f(l)] \equiv \sum_{l=0}^{NL} \binom{NL}{l} \left(\frac{1+a}{2}\right)^l \left(\frac{1-a}{2}\right)^{(NL-l)} f(l)\,, \tag{4.25}$$

so that $\mathbb{E} \equiv \mathbb{E}_l \mathbb{E}_\xi^{(l)}$. Indeed, for example, $\mathbb{E}[\xi_i^\mu] = 0$ and $\mathbb{E}[(\xi_i^\mu)^2] = (1+a)/2$.

Moreover, as we will see, for a natural introduction of the order parameters, it is useful to define the number of links, l, as the product of two independent binomial variables

$$l \doteq l_\eta l_\chi\,, \tag{4.26}$$

where the symbol $\doteq$ stands for the equality in distribution and

$$P(l_\eta) = \binom{N}{l_\eta} \sqrt{\frac{1+a}{2}}^{l_\eta} \sqrt{\frac{1-a}{2}}^{N-l_\eta}\,, \tag{4.27}$$

$$P(l_\chi) = \binom{L}{l_\chi} \sqrt{\frac{1+a}{2}}^{l_\chi} \sqrt{\frac{1-a}{2}}^{L-l_\chi}\,. \tag{4.28}$$

Of course, a product of two binomial variables is not a binomial variable itself, so at finite size this definition is not consistent; nevertheless, in the thermodynamic limit, the central limit theorem ensures that only the first two momenta of the distributions survive so that the definitions become consistent.

We also use the symbol $\langle\cdot\rangle$ to mean $\langle\cdot\rangle = \mathbb{E}\Omega(\cdot)$ and $\langle\cdot\rangle_G = \mathbb{E}_\xi^{(l)}\Omega(\cdot)$.

The main thermodynamical quantity of interest is the intensive pressure defined as

$$A(\alpha,\beta) = \lim_{N\to\infty} A_N(\alpha,\beta) = -\beta f(\alpha,\beta) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\log Z_{N,L}(\beta,\xi)\,. \tag{4.29}$$

Here $f(\alpha,\beta) = u(\alpha,\beta) - \beta^{-1}s(\alpha,\beta)$ is the free-energy density, $u(\alpha,\beta)$ the internal energy density and $s(\alpha,\beta)$ the entropy density.

Finally, we define two infinite (in the thermodynamic limit) sets of order parameters, the restricted overlaps, as

$$q_{12}^{l_\eta} = \frac{1}{l_\eta}\sum_i^{l_\eta} \sigma_i^1\sigma_i^2,$$

$$p_{12}^{l_\chi} = \frac{1}{l_\chi}\sum_\mu^{l_\chi} z_\mu^1 z_\mu^2\,, \tag{4.30}$$

which define the overlaps (restricted on sub-networks) between two replicas made up by parties with l_η and l_χ nodes, respectively.

4.3.2 Free energy interpolation and general strategy

In what follows we assume that no real external fields (as magnetic inputs or partial information submission for retrieval) act on the network, but fields insisting on each spin are strictly generated by other spins. Thus, the overall field felt by

an element of a given party is the sum (weighted through the couplings), of the states of the spins in the other party. Note that spins are connected in loops using the other party as a mirror, therefore, the equivalent analogue neural network is a recurrent network.

In this section we show that the free-energy can be calculated in specific cases (e.g. at the replica symmetrical level) by using a novel technique that has been developed in [60] for fully connected spin-glass models and extended in [37] to diluted ferromagnetic models. This technique introduces an external field acting on the system which "imitates" the internal, recurrently-generated input, by reproducing its average statistics. While the external, fictitious input does not reproduce the statistics of order two and higher, it represents correctly the averages. These external inputs are denoted as η and χ (one for each spin in each party) and are distributed following the Gaussian distributions with zero mean and whose variances scale according to the underlying topology (as a function of α, θ, γ) and coherently approaches zero when the network topology disappears.

In order to recover the second order statistics, the free-energy is interpolated smoothly between the case in which all fields are external, and all high order statistics is missing, and the case in which all fields are internal, describing the original network: Following the original Guerra's schemes [15, 60, 69, 70], this allows a powerful sum rule. We use an interpolating parameter $t \in [0, 1]$ for this morphing, such that for $t = 0$ the fields are all external and the calculation straightforward, while for $t = 1$ the original model is fully recovered.

In what follows, for the sake of clearness, we write $A = A_N(\alpha, \beta)$ even though α should be introduced only once the thermodynamic limit has been performed. The interpolating quenched pressure $\tilde{A}_N(\alpha, \beta, t)$ at finite N is then defined as

$$\tilde{A}_N(\alpha,\beta;t) = \frac{1}{N}\mathbb{E}\log\sum_{\{\sigma\}}\int\prod_{\mu}d\mu(z_\mu)\exp\left(\sqrt{t}\sqrt{\frac{\beta}{N^{1-\theta}}}\sum_{i,\mu}^{N,L}\xi_i^\mu\sigma_i z_\mu\right)\cdot$$
$$\cdot\exp\left(b\sqrt{1-t}\sum_{i}^{N}\sigma_i\eta_i\right)\exp\left(c\sqrt{1-t}\sum_{\mu}^{L}z_\mu\chi_\mu\right)\exp\left(\frac{d}{2}(1-t)\sum_{\mu}^{L}z_\mu^2\theta_\mu\right). \quad (4.31)$$

Throughout the paper, we assume that the limit

$$A(\alpha,\beta) = \lim_{N\to\infty}A_N(\alpha,\beta) = \lim_{N\to\infty}A_N(\alpha,\beta,t=1) \quad (4.32)$$

exists. The "interpolating fields" distributions are chosen to mimic the local fields behavior, so that η_i, χ_μ and θ_μ have zero value with probability $\sqrt{(1-a)/2}$, while, with probability $\sqrt{(1+a)/2}$, are normally distributed, except for θ_μ which assumes value 1[2]. Consequently, the number of active fields follows Equation (4.27). As for the constants b, c, d, they have to be chosen properly, as shown in the following.

The strategy for the evaluation of the pressure of the original model, $\tilde{A}(\alpha, \beta, t = 1)$, is to compute the t-streaming of $\tilde{A}(\alpha, \beta, t)$, namely $\partial_t\tilde{A}(\alpha, \beta, t = 1)$, and use the fundamental theorem of calculus to obtain

$$A_N(\alpha,\beta) = \tilde{A}_N(\alpha,\beta;t=1) = \tilde{A}_N(\alpha,\beta,t=0) + \int_0^1 dt'\left(\frac{d}{dt}\tilde{A}_N(\alpha,\beta;t)\right)_{t=t'}. \quad (4.33)$$

[2]Indeed, the presence the field θ_μ has much less physical meaning but simplifies the calculations.

When evaluating the streaming $\partial_t \tilde{A}$, we get the sum of four terms $(\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D})$; each comes as a consequence of the derivation of a corresponding exponential term appearing into Equation (4.31). In order to proceed we need to compute them explicitly:

$$\begin{aligned}
\mathcal{A} &= \frac{1}{N}\mathbb{E}\frac{\sqrt{\beta}}{2\sqrt{t}N^{(1-\theta)/2}}\sum_{i,\mu}^{l_\eta,l_\chi}\xi_i^\mu\omega(\sigma_i,z_\mu) = \frac{1}{N}\mathbb{E}\frac{\sqrt{\beta}}{2\sqrt{t}N^{(1-\theta)/2}}\sum_{i,\mu}^{l_\eta,l_\chi}\partial_{\xi_i^\mu}\omega(\sigma_i,z_\mu)\\
&= \frac{1}{N}\mathbb{E}\frac{\beta}{2N^{(1-\theta)}}\sum_{i,\mu}^{l_\eta,l_\chi}\left[\omega(\sigma_i^2,z_\mu^2)-\omega^2(\sigma_i,z_\mu)\right] =\\
&= \frac{1}{N}\frac{\beta}{2N^{(1-\theta)}}\sum_{l_\eta,l_\chi}P(l_\eta,l_\chi)l_\eta l_\chi\left[\langle z_\mu^2\rangle_G-\langle q_{12}^{l_\eta}p_{12}^{l_\chi}\rangle_G\right] =\\
&= \frac{1}{N}\frac{\beta}{2N^{(1-\theta)}}NL\left(\frac{1+a}{2}\right)\left[\langle z_\mu^2\rangle-\langle q_{12}^{l_\eta}p_{12}^{l_\chi}\rangle\right] = \frac{\alpha\beta}{2}\frac{\gamma}{2}\left[\langle z_\mu^2\rangle-\langle q_{12}^{l_\eta}p_{12}^{l_\chi}\rangle\right], \qquad (4.34)
\end{aligned}$$

where in the first passage we used integration by parts and, in the fourth, the factorization properties of the quenched averages [37, 71, 72, 73, 74] (which should be understood in the thermodynamic limit).

The same procedure can be used in the computation of the other terms, so to get:

$$\begin{aligned}
\mathcal{B} &= -\frac{1}{N}\mathbb{E}\frac{b}{2\sqrt{1-t}}\sum_{i}^{l_\eta}\eta_i\omega(\sigma_i) = -\frac{1}{N}\mathbb{E}\frac{b}{2\sqrt{1-t}}\sum_{i}^{l_\eta}\partial_{\eta_i}\omega(\sigma_i)\\
&= -\frac{1}{N}\mathbb{E}\frac{b^2}{2}\left(l_\eta-\sum_i^{l_\eta}\omega^2(\sigma_i)\right) = -\frac{b^2}{2N}\sum_{l_\eta}P(l_\eta)\,l_\eta\left(1-\langle q_{12}^{l_\eta}\rangle_G\right)\\
&= -\frac{b^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\left(1-\langle q_{12}^{l_\eta}\rangle\right)-\frac{b^2}{2}\sqrt{\frac{\gamma}{2}}N^{-\frac{\theta}{2}}\left(1-\langle q_{12}^{l_\eta}\rangle\right); \qquad (4.35)
\end{aligned}$$

$$\begin{aligned}
\mathcal{C} &= -\frac{1}{N}\mathbb{E}\frac{c}{2\sqrt{1-t}}\sum_{\mu}^{l_\chi}\chi_\mu\omega(z_\mu) = -\frac{1}{N}\mathbb{E}\frac{c}{2\sqrt{1-t}}\sum_{\mu}^{l_\chi}\partial_{\chi_\mu}\omega(z_\mu) =\\
&= -\frac{1}{N}\mathbb{E}\frac{c^2}{2}\left(\sum_\mu^{l_\chi}\omega(z_\mu^2)-\sum_\mu^{l_\chi}\omega^2(z_\mu)\right) = -\frac{c^2}{2N}\sum_{l_\chi}P(l_\chi)\,l_\chi\left(\sum_\mu^{l_\chi}\langle z_\mu^2\rangle_G-\langle p_{12}^{l_\chi}\rangle_G\right)\\
&= -\frac{\alpha c^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\left(\langle z^2\rangle-\langle p_{12}^{l_\chi}\rangle\right) = -\frac{\alpha c^2}{2}\sqrt{\frac{\gamma}{2}}N^{-\frac{\theta}{2}}\left(\langle z^2\rangle-\langle p_{12}^{l_\chi}\rangle\right); \qquad (4.36)
\end{aligned}$$

$$\mathcal{D} = -\frac{1}{N}\mathbb{E}\frac{d}{2}\sum_\mu^{l_\chi}\omega(z_\mu^2) = -\frac{\alpha d}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\langle z^2\rangle\,. \qquad (4.37)$$

Now, the t-streaming of the pressure reads off as

$$\begin{aligned}\frac{d\tilde{A}_N(\alpha,\beta,t)}{dt} &= \left[\frac{\alpha\beta}{2}N^\theta\left(\frac{1+a}{2}\right)-\frac{\alpha c^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}-\frac{\alpha d}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\right]\langle z^2\rangle + \\ &-\frac{\alpha\beta}{2}N^\theta\left(\frac{1+a}{2}\right)\langle q_{12}^{l_\eta}p_{12}^{l_\chi}\rangle+\frac{b^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\langle q_{12}^{l_\eta}\rangle+\frac{c^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\langle p_{12}^{l_\chi}\rangle + \\ &-\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\frac{b^2}{2}\,. \end{aligned} \tag{4.38}$$

4.3.3 Replica symmetric approximation and fluctuation source

As it is, this streaming encodes the whole full replica-symmetry-breaking complexity [64, 75] of the underlying glassy phase and it is intractable. Our plan is to split this derivative in two terms, one dealing with the averages of the order parameters and one accounting for their fluctuations. To this aim we introduce the *source of fluctuations*, $S(\alpha,\beta,t)$, as

$$S(\alpha,\beta,t)=-\frac{\alpha\beta}{2}\left(\frac{1+a}{2}\right)N^\theta\langle(q_{12}^{l_\eta}-\overline{q})(p_{12}^{l_\chi}-\overline{p})\rangle, \tag{4.39}$$

with

$$\overline{q}\equiv\mathbb{E}_{l_\eta}\overline{q}^{l_\eta},\quad \overline{p}\equiv\mathbb{E}_{l_\eta}\overline{p}^{l_\eta}\,. \tag{4.40}$$

Notice that the main order parameters $\overline{q}$ and $\overline{p}$ sum every overlap, each with its relative weight, on every possible subnetwork of the whole network according to the approaches [62, 63] and that they recover the standard order parameters of the Hopfield model when dilution is neglected [13, 60].

In order to relate Equation (4.39) to Equation (4.38), let us remember that we still have free parameters that can be chosen as [3]

$$\begin{aligned} b &= \sqrt{\alpha\beta\left(\frac{1+a}{2}\right)^{\frac{1}{2}}N^\theta\overline{p}}=\sqrt{\alpha\beta\overline{p}}\left(\frac{\gamma}{2}\right)^{\frac{1}{4}}N^{\frac{\theta}{4}},\\ c &= \sqrt{\beta\left(\frac{1+a}{2}\right)^{\frac{1}{2}}N^\theta\overline{q}}=\sqrt{\beta\overline{q}}\left(\frac{\gamma}{2}\right)^{\frac{1}{4}}N^{\frac{\theta}{4}},\\ d &= \beta N^\theta\left(\frac{1+a}{2}\right)^{\frac{1}{2}}(1-\overline{q})=\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{\frac{\theta}{2}}(1-\overline{q})\,, \end{aligned} \tag{4.41}$$

so to get

$$\frac{d\tilde{A}(\alpha,\beta;t)}{dt}=S(\alpha,\beta,t)+\frac{\alpha\beta}{2}\left(\frac{1+a}{2}\right)N^\theta\,\overline{p}(\overline{q}-1)=S(\alpha,\beta,t)+\frac{\alpha\beta\gamma}{4}\overline{p}(\overline{q}-1). \tag{4.42}$$

In the replica symmetric approximation, the order parameters do not fluctuate with respect to their quenched average in the thermodynamic limit as they get delta-distributed over their replica symmetric averages $\overline{q},\ \overline{p}$, which have been denoted with

[3] In particular, we choose d to cancel the $\langle z^2\rangle$ terms appearing in the first line of equation (4.38).

a bar. As a consequence, within this approximation, we can neglect the fluctuation source term $S(\alpha,\beta,t)$ and keep only the replica symmetric overlap averages in the expression (4.42) such that its integration is trivially reduced to a multiplication by one.

In order to obtain an explicit expression of the sum rule (4.33), we can then proceed to analyze the starting point for the "morphing", namely $\tilde{A}(\alpha,\beta;t=0)$, which can be calculated straightforwardly as it involves only one-body interactions:

$$\begin{aligned}
&\tilde{A}_N(\alpha,\beta,t=0) = \\
&= \frac{1}{N}\mathbb{E}\log\sum_{\{\sigma\}}\int\prod_{\mu}d\mu(z_\mu)\exp\left(\ b\sum_{i}^{N}\sigma_i\eta_i + c\sum_{\mu}^{L}z_\mu\chi_\mu + \frac{d}{2}\sum_{\mu}^{L}\theta_\mu z_\mu^2\right) = \\
&= \sqrt{\frac{\gamma}{2}}N^{-\frac{\theta}{2}}\log 2 + \sqrt{\frac{\gamma}{2}}N^{-\frac{\theta}{2}}\int d\mu(\eta)\log\cosh(\sqrt{\alpha\beta\overline{p}}\left(\frac{\gamma}{2}\right)^{\frac{1}{4}}N^{\frac{\theta}{4}}\ \eta) + \qquad (4.43)\\
&+\frac{\gamma\alpha\beta\bar{q}}{4}\frac{1}{1-\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{\frac{\theta}{2}}(1-\overline{q})} - \frac{\alpha}{2}\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{-\frac{\theta}{2}}\log\left(1-\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{\frac{\theta}{2}}(1-\overline{q})\right),
\end{aligned}$$

where we used

$$\begin{aligned}
&\frac{1}{N}\mathbb{E}_{l_\eta}\mathbb{E}_{\eta}^{l_\eta}\log\sum_{\{\sigma\}}\exp\left(b\sum_{i}^{N}\sigma_i\eta_i\right) = \\
&= \frac{1}{N}\mathbb{E}_{l_\eta}\mathbb{E}_{\eta}^{l_\eta}\sum_{i}^{N}\log 2\cosh(b\ \eta_i) = \frac{1}{N}\mathbb{E}_{l_\eta}l_\eta\ \mathbb{E}_\eta\log 2\cosh(b\ \eta) = \\
&= \left(\frac{1+a}{2}\right)^{\frac{1}{2}}\log 2 + \left(\frac{1+a}{2}\right)^{\frac{1}{2}}\int d\mu(\eta)\log\cosh(b\ \eta), \qquad (4.44)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{N}\mathbb{E}_{l_\chi}\mathbb{E}_{\chi}^{l_\chi}\log\int\prod_{\mu}d\mu(z_\mu)\exp\left(c\sum_{\mu}^{L}z_\mu\chi_\mu\right)\exp\left(\frac{d}{2}\sum_{\mu}^{L}\theta_\mu z_\mu^2\right) = \\
&= \frac{1}{N}\mathbb{E}_{l_\chi}\mathbb{E}_{\chi}^{l_\chi}\left(\sum_{\mu}\frac{c^2\chi_\mu^2}{2(1-d\theta_\mu)}\right) + \frac{1}{N}\mathbb{E}_{l_\chi}\mathbb{E}_{\chi}^{l_\chi}\frac{1}{2}\sum_{\mu}\log\left(\frac{1}{1-d\theta_\mu}\right) = \\
&= \frac{\alpha}{2}\sqrt{\left(\frac{1+a}{2}\right)}\left(\frac{c^2\sigma_\chi^2}{(1-d)}\right) - \frac{\alpha}{2}\sqrt{\left(\frac{1+a}{2}\right)}\log(1-d) = \\
&= \frac{\gamma\alpha\beta\bar{q}}{4}\frac{1}{1-\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{\frac{\theta}{2}}(1-\overline{q})} - \frac{\alpha}{2}\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{-\frac{\theta}{2}}\log\left(1-\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}}N^{\frac{\theta}{2}}(1-\overline{q})\right) \qquad (4.45)
\end{aligned}$$

Now, substituting the expression for $\tilde{A}_N(\alpha,\beta,t=0)$ of Equation(4.43) into (4.33), we obtain the replica-symmetric free energy (strictly speaking the mathematical

pressure) of the network as

$$\tilde{A}_N^{RS}(\alpha,\beta) = \tilde{A}_N(\alpha,\beta,t=0) + \left.\frac{d\tilde{A}_N^{RS}(\alpha,\beta,t}{dt}\right|_{t=0} =$$
$$= \sqrt{\frac{\gamma}{2}} N^{-\frac{\theta}{2}} \left[\log 2 + \int d\mu(\eta) \log\cosh\left(\sqrt{\alpha\beta\overline{p}}\left(\frac{\gamma}{2}\right)^{\frac{1}{4}} N^{\frac{\theta}{4}}\ \eta\right)\right] +$$
$$+\frac{\alpha\beta\gamma}{4}\,\bar{q}\,\frac{1}{1-\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}} N^{\frac{\theta}{2}}(1-\overline{q})} - \frac{\alpha}{2}\left(\frac{\gamma}{2}\right)^{\frac{1}{2}} N^{-\frac{\theta}{2}} \log\left(1-\beta\left(\frac{\gamma}{2}\right)^{\frac{1}{2}} N^{\frac{\theta}{2}}(1-\overline{q})\right) +$$
$$+\frac{\alpha\beta\gamma}{4}\overline{p}(\overline{q}-1) + \frac{\alpha\beta\gamma}{4} N^{-\theta}\,. \tag{4.46}$$

Despite the last expression is meant to hold in the thermodynamic limit, with a little mathematical abuse we left the explicit dependence on N to discuss some features of the solution: Equation (4.46) may look strange due to the strong presence of various powers of the volume size N, which in principle are potentially unwanted divergencies. We start noticing that, in the limit of zero dilution $\theta = 0$ and homogeneous distribution of fields $\gamma = 2$, the expression for the free-energy recovers the replica symmetric one of the analogue Hopfield model [60] (or digital one without retrieval [20]). Moreover, remembering the various topological regimes outlined in Section 3, we see that when the network changes the topological phase, for instance moving from a fully connected topology to a sparse graph, the coordination number may scale with the volume size or remain constant. These situations are deeply different from a thermodynamical viewpoint because, in order to have no negligible contributions to the free-energy, fields obtained by an extensive number of (finite) terms in the fully connected scenario must be (possibly) turned into fields obtained by a finite number of (infinite) terms in the dilute regime. As the topology changes, the fields must follow accordingly, which is equivalent to a (fast) noise rescaling with the volume size that is another standard approach to diluted network [76, 72, 77].

The physical free-energy is then obtained by extremizing this expression with respect to the order parameters; we only stress here that, as a general property of these neural networks/bipartite spin-glasses, the free-energy now obeys a min-max principle, which will not be deepened here (because it does not change the following procedure and it has been discussed in [60]). As a consequence, the following system defines the values of the overlaps (as functions of $\alpha,\ \beta$) that must be used into Equation (4.46)

$$\begin{aligned}
\frac{\partial\tilde{A}}{\partial\overline{q}} &= \frac{\alpha\beta(\frac{\gamma}{2})}{2}\left(\overline{p} - \frac{(\frac{\gamma}{2})^{\frac{1}{2}} N^{\frac{\theta}{2}}\beta\overline{q}}{(1-\beta(\frac{\gamma}{2})^{\frac{1}{2}} N^{\frac{\theta}{2}}(1-\overline{q}))^2}\right) = 0 \\
\frac{\partial\tilde{A}}{\partial\overline{p}} &= \frac{\alpha\beta(\frac{\gamma}{2})}{2}\left[\overline{q} - \int d\mu(\eta)\tanh^2(\sqrt{\alpha\beta\overline{p}}\left(\frac{\gamma}{2}\right)^{\frac{1}{4}} N^{\frac{\theta}{4}}\eta)\right] = 0\,,
\end{aligned} \tag{4.47}$$

by which

$$\overline{q} = \int d\mu(\eta)\tanh^2\left(\frac{\sqrt{\alpha\overline{q}}\beta(\frac{\gamma}{2})^{\frac{1}{2}} N^{\frac{\theta}{2}}}{1-\beta(\frac{\gamma}{2})^{\frac{1}{2}} N^{\frac{\theta}{2}}(1-\overline{q})}\,\eta\right)\,. \tag{4.48}$$

All the related models (e.g. Viana-Bray [78], Hopfield [56], Sherrington-Kirkpatrick [14]) display an ergodicity breaking associated with a second order phase transition

and presence of criticality. If we assume the same behavior even for the model investigated here, the self-consistency equation (4.48) can give hints on the critical line (in the parameter space) where ergodicity breaks down. In fact, when leaving the ergodic region (implicitly defined by $\overline{q} = 0,\ \overline{p} = 0$) the order parameters start growing (implicitly defining the critical line as the starting point) and, as continuity is assumed through the second order kind of transition, we can expand the r. h. s. of Equation (4.48) for low $\overline{q}$ and obtain a polynomial expression on both sides. Then, due to the principle of identity of polynomials, we can equate the two sides term by term obtaining

$$\beta_c = \frac{1}{\left(\frac{\gamma}{2}\right)^{\frac{1}{2}} N^{\frac{\theta}{2}} \left(1 + \sqrt{\alpha}\right)}, \tag{4.49}$$

which is the critical surface of the system.

Mirroring the discussion dealing with the free energy, we note that this result too is clearly a consequence of the choice (4.19) for the normalization factor that gives us an extensive thermodynamics. If we normalize choosing $D = N$, as it is usual in the Hopfield model [20], we obtain (turning to $T = 1/\beta$ which is most intuitive) $T_c = N^{\theta/2}(1 + \sqrt{\alpha})\sqrt{\gamma/2}$ (and recover the AGS line for $\theta = 0$ and $\gamma = 2$), such that the overall effect of increasing dilution is to reduce the value of the critical temperature because the couplings, on average, become weaker. In particular, in the finite connectivity regime ($\theta = 1$), the network is built of by N links instead of N^2 which, roughly speaking, implies a rescaling in the temperature proportional to $\sqrt{N}$ (coherently with a spin-glass behavior), as in the ferromagnetic counterpart its rescale is ruled by N instead of $\sqrt{N}$ [76] because the latter is a model defined through the first momentum, while the former by the variance.

Furthermore we stress that the system displays only one critical surface splitting the ergodic region from the spin glass and there are no further 'weak-transitions" for each sub-overlap, coherently with the scenario discussed in [79] for the similar case of the Viana-Bray model [78].

4.4 Fluctuation theory and critical behavior

The plan of this section is studying the regularity of the rescaled (and centered) overlap correlation functions.

The idea is as follows: If the system undergoes a second order phase transition, the (extensive) fluctuations of its order parameters should diverge on the critical surface (4.49), hence they should be described by meromorphic functions; from the poles of these functions it is possible to detect the critical surface. As a consequence, an explicit knowledge of these functions would confirm (or reject) the critical picture we obtained through the small overlap expansion of the previous section. However, obtaining them explicitly is not immediate and we sketch in what follows our strategy. At first, we define the (rescaled and centered) fluctuations of the order parameters as

$$\begin{aligned} Q_{ab}^{l_\eta} &= \sqrt{N}\left(q_{ab}^{l_\eta} - \overline{q}^{l_\eta}\right), \\ P_{ab}^{l_\eta} &= \sqrt{L}\left(p_{ab}^{l_\chi} - \overline{p}^{l_\eta}\right), \end{aligned} \tag{4.50}$$

such that, while $q_{ab}^{l_\eta} \in [-1,+1]$, $p_{ab}^{l_\eta} \in [-1,+1]$, $Q_{ab}^{l_\eta} \in \mathbb{R}$, $P_{ab}^{l_\eta} \in \mathbb{R}$, hence, the square of the latter may diverge as expected for second order phase transitions.

Nevertheless, obtaining them explicitly from the original Hamiltonian is prohibitive and we use another procedure, originally outlined in [70]: We evaluate these rescaled overlap fluctuations weighted with the non-interacting Hamiltonian in the Maxwell-Boltzman distribution, hence $\langle Q^2_{l_\eta 12}\rangle_{t=0}$, $\langle Q_{l_\eta 12} P_{l_\chi 12}\rangle_{t=0}$, $\langle P^2_{l_\chi 12}\rangle_{t=0}$, then we derive the streaming of a generic observable O (that is in principle a function of the spins of the parties and of the quenched memories), namely $\partial_t \langle O\rangle_t$ such that we know how to propagate $\langle O\rangle_{(t=0)}$ up to $\langle O\rangle_{(t=1)}$ (which is our goal), and finally we use this streaming equation (which turns out to be a dynamical system) on the Cauchy problem defined by $\langle Q^2_{l_\eta 12}\rangle_{t=0}$, $\langle Q_{l_\eta 12} P_{l_\chi 12}\rangle_{t=0}$, $\langle P^2_{l_\chi 12}\rangle_{t=0}$, obtaining the attended result. Once the procedure is completed, the simple analysis of the poles of $\langle Q^2_{l_\eta 12}\rangle_{t=1}$, $\langle Q_{l_\eta 12} P_{l_\chi 12}\rangle_{t=1}$, $\langle P^2_{l_\chi 12}\rangle_{t=1}$ will identify the critical surfaces of the system.

Starting with the study of the structure of the derivative, our aim is to compute the t-streaming for a generic observable O_s of s replicas. Calling

$$
\begin{aligned}
H_s \;=\; & \sum_{a=1}^{s} \Bigg\{ \sqrt{t}\sqrt{\frac{\beta}{N^{1-\theta}}} \sum_{i,\mu}^{l_\eta, l_\chi} \xi_i^\mu \sigma_i^a z_\mu^a + \sqrt{1-t}\; b \sum_{i}^{l_\eta} \eta_i \sigma_i^a + \sqrt{1-t}\; c \sum_{\mu}^{l_\chi} \chi_\mu z_\mu^a + \\
& + (1-t)\frac{d}{2}\theta_\mu \sum_{\mu}^{l_\chi} (z_\mu^a)^2 \Bigg\}\,,
\end{aligned}
\tag{4.51}
$$

such that

$$
\langle O\rangle_t = \frac{\int \prod_\mu^L d\mu(z_\mu) \sum_\sigma O \exp\left(-\beta H_s\right)}{\int \prod_\mu^L d\mu(z_\mu) \sum_\sigma \exp\left(-\beta H_s\right)},
$$

its t-streaming is

$$
\frac{d\langle O_s\rangle_t}{dt} = \frac{d}{dt}\mathbb{E}\frac{\sum_{\{\sigma\}} \int \prod_\mu d\mu(z_\mu) O_s e^{H_s}}{\sum_{\{\sigma\}} \int \prod_\mu d\mu(z_\mu) e^{H_s}} = \mathbb{E}\left[\Omega\left(O_s \frac{dH}{dt}\right)\right] - \mathbb{E}\left[\Omega(O_s)\Omega\left(\frac{dH}{dt}\right)\right]. \tag{4.52}
$$

In the last equation eight terms contribute. Let us call them A_1, B_1, C_1, D_1, A_2,

B_2, C_2, D_2 and compute them explicitly:

$$
\begin{aligned}
A_1 &= \frac{1}{2}\sqrt{\frac{\beta}{tN^{1-\theta}}}\mathbb{E}\sum_{\{\sigma\}}\int\prod_{\mu}d\mu(z_\mu)O_s\sum_a\sum_{i,\mu}^{l_\eta,l_\chi}\xi_i^\mu\sigma_i^a z_\mu^a\frac{e^{H_s}}{Z_s} = \\
&= \frac{1}{2}\sqrt{\frac{\beta}{tN^{1-\theta}}}\mathbb{E}\sum_{\{\sigma\}}\int\prod_{\mu}d\mu(z_\mu)O_s\sum_a\sum_{i,\mu}^{l_\eta,l_\chi}\sigma_i^a z_\mu^a\partial_{\xi_i^\mu}\frac{e^{H_s}}{Z_s} = \\
&= \frac{\beta}{2N^{1-\theta}}\mathbb{E}\left[\sum_{a,b}^{s}\sum_{i,\mu}^{l_\eta,l_\chi}\Omega(O_s\sigma_i^a z_\mu^a\sigma_i^b z_\mu^b)-\sum_{a,b}^{s}\sum_{i,\mu}^{l_\eta,l_\chi}\Omega(O_s\sigma_i^a z_\mu^a)\Omega(\sigma_i^b z_\mu^b)\right] = \\
&= \frac{\beta}{2N^{1-\theta}}\mathbb{E}\left[\sum_a l_\eta\sum_{\mu}^{l_\chi}\Omega(O_s(z^a)^2)+\sum_{a\neq b}^{s}l_\eta l_\chi\Omega(O_s q_{ab}^{l_\eta}p_{ab}^{l_\chi})+\right. \\
&\quad \left.-s\,l_\eta l_\chi\sum_a\Omega(O_s q_{s+1,a}^{l_\eta}p_{s+1,a}^{l_\chi})\right] = \\
&= \frac{\beta LN^\theta}{2}\left(\frac{1+a}{2}\right)\left[\sum_a\langle O_s(z^a)^2\rangle+\sum_{a\neq b}^{s}\langle O_s q_{ab}^{l_\eta}p_{ab}^{l_\chi}\rangle+\right. \\
&\quad \left.-s\sum_a\langle O_s q_{s+1,a}^{l_\eta}p_{s+1,a}^{l_\chi}\rangle\right]. \qquad (4.53)
\end{aligned}
$$

$$
\begin{aligned}
A_2 &= -\frac{1}{2}\sqrt{\frac{\beta}{tN^{1-\theta}}}\mathbb{E}\Omega O_s\sum_{\{\sigma\}}\int\prod_{\mu}d\mu(z_\mu)\sum_a\sum_{i,\mu}^{l_\eta,l_\chi}\xi_i^\mu\sigma_i z_\mu\frac{e^{H_s}}{Z_s} = \\
&= -\frac{1}{2}\sqrt{\frac{\beta}{tN^{1-\theta}}}\mathbb{E}\Omega O_s\sum_{\{\sigma\}}\int\prod_{\mu}d\mu(z_\mu)\sum_a\sum_{i,\mu}^{l_\eta,l_\chi}\sigma_i z_\mu\partial_{\xi_i^\mu}\frac{e^{H_s}}{Z_s} = \\
&= -\frac{\beta}{2N^{1-\theta}}\mathbb{E}\Omega O_s\left[\sum_{a,b}^{s}\sum_{i,\mu}^{l_\eta,l_\chi}\Omega(\sigma_i^a z_\mu^a\sigma_i^b z_\mu^b)-\sum_{a,b}^{s}\sum_{i,\mu}^{l_\eta,l_\chi}\Omega(\sigma_i^a z_\mu^a)\Omega(\sigma_i^b z_\mu^b)\right] = \\
&= -\frac{\beta}{2N^{1-\theta}}\mathbb{E}\left[\sum_{a<b}\sum_{\mu}^{l_\chi}\Omega(O_s\sigma_i^{s+1}z_\mu^{s+1}\sigma_i^{s+2}z_\mu^{s+2})+\sum_a^s\Omega(O_s)\Omega((z_\mu^a)^2)+\right. \\
&\quad \left.+\quad s^2\Omega(O_s\sigma_i^{s+1}z_\mu^{s+1}\sigma_i^{s+2}z_\mu^{s+2})\right] = \\
&= -\frac{\beta}{2N^{1-\theta}}\mathbb{E}\left[\frac{s(s-1)}{2}\,l_\eta l_\chi\Omega(O_s q_{s+1,s+2}^{l_\eta}p_{s+1,s+2}^{l_\chi})+sl_\eta l_\chi\Omega(O_s(z_\mu^{s+1})^2)+\right. \\
&\quad \left.-s^2\Omega(O_s q_{s+1,s+2}^{l_\eta}p_{s+1,s+2}^{l_\chi})\right] = \\
&= -\frac{\beta LN^\theta}{2}\left(\frac{1+a}{2}\right)\left[s\langle O_s q_{s+1,s+2}^{l_\eta}p_{s+1,s+2}^{l_\chi}\rangle-s\langle O_s(z^{s+1})^2\rangle\right]. \qquad (4.54)
\end{aligned}
$$

With analogous calculations

$$B_1 = -\frac{b^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}} N\left[\sum_{a<b}\langle O_s q_{ab}^{l_\eta}\rangle - s\langle O_s q_{s+1a}^{l_\eta}\rangle + s\langle O_s\rangle\right], \tag{4.55}$$

$$B_2 = \frac{b^2}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}} N\left[-s\langle O_s q_{s+1,s+2}^{l_\eta}\rangle + s\langle O_s\rangle\right], \tag{4.56}$$

$$C_1 = -\frac{c^2}{2}\alpha\left(\frac{1+a}{2}\right)^{\frac{1}{2}} N\left[\sum_a\langle O_s((z^a)^2)\rangle + \sum_{a<b}\langle O_s p_{a,b}^{l_\chi}\rangle - s\sum_a\langle O_s p_{s+1,a}^{l_\chi}\rangle\right], \tag{4.57}$$

$$C_2 = -\frac{c^2}{2}\alpha\left(\frac{1+a}{2}\right)^{\frac{1}{2}} N\left[s\langle O_s(z^{s+1})^2\rangle - \frac{s(s+1)}{2}\langle O_s p_{s+1,s+2}^{l_\chi}\rangle\right], \tag{4.58}$$

$$D_1 = -\frac{d\alpha N}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}}\sum_a\langle O_s(z^a)^2\rangle, \tag{4.59}$$

$$D_2 = \frac{d\alpha N}{2}\left(\frac{1+a}{2}\right)^{\frac{1}{2}} s\langle O_s(z^{s+1})^2\rangle. \tag{4.60}$$

Therefore, merging all these terms together, the streaming is

$$\frac{d}{dt}\langle O_s\rangle_t = \beta\sqrt{\alpha}\frac{\gamma}{2}\left(\sum_{a<b}^{s}\langle OQ_{ab}^{l_\eta}P_{ab}^{l_\chi}\rangle_t - s\sum_a^s\langle OQ_{a,s+1}^{l_\eta}P_{a,s+1}^{l_\chi}\rangle_t + \right. \\ \left. + \frac{s(s+1)}{2}\langle OQ_{s+1,s+2}^{l_\eta}P_{s+1,s+2}^{l_\chi}\rangle_t\right). \tag{4.61}$$

In order to control the overlap fluctuations, namely $\langle Q_{l_\eta 12}^2\rangle_{t=1}$, $\langle Q_{l_\eta 12}P_{l_\chi 12}\rangle_{t=1}$, $\langle P_{l_\chi 12}^2\rangle_{t=1}$, ..., noting that the streaming equation pastes two replicas to the ones already involved ($s=2$ so far), we need to study nine correlation functions. It is then useful to introduce them and refer to them by capital letters so to simplify their visualization:

$$\langle Q_{l_\eta 12}^2\rangle_t = A(t), \quad \langle Q_{l_\eta 12}Q_{l_\eta 13}\rangle_t = B(t), \quad \langle Q_{l_\eta 12}Q_{l_\eta 34}\rangle_t = C(t), \tag{4.62}$$

$$\langle Q_{l_\eta 12}P_{l_\chi 12}\rangle_t = D(t), \quad \langle Q_{l_\eta 12}P_{l_\chi 13}\rangle_t = E(t), \quad \langle Q_{l_\eta 12}P_{l_\chi 34}\rangle_t = F(t), \tag{4.63}$$

$$\langle P_{l_\chi 12}^2\rangle_t = G(t), \quad \langle P_{l_\chi 12}P_{l_\chi 13}\rangle_t = H(t), \quad \langle P_{l_\chi 12}P_{l_\chi 34}\rangle_t = I(t). \tag{4.64}$$

Let us now sketch their streaming. First, we introduce the operator "dot" as

$$\dot{O} = \frac{2}{\beta\sqrt{\alpha}\gamma}\frac{dO}{dt},$$

which simplifies calculations and shifts the propagation of the streaming from $t=1$ to $t=\beta\sqrt{\alpha}\gamma/2$. Using this we sketch how to write the streaming of the first two

correlations (as it works in the same way for any other):

$$\begin{aligned}
\dot{A} &= \langle Q^2_{l_\eta 12} Q_{l_\eta 12} P_{l_\chi 12}\rangle_t - 4\langle Q^2_{l_\eta 12} Q_{l_\eta 13} P_{l_\chi 13}\rangle_t + 3\langle Q^2_{l_\eta 12} Q_{l_\eta 34} P_{l_\chi 34}\rangle_t,\\
\dot{B} &= \langle Q_{l_\eta 12} Q_{l_\eta 13}\Big(Q_{l_\eta 12} P_{l_\chi 12} + Q_{l_\eta 13} P_{l_\chi 13} + Q_{l_\eta 23} P_{l_\chi 23}\Big)\rangle_t -\\
&- 3\langle Q_{l_\eta 12} P_{l_\chi 13}\Big(Q_{l_\eta 14} P_{l_\chi 14} + Q_{l_\eta 24} P_{l_\chi 24} + Q_{l_\eta 34} P_{l_\chi 34}\Big)\rangle_t + 6\langle Q_{l_\eta 12} P_{l_\chi 13} Q_{l_\eta 45} P_{l_\chi 45}\rangle_t
\end{aligned} \tag{4.65}$$

By assuming a Gaussian behavior, as in the strategy outlined in [70], we can write the overall streaming of the correlation functions in the form of the following differential system

$$\begin{aligned}
\dot{A} &= 2AD - 8BE + 6CF,\\
\dot{B} &= 2AE + 2BD - 4BE - 6BF - 6EC + 12CF,\\
\dot{C} &= 2AF + 2CD + 8BE - 16BF - 16CE + 20CF,\\
\dot{D} &= AG - 4BH + 3CI + D^2 - 4E^2 + 3F^2,\\
\dot{E} &= AH + BG - 2BH - 3BI - 3CH + 6CI + 2ED - 2E^2 - 6EF + 6F^2,\\
\dot{F} &= AI + CG + 4BH - 8BI - 8CH + 10CI + 2DF + 4E^2 - 16EF + 10F^2,\\
\dot{G} &= 2GD - 8HE + 6IF,\\
\dot{H} &= 2GE + 2HD - 4HE - 6HF - 6IE + 12IF,\\
\dot{I} &= 2GF + 2DI + 8HE - 16HF - 16IE + 20IF.
\end{aligned} \tag{4.66}$$

As we are interested in discussing criticality and not the whole glassy phase, it is possible to solve this system starting from the high noise region, once the initial conditions at $t = 0$ are known. As at $t = 0$ everything is factorized, the only needed check is by the correlations inside each party. Starting with the first party, we have to study A, B, C at $t = 0$. As only the diagonal terms give non-negligible contribution, it is immediate to work out this first set of starting points as

$$\begin{aligned}
A(0) &= \langle Q^2_{l_\eta 12}\rangle = N(\langle (q^{l_\eta}_{12})^2\rangle - 2\mathbb{E}_{l_\eta}\overline{q}^{l_\eta}\langle q^{l_\eta}_{12}\rangle + \mathbb{E}_{l_\eta}\overline{q}^{l_\eta} =\\
&= N\big(\mathbb{E}_{l_\eta}\frac{1}{l^2_\eta}\langle\sum_i^{l_\eta}(\sigma^1_i)^2(\sigma^2_i)^2\rangle + \overline{q}^2\big) = \left(\frac{1+a}{2}\right)^{-\frac{1}{2}} - N\overline{q}^2\\
&= \sqrt{\frac{2}{\gamma}}N^{\frac{\theta}{2}} - N\overline{q}^2,
\end{aligned} \tag{4.67}$$

$$B(0) = \langle Q_{l_\eta 12} Q_{l_\eta 13}\rangle = N(\langle q^{l_\eta}_{12} q^{l_\eta}_{13}\rangle - \overline{q}^2) = \sqrt{\frac{2}{\gamma}}N^{\frac{\theta}{2}}\overline{q} - N\overline{q}^2, \tag{4.68}$$

$$\begin{aligned}
C(0) &= \langle Q_{l_\eta 12} Q_{l_\eta 34}\rangle = N(\langle q^{l_\eta}_{12} q^{l_\eta}_{34}\rangle - \overline{q}^2) = N\big(\mathbb{E}_{l_\eta}\frac{1}{l^2_\eta}\langle\sum_i^{l_\eta}\sigma^1_i\sigma^2_i\sigma^3_i\sigma^4_i\rangle - N\overline{q}^2 =\\
&= \sqrt{\frac{2}{\gamma}}N^{\frac{\theta}{2}}\int d\mu(\eta)\tanh^4\left(\frac{\beta\sqrt{\alpha\overline{q}\frac{\gamma}{2}}N^{\theta/2}}{1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1-\overline{q})}\right) - N\overline{q}^2.
\end{aligned} \tag{4.69}$$

For the second party we need to evaluate G, H, I at $t = 0$. The only difference

with the first party is that $z_\mu^2 \neq 1$ as for the σ's.

$$\begin{aligned} G(0) &= \langle P^2_{l_\chi 12}\rangle = N(\langle (p_{12}^{l_\chi})^2\rangle - 2\mathbb{E}_{l_\chi}\overline{p}^{l_\chi}\langle p_{12}^{l_\chi}\rangle + \mathbb{E}_{l_\chi}\overline{p}^2_{l_\chi} = \\ &= N\big(\mathbb{E}_{l_\chi}\frac{1}{l_\chi^2}\sum_\mu^{l_\chi}\langle (z_\mu^1)^2\rangle_G\langle (z_\mu^2)^2\rangle_G - \overline{p}^2\big) = \\ &= \big(\alpha\frac{\gamma}{2}\big)^{-\frac{1}{2}}N^{\theta/2}\ \omega^2(z^2) - N\overline{p}^2, \qquad (4.70) \\ H(0) &= \langle P_{l_\chi 12}P_{l_\chi 13}\rangle = N(\langle p_{12}^{l_\chi}p_{13}^{l_\chi}\rangle - \overline{p}^2) = \\ &= \big(\alpha\frac{\gamma}{2}\big)^{-\frac{1}{2}}N^{\theta/2}\ \omega(z)\ \omega(z^2) - N\overline{p}^2, \qquad (4.71) \\ I(0) &= \langle P_{l_\chi 12}P_{l_\chi 34}\rangle = \\ &= N(\langle p_{12}^{l_\chi}p_{34}^{l_\chi}\rangle - \overline{p}^2) = \big(\alpha\frac{\gamma}{2}\big)^{-\frac{1}{2}}N^{\theta/2}\ \omega^2(z) - N\overline{p}^2. \qquad (4.72) \end{aligned}$$

Now, $\omega(z^2)$ and $\omega(z)$ are Gaussian integrals and can be explicitly calculated as

$$\begin{aligned} \omega(z) &= \frac{\int d\mu(z) z\exp\left(bz\chi + \frac{d}{2}z^2\right)}{\int d\mu(z)\exp\left(bz\chi + \frac{d}{2}z^2\right)} = \frac{b\langle\chi\rangle}{1-d} = 0, \qquad (4.73) \\ \omega(z^2) &= \frac{\int d\mu(z) z^2\exp\left(bz\chi + \frac{d}{2}z^2\right)}{\int d\mu(z)\exp\left(bz\chi + \frac{d}{2}z^2\right)} = \frac{1-d+b^2\langle\chi^2\rangle}{(1-d)^2} = \\ &= \frac{1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1-\overline{q}+\alpha\overline{p})}{\left(1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1-\overline{q})\right)^2}\,. \qquad (4.74) \end{aligned}$$

Finally, we have obviously $D(0) = E(0) = F(0) = 0$, because at $t = 0$ the two parties are independent. As we are interested in finding where ergodicity becomes broken (the critical line), we start propagating t (from 0 to 1) from the annealed region (high noise limit), where $\bar{q} \equiv 0$ and $\bar{p} \equiv 0$. It is immediate to check that, for the only terms that we need to consider, A, D, G (the other being strictly zero on the whole $t \in [0, 1]$), the starting points are:

$$A(0) = \sqrt{\frac{2}{\gamma}}N^{\frac{\theta}{2}} = \frac{1}{r}, \qquad (4.75)$$

$$D(0) = 0, \qquad (4.76)$$

$$G(0) = \frac{N^{\frac{\theta}{2}}}{\sqrt{\frac{\gamma}{2}}(1-\beta\sqrt{\frac{\gamma}{2}}N^{\frac{\theta}{2}})^2} = \frac{1}{rs^2}. \qquad (4.77)$$

Where we have defined $r = \sqrt{\frac{\gamma}{2}}N^{-\frac{\theta}{2}}$, $s = 1 - \beta\sqrt{\frac{\gamma}{2}}N^{\frac{\theta}{2}}$.

The evolution is ruled by

$$\dot{A} = 2AD \qquad (4.78)$$

$$\dot{D} = AG + D^2 \qquad (4.79)$$

$$\dot{G} = 2GD. \qquad (4.80)$$

Noticing that $\frac{\dot{A}}{G} = 0$ by substitution, and that $\frac{A(0)}{G(0)} = s^2$ we obtain immediately :

$$A(t) = G(t)s^2 = G(t)\left(1 - \beta\sqrt{\frac{\gamma}{2}}N^{\frac{\theta}{2}}\right)^2. \tag{4.81}$$

The system then reduces to two differential equations; calling $Y = D + Gs$, we have $\dot{Y} = \dot{D} + \dot{G}s = G^2s^2 + D^2 + 2GDs = Y^2$ with solution $Y(t) = \frac{Y(0)}{1-tY(0)}$, and $Y(0) = D(0) + G(0)s = \frac{1}{rs}$ by which we get

$$Y(t = \sqrt{\alpha}\beta\frac{\gamma}{2}) = \frac{1}{rs}\ \frac{1}{1 - \sqrt{\alpha}\beta\frac{\gamma}{2}(rs)^{-1}} = \frac{1}{\sqrt{\frac{\gamma}{2}}N^{-\theta/2}\left(1 - \beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1+\sqrt{\alpha})\right)}, \tag{4.82}$$

i.e. there is a regular behavior up to

$$\beta_c = \frac{1}{\sqrt{\frac{\gamma}{2}}N^{\frac{\theta}{2}}(1+\sqrt{\alpha})}\,, \tag{4.83}$$

which confirms the result obtained in Equation (4.49). Now, we can consider separately the evolution equation for G and D:

$$\dot{G} = 2G(t)\Big(Y(t) - sG(t)\Big) = \frac{2}{rs-t}G(t) - 2sG(t), \tag{4.84}$$

where we used $Y(t) = (rs - t)^{-1}$. Dividing both sides by G^2 and calling $Z = G^{-1}$ we get an ordinary first order differential equation for $Z(t)$:

$$-\dot{Z}(t) = 2Y(t)Z(t) - 2s = \frac{2}{rs-t}Z(t) - 2s. \tag{4.85}$$

that have the following solution for the initial condition $Z(0) = rs^2$:

$$Z(t) = 2s(rs - t) - \frac{1}{r}(rs - t)^2. \tag{4.86}$$

From $Z(t)$ we obtain $G(t)$, that is,

$$G(t) = \frac{1}{(rs-t)(s+\frac{t}{r})}. \tag{4.87}$$

Using Equation (4.81) and remembering that $D(t) = Y(t) - G(t)s$, we obtain the other overlap fluctuations

$$\langle Q^2_{l_\eta 12}\rangle = \frac{\left(1 - \beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}\right)^2}{\sqrt{\frac{\gamma}{2}}N^{-\theta/2}\left(1 - \beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1+\sqrt{\alpha})\right)\left(1 - \beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1-\sqrt{\alpha})\right)}, \tag{4.88}$$

$$\langle Q_{l_\eta 12}P_{l_\chi 12}\rangle = \frac{\sqrt{\alpha}\beta}{\left(1 - \beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1+\sqrt{\alpha})\right)\left(1 - \beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1-\sqrt{\alpha})\right)}, \tag{4.89}$$

$$\langle P^2_{l_\chi 12}\rangle = \frac{1}{\sqrt{\frac{\gamma}{2}}N^{-\theta/2}\left(1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1+\sqrt{\alpha})\right)\left(1-\beta\sqrt{\frac{\gamma}{2}}N^{\theta/2}(1-\sqrt{\alpha})\right)}. \quad (4.90)$$

A simple visual inspection of the formula above allows to confirm that the poles are located at

$$\beta\sqrt{\frac{\gamma}{2}}\,N^{\theta/2}\,(1+\sqrt{\alpha})=1,$$

confirming the heuristic result previously obtained. We can easily see furthermore that in the fully connected limit ($\gamma = 2$ and $\theta = 0$) we recover the result of [20].

4.5 Conclusions and outlooks

In this Chapter we introduced and solved, at the replica symmetric level, two disordered mean-field systems: the former provides a generalization of the analogue neural network by introducing dilution into its patterns encoding the memories, the latter is a bipartite and diluted spin-glass made up of a Gaussian party and an Ising party, respectively. From an applicative viewpoint (not discussed here, see e.g. [51]), the interest in these models raises in different contexts, but their peculiarity resides in the existence of sparse entries (instead of classical dilution on the neural network links as performed for instance earlier by Sompolinsky [80] through random graphs or recently by Coolen and coworkers [62] through small-worlds or scale-free architectures) which allows, when possible, parallel retrieval as for instance discussed in [52], [51], [53] and [54]. Interestingly, as we show, the Hamiltonians describing these systems are thermodynamically equivalent.

In our investigations we first considered the diluted analogue neural network and focused on the topological properties of the emergent weighted graph. We found an exact expression for the coupling distribution, showing that in the thermodynamic limit it converges to a central Gaussian distribution with variance scaling linearly with the system size N. We also calculated the average link probability which, as expected, depends crucially on the degree of dilution introduced. More precisely, by properly tuning it, the emergent structure displays an average coordination number $\bar{z}$ which can range from $\bar{z} = N$ (fully-connected regime) to $\bar{z} = \mathcal{O}(N)$ (constant link probability), to finite with $\bar{z} > 1$ (over-percolated network) or $\bar{z} < 1$ (under-percolated network).

Then, we moved to the thermodynamical analysis, where, through an interpolation scheme recently developed for fully connected Hebbian kernels [60], we obtained explicitly the replica symmetric free-energy coupled with its self-consistency equations. The overlaps, order parameters of the theory, turn out to be classical weighted sums of sub-overlaps defined on all possible sub-graphs (as for instance discussed in [37, 63]). Both a small overlap expansion of these self-consistencies, as well as a whole fluctuation theory developed for their rescaled correlations, confirm a critical behavior on a surface (in the $\alpha, \beta, \gamma, \theta$ hyperplane) that reduces to the well-known of Amit-Gutfreud-Sompolinsky when the dilution is sent to zero [20]. On the other hand, the net effect of entry dilution in bit-strings (which weakens the coupling strength) is to rescale accordingly the critical noise level at which ergodicity breaks down, as expected.

Without imposing retrieval through Lagrange multipliers (as for analogue patterns it is not a spontaneous phenomenon, see [60]) the system displays only two phases, an ergodic one (where all overlaps are zero) and a spin-glass one (where overlaps are non-zero), split by the second order critical surfaces (over which overlaps start being non-zero) which defines criticality.

The results reviewed in this Section have been published in [2].

Chapter 5

A Markov chain model for the analysis of clinical data

Last Chapters have been dedicated to coarse grain modeling of biological networks. The intent of the works presented there was to present solvable models that reproduce qualitative features of the biological systems. In this Chapter and in the next one, we move towards models that are meant to deal with real data.

Here we review the work presented in [3]. In this work a method for analyzing clinical data is provided and tested. This data are about the insurgence of tuberculosis (TB) infection and non-tuberculous mycobacterial (NTM) infection as a side effect in patients affected by rheumatoid arthritis (RA) that are treated with inhibitors of the tumor necrosis factor (TNF), a proinflammatory cytokine.

With the aim of discriminate between new born infections and reactivation of latent infections in the treated patients, a probabilistic model, that was previously proposed by R. S. Wallis in [81], is analytically studied and tested on real data.

5.1 The biomedical framework: rheumatoid arthritis, tumor necrosis factor inhibitors, tuberculosis and nontuberculous mycobacterial infections.

Over the last decades the improved understanding of the pathogenesis of chronic inflammatory diseases, together with a major advance in biotechnology, have accelerated the development of biological therapies, designed to neutralize specific targets that mediate and sustain the clinical manifestations of diseases. These compounds, mainly monoclonal antibodies (mAb) and fusion proteins, introduced a breakthrough in the management of different conditions including inflammatory rheumatologic disorders [82]. In this context, the first setting of application of the biological agents was RA, a chronic autoimmune disease affecting approximately 1% of the adult population [83]. If the disease is not treated adequately, progressive deformity can lead to loss of quality of life and reduce average life expectancy by about a decade [83]. Studies on the pathogenic mechanisms of RA have revealed that TNF is a cytokine playing a critical role in the inflammatory cascade that results in the irreversible joint damage typical of the disease [84]. Following these

discoveries, a series of clinical trials in patients with RA showed the therapeutic benefit of TNF blockade [85]. As a consequence, five biological agents engineered to block TNF actions are currently available: infliximab, adalimumab, golimumab, certolizumab pegol (all of them mAb), and etanercept (a receptor fusion protein) [86]. While being highly effective, TNF blockers have raised concerns about the potential for an increased susceptibility to infections, in particular the reactivation of latent tuberculosis (TB) infection [87, 88, 89, 81, 90]. Mycobacterium tuberculosis, the cause of human TB, can result in a metastable clinical latency lasting for decades. Much has been speculated about the structure of granuloma which should contain Mycobacteria, since murine models indicated that TNF was necessary for both formation and maintenance of granulomas [91]. However, subsequent studies on zebrafish model [92], monkeys [93], and humans [94, 95, 96, 97] challenged these data, demonstrating that the crucial role of TNF in the granuloma was indeed macrophage activation and stimulation of chemokine production. The reactivation of latent TB infection has been associated with all TNF inhibitors, hence pre-initiation screening procedures have been recommended, which have successfully reduced the number of reported cases [98], although current screening tools lack sensitivity and specificity [99, 100].

TNF blockers seem to increase also the risk of other granulomatous diseases, but little is known about the emergence of illness due to non-tuberculous Mycobacteria. These are a huge ensemble of pathogens (e.g. M. avium, M. abscessus, and so on [101]) and up to date, approximately 50 different mycobacterial species are considered to be etiological agents of human diseases and this number seems still growing [101]. Most cases these days occur in hosts with relatively intact immune responses. However, RA and other chronic diseases with pulmonary manifestations can predispose a person to NTM pulmonary disease [102] expressing as a possible serious complication, especially in immunosuppressed subjects. Thus, it is of utmost importance to study also the risk related to NTM, in the perspective to understand if a proper screening may be helpful in conferring a wider protection to the patients. This is particularly true, in that the TNF blockers appear to predispose both to disseminated and localized disease [103, 101], but also because these infections are increasing in prevalence, especially among women, which are more frequently affected by RA than men. In the present state of the art, the scenario for NTM diseases, with respect to TNF-blocking drugs, seems different from the TB counterpart: in particular, through extensive experimental screening, both Wallis and coworkers [90], and Winthrop and coworkers [103, 104] evidenced that latency and reactivation do not seem to play a crucial role in this context, yet a clear-cut picture is still missing.

Now, as far as TB is concerned, data collected through the Adverse Event Reporting System of the US Food and Drug Administration (FDA) in the time-window $1998 - 2002$, related to the two test-case drugs with a different mechanism of action, i.e. infliximab and etanercept, highlight that TB infections involve 54 over 10^5 patients treated with infliximab and 28 over 10^5 patients receiving etanercept [90]. Therefore, the question is: As the latency in TB can last decades, are these infections (in patients under therapy) new ones or are they reactivation of previously encountered pathogens due to a suppressed immune system? This kind of question underlies the awareness of a real need and disposal for extensive pre-screening procedures. Unfortunately, the answer is by far not trivial as, for TB, there are no

secure pathways to discriminate between a new infection or the raise of a previous one. Moreover, a clear methodology for finding latencies is still lacking. Furthermore, the rarity and different sizes of this infection in different countries (ranging from 5 over 10^5 in Sweden up to 140 over 10^5 in Romania [89]) implies that data analysis and its subsequent interpretation must be carefully performed.

As for NTM, still from FDA, through the post-marketing surveillance system (MedWatch) and through a further survey within the Emerging Infections Network of the Infectious Diseases Society of America (IDSA), Winthrop and coworkers reported a detailed study of possible correlations between the usage of TNF blockers and the emergence of NTM diseases: over a time-window of 8 years, they highlighted a higher prevalence of NTM diseases in patients treated with infliximab rather than etanercept [104].

In order to investigate possible correlations between the incidence of infections by such Mycobacteria and biological therapy, one could rely directly on the molecular details of TNF processing signal (which has been, at least partially, elucidated, see e.g. [105]), coupled to the underlying infliximab and etanercept mechanism of action, which could be achievable directly through molecular immunology approach. Beyond these "standard" strands, a completely different route can also be performed: Given the relative large amount of collected data, the problem can be considered from a purely inferential viewpoint, by-passing the underlying molecular immunology know-how (see also [106, 107]). According to this perspective, in Ref. [81, 90], an abstract (logical) environment for TB case has been defined, where patients can occupy one of the (following) five different states: (0) No infection, (1) New infection, (2) Latency, (3) Reactivation of a previous TB infection, (4) Post first TB encounter. Clearly, the patients starting the therapy (and hence belonging to the survey) can correspond to either state (0) or (2), because all the other states imply quantifiable sickness and the patient would then be treated for TB rather than RA. Then, at the end of the survey, a fraction of these patients will be in an illness state, i.e. either state (3) or (4). The transition rates between different states are assumed as free-parameters, whose values are estimated through numerical simulations: the best estimate is the one able to reproduce, with the smallest error, the experimental data. Remarkably, the probability of latent TB reactivation in patients treated with infliximab turned out to be an order of magnitude per unit of time higher than the same probability for patients trated with etanercept [81].

Here, we first formalize this approach in terms of Markov chains and we write the related Master equation in continuous-time limit, then we solve the model analytically and study its properties in full details. In this way we get the explicit expression for the number of patients $c(t)$ exhibiting a TB (re)activation, as a function of time t. One step forward, we check the robustness of our results through extensive Monte Carlo simulations and over the clinical data of the TB scenario, finding overall excellent agreement among all our results (and previous literature). Moreover, we find that different magnitudes for the probability of reactivation correspond to qualitative different behaviors for $c(t)$ (on the proper timescale), that is, the number of patients displaying active infection increases exponentially in time when using infliximab and linearly in time when using etanercept.

The analytical expression for the whole evolution of the system implies a great feasibility of the technique itself (e.g. we have the whole set of first integrals

and a clear picture of all the hidden symmetries) and also allows to address, in complete generality, several instances. In particular, we can finally consider generic NTM infections, where, interestingly, the scenario appears quite different from the TB counterpart: clinical data suggest that $c(t)$ (on the proper timescale) grows quadratically with time and this is recovered by our analytical picture only under the assumption of a negligible role played by latency reactivation. We check these findings also through extensive Monte Carlo runs, which are in full agreement too. Remarkably, this is very consistent with the present state of the art in the medical literature dealing with NTM.

As a final result, there are two types of conclusions which stem from our work: The former belongs to the world of modelers, while the latter to the world of clinicians.
From a mathematical perspective, the encouraging results of this approach may pave the way for the development of handily and fruitful instruments for physicians.
Much more carefully, in the clinician's counterpart, as this approach bypasses the whole underlying biological complexity, it may contribute to confirm, from a theoretical perspective, the current understanding of adverse events coupled to TNF-inhibitors and the consequent real need for screening procedures before undergoing biological therapies.

5.2 The mathematical model

In this section we formalize the scheme introduced in [81] and aimed to reproduce data of TB onset in patients treated with TNF inhibitors, with particular attention on infliximab (an anti-TNF mAb) and etanercept (a soluble TNF receptor). Seeking for clarity, in this section we mention only applications to the TB case, although, as we will see in Section 5.3.2, the approach is rather robust and can be successfully applied to the NTM case, too.

The model, whose structure is depicted in Figure 5.1, consists in identifying a set of possible states for the patient subjected to biological treatments, and in fixing the likelihood for the patient to change his/her state within a proper unit time [1].

The clinical states available to a test-patient are taken as follows (see Figure 1):

0 : Absence of infection;

1 : New infection (that after a time τ can give rise either to active TB or latent infection);

2 : Latent infection;

3 : Reactivated TB after latency;

4 : Active TB (that progress from new infection within a time τ, without an intervening period of latency).

[1] Clearly, on large samples, some patients may experience sudden incidents (e.g. death for other causes) or some others may assume both the drugs: the analysis has been previously purified from these cases [81, 90].

Moreover, each patient is assumed to change his/her state, following the corresponding transition probabilities, which constitute the model parameter set, and are meant over a proper unit time τ. Using t to label the time, these probabilities are:

L : Probability of having a latent infection at the beginning of the observation ($t = 0$), while, obviously, $(1 - L)$ is the probability of not having any infection at that moment;

N : Time rate of TB infection during the observational time;

P : Time rate at which a new TB infection becomes active TB; as a consequence, $(1 - P)$ is the time rate at which this new infections to give rise to a latent infection;

R : Time rate of reactivation of a latent TB infection.

As a hypothesis of the model only the parameters P and R may depend on the particular TNF inhibitor that the patient is taking.

We stress that such probabilistic framework is based on purely clinical variables.

On the experimental side, the available data consist in a collection of times (one for each patient) corresponding to the onset of TB (in its active phase, namely a detectable scenario), after the beginning of the treatment with TNF blockers. As a consequence, the only states which are possible to observe are the states 3 and 4. Unfortunately, as discussed in the introduction, these states (that account for ill patients) are not distinguishable one respect to the other by simply looking at the data (hence motivating both earlier studies [81, 90] and our machinery), however, some progress can be made using stochastic extremization. The idea resembles the standard maximum likelihood and consists in finding the best values for free parameters such that the theoretical curves collapse over the experimental data [108].

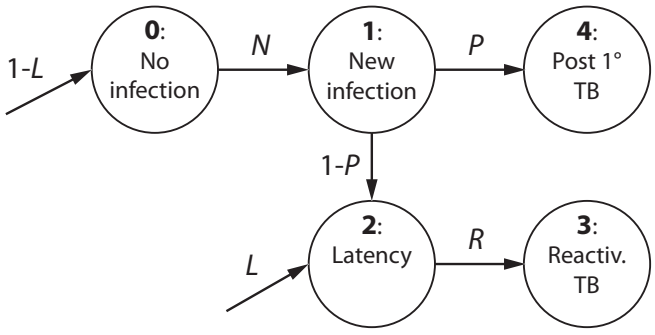


Figure 5.1. Symbolic representation of the Markov chain under investigation. States are represented as circles and numbered from 0 to 4 according to the scheme outlined in Section 5.2. The arrows $N, P, 1 - P, R$ represent the transition probabilities connecting two different states, while $1 - L$ and L represent the initial conditions on the states $0, 2$.

Markov chains and master equations

The model described in the previous section can be translated into a set of differential equations coding for the temporal evolution of the probability of patient's states (which can be compared to the corresponding fractions over a sample of patients given the large collection of data).

Being the states discrete, this can be accomplished in complete generality using Markov chains, namely a (discrete-time) probabilistic framework where the probability of being in a given state at a given time t depends only on the probability distribution over all the states at the previous time step $t-1$, and on the transition rates linking these states.

It is instructive to consider the illustrative Markov chain with only three states (A, B and C), non-null transition rates $w_{A\to B}$ and $w_{B\to C}$ and time step Δt, shown in Figure 5.2.

Note that, in the model, the probabilities of going from A to B and from B to C exist but not the opposite ($w_{B\to A}=0,\ w_{C\to B}=0$) hence, if the initial state is all concentrated in C, there will be no evolution, while if the starting point is spread among A and B, after enough time, the probability distribution will be peaked on C only (but in its *finite* temporal evolution resides our interest).

Now, the probability of remaining-at/moving-into the state B in the time interval Δt is given by the probability of already being in B (hence $p_B(t)$) plus the probability of arriving in B from A times the probability of being in A at the previous step (hence $w_{A\to B}p_A(t)$) minus the probability of leaving B to C times the probability of being in B at the previous step (hence $w_{B\to C}p_B(t)$); this concept can be written as follows:

$$p_B(t+\Delta t) = p_B(t) + p_A(t)\, w_{A\to B}\Delta t - p_B(t)\, w_{B\to C}\Delta t\,. \tag{5.1}$$

Since the mathematics for continuous variable differential equations is much more handily and does not change significantly the perspective if the time step is small with respect to the global time window[2], we are allowed to consider the time as a continuous variable. This can be achieved straightforwardly starting from the previous equation using a limit procedure:

$$\lim_{\Delta t\to 0}\frac{p_B(t+\Delta t)-p_B(t)}{\Delta t} = \frac{dp_B(t)}{dt} = p_A(t)w_{A\to B} - p_B(t)w_{B\to C}.$$

The evolution for the probability $p_B(t)$ is then ruled by the following differential equation, namely a "Master equation", which acts as a continuous counterpart of the Markov chain in the discrete-time case:

$$\frac{dp_B(t)}{dt} = p_A(t)w_{A\to B} - p_B(t)w_{B\to C}\,. \tag{5.2}$$

In general, for a system which can be in one of M arbitrary states, we need a $M\times M$ transition-rate matrix w (where $w_{i\to j}$ is the rate for the transition from state i to state j) and the Master equation takes the form

$$\dot{p}_i(t) \equiv \frac{dp_i}{dt} = \sum_{j=1}^{M} w_{j\to i}\, p_j(t) - \sum_{j=1}^{M} w_{i\to j}\, p_i(t)\,. \tag{5.3}$$

[2]Both the experimental data set considered here (for TB cases [90] and for NTM cases [103]) fulfill this requirement.

Finally, we switch to a form where the explicit timescale τ of the process appears directly in the equation, that is

$$\tau \dot{p}_i(t) = \sum_{j=1}^{M} W_{j\to i}\, p_j(t) - \sum_{j=1}^{M} W_{i\to j}\, p_i(t), \tag{5.4}$$

where $W_{i\to j} \sim w_{i\to j}\tau$ stands for the probability of transition from state i to j along the time interval τ.

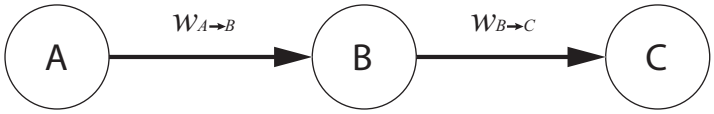


Figure 5.2. Toy Markov chain. From the state A there is a potential flux of probability at rate $w_{A\to B}$ toward the state B, hence we expect that, after a proper amount of time, a fraction of the probability p will be drained from A to B. The same holds for the situation linking C to B. After an infinite time the probability of having the patient in the state C is one, while it is zero for the states A, B.

Master equations for the model

Keeping in mind Figure 5.1, we can write down the system of differential equations describing the evolution of the five states earlier introduced as follows:

$$\begin{cases} \tau \dot{p}_0 = -N\, p_0(t), \\ \tau \dot{p}_1 = N p_0(t) - p_1(t), \\ \tau \dot{p}_2 = (1-P)\, p_1(t) - R\, p_2(t), \\ \tau \dot{p}_3 = R\, p_2(t), \\ \tau \dot{p}_4 = P\, p_1(t), \end{cases} \tag{5.5}$$

with initial conditions

$$\begin{cases} p_0(t=0) = 1 - L, \\ p_2(t=0) = L, \\ p_1(t=0) = p_3(t=0) \;=\; p_4(t=0) \;=\; 0. \end{cases} \tag{5.6}$$

The numbers indexing the probabilities mirror the enumeration of the previous section, that is, p_0 stands for the probability that a patient has never been affected by the infection, and so on. The parameter τ represents the typical time for a patient experiencing a new infection to either develop the disease or to fall into a latent state and it should be chosen according to the natural time-scale of the process described. For instance, for the TB case, the data collected suggest that τ is order of a few months [90, 109], and we set $\tau = 1\, month$ for the sake of simplicity and in agreement with previous works [81, 90].

Note that, as patients affected by active TB do not start RA therapy, we set $p_1(t=0) = p_3(t=0) = p_4(t=0) = 0$. Furthermore, the parameter L tunes the initial amount of latent-TB patients with respect to free-TB patients, such that for $L = 0$ all patients are healthy, while for $L = 1$ all patients display a latent TB infection; as we have no ways to discriminate between healthy and latent-infected patients, L is taken as a free parameter which can be estimated a posteriori comparing the solution of (5.5) with available data.

The solution of the system (5.5) can be easily obtained using first order ordinary differential equations theory and reads off as

$$\begin{cases} p_0(t) = (1-L)\, e^{-Nt/\tau}, \\ p_1(t) = N\frac{1-L}{1-N}\left(e^{-Nt/\tau} - e^{-t/\tau}\right), \\ p_2(t) = -\frac{(1-P)(1-L)N}{(1-N)(N-R)}\, e^{-Nt/\tau} + \left[L + \frac{(1-P)(1-L)N}{(N-R)(1-R)}\right]\, e^{-Rt/\tau} + \frac{(1-P)(1-L)N}{(1-N)(1-R)}\, e^{-t/\tau}, \\ p_3(t) = L + (1-P)(1-L) + \frac{R(1-P)(1-L)}{(1-N)(N-R)} e^{-Nt/\tau} - \left[L + \frac{(1-P)(1-L)N}{(1-R)(N-R)}\right] e^{-Rt/\tau} \\ \qquad -\frac{(1-P)(1-L)RN}{(1-N)(1-R)} e^{-t/\tau}, \\ p_4(t) = P\frac{1-L}{1-N}\left(1 - N - e^{-Nt/\tau} + N\, e^{-t/\tau}\right). \end{cases} \tag{5.7}$$

Of course, since the total amount of patients is conserved, $C_0 = p_0 + p_1 + p_2 + p_3 + p_4$ is an integral of motion, that is

$$0 = \frac{d}{dt} C_0 \Rightarrow C_0 = p_0(t) + p_1(t) + p_2(t) + p_3(t) + p_4(t) = \text{const.} \tag{5.8}$$

Beyond C_0, the system (5.5) admits another integral of motion C_1, namely

$$0 = \frac{d}{dt}\left[p_2(t) + p_3(t) + \frac{P-1}{P} p_4(t)\right] \Rightarrow C_1 = p_2(t) + p_3(t) + \frac{P-1}{P} p_4(t) = \text{const.} \tag{5.9}$$

This means that the rate of growth for patients in the latency branch (i.e. in states $2, 3$) equals the rate of growth for the rest of infected patients (i.e. in state 4) weighted by a factor $P^{-1} - 1$, so that the smaller P and the larger the difference between the related rates. The knowledge of integrals of motion can be very useful as they allow to obtain information in a very simple way; for instance, should P drop, then $p_4(t)$ would also decrease (or, analogously, $p_2(t) + p_3(t)$ would increase) in order to maintain C_1 constant. Given C_0 and C_1, other integrals of motion, which are combination of C_0 and C_1, can be trivially built. For example, $C_2 = C_0 - C_1$ fulfills

$$0 = \frac{d}{dt}\left[p_1(t) + p_0(t) + \frac{1}{P} p_4(t)\right] \Rightarrow C_2 = p_1(t) + p_0(t) + \frac{1}{P} p_4(t) = \text{const.} \tag{5.10}$$

We underline that this kind of investigation can be accomplished only through an analytical study of the system.

As discussed above, the fraction of active TB cases is given by the sum of the fraction of cases of direct TB after infection and of the fraction of cases with reactivated TB; namely, calling $c(t)$ the total fraction of cases, we have:

$$c(t) = p_3(t) + p_4(t)\,. \tag{5.11}$$

In the above quantity, the time dependence appears only through three different exponential decay terms ($e^{-Nt/\tau}$, $e^{-Rt/\tau}$, $e^{-t/\tau}$), which vanish at infinite time, so that the solution becomes a constant term equal to 1, meaning that, if we wait for a sufficient long (possibly infinite) time, all patients become sick (although, obviously, they can possibly die earlier due to reasons not related to RA/TB). In order to deepen the temporal evolution of these probabilities at relatively short times, it is useful to use a little bit of algebraic manipulation to distinguish constant terms from decaying terms, in such a way that we get

$$c(t) = 1 + k_1 e^{-t/\tau} + k_R e^{-Rt/\tau} + k_N e^{-Nt/\tau}, \tag{5.12}$$

where the three constants k_1, k_R and k_N are related to the physiological parameters by

$$\begin{cases} k_1 = \frac{N(1-L)(P-R)}{(1-N)(1-R)}, \\ k_R = -L + \frac{N(1-P)(1-L)}{(1-R)(R-N)}, \\ k_N = -\frac{(1-L)(R-PN)}{(1-N)(R-N)}. \end{cases} \tag{5.13}$$

Of course, from $c(t)$ one can derive the effective number of cases multiplying $c(t)$ by the overall number of treated patients.

Before turning attention to the fitting procedure, we stress that the analytical solution in Eq. (5.7) was successfully checked through numerical methods, i.e. fourth-order Runge-Kutta algorithm and Monte Carlo simulations[3].

5.3 Results

5.3.1 The TB-infection case

Having obtained the complete solution of the model and exploiting the available information on parameters, we now look for proper approximations able to highlight the effective behavior of $c(t)$ in cases of practical interest, starting with the TB scenario.

In particular, Eq. (5.12) can be reduced to a simpler form if we hypothesize that the probability N of TB infection is much smaller than all the other parameters, in agreement with studies on TB and with results found in [81] and, a posteriori, in the current work (see Table 5.1).

Hence, as a first approximation step, we assume $N \ll 1$ and $N/R \ll 1$ such that

[3]Here, with Monte Carlo simulation we mean a simulation in which a set of virtual patients evolves in time following the Markov chain of Figure 5.1 giving a sample of the evolution of the fraction of cases during time. In our simulations we set 10^6 virtual patients.

we can expand the solution, at the first order in N and N/R, as follows[4]:

$$\begin{aligned} k_1 &= (1-L)\left(P - R\,\frac{1-P}{1-R}\right) N[1+N+\mathcal{O}(N^2)], \\ k_R &= -L + \frac{(1-P)(1-L)}{R(1-R)}\,N\left[1+\frac{N}{R}+\mathcal{O}\left(\frac{N^2}{R^2}\right)\right], \\ k_N &= -(1-L)\left\{P[1+N+\mathcal{O}(N^2)] + (1-P)[1+N+\mathcal{O}(N^2)]\left[1+\frac{N}{R}+\mathcal{O}\left(\frac{N^2}{R^2}\right)\right]\right\}. \end{aligned}$$

Therefore, with some algebra and retaining only up-to-linear terms in N or N/R, we get

$$\begin{aligned} c(t) &\approx 1 + N(1-L)\frac{P-R}{1-R}e^{-t/\tau} + \left[-L + \frac{(1-P)(1-L)}{1-R}\frac{N}{R}\right]e^{-Rt/\tau} \\ &- (1-L)\left[1+N+\frac{N}{R}(1-P)\right]e^{-Nt/\tau}. \end{aligned} \tag{5.14}$$

Let us now move further and focus on the exponential terms. First, we notice that $1 > R > N$ and, consequently, we can neglect the term $e^{-t/\tau}$, as it decays much faster that both $e^{-Rt/\tau}$ and $e^{-Nt/\tau}$. Moreover, since the time range considered is ≈ 30 months and N is expected to be $\ll \tau/t \approx 10^{-1}$, we can expand $e^{-Nt/\tau}$ as $e^{-Nt/\tau} \approx 1 - Nt/\tau$, and considering only the leading dependence on t, we get

$$c(t) \approx 1 - (1-L)\left[1+N+\frac{N}{R}(1-P)\right] + \left[-L + \frac{(1-P)(1-L)}{1-R}\frac{N}{R}\right]e^{-Rt/\tau}. \tag{5.15}$$

As for $e^{-Rt/\tau}$, a similar approximation ($e^{-Rt/\tau} \approx 1 - Rt/\tau$) can be adopted as long as $R \lesssim 10^{-2}$, so to obtain the following linear approximation

$$c(t) \approx 1 + k_R(1 - Rt/\tau) + k_N. \tag{5.16}$$

Notice that a smaller (larger) estimate for τ would simply require a stricter (softer) condition on N and on R for the related linear expansions to hold (on the same time range); the model would not be affected and the parameters coupled with time, i.e. N, R, would be accordingly rescaled. As shown in Figure 5.3, the approximation (5.16) is rather good only for etanercept-treated patients, for which the best fit yields $R = 2.24 \cdot 10^{-2}$. On the other hand, if we consider infliximab-treated patients, the approximation (5.16) does not fit data, while using (5.15) we get a good overlap with data and the best fit yields $R = 2.12 \cdot 10^{-1}$, confirming that now Rt/τ is no longer small over the time window. All best fit coefficients are reported in Table 5.1.

We can estimate how sensitive $c(t)$ is with respect to the system parameters by deriving its analytic expression (see Eq. 5.12) with respect to N, P, R, L, respectively; in this way we get that, in the regime $N \ll (1, P, R, L)$, the most relevant parameter affecting the behavior of $c(t)$ is R. Another argument in favor of this claim is that, in the zero approximation of the solution (i.e. neglecting even terms $\mathcal{O}(N)$), P does not appear at all.

[4]Here and in the following we use the "big-O" Landau notation to characterize the growth rate of functions; more precisely, being $f(x)$ and $g(x)$ two arbitrary functions, we say $f(x) = \mathcal{O}(g(x))$ as $x \to 0$ if there exists a positive real number M such that $|f(x)| \leq M|g(x)|$.

In order to get further insight on the effect of infliximab and of etanercept on TB incidence, in Fig. 5.4 we plotted the model predictions for the percentage of patients having TB because of new infections ($p_4(t)$) or reactivations ($p_3(t)$).

To summarize, in the present context, we find that the most important difference between therapies based on infliximab or etanercept is that the former enhances TB reactivation more than the latter, in fact, we found $R_{INF} \sim 10\, R_{ETA}$, in agreement with [81]. Such a discrepancy implies even a qualitatively different behavior of $c(t)$ over the time-window considered: the number of infliximab-treated patients experiencing a TB infections grows exponentially in time, while for etanercept-treated patient the growth is linear.

However, as N is very small ($\sim 10^{-6}$) with respect to the experimental time-window (26 months), only the reactivation rate R can reasonably be inferred, while the measure of P, is affected by a much larger uncertainty, due to the lack of statistically influent number of new infections in the observational period. So possible effects due to differences in the rate of direct activation P can hardly be appreciated within the considered data set. In fact by plotting the solution and varying the order of magnitude of P (Figure 5.5(b) and 5.5(c)) and R (Figure 5.5(a)) it is possible to note the small effect of a coarse tune of P with respect to that of R. As a consequence the statement $P_{INF} \simeq P_{ETA}$ has to be claimed with more caution.

	TB-infection case: Best-fit coefficients.
L	$4.52 \cdot 10^{-4}$
N	$2.88 \cdot 10^{-6}$
$\mathbf{R}_{\mathrm{INF}}$	$2.12 \cdot 10^{-1}$
$\mathbf{P}_{\mathrm{INF}}$	$9.76 \cdot 10^{-1}$
$\mathbf{R}_{\mathrm{ETA}}$	$2.24 \cdot 10^{-2}$
$\mathbf{P}_{\mathrm{ETA}}$	$8.03 \cdot 10^{-1}$

Table 5.1. Tab. 1. TB-infection case: Best-fit coefficients. Best fit parameters obtained through the maximum likelihood method (see Eqs. 15, 16, respectively). Here we used $\tau = 1\, month$, consistently with clinical data. The fit was accomplished with the constraint that the parameters N and L are the same for both therapies, as they are drug-independent. The average relative error on these parameters is $\approx 9\%$.

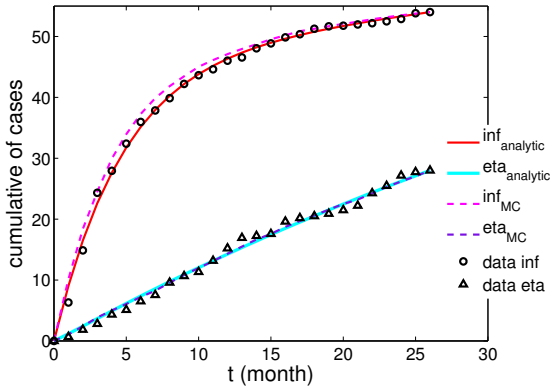


Figure 5.3. Comparison between experimental, analytical and numerical results for the TB-infection case. Cumulative number of patients undergoing active TB-infection. Experimental data from [81] (• for infliximab-treated patients and △ for etanercept-treated patients) are compared with the approximated analytical solution (see Eq. 5.15 and Eq. 5.16, respectively, solid curves) and with data from numerical simulation (dashed curves). The parameters used to draw the analytical curves correspond to the best-fit coefficients and are reported in Table 5.1. Notice that here we consider the extensive number of patients affected by TB over a population of $\mathcal{N} = 10^5$ treated patients, according to experimental results.

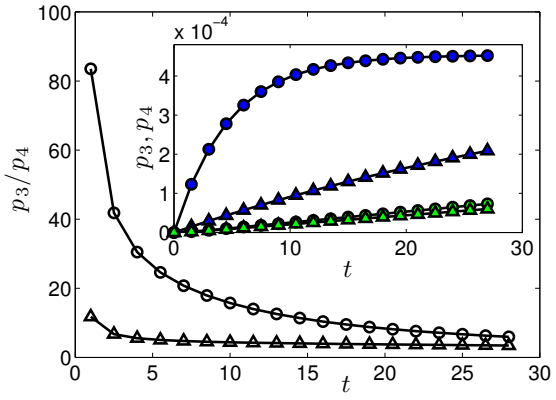


Figure 5.4. Probabilities p_3 and p_4 versus time for the TB-infection case. The probability p_3 of being in state 3 (TB reactivation) and p_4 of being in state 4 (Post 1° TB) are calculated from Eq. 5.7, using the parameters of Tab. 1. In the main plot we compare the ratio p_3/p_4 obtained from infliximab (•) and etanercept (△) parameters. In the inset, we depict each single probability, namely p_3 for infliximab (dark •) and for etanercept (dark △), p_4 for infliximab (bright •) and for etanercept (bright △); notice that the two sets of data for p_4 are partially overlapped. Lines are guide for the eye.

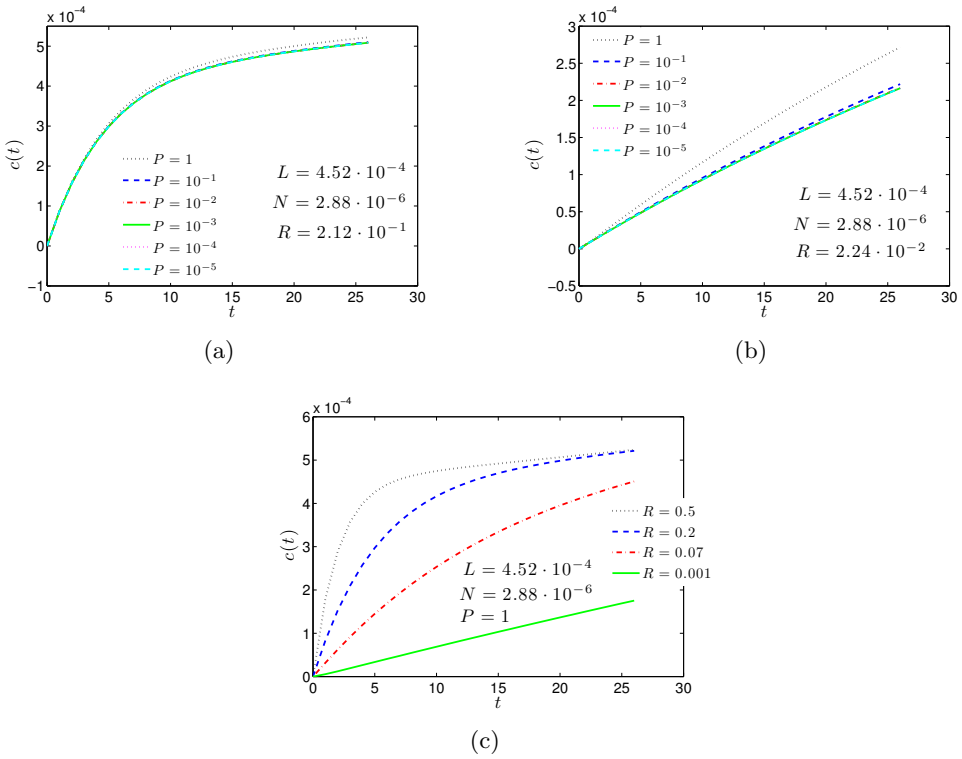


Figure 5.5. Time evolution of the complete solution $c(t)$ with the parameters fixed according to our best fits (see Table 5.1). Through the overlapping of the curves in panels (a) and (b) it is possible to note the mild role played by P.

5.3.2 The NTM-infection case

The mathematical model outlined in Section 5.2 can be applied to a general clinical scenario schematizable as in Figure 5.1. In the work presented in [3], the machinery has been applied to data reported in [103] and collected over the period $1999 - 2006$ where the insurgence of NTM infections are quantified. For details on the analysis of this data set we remand to [3]. Here we just report that the results seem to indicate that, for NTM, new infections are more important than reactivation and the difference between patients treated with infliximab and etanercept resides in the parameter P ($P_{INF} \sim 4P_{ETA}$).

5.4 Discussion

In this Chapter we formalized and extended a stochastic approach to data analysis (originally introduced in [88, 89, 81, 90]) for evidencing underlying correlations between adverse events and therapies based on immunosuppressants. In particular, the focus of our investigation concerns the risk of reactivation of latent mycobacterial infection in patients undergoing treatment with TNF-inhibitors.

We gave a clear and complete mathematical backbone to this approach, building it on explicit Markov processes, whose continuous-time limit yields the Master equation governing the evolution of the expected fraction of patients $c(t)$ exhibiting an active infection. We also solved the Master equation in all details finding an analytical expression for $c(t)$. Such mathematical developments make the original approach much more versatile and general: For instance, handling the complete (mathematical) solution allows to better account for reasonable approximations, tackling their control quantitatively (e.g. finding the proper timescales involved in the process or the integrals of motion constraining the evolution of the system). Furthermore, we can finally consider, within the same framework, different problems. In particular, we focused on TB and NTM infections emerging during anti-TNF therapies (infliximab and etanercept) according to data reported in [81, 103].

In the former case, we recovered previous findings showing that the rate of reactivation R of TB from a latent state to an active state plays a crucial role: being $R_{INF} \sim 10\, R_{ETA}$ we get qualitatively different behaviors for $c(t)$. More precisely, once fixed the observational time-window, for infliximab $c(t)$ grows exponentially with time, while for etanercept it grows linearly with time. Hence, these results sustain the need, for patients candidate to TNF blockers, to perform an accurate TB screening at baseline, irrespective of the type of antiTNF. Indeed, screening may decrease the risk of TB reactivation in such patients, while it is less clear what should be done to prevent NTM disease occurrence or progression in patients taking biologic agents. Importantly, for this purpose, we found that the comparison with experimental data allows to infer that reactivation of NTM infections plays a very minor role for both the therapies and that $c(t)$ grows quadratically with time.

We checked our results also against Monte Carlo simulation with excellent agreement.

Furthermore, our results are all consistent with recent experimental data and seem to indicate that TB and NTM infections are sustained by different pathogenetic mechanisms.

Non-tuberculous Mycobacteria are present in large numbers in the environment, including fresh water, aerosols, biofilms, and soils [110]. There are thus many opportunities for acquisition of NTM infection during ordinary daily activities, although the true incidence is not known. In contrast, nearly all transmission of Mycobacterium tuberculosis infection results from inhalation cough-generated aerosols from persons with active pulmonary TB. The annual risk of TB infection (ARTI) can be calculated from age-specific rates of tuberculin skin test reactivity; in most instances it is directly related to TB prevalence. Thus although the ARTI may reach 4% in highly TB-endemic regions such as South Africa, it is as low as 0.01% in much of Northern Europe and North America [111, 112, 113]. These epidemiologic findings are consistent with the results of our mathematical model, and underscore the interplay of microbial and host biology in determining the relative contributions of reinfection and reactivation to mycobacterial pathogenesis.

Hence, while the screening for TB is necessary prior to initiating biologics, when considering NTM only a watchful monitoring during the treatment is recommended. This finding is particularly relevant, since it allows to avoid screening for NTM infection, which is complicated by the poor sensitivity of chest radiograph and more expensive and invasive techniques, such as chest computed tomography scan and/or bronchoscopy, should be used.

It is worth stressing that this methodology, being based on very standard stochastic procedures, has the advantage to hold beyond the test case of Mycobacteria. We hope that this test-case may shed light to future developments of this sideline approach in figuring out adverse events of biological therapies.

Chapter 6

A Rep-Seq data analysis through Multivariate Gaussian Modeling for the study of Abs affinity maturation: preliminary results

In this Chapter we report some ideas on the analysis of an antibodies repertoire through multiparameter inference methods. The results reported here are still not published at the moment of the submission of this Thesis and we underline that they are meant as preliminary.

6.1 Multivariate Gaussian Modeling for protein families

A central problem in structural biology is that of predicting the tridimensional structure of a protein from the only knowledge of its amino acid sequence. A general recipe for the approach of the problem is still far to be found [114]. Nevertheless several strategies have been followed in order to restrict the problem. For example in [115], a phenomenological model for the folding of the primary structure is solved through replica method. Here we focus on the research field that deals with the statistical study of correlated substitutions within *multiple sequence alignments* (MSAs) of sequences classified in protein families.

A protein family is defined as a group of evolutionary related proteins that share a common ancestor. Protein in a family are said to be *homologous* and usually have the same functional role in different species. So, despite the differences in sequence, that can emerge from evolutionary noise, homologous proteins have, in general, rather convergent tridimensional structures.

The main idea beyond correlation analysis is that is that correlation patterns can be related to structural ones. In particular contacts between pairs of residues in the native structure could be related to the presence of strongly correlated substitutions

in two columns in the MSA of the protein family. Quoting [116], "the basic hypothesis connecting correlated substitution patterns and residue–residue contacts is very simple: If two residues of a protein or a pair of interacting proteins form a contact, a destabilizing amino acid substitution at one position is expected to be compensated by a substitution of the other position over the evolutionary timescale, in order for the residue pair to maintain attractive interaction".

The first attempts to use simple covariance analysis to predict residue-residue contacts ([117], [118]) partially succeeded in identifying some of them but presented a high false positive rate. As often happens in correlation analysis, the main reason beyond this mediocre performance is related to the difficulties in disentangling correlations caused by direct interactions from those produced by indirect ones.

Strategies to overcome the above cited problem have been developed. In particular a method called *Direct Coupling Analysis* (DCA) has been proposed in [119] and [116]. The main idea beyond DCA is that of inferring a probabilistic graphical model from the alignments and so to analyze the inferred interactions instead of the empirical correlations, so disentangling direct and indirect contributions to covariances. In the above cited works the choice of the probabilistic model to infer is based on *maximum entropy principle* [120], according to whom the least constrained model that reproduce single and pair sites amino acids frequencies in the alignment is the Inverse Potts model. This model allows to define a scalar quantity called *Direct Information* (DI) that is characteristic of every pair of positions in the alignment and that is correlated to the magnitude of the inferred interaction between them. High DI couples are in fact a good predictor of native contacts.

In [119] a message passing approach is used in order to solve the inference problem paying the price of slow computational times; in [116] the inference is performed within a mean field approximation that would be exact only in the limit of very weak interactions. Both works represented a breakthrough for the performances of contact predictions based on residues coevolution.

Starting from the ideas of the above cited works, some of the authors propose an even simpler probabilistic approach to the problem. That consists of identifying the relevant (real) variables with the fraction of times a given amino acid is present in a given position of an alignment of homologous proteins and to hypothesize that they fluctuate according to a simple multivariate gaussian distribution. This (strong) assumption permits to completely solve the inference problem by exact analytical computations. This method, that will be soon published in [121] under the name of *Multivariate Gaussian Modeling* (MGM), allows to achieve competitive performances for contact prediction. The fidelity of the method is highlighted in Figure 6.1 where the average *True Positives rate* (TP-rate, or specificity) of the Direct Information computed with both MGM and mean field Inverse Potts model is compared with that of the Mutual Information (MI) that represents a naïve pairs covariance analysis.

In the next Section we expose the mathematical details of MGM [122].

6.1.1 The mathematical method

In a MSA of P, sequences (whose length after the alignment is indicated with N) are formed by the 20 letters coding for the different amino acids, and may contain alignment gaps ("−"), such that the total alphabet size is $q = 21$.

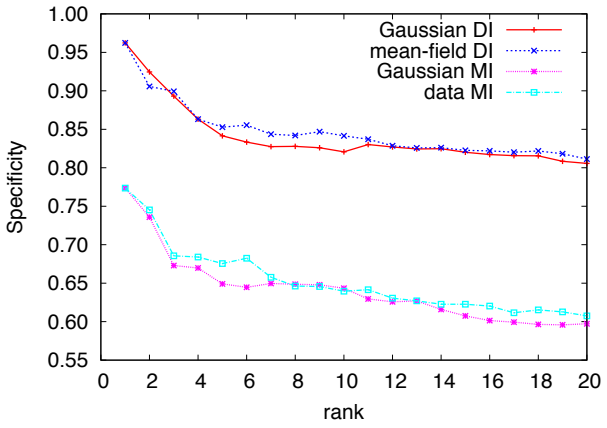


Figure 6.1. DI-ranking-based sensitivity-specificity curves (blue and red) and MI-ranking-based curves (light blue and magenta), averaged over 53 different protein families. The specificity as a function of the rank is defined as the fraction of true positive over the pairs with rank smaller than or equal to the indicated one. Gaussian and mean-field Potts method are found, up to small fluctuations, to perform equally well, with a significant and systematic improvement over MI-based methods. Figure from [121].

A MSA can be mapped in a binary data set composed by a $P \times N \times (q-1)$ object $(x_i^a)^\mu$ where the subscript $i \in \{1, \ldots, N\}$ runs over different amino acid residues, the superscript $a \in \{1, \ldots, q-1\}$, runs over different symbols and the superscript $\mu \in \{1, \ldots, P\}$ runs over different sequences. For example, if for protein μ, site i displays the $a-th$ letter, then $\vec{x}_i^\mu = (0, \ldots, 0, 1, 0, \ldots, 0)$ will be the unitary vector having only the $a-th$ non null component equal to 1. As in a MSA in every position a symbol (letter or *dash*) is present, while encoding it in a vector $\vec{x}_i^\mu$ of length $q = 21$, so one symbol is can be eliminated and the vector of length $q-1$ $\vec{x}_i^\mu = (0, \ldots, 0)$ will indicate a *dash* in the position i of the $\mu - th$ aligned sequence. The symbol that is eliminated while passing from the MSA to the x representation is arbitrary and we choose conventionally it to be the *dash*. We will refer to this freedom in the parametrization of the data as *gauge invariance*.

If the variable x are promoted to be real numbers ($x \in \mathbb{R}$), the assuming that the the sequences $(x_i^a)^\mu$ are drawn from a multivariate gaussian distribution, the log-likelihood of our model parameters $\{J, H\}$ given the data can be written as:

$$\begin{aligned}
\mathcal{L}(\{J,H\}|\text{DATA}) &= \sum_{\mu=1}^{P}\left[-\frac{1}{2}\sum_{ij}\vec{x}_i^{\mu}\cdot\hat{J}_{ij}\cdot\vec{x}_j^{\mu}+\sum_i\vec{H}_i\cdot\vec{x}_i^{\mu}-\log Z(J,H)\right] \\
&= \sum_{\mu=1}^{P}\left[-\frac{1}{2}\sum_{i,j,a,b}(x_i^a)^{\mu}\,J_{ij}^{ab}\left(x_j^b\right)^{\mu}+\sum_{i,a}H_i^a\,(x_i^a)^{\mu}-\log Z(J,H)\right] \\
&= P\left[-\frac{1}{2}\sum_{i,j,a,b}J_{ij}^{ab}\langle x_i^a x_j^b\rangle_{\mathrm{D}}+\sum_{i,a}H_i^a\langle x_i^a\rangle_{\mathrm{D}}+\right. \\
&\quad \left.-\langle\log Z(J,H)\rangle_{\mathrm{D}}\right] \qquad (6.1)
\end{aligned}$$

$$\begin{aligned}
Z(J,H) &= \int\prod_{i,a=1}^{N}dx_i^a\exp\left[-\frac{1}{2}\sum_{i,j,a,b}J_{ij}^{ab}x_i^a x_j^b+\sum_{i,a}H_i^a x_i^a\right] \\
&= \exp\left[-\frac{1}{2}\log\det J+\frac{1}{2}\sum_{i,j,a,b}H_i^a\left(J^{-1}\right)_{ij}^{ab}H_j^b+const\right], \qquad (6.2)
\end{aligned}$$

where we show both vectorial and component-wise notation and $\langle O\rangle_D := \frac{1}{P}\sum_{\mu=1}^{P}O^{\mu}$ is the empirical average of a generic observable O over the different experiments. Upon combining together Equations 6.1 and 6.2 we finally obtain up to irrelevant constants:

$$\begin{aligned}
\mathcal{L}(\{J,H\}|\text{DATA}) &= P\Bigg[\sum_{i,j,a,b}-\frac{J_{ij}^{ab}}{2}\langle x_i^a x_j^b\rangle_{\mathrm{D}}+\sum_{i,a}H_i^a\langle x_i^a\rangle_{\mathrm{D}} \\
&\quad -\sum_{i,j,a,b}\frac{H_i^a\left(J^{-1}\right)_{ij}^{ab}H_j^b}{2}+\frac{1}{2}\log\det J\Bigg] \qquad (6.3)
\end{aligned}$$

Maximum likelihood estimation of the parameters

Let us define the usual covariance matrix as

$$C_{ij}^{ab}=\langle x_i^a x_j^b\rangle_D-\langle x_i^a\rangle_D\langle x_j^b\rangle_D\,. \qquad (6.4)$$

We note that, while $\langle x_i^a\rangle_D$ represents the fraction of times the $a-th$ symbol is present in position i in the MSA, while the second one gives informations about the correlated variation of residues in the same $(i=j)$ or different $(i\neq j)$ positions.

We are now ready to maximize component-wise the log-likelihood $\mathcal{L}$ with respect to the model parameters J and H:

$$0 = -\frac{\partial\mathcal{L}}{\partial J_{ij}^{ab}}=\langle x_i^a x_j^b\rangle_D-\sum_{k,l,c,d}H_k^c\left(J^{-1}\right)_{ki}^{ca}\left(J^{-1}\right)_{jl}^{bd}H_l^d-J_{ij}^{-1}\,, \qquad (6.5)$$

$$0 = \frac{\partial\mathcal{L}}{\partial H_i^a}=\langle x_i^a\rangle_{\mathrm{D}}-\sum_{j,b}\left(J^{-1}\right)_{ij}^{ab}H_j^b\,. \qquad (6.6)$$

After inserting Equations 6.5 into Eq. 6.6 we get the final relation that gives the maximum likelihood (ML) estimation of the parameter J and H as a function of the data: Defining the usual empirical covariance matrix as

$$C_{ij}^{ab} = \langle x_i^a x_j^b \rangle_D - \langle x_i^a \rangle_D \langle x_j^b \rangle_D \,, \tag{6.7}$$

we get

$$J_{ij}^{ab} = (C^{-1})_{ij}^{ab} \,, \tag{6.8}$$

$$H_i^a = \sum_{j,b} J_{ij}^{ab} \langle x_j^b \rangle \,. \tag{6.9}$$

Direct Information

Once the parameters of the model have been inferred, one scalar quantity for every couple of positions i, j in the alignment can be determined in order to quantify the interaction strength of the sites. With this aim the Direct Information (DI) has been developed in order to be invariant under gauge transformations.

In order to do that, in the same spirit of what we have done in the previous sections, one could try to infer the ML model in a simplified setting where no interaction is present between variables. In this latter case the analogous of Equation 6.1 would be

$$\begin{aligned} \mathcal{L}(\{K,L\}|\text{DATA}) &= P\left[-\frac{1}{2}\sum_i \langle \vec{x}_i \cdot \hat{K}_i \cdot \vec{x}_i \rangle_D + \sum_i \vec{L}_i \cdot \langle \vec{x}_i \rangle_D - \langle \log Z(K,L) \rangle_D\right] = \\ &= P\left[-\frac{1}{2}\sum_{i,a,b} K_i^{ab} \langle x_i^a x_j^b \rangle + \sum_{i,a} L_i^a \langle x_i^a \rangle_D - \langle \log Z(K,L) \rangle_D\right] \end{aligned} \tag{6.10}$$

where the model parameter K is the analogous of J and L of H, and from the functional structure of the Likelihood function. the statistical independence among the different proteins is evident. After some simple algebra one gets that the ML estimation for $\{K, L\}$ is given by

$$\left(K^{-1}\right)_i^{ab} = \langle x_i^a x_i^b \rangle_D - \langle x_i^a \rangle_D \langle x_i^b \rangle_D \,, \tag{6.11}$$

$$L_i^a = \sum_b K_i^{ab} \langle x_i^b \rangle_D \,, \tag{6.12}$$

which are the analogous of Equations 6.8 and 6.9 respectively. Given the ML estimators K, L we can define the probability of a protein state x given K, L as

$$P^{\text{ind}}(x|K,L) \propto \prod_i \exp\left(-\frac{1}{2}\vec{x}_i \cdot \hat{K}_i \cdot \vec{x}_i + \hat{L}_i \cdot \vec{x}_i\right). \tag{6.13}$$

In a nutshell the idea is to compare an effective two-interacting-sites model whose interaction matrix is the J computed in Equation 6.8 with P^{ind}. Let us define the

probability distribution of the effective two-interacting-sites model as

$$
\begin{aligned}
&P_{ij}^{\rm dir}(\vec{x}_i,\vec{x}_j|\hat{J}_{ij},\hat{W}_i,\hat{W}_j,\vec{G}_i,\vec{G}_j) \propto \qquad (6.14)\\
&\propto \exp\left(-\vec{x}_i\cdot\hat{J}_{ij}\cdot\vec{x}_j-\frac{1}{2}(\vec{x}_i\cdot\hat{W}_i\cdot\vec{x}_i+\vec{x}_j\cdot\hat{W}_j\cdot\vec{x}_j)+(\vec{G}_i\cdot\vec{x}_i+\vec{G}_j\cdot\vec{x}_j)\right)\\
&\propto \exp\left(-\sum_{ab}J_{ij}^{ab}x_i^a x_j^b-\frac{1}{2}\sum_{ab}(W_i^{ab}x_i^a x_i^b+W_j^{ab}x_j^a x_j^b)+\sum_a(G_i^a x_i^a+G_j^a x_j^a)\right),
\end{aligned}
$$

where the parameters $\hat{W}_i,\vec{G}_i$ are chosen such that the single sites marginal of $P^{\rm dir}$ match the single sites probabilities $P^{\rm ind}$. We do this in two steps by imposing: (i) the equality of the connected second moment, (ii) the equality of the first moment. Let us note that the second connected component in a gaussian measure is just the inverse of the covariance matrix:

$$
\Sigma_{dir}:=\langle x_i^a x_j^b\rangle_{dir}-\langle x_i^a\rangle_{dir}\langle x_j^b\rangle_{dir} = \begin{pmatrix}\hat{W}_i & \hat{J}_{ij}\\ \hat{J}'_{ij} & \hat{W}_j\end{pmatrix}^{-1} \qquad (6.15)
$$

$$
\Sigma_{ind}=\langle x_i^a x_j^b\rangle_{ind}-\langle x_i^a\rangle_{ind}\langle x_j^b\rangle_{ind} = \begin{pmatrix}\hat{K}_i & 0\\ 0 & \hat{K}_j\end{pmatrix}^{-1}=\begin{pmatrix}\hat{K}_i^{-1} & 0\\ 0 & \hat{K}_j^{-1}\end{pmatrix} \qquad (6.16)
$$

note that in this representation Σ_{dir} and Σ_{ind} are $2(q-1)\times 2(q-1)$ matrices. We can now impose the equality of the diagonal part of the two matrices:

$$
\hat{K}_j^{-1} = \left(\hat{W}_j-\hat{J}'_{ij}\cdot\hat{W}_i^{-1}\cdot\hat{J}_{ij}\right)^{-1}, \qquad (6.17)
$$

$$
\hat{K}_i^{-1} = \left(\hat{W}_i-\hat{J}_{ij}\cdot\hat{W}_j^{-1}\cdot\hat{J}'_{ij}\right)^{-1}. \qquad (6.18)
$$

Such equations can be decoupled w.r.t. the variables $\hat{W}_i$ e $\hat{W}_j$:

$$
\hat{W}_j = \hat{K}_j+\hat{J}'_{ij}\left(\hat{K}_i+\hat{J}_{ij}\hat{W}_j^{-1}\hat{J}'_{ij}\right)^{-1}\hat{J}_{ij}, \qquad (6.19)
$$

$$
\hat{W}_i = \hat{K}_i+\hat{J}_{ij}\left(\hat{K}_j+\hat{J}'_{ij}\hat{W}_i^{-1}\hat{J}_{ij}\right)^{-1}\hat{J}'_{ij}; \qquad (6.20)
$$

and, after some manipulation, one obtains two matricial equations:

$$
\begin{aligned}
&A^2-A-M=0 \qquad (6.21)\\
&\qquad A = \hat{K}_i^{-1}\hat{W}_i\\
&\qquad M = \hat{K}_i^{-1}\hat{J}_{ij}\hat{K}_j^{-1}\hat{J}'_{ij},\\
&B^2-B-N=0 \qquad (6.22)\\
&\qquad B = \hat{K}_j^{-1}\hat{W}_j\\
&\qquad N = \hat{K}_j^{-1}\hat{J}'_{ij}\hat{K}_i^{-1}\hat{J}_{ij}.
\end{aligned}
$$

Solutions to Eq. 6.21 and Eq. 6.22 can be obtained by diagonalizing matrices M and N and solving $2(q-1)$ independent equations. We are now ready to express

the Direct Information as the Kullback-Leibler (KL) divergence of the two gaussian measures P^{dir} and P^{ind}:

$$\begin{aligned} DI(J,K)_{ij} &\equiv KL(P^{dir}||P^{ind}) = \\ &= \frac{1}{2}\Big\{\mathrm{tr}(\Sigma_{ind}^{-1}\Sigma_{dir}) + (\langle X\rangle_{ind} - \langle X\rangle_{dir})^T\Sigma_{ind}^{-1}(\langle X\rangle_{ind} - \langle X\rangle_{dir}) + \\ &\quad - \log\left(\frac{\det\Sigma_{dir}}{\det\Sigma_{ind}}\right) - 2(q-1)\Big\} \\ &= -\frac{1}{2}\log\left(\frac{\det\Sigma_{dir}}{\det\Sigma_{ind}}\right) \end{aligned} \tag{6.23}$$

since in our case $\langle X\rangle_{ind} = \langle X\rangle_{dir}$ and $\mathrm{tr}(\Sigma_{ind}^{-1}\Sigma_{dir}) = 2(q-1)$.

Data regularization (pseudocounts)

For the maximum likelihood estimations 6.8 and 6.9 of the parameters to be possible, the covariance matrix needs to be full rank in order to be inverted. As fluctuation in MSAs are generally limited, usually the experimental covariance matrix is rank deficient. To overcome this problem a regularization procedure has to be implemented. The simplest method for that is that of adding to the sample a number λ of fictitious sequences in which symbols in every site are fairly drawn from a flat distribution. This reduces to manipulate the data as:

$$\langle x_i^a\rangle_D \longrightarrow (1-\pi)\langle x_i^a\rangle_D + \pi\frac{1}{q}\,, \tag{6.24}$$

$$\langle x_i^a x_j^b\rangle_D \longrightarrow (1-\pi)\langle x_i^a x_j^b\rangle_D + \pi\frac{1}{q^2} \tag{6.25}$$

where the parameter

$$\pi \equiv \frac{\lambda}{P+\lambda}\,, \tag{6.26}$$

that is referred to as *pseudocount* parameter, naturally interpolates between the empirical ($\pi = 0$) and completely random ($\pi = 1$) data.

In [121] it is clarified how the use of a pseudocount based regularization in MGM is equivalent to the choice of a normal–inverse–Wishart prior (the conjugate prior of the multivariate gaussian distribution) over the parameters of the MGM.

6.1.2 Multivariate Gaussian Modeling for antibodies diversity: the general idea

The possibility of studying the features of the affinity (or of the neutralization power) of antibodies directed towards a certain antigen as a function of the sequence of the variable region is a fundamental issue in structural immunology.

Unfortunately, the size of the available sets of antibodies, for which both the sequence and the neutralization power toward an antigen is known, is, at the moment, as large as some tenth of antibodies. The typical size of an antibody's variable region is of order $N \sim 10^2$ amino acids; as the contributions of the amino acids to the neutralization power are not independent, the least structured function to describe it

would contain at least $(Nq)^2 \sim 10^6$ parameters. So up to now there is no possibility to estimate the parameters of such a function with the size of the available data sets.

Nevertheless the recent developing of sequencing techniques (*Deep Sequencing*, *Next Generation Sequencing*), that are able to produce in parallel up to $\sim 10^6$ sequence reads, have open the possibility of extensive experimental studies of the Abs repertoires in different living organisms. This kind of research is known under the the name of Rep-Seq (*Repertoire Sequencing*) experiments (see [123] for a review on the argument). These techniques have permitted for example to study the complete Igs repertoire of simple organisms such as the zebrafish, whose immune system has only ~ 300000 Abs producing B cells, hence determining its complete *antibodyome* (see [124] and [125] for the maximum entropy analysis of the data). In human the latter amount to $\sim 10^{9-10}$ cells, so that only limited samples of the entire repertoires are available up to now (see for example [126] for Rep-Seq experiment of Igs in human or [127] for a maximum entropy analysis of TCR repertoire in human).

As explained in more detail in Chapter 2, the features of the populations of B cells in host bodies are determined by the processes of genetic recombination, negative selection that eliminate self-directed antibodies, clonal expansion in response to different antigens that are or were present in the host body, random mutations (somatic hypermutation) and positive selection of antibodies that have an high affinity to antigens (affinity maturation). All these mechanisms interact in a complex manner to determine the B cell population present in a body.

In some cases it should be possible to manipulate Rep-Seq data sets in order to disentangle the processes described above with the aim of obtaining a sample whose evolution is mainly driven by affinity maturation and clonal expansion toward a specific antigen so that the resulting data set is highly correlated with the affinity (or neutralization power) of the antibodies. Once that this step is fulfilled, a probability distribution that is considered to have generated the sample can be inferred. If this probability distribution is indeed observed to be correlated with the affinity (or neutralization power), then it could be used as a proxy to study the features of the neutralization power as a function of the sequences and in principle to propose sequences of high neutralization power.

Moreover, as explained in Chapter 2, the affinity maturation is an evolutionary process in which different B cell clones compete for the antigen in the germinal centers. The study of the statistics of the population of sequences could also unveil interesting features of the fitness function in the space of sequences related to this process and in general of the evolutionary dynamics of the B cells. That could be of interest for a statistical population genetics analysis of the affinity maturation process, for example in the spirit of [128].

Within the above described scenario MGM is a useful tool as it interprets the x representation of the MSA as a (discrete) sampling from a gaussian distribution whose parameters can be inferred following the procedure outlined above. This distribution over the real x variable is of the form

$$P(x|J,H) = \frac{1}{Z(J,H)} e^{-E(x|J,H)} \,, \tag{6.27}$$

where the *energy* of the model is defined as

$$E(x|J,H) = -\frac{1}{2}\sum_{ij} \vec{x}_i^{\mu} \cdot \hat{J}_{ij} \cdot \vec{x}_j^{\mu} + \sum_i \vec{H}_i \cdot \vec{x}_i^{\mu} \tag{6.28}$$

which is the log-likelihood apart from the normalization factor $Z(J,H)$ defined in Equation (6.2).

6.2 Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing: a review of the experimental work

As exposed in Section 6.1.2, the idea of using a probability distribution inferred over a Rep-Seq dataset as a proxy for the affinity function could be tested in a case in which the Rep-Seq experiment is performed together with affinity measurements. This is the case of the work that has been published in [129]. We have used data from there to test the above exposed idea and the preliminary results are encouraging.

In this Section the experimental work is reviewed, while the preliminary results of our analysis are exposed in Section 6.3.

HIV-1 displays an enormous genetic diversity and in this resides a great part of the infection's strength. Despite this fact, from 10% to 25% of the patients develop cross-reactive neutralizing antibodies after several years of infection. These individuals are said to have a broadly neutralizing serum.

In a previous work [130] the authors of [129] isolated VRC01, VRC02 and VRC03, three similar broadly neutralizing antibodies (bnAbs) from a patient (donor 45) presenting a broadly neutralizing serum. This *VRC01-like Abs* have been seen to be bind gp120, a membrane glycoprotein used by the virus to attach the CD4 receptor on T lymphocytes, HIV target cells. To avoid neutralization by gp120 directed Abs, during evolution, the virus has developed a complex structure for gp120: Highly variable domains hide the site of attachment gp120-CD4 which is the only part of the protein that is under evolutionary pressure and that, for this reason, has relativley low freedom to mutate. There are structural experimental evidences that VRC01-like bind exactly this particular site on gp120 ([130] and [129]).

VRC01 neutralizes 90% of virus isolates with an average neutralization power $\mathrm{IC}_{50} \sim 0.3\,\mu$g/ml while its predicted unmutated germline ancsestor has low affinity for the antigen (dissociation constant in the millimolar range). All VRC01-like Abs display an high level of mutation ($\sim 30\%$ that has to be compared with the 5-15% of mutations in average Abs) from the inferred germilene. This underline that, for this kind of antibodies, the improvement of the neutralization power due to affinity maturation is substantial. This fact lead us to the idea that this could be a good system to study the affinity maturation process.

In [129] the authors isolated other VRC01-like antibodies from another donor (donor 74). An example of them is VRC-PG04. Couples of such bnAbs with the same unmutated germiline ancestor, from which they are mutated at about 30% and that come from different patients (for example VRC-PG04 from donor 74 and

VRC-01 from donor 45), are observed to be very different in sequence ($\sim 50\%$); this underline the fact that the (relatively) optimal sequence to bind the gp120-CD4 attachment site is not unique and that the evolutionary histories of populations of antibodies under analogous evolutionary pressures may be very different.

The authors performed a crystallographic study on VRC-PG04[1] in complex with the gp120. Comparing this with analogous studies on others VRC01-like Abs they argue that the most important region for the neutralization are the CDR2 on the heavy chain and the CDR3 on the light chain of the Ab.

All these facts underline that VRC01-like antibodies are an interesting system to study both on the fundamental and practical level. Indeed the authors decided to perform Rep-Seq experiments on donor 45 and donor 74 blood samples with the aim of sampling a part of the antibodies repertoires on these donors depending on the choice of the primers.

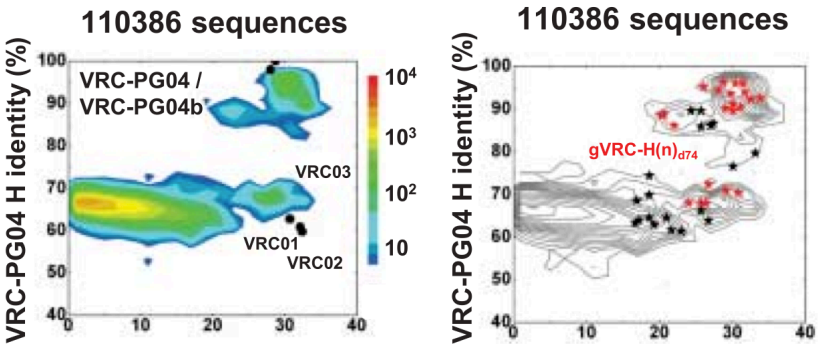
The authors choose 454-pyrosequencing as deep sequencing technique as it allow to sequence reads up to a length of 700 bp and so to sequence the whole variable region of both light and heavy chain. The side effect of this choice is that 454-pyrosequencing has a relatively high error rate (1/1000 bp on average); errors are concentrated on homopolymers.

Moreover, light and heavy chain are translated into different mRNAs molecules; as the sequencing technique capture the mRNA in the sample and mRNAs belonging to different cells are mixed during the procedure, it is only possible to reconstruct separately the light and heavy chain repertoire and there is no way to match the light and heavy chain belonging to the same antibody (B-cell clone).

Data of sequencing experiments on light and heavy chains for donor 45 and two experiments on heavy chains of donor 74 have been deposited to NCBI database. Reference to sequencing data can be found in the Acknowledgment of [129] or at http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?study=SRP006992.

The main results on the deep sequencing experiments are resumed in Figure 4 of [129], that we report here for clarity in Figure 6.2. The authors found out that, if heavy chains reads are classified by divergence from the inferred germline V gene and similarity with the nucleotide sequence coding for the heavy chain of VRC-PG04, they seem to form two clusters, one of which is closer to the already known VRC-PG04. Then authors measured the neutralization power of 45 successfully produced (out of 70 tried for production) chimeric Abs, in which VRC-PG04 light chain was coupled with heavy chains selected from the highly mutated ones in the sequenced set. The result of the neutralization measurements of 20 HIV-1 isolates, belonging to the clades A,B and C, is that heavy chains that are more similar to VRC-PG04 are in general prone to be (broadly) neutralizing (see Figure 6.2, Panel (b)), confirming that several VRC01-like antibodies are present in the sample and that they share some features with the known ones.

[1]Structure factors and coordinates for antibodies VRC03 and VRC-PG04 in complex with HIV-1 gp120 have been deposited with the Protein Data Bank under accession codes 3SE8 and 3SE9, respectively.



(a) (b)

Figure 6.2. Panel (a): Identity/divergence analysis of Rep-Seq experiment for heavy chains in donor 74 performed in [129]. On the horizontal axis the divergence from the IGHV1-2*02 germline gene of origin is reported while the identity to VRCPG-04 sequence is indicated on the vertical axis. Both quantities are computed at the nucleotides level. Panel (b): Neutralization power measurements on 45 Chimeric Abs whose heavy chain has been extracted from the Rep-Seq data set. Red stars correspond to Igs that neutralize at least one of the tested isolate viruses, while Black star indicate non-neutralizing Abs. Figure taken from [129]. See the publication for details.

6.3 The Multivariate Gaussian Modeling analysis: preliminary results

In the following we describe the analysis performed on Rep-Seq data on the heavy chain variable region of donor 74 from [129] . The work is still in progress and the result presented here are meant to be preliminary.

Donor 74 has been observed to produce the broadly neutralizing antibody VRC-PG04 that has IGHV1-2*02 and IGHJ2*01 as heavy chain V and J gene of origin (as the sequence presents as much as $\sim 30\%$ of mutation from the inferred germline, the D gene of origin is too short to be determined).

For our scopes the deposited raw data set is to be submitted to a bioinformatic analysis that will be described in details in [4]. Here we just underline that, starting form the one strand nucleotide sequences, our bioinformatic analysis selects the productive ones and returns a set of amino acid sequences each provided with its multiplicity, i.e. the number of times a nucleotide sequence coding for the same amino acid sequence is present in the set. We underline that, differently from [129], we only retain productive sequences, i.e. sequences for which the V and J genes are in frame and that do not present stop codons. This analysis generates a set 383267 productive amino acid sequences (which reduce to 191661 unique sequences) for the variable part of the heavy chain, provided with the inferred V and J gene of origin. According to the authors of [129], the sequencing primers have been chosen in such a way that, in principle, for all the sequenced reads, the inferred V gene of origin is in the family IGHV1.

Selecting the sequences whose V gene of origin is (one of the alleles of) IGHV1-2 gives a set of 72649 sequences (37839 unique), while the set of sequences that have IGHV1-2 and IGHJ2 as germline genes of origin consist of 6820 sequences (3258

unique).

It has to be kept in mind that sequences and and relative proportions of their abundances are strongly affected by sequencing errors and PCR biases. Moreover, despite the fact that the selection of the productive sequences reduces the number of sequences presenting errors, this procedure affects the relative proportions of sequences in the population. So we claim that systematic experimental errors are present and out of our control.

6.3.1 Clustering analysis

Broadly neutralizing antibodies individuated by Wu et al. in the blood sample of donor 74 are remarkably highly mutated from the inferred germline (IGHV1-2 and IGHJ2) . The identity/divergence analysis (see [129] and Figure 6.2) performed on the deep sequencing data suggests that sequences with inferred IGHV1-2 germline gene of origin (the same of VRC-PG04) should display the presence of a cluster of highly mutated sequences which is well separated from the cluster of typically mutated sequences; the same does not happens for Abs with a different IGHV inferred germline gene; IGHV1-2 related Abs of donor 74 seems thus to be a preferred system to study affinity maturation.

The effective presence of this clustering structure with a cluster more similar to the V and J germline genes and a highly mutated cluster more similar to the broadly neutralizing antibody VRC-PG04 was indeed verified though clustering algorithms as described in the following of this Section.

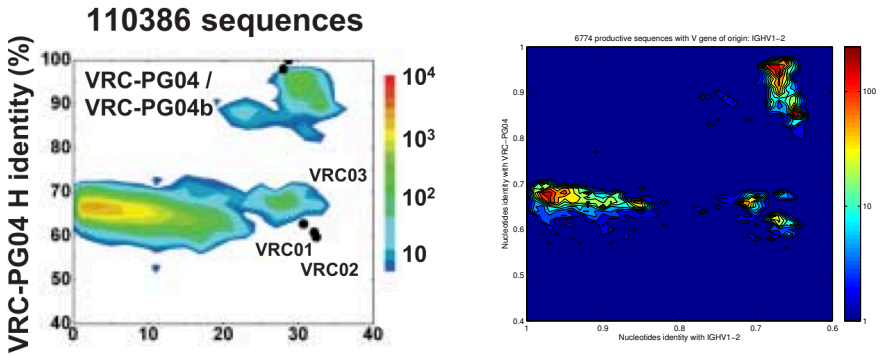
Identity/divergence analysis

The first check that needs to be done is that of performing the identity/divergence analysis on our set in which only productive sequences are retained. The result is shown in Figure 6.3 and it seems to confirm that the former analysis is coherent with the one performed in [129].

We also confirm that, as in [129], sequences that are more than 25% different form IGHV1-2 mainly have IGJV2 as inferred J gene of origin.

Interestingly almost all the sequences that display a divergence from the germline higher than 25% belong to the set that has IGHV1-2 and IGHJ2 as allelic origin. So, remarkably, also if VRC-PG04 had not been identified as a broadly neutralizing antibody, the identification of the signature of such a deep affinity maturation could make it possible to select the germline genes under evolutionary pressure only by divergence analysis.

The presence of a cluster of high divergence from the germline and high identity to VRC-PG04 suggests to deepen the clustering structure of the germline. That will be described in the next Section 6.3.1.



(a) Figure from [129]

(b) Figure from [4]

Figure 6.3. Comparison of the Identity/divergence analysis performed in [129] and [4]. See the caption of Figure 6.2 for explanation
Panel (a): Figure 6.2-(a) has been reported here for easier comparison.
Panel (b): Identity/divergence analysis performed in [4].

Clustering algorithms

Figure 6.3 shows that, if we describe the set by projecting it onto the two specific sequences (IGHV1-2*02 gene and of the heavy chain of VRC-PG04), sequences having IGHV1-2 as V gene in the inferred germline origin form at least two well separated clusters. Does this clustering configuration hold in the whole sequence space and not just after a proper projection? In order to check whether this is the case, the set of sequences with inferred germline in the families of IGHV1-2 and IGHJ2 has been submitted to a clustering algorithm that has been proposed in [131]. With this algorithm it is possible to determine the number of well defined clusters: In fact varying the parameter λ in the algorithm, the number of clusters in the output develops a plateau exactly when sequences are optimally separated. Figure 6.4 gives the indication that there are two well separated clusters.

Using the last information, we performed a clustering analysis on the same set of sequences (with IGHV1-2 and IGHJ2 allelic origin) by submitting different alignments to the Matlab algorithm `kmeans`, fixing at 2 the input parameter corresponding to the number of clusters. The consensus sequence of the two clusters were compared with the germline sequence (i.e. to the concatenated sequences of IGHV1-2*02 and IGHJ2*01 genes) and with the VRC-PG04 sequence. As shown in Figure 6.5, the consensus sequence of the normally mutated cluster is similar to the germline genes while for the highly mutated cluster the consensus sequence is similar to that of VRC-PG04. The size of the two obtained clusters are respectively of 3471 sequences (1874 unique) and 3349 sequences (1338 unique). In the following we will indicate with *clusterVJ* and *clusperPG04* respectively these two sets of sequences.

The different sets in which the sequences have been divided are resumed in Table 6.1.

set description	size	size (unique)
Productive IGHV1 origin	383267	191661
Productive IGHV1-2 origin	72649	37839
Productive IGHV1-2 and IGHJ2 origin	6820	3258
Productive IGHV1-2 and IGHJ2 origin - *clusterVJ*	3471	1874
Productive IGHV1-2 and IGHJ2 origin - *clusterPG04*	3349	1338

Table 6.1. Different sets in which sequences have been divided.

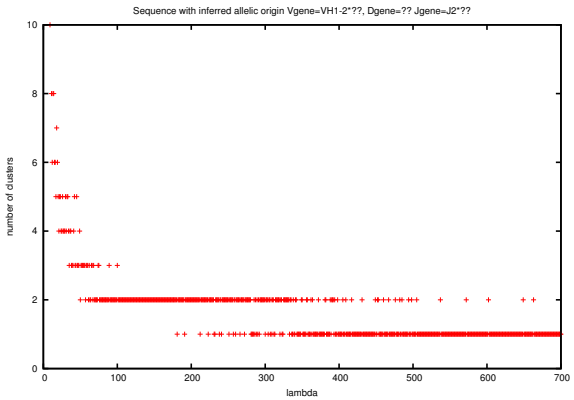


Figure 6.4. Number of clusters as a function of the parameter λ, see [131] for details. The plateau at 2 indicates that an optimal clustering is obtained with two clusters.

```
001  QAHLEQSGSGVKKPGASVRVSCWTSEDIFERTELIHWVRQAPGQGLEWIGWVKTVSGAVNFGS-
     |::| ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
001  QVQLVQSGSGVKKPGASVRVSCWTSEDIFERTELIHWVRQAPGQGLEWIGWVKTVTGAVNFGSP

064  DFR-RVSLTRDRDLFTAHMDIRGLTQGDTATYFCAR---Y--GQGWYFDLWGRGTL
     ||| |||||||||||||||||||||||||||||||    |  ||||||||||||||
065  DFRQRVSLTRDRDLFTAHMDIRGLTQGDTATYFCARQKFYTGGQGWYFDLWGRGTL


Identities = 101/110 (92%), Positives = 104/110 (95%)
001  QVQLVQSGAEVKKPGASVKVSCKASGYTF-TGYYMHWVRQAPGQGLEWMGWINPNSGGTNYAQK
     ||||||||: |||||||||::||| :|   | ||||||||||||||||||||||||||||||||
001  QVQLVQSGSGVKKPGASVRVSCWTSEDIFETGYYMHWVRQAPGQGLEWMGWINPNSGGTNYAQK

064  FQGRVTMTRDTSISTAYMELSRLRSDDTAVYYCARWYFDLWGRGTL
     |||||||||||||||||||||||||||||||||||||||||||||
065  FQGRVTMTRDTSISTAYMELSRLRSDDTAVYYCARWYFDLWGRGTL
```

Figure 6.5. Alignment of the consensus sequences of the two clusters with the sequences of VRC-PG04 (upper case) and the germline (lower case) .

6.3.2 Comparison between the inferred probability distribution and neutralization power measurements

As already pointed out, in [129] 70 sequenced heavy chains, mostly originating from IGHV1-2, were used to construct chimeric antibodies with the light chain of VRC-PG04. 45 of those have been successfully produced and tested for neutralization power against 20 HIV-1 virus belonging to clades A (6 viruses), B (8 viruses) and C (6 viruses). Although the neutralization power is a complex function of both light and heavy chain sequences (see [129], Figure 4A), these measurements can be used as an approximation of the contribution of the heavy chain to the neutralization power. The complexity of the interaction of light and heavy chain in determining the efficiency of an antibody is anyway an element to be kept in mind.

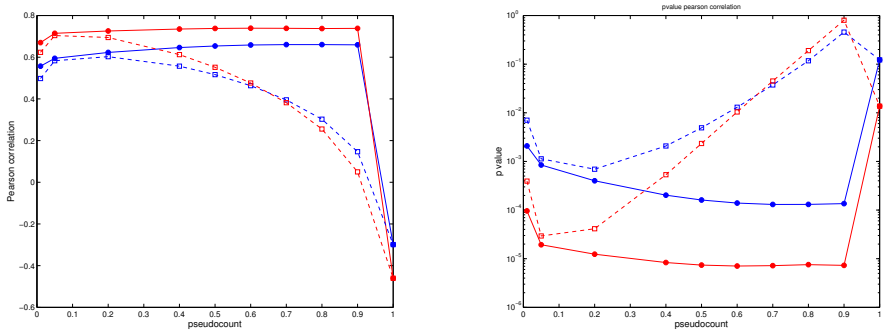
The above cited neutralization power measurements (the data relative to which are in Table S19 and S20 in the Supporting Information of [129]) can be compared with the statistical properties of the set. To do this, the highly mutated *clusterPG04* has been selected and a MGM has been inferred on this set considering each sequence with its relative multiplicity. For the 45 Abs that were tested for neutralization power, the IC_{50} has been compared with the energy of the inferred model. Inference performed with different values of the pseudocount parameter π display a significant Pearson correlation coefficient between the inferred energy and neutralization power of single viruses and with their (overall or over clade A viruses) average or minimum (See Figure 6.6 - full dots and continuous lines - and Figure 6.7). Note that values of IC_{50} that are reported in [129] as greater than 50 μg/ml are considered here to be equal to this value.

The above observation is very promising and it indicates that the MGM energy inferred on a proper Rep-Seq set could, in general, provide informations about the neutralization power.

A question that could arise after the last result is weather MGM is really needed to compare the repertoire structure with the neutralization power or if the latter is only for example simply a trivial function of some distance from the consensus sequence of the sample. To try to answer this question the above procedure has been repeated with a factorized MGM (in which non-diagonal J terms are set to zero). The result is that the energy inferred with the factorized model is less (and less significantly) correlated with the neutralization power then the energy learned with the complete correlated gaussian model (See Figure 6.6, empty squares and dashed lines).

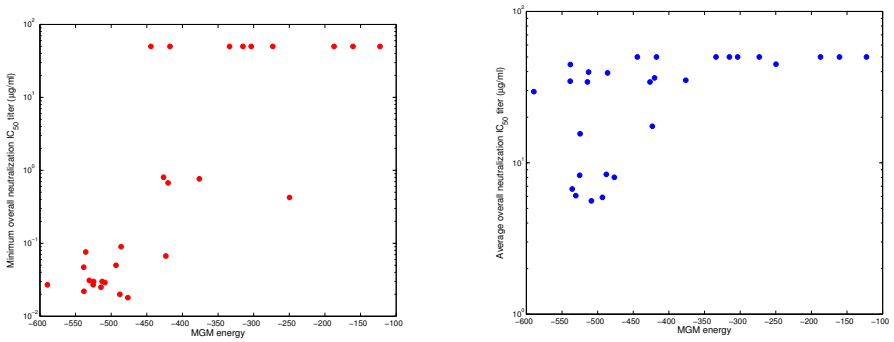
The message that should be learned out of the above result is that, as expected, not only single mutations but at least correlated pairs of mutation in the variable region sequence are needed in order to achieve affinity maturation. Special directions in the space of sequences (arising from combinations of the single aligned amino acids) that are more relevant for the affinity can be considered by learning a correlated MGM over the considered set.

The results reviewed in the current Section suggest to investigate the reasons for which the MGM seems to correlate the inferred probability distribution with the neutralization power. Which structural features of the Igs that are important for the neutralization power are captured by MGM? Next Sections try to answer this question.



(a) (b)

Figure 6.6. Panel (a): Scaling with the pseudocount of the Pearson correlation coefficient between the energy of the MGM learned on *clusterPG04* and the average (blue) and minimum (red) IC_{50} neutralization titer against the 20 tested HIV-1 viruses. Full dots and continuous lines correspond to a full MGM while empty squares and dashed lines to a factorized ($J \equiv 0$) MGM.
Panel (b): Scaling of the p-value of the Pearson correlation coefficient displayed in panel (a). The same symbolic conventions hold.
The complete MGM seems to significantly explain the neutralization power measurements. Consistently the correlation is not significant when $\pi = 1$, so when the model is learned on completely random data. Points with $\pi \sim 0$ refer to $\pi = 0.01$.



(a) (b)

Figure 6.7. An example of the comparison between the energy of the MGM inferred on *clusterPG04* and the average (Panel (a)) and minimum (Panel (b)) neutralization IC_{50} titer. In both showed examples the MGM is learned with pseudocount $\pi = 0.5$.

6.3.3 The antigen - heavy chain interaction

As the in the crystallographic structure the broadly neutralizing antibody VRC-PG04 is resolved in complex with the gp120 (its antigen) an interesting argument to investigate deals with the possibility of predicting some feature of the Ab-antigen binding by simply analyzing the population of sequences present in the sample.

Single site entropies and polarization change analysis

The simplest analysis that can be performed is that of the variability of the residues in the columns of the alignments. A measure of it are the entropies of the empirical distribution of residues in each each column in the alignment. The site entropies computed on *clusterPG04* have to be compared with the backround ones that encode the average variability of antibodies maturated from the same germline genes. The result are then analyzed for the residues that are aligned to those present in the PDB structure and with a focus on those that are considered to be in contact with gp120.

As a first try we approximated this background with the *clusterVJ*. As an alternative procedure, a library of the deposited antibodies sequences with inferred IGHV1-2 allelic origin has been constructed and assumed to represent the reference background of the IGHV part of the sequences. As shown in Figure **??**, although they display a certain level of incompatibility, both procedures give the convergent information that the amino acids in contact with the antigen have very different behavior with respect to the variability. Indeed some of the contact residues happens to be less variable in the mutated cluster than in the background. This is the case for Arg71, an amino acid which is observed to form a crucial interaction (salt bridge) with an Asp in the binding region of gp120. On the contrary other contact residues (most of which are in the CDR2 region) are observed to be more variable in the mutated cluster than in the background. An interpretation for that will be given in section 6.3.3.

The heterogeneity of the nature of the contacts between the antibody and the antigen does not seems to permit to individuate the contact residues by simply analyzing the single site entropies of the populations of antibodies.

We also analyzed the columns in the alignments distinguishing between the ones for which the most present residue changed from *clusterVJ* (or the background library) to *clusterPG04*. The results are shown in Figure **??** and show that most of the columns in the alignment that are considered to be in contact with the antigen have changed the most abundant amino acid within the formation of the highly mutated cluster. This is not anyway a strict constraint; Arg71 - that, as already highlighted, form a critical interaction with the antigen - is already present in the germline gene IGHV1-2 and is the most abundant residue in that position both in the background and in *clusterPG04*.

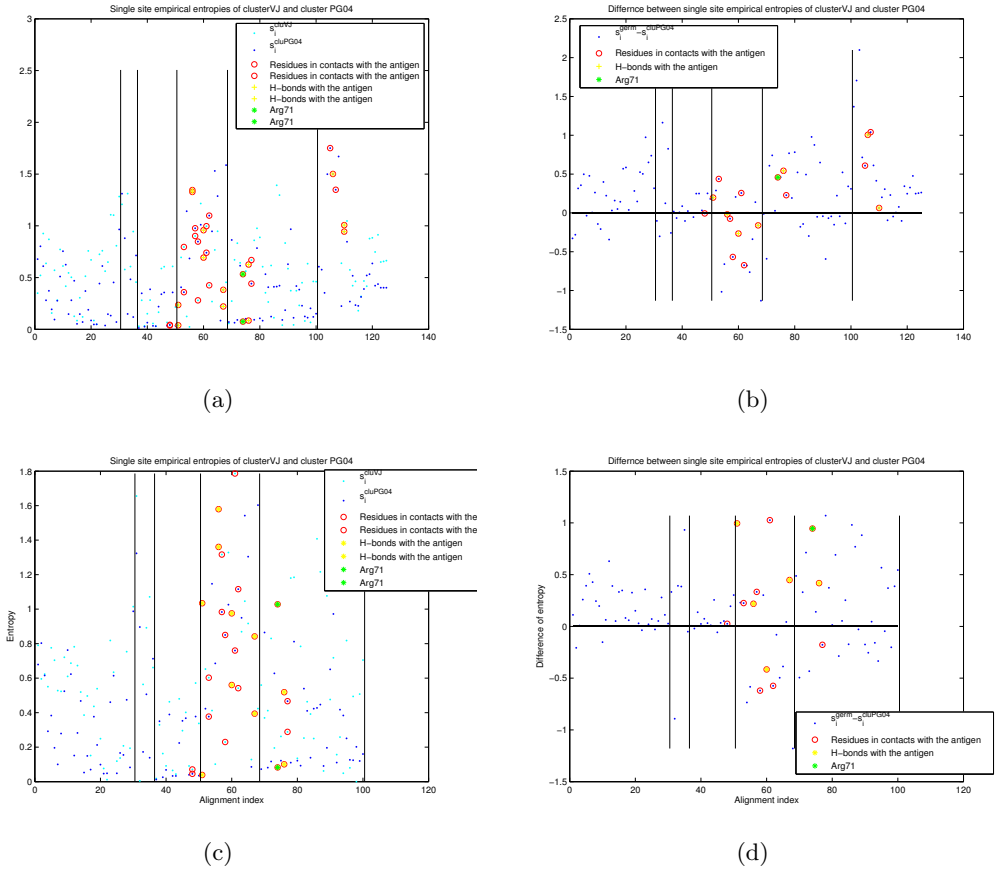


Figure 6.8. Panel (a): Single site empirical entropies for the *cluesterVJ* (light blue) and *clusterPG04* (dark blue).
Panel (b): Difference of the empirical single site entropies computed on *clusterVJ* and *clusterPG04*.
Panel (c) and (d): Same as in Panel (a) and (b) with the difference that the background entropies are computed on a library of deposited Abs sequences with IGHV1-2 inferred germline. In that case only the V region is displayed.
Entropies are displayed only for the columns of the alignment that do not display a dash in the sequence fo VRC-PG04. Highlighted points refer to columns in which the aligned residues of VRC-PG04 is in contact with the antigen. Vertical lines separate frameworks and complementary determining regions, following: FWR1, CDR1, FWR2, CDR2, FWR3, CDR3.

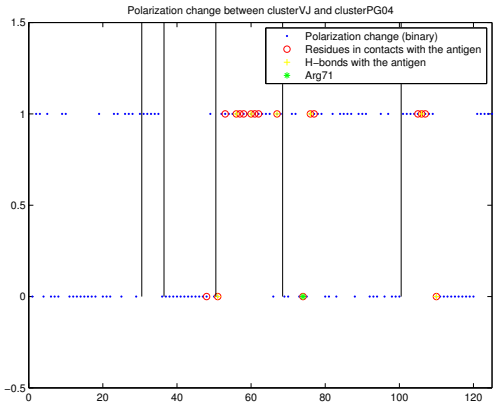


Figure 6.9. Residues are classified with a binary variable taking value 1 if the most abundant residue in column of the alignment changed from *clusterVJ* to *clusterPG04* and 0 otherwise. Vertical lines separate frameworks and complementary determining regions, following: FWR1, CDR1, FWR2, CDR2, FWR3, CDR3.

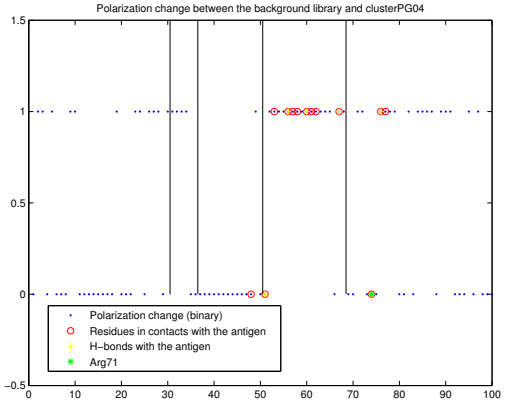


Figure 6.10. Residues are classified with a binary variable taking value 1 if the most abundant residue in column of the alignment changed from the background cluster to *clusterPG04* and 0 otherwise. Only IGHV region is considered. Vertical lines separate frameworks and complementary determining regions, following: FWR1, CDR1, FWR2, CDR2, FWR3, CDR3.

DCA analysis

The structure of the correlations between different columns in the MSA is another interesting feature of the populations that can be related to the details of the binding between the antibody and the antigen.

In fact sequences belonging to *clusterPG04* are peculiar for the emergence of a strong interaction pattern among a set of sites mostly belonging to the CDR2 region. The high DI signal in the CDR2 is more evident for high pseudocounts (see Figure 6.11).

Moreover, as shown in Figure 6.12, the high DI pattern of the CDR2 is characteristic of the highly mutated *clusterPG04* and is not present in *clusterVJ* that is more similar to the germline genes.

An explanation for this observation could relay in the fact that, in the tridimensional structure, the CDR2 loop appears to be shifted with respect to its typical position in order to permit the exposition of $\text{Arg71}_{\text{VRC-PG04}}$ that can so from the critical interaction $\text{Arg71}_{\text{VRC-PG04}}$ - $\text{Asp368}_{\text{gp120}}$. So while the presence of Arg in position 71 is mandatory to establish the interaction, as it can form a salt bridge with Asp, the residues in CDR2 do not feel a strong constraint and are more free to mutate, provided that they evolve in a correlated fashion preserving the exotic position of the loop.

The above explanation of the DI structure is still under investigation and needs to be confirmed by bioinformatic analysis. Anyway this is probably one of the structural characteristics of the population that are recovered by the MGM method in order to generate the results exposed in Section 6.3.2.

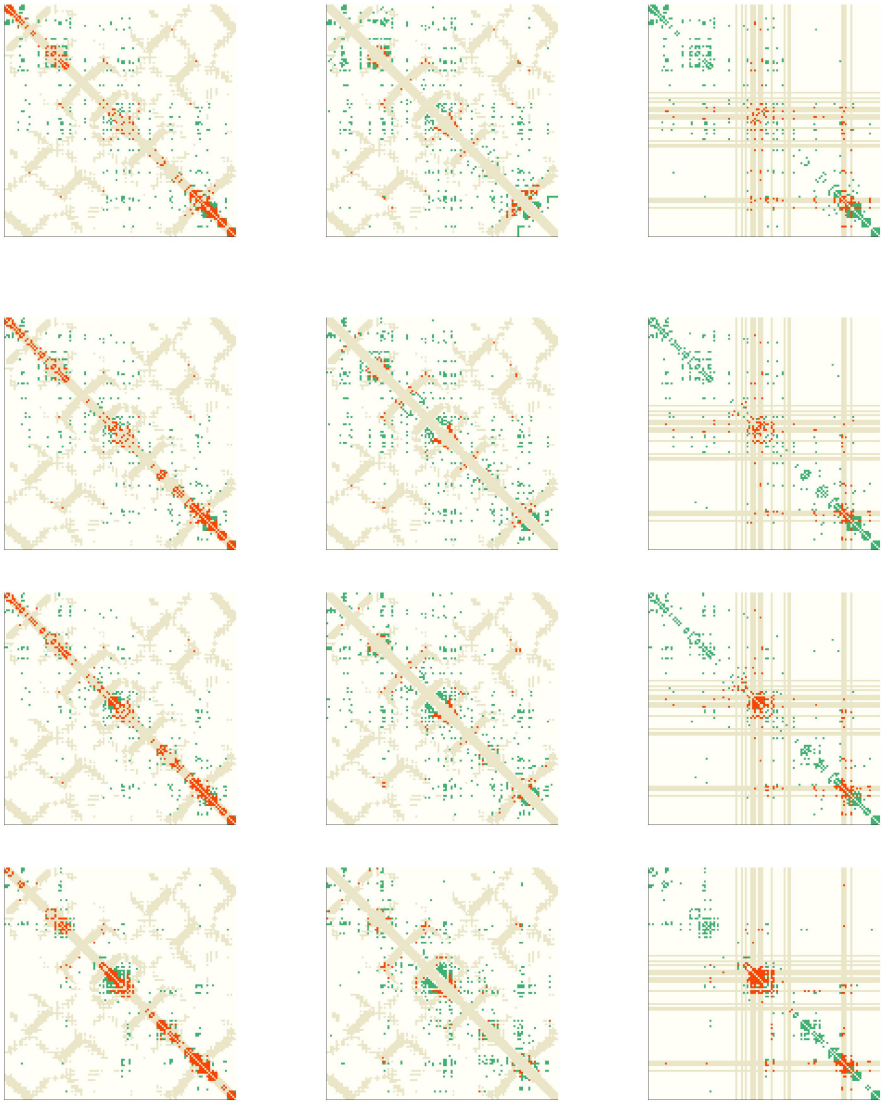


Figure 6.11. Direct Information map computed on the *clusterPG04*.
First column: the internal contact map of VRC-PG04 heavy chain is shown in grey. Two residues are considered to be in contact if at least a couple of atom is at distance lower than 8 Å. The first 300 couples with higher DI are displayed in red when they superpose to the internal contacts (true positive internal contact predictions) and in green when they do not (false positive internal contact predictions).
Second column: Same as in the first column but in red and green are displayed the first 300 couples of higher DI provided that they are at distance greater then 4 along the amino acids chain so that obvious contacts are not considered.
Third column: Grey lines highlight columns and raws referring to residues of the VRC-PG04 heavy chain than are in contact with the antigen (gp120). Dots represents the first 300 couples with higher DI and they are colored in red if at least one of the amino acid of the couple is in contact with the antigen and in green otherwise.
Different raws correspond to different values of the pseudocount parameter, pc. From top to bottom: 1^{th} raw: $\pi = 0.8$; 2^{nd} raw: $\pi = 0.5$; 3^{rd} raw: $\pi = 0.2$; 4^{th} raw: $\pi = 0.01$. In every plot the DI is computed by learning the gaussian model on the sequences belonging to *clusterPG04*.

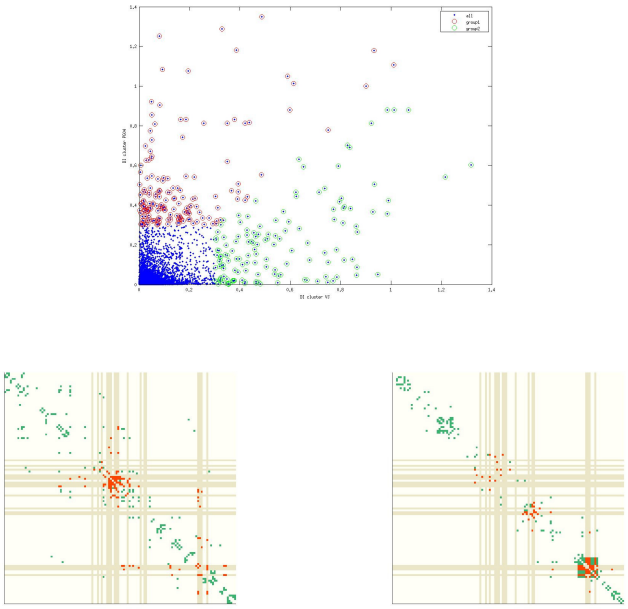


Figure 6.12. Comparison of the DI computed on *clusterPG04* and *clusterVJ* (see main text for details).
Upper panel: In the scatter plot the DI computed on *clusterPG04* (vertical axis) is compared with the DI computed on *clusterVJ* (orizontal axis). Circled couples correspond to pair of residues for which the DI>2.5 when computed on both clusters. Pairs for which the DI is higher when computed on cluster PG04 are circled in green (group 1), while those for which the opposite happens are circled in green (group 2).
Left panel: Pairs belonging to group 1 are compared with the contacts with the antigen as explained in Figure 6.11.
Right panel: Same as in left panel but for group 2.

6.3.4 Recovering the internal contacts

As MGM and gaussian DCA have been developed in order to recover internal native contacts in proteins by analyzing MSAs of homologous protein families, a natural question is whether it is possible to predict internal contacts of Igs variable domains through coevolution analysis in an individual's repertoire, so by performing DCA treating sets of sequenced Igs as a protein families.

As shown in Figure 6.11, sequences of *clusterPG04* treaded with the related multiplicities are not a good set for the prediction of the internal contacts. Only trivial contacts of amino acids that are close in the primary structure (sequence) are reproduced.

Further attempt can be done on other sets of sequences. As a result, as shown in Figures 6.13 and 6.14, the best set to reproduce internal contact is the largest one (all sequenced reads without any regards to the inferred germline, other, of course, than the constraint due to the selection of the primers that, as explain select Igs with inferred V gene in the family IGHV1) with a slight reweighting procedure that eliminates the effect of the sequences multiplicity from the analysis (see [116] for details on the reweighting procedure). In any case, also with this larger set, the method does provide a satisfactory performance in recovering the internal contacts of IGs variable domains. This fact is probably due to the relatively too low degree of variability that is present in the set that do not leave space for covariation of residues in contact.

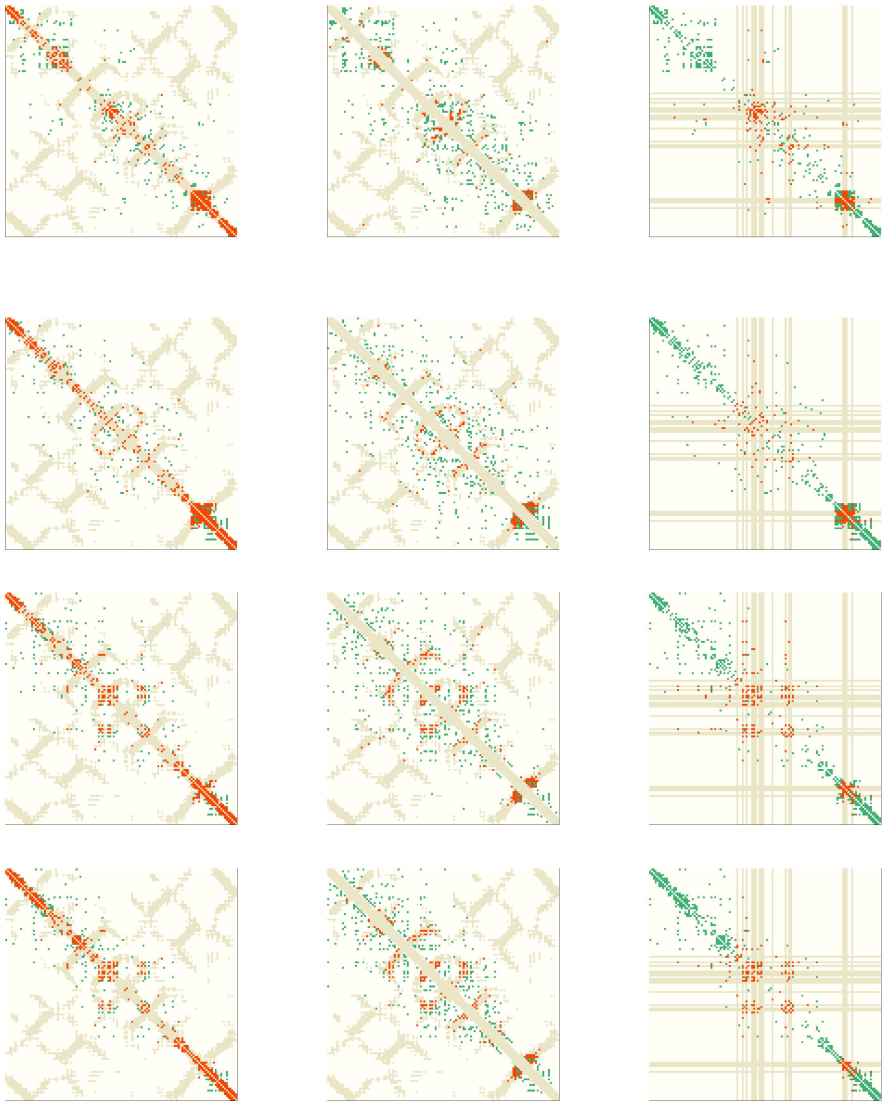


Figure 6.13. Direct information map computed on different set of sequences.
The explanation of the difference among different columns is the same as in Figure 6.11. Different raws correspond to different sets of sequences. From top to bottom: 1) Seqs with inferred IGHV1-2 and IGHJ2 germline genes; 2) Seqs with inferred IGHV1-2 and any IGHJ inferred germline genes 3) Seqs with any IGHV and IGHJ inferred germiline genes; 4) Seqs with any IGHV and IGHJ inferred germiline genes that are the result of a reweighting procedure with $\theta = 0.01$ (see [116]).
Every map has been computed with constant pseudocount parameter π=0.5.

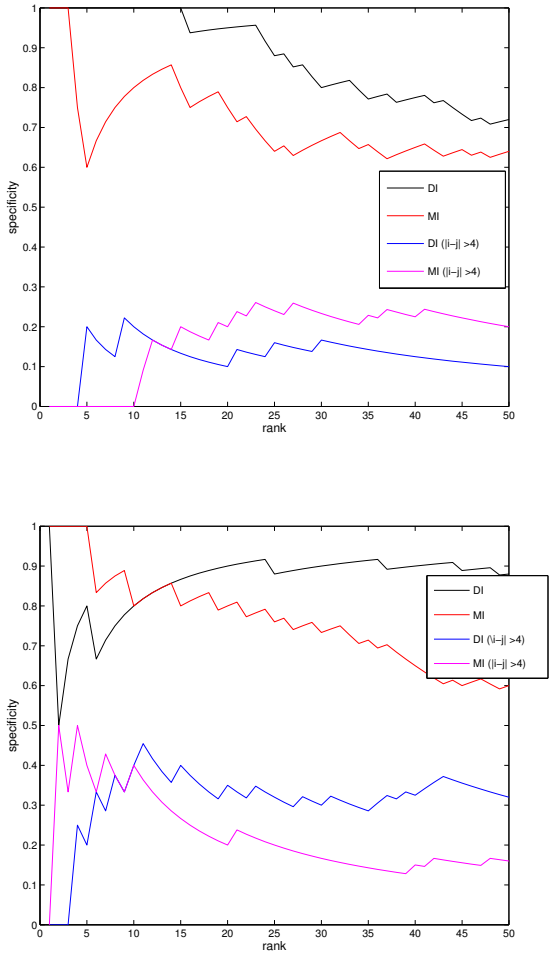


Figure 6.14. DI-ranking-based sensitivity-specificity curves: (*black* for all couples and *blue* for couples that are at distance larger than 4 amino acid along the primary structure) and MI-ranking-based curves (*red* for all couples and *magenta* for couples that are at distance larger than 4 amino acid along the primary structure). Specificity is defined as in Section 6.1.
Both panels refer to sequences with any IGHV and IGHJ inferred germline genes. Lower panel refers to the use of a reweighing procedure with parameter $\theta = 0.01$ (as in Figure 6.13, line 4), while in upper panel every sequence is counted with its multiplicity (as in Figure 6.13, line 3).

6.3.5 Conclusions

Here the preliminary results on an analysis performed on Rep-Seq data through MGM have been reported. They seems indicate that the chosen data set is very promising for the study of the affinity maturation as it presents Igs with many mutations with respect to the germline genes due to the evolutionary pressure provided by the presence of HIV-1 gp120 as an antigen. Moreover a set of highly mutated sequence can be identified by clustering algorithms and its presence allows to select the germline genes that are under strong evolutionary pressure in order to achieve affinity maturation.

Up to now the central result of the analysis is that, having learned a MGM on the cluster of highly mutated heavy chain sequences (*clusterPG04*), for some heavy chains tested for neutralization power, the inferred energy correlates with the neutralization titers. The result needs to be deepened but, if confirmed, it would allow to use Rep-Seq data to study the structure of the neutralization power as a function of the sequence.

As shown in [129], gp120 directed affinity maturation can walk through very different ways in the space of sequences. In fact bnAbs with the same inferred germline and convergent tridimensional structure but found in different donors display highly divergent sequences ($\sim 50\%$). This fact suggests that the neutralization power in the space of sequence can be imagined as a multivalley landscape and that a single valleys are explored by different patients. The perspective of an integration of multiple Rep-Seq experiment performed on different donors seems an interesting step forward for the understanding of the neutralization power optimization.

On the structural level, gaussian DCA analysis on Rep-Seq data does not seem to produce good result for the prediction of internal native contacts. Nevertheless functional DI patterns can be related to more complicated structural features that, for example, can be invloved in the optimization of the antigen binding.

Naïve analysis on the variability of the different positions in the MSA of the repertoire do not seem to represent a satisfactory tool for predicting antigen-heavy chain contacts. Probably the nature of the Ab-antigen interactions is too heterogeneous to display convergent variability patterns.

Bibliography

[1] E. Agliari, L. Asti, A. Barra, and L. Ferrucci. Organization and evolution of synthetic idiotypic networks. *Physical Review E*, 85(5):051909, 2012.

[2] E. Agliari, L. Asti, A. Barra, R. Burioni, and G. Uguzzoni. Analogue neural networks on correlated random graphs. *Journal of Physics A: Mathematical and Theoretical*, 45(36):365001, 2012.

[3] E. Agliari, L. Asti, A. Barra, R. Scrivo, G. Valesini, and R.S. Wallis. Application of a stochastic modeling to assess the evolution of tuberculous and non-tuberculous mycobacterial infection in patients treated with tumor necrosis factor inhibitors. *PloS one*, 8(1):e55017, 2013.

[4] L. Asti, P. Marcatili, A. Pagnani, and G. Uguzzoni. Multivariate Gaussian Modeling for antibodies affinity maturation. *in preparation.*

[5] E. Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31(1):253–258, 1925.

[6] M. Mezard and A. Montanari. *Information, physics, and computation.* Oxford University Press, 2009.

[7] W. Bialek. *Biophysics: searching for principles.* Princeton University Press, 2012.

[8] J.-P. Bouchaud and M. Potters. *Theory of financial risk and derivative pricing: from statistical physics to risk management.* Cambridge University Press, 2003.

[9] K. Huang. *Introduction to statistical physics.* CRC Press, 2001.

[10] L. Onsager. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Physical Review*, 65(3-4):117, 1944.

[11] R.J. Glauber. Time-dependent statistics of the ising model. *Journal of Mathematical Physics*, 4:294, 1963.

[12] W.K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[13] A.C.C. Coolen, R. Kühn, and P. Sollich. *Theory of neural information processing systems.* Oxford University Press, USA, 2005.

[14] D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(26):1792–1796, 1975.

[15] A. Barra, E. Mingione, and F. Guerra. Interpolating the Sherrington-Kirkpatrick replica trick. *Philosophical Magazine*, 92:78–97, 2012.

[16] G. Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979.

[17] F. Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in mathematical physics*, 233(1):1–12, 2003.

[18] M. Talagrand. The Parisi formula. *Annals of Mathematics*, pages 221–263, 2006.

[19] F. Guerra. About the overlap distribution in mean field spin glass models. *International Journal of Modern Physics B*, 10(13n14):1675–1684, 1996.

[20] D.J. Amit. *Modeling brain function: The world of attractor neural networks*. Cambridge Univ. Pr., 1992.

[21] A. Crisanti, D. J. Amit, and H. Gutfreund. Saturation level of the Hopfield model for neural network. *Europhysics Letters*, 2(4):337, 1986.

[22] C.D.C. Allen, T. Okada, and J.G. Cyster. Germinal-center organization and cellular dynamics. *Immunity*, 27(2):190–202, 2007.

[23] M. Meyer-Hermann, E. Mohr, N. Pelletier, Y. Zhang, G.D. Victora, and K.M. Toellner. A theory of germinal center b cell selection, division, and exit. *Cell Reports*, 2012.

[24] N.K. Jerne. Towards the network theory of the immune system. *Annales de l'Institut Pasteur Immunology*, 125:373–389, 1974.

[25] J. Stewart, F.J. Varela, and A. Coutinho. The relationship between connectivity and tolerance as revealed by computer simulation of the immune network: some lessons for an understanding of autoimmunity. *Journal of autoimmunity*, 2:15–23, 1989.

[26] F.J. Varela and A. Coutinho. Second generation immune networks. *Immunology Today*, 12(5):159–166, 1991.

[27] A. Barra and E. Agliari. A statistical mechanics approach to autopoietic immune networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(07):P07004, 2010.

[28] A. Barra and E. Agliari. Equilibrium statistical mechanics on correlated random graphs. *Journal of Statistical Mechanics: Theory and Experiment*, 2011:P02027, 2011.

[29] B.A. Shirley. *Protein stability and folding: Theory and practice*, volume 40. Humana Press New York:, 1995.

[30] R.J. Bagley, J.D. Farmer, S.A. Kauffman, N.H. Packard, A.S. Perelson, and I.M. Stadnyk. Modeling adaptive biological systems. *Biosystems*, 23(2):113–137, 1989.

[31] D. Farmer, A. Lapedes, N. Packard, and B. Wendroff. Evolution, games and learning. *Physica D*, 22:1986, 1986.

[32] J. Carneiro, A. Coutinho, J. Faro, and J. Stewart. A model of the immune network with BT cell co-operation. I—prototypical structures and dynamics. *Journal of Theoretical Biology*, 182(4):513–529, 1996.

[33] D.J. Amit, H. Gutfreund, and H. Sompolinsky. Information storage in neural networks with low levels of activity. *Physical Review A*, 35(5):2293, 1987.

[34] E. Agliari, C. Cioli, and E. Guadagnini. Percolation on correlated random networks. *Physical Review E*, 84:031120, 2011.

[35] M.E.J. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[36] D.S. Callaway, M.E.J. Newman, S.H. Strogatz, and D.J. Watts. Network robustness and fragility: Percolation on random graphs. *Physical review letters*, 85(25):5468, 2000.

[37] E. Agliari and A. Barra. A Hebbian approach to complex-network generation. *Europhysics Letters*, 94:10002, 2011.

[38] J. Stewart and F.J. Varela. Exploring the meaning of connectivity in the immune network. *Immunological reviews*, 110(1):37–61, 1989.

[39] J. Stewart and F.J. Varela. Morphogenesis in shape-space. Elementary meta-dynamics in a model of the immune network. *Journal of theoretical biology*, 153(4):477–498, 1991.

[40] E. Agliari, A. Barra, G. Del Ferraro, F. Guerra, and D. Tantari. Anergy in self-directed B lymphocytes from a statistical mechanics perspective. *arXiv preprint arXiv:1212.2574*, 2012.

[41] P. Erdős and A. Rényi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[42] B. Bollobás. *Random graphs*, volume 73. Cambridge university press, 2001.

[43] Sir F.M. Burnet et al. *The clonal selection theory of acquired immunity*, volume 3. Vanderbilt University Press Nashville, 1959.

[44] F.M. Burnet. A modification of Jerne's theory of antibody production using the concept of clonal selection. *CA: a Cancer Journal for Clinicians*, 26(2):119–121, 1976.

[45] A. Košmrlj, A.K. Chakraborty, M. Kardar, and E.I. Shakhnovich. Thymic selection of t-cell receptors as an extreme value problem. *Physical review letters*, 103(6):068103, 2009.

[46] A.K. Abbas, A.H.H. Lichtman, and S. Pillai. *Cellular and Molecular Immunology: with STUDENT CONSULT Online Access*. Elsevier Health Sciences, 2011.

[47] E. Agliari, A. Barra, S. Franz, and T. Sabetta. Some thoughts on the ontogenesis in Bi-cell immune networks. *Theoretical Biology, Heidelberg Proceedings, arXiv preprint arXiv:1012.2025*, 2012.

[48] B.-R. Chandrika and J. Stewart. Inverse analysis of empirical matrices of idiotypic network interactions. *Bulletin of Mathematical Biology*, 58(6):1123–1153, 1996.

[49] P.A. Cazenave. Idiotypic-anti-idiotypic regulation of antibody synthesis in rabbits. *Proceedings of the National Academy of Sciences*, 74(11):5122–5125, 1977.

[50] E. Agliari, A. Barra, F. Guerra, and F. Moauro. A thermodynamic perspective of immune capabilities. *Journal of Theoretical Biology*, 287:48–63, 2011.

[51] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, and F. Moauro. Multitasking associative networks. *Physical review letters*, 109(26):268101, 2012.

[52] E. Agliari, A. Barra, S. Bartolucci, A. Galluzzi, F. Guerra, and F. Moauro. Parallel processing in immune networks. *Physical Review E*, 87(4):042701, 2013.

[53] E. Agliari, A. Annibale, A. Barra, A.C.C. Coolen, and D: Tantari. Immune networks: multi-tasking capabilities at medium load. *arXiv preprint arXiv:1302.7259*, 2013.

[54] E. Agliari, A. Annibale, A. Barra, A.C.C. Coolen, and D. Tantari. Immune networks: multi-tasking capabilities near saturation. *Journal of Physics A: Mathematical and Theoretical*, 46(41):415003, 2013.

[55] A.C.C. Coolen. *The Mathematical theory of minority games - statistical mechanics of interacting agents*. Oxford University Press, 2005.

[56] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554, 1982.

[57] A. Bovier, A.C.D. Van Enter, and B. Niederhauser. Stochastic symmetry-breaking in a gaussian Hopfield model. *Journal of Statistical Physics*, 95(1):181–213, 1999.

[58] A.C.D. Enter and H.G. Schaap. Infinitely many states and stochastic symmetry in a gaussian Potts-Hopfield model. *Journal of Physics A: Mathematical and General*, 35:2581, 2002.

[59] A. Barra and F. Guerra. About the ergodicity in Hopfield analogical neural network. *Journal of Mathematical Physics*, 49:125217, 2008.

[60] A. Barra, G. Genovese, and F. Guerra. The replica symmetric approximation of the analogical neural network. *Journal of Statistical Physics*, 140(4):784–796, 2010.

[61] D. Bollé, T.M. Nieuwenhuizen, I.P. Castillo, and T. Verbeiren. A spherical Hopfield model. *Journal of Physics A: Mathematical and General*, 36:10269, 2003.

[62] B. Wemmenhove and A.C.C. Coolen. Finite connectivity attractor neural networks. *Journal of Physics A: Mathematical and General*, 36:9617, 2003.

[63] I.P. Castillo, B. Wemmenhove, J.P.L. Hatchett, A.C.C. Coolen, NS Skantzos, and T. Nikoletopoulos. Analytic solution of attractor neural networks on scale-free graphs. *Journal of Physics A: Mathematical and General*, 37:8789, 2004.

[64] M. Mezard, G. Parisi, and M.A. Virasoro. *Spin glass theory and beyond*, volume 9. World scientific Singapore, 1987.

[65] D.O. Hebb. *The organization of behavior: A neuropsychological theory*. Lawrence Erlbaum, 2002.

[66] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[67] S.N. Dorogovtsev and J.F.F. Mendes. *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press, USA, 2003.

[68] S.N. Dorogovtsev, A.V. Goltsev, and J.F.F. Mendes. Critical phenomena in complex networks. *Reviews of Modern Physics*, 80(4):1275, 2008.

[69] F. Guerra and F.L. Toninelli. The thermodynamic limit in mean field spin glass models. *Communications in Mathematical Physics*, 230(1):71–79, 2002.

[70] F. Guerra. Sum rules for the free energy in the mean field spin glass model. *Fields Institute Communications*, 30:161, 2001.

[71] M. Aizenman and P. Contucci. On the stability of the quenched state in mean-field spin-glass models. *Journal of Statistical Physics*, 92(5):765–783, 1998.

[72] A. Barra and L. De Sanctis. Stability properties and probability distributions of multi-overlaps in dilute spin glasses. *Journal of Statistical Mechanics: Theory and Experiment*, 2007:P08025, 2007.

[73] P. Contucci. Stochastic stability: a review and some perspectives. *Journal of Statistical Physics*, 138(1):543–550, 2010.

[74] S. Ghirlanda and F. Guerra. General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity. *Journal of Physics A: Mathematical and General*, 31:9149, 1998.

[75] H. Nishimori. *Statistical physics of spin glasses and information processing: an introduction*, volume 111. Oxford University Press, USA, 2001.

[76] E. Agliari, A. Barra, and F. Camboni. Criticality in diluted ferromagnets. *Journal of Statistical Mechanics: Theory and Experiment*, 2008:P10003, 2008.

[77] F. Guerra and F.L. Toninelli. The high temperature region of the Viana–Bray diluted spin glass model. *Journal of Statistical Physics*, 115(1):531–555, 2004.

[78] L. Viana and A.J. Bray. Phase diagrams for dilute spin glasses. *Journal of Physics C: Solid State Physics*, 18:3037, 1985.

[79] A. Barra. Driven transitions at the onset of ergodicity breaking in gauge-invariant complex networks. *International Journal of Modern Physics B*, 24(30):5995, 2010.

[80] H. Sompolinsky. Neural networks with nonlinear synapses and a static noise. *Physical Review A*, 34(3):2571, 1986.

[81] R.S. Wallis. Mathematical modeling of the cause of tuberculosis during tumor necrosis factor blockade. *Arthritis & Rheumatism*, 58:947–952, 2008.

[82] D. Tracey, L. Klareskog, E.H. Sasso, J.G. Salfeld, and P. Tak. Tumor necrosis factor antagonist mechanisms of action: a comprehensive review. *Pharmacology & Therapeutics*, 117:244–279, 2008.

[83] D.M. Lee and M.E. Weinblatt. Rheumatoid arthritis. *The Lancet*, 358:903–911, 2001.

[84] E.H. Choy and G.S. Panayi. Cytokine pathways and joint inflammation in rheumatoid arthritis. *The New England Journal of Medicine*, 344:907–916, 2001.

[85] M.J. Elliott, R.N. Maini, and M. Feldmann. TNF alpha blockade in rheumatoid arthritis: rationale, clinical out-comes and mechanisms of action. *International Journal of Immunopharmacology*, 17:141–145, 1995.

[86] P.P. Tak and J.R. Kalden. Advances in rheumatology: new targeted therapeutics. *Arthritis Research & Therapy*, 13(5):1–14, 2011.

[87] M.A. Gardam, E.C. Keystone, R. Menzies, S. Manners, E. Skamene, R. Long, and Vinh D.C. Anti-tumour necrosis factor agents and tuberculosis risk: Mechanisms of action and clinical management. *Lancet Infecteuous Diseases*, 3:148–155, 2003.

[88] R.S. Wallis and S. Ehlers. Tumor necrosis factor and granuloma biology: Explaining the differential infection risk of etanercept and infliximab. *Seminars in Arthritis and Rheumatism*, 34:34–38, 2005.

[89] R.S. Wallis. Tumour necrosis factor antagonists: structure, function, and tuberculosis risks. *The Lancet Infectious Diseases*, 8:601–611, 2008.

[90] R.S. Wallis, M.S. Broder, Y.J Wong, M.E. Hanson, and D.O. Beenhouwer. Granulomatous infectious diseases associated with tumor necrosis factor antagonists. *Clinical Infectious Diseases*, 38:1261–1265, 2004.

[91] H.L. Plessner, P. L. Lin, T. Kohno, J. S. Louie, D. Kirschner, J. Chan, and J.L. Flynn. Neutralization of tumor necrosis factor (TNF) by antibody but not tnf receptor fusion molecule exacerbates chronic murine tuberculosis. *The Journal of Infectious Diseases*, 195:1643–50, 2007.

[92] H. Clay, H.E Volkman, and L. Ramakrishnan. Tumor necrosis factor signaling mediates resistance to mycobacteria by inhibiting bacterial growth and macrophage death. *Immunity*, 29:283–94, 2008.

[93] P.L. Lin, A. Myers, L. Smith, C. Bigbee, M. Bigbee, C. Fuhrman, H. Grieser, I. Chiosea, N.N. Voitenek, S.V. Capuano, E. Klein, and J.L. Flynn. TNF neutralization results in disseminated disease during acute and latent M. tuberculosis infection with normal granuloma structure. *Arthritis & Rheumatism*, 62:340–350, 2010.

[94] J. Keane, S. Gershon, R.P. Wise, E. Mirabile-Levens, J. Kasznica, W.D. Schwieterman, J. N. Siegel, and M. M. Braun. Tuberculosis associated with infliximab, a tumor necrosis factor alpha-neutralizing agent. *The New England Journal of Medicine*, 345:1098–104, 2001.

[95] A. Iliopoulos, S. Kedikoglou, S. Laxanis, S. Kourouklis, and E. Katsaros. A case of tuberculous meningoencephalitis in a patient with Behcet's disease. *Clinical Rheumatology*, 25:121–2, 2006.

[96] E.N. Liberopoulos, A.A. Drosos, and M.S. Elisaf. Exacerbation of tuberculosis enteritis after treatment with infliximab. *American Journal of Medicine*, 113:615, 2002.

[97] J. C. Verhave, R. van Altena, M.J. Wijnands, and H.T. Roerdink. Tuberculous peritonitis during infliximab therapy. *The Netherlands Journal of Medicine*, 66:77–80, 2008.

[98] M.H. Schiff, G.R. Burmester, J.D. Kent, A.L. Pangan, H. Kupper, S.B. Fitzpatrick, and Donovan C. Safety analysis of adalimumab (humiraæ) in global clinical trials and US postmarketing surveillance of patients with rheumatoid arthritis. *Ann Rheum Dis*, 65:889–894, 2006.

[99] R. Scrivo, I. Sauzullo, F. Mengoni, G. Iaiani, A.R. Vestri, R. Priori, and et al. Serial interferon-γ release assays for screening and monitoring of tuberculosis infection during treatment with biologic agents. *Clinical Rheumatology*, 31:1567–1575, 2012.

[100] R. Scrivo, I. Sauzullo, F. Mengoni, R. Priori, M. Coppola, and A. Altri. Mycobacterial IFN-gamma release variations during long-term treatment with TNF blockers: lack of correlation with clinical outcome. *Journal of Rheumatology*, (in press), 2012.

[101] D. Wagner and L.S. Young. Nontuberculous mycobacteria infections: a clinical review. *Infection*, 31(257-270), 2003.

[102] D.E. Griffith, T. Aksamit, B.A. Brown-Elliott, A. Catanzaro, C. Daley, and F. et al. Gordin. An official ats/idsa statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am J Respir Crit Care Med*, 175:367–416, 2007.

[103] K.L. Winthrop, E. Chang, S. Yamashita, M.F. Iademarco, and P.A. LoBue. Nontuberculous mycobacteria infections and anti-tumor necrosis factor-α therapy. *Emerging Infectious Diseases Journal*, 15(10):1556–1561, 2009.

[104] K.L. Winthrop, S. Yamashita, S.E. Beekman, and P.M. Polgreen. Mycobacterial and other serious infections in patients receiving anti-TNF and other newly approved biological therapies; case-finding via the emerging infectious network. *Clinical Infectious Diseases*, 46:1738–1740, 2008.

[105] J.J. Peschon, D.S. Torrance, K.L. Stocking, M.B. Glaccum, C. Otten, C.R. Willis, K. Charrier, P.J. Morrissey, C.B. Ware, and Mohler K.M. TNF receptor-deficient mice reveal divergent roles for p55 and p75 in several models of inflammation. *The Journal of Immunology*, 160:943–948, 1998.

[106] S. Marino, D. Sud, H. Plessner, P.L. Lin, J. Chan, J.L. Flynn, and D.E. Kirschner. Differences in reactivation of tuberculosis induced from anti-tnf treatments are based on bioavailability in granulomatous tissue. *PLoS Computational Biology*, 3(10):1909–1924, 2007.

[107] M. Fallahi-Sichani, J.L. Flynn, J.J. Linderman, and D.E. Kirschner. Differential Risk of Tuberculosis Reactivation among Anti-TNF Therapies Is Due to Drug Binding Kinetics and Permeability. *The Journal of Immunology*, 188(7):3169–78, 2012.

[108] R. K. Pearson. *Exploring Data in Engineering, the Sciences, and Medicine*. Oxford University Press, 1952.

[109] M.D. Iseman and T.K. Marras. The importance of nontubercolous mycobacterial lung disease. *American Journal of Respiratory and Critical Care Medicine*, 178:999–1001, 2008.

[110] J.O. III Falkinham. Nontuberculous mycobacteria in the environment. *Clin Chest Med*, 23:529–551, 2002.

[111] I. Sutherland, M.A. Bleiker, J. Meijer, and K. Styblo. The risk of tuberculous infection in the Netherlands from 1967 to 1979. *Tubercle*, 64:241–253, 1983.

[112] K. Shanaube, C. Sismanidis, H. Ayles, N. Beyers, A. Schaap, and al. et. Annual risk of tuberculous infection using different methods in communities with a high prevalence of TB and HIV in Zambia and South Africa. *PLoS One*, 4:e7749, 2009.

[113] T.M. Daniel and S.M. Debanne. Estimation of the annual risk of tuberculosis infection for white men in the United States. *The Journal of Infectious Diseases*, 175:1535–1537, 1997.

[114] A Tramontano. *The ten most wanted solutions in protein bioinformatics*. CRC Press, 2005.

[115] S. Rabello, A.C.C. Coolen, C.J. Perez-Vicente, and F. Fraternali. A solvable model of the genesis of amino-acid sequences via coupled dynamics of folding and slow-genetic variation. *Journal of Physics A: Mathematical and Theoretical*, 41(28):285004, 2008.

[116] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D.S. Marks, C. Sander, R. Zecchina, J.N. Onuchic, T. Hwa, and M. Weigt. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proceedings of the National Academy of Sciences*, 108(49):E1293–E1301, 2011.

[117] R.A. White, H. Szurmant, Hoch J.A., and Hwa T. Features of protein–protein interactions in two-component signaling deduced from genomic libraries. *Methods Enzymols*, 422:75–101, 2007.

[118] J.M. Skerker, B.S. Perchuk, A. Siryaporn, E.A. Lubin, O. Ashenberg, M. Goulian, and M. T. Laub. Rewiring the specificity of two-component signal transduction systems. *Methods Enzymols*, 133:1043–1054, 2008.

[119] M. Weigt, R.A. White, H. Szurmant, J.A. Hoch, and T. Hwa. Identification of direct residue contacts in protein–protein interaction by message passing. *Proceedings of the National Academy of Sciences*, 106(1):67–72, 2009.

[120] D.J.C. MacKay. *Information theory, inference and learning algorithms*. Cambridge university press, 2003.

[121] C. Baldassi, M. Zamparo, C. Feinauer, A. Procaccini, R. Zecchina, and A. Pagnani. Multivariate Gaussian Modeling of protein families: Predicting residue contacts and protein-interaction partners, in preparation. 2013.

[122] Andrea Pagnani. Private communication. 2013.

[123] J. Benichou, R. Ben-Hamo, Y. Louzoun, and S. Efroni. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology*, 135(3):183–191, 2012.

[124] J.A. Weinstein, N. Jiang, R.A. White, D.S. Fisher, and S.R. Quake. High-throughput sequencing of the zebrafish antibody repertoire. *Science*, 324(5928):807–810, 2009.

[125] T. Mora, A.M. Walczak, W. Bialek, and C.G. Callan. Maximum entropy models for antibody diversity. *Proceedings of the National Academy of Sciences*, 107(12):5405–5410, 2010.

[126] K. Larimore, M.W. McCormick, H.S. Robins, and P.D. Greenberg. Shaping of human germline igh repertoires revealed by deep sequencing. *The Journal of Immunology*, 189(6):3221–3230, 2012.

[127] A. Murugan, T. Mora, A.M. Walczak, and C.G. Callan. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proceedings of the National Academy of Sciences*, 109(40):16161–16166, 2012.

[128] R.A. Neher and B.I. Shraiman. Competition between recombination and epistasis can cause a transition from allele to genotype selection. *Proceedings of the National Academy of Sciences*, 106(16):6866–6871, 2009.

[129] X. Wu, T. Zhou, J. Zhu, B. Zhang, I. Georgiev, C. Wang, X. Chen, N.S. Longo, M. Louder, K. McKee, et al. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science*, 333(6049):1593–1602, 2011.

[130] T. Zhou, I. Georgiev, X. Wu, Z.-Y. Yang, K. Dai, A. Finzi, Y. Do Kwon, J.F. Scheid, W. Shi, L. Xu, et al. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. *Science*, 329(5993):811–817, 2010.

[131] M. Bailly-Bechet, S. Bradde, A. Braunstein, A. Flaxman, L. Foini, and R. Zecchina. Clustering with shallow trees. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(12):P12010, 2009.