



International School for Advanced Studies
Molecular and Statistical Biophysics Group

COMPUTATIONAL APPROACHES TO COMPLEX BIOLOGICAL NETWORKS

Ph.D. Thesis of Giuseppe Facchetti

Supervisor: Dr. Claudio Altafini

October 2013

*a Tomáš
e alla nostra Trieste*

Contents

| | |
|--|-----------|
| Overview | 1 |
| PART I | |
| Perturbation of metabolic networks in Flux Balance Analysis | 3 |
| Introduction | 5 |
| 1 Constraint-based methods for metabolic networks | 9 |
| 1.1 Genome-scale metabolic networks | 9 |
| 1.2 Flux Balance Analysis | 12 |
| 1.3 Linear programming and duality theorem | 18 |
| 2 Drug synergisms for human metabolic diseases and cancer | 21 |
| 2.1 Background | 21 |
| 2.1.1 State of the art | 22 |
| 2.2 Materials and Methods | 25 |
| 2.2.1 The algorithm: bilevel optimization | 25 |
| 2.2.2 Algorithm for competing organisms | 29 |
| 2.2.3 Metabolic networks | 29 |
| 2.2.4 Selection procedure for the drugs | 30 |
| 2.3 Human metabolic diseases: results | 32 |
| 2.4 Human cancer: results | 39 |
| 2.5 Discussion | 43 |
| 3 A more realistic description of drug action: partial inhibition | 45 |
| 3.1 Motivations | 45 |
| 3.2 Materials and Methods | 46 |
| 3.2.1 The drug synergism problem | 46 |
| 3.2.2 The strong duality theorem and the need of Boolean variables | 50 |
| 3.2.3 The implementation of partial inhibition | 51 |
| 3.2.4 Inhibitions or activation of the objective reaction | 53 |
| 3.2.5 Cases of multiple equivalent solutions | 53 |
| 3.2.6 Metabolic networks | 55 |
| 3.3 Results | 55 |

| | | |
|----------------|---|------------|
| 3.3.1 | Screening for optimal drug combinations on <i>E.coli</i> core network | 55 |
| 3.3.2 | Drug interaction surfaces: three case studies | 59 |
| 3.3.3 | The problem due to the nonuniqueness of \mathbf{v}^{ut} | 61 |
| 3.3.4 | Computational performances. | 63 |
| 3.4 | Discussion and Conclusions | 64 |
| 4 | Transient dynamics in the metabolism of perturbed microorganisms | 67 |
| 4.1 | Motivations and state of the art | 67 |
| 4.1.1 | Experimental data from literature | 68 |
| 4.2 | Methods | 70 |
| 4.2.1 | The starting point of the transient | 70 |
| 4.2.2 | An heuristic criterion for long-term metabolic adjustment . . | 72 |
| 4.3 | Preliminary results | 74 |
| 4.4 | Discussions and perspectives | 77 |
| PART II | | |
| | Quantifying disorder in complex signed networks: from gene-regulatory to social networks | 79 |
| | Introduction | 81 |
| 5 | Theoretical framework: an overview | 83 |
| 5.1 | Ising spin-glass model | 83 |
| 5.2 | Optimization algorithm: gauge transformation | 85 |
| 5.3 | From spin glass to lossy data compression problem | 87 |
| 5.4 | Rate-distortion theorem | 90 |
| 6 | A rate-distortion theory for gene regulatory networks | 93 |
| 6.1 | Gene regulatory networks | 93 |
| 6.2 | Materials and Methods | 95 |
| 6.2.1 | Analyzed networks | 95 |
| 6.2.2 | Gene regulatory networks as constraint satisfaction problems | 96 |
| 6.2.3 | Null model of a gene regulatory network of given topology . . | 97 |
| 6.3 | Results and Discussion | 98 |
| 6.4 | Conclusions | 103 |
| 7 | Structural balance in large signed social networks | 105 |
| 7.1 | From structural balance to spin glasses | 105 |
| 7.2 | Materials and Methods | 107 |
| 7.2.1 | Three on-line large-signed social networks | 107 |
| 7.2.2 | Computation of global balance | 107 |
| 7.2.3 | Statistics | 110 |
| 7.3 | Results and Discussion | 111 |
| 7.3.1 | Global Structural Balance: the ground state | 111 |

| | |
|---|------------|
| <i>CONTENTS</i> | III |
| 7.3.2 Energy landscape characterization for Epinions and Slashdot | 116 |
| 7.4 Conclusions | 130 |
| PART III | |
| Diffusion and spatial gradients in the rod phototransduction pathway | 133 |
| Introduction | 135 |
| 8 Biology of rod photoreceptors | 137 |
| 8.1 Rods: function and geometry | 137 |
| 8.2 The signaling cascade | 139 |
| 8.3 Discs ageing | 143 |
| 9 Signaling efficiency and cGMP diffusion in rod photoreceptors | 145 |
| 9.1 Motivation | 145 |
| 9.2 Materials and Methods | 146 |
| 9.3 Gradient of efficiency along the OS | 146 |
| 9.3.1 Experimental results | 146 |
| 9.3.2 Modeling of the slope of the initial rising phase | 150 |
| 9.3.3 Single photon experiments | 153 |
| 9.4 Estimation of the diffusion coefficient of cGMP | 155 |
| 9.4.1 A parameter-free model | 156 |
| 9.5 Conclusions | 158 |
| Bibliography | 160 |

Overview

As more and more of the basic constituents of a cell (such as genes, proteins and chemical compounds) are being described and characterized, the field of Biology is showing a growing interest in describing the nature and the effect of their reciprocal connections. The need of understanding the links among biological components and of determining their interactions is one of the *raisons d'être* and of the driving forces behind the emergence of Systems Biology. This field of study uses an approach to biological and biomedical research which is more holistic than the traditional reductionist one. From a circuit of a single cellular process (for instance a signaling cascade) up to a large network of many cellular functions (e.g. all metabolic reactions), Systems Biology aims at providing a theoretical description of the principles that govern the living systems (for instance gene expression, chemotaxis, nutrients selection). Because of this holistic approach and because of the widely different level of complexity of the networks involved, many different mathematical methods have been developed during the years, based on physical, chemical and engineering approaches. The variety of these tools represents a very important aspect of Systems Biology.

Some of these computational methods are used in this thesis in order to investigate various properties of different biological systems. In particular, three research projects are presented in the three parts in which this thesis is split.

- I. The first part deals with the prediction of the perturbation of cellular metabolism induced by drugs. In particular, using a recently reconstructed model of the human metabolic network, we consider the problem of identifying the most selective drug synergisms for given therapeutic targets (in our case related to metabolic diseases and cancer). The problem is approached using Flux Balance Analysis as computational tool, a framework which provides a reliable quantitative description of the reaction fluxes of genome-wide metabolic networks.
- II. The second part of the thesis considers another class of biological networks, namely gene regulatory networks. In these systems, activation or inhibition

are the modes of regulation of the gene expression. The resulting signed graph model has remarkable analogies with an Ising spin glass: activation/deactivation are rephrased as positive/negative coupling between spins and the corresponding energy function is generated. Using this analogy, an analysis of the content of “disorder” and of the energy landscape can be carried out. The same theory has been also applied to much larger signed graphs drawn from other contexts, such as from social network theory.

- III. Finally, the last part of this thesis concerns the modeling of the signaling pathway of rod photoreceptors, in particular the investigation of the heterogeneity of the response of this signaling cascade when different parts of the cell are selectively stimulated with light. The electrophysiological experiments produced by our collaborators in the Neurobiology laboratory at SISSA have been analyzed with various dynamical systems: the ordinary and partial differential equations used to describe some steps of the phototransduction give an insight into the process of ageing of photoreceptors and into the role of the diffusion in the cytoplasm of these cells.

PART I

PERTURBATION OF METABOLIC NETWORKS IN FLUX BALANCE ANALYSIS

Introduction

Metabolism is probably the best known “organized” system at cellular level: understanding the chemical rules that govern the reactions which take place in the cell has helped in elucidating the connections between the various metabolic transformations. Linking the identified compounds (metabolites) through this set of chemical reactions, a network appears: nodes represent metabolites and directed edges reactions (each catalyzed by a specific enzyme). Thanks to the advanced high-throughput techniques and exploiting the genetic information provided by genome sequencing projects, metabolic networks are now more and more reliable and represent one of the main paradigms inside Systems Biology. Genome-scale metabolic networks have been reconstructed for many organisms, ranging from bacteria to fungi, from yeasts to the human cell.

In order to perform a characterization of these networks, many methods have been developed and used. In the following we focus on Flux Balance Analysis (FBA). In this constraint-based framework the minimization (or maximization) by linear programming of an heuristic cost functional over the stoichiometric constraints is used to compute an “effective rate” of each metabolic reaction. Several aspects of the metabolic networks have been investigated through FBA, like robustness with respect to reaction deletion and the effects of external perturbations (e.g. gene knockout, drug inhibition, change in the growing medium). In the same spirit, specific heuristic methods have been proposed to deal with these perturbation effects. Among these, it is worth mentioning Minimization Of Metabolic Adjustment (MOMA) and Regulatory On/Off Minimization (ROOM). While MOMA replaces the linear programming with a linear-quadratic programming, ROOM requires instead to resort to the more expensive mixed-integer linear programming (further details in the next chapter).

In the following chapters, after a brief description of genome-scale metabolic networks and FBA (Chapter 1), we present two applications of drug-induced network disturbance and a new heuristic criterion for describing the transient dynamics after a perturbation. Here is a more detailed description of the contents that follows.

Drug synergism for human diseases and cancer (Chapter 2)

In this chapter we investigate the effect of one or more drugs on the human metabolic network. Because of the topology of the network, the effect induced by a combination of drugs can be larger than the simple superposition of the perturbations induced by each single drug. This “nonlinear” effect is referred to as a *synergism*. Our aim is to exploit the synergism among already approved drugs in order to predict new multiple drug therapies. The repurposing of known drugs is indeed a very attractive task because many clinical results are already available for approved drugs. However, assessing the potential of multidrug therapies is a difficult task because of the combinatorial complexity (both theoretical and experimental) and because of the requirements on the selectivity of the therapy (i.e. it must have an effect on certain targets, but a limited side effect on the rest of the network).

To cope with this problem, we have developed a novel method for the systematic *in silico* investigation of synergistic effects of drugs on genome-scale metabolic networks. The algorithm (based on “bilevel optimization”, i.e. two nested optimization steps) efficiently finds the optimal combination of currently available drugs which guarantees the inhibition of an objective reaction (for instance the reaction responsible for a disease), while minimizing the side effect on the other cellular processes. In this project, we assume that a drug completely inhibits its target enzyme (i.e. it stops the metabolic reaction catalyzed by the enzyme). Two different applications are considered: finding drug synergisms for human metabolic diseases (like diabetes, obesity and hypertension) and finding antitumoral drug combinations with minimal side effect on the normal human cells. The results we obtain are consistent with some of the available therapeutic indications and predict new multiple drug treatments.

Perturbation of a network by drug partial inhibition (Chapter 3)

Within FBA the investigation of complex subtasks, such as finding an optimal combination of drugs, can be set up as a bilevel optimization problem. In the previous algorithm, in order to keep the linearity and convexity of these nested optimization problems, an ON/OFF description of the effect of the perturbation (i.e. Boolean variable) has been used. This restriction may not be always realistic: for instance, the inhibition of a reaction induced by a drug in fact may be partial rather than complete.

In this chapter we improve the previous formulation of the bilevel optimization problem overcoming the oversimplified ON/OFF modeling while preserving the linear nature of the problem. As guiding example, the same case study of drug synergism is considered: the search for the best multi-drug treatment which modulates an objective reaction and yields the minimal perturbation on the whole network. The inhibition induced by a drug (i.e. the interval $[0,1]$ instead of the set $\{0,1\}$) is described and tuned through a convex combination of a fixed number of Boolean variables (instead of a single Boolean variable). Compared to a simple ON/OFF modeling, the application of the algorithm to the core metabolism of *E.coli* highlights the possibility of finding both a broader spectrum of drug combinations and a

more fine-grade tuning of the selectivity. It is worth noting that this approach can be extended to any other FBA problem such as, for instance, strain improvement and cost functional identification.

Analysis of the transient dynamics in perturbed networks following a change in the growing medium (Chapter 4)

The description of the perturbation studied in the two previous projects are stationary. However, when a metabolic network is perturbed, it experiences a transient excursion before settling to the new steady state predicted by FBA. The criteria available in the current literature for perturbed networks (e.g. MOMA) provides only a snapshot of the initial part of this transient. The short-term metabolic adjustment described by MOMA consists typically in a temporary drop of the growth rate and in the activation of many reactions. Both effects are induced for example by a gene deletion or by a change of the carbon source in the growing medium. However, less attention has been given to the late dynamical adjustment of the metabolism that leads to the recovery of the proliferation rate.

Referring to experimental measurements reported in literature for *E.coli* cultures, we interpret this recovery as a stochastic sequence of small metabolic adjustments. The heuristic criterion we suggest to describe this recovery is based on concepts related to natural selection. The reactions transiently activated by the perturbation are progressively turned off when this provides a local improvement of the growth rate (the analogous of the so-called “fitness” in the evolutionary context). Applying iteratively this procedure together with the induced short-term adjustment (MOMA), we obtain a description of the recovery of the growth rate consistent with the experimental data we found in literature.

While the metabolic adjustment for describing the transient dynamics is work in progress (which requires still further analysis), the first two topics mentioned above have already been published in the following papers:

- G. Facchetti, M. Zampieri, C. Altafini *Predicting and characterizing selective multiple drug treatments for metabolic diseases and cancer*, BMC Systems Biology **6**, 115–130, (2012);
- G. Facchetti and C. Altafini *Partial inhibition and bilevel optimization in Flux Balance Analysis*, BMC Bioinformatics (2013), in press;

Chapter 1

Constraint-based methods for metabolic networks

1.1 Genome-scale metabolic networks

In recent years, genome-scale metabolic networks have represented an important paradigm of Systems Biology, well describing how interesting and relevant biological features can be deduced in spite of the complexity of the system [1, 2, 3, 4].

Metabolism is the set of chemical transformations among biomolecules within the cells of living organisms, necessary for the cells to grow and reproduce. It can be represented as a network whose nodes are the metabolites connected by single or multiple arrows (the reactions). More properly, it is a directed hypergraph [5, 6] which finds also a representation as a bipartite directed graph (see Figure 1.1). Each biochemical reaction is catalyzed by a specific enzyme. Enzymes are crucial for the metabolism because they allow organisms to drive desirable reactions that require energy and will not occur by themselves, by coupling them to spontaneous reactions that release energy. As enzymes act as catalysts, they allow these reactions to proceed quickly and efficiently. Moreover, enzymes allow the regulation of metabolic pathways in response to changes in the cell environment or signals from other cells. The presence or the absence of an enzyme makes the corresponding chemical transformation possible or not possible: therefore, the expression (non-expression) of a protein is correlated to the functioning (non-functioning) of a specific reaction. From this point of view, enzymes can be interpreted as the cellular switches of the catalyzed reactions.

Metabolism is formally divided into two categories: *catabolism* that breaks down organic nutrients, for example to produce energy in cellular respiration; *anabolism* that uses this energy to build components of cells such as aminoacids and nucleotides (the monomers for proteins and nucleic acids respectively). The chemical reactions of metabolism are usually organized into metabolic *pathways*, in which one compound

is transformed (or synthesized) through a series of steps into another compound, by a sequence of enzymes. In this perspective, metabolism shows a multi-level structure:

- ⊗ **cellular input and outputs:** metabolism is viewed as a complex machinery which transforms nutrients into biomolecules for cellular growth; waste is expelled from the cell to the environment;
- ⊗ **catabolism and anabolism:** the two main sectors of the metabolism, one for the processing of the nutrients and the other for the biosynthesis of macromolecule precursors (and special metabolites);
- ⊗ **pathways:** to a finer resolution, pathways are identified as set of reactions which serve a definite role, for instance the pathway for the synthesis of lipids, the pathway for degradation of complex carbohydrates etc.;
- ⊗ **individual reactions:** each transformation can form or break chemical bounds, can transfer a chemical group from one molecule to another, can oxidize or reduce a compound, can produce or consume energy etc. Except for very simple reactions, all these processes are catalyzed by an enzyme. According to the type of the catalyzed transformation, enzymes (and therefore reactions) are labeled with the universally accepted Enzyme Commission (EC) numbers.

Beside this universal structure, an important feature of metabolism is the similarity of the basic metabolic pathways and components among even vastly different species. The most evident example is the set of carboxylic acids that are best known as the intermediates in the citric acid cycle (the Krebs cycle); it is present in all known organisms, being found in species as diverse as the unicellular bacterium *Escherichia coli* and huge multicellular organisms like elephants. These striking similarities in the metabolic pathways are likely due to their early appearance in evolutionary history, and have been retained by natural selection because of their efficiency.

Nevertheless, each organism has some peculiarities which make necessary to reconstruct the whole network for all of them. This is performed by exploiting the information collected by various high-throughput experiments. Figure 1.2 depicts a workflow of the metabolism reconstruction indicating the methods and approaches used. The main data sources are the following:

1. *biochemistry*: the strongest evidence for the presence of a metabolic reaction is the isolation of the corresponding enzyme;
2. *genomic*: based on DNA sequence homology, functional assignments can be identified;
3. *physiology*: for instance, the ability of a organism to produce an aminoacid while fermented, may support the existence of the corresponding pathway;
4. *in silico modeling data*: in order to identify a gap, an *in silico* study of the consistency is often useful; as result, the presence of some reactions can be inferred.

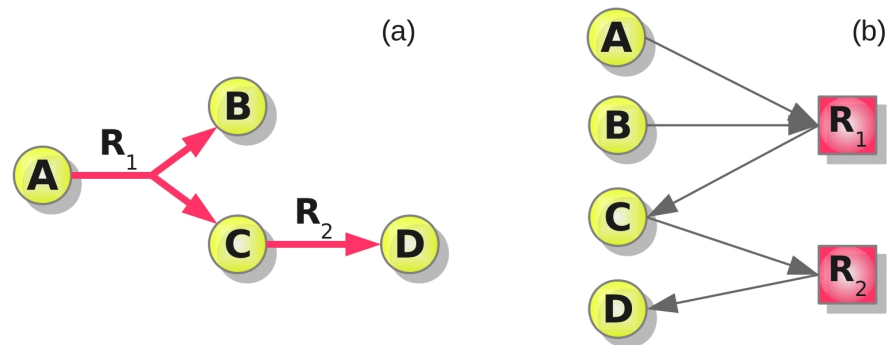


Figure 1.1: **Metabolic networks and hypergraphs.** Example of two equivalent representations of a metabolic network. Reaction R_1 (i.e. $A \rightarrow B + C$) and reaction R_2 (i.e. $C \rightarrow D$) are depicted as oriented hyperedges of a network (a) or as nodes of a bipartite directed graph (b).

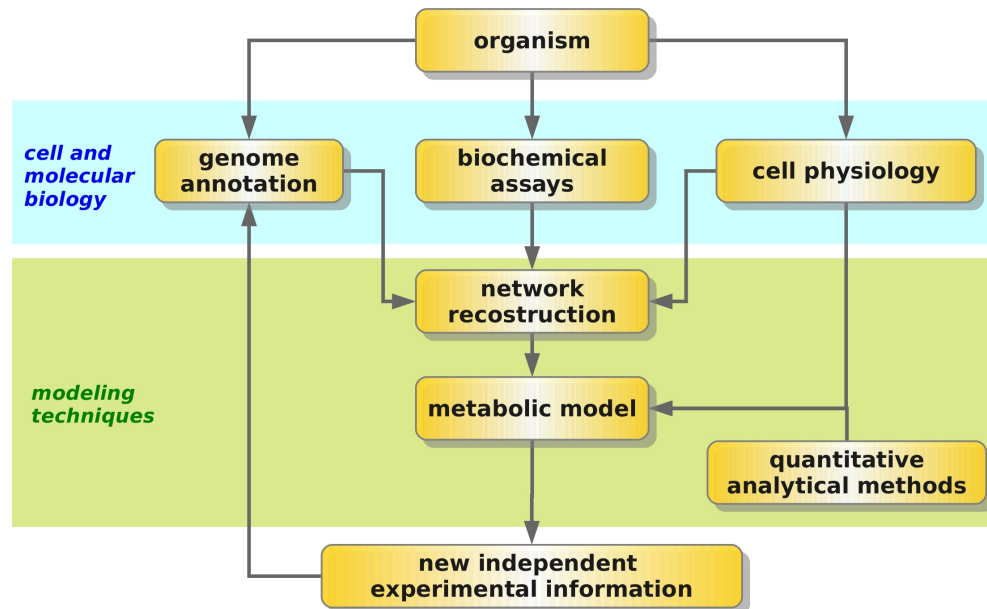


Figure 1.2: **Metabolism reconstructions.** A scheme of the overall process of genome-scale metabolic network reconstruction and subsequent model formation (adapted from [7]).

Table 1.1: Example of size of various reconstructed metabolic networks downloaded from the BIGG database.

| Organism | number of | | | Ref. |
|-----------------------------------|-------------|-----------|--------------|------|
| | metabolites | reactions | compartments | |
| <i>Helicobacter pylori</i> | 485 | 554 | 2 | [8] |
| <i>Staphylococcus aureus</i> | 655 | 743 | 2 | [9] |
| <i>Methanosarcina barkeri</i> | 628 | 690 | 2 | [10] |
| <i>Shewanella oneidensis</i> | 713 | 870 | 2 | [11] |
| <i>Mycobacterium tuberculosis</i> | 826 | 1025 | 2 | [12] |
| <i>Saccharomyces cerevisiae</i> | 1061 | 1266 | 3 | [13] |
| <i>Escherichia coli</i> | 1668 | 2382 | 3 | [14] |
| <i>Salmonella typhimurium</i> | 1802 | 2546 | 3 | [15] |
| <i>Homo sapiens</i> | 2766 | 3742 | 8 | [16] |

Reconstructed networks are published and are made available from different research groups together with additional important details (name and chemical formula of metabolites, stoichiometry of each reaction, EC number of the associated genes, code of the genes, pathway identification, biomass composition, etc.). The Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.jp/kegg/>), the Biochemical Genetic and Genomic (BIGG, bigg.ucsd.edu/), EcoCyc and MetaCyc (www.ecocyc.org/ and www.metacyc.org/) are just a few examples of databases where metabolic networks can be downloaded. Table 1.1 reports a list of some reconstructed networks available from the BIGG database. An example of such metabolic networks is that of *E.coli* reported in Fig. 1.3: it contains 1668 metabolites and 2382 reactions, distributed in 3 compartments (the compartmentalization is described denoting metabolites in different areas with different labels and introducing a transport process among them).

Making these networks available has significantly facilitated their spread and usage. Furthermore, new computational tools have been developed in order to obtain *in silico* predictions from them, and to try to collect more and more information about the principles which govern the structure and the function of cell metabolism. In particular, for its simplicity, FBA is a widely used framework that has contributed to exploit the knowledge contained in these metabolic reconstructions and to enlarge their spectrum of applications.

1.2 Flux Balance Analysis

FBA is a linear constraint-based framework for stoichiometric models of metabolic networks. We give here a short overview of this formalism. More details can be found in many reviews and books (for example [17, 18]).

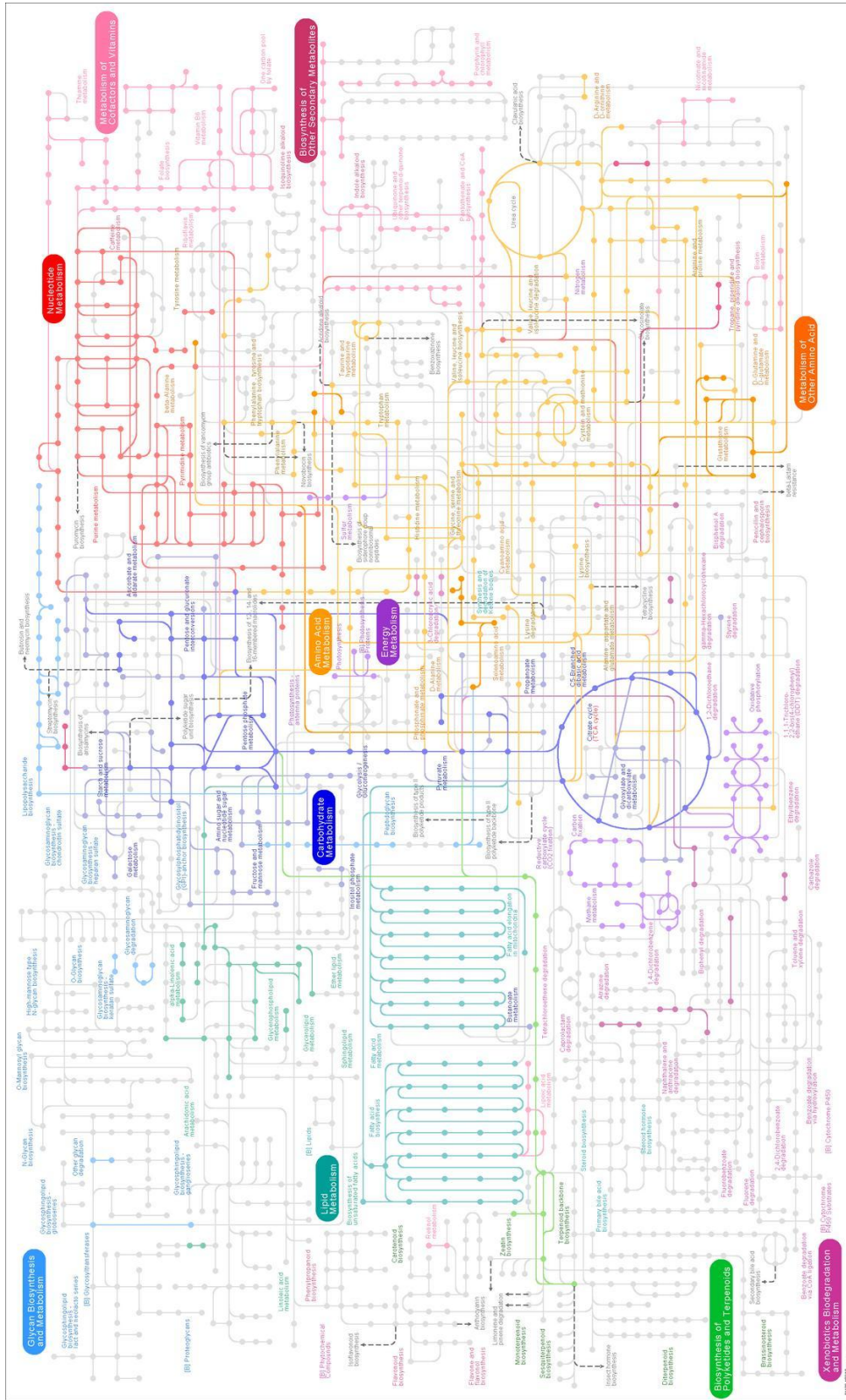
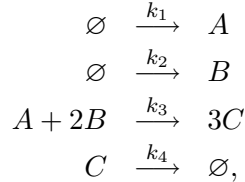


Figure 1.3: Example of a genome-scale reconstructed metabolic network. *E. coli* metabolic network contains 2382 reactions and 1668 metabolites, divided in 40 pathways.

No kinetic parameters

Viewed as a set of chemical reactions, the metabolism can be in principle treated as a dynamical system. Let us consider the simple network given by the following processes:



where \emptyset indicates that reactions 1, 2 and 4 consist in a transport from/to another compartment, and the parameter k_i represents the kinetic rate constant of the reaction i ($i = 1, \dots, 4$). The corresponding differential equations (based on the law of mass action) which describe the rate of variation of the concentrations of the three metabolites ($[A]$, $[B]$ and $[C]$) are:

$$\begin{aligned} d[A]/dt &= k_1 - k_3[A][B]^2; \\ d[B]/dt &= k_2 - 2k_3[A][B]^2; \\ d[C]/dt &= 3k_3[A][B]^2 - k_4[C]. \end{aligned}$$

Studying the dynamics requires the knowledge of the kinetic rate constants for all reactions, which is difficult for a network of thousands transformations (especially because the measurement or the theoretical calculation of a kinetic parameter is a very difficult task *in vivo*).

The main simplification introduced in FBA is that in the right-hand side of the differential equations the dependence on the concentrations of the metabolites is completely ignored and only a description of the reactions (with their stoichiometry) is maintained. In particular, each reaction is considered as a flux whose value indicates the rate of the reaction: in our example 4 fluxes are necessary, therefore a vector of 4 elements is enough to describe the reactions. We denote this vector by $\mathbf{v} = [v_1, v_2, v_3, v_4]^T$. Considering the stoichiometric coefficients, the dynamical systems can be rewritten in terms of \mathbf{v} as follows:

$$\begin{aligned} d[A]/dt &= v_1 - v_3; \\ d[B]/dt &= v_2 - 2v_3; \\ d[C]/dt &= 3v_3 - v_4, \end{aligned}$$

or equivalently

$$\begin{bmatrix} d[A]/dt \\ d[B]/dt \\ d[C]/dt \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} v_1 + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} v_2 + \begin{bmatrix} -1 \\ -2 \\ 3 \end{bmatrix} v_3 + \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} v_4 = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 3 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix},$$

where column j of the resulting 3×4 matrix contains the stoichiometric coefficients of reaction j .

Steady-state assumption

A second assumption is then introduced. As mentioned, a metabolic transformation takes place if the corresponding enzyme (i.e. gene) is expressed and active. Metabolic reactions are therefore subordinated to the transcriptional processes which occur on a much slower time scale. A few examples of typical time for the *E.coli* bacterium are here reported [19]:

| TRANSCRIPTIONAL PROCESSES | | |
|--------------------------------------|-----|----|
| transcribe a single gene | 60 | s |
| translate a single copy of a protein | 100 | s |
| METABOLIC PROCESSES | | |
| enzyme-substrate binding | 1 | ms |
| single chemical transformation | 50 | ms |

Ensemble processes occurring at the cellular scale are of course slower, but keep the same separation of scales. This fact implies that on the time scale of interest for most functions of a cell, the metabolic processes have usually reached their steady state: the derivative of the concentration of each metabolite is null. In this balanced perspective, the network is viewed as an electrical circuit where at each node (metabolite) the Kirchhoff's law holds. Our previous equation can then be rewritten as an algebraic equation:

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & -2 & 0 \\ 0 & 0 & 3 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix},$$

which can be interpreted as a constraint for the value of the fluxes \mathbf{v} .

Therefore, for a given network composed of m metabolites and r reactions ($m = 3$ and $r = 4$ in our toy example) the steady-state assumption can be compactly written as

$$\mathbf{S}\mathbf{v} = \mathbf{0}, \tag{1.1}$$

where the $m \times r$ matrix \mathbf{S} is called "stoichiometric matrix" ($S_{i,j}$ represents the stoichiometric coefficient of the i -th metabolite in the j -th reaction).

Lower and upper bounds

Thermodynamical constraints can be added to each reaction: an upper bound $U_i > 0$, which is related to the maximal enzyme activity and to the availability of nutrients; a lower bound L_i , which states the reversibility ($L_i < 0$) or irreversibility ($L_i = 0$) of the reaction:

$$L_i \leq v_i \leq U_i, \quad \forall i = 1, \dots, r. \tag{1.2}$$

Definition of a cost functional

Constraints (1.1) and (1.2) generate a convex and bound set $W \subset \mathbb{R}^r$ (a polytope) to which the flux vector \mathbf{v} has to belong (see Fig. 1.4a and b). In order to obtain the point in W which describes the “effective” fluxes of the network an optimization procedure is usually adopted. The definition of the cost functional is a key ingredient of this approach and there are many works which attempt to define new heuristic functionals capable of generating realistic metabolic fluxes (see for instance [20, 21, 22, 23, 24]). We present here the most common criteria.

Maximizing growth rate for unperturbed network. The rationale behind this criterion comes from an evolutionary point of view of the role of metabolism [20]. As mentioned, metabolism is a machinery by which nutrients are converted in energy and in the building blocks for the macromolecules, i.e. aminoacids for proteins, sugars for complex carbohydrates, nucleotides for DNA and RNA, lipids for membranes, etc. These building blocks are the constituents of the cell mass (called “biomass”) and are necessary for the cell to duplicate and proliferate. Through a chemical analysis of the biomass, the average cell composition in terms of these precursors can be estimated [25]. The heuristic assumption is therefore that any metabolic change which allows the cell to better exploit the nutrients and more efficiently synthesize the biomass components would have been favored by natural selection. According to the formalism we presented, the “biomass production” can be expressed as a cost functional $\phi(\mathbf{v})$ defined as a linear combination of the fluxes describing the precursors biosynthesis:

$$\phi(\mathbf{v}) = \mathbf{b}^T \mathbf{v},$$

where \mathbf{b} contains the experimental coefficients obtained from the chemical analysis of the biomass, i.e. the amount of precursors that is needed for unit of biomass. Indeed, together with the stoichiometric matrix, also the vector \mathbf{b} is normally included in a database of metabolic networks. Therefore, in case of unperturbed cells the metabolic fluxes \mathbf{v}^* can be easily found by the following optimization:

$$\mathbf{v}^* = \arg \max_{\mathbf{w} \in W} \mathbf{b}^T \mathbf{w}, \quad (1.3)$$

(see Figure 1.4c) where $\arg \max_{x \in Q} f(x)$ stands for the set of x for which the function f attains its maximal value in Q (or equivalently, its minimal value for “arg min”).

Minimizing adjustment in a perturbed network. When a metabolic network is perturbed, e.g. through a gene knockout or through the action of a drug, the steady state \mathbf{v}^* from (1.3) is no longer realistic. The most famous criterion for treating perturbed networks is MOMA (see Fig. 1.4(d)). According to [21], when a gene has been knocked-out the cell reacts by trying to minimize the adjustment with respect to the unperturbed situation defined by (1.3). Using the Euclidean L^2 -norm $\|\cdot\|_2$ to quantify the entity of the adjustment, this criterion translates in

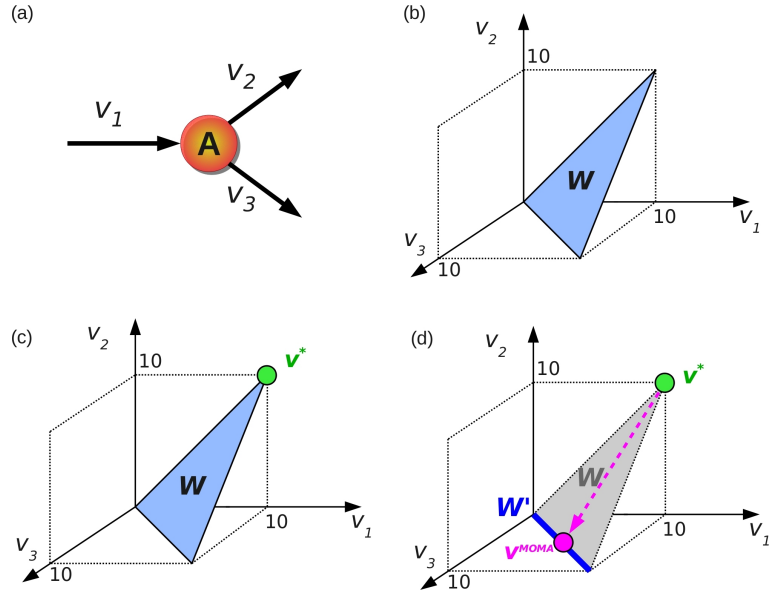


Figure 1.4: **Flux Balance Analysis.** (a): Sketch of a simple network where metabolite A is produced by reaction 1 and consumed by reactions 2 and 3; (b): Convex set (polytope W) defined by the constraints $v_1 - v_2 - v_3 = 0$ and $0 \leq v_i \leq 10$ for $i = 1, 2, 3$; (c): the green dot represents \mathbf{v}^* , the outcome of the optimization of the cost functional (biomass production) $\phi(\mathbf{v}) = 2v_2 + v_3$ (i.e. $\mathbf{b} = [0, 2, 1]^T$); (d): the perturbation of the network by removal of reaction 2 (knockout of the corresponding gene) reduces the set W to its subset W' . According to MOMA, the new optimum \mathbf{v}^{MOMA} is the orthogonal projection of the unperturbed solution \mathbf{v}^* on W' (magenta line and dot).

the following quadratic optimization problem

$$\mathbf{v}^{\text{MOMA}} = \arg \min_{\mathbf{w} \in W} \|\mathbf{w} - \mathbf{v}^*\|_2. \quad (1.4)$$

An alternative has been proposed in Ref. [22], where the minimization is performed on the number of adjusted reactions (ROOM), i.e. reactions whose flux adjustment with respect to \mathbf{v}^* is higher than a given threshold ε :

$$\mathbf{v}^{\text{ROOM}} = \arg \min_{\mathbf{w} \in W} \sum_i^r y_i, \quad (1.5)$$

where y_i is an integer variable equal to 1 if $|w_i - v_i^*| > \varepsilon$, 0 otherwise. Because of the integer nature of variables y_i , this approach belongs to the class of Integer Problems which have a higher computational complexity than Linear-Quadratic Programming.

Both heuristics provide reasonable predictions: while MOMA seems to provide a more reliable description of what happens shortly after a perturbation, ROOM seems to capture better a later (but not specified) stage in the metabolic adjustment.

1.3 Linear programming and duality theorem

The final step of FBA is an optimization problem. In particular, it consists of a convex optimization since both constraints and cost functional are linear (the convexity is preserved also in the case of MOMA based on the Euclidean L^2 norm). Therefore, linear optimization (known as Linear Programming, LP) can be straightforwardly applied. LP is based on very efficient algorithms which have been introduced since its development by Leonid Kantorovich in 1939. Kantorovich developed the earliest linear programming problems during the second World War to plan expenditures and returns of army in order to reduce costs. After the ending of the war, many industries used it in their daily planning. For military reasons, the method was kept secret until 1947 when George Dantzig published the *simplex method* and John von Neumann developed the theory of duality as a linear optimization solution, applying it in the field of game theory. Further details can be found in the books of J. Matoušek and B. Gärtner [26] and of A. Schrijver [27].

A standard way of describing a linear programming problem consists of the following three parts (in order to highlight the link with the FBA formulation presented above, the unknown variable is denoted by the vector $\mathbf{v} \in \mathbb{R}^r$):

- ⊙ **Constraints:** because of linearity, they can be written in general as a set of equalities

$$A_{\text{eq}}\mathbf{v} = \mathbf{c}_{\text{eq}},$$

and a set of inequalities

$$A_{\text{ineq}}\mathbf{v} \leq \mathbf{c}_{\text{ineq}}.$$

Without loss of generality, we can merge them in the following expression:

$$A\mathbf{v} \leq \mathbf{c}, \tag{1.6}$$

where $A = [A_{\text{eq}}^T, A_{\text{ineq}}^T]^T$ is a matrix and $\mathbf{c} = [\mathbf{c}_{\text{eq}}^T, \mathbf{c}_{\text{ineq}}^T]^T$ is a column vector.

- ⊙ **Nonnegative variables:** in standard form, only nonnegative values for the variables are considered:

$$v_i \geq 0 \quad \forall i = 1, \dots, r. \tag{1.7}$$

- ⊙ **Cost functional ϕ :** it is defined by the inner product with a given vector \mathbf{b} :

$$\phi(\mathbf{v}) = \mathbf{b}^T \mathbf{v}. \tag{1.8}$$

A linear programming problem may have no solution, one (and only one) solution, or an infinite number of solutions. This can be easily assessed by simple algorithms which scale polynomially with the size of the problem.

Table 1.2: **Dualization recipe:** nonnegativity and nonpositivity requirements for the variables are not counted among the constrains [26].

| | Primal problem | Dual problem |
|-----------------|--------------------------------|--------------------------------|
| Variables | v_1, v_2, \dots, v_r | w_1, w_2, \dots, w_s |
| Matrix | A | A^T |
| Right-hand side | \mathbf{c} | \mathbf{b} |
| Cost function | $\max \mathbf{b}^T \mathbf{v}$ | $\min \mathbf{c}^T \mathbf{w}$ |
| Constrains | i -th constraint has \leq | $w_i \geq 0$ |
| | i -th constraint has \geq | $w_i \leq 0$ |
| | i -th constraint has $=$ | $w_i \in \mathbb{R}$ |
| | $v_j \geq 0$ | j -th constraint has \geq |
| | $v_j \leq 0$ | j -th constraint has \leq |
| | $v_j \in \mathbb{R}$ | j -th constraint has $=$ |

Duality theory

Every linear programming problem, referred to as a *primal problem*, can be converted into a *dual problem* which provides an upper bound to the optimal value of the primal problem. In matrix form, if the primal problem is formulated in its standard form:

$$\text{Maximize } \mathbf{b}^T \mathbf{v} \quad \text{subject to} \quad A\mathbf{v} \leq \mathbf{c}, \quad \mathbf{v} \geq 0,$$

the corresponding symmetric dual problem is

$$\text{Minimize } \mathbf{c}^T \mathbf{w} \quad \text{subject to} \quad A^T \mathbf{w} \geq \mathbf{b}, \quad \mathbf{w} \geq 0,$$

where \mathbf{w} is the vector of dual variables.

Table 1.2 reports a summary of the rules for writing the dual of a generic formulation of the primal problem.

The importance of the dual problem lays in the Strong Duality Theorem of LP: [27]:

Theorem 1. *Let A be a matrix, and let \mathbf{b} and \mathbf{c} be vectors. Then*

$$\max\{\mathbf{b}^T \mathbf{v} \text{ such that } A\mathbf{v} \leq \mathbf{c}, \mathbf{v} \geq \mathbf{0}\} = \min\{\mathbf{c}^T \mathbf{w} \text{ such that } A^T \mathbf{w} \geq \mathbf{b}, \mathbf{w} \geq \mathbf{0}\},$$

provided that both sets are not empty.

This means that at the optimum, primal and dual cost functionals are equal. As a consequence, imposing the equality

$$\mathbf{b}^T \mathbf{v} = \mathbf{c}^T \mathbf{w}, \tag{1.9}$$

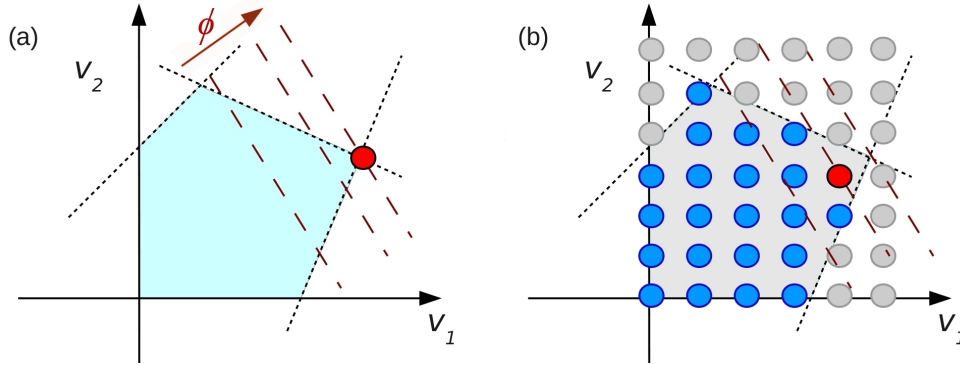


Figure 1.5: **LP vs MILP.** (a): Sketch of a LP maximization. This problem is the relaxed form of the MILP reported in (b). Relaxed solution (red dot) always lays on an edge of the polytope. (b): in MILP, gray dots are unfeasible points. The optimal MILP solution (red dot) cannot have a cost functional ϕ higher than the relaxed solution.

and adding the dual constraints to the primal constraints, the resulting polytope will be constituted only by the optimal solution of the primal problem.

In the following chapters this theorem is frequently applied in order to solve all bilevel optimizations (i.e. two nested optimizations) we obtain during the development of new algorithms.

Mixed Integer Linear Programming (MILP)

If some of the unknown variables are required to be integers, then the problem is called a Mixed Integer Linear Programming (MILP) problem (as mentioned ROOM belongs to this class, see Equation 1.5). In contrast to LP, these problems are generally NP-hard, i.e. the computational times scale more than any polynomial function of the number of integers. Advanced algorithms for solving integer linear programs include: (i) *cutting-plane method*; (ii) *branch and bound* (and its variations such as *branch and cut*, *branch and price* [28, 29]); (iii) *delayed column generation* (if the problem has some extra structure). One of the complications origins from the lack of convexity: indeed, the set of feasible points defined by the constraints is not convex (in particular it is not even connected since it is a set of isolated points; see Fig. 1.5b).

Normally, the optimization starts with the so-called “relaxed problem”: the LP relaxation of a $\{0,1\}$ integer program is the problem that arises by replacing the constraint where each variable must be 0 or 1 by a weaker constraint where each variable belong to the interval $[0, 1]$ (relaxed variables). The solution of the relaxed problem provides an upper (lower) bound on the maximized (minimized) cost functional (see Fig. 1.5). Then, depending on the procedure, the relaxed variables are converted one by one to integer. In the following we do not enter into the details of these algorithms: however, it is worth noting that the duality theorem does not apply to MILP. For this reason, approaches like ROOM, which refer to integer quantities cannot be used in any primal problem.

Chapter 2

Drug synergisms for human metabolic diseases and cancer

2.1 Background

In spite of the advances in molecular and computational biology, the discovery of new drugs still remains a challenging task which requires a very long period of research and development before any new compound can be commercialized. A possible alternative to the search of new active compounds is to make use of the unexploited properties of already available drugs, since a wide knowledge about both their therapeutic and toxicity effects has already been gathered during the study for their approval. In this perspective, a natural approach to broaden the range of applications of the existing drugs is to try to combine them in multiple drug therapies [30, 31, 32]. However, even though both the financial burden of conducting trials as well as the risk of adverse events in trial populations is expected to be sensibly lower for already approved drugs, so far the experimental investigations of multicomponent therapies have been quite limited [33, 34]. Major obstacles to this approach are the high number of possible combinations but also our limited understanding of the complex mode of action of a multidrug treatment. Indeed, multiple perturbations can show three types of interactions, which have been classified as synergistic, antagonistic and additive [35]. We focus our attention on the first type, where the use of drug combinations represents an enhancement with respect to the superposition of the single perturbations.

In order to identify synergistic effects, Ref. [33] experimentally investigated all pairs of a set of known drugs at different doses, obtaining a functional classification of the interactions by looking at their inhibitory effect. Various computational approaches, based on reconstructed genome-scale metabolic networks, have been also developed in Refs. [36, 37, 24, 38] for the identification of synergistic effects triggered by multiple drugs or multiple genetic perturbations (for example the so-called “synthetic lethality”). Unfortunately, a systematic evaluation of the effects of all possible combinations of drugs is unfeasible, because their number scales exponen-

tially with the number of chemicals taken into account (for an exhaustive search over 40 drugs, more than a trillion possible combinations should be tested). Moreover, a drug profile is given by both its therapeutic effect and by its side effect, the latter being related to its selectivity. For example, drugs such as anticancer agents have to selectively act only on tumor cells [39]. Similarly, metabolic diseases are induced by the imbalance of key metabolic pathways which, if modulated without affecting other vital functions, can rescue from the pathology [40, 41].

Referring to these specific needs, we have developed an algorithm based on the metabolic network of humans and on the comparison between the metabolic networks of human and cancer cells (as reconstructed recently in [2]) aimed at expanding the spectrum of applications of the existing drugs to new selective multidrug therapies. In particular, the algorithm has been applied to two different case studies:

- ⊙ finding drug synergisms for **metabolic diseases** (like diabetes, obesity and hypertension) on the human network [16];
- ⊙ finding **antitumoral** drug combinations (i.e. that are lethal for the cancer metabolic network) with minimal side effect on the normal human cell.

The computational method is based on FBA and relies on a bilevel optimization which, after reformulation through duality theory, allows the algorithm to efficiently search the interactions between drugs. Inspired by works such as Refs. [21, 22], we treat the inhibition of a metabolic reaction induced by a drug as the silencing of the gene which codes for the catalyzing enzyme (i.e. the simple ON/OFF description of the drug effect is adopted). In spite of this simplified framework, the synergistic effect resulting from multiple perturbations of the metabolic network is still well captured [37, 3] (see Figure 2.1 for a toy example). The selectivity of any drug treatment is correlated to its side effect, which is estimated as the number of stopped reactions (see Figure 2.1(c) *vs* (d)) plus a correction term for the known non-metabolic targets. Further details are described in the Section 2.2. It is worth noting that, because of the steady-state assumption of FBA, this formalism does not identify synergisms between drugs that manifest themselves as alterations of kinetic parameters and consequently of concentration of metabolites, like the case of drugs which act on the same linear path (as e.g. Trimethoprim and Sulfamethoxazole do on folate synthesis [42]). Similarly, interactions where for instance one drug inhibits the biodegradation of the others cannot be found by FBA-based methods [43].

2.1.1 State of the art

With respect to the available literature the procedure, we are proposing presents at least three important differences: (i) the synergisms are efficiently explored over all drug combinations without limiting only to pairwise combinations but without doing an exhaustive search, thanks to the application of duality theory; (ii) the multiple drug treatments suggested by the method guarantee both the inhibition of the chosen target (efficacy) and a minimal side effect on the other cellular functions

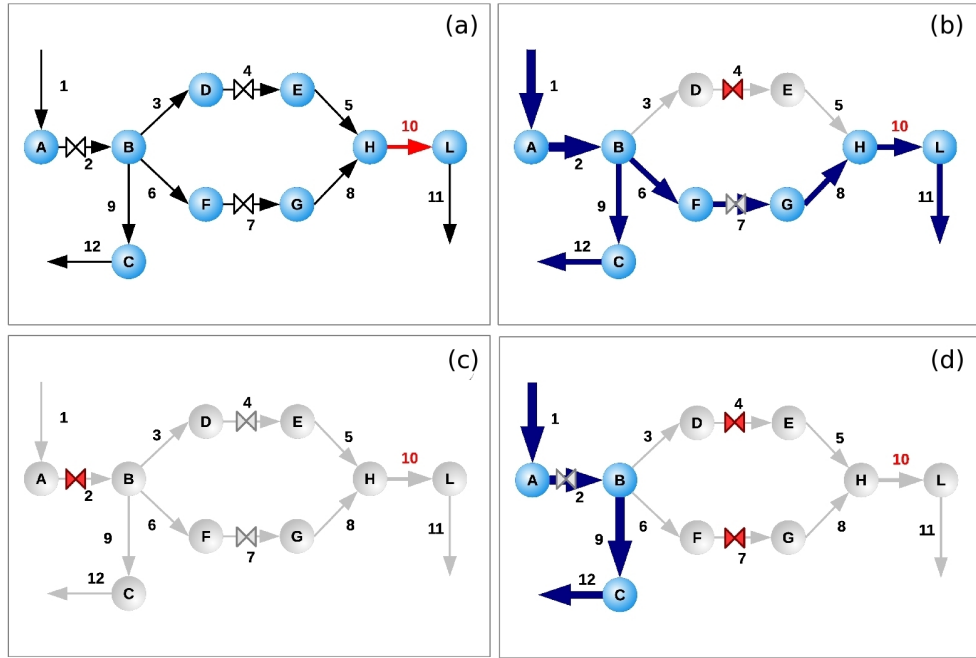


Figure 2.1: **Example of drug synergism in FBA.** For the toy network depicted in (a) the aim is to stop the objective reaction v_{10} (in red) by choosing a combination of drugs (the three valves “ \times ”) while blocking the minimum number of reactions other than v_{10} . In the drawing, blue arrows indicate active fluxes while gray arrows refer to stopped reactions; the valve is red if the drug is used, gray otherwise. Panel (b) shows how the use of a single drug (v_4 or v_7) does not stop the objective reaction, while the drug at v_2 blocks the objective reaction v_{10} but it also blocks all fluxes of the network (panel (c)). Therefore, the optimal drug combination blocking the objective function v_{10} with minimal side effect is given by the synergism of the two drugs acting at v_4 and v_7 (panel (d)). The comparison of panels (b) and (d) shows how a synergism is a behavior which cannot be simply inferred by the superposition of the effects of the single drugs, but that structurally depends from the topology of the network.

(selectivity); (iii) in our procedure, any metabolic process of the network can be chosen as possible disease and phenotype readout, not only cell growth as common in the FBA literature.

We present here a quick overview of the literature and of the main computational methods which investigate drug synergisms (or synergisms induced by other network perturbations, like gene knockout). We list them from the most similar to the most different with respect to our algorithm, underlying the main discrepancies.

OPTKNOCK AND OPTORF [24, 44]: the two works aim to find the best knockout (simple or multiple, up to a limit in cardinality fixed by the user) that maximizes the biosynthesis of a given metabolite while optimizing the biomass production. Both methods are based on FBA (OptOrf includes also some transcriptional reg-

ulations). The bilevel optimization contained in both algorithms has been reformulated through the duality theory, which assures the computational efficiency of the method. Indeed, we have been inspired by these works in developing of our algorithm.

ESSENTIALITY AND SYNTHETIC LETHALITY [36]: Through an algorithm which is similar to Optknock, this work identifies pairs, triple and some high-order gene deletions which are lethal in the *Escherichia coli* metabolism. Also in this method, the optimization is based on the biomass and does not consider any alternative objective reaction.

OPMET [38]: the aim is to search the gene knockout (of any cardinality) which stops a given objective reaction while inducing the minimum damage on the network (estimated as number of stopped reactions and unavailable metabolites). The approach used to model the network is similar to FBA but with a weaker assumption on the steady state: for instance, accumulation of metabolites is allowed. The search on the space of the combinations is performed dynamically through a branch-and-bound algorithm, refined with two filtering strategies. Properties of LP, like duality theory, which are useful for the efficiency of the algorithm have not been exploited.

EPISTASIS IN HUMAN METABOLISM [45]: gene epistasis has been here investigated through FBA formalism; no limitation on the cardinality of the gene deletions is imposed. However, since the search is based on an approximate determination of the elementary modes (called “pathway fragment generation”), solutions are suboptimal. An exact calculation of the elementary modes would have been computationally much more expensive and almost prohibitive for a large network such as the human. Moreover, no information on the side effect on the whole network is included.

EPISTASIS IN YEAST METABOLISM [37]: epistatic interactions are here studied performing an exhaustive search of all pairs of gene knockouts on the *S.cerevisiae* network modeled according to FBA; the effect is characterized through an epistasis indicator which quantifies the change of the biomass production of the multiple perturbation with respect to the two single perturbations. No objective reaction other than growth rate has been considered.

DRUG TARGETS IN CANCER [2]: also in this case only an exhaustive evaluation of all pairs of knockouts has been carried out; moreover, the side effect of the potential antitumoral treatments on the normal human cell is estimated in terms of ATP production, a central process in the metabolism. However, according to this evaluation criterion, all the antitumoral solutions we discuss below would induce the same impact on the human network, meaning that this definition of side effect is not able to discriminate among different drug treatments, while our definition does (see Table 2.5 in the following). Indeed, instead of considering a single reaction, our approach estimates the impact as a global loss of cellular functions, which in our opinion is reasonable since no tissue-specific network of the human metabolism is available at the moment.

2.2 Materials and Methods

2.2.1 The algorithm: bilevel optimization

The drug-synergism algorithm is developed in the framework of FBA; in particular, we suppose to have a metabolic network (with m metabolites and r reactions) and a set of d drugs which inhibit some reactions of this network (the set of targets of the k -th drug is indicated by \mathcal{T}_k). By decomposing any reversible reaction in a couple of irreversible reactions, we can always assume that fluxes have non-negative values. Then, the optimal synergism problem can be stated as follows:

Problem: *Given:*

- a metabolic network, which means a stoichiometric matrix $\mathbf{S} \in \mathbb{R}^{m \times r}$ and a vector of fluxes \mathbf{v} with upper-bounds \mathbf{U} , both laying in \mathbb{R}^r ;
- an objective reaction flux v_{obj} which has to be stopped;
- the set of d drugs together with their inhibition targets $\{\mathcal{T}_k\}_{k=1,\dots,d}$;

we want to find the subset of drugs such that it blocks v_{obj} causing the minimal side effect, i.e. a minimum perturbation on the overall reaction fluxes.

Of course, we are not interested in procedures which perform an exhaustive search in the space of all drugs combinations.

The space W of all possible steady-state fluxes \mathbf{v} as defined by FBA is

$$W := \{\mathbf{v} : \mathbf{S}\mathbf{v} = \mathbf{0}, \quad 0 \leq v_j \leq U_j \quad \forall j = 1, \dots, r\},$$

where v_j and U_j are the j -th components of the vectors \mathbf{v} and \mathbf{U} respectively. Given the inhibition targets of each drug, we assume that a drug inhibits completely the enzymes responsible for the targeted reactions, hence stopping the relative fluxes (the ON/OFF approach). Therefore, using a drug combination on the network means forcing (directly or indirectly) to zero some of the fluxes. We define a drug combination by means of the ‘‘inhibition vector’’ $\mathbf{h} \in \{0, 1\}^d$ such that

$$h_k = \begin{cases} 0 & \text{if drug } k \text{ induces no inhibition of its targets, i.e. drug } k \text{ is not used;} \\ 1 & \text{if drug } k \text{ induces 100\% inhibition of its targets, i.e. drug } k \text{ is used;} \end{cases}$$

Then, if the k -th drug targets the j -th reaction we write:

$$v_j \leq U_j(1 - h_k).$$

This reduces the space of feasible steady state fluxes to a subset $W(\mathbf{h}) \subset W$ (notice that $W(\mathbf{0}) \equiv W$).

We can now introduce a definition which quantifies the *side effect*, $\sigma(\mathbf{h})$, of a drug combination \mathbf{h} on the reaction fluxes. Clearly, there are many possible definitions. If we were dealing with microorganisms we could adapt to our situation the analogous MOMA or ROOM criteria [21, 22] modeling the perturbation on the fluxes (induced by knockout of one or more enzymes) with respect to the “wild type” fluxes. This requires, however, to know the unperturbed fluxes (the wild type reference) which for a microorganism corresponds to using the biomass production as cost functional of the FBA problem [20]. Unfortunately, for human metabolic network such a commonly accepted FBA criterion is unavailable, because human cells are normally not assumed to be maximizing their growth rate (they do not proliferate). Hence, an easy way to determine the metabolic fluxes for the unperturbed network (“wild type”) is not immediately available. Consequently, also the calculation of the perturbed fluxes and the quantification of the side effect are more ambiguous. Our choice in this work is to quantify the side effect as number of stopped reactions (more precisely, the number of reactions that cannot take place because of the drugs). This bypasses the lack of the reference fluxes and allows to quantify the number of cellular functions which are no longer available, regardless of the type of tissue to which the human cell belongs.

Since the available drugs we selected do not have only metabolic targets, a measure of the side effect based exclusively on metabolic reactions disregards the perturbation induced by the drugs on other cellular functions (for example signaling cascades, protein synthesis, etc). This additional information can be incorporated in the model weighting each drug variable h_k in the objective function according to its non-metabolic effects. Following this approach, we define:

$$\sigma(\mathbf{h}) := \min_{\mathbf{v} \in W(\mathbf{h})} \sum_{j=1}^r (1 - y_j) + \sum_{k=1}^d \beta_k h_k; \quad (2.1)$$

where the parameter β_k is an estimation of the non-metabolic perturbation induced by drug k and y_j is a binary variable such that $y_j = 0$ when the flux of reaction j is lower than a threshold ε ($\varepsilon = 0.1$ in this work); this threshold effect on the fluxes can be inserted in our problem by appending the following linear inequalities to the set of constraints of the problem:

$$\begin{aligned} \varepsilon y_j &\leq v_j & \forall j = 1, \dots, r; \\ U_j y_j &\geq v_j & \forall j = 1, \dots, r. \end{aligned}$$

In order to avoid a double count, for reversible reactions (split in two opposite fluxes) we impose also the additional constraint

$$y_j + y_l \leq 1 \quad \forall j \text{ and } l = 1, \dots, r \text{ such that } v_j \text{ and } v_l \text{ are opposite fluxes.}$$

Although the choice of the weights β_k is quite arbitrary, we have tried to evaluate both terms of (2.1) using a homogeneous criterion; this is achieved by relating

the values of β_k to the number of known non-metabolic targets of each drug. In particular we set:

$$\beta_k := \bar{\beta} \cdot [\text{number of non-metabolic targets of drug } k],$$

where the numerical coefficient $\bar{\beta}$ aims to capture both the direct and the indirect (i.e. propagated) effect of the perturbation across the non-metabolic functions of the cell. However, because of the lack of a model for the description of the propagation of this perturbative effect, we perform the estimation referring to the metabolic networks. In this perspective, $\bar{\beta}$ is estimated by the ratio between the total number of metabolic reactions that are stopped (directly and indirectly) by a drug and the total number of targets inhibited (directly) by the same drug. Referring to the human metabolic network and averaging over all drugs we selected, we obtain $\bar{\beta} = 7.7$.

The optimal solution can be described as follows:

Solution: For any subset of drugs (i.e. for any vector \mathbf{h}), we can find the set $W(\mathbf{h})$ and the perturbation $\sigma(\mathbf{h})$. By restricting to drug combinations which guarantee the inhibition of the objective reaction, i.e. such that $\max_{v \in W(\mathbf{h})}(v_{obj}) = 0$, the optimal solution \mathbf{h}^* is

$$\mathbf{h}^* = \arg \min_{\substack{\mathbf{h} \in \{0,1\}^d : \\ \max_{v \in W(\mathbf{h})}(v_{obj}) = 0}} \sigma(\mathbf{h}). \quad (2.2)$$

The resulting bilevel optimization is a min-max integer linear programming problem [46]. The inner problem adjusts the fluxes in order to achieve the *maximum* flow for the objective reaction when all fluxes are subjected to the inhibitions (drugs) imposed by the outer problem and to the stoichiometric constraints. The outer problem selects the combination of drugs which *minimizes* the side effect, restricting to those solutions of the inner problem which guarantee no flow for the objective reaction.

The bilevel optimization problem is the following:

$$\text{Minimize } \sum_{j=1}^r (1 - y_j) + \sum_{k=1}^d \beta_k h_k + b \sum_{k=1}^d h_k \quad \text{“outer problem”}$$

such that

$$\left[\begin{array}{ll} \text{Maximize } & v_{obj} \quad \text{“inner problem”} \\ \text{such that} & \\ & \sum_{j=1}^r S_{i,j} v_j = 0 \quad \forall i = 1, \dots, m \\ & v_j \leq U_j \quad \forall j = 1, \dots, r \\ & v_j \leq U_j (1 - h_k) \quad \forall j = 1, \dots, r, j \in \mathcal{T}_k; \end{array} \right]$$

$$v_{obj} = 0;$$

$$\varepsilon y_j \leq v_j \quad \forall j = 1, \dots, r;$$

$$U_j y_j \geq v_j \quad \forall j = 1, \dots, r;$$

$$y_j + y_l \leq 1 \quad \forall j = 1, \dots, r \text{ such that } v_l \text{ is the opposite flux of } v_j,$$

where the parameter $b \ll 1$ ($b = 0.001$ in this work) is introduced in the objective function of the outer problem in order to exclude the combinations containing redundant inhibitions and to avoid an “over-selection” of drugs.

To solve this bilevel optimization we apply the strong duality theorem (1.9) which consists in appending a list of constraints corresponding to the dual of the inner problem and setting the primal objective function equal to the dual. This leads to a single minimization problem. Dual variables are identified as follows (see Table 1.2 for details on their range of values): $\mu_1, \dots, \mu_m \in \mathbb{R}$ are the dual variables associated to the first m constraints of the inner problem (steady state); $\lambda_1, \dots, \lambda_r \in \mathbb{R}_+$ ($t := \sum_{j=1}^r |\mathcal{T}_j|$) are the dual variables associated to the second set of constraints of the inner problem (upper bounds); and $\delta_1, \dots, \delta_t \in \mathbb{R}_+$ are associated to drugs targets, i.e. to the third set of constraints of the inner problem.

The final optimization problem becomes:

$$\begin{aligned}
& \text{Minimize} && \sum_{j=1}^r (1 - y_j) + \sum_{k=1}^d \beta_k h_k + b \sum_{k=1}^d h_k \\
& \text{such that} && \\
& \sum_{j=1}^r S_{i,j} v_j = 0 && \forall i = 1, \dots, m; \\
& v_j \leq U_j && \forall j = 1, \dots, r; \\
& v_j \leq U_j (1 - h_k) && \forall j = 1, \dots, r, j \in \mathcal{T}_k; \\
& \sum_{i=1}^m S_{i,j} \mu_i + \lambda_j + \sum_{i=1}^{N_{t,j}} \delta_i \geq 0 && \forall j = 1, \dots, r, j \neq \{\text{obj}\}; \\
& \sum_{i=1}^m S_{i,j} \mu_i + \lambda_j + \sum_{i=1}^{N_{t,j}} \delta_i \geq 1; && \text{for } j = \text{obj}; \\
v_{obj} = & \sum_{i=1}^r U_i \left[\lambda_i + \delta_i \sum_{k \in \mathcal{T}_i} (1 - h_k) \right]; && (2.3) \\
& v_{obj} = 0; \\
& \varepsilon y_j \leq v_j && \forall j = 1, \dots, r; \\
& U_j y_j \geq v_j && \forall j = 1, \dots, r; \\
& y_j + y_l \leq 1 && \forall j = 1, \dots, r \mid v_l \text{ is the opposite flux of } v_j.
\end{aligned}$$

The key simplification is that the nonlinear terms $\delta_i h_k =: z_{ik}$ in (2.3) (the strong duality theorem equality) are exactly linearizable as follows:

$$\begin{aligned}
0 & \leq z_{ik} \leq \delta_i^{max} h_k \\
\delta_i - \delta_i^{max} (1 - h_k) & \leq z_{ik} \leq \delta_i
\end{aligned}$$

where δ_i^{max} is the upper bound for the dual variable δ_i (chosen arbitrarily high in the implementation).

2.2.2 Algorithm for competing organisms

The following version of the algorithm allows us to study the selectivity problem in a multi-network context, in which we would like to stop an objective reaction belonging to a first organism (in our case the proliferation rate w_{biomass} of a tumoral cell whose fluxes are denoted here by \mathbf{w} , stoichiometric matrix by $\tilde{\mathbf{S}} \in \mathbb{R}^{\tilde{m} \times \tilde{r}}$ and upper bounds by $\tilde{\mathbf{U}}$) while having a minimal effect on a second metabolic network (a normal human cell in our case). Since the two cells live in the same environment they experience the same drug inhibitions. The modified bilevel optimization problem is as follows:

$$\begin{array}{ll}
 \text{Minimize} & \sum_{i=1}^r (1 - y_i) + \sum_{k=1}^d \beta_k h_k + b \sum_{k=1}^d h_k & \text{“outer problem”} \\
 \text{such that} & & \\
 & \left[\begin{array}{ll}
 \text{Maximize} & w_{\text{biomass}} & \text{“inner problem”} \\
 \text{such that} & & \\
 & \sum_{i=1}^{\tilde{r}} \tilde{S}_{l,i} w_i = 0; \\
 & w_i \leq \tilde{U}_i; \\
 & w_i \leq \tilde{U}_i (1 - h_k); \\
 \sum_{j=1}^r S_{i,j} v_j = 0; \\
 v_j \leq U_j; \\
 v_j \leq U_j (1 - h_k); \\
 w_{\text{biomass}} = 0; \\
 \varepsilon y_j \leq v_j; \\
 U_j y_j \geq v_j.
 \end{array} \right]
 \end{array}$$

The constraints on the second network ($\mathbf{S}\mathbf{v} = \mathbf{0}$, thermodynamical bounds $v_j \leq U_j$ and drugs inequalities $v_j \leq U_j(1 - h_k)$) are located only in the outer problem. Therefore, they do not interfere with the inner problem and with the application of the duality theorem for the first network. This separation guarantees the constraint $\max(w_{\text{biomass}}) = 0$ and the optimality of the solution. Of course, also here the strong duality theorem is applied and the bilevel optimization is converted to a single step minimization.

The code for the algorithm has been developed in MATLAB (MathWorks R2010b). All MILPs have been performed using the ILOG-IBM CPLEX 12.1, under free academic license.

2.2.3 Metabolic networks

Two metabolic networks are used in this study: (i) human network developed in [16]; (ii) cancer network reconstructed in [2]. Some features of these two metabolic networks are listed in Table 2.1, together with the number of drugs currently approved.

Table 2.1: **Features of used metabolic networks.** The human network has been obtained from BIGG (bigg.ucsd.edu/), whereas the cancer network has been provided us by the authors of [2]. The number of reactions here reported is before the splitting of every reversible process in a pair of irreversible reactions.

| | Human [16] | Cancer [2] |
|-------------------------------|------------|------------|
| Number of reactions (r) | 2469 | 940 |
| Number of metabolites (m) | 1587 | 654 |
| Number of compartments | 8 | 8 |
| Number of pathways | 83 | 62 |
| Number of drugs (d) | 85 | 55 |

2.2.4 Selection procedure for the drugs

In order to generate realistic solutions, the available information about the existing drugs has been carefully filtered, selecting only inhibitions of metabolic human targets that have been experimentally proven (see Table 2.2). On the DrugBank database (www.drugbank.ca [47]) the following query is adopted:

1. the whole database contains 6708 drugs;
2. only *approved* drugs have been picked out, restricting the search to 1570 drugs;
3. we filter for drugs which act on *human* enzymes (identified by the EC number): the set reduces to 473 drugs;
4. we select only drugs for which an *inhibitory* effect on at least one enzyme of the human metabolism has been experimentally proven. The set reduces to 267 drugs (with the EC numbers of the inhibited enzymes). For these drugs, we count the number of non-metabolic targets, including also cases of agonism, antagonism and activation, since they all represent a perturbation to the regular functioning of the target (activation is not considered for metabolic targets because it does not affect the number of stopped reaction in our FBA formulation);
5. the reactions of the metabolic network directly inhibited by each drug are identified through the available correspondence between EC numbers of the inhibited enzymes and the gene codes first, and then through the correspondence between gene codes and metabolic reactions. During this step, it may happen that many genes, and hence many reactions, are associated to the same EC number. Therefore, although in the original database a drug inhibits only a single or a few targets, the number of metabolic reactions affected by the drug can be high. For instance, this is the case of Rosiglitazone (drug #7 in Table 2.2) which inhibits a single target, the long-chain-fatty-acid-CoA ligase an enzyme responsible for the binding of the acyl-CoA group to a long fatty acid chain. Since the substrate of this enzyme can be any carbon chain, regardless of the unsaturation (presence or not of double bonds C=C) and of the exact length (it just requires a chain longer than 12 carbons atoms), we end up with a drug which inhibits up to 60 metabolic targets;

Table 2.2: List of drugs selected from DrugBank database for human and cancer metabolic networks. Legend: M: number of metabolic targets; NM: number of non-metabolic targets; R: number of inhibited reactions of the human metabolic network; S.E.: side effect (according to Eq. (2.1)). Drugs marked with “X” act also on the cancer network. (*): Simvastatin and Atorvastatin are also present in this group (they induce the same metabolic inhibitions as Pravastatin but they have one and two non-metabolic targets, respectively).

| # | Drug name | M | NM | R | S.E. | cancer | # | Drug name | M | NM | R | S.E. | cancer |
|----|--------------------|---|----|----|-------|--------|----|----------------------|----|----|----|-------|--------|
| 1 | Vigabatrin | 1 | 1 | 3 | 17.7 | X | 41 | Pemetrexed | 4 | 0 | 8 | 15.0 | X |
| 2 | Chlormerodrin | 1 | 1 | 1 | 10.7 | X | 42 | Mycophenolic acid | 2 | 0 | 1 | 4.0 | X |
| 3 | Fenoprofen | 2 | 0 | 2 | 10.0 | X | 43 | Naftifine | 1 | 0 | 1 | 34.0 | X |
| 4 | Pravastatin(*) | 1 | 1 | 1 | 34.0 | X | 44 | Scopolamine | 1 | 5 | 3 | 42.5 | X |
| 5 | Maseprocol | 1 | 0 | 2 | 13.0 | | 45 | Sclegline | 2 | 0 | 5 | 23.0 | |
| 6 | Carbidopa | 1 | 0 | 5 | 20.0 | | 46 | Hexachlorophene | 1 | 1 | 4 | 9.7 | X |
| 7 | Rosiglitazone | 1 | 1 | 60 | 252.7 | X | 47 | Phenelzine | 6 | 1 | 9 | 22.7 | X |
| 8 | Zanamivir | 1 | 0 | 2 | 8.0 | | 48 | Sulfasalazine | 4 | 3 | 10 | 55.1 | X |
| 9 | Papaverine | 4 | 0 | 2 | 2.0 | | 49 | Minaprine | 2 | 7 | 6 | 77.9 | X |
| 10 | Udenafil | 1 | 0 | 3 | 5.0 | | 50 | Pentoxifyline | 4 | 2 | 21 | 43.4 | X |
| 11 | Methotrexate | 1 | 0 | 2 | 5.0 | X | 51 | Ribavirin | 3 | 1 | 5 | 16.7 | X |
| 12 | Pantoprazole | 1 | 0 | 1 | 1.0 | | 52 | Phenylbutazone | 3 | 0 | 2 | 10.0 | |
| 13 | Methoxyflurane | 3 | 4 | 1 | 31.8 | | 53 | Disulfiram | 2 | 0 | 9 | 16.0 | X |
| 14 | Dorzolamide | 2 | 0 | 4 | 4.0 | X | 54 | Zonisamide | 17 | 14 | 9 | 134.8 | X |
| 15 | Diethylcarbamazine | 3 | 0 | 4 | 23.0 | | 55 | Acetylsalicylic acid | 3 | 0 | 8 | 74.0 | X |
| 16 | Hydroxyurea | 4 | 0 | 4 | 10.0 | X | 56 | Ezetimibe | 2 | 1 | 2 | 11.7 | X |
| 17 | Carmustine | 1 | 0 | 2 | 3.0 | | 57 | Dipyridamole | 4 | 0 | 7 | 9.0 | X |
| 18 | Theophylline | 4 | 3 | 3 | 28.1 | | 58 | Mercaptopurine | 1 | 0 | 2 | 2.0 | X |
| 19 | Acarbose | 3 | 1 | 9 | 26.7 | | 59 | Auranofin | 1 | 1 | 3 | 10.7 | X |
| 20 | Floxuridine | 1 | 0 | 1 | 1.0 | X | 60 | Tioconazole | 1 | 0 | 1 | 34.0 | X |
| 21 | Valproic Acid | 2 | 1 | 5 | 30.7 | X | 61 | Trichlormethiazide | 4 | 1 | 5 | 12.7 | X |
| 22 | Indomethacin | 4 | 3 | 7 | 48.1 | X | 62 | Corulenin | 4 | 0 | 7 | 7.0 | X |
| 23 | Nitisione | 1 | 0 | 1 | 5.0 | | 63 | Mimosine | 2 | 1 | 5 | 15.7 | X |
| 24 | Droxidopa | 1 | 9 | 1 | 70.3 | X | 64 | Perhexiline | 2 | 0 | 52 | 113.0 | X |
| 25 | Physostigmine | 1 | 0 | 1 | 1.0 | | 65 | Orlistat | 2 | 1 | 10 | 22.7 | X |
| 26 | Carbamazole | 1 | 0 | 5 | 7.0 | | 66 | Leflunomide | 1 | 3 | 3 | 42.1 | X |
| 27 | Miglustat | 1 | 0 | 1 | 5.0 | | 67 | Trilestane | 2 | 2 | 2 | 18.4 | |
| 28 | Allopurinol | 1 | 0 | 3 | 11.0 | | 68 | Aminoguanidine | 1 | 0 | 2 | 4.0 | |
| 29 | Trimethoprim | 2 | 0 | 3 | 5.0 | X | 69 | Atovaquone | 3 | 0 | 1 | 6.0 | X |
| 30 | Gemcitabine | 3 | 0 | 85 | 64.0 | X | 70 | Carvedilol | 1 | 8 | 1 | 62.6 | X |
| 31 | Acetyldigitoxin | 1 | 0 | 1 | 1.0 | X | 71 | Desipramine | 1 | 12 | 1 | 98.4 | X |
| 32 | Cefdinir | 1 | 0 | 2 | 5.0 | X | 72 | Arsenic trioxide | 1 | 6 | 1 | 56.2 | X |
| 33 | Azelacic Acid | 2 | 1 | 3 | 14.7 | | 73 | Cyclizine | 1 | 1 | 2 | 10.7 | X |
| 34 | Pentostatin | 1 | 0 | 4 | 4.0 | X | 74 | Dazoxiben | 1 | 0 | 1 | 6.0 | |
| 35 | Diclofenac | 4 | 4 | 6 | 54.8 | X | 75 | Fomepizole | 4 | 0 | 15 | 33.0 | X |
| 36 | Quinacrine | 2 | 1 | 4 | 21.7 | X | 76 | Clomipramine | 1 | 5 | 4 | 57.5 | X |
| 37 | Monobenzene | 1 | 0 | 3 | 7.0 | | 77 | Bepitridil | 2 | 7 | 3 | 56.9 | X |
| 38 | Sulindac | 3 | 3 | 9 | 46.1 | X | 78 | Lithium | 3 | 2 | 5 | 46.4 | X |
| 39 | Amodiaquine | 1 | 0 | 1 | 5.0 | | 79 | Fidarestat | 1 | 0 | 7 | 13.0 | X |
| 40 | Thidronate | 1 | 1 | 3 | 13.7 | X | 80 | Carbenoxolone | 1 | 0 | 4 | 2.0 | X |
| | | | | | | | 81 | Canaline | 1 | 0 | 2 | 1.0 | X |
| | | | | | | | 82 | Myo-Inositol | 2 | 0 | 1 | 29.0 | X |
| | | | | | | | 83 | Voglibose | 1 | 0 | 4 | 7.0 | X |
| | | | | | | | 84 | Icatibant | 1 | 1 | 2 | 10.7 | X |
| | | | | | | | 85 | Tyloxapol | 1 | 0 | 2 | 6.0 | X |

6. drugs which have exactly the same metabolic targets are grouped; the final 85 groups are listed in Table 2.2. For each group only a representative is reported, namely the drug which has the minimal number of targets outside the metabolism.

2.3 Human metabolic diseases: results

In this section, we consider the inhibition of specific functions of human metabolism obtained without impairing other vital processes. First, the inhibitory effect of each single drug on the whole network has been calculated. Then, the following screening is performed: we systematically consider each reaction of the network as a potential objective function and we apply the algorithm, searching for the most selective synergism capable of blocking this objective reaction. Comparing this solution with the single-drug effects evaluated in advance, we distinguish three cases:

- a) the drug combination leads to a *new inhibition* since no single drug can stop the objective reaction;
- b) the objective reaction can be stopped also by a single drug but the drug combination is *more selective* (has a minor side effect);
- c) the objective reaction can be stopped also by a single drug and the multiple drug solution is *less selective* (this solution is not interesting because it triggers a larger side effect).

In all cases where a single or a multidrug solution is found, also all suboptimal solutions are hierarchically identified, iterating the procedure while excluding the current optimum, until the problem becomes unfeasible (i.e. no more solutions exist, capable of blocking that objective reaction). At the end of the screening, we obtained a set of 32 multicomponent solutions, ranging from combinations of two up to four compounds (see Table 2.3).

The following characterization of the synergistic effects is performed. For each combination we identify the set Y of metabolic reactions which cannot be stopped by any single drug of the combination, but which are stopped when all these drugs are used together. Then, the synergism is described by the vector $\mathbf{s} \in \{0, 1\}^r$, where $s_j = 1$ if reaction j belongs to Y . From the vectors \mathbf{s} of the 32 multiple drug solutions a matrix of distances is constructed and a cluster analysis is performed on these distances; the resulting distance-based tree (similar to a phylogenetic tree) is drawn in Figure 2.2 (upper panel). The synergisms are clustered in six classes (with clearly identifiable subclasses in some of them) labeled from “A” to “F”. This classification can be used to build also a proximity network for the drugs, linking those that belong to the same synergistic interaction. The outcome is drawn in Figure 2.2 (lower panel) and shows that the same clustering applies to the drugs involved in the synergisms. The result highlights how drugs can often be used in alternative one to the others: for example, the synergistic pairs of class C contain one drug among those labeled with the number 7 or 55 (Rosiglitazone or Acetylsalicylic acid)

Table 2.3: **Multiple drug solutions in the human metabolic network.** For each solution, the table reports the side effect $\sigma(\mathbf{h})$, the synergism size (i.e. the number of stopped reactions which exceeds the linear superposition of single drug effects), the ratio between these two quantities, and their classification (see Figure 2.2 and text for the clustering analysis). Drug numbers refer to Table 2.2. Bold font indicates that the solution (or part of it) has an experimental validation in literature.

| Drugs | Side eff. | Syn. | ratio | Class |
|--|-----------|------|-------|-------|
| Rosiglitazone (#7) - Quinacrine (#36) - Cerulenin (#62) - Tyloxapol (#85) | 363.8 | 91 | 25.0% | A |
| Rosiglitazone (#7) - Quinacrine (#36) - Orlistat (#65) - Tyloxapol (#85) | 377.7 | 91 | 24.0% | A |
| Rosiglitazone (#7) - Indomethacin (#22) - Cerulenin (#62) - Tyloxapol (#85) | 390.6 | 91 | 23.2% | A |
| Rosiglitazone (#7) - Diclofenac (#35) - Cerulenin (#62) - Tyloxapol (#85) | 397.5 | 91 | 22.8% | A |
| Rosiglitazone (#7) - Indomethacin (#22) - Orlistat (#65) - Tyloxapol (#85) | 404.5 | 91 | 22.4% | A |
| Rosiglitazone (#7) - Diclofenac (#35) - Orlistat (#65) - Tyloxapol (#85) | 411.4 | 91 | 22.1% | A |
| Rosiglitazone (#7) - Cerulenin (#62) | 298.9 | 52 | 17.3% | A |
| Rosiglitazone (#7) - Orlistat (#65) - | 312.8 | 52 | 16.6% | A |
| Indomethacin (#22) - Fomepizole (#75) | 84.7 | 1 | 1.1% | B |
| Naftifine (#43) - Acetylsalicylic acid (#55) | 116.0 | 6 | 5.1% | C |
| Acetylsalicylic acid (#55) - Tioconazole (#60) | 116.0 | 6 | 5.1% | C |
| Simvastatin/Pravastatin (#4) - Acetylsalicylic acid (#55) | 123.9 | 6 | 4.8% | C |
| Rosiglitazone (#7) - Tioconazole (#60) | 280.9 | 6 | 2.1% | C |
| Rosiglitazone (#7) - Naftifine (#43) | 280.9 | 6 | 2.1% | C |
| Simvastatin/Pravastatin (#4) - Rosiglitazone (#7) | 288.8 | 6 | 2.0% | C |
| Carbidopa (#6) - Droxidopa (#24) | 93.1 | 1 | 1.0% | D |
| Droxidopa (#24) - Selegiline (#45) | 96.1 | 1 | 1.0% | D |
| Droxidopa (#24) - Minaprine (#49) | 152.4 | 1 | 0.6% | D |
| Droxidopa (#24) - Zonisamide (#54) | 289.7 | 1 | 0.3% | D |
| Mycophenolic acid (#42) - Mercaptopurine (#58) | 11.0 | 5 | 45.4% | E |
| Ribavirin (#51) - Mercaptopurine (#58) | 23.9 | 5 | 20.9% | E |
| Udenafil (#10) - Mycophenolic acid (#42) - Mercaptopurine (#58) - | 18.0 | 7 | 38.8% | E |
| Mycophenolic acid (#42) - Dipyridamole (#57) - Mercaptopurine (#58) | 22.0 | 7 | 31.8% | E |
| Udenafil (#10) - Ribavirin (#51) - Mercaptopurine (#58) | 30.9 | 7 | 22.6% | E |
| Ribavirin (#51) - Dipyridamole (#57) - Mercaptopurine (#58) | 34.9 | 7 | 20.0% | E |
| Theophylline (#18) - Mycophenolic acid (#42) - Mercaptopurine (#58) | 41.7 | 7 | 16.7% | E |
| Mycophenolic acid (#42) - Pentoxifylline (#50) - Mercaptopurine (#58) | 53.8 | 7 | 13.0% | E |
| Theophylline (#18) - Ribavirin (#51) - Mercaptopurine (#58) | 54.6 | 6 | 10.9% | E |
| Pentoxifylline (#50) - Ribavirin (#51) - Mercaptopurine (#58) | 66.7 | 6 | 8.9% | E |
| Pentoxifylline (#50) - Arsenic trioxide (#72) | 118.2 | 17 | 14.3% | F |
| Cladribine (#16) - Pentoxifylline (#50) | 118.2 | 17 | 14.3% | F |
| Gemcitabine (#30) - Pentoxifylline (#50) | 157.8 | 15 | 9.5% | F |

in combination with one drug among number 4 or 43 or 60 (Pravastatin or Naftifine or Tioconazole; see Table 2.2 for all correspondences between names and numbers). Being the cardinality of class C equal to 6, we can deduce that these solutions are generated only by the combination of the pair and the triplet just mentioned. Three exceptions to the sharp clusterization of Figure 2.2 are represented by drugs labeled with the numbers 7, 22 and 50, respectively Rosiglitazone, Indomethacin and Pentoxifylline. Indeed Rosiglitazone targets many metabolic reactions (60, all in the fatty acid metabolism) which allow two types of interaction: class A for fatty acid activation and class C for cholesterol metabolism. On the other hand, Indomethacin causes only 7 inhibitions, some belonging to glycerolphospholipids metabolism and some others to pyruvate pathways: the first interact synergistically with drugs which target fatty acid reactions (class A), whereas the second can be combined with drugs acting on pyruvate metabolism (like Fomepizole, drug number 75, in class B). Finally, Pentoxifylline inhibits reactions both in the salvage pathway for nucleotides (which give synergisms in class E) and in pyrimidine catabolism (class F).

The analysis of the complete results obtained from the screening over all metabolic reactions is shown in Figure 2.3, where the stopped reactions are grouped on the ba-

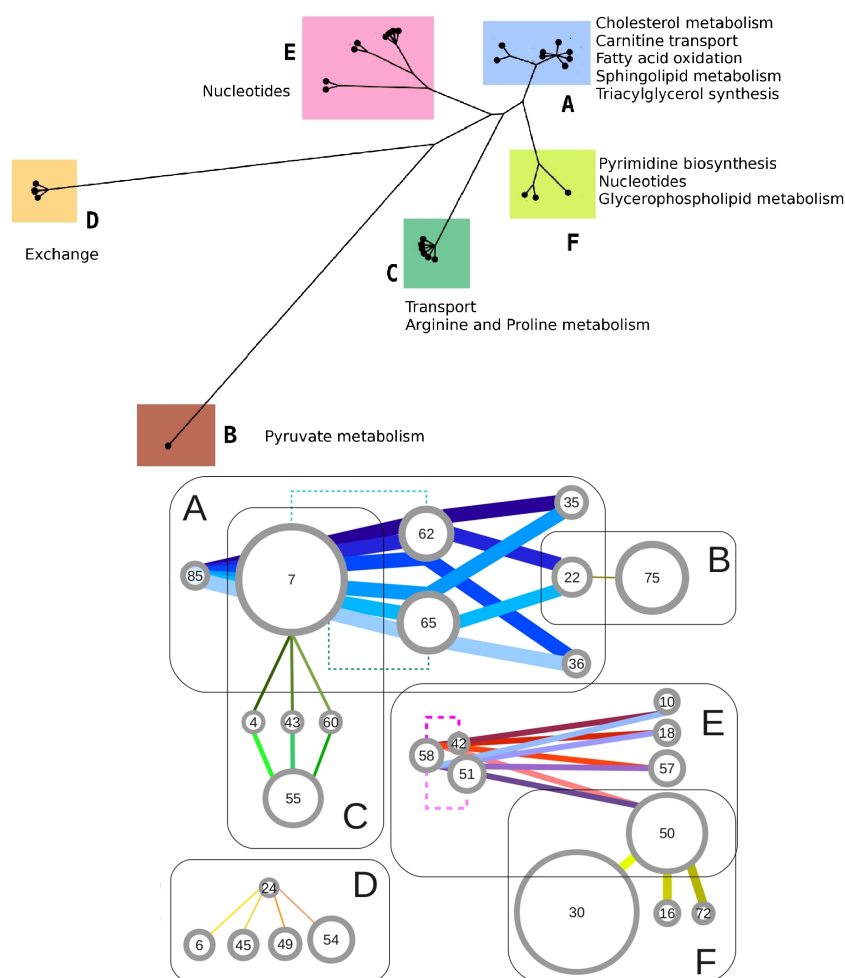


Figure 2.2: **Classification of the synergisms for the human metabolic network.** *Top panel:* Each leaf of the tree represents a multidrug solution that we have found. The layout of the graph is obtained through the same method used for phylogenetic trees (a distance tree, see text) and manifestly shows the clustering of these synergisms; the six clearly visible classes have been labeled with letters (from “A” to “F”). Names of the pathways mainly affected by each class are reported near the clusters. Details of the metabolic functions to which these classes of synergisms correspond are given in Figure 2.3. *Bottom panel:* This network of drugs represents a detailed characterization of the classes of synergisms. Each drug is indicated by a circle whose radius is proportional to the number of its direct targets (drugs are labeled with numbers according to Table 2.2). Each synergism is drawn as a colored line which connects the drugs involved (each synergism has its own color and the line thickness is proportional to the number of stopped reactions). Even in this more detailed representation, the six classes are still visible. Some subclasses can also be identified: drug pairs (7, 62) and (7, 65) in class A and drug pairs (42, 58) and (51, 58) in class E (indicated with dotted lines) exploit part of the synergism of the entire class; indeed these 4 pairs are the isolated leaves in the corresponding clusters in the top panel. Note the role of drugs 7, 22 and 50 in bridging classes A-C, A-B and E-F.

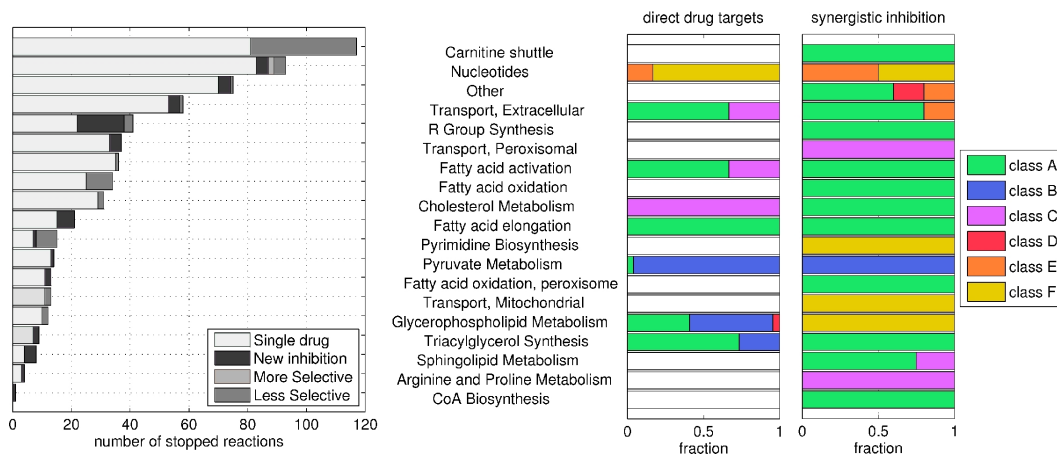


Figure 2.3: **Drug synergisms for the human metabolic network.** *Left panel:* For each affected pathway, the histogram reports the number of objective reactions which can be stopped; gray-scale bars represent reactions stopped only by a *Single drug* or multidrug solution, classified as *New inhibition* (meaning that no single drug is capable of triggering the inhibition), *More selective* and *Less selective* inhibitions (referring to the case where both single and multiple drug treatments are possible and the multiple one has respectively a lower and a higher side effect). *Right panels:* The two plots refer to multiple drug solutions only. For the same pathways, we report here the fraction of the direct drug targets and the fraction of the synergistic inhibitions which are induced by the six classes of synergisms (shown in Figure 2.2): the comparison between the two stacks shows that synergistic interactions can occur on pathways that are not direct targets of the drugs. See also Table 2.4.

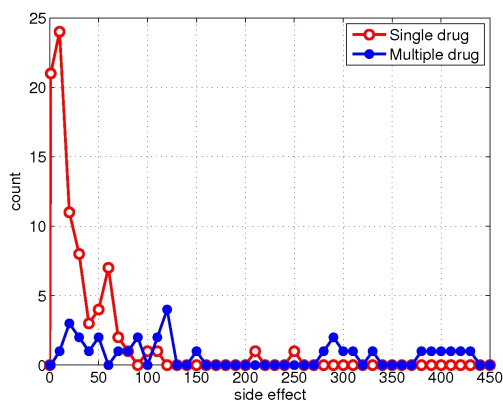


Figure 2.4: **Side effects comparison.** Histogram analysis of the size effect induced by single drug and by multiple drug solutions.

Table 2.4: **List of the inhibitions on human metabolism obtained by multiple drug solutions.** The reactions are sorted according to the pathway they belong. The solutions are classified as *New* inhibition, *More* selective and *Less* selective from a comparison with a possible single drug treatment; this information is reported in the third column. Last columns reports the class of the synergistic combination which cause the inhibition of the objective reaction (see Figure 2.2).

| OBJECTIVE REACTION | | SYNERG. | |
|--|----------------------------------|---------|--------|
| Name | Pathway | Type | Class. |
| Proline dehydrogenase | Arginine and Proline Metabolism | New | C |
| C160 transport into the mitochondria | Carnitine shuttle | Less | A |
| C161 transport into the mitochondria | Carnitine shuttle | Less | A |
| C180 transport into the mitochondria | Carnitine shuttle | Less | A |
| C181 transport into the mitochondria | Carnitine shuttle | Less | A |
| carnitine fatty-acyl transferase | Carnitine shuttle | Less | A |
| carnitine O-palmitoyltransferase | Carnitine shuttle | Less | A |
| carnitine O-stearoyl transferase | Carnitine shuttle | Less | A |
| carnitine octadecenoyl transferase | Carnitine shuttle | Less | A |
| carnitine transferase | Carnitine shuttle | Less | A |
| R group transport into the mitochondria | Carnitine shuttle | Less | A |
| transport into the mitochondria (carnitine) | Carnitine shuttle | Less | A |
| sterol O-acyltransferase 1 | Cholesterol Metabolism | Less | A |
| dephospho-CoA kinase | CoA Biosynthesis | New | A |
| 3',5'-Cyclic GMP exchange | Exchange | New | E |
| (R)-Pantothenate exchange | Exchange | New | A |
| Ceramide 1-phosphate exchange | Exchange | New | A |
| cholesterol ester exchange | Exchange | Less | A |
| L-Phenylalanine exchange | Exchange | New | D |
| Beta oxidation of fatty acid | Fatty acid oxidation | Less | A |
| Beta oxidation of long chain fatty acid | Fatty acid oxidation, peroxisome | New | A |
| Beta oxidation of long chain fatty acid | Fatty acid oxidation | Less | A |
| fatty acyl-CoA desaturase (n-C18:1CoA -> n-C18:2CoA) | Fatty acid elongation | New | A |
| stearoyl-CoA desaturase (n-C18:0CoA -> n-C18:1CoA) | Fatty acid elongation | New | A |
| choline phosphatase | Glycerophospholipid Metabolism | Less | F |
| phosphatidate cytidyltransferase | Glycerophospholipid Metabolism | Less | F |
| 5'-nucleotidase (GMP) | Nucleotides | More | E |
| CTP synthase (NH ₃) | Nucleotides | Less | F |
| cytidine kinase (ATP) | Nucleotides | New | F |
| cytidylate kinase (CMP),mitochondrial | Nucleotides | Less | F |
| GMP reductase | Nucleotides | New | E |
| guanylate cyclase | Nucleotides | New | E |
| guanylate kinase (GMP:ATP) | Nucleotides | New | E |
| nucleoside-diphosphate kinase (ATP:CDP), mitochondrial | Nucleotides | Less | F |
| ribonucleoside-diphosphate reductase (GDP) | Nucleotides | More | E |
| aspartate carbamoyltransferase (reversible) | Pyrimidine Biosynthesis | Less | F |
| carbamoyl-phosphate synthase (glutamine-hydrolysing) | Pyrimidine Biosynthesis | Less | F |
| CTP synthase (glutamine) | Pyrimidine Biosynthesis | Less | F |
| dihydroorotic acid dehydrogenase (quinone10) | Pyrimidine Biosynthesis | Less | F |
| dihydroorotase | Pyrimidine Biosynthesis | Less | F |
| orotate phosphoribosyltransferase | Pyrimidine Biosynthesis | Less | F |
| orotidine-5'-phosphate decarboxylase | Pyrimidine Biosynthesis | Less | F |
| hydroxyacylglutathione hydrolase | Pyruvate Metabolism | New | B |
| R group artificial flux | R Group Synthesis | New | A |
| R group to palmitate conversion | R Group Synthesis | Less | A |
| R total flux | R Group Synthesis | New | A |
| 3-Dehydroshinganine reductase | Sphingolipid Metabolism | New | A |
| Ceramide kinase | Sphingolipid Metabolism | New | A |
| dihydroceramide desaturase | Sphingolipid Metabolism | New | C |
| serine C-palmitoyltransferase | Sphingolipid Metabolism | New | A |
| ATP transporter, peroxisomal | Transport, Peroxisomal | New | C |
| cGMP transport (ATP-dependent) | Transport, Extracellular | New | E |
| cholesterol ester transporter | Transport, Extracellular | Less | A |
| cmp hs transport | Transport, Extracellular | New | A |
| cytidine facilitated transport in mitochondria | Transport, Mitochondrial | Less | F |
| Diphosphate transporter, peroxisome | Transport, Peroxisomal | New | C |
| fatty acid retinol efflux | Transport, Extracellular | New | A |
| intracellular transport | Transport, Mitochondrial | Less | F |
| NADP transporter, peroxisome | Transport, Peroxisomal | New | C |
| NADPH transporter, peroxisome | Transport, Peroxisomal | New | C |
| Pantothenate sodium symporter II | Transport, Extracellular | New | A |
| 1-acylglycerol-3-phosphate O-acyltransferase 1 | Triacylglycerol Synthesis | New | A |
| glycerol-3-phosphate acyltransferase | Triacylglycerol Synthesis | New | A |

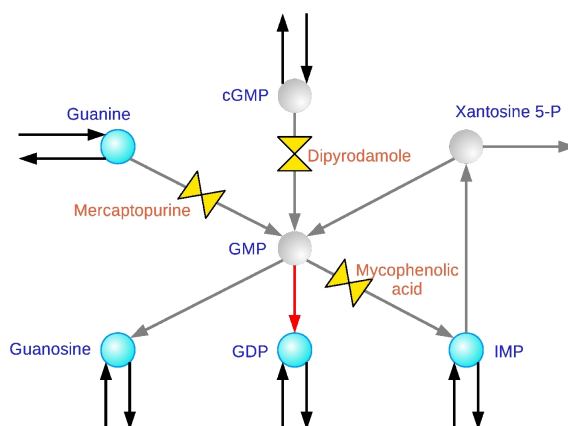


Figure 2.5: **Nonlinearity in the synergism: the example of Guanylate kinase.** The part of the human network here represented shows the nonlinear interaction when the three drug targets (the three valves, with the name of the drugs) are simultaneously inhibited: when this is the case, the objective reaction of Guanylate kinase (in red) is stopped. Gray arrows and gray circles indicate respectively stopped reactions and metabolites which become unavailable.

sis of the metabolic pathway to which they belong. This figure reports also the class and the targets of the drug combinations which induces the inhibition. As one can see, some synergisms occur between reactions which belong to different pathways: in particular, sphingolipids subsystems, CoA and pyrimidine biosynthesis contain reactions whose inhibitions are caused by interactions situated in other pathways, since none of the combined drugs have targets on them. Moreover, among the multiple drug solutions of Figure 2.3 (left histogram), there are several *new inhibitions* and a few *more selective* cases (a comparison of the side effects induced by single and multiple drug solutions is reported in Figure 2.4).

Among the cases of new inhibitions (see Table 2.4), the case of Guanylate kinase, although of no therapeutic interest, represents an easily visualizable example of the nonlinearity in the superposition of the effects as anticipated in the hypothetical situation presented in Figure 2.1. We consider the phosphorylation of GMP into GDP catalyzed by guanylate kinase as objective reaction. Since the blockage of GMP production will cause also the arrest of any transcription process, this inhibition constitutes only a toy example of synergism devoid of any practical value. For this problem, the algorithm proposes the combination of Mercaptopurine, Dipyridamole and Mycophenolic acid: the synergism takes place through the simultaneous inhibition of guanine phosphoribosyltransferase, 3',5'-cyclic-nucleotide phosphodiesterase and IMP dehydrogenase (see Figure 2.5 for a representation of the corresponding subnetwork). Indeed, these reactions are alternative ways of GMP biosynthesis. When and only when they are all blocked, Guanylate kinase lacks its substrate and stops as well.

The complete list of the objective reactions, with relative pathways and synergistic inhibitions, is reported in Table 2.4: this list contains many inhibitions in the fatty acid, cholesterol and carnitine transport pathways, which may represent solutions for obesity. In particular, concerning the case of hyperlipidemia diseases, the algorithm finds the combination of Rosiglitazone and Cerulenin (class A) as inhibitor of many reactions in the carnitine transferase and fatty acid desaturase pathways. This synergism has been reported in the literature for being active versus the biosynthesis of fatty acids in prostate tumors [48]: the mean IC_{50} are $45\mu\text{M}$ and $32\mu\text{M}$ for Rosiglitazone and Cerulenin alone, whereas it reduces to $5\mu\text{M}$ when they are combined: the authors claimed that this effect comes from the reduced production of fatty acids preventing the growth and differentiation of prostate cells. It is worth noting that we predict this combination also in three anticancer solutions (together with an additional target on palmitate conversion, see next section). Moreover, this pair is part of other eight synergisms on the human metabolism (class A), all concerning the same pathways .

Another significant example is represented by the inhibition of Dihydroceramide desaturase inside the human metabolic network. Ceramide is the hydrophobic membrane anchor of sphingolipids and is involved as a bioactive molecule in cell growth regulation, apoptosis, senescence, and diverse cell responses, particularly those linked to stress situations [49, 50]; recent studies have shown the role for ceramide biosynthesis in body weight regulation, energy expenditure, hence in the metabolic obesity syndrome [51]. For these reasons Dihydroceramide desaturase has been proposed as a promising potential target for metabolic diseases. Currently some specific inhibitors of Dihydroceramide desaturase are under investigation and development (for example GT11, XM462 and analogous [52, 53]) although there is no approved drug yet. Indeed in our model no single drug can stop this reaction. Our algorithm finds some possible multidrug treatments that block this reaction: among them, there are the synergistic pairs of Rosiglitazone plus Simvastatin (same group of Pravastatin, see Table 2.2), and Acetylsalicylic acid plus Atorvastatin (Pravastatin). Concerning the first synergism, clinical experiments have shown that combining these two drugs a significant reduction (about 30% less) of the intracellular accumulation of lipid is achieved [54]. Moreover, the authors of [55] investigate the adverse effect of single and combined therapies (hypoglycemia, body weight increase) and claim that adverse events are generally similar (the safety profile of Rosiglitazone was not adversely affected by the addition of Atorvastatin). Also our results predict a limited worsening of the adverse effect: indeed, after the combination with Atorvastatin the side effect of Rosiglitazone passes from 252.7 to 288.8 (see Table 2.2 and 2.3), i.e. it increases of about 14% only. For the same therapeutic purpose, the pair of Acetylsalicylic acid and Atorvastatin has been also studied. Clinical trials are currently ongoing [56] and some of them have shown promising results [57]. The rationale for this approach is based on the restoration of platelet sensitivity by reduction of the cholesterol levels. Moreover, as mentioned above, it is known that ceramide is involved in apoptosis. This combination has been tested for the treatment of prostate cancer: the results have shown a linear synergism between these two drugs [58].

Table 2.5: **Solutions for the cancer vs human problem.** We report the solutions and the side effects $\sigma(\mathbf{h})$ for the inhibition with approved drugs. Drug numbers refer to Table 2.2. Bold font indicates that the solution has an experimental validation.

| Solution nr. | Drugs | Side effect $\sigma(\mathbf{h})$ |
|--------------|------------------------------------|----------------------------------|
| 1 | Floxuridine (#20) | 1 |
| 2 | Mycophenolic acid (#42) | 4 |
| 3 | Trimethoprim (#29) | 5 |
| 4 | Methotrexate (#11) | 5 |
| 5 | Atovaquone (#69) | 6 |
| 6 | Tyloxapol (#85) | 6 |
| 7 | Ezetimibe (#56) | 12 |
| 8 | Pemetrexed (#41) | 15 |
| 9 | Ribavirin (#51) | 17 |
| 10 | Quinacrine (#36) | 22 |
| 11 | Myo-Inositol (#82) | 29 |
| 12 | Tioconazole (#60) | 34 |
| 13 | Naftifine (#43) | 34 |
| 14 | Simvastatin (#4) | 42 |
| 15 | Leflunomide (#66) | 42 |
| 16 | Auranofin (#59) - Fomepizole (#75) | 47 |
| 17 | Indomethacin (#22) | 48 |
| 18 | Diclofenac (#35) | 55 |
| 19 | Hydroxyurea (#16) | 56 |
| 20 | Arsenic trioxide (#72) | 56 |
| 21 | Gemcitabine (#30) | 99 |

2.4 Human cancer: results

A similar approach can be used to improve the *selectivity* and *specificity* of the treatment when dealing simultaneously with more than one type of cells. As presented in Section 2.2.2, our procedure can be modified in order to force the solution to preserve the metabolism of one cell while inhibiting an objective reaction of another. Indeed, a drug interaction can explore the differences in the topologies of the metabolic networks and, in this way, bypass the restrictions caused for instance by targets homology. This is crucial in case of anticancer therapy since tumoral and normal cells share the same genes.

For this purpose, we use the metabolic network of a generic human cancer assembled in [2] and the human metabolic network. We apply the modified version of the algorithm to the biomass reaction of the cancer network (which must be stopped), while minimizing the side effect on the regular human metabolism. As in the previous section, the procedure is iterated until the problem becomes unfeasible. The results are shown in Table 2.5 and Figure 2.6. The solutions are mainly single drugs which differ one from the other in terms of side effect on the human network. Many are known chemotherapeutic agents such as Floxuridine, Mycophenolic acid, Methotrexate, Pemetrexed, Ribavirin, Myo-Inositol, Sim-

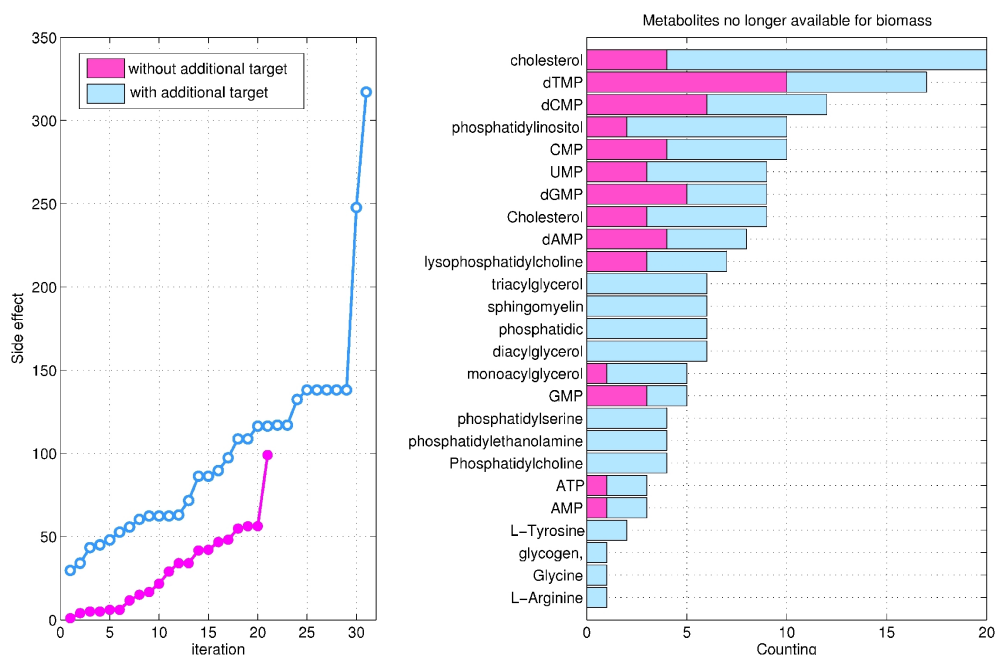


Figure 2.6: **Results on cancer *vs* human selectivity problem.** *Left panel:* The iterative application of the algorithm to the cancer *vs* human networks finds 21 solutions before becoming unfeasible (magenta line). Including the possibility of inhibiting an additional target, other 31 solutions are found (blue line). *Right panel:* The bars count the number of solutions which stop the biomass metabolite in the cancer metabolism (same color code). See Tables 2.5 and 2.6 for more details.

vastatin, Leflunomide, Indomethacin, Hydroxyurea, Arsenic trioxide, Gemcitabine [59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70]. Since only one synergism is present, between Fomepizole and Auranofin, these results suggest that approved drugs do not seem to induce significant interactions at the level of the metabolism.

In order to increase the range of putative synergisms, we have tried to use the algorithm to set up a search for targets potentially interacting with the currently available drugs. Instead of searching for the reactions which are synthetically lethal (as partially done, for instance, in [2]) and proposing them as potential new targets, we look for the reactions whose single inhibition may give a lethal synergism with any combination of the approved drugs. The problem is formally equivalent to the one we have already described: we search again for the optimal drug combination after having deleted a reaction in the cancer network (the same reaction is removed also from the human network; this reaction is called “additional target” since it will be the target of an additional new drug). We systematically consider each reaction of the cancer network as an additional target; the results of this screening are reported in Table 2.6. After a search in literature of possible inhibitors of the additional targets of the results, we identify some interesting solutions. For instance, the inhibition of

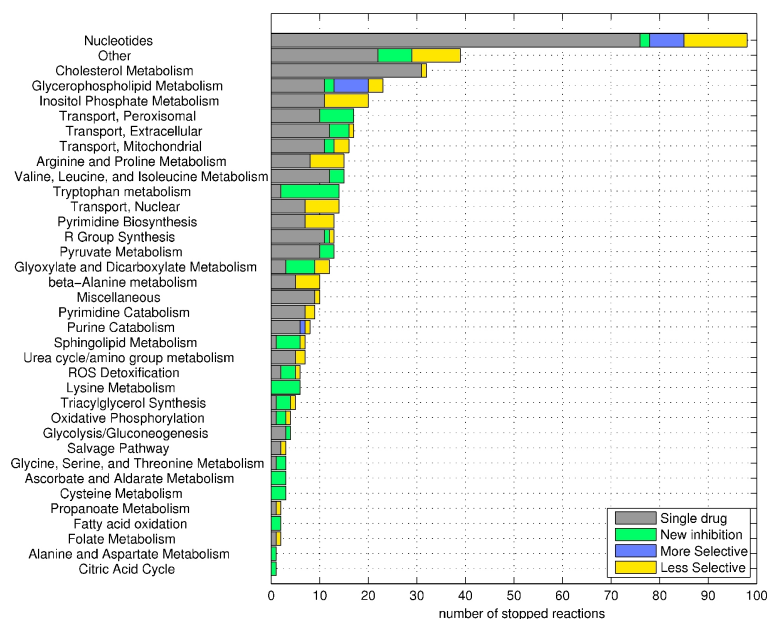


Figure 2.7: **Results on cancer metabolic network alone.** Like for the human network (see Figure 2.3), the plot reports the new inhibitions that synergisms make possible and the inhibitions with different selectivity with respect to a single drug treatment. Notice how the number of new inhibitions is significantly higher than in the human metabolism, meaning that the cancer pathways are less robust (and redundant) than their counterparts in the human network.

methenyltetrahydrofolate cyclohydrolase (combined with the use of Mimosine) can be induced by the experimental drug 5,6,7,8-tetrahydro- N^5, N^{10} -carbonylfolic acid [71]; therefore, its combination with Mimosine could represent a potential antitumor therapy. Also for other four additional targets there exist experimental inhibitors whose activity is reported in the literature. Concerning the drug solutions identified by the algorithm, the pair Rosiglitazone plus Cerulenin is proposed in combination with three possible additional targets. One is the palmitate fatty acid conversion: it is worth noting that the validated antitumoral activity of this pairs versus prostate cancer cells (as mentioned above) is due to the reduction of the synthesis of fatty acids [48]. In our prediction, indeed, among the metabolites which are no longer available because of this inhibition, there are cholesterol ester, mono-, di- and triacylglycerol. The other two additional targets are related to phosphatidylcholine. Part of the phosphatidylcholine pathway has been already identified as synthetic lethal [2], but without mentioning any possible exploitation. Our results suggest the use of a combined drug therapy as possible way to take advantage of this synthetic lethality.

2.5 Discussion

The field of drug combinatorics is largely unexplored experimentally and the potential of combined drug therapies is difficult to assess, mostly for lack of suitable systematic methodologies. To try to fill this gap, we have developed an algorithm which is capable of exploring efficiently the optimal synergisms among all possible drug combinations and of characterizing them in terms of side effect and selectivity. Indeed, the success of a drug discovery process depends on multiple aspects, not least on the fulfillment of requirements regarding selectivity and toxicity: for metabolic diseases the modulation of the key pathways without affecting the other vital functions can be instrumental for rescuing from the pathology. Similarly, anti-cancer compounds should only kill cancer cells without affecting normal ones. These requirements are rarely taken into account in standard computational approaches. The results we obtained by applying our algorithm to the human and tumor *vs* human metabolic networks show the possibility to take advantage of drug synergisms in proposing new therapies: the potentialities lay in the possibility to intervene with a different mechanism of action with respect to those that are currently available. In this enlarged repertoire of possibilities, we have identified examples of drug repurposing (some of them were previously demonstrated experimentally), a procedure which is becoming more and more attractive thanks to the reduced costs on the preclinical and clinical steps.

One of the main features of FBA-based knockout studies is that metabolic networks appear to be robust [75, 30], meaning that there seem to be a high degree of redundancy of the pathways inside a network (property alternatively reported as “nonessentiality” of the gene in Refs. [36, 45]). In the context of drug synergism, this property reflects into the presence of optimal solutions consisting of many drugs (in our case up to four, or even more if we consider the number of inhibition targets of each drug). For the same reason, the results show the necessity to extend the search to all possible drug combinations without limiting to those of low cardinality. This fact becomes significant especially when the drugs to combine present a high similarity in terms of inhibited targets; indeed, the characterization of the synergisms we have found shows a limited variety of possible interactions between the available drugs (only six classes were identified, see Figure 2.2). However, besides increasing the cardinality of the solutions, a very strong robustness may also reduce the total number of solutions because it makes more difficult to induce the simultaneous inhibition of all the redundant pathways; indeed, if the screening we have run on the human metabolism is applied also to the less robust cancer network (seen as stand-alone network), the number of possible inhibitions is much higher (see Figure 2.7 in comparison with Figure 2.3). Moreover, when we want to stop the biomass reaction of the cancer, mainly single drug solutions are found (see Table 2.5): this is again an index of the low redundancy of the cancer network and of the limited variety in the metabolic targets for the available drugs.

Nevertheless, adding the possibility to inhibit an extra target, we could identify some experimental compounds (other than the drugs from DrugBank) which may be

used as anticancer in a combination with the approved drugs (see Table 2.6). These examples show that predictive tools like the method we are proposing become more important if one considers also the possibility of combining active compounds which are not yet approved but for which a minimal characterization of the mechanism of action is available. In this perspective, experimental compounds which inhibit additional targets that are different from those affected by the approved drugs may represent a good chance for improving, through synergism, the spectrum of the whole set of the currently available drugs. Moreover, the application of our method can be extended to situations where multiple networks are compared and contrasted. It is expected that problems like this will become important as soon as tissue-specific networks of human metabolism and cancer-specific networks will become available in the near future.

In a broader perspective, if instead of confining our study only to the human network we consider also the metabolism of microorganisms, the exploitation of drug synergisms obtained with our algorithm can be useful in investigating a wide range of situations: (i) when specific enzyme inhibitors are not currently available, multiple drug solutions could represent an example of the *reprofiling* of existing drugs for new therapeutic indications; (ii) when the target enzyme has undergone a mutation rendering ineffective the original therapy, a synergistic solution may bypass the resistance acting on other enzymes and therefore help in *fighting resistance*; (iii) when the optimal synergism has no lethal impact while the single drug solution has. This *change in lethality* can be important for instance in cases of human-hosted bacteria producing toxic by-products: in order to save the useful symbiosis with these commensal bacteria, a selective (but not lethal) inhibition of the toxic processes must be pursued [76, 77].

Chapter 3

A more realistic description of drug action: partial inhibition

3.1 Motivations

In the drug synergisms problem presented in the previous chapter, we derived a bilevel optimization which was reformulated as a single optimization problem through the strong duality theorem of LP. In order to solve the nonlinearity which arose from this theorem (see Equation (2.3)), we described the inhibition induced by a drug as the knockout of the genes which code for the enzymes of the drug targets (i.e. by a Boolean variable). This approach was inspired by the literature on strain improvement [24, 44], which aims at the identification of the best knockout maximizing the biosynthesis of a key metabolite. However, if this ON/OFF description is correct for gene deletions, in the case of drug treatment (where the enzymes are inhibited by drugs) or gene mutation (on which one changes the enzymes activity) it represents only a rough approximation which may not constitute a realistic description of the biological effect. In addition, a complete inhibition of a disease-causing target may not represent the right therapeutic solution (in healthy cells, the level of each metabolite must be in a finite range rather than suppressed). It is in fact more plausible to assume that a drug acting on an enzyme leads to a partial loss of functionality of the latter, and hence to a partial inhibition of the corresponding reaction(s). Although expected to be a potential strategy in a multi-target approach, [32], partial inhibition has been considered only in a few computational works in the current literature. For example, studies like [78, 79] dealt with a small part of the network, modeling the kinetic reactions explicitly and solving them numerically. The partial inhibition then amounts, for instance, to a modulation of one or more kinetic parameters. Because of the complexity of the metabolic networks and because of the difficulty of knowing the kinetic parameters of all biochemical reactions, this approach cannot be applied at a genome-wide level. On the other hand, the authors

of [80] consider the whole network, but the (partial) inhibition is given as initial fixed parameter of the model and only the effect of the perturbation is quantified. A different approach is presented in [81] in the context of the prediction of new drug targets; these targets are identified through a two-stages FBA (which differs from a bilevel formulation because the two optimizations are not nested). However, the potential targets obtained with this method must be verified exhaustively, which may represent a problem for networks with more than the 26 reactions of the human hyperuricemia metabolic pathway considered in [81]. An alternative to overcome the ON/OFF oversimplification and to consider the modulation of the enzyme activity is to reformulate the bilevel optimization problem as a nonlinear (nonconvex) single optimization problem, but this leads to a more complicated situation from a numerical point of view [23].

Therefore, the aim of the present chapter is to introduce a novel algorithm which allows to provide a more realistic description of the partial inhibition induced by the drugs on large networks while still remaining within the framework of LP. Improving on the work from the previous chapter, we consider the problem of searching the optimal combination of drugs capable, through a synergistic effect, to *modulate* (inhibit or enhance) an objective reaction (i.e. a putative target for a disease) while inducing the minimal perturbation on the rest of the network. Replacing a single Boolean variable by a convex combination of a fixed number of Boolean variables, we are able to model the inhibition as any number belonging to a discretized representation of the interval $[0, 1]$. This approach preserves the linear nature of the final problem. Notice that the method we propose can be extended to any bilevel optimization which needs to deviate from the simple ON/OFF description.

In the following, we first formalize a more generalized version of the problem about drug synergisms; then, within this case study, we describe why Boolean variables are necessary in the reformulation of a bilevel optimization problem via the strong duality theorem of LP. After the presentation of the novel idea concerning the implementation of partial inhibition, the potentiality and limits of the proposed algorithm are discussed.

3.2 Materials and Methods

3.2.1 The drug synergism problem

As mentioned, in FBA the vector \mathbf{v} of the “effective” metabolic fluxes is obtained through the optimization of a certain cost functional. For unperturbed networks, the production of the macromolecular building blocks for the biomass (the growth rate, here indicated by the flux v_{biomass}) is often maximized [20]: we denote by \mathbf{v}^{ut} (ut=“untreated”; all symbols and variables are listed in Table 3.1) the reaction fluxes obtained after this optimization:

$$\mathbf{v}^{\text{ut}} = \arg \max_{\mathbf{v} \in W} v_{\text{biomass}}, \quad (3.1)$$

where $W = \{\mathbf{v} : \mathbf{S}\mathbf{v} = \mathbf{0} \text{ and } 0 \leq v_i \leq U_i, i = 1, \dots, r\}$ (as done in the previous chapter, all fluxes are rendered positive by decomposing any reversible reaction in a couple of irreversible reactions). These fluxes \mathbf{v}^{ut} can be nonunique [82] (an analysis of the case of multiple equivalent solutions is reported in Section 3.3.3). In the following, nevertheless, these unperturbed fluxes are considered as given parameters of the problem.

The drug synergisms problem we consider here is a generalization of the problem treated in the previous chapter. In particular, for a given metabolic network and for a given set of d drugs which inhibit some enzymes of this network (k -th drug has a set of target denoted by \mathcal{T}_k), we want to modulate a certain reaction (for example rendering its flux less than a given threshold) through a combination of these drugs, inducing the minimal effect on the rest of the network. In order to give a more clear presentation of the algorithm, we assume that a drug induces an identical fractional inhibition on all its targets. Therefore, the amount of inhibition by the k -th drug on its target reaction $i \in \mathcal{T}_k$, can be modeled by the linear constraint

$$v_i \leq U_i(1 - h_k), \quad (3.2)$$

where, U_i is the upper-bound of the flux v_i and where partial inhibition means $h_k \in [0, 1]$ (instead of $h_k \in \{0, 1\}$). Through this formalism we do not consider the allosteric interaction between two (or more) drugs on the same enzyme: indeed, our model simply takes the maximum inhibition over the set of drugs which affect the enzyme of reaction i :

$$v_i \leq U_i(1 - \max_{k:i \in \mathcal{T}_k} h_k).$$

Then, the vector $\mathbf{h} \in [0, 1]^d$ represents the drug treatment, i.e. the inhibition due to the drugs: for example, for $d = 3$, the vector $\mathbf{h} = [0.5, 0, 0.8]$ indicates that drug 2 is not used ($h_2 = 0$) while drugs 1 and 3 are used at dosages which cause respectively a 50% and 80% inhibition of their targets (hence, in (3.2), the reduction of their upper-bounds to 50% and 20% in the original values). For each choice of \mathbf{h} these inhibitions reduce the set W to a subset $W(\mathbf{h})$:

$$W(\mathbf{h}) = \{\mathbf{v} \in W \text{ such that } v_i \leq U_i(1 - h_k), \forall k = 1, \dots, d, \forall i \in \mathcal{T}_k\}.$$

The determination of the reaction fluxes $\mathbf{v}^{\text{tr}}(\mathbf{h})$ (tr=“treated”) for the drug-treated network is obtained through MOMA which has been shown to generate reasonable and realistic results for perturbed metabolism [21, 83, 84, 4, 85]. In order to apply the theory of linear programming, we use the definition of MOMA in terms of norm L^1 [86]. Then

$$\mathbf{v}^{\text{tr}}(\mathbf{h}) = \arg \min_{\mathbf{v} \in W(\mathbf{h})} \|\mathbf{v} - \mathbf{v}^{\text{ut}}\|_1. \quad (3.3)$$

In the following the side effect of a drug treatment is quantified in terms of the distance $\|\mathbf{v}^{\text{tr}}(\mathbf{h}) - \mathbf{v}^{\text{ut}}\|_1 = \sum_{i=1}^r |v_i^{\text{tr}}(\mathbf{h}) - v_i^{\text{ut}}|$ used in (3.3): the greater the distance, the bigger the impact of the drugs on the whole network.

Table 3.1: Symbol, value range, meaning and type of all quantities used in the algorithm description.

| NETWORK AND DRUGS | | | |
|---|---------------------------|---|--------------|
| symbol | range | description | type |
| r | \mathbb{N} | number of reactions | parameter |
| m | \mathbb{N} | number of metabolites | parameter |
| d | \mathbb{N} | number of drugs | parameter |
| S | $\mathbb{R}^{m \times r}$ | stoichiometric matrix of the network | parameter |
| $v_r^{\text{ut}}, \dots, v_r^{\text{ut}}$ | \mathbb{R}_+ | reaction fluxes of the untreated network | variable |
| $v_r^{\text{tr}}(\mathbf{h}), \dots, v_r^{\text{tr}}(\mathbf{h})$ | \mathbb{R}_+ | reaction fluxes after drug treatment \mathbf{h} | parameter |
| mod | $1, \dots, r$ | is the index of the reaction which must be modulated | parameter |
| W | - | set of feasible solutions for unperturbed network | parameter |
| $W(\mathbf{h})$ | - | set of feasible solutions for network perturbed by drug inhibition \mathbf{h} | parameter(*) |
| \mathcal{I}_k | - | set of targets of drug k ($k = 1, \dots, d$) | parameter |
| OUTER PROBLEM | | | |
| symbol | range | description | type |
| τ | $(0, 1)$ | threshold for the reaction that must be inhibited ($v_r^{\text{tr}} \leq \tau v_r^{\text{ut}}$) | parameter |
| P | > 1 | threshold for the reaction that must be activated ($v_r^{\text{tr}} \geq \tau v_r^{\text{ut}}$) | parameter |
| h_1, \dots, h_d | \mathbb{N}_0 | precision parameter | variable |
| $x_{1,0}, x_{1,1}, \dots, x_{d,P}$ | $\{0, 1\}$ | drug-inhibitions | variable |
| b | 10^{-3} | drug-dosage variables | variable |
| | | correction against an “overselection” of drugs | parameter |
| INNER PROBLEM | | | |
| symbol | range | description | type |
| v_1, \dots, v_r | \mathbb{R}_+ | reaction fluxes (primal problem and outer problem) | variable |
| U_1, \dots, U_r | \mathbb{R}_+ | upper-bounds for the reaction fluxes (of the primal problem) | parameter |
| a_1, \dots, a_r | \mathbb{R}_+ | absolute fluxes differences $a_i = v_i - v_i^{\text{ut}} $ (primal problem) | variable |
| μ_1, \dots, μ_m | \mathbb{R} | dual variables associated to the FBA steady-state constraint | variable |
| $\lambda_1, \dots, \lambda_r$ | \mathbb{R}_+ | dual variables associated to the unperturbed upper-bounds | variable |
| $\delta_1, \dots, \delta_t$ | \mathbb{R}_+ | dual variables associated to the drug inhibition; $t := \sum_{k=1}^d \mathcal{I}_k $ | variable |
| $\delta_1^{\text{max}}, \dots, \delta_t^{\text{max}}$ | \mathbb{R}_+ | upper-bounds of the dual variables $\delta_1, \dots, \delta_t$ | variable |
| $\alpha_1, \dots, \alpha_r$ | \mathbb{R}_+ | dual variables associated to the first absolute value inequalities | variable |
| β_1, \dots, β_r | \mathbb{R}_+ | dual variables associated to the second absolute value inequalities | variable |

(*): The set $W(\mathbf{h})$ is considered as parameter in the sense that its dependence on the variables \mathbf{h} is given.

The problem can be stated as follows:

Problem: *Given:*

- a metabolic network, which means a stoichiometric matrix $\mathbf{S} \in \mathbb{R}^{m \times r}$ and the upper-bounds $\mathbf{U} \in \mathbb{R}^r$, of the reaction fluxes \mathbf{v} ;
- the unperturbed fluxes \mathbf{v}^{ut} ;
- the set of d drugs together with their inhibition targets $\{\mathcal{T}_k\}_{k=1,\dots,d}$,
- the index (denoted by “mod”) of the objective reaction whose flux (v_{mod}) must be modulated;
- a threshold $\tau \in [0, 1]$ for the modulation constraint on v_{mod} ;

we want to find the inhibition $\mathbf{h} \in [0, 1]^d$ such that $v_{\text{mod}}^{\text{tr}}(\mathbf{h}) \leq \tau v_{\text{mod}}^{\text{ut}}$ and such that it causes the minimal side effect, i.e. the minimal distance $\|\mathbf{v}^{\text{tr}}(\mathbf{h}) - \mathbf{v}^{\text{ut}}\|_1$.

According to (3.3), for a given set of drugs (i.e. for a given inhibition vector \mathbf{h}), we can calculate both $\mathbf{v}^{\text{tr}}(\mathbf{h})$ (and then check whether $v_{\text{mod}}^{\text{tr}}(\mathbf{h}) \leq \tau v_{\text{mod}}^{\text{ut}}$) and the value of the side effect. Similarly to (2.2), the formulation of the problem is the following:

$$\mathbf{h} : \left[\begin{array}{l} \min \\ \mathbf{v}^{\text{tr}}(\mathbf{h}) = \arg \min_{\mathbf{v} \in W(\mathbf{h})} \|\mathbf{v} - \mathbf{v}^{\text{ut}}\|_1 \\ v_{\text{mod}}^{\text{tr}}(\mathbf{h}) \leq \tau v_{\text{mod}}^{\text{ut}} \end{array} \right] \|\mathbf{v}^{\text{tr}}(\mathbf{h}) - \mathbf{v}^{\text{ut}}\|_1. \quad (3.4)$$

The bilevel optimization (3.4) is a min-min linear program. The inner problem adjusts the fluxes so as to achieve the minimal metabolic adjustment (MOMA), subject to the drug inhibitions imposed by the outer problem and to the stoichiometric constraints. The outer problem selects the combination of drugs which has the minimum side effect and guarantees a modulated flux lower than the desired threshold.

Since we are looking for a minimum, the absolute value operation $a_i = |v_i - v_i^{\text{ut}}|$, necessary for the definition of the L^1 -norm, is obtained by adding the following linear constraints:

$$\begin{aligned} a_i &\geq +(v_i - v_i^{\text{ut}}); \\ a_i &\geq -(v_i - v_i^{\text{ut}}). \end{aligned}$$

The sum of a_i defines both the objective function of the inner and the outer problem. However in (3.4), at the optimal point of the inner problem (at the minimum of $\|\mathbf{v} - \mathbf{v}^{\text{ut}}\|_1$) we have that $\mathbf{v} = \mathbf{v}^{\text{tr}}(\mathbf{h})$, hence $\|\mathbf{v} - \mathbf{v}^{\text{ut}}\|_1$ is equal to the objective function of the outer problem. Notice that, despite of the common objective function, the two minimizations cannot be merged in a single optimization because of the additional constraint on v_{mod} contained in the outer problem. Indeed, calling B the set defined by the inequality $v_{\text{mod}}^{\text{tr}}(\mathbf{h}) \leq \tau v_{\text{mod}}^{\text{ut}}$, the following relationship holds:

$$\left\{ \arg \min_{\mathbf{v} \in W(\mathbf{h})} f(\mathbf{v}) \right\} \cap B \neq \left\{ \arg \min_{\mathbf{v} \in W(\mathbf{h}) \cap B} f(\mathbf{v}) \right\}.$$

Then, detailed equations of the bilevel optimization (3.4) are the following:

$$\begin{array}{ll}
\text{Minimize} & \sum_{i=1}^r a_i \quad \text{“outer problem”} \\
\text{such that} & \\
\left[\begin{array}{ll}
\text{Minimize} & \sum_{i=1}^r a_i \quad \text{“inner problem”} \\
\text{such that} & \\
& \sum_{j=1}^r S_{i,j} v_j = 0 \\
& v_i \leq U_i \\
& v_i \leq U_i(1 - h_k) \\
& +v_i - a_i \leq +v_i^{\text{ut}} \\
& -v_i - a_i \leq -v_i^{\text{ut}}
\end{array} \right] \\
v_{\text{mod}} & \leq \tau v_{\text{mod}}^{\text{ut}}.
\end{array}$$

3.2.2 The strong duality theorem and the need of Boolean variables

The obtained bilevel optimization is again solved by applying the strong duality theorem of LP (1.9). This leads to the following single minimization, in which Greek letters refer to dual variables (for clarity, we differentiate them according to the associated constraints of the primal problem, as detailed in Table 3.1):

$$\text{Minimize} \quad \sum_{i=1}^r a_i \quad \text{such that} \quad (3.5a)$$

$$\sum_{j=1}^r S_{i,j} v_j = 0 \quad \forall i = 1, \dots, m; \quad (3.5b)$$

$$v_i \leq U_i \quad \forall i = 1, \dots, r; \quad (3.5c)$$

$$v_i \leq U_i(1 - h_k) \quad \forall k = 1, \dots, d, i \in \mathcal{T}_k; \quad (3.5d)$$

$$v_i - a_i \leq +v_i^{\text{ut}} \quad \forall i = 1, \dots, r; \quad (3.5e)$$

$$-v_i - a_i \leq -v_i^{\text{ut}} \quad \forall i = 1, \dots, r; \quad (3.5f)$$

$$\sum_{i=1}^m S_{i,j} \mu_i + \lambda_j + \sum_{i:j \in \mathcal{T}_i} \delta_i + \alpha_j - \beta_j \geq 0 \quad \forall j = 1, \dots, r; \quad (3.5g)$$

$$\alpha_j + \beta_j \leq 1 \quad \forall j = 1, \dots, r; \quad (3.5h)$$

$$v_{\text{mod}} \leq \tau v_{\text{mod}}^{\text{ut}}; \quad (3.5i)$$

$$-\sum_{i=1}^r a_i = \sum_{i=1}^r \left[\lambda_i + U_i \delta_i \sum_{k:i \in \mathcal{T}_k} (1 - h_k) + (\alpha_i - \beta_i) v_i^{\text{ut}} \right], \quad (3.5j)$$

where (3.5a) specifies the objective function of the outer problem; equations (3.5b)–(3.5f) refer to the constraints of the original inner problem; (3.5g) and (3.5h) are the dual constraints; (3.5i) imposes the outer problem constraint on v_{mod} and (3.5j) is the duality theorem equality (1.9). However, this last equation is no longer a

linear constraint since it contains the product between the outer problem variable h_k and the dual variable δ_i ; hence, the problem can no longer be solved by a linear optimization. It is common to overcome this complication by restricting the h_k variables to Boolean values. In this case, in fact, the nonlinear terms $\delta_i h_k$ can be exactly linearized as follows:

$$\begin{aligned} z_{i,k} &:= \delta_i h_k; \\ 0 &\leq z_{i,k} \leq \delta_i^{max} h_k; \\ \delta_i - \delta_i^{max}(1 - h_k) &\leq z_{i,k} \leq \delta_i, \end{aligned} \tag{3.6}$$

where δ_i^{max} is the upper bound for the dual variable δ_i .

The restriction to Boolean variables saves the linear nature of the problem (which however requires now MILP) but it implies the assumption that drugs can only act as switches on the reactions, or equivalently, that we are considering only an ON/OFF model. This was indeed the approach we used in the study of the previous chapter.

3.2.3 The implementation of partial inhibition

In this section we propose a solution which can still use the duality theorem for solving the bilevel optimization while including the possibility of inducing a partial inhibition of the reactions targeted by the drugs. This requires to create a discretization of the interval $[0, 1]$ and to replace the ON/OFF action of each drug with $P + 1$ Boolean variables describing this discretization (P is a fixed parameter of the problem). For the k -th drug ($k = 1, \dots, d$) we introduce the set of Boolean variables $\{x_{k,n}\}_{n=0,\dots,P}$ and define the inhibition coefficient h_k by the following convex combination:

$$h_k := \frac{x_{k,0}}{2^P} + \sum_{n=1}^P \frac{x_{k,n}}{2^n}. \tag{3.7}$$

In (3.7) the integer P is related to the desired accuracy of the $[0, 1]$ discretization. Indeed, the factor h_k assumes values between 0 and 1 with precision 2^{-P} . Notice that for $P = 0$ we have the ON/OFF model of the previous section. We can replace now (3.2) with the following inequality:

$$v_i \leq U_i(1 - h_k) = U_i \left(1 - \frac{x_{k,0}}{2^P} - \sum_{n=1}^P \frac{x_{k,n}}{2^n} \right).$$

When the strong duality theorem is applied, the nonlinear terms are the $\delta_i h_k$ products. Expanding the product according to the definition in Eq. (3.7):

$$\delta_i h_k = \frac{\delta_i x_{k,0}}{2^P} + \sum_{n=1}^P \frac{\delta_i x_{k,n}}{2^n},$$

the nonlinearity is now spread over the products $\delta_i x_{k,n}$ with again $x_{k,n}$ a Boolean variables. Similarly to (3.6), we can write an equivalent set of linear inequalities:

$$\begin{aligned} z_{i,k,n} &:= \delta_i x_{k,n}; \\ 0 &\leq z_{i,k,n} \leq \delta_i^{max} x_{k,n}; \\ \delta_i - \delta_i^{max}(1 - x_{k,n}) &\leq z_{i,k,n} \leq \delta_i. \end{aligned}$$

Notice that any “representation” of the partial inhibition values can be used in place of (3.7). Let us imagine, for instance, that we would like h_k to have the same values obtained in the dose-response experiments for the determination of the half maximal inhibitory concentration (IC_{50}) of drug k (see Figure 3.1):

$$h_k \in \{0, \bar{h}_{k,0}, \bar{h}_{k,1}, \dots, \bar{h}_{k,P}, 1\};$$

(with $0 < \bar{h}_{k,i} < \bar{h}_{k,j} < 1, \forall i < j$) then we may define h_k by the following convex combination

$$h_k := \bar{h}_{k,0}x_{k,0} + (\bar{h}_{k,1} - \bar{h}_{k,0})x_{k,1} + \dots + (1 - \bar{h}_{k,P})x_{k,P}, \quad (3.8)$$

with a series of inequalities

$$x_{k,0} \leq x_{k,1} \leq x_{k,2} \dots \leq x_{k,P}.$$

Of course, the discretization (3.7) is the most efficient because, for a given number of Boolean variables ($P + 1$), it generates the maximal precision (2^{-P}). For this reason, in the following we will refer to (3.7) only.

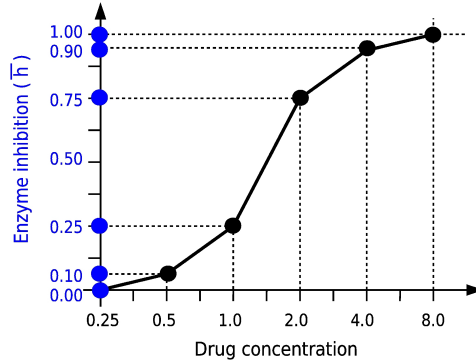


Figure 3.1: **Constructing the inhibition h from the experimental dose-response curve: an example.** The points of the curve are hypothetical experimental measurements of the effect of the drug k on the activity of the enzyme. The discretization of the curve can be used as basis for the discretization of the interval $[0, 1]$: therefore, referring to (3.8), we may define $h_k = 0.10x_{k,0} + 0.15x_{k,1} + 0.50x_{k,2} + 0.15x_{k,3} + 0.10x_{k,4}$.

3.2.4 Inhibitions or activation of the objective reaction

The evaluation of the effect on the fluxes induced by the drugs is performed through the MOMA formalism. It is known that this approach describes well the spreading across the network of the effect of the perturbation: many processes are down-regulated or up-regulated in order to adjust and compensate the effect of the perturbation (see for example [4]). For the same reason, a second intervention may amplify the deactivation (recovery) of a certain metabolic function that was down-regulated (activated) after the first perturbation [83]. In terms of multiple drug effect, this means that a drug synergism may induce both the inhibition and the activation of the reaction fluxes.

In our algorithm, through the choice of τ , one can select the exploitation of one of these two situations through a different constraint on v_{mod} in the outer problem: indeed, imposing as in the previous section

$$v_{\text{mod}} \leq \tau v_{\text{mod}}^{\text{ut}},$$

(for $0 \leq \tau < 1$) the algorithm identifies synergistic inhibitions, whereas requiring

$$v_{\text{mod}} \geq \tau v_{\text{mod}}^{\text{ut}},$$

(for $\tau > 1$) the algorithm generates drug interactions that up-regulates the objective reaction.

In the following, both versions are applied.

3.2.5 Cases of multiple equivalent solutions

In the development of our algorithm, we consider the possibility to have cases of multiple equivalent solutions (solutions with the same side effect on the network, i.e. the same distance $\|\mathbf{v}^{\text{ut}} - \mathbf{v}^{\text{tr}}\|_1$). It is worth noting that the use of the norm L^1 does not guarantee the uniqueness of the points at this minimal distance: indeed balls in L^1 and the polytope $W(\mathbf{h})$ are convex but not strictly convex sets. Unfortunately this limit can be overcome only passing to the L^2 formulation with the consequent loss of the linearity of the problem. However, we expect that such a type of situations are quite rare since they appear only when the hyperplane (or the intersection of some of them) which defines $W(\mathbf{h})$ and which realizes the minimum distance with respect to the vector \mathbf{v}^{ut} , is parallel to an edge (or face) of the L^1 -ball (see Figure 3.2).

Other (more common) cases of degeneracy of the solutions are avoided through a specific correction mechanism. For instance, if there exists a pair of drugs k and l such that $\mathcal{T}_k \subset \mathcal{T}_l$ (i.e. drug k inhibits enzymes which are already target by drug l), any solution which contains both drugs is equivalent to the solution without drug k (i.e. drug k is superfluous). Similar reasoning can be done between a lower and higher dosages of the same drug. Therefore, in order to prevent an ‘‘overselection’’ of drugs, we introduce an additional term in the objective function of the outer

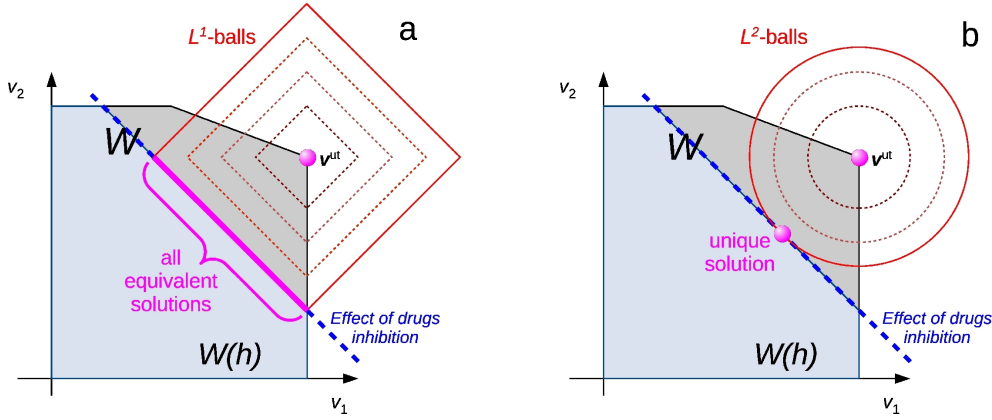


Figure 3.2: **Convexity of L^1 - and L^2 -formulations and uniqueness of the solution.** The two pictures report the unperturbed fluxes (\mathbf{v}^{ut}) in the original set W . When the inhibition \mathbf{h} is applied (dotted blue line), the set reduces to $W(\mathbf{h})$. (a) With the L^1 -norm, balls are not strictly convex, therefore cases of multiple equivalent solutions may appear. As one can see, these situations occur only when the hyperplane of $W(\mathbf{h})$ which realizes the minimum distance with respect to \mathbf{v}^{ut} is parallel to an edge of the L^1 -ball. (b) Conversely, L^2 balls are strictly convex and always lead to a unique solution.

problem. The new function becomes the following:

$$\sum_{i=1}^r a_i + b \sum_{k=1}^d h_k,$$

where the parameter b is chosen low enough (10^{-3} in our computations, much smaller than 10^3 , the common upper-bounds of the fluxes) in order to keep this term lower than the difference in the side effects and therefore to not change the order between non-equivalent solutions.

A similar approach is used to solve the redundancy in the definition of h_k . Indeed, since in (3.7) $x_{k,0}$ and $x_{k,1}$ have the same coefficient, the inhibition h_k does not change when swapping the values of these two variables. To avoid this degeneracy, we allow $x_{k,0}$ be 1 only if all the other $x_{k,j}$ (for $j = 1, \dots, P$) are 1. This is obtained through the following extra linear constraint:

$$x_{k,0} \leq \frac{1}{P} \sum_{j=1}^P x_{k,j}.$$

Since the problem is not strictly convex and since equivalent drug combinations may always appear (for instance, the combination of drug A which targets reaction 1 plus drug B which targets reactions 2 and 3 is equivalent to the combination of drug C which targets reaction 2 plus drug D which targets reactions 1 and 3), the problem of multiple equivalent solutions needs to be considered. In all these cases,

because of the numerical implementation of the algorithm, a random choice of one of these optimal solutions is taken. Because, by definition, all these equivalent solutions fulfill the conditions on the objective reaction and on the minimization of the side effect, their differences are irrelevant: indeed they concern only some other fluxes on which we do not have any specific requirements. For this reason any solution chosen by the implementation of the algorithm can be considered acceptable.

3.2.6 Metabolic networks

Six metabolic networks of different microorganisms are used in this study. They have been downloaded from BIGG database and their mean features are listed in Table 3.2. For the central carbon metabolism of *E.coli* the table reports also the number of drugs that have been selected from DrugBank database.

Table 3.2: **Main properties of the metabolic networks used in this work (from BIGG bigg.ucsd.edu/).** For the *E.coli* core network, drugs and respective targets have been selected from DrugBank database [47]: drugs which are common metabolites (for example Adenosine) have been discarded and drugs with the same targets have been grouped together.

| | <i>E.coli</i> core | <i>S.aureus</i> | <i>H.pylori</i> | <i>S.oneidensis</i> | <i>E.coli</i> | <i>S.typhimurium</i> |
|------------------------|--------------------|-----------------|-----------------|---------------------|---------------|----------------------|
| Nr. of reactions r | 95 | 575 | 513 | 696 | 1911 | 2224 |
| Nr. of metabolites m | 72 | 455 | 436 | 528 | 1337 | 1497 |
| Nr. of genes | 137 | 619 | 339 | 783 | 1261 | 1271 |
| Nr. of drugs d | 8 | - | - | - | - | - |
| Reference | [87] | [9] | [10] | [11] | [14] | [15] |

3.3 Results

3.3.1 Screening for optimal drug combinations on *E.coli* core network

The main scope of these calculations is to show the advantage given by the use of values of P higher than zero, i.e. of passing from the ON/OFF to a more accurate description. In order to better characterize its behavior (performing a large number of tests), we run the algorithm on the small network of the core metabolism of *E.coli*. A sketch of this network is depicted in Figure 3.3.

A set of tests have been carried out combining different values of τ and P , in particular:

$$\begin{aligned}\tau &\in \{0.0, 0.1, 0.5, 1.5, 2.0\}; \\ P &\in \{0, 1, 2\}.\end{aligned}$$

For each pair, we perform a screening that considers each metabolic reaction as objective process to be modulated (down- or up-regulated depending on the value of τ) and finds the most selective drug combination. The following characterization of

the solutions is performed. For a given P and for a given objective reaction v_{mod} , we consider the solutions \mathbf{h} at different values of τ . When the same drug combination is found for two values of τ (for example $\tau_1 = 0.1$ and $\tau_2 = 0.5$), the solution is considered valid only for the most stringent constraint ($\tau_1 = 0.1$, in the example; similarly, if $\tau_1 = 1.5$ and $\tau_2 = 2.0$ then the solution is associated to $\tau_2 = 2.0$ only). This procedure allows to consider only cases when passing to a weaker constraint on v_{mod} the severity (for instance the dosage) of the corresponding optimal drug treatment is reduced too. We analyze the results by looking at the following four indices.

Number of solutions: Figure 3.4(a) shows the total number of solutions we have found in the screening of all reactions at different P and τ . One can see that, when a complete stop of the objective reaction is required ($\tau = 0$) there is no significant advantage in increasing the precision P . However, when it is necessary to induce a more accurate modulation of the flux (inhibitory when $0 < \tau < 1$), higher values of P allow to find a larger number of solutions. Through the partial inhibition, indeed, we can find solutions which are closer to the desired threshold, whereas the simple ON/OFF model can mostly induce a complete stop of the objective reaction. A similar improvement can be also identified while passing from $\tau = 1.5$ to $\tau = 2.0$.

Cardinality of the solutions: More details are presented in Figure 3.4b, which reports the histogram of the cardinality of the solutions and their mean values for the case of $\tau = 0.5$ (averages for each value of τ are reported in Table 3.3). When the precision increases, the distribution of the cardinality shifts slightly to higher values, meaning that multiple drug treatments are slightly preferred.

Perturbation induced by the solutions: For each solution that we have identified during this screening, also the corresponding perturbation (i.e. the side effect $\|\mathbf{v}^{\text{tr}} - \mathbf{v}^{\text{ut}}\|_1$) can be evaluated. We calculate the frequency of these perturbation values (regardless of the value of τ). The result is shown in Figure 3.4(c). We notice that at higher precision, smaller perturbations become (slightly) more probable: as expected, for high values of P , the algorithm can modulate the inhibition more accurately and therefore reduce the impact on the network, while still satisfying the request on the flux of the objective reaction.

Nonlinearity exploited by the solutions: The interaction between drugs is normally interpreted as the deviation of the effect of combined drugs with respect to the linear superposition of the single drug perturbations. Therefore, similarly to the scaled epistasis measure presented in literature [37], a index of nonlinearity $\eta(\mathbf{h})$ can be defined on the basis of the flux of the objective reaction as follows.

Let $v_{\text{mod}}^{\text{tr}}(h_1, h_2, \dots, h_d)$ be the flux of the objective reaction at drug inhibition $\mathbf{h} = (h_1, h_2, \dots, h_d)$ and given by (3.3). Then,

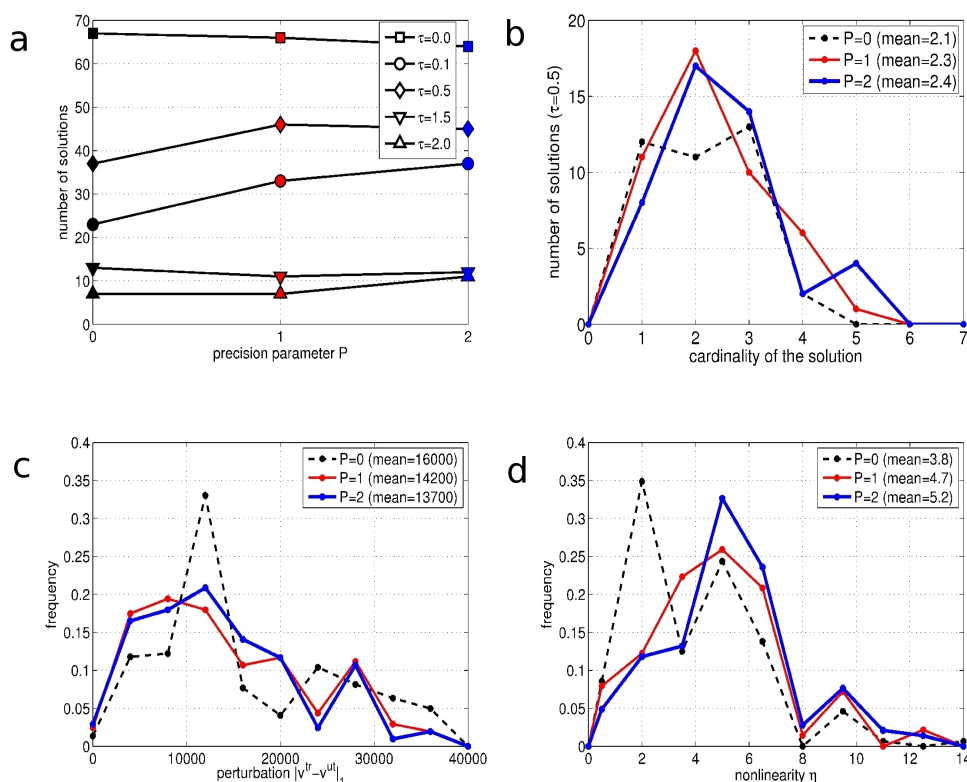


Figure 3.4: **Effect of the precision parameter P (results from the reactions screening).** (a) The plots show the number of solutions that are found for different values of τ (see legend) as function of P (color codes as in the following plots). As one can see, for $\tau > 0$, increasing the precision P , it is nearly always more frequent to find a drug combination which induces the sought modulation. (b) The plot is a detail of the curve in panel (a) for the case $\tau = 0.5$, and shows the frequency of the drug cardinality of the solutions (i.e. the number of drugs used in the solution). For larger P , the distribution is slightly shifted to higher cardinality (data for each τ are reported in Table 3.3). (c) Beside the number of solutions, higher precision produces more selective outcomes, i.e. with a lower side effect (as clearly shown by the mean values reported in the legend). The counting is performed over all different values of the threshold τ . (d) This plot shows the histogram of the values of the nonlinearity index $\eta(\mathbf{h})$ (as reported in (3.9)) calculated for all the solutions of the screening (still regardless the value of τ). From the curves and from the mean values reported in the legend, it is possible to see that higher amount of nonlinearity are obtained when P is increased.

Table 3.3: **Averaged cardinality of the drug combinations from the screening.** Figure 3.4(b) shows the details of the screening results for $\tau = 0.5$ and the corresponding legend reports the averaged value of the cardinality of the solutions which have been found; we summarize here the same quantities for all values of τ .

| Precision | $\tau = 0.0$ | $\tau = 0.1$ | $\tau = 0.5$ | $\tau = 1.5$ | $\tau = 2.0$ |
|------------------|--------------|--------------|--------------|--------------|--------------|
| $P = 0$ | 2.5 | 2.9 | 2.1 | 2.1 | 1.4 |
| $P = 1$ | 2.7 | 3.0 | 2.3 | 3.5 | 3.1 |
| $P = 2$ | 2.9 | 3.3 | 2.4 | 3.7 | 3.2 |

$$\eta(h_1, \dots, h_d) := \frac{v_{\text{mod}}^{\text{tr}}(h_1, 0, \dots, 0) + \dots + v_{\text{mod}}^{\text{tr}}(0, \dots, 0, h_d) - (d-1)v_{\text{mod}}^{\text{ut}} - v_{\text{mod}}^{\text{tr}}(h_1, \dots, h_d)}{v_{\text{mod}}^{\text{tr}}(0, 0, \dots, 0) - v_{\text{mod}}^{\text{tr}}(h_1, \dots, h_d)}, \quad (3.9)$$

since it holds $v_{\text{mod}}^{\text{tr}}(0, 0, \dots, 0) = v_{\text{mod}}^{\text{ut}}$. From this definition, $\eta = 0$ means linear behavior and $\eta > 0$ nonlinear. Therefore, for each solution of the screening, we calculate the corresponding $\eta(\mathbf{h})$ and we analyzed the distribution of its values (still ignoring the parameter τ): the result is shown in panel (d) of Figure 3.4. It is clear that increasing the value of P the nonlinearity index tends to be higher. It seems that, thanks to the higher precision, the algorithm may exploit more efficiently the nonlinearity property and, by consequence, it can limit the dosage of the drug and consequently reduce the perturbation.

3.3.2 Drug interaction surfaces: three case studies

For three of the solutions found through the screening procedure, we detail now the drug interactions exploited by the algorithm. In particular, we considered the synergisms in the inhibition of transketolase and of ribose-5-phosphate isomerase, and the synergism in the up-regulation of glutamate dehydrogenase (respectively indicated by codes TKT1, RPI and GLUDy in Figure 3.3). Each of the first two solutions contains a pair of drugs (Fomepizole plus Halofantrine and Fomepizole plus Hexachlorophene respectively). We explore the drug interaction surface changing the amount of inhibition induced by each compound, as could correspond in experiments to using different drug dosages (the interval $[0, 1]$ has been discretized using (3.7) with $P = 4$). The 2D surfaces are reported in Figure 3.5 panels (a) and (d). In the third case (up-regulation of glutamate dehydrogenase, panel (g)) the synergism is obtained combining three drugs (Nitrofurazone, Halofantrine and Pemetrexen); therefore, in order to have the 2D surface of interaction, the first drug is kept at the optimal inhibition value ($h_1 = 1$) and the combinations are explored changing the dosages of the remaining two drugs. Figure 3.5 reports also the nonlinearity index $\eta(\mathbf{h})$ as defined in (3.9): panels (b), (e) and (h) show that in all cases there is a clear enhancement of the effect when the drugs are combined.

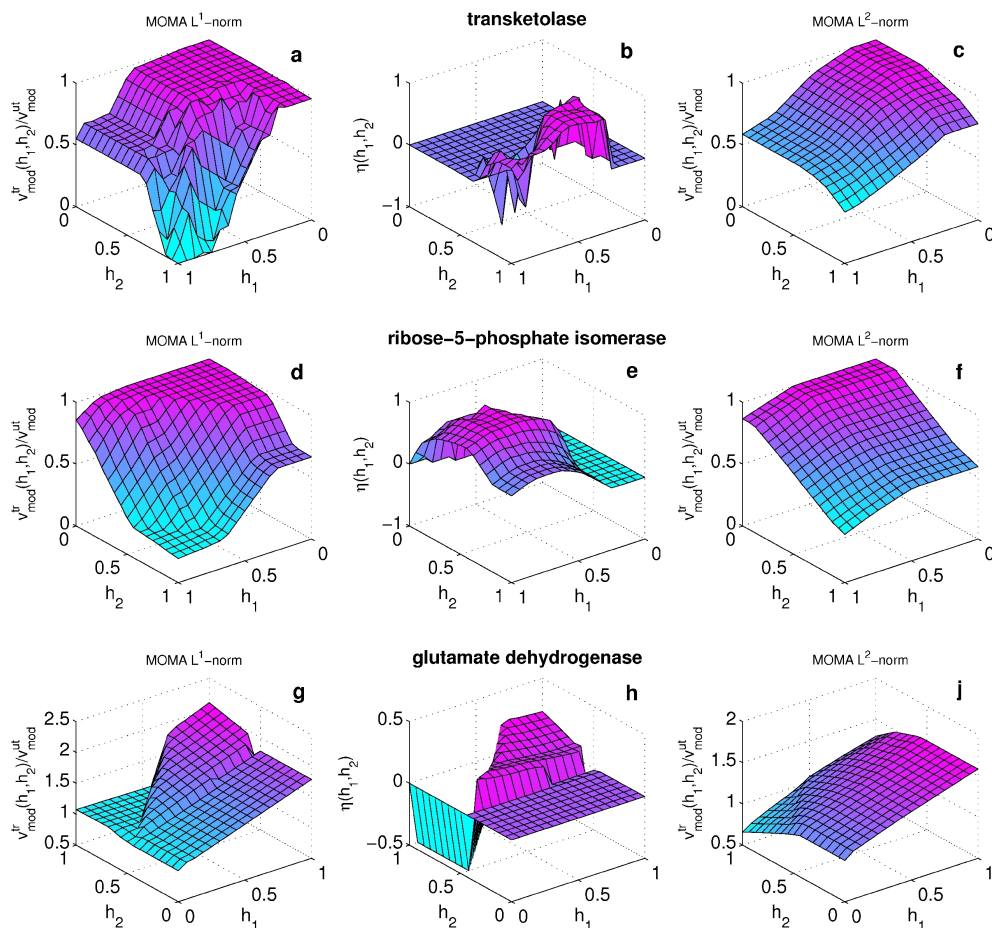


Figure 3.5: **Examples of surface of drug interaction.** Axes x and y report the inhibition coefficients h (dosage) for the two drugs; the z -axis reports the percentage of the flux through the modulated reaction after drug treatment calculated with L^1 formulation of MOMA with respect to the untreated value (left panels a, d and g), or the relative deviation, $\eta(\mathbf{h})$, of the effect induced by the drug combination from the linear superposition of the effects due to the single drugs as expressed by (3.9) (middle column, panels b, e and h). The right panels (c, f and j) show the calculations made with MOMA based on the L^2 -norm. The higher smoothness of the surface makes this formulation more reliable: however the surfaces obtained with our method (L^1 -norm) well reproduce the main characteristic of the synergism between the drugs. In the first two rows inhibitory synergisms are shown ($\tau < 1$), whereas the third row is an example of activating synergism ($\tau > 1$). Notice that for the sake of readability, in the plots of the last row, both x and y axes have been inverted.

These calculations has been performed with MOMA based on the L^1 -norm because, as already mentioned, it allows a definition of a linear function in the optimization problem. When compared to the surfaces obtained from the original quadratic formulation of MOMA (last column: panels (c), (f) and (j) of Figure 3.5), the results from L^1 -norm show some irregularity of the surfaces (which makes the original L^2 version more reliable) but the main features of the drug–drug interaction are still well described.

3.3.3 The problem due to the nonuniqueness of \mathbf{v}^{ut}

As mentioned after Eq. (3.1), the optimization of the biomass production may lead to a nonunique solution \mathbf{v}^{ut} . We presume that the nonuniqueness is mainly due to the presence of circulations (i.e. loops) and alternative equivalent pathways in the network, a problem which is well known in FBA and which is not ascribable to our algorithm. Nevertheless, we analyze here the behavior of the algorithm with respect to the variability of the unperturbed fluxes (one of the inputs of the algorithm).

Nonunique solution means that the set $M \subset W$ of solutions at the maximal growth rate $v_{\text{biomass}}^{\text{max}} = v_{\text{biomass}}^{\text{ut}}$ can be defined jointly by the standard FBA constraints and by the extra equality constraint on v_{biomass} :

$$\begin{aligned} \mathbf{S}\mathbf{v} &= \mathbf{0} \\ 0 &\leq v_i \leq U_i \quad \forall i = 1, \dots, r \\ v_{\text{biomass}} &= v_{\text{biomass}}^{\text{max}}. \end{aligned}$$

In order to perform a robustness analysis of the algorithm, we sample points in M and apply the following procedure:

- 1) We choose again ribose-5-phosphate isomerase as reaction of the *E.coli* core metabolic network to be modulated and we set $P = 2$, $\tau = 0.6$.
- 2) The reference fluxes of the “untreated” network with optimal biomass production are denoted by $\mathbf{v}^{\text{ut}}(0)$, whereas $\mathbf{v}^{\text{tr}}(0)$ is the corresponding solution of the algorithm (“treated” metabolic fluxes). The counter q is set equal to 0.
- 3) The counter q is increased. Starting from $\mathbf{v}^{\text{ut}}(q-1)$ we randomly choose a new vector of fluxes which has the same maximal growth rate (in practice we do a random walk in M). This is achieved by adding a Gaussian random vector to $\mathbf{v}^{\text{ut}}(q-1)$ and projecting the new point on M (the projection of a Gaussian random vector on a hyperplane is still Gaussian). This generates $\mathbf{v}^{\text{ut}}(q)$.
- 4) We run the algorithm using $\mathbf{v}^{\text{ut}}(q)$ and obtaining the corresponding $\mathbf{v}^{\text{tr}}(q)$.
- 5) Points 3) and 4) are repeated iteratively for 100 times.

In the end we have two set of vectors: $\{\mathbf{v}^{\text{ut}}(j)\}_{j=0,\dots,100}$ and the corresponding $\{\mathbf{v}^{\text{tr}}(j)\}_{j=0,\dots,100}$. For each pair (j, k) , $j, k = 0, \dots, 100$, $k > j$, we calculate the normalized distances:

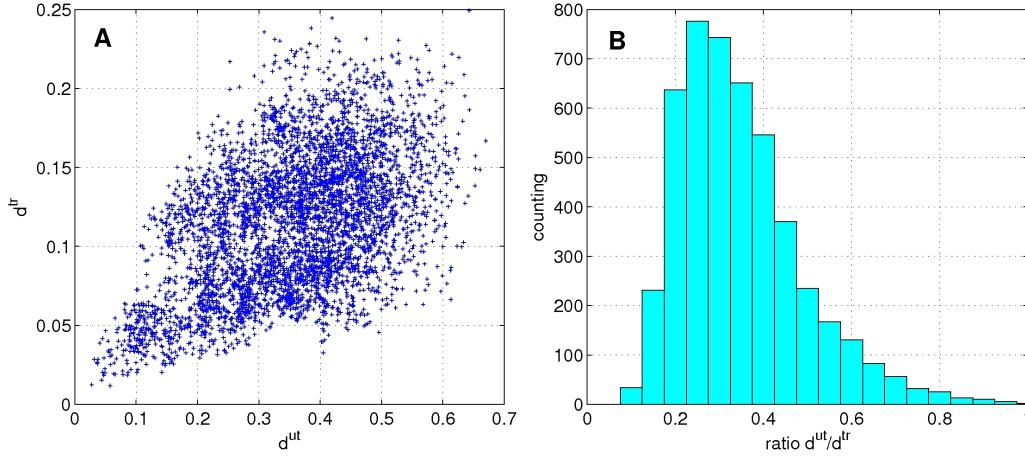


Figure 3.6: **Results from 100 different unperturbed fluxes \mathbf{v}^{ut} (4950 pairs) at the same biomass production $v_{\text{biomass}}^{\text{max}}$.** (A): Correlation between $d_{j,k}^{\text{ut}}$ and $d_{j,k}^{\text{tr}}$, ($j, k = 1, \dots, 100$, and $j > k$). (B): Distribution of the ratio $d_{j,k}^{\text{tr}}/d_{j,k}^{\text{ut}}$.

$$d_{j,k}^{\text{ut}} = \frac{\|\mathbf{v}^{\text{ut}}(j) - \mathbf{v}^{\text{ut}}(k)\|_1}{\|\mathbf{v}^{\text{ut}}(0)\|_1},$$

$$d_{j,k}^{\text{tr}} = \frac{\|\mathbf{v}^{\text{tr}}(j) - \mathbf{v}^{\text{tr}}(k)\|_1}{\|\mathbf{v}^{\text{ut}}(0)\|_1}.$$

Both normalizations are performed with respect to the same quantity $\|\mathbf{v}^{\text{ut}}(0)\|_1$. Results are reported in Figure 3.6. Panel A shows that there is a certain degree of correlation between the two distances.

The histogram analysis of the ratio between them (i.e. $d_{j,k}^{\text{tr}}/d_{j,k}^{\text{ut}}$) is presented in panel B. Since this ratio is always smaller than 1, these results indicate that starting from different \mathbf{v}^{ut} , the deviation among the corresponding solutions \mathbf{v}^{tr} of the algorithm is smaller than the variation of the \mathbf{v}^{ut} , meaning that the algorithm reduces (or at least it does not amplify) the variability of the starting point.

This can be justified by the fact that the algorithm is a “contraction”, i.e. denoting by \mathcal{F} the map corresponding to the algorithm, for any pair of vectors \mathbf{v} and \mathbf{w} , the following inequality holds:

$$\|\mathcal{F}(\mathbf{v}) - \mathcal{F}(\mathbf{w})\|_1 \leq \|\mathbf{v} - \mathbf{w}\|_1.$$

Indeed, the algorithm starts with the unperturbed convex set W of feasible fluxes defined by the standard FBA constraints. Then, this set is reduced to a proper convex subset $W(\mathbf{h})$ through the inhibition induced by the drugs (selected according to the modulation of the objective reaction and to the side effect). Finally, MOMA performs a projection of the unperturbed fluxes \mathbf{v}^{ut} on this subset. Therefore, as for any projection on a convex set in a metric space (the metrics being induced by the L^1 norm), the algorithm reduces distances.

3.3.4 Computational performances.

The proposed algorithm has been implemented in MATLAB (2012R) and the optimization has been performed using ILOG-IBM CPLEX 12.1 under academic license.

A benchmark study

First, the impact of the parameter P on the computational cost is evaluated on the core metabolism of *E.coli*: we choose one of the metabolic reactions for which we characterize the corresponding drug interaction surface (namely ribose-5-phosphate isomerase, see Figure 3.5) as reaction to be modulated ($\tau = 0.35$) and run the algorithm with different values of P from 0 to 5, recording the computational time required to find the solution. Since there are 8 drugs, the extremal values of P correspond to 8 and 48 Boolean variables in the whole problem. Notice that for $P = 5$, the accuracy on the definition of h_k is quite high ($2^{-P} = 1/32 < 5\%$).

In addition, we estimate the time needed to perform the evaluation of the inhibitory effect (MOMA) of a single drug combination as an average over a 20 random subsets of the 8 available drugs. From this value we can predict the approximate computational cost of an exhaustive search over all possible drug combinations (and dosages). The comparison of the performances of the algorithm with this estimation is plotted in panel (a) of Figure 3.7.

Performances on larger networks

Moreover, we run the algorithm on the metabolic network of the six microorganisms listed in Table 3.2. Our scope is to evaluate the impact of the size of the network (parametrized by the number of reactions r) on the computational performances. In order to limit the interference of other parameters, these calculations are carried out with the same objective reaction (in particular we still keep ribose-5-phosphate isomerase since it appears on all networks we have considered) at constant precision ($P = 2$) and threshold ($\tau = 0.35$), with the same number of drugs ($d = 8$), and choosing their inhibition targets in a random manner (unfeasible problems are ignored). However, since on very large networks it is quite unlikely to induce the sought modulation on the objective reaction when only 8 targets are inhibited, the number of targets of each drug is proportionally increased (on average the total number of inhibitions is approximately 6% of the total number of reactions). Because of the randomness in the choice of the targets, the computational times may present a significant variation. Therefore, Figure 3.7(b) shows the whole distributions of the computational time over 100 runs for each one of the six metabolic networks we considered. Finally, Figure 3.7(c) reports the mean and the standard deviation of these distributions as a function of the size of the network. On average, also for very large networks, the computational time is approximately one hour (on a 2.3 GHz CPU).

All these characterizations show the good performances of the algorithm.

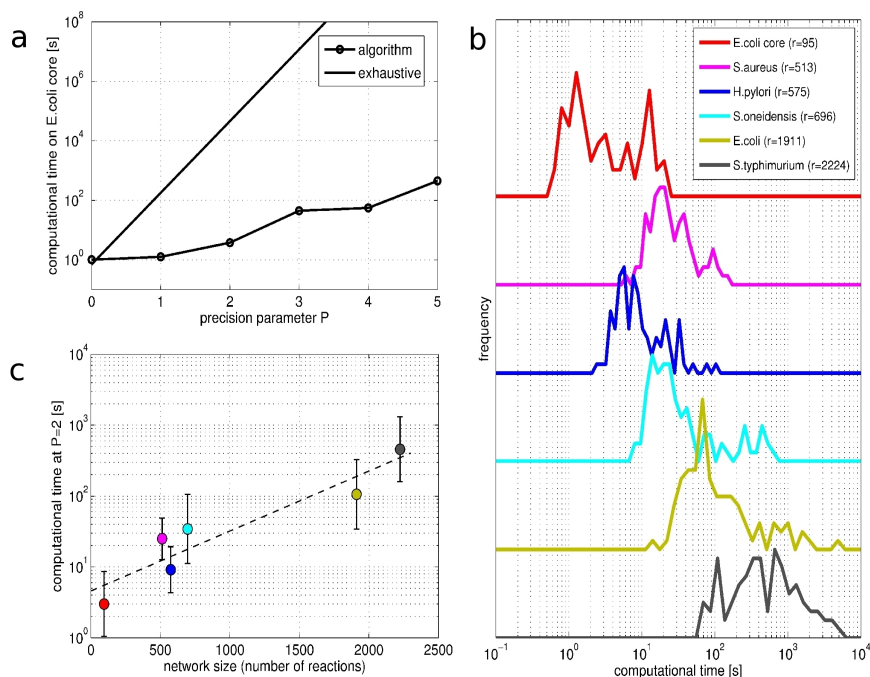


Figure 3.7: **Algorithm performances.** (a) Computational time of the algorithm compared with an estimation of the exhaustive search for different values of P on the *E. coli* core metabolism. (b) The influence of the size of the network on the performances is here reported as distribution of the computational time for the 6 metabolic networks listed in Table 3.2. Each distribution is built over 100 random samples obtained by changing the targets of a fixed number of drugs and keeping the precision parameter and the threshold constant ($P = 2$ and $\tau = 0.35$). (c): Averaged computational time (same color code of the corresponding distributions reported in panel (b)) and error bar as a function of the size of the network (expressed as the number of reactions r reported Table 3.2).

3.4 Discussion and Conclusions

Optimization is a concept widely used in many scientific fields; for instance, in systems biology, FBA makes use of it for discriminating reaction fluxes in large metabolic networks. Following the same philosophy, in order to cope with more complex situations, multiple optimization criteria can be needed simultaneously leading in some situations, like the one discussed in this work, to a bilevel optimization problem. The bilevel approach is promising for studying several features and applications of metabolic networks, for instance for identifying metabolic objective functions [23] or for studying perturbations around a nominal optimum [24, 44]. In the context of drug combinatorics, in order to efficiently solve the bilevel optimization, Boolean variables are commonly used in the outer problem. However, this ON/OFF description of the corresponding biological quantities may represent a very rough approximation, as it is the case for the (partial) inhibition induced by drugs

acting on the enzymes of a metabolic network.

In order to overcome this limitation, we propose an improvement on the formulation of the bilevel optimization in which a single Boolean variable is replaced by a convex combination of several Boolean quantities: in this manner the convex and linear nature of the problem is preserved and the description of the inhibitory effects becomes more realistic. Since the problem contains Boolean variables, the optimization falls in the MILP class: compared to LP, the *NP*-hard complexity of MILP [27] makes the new algorithm more expensive from a computational point of view. For the tasks at hand (see Figure 3.7), the algorithm behaves well also for large metabolic networks. The logarithm of the computational time scales linearly with the number of reactions, but with a small slope, so that on average the solution is found in a reasonable computational time, also for networks with around 2500 reactions and for $P = 2$.

For testing purposes, we run the algorithm on the central carbon metabolism of *E.coli* screening all reactions. We have found that increasing the number of Boolean variables used in the convex combination (the precision parameter P), it is more likely to find a solution which succeeds on the modulation of the objective reaction (see Figure 3.4(a)). In particular, partial inhibitions (i.e. modulations of the dosage of the drugs) are more frequent for multicomponent solutions (panel (b) of Figure 3.4): this result may be interpreted as a wider possibility, offered by the synergism, to calibrate a drug treatment according to the specific needs. Moreover our computations represent a confirmation on large networks of the expected, but still not verified, higher efficiency of multiple targets drug treatments in presence of partial inhibition [32]. In this perspective, the results show that this approach may also lead to treatments which are more selective (panel (c) of Figure 3.4).

A possible explanation can be found in the unexpected or hardly predictable drug synergism which are typical of complex systems such as metabolic networks, even in a simplified framework like FBA. In particular, concerning the synergistic interactions between drugs, the analysis done through the drug–drug interaction surface (Figure 3.5) reveals that nonlinear effects, not explained by superposition of the single drug perturbation, are significant and can be captured and exploited by the method proposed, unlike with a more coarse-grained ON/OFF description. We should mention, that the three case studies presented in Figure 3.5 do not pretend to have any clinical value: they have been selected only for the purpose of illustrating the method and the advantages it may give in the context of drug synergism and drug reprofiling for reconstructed metabolic networks.

It is worth noting that the problem of drug synergism we presented here can be read as a guiding example for a more general class of situations: indeed, the idea we have proposed for treating bilevel optimization can be applied to any other case which requires a more realistic modeling with respect to the oversimplified ON/OFF description, in biology as well as in all the other fields where LP is already used.

Chapter 4

Transient dynamics in the metabolism of perturbed microorganisms

4.1 Motivations and state of the art

Metabolic networks, investigated through the tool of FBA as in the previous chapters, deals only with one of the layers of the cellular functions, seen as an independent entity isolated from the rest of the regulatory machinery of the cell. This is obviously a drastic simplification and new frontiers start to appear when, for example, the metabolism is coupled with other internal functions of an organism, such as transcriptional networks [88, 89].

For these purposes, several researchers have studied the response of the metabolism of microorganisms after various perturbations [90, 91, 92, 37], for instance, a gene knockout or a change of the nutrients. It is known [93] that the microorganisms normally stop temporarily to proliferate after such perturbations. It has been hypothesized that during this period many more enzymes are expressed and activated than in standard conditions: this activation of this high number of metabolic pathways (the so-called *latent pathways* [4]) is considered as a “safety measure” that the cells adopt in order to survive under such stressful conditions [21, 83]. Indeed, these processes are less efficient with respect to those normally running and not all of them are necessary for the proliferation. Only later, once the survival is guaranteed, the cells start to regulate the metabolic reactions, selecting those which better perform the synthesis of the biomass; following this adjustment strategy, the proliferation rate can often be fully recovered. Figure 4.1 shows how a dynamics of this survival strategy may look like, according to [94].

As mentioned in Chapter 1, in FBA the perturbed networks are described using heuristic criteria such as MOMA and ROOM. While MOMA seems more reliable in the quantification of the effect shortly after the perturbation, ROOM seems better for a later stage of the metabolic adjustment. However, no procedure is available for

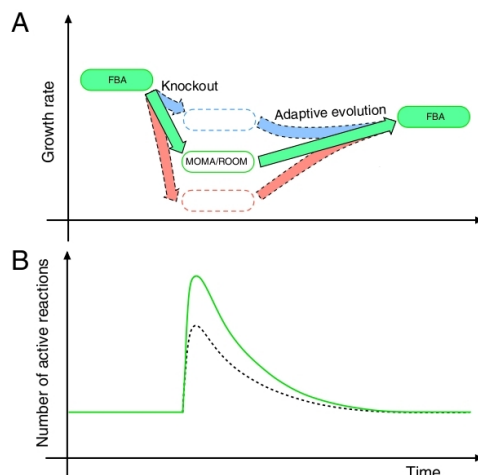


Figure 4.1: **Hypothetical dynamics of metabolic adaptation after a gene knockout [94].** (A): initial drop and recovery of the growth rate. (B): number of active reactions, which is high after the perturbation and progressively reduces during the adaptation. Dotted line indicates alternative trajectories.

the modeling of the entire transient dynamics and for describing the recovery of the growth rate that follows. For this reason, the scope of our work in this chapter is to develop a new heuristic criterion for FBA, able to predict the dynamics of the overall adjustment of the metabolism that a microorganism adopts after a perturbation.

4.1.1 Experimental data from literature

In order to present our procedure, we refer to a set of biological measurements available in the literature. In particular, we consider the data generated by Fong and Palsson [93]: their experiments have been carried out on *Escherichia coli* strains with different gene knockouts. After a culture for a night in a growing rich medium (i.e. with many types of nutrients such as carbohydrates, aminoacids, vitamins...) the microorganisms are transferred into a single carbon source medium. The authors have shown that, after this transfer, the microorganisms have a temporary reduction of the proliferation rate, which is recovered partially or almost totally depending on the carbon source and on the gene knockout (results are reported in Figure 4.2). The authors use FBA for both the wild type and the knockout strains in order to describe the growth rate in various growing media. However, as FBA is unable to describe the transient dynamics, the authors have found a good agreement only in terms of the final growth rate, without providing any description of the strategy adopted by the *E.coli* strains to get there.

In the following, the heuristic criterion for metabolic adjustment is first presented. A few (preliminary) results of its application to the central carbon metabolism of *E.coli* [87] are shown next.

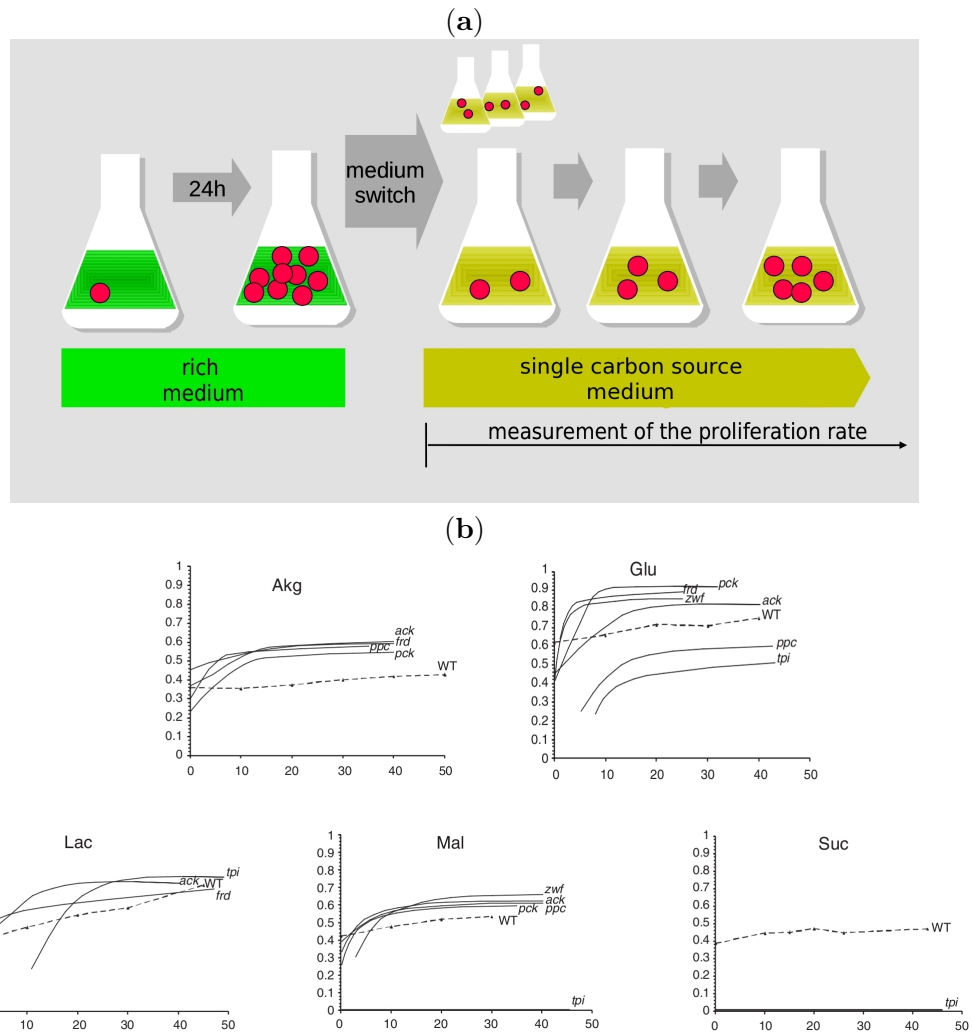


Figure 4.2: **Adaptive evolution experiments from [93].** (a): Scheme of the experimental protocol: after an overnight culture in a rich medium, the microorganism is transferred in a single carbon source medium. During the following days the growth rate is measured. (b): Each plot reports the growth rates (h^{-1}) of the *E.coli* strains with different gene knockouts in growing media with various carbon sources. Deleted genes: *ackA*=acetate kinase A, *frd*=fumarate reductase, *pck*=phosphoenolpyruvate carboxykinase, *ppc*=phosphoenolpyruvate carboxylase, *tpi*=triosephosphate isomerase; *zwf*=glucose 6-phosphate-1-dehydrogenase. Carbon sources: Akg= α -ketoglutarate; Glu=glucose; Lac=lactate; Mal=malate; Suc=succinate.

4.2 Methods

In this section, the following symbols are used: the metabolic network of *E.coli* is represented by the $m \times r$ stoichiometric matrix \mathbf{S} ; the lower and upper bounds of the reaction fluxes are \mathbf{L} and \mathbf{U} , respectively. Since nutrients inflows are among the fluxes considered in the metabolic network, \mathbf{L} and \mathbf{U} are related not only to the activation (biochemical or allosteric) and to the concentration of the enzymes (i.e. the gene expression), but also to the growing medium in which the microorganism is living. The constraints typical of FBA define the polytope

$$W(\mathbf{L}, \mathbf{U}) = \{\mathbf{v} : \mathbf{S}\mathbf{v} = \mathbf{0}, \mathbf{L} \leq \mathbf{v} \leq \mathbf{U}\},$$

here denoted as $W(\mathbf{L}, \mathbf{U})$ because it is viewed as function of the lower and upper bounds. For a given vector \mathbf{v} (the reaction fluxes), the growth rate G is defined by the inner product $\mathbf{b}^T \mathbf{v}$, where \mathbf{b} is the composition of the building blocks of the cellular biomass. This is the standard cost function of FBA [20].

In order to describe the adjustment of the metabolism when passing from rich medium to single carbon source, we will use a time-varying dynamics for the fluxes and the bounds, represented by the sequence

$$\{[\mathbf{v}^{(t)}, \mathbf{L}^{(t)}, \mathbf{U}^{(t)}]\}_{t=0,1,2,\dots},$$

where $\mathbf{v}^{(t)}$, $\mathbf{L}^{(t)}$ and $\mathbf{U}^{(t)}$ denote the reaction fluxes, the lower and upper bounds at time t , respectively (during the dynamics some components of \mathbf{L} and of \mathbf{U} can be set to zero because of the removal of the exchanged nutrients or because of the deactivation of the corresponding enzymes or genes).

4.2.1 The starting point of the transient

We explain here how the initial part of the experiments of Fong and Palsson can be described by the criteria already available.

The first day: The experiments start with an overnight in a rich medium; we indicate this initial condition by the polytope $W(\mathbf{L}^{(0)}, \mathbf{U}^{(0)})$, where the label (0) indicates the initial time $t = 0$. It is known that, for unperturbed networks, the maximization of the growth rate provides the fluxes

$$\mathbf{v}^* = \arg \max_{\mathbf{v} \in W(\mathbf{L}^0, \mathbf{U}^0)} \mathbf{b}^T \mathbf{v}.$$

Therefore, in the case of the wild type of *E.coli*, the initial metabolic state $\mathbf{v}^{(0)}$ is set equal to \mathbf{v}^* . On the contrary, in the case of microorganisms with a knockout (i.e. with $v_k = 0$, where k is the reaction whose enzyme is coded by the deleted gene), we know that MOMA provides a reliable description of the short-term metabolic adjustments. Therefore, since we are interested only in the first overnight in the rich medium, the metabolism of the microorganisms having a gene deletion is calculated as follows:

$$\mathbf{v}^{(0)} = \arg \min_{\mathbf{v} \in W(\mathbf{L}^0, \mathbf{U}^0) \cap \{v_k=0\}} \|\mathbf{v} - \mathbf{v}^*\|_2,$$

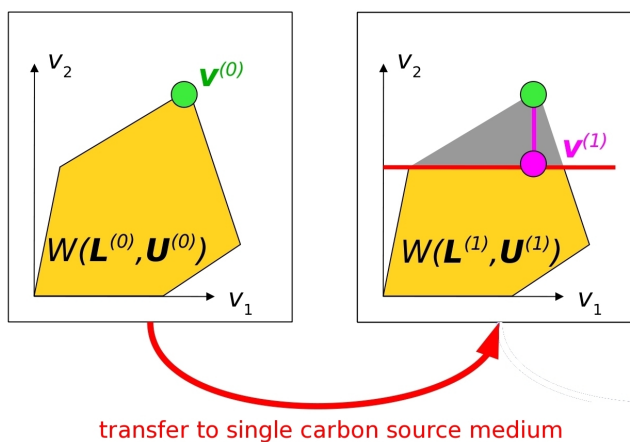


Figure 4.3: **Computational description of the initial step of the experiments.** The starting point in the rich medium is the vector of fluxes $\mathbf{v}^{(0)} \in W(\mathbf{L}^0, \mathbf{U}^0)$. After the transfer to a single carbon source medium, the metabolic fluxes $\mathbf{v}^{(1)}$ are calculated by MOMA (a orthogonal projection on the subset $W(\mathbf{L}^1, \mathbf{U}^1)$).

where $\|\cdot\|_2$ indicates the L^2 -norm.

The second day: At the beginning of the second day, the rich growing medium is substituted with a single carbon source medium: formally, the lower and upper bounds of the exchange reactions of the nutrients which are no longer available are set equal to zero. The polytope W is now defined by $\mathbf{L}^{(1)}$ and $\mathbf{U}^{(1)}$. As previously done, in order to obtain the one-day quantification of the metabolic adjustment of the microorganism (i.e. $\mathbf{v}^{(1)}$), the short term description provided by MOMA (with respect to $\mathbf{v}^{(0)}$) is considered again sufficiently reliable. We calculate then:

$$\mathbf{v}^{(1)} = \arg \min_{\mathbf{v} \in W(\mathbf{L}^1, \mathbf{U}^1)} \|\mathbf{v} - \mathbf{v}^{(0)}\|_2.$$

A computational description of this initial step is represented in Figure 4.3.

Both steps performed so far are standard in the literature of perturbed FBA. Proceeding further with these criteria is however impossible, unless we introduce some other mechanisms in the method, like the “reaction turn-off” described in the next section. It is worth noting that, among the possible ways used by a cell to stop a reaction, there is the gene resiliencing (transcriptional regulation of the gene expression) but also the allosteric interaction and the biochemical deactivation of the enzyme. For this reason, such cellular action will be simply called *turn-off* in the following.

4.2.2 An heuristic criterion for long-term metabolic adjustment

Generally, MOMA provides a low proliferation rate G and a higher number of active reactions N than FBA [21]. This is valid also for the fluxes $\mathbf{v}^{(1)}$ obtained for the second day of the adaptation (for which $G^{(1)} = \mathbf{b}^T \mathbf{v}^{(1)}$ and $N^{(1)} = |A(1)|$, where $A(s) = \{i : v_i^{(s)} \neq 0\}$ is the set of active reactions at time $t = s$ and $|A(s)|$ its cardinality), coherently with the low proliferation observed after the perturbation. Hence, in order to recover the growth rate, a reasonable strategy is that these survival-enabling “latent” pathways are progressively stopped. This progressive turn-off should be accompanied by a progressive restoring of the growth rate, up to a complete recovery (or more precisely, as complete as possible, given the carbon source and the knockout) as per FBA prediction.

Let us look at the move from $t = 1$ to $t = 2$ (expect for the first and second day, the time scale will be in arbitrary units). For instance, the cell turns off the enzyme of the reaction k (an active reaction, i.e. $k \in A(1)$). This means that $L_k^{(2)} = U_k^{(2)} = 0$ and that new changes are caused in the metabolism. Since we are interested to the short time horizon after this resilencing (i.e. only at the time 2), these changes can be described again by MOMA. For a given k , we can calculate the reaction fluxes as follows

$$\mathbf{w}[k] = \arg \min_{\mathbf{w} \in W(\mathbf{L}^2, \mathbf{U}^2)} \|\mathbf{w} - \mathbf{v}^{(1)}\|_2 \quad (4.1)$$

and the corresponding proliferation rate $g[k] = \mathbf{b}^T \mathbf{w}[k]$. Of course, this can be done for any generic time $t \geq 1$. Note that, although $\mathbf{w}[k]$ and $g[k]$ represent the metabolic fluxes and the growth rate respectively, we do not use the standard symbols \mathbf{v} and G because (4.1) provides only the k -th prediction over all $N^{(1)}$ possibilities.

Indeed, we need now to decide which value of k will be chosen by the cell.

Population dynamics. We may assume that a cell does not know the final metabolic state it will reach after several adjustments. For this reason, only instantaneous information can be used.

For simplicity, let us consider the case in which, at time t , the cell has two possible strategies $k = 1$ and $k = 2$ (see Figure 4.4) which lead temporarily to proliferation rates $g[1]$ and $g[2]$ respectively (calculated using (4.1)). Denoting with $Q(t)$ the population of the cell culture at time t and referring to the usual equation for its growth

$$dQ(t)/dt = g[k]Q(t),$$

after an arbitrary time Δt the population can be:

$$\begin{aligned} Q_1(t + \Delta t) &= Q(t)e^{g[1]\Delta t} && \text{if the cells turn off reaction 1;} \\ Q_2(t + \Delta t) &= Q(t)e^{g[2]\Delta t} && \text{if the cells turn off reaction 2.} \end{aligned}$$

We may suppose that natural selection would have advantaged the strategy that gives the highest instantaneous growth rate: as consequence, the ratio between

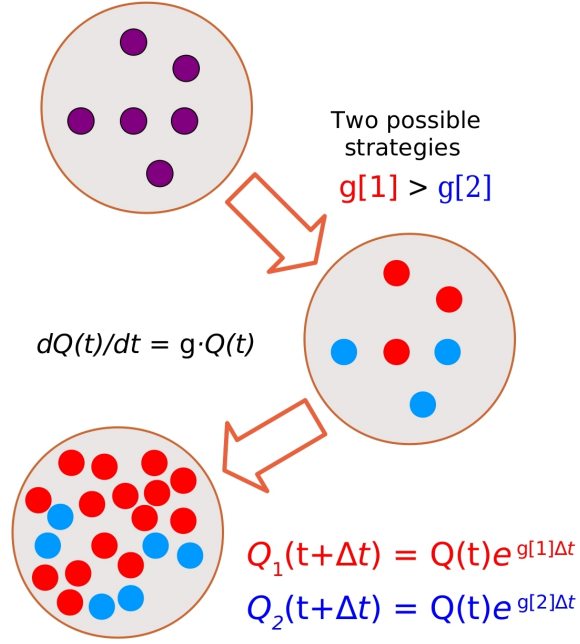


Figure 4.4: **Population dynamics.** Between two strategies that induce different effects, the more probable is that which induce a higher growth rate.

these populations indicates the probability \mathbb{P} for a cell to select the reaction to turn off: in particular we have

$$\mathbb{P}(k=1) = \frac{Q_1(t+\Delta t)}{Q_1(t+\Delta t) + Q_2(t+\Delta t)} = \frac{e^{g[1]\Delta t}}{e^{g[1]\Delta t} + e^{g[2]\Delta t}};$$

$$\mathbb{P}(k=2) = \frac{Q_2(t+\Delta t)}{Q_1(t+\Delta t) + Q_2(t+\Delta t)} = \frac{e^{g[2]\Delta t}}{e^{g[1]\Delta t} + e^{g[2]\Delta t}}.$$

Therefore, in our systems, among the $N(t)$ active reactions that a cell can turn off, the probability of turn off any reaction $k \in A(t)$ is given by

$$\mathbb{P}(k) = \frac{e^{\beta g[k]}}{\sum_{i \in A(t)} e^{\beta g[i]}}, \quad (4.2)$$

where we replace the arbitrary time Δt by the factor β . Indeed, (4.2) reminds the expression used in Statistical Physics [95] to calculate the probability distribution of the states of a system in which the denominator is the *partition function*

$$Z = \sum_{\mu} e^{\beta h(\mu)},$$

i.e. it is the summation over all possible states μ of the exponential of the Hamiltonian $h(\mu)$ multiplied by the Boltzmann factor $\beta = 1/k_B T$ (k_B denotes the Boltzmann constant and T the temperature). For this similarity, lower values of β (i.e. higher temperatures) would increase a random spread during the simulation over the possible cellular strategies (toward both lower and higher growth rates). In the following we arbitrarily set $\beta = 1/0.007$.

As the $g[k]$ are updated at each step, so are \mathbb{P} and Z . Clearly the partition function Z cannot be computed explicitly and we have to resort to a numerical simulation.

Simulation of the process. From Equation (4.2), given the state of the cell at time t (i.e. given $[\mathbf{v}^{(t)}, \mathbf{L}^{(t)}, \mathbf{U}^{(t)}]$), we can calculate the probability for all turn-off that a cell can adopt. The strategy is then chosen by generating a random index $x \in A(t)$ according to this probability distribution. This procedure is repeated until no turn-off provide any further advantage to the microorganism (i.e. until the growth rate cannot further increase).

Many random realizations (n) of such stochastic process are generated.

$$\{[\mathbf{v}^{(t)}(j), \mathbf{L}^{(t)}(j), \mathbf{U}^{(t)}(j)]\}_{t=0,1,2,\dots}, \quad j = 1, \dots, n$$

From each of them, we calculate the trajectories of $G^{(t)}(j)$ and of $N^{(t)}(j)$. The averages over the n samples generates the final dynamics $G(t)$ and $N(t)$. For direct comparison with experimental data, $G(t)$ is rescaled as in the paper of Fong and Palsson, i.e. such that it corresponds to the same substrate and oxygen uptake rates of the wild type in the same growing medium. This is achieved multiplying $G(t)$ by the ratio

$$\frac{v_{\text{biomass}}^{\text{Fong-Palsson}}}{\max_{W(\mathbf{L}^{(1)}, \mathbf{U}^{(1)})} v_{\text{biomass}}}.$$

4.3 Preliminary results

The method is applied to the *Escherichia coli* central carbon metabolism (details of this network can be found in Table 3.2 and Figure 3.3 of the previous chapter). The results, reported in Figure 4.5, show a good agreement with the experimental data of Figure 4.2(b): the cases where the growth rate was almost completely recovered have been progressively reproduced by our simulations. Similarly also for cases with a partial recovery. Moreover, also the number of active reactions is in agreement with the expected behavior reported in Figure 4.1, i.e. high at the beginning and reduced with time. The only difference concerns the dynamics of the wild type: for this strain, the prediction of the growth rate we obtain is always too high, although our starting value is much smaller than the experimental value. In spite of that, our final growth rate is always closer to the experimental data than the direct FBA computed by Fong and Palsson.

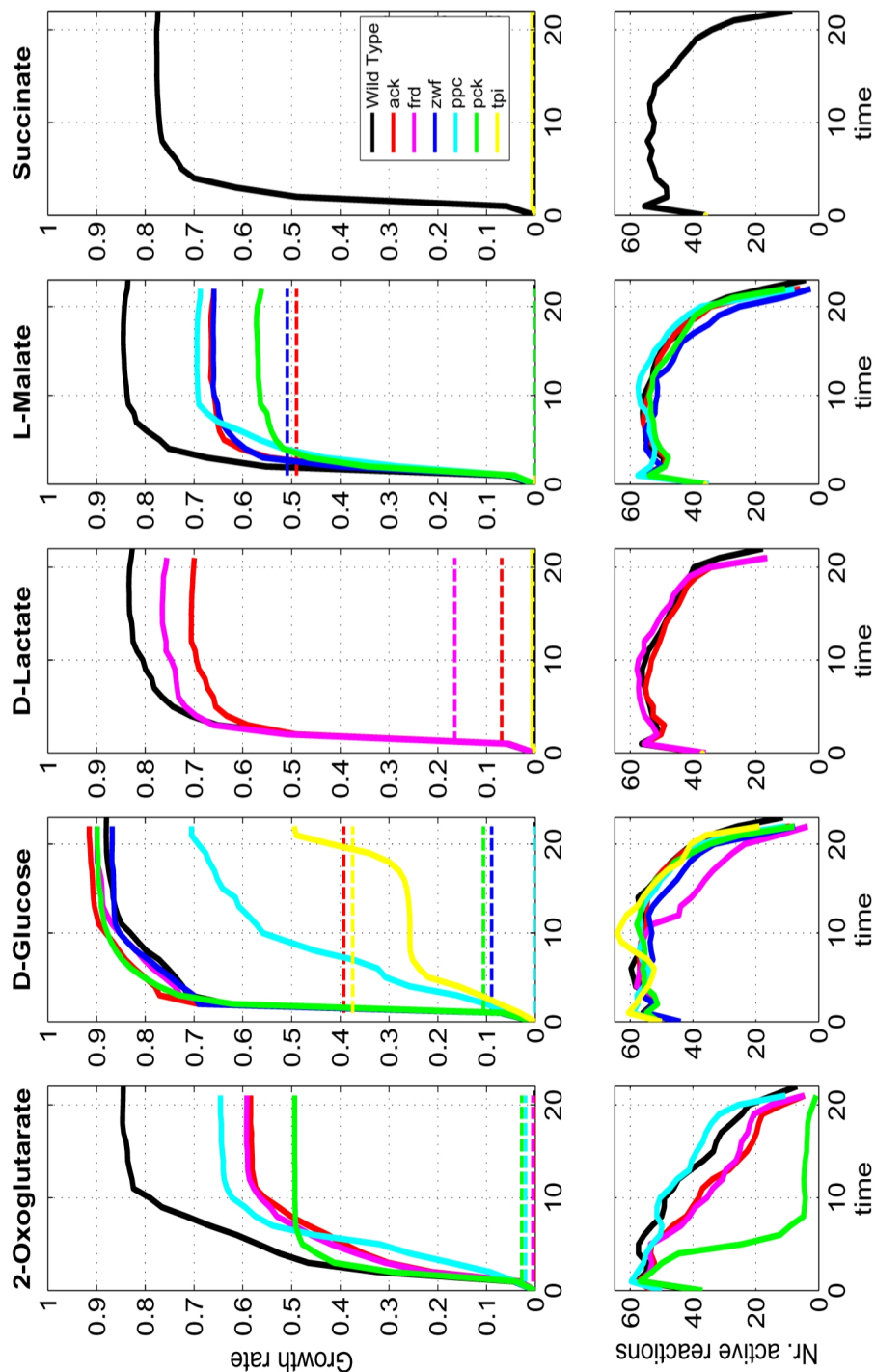


Figure 4.5: **Results of the simulations.** Mean value $G(t)$ (growth rate) and $N(t)$ (number of active reactions) over $n = 100$ random trajectories. Gene knockouts and growing media as in Figure 4.2b [93]. Dotted lines represent the growth rate predicted by ROOM.

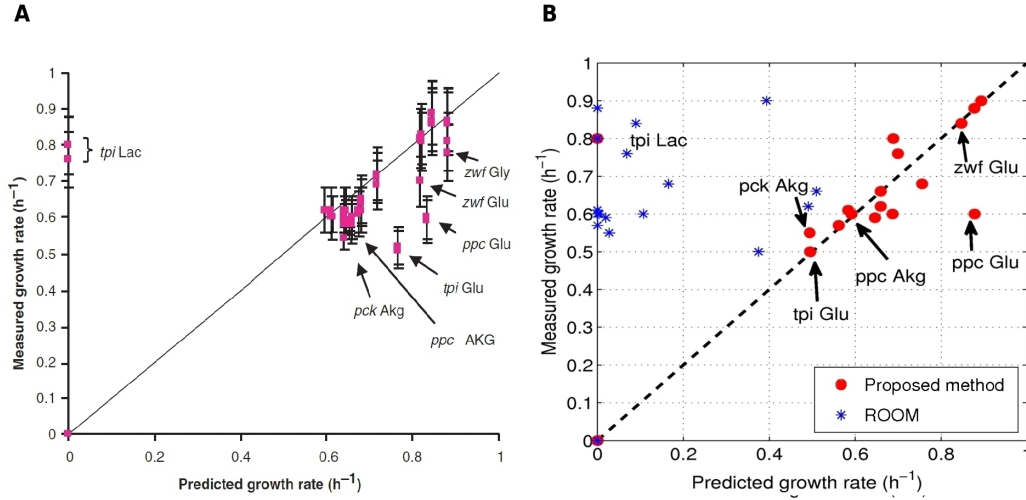


Figure 4.6: **Comparison with the theoretical prediction from the literature.** (A): correlation between experimental data and theoretical predictions indicated by Fong and Palsson [93]. The mean square errors is equal to 0.15 h^{-1} . (B): correlation between the same experimental data and our theoretical prediction: in our case the agreement is much better. The mean square error is 0.09 h^{-1} . Blue asterisks refer to ROOM predictions.

Indeed, we have also compared our results with the theoretical ones proposed in the paper of these authors. As mentioned, their *in silico* predictions concerned only the final point: in particular, the authors have calculated the maximal growth rate by the standard FBA criterion [20]. However, not all microorganisms reached the complete recovery of the growth rate: indeed, as reported in Figure 4.6A, the agreement they have obtained with respect to the experimental measurements is only partial. In our simulation, instead, just a few cases show a total recovery of the proliferation and a better agreement is thus obtained (see Figure 4.6B).

We calculate also the biomass production predicted by ROOM (after the change of the growth medium), since it is considered to be more reliable than MOMA in describing later stages in the metabolic adjustment. Indeed, with respect to results of MOMA (all with $G(0) = 0$, see initial points of the trajectories in Figure 4.5) the values of ROOM are higher (precisely, in the interval $[0, 0.5] \text{ h}^{-1}$, see dotted lines in Figure 4.5 and blue asterisks in Figure 4.6B). However, the fact that these values are randomly located between zero and the experimental data indicates that the time of the later stage to which ROOM refers is completely unclear and unpredictable. For this reason it cannot be considered a precise method for describing the transient dynamic of the metabolic adjustment.

4.4 Discussions and perspectives

These preliminary results indicate that the method we propose is able to generate a reliable description of the metabolic adjustment following a change of medium. Such good predictions are valid not only for what concerns the final growth rate but also for the whole trajectory. For this reason, we expect that our procedure can contribute to the understanding of the principles which control cellular metabolism. However, no measurement of genes regulation and of enzymes activation has been included in the paper we use as reference. In a more recent publication [94], the same research group of Palsson has performed a gene expression analysis during the adaptation of the metabolism in case of Lactate as carbon source. Several other papers (for instance [90, 91, 92]) report the experimental measurements of the rate of reactions of the central carbon metabolism (namely glycolysis and Krebs cycles) for *E.coli*. They represent then a set of new references potentially able to validate the method we have proposed here: for example, it may be possible to verify if the experimental sequences of the turned-off reactions are coherent with the theoretical predictions of our heuristic criterion.

Beside this, also the effect of the temperature parameter β should be investigated. Moreover, the probabilistic properties of the procedure viewed as a Markov process can be identified and characterized. The promising results obtained up to now encourage us to perform these additional studies.

PART II

QUANTIFYING DISORDER IN
COMPLEX SIGNED NETWORKS:
FROM GENE-REGULATORY TO SOCIAL NETWORKS

Introduction

In the previous part we have shown how FBA is able to capture the complexity of a metabolic network using a simple constraint-based model in which the kinetic parameters are completely neglected and only stoichiometric information is used. In the FBA approach, the “effective” fluxes are determined by the minimization (maximization) of a cost functional defined according to some heuristic criteria. In reality the yield of a reaction depends on the concentration and activity of the enzyme that catalyzes the reaction itself, hence on the expression of the gene that codes for the enzyme. This links the metabolic layer of a cell to the gene regulatory layer.

Unfortunately, the FBA formalism cannot be straightforwardly extended to the characterization of these regulations. In addition, our knowledge of the gene regulatory networks is largely incomplete even in the best studied prokaryotes, such as *E.coli*. For example in the RegulonDB database (the reference database for *E.coli* gene regulatory network) the coverage for what concerns the interactions among transcription factors and target genes is less than 50% of the genome (it is much worse for other organisms). For a vast fraction of these known interactions also a specification of the functional mode of action is available, namely transcriptional regulations are classified as activatory or inhibitory. This means that the gene network can be described as a signed graph, with nodes representing genes and signed edges the regulatory modes.

This model has similarities with the Ising spin-glass model used in Statistical Physics to describe disordered systems. As alternative, an Information Theory perspective can also be adopted. Both perspectives are explained in Chapter 5. Depending on the convenience of the investigation we need to perform, one of the two formalisms is taken as main reference. Even though nodes and edges are interpreted in a different manner in the two settings, the final computational issues are similar and related: hard optimization problems for which heuristic procedures are needed (LP is no longer suitable in these cases).

The main issue we investigate for our signed graphs is the amount of “true disorder” encoded in the network. In the Information Theory context of Lossy Source Compression codes used in Chapter 6 to deal with gene regulatory networks, this concept is called distortion. In Chapter 7 we discuss instead disorder in another

class of signed graphs drawn from social network theory. In this context the disorder is related to the notion of structural balance. The contents and the results presented in Chapters 6 and 7 are briefly summarized below.

Atypicality of regulatory networks (Chapter 6).

Gene regulatory networks, with their modes of transcriptional regulations (activation/inhibition), are treated as a lossy data compression problem for a coding system whose distortion is determined. For currently available gene networks, we show that the distortion is much lower than expected from null models, and that it is close to (when not lower than) the Shannon bound determined by the rate-distortion theorem.

These results highlight that the distribution of regulatory modes is highly atypical in the networks considered, and that this atypicality greatly helps in avoiding contradictory transcriptional actions.

Structural balance of social networks (Chapter 7).

According to Heider's theory of structural balance proposed in the 50's, social networks, whose edges represent friendly/hostile interactions among individuals, tend to be organized so as to avoid conflictual (called unbalanced in this context) situations represented as cycles of negative sign (i.e. frustrated cycles in the spin glass terminology). Using an algorithm for ground state calculation in large-scale Ising spin glasses, we compute the global level of balance of some very large on-line social networks: in spite of the "apparent disorder" present in these graphs, we verify that currently available networks are indeed extremely balanced (i.e. the frustration is very limited).

Moreover, by exploring the landscape of near-ground states we show that one of the networks we analyzed shows several distinct minima, separated by energy barriers determined by internally balanced sub-communities of users, a phenomenon similar to the replica-symmetry breaking of spin glasses. Passing from one valley to another requires to destroy the internal arrangement of these balanced sub-communities and then to reform it again.

The results of these chapters have been published in the following papers:

- G. Facchetti, G. Iacono, G. De Palo and C. Altafini. *A rate-distortion theory for gene regulatory networks and its application to logic gate consistency* Bioinformatics, **29**(9), 1166–1173, (2013).
- G. Facchetti, G. Iacono, C. Altafini. *Computing global structural balance in large-scale signed social networks*. Proc. Nat. Ac. Sci., **108**(52), 20953–20958, (2011).
- G. Facchetti, G. Iacono and C. Altafini. *Exploring the low-energy landscape of large-scale signed social networks*. Physical Review E, **86**, 036116 (2012).

Chapter 5

Theoretical framework: an overview

5.1 Ising spin-glass model

For us a spin glass is a signed graph in which a set of spins (nodes) are reciprocally interacting according to a defined topology (in Statistical Physics a regular lattice, for us a network of heterogeneous connectivity). In these models, each spin can exist in two or more states and the interactions between nearest neighbors can be of “ferromagnetic type” (represented as a positive weight on the edge) or “antiferromagnetic type” (represented as a negative weight on the edge). Each interaction has an energetic content which depends on the states of the interacting spins and on the value of the edge linking them: the sum of these energetic contributions defines an energy function called Hamiltonian h of the system. The study of h at different spin configurations may reveal important features of the network such as for instance clustering of sub-networks, appearance of spontaneous order (called, for instance, “magnetization” in case of spin of nuclei). The *Ising model* [96, 95] is one of the simplest cases of spin glass: spins have two possible states (which, without loss of generality, we consider now ± 1), see Fig. 5.1. Since we refer to pairwise undirected relationships only and since we assume they have all the same magnitude, a network of spins can be described by an undirected graph in which the n nodes are the spins and the m edges are their couplings. These couplings define the $n \times n$ adjacent matrix J of the network as follows:

$$J_{ij} = \begin{cases} +1 & \text{if there is a positive coupling between spin } i \text{ and spin } j; \\ 0 & \text{if there is no coupling between spin } i \text{ and spin } j; \\ -1 & \text{if there is a negative coupling between spin } i \text{ and spin } j. \end{cases} \quad (5.1)$$

As the signed graph is undirected, its adjacency matrix J is symmetric: $J_{ij} = J_{ji}$. Ignoring the single spin terms due to an external effect (e.g. an external magnetic field in case of spins of nuclei), the Hamiltonian $h(\mathbf{s})$ of the network J in a given

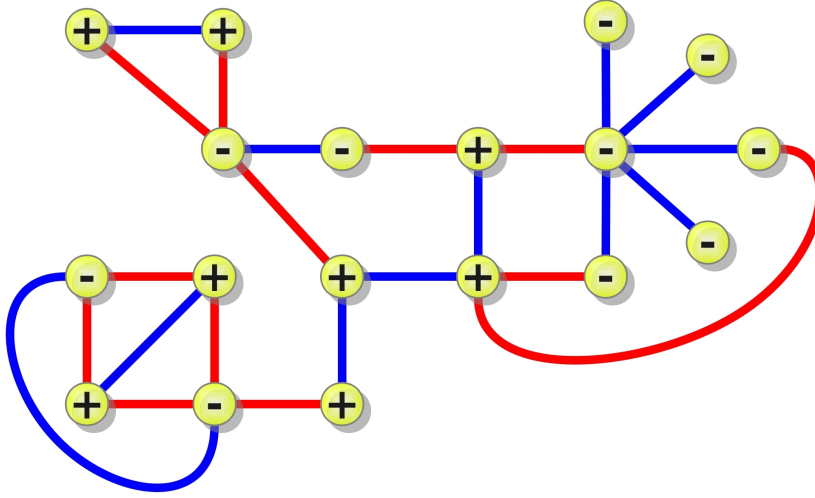


Figure 5.1: **Ising spin glass.** Example of an Ising spin glass on a planar network: spins are the nodes (“+” indicates $s_i = +1$, “-” indicates $s_i = -1$) and their coupling are the edges (blue: positive; red: negative). The assignment of the spins here reported corresponds to the minimal energy configuration. This network is not frustrated since all cycles are positive.

spin configuration $\mathbf{s} = [s_1 \dots s_n]^T$, $s_i \in \{\pm 1\}$, $i = 1, \dots, n$, is calculated as follows:

$$h(\mathbf{s}) = m - \frac{1}{2} \mathbf{s}^T J \mathbf{s} = \sum_{(i,j)} \frac{1 - J_{ij} s_i s_j}{2}, \quad (5.2)$$

where the summations runs over all adjacent pairs of nodes, i.e. it is a sum of $2m$ terms $(1 - J_{ij} s_i s_j)/2$ each of them equals to 0 or +1. The minimum of $h(\mathbf{s})$ is achieved in correspondence of a spin configuration \mathbf{s} for which as many as possible of the terms $(1 - J_{ij} s_i s_j)/2$ are null. When $J_{ij} > 0$ this requires $s_i = s_j$, when $J_{ij} < 0$ $s_i = -s_j$. The Hamiltonian has “true disorder” (or frustration) when not all terms of (5.2) can be “satisfied” (i.e. rendered $1 - J_{ij} s_i s_j = 0$) simultaneously. The Hamiltonian is the main tool for the characterization of the spin glass and of the associated network. Indeed, through the analysis of what is called “energy landscape” one may obtain information about the ground state and about other possible local minima of energy. As we will see, all problems presented in the following chapters need to search for the energetic minima and/or to study the main features of the energy landscape.

An exhaustive exploration of all configurations is prohibitive even for small networks (the number c of configurations scales exponentially with the number n of spins, in particular $c = 2^n$). Several methods have been proposed in order to collect as much information as possible (for instance the energy of the local minima and their locations) with a reasonable computational cost. In spite to some computational limits which start to appear with medium size systems (approximately over 10^3 spins), Monte Carlo with its large number of variations [97] is the most diffuse

method for this purpose. In the next section, we are going to present a heuristic algorithm based on the concept of gauge transformation which is capable of performing efficiently the energy minimization of an Ising spin glass.

Although we refer to spins such that $s_i \in \{\pm 1\}$, the presented formalization can be applied to any other pair of possible spin values (here called v_1 and v_2). For instance, in case of a signed graph representing a gene regulatory network, it would be more convenient to consider “expressed/unexpressed” as the two possible states, i.e. $v_1 = 1$ and $v_2 = 0$, respectively. Since we have two pairs of values (namely $\{-1, +1\}$ and $\{v_1, v_2\}$), we can always find a linear (bijective) map F converting one pair to the other. In particular:

$$\begin{aligned} F : \{-1, +1\} &\longrightarrow \{v_1, v_2\} \\ s_i &\longmapsto x_i \end{aligned}$$

and its inverse are given by the following expressions:

$$x_i = F(s_i) = \frac{(v_2 - v_1)s_i + (v_2 + v_1)}{2}; \quad (5.3)$$

$$s_i = F^{-1}(x_i) = \frac{2x_i - (v_2 + v_1)}{(v_2 - v_1)}. \quad (5.4)$$

Substituting (5.4) in (5.2), we can now calculate the same Hamiltonian with respect to the new “spin” variable \mathbf{x} .

5.2 Optimization algorithm: gauge transformation

We present here the main features of the heuristic algorithm used to find the minimum of $h(\mathbf{s})$. Computing this minimum is an NP-hard problem, i.e. the time of execution scales more than any polynomial function of the size of the problem. More details about our algorithm are available in [98].

Given the $(n \times n)$ adjacency matrix J (with $J_{ij} \in \{-1, 0, +1\}$), our approach to minimize the Hamiltonian of an Ising spin glass, i.e. to solve

$$\delta = \min_{\mathbf{s} \in \{-1, +1\}^n} h(\mathbf{s}) = \min_{\mathbf{s} \in \{-1, +1\}^n} m - \frac{1}{2} \mathbf{s}^T J \mathbf{s},$$

consists in finding a diagonal matrix $T_\sigma = \text{diag}(\sigma)$, $\sigma \in \{-1, +1\}^n$, such that $T_\sigma J T_\sigma$ has the least possible number of negative signs. Any such T_σ is a change of sign through a cut set of the graph of J , see example in Fig. 5.2. These operations are called switching equivalences in the theory of signed graphs [99], or gauge transformations in the theory of frustrated spin systems [100], and correspond to changes in the partial order of the orthants of \mathbb{R}^n in the theory of monotonic systems [101, 102].

The outcome of the algorithm is a gauge transformation of the adjacency matrix J into the equivalent J_σ :

$$J \rightarrow J_\sigma = T_\sigma J T_\sigma. \quad (5.5)$$

A system is exactly balanced if and only if there exists a σ such that $J_\sigma = T_\sigma J T_\sigma$ has no negative entries. Unlike the usual algorithms to find the ground states in spin systems, which explore the space of spin configurations \mathbf{s} , our optimization procedure works on the adjacency matrix J . Since $\mathbf{s} = T_\sigma \mathbf{1}$, in terms of the energy function (5.2) the two concepts are equivalent and (5.2) can be written as

$$h(\mathbf{s}) = m - \frac{1}{2} \mathbf{s}^T J \mathbf{s} = m - \frac{1}{2} \mathbf{1}^T T_\sigma J T_\sigma \mathbf{1}. \quad (5.6)$$

For $\mathbf{s} = \sigma$, from (5.6) it is easily seen that flipping spins is equivalent to applying the gauge transformation (5.5) to J . In spin glass theory [103] this corresponds to the notion that frustration δ is an invariant of transformations such as (5.5).

What makes the algorithm efficient is the observation that at each step identifying a gauge transformation which decreases $h(\mathbf{s})$ is straightforward, as it amounts to computing the rows/columns of the current $T_\sigma J T_\sigma$ having negative sum (i.e. the nodes of the graph of $T_\sigma J T_\sigma$ being adjacent to more negative than positive edges). Iterating the procedure (5.5) r times for signature vectors $\sigma_1, \dots, \sigma_r$ then means nesting r matrix multiplications:

$$J \rightarrow T_{\sigma_r} \dots T_{\sigma_1} J T_{\sigma_1} \dots T_{\sigma_r}. \quad (5.7)$$

Provided $\sigma_1, \dots, \sigma_r$ are chosen appropriately, the energy in (5.2) is monotonically decreasing. As the T_{σ_i} are all diagonal matrices, it is always possible to find a $\sigma \in \{-1, +1\}^n$ such that $T_\sigma = T_{\sigma_1} \dots T_{\sigma_r}$. Computationally, the fact that gauge transformations which decrease the energy are cheaply computable a priori (i.e. before each move) implies that only favorable directions are probed at each step. This approach (similar to the BKL algorithm [104] but of lower complexity) drastically decreases the computational time of a search, especially at low energies where favorable moves become rare. As long as $J_\sigma = T_\sigma J T_\sigma$ has rows/columns of negative sum, then the procedure (5.7) can be iterated for another step. The local

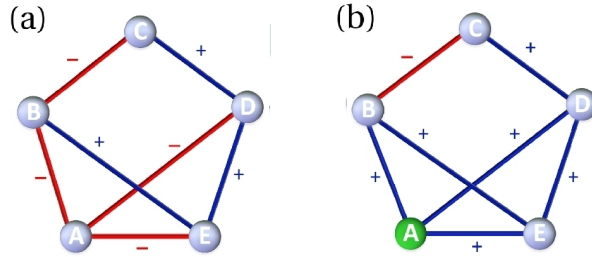


Figure 5.2: **Example of gauge transformation.** (a): The given initial graph with all spin state equal to +1. (b): Applying a sign change to all edges adjacent to the node A (in green) only one negative edge is left in the graph. Since we flip spin A, this optimum corresponds to choosing $s_A = -1$, $s_B = s_C = s_D = s_E = +1$. Hamiltonian (5.2) is then equal to 1 (only one negative edge is left).

search always stops in a minimum, corresponding to a J_σ having all non-negative row/columns sums. This minimum can be local or global. Perturbing locally the system (i.e. applying a gauge transformation which slightly increases the energy) allows to move away from the a local minimum and to continue the search. Millions of such perturbations (of various sizes) can be performed on a single run. After the run is terminated, the algorithm is randomly re-initialized and the whole procedure repeated. In Statistical Physics terminology, different runs of the procedure correspond to different *replicas*. A certain number of replicas reaching the low energy are stored for further processing. The best minimum over all replicas we computed is denoted δ_{up} (i.e. an upper bound of the true ground state δ).

Lower bounds on the ground state δ are available for exact branch-and-cut methods [28, 29] (which, however, can only deal efficiently with small-medium size instances) and for semidefinite programming approaches [105, 106, 107], for which the well-known Goemans–Williamson lower bound guarantee corresponds to $\delta_{low}/\delta_{up} \geq 0.8785$ [108]. An advantage of minimizing the negative entries of J_σ is that no matter if the procedure stops in a local or global minimum, the residual frustrations are straightforward to localize, as one of the optimal spin assignments (ground states) of the gauge transformed J_σ is always the state $\mathbf{1}$, i.e. the all spins up state (and by global symmetry, the all spins down state $-\mathbf{1}$). Furthermore, on an adjacency matrix composed largely by positive edges such as J_σ , it gets easier to try to associate to each remaining negative edge a cycle completed only by means of positive edges. As long as such frustrated cycles are built edge-disjoint, then the frustration they carry can be shown to be ineliminable [98]. This allows to compute also a lower bound δ_{low} for δ for each J_σ . In the concrete examples discussed in this thesis our ratio δ_{low}/δ_{up} always outperforms the aforementioned Goemans–Williamson ratio.

Once we have found the “optimal” gauge transformation T_σ (for which we recall that $\mathbf{1}$ is always a ground state), since $\mathbf{s} = T_\sigma \mathbf{1}$, reconstructing the optimal spin flip \mathbf{s}_0 in the original “quenching” J means “rewinding” all the gauge transformations performed. For (5.7) this amounts to write $\mathbf{s}_0 = T_{\sigma_1} \dots T_{\sigma_r} \mathbf{1}$.

5.3 From spin glass to lossy data compression problem

There is a consistent literature on the connections between the Statistical Physics of disordered systems and information theory, in particular between spin glass theory and Shannon coding theory. A typical approach consists in using some of the Statistical Physics tools developed for spin glasses in order to design new algorithms for a coding system or to analyze its properties [109, 110, 111]. The coding system could be for example a capacity-approaching lossy data compressor, such as those based on Low Density Parity Checks and affine methods [109, 110, 112].

In what follows, we show how to map an Ising spin-glass problem into a lossy data compression problem.

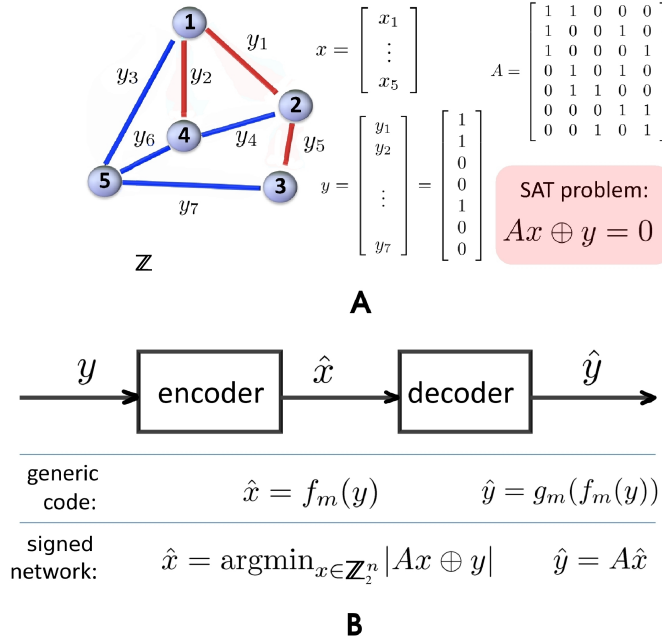


Figure 5.3: **Signed networks and coding systems.** (A): a toy example of signed graph and its formulation as a SAT problem. A negative (red) edge is mapped into $y_i = 1$, while a positive edge (blue) is mapped into $y_i = 0$. (B): rate-distortion scheme. In the regime $n < m$, the encoding/decoding scheme is normally referred to as a *lossy source compression problem* [113], as a length- n sequence $\hat{\mathbf{x}}$ is used to represent a length- m word \mathbf{y} . The distortion corresponds to the relative Hamming distance $d(\mathbf{y}, \hat{\mathbf{y}})/m$, where $\hat{\mathbf{y}} = A\hat{\mathbf{x}}$.

From the adjacent matrix J of a spin glass (5.1), we can first compute an $m \times n$ matrix A which contains all the information about the topology of the network (n the number of spins, m the number of edges). Each row of A identifies an edge, and has two nonzero entries (equal to 1) in correspondence of the two nodes linked by the edge, see Fig. 5.3A for an example.

The sign of the edges of the network can be represented by a vector $\mathbf{y} \in \mathbb{Z}_2^m$ as follows

$$y_k = \frac{1 - J_{ij}}{2} = \begin{cases} 0 & \text{if } J_{ij} = +1 \\ 1 & \text{if } J_{ij} = -1 \end{cases}, \quad \forall i > j \text{ such that } J_{ij} \neq 0, \quad (5.8)$$

for the triplets of indexes $\{k, (i, j)\}$ resulting from the $\text{vec}(\cdot)$ operation.

Similarly, after choosing $(v_1, v_2) = (0, 1)$, Equation (5.3) can be used to convert the vector of spins \mathbf{s} into the binary vector \mathbf{x} :

$$x_i = \frac{1 - s_i}{2} = \begin{cases} 0 & \text{if } s_i = +1 \\ 1 & \text{if } s_i = -1 \end{cases}, \quad i = 1, \dots, n. \quad (5.9)$$

According to the Ising glass interpretation, we need to search for the configuration of spins which has the minimum number of frustrations (i.e. the ground state). We can now translate the Hamiltonian minimization (a sum of m interactions, i.e. of m terms) into a constraint satisfaction (SAT) problem. In particular, searching for the assignments $\mathbf{x} \in \mathbb{Z}_2^n$ compatible with all the values of the given $\mathbf{y} \in \mathbb{Z}_2^m$ means solving a linear system of m equations over the field $(\mathbb{Z}_2, \oplus, \cdot)$

$$A\mathbf{x} \oplus \mathbf{y} = \mathbf{0}. \quad (5.10)$$

For example, the contribution to the Hamiltonian of two interacting spins (s_i and s_j) through the coupling J_{ij} is the term $(1 - J_{ij}s_i s_j)/2$ which is equal to 0 (energetically favored) or equal to +1 (energetically unfavored) according to number of negative signs over the three factors s_i , s_j and J_{ij} . In the same way, each row of (5.10) consists of a summation $x_i \oplus x_j \oplus y_k$ which is equal to +1 (constraint unsatisfied) according to the number of null values of the three terms x_i , x_j and y_k . Indeed (5.10) is known as “XORSAT problem” and it may or may not have a solution depending on \mathbf{y} . When (5.10) has no solution (it is said UNSAT in this case), then one can look for the $\mathbf{x} \in \mathbb{Z}_2^n$ that satisfies the maximal number of constraints of (5.10). This problem is called MAX-XORSAT, see [114], and is equivalent to the Hamiltonian minimization. In fact, denoting $d(\mathbf{y}, A\mathbf{x})$ the Hamming distance between \mathbf{y} and $A\mathbf{x}$, the following equalities hold:

$$h(\mathbf{s}) = \sum_{(i,j)} \frac{1 - J_{ij}s_i s_j}{2} = |A\mathbf{x} \oplus \mathbf{y}| = d(\mathbf{y}, A\mathbf{x}).$$

Then solving the MAX-XORSAT for a given \mathbf{y} means finding $\hat{\mathbf{x}}$ that minimizes $d(\mathbf{y}, A\mathbf{x})$:

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x} \in \mathbb{Z}_2^n} d(\mathbf{y}, A\mathbf{x}) = \operatorname{argmin}_{\mathbf{x} \in \mathbb{Z}_2^n} |A\mathbf{x} \oplus \mathbf{y}|. \quad (5.11)$$

In information theory, MAX-XORSAT problems are used to solve lossy source compression problems [106, 107]. In this context, \mathbf{y} is the original message, $\hat{\mathbf{x}}$ is the encoded (compressed) message and A defines the coding system (see Fig. 5.3B). The relative Hamming distance

$$D_{A,\mathbf{y}} = \frac{1}{m} \min_{\mathbf{x} \in \mathbb{Z}_2^n} d(\mathbf{y}, A\mathbf{x}) \quad (5.12)$$

is called the *distortion of the “word” \mathbf{y} associated to the “code” A* . We shall call *average distortion of the “code” associated to A* (denoted by D_A) the expectation of (5.12) over the entire alphabet \mathbb{Z}_2^m of \mathbf{y} . Assuming that the bits Y_i , $i = 1, \dots, m$, of the encoded message are independent random variables, each drawn from the same given probability distribution, $\operatorname{Prob}(Y_i = y) = p(y)$, then:

$$D_A = \mathbb{E}_{\mathbf{Y}}[D_{A,\mathbf{Y}}] = \frac{1}{m} \sum_{\mathbf{y} \in \mathbb{Z}_2^m} p(\mathbf{y}) d(\mathbf{y}, A\hat{\mathbf{x}}), \quad (5.13)$$

where, for each \mathbf{y} , $\hat{\mathbf{x}}$ solves (5.11).

5.4 Rate-distortion theorem

We now introduce a Shannon-type bound that allows to estimate the value of the distortion in correspondence of the “best” channel topology and of source sequences drawn from a probability distribution model. We start from the definition of distortion $D_{A,\mathbf{y}}$ of a word \mathbf{y} associated to the code A reported in (5.12) and we assume to have the probability distribution $\mathbb{P}(\mathbf{Y})$. The Shannon bound is given by the so-called rate-distortion theorem [113].

Denote $R = n/m$ the *rate* between the number of bits of the encoder (n) and that of the original source (m), $R < 1$. Consider an encoding function $f_m : \mathbb{Z}_2^m \rightarrow \{1, 2, \dots, 2^{mR}\}$ and a decoding function $g_m : \{1, 2, \dots, 2^{mR}\} \rightarrow \mathbb{Z}_2^m$, so that $\hat{y} = g_m(f_m(y))$. As in (5.13), the distortion associated with the code (i.e. to the pair $[f_m(\cdot), g_m(\cdot)]$) is the expectation with respect to the probability distribution $\mathbb{P}(\mathbf{Y})$:

$$D_{(f_m, g_m)} = \mathbb{E} \left[\frac{d(Y, \hat{Y})}{m} \right] = \sum_{\mathbf{y}} p(\mathbf{y}) \frac{d(\mathbf{y}, g_m(f_m(\mathbf{y})))}{m}. \quad (5.14)$$

Given R , as we vary the channel (i.e. $[f_m(\cdot), g_m(\cdot)]$) the pairs (R, D) deduced from (5.14) vary. In particular, a rate distortion pair (R, D) is said achievable if there exists a sequence of $[f_m(\cdot), g_m(\cdot)]$ for which

$$\lim_{m \rightarrow \infty} \mathbb{E} \left[\frac{d(Y, g_m(f_m(Y)))}{m} \right] \leq D.$$

The rate-distortion region C is the closure of the set of achievable rate-distortion pairs (R, D) . In particular we are interested in the *rate-distortion function* $R(D) = \min_{(R, D) \in C} R$, which is the minimum of rates R such that $(R, D) \in C$ for a given D , and in the *distortion-rate function* $D(R) = \min_{(R, D) \in C} D$ which is defined analogously, for a given R .

When the source has a Bernoulli distribution with probability q , i.e. $Y_i \sim \mathcal{B}(q)$, $i = 1, \dots, m$, then $R(D)$ can be calculated explicitly [113].

Theorem 2. *When the source is a Bernoulli distribution $\mathcal{B}(q)$, the rate-distortion function associated with the Hamming distortion (5.14) is*

$$R(D) = \begin{cases} H(q) - H(D) & 0 \leq D \leq \min(q, 1 - q) \\ 0 & D > \min(q, 1 - q) \end{cases} \quad (5.15)$$

where $H(D)$ is the binary entropy, $H(D) = -D \log_2 D - (1 - D) \log_2 (1 - D)$.

The expression (5.15) obviously identifies also $D(R)$ (see also Figure 5.4). For a given R , the value $D(R)$ is achieved by a “best” channel, not necessarily by all channels. In the case of a nonoptimal channel, the distortion for a $\mathcal{B}(q)$ source

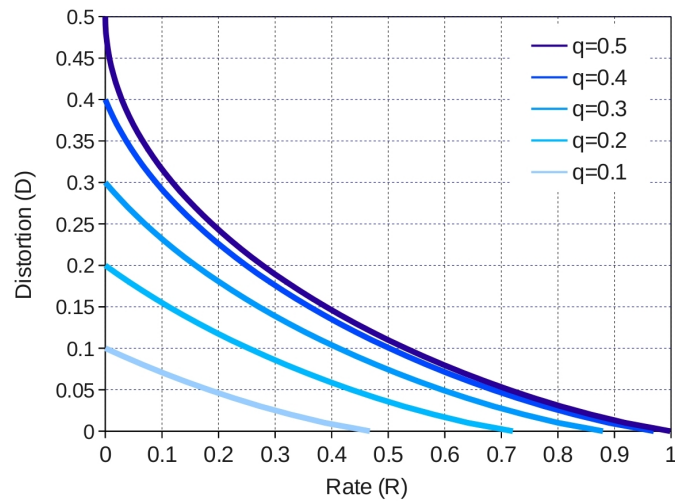


Figure 5.4: **Rate-distortion theorem.** Example of the rate-distortion function $R(D)$ from a Bernoulli distribution with parameter q from 0.1 to 0.5 (for symmetry reasons, the outcome is the same for q and $1 - q$).

(computed on this channel) must be strictly bigger than $D(R)$ by construction. Likewise, when applied to a specific source sequence, $D(R)$ is achieved only if the sequence is “typical” for the source probability distribution, while an odd, unusual sequence may instead deviate significantly from $D(R)$.

Chapter 6

A rate-distortion theory for gene regulatory networks

6.1 Gene regulatory networks

A gene regulatory network consists of a set of Transcription Factors (TF) regulating the expression of the genes of a given genome. As our knowledge of these regulatory mechanisms grows steadily, the problem of understanding the principles behind their organization and their functioning is becoming a crucial question of Systems Biology [115, 116, 117, 118, 119]. While the current literature is focused mainly on the topological aspects of a gene regulatory network, the perspective that we take in this work is to study whether the transcriptional regulatory modes (i.e. the activation/inhibition role of a TF on its target genes) of the gene networks currently available are distributed randomly on the given network or are organized according to some criteria.

To formalize this question and to try to understand the rules of this design, we use tools from information theory [113, 114] and, in particular, we treat a gene network as a “code” for which the signs of the regulatory actions constitute a particular “source word”, which can be compared to the typical words generated by a corresponding probabilistic model. As a comparison criterion we use the “level of coherence” of the regulatory actions along the network. Two regulatory orders emanating from a TF and acting on the same target gene (possibly through intermediate genes) are considered coherent when they induce the same behavior on the target gene (i.e. they both induce activation or repression); they are considered incoherent when they induce conflicting behaviors. As can be easily deduced from simple examples like FeedForward Loops (FFL, see Fig. 6.2), in a gene regulatory network incoherence is associated with negative (undirected) cycles on the signed graph having the genes as nodes, the regulations as edges, and the modes of the regulations as signs of the edges, see [98]. An undirected cycle is negative when it contains an odd number

6.2 Materials and Methods

6.2.1 Analyzed networks

Although large-scale gene regulatory networks are becoming available for several organisms, only for very few of them the signs of the regulations are available to date (see e.g. [121, 122, 123]). In this work we use the following three gene networks:

- ⊗ gene regulatory network of *Escherichia coli*, downloaded from *RegulonDB* database (<http://regulondb.ccg.unam.mx>). See [124].
- ⊗ gene regulatory network of *Saccharomyces cerevisiae*, discussed in [125] and downloadable from <http://www.weizmann.ac.il/mcb/UriAlon/>.
- ⊗ gene regulatory network for *Bacillus subtilis*, assembled by [126]. Downloaded from <http://dbtbs.hgc.jp/>.

Details like the number of genes n and regulations m for these three networks are in Table 6.1. A directed edge represents the action of a transcription factor on one of its target genes, and the sign means activation (+) or inhibition (-). As always in the gene network literature, these post-transcriptional regulations (the TF influencing the transcription of other genes are gene products) are represented at transcriptional level. Only a very few TFs are cataloged as acting both as activator and as repressor on the same gene. These regulations are disregarded in our study. Also the autoregulatory actions (i.e. self-loops of the signed graphs), which are abundant especially in the *E.coli* network, are not considered in the work.

In the second part of Table 6.1, the three gene networks are restricted to their largest bicomponent, i.e. to the subnetwork whose nodes are involved in undirected cycles. In this way, the leaves (and tree-like structures) of the undirected graph are dropped.

Table 6.1: **Main features of the analyzed gene regulatory networks.** n and m are the number of nodes and edges of the directed graph representing the gene network; rate R is the ratio nodes/edges; q is the fraction of negative edges.

| Network | n | m | R | q |
|---------------------|------|------|-------|-------|
| <i>E.coli</i> | 1461 | 3220 | 0.454 | 0.416 |
| <i>S.cerevisiae</i> | 690 | 1082 | 0.638 | 0.204 |
| <i>B.subtilis</i> | 918 | 1324 | 0.693 | 0.256 |

QUANTITIES RESTRICTED TO THE BICOMPONENT SUBNETWORKS.

| Network | n | m | R | q |
|---------------------|-----|------|-------|-------|
| <i>E.coli</i> | 892 | 2671 | 0.334 | 0.404 |
| <i>S.cerevisiae</i> | 336 | 736 | 0.457 | 0.234 |
| <i>B.subtilis</i> | 388 | 803 | 0.483 | 0.277 |

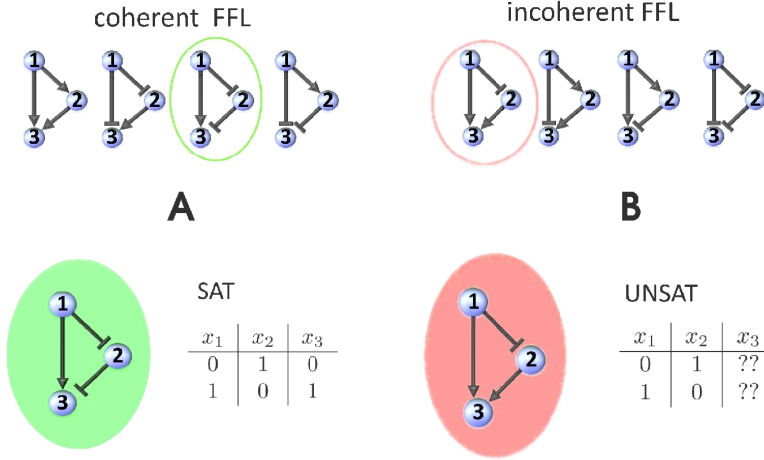


Figure 6.2: **FFL motifs and their steady state values.** The FFL is the simplest motif forming an undirected cycle in a graph. Of the 8 possible FFL, 4 are coherent (**A**) and 4 incoherent (**B**).

6.2.2 Gene regulatory networks as constraint satisfaction problems

In this section we are interested in formulating a description of the signed graph representing the gene regulatory network in terms of encoding/decoding systems. Consider a gene regulatory network composed of n nodes $\mathbf{x} = [x_1 \ \dots \ x_n]^T$ representing the genes and m directed edges $\mathbf{y} = [y_1 \ \dots \ y_m]^T$ representing regulatory actions of activation/inhibition of a gene on another gene. Assume both \mathbf{x} and \mathbf{y} are represented in Boolean terms: $x_i \in \mathbb{Z}_2 = \{0, 1\} = \{\text{“low”}, \text{“high”}\}$, and $y_i \in \mathbb{Z}_2 = \{0, 1\}$, where we use the convention that 0 stands for activation (i.e. “+”) and 1 for inhibition (i.e. “-”).

Let us make use of an $m \times n$ matrix A to describe the topology of the gene network. According to (5.10), finding gene expression assignments $\mathbf{x} \in \mathbb{Z}_2^n$ compatible with all the regulatory signs of a given $\mathbf{y} \in \mathbb{Z}_2^m$ means solving a linear system over the $(\mathbb{Z}_2, \oplus, \cdot)$ field:

$$A\mathbf{x} \oplus \mathbf{y} = \mathbf{0}. \quad (6.1)$$

The distortion of \mathbf{y} associated to the A is

$$D_{A,\mathbf{y}} = \frac{1}{m} \min_{\mathbf{x} \in \mathbb{Z}_2^n} d(\mathbf{y}, A\mathbf{x}), \quad (6.2)$$

where $d(\cdot, \cdot)$ denotes the Hamming distance.

Example: Feed-Forward Loops (FFL)

Let us consider as an example the FFL of Fig. 6.2. In a FFL, $n = m = 3$, hence

Table 6.2: **Coherent and incoherent FFL.** Counts of the FFL in the three transcriptional networks of Table 6.1. For *E.coli* and *B.subtilis* the FFL counts do not take into account the operonal structure of the DNA. When genes sharing an operon are lumped together, the fraction coherent/incoherent FFL is very similar. Notice that in many of these FFL, the intermediate node (i.e. x_2 in Fig. 6.2) has in-degree > 1 , meaning that the FFL are coupled. Other forms of coupling are present. For example many FFL share an edge. In *E.coli* this can be deduced also from the fact that the number of incoherent FFL is larger than the number of independent undirected negative cycles, see [123].

| Network | coherent FFL | incoherent FFL |
|---------------------|--------------|----------------|
| <i>E.coli</i> | 539 | 724 |
| <i>S.cerevisiae</i> | 39 | 34 |
| <i>B.subtilis</i> | 149 | 97 |

both alphabets \mathbb{Z}_2^n and \mathbb{Z}_2^m consist of 8 words and we may ask if for each choice of $\mathbf{y} \in \mathbb{Z}_2^3$ there is always an $\mathbf{x} \in \mathbb{Z}_2^3$ such that (5.10) is satisfied. In this case the connectivity matrix A is the following

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad (6.3)$$

and it is straightforward to verify that only in 4 of the 8 choices of \mathbf{y} Equation (6.1) is SAT. These cases (Fig. 6.2A) are known in the literature as *coherent* FFL while the 4 UNSAT cases (Fig. 6.2B) are called *incoherent* in [117]. In a coherent FFL $d(\mathbf{y}, A\hat{\mathbf{x}}) = 0$, while in an incoherent FFL $d(\mathbf{y}, A\hat{\mathbf{x}}) = 1$; therefore we have $D_{A,\mathbf{y}} = 1/3$. See Table 6.2 for a count of the FFL motifs in the gene networks of Table 6.1.

6.2.3 Null model of a gene regulatory network of given topology

Consider a gene regulatory network with n genes, m edges and a given topology. Let us denote Y_i a random variable associated to the sign of the regulatory actions between two genes, $i = 1, \dots, m$. Y_i takes values in the alphabet $\{0, 1\}$ (0 corresponds to “+” and 1 to “-”). If we know that $q = m^-/m$ is the fraction of negative signs in the true network, then to form a null model we can think of each edge sign as being drawn independently from a Bernoulli distribution with probability q for the symbol “1” (i.e. negative regulation), denoted $\mathcal{B}(q)$. $\mathbf{Y} = [Y_1 \dots Y_m]^T$, $Y_1, \dots, Y_m \in \mathcal{B}(q)$, is therefore the random variable for the gene regulatory network with the given topology. Since the Y_i are independent and identically distributed variables (i.i.d.), then each sample vector $\mathbf{y} = [y_1 \dots y_m]^T$ has a probability that factorizes: $p(\mathbf{Y} = \mathbf{y}) = \prod_{i=1}^m p(Y_i = y_i) = q^{|y|}(1-q)^{m-|y|}$.

6.3 Results and Discussion

Distortion of the analyzed gene networks

In the following we denote $D_{A,\mathbf{y}}^{\text{emp}}$ the distortion in correspondence of the true (“empirical”) edge sign assignments \mathbf{y} of a given network. The values reported in Table 6.3 are computed using (6.2). The upper and lower bounds computed for $D_{A,\mathbf{y}}^{\text{emp}}$ are fairly tight.

Needless to say, it is impossible to evaluate exactly quantities such as (5.13) which are computed exhaustively over $\mathbf{y} \in \mathbb{Z}_2^m$. To estimate D_A , we can repeat the same optimization as in (6.2) on a sufficiently large number of null models as described in Section 6.2.3. As can be seen in Table 6.3, $D_{A,\mathbf{y}}^{\text{emp}} < D_A$ for all three gene networks, meaning that the true gene networks have less distortion than the corresponding null models (coherently with the results reported in [98, 123]). Notice that only nodes involved in undirected cycles contribute to the distortion. In particular, then, intending the networks as undirected graphs and restricting to the bicomponents (i.e. dropping the nodes/edges not involved in undirected cycles) means changing the values of n and m , and hence of $D_{A,\mathbf{y}}^{\text{emp}}$ and of D_A , see second part of Table 6.3.

Both $D_{A,\mathbf{y}}^{\text{emp}}$ and D_A are obtained in correspondence of the given topology, described by the connectivity matrix A . If we allow also the topology to vary, then we can use the rate-distortion theorem of information theory [113] to determine the admissible region for the distortion D for all possible compression rates $R = n/m$ in correspondence of a $\mathcal{B}(q)$ source of edge sign words. As explained in the previous chapter, the boundary of such an admissible region represents a Shannon-type bound, and it is achieved in correspondence of a “best” network topology. For Bernoulli sources, the distortion D on such bound can be computed explicitly, see (5.15). In Fig. 6.3, the Shannon bound D is compared with $D_{A,\mathbf{y}}^{\text{emp}}$ and D_A for the three gene networks of Table 6.1. As can be observed, the values of $D_{A,\mathbf{y}}^{\text{emp}}$ are close to the corresponding D . In particular, once we restrict to the bicomponents (right panels in Fig. 6.3), $D_{A,\mathbf{y}}^{\text{emp}} < D$ in two of the three networks. On the contrary

Table 6.3: **Distortion of analyzed networks.** Lower and upper bounds of $D_{A,\mathbf{y}}^{\text{emp}}$ and D_A (distortion of the true edge sign assignment and of the null models, respectively).

| Network | $D_{A,\mathbf{y}}^{\text{emp}}$ | D_A |
|---------------------|---------------------------------|------------------|
| <i>E.coli</i> | [0.1134, 0.1152] | [0.1767, 0.2043] |
| <i>S.cerevisiae</i> | [0.0379, 0.0379] | [0.1077, 0.1091] |
| <i>B.subtilis</i> | [0.0536, 0.0536] | [0.1040, 0.1043] |

QUANTITIES RESTRICTED TO THE BICOMPONENT SUBNETWORKS

| | | |
|---------------------|------------------|------------------|
| <i>E.coli</i> | [0.1367, 0.1389] | [0.2130, 0.2462] |
| <i>S.cerevisiae</i> | [0.0557, 0.0557] | [0.1583, 0.1605] |
| <i>B.subtilis</i> | [0.0883, 0.0883] | [0.1713, 0.1718] |

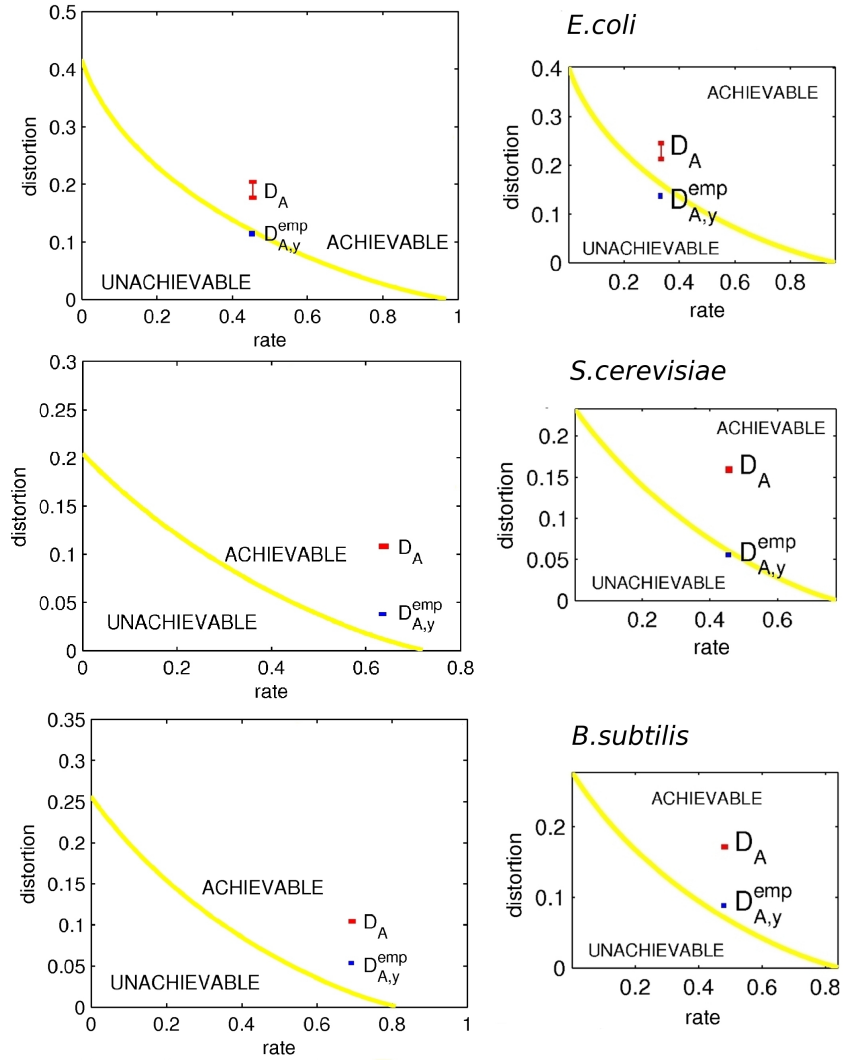


Figure 6.3: **Distortion and Shannon bounds.** *Left panels:* For the three gene networks of Table 6.1 the distortion in correspondence of the “true” edge signs \mathbf{y} ($D_{A,\mathbf{y}}^{\text{emp}}$) and in correspondence of the null models (D_A) is compared with the Shannon bound given by the rate-distortion theorem (yellow curve). Upper and lower bounds on $D_{A,\mathbf{y}}^{\text{emp}}$ and on D_A are normally very close, see Table 6.3 (one exception is D_A for *E.coli*). *Right panels:* the same quantities are shown for the bicomponent subnetworks (where nodes not involved in undirected cycles are dropped,). In this last case, $D_{A,\mathbf{y}}^{\text{emp}}$ is below the Shannon bound in 2 of the 3 networks (and very close in the third one). In all cases $D_{A,\mathbf{y}}^{\text{emp}} < D_A$ i.e. the distortion is atypical for a $\mathcal{B}(q)$ source of words (i.e. of edge sign assignments).

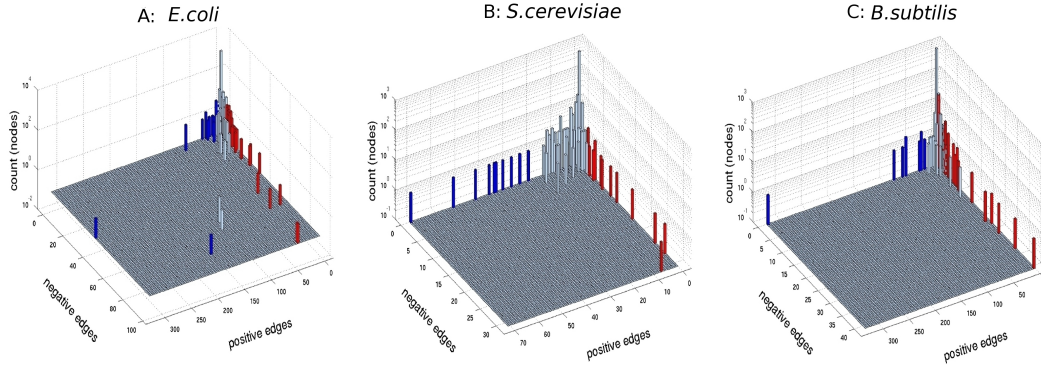


Figure 6.4: **Single-mode action of the TF.** The histograms show the number of positive and negative edges in the out-degree of the TF for the three networks. The histograms are highly skewed, meaning that the majority of TF have a single mode of action. In particular, the TF significantly single-mode (with respect to a cumulative binomial test, p -value 10^{-2}) are highlighted in color: blue for activators, red for repressors. See also Tables 6.4–6.6 for a list of the corresponding genes. For *E. coli*, the few TF having both positive and negative edges are well-known dual-mode regulators, such as *crp*, *fnr*, *ihf*, *fis*, *arcA* and *narL* (see Table 6.4).

$D_A \gg D$ in all three cases. The meaning is that in spite of non-optimal topologies, the distortions of our gene regulatory networks (which, from $D_{A,y}^{\text{emp}} < D_A$, we know to be much lower than expected) are also at the level expected for a “best” network topology.

In terms of Ising spin glasses, the distortion $D_{A,y}^{\text{emp}}$ has the interpretation of “frustration” encoded in the undirected cycles, i.e. of linearly independent undirected cycles having negative sign (meaning an odd number of inhibitions). Our result therefore implies that frustration is largely absent in these signed graphs. Notice that direct counts of the basic frustrated/non-frustrated motifs such as the incoherent/coherent FFL are largely inconclusive, see Table 6.2. The true distortion can be computed only genome-wide, and its calculation confirms that indeed conflictual orders are largely avoided.

Low distortion and single-mode TF

It is worth mentioning that the origin of the low distortion of the gene networks lies in the highly skewed distribution of the signs of the actions of the TF. As can be seen in Fig. 6.4, the vast majority of the TF tends to operate in a single-mode fashion on all their target genes. Dual-mode TF are statistically rare with respect to the null models. See Tables 6.4–6.6 for more details. While this skewness is expected, since the physical interaction mechanisms of an activator and of an inhibitor are normally different, its consequences for the regulation on a genome-wide scale have rarely been assessed, except on small motifs like FFL. Following the arguments of [127, 123], and in particular the notion of gauge equivalence, it can be shown that such a pattern is responsible for the limited amount of distortion of these networks.

Table 6.4: **TF with dual mode of action and TF enriched for positive or negative edges in *E.coli*.** The left part of the table shows all TF acting both as activator and repressor. Dual mode TF marked with an asterisk show a dominating action (i.e. they are enriched for positive or negative edges with respect to a cumulative binomial test, p -value 10^{-2}). The central and right part of the table list TF acting as activator or repressor on different target genes (enriched for positive or negative edges, same cumulative binomial test). In Fig. 6.4A TF enriched for positive edges are in blue and those enriched for negative edges in red.

| DUAL MODE | + | - | POSITIVE | + | - | NEGATIVE | + | - |
|------------------|-----|----|-----------------|-----|----|-----------------|----|----|
| torR | 9 | 2 | fhlA | 29 | 0 | ydeW | 0 | 8 |
| himD | 124 | 63 | rob | 12 | 0 | metJ | 0 | 13 |
| yabN | 2 | 3 | crp | 320 | 44 | ybbU | 0 | 10 |
| csgD | 8 | 1 | fnr | 176 | 84 | glpR | 0 | 9 |
| modE | 31 | 15 | appY | 9 | 0 | nagC | 1 | 15 |
| cysB | 17 | 6 | phoP | 23 | 5 | yjfQ | 0 | 7 |
| nagC* | 1 | 15 | evgA | 12 | 0 | hupA | 0 | 8 |
| marA | 19 | 6 | flhC | 75 | 4 | trpR | 0 | 11 |
| rcsB | 16 | 6 | malT | 10 | 0 | exuR | 0 | 7 |
| arcA* | 41 | 98 | glnG | 38 | 0 | arcA | 41 | 98 |
| narL* | 34 | 66 | yhiE | 30 | 0 | narL | 34 | 66 |
| tyrR | 2 | 8 | hyfR | 11 | 0 | pdhR | 5 | 13 |
| leuO | 10 | 1 | cbl | 9 | 0 | lexA | 0 | 39 |
| pdhR* | 5 | 13 | oxyR | 16 | 1 | yjeB | 0 | 16 |
| idnR | 4 | 1 | yhiX | 19 | 0 | yiaJ | 0 | 9 |
| rcsA | 15 | 6 | caiF | 10 | 0 | fruR | 9 | 24 |
| ompR | 6 | 6 | soxS | 25 | 0 | lrp | 29 | 52 |
| crp* | 320 | 44 | yaeG | 9 | 0 | purR | 0 | 30 |
| fruRt* | 9 | 24 | flhD | 75 | 4 | gatR | 0 | 6 |
| fnr* | 176 | 84 | phoB | 29 | 2 | mlc | 0 | 9 |
| lrp* | 29 | 52 | | | | gntR | 0 | 11 |
| phoP* | 23 | 5 | | | | uxuR | 0 | 6 |
| fis | 131 | 69 | | | | ycdC | 2 | 10 |
| asnC | 1 | 2 | | | | paaX | 0 | 12 |
| flhC* | 75 | 4 | | | | fur | 13 | 66 |
| nac | 6 | 8 | | | | agaR | 0 | 10 |
| narP | 17 | 17 | | | | deoR | 0 | 6 |
| dnaA | 3 | 6 | | | | argR | 5 | 31 |
| lrhA | 1 | 2 | | | | ybaD | 0 | 8 |
| ycdC* | 2 | 10 | | | | cytR | 0 | 10 |
| fur* | 13 | 66 | | | | iscR | 7 | 18 |
| fadR | 3 | 8 | | | | hns | 41 | 97 |
| citB | 6 | 1 | | | | gatR | 0 | 6 |
| argR* | 5 | 31 | | | | hupB | 0 | 8 |
| yhiW | 3 | 4 | | | | | | |
| rstA | 5 | 5 | | | | | | |
| oxyR* | 16 | 1 | | | | | | |
| yhcK | 1 | 7 | | | | | | |
| ybbI | 2 | 5 | | | | | | |
| iscR* | 7 | 18 | | | | | | |
| hns* | 41 | 97 | | | | | | |
| flhD* | 75 | 4 | | | | | | |
| himA | 124 | 63 | | | | | | |
| phoB* | 29 | 2 | | | | | | |
| cpxR | 34 | 21 | | | | | | |

Table 6.5: **TF enriched for positive or negative edges in *S.cerevisiae*.** The table lists the TF significantly single-mode (with respect to a cumulative binomial test, p -value 10^{-2}). In Fig. 6.4B the TF enriched for positive edges are shown in blue and those enriched for negative edges in red.

| POSITIVE | + | - | NEGATIVE | + | - |
|-----------------|----|---|-----------------|---|----|
| GCN4 | 53 | 0 | DAL80 | 0 | 17 |
| GLN3 | 29 | 0 | DAL80_GZF3 | 0 | 3 |
| MBP1_SWI6 | 36 | 0 | HDA1 | 0 | 5 |
| MSN2 | 35 | 0 | MIG1 | 0 | 27 |
| MSN4 | 32 | 0 | MIG2 | 0 | 12 |
| OAF1_PIP2 | 22 | 0 | MOT3 | 0 | 3 |
| STE12 | 71 | 0 | RGT1 | 1 | 4 |
| TEC1 | 44 | 0 | ROX1 | 1 | 8 |
| YAP1 | 38 | 0 | SKO1 | 0 | 3 |
| HAP2_3_4_5 | 25 | 0 | TUP1 | 0 | 24 |
| | | | UME6 | 7 | 31 |
| | | | XBP1 | 0 | 5 |
| | | | ARG80_ARG81 | 1 | 5 |
| | | | HDA1_TUP1 | 0 | 7 |
| | | | SSN6_TUP1 | 0 | 10 |

Table 6.6: **TF enriched for positive or negative edges in *B.subtilis*.** The table lists the TF significantly single-mode (with respect to a cumulative binomial test, p -value 10^{-2}). In Fig. 6.4C the TF enriched for positive edges are shown in blue and those enriched for negative edges in red.

| POSITIVE | + | - | NEGATIVE | + | - |
|-----------------|-----|---|-----------------|---|----|
| SigA | 326 | 0 | AbrB | 5 | 27 |
| SigB | 66 | 0 | AraR | 0 | 4 |
| SigD | 30 | 0 | CcpA | 3 | 41 |
| SigE | 83 | 0 | YsiA | 0 | 4 |
| SigF | 30 | 0 | CodY | 0 | 15 |
| SigG | 60 | 0 | CtsR | 0 | 5 |
| SigH | 24 | 0 | LexA | 0 | 33 |
| SigK | 59 | 0 | Fur | 0 | 23 |
| SigW | 33 | 0 | Hpr | 1 | 7 |
| | | | PerR | 1 | 7 |
| | | | PurR | 0 | 9 |
| | | | YvaN | 0 | 4 |
| | | | Rok | 0 | 9 |
| | | | SimR | 0 | 5 |
| | | | SpoIID | 9 | 22 |
| | | | SpoVT | 6 | 10 |
| | | | MtrB | 0 | 4 |
| | | | YdiH | 0 | 4 |
| | | | YlpC | 0 | 5 |
| | | | YrzC | 0 | 8 |

6.4 Conclusions

We have shown that the distortion of the currently available gene networks is much lower than the one of a typical sequence of the same code, and that it is comparable to (when not better than) the Shannon bound. This atypicality implies that in our gene networks the signs of the transcriptional regulations are highly organized and far from random. In particular, the origin of the low distortion can be traced in the scarcity of dual-mode TF, i.e. of TF acting both as activators and as repressors. Our calculation suggests a very practical reason for such an organization: single mode TF lower the distortion and hence help in avoiding conflictual transcriptional orders in which different TF induce contradictory actions on a downstream gene.

The multiplicity of TF regulating each gene implies that the logic with which these multiple TF combine their actions plays an important role on the regulation. Additional considerations on the consistency of the Boolean logics (canalizing such as AND-OR rather than XOR, non-canalizing) at the multi-input nodes, and how they are bound by distortion can be also drawn. Further details can be found in [128].

Chapter 7

Structural balance in large signed social networks

7.1 From structural balance to spin glasses

The method and the approach used in the previous chapter for the characterization of gene networks can be applied to a completely different context, like on-line signed social networks.

On-line social networks are examples of large-scale communities of interacting individuals in which local ties between users (friend, fan, colleague, but also friend/foe, trust/distrust, etc.) give rise to a complex, multidimensional web of aggregated social behavior [129, 130, 131, 132]. For such complex networks, the emergence of global properties from local interactions is an intriguing subject, so far investigated mostly at structural and topological level [133, 125, 134, 135, 130]. In social network theory [136, 137, 138], however, the content of the relationships is often even more important than their topology, and this calls for the development of novel analytical and computational tools, able to extrapolate content-related features out of the set of interactions of a social community. This is particularly challenging when, as in social networks retrieved from on-line media, the size of the community is very big, of the order of 10^5 individuals or higher.

A global property that has recently attracted some attention [139, 140, 129, 141] is determining the structural balance of a signed social network. Structural (or social) balance theory was first formulated by Heider in the 50's [142] in order to understand the structure and origin of tensions and conflicts in a network of individuals whose mutual relationships are characterizable in terms of friendship and hostility. According to the theory proposed by Heider, individuals tend to establish relations that avoid tensions: this can be condensed in the statement "*the enemy of my enemy is my friend*" (and similar, equivalent statements, see e.g. Figure 7.1a).

It was modeled in terms of signed graphs by Cartwright and Harary [143] (see

[138, 137] for an overview of the theory). The nodes of the graph represent users and the positive/negative edges their friendly/hostile relationships. It has been known for some time how to interpret structural balance on such networks [143]: the potential source of tensions are the cycles of the graph (i.e. the closed paths beginning and ending on the same node), notably those of negative sign (i.e. having an odd number of negative edges). It follows that the concept of balance is not related to the actual number of negative edges on the cycles but only to their parity, see Fig. 7.1 for an illustration on basic graphs. In particular, a signed graph is *exactly* balanced (i.e. tensions are completely absent) if and only if all its cycles are positive [143]. As such, structural balance is intrinsically a property of the network as a whole, not fragmentable into elementary subgraphs. Heiders's structural balance theory affirms that human societies tend to avoid tensions and conflictual relations. In a signed graph this translates into a level of balance higher than expected, given the network structure.

The first part of this work is focused on the characterization of some on-line networks through the determination at the global degree of structural balance: partial hints that social networks currently available are more balanced than expected are provided by both the small motif screening of [140, 129] and the spectral analysis of [144]. In particular, on the social networks analyzed in this work we show that the chance that a null model has the level of balance of the true networks is essentially equal to zero. For all three networks, the level of balance turns out to be even less than the Shannon bound one obtains developing a rate-distortion theory for the null models [109, 114, 112]. What makes our signed networks so balanced is the skewed distribution of the signs of the edges on the users: users with a large majority of friends but also users with a large majority of enemies are not causing any significant frustration. In terms of social network theory, this means that individuals manifestly recognized as enemies do not contribute to the structural tension of a community.

A second purpose of this work is to use the large amount of independent replicas obtained during the minimization procedure in order to describe the low-energy landscape of the system, i.e. of the landscape of near optimal balance of the social networks. For example if the system has multiple distinct valleys of nearly identical energy separated by energetic barriers, then it is said to have a replica symmetry breaking picture [145], meaning that ergodicity is broken and different cooling procedures lead to different isoenergetic (or nearly isoenergetic) minima separated by energy barriers. Some other times, instead, it may happen that the landscape is funneled around a single "dominant" valley, possibly surrounded by a plethora of local minima of little significance (and progressively higher energy). The two different pictures are both observed on two social networks considered in this study. Most interestingly, our analysis allows us to obtain insight into the origin of these differences in the landscape.

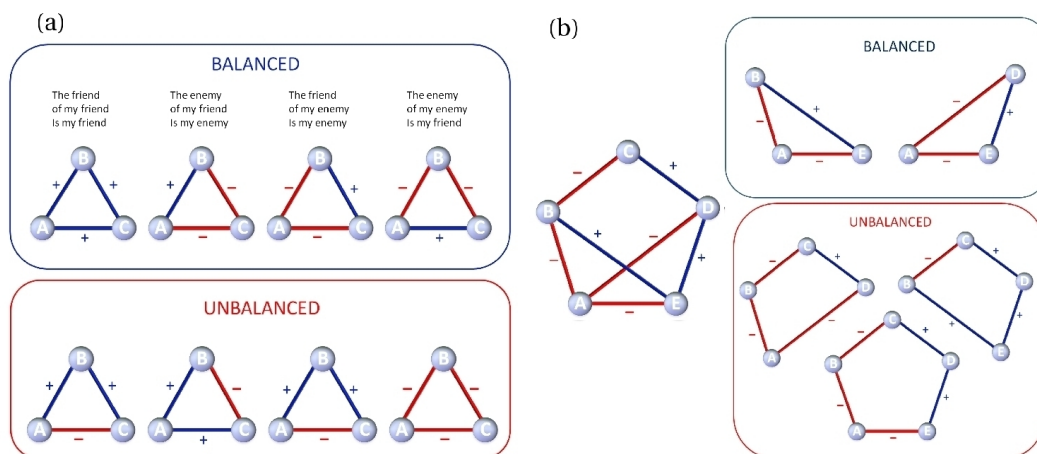


Figure 7.1: **Balanced and unbalanced graphs.** (a): Simplest cases of balance and unbalance: triangles. Users A and C are related directly and indirectly (through B). The sentences on the top connote this indirect relationship between A and C. Blue edges represent friendship, red hostility. The triangles are balanced when the direct and the indirect relationships have the same sign, unbalanced otherwise. (b): For generic graphs, testing all triangles may not give a satisfactory measure of the global balance. In the example the graph is not globally balanced, although all triangles are balanced.

7.2 Materials and Methods

7.2.1 Three on-line large-signed social networks

The three signed social networks analyzed in this work were downloaded from the Stanford Network Analysis Platform (<http://snap.stanford.edu/>) [139]:

- ⊗ *Epinions*: trust/distrust network among users of product review web site Epinions [146, 140];
- ⊗ *Slashdot*: friend/foes network of the technological news site Slashdot (Zoo feature) [147, 139];
- ⊗ *WikiElections*: election of admin among Wikipedia users [148].

The size (n) and number of edges (m) of these networks are given in Table 7.1. The edges of the networks are always considered as undirected. This process leads to only a limited number of sign inconsistencies between pairs of edges J_{ij} and J_{ji} . These inconsistencies are disregarded in our analysis.

7.2.2 Computation of global balance

From a computational point of view, verifying if a signed undirected network is exactly balanced is an easy problem, which can be answered in polynomial time

Table 7.1: **Signed social networks.** Data for the three networks characterized in this work (before and after symmetrization) and for their largest connected component. The networks are originally provided as directed graphs. n and m are the number of nodes and edges of the undirected graph, m^- and m^+ are the number of negative and positive edges of the networks. The compatible/incompatible edge pairs are pairs of directed edges J_{ij} and J_{ji} . They are compatible when $J_{ij} = J_{ji}$ and incompatible when $J_{ij} = -J_{ji}$. The latter are only a tiny fraction, and are disregarded in this study. This implies also that the “symmetrization” of the directed graphs is not changing notably the properties of the networks. Similarly, also the self-loops (i.e. the diagonal entries of the adjacency matrix) J_{ii} are not taken into account. $q = m^-/m$ is the probability of a negative edge and $R = n/m$ is the “rate of compression”. All three networks have a very large connected component, to which we focus our attention (c denotes the total number of connected components). Within a connected component, only nodes in a bicomponent are involved in cycles, and may potentially be unbalanced. The largest bicomponent of the largest connected component contains a significant fraction of the nodes ($n_{1,bicomp}$). $m_{1,bicomp}$ is the number of edges in the bicomponent.

| ORIGINAL NETWORKS | | | | | | |
|----------------------|--------|--------|--------|--------|---------------|----------|
| Network | n_d | m_d | m^- | m^+ | comp./incomp. | diagonal |
| <i>Epinions</i> | 131828 | 841372 | 123705 | 717667 | 254345 / 2703 | 573 |
| <i>Slashdot</i> | 82144 | 549202 | 124130 | 425072 | 93544 / 1949 | 0 |
| <i>WikiElections</i> | 8297 | 103591 | 21927 | 81664 | 5298 / 296 | 58 |

| AFTER SYMMETRIZATION | | | | | | |
|----------------------|--------|--------|--------|--------|-------|-------|
| Network | n | m | m^- | m^+ | q | R |
| <i>Epinions</i> | 131513 | 708507 | 118619 | 589888 | 0.167 | 0.186 |
| <i>Slashdot</i> | 82062 | 498532 | 117599 | 380933 | 0.236 | 0.165 |
| <i>WikiElections</i> | 7114 | 100321 | 21529 | 78792 | 0.214 | 0.071 |

| LARGEST CONNECTED COMPONENT | | | | | | | |
|-----------------------------|--------|--------|---------|---------|----------------|----------------|------|
| Network | n_1 | m_1 | m_1^- | m_1^+ | $n_{1,bicomp}$ | $m_{1,bicomp}$ | c |
| <i>Epinions</i> | 119070 | 701569 | 117759 | 583810 | 58174 | 640080 | 5566 |
| <i>Slashdot</i> | 82052 | 498527 | 117599 | 380928 | 51402 | 467511 | 6 |
| <i>WikiElections</i> | 7065 | 100295 | 21517 | 78778 | 4785 | 98015 | 24 |

[149, 144, 150]. When instead a graph is not exactly balanced, one can compute a *distance to exact balance*, i.e. a measure of the amount of unbalance in the network. The most plausible distance is given by the *least number of edges* that must be dropped (or changed of sign) in order for the graph to become exactly balanced [143, 151, 152]. Computing this distance (called the “line index of balance” in [151, 152]) is an NP-hard problem, equivalent to a series of well-known problems, such as

- computing the ground state of a (nonplanar) Ising spin glass [153];
- solving a MAX-CUT problem [106, 107];
- solving a MAX-2XORSAT problem [114].

The equivalence with energy minimization of a spin glass has for example been highlighted recently in [154] (see also [155, 156]). In this context, a negative cycle is denoted as *frustration*, and frustrations are the trademark of complex energy landscapes, with many local minima whose structure and organization has been so far explored only in special cases. For instance, the case studied in [156], the fully connected graph, is unrealistic for real social networks, which usually have heterogeneous connectivity degrees. As a matter of fact, for what concerns the on-line signed social networks currently available, only an analysis of local, low-dimensional motifs has been carried out so far [139, 140]. This amounts essentially to the enumeration of the triangles and to their classification into frustrated/non-frustrated, see [140, 129]. An alternative approach is taken in [144], where spectral properties of the Laplacian are investigated. For connected signed graphs, the magnitude of the smallest eigenvalue of the Laplacian is indicative of how unbalanced a network is, i.e. of how much frustration is encoded in the cycles of the networks.

Both approaches provide useful information in order to understand the structural balance of signed social networks. Yet, this information is partial and unsatisfactory. The small motif analysis, for example, only identifies the frustration on the smallest possible groups of interacting users, but overlooks more long-range conflicts associated to longer cycles (and larger communities), see Fig. 7.1(b) for an example. The spectral approach, on the contrary, gives an idea of the overall amount of frustration of the network, but does not provide any information on which relationships remain unbalanced. In terms of spin glasses, solving the problem globally and identifying the residual ineliminable tensions means computing the ground state(s) of a heterogeneous Ising spin glass, with bimodal bond distribution. We use the efficient heuristic method presented in Chapter 5 based on the application of equivalence transformations to the signed graph, called *gauge transformations*, to perform this minimization.

Practically, if $J_{ij} \in \{\pm 1\} = \mathbb{B}_2$ is the edge between the nodes s_i and s_j of the graph (call J the $n \times n$ symmetric matrix of entries J_{ij}), computing global balance means assigning a +1 or a -1 to all the nodes so as to minimize the energy functional (5.2). When J_{ij} represents friendship ($J_{ij} = +1$) each term in (5.2) gives a zero contribution if $s_i = s_j$ and a +1 contribution if $s_i = -s_j$, while when J_{ij} represents hostility ($J_{ij} = -1$) the summand is zero if $s_i = -s_j$ and +1 otherwise. The network is exactly balanced when there exists $\mathbf{s} \in \mathbb{B}_2^n$ such that all terms in (5.2) can be made simultaneously equal to zero. Computing the global balance of the network then means solving the following optimization problem

$$\delta = \min_{\mathbf{s} \in \mathbb{B}_2^n} h(\mathbf{s}) = \min_{\mathbf{s} \in \mathbb{B}_2^n} \left(m - \frac{1}{2} \mathbf{s}^T J \mathbf{s} \right). \quad (7.1)$$

In correspondence of $\mathbf{s}_0 = \operatorname{argmin}_{\mathbf{s} \in \mathbb{B}_2^n} h(\mathbf{s})$, the residual positive terms in (7.1) correspond to the least number of unbalanced pairwise relationships between nodes (i.e. the frustrations at the ground state).

7.2.3 Statistics

Z-statistics for δ_{up} with respect to a null model

To evaluate how significant the global balance δ_{up} is, this value is compared with the one obtained by means of a null model. We have constructed two different null models, both using the same topology of the original graphs, but assigning the edge signs by means of different probability distributions:

1. Bernoulli $\mathcal{B}(q)$. From Table 7.1, $q = m^-/m$ is the probability of a negative edge. In the null model each edge sign is drawn as a i.i.d. variable from $\mathcal{B}(q)$.
2. Hypergeometric distribution with parameters m, m^- . This null model corresponds to randomly reshuffling the signs of the edges on the nodes, while maintaining exactly the same number of positive/negative edges of the original graph.

In both cases 100 random model instances are generated for each of the three networks. Lower and upper bounds on δ are then computed by the algorithm described in Chapter 5. The mean values on these lower and upper bounds are denoted δ_{low}^{null} and δ_{up}^{null} .

Sign skewness

In order to evaluate if a node is enriched in positive or negative nodes, a p-value is calculated for each of the nodes, with respect to a null model network obtained by randomly shuffling the edges signs (over the entire network) while keeping the same fraction q of negative signs. Let s be a node with total degree k and positive degree k^+ , then the probability of finding for the node s an amount ℓ of positive edges less or equal to k^+ in the null model counterpart is given by the cumulative binomial

$$\text{Prob}[\ell \leq k^+] = \sum_{i=0}^{k^+} \binom{k}{i} (1-q)^i q^{k-i} \quad (7.2)$$

where, as before, $q = m^-/m$ is the probability of the event “negative edge” in the original network J . A vanishing p-value means that the node s has a significant amount of negative edges and a significant lack of positive edges. An analogous formula is used for the analysis of negative edge enrichment. The nodes with a p-value less or equal than a given significance threshold θ are classified as enriched. The sign skewness index ϕ then corresponds to the total number of edges belonging to nodes with p-value less than θ . A Z-score can be computed as follows:

$$\text{Z-score} = (\phi - \langle \phi_{null} \rangle) / \sigma_{null}$$

with $\langle \phi_{null} \rangle$ the average skewness index of 100 null models and σ_{null} the corresponding standard deviation.

7.3 Results and Discussion

First we discuss the characterization drawn from the minimal energy configuration only (the ground state) and secondly we move to consider the features of energy landscape obtained from the analysis of all replicas.

7.3.1 Global Structural Balance: the ground state

The local search algorithm is applied to the three on-line social networks of Table 7.1. Some ~ 4700 replicas were computed for Epinions, ~ 8000 for Slashdot and ~ 18000 for WikiElections. Each replica yields an estimate from above of the true value δ , denote it $\delta_{up}^{(k)}$. Let $\delta_{up} = \min_k \delta_{up}^{(k)}$ be the best value among all replicas. These values are given in Table 7.2, where also the corresponding lower bounds on δ , δ_{low} , are shown. That this algorithm scales well with size, and in fact that it can deal effectively with the signed social networks of dimension $\sim 10^5$ used in this work, is proved by the tiny gap left between δ_{low} and δ_{up} , which guarantees that the estimate for δ is accurate. For 2 of the 3 networks, we have essentially computed the true optimum, as $\delta_{low}/\delta_{up} > 0.99$, while the residual gap in the third network (Slashdot, $\delta_{low}/\delta_{up} > 0.95$) is most likely due to the lack of precision of the lower bound computation.

By definition, a local optimum of the Hamiltonian $h(\mathbf{s})$ is any \mathbf{s} such that for every user the majority of pairwise relationships are “satisfied”, i.e. yield a zero contribution to (5.2). Due to the ruggedness of the energy landscape, the number of local minima can be huge [155]. It is only by solving (7.1) that a local minimum becomes also a global optimum and, in the present context, acquires the meaning of balance value for the network. Since our computed δ_{up} is very close to the true δ , essentially all the residual conflicts in J_σ are ineliminable, i.e. they represent the real disorder of the problem. Due to the gauge equivalence, what holds in the ground state $\mathbf{1}$ for J_σ holds also in the configuration $\mathbf{s}_0 = \sigma$ for the original J . In the

Table 7.2: **Global balance of the networks.** δ_{low} and δ_{up} are the lower and upper bounds on the global balance. These are much lower than δ_{low}^{null} and δ_{up}^{null} , the corresponding average values of balance obtained on null models generated from a Bernoulli distribution $\mathcal{B}(q)$. The ratio δ/m (the more conservative δ_{up}/m for us) represents the distortion, i.e. the fraction of frustrated bipartite relationships in the global balance configuration \mathbf{s}_0 . For the values of δ_{low} and δ_{up} , the ratio δ_{low}/δ_{up} is much higher than the value achieved by popular semidefinite programming approaches to MAX-CUT (0.8785, see [108]), meaning that our ground-state algorithm is indeed quite efficient.

| Network | δ_{low} | δ_{up} | δ_{low}^{null} | δ_{up}^{null} | δ_{up}/m | δ_{low}/δ_{up} |
|----------------------|----------------|---------------|-----------------------|----------------------|-----------------|----------------------------|
| <i>Epinions</i> | 50452 | 50806 | 105247 | 105520 | 0.0717 | 0.9930 |
| <i>Slashdot</i> | 70014 | 73604 | 90346 | 106163 | 0.1476 | 0.9512 |
| <i>WikiElections</i> | 14194 | 14245 | 20878 | 20880 | 0.1420 | 0.9964 |

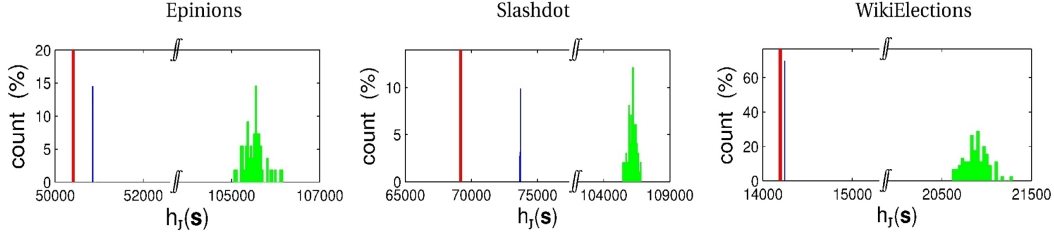


Figure 7.2: **Global balance and its statistical significance.** Comparison of $\delta_{up} = \min_k \delta_{up}^{(k)}$ with lower bounds and with null models generated from a Bernoulli distribution $\mathcal{B}(q)$. The lower bounds δ_{low} are shown in red, δ_{up} are in blue, and the distribution of δ_{up}^{null} in 100 null models in green. In each of the three networks $\delta_{up} \ll \delta_{up}^{null}$, meaning that the true networks are much less frustrated than expected from the null models (Z-test, with p-value $< 10^{-100}$). Furthermore, the interval of uncertainty of the optimal level of balance is very limited, since $\delta_{low}/\delta_{up} > 0.95$ ($\delta_{low}/\delta_{up} > 0.99$ for Epinions and WikiElections) and $\delta_{up} - \delta_{low} \ll \delta_{up}^{null} - \delta_{up}$.

optimal balance state \mathbf{s}_0 , a consistent fraction of users results to be completely free of tensions: from the 52.7% of WikiElections to the 83.7% for Epinions.

Statistical analysis of the level of balance

For a signed graph, the amount of frustration depends on the topology of the network, on the percentage of negative edges and, most of all, on their distribution over the graph. Unlike for spin glasses on regular lattices, for heterogeneous signed networks systematic predictions of the expected frustration, given the connectivity and the percentage of negative edges, are completely missing. We observe that in the three social networks we analyzed, the fraction of negative edges is always limited ($q = m^-/m = 16.7 \div 23.6\%$). In terms of spin glasses, this would correspond to a “partially ferromagnetic” quenching (more ferromagnetic than anti-ferromagnetic bonds). Obviously this leads to a lower frustration than in a spin glass with equally distributed edge signs. To evaluate if also the arrangements of the negative edges on the graph are favoring balance, we have to compare the sign arrangements on our networks with null models. In the null models we discuss here the edge signs are drawn as i.i.d. variables from a Bernoulli distribution with probability of negative sign equal to q . For each of the three networks, 100 randomizations were performed, and the corresponding δ_{low}^{null} and δ_{up}^{null} computed solving (7.1) by the same heuristic method used for the true networks (see Table 7.2). The distribution of the δ_{up}^{null} is compared with δ_{low} and δ_{up} of the true networks in Fig. 7.2. It can be observed that the null models are unavoidably much more frustrated than the real social networks (Z-test, p-value ~ 0). This means that indeed the organization of the signs in our social networks is such that tensions are largely avoided. Analogous results are obtained if the null models are constructed using a hypergeometric distribution, corresponding to reshuffling randomly the signs on the edges while keeping constant the ratio negative/positive edges.

For our networks, the property of being much more balanced than expected goes beyond the statistical significance of a Z-test on null models. As a matter of fact, δ_{up} is even lower than a Shannon-type bound which can be associated with the average frustration of our null models. For n and m sufficiently large, denoting $R = n/m$ the rate and $D = \delta/m$ the distortion, the rate-distortion Theorem (5.15) affirms that when the edge signs are drawn as i.i.d. variables from $\mathcal{B}(q)$ then the distortions achievable are in expectation lower bounded by the distortion-rate curves shown in Fig. 7.3, regardless of the topology of J . Distortions (and hence frustrations) that lie below this Shannon bound must be considered as obtained from edge sign assignments that are highly atypical for the probability “source” $\mathcal{B}(q)$. All three networks have sign arrangements that violate the Shannon bound. This confirms that the sign assignments of our networks are far from being typical for $\mathcal{B}(q)$, and in particular that their frustration is less than the frustration of a typical sequence of $\mathcal{B}(q)$ even if we arbitrarily modify the topology A . In Fig. 7.3 notice that, instead, the distortions δ_{low}^{null}/m and δ_{up}^{null}/m of the null models all lie above the Shannon bounds, as expected.

Skewness of the sign distributions and its social meaning

The feature that makes our networks so atypical is the skewness of the sign distribution on the individuals. In particular, the three networks have a significant fraction of nodes that are enriched for positive or negative edges (cumulative binomial test, p-value 10^{-5}), property not shared with the null models, see Fig. 7.4 and Table 7.3. Both fat tails of this sign distribution contribute to increase the balance of a network: the tail of positive edges, because users with many friends, have less enemies than expected from null models; the tail of negative edges for the opposite reason. A direct consequence of the sign skewness is that a considerable part of negative edges can be eliminated by means of gauge transformations, meaning that a vast fraction of the negative edges contribute only to the apparent disorder, not to the real frustration. On the contrary, the reduction of negative edges in the null models is always minimal, see Fig. 7.2 and Table 7.2.

That the reduction of negative edges in passing from J to J_σ is primarily due to users with high connectivity of enemies is confirmed in all three networks by the signed degree distributions of Fig. 7.5 (compare the degree distributions of negative edges in J and J_σ). In practice, a small fraction of individuals attracting a large number of negative edges contributes less to unbalance the social community than a homogeneous distribution of unfriendly relationships. The sociological interpretation of this fact is clear: unpopular individuals are easily “cast away” from the bulk of the community without creating much conflict within the community itself. Something similar does not happen for homogeneous distributions of the negative edges in the community. In conclusion, in all three networks analyzed, the local process of choosing friends/enemies induces a collective behavior which is strongly biased towards the creation of a disorder that is only apparent, thereby confirming the validity of Heider’s theory for this class of networks.

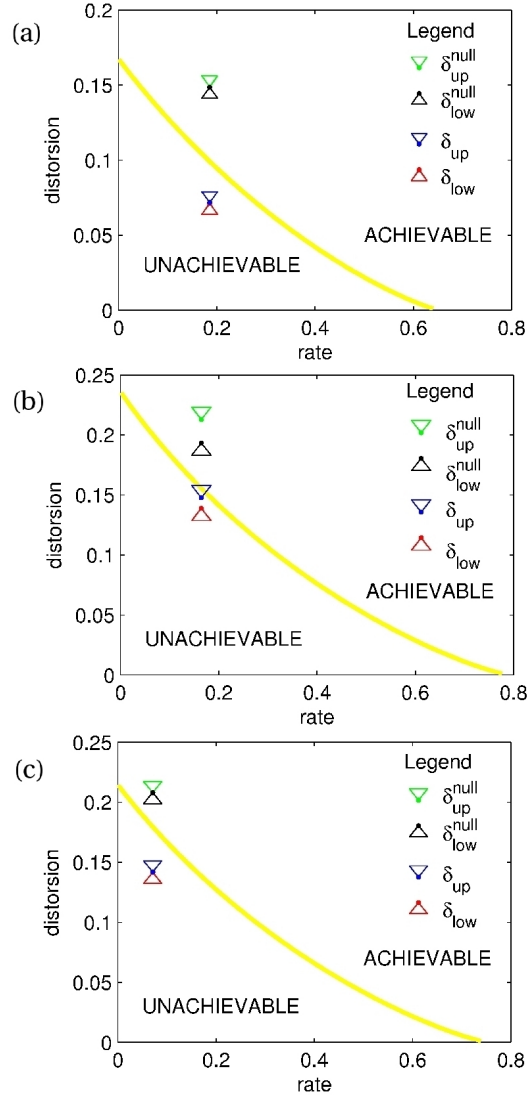


Figure 7.3: **Rate-distortion plots.** In the rate-distortion ($R = n/m$, $D = \delta/m$) plane, the yellow curves are the Shannon bounds of the rate-distortion theorem associated to a Bernoulli distribution $\mathcal{B}(q)$. The region above (below) the curve is achievable (unachievable) in expectation by an edge sign assignment drawn as a length- m sequence of i.i.d. variables from $\mathcal{B}(q)$, compressed to a length- n sequence, and then reconstructed. The distortion of the three true sign assignments (tip of the triangles, blue for δ_{up} and red for δ_{low} , partially overlapping) is less than this Shannon-type bound, meaning that these edge signatures are significantly away from a typical i.i.d. sequence from $\mathcal{B}(q)$. The signatures used in the null models of Table 7.2 (tip of the triangles, green for δ_{up}^{null} and black for δ_{low}^{null} , also partially overlapping) are instead in the achievable region.

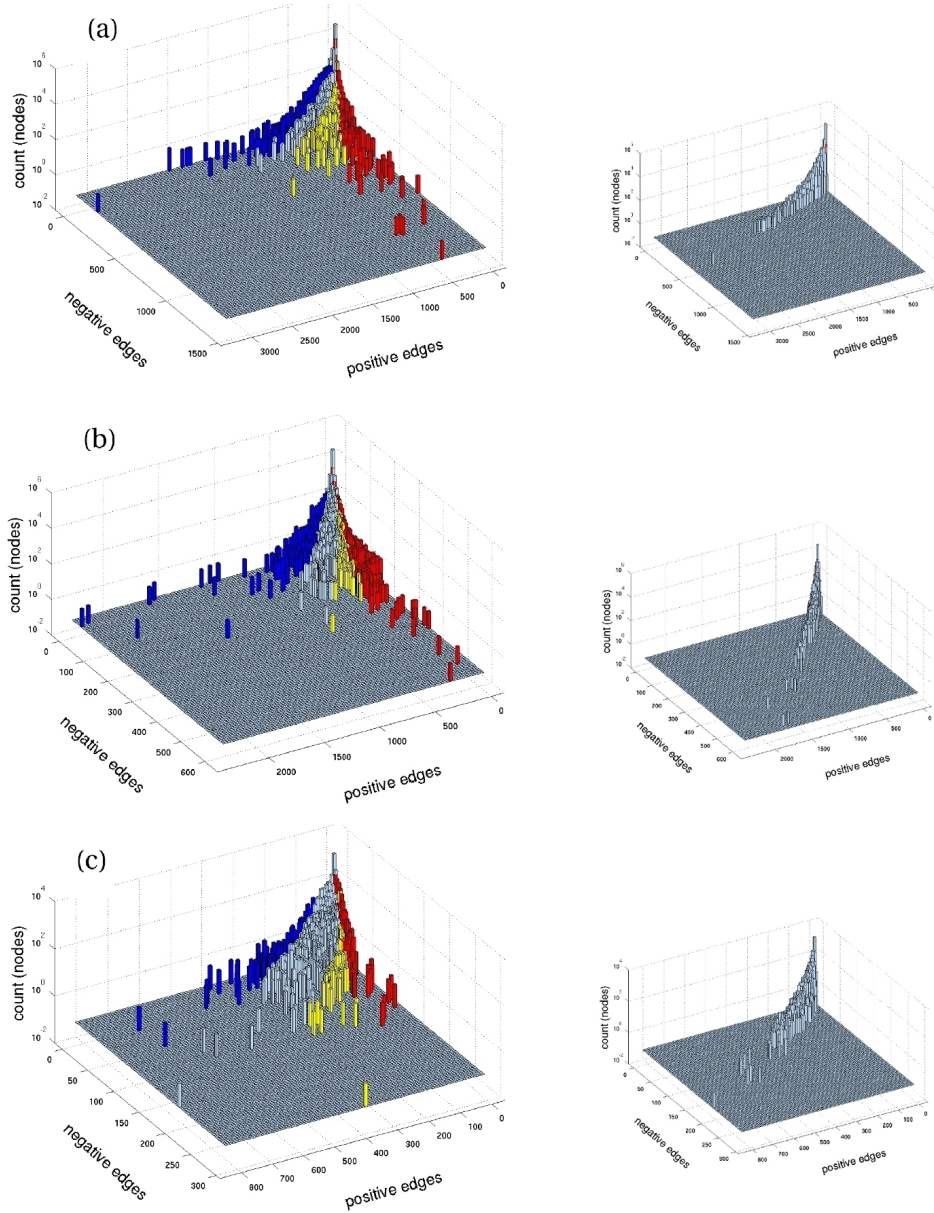


Figure 7.4: **Global balance and sign skewness.** *Left column:* All three networks have a significant percentage of nodes enriched for positive (blue) or negative edges (yellow and red), see Table 7.3. The sign skewness of a node is computed through a cumulative binomial test (p-value 10^{-5}). In particular, the nodes in red are adjacent to more negative than positive edges in J . Gauge-transforming these nodes reduces considerably the amount of negative edges of the networks while not altering their frustration. These histograms should be compared with the corresponding histogram for a null model (*right column*).

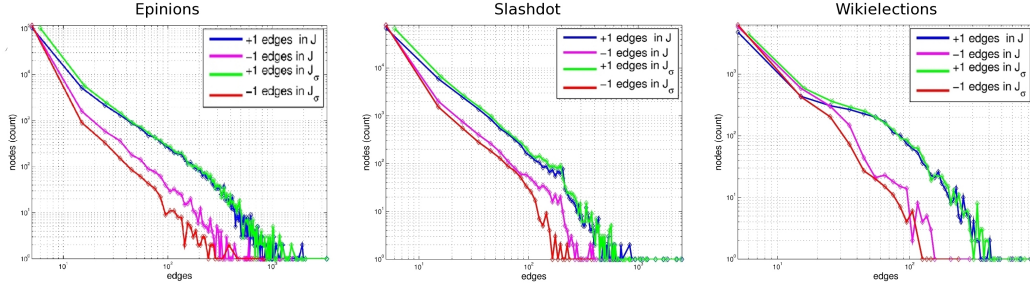


Figure 7.5: **Distribution of signed edges in the original J and in the optimal gauge transformed J_σ .** In all cases, passing from J to J_σ the number of nodes with high connectivity of negative edges (red tail of the distribution in Fig. 7.4) systematically decreases.

Table 7.3: **Sign skewness index.** The nodes significantly enriched for positive or negative edges are computed through the cumulative binomial models described in Section 7.2.3. Two significance thresholds are used: $\theta = 10^{-5}$ and $\theta = 0.01$. For the former, the 100 null models yield a very low ϕ_{null} , often equal to zero (this implies a very high skewness of the true networks). For $\theta = 0.01$ the Z-scores are still very high.

THRESHOLD $\theta = 10^{-5}$

| Network | nodes + | nodes - | tot. nodes (%) | ϕ | $\langle \phi_{null} \rangle$ | Z-score |
|----------------------|---------|---------|----------------|--------|-------------------------------|---------------|
| <i>Epinions</i> | 1619 | 2194 | 6.43 | 424247 | ~ 0 | $\sim \infty$ |
| <i>Slashdot</i> | 1121 | 1396 | 4.83 | 71781 | ~ 0 | $\sim \infty$ |
| <i>WikiElections</i> | 291 | 477 | 16.05 | 52879 | ~ 0 | $\sim \infty$ |

THRESHOLD $\theta = 0.01$

| Network | nodes + | nodes - | tot. nodes (%) | ϕ | $\langle \phi_{null} \rangle$ | Z-score |
|----------------------|---------|---------|----------------|--------|-------------------------------|---------|
| <i>Epinions</i> | 4148 | 6110 | 17.31 | 539839 | 16360 | 225.18 |
| <i>Slashdot</i> | 3363 | 3302 | 12.80 | 363950 | 10948 | 230.68 |
| <i>WikiElections</i> | 807 | 1021 | 38.20 | 82795 | 2682 | 116.10 |

7.3.2 Energy landscape characterization for Epinions and Slashdot

We now investigate the distribution of low-energy configurations corresponding to the different replicas computed by the optimization procedure. The minimization procedure of Section 5.2 is indeed repeated many times starting always from different initial points (independently and randomly chosen). Each run generates a replica: $\delta_{up}^{(k)}$ and $\mathbf{s}^{(k)}$ indicate the distortion and the spin configuration at the local minimum of replica k . Denote $\delta_{up} = \min_k \delta_{up}^{(k)}$ the best estimate obtained for the ground state energy over all replicas. Consider an energy band of width ϵ above δ_{up} . For both networks $\epsilon/\delta_{up} \sim 0.6\%$ is chosen, meaning a few hundreds of energy levels above

δ_{up} are considered (namely, $\epsilon = 310$ for Epinions and $\epsilon = 450$ for Slashdot). Only replicas reaching the interval $[\delta_{\text{up}}, \delta_{\text{up}} + \epsilon]$ are retained for further analysis. The number of such replicas is $r = 606$ for Epinions and $r = 5557$ for Slashdot.

This part of the work is limited to the networks of Epinions and Slashdot only, because of their peculiar features (WikiElection results are very similar to that obtained from Epinions).

Energy and Hamming distance distributions

The histogram of the values of energies $h(\mathbf{s})$ and the (relative) Hamming distances between these replicas are reported in Fig. 7.6 for the two networks. A first comparison of these energies shows a significant difference between the two networks: the histogram for Epinions reports only a single broad group whereas the histogram for Slashdot shows three peaks, see Fig. 7.6(a) and (b). If we compute the relative Hamming distances between each pair of replicas (Fig. 7.6(c) and (d)), the two networks still present a different behavior: for Epinions the minima are all close one to the other (Hamming distances are distributed like a single Gaussian peak); on the contrary, several peaks can be identified among the minima we have found for Slashdot.

It is worth noting that, since the energy has a global spin flip symmetry (i.e. $h(\mathbf{s}) = h(-\mathbf{s})$), the Hamming distances $d(\mathbf{s}^{(1)}, \mathbf{s}^{(2)})$ and $d(\mathbf{s}^{(1)}, -\mathbf{s}^{(2)})$ are not distinguishable. This amounts to saying that only relative distances between 0 and 1/2 can be considered (analogously to the Ising spin glass case). We follow this principle throughout the whole chapter.

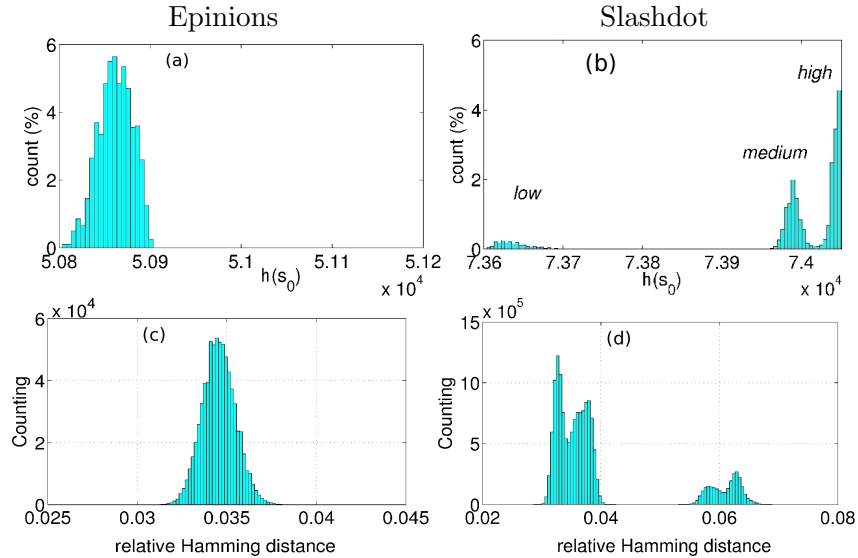


Figure 7.6: **Distribution of the energies and relative Hamming distances for all near-optimal replicas.** (a,c): Epinions. (b,d): Slashdot. The leftmost bar of histograms (a) and (b) corresponds to δ_{up} (blue line in Fig. 7.2).

Statistics over replicas

Since we can expect that each replica represents a different local minimum in the energy landscape, we can explore the distribution of the frequencies of the sign with which a spin appears in the r replicas: in particular, for each spin i we can define

$$\nu_i = \frac{1}{r} \min \left\{ \sum_{k=1}^r \frac{s_i^{(k)} - 1}{2}, \sum_{k=1}^r \frac{s_i^{(k)} + 1}{2} \right\}, \quad (7.3)$$

where the two summations represent the number of replicas in which the i -th spin has negative and positive sign respectively. In particular, for each value of ν between 0 and 1/2 (the histograms stop at $\nu = 1/2$ because of the global spin flip symmetry), the subset $W(\nu)$ of spins which have $\nu_i = \nu$ can be identified. For the subgraph corresponding to $W(\nu)$, we calculate:

$n(\nu)$: the number of spins;

$c(\nu)$: the number of connected components;

$z(\nu)$: the maximal size of these connected components.

Exploring the distribution of the frequency index $c(\nu)$ (see the middle panel in the first two rows of Fig. 7.7), we can observe that both networks have broad peaks around frequencies $\nu = 1/3$ and $\nu = 1/2$, meaning that the corresponding spins appear flipped in a half or in a third of the low-energy replicas. The peak at $\nu = 0$ corresponds to spins with equal orientation in all replicas, hence it constitutes a fixed “backbone” which does not contribute to the variability.

A characterization of the two peaks at $\nu = 1/2$ and $\nu = 1/3$ in the plot of the index $c(\nu)$ can be carried out through a probabilistic model. By construction, the replicas are statistically independent. We may also expect that the flipping components which belong to the peak at $\nu = 1/2$ are independent with respect to the flipping components around $\nu = 1/3$ (and viceversa). Under these assumptions, the two peaks can be modeled separately (the plot of $c(\nu)$ is then viewed as the superposition of two curves), considering the Binomial probability distribution $\text{Bin}(r, p)$ as the theoretical distribution for the variable $r\nu$ for each peak:

$$\mathbb{P}[r\nu = k] = \binom{r}{k} p^k (1-p)^{r-k} \quad (7.4)$$

(where the value of p is fixed at $p = 1/2$ and $p = 1/3$ for the two peaks). If $Y(k)$ is a random variable representing the number of connected components (over a population of Q flipping components, Q unknown) having $\nu = k/r$, and assuming that each connected component can flip independently, then

$$Y(k) \sim \text{Bin}(Q, \mathbb{P}[r\nu = k]).$$

By (7.4), its expectation value is

$$\mathbb{E}[Y(k)] = Q \binom{r}{k} p^k (1-p)^{r-k}, \quad (7.5)$$

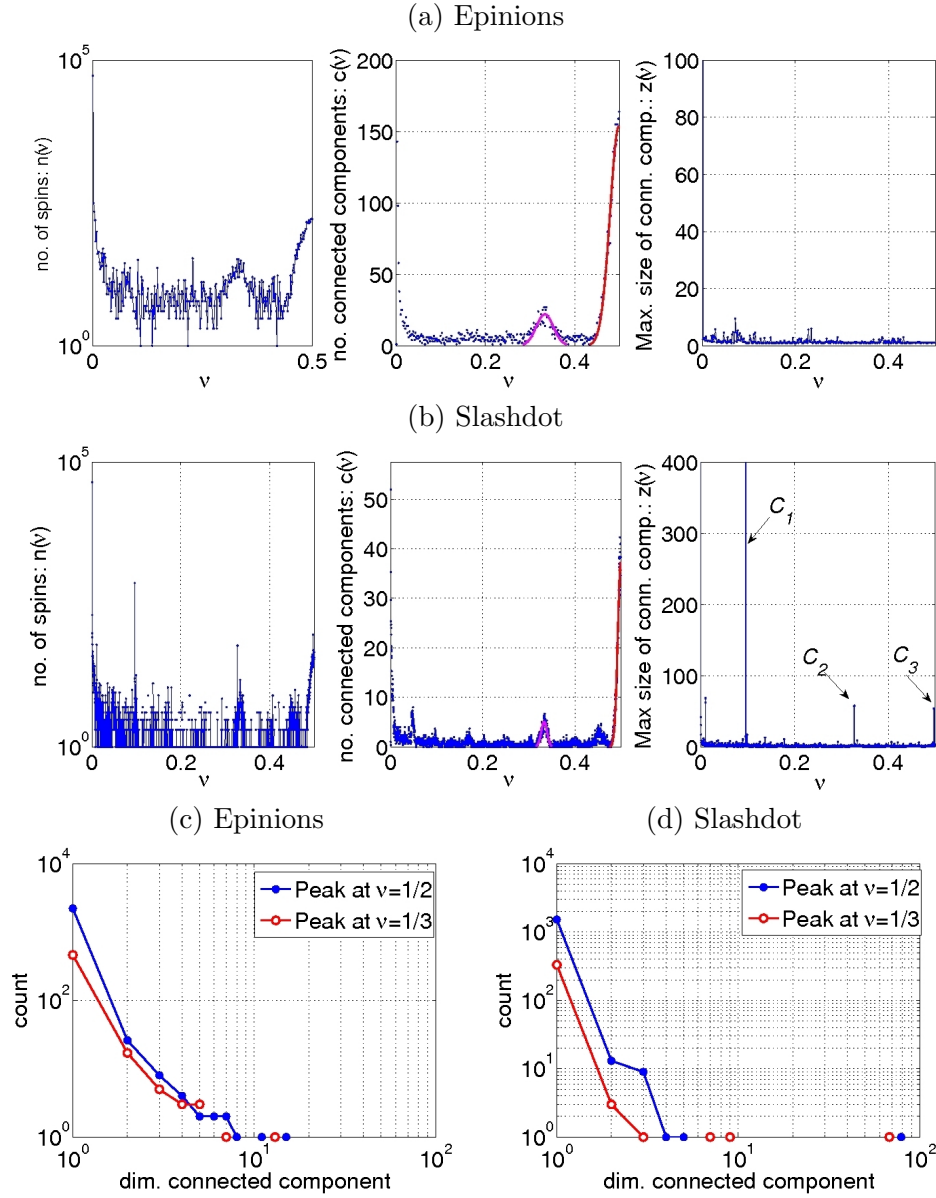


Figure 7.7: **Indices $n(\nu)$, $c(\nu)$ and $z(\nu)$ for the two networks.** (b): The letters C_1 , C_2 and C_3 indicate the three Slashdot communities identified in the spikes pointed by the arrows. Panels (c) and (d) report the dimensions of the connected components identified in the subgraph $W(\nu)$ of spins under the two peaks of $c(\nu)$ centered at $\nu = 1/3$ and at $\nu = 1/2$.

where the parameter Q can be obtained by linear regression, fitting the experimental data with k varying inside the support of the peak. Once a level of significance α has been fixed, we can select only spins which are in the central part of the fitted peak with p-value lower than α , as determined through a hypothesis test. The selected values of ν include only regions where the fitting error is smaller or at least comparable with the root mean square deviation (r.m.s.d.) of the data at each ν .

Extrapolating the connected subgraphs under the two peaks (Table 7.4 reports the results of the probabilistic model applied to the interpolation of these peaks), it is possible to obtain statistics of the recurrent motifs which form most of the variability of the spin configurations around a ground state. A catalogue of isoenergetic motifs (isoenergetic alternatives for a ground state) we identified through this interpolation is provided in Tables 7.5 and 7.6.

Isoenergetic motifs

Both the peaks at $\nu = 1/2$ and $\nu = 1/3$ described in Fig. 7.7 contain a large number of small disconnected components. For the two networks, the vast majority of these connected components are isoenergetic. In Fig. 7.8 these correspond to motifs having relative energy e^{rel} through the cut set equal to 0.5, where the relative energy of each motif \mathcal{M} is computed as the energy through the cut set that isolates the motif from the rest of the network, divided by the number of edges in the cut

Table 7.4: **Probabilistic binomial model for the spin frequencies.** Parameters of the fitting of the binomial probability distribution (7.5) to the broad peaks of $c(\nu)$ of spin frequencies $\nu = 1/2$ and $\nu = 1/3$, see Fig. 7.7 (second panel of (a) and (b)). Ranges are expressed as relative to the total number of replicas. For Epinions the selected intervals correspond respectively to $\nu \in [0.46, 0.50]$ and $\nu \in [0.31, 0.36]$. For Slashdot, the test yields the intervals $\nu \in [0.49, 0.50]$ and $\nu \in [0.32, 0.34]$.

| EPINIONS | | |
|-------------------------|------------------------|------------------------|
| <i>parameter</i> | $\nu = 1/2$ | $\nu = 1/3$ |
| k/r range for fitting | [263/606, 303/606] | [172/606, 232/606] |
| fitting parameter Q | 2378 | 642 |
| r.m.s.d. | 8 | 4 |
| α | 0.05 | 0.10 |
| selected region | [279/606, 303/606] | [187/606, 217/606] |
| no. selected components | 2336 | 528 |
| SLASHDOT | | |
| <i>parameter</i> | $\nu = 1/2$ | $\nu = 1/3$ |
| k/r range for fitting | [2648/5557, 2778/5557] | [1762/5557, 1942/5557] |
| fitting parameter q | 1740 | 447 |
| r.m.s.d. | 3 | 1 |
| α | 0.10 | 0.10 |
| selected region | [2717/5557, 2778/5557] | [1807/5557, 1897/5557] |
| no. selected components | 1656 | 419 |

set itself:

$$e_{\mathcal{M}}^{rel} = \sum_{\substack{i \in \mathcal{M} \\ j \notin \mathcal{M}}} (1 - J_{ij} s_i s_j) / (2 \sum_{i \in \mathcal{M}} c_i),$$

with c_i the connectivity through the cut set of the i -th node of the motif \mathcal{M} (c_i^+ and c_i^- are the corresponding numbers of +1 and -1 edges of c_i). Hence $e_{\mathcal{M}}^{rel} = 0$ means that with respect to the ground state the motif \mathcal{M} has all “satisfied” edges across the cut set, $e_{\mathcal{M}}^{rel} = 0.5$ means it has 50% of frustrated edges and $e_{\mathcal{M}}^{rel} = 1$ means 100% of edges frustrated. A catalog of the isoenergetic motifs under the two peaks is given in Tables 7.5 and 7.6.

While the presence of isoenergetic motifs under the $\nu = 1/2$ peak is straightforward to explain, the abundance of such motifs under the $\nu = 1/3$ peak is less obvious and requires an extra investigation. On what follows we restrict ourselves to the size-1 isoenergetic motifs of Epinions (in this case the motif \mathcal{M} is represented and indicated by its only node i). For each low energy replica σ (with the relative gauge-transformed adjacency matrix J_{σ}) and for each node i (size-1 isoenergetic motif under either the peak at $\nu = 1/2$ or that at $\nu = 1/3$) we calculate the percentage of edges which change sign with respect to the initial matrix J . This value is given by the ratio:

$$\bar{\rho}(\sigma, i) = \frac{\|J_{\sigma}^{(i,\cdot)} - J^{(i,\cdot)}\|_1}{2\|J^{(i,\cdot)}\|_1},$$

where $J^{(i,\cdot)}$ represents the i -th row of the matrix and $\|\cdot\|_1$ the L^1 -norm. We can easily attribute a plausible meaning to some of the values that this ratio can assume:

- $\bar{\rho}(\sigma, i) = 0$: the gauge transformation σ flips neither node i nor its neighbors;
- $\bar{\rho}(\sigma, i) = 1/2$: the gauge transformation σ flips only half (mostly one) of the neighbors of node i ;
- $\bar{\rho}(\sigma, i) = 1$: the gauge transformation σ flips the node i (or, less likely, all its neighbors).

In order to compare two differently populated peaks at $\nu = 1/2$ and $\nu = 1/3$, for each σ we must normalize the counts for the three cases:

$$\begin{aligned} \rho_0(\sigma) &= \frac{N_0}{N}; \\ \rho_{1/2}(\sigma) &= \frac{N_{1/2}}{N}; \\ \rho_1(\sigma) &= \frac{N_1}{N}, \end{aligned}$$

where N_k is the number of i such that $\bar{\rho}(\sigma, i)$ is equal to k and $N = \sum_k N_k$. As most pairs (σ, i) are such that $\bar{\rho}(\sigma, i)$ assume the values 0, 1/2, or 1, we have the “empirical” constraint:

$$\rho_0(\sigma) + \rho_{1/2}(\sigma) + \rho_1(\sigma) \approx 1 \quad \forall \sigma. \quad (7.6)$$

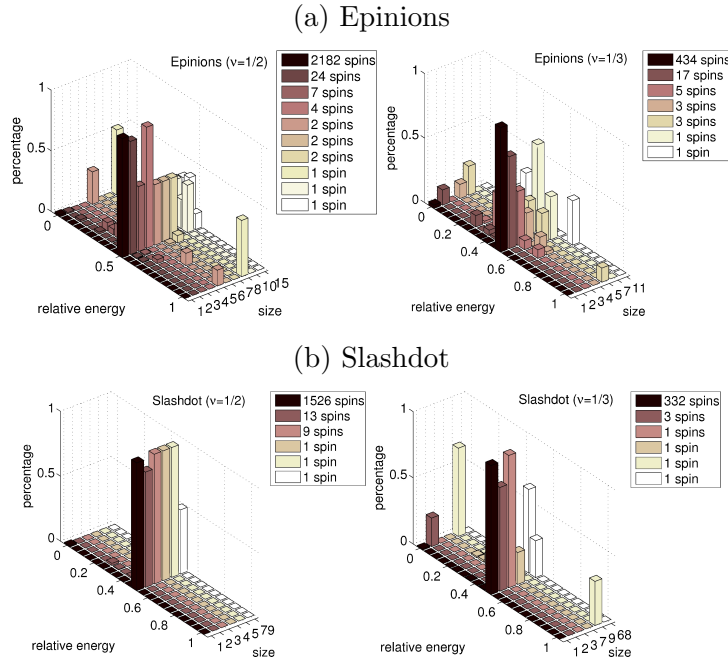


Figure 7.8: **Relative energy of the motifs flipping with frequencies $\nu = 1/2$ and $\nu = 1/3$.** These motifs are grouped according to their size. The gray-scale represents their number. In all histograms the relative energy (energy through the cut set divided by the corresponding number of edges) is concentrated around 0.5, meaning that half of the edges are frustrated in the ground state. Therefore, the peaks at $\nu = 1/2$ and $\nu = 1/3$ contain mostly isoenergetic motifs.

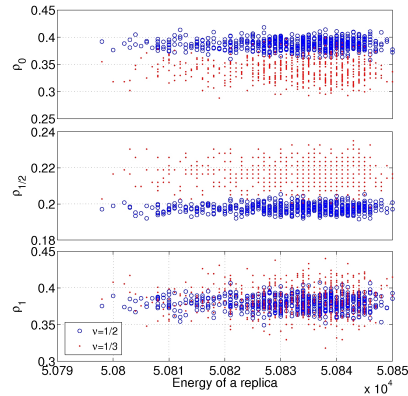

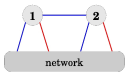
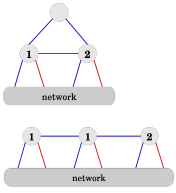



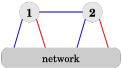
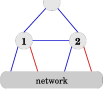
Figure 7.9: **Origin of the size-1 motifs for $\nu = 1/2$ and $\nu = 1/3$** (data from Epinions). Plots of the ρ_0 , $\rho_{1/2}$ and ρ_1 ratios for the 2172 isoenergetic size-1 motifs identified under the peak at $\nu = 1/2$ (circles) and for the 442 size-1 isoenergetic motifs belonging to the peak at $\nu = 1/3$ (points). The case $\rho_{1/2}$ (referring to gauge transformations that change 50% of the edges through the cut set) is more frequent under the $\nu = 1/3$ peak than under the $\nu = 1/2$ one.

Table 7.5: **Isoenergetic motifs identified under the peak at $\nu = 1/2$.** The values c_i^+ and c_i^- refer to the number of positive and negative edges from the i -th node to the rest of the network in the gauge transformed J_σ . The motifs are classified according to their size and then to the size of the cut set with the rest of the network. In the case marked with the asterisk, one of the triangles on the top presents a frustration. The pictures reported in the table represent examples of the corresponding motifs: blue line indicates a positive edge, red line a negative edge.

| motif \mathcal{M} | c_1^+ | c_1^- | c_2^+ | c_2^- | Epinions | Slashdot |
|---|---------|---------|---------|---------|-------------|-------------|
|  | 1 | 1 | - | - | 1928 | 1469 |
| | 2 | 2 | - | - | 180 | 155 |
| | 3 | 3 | - | - | 40 | 41 |
| | 4 | 4 | - | - | 15 | 14 |
| | 5 | 5 | - | - | 6 | 6 |
| | 6 | 6 | - | - | 2 | 6 |
| | 7 | 7 | - | - | 1 | 2 |
| | 8 | 8 | - | - | | 3 |
| | 9 | 9 | - | - | | 2 |
| | 17 | 17 | - | - | | |
| 24 | 24 | - | - | | | |
| 1 node: TOTAL | | | | | 2172 | 1698 |
|  | 1 | 1 | 1 | 1 | 13 | 5 |
| | 1 | 1 | 2 | 2 | 5 | 2 |
| | 1 | 1 | 3 | 3 | 2 | 2 |
| | 1 | 1 | 4 | 4 | 1 | 1 |
| | 1 | 1 | 5 | 5 | 2 | |
| | 2 | 2 | 2 | 2 | 1 | 1 |
| | 2 | 2 | 3 | 3 | | 1 |
| 2 nodes: TOTAL | | | | | 24 | 12 |
|  | 1 | 1 | 0 | 0 | 2 | 4* |
| | 2 | 2 | 0 | 0 | | 2 |
| | 4 | 4 | 0 | 0 | 1 | |
| | 1 | 1 | 1 | 1 | 1 | 2 |
| 1 | 1 | 2 | 2 | | 1 | |
| 3 nodes: TOTAL | | | | | 4 | 9 |

The comparison of the values obtained for Epinions is shown in Fig. 7.9. For ρ_1 (bottom panel) the plots at $\nu = 1/2$ and at $\nu = 1/3$ almost totally overlap. The top plot for ρ_0 which, as already mentioned, refers to both node and neighboring nodes unchanged in a replica, instead shows a systematic difference (less frequent in $\nu = 1/3$ than in $\nu = 1/2$). From Eq. (7.6), such a difference is compensated by an equal but opposite difference in the middle panel for $\rho_{1/2}$. Recalling that $\rho_{1/2}$ refers to the case in which 50% of the neighboring nodes of node i are flipped, then we can conclude that the peak at $\nu = 1/3$ appears to be due, at least to some extent, to a bulk of the network which is less “rigid” than under the $\nu = 1/2$ peak and hence allows for more frequent internal rearrangements. It is worth noting that this behavior is uniform across all low-energy replicas.

Table 7.6: **Isoenergetic motifs identified under the peak at $\nu = 1/3$.** Legend as in Table 7.5. In the case marked with a double asterisk the two nodes are interconnected by a negative edge.

| motif \mathcal{M} | c_1^+ | c_1^- | c_2^+ | c_2^- | Epinions | Slashdot |
|---|---------|---------|---------|---------|------------|------------|
|  | 1 | 1 | - | - | 367 | 245 |
| | 2 | 2 | - | - | 51 | 53 |
| | 3 | 3 | - | - | 17 | 13 |
| | 4 | 4 | - | - | 2 | 6 |
| | 5 | 5 | - | - | 1 | |
| | 6 | 6 | - | - | 1 | 1 |
| | 9 | 9 | - | - | 1 | |
| | 15 | 15 | - | - | 1 | |
| | 17 | 17 | - | - | 1 | |
| | 20 | 20 | - | - | 1 | |
| 1 node: TOTAL | | | | | 442 | 318 |
|  | 1 | 1 | 1 | 1 | 2 | |
| | 1 | 1 | 2 | 2 | 1 | |
| | 1 | 1 | 3 | 3 | 2 | |
| | 2 | 2 | 3 | 3 | | 1 |
| | 1 | 0 | 0 | 1 | 3 | 1 |
| | 2 | 1 | 0 | 1 | 1 | |
| | 1 | 0 | 0 | 1 | | 1** |
| 2 nodes: TOTAL | | | | | 9 | 3 |
|  | 1 | 0 | 0 | 1 | 1 | 1 |
| 3 nodes: TOTAL | | | | | 1 | 1 |

EPINIONS: a single valley

We can deduce from the histograms in Fig. 7.6 that all the minima of the Epinions network belong to the same valley. As we move away from the bottom of the valley (global minimum), the energy tends to grow. This feature can be deduced from the analysis of the distance among replicas versus energy of a replica reported in Fig. 7.10(a). The 2D bar plot shows that, moving away from the minimum of energy, the Hamming distance increases monotonically. This means that the valley is characterized by sufficiently regular ascending walls, and that the basin of attraction of the global minimum is rather broad.

The high degeneracy of the ground state, which is also suggested by the sample trajectories of Fig. 7.10(b), is a well-known feature of Ising spin glasses with bimodal bonds [157]. The flipping of the isoenergetic motifs we have identified (see Tables 7.5 and 7.6), may explain the broadening of the distribution of the Hamming distance. Apart from these degeneracies, the Epinions network contains many small disjoint motifs (scattered at all frequencies ν) whose flipping slightly increases the energy. Their cumulative effect is responsible for the energy difference between replicas, which smoothly grows moving away from the ground state.

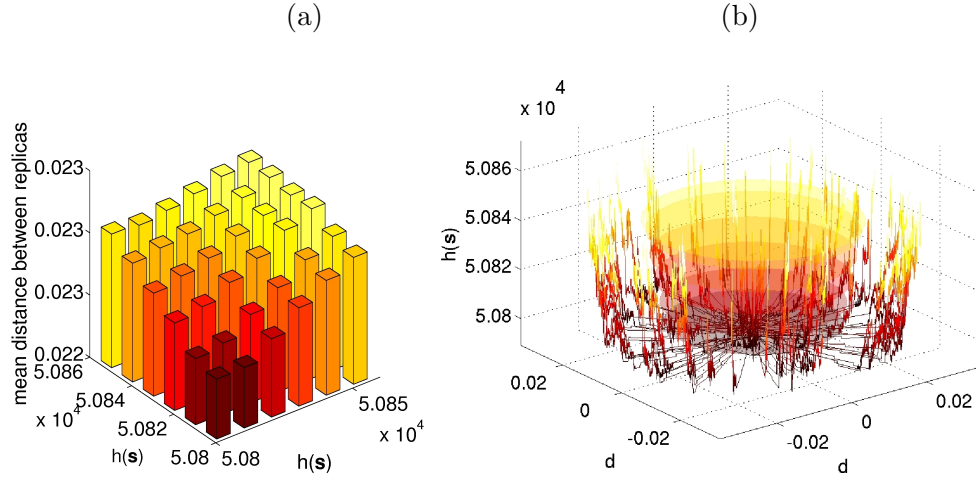


Figure 7.10: **Epinions energy landscape.** (a): Distance among replicas versus energy of a replica. The replicas of Fig. 7.6(a) are binned into 6 bins according to their energy, and the mean of the relative Hamming distances is computed. The vertical axis (and color code) represents the mean over bins of the relative distances. The replicas of least energy are also closer, and the distance grows regularly with the energy. (b): Sample minimal energy paths connecting the global and a local minimum. For visualization purposes, the trajectories are depicted as radially distributed according to a polar coordinate, with the global optimum placed in the origin. The vertical axis (and color code) represents the energy. The radius of the disks represent the average distance among the replicas in correspondence of the 6 bins of (a). The horizontal parts on the paths correspond to isoenergetic flips.

Table 7.7: **Slashdot balanced subnetworks:** nodes and edges of the 3 subgraphs C_i identified from Fig. 7.7. The row *rest* denotes the complement to $C_1 \cup C_2 \cup C_3$. Internal edges are those connecting two nodes of the same C_i , while external nodes are those connecting a node in C_i and one in C_j (or in *rest*). The C_i are visualized in Fig. 7.12.

| Subnetwork | Nodes | Edges | | | |
|-------------|-----------|---------|---------|---------|---------|
| | n_{C_i} | +1 int. | -1 int. | +1 ext. | -1 ext. |
| C_1 | 855 | 4169 | 1362 | 6268 | 6587 |
| C_2 | 62 | 11 | 107 | 143 | 148 |
| C_3 | 58 | 7 | 197 | 443 | 1009 |
| <i>rest</i> | 51073 | 345510 | 102856 | 6752 | 7259 |

SLASHDOT: competing valleys of near-optimal balance

For Slashdot, the low-energy landscape is markedly different: distinct peaks in the Hamming distances are clearly visible (Fig. 7.6) and are related to the three valleys observed in the energy histogram of Fig. 7.6(b). In fact, a scatter plot (Fig. 7.11(a)) reveals that while the two valleys at higher energy are nearby also in configuration space (cloud of points in the down-left corner in Fig. 7.11(a)) both of them are far away from the lowest valley (points in the upper-right corner). To confirm that indeed this multivalley profile is not due to undersampling of the low-energy landscape, the search for replicas in $[\delta_{\text{up}}, \delta_{\text{up}} + \epsilon]$ was performed a large number of times ($r = 5557$ for Slashdot whereas $r = 606$ for Epinions). Even increasing tenfold r , no new valley emerged for this network.

Looking at the spin frequencies of Slashdot (Fig. 7.7(b)) we can also observe a feature absent in Epinions, namely the presence of three sharp spike-like peaks for $z(\nu)$. If broad peaks usually contain small connected components, sharp, spike-like peaks in the plot of $z(\nu)$ are more likely to be associated to large connected components whose spins are simultaneously flipped in some of the replicas. A thorough analysis reveals that these three peaks correspond in fact to three connected subnetworks characterized by an high level of internal balance, higher than the rest of the network. In Table 7.7 we report some features of the corresponding subnetworks and in Fig. 7.12 we draw their adjacency matrices and graphs. These subnetworks are responsible for the formation of competing valleys of near-optimal balance described earlier. In fact, while the internal arrangement of their spins is usually frozen in each valley, their relative orientation with respect to the rest of the network may change passing from one valley to another, meaning that all spins of a subnetwork are simultaneously flipped, see Fig. 7.11(c). Computing a few sample trajectories from one valley to another (Fig. 7.11(b)), we can observe the presence of an energy barrier: the paths break the internal balance of some of the subnetworks before they are able to rearrange them again in another manner. In terms of our social networks, the formation of different valleys of near-optimal balance means the presence of possible alternative “alliances” between the majority of the users and a few internally balanced subnetworks of users (not necessarily friends among them). Fixing the relationship with one such subnetwork constraints the sign of the relationship with the other subnetworks. Different arrangements lead to slightly different global levels of balance for the whole network.

As already mentioned, a necessary condition for the spins of a subnetwork to have constant sign relative to each other in all low-energy replicas (which means also to be flipped simultaneously in all these near-optimal configurations) is that the subnetwork has to have a high level of balance internally, and a certain amount of frustration with respect to the rest of the network. At low energies, in fact, this favors a constant choice of spin orientation within the subnetwork which can however vary from valley to valley.

Denote C_1 , C_2 and C_3 the three subnetworks in Table 7.7 (and *rest* their complement in the original signed graph). By lumping together all nodes of each subnetwork, the corresponding matrix of adjacencies (blocks ordered as C_1 , C_2 , C_3 and

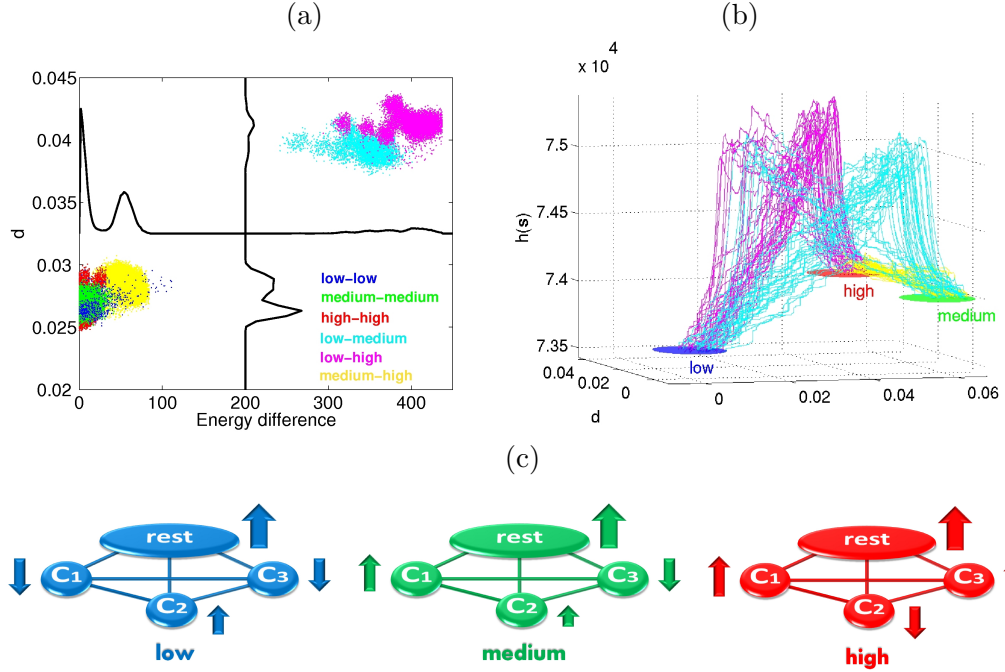


Figure 7.11: **Slashdot valleys** (a): The three valleys of minima corresponding to the three peaks in Fig. 7.6(b) are here called *low*, *medium* and *high*, according to their energy. For each pair of minima $\mathbf{s}^{(i)}$ and $\mathbf{s}^{(j)}$, the scatter plot shows differences in energy $|h(\mathbf{s}^{(i)}) - h(\mathbf{s}^{(j)})|$ versus the relative Hamming distance d for both intra and intervalley peaks. The two valleys *medium* and *high* are near and both more distant from the *low* valley. (b): A few sample trajectories connecting minima in different valleys. The radius of the disks corresponds to the average intravalley distance among the replicas. The height of the trajectories is indicative of the energy barrier between the valleys. To improve readability, the degenerate spin flips are not shown. (c): The three internally balanced communities C_i and their average magnetization in the three valleys of near-optimal balance.

Table 7.8: **Slashdot valleys**: $\langle \mathbf{s}_{C_i} \rangle_{low, medium, high}$ is the average magnetization of the subnetwork C_i (or of *rest*) in the three low-energy valleys of Fig. 7.11c.

| $\langle \mathbf{s}_{C_i} \rangle$ subnetwork | Valley | | |
|--|--------|--------|--------|
| | low | medium | high |
| C_1 | -0.712 | 0.712 | 0.712 |
| C_2 | 0.513 | 0.066 | -0.663 |
| C_3 | -0.517 | -0.517 | 0.506 |
| <i>rest</i> | 0.797 | 0.804 | 0.807 |

rest) is the following:

$$\mathcal{A} = \begin{bmatrix} 2807 & 16 & -206 & -129 \\ 16 & -96 & -1 & -19 \\ -206 & -1 & -190 & -359 \\ -129 & -19 & -359 & 242654 \end{bmatrix} \quad (7.7)$$

where the original amounts of positive and negative edges are:

$$\mathcal{A}(+/-) = \begin{bmatrix} 4169/1362 & 30/14 & 21/227 & 6217/6346 \\ 30/14 & 11/107 & 0/1 & 113/132 \\ 21/227 & 0/1 & 7/197 & 422/781 \\ 6217/6346 & 113/132 & 422/781 & 345510/102856 \end{bmatrix}. \quad (7.8)$$

Both C_2 and C_3 have a vast majority of internal negative edges, see Fig. 7.12 and Table 7.7. In correspondence of the optimal balance, the gauge-transformed adjacency matrix for the subnetworks is

$$\mathcal{A}_\sigma(+/-) = \begin{bmatrix} 5441/90 & 32/12 & 242/6 & 6726/5837 \\ 32/12 & 118/0 & 1/0 & 170/75 \\ 242/6 & 1/0 & 204/0 & 809/394 \\ 6726/5837 & 170/75 & 809/394 & 381167/67199 \end{bmatrix}$$

where it can be observed that the various C_i have very few residual negative edges (in particular C_2 and C_3 have none), meaning that indeed these subnetworks are internally balanced. If we look at the structure of the adjacency matrices of the C_i in Fig. 7.12, it is easy to understand why so many negative edges disappear in the gauge transformed \mathcal{A}_σ : the negative edges are all in correspondence of the same users. This is particularly visible in the adjacency matrix of C_1 : each row/column is highly skewed towards positive or negative edges. Such skewed sign distributions are the trademark for “apparent disorder”, i.e. negative edges which can be eliminated by means of gauge transformations and hence that do not spoil global balance. In Slashdot, users with a high number of negative edges are known as trolls [139]. Trolls do not add tension to the network, as they are unanimously tagged as foes by the other users. Looking carefully at Fig. 7.12 it is possible to observe that the subnetworks C_2 and C_3 , where negative edges are the vast majority, have nevertheless all positive cycles. In C_2 all cycles have length 3 and pass through the (positive) edge linking the two highly connected nodes. In C_3 , instead, cycles have length 4 and are composed of 4 negative edges.

Observing the pattern of signs (and sign flips) of the C_i in the 5557 low-energy replica of Slashdot, a high degree of regularity can be seen. Call $\mathbf{s}_{C_i}^{(j)}$ the spin configuration of the C_i subnetwork in the j -th low-energy replica, and n_{C_i} the number of spins of C_i . By construction, the relative Hamming distance within the subnetwork

$$d(\mathbf{s}_{C_i}^{(j)}, \mathbf{s}_{C_i}^{(k)}) = \left(1 - (\mathbf{s}_{C_i}^{(j)})^T \mathbf{s}_{C_i}^{(k)} / n_{C_i}\right) / 2$$

is always zero. The intervalley relative Hamming distance is shown in Fig. 7.11(a). The “average magnetization” within a subnetwork and within a valley is computed

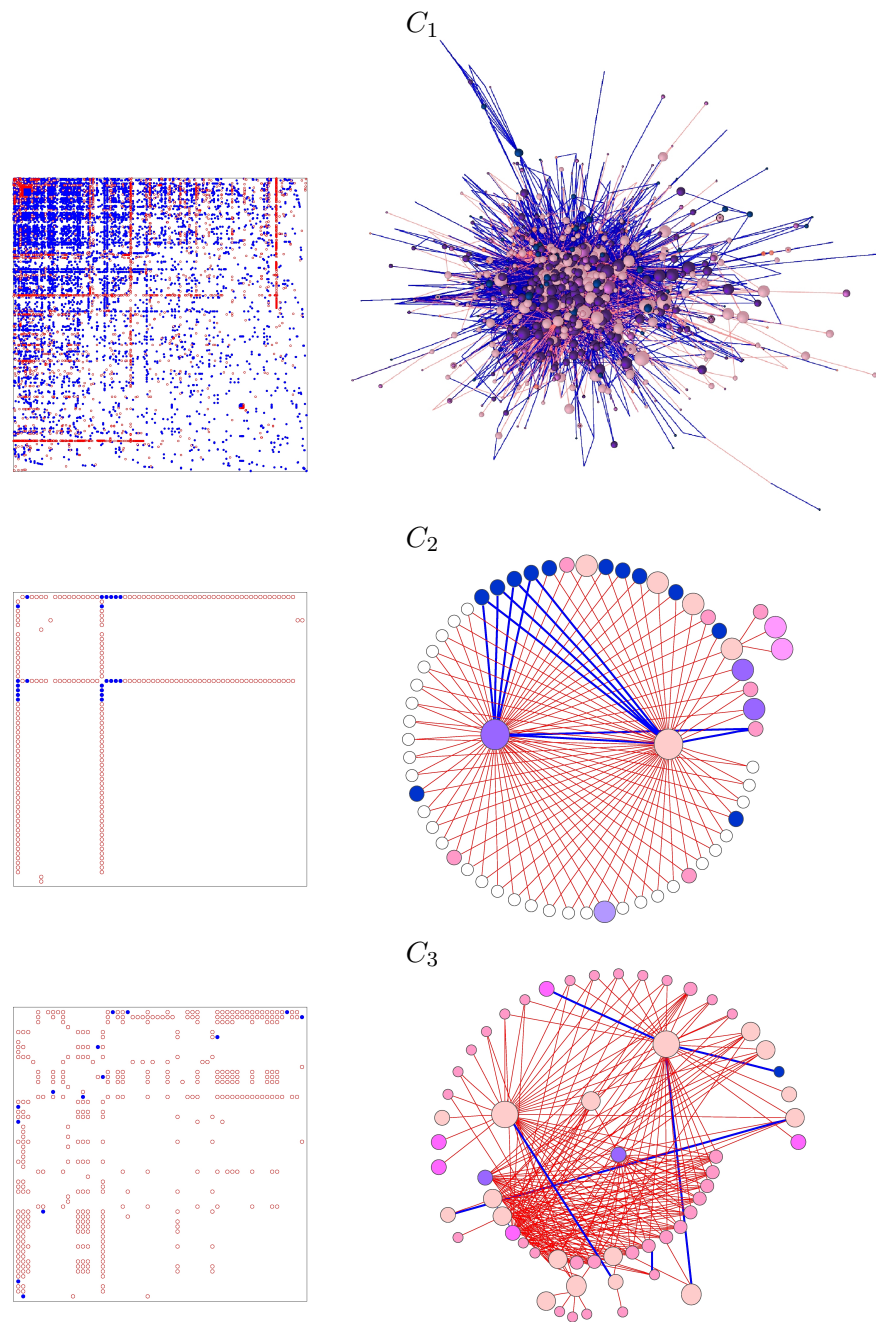


Figure 7.12: **Slashdot: the three internally balanced subnetworks C_1 , C_2 and C_3** of Table 7.7. The adjacency matrices (left column) and the corresponding signed graphs (right column) are shown. Blue (full) dots correspond to +1 edges, red (empty) to -1. In the corresponding graph, blue lines correspond to +1 edges, red lines to -1. The color-code for the nodes reflects the sum of the external edges: blue means positive sum, red negative, white no external edges.

restricting the computation to the \mathbf{s}_{C_i} spins and to the replicas falling into the valley. If r_{low} is the number of replicas in the *low* valley, then the average magnetization for the *low* valley is

$$\langle \mathbf{s}_{C_i} \rangle_{low} = \sum_{k \in C_i} \sum_{j \in low} s_k^{(j)} / (r_{low} n_{C_i}),$$

and similarly for $\langle \mathbf{s}_{C_i} \rangle_{medium}$ and $\langle \mathbf{s}_{C_i} \rangle_{high}$, $i = 1, 2, 3$. These average magnetizations are shown in Fig. 7.11(c) and Table 7.8. It can be observed that for example the spins \mathbf{s}_{C_1} are flipped passing from the *low* energy valley to the *medium/high* energy valleys. These community-wide flips do not modify the intravalley energy, but they alter the energy of the cut set between the C_i s and with the rest of the network. The increase in energy passing from one valley to another is a consequence of these changes. The cut sets between C_1 and C_2 and between C_2 and C_3 are always negligible, while the cut set between C_1 and C_3 can increase the energy considerably. This happens when $\langle \mathbf{s}_{C_1} \rangle$ and $\langle \mathbf{s}_{C_3} \rangle$ are not solidary, i.e. in the *medium* valley. This makes most of the difference in energy between *low* and *medium* valleys. In the *high* valley, instead, the magnetizations $\langle \mathbf{s}_{C_1} \rangle_{high}$ and $\langle \mathbf{s}_{C_3} \rangle_{high}$ are again aligned. However, both are flipped with respect to the *low* valley, and this reflects in the changes of energies in the cut sets relative to the *rest* of the network. Remember that *rest* contains a very large fraction of the network, and that it is highly biased towards positive magnetization (see Table 7.8), which is essentially uninfluenced by flipping of small communities. Nevertheless, the internal rearrangements inside *rest* due to the flips of \mathbf{s}_{C_1} and \mathbf{s}_{C_3} induce a consistent increase of frustration within *rest*. In summary, it appears that a feature useful for the creation of separated energy valleys is the presence of perfectly balanced subnetworks (like C_2 , C_3 and, to a large extent, also C_1). These can be composed of friends as well as of “declared” enemies: for what concerns structural balance, the trolls of C_2 and C_3 play exactly the same role as the users having many positive edges.

7.4 Conclusions

The point of view taken in this chapter, i.e. identify a signed social network with a spin glass (both are signed graphs), allows to pass from a local viewpoint of pairwise relationships between individuals to a global perspective of organization of the signs in a given social network. A byproduct of the approach is that a wide range of concepts developed for spin glasses can be used to investigate social networks. These include a “natural” energy functional whose global optimum corresponds to the value of structural balance in the network, a direct identification of frustration with the “social stress” at the basis of unbalance, and the possibility of evaluating near-optimal possible arrangements using the tools developed for the rough energy landscapes of disordered systems.

The first result we have presented in this chapter concerns the fact that currently available signed social networks are extremely balanced. In particular, individuals linked by a large majority of negative edges create mostly “apparent disorder”,

rather than true “frustration”. In the second part of the chapter we have shown examples of what can be achieved using a large number of low energy minima for the characterization of the same large-scale on-line networks.

In spite of a similar fraction of negative edges in the two larger networks (16.5% for Epinions and 23.9% for Slashdot) and of a similar low level of frustration (partially ferromagnetic behavior), two quite different low-energy landscapes have been identified: a single valley for Epinions and a set of three competing valleys for Slashdot. We associate this different qualitative behavior with the presence of exactly balanced (and highly connected) subnetworks, whose internal balance must be destroyed and reformed in order to pass from one valley to another. It is precisely the presence of these balanced subnetworks (and their arrangement with the rest of the network) that leads to a partial breaking of ergodicity in the system, and to the creation of energy barriers between competing near-optimal valleys. It is worth observing that these subnetworks are not necessarily composed by friends, making the determination of energy barriers like these a difficult task if one simply looks at the sign distributions on the edges and subdivides the nodes into clusters.

These results represent a clear demonstration that, for complex systems like large social networks, global properties cannot be inferred by local or mean features: more sophisticated analytical and computational tools must be developed and applied.

PART III

DIFFUSION AND SPATIAL GRADIENTS IN THE ROD PHOTOTRANSDUCTION PATHWAY

Introduction

In the works presented up to now, the focus has been on large networks and on a simplified description of the dynamics (or no dynamics at all). Indeed, the modeling approaches we have followed are based on some assumptions, for instance the steady-state assumption in FBA, which is possible because of the separation of time scales between metabolic and transcriptional processes. Similarly the Ising spin-glass models (and its formulation as data compression problem) provides a “static” description of gene regulatory and social networks. Another simplification we used so far is the absence of any spatial aspect. In FBA, for example, only a limited number of compartments are considered, together with the corresponding transport processes.

In this third part we are instead interested in characterizing the dynamical behaviour of a specific signaling pathway, namely phototransduction in rods of vertebrates. In particular we aim at using the photoresponse to light applied in a localized way along the cell body in order to understand the spatial gradient and to estimate the role of diffusion in the cytoplasm. For this purpose, the signaling pathway is here described by means of ordinary or partial differential equations. The number of molecular actors involved in the signaling is much smaller than in the complex networks of the previous chapters, hence we can effort to investigate the spatial and temporal dynamics of the system without having to deal with too many parameters.

Since several decades, phototransduction has been largely investigated and almost all the steps of the cascade have been elucidated (see Chapter 8). It has been also found that the efficiency of this signaling cascade depends on the part of the photoreceptor which is illuminated. However, an accurate modeling of this aspect is still missing. Since many mathematical models of the phototransduction have been already developed and published in the literature, our main goal is not to provide an additional omnicomprehensive model of this pathway, but to address the origin this spatial heterogeneity. In particular, our analysis and modeling are motivated by novel experiments carried out by Dr. Monica Mazzolini in the lab of prof. Vincent Torre in SISSA. These experiments in fact allow to look at localized light stimu-

lation with a much higher resolution and precision than in the existing literature (Chapter 9).

After my partial contribution to a first project on the adaptation of signaling pathways in rod photoreceptors and in olfactory neurons (not included in the thesis), a new paper about the efficiency gradient and diffusion is currently being submitted for publication.

- G. De Palo, G. Facchetti, M. Mazzolini, A. Menini, V. Torre and C. Altafini *Common dynamical features of sensory adaptation in photoreceptors and olfactory sensory neurons* Scientific Report, **3**, 1251–1258 (2013);
- M. Mazzolini, G. Facchetti, L. Andolfi, M. Lazzarino, R. Proietti, J. Treu, C. Altafini, E. Di Fabrizio, G. Rapp and V. Torre *Restricted spots of light reveal an efficiency gradient of the phototransduction cascade along rod outer segment*, submitted.

Chapter 8

Biology of rod photoreceptors

The biological research in the field of phototransduction began on the 40's: since then, an extremely vast literature has been published [158, 159, 160, 161, 162, 163]. Therefore, in this chapter we are going to give a brief summary of the state of the art, highlighting the links between cell biochemistry and cell geometry.

8.1 Rods: function and geometry

Vertebrates have developed specific neuronal cells for vision, cells that in evolved organisms like mammals are able to generate an electrophysiological response even to the absorption of a single photon. These cells are densely packed on the surface of the retina, forming a continuous film that maximizes the area in which photons can be captured. There are two types of photoreceptors: *cones* and *rods*. As hinted by the names, these cells differ in their geometry, conical and cylindrical, respectively (see Figure 8.1a). Cones confer color vision (there are three kinds of cones, each sensitive to the detection of red, green or blue color of light) and give fast response, whereas rods are slower but more sensitive to dim light (they provide night vision). This higher sensitivity of rods is caused by the higher amount of *rhodopsin*, the protein involved in the absorption of photons. Moreover, cones are less abundant (in humans there are about 6 millions cones in each retina) and are concentrated in the central part of the retina (the so-called *fovea*). On the contrary there are more the 120 millions of rods distributed around the fovea. Because of these differences, it is more easy to extract rods from the retina and to perform electrophysiological experiments on them: this is the reason why the most part of the research has been carried out on rods. Also in our work, we refer to this type of cell.

Rods consist of two parts: the *inner segment* (IS) and the *outer segment* (OS), see Figure 8.1(b). The former contains the nucleus and all organelles (especially a high number of mitochondria) for the common cellular functions, whereas the latter is the location where the phototransduction takes place. For this reason we focus our attention on the OS only. It is worth noting that light must first cross the IS before

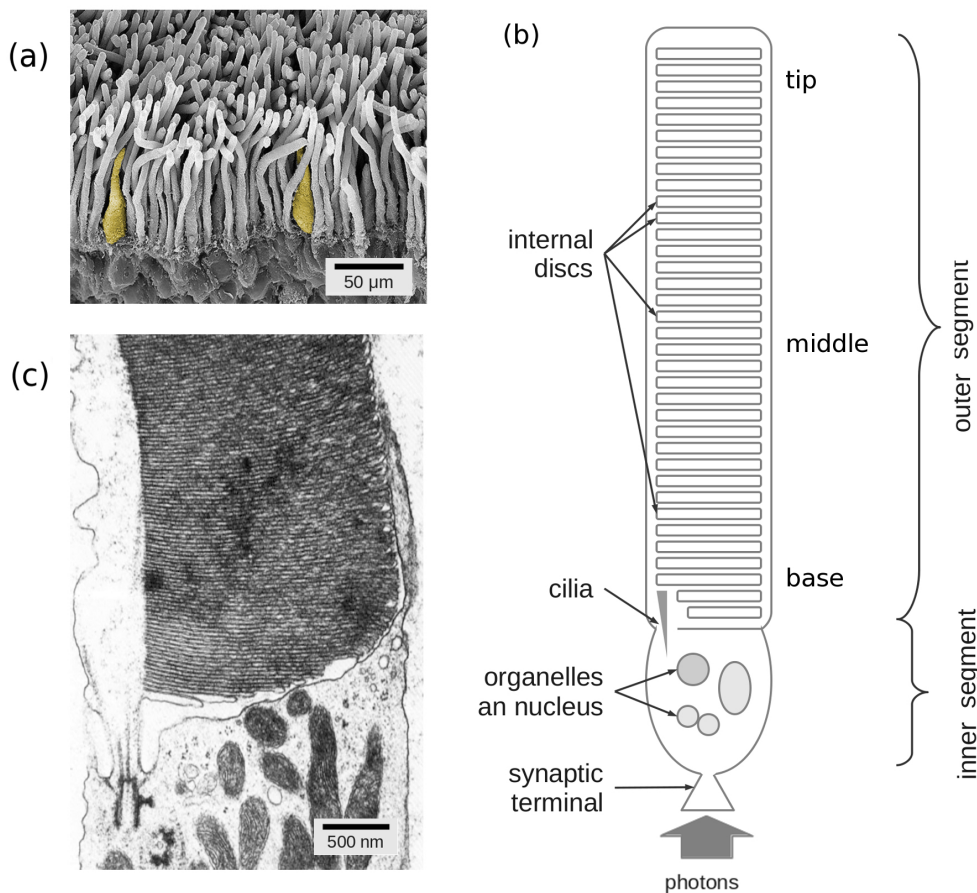


Figure 8.1: **Rod geometry** (a): Photoreceptors in the human retina. Most of the cells here are rod photoreceptors, but also two cone photoreceptors can be seen (yellow). (b): IS and OS of a rod photoreceptor: the communication between the two parts is assured by a motor cilia located at the interphase. (c): the OS is filled with internal discs which are produced at the base (close to the IS) by folding the membrane (image from *Rhesus monkey*, [163]).

being absorbed by the OS. In this manner, photons are mostly absorbed in the basal part of the OS (i.e. closer to the IS) where the phototransduction is more efficient (this gradient of efficiency is part of our investigation in the following chapter).

The OS contains a stack of thousands of so-called *discs* which are approximately 25 nm thick. Discs are densely packed: indeed, the distance between two adjacent discs and the distance between the disc border and the external membrane are both 20–25 nm only, see Figure 8.1(c). This dense packing assures an high concentration of rhodopsin. The same is valid for other proteins located on the disc surface and involved in the initial steps of the phototransduction. Therefore, in order to increase the probability of absorbing a photon, evolved and nocturnal species have a very long OS and thus a high content of rhodopsin. For example, while in toad and

salamander the length is about 20–30 μm , mammals like mouse or human have OS of 50–60 μm . In case of deep-sea fish the OS can even reach up to 100 μm length.

8.2 The signaling cascade

The geometry of the OS described above, is strictly correlated to the biochemical processes responsible for the detection of photons, amplification of the signal and generation of the photocurrent (due to the change in the membrane polarization at the synaptic terminal). We describe here the molecules and the corresponding reactions: all molecules are listed in Table 8.1 and the signaling cascade is depicted in Figure 8.3. Several negative feedbacks intervene to control the final response of the signaling pathways. Such pathways will represent the initial approach for our mathematical modeling (see Figure 9.6 in the next chapter).

Dark current

In absence of photons, the so-called *dark current* keeps circulating: indeed, without any light stimuli, ion channels are open and a flow of sodium and potassium ions enter in the OS and is pumped out from the IS by the Na^+/K^+ pump (see Fig. 8.2). Under this condition, a rod cell is not hyperpolarized and no synaptic activity is triggered (without light no signal to the brain is transferred). During illumination, instead, the current is reduced because of the induced phototransduction and a signal for the brain is generated. Although photoreceptors are neurons, they do not conduct action potentials, i.e. their synaptic activity induced by photons absorption does not generate a spike like a neuron of the central nervous system (the so-called “bursting”), but only a continuous modulated transduction current. In the neural network, downstream the rod, this induces a decrease in the release of the neurotransmitter glutamate to the postsynaptic cells.

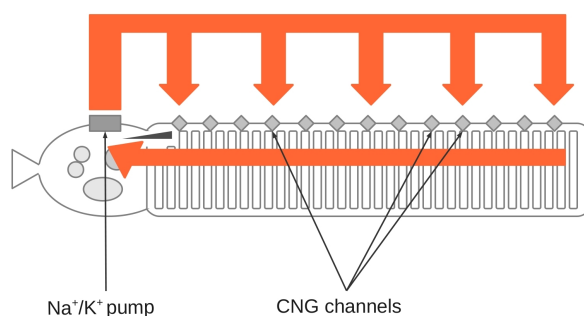


Figure 8.2: **Dark current in rod cells:** ions enter through the CNG channels located in the external membrane of the OS, migrate to the IS where they are pumped out by the Na^+/K^+ pump.

Table 8.1: Proteins and molecules of the rod phototransduction pathways.

| Name (abbrev.) | Location | Concentration | Diffusivity | Ref. |
|-----------------------------------|----------------|--------------------------|-------------------------------|------------|
| Rhodopsin (Rh) | internal discs | 25000 $\#/\mu\text{m}^2$ | 0.14 $\mu\text{m}^2/\text{s}$ | [158, 164] |
| G-protein (G) | internal discs | 2500 $\#/\mu\text{m}^2$ | 0.15 $\mu\text{m}^2/\text{s}$ | [158, 165] |
| Phosphodiesterase (E) | internal discs | 100 $\#/\mu\text{m}^2$ | 0.05 $\mu\text{m}^2/\text{s}$ | [158, 166] |
| guanylate cyclase (GC) | internal discs | 60 $\#/\mu\text{m}^2$ | n.a. | [158] |
| RGS9 complex (RGS9) | internal discs | 15 $\#/\mu\text{m}^2$ | n.a. | [167] |
| ions channels (CNG) | OS membrane | 100 $\#/\mu\text{m}^2$ | n.a. | [168] |
| cyclic GMP (cGMP) | cytoplasm | 3-5 μM | 70 $\mu\text{m}^2/\text{s}$ | [169] |
| calcium ions (Ca^{2+}) | cytoplasm | 200 nM | 15 $\mu\text{m}^2/\text{s}$ | [170] |
| Calmodulin (CM) | cytoplasm | 8 μM | n.a. | [170] |
| GC activating prot. (GCAP) | cytoplasm | 5 μM | n.a. | [171] |
| rhodopsin kinase (RK) | cytoplasm | 7 μM | n.a. | [158] |
| recoverin (Rec) | cytoplasm | 400 μM | n.a. | [158] |
| arrestin (Arr) | cytoplasm | 600 μM | n.a. | [158] |

Phototransduction

Vision begins with the photoisomerization of *11-cis-retinal* (the aldehyde of vitamin A), a molecule which is located inside rhodopsin. When this chromophore absorbs a photon, it isomerizes from the 11-cis state to the all-trans state. This change in the configuration induces a conformational rearrangement of the rhodopsin which becomes enzymatically active. There is a debate about the mobility of activated rhodopsin: Atomic Force Microscopy has shown that almost all rhodopsin molecules form a pseudocrystal structure (constituted by many dimers) which blocks the lateral mobility. However, a diffusion coefficient has been estimated to be nonzero, i.e. 0.14 $\mu\text{m}^2/\text{s}$ [164]. Moreover, the membrane of the internal disc is filled with other mobile proteins responsible for the amplification steps of phototransduction, namely *G-protein* (G, called also transducin) and *phosphodiesterase* (E), see Figure 8.3A. G-protein is a trimer (α , β and γ are the three subunits) bound to the membrane through two lipids, one from the α subunit and one from the β subunit. When a G-protein interacts with excited rhodopsin, a GTP molecule (guanosine-tri-phosphate) replaces a GDP molecule (guanosine-di-phosphate) in the α subunit, triggering the dissociation of the other two subunits (β and γ). It has been measured that within 1 second, each excited rhodopsin (i.e. each absorbed photon) activates tens of G-proteins [172]: this is the first amplification step of the phototransduction. GTP- α -G-protein, now linked to the membrane through a single lipid chain, has a greater mobility and migrate on the disc surface where it can encounter and activate one molecule of phosphodiesterase. The life time of this activated phosphodiesterase (the GTP- α -G-E complex) is determined by the hydrolysis of its GTP. This process is not fast, but it is catalyzed by the RGS9 enzyme which accelerates the deactivation by two orders of magnitude [173]. This first part of the cascade takes place on the surface of the disc membrane: it has been suggested that the reduction to a two-dimensional dynamics increases the probability of encounter of proteins and consequently increases the rate of the whole process [174, 175].

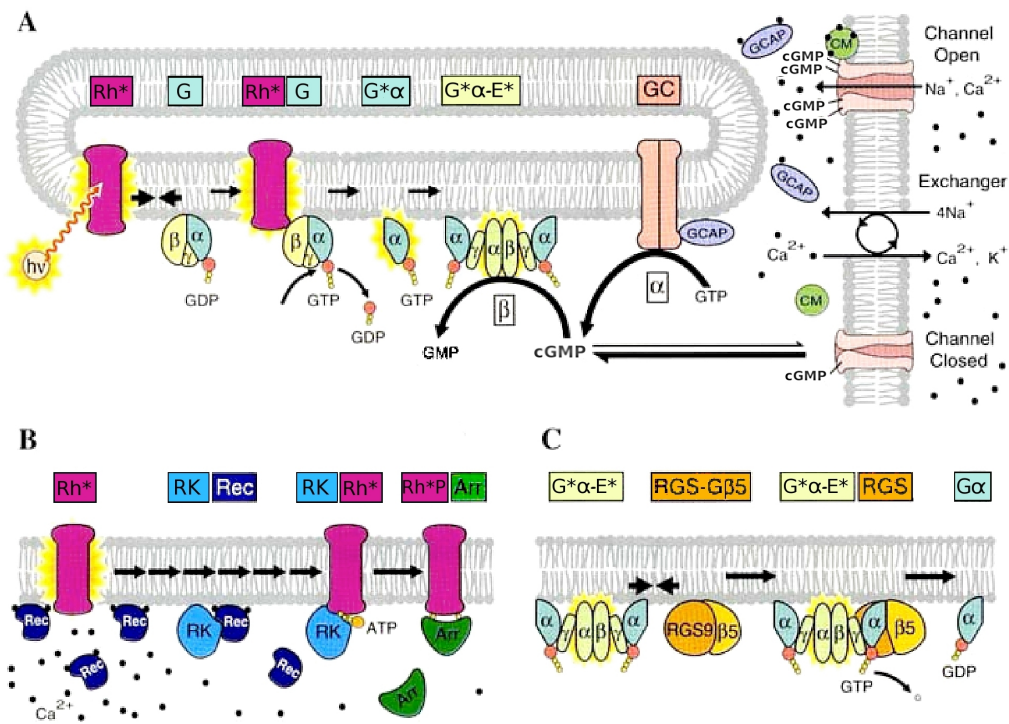


Figure 8.3: **Scheme of the phototransduction cascade in vertebrate rod photoreceptors** [158]. (A): Initial amplification steps of the cascade on the internal discs (the circular gray structure on the left part). From excited rhodopsin, α subunit of G-protein dissociates, phosphodiesterase is activated and cGMP is thus hydrolyzed. The right part of the scheme reports the processes which concern calcium ions: its entrance through the CNG channels regulated by cGMP, and the two feedbacks mediated by Calmodulin (on CNG channels) and by GCAP (on guanylate cyclase). (B): The picture shows the third calcium feedback which, involving rhodopsin kinase, recoverin and arrestin, leads to the complete switch off of rhodopsin and to the termination of the phototransduction. (C): the active complex GTP- α -G-E is deactivated by the RGS9 enzyme which catalyzes the hydrolysis of GTP to GDP and the consequent dissociation of the α subunit of the G-protein. The notation is the same as in Table 8.1; an asterisk indicates an activated protein.

On the contrary the following part of the cascade takes place mainly in the cytoplasm. The activated phosphodiesterase is indeed responsible for the hydrolysis of cyclic guanosine-mono-phosphate (cGMP), the second messenger of the signaling pathway, present in the cytoplasm of the OS. Because of the great catalytic effect of activated phosphodiesterase, also this step contributes to the amplification of the initial stimulus. Cyclic GMP is responsible for the opening of the calcium nucleotide

gated (CNG) channels. Under dark condition (i.e. without photon absorption), cytoplasm concentration of cGMP is about 3-5 μM : at this concentration approximately 10% of the CNG channels are open and both calcium and sodium ions enter in the OS. After light absorption, activated phosphodiesterase hydrolyzes cGMP closing the CNG channels and thus causing a drop of the circulating current. Suppression of the photocurrent causes the photoreceptors to hyperpolarise (thereby blocking the synaptic terminal) and the drop of concentration of calcium ions.

This lack of calcium is the starting point of three negative feedbacks which involve several steps of the cascade:

- ⊙ The most important feedback is the regulation of the biosynthesis of cGMP. Cyclic GMP is produced by guanylate cyclase (GC): this enzyme is linked to the membrane of the internal discs and is activated by the guanylate cyclase activation protein (GCAP). Since calcium ions bind and inhibit GCAP, in dark condition (high calcium concentration) the activity of GC is small. After light absorption, calcium is reduced and dissociates from GCAP [176]; GCAP is then no longer inhibited and thus GC starts to synthesize cGMP. This compensates the hydrolysis due to activated phosphodiesterase and then restores the initial concentration of cGMP.
- ⊙ Calcium ions act on the open CNG channels. Through a calmodulin-like protein, calcium binds to CNG and prevents the closure. Therefore, during phototransduction CNG channels tend to be closed and to reduce the dark current. Nevertheless, compared to the previous feedback, this effect is much weaker.
- ⊙ The third feedback acts on the origin of the phototransduction (i.e. on the activated rhodopsin) and its time scale is longer than the other two feedbacks. Recoverin, another calcium-binding protein, is normally bound to rhodopsin kinase when calcium is present. When the calcium level falls during phototransduction, calcium dissociates from recoverin releasing rhodopsin kinase which proceeds to phosphorylate the activated rhodopsin. This phosphorylation decreases the affinity between rhodopsin and G-protein (then slows down the initial amplification of the phototransduction). Finally, arrestin, another protein, binds and completely deactivates phosphorylated rhodopsin.

For the complete restoration of the dark conditions, all-trans retinal is transported from the OS to the pigment epithelial cells (other constituents of the retina), is converted to 11-cis retinal and then is transported back to the rods. Several other transports are facilitated by the motor cilia located between the IS and the OS. We mention here only three examples: ATP is produced in the IS by the mitochondria and then transported to the OS; similar situation is valid for GTP which is necessary for G-protein activation and for cGMP biosynthesis; rhodopsin is activated and phosphorylated in the OS and then is transported in the IS for being de-phosphorylated.

8.3 Discs ageing

As visible in Figure 8.1(c), internal discs are generated at the base of the OS by a folding of a membrane foil. Surprisingly, this process does not take place only once but it continues during the life of the rod photoreceptor. Indeed, by means of radiolabeled aminoacids a continuous formation of new discs at the base has been measured (see Figure 8.4), formation which is balanced by a continuous shedding at the tip where old discs are phagocytised by the retinal pigment epithelium [177]. This process has a turnover ranging from a few days in mice [178] to 6–7 weeks in frogs [178, 179].

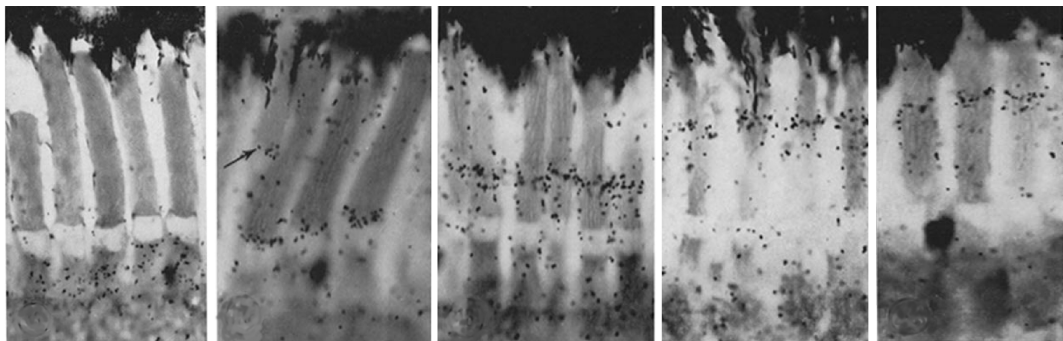


Figure 8.4: **Experimental evidence of discs renewal.** Radioautograph snapshots at different time from the retina of adult rats injected with methionine-³H. From left to right: just after injections, after 1 hour, 1 day, 7 days, 21 days and 24 days. The radioactive material incorporated in the disc proteins (the dark spots) moves from the base to the tip of the OS (images from [178]).

Electrophysiological experiments have shown that the phototransduction is less efficient when it involves old discs [180, 181], i.e. smaller photocurrents are measured. Nevertheless the biological and biophysical reasons for the degradation of aged discs are still unknown. The following hypotheses have been proposed:

- ⊙ **Lipids of the membrane.** An analysis of the lipidic composition of the membrane highlights that the cholesterol content decreases with age [182, 183]. It has been suggested that a higher level of cholesterol in young discs, although conferring more rigidity to the membrane (i.e. less motility of the embedded proteins), aggregates all proteins and increases the phototransduction efficiency. Similarly, the oxidation of lipids in older discs reduces the initial amplification step of the phototransduction [184].
- ⊙ **Transport to/from the IS.** An alternative hypothesis concerns the proteins and compounds involved in the cascade: referring to the fact that many proteins and key molecules (such as ATP) are regenerated and synthesized in the IS, the tip region (far from the IS) might have a lower efficiency if compared to the base (close to the IS). For instance, it has been measured that

phosphodiesterase is less abundant in the tip part of the OS [185]. For the same reason, the migration of ions from the tip part of the OS towards the sodium/potassium pump of the IS is more difficult and can generate a lower photocurrent [180].

- ⊙ **Surrogate antioxidant.** Every cell uses glutathione (GSH) as antioxidant. GSH is regenerated by a mechanism involving NADPH-dependent glutathione reductase. However, since this enzyme competes with the regeneration of retinol, rod photoreceptors are deprived of GSH. Therefore, the production of new discs and the elimination of old discs damaged by the photo-oxidation is interpreted as a surrogate antioxidant process necessary for preserving the cellular functions [186].

In order to evaluate the effect of ageing, we are interested in finding a correlation between the position of the discs in the OS and the efficiency of the phototransduction. For this reason, experimental illuminations which induce only the activation of the tip or only the activation of the base of the OS are reported in the following chapter: the obtained electrophysiological results are analyzed and reproduced through mathematical models.

Chapter 9

Signaling efficiency and cGMP diffusion in rod photoreceptors

9.1 Motivation

In most electrophysiological assays, rod photoreceptors are illuminated by a diffuse flash of light. During such experiments with a high light intensity, all the internal discs are stimulated (all CNG channels are closed and the photocurrent is saturated) and no spatial gradient appears along the OS. On the contrary, when a very dim light is used (i.e. a few photons are absorbed) only a few discs may be excited generating some spatial gradient (i.e. an inhomogeneous stimulation) along the OS. However, the location where this takes place is unknown. Consequently, diffuse light permits either uniform saturation or spatial inhomogeneities but never both at the same time. In order to control both, in this work we use localized illumination by tapered optical fibers (TOF). By placing a TOF in the proximity of the OS, the tiny beam of light emitted by the TOF can illuminate a reduced portion of the discs, leaving the remaining in dark-adapted conditions. While parameter distributed models based on the cylindrical geometry of the OS have already been published in the literature [187, 188, 189], none of them accounts for this spatial inhomogeneity. In particular, such experiments provide (i) information about the gradient of efficiency along the OS and (ii) evaluation of the diffusion of cGMP in the OS geometry.

Concerning the first point, several publications have already measured and evaluated a higher (resp. lower) efficiency of the signaling cascade in the basal part (resp. in the tip) of the OS [181, 180]. However, as mentioned in the previous chapter, a molecular explanation of this spatial gradient is still missing and so is an interpretation in terms of dynamical models.

Also estimates of the cGMP diffusivity have been published in literature. From the variation of the circulating current recorded from a truncated rod (i.e. which is losing the cytoplasmic cGMP), Koutals et al. [169] estimated a diffusion coefficient

of cGMP of approximately $70 \mu\text{m}^2/\text{s}$. Similar values have been obtained by other authors through model-based estimates, in which the diffusion coefficient is considered one of tens of parameters in a fitting to experiments [189, 169]. In our opinion, such high number of parameters to fit and the absence of localized illumination make the estimation of the diffusion less reliable. It is worth noting that the values usually reported in the literature refer to the so-called “longitudinal diffusion”, i.e. the 1D mobility of cGMP along the gap between the disc and the external membrane of the OS. Indeed, the stack of the internal discs slows down significantly the diffusion in the proximity of the external membrane of the OS: it has been estimated [170] that this “longitudinal” value is 6.5 times lower than the real 3D diffusion coefficient (usually called “radial diffusion”) that we are interested to determine.

Unless differently specified, all experiments reported in the following have been performed by Dr. Monica Mazzolini in the lab of Prof. Vincent Torre (Neurobiology Department at SISSA). Our contribution consists in the analysis of the experimental data and in the corresponding mathematical modeling.

9.2 Materials and Methods

Photocurrent recordings. Dissociated rods are obtained from adult male *Xenopus laevis* frogs. After mechanical isolation, the IS of an isolated and intact rod is drawn into a borosilicate glass electrode filled with Ringer’s solution. Rods are viewed under infrared light and stimulated with a 491 nm diffuse light laser. The restricted light is applied using a TOF. The photoresponses to brief flashes (5 ms) are recorded through the suction electrode. The sequence of the experimental steps is reported in Figure 9.1.

Characterization of the flash of light. In order to characterize the flash light exiting from the TOF, the following physical characterization has been performed by Dr. Laura Andolfi from CBM (Trieste). By using an ultrasensitive CCD camera we measure that the exiting light has a profile which illuminates about 5–6 μm (see Figure 9.2) and which corresponds to 100–120 discs of the OS (see, for instance, [180, 181, 190] for a comparison with the literature).

9.3 Gradient of efficiency along the OS

9.3.1 Experimental results

We compare photoresponses of rods exposed to conventional diffuse light and to localized illumination (by TOF). All the evoked electrophysiological currents are reported in Figure 9.3. In particular, we delivered localized illuminations at different locations of the OS (namely at the base, middle and tip) and with different light intensities (both saturating and non-saturating). Saturating lights are analyzed and plotted in Figure 9.4, whereas non-saturating lights are shown in Figure 9.5.

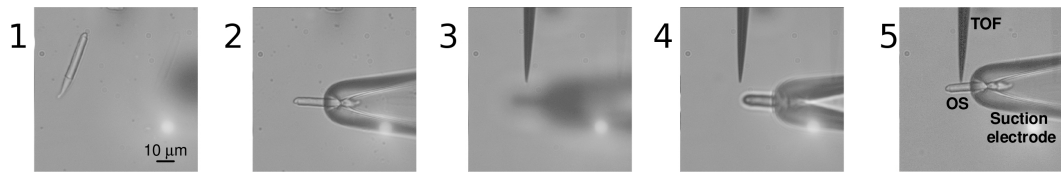


Figure 9.1: **Sequence of experimental steps.** The isolated rod (1) is fixed in the suction electrode (2). Then the TOF (3) is located close to the rod (4) and a localized part of the OS is illuminated (5).

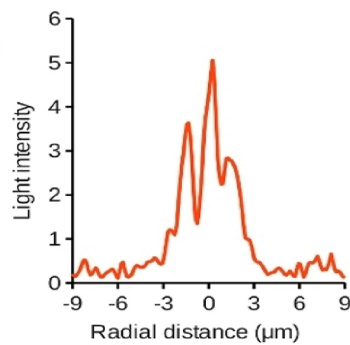


Figure 9.2: **Characterization of flash of light.** Profile of light from TOF: the size of the illuminated part is about 5–6 μm.

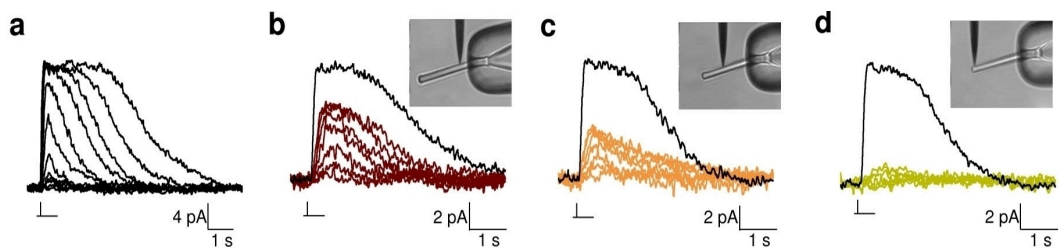


Figure 9.3: **Illumination with TOF at different positions along the OS.** (a): family of photoresponses to diffuse light of increasing intensity; (b): family of photoresponses obtained with the TOF illuminating the base of the OS (purple traces) compared to the photoresponse to a diffuse saturating flash of light (black trace); (c): as in (b) but with the TOF illuminating the middle of the OS and traces in orange; (d): as in (b) but with the TOF illuminating the tip of the OS and traces in green.

Saturating light

Photoresponses to pulses of saturating diffuse light and to saturating localized light (respectively black and colored traces of Fig. 9.4) exhibit different features:

- ⊙ **AMPLITUDE:** when only the base is illuminated (Fig. 9.4(b)) photoresponse has an amplitude equal to 0.52 ± 0.08 of that obtained by diffuse light. When only the middle (Fig. 9.4(c)) or only the tip (Fig. 9.4(d)) are illuminated this factor is 0.32 ± 0.04 or 0.15 ± 0.05 , respectively. These results are summarized in the histogram reported in panel (e) of Figure 9.4.
- ⊙ **TIME TO PEAK:** concerning the dynamics, photoresponses to diffuse light reach their peak in 190 ± 4 ms, while photoresponses to localized light reach the peak in 480 ± 40 ms, 560 ± 20 ms and 620 ± 80 ms at the base, middle and tip respectively (see Figure 9.4(f) for the summary).
- ⊙ **SATURATION TIME:** referring to the literature [191], we define saturation time (denoted by T_S) as the period in which all CNG channels remain closed, i.e. in which the photoresponse is saturated. Using diffuse saturating light of increasing intensity, T_S rises almost linearly, but not with localized illuminations (see Figure 9.4g).

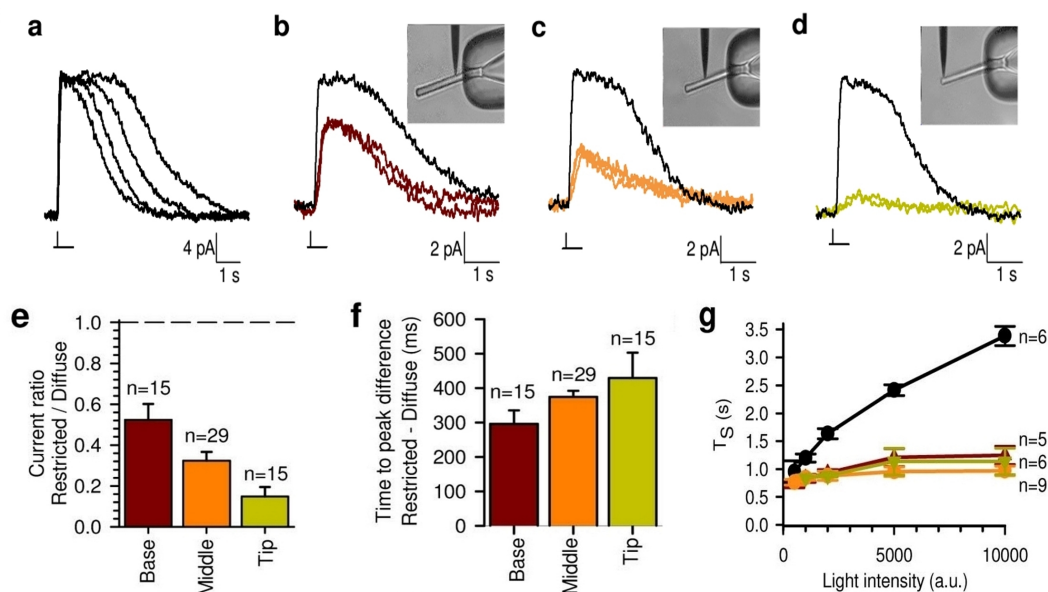


Figure 9.4: **Saturating illumination at different positions along the OS.** (a): photoresponses to saturating diffuse light at increasing intensity; (b, c, d): photocurrents to saturating light (at increasing intensity) localized at the base (purple traces), at the middle (orange traces) and at the the tip (green traces). Diffuse light in black; (e): average amplitude of maximal photoresponse (fraction of the maximal current with respect to diffuse light); n , number of mediated experiments; (f): average of time to peaks. (g): relationship between light intensity (mainly saturating) and saturation time T_S [191].

Non-saturating light

Curves in Figure 9.5(a), (b) and (c), obtained with non-saturating pulses of light, exhibit different amplitude and shape. While the difference in the maximal amplitude A is expected, the variation of the slope V of the rising phase (i.e. the maximal slope of the sigmoidal curve which can describe the initial rising phase of the photocurrent) is a novel experimental result. In particular, increasing the light intensity, the slope of the rising phase initially increases and then reaches a plateau (which is higher in case of base illumination): this is reported in Figure 9.5(d). Nevertheless, the slope of the rising phase and the amplitude of photocurrent are linearly correlated, regardless to the position (base, middle or tip) of the illumination, see in panel (e) of Figure 9.5.

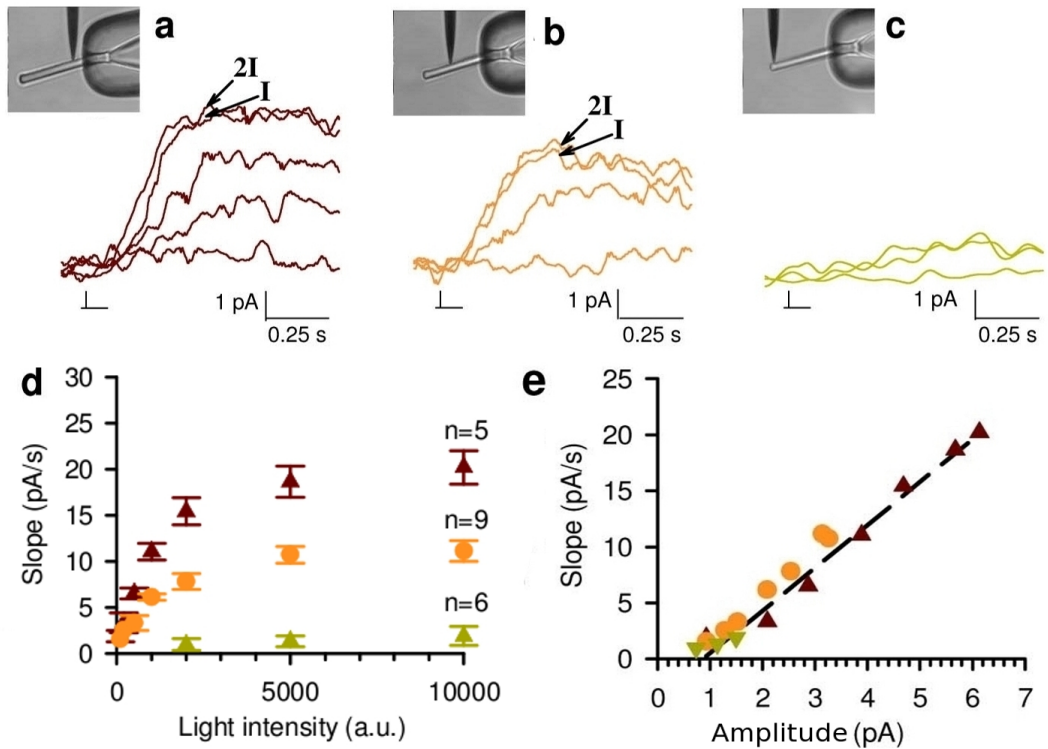


Figure 9.5: **Non-saturating illumination at different positions along the OS.** (a): family of photoresponses obtained with the TOF illuminating the base of the OS (purple traces); trace indicated by 2I was obtained by a flash of light with an intensity equal to twice that indicated by I. (b): as in (a) but with the TOF illuminating the middle of the OS and traces in orange; (c): as in (a) but with the TOF illuminating the tip of the OS and traces in green. (d): relationship between light intensity (in arbitrary unit) and slope V of the rising phase of photoresponses (same color code as panels a, b, and c); n , number of mediated experiments. (e): relationship between maximal amplitude A and slope V of the rising phase of the photoresponses.

Data interpretation and pathway simplification

We describe now the correlation between these experimental results and the steps of the phototransduction (described in Chapter 8). In order to do this, the complete signaling cascade is here reproduced as a network where 29 molecules are linked by 19 reactions (see Figure 9.6). In this pathway, the initial part of the cascade is colored in green (representing only the signaling molecules that are on the surface of the internal discs), whereas the second part is colored in red.

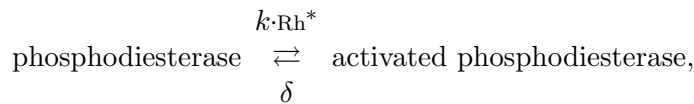
Because of the uniform density of the CNG channels along the OS [192, 168], the higher efficiency of the base with respect middle and tip cannot be justified by the signaling steps downstream of the cGMP hydrolysis. Moreover, since the rising phase of the photoresponses occurs before the intervention of the calcium feedbacks, the whole second part of the cascade (the red nodes in Figures 9.6) is ignored, meaning that it must be due to the part of the transduction depicted in green. The amplification step of this part involves three proteins: rhodopsin, G-protein and phosphodiesterase. Since the ratio of their concentrations is 250:25:1 (see Table 8.1) and the amplification ratio is 1:10:10, it is reasonable to hypothesize that saturation occurs for the less abundant protein, namely phosphodiesterase. In light of these considerations, our basic model describing the position-dependent TOF flash response includes only the activation-deactivation of phosphodiesterase, as reported in Figure 9.7. Furthermore, a progressive loss of phosphodiesterase in aged discs seems to be a good candidate for explaining the observed differences of efficiency of phototransduction between the base, middle and tip.

9.3.2 Modeling of the slope of the initial rising phase

We present here the mathematical model we use to interpret the slope V of the rising phase of the current responses at different light intensities and at different locations along the OS; through this approach, we infer new insight into the efficiency of phototransduction.

Derivation of the equation of the model

The part of the cascade we need to model is the loop reported Figure 9.7, i.e. the following single reversible reaction:



where k is the kinetic constant of phosphodiesterase activation by rhodopsin (excited by the localized photons and whose concentration is denoted by Rh^*) and δ is the kinetics of phosphodiesterase deactivation catalyzed by the RGS9 complex. The corresponding dynamic system is

$$\frac{dE^*}{dt} = k\text{Rh}^*(E_{\text{tot}} - E^*) - \delta E^*, \quad (9.1)$$

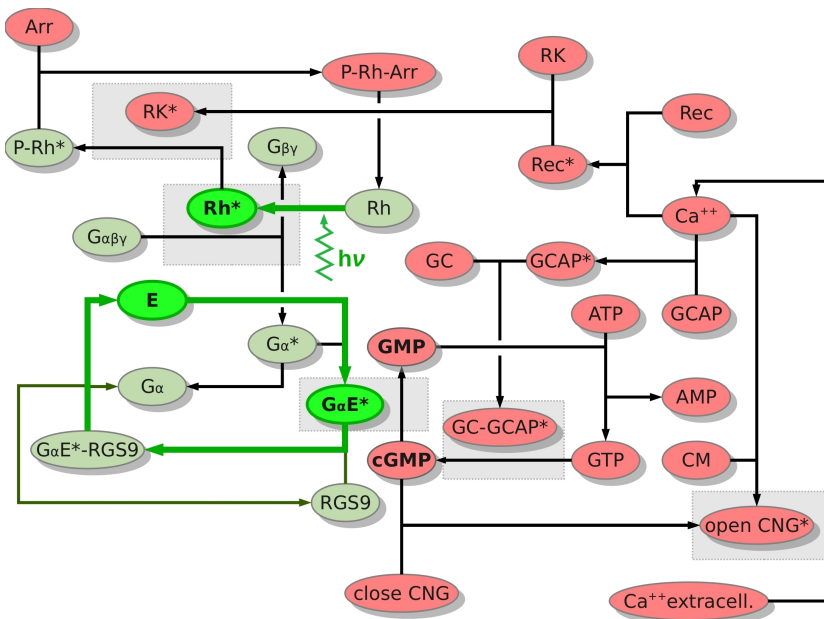


Figure 9.6: **Complete signaling pathway of the rod phototransduction.** The scheme of the cascade reported in Figure 8.3 can be described by this block diagram. Symbols refer to Table 8.1. In this representation, green nodes represent the molecules involved in the sequence of reactions that occur on the surface of the internal disc and that are responsible of the first amplification of the signal. Red nodes refer to the cytoplasm processes and the calcium feedbacks. Grey boxes include both the enzyme (an activated protein, marked with an asterisk) and the corresponding catalyzed reaction. Thick green nodes and green arrows indicate the simplifications that lead to the simplified scheme of Figure 9.7.

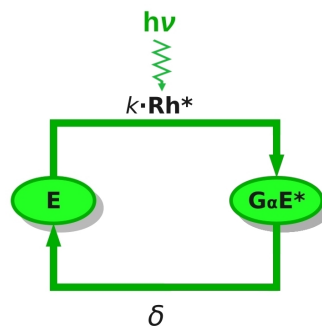


Figure 9.7: **Simplified scheme of phototransduction.** The scheme contains the basic ingredients we identified as responsible for the rising phase and for the gradient of efficiency along the OS (from base to tip). The parameters reported in this scheme are used in the mathematical model.

where E_{tot} is the total amount of phosphodiesterase on the discs where the localized illumination is absorbed, and E^* denotes the amount of activated phosphodiesterase. Thus, for a given Rh^* , the maximal amount of activated phosphodiesterase (for which $dE^*/dt = 0$) is

$$E_{\text{max}}^* = E_{\text{tot}} \frac{\text{Rh}^*}{\text{Rh}^* + \delta/k}. \quad (9.2)$$

The initial rising phase of the photoresponse is linked to the closure of CNG channels caused by the hydrolysis of cGMP. Since the rate of such hydrolysis is proportional to the concentration of activated phosphodiesterase,

$$\frac{d[\text{cGMP}]}{dt} = k_h[\text{cGMP}][E^*],$$

the maximal slope V of the rising phase observed in the current can be assumed to be proportional to E_{max}^* :

$$V \propto E_{\text{max}}^*.$$

Because of the high concentration of rhodopsin, also Rh^* can be assumed to be proportional to the light intensity I (i.e. $\text{Rh}^* = \alpha I$). Substituting these relations in (9.2), we obtained

$$V = V_{\text{max}} \frac{I}{I + K_s}, \quad (9.3)$$

where $V_{\text{max}} \propto E_{\text{tot}}$ and $K_s = \delta/k\alpha$. This expression provides the dependence of V from the light intensity I ; the parameters V_{max} and K_s are instead necessary to characterize the gradient of efficiency along the OS.

Fit of the experimental results

By using (9.3) to fit the relationship between the slope of the rising phase of the photocurrent and the light intensity (previously reported in panel (d) of Figure 9.5), we can obtain the values of V_{max} and of K_s at different positions along the OS (base, middle and tip). These fitting curves are reported in Figure 9.8(a) and the values of the fitting parameters are listed in Table 9.1.

We may notice that base, middle and tip exhibit the same values of K_s . From $K_s = \delta/k\alpha$, this means that the kinetics of the signaling cascade is the same regardless of the age of the discs. On the contrary, V_{max} is strongly dependent on the position; since it is related to the total amount of phosphodiesterase present in the illuminated discs, the value of V_{max} provides an estimate of the loss of this protein during the ageing of the internal discs.

Table 9.1: Fitting parameters for different positions along the OS.

| illuminated position | V_{max} (pA/s) | K_s (arbitrary units) |
|----------------------|-------------------------|-------------------------|
| Base | 27.7 ± 3.1 | 1080 ± 80 |
| Middle | 12.6 ± 3.7 | 1100 ± 170 |
| Tip | 2.5 ± 2.2 | 1100 ± 500 |

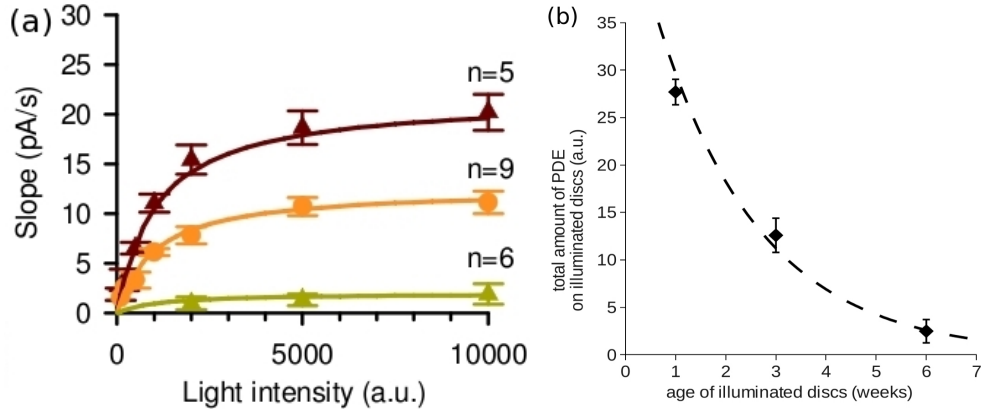


Figure 9.8: **Model results.** (a) Fitting of the slope of the rising phase by the mathematical model; experimental data of Figure 9.5(d) and fit by Equation (9.3). (b): Loss of phosphodiesterases: decay of E_{tot} in time and its fit with model (9.4) using $\gamma = 0.48 \text{ week}^{-1}$.

Rate of loss of phosphodiesterase

Although our experiments have only three locations (i.e. only three ages of the discs of the rod from *Xenopus laevis*, corresponding in particular to 1, 3 and 6 weeks approximately [178]) we can perform a rough estimation of the rate of loss of phosphodiesterase of the discs. In particular, the simple decay model can be used

$$dE_{\text{tot}}/dt = \gamma E_{\text{tot}}, \quad (9.4)$$

where γ is the characteristic time constant of the process of loss. Since $V_{\text{max}} \propto E_{\text{tot}}$, it can be rewritten as

$$dV_{\text{max}}/dt = \gamma V_{\text{max}}. \quad (9.5)$$

The fitting of the data (see Figure 9.8(b)) provides a value of $\gamma = 0.48 \text{ week}^{-1}$, meaning that a disc loses approximately 35% of its phosphodiesterase within one week. Electrophysiological experiments cannot provide validation of this model. Specific biological assays are needed (and not yet available).

9.3.3 Single photon experiments

In this section we show that our model is valid also in case of extremely low light intensity. In particular, we investigate the very dim photoresponses evoked by a single photon absorption.

Experimental results

Single photon responses are obtained by a series of N repetitions of dim flashes of light localized at the three different positions of the OS (base, middle and tip; $N=100, 50$ and 100 , respectively). With such a weak illumination, a response can be a failure (no absorption), a single or a few photoisomerations. The photocurrent measured in some repetitions are reported in panels (a), (b) and (c) of Figure 9.9.

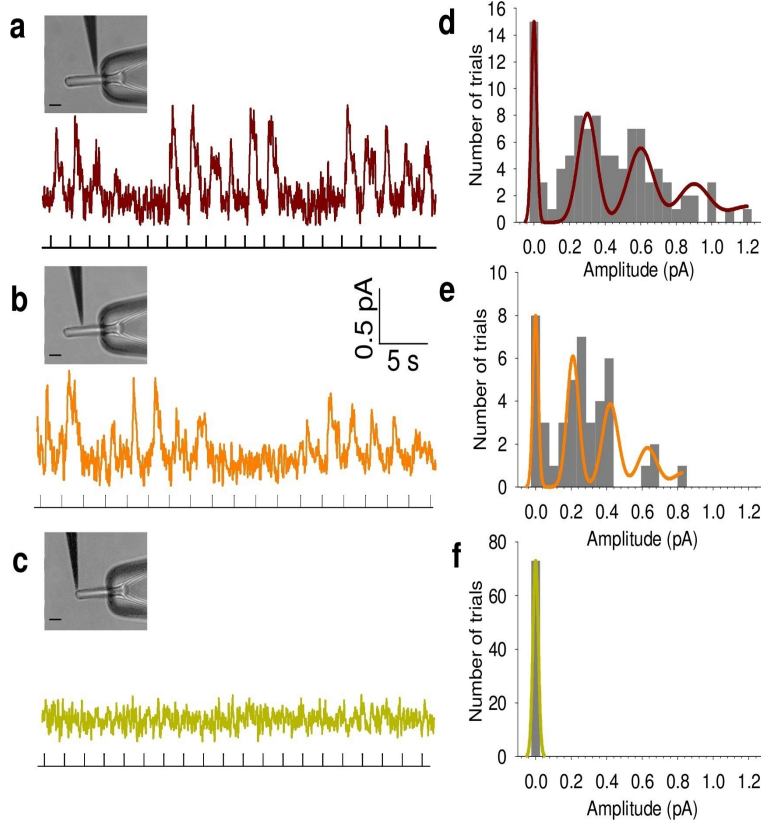


Figure 9.9: **Single photon responses.** (a,b,c): Photoresponses evoked by a series of repetitions of dim flashes of light to the base, middle and tip position of the OS (N is equal to 100, 50 and 100, respectively); (d,e,f): corresponding histograms of the amplitude of N evoked photoresponses fitted by (9.6) as explained in the text.

For each position of the TOF, we first estimate the maximal amplitudes A of the N photocurrents we have obtained (according to the procedure used in [193]). The histogram of these N values is fitted with the equation obtained from [194]:

$$p(A) = \sum_{k=0}^{\infty} \left(\frac{e^{-m} m^k}{k!} \right) \left[\frac{1}{2\pi(\sigma_0^2 + k\sigma_1^2)} \right]^{1/2} e^{-\frac{(A-ak)^2}{2\pi(\sigma_0^2 + k\sigma_1^2)}} \quad (9.6)$$

(see panels (d), (e) and (f) of Figure 9.9). The expected distribution of the observed amplitudes is the sum of the Gaussian components, where each component has a mean ka , variance $(\sigma_0^2 + k\sigma_1^2)$, and with an area re-scaled according to the Poisson equation $(e^{-m} m^k)/k!$ (where m is the mean number of photoisomerizations per flash). From the values of a , σ_0 and σ_1 used to provide the best fit of the histograms, we determine the single photon response to be 0.30 ± 0.11 pA for the base, 0.21 ± 0.12 pA for the middle and 0 pA (no response) for the tip.

Theoretical calculation

A consistent interpretation of these results can be provided by the proposed model. Using Equation (9.3) at a generic light intensity I , we calculate,

$$\frac{V^{\text{middle}}}{V^{\text{base}}} = \frac{V_{\text{max}}^{\text{middle}}}{V_{\text{max}}^{\text{base}}} \frac{I + K_s^{\text{base}}}{I + K_s^{\text{middle}}}.$$

For extremely small light intensity I (i.e. for a single photon absorption), this expression simplifies to

$$\frac{V^{\text{middle}}}{V^{\text{base}}} \simeq \frac{V_{\text{max}}^{\text{middle}} K_s^{\text{base}}}{V_{\text{max}}^{\text{base}} K_s^{\text{middle}}}.$$

Because of the linear correlation between the slope V and current amplitude A (Figure 9.8e), this ratio is equal to the ratio:

$$\frac{A^{\text{middle}}}{A^{\text{base}}} = \frac{V_{\text{max}}^{\text{middle}} K_s^{\text{base}}}{V_{\text{max}}^{\text{base}} K_s^{\text{middle}}}.$$

Using the values reported in Table 9.1 (and calculating the same expression for the ratio $A^{\text{tip}}/A^{\text{base}}$), we obtain

$$\begin{aligned} \frac{A^{\text{middle}}}{A^{\text{base}}} &= 0.5 \pm 0.2; \\ \frac{A^{\text{tip}}}{A^{\text{base}}} &= 0.0 \pm 0.1. \end{aligned}$$

According to this ratios, the theoretical single-photon responses at the base, middle and tip are 0.33, 0.17 and 0.01 pA, respectively. These values are sufficiently consistent with the experimental data mentioned above (0.30 ± 0.11 , 0.21 ± 0.12 and 0.0 pA). This agreement suggests that our model, although very simplified, is valid in a wide range of light intensities (from single photon up to saturating light).

9.4 Estimation of the diffusion coefficient of cGMP

In case of a saturating illumination at the base (i.e. when many phosphodiesterases are activated), the photocurrent shows a delay in reaching its maximal amplitude. Because of the geometry of the cytoplasm, such delay is presumably caused by the diffusion of the second messenger cGMP along the OS. Indeed, phosphodiesterase proteins (activated in the illuminated discs of the base) initially hydrolyze cGMP between the illuminated discs. This creates a gradient of concentration of cGMP along the OS, inducing the cGMP molecules present in the non-illuminated part of the OS to migrate towards the base (see Figure 9.10a).

By quantifying the correlation between this migration and the photocurrent delay, we can estimate the diffusion coefficient of cGMP. In particular we have developed a parameter-free model based on the geometry of the cytoplasm of a rod photoreceptor from *Xenopus laevis* (see Figure 9.10b).

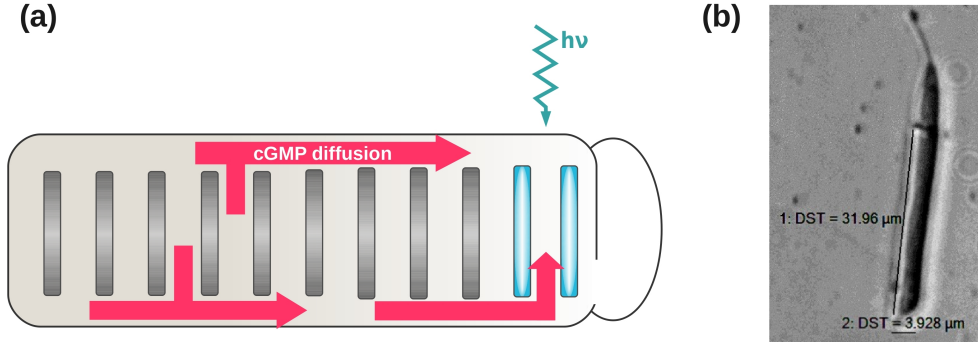


Figure 9.10: **Diffusion of cGMP inside the OS.** (a) Diffusion takes place in OS cytoplasm, i.e. through the space between discs and external membrane and in the gaps between adjacent discs; (b): measurement of length and diameter of a typical rod photoreceptor from *Xenopus laevis*.

9.4.1 A parameter-free model

Cyclic GMP diffuses in the cytoplasm of the OS; this domain is a cylinder with length $L = 32 \mu\text{m}$ and diameter $d = 4 \mu\text{m}$. Thanks to the radial symmetry, it can be reduced to a rectangle of dimension $L \times d/2$. Of course, the volume occupied by the internal discs is excluded. Our model describes the dynamics of the cGMP concentration (hereafter denoted g) along this domain during the initial part of the photoresponse (up to the peak time, i.e. 500 ms). The initial condition is a uniform concentration of cGMP equal to $3.5 \mu\text{M}$ [195]:

$$g(x, r, t = 0) = 3.5 \mu\text{M} \quad \forall (x, r) \in [0, 32 \mu\text{m}] \times [0, 2 \mu\text{m}].$$

The partial differential equation of cGMP dynamics is the Fick's law with the Laplacian expressed in cylindrical coordinates:

$$\frac{\partial g(x, r, t)}{\partial t} = D_{\text{cGMP}} \left[\frac{1}{r} \frac{\partial}{\partial r} \left(r \frac{\partial g(x, r, t)}{\partial r} \right) + \frac{\partial^2 g(x, r, t)}{\partial x^2} \right], \quad (9.7)$$

where D_{cGMP} is the diffusion coefficient of cGMP.

In case of localized illumination of the base of the OS, three boundary conditions are defined:

- ⊙ As reported in Figure 9.2, TOF illuminates about $5 \mu\text{m}$ of the base of the OS (from $x = 0$ to $x = 5 \mu\text{m}$). In these points, the activated phosphodiesterase hydrolyzes cGMP. Because of the high efficiency of the phototransduction at the base and because we are referring to saturating flashes, cGMP hydrolysis is very fast; the chemical reaction can be considered instantaneous and written as a Dirichlet condition:

$$g(x, r, t) = 0, \quad r = 0 \mu\text{m}, x \leq 5 \mu\text{m}. \quad (9.8)$$

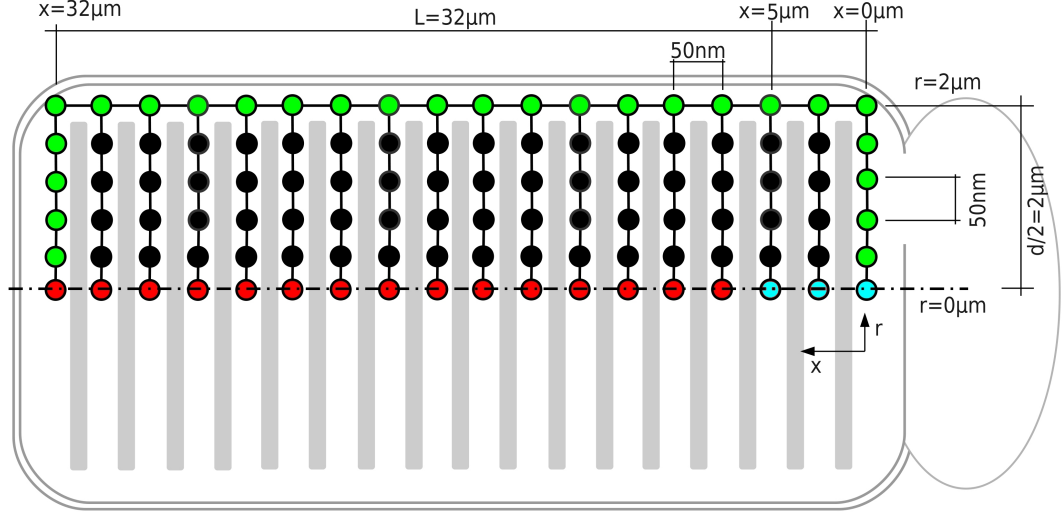


Figure 9.11: **Numerical grid for the discretization of the cytoplasm of the OS.** Black nodes are internal points for which (9.7) is applied, whereas color nodes have boundary conditions. In particular: blue nodes indicate phosphodiesterase activity (Dirichlet conditions given by (9.8)); red nodes refer to symmetric reflective boundary conditions (9.9); green nodes have reflective conditions because of the external membrane of the OS (9.10).

- ⊙ Due to symmetry reason, a reflective boundary conditions is used along the remaining part of the symmetry axis of the cylinder where the phosphodiesterases are not activated:

$$D_{cGMP} \frac{\partial g(x, r, t)}{\partial r} = 0, \quad r = 0 \mu\text{m}, x > 5 \mu\text{m}. \quad (9.9)$$

- ⊙ Along the three remaining sides of the rectangle, no migration can take place and then similar reflective conditions are used:

$$\begin{aligned} D_{cGMP} \frac{\partial g(x, r, t)}{\partial r} &= 0, & r &= 2 \mu\text{m}; \\ D_{cGMP} \frac{\partial g(x, r, t)}{\partial x} &= 0, & x &= 0 \mu\text{m}; \\ D_{cGMP} \frac{\partial g(x, r, t)}{\partial x} &= 0, & x &= 32 \mu\text{m}. \end{aligned} \quad (9.10)$$

In order to perform a numerical integration of the differential equation, the domain is discretized as a sequence of concatenated nodes at distance 50 nm from each other (the same distance between adjacent discs). Figure 9.11 reports the resulting grid.

The equation is integrated numerically for different values of the diffusion coefficient of cGMP ranging from 50 to 1000 $\mu\text{m}^2/\text{s}$. Each integration provides the drop

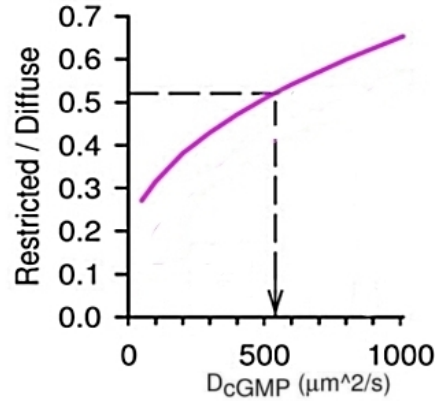


Figure 9.12: **Numerical integration of the model.** Theoretical ratio between localized and diffused saturating photoresponse (at $t = 500$ ms) *vs* value of D_{cGMP} in the case of illumination at the base of the OS. In order to reproduce the localized/diffuse ratio equal to 0.52 ± 0.08 , D_{cGMP} must assume a value of $550 \pm 50 \mu\text{m}^2/\text{s}$.

of concentration of cGMP along the whole domain. The ratio between this drop and the total amount of cGMP is the photocurrent ratio between localized and diffuse saturating light. The result is reported in Figure 9.12: considering that the photoresponse measured in case of base illumination reaches $52 \pm 8\%$ of the dark current (from Figure 9.4(e)), this model provides the estimation $D_{\text{cGMP}} = 500 \div 600 \mu\text{m}^2/\text{s}$. It is worth noting that such range may represent a lower bound: indeed, without the negative effect due to the calcium feedback, we overestimate the amplitude of the photoresponse and, because of the monotonicity of the curve in Figure 9.12, we underestimate the diffusion. In any case, our result is higher than the estimations available in the current literature, which range from 250 to 450 $\mu\text{m}^2/\text{s}$ (after correction of the “longitudinal diffusivity” [189, 169] by the factor of 6.5 [170]).

9.5 Conclusions

Rod photoreceptors and their signaling cascade evoked by light are well characterized: abundant information and results have been collected in many years of research. However, some details are still missing and, in some cases, various hypotheses have been suggested. For instance, a few of them concern the ageing of the discs (i.e. the gradient of efficiency of phototransduction along the OS) and the role of diffusion of the second messenger in the generation of the photocurrent. For this reason, the laboratory of prof. Vincent Torre has setup a new experimental technique based on the use of optical fibers and thus capable of generating a very localized illumination. Through such source of light, the efficiency of the signaling pathway along the OS (from base to tip) has been measured at different intensities of light, from very dim up to saturating flashes (see Figures 9.4 and 9.5). The maximal amplitude and the slope of the rising phase of the obtained photocurrents show a previously unobserved correlation with respect to the localized light intensity.

Starting from these results, our contribution has been the development of a mathematical model in order to justify the correlation between the different properties of the photocurrent. After a simplification of the signaling cascade, we have obtained an equation which reproduces the experimental data, which suggests that the gradient of efficiency can be attributed to the loss of phosphodiesterase proteins due to the ageing of the internal discs (see Figure 9.8). It is worth noting that, in spite of its simplified structure, our model is able to describe the efficiency at any light intensity, from saturating illumination down to single photon absorption.

The use of localized illumination has been utilized also to evaluate the mobility of cGMP in the cytoplasm. Using a parameter-free model applied to the more efficient situation (the illumination of the base of the OS), we have estimated the diffusion coefficient of cGMP to be about $550 \mu\text{m}^2/\text{s}$ (see Figure 9.12). Our result is higher than those reported in literature. However, as a result of this high diffusion coefficient, CNG channels near the tip of the OS can be closed by diffuse flashes of light, despite the very low level of phosphodiesterase proteins in the discs in that region.

Bibliography

- [1] A. Bordbar, A.M. Feist, R. Usaité-Black, J. Woodcock, B.O. Palsson, and I. Famili. A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. *BMC systems biology*, 5(1):180, 2011.
- [2] O. Folger, L. Jerby, C. Frezza, E. Gottlieb, E. Ruppín, and T. Shlomi. Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol*, 7, 2011.
- [3] E.S. Snitkin, D. Segrè, and T.F.C. Mackay. Epistatic interaction maps relative to multiple metabolic phenotypes. *PLoS Genetics*, 7(2), 2011.
- [4] S.P. Cornelius, J.S. Lee, and A.E. Motter. Dispensability of *Escherichia coli*'s latent pathways. *Proceedings of the National Academy of Sciences*, 108(8):3124, 2011.
- [5] Claude Berge and Edward Minieka. *Graphs and hypergraphs*, volume 7. North-Holland publishing company Amsterdam, 1973.
- [6] Steffen Klamt, Utz-Uwe Haus, and Fabian Theis. Hypergraphs and cellular networks. *PLoS computational biology*, 5(5):e1000385, 2009.
- [7] Markus W Covert, Christophe H Schilling, Iman Famili, Jeremy S Edwards, Igor I Goryanin, Evgeni Selkov, and Bernhard O Palsson. Metabolic modeling of microbial strains in silico. *Trends in biochemical sciences*, 26(3):179–186, 2001.
- [8] I. Thiele, T.D. Vo, N.D. Price, and B.O. Palsson. Expanded metabolic reconstruction of *Helicobacter pylori* (iT341 GSM/GPR): an in silico genome-scale characterization of single- and double-deletion mutants. *J Bacteriol.*, 187(16):5818–5830, 2005.
- [9] S.A. Becker and B.O. Palsson. Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiology*, 5 (8), 2005.
- [10] A.M. Feist, Palsson B.O. Scholten, J.C.M., F.J. Brockman, and T. Ideker. Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Molecular Systems Biology*, 2:doi:10.1038/msb4100046, 2006.
- [11] G.E. Pinchuk, E.A. Hill, J. De Ingeniis, X. Zhang, Osterman, and *et. al.* Constraint-based model of *Shewanella oneidensis* MR-1 metabolism: A tool for data analysis and hypothesis generation. *PLoS Comput Biol*, 6:e1000822, 2010.
- [12] N. Jamshidi and B.O. Palsson. Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the in silico strain inj661 and proposing alternative drug targets. *BMC Systems Biology*, 1 (26), 2007.
- [13] J. Forster, I. Famili, P. Fu, and B.O. J. Palsson. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res*, 13:244–253, 2003.

- [14] J. S. Edwards and B. O. Palsson. The *escherichia coli* mg1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA*, 97(10):5528–5533, 2000.
- [15] A. Raghunathan, J. Reed, S. Shin, B.O. Palsson, and S. Daefer. Constraint-based analysis of metabolic capacity of *Salmonella typhimurium* during host-pathogen interaction. *BMC Systems Biology*, 3 (38), 2009.
- [16] N.C. Duarte, S.A. Becker, N. Jamshidi, I. Thiele, and *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *PNAS*, 104 (6):1777–1782, 2007.
- [17] K. Raman and N. Chandra. Flux balance analysis of biological systems: applications and challenges. *Brief Bioinform*, 10 (4):435–449, 2009.
- [18] B.O. Palsson. *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, 2006.
- [19] Uri Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits.*, 2007, Chapman and Hall/CRC.
- [20] B.O. Palsson and A. Varma. Metabolic capabilities of *Escherichia coli* II: optimal growth pattern. *J Theor Biol*, 165:503–522, 1993.
- [21] D. Segré, D. Vitkup, and G.M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA*, 99 (23):15112–15117, 2002.
- [22] T. Shlomi, O. Berkman, and E. Ruppin. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci USA*, 102 (21):7695–7700, 2005.
- [23] A.P. Burgard and C.D. Maranas. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnology and bioengineering*, 82(6):670–677, 2003.
- [24] A.P. Burgard, P. Pharkya, and C.D. Maranas. OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng*, 84 (6):647–657, 2003.
- [25] John L Ingraham, Ole Maaløe, and Frederick Carl Neidhardt. *Growth of the bacterial cell*, volume 3. Sinauer Associates Sunderland, Massachusetts, 1983.
- [26] J. Matoušek and B. Gärtner. *Understanding and Using Linear Programming*. Springer, Berlin, 2000.
- [27] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, New York, 1986.
- [28] Frauke Liers, Michael Jünger, Gerhard Reinelt, and Giovanni Rinaldi. Computing exact ground states of hard ising spin glass problems by branch-and-cut. In A. K. Hartmann and H. Rieger, editors, *New Optimization Algorithms in Physics*, pages 47–69. Wiley-VCH Verlag GmbH, 2005.
- [29] Franz Rendl, Giovanni Rinaldi, and Angelika Wiegele. Solving Max-Cut to optimality by intersecting semidefinite and polyhedral relaxations. *Math. Programming*, 121(2):307, 2010.

- [30] J. Lehar, A. Krueger, G. Zimmermann, and A. Borisy. High-order combination effect and biological robustness. *Mol Syst Biol*, 4:doi:10.1038, 2008.
- [31] T.C. Chou. Theoretical basis, experimental design, and computerized simulation of synergism and antagonism in drug combination studies. *Pharmacol Rev*, 58 (3):621–681, 2006.
- [32] P. Csermely, V. Agoston, and S. Pongor. The efficiency of multi-target drugs: the network approach might help drug design. *Trends Pharmacol Sci*, 26:178–182, 2005.
- [33] P. Yeh, A.I. Tschumi, and R. Kishony. Functional classification of drugs by properties of their pairwise interaction. *Nature Gen*, 38 (4):489–494, 2006.
- [34] C.T. Keith, A.A. Borisy, and B.R. Stockwell. Multicomponent therapeutics for networked systems. *Nat Rev Drug Discov*, 4:1–8, 2005.
- [35] P. Yeh and R. Kishony. Networks from drug-drug surface. *Mol Syst Biol*, 3:85–87, 2007.
- [36] P.F. Suthers, A. Zomorodi, and C.D. Maranas. Genome-scale gene/reaction essentiality and synthetic lethality analysis. *Mol Syst Biol*, 5:301, 2005.
- [37] D. Segré, A. De Luna, G.M. Church, and R. Kishony. Modular epistasis in yeast metabolism. *Nat Genet*, 37 (1):77–83, 2005.
- [38] P. Sridhar, B. Song, T. Kahveci, and S. Ranka. Mining metabolic networks for optimal drug targets. *Pac Symp Biocomput*, 13:291–302, 2008.
- [39] W.G. Kaelin. The concept of synthetic lethality in the context of anticancer therapy. *Nat Rev Cancer*, 5:689–698, 2005.
- [40] K. Goh, M.E. Cusick, D. Valle, B. Childs, M. Vidal, and A.L. Barabási. The human disease network. *Proc Natl Acad Sci USA*, 104(21):8685–8690, 2007.
- [41] E. Ferrannini. Physiological and metabolic consequences of obesity. *Metabolism*, 44 (3):15–17, 1995.
- [42] A.J. Salter. Overview. trimethoprim-sulfamethoxazole: An assessment of more than 12 years of use. *Reviews of infectious diseases*, pages 196–236, 1982.
- [43] O. Weinreb, T. Amit, O. Bar-Am, and M.B.H. Youdim. Rasagiline: a novel anti-parkinsonian monoamine oxidase-B inhibitor with neuroprotective activity. *Progress in neurobiology*, 92(3):330–344, 2010.
- [44] J. Kim and J.L. Reed. OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst Biol*, 4:53, 2010.
- [45] M. Imielinski and B. Belta. Deep epistasis in human metabolism. *Chaos*, 20(2):026104, 2010.
- [46] S. Dempe. *Foundations of bilevel programming*. Kluwer Academic Publisher, New York, 2010.
- [47] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, et al. Drugbank 3.0: a comprehensive resource for omics research on drugs. *Nucleic acids research*, 39(suppl 1):D1035–D1041, 2011.

- [48] M. Mansour, D. Schwartz, R. Judd, B. Akingbemi, T. Braden, E. Morrison, J. Dennis, F. Bartol, A. Hazi, I. Napier, and A.B. Abdel-Mageed. Thiazolidinediones/PPAR γ agonists and fatty acid synthase inhibitors as an experimental combination therapy for prostate cancer. *International Journal of Oncology*, doi: 10.3892/ijo.2010.877, 2010.
- [49] C. Luberto and Y.A. Hannun. Ceramide in the eukaryotic stress response. *Trends Cell Biol*, 10(2):73–80, 2000.
- [50] L.M. Obeid and Y.A. Hannun. The ceramide-centric universe of lipid-mediated cell regulation: stress encounters of the lipid kind. *J Biol Chem*, 277:25847–25850, 2002.
- [51] G. Yang, L. Badeanlou, J. Bielawski, A.J. Roberts, Y.A. Hannun, and F. Samad. Central role of ceramide biosynthesis in body weight regulation, energy metabolism, and the metabolic syndrome. *Am J Physiol Endocrinol Metab*, 297(1):E211–E224, 2009.
- [52] G. Triola, G. Fabrias, M. Dragusin, L. Niederhausen, R. Broere, and *et al.* Specificity of the dihydroceramide desaturase inhibitor N-[(1R,2S)-2-Hydroxy-1-hydroxymethyl-2-(2-tridecyl-1-cyclopropenyl)ethyl]octanamide (GT11) in primary cultured cerebellar neurons. *Mol Pharmacol*, 66 (6):1671–1678, 2004.
- [53] J.A. Munoz-Olaya, X. Matabosch, C. Bedia, M. Egido-Gabs, J. Casas, and *et al.* Synthesis and biological activity of a novel inhibitor of dihydroceramide desaturase. *Chem Med Chem*, 3:946–953, 2009.
- [54] G. Llaverias, D. Lacasa, M. Viñals, M. Vázquez-Carrera, R.M. Sánchez, J.C. Laguna, and M. Alegret. Reduction of intracellular cholesterol accumulation in THP-1 macrophages by a combination of rosiglitazone and atorvastatin. *Biochemical pharmacology*, 68(1):155–163, 2004.
- [55] M.I. Freed, R. Ratner, S.M. Marcovina, M.M. Kreider, N. Biswas, B.R. Cohen, and J.D. Brunzell. Effects of rosiglitazone alone and in combination with atorvastatin on the metabolic abnormalities in type 2 diabetes mellitus. *The American journal of cardiology*, 90(9):947–952, 2002.
- [56] G. De Berardis, M. Sacco, V. Evangelista, A. Filippi, C.B. Giorda, G. Tognoni, U. Valentini, A. Nicolucci, *et al.* Aspirin and simvastatin combination for cardiovascular events prevention trial in diabetes (ACCEPT-D): design of a randomized study of the efficacy of low-dose aspirin in the prevention of cardiovascular events in subjects with diabetes mellitus treated with statins. *Trials*, 8(1):21, 2007.
- [57] P.W. Armstrong, B.H. Lorell, S. Nissen, and J. Borer. 97th meeting of the cardiovascular and renal drugs advisory committee July 18th and 19th, 2002. *Circulation*, 106(6):e9011–e9012, 2002.
- [58] T.J. Murtola, P. Pennanen, H. Syväälä, M. Bläuer, T. Ylikomi, and T.L.J. Tammela. Effects of simvastatin, acetylsalicylic acid, and rosiglitazone on proliferation of normal and cancerous prostate epithelial cells at therapeutic concentrations. *The Prostate*, 69(9):1017–1023, 2009.
- [59] G.R. Weiss, MB Garnick, RT Osteen, GD Steele Jr, RE Wilson, D. Schade, WD Kaplan, LM Boxt, K. Kandarpa, and RJ Mayer. Long-term hepatic arterial infusion of 5-fluorodeoxyuridine for liver metastases using an implantable infusion pump. *Journal of Clinical Oncology*, 1(5):337–344, 1983.

- [60] SB Carter, TJ Franklin, DF Jones, BJ Leonard, SD Mills, RW Turner, and WB Turner. Mycophenolic acid: an anti-cancer compound with unusual properties. *Nature*, 223:848–850, 1969.
- [61] H.L. Bleich, E.S. Boro, E. Frei III, N. Jaffe, M.H.N. Tattersall, S. Pitman, and L. Parker. New approaches to cancer chemotherapy with methotrexate. *New England Journal of Medicine*, 292(16):846–851, 1975.
- [62] M. Hazarika, R.M. White, J.R. Johnson, and R. Pazdur. FDA drug approval summaries: Pemetrexed (Alimta®). *The oncologist*, 9(5):482–488, 2004.
- [63] K.L.B. Borden and B. Culjkovic-Kraljacic. Ribavirin as an anti-cancer therapy: acute myeloid leukemia and beyond? *Leukemia & lymphoma*, 51(10):1805, 2010.
- [64] S. Lam, A. McWilliams, C. MacAulay, L. Wattenberg, E. Szabo, et al. A phase i study of myo-inositol for lung cancer chemoprevention. *Cancer Epidemiology Biomarkers & Prevention*, 15(8):1526–1531, 2006.
- [65] S.J. Cho, J.S. Kim, J.M. Kim, J.Y. Lee, H.C. Jung, and I.S. Song. Simvastatin induces apoptosis in human colon cancer cells and in tumor xenografts, and attenuates colitis-associated colon cancer in mice. *International Journal of Cancer*, 123(4):951–957, 2008.
- [66] P. Baumann, S. Mandl-Weber, A. Völkl, C. Adam, I. Bumeder, F. Oduncu, and R. Schmidmaier. Dihydroorotate dehydrogenase inhibitor A771726 (leflunomide) induces apoptosis and diminishes proliferation of multiple myeloma cells. *Molecular cancer therapeutics*, 8(2):366, 2009.
- [67] G. Zhang, C. Tu, G. Zhang, G. Zhou, and W. Zheng. Indomethacin induces apoptosis and inhibits proliferation in chronic myeloid leukemia cells. *Leukemia research*, 24(5):385–392, 2000.
- [68] U.M.H. Schrell, M.G. Rittig, M. Anders, F. Kiesewetter, R. Marschalek, U.H. Koch, and R. Fahlbusch. Hydroxyurea for treatment of unresectable and recurrent meningiomas. Inhibition of primary human meningioma cells in culture and in meningioma transplants by induction of the apoptotic pathway. *Journal of neurosurgery*, 86(5):845–852, 1997.
- [69] K.H. Antman. Introduction: the history of arsenic trioxide in cancer therapy. *The oncologist*, 6(Supplement 2):1–2, 2001.
- [70] H. von der Maase, SW Hansen, JT Roberts, L. Dogliotti, T. Oliver, MJ Moore, I. Boddagi, P. Albers, A. Knuth, CM Lippert, et al. Gemcitabine and cisplatin versus methotrexate, vinblastine, doxorubicin, and cisplatin in advanced or metastatic bladder cancer: results of a large, randomized, multinational, multicenter, phase iii study. *Journal of Clinical Oncology*, 18(17):3068–3077, 2000.
- [71] C. Temple Jr, L.L. Bennett Jr, J.D. Rose, R.D. Elliott, and J.H. Montgomery. Synthesis of pseudocofactor analogs as potential inhibitors of the folate enzymes. *Journal of Medicinal Chemistry*, 25(2):161–166, 1982.
- [72] M. Mahfouz, S. Johnson, and RT Holman. Inhibition of desaturation of palmitic, linoleic and eicosa-8, 11, 14-trienoic acids in vitro by isomeric cis-octadecenoic acids. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism*, 663(1):58–68, 1981.

- [73] T. Borges, HP Glauert, and LW Robertson. Perfluorodecanoic acid noncompetitively inhibits the peroxisomal enzymes enoyl-coa hydratase and 3-hydroxyacyl-coa dehydrogenase. *Toxicology and applied pharmacology*, 118(1):8–15, 1993.
- [74] K.Y. Tserng, S.J. Jin, and C.L. Hoppel. Spiropentaneacetic acid as a specific inhibitor of medium-chain acyl-coa dehydrogenase. *Biochemistry*, 30(44):10755–10760, 1991.
- [75] D. Deutscher, I. Meilijson, M. Kupiec, and E. Ruppín. Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nature Gen.*, 38 (9):993–998, 2006.
- [76] V. Mazumdar, E.S. Snitkin, S. Amar, and D. Segré. Metabolic network model of a human oral pathogen. *J Bacteriol.*, 191 (1):74–90, 2009.
- [77] L.V. Hooper and J.I. Gordon. Commensal host–bacterial relationships in the gut. *Science*, 292:1115–1118, 2001.
- [78] H. Peng, J. Wen, H. Li, J. Chang, and X. Zhou. Drug inhibition profile prediction for *nfκb* pathway in multiple myeloma. *PLoS one*, 6(3):e14750, 2011.
- [79] A. Salvador. Synergism analysis of biochemical systems. i. conceptual framework. *Mathematical biosciences*, 163(2):105–129, 2000.
- [80] H.G. Holzhütter. The generalized flux-minimization method and its application to metabolic networks affected by enzyme deficiencies. *Biosystems*, 83(2):98–107, 2006.
- [81] Z. Li, R.S. Wang, and X.S. Zhang. Two-stage flux balance analysis of metabolic networks for drug target identification. *BMC systems biology*, 5(Suppl 1):S11, 2011.
- [82] R Mahadevan, CH Schilling, et al. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic engineering*, 5(4):264, 2003.
- [83] A.E. Motter, N. Gulbahce, E. Almaas, and A.L. Barabási. Predicting synthetic rescues in metabolic networks. *Molecular Systems Biology*, 4(1), 2008.
- [84] Brett A Boghigian, John Armando, Daniel Salas, and Blaine A Pfeifer. Computational identification of gene over-expression targets for metabolic engineering of taxadiene production. *Applied microbiology and biotechnology*, 93(5):2063–2073, 2012.
- [85] Evan S. Snitkin and Daniel Segrè. Epistatic interaction maps relative to multiple metabolic phenotypes. *PLoS genetics*, 7(2):e1001294, 2011.
- [86] S.A. Becker, A.M. Feist, M.L. Mo, G. Hannum, B.Ø. Palsson, and M.J. Herrgård. Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox. *Nature protocols*, 2(3):727–738, 2007.
- [87] J.D. Orth, R.M.T. Fleming, and B.O. Palsson. Reconstruction and use of microbial metabolic networks: the core Escherichia coli metabolic model as an educational guide. *EcoSal*, doi: 10.1128/ecosal.10.2.1, 2009.
- [88] Markus J Herrgård, Baek-Seok Lee, Vasilij Portnoy, and Bernhard Ø Palsson. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *saccharomyces cerevisiae*. *Genome research*, 16(5):627–635, 2006.
- [89] José P Faria, Ross Overbeek, Fangfang Xia, Miguel Rocha, Isabel Rocha, and Christopher S Henry. Genome-scale bacterial transcriptional regulatory networks: reconstruction and integrated analysis with metabolic models. *Briefings in bioinformatics*, 2013.

- [90] Mai Li, Pei Yee Ho, Shanjing Yao, and Kazuyuki Shimizu. Effect of *suca* or *succ* gene knockout on the metabolism in *escherichia coli* based on gene expressions, enzyme activities, intracellular metabolite concentrations and metabolic fluxes by ^{13}C -labeling experiments. *Biochemical engineering journal*, 30(3):286–296, 2006.
- [91] Dong-Eun Chang, Soan Shin, Joon-Shick Rhee, and Jae-Gu Pan. Acetate metabolism in a *pta* mutant of *escherichia coli* w3110: Importance of maintaining acetyl coenzyme a flux for growth and survival. *Journal of bacteriology*, 181(21):6656–6663, 1999.
- [92] S Flores, G Gosset, N Flores, AA De Graaf, and F Bolivar. Analysis of carbon metabolism in *Escherichia coli* strains with an inactive phosphotransferase system by ^{13}C labeling and NMR spectroscopy. *Metabolic engineering*, 4(2):124–137, 2002.
- [93] Stephen S Fong and Bernhard Ø Palsson. Metabolic gene-deletion strains of *escherichia coli* evolve to computationally predicted growth phenotypes. *Nature genetics*, 36(10):1056–1058, 2004.
- [94] Qiang Hua, Andrew R Joyce, Bernhard Ø Palsson, and Stephen S Fong. Metabolic characterization of *escherichia coli* strains adapted to growth on lactate. *Applied and environmental microbiology*, 73(14):4639–4647, 2007.
- [95] Kerson Huang. *Introduction to statistical physics*. CRC Press, 2001.
- [96] Julia M Yeomans. *Statistical mechanics of phase transitions*. Oxford University Press, 1992.
- [97] Kurt Binder and Dieter W Heermann. *Monte Carlo simulation in statistical physics: an introduction*. Springer, 2010.
- [98] G. Iacono, F. Ramezani, N. Soranzo, and C. Altafini. Determining the distance to monotonicity of a biological network: a graph-theoretical approach. *IET Systems Biology*, 4:223–235, 2010.
- [99] T. Zaslavsky. Signed graphs. *Discrete Appl. Math.*, 4(1):47–74, 1982.
- [100] G Toulouse. Theory of the frustration effect in spin glasses : I. *Communications on Physics*, 2:115, 1977.
- [101] Eduardo D. Sontag. Monotone and near-monotone biochemical networks. *Syst. Synth. Biol.*, 1(2):59–87, 2007.
- [102] Hal L. Smith. Systems of ordinary differential equations which generate an order preserving flow. A survey of results. *SIAM Rev.*, 30(1):87–113, 1988.
- [103] Eduardo Fradkin, B. A. Huberman, and Stephen H. Shenker. Gauge symmetries in random magnetic systems. *Phys. Rev. B*, 18(9):4789–4814, 1978.
- [104] M. E. J. Newman and G. T. Barkema. *Monte Carlo Methods in Statistical Physics*. Clarendon Press, Oxford, 1999.
- [105] S. Burer, R.D.C. Monteiro, and Y. Zhang. Rank-two relaxation heuristics for max-cut and other binary quadratic programs. *SIAM Journal on Optimization*, 12(2):503–521, 2002.
- [106] Clayton W. Commander. Maximum cut problem, max-cut. In Christodoulos A. Floudas and Panos M. Pardalos, editors, *Encyclopedia of Optimization*, pages 1991–1999. Springer US, 2009.

- [107] P. Festa, P.M. Pardalos, M.G.C. Resende, and C.C. Ribeiro. Randomized heuristics for the max-cut problem. *Optimization Methods and Software*, 17(6):1033–1058, 2002.
- [108] MX Goemans and DP Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, 1995.
- [109] Stefano Ciliberti and Marc Mézard. The theoretical capacity of the parity source coder. *J. Stat. Mech.*, page P10003, 2005.
- [110] A. Montanari and R. Urbanke. Modern coding theory: The statistical mechanics and computer science point of view. In *Les Houches Summer School on Mathematical Statistical Physics*. Elsevier, 2007.
- [111] J.S. Yedidia, W.T. Freeman, and Y. Weiss. Constructing Free-Energy Approximations and Generalized Belief Propagation Algorithms. *IEEE Tr. Inf. Th.*, 51(7):2282–2312, 2005.
- [112] Martin J. Wainwright, Elitza Maneva, and Emin Martinian. Lossy Source Compression Using Low-Density Generator Matrix Codes: Analysis and Algorithms. *IEEE Tr. Inf. Th.*, 56(3):1351–1368, 2010.
- [113] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.
- [114] Marc Mezard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, Inc., New York, NY, USA, 2009.
- [115] Richard Bonneau. Learning biological networks: from modules to dynamics. *Nat. Chem. Biol.*, 4(11):658–664, 2008.
- [116] M. Cosentino Lagomarsino, P. Jona, B. Bassetti, and H. Isambert. Hierarchy and feedback in the evolution of the escherichia coli transcription network. *Proceedings of the National Academy of Sciences*, 104(13):5516–5520, 2007.
- [117] Shmoolik Mangan and Uri Alon. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences*, 100(21):11980–11985, 2003.
- [118] Stefan C Materna and Eric H Davidson. Logic of gene regulatory networks. *Current Opinion in Biotechnology*, 18(4):351 – 354, 2007.
- [119] I Shmulevich, ER Dougherty, and W Mang. From Boolean to probabilistic Boolean networks as models of genetic regulatory networks. *Proc. of the IEEE*, 90(11):1778–1792, 2002.
- [120] Shai S Shen-Orr, Ron Milo, Shmoolik Mangan, and Uri Alon. Network motifs in the transcriptional regulation network of escherichia coli. *Nature genetics*, 31(1):64–68, 2002.
- [121] Bhaskar DasGupta, German Andres Enciso, Eduardo D. Sontag, and Yi Zhang. Algorithmic and complexity results for decompositions of biological networks into monotone subsystems. *Biosystems*, 90(1):161–178, 2007.
- [122] A. Ma’ayan, A. Lipshtat, R. Iyengar, and Eduardo D. Sontag. Proximity of intracellular regulatory networks to monotone systems. *IET Syst. Biol.*, 2(3):103–112, 2008.

- [123] Giovanni Iacono and Claudio Altafini. Monotonicity, frustration, and ordered response: an analysis of the energy landscape of perturbed large-scale biological networks. *BMC Syst. Biol.*, 4:83, 2010.
- [124] Heladia Salgado, Socorro Gama-Castro, Martin Peralta-Gil, Edgar Díaz-Peredo, Fabiola Sánchez-Solano, Alberto Santos-Zavaleta, Irma Martínez-Flores, Verónica Jiménez-Jacinto, César Bonavides-Martínez, Juan Segura-Salazar, et al. Regulondb (version 5.0): *Escherichia coli* k-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Research*, 34(suppl 1):D394–D397, 2006.
- [125] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.
- [126] Nicolas Sierro, Yuko Makita, Michiel de Hoon, and Kenta Nakai. DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. *Nucleic Acids Res.*, 36(Suppl 1):D93–D96, 2008.
- [127] G. Facchetti, G. Iacono, and C. Altafini. Computing global structural balance in large-scale signed social networks. *PNAS*, 108:20953–8, 2011.
- [128] Giuseppe Facchetti, Giovanni Iacono, Giovanna De Palo, and Claudio Altafini. A rate-distortion theory for gene regulatory networks and its application to logic gate consistency. *Bioinformatics*, 29(9):1166–1173, 2013.
- [129] Michael Szell, Renaud Lambiotte, and Stefan Thurner. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31):13636–13641, 2010.
- [130] Mark E. J. Newman, Albert L. Barabási, and Duncan J. Watts, editors. *The Structure and Dynamics of Networks*. Princeton University Press, 2006.
- [131] T. Hogg, D. Wilkinson, G. Szabo, and M. Brzozowski. Multiple relationship types in online communities and social networks. In *Proc. of the AAAI Spring Symposium on Social Information Processing*, AAAI SIP 2008, 2008.
- [132] Gergely Palla, Albert-Laszlo Barabasi, and Tamas Vicsek. Quantifying social group evolution. *Nature*, 446:664, 2007.
- [133] Reka Albert and Albert-Laszlo Barabasi. Topology of evolving networks: local events and universality. *Phys.Rev.Lett.*, 85:5234, 2000.
- [134] Fernando Vega-Redondo. *Complex social networks*. Econometric Society monographs. Cambridge Univ. Press, Cambridge, 2007.
- [135] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.
- [136] Stephen P. Borgatti, Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. Network analysis in the social sciences. *Science*, 323(5916):892–895, 2009.
- [137] S. Wasserman and K. Faust. *Social Network Analysis: methods and applications*. Cambridge Univ. Press, 1994.
- [138] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets. Reasoning About a Highly Connected World*. Cambridge Univ. Press, Cambridge, 2010.

- [139] Jérôme Kunegis, Andreas Lommatzsch, and Christian Bauckhage. The slashdot zoo: Mining a social network with negative edges. In *18th International World Wide Web Conference*, pages 741–741. 18th Int WWW Conf., April 2009.
- [140] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In *ACM SIGCHI Conference on Human Factors in Computing Systems*, Atlanta, GA, April 2010. ACM SIGCHI.
- [141] Aravind Srinivasan. Local balancing influences global structure in social networks. *Proceedings of the National Academy of Sciences*, 108(5):1751–1752, 2011.
- [142] F. Heider. Attitudes and cognitive organization. *Journal of Psychology*, 21:107–122, 1946.
- [143] D. Cartwright and F. Harary. Structural balance: a generalization of Heider’s theory. *Psychological Review*, 63:277–292, 1956.
- [144] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. De Luca, and S. Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *SDM’10*, pages 559–559. SDM’10, 2010.
- [145] M. Mezard, G. Parisi, and M. Virasoro. *Spin glass theory and beyond*. World Scientific, 1986.
- [146] R. Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th international conference on World Wide Web*, WWW ’04, pages 403–412, 2004.
- [147] Cliff A.C. Lampe, Erik Johnston, and Paul Resnick. Follow the reader: filtering comments on Slashdot. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI ’07, pages 1253–1262, 2007.
- [148] Moira Burke and Robert Kraut. Mopping up: modeling Wikipedia promotion decisions. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, CSCW ’08, pages 27–36, 2008.
- [149] Frank Harary and Jerald A. Kabell. A simple algorithm to detect balance in signed graphs. *Mathematical Social Sciences*, 1(1):131 – 136, 1980.
- [150] John S. Maybee and Stuart J. Maybee. An algorithm for identifying Morishima and anti-Morishima matrices and balanced digraphs. *Mathematical Social Sciences*, 6(1):99 – 103, 1983.
- [151] F Harary. On the measurement of structural balance. *Behavioral Science*, 4(4):316–323, 1959.
- [152] F Harary. A matrix criterion for structural balance. *Naval Res. Logist. Quart.*, 7:195–199, 1960.
- [153] F Barahona. On the computational complexity of Ising spin glass models. *J. Phys A: Math. Gen.*, 15(10):3241–3253, 1982.
- [154] Serge Galam. Fragmentation versus stability in bimodal coalitions. *Physica A: Statistical and Theoretical Physics*, 230(1-2):174 – 188, 1996.
- [155] T. Antal, P. L. Krapivsky, and S. Redner. Dynamics of social balance on networks. *Phys. Rev. E*, 72(3):036121, Sep 2005.

- [156] Seth A. Marvel, Steven H. Strogatz, and Jon M. Kleinberg. Energy landscape of social balance. *Phys. Rev. Lett.*, 103(19):198701, Nov 2009.
- [157] Scott Kirkpatrick. Frustration and ground-state degeneracy in spin glasses. *Phys. Rev. B*, 16(10):4630–4641, 1977.
- [158] W. J. DeGrip, E. N. Pugh, and D.G. Stavenga. *Molecular mechanisms in visual transduction*. Elsevier, Berlin, 2000.
- [159] Marie E Burns and Denis A Baylor. Activation, deactivation, and adaptation in vertebrate photoreceptor cells. *Annual review of neuroscience*, 24(1):779–805, 2001.
- [160] Vadim Y Arshavsky, Trevor D Lamb, and Edward N Pugh Jr. G proteins and phototransduction. *Annual review of Physiology*, 64(1):153–187, 2002.
- [161] King-Wai Yau and Roger C Hardie. Phototransduction motifs and variations. *Cell*, 139(2):246–264, 2009.
- [162] David Krizaj and David R Copenhagen. Calcium regulation in photoreceptors. *Frontiers in bioscience: a journal and virtual library*, 7:d2023, 2002.
- [163] Roy H Steinberg, Steven K Fisher, and Don H Anderson. Disc morphogenesis in vertebrate photoreceptors. *Journal of Comparative Neurology*, 190(3):501–518, 1980.
- [164] V.I. Govardovskii, D.A. Korenyak, S.A. Shukolyukov, and L.V. Zueva. Lateral diffusion of rhodopsin in photoreceptor membrane: a reappraisal. *Molecular vision*, 15:1717, 2009.
- [165] V. Kerov and N.O. Artemyev. Diffusion and light-dependent compartmentalization of transducin. *Molecular and Cellular Neuroscience*, 46(1):340–346, 2011.
- [166] H. Muradov, K.K. Boyd, M. Haeri, V. Kerov, B.E. Knox, and N.O. Artemyev. Characterization of human cone phosphodiesterase-6 ectopically expressed in xenopus laevis rods. *Journal of Biological Chemistry*, 284(47):32662–32669, 2009.
- [167] W. He, C.W. Cowan, and T.G. Wensel. Rgs9, a GTPase accelerator for phototransduction. *Neuron*, 20(1):95–102, 1998.
- [168] JW Karpen, DA Loney, and DA Baylor. Cyclic GMP-activated channels of salamander retinal rods: spatial distribution and variation of responsiveness. *The Journal of Physiology*, 448(1):257–274, 1992.
- [169] Y. Koutalos, K. Nakatani, and KW Yau. Cyclic GMP diffusion coefficient in rod photoreceptor outer segments. *Biophysical journal*, 68(1):373–382, 1995.
- [170] Kei Nakatani, Chunhe Chen, and Yiannis Koutalos. Calcium diffusion coefficient in rod photoreceptor outer segments. *Biophysical journal*, 82(2):728–739, 2002.
- [171] Andreas Helten, Werner Säftel, and Karl-Wilhelm Koch. Expression level and activity profile of membrane bound guanylate cyclase type 2 in rod outer segments. *Journal of neurochemistry*, 103(4):1439–1446, 2007.
- [172] I.B. Leskov, V.A. Klenchin, J.W. Handy, G.G. Whitlock, V.I. Govardovskii, M.D. Bownds, T.D. Lamb, E.N. Pugh, and V.Y. Arshavsky. The gain of rod phototransduction: reconciliation of biochemical and electrophysiological measurements. *Neuron*, 27(3):525–537, 2000.

- [173] N.P. Skiba, J.A. Hopp, and V.Y. Arshavsky. The effector enzyme regulates the duration of G-protein signaling in vertebrate photoreceptors by increasing the affinity between transducin and rgs protein. *Journal of Biological Chemistry*, 275(42):32716, 2000.
- [174] David Axelrod and MD Wang. Reduction-of-dimensionality kinetics at reaction-limited cell surface receptors. *Biophysical journal*, 66(3):588–600, 1994.
- [175] FW Wiegel and C DeLisi. Evaluation of reaction rate enhancement by reduction in dimensionality. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 243(5):R475–R479, 1982.
- [176] Owen P Gross, Edward N Pugh Jr, and Marie E Burns. Calcium feedback to cgmp synthesis strongly attenuates single-photon responses driven by long rhodopsin lifetimes. *Neuron*, 76(2):370–382, 2012.
- [177] Linda Ruggiero, Mark P Connor, Jeannie Chen, Ralf Langen, and Silvia C Finnemann. Diurnal, localized exposure of phosphatidylserine by rod outer segment tips in wild-type but not *itgb5*^{-/-} or *mfige8*^{-/-} mouse retina. *Proceedings of the National Academy of Sciences*, 109(21):8145–8148, 2012.
- [178] Richard W Young. The renewal of photoreceptor cell outer segments. *The Journal of cell biology*, 33(1):61–72, 1967.
- [179] Eduardo Solessio, Shobana S Mani, Nicolas Cuenca, Gustav A Engbretson, Robert B Barlow, and Barry E Knox. Developmental regulation of calcium-dependent feedback in *Xenopus* rods. *The Journal of general physiology*, 124(5):569–585, 2004.
- [180] JL Schnapf. Dependence of the single photon response on longitudinal position of absorption in toad rod outer segments. *The Journal of Physiology*, 343(1):147–159, 1983.
- [181] TD Lamb, PA McNaughton, and K-W Yau. Spatial spread of activation and background desensitization in toad rod outer segments. *The Journal of physiology*, 319(1):463–496, 1981.
- [182] K Boesze-Battaglia, SJ Fliesler, and AD Albert. Relationship of cholesterol content to spatial distribution and age of disc membranes in retinal rod outer segments. *Journal of Biological Chemistry*, 265(31):18867–18870, 1990.
- [183] K Boesze-Battaglia, Todd Hennessey, and Arlene D Albert. Cholesterol heterogeneity in bovine rod outer segment disk membranes. *Journal of Biological Chemistry*, 264(14):8151–8155, 1989.
- [184] Beata Jastrzebska, Aleksander Debinski, Slawomir Filipek, and Krzysztof Palczewski. Role of membrane integrity on G-protein-coupled receptors: Rhodopsin stability and function. *Progress in lipid research*, 50(3):267–277, 2011.
- [185] Y Nishizawa, A Yamazaki, and J Usukura. Electron microscopic localization of PDE- α , - β and - δ in frog retina. *Investigative Ophthalmology and Visual Science*, 45(5):5353, 2004.
- [186] Barry S Winkler. An hypothesis to account for the renewal of outer segments in rod and cone photoreceptor cells: renewal as a surrogate antioxidant. *Investigative ophthalmology & visual science*, 49(8):3259–3261, 2008.

- [187] P. Bisegna, G. Caruso, D. Andreucci, L. Shen, V.V. Gurevich, H.E. Hamm, and E. DiBenedetto. Diffusion of the second messengers in the cytoplasm acts as a variability suppressor of the single photon response in vertebrate phototransduction. *Biophysical journal*, 94(9):3363–3383, 2008.
- [188] G. Caruso, P. Bisegna, D. Andreucci, L. Lenoci, V.V. Gurevich, H.E. Hamm, and E. DiBenedetto. Identification of key factors that reduce the variability of the single photon response. *Proceedings of the National Academy of Sciences*, 108(19):7804, 2011.
- [189] O.P. Gross, E.N. Pugh, and M.E. Burns. Spatiotemporal cGMP dynamics in living mouse rods. *Biophysical Journal*, 102(8):1775–1784, 2012.
- [190] Nigel Sim, Dmitri Bessarab, C Michael Jones, and Leonid Krivitsky. Method of targeted delivery of laser beam to isolated retinal rods by fiber optics. *Biomedical optics express*, 2(11):2926, 2011.
- [191] PD Calvert, VI Govardovskii, N. Krasnoperova, RE Anderson, J. Lem, and CL Makino. Membrane protein diffusion sets the speed of rod phototransduction. *Nature*, 411(6833):90–94, 2001.
- [192] Shu-ichi Watanabe and Gary Matthews. Regional distribution of cGMP-activated ion channels in the plasma membrane of the rod photoreceptor. *The Journal of neuroscience*, 8(7):2334–2337, 1988.
- [193] F Rieke and DA Baylor. Origin of reproducibility in the responses of retinal rods to single photons. *Biophysical Journal*, 75(4):1836–1857, 1998.
- [194] DA Baylor, TD Lamb, and King-Wai Yau. Responses of retinal rods to single photons. *The Journal of physiology*, 288:613, 1979.
- [195] D. Srivastava, R.L. Hurwitz, and D.A. Fox. Lead-and calcium-mediated inhibition of bovine rod cgmp phosphodiesterase: interactions with magnesium. *Toxicology and applied pharmacology*, 134(1):43–52, 1995.

ACKNOWLEDGMENTS

I would like to thank to my supervisor Claudio Altafini, for the great help he has given me during these four years, for teaching me how to perform a research, how to write papers, and for leaving me free to choose some projects.

Thanks to Vincent Torre for having involved me in his interesting research topic.

Thanks to the “mSBP” group, from all the faculty members to the secretary, from post-docs to students (both current and past): beside the courses, seminars, schools and sharing group, many dinners, drinks and games have been important for making me feel one of them. Thanks to Cristian, Alessandro, Giovanni, Angelo, Loredana, Alessandra, Barbara, Francesco, Emanuela, Jessica, Maria, Ivan, Sandro, Rolando, Trang, Fahimeh, Zahleh, Pilar, Salvatore, Luca, Duvan, Xiaojing, Ina, Marco, Andrea, Guido, Edoardo, Alejandro, Francesca and Richard.

Thanks to my officemate Gianpaolo for the useful discussions on the blackboard (after we got it); thanks to Daniele for showing me the terminal commands and for making my Erdisu card still useful; thanks to Fahimeh for the talks we had about nature and world situation. Thanks to Mattia for his support on my first project; thanks to Giovanna for the work we did together but also for her suggestions about photos.

Thanks to Monica for all the biological explanations and for the nice talks we had in the morning or together with Manuela.

Thanks to all SISSA staff who have helped me: Riccardo, Federica, Marco, Isabella and Alexandra.

Thanks to Sebastiano, Luca, Giancarlo, Marco, Alessio and Giorgia for the time spent together discussing about the future; thanks to Olga for her suggestions about my health, for the sweets and for the artichoke-cake.

Thanks to Luca for his moral support during these years but also for making me discover the wonderful world of biking, for having taught me how to dive and for having made me interested in Germany.

Thanks to the trainers of the CUS swimming course and to the swimmers which have encouraged me when I was not able to finish a lane: thanks to Renato, Stefano, Barbara, Karen, Silvia, Elena, Stefano, Giovanni, Enrico, Riccardo and Ignazio.

Thanks to Paolo, Laura, Marisa, Chantal, Judith and Carla of the British Film Club for the Wednesday conversations during the last two years.

Thanks to all “Assorbilibri”, the great reading group which made me discover many new books, visit regional art exhibitions and spend unforgettable evenings: thanks to Gabriella, Marinella, Alessandra, Marisa, Franca, Mirella, Giovanna, Massimo, Alessandro, Michele, Luisella, Maddalena, Loredana, Donatella, Patrizia and thanks also to Scilla. They are my family in Trieste.

Thanks to Oscar also for the photographic tours and advices, and for all the discussions and questions about life.

Thanks to all my friends around Italy: Francesco, Alessandro, Antonio, Petra, Ameriga, Emiliana and Anna which visited me and which are still friends even if I have been always bound here.

Thanks to my family for trusting me and my choices.

Thanks to Tomáš for having been the starting point of this Ph.D., for his always generous help, for the support the corrections and the suggestions he gave me during these four years, for the time spent with the “green friend”, for the buses we took together, for all carsolina, kren sandwiches, concerts... and for having made me love *our Trieste*.