HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Intracellular Networks
A Computational Systems Biology Perspective

Inaugural-Dissertation

zur Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

vorgelegt von
Armin Sadat Khonsari
aus Berlin

Düsseldorf, 27. Juni 2016

aus dem Institut für Mathematische Modellierung biologischer Systeme
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Markus Kollmann
Korreferent: Prof. Dr. Martin Lercher
Tag der mündlichen Prüfung: 12.09.2016

This thesis is based on following original papers:

Part I
The first part of this thesis is identical with the original paper [29]:

Armin Sadat Khonsari and Markus Kollmann,
Perception and regulatory principles of microbial growth control,
PLoS One, 10(5):e0126244 (2015)

Part II
Some of the ideas of the second part of this thesis can be found in the original paper [12]:

Christopher Blum, Nadia Heramvand, Armin Sadat Khonsari and Markus Kollmann,
Inferability of transcriptional networks from large scale gene deletion studies,
(submitted) (2016)

Intracellular Networks

A Computational Systems Biology Perspective

by

Armin Sadat Khonsari

Summary

Computational systems biology has emerged as a promising new biological field which studies the complexity of biological systems as a whole by integrating mathematical, computational and experimental approaches. The two applications presented in this thesis are concerned with different aspects of intracellular networks, elucidating the wide range of topics within this field.

In the first part the regulation of metabolic networks of fast growing microbes under fluctuating environmental nutrient availability is explored. Fast growth represents an effective strategy for microbial organisms to survive in competitive environments. To accomplish this task, cells must adapt their metabolism to changing nutrient conditions in a way that maximizes their growth rate. However, the regulation of the growth related metabolic pathways can be fundamentally different among microbes. Therefore, it was asked whether growth control by perception of the cell's intracellular metabolic state can give rise to higher growth than by direct perception of extracellular nutrient availability. The results of the computer simulation indicate that the intracellular perception is advantageous under situations where the up and down regulation of pathways cannot follow the fast changing nutrient availability in the environment. In this case, optimal regulation ignores any other nutrients except the most preferential ones, in agreement with the phenomenon of catabolite repression in prokaryotes. As a result, species that rely on intracellular perception gain a relevant fitness advantage in fluctuating nutrient environments, which enables survival by outgrowing competitors.

The second part focuses on the network inference of gene regulatory networks (GRN) and signal transduction networks (STN) from perturbation data. An important aspect in understanding organisms on a cellular level is the knowledge about the exact causal interaction network between biochemical components inside the cell. The inference of these GRN or STN exclusively from variations in the abundance of mRNA or phospho-proteins, respectively, in response to perturbations is experimentally more feasible on one hand but challenging due to high measurement noise on the other hand. Here, a novel machine learning technique in the field of network inference has been developed, which overcomes Gaussian measurement noise despite of only a few replicate experiments. The technique is based on the theory of probabilistic principle component analysis applied to partial correlations, which leads to a dimensionality reduction of the network inference problem. Knowledge about the structure of GRN and STN builds the groundwork for predictive models, which can be used to find new therapeutic targets in diseased cells or help to reprogram organisms in biotech applications.

Zusammenfassung

Computational systems biology hat sich als vielversprechendes neues Gebiet der Biologie etabliert, das durch die Integration mathematischer, computergestützter und experimenteller Ansätze die Komplexität biologischer Systeme als Ganzes untersucht. Die zwei in dieser Arbeit beschriebenen Anwendungen beschäftigen sich jeweils mit unterschiedlichen Aspekten von intrazellulären Netzwerken, was die Breite der unterschiedlichen Themen in diesem Gebiet verdeutlicht.

Im ersten Teil wird die Regulation von metabolischen Netzwerken von schnell wachsenden Mikroben in einer sich verändernden Nährstoffumgebung untersucht. Für mikrobielle Organismen in Konkurrenzsituationen stellt ein schnelles Wachstum eine effektive Überlebensstrategie dar. Um diese Aufgabe zu erfüllen, müssen Zellen ihren Stoffwechsel an sich verändernde Nährstoffbedingungen in einer Weise anpassen, die ihre Wachstumsrate maximiert. Jedoch kann bei Mikroben die Regulation von Stoffwechselwegen, die das Wachstum kontrollieren, fundamental unterschiedlich sein. Daher kam die Frage auf, ob eine Wachstumsregulation basierend auf der Wahrnehmung des intrazellulären Stoffwechselzustands zu einem höheren Wachstum führen kann als eine Regulation basierend auf der direkten Wahrnehmung der extrazellulären Nährstoffverfügbarkeit. Die Ergebnisse der Computersimulation zeigen, dass die intrazelluläre Wahrnehmung unter Situationen von Vorteil ist, wo das Herauf- und Herunterregulieren der Stoffwechselwege nicht den schnellen Veränderung der Nährstoffverfügbarkeit in der Umgebung folgen kann. In diesem Fall ignoriert die optimale Regulation alle anderen Nährstoffe mit Ausnahme des am meisten bevorzugten Nährstoffs, übereinstimmend mit dem Phänomen der Katabolitrepression in Prokaryoten. Infolgedessen gewinnen Spezies, die sich auf die intrazelluläre Wahrnehmung verlassen, einen Fitnessvorteil in fluktuierenden Nährstoffumgebungen, sodass das Überleben durch das Überwachsen der Konkurrenten gewährleistet ist.

Der zweite Teil beschäftigt sich mit der Netzwerk-Inferenz von Gen-regulatorischen-Netzwerken (GRN) und Signaltransduktions-Netzwerken (STN) anhand von experimentellen Störungsdaten. Ein wichtiger Aspekt um Organismen auf einer zellulären Ebene zu verstehen, ist das Wissen um das genaue kausale Interaktions-Netzwerk zwischen biochemischen Komponenten innerhalb einer Zelle. Die Inferenz von GRN bzw. STN ausschließlich aus Variationen in der Menge von mRNA bzw. Phosphoproteinen ist zwar experimentell einfacher durchführbar, jedoch wird das Vorhaben durch hohes Messrauschen erschwert. Deshalb wird in der vorliegenden Arbeit eine neuartige "Machine Learning" - Methode auf dem Gebiet der Netzwerk-Inferenz entwickelt, die weißes Messrauschen trotz nur weniger Wiederholungsexperimente bewältigt. Die Methode basiert auf der Anwendung der probabilistischen Hauptkomponentenanalyse (PPCA) auf partiellen Korrelationen, was zu einer Verringerung der Dimensionalität des Netzwerk-Inferenz-Problems führt. Das Wissen über die Struktur der GRN und STN bildet die Grundlage für Vorhersagemodelle, die verwendet werden können, um neue therapeutische Targets in erkrankten Zellen zu finden, oder um Organismen in Biotech-Anwendungen neu zu programmieren.

Contents

Part I

Perception and Regulatory Principles of Microbial Growth Control

Chapter 1

Introduction

One of the most essential aspects of living cells is growth and its associated control to fit the organisms' needs. In human, selection for fast and selfish growth can result in cancer, while it represents a very effective evolutionary strategy for microorganisms to survive in a competitive environment. The reproductive success of microbial organism depends on the fast and precise adjustment of their growth rate to the actual environmental condition [55]. The reason is that most microbes live in a highly competitive environment where fast and effective transfer of available nutrients into biomass can give a significant fitness advantage [45].

Selection for fast growth leads to phenomena such as overflow metabolism [26, 61, 62], where fast but wasteful conversion of glucose into biomass can be of advantage in comparison to the effective use of nutrients. The overflow metabolism of *E.coli* is also known as Crabtree effect in *S. cerevisiae* and as Warburg effect in cancer cells [37]. Another regulatory phenomena that is associated with fast growth and is commonly used among many bacteria and other microbes is carbon catabolite repression (CCR) [17,23,65]. To grow fast microbes selectively utilize preferred carbon sources in a hierarchical manner. In the presence of a preferred sugar such as glucose, CCR causes metabolic enzymes of alternative carbon sources to be expressed at low rate and can additionally reduce their activity.

There is strong evidence that growth dependent phenomena such as overflow metabolism or CCR are the consequence of a metabolic regulation or growth con-

trol in response to extracellular nutrient availability. Further, it seems possible that the perception of extracellular nutrient availability plays an important role in growth control [69], as it is the primary information cellular response is based on. We define two distinct types of perception, termed intracellular and extracellular perception. In the case of extracellular perception the cell regulates its metabolism exclusively in response to extracellular nutrient information, while in the case of intracellular perception microbes indirectly recognize nutrient availability by perceiving the intracellular metabolic state. The intracellular perception is motivated by experimental observations [18,32,68] of microbes, e.g. *E.coli*, which do not possess any extracellular carbohydrate receptors, like the Rgt2 and Snf3 glucose sensors of yeast [47,71]. These microbes should be capable of perceiving extracellular nutrient availability indirectly from intracellular metabolic states. Intuitively, the extracellular perception should lead to a more precise and fast adaptation to nutrient availability, since changes in the environment can be perceived faster and to higher accuracy. Here, the question arises whether exclusive intracellular perception can result in a growth benefit in presence of fast fluctuating nutrient concentrations. Following this question, we are interested in which frequency regimes the exclusive perception of intracellular nutrient concentration is evolutionary more beneficial than the exclusive perception of extracellular nutrient concentrations. Furthermore, what are the regulatory principles causing this benefit in average growth rate or fitness and can the regulatory phenomenon of carbon catabolite repression be understood by means of nutrient perception?

To give an answer to these questions and an explanation how the integration of the perception strategies for growth control contribute to shape growth rate in microorganisms, we will introduce a simplified replicator model for microbial growth. The replicator model consists of a minimal metabolic network, ribosomes, and a controller that can detect intracellular and extracellular metabolite concentrations. Optimal growth control is realized by minimizing the difference between the actual intracellular concentrations of metabolites and precursors and their desired concentrations, which is determined by the perceived nutrient availability. Using this simplified model we are able to show that growth control by perception of extracellular nutri-

ent concentrations is of selective advantage if environmental conditions change slowly over time. If environmental conditions change fast in comparison to the minimum generation time, gene regulation and protein turnover will lag behind and the model predicts that in this case sensing the intracellular precursor state is of advantage.

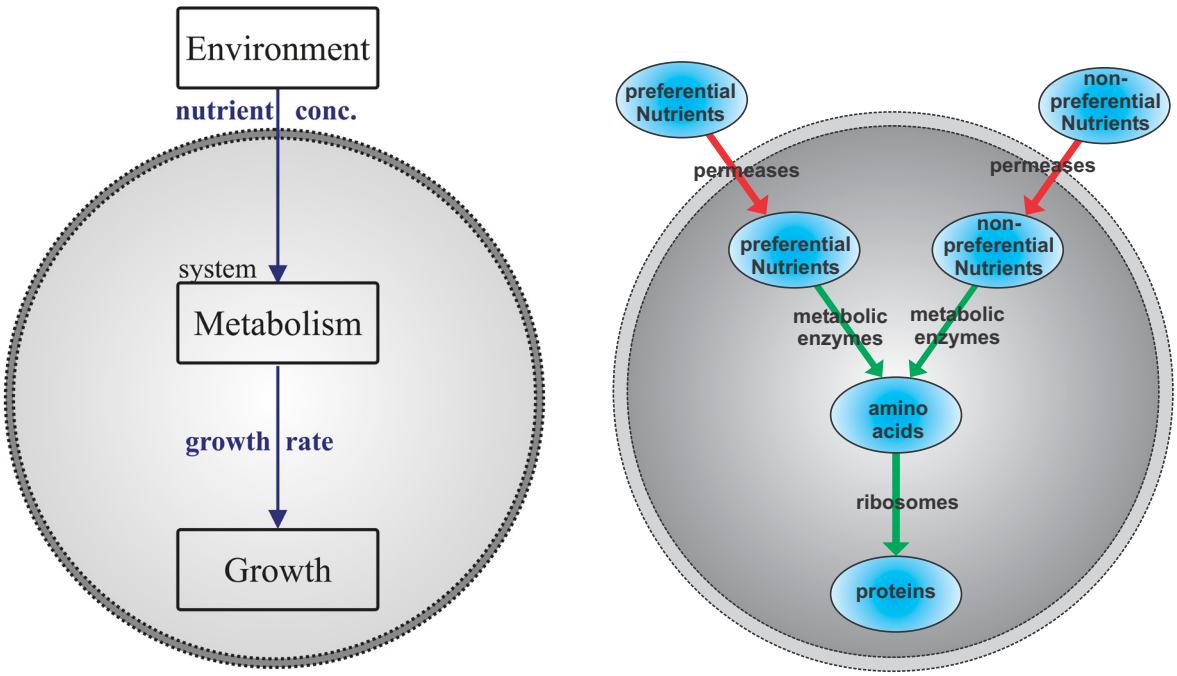
Chapter 2

Methods

2.1 Self-replicator model

The first step in modeling a system is to understand the main features which have a relevant effect on the studied phenomenon or scientific objective. These features are taken to construct the most simple model which still suffices to reproduce reality. In this study we are interested in fast growing unicellular organisms in changing environments with focus on cell metabolism and its regulation. Growth is a consequence of the underlying metabolic fluxes and growth rate is affected by changes in metabolic rate which in turn can be a result of environmental changes (see Figure 2-1(a)). In the following, we define growth by the amount of protein that is synthesized. Focusing exclusively on the protein content and thereby neglecting other cellular components is legitimate since the protein synthesis capacity of a cell remains approximately constant over time [56, 57].

The next question is how a real-life metabolism can be further simplified and generalized, to avoid inclusion of too many molecular details. Molenaar et al. [45] have successfully shown that simple self-replicating systems (self-replicators) qualitatively reproduce the regulation of major cellular components (protein, lipids, etc.) for unicellular organisms. The simplest self-replicator consists of ribosomes which synthesize themselves by means of precursors (real-life example [35]). In this work we rely on a slightly more complex architecture which is obtained by adding transporters and



(a) (b)

Figure 2-1: **The metabolism of the self-replicator shown in two possible representation.** (a) Block diagram: blocks symbolize processes and arrows associated inputs and outputs. The big dashed circle distinguishes between intracellular and extracellular processes. The process of growth is caused by the underlying metabolism which in turn depends on the nutrient availability in the environment. (b) Pool diagram: ellipses represent the protein and metabolite pools. Red arrows symbolize uptake transports and green arrows stand for metabolic pathway fluxes. The self-replicator consists of two metabolic pathways – one for preferential nutrients and one for non-preferential ones.

metabolic pathways to the simple self-replicator model.

The whole self-replicating system, as sketched in Figure 2-1(b), consists of a metabolic flux network, where metabolite pools are connected by biochemical reactions catalyzed by specific enzymes. For the sake of simplicity and without loss of generality, it is assumed that there are only two types of time varying nutrient components, namely a preferential sugar (PS) and a non-preferential sugar (NPS), which both can be growth limiting. All other compounds that are required for growth are assumed to be available in excess. Further, we assume that the self-replicating system will be situated in a surrounding that periodically switches between a PS and an NPS environment. As only two nutrient components change over time, our simplified

cell comprises two catabolic pathways. The external nutrients can be imported into the cell by specific permeases, where they are transformed into metabolic precursors, i.e. amino acids, as the only precursor in the system. Using amino acids, ribosomes synthesize the five distinct enzyme types that the self-replicator consists of, including themselves. These five enzymes constitute the total amount of proteins belonging to one self-replicator. Their relative share of the total protein amount influences the protein synthesis rate, i.e. growth rate.

Each metabolic pathway, represented by the fluxes and arrows in Figure 2-1(b), can be thought to be catalyzed by a group of enzymes with concentrations $\hat{E}_i$. The effective enzyme concentration of one whole pathway j is expressed as $E_j = \sum_i \hat{E}_i$. It is assumed that the maximum concentration of the proteome does not exceed a constant proteome density $E^{\max} = \sum_j E_j$ [42, 60]. The overall protein mass density of the whole population is defined as total protein mass M^{tot} of the population per total cell volume of the population V^{pop}.

$$E^{\max} := \frac{M^{\text{tot}}(t)}{V^{\text{pop}}(t)} = \text{const.} \tag{2.1}$$

Note that M^{tot} and V^{pop} are quantities that are measured in batch culture experiments and that $E^{\max}$ corresponds to the population averaged cellular protein concentration. In what follows, we assume that fast growing organisms are optimized for biomass production, an assumption which is strongly supported by recent experimental results [56, 68]. In order to describe the system dynamics with the necessary accuracy, we introduce a mathematical description for the metabolite and enzyme pool dynamics.

2.2 Metabolite pool dynamics

A metabolite pool is characterized by its mass density. The mass density $[X]$ of metabolite X is denoted as metabolite mass $m_X(t)$ per population volume V^{pop}.

$$[X](t) := \frac{m_X(t)}{V^{\text{pop}}(t)} \tag{2.2}$$

Alternatively, one can use the particle density $[\tilde{X}] := n_X/V^{\text{pop}}$, which is the amount of particle in *mol* over population volume in l. (This definition is utilized for flux balance analysis with the Matlab toolbox *cobra* [52].) All metabolite pool dynamics are defined by continuity equations. Furthermore the concentration and fluxes must always be positive, as it is hinted in Figure 2-2.

- Continuity: $\frac{d[X](t)}{dt} = v_{\text{in}}(t) - v_{\text{out}}(t)$
- Positivity: $[X](t) \geq 0$ and $v_i \geq 0$

The outflow rate $v_{\text{out}}(t)$ depends on the pool concentration $[X](t)$ in conjunction with the related enzyme concentration E_X, whereas there is no direct dependency to the inflow rate $v_{\text{in}}(t)$. Due to the existence of a single metabolic network, all pools are connected. This gives rise to interpret all inflow rates as an outflow rate of an upper pool $[Y](t)$. Hence for a linear pathway, it is sufficient to define the outflow rate as:

$$v_X(t) := v_{out}(t) = \frac{[X](t)}{K_M^X + [X](t)} \cdot k_X \cdot E_X(t) \quad , \tag{2.3}$$

where K_M^X is the Michaelis-Menten constant and k_X is the catalytic rate of the enzyme reaction. The inflow rate has the same expression as above with the only difference of being defined by the upper pool $[Y](t)$, i.e. $v_{\text{in}} = v_{\text{Y}}$. In order to work with normalized quantities, the relative mass λ_X of metabolite X is introduced by

$$\lambda_X(t) := \frac{m_X(t)}{M^{\text{tot}}(t)} = \frac{[X](t)}{E^{\max}} \quad . \tag{2.4}$$

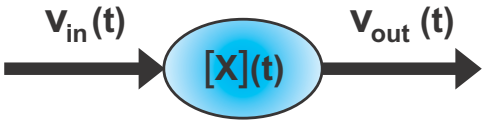


Figure 2-2: **Concentration dynamic of arbitrary metabolite X.** While the outflow rate $v_{\text{out}}(t)$ depends on the metabolite pool concentration $[X](t)$, the inflow rate $v_{\text{in}}(t)$ is independent of $[X](t)$ and is subject to an upstream pool.

The pool dynamics follow from the defined fluxes, the continuity equation, and the definition of the relative metabolite mass.

$$\frac{d}{dt}\lambda_X(t) = \frac{1}{E^{\max}} \cdot \frac{d}{dt}[X](t) = \frac{v_Y(t) - v_X(t)}{E^{\max}} \tag{2.5}$$

2.3 Enzyme pool dynamics: regulation and growth

A mathematical description of growth control can be obtained by determining the time dynamics of growth rate and the enzyme pools. Optimal growth control is achieved by regulating metabolic fluxes in a way that maximizes growth rate. The metabolic fluxes are driven by their related enzyme concentrations and extra- and intracellular metabolite concentrations. Since the latter is a not influenceable environmental factor, growth control exclusively means regulating enzyme concentrations. The optimal timing, by which this regulation is performed, is influenced by the growth rate. The reason is that the enzyme concentrations can be diluted or over-expressed due to growth. In the following, the proper quantitative definitions of growth, growth rate and regulation will be developed in order to obtain the basis for deriving their time dynamics.

2.3.1 Definition

To describe **cellular growth**, the protein mass is a better quantity than the corresponding concentration. The time evolution of the total protein mass $M^{\text{tot}}(t)$ is proportional to its cell population volume $V^{\text{pop}}(t)$, since we assume a constant total protein concentration $E^{\max}$. Consequently, $d_t E^{\max} = d_t \left(M^{\text{tot}}(t)/V_{\text{cell}}(t)\right) = 0$, despite

of increasing mass and volume. Hence, the total protein mass and the associated total protein mass flux are the appropriate quantities for describing cell growth and growth rate, respectively.

To describe the **regulatory dynamics** of the various enzyme pools we introduce the relative enzyme mass $\phi_j = M_j/M_{\text{tot}}$ by the ratio of the enzyme mass $M_j(t)$ of a metabolic pathway j to the total protein mass $M^{\text{tot}}(t)$. As the cellular system tends to maximize its growth rate, which is represented by the synthesized protein mass per time unit, optimal growth rate is a result of an optimized metabolism. In this model, the only way of tuning metabolism is by means of redistributing the enzyme concentrations $E_j(t)$ of metabolic pathways. This is due to a constant intracellular protein concentration, which is maintained by the cell to ensure efficiency of central cellular processes, such as protein folding [10, 42]. In analogy to the relative enzyme mass ϕ_j, one can define a relative enzyme concentration $E_j(t)/E^{\max}$, which can be shown to be related:

$$\phi_j(t) := \frac{M_j(t)}{M_{\text{tot}}(t)} = \frac{\left(M_j(t)/V^{\text{pop}}(t)\right)}{\left(M^{\text{tot}}(t)/V^{\text{pop}}(t)\right)} = \frac{E_j(t)}{E^{\max}} \quad , \tag{2.6}$$

where $M^{\text{tot}}(t) = \sum_j M_j(t)$ and $\sum_j \phi_j(t) = \phi^{\max} = 1$. Both quantities can likewise be used to describe metabolic regulation. But the relative enzyme mass $\phi_j(t)$ is more favorable, because it stands in direct relation to the definition of cellular growth, and will be used for the derivation of the regulatory dynamics below.

2.3.2 Regulation

The regulatory dynamics are obtained by taking the time derivative of the relative protein mass $\phi_j(t)$. The time derivative $d/dt\phi_j(t)$ depends on the derivatives of the total protein mass $d_t M^{\text{tot}}(t)$ and the pathway protein mass $d_t M_j(t)$. For this purpose

one can define the following useful relation between both mass quantities:

$$\gamma_j(t) := \frac{\frac{d}{dt}M_j(t)}{\frac{d}{dt}M^{\mathrm{tot}}(t)} \quad , \tag{2.7}$$

where $\sum_j \gamma_j(t) \equiv 1$. The *relative synthesis rate* $\gamma_j(t)$ is the synthesis rate of enzymes from pathway j with respect to the overall synthesis rate. It can be interpreted as the fraction of protein synthesis capacity that is assigned to enzyme j. This synthesis capacity can be generalized to other biological regulatory mechanisms, like the amount of mRNA , tRNA etc. . Deriving the relative protein mass and using relation Eq. (2.7) yields the ordinary differential equation for the regulatory dynamics.

$$\begin{aligned}
\frac{d}{dt}\left(\frac{M_j(t)}{M^{\mathrm{tot}}(t)}\right) &= \frac{M^{\mathrm{tot}}(t)\cdot(d_t M_j(t)) - M_j(t)\cdot(d_t M^{\mathrm{tot}}(t))}{(M^{\mathrm{tot}}(t))^2} \\
&= \frac{\frac{d}{dt}M^{\mathrm{tot}}(t)}{M^{\mathrm{tot}}(t)} \cdot \left(\frac{\frac{d}{dt}M_j(t)}{\frac{d}{dt}M^{\mathrm{tot}}(t)} - \frac{M_j(t)}{M^{\mathrm{tot}}(t)}\right) \\
\frac{d}{dt}\phi_j(t) &= \frac{\frac{d}{dt}M^{\mathrm{tot}}(t)}{M^{\mathrm{tot}}(t)} \cdot \left[\gamma_j(t) - \phi_j(t)\right]
\end{aligned} \tag{2.8}$$

The differential equation (Eq. (2.8)) describes the change of the relative enzyme mass for each pathway. This time-dependency of enzymatic resources represents the regulatory dynamics of a single cell, under the simplifying assumptions introduced before. The relative enzyme mass ϕ_j tends toward the synthesis rate ratio γ_j with the population size independent growth rate $v_{\mathrm{growth}} = d_t M^{\mathrm{tot}}/M^{\mathrm{tot}}$. Eq. (2.8) describes a growing cellular system that redistributes its protein synthesis capacity in regulatory manner, under the constraint $\sum_j \gamma_j(t) \equiv 1$.

There are three scenarios with respect to regulation. Using relation Eq. (2.8), one can find following interpretation:

1. **Dilution**: enzyme concentration decreases

   $\gamma_j(t) < \phi_j(t) \Leftrightarrow \frac{d}{dt}\frac{M_j(t)}{M^{\mathrm{tot}}(t)} < 0 \Leftrightarrow \frac{d}{dt}E_j(t) < 0$

   If the relative synthesis rate γ_j is smaller than the relative enzyme mass ϕ_j, the synthesis rate of enzyme j will be smaller than the growth rate. Hence, a dilution

effect will be initiated and relative enzyme mass and enzyme concentration E_j will decrease.

2. **Over-expression**: enzyme concentration increases
   $\gamma_j(t) > \phi_j(t) \Leftrightarrow \frac{d}{dt}\frac{M_j(t)}{M^{\text{tot}}(t)} > 0 \Leftrightarrow \frac{d}{dt}E_j(t) > 0$
   If the relative synthesis rate γ_j is larger than the relative mass ϕ_j, enzyme j will be synthesized faster than the rate the cell is growing. Hence, an over-expression effect will be initiated and relative enzyme mass and enzyme concentration E_j will increase.

3. **Homeostasis**: enzyme concentration stays constant
   $\gamma_j(t) = \phi_j(t) \Leftrightarrow \frac{d}{dt}\frac{M_j(t)}{M^{\text{tot}}(t)} = 0 \Leftrightarrow \frac{d}{dt}E_j(t) = 0$
   If the relative synthesis rate γ_j is as large as the relative mass ϕ_j, enzyme j will be synthesized exactly as fast as the rate the cell is growing. Hence, the cellular enzyme composition will be preserved and homeostasis is established - relative enzyme mass and enzyme concentration E_j will stay constant.

The system always tends to the third case, homeostasis, where following relation is established:

$$\frac{d}{dt}M_j(t) = \phi_j(t) \cdot \frac{d}{dt}M^{\text{tot}}(t) \tag{2.9}$$

2.3.3 Growth

In order to determine the time dependency of cellular protein mass growth, the following ordinary differential equation has to be solved:

$$\frac{d}{dt}M^{\text{tot}}(t) = \Big(\beta_R(t) \cdot k_R \cdot \phi_R(t)\Big) \cdot M^{\text{tot}}(t) \quad , \tag{2.10}$$

where $\beta_R(t) = ([\text{AA}])/(K_M^R + [\text{AA}])$ is the probability of amino-acid-binding to a ribosome and [AA] is the amino acid concentration. We do not consider the contribution of different amino acids because one type is sufficient for our phenomenological model, regarding previous assumptions. This differential equation, Eq. (2.10), represents exponential growth with a time-dependent growth rate $v_{\text{growth}}(t) := \beta_R(t) \cdot k_R \cdot \phi_R(t)$,

whereas $v_{\text{growth}}(t)$ is based on Michaelis-Menten kinetics of ribosomal translation. Solving this ordinary differential equation yields the following exponential growth relation.

$$M^{\text{tot}}(t) = M^{\text{tot}}(t_0) \cdot \exp\left(k_R \cdot \int_{t_0}^{t} \beta_R(t) \cdot \phi_R(t)\, dt\right) \tag{2.11}$$

Eq. (2.11) can be seen as microscopic view of cellular growth, where the population's protein mass is exponentially increased instead of the the number of cells. To transform Eq. (2.11) into a more classical macroscopical form of cell growth, one hast to introduce the relation $M^{\text{tot}}(t) = \langle M_{\text{cell}} \rangle \cdot n(t)$, where $n(t)$ denotes the number of cells in a population and $\langle M_{\text{cell}} \rangle$ is the average proteome mass of a single cell. Applying this relation and the connection $v_{\text{growth}}(t) = \ln 2 / t_D(t)$ between growth rate v_{growth} and cellular doubling time t_D yields the macroscopic view of cellular growth.

$$n(t) = n(t_0) \cdot 2^{\int_{t_0}^{t} \frac{1}{t_D(t)}\, dt} \tag{2.12}$$

Here, the time-dependent cellular doubling time is expressed as

$$t_D(t) = \frac{ln(2)}{\beta_R(t) \cdot k_R \cdot \phi_R(t)} \quad . \tag{2.13}$$

Eq. (2.12) and Eq. (2.13) show that the population size doubles by a time which depends on the amino acid concentration [AA] and relative protein mass *investment* in ribosomes. The more ribosomes and amino acids are present, the shorter is the cellular doubling time and the faster is cellular growth. Assuming stable proteins, the doubling time equals the response time t_R, i.e. the time a cell needs to respond properly to an environmental change.

2.4 Control system

A living cell can be regarded as a control system consisting of a system to be controlled, controller, actuator, and sensors. The system to be controlled is represented by the metabolic network, while the actuator can be seen as the protein synthesis

machinery, i.e. ribosomes which produce specific enzymes with a probability given by the relative protein synthesis rate γ_j. Next, sensors and controller have to be added to the model in order to complete the control system. In the following the process of sensing will be referred to as perception and there will be only two types as will be seen below, namely intracellular perception and extracellular perception. The explicit nature of the sensors are not important for our research question, since we are only interested in the effective information content of those. The controller yields cellular regulation, which must be inferred by an mathematical optimization process.

Having all pieces together, one can explain the dynamic steps of growth control by means of the control system sketched in Figure 2-3. The metabolic system takes up nutrients from the environment and metabolizes them into proteins, which in turn increase cell mass, i.e. the cell grows. This process is regulated by the controller which receives information about the nutrient availability from perception and hands over desired enzyme concentrations to the actuator, namely the ribosomes. The actuator implements this desired values by changing the actual enzyme concentration. This total control process is time dependent and hence explains the dynamic steps of growth control.

As mentioned, the modeled control system has a desired value and an actual value for all fluxes and enzyme concentrations. Former has to be distinguished from the optimal value. While the **desired value** is a quantity that the **actual value** aims for, the **optimal value** is a quantity which represents the global maximum or minimum of an objective function (here: growth rate). If and only if the desired value is determined under ideal conditions, it will be equal to the optimal value.

2.4.1 Desired value

The desired value, which represents the control of a system, depends strongly on the information quality of the surrounding environment, namely the extracellular nutrient concentration. This information is of utmost importance for the desired value's accuracy, that is the degree of optimality with respect to the control. Obviously, the measurement of extracellular nutrient concentration is more precise than the one of

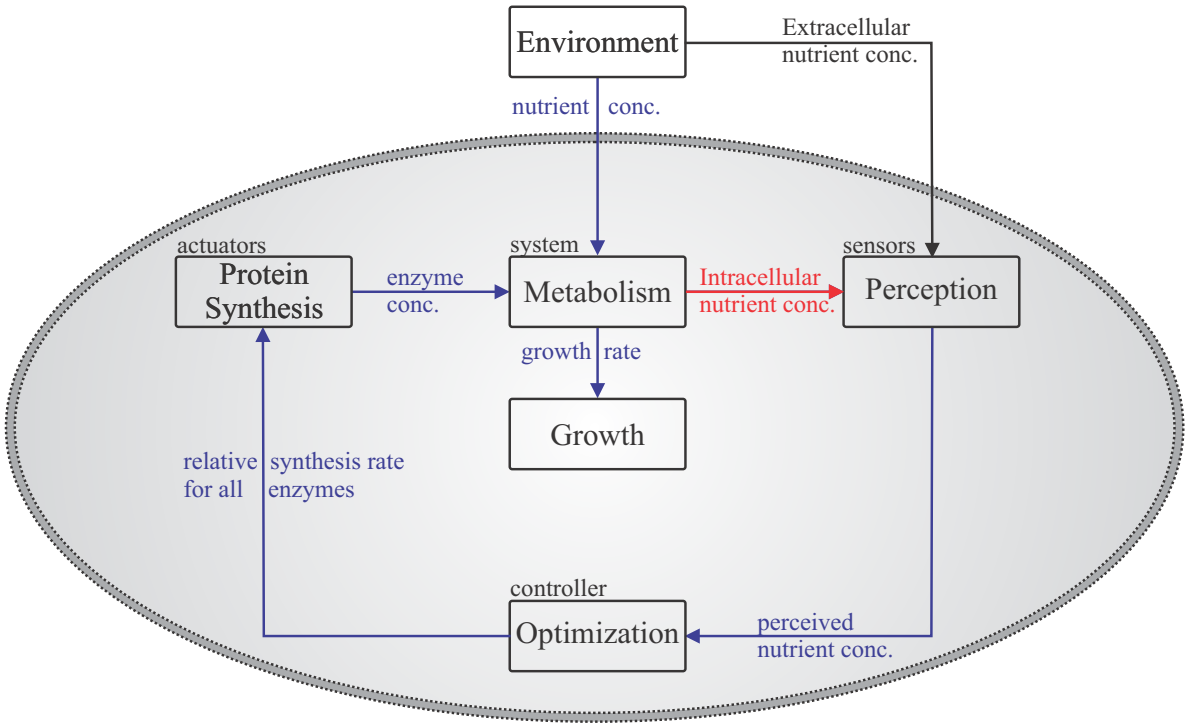


Figure 2-3: **Block diagram of the whole modeled replicating system.** This control system consists of a system to be controlled, namely the metabolic network, a controller, actuators and sensors for determining the metabolic pools' relative mass. Each block represents a process, which can contain sub-processes. While blue arrows represent input and output of the different processes, the red and black arrows represent the input for intracellular and extracellular perception, respectively.

intracellular nutrient concentration. On the other hand, a highly precise determination of the actual extracellular concentration can be disadvantageous with respect to growth, in the case of a rapidly changing environment.

Another important point for the determination of the desired value is a matching amount of enzymes to their associated metabolites. This *optimal resource allocation* prohibits the waste of enzymes in the case of enzyme overproduction and prevents from a non-optimal growth rate due to the mismatch between catalytic capacity of too less enzymes and the existing larger metabolite pool [16]. A non-optimal distribution of resources will always cause a decrease in growth rate compared to the optimal state. In physical terms, optimal resource allocation is defined as the condition, in which the metabolite net inflow rate v_{in} into a pool equals the catalytic net outflow

rate v_{out}.

$$\sum_i v_{in}^i(t) = \sum_k v_{out}^k(t) \tag{2.14}$$

This assumption or condition, respectively, implicates *balanced fluxes* and constant pool concentrations for the whole network, if the environment is regarded to be constant. Therefore, it is possible to consider the pool and flux dynamics as an stationary process, where the pool concentration and flux instantaneously adapt to an new environment by tuning enzyme concentrations to the according desired values.

2.4.2 Actual value

Balanced fluxes is a condition for optimality, but cannot always be achieved by the cell in reality. This is due to two major facts:

1. The information content is imprecise, e.g. because of the cell only measuring the intracellular nutrient concentration.

2. The change between different environments happens faster than the cells ability to adapt to the desired value.

Consequently, actual and desired value cannot always be identical, as it is in the case of a stationary process. It is appropriate to assume a stationary process in order to compute the desired values. But on the matter of determining the actual value, one must consider real dynamics of fluxes as well as pool concentrations.

2.4.3 Defining the desired value

The desired value $\phi_j^*(\hat{t})$ at time $\hat{t}$ is defined by the relative enzyme mass $\phi_j(t)$ which the system targets for if the environmental conditions would remain constant for $t > \hat{t}$. The proximity of the actual value to the desired value depends on how long the environment remains fixed relative to the response time t_R of the system. The two following limiting cases are possible, by defining T as the average time over which environmental conditions stay constant.

- **No adaptation ($t_R >> T$):** The desired value changes at any time.
- **Total adaptation ($t_R << T$):** The desired value remains fixed until full adaptation (homeostasis).

The desired value can be defined by the stationary case of the regulatory enzyme pool dynamics in Eq. (2.8). Above, it was assumed that optimal resource allocation or constant relative enzyme pools, respectively, is a state desired by the system. Therefore, the desired value ϕ_j^* follows from the condition

$$\frac{d}{dt}\phi_j(t) \stackrel{!}{=} 0$$

and corresponds to the relative enzyme mass synthesis rate γ_j at time t.

$$\phi_j^*(t) := \gamma_j(t) = \frac{\frac{d}{dt}M_j(t)}{\frac{d}{dt}M^{\text{tot}}(t)} \tag{2.15}$$

The cell implements the desired value by adjusting (regulating) the synthesis rate ratio γ_j, as Eq. (2.15) shows. The system drives the enzyme mass ratio ϕ_j towards the synthesis rate ratio, regardless of the initial condition of $\phi(t_0)$, i.e. pathway mass $M_j(t_0)$ and total mass $M^{\text{tot}}(t_0)$.

$$\frac{M_j(t) + M_j(t_0)}{M^{\text{tot}}(t) + M^{\text{tot}}(t_0)} \stackrel{t\to\infty}{\longrightarrow} \frac{\frac{d}{dt}M_j(t)}{\frac{d}{dt}M^{\text{tot}}(t)}$$

In summary, the synthesis rate ratio can be regarded as the control function of the cell. By having the knowledge of the ratio γ_j, it is possible to predict the state dynamics of the whole metabolic system. Of course, the control function has to depend on the extracellular nutrient concentrations and therefore on environmental conditions.

	relative mass flux		normalized **absolute** flux
Metabolite	$\frac{1}{E^{\max}} \cdot \frac{d}{dt}[X] = \frac{d}{dt}\frac{m_j(t)}{M^{\text{tot}}(t)}$	$\overset{!}{=}$	$\frac{1}{M^{\text{tot}}(t)} \cdot \frac{d}{dt} m_j(t)$
Enzyme	$\frac{1}{E^{\max}} \cdot \frac{d}{dt} E_j = \frac{d}{dt}\frac{M_j(t)}{M^{\text{tot}}(t)}$	$\neq$	$\frac{1}{M^{\text{tot}}(t)} \cdot \frac{d}{dt} M_j(t)$

Figure 2-4: **Relative mass flux and normalized absolute mass flux.** Metabolic reactions happen on a much faster time scale than the rate of protein synthesis. Consequently, the relative mass flux and normalized absolute mass flux are unequal for enzymes, while they are identical for metabolites.

2.5 Determining the optimal desired value

2.5.1 Relative and absolute mass fluxes

One has to distinguish between relative mass fluxes and normalized absolute mass fluxes, as shown in Figure 2-4 and Figure 2-5.

$$v_5 = \sum_j \frac{\frac{d}{dt} M_j(t)}{M^{\text{tot}}(t)} \neq \sum_j \frac{d}{dt} \frac{M_j(t)}{M^{\text{tot}}(t)} = 0$$

While both quantities are identical for the metabolite fluxes v_1, v_2, v_3, v_4, they are totally different for protein mass fluxes. The reason for this is the following assumption: the time dependent change of the metabolite pools happens on a much faster scale than the rate of protein synthesis. Therefore, the protein pathway mass M_j and the total protein mass M^{tot} can be regarded as constants for the time dynamics of metabolite mass $m_X(t)$.

2.5.2 Objective function and stoichiometric matrix

It is assumed that the metabolic network is optimized in such a way that the system's growth rate is maximized [16, 19, 28]. In order to determine the desired value $\phi^*(t)$ of the relative protein mass at time t, the metabolic network, with normalized absolute mass fluxes $(dm_j/dt)/M^{\text{tot}}$ and $(dM^{\text{tot}}/dt)/M^{\text{tot}}$, has to be optimized with respect to

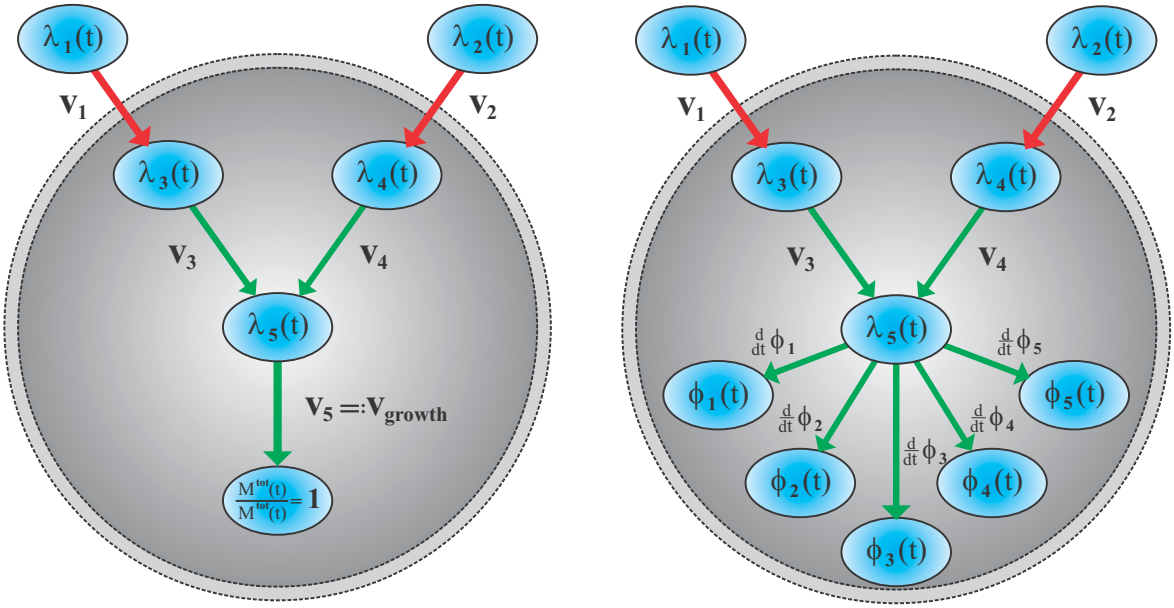


Figure 2-5: **Schematic figure of the simplified metabolic network.** (a) *Growth*: the arrows represent normalized absolute mass fluxes, while the metabolite and protein pools are quantified by normalized absolute mass. The growth rate $v_5 = v_{\text{growth}}$ is an absolute mass flux, since growth can only be understood in absolute terms. The normalization $1/M^{\text{tot}}$ is utilized to keep quantities independent of population size. (b) *Regulation*: the arrows represent relative mass fluxes, while the metabolite and protein pools are quantified by their relative mass. The self-replicator distributes its constrained protein resources between permeases ϕ_1, ϕ_2, metabolic enzymes ϕ_3, ϕ_4, and ribosomes ϕ_5. The enzyme synthesis acts as an feedback loop on the metabolic network, since metabolic fluxes v_j depend on enzyme levels $v_j \propto \phi_j$.

its growth rate (see Figure 2-5(a)). This optimization has to be applied for each time t. The growth rate v_{growth}, corresponding to fitness, is defined as normalized absolute protein mass flux (synthesis rate):

$$v_{\text{growth}} := v_5 = \frac{\left(\frac{d}{dt} M^{\text{tot}}(t)\right)}{M^{\text{tot}}(t)} \quad . \tag{2.16}$$

Absolute fluxes are of paramount importance, since growth can only be understood in absolute terms. Normalized fluxes, specifically the normalized protein flux, are utilized because they are independent of population size.

A stoichiometric matrix S with three metabolites and five fluxes can be formulated

for this metabolic network. As defined above, inflow fluxes are positive and outflow fluxes are negative.

$$S := \begin{pmatrix} +1 & 0 & -1 & 0 & 0 \\ 0 & +1 & 0 & -1 & 0 \\ 0 & 0 & +1 & +1 & -1 \end{pmatrix} \tag{2.17}$$

Using this matrix, the metabolite pool dynamic can be expressed as:

$$\frac{d}{dt}\vec{m}(t) = \begin{pmatrix} +1 & 0 & -1 & 0 & 0 \\ 0 & +1 & 0 & -1 & 0 \\ 0 & 0 & +1 & +1 & -1 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{pmatrix} \quad , \tag{2.18}$$

where

$$\vec{m}(t) := \begin{pmatrix} \lambda_3(t) \\ \lambda_4(t) \\ \lambda_5(t) \end{pmatrix} \quad . \tag{2.19}$$

2.5.3 Optimization conditions

The mathematical problem is to find the desired values $\phi_j^*(t)$ at time t for a given set of actual values of the metabolite pools and extracellular nutrient concentration. The actual values of the enzyme pools are not relevant for this purpose, since the cellular system drives towards the desired value, regardless of initial conditions of the enzyme pools. Since the growth rate is a flux, the desired relative protein masses ϕ_j^* need to be expressed in terms of desired metabolite fluxes v_j^*.

$$v_j^*(t) := \alpha_j^*(t) \cdot \phi_j^*(t) \quad , \tag{2.20}$$

where

$$\alpha_j^*(t) := \frac{\lambda_j^*(t)}{\frac{K_M^{(j)}}{E^{\max}} + \lambda_j^*(t)} \cdot k_j \quad \text{for } j = 1, ..., 5 \; . \tag{2.21}$$

The relative metabolite masses ${\lambda^*}_j$ at time t represent perceived values, which can differ with respect to the type of perception and need not to be equal to the real values $\lambda_j(t)$. The desired relative protein masses can be obtained by maximizing the growth rate for each time t under following conditions.

- **Positive fluxes:** The fluxes are constrained to be positive. (By constraining the lower boundary to a non zero value, one could simulate a basal enzyme expression level, which is not done here.)

- **Optimal resource allocation:** This condition implicates constant metabolite pools and hence balanced fluxes, as can be seen by setting the time derivative of all metabolite masses to zero (see Eq. (2.18)).

$$\frac{d}{dt}\vec{m}(t) \stackrel{!}{=} 0$$

$$S \cdot \vec{v}^*(t) \stackrel{!}{=} 0 \tag{2.22}$$

- **Proteome density conservation** (Molecular Crowding [10,42,56]) : The total amount of all enzyme pools summed up together is restricted, which arises from the assumed constant total enzyme concentration $E^{\max}$. Therefore, the sum of all relative protein mass is restricted by one,

$$\sum_{j=1}^{5} \phi_j^* = 1 \quad ,$$

from which follows, using Eq. (2.20),

$$\frac{v_1^*(t)}{\alpha_1^*(t)} + \frac{v_2^*(t)}{\alpha_2^*(t)} + \frac{v_3^*(t)}{\alpha_3^*(t)} + \frac{v_4^*(t)}{\alpha_4^*(t)} + \frac{v_5^*(t)}{\alpha_5^*(t)} = 1\,. \tag{2.23}$$

This density conservation constrains the allocation of cellular resources [1, 56, 57]. Our model basically incorporates a three component partition of the proteome [68], namely permeases ϕ_1, ϕ_2, metabolic enzymes ϕ_3, ϕ_4, and ribosomes ϕ_5. The cellular system has to distribute its constrained protein resources be-

tween those three components.

2.6 Perception

Perception is the key to proper regulation. Depending on the perceived extracellular nutrient availability, the system's controller regulates its metabolism differently. We define two kinds of perception, namely the extracellular and intracellular perception. In the case of extracellular perception the cell regulates its metabolism exclusively in response to extracellular nutrient information, while in the case of intracellular perception the opposite holds. In the latter case the cell indirectly recognizes nutrient availability by perceiving intracellular metabolic information.

Looking at Figure 2-3 one understands why extracellular perception effectively has to act as a feedforward loop while intracellular perception acts as feedback loop on the regulation. Assuming extracellular perception, the information about changes in external nutrient availability have already entered the controller before the cell is able to take them up. Thus, pathways are regulated in response to changes in the environment, even before nutrients enter the metabolism. Contrarily assuming intracellular perception, the information about external nutrients enters the controller not before nutrients have already been transported inside the cell. Thus, pathways are regulated in response to changes in intracellular nutrient concentrations, some time after the nutrient availability has changed in the environment. The cell indirectly perceives its environment and slowly adapts by a feedback control mechanisms.

The incorporation of perception into the above presented mathematical context is done by defining two types of proteome density conservation (see Eq. (2.23)) according to both perception types. Since, intracellular perception is equivalent to an exclusive information about intracellular metabolite pools, only the intracellular quantities α_3, α_4, α_5 enter the conservation equation of a system with intracellular perception.

$$\frac{v_1^*(t)}{\alpha_3(t)} + \frac{v_2^*(t)}{\alpha_4(t)} + \frac{v_3^*(t)}{\alpha_5(t)} + \frac{v_4^*(t)}{\alpha_5(t)} + \frac{v_5^*(t)}{\alpha_5(t)} = 1 \tag{2.24}$$

Extracellular perception is equivalent to an exclusive information about the extracellular nutrient availability. Therefore, only the extracellular quantities α_1 and α_2 enter the conservation equation of a system with extracellular perception.

$$\frac{v_1^*(t)}{\alpha_1(t)} + \frac{v_2^*(t)}{\alpha_2(t)} + \frac{v_3^*(t)}{\alpha_1(t)} + \frac{v_4^*(t)}{\alpha_2(t)} + \frac{v_5^*(t)}{(\alpha_1(t) + \alpha_2(t))} = 1 \tag{2.25}$$

2.7 Determining the actual value: protein synthesis & metabolism

To determine the actual values of the enzyme and metabolite pools, the metabolic network (Figure 2-5(b)) with relative mass fluxes λ_j and ϕ_j has to be used. The actual system can be modeled by a system of 10 coupled ordinary differential equations:

$$\frac{d}{dt}\lambda_j(t) = \alpha_Y(t) \cdot \phi_Y(t) \; - \; \alpha_j(t) \cdot \phi_j(t) \tag{2.26}$$

$$\frac{d}{dt}\phi_j(t) = v_{\text{growth}}(t) \cdot \left[\phi_j^*(t) - \phi_j(t)\right] \, , \tag{2.27}$$

where

$$v_{\text{growth}}(t) = \alpha_5(t) \cdot \phi_5(t)$$

and

$$\alpha_j(t) = \frac{\lambda_j(t)}{\frac{K_M^{(j)}}{E^{\max}} + \lambda_j(t)} \cdot k_j \quad .$$

Here, the index Y denotes the upstream metabolites and enzymes.

2.8 Simulation

To evaluate the fitness benefit due to perception in dependency of environmental fluctuations, a competing species experiment in a fluctuating environment was simulated. While each species exclusively perceives its environment according to intracellular or extracellular perception, the metabolic and regulatory mechanisms are similarly based on the above presented mathematical model. Hence, the only difference between both

species is the perception type.

The computer simulation of each species is implemented according to the block diagram Figure 2-3, which produces the dynamic behavior of growth rate, enzyme and metabolite concentrations, and relative protein synthesis rate (control function). The metabolic network is regulated by a flux balance analysis (FBA) based optimization process [28, 52, 59] (control process), which maximizes cellular growth rate [19] with respect to constant proteome density [42, 60] and optimal enzyme-resource allocation [16]. Particularly, our simulation of the pool dynamics can be regarded as some type of dynamic FBA with quasi-steady-state assumption. This assumption includes discretizing the time into time intervals Δt of constant growth rate $v_{\text{growth}} = \text{const.}$ and regulation (control) $\gamma_j = \text{const.}$, whereas the former is kept constant for the enzyme dynamics only. During an interval Δt, the enzyme and metabolite levels are variable and determined by the system of coupled differential equations, Eq. (2.26) and Eq. (2.27). At the end of each time step Δt the controller computes the desired enzyme levels γ_j by linear programing within FBA on the basis of the perceived metabolite levels λ_j^* (Eq. (2.24) or Eq. (2.25)). Finally, the updated growth rate v_{growth} and regulation γ_j are taken to repeat this procedure for the next time step. The difference of our simulation to conventional dynamic FBA [34, 38] is the notion of a control system, which is rather an element of cybernetic modeling [70].

To obtain regulatory and growth dynamics of the cell which are independent of initial conditions, the simulation operates until both species show a stable periodic behavior. The process of obtaining a stable periodic behavior simulates an evolutionary process in which the cell adapts to an environment with highly predictable fluctuations in nutrient availability. Having attained stability, one periodic growth rate interval is taken to compute the average growth rate, which is the measure for fitness. The whole procedure is repeated for different fluctuation frequencies and therefore yields a frequency dependent plot of the average growth rate. In conclusion, the computer simulation delivers a frequency dependent plot of the species' fitness as well as the underlying dynamic behavior of metabolism and regulation.

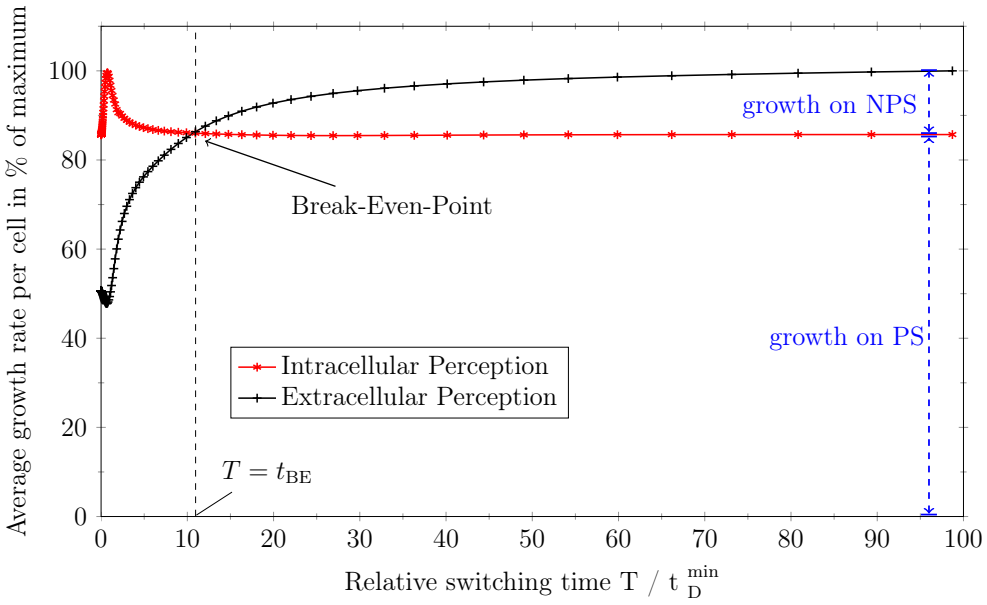
Chapter 3

Results

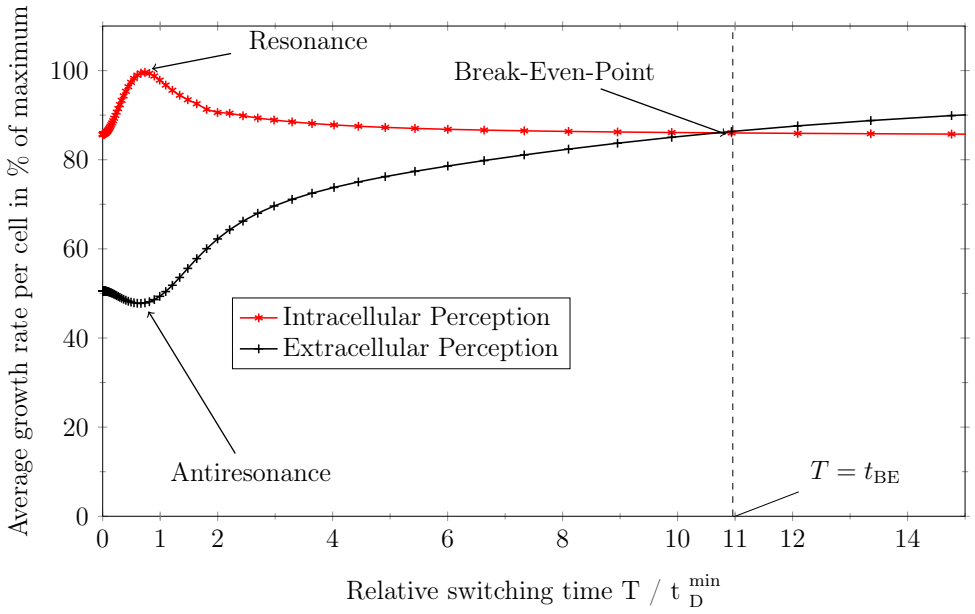
3.1 Simulation: average growth rate for different switching times

To determine the frequency regimes in which the intracellular perception is evolutionary more beneficial than the extracellular perception, the average growth rate of the *intracellular perceiving system* (IPS) and *extracellular perceiving system* (EPS) was plotted against the relative switching time $T/t_D^{\min}$, as can be seen in Figure 3-1.

The modeled self-replicators live in a highly predictable ecology, which fluctuates between two environments, namely the non-preferential sugar (NPS) and the preferential sugar (PS) environment. Both sugar types are always present, whereas their concentration fluctuates with respect to the environment. In the NPS environment the NPS possess 50% of the maximum sugar concentration, while the concentration of the PS is as low as 0.25%, which can be regarded as zero. In the PS environment the NPS concentration immediately decreases to 0.25%, while the PS transfers to a maximum concentration of 100%. For preferentiality in vitro, a difference between maximum PS and NPS concentration is not necessary, because changes in fluxes are caused by the quality of the sugar types, i.e the uptake rate. Nevertheless, this model feature guarantees sugar preferentiality in silico without loss of generality. The duration of one environment, PS or NPS, is called the switching time T. The reciprocal



(a)



(b)

Figure 3-1: **Simulation results of competing species experiment in a fluctuating nutrient environment.** Average growth rate for different relative switching times $T/t_D^{\min}$ and perception types, whereas $t_D^{\min}$ denotes the minimum cellular doubling time. The average growth rate is normalized by its maximal observable value for the sake of generality. The dashed black line at the break-even point t_{BE} divides fluctuating environments in regimes of fast $T = [0, t_{\mathrm{BE}}]$ and slow $T =]t_{\mathrm{BE}}, 100]$ fluctuations. (a) Average growth rate for the interval $T/t_D^{\min} = [0, 100]$. While the self-replicator with intracellular perception only grows on preferential sugar (PS), the one with extracellular perception also grows on non-preferential sugar (NPS). These contributions to the average growth rate can be seen for the steady state value. (b) Average growth rate for the interval $T/t_D^{\min} = [0, 15]$.

value of the switching time is exactly the frequency $f := 1/T$ of the fluctuations.

To gain a more general view all time quantities are normalized by the minimum cellular response time $t_R^{\min} = const.$, which corresponds to the time the cellular system needs to adapt to a constant PS environment. The response time is defined as the time, the cellular system needs to finish 50% of its regulatory work. Specifically it is the average time, the relative enzyme masses $\vec{\phi}$ need to reach half the way between initial $\vec{\phi}(t_0)$ and desired value $\vec{\phi}^*$.

$$\vec{\phi}(t_R) := \frac{1}{2} \cdot (\vec{\phi}^* - \vec{\phi}(t_0)) \tag{3.1}$$

Here, the initial values at time t_0 are the steady state values in the NPS environment. While the minimum response time gives an upper speed limit of cellular adaptation to changing nutrient availability, the cellular doubling time is experimentally more accessible. Assuming no protein degradation, the minimum response time $t_R^{\min}$ measures approximately the time a cell needs to double itself once in a constant PS environment, i.e. the minimum cellular doubling time $t_D^{\min} = const.$ (minimum generation time) [4]. This minimum doubling time $t_D^{\min}$ is constant and corresponds to the maximum growth rate that is achievable. Hence, normalization by the minimum response time can be interpreted as normalization by the minimum cellular doubling time generating the relative switching time $T/t_D^{\min}$ and relative time $t/t_D^{\min}$. These quantitates produce an organism-independent view on average growth rate and regulatory dynamics, which makes Figure 3-1 valid for all exponentially growing microorganisms.

Each point in Figure 3-1 represents the average growth rate for a given relative switching time, that is for a given fluctuation frequency. The average growth rate $\overline{v}_{\text{growth}}(T)$ is defined as the time integral over the growth rate dynamics $v_{\text{growth}}^{(T)}(t)$ divided by one period of fluctuations, specifically twice the switching time.

$$\overline{v}_{\text{growth}}(T) := \frac{1}{2T} \int_{t_0}^{t_0+2T} v_{\text{growth}}^{(T)}(t)\, dt \tag{3.2}$$

There are four switching time points, which are of interest for a qualitative analysis of the average growth rate. These are (i) T approaching zero, (ii) T around the

minimum response time (minimum doubling time), (iii) T at the break-even point t_{BE}, and (iv) T approaching infinity. The break-even point divides Figure 3-1 into two regimes, which are the fast fluctuating regime $T \in]0, t_{BE}]$ and the slowly fluctuating regime $T \in [t_{BE}, \infty[$. Inside the first regime the IPS has a larger average growth rate, whereas the EPS grows faster in the second one. For infinitely large switching times, the cells go into steady state. The steady state average growth values can be assigned to contributions due to full adaptation to the PS or NPS environment. As will be seen below, the IPS only adapts to the PS environment, which is equivalent to a cellular system under permanent carbon catabolite repression. Therefore, its steady state value in average growth rate is the contribution $\overline{v}^{PS}_{\text{growth}}$ due to exclusive PS adaptation. The EPS adapts fully to both sugar types when in steady state and will only utilize carbon catabolite repression if there are relevant amounts of PS in the environment. Thus, the difference between steady state values of EPS and IPS is exactly the contribution $\overline{v}^{NPS}_{\text{growth}}$ caused by adapting completely to NPS surrounding. The contribution to exclusive adaptation to the PS environment has to be larger than the one for the NPS environment, because this is actually the definition of sugar preferentiality. In the here presented environmental example, $\overline{v}^{NPS}_{\text{growth}} = 15\%$ and $\overline{v}^{PS}_{\text{growth}} = 85\%$. In conclusion, the intracellular perception, yielding permanent carbon catabolite repression, is evolutionary more beneficial for switching times $T \in]0, t_{BE}]$ and the extracellular perception is more beneficial for $T \in [t_{BE}, \infty[$.

3.2 Simulation: actual value

To understand the underlying regulatory principles of the results of Figure 3-1, the control, enzyme pool, metabolite pool and growth rate dynamics were analyzed at representative relative switching time values. The control dynamics can be understood as the dynamics of the relative protein synthesis rate $\vec{\gamma}(t)$.

3.2.1 Mixed environments ($T \to 0$)

If the switching time converges towards zero, the cellular system will no longer be able to distinguish between the two environments. Therefore, the cell will perceive a mixed environment. Further, the cell has no time to adapt to any individual environment, since the nutrient fluctuations are much faster than the minimum response time ($T << t_R^{\min}$). There are two regulatory ways to handle this situation, used by the EPS and the IPS, respectively. First, the cell can go into a mixed state, which responds to both environments at the same time. Because of limited resources, according to constant proteome density, the cell adapts partly and gains only half, 50%, of its possible average growth rate (see Figure 3-1 for $T/t_D^{\min} \to 0$). This is the regulatory principle of the EPS. Secondly, the cell can go into and stay in the state of the preferential sugar (PS) environment. This gives rise to no nutrient uptake in the NPS environment and a maximum nutrient uptake in the PS environment. Due to this one-sided adaptation to the PS, the cell gains an average growth rate below the maximum(100%) but higher than 50%. This is the regulatory principle of the IPS.

3.2.2 Resonance and antiresonance point ($T = \tau \approx t_R^{\mathbf{min}} \approx t_D^{\mathbf{min}}$)

If the switching time approaches the minimum response time $t_R^{\min}$ approximated by the minimum cellular doubling time $t_D^{\min}$, the regulatory effects will be observable. By approaching $t_D^{\min}$ another quantity becomes relevant, namely the time delay τ due to nutrient signaling, which is considered to be approximately equal to $t_D^{\min}$. This signaling time delay reflects the adaptation kinetics of the underlying metabolic network and thus is present in both systems, EPS as well as IPS. After changing the relative enzyme masses, it takes this time delay to observe an effect on the growth rate. Thus, any regulatory action will take effect only after τ. Moreover, the IPS needs this time to perceive its surrounding, before even being able to take any proper regulatory steps.

While the EPS perceives its nutrient environment in an exact and instantaneous manner, the IPS has a limited and delayed vision of its surrounding (see Figure 3-2(a),

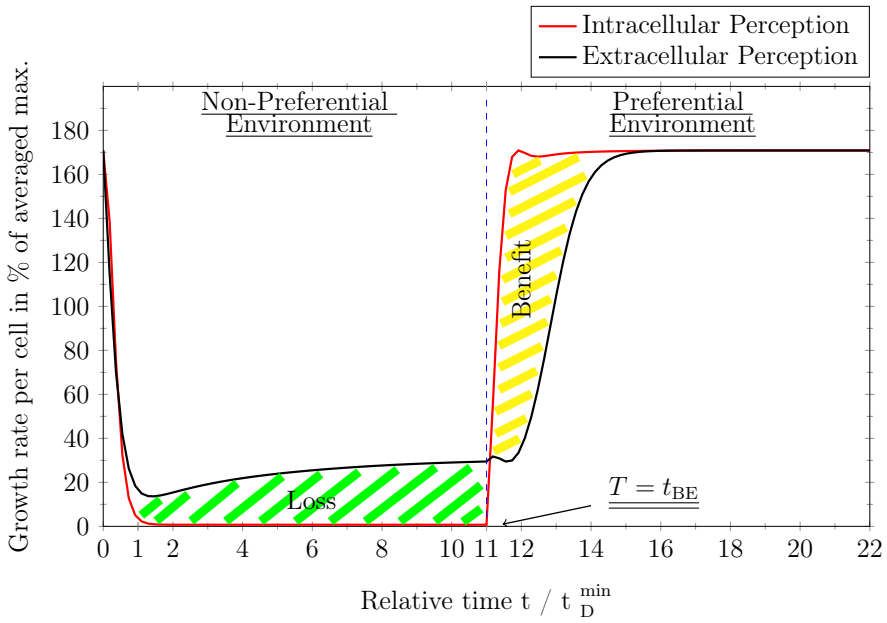
where the IPS does not grow at all in the NPS environment and Figure 3-3(a), where there is no NPS uptake at all). There are two main features that distinguish the IPS from the EPS or the intracellular perception from the extracellular perception, respectively. First, the IPS has to wait for a time delay τ until the nutrient signaling affects the intracellular metabolic pools, in order to sense what has happened externally. Secondly, the IPS deactivates the NPS pathway, which prevents the system to perceive NPS. Hence, the IPS is not able to sense the switching between environments with $T \leq t_D^{\min}$. Based on these perception types, the EPS adapts to each individual environment whereas the IPS adapts to the one with PS, only. Additionally, the IPS prepares itself for an increased PS uptake by hyper-up-regulation of PS uptake transporters during the NPS environment. This increased PS uptake only occurs for a short time interval (see Figure 3-3(b)), so that an environmental change with a switching time similar to the signaling time delay produces a resonance effect (see Figure 3-2(b)).

Considering a switching time that equals the signaling time delay $T = \tau \approx t_D^{\min}$, the EPS yields an antiresonance effect with the wrong pathway regulation at the wrong time. This effect generates the worst average growth rate possible ($< 50\%$), whereas the growth is even smaller than for adapting to both sugar types simultaneously in a mixed environment. In contrary, the IPS supplies the perfect regulation with the best possible result (100%), resulting from a resonance effect. Concluding, if there is no time for regulation to act on growth rate, it will be beneficial to focus on PS and use the PS gap phase to prepare for the PS environment.

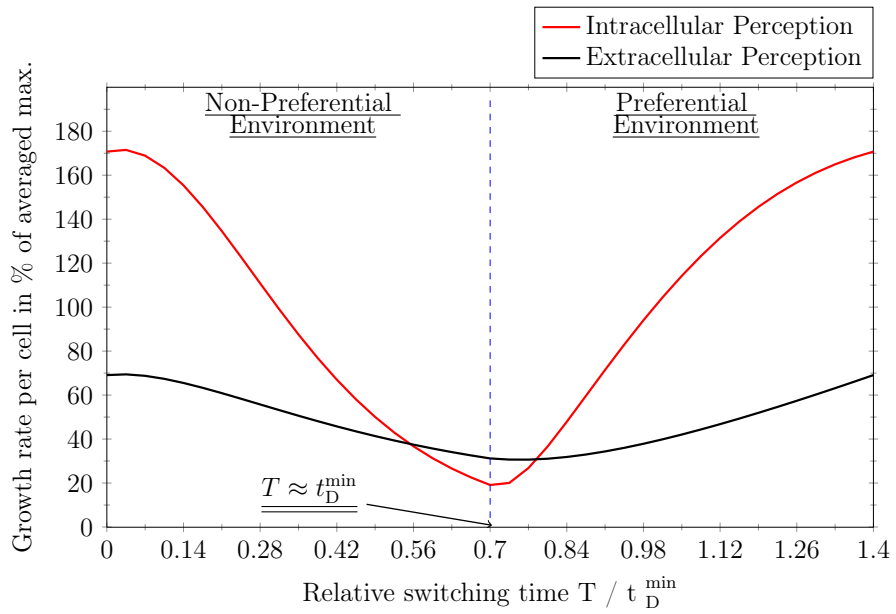
3.2.3 Break-even point ($T = t_{BE}$)

If the switching time T approaches the break-even point t_{BE}, the EPS will approach the IPS in average growth rate. Specifically, the Growth Benefit, due to adaptation to the NPS environment, will exceed its associated growth loss.

The IPS perceives correctly the extracellular PS concentration in both environments and imposes an constant activation of the PS pathway. On the other hand, the EPS alternately activates and deactivates NPS and PS pathways to adapt to the



(a)



(b)

Figure 3-2: **Growth rate dynamics at the break-even point and resonance point.** The plot shows one period $2T$ of fluctuations between non-preferential and preferential environment, whereas the dashed black line separates both environments (periodic boundary conditions). Time t is normalized by the minimum cellular doubling time $t_D^{\min}$. (a) Growth benefit and loss of intracellular perception due to exclusive adaptation to preferential sugar. The area between both graphs is the measure for benefit and cost relative to both perception types. (b) Growth dynamics at the resonance point $T/t_D^{\min} = 0.7 \approx 1$. The large amplitude of the growth rate fluctuations for intracellular perception leads to an optimal average performance and is caused by the resonance of cellular response time with switching time T between environments.

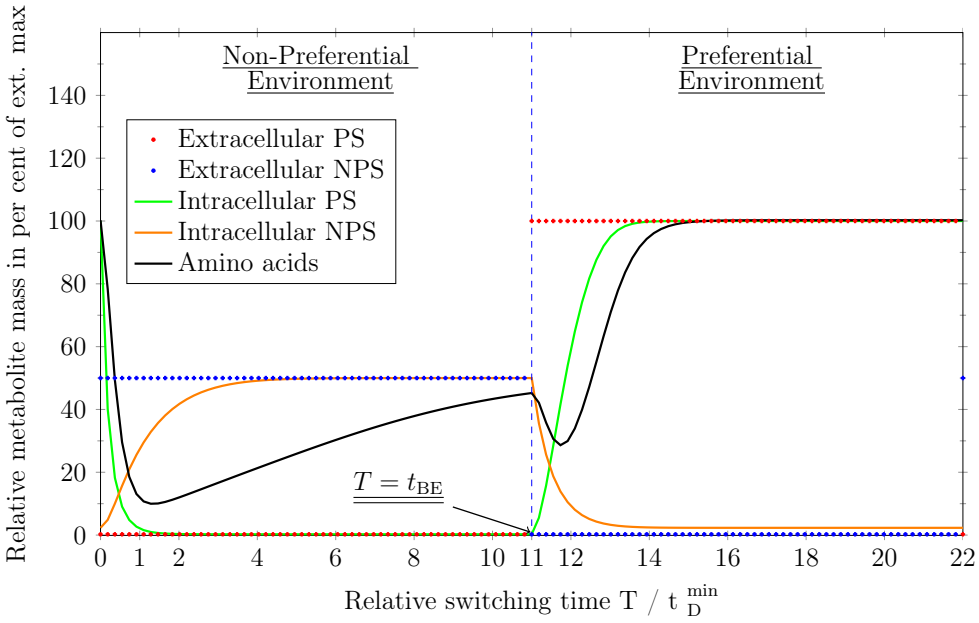
environment. There is enough time for full response and the EPS can implement the control function (desired value) into reality. Nevertheless, the IPS stays fixed inside the state of hyper-up-regulation throughout the whole NPS environment. After its amino acid pool becomes zero, there is no driving growth rate to regulate pathways.

To understand, why EPS and IPS approach the same average growth rate at the break-even point one has to understand the concept of growth benefit and growth loss due to the underlying regulatory strategy. Simply spoken, the IPS only uses the PS pathway to grow, while the EPS uses both pathways. To decide which of the strategies is more favorable, one has to explain for which cases using two pathways is more favorable than only one. The advantage of the IPS is that it does not need to adapt to the PS, so it gains a maximal growth rate while the EPS still is adapting to the new environment. This represents a growth benefit for the IPS (see Figure 3-2). The advantage of the EPS is that it can also grow in the NPS while the IPS goes into a type of growth arrest. This represents a growth loss for the IPS (see Figure 3-2). Concluding, these two growth effect are exactly equal at the break-even point.

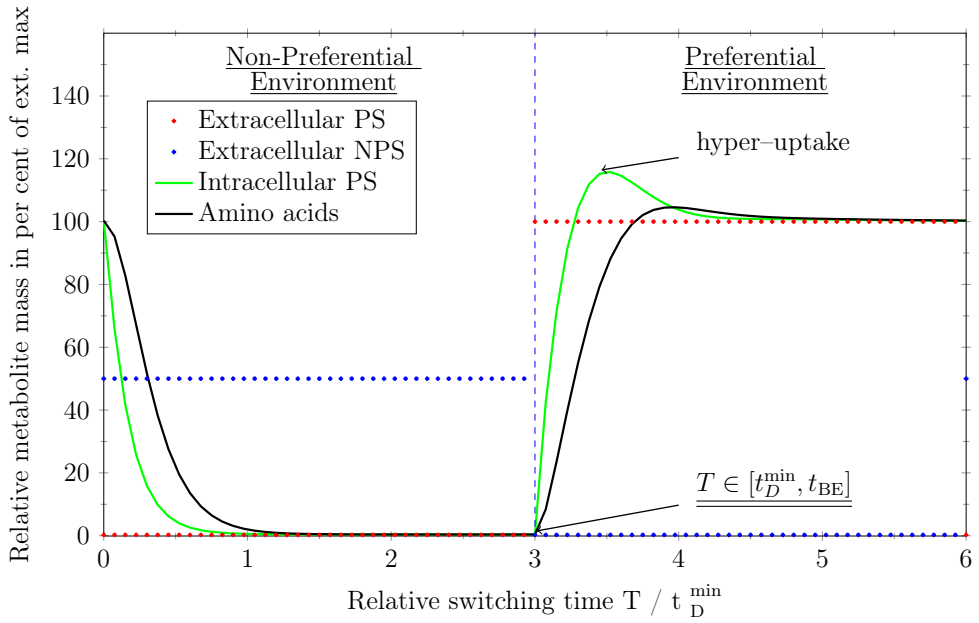
3.2.4 Steady state ($T \to \infty$) & limits of the model

If the switching time becomes larger than the break-even point and approaches infinity, the extracellular perception and therefore the EPS will have the dominant strategy. The cells enter steady state, therefore balanced fluxes, optimal resource allocation and constant metabolite pools are realized. The latter feature can be seen in Figure 3-3, where the metabolite pool concentrations converge to the one of extracellular nutrients.

A switching time that lasts an infinitely long time is the equivalent of an nutrient environment that stays constant and does not fluctuate at all. On one hand, the EPS imposes full adaptation to the respective nutrient environments, which intuitively makes sense for an infinitely large switching time. On the other hand, the IPS only adapts to the PS and thus resides in growth arrest during NPS surrounding (Figure 3-2(a)). The IPS traps itself in using PS until this resource is exhausted. More reasonable, the resulting drop in growth rate should promote the transition to strin-



(a)



(b)

Figure 3-3: **Metabolite pool dynamics.** The plot shows one period $2T$ of fluctuations between non-preferential and preferential environment, whereas the dashed black line separates both environments (periodic boundary conditions). Time t is normalized by the minimum cellular doubling time $t_D^{\min}$. (a) Extracellular perception at break-even point: both sugar types, preferential (PS) and non-preferential (NPS), are taken up. The condition of constant metabolite pools, caused by optimal enzymatic resource allocation, is approached for switching times T larger than the break-even point t_{BE}. (b) Intracellular perception at $T/t_D^{\min} = 3$ between resonance point and break-even point: only PS is taken up with an increased PS uptake during the PS environment, which is the cause for the optimal growth at the resonance point.

gent response, right after the break-even point t_{BE}. Stringent response would enable the IPS to activate the NPS pathway by bypassing the limited nutrient perception. Then, the average growth rate of IPS would probably converge to the one of the EPS.

Since our research questions is asking for regulatory principles in fluctuating environments, the case of infinitely large switching time is not relevant. It is only necessary to understand the limits of this model. After the break-even point t_{BE}, the model system makes no valid predictions for the IPS. It has no stringent response and thus can theoretically grow on the smallest amount of PS, which is 0.25% in the here presented example. This is a physiological unrealistic case. In order to add stringent response to the model, a constraint on the minimal detectable nutrient concentration could be introduced. If the PS concentration goes below this constraint, stringent response will be turned on.

Chapter 4

Discussion

This study indicates that indirect intracellular perception of extracellular nutrient availability can give rise to a growth benefit under situations where the up and down regulation of pathways cannot follow the fast changes of the nutrient environment. Although intracellular perception carries less information about the actual environmental conditions, this regulatory mechanism enables exponentially growing organisms to gain maximal average growth if nutrient concentrations fluctuate on timescales comparable to the minimum generation time.

In our simulation, a system with intracellular perception responds to strong fluctuations by keeping preferential nutrient pathways activated and non-preferential pathways inactivated. As a result the cell can take up preferential nutrients as soon as they are available without any prior regulation. This regulatory strategy is a good example for *minimal adjustment*. According to Schuetz et al. [55] there is a trade-off between *optimality* under one given condition and *minimal adjustment* between different conditions, i.e. Pareto optimality [49]. In other words, cells will tune metabolic pathways to obtain optimal growth if surrounded by a constant environment. Contrarily, in a fluctuating environment, cells will regulate their pathways to respond to environmental changes by minimal adjustment of pathways. In this sense, intracellular perception gives rise to a regulation of *minimal adjustment*, which is dominant under fast environmental changes. Additionally, our results show that the notion of optimality is also given under fluctuating conditions, since minimal adjustment is a

consequence of maximizing an objective function averaged over the range of conditions.

Moreover, our model of intracellular perception is in agreement with the phenomenon of carbon catabolite repression [17, 23], if cells are not able to distinguish between different conditions anymore, i.e. the fluctuation frequency approaches infinity. This situation is equivalent to a mixed constant environment. While carbon catabolite repression reflects the cell's affinity to preferential sugars in a stable mixed nutritional surrounding, our results indicate that this mechanism holds under fast fluctuations (around the minimum generation time) as well. To our knowledge, CCR has not been obtained from an mathematical optimization process, before.

Furthermore, our simulation of the growth dynamics produced a break-even point, where the average growth rate of the IPS and EPS are equal (Figure 3-1). At this point the growth benefit of the IPS in the preferential environment matches the growth loss in the non-preferential environment. Growth benefit and loss arise from the exclusive adaptation to the PS environment (Figure 3-2(a)). This is in agreement with the experimental work of Mitchell et al. [44], who have observed anticipation of environmental changes in the sugar metabolism of *E.coli* and *S.cerevisiae*. Mitchell et al. classified the regulatory response to environmental changes into direct and anticipatory regulation, whereas the former regulates its metabolism in direct response and the latter in advanced preparation. Further, they state that an anticipatory response will be evolutionary beneficial if "the benefit gained from anticipation exceeds the cost of early preparation". We can identify the anticipatory regulation with the IPS and the direct regulation with the EPS. As we have shown intracellular perception yields a preparation for the PS environment during the NPS environment, which can be regarded as an anticipatory behavior. Especially, the hyper-up-regulation of the PS uptake transporter in the presence of NPS environment, which results in the resonance peak of the average growth rate, serves as a good example for anticipatory regulation. This course of action is only beneficial for fluctuating environments with frequency smaller than the break-even frequency. Thus, anticipatory behavior in a highly predictable fluctuating environment can be understood by limited and delayed

intracellular perception.

Using our phenomenological computer model, we further showed that extracellular perception is of selective advantage under slow environmental fluctuations. However, it is reasonable to assume that intracellular perception always contributes to some extent to growth control. This hypothesis is supported by the observations of New et al. [51], who have shown that wild *S. cerevisiae* strains divide into sub-populations of specialist and generalists according to their growth rate related response time (lag phase). Generalist will adapt faster to a new carbon environment than specialists if the environment changes from a preferential to a non-preferential carbon source. Our results in Figure 3-2(a) for the non-preferential regime exhibit the same relation between growth regulation by means of extracellular perception (EPS) and intracellular perception (IPS). The EPS, like the generalists, adapts faster to the non-preferential environment than the IPS. In this context generalist could be seen as microbes whose growth control mainly depends on extracellular perception, while the contribution of intracellular perception has an bigger impact on the specialist's growth control. Although, both perception types can be utilized by microorganisms, their contribution to growth control can be differently depending on the individual evolutionary background.

Regarding the IPS, an interesting result of our simulation is the existence of a resonance peak for fluctuations around the minimum generation time. At this peak, the time delay in nutrient perception equals the switching time between environments resulting in optimal fitness. The data-based mathematical model of Mitchell and Pilpel [43] supports our finding as their cellular system shows a fitness peak around $1 - 2.5$ generation times.

To summarize, our work indicates that intracellular perception is of selective advantage and gives rise to CCR in oscillating environments, so that microbes specialize on the preferential nutrient and anticipate it in its absence. In general, intracellular perception could be a fundamental regulatory principle of minimal adjustment to changing conditions. Although our study is limited to a purely qualitative conclusion, due to the simplicity of our approach, the presented model is sufficient to gain insight

in the fundamental differences of microbial growth control. In following projects, it would be worthwhile to test our simulation with real metabolic networks, like from the model organisms *E.coli* or *S.cerevisiae*. Moreover, experimental evidence, i.e. competing species experiments, is needed to confirm our theory of the dominance of intracellular perception under fast fluctuations.

Part II

Inference of Biological Network Structure from Perturbation Data

Chapter 5

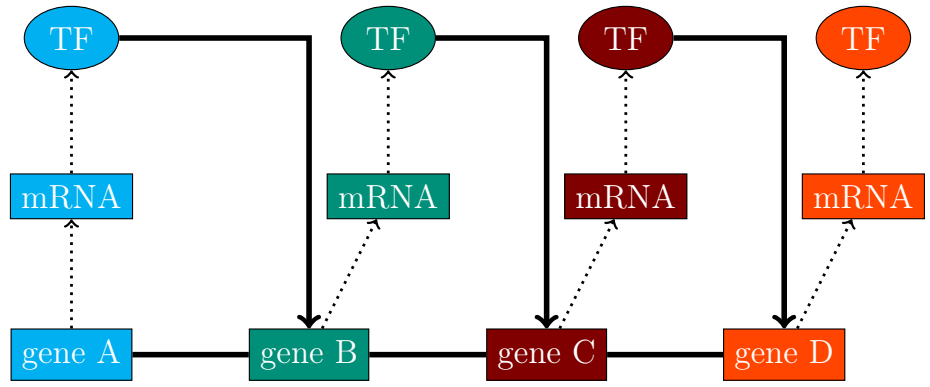
Introduction

In part 1 of this thesis, the regulation of metabolic networks was examined, while in part 2 the focus lies on the other two types of intracellular biological networks, i.e. signal transduction networks (STN) and gene regulatory networks (GRN). From a methodological perspective, in this second part the network structure and its interaction rate parameters are unknown, while the previous part assumed that the underlying molecular interactions are known. In order to predict or even control complex regulatory behavior of cells, like the cell response to environmental changes, the underlying molecular interaction pattern must be understood. This is exactly the general topic and purpose of part 2 of this thesis.

5.1 Biological networks

5.1.1 Gene regulatory networks

Gene regulatory network (GRN) control the main aspects of life, like cell differentiation, metabolism, the cell cycle and signal transduction [27]. They illustrate how the expression of genes are indirectly controlled by others and organized in a network like manner. Gene regulation is mostly a response to changes in the cellular environment, which can be triggered by growth factors or nutrient availability, but also by intracellular changes like DNA damage. Cellular response to these changes



(a) Schematic view of a simple gene regulatory network. Genes are transcribed to mRNA, which are translated to transcriptions factors (TF). The TF can promote (activate) or repress (deactivate) gene expression of downstream genes.

(b) Mathematical graph representation: The whole transcriptional and translational component belonging to one gene are effectively combined in a single node which can influence other nodes. Filled nodes stand for external perturbation, like gene knockout or changes in the genes' expression levels.

Figure 5-1: **Example of a simple linear gene regulatory network.** Bold arrows represent causal interaction between molecular components, i.e. transcriptions factors interact with genes which can activate or deactivate gene expression.

consists of co-expressing (activating) a set of genes that code for proteins associated with a specific cellular response. Co-expressed genes are often co-regulated by the same transcription factor [3]. In bacteria co-expressed genes are often organized in operons, which only have a single promoter region. Thus a single master regulator, i.e. transcription factor (TF), can initiate transcription of the whole gene set. The Lac operon of E.coli is a famous example [2,24], which codes for the protein machinery that takes up lactose from the environment and metabolizes it. In eukaryotic cells the loci of co-expressed genes are scattered over the whole genome, but still most genes are activated by a few master regulators (e.g. p53). The notion of master regulators and sparse networks otherwise leads to the assumption that GRN are scale free networks. In fact, most biological networks are assumed to be scale free networks, which means that the number of interactions k (degree) per node is distributed according

to a power law $p(k) \propto k^{-\gamma}$ [8]. This degree distribution implies very few master regulator genes, called hubs, that regulate the vast amount of genes, while most genes are linked very sparsely to the rest of the GRN.

The direct causal molecular interplay between genes in a regulatory network is highly complex, due to different types of biochemical and physical reactions and interactions that effectively cause gene A to activate (promote) gene B - as it is sketched in Figure 5-1(a). Measuring these interplays directly would involve a variety of high through-put experiments that are specialized to measure the different molecular components. For a genome wide approach this is mostly not feasible financially as well as experimentally. State of the art RNA sequencing techniques [63] produce more easily and accurate genome wide transcriptome data, which can be utilized to infer gene regulatory networks by exclusively focusing on one molecular component of the complex regulatory interplay .

By focusing only on the mRNA as a proxy, it is possible to mathematically represent complex GRN as a graph with directed edges, the so called directed graphical model (Bayesian networks). As sketched in Figure5-1(b) network nodes summarize the whole molecular machinery that is involved in interacting with the next genes. In the simplified example of Figure5-1 a chain is sketched in which node A activates (promotion of gene expression) or deactivates (suppression of gene expression) node B. The interaction is symbolized by an direct edge from node A to node B, indicating that the interaction is not reversible. Probabilistically, each node is denoted by a random variable which is chosen to be the mRNA abundance of the associated gene. Directed edges, also called links, between nodes represent conditional dependencies between random variables.

Choosing mRNA as a proxy for gene interactions has the disadvantage that any approximation bears, namely the uncertainty in the results. This will be discussed in more detail in the next section 5.2. But first signal transduction networks will be shortly introduced. Despite different characteristics, they have enough similarities to be inferred by the same method introduced in chapter 6.

5.1.2 Signal transduction networks

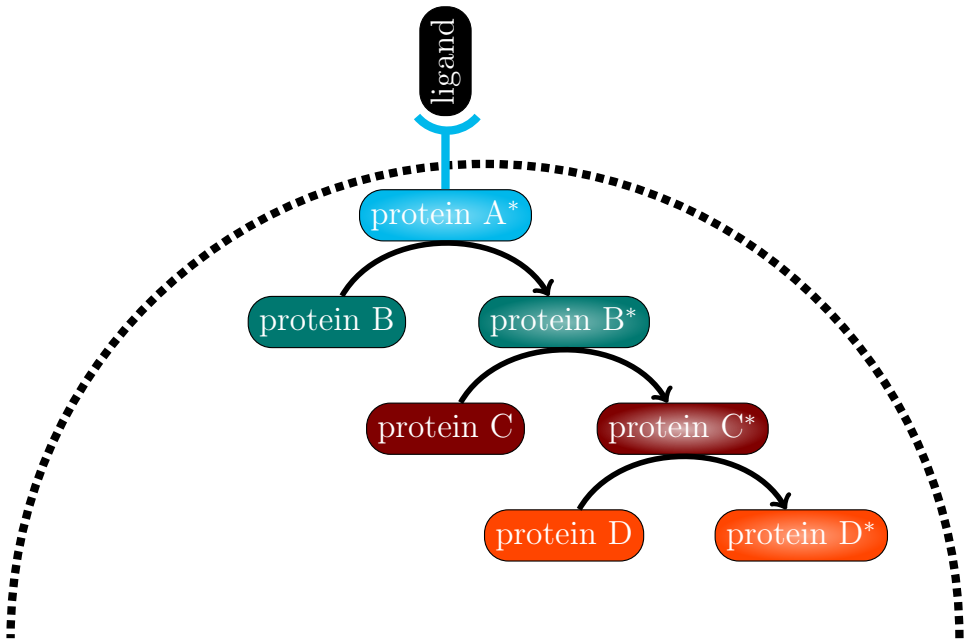
Cells of multicellular organisms need to communicate with one another in order to function properly and to respond to environmental changes as a whole. Therefore, there is a division of labor among differentiated cells inside a multicellular organism, where different cell types respond differently to the same extracellular signal. The differentiation of a cell is manifested by the expressed and not expressed genes, keeping in mind that all cells posses approximately the same DNA. Hence, signaling proteins and pathways depend on the cell type or differentiation, respectively.

While extracellular signal molecules mediate the communication between cells, intracellular signaling pathways forward the signal into the cell and initiate cellular response. An intracellular signaling pathway basically consist of three parts, namely reception, signal transduction, and initiation of cellular response. In the first step extracellular signal molecules (ligands) bind to receptor proteins on the outside of the plasma membrane (see Fig.5-2(a)), and activate intracellular signaling proteins. During signal transduction, signaling proteins are activated or inhibited (deactivated) in a cascade like manner which results in signal propagation through the cell. These signaling proteins act as molecular switches which mostly are activated by either phosphorylation or GTP binding. During signal transduction via a phosphorylation cascade, upstream signaling proteins act as kinases which phosphorylate downstream proteins. In the last step cellular response to the extracellular stimulation is initiated by activating effector proteins. Effector proteins can alter metabolism, gene expression, or cell morphology and movement. In the case of gene regulation, whole gene regulatory networks are usually activated.[1]

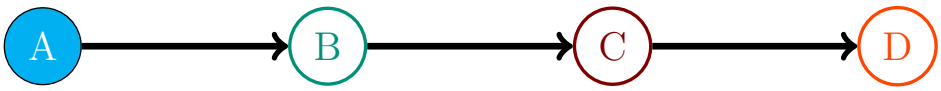
The phrase pathway in signaling pathway is somehow misleading, since in reality it is a complex network rather than a linear interaction scheme. In addition to the pathway's own complexity, there are cross-talks between different pathways associated with other extra- and intracellular signals. An example is the Akt-pathway, which has crosstalk to the MAPK pathway [6].

A major challenge is to infer protein-protein interactions of signal transduction

[1]This whole paragraph is based on [2]



(a) Schematic view of a simple signal transduction network, i.e. a phosphorylation cascade. Signaling proteins have an active A^* and inactive state A, whereas active signaling proteins can activate inactive proteins downstream. Ligands can bind extracellularly to membrane receptors to transmit a signal to the cytoplasm. The dashed semi-circle symbolizes the cell membrane.



(b) Mathematical graph representation: Filled nodes stand for external perturbations, e.g. inhibiting the kinase activity by drugs. The node activity is given by the abundance of active phosphorylated signaling proteins.

Figure 5-2: **Example of a simple linear signal transduction network, i.e phosphorylation cascade.** Bold arrows represent causal interaction between molecular components, i.e. active signaling proteins which can activate (phosphorylate) or deactivate (de-phosphorylate) downstream proteins.

from variations in the phosphor-protein abundance in response to inhibitory drugs and activating ligands. A mathematical representation for signal transduction networks are directed graphical models (Bayesian networks) that can explain direct causal interactions between signaling proteins. In Figure Fig.5-2 a simplified sketch of a phosphorylation cascade can be seen, where activated phospho-protein A^* serves as a kinase for deactivated (de-phosphorylated) signaling protein B. In the graph representation, each active phospho-protein is symbolized by a network node, while the respective protein abundance is understood as random variable - reflecting the probabilistic nature of the problem. Causal interactions between signaling proteins, i.e. phosphorylations, are mathematically captured by directed links representing conditional dependencies between random variables, i.e. protein abundance.

5.2 Network inference

5.2.1 The purpose of network inference

The goal of network inference is to reverse engineer the network structure from node activity data. The phrase "reverse" indicates that node activities are actually the effect of the networks' complex interaction scheme. They can naturally be computed if the engineered system, determined by network structure and interaction parameters, is known. Therefore, finding the real cause to the measured effects and thereby gaining knowledge about the underlying interaction scheme is the fundamental goal of network inference.

The systems of interest in this work are gene regulatory networks (GRN) and signal transduction networks (STN), both engineered by evolution and therefore a priori unknown. In these two biological systems the measured effect is the mRNA or phospho-protein abundance, respectively. Going forward they will be referred to as node activity, according to the terminology of graphical models.

To distinguish cause from pure correlation is a main problem which all reverse engineering tasks have in common. A correlation measures the statistical relationship

between two quantities and therefore can only predict the mutual behavior but not the causal reason for this behavior [15] . The ice cream sales - swimming pool drawing association is a famous example from marketing. Ice cream consumption correlates positively with the number of deaths by drowning. This is a pure correlation, since the common cause is an increase in temperature which leads to both effects [14]. In a deterministic system knowing the cause implies knowing the exact effect, while the reverse implication does not hold in general. Inference from node activities as the only source of information gives rise to correlation relations between the nodes. Additional information is needed to distinguish causal relations from pure correlations. The stochastic nature of biological systems poses an additional obstacle, so that only the probability of an effect or event can be determined. For that reason node activities are treated as random variables which can be described by an stochastic process.

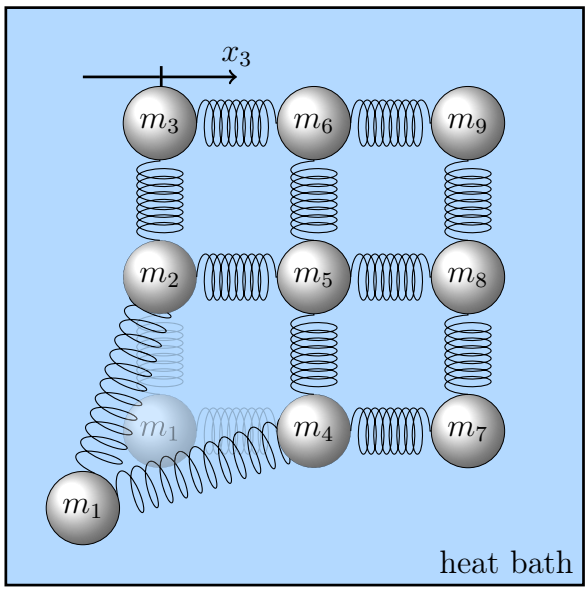
Concluding, network inference reconstructs the underlying network from node activity data, whereas causal relations have to be distinguished from pure correlations. The stochasticity of node activity data leads to a probabilistic interpretation of the inferred network. To infer biological networks exclusively from variations in the abundance of selected molecular components in response to systematic perturbations, is the goal of the here presented work. In the next subsection some network specific obstacles will be explained, that will be tackled in the method chapter 6.

5.2.2 Fundamental concepts of network inference

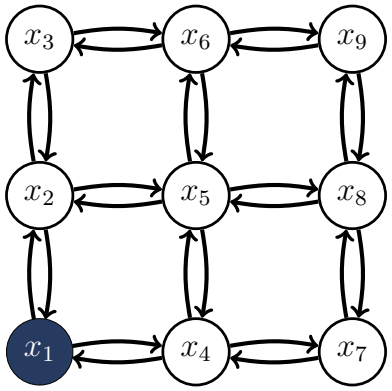
Perturbation experiments generate correlation data

Controlled[2] perturbations influence the network and generate correlation data. To understand this influence, it is of advantage to begin with an intuitive view on the problem. An intuitively accessible analogy is the spring-mass network, capable of describing the effect and the necessity of controlled perturbations. Figure5-3 shows an example of such spring-mass network which has nine metal balls connected by springs with specific spring constant, representing the link strength between nodes.

[2] Here the phrase "controlled" is chosen to distinguish controlled perturbations from random perturbations due to thermal noise and to emphasize the directed force on only one node.



(a) Spring-mass network in a heat bath. Ball m_1 is perturbed from its point of rest, which is illustrated by a transparent node m_1. Due to the perturbation the spring of m_1 attached to m_2 and m_4 is under tension, so that a restoring force back to the point of rest is present.



(b) Graphical model representation of the spring-mass network. The blue filling of node x_1 symbolizes the perturbation.

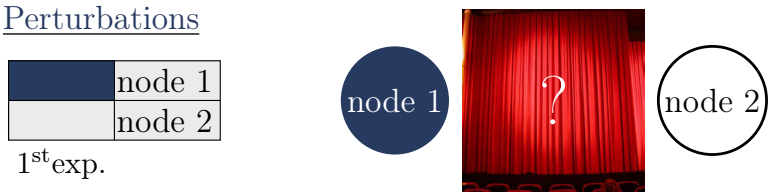
Figure 5-3: **Spring-mass network as a demonstrative example from physics.** Each metal ball of mass m_j is connected to the other balls by the use of springs. By deviating (perturbing) ball m_1 from its point of rest, all network nodes will oscillate around their points of rest leading to information flow through the whole network. The heat bath, illustrated by the blue background color, generates fast random perturbations on each node in addition to the controlled perturbation of ball m_1.

The measured node activity of each ball is given by its relative position x to its point of rest. Obviously, if no external forces act on the balls, there will be no sufficient data information to infer how each ball is connected to the rest of the network. In other words, in the presence of no perturbations it will be impossible to infer the network structure from node activity data, since there is no information flow propagating through the network. Perturbing in the sense of deviating one of the metal balls from its point of rest, as is sketched for mass m_1 in Figure5-3, will influence the rest of the network balls to oscillate around their point of rest. It is very important that these perturbations propagate through the whole network and thereby affect all nodes that can be affected by a single perturbation. This requires a long lasting and relative constant effect on the network as long as a controlled perturbation is applied.

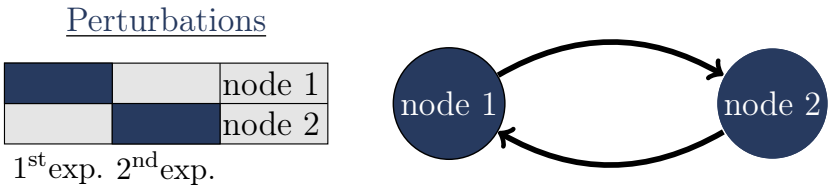
To bring this analogy closer to the statistic treatment of GRN and STN, one can consider a statistic sample set of many such identical and independent spring networks. It is only possible to measure all networks from the sample set separately at a specific point in time - generating the so called replicate experiments. The deviations for each system will be different, depending on the system's state during measurement. However, by means of statistics it is possible to compute the mean standard deviation or variance of each node in response to the single perturbations. Comparing these variances by determining the covariance gives rise to correlation associations between node pairs.

Coming back to the biological networks, the point of rest in the analogy can represent constant gene expression levels in a GRN or a constant growth signal in STN, which must not necessarily be zero. Like in the analogy systematic perturbations are crucial for network inference, otherwise there is no information flow through the network. Examples for controlled perturbations in GRN are transcriptional perturbations like gene over expression [21] or gene knockdowns with RNAi [13], while drugs, i.e. pharmacological inhibitors, can specifically inhibit phospho-proteins in STN, e.g. rapamycin inhibits signaling protein mTOR [50]. These controlled perturbations deviate steady state node activities slightly from their steady state values, whereas replicate experiments are required to see this deviation. In other words, the expected steady state value is the mean node activity of the replicates, while the variance represents the deviation from that value. The covariances of different nodes give rise to positive and negative correlations that describe mutual behavior of nodes to the systematically applied perturbations. Hence, the information about the network structure lies in the deviation from the steady state node activity behavior.

Summarizing, controlled perturbations deviate node activities from their steady state values and propagate through the whole network. As a result of this propagation correlation associations between node pairs can be discovered.



(a) Incomplete data set: Not all molecular components (nodes) have been perturbed in single perturbation experiments.



(b) Complete data set: All nodes have been perturinferredbed.

Figure 5-4: **Systematic perturbation experiments are needed to infer all direct causal molecular interactions from abundance measurements.**

Causal relations and inferability

Systematic perturbations contain information about causal relations How is it possible to reconstruct the underlying network from correlation data? As mentioned above additional information is required to infer causal relation between nodes instead of correlating associations. This additional information is obtained by systematic perturbations of the network consisting of single controlled perturbation experiments. In this systematic way, all networks nodes are perturbed step by step and the respective network responses are observed to produce a complete data set. The additional information lies in knowing the exact targeted node of each of these perturbation experiments, whereas knowledge about the exact effect on targeted nodes are not necessary. This will be derived analytically in the next chapter, but the spring-mass network analogy can help to give an intuitive explanation.

A direct causal link between mass 1 (node 1) and mass 2 (node 2) can be inferred if a significant part of the variation in the relative position of mass 2 can be explained by the one of mass 1, but not by the variations of other network nodes. In plain terms, it must be shown that the movement of mass 2 around its point of rest is

immediately influenced by the oscillation of mass 1 and is not mediated indirectly through oscillations of other masses in the network. The significance of variations can easily be derived from the amplitude (magnitude of variance) if all spring constants are identical, that is if all links share the same link strength. Otherwise, it will be more complicated but still possible, if a complete perturbation data set is available. A complete perturbation data set can be taken to compare the variances of each node from different perspectives, according to different perturbation targets. This allows one to distinguish between causal links and pure correlations as will be shown in chapter 6.

Incomplete data set leads to non-identifiability problem A complete perturbation data set enables one to infer all causal links in a network, while an incomplete data set represents an underdetermined system with not inferable causal links. A network of size[3] D has D^2 possible directed links denoted by D^2 unknown link strength parameters, which describe the network structure. In other words, the network inference problem has D^2 degrees of freedom, that have to be constrained by perturbation data. Each perturbation experiment contains activity data from D nodes, with the result that D unknown parameters can be determined. Hence, $P = D$ single perturbations experiments are necessary to uniquely identify all parameters, so that there is one and only one parameter set describing the network dynamics and structure. Therefore, a network can be uniquely inferred if there is a complete data set with as many perturbations experiments as nodes in the network. In the case of an incomplete data set $P < D$ with less perturbations than network nodes, the system is underdetermined leading to non-identifiable parameters or non-inferable links, respectively. Non-inferable links in this sense mean that it is not possible to distinguish between a direct causal link and a pure correlation association. Figure 5-4 shows the inferability problem for a network of size $D = 2$. By solely perturbing node 1, only a possible outgoing link from node 1 to node 2 can be inferred, while there is no information to infer a possible reverse link. The link from node 2 to node 1 is

[3]The network size denotes the number of nodes in a network.

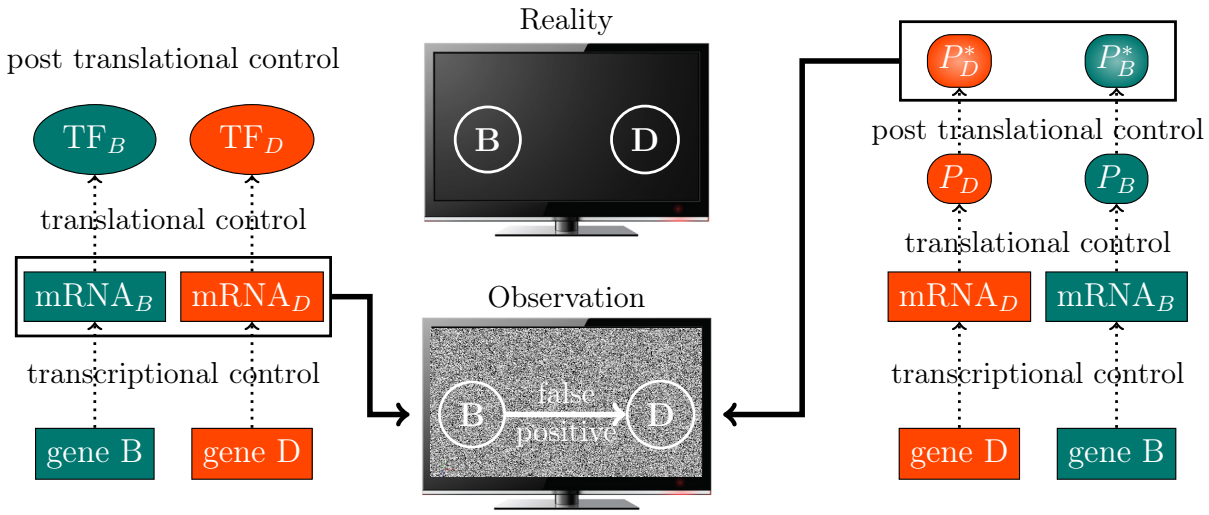


Figure 5-5: **Measurement noise is increased due to measuring only one molecular component, leading to more observed false positives.** By measuring only mRNA abundance in GRN or phospho-protein abundance in STN, two of the three control mechanisms are neglected. As a result the false positive rate of network inference algorithm increases.

called "non-inferable". This is an example for an incomplete data set leading to an underdetermined problem. To infer the whole network the second node has to be perturbed as well, like in Figure5-4(b). Concluding, the non-inferability problem is equivalent to the non-identifiability problem, whereas a complete perturbation data set is indispensable to uniquely infer the whole network.

Measurement noise

Measurement noise consists mostly of biological noise and to some smaller extend to technical noise [31], whereas biological noise is a consequence of thermal noise and other biological uncertainties. In biological networks one has to distinguish between two different kinds of perturbations, namely the controlled perturbations mentioned so far and the fast changing random perturbations due to thermal noise. Random thermal perturbations act and change on a much faster time scale than the controlled ones. They can be modeled by a heat bath in the spring network analogy (see Figure5-3), which shakes the metal balls around their respective points of rest even if no controlled perturbations are applied. In biological networks thermal perturbations

are recognized as gene expression noise, adding uncertainty to the steady state mRNA level as well as to the protein abundance [7].

Another source of uncertainty, which belongs to biological noise, is the fact that one molecular component, namely mRNA or phospho-protein abundance is chosen as proxy for the more complicated underlying interaction scheme. The mRNA abundance is regarded as a proxy for gene regulatory interactions of a GRN, but is actually just a measure for abundance changes caused by transcriptional control. Changes of gene expression due to translational and post-translational regulation are neglected in this way, and manifest an source of uncertainty. A similar problem arises for STN if only phospho-protein abundance is measured and included to the network inference model. Signal tranduction can be understood as a cascade of post-translational modifications to the signaling proteins. In the broader sense phospho-protein abundance is a measure for changes in the post-translational control. Variation in the abundance influenced by transcriptional or translation control are ignored by this approach, leading to another source of uncertainty similar to the one in GRN.

To Conclude, by choosing a measure that neglects parts of the complex interaction scheme, uncertainty arises that manifest as noise in the data. A consequence is that false positive links will be inferred, as sketched in Figure5-5.

5.3 Overview over network inference methods

Following, a short summary of correlation based linear network inference methods will be given. Linear models are less prone to overfitting due to noisy data and explain major phenomenons. On the other hand they can only describe two-body interaction, which means that more complicated interaction schemes involving 2 nodes or more to activate a target node are not captured in the model.

There are many different approaches to network inference. Mutual information, correlation coefficients and probabilistic graphical models, like Bayesian or Markov networks, are some prominent examples [39]. Certainly, it is possible to categorize these methods as undirected and directed network models or as generative and non-generative models, whereas most sophisticated methods stand somewhere in-between these categories. From a logical perspective, most network inference methods can be divided into two categories, namely inductive approaches based on similarity matrices[4] between each pair of nodes and deductive approaches based on modeling the node activity as a effect of a hypothesis network. The inductive inference starts with the observed empirical effect, i.e. similar node activity behavior, and reconstructs the links which are the general causation for this behavior. The deductive inference begins with a potential cause , i.e. a hypothesis network, which can reproduce the measured node activities. The here presented inductive network inference method is correlation based, so that some correlation based concepts will be shortly summarized, next.

5.3.1 Correlation coefficients

The Pearson correlation coefficient is a linear measure of association between two random variables. The phrase random emphasizes the fact that neither of these variables can be controlled in experiments, in contrast to a regression problem with one variable depending on a controlled variable [15]. Given two random variables x_j and x_i representing node activities as well as N replicate measurements x_{jn} and x_{in},

[4] so called measures of association

The Pearson correlation coefficient r for a data sample is expressed as:

$$r_{ji} = \frac{\sum_{n=1}^{N} (x_{jn} - \bar{x}_j) \cdot (x_{in} - \bar{x}_i)}{\left(\sum_{n=1}^{N} [x_{jn} - \bar{x}_j]^2\right)^{\frac{1}{2}} \cdot \left(\sum_{n=1}^{N} [x_{in} - \bar{x}_i]^2\right)^{\frac{1}{2}}} \tag{5.1}$$

where n denotes the replicate experiment and $\bar{x}_j$ a long with $\bar{x}_j$ are sample means of the respective random variables.

$$\bar{x}_j = \frac{1}{N} \sum_{n=1}^{N} x_{jn} \tag{5.2}$$

The correlation coefficient can be understood as a normalized covariance with values in the interval $[-1, 1]$. Therefore, the sample correlation coefficient can be written as sample covariance s_{ji} normalized by the sample standard deviations s_j and s_i of the node activity measurements [15].

$$r_{ji} = \frac{s_{ji}}{s_j s_i} \tag{5.3}$$

$$= \frac{1}{N-1} \sum_{n=1}^{N} \left(\frac{(x_{jn} - \bar{x}_j)}{s_j}\right) \cdot \left(\frac{(x_{in} - \bar{x}_i)}{s_i}\right) . \tag{5.4}$$

with sample covariance

$$s_{j,i} = \frac{1}{N-1} \sum_{n=1}^{N} (x_{jn} - \bar{x}_j) \cdot (x_{in} - \bar{x}_i) \ , \tag{5.5}$$

and sample standard deviations

$$s_j = \left(\frac{1}{N-1} \sum_{n=1}^{N} (x_{jn} - \bar{x}_j)^2\right)^{\frac{1}{2}} \tag{5.6}$$

The actual information about node activity associations lies in the covariance, so that it will be sufficient to focus on this quantity for the ongoing discussion. For a network of size D, it is convenient to introduce a covariance matrix containing all associations between pairs of network nodes on the off-diagonal and variances on the

diagonal. Accordingly, the sample covariance matrix is

$$S := \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1D} \\ s_{21} & s_{22} & & \\ \vdots & & \ddots & \vdots \\ s_{D1} & & & s_{DD} \end{pmatrix} , \tag{5.7}$$

and the population covariance matrix is written as

$$C := cov\left[\vec{x}\right] = \begin{pmatrix} cov[x_1, x_1] & cov[x_1, x_2] & \dots & cov[x_1, x_D] \\ cov[x_2, x_1] & cov[x_2, x_2] & & \\ \vdots & & \ddots & \vdots \\ cov[x_D, x_1] & & & cov[x_D, x_D] \end{pmatrix} , \tag{5.8}$$

with

$$cov[x_j, x_i] = E\left[x_j \cdot x_i\right] - E[x_i] \cdot E[x_i] \quad . \tag{5.9}$$

The sample measure is an approximation for the population measure and converges in the limit of very large[5] replicate data sets $N \to \infty$. The population measure can be regarded as the model quantity coming from theory.

To conclude, the covariance matrix already establishes a very simple correlation network, which is bidirectional or undirected and is unable to distinguish causal links from pure correlations in the case of GRN and STN.

[5]The phrase *Big Data* is used as a synonym in this case.

5.3.2 Inverse covariance matrix - partial correlation

A more sophisticated technique to infer pairwise associations is obtained by examining the inverse covariance matrix C^{-1}, whose elements are called partial correlation coefficients. Like in the case of correlations, partial correlations can only infer linear associations between a pair of random variables. Unlike correlations, partial correlations are calculated by considering the whole network or as Adi Raveh stated in his paper [54] from 1985 :

> " In contrast to the elements of $[C]$, the elements of the inverse [covariance matrix C^{-1}] usually change as additional [random] variables are added to or deleted from the set . [...] The inverse correlation matrix behaves in a truly multivariate fashion, rather than merely in a multibivariate fashion, as does [the correlation] itself." [54]

An easy clarifying example is the expression for the inverse C^{-1} of a two by two non-singular covariance matrix C, representing a network with two nodes or random variables, respectively.

$$C^{-1} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}^{-1} = \frac{1}{c_{11}c_{22} - c_{21}c_{12}} \cdot \begin{pmatrix} c_{22} & -c_{12} \\ -c_{21} & c_{11} \end{pmatrix} \tag{5.10}$$

The elements of C^{-1} are considered as partial correlations, which are computed by dividing by the determinant. The determinant is the main reason that accounts for the whole network and hence distinguishes correlations from partial ones. Partial correlations in the network can be interpreted as the immediate correlation between two nodes after removing the contribution of the remaining network on them, i.e. removing the influence of confounding[6] nodes.

The question that arises now is for which type of network structures and data, the inverse covariance matrix and partial correlations can be regarded as causal interactions? In other words, for which case does the inverse covariance matrix determine

[6]Assume a network $A \leftarrow C \rightarrow B$, then the mutual parent C of A and B is called confounding factor

the network structure? To answer this question the spring-mass system from Figure 5-3 will be considered again to heuristically derive the requirements. Because the spring mass system is situated in a heat bath, it can be regarded as a system in thermodynamic equilibrium. Thermodynamic equilibrium or equilibrium in general is defined as the absence of a net energy flow, whereas random energy fluctuation around a constant mean can be present. For the spring mass network this means that the system is in equilibrium as long as only random perturbations due to thermal noise are present. In contrast, the system will be in non-equilibrium if a net energy flow is present that can be initiated by controlled perturbations from outside the system. Controlled perturbations propagate through the whole network and thereby push the system out of equilibrium.

The probability of a certain system state, i.e. network structure, in the thermal equilibrium is given by the Gibbs-Boltzmann distribution

$$p(A) = \frac{1}{Z} \cdot \exp\left(-\frac{U(A)}{k_B T}\right) \quad , \tag{5.11}$$

with link strength matrix, also called interaction matrix, which is related to the spring constants.

$$A = \begin{pmatrix} A_{11} & \dots & A_{1D} \\ \vdots & \ddots & \vdots \\ A_{D1} & \dots & A_{DD} \end{pmatrix} \quad . \tag{5.12}$$

While the off-diagonal elements of the link strength matrix denote the strength of direct interactions between pairs of nodes, the diagonal elements represent a restoring force, like the degradation rate in GRN and STN. The network structure is determined by the pattern that zero elements of the interaction matrix generate, whereas $A_{ji} = 0$ means no direct influence of node i upon node j. Partition matrix Z is the normalizing constant that sums over all possible states and is responsible for the Gibbs-Boltzmann distribution being a real probability distribution with values in the interval $[0, 1]$.

$$Z = \sum_{A} \exp\left(-\frac{U(A)}{k_B T}\right) \tag{5.13}$$

Quantity $U(A)$ in eq.5.11 is called Gibbs measure in the general theory of Markov random fields [30]. Historically, it was introduced by physicist Ising in his model of ferromagnetism, where it is regarded as an energy function in the thermodynamic units $k_B \cdot T$ with k_B being the Boltzmann constant and T the temperature. The Gibbs measure $U(A)$ for the spring-mass system of network size D with node activity x_j can be formulated as

$$U(A) = -\sum_{j,i=1}^{D} x_j A_{ji} x_i \;=\; -\vec{x}^T \cdot A \cdot \vec{x} \quad , \tag{5.14}$$

where $\vec{x}^T = (x_1, \ldots, x_D)$ is the transpose of $\vec{x}$. Inserting this Gibbs measure in eq5.11 yields the probability $p(A)$ of the network having structure A

$$p(A) = \frac{1}{Z} \cdot \exp\left(\frac{\vec{x}^T \cdot A \cdot \vec{x}}{k_B T}\right) \quad . \tag{5.15}$$

By neglecting the constants of this equation and adding an mean node activity vector$\vec{\mu}$

$$p(A) \propto \exp\left[-\frac{1}{2}\left(\vec{x}-\vec{\mu}\right)^T (-A)\left(\vec{x}-\vec{\mu}\right)\right] \tag{5.16}$$

one can see the similarity to a multivariate Gaussian distribution

$$p(\vec{x}|\vec{\mu}, C) \propto \exp\left[-\frac{1}{2}\left(\vec{x}-\vec{\mu}\right)^T C^{-1}\left(\vec{x}-\vec{\mu}\right)\right] \quad . \tag{5.17}$$

By comparing eq.5.16 with eq.5.17, one can see the connection between the inverse covariance matrix C^{-1} and the interaction matrix A. To answer the question from the beginning of this section, namely for which conditions A is determined by C^{-1}, it is sufficient to know the features of the covariance matrix of a Gaussian distribution. For a multivariate Gaussian distribution the covariance matrix must be symmetric $C = C^T$ and positive-definite. Its inverse C^{-1} must also be symmetric and positive-definite ([11] p.688/689). An additional obvious restriction for the sample data is that $\vec{x}$ must be distributed according to a joint Gaussian distribution [36]. Applying

this conditions to answer above question, one can summarize:

$$C^{-1} = -A \tag{5.18}$$

is only valid for inferring the real causal network structure, if

1. the network structure is bidirectional $A = A^T$, so that an interaction is regarded as a reversible process,

2. A is negative-definite ($-A$ must be positive-definite), and

3. the node activity data must be distributed according to a Gaussian distribution around mean $\vec{\mu}$. This is only valid for a system in equilibrium, which was an initial condition for the heuristic derivation above.

The second condition, positive definiteness [7], means that all eigenvalues are positive and not equal to zero, hence A is invertible. Exclusive positive eigenvalues are necessary since they ensure that node activities x_j remain positive after interacting with the network. Node activities represent abundance quantities like concentration or copy number, therefore they must always be positive to make physically sense. Hence, positive definiteness is a general condition that must be valid for all network inference methods.

To conclude, the inverse covariance matrix and partial correlations do not determine the network structure of GRN and STN, since these biological networks are not bidirectional. Additionally, controlled perturbation experiments are usually performed, which contradicts another essential condition. Controlled perturbations are needed to generate an information flow through the network, as explained in previous sections. However, partial correlations for GRN and GRN can still be computed as long as the covariance matrix is positive definite, which is a sufficient condition for matrix inversion. In this case it is of paramount importance that these partial cor-

[7] Geometrically a positive definite matrix maps any vector on the positive subspace, where all vector elements stay positive. Therefore, it only stretches and distorts the vector - much like the multiplication of a positive real number does in a one dimension.

relations are not confused with real causal interaction - this is another example for undirected networks.

5.3.3 Response matrix

An approach that can infer direct network links from controlled perturbation data, if the perturbation strength is known, is the response matrix method [9, 20] . The local response G_{ji} denotes how changes δx_i in the activity of node i influence changes δx_j in the activity of nodes j.

$$G_{ji} := \frac{\delta x_j}{\delta x_i} \tag{5.19}$$

Consequently, by determining the response matrix G, one can infer the underlying network structure A, which is not necessarily the same quantity.

In following the relation between response matrix G and interaction matrix A will be derived, to show the limits of the response matrix method. Assuming no random perturbations and no measurement noise, one can formulate the steady state $(d\vec{x}/dt = 0)$ equation for node activities $\vec{x}$ as:

$$A \cdot \vec{x} + \vec{u} \stackrel{!}{=} 0 \tag{5.20}$$

$$\Leftrightarrow \vec{x} = -A^{-1} \cdot \vec{u} \quad , \tag{5.21}$$

where $\vec{u} = (u_1, \ldots, u_D)^T$ is the vector with controlled perturbations for a network of size D. Above equation 5.21 can be translated into an element-wise notation:

$$x_j = -A^{-1}_{ji} \cdot u_i \quad , \tag{5.22}$$

where A^{-1}_{ji} stands for element (j, i) of the inverse interaction matrix A_{ji}. Assuming the system stays in equilibrium, the direct (node activity) response δx_j of node j to a perturbation δu_i exclusively acting on node i is

$$\delta x_j = -A^{-1}_{ji} \cdot \delta u_i \quad . \tag{5.23}$$

Since the system musst stay in (close to) equlibrium, the notion of δx_j and δu_i can be regarded as some kind of infinitesimal small virtual displacement away from the equilibrium similar to the one in theoretical physics. Experimentally this can only be realized by small perturbations, which do not push the system to far away from its expected value and therefore permanently into another equlibrium state. Next, the perturbation effect or response δx_i of the perturbed node itself due to degradation rate A_{ii} and a net feedback from the whole network, can be written as:

$$\begin{aligned} \delta x_i &= -A_{ii}^{-1} \cdot \delta u_i \\ \Leftrightarrow \delta u_i &= -\frac{\delta x_i}{A_{ii}^{-1}} \qquad (5.24) \\ \Leftrightarrow A_{ii}^{-1} &= -\frac{\delta x_i}{\delta u_i}\,. \qquad (5.25) \end{aligned}$$

Finally it is possible to connect the interaction matrix with the response matrix by

$$G_{ji} = \frac{\delta x_j}{\delta x_i} = \frac{A_{ji}^{-1}}{A_{ii}^{-1}} \qquad (5.26)$$

However, trying to determine the absolute value A_{ji}^{-1} leads to the problem that information about the perturbation strength δu_i musst be available, which is usually not the case in biological networks.

$$A_{ji}^{-1} = \frac{\delta x_j}{\delta u_i} = G_{ji} \cdot A_{ii}^{-1} \qquad (5.27)$$

From a different perspective, this lack of information about δu_i is equivalent to the unknown diagonal elements A_{ii} (e.g. degradation rate) of the interaction matrix A, which could be computed if δu_i was known.

Concluding, the response matrix will only infer the network structure correctly, if there is experimental information about the exact perturbation strength. In biological networks, i.e. GRN and STN, information about the strength of perturbations is usually not available. There are a couple of very controversial [53] approximations to eq.5.27, which try inference without knowledge of the pertubation strength [9,20]. For

example is this achieved by setting the diagonal elements of the interaction matrix to $A_{ii} = -1$ [9], leading to infered link strengths that are normalized wrongly and are therefore not comparable with each other. As a result the slightest measurement noise will decrease performance drastically [53].

Chapter 6

Model and theory

6.1 Objective

The weakness of the inference methods described in the previous chapter, motivates to create a new improved network inference method which can infer the underlying network structure from correlation data. In detail, it should infer direct causal links between observed nodes, from controlled perturbation data. Additionally, it will be assumed that the target node of each perturbation experiment is known, whereas the exact perturbation strength is unknown. Further, this new method shall not be biased by approximating the diagonal elements of the interaction matrix, as it is the case for the response matrix methods. This can be achieved by focusing on the relative link strengths A_{ji}/A_{jj} with respect to the unknown restoring force (degradation rate), as will be shown later in this chapter. The relative link strengths can be used to infer the network structure but do not lead to link strength parameter estimation. In this work, the goal is to infer the network structure, whereas parameter estimation is considered to be a different topic.

In mathematical terms, the task is to infer structure G of a network with size D from $P \leq D$ controlled single perturbation experiments X^p , each having N replicates $X^p = \{\vec{x}_1, \ldots, \vec{x}_n, \ldots, \vec{x}_N\}$. Here, $\vec{x} = (x_1, \ldots, x_D)^T$ is the vector of random variables denoting the activity of all observed nodes in the network. The graph, which is also

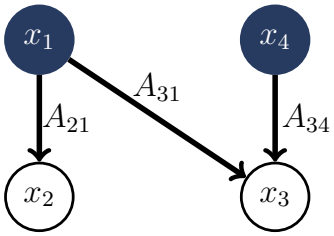


Figure 6-1: **Example network of size $D = 4$ with two perturbed nodes**. Node i acts on node j with a link strength (interaction strength) A_{ji}. Node activities are modeled by random variables $\vec{x} = (x_1, \ldots, x_D)^T$.

called network structure, is given by

$$G = \begin{pmatrix} G_{11} & \ldots & G_{1D} \\ \vdots & \ddots & \vdots \\ G_{D1} & \ldots & G_{DD} \end{pmatrix} \quad , \tag{6.1}$$

where $G_{ji} \in \{0, 1\}$ denotes a directed link from node i to node j (**i→j**). Following, the network structure or graph denoted by matrix G will be distinguished from the link strength matrix A.

$$A = \begin{pmatrix} A_{11} & \ldots & A_{1D} \\ \vdots & \ddots & \vdots \\ A_{D1} & \ldots & A_{DD} \end{pmatrix} \quad , \tag{6.2}$$

where $A_{ji} \in \mathbb{R}$ denotes the directed link strength from node i upon node j (**i→j**). While the network structure G is boolean with zero entries standing for no links and ones for direct links, the link strength matrix A shows how strong the effect of one node is upon the other. Due to measurement noise link strength estimations are hardly ever set link strengths A_{ji} exactly to zero, so that the inferred link strength matrix A^* can not determine the network structure G without any further information. Additional information can be a cutoff which sets weak links to zero or prior structural information like network sparsity.

Before starting to derive the new inference method a short overview of this chapter will be provided. First the inference problem of finding the whole network will be simplified by inferring the incoming links of a single node, so that the whole net-

work can be reconstructed step by step. This simplification will be probabilistically formulated in analogy to a Bayesian Network setting, where direct incoming links are referred to as random variables being conditional dependent on their immediate parents. In section 6.3 the node interactions will be modeled by a linear differential equation in steady state, which will lead to an expression for the population covariance matrix. This theoretically derived covariance matrix stands in contrast to the sample covariance matrix from the data samples. Assuming that all possible links exist, the so called total connectivity assumption, a maximum likelihood (ML) estimate $\widehat{A}^* = A^*_{ji}/A^*_{jj}$ for the relative link strength $\widehat{A}_{ji} = A_{ji}/A_{jj}$ will be derived. Finally, a Markov chain Monte Carlo simulation draws a set of most likely network structures $\{G\}$ from the posterior distribution $p(G|\text{data}, A^*)$ over sparse network structures given the link strength ML estimate A^*. The posterior distribution will be obtained by introducing different sparsity priors $(L0, L1, L2)$, which restrict model complexity with different emphasis.

6.2 From correlation data to direct causal incoming links

Considering a complete perturbation data set ($P = D$), the inference of the whole network A at once can be simplified into inferring the incoming links of each node step by step. In each step the incoming links $A_{j,\text{row}} := (A_{j1}, \ldots, A_{jD})$ of node j can be computed from the mean sample covariance matrix

$$\bar{S}^{(p\neq j)} := \frac{1}{P-1} \cdot \sum_{p=1, p\neq j}^{P} S^{(p)} \overset{P=D}{=} \frac{1}{D-1} \cdot \sum_{p=1, p\neq j}^{D} S^{(p)} , \tag{6.3}$$

where $S^{(p)}$ is the sample covariance matrix of the p^{th} perturbation experiment in which only node p is perturbed and the node activity for all nodes is measured in N replicate experiments. By averaging over all perturbation experiment but the one where node j is perturbed, the correlation information between all other network nodes will be

destroyed. Any information about existing indirect paths is destroyed by taking the mean $\bar{S}^{(p\neq j)}$ of sample covariance matrices $S^{(p)}$. The remaining correlations are direct ones between each perturbed node i and the unperturbed node j, whereas these correlations are either due to direct causal influence from the network upon node j or due to measurement noise. Since the perturbed target nodes are known, correlations between any perturbed node i and the one unperturbed node j, can be regarded as directed incoming links of latter. Concluding, to infer the incoming links of each node j step by step average correlation information from all perturbation experiments but the one where j is perturbed is utilized. In simple words, the incoming links of node j can be computed if all other nodes are perturbed as it is sketched in Figure 6-2 .

For an incomplete perturbation data set ($P < D$) network inference becomes a non-identifiable[1] problem, because the degrees of freedom are larger than the number of constraints given by single perturbations (see section 5.2.2). From the perspective of correlations, not all indirect paths are destroyed, so that possible direct incoming links to node j can be a result of pure correlations. Nevertheless, given an incomplete data set the mean sample covariance is

$$\bar{S}^{(p\neq j)} := \frac{1}{Q} \cdot \sum_{p=1, p\neq j}^{P} S^{(p)} \quad , \tag{6.4}$$

where $Q = P - 1$ for j being a removed perturbation experiment and $Q = P$ if there is no perturbation experiment available for node j.

Probabilistic view by means of Bayesian networks: The notion of inferring the incoming links of one node j independent of the remaining network structure, can be probabilistically formulated in analogy to a Bayesian network (BN) setting. Bayesian networks belong to the family of linear probabilistic graphical models ([11] p.359) and are typically applied in order to model directed causal interactions. They are defined as directed acyclic graphs with random variables standing for node ac-

[1]The non-identifiability problem can be solved by the "rescaling" concept, which is a numerical approach introduced in our paper [12]. The rescaling concept is not part of this PhD thesis, so that the non-identifiability problem for in complete data sets will not be sufficiently covered in the here presented work.

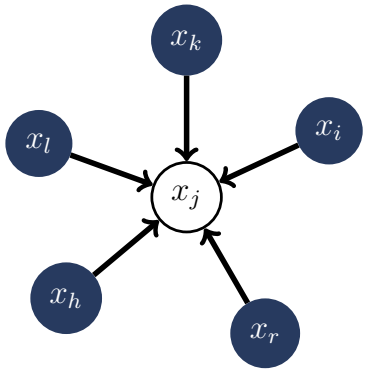


Figure 6-2: **Inference of the whole network can be simplified by inferring the incoming links of each node step by step.** In an approach analogous to Bayesian networks, immediate parents of nodes j can be formulated as conditional dependencies. Parents must be perturbed to infer in-coming links of node j, whereas data where j is perturbed is removed. Perturbed nodes are filled with blue, while not perturbed ones are white.

tivities. Direct links (or edges) between nodes characterize the conditional probabilistic independence of one node to the rest of the network given its direct parent nodes [11, 48]. As a result the joint probability $p\left(\vec{x}|A,G\right)$ factorizes into a product of conditional probabilities $p(x_j|G_{j,row}, A_{j,row}, Pa_G(x_j))$.

$$p\left(\vec{x}|G,A\right) = p\left(x_1,\ldots,x_D|G,A\right) = \prod_{j=1}^{D} p(x_j|G_{j,row}, A_{j,row}, Pa_G(x_j)) \quad , \tag{6.5}$$

where the probability of node activity x_j only depends on the immediate parents' node activities $Pa_G(x_j)$, which are given by the network structure G. Actually, the immediate parents $Pa_G(x_j)$ are determined by the incoming links denoted by $G_{j,row}$ and $A_{j,row}$, so that the conditional probabilities can be understood as conditional independent parts of the network.

Hence, the approach of inferring the incoming links step by step can be heuristically connected to the Baysian network description, by identifying the parents $Pa_G(x_j)$ of node j as nodes that are perturbed and node j itself as the only unperturbed node for a complete data set (see Figure 6-2). Further, total connectivity will be assumed $G = \mathbb{1}$, i.e. $G_{ji} = 1 \forall j, i$, where each node is connected to all other

nodes in the network.

$$p(x_j|G, A_{j,row}, Pa_G(x_j)) \overset{G=\mathbb{1}}{=} p(x_j|A_{j,row}, \{x_{i\neq j}\}) \widehat{=} p(\bar{S}^{(p\neq j)}|A_{j,row},) \\ = p(\{X^{p\neq j}\}|A_{j,row}) \tag{6.6}$$

where the parents $Pa_G(x_j) = \{x_{i\neq j}\}$ are the whole network without node j. In eq.6.6 the parents are identified with the perturbed nodes p, while the activity date x_j in BN is identified with the perturbation data $\{X^{p\neq j}\}$, which is equivalent to the notation with the mean sample covariance notation $\bar{S}^{(p\neq j)}$. Now, the joint probability of the whole network can be understood as the product of conditional probabilities of part of the network, namely the incoming links of each nodes.

$$\begin{aligned} & p\left(\{X^p\}|G=\mathbb{1}, A\right) \\ =& p\left(\{X^{p\neq 1}\}, \ldots, \{X^{p\neq D}\}|G=\mathbb{1}, A\right) = \prod_{j=1}^{D} p(\{X^{p\neq j}\}|A_{j,row}) \\ =& p\left(\bar{S}^{p\neq 1}, \ldots, \bar{S}^{p\neq D}|G=\mathbb{1}, A\right) = \prod_{j=1}^{D} p(\bar{S}^{p\neq j}|A_{j,row}) \end{aligned} \tag{6.7}$$

In other words, eq.6.7 describes how the joint likelihood of the whole perturbation data given the whole network A is equal to the product of conditional likelihoods of unperturbed node j given only its incoming links $A_{j,row}$ and the perturbed nodes' data in the form of $\bar{S}^{(p\neq j)}$. Consequently, determining the likelihood of part of the network A_{ji} can be achieved without considering the remaining network. This is exactly the approach which will be implemented in the upcoming sections to compute a maximum likelihood estimate for the incoming link strengths leading to the whole link strength matrix A. Notice that in contrast to Bayesian networks, the total network can comprise cycles, since each conditional likelihood is conditioned on differently averaged data represented by $\bar{S}^{(p\neq j)}$.

6.3 A probabilistic view

The goal of this section is to formulate a linear stochastic model which describes the steady state activity of the network in order to obtain a theoretical expression for the population covariance matrix C. As mentioned in the previous section 5.2.2, the information about the link strength of an interaction will be contained in the co-variations of node activities, if controlled perturbation are applied. Consequently, the mean or expectation of node activities do not contain any information for reconstructing the network.

In detail, the model describes node activity deviations $\vec{x} - \vec{\mu}$ from the steady state value μ as a linear process:

$$\frac{d\vec{x}(t)}{dt} = A \cdot \left[\vec{x}(t) - \vec{\mu}\right] + \vec{u} \stackrel{!}{=} 0 \quad , \tag{6.8}$$

where the steady state value is regarded as the expectation value $\mu := E[\vec{x}]$ and perturbation vector $\vec{u} = (u_1, \ldots, u_D)^T$ consists of single controlled perturbations on each network node. Controlled perturbations must be applied long enough to propagate through the whole network, which is implemented by $\vec{u}$ being constant in time or independent of time, respectively. The link strength matrix A is considered to be a negative definite matrix, so that $-A$ is positive definite like in the case of partial correlations (see section 5.3.2). The definiteness condition ensures that the link strength matrix is invertible and that its inverse A^{-1} is negative definite as well. Since the process in eq.(6.8) is actually a linear stochastic one, the variables in above equation will be substituted by random variables denoted by replicate index n. After rearranging eq.(6.8) with respect to node activities $\vec{x}_n$ and adding Gaussian noise ϵ_n, the generative view of this linear Gaussian model will be:

$$\vec{x}_n = -A^{-1} \cdot \vec{u}_n(\vec{a}_n) + \vec{\mu} + \vec{\epsilon}_n \quad , \tag{6.9}$$

where $\vec{x}_n$ is the measured node activity with technical measurement noise $\vec{\epsilon}_n = (\epsilon_{1n}, \ldots \epsilon_{Dn})^T$. The technical noise is assumed to be Gaussian distributed with zero

population mean $E[\vec{\epsilon}] = 0$ and covariance matrix $cov[\vec{\epsilon}] = \sigma_\epsilon^2 \cdot I_D$, whereas I_D denoting a $D \times D$ identity matrix and σ_ϵ the common standard deviation. In other words, the technical noise has the same uncertainty effect σ_ϵ on all network nodes leading to noisy observations even in the absence of external controlled perturbations. However, this uncertainty effect vanishes on average for large replicate numbers (large sample size), which can represent the statistical population size sufficiently. In the generative view, one can define the noise random variable as

$$\vec{\epsilon}_n = \sigma_\epsilon \cdot I_D \cdot \vec{\xi}_n = \sigma_\epsilon \cdot \vec{\xi}_n \tag{6.10}$$

$$p\left(\vec{\epsilon}\right) = \mathcal{N}\big(\vec{\epsilon} \mid 0\ ,\ \sigma_\epsilon^2 \cdot I_D\big) \quad , \tag{6.11}$$

where random variable $\vec{\xi}_n$ is sampled from the standard normal distribution $p(\vec{\xi}) = \mathcal{N}\big(\vec{\xi} \mid 0, I_D\big)$. Further, the effect of perturbations $\vec{u}(\vec{a}_n)$ on the network are assumed to be linear and Gaussian with an zero expectation $E[\vec{u}] = 0$ (population mean) and a covariance matrix $cov(\vec{u}) = B \cdot B^T$

$$\vec{u}_n = B \cdot \vec{a}_n \tag{6.12}$$

$$p(\vec{u}) = \mathcal{N}\big(\vec{u} \mid 0\ ,\ B \cdot B^T\big) \quad , \tag{6.13}$$

where random variable $\vec{a}_n$ is sampled from the standard normal distribution $p(\vec{a}) = \mathcal{N}\big(\vec{a} \mid 0, I_D\big)$. In the here presented model, perturbations represent small stochastic deviations from the mean node activity and can therefore only affect the standard deviation. This is the reason why $E[\vec{u}]$ is set to zero and is valid without loss of generality. In the general case of a non zero expected perturbation, the expectation value of node activities $\mu = E[\vec{x}]$ can be formulated with respect to the expectation value $E[\vec{u}]$ of perturbations.

$$\vec{\mu} = E[\vec{x}] = J^{-1} \cdot E[\vec{u}] \quad , \tag{6.14}$$

The definition of matrix $B \in \mathbb{R}^{D \times D}$ and hence $cov[\vec{u}]$ depend both on the experimental implementation of perturbations. It will be assumed that the effect of different

perturbations u_k are independent, which results in zero covariance $cov(u_k, u_i) = 0$ for all $k \neq j$ between single perturbation experiments. The uncertainty of the perturbation effect on the nodes is the same for all nodes in the network, i.e. the variances are identical $var(u_k) = var(u_i) \equiv \sigma_u^2$ for all k, j . Further, the complete perturbation data set $\{X^{(p)}\}$ will be arranged to a data set $\{X^{(p\neq j)}\}$ which includes all perturbations except for node j, whose incoming links should be infered by the sample covariance matrix $S^{(p\neq j)}$ (see section 6.2). These conditions lead to the exact expression for the standard deviation matrix:

$$B := \sigma_u \cdot I_B = \sigma_u \cdot \sum_{p=1, p\neq j}^{P} I^{(p)} \tag{6.15}$$

$$I_B := \sum_{p=1, p\neq j}^{P} I^{(p)} \quad , \tag{6.16}$$

where $I^{(p)} \in \mathbb{R}^{D\times D}$ is a diagonal matrix with the only non-zero entry $I^{(p)}_{pp} = 1$ signifying the perturbed target node p of a single perturbation experiment.

$$I^{(p)} := \text{diag}\left(0, \ldots, 0, I^{(p)}_{pp} = 1, 0, \ldots, 0\right) \overset{\text{e.g. } p=2}{\overset{D=3}{=}} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{6.17}$$

The standard deviation matrix B of perturbations contains two types of information about the controlled perturbations, namely the unknown strength of perturbations σ_u and the known target of perturbations I_B. Finally, the general form of the generative view is obtained by inserting eq.(6.12) and (6.10) into eq.(6.9), which yields

$$\vec{x}_n = -\underbrace{A^{-1} \cdot B \cdot \vec{a}_n}_{\text{standard deviation}} + \underbrace{\vec{\mu}}_{\text{mean}} + \underbrace{\sigma_\epsilon \cdot \vec{\xi}_n}_{\text{noise}} \tag{6.18}$$

It can be understood as the sum of standard deviation due to perturbations, node activity mean, and random Gaussian noise.

To obtain an expression for the population covariance matrix C of the network activity $\vec{x}_n$, eq.(6.18) can be formulated in a probabilistic way. Since the sum of two

Gaussian is again a Gaussian distributions, joint probability distribution $p(\vec{x})$ must be multivariate Gaussian distributed, too.

$$p(\vec{x}) = \mathcal{N}\big(\vec{x} \mid E[\vec{x}], cov[\vec{x}]\big) \quad , \tag{6.19}$$

where, considering $E[\vec{a}] = 0$, $E[\vec{\epsilon}] = 0$, the expectation or population mean of the node activities is

$$E[\vec{x}] = E[A^{-1} \cdot B \cdot \vec{a}_n + \vec{\mu} + \vec{\epsilon}_n] = \vec{\mu} \quad , \tag{6.20}$$

and, by considering $E[\vec{a} \cdot \vec{\epsilon}] = E[\vec{a}] \cdot E[\vec{\epsilon}]$, $cov[\vec{a}] = E[\vec{a}\vec{a}^T] = I_D$, the covariance matrix is

$$\begin{aligned} C := cov[\vec{x}] &= E\left[(\vec{x}_n - E[\vec{x}_n]) \cdot (\vec{x_n} - E[\vec{x}_n])^T\right] \\ &= E\left[(A^{-1} \cdot B \cdot \vec{a}_n + \vec{\epsilon}_n) \cdot (A^{-1} \cdot B \cdot \vec{a}_n + \vec{\epsilon}_n)^T\right] \\ \Rightarrow C &= A^{-1} \underbrace{BB^T}_{cov(\vec{u})} \left(A^{-1}\right)^T + \sigma_\epsilon^2 \cdot I_D \quad . \end{aligned} \tag{6.21}$$

Concluding, the theoretically derived population covariance matrix C depends on two model parameters, which are the inverse link strength matrix A^{-1} and the variance of technical noise σ_ϵ^2. These model parameters have to be inferred from training data, i.e. systematic perturbation data, whereas inferring the link strength matrix A and hence the network structure is the actual goal. The joint Gaussian distribution, which is used to define the likelihood function in the upcoming section, is

$$\begin{aligned} p\left(\vec{x}|A^{-1}\right) &= \mathcal{N}\left(\vec{x} \mid \vec{\mu}, C\left(A^{-1}, \sigma_\epsilon^2\right)\right) \\ &= \left(\frac{1}{2\pi}\right)^{D/2} \cdot \left(\frac{1}{det(C)}\right)^{1/2} \exp\left[-\frac{1}{2} \cdot (\vec{x}_n - \vec{\mu})^T C^{-1} (\vec{x}_n - \vec{\mu})\right] \end{aligned} \tag{6.22}$$

Finally, notice that the generative view established by eq.(6.18), can and will be used to produce synthetic data, which is a typical way to assess the performance of network inference algorithms.

6.4 Maximum likelihood estimate of link strength assuming total network connectivity

The goal of this section is to determine the principle components of the inverse mean sample covariance matrix $(\bar{S}^{(p\neq j)})^{-1}$, which can be identified with the maximum likelihood estimate of the incoming links $A_{j,\text{row}}$ of the not perturbed node j. As mentioned in section 6.2, any information about indirect paths is destroyed by taking the mean $\bar{S}^{(p\neq j)}$, so that the remaining associations are due to direct links from perturbed nodes i to the unperturbed node j. In other words, the remaining real information left in $\bar{S}^{(p\neq j)}$ is the one about the incoming links $A_{j,\text{row}}$ of node j, whereas measurement noise distorts this information. Further, this incoming link $A_{j,\text{row}}$ can be found to be the principle components (PC) of the partial correlation matrix $(\bar{S}^{(p\neq j)})^{-1}$, i.e. the principle partial correlations, while the remaining components are due to pure measurement noise.

The general idea of principle component analysis (PCA) is to map correlation data associations from the data space $\mathbb{R}^D$ of dimension D into the principle subspace $\mathbb{R}^M$ of dimension $M \leq D$. The principle subspace $\mathbb{R}^M$ represents the subspace of real correlations associations, while the complement subspace $\mathbb{R}^{M_0} := \mathbb{R}^{D-M}$ represents the space of correlation associations due to "noise". By looking at the eigendecomposition (spectral decomposition) of S one can easily show that the principle subspace of the inverse mean sample covariance matrix $(S)^{-1}$ (partial correlations) is identical to the "noise" subspace of S. According to PCA [11], the eigenvectors corresponding to the largest M eigenvalues of S span the principle subspace of S, while compliment subspace $\mathbb{R}^{M_0}$ is spanned by the eigenvectors of the M_0 smallest eigenvalue. By looking

at the eigendecomposition (spectral decomposition) of S and of its inverse $(S)^{-1}$,

$$S = \underbrace{\sum_{i=1}^{M} \lambda_i \vec{U}_i \vec{U}_i^T}_{\substack{\textbf{principle subspace of} \\ \textbf{correlations } S}} + \underbrace{\sum_{l=1}^{M_0} \lambda_l \vec{U}_l \vec{U}_l^T}_{\substack{\textbf{principle subspace of} \\ \textbf{partial correlations } (S)^{-1}}} \tag{6.23}$$

$$(S)^{-1} = \overbrace{\sum_{i=1}^{M} \frac{1}{\lambda_i} \vec{U}_i \vec{U}_i^T}^{\underline{\textbf{correlations } S}} + \overbrace{\sum_{l=1}^{M_0} \frac{1}{\lambda_l} \vec{U}_l \vec{U}_l^T}^{\underline{\textbf{partial correlations } (S)^{-1}}} , \tag{6.24}$$

it is obvious that the principle subspace of partial correlations $(S)^{-1}$ is spanned by the M_0 eigenvectors of the smallest eigenvalues of S, since the reciprocal values are the largest eigenvalues of the partial correlation matrix. This feature enables one to determine the principle subspace of partial correlations by determining the eigenvectors U_{M_0} of S.

Each component or random variable in the PCA framework stands for the activity x_i of one node in the network inference framework, so that the dimension of the data space is determined by the number of nodes D in the network. Further, the dimension M of the principle subspace of correlations can be identified with the number of perturbed nodes in the reduced data set represented by mean sample covariance matrix $\bar{S}^{(p \neq j)}$ (for a complete data set $M = P-1$). Consequently, the dimension of the principle subspace of partial correlations is identified with the number of unperturbed nodes, which is $M_0 = 1$ for a complete data set. In summary, the PCA framework can be used to distinguish the relevant partial correlation associations in the reduced data set from the associations due to measurement noise.

6.4.1 Generalized probabilistic principle component analysis

In the following, a generalization of probabilistic principle component analysis (PPCA) [11,58] will be introduced which is motivated by above notion of complementary principle subspaces $\mathbb{R}^M$ and $\mathbb{R}^{M_0}$ and which will enable one to determine the principle components of the inverse mean sample covariance matrix $(\bar{S}^{(p \neq j)})^{-1}$. Moreover, these

principle components, i.e. principle partial correlations, will be shown to be the incoming links of the unperturbed nodes in the data sample, which generates ($\bar{S}^{(p\neq j)}$).

The regular PPCA introduced by Bishop and Tipping [58] has a generative view of the observed data $\vec{x}$, that is naturally understood as a map from a lower dimensional principle subspace $\mathbb{R}^M$ into the data space $\mathbb{R}^D$. The main idea is that the observed data $\vec{x}_n \in \mathbb{R}^D$ can be generated by the sum of unobserved (hidden) random variable $\vec{z} \in \mathbb{R}^M$ from a lower dimensional subspace, called principle subspace, and multivariate Gaussian noise from the complementary subspace. The unobserved variables $\vec{z}$ represent the part of the observed data that is due to the actual examined (physical) process, which could have been measured directly in the absence of noise, i.e. other processes that interact with the examined process. Since subspace $\mathbb{R}^M$ represents the actual examined process it is called *principle* subspace. On the other hand, the noise random variables from the complementary subspace $\mathbb{R}^{M_0}$ represent uncontrollable stochastic processes, which intermingle with the actual examined process, thereby distorting measurements.

The method of PPCA determines a new set of basis vectors for the observed data space $\mathbb{R}^D$ that enables one to separate observation due to the actual examined process from observations due to random noise. Hence it is possible to define principle random variables and noise random variable, which can explain the observed data. The new basis is given by the column vectors of a matrix $W_M \in \mathbb{R}^{D\times M}$, which maps any vector $\vec{z}$ from the principle subspace onto an observation $\vec{x}^{(\text{real})}$ without noise in the higher dimensional data space

$$\begin{aligned} &W_M \colon \mathbb{R}^M \to \mathbb{R}^D \\ &W_M \colon \vec{z}_n \mapsto \left(\vec{x}_n^{(\text{real})} - \vec{\mu}\right) \quad , \end{aligned} \tag{6.25}$$

where $\vec{\mu}$ is the expectation value $E[\vec{x}_n]$. The generative view of the noisy observation $\vec{x}_n$ using matrix W_M is then given by

$$\vec{x}_n - \vec{\mu} = W_M \cdot \vec{z}_n + \vec{\epsilon}_n \quad , \tag{6.26}$$

where $\vec{z}_n \in \mathbb{R}^M$ and $\vec{x}_n \in \mathbb{R}^D$ as in the previous section. The generative eq.(6.26) can be reformulated using the orthogonal projection matrix $I_M \in \mathbb{R}^{D\times D}$ and the matrix $W_D \in \mathbb{R}^{D\times D}$

$$\vec{x}_n - \vec{\mu} = W_D \cdot I_M \cdot \vec{a}_n + \vec{\epsilon}_n \quad , \tag{6.27}$$

where $\vec{a}_n \in \mathbb{R}^D$ is a vector in data space, which is sampled from the standard normal distribution $\mathcal{N}(\vec{a}|0, I_D)$. Matrix $W_D \in \mathbb{R}^{D\times D}$ consists of M first column vectors, that span the principle subspace $\mathbb{R}^M$, and $M_0 = D - M$ remaining column vectors, which span the complementary subspace $\mathbb{R}^{M_0}$ leading to a set of basis vectors of the whole data space $\mathbb{R}^D = \mathbb{R}^M \oplus \mathbb{R}^{M_0}$.

$$W_D := \Big(\; W_M \;\Big|\; W_{M_0} \; \Big) = \Big(\; W_M \;\Big|\; W_{D-M} \; \Big) \quad , \tag{6.28}$$

with block matrix $W_{M_0} \in \mathbb{R}^{D\times M_0}$ consisting of the basis vectors that span the complementary subspace $\mathbb{R}^{M_0}$. The orthogonal projection matrix $I_M = I_M \cdot I_M$ ($I_M \in \mathbb{R}^{D\times D}$) is a diagonal matrix, with the first M diagonal elements equal to one, while the remaining elements are zero.

$$I_M := \mathrm{diag}\Big(\overbrace{\underbrace{1,\ldots,1}_{M}, \underbrace{0,\ldots,0}_{M_0}}^{D} \Big) \tag{6.29}$$

The matrix product of W_D and I_M like in eq.(6.27), produces a map from the data space into the principle subspace.

$$W_D \cdot I_M := \Big(\; W_M \;\Big|\; 0 \; \Big) \tag{6.30}$$

Consequently, the formulation of eq.(6.27) is equivalent to the standard PPCA formulation eq.(6.26). In contrast to standard PPCA, the alternative formulation uses a map within the data space

$$\begin{aligned} &W_M \cdot I_M \colon \mathbb{R}^D \to \mathbb{R}^D \\ &W_M \cdot I_M \colon \vec{a}_n \mapsto \left(\vec{x}_n^{(\mathrm{real})} - \vec{\mu}\right) \quad , \end{aligned} \tag{6.31}$$

so that the subspace is expressed in terms of the basis vectors that span the data space. Further, the image $\mathrm{im}(W_M \cdot I_M)$ is equal to the principle subspace, while the nullspace $\mathrm{null}(W_M \cdot I_M)$ generates the complementary subspace.

The generative view of perturbation experiments eq.(6.18) looks very similar to the alternative PPCA formulation eq.(6.27). By identifying W_D with the inverse interaction matrix A^{-1} and the orthogonal projection I_M by the diagonal standard deviation matrix $B = \sigma_u \cdot I_B$,

$$\begin{aligned} \vec{x}_n - \vec{\mu} &= -A^{-1} \cdot \quad B \quad \cdot \vec{a}_n + \vec{\epsilon}_n \\ &= -A^{-1} \cdot \sigma_u \cdot I_B \cdot \vec{a}_n + \vec{\epsilon}_n \quad , \end{aligned} \tag{6.32}$$

one can see the similarity. The orthogonal projection matrix I_B is a diagonal matrix, with diagonal elements

$$(I_B)_{ii} = \begin{cases} 0 & \text{node } i \text{ unperturbed} \\ 1 & \text{node } i \text{ perturbed} \end{cases} \quad . \tag{6.33}$$

An example for I_B for a complete perturbation data set $P = D$ (see section 6.2 & section5.2.2) of a network of size $D = 3$, where perturbation data of node $j = 2$ is removed, is

$$I_B \overset{\substack{P=D,j=2 \\ D=3}}{=} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad . \tag{6.34}$$

The differences between the orthonormal project matrices I_M and I_B are that the non-zero elements of I_B must not be ordered contrary to the elements of I_M. This is actually not a real difference, since by changing the order of random variable notations I_B could be transformed into an ordered version like I_M. Regardless of the order of diagonal elements, both projections map onto the principle subspace of correlations.

In an additional step the alternative PPCA formulation can be generalized in a way that only distinguishes between complementary subspaces. This is achieved by introducing biological noise $\vec{b}_n \sim \mathcal{N}\left(\vec{b}|0, \sigma_\eta^2 \cdot I_D\right)$, which propagates through the

whole network and acts similar to perturbations but with less "strength". The biological noise has a standard deviation σ_η that is much smaller than the one, σ_u, caused by controlled perturbations. Hence, it has only relevance for the unperturbed nodes, while it can be neglected for the perturbed nodes. As a consequence, one can add an biological noise term to eq.(6.32) which only acts on the complementary subspace $\mathbb{R}^{M_0}$.

$$\begin{aligned}\vec{x}_n - \vec{\mu} &= -A^{-1} \cdot \quad B \quad \cdot \vec{a}_n \; - \; A^{-1} \cdot \quad \overline{B} \quad \cdot \vec{\eta}_n \\ &= -A^{-1} \cdot \sigma_u I_B \cdot \vec{a}_n \; - \; A^{-1} \cdot \sigma_\eta \overline{I}_{\overline{B}} \cdot \vec{\eta}_n \quad , \end{aligned} \tag{6.35}$$

where $\overline{I}_{\overline{B}}$ is introduced to be the matrix complement of projection I_B with $I_B + \overline{I}_{\overline{B}} = I_D$ and $\vec{\eta}_n \sim \mathcal{N}(\vec{\eta}|0, I_D)$ is sampled from the standard normal distribution. The complement $\overline{B} = \overline{I}_{\overline{B}}$ can be understood as the standard deviation matrix of the biological noise acting only on the complementary subspace $\mathbb{R}^{M_0}$. Matrix $\overline{B} = \sigma_\eta \cdot \overline{I}_{\overline{B}}$ is a diagonal matrix with elements equal to σ_η for unperturbed nodes and zero elements otherwise. Since it represents unperturbed nodes affected by biological noise, it has no factor σ_u like in the case of B.

$$\overline{B}_{jj} = \begin{cases} \sigma_\eta & \text{node } j \text{ unperturbed} \\ 0 & \text{node } j \text{ perturbed} \end{cases} \quad . \tag{6.36}$$

In eq.(6.35) the technical noise term $\vec{\epsilon}_n = \sigma_\epsilon \cdot I_D \cdot \vec{\xi}_n$ has been neglected since it is much smaller than biological noise, as explained in section 5.2.2. To summarize the magnitude of controlled perturbations σ_u, biological noise σ_η, and technical noise σ_ϵ with respect to each other, one can write following relation

$$\sigma_u \gg \sigma_\eta \gg \sigma_\epsilon \quad . \tag{6.37}$$

For an ordered sequence of perturbed and unperturbed nodes, one can write following identity

$$B + \overline{B} = \mathrm{diag}\Big(\overbrace{\underbrace{\sigma_u, \ldots, \sigma_u}_{M}, \underbrace{\sigma_\eta, \ldots, \sigma_\eta}_{M_0}}^{D} \Big) \quad , \tag{6.38}$$

whereas $M_0 = 1$ and $M = P - 1$ for an complete data set. The general form of PPCA can now be understood as a sum of two terms, where the first term represents a map from data space onto the principle subspace of correlations

$$A^{-1} \cdot I_B := \left(\begin{array}{c|c} A_M^{-1} & A_{M_0}^{-1} \end{array} \right) \cdot I_B = \left(\begin{array}{c|c} A_M^{-1} & 0 \end{array} \right) \quad , \tag{6.39}$$

and where the second term stands for a map from the data space onto the complementary subspace.

$$A^{-1} \cdot \overline{I}_{\overline{B}} := \left(\begin{array}{c|c} 0 & A_{M_0}^{-1} \end{array} \right) \quad . \tag{6.40}$$

As in section 6.3 the generative view in eq.(6.35) can be transformed into the probabilistic expression, resulting in a joint Gaussian distribution $\mathcal{N}(\vec{x}|\mu, C)$. This Gaussian distribution has the same expression like before, but with a different expression for the population covariance matrix C, which will derived similarly to the previous section 6.3.

$$\begin{aligned} C := cov[\vec{x}] &= E\left[(\vec{x}_n - E[\vec{x}_n]) \cdot (\vec{x_n} - E[\vec{x}_n])^T \right] \\ &= E\left[(A^{-1} \cdot B \cdot \vec{a}_n + A^{-1} \cdot \overline{B} \cdot \vec{\eta}_n) \cdot (A^{-1} \cdot B \cdot \vec{a}_n + A^{-1} \cdot \overline{B} \cdot \vec{\eta}_n)^T \right] \\ \Rightarrow C &= \underbrace{A^{-1} B \cdot B^T \left(A^{-1}\right)^T}_{\text{principle subspace of } C} + \underbrace{A^{-1} \overline{B} \cdot \overline{B}^T \left(A^{-1}\right)^T}_{\text{principle subspace of } C^{-1}} \quad , \end{aligned} \tag{6.41}$$

where folowing relations have been applied: $E[\vec{a}_n] = 0 = E[\vec{\eta}]$ and $E\left[\vec{a}_n \cdot \vec{a}_n^T\right] = I_D = E[\vec{\eta}_n \cdot \vec{\eta}_n^T]$. By comparing the modeled population covariance matrix from eq.(6.41) with the mean sample covariance matrix from eq.(6.23) one can already guess, that the first and second term can be represented with respect to complementary sets of eigenvectors of covariance matrix C.

After having obtained the general form for the covariance matrix C, an expression

for the inverse population covariance matrix C^{-1} (partial correlations) will be derived by taking the inverse of eq.(6.41).

$$\begin{aligned} C = & \; A^{-1}\left[B\cdot B^T \; + \; \overline{B}\cdot\overline{B}^T\right]\left(A^{-1}\right)^T \\ = & \; A^{-1}\left[\sigma_u B \; + \; \sigma_\eta \overline{B}\right]\left(A^{-1}\right)^T \\ \overset{eq.(6.38)}{=} & \; A^{-1}\cdot \mathrm{diag}\Big(\overbrace{\underbrace{\sigma_u^2,\ldots,\sigma_u^2}_{M},\underbrace{\sigma_\eta^2,\ldots,\sigma_\eta^2}_{M_0}}^{D}\Big)\cdot\left(A^{-1}\right)^T \end{aligned} \tag{6.42}$$

This new expression for C can be easily inverted, keeping in mind that $-A$ is positive definite. Without loss of generality it is assumed that the diagonal matrix elements of $B\cdot B^T \; + \; \overline{B}\cdot\overline{B}^T$ follow an ordered sequence, which makes it visually more simple. From eq.(6.42) follows

$$\begin{aligned} C^{-1} &= A^T\cdot \mathrm{diag}\Big(\overbrace{\underbrace{\frac{1}{\sigma_u^2},\ldots,\frac{1}{\sigma_u^2}}_{M},\underbrace{\frac{1}{\sigma_\eta^2},\ldots,\frac{1}{\sigma_\eta^2}}_{M_0}}^{D}\Big)\cdot A \\ &= A^T\cdot\left[\frac{1}{\sigma_u^2}I_B \; + \; \frac{1}{\sigma_\eta^2}\overline{I}_{\overline{B}}\right]\cdot A \\ &= A^T\cdot\left[\left(\frac{1}{\sigma_u}I_B^T\right)\cdot\left(\frac{1}{\sigma_u}I_B\right) \; + \; \left(\frac{1}{\sigma_\eta}\overline{I}_{\overline{B}}^T\right)\cdot\left(\frac{1}{\sigma_\eta}\overline{I}_{\overline{B}}\right)\right]\cdot A \\ \Rightarrow C^{-1} &= \frac{1}{\sigma_u^2}\cdot\underbrace{A^T\cdot I_B^T\cdot I_B\cdot A}_{\text{noise subspace}} + \frac{1}{\sigma_\eta^2}\cdot\underbrace{A^T\cdot\overline{I}_{\overline{B}}^T\cdot\overline{I}_{\overline{B}}\cdot A}_{\text{principle subspace of } C^{-1}} \end{aligned} \tag{6.43}$$

Eq.(6.43) states the general expression for the inverse population covariance matrix, representing partial correlations. As mentioned before, the principle subspace of C^{-1} is given by the unperturbed nodes denoted by the image of projection matrix $\overline{I}_{\overline{B}}$. Therefore, the first term of eq.(6.43) can be regarded as noise subspace with *not important components*, which can be understood by partial correlations between perturbed nodes. The information about partial correlations between perturbed nodes is destroyed as explained in section 6.2, so that any observed partial correlations between perturbed nodes are regarded as noise. Further, one can see from the prefactor $1/\sigma_u \ll 1/\sigma_\eta$ that the second term of eq.(6.43) is the dominating part, while the first

term approaches zero for very large controlled perturbations σ_u. Since, the goal is to model the inverse mean population covariance matrix denoted by C^{-1}, a couple of approximation will be done to the the first term. This will simplify eq.(6.43) and bring the formulation closer to a PPCA representation of partial correlations.

$$\begin{aligned} C^{-1} &\approx \frac{1}{\sigma_u^2} \cdot \operatorname{diag}\left(A^T I_B^T I_B A\right) \; + \; \frac{1}{\sigma_\eta^2} \cdot A^T \cdot \overline{I}_{\overline{B}}^T \cdot \overline{I}_{\overline{B}} \cdot A \\ &\approx \frac{1}{\sigma_u^2} \cdot \operatorname{diag}\left(\operatorname{mean}\left[\operatorname{diag}\left(A^T I_B^T I_B A\right)\right]\right) \; + \; \frac{1}{\sigma_\eta^2} \cdot A^T \cdot \overline{I}_{\overline{B}}^T \cdot \overline{I}_{\overline{B}} \cdot A \\ &\approx \frac{1}{\sigma_u^2} \cdot \alpha^2 \cdot I_D \; + \; \frac{1}{\sigma_\eta^2} \cdot A^T \cdot \overline{I}_{\overline{B}}^T \cdot \overline{I}_{\overline{B}} \cdot A \\ &\approx \frac{\alpha^2 \cdot \sigma_\eta^2}{\sigma_u^2} \cdot I_D \; + \; A^T \cdot \overline{I}_{\overline{B}}^T \cdot \overline{I}_{\overline{B}} \cdot A \quad , \end{aligned} \tag{6.44}$$

where α is an estimate for the noise subspace, which can be used for assessing the limits of this approximation, like in the case of regular PPCA. The fraction σ_η^2/σ_u^2 can be regarded as a noise-to-signal ratio, where perturbation variance σ_u^2 is the signal and the variance σ_η represents biological noise. For a signal which is much larger than the noise σ_η the first term of eq.(6.44) converges to zero.

To get closer to the final PPCA formulation of partial correlations, one should understand the effect of the projection matrix $\overline{I}_{\overline{B}}$ on the interaction matrix A, whereas an ordered sequence of M perturbed and M_0 unperturbed nodes is considered without loss of generality.

$$\overline{I}_{\overline{B}} \cdot A := \overline{I}_{\overline{B}} \cdot \left(\frac{A_M}{A_{M_0}} \right) = \left(\frac{0}{A_{M_0}} \right) \quad , \tag{6.45}$$

where $A_{M_0} \in \mathbb{R}^{M_0 \times D}$ is a block matrix whose rows represent the incoming link strength of unperturbed nodes or, in terms of PCA, whose row vectors span the principle subspace of C^{-1}, respectively. For a complete dataset block matrix A_{M_0} reduces to the row vector $A_{j,\mathrm{row}}$, whose elements are the incoming links of unperturbed node j introduced in section 6.2. On the other hand, the rows of $A_M \in \mathbb{R}^{M \times D}$ stand for the incoming link strengths of the perturbed nodes, which span the complementary

subspace and are assumed to be due to noise only. Note, that A_M does not stand for part of the real network, it only reflects the observation for the reduced data set leading to $\bar{S}^{(p\neq j)}$. Contrary, A_{M_0} represents the real link strengths of unperturbed links.

The eq.(6.44) can be reformulated exclusively by means of block matrix A_{M_0}, similarly to the reformulation of eq.(6.26) to eq.(6.27), but only reverse. Matrix A_{M_0} is a map from the principle subspace $\mathbb{R}^{M_0}$ of partial correlations onto the higher dimensional (inverse) data space $\mathbb{R}^D$, that is $A_{M_0} : \mathbb{R}^{M_0} \to \mathbb{R}^D$. By utilizing this perspective, the inverse population covariance matrix will be

$$C^{-1} = A_{M_0}^T \cdot A_{M_0} \; + \; \sigma_{\text{nsr}}^2 \cdot I_D \quad , \tag{6.46}$$

where $\sigma_{\text{nsr}}^2 := \alpha^2 \cdot \sigma_\eta^2/\sigma_u^2$. Note, that for convenience the equal sign has been used in eq.(6.46), despite it is still an approximation. Finally, the joint Gaussian probability distribution from eq.(6.22) can naturally be expressed with respect to the inverse covariance matrix C^{-1} by applying identity $\det(C) = 1/\det(C^{-1})$ for C being a non singular quadratic matrix.

$$\begin{aligned} p\left(\vec{x}|A_{M_0}\right) &= \mathcal{N}\big(\vec{x} \mid \vec{\mu}, C^{-1}\left(A_{M_0}, \sigma_{\text{nsr}}^2\right)\big) \\ &= \left(\frac{1}{2\pi}\right)^{D/2} \cdot \left[\det(C^{-1})\right]^{1/2} \cdot \exp\left[-\frac{1}{2} \cdot (\vec{x}-\vec{\mu})^T C^{-1} (\vec{x}-\vec{\mu})\right] \end{aligned} \tag{6.47}$$

This probability function does depend directly on the link strength matrix A_{M_0}, since the expression for the inverse mean population covariance matrix C^{-1} from (6.46) does depend on A_{M_0}.

6.4.2 Maximum likelihood PCA for partial correlations

In order to infer the incoming link strengths A_{M_0} of unperturbed nodes from the reduced data set $\{X^{(p\neq j)}\}$ a maximum likelihood approach similar to the one from PPCA [11] will be implemented, next. Assuming data $\{X^{(p\neq j)}\}$ to be independently

distributed, the full expression of the likelihood function is

$$
\begin{aligned}
& p\left(\{X^{(p\neq j)}\}|A_{M_0},\sigma^2_{\mathrm{nsr}},\vec{\mu}\right) \\
&= \prod_{p=1,p\neq j}^{Q}\prod_{n=1}^{N}\left(\frac{1}{2\pi}\right)^{D/2}\left[\det(C^{-1})\right]^{1/2}\exp\left[-\frac{1}{2}\cdot\left(\vec{x}_n^{(p)}-\vec{\mu}\right)^T C^{-1}\left(\vec{x}_n^{(p)}-\vec{\mu}\right)\right] \\
&= \prod_{l=1}^{N\cdot Q}\left(\frac{1}{2\pi}\right)^{D/2}\left[\det(C^{-1})\right]^{1/2}\exp\left[-\frac{1}{2}\cdot\left(\vec{x}_l-\vec{\mu}\right)^T C^{-1}\left(\vec{x}_l-\vec{\mu}\right)\right], \qquad (6.48)
\end{aligned}
$$

where $Q = P-1$ for j being a removed perturbation experiment and $Q = P$ if there is no perturbation experiment for node j available (incomplete data set). The number of included perturbation experiments, denoted by Q, and the number of replicate experiments N for each perturbation experiment is combined in the last step in the above equation. The corresponding log likelihood $\mathcal{L}$ can be written as

$$
\begin{aligned}
& \mathcal{L} := \ln\left(p\left(\{X^{(p\neq j)}\}|A_{M_0},\sigma^2_{\mathrm{nsr}},\vec{\mu}\right)\right) \\
&= -\frac{NQ}{2}\left(D\ln(2\pi) - \ln\left[\det(C^{-1})\right] + \frac{1}{NQ}\sum_{n=1}^{N\cdot Q}\left(\vec{x}_l-\vec{\mu}\right)^T C^{-1}\left(\vec{x}_l-\vec{\mu}\right)\right). \qquad (6.49)
\end{aligned}
$$

The next steps are constituted of maximizing above log likelihood $\mathcal{L}$ with respect to its unknown parameters, namely the population mean $\vec{\mu}$, the partial link strength matrix $A_{M_0} \in \mathbb{R}^{M_0\times D}$ and the noise-to-signal ratio σ^2_{nsr}. Maximizing $\mathcal{L}$ with respect to the population mean $\vec{\mu}$ yields the sample mean over the reduced data set $\{X^{(p\neq j)}\}$.

$$
\vec{\mu}_{\mathrm{ML}} = \frac{1}{NQ}\sum_{l=1}^{NQ}\vec{x}_l =: \overline{\vec{x}} \qquad (6.50)
$$

By substituting the population mean $\vec{\mu}$ with the sample mean $\overline{\vec{x}}$ the log likelihood of eq.(6.49) can be expressed through the mean sample covariance matrix $\bar{S}^{(p\neq j)}$ introduced in section 6.2.

$$
\begin{aligned}
\mathcal{L} &= \ln\left(p\left(\bar{S}^{(p\neq j)}|A_{M_0},\sigma^2_{\mathrm{nsr}},\vec{\mu}_{\mathrm{ML}}\right)\right) \\
&= -\frac{NQ}{2}\Big(D\ln(2\pi) - \ln\left[\det(C^{-1})\right] + \mathrm{Tr}\left[C^{-1}\cdot\bar{S}^{(p\neq j)}\right]\Big), \qquad (6.51)
\end{aligned}
$$

where the mean sample covariance matrix, which is computed over the reduced data set and can also be understood as the mean over sample covariance matrices $S^{(p)}$ of each perturbation experiment.

$$\begin{aligned}\bar{S}^{(p\neq j)} :=& \frac{1}{NQ}\sum_{l=1}^{NQ}\left(\vec{x}_l - \overline{\vec{x}}\right)\left(\vec{x}_l - \overline{\vec{x}}\right)^T \\ =& \frac{1}{Q}\sum_{p=1}^{Q}\frac{1}{N}\sum_{n=1}^{N}\left(\vec{x}_n^{(p)} - \overline{\vec{x}}^{(p)}\right)\left(\vec{x}_n^{(p)} - \overline{\vec{x}}^{(p)}\right)^T = \frac{1}{Q}\sum_{p=1}^{Q} S^{(p)} \end{aligned} \tag{6.52}$$

The derivative [2] of the log likelihood function $\mathcal{L}$ of eq.(6.51) with respect to A_{M_0} will be determined and set equal to zero to obtain the maximum likelihood estimate $A^*_{M_0}$ under the total connectivity assumption. It is sufficient to compute the derivative of the second and third term of $\mathcal{L}$ separately, while the first term can be neglected since it does not depend on A_{M_0}. Keeping in mind that the expression for the inverse covariance matrix C^{-1} is given by eq.(6.46), the derivatives of the second and third term are

$$\frac{d}{dA_{M_0}} \ln\left[\det\left(A_{M_0}^T \cdot A_{M_0} \;+\; \sigma^2_{\text{nsr}} \cdot I_D\right)\right] = 2 \cdot A_{M_0} \cdot C \tag{6.53}$$

$$\frac{d}{dA_{M_0}} \operatorname{Tr}\left[\left(A_{M_0}^T \cdot A_{M_0} \;+\; \sigma^2_{\text{nsr}} \cdot I_D\right) \cdot \bar{S}^{(p\neq j)}\right] = 2 \cdot A_{M_0} \cdot \bar{S}^{(p\neq j)} \quad . \tag{6.54}$$

Combining the single derivative results yields the derivative of the total log likelihood function.

$$\frac{d\mathcal{L}\left(A_{M_0}\right)}{dA_{M_0}} = QN\left(A_{M_0} \cdot C \;-\; A_{M_0} \cdot \bar{S}^{(p\neq j)}\right) \stackrel{!}{=} 0 \tag{6.55}$$

$$\Leftrightarrow \quad A_{M_0} \cdot C = A_{M_0} \cdot \bar{S}^{(p\neq j)}$$

$$\Leftrightarrow \quad \left(\bar{S}^{(p\neq j)}\right)^{-1} \cdot C \cdot A_{M_0}^T = A_{M_0}^T \tag{6.56}$$

The last equation,eq.(6.56), depends on the inverse mean sample covariance matrix $\left(\bar{S}^{(p\neq j)}\right)^{-1}$, representing observed partial correlations in the reduced data set. There

[2] In this context, the first derivative is sufficient to find the maximum, since the likelihood function is considered to be a Gaussian with a single maximum.

are three different solutions to this equation, namely $A_{M_0} = 0$, $C = \bar{S}^{(p \neq j)}$ and a third case with $A_{M_0} \neq 0$ and $C \neq \bar{S}^{(p \neq j)}$, which is the interesting one. The second case $C = \bar{S}^{(p \neq j)}$ does not reduce the dimensionality like it is the case in Principle Component Analysis (PCA), which means that it is not in agreement with the notion of a underlying principle subspace. To reduce the dimensionality, a singular value decomposition (SVD) of link strength matrix $A_{M_0} \in \mathbb{R}^{M_0 \times D}$ ($M_0 < D$) will be introduced

$$\begin{aligned} A_{M_0} &= V^T \Sigma U \\ A_{M_0}^T &= U^T \Sigma V \quad , \end{aligned} \tag{6.57}$$

where U is an orthonormal $M_0 \times D$ matrix with $UU^T = I_{M_0}$, V is an orthonormal $M_0 \times M_0$ matrix with $V^T V = I_M = VV^T$, representing a rotation in the principle subspace of partial correlations, and $\Sigma = \mathrm{diag}(l_1, ..., l_{M_0})$ is a $M_0 \times M_0$ diagonal matrix with singular values l_j. Matrix U is chosen to be a row vector matrix, corresponding to the rows of A_{M_0}, which stand for the incoming links of the unperturbed nodes. Using the SVD and the identity $V \cdot V^T = I_{M_0}$ the mean population covariance matrix C has the following form

$$C = \left(A_{M_0}^T \cdot A_{M_0} \; + \; \sigma_{\mathrm{nsr}}^2 \cdot I_D \right)^{-1} \tag{6.58}$$

$$= \left(\sigma_{\mathrm{nsr}}^2 \cdot I_D \; + \; U^T \Sigma^2 U \right)^{-1} \tag{6.59}$$

Above expression, eq.(6.59), can be rewritten by utilizing the Woodbury identity[3], which leads to

$$\begin{aligned} C &= \sigma_{\mathrm{nsr}}^{-2} I_D - \sigma_{\mathrm{nsr}}^{-2} I_D U^T \left[\Sigma^{-2} + U \sigma_{\mathrm{nsr}}^{-2} I_D U^T \right]^{-1} U \sigma_{\mathrm{nsr}}^{-2} I_D \\ &= \sigma_{\mathrm{nsr}}^{-2} I_D - U^T \sigma_{\mathrm{nsr}}^{-2} I_{M_0} \left[\Sigma^{-2} + U U^T \sigma_{\mathrm{nsr}}^{-2} I_{M_0} \right]^{-1} \sigma_{\mathrm{nsr}}^{-2} I_{M_0} U \\ \Rightarrow C &= \sigma_{\mathrm{nsr}}^{-2} I_D - U^T \sigma_{\mathrm{nsr}}^{-2} I_{M_0} \left[\Sigma^{-2} + \sigma_{\mathrm{nsr}}^{-2} I_{M_0} \right]^{-1} \sigma_{\mathrm{nsr}}^{-2} I_{M_0} U \quad , \end{aligned} \tag{6.60}$$

[3]Woodbury identity: $(L + U^T K U)^{-1} = L^{-1} - L^{-1} U^T (K^{-1} + U L^{-1} U^T)^{-1} U L^{-1}$ with $L : D \times D$ matrix, $U^T : D \times M_0$ matrix, $U^T : M_0 \times D$ matrix, and $K : M_0 \times M_0$ matrix

where identity matrix $I_{M_0} \in \mathbb{R}^{M_0 \times M_0}$ should be distinguished from identity matrix $I_D \in \mathbb{R}^{D \times D}$. Further, $I_D U^T = U^T = U^T I_M$ and $\sigma_{\text{nsr}} U^T = U^T \sigma_{\text{nsr}}$ has been applied.

After having obtained the SVD eq.(6.57) and the associated expression for C in eq.(6.60), one can return to eq.(6.56) and insert those equations.

$$\begin{aligned}
&\left(\bar{S}^{(p\neq j)}\right)^{-1}\left[\sigma_{\text{nsr}}^{-2} I_D - U^T \sigma_{\text{nsr}}^{-2} I_{M_0}\left[\Sigma^{-2} + \sigma_{\text{nsr}}^{-2} I_{M_0}\right]^{-1} \sigma_{\text{nsr}}^{-2} I_{M_0} U\right] U^T \Sigma V = U^T \Sigma V \\
&\Leftrightarrow \left(\bar{S}^{(p\neq j)}\right)^{-1} U^T \sigma_{\text{nsr}}^{-2} I_{M_0} - \left(\bar{S}^{(p\neq j)}\right)^{-1} U^T \sigma_{\text{nsr}}^{-2} I_{M_0}\left[\Sigma^{-2} + \sigma_{\text{nsr}}^{-2} I_{M_0}\right]^{-1} \sigma_{\text{nsr}}^{-2} I_{M_0} = U^T \\
&\Leftrightarrow \left(\bar{S}^{(p\neq j)}\right)^{-1} U^T \left[\sigma_{\text{nsr}}^{-2} I_{M_0} - \sigma_{\text{nsr}}^{-2} I_{M_0}\left[\Sigma^{-2} + \sigma_{\text{nsr}}^{-2} I_{M_0}\right]^{-1} \sigma_{\text{nsr}}^{-2} I_{M_0}\right] = U^T
\end{aligned} \tag{6.61}$$

After above manipulations, the Woodbury identity will be applied again to eq.(6.61) to arrive at a more compact form.

$$\begin{aligned}
&\left(\bar{S}^{(p\neq j)}\right)^{-1} U^T \left[\Sigma^2 + \sigma_{\text{nsr}}^2 I_{M_0}\right]^{-1} = U^T \\
&\Rightarrow \left(\bar{S}^{(p\neq j)}\right)^{-1} U_{M_0}^T = \left[\Sigma_{M_0}^2 + \sigma_{\text{nsr}}^2 I_{M_0}\right] U_{M_0}^T = \Lambda_{M_0} U_{M_0}^T \quad ,
\end{aligned} \tag{6.62}$$

Obviously, eq.(6.62) is an eigenvalue equation where the eigenvalues, denoted by diagonal matrix Λ_{M_0}, and eigenvectors, contained as row vectors in matrix U_{M_0}, of $\left(\bar{S}^{(p\neq j)}\right)^{-1}$ must be determined. As a consequence, the problem of maximizing the likelihood function $\mathcal{L}$ with respect to link strength matrix A_{M_0} can be understood as an eigenvalue problem of the inverse mean sample covariance matrix $\left(\bar{S}^{(p\neq j)}\right)^{-1}$, i.e. the observed partial correlations. The index M_0 has been added to U, i.e. U_{M_0}, to indicate that $U_{M_0}^T$ contains only the principle eigenvectors that span the principle subspace of $\left(\bar{S}^{(p\neq j)}\right)^{-1}$.

To infer the incoming links A_{M_0} of the unperturbed nodes, the first M_0 largest eigenvalues and their corresponding eigenvectors have to be computed by eq.(6.62). From the eigenvalues Λ_{M_0} it it possible to determine the singular values Σ_{M_0}

$$\Sigma_{M_0} = \left[\Lambda_{M_0} - \sigma_{\text{nsr}}^2 I_{M_0}\right]^{\frac{1}{2}} \tag{6.63}$$

Finally, the expression for the maximum likelihood estimate $A_{M_0}^*$ for the incoming

link strengths of the unperturbed nodes is obtained by inserting the singular value $\Sigma_{M_0} \in \mathbb{R}^{M_0 \times M_0}$ and eigenvectors $U_{M_0} \in \mathbb{R}^{M_0 \times D}$ into the SVD of A_{M_0} of eq.(6.57).

$$A^*_{M_0} = V^T_{M_0} \cdot \left[\Lambda_{M_0} - \sigma^2_{\text{nsr}} I_{M_0}\right]^{\frac{1}{2}} \cdot U_{M_0} \tag{6.64}$$

The orthonormal rotation matrix V_{M_0} can not be determined by the here presented maximum likelihood approach, which similarly occurs in the case of regular PPCA [58]. In other words, the exact values of $A^*_{M_0}$ are non-identifiable with respect to a rotation in the principle subspace spanned by the row vectors of matrix U_{M_0}. The reason behind is the rotational invariance of the modeled inverse covariance matrix C^{-1} , which was derived in eq.(6.59). Following the same derivation as for regular PPCA [58], it can be shown that Λ_{M_0} comprises the M_0 largest eigenvalues. Further, in accordance with [58] , the noise-to-signal ratio σ^2_{nsr} is the average over the $M = D - M_0$ discarded eigenvalues $\{\lambda_i\}$ of the complementary noise subspace $\mathbb{R}^M$.

$$\sigma^2_{\text{nsr}} = \frac{1}{D - M_0} \sum_{i=M_0+1, i\neq j}^{D} \lambda_i \tag{6.65}$$

In the following, above solution will be specified for two different cases, namely the complete and incomplete data set introduced in section5.2.2

Complete data set $M_0 = 1$ ($P = D$): As a reminder, the complete data set consists of as many single controlled perturbation experiments P as nodes in the network D. In this case the corresponding reduced set $\{X^{(p\neq j)}\}$ comprises all perturbation experiments except for the one of node j, whose incoming links can be inferred. From a linear algebra perspective, the principle subspace has only one dimension, i.e. $M_0 = 1$. Consequently, the ML estimated link strength matrix $A^*_{M_0}$ from eq.(6.64) reduces to the row vector $A^*_{j,\text{row}}$, representing the incoming link strength of node j

under the total connectivity assumption.

$$A^*_{j,\mathrm{row}} = v \cdot \left[\lambda_j - \sigma^2_{\mathrm{nsr}}\right]^{\frac{1}{2}} \cdot U_{j,\mathrm{row}} \tag{6.66}$$

$$\left(A^*_{j1} \dots A^*_{jD}\right) = v \cdot \left[\lambda_j - \sigma^2_{\mathrm{nsr}}\right]^{\frac{1}{2}} \cdot \left(U_{j1} \dots U_{jD}\right) \quad , \tag{6.67}$$

where the rotational matrix V_{M_0} reduces to scalar $v \in \{-1, 1\}$ and λ_j denotes the only eigenvalue of $\bar{S}^{(p\neq j)}$ associated with the principle subspace, which is spanned by the only eigenvector $U_{j,\mathrm{row}}$. The non-identifiability problem that undetermined scalar v states can be solved by considering the relative link strength, which is the link strength A^*_{ji} relative to the link strength A^*_{jj} of node j on itself, i.e. the degradation rate or restoring force, respectively.

$$\widehat{A}^*_{ji} := \frac{A^*_{ji}}{A^*_{jj}} = \frac{v \cdot \left[\lambda_j - \sigma^2_{\mathrm{nsr}}\right]^{\frac{1}{2}} \cdot U_{ji}}{v \cdot \left[\lambda_j - \sigma^2_{\mathrm{nsr}}\right]^{\frac{1}{2}} \cdot U_{jj}} = \frac{U_{ji}}{U_{jj}} \tag{6.68}$$

$$\Rightarrow \widehat{A}^*_{jj} = 1 \tag{6.69}$$

The relative link strength quantity does not determine the real strength of interactions, but it is a sufficient measure to infer the network structure. In contrast to the response matrix method (see section5.3.3), the inferred link strengths of the here presented method are comparable, which makes it possible to use it to infer the network structure. This comparability is restricted to networks in which the real restoring force A_{jj} is the same for all nodes.

A very interesting consequence of the relative link strength $\widehat{A}^*_{ji}$ for the complete data set, is that it only depends on the eigenvector $U_{j,\mathrm{row}}$. The eigenvectors of the covariance matrix $\bar{S}^{(p\neq j)}$ and its inverse $(\bar{S}^{(p\neq j)})^{-1}$ are the same, as mentioned in the beginning of this section (see eq.(6.23)). Therefore, one can infer the relative link strength $\widehat{A}^*_{ji}$ by computing the eigenvectors of the covariance matrix $\bar{S}^{(p\neq j)}$, whereas the transposed eigenvectors corresponding to the smallest eigenvalue of $\bar{S}^{(p\neq j)}$ must be chosen. This different approach can be practically more easily applicable, if $\bar{S}^{(p\neq j)}$ is close to singularity. Concluding, it is possible to infer the whole network by inferring the incoming links of each node separately step by step using P different mean sample

covariance matrices $\bar{S}^{(p\neq j)}$.

Incomplete data set $M_0 > 1$ ($P < D$): The complete data set case is the focus of this work, nevertheless the here presented framework is capable of treating incomplete data. For incomplete data the dimension M_0 of the principle subspace increases to multidimensional subspace, so that eq.(6.64) will have the form

$$\begin{pmatrix} A^*_{11} & \cdots & A^*_{1D} \\ \vdots & \ddots & \vdots \\ A^*_{M_0 1} & \cdots & A^*_{M_0 D} \end{pmatrix} = \begin{pmatrix} V^T_{11} & \cdots & V^T_{1M_0} \\ \vdots & \ddots & \vdots \\ V^T_{M_0 1} & \cdots & V^T_{M_0 M_0} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & & 0 \\ & \ddots & \\ 0 & & \Sigma_{M_0 M_0} \end{pmatrix} \begin{pmatrix} U_{11} & \cdots & U_{1D} \\ \vdots & \ddots & \vdots \\ U_{M_0 1} & \cdots & U_{M_0 D} \end{pmatrix}$$

Then each element of $A^*_{M_0}$ has the following expression

$$A^*_{ji} = \sum_{k=1}^{M_0} V^T_{jk} \cdot \Sigma_{kk} \cdot U_{ki} \quad , \tag{6.70}$$

for $j \in \{1, \ldots, M_0\}$ denoting the unperturbed nodes in the reduced data set and $i \in \{1, \ldots, D\}$ going over all nodes. Consequently, the relative link strength $\widehat{A}^*_{ji}$ for an incomplete data set is obtained by

$$\widehat{A}^*_{ji} = \frac{\sum_{k=1}^{M_0} V^T_{jk} \cdot \Sigma_{kk} \cdot U_{ki}}{\sum_{k=1}^{M_0} V^T_{jk} \cdot \Sigma_{kk} \cdot U_{kj}} \tag{6.71}$$

As can be seen from above equation, the problem of the non-identifiable rotational vector V_{M_0} is not solved by using the relative link strength for an incomplete data set. However, it is possible to use a numerical approach, called "rescaling method", as well as some assumptions for V_{M_0} and Σ_{M_0} to arrive at an identifiable problem. As mentioned earlier the incomplete data case and hence this numerical approach is not part of the here presented work and was only summarized for the sake of completeness.

6.5 Markov chain Monte Carlo sampling over posterior distribution of network structures

The goal of this section is to find the most probable network structure and hence to improve the network inference results of the maximum likelihood estimate introduced in the former section, section 6.4. Adding prior information about network sparsity, enables one to draw samples from the posterior over network structures (given the link strength) by means of Markov chain Monte Carlo (MCMC) sampling. In section 6.4 the link strengths A^*_{ji} between nodes were estimated from correlation information by assuming that every node is affected by the total network, i.e. total connectivity of the network structure $G = \mathbb{1}$. The next step is to find an estimate for the network structure G, which is most likely considering the link strength ML estimate A^* and the data. This way it is possible to find more precisely the real interactions (links) in contrast to A^* which incorporates the total connectivity assumption.

6.5.1 The likelihood function in dependence of the network structure

To express the likelihood function with respect to a variable network structure or graph G, one can return to eq.(6.7). Assuming that the link strength A is given by the maximum likelihood estimate A^* and that the structure G is a variable parameter, the joint likelihood can be written as:

$$\begin{aligned}
&p\left(\{X^p\}|G, A^*\right) \\
&= p\left(\{X^{p\neq 1}\}, \ldots, \{X^{p\neq D}\}|G, A^*\right) = \prod_{j=1}^{D} p(\{X^{p\neq j}\}|G_{j,row}, A^*_{j,row}) \\
&= p\left(\bar{S}^{p\neq 1}, \ldots, \bar{S}^{p\neq D}|G, A^*\right) = \prod_{j=1}^{D} p(\bar{S}^{p\neq j}|G_{j,row}, A^*_{j,row}) \\
&= p\left(\bar{S}^{p\neq 1}, \ldots, \bar{S}^{p\neq D}|G\right) = \prod_{j=1}^{D} p(\bar{S}^{p\neq j}|G_{j,row}) \qquad (6.72)
\end{aligned}$$

In the last step of eq.(6.72) the notation of the joint and conditional likelihood functions are simplified by neglecting to write down explicitly the A^* and $A^*_{\mathrm{j,row}}$ dependency. In the current case, the network structure $G \in \{0,1\}^{D\times D}$ can be arranged in any possible way, while the total connectivity assumption in the previous section led to the structure $G = \mathbb{1}$. In other words, the conditional likelihood distribution $p(\bar{S}^{p\neq j}|G_{j,row})$ given a certain incoming network structure $G_{j,row}$ of node j, expresses the probability that the activity X_j of node j can be explained by the direct interaction of a subset of network nodes, i.e. the immediate parents $Pa_{G_{j,row}}(X_j)$. Therefore, structure G is the model parameter, which has to be inferred from training data. Further, the following derivations are based on a complete data set, so that the maximum likelihood estimate $A^*_{\mathrm{j,row}} \equiv A^*_{M_0=1}$ of eq.(6.67) will be applied.

The conditional likelihood function $p(\bar{S}^{p\neq j}|G_{j,row}, A^*_{\mathrm{j,row}})$ of the mean sample covariance matrix $\bar{S}^{p\neq j}$ given the incoming link structure $G_{j,row}$ and the incoming link strength ML estimate $A^*_{\mathrm{j,row}}$ can be derived from the log likelihood expression in eq.(6.51)

$$\begin{aligned} & p\left(\bar{S}^{p\neq j}|G_{j,row}, A^*_{j,row}\right) \\ &= \left(\frac{1}{2\pi}\right)^{\frac{DNQ}{2}} \cdot \left(\det\left[C^{-1}\right]\right)^{\frac{NQ}{2}} \cdot \exp\left(-\frac{NQ}{2} \cdot \mathrm{Tr}\left[C^{-1}\cdot \bar{S}^{p\neq j}\right]\right) \end{aligned} \tag{6.73}$$

This likelihood function can be understood as a measure which compares the modeled correlations represented by C^{-1} with the observed correlations given by $\bar{S}^{p\neq j}$. The closer C gets to $\bar{S}^{p\neq j}$, the higher the probability, and vice versa. Hence , the conditional likelihood could also be expressed as the likelihood $p\left(\bar{S}^{p\neq j}|C^{-1}\left(G_{j,row}, A^*_{j,row}\right)\right)$ of the sample covariance matrix given the modeled population covariance matrix. At the first glance above equation, eq.(6.73), looks similar to the likelihood function which was maximized to obtain the link strength $A^*_{\mathrm{j,row}}$, i.e. eq.(6.51). However, the inverse population covariance matrix C^{-1} of eq.(6.73) depends on the constant row vector $A^*_{\mathrm{j,row}} = \left(A^*_{j1}\dots A^*_{jD}\right)$ and the variable structure $G_{j,\mathrm{row}} = (G_{j1}\dots G_{jD})$. This general expression for the inverse population covariance matrix is obtained from

eq.(6.46) by forcing the incoming structure upon the ML link strength.

$$C^{-1} = \left(A^*_{j,\mathrm{row}} \circ G_{j,\mathrm{row}}\right)^T \cdot \left(A^*_{j,\mathrm{row}} \circ G_{j,\mathrm{row}}\right) + \sigma^2_{\mathrm{nsr}} \cdot I_D \quad , \tag{6.74}$$

where " $\circ$ " is the Hadamard product, which stands for an element wise matrix multiplication. By incorporating the element wise multiplication of the structure $G_{j,\mathrm{row}}$ to link strength $A^*_{j,\mathrm{row}}$, incoming link strengths can be set exactly to zero. Notice that eq.(6.46) and eq.(6.74) are identical under the total connectivity assumption, where each element of $G_{ji} = 1$ is equal to one and hence can be neglected. This general expression can be further specified by inserting the ML solution $A^*_{j,\mathrm{row}}$ of the link strength given by eq.(6.67), leading to

$$C^{-1} = \left(\lambda_j - \sigma^2_{\mathrm{nsr}}\right) \cdot \left(U_{j,\mathrm{row}} \circ G_{j,\mathrm{row}}\right)^T \cdot \left(U_{j,\mathrm{row}} \circ G_{j,\mathrm{row}}\right) + \sigma^2_{\mathrm{nsr}} \cdot I_D \quad . \tag{6.75}$$

The non identifiable parameter $v \in \{-1, 1\}$ drops out, since it enters quadratically above expression, whereas σ^2_{nsr} is given by eq.(6.65) for a one dimensional principle subspace, i.e. $M_0 = 1$.

$$\sigma^2_{\mathrm{nsr}} = \frac{1}{D-1} \sum_{i=2, i\neq j}^{D} \lambda_i \tag{6.76}$$

As a reminder, λ_j and $U_{j,\mathrm{row}}$ are the largest eigenvalue and the corresponding eigenvector of the inverse $(\bar{S}^{p\neq j})^{-1}$, while λ_i in above sum represents the discarded eigenvalues of the noise subspace complementary to the principle subspace.

6.5.2 Sparsity prior and posterior probability

Biological networks like gene regulatory or signal transduction network are usually sparse [33], meaning that direct links between nodes are rare. In addition to the available training data, this information about the network structure can enhance network inference results. The common procedure to include additional general information, like network sparsity, to the inference setting is achieved by introducing a sparsity prior probability distribution. This prior $p(G)$ on the network structure is

multiplied to the likelihood $p\left(\bar{S}^{p\neq 1},\ldots,\bar{S}^{p\neq D}|G\right)$ to arrive at the posterior probability $p\left(G|\bar{S}^{p\neq 1},\ldots,\bar{S}^{p\neq D}\right)$, which measures the probability of a certain structure G given the data (in the form of the mean sample covariance matrices). According to Bayes' theorem the posterior takes the form

$$p\left(G|\bar{S}^{p\neq 1},\ldots,\bar{S}^{p\neq D}\right)=\frac{p\left(\bar{S}^{p\neq 1},\ldots,\bar{S}^{p\neq D}|G\right)\cdot p(G)}{p\left(\bar{S}^{p\neq 1},\ldots,\bar{S}^{p\neq D}\right)} \tag{6.77}$$

$$\Rightarrow\ p\left(G|\bar{S}^{p\neq 1},\ldots,\bar{S}^{p\neq D}\right)\propto p\left(\bar{S}^{p\neq 1},\ldots,\bar{S}^{p\neq D}|G\right)\cdot p(G)\quad . \tag{6.78}$$

As will be seen later, the unknown normalizing constant embodied by the data probability $p\left(\bar{S}^{p\neq 1},\ldots,\bar{S}^{p\neq D}\right)$ will disappear if applied to the MCMC framework, making eq.(6.78) to the more relevant one. Further, this joint posterior probability from above has a formulation based on the conditional probabilities introduced earlier for the likelihood function. The joint posterior probability takes the form

$$\begin{aligned}
&p\left(G|\bar{S}^{p\neq 1},\ldots,\bar{S}^{p\neq D}\right)\\
&=\prod_{j=1}^{D}p\left(G_{j,\mathrm{row}}|\bar{S}^{p\neq j}\right)\propto\left[\prod_{j=1}^{D}p\left(\bar{S}^{p\neq j}|G_{j,\mathrm{row}}\right)\right]\cdot\left[\prod_{j=1}^{D}p\left(G_{j,\mathrm{row}}\right)\right] &(6.79)\\
&=\prod_{j=1}^{D}p\left(G_{j,\mathrm{row}}|\bar{S}^{p\neq j}\right)\propto\prod_{j=1}^{D}p\left(\bar{S}^{p\neq j}|G_{j,\mathrm{row}}\right)\cdot p\left(G_{j,\mathrm{row}}\right)\quad . &(6.80)
\end{aligned}$$

The expression of eq.(6.80) shows that instead of introducing a prior $p(G)$ for the whole network G it is possible to create a prior $p\left(G_{j,\mathrm{row}}\right)$ which acts only on the incoming link structure $G_{j,\mathrm{row}}$. Therefrom once again, the inference problem of the whole network can be simplified to one of inferring the incoming links of each unperturbed node step by step.

$$p\left(G_{j,\mathrm{row}}|\bar{S}^{p\neq j}\right)\propto p\left(\bar{S}^{p\neq j}|G_{j,\mathrm{row}}\right)\cdot p\left(G_{j,\mathrm{row}}\right) \tag{6.81}$$

In contrast to the likelihood approach the posterior includes additional network information, established by the prior.

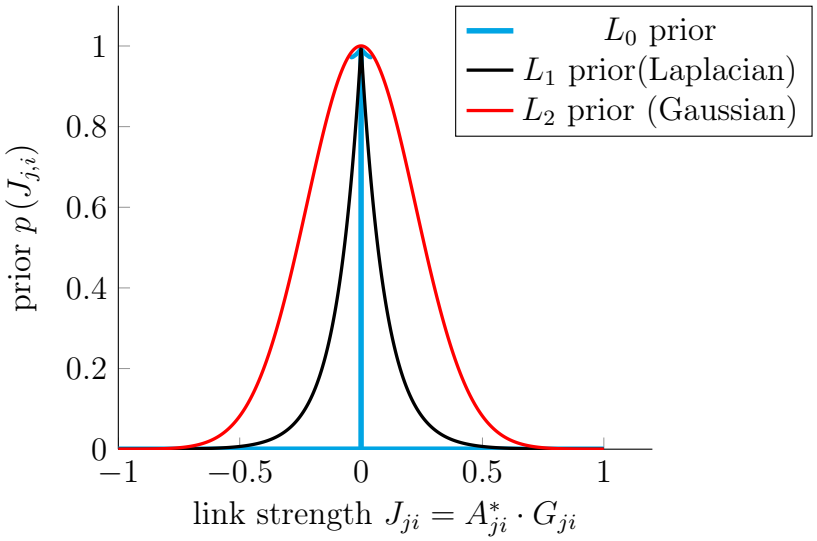


Figure 6-3: **Sparsity prior distributions without normalization for a link J_{ji} from node i to j given the regularization coefficient $\gamma = 10$.** The L_0 norm imposes the strongest sparsity condition upon the link strength J_{ji}, since this function is strongly peaked for $J_{ji} = 0$ and otherwise $J_{ji} \neq 0$ close to to zero, i.e. $\exp(-\gamma)$. The discontinuous jump of the L_0 prior at $J_{ji} = 0$ is visualized by the blue arrow. The other two priors constitute weaker sparsity constraints than the L_0 prior. The L_1 prior follows a Laplace distribution, which is distributed closer around its peak than the Gaussian distribution of the L_2 prior. Notice, that $p(J_{j,\text{row}}) = \prod_{i=1}^{D} p(J_{ji})$ and $p(J_{ji}) \propto \exp\left[-\gamma \cdot \|J_{j,i}\|_l^l\right]$, with l denoting the norm.

Network sparsity priors

There are many priors that can induce network sparsity, nonetheless the main feature they all have in common is a high probability for the absence of any link G_{ji}. As a consequence links must be properly supported by the training data, i.e. possess a high likelihood, to be inferred as existing link. In terms of machine learning these sparsity priors lead to a reduction of model complexity, which is manifested in a reduced number of parameters G_{ji} used to explain the observed data. In the following the L_0 sparsity prior and for comparison only three more priors will be introduced.

In this work the main interest lies in the L_0 sparsity prior based on the L_0 norm regularizer, because it represents the strongest possible sparsity condition. The L_0 prior drives link strengths that are not supported by data to exactly zero resulting in a clear structure without the need of a cutoff, as it is the case for the maximum

likelihood (ML) estimate $\widehat{A}^*$. This kind of cutoff stands for a threshold under which inferred link strengths are considered to be zero and hence to not exist. Before writing down the prior expressions, one should be aware that by considering the Hadamard product between the structure G and the ML estimate A^* in eq.(6.73) a novel link strength matrix J has been established.

$$J := A^* \circ G \quad , \quad J_{j,\text{row}} := A^*_{j,\text{row}} \circ G_{j,\text{row}} \tag{6.82}$$

Notice that the priors act on this new interaction strength J and not exclusively on the structure G - only in the case of the L_0 prior both dependencies are identical. Finally the L_0 prior which acts on the incoming links of node j takes the form

$$\begin{aligned} p\left(G_{j,\text{row}}\right) &= \frac{1}{c} \cdot \exp\left[-\gamma \cdot \left\|G_{j,\text{row}}\right\|_0\right] \\ &\propto \exp\left[-\gamma \cdot \left\|G_{j,\text{row}}\right\|_0\right] \end{aligned} \tag{6.83}$$

where γ is the regularization coefficient or strength parameter, which measures how strong prior information is imposed upon the inference algorithm. The L_0 matrix norm[4] is given by the sum over all incoming structure elements $G_{ji} \in \{0, 1\}$

$$\left\|G_{j,\text{row}}\right\|_0 = \sum_{i=1}^{D} G_{ji} = \sum_{i=1}^{D} \left\|J_{ji}\right\|_0 \tag{6.84}$$

and c is the the normalization constant, so that $\int p\left(J_{j,\text{row}}\right) dJ_{j,\text{row}} \stackrel{!}{=} 1$. The L_0 prior will be compared to the L_1 and L_2 norm priors, which will be formulated next. The

[4] L_0 matrix norm $\|J\|_l = \sum_{j,i=1}^{D} \delta(J_{ji})$, where δ is the Kronecker delta with $\delta(J_{ji}) = 1$ if $J_{ji} \neq 0$ and $\delta(J_{ji}) = 0$ if $J_{ji} = 0$. In this sense, the L_0 matrix norm measure the number of non-zero elements of matrix J.

L_1 prior [41] for the incoming link strengths $J_{i,\text{row}}$ considering expression eq.(6.67) is

$$p\left(G_{j,\text{row}}\right) \equiv p\left(G_{j,\text{row}}|A^*_{j,\text{row}}\right) = \frac{\gamma}{2} \cdot \exp\left[-\gamma \left\|J_{j,\text{row}}\right\|_1\right] \tag{6.85}$$

$$\propto \exp\left[-\gamma \cdot \left\|J_{j,\text{row}}\right\|_1\right]$$

$$p\left(G_{j,\text{row}}\right) \propto \exp\left[-\gamma \cdot \left[\lambda_j - \sigma^2_{\text{nsr}}\right]^{\frac{1}{2}} \cdot \left\|U_{j,\text{row}} \circ G_{j,\text{row}}\right\|_1\right] \quad , \tag{6.86}$$

where the explicit notation of $A^*_{j,\text{row}}$ has been omitted from the prior $p\left(G_{j,\text{row}}\right)$, indicated by " $\equiv$ " in the first equation line. Due to the L_1 norm[5] the non-identifiable rotation parameter $v \in \{0, 1\}$ vanishes. The L_1 prior has the form of a Laplace distribution, which is sharply peaked around a zero link strength(see Figure 6-3). The disadvantage of the L_1 norm is that it expresses a weaker condition for sparsity than the L_0 norm, so that unlikely link strengths are not driven to zero [46, 50, 66]. The same behavior will be observed if the L_2 prior is applied, which poses an even weaker constraint due to a broader distribution around zero. The L_2 prior is a Gaussian distribution and the common norm to prevent overfitting in linear regression problems ([11], p.10). The L_2 sparsity prior takes on the following expression

$$p\left(G_{j,\text{row}}\right) \equiv p\left(G_{j,\text{row}}|A^*_{j,\text{row}}\right) = \left(\frac{\gamma}{\pi}\right)^{\frac{1}{2}} \cdot \exp\left[-\gamma \cdot \left\|J_{j,\text{row}}\right\|_2^2\right] \tag{6.87}$$

$$\propto \exp\left[-\gamma \cdot \left\|J_{j,\text{row}}\right\|_2^2\right]$$

$$p\left(G_{j,\text{row}}\right) \propto \exp\left[-\gamma \cdot \left[\lambda_j - \sigma^2_{\text{nsr}}\right] \cdot \left\|U_{j,\text{row}} \circ G_{j,\text{row}}\right\|_2^2\right] \quad . \tag{6.88}$$

For the use of above priors in the MCMC framework only the proportionality relations eq.(6.83),(6.86) & (6.88) are of interest, since the normalization constants will be reduced with respect to ratios of priors. Figure 6-3 shows a plot of the three priors without any normalization, so that the plotted functions in this figure can not be regarded as probability density functions whose integral is one. In this form the priors can be regarded as weight functions, which reduce model complexity.

[5] Matrix norm $\|J\|_l = \left(\sum_{j,i=1}^{D} |J_{ji}|^l\right)^{1/l}$

6.5.3 Markov chain Monte Carlo sampling over posterior

As the title of this section has already indicated a sample of network structures from the derived posterior probability distribution will be drawn using Markov chain Monte Carlo (MCMC) sampling. Having obtained a sample of network structures representing the posterior sufficiently, it will be possible to determine the sample mean of the posterior distribution, i.e. the average network structure. Finally, the mean value of each network link serves as a score, which measures the inferred probability that this link exists given the perturbation data.

Sampling over the whole posterior distribution leads to a set of possible network structures $\{G^{(t)}\}$, which can be used to score the existence of each link with the relative frequency of its appearance within the sample set. Alternatively, one could determine the maximum posterior (MAP) network structure by an optimization procedure like Simulated Annealing [5, 46]. However, the MAP approach has the disadvantage of yielding only one probable network structure, whereby neglecting alternative structures of similar probability. Especially in the presence of measurement noise not all links are sufficiently supported by data. Therefore, inferring a set of probable structures will lead to a deeper insight into the underling network structure than the inference of a single MAP solution [46, 50].

In the following, the applied MCMC algorithm will be briefly explained with the help of the pseudocode of Algorithm 1. The focus of this section is to explain how MCMC sampling is applied to above derived network inference framework. For a detailed mathematical explanation of MCMC in the machine learning context the reader is referred to [5]. The here applied MCMC simulation has the special feature of sampling over the incoming structure independent of the rest of the network. Remember that the problem of inferring the whole network at once has been simplified to one of inferring the incoming links of each node step by step. Therefore, it is possible to sample over the incoming structure rather than over the whole network, that is to sample over the conditional posterior $p\left(G_{j,\mathrm{row}}|\bar{S}^{(p\neq j)}\right)$ rather than over the joint posterior $p\left(G|\bar{S}^{(p\neq 1)},\ldots,\bar{S}^{(p\neq D)}\right)$.

Algorithm 1 Metropolis algorithm for inferring the incoming links of each node j step by step

1: **for** $j \leftarrow 1$ **to** D **do**
2: **for** $t \leftarrow 1$ **to** $MCMC$ **do**
3: Produce pseudo random numbers:
4: **Sample** $u \leftarrow U_{[0,1]}$ ▷ Sample from uniform distribution of interval [0,1]
5: **Sample** $i \leftarrow U_{[1,2,\ldots,D]|i\neq j}$ ▷ Sample from uniform distribution over all nodes $i \neq j$. $J_{ji} : i \rightarrow j$
6: Monte Carlo step:
7: $Z_{j,\text{row}} = G^{(t)}_{j,\text{row}}$ ▷ $G^{(t)}_{j,\text{row}}$: accepted network structure of MCMC step t
8: $Z_{ji} = 1 - Z_{ji}$ ▷ Flip (remove or add) 1 incoming link → proposal structure $Z_{j,\text{row}}$ of MCMC step t
9: Compute inverse pop. covariance C for use in likelihood function :
10: $C^{-1}_{prop} = \left(A^*_{j,\text{row}} \circ Z_{j,\text{row}}\right)^T \cdot \left(A^*_{j,\text{row}} \circ Z_{j,\text{row}}\right) + \sigma^2_{\text{nsr}} \cdot I_D$ ▷ For proposal structure
11: $C^{-1}_{old} = \left(A^*_{j,\text{row}} \circ G^{(t)}_{j,\text{row}}\right)^T \cdot \left(A^*_{j,\text{row}} \circ G^{(t)}_{j,\text{row}}\right) + \sigma^2_{\text{nsr}} \cdot I_D$ ▷ For last accepted structure
12: Determine acceptance probability $a\left(Z_{j,\text{row}}, G^{(t)}_{j,\text{row}}\right)$:
13: $I\left(C^{-1}_{prop}, C^{-1}_{old}\right) = \dfrac{p\left(Z_{j,\text{row}}|\bar{S}^{p\neq j}\right)}{p\left(G^{(t)}_{j,\text{row}}|\bar{S}^{p\neq j}\right)} = \dfrac{p\left(\bar{S}^{p\neq j}|Z_{j,\text{row}}\right)}{p\left(\bar{S}^{p\neq j}|G^{(t)}_{j,\text{row}}\right)} \cdot \dfrac{p\left(Z_{j,\text{row}}\right)}{p\left(G^{(t)}_{j,\text{row}}\right)}$ ▷ Posterior probability **ratio**
14: $a\left(Z_{j,\text{row}}, G^{(t)}_{j,\text{row}}\right) = \min\left\{1\ ,\ I\left(C^{-1}_{prop}, C^{-1}_{old}\right)\right\}$ ▷ Acceptance probability
15: Check acceptance of proposal structure :
16: **if** $u < a$ **then**
17: $G^{(t+1)}_{j,\text{row}} = Z_{j,\text{row}}$ ▷ Proposal structure is accepted
18: **else**
19: $G^{(t+1)}_{j,\text{row}} = G^{(t)}_{j,\text{row}}$ ▷ Proposal structure is **not** accepted
20: **end if**
21: **end for**
22: **end for**

In Algorithm 1 a MCMC simulation is executed for each node, indicated by the loop over unperturbed node j. The MCMC sampling procedure can be regarded as a random walk in the solution space of incoming structures, whereas each new proposal structure is generated only from the last accepted structure as it is the case in a Markov chain. The incoming proposal structure $Z_{j,\text{row}}$ is obtained by adding or removing (flipping) one incoming link to or from the last accepted incoming structure $G^{(t)}_{j,\text{row}}$, whereas t denotes the Monte Carlo step (Algorithm 1 line 8). The flipped incoming link i is drawn from a uniform[6] distribution, so that each possible Monte Carlo step is equally likely. Therefore, the random walk in the space of proposal structures can be regarded as a *symmetric random walk* justifying the use of the

[6]As the diagonal elements of network structure, i.e. the degradation rate, are assumed to always exist, they are fixed to one $G^{(t)}_{jj} \equiv 1$.

Metropolis algorithm [5]. In the next step the conditional posterior ratio of the proposed structure $p(Z_{j,\mathrm{row}}|\bar{S}^{p\neq j})$ relative to the last accepted structure $p(G^{(t)}_{j,\mathrm{row}}|\bar{S}^{p\neq j})$ is computed, which is equal to the associated likelihood ratio times the prior ratio.

$$\frac{p\left(Z_{j,\mathrm{row}}|\bar{S}^{p\neq j}\right)}{p\left(G^{(t)}_{j,\mathrm{row}}|\bar{S}^{p\neq j}\right)} = \frac{p\left(\bar{S}^{p\neq j}|Z_{j,\mathrm{row}}\right)}{p\left(\bar{S}^{p\neq j}|G^{(t)}_{j,\mathrm{row}}\right)} \cdot \frac{p\left(Z_{j,\mathrm{row}}\right)}{p\left(G^{(t)}_{j,\mathrm{row}}\right)} \tag{6.89}$$

As mentioned above all constant terms, i.e. the data probability $p\left(\bar{S}^{p\neq j}\right)$ and normalization constant of the different priors, vanish within the ratio term. This has the advantage that the posterior probability can even be utilized in the case of undetermined quantities like the data probability. The likelihood ratio takes the form

$$\begin{aligned}&\frac{p\left(\bar{S}^{p\neq j}|Z_{j,\mathrm{row}}\right)}{p\left(\bar{S}^{p\neq j}|G^{(t)}_{j,\mathrm{row}}\right)}\\ &= \left[\frac{\det\left(C^{-1}_{\mathrm{prop}}\right)}{\det\left(C^{-1}_{\mathrm{old}}\right)}\right]^{\frac{NQ}{2}} \cdot \exp\left[\frac{NQ}{2}\cdot\left(-\operatorname{Tr}\left[C^{-1}_{\mathrm{prop}}\cdot\bar{S}^{p\neq j}\right]+\operatorname{Tr}\left[C^{-1}_{\mathrm{old}}\cdot\bar{S}^{p\neq j}\right]\right)\right]\end{aligned} \tag{6.90}$$

where C^{-2}_{old} denotes the inverse population covariance matrix of the last accepted structure $G^{(t)}_{j,\mathrm{row}}$ and ${C^{-1}_{\mathrm{prop}}}^{-1}$ the one of the proposal structure $Z_{j,\mathrm{row}}$. The inverse covariance matrix is given by eq.(6.75). To have a complete picture the prior ratios for the three different prior can be found below. The prior ratio for the L_0 prior can be written as

$$\frac{p\left(Z_{j,\mathrm{row}}\right)}{p\left(G^{(t)}_{j,\mathrm{row}}\right)} = \exp\left[\gamma\left(-\left\|Z_{j,\mathrm{row}}\right\|_0+\left\|G^{(t)}_{j,\mathrm{row}}\right\|_0\right)\right] \quad . \tag{6.91}$$

The prior ratio for the L_1 prior has the form

$$\begin{aligned}&\frac{p\left(Z_{j,\mathrm{row}}\right)}{p\left(G^{(t)}_{j,\mathrm{row}}\right)}\\ &= \exp\left[\gamma\cdot\left[\lambda_j-\sigma^2_{\mathrm{nsr}}\right]^{\frac{1}{2}}\cdot\left(-\left\|U_{j,\mathrm{row}}\circ Z_{j,\mathrm{row}}\right\|_1+\left\|U_{j,\mathrm{row}}\circ G^{(t)}_{j,\mathrm{row}}\right\|_1\right)\right] \quad .\end{aligned} \tag{6.92}$$

Finally, the prior ratio for the L_2 prior is given by

$$\begin{aligned}
&\frac{p\left(Z_{j,\mathrm{row}}\right)}{p\left(G^{(t)}_{j,\mathrm{row}}\right)} \\
&= \exp\left[\gamma \cdot \left[\lambda_j - \sigma^2_{\mathrm{nsr}}\right] \cdot \left(-\left\|U_{j,\mathrm{row}} \circ Z_{j,\mathrm{row}}\right\|_2^2 + \left\|U_{j,\mathrm{row}} \circ G^{(t)}_{j,\mathrm{row}}\right\|_2^2\right)\right] \ . \qquad (6.93)
\end{aligned}$$

As can be seen from above expressions neither of the prior ratios nor the likelihood ratio depend on their normalization constants. In the next step of the Metropolis algorithm (Algorithm 1 line 14) the acceptance probability a of the proposal structure will be set to one if the proposal posterior is larger than the posterior of the last accepted structure. In this case the proposal structure $Z_{j,\mathrm{row}}$ is simply accepted (Algorithm 1 line 17). If the opposite case $p(Z_{j,\mathrm{row}}|\bar{S}^{p\neq j}) < p(G^{(t)}_{j,\mathrm{row}}|\bar{S}^{p\neq j})$ the proposal structure will be accepted with a probability equal to the posterior ratio.

Concluding, the here presented network inference method produces a set of probable network structures by sampling over the conditional posterior probability of network structures given a complete perturbation data set. The problem of sampling over the whole space of networks structures is simplified to the one of sampling over the incoming link structures of each node separately. In other words there are D independent MCMC simulations for a network of size D, whereas each MCMC simulation scans over all possible incoming link structures.

Chapter 7

Performance Assessment

The goal of this section is to assess the performance of the network inference algorithm derived in the previous sections with the help of synthetic data generated from a signal transduction network.

7.1 Synthetic data

The here used signal transduction network[1] is a combination of the four epidermal growth factor receptors ErbB1 (EGFR), ErbB2, ErbB3, ErbB4 [67] with the MAP kinase signaling cascade ([64] p.175). Additionally, four ligands that can bind to the ErbB receptor family are included as nodes to the signaling network in accordance with [48,67], which yields a network of 14 nodes and 16 links as can be seen in Fig.7-1. For the sake of brevity this network will be referred to as *EGFR network* in the following. The EGFR network of Figure 7-1 can be transformed into a structure matrix G and a link strength matrix A representing the "gold-standard", which the inferred network structure will be compared to below.

$$G_{ji} = \begin{cases} 1 & \text{for existing interactions} \\ 1 & \text{for } j = i \\ 0 & \text{no interaction} \end{cases} \tag{7.1}$$

[1] This specific signal transduction network was used by Mukherjee et al. (2008) [48] to generate synthetic data and assess their own network inference algorithm.

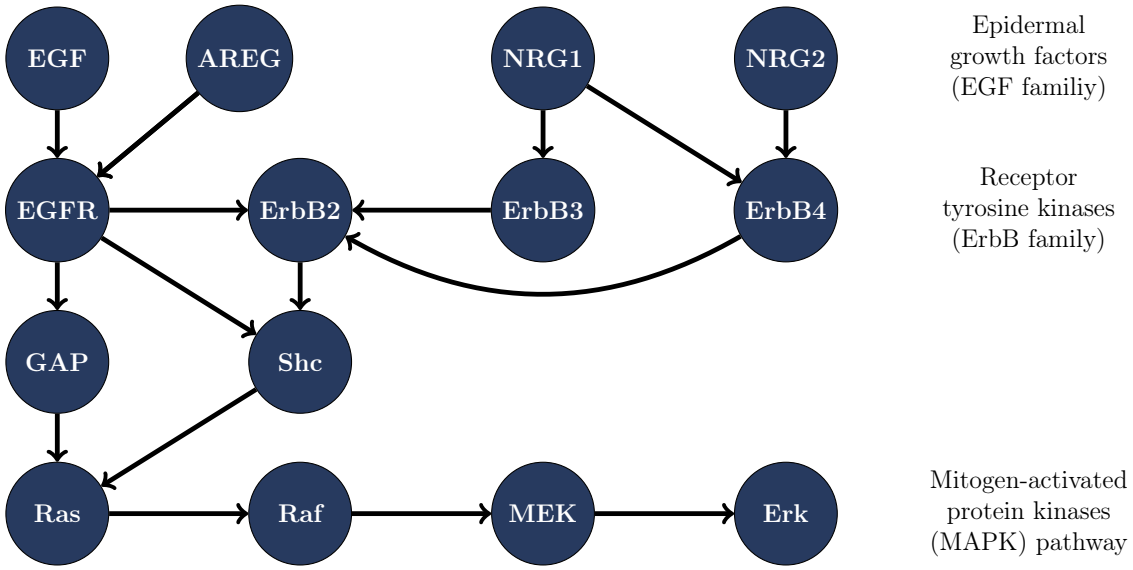


Figure 7-1: **The EGFR network used to generate synthetic data.** The data generating network structure is taken from [48] and consists of growth factors from the epidermal growth factor family, which activate receptor tyrosine kinases (RTKs) from the ErbB family (EGFR, ErbB2, ErbB3, ErbB4). The receptor proteins are bound within the cell membrane, so that a growth factor can bind to the receptors extracellularly in order to mediate the signal to intracellular phospho-proteins [67]. Among others, the activation of the RTKs initiates the phosphorylation cascade of the mitogen-activated protein kinases (MAPK) pathway, which follows the order Ras, Raf, MEK, and Erk. The MAPK pathway initiates cell proliferation as cellular response to the growth signal [64]. The blue color of all nodes indicates that all nodes are perturbed in single experiments. Further, the protein abundance (node activity) of the whole network is measured (observed). Arrows in the here presented EGFR network represent an activation process, e.g. activation of the kinase activity by means of upstream phosphorylation.

The link strengths A_{ji} of the existing links G_{ji} will be assumed to be

$$A_{ji} = \begin{cases} 1 & \text{for existing interactions} \\ -1.5 & \text{for } j = i \text{ i.e. degradation rate} \\ 0 & \text{no interaction} \end{cases} . \tag{7.2}$$

Note that these values do not represent the real biological link strengths, i.e. reaction rates, so that this artificial network serves only as a proof of principle. However, the objective of the here presented network inference algorithm is to infer the correct network structure, while inferring the reaction parameters, i.e. the link strengths, is

not part of it. Hence, using real reaction rates for data generation is not imperative. Further, the link strength matrix is chosen to be negative definite, which ensures that node activities, e.g. phospho-proteins' abundance, stay positive and do not diverge with respect to the stochastic process used to generate the synthetic data. To produce synthetic data the interplay of the molecular components in the EGFR network will be assumed to follow the linear stochastic process introduced by eq.(6.18). As reminder this process took the mathematical form

$$\vec{x}_n^{(p)} = -A^{-1} \cdot B^{(p)} \cdot \vec{a}_n \; + \; \sigma \cdot \vec{\phi}_n \quad , \tag{7.3}$$

where n denotes the replicate experiment and p the perturbation experiment and σ is assumed to be all types of measurement noise (not only technical noise). The mean node activity $\vec{\mu}$ is set to zero, since its value is irrelevant for the inference algorithm and therefore can take any arbitrary value (see section 5.2.2 and 6.3). Further, $\vec{a}_n$ and $\vec{\phi}_n$ are drawn from the standard normal distribution. To obtain a complete data set $X := \{\vec{x}_n^{(p)}\}_{\forall n,p}$ of single perturbation experiments ($P = D$), the perturbation standard deviation matrix $B^{(p)}$ is chosen to be

$$B^{(p)} \equiv \text{diag}\left(0, \ldots, 0, B_{pp}^{(p)} = \sigma_u, 0, \ldots, 0\right) \tag{7.4}$$

for each perturbation experiment, i.e. only one node at a time is perturbed. The standard deviation of measurement noise σ and of perturbations σ_u are chosen relative to each other, so that the noise-to-signal ratio nsr $= \sigma/\sigma_u$ is the relevant quantity. This enables one to generate complete data sets for different nsr levels, making it possible to assess the network inference algorithm under the influence of different measurement noise levels with respect to the signal σ_u.

As as special case one complete data set with a noise-to-signal ratio nsr $= 0.15$ and a replicate number $N = 4$ is chosen, on which the here presented network inference algorithm is assessed. A replicate number of only a few experiments is the typical size that one encounters for real state of the art perturbation experiments of signal transduction networks [25]. Given the $N = 4$ and the EGFR network, the net-

work inference algorithm reconstructs the network structure perfectly for nsr < 0.1. Above nsr > 0.2 the number of false positive inferred links approaches the number of true positive links, which is equivalent to a 50% chance of inferring a true positive link. Therefore, the choice of nsr $= 0.15$ lies exactly in the interesting regime nsr $\in [0.1, 0.2]$ in which the choice of prior knowledge and MCMC sampling can boost the performance compared to a network inference based on the maximum likelihood (ML) estimate of the relative link strength. The reader should be aware, that this nsr regime depends on the number of replicates and the examined network, i.e. size and structure. Obviously, in the case of larger replicate numbers, the algorithm will be able to cope with higher measurement noise.

7.2 MCMC convergence and prior sensitivity

Before being able to infer the structure with the algorithm, one has to adjust the parameters of the MCMC sampling and the prior strength parameter γ. The MCMC simulation is governed by the total number of Monte Carlo (MC) steps, denoted by $MCMC$, the burn-in interval, and the subsampling interval [48]. The burn-in interval is the initial period of MC steps until the MCMC simulation converges to its stationary distribution [22]. It is discarded so that the MCMC sampling does not depend on the initial condition, namely the initial network structure. The subsampling interval must be set large enough so that the sample points can be regarded as statistically independent, i.e. one sampled structure does not depend on the one before. Finally, the total number of Monte Carlo steps $MCMC$ ensures that there is a large enough sample set with respect to the discarded burn-in and the subsampling interval, which can represent the stationary distribution, namely the posterior over network structures, sufficiently.

The here utilized way to select these parameters is by plotting the MCMC convergence curve [40], which is the model error of structure $G^{(t)}$ given the complete data set in dependence of the Monte Carlo step t as shown in Figure 7-2. The error function is defined as the negative logarithm of the joint posterior distribution

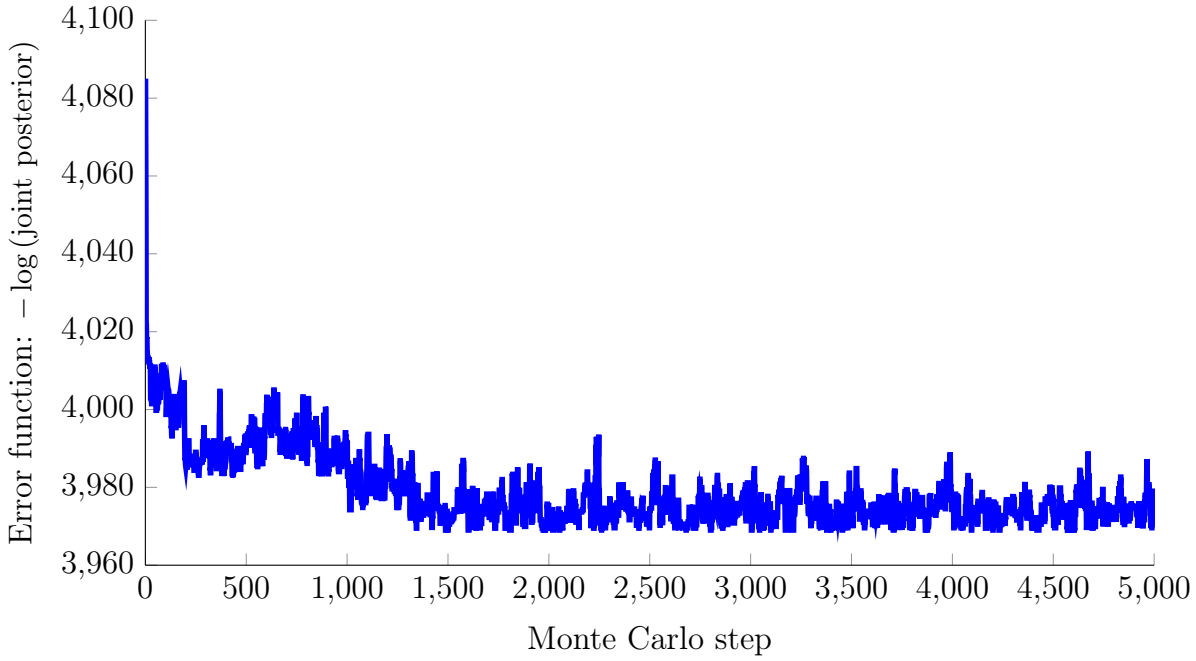


Figure 7-2: **Error function of the whole network structure G without normalization constant in dependency of MCMC steps shows the convergence of the MCMC samples to the stationary distribution.** The convergence curve is shown for the MCMC algorithm with an L_0 prior, whereas the other prior have a similar convergence behavior. After 2000 MC steps the MCMC simulation converges to its stationary distribution independent of the initial network structure. Therefore, a burn-in of 5000 and total length of $MCMC = 50000$ is sufficient to obtain a representative sample of the posterior distribution over network structures. The variance of the the error function around the stationary mean depends on the size of prior strength γ and of course as well on the prior type. In detail the error variance depends on the posterior variance, so that a larger γ generates a larger error variance.

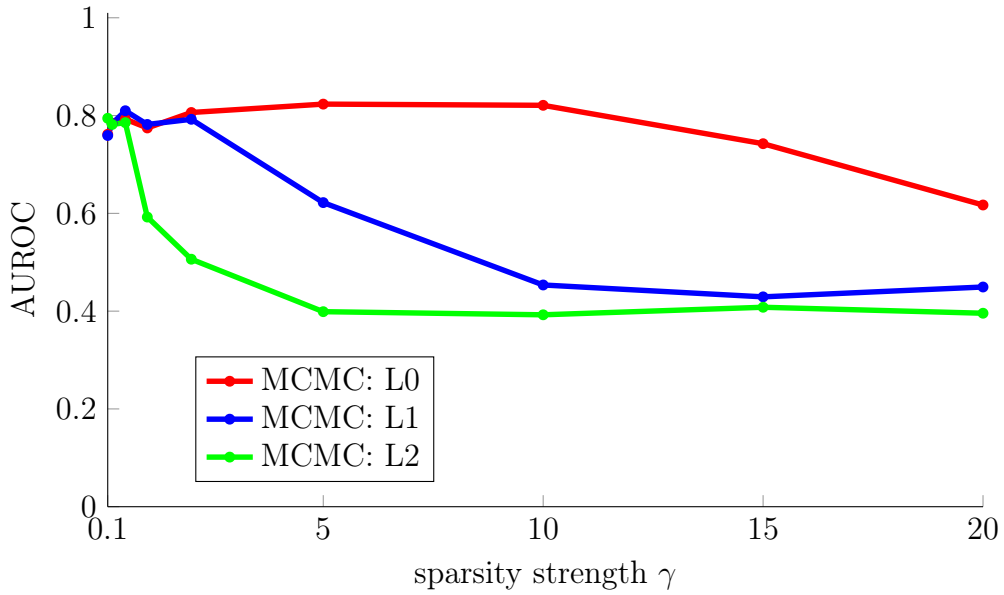
$p(G|\{X^{(p)}\})$, whereas the L_0 prior is chosen with the prior strength γ that maximize the area under the *receiver operating characteristic* (AUROC) curve as will be shown below. To remind the reader, there are $D = 14$ different MCMC simulations, one for the incoming links of each node represented by the conditional posterior distribution $p(G_{\text{j,row}}|X^{(p\neq j)})$ as derived in chapter 6. The MCMC convergence Figure 7-2 views the negative logarithm of the joint posterior, which is the product of the conditional posteriors, i.e. $p(G|\{X^{(p)}\}) = \prod_{j=1}^{D} p(G_{\text{j,row}}|X^{(p\neq j)})$. As can be observed from Figure 7-2 the MCMC simulation converges after about 2000 MC steps. This result holds as well for randomly chosen initial structures, where convergence occurs after a few thousand MC steps. To be on the save side a burn-in interval of 5000 MC steps is chosen

and the subsampling interval is set to 100 MC steps according to [40]. With a total number of $MCMC = 50,000$ Monte Carlo steps a subsample set of 450 structures is drawn from the joint posterior distribution.

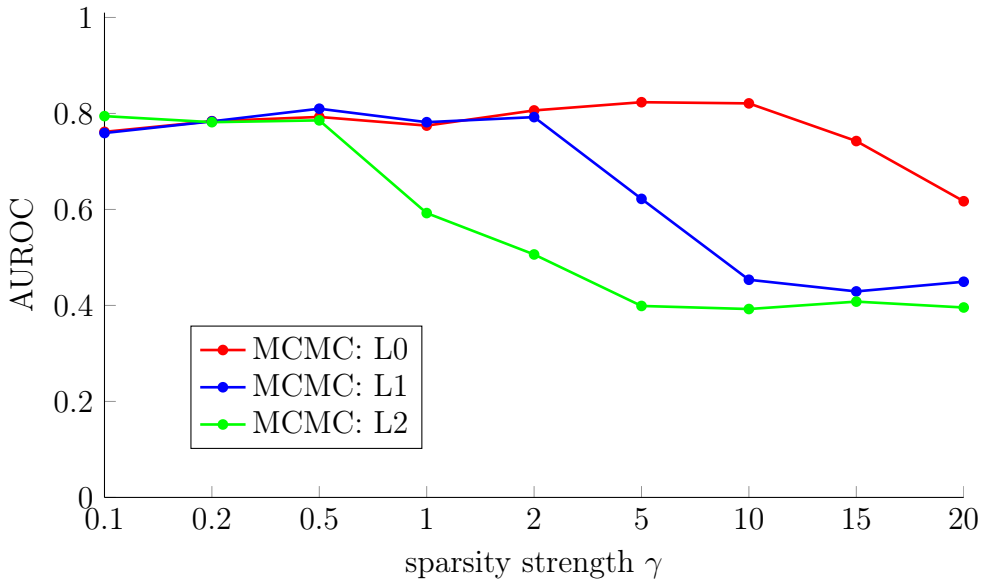
Having obtained the MCMC setting one can turn to determining the prior strength parameter γ. As mentioned γ is set to the value that maximizes the area under the receiver operating characteristic (AUROC) curve [48]. The ROC curve plots the true positive rate against the false positive rate for different thresholds [41]. This threshold serves as a decision boundary, which is compared to the inferred confidence score of each link, so that confidence scores above the threshold are regarded as inferred link whereas links with a score below the threshold are regarded as no link. The confidence score[2] for a link inferred by the here presented algorithm is the relative frequency of link appearance within the subsample set of 450 network structures gained from the MCMC simulation. The area under the ROC curve can be used to summarize the ROC performance by a single scalar, namely the AUROC value, which is computed for different γ value to generate Figure 7-3. The best possible ROC performance corresponds to an AUROC value of one, while the worst case is stated by a zero AUROC value. But usually a performance that goes below AUROC $= 0.5$ is already not reliable. Therefore, setting the γ value to the one that maximizes the AUROC value is a reasonable way to determine the unknown prior strength γ - among others [41].

From Figure 7-3(b) one obtains the maximum AUROC values for the here presented network inference algorithm depending on the chosen prior. The maximum AUROC for the L_0 prior is achieved by setting $\gamma_{L_0} = 5$, while $\gamma_{L_1} = 0.5$ and $\gamma_{L_2} = 0.1$ lead to a maximum AUROC for the L_1 and L_2 prior, respectively. Another interesting feature of the AUROC curves is gained by observing the sensitivity of AUROC performance with respect to a wide range of γ choices. As it is shown in Figure 7-3(a), the performance of the algorithm with a L_0 prior is less sensitive towards the choice of γ than its performance with a L_2 prior, whereas the $L1$ prior is situated somewhere in-between as Figure 7-3(a) indicates. This robustness in terms of parameter sensi-

[2] The confidence score will be explained in more detail in the next subsection.



(a) The sensitivity of inference results represented by AUROC with respect to the prior strength parameter γ.The AUROC performance of L_0 prior is the most robust one with respect to the choice of γ, while the L_2 prior is very sensitive.



(b) Determine $\gamma_{\max}$, so that the AUROC value is maximal. Maximum AUROC values are generated for $\gamma_{L_0} = 5$, while $\gamma_{L_1} = 0.5$ and $\gamma_{L_2} = 0.1$. Notice, that scaling of γ axis has been changed to be able to easily compare the AUROC values for different γ values.

Figure 7-3: **Area under the receiver operating characteristic (AUROC) curve in dependency of the sparsity strength parameter γ.** Maximum AUROC values for the MCMC network inference algorithm depending on the chosen prior as well as the sensitivity of the AUROC results regarding the choice of the parameter γ.

tivity is another[3] advantage which makes the L_0 prior more preferential for network inference than the other priors.

7.3 Network inference

In this last section the different subtypes of the here presented novel network inference algorithm will be compared among each other by means of ROC performance for the previously introduced synthetic data set. In detail, the algorithm based on the ML estimate of the relative link strength (*NetInf_ML*) from section 6.4 will be compared to the algorithm based on MCMC sampling over the posterior of network structures (*NetInf_MCMC*) from section 6.5 . Furthermore, the effect of the different priors on the performance of *NetInf_MCMC* will be analyzed. For the sake of convenience each of this algorithm subtypes is given an abbreviation, namely *NetInf_ML* for the maximum likelihood algorithm and *NetInf_MCMC* for the MCMC algorithm. The different priors of latter algorithm are denoted by *NetInf_MCMC:L0*, *NetInf_MCMC:L1*, and *NetInf_MCMC:L2*.

Before starting the comparison, the definition and interpretation of ROC curves will be briefly summarized and the used link confidence scores for each of the algorithms will be explained. As mentioned above the receiver operating characteristic (ROC) curve plots the true positive rate (TPR) against the false positive rate (FPR) for different thresholds [39, 41].

$$TPR = \frac{\#\text{ true positives}}{\#\text{ true positives} + \#\text{ false negatives}} \tag{7.5}$$

$$FPR = \frac{\#\text{ false positives}}{\#\text{ false positives} + \#\text{ true negatives}} \tag{7.6}$$

The threshold represents a decision boundary that is used to discriminate the confidence scores of each inferred link as link or no link in a boolean manner. The

[3]The previously mentioned advantage of using a L_0 prior for network sparsity is its ability to drive unlikely links, that are not supported by data, exactly to zero in contrast to the other priors.

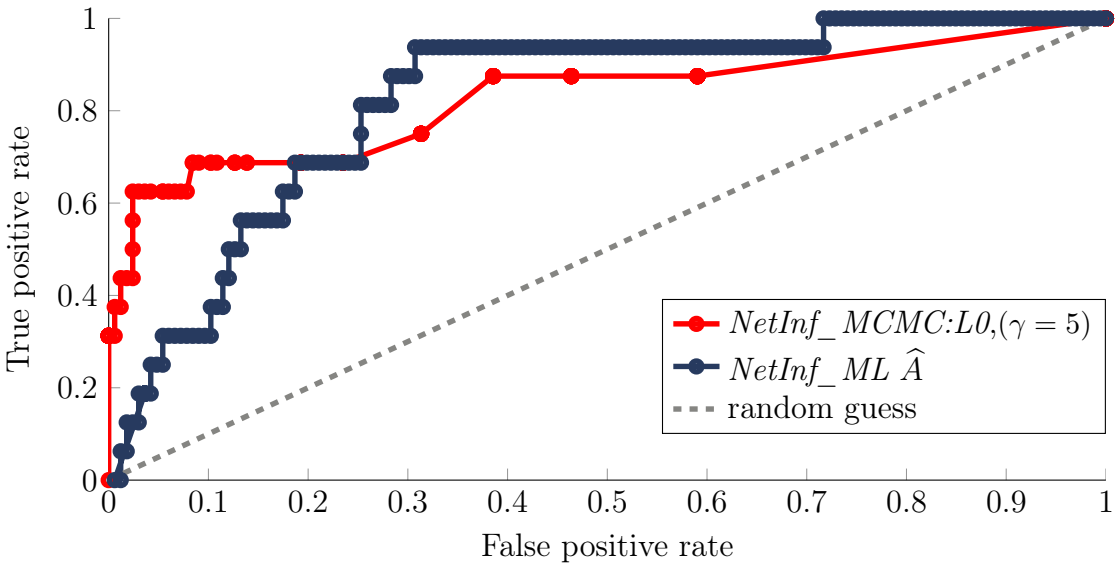
confidence score can be interpreted as a measure that evaluates how confident the network inference algorithm is about the inferred structure. In case of *NetInf_ ML* the confidence score of each link is simply the absolute value of the inferred relative link strength $\text{score}_{\text{ML}}(\widehat{A}_{ji}) = |\widehat{A}_{ji}|$. Hence in this case, a large relative link strength means more confidence and a small link strength less, whereas the absolute value neglects whether the link stands for an activating or inhibitory process. On the the other hand, the confidence score in the case of *NetInf_ MCMC* is the relative frequency with which a link appears in the MCMC sample set, i.e. among the $L = 450$ network structures $G^{(l)}$.

$$\text{score}_{\text{MCMC}}(G_{ji}) = \frac{1}{L} \cdot \sum_{l=1}^{L} G_{ji}^{(l)} \quad , \tag{7.7}$$

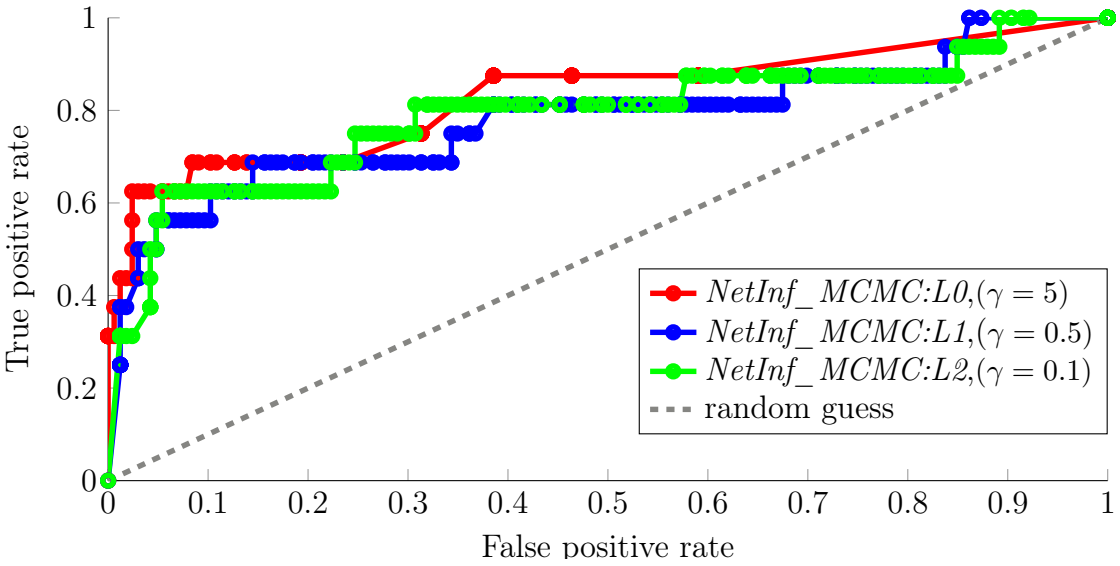
where L denotes the total number of sample structures gained from the MCMC sampling and $G^{(l)}$ denotes the l^{th} sample structure. Hence in this case, a link that appears in most of the sampled structures is weighted with a higher score than a link that appears in a few sample structures, only.

In Figure 7-4(a) one can see the ROC curves of *NetInf_ MCMC:L0* compared to *NetInf_ ML* for the synthetic data of the EGFR network at a noise-to-signal ratio $nsr = 15\%$ and only $N = 4$ replicates. First of all, one can recognize that both algorithms perform significantly better than an inference algorithm based of random guess qualifier, indicated by the diagonal dashed line. Furthermore, *NetInf_ MCMC:L0* performs with significantly higher recall than *NetInf_ ML*. In detail, *NetInf_ MCMC:L0* infers 70% of the total network structure with only a 10% false positive rate, while for 60% of the network the false positive rate is even smaller, namely at $FPR = 2\%$. In contrast, *NetInf_ ML* infers 70% of the total network structure with a 20%, i.e. twice as much as *NetInf_ MCMC:L0*. For 60% of the network *NetInf_ ML* takes a false positive rate of about 18%, which is six times higher than the one from *NetInf_ MCMC:L0*.

In Figure 7-4(b) the performance of *NetInf_ MCMC* with different priors is compared. The general impression is that *NetInf_ MCMC:L1* and *NetInf_ MCMC:L2*



(a) Comparison in the ROC performance of *NetInf_ MCMC:L0* compared to *NetInf_ ML*. The *NetInf_ MCMC:L0* algorithm performs with significantly higher recall than *NetInf_ ML*.



(b) Comparison in the ROC performance of *NetInf_ MCMC* with the different priors (L_0, L_1, L_2). The performance difference among the priors is much smaller than in the case of *NetInf_ MCMC:L0* and *NetInf_ ML*.

Figure 7-4: **ROC curves for the *NetInf_ MCMC NetInf_ ML* network inference algorithms:** for the synthetic data of the EGFR network at a noise-to-signal ratio $nsr = 15\%$ and only $N = 4$ replicates.

perform very similar to *NetInf_MCMC:L0* with respect to their ROC curves. In other words, the performance difference among the priors is much smaller than in the case of *NetInf_MCMC:L0* and *NetInf_ML*. In detail, *NetInf_MCMC:L0* has a higher recall till 70% of the total inferred network, which means that at 70% TPR *NetInf_MCMC:L1* has FPR of about 15% and *NetInf_MCMC:L2* an even worse FPR of about 23%. At a TPR of 60% *NetInf_MCMC:L1* and *NetInf_MCMC:L2* have a FPR below 10% and hence perform more similar to *NetInf_MCMC:L0*.

Concluding, the here presented network inference framework, derived in section 6.4 & 6.5, is capable of handling noisy data with only a few replicates. Adding prior knowledge about network sparsity in form of a L_0, L_1 or L_2 prior does enhance the performance, so that the rate of inferring true links (TPR) is much higher, up to 30 times, than inferring not existing links (FPR). Although, the *NetInf_MCMC* creates similar ROC curve results for all three priors the L_0 prior is superior. This superiority comes from the fact that the L_0 prior drives unlikely links to or very close to zero, which practically makes the decision to classify a link as real more viable. The ROC representation does not clearly reflect this feature, since it shows results for a wide range of confidence thresholds. If the the real network is unknown, i.e. there is no acceptable gold standard as it is the case for real biological data, the confidence scores of *NetInf_MCMC:L0* allow to choose a proper confidence threshold more easily. In contrast, by using the L_1 and L_2 priors for real biological data, the problem of setting up the proper confidence threshold arises, since their scores are not well distinguishable.

Chapter 8

Conclusion

In this study a novel machine learning approach for the inference of biological network structure from controlled perturbation data has been introduced, which reduces the dimensionality of the network inference problem of the whole network to one of inferring the incoming links of each node separately. As a consequence, the approach based on the probabilistic principle component analysis (PPCA) of partial correlations, distinguishes direct causal links between observed nodes from pure correlation associations and overcomes even Gaussian measurement noise despite of only a few replicate experiments.

By rearranging a complete data set into D different reduced data sets, the network inference problem of the whole network (of size D) is transformed to inferring incoming links of each node separately - leading to the whole network structure step by step. Each reduced data set contains all available single perturbation experiments except for one, so that all nodes but one are perturbed. The main idea (which leads to a reduced dimensionality), is that the relevant partial correlations in such a reduced data set are the ones from the perturbed nodes to the single unperturbed node. Existing partial correlations between other nodes are "destroyed" by the very act of controlled perturbations. Further, the relevant partial correlations can be unambiguously regarded as directed links due to the simple fact that the perturbation signal flows from a perturbed node to the unperturbed one. Hence, the relevant information contained in each reduced data set is the information about the incoming links of the

unperturbed node.

To extract the relevant information from the reduced data set a maximum likelihood (ML) approach was chosen which is based on the notion of probabilistic principle component analysis (PPCA) [58]. In contrast to regular PPCA, this approach infers the principle components of the inverse covariance matrix (partial correlations), which are the incoming links of the unperturbed node. Simply spoken, the likelihood function in this approach compares the modeled partial correlation, represented by the inverse population covariance matrix C^{-1}, with the partial correlations S^{-1} of the reduced data set. By modeling the effect of perturbations upon each node by a linear Gaussian stochastic process, a relation between the network structure, given in form of an interaction matrix, and the inverse covariance matrix C^{-1} was found. As a consequence the likelihood function can be maximized with respect to the interaction matrix. Further, it was shown that the notion of a principle subspace of partial correlations (inverse covariance matrix) can be identified with the inference of the incoming links of unperturbed nodes in the reduced data setting. This novel idea of PPCA of partial correlations creates an expression for the modeled C^{-1}, which depends only on the incoming network structure of the unperturbed node, thereby explaining the relevant information of the reduced data set as the principle subspace of partial correlations.

To further improve inference results, different prior distributions that induce network sparsity were included to the likelihood framework, so that a posterior distribution over incoming network structures was established. By means of Markov chain Monte Carlo (MCMC) sampling the whole solution space of the posterior can be examined, resulting in a sample set of probable network structures. Rather than just obtaining one possible structure as in the ML approach, the MCMC approach has the advantage of offering a whole set of possible structures. Lastly, a very important consequence of the dimensionality reduction for the MCMC algorithm is that it only samples over the incoming network structure of each unperturbed node. This feature enables faster MCMC convergence, which makes it possible to apply the MCMC technique to larger networks. This is the advantage of the here presented MCMC approach

compared to methods that sample over the whole network structure like [46, 50].

Comparing the ROC performance of MCMC and ML approach for synthetic data of a EGFR signal transduction network, confirms the improved performance of MCMC network inference algorithm. For a recall of 60% of the network the false positive rate (FPR) of the ML approach lies up to nine times above the FPR of the MCMC approach, so that latter approach leads to a more precise inference result. This result is obtained for a synthetic data set with a noise-to-signal ratio (nsr) of 15% and only 4 replicate experiments, representing the typical available experimental data in the presence of high noise level [72]. In general, both approaches infer the exact correct network in the presence of low or medium noise-to-signal ratios of up to 10% given only a few replicates. Additionally, the ROC performance of the MCMC approach with respect to the L_0, L_1, and L_2 norm sparsity priors revealed only minor differences. However, the inference results by means of L_1, and L_2 norm are very sensitive regarding the choice of the sparsity strength parameter γ. In contrast, the L_0 norm prior produces robust results for a wide range of γ values, thereby simplifying the parameter choice in practical applications. Furthermore, the fact that the L_0 norm produces a clear threshold for distinguishing inferred link from "noise", was reproduced by the MCMC approach.

Although, the here presented theory is derived for a complete data set, comprising the perturbation of all network nodes, it provides the framework to be extended to a incomplete data sets. Further, the performance assessment could be seen only as a proof of principle as long as it has not been applied to real experimental data. However, experimental data has the flaw of not knowing the real underlying network structure, making it difficult to assess the algorithm with respect to a gold standard. Nonetheless, in future work it would be worthwhile to apply the method to different types of networks with synthetic and experimental data.

To summarize, in this thesis a novel machine learning technique in the field of network inference has been developed, which overcomes Gaussian measurement noise despite of only a few replicates. The main achievement is established by the here introduced theory of *probabilistic component analysis of partial correlations*, which

leads to a dimensionality reduction of the network inference problem. In general, the here presented method creates insight in the exact causal molecular interplay of gene regulatory or signal transduction network, which is by far superior to the pure knowledge about their correlations. Knowledge about these interplays builds the groundwork for predictive models, which enable one to find new therapeutic targets in diseased cells or help to reprogram organisms to express a desired phenotype in biotech applications.

Bibliography

[1] Roi Adadi, Benjamin Volkmer, Ron Milo, Matthias Heinemann, and Tomer Shlomi. Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. *PLoS Comput Biol*, 8(7):e1002575, 2012.

[2] Bruce Alberts. *Molecular biology of the cell*. Garland Science, New York, 5th edition, 2008.

[3] Dominic J Allocco, Isaac S Kohane, and Atul J Butte. Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinformatics*, 5:18, Feb 2004.

[4] Uri Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman and Hall/CRC, 1 edition, 2006.

[5] Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to mcmc for machine learning. *Machine learning*, 50(1-2):5–43, 2003.

[6] Roi Avraham and Yosef Yarden. Feedback regulation of egfr signalling: decision making by early and delayed loops. *Nat Rev Mol Cell Biol*, 12(2):104–17, Feb 2011.

[7] Arren Bar-Even, Johan Paulsson, Narendra Maheshri, Miri Carmi, Erin O'Shea, Yitzhak Pilpel, and Naama Barkai. Noise in protein expression scales with natural protein abundance. *Nat Genet*, 38(6):636–43, Jun 2006.

[8] Albert-László Barabási and Zoltán N Oltvai. Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 5(2):101–13, Feb 2004.

[9] Baruch Barzel and Albert-László Barabási. Network link prediction by global silencing of indirect correlations. *Nat Biotechnol*, 31(8):720–5, Aug 2013.

[10] Q K Beg, A Vazquez, J Ernst, M A de Menezes, Z Bar-Joseph, A-L Barabási, and Z N Oltvai. Intracellular crowding defines the mode and sequence of substrate uptake by escherichia coli and constrains its metabolic activity. *Proc Natl Acad Sci U S A*, 104(31):12663–8, Jul 2007.

[11] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 1st edition, 2006.

[12] Christopher Blum, Nadia Heramvand, Armin Sadat Khonsari, and Markus Kollmann. Inferability of transcriptional networks from large scale gene deletion studies. *(submitted)*, 2016.

[13] Michael Boutros and Julie Ahringer. The art and design of genetic screens: Rna interference. *Nat Rev Genet*, 9(7):554–66, Jul 2008.

[14] Milly Casey-Campbell and Martin L Martens. Sticking it all together: A critical assessment of the group cohesion–performance literature. *International Journal of Management Reviews*, 11(2):223–246, 2009.

[15] Christopher Chatfield. *Statistics for technology*. Chapman and Hall, 3rd edition, 1983.

[16] Erez Dekel and Uri Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050):588–92, Jul 2005.

[17] Josef Deutscher. The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol*, 11(2):87–93, Apr 2008.

[18] Josef Deutscher, Christof Francke, and Pieter W Postma. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev*, 70(4):939–1031, Dec 2006.

[19] J S Edwards, R U Ibarra, and B O Palsson. In silico predictions of escherichia coli metabolic capabilities are consistent with experimental data. *Nat Biotechnol*, 19(2):125–30, Feb 2001.

[20] Soheil Feizi, Daniel Marbach, Muriel Médard, and Manolis Kellis. Network deconvolution as a general method to distinguish direct dependencies in networks. *Nat Biotechnol*, 31(8):726–33, Aug 2013.

[21] Timothy S Gardner, Diego di Bernardo, David Lorenz, and James J Collins. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science*, 301(5629):102–5, Jul 2003.

[22] Charles J Geyer. Practical markov chain monte carlo. *Statistical Science*, pages 473–483, 1992.

[23] Boris Görke and Jörg Stülke. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol*, 6(8):613–24, Aug 2008.

[24] Boris Görke and Jörg Stülke. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol*, 6(8):613–24, Aug 2008.

[25] Steven M Hill, Laura M Heiser, Thomas Cokelaer, Michael Unger, Nicole K Nesser, Daniel E Carlin, Yang Zhang, Artem Sokolov, Evan O Paull, Chris K Wong, Kiley Graim, Adrian Bivol, Haizhou Wang, Fan Zhu, Bahman Afsari, Ludmila V Danilova, Alexander V Favorov, Wai Shing Lee, Dane Taylor, Chenyue W Hu, Byron L Long, David P Noren, Alexander J Bisberg, HPN-DREAM Consortium, Gordon B Mills, Joe W Gray, Michael Kellen, Thea Norman, Stephen Friend, Amina A Qutub, Elana J Fertig, Yuanfang Guan, Mingzhou Song, Joshua M Stuart, Paul T Spellman, Heinz Koeppl, Gustavo Stolovitzky, Julio Saez-Rodriguez, and Sach Mukherjee. Inferring causal molecular networks: empirical assessment through a community-based effort. *Nat Methods*, 13(4):310–8, Apr 2016.

[26] N Hollywood and H W Doelle. Effect of specific growth rate and glucose concentration on growth and glucose metabolism of escherichia coli k-12. *Microbios*, 17(67):23–33, 1976.

[27] Guy Karlebach and Ron Shamir. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol*, 9(10):770–80, Oct 2008.

[28] Kenneth J Kauffman, Purusharth Prakash, and Jeremy S Edwards. Advances in flux balance analysis. *Curr Opin Biotechnol*, 14(5):491–6, Oct 2003.

[29] Armin S Khonsari and Markus Kollmann. Perception and regulatory principles of microbial growth control. *PLoS One*, 10(5):e0126244, 2015.

[30] Ross Kindermann, James Laurie Snell, et al. *Markov random fields and their applications*, volume 1. American Mathematical Society Providence, RI, 1980.

[31] Lev Klebanov and Andrei Yakovlev. How high is the level of technical noise in microarray data? *Biol Direct*, 2:9, 2007.

[32] Oliver Kotte, Judith B Zaugg, and Matthias Heinemann. Bacterial adaptation through distributed sensing of metabolic fluxes. *Mol Syst Biol*, 6:355, 2010.

[33] Robert D Leclerc. Survival of the sparsest: robust gene networks are parsimonious. *Mol Syst Biol*, 4:213, 2008.

[34] Jong Min Lee, Jong Min Lee, Erwin P Gianchandani, James A Eddy, and Jason A Papin. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput Biol*, 4(5):e1000086, May 2008.

[35] Tracey A Lincoln and Gerald F Joyce. Self-sustained replication of an rna enzyme. *Science*, 323(5918):1229–32, Feb 2009.

[36] Po-Ling Loh, Martin J Wainwright, et al. Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *The Annals of Statistics*, 41(6):3022–3049, 2013.

[37] Sophia Y Lunt and Matthew G Vander Heiden. Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. *Annu Rev Cell Dev Biol*, 27:441–64, 2011.

[38] Radhakrishnan Mahadevan, Jeremy S Edwards, and Francis J Doyle, 3rd. Dynamic flux balance analysis of diauxic growth in escherichia coli. *Biophys J*, 83(3):1331–40, Sep 2002.

[39] Daniel Marbach, James C Costello, Robert Küffner, Nicole M Vega, Robert J Prill, Diogo M Camacho, Kyle R Allison, DREAM5 Consortium, Manolis Kellis, James J Collins, and Gustavo Stolovitzky. Wisdom of crowds for robust gene network inference. *Nat Methods*, 9(8):796–804, Aug 2012.

[40] B Mau, M A Newton, and B Larget. Bayesian phylogenetic inference via markov chain monte carlo methods. *Biometrics*, 55(1):1–12, Mar 1999.

[41] Patricia Menéndez, Yiannis A I Kourmpetis, Cajo J F ter Braak, and Fred A van Eeuwijk. Gene regulatory networks from multifactorial perturbations using graphical lasso: application to the dream4 challenge. *PLoS One*, 5(12):e14147, 2010.

[42] Agnès Miermont, François Waharte, Shiqiong Hu, Megan Nicole McClean, Samuel Bottani, Sébastien Léon, and Pascal Hersen. Severe osmotic compression triggers a slowdown of intracellular signaling, which can be explained by molecular crowding. *Proc Natl Acad Sci U S A*, 110(14):5725–30, Apr 2013.

[43] Amir Mitchell and Yitzhak Pilpel. A mathematical model for adaptive prediction of environmental changes by microorganisms. *Proc Natl Acad Sci U S A*, 108(17):7271–6, Apr 2011.

[44] Amir Mitchell, Gal H Romano, Bella Groisman, Avihu Yona, Erez Dekel, Martin Kupiec, Orna Dahan, and Yitzhak Pilpel. Adaptive prediction of environmental changes by microorganisms. *Nature*, 460(7252):220–4, Jul 2009.

[45] Douwe Molenaar, Rogier van Berlo, Dick de Ridder, and Bas Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol*, 5:323, 2009.

[46] Evan J Molinelli, Anil Korkut, Weiqing Wang, Martin L Miller, Nicholas P Gauthier, Xiaohong Jing, Poorvi Kaushik, Qin He, Gordon Mills, David B Solit, Christine A Pratilas, Martin Weigt, Alfredo Braunstein, Andrea Pagnani, Riccardo Zecchina, and Chris Sander. Perturbation biology: inferring signaling networks in cellular systems. *PLoS Comput Biol*, 9(12):e1003290, 2013.

[47] Hisao Moriya and Mark Johnston. Glucose sensing and signaling in saccharomyces cerevisiae through the rgt2 glucose sensor and casein kinase i. *Proc Natl Acad Sci U S A*, 101(6):1572–7, Feb 2004.

[48] Sach Mukherjee and Terence P Speed. Network inference using informative priors. *Proc Natl Acad Sci U S A*, 105(38):14313–8, Sep 2008.

[49] Deepak Nagrath, Marco Avila-Elchiver, Francois Berthiaume, Arno W Tilles, Achille Messac, and Martin L Yarmush. Integrated energy and flux balance based multiobjective framework for large-scale metabolic networks. *Ann Biomed Eng*, 35(6):863–85, Jun 2007.

[50] Sven Nelander, Weiqing Wang, Björn Nilsson, Qing-Bai She, Christine Pratilas, Neal Rosen, Peter Gennemark, and Chris Sander. Models from experiments: combinatorial drug perturbations of cancer cells. *Mol Syst Biol*, 4:216, 2008.

[51] Aaron M New, Bram Cerulus, Sander K Govers, Gemma Perez-Samper, Bo Zhu, Sarah Boogmans, Joao B Xavier, and Kevin J Verstrepen. Different levels of catabolite repression optimize growth in stable and variable environments. *PLoS Biol*, 12(1):e1001764, Jan 2014.

[52] Jeffrey D Orth, Ines Thiele, and Bernhard O Palsson. What is flux balance analysis? *Nat Biotech*, 28(3):245–248, 03 2010.

[53] Lior Pachter. The network nonsense of albert-lászló barabási. *url: https://liorpachter.wordpress.com/2014/02/10/the-network-nonsense-of-albert-laszlo-barabasi/*, Feb 2014 (last checked: 2016-05-02).

[54] Adi Raveh. On the use of the inverse of the correlation matrix in multivariate data analysis. *The American Statistician*, 39(1):39–42, 1985.

[55] Robert Schuetz, Nicola Zamboni, Mattia Zampieri, Matthias Heinemann, and Uwe Sauer. Multidimensional optimality of microbial metabolism. *Science*, 336(6081):601–4, May 2012.

[56] Matthew Scott, Carl W Gunderson, Eduard M Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330(6007):1099–102, Nov 2010.

[57] Matthew Scott and Terence Hwa. Bacterial growth laws and their applications. *Curr Opin Biotechnol*, 22(4):559–65, Aug 2011.

[58] Michael E Tipping and Christopher M Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.

[59] Amit Varma and Bernhard Palsson. Metabolic flux balancing: Basic concepts, scientific and practical use. *Nature Biotechnology*, 12(10):994–998, 1994.

[60] Alexei Vazquez, Qasim Beg, Marcio deMenezes, Jason Ernst, Ziv Joseph, Albert Barabasi, Laszlo Boros, and Zoltan Oltvai. Impact of the solvent capacity constraint on e. coli metabolism. *BMC Systems Biology*, 2(1):7, 2008.

[61] G N Vemuri, E Altman, D P Sangurdekar, A B Khodursky, and M A Eiteman. Overflow metabolism in escherichia coli during steady-state growth: transcriptional regulation and effect of the redox ratio. *Appl Environ Microbiol*, 72(5):3653–61, May 2006.

[62] G N Vemuri, M A Eiteman, J E McEwen, L Olsson, and J Nielsen. Increasing nadh oxidation reduces overflow metabolism in saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, 104(7):2402–7, Feb 2007.

[63] Zhong Wang, Mark Gerstein, and Michael Snyder. Rna-seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10(1):57–63, Jan 2009.

[64] Robert A. Weinberg. *The biology of cancer*. Garland science, Taylor Francis Group, LLC, 2007.

[65] Steen Lund Westergaard, Ana Paula Oliveira, Christoffer Bro, Lisbeth Olsson, and Jens Nielsen. A systems biology approach to study glucose repression in the yeast saccharomyces cerevisiae. *Biotechnol Bioeng*, 96(1):134–45, Jan 2007.

[66] Jason Weston, André Elisseeff, Bernhard Schölkopf, and Mike Tipping. Use of the zero norm with linear models and kernel methods. *The Journal of Machine Learning Research*, 3:1439–1461, 2003.

[67] Y Yarden and M X Sliwkowski. Untangling the erbb signalling network. *Nat Rev Mol Cell Biol*, 2(2):127–37, Feb 2001.

[68] Conghui You, Hiroyuki Okano, Sheng Hui, Zhongge Zhang, Minsu Kim, Carl W Gunderson, Yi-Ping Wang, Peter Lenz, Dalai Yan, and Terence Hwa. Coordination of bacterial proteome with metabolism by cyclic amp signalling. *Nature*, 500(7462):301–6, Aug 2013.

[69] Hyun Youk and Alexander van Oudenaarden. Growth landscape formed by perception and import of glucose in yeast. *Nature*, 462(7275):875–9, Dec 2009.

[70] Jamey Young, Kristene Henne, John Morgan, Allan Konopka, and Doraiswami Ramkrishna. Cybernetic modeling of metabolism: towards a framework for rational design of recombinant organisms. *Chemical Engineering Science*, 59:5041–5049, 2004.

[71] Shadia Zaman, Soyeon Im Lippman, Xin Zhao, and James R Broach. How saccharomyces responds to nutrients. *Annu Rev Genet*, 42:27–81, 2008.

[72] Fan Zhu and Yuanfang Guan. Predicting dynamic signaling network response under unseen perturbations. *Bioinformatics*, 30(19):2772–8, Oct 2014.

List of Figures

Acknowledgments

At the end I want to express my gratitude to all those people who helped and supported me to finish this thesis. Especially I would like to thank:

Markus Kollmann and Martin Lercher for being my supervisors and supporting me with my research projects. Thank you for introducing me to systems biology and letting me take part at one of the most interesting research fields.

My friends and colleagues at our institute at the Heinrich-Heine University Nima Abedpour, Nadia Heramvand, Linlin Zhao, Christopher Blum for the interesting scientific discussions and of course your good company. I really enjoyed our joint travel to the summer school in Taormina in 2015. I also want to thank the pure biologists from our institute for giving me the opportunity to experience real biological experiments. Special thanks goes to Peter Thul, who helped me out in my first months as well as to Petra Kolkhof for helping me with some experiments.

My friends and former fellow students from the physics department Marius Blecher, Luis Buslay, Mark Ludwig, Arash Nikoubashman, Adrian Jaspers and everyone else that I have forgotten to mention for the good times we spent in and outside the university.

My family for steadily supporting my studies and making this thesis possible in the first place.

An Nguyen for your love and care.

Ich versichere an Eides Statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der "Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine- Universität Düsseldorf" erstellt worden ist.

Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Armin Sadat Khonsari

Düsseldorf, 27. Juni 2016