

Alma Mater Studiorum - Università di Bologna

# **DOTTORATO DI RICERCA IN FISICA**

**Scuola di Scienze**

**XXVII Ciclo - A.A. 2011/2012**

Settore concorsuale di afferenza: 02/B3

Settore scientifico disciplinare: FIS/07

**Titolo:**

## **STATISTICAL MECHANICS FORMALISM AND METHODS FOR THE ANALYSIS OF REAL NETWORKS**

Presentata da:

**Giulia Menichetti**

Coordinatore Dottorato:  
Prof. Fabio Ortolani

Relatore:  
Prof. Daniel Remondini

Esame finale anno 2015

# Contents

<b>1</b>	<b>Network Ensembles</b>	<b>4</b>
1.1	Introduction to network ensembles . . . . .	4
1.1.1	Exponential random graph models . . . . .	6
1.2	Biological applications of network entropy . . . . .	14
1.2.1	Omics Data . . . . .	16
1.2.2	Modelling biological networks: the role of network entropy	20
1.2.3	Results . . . . .	24
1.3	Conclusions . . . . .	30
<b>2</b>	<b>Multilayer networks and multiplex networks</b>	<b>34</b>
2.1	Introduction to multilayer networks and multiplex networks . . .	34
2.2	Measures on multiplex networks . . . . .	41
2.2.1	Single-layer observables . . . . .	42
2.2.2	Total overlap and total weighted overlap of the multiplex networks . . . . .	45
2.2.3	Multilink observables . . . . .	47
2.2.4	Overlap multiplicity $\nu$ . . . . .	52

2.3	Canonical weighted multiplexes ensembles or exponential weighted multiplexes . . . . .	55
2.3.1	Uncorrelated and correlated canonical multiplex ensembles	57
2.3.2	Unweighted multiplex . . . . .	59
2.3.3	Weighted multiplex . . . . .	72
2.4	Weighted real data . . . . .	86
2.4.1	APS dataset . . . . .	88
2.4.2	Gene-Gene Duplex . . . . .	95
2.5	Conclusions . . . . .	100
<b>3</b>	<b>Control Theory</b>	<b>108</b>
3.1	Introduction to control theory . . . . .	108
3.2	Definition of controllability and structural controllability of a network . . . . .	109
3.2.1	Matching, Maximum Matching and Perfect Matching . . .	113
3.2.2	Real networks and null models . . . . .	115
3.3	Cavity method and Belief Propagation: a short introduction . . .	116
3.3.1	Disordered systems . . . . .	117
3.3.2	Belief propagation on factor graphs . . . . .	130
3.4	The BP approach to the maximum matching problem . . . . .	133
3.4.1	The BP equations . . . . .	136
3.4.2	Free energy and energy of the problem . . . . .	140
3.4.3	The $\beta \rightarrow \infty$ limit . . . . .	141
3.4.4	BP/MS Equations in an ensemble of random networks with given degree distribution . . . . .	142

3.5 Controllability and minimal degrees . . . . . 145

3.5.1 Sufficient condition for the full controllability of networks . 146

3.5.2 Conditions for the full controllability of random networks . 148

3.5.3 Improving the controllability of a network . . . . . 152

3.6 Controllability of multiplex networks . . . . . 154

3.6.1 The structural controllability of a multiplex network . . . 154

3.6.2 Mapping to a constraint Maximum Matching Problem . . 157

3.6.3 BP Equations . . . . . 158

3.6.4 The limit  $\beta \rightarrow \infty$  . . . . . 164

3.6.5 Controllability of uncorrelated multiplex networks with given  
in-degree and out-degree distribution . . . . . 166

3.6.6 Phase transition in the controllability of Poisson multiplex  
networks . . . . . 167

3.6.7 Dependence on the correlation between low in-degree and  
out-degree nodes in the different layers . . . . . 174

3.7 Conclusion . . . . . 178

**4 Supplementary Information 183**

4.1 APS dataset . . . . . 183

4.1.1 The two datasets . . . . . 183

4.1.2 Total overlap and total weighted overlap of the multiplex  
networks . . . . . 186

4.1.3 Degree and multidegree distribution of the two multiplex  
networks . . . . . 188

4.1.4 Weighted network properties of single layers . . . . . 190

4.1.5	Statistical analysis of the properties of multilinks in the CoCo-PRL/PRE multiplex network . . . . .	192
4.1.6	Statistical analysis of the properties of multilinks in the CoCi-PRE multiplex network . . . . .	195
4.1.7	$\Psi$ and $\Xi$ . . . . .	197
4.2	Weighted multiplex ensembles . . . . .	202
4.2.1	Examples of uncorrelated weighted multiplex networks . . . . .	202
4.2.2	Examples of correlated weighted multiplex networks . . . . .	206
4.3	Controllability and minimal degrees . . . . .	212
4.3.1	Stability condition . . . . .	212
4.3.2	Number of driver nodes . . . . .	215
4.3.3	Improving the controllability of scale-free networks . . . . .	216
4.3.4	Poisson networks . . . . .	217
4.3.5	Improving the controllability of Poisson networks . . . . .	219
4.4	Controllability of multiplex networks . . . . .	221
4.4.1	Stability condition . . . . .	221
4.4.2	Correlations . . . . .	223

# Introduction

This PhD thesis describes topics of network theory along three main themes. In Ch. 1 a general introduction to network ensembles is given, and the relations with “standard” equilibrium statistical mechanics are described. The canonical ensembles of networks, i.e. ensembles where some selected constraints are satisfied on average, provide answers to questions such as: how many networks satisfy the considered constraints, and how these constraints affect more complex observables and behaviours. Moreover, the entropy measure for the canonical ensemble is fundamental for the computation of the probability marginals and its interpretation has been exploited for relevant applications. In particular, network entropy, corresponding to the logarithm of the number of typical graphs in the considered ensemble, has been the starting point for further studies on biological networks integrated with different types of omics data. We modelled the statistical properties of the integrated PPI-signalling-mRNA expression networks in different cases (i.e. cancer studies and ageing studies) interpreting the network entropy measure as the extent of the *parameter space* allowed to the cell, in terms of cell phenotypes or clonality. The major results of this chapter have been reported in two papers: the first one is Menichetti and Remondini (2015) and the second one is currently under submission.

Multilayer networks and, in particular, multiplex networks (in which different networks share the same nodes) are the main topic of Ch. 2. Multilayer networks were introduced to evaluate and quantify the correlations between interdependent networks or, moreover, to thoroughly describe also a single network in which different kinds of interaction are represented. For example, in biological systems, gene, protein and metabolite networks have strong correlations and interdependencies that cannot be fully pictured in terms of single graphs. In this chapter a fully description of the main observables related to multiplex networks is given. We consider the formalism of multilinks in order to fully characterise link overlap in maximum-entropy multiplex ensembles. We showed some real examples in which relevant information could be uncovered only by considering the multiplex nature of the given system. The first presented case is APS, i.e. citation and collaborations networks from different journals of the American Physical Society. We built different types of duplex networks and thank to the multilink observables we discovered some relevant patterns in the citation-collaboration behaviours. The second example is a biological duplex network: starting from a case-control study on colorectal cancer, one layer is related to normal samples, while the other one to cancer samples. Also in this situation, multilinks highlight significant differences and nontrivial similarities between healthy and cancer biological processes. Moreover, on this particular duplex, we tested our null models and the related algorithms. The results in this chapter are described in Menichetti et al. (2014a) and Menichetti et al. (2014b).

The last chapter is completely dedicated to control theory and its relation with network theory. Control theory has a wide range of applications, from drug discovery to the study of biomass flows, or furthermore, to the description of

dynamical process in the brain. We define the controllability of a network as the possibility to drive its dynamical state to any desired state by applying different external signals only to a subset of nodes defined as *driver nodes*. In this chapter, the main concepts and calculations of this theory are presented, and a short introduction to cavity method / belief propagation is given. Moreover, we characterise how the structural controllability of a network is affected by the fraction of low in-degree and low out-degree nodes. Finally, we present a novel approach to the controllability of multiplex networks since, in the last years, large attention has been given to the dynamics taking place on multiplex networks but no studies consider their controllability. We studied the case in which the driver nodes are forced to be the same in each layer. As expected, a multiplex network is more demanding in terms of controllability than the situation in which each layer is considered separately. Anyway, the introduction of some correlations in the low degree nodes can reduce this gap. Moreover, in the case of Poisson duplex networks, small variations of the average degree can cause discontinuities in the number of driver nodes. In this chapter we collected all the preliminary theoretical work needed to fully characterise real data. These results are described in two papers, the first one is Menichetti et al. (2014c) while the second one, related to multiplex controllability, is in progress.

In summary, this thesis provides a thoroughly theoretical background in network theory and shows novel applications to real problems and data.

# 1 Network Ensembles

## 1.1 Introduction to network ensembles

Equilibrium statistical mechanics provides a general framework for the development of null models in network theory. Modelling, as always, helps us in the understanding of the important features characterising the network structure and the interplay with processes that take place on it (e.g. the flow of traffic on the Internet, the spread of a disease over a social network). Dynamical processes are affected by the phenomenological quantities characterising the underlying network (Park and Newman, 2004). In particular, the study of community structures and motifs has become very popular both in social systems and omics studies (e.g. KEGG pathways).

Are these higher order characteristics explainable in terms of low level features such as the degree sequence of our network or are they additional structural patterns? In statistical mechanics the essential concept of ensemble is defined as a large number of copies of a system (sometimes an infinity), considered all at once, each of which represents a possible state in which the real system might be in (microstate). Also in network theory, ensemble models are those that do not focus on a single network, but consider a probability distribution over many pos-

sible networks. Moreover, in network ensembles with specified constraints, all the other features become completely random. We define these network ensembles *randomised*: for a given real network, different randomised networks ensembles can be generated, depending on the structural characteristics of the network we want to consider as constraints (Bianconi, 2009). These ensembles have been introduced and systematically used as reference to identify non-random patterns in real network and to reveal how a *typical graph* with given properties looks like (Squartini and Garlaschelli, 2011; Hartmann and Weigt, 2005).

An observed real network is then considered just as a single realisation of a larger statistical ensemble gathering all the possible realisations compatible with some defined features. Mathematically speaking, a statistical ensemble of networks can be defined as a set of graphs  $\mathcal{G}$  where, for each graph  $G \in \mathcal{G}$  a probability  $P(G)$  is defined. We are mainly interested in the so-called *maximum-entropy graph ensembles* with given constraints. The concept of entropy shows up naturally in many different situations and theories, starting from the first probabilistic interpretation of thermodynamic entropy given by Ludwig Boltzmann, and becoming a key-concept of information theory and hypothesis testing over large deviations (Greven et al., 2003). From the point of view of statistical mechanics, given a set of macroscopic variables, entropy gives us a measure of the spreading out of probability over different possible microscopic states.

The method of the maximisation of network entropy with given constraints provides  $\{P(G)\}$  and the analytical expression for the marginal probabilities  $\{p_{ij}\}$  (probability of having a link between node  $i$  and node  $j$ ). Expected values of quantities of interest can be calculated analytically, without sampling the configuration space as in the huge time-consuming local rewiring algorithm (micro-

canonical approach). We can distinguish, then, the main properties explained by the given constraints from those more non-trivial. In Squartini and Garlaschelli (2011) a comparison among the main procedures of randomisation for real networks is presented.

Once calculated  $\{P(G)\}$ , the related entropy value can be interpreted as a measure of the level of organisation and *order*. A real network is characterised by a collection of features that we want to investigate with our models. If we want to assess the role of these structural features a useful recipe is considering a subsequent series of randomised networks ensembles, with an increasing number of structural constraints shared with the real network. The entropy value of these network ensembles decreases with the addition of more constraints. Furthermore, we can evaluate how selective a particular constraint is if it produces a significant difference in the subsequent entropy values (Bianconi, 2009).

In the following we present an introduction to the main topic of exponential random graph models or so-called *canonical network ensembles* and some hints of the *micro-canonical network ensembles*.

### 1.1.1 Exponential random graph models

We talk of *exponential random graph* when we consider the distribution over a specified set of graphs that maximises the entropy with given constraints. It is literally the analogue of the Boltzmann distribution of a physical system over its microstates at finite temperature (Park and Newman, 2004). Like all maximum entropy ensembles, it gives the best prediction of an unknown quantity, given a set of enforced constraints (Jaynes, 1957; Cover and Thomas, 2006).

In this introduction to the subject, for the sake of simplicity, we consider just undirected simple graphs (at most a single edge between any pair of vertices), without self-loops.

Let's consider a set  $\mathcal{G}$  of graphs with the same number of nodes  $N$ . We call  $G$  a graph in our set of graphs and define  $P(G)$  as the probability of that graph within our ensemble. Each graph  $G$  is identified by the so-called adjacency matrix  $\{a_{ij}\}$ , where each  $a_{ij}$  could be 0 (event *no link*) or 1 (event *link*). The sum over all the graphs  $G$  of the ensemble  $\mathcal{G}$  is then

$$\sum_G = \sum_{\{a_{ij}\}} = \prod_{i < j} \sum_{a_{ij}=0}^1 \quad (1.1)$$

$P(G)$  is chosen such that the expectation value of each our observables  $\{O_i\}$  is equal to its observed value. The best choice of probability distribution, as previously explained, is given by the maximisation of the Gibbs entropy.

$$S = - \sum_{G \in \mathcal{G}} P(G) \log(P(G)) \quad (1.2)$$

subject to the constraints

$$\sum_G P(G) O_i(G) = \langle O_i \rangle \quad (1.3)$$

$$\sum_G P(G) = 1 \quad (1.4)$$

where  $O_i(G)$  is the value of  $O_i$  in the graph  $G$ .

For this kind of maximisation problem with constraints, we introduce the La-

grangian multipliers  $\alpha, \{\lambda_i\}$  and we solve the following equation

$$\frac{\partial}{\partial P(G)} \left[ S + \alpha \left( 1 - \sum_G P(G) \right) + \sum_i \lambda_i \left( \langle O_i \rangle - \sum_G P(G) O_i(G) \right) \right] = 0 \quad (1.5)$$

for all graphs  $G$ . This leads to the solution

$$P(G) = \frac{e^{-H(G)}}{Z} \quad (1.6)$$

where  $H(G)$  is the graph Hamiltonian, defined as  $H(G) = \sum_i \lambda_i o_i(G)$ , and  $Z$  is the partition function, defined as  $Z = e^{\alpha+1} = \sum_G e^{-H(G)}$ . Following the path given by the conventional statistical mechanics we introduce the free energy as  $F = -\log Z$ . If we substitute in Eq. 1.2 the probability distribution given by Eq. 1.6, we obtain

$$S(G) = \langle H(G) \rangle + \log Z = \langle H(G) \rangle - F \quad (1.7)$$

This equation looks familiar and similar to the physical equation  $F = U - TS$  (even if the true parallelism is with the gran canonical ensemble).

We consider now one of the most popular ensembles, the conjugate-canonical ensemble of the so-called *configuration model*. This is one of most important model used in network theory because it encodes the main topological notions of a real network. It specifies only local constraints, namely, the degree  $k_i$  (the number of incident edges) of each vertex ( $i = 1, \dots, N$ ). The Hamiltonian for this model expresses the constraints upon the degrees

$$H = \sum_i \lambda_i k_i \quad (1.8)$$

Using the formalism of the adjacency matrix, each  $k_i$  is equal to  $k_i = \sum_{j \neq i} a_{ij}$ .  $H$  then becomes  $H = \sum_{ij} \lambda_i a_{ij} = \sum_{i < j} (\lambda_i + \lambda_j) a_{ij}$ . The computation of the partition function is

$$Z = \sum_{\{a_{ij}\}} \exp \left( - \sum_{i < j} (\lambda_i + \lambda_j) a_{ij} \right) \quad (1.9)$$

$$= \prod_{i < j} (1 + e^{-(\lambda_i + \lambda_j)}) = \prod_{i < j} \mathcal{Z}_{ij} \quad (1.10)$$

The probability  $P(G)$  of a graph in this ensemble can be written as

$$P(G) = \frac{e^{-\sum_{i < j} (\lambda_i + \lambda_j) a_{ij}}}{\prod_{i < j} (1 + e^{-(\lambda_i + \lambda_j)})} \quad (1.11)$$

and the free energy is

$$F = - \sum_{i < j} \log (1 + e^{-(\lambda_i + \lambda_j)}) \quad (1.12)$$

The probability  $p_{ij}$  of a link  $i, j$  is

$$p_{ij} = \langle a_{ij} \rangle = \frac{\partial F}{\partial (\lambda_i + \lambda_j)} = \frac{1}{e^{\lambda_i + \lambda_j} + 1} = \frac{e^{-(\lambda_i + \lambda_j)}}{1 + e^{-(\lambda_i + \lambda_j)}} \quad (1.13)$$

In this ensemble these marginal probabilities show some *natural correlations* given by the degree sequence, i.e.  $p_{ij} \neq f(\lambda_i) f(\lambda_j)$  (Bianconi, 2009).

Equations 1.10, 1.12 and 1.13 recall a physical grand canonical ensemble where edges become particles and pairs of vertices are single-particle states. The partition function  $Z$  can be expressed as the productory of  $N(N - 1)/2$  single-particle state partition functions  $\mathcal{Z}_{ij}$ . The exponential random graph models are generally called *canonical network ensembles* because in network theory we mainly consider the distinction between *soft* and *hard* constraints. We define *microcanonical network ensembles* by imposing a set of hard constraints that

must be satisfied by each network in the ensemble, while canonical network ensembles are considered those who satisfy soft constraints, i.e., the constraints are satisfied on average.

Using an approach more similar to the physical grand canonical ensemble, the partition function  $Z$  can be expressed as

$$\begin{aligned} Z &= \sum_{\{a_{ij}\}} \exp \left( e^\mu L - \sum_i \lambda_i k_i \right) = \sum_{\{a_{ij}\}} \exp \left( \mu \sum_{i<j} a_{ij} - \sum_{i<j} (\lambda_i + \lambda_j) a_{ij} \right) \\ &= \prod_{i<j} (1 + e^{\mu - (\lambda_i + \lambda_j)}) \end{aligned} \quad (1.14)$$

where  $L$  express the number of edges in the network. In this formalism  $Z$  plays the role of *gran partition function* and the previous free energy  $F$  can be considered as *gran potential*. Usually, the chemical potential  $\mu$  is not considered explicitly since it can be considered as an additional constant in the hamiltonian (Garlaschelli and Loffredo, 2006).

Following the usual parallelism, the link probability  $p_{ij}$  behaves like the average occupation number of a specific single-particle state. Moreover, Eq. 1.13 recalls the Fermi-Dirac distribution: simple graphs have each single-particle state occupied at most by one particle, according to the Pauli exclusion principle. Therefore, simple graphs have many similarities with systems of non-interacting fermions (Park and Newman, 2004).

For the sake of completeness, we make a small reference to weighted networks. We consider the situation where the weight of a link  $a_{ij}$  can take only integer values. This is not a very restrictive constraint since any finite network with weights of the links taking rational numbers can be easily reduced to a network of integer weights (Bianconi, 2009). These integer values can be considered as an arbitrary

number of unitary links between any pair of nodes. Weighted networks have a correspondence with Bose-Einstein gas. Fixing on average a specific *strenght* sequence we obtain

$$Z = \prod_{i < j} \left( \frac{1}{1 - e^{-(\lambda_i + \lambda_j)}} \right) \quad (1.15)$$

$$w_{ij} = \langle a_{ij} \rangle = \frac{1}{e^{\lambda_i + \lambda_j} - 1} = \frac{e^{-(\lambda_i + \lambda_j)}}{1 - e^{-(\lambda_i + \lambda_j)}} \quad (1.16)$$

$$p_{ij} = \langle \theta(a_{ij}) \rangle = e^{-(\lambda_i + \lambda_j)} \quad (1.17)$$

$$\pi_{ij}(a_{ij}) = \langle \delta(a_{ij}) \rangle = e^{-(\lambda_i + \lambda_j) a_{ij}} (1 - e^{-(\lambda_i + \lambda_j)}) \quad (1.18)$$

$$P(G) = \prod_{i < j} \pi_{ij}(a_{ij}) \quad (1.19)$$

where  $w_{ij}$  is usually called average weight and is the expected number of unitary links between nodes  $i$  and  $j$ , and  $\pi_{ij}(a_{ij})$  is the probability of having weight  $a_{ij}$  between nodes  $i$  and  $j$  with  $p_{ij} = \sum_{a_{ij} \neq 0} \pi_{ij}(a_{ij})$ .

We present now another useful approach for canonical ensembles, especially considered in our biological models. The main role here is played by the marginals  $p_{ij}$ , the probability of having a link between node  $i$  and node  $j$ . Each undirected simple graph  $G$ , belonging to a canonical ensemble, is described by its probability distribution  $P(G)$ , defined by its adjacency matrix  $\{a_{ij}\}$

$$P(G) = \prod_{i < j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1 - a_{ij}} \quad (1.20)$$

where a link between nodes  $i$  and  $j$  is present with probability  $p_{ij}$ , otherwise absent with probability  $(1 - p_{ij})$ . Considering Eq. 1.11 and Eq. 1.13 this equality appears straight clear. The matrix elements appear as independent and uncorrelated random parameters. Defining the log-likelihood function as  $L =$

$-\log(P(G))$ , entropy  $S$  becomes nothing more than the average log-likelihood over the probability distributions of the marginals

$$S = \langle L \rangle = - \sum_{i < j} p_{ij} \log p_{ij} - \sum_{i < j} (1 - p_{ij}) \log(1 - p_{ij}) \quad (1.21)$$

The entropy of a canonical ensemble is considered as the logarithm of the number of *typical* networks and it takes exactly the form of a Shannon entropy (Anand and Bianconi, 2009).

In the same way we previously computed  $P(G)$  we now maximise  $S$  with some constraints in order to find  $\{p_{ij}\}$ . The two expressions of entropy in Eq. 1.2 and 1.21 are exactly equivalent. This can be proved performing the following calculation

$$\begin{aligned} S &= - \sum_{\{a_{ij}\}} \prod_{i < j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1 - a_{ij}} \log \left( \prod_{k < l} p_{kl}^{a_{kl}} (1 - p_{kl})^{1 - a_{kl}} \right) \\ &= - \sum_{i < j} p_{ij} \log p_{ij} - \sum_{i < j} (1 - p_{ij}) \log(1 - p_{ij}) \end{aligned}$$

We suppose that our ensemble is subjected to  $\kappa = 1 \dots M$  structural constraints, i.e.

$$f_k(\{p_{ij}\}) = F_\kappa \quad (1.22)$$

where  $f_k(\{p_{ij}\})$  is a constraint function on the probability matrix  $f_k(\{p_{ij}\})$ .

The link probabilities are provided by the maximisation of the Shannon entropy subjected to our constraints. The marginal probabilities  $p_{ij}$  are given as the solution to the system of equations

$$\frac{\partial}{\partial p_{ij}} \left\{ S + \sum_{\kappa=1}^M \lambda_\kappa (F_\kappa - f_\kappa(\{p_{ij}\})) \right\} = 0 \quad (1.23)$$

For the configuration ensemble we have  $N$  constraints given by

$$k_i = \sum_{j \neq i} p_{ij} \quad (1.24)$$

In order to calculate  $p_{ij}$  we introduce the function

$$F^* = - \sum_{i < j} p_{ij} \log p_{ij} - \sum_{i < j} (1 - p_{ij}) \log(1 - p_{ij}) + \sum_i \lambda_i \left( k_i - \sum_j p_{ij} \right) \quad (1.25)$$

and we impose

$$\frac{\partial F^*}{\partial p_{ij}} = \log \frac{1 - p_{ij}}{p_{ij}} - (\lambda_i + \lambda_j) = 0 \quad (1.26)$$

The marginal probabilities result

$$p_{ij} = \frac{e^{-(\lambda_i + \lambda_j)}}{1 + e^{-(\lambda_i + \lambda_j)}} = \frac{z_i z_j}{1 + z_i z_j} \quad (1.27)$$

with the variables  $z_i = e^{-\lambda_i}$ , which are commonly referred to as *hidden variables*.

The probabilities in Eq. 1.27 and Eq. 1.13 are exactly equivalent.

Lastly, we make a few considerations about microcanonical network ensembles.

These ensembles are composed by all those networks which satisfy exactly the constraints. Following the approach presented in Bianconi (2009) one may introduce a *partition function*  $Z$  that counts the number of networks which fulfil the requirements. The main equations for undirected simple networks in the configuration model are

$$Z = \sum_{\{a_{ij}\}} \prod_{i=1}^N \delta \left( \sum_j a_{ij} - k_i \right) \exp \left( \sum_{i < j} h_{ij} a_{ij} \right) \quad (1.28)$$

$$\Sigma = \frac{1}{N} \log Z |_{h_{ij}=0 \ \forall (i,j)} \quad (1.29)$$

where  $\Sigma$  is defined as *entropy per node* (Bianconi, 2009). As long as we consider network ensembles with an extensive number of constraints the microcanonical entropy per node  $\Sigma$  and the canonical entropy per node  $S/N$  are not equal in the thermodynamic limit (Anand and Bianconi, 2009, 2010). In both the situations if two graphs satisfy in the same way the constraints they will have equal probabilities (i.e.  $P(G_1) = P(G_2)$ ). Anyway, the microcanonical ensemble defines a null probability for all those graphs in which the constraints are not matched exactly, while for the canonical ensemble all possible graphs can occur (constraints on the average values). From this perspective canonical ensembles are more robust to errors in the original data: the *true* graph will never appear in a microcanonical model based on the observed and biased data (Squartini and Garlaschelli, 2011). If we consider the uncertainty affecting biological data the canonical approach should be the best one.

## 1.2 Biological applications of network entropy

Biological systems can be seen as complex systems that translate genomic information into phenotypes Pagel and Pomiankowski (2008); De Las Rivas and Fontanillo (2010). A useful approach is to describe these systems as networks, with the system elements (eg. genes, proteins) as nodes, and the relationships between them (eg. transcription or protein-protein interaction) as edges (Barabasi and Oltvai, 2004; Alm and Arkin, 2003). An important class of biological networks comprises the protein-protein interaction networks (PPI, Vidal et al. (2011); Cerami et al. (2011); Szklarczyk et al. (2011)): edges in these networks describe interactions between proteins that are part of the same physical

complex or post-translational modifications mediating signal transduction flows. Networks of interacting proteins can be thought as characterizing the cell phenotypes given their genetic and transcriptomic profile. These and other interactions are also encoded into functional pathways, such as signalling and metabolic pathways, as are mapped for example in KEGG database (Kyoto Encyclopaedia of Genes and Genomes, [www.genome.jp/KEGG](http://www.genome.jp/KEGG)). In our study we are interested in the integration between the transcriptomic and the interactomic data, thus the statistical properties of integrated PPI-signalling-mRNA expression networks seem to be good observables to investigate systemic pathologies such as cancer and ageing (Teschendorff and Severini, 2010; Barea and Bonatto, 2009). This approach can be more informative than analyzing gene expression data on its own. Indeed, integrative PPI-mRNA expression studies have helped to tease out relevant patterns of expression variation in the contextual framework of signalling pathways and protein complexes (Pagel and Pomiankowski, 2008; West et al., 2012; van Wieringen and van der Vaart, 2011).

Using the tools presented in Sec. 1.1 we have the chance to build up a thorough biological network model. Thanks to some suitable constraints encoding the most relevant network features, we can evaluate the information content of biological structures, and moreover, we can apply specific methods for time-dependent and time-independent data (Anand and Bianconi, 2010; Bianconi et al., 2009).

Our approach relies on the theory of network ensembles with given topology (encoded in the degree sequence) and metrics (represented by distance between values assigned to the nodes): the PPI-signalling structure is embedded in the network topology, while mRNA expression data define the values assigned to the nodes.

We studied two biological phenomena that encode different landscapes of cellular perturbation, namely cancer and ageing in humans, and whose datasets were characterized by a different experimental design (case-control studies and a time series built on samples of different age). Network entropy approach offers a new perspective to the study of such phenomena, highlighting a more systemic behaviour of the cell beyond single-element analysis, but nonetheless it can be applied at several scales, from a whole-cell point of view (the full network) to single biological pathways characterising the main cell processes like metabolism and signalling (subnetworks defined by a priori biological knowledge), up to single nodes (genes/proteins in the network).

### 1.2.1 Omics Data

#### **PPI-signalling network**

In order to define a network in which the nodes (namely proteins, measured by their mRNA transcription profile) could be adequately annotated both in terms of their biological function and their potential interactions, we considered only the genes that were annotated both in KEGG database and in PathwayCommons ([www.pathwaycommons.org](http://www.pathwaycommons.org)) PPI network.

We started considering the protein-protein interaction network extracted from the Pathway Commons database regarding *Homo Sapiens* proteins. The initial PPI network contained 11604 nodes and 420601 links: after self-interaction and redundant annotation removal we obtained a giant component of 11394 nodes and 420516 links. Since we used different gene expression datasets on different microarray platforms, we considered the intersection of the PPI protein IDs with

the gene annotations of each microarray platform, considering only the genes that had also a known annotation in the KEGG database. In this way, each network could be further divided considering nodes annotated into each single KEGG pathway.

This procedure produced different networks for each considered platform, with a number of nodes ranging from 2000 to 3000.

## **Cancer datasets**

The analysis has been performed onto four datasets by downloading the normalised data from GEO Omnibus ([www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo)).

The first dataset (referred to as “Colon”, GEO accession number GSE4183 (Gy-  
orffy et al., 2009)) is composed by 8 normal colon biopsies and 15 colorectal cancer samples.

The second dataset is related to Ewing’s sarcoma (“Ewing” dataset, GEO accession number GSE12102 Scotlandi et al. (2009)), consisting of 30 primary and 7 metastatic tumour samples. Other two dataset refer to breast cancer samples: in the first we have primary tumour specimens that developed metastasis or not (97 and 28 samples respectively, referred to as “Met”, GEO accession number GSE2990 Sotiriou et al. (2006); Loi et al. (2007)), while in the second there are primary tumour biopsies that relapsed or not (107 and 179 samples respectively, referred to as “REL”, GEO accession number GSE2034 Wang et al. (2005)).

Colon and Ewing datasets are both profiled with the Affymetrix U133 plus 2 microarray platform, and the intersection with the PPI network and the KEGG database resulted in a network with 2835 nodes.

Rel and Met datasets are both profiled with the Affymetrix U133 A microarray platform, and the intersection with the PPI network and the KEGG database resulted in the a network of 2618 nodes.

In each dataset, a restricted gene list (and a corresponding reduced network) was obtained by performing a Student's T test for uncoupled samples over the two groups in which each dataset is divided into. The main purpose of this selection is to evaluate the behaviour of the network entropy measure for a subset of nodes that significantly behave differently in the two groups, as compared to the full set of available nodes in the network.

For the Colon dataset we applied a  $P < 0.05$  significance threshold plus Benjamini-Hochberg post-hoc correction, obtaining a subnetwork of 312 nodes. For the Ewing and the Breast datasets we only applied a  $P < 0.05$  significance threshold, obtaining a network with 136 nodes for Ewing, 151 and 313 nodes for Met and Rel datasets respectively, since almost no genes would have passed the post-hoc correction. This is probably due to the fact that in these datasets the differences between groups are less pronounced than in a normal-cancer comparison, as described in the related papers from which the data were collected.

Since we can calculate the network entropy value for each sample, we obtain 23 entropy values for Colon, 37 for Ewing, 125 for Met and 286 for Rel datasets, both for the full network (that will be used for single-node entropy calculation, as described below) and the 5% significance gene selection.

In order to estimate significant differences between the groups, as a typical *case-control* study design, since the null distribution of network entropy values is not known in advance for arbitrary networks, we performed nonparametric Wilcoxon

rank sum tests between the entropy values for each group.

## **Ageing dataset**

We considered a cross-sectional study (time series) of 25 whole-genome expression profiles of T lymphocytes extracted from healthy males of ages spanning typical adult human lifespan (from 25 to 97 years, see Remondini et al. (2010) for further details). This dataset is naturally divided into 5 age groups: A) 25-34 y (mean = 29.6 y); B) 43-46 y (mean = 44 y); C) 55-62 y (mean = 58.2 y); D) 70-79 y (mean = 74.2 y); E) 92-97 y (mean = 94.4 y).

The gene expression dataset (obtained through a custom array, see Remondini et al. (2010)) after processing is composed by 13103 probes x 25 age samples. The intersection with the giant component of Pathway Commons data and the KEGG database results in a PPI network of 1976 nodes, used for single-node entropy analysis. A restricted gene list was obtained by performing a 1-way Anova over the age groups, in order to look for genes significantly changing expression profile in time. With a  $P < 0.05$  significance threshold plus Benjamini-Hochberg post-hoc correction we obtained a subnetwork of 217 nodes. We applied the same significance threshold considered in the original paper in order to compare the results obtained by gene expression analysis and the results obtained by this network entropy approach.

We obtained 25 network entropy values (one for each sample) both for the whole network and for the 5% significance gene selection. Also in this case we applied nonparametric test for network entropy comparisons, namely Kruskal-Wallis test over the 5 age groups, to define a subgroup of genes significantly changing expression profile over the whole time series, and Wilcoxon rank sum

test for comparison between any two groups.

## 1.2.2 Modelling biological networks: the role of network entropy

Based on the formalism developed in Sec. 1.1.1, we apply the concept of entropy of network ensembles to a real biological situation. In our case study, each sample can be described by a network of  $N$  nodes (adjacency matrix  $\{a_{ij}\}$ ) and by an additional distance matrix  $\{d_{ij}^a\}$  (where the label  $a$  identifies the considered sample). The first set of observables is related to the network topological structure, and is given by the degree sequence of the PPI network, namely the  $N$ -dimensional vector of the connectivity degree of each node:  $\{k_i\}$ ,  $i = 1, \dots, N$ , with  $k_i = \sum_{j \neq i} a_{ij}$ . Since we consider a network (and calculate an entropy value) for each sample, these topological constraints are equal for all the samples<sup>1</sup>. The second set of observables is related to the distance values of the network, expressing metric relations between nodes: assigning to the nodes of each sample the values of mRNA expression of the corresponding genes in the selected microarray  $g_i^a$  (with index  $i$  ranging over all the nodes and index  $a$  ranging over all the samples of the dataset), we define  $d_{ij}^a$  as the euclidean distance of the gene expression values, i.e.

$$d_{ij}^a = \sqrt{(g_i^a - g_j^a)^2} = |g_i^a - g_j^a| \quad (1.30)$$

We collect all these values into an histogram with  $N_b$  bins, with a number of bins equal to the square root of the number of nodes in the network:  $N_b = \sqrt{N}$  (a

---

<sup>1</sup>Measured on the same microarray platform.

reasonable choice considering the sparsity of the PPI network and of its subsets). For each couple of genes we have a particular distance value but not necessarily a link in the PPI network. The second set of network observables refers to the number of the PPI links whose distance values fall in a given bin. For each distance bin we count the number of these PPI links related to it and we fix them on average. We remark that this set of observables is specific for each sample, being related to its expression profile.

The entropy of network ensembles follows from Eq. 1.21, where as previously defined,  $p_{ij}$  represents the probability of having a link between node  $i$  and node  $j$ . In a generic graph of this ensemble, a link  $a_{ij}$  is present with probability  $p_{ij}$ , otherwise absent with probability  $(1 - p_{ij})$ .

We define the *spatial ensemble* as an ensemble of network obtained by enforcing the constraints on the degree sequence  $\{k_i\}$  and on the number  $B_l$  of PPI links belonging to each distance bin,  $d_{ij} \in I_l$ , described by the following equations:

$$k_i = \sum_j^N p_{ij}; \quad i = 1, \dots, N \quad (1.31)$$

$$B_l = \sum_{i < j}^N \chi_l(d_{ij}) p_{ij}; \quad l = 1, \dots, N_b \quad (1.32)$$

where  $N$  is the number of nodes in the network,  $N_b$  is the number of bins considered for the empirical distribution of distances, and  $\chi_l$  is the characteristic function of each bin of width  $(\Delta d)_l$ :  $\chi_l(x) = 1$  if  $x \in [d_l, d_l + (\Delta d)_l]$ ,  $\chi_l(x) = 0$  otherwise.

The probability matrix  $\{p_{ij}\}$  is obtained by the constrained maximization of the entropy function (Eq. 1.23), as described in the following equation:

$$\frac{\partial}{\partial p_{ij}} \left\{ S + \sum_i^N \lambda_i \left( k_i - \sum_j p_{ij} \right) + \sum_l^{N_b} g_l \left( B_l - \sum_{i < j} \chi_l(d_{ij}) p_{ij} \right) \right\} = 0$$

where  $\lambda_i$  and  $g_l$  are the the Lagrangian multipliers related to our constraints.

For each  $(i, j)$  the resulting marginal probability is

$$p_{ij} = \sum_l^{N_b} \chi_l(d_{ij}) \frac{e^{-(\lambda_i + \lambda_j + g_l)}}{1 + e^{-(\lambda_i + \lambda_j + g_l)}} = \sum_l^{N_b} \chi_l(d_{ij}) \frac{z_i z_j W_l}{1 + z_i z_j W_l} \quad (1.33)$$

where  $z_i = e^{-\lambda_i}$ ,  $W_l = e^{-g_l}$ , commonly known as hidden variables, are functions of the Lagrangian multipliers  $\lambda_i$  and  $g_l$ . If we consider only the constraints on the degree sequence stated in Eq. 1.31 we come back to the so called *configuration ensemble* that was fully explained in the previous section. The number of constraints for the configuration ensemble is  $N$ , while for the spatial ensemble it is  $N + N_b$ . The additional Lagrangian multipliers  $\{W_l\}$  contain information about gene expression profiles, modulating the probability  $p_{ij}$  of having a link between node  $i$  and node  $j$  with a given expression difference  $d_{ij}$ . We remark that a significant difference between the network entropy calculated in the spatial and configuration ensembles reflects the relevance of the information encoded in the gene expression data, as will be the case for all of our analyses. In particular, what matters is how this genetic information is filtered by the PPI network.

The canonical ensemble deriving from a real instance gives an entropy value that is considered as the logarithm of the number of “typical” networks in this ensemble, given the constraints to be satisfied on average by each network belonging to the ensemble (Anand and Bianconi, 2009).

Considering the link probabilities  $p_{ij}$  obtained for the full PPI network, it is

also possible to define a single-node entropy-like measure (in analogy with the Shannon entropy) for the  $i$ -th node. Since  $p_{ij} \geq 0 \forall (i, j)$ , and since the relation  $\sum_{j \neq i} p_{ij} = k_i$  is enforced for each node  $i$ , we can define an entropy-like measure  $S_i$  as follows:

$$S_i = - \sum_j p'_{ij} \log p'_{ij} \quad p'_{ij} = \frac{p_{ij}}{k_i} \quad (1.34)$$

Given the single node entropy values  $\{S_i\}$  for each sample, we checked by a non-parametric Wilcoxon rank sum test for significant differences at a single node level between the groups of our datasets. Since we know the KEGG annotation for each gene of our network, we also performed a functional analysis of specific biochemical pathways, based on enrichment analysis of pathways by genes significantly changing their single-node entropy value  $S_i$ . In this way the entropy analysis could be scaled from the full PPI network to single-node and single-pathway level.

Taking advantage of the a priori biological knowledge available from the KEGG database, we remark that it is indeed possible to obtain several subnetworks of the initial PPI network: at a first level, the genes annotated in the PPI can be divided into 6 functional groups, that can be further subdivided into 42 metapathways, and again into 191 KEGG biological pathways (see Fig. 1.3). We decided to apply our analysis at the pathway level, in order to gain more information on the single known biological mechanisms described into the KEGG database.

For the calculation of the entropy values, the link probabilities  $p_{ij}$  and the Lagrangian multipliers, we developed an iterative algorithm: given a random starting guess for the value of the lagrangian multipliers  $\{z_i\}$  and  $\{W_l\}$ , the  $p_{ij}$  values are calculated according to Eq. 1.33. These values are then substituted in the constraint equations 1.31 and 1.32 for the updating of the lagrangian multipli-

Table 1.1: Cancer datasets: median values of the network entropy groups  $S_1$  and  $S_2$  as pictured in Fig. 1.1(a), in a typical case-control design. With  $p_W$  we consider the p-value given by the Wilcoxon ran sum test.

	$S_1$	$S_2$	Size	$p_W$
Colon	13.0349	13.0680	312	$4.35 \cdot 10^{-4}$
Ewing	8.8057	8.7569	136	0.0023
Met	9.9483	9.9159	151	$4.12 \cdot 10^{-4}$
Rel	15.4700	15.4664	313	0.0197

ers. The process is repeated upon convergence. We checked by random sampling that the application of the iterative algorithm for different initial guesses leads to the same final entropy values (since under these constraints it is a convex function that admits an unique maximum). The threshold for the convergence of the algorithm was set to  $10^{-5}$ , and we remark that every significant change in entropy values was at least of a order of magnitude higher, thus the chosen precision is not affecting our results. This algorithm is available in Matlab code.

### 1.2.3 Results

#### Network entropy

The first analysis consisted in comparing the entropy values for the samples belonging to the different classes (see Figure 1.1). For the Colon dataset (Fig. 1.1, Panel *a*) we see a significant increase of network entropy  $S$  between normal and cancer samples ( $P = 0.00043$ ) when considering the selection of genes which expression profile differed between normal and cancer samples. At a full-network level, the same trend is observed, but the result is weakly non significant ( $P = 0.057$ ) We interpret this result as an increase in cell deregulation when passing from normal to cancer cell, reflected in a higher “phenotypic space” available, since many regulation mechanisms (eg. related to cell cycle, apoptosis or DNA

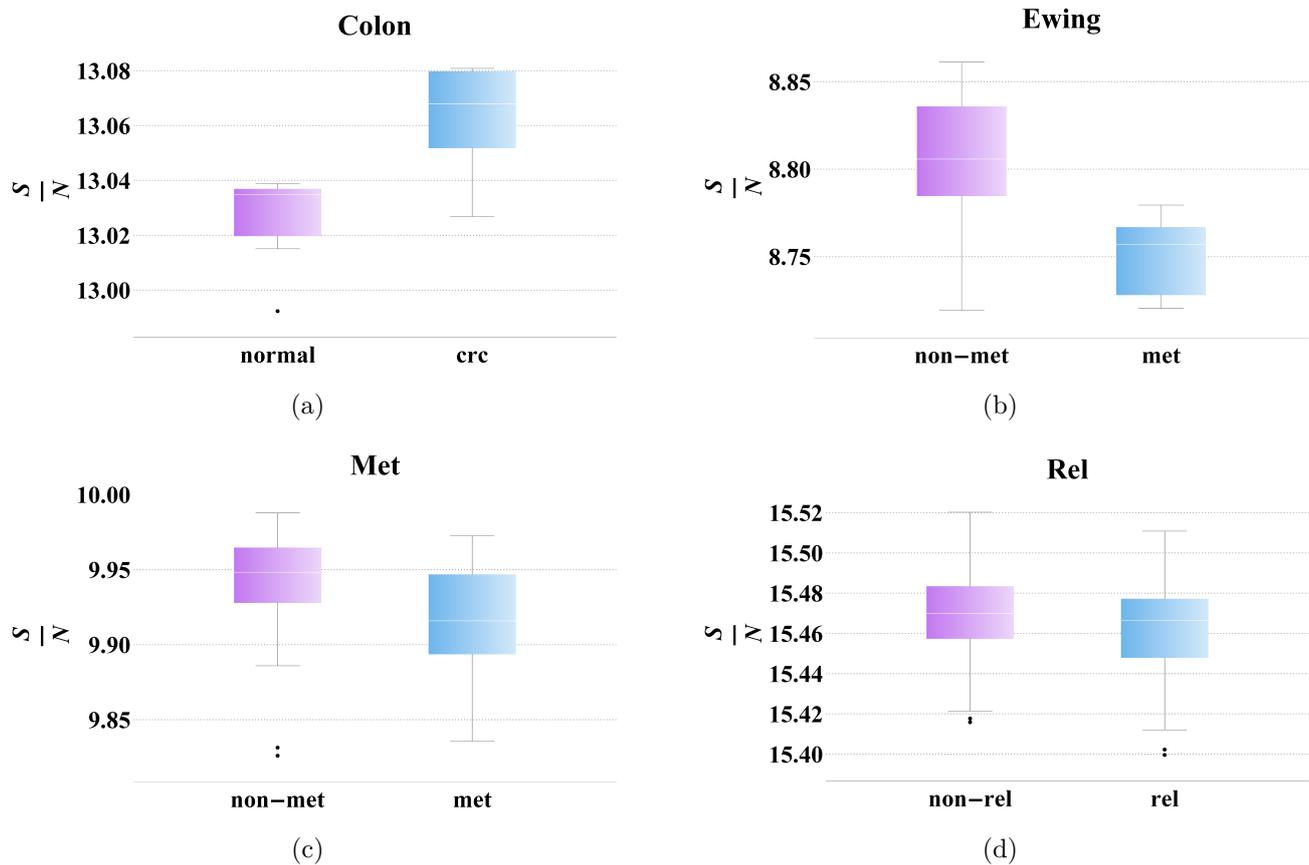


Figure 1.1: Boxplots for the network entropy values in the studied cancer databases. Panel a: colon cancer, normal vs. cancer samples. In this case cancer samples have a significantly higher entropy. Figure b, c, d: Ewing sarcoma, metastatic and relapsing breast cancer databases, respectively. In b, c, d cases a primary tumour samples are compared with tumour samples that relapsed or developed metastasis during disease progression. In these cases entropy has a significantly higher value in the primary tumour groups.

repair) are lost in a cancer cell (Hanahan and Weinberg, 2000).

If we consider single-node entropy, we find 665 genes (over 2835) with a significant difference between normal and cancer samples (see supplementary file). The single genes with highest significance are involved in known cancer-related pathways, such as “WNT”, “MAPK”, “Notch” and “Cell communication” pathways. The role of the genes which single-node entropy is differing significantly between normal and cancer samples can be better understood at a KEGG pathway level:

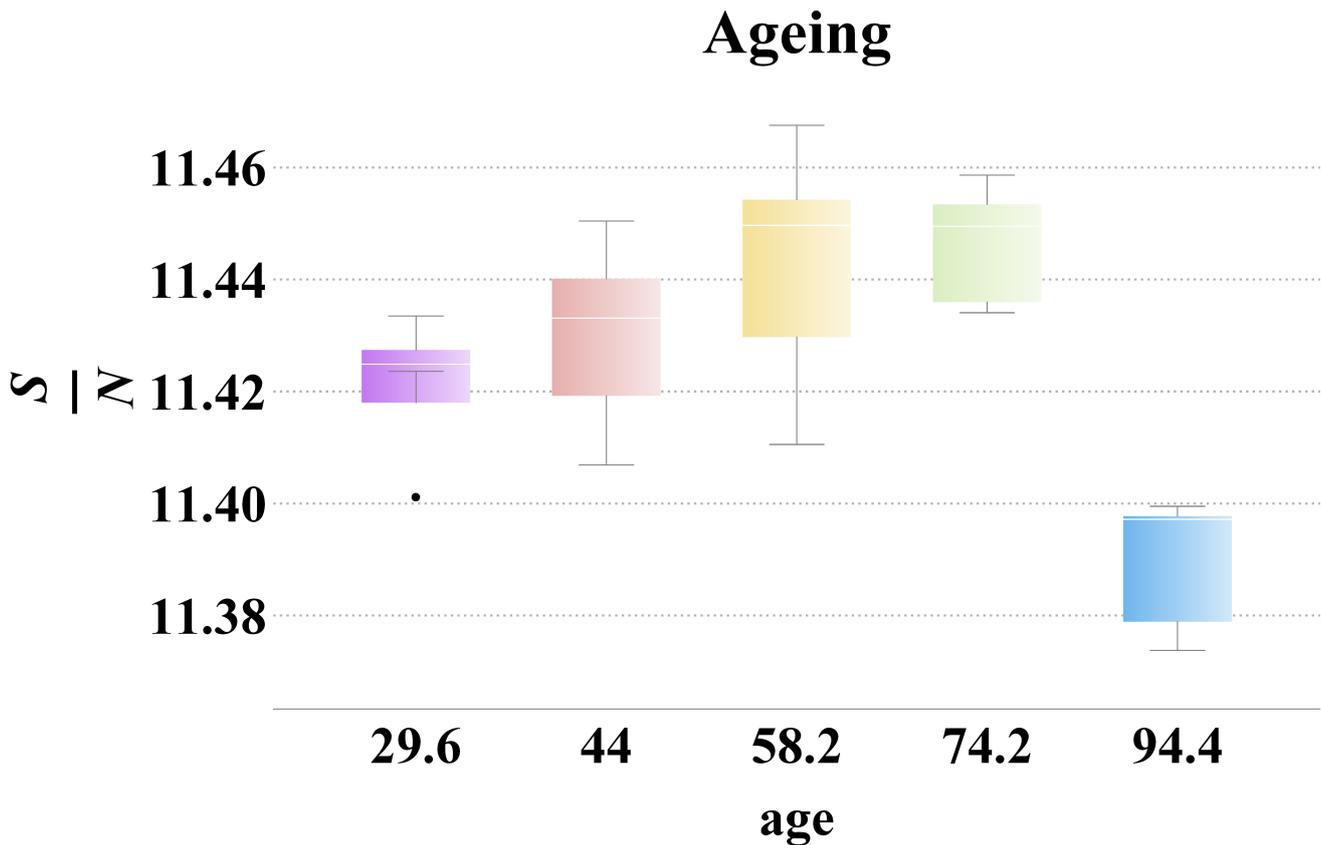


Figure 1.2: Boxplots for the network entropy values in the studied ageing database. Successfully aged people have a significantly lower entropy. We show the results for the Kruskal-Wallis test over the five age groups and for the Wilcoxon rank sum test for each pair of groups in Tab. 1.2.

an analysis based on the hypergeometric distribution (i.e. counting the number of genes with significant differences in single-node entropy for a particular pathway, given the total number of significant variations in the whole network) shows that 25 (over 191) pathways are significantly enriched ( $P < 0.05$ ), among which “Oxidative phosphorylation”, “Focal adhesion”, “TCA cycle”, “Cell communication”, “Apoptosis”, “Cell adhesion molecules” with a clear involvement in cancer progression both at a signalling and at a metabolic level (Hanahan and Weinberg, 2000).

For the other class of comparisons, between primary and secondary cancers

(metastatic or relapsing) we find instead a significant decrease in network entropy ( $P = 0.014$  for the Ewing dataset,  $P=0.00041$  for MET dataset,  $P=0.02$  for the REL dataset). In this case, the change from a primary cancer to a metastatic or relapsing state implies an evolutionary selection, since some specific steps need to occur, e.g. regarding epithelial-to-mesenchymal transition mechanisms (Brabletz and Brabletz, 2010) or adaptation to pharmacoresistance, or clonal selection induced by therapy. The reduction in “phenotypic” space is thus a measure of this phenomenon. In the Ewing dataset, 142 genes have a significant difference in single-node entropy  $S_i$  ( $P < 0.05$ ) between primary and metastatic samples, involved in many pathways, with a large majority of lipid metabolism pathway. A significance analysis at KEGG pathway level produces 33 significantly enriched pathways, such as “Glycolysis/Gluconeogenesis”, “Pentose phosphate”, “Galactose metabolism”, “Glycosphingolipid biosynthesis”, but also “Cell communication”, “Focal adhesion” and “ECM-receptor interaction” that might be involved in metastatic processes such as cell migration. In the MET dataset, 342 genes have a significant difference in single-node entropy. Even if the cell type is different (primary breast cancer) many pathways are the same as for the Ewing dataset, in particular related to the lipid metabolism. Functional analysis highlights 48 enriched pathways, among which “Glycolysis/Gluconeogenesis”, “Galactose metabolism”, “Glycosphingolipid biosynthesis” as for Ewing dataset, and also pathways such as “Cell adhesion molecules” that can be again related to metastatic progression. For the REL dataset, 331 genes had a significant difference in single-node entropy, and 23 pathways were functionally enriched with a  $P < 0.05$ . Among these pathways, some of them are related to metabolism (“Ether lipid biosynthesis”, “Biosynthesis of steroids”, “Pyrim-

Table 1.2: Ageing dataset: in the upper part of the table we show the median values for the five network entropy age groups as pictured in Fig. 1.2. With  $p_K$  we consider the p-value given by the Kruskal-Wallis test over the five age groups. In the lower table we show the results for the Wilcoxon rank sum test for each pair of groups.

$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$p_K$
11.4249	11.4331	11.4497	11.4495	11.3972	0.0028
$p_W$	group 1	group 2	group 3	group 4	group 5
group 1		0.5476	0.0952	0.0079	0.0079
group 2			0.4206	0.1508	0.0079
group 3				1	0.0079
group 4					0.0079

idine metabolism”), but also to specific functions such as “RNA polymerase”, “DNA polymerase”, “Proteasome”, “Cell adhesion molecules” and “Metabolism of xenobiotics by cytochrome P450”.

We remark that the pathways involved in a change in entropy, as shown above, are very different from the pathways that can be obtained by an identical functional analysis performed on genes with a significant change in gene expression (thus related to gene up or downregulation) reflecting the different information encoded in network entropy at whole-cell and single-node level (data not shown).

For the Ageing dataset, we exploited the time series design by applying a Kruskal-Wallis test over the age groups, in order to evaluate significant changes in network entropy over the whole life span. The trend for the five groups was significantly different ( $P=0.0028$ , see Fig. 1.2). In particular, among the 5 age groups a multiple testing by ranksum revealed that only the oldest age group showed a significantly different behaviour, with a lower Network Entropy than the other age groups. The last age group is related to successfully ageing people,

Table 1.3: Pathway analysis: number of significant genes and pathways based on the single node entropy variations. For the genes we applied a Wilcoxon rank sum test in the usual case-control setup. For the pathways we performed an enrichment analysis, highlighting those paths enriched by genes significantly changing their single-node entropy value.

	Significant genes	Significant pathways
Colon	665	25
Ewing	142	33
Met	342	48
Rel	331	23
Ageing	290	16

since their age is larger than average life expectancy, thus it represents a very selected group from an epidemiologic point of view. Its different value in network entropy could be explained in two ways, that our data do not allow to distinguish: first, the successfully ageing group represents a selection, in terms of phenotype, over the human population. Thus the reduced entropy highlights their peculiar expression profile. As a second hypothesis, the oldest group shows a smaller plasticity in terms of the possible phenotypic profiles that the cells can assume. This aspect can be related to the “frail” phenotype (Ferrucci et al., 2008; Fried et al., 2004), for which old people are less capable of adaptation, both from a psychological and from a physical point of view. For the single-node entropy and functional enrichment analysis we considered a comparison between the youngest and the oldest age group, representing the two extremes of our time series: a rank sum test found 290 (over 1976) genes with a significant difference in  $S_i$  ( $P < 0.05$ ). The KEGG pathways mostly enriched by significant genes are in part related to the specific cell type, ie lymphocytes (“T cell receptor signalling”, “B cell receptor signalling”, “hematopoietic cell lineage”), metabolic pathways (“Androgen and estrogen metabolism”, “Biotin metabolism”, “Histidine metabolism”), and path-

ways involved in cellular degradation/production machinery ("Proteasome"), in particular at the nucleolar level, such as "Ribosome" and "DNA polymerase" that are known to be altered during ageing (Bellavista et al., 2014; Lempiainen and Shore, 2009)

## 1.3 Conclusions

In this chapter we have introduced the main concepts of Statistical Mechanics of network ensembles. As expected, there are strong analogies with the usual statistical physics, especially considering the parallelism with fermionic and bosonic distributions (in the unweighted and weighted network cases respectively). On the other hand, the same formalism can be easily interpreted from the point of view of Information theory.

This approach defines correct and unbiased null models of networks, giving the chance to quantify which network features are peculiar to the studied system and which are simply due to randomness.

Moreover, we have shown how this approach provides observables for real data, such as the measure of network entropy that we applied to omics data, in a typical setup of Systems Biology.

In the presented biological study the measure of network entropy successfully integrates the topological information encoded in the protein interaction network with gene expression profiling. This measure is introduced to characterise different levels of cellular perturbation, namely the comparison between healthy and cancer samples, primary and metastatic cancer samples, and a time series of healthy samples with different ages across the whole human lifespan. This

measure estimates the number of networks that satisfy given constraints, in our case the degree sequence of the protein network and the distribution of the link distances as given by the difference in expression between genes, and can be interpreted as the extent of the “parameter space” allowed to the cell in a given state in terms of gene expression plasticity, or also in terms of different cell phenotypes (in terms of cell clonality for the case of cancer).

Different case studies help to clarify this interpretation. Regarding the comparison between healthy and cancer cells, we observe an increase of network entropy, possibly due to a larger deregulation of the biological mechanisms and functions involved or to an increase in cell phenotypical diversity. When we consider primary and metastatic (or relapsing) samples, network entropy shows a significant decrease instead, reflecting the canalisation or the evolution (in terms of clonal extent or gene expression profile) necessary to achieve this specific state. In a time series of ageing people, we see a sharp decrease of network entropy for the successful ageing group (with an age larger than typical life expectancy) that could also in this case represent a sort of selection of specific ageing phenotypes.

The formalism allows to define a measure of entropy at different scales, from single gene to biological pathways, that highlights how the changes in entropy are specific for the biological function and the experimental design (case-control, cell-type) considered. This method provides a different perspective on the analysis of gene expression data, integrating single-gene expression measurements and functional relationships between genes due to biological functions inside the cell. The entropy measure  $S$  seems an observable sensitive enough to evaluate the effect of physiological perturbations such as the changes occurring during the cellular ageing process, and also the differences between cancer subtypes before

the progression to metastatic and relapsing phenotypes. The statistical significance of  $S$  resulted independent on network properties, such as the number of nodes, and increased when a selected subset was considered, thus reflecting the biological relevance of the data used.

We were able to scale the analysis at different levels based on a priori biological knowledge (as obtained from KEGG database) so to apply the analysis to specific biological functions and pathways.

The approach can be generalised to other systems as well, considering different networks for the topological constraints, like transcription or metabolic networks, different high-throughput observables, for example methylation states or metabolic compounds, and finally considering different metrics, like correlation or mutual information, to define the weights of the network.

Metabolism	Genetic Information Processing	Environmental Information Processing	Cellular Processes	Organismal Systems	Human Diseases
Carbohydrate metabolism	Transcription	Membrane transport	Transport and catabolism	Immune system	Cancers
Energy metabolism	Translation	Signal transduction	Cell motility	Endocrine system	Immune diseases
Lipid metabolism	Folding, sorting and degradation	Signaling molecules and interaction	Cell growth and death	Circulatory system	Neurodegenerative diseases
Nucleotide metabolism	Replication and repair		Cell communication	Digestive system	Substance dependence
Amino acid metabolism				Excretory system	Cardiovascular diseases
Metabolism of other amino acids				Nervous system	Endocrine and metabolic diseases
Glycan biosynthesis and metabolism				Sensory system	Infectious diseases: Bacterial
Metabolism of cofactors and vitamins				Development	Infectious diseases: Viral
Metabolism of terpenoids and polyketides				Environmental adaptation	Infectious diseases: Parasitic
Biosynthesis of other secondary metabolites					
Xenobiotics biodegradation and metabolism					
Reaction module maps					
Chemical structure transformation maps					

Figure 1.3: KEGG database: the genes annotated in the PPI can be grouped following KEGG into 6 functional groups, further subdivided into 42 metapathways, and finally into 191 pathways (data not shown)

# 2 Multilayer networks and multiplex networks

## 2.1 Introduction to multilayer networks and multiplex networks

Network theory investigates the global topology and structural patterns of the interactions among the constituent elements of a number of complex systems including social groups, infrastructure and technological systems, the brain and biological networks (Albert and Barabási, 2002; Newman, 2003; Boccaletti et al., 2006; Fortunato, 2010). Over the last fifteen years, a large body of literature has attempted to disentangle noise and stochasticity from non-random patterns and mechanisms, in an attempt to gain a better understanding of how these systems function and evolve. Further advances in the study of complex systems has clarified that in order to understand the complexity of a large variety of systems is not enough to consider single networks, but it is necessary to describe the complex set of interactions between different networks by adopting the framework of *multilayer networks*. Multilayer networks are formed by a set  $M$  of layers constituted by single networks, and by interlinks connecting the nodes in the

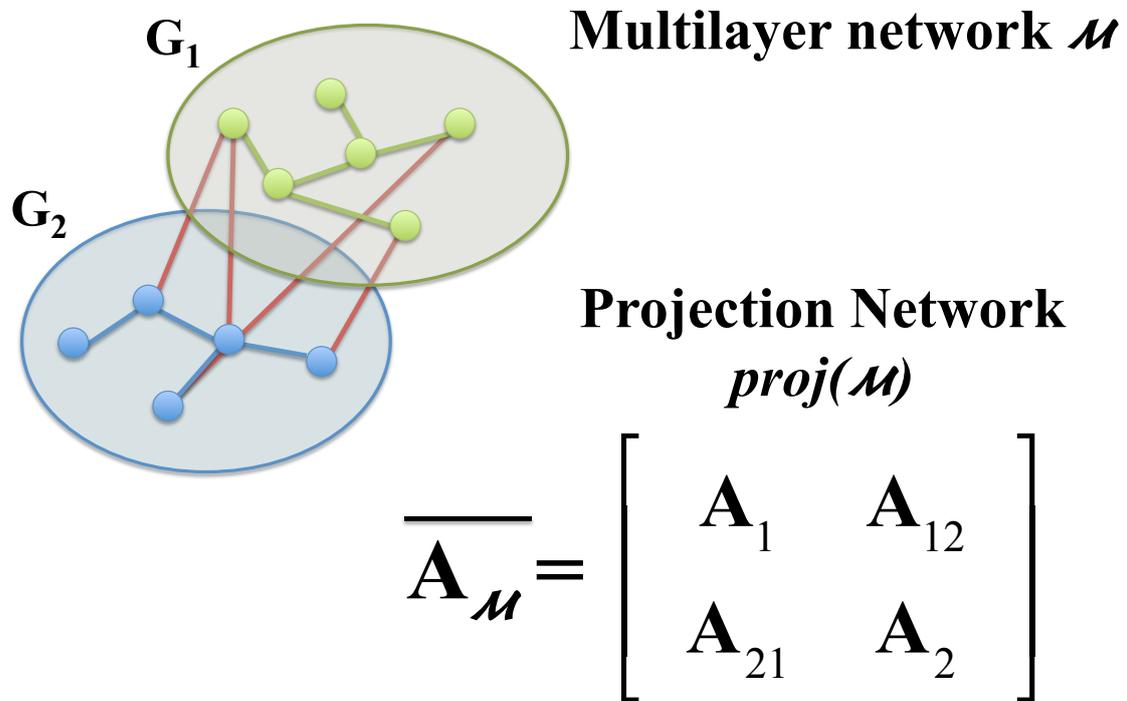


Figure 2.1: A representation of a generic multilayer network  $\mathcal{M} = (\mathcal{G}, \mathcal{C})$  composed by two graphs:  $G_1$  and  $G_2$ . The *interlayer* connections are in red while the *intra-layer* connections are in green for graph  $G_1$  and in blue for graph  $G_2$ . The adjacency matrix of the related projection network  $\text{proj}(\mathcal{M}) = (X_{\mathcal{M}}, E_{\mathcal{M}})$  is displayed in the lower-right corner.

different layers. Formally, a multilayer graph is described by a pair  $\mathcal{M} = (\mathcal{G}, \mathcal{C})$  where  $\mathcal{G} = \{G_\alpha; \alpha \in \{1, \dots, M\}\}$  is a set of graphs  $G_\alpha = (X_\alpha, E_\alpha)$  (called layers) and by

$$\mathcal{C} = \{E_{\alpha\beta} \subseteq X_\alpha \times X_\beta; \alpha, \beta \in \{1, \dots, M\}, \alpha \neq \beta\}$$

defining the set of interconnections between nodes of different  $G_\alpha$  and  $G_\beta$  ( $\alpha \neq \beta$ ). The elements of  $E_{\alpha\beta}$  ( $\alpha \neq \beta$ ) are called *interlayer* connections (see Fig. 2.1, red edges) while the elements of each  $E_\alpha$  are called *intralayer* connections (see Fig. 2.1, green edges for graph  $G_1$  and blue edges for graph  $G_2$ ). We denote by  $X_\alpha = \{x_1^\alpha, \dots, x_{N_\alpha}^\alpha\}$  the set of nodes of the layer  $G_\alpha$  and by  $A^\alpha = \{a_{ij}^\alpha\} \in \mathbb{R}^{N_\alpha \times N_\alpha}$ . Furthermore, associated with  $E_{\alpha\beta}$  we define a similar adjacency matrix  $A^{\alpha\beta} = \{a_{ij}^{\alpha\beta}\} \in \mathbb{R}^{N_\alpha \times N_\beta}$ . The *projection network*  $\text{proj}(\mathcal{M}) = (X_{\mathcal{M}}, E_{\mathcal{M}})$ , related to the multilayer  $\mathcal{M}$ , is given by

$$X_{\mathcal{M}} = \cup_{\alpha=1}^M X_\alpha \quad E_{\mathcal{M}} = \left( \cup_{\alpha=1}^M E_\alpha \right) \cup \left( \cup_{\alpha, \beta=1, \alpha \neq \beta}^M E_{\alpha\beta} \right)$$

Its adjacency matrix is indicated as  $\bar{A}_{\mathcal{M}}$  (see Fig. 2.1). In biological fields there are many interesting examples well modelled by multilayer networks. For example, the biological functionality of the cells can be described by a multilayer network involving at least metabolic, protein interaction and transcription network layers. Moreover, the so-called *systems medicine* seems naturally embedded in a multilayer network (see Fig. 2.2). The definition of systems medicine has been forged with the introduction of complex network methodology in biomedicine: it involves a systemic view of the organism where the various elements building living beings are considered in their interplay. Systems medicine looks at

multilayer networks as possible tools for combining the characterisation of the main constituents of the cell: genes, proteins and metabolites. Up to now, many different complex networks have been studied, e.g. gene-gene coexpression networks, protein-protein interaction networks, metabolite-metabolite coexpression network. Each one has been considered separately, not including the strong correlations and interdependencies with the other complex networks. The representation of the cell, and moreover of the living being, as interdependent layers may give a new insight about the appearance of systemic pathological conditions. Furthermore, focusing on the interdependencies among genes and proteins, we can build a multilayer network encoding both experimental setup (coexpression matrices from experimental profiles) and annotated reactions (protein-protein interaction network, Recon X). This structure naturally pictures the gene control upon the production of proteins, turning into catalysers of the metabolic reactions. Furthermore, a multivariate statistics and an integrated clustering can be performed.

Multilayer networks can be distinguished in multiplex networks (Szell et al., 2010; Cardillo et al., 2013; Nicosia and Latora, 2014; Donges et al., 2009) and interacting networks of networks (Gao et al., 2012; Bianconi and Dorogovtsev, 2014). In interacting networks of networks the nodes in the different layers represent different elements of the system. For example, in the cell, metabolites, proteins and transcription factors remain distinct biological entities. In a *multiplex network*, instead, the same set of nodes forms  $M$  networks, one in each layer corresponding to different types of interactions. Examples of multiplex networks include:

- **social networks:** the same individuals can be connected through different types of social ties originating from friendship, collaboration, or family

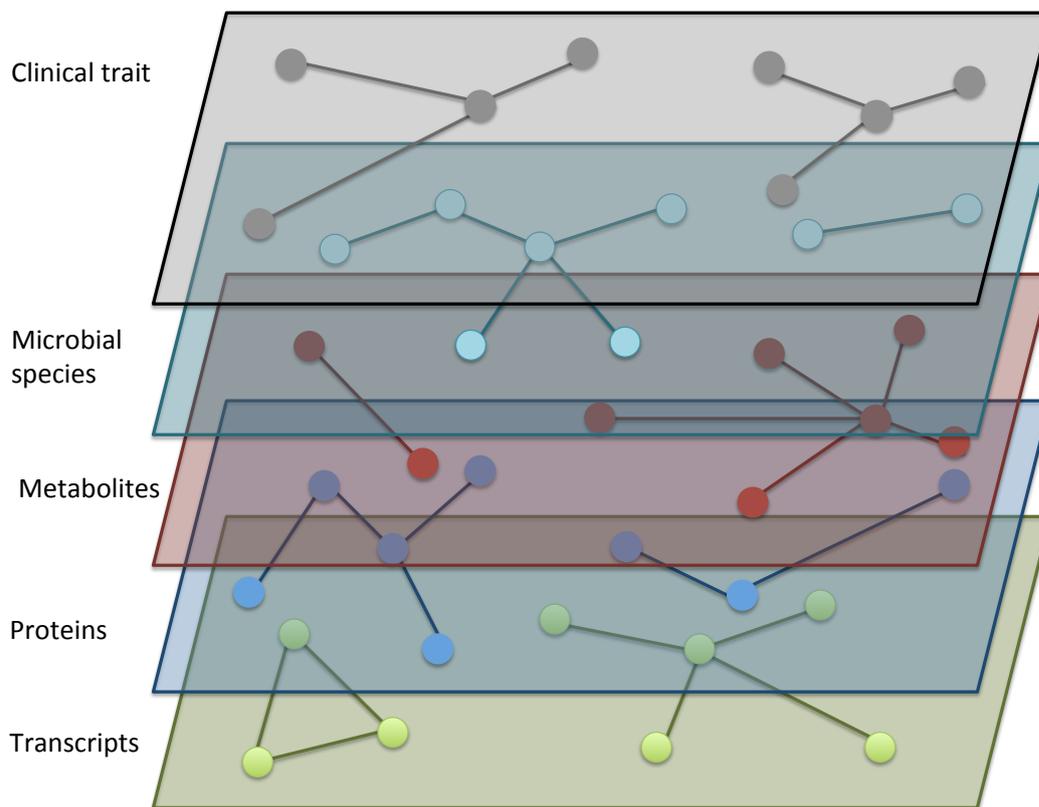


Figure 2.2: Systems medicine: a multilayer network point of view

relationships (Szell et al., 2010)

- **air transportation networks:** different airports can be connected through flights of different companies (Cardillo et al., 2013; Nicosia and Latora, 2014)
- **brain networks:** different regions can be seen as connected by the functional and structural neural networks (Bullmore and Sporns, 2009; Castellani et al., 2014)

Most of the studies so far conducted on multiplex networks have been concerned with the empirical analysis of a wide range of systems (Szell et al., 2010; Cardillo et al., 2013; Donges et al., 2009; Morris and Barthelemy, 2012), the

modeling of their underlying structures (Battiston et al., 2014; Halu et al., 2013; Mucha et al., 2010), and the description of new critical phenomena and processes occurring on them (Buldyrev et al., 2010; Baxter et al., 2012; Gómez et al., 2013; Brummitt et al., 2012). In particular it has been found that multiplex networks encode in their structure important correlations: we can distinguish for example between degree correlations (Min et al., 2014; Nicosia et al., 2014) determining whether a hub in a network is also an hub in another network, overlap determining to what extent any two nodes of the network are linked in several networks at the same time (Szell et al., 2010; Cardillo et al., 2013; Bianconi, 2013; Halu et al., 2014), or pairwise activity correlations measuring if the presence of a node in one network is correlated with the presence of another node in the same network (Nicosia and Latora, 2014). Many multiplex networks are also weighted, i.e. the links between the nodes not only are distinguished by the type of interaction linking the nodes, but also by the intensity of these interactions.

Despite this growing interest in multiplex networks, a fundamental question still remains largely unanswered: what is the advantage of a full-fledged analysis of complex systems that takes all their interacting layers into account, over more traditional studies that represent such systems as single networks with only one layer? To answer this question, one should demonstrate that novel and relevant information can be uncovered only by taking the multiplex nature of complex systems directly into account, and would instead remain undetected if individual layers were analysed in isolation. In the following, an attempt is made to offer a possible solution to this problem within the context of weighted multiplex networks, presenting two data-sets from the real world:

**American Physical Society (APS)** citation and collaborations networks from

different journals of the American Physical Society

**Gene-gene duplex** two gene-gene networks extracted using the gene expression of a pool of cancer patients and a pool of healthy subjects respectively for each layer

The results show how in these systems it is possible that the weights of the links are correlated with the pattern of overlap observed between the links of different layers. It is therefore very important, as previously explained in Ch. 1, to propose *maximum-entropy multiplex ensembles* with given constraints (Park and Newman, 2004; Bianconi, 2008; Bianconi et al., 2008; Anand and Bianconi, 2009; Annibale et al., 2009; Squartini et al., 2011; Garlaschelli, 2009; Squartini and Garlaschelli, 2011; Garlaschelli and Loffredo, 2009; Sagarra et al., 2013, 2014; Del Genio et al., 2010; Zlatic et al., 2009): these models can be used to generate multiplex networks with different types of correlations. These models, on one side can be used to simulate dynamical processes on different multiplex network topologies, on the other side, similarly to what happens for single networks, their entropy (Bianconi, 2008; Bianconi et al., 2008) can be used to evaluate the information content of some of their properties (see Sec. 1.2, Bianconi et al. (2009)). Here we provide the theoretical framework to generate null models for multiplex networks, using the combined tools of canonical network models (exponential random graphs) and the recently introduced concept (Bianconi, 2013) of multi-links, that is able to distinguish between different patterns of overlap of the links in the multiplex network.

This chapter is structured as follows: in Sec. 2.2 we give a general introduction to the main observables characterising multiplex networks; in Sec. 2.3 we present a large series of null models for uncorrelated and correlated multiplex networks,

both unweighted and weighted; in Sec. 2.4 we analyse two real datasets, i.e. the American Physical Society dataset and a biological case study.

## 2.2 Measures on multiplex networks

In a large variety of cases real multiplex networks show a significant overlap of the links through different layers, meaning that the number of links present at the same time in two layers or more is not negligible with respect to the total number of links in the different layers. This is one of the most important kind of correlation that we observe in multiplex networks, along with the deriving correlation in the node connectivity pattern through different layers. Social communications and interactions offer a natural landscape for this kind of correlation: we usually communicate with our friends in different ways, i.e. by phone-calls, by e-mail, by instant messaging. These means of communications are nothing more than different layers, different graphs with same nodes. If we pick , for example, the layer of phone calls and the layer of instant messaging, it is very likely to observe a high link overlap, and moreover, a non-trivial local overlap of links related to a given node. One way to fully characterised the link overlap is by the introduction of the so-called *multilinks*. The multilink formalism was introduced in Bianconi (2013) for unweighted multiplex networks. In this section we present the more general approach for weighted multiplex networks and we give a comparison with the usual single-layer measures. We define all the observables for weighted multiplex, considering the unweighted situation as a particular case.

A weighted multiplex is formed by  $N$  nodes connected by  $M$  weighted networks  $G_\alpha$ , with  $\alpha = 1, \dots, M$ . A multiplex can be represented as  $\vec{G} = (G_1, G_2, \dots, G_\alpha, \dots, G_M)$

where each network  $G_\alpha$  is fully described by the weighted adjacency matrix of elements  $a_{ij}^\alpha$ , with  $a_{ij}^\alpha > 0$  if there is a link of weight  $a_{ij}^\alpha$  between node  $i$  and node  $j$  in layer  $\alpha$ , otherwise we have  $a_{ij}^\alpha = 0$ .

As previously explained in Ch. 1, in order to simplify the treatment of the weighted multiplex, we suppose that the weight of the link between any pair of nodes  $(i, j)$ ,  $a_{ij}^\alpha$  can only assume integer values. This is a legitimate assumption because in a large number of weighted multiplexes the weights of the links can be considered as multiples of a minimal weight. Moreover, for the sake of simplicity we consider only networks without tadpoles and with a symmetric adjacency matrix  $\{a_{ij}^\alpha\}$ , i.e. undirected networks. The generalisation of our approach to directed multiplex networks is straightforward.

Since each layer of the multiplex is a weighted network, we can introduce the so-called *total strength*,  $S_\alpha$  that takes into account the total weight of the links in layer  $\alpha$ . The expression for  $S_\alpha$  is

$$S_\alpha = \sum_{i < j} a_{ij}^\alpha. \quad (2.1)$$

The *total number of links*  $L_\alpha$  for a specific layer  $\alpha$  is strictly related to Eq. 2.1, and is given by

$$L_\alpha = \sum_{i < j} \theta(a_{ij}^\alpha). \quad (2.2)$$

### 2.2.1 Single-layer observables

Each single layer  $\alpha$  of the multiplex network is a weighted network (Barrat et al., 2004; Almaas et al., 2004): for each layer we can characterise the topo-

logical quantities (such as the degree distribution) but also the heterogeneous interactions between the nodes. Interesting weights-topology correlations are usually a signature of the given network. These correlations can be revealed by measuring the following three quantities:

- the degree  $k_i^\alpha$  of a node  $i$  in layer  $\alpha$ ,
- the strength  $s_i^\alpha$  of node  $i$  in layer  $\alpha$ ;
- the inverse participation ratio  $Y_i^\alpha$  of node  $i$  in layer  $\alpha$ .

These quantities can be expressed in terms of the adjacency matrix elements respectively as

$$k_i^\alpha = \sum_{j \neq i} \theta(a_{ij}^\alpha), \quad (2.3)$$

where the function  $\theta(x) = 1$  if  $x > 0$  otherwise  $\theta(x) = 0$ ;

$$s_i^\alpha = \sum_{j \neq i} a_{ij}^\alpha, \quad (2.4)$$

and

$$Y_i^\alpha = \sum_{j \neq i} \left( \frac{a_{ij}^\alpha}{s_i^\alpha} \right)^2. \quad (2.5)$$

Moreover here we introduce for further convenience the quantity  $u_i^\alpha$

$$u_i^\alpha = Y_i^\alpha (s_i^\alpha)^2 = \sum_{j \neq i} (a_{ij}^\alpha)^2, \quad (2.6)$$

which indicates the sum of the squares of the weights incident to a node. Similarly to what happens for single networks (Barrat et al., 2004; Almaas et al., 2004), in any given layer  $\alpha$ , the strength  $s_i^\alpha$  of a node indicates the sum of the weights of the links of node  $i$  in layer  $\alpha$ , while the inverse participation ratio  $Y_i^\alpha$  indicates how unevenly the weights of the links of node  $i$  in layer  $\alpha$  are distributed. The inverse of  $Y_i^\alpha$  has a range between 1 and  $k_i^\alpha$ . The extremes of the interval correspond respectively to an uniform weight distribution across the links of the node  $i$  in the layer  $\alpha$ , i.e.  $a_{ij}^\alpha = s_i^\alpha/k_i^\alpha$ , that means  $(Y_i^\alpha)^{-1} = k_i^\alpha$ , and to the opposite situation, i.e.  $(Y_i^\alpha)^{-1} \approx 1$ , when one particular link of the node  $i$  has a prevailing weight, i.e.  $a_{ir}^\alpha \gg a_{ij}^\alpha$  for every  $j \neq r$ . In these terms  $Y_i^\alpha$  characterises the effective number of links of node  $i$  in layer  $\alpha$ .

It is a standard procedure in network theory to evaluate the averages of the strength and the partition ratio of the weights of the links conditioning on the degree of the node. In a multiplex, we will then consider the following quantities

$$\begin{aligned}
 s_\alpha(k) &= \langle s_i^\alpha \delta(k_i^\alpha, k) \rangle = \frac{1}{N_k^\alpha} \sum_i s_i^\alpha \delta(k_i^\alpha, k) \\
 Y_\alpha(k) &= \langle Y_i^\alpha \delta(k_i^\alpha, k) \rangle = \frac{1}{N_k^\alpha} \sum_i Y_{i,\alpha} \delta(k_i^\alpha, k)
 \end{aligned} \tag{2.7}$$

where  $N_k^\alpha$  indicates the number of nodes of degree  $k$  in layer  $\alpha$ . When considering  $s_k^\alpha$ , similarly to what happens in general on single networks, we can expect a scaling of the type

$$s_\alpha(k) \propto k^{\beta_\alpha}, \tag{2.8}$$

with  $\beta_\alpha \geq 1$ . We can distinguish Barrat et al. (2004) between two main scenarios depending on the value of the exponent. For  $\beta_\alpha = 1$  the average strength of

nodes of degree  $k$  increases linearly with  $k$ . This means that the average weight of the links incident to a node does not depend on the degree of the node, at least if we consider only distinguishable links ( for a treatment of the case of undistinguishable links see Sagarra et al. (2013, 2014)). For  $\beta_\alpha > 1$  hubs tend to have in average links with greater weight than low connectivity nodes. In a multiplex, we might have that the weights in the different layers are distributed differently. Therefore we might observe in some layers a superlinear growth of the  $s_\alpha(k)$  with the degree in that layer, while in other layers we can observe a linear dependence of the strengths on the degree. When considering single weighted networks it has been observed that in many cases the inverse participation ratio scales as an inverse power-law of the degree of the node (Almaas et al., 2004). In the multiplex scenario, this would imply

$$Y_\alpha(k) \propto \frac{1}{k^{\lambda_\alpha}}, \quad (2.9)$$

where the exponent  $\lambda_\alpha \leq 1$  might change from one layer to another layer. The exponent  $\lambda_\alpha = 1$  indicates that all the weights incident to any node are equal, while the exponent  $\lambda_\alpha = 0$  would imply the opposite scenario where for every node, one of the weights incident to them is significantly higher than the other weights.

### 2.2.2 Total overlap and total weighted overlap of the multiplex networks

In order to characterise the overlap existing between the links of the multiplex networks, we define the total overlap  $O^{\alpha,\alpha'}$  between layer  $\alpha$  and layer  $\alpha'$  as the

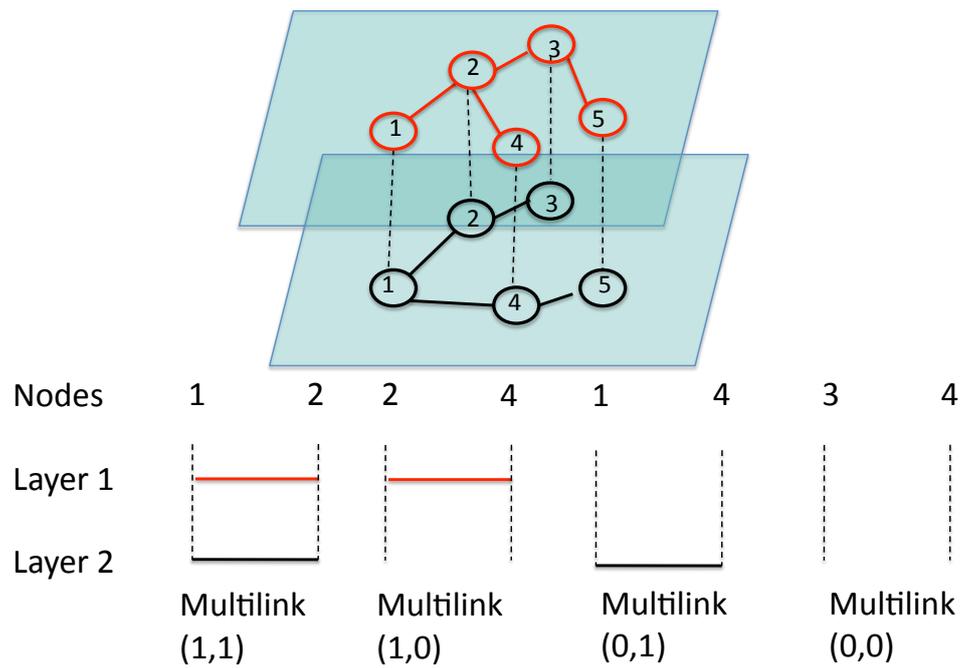


Figure 2.3: Schematic view of a duplex (multiplex formed by two networks) where any pair of nodes is linked by a different multilink  $\vec{m}$ .

total number of pair of nodes  $(i, j)$  connected both in layer  $\alpha$  and in layer  $\alpha'$ , i.e.,

$$O^{\alpha, \alpha'} = \sum_{i < j} \theta(a_{ij}^{\alpha}) \theta(a_{ij}^{\alpha'}), \quad (2.10)$$

where  $\theta(x) = 1$  if  $x > 0$  and  $\theta(x) = 0$  otherwise. This definition can be extended to weighted multiplex networks by defining the *total weighted overlap*  $O^{(w)\alpha, \alpha'}$  between layer  $\alpha$  and layer  $\alpha'$  as

$$O^{(w)\alpha, \alpha'} = \sum_{i < j} \min \left( \frac{w_{ij}^{\alpha}}{w_{max}^{\alpha}} \frac{w_{ij}^{\alpha'}}{w_{max}^{\alpha'}} \right), \quad (2.11)$$

where  $w_{max}^{\alpha}$  is the maximal weight in layer  $\alpha$ .

### 2.2.3 Multilink observables

It has been recently shown (Bianconi, 2013) that *multilinks* are the most natural way to describe and generate multiplex networks with overlap of the links. We say that two nodes are connected by a multilink  $\vec{m} = (m_1, m_2, \dots, m_{\alpha}, \dots, m_M)$  with  $m_{\alpha} = 0, 1$  if they are connected in every layer  $\alpha$  such that  $m_{\alpha} = 1$  and not connected in every layer  $\alpha$  where  $m_{\alpha} = 0$ . In figure 2.3 we show an example of a multiplex formed by two layers where each pair of node is linked by a given multilink. In order to indicate if a multilink  $\vec{m}$  is present or not between two given nodes  $i$  and  $j$  we can introduce a multiadjacency matrix  $\mathbf{A}^{\vec{m}}$  with elements  $A_{ij}^{\vec{m}}$  equal to 1 if there is a multilink  $\vec{m}$  between node  $i$  and node  $j$  and zero otherwise.

In terms of the weighted adjacency matrices  $\mathbf{a}^{\alpha}$  of the multiplex the elements  $A_{ij}^{\vec{m}}$

of the multiadjacency matrix  $\mathbf{A}^{\vec{m}}$  are given by

$$A_{ij}^{\vec{m}} = \prod_{\alpha=1}^M [\theta(a_{ij}^{\alpha})m_{\alpha} + (1 - \theta(a_{ij}^{\alpha}))(1 - m_{\alpha})] \quad (2.12)$$

where  $\theta(x) = 1$  if  $x > 0$ , otherwise  $\theta(x) = 0$ . The multilink  $\vec{m} = \vec{0}$  between two nodes represents the situation in which in all the layers of the multiplex the two nodes are not directly linked.

The multiadjacency matrices are  $2^M$  but there are only  $2^M - 1$  independent multiadjacency matrices because the normalisation condition

$$\sum_{\vec{m}} A_{ij}^{\vec{m}} = 1, \quad (2.13)$$

is satisfied for any pair of nodes  $(i, j)$ . Furthermore, since the multiadjacency matrices have elements  $A_{ij}^{\vec{m}} = 0, 1$ , the above condition implies that between any pair of nodes  $(i, j)$  there can be only one multilink  $\vec{m}$ . We indicate the type of this multilink as

$$\vec{m} = \vec{m}^{ij} = (\theta(a_{ij}^1), \theta(a_{ij}^2), \dots, \theta(a_{ij}^{\alpha}), \dots, \theta(a_{ij}^M)), \quad (2.14)$$

where  $\theta(x) = 1$  if  $x > 0$  and otherwise  $\theta(x) = 0$ . The multilink  $\vec{m}$  is characterised by the *overlap multiplicity*  $\nu(\vec{m}) = \sum_{\alpha} m_{\alpha}$  indicating that the multilink  $\vec{m}$  links two pair of nodes by  $\nu(\vec{m})$  links. Using the multiadjacency matrices it is possible to define the *multidegree*  $\vec{m}$ ,  $k_i^{\vec{m}}$  of node  $i$ , given by

$$k_i^{\vec{m}} = \sum_{j \neq i} A_{ij}^{\vec{m}}, \quad (2.15)$$

indicating how many multilinks  $\vec{m}$  are connected to node  $i$ . Consider for example the social multiplex network where people interact by two means of communication (mobile-phone, email). The multidegree  $k_i^{(1,1)}$  indicates the number of friends of node  $i$  that communicate with node  $i$  both by email and mobile phone,  $k_i^{(1,0)}$  indicates the number of friends of node  $i$  that only communicate with node  $i$  by mobile-phone and  $k_i^{(0,1)}$  indicates the number of friends of node  $i$  that only communicate with node  $i$  by email. Moreover, we define also the more global  $L^{\vec{m}}$ , i.e. the *total number of multilinks  $\vec{m}$*

$$L^{\vec{m}} = \sum_{i < j} A_{ij}^{\vec{m}}, \quad (2.16)$$

For a given weighted multiplex network we can study the relation between weights and multilinks introducing, at first, the *total multistrength  $\vec{m}$* ,  $S_{\alpha}^{\vec{m}}$  in a layer  $\alpha$  such that  $m_{\alpha} > 0$  as

$$S_{\alpha}^{\vec{m}} = \sum_{i < j} a_{ij}^{\alpha} A_{ij}^{\vec{m}}. \quad (2.17)$$

Given a particular multilink  $\vec{m}$ , this quantity indicates the total weight in layer  $\alpha$  of multilinks  $\vec{m}$  and it is properly defined whenever  $m_{\alpha} > 0$ . The number of total multistrengths  $\vec{m}$  that we can define in a multiplex of  $M$  layers is given by  $K = M2^{M-1}$ . In fact we have that the total multistrength  $S_{\alpha}^{\vec{m}}$  is non-trivial only for multilinks  $\vec{m}$  where  $m_{\alpha} = 1$ , while for the remaining layers  $\beta$  the value of  $m_{\beta}$  can be either zero or one.

Moreover we can define the *multistrength  $\vec{m}$* ,  $s_{i,\alpha}^{\vec{m}}$  of node  $i$  in layer  $\alpha$  such that

$m_\alpha > 0$ , as

$$s_{i,\alpha}^{\vec{m}} = \sum_{j \neq i} a_{ij}^\alpha A_{ij}^{\vec{m}} \quad (2.18)$$

and the *inverse multi participation ratio*  $\vec{m}$ ,  $Y_{i,\alpha}^{\vec{m}}$  of node  $i$  in layer  $\alpha$  such that  $m_\alpha > 0$  as

$$Y_{i,\alpha}^{\vec{m}} = \sum_{j \neq i} \left( \frac{a_{ij}^\alpha A_{ij}^{\vec{m}}}{\sum_r a_{ir}^\alpha A_{ir}^{\vec{m}}} \right)^2. \quad (2.19)$$

Using the same argument used to evaluate the number of total multistrengths  $\vec{m}$ , it is easy to prove that the number of local multistrength  $\vec{m}$  and the number of multi participation ratio  $\vec{m}$  are given by  $NM2^{M-1}$ . Moreover here we introduce  $u_i^{\alpha, \vec{m}}$ , the sum of the squares of the weights incident to a node  $i$  in layer  $\alpha$  and belonging to a certain type of multilink, as

$$u_{i,\alpha}^{\vec{m}} = Y_{i,\alpha}^{\vec{m}} (s_{i,\alpha}^{\vec{m}})^2 = \sum_{j \neq i} (a_{ij}^\alpha A_{ij}^{\vec{m}})^2. \quad (2.20)$$

In multiplex weighted networks, it was found that multistrengths and inverse multi partition ratio can have a different scaling behavior depending on the type of multilink. In fact the average quantities

$$\begin{aligned} s_\alpha^{\vec{m}}(k^{\vec{m}}) &= \left\langle s_i^{\alpha, \vec{m}} \delta(k_i^{\vec{m}}, k^{\vec{m}}) \right\rangle \\ Y_\alpha^{\vec{m}}(k^{\vec{m}}) &= \left\langle Y_i^{\alpha, \vec{m}} \delta(k_i^{\alpha, \vec{m}}, k^{\vec{m}}) \right\rangle \end{aligned} \quad (2.21)$$

are expected to scale like

$$\begin{aligned} s_{\alpha}^{\vec{m}}(k^{\vec{m}}) &\propto (k^{\vec{m}})^{\beta_{\alpha,\vec{m}}}, \\ Y_{\alpha}^{\vec{m}}(k^{\vec{m}}) &\propto (k^{\vec{m}})^{-\lambda_{\alpha,\vec{m}}} \end{aligned} \quad (2.22)$$

with  $\beta_{\alpha,\vec{m}} \geq 1$  and positive  $\lambda_{\alpha,\vec{m}} \leq 1$ . The significance dependence of these exponents as a function of the multilink type  $\vec{m}$ , i.e. on the presence of a certain pattern of overlap or absence of it, indicates the rich interplay between the topology of the weighted networks and their weights. For example in the CoCi-PRE duplex described in Sec. 2.4.1, formed by authors of PRE that in one layer are connected by collaborations and on the other layer are connected by citations of each other work, the weight-topology correlation is revealed by the different exponent of the multistrength in the citation network calculated either in presence of the overlap of the links in the two layers or in absence of it. This reveals the tendency of scientific authors of PRE to cite more the scientists of high multidegree that are their co-authors than the scientists with the same multidegree that are not their co-authors. These correlations between weights and overlap patterns are a very general type of correlation likely to exist in large set of multiplex dataset with significant overlap of the links. It is therefore very important to be able to construct null models for multiplex networks with the desired level of correlations between weights and overlap of the links, i.e. with given weighted properties of the multilinks.

### 2.2.4 Overlap multiplicity $\nu$

Using multilinks  $\vec{m}$  can be numerically viable only for weighted multiplex networks with a number  $M$  of layers such that  $M \ll \log(N)$ . As long as this condition is not met, it is more efficient to study the properties of the  $\nu$ -multilinks. The  $\nu$ -multilinks are only characterised by their overlap multiplicity  $\nu$ , i.e. the  $\nu$ -multilinks are all the multilinks that connects two nodes of the multiplex with  $\nu$  links in  $\nu$  different layers. Therefore in a multiplex social networks, where the layers correspond to the means of communication between two people, node  $i$  and node  $j$  are linked by a  $\nu$ -multilink if they can communicate by a maximum of  $\nu$  means of communication, independently on the identity of these. For example two people that communicate in Twitter and Facebook are linked by a  $\nu$ -multilink with  $\nu = 2$ , and the same is true for two people interacting by mobile phone and email.

We can therefore define the  $\nu$ -multiadjacency matrices  $\mathbf{A}^\nu$  with elements  $A_{ij}^\nu = 0, 1$  given by

$$\begin{aligned} A_{ij}^\nu &= \sum_{\vec{m} | \nu(\vec{m}) = \nu} A_{ij}^{\vec{m}} \\ &= \sum_{\vec{m} | \nu(\vec{m}) = \nu} \prod_{\alpha=1}^M [\theta(a_{ij}^\alpha) m_\alpha + (1 - \theta(a_{ij}^\alpha))(1 - m_\alpha)], \end{aligned} \quad (2.23)$$

and  $\nu = 0, 1, 2, \dots, M$ . The  $\nu$ -adjacency matrices are not all independent, since between any two nodes there can be just one type of  $\nu$ -mutlilink, i.e.

$$\sum_{\nu=0}^M A_{ij}^\nu = 1. \quad (2.24)$$

Therefore we can consider as independent variables only the  $\nu$ -adjacency matrices

corresponding to the non trivial  $\nu$ -multilinks with  $\nu = 1, 2 \dots, M$ . Moreover we call with  $\nu^{ij}$  the type of  $\nu$ -multilink connecting node  $i$  with node  $j$ , i.e. we have

$$A_{ij}^{\nu^{ij}} = 1 \quad (2.25)$$

for all pairs of nodes  $(i, j)$ . The number of distinct and non trivial  $\nu$ -multilinks with  $\nu \neq 0$  is given by  $M$ , hence the  $\nu$ -properties of the networks are only polynomial with  $M$  while the full multilink properties are growing exponentially with  $M$ . Modelling networks with given  $\nu$ -multilinks properties is therefore convenient when considering multiplex networks with large number of layers  $M$ . Given the definition of  $\nu$ -multiadjacency matrices it is straightforward to define the  $\nu$ -multidegree  $k_i^\nu$  of node  $i$ , given by

$$k_i^\nu = \sum_{j=1}^N A_{ij}^\nu \quad (2.26)$$

indicating the number of neighbors of node  $i$  that are connected to node  $i$  by a  $\nu$ -multilink, with  $\nu = 0, 1, 2 \dots, M$ . The total number of  $\nu$ -multilinks is trivially given by

$$L^\nu = \sum_{i < j} A_{ij}^\nu \quad (2.27)$$

If we consider the weighted properties of the  $\nu$ -multilink for a given layer  $\alpha$ , we can define the  $\nu$ -total strength  $S_\alpha^\nu$ , the  $\nu$ -multistrength sequence  $\{s_{i,\alpha}^\nu\}$  and the

$\nu$ -inverse multi participation ratio  $\{Y_{i,\alpha}^\nu\}$ , as in the following,

$$S_\alpha^\nu = \sum_{i < j} a_{ij}^\alpha A_{ij}^\nu \quad (2.28)$$

$$s_{i,\alpha}^\nu = \sum_{j \neq i} a_{ij}^\alpha A_{ij}^\nu$$

$$Y_{i,\alpha}^\nu = \sum_{j \neq i} \left( \frac{a_{ij}^\alpha A_{ij}^\nu}{\sum_r a_{ir}^\alpha A_{ir}^\nu} \right)^2. \quad (2.29)$$

Moreover, we can introduce the quantities  $u_i^{\alpha,\nu}$ , indicating the sum of the squares of the weights incident to a node  $i$  in layer  $\alpha$  and belonging to a certain type of  $\nu$ -multilink, as

$$u_{i,\alpha}^\nu = Y_{i,\alpha}^\nu (s_{i,\alpha}^\nu)^2 = \sum_{j \neq i} (a_{ij}^\alpha A_{ij}^\nu)^2. \quad (2.30)$$

Similarly to what described in the previous paragraph, we can evaluate the correlations between the weights and the pattern of overlap between the links by measuring the exponents  $\beta_{\alpha,\nu}$  and  $\xi_{\alpha,\nu}$ , determining the scaling

$$\begin{aligned} s_\alpha^\nu(k^\nu) &\propto (k^\nu)^{\beta_{\alpha,\nu}}, \\ Y_\alpha^\nu(k^\nu) &\propto (k^\nu)^{-\xi_{\alpha,\nu}} \end{aligned} \quad (2.31)$$

of the average quantities  $s_\alpha^\nu(k^\nu)$  and  $Y_\alpha^\nu(k^\nu)$  given by

$$\begin{aligned} s_\alpha^\nu(k^\nu) &= \langle s_i^{\alpha,\nu} \delta(k_i^\nu, k^\nu) \rangle \\ Y_\alpha^\nu(k^\nu) &= \langle Y_i^{\alpha,\nu} \delta(k_i^{\alpha,\nu}, k^\nu) \rangle. \end{aligned} \quad (2.32)$$

## 2.3 Canonical weighted multiplexes ensembles or exponential weighted multiplexes

Null models for weighted multiplex networks can be constructed using the formalism of canonical network ensembles also known as exponential random graphs (Park and Newman, 2004; Anand and Bianconi, 2009; Squartini et al., 2011). These ensembles of networks generate the least biased set of networks satisfying a set of constraint on average. In fact, these ensembles are derived by a maximal entropy approach conditioned to a series of structural constraints. The entropy of these ensembles and of the correspondent microcanonical ensembles enforcing the corresponding hard constraints (Bianconi, 2008; Bianconi et al., 2009), can be used to quantify the level of information encoded in the structural constraints that are imposed to the networks. In Bianconi (2013); Halu et al. (2014) this approach was taken to model simple multiplex networks. Here we show how this framework can be applied to model weighted multiplex networks.

A weighted multiplex ensemble is defined once the probability  $P(\vec{G})$  of any possible weighted multiplex is given. We can build a canonical multiplex ensemble by maximizing the entropy  $\mathcal{S}$  of the ensemble given by

$$\mathcal{S} = - \sum_{\vec{G}} P(\vec{G}) \log P(\vec{G}) \quad (2.33)$$

under the condition that the soft constraints we want to impose are satisfied. We assume to have  $K$  of such constraints determined by the conditions

$$\sum_{\vec{G}} P(\vec{G}) F_{\mu}(\vec{G}) = C_{\mu} \quad (2.34)$$

for  $\mu = 1, 2, \dots, K$ , where  $F_\mu(\vec{G})$  determines one of the structural constraints that we want to impose to the multiplex. Therefore, the maximal-entropy multiplex ensemble satisfying the constraints given by Eqs. 2.34 is the solution of the following system of equations

$$\frac{\partial}{\partial P(\vec{G})} \left[ \mathcal{S} - \sum_{\mu=1}^K \lambda_\mu \sum_{\vec{G}} F_\mu(\vec{G}) P(\vec{G}) - \Lambda \sum_{\vec{G}} P(\vec{G}) \right] = 0, \quad (2.35)$$

where the Lagrangian multiplier  $\Lambda$  enforces the normalisation of the  $P(\vec{G})$  probability distribution, and the Lagrangian multiplier  $\lambda_\mu$  enforces the constraint  $\mu$ . Therefore we get that the probability of a multiplex  $P(\vec{G})$  in a canonical multiplex ensemble is given by

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{\mu} \lambda_\mu F_\mu(\vec{G}) \right] \quad (2.36)$$

where the normalisation constant  $Z = \exp(1+\Lambda)$  is called the “partition function” of the canonical multiplex ensemble and is fixed by the normalisation condition on  $P(\vec{G})$ . The values of the Lagrangian multipliers  $\lambda_\mu$  are determined by imposing the constraints given by Eq. 2.34, assuming for the probability  $P(\vec{G})$  the structural form given by Eq. 2.36. From the definition of the partition function  $Z$  and Eq. 2.36, it can be easily shown that the Lagrangian multipliers  $\lambda_\mu$  can be expressed as the solutions of the following set of equations,

$$C_\mu = - \frac{\partial \log Z}{\partial \lambda_\mu}. \quad (2.37)$$

We call the entropy  $\mathcal{S}$  of the canonical multiplex ensemble the *Shannon entropy* of the ensemble.

Further on, we can define the marginal probability for a specific value of the element  $a_{ij}^\alpha$  as

$$\pi_{ij}^\alpha(a_{ij}^\alpha = w) = \sum_{\vec{G}} P(\vec{G}) \delta(a_{ij}^\alpha, w) \quad (2.38)$$

where  $\delta(x, y)$  stands for the Kronecker delta. The marginal probabilities  $\pi_{ij}^\alpha(a_{ij}^\alpha)$  sum up to one

$$\sum_{a_{ij}^\alpha=0}^{\infty} \pi_{ij}^\alpha(a_{ij}^\alpha) = 1 \quad (2.39)$$

We can compute also the average weight  $\langle a_{ij}^\alpha \rangle$  between node  $i$  and node  $j$  that is

$$\langle a_{ij}^\alpha \rangle = \sum_{\vec{G}} P(\vec{G}) a_{ij}^\alpha = \sum_{a_{ij}^\alpha=0}^{\infty} a_{ij}^\alpha \pi_{ij}^\alpha(a_{ij}^\alpha) \quad (2.40)$$

In the layer  $\alpha$  a link between two nodes  $i$  and  $j$  exists with probability  $p_{ij}^\alpha$ , that is related with all the possible weights different from zero

$$p_{ij}^\alpha = \sum_{\vec{G}} P(\vec{G}) \theta(a_{ij}^\alpha) = \sum_{a_{ij}^\alpha \neq 0}^{\infty} \pi_{ij}^\alpha(a_{ij}^\alpha). \quad (2.41)$$

### 2.3.1 Uncorrelated and correlated canonical multiplex ensembles

The multiplex ensembles can be distinguished between uncorrelated and correlated multiplex ensembles. For uncorrelated multiplex ensembles, the probability of a multiplex  $P(\vec{G})$  is factorizable into the probability  $P_\alpha(G_\alpha)$  of each single network  $G_\alpha$  at layer  $\alpha$ , i.e.

$$P(\vec{G}) = \prod_{\alpha=1}^M P_\alpha(G_\alpha). \quad (2.42)$$

Therefore, the entropy  $\mathcal{S}$  of any uncorrelated multiplex ensemble given by Eq. 2.33 with  $P(\vec{G})$  given by Eq. 2.42 is additive in the number of layers, i.e.

$$\mathcal{S} = \sum_{\alpha=1}^M \mathcal{S}_\alpha = - \sum_{\alpha=1}^M \sum_{G_\alpha} P_\alpha(G_\alpha) \log P_\alpha(G_\alpha) \quad (2.43)$$

As a consequence of these relations, when each constraint depends on a single network  $G_\alpha$  in a layer  $\alpha$  the resulting multiplex ensemble is uncorrelated.

Example of these types of constraints are the total strengths  $S^\alpha$  in each layer  $\alpha$ , the strength  $s_i^\alpha$  of the generic node  $i$  in layer  $\alpha$ , or the degree  $k_i^\alpha$  of the node  $i$  in layer  $\alpha$ .

In these ensembles of multiplex networks we have that the presence of a link in a layer  $\alpha$  is uncorrelated with the presence of a link between the same two nodes in a layer  $\beta \neq \alpha$ . Therefore we have

$$\langle a_{ij}^\alpha a_{ij}^\beta \rangle = \langle a_{ij}^\alpha \rangle \langle a_{ij}^\beta \rangle. \quad (2.44)$$

In correlated multiplex networks, instead the probability of a multiplex does not factorize into the probabilities of the single networks that constitute the multiplex network. We have in this case

$$P(\vec{G}) \neq \prod_{\alpha=1}^M P_\alpha(G_\alpha). \quad (2.45)$$

and as a consequence of this there is at least a pair of nodes  $(i, j)$  and layers  $\alpha, \beta$  such that the weights of the links connecting node  $i$  and node  $j$  is layer  $\alpha$  and

layer  $\beta$  are correlated, i.e.

$$\langle a_{ij}^\alpha a_{ij}^\beta \rangle \neq \langle a_{ij}^\alpha \rangle \langle a_{ij}^\beta \rangle. \quad (2.46)$$

Example of constraints that generate correlated multiplex ensembles are constraints on the multidegree sequence or the multistrength sequence.

In the following we provide many examples of uncorrelated and correlated multiplex network ensemble. In particular, the first section (Sec. 2.3.2) is completely dedicated to unweighted multiplex ensembles, a necessary precondition to weighted multiplex network ensembles.

### 2.3.2 Unweighted multiplex

We summarise here the main results related to unweighted multiplex ensembles. Only for this section we consider  $N$  nodes connected by  $M$  unweighted networks  $G_\alpha$ . Each network  $G_\alpha$  is fully described by its adjacency matrix of elements  $a_{ij}^\alpha$ , with  $a_{ij}^\alpha = 1$  if there is a link between node  $i$  and node  $j$  in layer  $\alpha$ , otherwise we have  $a_{ij}^\alpha = 0$ .

#### Examples of uncorrelated unweighted multiplex ensembles

##### Multiplex ensembles with given expected total number of links in each layer

As a first example of uncorrelated multiplex, we consider the case in which we fix the average number of links in each layer  $\alpha$  to be equal to  $L^\alpha$ . We have  $K = M$  constraints in the system, indicated with a label  $\alpha = 1, 2, \dots, M$ . These

constraints are given by

$$\sum_{\vec{G}} F_\alpha(\vec{G}) P(\vec{G}) = \sum_{\vec{G}} \left( \sum_{i < j} a_{ij}^\alpha \right) P(\vec{G}) = L^\alpha. \quad (2.47)$$

The probability distribution of a multiplex in this ensemble is given by Eq. 2.36 that reads in this case,

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{\alpha=1}^M \kappa_\alpha \sum_{i < j} a_{ij}^\alpha \right], \quad (2.48)$$

where the partition function  $Z$  can be expressed explicitly as

$$\begin{aligned} Z &= \sum_{\vec{G}} \exp \left[ - \sum_{\alpha=1}^M \kappa_\alpha \sum_{i < j} a_{ij}^\alpha \right] \\ &= \prod_{\alpha=1}^M \left[ (1 + e^{-\kappa_\alpha})^{\binom{N}{2}} \right]. \end{aligned} \quad (2.49)$$

The Lagrangian multipliers  $\kappa_\alpha$  defining the probability of the multiplex  $P(\vec{G})$ , are fixed by the conditions

$$L^\alpha = - \frac{\partial \log Z}{\partial \kappa_\alpha} = \binom{N}{2} \frac{e^{-\kappa_\alpha}}{1 + e^{-\kappa_\alpha}}. \quad (2.50)$$

For unweighted multiplex network the probability of having a link between node  $i$  and node  $j$  in layer  $\alpha$ ,  $p_{ij}^\alpha$ , is simply equal to  $\langle a_{ij}^\alpha \rangle$ , i.e.,

$$p_{ij}^\alpha = p^\alpha = \frac{L^\alpha}{\binom{N}{2}} = \frac{e^{-\kappa_\alpha}}{1 + e^{-\kappa_\alpha}} \quad (2.51)$$

Fixed  $\{p_{ij}^\alpha\}$ , the probability of a multiplex in this ensemble follows Eq. 2.42, and is given by

$$P(\vec{G}) = \prod_{\alpha=1}^M \prod_{i<j} (p_{ij}^\alpha)^{a_{ij}^\alpha} (1 - p_{ij}^\alpha)^{1-a_{ij}^\alpha} \quad (2.52)$$

$$= \prod_{\alpha=1}^M (p^\alpha)^{\ell^\alpha} (1 - p^\alpha)^{\binom{N}{2} - \ell^\alpha} \quad (2.53)$$

in agreement with Eq. 2.41 that we have previously met in Ch. 1. With  $\ell^\alpha$  we express the desired number of links for each layer  $\alpha$ . The last expression properly normalises if we take into account the multiplicity of a generic graph with given  $\ell^\alpha$ , i.e.

$$\sum_{\{a_{ij}^\alpha\}} \prod_{\alpha=1}^M \prod_{i<j} (p_{ij}^\alpha)^{a_{ij}^\alpha} (1 - p_{ij}^\alpha)^{1-a_{ij}^\alpha} = \sum_{\{\ell^\alpha\}} \prod_{\alpha=1}^M \binom{N(N-1)/2}{\ell^\alpha} (p^\alpha)^{\ell^\alpha} (1 - p^\alpha)^{\binom{N}{2} - \ell^\alpha} = 1 \quad (2.54)$$

In this uncorrelated ensemble the probability  $P(\vec{G})$  is nothing more than the product of  $M$  binomial probability distributions. The entropy  $\mathcal{S}$  of this canonical multiplex ensemble is given by Eq. 2.33 that can be rearranged in

$$\mathcal{S} = - \sum_{\alpha=1}^M \sum_{i<j} p_{ij}^\alpha \log p_{ij}^\alpha + (1 - p_{ij}^\alpha) \log (1 - p_{ij}^\alpha) \quad (2.55)$$

Using the marginals  $p_{ij}^\alpha$  given by Eq. 2.51 we simplify the previous formula as

$$\mathcal{S} = - \binom{N}{2} \sum_{\alpha=1}^M [p^\alpha \log p^\alpha + (1 - p^\alpha) \log (1 - p^\alpha)] \quad (2.56)$$

If the number of layers  $M$  is finite, applying the Stirling's approximation in the large  $N$  limit we get

$$\mathcal{S} = \sum_{\alpha=1}^M \log \left[ \binom{N(N-1)/2}{L^\alpha} \right]. \quad (2.57)$$

where Eq. 2.51 was used.

It is instructive to calculate the average global overlap  $\langle O^{\alpha,\alpha'} \rangle$  between two layers  $\alpha$  and  $\alpha'$ , following from Eq. 2.10. The computation easily gives

$$\begin{aligned} \langle O^{\alpha,\alpha'} \rangle &= \sum_{i<j} p_{ij}^{\alpha} p_{ij}^{\alpha'} \\ &= \binom{N}{2} p^{\alpha} p^{\alpha'} \\ &= \frac{2L^{\alpha} L^{\alpha'}}{N(N-1)} \end{aligned} \quad (2.58)$$

As expected for uncorrelated multiplex ensembles  $\langle a_{ij}^{\alpha} a_{ij}^{\alpha'} \rangle = \langle a_{ij}^{\alpha} \rangle \langle a_{ij}^{\alpha'} \rangle$ . Moreover, if  $L^{\alpha} = \mathcal{O}(N) \forall \alpha = 1, 2, \dots, M$  then  $\langle O^{\alpha,\alpha'} \rangle$  is a finite number in the large network limit and so, the overlap of links in this limit becomes a totally negligible phenomena:  $\langle O^{\alpha,\alpha'} \rangle$  is in fact much smaller than both  $L^{\alpha}$  and  $L^{\alpha'}$  (Bianconi, 2013).

### Multiplex ensemble with expected degree sequence in each layer

As possible constraints we can consider the expected degree sequence  $\{k_i^{\alpha}\}$  for each layer  $\alpha$ . The number of constraints in this case is  $K = M \times N$ , indicated with a label  $\alpha = 1, 2, \dots, M$ . These constraints are given by

$$\sum_{\vec{G}} F_{i,\alpha}(\vec{G}) P(\vec{G}) = \sum_{\vec{G}} \left( \sum_{j \neq i} a_{ij}^{\alpha} \right) P(\vec{G}) = k_i^{\alpha} \quad (2.59)$$

The probability of a multiplex  $P(\vec{G})$  is given by Eq. 2.36 that in this case can be written as

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{\alpha=1}^M \sum_i \mu_{i,\alpha} \sum_{j \neq i} a_{ij}^{\alpha} \right] \quad (2.60)$$

where the partition function  $Z$  can be expressed explicitly as

$$Z = \sum_{\vec{G}} \exp \left[ - \sum_{\alpha=1}^M \sum_i \mu_{i,\alpha} \sum_{j \neq i} a_{ij}^\alpha \right] \quad (2.61)$$

$$= \prod_{\alpha=1}^M \prod_{i < j} (1 + e^{-(\mu_{i,\alpha} + \mu_{j,\alpha})}) \quad (2.62)$$

and the Lagrangian multipliers  $\mu_{i,\alpha}$  are fixed by the condition

$$k_i^\alpha = - \frac{\partial \log Z}{\partial \mu_{i,\alpha}} = \sum_{j \neq i} \frac{e^{-(\mu_{i,\alpha} + \mu_{j,\alpha})}}{1 + e^{-(\mu_{i,\alpha} + \mu_{j,\alpha})}}. \quad (2.63)$$

The probability of having a link between node  $i$  and node  $j$  in layer  $\alpha$ ,  $p_{ij}^\alpha$ , reads

$$p_{ij}^\alpha = \langle a_{ij}^\alpha \rangle = \frac{e^{-(\mu_{i,\alpha} + \mu_{j,\alpha})}}{1 + e^{-(\mu_{i,\alpha} + \mu_{j,\alpha})}} \quad (2.64)$$

The probability of a given multiplex  $P(\vec{G})$  and the related entropy  $\mathcal{S}$  follow, respectively, from Eq. 2.52 and Eq. 2.55, substituting Eq. 2.64.

## Examples of correlated unweighted multiplex ensembles

### Multiplex ensemble with given expected total number of multilinks $\vec{m}$

We consider a correlated multiplex ensemble, in which we fix the total number of multilinks  $\vec{m}$ , given by  $L^{\vec{m}}$ . The number of possible constraints is equal to  $K = 2^M - 1$  because, as previously mentioned, the number of different multilinks is  $2^M$  but only  $2^M - 1$  are independent thank to the normalisation condition.

These constraints are given by

$$\sum_{\vec{G}} F^{\vec{m}}(\vec{G}) P(\vec{G}) = \sum_{\vec{G}} \left( \sum_{i < j} A_{ij}^{\vec{m}} \right) P(\vec{G}) = L^{\vec{m}}, \quad (2.65)$$

where the multiadjacency matrix element  $A_{ij}^{\vec{m}}$  is defined in Eq. 2.12, and for the particular case of an unweighted multiplex turns into

$$A_{ij}^{\vec{m}} = \prod_{\alpha=1}^M [a_{ij}^{\alpha} m_{\alpha} + (1 - a_{ij}^{\alpha})(1 - m_{\alpha})] \quad (2.66)$$

The canonical probability  $P(\vec{G})$  of a multiplex in the ensembles is given by the general expression given in Eq. 2.36 that in this case becomes

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{\vec{m} \neq \vec{0}} \kappa^{\vec{m}} \sum_{i < j} A_{ij}^{\vec{m}} \right] \quad (2.67)$$

where the partition function  $Z$  is given by

$$Z = \mathcal{Z}^{\binom{N}{2}} \quad (2.68)$$

where

$$\mathcal{Z} = 1 + \sum_{\vec{m} \neq \vec{0}} e^{-\kappa^{\vec{m}}} \quad (2.69)$$

The Lagrangian multipliers  $\kappa^{\vec{m}}$  ( $\vec{m} \neq \vec{0}$ ) are fixed by the conditions

$$- \frac{\partial \log Z}{\partial \kappa^{\vec{m}}} = L^{\vec{m}}, \quad (2.70)$$

which yields

$$L^{\vec{m}} = \binom{N}{2} \frac{e^{-\kappa^{\vec{m}}}}{1 + \sum_{\vec{m} \neq \vec{0}} e^{-\kappa^{\vec{m}}}} = \binom{N}{2} \langle A_{ij}^{\vec{m}} \rangle \quad (2.71)$$

The probability of a multilink  $\vec{m}$  between node  $i$  and node  $j$  is  $p_{ij}^{\vec{m}} = \langle A_{ij}^{\vec{m}} \rangle$  and in this ensemble it is independent on the pair of nodes  $(i, j)$ , i.e.  $p_{ij}^{\vec{m}} = p^{\vec{m}}$  and

$$p^{\vec{m}} = \frac{e^{-\kappa^{\vec{m}}}}{1 + \sum_{\vec{m} \neq \vec{0}} e^{-\kappa^{\vec{m}}}} = \frac{L^{\vec{m}}}{\binom{N}{2}} \quad (2.72)$$

The particular case  $\vec{m} = \vec{0}$  depends on the other multilinks, namely,

$$p^{\vec{0}} = 1 - \sum_{\vec{m} \neq \vec{0}} p_{ij}^{\vec{m}} = \frac{1}{Z} \quad (2.73)$$

$$L^{\vec{0}} = \binom{N}{2} - \sum_{\vec{m} \neq \vec{0}} L^{\vec{m}} = \binom{N}{2} p^{\vec{0}} \quad (2.74)$$

Fixed  $\{p_{ij}^{\vec{m}}\}$ , the probability of a multiplex  $P(\vec{G})$  can be rearranged as a function of the marginal probabilities  $p_{ij}^{\vec{m}}$ , i.e.

$$P(\vec{G}) = \prod_{i < j} \prod_{\vec{m}} (p_{ij}^{\vec{m}})^{A_{ij}^{\vec{m}}} \quad (2.75)$$

$$= \prod_{\vec{m}} (p^{\vec{m}})^{\ell^{\vec{m}}} \quad (2.76)$$

where  $\{\ell^{\vec{m}}\}$  is the desired set of total multilink. The last expression correctly satisfies the normalisation condition if we consider the right multiplicity of a generic state with a given sequence of  $\{\ell^{\vec{m}}\}$ , i.e.

$$\sum_{\{a_{ij}^{\alpha}\}} \prod_{i < j} \prod_{\vec{m}} (p_{ij}^{\vec{m}})^{A_{ij}^{\vec{m}}} = \sum_{\{\ell^{\vec{m}}\}} \frac{N(N-1)/2}{\prod_{\vec{m}} \ell^{\vec{m}}!} \prod_{\vec{m}} (p^{\vec{m}})^{\ell^{\vec{m}}} = 1 \quad (2.77)$$

like a proper multinomial probability distribution.

Finally, the entropy  $\mathcal{S}$  of this ensemble can be calculated starting from its defi-

inition Eq. 2.33 as a function of the probability marginals  $p_{ij}^{\vec{m}}$ , i.e.

$$\mathcal{S} = - \sum_{i < j} \sum_{\vec{m}} p_{ij}^{\vec{m}} \log p_{ij}^{\vec{m}} \quad (2.78)$$

$$= - \binom{N}{2} \sum_{\vec{m}} p^{\vec{m}} \log p^{\vec{m}} \quad (2.79)$$

where  $p^{\vec{m}}$  is given by Eq. 2.72. If the number of layers  $M$  is finite, we finally get (Stirling's approximation in the large  $N$  limit)

$$\mathcal{S} = \log \frac{N(N-1)/2}{\prod_{\vec{m}} (L^{\vec{m}}!)} \quad (2.80)$$

where Eq. 2.72 was used. We can now evaluate the average global overlap  $\langle O^{\alpha, \alpha'} \rangle$  from Eq. 2.10 for a comparison with the results for uncorrelated multiplex ensembles (Eq. 2.58). We get

$$\langle O^{\alpha, \alpha'} \rangle = \sum_{\vec{m} | m_{\alpha} = m_{\alpha'} = 1} \sum_{i < j} p_{ij}^{\vec{m}} \quad (2.81)$$

and for the particular case of a duplex we have only one possible situation, i.e.

$$\langle O \rangle = L^{11} = \binom{N}{2} p^{11} \quad (2.82)$$

This quantity can be significant even for sparse networks. Assuming  $L^{11}$ ,  $L^{01}$  and  $L^{10}$  to be proportional to  $N$  implies an overlap not irrelevant (Bianconi, 2013).

### **Multiplex ensemble with given expected total number of $\nu$ -multilinks**

In presence of many layers  $M$  we can consider as constraints the average total number of  $\nu$ -multilinks  $L^{\nu}$  with  $\nu = 1, 2, \dots, M$ . With respect to the previ-

ous case, now the number of constraints is sensibly reduced and is given by  $M$  constraints, i.e.,

$$\sum_{\vec{G}} F^\nu(\vec{G}) P(\vec{G}) = \sum_{\vec{G}} \left( \sum_{i<j} A_{ij}^\nu \right) P(\vec{G}) = L^\nu, \quad (2.83)$$

where  $A_{ij}^\nu$  is defined by Eq. 2.23. The canonical probability  $P(\vec{G})$  of a multiplex in the ensembles follows from the general expression given in Eq. 2.36 that in this case becomes

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{\nu=1}^M \omega^\nu \sum_{i<j} A_{ij}^\nu \right] \quad (2.84)$$

where the partition function  $Z$  is given by

$$Z = \mathcal{Z}^{\binom{N}{2}} \quad (2.85)$$

where

$$\mathcal{Z} = 1 + \sum_{\nu=1}^M \binom{M}{\nu} e^{-\omega^\nu} \quad (2.86)$$

The Lagrangian multipliers  $\omega^\nu$  ( $\nu \neq 0$ ), are fixed by the conditions

$$- \frac{\partial \log Z}{\partial \omega^\nu} = L^\nu, \quad (2.87)$$

which yields

$$L^\nu = \binom{N}{2} \frac{1}{\mathcal{Z}} \binom{M}{\nu} e^{-\omega^\nu} \quad (2.88)$$

The probability of having a  $\nu$ -multilink  $p_{ij}^\nu$  in this ensemble is independent on the pair of nodes  $(i, j)$ , namely

$$p^\nu = \langle A_{ij}^\nu \rangle = \frac{L^\nu}{\binom{N}{2}} = \frac{1}{\mathcal{Z}} \binom{M}{\nu} e^{-\omega^\nu} \quad (2.89)$$

We can even consider the probability of a given multilink  $\vec{m}$  (independent on the pair of nodes  $(i, j)$  too)

$$p^{\vec{m}} = \langle A_{ij}^{\vec{m}} \rangle = \frac{1}{\mathcal{Z}} e^{-\omega^{\nu(\vec{m})}} = \frac{p^{\nu(\vec{m})}}{\binom{M}{\nu(\vec{m})}} \quad (2.90)$$

where, for a given  $\vec{m}$ ,  $\nu(\vec{m}) = \sum_\alpha m_\alpha$ .

The particular case for  $\nu = 0$  depends on the other  $\nu$ -multilinks, i.e.

$$p^0 = 1 - \sum_{\nu=1}^M p_{ij}^\nu = \frac{1}{\mathcal{Z}} \quad (2.91)$$

$$L^0 = \binom{N}{2} - \sum_{\nu=1}^M L^\nu = \binom{N}{2} p^0 \quad (2.92)$$

The probability of a multiplex  $P(\vec{G})$  can be rearranged as a function of the marginal probabilities  $p^\nu$ , i.e.

$$P(\vec{G}) = \prod_{i < j} \prod_{\nu=0}^M \left( \frac{p_{ij}^\nu}{\binom{M}{\nu}} \right)^{A_{ij}^\nu} \quad (2.93)$$

$$= \prod_{\nu=0}^M \left( \frac{p^\nu}{\binom{M}{\nu}} \right)^{\ell^\nu} \quad (2.94)$$

where  $\{\ell^\nu\}$  is the desired sequence of  $\nu$ -multilinks. The last expression correctly satisfies the normalisation condition if we take into account the multiplicity of a

generic state with a given sequence of  $\{L^\nu\}$

$$\sum_{\{a_{ij}^\nu\}} \prod_{i<j} \prod_{\nu=0}^M \left( \frac{p_{ij}^\nu}{\binom{M}{\nu}} \right)^{A_{ij}^\nu} = \sum_{\{\ell^\nu\}} \frac{N(N-1)/2}{\prod_{\nu} \ell^\nu!} \left[ \prod_{\nu=0}^M \binom{M}{\nu}^{\ell^\nu} \right] \prod_{\nu=0}^M \left( \frac{p^\nu}{\binom{M}{\nu}} \right)^{\ell^\nu} \quad (2.95)$$

The previous multiplicity tells us that not only we choose in how many ways the  $\nu$ -multilinks are distributed across the couples of nodes, but for each  $\nu$ -multilink we have  $\binom{M}{\nu}$  multilinks  $\vec{m}_{\nu(\vec{m})=\nu}$  that can be represented, i.e. different patterns of *link activities* that correspond to the same overlap multiplicity.

Starting from Eq. 2.33 the entropy  $\mathcal{S}$  of this ensemble can be calculated as a function of the probability marginals  $p_{ij}^\nu$ , i.e.

$$\mathcal{S} = - \sum_{i<j} \sum_{\nu=0}^M p_{ij}^\nu \log(p_{ij}^\nu) + \sum_{i<j} \sum_{\nu=0}^M p_{ij}^\nu \log \binom{M}{\nu} \quad (2.96)$$

$$= - \binom{N}{2} \sum_{\nu=0}^M p^\nu \log(p^\nu) + \binom{N}{2} \sum_{\nu=0}^M p^\nu \log \binom{M}{\nu} \quad (2.97)$$

When  $M$  is finite we can always approximate the last expression thank to the Stirling's Approximation, i.e.

$$\mathcal{S} = \log \left( \frac{\binom{N}{2}!}{\prod_{\nu=0}^M L^\nu!} \prod_{\nu=0}^M \binom{M}{\nu}^{L^\nu} \right) \quad (2.98)$$

where Eq. 2.89 was considered.

### Multiplex ensemble with given expected multidegree sequence

Here we consider another level of coarse-graining for the multiplex network and we study correlated unweighted multiplex in which we fix the average multidegree

sequence  $k_i^{\vec{m}}$  for each node  $i$ , for a given multilink  $\vec{m}$ . Following the previous line of reasoning, we can express  $N \times (2^M - 1)$  constraints.

These constraints are given by

$$\sum_{\vec{G}} F_i^{\vec{m}}(\vec{G}) P(\vec{G}) = \sum_{\vec{G}} \left( \sum_{j \neq i} A_{ij}^{\vec{m}} \right) P(\vec{G}) = k_i^{\vec{m}}, \quad (2.99)$$

with  $i = 1, \dots, N$  and  $\vec{m} = (m_1, m_2, \dots, m_\beta, \dots, m_M)$  with  $m_\beta = 0, 1$ . The canonical probability  $P(\vec{G})$  of the multiplex in the ensemble is

$$\begin{aligned} P(\vec{G}) &= \frac{1}{Z} \exp \left[ - \sum_{\vec{m} \neq \vec{0}} \sum_i \mu_i^{\vec{m}} \sum_{j \neq i} A_{ij}^{\vec{m}} \right] \\ &= \frac{1}{Z} \prod_{i < j} \exp \left[ - \sum_{\vec{m} \neq \vec{0}} (\mu_i^{\vec{m}} + \mu_j^{\vec{m}}) A_{ij}^{\vec{m}} \right], \end{aligned} \quad (2.100)$$

where the partition function  $Z$  can be expressed explicitly as

$$Z = \prod_{i < j} \mathcal{Z}_{ij} \quad (2.101)$$

where

$$\mathcal{Z}_{ij} = 1 + \sum_{\vec{m} \neq \vec{0}} e^{-(\mu_i^{\vec{m}} + \mu_j^{\vec{m}})} \quad (2.102)$$

The Lagrangian multipliers  $\mu_i^{\vec{m}}$  ( $\vec{m} \neq \vec{0}$ ) are fixed by the conditions

$$- \frac{\partial \log Z}{\partial \mu_i^{\vec{m}}} = k_i^{\vec{m}} = \sum_{j \neq i} \langle A_{ij}^{\vec{m}} \rangle, \quad (2.103)$$

The probability of a multilink  $\vec{m}$  between node  $i$  and node  $j$  is  $p_{ij}^{\vec{m}} = \langle A_{ij}^{\vec{m}} \rangle$  and

it is given by

$$p_{ij}^{\vec{m}} = \frac{e^{-(\mu_i^{\vec{m}} + \mu_j^{\vec{m}})}}{1 + \sum_{\vec{m} \neq \vec{0}} e^{-(\mu_i^{\vec{m}} + \mu_j^{\vec{m}})}} \quad (2.104)$$

As previously, the measures related to  $\vec{m} = \vec{0}$  depend on the other multilinks, namely,

$$p_{ij}^{\vec{0}} = 1 - \sum_{\vec{m} \neq \vec{0}} p_{ij}^{\vec{m}} = \frac{1}{\mathcal{Z}_{ij}} \quad (2.105)$$

$$k_i^{\vec{0}} = N - 1 - \sum_{\vec{m} \neq \vec{0}} k_i^{\vec{m}} \quad (2.106)$$

The probability of given multiplex in this ensemble  $P(\vec{G})$  and the related entropy value  $\mathcal{S}$  follow from Eq. 2.75 and Eq. 2.78.

### Multiplex ensemble with given expected $\nu$ -multidegree sequence

Considering now the  $\nu$ -multilinks we convert the previous model into a new one, i.e. we fix the given expected  $\nu$ -multidegree sequence, with a number of constraints equal to  $M \times N$ :

$$\sum_{\vec{G}} F_i^\nu(\vec{G}) P(\vec{G}) = \sum_{\vec{G}} \left( \sum_{j \neq i} A_{ij}^\nu \right) P(\vec{G}) = k_i^\nu, \quad (2.107)$$

The canonical probability  $P(\vec{G})$  of the multiplex in the ensembles reads

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{i < j} \sum_{\nu=1}^M (\omega_i^\nu + \omega_j^\nu) A_{ij}^\nu \right] \quad (2.108)$$

where the partition function  $Z$  is given by

$$Z = \prod_{i < j} \mathcal{Z}_{ij} \quad (2.109)$$

where

$$\mathcal{Z}_{ij} = 1 + \sum_{\nu=1}^M \binom{M}{\nu} e^{-(\omega_i^\nu + \omega_j^\nu)} \quad (2.110)$$

The Lagrangian multipliers  $\omega_i^\nu$  ( $\nu \neq 0$ ), are fixed by the conditions

$$-\frac{\partial \log Z}{\partial \omega_i^\nu} = k_i^\nu = \sum_{j \neq i} \langle A_{ij}^\nu \rangle \quad (2.111)$$

The probability of a  $\nu$ -multilink between node  $i$  and node  $j$  is  $p_{ij}^\nu = \langle A_{ij}^\nu \rangle$  and it is given by

$$p_{ij}^\nu = \langle A_{ij}^\nu \rangle = \frac{1}{\mathcal{Z}_{ij}} \binom{M}{\nu} e^{-(\omega_i^\nu + \omega_j^\nu)} \quad (2.112)$$

The probability of a given multilink  $\vec{m}$  between node  $i$  and node  $j$

$$p_{ij}^{\vec{m}} = \langle A_{ij}^{\vec{m}} \rangle = \frac{1}{\mathcal{Z}_{ij}} e^{-(\omega_i^\nu + \omega_j^\nu)} = \frac{p_{ij}^{\nu(\vec{m})}}{\binom{M}{\nu}} \quad (2.113)$$

Finally, the probability of a given multiplex  $P(\vec{G})$  and the consequent entropy value  $\mathcal{S}$  follow from Eqs. 2.93, 2.96.

### 2.3.3 Weighted multiplex

We present now the most useful null models for weighted multiplex networks, both uncorrelated and correlated. We start considering examples related to the measure of total strength (a proper warm-up) and then we show the main calculations for those ensembles that enforce in the same time constraints on the strength sequence and on the degree sequence. For the analysis of real data these ensembles are quite essential to study weight-topology correlations. For

additional examples of weighted multiplex ensembles see Sec. 4.2.

In Sec. 2.3.3, 2.4.2 we illustrate a null model with given expected multidegree sequence  $\{k_i^{\vec{m}}\}$  and given expected multistrength sequence  $\{s_{i,\alpha}^{\vec{m}}\}$  applied to biological data.

## Examples of uncorrelated weighted multiplex ensembles

### Multiplex ensembles with given expected total strength in each layer

As a first example of uncorrelated weighted multiplex, we consider the case in which we fix the average strength in each layer  $\alpha$  to be equal to  $S^\alpha$ . In this case we have  $K = M$  constraints in the system, indicated with a label  $\alpha = 1, 2, \dots, M$ . These constraints are given by

$$\sum_{\vec{G}} F_\alpha(\vec{G}) P(\vec{G}) = \sum_{\vec{G}} \left( \sum_{i < j} a_{ij}^\alpha \right) P(\vec{G}) = S^\alpha. \quad (2.114)$$

The probability distribution of a multiplex in this ensemble is given by Eq. 2.36 that reads in this case,

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{\alpha=1}^M \lambda_\alpha \sum_{i < j} a_{ij}^\alpha \right], \quad (2.115)$$

where the partition function  $Z$  can be expressed explicitly as

$$\begin{aligned} Z &= \sum_{\vec{G}} \exp \left[ - \sum_{\alpha=1}^M \lambda_\alpha \sum_{i < j} a_{ij}^\alpha \right] \\ &= \prod_{\alpha=1}^M \left[ \left( \frac{1}{1 - e^{-\lambda_\alpha}} \right)^{\binom{N}{2}} \right]. \end{aligned} \quad (2.116)$$

The Lagrangian multipliers  $\lambda_\alpha$  defining the probability of the multiplex  $P(\vec{G})$ , are fixed by the conditions

$$S_\alpha = -\frac{\partial \log Z}{\partial \lambda_\alpha} = \binom{N}{2} \frac{e^{-\lambda_\alpha}}{1 - e^{-\lambda_\alpha}}. \quad (2.117)$$

Finally the average weight  $\langle a_{ij}^\alpha \rangle$  can be evaluated from Eq. 2.40 and is given by

$$\langle a_{ij}^\alpha \rangle = \frac{S_\alpha}{\binom{N}{2}}, \quad (2.118)$$

that is equivalent to say  $S_\alpha = \sum_{i<j} \langle a_{ij}^\alpha \rangle$ .

From Eq. 2.38 we write the marginal probabilities  $\pi(a_{ij}^\alpha)$  in this specific multiplex ensemble as

$$\pi_{ij}^\alpha(a_{ij}^\alpha) = e^{-\lambda_\alpha a_{ij}^\alpha} (1 - e^{-\lambda_\alpha}). \quad (2.119)$$

Moreover, from Eq. 2.41 the probability  $p_{ij}^\alpha$  of having a positive weight  $a_{ij}^\alpha > 0$  of the link between node  $i$  and node  $j$  in layer  $\alpha$  is independent on the pair of nodes  $(i, j)$ , i.e.  $p_{ij}^\alpha = p^\alpha$  and is given by

$$p^\alpha = e^{-\lambda_\alpha}. \quad (2.120)$$

We observe that we can write the Eq. 2.36 in terms of marginal probabilities  $\pi_{ij}^\alpha(a_{ij}^\alpha)$ , namely

$$P(\vec{G}) = \prod_{\alpha=1}^M \prod_{i<j} \pi_{ij}^\alpha(a_{ij}^\alpha). \quad (2.121)$$

Therefore the entropy  $\mathcal{S}$  of this canonical multiplex ensemble is given by Eq. 2.33

and in this special case can be written as

$$\mathcal{S} = - \sum_{\alpha=1}^M \sum_{i < j} \sum_{a_{ij}^{\alpha}=0}^{\infty} \pi_{ij}^{\alpha}(a_{ij}^{\alpha}) \log(\pi_{ij}^{\alpha}(a_{ij}^{\alpha})). \quad (2.122)$$

Using the marginals  $\pi_{ij}^{\alpha}(a_{ij}^{\alpha})$  given by Eqs. 2.119 and Eq. 2.117 the entropy can be rearranged as

$$\begin{aligned} \mathcal{S} = & \sum_{\alpha=1}^M \left[ \left( \binom{N}{2} + S_{\alpha} \right) \log \left( \binom{N}{2} + S_{\alpha} \right) \right. \\ & \left. - S_{\alpha} \log S_{\alpha} - \binom{N}{2} \log \binom{N}{2} \right] \end{aligned} \quad (2.123)$$

If the number of layers  $M$  is finite, applying the Stirling's approximation in the large  $N$  limit we get

$$\mathcal{S} = \sum_{\alpha=1}^M \log \left[ \binom{\binom{N}{2} + S_{\alpha}}{\binom{N}{2}} \right]. \quad (2.124)$$

### **Multiplex ensembles with given expected strength sequence and degree sequence in each layer**

We fix the expected strength  $s_i^{\alpha}$  and the expected degree  $k_i^{\alpha}$  of every node  $i$ , in each layer  $\alpha$ . We have  $K = M \times 2N$  constraints in the system. These constraints are given by

$$\begin{aligned} \sum_{\vec{G}} F_{i,\alpha}(\vec{G}) P(\vec{G}) &= \sum_{\vec{G}} \left( \sum_{j \neq i} a_{ij}^{\alpha} \right) P(\vec{G}) = s_i^{\alpha} \\ \sum_{\vec{G}} F_{i,\alpha}(\vec{G}) P(\vec{G}) &= \sum_{\vec{G}} \left( \sum_{j \neq i} \theta(a_{ij}^{\alpha}) \right) P(\vec{G}) = k_i^{\alpha}, \end{aligned} \quad (2.125)$$

with  $\alpha = 1, 2, \dots, M$ . We introduce the Lagrangian multipliers  $\lambda_{i,\alpha}$  for the first set of  $N \cdot M$  constraints and the Lagrangian multipliers  $\omega_{i,\alpha}$  for the second set of  $N \cdot M$  constraints. Therefore, the probability  $P(\vec{G})$  of a multiplex in this

ensemble, of general expression given by Eq. 2.36, in this specific example is given by

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{\alpha=1}^M \sum_i \lambda_{i,\alpha} \sum_{j \neq i} a_{ij}^\alpha - \sum_{\alpha=1}^M \sum_i \omega_{i,\alpha} \sum_{j \neq i} \theta(a_{ij}^\alpha) \right]$$

where the partition function  $Z$  can be expressed explicitly as

$$\begin{aligned} Z &= \sum_{\vec{G}} \exp \left[ - \sum_{\alpha=1}^M \sum_i \sum_{j \neq i} (\lambda_{i,\alpha} a_{ij}^\alpha + \omega_{i,\alpha} \theta(a_{ij}^\alpha)) \right] \\ &= \prod_{\alpha=1}^M \prod_{i < j} \left( 1 + \frac{e^{-(\omega_{i,\alpha} + \omega_{j,\alpha}) - (\lambda_{i,\alpha} + \lambda_{j,\alpha})}}{1 - e^{-(\lambda_{i,\alpha} + \lambda_{j,\alpha})}} \right), \end{aligned} \quad (2.126)$$

and the Lagrangian multipliers are fixed by the conditions

$$\begin{aligned} s_i^\alpha &= - \frac{\partial \log Z}{\partial \lambda_{i,\alpha}} \\ k_i^\alpha &= - \frac{\partial \log Z}{\partial \omega_{i,\alpha}} \end{aligned} \quad (2.127)$$

The average weight of the link  $(i, j)$  in layer  $\alpha$ , i.e.  $\langle a_{ij}^\alpha \rangle$ , is given by Eq. 2.40 that in this case reads

$$\langle a_{ij}^\alpha \rangle = \frac{e^{-(\omega_{i,\alpha} + \omega_{j,\alpha}) + (\lambda_{i,\alpha} + \lambda_{j,\alpha})}}{(e^{\lambda_{i,\alpha} + \lambda_{j,\alpha}} - 1)(e^{-(\omega_{i,\alpha} + \omega_{j,\alpha})} + e^{\lambda_{i,\alpha} + \lambda_{j,\alpha}} - 1)} \quad (2.128)$$

From Eq. 2.38 we write the marginal probabilities  $\pi_{ij}^\alpha(a_{ij}^\alpha)$  for this specific ensemble that is given by

$$\pi_{ij}^\alpha(a_{ij}^\alpha) = \frac{e^{-(\lambda_{i,\alpha} + \lambda_{j,\alpha}) a_{ij}^\alpha - (\omega_{i,\alpha} + \omega_{j,\alpha}) \theta(a_{ij}^\alpha)} (1 - e^{-(\lambda_{i,\alpha} + \lambda_{j,\alpha})})}{1 + e^{-(\lambda_{i,\alpha} + \lambda_{j,\alpha})} (e^{-(\omega_{i,\alpha} + \omega_{j,\alpha})} - 1)}. \quad (2.129)$$

Moreover, from Eq. 2.41 the probability  $p_{ij}^\alpha$  that the link  $(i, j)$  in layer  $\alpha$  has

weight different from zero is given by

$$p_{ij}^\alpha = \frac{e^{-(\omega_{i,\alpha} + \omega_{j,\alpha})}}{e^{-(\omega_{i,\alpha} + \omega_{j,\alpha})} + e^{\lambda_{i,\alpha} + \lambda_{j,\alpha}} - 1} \quad (2.130)$$

The probability of a multiplex in this ensemble is given by Eq. 2.121 with the marginals  $\pi_{ij}^\alpha(a_{ij}^\alpha)$  given by Eq. 2.129. The entropy  $\mathcal{S}$  of this canonical multiplex ensemble is given by Eq. 2.122.

## Examples of correlated weighted multiplex ensembles

### Multiplex ensembles with given expected total multistrength $S_\alpha^{\vec{m}}$

Here we consider a correlated weighted multiplex ensemble, in which we fix the total multistrength  $\vec{m}$ , given by  $S_\alpha^{\vec{m}}$  for a layer  $\alpha$  such that  $m_\alpha = 1$ . Since the number of the possible multistrengths  $\vec{m}$  in layer  $\alpha$  are given by  $M \cdot 2^{M-1}$ , this gives a number of constraints that is equal to  $K = M \cdot 2^{M-1}$ . These constraints are given by

$$\sum_{\vec{G}} F_\alpha^{\vec{m}}(\vec{G}) P(\vec{G}) = \sum_{\vec{G}} \left( \sum_{i < j} A_{ij}^{\vec{m}} a_{ij}^\alpha \right) P(\vec{G}) = S_\alpha^{\vec{m}}, \quad (2.131)$$

where the multiadjacency matrix element  $A_{ij}^{\vec{m}}$  is defined in Eq. 2.12. The canonical probability  $P(\vec{G})$  of the multiplex in the ensembles is given by the general expression given in Eq. 2.36 that in this case becomes

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{\vec{m} \neq \vec{0}} \sum_{\alpha=1}^M \lambda_\alpha^{\vec{m}} \sum_{i < j} A_{ij}^{\vec{m}} a_{ij}^\alpha \right] \quad (2.132)$$

where the partition function  $Z$  is given by

$$Z = \mathcal{Z}^{\binom{N}{2}} \quad (2.133)$$

where

$$\mathcal{Z} = \sum_{\vec{m}} \prod_{\alpha=1}^M \left( \frac{e^{-\lambda_{\alpha}^{\vec{m}}}}{1 - e^{-\lambda_{\alpha}^{\vec{m}}}} \right)^{m_{\alpha}} \quad (2.134)$$

Without loss of generality, if  $m_{\alpha} = 0$  we put  $\lambda_{\alpha}^{\vec{m}} = 1/2$ . We can do this because the probability of a multiplex does not depend on any of these values, and we need to define them only for simplifying the notation. The Lagrangian multipliers  $\lambda_{\alpha}^{\vec{m}}$  with  $m_{\alpha} = 1$ , are fixed by the conditions

$$-\frac{\partial \log Z}{\partial \lambda_{\alpha}^{\vec{m}}} = S_{\alpha}^{\vec{m}}, \quad (2.135)$$

which yields

$$S_{\alpha}^{\vec{m}} = \binom{N}{2} \frac{1}{\mathcal{Z}} \left( \frac{1}{1 - e^{-\lambda_{\alpha}^{\vec{m}}}} \right) \prod_{\beta=1}^M \left( \frac{e^{-\lambda_{\beta}^{\vec{m}}}}{1 - e^{-\lambda_{\beta}^{\vec{m}}}} \right)^{m_{\beta}}. \quad (2.136)$$

We now indicate with  $\vec{a}_{ij}$  the vector  $(a_{ij}^1, a_{ij}^2, \dots, a_{ij}^{\alpha}, \dots, a_{ij}^M)$ . The probability of a multiplex  $P(\vec{G})$  can be rewritten as

$$P(\vec{G}) = \prod_{i < j} \pi_{ij}(\vec{a}_{ij}), \quad (2.137)$$

with

$$\pi_{ij}(\vec{a}_{ij}) = \frac{e^{-\sum_{\alpha=1, M} \lambda_{\alpha}^{\vec{m}^{ij}} a_{ij}^{\alpha}}}{\mathcal{Z}}, \quad (2.138)$$

where  $\vec{m}^{ij} = (m_1^{ij}, \dots, m_{\alpha}^{ij}, \dots, m_m^{ij})$  with  $m_{\alpha}^{ij} = \theta(a_{ij}^{\alpha})$ . With  $\pi_{ij}(\vec{a}_{ij})$  we define, for a position  $ij$ , the probability of a particular sequence of weights on the layers.

The normalization condition is fulfilled

$$\sum_{\vec{a}_{ij}} \pi_{ij}(\vec{a}_{ij}) = 1. \quad (2.139)$$

Further on we can compute the average weight of the link  $ij$  on the multilink  $\vec{m}$ , in the layer  $\alpha$

$$\langle a_{ij}^\alpha A_{ij}^{\vec{m}} \rangle = \sum_{\vec{G}} a_{ij}^\alpha A_{ij}^{\vec{m}} P(\vec{G}) = \sum_{\vec{a}_{ij}} a_{ij}^\alpha A_{ij}^{\vec{m}} \pi_{ij}(\vec{a}_{ij}). \quad (2.140)$$

Using Eq. 2.138 for the explicit expression of  $\pi(\vec{a}_{ij})$  and comparing the results with Eq. 2.136 it is easy to show that

$$\langle a_{ij}^\alpha A_{ij}^{\vec{m}} \rangle = \frac{S_\alpha^{\vec{m}}}{\binom{N}{2}}. \quad (2.141)$$

The probability of a multilink  $\vec{m}$  between node  $i$  and node  $j$ ,  $p_{ij}^{\vec{m}} = \langle A_{ij}^{\vec{m}} \rangle$  in this ensemble is independent on the pair of nodes  $(i, j)$ . Therefore we have  $p_{ij}^{\vec{m}} = p^{\vec{m}}$  with

$$p^{\vec{m}} = \frac{\prod_{\alpha=1}^M \left( \frac{e^{-\lambda_\alpha^{\vec{m}}}}{1 - e^{-\lambda_\alpha^{\vec{m}}}} \right)^{m_\alpha}}{\mathcal{Z}}, \quad (2.142)$$

where the normalization condition is fulfilled, namely,

$$\sum_{\vec{m}} p_{ij}^{\vec{m}} = 1. \quad (2.143)$$

Moreover, the relationship between  $p_{ij}^{\vec{m}}$  and the probabilities  $\pi_{ij}(\vec{a}_{ij})$  is

$$\sum_{\vec{a}_{ij}} A_{ij}^{\vec{m}} \pi_{ij}(\vec{a}_{ij}) = p_{ij}^{\vec{m}}. \quad (2.144)$$

Finally, the probability of a multiplex  $P(\vec{G})$  is given by Eq. 2.137 and the entropy  $\mathcal{S}$  of this ensemble can be calculated starting from its definition Eq. 2.33, giving

$$\mathcal{S} = - \sum_{i < j} \sum_{\vec{a}_{ij}} \pi_{ij}(\vec{a}_{ij}) \log \pi_{ij}(\vec{a}_{ij}). \quad (2.145)$$

### Multiplex ensembles with given expected $\nu$ -total strength $S_\alpha^\nu$

In presence of many layers  $M$  we can consider as constraints the average  $\nu$ -total strength  $S_\alpha^\nu$  with  $\nu = 1, 2, \dots, M$ . With respect to the case with fixed total average multistrength, the number of constraints is dramatically reduced and is given by  $M^2$  constraints

$$\sum_{\vec{G}} F_\alpha^\nu(\vec{G}) P(\vec{G}) = \sum_{\vec{G}} \left( \sum_{i < j} a_{ij}^\alpha A_{ij}^\nu \right) P(\vec{G}) = S_\alpha^\nu. \quad (2.146)$$

The probability  $P(\vec{G})$  of the multiplex network, is therefore given in terms of  $M^2$  Lagrangian multipliers  $\lambda_\alpha^\nu$ , i.e.

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{\nu=1}^M \sum_{\alpha=1}^M \lambda_\alpha^\nu \sum_{i < j} A_{ij}^\nu a_{ij}^\alpha \right] \quad (2.147)$$

where the partition function  $Z$  is given by  $Z = \mathcal{Z}^{\binom{N}{2}}$  with

$$\mathcal{Z} = \sum_{\nu=0}^M \sum_{\vec{m} | \nu(\vec{m}) = \nu} \prod_{\alpha=1}^M \left( \frac{e^{-\lambda_\alpha^\nu}}{1 - e^{-\lambda_\alpha^\nu}} \right)^{m_\alpha}. \quad (2.148)$$

The Lagrangian multipliers  $\lambda_\alpha^\nu$  are fixed fixed by the constraints Eq.2.147 that can be also expressed as

$$- \frac{\partial \log Z}{\partial \lambda_\alpha^\nu} = S_\alpha^\nu. \quad (2.149)$$

The probability  $P(\vec{G})$  of the multiplex network is given by Eq. 2.137 and the entropy of the ensemble takes the simple expression given by Eq. 2.145 where  $\pi_{ij}(\vec{a}_{ij})$  is given by

$$\pi_{ij}(\vec{a}_{ij}) = \frac{e^{-\sum_{\alpha=1}^M \lambda_{\alpha}^{\nu^{ij}} a_{ij}^{\alpha}}}{\mathcal{Z}}. \quad (2.150)$$

Finally the probability  $p^{\nu}$  of a  $\nu$ -multilink between any two nodes of the multiplex network is given by

$$p^{\nu} = \frac{1}{\mathcal{Z}} \prod_{\alpha=1}^M \left( \frac{e^{-\lambda_{\alpha}^{\nu}}}{1 - e^{-\lambda_{\alpha}^{\nu}}} \right)^{m_{\alpha}}, \quad (2.151)$$

while we have that the average weight of a  $\nu$  multilink is given by

$$\begin{aligned} \langle a_{ij}^{\alpha} A_{ij}^{\nu} \rangle &= \frac{S_{\alpha}^{\nu}}{\binom{N}{2}} = \frac{1}{\mathcal{Z}} \left( \frac{1}{1 - e^{-\lambda_{\alpha}^{\nu}}} \right) \times \\ &\times \sum_{\vec{m} | \nu(\vec{m}) = \nu} m_{\alpha} \prod_{\beta=1}^M \left( \frac{e^{-\lambda_{\beta}^{\nu}}}{1 - e^{-\lambda_{\beta}^{\nu}}} \right)^{m_{\beta}}. \end{aligned} \quad (2.152)$$

### **Multiplex ensembles with given expected multidegree sequence $\{k_i^{\vec{m}}\}$ and given expected multistrength sequence $\{s_{i,\alpha}^{\vec{m}}\}$**

In many applications it is important to consider the weighted multiplex networks in which we fix at the same time the average multidegree sequence  $k_i^{\vec{m}}$  and the average multistrength sequence  $s_{i,\alpha}^{\vec{m}}$ . The number of independent constraints is therefore  $K = (2^M - 1) \cdot N + (2^{M-1}) \cdot M \cdot N$ .

In particular, the constraints we are imposing are the following,

$$\begin{aligned}\sum_{\vec{G}} F_{i,\alpha}^{\vec{m}}(\vec{G})P(\vec{G}) &= \sum_{\vec{G}} \left( \sum_{j \neq i} A_{ij}^{\vec{m}} a_{ij}^\alpha \right) P(\vec{G}) = s_{i,\alpha}^{\vec{m}} \\ \sum_{\vec{G}} F_i^{\vec{m}}(\vec{G})P(\vec{G}) &= \sum_{\vec{G}} \left( \sum_{j \neq i} A_{ij}^{\vec{m}} \right) P(\vec{G}) = k_i^{\vec{m}}.\end{aligned}\quad (2.153)$$

The canonical probability  $P(\vec{G})$  of the multiplex in the ensemble becomes

$$\begin{aligned}P(\vec{G}) &= \frac{1}{Z} \exp \left[ - \sum_{\vec{m} \neq \vec{0}} \sum_i \sum_{j \neq i} \left( \omega_i^{\vec{m}} A_{ij}^{\vec{m}} + \sum_{\alpha=1}^M \lambda_{i,\alpha}^{\vec{m}} A_{ij}^{\vec{m}} a_{ij}^\alpha \right) \right] \\ &= \frac{1}{Z} \exp \left[ - \sum_{i < j} \sum_{\vec{m} \neq \vec{0}} (\omega_i^{\vec{m}} + \omega_j^{\vec{m}}) A_{ij}^{\vec{m}} \right] \times \\ &\times \exp \left[ - \sum_{i < j} \sum_{\vec{m} \neq \vec{0}} \sum_{\alpha=1}^M (\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}}) A_{ij}^{\vec{m}} a_{ij}^\alpha \right]\end{aligned}\quad (2.154)$$

The partition function  $Z$  can be expressed explicitly as

$$Z = \prod_{i < j} \mathcal{Z}_{ij} \quad (2.155)$$

where  $\mathcal{Z}_{ij}$  is given by

$$\mathcal{Z}_{ij} = 1 + \sum_{\vec{m} \neq \vec{0}} e^{-(\omega_i^{\vec{m}} + \omega_j^{\vec{m}})} \prod_{\alpha=1}^M \left( \frac{e^{-(\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}})}}{1 - e^{-(\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}})}} \right)^{m_\alpha} \quad (2.156)$$

The Lagrangian multipliers are fixed by the conditions

$$\begin{aligned}-\frac{\partial \log Z}{\partial \lambda_{i,\alpha}^{\vec{m}}} &= s_{i,\alpha}^{\vec{m}} = \sum_{j \neq i} \langle a_{ij}^\alpha A_{ij}^{\vec{m}} \rangle, \\ -\frac{\partial \log Z}{\partial \omega_i^{\vec{m}}} &= k_i^{\vec{m}} = \sum_{j \neq i} \langle A_{ij}^{\vec{m}} \rangle.\end{aligned}\quad (2.157)$$

We can calculate the probability of a vector  $\vec{a}_{ij} = (a_{ij}^1, a_{ij}^2, \dots, a_{ij}^M)$  characterizing the weights of the links between node  $i$  and node  $j$  in all the layers, getting

$$\pi_{ij}(\vec{a}_{ij}) = \frac{e^{-(\omega_i^{\vec{m}^{ij}} + \omega_j^{\vec{m}^{ij}})}}{\mathcal{Z}_{ij}} e^{-\sum_{\alpha=1, M} (\lambda_{i,\alpha}^{\vec{m}^{ij}} + \lambda_{j,\alpha}^{\vec{m}^{ij}}) a_{ij}^\alpha} \quad (2.158)$$

Further on we can compute the average weight of the link  $ij$  on the multilink  $\vec{m}$ , in the layer  $\alpha$  and the probability of a multilink  $\vec{m}$  between node  $i$  and node  $j$ ,  $p_{ij}^{\vec{m}} = \langle A_{ij}^{\vec{m}} \rangle$ , respectively,

$$\begin{aligned} \langle a_{ij}^\alpha A_{ij}^{\vec{m}} \rangle &= \frac{e^{-(\omega_i^{\vec{m}} + \omega_j^{\vec{m}})}}{\mathcal{Z}_{ij}} \left( \frac{1}{1 - e^{-(\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}})}} \right) \times \\ &\times \prod_{\beta=1}^M \left( \frac{e^{-(\lambda_{i,\beta}^{\vec{m}} + \lambda_{j,\beta}^{\vec{m}})}}{1 - e^{-(\lambda_{i,\beta}^{\vec{m}} + \lambda_{j,\beta}^{\vec{m}})}} \right)^{m_\beta} \end{aligned} \quad (2.159)$$

$$p_{ij}^{\vec{m}} = \frac{e^{-(\omega_i^{\vec{m}} + \omega_j^{\vec{m}})}}{\mathcal{Z}_{ij}} \prod_{\alpha=1}^M \left( \frac{e^{-(\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}})}}{1 - e^{-(\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}})}} \right)^{m_\alpha} \quad (2.160)$$

Finally, the probability  $P(\vec{G})$  of a multiplex network  $\vec{G}$  in this ensemble is given by Eq. 2.137 and the entropy of the ensemble takes the simple expression given by Eq. 2.145, with the marginal probabilities from Eq. 2.158

### **Multiplex ensembles with given expected $\nu$ -multidegree sequence $\{k_i^\nu\}$ and expected $\nu$ -multistrength sequence $\{s_{i,\alpha}^\nu\}$**

In a multiplex networks formed by many layers, an efficient way to consider both topological and weighted properties of the multilayer structure is to construct multiplex networks with given expected  $\nu$ -multidegree sequence  $\{k_i^\nu\}$  and expected  $\nu$ -multistrength sequence  $\{s_{i,\alpha}^\nu\}$ . The  $N \cdot M \cdot (M + 1)$  constraints are

given by

$$\begin{aligned}\sum_{\vec{G}} F_{i,\alpha}^\nu(\vec{G})P(\vec{G}) &= \sum_{\vec{G}} \left( \sum_{j \neq i} a_{ij}^\alpha A_{ij}^\nu \right) P(\vec{G}) = s_{i,\alpha}^\nu \\ \sum_{\vec{G}} F_i^\nu(\vec{G})P(\vec{G}) &= \sum_{\vec{G}} \left( \sum_{j \neq i} A_{ij}^\nu \right) P(\vec{G}) = k_i^\nu,\end{aligned}\quad (2.161)$$

with  $i = 1, 2, \dots, N$ ,  $\alpha = 1, 2, \dots, M$  and  $\nu = 1, 2, \dots, M$ . The canonical probability  $P(\vec{G})$  of the multiplex in this ensemble can be expressed in terms of the Lagrangian multipliers  $\lambda_{j,\alpha}^\nu$  and  $\omega_i^\nu$ , i.e.

$$\begin{aligned}P(\vec{G}) &= \frac{1}{Z} \exp \left[ - \sum_{i < j} \sum_{\nu=1}^M (\omega_i^\nu + \omega_j^\nu) A_{ij}^\nu \right] \times \\ &\times \exp \left[ - \sum_{i < j} \sum_{\nu=1}^M \sum_{\alpha=1}^M (\lambda_{i,\alpha}^\nu + \lambda_{j,\alpha}^\nu) A_{ij}^\nu a_{ij}^\alpha \right],\end{aligned}\quad (2.162)$$

where the partition function  $Z$  is given by

$$Z = \prod_{i < j} \mathcal{Z}_{ij}, \quad (2.163)$$

with

$$\mathcal{Z}_{ij} = 1 + \sum_{\nu=1}^M e^{-(\omega_i^\nu + \omega_j^\nu)} \sum_{\vec{m} | \nu(\vec{m}) = \nu} \prod_{\alpha=1}^M \left( \frac{e^{-(\lambda_{i,\alpha}^\nu + \lambda_{j,\alpha}^\nu)}}{1 - e^{-(\lambda_{i,\alpha}^\nu + \lambda_{j,\alpha}^\nu)}} \right)^{m_\alpha} \quad (2.164)$$

The Lagrangian multipliers are fixed by the conditions Eq. 2.161 that can be also written in terms of the partial derivatives of the partition function as

$$\begin{aligned}- \frac{\partial \log Z}{\partial \lambda_{i,\alpha}^\nu} &= s_{i,\alpha}^\nu = \sum_{j \neq i} \langle a_{ij}^\alpha A_{ij}^\nu \rangle, \\ - \frac{\partial \log Z}{\partial \omega_i^\nu} &= k_i^\nu = \sum_{j \neq i} \langle A_{ij}^\nu \rangle.\end{aligned}\quad (2.165)$$

As in the previous cases, the probability  $P(\vec{G})$  of a multiplex network  $\vec{G}$  is given

by Eq. 2.137. The entropy of this ensemble takes the same expression given by Eq. 2.145 with  $\pi_{ij}(\vec{a}_{ij})$  given by

$$\pi_{ij}(\vec{a}_{ij}) = \frac{e^{-(\omega_i^{\nu ij} + \omega_j^{\nu ij})}}{\mathcal{Z}_{ij}} e^{-\sum_{\alpha=1, M} (\lambda_{i, \alpha}^{\nu ij} + \lambda_{j, \alpha}^{\nu ij}) a_{ij}^{\alpha}} \quad (2.166)$$

The probability  $p_{ij}^{\nu}$  that the node  $i$  and the node  $j$  are linked by a  $\nu$ -multilink is given by

$$p_{ij}^{\nu} = \frac{e^{-(\omega_i^{\nu} + \omega_j^{\nu})}}{\mathcal{Z}_{ij}} \sum_{\vec{m} | \nu(\vec{m}) = \nu} \prod_{\alpha=1}^M \left( \frac{e^{-(\lambda_{i, \alpha}^{\nu} + \lambda_{j, \alpha}^{\nu})}}{1 - e^{-(\lambda_{i, \alpha}^{\nu} + \lambda_{j, \alpha}^{\nu})}} \right)^{m_{\alpha}} \quad (2.167)$$

Finally, the average weight of the link  $a_{ij}^{\alpha}$  belonging to a  $\nu$ -multilink is given by

$$\begin{aligned} \langle a_{ij}^{\alpha} A_{ij}^{\nu} \rangle &= \frac{e^{-(\omega_i^{\nu} + \omega_j^{\nu})}}{\mathcal{Z}_{ij}} \left( \frac{1}{1 - e^{-(\lambda_{i, \alpha}^{\nu} + \lambda_{j, \alpha}^{\nu})}} \right) \times \\ &\times \sum_{\vec{m} | \nu(\vec{m}) = \nu} m_{\alpha} \prod_{\beta=1}^M \left( \frac{e^{-(\lambda_{i, \beta}^{\nu} + \lambda_{j, \beta}^{\nu})}}{1 - e^{-(\lambda_{i, \beta}^{\nu} + \lambda_{j, \beta}^{\nu})}} \right)^{m_{\beta}} \end{aligned} \quad (2.168)$$

### Sampling multiplex ensembles with given expected multidegree sequence

$\{k_i^{\vec{m}}\}$  and given expected multistrength sequence  $\{s_{i, \alpha}^{\vec{m}}\}$

Here we want to discuss how the theoretical framework described in the previous section can be used to generate weighted multiplex networks sampled from a multiplex network ensemble. We have chosen to focus specifically on the case of a multiplex network ensemble in which the given expected multidegree sequence  $\{k_i^{\vec{m}}\}$  and the given expected multistrength sequence  $\{s_{i, \alpha}^{\vec{m}}\}$  are constrained, but the framework we outline here of this case can be easily extended to the other ensembles discussed in the previous section. Given Eqs. 2.160, 2.158, the probability  $\pi_{ij}(\vec{a}_{ij})$  can be expressed as a function of the probability  $p_{ij}^{\vec{m}}$  of a multilink

$\vec{m}$  between node  $i$  and node  $j$ , namely

$$\pi_{ij}(\vec{a}_{ij}) = p_{ij}^{\vec{m}^{ij}} \prod_{\alpha=1}^M \left( \left[ e^{-(\lambda_{i,\alpha}^{\vec{m}^{ij}} + \lambda_{j,\alpha}^{\vec{m}^{ij}})} \right]^{a_{ij}^{\alpha} - 1} \left[ 1 - e^{-(\lambda_{i,\alpha}^{\vec{m}^{ij}} + \lambda_{j,\alpha}^{\vec{m}^{ij}})} \right] \right)^{m_{\alpha}^{ij}} \quad (2.169)$$

The productory in Eq. 2.169 is the conditional probability of the multiweight  $\vec{a}_{ij}$ , given the multilink  $\vec{m}^{ij}$ . The new expression for  $\pi_{ij}(\vec{a}_{ij})$  suggests a way for sampling networks from the distribution given by Eq. 2.137, with  $\pi_{ij}(\vec{a}_{ij})$  given by Eq. 2.169. In fact for sampling a multiplex network from this particular ensemble, we draw a multilink  $\vec{m}$  with probability  $p_{ij}^{\vec{m}}$  for each couple of nodes  $i$  and  $j$ . Subsequently, given a particular multilink, whenever  $m_{\alpha} = 1$  we draw the additional weight  $a_{ij}^{\alpha} - 1$  from a geometric distribution with parameter  $1 - e^{-(\lambda_{i,\alpha}^{\vec{m}^{ij}} + \lambda_{j,\alpha}^{\vec{m}^{ij}})}$  and  $a_{ij}^{\alpha} \geq 1$ .

Following Eqs. 2.157 we wrote a Matlab code that produces the Lagrangian multipliers and calculates the entropy value of the ensemble. The algorithm runs until it finds convergence with precision  $10^{-4}$  (this value can be always improved).

## 2.4 Weighted real data

Just as with single networks, links between nodes may have a different weight, reflecting their intensity, capacity, duration, intimacy or exchange of services (Granovetter, 1973). The role played by the weights in the functioning of many networks, and especially the relative benefits of weak and strong ties in social networks, have been the subject of a longstanding debate (Granovetter, 1973; Onnela et al., 2007; Karsai et al., 2014). Moreover, it has been shown that, in single networks, the weights can be distributed in a heterogeneous way, as

a result of the non-trivial effects that the structural properties of the networks have on them (Barrat et al., 2004; Barthelemy et al., 2003; Almaas et al., 2004; Serrano et al., 2009). In particular, correlations between weights and structural properties of single networks can be uncovered by the analysis of strength-degree correlations (Barrat et al., 2004) and by the distribution of the weights of the links incident upon the same node Almaas et al. (2004). To characterize weighted networks, it is common practice to measure the following quantities: i) the average strength of nodes of degree  $k$ , i.e.  $s = s(k)$ , describing how weights are distributed in the network; and ii) the average inverse participation ratio of the weights of the links incident upon nodes of degree  $k$ , i.e.  $Y = Y(k)$ , describing how weights are distributed across the links incident upon nodes of degree  $k$ . Here we show that these two quantities do not capture the full breadth of the information encoded in multiplex networks. Indeed, a full-fledged analysis of the properties of multiplex networks is needed that takes the multiple interacting and co-evolving layers simultaneously into account.

In particular, in the following section different multiplex networks have been extracted from the APS dataset in order to investigate the correlation between the weights of the links and the overlap of the links in different layers. In the multiplex networks formed by the PRE authors in which the scientists are linked if they collaborated with each other and if they cite each other, has been shown to display a statistical significant difference between the way scientists cite their collaborators and the way scientists cite non-collaborators. This result shows that in this as in other systems it is possible that the weights of the links are correlated with the pattern of overlap observed between the links of different layers.

### 2.4.1 APS dataset

To provide empirical evidence that weighted properties of multilinks are fundamental for properly assessing weighted multiplex networks, we focus on the networks of the authors of papers published in the journals of the American Physical Society (APS), and analyse the scientific collaboration network and the citation network connecting the same authors. These networks are intrinsically weighted since any two scientists can co-author more than one paper and can cite each other's work several times. A large number of studies have analysed similar bibliometric datasets drawing upon network theory (Redner, 1998; Park and Newman, 2004; Newman, 2001; Radicchi et al., 2009, 2008). Unlike these studies, here we investigate the APS bibliometric dataset using the framework of multiplex networks that allows us to explore novel properties of the collaboration and citation networks. In particular, we show that multistrength and the inverse multiparticipation ratio enable new relevant information to be extracted from the APS dataset and that this information extends beyond what is encoded in the strength and inverse participation ratio of single layers. Finally, based on the entropy of multiplex ensembles, we propose an indicator  $\Xi$  to evaluate the additional amount of information that can be extracted from the weighted properties of multilinks in multiplex networks over the information encoded in the properties of their individual layers analysed separately.

#### **Empirical evidence of weighted properties of multilinks**

In this section, we will draw on the measures introduced above and provide empirical evidence that, in weighted multiplex networks, weights can be correlated with the multiplex structure in a non-trivial way. To this end, we analyze

the bibliographic dataset that includes all articles published in the APS journals (i.e., Physical Review Letters, Physical Review, and Reviews of Modern Physics) from 1893 to 2009. Of these articles, the dataset includes their citations as well as the authors. Here, we restrict our study only to articles published either in Physical Review Letters (PRL) or in Physical Review E (PRE) and written by ten or fewer authors,  $n_p \leq 10$ . We constructed multiplex networks in which the nodes are the authors and links between them have a two-fold nature: scientific collaborations with weights defined as in Newman (2001) (see Sec. 4.1), and citations with weights indicating how many times author  $i$  cited author  $j$ .

In particular, we created the following two duplex networks (i.e. multiplex networks with  $M = 2$ ):

1. **CoCo-PRL/PRE:** *collaborations among PRL and PRE authors.* The nodes of this multiplex network are the authors with articles published both in PRL and PRE (i.e., 16,207 authors). These nodes are connected in layer 1 through weighted undirected links indicating the strength of their collaboration in PRL (i.e., co-authorship of PRL articles). The same nodes are connected in layer 2 through weighted undirected links indicating the strength of their collaboration in PRE (i.e., co-authorship of PRE articles).
2. **CoCi-PRE:** *collaborations among PRE authors and citations to PRE articles.* The nodes of this multiplex network are the authors of articles published in PRE (i.e., 35,205 authors). These nodes are connected in layer 1 through weighted undirected links indicating the strength of their collaboration in PRE (i.e., co-authorship of PRE articles). The same nodes are connected in layer 2 through weighted directed links indicating how many times an author (with articles in PRE) cited another author's work, where

citations are limited to those made to PRE articles.

Both these multiplex networks show a significant overlap of links and a significant correlation between degrees of nodes as captured by the Pearson correlation coefficient  $\rho$  (see Sec. 4.1). This finding supports the hypothesis that the two layers in each of the multiplex networks are correlated. That is, the existence of a link between two authors in one layer is correlated with the existence of a link between the same authors in the other layer. Moreover, the multidegrees of the multiplex networks are broadly distributed, and the hubs in the scientific collaboration network tend to be also the hubs in the citation network (see Sec. 4.1).

In the case of the CoCo-PRL/PRE network, multilinks  $\vec{m} = (1, 0)$ ,  $\vec{m} = (0, 1)$  and  $\vec{m} = (1, 1)$  refer to collaborations only in PRL, only in PRE, and in both PRL and PRE, respectively. Moreover, to distinguish between the weights used when evaluating multistrength, we have  $\alpha = PRL$  or  $\alpha = PRE$ . Results indicate that multistrength and the inverse multiparticipation ratio behave according to Eq. 2.22 (see Fig. 2.4). The difference between exponents  $\beta_{\vec{m}, PRL}$  for  $\vec{m} = (1, 0)$  and  $\vec{m} = (1, 1)$  is not statistically significant. Nevertheless, there is a statistically significant difference between the average weights of multilinks  $(1, 0)$  and  $(1, 1)$  in the PRL layer. As to the inverse multiparticipation ratio, there is a significant variation in the exponents,  $\lambda_{(1,0), PRL} = 0.84 \pm 0.03$  and  $\lambda_{(1,1), PRL} = 0.74 \pm 0.05$  (see Fig. 2.4, bottom left panel). This suggests that the weights of the collaborative links between co-authors of both PRL and PRE articles are distributed more heterogeneously than the weights of collaborative links between co-authors of articles published only in PRL (see Sec. 4.1 for details on the statistical tests). Similar results were found for multistrengths evaluated in the PRE layer (see

Fig. 2.4, right panels).

These findings clearly indicate that the partial analysis of individual layers would fail to uncover the fact that the average weight of the link between authors that collaborated both on PRL and PRE articles is significantly larger than the average weight of the link between authors that collaborated only on articles published in one journal. Moreover, the difference in functional behaviour of the multipartition ratio across layers could not be captured if layers were analysed separately.

In the case of the CoCi-PRE network, there are even more significant differences between the properties of the multilinks than in the previous network. In the CoCi-PRE network the functional behaviour of multistrength also depends on the type of multilink. Figure 2.5 shows the average multistrength in the CoCi-PRE network. To distinguish between the weights used to measure multistrength, we have layer  $\alpha = col$ , which refers to the collaboration network constructed on PRE articles, and layer  $\alpha = cit$ , which refers to the citation network between PRE articles, where a distinction is also made between incoming (*in*) and outgoing (*out*) links. First, in the scientific collaboration network, exponents  $\beta_{\vec{m},col}$  are not statistically different, but the average weight of multilink  $(1, 1)$  is larger than the average weight of multilinks  $(1, 0), in$  and  $(1, 0)out$ . Moreover, exponents  $\lambda_{(1,0),col,in}$  and  $\lambda_{(1,0),col,out}$  are larger than exponents  $\lambda_{(1,1),col,in}$ ,  $\lambda_{(1,1),col,out}$ , indicating that the weights of authors' collaborative links with other cited/citing authors are distributed more heterogeneously than the weights of authors' collaborative links with other authors with whom there are no links in the citation network. Second, in the citation network multistrengths follow a distinct functional behaviour depending on the different type of multilink, and are characterised by

different  $\beta_{\vec{m},cit,in/out}$  exponents. In fact the fitted values of these exponents are given by  $\beta_{(1,1)cit,in} = 1.30 \pm 0.07$ ,  $\beta_{(1,1)cit,out} = 1.32 \pm 0.08$ ,  $\beta_{(0,1)cit,in} = 1.11 \pm 0.01$ , and  $\beta_{(0,1)cit,out} = 1.10 \pm 0.02$ . This implies that, on average, highly cited authors are cited by their co-authors to a much greater extent than is the case with poorly cited authors. A similar, though much weaker effect was also found for the citations connecting authors that are not collaborators. Furthermore, in the citation layer the inverse multiparticipation ratio for multilink (1, 1) is always larger than the inverse multiparticipation ratio for multilinks (1, 0) and (0, 1) (see Sec. 4.1 for details on the statistical test). Finally, when single layers were analysed separately, we found  $\beta_{col} = 1.03 \pm 0.04$  in the collaboration network, and  $\beta_{cit,in} = 1.13 \pm 0.02$  and  $\beta_{cit,out} = 1.14 \pm 0.03$  in the citation network. This indicates that in the citation network strength grows super-linearly as a function of degree, i.e., weights are not distributed uniformly. Nevertheless, correlations between weights and types of multilinks cannot be captured if the two individual layers are studied separately.

### **Assessing the informational content of weighted multilinks**

Recent research on single networks has shown that the entropy of network ensembles provides a very powerful tool for quantifying their complexity (Park and Newman, 2004; Johnson et al., 2010; Bianconi et al., 2009). Here, we propose a theoretical framework based on the entropy of multiplex ensembles for assessing the amount of information encoded in the weighted properties of multilinks. Multiplex weighted network ensembles can be defined as the set of all weighted multiplex networks satisfying a given set of constraints, such as the expected degree sequence and the expected strength sequence in every layer of the multiplex

network, or the expected multidegree sequence and the expected multistrength sequence. As we showed in the previous sections, a set of constraints imposed upon the multiplex network ensemble uniquely determines the probability  $P(\vec{G})$  of the multiplex networks in the ensemble. The entropy  $\mathcal{S}$  of the multiplex ensemble follows from Eq. 2.33 and it indicates the logarithm of the typical number of multiplex networks in the ensemble. The smaller the entropy, the larger the amount of information stored in the constraints imposed on the network. The entropy can be regarded as an unbiased way to evaluate the informational value of these constraints.

In order to gauge the information encoded in a weighted multiplex network with respect to a null model, we define the indicator  $\Psi$ , which quantifies how much information is carried by the weight distributions of a weighted multiplex ensemble. In particular,  $\Psi$  compares the entropy of a weighted multiplex ensemble  $\mathcal{S}$  with the entropy of a weighted multiplex ensemble in which the weights are distributed homogeneously. Therefore,  $\Psi$  can be defined as

$$\Psi = \frac{|\mathcal{S} - \langle \mathcal{S} \rangle_{\pi(w)}|}{\langle (\delta \mathcal{S})^2 \rangle_{\pi(w)}}, \quad (2.170)$$

where  $\langle (\delta \mathcal{S})^2 \rangle_{\pi(w)}$  is the standard deviation, and the average  $\langle \dots \rangle_{\pi(w)}$  is calculated over multiplex networks with the same structural properties but with weights distributed homogeneously. In particular, when the weight distribution is randomized, the multiplex networks are constrained in such a way that each link must have a minimal weight (i.e.,  $w_{ij} \geq 1$ ), while the remaining of the total weight is distributed randomly over the links. In all the considered network ensembles we have assumed that the weights of the links can only take values that

are multiple of a minimal weight. This assumption is by no means a limitation of this approach because for every finite network, there is always a minimal weight in the network such that this hypothesis is verified.

In order to evaluate the amount of information encoded in the weight of links in single layers and compare it to the information supplied by multistrength, we consider the following undirected multiplex ensembles:

- *Correlated weighted multiplex ensemble.* In this ensemble, we fix the expected multidegree sequence  $\{k_i^{\vec{m}}\}$ , and we set the expected multistrength sequence  $\{s_i^{\vec{m},\alpha}\}$  to be

$$s_i^{\vec{m},\alpha} = c_{\vec{m},\alpha} (k_i^{\vec{m},\alpha})^{\lambda_{\vec{m},\alpha}} \quad (2.171)$$

for every layer  $\alpha$ . We call  $\Psi^{corr}$  the  $\Psi$  calculated from this ensemble.

- *Uncorrelated weighted multiplex ensemble.* In this ensemble, we set the expected degree  $k_i^\alpha$  of every node  $i$  in every layer  $\alpha = 1, 2$  to be equal to the sum of the multidegrees (with  $m_\alpha = 1$ ) in the correlated weighted multiplex ensemble. We set the expected strengths  $s_i^\alpha$  of every node  $i$  in every layer  $\alpha$  to be equal to the sum of the multistrengths of node  $i$  in layer  $\alpha$  in the correlated weighted multiplex ensemble. We call  $\Psi^{corr}$  the  $\Psi$  calculated from this ensemble.

In the correlated weighted multiplex ensemble the properties of the multilinks are accounted for, while in the uncorrelated weighted multiplex ensemble the different layers of the multiplex networks are analysed separately (see see Sec. 4.1 for the details). Finally, to quantify the additional amount of information carried by the correlated multiplex ensemble with respect to the uncorrelated

multiplex ensemble, we define the indicator  $\Xi$  as

$$\Xi = \frac{\Psi^{corr}}{\Psi^{uncorr}}. \quad (2.172)$$

As an example of a possible application of the indicator  $\Xi$ , we focus on a case inspired by the CoCi-PRE multiplex network, where we consider different exponents  $\beta_{\vec{m},\alpha,in/out}$  for different multilinks. First, we created the correlated multiplex ensemble with power-law multidegree distributions  $P(k^{\vec{m}}) = C(k^{\vec{m}})^{-\gamma_{\vec{m}}}$  with exponents  $\gamma_{(1,m_2)} = 2.6$  for  $m_2 = 0, 1$  and  $\gamma_{(0,1),(in/out)} = 1.9$  (where for multidegree  $(0, 1)$  we imposed a structural cut-off). Multistrengths satisfy Eq. 4.11, with  $c_{\vec{m},\alpha} = 1$  and  $\beta_{(1,m_2),1} = 1$ , for  $m_2 = 0, 1$ ;  $\beta_{(1,1),2} = 1.3$ , and  $\beta_{(0,1),2} = 1.1$ . Second, for the second layer, we created the uncorrelated version of the multiplex ensemble which is characterized by a super-linear dependence of the average strength on the degree of the nodes. We then measured  $\Psi$  as a function of network size  $N$  for these different ensembles. Numerically, the average  $\langle \dots \rangle_{\pi(w)}$  was evaluated from 100 randomizations. Figure 2.6 shows that  $\Psi$  increases with network size  $N$  as a power law, and that  $\Xi$  fluctuates around an average value of 1.256. These findings indicate that a significant amount of information is contained in multistrength and cannot be extracted from individual layers separately. Similar results, not shown here, were obtained with a correlated weighted multiplex ensemble characterized by non-trivial inverse multiparticipation ratios.

## 2.4.2 Gene-Gene Duplex

In this section we analyse a dataset of gene expression profiles from human cancer and healthy subjects, using the framework of multilayer networks. For

this analysis, we construct a duplex based on whole-genome gene expression data, as taken from Geo Omnibus Database (NCBI (2014), *GSE4183* dataset). We have already met this dataset in Sec. 1.2.1 and we referred to it as “Colon”. From this dataset, a subset of 2835 genes was chosen, known to have a clear biological role (i.e. belonging to known functional pathways as annotated in the KEGG database KEGG (2014)) and with potential interactions between each other (as annotated in PathwayCommons Protein-Protein Interaction network database, Commons (2014)). In one layer, the network is reconstructed from gene expression correlation of  $N_N = 8$  normal colon samples, while in the other layer  $N_C = 15$  cancer samples are considered. We define as  $e_{ij}^\alpha$  the gene expression value for layer  $\alpha$  (normal N or cancer C), in which  $i$  is the gene index (ranging from 1 to 2835) and  $j$  refers to the sample (ranging from 1 to 8 for normal samples dataset, and from 1 to 15 for cancer samples dataset).

Nonparametric Kendall’s  $\tau$  is used in order to evaluate the correlation between genes, and in each layer a network is obtained by a thresholding on the absolute value of  $\tau$  that keeps about  $\approx 10\%$  of the possible links ( $\tau_1 = 0.5$  and  $\tau_2 = 0.4$  for normal and cancer samples respectively).

We also associate a weight  $a_{ij}^\alpha$  to each duplex link, obtained from gene expression values of the normal and cancer groups. We calculate the average value over all samples for each gene in both layers, namely,

$$\langle e_i^\alpha \rangle = \frac{1}{N_\alpha} \sum_{k=1}^{N_\alpha} e_{ik}^\alpha \quad (2.173)$$

with  $\alpha = N, C$  and define the weights on each layer as the absolute difference between all gene couples

$$a_{ij}^\alpha = |\langle e_i^\alpha \rangle - \langle e_j^\alpha \rangle| \quad \forall i, j = 1, \dots, 2835. \quad (2.174)$$

The weights have been discretized as follows: given the minimum and maximum over all values of  $a_{ij}^\alpha$  (from the union of cancer and normal samples distance matrices), we performed a uniform binning with 100 bins in this interval, thus obtaining 100 possible values for the weights  $a_{ij}^\alpha$ . This duplex encodes in its topology all the connections among those genes with highly correlated or anti-correlated gene expression profiles. Moreover, the weight distribution describes their distances in terms of mean gene expression values. These kinds of information are essentially different: for example, two genes can be highly correlated in their trends across the samples but one could be much more expressed than the other one.

We can integrate different aspects of gene expression data sets thanks to network approaches, and furthermore, we can investigate different experimental setups thanks to multiplex networks tools. The analysis of the multiplex network we have constructed, formed by one layer for the normal samples and one layer for those patients with colorectal cancer, can help us understand if there is a backbone of highly correlated genes that are conserved after the onset of the cancer disease. Moreover, we can characterise all the interactions that are specific for the two conditions.

In order to understand how the weights of the links in a selected layer are related to different multilinks we consider the distributions  $\{s_{i,\alpha}^{\vec{m}}/k_i^{\vec{m}}\}$ , i.e. for each

node we calculate the average weight of its interactions, classified according to the multilinks. In Fig. 2.9 we show these distributions, for a given layer  $\alpha = 1, 2$  and a given multilink  $\vec{m}$ . In both layers, the distribution of average weights related to multilink  $(1, 1)$  is significantly different from that one of the specific layer (i.e. multilink  $(1, 0)$  or  $(0, 1)$ ), with a lower mean value and median of the distribution. For layer 1, we compared the distributions  $\{s_{i,1}^{(1,1)}/k_i^{(1,1)}\}$  and  $\{s_{i,1}^{(1,0)}/k_i^{(1,0)}\}$  using a Wilcoxon rank sum test, a nonparametric test for equality of population medians. The p-value is highly significant ( $3.88 \cdot 10^{-22}$ ) and the two mean values are, respectively,  $\langle \{s_{i,1}^{(1,1)}/k_i^{(1,1)}\} \rangle = 19.36$  and  $\langle \{s_{i,1}^{(1,0)}/k_i^{(1,0)}\} \rangle = 20.92$ . For layer 2, the layer related to cancer samples, the rank sum test is always significant but with a less dramatic p-value ( $5.23 \cdot 10^{-8}$ ). The mean values for this layer are respectively  $\langle \{s_{i,2}^{(1,1)}/k_i^{(1,1)}\} \rangle = 19.54$  and  $\langle \{s_{i,2}^{(0,1)}/k_i^{(0,1)}\} \rangle = 20.46$ .

We studied the relation between the weights of the set of overlapping links,  $\{a_{ij,1}^{(1,1)}\}$  and  $\{a_{ij,2}^{(1,1)}\}$ . A linear fitting shows that these weights are almost identical, with a relation  $a_{ij,2}^{(1,1)} = 0.94 \cdot a_{ij,1}^{(1,1)} + 3.30$  ( $R^2 = 0.92$ ). This result is not trivial, since genes could be correlated (preserving the links) but expressed in a different way (i.e. with different weights) in healthy and cancer samples, and highlights the existence of a backbone of genes (and related biological processes) that are conserved during the disease progression, possibly due to their fundamental functional role.

The main goal here is the creation of a null model for such a multiplex real instance, in order to provide an example of possible application of the theoretical framework here developed to model real datasets. In order to generate a

null model, we will construct a network ensemble with given multidegree sequence and multistrength sequence and generate multiplex networks out of this ensemble with the desired structural properties. Sampling multiplex networks from their ensembles will offer the opportunity of comparing our real biological structure with some compatible instances. Moreover, the entropy measure gives us the logarithm of the number of “typical” duplex networks in the ensemble, a value that can be used to compare different experimental setups and clinical conditions, evaluating what is the level of information encoded in the selected structural properties of biological networks.

### **Comparison between the null model and the biological case study**

We compare now the structural properties of our biological case study with the networks with the same multidegree sequence and multistrength sequence generated by sampling the corresponding multiplex network ensemble. Starting from our biological duplex network described in section 2.4.2, at first we calculated the Lagrangian multipliers needed for  $\{p_{ij}^{\vec{m}}\}$  and  $\{\pi_{ij}(\vec{a}_{ij})\}$ , secondly we generated 100 different duplex networks. We checked the average values and fluctuations across our 100 duplexes. In Fig. 2.7 we compare the behavior of the average values across the duplexes with the related real values, the assumed fixed average values of the canonical ensemble. We found that the multidegrees and the multistrengths are equal in average to the constrained values showing that the multiplex network framework is able to reproduce well these properties. Nevertheless from sample to sample the individual structural properties of the nodes (their multidegrees and their multistrengths) might fluctuate. In Figure

2.8 we investigate the role of the fluctuations by plotting the histogram of the  $z$ -scores of values of the multidegrees or of the multistrengths for single nodes, in the layer 2 of the duplex networks, the cancer layer. These distributions are calculated over the 100 multiplex networks sampled by this ensemble.

## 2.5 Conclusions

In this chapter we gave an exhaustive description of multilayer networks, and in particular, of multiplex networks. The main observables and the related null models are displayed. To study the correlations generated by the overlap of the links in different layers we use the multilink-formalism. For the particular case of weighted multiplex networks, we have shown that significant correlations across layers are present, and moreover, that weights are closely correlated with the multiplex network structure. Thank to the introduced observables we proved that many properties of multiplex networks cannot be reduced and predicted by the measures obtained on single layers. These weighted multiplex properties capture the crucial role played by multilinks in the distribution of weights, i.e., the extent to which there is a link connecting each pair of nodes in every layer of the multiplex network. To illustrate these findings we presented two datasets. In the first one we analysed the weighted properties of multilinks in two multiplex networks constructed by combining the co-authorship and citation networks involving the authors included in the APS dataset. Based on the entropy of multiplex ensembles, we developed a theoretical framework for evaluating the information encoded in weighted multiplex networks, and proposed the indicator  $\Psi$  for quantifying the information that can be extracted from a given

dataset with respect to a null model in which weights are randomly distributed across links. Finally, we proposed a new indicator  $\Xi$  that can be used to evaluate the additional amount of information that the weighted properties of multilinks provide over the information contained in the properties of single layers.

The second dataset is related to omics studies, i.e. gene expression profiles of healthy people and subjects affected by CRC. We showed that multiplex observables highlight significant differences and nontrivial similarities between biological processes in healthy and cancer cells in this gene expression profiling dataset. Moreover, we tested the performance of our null models on this particular dataset.

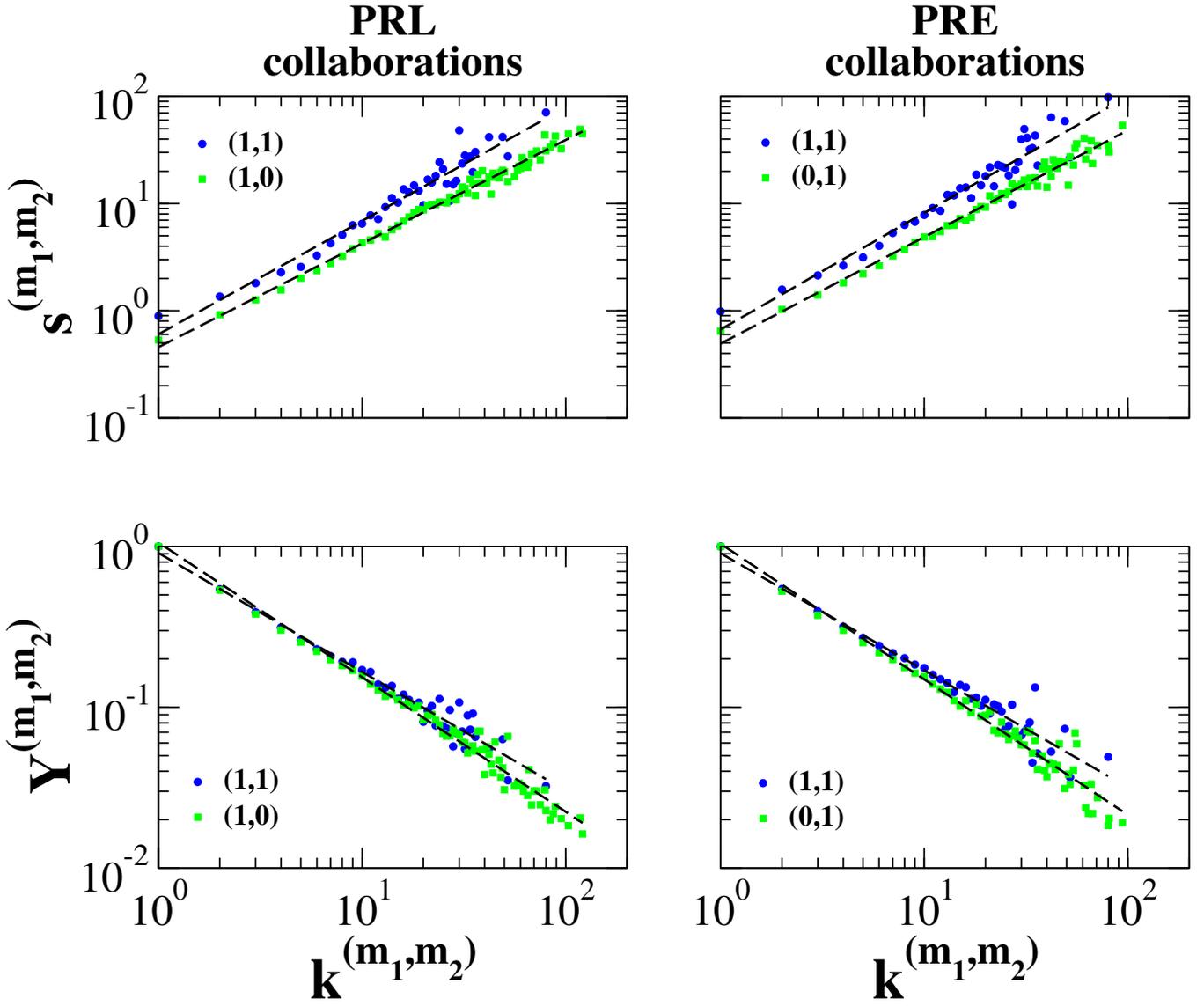


Figure 2.4: Average multistrength and average inverse multiplicity ratio versus multidegree in the CoCo-PRE/PRL multiplex network. The average multistrengths and the average inverse multiplicity ratios are fitted by a power-law distribution of the type described in Eq. 4.8 (fitted distributions are here indicated by black dashed lines). Statistical tests for the collaboration network of PRL suggest that the exponents  $\beta_{\vec{m},1}$  defined in Eq. 4.8 are the same, while exponents  $\lambda_{\vec{m},PRL}$  are significantly different. Similar results can be obtained for the exponents in the PRE collaboration layer. Nevertheless, multistrengths  $s^{(1,1),\alpha}$  are always larger than multistrengths  $s^{(1,0),PRL}$  and  $s^{(0,1),PRE}$ , when multistrengths are calculated over the same number of multilinks, i.e.,  $k^{(1,1)} = k^{(1,0)} = k^{(0,1)}$  (see Sec. 4.1 for the statistical test on this hypothesis).

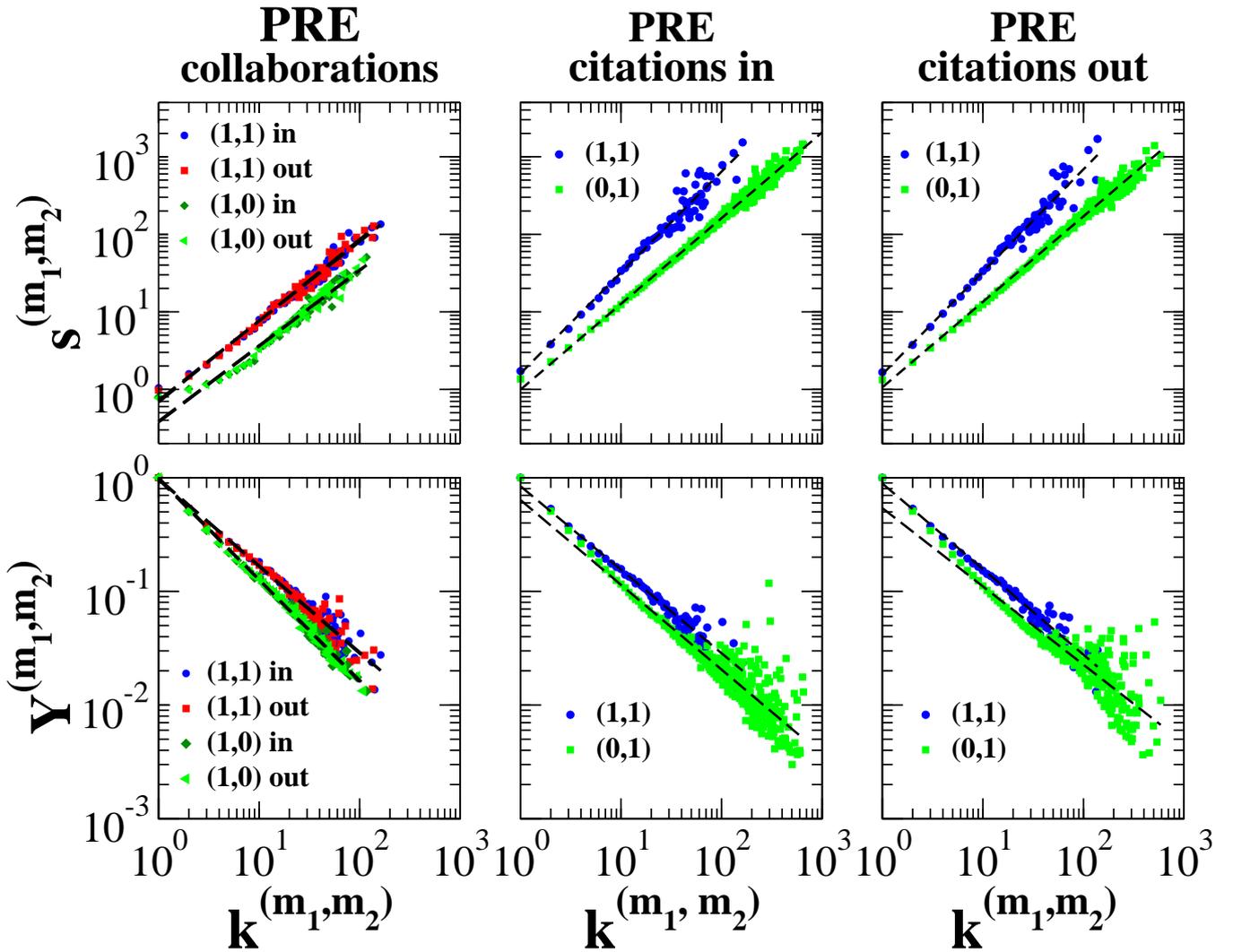


Figure 2.5: Properties of multilinks in the weighted CoCi-PRE multiplex network. In the case of the collaboration network, the distributions of multistrengths versus multidegrees always have the same exponent, but the average weight of multilinks (1,1) is larger than the average weight of multilinks (1,0). Moreover, the exponents  $\lambda_{(1,0),col,in}$ ,  $\lambda_{(1,0),col,out}$  are larger than exponents  $\lambda_{(1,1),col,in}$ ,  $\lambda_{(1,1),col,out}$ . In the case of the citation layer, both the incoming multistrengths and the outgoing multistrengths have a functional behavior that varies depending on the type of multilink. Conversely, the average inverse multiplicity ratio in the citation layer does not show any significant change of behavior when compared across different multilinks.

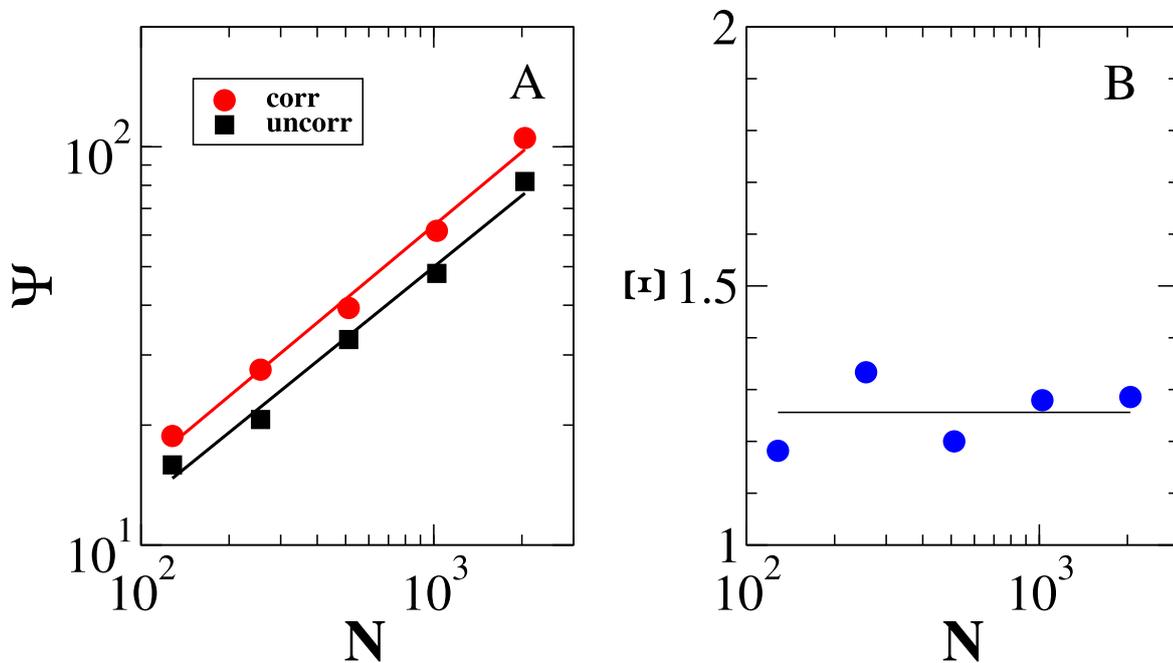


Figure 2.6: (A) Value of the indicator  $\Psi$  defined in Eq. 2.170 indicating the amount of information carried by the correlated and the uncorrelated multiplex ensembles of  $N$  nodes with respect to a null model in which the weights are distributed uniformly over the multiplex network. (B) Value of the indicator  $\Xi$  defined in Eq. 2.172 indicating the additional amount of information encoded in the properties of multilinks in the correlated multiplex ensemble with respect to the corresponding uncorrelated multiplex ensemble. The solid line refers to the average value of  $\Xi$  over the different multiplex network sizes.

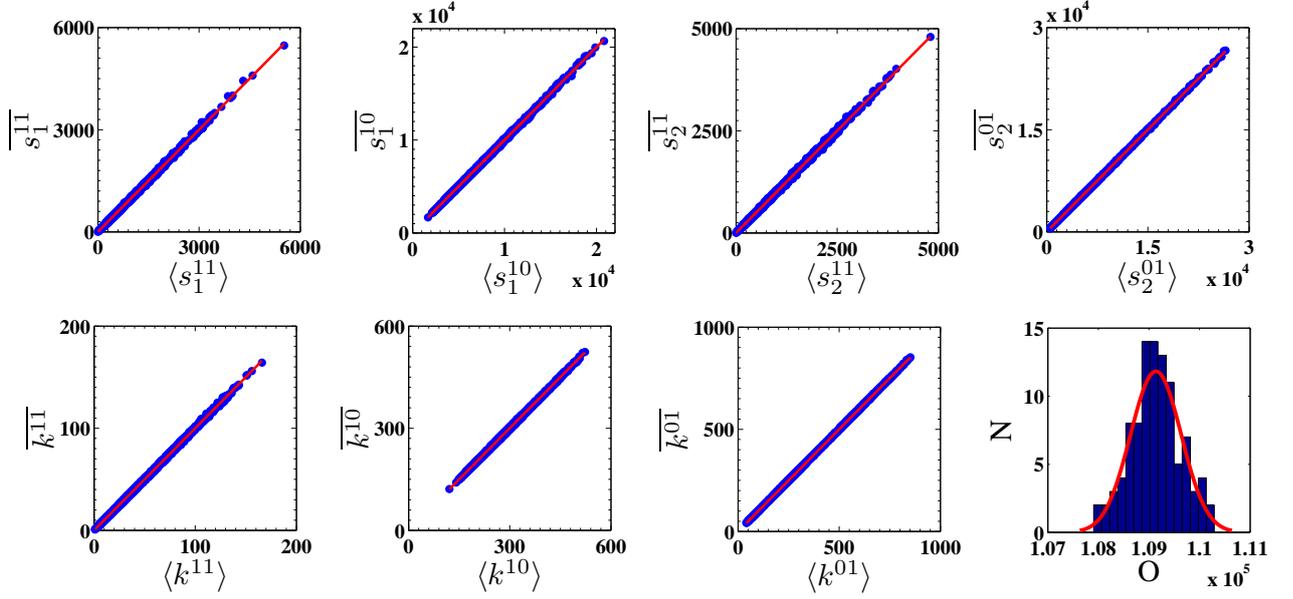


Figure 2.7: Comparison of the real values of multistrength and multidegree sequence with their related average values calculated over 100 instances. Angular brackets ( $\langle \dots \rangle$ ) indicate the real values (the fixed average values of the canonical ensemble), while overbar ( $\overline{\dots}$ ) defines the average measure over the 100 duplexes. Considering the relative error between the real values and the average values for each node,  $\Delta E_i = (\bar{x}_i - \langle x_i \rangle) / \langle x_i \rangle$   $i = 1, \dots, 2, 835$ , the average absolute relative error  $\langle |\Delta E| \rangle$ , over all nodes for each measure, ranges from a minimum of 0.5% to a maximum of 2.4%. In the last panel we display the distribution of the 100 measures of the overlap between the two layers (in the real duplex this value was 109,056 links). The red line is a Gaussian distribution with the same mean and variance as the empirical distribution.

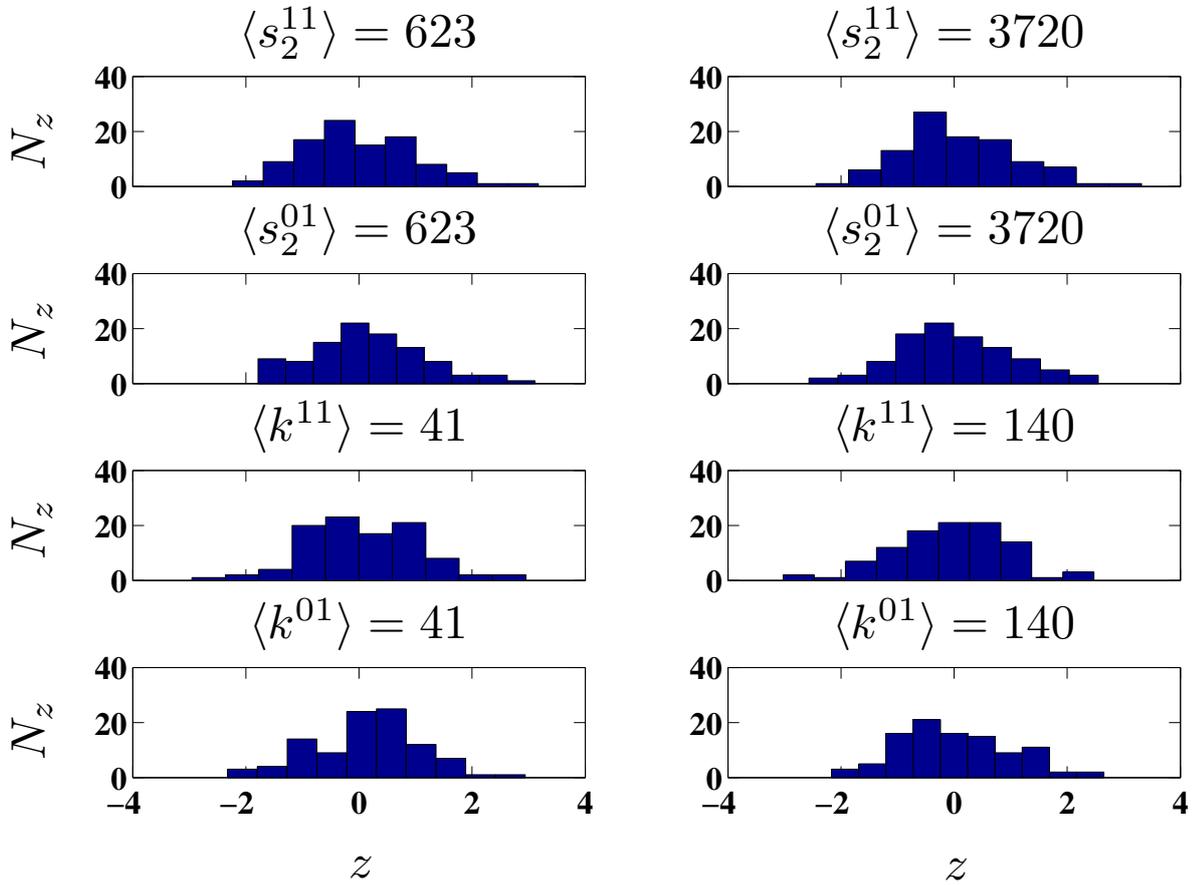


Figure 2.8: Distributions of the  $z$ -scores  $\{z_i\}$  related to some fixed values of multistrength and multidegree, in Layer 2 (colorectal cancer layer). In each panel we display the 100 values of  $z$  across the sampled instances (gathered in 10 bins), for a chosen node with that assigned value of multistrength or multidegree. Similar results are also found for Layer 1 (normal samples).

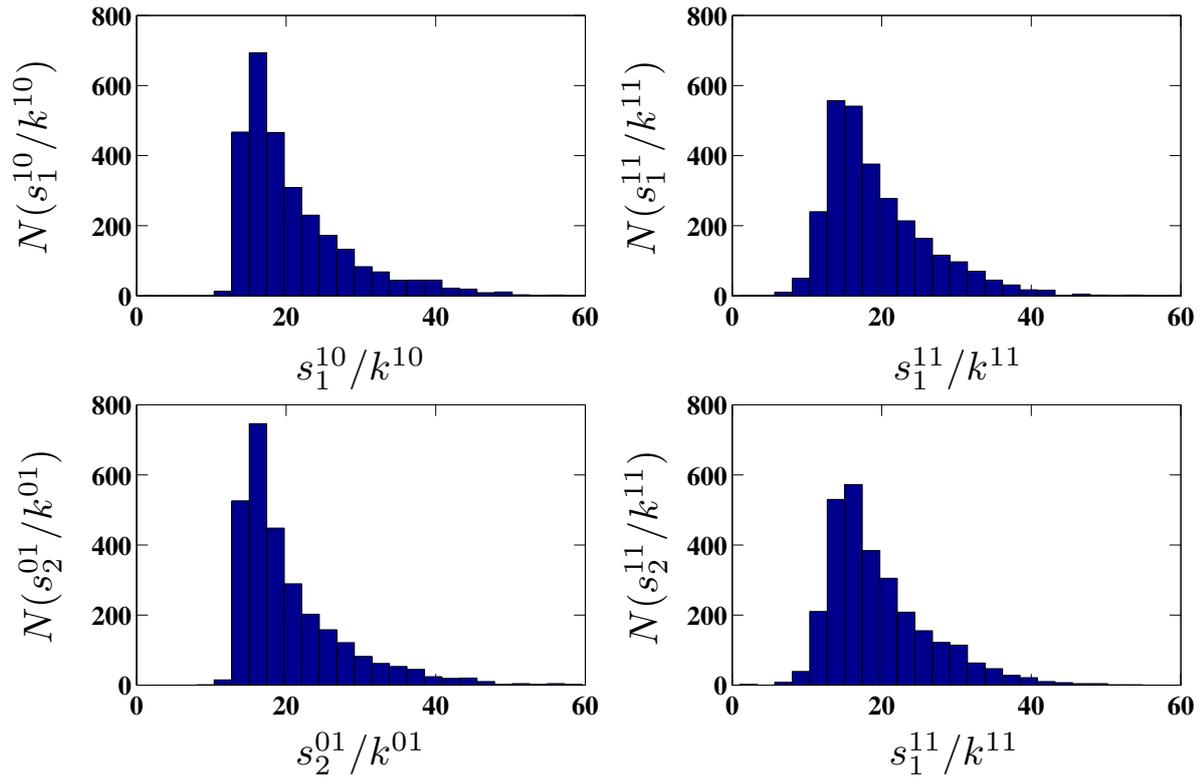


Figure 2.9: Biological case study: we display the distributions  $\{s_{i,\alpha}^{\vec{m}}/k_i^{\vec{m}}\}$ , i.e. the average weight of each node's interactions, classified according to the multilinks.

# 3 Control Theory

## 3.1 Introduction to control theory

Control theory is a well known branch of engineering with applications to a large number of disciplines ranging from medicine and drug discovery (Csermely et al., 2013), to the characterisation of dynamical processes in the brain (Bullmore and Sporns, 2009; Bonifazi et al., 2009; Power et al., 2011), or the evaluation of risk in financial markets (Delpini et al., 2013). Control theory is even involved in the study of biomass flows in ecological systems. A dynamical system is considered *controllable* if, given a designed choice of inputs, it can be driven from any initial state to any chosen final state in finite time. The dynamical rules are embedded in a network representing the interactions of the different components of the system. The question of control placement (Ruths and Ruths, 2014), namely, which nodes to control with external inputs in order to achieve the controllability for the entire system, has many fall-outs in practical studies of dynamical networks. In the typical EEG experimental design for instance, electrodes can be used to stimulate brain voxels. In this brain model each node is a cortical voxel, and its state is given by the level of excitation at a specific point in time. The correlation in the EEG signals usually determines if one brain region

is connected to another. For social sciences the mapping of sentiment or opinion towards a particular subject is of great interest: considering people as nodes and their relationships as edges (friendship, influence, authority) is possible to map the opinion spreading over the population. The key-role of some influential individuals is again matter of control placement and of particular appealing for advertising (Ruths and Ruths, 2014).

In general, the dynamical information and the topological features are not always available and well integrated and this is the main difficulty in dealing with control theory. The interplay between network structure and its related controllability has been recently analysed in the crucial papers (Liu et al., 2011, 2012). Topology plays a fundamental role in affecting dynamical processes on a given network and it seems as well playing a key role in network controllability.

## 3.2 Definition of controllability and structural controllability of a network

Controllability studies the relationship between the state of a given system and its inputs (Iglesias and Ingalls, 2009). We start considering the simplest picture, a given system well described by canonical linear time-invariant dynamics. Many real systems are better represented by nonlinear dynamics but we consider the linear approach as a natural prerequisite of the nonlinear controllability problem. Consider the system

$$\frac{d\mathbf{x}(t)}{dt} = A\mathbf{x} + B\mathbf{u}, \quad (3.1)$$

in which each element of the vector  $\mathbf{x}(t)$ , namely,  $x_i(t)$  with  $i = 1, 2, \dots, N$ , represents the state of node  $i$  at time  $t$ . This vector is therefore called the *state vector*. A good example is given by gene regulatory network where each  $x_i(t)$  gives the transcription factor concentration of gene  $i$  at time  $t$ . The matrix  $A$  is a  $N \times N$  (asymmetric) matrix defined as the *state matrix* of the system. It encodes the directed weighted relation of each couple of system component and the element  $a_{ij}$  is associated with the link ( $j \rightarrow i$ ) on a directed network  $G(A)$ . Always considering a gene regulatory network, the interaction can be both positive (excitatory) or negative (inhibitory). Whenever all the links have unit strength the state matrix  $A$  becomes the transpose of the adjacency matrix of  $G(A)$ .

$B$  is a  $N \times M$  matrix, so-called *input matrix*, defining which nodes are controlled by the outside ( $M \leq N$ ). The  $M$  external signals are indicated by the vector  $\mathbf{u}(t)$  of elements  $u_\alpha$  and  $\alpha = 1, 2 \dots M$ . A system is *controllable* if the control input  $\mathbf{u}(t)$  is able to drive the state  $\mathbf{x}(t)$  from any initial condition  $\mathbf{x}_0(t_0) \in \mathbb{R}^N$  to a final configuration  $\mathbf{x}_1(t_1) \in \mathbb{R}^N$  (Iglesias and Ingalls, 2009).

We define as *driver nodes*  $N_D$  those controlled nodes which do not share input signals and whose control determines the dynamics of the system. The main goal is the identification of the minimal set of driver nodes which guarantees full controllability.

For any given realisation of  $A$  and  $B$ , the dynamical system is controllable if it satisfies Kalman's controllability rank condition, i.e. the controllability matrix  $C = (B, AB, A^2B, \dots, A^{N-1}B)$  is full rank ( $\text{rank}(C)=N$ ). For a simple example of controllability see Fig. 3.1.

In addition to the fact that the verification of Kalman's condition can be computationally very demanding for large systems, in most real systems the notion

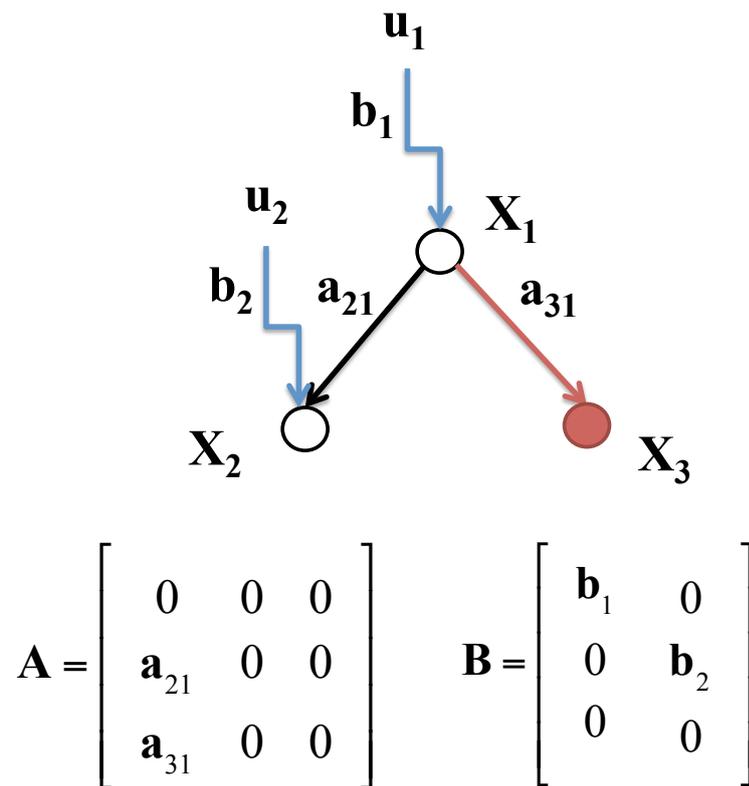


Figure 3.1: Controllability of a small system: the linear dynamics of Eq. 3.1 is defined by matrix  $A$  (*state matrix*) and by matrix  $B$  (*input matrix*). The system is controlled by two inputs  $\mathbf{u} = (u_1(t), u_2(t))^t$  and the input signals are displayed in blue. The controllability matrix  $C = (B, AB, A^2B)$  has rank 3, meaning that Kalman's controllability rank condition is satisfied. Link directions follow the transpose of matrix  $A$ . Maximum matching links ( $a_{31}$ ) and matched nodes ( $X_3$ ) are shown in red. Empty nodes are unmatched but connecting them to input signals yields full control of the system.

of exact controllability is unusable since the entries of  $A$  and  $B$  are not perfectly known. As an alternative, if we assume that the non-zero matrix elements of  $A$  and  $B$  are free parameters, we can consider the concept of *structural controllability* (Lin, 1974). The system is structurally controllable if for any choice of the free parameters in  $A$  and  $B$ , except for a variety of zero Lebesgue measure in

the parameter space,  $C$  is full rank (Lin, 1974). Since structural controllability only distinguishes between zero and non-zero entries of the matrices  $A$  and  $B$ , a given directed network is structurally controllable if it is possible to determine the input nodes (i.e. the position of the non-zero entries of the matrix  $B$ ) in a way to control the dynamics described by any realisation of the matrix  $A$  with the same non-zero elements, except for atypical realisations of zero measure. A network is therefore structurally controlled by identifying the minimum number of driver nodes.

In his seminal work, Lin defined the corresponding network analog of the linear system 3.1, the so-called control-augmented graph  $G(A, B)$ , formed by adding *control nodes* to the original network  $G(A)$  (one for each control input). His fundamental result states that a linear system  $(A, B)$  is structurally controllable if and only  $G(A, B)$  is spanned by *cacti*. A cactus is the key structure of Lin's Structural Controllability Theorem and it is composed by elementary substructures, namely, (see Fig. 3.2)

**Stem**  $\rightarrow$  elementary path starting from an *input vertex* or *control node* called *root* and ending in a terminal vertex called *top*

**Bud**  $\rightarrow$  elementary cycle with an additional edge  $e$  called *distinguished edge of the bud* (it ends but not begins in a vertex of the cycle)

**U – rooted factorial connection**  $\rightarrow$  union of vertex-disjoint stems and elementary cycles spanning  $G(A, B)$

A *cactus* is then composed by a main stem with some buds attached only by their distinguished edges (no distinguished edge starts at the top of the main stem). Finally, “a *cacti* is a set of vertex-disjoint cacti” (Liu et al., 2011). Removing

any edge from the cacti breaks the controllability of the system. This structure is therefore *minimal* and is equivalent to the irreducibility condition on  $[A, B]$  (Ruths and Ruths, 2014).

Liu et al. (2011), starting from Lin’s Structural Controllability Theorem, mapped the problem of finding the minimal set of driver nodes into a maximum matching problem. The so-called “Minimum Input Theorem” identifies the driver nodes with the unmatched nodes. In most cases the structure of  $B$  is not known a priori, this means that, once studied  $G(A)$ , a suitable choice of input connections can guarantee the controllability of the system. Applying the maximum matching algorithm we gather the set of disjoint simple paths and loops that maximally cover  $G(A)$ . We connect a single input to each unmatched node, building then a spanning cacti of  $G(A, B)$ . Whenever a cycle does not have a distinguished edge connecting it to a stem, we build a new bud taking one of the input nodes and connecting it to any node of the cycle (no further input nodes are needed). For further explanations see Fig. 3.2.

### 3.2.1 Matching, Maximum Matching and Perfect Matching

For an undirected network a matching is defined as a set of edges without common vertices. A node is considered *matched* if it is incident to an edge in the matching, otherwise it is called *unmatched*. We have a “maximum matching” whenever a matching reaches the maximum cardinality, while we define “perfect matching” a matching which matches all the nodes in the graph. These definitions naturally apply to bipartite networks where the famous Hopcroft-Karp algorithm finds the maximum matching in  $O(\sqrt{V}E)$  with  $V$  the number of nodes and  $E$  the number of edges.

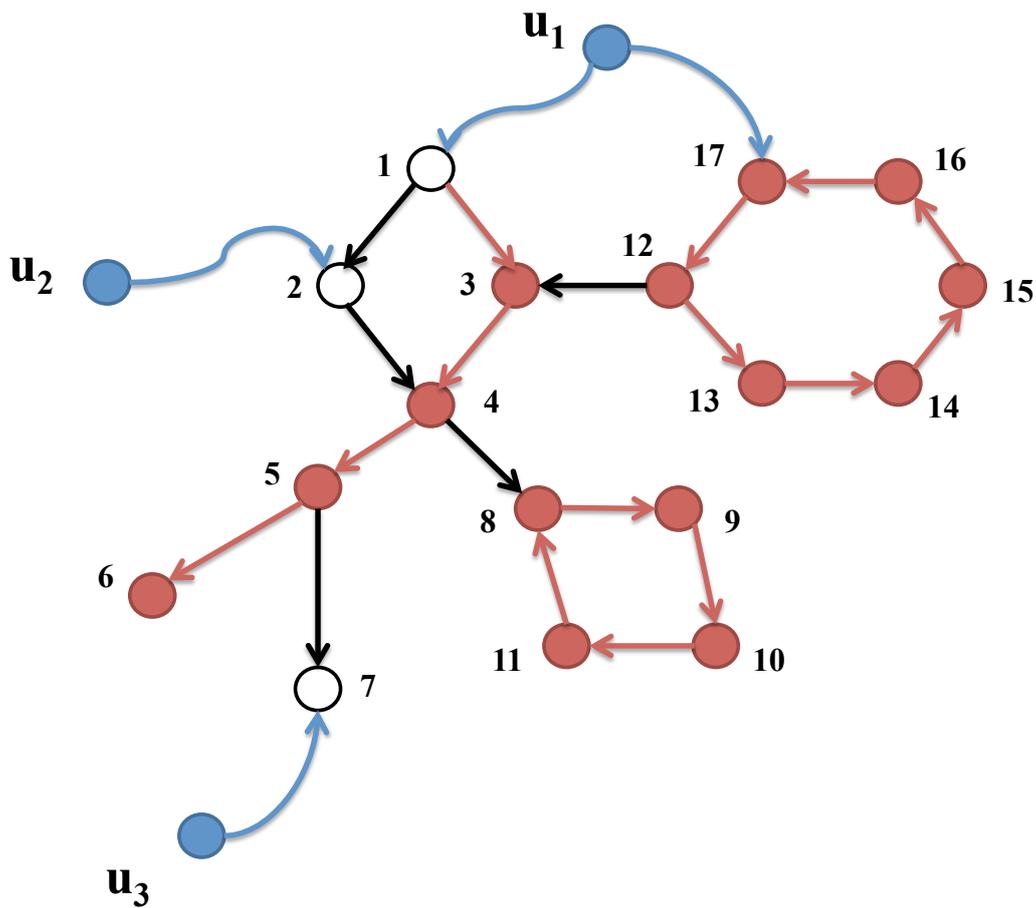


Figure 3.2: Control-augmented graph  $G(A, B)$ : input vertices and links are marked in blue and in this particular system we have three control nodes driving three stems, namely,  $(u_1, 1, 3, 4, 5, 6)$ ,  $(u_2, 2)$  and  $(u_3, 7)$ . Moreover, we have two loops,  $(12, 13, 14, 15, 16, 17)$  and  $(8, 9, 10, 11)$ . Loop  $(8-9-10-11)$  has its distinguished edge  $(4, 8)$ , forming together a bud. Loop  $(12, 13, 14, 15, 16, 17)$  does not have a distinguished edge so we connect it to input  $u_1$ , building another bud. The U-rooted factorial connection is composed by:  $(u_1, 1, 3, 4, 5, 6)$ ,  $(u_2, 2)$ ,  $(u_3, 7)$ ,  $(12, 13, 14, 15, 16, 17)$ ,  $(8, 9, 10, 11)$ . The spanning cacti is composed by one main cactus given by  $(u_1, 1, 3, 4, 5, 6)$ ,  $(4, 8)$ ,  $(u_1, 17)$  and  $(12, 13, 14, 15, 16, 17)$ . The cactus  $(u_2, 2)$  and the other cactus  $(u_3, 7)$  are simply stems. Red edges belong to one the possible maximum matching of  $G(A)$ . We can appreciate how this algorithm finds a set of disjoint paths and circles covering  $G(A)$ : no two edges share a common starting node or a common ending node. Red nodes are defined as *matched* while empty nodes 1, 2, and 7 are defined as *unmatched*. Each control node is then associated to one unmatched node (Ruths and Ruths, 2014).

In directed networks the definition of matching considers the existence of starting vertices and ending vertices. Therefore in this version of matching no two edges share a common starting or ending vertex. A node is matched whenever is pointed by an edge in the matching, otherwise, it is unmatched (see Fig. 3.1 and Fig. 3.2). The maximum matching of a directed network  $G(A)$  can be computed always with the Hopcroft-Karp algorithm (Hopcroft and Karp, 1973), considering a bipartite version of the initial network. The new bipartite graph is created with twice the number of nodes of the original network: the nodes are divided in two groups, the first group  $V^+ = \{v_1^+, \dots, v_N^+\}$  is composed by the original vertices seen as out-vertices, while in the second group  $V^- = \{v_1^-, \dots, v_N^-\}$  we collect the in-vertices. The original edges are then translated in new links starting from  $V^+$  and ending in  $V^-$ , in other words, an edge  $(v_i^+, v_j^-)$  corresponds to the link  $i \rightarrow j$  in  $G(A)$  (Liu et al., 2011).

The peculiarity of directed matching is that its edges form elementary paths and circles. In elementary paths every node is matched except the first one, while in circles all nodes are matched, giving then an example of perfect matching (see Fig. 3.2). These features gave the chance to map Lin's Structural Controllability Theorem into the matching problem on directed networks. Once obtained the maximum matching for the given network we can connect an input to each unmatched node.

### 3.2.2 Real networks and null models

In Liu et al. (2011) a set of different kinds of real networks was analysed and for each one the density of driver nodes was obtained applying the maximum matching algorithm. One of the main findings regards the role of hubs in the

controllability of a given network. The fraction of driver nodes is significantly higher among nodes with low degree than among hubs. Moreover, in all the cases the average degree of driver nodes  $\langle k_D \rangle$  is either importantly smaller than or comparable to the average degree of the network  $\langle k \rangle$  (where  $\langle k \rangle = 2 \langle k_{in} \rangle = 2 \langle k_{out} \rangle$ ). Randomisation enlightens the interplay between topological features and controllability (see Ch. 1). A full randomisation procedure keeps the number of nodes  $N$  and the number of links  $L$  unchanged. This procedure turns the real network into a directed *Erdos-Renyi* random network. The results displayed in Liu et al. (2011) indicate how a full randomisation removes the topological features that are related to controllability. On the other hand, a degree-preserving randomisation that keeps both the  $k_{in}$  and the  $k_{out}$  for each node (Maslov, 2014) does not modify in a significant way the original number of driver nodes  $N_D$ . The joint degree distribution  $P(k_{in}, k_{out})$  seems to strongly characterise the controllability of the underlying network. So fixed  $P(k_{in}, k_{out})$  the real network is just a given instance of a more general ensemble of networks (see Ch. 1). Analytical methods such as the cavity method give us the chance to calculate the average density of driver nodes  $n_D$  over all network instances compatible with the given  $P(k_{in}, k_{out})$ .

### 3.3 Cavity method and Belief Propagation: a short introduction

In this section the essential prerequisites of statistical mechanics for the analysis of the matching problem on directed networks are presented. In the following sections a similar formalism for the computation of the number of driver nodes, using belief propagation and population dynamics, will be covered.

### 3.3.1 Disordered systems

Among the methods for disordered systems replica method and cavity method are the most common. In particular, given the context of this thesis we are interested to the latter. An example of disordered system is given by the Ising model on a random graph, a version of the spin glass problem in the dilute case. We consider a random graph  $G = (V, E)$  drawn from of the ensemble of *Erdos-Renyi* networks with  $N$  nodes and average degree  $c$  (Hartmann and Weigt, 2005). The vertices are labelled as  $V = \{1, 2, \dots, N\}$  and we assume that the edges are drawn independently with probability  $p = c/N$ . The matrix  $J$  becomes the adjacency matrix of  $G$  and its elements are

$$J_{ij} = \begin{cases} 1 & \text{if } \{ij\} \in E \\ 0 & \text{if } \{ij\} \notin E \end{cases}$$

The probability of a given graph belonging to the ensemble reflects the independence of different edges and can be written as

$$P(J) = \prod_{i < j} [(1 - p)\delta(J_{ij}) + p\delta(J_{ij} - 1)] \quad (3.2)$$

The Hamiltonian is given by (Hartmann and Weigt, 2005)

$$H_J(\sigma_1, \dots, \sigma_N) = - \sum_{i < j} J_{ij} \sigma_i \sigma_j \quad (3.3)$$

where the spins are expressed with the popular notation  $\sigma_i = \pm 1$ ,  $i = 1, \dots, N$ . Eq. 3.3 is characterised by the microscopic configuration  $\mathcal{C} = \{\sigma_i\}$  and by the quenched disorder  $J$ . Quenched disorder means practically that the disorder is fixed and it is not subject to thermal fluctuations (Zagordi, 2007). In physical terms this means that the time scale characterising changes in the interactions is much longer than that one regarding changes in the dynamical variables. Eq. 3.3 and the related free energy density depend on a specific realisation of the disorder (matrix  $J$ ), i.e.

$$f_J = -\frac{1}{\beta N} \log Z_J \quad (3.4)$$

where  $N$  is the system size and  $Z_J$  is the realisation-dependent partition function. Anyway, from physical considerations the typical features of a disordered system do not change going from realisation to realisation (Zagordi, 2007). This *selfaverageness* property explains how, for large systems, the studied properties do not depend on the given instance of  $J$ , i.e.

$$f_\infty(\beta) = \lim_{N \rightarrow \infty} f_J(\beta, N) \quad (3.5)$$

Furthermore, the mean value of a self averaging quantity over the disorder, such as the free energy density, is well defined and corresponds with the thermodynamic limit:

$$f = \lim_{N \rightarrow \infty} \frac{1}{\beta N} \langle \log Z_J \rangle_J = f_\infty(\beta) \quad (3.6)$$

This means that  $f$  should have small fluctuations (of the order of  $1/N$ ), a typical situation in case of short range interactions (Zagordi, 2007).

### Quenched average and annealed average

The so-called *quenched average* is given by

$$f = \langle f_J \rangle_J = \int dJP(J) F_J = -\frac{1}{\beta N} \int dJP(J) \log Z_J \quad (3.7)$$

Integrating the logarithm of the partition function shows many difficulties and the famous *replica trick* was developed to avoid this issue (in this short introduction we are not presenting the replica method). Another possible average is the so-called *annealed average*, namely,

$$-\frac{1}{\beta N} \log \int dJP(J) Z_J \quad (3.8)$$

where we calculate the logarithm of  $\langle Z_J \rangle_J$ . This kind of approximation often leads to wrong results (Zagordi, 2007).

### Cavity method

The cavity method (or Bethe-Peierls approximation) is a useful approach to improve the classic mean-field approximation. In the usual mean-field the degrees of freedom of a single spin are treated exactly while all the other variables are replaced by their average values (Nishimori and Ortiz, 2010). In the cavity method the configuration of nearest neighbours of a given spin are considered with no approximation and all the spins beyond those neighbours are approximated by their mean values. The neighbouring spins then feel effective fields that express the influence of spins beyond them.

This method is characterised by the locality of the dependencies among its variables and this is a common feature of many other inference problems. This common trait belongs especially to the so-called graphical models: Bayesian networks, Markov random fields and factor graphs. With these graphical models we can easily represent systems of Ising spins or constraint satisfaction problems (e.g. random XOR-SAT), that are indeed related. For a general introduction to graphical models and belief propagation we refer to Yedidia et al. (2001). Belief propagation or *BP* is an efficient local message passing algorithm for the resolution of inference problems represented by graphical models. Moreover, it can only converge to a fixed point that is also a stationary point of the Bethe approximation of free energy. The main purpose of the BP algorithm is the computing of marginal probabilities or *beliefs*, at least approximately (when the graph has no loops the algorithm is exact), in a time growing linearly with the number of nodes in the system and not exponentially (Yedidia et al., 2001). The same algorithm, named in different ways, has been repeatedly rediscovered in a large variety of scientific fields: the forward-backward algorithm, the Viterbi algorithm, iterative decoding algorithms for Gallager codes and turbocodes, Pearl's belief algorithm for Bayesian network, the Kalman filter, and finally, the transfer-matrix approach in physics (Yedidia et al., 2001).

Unlike the replica method, the cavity method gives results both for the quenched average and the single graph. The results on a single instance are fundamental for the parallelism with message passing algorithms. For the considered example, i.e. Ising spins on a random graph, the two methods give exactly the same results (replica symmetric regime).

The main and fundamental assumption of this method is the *locally tree-like*

structure of the underlying graph defined by  $J$ , meaning that almost all loops are of length  $\mathcal{O}(\log N)$  and therefore, in the thermodynamic limit, they go to infinity.

For the sake of simplicity we consider then a tree-graph. This kind of structure allows the calculation of the partition function via an iterative scheme (Hartmann and Weigt, 2005). Selecting an arbitrary node  $i$  we introduce the restricted partition function  $Z_i(\sigma_i)$ , i.e.,

$$Z_i(\sigma_i) = \sum_{\{\sigma_k, k \neq i\}} e^{\beta \sum_{l < m} J_{lm} \sigma_l \sigma_m} \quad (3.9)$$

where the spin  $\sigma_i$  is fixed to  $+1$  or  $-1$  and  $Z = Z_i(1) + Z_i(-1)$ . The subtrees that are rooted in the neighbours  $j$  of  $i$  are independent and disconnected if we remove vertex  $i$  and all its links. Eq. 3.9 can be then rewritten as a function of the restricted partition functions  $Z_{j \rightarrow i}(\sigma_j)$  of the subtrees *above*  $i$  (Hartmann and Weigt, 2005), namely,

$$\begin{aligned} Z_i(\sigma_i) &= \sum_{\{\sigma_j, j \in N(i)\}} e^{\beta \sigma_i \sum_{j \in N(i)} \sigma_j} \prod_{j \in N(i)} Z_{j \rightarrow i}(\sigma_j) \\ &= \prod_{j \in N(i)} \sum_{\sigma_j} e^{\beta \sigma_i \sigma_j} Z_{j \rightarrow i}(\sigma_j) \end{aligned} \quad (3.10)$$

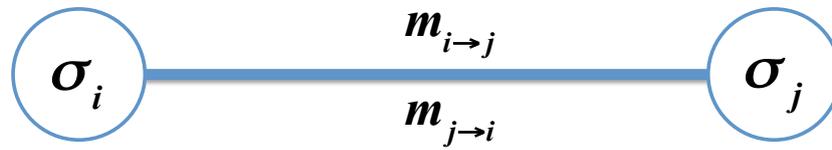
where  $N(i)$  defines the neighbourhood of node  $i$ . Recursively,  $Z_{j \rightarrow i}(\sigma_j)$  depends on the subtrees above vertex  $j$ , i.e.,

$$\begin{aligned} Z_{j \rightarrow i}(\sigma_j) &= \sum_{\{\sigma_k, k \in N(j) \setminus i\}} e^{\beta \sigma_j \sum_{k \in N(j) \setminus i} \sigma_k} \prod_{k \in N(j) \setminus i} Z_{k \rightarrow j}(\sigma_k) \\ &= \prod_{k \in N(j) \setminus i} \sum_{\sigma_k} e^{\beta \sigma_j \sigma_k} Z_{k \rightarrow j}(\sigma_k) \end{aligned} \quad (3.11)$$

In Yedidia et al. (2001) the authors give an example of belief propagation

### Markov Random Fields

$$\psi_{ij}(\sigma_i, \sigma_j) = e^{\beta J_{ij} \sigma_i \sigma_j}$$



### Factor Graph

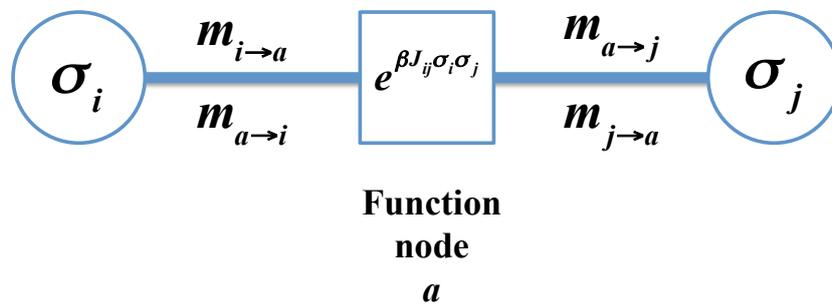


Figure 3.3: Ising model on a random graph: representation of a single couple of spins  $i$  and  $j$  in two different graphical models. For pairwise Markov random fields each spin is a node variable (circle) with possible states  $\pm 1$ . Each link is associated with a compatibility function  $\psi_{ij}(\sigma_i, \sigma_j) = e^{\beta J_{ij} \sigma_i \sigma_j}$ . The factor graph introduces instead a function node  $a$  (square) representing  $f_a(\sigma_i, \sigma_j) = e^{\beta J_{ij} \sigma_i \sigma_j}$ .

implemented on pairwise Markov random fields (MRF's). The main result that we want to mention about MRF's is that they can reproduce our problem of Ising spins on a network. Each spin variable is seen as a node variable with possible states  $\pm 1$ . Each edge  $(i, j)$  is associated with a *compatibility function*  $\psi_{ij}(\sigma_i, \sigma_j)$  through the expression  $\log \psi_{ij}(\sigma_i, \sigma_j) = \beta J_{ij} \sigma_i \sigma_j$  (see Fig. 3.3). The overall joint

probability of the pairwise Markov random fields is

$$p(\{\sigma_i\}) = \frac{1}{Z} \prod_{(i,j) \in E} \psi_{ij}(\sigma_i, \sigma_j) \quad (3.12)$$

that is equivalent to the usual Boltzmann probability distribution. For the BP algorithm implemented on our graph two main definitions are introduced, the *belief*  $p(\sigma_i)$  and the message update rule, i.e.,

$$p(\sigma_i) = K \prod_{j \in N(i)} m_{j \rightarrow i}(\sigma_i) \quad (3.13)$$

$$m_{j \rightarrow i}(\sigma_i) = \sum_{\sigma_j} \psi_{ij}(\sigma_i, \sigma_j) \prod_{k \in N(j) \setminus i} m_{k \rightarrow j}(\sigma_j) \quad (3.14)$$

These last equations are connected to Eq. 3.10 and Eq. 3.11 if we consider

$$m_{j \rightarrow i} = \sum_{\sigma_j} e^{\beta \sigma_i \sigma_j} Z_{j \rightarrow i}(\sigma_j) \quad (3.15)$$

and  $p(\sigma_i) = Z_i(\sigma_i)/Z$  (the normalisation constant is then  $K = 1/Z$ ).

The introduction of the *cavity fields*  $h_{j \rightarrow i}$  simplifies the hierarchy of Eqs. 3.10, 3.11. They are defined as

$$h_{j \rightarrow i} = \frac{1}{2\beta} \log \frac{Z_{j \rightarrow i}(+1)}{Z_{j \rightarrow i}(-1)} = \frac{1}{2\beta} \sum_{k \in N(j) \setminus i} \log \frac{\cosh \beta(h_{k \rightarrow j} + 1)}{\cosh \beta(h_{k \rightarrow j} - 1)} \quad (3.16)$$

In a similar way the *local effective field* is defined as

$$h_i = \frac{1}{2\beta} \log \frac{Z_i(+1)}{Z_i(-1)} = \frac{1}{2\beta} \sum_{j \in N(i)} \log \frac{\cosh \beta(h_{j \rightarrow i} + 1)}{\cosh \beta(h_{j \rightarrow i} - 1)} \quad (3.17)$$

The interpretation of cavity fields and of the effective field is easier if we con-

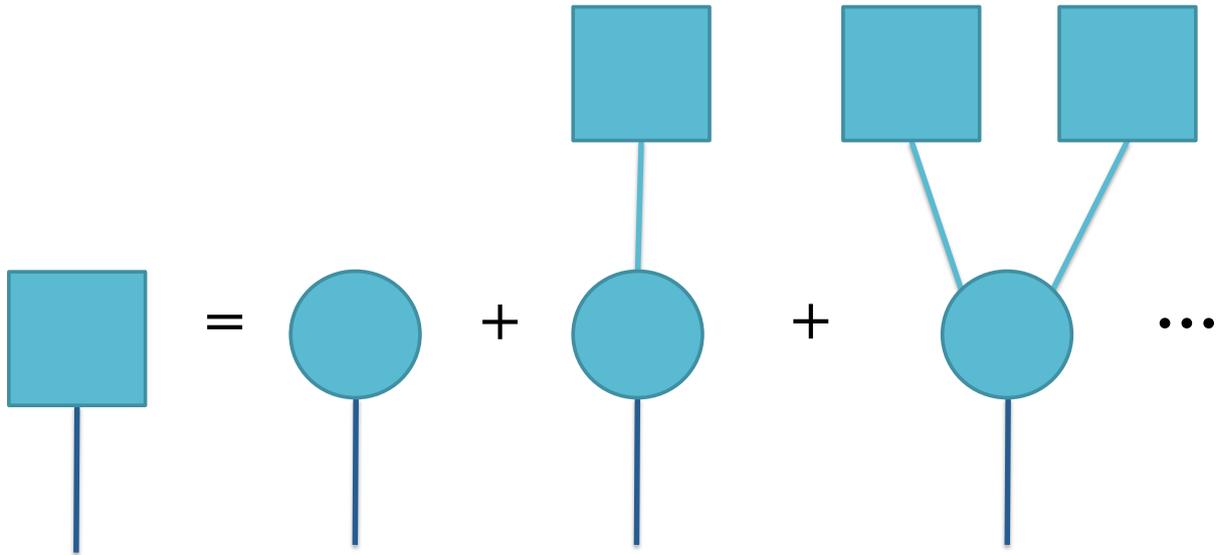


Figure 3.4: Graphical solution of the cavity-field distribution (see Eq. 3.26): a given cavity field is generated from the cavity fields of  $k$  neighbours, where  $k$  is the so-called *excess degree* and follows the nearest neighbour degree distribution. Following a generic link in a given direction, its associated cavity field (square, on the left) is defined considering all the possible contributions from the cavity fields (squares, on the right) acting on the spin (circle) at its “ending vertex” (from the right to the left, the ending vertex has excess degree equal to 0, 1, 2)

sider the marginal probabilities  $p(\sigma_i)$  and  $p_{j \rightarrow i}(\sigma_j)$

$$p(\sigma_i) = \frac{Z_i(\sigma_i)}{Z} = \frac{e^{\beta h_i \sigma_i}}{2 \cosh \beta h_i} \quad (3.18)$$

$$p_{j \rightarrow i}(\sigma_j) = \frac{Z_{j \rightarrow i}(\sigma_j)}{Z_{j \rightarrow i}(1) + Z_{j \rightarrow i}(-1)} = \frac{e^{\beta h_{j \rightarrow i} \sigma_j}}{2 \cosh \beta h_{j \rightarrow i}} \quad (3.19)$$

Eq. 3.18 tells us that spin  $i$  behaves like a single spin in an external field  $h_i$ .

The same scenario leads to  $Z_i(\sigma_i) = Ae^{\beta h_i \sigma_i}$  with  $A$  a positive constant, and to the usual expression for local magnetisation given by  $m_i = \tanh \beta h_i$ . Similar arguments work for Eq. 3.19 where now the global effect on spin  $j$  is considered without its neighbour  $i$  and  $Z_{j \rightarrow i}(\sigma_j) = Be^{\beta h_{j \rightarrow i} \sigma_j}$  with  $B$  a positive constant.

It is convenient to introduce also the joint probability distribution for spin  $i$  and  $j$ ,  $p(\sigma_i, \sigma_j)$ . Following Yedidia et al. (2001) we get

$$\begin{aligned}
p(\sigma_i, \sigma_j) &= K \psi_{ij}(\sigma_i, \sigma_j) \prod_{k \in N(i) \setminus j} m_{k \rightarrow i}(\sigma_i) \prod_{l \in N(j) \setminus i} m_{l \rightarrow j}(\sigma_j) \\
&= \frac{1}{Z} e^{\beta \sigma_i \sigma_j} Z_{i \rightarrow j}(\sigma_i) Z_{j \rightarrow i}(\sigma_j) \\
&= \frac{(Z_{i \rightarrow j}(1) + Z_{i \rightarrow j}(-1))(Z_{j \rightarrow i}(1) + Z_{j \rightarrow i}(-1))}{Z} e^{\beta \sigma_i \sigma_j} p_{i \rightarrow j}(\sigma_i) p_{j \rightarrow i}(\sigma_j) \\
&= \frac{(2B_1 \cosh \beta h_{i \rightarrow j})(2B_2 \cosh \beta h_{j \rightarrow i})}{Z} e^{\beta \sigma_i \sigma_j} \frac{e^{\beta h_{i \rightarrow j} \sigma_i}}{2 \cosh \beta h_{i \rightarrow j}} \frac{e^{\beta h_{j \rightarrow i} \sigma_j}}{2 \cosh \beta h_{j \rightarrow i}} \\
&= \frac{B_1 B_2 e^{\beta \sigma_i \sigma_j} e^{\beta h_{i \rightarrow j} \sigma_i} e^{\beta h_{j \rightarrow i} \sigma_j}}{Z_{i \rightarrow j}(1)(\sum_{\sigma_j} e^{\beta \sigma_j} Z_{j \rightarrow i}(\sigma_j)) + Z_{i \rightarrow j}(-1)(\sum_{\sigma_j} e^{-\beta \sigma_j} Z_{j \rightarrow i}(\sigma_j))} \\
&= \frac{e^{\beta \sigma_i \sigma_j} e^{\beta h_{i \rightarrow j} \sigma_i} e^{\beta h_{j \rightarrow i} \sigma_j}}{\sum_{\sigma_i, \sigma_j} e^{\beta \sigma_i \sigma_j} e^{\beta h_{i \rightarrow j} \sigma_i} e^{\beta h_{j \rightarrow i} \sigma_j}} \tag{3.20}
\end{aligned}$$

where we considered  $Z_{i \rightarrow j}(\sigma_i) = B_1 e^{\beta h_{i \rightarrow j} \sigma_i}$  and  $Z_{j \rightarrow i}(\sigma_j) = B_2 e^{\beta h_{j \rightarrow i} \sigma_j}$ . This joint probability distribution  $p(\sigma_i, \sigma_j)$  is properly normalised and satisfies  $p(\sigma_i) = \sum_{\sigma_j} p(\sigma_i, \sigma_j)$ . Moreover, we can calculate the total joint probability distribution  $p(\{\sigma\})$  as a function of  $\{p(\sigma_i)\}$  and  $\{p(\sigma_i, \sigma_j)\}$ , i.e.

$$p(\{\sigma\}) = \frac{\prod_{(i,j) \in E} p(\sigma_i, \sigma_j)}{\prod_i p(\sigma_i)^{k_i - 1}} \tag{3.21}$$

where  $k_i$  is the number of neighbours of spin  $i$ . This computation is exact whenever the graph has no loops and it becomes equivalent to the usual Boltzmann

probability distribution for a given configuration  $\{\sigma\}$ , i.e,

$$p(\{\sigma\}) = \frac{1}{Z} e^{-\beta H_J(\sigma_1, \dots, \sigma_N)} \quad (3.22)$$

Solving Eq. 3.21 the correct normalisation constant  $1/Z$  is obtained considering

$$\frac{Z^{2|E|-N}}{Z^{|E|}} = \frac{Z^{2(N-1)-N}}{Z^{(N-1)}} \quad (3.23)$$

where  $2|E| - N = \sum_i k_i - 1$  and  $|E| = N - 1$  (we consider a tree-graph).

Whenever we deal with random graphs locally tree-like the factorisation of the restricted partition function  $Z_i(\sigma_i)$  in subtrees is not correct ( see Eq. 3.10). Anyway,  $Z_i(\sigma_i)$  depends on the joint partition function  $Z_{N(i)|i}(\{\sigma_j\})$ , where  $\{\sigma_j\}$  are all the neighbours of node  $i$ , with  $i$  removed from the network (Hartmann and Weigt, 2005). The *cavity graph*  $G_i$  has origins from the initial graph  $G$  where node  $i$  and its incident links have been removed. The requested locally tree-like structure helps us: given a graph where almost all loops have length of  $\mathcal{O}(\log N)$ , in  $G_i$  the neighbours  $N(i)$  become really distant from each other and they can be considered practically uncorrelated whenever  $N \gg 1$ . The joint marginal distribution of these spins then factorises.

The field equations Eq. 3.16 and Eq. 3.17 determine a self-consistent system and they are solved in a similar way to the various message passing algorithms. In practice, for a given graph, starting with some initial conditions for  $\{h_{j \rightarrow i}\}$ , for each link  $(i, j)$ , for each direction of the edge, we iterate Eq. 3.16 till convergence. Once obtained all the  $\{h_{j \rightarrow i}\}$ , thank to Eq. 3.17 we calculate the effective fields  $\{h_i\}$  and then the physical behaviour of our system becomes fully characterised. We have just given the basic tools to solve our problem for a particular graph

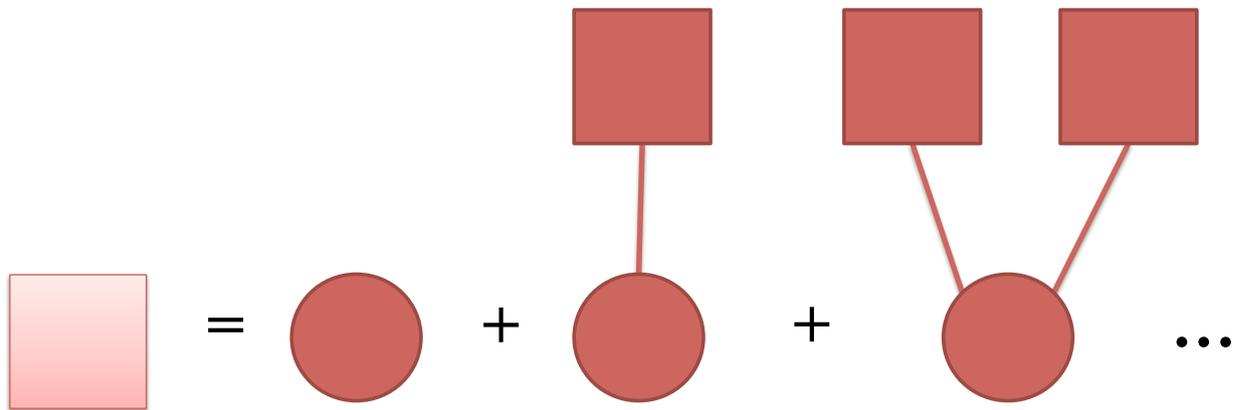


Figure 3.5: Graphical solution of the physical effective-field distribution (see Eq. 3.27): an effective field is generated from the cavity fields of all its neighbours  $k$ , where  $k$  follows the usual degree distribution  $P(k)$ . The light-red square represents the effective field acting on a generic spin (circle), while the dark-red squares are the cavity fields, considered in all their possible contributions to the given spin (from the right to the left, the considered node has degree equal to 0, 1, 2).

(associated with a given  $J$ ) of the ensemble. First of all, we introduce the *cavity-field* probability distribution as

$$P_{cav}(h) = \frac{1}{cN} \sum_{\{i,j\} \in E} [\delta(h - h_{i \rightarrow j}) + \delta(h - h_{j \rightarrow i})] \quad (3.24)$$

where each edge gives two contributes to the fields (Hartmann and Weigt, 2005).

We can see Eq. 3.24 as the normalised histogram over all cavity fields. Moreover,

also an *effective – field* distribution can be considered, namely,

$$P(h) = \frac{1}{N} \sum_{i \in V} \delta(h - h_i) \quad (3.25)$$

In the thermodynamic limit both the cavity-field distribution and the effective-field distribution can be determined by self-consistent equations, given the equivalence of every links in a random graph. How can we write down these equations? Thank to their graphical representation in Fig. 3.4 and in Fig. 3.5. This method is similar to the usual procedure considered in network theory for the computation of the giant component for uncorrelated random networks with the generating functions.

Let's start with the computation of the cavity fields: a generic  $h_{j \rightarrow i}$  depends on the cavity fields  $\{h_{k \rightarrow j}\}$  of  $k$  neighbours, where  $k$  follows the nearest neighbour probability distribution, or equivalently,  $k$  can be considered as the excess degree of node  $j$  with respect to edge  $(i, j)$  (Hartmann and Weigt, 2005). These arguments lead to

$$\begin{aligned} P_{cav}(h) &= \sum_{k=1}^{\infty} \frac{kP(k)}{\langle k \rangle} \int \prod_{i=1}^{k-1} dh_i P_{cav}(h_i) \delta \left( h - \frac{1}{2\beta} \sum_{i=1}^{k-1} \log \frac{\cosh \beta(h_i + 1)}{\cosh \beta(h_i - 1)} \right) \\ &= \sum_{k=0}^{\infty} \frac{(k+1)P(k+1)}{\langle k \rangle} \int \prod_{i=1}^k dh_i P_{cav}(h_i) \delta \left( h - \frac{1}{2\beta} \sum_{i=1}^k \log \frac{\cosh \beta(h_i + 1)}{\cosh \beta(h_i - 1)} \right) \\ &= \sum_{k=0}^{\infty} e^{-c} \frac{c^k}{k!} \int \prod_{i=1}^k dh_i P_{cav}(h_i) \delta \left( h - \frac{1}{2\beta} \sum_{i=1}^k \log \frac{\cosh \beta(h_i + 1)}{\cosh \beta(h_i - 1)} \right) \end{aligned} \quad (3.26)$$

For  $P(h)$  the contributes come from all the neighbours of the considered node,

meaning that we need to consider the usual degree distribution  $P(k)$ , namely,

$$\begin{aligned} P(h) &= \sum_{k=0}^{\infty} P(k) \int \prod_{i=1}^k dh_i P_{cav}(h_i) \delta \left( h - \frac{1}{2\beta} \sum_{i=1}^k \log \frac{\cosh \beta(h_i + 1)}{\cosh \beta(h_i - 1)} \right) \\ &= \sum_{k=0}^{\infty} e^{-c} \frac{c^k}{k!} \int \prod_{i=1}^k dh_i P_{cav}(h_i) \delta \left( h - \frac{1}{2\beta} \sum_{i=1}^k \log \frac{\cosh \beta(h_i + 1)}{\cosh \beta(h_i - 1)} \right) \end{aligned} \quad (3.27)$$

For the characteristics of the Poissonian degree distribution  $P_{cav}(h) = P(h)$ , even if the fields on single vertices do not coincide. These equations give an iterative recipe for the estimate of  $P(h)$  and the solution follows a fixed-point procedure. In particular, for  $0 < T < T_c$  (where  $T_c$  is a critical temperature, average-degree dependent, above which the model behaves as a paramagnet), we can go ahead just numerically using a very popular algorithm called *population-dynamics* (Hartmann and Weigt, 2005). We start considering a large population  $\{h_1, \dots, h_M\}$  of  $M \gg 1$  fields, representing the field-distribution and firstly initialised randomly. We then run the algorithm that iteratively replaces some fields inside the population, until convergence. The following pseudo-code represents the main path of the population dynamics:

**do**

draw  $k$  from a Poisson distribution  $e^{-c} c^k / k!$

select randomly  $k + 1$  indices  $i, i_1, \dots, i_k \in \{1, \dots, M\}$

replace  $h_i^{old}$  with  $h_i^{new} = \frac{1}{2\beta} \sum_{l=1}^k \log \frac{\cosh \beta(h_{i_l} + 1)}{\cosh \beta(h_{i_l} - 1)}$

**while** not converged

**return**  $(h_1, \dots, h_M)$

The convergence of this algorithm means that the statistical properties of the population (especially its histogram) become constant up to negligible fluctua-

tions.

We give now an example of the computation a global quantity such as energy. Energy can be calculated for a single network or it can be considered as the energy density for the network ensemble.

For a single instance we have

$$\begin{aligned}\langle H_J \rangle_T &= - \sum_{i < j} J_{ij} \langle \sigma_i \sigma_j \rangle_T \\ &= - \sum_{i < j} J_{ij} \frac{\sum_{\sigma_i, \sigma_j} \sigma_i \sigma_j e^{\beta \sigma_i \sigma_j} e^{\beta h_{i \rightarrow j} \sigma_i} e^{\beta h_{j \rightarrow i} \sigma_j}}{\sum_{\sigma_i, \sigma_j} e^{\beta \sigma_i \sigma_j} e^{\beta h_{i \rightarrow j} \sigma_i} e^{\beta h_{j \rightarrow i} \sigma_j}}\end{aligned}\quad (3.28)$$

where Eq. 3.20 has been used. Averaging over the graph ensemble we obtain the energy density, i.e.

$$e = \frac{\langle \langle H_J \rangle_T \rangle_J}{N} = -\frac{c}{2} \int dh_1 dh_2 P_{cav}(h_1) P_{cav}(h_2) \frac{\sum_{\sigma_1, \sigma_2} \sigma_1 \sigma_2 e^{\beta \sigma_1 \sigma_2} e^{\beta h_1 \sigma_1} e^{\beta h_2 \sigma_2}}{\sum_{\sigma_1, \sigma_2} e^{\beta \sigma_1 \sigma_2} e^{\beta h_1 \sigma_1} e^{\beta h_2 \sigma_2}}\quad (3.29)$$

### 3.3.2 Belief propagation on factor graphs

Another variant of the belief propagation algorithm is implemented over the previously mentioned factor graphs. A factor graph is one of the possible graphical models together with Bayesian networks and pairwise Markov random fields. A factor graph is essentially a bipartite graph containing two types of nodes called function nodes (set  $F$ ) and variable nodes (set  $V$ ). In the previous section we mentioned pairwise Markov random fields where there is just a single type of nodes, i.e. variable nodes, and each connection between a pair of nodes defines a compatibility function. The main feature of all these graphical models is the factorisation of the overall joint probability (see Eq. 3.12). For factor graphs the

joint probability reads

$$p(\mathbf{x}) \propto \prod_{a \in F} f_a(\mathbf{x}_a) \quad (3.30)$$

where  $\mathbf{x}_a$  represents the vector of variable nodes in the neighbourhood of function node  $a$ , in one of the possible configurations of their given states, i.e.  $\mathbf{x}_a$  contains the arguments of function  $a$ . Two kinds of nodes need two different types of messages running over the factor graphs, differently from pairwise Markov random fields where we updated just one type of message. These messages are formally defined as

**variable node  $v \rightarrow$  function node  $a$**

$$m_{v \rightarrow a}(x_v) = \prod_{b \in N(v) \setminus a} m_{b \rightarrow v}(x_v) \quad (3.31)$$

**function node  $b \rightarrow$  variable node  $v$**

$$m_{b \rightarrow v}(x_v) = \sum_{\mathbf{x}_b \setminus x_v} f_b(\mathbf{x}_b) \prod_{t \in N(b) \setminus v} m_{t \rightarrow b}(x_t) \quad (3.32)$$

where for  $x_v$  we denote one of the possible states of variable node  $v$  and  $\sum_{\mathbf{x}_b \setminus x_v}$  means summing over the states of all the neighbours of  $b$  except node  $v$ .

We define then two marginal distributions for each kind of nodes:

**variable node**

$$p(x_v) \propto \prod_{a \in N(v)} m_{a \rightarrow v}(x_v) \quad (3.33)$$

**function node**

$$p(\mathbf{x}_a) \propto f_a(\mathbf{x}_a) \prod_{v \in N(a)} m_{v \rightarrow a}(x_v) \quad (3.34)$$

The factor graph gives an alternative representation of the Ising model on a random graph. Each interaction term becomes a function node and it is represented by a square vertex, while each spin is represented by a circular variable node (see Fig. 3.3). In this model every function node has degree two, i.e. each function node is associated with just one link and connected only to two variable nodes. Let's consider the situation pictured in Fig. 3.3. Combining Eq. 3.31 with Eq. 3.32 we find

$$m_{i \rightarrow a}(\sigma_i) = \prod_{b \in N(i) \setminus a} \sum_{\sigma_k} e^{\beta \sigma_i \sigma_k} m_{k \rightarrow b}(\sigma_k) \quad (3.35)$$

where with  $\sigma_k$  we denote one of the spins connected to  $\sigma_i$  in the real graph. On the factor graph this spin  $\sigma_k$  is linked to  $\sigma_i$  through a generic function node  $b$ . We modify  $b \in N(i) \setminus a$  in  $k \in N(i) \setminus j$  obtaining

$$m_{i \rightarrow a}(\sigma_i) = \prod_{k \in N(i) \setminus j} \sum_{\sigma_k} e^{\beta \sigma_i \sigma_k} m_{k \rightarrow b}(\sigma_k) \quad (3.36)$$

$$(3.37)$$

This last equation is equivalent to  $Z_{i \rightarrow j}(\sigma_i)$  (see Eq. 3.11). Finally, we use Eq. 3.34 to evaluate  $p(\sigma_i, \sigma_j)$  (see Eq. 3.20), namely,

$$p(\sigma_i, \sigma_j) \propto e^{\beta \sigma_i \sigma_j} m_{i \rightarrow a}(\sigma_i) m_{j \rightarrow a}(\sigma_j) \quad (3.38)$$

in agreement with Eq. 3.20.

## 3.4 The BP approach to the maximum matching problem

### The maximum matching problem

The maximum matching problem can be treated by statistical mechanics techniques (Liu et al., 2011; Zdeborová and Mézard, 2006; Altarelli et al., 2011; Mézard and Parisi, 2001; Martin et al., 2001; Hartmann and Weigt, 2005) such as the cavity method. As previously introduced, a matching  $M$  of a directed graph is a set of directed edges without common start or end vertices, and it is maximum when it contains the maximum possible number of edges. The problem of finding a maximum matching of a directed graph can be cast on a statistical mechanics problem, by introducing variables  $s_{ij} \in \{1, 0\}$  on each directed link from node  $i$  to node  $j$ , indicating whether the directed link is in  $M$  ( $s_{ij} = 1$ ) or not ( $s_{ij} = 0$ ). The configurations of variables  $\{s_{ij}\}$  have to satisfy the following matching condition,

$$\sum_{j \in \partial_+ i} s_{ij} \leq 1, \quad \sum_{j \in \partial_- i} s_{ji} \leq 1, \quad (3.39)$$

where  $\partial_- i$  indicates the set of nodes  $j$  that point to node  $i$  in the directed network, and  $\partial_+ i$  indicates the set of nodes  $j$  that are pointed by node  $i$  (see Fig. 3.6). If these constraints are satisfied each node  $i$  of the network has at most one in-coming link that is matched, (i.e. one neighbour  $j \in \partial_- i$  such that  $s_{ji} = 1$ ) and at most one outgoing link (one neighbour  $j \in \partial_+ i$  such that  $s_{ij} = 1$ ) that is

matched. Moreover the variables  $\{s_{ij}\}$  should minimise the energy function

$$\begin{aligned}
E &= 2 \sum_{i=1}^N \left( 1 - \sum_{j \in \partial_- i} s_{ji} \right) \\
&= \sum_{i=1}^N \left( 1 - \sum_{j \in \partial_- i} s_{ji} \right) + \sum_{i=1}^N \left( 1 - \sum_{j \in \partial_+ i} s_{ij} \right) \\
&= 2N_D
\end{aligned} \tag{3.40}$$

$$= 2(N - |M|) \tag{3.41}$$

where  $N_D$  is the number of unmatched nodes in the network and this number also determines the minimum number of driver nodes required to fully control the network.  $|M|$  is the cardinality of the matching  $M$  and is also equal to the number of matched nodes: each one of these nodes is associated to an ending vertex of a matched link. Moreover, we remark

$$\sum_{i=1}^N \sum_{j \in \partial_- i} s_{ji} = \sum_{i=1}^N \sum_{j \in \partial_+ i} s_{ij} = |M| \tag{3.42}$$

The  $\beta \rightarrow \infty$  limit corresponds to the ground state, in other words the situation in which the matching is maximum and the number of driver nodes is minimum.

We aim at finding the distribution  $P(\{s_{ij}\})$  given by

$$\begin{aligned}
P(\{s_{ij}\}) &= \frac{e^{-\beta E}}{Z} \prod_{i=1}^N \theta \left( 1 - \sum_{j \in \partial_+ i} s_{ij} \right) \\
&\quad \times \prod_{i=1}^N \theta \left( 1 - \sum_{j \in \partial_- i} s_{ji} \right)
\end{aligned} \tag{3.43}$$

where  $\theta(x) = 1$  for  $x \geq 0$  and  $\theta(x) = 0$  for  $x < 0$  and where  $Z$  is the normalization constant, that corresponds to the partition function of the statistical mechanics problem. In particular our goal is to find this distribution in the limit  $\beta \rightarrow \infty$

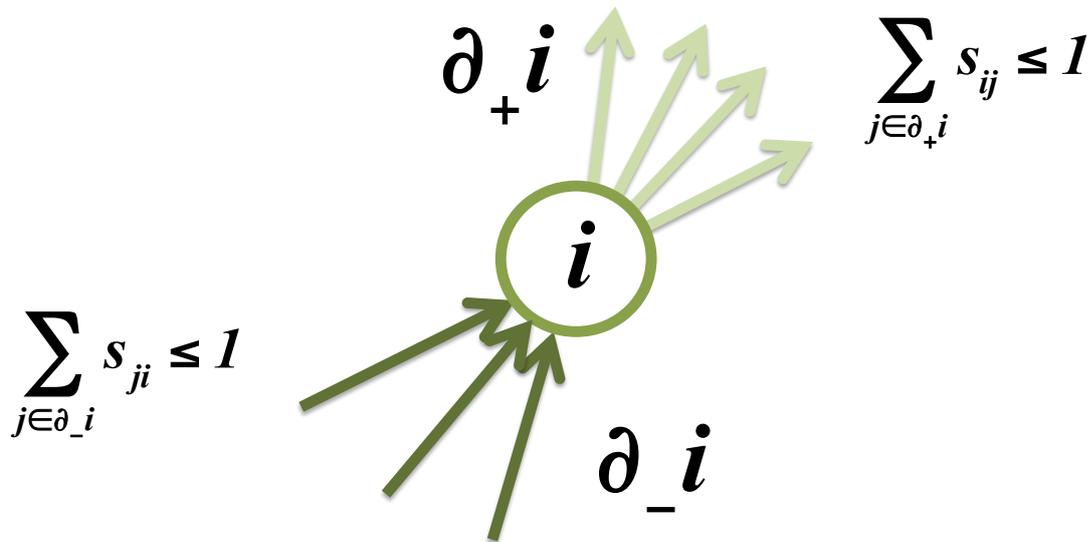


Figure 3.6: Once fixed node  $i$  we define two sets: the first one is  $\partial_- i$  and it gathers all those nodes pointing to node  $i$ ; the second one is  $\partial_+ i$  and it indicates all those nodes pointed by node  $i$ . To each link we associate a variable  $s_{ij}$  or  $s_{ji}$  depending on the direction of the considered edge. For instance, let's consider a generic link  $i \rightarrow j$ : the related variable  $s_{ij}$  is 1 if  $i \rightarrow j$  belongs to the matching, 0 otherwise.

in order to characterise the optimal (i.e. the maximum-sized) matching in the network. The free-energy density of the problem  $f(\beta)$  is defined as

$$\beta N f(\beta) = -\ln Z, \quad (3.44)$$

and the energy of the problem is therefore given by

$$E = \frac{\partial[\beta N f(\beta)]}{\partial \beta}. \quad (3.45)$$

### 3.4.1 The BP equations

The distribution  $P(\{s_{ij}\})$  on a locally tree-like network can be solved by the BP message passing method by finding the messages that nearby nodes sent to each other. In Zdeborová and Mézard (2006) the problem of matching in undirected random networks was solved using a factor graph: each node of the original network becomes a function node and each link is represented by a variable node. This representation is easily explained: the matching variables (i.e.  $s_{ij}$ ) are associated with the links of network but the constraint functions (similar to Eqs. 3.39) and the Boltzmann factor work in the neighbourhood of each original node.

Moreover, a generic link  $\ell = (i, j)$  has always degree 2, meaning that it has a connection only with function node  $i$  and function node  $j$ , and no further constraints are applied to this variable node. For this link then, the outgoing message to function node  $j$  is equivalent to the incoming message from function node  $i$  i.e.  $m_{\ell \rightarrow j}(s_\ell) = m_{i \rightarrow \ell}(s_\ell)$  (see Eq. 3.31). The formalism used in Zdeborová and Mézard (2006) is then rearranged in messages running only between couples of function nodes, i.e. for node  $i$  and  $j$  we get  $m_{i \rightarrow j}(s_{ij})$  and  $m_{j \rightarrow i}(s_{ij})$ .

For the particular case of directed graphs we distinguish between messages going in the direction of the link,  $m_{i \rightarrow j}(s_{ij})$ , and messages going in the opposite direction of the link,  $\hat{m}_{i \rightarrow j}(s_{ji})$ . Moreover, as explained by Eqs. 3.39, the neighbourhood of each node  $\partial i$  is divided in  $\partial_+ i$  and  $\partial_- i$ . We are going to use normalised messages as in the custom, i.e.  $P_{i \rightarrow j}(s_{ij})$  and  $\hat{P}_{i \rightarrow j}(s_{ji})$ . Following directly from

Eqs. 3.31, 3.32, the BP equations for these messages read

$$\begin{aligned}
P_{i \rightarrow j}(s_{ij}) &= \frac{1}{\mathcal{D}_{i \rightarrow j}} \sum_{\{s_{ik}\}_{k \in \partial_+ i \setminus j}} \theta \left( 1 - \sum_{k \in \partial_+ i} s_{ik} \right) \\
&\quad \times \exp \left[ -\beta \left( 1 - \sum_{k \in \partial_+ i} s_{ik} \right) \right] \\
&\quad \times \prod_{k \in \partial_+ i \setminus j} \hat{P}_{k \rightarrow i}(s_{ik}), \\
\hat{P}_{i \rightarrow j}(s_{ji}) &= \frac{1}{\hat{\mathcal{D}}_{i \rightarrow j}} \sum_{\{s_{ki}\}_{k \in \partial_- i \setminus j}} \theta \left( 1 - \sum_{k \in \partial_- i} s_{ki} \right) \\
&\quad \times \exp \left[ -\beta \left( 1 - \sum_{k \in \partial_- i} s_{ki} \right) \right] \\
&\quad \times \prod_{k \in \partial_- i \setminus j} P_{k \rightarrow i}(s_{ki}), \tag{3.46}
\end{aligned}$$

where  $\mathcal{D}_{i \rightarrow j}$  and  $\hat{\mathcal{D}}_{i \rightarrow j}$  are normalisation constants. At function node  $i$  for messages  $P_{i \rightarrow j}(s_{ij})$  we apply the function

$$f_i^+(s_{\{ik\}}, k \in \partial_+ i) = \theta \left( 1 - \sum_{k \in \partial_+ i} s_{ik} \right) \exp \left[ -\beta \left( 1 - \sum_{k \in \partial_+ i} s_{ik} \right) \right] \tag{3.47}$$

while for messages  $\hat{P}_{i \rightarrow j}(s_{ji})$  we consider the function

$$f_i^-(s_{\{ki\}}, k \in \partial_- i) = \theta \left( 1 - \sum_{k \in \partial_- i} s_{ki} \right) \exp \left[ -\beta \left( 1 - \sum_{k \in \partial_- i} s_{ki} \right) \right] \tag{3.48}$$

Similarly to Eqs. 3.16, 3.19, the messages  $\{P_{i \rightarrow j}(s_{ij}), \hat{P}_{i \rightarrow j}(s_{ji})\}$  can be parametrised by the cavity fields  $h_{i \rightarrow j}$  and  $\hat{h}_{i \rightarrow j}$  defined by

$$\frac{P_{i \rightarrow j}(1)}{P_{i \rightarrow j}(0)} = e^{\beta h_{i \rightarrow j}} \quad \frac{\hat{P}_{i \rightarrow j}(1)}{\hat{P}_{i \rightarrow j}(0)} = e^{\beta \hat{h}_{i \rightarrow j}} \tag{3.49}$$

and, thank to the normalisation condition, rearranged in

$$P_{i \rightarrow j}(s_{ij}) = \frac{e^{\beta h_{i \rightarrow j} s_{ij}}}{1 + e^{\beta h_{i \rightarrow j}}} = e^{\beta h_{i \rightarrow j} s_{ij}} P_{i \rightarrow j}(0) \quad (3.50)$$

$$\hat{P}_{i \rightarrow j}(s_{ji}) = \frac{e^{\beta \hat{h}_{i \rightarrow j} s_{ji}}}{1 + e^{\beta \hat{h}_{i \rightarrow j}}} = e^{\beta \hat{h}_{i \rightarrow j} s_{ji}} \hat{P}_{i \rightarrow j}(0) \quad (3.51)$$

In terms of the cavity fields, Eqs. 3.46 reduce to the following set of equations,

$$\begin{aligned} h_{i \rightarrow j} &= -\frac{1}{\beta} \log \left( e^{-\beta} + \sum_{k \in \partial_+ i \setminus j} e^{\beta \hat{h}_{k \rightarrow i}} \right), \\ \hat{h}_{i \rightarrow j} &= -\frac{1}{\beta} \log \left( e^{-\beta} + \sum_{k \in \partial_- i \setminus j} e^{\beta h_{k \rightarrow i}} \right). \end{aligned} \quad (3.52)$$

that were first derived in Liu et al. (2011) for this problem. These last equations follow from Eqs. 3.46 introducing Eqs 3.50, 3.51, namely

$$\begin{aligned} P_{i \rightarrow j}(0) &= \frac{1}{\mathcal{D}_{i \rightarrow j}} \left[ e^{-\beta} + \sum_{k \in \partial_+ i \setminus j} e^{\beta \hat{h}_{k \rightarrow i}} \right] \prod_{k \in \partial_+ i \setminus j} \hat{P}_{k \rightarrow i}(0) \\ P_{i \rightarrow j}(1) &= \frac{1}{\mathcal{D}_{i \rightarrow j}} \prod_{k \in \partial_+ i \setminus j} \hat{P}_{k \rightarrow i}(0) \\ \hat{P}_{i \rightarrow j}(0) &= \frac{1}{\hat{\mathcal{D}}_{i \rightarrow j}} \left[ e^{-\beta} + \sum_{k \in \partial_- i \setminus j} e^{\beta h_{k \rightarrow i}} \right] \prod_{k \in \partial_- i \setminus j} P_{k \rightarrow i}(0) \\ \hat{P}_{i \rightarrow j}(1) &= \frac{1}{\hat{\mathcal{D}}_{i \rightarrow j}} \prod_{k \in \partial_- i \setminus j} P_{k \rightarrow i}(0) \end{aligned}$$

and finally considering Eqs. 3.49.

In the Bethe approximation, in a similar way to Eq. 3.21, the probability distribution  $P(\{s_{ij}\})$  is given by

$$P_{\text{Bethe}}(\{s_{ij}\}) = \prod_{i=1}^N P_i(\underline{S}_i) \left( \prod_{\langle i,j \rangle} P_{ij}(s_{ij}) \right)^{-1} \quad (3.53)$$

where  $P_i(\underline{S}_i)$  and  $P_{ij}(s_{ij})$  are the marginal distribution over the nodes and the links of the network, that can be computed in terms of the cavity messages  $P_{i \rightarrow j}(s_{ij})$ ,  $\hat{P}_{i \rightarrow j}(s_{ji})$ , or equivalently in terms of the cavity fields  $h_{i \rightarrow j}$  and  $\hat{h}_{i \rightarrow j}$ . In particular,  $\underline{S}_i = (s_{\{ik\}}, k \in \partial_+ i) \cup (s_{\{ki\}}, k \in \partial_- i)$  and  $P_i(\underline{S}_i)$  follows the marginal distribution for function nodes (see Eq. 3.34). The variable node  $s_{ij}$  follows instead the marginal probability given by Eq. 3.33. The marginal probabilities read

$$P_i(\underline{S}_i) = \frac{e^{-\beta[(1-\sum_{k \in \partial_+ i} s_{ik}) + (1-\sum_{k \in \partial_- i} s_{ki})]}}{\mathcal{C}_i} \quad (3.54)$$

$$\begin{aligned} & \times \theta \left( 1 - \sum_{k \in \partial_+ i} s_{ik} \right) \theta \left( 1 - \sum_{k \in \partial_- i} s_{ki} \right) \\ & \times \prod_{k \in \partial_+ i} \hat{P}_{k \rightarrow i}(s_{ik}) \prod_{k \in \partial_- i} P_{k \rightarrow i}(s_{ki}) \\ P_{ij}(s_{ij}) & = \frac{1}{\mathcal{C}_{ij}} P_{i \rightarrow j}(s_{ij}) \hat{P}_{j \rightarrow i}(s_{ij}) \end{aligned} \quad (3.55)$$

where  $\mathcal{C}_i$  and  $\mathcal{C}_{ij}$  are normalization constant given by

$$\begin{aligned} \mathcal{C}_i & = \left( e^{-\beta} + \sum_{k \in \partial_+ i} e^{\beta \hat{h}_{k \rightarrow i}} \right) \left( e^{-\beta} + \sum_{k \in \partial_- i} e^{\beta h_{k \rightarrow i}} \right) \\ & \times \prod_{k \in \partial_+ i} \hat{P}_{k \rightarrow i}(0) \prod_{k \in \partial_- i} P_{k \rightarrow i}(0) \end{aligned} \quad (3.56)$$

$$\mathcal{C}_{ij} = (1 + e^{\beta(h_{i \rightarrow j} + \hat{h}_{j \rightarrow i})}) P_{i \rightarrow j}(0) \hat{P}_{j \rightarrow i}(0). \quad (3.57)$$

### 3.4.2 Free energy and energy of the problem

The free energy of the problem can be found by evaluating the Gibbs free energy  $F_{Gibbs}$  given by

$$\beta F_{Gibbs} = \sum_{\{s_{ij}\}} P(\{s_{ij}\}) \log \left( \frac{P(\{s_{ij}\})}{e^{-\beta E} \psi(\{s_{ij}\})} \right) \quad (3.58)$$

This function assumes its minimal value  $-\log Z$  when  $P(\{s_{ij}\}) = P(\{s_{ij}\})^B = e^{-\beta E} \psi(\{s_{ij}\})/Z$ , where  $\psi(\{s_{ij}\})$  indicates the constraints

$$\psi(\{s_{ij}\}) = \prod_{i=1}^N \left[ \theta \left( 1 - \sum_{j \in \partial_+ i} s_{ij} \right) \theta \left( 1 - \sum_{j \in \partial_- i} s_{ji} \right) \right]. \quad (3.59)$$

and  $P(\{s_{ij}\})^B$  is the classical Boltzmann's law. From the point of view of Information theory, minimising the Gibbs free energy is equivalent to evaluate the minimum of a Kullback-Leiber distance between  $P(\{s_{ij}\})^B$  and a trial distribution  $P(\{s_{ij}\})$ , i.e.

$$D(P(\{s_{ij}\}) || P(\{s_{ij}\})^B) = \sum_{\{s_{ij}\}} P(\{s_{ij}\}) \log \frac{P(\{s_{ij}\})}{P(\{s_{ij}\})^B} \quad (3.60)$$

In Bethe approximation, in a situation of in which our graph is singly connected, we can use Eq. 3.53, and the related Eqs. 3.54, 3.55. The fixed-point solutions of the BP equations 3.46, when the graph has no loops, determine the exact marginal probabilities. We express then  $P(\{s_{ij}\})^B$  in terms of Eq. 3.53, that implies

$$Z = \frac{\prod_i \mathcal{C}_i}{\prod_{\langle i,j \rangle} \mathcal{C}_{ij}} \quad (3.61)$$

From the previous equations we can write the Gibbs free energy as

$$\beta F_{Bethe} = \sum_{\langle i,j \rangle} \log(\mathcal{C}_{ij}) - \sum_{i=1}^N \log(\mathcal{C}_i). \quad (3.62)$$

Inserting Eqs. 3.56, 3.57 into 3.62, we obtain the free energy of this matching problem, given by Liu et al. (2011), i.e.

$$\begin{aligned} \beta N f(\beta) = & - \sum_{i=1}^N \left( e^{-\beta} + \sum_{k \in \partial_+ i} e^{\beta \hat{h}_{k \rightarrow i}} \right) \\ & - \sum_{i=1}^N \left( e^{-\beta} + \sum_{k \in \partial_- i} e^{\beta h_{k \rightarrow i}} \right) \\ & + \sum_{\langle i,j \rangle} \ln \left( 1 + e^{\beta(h_{i \rightarrow j} + \hat{h}_{j \rightarrow i})} \right). \end{aligned} \quad (3.63)$$

Using Eq.3.45 we get the energy

$$\begin{aligned} E = & \sum_{i=1}^N \left[ \frac{e^{-\beta} - \sum_{k \in \partial_+ i} \hat{h}_{k \rightarrow i} e^{\beta \hat{h}_{k \rightarrow i}}}{e^{-\beta} + \sum_{k \in \partial_+ i} e^{\beta \hat{h}_{k \rightarrow i}}} \right] \\ & + \sum_{i=1}^N \left[ \frac{e^{-\beta} - \sum_{k \in \partial_- i} h_{k \rightarrow i} e^{\beta h_{k \rightarrow i}}}{e^{-\beta} + \sum_{k \in \partial_- i} e^{\beta h_{k \rightarrow i}}} \right] \\ & + \sum_{\langle i,j \rangle} \frac{(h_{i \rightarrow j} + \hat{h}_{j \rightarrow i}) e^{\beta(h_{i \rightarrow j} + \hat{h}_{j \rightarrow i})}}{1 + e^{\beta(h_{i \rightarrow j} + \hat{h}_{j \rightarrow i})}}. \end{aligned} \quad (3.64)$$

### 3.4.3 The $\beta \rightarrow \infty$ limit

In the  $\beta \rightarrow \infty$  limit, the energy of a maximum matching can be written as follows

$$\begin{aligned} E = & - \sum_{i=1}^N \max \left[ -1, \max_{k \in \partial_+ i} \hat{h}_{k \rightarrow i} \right] - \sum_{j=1}^N \max \left[ -1, \max_{k \in \partial_- i} h_{k \rightarrow i} \right] \\ & + \sum_{\langle i,j \rangle} \max \left[ 0, h_{i \rightarrow j} + \hat{h}_{j \rightarrow i} \right] \end{aligned} \quad (3.65)$$

in which for each directed link  $(i, j)$  the cavity fields  $\{h_{i \rightarrow j}, \hat{h}_{i \rightarrow j}\}$  satisfy the zero-temperature Belief Propagation equations, also known as Max-Sum (MS) equations,

$$h_{i \rightarrow j} = -\max \left[ -1, \max_{k \in \partial_+ i \setminus j} \hat{h}_{k \rightarrow i} \right], \quad (3.66a)$$

$$\hat{h}_{i \rightarrow j} = -\max \left[ -1, \max_{k \in \partial_- i \setminus j} h_{k \rightarrow i} \right], \quad (3.66b)$$

where in these equations when node  $i$  has only one outgoing link pointing to node  $j$ , i.e.  $|\partial_+ i| = 1$  we assume  $h_{i \rightarrow j} = 1$ ; similarly, when node  $i$  has only one incoming link coming from node  $j$ , i.e.  $|\partial_- i| = 1$  we assume  $\hat{h}_{i \rightarrow j} = 1$ . In the infinite size limit, the MS equations are closed for cavity fields with support either on  $\{-1, 1\}$  or on  $\{-1, 0, 1\}$  (Zdeborová and Mézard, 2006; Liu et al., 2011; Altarelli et al., 2011). When multiple solutions coexist, the dynamically stable solutions of minimum energy are the correct solutions of the maximum matching problem. As previously explained, the fields are sent in the same direction  $h_{i \rightarrow j}$  or in the opposite direction  $\hat{h}_{i \rightarrow j}$  of the links and indicate the following messages (Zdeborová and Mézard, 2006):  $h_{i \rightarrow j} = \hat{h}_{i \rightarrow j} = 1$  indicates *match me*,  $h_{i \rightarrow j} = \hat{h}_{i \rightarrow j} = -1$  indicates *do not match me*, finally  $h_{i \rightarrow j} = \hat{h}_{i \rightarrow j} = 0$  indicates *do what you want*.

### 3.4.4 BP/MS Equations in an ensemble of random networks with given degree distribution

Eq. 3.64 holds on a single directed graph. We can go further and calculate the average energy density over an ensemble of networks with given  $P(k_{in}, k_{out})$ . This situation is similar to the previous one related to the Ising model on random

graphs (see Sec. 3.3.1). The self-consistent equations in this case read

$$\begin{aligned}
 \mathcal{P}(h) &= \sum_{k_{in}=0}^{\infty} \sum_{k_{out}=1}^{\infty} \frac{k_{out}}{\langle k_{out} \rangle} P(k_{in}, k_{out}) \int \prod_i^{k_{out}-1} d\hat{h}_i \hat{\mathcal{P}}(\hat{h}_i) \delta \left[ h + \frac{1}{\beta} \log \left( e^{-\beta} + \sum_i^{k_{out}-1} e^{\beta \hat{h}_i} \right) \right] \\
 &= \sum_{k_{out}=1}^{\infty} \frac{k_{out}}{\langle k_{out} \rangle} P(k_{out}) \int \prod_i^{k_{out}-1} d\hat{h}_i \hat{\mathcal{P}}(\hat{h}_i) \delta \left[ h + \frac{1}{\beta} \log \left( e^{-\beta} + \sum_i^{k_{out}-1} e^{\beta \hat{h}_i} \right) \right] \\
 \hat{\mathcal{P}}(\hat{h}) &= \sum_{k_{in}=1}^{\infty} \sum_{k_{out}=0}^{\infty} \frac{k_{in}}{\langle k_{in} \rangle} P(k_{in}, k_{out}) \int \prod_i^{k_{in}-1} dh_i \mathcal{P}(h_i) \delta \left[ \hat{h} + \frac{1}{\beta} \log \left( e^{-\beta} + \sum_i^{k_{in}-1} e^{\beta h_i} \right) \right] \\
 &= \sum_{k_{in}=1}^{\infty} \frac{k_{in}}{\langle k_{in} \rangle} P(k_{in}) \int \prod_i^{k_{in}-1} dh_i \mathcal{P}(\hat{h}_i) \delta \left[ \hat{h} + \frac{1}{\beta} \log \left( e^{-\beta} + \sum_i^{k_{in}-1} e^{\beta h_i} \right) \right]
 \end{aligned}$$

These equations are solved numerically by the algorithm of population dynamics, with a philosophy similar to Sec. 3.3.1. In the limit  $\beta \rightarrow \infty$  in which we look for the optimal matching we have that these distributions can be written as a sum of three delta functions, i.e.

$$\begin{aligned}
 \mathcal{P}(h) &= w_1 \delta(h-1) + w_2 \delta(h+1) + w_3 \delta(h) \\
 \hat{\mathcal{P}}(\hat{h}) &= \hat{w}_1 \delta(\hat{h}-1) + \hat{w}_2 \delta(\hat{h}+1) + \hat{w}_3 \delta(\hat{h}),
 \end{aligned} \tag{3.67}$$

where the variables  $\{w_1, w_2, w_3\}$  and the variables  $\{\hat{w}_1, \hat{w}_2, \hat{w}_3\}$  must satisfy the following normalisation conditions,  $w_1 + w_2 + w_3 = 1$  and  $\hat{w}_1 + \hat{w}_2 + \hat{w}_3 = 1$ .

The MS equations 3.66 can be written as equations for the set of probabilities

$\{w\}, \{\hat{w}\}$  obtaining

$$\begin{aligned}
w_1 &= \sum_k \frac{k}{\langle k \rangle_{out}} P^{out}(k) (\hat{w}_2)^{k-1} \\
w_2 &= \sum_k \frac{k}{\langle k \rangle_{out}} P^{out}(k) [1 - (1 - \hat{w}_1)^{k-1}] \\
\hat{w}_1 &= \sum_k \frac{k}{\langle k \rangle_{in}} P^{in}(k) (w_2)^{k-1} \\
\hat{w}_2 &= \sum_k \frac{k}{\langle k \rangle_{in}} P^{in}(k) [1 - (1 - w_1)^{k-1}], \tag{3.68}
\end{aligned}$$

with  $w_3 = 1 - w_1 - w_2$  and  $\hat{w}_3 = 1 - \hat{w}_1 - \hat{w}_2$ . Moreover, the energy given by Eq. 3.64 in the  $\beta \rightarrow \infty$  can be expressed in terms of the distributions  $\{w_i\}$  and  $\{\hat{w}_i\}$  obtaining,

$$\begin{aligned}
\frac{E}{N} &= \sum_k P^{out}(k) \left\{ (\hat{w}_2)^k - [1 - (1 - \hat{w}_1)^k] \right\} \\
&\quad \sum_k P^{in}(k) \left\{ (w_2)^k - [1 - (1 - w_1)^k] \right\} \\
&\quad + \langle k \rangle_{in} [\hat{w}_1(1 - w_2) + w_1(1 - \hat{w}_2)]. \tag{3.69}
\end{aligned}$$

In other words, the fraction of driver nodes  $n_D = E/(2N)$  in the network can be simply expressed in terms of the distributions  $\{w_i\}$  and  $\{\hat{w}_i\}$ . Eqs. 3.68 can have multiple solutions for the variables  $\{w_i\}$  and  $\{\hat{w}_i\}$ . In order to select the correct solution of the matching problem one should ensure that the following three conditions are satisfied.

- i) The sets  $\{w_i\}$  and  $\{\hat{w}_i\}$  must indicate two probability distributions;*
- ii) The solution should be stable:* The solution of the system of Eqs. 3.68 should be stable under small perturbation of the values of the distributions  $\{w_i\}$  and  $\{\hat{w}_i\}$ .
- iii) Find the optimal stable solution:* If the system of Eqs. 3.68 has more than

one solution that satisfies both conditions *i)* and *ii)*, in order to find the optimal matching one should select the solution with lowest energy  $E$  (when  $T = 0$  the message passing algorithm could give spurious solutions; we verified the agreement of the results given by the Hopcroft-Karp algorithm with the stable solutions with lowest energy).

### 3.5 Controllability and minimal degrees

Liu et al. (2011) characterise in detail the set of driver nodes for real networks and for ensembles of networks with given in-degree and out-degree distribution. By analysing scale-free networks with minimum in-degree and minimum out-degree equal to 1 they have found that the smaller is the power-law exponent  $\gamma$  of the degree distribution, the larger is the fraction of driver nodes in the network. This result has prompted the authors of Liu et al. (2011) to say that the higher is the heterogeneity of the degree distribution, the less controllable is the network. In this section we explore the role of low in-degree and low out-degree nodes in the controllability of networks. In the following, we show how changing the fraction of nodes with in-degree and out-degree less than 3, conditions the number of driver nodes of a network in a dramatic way. In particular, if the minimum in-degree and the minimum out-degree of a network are both greater than 2 then any network, independently on the level of heterogeneity of the degree distribution, is fully controllable by an infinitesimal fraction of nodes. Therefore the heterogeneity of the network is not the only element determining the number of driver nodes in the network and that this number is very sensible on the fraction of low in-degree low out-degree nodes of the network. This result allows

us to propose a method to improve the controllability of networks by decreasing the density of nodes with in-degree and out-degree less than 3, adding links to the network.

### 3.5.1 Sufficient condition for the full controllability of networks

Let us now show that for any network topology if the in-degree and the out-degree of the network is greater than 2 the fraction of driver nodes is zero. First we observe that the configuration in which all fields are zero, i.e.  $h_{i \rightarrow j} = \hat{h}_{i \rightarrow j} = 0$ , is an allowed solution of the Eqs. 3.66a, 3.66b as soon as the minimum in-degree and minimum out-degree equal to 1. In fact if a node has in-degree 1 this link must be matched, and a similar situation occurs for the nodes with out-degree 1, generating a set of hard constraints incompatible with the configuration in which all the fields are zero, while if the minimum in-degree or out-degree of the network is greater than 1, all the nodes can be matched in a variety of ways therefore all the fields can be equal to zero. This solution corresponds to a fraction of driver nodes  $n_D = 0$  if the minimum in-degree and the minimum out-degree are greater than 1. This solution is also stable if, when we change a single field from zero to a value different from zero, the perturbation does not propagate in the network. Suppose that  $\hat{h}_{k \rightarrow i}$  is changed, say, from 0 to 1, meaning that the message is *match me*, then all the nodes  $j \in \partial_+ i$  neighbor of  $i$  and different from  $k$  receive a message *do not match me*. But if all the nodes  $j$  have more than 2 incoming links, also if the link  $(j, k)$  is not matched they can still send to their incoming neighbors the messages *do what you want* since there are different ways in which the matching can be achieved and they do not have to impose to any of their

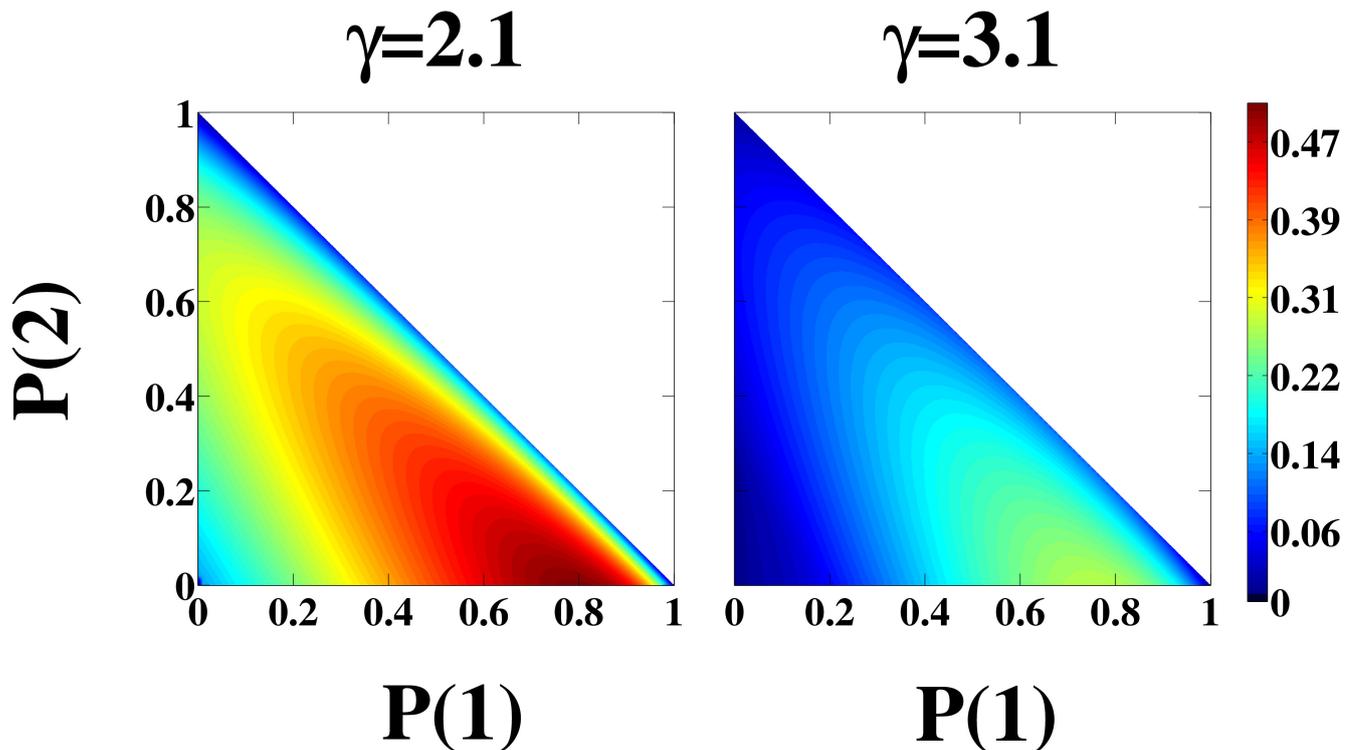


Figure 3.7: Heat map representing the density of driver nodes  $n_D$  as a function of the parameters  $P(1)$  and  $P(2)$  for networks of  $N = 10^6$  nodes with degree distribution given by Eq. 3.71 and  $\gamma = 2.1$  (left), 3.1 (right). The density  $n_D$  is obtained by numerically solving the BP/MS equations for an ensemble of networks with given degree distribution. The region in which  $P(1) + P(2) > 1$  is non-physical.

other links to be matched. Therefore the perturbation does not propagate in the network. A similar argument holds for a change of the field  $h_{k \rightarrow i}$  to 1 which does not propagate if the out-degree of the network is greater than 2. This stability argument shows that for every tree-like network for which the BP/MS equations are valid, if the in-degree and the out-degree of the network is greater than 2 then the density of driver nodes is  $n_D = 0$ . Note that this is a sufficient condition for the stability of the  $n_D = 0$  solution but more stringent conditions are discussed in the following for networks with given degree distribution.

### 3.5.2 Conditions for the full controllability of random networks

In the following we focus on ensembles of random networks with given in-degree and out-degree distribution  $P^{in}(k)$  and  $P^{out}(k)$ . In this case, it is possible to write the BP/MS equations and the energy in terms of the probabilities  $w_i \in [0, 1]$  and  $\hat{w}_i \in [0, 1]$  with  $i = 1, 2, 3$  that the cavity fields  $h_{i \rightarrow j}$  and  $\hat{h}_{i \rightarrow j}$  are respectively given by  $\{1, -1, 0\}$ . From the BP/MS equations of the matching problem on random networks with given degree distribution, we found that the solution  $n_D = 0$  is allowed if and only if  $P^{in/out}(0) = P^{in/out}(1) = 0$ . The replica-symmetric cavity equations are supposed to give the correct solution to the maximum matching problem if no instabilities take place. By analysing the stability condition of the BP/MS equations, we find that the stability conditions for this solution in an ensemble of networks with given in-degree and out-degree sequence, are

$$P^{out}(2) < \frac{\langle k \rangle_{in}^2}{2\langle k(k-1) \rangle_{in}}, \quad P^{in}(2) < \frac{\langle k \rangle_{in}^2}{2\langle k(k-1) \rangle_{out}}. \quad (3.70)$$

In particular when the minimum in-degree and the minimum out-degree of scale-free networks are both greater than 2, i.e.  $P^{in/out}(0) = P^{in/out}(1) = P^{in/out}(2) = 0$ , the fraction of driver nodes is zero in the thermodynamic limit, for any choice of the degree distribution with this property. By changing the minimum in-degree and minimum out-degree of the network the number of driver nodes can change dramatically, independently of the tail of the degree distribution and the level of degree heterogeneity.

In order to use the above calculation to estimate the role of low-degree nodes on

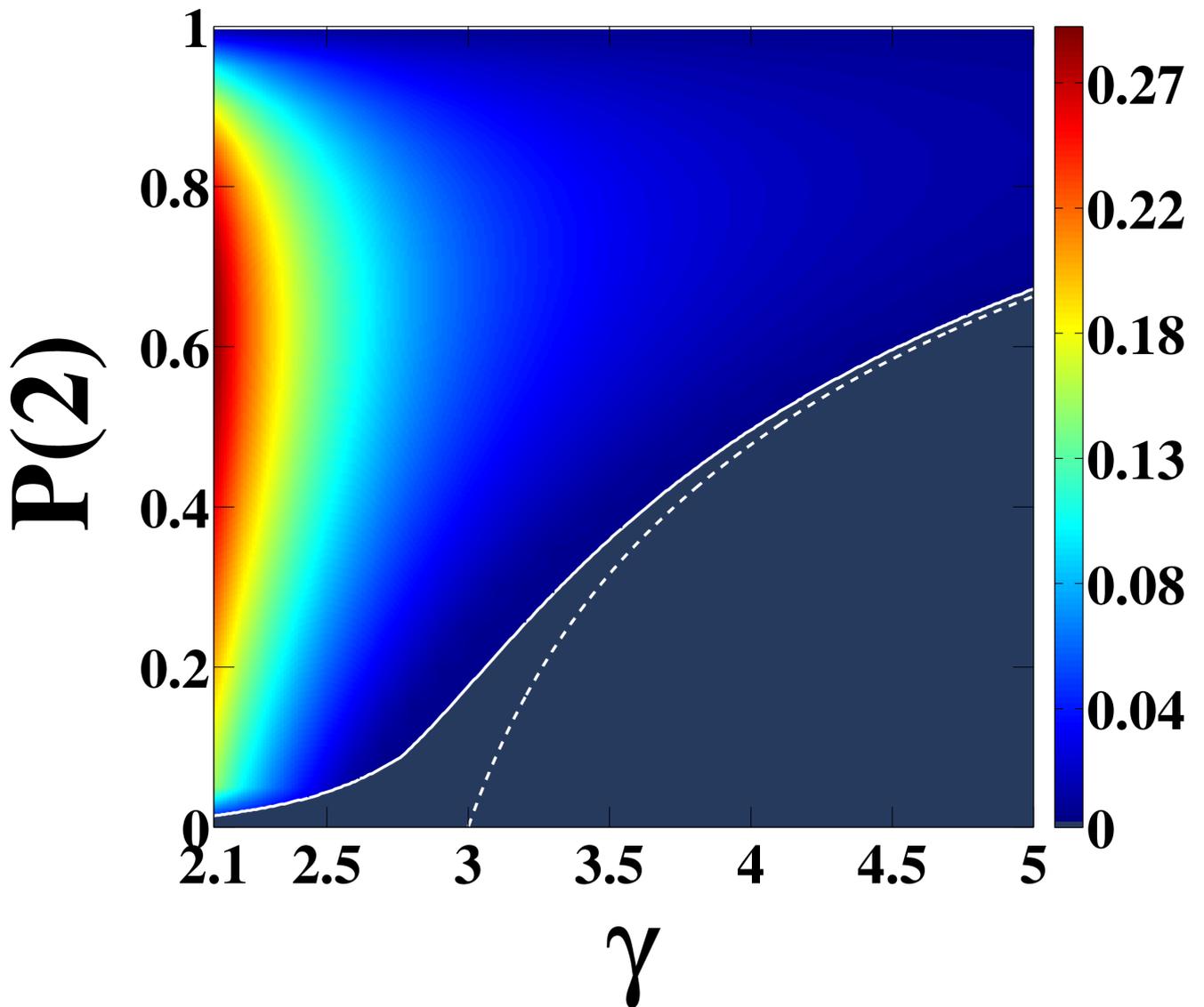


Figure 3.8: Phase diagram of the density of driver nodes  $n_D$  as a function of the parameters  $\gamma$  and  $P(2)$  for networks of  $N = 10^6$  nodes with degree distribution given by Eq. (3.71 and  $P(1) = 0$ . The density  $n_D$  is obtained by numerically solving the BP/MS equations for an ensemble of networks with given degree distribution. The solid lines indicate the stability lines for  $N = 10^6$ , the dotted lines indicate the stability lines in the limit  $N \rightarrow \infty$ .

the fate of the zero-energy solution in finite networks, we consider uncorrelated random graphs with the following power-law degree distribution

$$P^{in}(k) = P^{out}(k) = \begin{cases} P(1) & \text{if } k = 1 \\ P(2) & \text{if } k = 2 \\ Ck^{-\gamma} & \text{if } k \in [3, K] \end{cases} \quad (3.71)$$

with  $C$  a constant determined by normalization and maximum degree

$$K = \min(\sqrt{N}, \{[1 - P(1) - P(2)]N\}^{1/(\gamma-1)}) \quad \gamma > 2$$

$$K = \min(N^{1/\gamma}, \{[1 - P(1) - P(2)]N\}^{1/(\gamma-1)}) \quad \gamma \in (1, 2]$$

that is the minimum between the structural cutoff (Boguñá et al., 2004; Seyed-Allaei et al., 2006) of the network and the natural cutoff of the degree distribution. These networks can be generated numerically using the configuration model. As long as  $P(1) = P(2) = 0$ , the density of driver nodes goes to zero ( $n_D \rightarrow 0$ ) for any exponent  $\gamma > 1$ . More generally, the density  $n_D$  of driver nodes changes dramatically as a function of  $P(1)$  and  $P(2)$  as shown by the heat map in Fig. 3.7 for  $\gamma = 2.1, 3.1$ . Moreover, in Fig. 3.8, we plot the phase diagram for  $P(1) = 0$  indicating the region where the solution  $n_D = 0$  is stable both for a finite network of  $N = 10^6$  nodes (white solid line) and for  $N \rightarrow \infty$  (white dotted line). Note that, for  $\gamma \in (2, 3]$ , stability line converges quite slowly to zero in the infinite size limit.

A confirmation of the validity of this scenario is reported in Fig. 3.9 from a direct comparison of the theoretical results in the ensemble of networks with given degree distribution, with those obtained by the BP algorithm or by computing

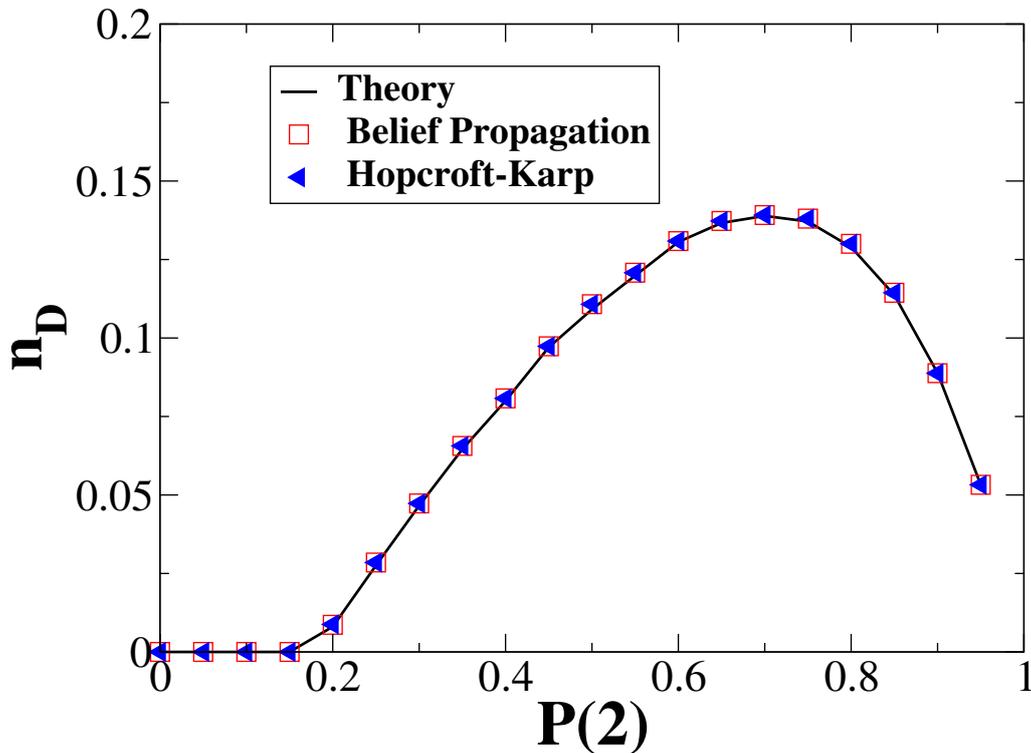


Figure 3.9: Density of driver nodes  $n_D$  as a function of  $P(2)$  for in-degree and out-degree distributions as in Eq. 3.71 with  $P(1) = 0$  and  $\gamma = 2.3$ . The fraction of driver nodes computed with the BP/MS algorithm on a network of  $N = 10^4$  nodes (averaged over 50 network realizations) is compared with the exact results obtained using the Hopcroft-Karp algorithm for maximum matching (Hopcroft and Karp, 1973) and with the theoretical expectation for the density  $n_D$  in an ensemble of random networks with the same degree distribution.

explicitly the maximum matching using the Hopcroft-Karp algorithm (Hopcroft and Karp, 1973) finding very good agreement. Fig. 3.9 also shows that  $n_D$  vanishes by decreasing  $P(2)$ . From our numerical results (see Sec. 4.3.2), in the region in which the solution  $n_D = 0$  is stable and we are far from the stability transition, both algorithms give a zero number of driver nodes  $N_D = 0$ , meaning that all the nodes are matched, and therefore a single external input can be used to control the network.

### 3.5.3 Improving the controllability of a network

These results suggest a simple and very effective way to improve the controllability of a network, by decreasing the fraction of nodes with in-degree and out-degree equal to 0, 1 and 2. Starting from a network with given degree distribution, we first add links starting from any node of out-degree equal to 0 (if present in the network) and randomly attached to any other node of the network, or starting from any random node of the network and ending to nodes of in-degree 0. When there are no more nodes with in-degree or out-degree equal to 0, we repeat the process of random addition of links to nodes with in-degree or out-degree equal to 1 and 2. At the end of the process the minimum in-degree of the network and the minimum out-degree is equal to 3.

Fig. 3.10A shows the reduction in the fraction of driver nodes  $n_D(\Delta L)$  compared to the original one  $n_D(0)$  due to the addition of a fraction  $\Delta L/L_0$  of directed links to a network with pure power-law degree distribution and structural cutoff. It is clear that by lowering the ratio of low in-degree and low out-degree nodes it is possible to reach full controllability of the network. However this can be costly, since for a given network the number of links that need to be added can be a significant fraction of the initial number of links. Nevertheless, by means of this link-addition process, the number of driver nodes decreases steadily and, for example, in the case considered in Fig. 3.10 the number of driver nodes is decreased by 50% just by adding a 12% of links. Finally we have measured how other properties of the network change during this procedure, observing that the clustering coefficient does not change significantly while the average distance decreases. In Sec.4.3.3 we give an example in which this procedure is much more efficient: we consider a network with the previous number of nodes and initial

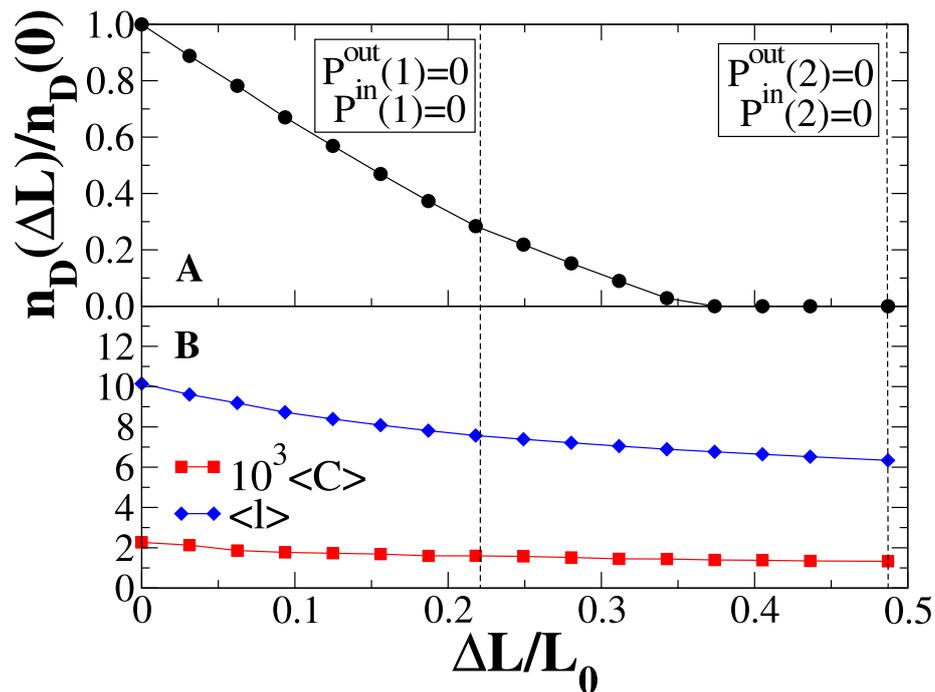


Figure 3.10: Fraction of driver nodes  $n_D(\Delta L)/n_D(0)$  (panel A), average clustering coefficient  $\langle C \rangle$  and average distance  $\langle l \rangle$  (panel B) of the network as a function of the fraction of added links to low degree nodes. The results are obtained from the BP/MS algorithm. The initial network is a power-law network with in-degree distribution equal to the out-degree distribution,  $N = 10^4$  nodes, and power-law exponent  $\gamma = 2.3$ . The symbol  $\Delta L$  indicates the number of added links to the network, whereas  $L_0$  indicates the initial number of links of the network.

average degree, but with the degree distributions with a power-law exponent  $\gamma = 3$ . Note that this procedure can also be applied to networks with other degree distributions as Poisson networks (Sec 4.3.5).

## 3.6 Controllability of multiplex networks

In this section our purpose is the extension of the concept of structural controllability to multiplex networks, thoroughly described in Ch. 2. This section describes the preliminary theoretical work requested for the data analysis that we are currently designing.

A multiplex network  $\vec{G} = (G_1, G_2, \dots, G_K)$  is formed by  $K$  networks  $G_\alpha = (V_\alpha, E_\alpha)$  with  $\alpha = 1, 2, \dots, K$  describing how the  $N$  nodes of the vertex set  $V_\alpha$  interact in each of the layers. We indicate by  $(i, \alpha)$  with  $i = 1, 2, \dots, N$  the nodes in the set  $V_\alpha$  with  $\alpha = 1, 2, \dots, K$  and we call replica nodes the nodes  $(i, \alpha)$  with fixed value of  $i$  and different value of  $\alpha$ . Our goal is to find the minimal number of driver nodes that need to be stimulated by independent signals in order to drive the dynamical state of the multiplex network to any desired state. Moreover, we impose that the independent external signals are applied only to replica nodes (see Fig. 3.11). This design seems particularly interesting especially regarding a possible integration of different types of brain networks. For example, for the same white-matter/grey-matter interface partition we can have data of diffusion MRI, the so-called “structural brain”, and data coming from the analysis of functional MRI, also known as “functional brain”.

### 3.6.1 The structural controllability of a multiplex network

We consider a multiplex networks in which every node  $i = 1, 2, \dots, N$  has a replica node in each layer and every layer is formed by a directed network between the corresponding replica nodes. We assume that each replica node can have a different dynamical state and can send different signals in the different networks

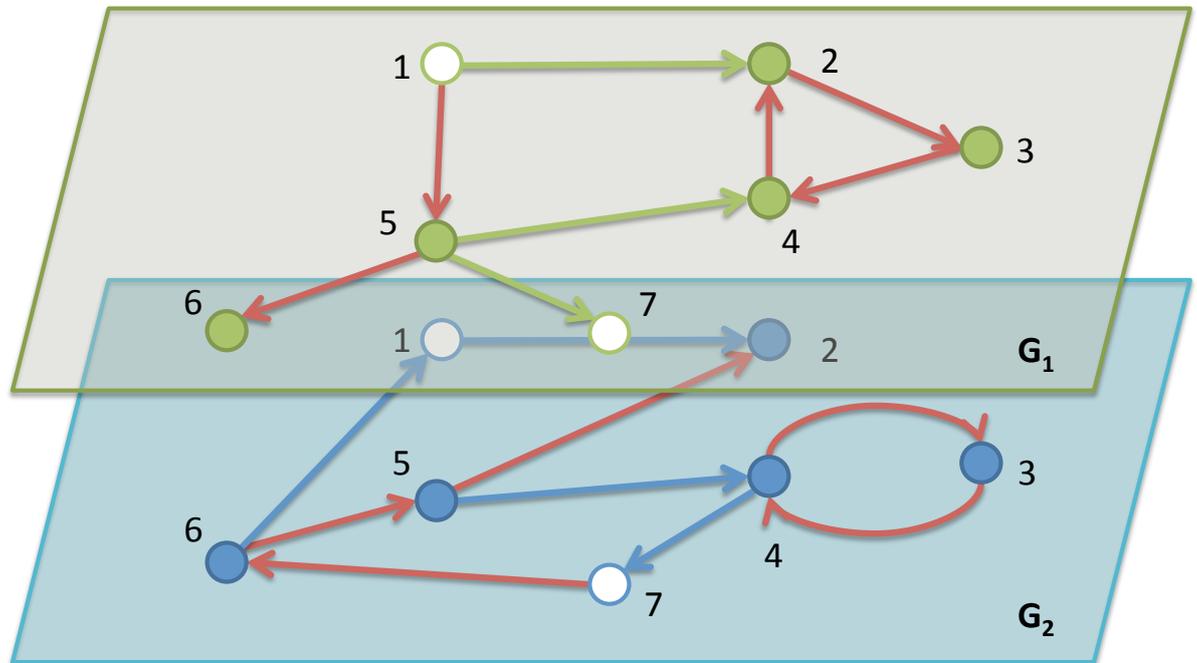


Figure 3.11: Controllability of a duplex network: in the two layers we force the driver nodes (white nodes) to be the same. Red links define the matching in each layer.

(each layers is characterised by a different dynamical process). In this case the controllability of the layers can be treated by control theory methods used for single layer taken in isolation. Looking back to the definitions of Sec. 3.2, for each network  $G^\alpha = (V^\alpha, E^\alpha)$  associated with layer  $\alpha$  the dynamical state can be controlled by applying  $M^\alpha$  independent signals to the driver nodes, according to

the equation

$$\frac{d\mathbf{x}^\alpha(t)}{dt} = A^\alpha \mathbf{x}^\alpha + B^\alpha \mathbf{u}^\alpha, \quad (3.72)$$

in which the vector  $\mathbf{x}^\alpha(t)$ , of elements  $x_i^\alpha(t)$  with  $i = 1, 2, \dots, N$ , represents the dynamical state of the network in layer  $\alpha$ ,  $A^\alpha$  is the  $N \times N$  state matrix of layer  $\alpha$ , and  $B^\alpha$  is the  $N \times M^\alpha$  input matrix describing the interaction between the replica nodes of the layers and  $M^\alpha \leq N$  external signals, indicated by the vector  $\mathbf{u}^\alpha(t)$  of elements  $u_\mu^\alpha$  and  $\mu = 1, 2 \dots M^\alpha$ . Each layer of the multiplex networks can be structurally controlled by identifying a minimum number of driver nodes. Here we make the assumption that in the multiplex network, the driver nodes must be the same in the different layers supervising the entire multiplex network at the same time. For the sake of simplicity, we consider a multiplex network formed by two layers. Finding the driver nodes of the duplex network, can be mapped to a matching problem where

- every node has at most one matched incoming link
- every node has at most one matched outgoing link
- any two replica nodes either have no matched incoming links on each layer or have one matched incoming link in each layer

This problem can be studied by statistical mechanics tools on ensemble of duplex networks and on single duplex network realisations by the use of the cavity method and the Belief Propagation algorithm providing the minimal number of driver nodes.

### 3.6.2 Mapping to a constraint Maximum Matching Problem

We consider a duplex network with layers  $\alpha = 1, 2$ . We impose that the driver nodes in the two networks are replica nodes and we minimise the number of driver nodes in the multiplex network. In order to build an algorithm able to find the driver nodes in the multiplex satisfying our constraint we consider the variables  $s_{ij}^\alpha = 1, 0$  indicating respectively if the directed link from node  $(i, \alpha)$  to node  $(j, \alpha)$  in layer  $\alpha = 1, 2$  is matched or not.

In the two layer of the multiplex we want to have a matching, i.e. the following constraints must always be satisfied.

$$\begin{aligned} \sum_{j \in \partial_+^\alpha i} s_{ij}^\alpha &\leq 1, \\ \sum_{j \in \partial_-^\alpha i} s_{ji}^\alpha &\leq 1. \end{aligned} \quad (3.73)$$

In addition, we have the following constraints

$$\sum_{j \in \partial_-^{[1]} i} s_{ji}^{[1]} = \sum_{j \in \partial_-^{[2]} i} s_{ji}^{[2]}. \quad (3.74)$$

These condition impose that each node has either two unmatched replicas or two matched replicas in the two networks of the duplex. We consider the energy of the problem  $E(\beta)$  given by

$$\begin{aligned} E(\beta) &= \sum_\alpha \sum_i \left( 1 - \sum_{j \in \partial_-^\alpha i} s_{ji}^\alpha \right) \\ &= \sum_\alpha \sum_i E_i^\alpha, \end{aligned} \quad (3.75)$$

considering now

$$E_i^\alpha = 1 - \sum_{j \in \partial_+^\alpha i} s_{ij}^\alpha. \quad (3.76)$$

### 3.6.3 BP Equations

In Sec. 3.4 we gave the general formalism of the cavity method for the matching problem over single directed graph and the related generalisation to an ensemble of networks. The Boltzmann distribution  $P(\{s_{ij}\})$  for this problem is given by

$$\begin{aligned} P(\{s_{ij}\}) &= \frac{e^{-\beta E}}{Z} \prod_{\alpha} \prod_{i=1}^N \theta \left( 1 - \sum_{j \in \partial_+^\alpha i} s_{ij}^\alpha \right) \\ &\quad \times \prod_{\alpha} \prod_{i=1}^N \theta \left( 1 - \sum_{j \in \partial_-^\alpha i} s_{ji}^\alpha \right) \\ &\quad \times \delta \left( \sum_{i \in \partial_-^{[1]} j} s_{ij}^{[1]}, \sum_{i \in \partial_-^{[2]} j} s_{ij}^{[2]} \right). \end{aligned} \quad (3.77)$$

where  $\theta(x) = 1$  for  $x \geq 0$  and  $\theta(x) = 0$  for  $x < 0$ ,  $\delta(x)$  is the Kronecker delta, and where  $Z$  is the usual partition function of the statistical mechanics problem. Subsequently, we perform the limit  $\beta \rightarrow \infty$  in order to characterise the maximum-sized matching in the two layers, enforcing Eqs. 3.73, 3.74. The distribution  $P(\{s_{ij}\})$  on a locally tree-like network can be solved by the BP message passing method by finding the messages that nearby nodes sent to each other. In particular we distinguish between messages that in layer  $\alpha$  are going in the direction of the link,  $P_{i \rightarrow j}^\alpha(s_{ij})$ , and the messages that in layer  $\alpha$  are going in the opposite direction of the link,  $\hat{P}_{i \rightarrow j}^\alpha(s_{ji})$ . We have then four types of messages and their BP equations are

$$\begin{aligned}
P_{i \rightarrow j}^{[1]}(s_{ij}^{[1]}) &= \frac{1}{\mathcal{D}_{i \rightarrow j}^{[1]}} \sum_{\{s_{ik}^{[1]}\} | k \in \partial_+^{[1]} i \setminus j} \sum_{\{s_{ik}^{[2]}\} | k \in \partial_+^{[2]} i} \left\{ \theta \left( 1 - \sum_{k \in \partial_+^{[1]} i} s_{ik}^{[1]} \right) \right. \\
&\quad \theta \left( 1 - \sum_{k \in \partial_+^{[2]} i} s_{ik}^{[2]} \right) \exp \left[ -\beta \left( 1 - \sum_{k \in \partial_+^{[1]} i} s_{ik}^{[1]} \right) \right] \exp \left[ -\beta \left( 1 - \sum_{k \in \partial_+^{[2]} i} s_{ik}^{[2]} \right) \right] \\
&\quad \left. \prod_{k \in \partial_+^{[1]} i \setminus j} \hat{P}_{k \rightarrow i}^{[1]}(s_{ik}^{[1]}) \prod_{k \in \partial_+^{[2]} i} \hat{P}_{k \rightarrow i}^{[2]}(s_{ik}^{[2]}) \right\}, \\
P_{i \rightarrow j}^{[2]}(s_{ij}^{[2]}) &= \frac{1}{\mathcal{D}_{i \rightarrow j}^{[2]}} \sum_{\{s_{ik}^{[2]}\} | k \in \partial_+^{[2]} i \setminus j} \sum_{\{s_{ik}^{[1]}\} | k \in \partial_+^{[1]} i} \left\{ \theta \left( 1 - \sum_{k \in \partial_+^{[2]} i} s_{ik}^{[2]} \right) \right. \\
&\quad \theta \left( 1 - \sum_{k \in \partial_+^{[1]} i} s_{ik}^{[1]} \right) \exp \left[ -\beta \left( 1 - \sum_{k \in \partial_+^{[2]} i} s_{ik}^{[2]} \right) \right] \exp \left[ -\beta \left( 1 - \sum_{k \in \partial_+^{[1]} i} s_{ik}^{[1]} \right) \right] \\
&\quad \left. \prod_{k \in \partial_+^{[2]} i \setminus j} \hat{P}_{k \rightarrow i}^{[2]}(s_{ik}^{[2]}) \prod_{k \in \partial_+^{[1]} i} \hat{P}_{k \rightarrow i}^{[1]}(s_{ik}^{[1]}) \right\}, \\
\hat{P}_{i \rightarrow j}^{[1]}(s_{ji}^{[1]}) &= \frac{1}{\hat{\mathcal{D}}_{i \rightarrow j}^{[1]}} \sum_{\{s_{ki}^{[1]}\} | k \in \partial_-^{[1]} i \setminus j} \sum_{\{s_{ki}^{[2]}\} | k \in \partial_-^{[2]} i} \left\{ \theta \left( 1 - \sum_{k \in \partial_-^{[1]} i} s_{ki}^{[1]} \right) \right. \\
&\quad \times \theta \left( 1 - \sum_{k \in \partial_-^{[2]} i} s_{ki}^{[2]} \right) \delta \left( \sum_{i \in \partial_-^{[1]} j} s_{ij}^{[1]}, \sum_{i \in \partial_-^{[2]} j} s_{ij}^{[2]} \right) \\
&\quad \left. \times \prod_{k \in \partial_-^{[1]} i \setminus j} P_{k \rightarrow i}^{[1]}(s_{ki}^{[1]}) \prod_{k \in \partial_-^{[2]} i} P_{k \rightarrow i}^{[2]}(s_{ki}^{[2]}) \right\}, \\
\hat{P}_{i \rightarrow j}^{[2]}(s_{ji}^{[2]}) &= \frac{1}{\hat{\mathcal{D}}_{i \rightarrow j}^{[2]}} \sum_{\{s_{ki}^{[2]}\} | k \in \partial_-^{[2]} i \setminus j} \left\{ \theta \left( 1 - \sum_{k \in \partial_-^{[2]} i} s_{ki}^{[2]} \right) \right. \\
&\quad \times \sum_{\{s_{ki}^{[1]}\} | k \in \partial_-^{[1]} i} \left[ \theta \left( 1 - \sum_{k \in \partial_-^{[1]} i} s_{ki}^{[1]} \right) \delta \left( \sum_{i \in \partial_-^{[1]} j} s_{ij}^{[1]}, \sum_{i \in \partial_-^{[2]} j} s_{ij}^{[2]} \right) \right. \\
&\quad \left. \left. \times \prod_{k \in \partial_-^{[2]} i \setminus j} P_{k \rightarrow i}^{[2]}(s_{ki}^{[2]}) \prod_{k \in \partial_-^{[1]} i} P_{k \rightarrow i}^{[1]}(s_{ki}^{[1]}) \right] \right\} \quad (3.78)
\end{aligned}$$

where  $\mathcal{D}_{i \rightarrow j}^\alpha$  and  $\hat{\mathcal{D}}_{i \rightarrow j}^\alpha$  are normalisation constants. Similarly to Eqs. 3.49 previously defined for a single network, the messages  $\{P_{i \rightarrow j}^\alpha(s_{ij}^\alpha), \hat{P}_{i \rightarrow j}^\alpha(s_{ji}^\alpha)\}$  can be parametrised by the cavity fields  $h_{i \rightarrow j}$  and  $\hat{h}_{i \rightarrow j}$ , defined by

$$\frac{P_{i \rightarrow j}^\alpha(1)}{P_{i \rightarrow j}^\alpha(0)} = e^{\beta h_{i \rightarrow j}} \quad \frac{\hat{P}_{i \rightarrow j}^\alpha(1)}{\hat{P}_{i \rightarrow j}^\alpha(0)} = e^{\beta \hat{h}_{i \rightarrow j}} \quad (3.79)$$

and rearranged in

$$P_{i \rightarrow j}^\alpha(s_{ij}^\alpha) = \frac{e^{\beta h_{i \rightarrow j} s_{ij}^\alpha}}{1 + e^{\beta h_{i \rightarrow j}}} = e^{\beta h_{i \rightarrow j} s_{ij}^\alpha} P_{i \rightarrow j}^\alpha(0) \quad (3.80)$$

$$\hat{P}_{i \rightarrow j}^\alpha(s_{ji}^\alpha) = \frac{e^{\beta \hat{h}_{i \rightarrow j} s_{ji}^\alpha}}{1 + e^{\beta \hat{h}_{i \rightarrow j}}} = e^{\beta \hat{h}_{i \rightarrow j} s_{ji}^\alpha} \hat{P}_{i \rightarrow j}^\alpha(0) \quad (3.81)$$

In terms of the cavity fields, Eqs. 3.78 reduce to the following set of equations,

$$\begin{aligned} h_{i \rightarrow j}^\alpha &= -\frac{1}{\beta} \log \left( e^{-\beta} + \sum_{k \in \partial_+^\alpha i \setminus j} e^{\beta \hat{h}_{k \rightarrow i}^\alpha} \right), \\ \hat{h}_{i \rightarrow j}^{[1]} &= -\frac{1}{\beta} \log \left( \frac{1}{\sum_{k \in \partial_-^{[2]} i} e^{\beta h_{k \rightarrow i}^{[2]}}} + \sum_{k \in \partial_-^{[1]} i \setminus j} e^{\beta h_{k \rightarrow i}^{[1]}} \right), \\ \hat{h}_{i \rightarrow j}^{[2]} &= -\frac{1}{\beta} \log \left( \frac{1}{\sum_{k \in \partial_-^{[1]} i} e^{\beta h_{k \rightarrow i}^{[1]}}} + \sum_{k \in \partial_-^{[2]} i \setminus j} e^{\beta h_{k \rightarrow i}^{[2]}} \right), \end{aligned} \quad (3.82)$$

These last equations are derived with the usual recipe: they follow from Eqs.

3.78 introducing Eqs 3.80, 3.81, i.e.

$$\begin{aligned}
P_{i \rightarrow j}^{[1]}(0) &= \frac{1}{\mathcal{D}_{i \rightarrow j}^{[1]}} \left[ e^{-\beta} + \sum_{k \in \partial_+ i^{[1]} \setminus j} e^{\beta \hat{h}_{k \rightarrow i}^{[1]}} \right] \prod_{k \in \partial_+ i^{[1]} \setminus j} \hat{P}_{k \rightarrow i}^{[1]}(0) \\
&\times \left[ e^{-\beta} + \sum_{k \in \partial_+ i^{[2]} \setminus j} e^{\beta \hat{h}_{k \rightarrow i}^{[2]}} \right] \prod_{k \in \partial_+ i^{[2]} \setminus j} \hat{P}_{k \rightarrow i}^{[2]}(0) \\
P_{i \rightarrow j}^{[1]}(1) &= \frac{1}{\mathcal{D}_{i \rightarrow j}^{[1]}} \prod_{k \in \partial_+ i^{[1]} \setminus j} \hat{P}_{k \rightarrow i}^{[1]}(0) \left[ e^{-\beta} + \sum_{k \in \partial_+ i^{[2]} \setminus j} e^{\beta \hat{h}_{k \rightarrow i}^{[2]}} \right] \prod_{k \in \partial_+ i^{[2]} \setminus j} \hat{P}_{k \rightarrow i}^{[2]}(0) \\
P_{i \rightarrow j}^{[2]}(0) &= \frac{1}{\mathcal{D}_{i \rightarrow j}^{[2]}} \left[ e^{-\beta} + \sum_{k \in \partial_+ i^{[2]} \setminus j} e^{\beta \hat{h}_{k \rightarrow i}^{[2]}} \right] \prod_{k \in \partial_+ i^{[2]} \setminus j} \hat{P}_{k \rightarrow i}^{[2]}(0) \\
&\times \left[ e^{-\beta} + \sum_{k \in \partial_+ i^{[1]} \setminus j} e^{\beta \hat{h}_{k \rightarrow i}^{[1]}} \right] \prod_{k \in \partial_+ i^{[1]} \setminus j} \hat{P}_{k \rightarrow i}^{[1]}(0) \\
P_{i \rightarrow j}^{[2]}(1) &= \frac{1}{\mathcal{D}_{i \rightarrow j}^{[2]}} \prod_{k \in \partial_+ i^{[2]} \setminus j} \hat{P}_{k \rightarrow i}^{[2]}(0) \left[ e^{-\beta} + \sum_{k \in \partial_+ i^{[1]} \setminus j} e^{\beta \hat{h}_{k \rightarrow i}^{[1]}} \right] \prod_{k \in \partial_+ i^{[1]} \setminus j} \hat{P}_{k \rightarrow i}^{[1]}(0) \\
\hat{P}_{i \rightarrow j}^{[1]}(0) &= \frac{1}{\hat{\mathcal{D}}_{i \rightarrow j}^{[1]}} \left[ 1 + \sum_{k \in \partial_- i^{[1]} \setminus j} e^{\beta h_{k \rightarrow i}^{[1]}} \sum_{k \in \partial_- i^{[2]}} e^{\beta h_{k \rightarrow i}^{[2]}} \right] \prod_{k \in \partial_- i^{[1]} \setminus j} P_{k \rightarrow i}^{[1]}(0) \prod_{k \in \partial_- i^{[2]}} P_{k \rightarrow i}^{[2]}(0) \\
\hat{P}_{i \rightarrow j}^{[1]}(1) &= \frac{1}{\hat{\mathcal{D}}_{i \rightarrow j}^{[1]}} \sum_{k \in \partial_- i^{[2]}} e^{\beta h_{k \rightarrow i}^{[2]}} \prod_{k \in \partial_- i^{[1]} \setminus j} P_{k \rightarrow i}^{[1]}(0) \prod_{k \in \partial_- i^{[2]}} P_{k \rightarrow i}^{[2]}(0) \\
\hat{P}_{i \rightarrow j}^{[2]}(0) &= \frac{1}{\hat{\mathcal{D}}_{i \rightarrow j}^{[2]}} \left[ 1 + \sum_{k \in \partial_- i^{[2]} \setminus j} e^{\beta h_{k \rightarrow i}^{[2]}} \sum_{k \in \partial_- i^{[1]}} e^{\beta h_{k \rightarrow i}^{[1]}} \right] \prod_{k \in \partial_- i^{[2]} \setminus j} P_{k \rightarrow i}^{[2]}(0) \prod_{k \in \partial_- i^{[1]}} P_{k \rightarrow i}^{[1]}(0) \\
\hat{P}_{i \rightarrow j}^{[2]}(1) &= \frac{1}{\hat{\mathcal{D}}_{i \rightarrow j}^{[2]}} \sum_{k \in \partial_- i^{[1]}} e^{\beta h_{k \rightarrow i}^{[1]}} \prod_{k \in \partial_- i^{[2]} \setminus j} P_{k \rightarrow i}^{[2]}(0) \prod_{k \in \partial_- i^{[1]}} P_{k \rightarrow i}^{[1]}(0)
\end{aligned}$$

and finally considering Eqs. 3.79. The equations for  $P_{i \rightarrow j}^\alpha(s_{ij}^\alpha)$  clearly factorise in two terms, one for each layer.

For our model the Bethe approximation of the probability distribution  $P(\{s_{ij}\})$  reads

$$P_{Bethe}^{duplex}(\{s_{ij}\}^{[1]}, \{s_{ij}\}^{[2]}) = \frac{\prod_{i=1}^N P_i(\underline{S}_i^{[1],+}) P_i(\underline{S}_i^{[2],+}) P_i(\underline{S}_i^{[1,-]}, \underline{S}_i^{[2,-]})}{\left( \prod_{\langle i,j \rangle} P_{ij}(s_{ij}^{[1]}) P_{ij}(s_{ij}^{[2]}) \right)} \quad (3.83)$$

where  $P_i(\underline{S}_i^{[\alpha],+})$  is the marginal distribution over node  $i$  considering everything that is pointed by  $i$  in layer  $\alpha$ , while  $P_i(\underline{S}_i^{[1,-]}, \underline{S}_i^{[2,-]})$  is the marginal distribution for node  $i$  taking into account everything that points to  $i$  in both layers. Finally,  $P_{ij}(s_{ij}^\alpha)$  is the usual marginal distribution over link  $(i, j)$  in layer  $\alpha$ . The explicit expressions for these marginal probabilities read

$$\begin{aligned}
 P_i(\underline{S}_i^{[\alpha],+}) &= \frac{1}{\mathcal{C}_i^{+,\alpha}} \exp \left[ -\beta \left( 1 - \sum_{k \in \partial_+^{[\alpha]} i} s_{ik}^{[\alpha]} \right) \right] \theta \left( 1 - \sum_{k \in \partial_+^\alpha i} s_{ik}^\alpha \right) \prod_{k \in \partial_+ i^{[\alpha]}} \hat{P}_{k \rightarrow i}^{[\alpha]}(s_{ik}^{[\alpha]}) \\
 P_i(\underline{S}_i^{[1,-]}, \underline{S}_i^{[2,-]}) &= \frac{1}{\mathcal{C}_i^-} \theta \left( 1 - \sum_{j \in \partial_-^{[1]} i} s_{ji}^{[1]} \right) \theta \left( 1 - \sum_{j \in \partial_-^{[2]} i} s_{ji}^{[2]} \right) \delta \left( \sum_{i \in \partial_-^{[1]} j} s_{ij}^{[1]}, \sum_{i \in \partial_-^{[2]} j} s_{ij}^{[2]} \right) \\
 &\quad \times \prod_{k \in \partial_- i^{[1]}} P_{k \rightarrow i}^{[1]}(s_{ki}^{[1]}) \prod_{k \in \partial_- i^{[2]}} P_{k \rightarrow i}^{[2]}(s_{ki}^{[2]}) \\
 P_{ij}(s_{ij}^\alpha) &= \frac{1}{\mathcal{C}_{ij}^\alpha} P_{i \rightarrow j}^\alpha(s_{ij}^\alpha) \hat{P}_{j \rightarrow i}^\alpha(s_{ij}^\alpha)
 \end{aligned} \tag{3.84}$$

where  $\mathcal{C}_i^{+,\alpha}$ ,  $\mathcal{C}_i^-$ , and  $\mathcal{C}_{ij}^\alpha$  are normalisation constant given by

$$\begin{aligned}
 \mathcal{C}_i^{+,\alpha} &= \left( e^{-\beta} + \sum_{k \in \partial_+ i^\alpha} e^{\beta \hat{h}_{k \rightarrow i}^\alpha} \right) \prod_{k \in \partial_+ i^\alpha} \hat{P}_{k \rightarrow i}^\alpha(0) \\
 \mathcal{C}_i^- &= \left[ 1 + \sum_{k \in \partial_- i^{[1]}} e^{\beta h_{k \rightarrow i}^{[1]}} \sum_{k \in \partial_- i^{[2]}} e^{\beta h_{k \rightarrow i}^{[2]}} \right] \prod_{k \in \partial_- i^{[1]}} P_{k \rightarrow i}^{[1]}(0) \prod_{k \in \partial_- i^{[2]}} P_{k \rightarrow i}^{[2]}(0) \\
 \mathcal{C}_{ij}^\alpha &= (1 + e^{\beta(h_{i \rightarrow j}^\alpha + \hat{h}_{j \rightarrow i}^\alpha)}) P_{i \rightarrow j}^\alpha(0) \hat{P}_{j \rightarrow i}^\alpha(0).
 \end{aligned} \tag{3.85}$$

The free energy  $\beta F = -\log Z$  is then computed following the expression of the partition function  $Z$  by means of the normalisation constants, namely

$$Z = \frac{\prod_i \mathcal{C}_i^{+,[1]} \mathcal{C}_i^{+,[2]} \mathcal{C}_i^-}{\prod_{\langle i,j \rangle} \mathcal{C}_{ij}^{[1]} \mathcal{C}_{ij}^{[2]}} \tag{3.86}$$

The free energy reads

$$\begin{aligned}
-\beta F &= \sum_{\alpha} \sum_{i=1}^N \left[ \ln \left( e^{-\beta} + \sum_{k \in \partial_{+}^{\alpha} i} e^{\beta \hat{h}_{k \rightarrow i}^{\alpha}} \right) \right] \\
&+ \sum_{i=1, N} \ln \left( 1 + \sum_{k \in \partial_{-}^A i} e^{\beta h_{k \rightarrow i}^A} \sum_{k' \in \partial_{-}^B i} e^{\beta h_{k' \rightarrow i}^B} \right) \\
&- \sum_{\alpha} \sum_{\langle i, j \rangle_{\alpha}} \ln \left( 1 + e^{\beta (h_{i \rightarrow j}^{\alpha} + \hat{h}_{j \rightarrow i}^{\alpha})} \right), \tag{3.87}
\end{aligned}$$

Finally, the energy  $E = \frac{\partial \beta F}{\partial \beta}$  of the model is given by

$$\begin{aligned}
E &= \sum_{\alpha} \sum_{i=1}^N \left[ \frac{e^{-\beta} - \sum_{k \in \partial_{+}^{\alpha} i} \hat{h}_{k \rightarrow i}^{\alpha} e^{\beta \hat{h}_{k \rightarrow i}^{\alpha}}}{1 + \sum_{k \in \partial_{+}^{\alpha} i} e^{\beta \hat{h}_{k \rightarrow i}^{\alpha}}} \right] \\
&- \sum_{i=1}^N \frac{\sum_{k \in \partial_{-}^{[1]} i} h_{k \rightarrow i}^{[1]} e^{\beta h_{k \rightarrow i}^{[1]}} \sum_{k' \in \partial_{-}^{[2]} i} e^{\beta h_{k' \rightarrow i}^{[2]}}}{1 + \sum_{k \in \partial_{-}^{[1]} i} e^{\beta h_{k \rightarrow i}^{[1]}} \sum_{k' \in \partial_{-}^{[2]} i} e^{\beta h_{k' \rightarrow i}^{[2]}}} \\
&- \sum_{i=1}^N \frac{\sum_{k \in \partial_{-}^{[1]} i} h_{k \rightarrow i}^{[1]} e^{\beta h_{k \rightarrow i}^{[1]}} \sum_{k' \in \partial_{-}^{[2]} i} h_{k' \rightarrow i}^{[2]} e^{\beta h_{k' \rightarrow i}^{[2]}}}{1 + \sum_{k \in \partial_{-}^{[1]} i} e^{\beta h_{k \rightarrow i}^{[1]}} \sum_{k' \in \partial_{-}^{[2]} i} e^{\beta h_{k' \rightarrow i}^{[2]}}} \\
&+ \sum_{\alpha} \sum_{\langle i, j \rangle_{\alpha}} \frac{(h_{i \rightarrow j}^{\alpha} + \hat{h}_{j \rightarrow i}^{\alpha}) e^{\beta (h_{i \rightarrow j}^{\alpha} + \hat{h}_{j \rightarrow i}^{\alpha})}}{1 + e^{\beta (h_{i \rightarrow j}^{\alpha} + \hat{h}_{j \rightarrow i}^{\alpha})}}. \tag{3.88}
\end{aligned}$$

In the limit  $\beta \rightarrow \infty$  the BP or (Max-Sum) equations determining the values of these fields are given by

$$\begin{aligned}
h_{i \rightarrow j}^{\alpha} &= -\max \left\{ -1, \max_{k \in \partial_{+}^{\alpha} i \setminus j} \hat{h}_{k \rightarrow i}^{\alpha} \right\} \\
\hat{h}_{i \rightarrow j}^{[1]} &= -\max \left\{ \max_{k \in \partial_{-}^{[1]} i \setminus j} h_{k \rightarrow i}^{[1]}, \min \left[ 0, \max_{k \in \partial_{-}^{[2]} j} h_{k \rightarrow i}^{[2]} \right] \right\} \\
\hat{h}_{i \rightarrow j}^{[2]} &= -\max \left\{ \max_{k \in \partial_{-}^{[2]} i \setminus j} h_{k \rightarrow i}^{[2]}, \min \left[ 0, \max_{k \in \partial_{-}^{[1]} j} h_{k \rightarrow i}^{[1]} \right] \right\}. \tag{3.89}
\end{aligned}$$

These equations close on the set of values  $\{-1, 0, 1\}$  for the fields  $h_{i \rightarrow j}^{\alpha}$  and  $\hat{h}_{i \rightarrow j}^{\alpha}$ .

The energy  $E$  can also be expressed in terms of the fields and is given by

$$\begin{aligned}
E = & - \sum_{\alpha} \sum_{i=1}^N \max \left[ -1, \max_{k \in \partial_{+}^{\alpha} i} \hat{h}_{k \rightarrow i}^{\alpha} \right] \\
& - \sum_{i=1}^N \max \left[ 0, \max_{k \in \partial_{-}^{[1]} i} h_{k \rightarrow i}^{[1]} + \max_{k \in \partial_{-}^{[2]} i} h_{k \rightarrow i}^{[2]} \right] \\
& + \sum_{\alpha} \sum_{\langle i, j \rangle} \max \left[ 0, h_{i \rightarrow j}^{\alpha} + \hat{h}_{j \rightarrow i}^{\alpha} \right]
\end{aligned} \tag{3.90}$$

### 3.6.4 The limit $\beta \rightarrow \infty$

The derivation of  $P_{\alpha}(h^{\alpha})$  and  $\hat{P}_{\alpha}(\hat{h}^{\alpha})$  for  $\beta$  finite is pretty similar to the procedures exposed in Sec. 3.4.4. The distribution of the fields over this ensemble of networks for  $\beta \rightarrow \infty$  is given by

$$\begin{aligned}
P_{\alpha}(h^{\alpha}) &= w_1^{\alpha} \delta(h^{\alpha} - 1) + w_2^{\alpha} \delta(h^{\alpha} + 1) + w_3^{\alpha} \delta(h^{\alpha}), \\
\hat{P}_{\alpha}(\hat{h}^{\alpha}) &= \hat{w}_1^{\alpha} \delta(\hat{h}^{\alpha} - 1) + \hat{w}_2^{\alpha} \delta(\hat{h}^{\alpha} + 1) + \hat{w}_3^{\alpha} \delta(\hat{h}^{\alpha}),
\end{aligned} \tag{3.91}$$

where the probabilities  $w_1, w_2, w_3$  are normalized  $w_1 + w_2 + w_3 = 1$  as well as the probabilities  $\hat{w}_1, \hat{w}_2, \hat{w}_3$  that satisfy  $\hat{w}_1 + \hat{w}_2 + \hat{w}_3 = 1$ . The BP equations can be written as equations for the set of probabilities  $\{w^{[1]}\}, \{\hat{w}^{[1]}\}, \{w^{[2]}\}, \{\hat{w}^{[2]}\}$  obtaining

$$\begin{aligned}
w_1^\alpha &= \sum_{k^\alpha} \frac{k^\alpha}{\langle k^\alpha \rangle_{out}} P_\alpha^{out}(k^\alpha) (\hat{w}_2^\alpha)^{k-1} \\
w_2^\alpha &= \sum_{k^\alpha} \frac{k^\alpha}{\langle k^\alpha \rangle_{out}} P_\alpha^{out}(k^\alpha) [1 - (1 - \hat{w}_1^\alpha)^{k-1}] \\
w_3^\alpha &= 1 - w_1^\alpha - w_2^\alpha \\
\hat{w}_3^\alpha &= 1 - \hat{w}_1^\alpha - \hat{w}_2^\alpha \\
\hat{w}_1^{[1]} &= \sum_{k^{[1]}, k^{[2]}} \frac{k^{[1]}}{\langle k^{[1]} \rangle_{in}} P^{in}(k^{[1]}, k^{[2]}) (w_2^{[1]})^{k^{[1]}-1} [1 - (1 - w_1^{[2]})^{k^{[2]}}] \\
\hat{w}_2^{[1]} &= \sum_{k^{[1]}, k^{[2]}} \frac{k^{[1]}}{\langle k^{[1]} \rangle_{in}} P^{in}(k^{[1]}, k^{[2]}) \left[ 1 - (1 - w_1^{[1]})^{k^{[1]}-1} + (1 - w_1^{[1]})^{k^{[1]}-1} (w_2^{[2]})^{k^{[2]}} \right] \\
\hat{w}_1^{[2]} &= \sum_{k^{[1]}, k^{[2]}} \frac{k^{[2]}}{\langle k^{[2]} \rangle_{in}} P^{in}(k^{[1]}, k^{[2]}) (w_2^{[2]})^{k^{[2]}-1} [1 - (1 - w_1^{[1]})^{k^{[1]}}] \\
\hat{w}_2^{[2]} &= \sum_{k^{[1]}, k^{[2]}} \frac{k^{[2]}}{\langle k^{[2]} \rangle_{in}} P^{in}(k^{[1]}, k^{[2]}) \left[ 1 - (1 - w_1^{[2]})^{k^{[2]}-1} + (1 - w_1^{[2]})^{k^{[2]}-1} (w_2^{[1]})^{k^{[1]}} \right]
\end{aligned} \tag{3.92}$$

Finally, the energy density in this limit becomes

$$\begin{aligned}
n_D &= \frac{E}{N} = \sum_\alpha \sum_k P_\alpha^{out}(k^\alpha) \left\{ (\hat{w}_2^\alpha)^{k^\alpha} - [1 - (1 - \hat{w}_1^\alpha)^{k^\alpha}] \right\} \\
&\quad - \sum_{k^{[1]}, k^{[2]}} P^{in}(k^{[1]}, k^{[2]}) [1 - (1 - w_1^{[1]})^{k^{[1]}}] [1 - (w_2^{[2]})^{k^{[2]}}] \\
&\quad - \sum_{k^{[1]}, k^{[2]}} P^{in}(k^{[1]}, k^{[2]}) [1 - (1 - w_1^{[2]})^{k^{[2]}}] [1 - (w_2^{[1]})^{k^{[1]}}] \\
&\quad + \sum_\alpha \langle k^\alpha \rangle_{in} [\hat{w}_1^\alpha (1 - w_2^\alpha) + w_1^\alpha (1 - \hat{w}_2^\alpha)]
\end{aligned} \tag{3.93}$$

### 3.6.5 Controllability of uncorrelated multiplex networks with given in-degree and out-degree distribution

Let us consider the case of uncorrelated duplex networks in which the degree of the same node in different layers are uncorrelated and there is no significant overlap of the links. In this case each layer is formed by a network built using the configuration model. If we define the generating functions,

$$\begin{aligned}
 G_0^{\alpha,in}(z) &= \sum_k P_\alpha^{in}(k) z^k, \\
 G_1^{\alpha,in}(z) &= \sum_k \frac{k}{\langle k^\alpha \rangle} P_\alpha^{in}(k) z^{k-1}, \\
 G_0^{\alpha,out}(z) &= \sum_k P_\alpha^{out}(k) z^k, \\
 G_1^{\alpha,out}(z) &= \sum_k \frac{k}{\langle k^\alpha \rangle} P_\alpha^{out}(k) z^{k-1},
 \end{aligned} \tag{3.94}$$

with  $\alpha = 1, 2$ , then the BP equations can be rewritten in terms of the probabilities  $\{w_i^\alpha\}_{i=1,2,3}$  and  $\{\hat{w}_i^\alpha\}_{i=1,2,3}$ . By means of these generating functions the BP equation read as in the following

$$\begin{aligned}
w_1^\alpha &= G_1^{\alpha,out}(\hat{w}_2^\alpha), \\
w_2^\alpha &= [1 - G_1^{\alpha,out}(1 - \hat{w}_1^\alpha)], \\
w_3^\alpha &= 1 - w_1^\alpha - w_2^\alpha, \\
\hat{w}_3^\alpha &= 1 - \hat{w}_1^\alpha - \hat{w}_2^\alpha, \\
\hat{w}_1^{[1]} &= G_1^{[1],in}(w_2^{[1]}) [1 - G_0^{[2],in}(1 - w_1^{[2]})], \\
\hat{w}_2^{[1]} &= 1 - G_1^{[1],in}(1 - w_1^{[1]}) (1 - G_0^{[2],in}(w_2^{[2]})), \\
\hat{w}_1^{[2]} &= G_1^{[2],in}(w_2^{[2]}) [1 - G_0^{[1],in}(1 - w_1^{[1]})], \\
\hat{w}_2^{[2]} &= 1 - G_1^{[2],in}(1 - w_1^{[2]}) (1 - G_0^{[1],in}(w_2^{[1]}))
\end{aligned} \tag{3.95}$$

It follows that the density of driver nodes can be rearranged as

$$\begin{aligned}
n_D &= \sum_\alpha \{ G_0^{\alpha,out}(\hat{w}_2^\alpha) - [1 - G_0^{\alpha,out}(1 - \hat{w}_1^\alpha)] \} \\
&\quad - \{ [1 - G_0^{[1],in}(1 - w_1^{[1]})][1 - G_0^{[2],in}(w_2^{[2]})] \\
&\quad + [1 - G_0^{[2],in}(1 - w_1^{[2]})][1 - G_0^{[1],in}(w_2^{[1]})] \} \\
&\quad + \sum_\alpha \langle k^\alpha \rangle_{in} [\hat{w}_1^\alpha(1 - w_2^\alpha) + w_1^\alpha(1 - \hat{w}_2^\alpha)].
\end{aligned} \tag{3.96}$$

### 3.6.6 Phase transition in the controllability of Poisson multiplex networks

We consider now the case of two Poisson networks with the same in/out average degree in the two layers. The average in/out degree in the different layers is called

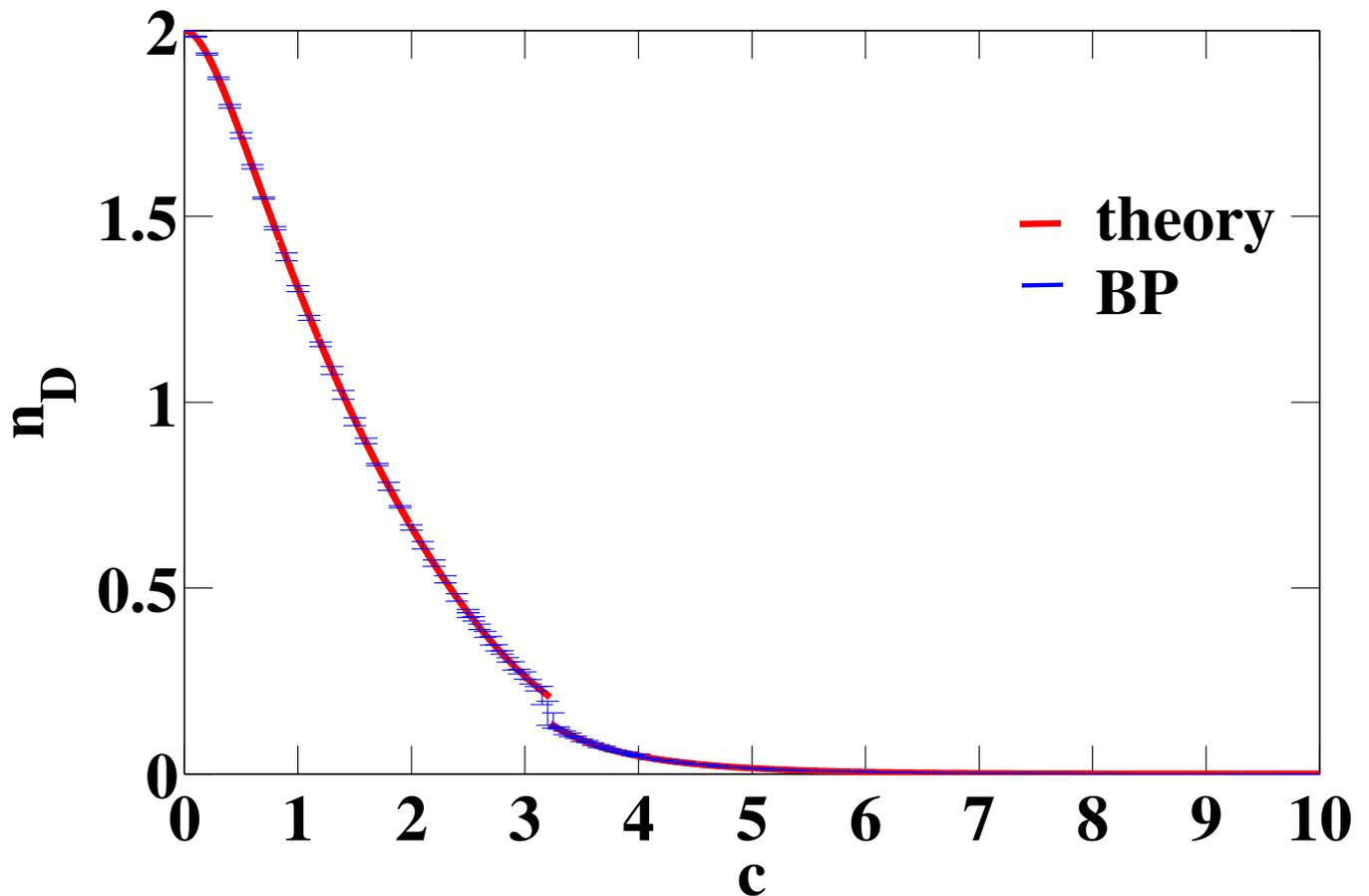


Figure 3.12: Fraction  $n_D = E/N$  of driver nodes in a Poisson duplex network with  $\langle k^{[1]} \rangle_{in} = \langle k^{[1]} \rangle_{out} = \langle k^{[2]} \rangle_{in} = \langle k^{[2]} \rangle_{out} = c$ , plotted as a function of the average degree  $c$ . The red line is the numerical solution of Eqs. 3.99 - 3.100. The points indicate the average BP results obtained over 5 single realisations of the Poisson duplex networks with average degree  $c$  and  $N = 10^4$ . For every point corresponding to the average BP result, the error bar indicates the interval of one standard deviation from the mean.

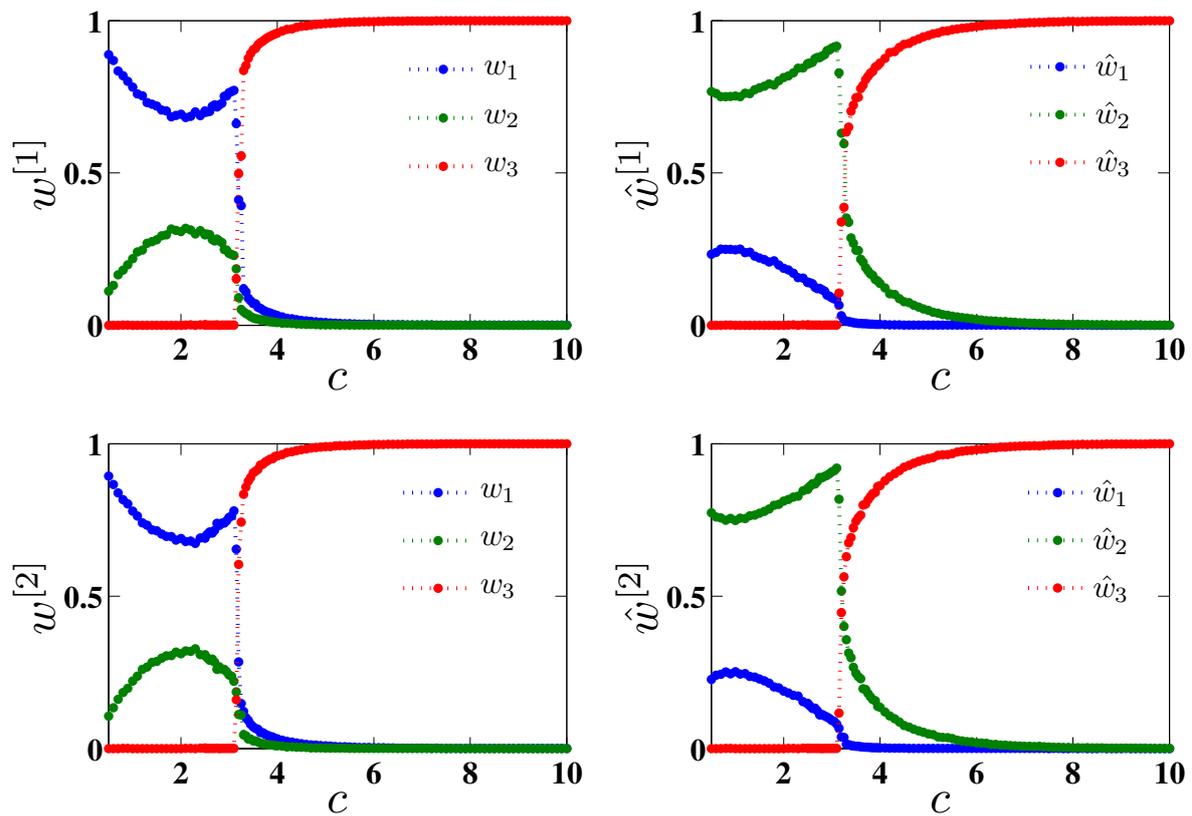


Figure 3.13: Values of the probabilities  $\{w_i\}_{i=1,2,3}$  and  $\{\hat{w}_i\}_{i=1,2,3}$  for a Poisson duplex network with  $\langle k^{[1]} \rangle_{in} = \langle k^{[1]} \rangle_{out} = \langle k^{[2]} \rangle_{in} = \langle k^{[2]} \rangle_{out} = c$ , plotted as a function of the average degree  $c$ . These probabilities are calculated directly from BP results obtained over 5 single realizations of these duplex networks with average degree  $c$  and  $N = 10^4$ .

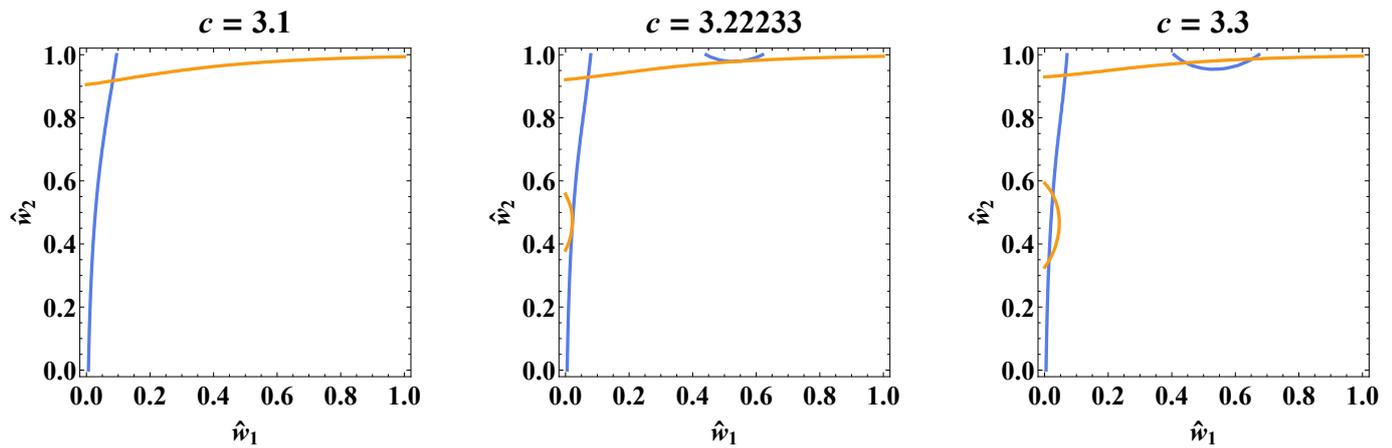


Figure 3.14: Two directed Poisson networks with the same in/out average degree in the two layers equal to  $c$  and the same driver nodes: plotting Eq. 3.99 and Eq. 3.100 we show a phase transition for  $c = 3.22$ .

*c.* In other words, we have  $\langle k^{[1]} \rangle_{in} = \langle k^{[1]} \rangle_{out} = \langle k^{[2]} \rangle_{in} = \langle k^{[2]} \rangle_{out} = c$ . In this case we might assume  $w_i^{[1]} = w_i^{[2]}$  and  $\hat{w}_i^{[1]} = \hat{w}_i^{[2]}$  and then, the BP equations become

$$\begin{aligned}
 w_1 &= e^{-c(1-\hat{w}_2)}, \\
 w_2 &= [1 - e^{-c\hat{w}_1}], \\
 w_3 &= 1 - w_1 - w_2, \\
 \hat{w}_3 &= 1 - \hat{w}_1 - \hat{w}_2, \\
 \hat{w}_1 &= e^{-c(1-w_2)} [1 - e^{-cw_1}], \\
 \hat{w}_2 &= [1 - e^{-cw_1} + e^{-cw_1} e^{-c(1-w_2)}].
 \end{aligned}
 \tag{3.97}$$

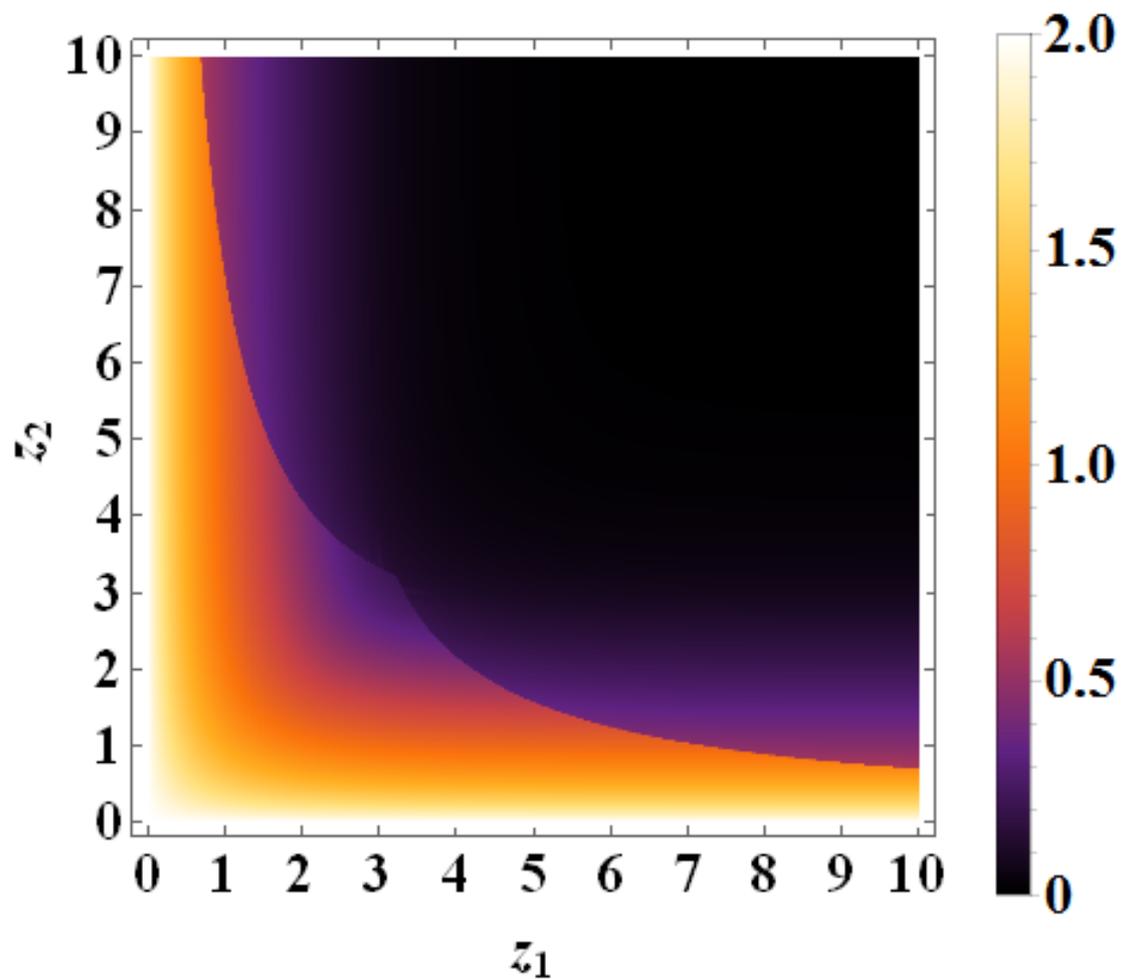


Figure 3.15: Phase diagram of the Poisson duplex network with average degrees  $\langle k^{[1]} \rangle_{in} = \langle k^{[1]} \rangle_{out} = z_1$  and  $\langle k^{[2]} \rangle_{in} = \langle k^{[2]} \rangle_{out} = z_2$ . The colour code indicates the density of driver nodes  $n_D = E/N$  in the duplex network.

The density of driver nodes is given in this case by

$$\begin{aligned}
n_D &= 2 \left[ e^{-c(1-\hat{w}_2)} - 1 + e^{-c\hat{w}_1} \right] \\
&\quad - 2 \left[ 1 - e^{-cw_1} \right] \left[ 1 - e^{-c(1-w_2)} \right] \\
&\quad + 2c \left[ \hat{w}_1(1-w_2) + w_1(1-\hat{w}_2) \right].
\end{aligned} \tag{3.98}$$

We can further simplify the BP equations considering only

$$\hat{w}_1 = e^{-ce^{-c\hat{w}_1}} \left[ 1 - e^{-ce^{-c(1-\hat{w}_2)}} \right], \tag{3.99}$$

$$\hat{w}_2 = \left[ 1 - e^{-ce^{-c(1-\hat{w}_2)}} + e^{-ce^{-c(1-\hat{w}_2)}} e^{-ce^{-c\hat{w}_1}} \right], \tag{3.100}$$

$$0 \leq \hat{w}_1 + \hat{w}_2 \leq 1$$

From the solution of these equations it is possible to find out a phase transition occurring at  $c = c^*$  where the number of driver nodes  $n_D$  of the network has a discontinuity (see Fig. 3.12). The value of the average degree  $c^*$  at which this discontinuity occurs, can be found by imposing that the two curves of the plane  $\hat{w}_1, \hat{w}_2$  given by

$$\hat{w}_1 = h_1(\hat{w}_1, \hat{w}_2) = e^{-ce^{-c\hat{w}_1}} \left[ 1 - e^{-ce^{-c(1-\hat{w}_2)}} \right] \tag{3.101}$$

$$\hat{w}_2 = h_2(\hat{w}_1, \hat{w}_2) = \left[ 1 - e^{-ce^{-c(1-\hat{w}_2)}} + e^{-ce^{-c(1-\hat{w}_2)}} e^{-ce^{-c\hat{w}_1}} \right] \tag{3.102}$$

$$\tag{3.103}$$

for  $c = c^*$  are tangent to each other at their interception. This point is found by

imposing that the Eqs. 3.101 – 3.102 are satisfied together with the equation

$$|J| = 0, \quad (3.104)$$

with  $J$  indicating the Jacobian of the system of equations  $\hat{w}_1 = h_1(\hat{w}_1, \hat{w}_2)$  and  $\hat{w}_2 = h_2(\hat{w}_1, \hat{w}_2)$  given by

$$J = \begin{pmatrix} 1 - \frac{\partial h_1(\hat{w}_1, \hat{w}_2)}{\partial \hat{w}_1} & \frac{\partial h_1(\hat{w}_1, \hat{w}_2)}{\partial \hat{w}_2} \\ \frac{\partial h_2(\hat{w}_1, \hat{w}_2)}{\partial \hat{w}_1} & 1 - \frac{\partial h_2(\hat{w}_1, \hat{w}_2)}{\partial \hat{w}_2} \end{pmatrix}.$$

Imposing that the three Eqs. 3.101 – 3.102 – 3.104 are simultaneously satisfied, the solution  $c^* = 3.222326106$  is found (see Fig. 3.14). For  $c < c^*$  we observe that  $w_3 = \hat{w}_3 = 0$ . For  $c > c^*$  we observe a discontinuity in both  $w_3$  and  $\hat{w}_3$ . Finally for  $c > c^*$  since the functions  $h_1(\hat{w}_1, \hat{w}_2)$  and  $h_2(\hat{w}_1, \hat{w}_2)$  are analytic, we observe a singularity of the type

$$\begin{aligned} w_3 - w_3^* &\propto (c - c^*)^{1/2} \\ \hat{w}_3 - \hat{w}_3^* &\propto -(c - c^*)^{1/2}, \end{aligned} \quad (3.105)$$

showing that this transition has the order parameters  $w_3$  and  $\hat{w}_3$  and that is hybrid (see Fig. 3.13). In fact taking the variations of the Eqs. 3.101 – 3.102 with respect to a change in the value of the average degree  $c$  around the value  $c = c^*$ , and expanding these equations up to the second order, one observes that each of the variations  $\delta\hat{w}_1 = \hat{w}_1 - \hat{w}_1^*$  and  $\delta\hat{w}_2 = \hat{w}_2 - \hat{w}_2^*$  are both proportional to  $\sqrt{\delta c} = (c - c^*)^{1/2}$ . A similar argument applies to  $\delta w_1$  and  $\delta w_2$ . Therefore the maximum matching problem on multilayer networks can display an hybrid transition in the case of a duplex formed by two Poisson networks. This hybrid

phase transition with a square root singularity is in the same universality class of the emergence of the mutually connected component in multiplex networks (Buldyrev et al., 2010; Baxter et al., 2012; Boccaletti et al., 2014). We guess it can indicate that multilayer networks can display an increased fragility to random damage with respect to single layers, and that abrupt discontinuities in the number of driver nodes can result for a small change in the multiplex network topology. Moreover, we considered also duplex networks formed by two Poisson networks with different average degree, i.e. with  $\langle k^{[1]} \rangle_{in} = \langle k^{[1]} \rangle_{out} = z_1$  and  $\langle k^{[2]} \rangle_{in} = \langle k^{[2]} \rangle_{out} = z_2$  we solved numerically the BP equations in this case. The phase diagram is shown in Figure 3.8

### 3.6.7 Dependence on the correlation between low in-degree and out-degree nodes in the different layers

As we previously showed in Sec. 3.5, the number of low in-degree nodes and out-degree nodes is able to modulate the number of driver nodes in single layers. In fact, if the minimum in-degree and out degree are both greater than 2 then the network is fully controllable. These results derive from stability consideration of the BP equations. Also when considering the multiplex controllability, and solving the corresponding BP equations, it is important to make stability considerations in the same way we previously explained for a single network. We evaluated the criterion for  $n_D = 0$  also in this case and we found that the

stability conditions read

$$\begin{aligned}
2 \frac{\langle k^{[1]}(k^{[1]} - 1) \rangle_{in} P_{[1]}^{out}(2)}{\langle k^{[1]} \rangle_{in} \langle k^{[1]} \rangle_{out}} &< 1 \\
2 \frac{\langle k^{[2]}(k^{[2]} - 1) \rangle_{in} P_{[2]}^{out}(2)}{\langle k^{[2]} \rangle_{in} \langle k^{[2]} \rangle_{out}} &< 1.
\end{aligned} \tag{3.106}$$

When  $P_{[1]}^{in}(k) = P_{[1]}^{out}(k) = P_{[2]}^{in}(k) = P_{[2]}^{out}(k) = P(k)$  we have just one stability criterion for this particular solution and it reads

$$P(2) < \frac{\langle k \rangle^2}{2 \langle k(k-1) \rangle}. \tag{3.107}$$

recovering the result for a single network with  $P^{in}(k) = P^{out}(k)$  (for further details about the calculations see Sec. 4.4.1).

We made a comparison between the results of Sec. 3.5 regarding the phase diagram presented in Fig. 3.8 for power law networks with minimum degree 2, and the same situation for a duplex network ( $P_{[1]}^{in}(k) = P_{[1]}^{out}(k) = P_{[2]}^{in}(k) = P_{[2]}^{out}(k) = P(k)$ ). In Fig. 3.16 on the left, we present the results for a duplex network and on the right, we display the results for two single networks, controlled separately. As expected, the stability criterion works in the same way for the two systems but the number of driver nodes needed to control a duplex network is higher. Controlling a duplex network is then definitely more demanding. In order to investigate further this situation and the role of low degree nodes in determining the controllability of a network we consider now *correlated* duplex networks in which nodes of low degree in one layer are also likely to be nodes of low degree in the other layer. As showed, when such degree correlations are absent, the number of driver nodes of a duplex is always higher then the sum of driver nodes

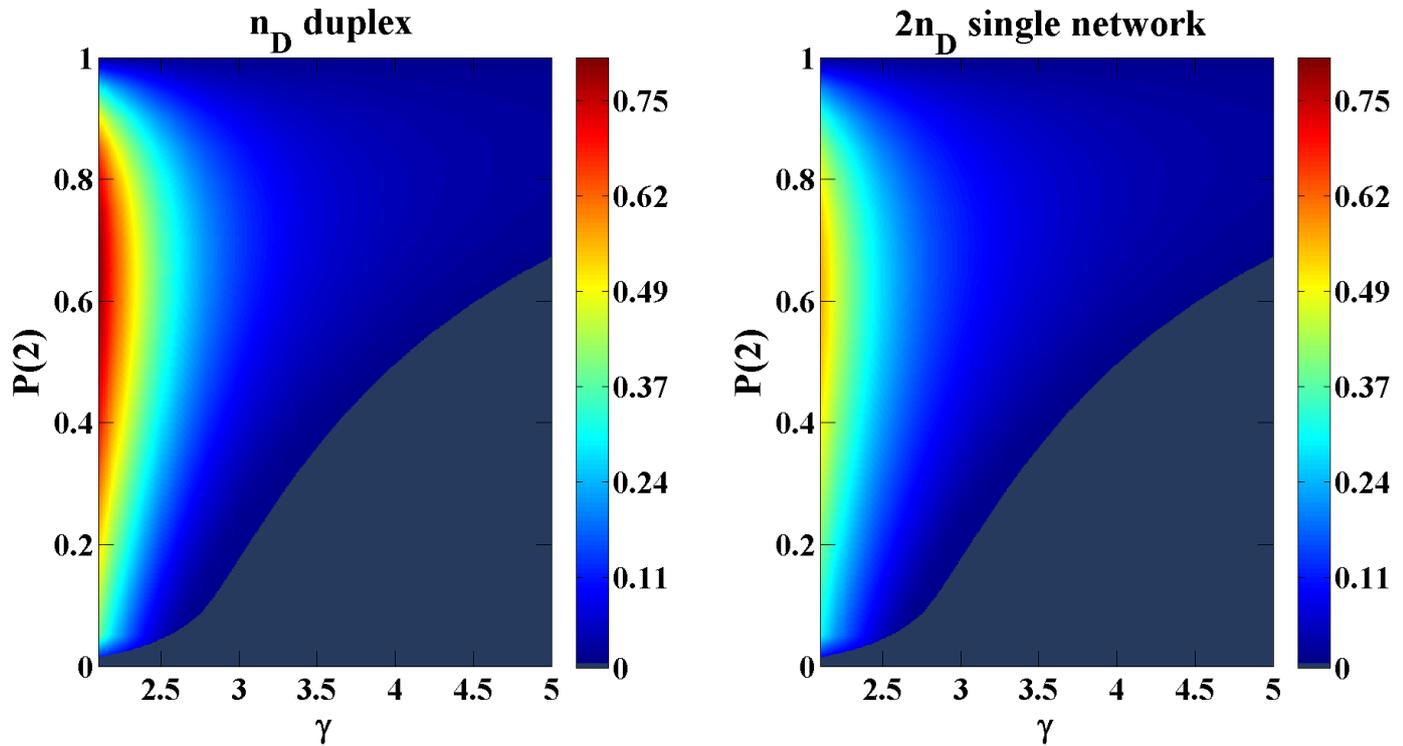


Figure 3.16: On the left: phase diagram of the density of driver nodes  $n_D$  as a function of the parameters  $\gamma$  and  $P(2)$  for duplex networks of  $N = 10^6$  nodes with degree distribution given by Eq. 3.71 and  $P(1) = 0$ . On the right: phase diagram of the density of driver nodes  $n_D$  as a function of the parameters  $\gamma$  and  $P(2)$  for two separate networks of  $N = 10^6$  nodes with degree distribution given by Eq. 3.71 and  $P(1) = 0$ .

in the two layers. But how does this difference change if the low degree nodes of one layer are also the low degree nodes of the other layer? Looking back to Eqs. 3.92, 3.93 we introduce correlations modulating  $P^{in}(k^{[1]}, k^{[2]})$ . For low-degree correlations we define  $P^{in}(k^{[1]}, k^{[2]})$  as

$$P^{in}(k^{[1]}, k^{[2]}) = \begin{cases} p\delta_{k^{[2]}, k^{[1]}}P(k^{[1]}) + (1-p)P(k^{[1]})P(k^{[2]}), & \text{if } k^{[1]} \leq 2 \\ (1-p)P(k^{[1]})P(k^{[2]}), & \text{if } k^{[1]} > 2 \quad k^{[2]} \leq 2 \\ p\frac{P(k^{[2]})}{C}P(k^{[1]}) + (1-p)P(k^{[1]})P(k^{[2]}), & \text{if } k^{[1]} > 2 \quad k^{[2]} > 2 \end{cases}$$

where  $C = 1 - \sum_{k \leq 2} P(k)$  and we consider in particular  $P(k^{[1]}) = P(k^{[2]}) = P(k)$ . The probability  $p$  modulates the balance between the correlated scenario and the classical uncorrelated situation. We studied also a *total-degree-correlation*, namely, a degree correlation over the total range of the degree distributions in the two layers. This total degree correlation is defined as

$$P^{in}(k^{[1]}, k^{[2]}) = p\delta_{k^{[2]}, k^{[1]}}P(k^{[1]}) + (1-p)P(k^{[1]})P(k^{[2]})$$

where the probability  $p$  plays the same role and again,  $P(k^{[1]}) = P(k^{[2]}) = P(k)$ . For further computations regarding the density of driver nodes for these correlated duplex networks see Sec. 4.4.2.

We consider at first the analytical solutions for a duplex formed by two Poisson networks with the same average degree  $c$ . Varying the low-degree correlation thanks to  $p$  we change the profile of the density of driver nodes. In Fig. 3.17 with  $p = 0$  we recover the previous trend for uncorrelated duplex networks. The more we increase  $p$  the lower the density of driver nodes becomes and the smaller

the discontinuity gap appears. Anyway, we never reach the profile given by two separated poisson networks (the black dashed line in comparison with the grey line for  $p = 1$ ).

Moreover, we validated the scenario of low-degree correlations and total-degree correlations thank to BP simulations. In Fig. 3.18 and Fig. 3.19, we consider, respectively, directed poisson networks with increasing average degree  $c$  ( $N = 10^4$ ) and power law networks with  $\gamma = 2.1, \dots, 3$  (directed power law networks with  $k_{min}^{in} = k_{min}^{out} = 1$  and structural cutoff  $k_{max}^{in} = k_{max}^{out} = \sqrt{N \langle k \rangle_{in/out}}$  with  $N = 10^4$ ). Furthermore, we consider the total-degree correlation for  $p = 1$ , i.e. the situation in which the replica nodes in the two layers have exactly the same degree. In addition to a good agreement between analytical results and BP simulations we found that no significant improvement on the number of driver nodes can be made by correlating also the nodes with in degree greater than 2. This result highlights once more the importance of low degree nodes also for the controllability of multiplex networks.

## 3.7 Conclusion

In this chapter we gave a general introduction to control theory and we showed how network theory plays a major role in the so-called *structural controllability*. Moreover we presented the cavity method in some of its different versions. Using these tools we have shown how the structural controllability of a network depends strongly on the fraction of low in-degree and low out-degree nodes. For any uncorrelated directed network with given in-degree and out-degree distribution, the minimum fraction of driver nodes is zero, i.e.  $n_D = 0$ , if the in-degrees and

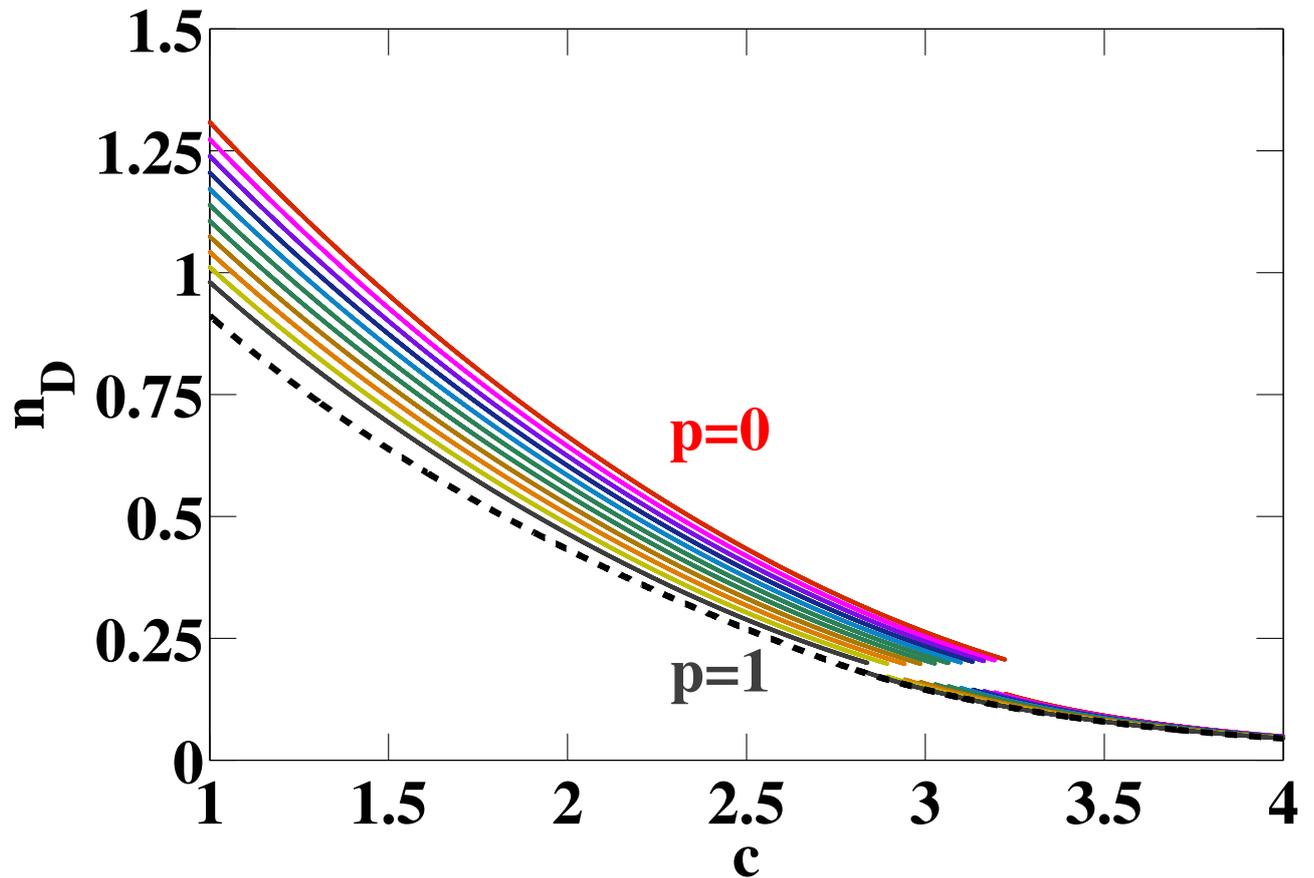


Figure 3.17: Low-degree correlation for a duplex network composed by two Poisson networks with the same average degree  $c$ : we display the density of the driver nodes  $n_D$  as function of  $c$  and  $p$ . The result for two separated Poisson networks is shown in black (dashed curve) while the situation of uncorrelated layers is shown in red.

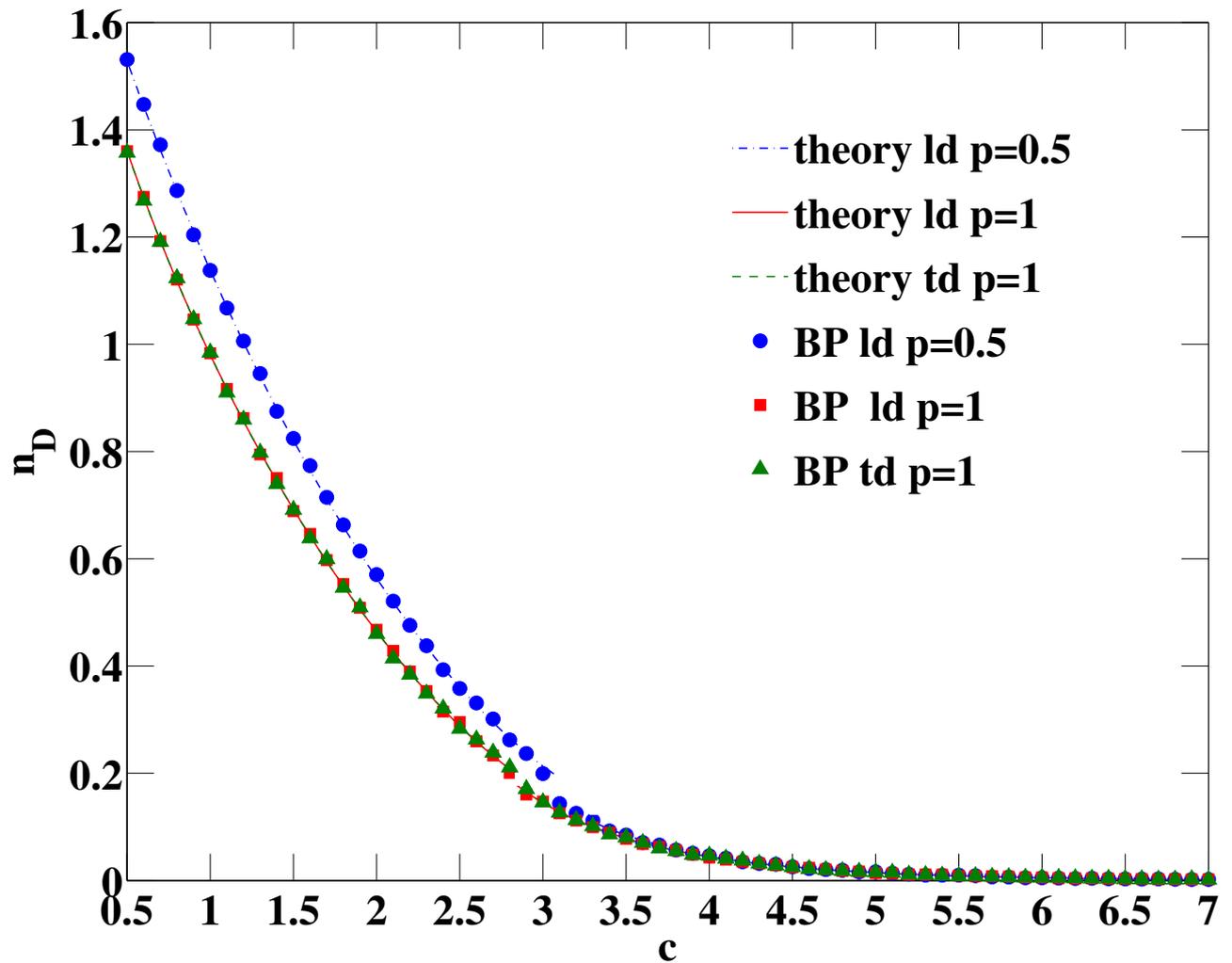


Figure 3.18: Poisson duplex networks: BP results and comparison with the analytical results for low-degree (ld) and total-degree (td) correlations. We considered networks with  $N = 10^4$  and each point is the average over 5 BP simulations.

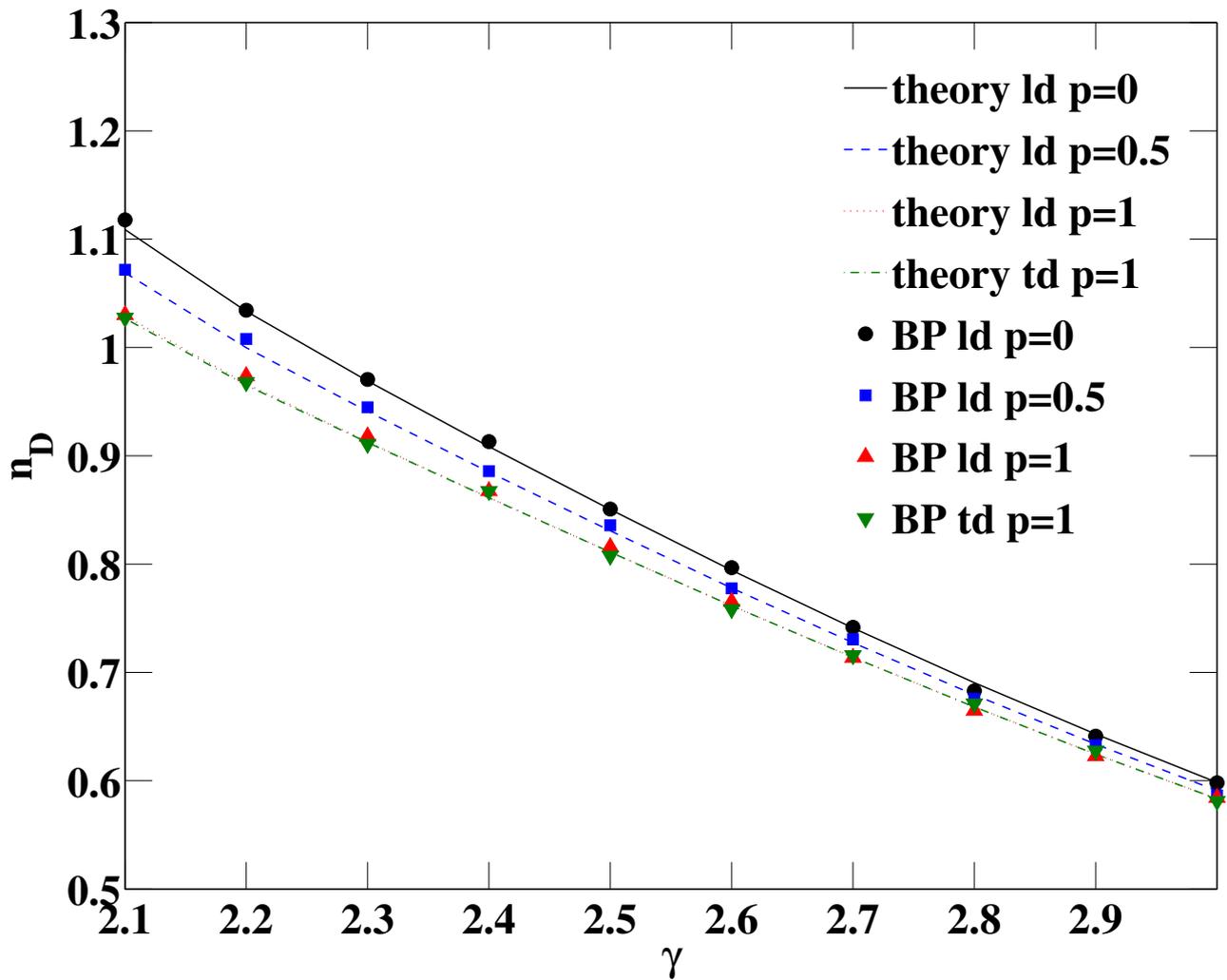


Figure 3.19: Powerlaw duplex networks with structural cutoff: BP results and comparison with the analytical results for low-degree (ld) and total-degree (td) correlations. We considered networks with  $N = 10^4$  and each point is the average over 20 BP simulations.

the out-degrees of all nodes are both greater than 2. For the relevant class of networks with power-law degree distribution, the number of driver nodes can change dramatically by changing the fraction of nodes with in-degree and out-degree equal to 1 or 2. Moreover, we extended the formalism of control theory to multiplex networks, introducing a novel approach: we have characterised the controllability of multiplex networks in which the driver nodes are forced to be the same in each layer. As expected, the multiplex network controllability is more demanding than the controllability of single layers and it is possible to observe discontinuity in the number of driver nodes as a function of the multiplex network topology, as in the case a a multiplex network formed by two Poisson networks. Finally, the introduction of structural correlations in the multiplex networks affecting the low-degrees can reduce the number of driver nodes requested.

# 4 Supplementary Information

## 4.1 APS dataset

We gather here the major details and statistics regarding the APS dataset and the related multiplex networks. For further images picturing the degree and multidegree distributions, and moreover, the relations between single-layer observables, we refer to Menichetti et al. (2014a) and its supplementary.

### 4.1.1 The two datasets

We have considered the American Physical Society (APS) research data that is organised into two main datasets:

- *Article metadata*: for each article the metadata includes DOI, journal, volume, issue, first page and last page, article id and number of pages, title, authors, affiliations, publication history, PACS codes, table of contents, heading, article type, and copyright information.
- *Citing article pairs*: this dataset consists of pairs of APS articles that cite each other. Each pair is represented by a pair of DOIs. The first id cites the second id.

In the APS metadata an author is usually identified by given name, middle name, and surname. In different articles, the same author can appear with his/her full name or with his/her initials. To deal with this issue, we decided to identify a specific author with the initials of his/her given name and middle name and with his/her full surname. We restricted our analysis to the article metadata and citing article pairs that relate only to PRL and PRE. The total number of PRL articles is 95,516 and the total number of PRL authors is 117,412. The total number of PRE articles is 35,944 and the total number of PRE authors is 36,171. The number of authors that published both in PRE and PRL is equal to 17,470.

Among the papers published in PRE and PRL, we focused our study only on those containing a number of authors  $n_p \leq 10$ . This excludes most of the experimental high-energy collaborations that are typically characterised by a number of authors of a different order of magnitude. We decided to place such a cut-off to the maximum number of authors allowed per paper to avoid biases due to very large publications. Given the cut-off, our study thus becomes limited to 35,766 PRE articles (99.5 %) and 35,205 PRE authors (97.3 %) on the one hand, and 89,245 PRL articles (93.4 %) and 92,436 PRL authors (78.7 %) on the other. The intersection of these two datasets includes 16,207 authors (i.e., 92.8 % of the previous intersection).

We analysed two types of interaction between APS authors: scientific collaborations and citations, with weights defined as follows.

- *Collaborations*: two authors are connected if they co-authored at least one paper. The collaborative interaction between author  $i$  and author  $j$  is defined as in Newman (2001), Barrat et al. (2004), i.e., the undirected adjacency matrix element  $a_{ij}$  is given by

$$a_{ij} = \sum_{p \in I} \frac{\delta_i^p \delta_j^p}{n_p - 1} \quad i \neq j \quad (4.1)$$

$$a_{ii} = 0, \quad (4.2)$$

where the index  $p$  indicates an article in the dataset  $I$ ,  $n_p$  indicates the number of authors of article  $p$  and  $\delta_i^p = 1$  if node  $i$  is an author of article  $p$ , and  $\delta_i^p = 0$  otherwise. The resulting network is undirected and without self-loops.

- *Citations*: two authors are connected by a directed link if one author cites the other one. In this case, the element  $a_{ij}$  of the directed adjacency matrix indicating how many times node  $i$  cites node  $j$  is given by

$$a_{ij} = \sum_{p, p' \in I} \delta_i^p \delta_j^{p'} b_{p, p'}, \quad (4.3)$$

where  $b_{p, p'} = 1$  if article  $p$  cites article  $p'$ , and  $b_{p, p'} = 0$  otherwise. Moreover  $\delta_i^p$  is defined as above and indicates whether  $i$  is author of article  $p$  ( $\delta_i^p = 1$ ) or not ( $\delta_i^p = 0$ ). The resulting network is directed and with self-loops.

We constructed the following two duplex networks:

1. **CoCo-PRL/PRE**: *collaborations among PRL and PRE authors*. The nodes of this multiplex network are the authors who published articles both in PRL and PRE (i.e., 16,207 authors). These nodes are connected

in layer 1 through weighted undirected links indicating the strength of their collaboration in PRL (i.e., co-authorship of PRL articles). The same nodes are connected in layer 2 through weighted undirected links indicating the strength of their collaborations in PRE (i.e., co-authorship of PRE articles).

2. **CoCi-PRE:** *collaborations among PRE authors and citations to PRE articles.* The nodes of this multiplex network are the authors of articles published in PRE (i.e., 35,205 authors). These nodes are connected in layer 1 through weighted undirected links indicating the strength of their collaboration in PRE (i.e., co-authorship of PRE articles). The same nodes are connected in layer 2 through weighted directed links indicating how many times an author (with articles in PRE) cited another author’s work, where citations are limited to those made to PRE articles.

### 4.1.2 Total overlap and total weighted overlap of the multiplex networks

In order to characterise the overlap existing between the links of the multiplex networks, we consider the total overlap  $O^{\alpha,\alpha'}$  between layer  $\alpha$  and layer  $\alpha'$  and its related total weighted overlap  $O^{(w)\alpha,\alpha'}$  defined, respectively in Eqs. 2.10, 2.11. Table 4.1 reports details on the total overlap and total weighted overlap, and indeed shows that our multiplex networks are characterized by a significant overlap of links.

Table 4.1: Total overlap and total weighted overlap in the CoCo-PRL/PRE and CoCi-PRE multiplex networks.

Dataset	Layer	Total overlap %	Total weighted overlap %
CoCo-PRL/PRE	PRL	35.75	28.35
CoCo-PRL/PRE	PRE	39.10	33.84
CoCi-PRE	coll	39.51	14.24
CoCi-PRE	cit	12.64	20.76

Table 4.2: Kendall  $\tau$  coefficient measuring the correlations between the degrees in the different layers and the strengths in the different layers in the CoCi-PRE multiplex network.

$\tau$	$k_1$	$k_2^{in}$	$k_2^{out}$	$\tau$	$s_1$	$s_2^{in}$	$s_2^{out}$
$k_1$	1	0.36	0.37	$s_1$	1	0.53	0.53
$k_2^{in}$	0.36	1	0.38	$s_2^{in}$	0.53	1	0.41
$k_2^{out}$	0.37	0.38	1	$s_2^{out}$	0.53	0.41	1

Table 4.3: Kendall  $\tau$  coefficient measuring the correlations between the degrees in the different layers and the strengths in the different layers in the CoCo-PRL/PRE multiplex network.

$\tau$	$k_1$	$k_2$	$\tau$	$s_1$	$s_2$
$k_1$	1	0.44	$s_1$	1	0.37
$k_2$	0.44	1	$s_2$	0.37	1

### 4.1.3 Degree and multidegree distribution of the two multiplex networks

The nodes  $i = 1, 2, \dots, N$  of the multiplex networks have degrees  $k_i^1$  in layer 1 and  $k_i^2$  in layer 2. Moreover, we can define the multidegree  $k_i^{\vec{m}}$  of a generic node  $i$  as the sum of the multilinks  $\vec{m}$  incident on it. We observe that, since we always have

$$k_i^{\vec{0}} = (N - 1) - \sum_{\vec{m} \neq \vec{0}} k_i^{\vec{m}}, \quad (4.4)$$

we can therefore restrict the analysis to multidegrees  $\vec{m} \neq \vec{0}$ . The degree and multidegree for both the CoCo-PRL/PRE and the CoCi-PRE multiplex networks are characterised by broad distributions. Moreover, in both duplex networks, the degrees each author has in the two layers are positively correlated, as indicated by the Kendall correlation coefficient between degrees (See Tables 4.2, 4.3 ). Finally, also multidegrees in the multiplex networks are correlated, as indicated by their Kendall coefficients (See Tables 4.4, 4.5). This correlation coefficient are calculated both for degree and multidegrees. In what follows, we give the definition in the case of two degree sequences. The extension to multidegree sequences is straightforward.

The Kendall's  $\tau$  correlation coefficient between the degree sequences  $\{k_i^\alpha\}$  and  $\{k_i^\beta\}$  in the two layers  $\alpha$  and  $\beta$  is a measure that takes into account the sequence of ranks  $\{x_i^\alpha\}$  and  $\{x_i^\beta\}$ . A pair of nodes  $i$  and  $j$  are concordant if their ranks have the same order in the two sequences, i.e.,  $(x_i^\alpha - x_j^\alpha)(x_i^\beta - x_j^\beta) > 0$ ; otherwise, they are discordant. The Kendall's  $\tau$  is defined in terms of the number  $n_c$  of

concordant pairs and the number  $n_d$  of discordant pairs, and is given by

$$\tau = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}}, \quad (4.5)$$

where  $n_0 = 1/2N(N - 1)$  and the terms  $n_1$  and  $n_2$  account for the degeneracy of the ranks and are given by

$$\begin{aligned} n_1 &= \frac{1}{2} \sum_n u_n(u_n - 1), \\ n_2 &= \frac{1}{2} \sum_n v_n(v_n - 1), \end{aligned} \quad (4.6)$$

where we call  $u_n$  the number of nodes in the  $n$ th tied group of the degree sequence  $\{k^\alpha\}$ , and we call  $v_n$  the number of nodes in the  $n$ th tied group of the degree sequence  $\{k^\beta\}$ .

Table 4.4: Kendall's  $\tau$  coefficient measuring the correlations between multidegrees in the CoCi-PRE multiplex network.

$\tau$	$k_{11}^{in}$	$k_{11}^{out}$	$k_{10}^{in}$	$k_{10}^{out}$	$k_{01}^{in}$	$k_{01}^{out}$
$k_{11}^{in}$	1	0.63	-0.06	0.13	0.54	0.39
$k_{11}^{out}$	0.63	1	0.12	-0.02	0.39	0.49
$k_{10}^{in}$	-0.06	0.12	1	0.68	0.03	0.12
$k_{10}^{out}$	0.13	-0.02	0.68	1	0.13	0.09
$k_{01}^{in}$	0.54	0.39	0.03	0.13	1	0.34
$k_{01}^{out}$	0.39	0.49	0.12	0.09	0.34	1

Table 4.5: Kendall’s  $\tau$  coefficient measuring the correlations between multidegrees in the CoCo-PRL/PRE multiplex network.

$\tau$	$k_{11}$	$k_{10}$	$k_{01}$
$k_{11}$	1	0.13	0.13
$k_{10}$	0.13	1	0.23
$k_{01}$	0.13	0.23	1

#### 4.1.4 Weighted network properties of single layers

Here we report the weighted network properties of the single layers of our multiplex networks. In general, the average strength  $s_k^\alpha$  of nodes with degree  $k$  in layer  $\alpha$  and the average inverse participation ratio  $Y_k^\alpha$  of nodes with degree  $k$  in layer  $\alpha$  are described by the functional behaviour

$$\begin{aligned} s_k^\alpha &\propto k^{\beta_\alpha}, \\ Y_k^\alpha &\propto \frac{1}{k^{\lambda_\alpha}}. \end{aligned} \quad (4.7)$$

We have considered both the CoCo-PRL/PRE dataset and the CoCi-PRE dataset and fitted  $s_k^\alpha$  and  $Y_k^\alpha$  according to this expected power-law behaviour.

The exponents shown in Table 4.6 and Table 4.7 have been computed with the method “regression”, function of Matlab MathWorks (2015). This function performs a multiple linear regression, and for each coefficient gives the 95% confidence interval. In the tables, we show also the coefficient of determination  $R^2$  indicating how well the power-law trend fits the data.

As shown by Table 4.6, the CoCo-PRL/PRE multiplex network is charac-

Table 4.6: CoCo-PRL/PRE multiplex network: power-law exponents  $\lambda_\alpha$  and  $\beta_\alpha$  determining the functional behaviour for the average strength  $s_k^\alpha$  of nodes with degree  $k$  and the average inverse participation ratio  $Y_k^\alpha$ , with  $\alpha$  corresponding to the collaboration layer in PRL or in PRE.

layer	$\alpha$	$\beta_\alpha$	$R^2$	$\lambda_\alpha$	$R^2$
PRL	1	$0.96 \pm 0.04$	0.96	$0.84 \pm 0.03$	0.97
PRE	2	$1.01 \pm 0.05$	0.96	$0.80 \pm 0.05$	0.94

Table 4.7: CoCi-PRE multiplex network: power-law exponents  $\lambda_\alpha$  and  $\beta_\alpha$  determining the functional behaviour for the average strength  $s_k^\alpha$  of nodes with degree  $k$  and the average inverse participation ratio  $Y_k^\alpha$ , with  $\alpha$  corresponding to the collaboration layer or to the citation layer, both for PRE. For the citation layer, we consider separately the in-behaviour and the out-behaviour.

layer	$\alpha$	$\beta_\alpha$	$R^2$	$\lambda_\alpha$	$R^2$
Co	1	$1.03 \pm 0.04$	0.96	$0.79 \pm 0.04$	0.94
Ci <sub>in</sub>	2,in	$1.13 \pm 0.02$	0.98	$0.72 \pm 0.03$	0.85
Ci <sub>out</sub>	2,out	$1.14 \pm 0.03$	0.97	$0.70 \pm 0.04$	0.83

terized by a linear behavior of average strength as a function of the degree of nodes. Table 4.7 shows that the CoCi-PRE multiplex network is characterized by a linear behavior of average strength as a function of the degree of nodes in the collaboration network, and by a super-linear behavior in the citation network.

### 4.1.5 Statistical analysis of the properties of multilinks in the CoCo-PRL/PRE multiplex network

In this subsection, we discuss in detail the results of our statistical analysis of the properties of multilinks in the CoCo-PRL/PRE multiplex network. In particular, we focus on the average multistrength of nodes with a given multidegree, i.e.,  $s^{\vec{m},\alpha}(k^{\vec{m}}) = \langle s_i^{\vec{m},\alpha} \delta(k_i^{\vec{m}}, k^{\vec{m}}) \rangle$ , and the average inverse multiparticipation ratio of nodes with a given multidegree, i.e.,  $Y^{\vec{m},\alpha}(k^{\vec{m}}) = \langle Y_i^{\vec{m},\alpha} \delta(k_i^{\vec{m}}, k^{\vec{m}}) \rangle$ . These quantities are expected to scale as

$$\begin{aligned} s^{\vec{m},\alpha}(k^{\vec{m}}) &= e^{q^{\vec{m},\alpha}} (k^{\vec{m}})^{\beta_{\vec{m},\alpha}} \\ Y^{\vec{m},\alpha}(k^{\vec{m}}) &= e^{p^{\vec{m},\alpha}} \frac{1}{(k^{\vec{m}})^{\lambda_{\vec{m},\alpha}}}, \end{aligned} \quad (4.8)$$

with exponents  $\beta_{\vec{m},\alpha} \geq 1$  and  $\lambda_{\vec{m},\alpha} \leq 1$ . We have computed these exponents with the method “regression”, function of Matlab MathWorks (2015) This function performs a multiple linear regression, and for each coefficient gives the 95% confidence interval. We have also computed the coefficient of determination  $R^2$  indicating how well the power-law trend fits the data. For a complete list of the exponents characterizing multistrength and the inverse multiparticipation ratio, see Table 4.8. In what follows, we will label the *PRL* collaboration layer as  $\alpha = 1$  and the *PRE* collaboration layer as  $\alpha = 2$ .

## Statistical analysis of the average multistrengths in the CoCo-PRL/PRE multiplex network

In the CoCo-PRL/PRE multiplex network, the fitted exponents  $\beta_{\vec{m},1}$  for multilinks  $\vec{m} = (1, 1)$  and  $\vec{m} = (1, 0)$  and the fitted proportionality constants in Eq. 4.8 are not significantly different. However, we can perform a paired samples Student's t-test to show how the average multistrength per fixed multidegree  $s^{\vec{m},\alpha}(k^{\vec{m}})$  is significantly higher for multilinks  $(1, 1)$  than multilinks  $(1, 0)$ . We have identified pairs of average multistrength  $s^{(1,1),1}(k^{(1,1)})$  and  $s_k^{(1,0),1}(k^{(1,0)})$ , corresponding to the same multidegree value  $k^{(1,1)} = k^{(1,0)} = k$ . The paired samples Student's t-test returns a test decision for the null hypothesis that the values  $\log(s^{(1,1),1}(k)/s^{(1,0),1}(k))$  come from a normal distribution with mean zero and variance from the data. In our case, the null hypothesis is rejected with a p-value equal to  $2.90 \cdot 10^{-16}$ . Furthermore,  $\langle \log(s^{(1,1),1}(k)/s^{(1,0),1}(k)) \rangle$  is equal to 0.53. This analysis suggests that for a particular value of degree  $k$  the related  $s^{(1,1),1}(k)$  is higher than  $s^{(1,0),1}(k)$ , and the two average multistrengths satisfy the relation  $s^{(1,1),1}(k) \approx e^{0.53} s^{(1,0),1}(k)$ . Similar results were obtained in the case of the multistrengths on the second layer indicating the collaboration network on *PRE* articles. The null hypothesis is rejected with a p-value equal to  $8.98 \cdot 10^{-15}$ , and  $\langle \log(s^{(1,1),2}(k)/s^{(0,1),2}(k)) \rangle$  is equal to 0.57.

## Statistical analysis of the average inverse multiparticipation ratio in the CoCo-PRL/PRE multiplex network

In the *PRL* layer the fitted exponents  $\lambda_{\vec{m},\alpha}$  are significantly different. The weights regarding multilinks  $(1, 1)$  are distributed more heterogeneously than the weights regarding multilinks  $(1, 0)$ . A similar situation is found also in the

*PRE* layer. The paired Student's t-test is also useful to understand the properties of the average inverse multiparticipation ratio. In addition to the fitted exponents, we can perform a t-test as we did previously considering now  $Y^{\vec{m},\alpha}(k^{\vec{m}})$ . This test underlines how the inverse multiparticipation ratios regarding multilinks (1,1) are significantly higher than those regarding multilinks (1,0) or (0,1). In the case  $Y^{(1,1),1}(k)$  vs  $Y^{(1,0),1}(k)$ , the t-test gives a p-value equal to 0.002 and an average value  $\langle \log(Y^{(1,1),1}(k)/Y^{(1,0),1}(k)) \rangle = 0.11$ . In the case  $Y^{(1,1),2}$  vs  $Y^{(0,1),2}(k)$ , the p-value is equal to  $6.64 \cdot 10^{-6}$ , and the average value is  $\langle \log(Y^{(1,1),2}(k)/Y^{(0,1),2}(k)) \rangle = 0.19$ .

Table 4.8: CoCo-PRL/PRE multiplex network: power-law exponents  $\lambda_{\vec{m}}$  and  $\beta_{\vec{m}}$  and parameters  $p_{\vec{m}}$ ,  $q_{\vec{m}}$  determining the functional behavior for average multistrength of nodes with a given multidegree,  $s^{\vec{m},\alpha}(k^{\vec{m}})$ , and for average inverse multiparticipation ratio of nodes with a given multidegree,  $Y^{\vec{m},\alpha}(k^{\vec{m}})$ , with  $\alpha$  corresponding to the collaboration layer in PRL (1) or in PRE (2). The value of the determination coefficient  $R^2$  for the power-law fits is also reported.

$\vec{m}, \alpha$	$\beta_{\vec{m}}$	$q_{\vec{m}}$	$R^2$	$\lambda_{\vec{m}}$	$p_{\vec{m}}$	$R^2$
(1,1), 1	$1.06 \pm 0.09$	$-0.51 \pm 0.26$	0.94	$0.74 \pm 0.05$	$-0.09 \pm 0.16$	0.95
(1,0), 1	$0.97 \pm 0.03$	$-0.78 \pm 0.10$	0.99	$0.84 \pm 0.03$	$0.06 \pm 0.12$	0.97
(1,1), 2	$1.09 \pm 0.10$	$-0.40 \pm 0.29$	0.93	$0.73 \pm 0.06$	$-0.10 \pm 0.19$	0.93
(0,1), 2	$1.00 \pm 0.04$	$-0.71 \pm 0.14$	0.98	$0.84 \pm 0.05$	$0.04 \pm 0.16$	0.95

### 4.1.6 Statistical analysis of the properties of multilinks in the CoCi-PRE multiplex network

We analyzed the average multistrength of nodes with a given multidegree, i.e.,  $s^{\vec{m},\alpha,(in/out)}(k^{\vec{m}(in/out)}) = \left\langle s_i^{\vec{m},\alpha,(in/out)} \delta(k_i^{\vec{m},(in/out)}, k^{\vec{m},(in/out)}) \right\rangle$ , and the average inverse multiparticipation ratio of nodes with a given multidegree, i.e.,  $Y^{\vec{m},\alpha,(in/out)}(k^{\vec{m},(in/out)}) = \left\langle Y_i^{\vec{m},\alpha,(in/out)} \delta(k_i^{\vec{m},(in/out)}, k^{\vec{m},(in/out)}) \right\rangle$ , where a distinction was made between incoming and outgoing links in the citation layer. These quantities are expected to scale as

$$\begin{aligned}
 s^{\vec{m},1,(in,out)}(k^{\vec{m},(in,out)}) &= e^{q^{\vec{m},(in,out),1}} (k^{\vec{m},(in/out)})^{\beta_{\vec{m},1,(in/out)}} \\
 s^{\vec{m},2,(in/out)}(k^{\vec{m},(in/out)}) &= e^{q^{\vec{m},2,(in/out)}} (k^{\vec{m},(in/out)})^{\beta_{\vec{m},2,(in/out)}} \\
 Y^{\vec{m},1,(in/out)}(k^{\vec{m},(in/out)}) &= e^{p^{\vec{m},1,(in/out)}} \frac{1}{(k^{\vec{m},(in/out)})^{\lambda_{\vec{m},1,(in/out)}}} \\
 Y^{\vec{m},2,(in/out)}(k^{\vec{m},(in/out)}) &= e^{p^{\vec{m},2,(in/out)}} \frac{1}{(k^{\vec{m},(in/out)})^{\lambda_{\vec{m},2,(in/out)}}}, \quad (4.9)
 \end{aligned}$$

with exponents  $\beta_{\vec{m},\alpha,(in/out)} \geq 1$  and  $\lambda_{\vec{m},\alpha,(in/out)} \leq 1$ . We have computed these exponents with the method “regression”, function of Matlab MathWorks (2015). This function performs a multiple linear regression, and for each coefficient gives the 95% confidence interval. We have also computed the coefficient of determination  $R^2$  indicating how well the power-law trend fits the data. The complete list of the exponents and the multiplication constants characterising multistrength and the inverse multiparticipation ratio can be found in Table 4.9 together with the corresponding values of  $R^2$ . In what follows we will label the *PRE* collaboration layer as  $\alpha = 1$  and the *PRE* citation layer as  $\alpha = 2$ .

## The statistical analysis of the average multistrengths in the CoCi-PRE multiplex network

In the CoCi-PRE multiplex network, we can perform, at first, a statistical analysis of the multistrengths in the collaboration layer. The fitted exponents  $\beta_{(1,1),1,in}$ ,  $\beta_{(1,1),1,out}$ ,  $\beta_{(1,0),1,in}$  and  $\beta_{(1,0),1,out}$  are not significantly different. Conversely, the fitted intercepts of the log-log plot, regarding multilinks  $(1, 1)$ ,  $(in/out)$  are significantly different from the intercept for multilinks  $(1, 0)$ ,  $(in/out)$ .

From a paired samples Student's t-test, in the same way as we did for the average multistrengths in the CoCo-PRL/PRE multiplex network, we obtained that both  $s^{(1,1),1,in}(k)$  and  $s^{(1,1),1,out}(k)$  are significantly higher than  $s^{(1,0),1,in}(k)$  and  $s^{(1,0),1,out}(k)$ . In the case  $s^{(1,1),1,in}(k)$  vs  $s^{(1,0),1,in}(k)$ , we have a p-value equal to  $1.91 \cdot 10^{-35}$  and an average value  $\langle \log (s^{(1,1),1,in}(k)/s^{(1,0),1,in}(k)) \rangle = 0.78$ . In the case  $s^{(1,1),1,out}(k)$  vs  $s^{(1,0),1,out}(k)$ , we have a p-value equal to  $9.93 \cdot 10^{-30}$  and an average value  $\langle \log (s^{(1,1),1,out}(k)/s^{(1,0),1,out}(k)) \rangle = 0.80$ .

Based on the fitted parameters and the Student's t-test, the data suggest that both multidegrees for multilinks  $(1, 1)$  and multilinks  $(1, 0)$  have a linear relation with their own multistrengths in the collaboration layer, and that multistrengths  $(1,1)$  are related to multistrengths  $(1,0)$  by a multiplicative constant. In the citation layer, the fitted exponents  $\beta_{\vec{m},in/out}$  indicate a super-linear scaling, and are significantly different (see Table 4.9).

## The statistical analysis of the inverse multiparticipation ratio in the CoCi-PRE multiplex network

For the collaboration layer, comparing  $\lambda_{(1,1),1,in}$  with  $\lambda_{(1,0),1,in}$ , and  $\lambda_{(1,1),1,out}$  with  $\lambda_{(1,0),1,out}$ , the confidence intervals of these fitted exponents do not over-

lap for a narrow window. Performing the t-test as usual, we found that the inverse multiparticipation ratio for multilinks (1, 1) is always larger than the inverse multiparticipation ratio for multilinks (1, 0). In the case  $Y^{(1,1),1,in}(k)$  vs  $Y^{(1,0),1,in}(k)$ , the t-test gives a p-value equal to  $3.70 \cdot 10^{-17}$  and an average value  $\langle \log(Y^{(1,1),1,in}(k)/Y^{(1,0),1,in}(k)) \rangle = 0.39$ . In the case  $Y^{(1,1),1,out}$  vs  $Y^{(1,0),1,out}(k)$ , the p-value is equal to  $5.48 \cdot 10^{-19}$  and the average value is  $\langle \log(Y^{(1,1),1,out}(k)/Y^{(1,0),1,out}(k)) \rangle = 0.33$ .

In the *in*- and *out*-citation layers, the fitted exponents  $\lambda_{\vec{m},2,(in/out)}$  regarding multilinks (1, 1) are not significantly different from those regarding multilinks (0, 1). Nevertheless, the paired Student's t-test shows how the inverse multiparticipation ratio for multilinks (1, 1) is always larger than the inverse multiparticipation ratio for multilinks (0, 1). In the case  $Y^{(1,1),2,in}(k)$  vs  $Y^{(0,1),2,in}(k)$ , the t-test gives a p-value equal to  $7.60 \cdot 10^{-21}$  and an average value  $\langle \log(Y^{(1,1),2,in}(k)/Y^{(0,1),2,in}(k)) \rangle = 0.34$ . In the case  $Y^{(1,1),2,out}(k)$  vs  $Y^{(0,1),2,out}(k)$ , the p-value is equal to  $1.12 \cdot 10^{-15}$  and the average value is  $\langle \log(Y^{(1,1),2,out}(k)/Y^{(0,1),2,out}(k)) \rangle = 0.34$ .

#### 4.1.7 $\Psi$ and $\Xi$

As an example of a possible application of the indicators  $\Psi$  and  $\Xi$ , we analyze a case inspired by the CoCi-PRE multiplex network. Due to the numerical limitations of the programs that are able to evaluate the entropy of multiplex ensembles, we perform a finite-size analysis of the indicators  $\Psi$  and  $\Xi$  as a function of the size of the multiplex network  $N = 128, 256, \dots, 2048$ . In particular, we consider the following undirected multiplex ensembles:

- *Correlated weighted multiplex ensemble.* First, we create the correlated multiplex ensemble with power-law expected multidegree distributions with

Table 4.9: CoCi-PRE multiplex network: power-law exponents  $\lambda_\alpha$  and  $\beta_\alpha$  and parameters  $p_{\vec{m}}$ ,  $q_{\vec{m}}$  determining the functional behavior for the average multistrength of nodes with a given multidegree  $s^{\vec{m},\alpha,(in/out)}(k^{\vec{m}(in/out)})$  and the average inverse multiparticipation ratio of nodes with a given multidegree,  $Y^{\vec{m},\alpha,(in/out)}(k^{\vec{m},(in,out)})$ , with  $\alpha$  corresponding to the collaboration layer (1) or to the citation layer (2), both for PRE. The coefficient of determination  $R^2$  determining the quality of the power-law fit is also reported

$\vec{m}, \alpha, \text{in/out}$	$\beta_{\vec{m}}$	$q_{\vec{m}}$	$R^2$	$\lambda_{\vec{m}}$	$p_{\vec{m}}$	$R^2$
(1,1), 1, in	$1.02 \pm 0.04$	$-0.33 \pm 0.15$	0.97	$0.76 \pm 0.05$	$-0.03 \pm 0.16$	0.94
(1,1), 1, out	$1.05 \pm 0.04$	$-0.38 \pm 0.14$	0.98	$0.77 \pm 0.05$	$-0.03 \pm 0.17$	0.94
(1,0), 1, in	$0.98 \pm 0.05$	$-0.97 \pm 0.17$	0.96	$0.88 \pm 0.03$	$-0.03 \pm 0.09$	0.99
(1,0), 1, out	$0.97 \pm 0.05$	$-0.95 \pm 0.16$	0.97	$0.90 \pm 0.02$	$0.00 \pm 0.06$	0.99
(1,1), 2, in	$1.30 \pm 0.07$	$0.47 \pm 0.25$	0.95	$0.73 \pm 0.05$	$-0.17 \pm 0.16$	0.94
(1,1), 2, out	$1.32 \pm 0.08$	$0.45 \pm 0.26$	0.95	$0.74 \pm 0.04$	$-0.50 \pm 0.20$	0.80
(0,1), 2, in	$1.11 \pm 0.01$	$-0.01 \pm 0.07$	0.99	$0.75 \pm 0.05$	$-0.12 \pm 0.16$	0.95
(0,1), 2, out	$1.10 \pm 0.02$	$0.06 \pm 0.09$	0.98	$0.69 \pm 0.05$	$-0.62 \pm 0.22$	0.77

exponents  $\gamma^{(1,m_2)} = 2.6$  for  $m_2 = 0, 1$  and  $\gamma^{(0,1)} = 1.9$  (for multidegree (0, 1) we impose a structural cut-off). In particular, in order to avoid the effects of fluctuations in the expected multidegree sequence, we rank the multidegrees as  $r = 1, 2, \dots, N$ , and take the sequence in which the multidegree  $k_r^{\vec{m}}$  of rank  $r$  is defined by

$$\frac{r}{N} = \int_{k_r^{\vec{m}}}^K P(k^{\vec{m}}) dk^{\vec{m}}, \quad (4.10)$$

where we take the maximal cut-off  $K = N^{1/\gamma^{\vec{m}}}$  for  $\gamma^{\vec{m}} > 2$  and  $K = \sqrt{\langle k^{\vec{m}} \rangle N}$  for  $\gamma^{\vec{m}} < 2$ . Note that this is possible because the expected multidegrees are real values. Moreover, the expected multistrengths are assumed to satisfy

$$s_i^{\vec{m},\alpha} = c_{\vec{m},1}(k^{\vec{m},\alpha})^{\lambda_{\vec{m},\alpha}}, \quad (4.11)$$

with  $c_{\vec{m},\alpha} = 1$ ,  $\beta^{(1,m_2),1} = 1$  for  $m_2 = 0, 1$ ,  $\beta^{(1,1),2} = 1.3$ , and  $\beta^{(0,1),2} = 1.1$ .

- *Uncorrelated weighted multiplex ensemble.* In this ensemble, we set the expected degree  $k_i^\alpha$  of every node  $i$  in every layer  $\alpha = 1, 2$  to be equal to the sum of the expected multidegrees (with  $m_\alpha = 1$ ) in the correlated weighted multiplex ensemble. Moreover, we set the expected strengths  $s_i^\alpha$  of every node  $i$  in every layer  $\alpha$  to be equal to the sum of the expected multistrengths of node  $i$  in layer  $\alpha$  in the correlated weighted multiplex ensemble.

We measure the indicator  $\Psi$  that compares the entropy of a weighted multiplex ensemble  $\mathcal{S}$  with the entropy of a weighted multiplex ensemble in which weights are distributed homogeneously. Therefore,  $\Psi$  can be defined as

$$\Psi = \frac{|\mathcal{S} - \langle \mathcal{S} \rangle_{\pi(w)}|}{\langle (\delta \mathcal{S})^2 \rangle_{\pi(w)}}, \quad (4.12)$$

where the average  $\langle \dots \rangle_{\pi(w)}$  is calculated over multiplex networks with the same structural properties but with weights distributed homogeneously. In particular, when the weight distribution is randomized, the multiplex networks are constrained in such a way that each link must have a minimal weight (i.e.  $w_{ij} > 1$ ),

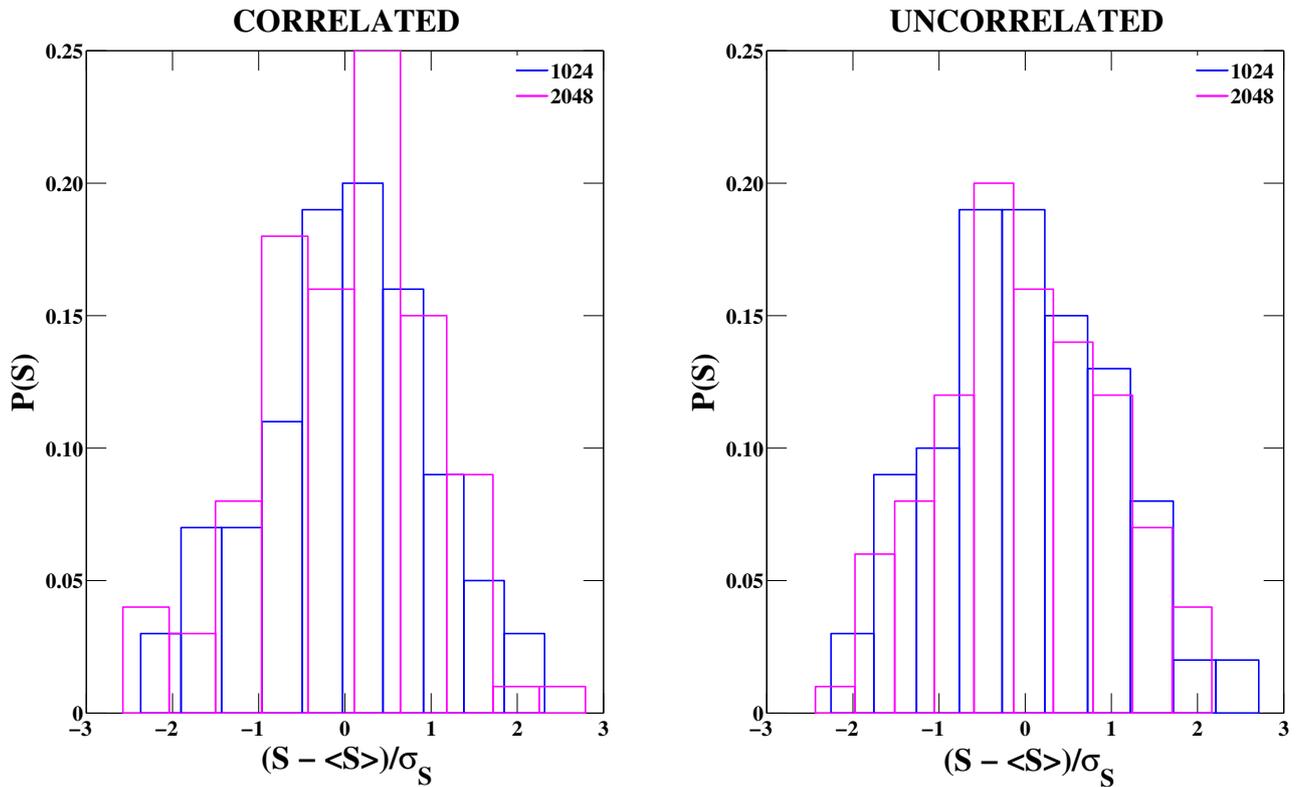


Figure 4.1: The  $P(\mathcal{S})$  distribution in the null models for correlated and uncorrelated multiplex ensembles in which the weights are distributed uniformly over the links of the multiplex network. The  $P(\mathcal{S})$  distributions are calculated over 100 randomizations of the weights for multiplex networks of  $N = 1024$  and  $N = 2048$  nodes.

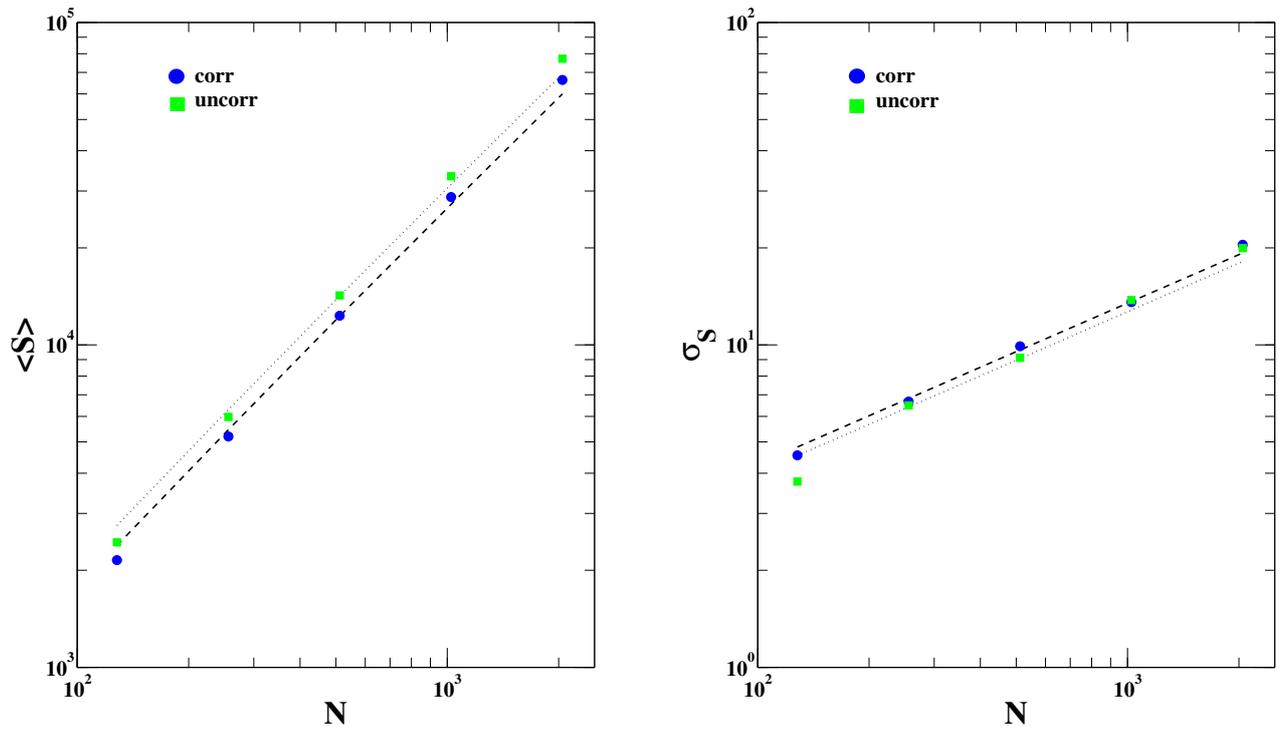


Figure 4.2: The mean  $\langle \mathcal{S} \rangle$  and variance  $\sigma_S$  as a function of the system size  $N$  for the null models of correlated and uncorrelated multiplex ensembles in which the weights are distributed uniformly over the links of the multiplex network. The solid lines indicate the fit of the data in which we assume  $\langle \mathcal{S} \rangle = aN \log N$  and  $\sigma_S = b\sqrt{N}$ .

while the remaining of the total weight is distributed randomly across links. When numerically evaluating  $\langle \dots \rangle_{\pi(w)}$ , we obtain the average over 100 weight randomizations.

The distribution  $P(\mathcal{S})$  of the entropy  $\mathcal{S}$  calculated over these randomizations, both for the uncorrelated weighted multiplex ensemble and for the correlated weighted multiplex ensemble, is shown in Fig. 4.1. In both cases, we observe a distribution that can be fitted by a Gaussian function with mean and variance scaling as  $\langle \mathcal{S} \rangle \propto N \log N$  and  $\langle (\delta \mathcal{S})^2 \rangle_{\pi(w)} \propto \sqrt{N}$  (See Fig. 4.2). We call  $\Psi^{corr}$  the indicator  $\Psi$  calculated on the correlated multiplex ensemble and indicate with  $\Psi^{uncorr}$  the indicator  $\Psi$  calculated on the corresponding uncorrelated multiplex ensemble. Finally, to quantify the additional amount of information carried by the correlated multiplex ensemble with respect to the uncorrelated multiplex ensemble, we measure the indicator  $\Xi$  as

$$\Xi = \frac{\Psi^{corr}}{\Psi^{uncorr}}. \quad (4.13)$$

The finite-size scaling of  $\Psi^{corr}$ ,  $\Psi^{uncorr}$  and  $\Xi$  are shown in Fig. 4 in the manuscript.

## 4.2 Weighted multiplex ensembles

### 4.2.1 Examples of uncorrelated weighted multiplex networks

#### Multiplex ensembles with given expected strength sequence in each layer

We consider here the multiplex ensemble in which we fix the expected strength  $s_i^\alpha$  of every node  $i$ , in each layer  $\alpha$ . We have  $K = M \cdot N$  constraints in the system

indicated with a label  $\alpha = 1, 2, \dots, M$ . These constraints are given by

$$\sum_{\vec{G}} F_{i,\alpha}(\vec{G}) P(\vec{G}) = \sum_{\vec{G}} \left( \sum_{j \neq i} a_{ij}^\alpha \right) P(\vec{G}) = s_i^\alpha \quad (4.14)$$

The probability of a multiplex  $P(\vec{G})$  is given by Eq.(2.36) that in this case can be written as

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{\alpha=1}^M \sum_i \lambda_{i,\alpha} \sum_{j \neq i} a_{ij}^\alpha \right] \quad (4.15)$$

where the partition function  $Z$  can be expressed explicitly as

$$\begin{aligned} Z &= \sum_{\vec{G}} \exp \left[ - \sum_{\alpha=1}^M \sum_i \lambda_{i,\alpha} \sum_{j \neq i} a_{ij}^\alpha \right] \\ &= \prod_{\alpha=1}^M \prod_{i < j} \left[ 1 - e^{-(\lambda_{i,\alpha} + \lambda_{j,\alpha})} \right]^{-1}, \end{aligned} \quad (4.16)$$

and the Lagrangian multipliers  $\lambda_{i,\alpha}$  are fixed by the condition

$$s_i^\alpha = - \frac{\partial \log Z}{\partial \lambda_{i,\alpha}} = \sum_{j \neq i} \frac{e^{-(\lambda_{i,\alpha} + \lambda_{j,\alpha})}}{1 - e^{-(\lambda_{i,\alpha} + \lambda_{j,\alpha})}}. \quad (4.17)$$

The average weight  $\langle a_{ij}^\alpha \rangle$  given by Eq. (2.40) can be calculated explicitly as a function of the Lagrangian multipliers, giving

$$\langle a_{ij}^\alpha \rangle = \frac{e^{-(\lambda_{i,\alpha} + \lambda_{j,\alpha})}}{1 - e^{-(\lambda_{i,\alpha} + \lambda_{j,\alpha})}}, \quad (4.18)$$

which implies, together with Eq. (4.17),  $s_i^\alpha = \sum_{j \neq i} \langle a_{ij}^\alpha \rangle$ .

From Eq. (2.38) we write the marginal probabilities  $\pi_{ij}^\alpha(a_{ij}^\alpha)$  for specific weight  $a_{ij}^\alpha$  as

$$\pi_{ij}^\alpha(a_{ij}^\alpha) = e^{-(\lambda_{i,\alpha} + \lambda_{j,\alpha}) a_{ij}^\alpha} (1 - e^{-(\lambda_{i,\alpha} + \lambda_{j,\alpha})}), \quad (4.19)$$

i.e. the weight of a link is distributed exponentially, with a mean that depends both on the pair of linked nodes  $(i, j)$  and on the layer  $\alpha$ . Moreover, from Eq. (2.41) we can evaluate the probability  $p_{ij}^\alpha$  of having a weight different from zero that is given by

$$p_{ij}^\alpha = e^{-(\lambda_{i,\alpha} + \lambda_{j,\alpha})}. \quad (4.20)$$

Finally the the probability of a multiplex in this ensemble is given by Eq. (2.121) with the marginals  $\pi_{ij}^\alpha(a_{ij}^\alpha)$  given by Eq. (4.19). Therefore the entropy  $\mathcal{S}$  of this canonical multiplex ensemble is given by Eq. (2.122) with the marginals  $\pi_{ij}^\alpha(a_{ij}^\alpha)$  given by Eq. (4.19)

### **Multiplex ensembles with given expected strength sequence, given expected degree sequence and given expected sequences $\{u_i^\alpha\}$ in each layer**

The last example of uncorrelated multiplex that we will consider is the one in which we fix the expected strength  $s_i^\alpha$ , the expected degree  $k_i^\alpha$  and the expected  $u_i^\alpha$  of every node  $i$  in each layer  $\alpha$ . We have  $K = M \times 3N$  constraints in the system. These constraints are given by

$$\begin{aligned} \sum_{\vec{G}} F_{i,\alpha}(\vec{G})P(\vec{G}) &= \sum_{\vec{G}} \left( \sum_{j \neq i} a_{ij}^\alpha \right) P(\vec{G}) = s_i^\alpha \\ \sum_{\vec{G}} F_{i,\alpha}(\vec{G})P(\vec{G}) &= \sum_{\vec{G}} \left( \sum_{j \neq i} \theta(a_{ij}^\alpha) \right) P(\vec{G}) = k_i^\alpha \\ \sum_{\vec{G}} F_{i,\alpha}(\vec{G})P(\vec{G}) &= \sum_{\vec{G}} \left( \sum_{j \neq i} (a_{ij}^\alpha)^2 \right) P(\vec{G}) = u_i^\alpha \end{aligned} \quad (4.21)$$

with  $\alpha = 1, 2, \dots, M$ . We introduce the Lagrangian multipliers  $\lambda_{i,\alpha}$  for the first set of  $N \times M$  constraints, the Lagrangian multipliers  $\omega_{i,\alpha}$  for the second set of  $N \times M$  constraints and the Lagrangian multipliers  $z_{i,\alpha}$  for the third set of  $N \times M$  constraints. Therefore, the probability  $P(\vec{G})$  of a multiplex in this ensemble, of general expression given by Eq. (2.36), in this specific example is given by

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{\alpha=1}^M \sum_i \lambda_{i,\alpha} \sum_{j \neq i} a_{ij}^\alpha - \sum_{\alpha=1}^M \sum_i \omega_{i,\alpha} \sum_{j \neq i} \theta(a_{ij}^\alpha) - \sum_{\alpha=1}^M \sum_i z_{i,\alpha} \sum_{j \neq i} (a_{ij}^\alpha)^2 \right]$$

If we define as  $I_{ij}^\alpha$  the series

$$I_{ij}^\alpha = \sum_{a_{ij}^\alpha=1}^{S^\alpha} \exp \left[ -(\lambda_{i,\alpha} + \lambda_{j,\alpha}) a_{ij}^\alpha - (z_{i,\alpha} + z_{j,\alpha}) (a_{ij}^\alpha)^2 \right], \quad (4.22)$$

where  $S^\alpha = \sum_{i=1}^N s_i^\alpha$ . The sum  $I_{ij}^\alpha$  is convergent when  $(z_{i,\alpha} + z_{j,\alpha}) > 0$ , the partition function  $Z$  can be expressed as

$$Z = \prod_{\alpha=1}^M \prod_{i < j} \left[ 1 + e^{-(\omega_{i,\alpha} + \omega_{j,\alpha})} I_{ij}^\alpha \right] \quad (4.23)$$

The Lagrangian multipliers are fixed by the conditions

$$\begin{aligned} -\frac{\partial \log Z}{\partial \lambda_{i,\alpha}} &= s_i^\alpha \\ -\frac{\partial \log Z}{\partial \omega_{i,\alpha}} &= k_i^\alpha \\ -\frac{\partial \log Z}{\partial z_{i,\alpha}} &= u_i^\alpha \end{aligned} \quad (4.24)$$

The average weight of the link  $(i, j)$  in layer  $\alpha$ , i.e.  $\langle a_{ij}^\alpha \rangle$ , is given by Eq. (2.40)

that in this case reads

$$\begin{aligned} \langle a_{ij}^\alpha \rangle &= \frac{e^{-(\omega_{i,\alpha} + \omega_{j,\alpha})}}{[1 + e^{-(\omega_{i,\alpha} + \omega_{j,\alpha})} I_{ij}^\alpha]} \times \\ &\times \left[ \sum_{a_{ij}^\alpha=1}^{S^\alpha} a_{ij}^\alpha \exp\left(-(\lambda_{i,\alpha} + \lambda_{j,\alpha}) a_{ij}^\alpha - (z_{i,\alpha} + z_{j,\alpha})(a_{ij}^\alpha)^2\right) \right] \end{aligned}$$

From Eq. (2.38) we write the marginal probabilities  $\pi_{ij}^\alpha(a_{ij}^\alpha)$  for this specific ensemble that is given by

$$\pi_{ij}^\alpha(a_{ij}^\alpha) = \frac{e^{-(\lambda_{i,\alpha} + \lambda_{j,\alpha}) a_{ij}^\alpha - (\omega_{i,\alpha} + \omega_{j,\alpha}) a_{ij}^\alpha - (z_{i,\alpha} + z_{j,\alpha})(a_{ij}^\alpha)^2}}{[1 + e^{-(\omega_{i,\alpha} + \omega_{j,\alpha})} I_{ij}^\alpha]} \quad (4.25)$$

Moreover, from Eq. (2.41) the probability  $p_{ij}^\alpha$  that the link  $(i, j)$  in layer  $\alpha$  has weight different from zero is given by

$$p_{ij}^\alpha = \frac{e^{-(\omega_{i,\alpha} + \omega_{j,\alpha})} I_{ij}^\alpha}{[1 + e^{-(\omega_{i,\alpha} + \omega_{j,\alpha})} I_{ij}^\alpha]} \quad (4.26)$$

The probability of a multiplex in this ensemble is given by Eq. (2.121) with the marginals  $\pi_{ij}^\alpha(a_{ij}^\alpha)$  given by Eq. (4.25) while the entropy  $\mathcal{S}$  of this canonical multiplex ensemble is given by Eq. (2.122) with the marginals  $\pi_{ij}^\alpha(a_{ij}^\alpha)$  given by Eq. (4.25)

## 4.2.2 Examples of correlated weighted multiplex networks

### Multiplex ensembles with given expected multistrength sequence $\{s_{i,\alpha}^{\vec{m}}\}$

We study here a correlated weighted multiplex ensemble in which we fix the average strength sequence  $s_{i,\alpha}^{\vec{m}}$  for each node  $i$ , in each layer  $\alpha$  such that  $m_\alpha = 1$ , for a given multilink  $\vec{m}$ . Following the previous line of reasoning, we can express

properly just  $N \cdot M \cdot 2^{M-1}$  constraints.

These constraints are given by

$$\sum_{\vec{G}} F_{i,\alpha}^{\vec{m}}(\vec{G}) P(\vec{G}) = \sum_{\vec{G}} \left( \sum_{j \neq i} A_{ij}^{\vec{m}} a_{ij}^\alpha \right) P(\vec{G}) = s_{i,\alpha}^{\vec{m}}, \quad (4.27)$$

with  $i = 1, \dots, N$ ,  $\vec{m} = (m_1, m_2, \dots, m_\beta, \dots, m_M)$  with  $m_\beta = 0, 1$  and finally  $\alpha = 1, \dots, M$  with the condition  $m_\alpha = 1$ . The canonical probability  $P(\vec{G})$  of the multiplex in the ensemble is

$$\begin{aligned} P(\vec{G}) &= \frac{1}{Z} \exp \left[ - \sum_{\vec{m} \neq \vec{0}} \sum_{\alpha=1}^M \sum_i \lambda_{i,\alpha}^{\vec{m}} \sum_{j \neq i} A_{ij}^{\vec{m}} a_{ij}^\alpha \right] \\ &= \frac{1}{Z} \prod_{i < j} \exp \left[ - \sum_{\vec{m} \neq \vec{0}} \sum_{\alpha=1}^M (\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}}) A_{ij}^{\vec{m}} a_{ij}^\alpha \right], \end{aligned} \quad (4.28)$$

where the partition function  $Z$  can be expressed explicitly as

$$Z = \prod_{i < j} \mathcal{Z}_{ij} \quad (4.29)$$

where

$$\mathcal{Z}_{ij} = \sum_{\vec{m}} \prod_{\alpha=1}^M \left( \frac{e^{-(\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}})}}{1 - e^{-(\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}})}} \right)^{m_\alpha}, \quad (4.30)$$

The Lagrangian multipliers  $\lambda_{i,\alpha}^{\vec{m}}$ , with  $\alpha$  such that  $m_\alpha = 1$ , are fixed by the conditions

$$- \frac{\partial \log Z}{\partial \lambda_{i,\alpha}^{\vec{m}}} = s_{i,\alpha}^{\vec{m}} = \sum_{j \neq i} \langle a_{ij}^\alpha A_{ij}^{\vec{m}} \rangle, \quad (4.31)$$

where  $\langle a_{ij}^\alpha A_{ij}^{\vec{m}} \rangle$  is the average weight of the link between node  $i$  and node  $j$  on the multilink  $\vec{m}$ , in the layer  $\alpha$ . This quantity can be computed as

$$\langle a_{ij}^\alpha A_{ij}^{\vec{m}} \rangle = \frac{1}{\mathcal{Z}_{ij}} \left( \frac{1}{1 - e^{-(\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}})}} \right) \prod_{\beta=1}^M \left( \frac{e^{-(\lambda_{i,\beta}^{\vec{m}} + \lambda_{j,\beta}^{\vec{m}})}}{1 - e^{-(\lambda_{i,\beta}^{\vec{m}} + \lambda_{j,\beta}^{\vec{m}})}} \right)^{m_\beta}.$$

We can calculate the probability of a vector  $\vec{a}_{ij} = (a_{ij}^1, a_{ij}^2, \dots, a_{ij}^M)$  characterizing the weights of the links between node  $i$  and node  $j$  in all the layers, getting

$$\pi_{ij}(\vec{a}_{ij}) = \frac{1}{\mathcal{Z}_{ij}} e^{-\sum_{\alpha=1, M} (\lambda_{i,\alpha}^{\vec{m}^{ij}} + \lambda_{j,\alpha}^{\vec{m}^{ij}}) a_{ij}^\alpha}. \quad (4.32)$$

These probabilities satisfy the normalisation condition given by Eq. (2.139). The probability  $p_{ij}^{\vec{m}}$  of a multilink  $\vec{m}$  between the node  $i$  and the node  $j$  is given by

$$p_{ij}^{\vec{m}} = \langle A_{ij}^{\vec{m}} \rangle = \frac{1}{\mathcal{Z}_{ij}} \prod_{\alpha=1}^M \left( \frac{e^{-(\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}})}}{1 - e^{-(\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}})}} \right)^{m_\alpha}, \quad (4.33)$$

where these probabilities satisfy the normalisation condition given by Eq. (2.143) and are related to the probabilities  $\pi_{ij}^{\vec{m}}(\vec{a}_{ij})$  given by Eq. (4.32), by Eq. (2.144).

Probability  $P(\vec{G})$  and entropy  $\mathcal{S}$  follow Eqs. (2.137), (2.145) respectively.

### **Multiplex ensembles with given expected $\nu$ -multistrength sequence $\{s_{i,\alpha}^\nu\}$**

In the case in which one wants to describe multiplex networks with many layers  $M$ , one can consider to fix the average  $\nu$ -multistrength sequence  $\{s_{i,\alpha}^\nu\}$  with  $i = 1, 2, \dots, N$  and  $\nu = 1, 2, \dots, M$ . Therefore, the number of constraints of

the previous example is reduced to just  $N \cdot M^2$  soft constraints given by

$$\sum_{\vec{G}} F_{i,\alpha}^\nu(\vec{G}) P(\vec{G}) = \sum_{\vec{G}} \left( \sum_{j \neq i} a_{ij}^\alpha A_{ij}^\nu \right) P(\vec{G}) = s_{i,\alpha}^\nu. \quad (4.34)$$

In this case, the probability  $P(\vec{G})$  of a multiplex network  $\vec{G}$  in this ensemble is expressed in terms of the  $N \times M^2$  Lagrangian multipliers  $\lambda_{i,\alpha}^\nu$  and is given by

$$P(\vec{G}) = \frac{1}{Z} \exp \left[ - \sum_{i < j} \sum_{\nu=1}^M \sum_{\alpha=1}^M (\lambda_{i,\alpha}^\nu + \lambda_{j,\alpha}^\nu) A_{ij}^\nu a_{ij}^\alpha \right],$$

where the partition function  $Z$  can be expressed as

$$Z = \prod_{i < j} \mathcal{Z}_{ij} \quad (4.35)$$

with

$$\mathcal{Z}_{ij} = \sum_{\nu=0}^M \sum_{\vec{m} | \nu(\vec{m}) = \nu} \prod_{\alpha=1}^M \left( \frac{e^{-(\lambda_{i,\alpha}^\nu + \lambda_{j,\alpha}^\nu)}}{1 - e^{-(\lambda_{i,\alpha}^\nu + \lambda_{j,\alpha}^\nu)}} \right)^{m_\alpha}. \quad (4.36)$$

The Lagrangian multipliers are fixed by the conditions in Eq. (4.34), or equivalently by

$$- \frac{\partial \log Z}{\partial \lambda_{i,\alpha}^\nu} = s_{i,\alpha}^\nu = \sum_{j \neq i} \langle a_{ij}^\alpha A_{ij}^\nu \rangle. \quad (4.37)$$

The probability  $P(\vec{G})$  of a multiplex network  $\vec{G}$  in this ensemble is given by Eq. (2.137) and the entropy of the ensemble takes the simple expression given by Eq. (2.145) where  $\pi_{ij}(\vec{a}_{ij})$  is given by

$$\pi_{ij}(\vec{a}_{ij}) = \frac{1}{\mathcal{Z}_{ij}} e^{-\sum_{\alpha=1, M} (\lambda_{i,\alpha}^{\nu ij} + \lambda_{j,\alpha}^{\nu ij}) a_{ij}^\alpha}. \quad (4.38)$$

Finally, the probability  $p_{ij}^\nu$  that the node  $i$  and the node  $j$  are linked by a  $\nu$ -

multilink is given by

$$p_{ij}^\nu = \frac{1}{\mathcal{Z}_{ij}} \sum_{\vec{m} | \nu(\vec{m}) = \nu} \prod_{\alpha=1}^M \left( \frac{e^{-(\lambda_{i,\alpha}^\nu + \lambda_{j,\alpha}^\nu)}}{1 - e^{-(\lambda_{i,\alpha}^\nu + \lambda_{j,\alpha}^\nu)}} \right)^{m_\alpha}, \quad (4.39)$$

while the average weight of the link  $a_{ij}^\alpha$  belonging to a  $\nu$ -multilink is given by

$$\begin{aligned} \langle a_{ij}^\alpha A_{ij}^\nu \rangle &= \frac{1}{\mathcal{Z}_{ij}} \left( \frac{1}{1 - e^{-(\lambda_{i,\alpha}^\nu + \lambda_{j,\alpha}^\nu)}} \right) \times \\ &\times \sum_{\vec{m} | \nu(\vec{m}) = \nu} m_\alpha \prod_{\beta=1}^M \left( \frac{e^{-(\lambda_{i,\beta}^\nu + \lambda_{j,\beta}^\nu)}}{1 - e^{-(\lambda_{i,\beta}^\nu + \lambda_{j,\beta}^\nu)}} \right)^{m_\beta}. \end{aligned} \quad (4.40)$$

**Multiplex ensembles with given expected multidegree sequence  $\{k_i^{\vec{m}}\}$ , given expected multistrength sequence  $\{s_{i,\alpha}^{\vec{m}}\}$  and given expected sequence  $\{u_{i,\alpha}^{\vec{m}}\}$**

As a fourth case of correlated weighted multiplex ensemble, we consider the case in which we fix the average multidegree  $k_i^{\vec{m}}$  of node  $i$ , for each node  $i = 1, \dots, N$ , for  $\vec{m} \neq \vec{0}$ . Moreover, for each node  $i$  in layer  $\alpha$  we impose the average multistrength  $s_{i,\alpha}^{\vec{m}}$  and the second moment of the weights incident to it and belonging to a multilink  $\vec{m}$ , i.e.  $u_{i,\alpha}^{\vec{m}}$ . The number of independent constraints is therefore  $K = (2^M - 1) \cdot N + 2^M \cdot M \cdot N$ .

In particular, the constraints we are imposing are the following,

$$\begin{aligned} \sum_{\vec{G}} F_{i,\alpha}^{\vec{m}}(\vec{G}) P(\vec{G}) &= \sum_{\vec{G}} \left( \sum_{j \neq i} A_{ij}^{\vec{m}} a_{ij}^\alpha \right) P(\vec{G}) = s_{i,\alpha}^{\vec{m}} \\ \sum_{\vec{G}} F_i^{\vec{m}}(\vec{G}) P(\vec{G}) &= \sum_{\vec{G}} \left( \sum_{j \neq i} A_{ij}^{\vec{m}} \right) P(\vec{G}) = k_i^{\vec{m}} \\ \sum_{\vec{G}} F_{i,\alpha}^{\vec{m}}(\vec{G}) P(\vec{G}) &= \sum_{\vec{G}} \left( \sum_{j \neq i} (A_{ij}^{\vec{m}} a_{ij}^\alpha)^2 \right) P(\vec{G}) = u_{i,\alpha}^{\vec{m}} \end{aligned} \quad (4.41)$$

The canonical probability  $P(\vec{G})$  of the multiplex in the ensembles is

$$\begin{aligned}
 P(\vec{G}) &= \frac{1}{Z} \exp \left[ - \sum_{i < j} \sum_{\vec{m} \neq \vec{0}} (\omega_i^{\vec{m}} + \omega_j^{\vec{m}}) A_{ij}^{\vec{m}} \right] \times \\
 &\times \exp \left[ - \sum_{i < j} \sum_{\vec{m} \neq \vec{0}} \sum_{\alpha=1}^M (\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}}) A_{ij}^{\vec{m}} a_{ij}^\alpha \right] \\
 &\times \exp \left[ - \sum_{i < j} \sum_{\vec{m} \neq \vec{0}} \sum_{\alpha=1}^M (z_{i,\alpha}^{\vec{m}} + z_{j,\alpha}^{\vec{m}}) A_{ij}^{\vec{m}} (a_{ij}^\alpha)^2 \right]
 \end{aligned} \tag{4.42}$$

The partition function  $Z$  can be expressed explicitly as

$$\begin{aligned}
 Z &= \prod_{i < j} \mathcal{Z}_{ij} \\
 &= \prod_{i < j} \left( 1 + \sum_{\vec{m} \neq \vec{0}} e^{-(\omega_i^{\vec{m}} + \omega_j^{\vec{m}})} \prod_{\alpha=1}^M \left( I_{ij}^{\vec{m},\alpha} \right)^{m_\alpha} \right)
 \end{aligned} \tag{4.43}$$

where  $I_{ij}^{\vec{m},\alpha}$  is given by

$$I_{ij}^{\vec{m},\alpha} = \sum_{a_{ij}^\alpha=1}^{S^{\vec{m},\alpha}} \exp \left[ -(\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}}) a_{ij}^\alpha - (z_{i,\alpha}^{\vec{m}} + z_{j,\alpha}^{\vec{m}}) (a_{ij}^\alpha)^2 \right],$$

where  $S^{\vec{m},\alpha} = \sum_{i=1}^N s_{i,\alpha}^{\vec{m}}$ . The Lagrangian multipliers are fixed by the conditions

$$\begin{aligned}
 -\frac{\partial \log Z}{\partial \lambda_{i,\alpha}^{\vec{m}}} &= s_{i,\alpha}^{\vec{m}} = \sum_{j \neq i} \langle a_{ij}^\alpha A_{ij}^{\vec{m}} \rangle, \\
 -\frac{\partial \log Z}{\partial \omega_i^{\vec{m}}} &= k_i^{\vec{m}} = \sum_{j \neq i} \langle A_{ij}^{\vec{m}} \rangle, \\
 -\frac{\partial \log Z}{\partial z_{i,\alpha}^{\vec{m}}} &= u_{i,\alpha}^{\vec{m}} = \sum_{j \neq i} \langle (a_{ij}^\alpha)^2 A_{ij}^{\vec{m}} \rangle
 \end{aligned} \tag{4.44}$$

The average weight  $\langle a_{ij}^\alpha A_{ij}^{\vec{m}} \rangle$  of the multilink  $\vec{m}$  between nodes  $i$  and  $j$  in the layer  $\alpha$  and the probability  $p_{ij}^{\vec{m}}$  of a multilink  $\vec{m}$  between node  $i$  and node  $j$  are

given respectively by

$$\begin{aligned}
\langle a_{ij}^\alpha A_{ij}^{\vec{m}} \rangle &= -\frac{e^{-(\omega_i^{\vec{m}} + \omega_j^{\vec{m}})}}{\mathcal{Z}_{ij}} \left( \frac{\partial I_{ij}^{\vec{m},\alpha}}{\partial (\lambda_{i,\alpha}^{\vec{m}} + \lambda_{j,\alpha}^{\vec{m}})} \right) \times \\
&\quad \times \prod_{\beta \neq \alpha}^M \left( I_{ij}^{\vec{m},\beta} \right)^{m_\beta} \\
p_{ij}^{\vec{m}} &= \frac{e^{-(\omega_i^{\vec{m}} + \omega_j^{\vec{m}})}}{\mathcal{Z}_{ij}} \prod_{\alpha=1}^M \left( I_{ij}^{\vec{m},\alpha} \right)^{m_\alpha}
\end{aligned} \tag{4.45}$$

The probability of a specific multiweight  $\vec{a}_{ij}$  in the between the nodes  $(i, j)$  is

$$\begin{aligned}
\pi_{ij}(\vec{a}_{ij}) &= \frac{e^{-(\omega_i^{\vec{m}^{ij}} + \omega_j^{\vec{m}^{ij}})}}{\mathcal{Z}_{ij}} e^{-\sum_{\alpha=1, M} (\lambda_{i,\alpha}^{\vec{m}^{ij}} + \lambda_{j,\alpha}^{\vec{m}^{ij}}) a_{ij}^\alpha} \times \\
&\quad \times e^{-\sum_{\alpha=1, M} (z_{i,\alpha}^{\vec{m}^{ij}} + z_{j,\alpha}^{\vec{m}^{ij}}) (a_{ij}^\alpha)^2}
\end{aligned} \tag{4.46}$$

As previously, probability  $P(\vec{G})$  and entropy  $\mathcal{S}$  follow Eqs. (2.137), (2.145) respectively.

## 4.3 Controllability and minimal degrees

### 4.3.1 Stability condition

Here we consider the stability of the replica-symmetric solution of Eqs. (3.68) (see e.g. Montanari and Ricci-Tersenghi (2003); Rivoire et al. (2003); Castellani et al. (2005); Lucibello and Ricci-Tersenghi (2014) for discussions on the RS stability). The replica symmetry assumes that all cavity fields have the same distributions  $\mathcal{P}(h)$  and  $\hat{\mathcal{P}}(\hat{h})$ , that in the zero temperature limit can be parametrized by mixtures of delta functions. If we relax such assumption, we have to enlarge the functional space by considering distributions  $\mathcal{Q}[\mathcal{P}]$  and  $\hat{\mathcal{Q}}[\hat{\mathcal{P}}]$

of cavity field distributions. There are two ways in which the replica-symmetric solution can be recovered in this enlarged functional space: 1)  $\mathcal{Q}[\mathcal{P}] = \delta[\mathcal{P} - \mathcal{P}^*]$  with  $\mathcal{P}^*(h) = \sum_{\alpha} w_{\alpha} \delta(h - h_{\alpha})$ , and 2)  $\mathcal{Q}[\mathcal{P}] = \sum_{\alpha} w_{\alpha} \delta[\mathcal{P} - \delta(h - h_{\alpha})]$ . For the sake of simplicity in this thesis we are going to show just the first approach. In this situation, the replica symmetric solution can become unstable towards a functional  $\mathcal{Q}$  with non-zero variance and this corresponds to the dynamical instability of the solutions under iteration of the Eqs. (3.68). In other words, the instability means that the distribution of cavity fields does not actually concentrate around discrete values, therefore the corresponding solution is not reachable from any finite temperature. In order to evaluate this type of instability we compute the Jacobian of the system of Eqs. (3.68) and impose that all its eigenvalues have modulus less than one. The  $6 \times 6$  Jacobian matrix reads

$$J = \begin{pmatrix} 0 & 0 & 0 & 0 & G'_{1,out}(\hat{w}_2) & 0 \\ 0 & 0 & 0 & G'_{1,out}(1 - \hat{w}_1) & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & G'_{1,in}(w_2) & 0 & 0 & 0 & 0 \\ G'_{1,in}(1 - w_1) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 \end{pmatrix}. \quad (4.47)$$

where

$$\begin{aligned}
G_{1,in}(x) &= \sum_k \frac{k}{\langle k \rangle_{in}} P^{in}(k) x^{k-1} \\
G'_{1,in}(x) &= \sum_k \frac{k(k-1)}{\langle k \rangle_{in}} P^{in}(k) x^{k-2} \\
G_{1,out}(x) &= \sum_k \frac{k}{\langle k \rangle_{out}} P^{out}(k) x^{k-1} \\
G'_{1,out}(x) &= \sum_k \frac{k(k-1)}{\langle k \rangle_{out}} P^{out}(k) x^{k-2},
\end{aligned} \tag{4.48}$$

with  $\langle k \rangle_{in} = \langle k \rangle_{out}$ . Two eigenvalues are zero, the other four have degenerate modulus, therefore the stability conditions are

$$\begin{aligned}
G'_{1,in}(1-w_1)G'_{1,out}(\hat{w}_2) &< 1, \\
G'_{1,out}(1-\hat{w}_1)G'_{1,in}(w_2) &< 1.
\end{aligned} \tag{4.49}$$

By considering the zero-energy solution  $w_1 = w_2 = \hat{w}_1 = \hat{w}_2 = 0$  and  $w_3 = \hat{w}_3 = 1$ , emerging for  $P^{in}(1) = P^{out}(1) = 0$ , the stability criteria implies the condition Eq. 3.70 that we rewrite here for convenience,

$$P^{out}(2) < \frac{\langle k \rangle_{in}^2}{2\langle k(k-1) \rangle_{in}}, \quad P^{in}(2) < \frac{\langle k \rangle_{in}^2}{2\langle k(k-1) \rangle_{out}}. \tag{4.50}$$

Notice that for  $P^{in}(1) = P^{out}(1) = 0$  there is also the zero energy solution  $w_1 = 0, w_2 = 1, \hat{w}_1 = 1, \hat{w}_2 = 0$  and the symmetric solution  $w_1 = 1, w_2 = 0, \hat{w}_1 = 0, \hat{w}_2 = 1$ . The first solution is stable when the stability conditions given by Eqs. (4.59) are satisfied, i.e. when

$$G'_{1,in}(1)G'_{1,out}(0) = \frac{\langle k(k-1) \rangle_{in}}{\langle k \rangle_{in}} \frac{2P^{out}(2)}{\langle k \rangle_{out}} < 1, \tag{4.51}$$

the second solution is stable when the following condition is satisfied

$$G'_{1,in}(0)G'_{1,out}(1) = \frac{\langle k(k-1) \rangle_{out}}{\langle k \rangle_{out}} \frac{2P^{in}(2)}{\langle k \rangle_{in}} < 1. \quad (4.52)$$

Therefore, when  $P^{in}(k) = P^{out}(k)$ , these solutions are stable under the same conditions in which the solution  $w_1 = w_2 = \hat{w}_1 = \hat{w}_2 = 0$  is stable, and all these solutions correspond to the same value of the energy density  $E/N = 0$ .

### 4.3.2 Number of driver nodes

The BP equations solving the maximum matching problem on a random network ensemble are expected to give the correct value for density of driver nodes in the limit of large networks  $N \rightarrow \infty$ . In particular, in the region in which BP predicts a zero fraction of driver nodes  $n_D$ , the BP algorithm does not guarantee that the exact number of driver nodes is zero, i.e.  $N_D = 0$ . Nevertheless in our simulations, by running the Hopcroft-Karp algorithm (Hopcroft and Karp, 1973) on finite networks in the region where BP predicts a zero fraction of driver nodes, i.e.  $n_D = 0$ , we have always found that, as soon as we are sufficiently far from the boundary of the region defined by the stability conditions, the networks have a number of driver nodes equal to zero, i.e.  $N_D = 0$ . In Fig. 4.3 we show the histogram of the results obtained by the Hopcroft-Karp algorithm corresponding to the points of Fig. 3.9 with predicted zero fraction, i.e.  $n_D = 0$  of driver nodes.

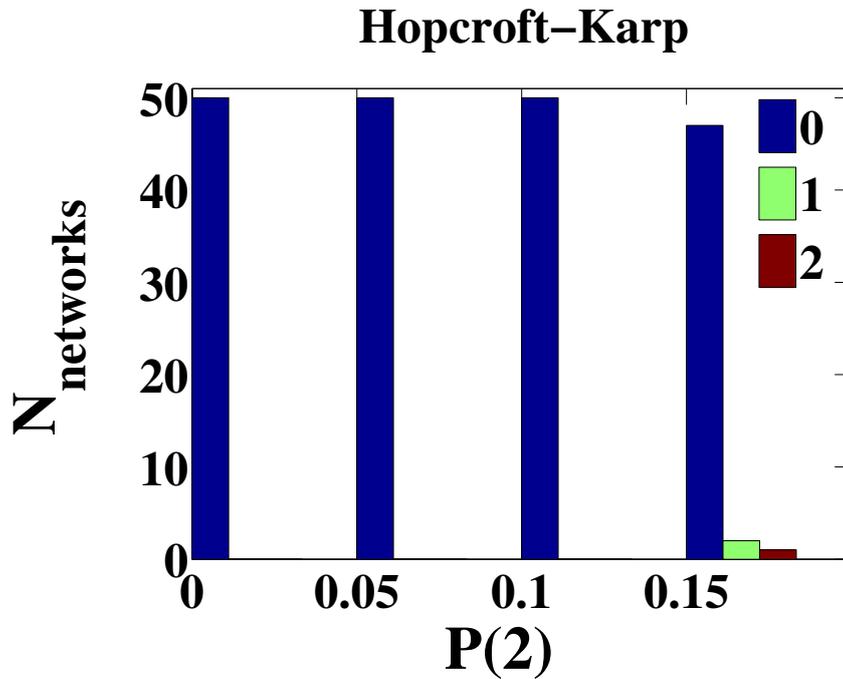


Figure 4.3: Histograms showing the number of network realisations that, out of a total of 50 realisations, show a certain number of driver nodes  $N_D$  in the region of phase space in which BP predicts zero fraction of driver nodes  $n_D = 0$ . The different histograms are displayed as a function of  $P(2)$  for in-degree and out-degree distributions as in Eq. 3.71 of the main text with  $P(1) = 0$  and  $\gamma = 2.3$ . The size of the networks is of  $N = 10^4$ . The histogram refers to the exact matching algorithm by Hopcroft and Karp (1973). As long as we are far from the stability conditions  $P(2) = 0.181947$ , these results show that the expected number of driver nodes is consistent with  $N_D = 0$ .

### 4.3.3 Improving the controllability of scale-free networks

In the Sec. 3.5.3 we gave an example of a power-law network with in-degree distribution equal to out-degree distribution,  $N = 10^4$  nodes, and power-law exponent  $\gamma = 2.3$ . We showed that in this particular case our recipe was quite demanding in terms of fraction of links needed to reach the full controllability of the network. Nevertheless, if we keep the same initial average degree and we consider the degree distributions with a power-law exponent  $\gamma = 3$ , implying

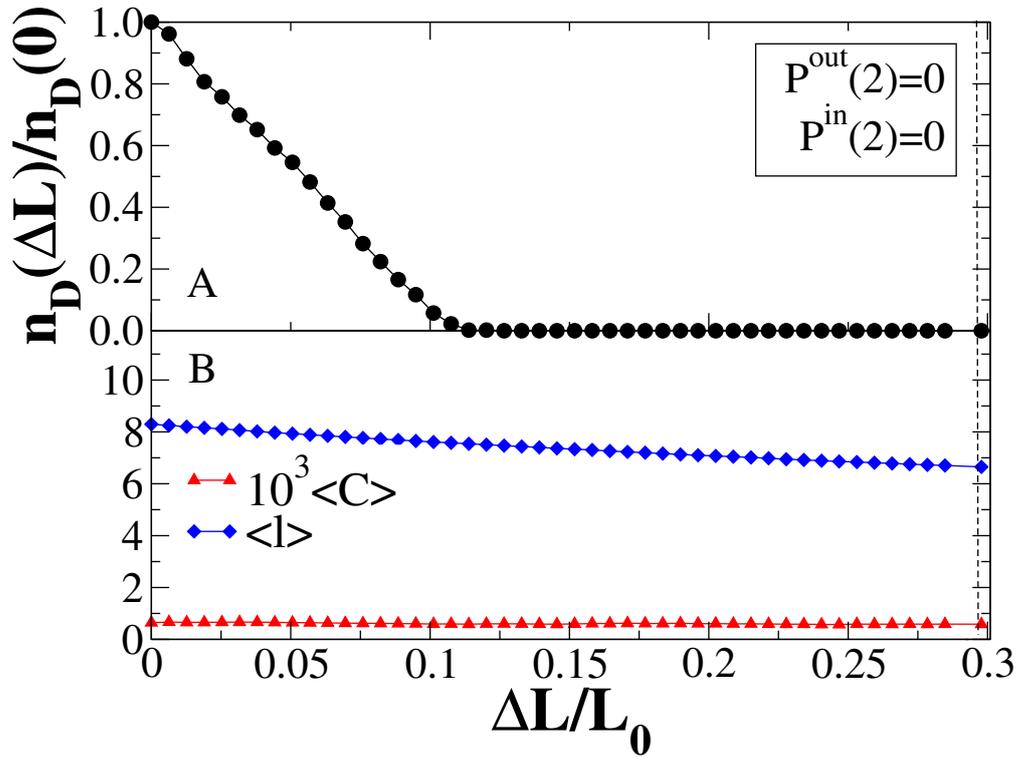


Figure 4.4: Fraction of driver nodes  $n_D(\Delta L)/n_D(0)$  (panel A) average clustering coefficient  $\langle C \rangle$  and average distance  $\langle l \rangle$  (panel B) of the network as a function of the fraction of added links to low degree nodes. The results are obtained solving the MS equations. The initial network is a power-law network with in-degree distribution equal to out-degree distribution,  $N = 10^4$  nodes, and power-law exponent  $\gamma = 3$ . The symbol  $\Delta L$  indicates the number of added links to the network, whereas  $L_0$  indicates the initial number of links of the network.

that we start from a minimum in-degree and our-degree equal to 2, the fraction of links for the full controllability drops to 13% (see Fig. 4.4).

#### 4.3.4 Poisson networks

In Sec. 3.5.2 we have assessed the role of low-degree nodes in the controllability of networks, especially considering uncorrelated random graphs with power-law degree distribution. We consider now Poisson networks with the following degree distribution

$$P^{in}(k) = P^{out}(k) = \begin{cases} P(1) & \text{if } k = 1 \\ P(2) & \text{if } k = 2 \\ C \frac{\lambda^k}{k!} & \text{if } k \in [3, \infty] \end{cases} \quad (4.53)$$

with  $C$  a constant determined by normalization. We especially focus on the situation in which  $P(1) = 0$  and the stability condition for the solution  $\{w_1, w_2, w_3\} = \{0, 0, 1\}$ ,  $\{\hat{w}_1, \hat{w}_2, \hat{w}_3\} = \{0, 0, 1\}$  reads

$$P(2) \leq \frac{\langle k \rangle^2}{2(\langle k^2 \rangle - \langle k \rangle)} \quad (4.54)$$

where  $\langle k \rangle$  and  $\langle k^2 \rangle$  can be easily expressed as

$$\langle k \rangle = 2P(2) + (1 - P(2)) \frac{\lambda(e^\lambda - 1 - \lambda)}{e^\lambda - 1 - \lambda - \lambda^2/2} \quad (4.55)$$

$$\langle k^2 \rangle = 4P(2) + (1 - P(2)) \frac{e^\lambda(\lambda + \lambda^2) - \lambda - 2\lambda^2}{e^\lambda - 1 - \lambda - \lambda^2/2} \quad (4.56)$$

In Fig. 4.5 we show the phase diagram pointing out the fraction of driver nodes  $n_D$  as a function of the parameters  $\lambda$  and  $P(2)$ . The dark grey area defines the region where the zero-energy solution is stable, hence the network has an infinitesimal fraction of driver nodes ( $n_D = 0$ ). Outside this region, the minimum fraction of driver nodes necessary for a full network control is displayed (lowest stable solution of the MS equations).

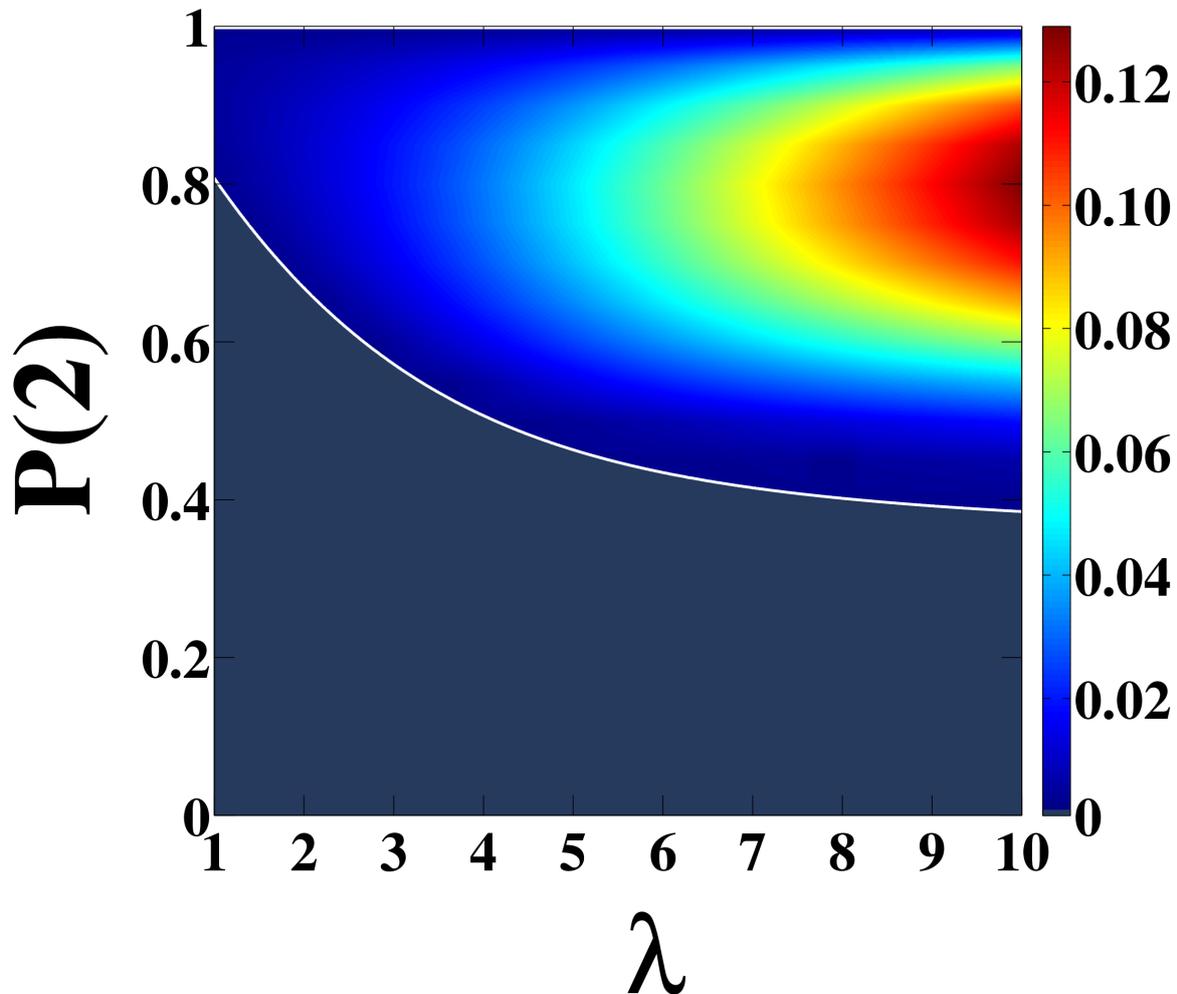


Figure 4.5: Phase diagram indicating the density of driver nodes  $n_D$  (indicated according to the color code on the left) as a function of the parameters  $\lambda$  and  $P(2)$  for networks of nodes with degree distribution given by Eq. (4.53) and  $P(1) = 0$ . The density of driver nodes is obtained by numerically solving Eqs. (3.68). The solid line indicates the stability line.

### 4.3.5 Improving the controllability of Poisson networks

In Sec. 3.5.2 we have described an algorithm that can improve the controllability of networks by adding links to it and reducing the number of nodes with in-degree and out-degree smaller than 3. While in Sec. 3.5.2 we show that such algorithm can be used to improve the controllability of scale-free networks, here we show that the same algorithm can be used to improve the controllability also

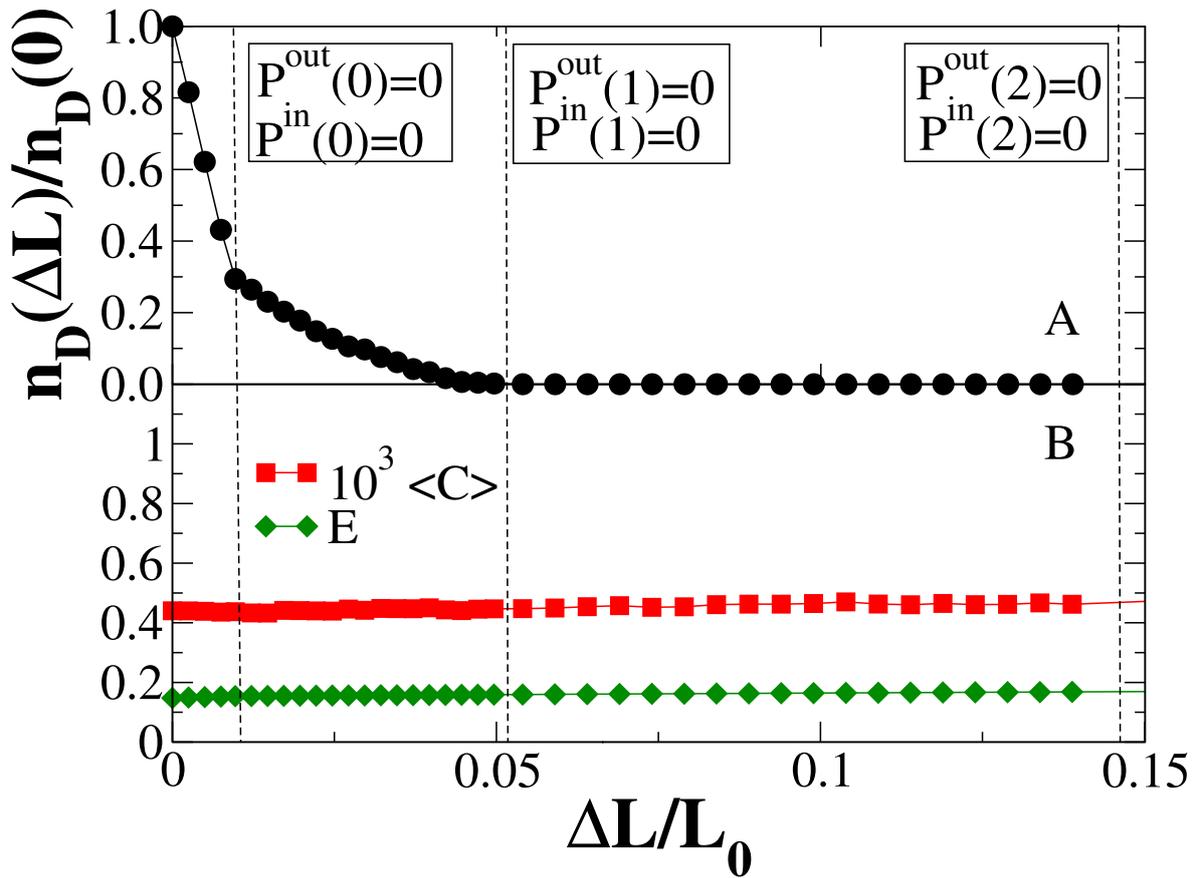


Figure 4.6: Fraction of driver nodes  $n_D(\Delta L)/n_D(0)$  (panel A) average clustering coefficient  $\langle C \rangle$  and efficiency  $E$  (panel B) of the network as a function of the fraction of added links to low degree nodes. The results are obtained solving the MS equations with the Belief Propagation algorithm. The initial network is a Poisson network with in-degree distribution equal to out degree distribution,  $N = 10^4$  nodes, and average degree  $c = 4$ . The symbol  $\Delta L$  indicates the number of added links to the network, whereas  $L_0$  indicates the initial number of links of the network. The links are added to low degree nodes in the following way. First links are added to nodes of in-degree and out-degree 0, then links are added to nodes of in-degree and out-degree 1, and finally to nodes of in-degree and out-degree 2. This strategy can be used to increase the controllability of networks.

of Poisson networks. In fact this approach can be applied to networks with any type of degree distribution. In Figure 4.6 we display the fraction  $n_D(\Delta L)$  of driver nodes when we add  $\Delta L$  links in the network divided by its initial value  $n_D(0)$  where the network has a Poisson degree distribution and average degree

$c = 4$ . We note that in this case the fraction of links that need to be added to have full controllability is of the order of 5%. Here we have chosen to display the efficiency  $E$  instead of the average distance  $\langle l \rangle$  because the network, specially at the beginning, is not fully connected.

When  $P^{in}(1) = P^{out}(1) = 0$  the displayed network has  $P^{in}(2) = P^{out}(2) \approx 0.21$  and it becomes fully controllable.

## 4.4 Controllability of multiplex networks

### 4.4.1 Stability condition

We compute here the Jacobian of the system of Eqs. (3.95) and impose that all its eigenvalues have modulus less than one. We avoid to consider  $w_3^\alpha$  and  $\hat{w}_3^\alpha$  because they influence only the number of null eigenvalues (4 eigenvalues upon 12). The  $12 \times 12$  Jacobian matrix becomes  $8 \times 8$  and it reads

$$J = \begin{pmatrix} 0 & 0 & 0 & G_1^{[1],out,'}(w_2^{[1]}) & \\ 0 & 0 & G_1^{[1],out,'}(1 - w_1^{[1]}) & 0 & \\ 0 & G_1^{[1],in,'}(w_2^{[1]})(1 - G_0^{[2],in}(1 - w_1^{[2]})) & 0 & 0 & \\ G_1^{[1],in,'}(1 - w_1^{[1]})(1 - G_0^{[2],in}(w_2^{[2]})) & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \\ G_1^{[2],in}(w_2^{[2]}) \langle k \rangle_{[1],in} G_1^{[1],in}(1 - w_1^{[1]}) & 0 & 0 & 0 & \\ 0 & G_1^{[2],in}(1 - w_1^{[2]}) \langle k \rangle_{[1],in} G_1^{[1],in}(w_2^{[1]}) & 0 & 0 & \end{pmatrix}$$

$$\begin{array}{cccc}
 & 0 & 0 & 0 & 0 \\
 & 0 & 0 & 0 & 0 \\
 G_1^{[1],in}(w_2^{[1]}) \langle k \rangle_{[2],in} & G_1^{[2],in}(1-w_1^{[2]}) & 0 & 0 & 0 \\
 \dots & 0 & G_1^{[1],in}(1-w_1^{[1]}) \langle k \rangle_{[2],in} & G_1^{[2],in}(w_2^{[2]}) & 0 \\
 & 0 & 0 & 0 & G_1^{[2],out,'}(\hat{w}_2^{[2]}) \\
 & 0 & 0 & G_1^{[2],out,'}(1-\hat{w}_1^{[2]}) & 0 \\
 & 0 & G_1^{[2],in,'}(w_2^{[2]})(1-G_0^{[1],in}(1-w_1^{[1]})) & 0 & 0 \\
 G_1^{[2],in,'}(1-w_1^{[2]})(1-G_0^{[1],in}(w_2^{[1]})) & 0 & 0 & 0 & 0
 \end{array} \Bigg)$$

where, in similar way to Sec 4.3.1, we define

$$\begin{aligned}
 G_0^{\alpha,in}(x) &= \sum_k P_\alpha^{in}(k) x^k \\
 G_0^{\alpha,out}(x) &= \sum_k P_\alpha^{out}(k) x^k \\
 G_1^{\alpha,in}(x) &= \sum_k \frac{k}{\langle k^\alpha \rangle_{in}} P_\alpha^{in}(k) x^{k-1} \\
 G_1^{\alpha,in,'}(x) &= \sum_k \frac{k(k-1)}{\langle k^\alpha \rangle_{in}} P_\alpha^{in}(k) x^{k-2} \\
 G_1^{\alpha,out}(x) &= \sum_k \frac{k}{\langle k^\alpha \rangle_{out}} P_\alpha^{out}(k) x^{k-1} \\
 G_1^{\alpha,out,'}(x) &= \sum_k \frac{k(k-1)}{\langle k^\alpha \rangle_{out}} P_\alpha^{out}(k) x^{k-2}, \tag{4.57}
 \end{aligned}$$

Considering in particular the solution  $w_1^\alpha = \hat{w}_1^\alpha = w_2^\alpha = \hat{w}_2^\alpha = 0$  and  $w_3^\alpha = \hat{w}_3^\alpha =$

1, emerging for  $P_\alpha^{in}(1) = P_\alpha^{out}(1) = 0$  the Jacobian matrix changes in

$$J = \begin{pmatrix} 0 & 0 & 0 & \frac{2P_{[1]}^{out}(2)}{\langle k^{[1]} \rangle_{out}} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\langle k^{[1]}(k^{[1]}-1) \rangle_{out}}{\langle k^{[1]} \rangle_{out}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{\langle k^{[1]}(k^{[1]}-1) \rangle_{in}}{\langle k^{[1]} \rangle_{in}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{2P_{[2]}^{out}(2)}{\langle k^{[2]} \rangle_{out}} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{\langle k^{[2]}(k^{[2]}-1) \rangle_{out}}{\langle k^{[2]} \rangle_{out}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{\langle k^{[2]}(k^{[2]}-1) \rangle_{in}}{\langle k^{[2]} \rangle_{in}} & 0 & 0 & 0 \end{pmatrix} \quad (4.58)$$

Four eigenvalues are zero, the other four have degenerate modulus, therefore the stability conditions are

$$\begin{aligned} 2 \frac{\langle k^{[1]}(k^{[1]}-1) \rangle_{in}}{\langle k^{[1]} \rangle_{in}} \frac{P_{[1]}^{out}(2)}{\langle k^{[1]} \rangle_{out}} &< 1 \\ 2 \frac{\langle k^{[2]}(k^{[2]}-1) \rangle_{in}}{\langle k^{[2]} \rangle_{in}} \frac{P_{[2]}^{out}(2)}{\langle k^{[2]} \rangle_{out}} &< 1. \end{aligned} \quad (4.59)$$

When  $P_{[1]}^{in}(k) = P_{[1]}^{out}(k) = P_{[2]}^{in}(k) = P_{[2]}^{out}(k)$  we have just one stability criterion for this particular solution and it reads

$$P(2) < \frac{\langle k \rangle^2}{2 \langle k(k-1) \rangle} \quad (4.60)$$

#### 4.4.2 Correlations

We give here the calculations for the density of driver nodes when degree correlations between the two layer are introduced (see Sec. 3.6.7).

## Low-degree correlation

We introduce in Eq. 3.92 and Eq. 3.93 the joint probability

$$P^{in}(k^{[1]}, k^{[2]}) = \begin{cases} p\delta_{k^{[2]}, k^{[1]}}P(k^{[1]}) + (1-p)P(k^{[1]})P(k^{[2]}), & \text{if } k^{[1]} \leq 2 \\ (1-p)P(k^{[1]})P(k^{[2]}), & \text{if } k^{[1]} > 2 \quad k^{[2]} \leq 2 \\ p\frac{P(k^{[2]})}{C}P(k^{[1]}) + (1-p)P(k^{[1]})P(k^{[2]}), & \text{if } k^{[1]} > 2 \quad k^{[2]} > 2 \end{cases}$$

where  $C = 1 - \sum_{k \leq 2} P(k)$ . For simplicity's sake we consider  $P(k^{[1]}) = P(k^{[2]}) = P(k)$ . The probability  $p$  modulates the strength of the correlation. The modified equations in 3.92 become

$$\begin{aligned} \hat{w}_1 &= p \left[ \frac{P(1)}{\langle k \rangle} w_1 + \frac{2P(2)}{\langle k \rangle} w_2 (1 - (1 - w_1)^2) + (G_1(w_2) - \frac{P(1)}{\langle k \rangle} \right. \\ &\quad \left. - \frac{2P(2)}{\langle k \rangle} w_2) (1 - G_0^t(1 - w_1)) \right] + (1-p)G_1(w_2) [1 - G_0(1 - w_1)] \\ \hat{w}_2 &= p \left[ \frac{P(1)}{\langle k \rangle} w_2 + \frac{2P(2)}{\langle k \rangle} (w_1 + w_2^2(1 - w_1)) + 1 - \frac{P(1)}{\langle k \rangle} - \frac{2P(2)}{\langle k \rangle} \right. \\ &\quad \left. - (G_1(1 - w_1) - \frac{P(1)}{\langle k \rangle} - \frac{2P(2)}{\langle k \rangle} (1 - w_1)) (1 - G_0^t(w_2)) \right] \\ &\quad + (1-p) [1 - G_1(1 - w_1) + G_1(1 - w_1)G_0(w_2)] \end{aligned} \quad (4.61)$$

where  $G_0^t(x) = \sum_{k \geq 3} \frac{P(k)}{C} x^k$ .

Finally, Eq. 3.93 is modified in

$$\begin{aligned} n_D &= 2 \{ G_0(\hat{w}_2) - [1 - G_0(1 - \hat{w}_1)] \} + 2\langle k \rangle [\hat{w}_1(1 - w_2) + w_1(1 - \hat{w}_2)] \\ &\quad - 2(1-p) \{ [1 - G_0(1 - w_1)][1 - G_0(w_2)] \} \\ &\quad - 2p \{ P(1)w_1(1 - w_2) + P(2)(1 - (1 - w_1)^2)(1 - w_2^2) + \\ &\quad \quad C(1 - G_0^t(1 - w_1))(1 - G_0^t(w_2)) \} \end{aligned} \quad (4.62)$$

## Total correlation

We introduce in Eq. 3.92 and Eq. 3.93 the joint probability

$P^{in}(k^{[1]}, k^{[2]}) = p\delta_{k^{[2]}, k^{[1]}}P(k^{[1]}) + (1 - p)P(k^{[1]})P(k^{[2]})$  For simplicity's sake we consider  $P(k^{[1]}) = P(k^{[2]}) = P(k)$ . The probability  $p$  modulates the strength of the correlation. The modified equations in 3.92 becomes

$$\begin{aligned}
 \hat{w}_1 &= p[G_1(w_2) - (1 - w_1)G_1(w_2(1 - w_1))] \\
 &\quad + (1 - p)G_1(w_2)[1 - G_0(1 - w_1)] \\
 \hat{w}_2 &= p[1 - G_1(1 - w_1) + w_2G_1(w_2(1 - w_1))] \\
 &\quad + (1 - p)[1 - G_1(1 - w_1) + G_1(1 - w_1)G_0(w_2)] \tag{4.63}
 \end{aligned}$$

Finally, Eq. 3.93 is modified in

$$\begin{aligned}
 n_D &= 2\{G_0(\hat{w}_2) - [1 - G_0(1 - \hat{w}_1)]\} + 2\langle k \rangle [\hat{w}_1(1 - w_2) + w_1(1 - \hat{w}_2)] \\
 &\quad - 2(1 - p)\{[1 - G_0(1 - w_1)][1 - G_0(w_2)]\} \\
 &\quad - 2p\{1 - G_0(1 - w_1) - G_0(w_2) + G_0(w_2(1 - w_1))\} \tag{4.64}
 \end{aligned}$$

# Bibliography

- G. Menichetti, D. Remondini, Entropy of a network ensemble: definitions and applications to genomic data, *Theoretical Biology Forum* 1 (2015) 107.
- G. Menichetti, D. Remondini, P. Panzarasa, R. J. Mondragón, G. Bianconi, Weighted multiplex networks., *PloS one* 9 (2014a) e97857.
- G. Menichetti, D. Remondini, G. Bianconi, Correlations between weights and overlap in ensembles of weighted multiplex networks, *Phys. Rev. E* 90 (2014b) 062817.
- G. Menichetti, L. Dall'Asta, G. Bianconi, Network controllability is determined by the density of low in-degree and out-degree nodes, *Phys. Rev. Lett.* 113 (2014c) 078701.
- J. Park, M. E. J. Newman, Statistical mechanics of networks, *Physical Review E* 70 (2004) 066117.
- G. Bianconi, Entropy of network ensembles, *Phys. Rev. E* 79 (2009) 036114.
- T. Squartini, D. Garlaschelli, Analytical maximum-likelihood method to detect patterns in real networks, *New Journal of Physics* 13 (2011) 83001.

- A. K. Hartmann, M. Weigt, Phase Transitions in Combinatorial Optimization Problems, Weinheim, Wiley-VCH edition, 2005.
- A. Greven, G. Keller, G. Warnecke, Entropy, Princeton series in applied mathematics, Princeton University Press, 2003.
- E. T. Jaynes, Information theory and statistical mechanics, Phys. Rev. 106 (1957) 620–630.
- T. Cover, J. Thomas, Elements of Information Theory, Wiley, 2006.
- D. Garlaschelli, M. I. Loffredo, Multispecies grand-canonical models for networks with reciprocity, Phys. Rev. E 73 (2006) 015101.
- K. Anand, G. Bianconi, Entropy measures for networks: Toward an information theory of complex topologies, Phys. Rev. E 80 (2009) 45102.
- K. Anand, G. Bianconi, Gibbs entropy of network ensembles by cavity methods, Phys. Rev. E 82 (2010) 011116.
- M. Pagel, A. Pomiankowski, Evolutionary genomics and proteomics, Sinauer Associates, 2008.
- J. De Las Rivas, C. Fontanillo, Protein–protein interactions essentials: Key concepts to building and analyzing interactome networks, PLoS Computational Biology 6 (2010) e1000807.
- A.-L. Barabasi, Z. N. Oltvai, Network biology: understanding the cell’s functional organization, Nat Rev Genet 5 (2004).
- E. Alm, A. Arkin, Biological networks, Curr Opin Struct Biol 13 (2003) 193–202.

- M. Vidal, M. Cusick, A. Barabasi, Interactome networks and human disease, *Cell* 144 (2011) 989–98.
- E. Cerami, B. Gross, E. Demir, I. Rodchenkov, O. Babur, N. Anwar, N. Schultz, G. Bader, C. Sander, Pathway commons, a web resource for biological pathway data, *Nucleic Acids Res* 39 (2011) D685–D690.
- D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguetz, T. Doerks, M. Stark, J. Muller, P. Bork, L. Jensen, C. von Mering, The string database in 2011: functional interaction networks of proteins, globally integrated and scored, *Nucl. Acids Res.* 39 (2011) D561–D568.
- A. Teschendorff, S. Severini, Increased entropy of signal transduction in the cancer metastasis phenotype, *BMC Systems Biology* 4 (2010) 104.
- F. Barea, D. Bonatto, Aging defined by a chronologic-replicative protein network in *saccharomyces cerevisiae*: an interactome analysis, *Mech Ageing Dev* 130 (2009) 444–60.
- J. West, G. Bianconi, S. Severini, A. Teschendorff, Differential network entropy reveals cancer system hallmarks, *Scientific Reports* (2012) 10.1038/srep00802.
- W. N. van Wieringen, A. W. van der Vaart, Statistical analysis of the cancer cell's molecular entropy using high-throughput data, *Bioinformatics* 27 (2011) 556–563.
- G. Bianconi, P. Pin, M. Marsili, Assessing the relevance of node features for network structure., *Proceedings of the National Academy of Sciences of the United States of America* 106 (2009) 11433–8.

- B. Györfy, B. Molnar, H. Lage, Z. Szallasi, A. Ecklund, Evaluation of microarray preprocessing algorithms based on concordance with rt-pcr in clinical samples, *PLoS One* 4 (2009) e5645.
- K. Scotlandi, D. Remondini, G. Castellani, M. Manara, F. Nardi, L. Cantiani, M. Francesconi, M. Mercuri, A. Caccuri, M. Serra, M. Knuutila, P. Picci, Overcoming resistance to conventional drugs in ewing sarcoma and identification of molecular predictors of outcome, *J. Clin. Oncol.* 27 (2009) 2209–16.
- C. Sotiriou, P. Wirapati, S. Loi, A. Harris, S. Fox, J. Smeds, H. Nordgren, P. Farmer, V. Praz, B. Haibe-Kains, C. Desmedt, D. Larsimont, F. Cardoso, H. Peterse, D. Nuyten, M. Buyse, M. J. Van de Vijver, J. Bergh, M. Piccart, M. Delorenzi, Gene expression profiling in breast cancer: Understanding the molecular basis of histologic grade to improve prognosis, *Journal of the National Cancer Institute* 98 (15 February 2006) 262–272.
- S. Loi, B. Haibe-Kains, C. Desmedt, F. Lallemand, A. M. Tutt, C. Gillet, P. Ellis, A. Harris, J. Bergh, J. A. Foekens, J. G. Klijn, D. Larsimont, M. Buyse, G. Bontempi, M. Delorenzi, M. J. Piccart, C. Sotiriou, Definition of clinically distinct molecular subtypes in estrogen receptor $\alpha$ -positive breast carcinomas through genomic grade, *Journal of Clinical Oncology* 25 (April 1, 2007) 1239–1246.
- Y. Wang, J. G. Klijn, Y. Zhang, A. M. Sieuwerts, M. P. Look, F. Yang, D. Tantalov, M. Timmermans, M. E. Meijer-van Gelder, J. Yu, T. Jatko, E. M. Berns, D. Atkins, J. A. Foekens, Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer, *Lancet* 365 (2005) 671–679.

- D. Remondini, S. Salvioli, M. Francesconi, M. Pierini, D. J. Mazzatti, J. R. Powell, I. Zironi, F. Bersani, G. Castellani, C. Franceschi, Complex patterns of gene expression in human t cells during in vivo aging, *Mol. BioSyst.* 6 (2010) 1983–1992.
- D. Hanahan, R. Weinberg, The hallmarks of cancer, *Cell* 100 (2000) 57–70.
- S. Brabletz, T. Brabletz, The zeb/mir-200 feedback loop—a motor of cellular plasticity in development and cancer?, *EMBO Reports* 11 (2010) 670–677.
- L. Ferrucci, F. Giallauria, D. Schlessinger, Mapping the road to resilience: Novel math for the study of frailty, *Mechanisms of Ageing and Development* 129 (2008) 677–679.
- L. Fried, L. Ferrucci, J. Darer, J. Williamson, , G. Anderson, Untangling the concepts of disability, frailty, and comorbidity: Implications for improved targeting and care, *Journal of Gerontology* 59 (2004) 255–263.
- E. Bellavista, M. Martucci, F. Vasuri, A. Santoro, M. Mishto, A. Kloss, E. Capizzi, A. Degiovanni, C. Lanzarini, D. Remondini, A. Dazzi, S. Pellegrini, M. Cescon, M. Capri, S. Salvioli, A. D’Errico-Grigionic, B. Dahlmannd, G. Grazi, C. Franceschi, Lifelong maintenance of composition, function and cellular/ subcellular distribution of proteasomes in human liver, *Mechanisms of Ageing and Development* 141-142 (2014) 26–34.
- H. Lempiainen, D. Shore, Growth control and ribosome biogenesis, *Current Opinion in Cell Biology* 21 (2009) 855–863.
- R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, *Reviews of Modern Physics* 74 (2002) 47–97.

- M. E. J. Newman, The structure and function of complex networks (2003) 58.
- S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: Structure and dynamics, *Physics reports* 424 (2006) 175–308.
- S. Fortunato, Community detection in graphs, *Physics Reports* 486 (2010) 75–174.
- M. Szell, R. Lambiotte, S. Thurner, Multirelational organization of large-scale social networks in an online world., *Proceedings of the National Academy of Sciences of the United States of America* 107 (2010) 13636–41.
- A. Cardillo, J. Gómez-Gardeñes, M. Zanin, M. Romance, D. Papo, F. del Pozo, S. Boccaletti, Emergence of network features from multiplexity., *Scientific reports* 3 (2013) 1344.
- V. Nicosia, V. Latora, Measuring and modelling correlations in multiplex networks (2014) 21.
- J. F. Donges, Y. Zou, N. Marwan, J. Kurths, Complex networks in climate dynamics, *The European Physical Journal Special Topics* 174 (2009) 157–179.
- J. Gao, S. V. Buldyrev, H. E. Stanley, S. Havlin, Networks formed from interdependent networks, *Nature physics* 8 (2012) 40–48.
- G. Bianconi, S. N. Dorogovtsev, Multiple percolation transitions in a configuration model of a network of networks, *Physical Review E* 89 (2014) 062814.
- E. Bullmore, O. Sporns, Complex brain networks: graph theoretical analysis of structural and functional systems., *Nature reviews. Neuroscience* 10 (2009) 186–98.

- G. Castellani, N. Intrator, D. Remondini, Systems biology and brain activity in neuronal pathways by smart device and advanced signal processing., *Frontiers in genetics* 5 (2014) 253.
- R. G. Morris, M. Barthelemy, Transport on Coupled Spatial Networks, *Physical Review Letters* 109 (2012) 128703.
- F. Battiston, V. Nicosia, V. Latora, Structural measures for multiplex networks, *Physical Review E* 89 (2014) 032804.
- A. Halu, R. J. Mondragón, P. Panzarasa, G. Bianconi, Multiplex PageRank., *PloS one* 8 (2013) e78293.
- P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, J.-P. Onnela, Community structure in time-dependent, multiscale, and multiplex networks., *Science (New York, N.Y.)* 328 (2010) 876–8.
- S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, S. Havlin, Catastrophic cascade of failures in interdependent networks., *Nature* 464 (2010) 1025–8.
- G. J. Baxter, S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes, Avalanche Collapse of Interdependent Networks, *Physical Review Letters* 109 (2012) 248701.
- S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, A. Arenas, Diffusion Dynamics on Multiplex Networks, *Physical Review Letters* 110 (2013) 028701.
- C. D. Brummitt, R. M. D’Souza, E. A. Leicht, Suppressing cascades of load in

- interdependent networks., Proceedings of the National Academy of Sciences of the United States of America 109 (2012) E680–9.
- B. Min, S. D. Yi, K.-M. Lee, K.-I. Goh, Network robustness of multiplex networks with interlayer degree correlations, Phys. Rev. E 89, 042811 2014 (2014).
- V. Nicosia, G. Bianconi, V. Latora, M. Barthelemy, Nonlinear growth and condensation in multiplex networks, Physical Review E 90 (2014) 042807.
- G. Bianconi, Statistical Mechanics of Multiplex Ensembles: Entropy and Overlap, Physical Review E 87 (2013) 062806.
- A. Halu, S. Mukherjee, G. Bianconi, Emergence of overlap in ensembles of spatial multiplexes and statistical mechanics of spatial interacting network ensembles, Physical Review E 89 (2014) 12806.
- G. Bianconi, The entropy of randomized network ensembles, EPL (Europhysics Letters) 81 (2008) 28005.
- G. Bianconi, A. C. C. Coolen, C. J. Perez Vicente, Entropies of complex networks with hierarchically constrained topologies., Physical review. E, Statistical, nonlinear, and soft matter physics 78 (2008) 016114.
- A. Annibale, A. C. C. Coolen, L. P. Fernandes, F. Fraternali, J. Kleinjung, Tailored graph ensembles as proxies or null models for real networks I: tools for quantifying structure, Journal of Physics A: Mathematical and Theoretical 42 (2009) 485001.
- T. Squartini, G. Fagiolo, D. Garlaschelli, Randomizing world trade. II. A weighted network analysis, Physical Review E 84 (2011) 46118.

- D. Garlaschelli, The weighted random graph model, *New Journal of Physics* 11 (2009) 73005.
- D. Garlaschelli, M. I. Loffredo, Generalized bose-fermi statistics and structural correlations in weighted networks, *Physical review letters* 102 (2009) 38701.
- O. Sagarra, C. J. Pérez-Vicente, A. Díaz-Guilera, Statistical mechanics of multi-edge networks, *Physical Review E* 88 (2013) 062806.
- O. Sagarra, F. Font-Clos, C. J. Pérez-Vicente, A. Díaz-Guilera, The configuration multi-edge model: Assessing the effect of fixing node strengths on weighted network magnitudes, *EPL* 107 (2014).
- C. I. Del Genio, H. Kim, Z. Toroczkai, K. E. Bassler, Efficient and exact sampling of simple graphs with given arbitrary degree sequence, *PloS one* 5 (2010) e10012.
- V. Zlatic, G. Bianconi, A. Díaz-Guilera, D. Garlaschelli, F. Rao, G. Caldarelli, On the rich-club effect in dense and weighted networks, *The European Physical Journal B* 67 (2009) 271–275.
- A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks., *Proceedings of the National Academy of Sciences of the United States of America* 101 (2004) 3747–52.
- E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai, A.-L. Barabási, Global organization of metabolic fluxes in the bacterium *Escherichia coli.*, *Nature* 427 (2004) 839–43.

- M. Granovetter, The strength of weak ties, *American Journal of Sociology* 78 (1973) 1360–1380.
- J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási, Structure and tie strengths in mobile communication networks., *Proceedings of the National Academy of Sciences of the United States of America* 104 (2007) 7332–6.
- M. Karsai, N. Perra, A. Vespignani, Time varying networks and the weakness of strong ties., *Scientific reports* 4 (2014) 4001.
- A. Barrat, M. Barthelemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks., *Proceedings of the National Academy of Sciences of the United States of America* 101 (2004) 3747–52.
- M. Barthelemy, B. Gondran, E. Guichard, Spatial structure of the internet traffic, *Physica A: Statistical Mechanics and its Applications* 319 (2003) 633–642.
- M. A. Serrano, M. Boguñá, A. Vespignani, Extracting the multiscale backbone of complex weighted networks., *Proceedings of the National Academy of Sciences of the United States of America* 106 (2009) 6483–8.
- S. Redner, How popular is your paper? An empirical study of the citation distribution, *The European Physical Journal B* 4 (1998) 131–134.
- M. E. J. Newman, The structure of scientific collaboration networks, *Proceedings of the National Academy of Sciences* 98 (2001) 404–409.
- F. Radicchi, S. Fortunato, B. Markines, A. Vespignani, Diffusion of scientific credits and the ranking of scientists, *Physical Review E* 80 (2009) 056103.

- F. Radicchi, S. Fortunato, C. Castellano, Universality of citation distributions: toward an objective measure of scientific impact., *Proceedings of the National Academy of Sciences of the United States of America* 105 (2008) 17268–72.
- S. Johnson, J. J. Torres, J. Marro, M. A. Muñoz, Entropic Origin of Disassortativity in Complex Networks, *Physical Review Letters* 104 (2010) 108702.
- NCBI, Gene Expression Omnibus, 2014.
- KEGG, Kyoto Encyclopedia of Genes and Genomes, 2014.
- P. Commons, A Resource for Biological Pathway Analysis, 2014.
- P. Csermely, T. Korcsmáros, H. J. M. Kiss, G. London, R. Nussinov, Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review., *Pharmacology & therapeutics* 138 (2013) 333–408.
- P. Bonifazi, M. Goldin, M. A. Picardo, I. Jorquera, A. Cattani, G. Bianconi, A. Represa, Y. Ben-Ari, R. Cossart, GABAergic hub neurons orchestrate synchrony in developing hippocampal networks, *Science* 326 (2009) 1419.
- J. D. Power, A. L. Cohen, S. M. Nelson, G. S. Wig, K. A. Barnes, J. A. Church, A. C. Vogel, T. O. Laumann, F. M. Miezin, B. L. Schlaggar, S. E. Petersen, Functional network organization of the human brain., *Neuron* 72 (2011) 665–78.
- D. Delpini, S. Battiston, M. Riccaboni, G. Gabbi, F. Pammolli, G. Caldarelli, Evolution of controllability in interbank networks., *Scientific reports* 3 (2013) 1626.

- J. Ruths, D. Ruths, Control profiles of complex networks, *Science* 343 (2014) 1373–1376.
- Y.-Y. Liu, J.-J. Slotine, A.-L. Barabási, Controllability of complex networks., *Nature* 473 (2011) 167–73.
- Y.-Y. Liu, E. Csóka, H. Zhou, M. Pósfai, Core percolation on complex networks, *Phys. Rev. Lett.* 109 (2012) 205703.
- P. A. Iglesias, B. P. Ingalls, *Control Theory and Systems Biology*, Boston, the mit pr edition, 2009.
- C. T. Lin, Structural controllability, *IEEE Trans. Auto. Contr.* 19 (1974) 201.
- J. E. Hopcroft, M. Karp, An  $n^5/2$  algorithm for maximum matchings in bipartite graphs, *SIAM Journal of Computing* 2 (1973) 225–231.
- S. Maslov, Degree-preserving random rewiring and correlation profile of a complex network, 2014.
- O. Zagordi, *STATISTICAL PHYSICS METHODS IN COMPUTATIONAL BIOLOGY*, Ph.D. thesis, SISSA, 2007.
- H. Nishimori, G. Ortiz, *Elements of Phase Transitions and Critical Phenomena*, Oxford Graduate Texts, OUP Oxford, 2010.
- J. S. Yedidia, W. T. Freeman, Y. Weiss, *Understanding Belief Propagation and its Generalizations* (2001).
- L. Zdeborová, M. Mézard, The number of matchings in random graphs, *Journal of Statistical Mechanics: Theory and Experiment* 2006 (2006) P05003.

- F. Altarelli, A. Braunstein, A. Ramezanzpour, R. Zecchina, Stochastic matching problem, *Physical review letters* 106 (2011) 190601.
- M. Mézard, G. Parisi, The Bethe lattice spin glass revisited, *The European Physical Journal B-Condensed Matter and Complex Systems* 20 (2001) 217–233.
- O. C. Martin, R. Monasson, R. Zecchina, Statistical mechanics methods and phase transitions in optimization problems, *Theoretical computer science* 265 (2001) 3–67.
- M. Boguñá, R. Pastor-Satorras, A. Vespignani, Cut-offs and finite size effects in scale-free networks, *The European Physical Journal B-Condensed Matter and Complex Systems* 38 (2004) 205–209.
- H. Seyed-Allaei, G. Bianconi, M. Marsili, Scale-free networks with an exponent less than two, *Physical Review E* 73 (2006) 46113.
- S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña Nadal, Z. Wang, M. Zanin, The structure and dynamics of multilayer networks, *Physics Reports* 544 (2014) 1–122.
- MathWorks, Multiple linear regression - MATLAB regress, 2015.
- A. Montanari, F. Ricci-Tersenghi, On the nature of the low-temperature phase in discontinuous mean-field spin glasses, *The European Physical Journal B-Condensed Matter and Complex Systems* 33 (2003) 339–346.
- O. Rivoire, G. Biroli, O. C. Martin, M. Mézard, Glass models on Bethe lattices,

The European Physical Journal B-Condensed Matter and Complex Systems  
37 (2003) 55–78.

T. Castellani, F. Krzakala, F. Ricci-Tersenghi, Spin glass models with ferromagnetically biased couplings on the Bethe lattice: analytic solutions and numerical simulations, The European Physical Journal B-Condensed Matter and Complex Systems 47 (2005) 99–108.

C. Lucibello, F. Ricci-Tersenghi, The statistical mechanics of random set packing and a generalization of the Karp-Sipser algorithm, International Journal of Statistical Mechanics 2014 (2014).