Universiteit
Leiden
The Netherlands

On metrics and models for multiplex networks

Gemmetto, V.

Citation
Gemmetto, V. (2018, January 16). *On metrics and models for multiplex networks*. Retrieved from https://hdl.handle.net/1887/61132

Version:	Not Applicable (or Unknown)
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/61132

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The following handle holds various files of this Leiden University dissertation:
http://hdl.handle.net/1887/61132

Author: Gemmetto, V.
Title: On metrics and models for multiplex networks
Issue Date: 2018-01-16

Chapter 3

Multiplex network reconstruction

The characterization of various properties of real-world systems requires the knowledge of the underlying network of connections among the system's components. Unfortunately, in many situations the complete topology of this network is empirically inaccessible, and one has to resort to probabilistic techniques to infer it from limited information. While network reconstruction methods have reached some degree of maturity in the case of single-layer networks (where nodes can be connected only by one type of links), the problem is practically unexplored in the case of multiplex networks, where several interdependent layers, each with a different type of links, coexist. Even the most advanced network reconstruction techniques, if applied to each layer separately, fail in replicating the observed inter-layer dependencies making up the whole coupled multiplex. Here we develop a methodology to reconstruct a class of correlated multiplexes which includes the World Trade Multiplex as a specific example we study in detail. Our method starts from any reconstruction model that successfully reproduces some desired marginal properties, including node strengths and/or node degrees, of each layer separately. It then introduces the minimal dependency structure required to replicate an additional set of higher-order properties that quantify the portion of each node's degree and each node's strength that is shared and/or reciprocated across pairs of layers. These properties are found to provide empirically robust measures of inter-layer coupling. Our method allows joint multi-layer connection probabilities to be reliably reconstructed from marginal ones, effectively bridging the gap between single-layer properties and truly multiplex information.

The results presented in this chapter have been published in the following reference:
V. Gemmetto, D. Garlaschelli, *arXiv*:1709.03918 (2017).

3.1 Introduction

In the last twenty years, the study of complex networks acquired importance as it could significantly increase our understanding of many real-world systems [1, 2, 3], ranging from the global airport infrastructure [4] to biological systems like the brain [5]. Indeed, it is easy to realize that several systems, including the ones just mentioned, share a common abstract representation in terms of nodes connected by links, i.e. in terms of graphs or networks.

However, a more careful analysis shows that a simple network representation is often not enough to fully capture the whole complexity of the aforementioned systems [6]. For instance, the presence of different airline companies significantly affects the air transportation landscape [7, 8]. Similarly, the human body can be thought of as a set of interdependent networks where several complex physiological systems, e.g. the nervous and the cardiovascular ones, constantly interact [9].

For this reason, the concepts of multiplex and interdependent networks have been developed. In a multiplex network, a given set of nodes is connected through different modes of interactions; the system is therefore represented as a coloured-edge or layered graph [10], where each layer contains the same set of "replica nodes". Interdependent networks are instead composed of two or more interconnected networks, where each node of any graph is dependent on one or more nodes belonging to the other(s) [6].

Several studies have focused on the analysis of structural aspects of these multi-graphs [11, 12, 13]. In particular, the analysis of the overlap between layers of a multiplex network can provide valuable information in order to better understand some dynamical processes that occur on top of those systems [14, 15] or possible failure cascades [6]. Moreover, the presence of dipendencies between layers crucially affects the systemic risk associated to these networks, for instance in the case of financial or economic systems [16, 17]. It must be pointed out that, in order to study the aforementioned dynamical processes, the full graph structure is required, even in the case of monoplex networks. In general, however, confidentiality issues or limitedness of the topological information may not allow the knowledge of the entire network, but only that of partial information about the nodes (for instance, the degrees of all or some of the vertices, or the strengths and the density). Various *network reconstruction* methods have therefore been developed, in order to successfully infer the full topological structure of graphs starting from incomplete information [18, 19, 20, 21, 22, 23, 24, 25]. Unfortunately, the current methodologies are applicable only to single-layer networks, leaving an important gap open in the study of multiplex networks. If these techniques were applied to each layer of a multiplex separately, they would by construction fail in replicating the empirical coupling between layers.

Our main goal in the present chapter is that of developing a satistifactory methodology for the reconstruction of multiplex networks from partial information. Our approach is guided by the following consideration. Clearly, a single-layer network can be seen as a particularly simple case of a multiplex, i.e. in the limit

when the number of layers is one. Then, from an entirely general point of view, a method to reconstruct multiplex networks may fail as a result of (a combination of) two factors. On one hand, the method may be unsuccessful because the properties of (some of) the layers are incorrectly reconstructed. This may be due to the method failing on each layer separately, a circumstance that strongly indicates an intrinsic unreliability of the reconstruction model itself, even when applied in the single-layer limit. On the other hand, the method may succeed in replicating the marginal properties of each layer separately, while it may fail in replicating the interdependencies among layers. In the former case we do not learn anything useful about whether and how the method can be improved. By contrast, the latter situation is quite informative, as it indicates that, if the reconstruction model could be generalized in such a way that its marginal single-layer properties are maintained, while at the same time its inter-layer ones are made more realistic, then it would become an acceptable method for reconstructing multiplexes with coupled layers.

Following the above reasoning, we put ourselves in the latter situation and assume that the empirical multiplex is taken from a class of multiplex networks for which a 'marginal' method capable of reliably reconstructing each layer separately exists. Then, we investigate how a generalized and coupled multiplex method with the same marginal properties can be constructed. Building on the recent literature on single-layer network reconstruction methods, we select the World Trade Multiplex (WTM) as the ideal empirical candidate for our analysis. The nodes of this multiplex are countries of the world, whereas links represent trade relationships, disaggregated into different commodities. Each commodity gives rise to a separate layer. The links in each layer are in principle directed (from the exporter to the importer) and weighted (by the dollar value of the trade relationship), even though they are often projected into undirected and/or unweighted ones. The empirical properties of the WTM have been studied extensively [26, 27, 28, 29, 30, 31]. If all the commodities are aggregated together, one obtains a single-layer projection documenting the total trade fluxes among countries [27, 28, 32, 33]. In the representation considered here, we use data from Ref. [34, 35] reporting $N = 207$ countries trading in $M = 96$ different commodities, each representing a given layer of the multiplex.

The WTM fulfills our criterion stated above, because it has been shown that each of its layers is very closely replicated by a model that takes only local node information as input. Indeed, the purely binary structure of each layer of the WTM can be replicated starting from the knowledge of the degree of each node in that layer [27] (Binary Configuration Model [36, 37]), while the weighted structure can be successfully replicated from the knowledge of both the strength and the degree of each node in that layer [29] (Enhanced Configuration Model [22, 37]). More relaxed reconstruction models [23, 24, 25], which are discussed later in the paper, have also been shown to successfully replicate the properties of the World Trade network. At the same time, it has been shown that the knowledge of the strength and degree of each node in each layer is not enough to replicate the coupling be-

tween layers [30, 31], illustrating that even if the marginal reconstruction method is successful in each and every layer separately, it fails in replicating the multiplex as a whole.

Our strategy in this chapter is that of devising a way to preserve the good marginal properties of single-layer reconstruction methods, while at the same time introducing a minimal but effective coupling such that, additionally, various robust inter-layer properties of the multiplex are also replicated. The structure of the chapter is as follows. In sec. 3.2 we introduce some preliminary concepts that constrain the range of possible multiplex reconstruction models. In sec. 3.3 we focus on the case of binary multiplexes (both undirected and directed) and develop a multiplex reconstruction method in that case. In sec. 3.4 we move on to weighted multiplexes (again, both undirected and directed) and develop the weighted counterpart of the reconstruction method. Finally, in sec. 4.6 we make some concluding remarks.

3.2 Preliminaries

This section establishes some useful criteria which constrain the features of the multiplex reconstruction model we are after.

3.2.1 Beyond inter-layer degree correlations

To reliably reconstruct a multiplex, we need to identify useful target properties that accurately capture the inter-layer coupling. Various notions of inter-layer overlap have been developed in the literature, for instance in terms of *correlation of layer activity* [11] and *overlapping degree* [12]. In single-layer networks, degree correlation is usually computed by looking at the average degree of the first neighbours of a node having a certain degree (*average nearest neighbour degree*). In the same spirit, notions of multiplex assortativity or inter-layer degree correlation have been developed [11, 38, 39]. The inter-layer degree correlation function has been defined as:

$$\overline{k}^{\alpha}\left(k^{\beta}\right) = \sum_{k^{\alpha}} k^{\alpha} P\left(k^{\alpha}|k^{\beta}\right) \tag{3.1}$$

where $P\left(k^{\alpha}|k^{\beta}\right)$ is the probability that a node having a given degree k^{β} in layer β has degree k^{α} in layer α.

We have shown in the previous chapters that the above quantity is unfortunately not informative about the component of inter-layer coupling that is not due to the degree distribution of the various layers [30]. For instance, if the same node is a hub in multiple layers (a property that gives rise to positive inter-layer assortativity), it will automatically produce a significant overlap of links across these layers, even if links in different layers are drawn completely independently. Such an overlap should therefore not be taken as a genuine measure of

statistical dependency across layers. This spurious effect increases with increasing intra-layer density and increasing heterogeneity of local node properties like degrees and strengths. In order to detect 'true' inter-layer dependencies that are not merely explained by chance, density, or by the local properties of individual nodes, one can construct maximum-entropy null models of multiplexes with independent layers and given node properties [36, 40, 41]. In these null models, in each layer every node has - on average - the same degree (for binary networks), or strength (for weighted networks), that it has in the real multiplex [30]. Apart from these constraints, the maximum-entropy multiplex ensemble is completely random and no dependency is introduced among layers. The expectation values of the multiplexity over the null ensemble can be calculated exactly and used to filter out the undesired effects from the measured values. In the first chapter, we have therefore defined new metrics that quantify the intensity of coupling among layers of an undirected multiplex network, introducing the concept of *multiplexity* [30]. We have used these metrics to extensively document the empirical properties of real-world systems such as the World Trade Multiplex (WTM) [26, 29, 30] and the European Airport Multiplex [7]. We concluded that much of the apparent multiplexity observed among the layers is actually explained by the local properties of nodes. Still, we found a significant level of measured remaining overlap, which quantifies the residual, 'genuine' multiplexity structure of the WTM.

Whenever it is important to take into account the directionality of the connections in a graph [42], the aforementioned approach can be extended to directed multi-layer networks [31]. We found that, in the directed case, the inter-layer 'link overlap' can manifest itself in terms of both the 'alignment' (a phenomenon that we called *multiplexity* in analogy with the undirected case [30]) and the 'anti-alignment' (a phenomenon that we called *multireciprocity* as a generalization of the ordinary reciprocity for single-layer networks [43, 44]) of links across layers. Since in each layer links are allowed in both directions between any two nodes, the alignment and the anti-alignment of links across layers do not conflict with each other and can actually coexist.

3.2.2 A multiplex model with dyadic independence

Our aim is that of introducing a minimal but realistic multiplex model that can reproduce the observed inter-layer dependencies reported above. Unlike the null models considered therein, the multiplex model should be characterized by non-trivial joint probabilities of connection involving multiple layers. We want to develop one such model for binary multiplexes, and one for weighted multiplexes, in both the undirected and directed case.

To keep the model as simple as possible, we assume *dyadic independence*: the presence (and weight) of a link connecting a pair of nodes in a given layer does not depend on the presence (and weight) of a link connecting a *different* pair of nodes in the same or in any other layer, although it does depend on the presence (and weight) of the links connecting the same pair of nodes in other layers. If

we introduce the term *multidyad* to denote a single pair of nodes 'replicated' over all layers of the multiplex (i.e. the set of all single-layer dyads involving the same two nodes), the above assumption might be referred to as *multidyadic independence*. Note that, in directed and binary single-layer networks, a dyad formed by two nodes i and j can have 4 different topologies (a single link from i to j, a single link from i to j, two reciprocal links between i and j, or no link at all). This implies that, in a directed binary multiplex with M layers, a multidyad can have 4^M possible topologies. In a directed and weighted single-layer network, even assuming that the weights are non-negative integer numbers (as often done in previous approaches), a dyad can already have an infinity of possible weight-dependent configurations. Correspondingly, a multidyad in a multiplex with M layers would have an infinite number, 'raised to the Mth power', of configurations. Analogously, similar considerations hold for the undirected case, with the only difference that a dyad in a single-layer unweighted graph can now have 2 possible distinct values (a link between i and j, or no link at all).

The assumption of multidyadic independence only restricts the topological properties that individual layers can have, but does not restrict the range of possible dependencies among layers of the multiplex. Moreover, many single-layer networks have been in fact shown to have a structure consistent with dyadic independence [27, 28, 29, 36]. This property is also confirmed by the success of network reconstruction techniques that, as the one we will introduce here, assume dyadic independence [20, 21, 45, 46]. An important example is given precisely by the WTM, whose single-layer structure is largely consistent with dyadic independence [27, 29].

3.3 Binary multiplex model

In this Section, we develop our analytical framework and show the results of the application of such a theoretical model to a real-world system, namely the binarized version of the International Trade Multiplex [30, 35].

Let us consider the marginal - i.e. unconditional on the presence of any other link in any layer - probability that a (possibly directed) link from node i to node j exists in layer α:

$$p_{ij}^{\alpha} \equiv P(a_{ij}^{\alpha} = 1) = \langle a_{ij}^{\alpha} \rangle \tag{3.2}$$

(here and in the rest of the chapter, angular brackets do not denote expected values under a *null* model with *independent* layers – as in the previous chapters – but ensemble averages over a *realistic* multiplex model with *dependent* layers). Due to our assumption of multidyadic independence, the relevant information that is marginalized in the probability p_{ij}^{α} does not involve other pairs of nodes (joint probabilities involving multiple pairs of nodes would in any case factorize into products of marginal probabilities of invididual pairs of nodes), but it does involve other layers.

In other words, p_{ij}^{α} does not contain information about the inter-layer dependencies that we want to model. As such, it can be chosen to be specified by any convenient single-layer network model that satisfactorily reproduces the topological properties of layer α. This marginal model is not actually an essential ingredient of our multiplex model and can be in some sense 'outsourced'. For instance, it can be chosen to be a proper null model: an appropriate choice would be the (undirected or directed) Configuration Model [47], i.e. the ensemble of networks satisfying on average the empirical degree sequence observed in that specific layer α. It has indeed been shown [27, 29, 36] that this model is able to reliably replicate the topological properties of each layer of many real multilayer networks, including the World Trade Web itself [48, 49]. Hence, defining the values p_{ij}^{α} as the link probabilities, for each layer separately, deriving from the Configuration Model is the most straightforward choice. As a byproduct, this choice illustrates that the previously introduced multiplex assortativity metrics (Eq. (3.1)) are not informative about the inter-layer coupling of interest for our analysis, because they are completely reabsorbed into the dyadic probabilities p_{ij}^{α}; hence, these measures simply refer to a different kind of dependency between layers.

We now come to the definition of the true building blocks of our model of multiplexes with dependent layers. Indeed, the assumption that layers are dependent implies that joint probabilities involving the same pairs of nodes but different layers should *not* trivially factorize into products of marginal probabilities of the type p_{ij}^{α}. We therefore need to introduce generic joint probabilities that involve multiple layers. In general, even if we are assuming multidyadic independence, for each pair of nodes we should consider the joint probabilities of all combinations of links across all layers together, i.e. (in the jargon of multiplex networks [41]) the probabilities of all possible *multilinks* involving the same two nodes. As we mentioned, in multiplexes with directed links a multidyad can have 4^M possible topologies, i.e. 4^M possible multilinks. For each pair of nodes, fully specifying the joint connection probabilities across all layers would require the specification of a different probability for each of these multilinks, with the only constraint that the 4^M probabilities sum up to one. This would lead to the definition of $4^M - 1$ probabilities. While this operation is feasible and insightful in the most studied case of a multiplex with two layers only, it becomes increasingly challenging (and decreasingly transparent) as M increases.

By contrast, we want to keep our approach feasible and useful (both from a modelling and from a network reconstruction perspective) even in the case of a very large number of layers, for which our formalism based on multiplexity and multireciprocity matrices fully shows its advantages. Therefore we take the following parsimonious approach. For a given pair of nodes, we start from the definition of two joint (and conditional) probabilities that fully characterize both the multiplexity and the reciprocity properties of a single pair of layers, and then consider the set of such probabilities for all the M^2 pairs of layers (including a layer with itself) of the multiplex. This leads to a set of only $2M^2$ probabilities

defining the directed multiplex model (but still for a single pair of nodes). This set represents the relevant projection (or marginalization) of the full set of $4^M - 1$ multilink probabilities. The quadratic (as opposed to exponential) growth of the number of probabilities with the number of layers makes our approach appealing and manageable. Moreover, we will show that, at least in the empirical case study considered here, the conditional probabilities are approximately independent of the particular pair of nodes, making the information contained in the multiplexity and multireciprocity matrices sufficient in order to fully characterize the dependencies among the layers. Remarkably, this also means that the number of relevant probabilities remains $2M^2$ (equal to the total number of entries in the multiplexity and multireciprocity matrices) independently of the number N of nodes in the multiplex. Similar considerations can be made for the undirected systems.

We recall that, as reported in Chapter 1 and in [30], in the undirected binary case the multiplexity reads:

$$m_b^{\alpha\beta} = \frac{2\sum_i \sum_{j<i} \min\{a_{ij}^\alpha, a_{ij}^\beta\}}{L^\alpha + L^\beta} = \frac{2\sum_i \sum_{j<i} a_{ij}^\alpha a_{ij}^\beta}{L^\alpha + L^\beta} = \frac{2L^{\alpha \rightrightarrows \beta}}{L^\alpha + L^\beta} \tag{3.3}$$

where a_{ij}^α are the entries of the adjacency matrices of the various layers, $L^\alpha = \sum_{i<j} a_{ij}^\alpha$ is the number of links in that layer and $L^{\alpha \rightrightarrows \beta}$ counts the number of links present in both layers α and β between the same pairs of nodes. This notation is somewhat redundant at this stage, but on the other hand it allows for an easier generalization to the directed case, as we will show later. So, $m_b^{\alpha\beta}$ ranges between 0 and 1 and represents a normalized overlap between pairs of layers of a multiplex. As mentioned in the introduction of this chapter, in the directed case we must take into account both the 'aligned' and the 'anti-aligned' overlap. Hence, in the second chapter we defined the binary directed multiplexity and multireciprocity [31] respectively as:

$$m_b^{\alpha\beta} = \frac{2\sum_i \sum_{j\neq i} \min\{a_{ij}^\alpha, a_{ij}^\beta\}}{L^\alpha + L^\beta} = \frac{2\sum_i \sum_{j\neq i} a_{ij}^\alpha a_{ij}^\beta}{L^\alpha + L^\beta} = \frac{2L^{\alpha \rightrightarrows \beta}}{L^\alpha + L^\beta} \tag{3.4}$$

and

$$r_b^{\alpha\beta} = \frac{2\sum_i \sum_{j\neq i} \min\{a_{ij}^\alpha, a_{ji}^\beta\}}{L^\alpha + L^\beta} = \frac{2\sum_i \sum_{j\neq i} a_{ij}^\alpha a_{ji}^\beta}{L^\alpha + L^\beta} = \frac{2L^{\alpha \leftrightarrows \beta}}{L^\alpha + L^\beta} \tag{3.5}$$

where $L^{\alpha \rightrightarrows \beta}$ represents the number of directed links present in both the considered layers between the same pairs of nodes, while $L^{\alpha \leftrightarrows \beta}$ counts the number of directed links present in α which are reciprocated in β, over all the possible pairs of vertices.

In the previous sections we stressed the importance of the inter-layer link coupling for the characterization of a real-world multiplex. We now pave the way for realistic (undirected and directed) binary models that can capture the observed features in the particular case of the World Trade Multiplex. Once more, one should not confuse these realistic models with the null models used in other contexts [50].

3.3.1 Undirected binary model

We start with the definitions of the measures we will focus on in our analysis. The empirical single-layer degree reads:

$$k_i^\alpha = \sum_{j \neq i} a_{ij}^\alpha. \tag{3.6}$$

Moreover, we can introduce the first of the new quantities that will allow us to properly describe the inter-layer coupling of a multiplex, namely the empirical *multiplexed degree*:

$$k_i^{\alpha \rightrightarrows \beta} = \sum_{j \neq i} a_{ij}^\alpha a_{ij}^\beta. \tag{3.7}$$

If we look at Eq. (3.3), we immediately see that, as compared to the global quantity $m_b^{\alpha,\beta}$, the multiplexed degree $k_i^{\alpha \rightrightarrows \beta}$ provides an even more detailed, local quantification of the multiplexity.

In what follows, we first establish an empirically robust pattern displayed by $k_i^{\alpha \rightrightarrows \beta}$ and then select it as one of the target properties that a multiplex reconstruction model should replicate, in addition to the desired marginal single-layer network properties. Figure 3.1 reports the scatter plot of $k_i^{\alpha \rightrightarrows \beta}$ versus k_i^β for four pairs of commodities (blue points). We clearly see an approximate linear trend of the type

$$k_i^{\alpha \rightrightarrows \beta} \approx u^{\alpha\beta} k_i^\beta. \tag{3.8}$$

Similar plots can be observed for the other pairs of layers as well (not shown). The robustness of this pattern motivates us to look for a multiplex model able to replicate it.

We define the joint probability $p_{ij}^{\alpha \rightrightarrows \beta}$ for the simultaneous presence of a link from node i to node j in layer α and of a corresponding link in layer β:

$$p_{ij}^{\alpha \rightrightarrows \beta} \equiv P(a_{ij}^\alpha = 1 \cap a_{ij}^\beta = 1) = \langle a_{ij}^\alpha a_{ij}^\beta \rangle = p_{ij}^{\beta \rightrightarrows \alpha}. \tag{3.9}$$

Using $p_{ij}^{\alpha \rightrightarrows \beta}$ and the aforementioned p_{ij}^β we can also obtain the *conditional* probability $u_{ij}^{\alpha\beta}$ that a link from i to j exists in layer α, given that the corresponding link exists in layer β:

$$u_{ij}^{\alpha\beta} \equiv P(a_{ij}^\alpha = 1 | a_{ij}^\beta = 1) = p_{ij}^{\alpha \rightrightarrows \beta} / p_{ij}^\beta. \tag{3.10}$$

We call $u_{ij}^{\alpha\beta}$ the *multiplexity probability*. Note that, while $p_{ij}^{\alpha \rightrightarrows \beta}$ is symmetric under the exchange of α and β, $u_{ij}^{\alpha\beta}$ is not; indeed, we have:

$$p_{ij}^{\alpha \rightrightarrows \beta} = u_{ij}^{\alpha\beta} p_{ij}^\beta = u_{ij}^{\beta\alpha} p_{ij}^\alpha = p_{ij}^{\beta \rightrightarrows \alpha}. \tag{3.11}$$

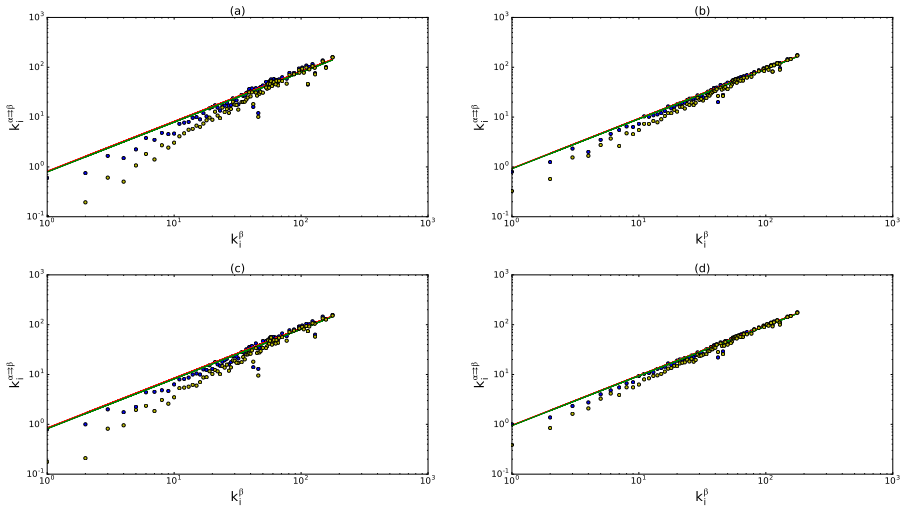

Figure 3.1: **Degree of layer β versus inter-layer multiplexed degree for 4 different pairs of commodities**: inorganic chemicals (a), plastics (b), iron and steel (c), electric machinery (d) versus trade in cereals. Blue dots: real data; yellow dots: expected multiplexed degree according to the uncorrelated model; lower green line: expected trend according to (3.24); upper red line (when discernible): best fit. In all the cases, $R^2 > 0.93$, for both the curves. It should be noted that we fit the empirical data with lines of the form $y = a \cdot x$, and only after we plot the results in log-log scale.

Furthermore $p_{ij}^{\alpha \rightrightarrows \beta}$ depends, at least in the general case, both on the pair of nodes and on the pair of layers. Given the previous definitions, the expected value of the multiplexed degree becomes:

$$\langle k_i^{\alpha \rightrightarrows \beta} \rangle = \sum_{j \neq i} \langle a_{ij}^{\alpha} a_{ij}^{\beta} \rangle = \sum_{j \neq i} p_{ij}^{\alpha \rightrightarrows \beta} = \sum_{j \neq i} u_{ij}^{\alpha\beta} p_{ij}^{\beta} = \sum_{j \neq i} u_{ij}^{\beta\alpha} p_{ij}^{\alpha}. \tag{3.12}$$

The main goal consists in understanding the structure of $u_{ij}^{\alpha\beta}$, which is the crucial quantity responsible for the coupling among layers. By contrast, as already said before, p_{ij}^{β} can in general be left largely unspecified as it can be chosen to be any single-layer network model that satifactorily reproduces a set of desired marginal topological properties of layer β, irrespective of the coupling with the other layers. The only basic property we require from p_{ij}^{β} is that the degree sequence is among such desired properties, or in other words that, for each node i and each layer β, the expected degree $\langle k_i^{\beta} \rangle = \sum_{j \neq i} p_{ij}^{\beta}$ satisfactorily replicates the empirical degree

k_i^β:

$$\langle k_i^\beta \rangle = \sum_{j \neq i} p_{ij}^\beta \approx k_i^\beta \quad \forall i. \tag{3.13}$$

For instance, if the Binary Configuration Model [36, 37] is chosen as the marginal single-layer reconstruction method, the above criterion is strictly verified, since that model assumes that the degree of each node is known and that the p_{ij}^β can be *constructed* as the maximum-entropy probability such that

$$\langle k_i^\beta \rangle = \sum_{j \neq i} p_{ij}^\beta = k_i^\beta \quad \forall i. \tag{3.14}$$

Other marginal reconstruction methods, which relax the hypothesis that the degree of each node is known, use other node-specific pieces of information, plus some proxy of the overall network density, to construct a p_{ij}^β such that Eq. (3.13) is in any case realized [23, 24, 25, 33, 51]. The above examples have all been shown to provide reliably reconstructed networks [23, 24, 25].

The presence of a nontrivial $u_{ij}^{\alpha\beta}$ in the present multiplex model implies that any p_{ij}^β coming from a single-layer model should be interpreted as a marginal probability resulting from a more realistic model where the presence of links across all layers is governed by a joint distribution for the entire multiplex. In other words, $u_{ij}^{\alpha\beta}$ allows us to extend any desired single-layer model to a truly multiplex model with nontrivial coupling among layers. The trivial case of independent layers can be easily recovered by setting:

$$\left[u_{ij}^{\alpha\beta}\right]_{unc} = p_{ij}^\alpha \tag{3.15}$$

since here the presence of the link in layer β does not affect the connection probability in layer α. In such a case, the expected multiplexed degree becomes:

$$\langle k_i^{\alpha \rightrightarrows \beta} \rangle_{unc} = \sum_{j \neq i} p_{ij}^\alpha p_{ij}^\beta. \tag{3.16}$$

From Eq. (3.3) it should be noted that, if such an uncoupled model were used to generate the multiplex, the expected value of the multiplexity $m_b^{\alpha\beta}$ would be zero. Yet, if Eq. (3.13) holds, then the inter-layer degree correlation function defined in Eq. (3.1) would be replicated. This shows that such a correlation function is not informative about the genuine inter-layer dependencies which go beyond the degree-degree correlations across the layers of the multiplex. By contrast, the multiplexity $m_b^{\alpha\beta}$ is, confirming the argument that led us to its introduction in Chapter 1.

To build a minimal model that can reproduce the observed level of similarity (i.e., multiplexity) between layers of the multiplex, we require that the robust empirical trend encapsulated in Eq. (3.8) is replicated. Looking at Eqs. (3.12)

and (3.13), and imposing Eq. (3.8), this requirement implies that the conditional probability $u_{ij}^{\alpha\beta}$ should be approximately independent of the pair of nodes:

$$u_{ij}^{\alpha\beta} = \frac{p_{ij}^{\alpha\rightrightarrows\beta}}{p_{ij}^{\beta}} = \frac{\langle a_{ij}^{\alpha} a_{ij}^{\beta}\rangle}{\langle a_{ij}^{\beta}\rangle} \approx u^{\alpha\beta}. \tag{3.17}$$

Since the transformation $i \mapsto j$ together with $\alpha \mapsto \beta$ keeps the quantities unaffected, we also have

$$u^{\alpha\beta}\langle a_{ij}^{\beta}\rangle \approx \langle a_{ij}^{\alpha} a_{ij}^{\beta}\rangle = \langle a_{ij}^{\beta} a_{ij}^{\alpha}\rangle \approx q^{\beta\alpha}\langle a_{ij}^{\alpha}\rangle. \tag{3.18}$$

Summing over i and j, we get

$$u^{\alpha\beta} L^{\beta} \approx u^{\beta\alpha} L^{\alpha}. \tag{3.19}$$

and from (3.17) we immediately have

$$u^{\alpha\beta}\langle a_{ij}^{\beta}\rangle \approx \langle a_{ij}^{\alpha} a_{ij}^{\beta}\rangle. \tag{3.20}$$

Summing over i and j and inverting, we obtain

$$u^{\alpha\beta} \approx \frac{\sum_i \sum_{i<j} \langle a_{ij}^{\alpha} a_{ij}^{\beta}\rangle}{\sum_i \sum_{i<j} \langle a_{ij}^{\beta}\rangle} = \frac{\sum_i \sum_{i<j} a_{ij}^{\alpha} a_{ij}^{\beta}}{L^{\beta}}. \tag{3.21}$$

The above relations allow us to express twice the inverse of (3.3) as

$$\frac{2}{m_b^{\alpha\beta}} = \frac{L^{\alpha} + L^{\beta}}{\sum_i \sum_{i<j} a_{ij}^{\alpha} a_{ij}^{\beta}} \approx \frac{1}{u^{\alpha\beta}} + \frac{1}{u^{\beta\alpha}} \tag{3.22}$$

where $m_b^{\alpha\beta}$ is measured from the multiplex data while $u^{\alpha\beta}$ is derived from the slope of the empirical linear relationship between k_i^{β} and $k_i^{\alpha\rightrightarrows\beta}$. Thus, we find that $m_b^{\alpha\beta}$ is approximately the harmonic mean of the conditional probabilities $u^{\alpha\beta}$ and $u^{\beta\alpha}$. Applying Eq. (3.19) to the previous expression, we get:

$$\begin{aligned} \frac{2}{m_b^{\alpha\beta}} &\approx \frac{1}{u^{\alpha\beta}}\left(1 + \frac{u^{\alpha\beta}}{u^{\beta\alpha}}\right) \\ &\approx \frac{1}{u^{\alpha\beta}}\left(1 + \frac{L^{\alpha}}{L^{\beta}}\right) \\ &= \frac{L^{\alpha} + L^{\beta}}{u^{\alpha\beta} L^{\beta}} \end{aligned} \tag{3.23}$$

Hence, the value of the slope in the plots of $k_i^{\alpha\rightrightarrows\beta}$ vs k_i^{β} is predicted to be

$$u^{\alpha\beta} \approx \frac{L^{\alpha} + L^{\beta}}{2L^{\beta}} m_b^{\alpha\beta} \tag{3.24}$$

Indeed, in Figure 3.1 we show that the best fit curves almost coincide with the expected ones having slope calculated independently from Eq. (3.24). Futhermore, we also show (yellow dots) that the model assuming independent layers as in Eqs. (3.15) and (3.16) produces values of the multiplexed degree that are systematically lower than the empirical ones.

From the previous analysis, it turns out phenomenologically that the minimal model one can design in order to reproduce the (local) observed values of the multiplexed degree requires only the (global) information about the total number of multiplexed links $L^{\alpha\rightrightarrows\beta}$ for any ordered pair of layers (α, β) (together with the aforementioned degree sequences in each layer).

In other words, a reliable network reconstruction method for the class of multiplexes we are focusing on here requires as input information a reconstruction model that works successfully on each layer separately, plus the $M(M-1)/2$ values of $L^{\alpha\rightrightarrows\beta}$, for all pairs of layers. These values are the numerators of the entries of the so-called (binary) *multiplexity matrix* [30]. If the reconstruction model is chosen to be the Configuration Model, then the overall input information reduces to the degree sequence $\vec{k}^{\alpha}$ for each layer α, plus the values $L^{\alpha\rightrightarrows\beta}$ for each pair of layers.

3.3.2 Directed binary model

As said in the introductive section, in the directed case we should take into account that the inter-layer coupling can intervene both in terms of alignment and anti-alignment. Hence, we have not only to extend the notion of *multiplexed degree* to the directed case, but also to introduce the quantity dubbed *multireciprocated degree*. It is indeed straightforward to exploit the same approach to analyse the patterns of multiplexity and multireciprocity in the directed case. The main difference w.r.t. the undirected case will consist in the definition of two separate conditional probabilities. We start defining the quantities that we will measure on the real multiplex network, namely the in-degree:

$$k_i^{\alpha,in} = \sum_{j\neq i} a_{ji}^{\alpha}; \tag{3.25}$$

and the out-degree:

$$k_i^{\alpha,out} = \sum_{j\neq i} a_{ij}^{\alpha}. \tag{3.26}$$

In analogy with the undirected model, we assume we can start from a marginal single-layer model characterized by the probability $p_{ij}^{\beta} = \langle a_{ij}^{\beta} \rangle$ that a *directed* link from node i to node j exists. The only thing we require from p_{ij}^{β} is that it reliably

replicates the in- and out-degree of each node i in layer β:

$$\langle k_i^{\alpha,in} \rangle = \sum_{j \neq i} p_{ji}^{\alpha} \approx k_i^{\alpha,in} \quad \forall i \tag{3.27}$$

$$\langle k_i^{\alpha,out} \rangle = \sum_{j \neq i} p_{ij}^{\alpha} \approx k_i^{\alpha,out} \quad \forall i, \tag{3.28}$$

generalizing the corresponding criterion in Eq. (3.13).

We also define the multiplex quantities that extend the ones introduced in the undirected case, i.e. the *multiplexed degree*:

$$k_i^{\alpha \rightrightarrows \beta} = \sum_{j \neq i} a_{ij}^{\alpha} a_{ij}^{\beta}. \tag{3.29}$$

and the *multireciprocated degree.*

$$k_i^{\alpha \rightleftarrows \beta} = \sum_{j \neq i} a_{ij}^{\alpha} a_{ji}^{\beta}. \tag{3.30}$$

It is possible to generalize the argument explained in the previous subsection; also in this case we find that $k_i^{\alpha \rightrightarrows \beta}$ and $k_i^{\alpha \rightleftarrows \beta}$ are in almost-linear relation with, respectively, $k_i^{\beta,out}$ (not shown, as it is very similar to the undirected case) and $k_i^{\beta,in}$ (Figure 3.2, blue dots), therefore we can set:

$$k_i^{\alpha \rightrightarrows \beta} \approx u^{\alpha\beta} k_i^{\beta,out} \tag{3.31}$$

and

$$k_i^{\alpha \rightleftarrows \beta} \approx v^{\alpha\beta} k_i^{\beta,in}. \tag{3.32}$$

The presence of two different multiplex quantities leads to the definition of two distinct joint probabilities:

$$p_{ij}^{\alpha \rightrightarrows \beta} \equiv P(a_{ij}^{\alpha} = 1 \cap a_{ij}^{\beta} = 1) = \langle a_{ij}^{\alpha} a_{ij}^{\beta} \rangle = p_{ij}^{\beta \rightrightarrows \alpha} \tag{3.33}$$

gives the probability for the simultaneous presence of a link from node i to node j in layer α and of a corresponding link (with the same direction) in layer β, while:

$$p_{ij}^{\alpha \rightleftarrows \beta} \equiv P(a_{ij}^{\alpha} = 1 \cap a_{ji}^{\beta} = 1) = \langle a_{ij}^{\alpha} a_{ji}^{\beta} \rangle = p_{ji}^{\beta \rightleftarrows \alpha} \tag{3.34}$$

is the probability of having a link from node i to node j in layer α and a link in the opposite direction in layer β. Consequently, from these joint probabilities and the marginal single-layer probabilities we can derive the two separate *conditional* probability $u_{ij}^{\alpha\beta}$ that a link from i to j exists in layer α, given that the corresponding link exists in layer β:

$$u_{ij}^{\alpha\beta} \equiv P(a_{ij}^{\alpha} = 1 | a_{ij}^{\beta} = 1) = p_{ij}^{\alpha \rightrightarrows \beta} / p_{ij}^{\beta}. \tag{3.35}$$

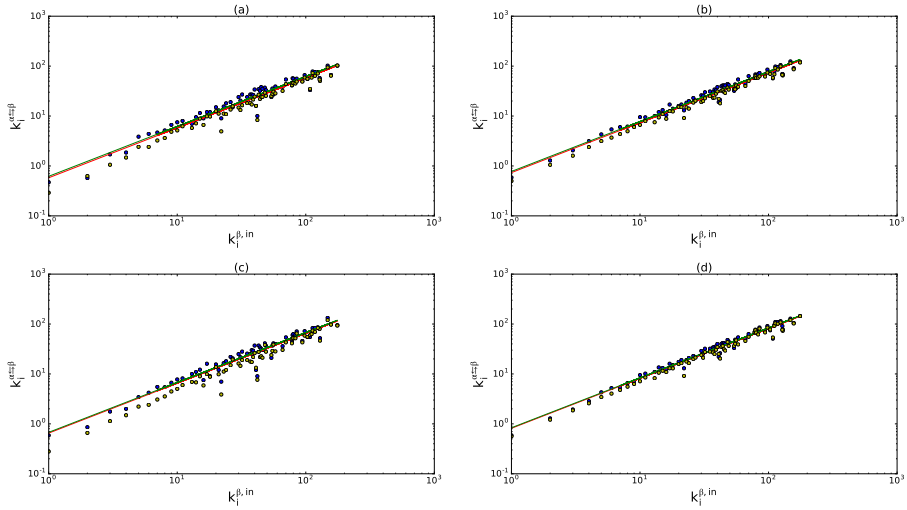

Figure 3.2: **In-degree of layer β versus inter-layer multireciprocated degree for 4 different pairs of commodities:** inorganic chemicals (a), plastics (b), iron and steel (c), electric machinery (d) versus trade in cereals. Blue dots: real data; yellow dots: expected multireciprocated degree according to the uncorrelated model; lower green line: expected trend according to (3.40); upper red line (when discernible): best fit. In all the cases, $R^2 > 0.95$, for both the curves. It should be noted that we fit the empirical data with lines of the form $y = a \cdot x$, and only after we plot the results in log-log scale.

and $v_{ij}^{\alpha\beta}$ representing the probability of having a link from i to j in α, given that a link from j to i exists in layer β:

$$v_{ij}^{\alpha\beta} \equiv P(a_{ij}^{\alpha} = 1 | a_{ji}^{\beta} = 1) = p_{ij}^{\alpha \rightleftarrows \beta} / p_{ji}^{\beta}. \tag{3.36}$$

We call $u_{ij}^{\alpha\beta}$ the *multiplexity probability* and $v_{ij}^{\alpha\beta}$ the *multireciprocity probability*. These probabilities lead to the separate notions of expected *multiplexed* and *multireciprocated degree*, defined respectively as:

$$\langle k_i^{\alpha \rightrightarrows \beta} \rangle = \sum_{j \neq i} \langle a_{ij}^{\alpha} a_{ij}^{\beta} \rangle = \sum_{j \neq i} p_{ij}^{\alpha \rightrightarrows \beta} = \sum_{j \neq i} u_{ij}^{\alpha\beta} p_{ij}^{\beta} = \sum_{j \neq i} u_{ij}^{\beta\alpha} p_{ij}^{\alpha} \tag{3.37}$$

and:

$$\langle k_i^{\alpha \rightleftarrows \beta} \rangle = \sum_{j \neq i} \langle a_{ij}^{\alpha} a_{ji}^{\beta} \rangle = \sum_{j \neq i} p_{ij}^{\alpha \rightleftarrows \beta} = \sum_{j \neq i} v_{ij}^{\alpha\beta} p_{ji}^{\beta} = \sum_{j \neq i} v_{ij}^{\beta\alpha} p_{ji}^{\alpha} \tag{3.38}$$

Analogously to the undirected case, $u_{ij}^{\alpha\beta}$ and $v_{ij}^{\alpha\beta}$ are driving the real coupling among the layers of the system, while the single-layer probabilities p_{ij}^{α} can be freely

chosen starting from any network model that correctly reproduces the marginal topology of the considered layer. For instance, we may choose the Directed Configuration Model [36, 37], for which Eqs. (3.27) and (3.28) hold with a strict equality sign, or some of its relaxed versions that assume less input information [23, 24, 25].

With the same reasoning of the previous subsection, it is possible to show that the value of the slope in the plots of $k_i^{\alpha\rightrightarrows\beta}$ vs $k_i^{\beta,out}$ is predicted to be:

$$u^{\alpha\beta} \approx \frac{L^\alpha + L^\beta}{2L^\beta} m_b^{\alpha\beta} \tag{3.39}$$

while the slope in the plots of $k_i^{\alpha\rightleftarrows\beta}$ vs $k_i^{\beta,in}$ is, according to the model:

$$v^{\alpha\beta} \approx \frac{L^\alpha + L^\beta}{2L^\beta} r_b^{\alpha\beta}. \tag{3.40}$$

As shown in Figure 3.2 for the multireciprocated degree (the corresponding plot referred to the multiplexed degree is not reported, being however very similar to the undirected case), the best fit curves are well modelled by the expected ones. We also show the results of the uncorrelated model, producing again values of the multireciprocated degree which are systematically lower than the observed values.

It turns therefore out that an appropriate multiplex reconstruction method for the class of directed multi-layer networks we are considering is based on the information about the in- and out-degree sequences of each layer combined with the entries of the matrices $L^{\alpha\rightrightarrows\beta}$ and $L^{\alpha\rightleftarrows\beta}$ for any pair of layers.

3.4 Weighted multiplex model

In the case of weighted multiplex networks, the marginal (i.e. single-layer) quantity we will focus on is the the weight w_{ij}^α associated to any (possible directed) link between i and j in layer α, together with its expected value $\langle w_{ij}^\alpha \rangle$. At the same time, we can still consider the link probability p_{ij}^α, representing the chance that nodes i and j are connected by a link, irrespective of the weight of the latter. Since the assumption of multidyadic independence still holds, the information provided by $\langle w_{ij}^\alpha \rangle$ and p_{ij}^α does not involve other pairs of nodes other than (i, j).

As the marginal quantities $\langle w_{ij}^\alpha \rangle$ are not influenced by the inter-layer coupling that we will add, they can therefore be considered as expectation values provided by any model able to correctly reproduce the weighted structure of layer α. However, in order to correctly reproduce the entire multiplex, we need to employ a single-layer model which has been proved to be reliable; it has been shown [22] that the Weighted Configuration Model [52] is not capable of reproducing both the topology and the weighted structure of a network, as it gives rise to almost complete graphs. Instead, we can think of the marginal values as stemming from the Enhanced Configuration Model [29, 37] - constraining both the degree and

the strength sequence of the observed graph, i.e.

$$\langle s_i^\alpha \rangle = \sum_{j \neq i} \langle w_{ij}^\alpha \rangle = s_i^\alpha \quad \forall i, \tag{3.41}$$

$$\langle k_i^\alpha \rangle = \sum_{j \neq i} p_{ij}^\alpha = k_i^\alpha \quad \forall i. \tag{3.42}$$

Similar to the binary case, these constraints can be relaxed in such a way that the required input information is considerably reduced. For instance, the methods proposed in refs. [23, 24, 25] require much less input information but are still such that

$$\langle s_i^\alpha \rangle = \sum_{j \neq i} \langle w_{ij}^\alpha \rangle \approx s_i^\alpha \quad \forall i, \tag{3.43}$$

$$\langle k_i^\alpha \rangle = \sum_{j \neq i} p_{ij}^\alpha \approx k_i^\alpha \quad \forall i, \tag{3.44}$$

and have recently been found to provide the best reconstruction methods for monoplex weighted networks from limited information [45, 46].

In the weighted case, the assumption of dependency between layers means that the joint probability of observing a given weight w_{ij}^α between i and j in layer α together with a weight w_{ij}^β in β does not factorize into two separate single-layer probabilities. In previous studies [53] this issue has been tackled by introducing the concept of *multistrength*; however, as already explained for the binary case, this approach is practically feasible only in the case of multiplex networks with a (very) limited number of layers.

On the contrary, our multiplex reconstruction technique appears to be useful also when applied to multigraphs possessing a larger number of layers, as it requires as input the strength sequence of the various layers and the multiplexity/-multireciprocity matrices (both growing like M^2). This quadratic growth in the number of layers (opposed to the exponential growth shown by the multistrength method), combined with the phenomenological observation that the conditional probabilities are again independent of the considered pair of nodes, makes our approach very promising.

As we said, our reconstruction method builds on the notions of *weighted multiplexity and multireciprocity*; in particular, in the undirected case we will exploit the measures of weighted multiplexity introduced in Chapter 1:

$$m_w^{\alpha\beta} = \frac{2\sum_i \sum_{j<i} \min\{w_{ij}^\alpha, w_{ij}^\beta\}}{W^\alpha + W^\beta} = \frac{2W^{\alpha \rightrightarrows \beta}}{W^\alpha + W^\beta} \tag{3.45}$$

where w_{ij}^α are the entries of the weighted adjacency matrices of the various layers, $W^\alpha = \sum_{i<j} w_{ij}^\alpha$ is the total weight associated to the links in that layer and $W^{\alpha \rightrightarrows \beta}$ represents the "shared weight" between α and β. In analogy with the binary case, $m_w^{\alpha\beta}$ ranges between 0 and 1 and represents a normalized weighted

overlap between pairs of layers of the multi-graph. In the directed case, instead, we have to consider the overlap in both the directions. In Chapter 2 we defined the weighted directed multiplexity and multireciprocity respectively as:

$$m_w^{\alpha\beta} = \frac{2\sum_i \sum_{j\neq i} \min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}}{W^{\alpha} + W^{\beta}} = \frac{2W^{\alpha\rightrightarrows\beta}}{W^{\alpha} + W^{\beta}} \tag{3.46}$$

and

$$r_w^{\alpha\beta} = \frac{2\sum_i \sum_{j\neq i} \min\{w_{ij}^{\alpha}, w_{ji}^{\beta}\}}{W^{\alpha} + W^{\beta}} = \frac{2W^{\alpha\rightleftarrows\beta}}{W^{\alpha} + W^{\beta}} \tag{3.47}$$

where $W^{\alpha\rightrightarrows\beta}$ is the "shared total weight" between the considered layers, and $W^{\alpha\leftrightarrows\beta}$ is the "shared reciprocated weight" between α and β.

In the following sections we will show a method to reconstruct the World Trade Multiplex from single-layer information exploiting the knowledge of the aforementioned multiplexity and multireciprocity matrices.

3.4.1 Undirected weighted model

In this section, we will focus on the relation between the single-layer strength, defined as:

$$s_i^{\alpha} = \sum_{j\neq i} w_{ij}^{\alpha} \tag{3.48}$$

and the *multiplexed strength*, for any ordered pair of layers:

$$s_i^{\alpha\rightrightarrows\beta} = \sum_{j\neq i} w_{ij}^{\alpha\rightrightarrows\beta} \equiv \min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\} \tag{3.49}$$

where $w_{ij}^{\alpha\rightrightarrows\beta}$ is the multiplexed component of the weights associated to the links between i and j in layers α and β. In particular, $s_i^{\alpha\rightrightarrows\beta}$ is the multiplex quantity allowing us to describe the inter-layer weighted coupling. Figure 3.3 reports the relation between s_i^{β} and $s_i^{\alpha\rightrightarrows\beta}$ for various pairs of commodities of the World Trade Multiplex; a clear empirical trend is exhibited (blue points), that can be approximated as:

$$s_i^{\alpha\rightrightarrows\beta} \approx U^{\alpha\beta} s_i^{\beta}. \tag{3.50}$$

Our goal will consist in designing the minimal model able to capture this empirical evidence.

In this perspective, we define the corresponding expected quantities $\langle w_{ij}^{\alpha\rightrightarrows\beta}\rangle$ and $\langle w_{ij}^{\alpha}\rangle$; in particular, the multiplexed component can be written in terms of a

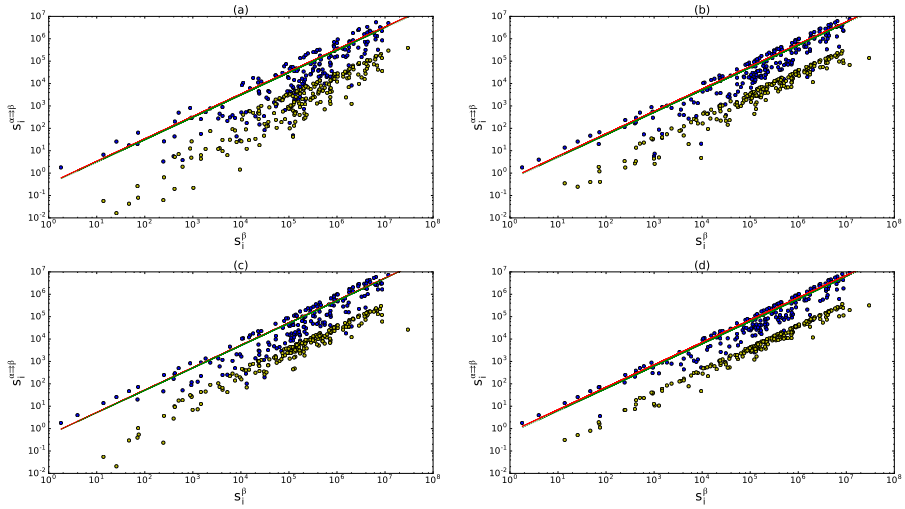

Figure 3.3: **Strength of layer β versus inter-layer multiplexed strength for 4 different pairs of commodities:** inorganic chemicals (a), plastics (b), iron and steel (c), electric machinery (d) versus trade in cereals. Blue dots: real data; yellow dots: expected multiplexed strength according to the uncorrelated model; lower green line: expected trend according to (3.59); upper red line (when discernible): best fit. In all the cases, $R^2 > 0.92$, for both the curves. It should be noted that we fit the empirical data with lines of the form $y = a \cdot x$, and only after we plot the results in log-log scale.

joint probability, in order to keep the same structure adopted for the binary case:

$$\begin{aligned}
\langle w_{ij}^{\alpha \rightrightarrows \beta} \rangle &= \langle \min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\} \rangle = \\
&= \sum_{w=1}^{\infty} P\left(\min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\} \geq w\right) = \\
&= \sum_{w=1}^{\infty} P\left(w_{ij}^{\alpha} \geq w \cap w_{ij}^{\beta} \geq w\right) = \\
&= \sum_{w=1}^{\infty} U_{ij}^{\alpha\beta}\left(w_{ij}^{\alpha} \geq w | w_{ij}^{\beta} \geq w\right) P\left(w_{ij}^{\beta} \geq w\right) \qquad (3.51)
\end{aligned}$$

where $U_{ij}^{\alpha\beta}$ is now the probability of observing a weight w_{ij}^{α} in α larger than w given that a weight w_{ij}^{β} larger than w has been observed in β.

As mentioned, the phenomenological observation shows that the conditional probability defined in (3.51) is actually independent from the considered pair of

nodes:

$$U_{ij}^{\alpha\beta} = \frac{\langle \min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}\rangle}{\langle w_{ij}^{\beta}\rangle} \approx U^{\alpha\beta} \tag{3.52}$$

Applying the same transformations $i \mapsto j$ and $\alpha \mapsto \beta$ we get:

$$\begin{aligned} U^{\alpha\beta}\langle w_{ij}^{\beta}\rangle &\approx \langle \min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}\rangle = \\ &= \langle \min\{w_{ij}^{\beta}, w_{ij}^{\alpha}\}\rangle \approx U^{\beta\alpha}\langle w_{ij}^{\alpha}\rangle \end{aligned} \tag{3.53}$$

Summing (3.53) over i and j, we have:

$$U^{\alpha\beta}W^{\beta} = U^{\beta\alpha}W^{\alpha} \tag{3.54}$$

Similarly, inverting (3.52) we obtain:

$$U^{\alpha\beta}\langle w_{ij}^{\beta}\rangle \approx \langle \min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}\rangle \tag{3.55}$$

and summing the previous expression, as in the binary case:

$$\begin{aligned} U^{\alpha\beta} &= \frac{\sum_i \sum_{j<i}\langle \min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}\rangle}{\sum_i \sum_{j<i}\langle w_{ij}^{\beta}\rangle} = \\ &= \frac{\sum_i \sum_{j<i} \min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}}{W^{\beta}} \end{aligned} \tag{3.56}$$

Therefore we get:

$$\frac{2}{m_w^{\alpha\beta}} = \frac{2\left(W^{\alpha} + W^{\beta}\right)}{2\sum_i \sum_{j<i} \min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}} = \frac{1}{U^{\alpha\beta}} + \frac{1}{U^{\beta\alpha}} \tag{3.57}$$

where $m_w^{\alpha\beta}$ represents the entry of the weighted multiplexity matrix and $U^{\alpha\beta}$ is derived from the empirical relationship between s_i^{β} and $s_i^{\alpha\rightrightarrows\beta}$. In analogy with the binary case, $m_w^{\alpha\beta}$ is therefore the harmonic mean of the conditional probabilities $U^{\alpha\beta}$ and $U^{\beta\alpha}$, as previously defined. Applying (3.54) to the previous expression, we get:

$$\begin{aligned} \frac{2}{m_w^{\alpha\beta}} &= \frac{1}{U^{\alpha\beta}}\left(1 + \frac{U^{\alpha\beta}}{U^{\beta\alpha}}\right) = \\ &= \frac{1}{U^{\alpha\beta}}\left(1 + \frac{W^{\alpha}}{W^{\beta}}\right) = \\ &= \frac{W^{\alpha} + W^{\beta}}{U^{\alpha\beta}W^{\beta}} \end{aligned} \tag{3.58}$$

Thus, the value of the angular coefficient in the plots $s_i^{\alpha\rightrightarrows\beta}$ vs s_i^{β} should be, in the weighted case:

$$U^{\alpha\beta} = \frac{W^{\alpha} + W^{\beta}}{2W^{\beta}} m_w^{\alpha\beta} \tag{3.59}$$

in perfect analogy with the unweighted case. Indeed, in Figure 3.3 we show the comparison between the actual fit lines and the expected ones according to (3.59): the agreement is clear and robust across different pairs of commodities.

Therefore, in analogy to the unweighted case, here the minimal model suitable to reproduce the observed values of pairwise weighted multiplexity is based on the total multiplexed weight $W^{\alpha\rightrightarrows\beta}$ for any ordered pair of layers (α, β), accompanied by the strength sequences measured in any layer. We indeed show that any model that does not take into account some sort of weighted coupling between layers would not be sufficient, as shown by the results provided by the uncorrelated model (yellow dots in Figure 3.3).

3.4.2 Directed weighted model

Also in the weighted case it is possible to extend the analysis to the directed case. Here, the main goal consists in the study of the relation between single-layer metrics and inter-layer weighted quantities, in order to model them exploiting the notions of directed multiplexity and multireciprocity introduced before.

We have to define two distinct strengths, namely the out-strength:

$$s_i^{\alpha,out} = \sum_{j\neq i} w_{ij}^{\alpha} \tag{3.60}$$

and the in-strength:

$$s_i^{\alpha,in} = \sum_{j\neq i} w_{ji}^{\alpha} \tag{3.61}$$

Moreover, also the multiplex quantities will split into two separate metrics, i.e. the *multiplexed strength*:

$$s_i^{\alpha\rightrightarrows\beta} = \sum_{j\neq i} w_{ij}^{\alpha\rightrightarrows\beta} \equiv \min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\} \tag{3.62}$$

and the *multireciprocated strength*:

$$s_i^{\alpha\rightleftarrows\beta} = \sum_{j\neq i} w_{ij}^{\alpha\rightleftarrows\beta} \equiv \min\{w_{ij}^{\alpha}, w_{ji}^{\beta}\} \tag{3.63}$$

where $w_{ij}^{\alpha\rightrightarrows\beta}$ is the multiplexed component of the weights associated to the directed links from i to j in layers α and β, and $w_{ij}^{\alpha\rightleftarrows\beta}$ is the reciprocated component. $s_i^{\alpha\rightrightarrows\beta}$ and $s_i^{\alpha\rightleftarrows\beta}$ are the metrics that will allow us to analyse and model the inter-layer coupling of the weighted World Trade Multiplex.

We empirically observe that the relations between $s_i^{\alpha,out}$ and $s_i^{\alpha\rightrightarrows\beta}$ (not shown), and $s_i^{\alpha,in}$ and $s_i^{\alpha\rightleftarrows\beta}$ (Figure 3.4, blue points) are both linearly approximated; hence:

$$s_i^{\alpha\rightrightarrows\beta} \approx U^{\alpha\beta} s_i^{\beta,out} \tag{3.64}$$

and

$$s_i^{\alpha \rightleftarrows \beta} \approx V^{\alpha\beta} s_i^{\beta,in}. \tag{3.65}$$

With the same reasoning developed for the undirected case, it is possible to derive

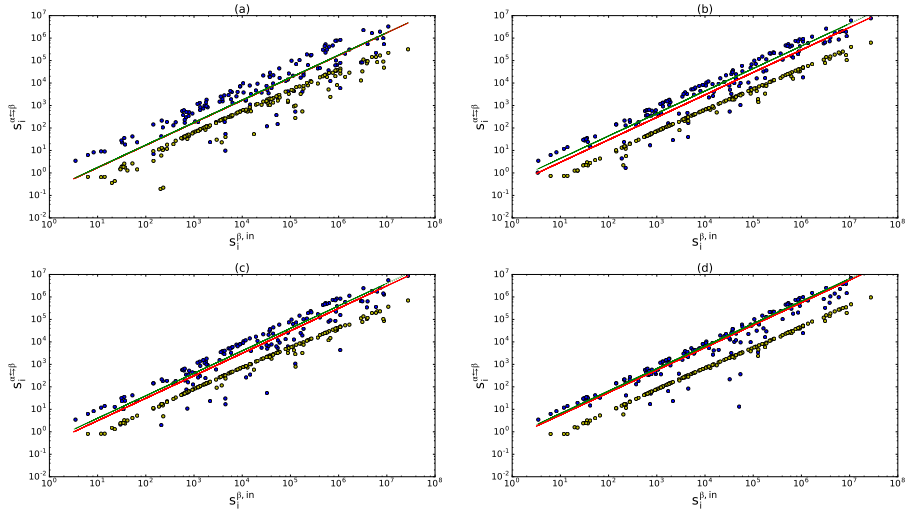

Figure 3.4: **In-strength of layer β versus inter-layer multireciprocated strength for 4 different pairs of commodities:** inorganic chemicals (a), plastics (b), iron and steel (c), electric machinery (d) versus trade in cereals. Blue dots: real data; yellow dots: expected multireciprocated strength according to the uncorrelated model; lower green line: expected trend according to (3.24); upper red line (when discernible): best fit. In all the cases, $R^2 > 0.95$, for both the curves. It should be noted that we fit the empirical data with lines of the form $y = a \cdot x$, and only after we plot the results in log-log scale.

the expected value of the angular coefficient $U^{\alpha\beta}$ and $V^{\alpha\beta}$, exploiting the notion of conditional probability; we obtain that the model predicts:

$$U^{\alpha\beta} = \frac{W^\alpha + W^\beta}{2W^\beta} m_w^{\alpha\beta} \tag{3.66}$$

and

$$V^{\alpha\beta} = \frac{W^\alpha + W^\beta}{2W^\beta} r_w^{\alpha\beta} \tag{3.67}$$

where $m_w^{\alpha\beta}$ and $r_w^{\alpha\beta}$ are the corresponding entries of, respectively, the multiplexity and multireciprocity matrices. The results of the fit of the model to the

World Trade Multiplex are shown in Figure 3.4. The model is able to satisfactorily reproduce the values of multireciprocated strengths starting from the single-layer in-strengths (similar results are obtained for the relation between the out-strength and the multiplexed strength), while an uncorrelated model (i.e., without introducing any sort of dependency between layers) cannot capture the phenomenological observation.

Hence, in the weighted directed case the most inexpensive reconstruction model builds on the knowledge of the in- and out-strength sequence of the different layers plus the $M \times M$ multiplexity and multireciprocity matrices.

3.5 Conclusions

The reconstruction of multiplex properties in multi-layer networks from single-layer information is an important and so far unfaced problem. Indeed, in the multiplex case the limitedness of information about the full topology may affect only some of the layers; hence, any tool allowing us to infer inter-layer node-specific properties from the known information related to some particular layer is theoretically interesting and practically useful. In this chapter we have provided a possible solution to this issue by means of the new quantities dubbed *multiplexed* and *multireciprocated degrees and strengths*, directly stemming from the previously defined *multiplexity* and *multireciprocity*. Our reconstruction technique builds on methods that have been shown to be well-grounded in the single-layer case. Indeed, previous studies highlighted that it is possible to correctly reproduce the topological structure of real-world graphs starting from limited information about, for instance, the strengths and the density of the considered system.

In this chapter we have extended the notion of network reconstruction to the case of multi-layer systems, in particular proving that a trustworthy reconstruction method can be based on the knowledge of (possibly in turn reconstructed) degrees or strengths of the single layers, combined with the compact and usually fixed-over-time multiplexity and multireciprocity matrices. Furthermore, our methodology works for both binary and weighted networks and it is able to take into account also the potential directionality of the links.

We must however stress that this technique is successfully applicable to systems exhibiting two main features. First, the single layers should be reproducible via the Configuration Model (or the Enhanced Configuration Model in the weighted case), such that the entire topology could be reconstructed just from the knowledge of the degrees of the single nodes (respectively, from the strengths). Second, the conditional probabilities of observing a link in any layer given that a link exists between the same pair of nodes in a different layer should be independent of the considered nodes: in other words, such probabilities (that we called *multiplexity and multireciprocity probabilities*) should be common for all the nodes and dependent only on the pair of layers we are focusing on. Although these assumptions significanly restrict the range of systems that can be successfully re-

constructed through our method, we highlight that one of most crucial economic networks, namely the World Trade Multiplex (incidentally, strongly suffering of the problem of missing data), belongs to this class of multi-layer networks.

Moreover, we have shown that the measures of multiplexed or multireciprocated degrees and strengths can give information about the coupling between layers. We have indeed explained that, by means of the aforementioned quantities, it is possible to acquire more refined notions of inter-layer coupling; multiplexed and multireciprocated degrees and strengths can therefore be thought of as new measures of multiplex assortativity, expressing the coupling caused by dependencies different than the simple correlation between the degree or strength distributions.

Future steps in the design of reconstruction techniques are needed in order to further generalize the aforementioned methods. Nevertheless, our findings show that the multiplexity and multireciprocity matrices allow us to reconstruct the joint connection probabilities from the marginal ones, hence bridging the gap between single-layer information and truly multiplex properties.

Bibliography

[1] A.-L. Barabási, R. Albert (1999) 'Emergence of scaling in random networks', *Science* **286** (5439), 509

[2] R. Albert, A.-L. Barabási (2002) 'Statistical mechanics of complex networks', *Review of Modern Physics* **74**, 47

[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D. U. Hwang (2006) 'Complex networks: structure and dynamics', *Physics Reports* **424** (4), 175

[4] R. Guimerà, L. A. N. Amaral (2004) 'Modeling the world-wide airport network', *European Physics Journal B* **38**, 381

[5] E. Bullmore, O. Sporns (2009) 'Complex brain networks: graph theoretical analysis of structural and functional systems', *Nature Review Neuroscience* **10**, 186

[6] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, S. Havlin (2010) 'Catastrophic cascade of failures in interdependent networks', *Nature* **464** (7291), 1025

[7] A. Cardillo, J. Gómez-Gardeñes, M. Zanin, M. Romance, D. Papo, F. del Pozo, S. Boccaletti (2013) 'Emergence of network features from multiplexity', *Scientific Reports* **3**, 1344

[8] A. Cardillo, M. Zanin, J. Gómez-Gardeñes, M. Romance, A. J. García del Amo, S. Boccaletti (2012) 'Modeling the multi-layer nature of the European Air Transport Network: resilience and passengers re-scheduling under random failures', *European Physics Journal Special Topics* **215**, 23

[9] A. Bashan, R. P. Bartsch, J. W. Kantelhardt, S. Havlin, P. C. Ivanov (2012) 'Network physiology reveals relations between network topology and physiological function', *Nature Communications* **3**, 702

[10] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, M. A. Porter (2014) 'Multilayer networks', *Journal of Complex Networks* **2** (3), 203

[11] V. Nicosia, V. Latora (2015) 'Measuring and modeling correlations in multiplex networks', *Physical Review E* **92** (3), 032805

[12] F. Battiston, V. Nicosia, V. Latora (2014) 'Structural measures for multiplex networks', *Physical Review E* **89** (3), 032804

[13] M. De Domenico, A. Sole-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gomez, A. Arenas (2013) 'Mathematical formulation of multilayer networks', *Physical Review X* **3** (4), 041022

[14] S. Gomez, A. Diaz-Guilera, J. Gómez-Gardeñes, C. J. Perez-Vicente, Y. Moreno, A. Arenas (2013) 'Diffusion dynamics on multiplex networks', *Physical Review Letters* **110** (2), 028701

[15] E. Cozzo, R. A. Banos, S. Meloni, Y. Moreno (2013) 'Contact-based social contagion in multiplex networks', *Physical Review E* **88** (5), 050801

[16] L. Bargigli, G. di Iasio, L. Infante, F. Lillo, F. Pierobon (2014) 'The multiplex structure of interbank networks', *Quantitative finance* **15** (4), 673

[17] S. Poledna, J. L. Molina-Borboa, S. Martinez-Jaramillo, M. van der Leij, S. Thurner (2015) 'The multi-layer network nature of systemic risk and its implications for the costs of financial crises', *Journal of Financial Stability* **20**, 70

[18] A. Clauset, C. Moore, M. E. J. Newman (2008) 'Hierarchical structure and the prediction of missing links in networks', *Nature* **453**, 98

[19] I. Mastromatteo, E. Zarinelli, M. Marsili (2012) 'Reconstruction of financial networks for robust estimation of systemic risk', *Journal of Statistical Mechanics* **2012**, P03011

[20] N. Musmeci, S. Battiston, G. Caldarelli, M. Puliga, A. Gabrielli (2013) 'Bootstrapping topological properties and systemic risk of complex networks using the fitness model', *Journal of Statistical Physics* **151**, 720

[21] G. Caldarelli, A. Chessa, A. Gabrielli, F. Pammolli, M. Puliga (2013) 'Reconstructing a credit network', *Nature Physics* **9**, 125

[22] R. Mastrandrea, T. Squartini, G. Fagiolo, D. Garlaschelli (2014) 'Enhanced reconstruction of weighted networks from strengths and degrees', *New Journal of Physics* **16**, 043022

[23] G. Cimini, T. Squartini, A. Gabrielli, D. Garlaschelli (2015) 'Estimating topological properties of weighted networks from limited information', *Physical Review E*, **92** (4), 040802

[24] G. Cimini, T. Squartini, D. Garlaschelli, A. Gabrielli (2015) 'Systemic risk analysis on reconstructed economic and financial networks', *Scientific Reports* **5** 15758

[25] T. Squartini, G. Cimini, A. Gabrielli, D. Garlaschelli (2017) 'Network reconstruction via density sampling', *Applied Network Science* **2** (1), 3

[26] M. Barigozzi, G. Fagiolo, D. Garlaschelli (2010) 'Multinetwork of international trade: a commodity-specific analysis', *Physical Review E* **81** (4), 046104

[27] T. Squartini, G. Fagiolo, D. Garlaschelli (2011) 'Randomizing world trade. I. A binary network analysis', *Physical Review E* **84** (4), 046117

[28] T. Squartini, G. Fagiolo, D. Garlaschelli (2011) 'Randomizing world trade. II. A weighted network analysis', *Physical Review E* **84** (4), 046118

[29] R. Mastrandrea, T. Squartini, G. Fagiolo, D. Garlaschelli (2014) 'Reconstructing the world trade multiplex: the role of intensive and extensive biases', *Physical Review E* **90** (6), 062804

[30] V. Gemmetto, D. Garlaschelli (2015) 'Multiplexity versus correlation: the role of local constraints in real multiplexes', *Scientific Reports* **5**, 9120

[31] V. Gemmetto, T. Squartini, F. Picciolo, F. Ruzzenenti, D. Garlaschelli (2016) 'Multiplexity and multireciprocity in directed multiplexes', *Physical Review E* **94** (4), 042316

[32] M. A. Serrano, M. Boguñá (2003) 'Topology of the world trade web', *Physical Review E* **68** (1), 015101

[33] D. Garlaschelli, M. I. Loffredo (2005) 'Structure and evolution of the world trade network', *Physica A* **355** (1), 138

[34] World Custom Organization. available at: `http://www.wcoomd.org`

[35] G. Gaulier, S. Zignago (2010) 'BACI: international trade database at the product-level (the 1994-2007 version)', *CEPII Working Paper* **23**

[36] T. Squartini, D. Garlaschelli (2011) 'Analytical maximum-likelihood method to detect patterns in real networks', *New Journal of Physics* **13**, 083001

[37] T. Squartini, R. Mastrandrea, D. Garlaschelli (2015) 'Unbiased sampling of network ensembles', *New Journal of Physics* **17**, 023052

[38] K. M. Lee, J. Y. Kim, W. K. Cho, K. I. Goh, I. M. Kim (2012) 'Correlated multiplexity and connectivity of multiplex random networks', *New Journal of Physics* **14**, 033027

[39] V. Nicosia, G. Bianconi, V. Latora, M. Barthelemy (2013) 'Growing multiplex networks', *Physical review Letters* **111** (5), 058701

[40] D. Garlaschelli, M. I. Loffredo (2008) 'Maximum likelihood: extracting unbiased information from complex networks', *Physical Review E* **78** (1), 015101

[41] G. Bianconi (2013) 'Statistical mechanics of multiplex networks: entropy and overlap', *Physical Review E* **87** (6), 062806

[42] H. Ebel, L. I. Mielsch, S. Bornholdt (2002) 'Scale-free topology of e-mail networks', *Physical Review E* **66** (3), 035103

[43] D. Garlaschelli, M. I. Loffredo (2004) 'Patterns of link reciprocity in directed networks', *Physical Review Letters* **93** (26), 268701

[44] T. Squartini, F. Picciolo, F. Ruzzenenti, D. Garlaschelli (2013) 'Reciprocity of weighted networks', *Scientific Reports* **3**, 2729

[45] P. Mazzarisi, F. Lillo (2017) 'Methods for reconstructing interbank networks from limited information: a comparison', in: Abergel, F. et al. (eds) *Econophysics and sociophysics: recent progress and future directions*, New Economic Windows. Springer

[46] K. Anand, et al. (2017) 'The missing links: a global study on uncovering financial network structures from partial data', *Journal of Financial Stability*

[47] S. Maslov, K. Sneppen (2002) 'Specificity and stability in topology of protein networks', *Science* **296** (5569), 910

[48] G. Fagiolo, J. Reyes, S. Schiavo (2010) 'The evolution of the world trade web: a weighted network analysis', *Journal of Evolutionary Economics* **20** (4), 479

[49] G. Fagiolo, J. Reyes, S. Schiavo (2008) 'On the topological properties of the world trade web: a weighted network analysis', *Physica A* **387** (15), 3868

[50] J. Park, M. E. J. Newman. (2004) 'Statistical mechanics of networks', *Physical Review E* **70** (6), 066117

[51] D. Garlaschelli, M. I. Loffredo (2004) 'Fitness-dependent topological properties of the world trade web' *Physical Review Letters* **93** (18), 188701

[52] M. A. Serrano, M. Boguñá (2005) 'Weighted configuration model', *AIP Conference Proceedings* **776**, 101

[53] G. Menichetti, D. Remondini, P. Panzarasa, R. J. Mondragón, G. Bianconi (2014) 'Weighted multiplex networks', *PLoS ONE* **9** (6), e97857