On metrics and models for multiplex networks

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 16 januari 2018
klokke 11.15 uur

door

Valerio Gemmetto

geboren te Alba (Italië)
in 1989

Promotor:	Prof. dr. ir. W. van Saarloos
Co-promotor:	Dr. D. Garlaschelli

Promotiecommissie:	Prof. dr. G. Caldarelli (IMT Lucca, Italy)
	Dr. M.A. Serrano (Universitat de Barcelona, Spain)
	Prof. dr. E.R. Eliel
	Prof. dr. W.Th.F. den Hollander
	Prof. dr. J.M. van Ruitenbeek
	Dr. M. Emmerich

This work was supported by the EU project MULTIPLEX (contract 317532) and by the Dutch Econophysics Foundation (Stichting Econophysics, Leiden, the Netherlands).

The cover shows a depiction of "The Reach" by JK Rofling (https://www.jkrofling.com/).

Casimir PhD series Delft-Leiden 2017-45
ISBN 978-90-8593-329-8

To Mom, Dad and Giorgia

Contents

Introduction

The explosion of the World Wide Web at the end of the 20th century boosted the interest in the comprehension of how real-world networks work. Indeed, before this boom the study of interconnected systems was limited to the social sciences [1, 2, 3] and mathematics [4]. The former used to employ the notion of networks to represent human relations like friendships and sexual contacts; the latter focused on the analysis of graphs, namely abstract systems composed of interacting units. The attention of physicists was, instead, mostly driven by the large size and the dynamic, self-organizing nature of these systems. In particular, the statistical physics approach proved to be convenient as it could help to understand the emergent macroscopic phenomena in terms of the microscopical interactions between the basic elements of the system. Moreover, the presence of common features shared by very diverse systems - like connectivity patterns characterized by large fluctuations, scale-free topology, etc. - asked for general modeling principles, typical of the physics community.

In this context, many efforts have been made to fully characterize the static structure of real interconnected systems, to analyze and model their growth and to study dynamical processes acting on top of such networks [5, 6, 7]. However, these studies can be affected by the noise and randomness associated to the considered systems; this observation, combined with the steadily increasing availability of "big data" [8], highlighted the need for methods that allow the extraction of the meaningful information from the - sometimes massive - real-world systems. This issue has been faced by means of the introduction of null models, i.e. benchmarks to which the observed networks could be compared. In particular, a successful set of such null models is represented by the maximum-entropy models [9], that proved their effectiveness in the grasp of (sometimes highly hidden) network patterns.

Despite all these efforts, scientists soon realized that something was still missing in the full understanding of many networked systems. For instance, the famous blackout that occurred in Italy in 2003 which involved almost its entire power grid could not be explained in terms of the usual network theory; however, a seminal work by H. E. Stanley, S. Havlin and collaborators [10] showed that such an event could be modelled as a failure cascade in interdependent networks, given that the cause of the large disruption was the interplay between the damage of the power stations and the resulting failure of the corresponding Internet communication

network, which in turn determined breakdowns of the still working power grid nodes.

Analogously, the interaction between the spreading of an epidemic and the information awareness useful to prevent the disease can be modelled as the coupling of two distinct dynamical processes acting on two overlapping networks: indeed, while the infection spreads on the network of physical contacts, the information propagates on the layer of virtual social encounters between the same individuals. It has been shown [11] that the presence of such a twofold structure can explain the effectiveness of tools like Facebook and Twitter in controlling and reducing the effects of seasonal influenza-like diseases.

These examples showed that many real systems can be suitably represented as different networks coupled with each other; this led to the notion of interdependent and multi-layer complex networks. Interdependent networks are systems composed by two or more distinct networks, where each node of any graph is dependent on one or more nodes belonging to the other(s) [10, 12], such as the power grid and the Internet system in the aforementioned example. Multi-layer networks, instead, are systems where a set of nodes is connected via distinct types of interaction, each represented by a different layer; in general, each element can be connected through intra-layer (i.e. within the layer) and inter-layer links (that is, edges connecting nodes in different layers) [13, 14]. In this thesis, we will focus on multiplex networks, which is a specific class of multi-layer systems where all layers consist of the same nodes [1].

So far, several aspects of multiplex networks have been investigated: their static structure [15, 16], possible growth mechanisms [17], community structure [18, 19] and dynamical processes occurring on top of them like diffusion [20] or epidemic spreading [21] are just some examples. Nevertheless, as mentioned for the single-layer case, the need of null models to extract the relevant information from these complex systems is still both significant and urgent, as it could help scientists in different fields to fully capture the essence of various real systems.

This thesis is meant to fill this gap. Specifically, we will extend the concept of null models as canonical ensembles of multi-graphs with given constraints and present new metrics able to characterize real-world layered systems based on their correlation patterns. We will make extensive use of the maximum-entropy method in order to find the analytical expression of the expectation values of several topological quantities; furthermore, we will employ the maximum-likelihood method to fit the models to real datasets. One of the main contributions of the present work is providing models and metrics that can be directly applied to real data, even in the case of multi-graphs exhibiting a large number of layers, unlike other work [22] that is, instead, limited to systems with a (very) small set of layers.

The thesis is divided into five chapters, each of them focusing on a different aspect related to the analysis of multiplex networks, although the last two chapters present results that are valid both for monoplex and multiplex systems.

[1] We will keep using the terms "multiplex networks", "multi-layer networks" and "multi-networks" as synonyms, although this is not strictly correct [13].

In Chapter 1 we focus on undirected multiplex networks. We provide new measures of correlation between the layers of a multi-graph, both for binary and weighted systems. Moreover, we highlight the importance of employing null models to distinguish between the information encoded in the node-specific properties and the one related to the higher-order interactions between the elements composing the network. In particular, we point out that the use of homogeneous random benchmarks can lead to misleading results, while heterogeneous null models are theoretically more appropriate and practically more reasonable. We test our measures and models on real-world networks, showing that our approach is able to characterize the considered systems based on their correlation patterns.

In Chapter 2 we shift our focus to directed multiplex networks. We show that the extension of the structural quantities developed in the previous chapter is not trivial, as the directionality of the connections implies that the interdependencies between layers are twofold: indeed, in addition to the tendency of links of different layers to align as the result of the above mentioned multiplexity, there exists also a complementary tendency to anti-align as the result of the so-called multi-reciprocity, expressing the propensity of links in one layer to be reciprocated by opposite links in a different layer. Furthermore, we provide a thorough analysis of the World Trade Multiplex; this system, representing the import-export connections among countries trading in different commodities, is indeed one of the best examples of directed weighted multi-layer graph. We point out that our investigation can have a significant impact on the development of product taxonomies and on the improvement of the existing algorithms to establish the complexity of products and competitiveness of countries.

In Chapter 3 we exploit the quantities introduced in the previous chapters to provide a new network reconstruction method applicable to multi-layer graphs. Missing data or confidentiality issues may limit our knowledge of the entire set of connections of a real network; hence, the problem of recovering the full topology of a network from partial information is a very hot topic, but such reconstructive techniques have not been extended to the multiplex case yet, to the best of our knowledge. Here we face this issue, providing a method able to infer the connections of any node in a given layer from the same information referred to a different layer. It turns out that this methodology, applicable to a specific class of multi-layer networks, can be successfully employed to reconstruct the World Trade Multiplex.

In the previous chapters we have shown that the null models are crucial in order to properly characterize the correlation patterns of these layered systems and to overcome the possible scarcity of topological information. In Chapter 4, instead, we illustrate that the maximum-entropy models also allow us to find the so-called backbone of a real network, i.e. the information which is irreducible to the single-node properties and is therefore peculiar to the network itself. This technique can be useful for vizualization purposes and may have a beneficial impact on further analyses such as community detection, both in terms of results and computational resources. We apply our filtering technique to several real systems (both single-

layer and multiplex) and compare the results with other state-of-the-art methods, showing that it is able to provide interesting insights in the analysis of very diverse real-world networks.

Finally, in Chapter 5 we move our attention to a different dataset, namely the scientific publication system. We exploit the innovative ScienceWISE platform [23] connected to the arXiv repository [24] to extract information about physics manuscripts and the scientific concepts therein. It turns out that this system has a straightforward representation in terms of a bipartite network, i.e. a graph composed by two distinct types of nodes such that an edge can exist only between nodes of different type (in our case, articles and concepts). From this bipartite network it is possible to build a unipartite (articles-only) graph, where any link stands for the similarity between two papers in terms of content. The application of a community detection algorithm allows us to make conclusions about specificities in the approach employed by authors to classify their articles; furthermore, we provide deeper interpretations of the notion of ground-truth.

We end the thesis with some concluding remarks and future perspectives on the design and application of maximum-entropy models to multiplex networks.

Bibliography

[1] J. L. Moreno (1934) 'Who shall survive: foundations of sociometry, group psychotherapy and sociodrama', Beacon House, New York

[2] M. S. Granovetter (1973) 'The strength of weak ties', *American Journal of Sociology* **78**, 1360

[3] S. Wasserman, K. Faust (1994) 'Social network analysis', Cambridge University Press (Cambridge, New York)

[4] P. Erdős, A. Rényi (1960) 'On the evolution of random graphs', *Publ. Math. Inst. Hung. Acad. Sci.* **5**, 17

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

[6] A. Barrat, M. Barthelemy, A. Vespignani (2008) 'Dynamical processes on complex networks', Cambridge University Press (Cambridge)

[7] G. Caldarelli (2007) 'Scale-free networks: complex webs in nature and technology', Oxford University Press (Oxford)

[8] C. Lynch (2008) 'Big data: how do your data grow?', *Nature* **455**, 28

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

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

[11] C. Granell, S. Gomez, A. Arenas (2013) 'Dynamical interplay between awareness and epidemic spreading in multiplex networks', *Physical Review Letters* **111** (12), 128701

[12] J. Gao, S. V. Buldyrev, H. E. Stanley, S. Havlin (2012) 'Networks formed from interdependent networks', *Nature Physics* **8** (1), 40

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

[14] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, M. Zanin (2014) 'The structure and dynamics of multilayer networks', *Physics Reports* **544** (1), 1

[15] M. Szell, R. Lambiotte, S. Thurner (2010) 'Multirelational organization of large-scale social networks in an online world', *Proceedings of the National Academy of Sciences USA* **107** (31), 13636

[16] 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

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

[18] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, J.-P. Onnela (2010) 'Community structure in time-dependent, multiscale, and multiplex networks', *Science* **328** (5980), 876

[19] M. De Domenico, A. Lancichinetti, A. Arenas, M. Rosvall (2015) 'Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems', *Physical Review X* **5** (1), 011027

[20] 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

[21] C. Granell, S. Gomez, A. Arenas (2014) 'Competing spreading processes on multiplex networks: awareness and epidemics', *Physical Review E* **90** (1), 012808

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

[23] `http://sciencewise.info/`

[24] `http://arxiv.org`

Chapter 1

Undirected multiplex networks

Several systems can be represented as multiplex networks, i.e. in terms of a superposition of various graphs, each related to a different mode of connection between nodes. Hence, the definition of proper mathematical quantities aiming at capturing the added level of complexity of those systems is required. Various steps in this direction have been made. In the simplest case, dependencies between layers are measured via correlation-based metrics, a procedure that we show to be equivalent to the use of completely homogeneous benchmarks specifying only global constraints. However, this approach does not take into account the heterogeneity in the degree and strength distributions, which is instead a fundamental feature of real-world multiplexes. In this chapter, we compare the observed dependencies between layers with the expected values obtained from maximum-entropy reference models that appropriately control for the observed heterogeneity in the degree and strength distributions. This information-theoretic approach results in the introduction of novel and improved multiplexity measures that we test on different datasets, i.e. the International Trade Network and the European Airport Network. Our findings confirm that the use of homogeneous benchmarks can lead to misleading results, and highlight the important role played by the distribution of hubs across layers.

The results presented in this chapter have been published in the following reference:
V. Gemmetto, D. Garlaschelli, *Scientific Reports*, **5**, 9120 (2015).

1.1 Introduction

The study of networks allows scientists to suitably represent and analyze biological, economic and social systems as a set of units (nodes) connected by edges (links) symbolizing interactions [1, 2, 3, 4].

However, this approach may actually lead to an oversimplification: indeed, several systems are composed by units connected by multiple kinds of interaction. In such systems, the same set of nodes is joined by various types of links, each of those representing a different mode of connection [5]. The simplest way to analyse such systems is the aggregation of the various levels in a single network, but it turns out that such a simplification may discard fundamental information about the real topology of the network and therefore about possible dynamical processes acting on the system [6]. For instance, such an aggregation may result in a loss of information about the distribution of the hubs across layers, which is instead crucial for the control of several processes arising on an interdependent network [7]. Then, in order to solve such an issue, in the last few years the study of multi-layer networks has been pursued. In this context, new quantities aiming at mathematically analyzing multi-level networks have been provided [8, 9, 10, 11]; furthermore, models of growth [12, 13, 14] and dynamical processes occurring on multiplexes, such as epidemic spreading [15], diffusion [16], cooperation [17] and information spreading [18] have been designed.

In this chapter, we follow the path towards the definition of measures that can be applied to multi-level networks, in order to characterize significant structural properties of these systems, in particular focusing on the analysis of the dependencies between layers. We argue that, in order to properly characterize such dependencies, a comparison between the observed correlation and some notion of expected correlation is required. We therefore exploit the concept of multiplex ensemble [19, 20, 21], aiming at the definition of suitable null models for multi-layer complex networks, in order to compare the observed overlap between layers with the expected overlap one would find in a random superposition of layers with the same node-specific properties. In particular, since our purpose is precisely that of measuring such dependencies, we will consider uncorrelated multiplex ensembles, in order to define a null model for the real system so that it is possible to compare the observed correlations with reference models where the overlap between layers is actually randomized and, at the same time, important node-specific properties of the real network are preserved.

Various efforts have already been made about the study of correlations in multi-level networks [22, 23, 24], but the comparison of the observed results with the expected ones has generally been based on a - sometimes implicit - assumption: the benchmark was a completely homogeneous graph. In particular, here we show that correlation-based measures of inter-layer dependency (of the type used e.g. in ref.[22]) build on an implicit assumption of homogeneity, which in the unweighted case is equivalent to the choice of the Random Graph as null model. Similarly, for weighted networks, the chosen benchmark was equivalent to the Weighted

Random Graph, where the weight distribution is independent from the considered pair of nodes [25].

However, this assumption of uniformity in the probability distributions strongly contrasts with the observed findings in real-world complex systems. Indeed, one of the most well-known features of complex networks is their heterogeneity [26], both in the degree distribution and in the strength distribution; it is therefore crucial to take this aspect into account when proper null models for graphs are designed. Moreover, it has been recently shown that, in multiplex networks, the correlation between degrees (and strengths) of nodes across different layers is also an important structural feature that can have strong effects on the dynamics [7, 27]. Ultimately, such inter-layer degree correlations determine the distribution of hubs across layers, i.e. whether the same nodes tend to be hubs across many layers, or whether different layers are characterized by different hubs. We therefore aim at measuring multiplexity in terms of the "residual" inter-layer dependencies that persist after we filter out, for each layer separately, the effects induced by the heterogeneity of the empirical degree (for unweighted networks) or strength (for weighted graphs) distribution. We show that such a refinement can completely change the final findings and lead to a deeper understanding of the actual dependencies observed between layers of a real-world multiplex.

First, we introduce a new *"absolute" measure of multiplexity* designed to quantify the overlap between layers of a multi-level complex network. Second, we derive the expression of the expected value of such a quantity, both in the binary and in the weighted case, for randomized networks, by enforcing different constraints. Third, we combine the "absolute" multiplexity and its expected value into a *filtered, "relative" measure of multiplexity* that has the desired properties. We finally apply our measures to two different real-world multiplexes, namely the World Trade Multiplex Network and the European Airport Network, showing that the analysis of the dependencies between layers can actually make some important structural features of these systems explicit.

Indeed, while the former shows significant correlations between layers (i.e., traded commodities), in the latter almost no overlap can generally be detected, thus clearly defining two opposite classes of multiplexes based on the observed correlations. Furthermore, we will link such a behaviour with the distribution of the hubs across layers, hence providing a straightforward explanation to the observed findings.

1.2 Methods

1.2.1 Null models

It is possible to design null models for multi-level networks as maximum-entropy ensembles on which we enforce a given set of constraints [21]. In particular, we exploit the concept of uncorrelated multiplex ensemble, so that the definition of

proper null models for the considered multiplex reduces to the definition of an indipendent null model for any layer of the system. In order to do this, we take advantage of the concept of canonical network ensemble, or exponential random graph [28], i.e. the maximum-entropy family of graphs satisfying a set of constraints on average. In this context the resulting randomized graph preserves only part of the topology of the considered real-world network and is entirely random otherwise, thus it can be employed as a proper reference model. However, fitting such previously defined models to real datasets is hard, since it is usually computationally demanding as it requires the generation of many randomized networks whose properties of interest have to be measured.

In this perspective, we exploit a fast and completely analytical maximum-entropy method, based on the maximization of the likelihood function [29, 30, 31], which provides the exact probabilities of occurrence of random graphs with the same average constraints as the real network. From such probabilities it is then possible to compute the expectation values of the properties we are interested in, such as the average link probability or the average weight associated to the link established between any two nodes. While the adoption of such a method is not strictly required when dealing with global constraints like the total number of links observed in a network (the so-called Random Graph), it becomes crucial when facing the problem of enforcing local constraints such as the degree sequence or the strength sequence (Binary or Weighted Configuration Model). More information about such null models can be found in the following subsections and appendices.

1.2.2 Homogeneous null models

The simplest null model for a binary multiplex is an independent superposition of layers in which each layer is a Random Graph (RG) [28], which enforces as constraint the expected number of links in that layer. Such model, therefore, provides a unique expected probability p^α that a link between any two nodes is established in layer α: however, such a reference model completely discards any kind of heterogeneity in the degree distributions of the layers, resulting in graphs where each node has on average the same number of connections, inconsistently with the observed real networks. Thus, the probability of connection between any two nodes in layer α is uniformly given by:

$$p^\alpha = \frac{L^\alpha}{N(N-1)/2} \tag{1.1}$$

where L^α is the total number of links actually observed in layer α.

Similar considerations apply to weighted networks and the related Weighted Random Graph (WRG) [25], i.e. the straightforward extension of the previous Random Graph to weighted systems; in such a null model, the probability of having a link of weight w between two nodes i and j is independent from the choice of the nodes and only depends on the total weight observed in a layer and on the number of nodes.

Analogously to the corresponding Binary Random Graph, also this kind of null model discards the simultaneous presence of nodes with high and low values of the strengths (that is, a high or low sum of the weights associated to links incident on that node).

1.2.3 Heterogeneous null models

To take into account the heterogeneity of the real-world networks, in the unweighted case we consider a null model where the multiplex is an independent superposition of layers, each of which is a (Binary) Configuration Model (BCM) [32], i.e. an ensemble of networks satisfying on average the empirical degree sequence observed in that specific layer. Since we make use of the canonical ensembles, it is possible to obtain from the maximum-likelihood method each probability p_{ij}^{α} that nodes i and j are connected in layer α (notice that such value p_{ij}^{α} is basically the expectation value of a_{ij}^{α} under the chosen Configuration Model). Similarly, as a null model for a weighted multiplex we consider an independent superposition of layers, each described by the Weighted Configuration Model (WCM) [33]: here, for each layer separately, the enforced constraint is the strength sequence as observed in the real-world multiplex. In this view, the likelihood maximization provides the expectation value of each weight w_{ij}^{α} for any pair of nodes i and j as supplied by the Weighted Configuration Model. It is worth noticing that enforcing the degree sequence (respectively, the strength sequence in the weighted case) automatically leads to the design of a null model where also the total number of links (respectively, the total weight) of the network is preserved. In the appendices attached to this chapter we will provide equations generalizing, for instance, equation (1.1), whose solution allows then to derive the analytical expression of the expected link probability p_{ij}^{α} and, in the weighted case, the expected link weight w_{ij}^{α}. In order to do this, we make use of a set of N auxiliary variables x_i^{α} for any layer α, which are proportional to the probability of establishing a link between a given node i and any other node (or, respectively for the weighted case, establishing a link characterized by a given weight), being therefore directly informative on the expected probabilities p_{ij}^{α} (or, respectively, the expected weights w_{ij}^{α}).

Before introducing our measures of multiplexity, we make an important preliminary observation. Simple measures of inter-layer dependency are based on correlation metrics, which in turn rely on an assumption of uniformity, such assumption being ultimately equivalent to the choice of a uniform Random Graph as a null model; this will strengthen the choice of emplying heterogeneous benchmarks throughout the entire thesis. We illustrate this result in more detail in the appendices.

1.2.4 Multiplexity

When unweighted networks are considered, we define the "absolute" binary multiplexity between any two layers α and β as:

$$m_b^{\alpha\beta} = \frac{2\sum_{i<j}\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}}{L^{\alpha} + L^{\beta}} \tag{1.2}$$

where L^{α} is the total number of links observed in layer α and $a_{ij}^{\alpha} = 0, 1$ depending on the presence of the link between nodes i and j in layer α. Such a quantity represents a normalized overlap between any pair of layers and can therefore be thought of as a normalized version of the global overlap introduced in [21].

The previous definition can be easily extended to weighted multiplex networks. We define the "absolute" weighted multiplexity as:

$$m_w^{\alpha\beta} = \frac{2\sum_{i<j}\min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}}{W^{\alpha} + W^{\beta}} \tag{1.3}$$

where w_{ij}^{α} represents the weight of the link between nodes i and j in layer α and W^{α} is the total weight related to the links in that layer. Both (1.2) and (1.3) range in $[0, 1]$, are maximal when layers α and β are identical - that is, if there is complete similarity between those two layers - and minimal when they are totally different; in this perspective, they evaluate the tendency of nodes to share links in distinct layers.

However, the above absolute quantities are uninformative without a comparison with the value of multiplexity obtained when considering a null model. We may indeed measure high values of multiplexity between two layers due to the possibly large observed values of density, without any significant distinction between real dependence and overlap imposed by the presence of many links in each layer (thus forcing an increase in the overlap itself).

Furthermore, we cannot draw a clear conclusion about the amount of correlation between layers by just looking at the observed value, since such a measure is not universal and, for instance, no comparison between different multiplexes can be done based on the raw "absolute" multiplexity.

We therefore introduce the following "relative" or rescaled quantity along the lines of refs. [34, 35]:

$$\mu^{\alpha\beta} = \frac{m^{\alpha\beta} - \langle m^{\alpha\beta}\rangle}{1 - \langle m^{\alpha\beta}\rangle} \tag{1.4}$$

where $m^{\alpha\beta}$ is the value measured for the observed real-world multiplex and $\langle m^{\alpha\beta}\rangle$ is the value expected under a suitably chosen null model. The main null models that we will consider are respectively the Random Graph (RG) and the Binary Configuration Model (BCM) in the unweighted case, the Weighted Random Graph (WRG) and the Weighted Configuration Model (WCM) in the weighted case. We will characterize them in more detail in the appendices following this chapter.

This rescaled quantity is now directly informative about the real correlation between layers: in this context, positive values of $\mu^{\alpha\beta}$ represent positive correlations, while negative values are associated to anticorrelated pairs of layers; furthermore, pairs of uncorrelated layers show multiplexity values comparable with 0.
One of the motivations of the present work is the consideration that, in the binary case, when the Random Graph is considered as a null model, the previous quantity (1.4) can actually be reduced to the standard correlation coefficient between the entries of the adjacency matrix referred to any two layers α and β of a multi-level graph, defined as:

$$Corr\{a_{ij}^{\alpha}, a_{ij}^{\beta}\} = \frac{\langle a_{ij}^{\alpha} a_{ij}^{\beta}\rangle - \langle a_{ij}^{\alpha}\rangle\langle a_{ij}^{\beta}\rangle}{\sigma_{\alpha}\sigma_{\beta}} \tag{1.5}$$

In the appendices, we show that the previous expression is nothing but a different normalization of the rescaled binary multiplexity defined in (1.4):

$$Corr\{a_{ij}^{\alpha}, a_{ij}^{\beta}\} = F\cdot\left(m^{\alpha\beta} - \langle m^{\alpha\beta}\rangle\right) \tag{1.6}$$

where F is a factor depending on L^{α}, L^{β} and N.

1.3 Results

1.3.1 Binary analysis

We validate our definitions applying them to two different real-world multiplexes: the World Trade Multiplex (WTM) ($N = 207$ countries, $M = 96$ layers representing traded commodities), available as a weighted multi-level network, and the European Airport Network ($N = 669$ airports, $M = 130$ airlines), provided as an unweighted system. A more detailed description of the International Trade dataset, which is one of the main focuses of the entire thesis, can be found in the appendix following this chapter.

The implementation of the concept of multiplexity to different networks can lead to completely divergent results, according to the structural features of the considered systems. Indeed, the application of (1.2) to the WTM leads to the color-coded multiplexity matrix $\mathbf{M}_b$ shown in Figure 1.1(a). Such an array generally shows very high overlaps between layers, i.e. between different classes of commodities, pointing out that usually each country tends to import from or export to the same set of countries almost independently of the traded items; this is true in particular for most of the edible products (layers characterized by commodity codes ranging from 1 to 22, as listed in the aforementioned appendix.

In order to have a complete picture of the dependencies between layers of the considered systems, we have to compare our findings with the overlaps expected for multiplexes having only some of the properties in common with the observed ones. The simplest benchmark, as well as the most widely used, is the Random Graph (RG), which discards, as we said, any kind of heterogeneity in the degree

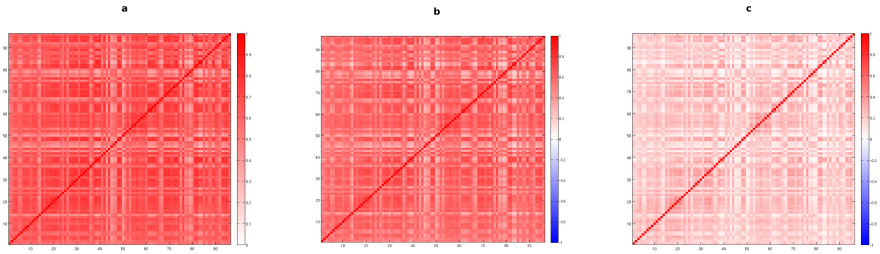

Figure 1.1: **Analysis of the binary multiplexity between layers of the World Trade Multiplex in 2011.** Color-coded matrices with entries given by $m_b^{\alpha\beta}$ (a), $\mu_{RG}^{\alpha\beta}$ (b) and $\mu_{BCM}^{\alpha\beta}$ (c) for any pair of layers (commodities).

distributions of the layers. When we compute $\mu_{RG}^{\alpha\beta}$ for the World Trade Network, we obtain the multiplexity matrix shown in Figure 1.1(b). The plot clearly shows that most of the correlations are still present: this layer-homogeneous null model, together with the presence of comparable densities across the various layers, does not significantly affect the expected overlaps. So far, we have discarded heterogeneity in our null models. However, this can considerably affect the significance of our findings. Therefore, we introduce heterogeneity in the degree distribution within the reference model by means of the previously defined (Binary) Configuration Model (BCM). This way, it is actually possible to detect only the non-trivial dependencies, therefore discarding all the overlaps simply due to the possibly high density of the layers, that would otherwise increase the observed interrelations even if no real correlation is actually present.

This is exactly what happens when the World Trade Network is analyzed. Indeed, as shown in Figure 1.1(c), we find out that a significant amount of the binary overlap observed in this network is actually due to the information included in the degree sequence of the various layers, rather than to a real dependence between layers. This method is therefore able to detect the really meaningful similarity between layers, discarding the trivial overlap caused by the presence, for instance, of nodes having a high number of connections in most of the layers. This non-significant overlap is thus filtered out by our procedure. Such observations clearly show that the Random Graph is not the most proper reference model in order to obtain an appropriate representation of crucial properties of such multilevel systems.

We now note that linear correlations have been used in the literature to produce dendrograms [22, 36]. As we mentioned, the use of linear correlations corresponds to the choice of the Random Graph as null model. Here, we can instead make use of $\mu_{BCM}^{\alpha\beta}$ to implement an improved hierarchical clustering procedure,

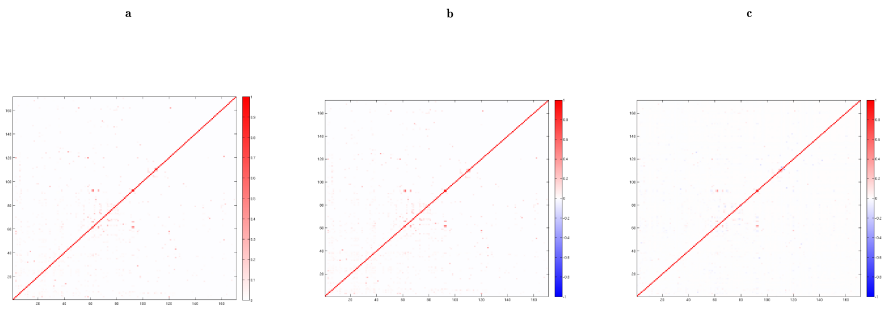

Figure 1.2: **Analysis of the binary multiplexity between layers of the European Airport Network.** Color-coded matrices with entries given by $m_b^{\alpha\beta}$ (a), $\mu_{RG}^{\alpha\beta}$ (b) and $\mu_{BCM}^{\alpha\beta}$ (c) for any pair of layers (airlines).

as reported in the appendix.

A completely different behaviour can be observed for the European Airport System. Indeed, low values of multiplexity observed for such a network (Figure 1.2(a)) illustrate nearly no overlap between most of the layers: this highlights the well-known tendency of airline companies to avoid superpositions between routes with other airlines. In Figure 1.2(b) we show the residual correlations obtained after the application of the Random Graph: almost no difference can be perceived with respect to Figure 1.2(a), since the expected overlap in this case is very small, due to the very low densities of the various layers. We should point out that the Random Graph is not a proper reference model for this real-world network, since the assumption of uniformity in the degree of the different nodes (i.e., airports) is actually far from the observed structure of such a system, as we will highlight later. Nevertheless, in Figure 1.2(c) we show that, at first glance, the adoption of the Configuration Model does not look strictly required when the European Airport Network is considered, except for a more suitable mathematical approach, since the overall matrix looks apparently similar to the previous Figure 1.2(b). However, the presence of a larger number of negative values of multiplexity and the simultaneous disappearance of most of the significantly high values highlight once more the anti-correlated character of such a system, and this crucial structural property of the airport multiplex network was not fully revealed by the application of the Random Graph.

In this case, a dendrogram designed form matrices reported in Figure 1.2 would not be meaningful, since most of the layers meet at a single root level, due to the very low correlation observed between them.

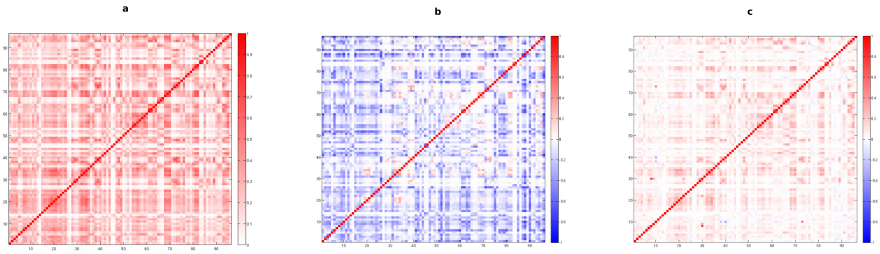

Figure 1.3: **Analysis of the weighted multiplexity between layers of the World Trade Multiplex in 2011.** Color-coded matrices with entries given by $m_w^{\alpha\beta}$ (a), $\mu_{WRG}^{\alpha\beta}$ (b) and $\mu_{WCM}^{\alpha\beta}$ (c) for any pair of layers (commodities).

1.3.2 Weighted analysis

Since the International Trade Network is represented by a weighted multiplex, the analysis of weighted overlaps between layers of that system can be performed, in order to obtain more refined information about the dependencies between different classes of commodities. We should indeed point out that, for the World Trade Web, while the binary overlaps provided by (1.2) only supply information about the dependencies between the topologies of the various layers representing trade in different commodities, the weighted multiplexity defined in (1.3) is able to detect patterns of correlation between quantities of imported and exported classes of items. In this perspective, observing high correlations is therefore more unlikely. This is due, mathematically, to the functional form of the definition of the multiplexity given in (1.3), which is significantly dependent on the balance between weights of the corresponding links in different layers; such a property, therefore, tends to assign higher correlations to pairs of commodities characterized by similar global amount of trade, as we want.

In Figure 1.3(a) we show the color-coded matrix $\mathbf{M}_w$ associated to the raw values of weighted multiplexity as observed in the International Trade Network: clear dependencies between different layers are still present, but a comparison with its corresponding binary matrix $\mathbf{M}_b$(shown in Figure 1.1(a)) explicitly reveals that, while some pairs of layers are significantly overlapping, several pairs of commodities are now actually uncorrelated, as expected when the weights of the links are taken into account. In order to provide information about the relation between the observed dependencies between layers and the expected ones under a given benchmark, as a first estimate, we calculate $\mu_{WRG}^{\alpha\beta}$, therefore considering the corresponding Weighted Random Graph (WRG) as a reference for our real-world network. Our findings show, in Figure 1.3(b), a strongly uncorrelated behavior associated to most of the pairs of commodities, in contrast with our intuitive

expectations based on the results obtained in the binary case.

We then compare the observed multiplexity with its expected values under the Weighted Configuration Model (WCM). Results, shown in Figure 1.3(c), exhibit a completely different behavior with respect to Figure 1.3(b), thus highlighting once more the importance of taking into account the heterogeneity in the weight and degree distributions within the considered null model. Indeed, we observe that, exploiting this more suitable reference, several pairs are still correlated, even in the weighted case, some of them are actually uncorrelated, as expected by looking at the corresponding binary matrix (Figure 1.1(c)), and only a few, with respect to the Weighted Random Graph case, remain anti-correlated. In general, however, the dependencies between layers in the weighted case are less noticeable, as we can see from a comparison between the matrices shown in Figures 1.1(c) and 1.3(c).

1.3.3 Hubs distribution

The different behaviours observed for the two considered multiplexes can be, at least partly, explained in terms of distribution of the hubs across layers. As we show in Figure 1.4(a) and 1.4(b), generally any two layers of the World Trade Multiplex exhibit the same set of hubs (which in this particular case are represented by the richest and most industrialized countries). Indeed, the two network layers plotted in the Figure are, already from visual inspection, very similar to each other. This property produces a high dependence between layers, since the overlap is increased by the multiple presence of links in the various layers connecting nodes to the hubs.

It is possible to show that this hubs distribution, leading to the higher overlap between layers, is strongly correlated to the relation existing between the hidden variables x_i associated to each node in the different layers (we provide further details about such variables in appendices). Indeed, as shown in Figure 1.4(c), for the considered pair of layers (but several pairs actually exhibit the same behaviour) such a trend can be clearly represented by a straight line, thus pointing out that nodes with higher x_i in one layer (hence, with higher probability of establishing a link with any other node in that layer) generally also have higher x_i in a different layer.

However, when the European Airport Network is considered, an opposite trend can be observed, thus a clear explanation of the small measured overlap applies; indeed, Figures 1.5(a) and 1.5(b) show that in this case the layers can be approximated to star-like graphs, with a single, largely connected hub and several other poorly connected nodes. Though, the hub is in general different for almost any considered layer, since each airline company is based on a different airport: in the considered pair of layers, hubs are represented by Rome - Fiumicino airport (FCO) for Alitalia and Amsterdam - Schiphol airport (AMS) for KLM. Such a property decreases significantly the overlap between layers, thus leading to the matrices previously shown in Figure 1.2.

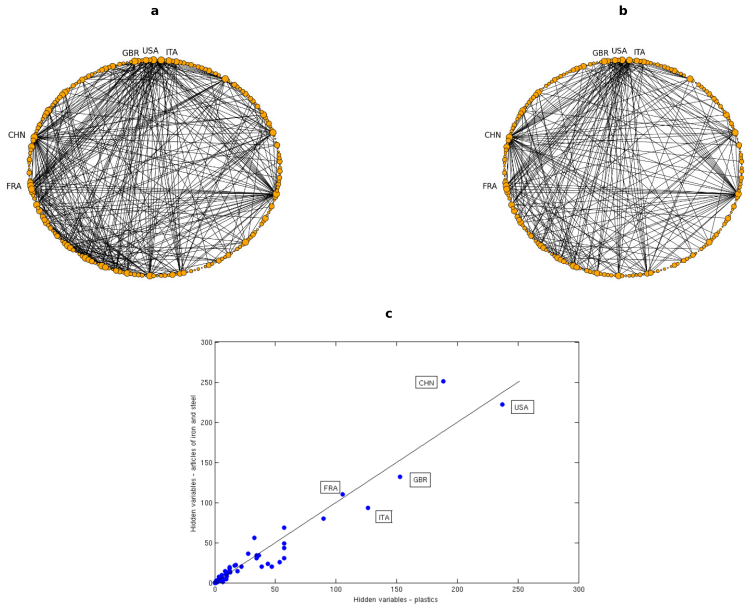

Figure 1.4: **Hubs distribution in the World Trade Multiplex.** Top panels: graphs representing two layers of the system, respectively those associated to trade in plastic (a) and articles of iron and steel (b); nodes represent trading countries; size of a node is proportional to its degree in that layer. Only links associated to a trade larger than 100 millions dollars are reported. Bottom panel: scatter plot of the hidden variables x_i relative to each of the nodes for the same two layers; the black line represents the identity line.

Similar considerations can be done when looking at Figure 1.5(c), where the scatter plot of the hidden variables associated to the nodes in two different layers is shown. We observe that no linear trend can be inferred, since only the two hubs stand out from the bunch of the other airports (which are actually characterized by different values of x_i, even though this cannot be fully appreciated). It is anyway clear that the hub of one layer, characterized by the highest value x_i (hence, with the highest probability of establishing a link with any other node in that layer) is a poorly connected node in a different layer, being characterized by a small value of x_i.

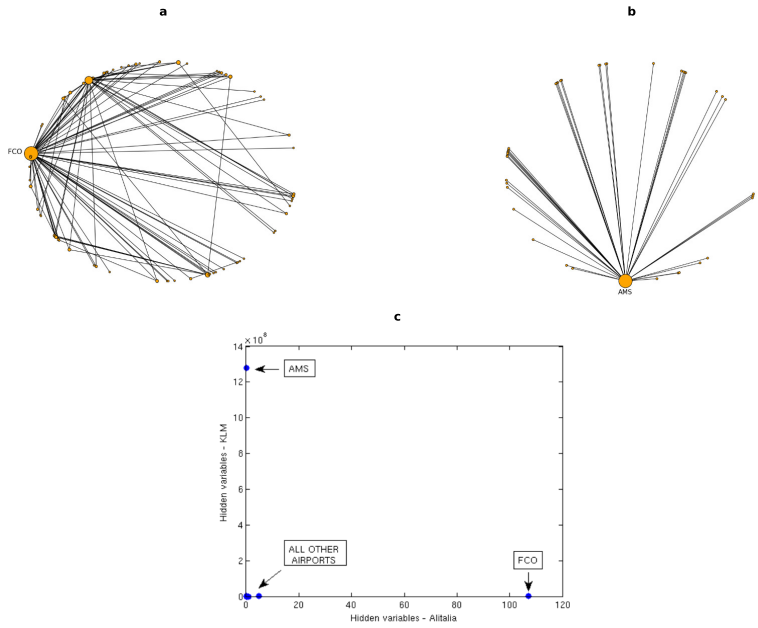

Figure 1.5: **Hubs distribution in the European Airport multiplex.** Top panels: graphs representing two layers of the system, respectively those associated to Alitalia airline (a) and KLM airline (b); nodes represent european airports; size of a node is proportional to its degree in that layer. All the observed links are reported. Bottom panel: scatter plot of the hidden variables x_i relative to each of the nodes for the same two layers.

1.4 Discussion

In the last few years the multiplex approach has revealed itself as a useful framework to study several real-world systems characterized by elementary units linked by different kinds of connection. In this context, we have introduced new measures aiming at analyzing dependencies between layers of the network, both for binary and weighted multi-graphs. We showed that our measures of multiplexity are able to extract crucial information from both sparse and dense networks by testing it on different real-world multi-layer systems. We clearly found that a distinction can be done based on the degree of overlap between links in different layers. For instance, we showed that some multiplexes exhibit small overlap between links in different layers, since just a limited number of nodes are active in many layers, while most of them participate to one or few layers. However, for other systems, such as the World Trade Multiplex Network (WTM), most of the pairs of nodes are connected in several layers, so that such multiplexes exhibit large overlap

between layers. Furthermore, we found that the multiplexity can also provide interesting information about the distribution of hubs across the various layers; indeed, systems characterized by nodes having many connections in most of the layers, such as the WTM, tend to show higher values of raw binary multiplexity. On the other hand, in other networks exhibiting values of multiplexity for most of the pairs of layers close to 0, a node with a low degree in a given layer may represent a hub in a different layer: the European Airport Network is a clear prototype of such systems.

Our findings suggest that adopting proper null models for multi-level networks, enforcing constraints taking into account dependencies between layers, is required in order to suitably model such real-world systems.

Further research in this direction, including the studies reported in the following chapters, will hopefully provide a better understanding of the role of local constraints in real-world multi-level systems.

Appendix

1.A Uncorrelated null models for multi-layer networks

We define the multiplex $\overrightarrow{G} = (G^1, G^2, \ldots, G^M)$ as the superposition of M layers G^α ($\alpha = 1, 2, \ldots, M$), each of them represented by a (possibly weighted) network sharing the same set of N nodes with the other ones, although we do not require that all the vertices are active in each layer. Therefore, multiplex ensembles can be defined by associating a probability $P(\overrightarrow{G})$ to each multi-network, so that the entropy S of the ensemble is given by:

$$S = -\sum_{\overrightarrow{G}} P(\overrightarrow{G}) \ln P(\overrightarrow{G}) \tag{1.7}$$

It is then possible to design null models for multi-level networks by maximizing such an entropy after the enforcement of proper constraints. In this context, previous works, mentioned in the main text, introduced the concepts of correlated and uncorrelated multiplex ensembles, based on the possibility to introduce correlations between layers within the null models. In particular, for an uncorrelated ensemble the probability of a given multiplex can be factorized into the probabilities of each single-layer network G^α belonging to that multiplex, as the links in any two layers α and β are uncorrelated; thus, it is given by:

$$P(\overrightarrow{G}) = \prod_{\alpha=1}^{M} P^\alpha(G^\alpha) \tag{1.8}$$

Instead, if we want to take into account correlations between layers, the previous relation (1.8) does not hold.

As stated in the main text, since our purpose is precisely that of measuring such correlations, we are going to consider the former type of ensemble, in order to define a null model for the real system so that it is possible to compare the observed correlations with reference models where the overlap between layers is actually randomized and, at the same time, important properties of the real network are preserved.

In this perspective, therefore, the definition of proper null models for the considered multiplex reduces to the definition of an indipendent null model for any layer of the system. In order to do this, we take advantage of the concept of canonical network ensemble, or exponential random graph, i.e. the randomized family of graphs satisfying a set of constraints on average. In this context the resulting randomized graph preserves only part of the topology of the considered real-world network and is entirely random otherwise, thus it can be employed as a proper reference model.

However, fitting such previously defined models to real datasets is hard, since it is usually computationally demanding as it requires the generation of many randomized networks whose properties of interest have to be measured. In this perspective, we make use of a fast and completely analytical maximum-entropy method, combined with the maximization of the likelihood function, which provides the exact probabilities of occurrence of random graphs with the same average constraints as the real network. From such probabilities it is then possible to compute the expectation values of the properties we are interested in, such as the average link probability or the average weight associated to the link established between any two nodes. This procedure is general enough to be applied to any network, including the denser ones, and does not require the sampling of the configuration space in order to compute average values of the quantities of interest. While the adoption of such a method is not strictly required when dealing with global constraints like the total number of links observed in a network, it becomes crucial when facing the problem of enforcing local constraints such as the degree sequence or the strength sequence.

Indeed, so far the most widely used graph null model has been represented by the Random Graph (RG), which enforces on average as constraint the expected number of links in the network. Such model, therefore, provides a unique expected probability p^α that a link between any two nodes is established in layer α: however, as we said, such a reference model completely discards any kind of heterogeneity in the degree distributions of the layers, resulting in graphs where each node has on average the same number of connections, inconsistently with the observed real networks. Thus, the probability of connection between any two nodes in layer α is uniformly given by:

$$p^\alpha = \frac{L^\alpha}{N(N-1)/2} \tag{1.9}$$

where L^α is the total number of links actually observed in layer α:

$$L^\alpha = \sum_{i<j} a_{ij}^\alpha \tag{1.10}$$

and $a_{ij}^\alpha = 0, 1$ depending on the presence of the link between nodes i and j in layer α.

Similar considerations apply to weighted networks and the related Weighted Random Graph (WRG), i.e. the straightforward extension of the previous Random Graph to weighted systems; in such a null model, the probability of having a link of weight w between two nodes i and j is independent from the choice of the nodes, and it is given by the following geometric distribution:

$$P(w^\alpha) = (p^\alpha)^w (1 - p^\alpha) \tag{1.11}$$

where the maximum-likelihood method shows that the optimal value of the parameter p^α is given by:

$$p^\alpha = \frac{2W^\alpha}{N(N-1) + 2W^\alpha} \tag{1.12}$$

with W^α defined as the total weight observed in layer α (w_{ij}^α is the weight associated to the link between nodes i and j in the same layer):

$$W^\alpha = \sum_{i<j} w_{ij}^\alpha \tag{1.13}$$

Similarly to the corresponding binary random graph, also this kind of null models discards the simultaneous presence of nodes characterized by high and low values of the strengths (that is, by a high or low sum of the weights associated to links incident on that node).

To take into account the heterogeneity of the real-world networks within the null models, in the unweighted case we consider the Binary Configuration Model (BCM), i.e. the ensemble of networks satisfying on average a given degree sequence. Since we make use of the canonical ensembles, it is possible to obtain from the maximum-likelihood method each probability p_{ij}^α that nodes i and j are connected in layer α (notice that such value p_{ij}^α is basically the expectation value of a_{ij}^α under the chosen Configuration Model). Similarly, for weighted graphs the Weighted Configuration Model (WCM) can be defined: here, the enforced constraint is represented by the strength sequence as observed in the real-world network. In this view, the likelihood maximization provides the expectation value of each weight w_{ij}^α for any pair of nodes i and j as supplied by the Weighted Configuration Model. It is worth noticing that enforcing the degree sequence (respectively, the strength sequence in the weighted case) automatically leads to the design of a null model where also the total number of links (respectively, the total weight) of the network is preserved. In the following section, we will provide

equations generalizing equations (1.9) and (1.12), whose solution allows then to derive the analytical expression of the expected link probability p_{ij}^{α} and, in the weighted case, the expected link weight w_{ij}^{α}. In order to do this, we make use of a set of N auxiliary variables x_i^{α} for any layer α, which are proportional to the probability of establishing a link between a given node i and any other node (or, respectively for the weighted case, establishing a link characterized by a given weight), being therefore directly informative on the expected probabilities p_{ij}^{α} (or, respectively, the expected weights w_{ij}^{α}).

1.B Maximum-likelihood method

We now briefly explain the maximum-likelihood method (more details about this technique can be found in the appendix associated to Chapter 2, where it is also extended to the directed case). In the binary case, when the observed degree sequence represents the property that we want to preserve (i.e., in the so-called configuration model), the method reduces to finding the solution to following set of N coupled nonlinear equation, independently for each layer $\alpha = 1, 2, \ldots, M$:

$$\sum_{i<j} \frac{x_i^{\alpha} x_j^{\alpha}}{1 + x_i^{\alpha} x_j^{\alpha}} = k_i^{\alpha} \quad \forall i = 1, 2, \ldots, N \tag{1.14}$$

where k_i^{α} is the observed degree of node i in layer α and the unknown variables of the equation are the so-called N hidden variables associated to that layer.

Thus, the expected link probability p_{ij}^{α} is given by, for any pair of nodes (i, j) in any layer α:

$$p_{ij}^{\alpha} = \frac{x_i^{\alpha} x_j^{\alpha}}{1 + x_i^{\alpha} x_j^{\alpha}} \tag{1.15}$$

which is therefore the generalization of the expression (1.9) in the previous section. We can therefore see that such hidden variables x_i^{α} are proportional to the expected link probability p_{ij}^{α} in a given layer α: a higher value of x_i^{α} will correspond to a higher expected probability of observing a link between i and any other node $j \neq i$, and vice-versa.

Similarly, for weighted multiplexes, we can enforce the strength sequence observed in a real network on a network ensemble, thus designing a proper null model where the strength sequence of the considered real-world network is preserved, while the other properties are randomized. In this context, the maximum-likelihood method for weighted graphs reduces to solving a set of N coupled nonlinear equations. For any node i in any layer α, we have:

$$\sum_{i<j} \frac{x_i^{\alpha} x_j^{\alpha}}{1 - x_i^{\alpha} x_j^{\alpha}} = s_i^{\alpha} \tag{1.16}$$

where s_i^α is the observed strength of node i in layer α and the unknown variables of the equation are, again, the N hidden variables associated to the considered layer.

Thus, the expected link weight w_{ij}^α is given by, for any pair of nodes (i, j):

$$w_{ij}^\alpha = \frac{x_i^\alpha x_j^\alpha}{1 - x_i^\alpha x_j^\alpha} \tag{1.17}$$

hence generalizing the corresponding equation (1.12). In this case, the computed hidden variables x_i^α are proportional to the expected link weight w_{ij}^α in a given layer α; a higher value of x_i^α will therefore correspond to a higher expected link weight between i and any other node $j \neq i$, and vice-versa.

We can now derive the expression for the expectation values of the binary and weighted multiplexity defined in the main text.

1.C Binary multiplexity

When the unweighted networks are considered we have defined the "absolute" binary multiplexity between any two layers α and β as:

$$m_b^{\alpha\beta} = \frac{2\sum_{i<j} \min\{a_{ij}^\alpha, a_{ij}^\beta\}}{L^\alpha + L^\beta} \tag{1.18}$$

with the previously introduced notation.

As we said, this quantity is informative only after a comparison with the value of binary multiplexity obtained when considering a null model. We have therefore introduced the following transformed or rescaled quantity:

$$\mu_b^{\alpha\beta} = \frac{m_b^{\alpha\beta} - \langle m_b^{\alpha\beta}\rangle}{1 - \langle m_b^{\alpha\beta}\rangle} \tag{1.19}$$

where $m_b^{\alpha\beta}$ is the value measured for the observed real-world multiplex and $\langle m_b^{\alpha\beta}\rangle$ is the value expected under the chosen null model. We will show in the next section that, when the Random Graph is considered as a null model, the previous quantity (1.19) is actually the correlation coefficient between the entries of the adjacency matrix referred to any two layers α and β of a multi-level graph.

We should point out that the raw intra-layer multiplexity $m_b^{\alpha\alpha}$ always leads to a measured value equal to 1, representing complete similarity between any layer and itself. However, the rescaled intra-layer multiplexity $\mu_{BCM}^{\alpha\alpha}$ actually leads to an indeterminate value; therefore, we choose to set this value by construction equal to 1 too, for sake of clarity.

In order to compute $\mu_b^{\alpha\beta}$ we should then calculate the expected multiplexity under the chosen null model, that is:

$$\langle m_b^{\alpha\beta}\rangle = \frac{2\sum_{i<j}\langle \min\{a_{ij}^\alpha, a_{ij}^\beta\}\rangle}{\langle L^\alpha\rangle + \langle L^\beta\rangle} \tag{1.20}$$

However, since both the considered null models preserve the average number of links in each layer as constraint, we have just to evaluate the analytical expression for the expected value of the minimum of two variables. In the unweighted case, this is easy because it reduces to the evaluation of the expected minimum between two indipendent, binary variables. In particular, when the Configuration Model is considered (the extension to the Random Graph is straightforward), the probability that a link exists between nodes i and j is given by the mass probability function of a Bernoulli-distributed variable:

$$P(a_{ij}^{\alpha}) = p_{ij}^{a_{ij}}(1-p_{ij})^{(1-a_{ij})} \tag{1.21}$$

Therefore, we have for the configuration model:

$$\begin{aligned}
\langle \min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}\rangle_{BCM} &= \sum_{a_{ij}^{\alpha}, a_{ij}^{\beta}} \min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\} P\left(\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}\right) = \\
&= 0 \cdot P\left(\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\} = 0\right) + 1 \cdot P\left(\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\} = 1\right) = \\
&= P\left(\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\} = 1\right) = \\
&= P\left(a_{ij}^{\alpha} = 1\right) P\left(a_{ij}^{\beta} = 1\right) = \\
&= p_{ij}^{\alpha} p_{ij}^{\beta}
\end{aligned} \tag{1.22}$$

and similarly for the Random Graph:

$$\langle \min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}\rangle_{RG} = p^{\alpha} p^{\beta} \tag{1.23}$$

where we define p^{α} as the fraction of links actually present in that layer, as we have already done before:

$$p^{\alpha} = \frac{L^{\alpha}}{N(N-1)/2} \tag{1.24}$$

It is now possible to compute the analytical expression for the rescaled multiplexity. We obtain for the Random Graph:

$$\mu_{RG}^{\alpha\beta} = \frac{2\sum_{i<j}\left(\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\} - p^{\alpha}p^{\beta}\right)}{\sum_{i<j}\left(a_{ij}^{\alpha} + a_{ij}^{\beta} - 2p^{\alpha}p^{\beta}\right)} \tag{1.25}$$

and for the Binary Configuration Model:

$$\mu_{BCM}^{\alpha\beta} = \frac{2\sum_{i<j}\left(\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\} - p_{ij}^{\alpha}p_{ij}^{\beta}\right)}{\sum_{i<j}\left(a_{ij}^{\alpha} + a_{ij}^{\beta} - 2p_{ij}^{\alpha}p_{ij}^{\beta}\right)} \tag{1.26}$$

1.C.1 Binary multiplexity: z-scores

As we have already said, such rescaled quantities provide proper information about the similarity between layers of a multiplex, by evaluating the dependencies measured in a real network with respect to what we would expect, on average, for an ensemble of multi-level networks sharing only some of the topological properties of the observed one. However, we cannot understand, from the obtained values of multiplexity itself, whether the observed value of $m_b^{\alpha\beta}$ is actually compatible with the expected one, as $\mu_{BCM}^{\alpha\beta}$ (and the correspondig value related to the Random Graph) does not provide any information about the standard deviation associated to the expected value of multiplexity.

In order to solve this issue, we introduce the z-score associated to the previously defined multiplexity:

$$z\left[m_b^{\alpha\beta}\right] = \frac{m_b^{\alpha\beta} - \langle m_b^{\alpha\beta}\rangle}{\sigma\left[m_b^{\alpha\beta}\right]} \tag{1.27}$$

where $m_b^{\alpha\beta}$ is the measured multiplexity between a given pair of layers on the real-world network, $\langle m_b^{\alpha\beta}\rangle$ is the value expected under the chosen null model and $\sigma[m_b^{\alpha\beta}]$ is the related standard deviation. The z-score, therefore, shows by how many standard deviations the observed value of multiplexity differs with respect to the expected one for any pair of layers. In particular, in the binary case such a quantity becomes:

$$z\left[m_b^{\alpha\beta}\right] = \frac{\sum_{i<j}\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\} - \sum_{i<j}\langle min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}\rangle}{\sigma\left[\sum_{i<j}\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}\right]} \tag{1.28}$$

Interestingly, not only the expected value, but even the standard deviation can be calculated analytically. Indeed:

$$\sigma^2\left[\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}\right] = \langle\min^2\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}\rangle - \langle\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}\rangle^2 \tag{1.29}$$

Exploiting again the binary character of the two indipendent variables a_{ij}^{α} and a_{ij}^{β}, the expected value of the square of the minimum becomes for the Configuration Model:

$$\begin{aligned}
\langle\min^2\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}\rangle_{BCM} &= \sum_{a_{ij}^{\alpha}, a_{ij}^{\beta}} \min^2\{a_{ij}^{\alpha}, a_{ij}^{\beta}\} P\left(\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}\right) = \\
&= 0\cdot P\left(\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\} = 0\right) + 1\cdot P\left(\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\} = 1\right) = \\
&= P\left(\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\} = 1\right) = \\
&= P\left(a_{ij}^{\alpha} = 1\right) P\left(a_{ij}^{\beta} = 1\right) = \\
&= p_{ij}^{\alpha} p_{ij}^{\beta}
\end{aligned} \tag{1.30}$$

Therefore, the standard deviation, required in order to evaluate the z-score associated to the multiplexity, is given by:

$$\sigma\left[\sum_{i<j}\min\{a_{ij}^{\alpha},a_{ij}^{\beta}\}\right]=\sqrt{\sum_{i<j}\left[p_{ij}^{\alpha}p_{ij}^{\beta}-\left(p_{ij}^{\alpha}p_{ij}^{\beta}\right)^{2}\right]} \tag{1.31}$$

The analytical value of the z-score related to the binary multiplexity, when the Configuration Model is taken into account, is then:

$$z_{BCM}^{\alpha\beta}=\frac{\sum_{i<j}\min\{a_{ij}^{\alpha},a_{ij}^{\beta}\}-\sum_{i<j}p_{ij}^{\alpha}p_{ij}^{\beta}}{\sqrt{\sum_{i<j}\left[p_{ij}^{\alpha}p_{ij}^{\beta}-\left(p_{ij}^{\alpha}p_{ij}^{\beta}\right)^{2}\right]}} \tag{1.32}$$

Extending such results to the Random Graph is immediate, since everything reduces to a change in the definition of the probability of observing a link between any given pair of nodes in each layer. Hence, the z-score associated to the binary multiplexity according to the binary Random Graph is given by:

$$z_{RG}^{\alpha\beta}=\frac{\sum_{i<j}\min\{a_{ij}^{\alpha},a_{ij}^{\beta}\}-\sum_{i<j}p^{\alpha}p^{\beta}}{\sqrt{\sum_{i<j}\left[p^{\alpha}p^{\beta}-\left(p^{\alpha}p^{\beta}\right)^{2}\right]}} \tag{1.33}$$

where we used the previous definitions for p^{α} and p^{β}.

We should point out that such z-scores should in principle be defined only if the associated property (in this case, $\mu_{BCM}^{\alpha\beta}$) is normally distributed; nevertheless, even if such assumption does not occur, they provide important information about the consistency between observed and randomized values. It is worth saying that these z-scores provide a different kind of information with respect to the previous multiplexities. Mathematically, the only correlation between, for example, $\mu_{BCM}^{\alpha\beta}$ and the corresponding $z_{BCM}^{\alpha\beta}$ is the sign concordance; furthermore, the z-score is useful in order to understand whether, for instance, values of multiplexity close to 0 are actually comparable with 0, so that we can consider those two layers as uncorrelated, or they are instead significantly unexpected, although very small. In this perspective, we should not expect a particular relation between such two variables $\mu_{BCM}^{\alpha\beta}$ and $z_{BCM}^{\alpha\beta}$ (or, respectively, $\mu_{RG}^{\alpha\beta}$ and $z_{RG}^{\alpha\beta}$).

1.C.2 Relationship with the correlation coefficient

A possible definition of correlation between layers of a multiplex builds on the standard correlation coefficient:

$$Corr\{a_{ij}^{\alpha},a_{ij}^{\beta}\}=\frac{\langle a_{ij}^{\alpha}a_{ij}^{\beta}\rangle-\langle a_{ij}^{\alpha}\rangle\langle a_{ij}^{\beta}\rangle}{\sigma_{\alpha}\sigma_{\beta}} \tag{1.34}$$

Hence, a value of correlation equal to 0 represents a pair of uncorrelated layers only if the probability distributions of a_{ij}^{α} and a_{ij}^{β} are independent from the chosen node, that is, if all the edges in a certain layer are statistically equivalent. However, this leads to a probability of establishing a given link which is common to each pair of nodes, and this is the assumption behind the Random Graph.

In this context, it is then possible to show that, when the Binary Random Graph is taken into consideration, our novel measure of multiplexity can be reduced to the usual definition of correlation coefficient. Indeed, we have:

$$\begin{aligned}\langle a_{ij}^{\alpha} a_{ij}^{\beta} \rangle &= \frac{2\sum_{i<j} a_{ij}^{\alpha} a_{ij}^{\beta}}{N(N-1)} = \\ &= \frac{2\sum_{i<j} \min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}}{L^{\alpha}+L^{\beta}} \frac{L^{\alpha}+L^{\beta}}{N(N-1)} = \\ &= m_b^{\alpha\beta} \frac{L^{\alpha}+L^{\beta}}{N(N-1)} \end{aligned} \tag{1.35}$$

Moreover, the average value of a_{ij}^{α} over all the pairs of nodes in layer α is given by:

$$\langle a_{ij}^{\alpha} \rangle = \frac{2L^{\alpha}}{N(N-1)} \tag{1.36}$$

and similarly for layer β:

$$\langle a_{ij}^{\beta} \rangle = \frac{2L^{\beta}}{N(N-1)} \tag{1.37}$$

Hence,

$$\langle a_{ij}^{\alpha} \rangle \langle a_{ij}^{\beta} \rangle = \frac{4L^{\alpha}L^{\beta}}{N^2(N-1)^2} \tag{1.38}$$

On the contrary, the expected value of multiplexity under random graph is given by:

$$\begin{aligned}\langle m_b^{\alpha\beta} \rangle &= \frac{2\sum_{i<j} p^{\alpha} p^{\beta}}{L^{\alpha}+L^{\beta}} = \\ &= \frac{N(N-1)}{L^{\alpha}+L^{\beta}} \frac{2L^{\alpha}}{N(N-1)} \frac{2L^{\beta}}{N(N-1)} = \\ &= \frac{1}{N(N-1)} \frac{4L^{\alpha}L^{\beta}}{L^{\alpha}+L^{\beta}} \end{aligned} \tag{1.39}$$

There is therefore a direct relation between $\langle a_{ij}^{\alpha} \rangle \langle a_{ij}^{\beta} \rangle$ and $\langle m_b^{\alpha\beta} \rangle$:

$$\begin{aligned}\langle a_{ij}^{\alpha} \rangle \langle a_{ij}^{\beta} \rangle &= \frac{4L^{\alpha}L^{\beta}}{N^2(N-1)^2} = \\ &= \langle m_b^{\alpha\beta} \rangle \frac{L^{\alpha}+L^{\beta}}{N(N-1)} \end{aligned} \tag{1.40}$$

Furthermore, we need to derive the expression for the standard deviation σ_α and σ_β:

$$\begin{aligned}\sigma_\alpha &= \sqrt{\langle \left(a_{ij}^\alpha\right)^2\rangle - \langle a_{ij}^\alpha\rangle^2} = \\ &= \sqrt{\langle a_{ij}^\alpha\rangle \left(1 - \langle a_{ij}^\alpha\right)} = \\ &= \sqrt{\frac{2L^\alpha}{N(N-1)}\left[1 - \frac{2L^\alpha}{N(N-1)}\right]}\end{aligned} \tag{1.41}$$

and analogously for β. Hence, the correlation coefficient between a_{ij}^α and a_{ij}^β is given by:

$$\begin{aligned}Corr\{a_{ij}^\alpha, a_{ij}^\beta\} &= \frac{\frac{L^\alpha+L^\beta}{N(N-1)} m_b^{\alpha\beta} - \frac{L^\alpha+L^\beta}{N(N-1)}\langle m_b^{\alpha\beta}\rangle}{\frac{2}{N(N-1)}\sqrt{L^\alpha L^\beta\left(1-\frac{2L^\alpha}{N(N-1)}\right)\left(1-\frac{2L^\beta}{N(N-1)}\right)}} \\ &= \frac{\left(L^\alpha + L^\beta\right)\left(m_b^{\alpha\beta} - \langle m_b^{\alpha\beta}\rangle\right)}{2\sqrt{L^\alpha L^\beta\left(1-\frac{2L^\alpha}{N(N-1)}\right)\left(1-\frac{2L^\beta}{N(N-1)}\right)}}\end{aligned} \tag{1.42}$$

It is therefore clear that, apart from a different normalization factor (depending on L^α and L^β), our definition of binary rescaled multiplexity, when the Random Graph is considered as null model, reduces to the usual correlation coefficient (1.34).

However, such a property does not hold when a different reference model, such as the Configuration Model, is considered.

1.D Weighted multiplexity

In the main text, we have also extended the previous definitions to weighted multiplex networks. We have defined the "absolute" weighted multiplexity as:

$$m_w^{\alpha\beta} = \frac{2\sum_{i<j}\min\{w_{ij}^\alpha, w_{ij}^\beta\}}{W^\alpha + W^\beta} \tag{1.43}$$

where w_{ij}^α represents the weight of the link between nodes i and j in layer α and W^α is the total weight related to the links in that layer.

Furthermore, we have defined the following transformed or rescaled quantity:

$$\mu_w^{\alpha\beta} = \frac{m_w^{\alpha\beta} - \langle m_w^{\alpha\beta}\rangle}{1 - \langle m_w^{\alpha\beta}\rangle} \tag{1.44}$$

where $\langle m_w^{\alpha\beta}\rangle$ is the value measured for the observed real-world network and $\langle m_w^{\alpha\beta}\rangle$ is the value expected under the considered reference model. Again, the sign of

$\mu_w^{\alpha\beta}$ is then directly informative about the weighted dependency existing between layers.

In this context, the expected value of weighted multiplexity is given by:

$$\langle m_w^{\alpha\beta} \rangle = \frac{2\sum_{i<j}\langle \min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}\rangle}{\langle W^{\alpha}\rangle + \langle W^{\beta}\rangle} \tag{1.45}$$

However, since both the Weighted Random Graph and the Weighted Configuration Model preserve the average total weight associated to the links in each layer as constraint, also in this case we just need to evaluate the analytical expression for the expected value of the minimum of two variables; the only difference with respect to the binary description is related to a change in the underlying probability distribution.

Indeed, in the weighted case, when the Weighted Configuration Model is considered (again, the extension to the Weighted Random Graph is straightforward) such variables are distributed according to a geometrical distribution:

$$P(w_{ij}^{\alpha}) = p_{ij}^{w_{ij}^{\alpha}}(1 - p_{ij}^{\alpha}) \tag{1.46}$$

In order to quantify such an expectation value, we exploit the cumulative distribution of the minimum between the considered variables:

$$\begin{aligned} P\left(\min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\} \geq w\right) &= P\left(w_{ij}^{\alpha} \geq w\right) P\left(w_{ij}^{\beta} \geq w\right) = \\ &= \left(p_{ij}^{\alpha} p_{ij}^{\beta}\right)^{w} \end{aligned} \tag{1.47}$$

Thus, the expected minimum, under Weighted Configuration Model, becomes:

$$\begin{aligned} \langle \min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}\rangle_{WCM} &= \sum_{w'} w'[P\left(\min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\} \geq w'\right) + \\ &- P\left(\min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\} \geq w' + 1\right)] = \\ &= \sum_{w'} w' \left[\left(p_{ij}^{\alpha} p_{ij}^{\beta}\right)^{w'} - \left(p_{ij}^{\alpha} p_{ij}^{\beta}\right)^{w'+1}\right] = \\ &= \frac{p_{ij}^{\alpha} p_{ij}^{\beta}}{1 - p_{ij}^{\alpha} p_{ij}^{\beta}} \end{aligned} \tag{1.48}$$

and, for the Weighted Random Graph:

$$\langle \min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}\rangle_{WRG} = \frac{p^{\alpha} p^{\beta}}{1 - p^{\alpha} p^{\beta}} \tag{1.49}$$

where we define p^{α}, according to the likelihood maximization, as:

$$p^{\alpha} = \frac{W^{\alpha}}{W^{\alpha} + N(N-1)/2}, \tag{1.50}$$

We can now compute the analytical expression for the rescaled multiplexity, according to both the chosen null models. We obtain for the Weighted Random Graph (WRG):

$$\mu^{\alpha\beta}_{WRG} = \frac{2\sum_{i<j}\left(\min\{w^{\alpha}_{ij}, w^{\beta}_{ij}\} - \frac{p^{\alpha}p^{\beta}}{1-p^{\alpha}p^{\beta}}\right)}{\sum_{i<j}\left(w^{\alpha}_{ij} + w^{\beta}_{ij} - 2\frac{p^{\alpha}p^{\beta}}{1-p^{\alpha}p^{\beta}}\right)} \tag{1.51}$$

and for the Weighted Configuration Model (WCM):

$$\mu^{\alpha\beta}_{WCM} = \frac{2\sum_{i<j}\left(\min\{w^{\alpha}_{ij}, w^{\beta}_{ij}\} - \frac{p^{\alpha}_{ij}p^{\beta}_{ij}}{1-p^{\alpha}_{ij}p^{\beta}_{ij}}\right)}{\sum_{i<j}\left(w^{\alpha}_{ij} + w^{\beta}_{ij} - 2\frac{p^{\alpha}_{ij}p^{\beta}_{ij}}{1-p^{\alpha}_{ij}p^{\beta}_{ij}}\right)} \tag{1.52}$$

with the previously defined notation.

1.D.1 Weighted multiplexity: z-scores

Furthermore, we can extend to the weighted case the analysis of the z-scores associated to the values of multiplexity as defined in (1.44). We can define it in the usual way:

$$z\left[m^{\alpha\beta}_{w}\right] = \frac{\sum_{i<j}\min\{w^{\alpha}_{ij}, w^{\beta}_{ij}\} - \sum_{i<j}\langle\min\{w^{\alpha}_{ij}, w^{\beta}_{ij}\}\rangle}{\sigma\left[\sum_{i<j}\min\{w^{\alpha}_{ij}, w^{\beta}_{ij}\}\right]} \tag{1.53}$$

Since:

$$\sigma^2\left[\min\{w^{\alpha}_{ij}, w^{\beta}_{ij}\}\right] = \langle\min^2\{w^{\alpha}_{ij}, w^{\beta}_{ij}\}\rangle - \langle\min\{w^{\alpha}_{ij}, w^{\beta}_{ij}\}\rangle^2 \tag{1.54}$$

we just have to compute the analytical expression for the expected value of the square of minimum bewteeen w^{α}_{ij} and w^{β}_{ij}. Then, following the same procedure adopted for (1.48) we find:

$$\begin{aligned}
\langle\min^2\{w^{\alpha}_{ij}, w^{\beta}_{ij}\}\rangle_{WCM} &= \sum_{w'}(w')^2\left[P\left(\min\{w^{\alpha}_{ij}, w^{\beta}_{ij}\} \geq w'\right) + \right. \\
&- \left. P\left(\min\{w^{\alpha}_{ij}, w^{\beta}_{ij}\} \geq w'+1\right)\right] = \\
&= \sum_{w'}(w')^2\left[\left(p^{\alpha}_{ij}p^{\beta}_{ij}\right)^{w'} - \left(p^{\alpha}_{ij}p^{\beta}_{ij}\right)^{w'+1}\right] = \\
&= \frac{p^{\alpha}_{ij}p^{\beta}_{ij} + \left(p^{\alpha}_{ij}p^{\beta}_{ij}\right)^2}{\left(1-p^{\alpha}_{ij}p^{\beta}_{ij}\right)^2}
\end{aligned} \tag{1.55}$$

and therefore the standard deviation is:

$$\sigma\left[\sum_{i<j}\min\{w_{ij}^{\alpha},w_{ij}^{\beta}\}\right]=\sqrt{\sum_{i<j}\left[\frac{p_{ij}^{\alpha}p_{ij}^{\beta}+\left(p_{ij}^{\alpha}p_{ij}^{\beta}\right)^{2}}{\left(1-p_{ij}^{\alpha}p_{ij}^{\beta}\right)^{2}}-\frac{\left(p_{ij}^{\alpha}p_{ij}^{\beta}\right)^{2}}{\left(1-p_{ij}^{\alpha}p_{ij}^{\beta}\right)^{2}}\right]} \tag{1.56}$$

Finally, the z-score associated to the weighted multiplexity under Weighted Configuration Model is therefore given by:

$$z_{WCM}^{\alpha\beta}=\frac{\sum_{i<j}\min\{w_{ij}^{\alpha},w_{ij}^{\beta}\}-\sum_{i<j}\frac{p_{ij}^{\alpha}p_{ij}^{\beta}}{1-p_{ij}^{\alpha}p_{ij}^{\beta}}}{\sqrt{\sum_{i<j}\frac{p_{ij}^{\alpha}p_{ij}^{\beta}}{\left(1-p_{ij}^{\alpha}p_{ij}^{\beta}\right)^{2}}}} \tag{1.57}$$

Analogously, we get:

$$z_{WRG}^{\alpha\beta}=\frac{\sum_{i<j}\min\{w_{ij}^{\alpha},w_{ij}^{\beta}\}-\sum_{i<j}\frac{p^{\alpha}p^{\beta}}{1-p^{\alpha}p^{\beta}}}{\sqrt{\sum_{i<j}\frac{p^{\alpha}p^{\beta}}{(1-p^{\alpha}p^{\beta})^{2}}}} \tag{1.58}$$

for the Weighted Random Graph, where we used the previous definitions for p^{α} and p^{β}.

1.E Additional results

As we stated in the main text, in order to have a better understanding of the correlations between layers, it is possible to implement a hierarchical clustering procedure starting from each of the aforementioned multiplexity matrices. However, we have to define a notion of distance between layers, starting from our notion of dependency. We can define a distance $d^{\alpha\beta}$ between any pair of commodities in the following way:

$$d^{\alpha\beta}=\sqrt{\frac{1-\mu_{BCM}^{\alpha\beta}}{2}}. \tag{1.59}$$

where we chose to consider, for instance, the transformed multiplexity under Binary Configuration Model. Hence, the maximum possible distance $d^{\alpha\beta}$ between any two layers is 1 (when layers α and β show multiplexity $\mu_{BCM}^{\alpha\beta}=1$), while the minimum one is 0 (corresponding to $\mu_{BCM}^{\alpha\beta}=-1$). We can therefore represent the layers of the multiplex as the leaves of a taxonomic tree, where highly correlated communities meet at a branching point which is closer to baseline level. In Figure 1.6 we show the dendrogram obtained by applying the Average Linkage Clustering Algorithm to the matrix representing values of multiplexity $\mu_{BCM}^{\alpha\beta}$ for the World Trade Multiplex (WTM). We can see that some groups of similar commodities are

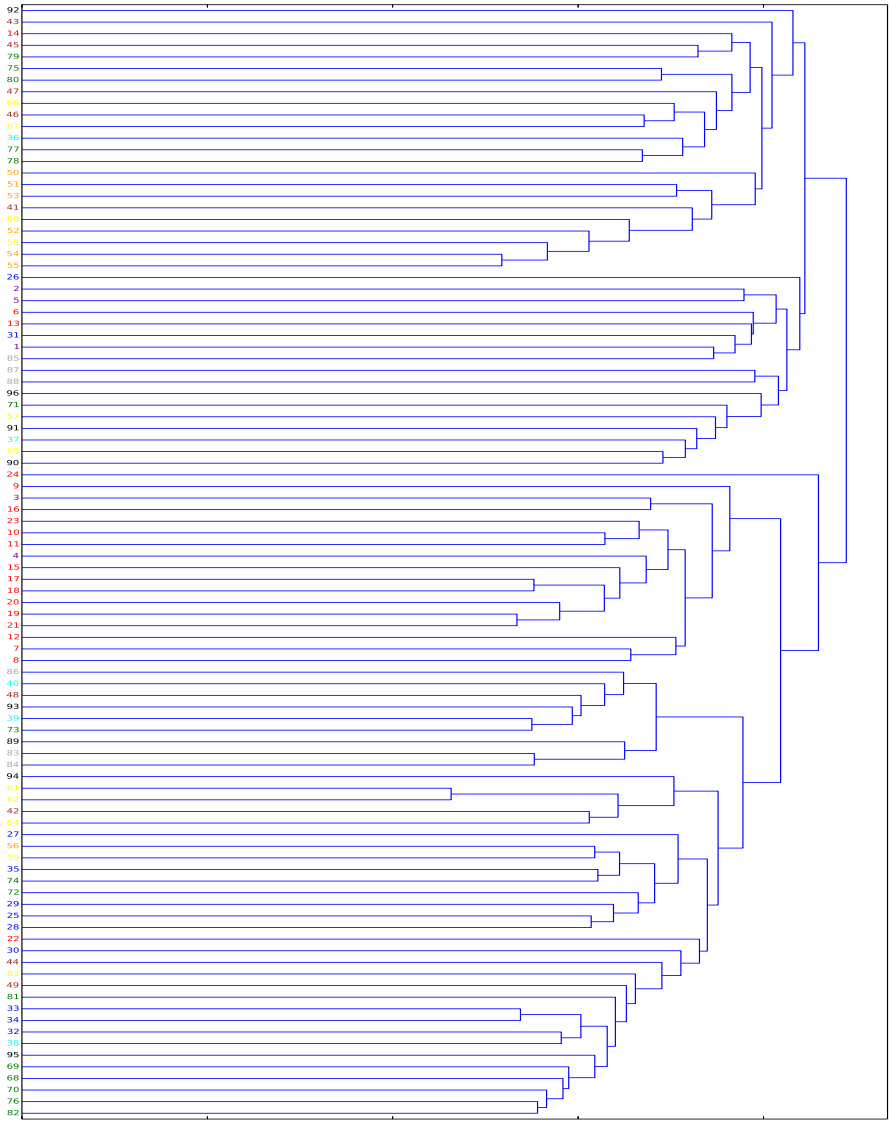

Figure 1.6: **Dendrogram of commodities traded in 2011 in the WTM as obtained applying the Average Linkage Clustering Algorithm to the binary rescaled multiplexity $\mu^{\alpha\beta}_{BCM}$.** Colors of the leaves represent different classes of commodities, as reported in the last Section of this Appendix.

clearly visible (for instance, the group of edible commodities can be easily identified), while in other cases apparently distant commodities are grouped together, pointing out that some unexpected dependencies are present. The dendrogram reported in Figure 1.6 therefore represents a refinement of the taxonomic tree reported in previous studies, where the usual correlation coefficient was employed to define the dependency between layers. Similar dendrograms can be designed starting from the matrices representing values of $\mu_{RG}^{\alpha\beta}$ or weighted multiplexity $\mu_{WRG}^{\alpha\beta}$ and $\mu_{WCM}^{\alpha\beta}$.

Moreover, it is possible to perform the same analysis on the European Airport Network. However, a dendrogram in this case would not be meaningful, since most of the layers meet at a single root level, due to the very low correlation observed between them.

As we said, color-coded multiplexity matrices, as shown in the main text, are useful in order to detect the meaninful dependencies between layers in a multiplex, but they do not supply any information about the discrepancy of the observed values from the corresponding expected ones. Hence, the introduction of suitable z-scores associated to the previously defined quantities is required. Moreover, it is worth reminding that the information provided by (1.26) (respectively (1.25) for the Random Graph) is not necessarily connected to that supplied by (1.32) (respectively, (1.33)) Indeed, while the multiplexity by itself detects the degree of correlation between layers of a multi-level network, the corresponding z-scores reveal how significant those values actually are with respect to our expectations. In Figure 1.7(a) we show, for the World Trade Multiplex (WTM), the scatter

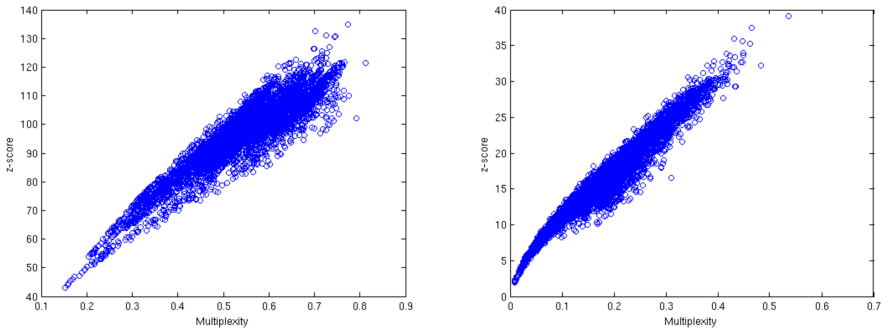

Figure 1.7: **Significance of the binary multiplexity values for the World Trade Multiplex.** Scatter plots of binary multiplexity values $\mu_b^{\alpha\beta}$ vs the corresponding z-score for each pair of layers, respectively for Random Graph (a) and Binary Configuration Model (b), for the WTM.

plot of the values of binary multiplexity versus the corresponding z-scores, after comparing the observed values with the expected ones under Random Graph. We show that observed very large values of z-scores reveal a high significance of

the previously obtained overlaps; such a consideration therefore points out that even the pairs of layers showing low (but positive) values of multiplexity cannot actually be considered as uncorrelated. Furthermore, a clear correlation between $\mu_{RG}^{\alpha\beta}$ and $z_{RG}^{\alpha\beta}$ can be observed, thus large values of binary multiplexity correspond to large z-scores, and vice-versa.

Similar considerations can be done when the Binary Configuration Model is considered as a benchmark. Indeed, as we show in Figure 1.7(b), a large correlation between $\mu_{BCM}^{\alpha\beta}$ and $z_{BCM}^{\alpha\beta}$ is still present when we consider the WTM; moreover, since almost all the z-scores are higher than the widely used critical value $z_{BCM}^{*} = 2$ (so that almost no pair of layers shows a multiplexity lying within 2 standard deviations form the expected value), we highlight that most of the pairs therefore exhibit unexpectedly high correlations with respect to the corresponding average value obtained when randomizing the real-world layers according to the Configuration Model, similarly to what we found before for the Random Graph.

However, if we look at the absolute values of such z-scores, we observe that the significance of the values of multiplexity under Random Graph ($\mu_{RG}^{\alpha\beta}$) is generally much higher than that measured under Binary Configuration Model ($\mu_{BCM}^{\alpha\beta}$). This property, which will still be true in the following Figures, is actually not surprising, since the Configuration Model enforces more constraints and therefore leads to higher similarity with the real network w.r.t the Random Graph.

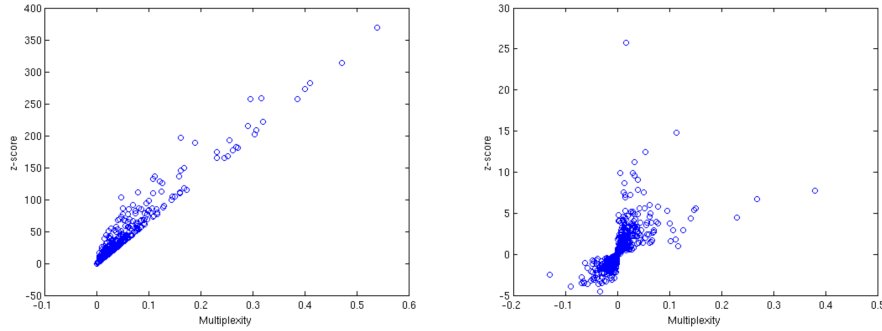

Figure 1.8: **Significance of the binary multiplexity values for the European Airport Network.** Scatter plots of binary multiplexity values $\mu_b^{\alpha\beta}$ vs the corresponding z-score for each pair of layers, respectively for Random Graph (a) and Binary Configuration Model (b).

A different trend can be observed when the European Airport Network is taken into account (Figure 1.8(a)). Indeed, it is still clear a high correlation between values of multiplexity and their respective z-scores when the Random Graph is considered. However, many z-scores associated to multiplexities close to 0, in this case, are now close to 0 themselves, therefore suggesting that many pairs of

layers (i.e. airline companies) may actually be anti-correlated rather than simply uncorrelated. In this case, the adoption of a more refined null model is then crucial in order to deeply understand the structural properties of such a system.

When the Binary Configuration Model is considered as benchmark, however, the analysis of the corresponding scatter plots dramatically changes. However, as we said, these results are strongly dependent on the considered network. Indeed, Figure 1.8(b) exhibits a completely different trend with respect, for instance, to the corresponding Figure 1.7(b) (related to the World Trade Multiplex): no correlation between $\mu^{\alpha\beta}_{BCM}$ and $z^{\alpha\beta}_{BCM}$ can be observed in this case, so that the same value of multiplexity can be either associated to a low z-score (thus being compatible with the expected value under the chosen Configuration Model) or to very high z-scores (hence unexpectedly different from the model's expectation). Moreover, Figure 1.8(b) clearly shows the sign-concordance existing between the multiplexity and the associated z-score that we pointed out in the previous Section. However, no other clear trend can be inferred from such a plot, therefore pointing out the importance of taking into account both the quantities ($\mu^{\alpha\beta}_{BCM}$ and $z^{\alpha\beta}_{BCM}$) in order to have a complete understanding of the correlations between layers of a multiplex.

Furthermore, we should highlight once more that, in terms of absolute z-scores values, the significance of the values of multiplexity under Random Graph ($\mu^{\alpha\beta}_{RG}$) is usually much higher than that observed after the comparison with the Configuration Model ($\mu^{\alpha\beta}_{BCM}$), as we have already found before for the WTM.

Similarly, we can analyze the patterns of correlations resulting from the z-scores associated to the weighted multiplexity, as defined in (1.58) and (1.57). In Figure 1.9(a) we show the relation between the values of weighted multiplexity for any pair of layers and the related z-score, computed with respected to the expected multiplexity according to the Weighted Random Graph. The sign concordance is still clear, but the correlation between $\mu^{\alpha\beta}_{WRG}$ and $z^{\alpha\beta}_{WRG}$ is much less sharp with respect to the corresponding binary case, especially for negative values of multiplexity.

Even more so, such a weak correlation between weighted multiplexity and the corresponding z-score completely disappears when the considered benchmark is the Weighted Configuration Model (Figure 1.9(b)): in this case the same value of $\mu^{\alpha\beta}_{WCM}$ may correspond to z-scores even characterized by different orders of magnitude, thus pointing out once more the importance of the introduction of a notion of standard deviation referred to the average $\langle \mu^{\alpha\beta}_{WCM} \rangle$. Indeed, the same value of observed multiplexity can actually be either extremely unexpected or in full agreement with the null model's prediction.

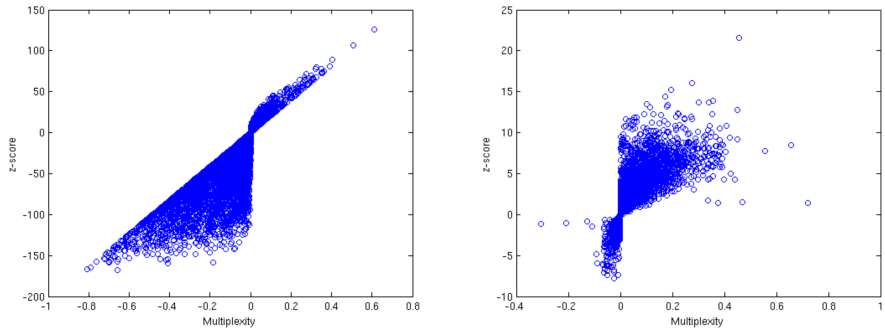

Figure 1.9: **Significance of the weighted multiplexity values for the World Trade Multiplex.** Scatter plots of weighted multiplexity values $\mu_w^{\alpha\beta}$ vs the corresponding z-scores for each pair of layers, respectively for Weighted Random Graph (a) and Weighted Configuration Model (b).

1.F International Trade Multiplex Network: list of layers

Throughout this thesis we meticulously analyze the World Trade Multiplex Network (WTM), as provided by the BACI database mentioned in the main text. The data provide information about import and export between $N = 207$ countries (we focus in particular on the year 2011) and turn out to have a straightforward representation in terms of multi-layered network; it is indeed possible to disaggregate the global trade between any two countries into the import and export in a given commodity, so that the global trade system can be thought of as the superposition of all the layers. The network is then composed by 207 countries and $M = 96$ different commodities, according to the standard international classification HS1996 (the list of commodities is reported below in Table 1.1). While the aggregated network shows a density higher than 55%, the various layers are characterized by densities from 6% (related to trade in silk) to 45% (for import-export of mechanical appliances and parts thereof). Such heterogeneity may suggest that a multiplex analysis is therefore required. Interestingly, in this case each of the layers is represented by a weighted network, where the weight associated to any link in a layer stands for the amount of money exchanged by a given pair of countries in that layer (i.e., commodity).

	Commodity	
01	Live animals	●
02	Meat and edible meat offal	●
03	Fish, crustaceans and acquatic invertebrates	●
04	Dairy produce; birs eggs; honey and other edible animal products	●
05	Other products of animal origin	●
06	Live trees, plants; bulbs, roots; cut flowers and ornamental foliage tea and spices	●
07	Edible vegetables and certain roots and tubers	●
08	Edible fruit and nuts; citrus fruit or melon peel	●
09	Coffee, tea, mate and spices	●
10	Cereals	●
11	Milling products; malt; starch; inulin; wheat gluten	●
12	Oil seeds and oleaginous fruits; miscellaneous grains, seeds and fruit; industrial or medicinal plants; straw and fodder	●
13	Lac; gums, resins and other vegetable sap and extracts	●
14	Vegetable plaiting materials and other vegetable products	●
15	Animal, vegetable fats and oils, cleavage products, etc.	●
16	Edible preparations of meat, fish, crustaceans, mollusks or other aquatic invertebrates	●
17	Sugars and sugar confectionary	●
18	Cocoa and cocoa preparations	●
19	Preparations of cereals, flour, starch or milk; bakers wares	●
20	Preparations of vegetables, fruit, nuts or other plant parts	●
21	Miscellaneous edible preparations	●
22	Beverages, spirits and vinegar	●
23	Food industry residues and waste; prepared animal feed	●
24	Tobacco and manufactured tobacco substitutes	●
25	Salt; sulfur; earth and stone; lime and cement plaster	●
26	Ores, slag and ash	●
27	Mineral fuels, mineral oils and products of their distillation; bitumin substances; mineral wax	●
28	Inorganic chemicals; organic or inorganic compounds of precious metals, of rare-earth metals, of radioactive elements or of isotopes	●

29	Organic chemicals	●
30	Pharmaceutcal products	●
31	Fertilizers	●
32	Tanning or dyeing extracts; tannins and derivatives; dyes, pigments and coloring matter; paint and varnish; putty and other mastics; inks	●
33	Essential oils and resinoids; perfumery, cosmetic or toilet preparations	●
34	Soap; waxes; polish; candles; modeling pastes; dental preparations with basic of plaster	●
35	Albuminoidal substances; modified starch; glues; enzymes	●
36	Explosives; pyrotechnic products; matches; pyrophoric alloys; certain combustible preparations	●
37	Photographic or cinematographic goods	●
38	Miscellaneous chemical products	●
39	Plastics and articles thereof	●
40	Rubber and articles thereof	●
41	Raw hides and skins (other than furskins) and leather	●
42	Leather articles; saddlery and harness; travel goods, handbags and similar; articles of animal gut (not silkworm gut)	●
43	Furskins and artificial fur; manufactures thereof	●
44	Wood and articles of wood; wood charcoal	●
45	Cork and articles of cork	●
46	Manufactures of straw, esparto or other plaiting materials; basketware and wickerwork	●
47	Pulp of wood or of other fibrous cellulosic material; waste and scrap of paper and paperboard	●
48	Paper and paperboard and articles thereof; paper pulp articles	●
49	Printed books, newspapers, pictures and other products of printing industry; manuscripts, typescripts	●
50	Silk, including yarns and woven fabric thereof	●
51	Wool and animal hair, including yarn and woven fabric	●
52	Cotton, including yarn and woven fabric thereof	●
53	Other vegetable textile fibers; paper yarn and woven fabrics of paper yarn	●
54	Manmade filaments, including yarns and woven fabrics	●
55	Manmade staple fibers, including yarns and woven fabrics	●

56	Wadding, felt and nonwovens; special yarns; twine, cordage, ropes and cables and article thereof	🟠
57	Carpets and other textile floor coverings	🟠
58	Special woven fabrics; tufted textile fabrics; lace; tapestries; trimmings; embroidery	🟡
59	Impregnated, coated, covered or laminated textile fabrics; textile articles for industrial use	🟡
60	Knitted or crocheted fabrics	🟡
61	Apparel articles and accessories, knitted or crocheted	🟡
62	Apparel articles and accessories, not knitted or crocheted	🟡
63	Other textile articles; needlecraft sets; worn clothing and worn textile articles; rags	🟡
64	Footwear, gaiters and the like and parts thereof	🟡
65	Headgear and parts thereof	🟡
66	Umbrellas, walking sticks, seat sticks, riding crops, whips, and parts thereof	🟡
67	Prepared feathers, down and articles thereof; artificial flowers; articles of human hair	🟡
68	Articles of stone, plaster, cement, asbestos, mica or similar materials	🟢
69	Ceramic products	🟢
70	Glass and glassware	🟢
71	Pearls, precious stones, metals, coins, etc.	🟢
72	Iron and steel	🟢
73	Articles of iron and steel	🟢
74	Copper and articles thereof	🟢
75	Nickel and articles thereof	🟢
76	Aluminum and articles thereof	🟢
77	Lead and articles thereof	🟢
78	Zinc and articles thereof	🟢
79	Tin and articles thereof	🟢
80	Other base metals; cermets; articles thereof	🟢
81	Tools, implements, cutlery, spoons and forks of base metal and parts thereof	🟢
82	Miscellaneous articles of base metal	🟢
83	Nuclear reactors, boilers, machinery and mechanical appliances; parts thereof	⚫
84	Electric machinery, equipment and parts; sound equipment; television equipment	⚫
85	Railway or tramway; locomotives, rolling stock, track fixtures and parts thereof; mechanical and electromechanical traffic signal equipment	⚫

86	Vehicles (not railway, tramway, rolling stock); parts and accessories	●
87	Aircraft, spacecraft, and parts thereof	●
88	Ships, boats and floating structures	●
89	Optical, photographic, cinematographic, measuring, checking, precision, medical or surgical instruments/apparatus; parts and accessories	●
90	Clocks and watches and parts thereof	●
91	Musical instruments; parts and accessories thereof	●
92	Arms and ammunition, parts and accessories thereof	●
93	Furniture; bedding, mattresses, cushions, etc.; other lamps and light fitting, illuminated signs and nameplates, prefabricate buildings	●
94	Toys, games and sports equipment; parts and accessories	●
95	Miscellaneous manufactured articles	●
96	Works of art, collectors pieces and antiques	●

Table 1.1: **List of commodities of the WTM, according to the standard international classification HS1996, and associated codes, as provided by the BACI-Comtrade dataset.** In the first column we show the number representing each product. In the third column we divide such commodities in classes of similar traded items, each of them being represented by a different colored circle; colors are the same as reported in the dendrogram in Figure 1.6.

Bibliography

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

[2] M. E. J. Newman (2003) 'The structure and function of complex networks', *SIAM Review* **45**, 167

[3] D. J. Watts, S. H. Strogatz (1998) 'Collective dynamics of "small-world" networks', *Nature* **393**, 440

[4] S. Fortunato (2010) 'Community detection in graphs', *Physics Reports* **486** (3), 75

[5] S. Wasserman, K. Faust (1994) 'Social network analysis', Cambridge University Press (Cambridge, New York)

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

[7] F. Radicchi (2014) 'Driving interconnected networks to supercriticality' *Physical Review X* **4** (2), 021014

[8] M. Szell, R. Lambiotte, S. Thurner (2010) 'Multirelational organization of large-scale social networks in an online world', *Proceedings of the National Academy of Sciences USA* **107** (31), 13636

[9] 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

[10] R. G. Morris, M. Barthelemy (2012) 'Transport on coupled spatial networks', *Physical Review Letters* **109** (12), 128703

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

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

[13] J. Y. Kim, K.-I. Goh (2013) 'Coevolution and correlated multiplexity in multiplex networks', *Physical Review Letters* **111** (5), 058702

[14] V. Nicosia, G. Bianconi, V. Latora, M. Barthelemy (2014) 'Non-linear growth and condensation in multiplex networks' *Physical Review E* **90**(4), 042807

[15] A. Saumell-Mendiola, M. A. Serrano, M. Boguñá (2012) 'Epidemic spreading on interconnected networks', *Physical Review E* **86** (2), 026106

[16] 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

[17] J. Gómez-Gardeñes, I. Reinares, A. Arenas, L.-M. Floria (2012) 'Evolution of cooperation in multiplex networks', *Scientific Reports* **2**, 620

[18] E. Estrada, J. Gómez-Gardeñes (2014) 'Communicability reveals a transition to coordinated behavior in multiplex networks', *Physical Review E* **89** (4), 042819

[19] M. E. J. Newman, S. H. Strogatz, D. J. Watts (2001) 'Random graphs with arbitrary degree distributions and their applications', *Physical Review E* **64** (2), 026118

[20] J. Park, M. E. J. Newman (2003) 'Origin of degree correlations in the Internet and other networks', *Physical Review E* **68** (2), 026112

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

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

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

[24] 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

[25] D. Garlaschelli (2009) 'The weighted random graph model', *New Journal of Physics* **11**, 073005

[26] A. Barrat, M. Barthelemy, A. Vespignani (2008) 'Dynamical processes on complex networks', Cambridge University Press (Cambridge)

[27] B. Min, S. Do Yi, K.-M. Lee, K.-I. Goh (2014) 'Network robustness of multiplex networks with interlayer degree correlations', *Physical Review E* **89** (4), 042811

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

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

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

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

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

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

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

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

[36] R. Mantegna (1999) 'Hierarchical structure in financial markets', *European Physics Journal B* **11** (1), 193

Chapter 2

Directed multiplex networks

Real-world multi-layer networks feature nontrivial dependencies among links of different layers. Here we argue that, if links are directed, dependencies are twofold. Besides the ordinary tendency of links of different layers to align as the result of 'multiplexity', there is also a tendency to anti-align as the result of what we call 'multireciprocity', i.e. the fact that links in one layer can be reciprocated by *opposite* links in a different layer. Multireciprocity generalizes the scalar definition of single-layer reciprocity to that of a square matrix involving all pairs of layers. We introduce multiplexity and multireciprocity matrices for both binary and weighted multiplexes and validate their statistical significance against maximum-entropy null models that filter out the effects of node heterogeneity. We then perform a detailed empirical analysis of the World Trade Multiplex (WTM), representing the import-export relationships between world countries in different commodities. We show that the WTM exhibits strong multiplexity and multireciprocity, an effect which is however largely encoded into the degree or strength sequences of individual layers. The residual effects are still significant and allow to classify pairs of commodities according to their tendency to be traded together in the same direction and/or in opposite ones. We also find that the multireciprocity of the WTM is significantly lower than the usual reciprocity measured on the aggregate network. Moreover, layers with low (high) internal reciprocity are embedded within sets of layers with comparably low (high) mutual multireciprocity. This suggests that, in the WTM, reciprocity is inherent to groups of related commodities rather than to individual commodities. We discuss the implications for international trade research focusing on product taxonomies, the product space, and fitness/complexity metrics.

The results presented in this chapter have been published in the following reference:
V. Gemmetto, T. Squartini, F. Picciolo, F. Ruzzenenti, D. Garlaschelli, *Physical Review E*, **94** (4), 042316 (2016).

2.1 Introduction

Several real-world systems are composed by intricately interconnected units, thus exhibiting a nontrivial network structure. The behaviour and dynamics of such systems are strongly dependent on how information can propagate throughout the network. Both the directionality and the intensity of connections crucially affect this process, and should possibly be incorporated in the network description. For instance, most of the communication relations among individuals, such as exchanges of letters, e-mails or texts, are intrinsically directional and are therefore best represented as directed networks [1]. Furthermore, such interactions typically have heterogeneous intensities, calling for a description in terms of weighted networks [2].

Recently, it has been realized that many real-world systems often require an even more detailed representation, because a given set of units can be connected by different kinds of relations. This property can be abstractly captured in terms of so-called *edge-colored graphs* (where links of different colors are allowed among the same set of nodes) or equivalenlty *multi-layer* or *multiplex networks* (where the same set of nodes is replicated in multiple layers, each of which is an ordinary network) [3, 4]. The nontrivial properties of these systems, with respect to ordinary single-layer ('monochromatic' or 'monoplex') networks, arise from the fact that the various layers are interdependent and the presence of a link in one layer can influence the presence of a link in a different layer. A clear example is represented by the different kinds of relationships existing between employees in a university department [5], where individuals can be connected by co-authorship, common leisure activities, on-line social networks etc. The interdependence of layers implies that the topological properties usually defined for monoplex networks admit nontrivial generalizations to multiplex networks, and that some properties which are uninteresting, or even undefined, for single-layer networks become relevant for multiplexes.

This chapter introduces novel metrics characterizing the dependencies among layers in multiplexes *with directed links*. While various measures of inter-layer overlap for multiplexes have already been introduced [6, 7], they suffer from two main limitations. First, most definitions are available only for multiplexes with undirected links, and their straightforward generalization to the directed case would overshadow important properties that are inherent to directed networks, most importantly the reciprocity (which is one of our main focuses here). Second, even in 'trivial' multiplexes where there is no dependence among layers (i.e. in independent superpositions of single-layer networks with the same set of nodes), a certain degree of inter-layer overlap can be created entirely by chance. This effect becomes more pronounced as the density of the single-layer networks increases and as the correlation among single-node properties (like degrees or strengths) across layers increases. For instance, if a node is a hub in multiple layers, there

is an increased chance of overlap among these layers, even if the presence of links in one layer is assumed not to influence the presence of links in another layer.

The two limitations discussed above highlight the need to define metrics that appropriately filter out both global (network-wide) and local (node-specific) density effects. Correlation-based measures of inter-layer overlap have been proposed with this aim in mind [8]. However, as recently pointed out [9], correlation-based metrics for multiplexes are not a correct solution in general, because they implicitly assume that edges observed between different pairs of nodes are sampled from the same probability distribution. This assumption is strongly violated in real-world networks, whose markedly heterogeneous topology is a signature of very different probabilities for edges emanating from different nodes, e.g. the probability of links being found around more important nodes is clearly different from the probability of links being found around less important nodes.

The above considerations motivate us to introduce new multiplexity metrics that explicitly take the directionality of links into account and appropriately filter out the spurious effects of chance, while controlling for the extreme heterogeneity of empirical node-specific properties. In this chapter we carry out this program by extending recent 'filtered' definitions of multiplexity [9], originally defined for undirected links, to the case of directed links. Although this might seem a straightforward procedure at first, we will in fact show that it requires different null models, triggers novel concepts, and leads to new quantities that are undefined in the undirected case. Indeed, while in the undirected case there is only one possible notion of dependency among links in different layers, in the directed case there are two possibilities, depending on whether links are 'aligned' or 'anti-aligned'.

Aligned links between two layers are observed when a directed link from node i to node j exists in both layers. This situation is the straightforward analogue of what can happen in undirected multiplex networks, and is a signature of the fact that the connection from i to j is relevant for multiple layers. We will denote this effect simply as (directed) *multiplexity*, in analogy with the undirected case [9], and will study it in the general case of an arbitrary number of layers. By contrast, anti-aligned links form between two layers when a link from node i to node j in one layer is *reciprocated* by an opposite link from node j to node i in the other layer. This situation does not have a counterpart in the case of undirected multiplexes and leads us to the definition of the novel concept of *multireciprocity*, i.e. the generalization of the popular concept of reciprocity to the case of multiplex networks.

In monoplex networks - either binary [10] or weighted [11] - reciprocity is defined as the tendency of vertex pairs to form mutual connections. This property, which is one of the best studied properties of single-layer directed networks, can crucially affect various dynamical processes such as diffusion [12], percolation [13]

and growth [14, 15]. For instance, the presence of directed, reciprocal connections can lead to the establishment of functional communities and hierarchies of groups of neurons in the cerebral cortex [16].

In binary graphs, a simple measure of reciprocity is the ratio of the number of reciprocated links (i.e. realized links for which the link pointing in the opposite direction between the same two nodes is also realized) to the total number of directed links. However, it has been shown [10] that this measure is not *per se* informative about the actual tendency towards reciprocation, because even in a random network a certain number of reciprocated links will appear. So the number of observed mutual interactions has to be compared with the expected number obtained for a given random null model, if one wants to understand whether mutual links are present in the real network significantly more (or less) often than in the random benchmark [17]. It is therefore crucial to make use of proper null models for networks. Since in most real-world directed networks the distribution of the number of in-coming and out-going links (i.e. the in-degree and out-degree) of nodes is very broad, an appropriate null model should fix the in- and out-degrees of all nodes equal to their observed values. The null model of directed networks with given in- and out-degrees often goes under the name of directed binary configuration model (DBCM) [18]. The rationale underlying the DBCM is the consideration that the in- and out-degree of a node might reflect some intrinsic 'size', or other characteristic, of that node; therefore a null model tailored for a specific network should preserve the observed degree heterogeneity. Conveniently, the DBCM is also the correct null model to use when measuring the *multiplexity* among layers of a multiplex with directed links. Indeed, the DBCM is the directed generalization of the undirected binary configuration model used in the previous chapter [9] for the definition of appropriately filtered, undirected multiplexity metrics. This nicely implies that we can use the DBCM as a single null model in our analysis of both multiplexity and multireciprocity.

Recently, the definition of reciprocity has been extended to weighted networks [11]. A simple measure of weighted reciprocity is the ratio of 'total reciprocated link weight' to total link weight, where the reciprocated link weight is defined, for any two reciprocated links, as the minimum weight of the two links. Similarly to the binary case, some level of weighted reciprocity can be generated purely by chance. So the empirical measure has to be compared to its expected value under a proper null model, represented in this case by a random weighted network where each node has the same in-strength and out-strength (i.e. total in-coming link weight and total out-going link weight, respectively) as in the real network. This null model is sometimes called the directed weighted configuration model (DWCM) [19] and, conveniently, is also the relevant null model (generalizing its undirected counterpart [9]) to study the multiplexity in presence of weighted directed links.

We stress that the concept of reciprocity has not been generalized to multiplex networks yet. Our definition of multireciprocity represents the first step in this direction and captures the tendency of a directed link in one layer of a multiplex to be reciprocated by an opposite link in a possibly *different* layer. While ordinary reciprocity can be quantified by a scalar quantity, multireciprocity requires a square matrix where all the possible pairs of layers are considered. Similarly, the multiplexity also requires a square matrix. Together, the multiplexity matrix and the multireciprocity matrix represent the two 'directed' extensions of the undirected multiplexity matrix that has been introduced in Chapter 1 to characterize undirected (either binary or weighted) multiplexes.

The rest of the chapter is organized as follows. In Sec. 2.2 we introduce our methods, null models and main definitions for both binary and weighted multiplexes. In Sec. 2.3 we apply our techniques to the analysis of the World Trade Multiplex (WTM), a directed weighted multiplex representing the import-export relations between countries of the world in different products. We identify a number of empirical properties of the WTM that are impossible to access via the usual aggregate (monoplex) analysis of the network of total international trade. We finally conclude the chapter in Sec. 2.4, where we discuss some important implications of our results, both for the general study of multiplex networks and for more specific research questions in international trade economics. Several necessary technical details are given in the follwing appendices.

2.2 Multiplexity and Multireciprocity metrics

In this section we give definitions of (directed) multiplexity and multireciprocity metrics for both binary and weighted multiplexes. These definitions require, as a preliminary step, the introduction of appropriate null models. In turn, null models require the choice of a convenient notation. We address these points in the resulting order.

We represent a directed multiplex $\overrightarrow{G} = (G^1, \ldots, G^M)$ as the superposition of M directed networks (layers) G^α ($\alpha = 1, \ldots, M$), all sharing the same set of N nodes [3]. Links can be either binary or weighted. In the binary case, each layer α is represented by a $N \times N$ binary adjacency matrix $G^\alpha = (a_{ij}^\alpha)_{i,j=1}^N$, where $a_{ij}^\alpha = 0, 1$ depending on whether a directed link from node i to node j is absent or present, respectively. In the weighted case, each layer α is represented by a $N \times N$ non-negative integer adjacency matrix $G^\alpha = (w_{ij}^\alpha)_{i,j=1}^N$, where $w_{ij}^\alpha = 0, 1, \ldots \infty$ is the weight of the directed link from node i to node j ($w_{ij}^\alpha = 0$ indicating the absence of such link). We denote by $\mathcal{G}_N$ the set of all (binary or weighted) single-layer graphs with N nodes, and by $\mathcal{G}_N^M \equiv (\mathcal{G}_N)^M$ the set of all (binary or weighted) M-layer multiplexes with N nodes.

2.2.1 Null models of multiplex networks: maximum entropy and maximum likelihood

Since our purpose is that of measuring correlations between directed links (possibly, in opposite directions) in different layers, we define independent reference models for each layer of the multiplex, thus creating an uncorrelated null model for the entire multiplex [7, 9]. This means that, if $\mathcal{P}(\overrightarrow{G}|\overrightarrow{\theta})$ denotes the joint probability of the entire multiplex $\overrightarrow{G} \in \mathcal{G}_N^M$ (given a set of constraints enforced via the vector $\overrightarrow{\theta}$ of parameters, see Appendix 2.A) and

$$P^{\alpha}(G^{\alpha}|\overrightarrow{\theta^{\alpha}}) \equiv \underbrace{\sum_{G^1 \in \mathcal{G}_N} \cdots \sum_{G^{\beta} \in \mathcal{G}_N} \cdots \sum_{G^M \in \mathcal{G}_N}}_{\beta \neq \alpha} \mathcal{P}(\overrightarrow{G}|\overrightarrow{\theta}) \tag{2.1}$$

denotes the (marginal) probability for the single-layer graph $G^{\alpha} \in \mathcal{G}_N$ (given a set of layer-specific constraints enforced via the partial vector $\overrightarrow{\theta^{\alpha}}$), we require the null model to obey the factorization property

$$\mathcal{P}(\overrightarrow{G}|\overrightarrow{\theta}) = \prod_{\alpha=1}^{M} P^{\alpha}(G^{\alpha}|\overrightarrow{\theta^{\alpha}}). \tag{2.2}$$

The above property ensures that the definition of the null model for the entire multiplex reduces to the definition of independent null models for each layer separately (see Appendix 2.A for a rigorous derivation).

In the case of binary multiplexes, the null model we want to use to control for the heterogeneity of nodes in each layer is, as we have already mentioned, the Directed Binary Configuration Model (DBCM) [20, 21], defined as the ensemble of binary networks with given in-degree and out-degree sequences. At this point, we have to make a major decision, since the DBCM can be implemented either microcanonically or canonically.

In the microcanonical approach, node degrees are "hard", i.e. enforced sharply on each realization. The most popular microcanonical implementation of the DBCM is based on the random degree-preserving rewiring of links [18] (a.k.a. the Local Rewiring Algorithm), which unfortunately introduces a bias. This bias arises because, if the degree distribution is sufficiently broad (as in most real-world cases), the randomization process explores the space of possible network configurations not uniformly, giving higher probability to the configurations that are "closer" to the initial one [22] (more details are given in Appendix 2.B). Another possible microcanonical implementation, based on the random matching of "edge stubs" (half links) to the nodes, creates undesired self-loops and multiple edges [18, 23]. Besides these limitations, microcanonical approaches are computationally demanding. Indeed, in order to measure the expected value of any

quantity of interest, it is necessary to generate several randomized networks, on each of which the quantity needs to be calculated. This sampling method is *per se* very costly, and even more so in the case of multiplex networks, due to the presence of several layers requiring a further multiplication of iterations (see Appendix 2.B).

By contrast, in the canonical implementation [20, 21] of the DBCM the in- and out-degrees are "soft", i.e. preserved only on average. The resulting probability distribution over the ensemble of possible graphs is obtained analytically by maximizing the entropy subject to the enforced constraints [20, 24, 25, 26] (see Appendix 2.A for details). This procedure leads to the class of models also known as Exponential Random Graphs or $p^\star$ models [27, 28, 29]. In order to fit such exponential random graphs to real-world networks, we adopt an exact, unbiased and fast method [20, 21] based on the Maximum Likelihood principle [30]. The method is summarized in Appendix 2.B and implemented in our analysis using the so-called MAX&SAM ("Maximize and Sample") algorithm [21]. The latter yields the exact probabilities of occurrence of any graph in the ensemble and the explicit expectation values of the quantities of interest. This has the enormous advantage that an explicit sampling of graphs is not required: expectation values are calculated analytically and not as sample averages. In particular, the probability p_{ij}^{α} that a link from node i to node j is realized in layer α ($a_{ij}^{\alpha} = 1$) can be easily calculated. From the set of all such probabilities, the expected value of - for instance - the multireciprocity can be computed analytically and directly compared with the empirical value, in order to obtain a filtered measure.

We now come to the case of multiplexes with weighted links. In this case we want the enforced constraints to be the in-strength and out-strength sequences of the real network, separately for each layer. The corresponding model is sometimes referred to as the Directed Weighted Configuration Model (DWCM) [11]. As for the binary case, we want to build the null model canonically as a maximum-entropy ensemble of weighted networks, leading to a weighted Exponential Random Graph model [20, 11]. The implementation we use is again based on the MAX&SAM algorithm [21], which in this case calculates the exact probability that, in the null model, the weight of the directed link connecting node i to node j in layer α has a particular value w_{ij}^{α}, for each pair of nodes and each layer. From this probability, the expected weighted multireciprocity can be computed analytically and compared with the empirical one, thus producing a filtered value that, in this case as well, does not require the explicit sampling of graphs.

2.2.2 Binary multiplexity and multireciprocity

Our first set of main definitions are specific for multiplexes with binary links. Consider a directed and binary multiplex $\vec{G}$ with M layers. We quantify the similarity and reciprocity between any two layers α and β by defining the binary

multiplexity $m_b^{\alpha\beta}$ and *multireciprocity* $r_b^{\alpha\beta}$ as follows:

$$m_b^{\alpha\beta} = \frac{2\sum_i \sum_{j\neq i} \min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}}{L^{\alpha} + L^{\beta}} = \frac{2L^{\alpha\rightrightarrows\beta}}{L^{\alpha} + L^{\beta}}, \tag{2.3a}$$

$$r_b^{\alpha\beta} = \frac{2\sum_i \sum_{j\neq i} \min\{a_{ij}^{\alpha}, a_{ji}^{\beta}\}}{L^{\alpha} + L^{\beta}} = \frac{2L^{\alpha\rightleftarrows\beta}}{L^{\alpha} + L^{\beta}}, \tag{2.3b}$$

where $L^{\alpha} = \sum_i \sum_{j\neq i} a_{ij}^{\alpha}$ represents the total number of directed links in layer α (analogously for layer β), $L^{\alpha\rightrightarrows\beta} = \sum_i \sum_{j\neq i} \min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}$ is the number of links of layer α that are multiplexed in layer β (clearly, $L^{\alpha\rightrightarrows\beta} = L^{\beta\rightrightarrows\alpha}$), and $L^{\alpha\rightleftarrows\beta} = \sum_i \sum_{j\neq i} \min\{a_{ij}^{\alpha}, a_{ji}^{\beta}\}$ is the number of links of layer α that are reciprocated in layer β (clearly, $L^{\alpha\rightleftarrows\beta} = L^{\beta\rightleftarrows\alpha}$). Note that possible self-loops (terms of the type a_{ii}^{α}) are deliberately ignored because they are indistinguishable from links pointing in the opposite direction, thus making their contribution to either multiplexity or multireciprocity undefined.

Equations (2.3) can be regarded as defining the entries of two $M \times M$ matrices, which we will call the *binary multiplexity matrix* $\mathbf{M}_b = (m_b^{\alpha\beta})_{\alpha\beta}$ and the *binary multireciprocity matrix* $\mathbf{R}_b = (r_b^{\alpha\beta})_{\alpha\beta}$ respectively. The matrices $\mathbf{M}_b$ and $\mathbf{R}_b$ represent the two natural extensions, to the case of directed multiplexes, of the single binary multiplexity matrix introduced in Chapter 1 [9] for undirected binary multiplexes. Both matrices provide information about the 'overlap' between directed links connecting pairs of nodes in different layers. Their entries range in $[0, 1]$ and are maximal only when layers α and β are respectively identical (i.e. $a_{ij}^{\alpha} = a_{ij}^{\beta}$ for all $i \neq j$) and fully 'multireciprocated' (i.e. $a_{ij}^{\alpha} = a_{ji}^{\beta}$ for all $i \neq j$). The matrix $\mathbf{M}_b$ has by construction a unit diagonal, since the intra-layer multiplexity trivially has the maximum value $m_b^{\alpha\alpha} = 1$ for all α. By contrast, the diagonal of $\mathbf{R}_b$ is nontrivial and of special significance, as the intra-layer multireciprocity $r_b^{\alpha\alpha}$ reduces to the ordinary definition of binary reciprocity for monoplex networks [10].

For 'trivial', uncorrelated multiplexes made of sparse non-interacting layers with narrow degree distributions, the matrix $\mathbf{M}_b$ would asymptotically (i.e. in the limit of large N, but not necessarily large M) be the $M \times M$ identity matrix, and the matrix $\mathbf{R}_b$ would asymptotically be a $M \times M$ diagonal matrix. This is because, in presence of sparse uncorrelated layers without hubs, the chance of a link in one layer 'overlapping' with a (mutual) link in a different layer is negligible. For finite and/or dense networks and/or broad degree distributions, however, positive values of $m_b^{\alpha\beta}$ and $r_b^{\alpha\beta}$ (with $\alpha \neq \beta$) can be produced entirely by chance even in a multiplex with no dependencies among layers. For instance, if the same node is a hub in multiple layers, the chance of a large overlap of links among all pairs of such layers is very high, even if the layers are non-interacting.

The above considerations imply that, in order to extract statistically significant information about the tendency towards multiplexity and multireciprocity in a

real-world multiplex, it becomes necessary to compare the empirical values of $m_b^{\alpha,\beta}$ and $r_b^{\alpha,\beta}$ with the corresponding expected values calculated under the chosen null model of independent multiplexes with given degrees (i.e. the DBCM). Hence, we introduce the transformed (i.e., rescaled) binary multiplexity and multireciprocity matrices with entries

$$\mu_b^{\alpha\beta} = \frac{m_b^{\alpha\beta} - \langle m_b^{\alpha\beta}\rangle_{\mathrm{DBCM}}}{1 - \langle m_b^{\alpha\beta}\rangle_{\mathrm{DBCM}}} \qquad (\alpha \neq \beta), \tag{2.4a}$$

$$\rho_b^{\alpha\beta} = \frac{r_b^{\alpha\beta} - \langle r_b^{\alpha\beta}\rangle_{\mathrm{DBCM}}}{1 - \langle r_b^{\alpha\beta}\rangle_{\mathrm{DBCM}}}, \tag{2.4b}$$

where $\langle\cdot\rangle_{\mathrm{DBCM}}$ denotes the expected value under the DBCM. Note that, since $\langle m_b^{\alpha\alpha}\rangle_{DBCM} = m_b^{\alpha\alpha} = 1$ for all α, we formally set the diagonal terms $\mu_b^{\alpha\alpha} \equiv 1$, as the definition (2.4a) would produce an indeterminate expression if extended to $\alpha = \beta$. The explicit calculation of the above expected values is provided in Appendix 2.C and more details are provided later in this section.

The filtered quantities (2.4) are directly informative about the presence of dependencies between layers. Positive values represent higher-than-expected multiplexity or multireciprocity (correlated or 'attractive' pairs of layers), while negative values represent lower-than-expected quantities (anticorrelated or 'repulsive' pairs of layers). Pairs of uncorrelated ('noninteracting') layers are characterized by multiplexity and multireciprocity values comparable with 0. In principle, a layer that is uncorrelated with all other layers can be separated from the multiplex and analysed separately from it.

The choice of the denominator of (2.4a) and (2.4b), *a priori* not obvious, guarantees that the maximum value for the transformed multiplexity and multireciprocity is 1. Moreover, it ensures that $\rho_b^{\alpha\alpha}$ reduces to the rescaled reciprocity ρ_b defined for single-layer networks [10]. It should also be noted that the multiplexity defined in (2.3a) is just the normalized version of the inter-layer overlap introduced in [6] and [7], extended to directed multiplex networks. In this context, the novel contribution that we give is the comparison with a null model. Indeed, while (2.3a) only provides information about the raw similarity of the layers, which is strongly density-dependent, the transformed measure (2.4a) is mapped to a universal interval. In combination with the z-scores that we introduce later, it can be used to consistently compare the statistical significance of the multiplexity of different systems. The quantity defined in (2.3b), which focuses explicitly on the reciprocity properties of the multiplex, has never been introduced before, along with its transformed quantity defined in (2.4b). The latter can be used for a consistent comparison of the multireciprocity of multiplexes with different densities.

The calculation of the expected values of $m_b^{\alpha\beta}$ and $r_b^{\alpha\beta}$ under the DBCM can be carried out analytically using the MAX&SAM method [21], with no need to actually randomize the empirical network or numerically sample the null model ensemble. Ultimately, the calculation requires the computation of the expected value of the minimum between two binary random variables (see Appendix 2.C). If $p_{ij}^{\alpha} \equiv \langle a_{ij}^{\alpha} \rangle_{\mathrm{DBCM}}$ denotes the probability that, under the DBCM, a directed link is realized from node i to node j in layer α, then the adjacency matrix entry a_{ij}^{α} is described by the Bernoulli mass probability function

$$P(a_{ij}^{\alpha}) = (p_{ij}^{\alpha})^{a_{ij}^{\alpha}} (1 - p_{ij}^{\alpha})^{(1-a_{ij}^{\alpha})}. \tag{2.5}$$

Using the above equation, and given the explicit expression for p_{ij}^{α}, it is possible to calculate $\mu_b^{\alpha\beta}$ and $\rho_b^{\alpha\beta}$ analytically as reported in the aforementioned Appendix 2.C.

It is instructive to compare the multivariate quantities measured on the multiplex with the corresponding scalar quantities defined on the aggregate monoplex network obtained by combining all layers together. This comparison can highlight the gain of information resulting from the multiplex representation, with respect to the ordinary monoplex projection where all the distinct types of links are treated as equivalent. The binary aggregate monoplex can be defined in terms of the adjacency matrix with entries

$$a_{ij}^{\mathrm{mono}} = 1 - \prod_{\alpha=1}^{M} (1 - a_{ij}^{\alpha}) = \begin{cases} 1 & \text{if } \exists \alpha : a_{ij}^{\alpha} = 1 \\ 0 & \text{otherwise} \end{cases} . \tag{2.6}$$

For the quantities we defined so far, the only meaningful comparison between the multiplex and the aggregate network can be done in terms of the reciprocity, because the multiplexity of the aggregate is $m_b^{\mathrm{mono}} = 1$ by construction. The single, global reciprocity of the aggregated monoplex network is given by

$$r_b^{\mathrm{mono}} = \frac{\sum_i \sum_{j \neq i} \min\{a_{ij}^{\mathrm{mono}}, a_{ji}^{\mathrm{mono}}\}}{L^{\mathrm{mono}}} \tag{2.7}$$

where $L^{\mathrm{mono}} = \sum_i \sum_{j \neq i} a_{ij}^{\mathrm{mono}}$. Similarly, it is possible to define the corresponding filtered quantity ρ_b^{mono}, in analogy with (2.4b).

The transformed quantities $\mu_b^{\alpha\beta}$ and $\rho_b^{\alpha\beta}$ defined in (2.4) capture the similarity and reciprocity between layers of a multiplex via a comparison of the empirical values with the expected values under a null model. However, those quantities do not consider any information about the variances of the values of multiplexity and multireciprocity under the null model, thus giving no direct information about statistical significance. In particular, even multiplexes sampled from the null model with independent layers would be characterized by small, but in general nonzero, values of $\mu_b^{\alpha\beta}$ and $\rho_b^{\alpha\beta}$. This makes it difficult to disentangle, for

an observed real-world multiplex, weak inter-layer dependencies from pure noise. Moreover, the random fluctuations around the expectation values will be in general different for different pairs of layers, potentially making the comparison of the values of $\mu_b^{\alpha\beta}$ and $\rho_b^{\alpha\beta}$ for different pairs of layers misleading. To overcome these limitations, we define the z-scores associated to $m_b^{\alpha\beta}$ and $r_b^{\alpha\beta}$ as:

$$z\big(m_b^{\alpha\beta}\big) = \frac{m_b^{\alpha\beta} - \langle m_b^{\alpha\beta}\rangle_{\mathrm{DBCM}}}{\sqrt{\langle (m_b^{\alpha\beta})^2\rangle_{\mathrm{DBCM}} - \langle m_b^{\alpha\beta}\rangle_{\mathrm{DBCM}}^2}}, \tag{2.8a}$$

$$z\big(r_b^{\alpha\beta}\big) = \frac{r_b^{\alpha\beta} - \langle r_b^{\alpha\beta}\rangle_{\mathrm{DBCM}}}{\sqrt{\langle (r_b^{\alpha\beta})^2\rangle_{\mathrm{DBCM}} - \langle r_b^{\alpha\beta}\rangle_{\mathrm{DBCM}}^2}}. \tag{2.8b}$$

As for the quantities defined in (2.4), it is possible to obtain an analytical expression for the z-scores as well. This is shown in detail in Appendix 2.C.

Each z-score in (2.8) has the same sign as the corresponding quantity in (2.4), since the numerator is the same and both have positive denominators. However, except for the common sign, the two sets of quantities can have *a priori* very different values. In particular, the z-scores count the number of standard deviations by which the observed raw quantities deviate from their expected values under the null model. As such, they are useful in order to understand whether small measured values of $\mu_b^{\alpha\beta}$ or $\rho_b^{\alpha\beta}$ are actually consistent with zero within a small number of standard deviations, in which case we can consider the layers α and β as uncorrelated. We point out that, in general, z-scores have a clear statistical interpretation only if their distribution is Gaussian under repeated realizations of the model. In our case, although the quantities $m_b^{\alpha\beta}$ and $r_b^{\alpha\beta}$ are not truly normally distributed under the null model, they are defined as the sum of many independent 0/1 random variables (of the type $\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}$ or $\min\{a_{ij}^{\alpha}, a_{ji}^{\beta}\}$ respectively), which all have variance in the interval $(0, 1/4]$ and are thus approximately described by a central limit theorem ensuring an asymptotic convergence to the normal distribution. We can therefore consider as statistically significant all the z-scores having an absolute value larger than a given threshold, which we set at $z_c = 2$. This selects the observed pairs of layers with values of multiplexity and/or multireciprocity that differ from their expectation values by more than 2 standard deviations, i.e. with $|z| > z_c$.

2.2.3 Weighted multiplexity and multireciprocity

We now move to our second set of definitions, valid for weighted multiplexes. In analogy with (2.3), we define the *weighted* multiplexity and multireciprocity

matrices $\mathbf{M}_w$ and $\mathbf{R}_w$ having entries

$$m_w^{\alpha\beta} = \frac{2\sum_i \sum_{j\neq i} \min\{w_{ij}^\alpha, w_{ij}^\beta\}}{W^\alpha + W^\beta} = \frac{2W^{\alpha\rightrightarrows\beta}}{W^\alpha + W^\beta}, \tag{2.9a}$$

$$r_w^{\alpha\beta} = \frac{2\sum_i \sum_{j\neq i} \min\{w_{ij}^\alpha, w_{ji}^\beta\}}{W^\alpha + W^\beta} = \frac{2W^{\alpha\rightleftarrows\beta}}{W^\alpha + W^\beta}, \tag{2.9b}$$

where $W^\alpha = \sum_i \sum_{j\neq i} w_{ij}^\alpha$ is the total weight of the links in layer α (analogously for layer β), $W^{\alpha\rightrightarrows\beta} = \sum_i \sum_{j\neq i} \min\{w_{ij}^\alpha, w_{ij}^\beta\}$ is the total link weight of layer α that is multiplexed in layer β (clearly, $W^{\alpha\rightrightarrows\beta} = W^{\beta\rightrightarrows\alpha}$), and $W^{\alpha\rightleftarrows\beta} = \sum_i \sum_{j\neq i} \min\{w_{ij}^\alpha, w_{ji}^\beta\}$ is the total link weight of layer α that is reciprocated in layer β (clearly, $W^{\alpha\rightleftarrows\beta} = W^{\beta\rightleftarrows\alpha}$). The matrices $\mathbf{M}_w$ and $\mathbf{R}_w$ represent the two generalizations, for directed multiplexes, of the weighted multiplexity matrix introduced in the previous chapter and in [9] for undirected weighted multiplexes. Like their binary counterparts, both matrices have entries in the range $[0, 1]$, the maximum value being attained by identical ($w_{ij}^\alpha = w_{ij}^\beta$ for all i, j) and fully 'multireciprocated' ($w_{ij}^\alpha = w_{ji}^\beta$ for all i, j) layers respectively. In analogy with the corresponding binary case, the diagonal of $\mathbf{M}_w$ has all unit entries while that of $\mathbf{R}_w$ has entries that coincide with the recent definition of reciprocity for weighted monoplex networks [11].

In this case as well, for trivial multiplexes with sparse noninteracting layers and narrow strength distributions, the two matrices are expected to be asymptotically diagonal. However, this is no longer true in presence of dense layers and/or for broad strength distributions, and we therefore need a comparison of the raw quantities with their expected value under a null model (now the DWCM). This consideration leads us to introduce the transformed weighted multiplexity and multireciprocity matrices with entries

$$\mu_w^{\alpha\beta} = \frac{m_w^{\alpha\beta} - \langle m_w^{\alpha\beta}\rangle_{\mathrm{DWCM}}}{1 - \langle m_w^{\alpha\beta}\rangle_{\mathrm{DWCM}}}, \tag{2.10a}$$

$$\rho_w^{\alpha\beta} = \frac{r_w^{\alpha\beta} - \langle r_w^{\alpha\beta}\rangle_{\mathrm{DWCM}}}{1 - \langle r_w^{\alpha\beta}\rangle_{\mathrm{DWCM}}}, \tag{2.10b}$$

where $\langle\cdot\rangle_{\mathrm{DWCM}}$ denotes the expected value under the DWCM. As in the binary case, we can derive an analytical expression for the expected values that ultimately requires the expectation of the minimum of w_{ij}^α and w_{ij}^β (or w_{ji}^β). This is done in Appendix 2.D. It turns out that, under the DWCM, the distribution of link weights is geometrical [11, 21]:

$$P(w_{ij}^\alpha) = (p_{ij}^\alpha)^{w_{ij}^\alpha}(1 - p_{ij}^\alpha), \tag{2.11}$$

where p_{ij}^α denotes again the probability that a directed link (of any positive weight) from node i to node j is realized in layer α. The above probability can be used

to calculate $\mu_w^{\alpha\beta}$ and $\rho_w^{\alpha\beta}$ analytically as discussed in Appendix 2.D.

The weighted multireciprocity of the multiplex can be conveniently compared with the weighted reciprocity of the aggregated monoplex network. The link weights of the latter are defined by

$$w_{ij}^{\text{mono}} = \sum_{\alpha=1}^{M} w_{ij}^{\alpha}, \tag{2.12}$$

and the associated aggregate weighted reciprocity [11] is

$$r_w^{\text{mono}} = \frac{\sum_i \sum_{j\neq i} \min\{w_{ij}^{\text{mono}}, w_{ji}^{\text{mono}}\}}{W^{\text{mono}}} \tag{2.13}$$

(where $W^{\text{mono}} = \sum_i \sum_{j\neq i} w_{ij}^{\text{mono}}$). The corresponding filtered value ρ_w^{mono} can be defined as in (2.10b).

In analogy with the binary case, it is possible to define the z-scores associated to $m_w^{\alpha\beta}$ and $r_w^{\alpha\beta}$ as follows:

$$z\big(m_w^{\alpha\beta}\big) = \frac{m_w^{\alpha\beta} - \langle m_w^{\alpha\beta}\rangle_{\text{DWCM}}}{\sqrt{\langle (m_w^{\alpha\beta})^2\rangle_{\text{DWCM}} - \langle m_w^{\alpha\beta}\rangle_{\text{DWCM}}^2}}, \tag{2.14a}$$

$$z\big(r_w^{\alpha\beta}\big) = \frac{r_w^{\alpha\beta} - \langle r_w^{\alpha\beta}\rangle_{\text{DWCM}}}{\sqrt{\langle (r_w^{\alpha\beta})^2\rangle_{\text{DWCM}} - \langle r_w^{\alpha\beta}\rangle_{\text{DWCM}}^2}}. \tag{2.14b}$$

The explicit analytical expressions for these z-scores are calculated in Appendix 2.D. Again, the z-scores (2.14) have the same signs as the corresponding quantities (2.10), but in addition they allow to test for statistical significance using e.g. a threshold of $z_c = 2$.

2.3 Empirical analysis of the World Trade Multiplex

In this section, we apply the framework defined so far to the analysis of a real-world system. This system is the World Trade Multiplex (WTM), defined as the multi-layer network representing the directed trade relations between world countries in different commodities. At both the binary and the weighted level, the structure of the aggregate (monoplex) version of this network is well studied [31, 32, 33], as well as that of many of its layers separately [8, 34]. However, much less is known about the inter-layer dependencies in the WTM. In particular, an assessment of the inter-layer couplings that are not simply explained by

the local topological properties of the WTM has been carried out only for the undirected version of the network [9]. Given the importance of the directionality of trade flows, especially at the disaggregated level of individual commodities, it is therefore important to carry out a directed analysis of the WTM. The tools we have introduced in the previous section allow us to make this step and arrive at a novel characterization of the WTM where the undirected multiplexity properties documented in the previous chapter [9] are resolved into their two directed components, namely multiplexity and multireciprocity. These results have important potential implications for problems related to research on international trade, such as the definition of trade-based 'product taxonomies' [8], the construction of the 'product space' [35], and the calculation of 'fitness and complexity' metrics [36]. These points are discussed later in sec. 2.4.

2.3.1 Data

We use the already mentioned BACI-Comtrade dataset [37] where international trade flows among all countries of the world are disaggregated into different commodity classes at the 2-digit resolution level, defined as in the standard HS1996 classification [38] of traded goods. Here we take into account the directionality of trade, hence distinguishing between import and export. As explained in the previous chapter, it is possible to represent this dataset as a multiplex as in [8, 9, 34]. In particular, we will consider a multi-layer representation defined by $N = 207$ nodes (countries) and $M = 96$ layers (commodities), for the year 2011. Since each trade exchange is reported by both the importer and the exporter (and the two values may in general differ), the dataset uses a reconciliation procedure to get a unique value for each flow (see [37] for details). All the resulting trade volumes are expressed in thousands of dollars in the dataset. Since our approach works for integer link weights, all the reported trade values have been rescaled by first dividing by 10 (for computational reasons) and then rounding to the closest integer. This defines our integer link weights $\{w_{ij}^{\alpha}\}$ for all layers. For each entry w_{ij}^{α}, we then define $a_{ij}^{\alpha} = 1$ if $w_{ij}^{\alpha} > 0$ and $a_{ij}^{\alpha} = 0$ otherwise. We point out that the rounding procedure does not significantly affect the structure of the system under study, as the percentage of original links which are lost (i.e. rounded to zero) is negligible.

From the multiplex trade flows we also compute the aggregate binary and weighted links a_{ij}^{mono} and w_{ij}^{mono} between any two countries i and j in the collapsed monoplex trade network, as in (2.6) and (2.12) respectively. This allows us to compare the multiplex structure of trade with the aggregate one and highlight relevant information that is lost in the aggregation procedure. For instance, for both the binary and the weighted representation of the system, we can compare the values of the multireciprocity matrix measured on the commodity-resolved multiplex with the usual scalar reciprocity measured on the monoplex aggregate trade network.

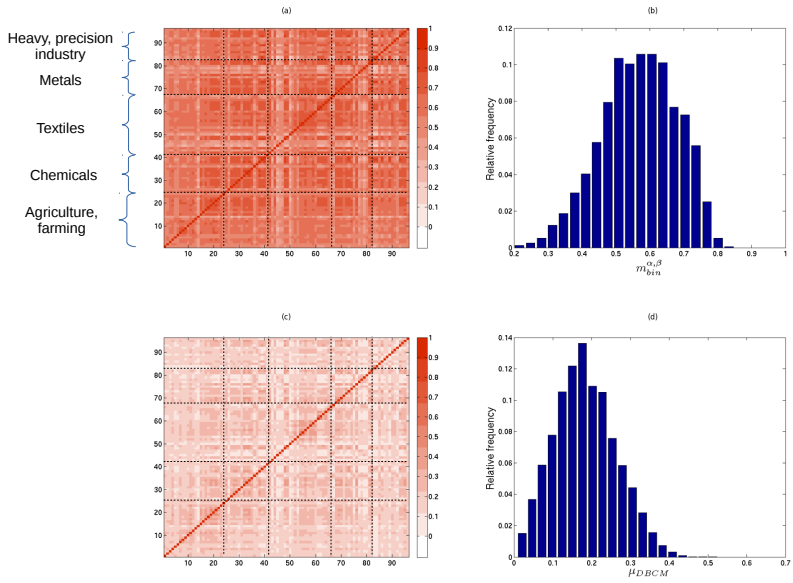

Figure 2.1: **Analysis of the binary multiplexity between layers of the WTM.** Top panels: color-coded binary multiplexity matrix $\mathbf{M}_b$ (a) and corresponding distribution of off-diagonal multiplexity values $m_b^{\alpha\beta}$ (with $\alpha \neq \beta$) (b). Bottom panels: same as for the top panels, but with raw binary multiplexity $m_b^{\alpha\beta}$ replaced by rescaled binary multiplexity $\mu_b^{\alpha\beta}$.

2.3.2 Binary analysis

We start with a binary analysis of the WTM, thus taking into account only the topology of the various layers while disregarding the information about trade volumes. In Figure 2.1(a) we show the color-coded binary multiplexity matrix $\mathbf{M}_b$. Next to it, in Figure 2.1(b) we show the corresponding frequency distribution of off-diagonal matrix entries $m_b^{\alpha\beta}$ (with $\alpha \neq \beta$). In calculating the frequencies, we discard the diagonal entries because they trivially evaluate to $m_b^{\alpha\alpha} = 1$, as discussed above. High values of multiplexity are observed for most of the pairs of commodities. This result is in agreement with what has been reported in [9] (see Chapter 1) on the basis of an undirected analysis of the WTM where imports and exports between any two countries were combined together into a single trade link.

As we mentioned, the multiplexity matrix $\mathbf{M}_b$ would be asymptotically diag-

onal for trivial multiplexes with sparse non-interacting layers and narrow degree distributions. However, since the layers of the WTM are very dense and their degree distributions significantly broad [8, 9, 34], this system is an ideal case study requiring the use of a null model in order to assess the presence of a genuine coupling among layers. In Figure 2.1(c) we show the color-coded matrix of rescaled multiplexity values $\mu_b^{\alpha\beta}$, which control for the effects of the heterogeneity of the layer-specific in- and out-degree sequences. Similarly, in Figure 2.1(d) we show the corresponding distribution of off-diagonal entries. We find that, after controlling for the degrees, a significant amount of correlation is destroyed. However all the values are still strictly positive, indicating a tendency of all pairs of commodities to be 'traded together'. The statistical significance of this result is discussed later in terms of z-scores.

We now move to the analysis of multireciprocity. It is known that, when the aggregate trade in all commodities is considered, the binary monoplex representation of the World Trade Network exhibits a high level of reciprocity [10, 39, 40]. It is interesting to see whether such a property is preserved also at the multiplex level, and how the values compare with the aggregate case. Figure 2.2(a) shows the color-coded binary multireciprocity matrix $\mathbf{R}_b$ and Figure 2.2(b) the corresponding distribution of off-diagonal entries [1], with a superimposed delta function indicating the value of the binary reciprocity r_b^{mono} of the aggregate monoplex network as a comparison. The results are comparable with those found above for the multiplexity. Also in this case, the high multireciprocity values are consistent with the high multiplexity values found for the undirected representation of the WTM [9] (where pairs of reciprocated links in each layer are merged into single undirected links). However, for the multireciprocity this result is much less trivial than for the multiplexity, given the chosen level of disaggregation into many commodity classes. Indeed one would expect that, at such a relatively high resolution, it should be not very likely (at least not as likely as in the undirected representation) that the same commodity is traded "back and forth", i.e. both ways between the same two countries. In any case we do find, in accordance with what we expect, that for all pairs of commodities the multireciprocity is significantly smaller than the reciprocity r_b^{mono} of the aggregate monoplex. This means that, as layers are aggregated, there is a bigger relative increment (with respect to individual layers) in the overall number of reciprocated links than in the total number of links.

As an interesting result, the intra-layer reciprocity values $r_b^{\alpha\alpha}$ lying along the diagonal of the multireciprocity matrix are found to be very similar to the values of the matrix entries $r_b^{\alpha\beta}$ lying close to the diagonal. Indeed, in the matrix plot of Figure 2.2(a) the diagonal is visually indistinguishable from the entries of the ma-

[1] We discard the diagonal entries in order to make the distribution compatible with the corresponding distribution for the multiplexity shown above; in any case, if the diagonal entries are included, the distribution looks very similar.

trix that are "nearby". Given the order of the commodities in the matrix (as shown in the appendix of Chapter 1), these nearby entries represent the multireciprocity between pairs of similar commodities. This result means that the high reciprocity of the aggregate trade monoplex does *not* arise from the superposition of layers with high internal reciprocity and low mutual multireciprocity (as would be the case in presence of an approximately diagonal multireciprocity matrix). Rather, we find that a trade flow in one commodity α tends to be reciprocated by comparable trade flows in several different commodities, including (but not dominated by) the same commodity α and many other related commodities. Specifically, it can be seen from Figure 2.2(a) that layers characterized by low (high) values of internal reciprocity are embedded within groups of layers with low (high) mutual multireciprocity. This suggests that the level of reciprocity in international trade is not an intrinsic property of individual commodities, but rather a property of whole groups of mutually reciprocated commodities with comparable multireciprocity values.

In Figure 2.2(c) and (d) we show the color-coded binary rescaled multireciprocity matrix and the corresponding distribution of off-diagonal entries $\rho_b^{\alpha\beta}$ (with $\alpha \neq \beta$). The relatively small values (with respect to the non-rescaled quantities) indicate that, in analogy with what we found for the multiplexity, the apparent correlation between the topology of pairs of layers is largely encoded in the relatedness of the degree sequences of such pairs. For the vast majority of pairs of commodities the multireciprocity is still lower than that measured on the aggregate network. However, all pairs of layers preserve a positive residual multireciprocity, the statistical significance of which is studied later in our z-score analysis.

When we look at the multiplexity matrix in Figure 2.1(a) and the corresponding multireciprocity matrix in Figure 2.2(a), we see the appearance of similar patterns. Such similarity is further investigated in Figure 2.3(a), where we report the scatter plots of pairwise multireciprocity values versus the corresponding multiplexity values. We observe a roughly linear trend, which is however lost when we look at the filtered values, as shown in Figure 2.3(b). We see that, in the latter case, the relationship between $\rho_b^{\alpha\beta}$ and $\mu_b^{\alpha\beta}$ is non-linear and significantly scattered. Although the presence of a non-linear relation may be related to the particular choice of normalization adopted in (2.4), we point out that the entity of the scatter is so big that it is not possible to retrieve the value of multiplexity from the multireciprocity, and vice-versa. This illustrates that the two quantities convey different pieces of information that are irreducible to each other.

Similar considerations apply to the z-scores. In Figure 2.4(a) and 2.4(b) we show the empirical relation between the transformed multiplexity and multireciprocity and their corresponding z-scores: it is worth recalling that the informa-

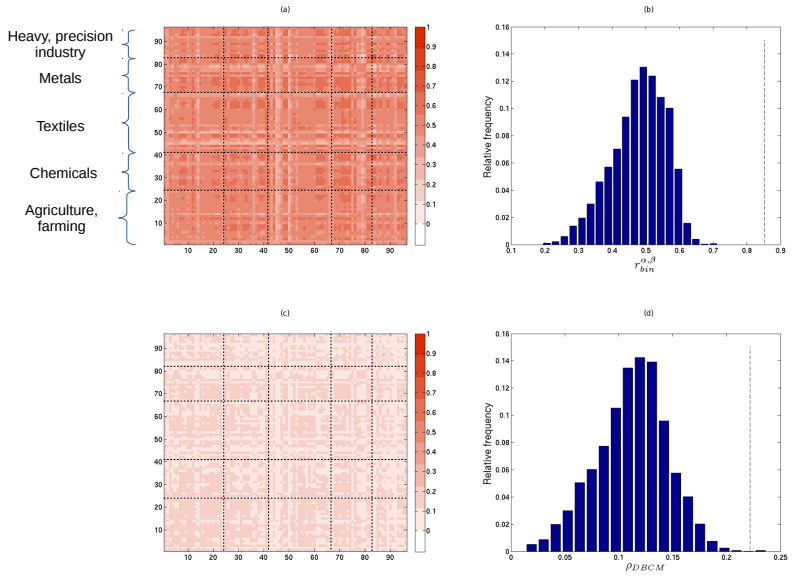

Figure 2.2: **Analysis of the binary multireciprocity between layers of the WTM.** Top panels: color-coded binary multireciprocity matrix $\mathbf{R}_b$ (a) and corresponding distribution of off-diagonal multireciprocity values $r_b^{\alpha\beta}$ (with $\alpha \neq \beta$) (b). Bottom panels: same as for the top panels, but with raw binary multireciprocity $r_b^{\alpha\beta}$ replaced by rescaled binary multireciprocity $\rho_b^{\alpha\beta}$. The dashed lines represent the value of (raw and rescaled) binary reciprocity r_b^{mono} and ρ_b^{mono} of the aggregated monoplex network.

tion provided by these two quantities can be *a priori* different, given the lack of information about the standard deviation in the rescaled multiplexity and multireciprocity metrics. Empirically, we however find a strong correlation between these quantities, indicating that large values of binary multiplexity or multireciprocity correspond to large z-scores, and vice-versa. Moreover, even the smallest z-scores (those found for the pairs of layers showing very low multiplexity or multireciprocity) are still quite high (i.e. positive and larger than $z_c = 2$) in terms of statistical significance. This means that even the pairs of layers with smallest multiplexity or multireciprocity should be considered as significantly and positively correlated. We therefore conclude that, at a binary level, every commodity of the WTM tends to be traded together with all other commodities, both in the same and in the opposite direction. As we show below, this is no longer the case when the weighted

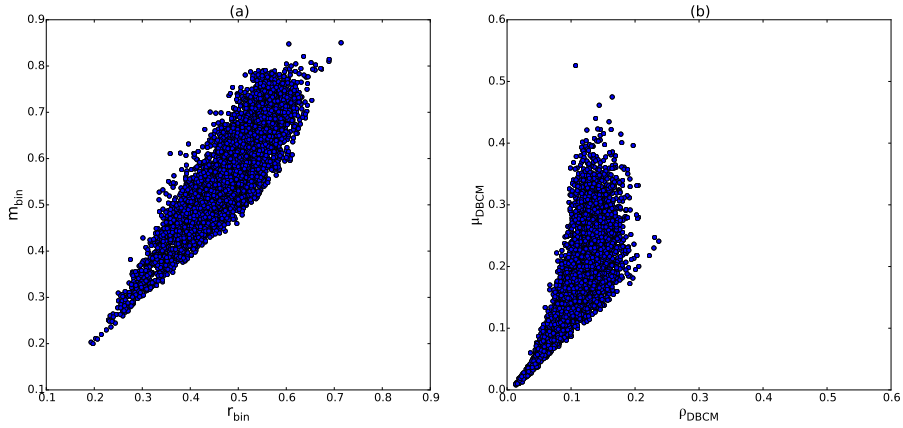

Figure 2.3: **Relation between the values of binary multiplexity and multireciprocity for the WTM.** Scatter plots of off-diagonal binary multireciprocity values versus off-diagonal binary directed multiplexity values. Left: raw values ($r_b^{\alpha\beta}$ vs $m_b^{\alpha\beta}$); right: rescaled values ($\rho_b^{\alpha\beta}$ vs $\mu_b^{\alpha\beta}$).

version of the multiplex is considered.

Figure 2.4(c) shows the relation existing between $z(r_b^{\alpha\beta})$ and $z(m_b^{\alpha\beta})$ for each pair of layers. If we compare this figure with Figure 2.3, we see that in this case the trend is more linear, although the scatter is again quite large. This confirms that it is not possible to recover the values of multiplexity from those of multireciprocity, and vice-versa.

2.3.3 Weighted analysis

We now perform a weighted analysis of the World Trade Multiplex, by taking into account the values of import and export observed between countries.

In Figure 2.5(a) and (b) we show the color-coded weighted directed multiplexity matrix $\mathbf{M}_w$ and the distribution of its off-diagonal entries. We clearly see that, even though several pairs of commodities are still strongly overlapping, the multiplexity distribution is concentrated over a range of significantly smaller values with respect to the corresponding binary distribution. Indeed, the notion of weighted multiplexity, by involving the minimum of the weights of two reciprocated links, provides a stricter criterion with respect to the unweighted case. In particular, for any pair of nodes and any pair of layers, it is more unlikely to achieve the maximum weighted value $\min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}$ than the maximum binary

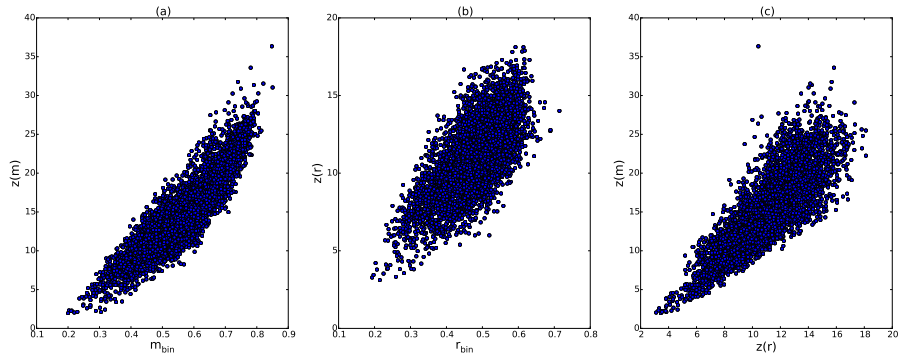

Figure 2.4: **Analysis of significance of the values of binary multiplexity and multireciprocity for the WTM.** Left: binary transformed multiplexity $\mu_b^{\alpha\beta}$ versus its corresponding z-score $z(m_b^{\alpha\beta})$; center: binary transformed multireciprocity $\rho_b^{\alpha\beta}$ versus its corresponding z-score $z(r_b^{\alpha\beta})$; right: $z(r_b^{\alpha\beta})$ vs $z(m_b^{\alpha\beta})$. Only off-diagonal values are reported.

value $\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}$. Lower values of multiplexity with respect to Figure 2.1(a) are therefore expected. We also expect to find a similar reduction for the multireciprocity later on.

In Figure 2.5(c) and (d) we report the color-coded weighted rescaled multiplexity matrix and the corresponding distribution of off-diagonal entries $\mu_w^{\alpha\beta}$. The fact that many values are now mapped to zero means that a significant component of the overlap between commodities can be explained simply in terms of the correlated strength sequences of the various layers. Importantly, we see that some pairs of layers actually exhibit negative rescaled multiplexity, even though the distribution is far from symmetric. This result, which is only visible in the weighted analysis, means that there are pairs of commodities for which the observed trade multiplexity is actually *lower* than expected under the null model: these commodities prefer 'not to be traded together'.

We then analyze the weighted multireciprocity of the WTM. Recently, it has been shown that the aggregated version of the network has a strong weighted reciprocity [11], a result that we can now complement with the analysis of the disaggregated multiplex. In Figure 2.6(a) and (b) we report the color-coded weighted multireciprocity matrix $\mathbf{R}_w$, along with the distribution of its off-diagonal entries. In analogy with the binary case, we see that the aggregated network exhibits a reciprocity which is significantly higher than the multireciprocity associated to any individual pair of layers. Yet several pairs of commodities are characterized by a substantial level of multireciprocity. In Figure 2.6(c) and (d) we show the

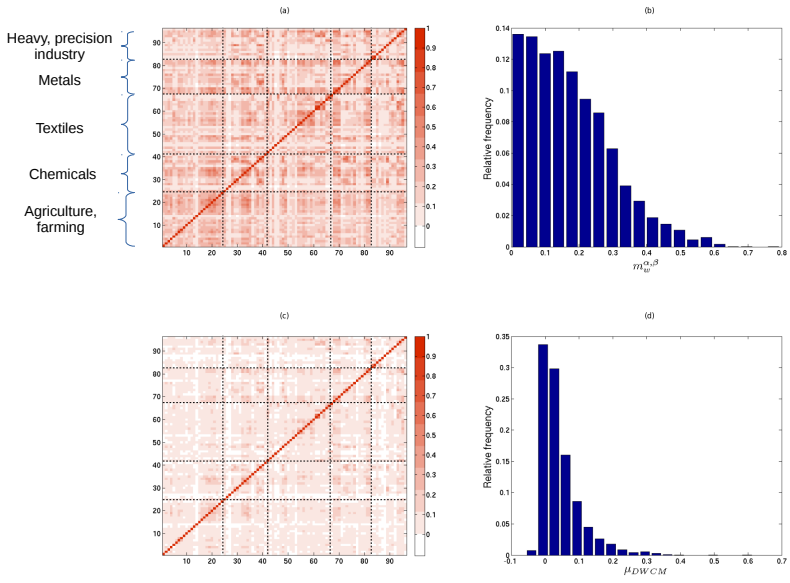

Figure 2.5: **Analysis of the weighted multiplexity between layers of the WTM.** Top panels: color-coded weighted multiplexity matrix $\mathbf{M}_w$ (a) and corresponding distribution of off-diagonal multiplexity values $m_w^{\alpha\beta}$ (with $\alpha \neq \beta$) (b). Bottom panels: same as for the top panels, but with raw weighted multiplexity $m_w^{\alpha\beta}$ replaced by rescaled weighted multiplexity $\mu_w^{\alpha\beta}$. Note that, in panel (c), white entries represent negative values.

corresponding results for the rescaled weighted multireciprocity $\rho_w^{\alpha,\beta}$. We see that many values become close to zero and some become negative, in analogy with the behaviour of the multiplexity. The identification of pairs of layers with negative rescaled multireciprocity indicates that the corresponding commodities 'prefer not to be traded in opposite directions', in contrast with the results we found in the binary analysis.

In Figure 2.7 we compare the weighted multireciprocity and the weighted multiplexity. When we consider the raw values (a), we observe a clear linear trend (although more scattered than in the corresponding unweighted case). The trend becomes even more robust, and less noisy, for the filtered values, as shown in (b). In both panels, the most significant commodities (both in terms of trade volumes and economic relevance) mainly lie along the diagonal, while the outliers represent

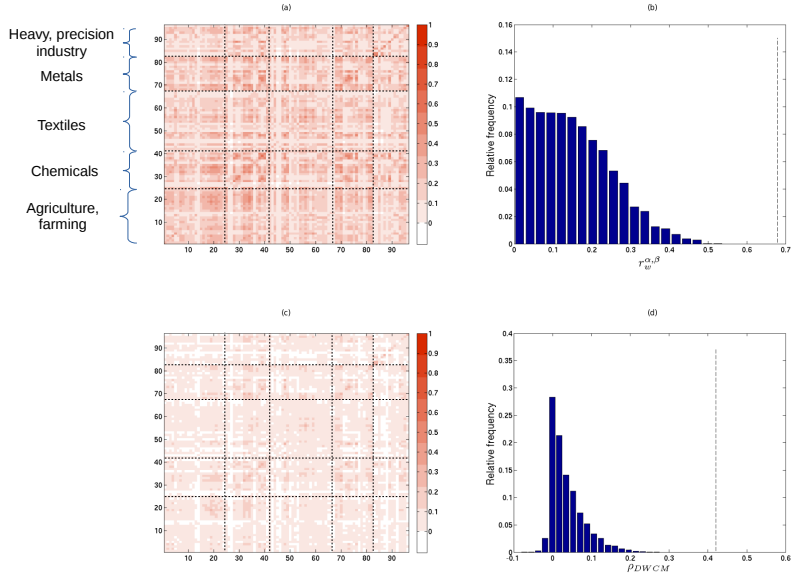

Figure 2.6: **Analysis of the weighted multireciprocity between layers of the WTM.** Top panels: color-coded weighted multireciprocity matrix $\mathbf{R}_w$ (a) and corresponding distribution of off-diagonal multireciprocity values $r_w^{\alpha\beta}$ (with $\alpha \neq \beta$) (b). Bottom panels: same as for the top panels, but with raw weighted multireciprocity $r_w^{\alpha\beta}$ replaced by rescaled weighted multireciprocity $\rho_w^{\alpha\beta}$. The dashed lines represent the value of (raw and rescaled) weighted reciprocity r_w^{mono} and ρ_w^{mono} of the aggregated monoplex network. Note that, in panel (c), white entries represent negative values.

less relevant products (for instance, some textiles or less traded craft goods). We also see pairs of commodities whose multireciprocity is similar to the reciprocity of the aggregate trade network. These commodities, such as cereals and heavy industry products, are not necessarily the most traded ones, still they better represent the reciprocity patterns of total trade among countries, possibly because they give the main contribution to the reciprocity of the aggregated network.

Quantitatively, another important difference between the binary and the weighted approach lies in the statistical significance of the values of multiplexity and multireciprocity, as we can see from the analysis of the z-scores (Figure 2.8). Indeed, in the unweighted case we found that even the smallest values of $\mu_b^{\alpha\beta}$ and $\rho_b^{\alpha\beta}$

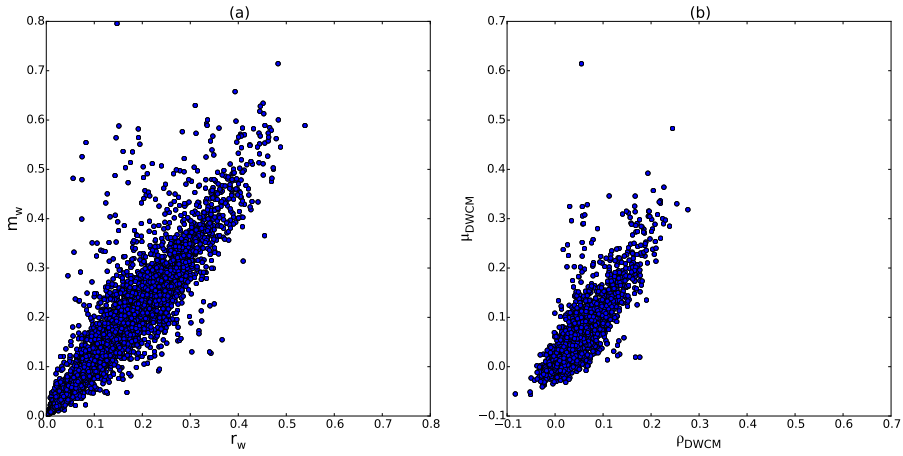

Figure 2.7: **Relation between the values of weighted multiplexity and multireciprocity for the WTM.** Scatter plots of off-diagonal weighted multireciprocity values versus off-diagonal weighted directed multiplexity values. Left: raw values ($r_w^{\alpha\beta}$ vs $m_w^{\alpha\beta}$); right: rescaled values ($\rho_w^{\alpha\beta}$ vs $\mu_w^{\alpha\beta}$).

are significant, as the corresponding z-scores are larger than the critical value z_c. Instead, here we observe almost no correlation (except for the aforementioned sign concordance) between weighted multiplexity or multireciprocity and the corresponding z-scores (see Fig. 2.8(a) and 2.8(b) respectively). Indeed, the same value of $\mu_b^{\alpha\beta}$ or $\rho_b^{\alpha\beta}$ may even correspond to z-scores with different orders of magnitude. This means that, even for two pairs of layers with the same observed value of weighted multiplexity or multireciprocity, the statistical significance of the inter-layer coupling can be very different. Moreover, the absolute value of many weighted z-scores is found below the significance threshold $z_c = 2$, identifying pairs of uncorrelated layers (a result that is unobserved in the binary case). Finally, many pairs of commodities have a negative z-score below $-z_c$ for the multiplexity and/or multireciprocity. For these pairs, the tendency *not* to be traded in the same direction and/or in opposite direction is statistically validated and confirms a difference with respect to the binary case.

As a final result, in Figure 2.8(c) we show the relation existing between $z\big(m_w^{\alpha\beta}\big)$ and $z\big(r_w^{\alpha\beta}\big)$. We find an overall level of correlation which however leaves room for a significant scatter of points around the identity line. This scatter is big enough to imply that, for a given significance threshold z_c, the pairs of commodities can be partitioned in the following five classes:

1. a few pairs of commodities that tend to be traded in the same direction

($z(m_w^{\alpha\beta}) > z_c$) but not in opposite directions ($z(r_w^{\alpha\beta}) < -z_c$): examples are apparel articles vs ships and boats; food industry residues, prepared animal feed vs ores, slag and ash;

2. a few pairs of commodities that tend to be traded in opposite directions ($z(r_w^{\alpha\beta}) > z_c$) but not in the same direction ($z(m_w^{\alpha\beta}) < -z_c$): examples are ores, slag and ash vs footwear and gaiters; apparel articles vs ores, slag and ash;

3. a moderately-sized group of pairs of commodities that tend to be traded neither in the same direction ($z(m_w^{\alpha\beta}) < -z_c$) nor in opposite ones ($z(r_w^{\alpha\beta}) < -z_c$): examples are raw hides and skins vs arms and ammunitions; tobacco vs ships and boats;

4. a large group of pairs of commodities for which there is no statistically significant tendency in at least one of the two directions ($|z(m_w^{\alpha\beta})| < z_c$ and/or $|z(r_w^{\alpha\beta})| < z_c$): examples are tobacco vs inorganic chemicals; explosives, pyrotechnic products vs vehicles (note that this class can be further split in sub-classes where commodities are uncorrelated in one direction but correlated in different ways in the other direction);

5. a very large group of pairs of commodities that tend to be traded both in the same direction ($z(m_w^{\alpha\beta}) > z_c$) and in opposite ones ($z(r_w^{\alpha\beta}) > z_c$): examples are sugar vs cocoa; soap, waxes, candles vs sugar.

It should be noted that, in contrast with the above classification, the binary analysis concluded that all pairs of commodities belong to the last class only.

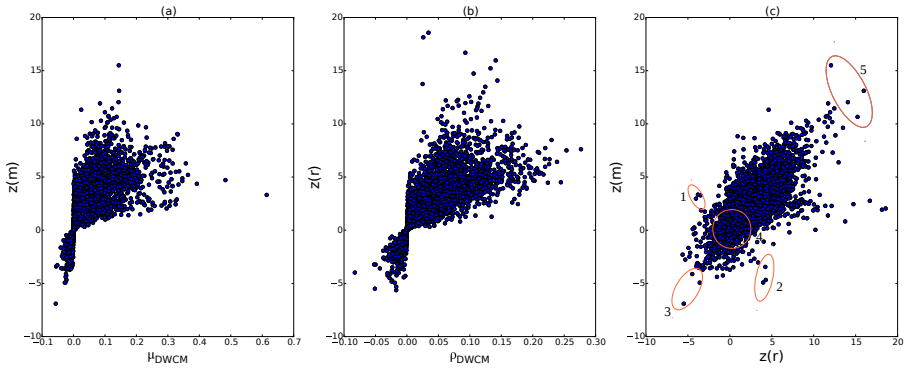

Figure 2.8: **Analysis of significance of the values of weighted multiplexity and multireciprocity for the WTM.** Left: weighted transformed multiplexity $\mu_w^{\alpha\beta}$ versus its corresponding z-score $z(m_w^{\alpha\beta})$; center: weighted transformed multireciprocity $\rho_w^{\alpha\beta}$ versus its corresponding z-score $z(r_w^{\alpha\beta})$; right: $z(r_w^{\alpha\beta})$ vs $z(m_w^{\alpha\beta})$. In panel (c), numbered circles correspond to the bullet points reported in Sec. 2.3.3. Only off-diagonal values are reported.

2.4 Discussion and conclusions

The study of multi-layer networks has received substantial attention in the last few years, leading to the introduction of several novel quantities characterizing the structure of multiplexes as well as the behaviour of several dynamical processes taking place on them. The aim of all these studies is that of highlighting the role of the inter-layer couplings, the latter being the ultimate reason why layers of a multiplex should be analyzed together in the first place, rather than separately. In this chapter we have argued that even the *simplest* definitions of inter-layer coupling, based merely on the structural overlap of links across layers, are strongly biased by the density, finiteness, and heterogeneity of the network. We have shown that controlling for the above effects requires a quite elaborate statistical treatment. Focusing on multiplexes with (binary or weighted) directed links, we have introduced maximum-entropy multiplex ensembles with given node properties as the unbiased null models serving as a benchmark for the empirically observed properties. We have then defined novel multiplexity and multireciprocity metrics, respectively quantifying the tendency of pairs of links to 'align' and/or 'anti-align' across each pair of layers of a real-world directed multiplex. Since links can exist in both directions in every layer, the possible tendencies of forming aligned (multiplexed) and anti-aligned (multireciprocated) links do not conflict with each other and can actually coexist. Both multiplexity and multireciprocity are matrix-valued, as they represent the possible couplings among all pairs of layers. While multiplexity is a natural extension of the corresponding definition

for undirected multiplexes, multireciprocity is a novel concept representing a non-trivial extension of the notion of single-layer reciprocity to multi-layer networks.

We believe that our results can be of value for several applications. For instance, they provide a statistically rigorous way to identify possible (groups of) layers that are uncorrelated from the other layers, thus allowing to simplify the whole multiplex into mutually independent sub-systems with smaller numbers of layers. This problem has received significant attention recently [41, 42]. Our finding of a strong influence of the local node properties on the overall level of inter-layer coupling suggests that many of the results found with alternative techniques that do not control for these effects might be subject to an uncontrolled level of bias.

Other more specific applications are relevant for the specific case study of the WTM. In extreme summary, our detailed analysis of this system confirmed that its multiplex structure contains much more information than the aggregated network of total trade does. At a binary level, we found that all pairs of commodities tend to be traded together between countries, both in the same direction (high multiplexity) and in opposite directions (high multireciprocity). At a weighted level, this result only holds for a subset of pairs of commodities. Other commodity pairs are not correlated and others even tend to avoid being traded together in the same direction and/or in opposite ones. The multireciprocity structure of the WTM highlights a tendency of groups of commodities to have a comparably high mutual reciprocity, of the same entity of the internal single-layer reciprocity of these commodities. When aggregated into the monoplex network of total international trade, the WTM has a resulting reciprocity that is much bigger than the multireciprocity among its constituent layers.

In the light of the above results, our approach has implications relevant to various directions in international trade research. In particular, it indicates concrete ways to refine existing measures of inter-commodity correlation or similarity that are widely used to construct, among others, 'product taxonomies' [8], the 'product space' [35] and 'fitness and complexity' metrics [36]. All these applications are briefly explained below.

Inter-commodity correlation metrics have been introduced to quantify the coupling among layers of the WTM [8], with the goal of constructing 'product taxonomies' that reflect empirical trade similarities, as opposed to pre-defined product categories. However, as already pointed out inthe first chapter and in [9], correlation metrics make an implicit and totally unrealistic assumption of structural homogeneity of the network, by interpreting all the edges of a layer as independent observations drawn from the same probability distribution. Our results provide alternative metrics of inter-layer coupling that replace the homogeneity assumption with a much more realistic null model that accurately controls for

the observed degree of node heterogeneity in each layer. The use of our metrics is likely to change the structure of correlation-based product taxonomies significantly.

The 'product space' is defined as a network of commodities connected by links whose weight quantifies the tendency of a pair of commodities to be traded together (in the same direction) between the same two countries [35]. Our results clearly indicate that, to be statistically reliable, such an analysis should include a way to filter out the strong empirical heterogeneity of node degrees and/or node strengths. Moreover, they highlight a second layer of information that should be relevant for the product space construction, namely the fact that, besides the tendency of pairs of commodities to be traded together in the same direction (multiplexity), there can be a substantial tendency of being traded in the opposite direction (multireciprocity). We found that these two effects have a comparable magnitude. We also found that pairs of commodities with approximately the same multiplexity can be characterized by very different levels of multireciprocity. This suggests that neglecting multireciprocity in the construction of the product space can represent a substantial loss of information.

Finally, the 'fitness and complexity' approach focuses on the bipartite network of countries and their exported products, and uses the structure of this network to recursively define metrics of product complexity and country competitiveness (fitness) [36]. This method can reveal the 'hidden' potential of countries that is not (yet) reflected in their current GDP levels. Clearly, the output of this approach entirely depends on how the bipartite country-product matrix is constructed. This matrix is ultimately a projection of the WTM but is generally filtered using a null model based on the concept of 'revealed comparative advantage' [43], which however operates at the aggregate country-product level and not at the level of the underlying multiplex. As such, it does not control for the size of importers. Our approach provides a way to enforce a more accurate null model on the original WTM and obtain an alternative bipartite country-product projection.

We believe that all the research directions outlined above deserve future explorations and we expect the results reported in this chapter to be of use.

Appendix

2.A Maximum-entropy method for multiplex networks

As in the previous chapter, we define null models of multiplexes as canonical maximum-entropy ensembles satisfying a given set $\overrightarrow{\mathcal{C}}$ of $\mathcal{K}$ constraints on aver-

age. If $G^\alpha \in \mathcal{G}_N$ denotes the graph realized in layer α of the multiplex (recall that $\mathcal{G}_N$ is the set of all directed monoplex graphs with N nodes), and if $\overrightarrow{G} \in \mathcal{G}_N^M$ denotes the entire multiplex (where $\mathcal{G}_N^M$ is the set of all directed multiplex graphs with N nodes and M layers), we write $\overrightarrow{G} = (G^\alpha)_{\alpha=1}^M$. Now let $\overrightarrow{\mathcal{C}}$ denote a vector-valued function on $\mathcal{G}_N^M$, evaluating to $\overrightarrow{\mathcal{C}}(\overrightarrow{G})$ on the particular multiplex $\overrightarrow{G}$. The vector $\overrightarrow{\mathcal{C}}(\overrightarrow{G})$ is to be regarded as a set of structural properties measured on $\overrightarrow{G}$.

A canonical ensemble of binary (weighted) directed multiplex networks with the *soft constraint* $\overrightarrow{\mathcal{C}}$ is specified by a probability distribution $\mathcal{P}\big(\overrightarrow{G}\,|\,\overrightarrow{\theta}\big)$ on $\mathcal{G}_N^M$, where $\overrightarrow{\theta}$ is a vector of Lagrange multipliers required to enforce a desired expected value

$$\langle \overrightarrow{\mathcal{C}} \rangle_{\overrightarrow{\theta}} = \sum_{\overrightarrow{G} \in \mathcal{G}_N^M} \mathcal{P}\big(\overrightarrow{G}\,|\,\overrightarrow{\theta}\big)\, \overrightarrow{\mathcal{C}}(\overrightarrow{G}) \tag{2.15}$$

of $\overrightarrow{\mathcal{C}}$. Note that both $\overrightarrow{\theta}$ and $\overrightarrow{\mathcal{C}}$ are vectors of numbers with the same (but model-dependent) dimension $\mathcal{K}$, while $\overrightarrow{G}$ is always an M-dimensional vector of graphs. Obviously, an additional constraint on the probability is the normalization condition

$$\sum_{\overrightarrow{G} \in \mathcal{G}_N^M} \mathcal{P}\big(\overrightarrow{G}\,|\,\overrightarrow{\theta}\big) = 1 \qquad \forall\ \overrightarrow{\theta}. \tag{2.16}$$

We want our ensembles to produce multiplexes with independent layers. This requirement corresponds to the enforcement of separate constraints on the different layers, i.e. $\overrightarrow{\mathcal{C}} = (\overrightarrow{C^\alpha})_{\alpha=1}^M$, where $\overrightarrow{C^\alpha}$ is a K^α-dimensional vector of structural properties of the network in layer α only, evaluating to $\overrightarrow{C^\alpha}(G^\alpha)$ on the particular single-layer graph G^α. This leads to a separation in the corresponding Lagrange multipliers, i.e. $\overrightarrow{\theta} = (\overrightarrow{\theta^\alpha})_{\alpha=1}^M$. K^α is the dimension of both $\overrightarrow{C^\alpha}$ and $\overrightarrow{\theta^\alpha}$, and we must have $\sum_{\alpha=1}^M K^\alpha = \mathcal{K}$. Consequently, we can express the entropy of the ensemble of multiplex networks as

$$\begin{aligned} \mathcal{S}\big(\overrightarrow{\theta}\big) &\equiv -\sum_{\overrightarrow{G} \in \mathcal{G}_N^M} \mathcal{P}\big(\overrightarrow{G}\,|\,\overrightarrow{\theta}\big) \ln \mathcal{P}\big(\overrightarrow{G}\,|\,\overrightarrow{\theta}\big) \\ &= \sum_{\alpha=1}^M S^\alpha\big(\overrightarrow{\theta^\alpha}\big), \end{aligned} \tag{2.17}$$

where

$$S^\alpha\big(\overrightarrow{\theta^\alpha}\big) \equiv -\sum_{G^\alpha \in \mathcal{G}_N} P^\alpha\big(G^\alpha|\overrightarrow{\theta^\alpha}\big) \ln P^\alpha\big(G^\alpha|\overrightarrow{\theta^\alpha}\big) \tag{2.18}$$

is the entropy of the ensemble of monoplex graphs for the individual layer α, with $P^\alpha\big(G^\alpha|\overrightarrow{\theta^\alpha}\big)$ subject to the normalization condition

$$\sum_{G^\alpha\in\mathcal{G}_N} P^\alpha\big(G^\alpha|\overrightarrow{\theta^\alpha}\big)=1 \qquad \forall\ \overrightarrow{\theta^\alpha} \qquad \alpha=1,M. \tag{2.19}$$

At this point, we want to maximize the entropy $\mathcal{S}\big(\overrightarrow{\theta}\big)$, subject to the soft constraint $\overrightarrow{C}$, to find the functional form of $\mathcal{P}\big(\overrightarrow{G}|\overrightarrow{\theta}\big)$ we are looking for. Equation (2.17) ensures that the maximization of $\mathcal{S}\big(\overrightarrow{\theta}\big)$, subject to (2.15), reduces to the maximization of each single-layer entropy $S^\alpha\big(\overrightarrow{\theta^\alpha}\big)$, subject to

$$\langle\overrightarrow{C^\alpha}\rangle_{\overrightarrow{\theta^\alpha}}=\sum_{G^\alpha\in\mathcal{G}_N} P^\alpha\big(G^\alpha|\overrightarrow{\theta^\alpha}\big)\overrightarrow{C^\alpha}(G^\alpha), \tag{2.20}$$

separately. Therefore the probability $\mathcal{P}\big(\overrightarrow{G}|\overrightarrow{\theta}\big)$ maximizing $\mathcal{S}\big(\overrightarrow{\theta}\big)$ reduces to the product of all single-layer probability distributions of the type $P^\alpha\big(G^\alpha|\overrightarrow{\theta^\alpha}\big)$, each of which should separately maximize the corresponding entropy $S^\alpha\big(\overrightarrow{\theta^\alpha}\big)$.

The general solution to the problem of maximizing $S^\alpha\big(\overrightarrow{\theta^\alpha}\big)$, subject to (2.20), for single-layer networks, leads in our notation to the probability distribution

$$P^\alpha\big(G^\alpha|\overrightarrow{\theta^\alpha}\big)=\frac{e^{-H^\alpha(G^\alpha|\overrightarrow{\theta^\alpha})}}{Z\big(\overrightarrow{\theta^\alpha}\big)}, \tag{2.21}$$

where

$$H^\alpha\big(G^\alpha|\overrightarrow{\theta^\alpha}\big)=\overrightarrow{\theta^\alpha}\cdot\overrightarrow{C^\alpha}(G^\alpha) \tag{2.22}$$

is the *graph Hamiltonian* (the dot indicating a scalar product, i.e. a linear combination of the enforced constraints) and

$$Z\big(\overrightarrow{\theta^\alpha}\big)=\sum_{G^\alpha\in\mathcal{G}_N} e^{-H^\alpha(G^\alpha|\overrightarrow{\theta^\alpha})} \tag{2.23}$$

is the *partition function* (representing the normalizing constant for the probability).

Equation (2.23), and consequently (2.21), leads to different explicit functional forms depending on the choice of the constraint(s), i.e. depending on the functional form of $\overrightarrow{C^\alpha}(G^\alpha)$. In the following Sections we explicitly discuss the cases of the Directed Binary Configuration Model (where the constraints are the in- and out-degrees of all nodes in each layer α) and of the Directed Weighted Configuration Model (where the constraints are the in- and out-strenghts of all nodes in

each layer α), respectively.

Once an explicit expression for each $P^{\alpha}\big(G^{\alpha}|\overrightarrow{\theta^{\alpha}}\big)$ is found, we can find the final expression for the whole multiplex probability in the null model:

$$\mathcal{P}\big(\overrightarrow{G}\,|\,\overrightarrow{\theta}\,\big) = \prod_{\alpha=1}^{M} \frac{e^{-H^{\alpha}(G^{\alpha}|\overrightarrow{\theta^{\alpha}})}}{Z\big(\overrightarrow{\theta^{\alpha}}\big)} = \frac{e^{-\mathcal{H}(\overrightarrow{G}\,|\,\overrightarrow{\theta}\,)}}{\mathcal{Z}\big(\overrightarrow{\theta}\big)}, \tag{2.24}$$

where

$$\mathcal{H}(\overrightarrow{G}\,|\,\overrightarrow{\theta}\,) \equiv \sum_{\alpha=1}^{M} H^{\alpha}(G^{\alpha}|\overrightarrow{\theta^{\alpha}}) \tag{2.25}$$

and

$$\mathcal{Z}\big(\overrightarrow{\theta}\big) \equiv \prod_{\alpha=1}^{M} Z\big(\overrightarrow{\theta^{\alpha}}\big). \tag{2.26}$$

The last three equations rephrase the independence of all layers explicitly.

2.B Maximum-likelihood method for multiplex networks

The maximization of the entropy is a constrained, *functional* maximization of $\mathcal{S}\big(\overrightarrow{\theta}\big)$ in the space of probability distributions. As such, its result is the *functional form* of the maximum-entropy distribution $\mathcal{P}\big(\overrightarrow{G}\,|\,\overrightarrow{\theta}\,\big)$, given by (2.24), but not its *numerical values*. In fact, the distribution depends on the whole vector of parameters $\overrightarrow{\theta}$, and any expectation value calculated analytically using the explicit expression of $\mathcal{P}\big(\overrightarrow{G}\,|\,\overrightarrow{\theta}\,\big)$ can only be evaluated numerically after a value of $\overrightarrow{\theta}$ is specified. This leads to the problem of choosing $\overrightarrow{\theta}$. Since we are interested in the case where all layers of the multiplex are independent, choosing a value of $\overrightarrow{\theta}$ reduces to the problem of choosing $\overrightarrow{\theta^{\alpha}}$ separately for each layer.

The problem of finding the parameter values of a maximum-entropy model of single-layer networks has been solved in the general case using the maximum likelihood principle. In our notation here, this solution can be restated as follows. Let G^{α}_{*} denote, among all graphs $G^{\alpha} \in \mathcal{G}_N$, the particular *empirical* network realized in layer α of the multiplex. Given G^{α}_{*}, the log-likelihood function

$$\mathcal{L}^{\alpha}\big(\overrightarrow{\theta^{\alpha}}\big) \equiv \ln P\big(G^{\alpha}_{*}|\overrightarrow{\theta^{\alpha}}\big) \tag{2.27}$$

represents the log of the probability to generate the empirical graph G^{α}_{*}, given a value of $\overrightarrow{\theta^{\alpha}}$. The maximum likelihood principle states that the optimal choice for

$\overrightarrow{\theta^\alpha}$ is the one that maximizes the chances to obtain G^α_* from the model, i.e. the one that maximizes $\mathcal{L}^\alpha\big(\overrightarrow{\theta^\alpha}\big)$. Let this parameter choice be denoted by $\overrightarrow{\theta^\alpha_*}$, where

$$\overrightarrow{\theta^\alpha_*} \equiv \arg\max_{\overrightarrow{\theta^\alpha}} \mathcal{L}^\alpha\big(\overrightarrow{\theta^\alpha}\big). \tag{2.28}$$

As a general result, the value $\overrightarrow{\theta^\alpha_*}$ defined above is such that

$$\langle \overrightarrow{C^\alpha} \rangle_{\overrightarrow{\theta^\alpha_*}} = \overrightarrow{C^\alpha}(G^\alpha_*), \tag{2.29}$$

i.e. the expectation value of each constraint coincides with the empirical value measured on the empirical network G^α_*. This is precisely the outcome we desire, given that our ultimate goal is the construction of ensembles of networks with the same numerical value of the constraints as in the real network.

From a practical point of view, eqs. (2.28) and (2.29) represent two equivalent ways to determine $\overrightarrow{\theta^\alpha_*}$. The former requires the maximization of a scalar function over a K^α-dimensional space, while the latter requires the solution of a system of K^α nonlinear coupled equations. For various choices of the graph ensemble $\mathcal{G}_N$ and of the constraints $\overrightarrow{C^\alpha}$ (including those required for our analysis), both approaches are implemented in the MAX&SAM algorithm (see references in the main text of this Chapter). More details are given in Sections 2.C and 2.D. Once the value $\overrightarrow{\theta^\alpha_*}$ is found, it is used to find the *numerical value* $P\big(G^\alpha|\overrightarrow{\theta^\alpha_*}\big)$ of the probability of any graph $G^\alpha \in \mathcal{G}_N$. So, while the maximization of the entropy generates the functional form of the graph probability, the maximization of the likelihood fixes its numerical values. If X^α denotes any single-layer structural property X of interest, the above procedure allows us to evaluate the expected value

$$\langle X^\alpha \rangle \equiv \langle X^\alpha \rangle_{\overrightarrow{\theta^\alpha_*}} = \sum_{G^\alpha \in \mathcal{G}_N} P\big(G^\alpha|\overrightarrow{\theta^\alpha_*}\big) X^\alpha(G^\alpha) \tag{2.30}$$

(and similarly the standard deviation) of X^α explicitly over the desired ensemble. For many properties of interest, the expected value (2.30) can be calculated analytically given the explicit expression of $P\big(G^\alpha|\overrightarrow{\theta^\alpha_*}\big)$, without the need to sample the graph ensemble explicitly. For more complicated properties, one can instead use the knowledge of $P\big(G^\alpha|\overrightarrow{\theta^\alpha_*}\big)$ to sample graphs from the ensemble in an unbiased way and then calculate expectations as sample averages.

The multiplexity and multireciprocity metrics introduced in the main text are not single-layer properties like X^α, as they require measurements on multiple layers simultaneously. We therefore need to generalize eq. (2.30) to the case of an arbitrary multiplex quantity $\mathcal{X}$, evaluating to $\mathcal{X}(\overrightarrow{G})$ on a specific multiplex

$\overrightarrow{G} \in \mathcal{G}_N^M$, as follows:

$$\langle \mathcal{X} \rangle \equiv \langle \mathcal{X} \rangle_{\overrightarrow{\theta_*}} = \sum_{\overrightarrow{G} \in \mathcal{G}_N^M} \mathcal{P}\big(\overrightarrow{G} | \overrightarrow{\theta_*}\big) \mathcal{X}(\overrightarrow{G}) \tag{2.31}$$

where $\overrightarrow{\theta_*} = (\overrightarrow{\theta_*^\alpha})_{\alpha=1}^M$ contains the Langrange multipliers (2.28) for all layers and $\overrightarrow{G_*} = (G_*^\alpha)_{\alpha=1}^M \in \mathcal{G}_N^M$ denotes the whole empirical multiplex. Both the expected values and the standard deviations of multiplexity and multireciprocity can be calculated explicitly, and we will therefore follow the analytical approach, which is exact and faster than the sampling approach (see Sections 2.C and 2.D).

From a computational point of view, the above canonical approach based on soft constraints has many benefits with respect to the microcanonical approach with hard constraints. Indeed, the microcanonical approach cannot be controlled analytically, and necessarily requires sampling many randomized multiplexes explicitly from the ensemble. Generating even only a single randomized multiplex requires the iteration of many random constraint-preserving 'rewiring moves', which is computationally costly. Such a procedure must be repeated several times, to produce a large sample of R randomized multiplexes, on each of which any topological property X of interest has to be calculated. Finally, a sample average should be performed to obtain an estimate of $\langle X \rangle$.

For instance, on single-layer networks with constrained degree sequence one should iterate the so-called 'local rewiring algorithm' that preserves the degrees while randomizing the network. On a monoplex network with L links, the above approach would require a computational time of order $O(L)$, only to generate a single realization of the randomized network. On such a realization, one would then need to measure X (for instance the monoplex reciprocity), which would require a certain time T_X. The total time needed for a single realization would therefore be $T_X + O(L)$, and for all realizations $R \cdot T_X + O(R \cdot L)$.

In a multiplex network with M layers, the corresponding time required to generate a single randomized multiplex would in principle be of order $O(\sum_{\alpha=1}^M L^\alpha)$, where L^α is the number of links in the α-th layer. However, if layers are independent in the null model, the randomization could (if computational resource allows) be run in parallel on the different layers, thus reducing the above time to $O(\bar{L})$ where $\bar{L}$ is the average number of links per layer, which does not scale with M. However, the calculation of multiplex quantities X (e.g. the multireciprocity) which would require a time T_X for a single layer (e.g. the monoplex reciprocity) would now need to be iterated for each pair of layers, thus requiring a time $O(T_X \cdot M^2)$. In total, this means that the total microcanonical computational time for a multiplex is $T_{\text{mic}} = O(R \cdot T_X \cdot M^2) + O(R \cdot L)$, before carrying out the final sample averages.

By contrast, our canonical approach does not require the sampling of any multiplex. For individual layers, the calculation of the expected value of most properties of interest basically requires replacing the adjacency matrix of the network with the corresponding expected matrix (or more complicated replacements that in any case require a comparable calculation time). Therefore calculating the expected value $\langle X \rangle$ takes the same time T_X that it would take for the empirical property X to be calculated on the real system. The same holds true for the entire multiplex. Therefore the total canonical time needed is $T_{\rm can} = O(T_X \cdot M^2) + T_{\mathcal{L}}$, where $T_{\mathcal{L}}$ is the one-off time required to preliminary maximize the likelihood (possibly of each layer in parallel) defined in (2.27).

As already mentioned above, the time $T_{\mathcal{L}}$ required to maximize the likelihood function can be proxied by the time required to solve a system of coupled, nonlinear equations ($2N$ equations in the case of directed networks, as shown below). However, since such systems can be further simplified by rewriting them only in terms of the sequences of distinct directed degrees/strengths (which are always less than $2N$), the computational time drops to the order of seconds or minutes (depending on the chosen constraints) for each layer. Moreover, further analyses on synthetic networks have shown that this time scales roughly quadratically with the number of nodes; this is anyway considerably shorter than the corresponding total microcanonical time $T_{\rm mic}$ estimated above.

Besides the computational advantages described above, the canonical approach has the statistical advantage of being a truly unbiased method, in the sense that its maximum-entropy nature implies that no preference is given to specific graph configurations, other than on the basis of the enforced constraints. So unbiasedness is ensured by the maximum degree of randomness encoded in the graph probability, given the constraints. By constrast, microcanonical approaches are not guaranteed to ensure the same property. In the microcanonical case, unbiasedness means that the realizations of the network should be sampled uniformly (i.e. with exactly the same probability) from the whole set of configurations compatible with the constraints. Ensuring uniform sampling is highly nontrivial and often impossible. For instance, in the case of graphs with fixed degree sequence, it can be proved that the local rewiring algorithm is biased, as it preferentially samples configurations that are 'close' to the empirical one. Previous studies showed that it is in principle possible to remove this bias, by calculating the so-called 'mobility' function (which is a quantity that depends on the current configuration being randomized) and accepting the 'next' randomized configurations with a probability that depends on the mobility itself. This requirement further increases, and by a large extent, the already heavy computational requirements of the microcanonical approach, because the mobility should be continuously recalculated during the randomization process.

2.C Directed Binary Configuration Model

In this Section we explicitly discuss the DBCM model, obtained through the maximum entropy and maximum likelihood methods in the specific case where $\mathcal{G}_N$ contains all binary directed graphs with N nodes and $\overrightarrow{C^\alpha}$ is a vector of dimension $K^\alpha = 2N$ containing the out-degree k_i^{out} and the in-degree k_i^{in} of all nodes ($i = 1, N$). Correspondingly, the $2N$-dimensional vector $\overrightarrow{\theta^\alpha}$ contains the associated Lagrange multipliers ϕ_i^α and χ_i^α for all nodes. Note that we enforce the in- and out-degree sequences on all layers, which means that, as a function, $\overrightarrow{C^\alpha} = (\overrightarrow{k^{out}}, \overrightarrow{k^{in}})$ is the same for all α. However, the numerical values of the degrees in different layers will in general be different, i.e. $\overrightarrow{C}(G^\alpha) \neq \overrightarrow{C}(G^\beta)$ for $\alpha \neq \beta$, thus $\overrightarrow{\theta^\alpha} = (\overrightarrow{\phi^\alpha}, \overrightarrow{\chi^\alpha})$ must still depend on α explictly.

For single-layer networks, this model has been fully discussed. Here we simply summarize the main steps leading to the final expressions for the expected binary multiplexity and binary reciprocity. Using the notation introduced in the main text and in Appendix 2.A, the single-layer Hamiltonian (2.22) reads

$$\begin{aligned} H\big(G^\alpha|\overrightarrow{\phi^\alpha}, \overrightarrow{\chi^\alpha}\big) &= \overrightarrow{\phi^\alpha} \cdot \overrightarrow{k^{out}}(G^\alpha) + \overrightarrow{\chi^\alpha} \cdot \overrightarrow{k^{in}}(G^\alpha) = \\ &= \sum_{i=1}^{N} \big[\phi_i^\alpha k_i^{out}(G^\alpha) + \chi_i^\alpha k_i^{in}(G^\alpha)\big] = \\ &= \sum_{i=1}^{N} \sum_{j\neq i} \big(\phi_i^\alpha + \chi_j^\alpha\big) a_{ij}^\alpha \end{aligned} \tag{2.32}$$

and the partition function (2.23) can be calculated as:

$$\begin{aligned} Z\big(\overrightarrow{\phi^\alpha}, \overrightarrow{\chi^\alpha}\big) &= \prod_{i=1}^{N} \prod_{j\neq i} \big(1 + e^{-\phi_i^\alpha - \chi_j^\alpha}\big) \\ &= \prod_{i=1}^{N} \prod_{j\neq i} \big(1 + x_i^\alpha y_j^\alpha\big), \end{aligned} \tag{2.33}$$

where we have set $x_i^\alpha \equiv e^{-\phi_i^\alpha}$ and $y_i^\alpha \equiv e^{-\chi_i^\alpha}$. This implies that the probability (2.21) can be written explicitly as

$$\begin{aligned} P^\alpha\big(G^\alpha|\overrightarrow{\phi^\alpha}, \overrightarrow{\chi^\alpha}\big) &= \prod_{i=1}^{N} \prod_{j\neq i} \frac{(x_i^\alpha y_j^\alpha)^{a_{ij}^\alpha}}{1 + x_i^\alpha y_j^\alpha} \\ &= \prod_{i=1}^{N} \prod_{j\neq i} (p_{ij}^\alpha)^{a_{ij}^\alpha} (1 - p_{ij}^\alpha)^{1 - a_{ij}^\alpha}, \end{aligned} \tag{2.34}$$

where

$$p_{ij}^{\alpha} = \frac{x_i^{\alpha} y_j^{\alpha}}{1 + x_i^{\alpha} y_j^{\alpha}} \tag{2.35}$$

is the probability of a directed link from i to j in layer α. Equation (2.34) shows that the random variable a_{ij}^{α} is drawn, for all $i \neq j$, from a Bernoulli distribution with success probability p_{ij}^{α}.

Given the real-world multiplex $\overrightarrow{G}_* = (G_*^{\alpha})_{\alpha=1}^{M}$, the single-layer log-likelihood function (2.27) to be maximized is then given by

$$\begin{aligned} \mathcal{L}\big(\overrightarrow{x^{\alpha}}, \overrightarrow{y^{\alpha}}\big) &= \sum_{i=1}^{N} \big[k_i^{out}(G_*^{\alpha}) \ln x_i^{\alpha} + k_i^{in}(G_*^{\alpha}) \ln y_i^{\alpha}\big] + \\ &\quad - \sum_{i=1}^{N} \sum_{j \neq i} \ln\big(1 + x_i^{\alpha} y_j^{\alpha}\big), \end{aligned} \tag{2.36}$$

and the equivalent set of $2N$ coupled nonlinear equations (2.29) to be solved is

$$\sum_{j \neq i} \frac{x_i^{\alpha} y_j^{\alpha}}{1 + x_i^{\alpha} y_j^{\alpha}} = k_i^{out}(G_*^{\alpha}) \qquad \forall i = 1, N \tag{2.37}$$

$$\sum_{j \neq i} \frac{x_j^{\alpha} y_i^{\alpha}}{1 + x_j^{\alpha} y_i^{\alpha}} = k_i^{in}(G_*^{\alpha}) \qquad \forall i = 1, N. \tag{2.38}$$

Once found, the values of $\{x_i^{\alpha}\}$ and $\{y_i^{\alpha}\}$ providing the unique solution to the above problem can be put back in eqs. (2.34) and (2.35), allowing us to analytically calculate the expected values $\langle\cdot\rangle_{\text{DBCM}}$ of the quantities of interest via the corresponding probabilities p_{ij}^{α} (where for simplicity we drop the asterisk indicating that p_{ij}^{α} is evaluated at the specific values that maximize the likelihood).

In particular, we can calculate the rescaled metrics of multiplexity and multireciprocity defined in the main text as follows. First of all, since in the DBCM the in- and out-degrees of all nodes in all layers are equal to their expected values, we necessarily have $\langle L^{\alpha}\rangle_{\text{DBCM}} = L_*^{\alpha}$ for all α, where $L_*^{\alpha} \equiv L^{\alpha}(G_*^{\alpha})$ is the number of links of the observed, layer-specific graph G_*^{α}. This means that L^{α} is a constrained quantity, and we therefore expect the denominators of the aforementioned quantities to fluctuate around their expected values $L_*^{\alpha} + L_*^{\beta}$ much less than how the numerators fluctuate around the corresponding expected values. We therefore approximate their expected values as follows:

$$\langle m_b^{\alpha,\beta}\rangle_{\text{DBCM}} = \frac{2\langle L^{\alpha \rightrightarrows \beta}\rangle_{\text{DBCM}}}{L_*^{\alpha} + L_*^{\beta}} \qquad (\alpha \neq \beta), \tag{2.39a}$$

$$\langle r_b^{\alpha,\beta}\rangle_{\text{DBCM}} = \frac{2\langle L^{\alpha \rightleftarrows \beta}\rangle_{\text{DBCM}}}{L_*^{\alpha} + L_*^{\beta}}. \tag{2.39b}$$

Consequently,

$$\mu_b^{\alpha,\beta} = \frac{2L_*^{\alpha\rightrightarrows\beta} - 2\langle L^{\alpha\rightrightarrows\beta}\rangle_{\mathrm{DBCM}}}{L_*^{\alpha} + L_*^{\beta} - 2\langle L^{\alpha\rightrightarrows\beta}\rangle_{\mathrm{DBCM}}} \qquad (\alpha \neq \beta),$$
$$\rho_b^{\alpha,\beta} = \frac{2L_*^{\alpha\rightleftarrows\beta} - 2\langle L^{\alpha\rightleftarrows\beta}\rangle_{\mathrm{DBCM}}}{L_*^{\alpha} + L_*^{\beta} - 2\langle L^{\alpha\rightleftarrows\beta}\rangle_{\mathrm{DBCM}}}.$$

Since a_{ij}^{α} and a_{ij}^{β} (for $\beta \neq \alpha$), and similarly a_{ij}^{α} and a_{ji}^{β} (for any β), are independently drawn from two Bernoulli distributions, the expected values of $\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}$ (with $\beta \neq \alpha$) and $\min\{a_{ij}^{\alpha}, a_{ji}^{\beta}\}$ are easily calculated as

$$\langle \min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}\rangle_{\mathrm{DBCM}} = p_{ij}^{\alpha} p_{ij}^{\beta} \qquad (\alpha \neq \beta), \tag{2.40a}$$
$$\langle \min\{a_{ij}^{\alpha}, a_{ji}^{\beta}\}\rangle_{\mathrm{DBCM}} = p_{ij}^{\alpha} p_{ji}^{\beta}, \tag{2.40b}$$

as shown in Chapter 1 for the undirected case. Therefore the final expressions for the transformed multiplexity and multireciprocity are:

$$\mu_b^{\alpha,\beta} = \frac{2L_*^{\alpha\rightrightarrows\beta} - 2\sum_i \sum_{j\neq i} p_{ij}^{\alpha} p_{ij}^{\beta}}{L_*^{\alpha} + L_*^{\beta} - 2\sum_i \sum_{j\neq i} p_{ij}^{\alpha} p_{ij}^{\beta}} \quad (\alpha \neq \beta) \tag{2.41a}$$
$$\rho_b^{\alpha,\beta} = \frac{2L_*^{\alpha\rightleftarrows\beta} - 2\sum_i \sum_{j\neq i} p_{ij}^{\alpha} p_{ji}^{\beta}}{L_*^{\alpha} + L_*^{\beta} - 2\sum_i \sum_{j\neq i} p_{ij}^{\alpha} p_{ji}^{\beta}}, \tag{2.41b}$$

where the probabilities are defined according to Eq. (2.35).

Similarly, we need to calculate the z-scores associated to our metrics. To do this, we need to calculate the standard deviations of $m_b^{\alpha\beta}$ and $r_b^{\alpha\beta}$ at the denominator of the z-scores. Neglecting again the fluctuations of the constrained quantities L^{α} and L^{β} around their average values (with respect to the fluctuations of the unconstrained quantities), and since all pairs of nodes are independent, we calculate the variances of $m_b^{\alpha\beta}$ and $r_b^{\alpha\beta}$ in a way similar to what we did for the expressions in eq. (2.42):

$$\mathrm{Var}[m_b^{\alpha\beta}] = \frac{4\sum_i \sum_{j\neq i} \mathrm{Var}[\min\{a_{ij}^{\alpha}, a_{ij}^{\beta}\}]}{(L_*^{\alpha} + L_*^{\beta})^2} \qquad (\alpha \neq \beta),$$
$$\mathrm{Var}[r_b^{\alpha\beta}] = \frac{4\sum_i \sum_{j\neq i} \mathrm{Var}[\min\{a_{ij}^{\alpha}, a_{ji}^{\beta}\}]}{(L_*^{\alpha} + L_*^{\beta})^2}.$$

Now we note that the minimum of two 0/1 quantities is also a 0/1 quantity. This implies that the square of the minimum is equal to the minimum itself, and that the expected square of the minimum is equal to the expected value of the

minimum. In formulas:

$$\langle \min^2\{a_{ij}^\alpha, a_{ij}^\beta\}\rangle_{\rm DBCM} = p_{ij}^\alpha p_{ij}^\beta \qquad (\alpha \neq \beta), \tag{2.43}$$

$$\langle \min^2\{a_{ij}^\alpha, a_{ji}^\beta\}\rangle_{\rm DBCM} = p_{ij}^\alpha p_{ji}^\beta. \tag{2.44}$$

It then follows that the variance of the minimum is

$$\begin{aligned} {\rm Var}\big[\min\{a_{ij}^\alpha, a_{ij}^\beta\}\big] &= p_{ij}^\alpha p_{ij}^\beta (1 - p_{ij}^\alpha p_{ij}^\beta) \qquad (\alpha \neq \beta), \\ {\rm Var}\big[\min\{a_{ij}^\alpha, a_{ji}^\beta\}\big] &= p_{ij}^\alpha p_{ji}^\beta (1 - p_{ij}^\alpha p_{ji}^\beta). \end{aligned}$$

Putting these expressions into those for ${\rm Var}[m_b^{\alpha,\beta}]$ and ${\rm Var}[r_b^{\alpha,\beta}]$, and taking the square root to obtain the standard deviations, we finally arrive at the explicit calculation of the z-scores:

$$\begin{aligned} z\big(m_b^{\alpha\beta}\big) &= \frac{L_*^{\alpha\rightleftarrows\beta} - \sum_i \sum_{j\neq i} p_{ij}^\alpha p_{ij}^\beta}{\sqrt{\sum_i \sum_{j\neq i} p_{ij}^\alpha p_{ij}^\beta (1 - p_{ij}^\alpha p_{ij}^\beta)}} \quad (\alpha \neq \beta) \\ z\big(r_b^{\alpha\beta}\big) &= \frac{L_*^{\alpha\rightleftarrows\beta} - \sum_i \sum_{j\neq i} p_{ij}^\alpha p_{ji}^\beta}{\sqrt{\sum_i \sum_{j\neq i} p_{ij}^\alpha p_{ji}^\beta (1 - p_{ij}^\alpha p_{ji}^\beta)}} \end{aligned}$$

From a direct comparison between the above equations and Eqs. (2.41), we immediately observe the sign concordance reported in Chapters 1 and 2.

2.D Directed Weighted Configuration Model

Here we consider the DWCM model, obtained when $\mathcal{G}_N$ contains all weighted directed graphs (with non-negative integer edge weights) with N nodes and $\overrightarrow{C^\alpha}$ is a vector of dimension $K^\alpha = 2N$ containing the out-strength s_i^{out} and the in-strength s_i^{in} of all nodes ($i = 1, N$). The $2N$-dimensional vector $\overrightarrow{\theta^\alpha}$ contains the associated Lagrange multipliers ϕ_i^α and χ_i^α. As for the DBCM, $\overrightarrow{C^\alpha} = (\overrightarrow{s^{out}}, \overrightarrow{s^{in}})$ is the same function for all α. However, the numerical values $\overrightarrow{\theta^\alpha} = (\overrightarrow{\phi^\alpha}, \overrightarrow{\chi^\alpha})$ still depend on α.

For single-layer networks, the Hamiltonian (2.22) reads

$$\begin{aligned} H\big(G^\alpha|\overrightarrow{\phi^\alpha}, \overrightarrow{\chi^\alpha}\big) &= \overrightarrow{\phi^\alpha} \cdot \overrightarrow{s^{out}}(G^\alpha) + \overrightarrow{\chi^\alpha} \cdot \overrightarrow{s^{in}}(G^\alpha) = \\ &= \sum_{i=1}^N \big[\phi_i^\alpha s_i^{out}(G^\alpha) + \chi_i^\alpha s_i^{in}(G^\alpha)\big] = \\ &= \sum_{i=1}^N \sum_{j\neq i} \big(\phi_i^\alpha + \chi_j^\alpha\big) w_{ij}^\alpha \end{aligned} \tag{2.45}$$

and the partition function (2.23) can be calculated as:

$$\begin{aligned} Z(\overrightarrow{\phi^\alpha}, \overrightarrow{\chi^\alpha}) &= \prod_{i=1}^{N} \prod_{j \neq i} \left(1 - e^{-\phi_i^\alpha - \chi_j^\alpha}\right)^{-1} \\ &= \prod_{i=1}^{N} \prod_{j \neq i} \left(1 - x_i^\alpha y_j^\alpha\right)^{-1}, \end{aligned} \tag{2.46}$$

where we have set $x_i^\alpha \equiv e^{-\phi_i^\alpha}$ and $y_i^\alpha \equiv e^{-\chi_i^\alpha}$. This implies that the probability (2.21) can be written as

$$\begin{aligned} P^\alpha\left(G^\alpha | \overrightarrow{\phi^\alpha}, \overrightarrow{\chi^\alpha}\right) &= \prod_{i=1}^{N} \prod_{j \neq i} (x_i^\alpha y_j^\alpha)^{w_{ij}^\alpha} (1 - x_i^\alpha y_j^\alpha) \\ &= \prod_{i=1}^{N} \prod_{j \neq i} (p_{ij}^\alpha)^{w_{ij}^\alpha} (1 - p_{ij}^\alpha), \end{aligned} \tag{2.47}$$

where

$$p_{ij}^\alpha = x_i^\alpha y_j^\alpha \tag{2.48}$$

denotes again the probability that a directed link (of any positive weight) from node i to node j is realized in layer α. Equation (2.47) gives the interpretation of w_{ij}^α as a geometrically distributed variable, constructed as the iteration of many random events, each defined as incrementing w_{ij}^α by one, starting from $w_{ij}^\alpha = 0$. In this interpretation, p_{ij}^α is the elementary probability of a 'success' event, and the probability that $w_{ij}^\alpha = w$ coincides with the probability $(p_{ij}^\alpha)^w (1 - p_{ij}^\alpha)$ of having w consecutive successes followed by one failure. This leads precisely to a geometric distribution.

The single-layer log-likelihood function (2.27) to be maximized is now given by

$$\begin{aligned} \mathcal{L}(\overrightarrow{x^\alpha}, \overrightarrow{y^\alpha}) &= \sum_{i=1}^{N} \left[s_i^{out}(G_*^\alpha) \ln x_i^\alpha + s_i^{in}(G_*^\alpha) \ln y_i^\alpha \right] + \\ &\quad + \sum_{i=1}^{N} \sum_{j \neq i} \ln\left(1 - x_i^\alpha y_j^\alpha\right), \end{aligned} \tag{2.49}$$

and the corresponding equations (2.29) are

$$\sum_{j \neq i} \frac{x_i^\alpha y_j^\alpha}{1 - x_i^\alpha y_j^\alpha} = s_i^{out}(G_*^\alpha) \qquad \forall i = 1, N \tag{2.50}$$

$$\sum_{j \neq i} \frac{x_j^\alpha y_i^\alpha}{1 - x_j^\alpha y_i^\alpha} = s_i^{in}(G_*^\alpha) \qquad \forall i = 1, N. \tag{2.51}$$

The expected values $\langle\cdot\rangle_{\rm DWCM}$ of the relevant quantities can be found through eqs. (2.47) and (2.48), evaluated at the values of $\{x_i^\alpha\}$ and $\{y_i^\alpha\}$ that solve the above problem (again, in what follows we drop the asterisk indicating that p_{ij}^α is evaluated at the specific values that maximize the likelihood).

We start with the calculation of the expected values of the multiplexity and multireciprocity metrics defined in the main text. In analogy with what we did for the DBCM, we expect the (constrained) denominators of both the metrics to fluctuate much less than the (unconstrained) numerators and we therefore replace the denominators with their expected values $W_*^\alpha + W_*^\beta$. We therefore write

$$\langle m_w^{\alpha\beta}\rangle_{\rm DWCM} = \frac{2\langle W^{\alpha\rightrightarrows\beta}\rangle_{\rm DWCM}}{W_*^\alpha + W_*^\beta}, \tag{2.52a}$$

$$\langle r_w^{\alpha\beta}\rangle_{\rm DWCM} = \frac{2\langle W^{\alpha\rightleftarrows\beta}\rangle_{\rm DWCM}}{W_*^\alpha + W_*^\beta} \tag{2.52b}$$

and

$$\begin{aligned}\mu_w^{\alpha\beta} &= \frac{2W_*^{\alpha\rightrightarrows\beta} - 2\langle W^{\alpha\rightrightarrows\beta}\rangle_{\rm DWCM}}{W_*^\alpha + W_*^\beta - 2\langle W^{\alpha\rightrightarrows\beta}\rangle_{\rm DWCM}} \qquad (\alpha\neq\beta),\\ \rho_w^{\alpha\beta} &= \frac{2W_*^{\alpha\rightleftarrows\beta} - 2\langle W^{\alpha\rightleftarrows\beta}\rangle_{\rm DWCM}}{W_*^\alpha + W_*^\beta - 2\langle W^{\alpha\rightleftarrows\beta}\rangle_{\rm DWCM}}.\end{aligned}$$

Since w_{ij}^α and w_{ij}^β (for $\beta\neq\alpha$), and similarly w_{ij}^α and w_{ji}^β (for any β), are independenlty drawn from two geometric distributions, the expected values of $\min\{w_{ij}^\alpha, w_{ij}^\beta\}$ (with $\beta\neq\alpha$) and $\min\{w_{ij}^\alpha, w_{ji}^\beta\}$ are easily calculated as

$$\langle\min\{a_{ij}^\alpha, a_{ij}^\beta\}\rangle_{\rm DBCM} = \frac{p_{ij}^\alpha p_{ij}^\beta}{1-p_{ij}^\alpha p_{ij}^\beta} \qquad (\alpha\neq\beta), \tag{2.53a}$$

$$\langle\min\{a_{ij}^\alpha, a_{ji}^\beta\}\rangle_{\rm DBCM} = \frac{p_{ij}^\alpha p_{ji}^\beta}{1-p_{ij}^\alpha p_{ji}^\beta}. \tag{2.53b}$$

Therefore the transformed multiplexity and multireciprocity read

$$\mu_w^{\alpha\beta} = \frac{2W_*^{\alpha\rightrightarrows\beta} - 2\sum_i\sum_{j\neq i}\frac{p_{ij}^\alpha p_{ij}^\beta}{1-p_{ij}^\alpha p_{ij}^\beta}}{W_*^\alpha + W_*^\beta - 2\sum_i\sum_{j\neq i}\frac{p_{ij}^\alpha p_{ij}^\beta}{1-p_{ij}^\alpha p_{ij}^\beta}} \quad (\alpha\neq\beta) \tag{2.54a}$$

$$\rho_w^{\alpha\beta} = \frac{2W_*^{\alpha\rightleftarrows\beta} - 2\sum_i\sum_{j\neq i}\frac{p_{ij}^\alpha p_{ji}^\beta}{1-p_{ij}^\alpha p_{ji}^\beta}}{W_*^\alpha + W_*^\beta - 2\sum_i\sum_{j\neq i}\frac{p_{ij}^\alpha p_{ji}^\beta}{1-p_{ij}^\alpha p_{ji}^\beta}}, \tag{2.54b}$$

where the probabilities are defined in Eq. (2.48).

We then calculate the z-scores. Following an argument similar to the binary case, we write

$$\mathrm{Var}[m_w^{\alpha\beta}] = \frac{4\sum_i \sum_{j\neq i} \mathrm{Var}[\min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}]}{(W_*^{\alpha} + W_*^{\beta})^2} \qquad (\alpha \neq \beta),$$
$$\mathrm{Var}[r_w^{\alpha\beta}] = \frac{4\sum_i \sum_{j\neq i} \mathrm{Var}[\min\{w_{ij}^{\alpha}, w_{ji}^{\beta}\}]}{(W_*^{\alpha} + W_*^{\beta})^2}.$$

After calculating the variance of the minimum of two geometrically distributed random variables, we get

$$\mathrm{Var}\big[\min\{w_{ij}^{\alpha}, w_{ij}^{\beta}\}\big] = \frac{p_{ij}^{\alpha} p_{ij}^{\beta}}{\big(1 - p_{ij}^{\alpha} p_{ij}^{\beta}\big)^2} \qquad (\alpha \neq \beta),$$
$$\mathrm{Var}\big[\min\{a_{ij}^{\alpha}, a_{ji}^{\beta}\}\big] = \frac{p_{ij}^{\alpha} p_{ji}^{\beta}}{\big(1 - p_{ij}^{\alpha} p_{ji}^{\beta}\big)^2}.$$

Combining all the relevant expressions together, we get for the z-scores:

$$z\big(m_w^{\alpha\beta}\big) = \frac{W_*^{\alpha\rightleftarrows\beta} - \sum_i \sum_{j\neq i} \frac{p_{ij}^{\alpha} p_{ij}^{\beta}}{1 - p_{ij}^{\alpha} p_{ij}^{\beta}}}{\sqrt{\sum_i \sum_{j\neq i} \frac{p_{ij}^{\alpha} p_{ij}^{\beta}}{\big(1 - p_{ij}^{\alpha} p_{ij}^{\beta}\big)^2}}} \qquad (\alpha \neq \beta)$$
$$z\big(r_w^{\alpha\beta}\big) = \frac{W_*^{\alpha\rightleftarrows\beta} - \sum_i \sum_{j\neq i} \frac{p_{ij}^{\alpha} p_{ji}^{\beta}}{1 - p_{ij}^{\alpha} p_{ji}^{\beta}}}{\sqrt{\sum_i \sum_{j\neq i} \frac{p_{ij}^{\alpha} p_{ji}^{\beta}}{\big(1 - p_{ij}^{\alpha} p_{ji}^{\beta}\big)^2}}}$$

in analogy with the results shown in the first chapter for the undirected case. Comparing with Eqs. (2.54), we confirm the concordance of the sings.

Bibliography

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

[2] M. E. J. Newman (2004) 'Analysis of weighted networks', *Physical Review E* **70** (5), 056131

[3] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, M. Zanin (2014) 'The structure and dynamics of multilayer networks', *Physics Reports* **544** (1), 1

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

[5] M. Magnani, B. Micenková, L. Rossi (2013) 'Combinatorial analysis of multiple networks', *arXiv*:1303.4986

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

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

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

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

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

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

[12] L. A. Meyers, M. E. J. Newman, B. Pourbohloul (2006) 'Predicting epidemics on directed contact networks', *Journal of Theoretical Biology* **240** (3), 400

[13] M. Boguñá, M. Serrano (2005) 'Generalized percolation in random directed networks', *Physical Review E* **72** (1), 016106

[14] M. Schnegg (2006) 'Reciprocity and the emergence of power laws in social networks', *International Journal of Modern Phtsics C* **17**, 1067

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

[16] C. Zhou, L. Zemanová, G. Zamora, C. C. Hilgetag, J. Kurths (2006) 'Hierarchical organization unveiled by functional connectivity in complex brain networks', *Physical Review Letters* **97** (23), 238103

[17] M. E. J. Newman, S. Forrest, J. Balthrop (2002) 'Email networks and the spread of computer viruses', *Physical Review E* **66** (3), 035101

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

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

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

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

[22] E. S. Roberts, A. C. C. Coolen (2012) 'Unbiased degree-preserving randomization of directed binary networks', *Physical Review E* **85** (4), 046103

[23] M. E. J. Newman, S. H. Strogatz, D. J. Watts (2001) 'Random graphs with arbitrary degree distributions and their applications', *Physical Review E* **64** (2), 026118

[24] G. L. Robins, P. E. Pattison, Y. Kalish, D. Lusher (2007) 'An introduction to exponential random (p^*) models for social networks', *Social Networks* **29** (2), 173

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

[26] J. Park, M. E. J. Newman (2003) 'Origin of degree correlations in the Internet and other networks', *Physical Review E* **68** (2), 026112

[27] P. W. Holland, S. Leinhardt (1981) 'An exponential family of probability distributions for directed graphs', *Journal of the American Statistical Association* **76** (373), 33

[28] S. Wasserman, K. Faust (1994) 'Social network analysis', Cambridge University Press, Cambridge, New York

[29] T. A. B. Snijders, P. E. Pattison, G. L. Robins, M. S. Handcock (2006) 'New specifications for exponential random graph models', *Sociological Methodology* **36** (1), 99

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

[31] D. Garlaschelli, M. I. Loffredo (2004) 'Fitness-dependent topological properties of the World Trade Web', *Physical Review Letters* **93** (18), 188701

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

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

[34] 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

[35] C. A. Hidalgo, B. Klinger, A.-L. Barabási, R. Hausmann (2007) 'The product space conditions the development of nations', *Science* **317**, 482

[36] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli, L. Pietronero (2012) 'A new metrics for countries' fitness and products' complexity', *Scientific Reports* **2**, 723

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

[38] http://www. wcoomd.org

[39] F. Ruzzenenti, D. Garlaschelli, R. Basosi (2010) 'Complex networks and symmetry II: reciprocity and evolution of world trade', *Symmetry* **2** (3), 1710

[40] F. Picciolo, T. Squartini, F. Ruzzenenti, R. Basosi, D. Garlaschelli (2012) 'The role of distances in the World Trade Web', *Proceedings of the 8th International Conference on Signal Image Technology and Internet Based Systems (SITIS)*, 784

[41] M. de Domenico, V. Nicosia, A. Arenas, V. Latora (2015) 'Structural reducibility of multilayer networks', *Nature Communications* **6**, 6864

[42] J. Iacovacci, Z. Wu, G. Bianconi (2015) 'Mesoscopic structures reveal the network between the layers of multiplex data sets', *Physical Review E* **92** (4), 042806

[43] B. Balassa (1965) 'Trade liberalization and "revealed" comparative advantage', *Manchester School* **33**, 99

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:

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

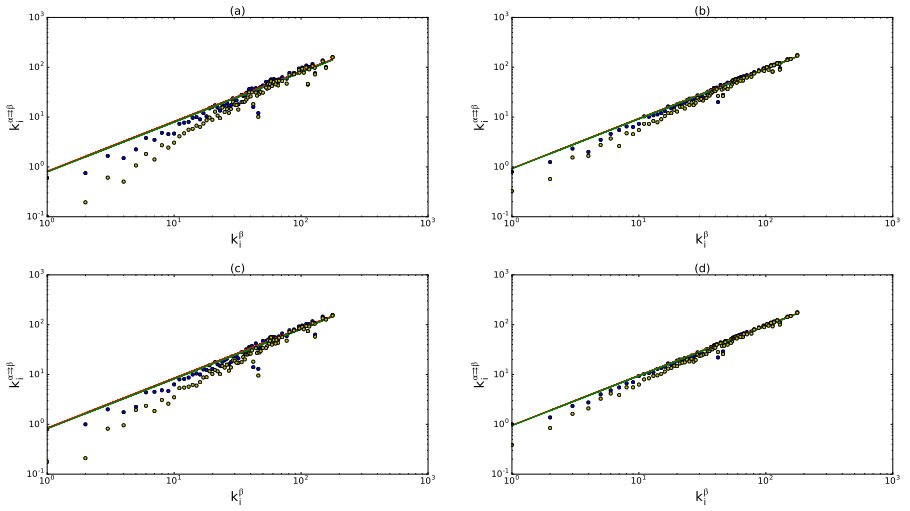

Figure 3.1: **Degree of layer β versus inter-layer multiplexed degree for 4 different pairs of commodities**: inorganic chemicals (a), plastics (b), iron and steel (c), electric machinery (d) versus trade in cereals. Blue dots: real data; yellow dots: expected multiplexed degree according to the uncorrelated 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:

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

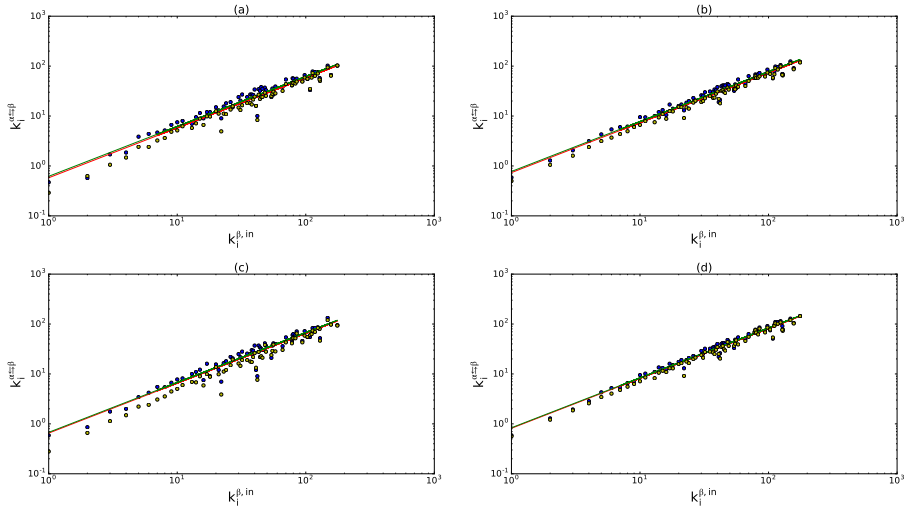

Figure 3.2: **In-degree of layer β versus inter-layer multireciprocated degree for 4 different pairs of commodities:** inorganic chemicals (a), plastics (b), iron and steel (c), electric machinery (d) versus trade in cereals. Blue dots: real data; yellow dots: expected multireciprocated degree according 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

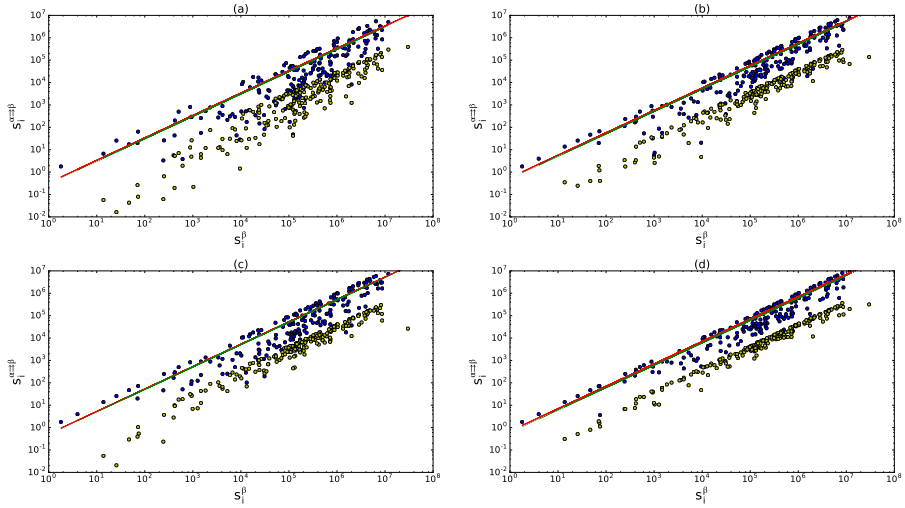

Figure 3.3: **Strength of layer β versus inter-layer multiplexed strength for 4 different pairs of commodities:** inorganic chemicals (a), plastics (b), iron and steel (c), electric machinery (d) versus trade in cereals. Blue dots: real data; yellow dots: expected multiplexed strength according to the 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

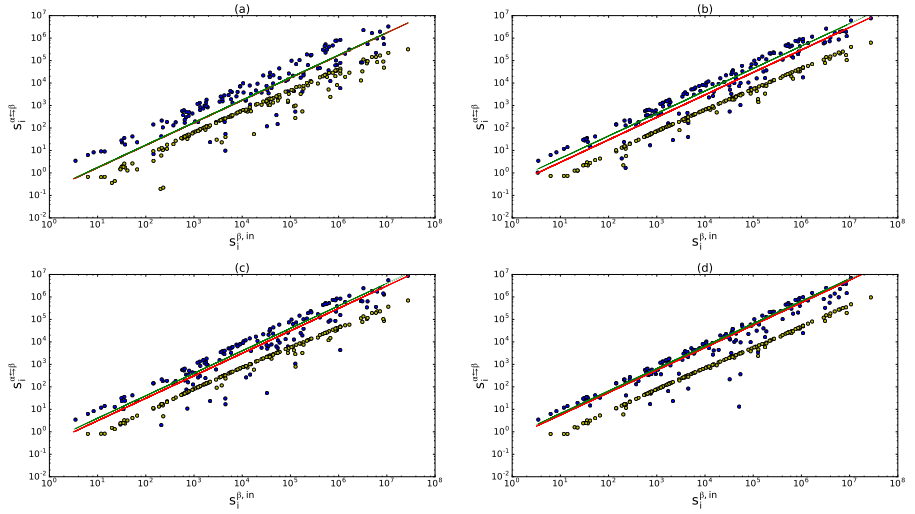

Figure 3.4: **In-strength of layer β versus inter-layer multireciprocated strength for 4 different pairs of commodities:** inorganic chemicals (a), plastics (b), iron and steel (c), electric machinery (d) versus trade in cereals. Blue dots: real data; yellow dots: expected multireciprocated strength 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

Chapter 4

Backbone extraction

Networks provide an informative, yet non-redundant description of complex systems only if links represent truly dyadic relationships that cannot be directly traced back to node-specific properties such as size, importance, or coordinates in some embedding space. In any real-world network, some links may be reducible, and others irreducible, to such local properties. This dichotomy persists despite the steady increase in data availability and resolution, which actually determines an even stronger need for filtering techniques aimed at discerning essential links from non-essential ones. Here we introduce a rigorous method that, for any desired level of statistical significance, outputs the network backbone that is irreducible to the local properties of nodes, *i.e.* their degrees and strengths. Unlike previous approaches, our method employs an exact maximum-entropy formulation guaranteeing that the filtered network encodes only the links that cannot be inferred from local information. Extensive empirical analysis confirms that this approach uncovers essential backbones that are otherwise hidden amidst many redundant relationships and inaccessible to other methods. For instance, we retrieve the hub-and-spoke skeleton of the US airport network and many specialised patterns of international trade. Being irreducible to local transportation and economic constraints of supply and demand, these backbones single out genuinely higher-order wiring principles.

The results presented in this chapter have been published in the following reference:
V. Gemmetto, A. Cardillo, D. Garlaschelli, *arXiv*:1706.00230 (2017).

4.1 Introduction

Over the last two decades, networks have become a standard tool of analysis of many complex systems. A network can represent any instance of a natural, social or technological system as a collection of *nodes* (or *vertices*) connected by *links* (or *edges*) [1, 2, 3, 4, 5, 6, 7, 8]. In many cases, it is possible to quantify the magnitude of interaction between two nodes, and encode it as a numerical value – the *weight* – attached to the link connecting them. In these cases, we refer to those networks as *weighted* [9]. Weighted networks are the focus of this chapter.

A recurrent and robust feature observed in the vast majority of real-world networks is a striking heterogeneity of the topological properties of different nodes. For instance, both the number of connections of a node (the so-called *degree*) and the total weight of these connections (the so-called *strength*) differ greatly across nodes in a network. Typically, the empirical distribution of both quantities in a given network is a power law, or more generally a broad distribution with 'fat tails'. Various models have been introduced in order to propose explanations for the origin of such distributions in real networks. Regardless of the possible mechanisms generating it, the observed heterogeneity of nodes has immediate consequences for the way real networks can or should be analysed. For instance, topological quantities averaged or summed over nodes (such as the total number of links, or the total link weight) are typically not informative about the whole network structure. Consequently, many structural quantities (such as the *clustering coefficient* measuring the relative abundance of triangles) should be defined *locally* and interpreted conditionally on the values of the degree and/or strength of nodes.

This chapter focuses on another important and well known consequence of the heterogeneity of nodes: namely, the impossibility of using a single reference value, or equivalently a unique global threshold, to assess the importance of different links in a given weighted network. If nodes have different strengths, global thresholds do not work, due to varying levels of statistical significance: a light-weighted link connecting two nodes with low strength may be even more significant than a heavy one connecting nodes with high strength [10]. This calls for techiques to assess the statistical significance of links based on the local properties of nodes. This problem arises in a variety of circumstances. One of the most important and recurrent examples is *graph filtering*, *i.e.* the identification of the most relevant links and the subsequent elimination of the least significant ones. Of course, whether a link qualifies as 'relevant' is largely problem-dependent and ultimately relates to the specific reason why the network is being filtered in the first place. Such a reason may be of practical or fundamental character. In this chapter, we propose a novel method of graph filtering whose motivation encompasses both aspects.

At a practical level, a widespread reason for filtering is the necessity of keeping real graphs *sparse* in order to directly and easily pinpoint the main relationships between the units of the system they represent. The recent data deluge associated

with the so-called *Big Data* era [11] has bolstered the tendency to represent complex systems as networks [12, 13, 14, 15]. Unfortunately, the huge opportunities associated to the availability of big data are not free of charges. The increase of information, in fact, usually produces a raise in the number of reported connections, which in turn increases the computational complexity of most algorithms used for network analysis and visualization. Effectively, this makes real networks harder to render, characterize and ultimately get a grip on. To continue using networks to retrieve the essential *backbone* of a complex system, uncluttering techniques are therefore needed.

At a more fundamental level, and irrespective of computational aspects, there is a subtler and less contemplated, yet very important reason for graph filtering. While this is seldom acknowledged, the network representation of a real system is non-redundant, thus really necessary, only if the presence and/or magnitude of the pairwise interactions between the units of the system cannot be entirely inferred from node-specific properties – such as size, importance, position in some underlying geometry, etc. Indeed, if node-specific properties were enough to characterize, infer or reconstruct the relationships among pairs of nodes (for instance if the network were a regular lattice embedded in space and the coordinates of nodes were given, or if the topology were a function of only the sizes of nodes and such sizes were known), then the network representation, while still correct, would be redundant. In general, some of the links of a given network may be 'reducible', and others 'irreducible', to node-specific properties. If so, the former would be in some sense unsurprising, while the latter would be much more interesting and provide truly dyadic information. This possibility calls for the introduction of *graph filtering techniques that allow irreducible links to be discerned from reducible ones, thus highlighting the non-redundant backbone of a network.* One should at this point note that recently, motivated by the fact that certain (e.g. financial) networks cannot be empirical observed in their entirety because of privacy or confidentiality reasons, there has been a proliferation of techniques devised for reconstructing the hidden topology of such networks from partial information [16, 17]. A specific class of methods have significantly increased the level of predictability of network structure from local, node-specific information [18, 19, 20, 21, 22], thus showing that, indeed, real networks can have a considerably big 'reducible' component.

Taken together, the above two considerations imply that, as the empirical availability of node-related information increases and the toolkit for reconstructing networks from partial information expands, the question of what makes up the non-redundant, genuinely dyadic properties of a network becomes more important and more difficult to answer. In this chapter we focus on the problem of identifying the *irreducible backbone* of a weighted network from a novel, rigorous standpoint. We define such backbone as the collection of links that cannot be reconstructed via unbiased inference from the knowledge of the local topological properties of nodes. This new definition guarantees that, by construction, the network backbone only encodes truly dyadic information. We identify the irreducible backbone by constructing, for a given empirical network, a corresponding unbi-

ased maximum-entropy null model where both the degree and the strength of each node are preserved as ensemble averages. Such maximum-entropy model is described by the generalized Bose-Fermi distribution [23] and is called the *Enhanced Configuration Model* (ECM) [18] because it enhances the Weighted Configuration Model (*i.e.* the maximum-entropy ensemble of weighted networks with given node strengths) by adding the degree sequence as an extra constraint.

The choice of the ECM as our null model is motivated by a recent series of theoretical and empirical results focusing on the problem of network reconstruction [18, 19, 20, 21, 22]. These studies have shown that the information encoded in the degrees and strengths of the nodes can be used to replicate many higher-order properties of real-world networks [18, 19]. The knowledge of both degrees and strengths is crucial to this purpose. Indeed, using only the degrees (as in the Binary Configuration Model [24, 25]) would provide no information about link weights, while using only the strengths (as in the Weighted Configuration Model [25, 26]) would lead to an exceedingly high density of links resulting, tipically, in almost complete graphs. In conclusion, these studies show that, in presence of only local node-specific information, the best unbiased inference about the entire structure of a weighted network is the one obtained using both the strengths and the degrees of all nodes as input.

It is important to stress that, although the recent progress in the area of network reconstruction is a strong motivation for the present work, our approach goes in a direction that is entirely opposite to that of network reconstruction techniques. Indeed, in the network reconstruction approach, links are unknown and are inferred (i.e. *created*) probabilistically from the knowledge of node-specific quantities, trusting the resulting ensemble of random graphs as the best guess about the structure of the unobserved network. By constrast, in our approach the full network is already known from the beginning and links are *removed* if they are consistent with the random ensemble, which here acts as a filter (i.e. a *null* model) rather than an inference tool (i.e. a *generative* model). Thus, the links that are generated in the network reconstruction approach are precisely those that are removed here. In this entirely opposite perspective, we have to complement the generative, probabilistic toolkit of maximum entropy ensembles with the introduction of a new, statistic toolkit of hypothesis testing.

The rest of the chapter is organized as follows. In sec. 4.2 we discuss the novel ingredients of our framework with respect to previous graph filtering approaches and we further elaborate on the above key difference between our method and the problem of network reconstruction. In sec. 4.3 we describe our method for the extraction of irreducible network backbones in detail. In sec. 4.4 we show the results of an extensive empirical analysis using our method. In sec. 4.5 we discuss further extensions of our method to directed and bipartite networks. Finally, in sec. 4.6 we make some concluding remarks, followed by some additional results in appendix.

4.2 Relation to previous work

Despite our motivation, *i.e.* the identification of the irreducible network backbone that cannot be inferred from the local topological properties, is quite novel, our work necessarily relates to previous literature. Under different names (e.g. thresholding, filtering, pruning, sparsification, backbone extraction, statistical validation, etc.), several algorithms have been proposed in order to remove links from a network. In general, the available approaches can be grouped in two main categories: *coarse graining* and *edge removal* methods. The former tend to merge together nodes with similar properties, thereby providing a hierarchical, multiscale view of the system [27, 28]. The latter fix instead the scale and proceed by removing connections. Edge removal methods split further into two sub-categories: *pruning* and *sparsification* techniques. Pruning approaches aim at removing connections to unveil some hidden structure/property of the system that is considered to be unknown *a priori*. On the other hand, sparsification techniques remove connections while preserving some property of the original system, thus aiming at retrieving comparable information but at a cheaper computational cost [29, 30, 31].

Among the pruning solutions, the most straightforward one is thresholding and its most recent variations (see e.g. [32]). Removing all the edges having a weight, w, lighter than a given value, w_t, produces systems surely sparser but at the cost of losing all their "weak ties" [10] and, more importantly, losing the weight heterogeneity which represents one of the hallmarks of complex systems [9]. Despite such serious limitations, thresholding has been used extensively, for example, in brain networks [5]. Another notorious technique is the extraction of the Minimum Spanning Tree (MST) albeit it delivers an over simplification of the system because it destroys many features (cycles, clustering and so on) [33, 34, 35]. Other pruning techniques are *link validation* methods, which produce what is sometimes called a *statistically validated network* [36]. These are the most similar to our method proposed here, yet still different, because in general the statistically validated network is not guaranteed to be irreducible (e.g. if the full information about strengths and degrees is not specified) or unbiased (e.g. if the procedure is not maximum-entropy). To the best of our knowledge, in fact, validation of empirical networks against maximum-entropy ensembles with constrained strenghts and degrees has not be done yet.

On the other hand, the idea of sparsification relies on the (often implicit) assumption that the properties of the original network, especially those to be preserved, are statistically significant and therefore worth preserving in the first place. This means that sparsification techniques require, at least in principle, that a preliminary filtering has already taken place. In this sense, the method we propose here is a filtering method that can be used for link pruning, link validation, and as a preliminary step for other sparsification methods.

Coming to a closer inspection of the available techniques, three methods are most directly related to what we are going to develop in this chapter: the Dis-

parity Filter (DF) introduced by Serrano *et al.* [37], the so-called GloSS method proposed by Radicchi *et al.* [38], and a method recently developed by Dianati [39]. Some details of these methods are briefly recalled in the remainder of this section. For completeness, we also stress the main differences between the filtering technique that will be introduced here and recent network reconstruction approaches that, while apparently related, go precisely in the opposite direction.

4.2.1 The Disparity Filter

The DF is very close in spirit to our method, since it constrains both the strength and the degree of each node [37]. However, there are crucial differences that we now explain.

The DF assumes that, in the null model, the strength s_i of each node i is redistributed uniformly at random over the k_i links of that node. The resulting criterion for establishing whether a link having weight w_{ij}^* satisfies the null hypothesis requires the computation of the following p-value

$$\gamma_{ij} = 1 - \int_0^{w_{ij}^*} \rho(w|s_i, k_i)\, dw\,, \tag{4.1}$$

and its comparison with a chosen critical value $\widetilde{\gamma}$. Here $\rho(w|s_i, k_i)$ is probability (density) that a link has weight w, under the null hypothesis that the value s_i is partitioned uniformly at random into k_i terms. By contrast, we will see that the correct maximum-entropy probability derived in our approach does not correspond to such uniformly random partitioning, as it collectively depends also on the strengths and degrees of all other nodes. This means that the DF introduces some bias. This bias can be understood by noticing that, when distributing the strength of a node at random over its links, it disregards the strength of the nodes at the other end of these links, thus effectively 'flattening' the weights received by these nodes. For each node i, this creates randomized weights that tend to be too small around high-strength neighbours and too high around low-strength neighbours. As a consequence, as we will confirm later, the DF has a bias towards retaining heavier connections.

From a mathematical point of view, the above problem is manifest in the fact that, despite $w_{ij}^* = w_{ji}^*$ (we are considering undirected networks for the moment), eq. (4.1) is not symmetric under the exchange of i and j, and in general one has $\gamma_{ij} \neq \gamma_{ji}$. To partly compensate for this, the DF is usually applied twice, from the perspective of each of the two endpoints of an edge. However, it should be noted that, in general, the resulting randomized weights cannot be actually generated in any network, as it is not possible to produce randomized link weights that realize the null hypothesis for all nodes simultaneously. So, in the end, the statistical test is based on an ill-defined null hypothesis.

Another evidence of the above problem is the fact that, under the correct maximum-entropy model, the level of heterogeneity of the weights of the links

incident on a node (as measured for instance by the so-called 'disparity') does not take on the values one usually expects under the assumption of uniform randomness and is strongly biased towards other values [23]. A consequent bias must necessarily arise in the p-values as calculated above.

A final difference with respect to our method is that the DF enforces the degrees and strengths sharply (i.e. as in the *microcanonical* ensemble in statistical physics), whereas we enforce them only as ensemble averages (i.e. as in the *canonical* ensemble) [40]. The microcanonical implementation, even if carried out in a correct and unbiased way, implies statistical dependencies between the weights of all edges in the null model, because these weights must add up to a deterministic value. This in turn implies that the statistical test cannot be carried out separately for each edge. In our canonical implementation of the null model, all edges are instead independent, a property that allows us to consistently carry out the statistical test for each node separately, even if, as we mentioned, our p-value for each link depends on the degrees and strengths of all other nodes in the network, as desired. The recent results about the non-equivalence of microcanonical and canonical ensembles of random graphs with given strengths and/or degrees [40, 41, 42] imply that the two approaches remain different even in the limit of large network size, and must therefore lead to different results.

4.2.2 The GloSS method

In the GloSS method [38], the null model used to assign p-values to edges is a network with exactly the same topology of the original network and with link weights randomly drawn from the empirical weight distribution $P(w)$. This effectively means that the observed link weights are randomly reshuffled over the existing, fixed topology. Unlike the local criterion of the DF, this choice results in a *global* null model. Indeed, since links are fixed and they all have the same probability of being assigned a given weight, the statistical test is the same for every existing edge, and the method effectively reduces to selecting the strongest weight only, thus setting a global threshold which depends on the desired confidence level. This leads us back to the problem of global thresholds being inappropriate for networks with strong heterogeneity.

Another, related problem with GloSS is the fact that it conceives the topology and the weights as two separate, or separable, network properties. This is however hard to justify, since the topology is encoded in the adjacency matrix whose entries $\{a_{ij}\} = \{\Theta(w_{ij})\}$ are binary projections of the link weights $\{w_{ij}\}$, and thus entirely dependent on the latter. Indeed, as a result of this decoupling, the null model turns light links into heavy ones and viceversa, irrespective of the importance of the end-point vertices. In other words, it views the weight distribution as unconditional on the strengths of the end-point vertices. The strengths are viewed as the result of, rather than a constraints for, a random realization of weights. The resulting expected strengths are indeed proportional to the observed degrees. One can partly relax the null model by globally reshuffling the weights

while simultaneously randomizing the topology in a degree-preserving way [43], but the method will still retain the proportionality between the expected strength and the degree of a node, and the underlying notion of complete separability of topology and weights.

It should be noted that the proportionality between strengths and degrees in the null model violates the strongly non-linear relationship observed between these two quantities in real-world networks [9, 20, 21]. Coming back to the network reconstruction problem, this implies that GloSS suffers from what we may call a *redundancy* problem: since many properties of real-world networks can be inferred from the empirical degrees and strengths of nodes (as we will briefly recall below), GloSS cannot ensure that the filtered network is irreducible to the knowledge of such node-specific properties. Indeed, such properties contain in general more information than what is retained in the null model. The resulting network backbone may therefore still contain redundant interactions.

4.2.3 The 'hairball' method

Dianati has recently proposed a different 'hairball' approach [39] where, unlike the methods discussed above, the null model is maximum-entropy based and therefore unbiased. The constraints imposed on entropy maximization are the strenghts of all nodes, but not the degrees. Dianati considers two distinct null models: a local one, acting on single links, named Marginal Likelihood Filter (MLF) and a global one, acting on the network as a whole, named Global Likelihood Filter (GLF). Both null models produce graphs which, for a given p-value, are quite alike.

Although the maximum-entropy nature of the filter introduced by Dianati fixes the problem of bias encountered by the other approaches, the method still suffers from the redundancy problem, even if in a direction in some sense opposite to that of GloSS. Concretely, constraining only the strength sequence in the null model corresponds to generating almost complete networks [25, 26]. This implies that the nonlinear empirical strength-degree relation is again violated, here because the degree of each node tends to saturate to the maximum allowed value and is therefore independent of the strength. Once more, this does not guarantee that the filtered network is irreducible to the knowledge of the degrees and strenghts of nodes. Indeed, the weights are redistributed among virtually all the possible pairs of nodes, which also means that the empirical non-zero link weights are systematically larger than those generated by the null model. This implies that the filter tends to retain too many spurious links.

4.2.4 Network reconstruction methods

As we mentioned in sec. 4.1, there has recently been a flourishing of methods to reconstruct networks from partial information. Among the motivations for the introduction of these methods, a prominent one is the frequent lack of transparency about real-world financial networks or the lack of fine-grained data about social

contact networks. Fortunately, our specific need in this chapter allows us to restrict to a well defined class of network reconstruction techniques, among the zoo of available ones [16].

Since our main goal is that of separating truly dyadic properties from local node-specific ones, the relevant class of network reconstruction models that can guide the definition of our method is the one that assumes that the partial information available about the network is local and node-specific. The methods in this class start from the knowledge of the strengths and/or degrees (of all or a subset of the nodes) and infer the overall network topology by constructing a maximum-entropy (i.e. maximally unbiased) ensemble of graphs consistent with these properties [18, 19, 20, 21, 22]. The final output is an ensemble of random graphs where links are created with some probability.

It turns out that, in order to achieve a reliable reconstruction of an unknown weighted network, the optimal set of node-specific properties to be enforced as constraints is given precisely by the degrees and the strengths of all nodes [18]. Indeed, including only the degrees would predict the topology fairly well, but would provide no information about link weights [24, 25]. By contrast, as we have already mentioned, including only the strengths would produce networks that are overly dense and tend to be almost completely connected [25, 26]. This is the result of the fact that strengths contain no information about the bare topology of the network, and ensembles constructed from node strengths tend to dilute the total link weight among basically all pairs of nodes. Statistical tests confirm that the information contained in the degrees is indeed irreducible to that contained in the strengths [18]. Indeed, it has been shown that, if the degrees of nodes are not empirically accessible, the success of the reconstruction method entirely depends on how well one is able to preliminarily obtain reliable estimates of the degrees, before constructing maximum-entropy null models that simultaneously preserve the (observed) strengths and the (inferred) degrees [19, 20, 21, 22].

The above results lead us to choose the ECM, where both strengths and degrees are enforced, as the most appropriate ensemble for our filtering purposes, because it provides the most accurate inference possible about a weighted network, if only local node-specific properties can be accessed.

It is important to stress again that, although our filtering framework will largely build upon the mathematical properties of the ECM, its goal is basically opposite to that of the network reconstruction methods based on the same model. Indeed, in our approach we do know the entire network, and our aim is that of filtering out what might be inferred about it if we were given only local information. To this end, we use the ECM to remove (not create) connections from the real network. The output is therefore not the ensemble of random graphs, but precisely the opposite, i.e. the sets of nodes/edges of the original network that are not compatible (within any desired level of statistical significance) with the random graph model embodying the null hypothesis. In order to carry out this program, we indeed need to introduce new ingredients to the ECM framework, in particular the calculation of a p-value, separately for each link, based on the

observed link weight and of the likelihood of subgraphs of the original network. These steps will enable us to define both local and global new schemes for graph filtering.

Again, the rationale behind this entirely new approach is that, if connections can be predicted based only on local properties (i.e. via network reconstruction), they are in some sense redundant, i.e. reducible to a list of node properties, and the network representation is therefore unnecessary. We thus aim at finding the graph that is non-redundant, i.e. maximally unlikely to be produced via network reconstruction. This approach pushes the field of graph filtering into a new direction.

4.3 Extraction of irreducible backbones: the ECM filter

In the rest of this chapter, we aim at combining together the good ingredients of previous methods, while overcoming their most important limitations. In particular, we want to retain the unbiasedness of the maximum-entropy approach proposed by Dianati while keeping the empirical non-linear relationship between strengths and degrees as in the DF.

We therefore introduce a new filtering method based on the comparison between a given real-world weighted network and a canonical *maximum-entropy* ensemble of weighted networks having (on average) the same degree sequence and the same strength sequence as the real network, *i.e.* the ECM. Our model can be fully characterized analytically, a property that allows us to explicitly calculate the exact p-value for each realized edge in the original network. Unlike the DF, our maximum-entropy construction ensures consistency from the point of view of both nodes at the endpoint of an edge and makes the null hypothesis realizable by the networks in the statistical ensemble. It also makes different edges statistically independent, thus justifying the establishment of a separate test for each observed link. We remark that, unlike all previous approaches, the use of the ECM ensures that the filtered network cannot be retrieved by any impartial and unbiased network reconstruction method that starts from local information. This also fixes the redundancy problem of other methods.

Below, we first briefly review the definition and main properties of the ECM [18, 40], which has been originally introduced under the name of *Bose-Fermi* ensemble [23], and then provide a new recipe to use it for the purpose of graph filtering.

4.3.1 The Enhanced Configuration Model or Bose-Fermi Ensemble

Very generally, a maximum-entropy model is a canonical ensemble described by a probability distribution $P(G)$ (over the microscopic configurations $\{G\}$ of the system) that maximizes the Shannon-Gibbs entropy $S = -\sum_G P(G) \ln P(G)$, while

satisfying a given set of macroscopic constraints enforced as ensemble averages [44]. The formal solution to this problem is a Boltzmann-like (*i.e.* exponential) probability function of the form $P(G) = Z^{-1}e^{-H(G)}$, whose negative exponent, sometimes (improperly) termed "Hamiltonian (function)" $H(G)$, is a linear combination of the constraints and whose normalization constant is the inverse of the so-called "partition function" $Z = \sum_G e^{-H(G)}$. We are now going to define these quantities rigorously for the case of interest to us.

We consider an ensemble $\mathcal{W}$ of undirected, unipartite weighted networks with a fixed number N of nodes (an explicit generalization to the case of directed and bipartite networks is provided later in sec. 4.5). Each element of $\mathcal{W}$ is a weighted graph, uniquely specified by a $N \times N$ symmetric matrix $\boldsymbol{W}$ whose entry $w_{ij} = w_{ji}$ represents the weight of the link connecting node i to node j ($w_{ij} = 0$ means that i and j are not connected). Without loss of generality, we assume integer weights ($w_{ij} = 0, 1, 2, \dots$) and no self-loops ($w_{ii} = 0$ for all i). Starting from the matrix $\boldsymbol{W}$, one can construct the *adjacency matrix* $\boldsymbol{A}(\boldsymbol{W})$ whose entry is defined as $a_{ij}(\boldsymbol{W}) = \Theta(w_{ij})$, *i.e.* $a_{ij}(\boldsymbol{W}) = 1$ if $w_{ij} > 0$ and $a_{ij}(\boldsymbol{W}) = 0$ if $w_{ij} = 0$. Given a network $\boldsymbol{W}$, the *strength* of node i is defined as $s_i(\boldsymbol{W}) = \sum_{j \neq i} w_{ij}$ and the *degree* of node i is defined as $k_i(\boldsymbol{W}) = \sum_{j \neq i} a_{ij}(\boldsymbol{W})$.

Let us consider an empirical network, $\boldsymbol{W}^*$, that we would like to filter. We define $s_i^* \equiv s_i(\boldsymbol{W}^*)$ and $k_i^* \equiv k_i(\boldsymbol{W}^*)$, so that the resulting *empirical strength and degree sequences* are denoted as $\vec{k}^*$ and $\vec{s}^*$ respectively. We look for the probability distribution P over graphs that maximizes the entropy, under the constraint that the expected degree and strength of each node equal the empirical values, e.g.

$$\langle \vec{k} \rangle = \vec{k}^*, \quad \langle \vec{s} \rangle = \vec{s}^*. \tag{4.2}$$

The above requirement introduces a Lagrange multiplier, which for later convenience we denote as $-\ln x_i$ (with $x_i > 0$ for all i), for each expected degree $\langle k_i \rangle$ and another multiplier, denoted as $-\ln y_i$ (with $0 < y_i < 1$ for all i), for each expected strength $\langle s_i \rangle$. The graph probability we are looking for will depend on these $2N$ parameters, which we array in two N-dimensional vectors $\vec{x}$ and $\vec{y}$. We require that such probability, denoted as $P(\boldsymbol{W}|\vec{x}, \vec{y})$ from now on, maximizes the Shannon-Gibbs entropy

$$S(\vec{x}, \vec{y}) = -\sum_{\boldsymbol{W} \in \mathcal{W}} P(\boldsymbol{W}|\vec{x}, \vec{y}) \, \ln P(\boldsymbol{W}|\vec{x}, \vec{y}) \tag{4.3}$$

subject to the constraints in (4.2) and to the normalization condition

$$\sum_{\boldsymbol{W} \in \mathcal{W}} P(\boldsymbol{W}|\vec{x}, \vec{y}) = 1. \tag{4.4}$$

The solution to the above constrained maximization problem is found to be [23, 18, 40] the probability

$$P(\boldsymbol{W}|\vec{x}, \vec{y}) = \frac{e^{-H(\boldsymbol{W}|\vec{x}, \vec{y})}}{Z(\vec{x}, \vec{y})} = \prod_{i=1}^{N} \prod_{j<i} q_{ij}(w_{ij}), \tag{4.5}$$

where we have introduced the Hamiltonian

$$\begin{aligned} H\left(\boldsymbol{W}|\,\vec{x},\vec{y}\right) &= -\sum_{i=1}^{N}\left[k_i\left(\boldsymbol{W}\right)\ln x_i + s_i\left(\boldsymbol{W}\right)\ln y_i\right] \\ &= -\sum_{i=1}^{N}\sum_{j<i}\left[\Theta(w_{ij})\ln(x_i x_j) + w_{ij}\ln(y_i y_j)\right], \end{aligned} \tag{4.6}$$

the partition function

$$\begin{aligned} Z(\vec{x},\vec{y}) &= \sum_{\boldsymbol{W}\in\mathcal{W}} e^{-H(\boldsymbol{W}|\vec{x},\vec{y})} \\ &= \prod_{i=1}^{N}\prod_{j<i}\frac{1-y_i y_j + x_i x_j y_i y_j}{1-y_i y_j}, \end{aligned} \tag{4.7}$$

and the probability that a link between nodes i and j has weight w:

$$\begin{aligned} q_{ij}(w) &\equiv \frac{\left(x_i x_j\right)^{\Theta(w)}\left(y_i y_j\right)^{w}\left(1-y_i y_j\right)}{1-y_i y_j+x_i x_j y_i y_j} \\ &= \begin{cases} 1-p_{ij} & \text{if} \quad w=0 \\ p_{ij}\left(y_i y_j\right)^{w-1}\left(1-y_i y_j\right) & \text{if} \quad w>0 \end{cases}, \end{aligned} \tag{4.8}$$

with

$$p_{ij} \equiv 1-q_{ij}(0) = \frac{x_i x_j y_i y_j}{1-y_i y_j+x_i x_j y_i y_j} \tag{4.9}$$

representing the probability that nodes i and j are connected, irrespective of the weight $w_{ij}>0$ of the link connecting them.

The key quantity describing the above maximum-entropy ensemble is $q_{ij}(w)$, whose expression (4.8) has been first derived in [23] and denoted as *Bose-Fermi* distribution. The name comes from the fact that, as a result of enforcing both degrees and strenghts, the distribution combines features of the Bose-Einstein distribution, which is encountered when dealing with systems described by integer configurations such as $\boldsymbol{W}$, and the Fermi-Dirac distribution, which is encountered when dealing with systems described by binary configurations such as $\boldsymbol{A}(\boldsymbol{W})$.

We now come back to the real-world network $\boldsymbol{W}^*$ that we want to prune, and to its strength and degree sequences $\vec{s}^*$ and $\vec{k}^*$. Using the above form of $q_{ij}(w)$, $\langle\vec{s}\rangle$ and $\langle\vec{k}\rangle$ can be calculated explicity, both as functions of $\vec{x}$ and $\vec{y}$, so that the condition (4.2) can be rewritten explicitly [18, 40] as

$$k_i^* = \sum_{j\neq i}\frac{x_i x_j y_i y_j}{1-y_i y_j+x_i x_j y_i y_j} \quad \forall i, \tag{4.10}$$

$$s_i^* = \sum_{j\neq i}\frac{x_i x_j y_i y_j}{(1-y_i y_j+x_i x_j y_i y_j)(1-y_i y_j)} \quad \forall i. \tag{4.11}$$

The above system of $2N$ coupled nonlinear equations is solved by certain parameter values $(\vec{x}^*, \vec{y}^*)$. Equivalently, the values $(\vec{x}^*, \vec{y}^*)$ can be proven to coincide with the values that maximize the log-likelihood of the model [40, 45, 46], *i.e.*

$$(\vec{x}^*, \vec{y}^*) = \underset{\{x_i>0,\, 0<y_i<1\, \forall i\}}{\operatorname{argmax}} \mathcal{L}(\vec{x}, \vec{y}) \tag{4.12}$$

where the log-likelihood, $\mathcal{L}$, is defined as

$$\begin{aligned} \mathcal{L}(\vec{x}, \vec{y}) &= \ln P(\boldsymbol{W}^*|\vec{x}, \vec{y}), \\ &= \sum_{i=1}^{N} \sum_{j<i} \ln q_{ij}(w_{ij}^*) \\ &= \sum_{i=1}^{N} [k_i^* \ln x_i + s_i^* \ln y_i] \\ &\quad - \sum_{i=1}^{N} \sum_{j<i} \ln \frac{1 - y_i y_j + x_i x_j y_i y_j}{1 - y_i y_j}. \end{aligned} \tag{4.13}$$

Once $P(\boldsymbol{W}|\vec{x}, \vec{y})$ is evaluated at the parameter values $(\vec{x}^*, \vec{y}^*)$, we obtain the explicit maximum-entropy probability distribution $P(\boldsymbol{W}|\vec{x}^*, \vec{y}^*)$ that we were looking for. For ease of notation, once the values $(\vec{x}^*, \vec{y}^*)$ are inserted into Eqs. (4.8) and (4.9), we denote the resulting key probabilities as q_{ij}^* and p_{ij}^* respectively.

4.3.2 The local filter

We can now introduce our new filtering technique. Starting from the properties of the ECM summarized in the previous subsection, here we develop a novel statistical test that uses the ECM as null hypothesis and arrives at the explicit calculation of the p-value for the acceptance of each link in the network. We consider a local (and, as we argue later, more appropriate) version of our filtering method first, and then move on to a global one.

As we anticipated, our local filtering method is similar in spirit to the Disparity Filter (DF) introduced by Serrano *et al.* [37], as it is based on the calculation of a p-value γ_{ij}^* for each observed link of weight $w_{ij}^* > 0$, defined as the probability that the null model produces a weight $w_{ij} \geq w_{ij}^*$, and on the removal of links for which γ_{ij}^* is higher than a fixed critical value $\widetilde{\gamma}$. However, our method improves upon the DF by recalculating the p-values according to the maximum-entropy probability q_{ij}^* derived above. Since in our case weights are discrete, p-values

should be calculated by replacing the integral appearing in Eq. (4.1) with a sum:

$$\begin{aligned}
\gamma_{ij}^* &\equiv \text{Prob}(w_{ij} \geq w_{ij}^*) \\
&= \sum_{w \geq w_{ij}^*} q_{ij}^*(w) \\
&= \begin{cases} 1 & \text{if } w_{ij}^* = 0 \\ 1 - \sum_{w=0}^{w_{ij}^*-1} q_{ij}^*(w) & \text{if } w_{ij}^* > 0 \end{cases} .
\end{aligned} \tag{4.14}$$

If a link is actually present in the observed network, *i.e.* $w_{ij}^* > 0$, we have

$$\begin{aligned}
\gamma_{ij}^* &= 1 - \sum_{w=0}^{w_{ij}^*-1} q_{ij}^*(w) \\
&= 1 - \sum_{w=0}^{w^*-1} \frac{\left(x_i^* x_j^*\right)^{\Theta(w)} \left(y_i^* y_j^*\right)^w \left(1 - y_i^* y_j^*\right)}{1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*} \\
&= 1 - \frac{1 - y_i^* y_j^*}{1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*} \left[1 + x_i^* x_j^* \sum_{w=1}^{w^*-1} \left(y_i^* y_j^*\right)^w\right] \\
&= \frac{x_i^* x_j^* \left(y_i^* y_j^*\right)^{w_{ij}^*}}{1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*} \\
&= p_{ij}^* \left(y_i^* y_j^*\right)^{w_{ij}^*-1} .
\end{aligned} \tag{4.15}$$

The above quantity represents the probability of generating a link between nodes i and j with a weight equal to, or greater than, the observed weight w_{ij}^*. It can be seen from Eq. (4.15) that this probability coincides with the probability p_{ij}^* that a link of unit weight is established, times the probability $\left(y_i^* y_j^*\right)^{w_{ij}^*-1}$ that the weight is successfully incremented $w_{ij}^* - 1$ times (so that the total weight is at least w_{ij}^*), irrespective of whether possible attempts to further increment the weight beyond w_{ij}^* are successful or not.

Per se, γ_{ij}^* represents the p-value associated with the null hypothesis that the edge weight w_{ij}^* has been produced by mere chance, given the empirical strength and degree sequences $\vec{s}^*$ and $\vec{k}^*$. Links with a higher value of γ_{ij}^* are closer to compatibility with the null hypothesis. Therefore the quantity $1/\gamma_{ij}^* > 0$ can be viewed as a rescaling of the original weight $w_{ij}^* > 0$ that effectively reduces the absolute importance of large weights, if these are found between nodes with large strengths and/or degrees. In principle, this rescaling can already be considered a form of filtering, that keeps all edges but with modified weights. In practice, we are going to fix a threshold value (corresponding to a desired level of statistical significance) and retain only the edges for which $1/\gamma_{ij}^*$, rather than w_{ij}^*, is larger than the threshold. As we now show, this crucial step effectively replaces the problematic enforcement of a global threshold on the original weights with

the enforcement of a local threshold that controls for the strengths and degrees of nodes. In particular, we reject the null hypothesis, and therefore retain the observed link between nodes i and j as statistically significant, if γ^*_{ij} is smaller than a desired threshold $\widetilde{\gamma}$:

$$\gamma^*_{ij} < \widetilde{\gamma}. \tag{4.16}$$

Equivalently, using Eq. (4.15) the above homogeneous (global) threshold $\widetilde{\gamma}$ for γ^*_{ij} translates into the following heterogeneous (local) threshold $\widetilde{w}_{ij}$ for w^*_{ij}:

$$w^*_{ij} > 1 + \frac{\ln(\widetilde{\gamma}/p^*_{ij})}{\ln(y^*_i y^*_j)} \equiv \widetilde{w}_{ij}. \tag{4.17}$$

It should be noted that the term on the r.h.s. depends on x^*_i, x^*_j, y^*_i, y^*_j. In turn, these four parameters depend *on the entire empirical strength and degree sequences* $\vec{s}^*$ and $\vec{k}^*$ through Eqs. (4.10) and (4.11), or equivalently (4.12) and (4.13). So, unlike the DF, where the statistical significance of the observed edge weight w^*_{ij} is assessed against a null model that (upon double-checking from the point of view of both i and j) depends only on the endpoint properties s^*_i, k^*_i, s^*_j, k^*_j, here the statistical test for w^*_{ij} depends on the degrees and strengths *of all nodes in the network*. This is a desirable property, following from the maximum-entropy nature of our model whereby the specified constraints *collectively* determine the probability of each graph, and ultimately each edge, in the ensemble.

Summing up, our local filtering method is very simple: given the empirical network $\boldsymbol{W}^*$ with strength sequence $\vec{s}^*$ and degree sequence $\vec{k}^*$, we

- find the values $(\vec{x}^*, \vec{y}^*)$ through Eqs. (4.10) and (4.11), or equivalently (4.12) and (4.13) (efficient algorithms serving this purpose have been devised [40] and coded [47, 48]);
- retain only the links (along with their weight w^*_{ij}) that realize Eq. (4.16), or equivalently (4.17), for a given value of the threshold $\widetilde{\gamma}$ (a generally accepted reference choice is $\widetilde{\gamma} = 0.05$, although we will show results for a wide range of values of $\widetilde{\gamma}$).

We refer to the resulting pruned network as the *local backbone* of $\boldsymbol{W}^*$ and denote it in terms of the ($\widetilde{\gamma}$-dependent) matrix $\boldsymbol{\Sigma}^{\text{local}}(\widetilde{\gamma})$ with entries $\sigma^{\text{local}}_{ij}(\widetilde{\gamma})$. Clearly, the extreme cases are $\boldsymbol{\Sigma}^{\text{local}}(1) = \boldsymbol{W}^*$ (all links of the original network being preserved) and $\boldsymbol{\Sigma}^{\text{local}}(0) = \boldsymbol{0}$, the latter denoting a matrix will all zero entries, *i.e.* an empty graph.

4.3.3 The global filter

So far, we have implemented the filter locally by computing the significance of each link γ_{ij} and comparing it with a given critical p-value $\widetilde{\gamma}$. However, in analogy

with [39], it is in principle possible to apply the same filter in a global, whole-graph fashion. In this context, we assume that the most significant *global backbone* $\boldsymbol{\Sigma}^{\text{global}}(\widetilde{L})$ with $\widetilde{L} \leq L^*$ links (where L^* is the empirical number of links in the original network $\boldsymbol{W^*}$) is the minimum-likelihood subgraph among the set of all subgraphs of the original network having $\widetilde{L}$ edges. This is equivalent to claiming that $\boldsymbol{\Sigma}^{\text{global}}(\widetilde{L})$ is the subnetwork which is least likely to be generated by pure chance. For each weighted subgraph $\boldsymbol{\Sigma}$ of the observed network $\boldsymbol{W^*}$, the likelihood is

$$P(\boldsymbol{\Sigma}|\boldsymbol{W^*}) = \prod_{i<j} \left[q_{ij}(\sigma_{ij})\right]^{a^*_{ij}} = \prod_{i<j} \left[q_{ij}(w^*_{ij})\right]^{a^*_{ij}} , \tag{4.18}$$

where σ_{ij} is the weight of the link between i and j in the subgraph $\boldsymbol{\Sigma}$ and $a^*_{ij} = 0, 1$ is the element of the adjacency matrix of the original graph $\boldsymbol{W^*}$. The global backbone (for given $\widetilde{L}$) is then defined as

$$\boldsymbol{\Sigma}^{\text{global}}(\widetilde{L}) = \underset{\boldsymbol{\Sigma}: L(\boldsymbol{\Sigma})=\widetilde{L}}{\text{argmin}} P(\boldsymbol{\Sigma}|\boldsymbol{W^*}), \tag{4.19}$$

where $L(\boldsymbol{\Sigma})$ denotes the number of links in the subgraph $\boldsymbol{\Sigma}$.

Given $\widetilde{L}$, the minimum of the likelihood is achieved by the $\widetilde{L}$ smallest factors of the product in Eq. (4.18). Hence, the entries of $\boldsymbol{\Sigma^*}(\widetilde{L})$ are easily found to be

$$\sigma^{\text{global}}_{ij}(\widetilde{L}) = \begin{cases} w^*_{ij} & \text{if } (i,j) \in \lambda_{\widetilde{L}}(\boldsymbol{W^*}) \\ 0 & \text{otherwise} \end{cases} ,$$

where $\lambda_{\widetilde{L}}(\boldsymbol{W^*})$ is the set of the $\widetilde{L}$ least likely links, *i.e.* those with the smallest probabilities $q_{ij}(w^*_{ij})$.

Note that, while the local filter selects links based on statistical significance, this is not the case for the global one. Nonetheless, it is worth comparing the local backbone, for a given $\widetilde{\gamma}$, with the global one obtained using a value of $\widetilde{L}$ giving as many links as the local backbone. This effectively establishes a relationship between $\widetilde{L}$ and $\widetilde{\gamma}$. It then becomes clear that the difference between the local filter and the global one is the fact that the latter selects the $\widetilde{L}$ links for which the probability mass function $\text{Prob}(w_{ij} = w^*_{ij})$ is minimum, while the former selects the $\widetilde{L}$ links for which the *cumulative* probability function $\text{Prob}(w_{ij} \geq w^*_{ij})$ is minimum. One can at this point note that, as in the usual construction of one-sided tests and the associated p-values in statistics, the use of the cumulative probability is much more reasonable, as it makes more sense to define compatibility with the null model in terms of the chance that the edge weight is equal to *or larger than*, rather than only equal to, the empirical one. We therefore claim that the local method should be preferred over the global one. We also note that, if the global filter were redefined in terms of cumulative probability, the two methods would coincide. Nevertheless, in the following we measure also the performances of the two methods and present an empirical *a posteriori* confirmation of our claim.

4.4 Empirical analysis

In this Section we gauge the performances of the ECM filter and compare its filtering power with that of the disparity and GloSS methods. All three methods are based on null models that, by preserving (among other properties) the degree sequence of the original network, automatically preserve the link density and therefore allow for a consistent comparison. We do not include the method by Dianati in this comparison since, as we mentioned, it does not preserve the link density and therefore tends to retain too many spurious links, many of which would be reducible to the knowledge of node degrees. After that, we compare the networks filtered with the local and global versions of ECM. Finally, we show how our filter is able to dig out interesting hidden patterns by presenting the results obtained for the time-varying World Trade and US airport networks.

4.4.1 Data

Here we provide a short description of the datasets used, and in Tab. 4.1 we list their fundamental topological features.

Domestic flights in the U.S.A. A node corresponds to an airport of U.S.A. and a link between two airports exists if there is a direct flight connecting them. The weight of a link indicates the number of passengers transiting between two airports [9].

Florida Bay Foodweb The network describes the trophic interactions between species during the dry season in the South Florida Bay ecosystem. The data have been collected from the ATLSS Project by the University of Maryland [49]. Nodes correspond to species and links represent the carbon flows (mg C y^{-1} m^{-2}) among them.

Star Wars movies The data portrait the interactions between the characters of the Star Wars films saga. Each node represents a character of the cast, while a link connects two characters if they both speak in the same scene and the weight counts the number of different scenes that they share across the seven episodes of the saga [50].

World Trade snapshots The networks represent the trading volumes between countries in the period between the years 1998 and 2011. Each year is encoded as a distinct network. A node indicates a country, while the weight of a link denotes the gross trade volume (measured in thousands of US dollars) between two countries [51, 52].

World Trade Multiplex Multiplex representation of trade volumes for year 2011. Each layer represents a different commodity [53]; we focus on four products,

Network	N	L	$\rho\,(\%)$
US airports network	426	2439	2.69
Florida Bay foodweb	126	1969	25.00
Star Wars network:			
All merged	111	444	7.27
All	112	450	7.24
Full int	110	398	6.64
Mentions	113	817	12.91
World trade network:			
Year 1998	208	10210	47.43
Year 1999	208	10904	50.65
Year 2000	208	11778	54.71
Year 2001	208	12256	56.93
Year 2002	208	12523	58.17
Year 2003	208	12796	59.44
Year 2004	208	12921	60.02
Year 2005	208	13145	61.06
Year 2006	208	13146	61.06
Year 2007	208	13230	61.45
Year 2008	208	13489	62.66
Year 2009	208	13360	62.06
Year 2010	208	13321	61.88
Year 2011	208	12956	60.18
World trade multiplex:			
Fish	207	4628	21.71
Cereals	207	3474	16.29
Fuel/oil	207	5711	26.79
Iron	207	5348	25.08

Table 4.1: **Topological characteristics of the datasets used.** For each network, we report the number of nodes N, of edges L and the link density ρ for the non filtered case.

namely: Fish, crustaceans and acquatic invertebrates (FISH); Cereals (CER); Mineral fuels, mineral oils and products of their distillation, bitumin substances, mineral wax (FUEL/OIL); Iron and steel (IRON) [51, 52].

4.4.2 Typical results and comparison with other methods

When it comes to performances, a good filtering technique should, ideally, be able to prune as many connections as possible while preserving the highest amount of

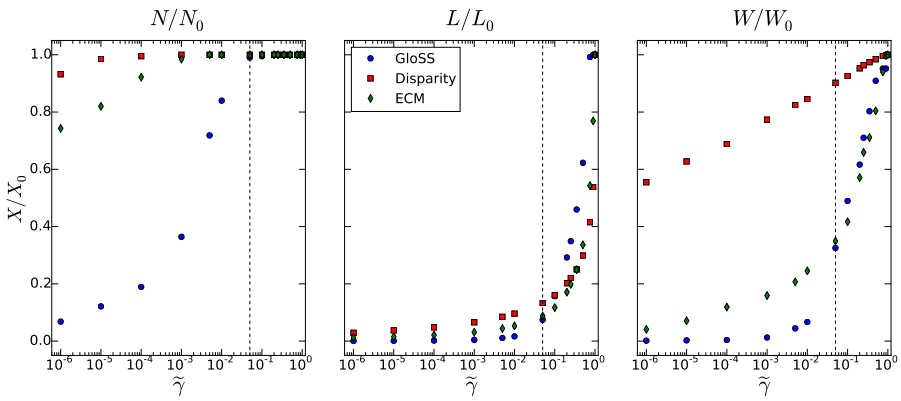

Figure 4.1: **Effects of filtering on three topological indicators, for the International Trade Network in 2011**. Fraction of nodes N/N_0 (left), links L/L_0 (center), and total weight W/W_0 (right) as a function of the p-value $\widetilde{\gamma}$. Red squares refer to the disparity filter, blue dots to GloSS, and green diamonds to ECM.

information and avoiding the breakup of the system. A good way to measure such ability is computing the fraction of a given quantity X preserved after filtering, X/X_0, where X_0 denotes the same quantity measured in the original network. We consider three indicators: number of nodes (N), of edges (L) and the total weight (W) respectively. The behaviour of these indicators for different filtering intensities (*i.e.* the p-values) for the International Trade Network in 2011 is displayed in Fig. 4.1.

In the left panel, we notice how all methods return networks with no isolated nodes up to $\widetilde{\gamma} \simeq 0.06$. Below such value, the disparity filter appears to be the most conservative method because its local nature tends to avoid the pruning of all the connections of a node. At the other extreme, despite imposing the conservation of the initial topology, GloSS is the most aggressive method, isolating more than 20% of nodes for $\widetilde{\gamma} < 0.05$. The ECM filter, instead, stays in between these boundaries and achieves a trade-off between its aggressive and conservative counterparts. In the strong filtering regime, which corresponds to the typical accepted range of p-values $\widetilde{\gamma} < 0.05$, the established hierarchy holds also for L/L_0 and W/W_0. The scenario changes, instead, for $\widetilde{\gamma} > 0.05$. In this regime ECM prunes out more connections than GloSS (central panel) as well as heavier than disparity ones (right panel). More specifically, since GloSS redistributes only the weights keeping the topology unaltered, this artificially boosts the significance of each link making it harder to remove. The stark difference in the behaviour of W/W_0 for the DF is, instead, the hallmark of bias towards heavier edges. Although the preservation of heavy connections might seem an advantage this is not always

the case. As we will show later, heavier connections tend to conceal interesting features of the system such as its mesoscopic structure (like, for instance, the presence of communities [54, 55]).

4.4.3 Local versus global filtering

Besides comparing the ECM filter with other existing solutions, it is worth comparing also its *global* and *local* versions. Given a certain p-value $\widetilde{\gamma}$, to properly compare the two implementations we first produce the local backbone, which results in a certain number $\widetilde{L}$ of links, and then rank all the edges of the original network according to their link probability q^*_{ij} using Eq. (4.8). We thus obtain the global backbone by retaining only the first $\widetilde{L}$ links. Finally, to study the differences between the two filtering approaches, in Fig. 4.2 we display the fraction of nodes, edges and total weight with respect to the p-value for the International Trade Network in 2011. By construction, the trends showing the fractions of retained links coincide. Furthermore, the analysis of Fig. 4.2 denotes no qualitative difference between the fraction of preserved nodes in the two methods. This is however not the case when we consider the residual total weight: indeed, we observe that the global ECM filter preserves significantly more weight than the local one, in particular for p-values higher than 10^{-4}.

A portrait of the differences between the local and global filter can be found in Fig. 4.3 and Tab. 4.2 (results for other datasets can be found in the appendix associated to this chapter), where we display the case of US airports dataset. The most striking feature of Fig. 4.3 is the stark difference between the local network (right panel) and the other two. In the local network, in fact, we observe the emergence of a clear hub-and-spoke pattern [56]. Indeed, the list of the 20 heaviest edges (Tab. 4.2) confirms that there is not very much difference between the global network and the original one in terms of backbone, while the difference becomes much stronger in the local case with the appearance of many connections among global "tier-1" hubs like New York and San Francisco and "tier-2" airports like Austin, Cleveland and Indianapolis, just to name a few as clearly shown in the graphs displayed in Fig. 4.3.

A more detailed analysis of the similarity between the local and global networks as a function of the p-value is provided by computing the Jaccard score J [57]. This score quantifies the similarity between two sets A and B by computing the ratio between the cardinality of the intersection and the cardinality of the union, *i.e.*

$$J = \frac{|A \cap B|}{|A \cup B|} \,. \tag{4.20}$$

A value $J = 1$ indicates that A and B are exactly the same set, while a value $J = 0$ denotes that the sets are completely different. In our case, we calculate J for the sets of edges belonging to the local and global networks computed using different

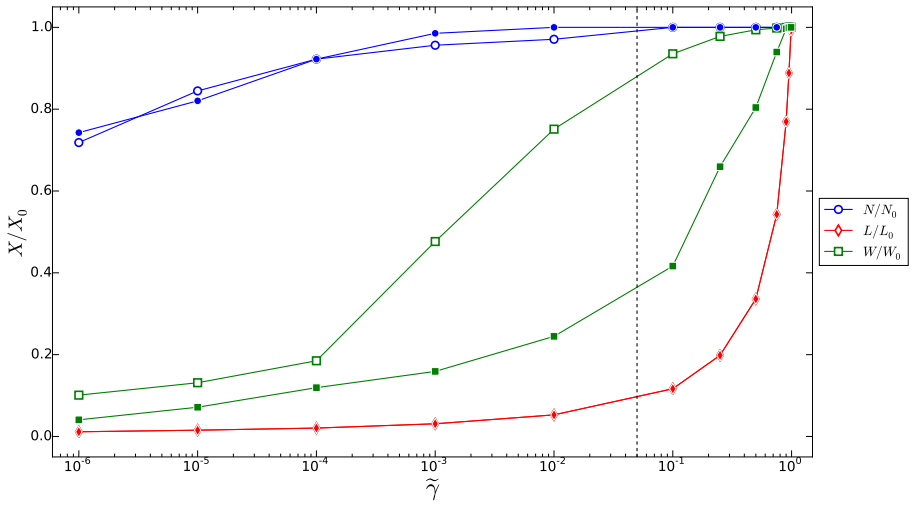

Figure 4.2: **Effects of the global and local implementations of the ECM filter on three topological indicators.** Fraction of nodes N/N_0 (blue dots), links L/L_0 (red diamonds), and total weight W/W_0 (green squares) as a function of the p-value $\widetilde{\gamma}$. Filled symbols refer to the local filter and empty symbols to the global one.

values of $\widetilde{\gamma}$. The results for all the datasets are visible in Fig. 4.4. The shape of J versus $\widetilde{\gamma}$ highlights two distinct behaviours. In one case, the similarity between local and global networks tends to fade away monotonically as we increase the aggressivity of the filtering. In the other case, the two networks initially tend to differentiate and become more and more alike thereafter. The World Trade Network is an example of the latter behaviour. As we increase the aggressivity of the filter, we observe the presence of a minimum of similarity around $\widetilde{\gamma} \simeq 0.35$ followed by an increase up to $J \approx 0.8$ for $\widetilde{\gamma} = 10^{-6}$. One culprit of such behaviour is that the original networks are, in general, denser than the others ($\langle\rho\rangle \gtrsim 58\%$ for time-varying and $\langle\rho\rangle \gtrsim 22\%$ for single commodities). For $\widetilde{\gamma} > 0.35$ (which corresponds to a low level of statistical significance), the ECM filters (local and global) tend to prune out different links as suggested by the decrease of J. Instead, the behavior of J for $\widetilde{\gamma} \leq 0.35$ (which includes all acceptable ranges of significance) suggests the emergence of a backbone shared by both networks which is resilient to pruning, resulting in an increase of J.

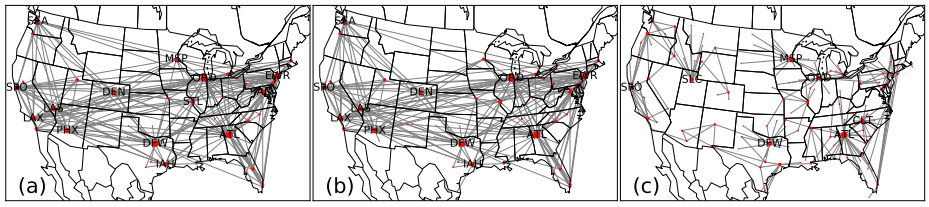

Figure 4.3: **Heaviest links in the US Airport Network.** Visual representation of the 200 heaviest links in the original network (left), ECM global filter (center) and ECM local (right) for the US Airport Network. The local filtering is obtained using $\widetilde{\gamma} = 0.05$ and the global filtering is constructed such that the number of links is the same as in the local one.

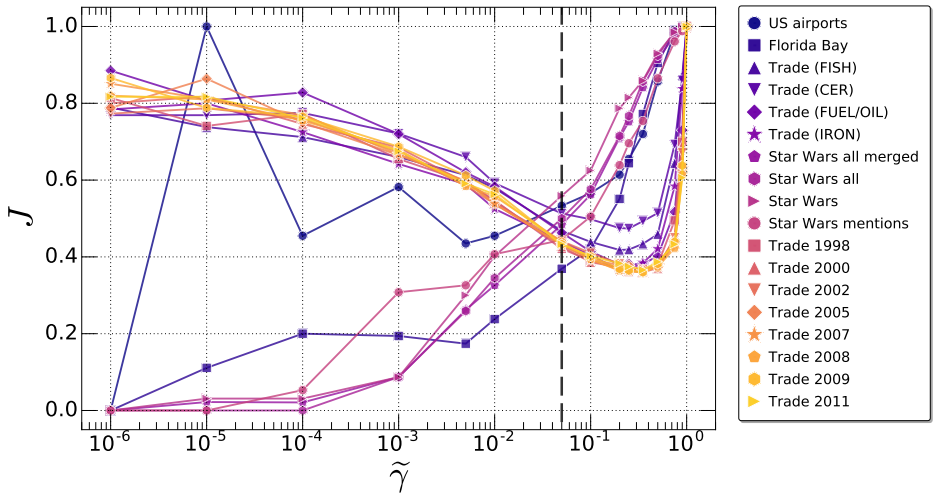

Figure 4.4: **Similarity between the local and global ECM filters.** Jaccard score J of local and global filtered backbones obtained at different p-values $\widetilde{\gamma}$ for all the datasets considered in our study. The vertical dashed line denotes $\widetilde{\gamma} = 0.05$.

4.4.4 The filter at work on multiplex networks

The presence of a similar trend in the Jaccard score between global and local backbones of yearly and single commodities leads us to investigate the effect that the aggregation of multiple commodities has on the extraction of the backbone. A *multiplex* network representation [53, 58] provides the natural way to study such an effect. It has been proven, in fact, that the topological properties of single layers and aggregate networks may differ a lot [59]; whilst in some cases, the multiplex can be *reduced*, deleting entire layers without losing information [60].

Orginal network	Local filter	Global filter
Los Angeles - San Francisco	Las Vegas - Los Angeles	Los Angeles - San Francisco
Las Vegas - Los Angeles	Boston - New York	Las Vegas - Los Angeles
Los Angeles - Phoenix	Seattle - San Francisco	Los Angeles - Phoenix
New York - Chicago	New York - Ft Lauderdale	New York - Chicago
Los Angeles - Chicago	San Diego - San Francisco	Los Angeles - Chicago
Dallas - Houston	Los Angeles - Sacramento	Dallas - Houston
New York - Los Angeles	Portland - San Francisco	New York - Los Angeles
Chicago - San Francisco	Houston - New Orleans	Chicago - San Francisco
Atlanta - New York	Kansas City - Chicago	Atlanta - New York
Boston - New York	Dallas - San Antonio	Boston - New York
New York - Washington	Austin - Dallas	New York - Washington
Dallas - Los Angeles	Houston - San Antonio	Dallas - Los Angeles
Seattle - San Francisco	Austin - Houston	Seattle - San Francisco
Las Vegas - San Francisco	Cleveland - Chicago	Las Vegas - San Francisco
New York - San Francisco	New York - West Palm B.	New York - San Francisco
New York - Ft Lauderdale	Albuquerque - Phoenix	New York - Ft Lauderdale
Minneapolis - Chicago	Spokane - Seattle	Minneapolis- Chicago
San Diego - San Francisco	Indianapolis - Chicago	San Diego - San Francisco
Los Angeles - Sacramento	Atlanta - Jacksonville	Los Angeles - Sacramento
Denver - Chicago	Reno - San Francisco	Denver - Chicago

Table 4.2: **Heaviest connections in the US Airport Network.** List of the 20 heaviest links in the US Airport Network in the original network (left column), and after applying the local (center) and global (right) ECM filters.

It is therefore reasonable to ask whether filtering the layers first, and projecting them onto a single layer then, produces a filtered backbone whose structural properties are different from those of the network obtained inverting the order of these operations.

In Fig. 4.5, we report the evolution of four topological indicators, namely: Jaccard score (J), number of edges (L), size of the giant component (S) and size of the mutually connected component (S_i), with respect to $\widetilde{\gamma}$. In Fig. 4.5(a) we display the similarity of the backbones using the Jaccard score J. We can gauge the similarity either in a topological sense (the same link existing in both backbones) or in a weighted one (the link existing in both backbones with the same weight w). Except for the case where no filtering is performed, the weighted similarity is always smaller than the topological one, suggesting that the connection among the same countries is significative for one specific commodity but not in the remaining ones. Additionally, after an initial increase, for $\widetilde{\gamma} < 0.01$ the difference between the topological and weighted similarities remains more or less constant. In general, we observe that filtering before projecting returns a network which has fewer edges L (Fig. 4.5(b)) and a smaller giant component S (Fig. 4.5(c)).

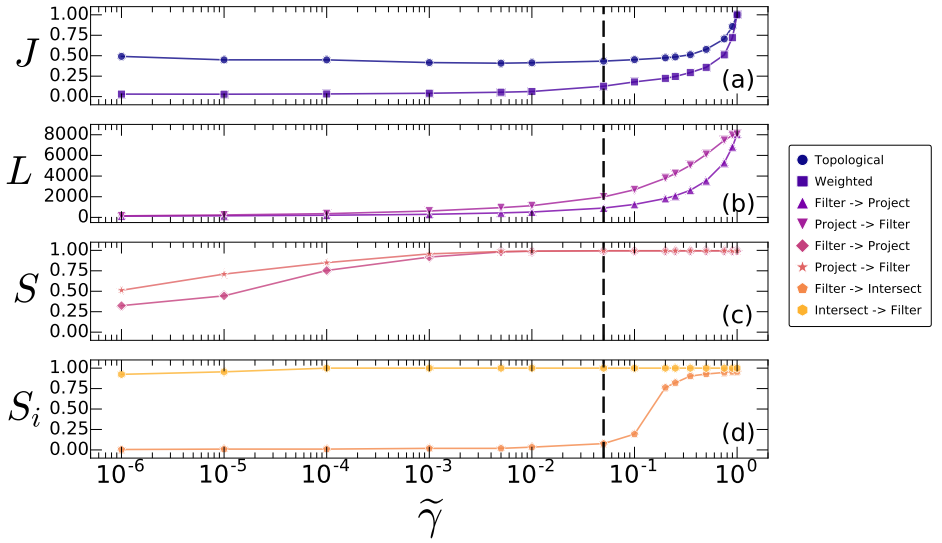

Figure 4.5: **Results of the ECM filter applied to the World Trade Multiplex.** Effect of multiplexity on the extraction of the weighted backbone. From top to bottom, we report the Jaccard score J (a), number of edges L (b), size of the giant component S (c) and size of the mutually connected component S_i (d) as a function of p-value $\widetilde{\gamma}$. The quantities L, S and S_i are computed in networks obtained filtering each layer first and projecting them then (Filter → Project/Intersect) or in the inverse order case (Project/Intersect → Filtering).

The existence of a giant component is crucial for the appearance of several collective phenomena like synchronization and spreading, just to cite a few [61]. In multiplex networks, besides the giant component in the single layers and in the aggregate network, the so-called *mutually connected component* (here defined as the network obtained projecting only the edges appearing in all the layers) plays a key role in the emergence of collective phenomena as well [62]. In Fig. 4.5(d), we compute the size of the mutually connected component, S_i, of the networks obtained filtering the layers first, extracting the intersection of the edge sets then and projecting them finally (orange pentagons). We also show the result obtained by computing the intersection first, projecting the layers and filtering the aggregate network then (yellow hexagons). As we can see from panel (d), the behavior of these two quantities with respect to $\widetilde{\gamma}$ is completely different. In particular, for $\widetilde{\gamma} < 0.2$ one case is above the critical percolation threshold while the other is already completely fragmented [63]. Moreover, a visual representation of the original networks and their respective backbones for some commodities is available in the appendix.

4.4.5 'Irreducible patterns' revealed by the method

Finally, we illustrate several results showing that the ECM-filtered backbones can unveil significant information about real-world systems that would otherwise remain hidden or not completely revealed by the other methods. We provide also some interpretation of the uncovered patterns. For the sake of brevity, here we discuss only the results obtained for the US airports and the International Trade networks, albeit similar conclusions can be drawn from the analysis of the other datasets as well, as shown in the appendix.

We start from US airports and refer to Fig. 4.6, where we show the original network (panel a) compared with the results of three filtering techniques: disparity (panel b), GloSS (panel c), and the local ECM filter (panel d). For all three methods, the backbones are extracted using always the same p-value $\widetilde{\gamma} = 0.05$ for consistency and, to facilitate visual comparison, we display only the first 200 heaviest connections. At first glance, we notice a stark difference between the three backbones. More specifically, the disparity backbone is akin to the original network, displaying several long-range connections between airports like Atlanta (ATL), Chicago (ORD), Newark (EWR) and Los Angeles (LAX) to mention a few. All these airports are among the top 12 in terms of the number of passengers in 2002 as reported by the Federal Aviation Administration (FAA) of the United States [64]. The pattern of connections resembles therefore a *point-to-point* one [8]. Qualitatively, this is in agreement with the tendency of the disparity filter to preserve heavier connections as reported in Fig. 4.1. This is not the case for GloSS (panel c) and ECM (panel d). The former returns a very sparse backbone having less than 200 edges (so all the retrieved edges are shown in this case) where, despite the emergence of some star-like structures centered around Atlanta (ATL), Minneapolis (MSP) and Dallas (DFW), there are almost no connections at all in the west of the country. Such a result is clearly undesirable, at it would imply no relevant connections for many US states at the chosen p-value, resulting in a heavily fragmented network. By contrast, the ECM filter displays a much more spread-out pattern consisting of several local hubs, not directly connected to each other. This corresponds to the so called *hub-and-spoke*, a well known structure observed in many spatial systems and indeed used in the airline system [8, 56]. The hub-and-spoke structure is usually the result of a design aimed at minimizing the operational cost, here emphasizing the role of regional hubs as Salt Lake City (SLC), Minneapolis (MSP), Portland (PDX), Charlotte (CLT) and St. Louis (STL), to name a few. In other words, the ECM filter uncovers the cost-oriented hub-and-spoke structure of US airports that is hidden within large-flow point-to-point patterns. Importantly, all US states are connected in the ECM backbone, making the resulting structure overall connected and hence much more acceptable in terms of transportation constraints.

The case of the International Trade Network (in the year 2011) exhibits trends similar to US airports. In particular, as reported in Fig. 4.7, the disparity backbone and the original network (panels b and a) look very alike, having China

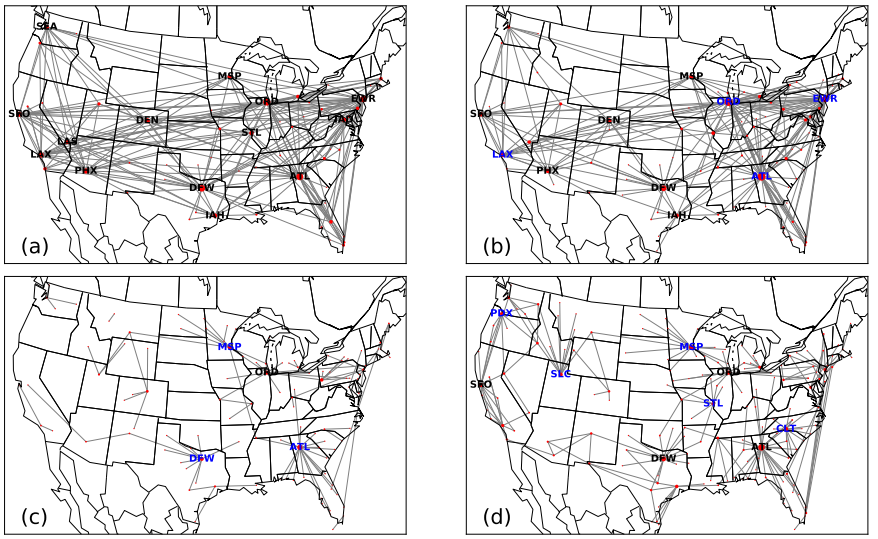

Figure 4.6: **Filtered backbones of the US airport network.** In each map, we display the top 200 heaviest connections for the original (a), disparity (b), GloSS (c) and ECM (d) networks. All the filtered backbones have been obtained using a common p-value equal to $\widetilde{\gamma} = 0.05$.

(CHN) and USA playing the role of global juggernauts since they embody together the 32.5% of all connections. We also notice the role of global broker/middleman played by Europe as well as the presence of members of G8 as Russia (RUS) and Japan (JPN), together with some G20 members like India (IND), South Korea (KOR), Brazil (BRA), South Africa (ZAF), Australia (AUS) and Indonesia (IDN). However, the complete absence of connections either within or towards African countries (except for South Africa) looks quite unrealistic. The backbone obtained using GloSS, albeit resembling the original one, looks more like a star with Europe at its center, in line with its geographical, political and technological role. Unfortunately, in the map depicted by GloSS it is hard to discriminate any local relationship between neighbouring countries which surely exists due to their tight related historical development. As in the case of airports, the scenario depicted by ECM (panel d) is rather different from the previous two and is the least predictable from the original network. Some of the features captured by disparity and GloSS can still be found in the ECM backbone, like the prominent role of USA, China and Europe on the global checkerboard. Others are captured by ECM only, and can thus be considered the hallmarks of ECM itself. We observe the emergence of many more "spheres of influence" characterized by an interaction pattern which is stronger with neighbouring countries in close analogy with what observed for airports. For example, Russia loses its global role and becomes

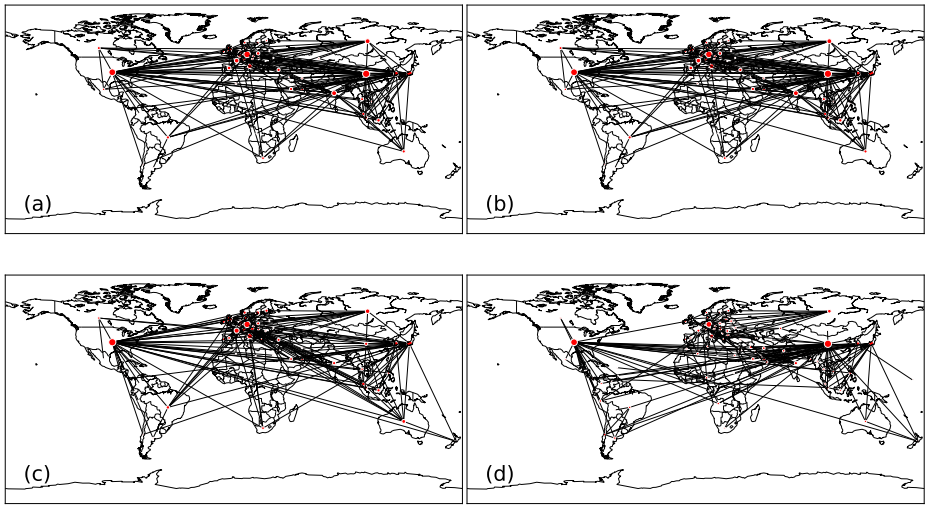

Figure 4.7: **Filtered backbones for the International Trade Networks network in year 2011.** In each map, we display the top 200 heaviest connections for the original (a), disparity (b), GloSS (c) and ECM (d) networks. All the filtered backbones have been obtained using a p-value, $\widetilde{\gamma}$, equal to 0.05.

almost exclusively a partner of European countries. Brazil has more connections with other South American countries. African countries other than South Africa like Nigeria (NGA), Angola (AGO) and North African countries such as Morocco (MAR) and Egypt (EGY) appear. Australia becomes more pivotal in the South Pacific. An unexpected trait highlighted by ECM is the brokering role between USA and China played by Middle Eastern countries like Saudi Arabia (SAU). Finally, Europe loses its role of global broker and becomes a more independent player.

Finally, we comment on the ability of ECM to identify relevant features *per se*. As an illustrative example, we consider the time evolution of the International Trade Network in the period 1998–2011 displayed in Fig. 4.8. In 1998, we can distinguish basically six "centers of influence". Two of them (namely USA and France (FRA)) act as global partners, while the other four, *i.e.* Russia (RUS), South Africa (ZAF), Australia (AUS) and Japan (JPN), appear instead to play a more "local" role. As time passes, we notice the rise of some countries and the fall of others. For example, around year 2002 we notice the growth of China (CHN), South Korea (KOR) and India (IND). In 2006 France (FRA) has considerably lost its original influence while New Zealand (NZL) plays a prominent role among the Pacific islands compartment (though showing connections which are less relevant in terms of exchanged volumes); moreover, China still exhibits a startling

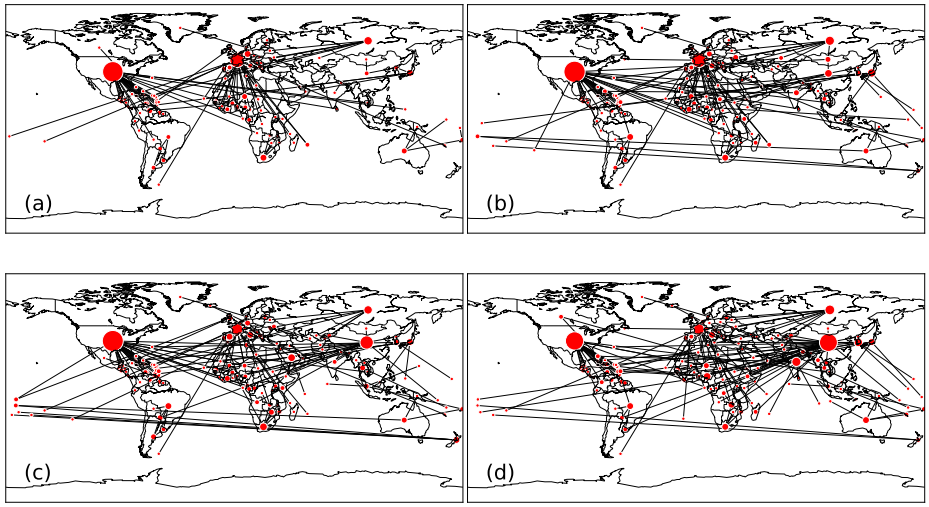

Figure 4.8: **Time evolution of the International Trade Network, filtered according to the local ECM filter.** Top left: 1998; top right: 2002; bottom left: 2006; bottom right: 2011. Figures refer to a critical p-value of 10^{-6}.

development. Finally, in 2011 China and USA appear to be equally influential centers of trade. In particular, China gains several connections with the African countries, to the detriment of France and other European countries.

4.5 Additional specifications of the method

We now illustrate how our method can be extended to different ensembles of weighted networks. For the sake of brevity, we only provide the mathematical expressions and do not show explicit empirical analyses.

4.5.1 Extension to directed networks

Let us consider the set $\mathcal{W}$ of *directed* weighted networks with N nodes, each of which is described by a $N \times N$ weight matrix $\boldsymbol{W}$ that is not necessarily symmetric and has non-negative integer entries. The constraints we impose are now the out-degree k_i^{out}, the in-degree k_i^{in} (defined as the number of out-going and in-coming links of node i, respectively), the out-strength s_i^{out} and the in-strength s_i^{in} (defined as the total out-going and in-coming weight of the links of node i, respectively). If we denote empirical values by asterisks, enforcing these constraints on average

means requiring

$$\langle \vec{k}^{\,\mathrm{out}} \rangle = \vec{k}^{\,\mathrm{out}*}, \quad \langle \vec{k}^{\,\mathrm{in}} \rangle = \vec{k}^{\,\mathrm{in}*}, \tag{4.21}$$

$$\langle \vec{s}^{\,\mathrm{out}} \rangle = \vec{s}^{\,\mathrm{out}*}, \quad \langle \vec{s}^{\,\mathrm{in}} \rangle = \vec{s}^{\,\mathrm{in}*}, \tag{4.22}$$

which results in the Hamiltonian

$$\begin{aligned} H \quad & (\boldsymbol{W} | \vec{x}^{\,\mathrm{out}}, \vec{x}^{\,\mathrm{in}}, \vec{y}^{\,\mathrm{out}}, \vec{y}^{\,\mathrm{in}}) = \\ = \quad & -\sum_{i=1}^{N} \sum_{j \neq i} \left[\Theta(w_{ij}) \ln(x_i^{\mathrm{out}} x_j^{\mathrm{in}}) + w_{ij} \ln(y_i^{\mathrm{out}} y_j^{\mathrm{in}}) \right], \end{aligned} \tag{4.23}$$

where x_i^{out}, x_i^{in}, y_i^{out}, y_i^{in} are Lagrange multipliers coupled to k_i^{out}, k_i^{in}, s_i^{out}, s_i^{in} respectively. A straightforward modification of the calculation we showed for the undirected case leads to the following results. The graph probability that maximizes the Shannon-Gibbs entropy subject to the above constraints is

$$P(\boldsymbol{W} | \vec{x}^{\,\mathrm{out}}, \vec{x}^{\,\mathrm{in}}, \vec{y}^{\,\mathrm{out}}, \vec{y}^{\,\mathrm{in}}) = \prod_{i=1}^{N} \prod_{j \neq i} q_{ij}(w_{ij}), \tag{4.24}$$

where

$$q_{ij}(w) = \begin{cases} 1 - p_{ij} & \text{if} \quad w = 0 \\ p_{ij} \left(y_i^{\mathrm{out}} y_j^{\mathrm{in}} \right)^{w-1} \left(1 - y_i^{\mathrm{out}} y_j^{\mathrm{in}} \right) & \text{if} \quad w > 0 \end{cases},$$

is the probability that the *directed* link from node i to node j has weight w, and

$$p_{ij} \equiv 1 - q_{ij}(0) = \frac{x_i^{\mathrm{out}} x_j^{\mathrm{in}} y_i^{\mathrm{out}} y_j^{\mathrm{in}}}{1 - y_i^{\mathrm{out}} y_j^{\mathrm{in}} + x_i^{\mathrm{out}} x_j^{\mathrm{in}} y_i^{\mathrm{out}} y_j^{\mathrm{in}}} \tag{4.25}$$

is the probability that a directed link from node i to node j exists, irrespective of its weight.

Given a real directed network $\boldsymbol{W}^*$, the values of the Lagrange multipliers are found by maximizing the log-likelihood

$$\ln P(\boldsymbol{W^*} | \vec{x}^{\,\mathrm{out}}, \vec{x}^{\,\mathrm{in}}, \vec{y}^{\,\mathrm{out}}, \vec{y}^{\,\mathrm{in}}) \tag{4.26}$$

or, equivalently, as the solution to the following $4N$ coupled equations:

$$\begin{aligned} k_i^{\,\mathrm{out}*} &= \sum_{j \neq i} \frac{x_i^{\mathrm{out}} x_j^{\mathrm{in}} y_i^{\mathrm{out}} y_j^{\mathrm{in}}}{1 - y_i^{\mathrm{out}} y_j^{\mathrm{in}} + x_i^{\mathrm{out}} x_j^{\mathrm{in}} y_i^{\mathrm{out}} y_j^{\mathrm{in}}} \\ k_i^{\,\mathrm{in}*} &= \sum_{j \neq i} \frac{x_j^{\mathrm{out}} x_i^{\mathrm{in}} y_j^{\mathrm{out}} y_i^{\mathrm{in}}}{1 - y_j^{\mathrm{out}} y_i^{\mathrm{in}} + x_j^{\mathrm{out}} x_i^{\mathrm{in}} y_j^{\mathrm{out}} y_i^{\mathrm{in}}} \\ s_i^{\,\mathrm{out}*} &= \sum_{j \neq i} \frac{x_i^{\mathrm{out}} x_j^{\mathrm{in}} y_i^{\mathrm{out}} y_j^{\mathrm{in}}}{(1 - y_i^{\mathrm{out}} y_j^{\mathrm{in}} + x_i^{\mathrm{out}} x_j^{\mathrm{in}} y_i^{\mathrm{out}} y_j^{\mathrm{in}})(1 - y_i^{\mathrm{out}} y_j^{\mathrm{in}})} \\ s_i^{\,\mathrm{in}*} &= \sum_{j \neq i} \frac{x_j^{\mathrm{out}} x_i^{\mathrm{in}} y_j^{\mathrm{out}} y_i^{\mathrm{in}}}{(1 - y_j^{\mathrm{out}} y_i^{\mathrm{in}} + x_j^{\mathrm{out}} x_i^{\mathrm{in}} y_j^{\mathrm{out}} y_i^{\mathrm{in}})(1 - y_j^{\mathrm{out}} y_i^{\mathrm{in}})} \end{aligned}$$

Once the parameter values are found, the p-value for the weight $w_{ij}^* > 0$ of the realized directed link from node i to node j reads

$$\gamma_{ij}^* \equiv \text{Prob}(w_{ij} \geq w_{ij}^*) = p_{ij}^* \left(y_i^{\text{out}*} y_j^{\text{in}*}\right)^{w_{ij}^*-1}. \tag{4.27}$$

As before, the local filter proceeds by retaining only the links for which the p-value γ_{ij}^* is smaller than a chosen critical value $\widetilde{\gamma}$. The global filter would employ a similar criterion based on the probability mass function, rather than on the cumulative probability function, but this is expected to lead to poorer results, as already discussed for the undirected case.

4.5.2 Extension to bipartite networks

We then assume that $\boldsymbol{W}^*$ is a *bipartite*, undirected, weighted network, and $\mathcal{W}$ the corresponding ensemble. Each network in the ensemble has two layers, one with N_1 nodes and one with N_2 nodes. Links are only allowed across layers, not within them. For each node i, one can still define the degree k_i and strength s_i as for an ordinary (*i.e.* unipartite) undirected graph. The main difference with respect to the unipartite case is the fact that all graphs $\boldsymbol{W}$ that do not have a bipartite structure are excluded from $\mathcal{W}$ and from the calculations.

There is, however, a trick that allows us to map (exactly) the ensemble of bipartite undirected graphs to the ensemble of unipartite directed graphs considered above. The trick consists in assigning an arbitrary but common direction (say, from layer 1 to layer 2) to all the links in the original bipartite network $\boldsymbol{W}^*$. Then, the resulting directed network can be treated as a unipartite one with $N = N_1 + N_2$ nodes and the procedure described above for directed networks can be applied. At the end, the direction of the links that are retained by the filter is simply discarded, and one correctly obtains the irreducible backbone of the original bipartite undirected network.

The above mapping between a bipartite undirected graph and a unipartite directed graph (and the corresponding null models) is exact because, after assigning a direction to the links, all nodes in (say) layer 1 have $k_i^{in} = 0$ and $k_i^{out} > 0$, while all nodes in (say) layer 2 have $k_i^{out} = 0$ and $k_i^{in} > 0$ (if we had $k_i^{in} = 0$ and $k_i^{out} = 0$, node i would be disconneted from all other nodes, and we would have discarded it). Similar conditions hold for the in- and out-strenghts. If we now apply the null model for weighted directed unipartite networks described in the previous subsection, the zero in- and out-degrees (and strengths) will be kept to zero, as there is no other way to enforce that their expected value is zero. This ensures that the nodes in each layer do not receive connections from other nodes in the same layer, so that the bipartite structure is preserved in the null model as desired.

4.6 Conclusions

The ever-increasing availability of 'big data' has spurred the use of networks as a powerful way to capture the relevant features of complex systems. However, when the flood of information becomes overwhelming, the advantages of a network representation tend to fade away and the possibility to discriminate the essential structure of the system (*i.e.* its backbone) deteriorates considerably. To preserve sparsity and non-redundancy of networks, several filtering techniques have been developed so far.

At the same time, recent improvements in network reconstruction techniques have emphasized that many structural features of real-world networks can be reliably estimated from the knowledge of the local node-specific topological properties, namely the degrees and strengths of nodes. This means that the truly dyadic relationships, *i.e.* those that are irreducible to node properties, may be hidden amidst a majority of redundant ones. In particular, recent results have shown that the "first-order approximation" for many networks with heterogeneous nodes is the ECM, while any other feature not directly encapsulated in the size of nodes (like higher-than-expected preference for specific connections, dependence on geographic or other distances, presence of communities and motifs, etc.) is expected to be immediately visible at the next order.

Based on the above considerations, we have introduced a method that filters out the first-order local effects embodied in the strength and degree sequence, thus highlighting the truly dyadic (and higher-order) patterns relating nodes to each other. We found that, while before applying the filter many networks display similar properties (precisely because their first-order structure is well approximated by the ECM), after applying the filter they show significant differences, presumably because higher-order features arise as network-specific effects.

Importantly, since strengths and degrees are the maximal set of local node-specific properties that can be defined in any weighted network, our approach is guaranteed to identify the connections that are by construction impossible to infer on the basis of node-specific properties alone and that cannot be recovered by any network reconstruction method based only on local node properties.

The comparison of the performances of the ECM, disparity and GloSS methods shows that the ECM filter outperforms its competitors. We have also examined the structural differences between the backbones retrieved from the global and local implementations of our filter and the role of the order of filtering and aggregating in multiplex networks. We have applied the ECM filter to the analysis of several empirical datasets and illustrated how successfully it extracts relevant hidden features like the hub-and-spoke structure of the US airport network and the evolution in time of the most relevant "spheres of influence" across world trade.

Finally, we have shown that the ECM filter can be applied to different kinds of weighted networks (e.g. undirected, directed, bipartite) and therefore constitutes a valuable tool for the analysis of any networked system where the excess of information hinders the identification of the essential backbone of interactions.

We believe that the approach introduced in this chapter advances significantly the state of the art in the field of graph filtering by creating a new paradigm based on irreducibility to the output of network reconstruction, thereby pushing the field of graph filtering into a promising, yet unexplored direction.

Appendix

The present Appendix contains the results that have not been displayed in Chapter 4, grouped by datasets.

4.A World Time-varying Trade Network

Considering the trading volumes between countries in the period between the years 1998 and 2011 we can build a time-varying network where each time snapshot corresponds to a given year. A node indicates a country and the weight of a link denotes the gross trade volume between two countries. In Fig. 4.9 we show the filtered graphs for the year 2011. Panel (a) is the local filter case obtained considering p-value, $\widetilde{\gamma}$, equal to 10^{-6}. Panel (b) is the global case obtained choosing the first L' least likely links such that the number of edges in the two networks is the same. At a glance, we see that most of the significant connections are shared by both graphs.

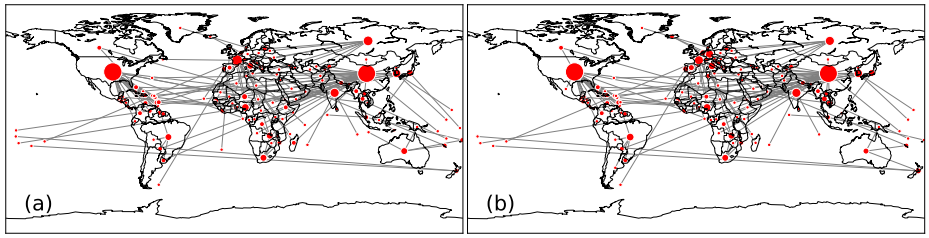

Figure 4.9: **2011 World Trade Network filtered using local (a) and global (b) ECM filters.** The local network is obtained using $\widetilde{\gamma} = 10^{-6}$. The number of links, L, in both graphs is 149. The size of the nodes is proportional to their degree.

In Tab. 4.3 we report the list of the twenty most significant links according to local ECM, global ECM, GloSS and Disparity filters for year 1998. The names of the countries are represented using the ISO 3166-1 alpha-3 standard encoding. Since we are looking at the first twenty most significant links out of about 200, it is reasonable that the four lists are pretty much similar. Interestingly, the most significant connection in GloSS and Disparity Filter (DF) (*i.e.* that between Canada (CAN) and USA (USA)) is not even present among the most significative

of ECM filters. DF, instead, fails to identify the links between France (FRA) and French Polynesia (PYF) and also between Belgium-Luxemburg (BLX) and Central African Republic (CAF) to give an example. More in general, DF tends to assign an eccessive relevance to USA placing fifteen out of twenty connections with USA in the list while for the other methods such number drops to just half of the connections. Finally, it is worth mentioning that with the sole exception of DF, we are not able to find trace of China in 1998 table in accordance with the not predominant role played by such country at that time. In Tab. 4.4 we

Rank	Local filter	Global filter	GloSS	Disparity
1	DNK GRL	DNK GRL	CAN USA	CAN USA
2	JAM USA	JAM USA	DNK GRL	MEX USA
3	HND USA	HND USA	TCA USA	BLR RUS
4	HTI USA	BLR RUS	COM FRA	DOM USA
5	VGB RUS	DOM USA	HTI USA	HND USA
6	BLR RUS	HTI USA	VGB RUS	AUT DEU
7	FRA PYF	VGB RUS	BLX CAF	CRI USA
8	DOM USA	FRA PYF	JAM USA	CZE DEU
9	BLX CAF	CRI USA	FRA PYF	JAM USA
10	FRA MDG	GAB USA	CPV PRT	VGB RUS
11	GAB USA	FRA MDG	AND ESP	JPN USA
12	AND ESP	AND ESP	FRA MDG	COL USA
13	CRI USA	BLX CAF	HND USA	HTI USA
14	ALB ITA	ALB ITA	GRD USA	GTM USA
15	TCA USA	GTM USA	ALB ITA	GAB USA
16	CPV PRT	NIC USA	BLR RUS	IRL GBR
17	COM FRA	BHS USA	GAB USA	CHN USA
18	NIC USA	ANT VEN	FRA WLF	ECU USA
19	BHS USA	CPV PRT	CRI USA	ISR USA
20	GTM USA	TTO USA	DOM USA	TCA USA

Table 4.3: **Most significant links of the World Trade Web in 1998.** List of the twenty most significant connections in the International Trade Network for year 1998 according to the local and global ECM filters, GloSS and Disparity Filter.

list the same kind of information displayed in Tab. 4.3 but for year 2011. At first glance, something catches our attention, namely the presence of the following "*bizarre*" connections: Antigua and Barbuda (ATG) and Nigeria (NGA), Algeria (DZA) and Saint Kitts and Nevis (KNA), Barbados (BRB) and Nigeria, Turks and Caicos Islands (TCA) and USA. All methods indicate those connections as relevant. However, a deeper analysis of the data revealed the presence of errors in

the records of trade volumes among such countries. Some of the endpoints of these anomalous connections are very small Caribbean or Pacific countries, showing sheer trade volumes (for certain commodities) higher than those between China and USA for example. In the light of these findings, filtering can be thought of not exclusively as a way to recognize relevant connections, but also as a method to validate them. Finally, contrary to the Disparity Filter, ECM and GloSS identify additional wrong entries in the dataset such as: Bermuda (BMU) and South Korea, South Korea and Liberia (LBR), Nigeria and Niue (NIU) and Cocos (Keeling) Islands (CCK) and India (IND). We have checked for the presence of such anomalous connections across all our time-varying data, and we have found that such mistakes are present only in the years 2009, 2010 and 2011. Among other noticeable - and meaningful - connections spotted by ECM filter, instead, we find: South Africa (ZAF) and Zimbawe (ZWE), Denmark (DNK) and Greenland (GRL), China (CHN) and Mongolia (MNG), Albania (ALB) and Italy (ITA) just to cite a few.

The scenario becomes more interesting by looking at Tab. 4.5, i.e. the list of the twenty heaviest links according to our global and local methods. Here, in fact, we can see how the link between China (CHN) and USA (USA), which is the heaviest in the original network, has disappeared from both filtered networks. Conversely, the relation between Russian Federation (RUS) and Ukraine (UKR) clearly emerges in the filtered networks as one of the most important ones. Another curious feature is the vanishing of Germany (DEU) from the column of local filter albeit it appears in ten out of twenty positions available in the original networks. Finally, we observe the presence of a link between Italy (ITA) and Libya (LBY) which have a strong historic and economic relation due to the past role of Libya as one of the colonies of Italy during the beginning of the 20^{th} century.

4.B World Trade Multiplex Network

The results displayed in Figures 4.10 - 4.12 show that the ECM filter can be useful to detect patterns in the International Trade Multiplex, namely the multilayer network where each node denotes a country and each layer represents the trade in a given commodity, that is one of the main focuses of this thesis. In particular, here we consider the year 2011. The original layers do not exhibit any evident difference among each other, due to the large density of this disaggregated representation; the filtered ones provide, instead, some relevant information, such as the appearance of Norway and Russia as hubs, respectively in the trade in fish and fuels/oil.

4.C US Airport Network

In the US airport network, the first result reported in Fig. 4.13 is the behaviour of the fractions of nodes, edges and total weight as a function of the p-value. As

Rank	Local filter	Global filter	GloSS	Disparity
1	ATG NGA	ATG NGA	ATG NGA	ATG NGA
2	CHN PRK	CHN PRK	CHN PRK	MEX USA
3	DZA KNA	DZA KNA	NGA NIU	CAN USA
4	NPL IND	NPL IND	BRA LCA	TCA USA
5	BRB NGA	BRB NGA	DZA KNA	NPL IND
6	TCA USA	TCA USA	NPL IND	BRB NGA
7	AND ESP	AND ESP	AND ESP	BLR RUS
8	ZAF ZWE	ZAF ZWE	BRB NGA	DOM USA
9	BRA LCA	BRA LCA	CCK IND	TCD USA
10	DOM USA	DOM USA	DNK GRL	AND ESP
11	DNK GRL	ABW USA	TCD USA	AUT DEU
12	TCD USA	TCD USA	ZAF ZWE	HND USA
13	ABW USA	DNK GRL	BTN IND	ABW USA
14	ALB ITA	HND USA	DOM USA	GTM USA
15	HND USA	ALB ITA	COM FRA	CRI USA
16	JAM USA	CHN SDN	PRT STP	ALB ITA
17	CHN SDN	JAM USA	ALB ITA	CZE DEU
18	CHN MNG	CHN MNG	ABW USA	BTN IND
19	BTN IND	KOR LBR	COK NZL	COL USA
20	BMU KOR	BTN IND	JAM USA	SLV USA

Table 4.4: **Most significant links of the World Trade Web in 2011.** List of the twenty most significant connections in the International Trade Network for year 2011 according to the local and global ECM filters, GloSS and Disparity filter. Green cells correspond to connections displaying false volumes. Blue (orange) cells correspond to erroneous connections identified only by ECM (GloSS) filter.

we can see, for $\widetilde{\gamma} = 0.05$ the filter is able to remove around 70% of the connections while retaining about 20% of the total information (i.e. total weight). In Fig. 4.14 we display the pruned networks (according to both the versions of the ECM filtered) with all their links. Despite being more "noisy", the difference between the structures of local and global networks remains clearly distinguishable. In particular, we observe the persistence of many links among principal aiports due to their heavy weights. The absence of such links in the local network enables the emergence of the hub-and-spoke structure mentioned in the main text.

In addition to the visual comparison between local and global filters, we propose the comparison between the twenty most significant edges (Tab. 4.6). The most striking fact about the names listed in this table is that they correspond mainly to very small towns. The only exception is link number 17 in the global network corresponding to the connection between Miami and Key West which is

Rank	Original	Local	Global
1	CHN USA	RUS UKR	CAN USA
2	CHN JPN	USA VEN	MEX USA
3	CAN USA	BLR RUS	AUT DEU
4	MEX USA	COL USA	RUS UKR
5	CHN KOR	AGO CHN	USA VEN
6	FRA DEU	JPN PAN	DEU HUN
7	CHN DEU	CHN OMN	BLR RUS
8	DEU NLD	ECU USA	COL USA
9	JPN USA	LTU RUS	ARG BRA
10	DEU ITA	CRI USA	JPN QAT
11	BLX NLD	AZE ITA	AGO CHN
12	DEU GBR	GTM USA	PRT ESP
13	BLX DEU	CHN SDN	JPN PAN
14	DEU USA	FRA TUN	CHN AMN
15	DEU CHE	HND USA	ECU USA
16	AUS CHN	KOR LBR	LTU RUS
17	JPN KOR	TTO USA	AUS NZL
18	AUT DEU	KOR MHL	CRI USA
19	BLX FRA	ITA LBY	DEU SVN
20	DEU POL	CHN MNG	AZE ITA

Table 4.5: **Heaviest links in World Trade Web in 2011.** List of the twenty heaviest connections in the International Trade Network (2011) in the original network, and according to the local and global ECM filters.

very important, among many factors, for tourism. This is completely different from what can be seen in the main text (where the links were ranked according to the weight of the connections, rather than according to the p-value), where all the connections listed are among main airports with New York and Los Angeles playing prominent roles in both the global and the local network.

4.D Florida Bay Food Web

The analysis of the filtering power of local ECM on the Florida Bay dry dataset on the usual three topological indicators displays a behaviour not very different from the other cases. However, for $\widetilde{\gamma} = 0.05$ we observe a slightly higher value of $\frac{W}{W_0}$ than for airports accompanied by a steeper decrease of if for lower values of $\widetilde{\gamma}$.

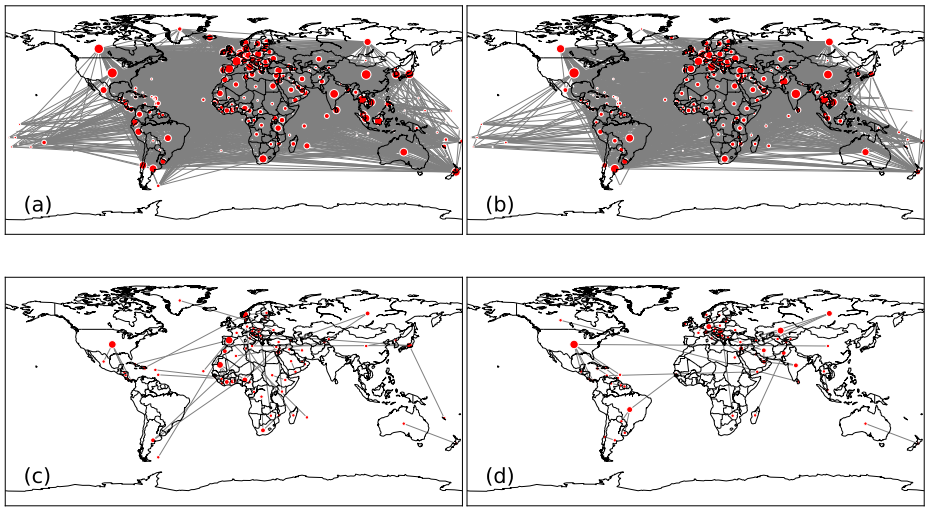

Figure 4.10: **Original and ECM filtered trade multiplex of fish and cereals.** Panels a-c refer to the original (a) and local ECM filter (c) fish and crustaceans commodity. Panels b-d account for cereals, instead. The original networks are built using the 2011 data, and have been filtered considering $\widetilde{\gamma} = 10^{-5}$.

4.E Star Wars

We conclude our portfolio of datasets with the Star Wars movie saga one. The effects of filter aggressivity on topological quantities shown in Fig. 4.16 are in line with similar results for other datasets. However the amount of retained information for $\widetilde{\gamma} = 0.05$ is much higher than any other case. This is probably due to the fact that these networks are already very sparse and therefore the statistical significance of their links is high. Considering the full interactions dataset, the visual inspection (Fig. 4.17) of the original and filtered networks permits to identify those characters playing a key role. In particular, the centrality of Darth Vader, Luke Skywalker and Obi-Wan Kenobi clearly increases while for other characters like C3P0 and Jar Jar Binks this is the opposite, showing once more the usefulness of the ECM filter.

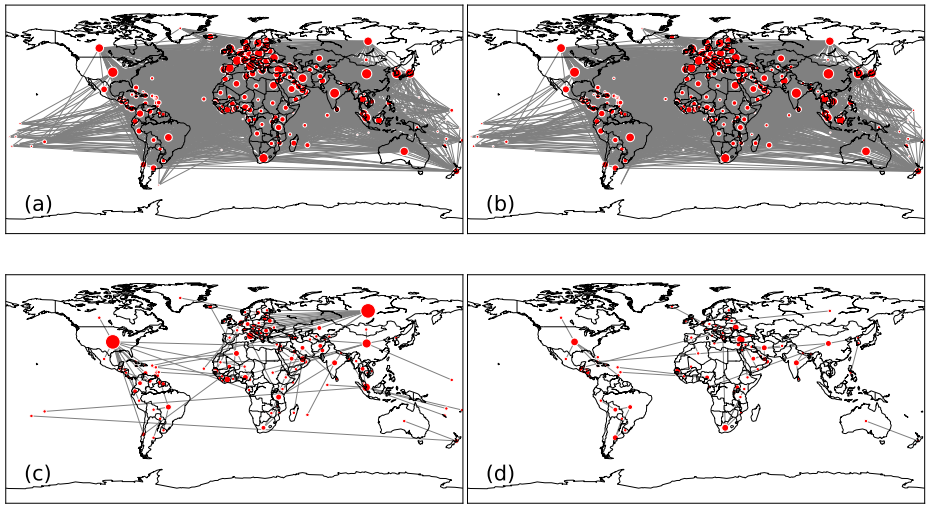

Figure 4.11: **Original and ECM filtered trade multiplex of fuels/oils and iron/steel.** Original (panels a-b) and ECM filtered (panels c-d) trade multiplex of fuels and oils (panels a-c) and iron and steel (panels b-d). The original networks are built using the 2011 data, and have been filtered considering $\widetilde{\gamma} = 10^{-5}$.

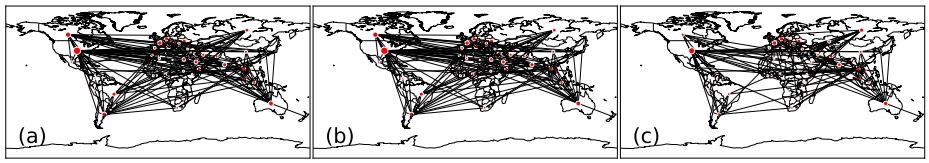

Figure 4.12: **Visual representation of the 200 heaviest links in the original network (a), ECM global (b) and local (c) filters for the 2011 Trade in cereals.** The local filtering is obtained using $\widetilde{\gamma} = 0.05$, the global one is constructed such that $L_{GLOB} = L_{LOC}$.

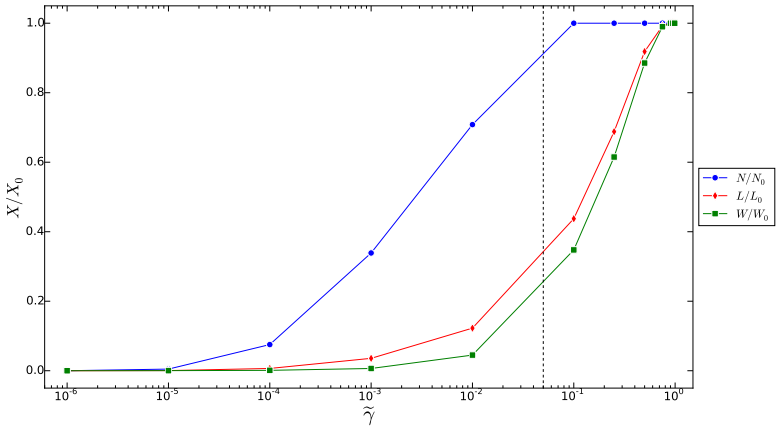

Figure 4.13: **Effect of filtering on the US-airport dataset.** We report the fraction of the number of nodes N/N_0 (blue dots), edges L/L_0 (red diamonds) and total weight W/W_0 (green squares) as a function of the p-value $\widetilde{\gamma}$. The vertical line corresponds to $\widetilde{\gamma} = 0.05$.

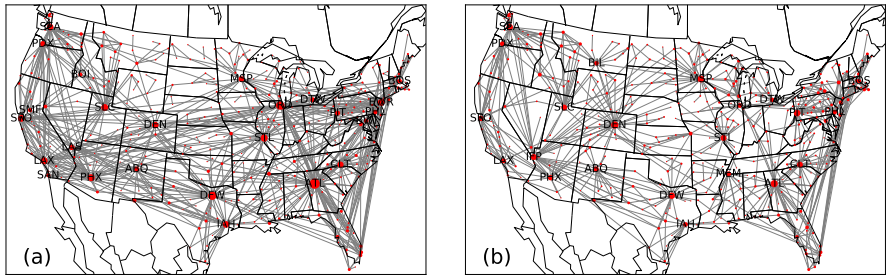

Figure 4.14: **Graphs obtained after filtering according to the global (left) and local (right) ECM method, for the US Airport Network.** In the local filter, results refer to $\widetilde{\gamma} = 0.05$; in the global one, we choose the most unlikely links such that the number of edges in the two panels is the same (in this case, 764 links).

Rank	Local filter	Global filter
1	Nantucket - Hyannis	Nantucket - Hyannis
2	Bedford - Trenton	Bedford - Trenton
3	Fort Dodge - Mason City	Fort Dodge - Mason City
4	Alpena - Sault Ste Marie	Alpena - Sault Ste Marie
5	Devils Lake - Jamestown	Lewiston - Pullman
6	Hot Springs - Harrison	Spokane - Seattle
7	Denver - Bullhead City	Eau Claire - Rhinelander
8	Kingman - Prescott	Devils Lake - Jamestown
9	Brookings - Huron	Hancock - Marquette
10	Melbourne - Oshkosh	Hot Springs - Harrison
11	Hancock - Marquette	Columbus Starkville WestPt - Tupelo
12	El Dorado - Jonesboro	Springfield - Quincy
13	Eau Claire - Rhinelander	Grand Rapids - Saint Cloud
14	Havre - Lewistown	Kingman - Prescott
15	Grand Rapids - Saint Cloud	Friday Harbor - Lopez Island
16	Clovis - Hobbs	Idaho Falls - Pocatello
17	Riverton - Worland	Key West - Miami
18	Lewiston - Pullman	Brookings - Huron
19	North Platte - Norfolk	El Dorado - Jonesboro
20	Manhattan - Salina	Melbourne - Oshkosh

Table 4.6: **Most significant links of the US Airport Network.** List of the 20 most significant connections in the US Airport Network according to the local and global ECM filters. The local network is filtered considering $\widetilde{\gamma} = 0.05$.

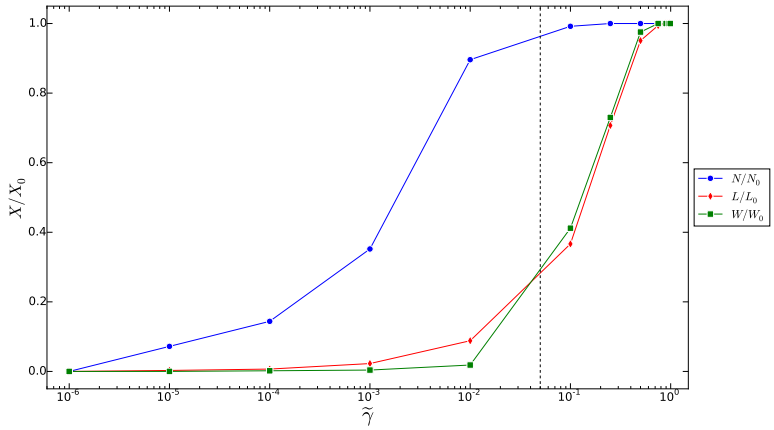

Figure 4.15: **Effect of filtering on the Florida Bay food web dataset.** We report the fraction of the number of nodes N/N_0 (blue dots), edges L/L_0 (red diamonds) and total weight W/W_0 (green squares) as a function of the p-value $\widetilde{\gamma}$. The vertical line corresponds to $\widetilde{\gamma} = 0.05$.

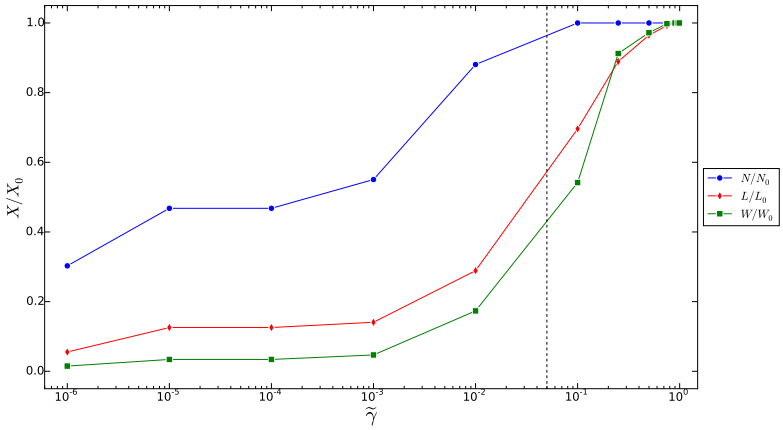

Figure 4.16: **Effect of filtering on the Star Wars interactions network.** We report the fraction of the number of nodes N/N_0 (blue dots), edges L/L_0 (red diamonds) and total weight W/W_0 (green squares) as a function of the p-value $\widetilde{\gamma}$. The vertical line corresponds to $\widetilde{\gamma} = 0.05$.

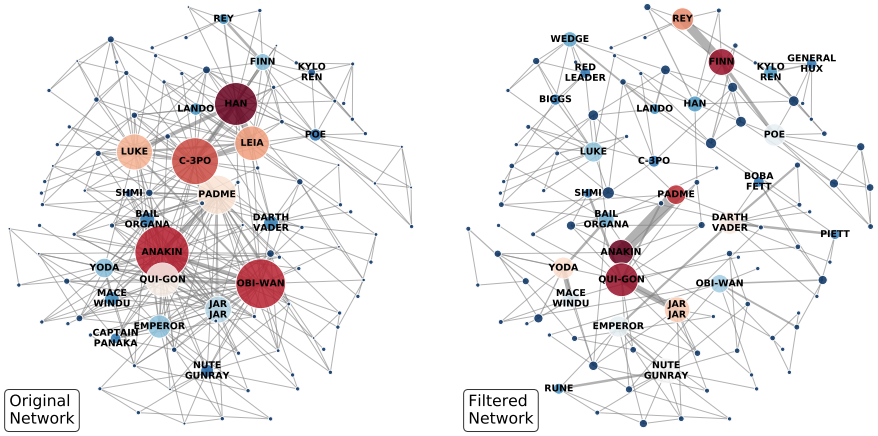

Figure 4.17: **Visual representation of the original and filtered Star Wars network.** Original graph showing the interactions among Star Wars characters (left) and corresponding pruned graph (right), according to the local ECM filter with p-value equal to $\widetilde{\gamma} = 0.05$. The size of the nodes is proportional to their degrees. We highlight also the most important connections.

Bibliography

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

[2] M. E. J. Newman (2010) 'Networks', Oxford University Press

[3] A.-L. Barabási (2011) 'The network takeover', *Nature Physics* **8**, 14

[4] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, Y. Aberg (2001) 'The web of human sexual contacts', *Nature* **411**, 907

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

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

[7] R. Guimerá, B. Uzzi, J. Spiro, L. A. N. Amaral (2005) 'Team assembly mechanisms determine collaboration network structure and team performance', *Science* **308**, 697

[8] M. Barthelemy (2011) 'Spatial networks', *Physics Reports* **499** (1), 1

[9] A. Barrat, M. Barthelemy, R. Pastor-Satorras, A. Vespignani (2004) 'The architecture of complex weighted networks', *Proceedings of the National Academy of Sciences USA* **101** (11), 3747

[10] M. S. Granovetter (1973) 'The strength of weak ties', *American Journal of Sociology* **78**, 1360

[11] C. Lynch (2008) 'Big data: how do your data grow?', *Nature* **455**, 28

[12] V. D. Blondel, A. Decuyper, G. Krings (2015) 'A survey of results on mobile phone datasets analysis', *EPJ Data Science* **4**, 10

[13] J. P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási (2007) 'Structure and tie strengths in mobile communication networks', *Proceedings of the National academy of Sciences USA* **104** (18), 7332

[14] J. Menche, A. Sharma, M. Kitsak, S. D. Ghiassian, M. Vidal, J. Loscalzo, A.-L. Barabási (2015) 'Uncovering disease-disease relationships through the incomplete interactome', *Science* **347**, 1257601

[15] S. A. Myers, J. Leskovec (2014) 'The bursty dynamics of the Twitter information network', *Proceedings of the 23rd International Conference on World Wide Web*, 913

[16] T. Squartini *et al.* , in preparation

[17] T. Squartini, D. Garlaschelli, Springer, forthcoming

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

[19] 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

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

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

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

[23] D. Garlaschelli, M. I. Loffredo (2009) 'Generalized Bose-Fermi statistics and structural correlations in weighted networks', *Physical Review Letters* **102** (3), 038701

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

[25] T. Squartini, G. Fagiolo, D. Garlaschelli (2011) 'Null models of economic networks: the case of the world trade web', *Journal of Economic Interaction and Coordination* **8** (1), 75

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

[27] B. J. Kim (2004) 'Geographical coarse graining of complex networks', *Physical Review Letters* **93** (16), 168701

[28] D. Gfeller, P. De Los Rios (2007) 'Spectral coarse graining of complex networks', *Physical Review Letters* **99** (3), 38701

[29] M. Hamann, G. Lindner, H. Meyerhenke, C. L. Staudt, D. Wagner (2016) 'Structure-preserving sparsification methods for social networks', *Social Network Analysis and Mining* **6**, 22

[30] A. Gionis, P. Rozenshtein, N. Tatti, E. Terzi (2017) 'Community-aware networks sparsification', *arXiv*:1701.07221

[31] M. Coscia, F. Neffke (2017) 'Network backboning with noisy data', *Proceedings of the 33rd International Conference on Data Engineering (ICDE)*, 425

[32] F. De Vico Fallani, V. Latora, M. Chavez (2017) 'A topological criterion for filtering information in complex brain networks', *PloS Computational Biology* **13** (1), e1005305

[33] J. B. Kruskal (1956) 'On the shortest spanning subtree of a graph and the travelling salesman problem', *Proceedings of the American Mathematical Society* **7**, 48

[34] P. J. Macdonald, E. Almas, A.-L. Barabási (2005) 'Minimum spanning trees of weighted scale-free networks', *Europhysics Letters* **72** (2), 308

[35] S. Scellato, A. Cardillo, V. Latora, S. Porta (2006) 'The backbone of a city', *European Physics Journal B* **50**, 221–225

[36] M. Tumminello, S. Micciché, F. Lillo, J. Piilo, R. N. Mantegna (2011) 'Statistically validated networks in bipartite complex systems', *PloS ONE* **6** (3), e17994

[37] M. A Serrano, M. Boguñá, A. Vespignani (2009) 'Extracting the multiscale backbone of complex weighted networks', *Proceedings of the National Academy of Sciences USA* **106** (16), 6483

[38] F. Radicchi, J. J. Ramasco, S. Fortunato (2011) 'Information filtering in complex weighted networks', *Physical Review E* **83** (4), 046101

[39] N. Dianati (2016) 'Unwinding the hairball graph: pruning algorithms for weighted complex networks', *Physical Review E* **93** (1), 012304

[40] T. Squartini, R. Mastrandrea, D. Garlaschelli (2105) 'Unbiased sampling of network ensembles', *New Journal of Physics* **17**, 023052

[41] T. Squartini, D. de Mol, F. den Hollander, D. Garlaschelli (2015) 'Breaking of ensemble equivalence in networks', *Phisical Review Letters* **115** (26), 268701

[42] D. Garlaschelli, F. den Hollander, A. Roccaverde (2016) 'Ensemble nonequivalence in random graphs with modular structure', *Journal of Physics A: Mathematical and Theoretical* **50** (1), 015001

[43] T. Opsahl, V. Colizza, P. Panzarasa, J. J. Ramasco (2008) 'Prominence and control: the weighted rich-club effect', *Physical Review Letters* **101** (16), 168702

[44] J. Park, M. E. J. Newman (2004) 'Statistical mechanics of networks', *Physical Review E* **70** (6), 66117

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

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

[47] MAX & SAM package. Available at: `http://www.mathworks.it/matlabcentral/fileexchange/46912-max-sam-package-zip`

[48] `http://drive.google.com/drive/folders/0B_rBKSwFTur3M0tvd0w4dW45aE0`, `http://drive.google.com/open?id=0B_rBKSwFTur3M0tvd0w4dW45aE0&authuser=0`

[49] R. E. Ulanovicz, C. Bondavalli, M. S. Egnotovich (1998) 'Network analysis of trophic dynamics in South Florida ecosystem, FY 97: the Florida Bay Ecosystem', available at: `https://networkdata.ics.uci.edu/netdata/html/floridaFoodWebs.html`

[50] E. Gabasova (2015) 'The Star Wars social network', Evelina Gabasova's Blog. Data available at: `https://github.com/evelinag/StarWars-social-network/tree/master/networks`

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

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

[53] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, M. Zanin (2014) 'The structure and dynamics of multilayer networks', *Physics Reports* **544** (1), 1

[54] J. Liebig, A. Rao (2016) 'Fast extraction of the backbone of projected bipartite networks to aid community detection', *Europhysics Letters* **113** (2), 28003

[55] S. Fortunato (2010) 'Community detection in graphs', *Physics Reports* **486** (3), 75

[56] D. L. Bryan, M. E. O'Kelly (1999) 'Hub-and-spoke networks in air transportation: an analytical review', *Journal of Regional Science* **39**, 275

[57] P. Jaccard (1902) 'Lois de distribution florale dans la zone alpine', *Bulletin de la Socieéteé Vaudoise des Sciences Naturelles* **38**, 69

[58] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, A. Arenas (2013) 'Mathematical formulation of multilayer networks', *Physical Review X* **3** (4), 41022

[59] 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

[60] M. De Domenico, V. Nicosia, A. Arenas, V. Latora (2015) 'Structural reducibility of multilayer networks', *Nature Communications* **6**, 6864

[61] A. Barrat, M. Barthelemy, A. Vespignani (2008) 'Dynamical processes on complex networks', Cambridge University Press, Cambridge

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

[63] D. Stauffer, A. Aharony (1994) 'Introduction to percolation theory', CRC Press, Taylor & Francis

[64] Federal Aviation Administration, Passenger Boarding (Enplanement) and All-Cargo Data for U.S. Airports. Available at: `https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/previous_years/#2002`

Chapter 5

Scientific publications network

Community detection techniques are widely used to infer hidden structures within interconnected systems. Despite demonstrating high accuracy on benchmarks, they reproduce the external classification for many real-world systems with a significant level of discrepancy. A widely accepted reason behind such an outcome is the unavoidable loss of non-topological information (such as node attributes) encountered when the original complex system is converted into a network. In this chapter we systematically show that the observed discrepancies may also be caused by a different reason: the external classification itself. For this end we use scientific publication data which i) exhibit a well defined modular structure and ii) hold an expert-made classification of research articles. Having represented the articles and the extracted scientific concepts both as a bipartite network and as its unipartite projection, we applied modularity optimization to uncover the inner thematic structure. The resulting clusters are shown to partly reflect the author-made classification, although some significant discrepancies are observed. A detailed analysis of these discrepancies shows that they may carry essential information about the system, mainly related to the use of similar techniques and methods across different (sub)disciplines, that is otherwise omitted when only the external classification is considered.

The results presented in this chapter have been published in the following reference:
V. Palchykov, V. Gemmetto, A. Boyarsky, D. Garlaschelli, *EPJ Data Science*, **5**, 28 (2016).

5.1 Introduction

A conflict between two members of a relatively small university organization that happened more than 40 years ago [1] has attracted a lot of attention in the scientific community so far [2]. A confrontation during the conflict resulted in a fission of the organization, known as Zachary's karate club, into two smaller groups, gathered around the president and the instructor of the club, respectively. Predicting the sizes and compositions of the resulting factions, given the structure of the social interaction network before the split, attracted a lot of attention. This puzzle, supplemented by the known outcome, makes this system among the best studied benchmarks to test community detection algorithms [3]. Having verified a high level performance on the aforementioned system and on other benchmarks [4], community detection algorithms have then been massively applied to uncover tightly connected modules within large real-world systems. This allowed scientists to identify, for instance, Flemish- and French-speaking communities in Belgium using mobile phone communication networks [5], detect functional regions in the human or animal brain from neural connectivity [6], observe the emergence of scientific disciplines [7] and investigate the evolution of science using citation patterns and article metadata [8, 9, 10].

A bird's eye view on the identified clusters in real-world systems certifies their meaningfulness. However, an in-depth quantitative validation of the community structure requires its comparison with an external classification of the nodes, which is accessible only for a limited number of large systems. Examples include crowd-sourced tag assignments for software packages [11], product categories for Amazon copurchasing networks [12], declared group membership for various online social networks [13, 14] and publication venues for coauthorship networks in the computer science literature [13]. Surprisingly, significant discrepancies have been identified between the extracted grouping of nodes and their external classification for these systems [11, 15]. This message remains robust independently of the system under investigation and the technique used to uncover its community structure, and calls for a detailed inspection of such discrepancies in order to understand the reasons behind them.

One of the possible reasons concerns the strong simplification that occurs during the projection of the original complex system into a network. This projection may omit some crucial information that cannot be encoded into the structural connection pattern [11]. The missing information may correspond to age or gender of individuals in social networks [16, 17] or geographical position of the nodes within spatially embedded systems [18]. Following this direction, several algorithms [19, 20] have been developed in order to handle specific nodes attributes, beside the usual connectivity patterns. Such approaches have been shown to identify groups of nodes that more closely reproduce the external classification in real-world systems [20] than the techniques that rely on the connectivity patterns only.

In this chapter we argue that, independently of the aforementioned issue, the

supposedly poor performance of community detection algorithms may be caused by the external classification itself and its misinterpretation. For instance, a system may possess several alternative classification schemes, such as thematic and methodological groupings in a system of scientific publications or in academic coauthorship networks [21]. In such situation, the discrepancies between the community detection results and a single accessible classification (e.g. based on thematic similarity) may carry, instead, meaningful information (e.g. about methodological similarity), therefore providing an added value to the system understanding.

Here we explore this idea by performing a detailed analysis of a scientific publication record system. This system may be simplified into a structural network representation, where the nodes correspond to scientific articles, and the links represent the relationship between them. There are various possibilities to map these relationships: direct citation [22], cocitation and bibliographic coupling [23] or content related similarities [24, 25]. In this chapter we focus on the latter, considering scientific terms or concepts that appear within the articles. Performing community detection on the corresponding network, we compare the results with an expert made classification of these articles, considering both similarities and discrepancies between the two different partitions. Then we investigate the main reasons causing the most notable deviations.

This chapter is organized as follows. In the section 5.2 we present the dataset used; in sec. 5.3 we introduce the methodology used to build the networks, extract the partitions and compare them with the external classification. Finally, in sections 5.4 and 5.5 we present our findings and discuss them.

5.2 Data

We investigate a collection of scientific manuscripts submitted to e-print repository `arXiv` [26] during the years 2013 and 2014. During the submission process, the authors were requested to classify the manuscript according to the `arXiv` classification scheme by assigning at least one category to it. In our analysis we are focussed only on the articles that have been assigned to a single category, restricting ourself to the field of physics. Moreover, the collections of manuscripts submitted during the years 2013 and 2014 will be considered separately, eliminating the possible issues related to the temporal evolution of research disciplines. The resulting datasets consist of 36386 articles submitted during 2013 and 41848 articles submitted during 2014, and will be referred below (together with the extracted contents) as the `arxivPhys2013` and `arxivPhys2014` datasets, respectively. The numbers of articles belonging to each category are shown in Tab. 5.1.

Each article is represented by a set of scientific concepts that characterize its content, i.e. specific words or combinations of them. The concepts have been identified within the full text by the `ScienceWISE.info` platform (`SW`). `SW`

category	n^{s}_{2013}	n^{m}_{2013}	n^{s}_{2014}	n^{m}_{2014}
`nucl-th`	648	1628	766	1210
`nucl-ex`	315	924	324	736
`hep-ph`	2625	3935	3116	2885
`hep-ex`	602	1726	706	1225
`hep-lat`	352	695	419	417
`hep-th`	1787	3717	2316	2960
`gr-qc`	1118	2782	1527	2204
`astro-ph`	10984	3023	11445	2437
`physics`	4452	6479	5711	4880
`cond-mat`	10549	4609	11397	3538
`nlin`	392	327	522	905
`quant-ph`	2558	3240	3187	2471
`math-ph`	0	3789	412	2668

Table 5.1: **Distribution of articles among categories.** The number of manuscript submitted during the year y that have been assigned to a given category only ($n^{\mathrm{s}}_{\mathrm{y}}$) or to the category and at least one another ($n^{\mathrm{m}}_{\mathrm{y}}$). List of categories: theoretical and experimental nuclear physics (`nucl-th` and `nucl-ex`, respectively), four branches of high energy physics (`hep-ph`: phenomenology, `hep-ex`: experiment, `hep-lat`: lattice and `hep-th`: theory), general relativity and quantum cosmology (`gr-qc`), astrophysics (`astro-ph`), physics (`physics`), condensed matter physics (`cond-mat`), nonlinear science (`nlin`), quantum physics (`quant-ph`) and mathematical physics (`math-ph`).

is a web service connected to the main online repositories such as `arXiv`, whose peculiarity is a bottom-up approach in the management of scientific concepts [27]. The initially created scientific ontology was followed by a continuous editing by the users, for instance by adding new concepts, definitions and relationships. This crowd-sourced procedure leads to the most comprehensive vocabulary of scientific concepts in the domain of physics. Such vocabulary takes care of synonyms that refer to the same concepts and it includes physics concepts explicitely labeled as generic like `mass` or `energy`, or more specific ones like `community detection`. Both are the results of crowd-sourcing by the registered expert-users.

The number k of concepts significantly vary among the manuscripts, reaching up to $k_{\max} \sim 400$ for review articles. The average number of identified concepts $\langle k \rangle$ per article, together with some other characteristics of the datasets `arxivPhys2013` and `arxivPhys2014`, are shown in Tab. 5.2.

	N	V	V_{gen}	$\langle k \rangle$	L_{idf}	L_{bp}
`arxivPhys2013`	36386	12200	347	37	3.3×10^8	1.3×10^6
`arxivPhys2014`	41848	12728	344	38	4.5×10^8	1.6×10^6

Table 5.2: **Basic characteristics of the datasets.** Total number of articles (N), total number of identified concepts (V) and the number of generic ones (V_{gen}) among them; $\langle k \rangle$ gives the average number of non-generic concepts within arbitrary chosen article. The number of links in a unipartite network (provided that the generic concepts are excluded) L_{idf} is two orders of magnitude larger than the corresponding number of links in bipartite networks (L_{bp})[1]. This results in significant differences in computational resources needed to perform community detection analysis.

5.3 Methods

The dataset may be represented as a network, whose nodes correspond to articles. Two nodes i and j are connected by a link if the corresponding articles share at least a single common concept. The resulting networks are extremely dense, covering almost 90% of all possible network connections; this number may be reduced to 50% if the generic concepts are ignored (see Tab.5.2). Below, to save the computational resources, we will ignore the generic concepts in our analysis. The weight of the link between two manuscripts is designed to reflect the level of content similarity between two articles, i.e. the overlap between the respective lists of concepts. Different concepts, however, may contribute differently to the similarity among two articles. Indeed, sharing a widely used concept should affect the similarity between two articles differently than sharing a specific one, suggesting that specific concepts should have a higher impact on the similarity. Each concept c in the dataset is therefore weighted according to its occurrence, which may be accounted for by the so-called $idf(c)$ factor [28]:

$$idf(c) = \log \frac{N}{N(c)}. \tag{5.1}$$

Here N is the total number of articles and $N(c)$ is the number of articles that contain concept c. As mentioned above, among the V concepts identified by `SW`, we will consider only the specific ones, discarding the V_{gen} generic concepts. The content of each article can be therefore expressed by means of a $(V - V_{\text{gen}})$-dimensional concept vector $\vec{v}_i$. The element v_{ic} of the concept vector of the article i has non-zero value equal to $idf(c)$ only if the concept c appears within the article i and equals zero otherwise.

The similarity between the contents of two articles i and j, and the link weight w_{ij} between the corresponding nodes, may then be estimated by the cosine simi-

[1] These represent, in all the cases, roughly the 60% of all the links, i.e. including also the contribution given by the generic concepts.

larity between the two concept vectors $\vec{v}_i$ and $\vec{v}_j$ as follows:

$$w_{ij} = \frac{\vec{v}_i \cdot \vec{v}_j}{|\vec{v}_i||\vec{v}_j|}. \tag{5.2}$$

The resulting network will be referred below as the `idf` representation of the data.

Alternatively to `idf` representation, the dataset may be mapped into a bipartite network. Such network consists of the nodes of two types that correspond to manuscripts and scientific concepts, respectively. The unweighted links in the simplest case reflect the appearance of a concept within the article. This network will be referred below to as a `bp` representation of the data, and the usage of the two alternative representation will serve the robustness of our results. The number of links ($L_{\rm idf}$, $L_{\rm bp}$) of these networks are shown in Tab. 5.2. As one may see, the number of links in `bp` representation is about two orders of magnitude smaller than the number of links in the corresponding `idf` representation. This has significant consequences on the run-time and memory used to analyse the networks.

Indeed, the run-time t of the Louvain algorithm scales with the number of links L of the considered network. Since empirically in the bipartite representation $L_{bp} \sim O(N)$ while in the unipartite case $L_{idf} \sim O(N^2)$, this reflects in much different computational resources required to perform the community detection. Moreover, here we point out that the bipartite representation is the most natural and suitable characterization of the dataset, since the null model behind such representation of the data is definitely more correct. In fact, the bipartite null model is consistent with the constraints on both the types of node (number of papers per concept and concepts per article). This feature is instead lost when the system is projected into a unipartite network, since the previous constraints are not matched any more. Furthermore, the bipartite representation and null model already take into account the presence of more frequent concepts, sparing us the use of any `idf` factor. In this context, we therefore propose the use of the bipartite representation as a possible alternative to the more widespread `idf` (or `tf-idf`) unipartite representation.

In order to find a unipartite network partition, we will maximize a modularity function [29]. To deal with bipartite networks, we adopt a co-clustering approach [30] and Barber's generalization of modularity [31].

In both cases, we assume that each article may belong to a single cluster only, hence exploiting the notion of non-overlapping communities. Furthermore, the co-clustering approach makes stronger restrictions on a bipartite partition, compared to a unipartite one. Indeed, the resulting clusters of a bipartite partition consist of both articles and related concepts, and we assume that each concept belongs to a single cluster as well. Such restriction may be relaxed, for instance by using alternative ways to generalize modularity for bipartite network [32] or by employing stochastic block model techniques [33]. However, we will consider co-clustering of bipartite networks since it allows us to straightforwardly employ the same greedy optimization algorithm [5] for the networks of both types.

The restriction towards a single algorithm is also caused by the result [11] that i) the selected algorithm is among the ones that perform best on real-world networks and ii) the major influence on the accuracy is related to the dataset itself rather than the algorithm. Due to the stochastic origin of this algorithm, it has been applied 100 times for unipartite networks and 1000 times for bipartite ones (due to the significantly different number of links and, therefore, the required computational resources). Among the detected partitions, for each network we will select the single partition that corresponds to the highest value of modularity; this partition will be referred below as the optimal partition for each network.

5.4 Results

A partition of a bipartite network consists of clusters that contain both articles and scientific terms (concepts), while clusters of a unipartite network partition consist of articles only. To compare both unipartite and bipartite partitions with the external article classification, we will be focussed only on the articles that fall into each cluster. Thus, by referring below to a cluster of bipartite partition we mean the set of articles that belong to the specified cluster. In this perspective, the external classification of the articles is represented by the `arXiv` standard split into different subject classes or categories (`astro-ph`, `cond-mat`, etc.).

Then, given two partitions P and Q of the same network (for instance a detected network partition and the `arXiv` classification), an initial comparison between them has been performed using an information-based symmetrically normalized mutual information:

$$I_{\mathrm{N}}(P,Q) = \frac{2I(P,Q)}{H(P)+H(Q)}. \tag{5.3}$$

Here $I(P,Q)$ is the mutual information [34] between two partitions P and Q, and $H(P)$ is the entropy of partition P. The normalized mutual information $I_{\mathrm{N}}(P,Q)$ may vary between 0 and 1. A value of 0 indicates that the two partitions have no information in common, while a value of 1 corresponds to identical partitions. In Tab. 5.3 we show the level of similarity between each optimal partition and the `arXiv` classification ones. The reported values of normalized mutual information indicate the existence of some common information between automatically identified clusters of articles (both in the bipartite and unipartite cases) and the author based classification. However, the values being quite far from the possible maximum of 1 reflect evidence for some discrepancies between the partitions. Below we perform a detailed analysis of these discrepancies. Here we will show the results for the `arxivPhys2013` dataset; similar findings can be observed in the `arxivPhys2014` case and are shown in the following appendix.

The first difference is observed in the numbers of detected clusters and of `arXiv` subject classes: while the number of categories in the `arXiv` classification

	idf	bp
arxivPhys2013	0.60 ± 0.02	0.56 ± 0.03
arxivPhys2014	0.55 ± 0.00	0.54 ± 0.02

Table 5.3: **Similarity between network partitions and external classification.** Average value of the normalized mutual information I_{N} (5.3) between a partition of each network representation and arXiv classification of the articles and the corresponding standard deviations. Both bp and idf partitions demonstrate similar value of closeness to arXiv classification.

scheme is 12 [2], the number of clusters in our partitions is only equal to 4 in the idf and to 6 in the bp network representations, respectively[3]. Indeed, the articles of some different arXiv categories tend to belong to a single cluster. This may be clearly observed in Fig. 5.1 that shows the fraction of articles of each arXiv category belonging to each cluster in the resulting partitions. This merger is especially visible for different high energy physics (hep) categories (hep-ph, hep-ex, hep-lat and hep-th): in the idf partition, almost 99% of all these articles fell into a single cluster, independently of the sub-field. This result, despite deviating from the arXiv classification scheme, is reasonable since we observe a union of almost all papers about high energy physics, no matter if they deal with experimental or theoretical issues.

Instead, in the bp partition the articles of the four hep categories are almost entirely distributed among two clusters, focussed on experimental and theoretical issues, respectively. The first of them joins 95% of all articles that belong to experimental categories (hep-ph, hep-ex or hep-lat), while the second one contains 94% of all theoretical (hep-th) articles. Thus, the presence of more clusters within the bipartite network partition allows us to identify methodologically different clusters of articles within the hep categories, in particular dividing theoretical papers from experimental ones.

Even though the split of hep articles into two groups may be simply explained by the different approaches used to study the phenomena, a further result can be observed from Fig. 5.1: in the bipartite network partition, hep-th articles tend to form a single cluster with the articles that belong to general relativity and quantum cosmology (category gr-qc) rather than with the other high energy physics articles, thus appearing to be more similar to gr-qc papers rather than to the other hep ones. Intuitively, indeed, we know that both hep-th and gr-qc both focus mostly on general relativity, while the other hep categories focus on particle physics [4].

[2] In fact, there are 13 physics categories in arXiv classification scheme, but there is no single article in arxivPhys2013 dataset that belong to math-ph category only.

[3] By performing a detailed comparison we ignore all single-node clusters, which contain the articles for which no concept has been identified.

[4] Indeed, it is very likely that nowadays the hep- categories would be split in multiple subcategories (namely hep-th, hep-lat, etc.). However, here we point out that our study (in particular in the

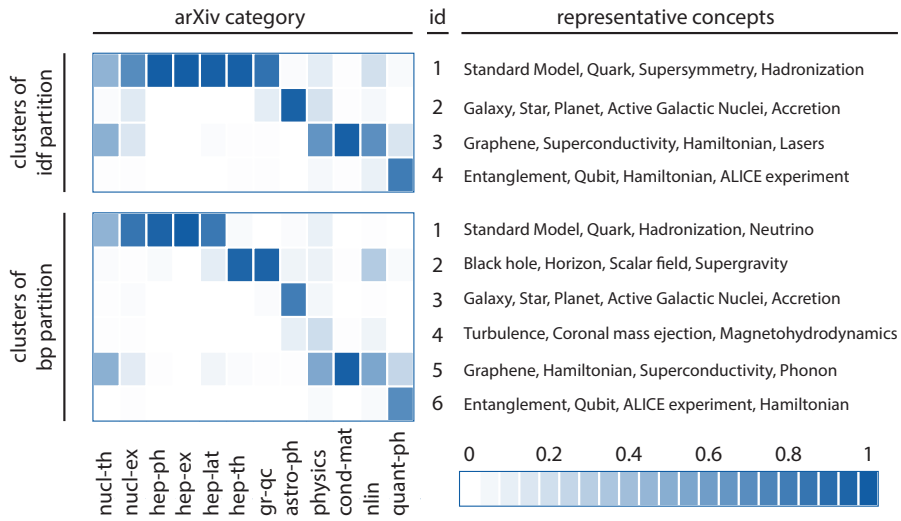

Figure 5.1: **Inner composition of arxivPhysics2013 partitions.** The color of each cell accounts for the fraction of articles of a given category belonging to a cluster (each column sums to 1). The articles of the same categories tend to incorporate into single clusters as justified by the clearly visible block-diagonal structure of both `idf` and `bp` partitions. Nevertheless, the split of some categories into distinct clusters may be observed. For instance, the articles of `nucl-th` category are roughly equally split among `hep-` and `cond-mat`-dominated categories. On the right, the most representative concepts for each cluster are shown.

Such a relatedness between the articles of the two theoretical physics categories (`hep-th` and `gr-qc`) may be verified independently by a category co-occurrence analysis. To show this, we will use the complementary part of the investigated dataset. This set consists of all articles that have been submitted to `arXiv` during the same 2013 year, but for which the authors have assigned at least two different categories. Thus, no article of this set overlaps with the clustered `arxivPhys2013` collection. Irrespective of the details of the decision-making process through which authors assign multiple categories, this multiplicity reflects the authors' decision that the scope of the article can not be properly covered by a single category of a given classification scheme. Whilst several categories may cover the scope of a single research article, the co-occurrence of the same two categories in a significant fraction of articles may reflect some hidden relationships between them. The corresponding empirical co-occurrence matrix is shown in Fig. 5.2 and indicates

bipartite case) shows that `hep-th` looks actually more similar to `gr-qc` than to the other `hep-` classes. This therefore seems to strengthen the apparently counterintuitive choice of dividing the high energy articles in different primary classes.

the fraction of articles of a given category that have been co-submitted to the other categories. The diagonal elements of this matrix indicate the fraction of articles of each category that have been assigned to a single category by the author(s), i.e. the articles of the `arxivPhys2013` dataset. A normalization procedure has been performed such that each column of the matrix sums to 1.

Fig. 5.2 confirms that the `hep-th` subject class is indeed more related to the `gr-qc` class than to the other `hep` categories: `hep-th` co-occurred with `gr-qc` in 1721 articles, and with all other `hep` categories in only 1286 articles, even though the number of the corresponding `hep` papers (`hep-ph`, `hep-ex`, `hep-lat`) exceeds the number of `gr-qc` ones threefold. This high level of relatedness between `hep-th` and `gr-qc` categories justifies the merging of the articles of these categories into a single cluster and indicates the meaningful deviation from the `arXiv` classification scheme. It is worth to mention that in the `idf` partition, where all `hep` category articles tend to belong to a single cluster, the same cluster is supplemented by 87% of all `gr-qc` articles, in agreement with the result observed above. Moreover such a tendency is not restricted to the dataset for the selected year: it has also been observed for the `arxivPhys2014` one (as shown in the appendix).

The same approach explains the presence of a significant fraction of `physics`, non-linear (`nlin`) and quantum physics (`quant-ph`) articles into the `cond-mat` clusters. It also allows us to understand a possible reason why nuclear physics articles (both theory and experiment) occur significantly within the `hep` clusters. However, it cannot explain the presence of roughly one half of `nucl-th` articles into the condensed matter cluster (cluster No. 3 in `idf` and No. 5 in `bp` partitions) in both network representations. The latter deviation from the article classification, which is not explained by category co-occurrence, does not exclude that similarities between these topics exist but are considered not strong enough by the authors to label the articles with both subject classes. To uncover the possible essence of these similarities, we examine the top representative concepts that characterize the `nucl-th` articles that belong to the two different clusters, see Table 5.4. In both cases, the top representative concepts contain the ones that characterize the object of investigation within theoretical nuclear physics, such as `Isotope`, `Isospin` or `Nuclear matter`. However, one may clearly identify method-related concepts, such as `Hartree-Fock`, `Hamiltonian`, `Mean field` and `Random phase approximation`, among the top representative concepts of articles in the `cond-mat` cluster. These concepts clearly characterize methods that are widely used in condensed matter physics research, and that have not been identified among top concepts in any other cluster. This result emphasizes the ability of scientific concepts found within research articles to highlight not only topics focussed on the same objects, but also methodologically similar research directions.

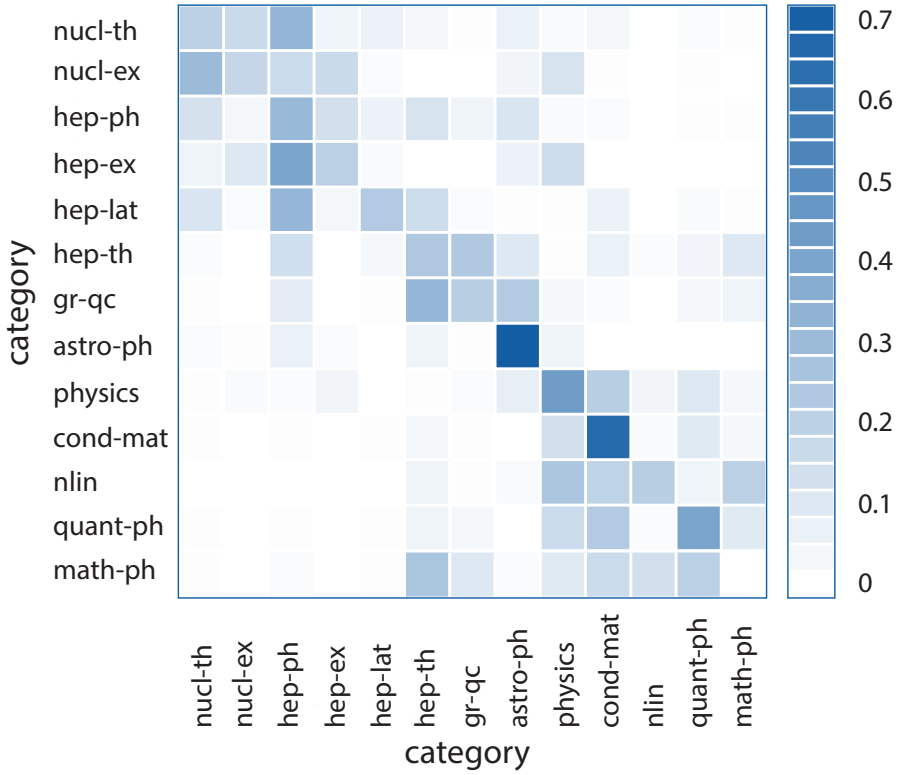

Figure 5.2: **Co-occurrence matrix of `arXiv` categories during year 2013.** Built on the complementary dataset to `arxivPhys2013`, this matrix reflects the relationships between `arXiv` categories and allows to justify the meaningfulness of some remarkable discrepancies, like the merger of `hep-th` and `gr-qc` articles. Each non-diagonal element reflects the fraction of articles in which two specified categories have co-occurred. The diagonal cells represent the fractions of articles that have been assigned to a single category, i.e. they concern the articles of the `arxivPhys2013` dataset. A normalization procedure has been performed such that each row of the matrix sums to 1. Thus, the aforementioned fractions correspond to the fractions of manuscripts that have been labeled with a given category.

%	Concept (cluster no. 1)	%	Concept (cluster no. 3)
43	Hadronization	55	Isotope
39	Isospin	53	Hamiltonian
37	Pion	39	Hartree-Fock
33	Degree of freedom	36	Quadrupole
32	Heavy ion collision	34	Isospin
31	Quark	31	Nuclear matter
29	Chirality	30	Degree of freedom
29	Hamiltonian	28	Mean field
29	Nuclear matter	26	Harmonic oscillator
26	Coupling constant	25	Spin orbit

Table 5.4: **Representative concepts of two groups of articles categorized as `nucl-th`.** The left side of the table represents the group of articles that fell into the `hep` dominated cluster (no. 1) in `idf` partition. The right side – the other group: the `nucl-th` articles that fell into the `cond-mat` dominated cluster (no. 3). For each group, the numbers next to the concepts give the percentage of articles in which the concept has been identified. The table allows us to make a suggestion that the two groups of articles significantly differ by the methods used to investigate nuclear matter.

5.5 Conclusions

The differences between the outcomes of community detection algorithms and possible external classifications may have various reasons. The most notable of them concern a possible failure of the considered algorithm or the unavoidable loss of data about real complex systems determined by their representation as networks. To deal with the first issue, algorithms are heavily tested on benchmarks, while the second issue is still under investigation [20]. In this chapter, we emphasize a third possible reason behind such discrepancies, i.e. the fact that the external classification itself may possess its own limitations. For this reason we performed a detailed investigation of a scientific publication system which i) may be naturally represented as a network and ii) owns an external author-made classification of scientific articles. While, indeed, some discrepancies are caused by the lack of data (for instance in the case of the articles for which no concept has been identified), we argue that the most remarkable of them may reflect real commonalities across different subject classes. Academic publications are traditionally categorized and classified[5] according to objects or phenomena under investigation. The same phenomena, however, may be explored using various approaches, experimental observation and theoretical modeling being among them.

[5]Document classification and categorization are different processes: classification refers to the assignment of one or more predefined categories to a document, while categorization refers to the process of dividing the set of documents into priory unknown groups whose members are in some way similar to each other [35].

On the other hand, the phenomena that belong to different research topics may be investigated using the same methods, composing the core of the interdisciplinary research. Thus, a more comprehensive classification or research articles may be represented by a two layer categorization scheme, where one layer reflects phenomena or objects while the other one stands for the methods of investigation. Usually, these two layers are not taken equally into account. The expert made classification may include rather a strong bias towards the object layer. The reasons involve the classification scheme itself and the limited knowledge about all other research disciplines that employ the same methods. Instead, automatic concept-based categorization has no direct preference for any of the layers: the extracted concepts correspond both to phenomena and methods, and the algorithm has no information about the possible division of the concepts. Thus, the observed discrepancies may reflect the dominance of the methodological layer over the other one, which corresponds to phenomena or objects. Similar results have been previously observed within the collaboration network of scientists at Santa Fe Institute [21], where, besides the expected grouping around common topics, some methodologically driven clusters have been observed.

This shows that the failure in reproducing an external classification may indicate a genuinely more complicated organization within the system, in addition to the lack of data or algorithmic mistakes. Besides developing sophisticated algorithms to deal with real systems, we should therefore keep in mind that some observed discrepancies may go beyond the standard classification and carry important information about the system under study. We believe that similar results may be observed in other systems. Indeed, the ground truth necessarily follows from a given classification criterion; however, the considered data may contain more than that single type of information (perhaps in conflict one with each other). In general, therefore, it may happen that what we consider as the ground truth is just one of the possible reference points, rather than some absolute truth. Understanding the information employed to define the so-called ground truth is therefore crucial in order to perform a proper comparison between external classification and automatically retrieved communities.

Appendix

5.A Scientific publications network in 2014

Here we show the results of the community detection algorithm to the so-called `arxivPhys2014` dataset, representing the content-relations between 41848 scientific articles that have been assigned to a single physics category, submitted to `arXiv` in 2014; our findings are reported in Figure 5.3 (top panel). The partitions obtained through the Louvain algorithm are very similar to those observed for the `arxivPhys2013` dataset: we see that, also in this case, the manuscripts belonging to the same category tend to merge into single clusters as illustrated

by the block-diagonal structure of both `idf` and `bp` clusterings. Still, the split of some categories into different communities may be observed, such as `nucl-th` and `math-ph`.

Furthermore, we can justify our results based on the co-occurrence matrix reported in Figure 5.3 (bottom panel). This matrix, built on the complementary dataset of `arxivPhys2014`, namely the set of articles showing more than one physics category, reflects the relations between the various `arXiv` categories in 2014 and can therefore explain the reason of some of the observed discrepancies, such as the union of `hep-th` and `gr-qc` manuscripts.

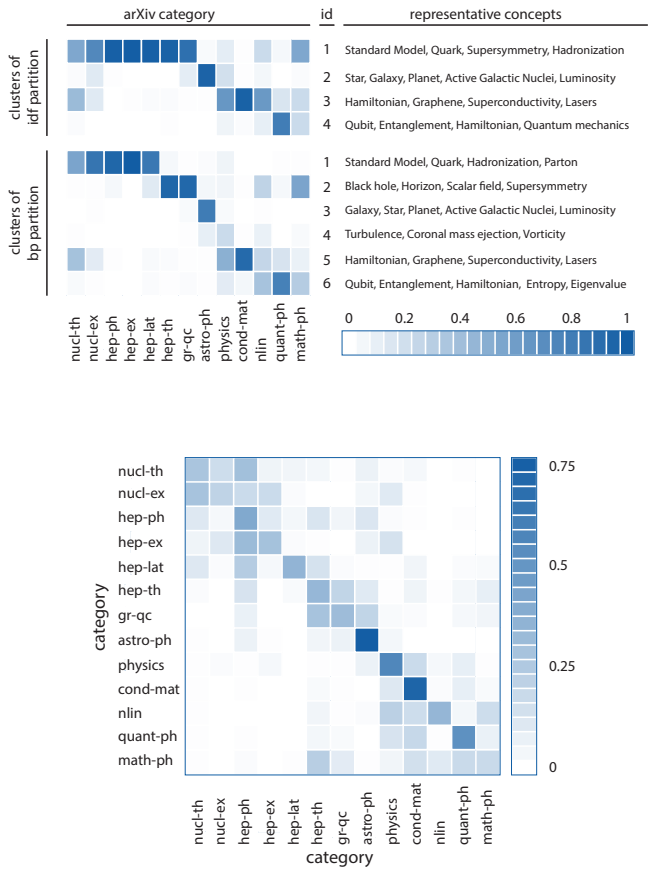

Figure 5.3: **Results of the analysis for the `arxivPhys2014` dataset.** Top: inner composition of the obtained partitions. The color of each cell accounts for the fraction of articles of a given category belonging to a cluster (each column sums to 1); the articles of the same categories tend to incorporate into single clusters as justified by the clearly visible block-diagonal structure of both `idf` and `bp` partitions. Bottom: co-occurrence matrix of `arXiv` categories during year 2014. Each non-diagonal element reflects the fraction of articles in which two specified categories have co-occurred; the diagonal cells represent the fractions of articles that have been assigned a single category, i.e. they concern the articles of the `arxivPhys2014` dataset. A normalization procedure has been performed such that each row of the matrix sums to 1.

Bibliography

[1] W. W. Zachary (1977) 'An information flow model for conflict and fission in small groups', *Journal of Anthropological research*, 452

[2] M. E. J. Newman (2012) 'Communities, modules and large-scale structure in networks', *Nature Physics* **8** (1), 25

[3] S. Fortunato (2010) 'Community detection in graphs', *Physics Reports* **486** (3), 75

[4] A. Lancichinetti, S. Fortunato, F. Radicchi (2008) 'Bechmark graphs for testing community detection algorithms', *Physical Review E* **78** (4), 046110

[5] V. Blondel, J.-L. Guillame, R. Lambiotte, E. Lefebvre (2008) 'Fast unfolding of communities in large networks', *Journal of Statistical Mechanics: Theory and Experiment* **2008** (10), P10008

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

[7] N. Shibata, Y. Kajikawa, Y. Takeda, K. Matsushima (2008) 'Detecting emerging research fronts based on topological measures in citation networks of scientific publications', *Technovation* **28** (11), 758

[8] M. Herrera, D. C. Roberts, N. Gulbahce (2010) 'Mapping the evolution of scientific fields', *PloS ONE* **5** (5), e10355

[9] M. Rosvall, C. T. Bergstrom (2010) 'Mapping change in large networks', *PloS ONE* **5** (1), e8694

[10] P. Chen, S. Redner (2010) 'Community structure of the physical review citation network', *Journal of Informetrics* **4** (3), 278

[11] D. Hric, R. K. Darst, S. Fortunato (2014) 'Community detection in networks: structural communities versus ground truth', *Physical Review E* **90** (6), 062805

[12] J. Leskovec, L. A. Adamic, B. A. Huberman (2007) 'The dynamics of viral marketing', *ACM Transactions on the Web (TWEB)* **1** (1), 5

[13] L. Backstrom, D. Huttenlocher, J. Kleinberg, X. Lan (2006) 'Group formation in large social networks: membership, growth, and evolution', *Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining*, 44

[14] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, B. Bhattacharjee (2007) 'Measurements and analysis of online social networks', *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 29

[15] J. Yang, J. Leskovec (2015) 'Defining and evaluating network communities based on ground-truth', *Knowledge and Information Systems* **42** (1), 181

[16] V. Palchykov, K. Kaski, J. Kertész, A. L. Barabási, R. I. M. Dunbar (2012) 'Sex differences in intimate relationships', *Scientific Reports* **2**, 370

[17] L. Kovanen, K. Kaski, J. Kertész, J. Saramäki (2013) 'Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences', *Proceedings of the National Academy of Sciences USA* **110** (45), 18070

[18] P. Expert, T. Evans, V. Blondel, R. Lambiotte (2011) 'Uncovering space-independent communities in spatial networks', *Proceedings of the National Academy of Sciences USA* **108** (19), 7663

[19] C. Bothorel, J. D. Cruz, M. Magnani, B. Micenkova (2015) 'Clustering attributed graphs: models, measures and methods', *Network Science* **3** (3), 408

[20] M. E. J. Newman, A. Clauset (2016) 'Structure and inference in annotated networks', *Nature Communications* **7**, 11863

[21] M. Girvan, M. E. J. Newman (2002) 'Community structure in social and biological networks', *Proceedings of the National Academy of Sciences USA* **99** (12), 7821

[22] L. Waltman, N. J. Van Eck (2012) 'A new methodology for constructing a publication-level classification system of science', *Journal of the American Society for Information Science and Technology* **63** (12), 2378

[23] K. W. Boyack, R. Klavans (2010) 'Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately?', *Journal of the American Society for Information Science and Technology* **61** (12), 2389

[24] K. W. Boyack, D. Newman, R. J. Duhon, R. Klavans, M. Patek, J. R. Biberstine, B. Schijvenaars, A. Skupin, N. Ma, K. Börner (2011) 'Clustering more than two million biomedical publications: comparing the accuracies of nine text-based similarity approaches', *PloS ONE* **6** (3), e18029

[25] P. Glenisson, W. Glänzel, F. Janssens, B. De Moor (2005) 'Combining full text and bibliometric information in mapping scientific disciplines', *Information Processing & Management* **41** (6), 1548

[26] An electronic archive and distribution server for research articles, `http://arxiv.org`

[27] R. Prokofyev, G. Demartini, A. Boyarsky, O. Ruchayskiy, P. Cudré-Mauroux (2013) 'Ontology-based word sense disambiguation for scientific literature', *Advances in Information Retrieval*, 594

[28] K. S. Jones (1973) 'Index term weighting', *Information Storage and Retrieval* **9** (11), 619

[29] M. E. J. Newman, M. Girvan (2004) 'Finding and evaluating community structure in networks', *Physical Review E* **69** (2), 026113

[30] D. M. Blei, A. Y. Ng, M. I. Jordan (2003) 'Latent Dirichlet Allocation', *Journal of Machine Learning Research* **3**, 993

[31] M. J. Barber (2007) 'Modularity and community detection in bipartite networks', *Physical Review E* **76** (6), 066102

[32] R. Guimerá, M. Sales-Pardo, L. A. N. Amaral (2007) 'Module identification in bipartite and directed networks', *Physical Review E* **76** (3), 036102

[33] D. B. Larremore, A. Clauset, A. Z. Jacobs (2014) 'Efficiently inferring community structure in bipartite networks', *Physical Review E* **90** (1), 012805

[34] M. Meilă (2007) 'Comparing clusters – an information based distance', *Journal of Multivariate Analysis* **98** (5), 873

[35] E. K. Jacob (2004) 'Classification and categorization: a difference that makes a difference', Graduate School of Library and Information Science. University of Illinois at Urbana-Champaign

Concluding remarks

In this thesis, we have focused on multi-layer complex networks, developing various maximum-entropy models and showing their application in the extraction of relevant patterns of several real systems. We have therefore exploited concepts stemming from theoretical and statistical physics, such as entropy and partition functions, to design null models for networked systems as canonical ensembles with specified constraints.

The definition of random benchmarks in terms of canonical ensembles is by no means new, as it has already been introduced at the beginning of the 21th century for single-layer networks and extended to the multiplex case a few years later. Our main contribution, clearly illustrated in this work, consisted in the fit of such maximum-entropy models to real-world systems via the so-called maximum-likelihood method. This allowed us to deeply analyze systems composed by a large number of layers without incurring into significant computational limitations.

We have pointed out that the aforementioned models can be employed for different - and sometimes even opposite - purposes, ranging from the phenomenological modelling of observed networks to the statistical inference and data filtration. In this context, we have indeed shown that they are able to inform us about the genuine correlations between layers of a multiplex; the flexibility of these models was also exhibited by their application to directed and weighted graphs, as illustrated in Chapters 1 and 2. Furthermore, in Chapter 3 we have pointed out that such metrics can overcome the problem of limitedness of topological information, leading to the design of new multiplex reconstruction methods able to infer the inter-layer topology from partial information. We have then highlighted that, strikingly, the same models that can be used for the previous inference problems may also be employed for the opposite task, namely the data filtration. This observation led us to the development of an original and successful graph pruning method (Chapter 4). In conclusion, we have also focused on a scientific publications system, thanks to the collaboration with the ScienceWISE platform, that allowed us to connect scientific manuscripts based on their content. In Chapter 5 we have shown that this system can be effectively tackled in the network theory framework; a better comprehension of this system, involving the whole scientific community, may come from the inclusion of other layers of interactions, such as adding the information about the citations between articles. These analyses can

therefore provide insights into the global scientific landscape.

As previously stated, one of the cornerstones of the thesis was the focus on real-world systems. Indeed, we have tested all our new metrics and models to observed networks, ranging from the economic sector to the infrastructural one. Moreover, we have highlighted that a better understanding of some of these systems is strictly connected to the use of the multiplex approach, as clearly shown for instance in the case of the World Trade Network. This approach can therefore provide a significant added value to the usual "monoplex" network theory.

Our findings point out once more the power of the maximum-entropy method, especially when coupled to the maximum-likelihood approach, and show their relevance with respect to various fields. These results can therefore be considered as the building blocks of further research in the direction of more advanced maximum-entropy network models, for instance with the introduction of inter-layer correlation within the benchmarks and the applications of similar reference models to different fields, ranging from the financial sector to the biological systems.

Samenvatting

Allerlei realistische collecties van gegevens kunnen weergegeven worden door *multiplex netwerken* (ook bekend als meerlaagse netwerken of multigrafen): superposities van netwerken die elk op een andere manier de verbindingen tussen knooppunten leggen.

De betrouwbaarheid van zulke weergaves kan worden aangetast door ruis en willekeurigheid, terwijl toch de behoefte bestaat aan het extraheren van zinvolle informatie uit deze - soms enorme - verzamelingen van gegevens. Aan deze behoefte wordt voldaan door de introductie van nulmodellen, ofwel: referentiemodellen, waarmee de waargenomen netwerken vergeleken kunnen worden. We breiden het idee van nulmodellen als kanonieke ensembles van netwerken met bepaalde randvoorwaarden uit naar het multiplex geval en presenteren nieuwe meettechnieken waarmee we gelaagde systemen kunnen karakteriseren op basis van de correlatiepatronen. We maken uitgebreid gebruik van het *maximale-entropie principe* om analytische uitdrukkingen voor de verwachtingswaarden van verschillende grootheden met betrekking tot de topologie van het netwerk te vinden; bovendien gebruiken we de *maximum-likelihood methode* om de modellen zo goed mogelijk de realistische gegevens te laten weergeven.

We behandelen eerst *ongerichte* multiplex netwerken. We introduceren nieuwe maten voor de correlaties tussen de lagen van een multigraaf, zowel voor binaire (ongewogen) als gewogen netwerken. Verder wijzen we op het belang van het gebruik van nulmodellen om de informatie die is gecodeerd in knooppunt-specifieke eigenschappen te onderscheiden van informatie die gerelateerd is aan hogere orde interacties tussen de elementen waaruit het netwerk bestaat. We maken duidelijk dat het gebruik van homogene willekeurige referentiemodellen kan leiden tot misleidende resultaten; heterogene nulmodellen zijn theoretisch geschikter en in de praktijk redelijker.

Vervolgens verleggen we de aandacht naar *gerichte* multiplex netwerken. We tonen aan dat de uitbreiding van de structuurgrootheden die zijn ontwikkeld voor ongerichte netwerken niet triviaal is, aangezien de gerichtheid van de verbindingen impliceert dat de afhankelijkheden tussen lagen van tweeërlei aard zijn: behalve dat er een tendens is verbindingen in verschillende lagen in dezelfde richting te laten wijzen, vanwege de zogenaamde multiplexiteit, is er ook een complementaire tendens verbindingen in verschillende lagen in tegengestelde richting te laten

wijzen, de zogenaamde *multireciprociteit*.

Verder stellen we een methode voor om gecorreleerde multiplex netwerken te reconstrueren, waarin de topologie uit gedeeltelijke informatie afgeleid wordt. Onze techniek bouwt voort op willekeurige reconstructiemodellen, die met succes een aantal gewenste eigenschappen van enkellaagsnetwerken reproduceren (zoals het gewicht van de verbindingen en/of het aantal buren van de knooppunten). Vervolgens wordt de minimale afhankelijkheidsstructuur geïntroduceerd die nodig is om een extra verzameling van hogere-orde interlaageigenschappen te repliceren.

We illustreren dat de maximale-entropie modellen het ook mogelijk maken om de zogenaamde ruggengraat van een netwerk te vinden. We introduceren een grondige methode die, voor elk gewenst niveau van statistische significantie, de subgraaf produceert die niet te reduceren is naar de lokale eigenschappen van de knooppunten van het netwerk. We laten zien dat, in tegenstelling tot eerdere methoden, de exacte maximale-entropie formulering garandeert dat het gefilterde netwerk alleen verbindingen bevat die niet kunnen worden afgeleid van lokale informatie.

In alle eerder genoemde gevallen testen we onze meettechnieken en modellen op verschillende realistische netwerken, met speciale aandacht voor het *World Trade Multiplex*: het netwerk dat de import-export verbindingen weergeeft tussen landen die met elkaar handelen in verschillende producten.

Tenslotte bestuderen we een andere dataset, namelijk het *netwerk van wetenschappelijke publicaties*. We tonen aan dat dit systeem een eenvoudige weergave heeft in termen van een bipartiet netwerk, ofwel een graaf die bestaat uit twee verschillende soorten knooppunten, waarbij er alleen verbindingen zijn tussen knooppunten van verschillend type (in ons geval, artikelen en wetenschappelijke concepten daarin). De toepassing van een *community detection* algoritme stelt ons in staat om conclusies te trekken over specifieke aanpakken van auteurs om hun artikelen te classificeren. Bovendien geven we diepgravender interpretaties van het begrip *gouden standaard*.

List of publications

[1] V. Gemmetto, A. Barrat, C. Cattuto (2014) 'Mitigation of infectious disease at school: targeted class closure vs school closure', *BMC Infectious Diseases* **14** (1), 695

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

[3] V. Palchykov, V. Gemmetto, A. Boyarsky, D. Garlaschelli (2016) 'Ground truth? Concept-based communities versus the external classification of physics manuscripts', *EPJ Data Science* **5** (1), 28

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

[5] A. Maulana, V. Gemmetto, D. Garlaschelli, I. Yevesyeva, M. Emmerich (2016) 'Modularity maximization in multiplex network analysis using many-objective optimization', *Proceedings of the 2016 IEEE Symposium Series on Computational Intelligence*, 1

[6] A. Martini, A. Lutov, V. Gemmetto, A. Magalich, A. Cardillo, A. Constantin, V. Palchykov, M. Khayati, P. Cudré-Mauroux, A. Boyarsky, O. Ruchayskiy, D. Garlaschelli, P. De Los Rios, K. Aberer (2017) 'ScienceWISE: topic modeling over scientific literature networks', *arXiv*:1612.07636

[7] V. Gemmetto, A. Cardillo, D. Garlaschelli (2017) 'Irreducible network backbones: unbiased graph filtering via maximum entropy', *arXiv*:1706.00230

[8] V. Gemmetto, D. Garlaschelli (2017) 'Reconstruction of multiplex networks with correlated layers', *arXiv*:1709.03918

[9] F. Picciolo, V. Gemmetto, F. Ruzzenenti (2017) 'A network analysis of the global energy market: an insight on the entanglement between crude oil and world economy', submitted

Curriculum vitæ

I was born in Alba (Italy) on the 10th of August 1989. I graduated in 2008 from the Leonardo Cocito high school in Alba, with specialization in mathematics and science. Afterward, I started my Bachelor Degree in Physics at University of Torino (Italy), where I graduated in 2011. In 2013 I graduated, from the same university, in Physics of Complex Systems. In November 2013 I moved to Leiden (the Netherlands), where I enrolled as a PhD candidate at the Instituut-Lorentz for Theoretical Physics. In November 2017 I moved to Amsterdam to work as a quantitative researcher at Duyfken Trading Knowledge.

Acknowledgements

This thesis includes most of the work I have done in the last four years. However, no man is an island, so this humble piece of science would have not been possible without the help and support of several people, who definitely deserve a mention.

First of all, I thank my supervisor Diego Garlaschelli. His vast knowledge and unlimited patience led me to the finish line; the freedom he gives to his collaborators and the relaxed environment he creates around himself helped me to increase my scientific and social skills. I also thank my promotor Wim van Saarloos for his cooperation during the last bits of PhD. I thank my former supervisors, Ciro Cattuto and Alain Barrat, for their mentoring and for introducing me to the research community. I thank Alex, Assaf and Vasyl, with whom I shared my entire PhD, for their guidance and friendship. I thank Elena, who arrived later but soon became an important figure in my life, a confidant and a good friend. I thank all the current and former members of my highly diversified group - Tiziano, Qi, Andrea, Janusz and all the undergraduate students - for the time spent together, and all the colleagues of the Institute with whom I shared coffee breaks and lunches. I thank my students, Ruben and Nedim, for their proactiveness and - sometimes misplaced - trust; in particular, I thank Nedim for the Dutch translation of the summary (together with Peter), and for reminding me not to judge a book from its cover. I thank all the co-authors that I have not mentioned so far, in particular Alessio, Franco, Francesco, Alexey, Andrea, Paolo, Asep and Michael. I thank Richard, with whom I shared my teaching duties. And I thank Fran, Marianne, Barry, Monique and all the secretary staff for their Dutch efficiency.

Life, however, does not consist only of science, and sometimes brings us to unknown and unpleasant territories. I therefore thank Lucie, who effectively helped me in the toughest moments. I thank my parents and my sister for their constant support and for reminding me that I will never be alone. And, finally, I thank Gabriele, the key to my peace of mind, who will always be the person without whom I would not be here.