



Analysing Mesoscale Structures in Economic and Financial Networks

PhD in Institutions, Markets and Technologies - Curriculum
in Economics, Management and Data Science

XXXI Cycle

By

Jeroen Allard van Lidth de Jeude

2019

The dissertation of Jeroen Allard van Lidth de Jeude is approved.

Program Coordinator: Prof. Massimo Riccaboni, IMT School for Advanced Studies Lucca

Supervisor: Prof. Guido Caldarelli, IMT School for Advanced Studies Lucca

Supervisor: Dr. Tiziano Squartini, IMT School for Advanced Studies Lucca

The dissertation of Jeroen Allard van Lidth de Jeude has been reviewed
by:

Dr. Naoki Masuda, University of Bristol

Dr. Angelo Bifone, Istituto Italiano di Tecnologia

**IMT School for Advanced Studies Lucca
2019**

Contents

List of Figures	vii
List of Tables	ix
Acknowledgements	x
Vita and Publications	xi
Abstract	xiii
1 Introduction	1
1.1 An introduction to network theory	4
1.2 Micro - meso - macro	6
1.3 A minimal introduction to null models	7
1.3.1 Recipe for maximum entropy models	13
1.3.2 Applications of maximum entropy	14
1.4 Economic and financial networks - an overview	15
1.4.1 Economic networks	15
1.4.2 Financial networks	16
1.4.3 Enriched Models	16
1.4.4 Methods and structure	17
1.5 New methods for mesoscale network structures	17
2 Reconstructing Mesoscale Network Structures	19
2.1 Null models for mesoscale network structures	20

2.2	Data	22
2.3	Methods	22
2.3.1	Modular structure	23
2.3.2	Null models	24
2.3.3	Model selection criteria	25
2.3.4	Recipe for reconstruction	28
2.4	Reconstructing mesoscale structures in economic and financial networks	29
2.5	Discussion	36
3	Detecting Mesoscale Network Structures	38
3.1	Detecting core-periphery and bipartite structures	39
3.2	Methods	40
3.2.1	The limitations of surprise	41
3.2.2	A bimodular surprise	45
3.2.3	Asymptotic results	47
3.3	Detecting mesoscale structures in socio-economic and financial networks	49
3.3.1	Economic networks	51
3.3.2	Financial networks	52
3.4	Discussion	59
4	Unraveling the Multilayer Structure of Corporate Networks	61
4.1	Corporate networks	62
4.2	Data	63
4.2.1	Networks	64
4.3	Network topology of the corporate multilayer network	66
4.3.1	Layer descriptions	66
4.3.2	Multiplex structure	68
4.3.3	Network structure and company characteristics	74
4.4	Discussion	75
5	Maximum Entropy Approach to Link Prediction in Bipartite Networks	77
5.1	Introduction	78

5.2	Methods	79
5.2.1	Link prediction methods	80
5.2.2	The Bipartite Configuration Model approach	82
5.2.3	Evaluation measures	83
5.3	Data	84
5.4	Results	85
5.5	Discussion	89
6	Conclusion	90
A	Null Models	95
B	Bimodular Surprise	103
B.1	Numerical approximations	103
B.2	Core classification - node degree	108
B.3	Weighted bimodular surprise	109
B.3.1	Method	109
B.3.2	Results	109
C	Multilayer Structure of Corporate Networks	111
C.1	Structural reducibility framework	111
C.2	Robustness of the multiplex regression	111
C.3	Node properties regression	112
C.4	Applications	113
C.4.1	Interwoven relations before a partial takeover: EON - RWE	113
C.4.2	Clustered auditor choice	113
D	Alternative Link Prediction Methods	117
D.1	Other link prediction methods in terms of quantities per node	117

List of Figures

1	Graph example	5
2	Scale of network structures	6
3	Mesoscale structures	7
4	Schematic bow-tie structure	21
5	World Trade Web bow-tie evolution	30
6	World Trade Web degree evolution	31
7	World Trade Web model fit	32
8	Dutch Interbank Network bow-tie evolution	33
9	Dutch Interbank Network model fit	34
10	Traditional surprise limitations - core-periphery graph	43
11	Traditional surprise limitations - k-star graph	44
12	Bimodular surprise approximations	48
13	Bimodular surprise examples	53
14	Further bimodular surprise examples	54
15	Core-periphery structure of the World Trade Web	55
16	Production network core-periphery structure	56
17	Core-periphery structure of e-MID	57
18	Mesoscale structure of e-MID	58
19	Multilayer structure of corporate networks	66
20	Core-periphery structure of German multiplex	69
21	Structural reducibility of the corporate multiplex	72

22	Bipartite graph	79
23	MovieLens link prediction performance	87
24	Venezuelan Banks and Assets link prediction	88
25	Bimodular surprise numerical approximation	105
26	Core classification of high degree nodes	108
27	Weighted surprise - core-periphery	110
28	Weighted surprise - bipartite	110
29	Network structure during EON-RWE merger	114
30	Preferential choice of audit firm	116

List of Tables

1	Mesoscale null models	26
2	Corporate multiplex description statistics	67
3	Layer overlap in the corporate multiplex	71
4	Multiplex network regression results	74
5	MovieLens data description	83
6	MovieLense prediction performance	86
7	Pseudocode of the PACO algorithm	107
8	Robustness of the multiplex network regression	112

Acknowledgements

This thesis is based on the work of the author and coauthors as follows:

- Chapter 2 was published as [123] in the journal *Complexity*, co-authored with Tiziano Squartini, Fabio Saracco, Guido Caldarelli (IMT School for Advanced Studies Lucca) and Riccardo Di Clemente at University College London.
- Chapter 3 is based on [122], under review at *EuroPhysics Letters*. This paper was co-authored with Tiziano Squartini and Guido Caldarelli.
- Chapter 4 is published in the *New Journal of Physics* as an invited article for a special issue on multilayer networks as [121]. This work was mainly done while I was visiting Tomaso Aste at University College London and also co-authored with Guido Caldarelli.
- The content of Chapter 5 is available on arXiv as [8]. This paper was co-authored with Margarita Baltakiene, Kęstutis Baltakys, Dario Cardamone, Federica Parisi, Tommaso Radicioni, Fabio Saracco, and Maddalena Torricelli.

All this work would not be here without the support of all my collaborators and IMT friends. I want to especially thank Tiziano Squartini and Guido Caldarelli for their support, and Tomaso Aste at UCL in London and Iman van Lelyveld at De Nederlanse Bank for hosting me while I took a break from beautiful Lucca.

Vita

- October 3, 1989** Born, Voorburg, The Netherlands
- 2013** B.Sc. in Applied Physics
Thesis:
*Testing the performance of a new Intra-Vascular
Ultrasound transducer*
Technische Universiteit Delft, Delft, The Netherlands
- 2014** M.Sc. in Complexity Science
Theses:
I) *A new lot sizing and scheduling heuristic for
multi-site biopharmaceutical production*
II) *Quantifying the intentions of university applicants
with Google Analytics*
University of Warwick, Coventry, United Kingdom
- 2014-2015** Research Fellow in Applied Bayesian Statistics
*Data-driven maintenance routines for high voltage
electricity assets*
National Grid Plc. & University of Warwick,
Royal Leamington Spa & Coventry, United Kingdom

Publications

(Pre-)Print articles

1. Jeroen van Lidth de Jeude, Riccardo Di Clemente, Guido Caldarelli, Fabio Saracco and Tiziano Squartini. "Reconstructing Mesoscale Network Structures". *Complexity* (2019). DOI:10.1155/2019/5120581
2. Jeroen van Lidth de Jeude, Guido Caldarelli and Tiziano Squartini. "Detecting Core-Periphery Structures by Surprise". arXiv preprint:1810.04717 (2018). (Under review at *Europhysics Letters*).
3. Jeroen van Lidth de Jeude, Tomaso Aste and Guido Caldarelli. "The Multilayer Structure of Corporate Networks". *New Journal of Physics* (2019). DOI: 10.1088/1367-2630/ab022d
4. M. Baltakiene, K. Baltakys, D. Cardamone, F. Parisi, T. Radicioni, M. Torricelli, J. A. van Lidth de Jeude and F. Saracco. "Maximum entropy approach to link prediction in bipartite networks". arXiv preprint:1805.04307 (2018).

Abstract

Modelling, characterising and detecting the structure of complex networks is of primary importance to understand the dynamics of the systems considered. This is especially true for economic and financial networks, whose structural organisation deeply affects their resilience to shock propagation. Many real world networks are characterised by the presence of *mesoscale structures*: while a lot of attention has been focused on the *community* structure, many real-world networks are characterised by *core-periphery*, *bow-tie* and *bipartite* structures, especially so economic and financial networks. In this thesis we present new methods to model and detect these mesoscale structures. We apply these methods to characterise the structure of real-world economic and financial networks. Using maximum entropy networks encoding different levels of information, we model the structure of the international trade network and of national interbank exposures networks. We find that constraining local information is enough to reconstruct the mesoscale structure of these networks: hence, we introduce a new method to detect statistically significant bimodular structures, based on the connectivity *within* and *between* network modules. We also apply our method to multiplex networks. In particular, to unravel different types of corporate networks, we construct a new multilayer dataset of company interactions: we find that the disaggregate network describes a small corporate world, but that these different company interactions are characterised by vastly different topological properties.

Chapter 1

Introduction

The great recession provides a prime example of the importance of networks in economics and finance. Why did a German bank suffer from a failing sub-prime mortgage in the United States? As we now know there are many steps between these events; reselling, repackaging, indirect exposures through various actors: a chain of events. Mapping this interconnectedness helps to untangle the complex system of our economy.

Networks have always existed in the economy of the world. Think of trading webs like the silk road, or the vast colonial trading networks that gave rise to the first stocks [51]. With network theory we can study the efficiency, robustness and resilience to shocks of these interacting systems. In recent years networks contributed a lot to the understanding of systemic risk in the economy [87, 17, 105]. Traditionally risk is mainly an individual attribute; e.g. the probability a company defaults or an individual fails to deliver is estimated from the state of its affairs. What crises illustrate is *systemic risk* rather than individual risks; the risk of a company default not due to the poor state of its own balance sheet, but due to external factors like a failing supplier. For the individuals in the system this is an indirect effect, that affects them by their connections to others. For the system as a whole, such risks are linked to the internal structure [83].

Spreading is a main aspect of systemic risk. Spreading phenomena are present in many fields, like epidemics in biology, and spreading of influence in social science. The systems in which spreading takes place are almost always heterogeneous systems. This means the subjects do not all have the same properties, but, more importantly, that the number of interactions characterizing their activity is broadly distributed across agents. The key here is that spreading is very much dependent on the structure of the interactions between the individuals. For example, in disease spreading there is an epidemic threshold of connectivity below which basically no spreading will occur, but if the connectivity exceeds this threshold the whole network will be affected [89].

Crisis or epidemics often start with shocks. One of the simplest examples of shocks in the financial system is defaults in a network of outstanding loans [5, 28, 1]. Financial institutions form a network through loans they get from each other. If in such a network one player defaults on its loan, this shock travels to the neighbours of this player, and might cause further defaults there. As such this shock can spread through the network.

The structure of a network plays a very important role in this spreading dynamics [83]. It is easy to imagine that a shock that hits a central player spreads more easily than a shock on poorly connected player [19]. However an interconnected core can also bring resilience in some cases, by "spreading out the pain", or dividing the impact of the shock to multiple neighbours [1]. Shocks are far from the only relevant phenomena which highlight the importance of network structure, but especially in an economic setting they provide a simple accessible illustration.

The structure of networks can be described in many ways [89]. On the microscale there exist a vast amount of centrality measures that describe various kinds of importance of the individual actor. Looking at the aggregate of such microscale properties can describe the network as a whole, i.e. the macroscale. The macroscale analyses question like whether the system consists of equal players, or of few important and many fractional players. In-between these two scales is what is proba-

bly most visibly recognizable for network structure: the mesoscale, the properties of groups of nodes. Is the network clustered into groups? Is there a group of central player who everyone connects to?

For many economic and financial systems the network structure on this scale resembles a staged hierarchy: at the centre we find a group of significantly interconnected players, outside this 'core' we find groups of players that are mainly only connected to this core group [84, 20]. An example is the structure of the network of banks [40]. In the centre are 'core' banks, the often called 'too big to fail' institutions, which trade a lot among themselves. In fact, network science has shown that the systemic risk comes from institution that are 'too central', rather than too big [18].

We can describe the structure of networks in many ways. Seemingly endless centrality measures have been defined to rank the importance of players, and we know a lot about the macroscale distributions of local properties, such as scale free distribution where the structure of the network looks similar at different levels of aggregation [89, 30]. On the mesoscale most attention focusses on community detection [52, 53, 73]; identifying clusters of nodes whose internal connectivity is significantly large. Communities however do not represent the only kind of mesoscale structures characterizing real world networks. Other examples are core-periphery structures, bow-tie structures and bipartite structures.

In this thesis we focus on such *mesoscale network structures*. In Chapter 2 we will analyse various alternative models to account for these structures and show the minimal information required to describe these systems. We illustrate the performance of these various models on networks from economics and finance. In Chapter 3 we build on the insights we gain from these real-world applications, and focus on the detection of these mesoscale structures. We will present a novel way to identify the different building blocks of the mesoscale structures.

With applications on real world data we can show the relevance of more theoretical work. In collaboration with the Dutch central bank and later the central bank of Mexico, we started to develop tools to analyse and quantify the mesoscale structure of interbank exposure networks.

Real world data however always brings challenges and often complications. In this work we discovered at a late stage of the project serious data quality issues. The methods we present in Chapters 2 and 3 have been applied to other financial and economic networks, as well as various other real-world networks. In fact, these methods are perfectly general and can be applied to any network structure.

Chapter 4 is instead completely devoted to the study of real-world corporate systems. We will unravel the structure of different types of links between companies, such as ownership ties and research collaborations. Analysing the networks of these different interactions in parallel will give us the opportunity to characterise the differences in structure of different types of company interactions.

Finally, in Chapter 5, we show a spin-off application of the network model framework we use in Chapter 2. We propose a new method for link prediction in bipartite graphs, such as graphs describing consumer purchases. In fact, we apply this method to predict missing links in movie preferences. First, however we introduce some network theory, and give an overview of the subject of economic and financial networks.

1.1 An introduction to network theory

We start here with a very brief introduction to network theory. For excellent, more comprehensive material we refer the reader to [31, 89].

Graph / Network

A *graph* or *network* consists of points or *nodes* which are connected by links or *edges* (we will use these terms interchangeably). Nodes represent the actors in the system; e.g. individuals or companies. An edge between two nodes indicates the existence of some relation between the two nodes. Edges can be undirected or directed, specifying the direction of the flow of the connection. The number of edges connected to a node is called the *node-degree*. As an example, in an interbank network, nodes indicate banks and nodes are linked to indicate the presence of a loan or

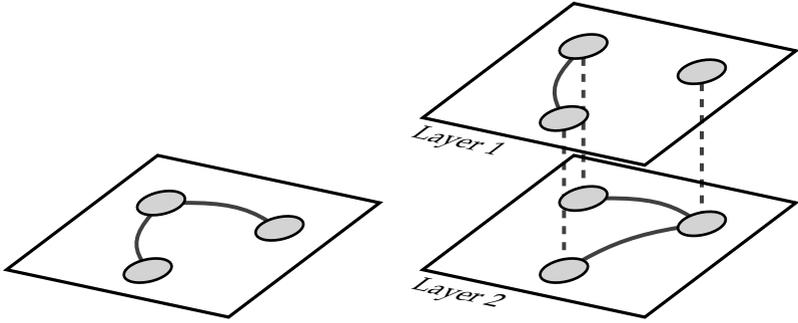


Figure 1: Example of a graph with three nodes and two edges (Left) and a multiplex graph with two layers (Right).

exposure between two banks.

Multilayer networks

Multilayer graphs can represent multiple types of relations in one system. A *layer* is the (sub-)graph which represents the network of a single type of interaction. The special case where the set of nodes is the same in all layers, is also called a *multiplex* graph. In the example of an interbank network, layers could represent different types of exposures, like short-term, long-term, mortgage, etc. Every layer then represents the connections specific to a single type of exposure. An *aggregate network* can be created by collapsing all layers into a single layer.

Network properties

Network properties describe different characteristics of the network, or of the nodes. The *link-density* is the general measure for how interconnected the network is. It is defined by the fraction observed edges over the number possible edges. The *average path length* is a measure of distance between any random selected pair of nodes. It is defined as the average number of edges that constitutes the shortest paths (chain of connected edges) between all pairs of nodes. There are endless more

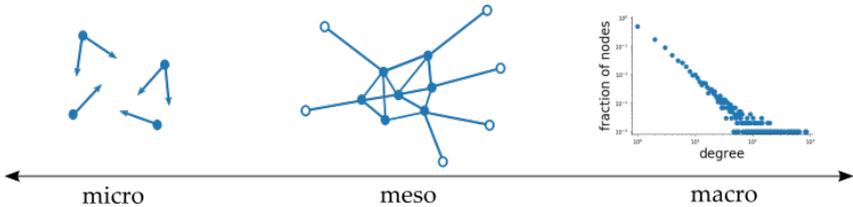


Figure 2: Scale of network structures: individual node properties on the *micro-scale*, the structure of groups of nodes on the *meso-scale*, and the distribution of microscale properties on the *macro-scale*.

network properties describing the centrality of nodes or the tendency to close triangle paths, but these are the main properties used in this thesis.

1.2 Micro - meso - macro

The structure of networks is defined at different scales. On the smallest scale, the *micro-scale*, we find characteristics of single nodes and edges; the degree of a node is such a microscale property. On the other side of the spectrum are *macro-scale* properties. These can be seen as the properties of the distribution of the micro-scale characteristics, e.g. the degree distribution. In-between these ends we have *meso-scale* structures. These consists of groups of nodes or edges, e.g. clusters.

Early networks science furthered a lot the micro and macro, characterizing nodes with the many centrality measures and characterising degree distributions. On the meso-scale, community detection has by far received the most attention. Network *communities* are usually defined as groups of nodes that have significant *intra*-cluster connectivity and a sparse *inter*-cluster connectivity. Many methods exist to detect communities, like the widely used modularity [88] or the more recent proposed surprise [4]. Community structure is found in many real world networks. In economic and financial networks modular structures have been identified as for example trading blocks in the worldwide trading networks [12] and clusters in the corporate board membership networks [125]. In fact, these network communities can often relate to well known

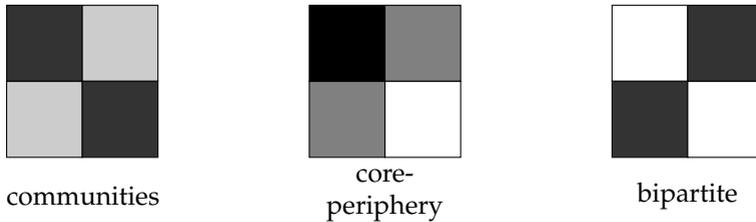


Figure 3: Adjacency matrix representation of some meso-scale structures, when the nodes in the matrix are properly ordered according to their clustering. Darker colours indicate higher edge-density. Blocks on the diagonal (from the top left to the bottom right) signify *intra-cluster* connectivity, off-diagonal blocks show *inter-cluster* connectivity. We formally introduce these structures in Section 2.3.1.

groups as sectors and product classes [113, 103].

Real-world networks however show more mesoscale structures than just communities, examples are *bow-tie*, *core-periphery* and *bipartite* structures. Opposite to communities, these structures **do** have a significant inter-cluster connectivity, see also Figure 3. We will formally introduce some of these structures in Section 2.3.1. Interbank networks have been shown to display a core-periphery like structure [70], while for example corporate ownership networks can show bow-tie structures [125]. Bipartite structures have significant inter-cluster connectivity and intra-cluster connectivity is very sparse. Note that here we intend bipartite structures in monopartite networks, not the more common bipartite graphs that describe the connectivity between two different classes where intra-cluster links are not allowed. Interbank networks might often have a core-periphery structure, but can also manifest bipartite structures [13].

1.3 A minimal introduction to null models

In this section we introduce various graph models used in this thesis. Null models provide a baseline for comparison of the characteristics of an observed system. Generally speaking, we can have multiple null models, incorporating different levels of information or characteristics

of the observed system. Null models allow significant structures to be revealed in observed networks, or equivalently, to identify which characteristics of the observed network can be explained by (the characteristics of) the null model. For example, if a null model generates exactly the same network as an observed one, we can 'use' the characteristics of the null model to describe the observed process.

A graph model is a method or set of rules to generate a network. In general a graph model does not generate a specific network configuration, but rather an ensemble of networks adhering to the set of model rules.

One of the simplest models is a Random Graph, which generates random graphs with a sole constraint on the overall edge density. The existence of an edge between every pair of nodes is simply proportional to the overall density; hence the name Random Graph. The Directed Random Graph (DRG) is the version of this model with directed edges. This model is further discussed in appendix A.

The Configuration Model (CM) is the name for a collection of models which is based on constraining the degree distribution or sequence. Therefore the ensemble of graphs is constrained by the degree sequence instead of the density as for the random graph. The CM can be understood as building the network up from the degree sequence, or the sequence of stubs. Randomly matching all these stubs together creates one configuration of the model. This random process can also be understood with the idea of entropy of the system.

Maximum entropy networks

We will discuss briefly how the maximum entropy principle helps to analyse networks. A key observation on this framework is that the maximum entropy procedure provides the least-biased distribution given our constraints.

Entropy is a measure in physics (Gibbs entropy) or information theory (Shannon entropy) that indicates 'uncertainty' or 'disorder' [71]. More precisely, entropy measures the multiplicity of a state. In terms of net-

works, this multiplicity can be seen as the number of possible configurations of nodes and edges that satisfy certain constraints. Disordered, or random states tend to have high multiplicity of equivalence states. This is why we can say maximizing entropy is equal to maximizing disorder [67]. We can think of maximally uncertain as least biased, by making no assumptions on distributions other than the constraints in the model.

Broadly speaking, the graph *ensemble* is the set of all network configurations that are compatible with a given set of constraints. A basic constraint on a binary graph can be the number of nodes of a graph. In this case the ensemble ranges from a graph with zero edges where all nodes are isolated, to a fully connected graph where all nodes are connected to all other nodes. The graph model would define how likely it is to encounter any of these configurations: it defines a *probability distribution* over the ensemble.

Boltzmann stated that systems in nature always tend to maximize entropy. So why do networks often show ordered hierarchical structures, instead of totally random structures? This observation indicates that there are forces acting on the system to favour specific configurations. The order appears because the system is not an isolated boundary-less system, but there are driving forces on the system that push the system into certain ordered states [67].

With a maximum entropy graph model we evaluate the probability of an observed network configuration under the null model. In this way we can identify significant structures with respect to the maximally random state of the model we define.

This framework is popular in network analysis because it defines the probability distribution over the graph ensemble in the least biased way, i.e. maximally random under the set of chosen constraints [89]. The flexibility on choosing the constraints of the model provide another advantage. One can analyse complex systems step by step, by imposing increasingly less trivial constraints to inspect a system's level of self-organisation.

Exponential random graph framework

We now formally introduce the Exponential Random Graph (ERG), the graph model resulting from entropy maximization. We define \mathcal{G} the grandcanonical ensemble of graphs. Grandcanonical signifies that we use soft constraints; the constraints must be satisfied by the ensemble on average. The set of constraints we impose on the model is $\vec{C}(\mathbf{G})$. Equivalent to graph \mathbf{G} we can also write the corresponding adjacency matrix \mathbf{A} . The Shannon entropy under a given set of constraints $\vec{C}(\mathbf{A})$ reads [109]

$$S = - \sum_{\mathbf{A} \in \mathcal{G}} P(\mathbf{A}) \ln P(\mathbf{A}). \quad (1.1)$$

The probability coefficient $P(\mathbf{A})$ is now assigned to every graph in the ensemble. This defines the constraint-optimization problem. The result is the well-known exponential distribution

$$P(\mathbf{A}|\vec{\theta}) = e^{-H(\mathbf{A},\vec{\theta})} / Z(\vec{\theta}), \quad (1.2)$$

where $\vec{\theta}$ expresses the Lagrange multipliers. The Hamiltonian expressing the imposed set of constraints is

$$H(\mathbf{A}, \vec{\theta}) = \vec{\theta} \cdot \vec{C}(\mathbf{A}), \quad (1.3)$$

and the partition function for normalization of the distribution function is

$$Z(\vec{\theta}) = \sum_{\mathbf{A} \in \mathcal{G}} e^{-H(\mathbf{A},\vec{\theta})}. \quad (1.4)$$

Degree-informed null models: Directed Configuration Model (DCM)

From this specification of the exponential random graph we define specific null models by setting the constraints $\vec{C}(\mathbf{A})$. We can make null models for observed graphs with the actual observed characteristics, rather than picking for example a standard degree distribution. As an illustration of this process we will now derive the Directed Configuration Model. This model constraints the directed degree sequence. This relative simple version of the configuration model has proven informative

in many cases [110, 102, 37, 38]. Further models will be introduced in Chapter 2 and formally defined in Appendix A.

The Directed Configuration Model (DCM) is obtained by restricting the degree sequence of the model. Typically we restrict the degree sequence to the observed degree sequence of the network that we intend to model. Let \mathbf{A}^* be the adjacency matrix of a real unweighed directed graph with entries a_{ij} indicating the presence of a link between node i and node j . The directed degree sequences are given by $\vec{k}^{in}(\mathbf{A}^*) = \sum_{j(\neq i)} a_{ij}$ for the in-degrees and $\vec{k}^{out}(\mathbf{A}^*) = \sum_{j(\neq i)} a_{ji}$ for the out-degrees. The constraints for the DCM are these directed degree sequences: $\vec{C}(\mathbf{A}) = \vec{k}^{in}(\mathbf{A}) + \vec{k}^{out}(\mathbf{A})$. With this specific constraint the Hamiltonian reads

$$H(\mathbf{G}, \vec{\alpha}, \vec{\beta}) = \sum_i (\alpha_i k_i^{in}(\mathbf{G}) + \beta_i k_i^{out}(\mathbf{G})) \quad (1.5)$$

$$= \sum_{i \neq j} (\alpha_i + \beta_j) a_{ij} \quad (1.6)$$

Upon defining $x_i \equiv e^{-\alpha_i}$ and $y_i \equiv e^{-\beta_i}$, the multipliers induce probability coefficients reading

$$p_{ij} = \frac{x_i y_j}{1 + x_i y_j} \quad (1.7)$$

to be numerically determined by solving the likelihood equations

$$\begin{cases} k_i^{in} = \langle k_i^{in} \rangle, \forall i \\ k_i^{out} = \langle k_i^{out} \rangle, \forall i \end{cases} \quad (1.8)$$

with the expected in- and out-degrees reading $\langle k_i^{in} \rangle = \sum_{j(\neq i)} p_{ij}$ and $\langle k_i^{out} \rangle = \sum_{j(\neq i)} p_{ji}$. Explicitly we usually solve the system of equation

$$\begin{cases} k_i^{in} = \sum_{j \neq i} \frac{x_j y_i}{1 + x_j y_i}, \forall i \\ k_i^{out} = \sum_{j \neq i} \frac{x_i y_j}{1 + x_i y_j}, \forall i \end{cases} \quad (1.9)$$

The Directed Random Graph model (DRG) can be recovered as a particular case of the DCM, obtained by posing $\alpha_i \equiv \alpha$ and $\beta_j \equiv \beta$ in Eq.

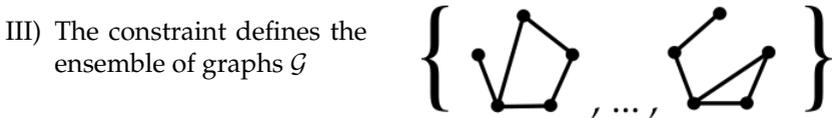
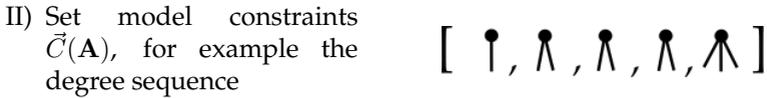
1.5. The only coefficient $p_{ij} \equiv p$ is determined by solving the equation $L = \langle L \rangle$ with $L = \sum_{i \neq j} a_{ij}$ and $\langle L \rangle = \sum_{i \neq j} p$, where L is the total number of edges in the graph.

Configuration model solutions

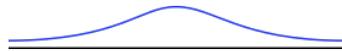
The CM can be implemented numerically or analytically. By randomly generating the graphs matching the node-stubs or randomizing a network respecting the degree sequence constraints, one can numerically generate graphs according to the model. The second method is the analytical approach by which we seek to specifically write down expressions that describe the ensemble characteristics. The computational method has limitations and can be computationally expensive [109]. Throughout this thesis we use a specific analytical method that solves the maximum entropy constraint optimization, solving the system of equation for the Lagrange estimators. This solution specifies directly expected values of network topological characteristics and all link probabilities. More details can be found in [109].

1.3.1 Recipe to implement maximum entropy null models for observed networks

To recap, the general procedure to create null models for real-world graphs using the maximum entropy framework is as follows:



IV) Solve maximum entropy constraint optimization to find probability distribution over the ensemble: probability per graph $P(\mathbf{A}|\vec{\theta})$



V) Find:

- all edge-probabilities p_{ij}
 \Rightarrow used for link prediction in Chapter 5
- observed graph probability or likelihood
 \Rightarrow used for reconstruction in Chapter 2



$$P(G^*|\vec{\theta})$$

1.3.2 Applications of maximum entropy

The maximum entropy framework has been proven very successful in a wide range of tasks [107]. A relevant application is the task of network reconstruction, where the objective is to reconstruct graphs from partial information. There are many situations in which part of the network structure is unknown, or at least not available. Confidentiality in the financial sector often prohibits an institution from revealing the counter parties involved in their deals or loans. This means the interaction network has to be reconstructed from for example the number of deals per institution (degree sequence) [7]. Recent works have shown the differences between reconstruction methods and which information is required for these models. For reconstructing credit networks, reconstruction from only the edge weights works well [99]. Analysis of worldwide exports (World Trade Web - WTW) showed information about edge weights might not be superior than the binary information on links when reconstructing the WTW network [108].

A task similar to reconstruction is missing link prediction. Whereas in reconstruction usually all individuals and the number of connections they have are known, in link prediction there is an entire subset of edges unknown. As a spin-off work of the maximum entropy framework, we will discuss link prediction in bipartite networks. Link prediction is the task of predicting or proposing most likely links in a graph. For bipartite networks this can be used for example as recommendation systems, suggesting (bipartite) links between consumers and products. In Chapter 5 we use link probabilities from the maximum entropy graph model to predict links in real world networks, such as film suggestions and bank-asset links.

Pattern detection is another useful application of the maximum entropy framework. We can for example determine the over- or under-representation of certain small network motifs in an observed network [110]. Coupled with a time dimension this allows to quantify the evolution of network structure over time [102]. We will use a similar approach in Chapter 3 when analysing changes in core-periphery structures of real

world networks.

1.4 Economic and financial networks - an overview

Many actions in economics and finance evolve around exchanges or transactions. Systems with many interacting agents can be seen as networks of interactions. When the agents are persons, networks can represent for example consumer behaviour, like linking customers to purchases. For agents that are companies, common networks are interbank network - loans between banks - and production networks, detailing the various steps in the production of goods. On a larger scale, one can see the aggregate of production networks as sectoral input-output networks, describing flow between sectors. Worldwide, the transnational flow of goods translates to the World Trade Web, the network of imports and exports between countries.

1.4.1 Economic networks

Studies of economic networks, like the world trade web, have increased the understanding of economic phenomena such as the industrial growth of countries. In many instances network analysis highlight also the fragility of the global economy. It has been shown that relatively small localized shocks can initiate avalanches in production networks describing supplier relations [34]. From the network of countries and their exports one can also identify industrial specialization among countries [112, 66]. For companies, or corporate networks, networks are for example used to map the chains of ownership of companies. These ownership networks have revealed a small corporate world; a large concentration of ultimate ownership in a small group of core companies in this complex network [125]. The relations between companies have also been described by financial ties.

1.4.2 Financial networks

Financial networks focus on the flow, or at least the exposure, of capital. The great financial recession has put a focus on the analysis of interconnectedness in the financial system. Networks provide a good framework to represent this interconnectedness as networks are build of relations. A lot of research has focussed on so called *interbank networks*. These describe transactions, or exposures between banks, but also wider financial institutions. Studies of these networks have identified as a key revelation the importance of second order effects in systemic risk. Meaning indirect network effects can be as important as direct shocks for the stability financial institutions [62, 57]. The interconnectedness also comes from indirect channels, rather than direct bilateral links such as a loan between two parties. This indirect effect is illustrated by so called overlapping exposures. A shock to an asset or sector will affect all parties that have an exposure to that asset or sector. This provides an indirect channel of shock propagation between financial institutions [29, 28]. A final type of financial networks are the ones describing stock markets. Stock prices can contain a lot of information about the state of companies and the wider markets. Networks constructed from the stocks can reveal significant patterns in this underlying information [84, 23, 119]

1.4.3 Enriched Models

For certain applications the best models of the (dynamics of the) network take into account some of the underlying economics in addition to the network structure [105]. Systemic risk in financial network can be approached from a pure network analysis, as an epidemic spreading process [87]. More sophisticated models however take into account the structure of the balance sheets of the banks [18, 83]. The network of worldwide imports and exports or economic growth of countries are other examples where mixed models perform well [68].

1.4.4 Methods and structure

The structure of these networks is generally described by significant topological characteristics. General descriptions as scale free properties and degree distributions can reveal overall clustered structures or homogeneous ones. Recently some doubt has been cast over the widespread classification of power-law distributions [39, 124]. Null models as discussed before are a powerful tool to reveal significant structure as small graph motifs and communities [11, 110] as also described in Section 1.3.2. Community detection has received a lot of attention in economic networks focussed on detecting specialisations of countries and companies [112]. The *mesoscale* structure of many financial and economic networks has been found different from communities. The network of world trade has been characterized as *core-periphery* [108], interbank networks are often found to exhibit some core-periphery structure [70], but also *bipartite structures* [13]. The flow of control in ownership networks can be described by a *bow-tie structures* [125]. For the detection and characterisation of these meso-scale structure different techniques are used. For core-periphery structure most recent studies use modifications of the Borgatti model [24]; a score function indicating the extent to which a given graph partition deviates from an ideal core-periphery structure. Such measures have limitations, as a high dependency on node degree and no embedded significance evaluation.

1.5 New methods for mesoscale network structures

In this thesis we will propose new methods for the analysis of mesoscale network structures, and use these methods to analyse various real-world economic and financial systems. In the next chapter we will clarify which null models fit best with the observed modular structures in real-world economic and financial networks. In Chapter 3 we propose a measure based on connectivity to try and overcome some of the limitations in detecting core-periphery structures with alternative methods similar to the

Borgatti model. Finally we look at the interplay of certain of these economic/financial networks. Almost exclusively this research looks at the single types of networks in isolation, instead of the entire ecosystem. As described before, the mesoscale structure is then often characterised as core-periphery or bow-tie [125, 132]. However many of these networks are inherently linked; e.g. a company subsidiary is both linked in the production chain and the ownership chain to the parent company. With new data on corporate networks we will disentangle the aggregate corporate network and reveal that the layers in this multiplex show different mesoscale structures.

Chapter 2

Reconstructing Mesoscale Network Structures

The content of this chapter was published in the journal Complexity, see [123].

In this chapter we look at null models for mesoscale network structures and the question of whether information on this modular structure should be incorporated in these models. When facing the problem of reconstructing complex mesoscale network structures, it is generally believed that (null) models encoding the nodes organization into modules must be employed. Here we focus on two block structures that characterize the empirical mesoscale organization of many real-world networks, i.e. the *bow-tie* and the *core-periphery* ones, with the aim of quantifying the minimal amount of topological information that needs to be enforced in order to reproduce the topological details of the former. Our analysis shows that constraining the network degree sequence is often enough to reproduce such structures, as confirmed by model selection criteria as AIC or BIC. Furthermore, as a byproduct, we enrich the toolbox for the analysis of bipartite networks. Both the bow-tie and the core-periphery structures, in fact, partition the networks into asymmetric blocks characterized by binary, directed connections, thus calling for the extension of a recently-proposed method to randomize *undirected*, bipartite networks

to the *directed* case.

2.1 Null models for mesoscale network structures

The analysis of mesoscale network structures is a topic of great interest within the community of network scientists: much attention, however, has been received by the community-detection topic [52, 53, 73], while the analysis of other meso-structures has remained far less explored.

In this chapter we aim to contribute to this stream of research, by exploring the effectiveness of null models that constrain only local information in reproducing complex meso-structures. When approaching such a problem it is, in fact, commonly believed that models encoding the nodes organization into modules must be employed: here we test this hypothesis, by comparing models that enforce topological information like the total number of links, the degree sequences and the reciprocity structure with their block-wise counterparts.

To this aim, we have considered real-world networks whose topological structure is *empirically* characterized by bow-tie and core-periphery structures: both are characterized by a central, cohesive subgraph surrounded by a loosely-connected set of nodes [41]. In the first case, however, the central part of the network has a fan-in and a fan-out component, respectively entering into and exiting from it.

In order to do so, we compare competing null models and apply model selection criteria in order to unambiguously determine the “winner”, i.e. the one carrying the right amount of information to account for the inspected structures. Remarkably, all null models considered in this chapter can be recovered within the same framework, i.e. the entropy-maximization one introduced in Chapter 1, which has been proven to be rather effective for both pattern detection and real-world networks reconstruction [109, 85].

When studying these mesoscale structures, it is evident that analysing the way nodes cluster together unavoidably leads to the analysis of the way such modules interact. The interactions *between* two clusters of nodes

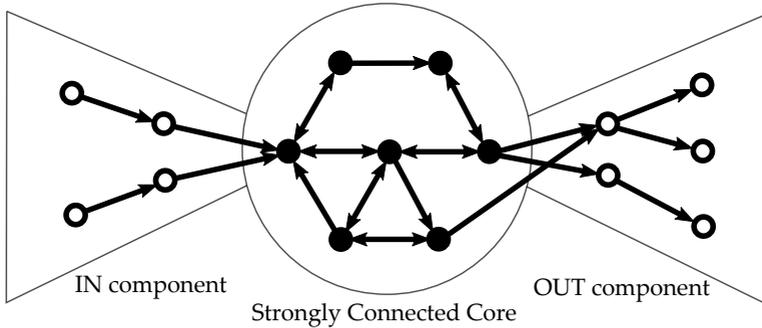


Figure 4: Schematic visualisation of a bow-tie network structure.

are described by a bipartite network. In fact, the adjacency matrix of the clustered structure is characterized by diagonal square blocks (i.e. the adjacency matrices of the modules themselves) and off-diagonal rectangular blocks (i.e. the adjacency matrices of the bipartite networks encoding their interactions). Among the many, available, network representations, the bipartite one has recently received much attention [115, 36]. This, in turn, has led to the definition of algorithms for randomizing [74, 46, 114, 104], reconstructing [111] or projecting [120, 103] *undirected*, bipartite networks. The directed case, however, has not yet been explored, thus calling for the definition of techniques to approach the study of this kind of networks as well. Our method will be employed to analyse economic and financial networks. More specifically, we will focus on two systems: the World Trade Web and the Dutch Interbank Network. As we will show, while the former can be empirically characterized by a partial bow-tie structure, the latter is characterized by the co-existence of a core-periphery-like structure and a proper bow-tie one, the second one carrying a larger amount of information about the system evolution than the first one.

2.2 Data

We consider two real world systems for the present analysis; one *economic* network, the World Trade Web, and one *financial* network, the Dutch Interbank Network.

The World Trade Web. We consider yearly bilateral data on exports and imports from the UN COMTRADE database [128], from 1992 to 2002. We limit ourselves to considering the World Trade Web (WTW hereafter) in its binary, directed representation at the aggregate level. In order to perform a temporal analysis and compare different years, we restrict ourselves to a balanced panel of $N = 162$ countries (present in the data throughout the considered interval). Accordingly, for a given year t , $a_{i,j}^t = 1$ ($a_{i,j}^t = 0$) means that country i has registered a non-null (null) export towards country j .

The Dutch Interbank Network. We consider a dataset where nodes are Dutch banks and a link from node i to node j indicates that bank i has an exposure larger than 1.5 million euros and with maturity shorter than one year, towards a creditor bank j [70]. We consider 44 quarterly snapshots of the Dutch Interbank Network (DIN hereafter), from 1998Q1 to 2008Q4. The last year in the sample represents the year during which the recent financial crisis became manifest.

2.3 Methods

Let us, first, define the modular structures that we consider in this chapter. The partitioning of the data with respect to these modules will be imposed by the specific modular structure, and will thus be fixed rather than fitted. We then continue to explain which competing null models we will use. Finally we introduce the model selection criteria used to compare the explanatory power of these competing models.

2.3.1 Modular structure

Let us now provide an algebraic representation of the mesoscale structures considered in this chapter, i.e. the bow-tie and the core-periphery ones, as described informally in 1. Networks whose topology is empirically characterised by a core-periphery structure can be represented as follows:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}^\bullet & \mathbf{A}^\top \\ \mathbf{A}^\perp & \mathbf{A}^\circ \end{pmatrix}; \quad (2.1)$$

the adjacency matrix \mathbf{A} is composed by four distinct blocks: while the square adjacency matrices \mathbf{A}^\bullet and \mathbf{A}° lying along the diagonal represent the core and the periphery modules, the two rectangular (in the most general case), off-diagonal matrices \mathbf{A}^\top and \mathbf{A}^\perp represent the (bipartite) networks through which they interact. Usually, the link densities of the matrices above satisfy the chain of relationships $c(\mathbf{A}^\bullet) > c(\mathbf{A}^\top) \simeq c(\mathbf{A}^\perp) > c(\mathbf{A}^\circ)$, i.e. the core module is (much) denser than the periphery module.

Notice that the two matrices \mathbf{A}^\top and \mathbf{A}^\perp bring genuinely different information: while the generic entry $a_{cp}^\top = 1$ ($a_{cp}^\top = 0$) indicates that a directed link from the node c in the core to the node p in the periphery is present (absent), the generic entry $a_{pc}^\perp = 1$ ($a_{pc}^\perp = 0$) indicates that a directed link from the periphery node p to the core node c is present (absent). In other words, in order to fully describe the topological structure of *one, directed* bipartite network, *two* matrices are, in fact, needed. Naturally, in case the network \mathbf{A} is undirected, $\mathbf{A}^\bullet = [\mathbf{A}^\bullet]^\top$, $\mathbf{A}^\circ = [\mathbf{A}^\circ]^\top$ and $\mathbf{A}^\top = [\mathbf{A}^\perp]^\top$, which restores the symmetry of the whole adjacency matrix (i.e. $\mathbf{A} = \mathbf{A}^\top$).

While the definition of core-periphery structure is quite intuitive, the definition of bow-tie structure, on the other hand, is based on the concept of node *reachability*: node i is reachable from node j if a path exists from node i to node j (a path being defined as a sequence of adjacent links connecting i with j). According to this definition, each node is assigned to one of the sets described in [129]. The definition of the three most

relevant ones follows:

- SCC: each node in the Strongly Connected Component (SCC) is reachable from any other node belonging to the SCC;
- IN: each node in the SCC is reachable from any node belonging to the IN-component;
- OUT: each node in OUT-component is reachable from any node belonging to the SCC.

According to the definitions above, networks whose topology is empirically characterised by a bow-tie structure can be represented by the following adjacency matrix

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}^i & \mathbf{A}^{\setminus} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^s & \mathbf{A}^{\setminus\setminus} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}^o \end{pmatrix} \quad (2.2)$$

the three blocks \mathbf{A}^s , \mathbf{A}^i and \mathbf{A}^o representing the SCC, IN- and OUT-component respectively. The off-diagonal matrices \mathbf{A}^{\setminus} and $\mathbf{A}^{\setminus\setminus}$, instead, represent the (bipartite) networks through which they interact.

Notice that the partitioning of the graph in these modules is imposed by our definition of the bow-tie structure. In our data applications the core-module of the core-periphery structure will also be taken as the SCC, as argued in more detail when we discuss the data characteristics.

2.3.2 Null models

To explore if specific information about nodes group membership is require to reconstruct these mesoscale structures, we consider two sets of null models:

- constraining local information,
- constraining local information **and** group membership of the nodes.

Throughout these sets we consider three classes of null models: *density informed*, *degree informed*, and *reciprocity informed*.

We use the null models from the maximum entropy framework as introduced in Chapter 1. The three classes of null models we consider put constraints on an increasing amount of information of the graph. As a basic model we use the *Directed Random Graph* (DRG), which puts a sole constraint on the edge density, denoted by l . The *Directed Configuration Model* (DCM) only uses the directed degree sequences; the in-degree sequence \vec{k}^{in} and out-degree sequence \vec{k}^{out} . When analysing directed networks, however, a non-trivial piece of information to be taken into account is represented by reciprocity [60]. Therefore we also use the *Reciprocated Configuration Model* (RCM) which restricts the reciprocated degree sequence (in-, out- and reciprocated-degrees of the nodes: \vec{k}^{in} , \vec{k}^{out} , \vec{k}^{rec}).

The models that also constrain the group membership of the nodes are denoted by the prefix *Block* (or the abbreviated B). The block version of the DRG is also known as the Stochastic Block Model (SBM). For the DCM and RCM the block versions are indicated as the *Block Configuration Model* (BCM) and the *Block Reciprocated Configuration Model* (BRCM). These models put exactly the same constraints as their corresponding non-block models, but are solved within all blocks of the core-periphery or bow-tie structure. Again, the block-structure is known in all cases, as we will expand on when introducing the data. In Appendix A we provide the full specification and derivation of these models.

To sum up, we will look at the following models, both with and without information about the modular structure of the network.

2.3.3 Model selection criteria

Although rising the number of parameters in order to better reproduce the observations is tempting, the risk of overfitting should be, nevertheless, avoided. Introducing the information of the group membership of

Without block information		With block information	
	Constraints		Constraints
DRG	l	SBM	l within and between blocks
DCM	$\vec{k}^{in}, \vec{k}^{out}$	BCM	$\vec{k}^{in}, \vec{k}^{out}$ within and between blocks
RCM	$\vec{k}^{in}, \vec{k}^{out}, \vec{k}^{rec}$	BRCM	$\vec{k}^{in}, \vec{k}^{out}, \vec{k}^{rec}$ within and between blocks

Table 1: Competing mesoscale null models

the nodes with the block models gives such a rise in parameters. A criterion to identify the best model out of a basket of possible ones is, thus, needed. In what follows, we will adopt the Akaike Information Criterion (AIC hereafter)

$$\text{AIC} = -2\mathcal{L}(\mathbf{A}) + 2K + \frac{2K(K+1)}{n-K-1} \quad (2.3)$$

and the Bayesian Information Criterion (BIC hereafter)

$$\text{BIC} = -2\mathcal{L}(\mathbf{A}) + K \ln n \quad (2.4)$$

whose first addendum, in both cases, is proportional to the log-likelihood of the null model under analysis, K is the number of parameters defining the model and n is the sample size (set, as usual, at $N(N-1)$). Both AIC and BIC are minimum for the best explanatory model in the basket [27].

In order to make Eqs. 2.3 and 2.4 more explicit, let us call B the number of blocks our network has been divided into (i.e. the *diagonal* blocks of the matrix \mathbf{A}). While the Directed Random Graph (DRG) is defined by just one parameter, $K_{DRG} = 1$, the Stochastic Block Model (SBM) is defined by $K_{SBM} = B^2$ parameter (as can be verified upon inspecting definitions 2.1 and 2.2). Specifying the degree sequence leads to a further rise of the number of parameters: the Directed Configuration Model (DCM) is, in fact, defined by $K_{DCM} = 2N$, and the Block Configuration Model (BCM) is defined by $K_{BCM} = 2NB$ (each node, in fact, “needs” two parameters per block: one for the in-degree and one for the out-degree). Accounting also for the information provided by the reciprocity requires a number of parameters to be specified that is $K_{RCM} = 3N$ for the Reciprocal Configuration Model (RCM) and $K_{BRCM} = 3NB$ for

the Block Reciprocal Configuration Model (BRCM - each node, in fact, “needs” three parameters per block).

The model selection framework based upon the two information criteria above allows the probability that a given model m is the best approximating model to be calculated as well, via the so-called *AIC weights* and *BIC weights*, defined as

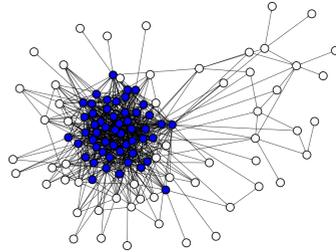
$$w_m = \frac{e^{-\Delta_m/2}}{\sum_m e^{-\Delta_m/2}} \quad (2.5)$$

with $\Delta_m = \text{AIC}_m - \min\{\text{AIC}_m\}_m$ and $\Delta_m = \text{BIC}_m - \min\{\text{BIC}_m\}_m$, respectively.

2.3.4 Recipe for reconstruction of mesoscale network structures

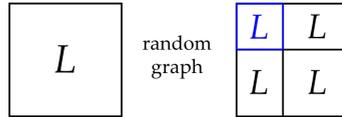
To recap, the general procedure to evaluate the reconstruction of mesoscale structures follows:

I) Observed graph G^* , with fixed modular partition

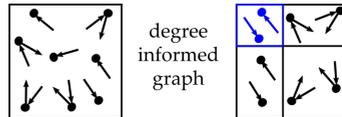


II) Model the graph with the maximum entropy framework using different constraints:
DRG & BlockDRG (or SBM)

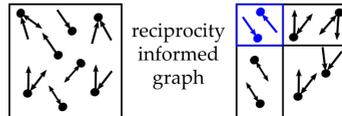
with block structure information



DCM & BlockDCM (or BCM)



RCM & BlockRCM (BRM)



III) Evaluate model fit with AIC and BIC model selection criteria to compensate for the rise in parameters.

$$\begin{array}{ll}
 P(G^* | \vec{\theta}_{DRG}) & P(G^* | \vec{\theta}_{SBM}) \\
 P(G^* | \vec{\theta}_{DCM}) & P(G^* | \vec{\theta}_{BCM}) \\
 P(G^* | \vec{\theta}_{RCM}) & P(G^* | \vec{\theta}_{BRM})
 \end{array}$$

2.4 Reconstructing mesoscale structures in economic and financial networks

The World Trade Web. Although the WTW has been deeply studied throughout the years [106, 55, 50, 86], the analysis of its mesoscale organization has, so far, received far less attention [10, 117]. Interestingly, checking for the applicability of the bow-tie definition provided above, the WTW appears as being partitioned into a SCC and an IN-component only, the OUT-component being completely missing (see Fig. 5). According to the algebraic representation introduced at the beginning of this chapter, the WTW mesoscale structure is represented by the following adjacency matrix

$$\mathbf{A}^{\text{WTW}} = \begin{pmatrix} \mathbf{A}^i & \mathbf{A}^j \\ \mathbf{0} & \mathbf{A}^s \end{pmatrix} \quad (2.6)$$

with $\mathbf{A}^i = \mathbf{0}$ throughout our temporal interval. This implies that the nodes belonging to the IN-component do not establish internal relationships, their links are pointing towards the SCC nodes only (via the \mathbf{A}^j block). Interestingly, the percentage of nodes belonging to the SCC steadily increases with time: from the 32% in 1992 to almost the 75% in 2002. Since the total number of nodes does not vary across the considered temporal interval, the IN-component shrinks accordingly. These results refine the picture drawn in [10], where only the largest connected component was considered.

From a macroeconomic point of view, the increasing number of nodes within the SCC may evidence a sort of ongoing globalization process [10]. It is interesting to notice that the inclusion of (whole subsets of) countries within the SCC seems to be related to the existence of trade agreements. Examples are provided by Commonwealth nations - all of which are part of the SCC since 1993 - European nations (EU as a whole joined the SCC in 1994, the same year of the EEA agreement) and the case of USA (NAFTA entered into force in 1994 as well). From a purely topological perspective, an interesting dynamics takes place: as shown in Fig. 6, the reciprocal degree of nodes belonging to the SCC keeps rising. Since

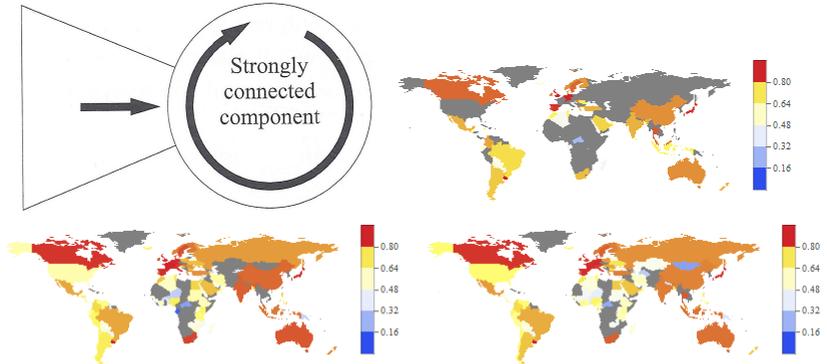


Figure 5: Top-left panel: the WTW bow-tie structure, composed by the SCC and the IN-component only. The other panels show clockwise the countries belonging to the SCC (in colors) and the countries belonging to the IN-component (in gray) in 1993, 1998 and 2002, respectively. Countries belonging to the SCC keep rising their reciprocated degree (see also Fig. 6); richest world countries (Canada, Europe, Japan - in dark red) are always characterized by the largest values of reciprocated degree.

all nodes are characterized by a rather stable in-degree value, this finding points out the tendency of such countries to reciprocate previously-established connections by creating new out-going links (i.e. to consolidate existing trade relationships). Beside revealing that the high value of reciprocity within the SCC is one of the causes behind the existence of a large number of paths within it, the overall effect of this dynamics seems, thus, to be that of fostering trade exchanges between the members of the SCC.

Let us now analyse what kind of topological information is actually needed in order to explain the mesoscale WTW structure. To this aim, let us summing up the observations about the actual structure of the WTW by imagining a *densely-connected, highly-reciprocated* SCC ($c(\mathbf{A}^s) \simeq r(\mathbf{A}^s) \simeq 0.8$ throughout our temporal interval).

The need of considering a block model becomes evident when com-

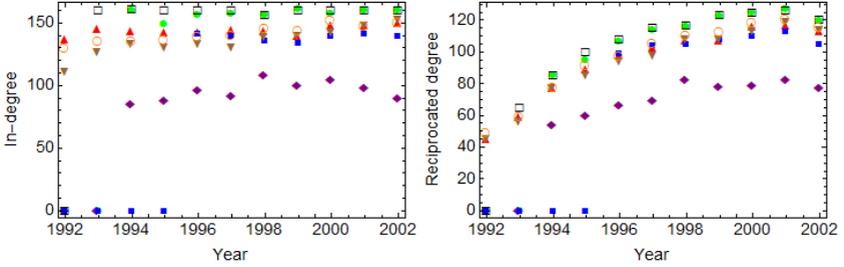


Figure 6: Dynamics of the in-degree (defined as $h_i = \sum_{j(\neq i)} a_{ji}$) and of the reciprocated degree (defined as $k_i^{\leftrightarrow} = \sum_{j(\neq i)} a_{ij}a_{ji}$) of a sample of countries (Italy, in green; Japan, in black; China, in red; Russia, in blue; India, in brown; USA, in purple; Australia, in orange): while the in-degree remains rather stable across time, the value of the reciprocated degree keeps rising once the country has joined the SCC. Such a dynamics can be interpreted as a signal of ongoing integration [10].

paring the homogeneous benchmark provided by the DRG with its block-wise counterpart, i.e. the SBM (see Fig. 7): the SBM outperforms the DRG since the network is “composed” by parts characterized by very different link-densities ($c(\mathbf{A}^s) \in [0.75, 0.9]$ and $c(\mathbf{A}^i) = 0$) that cannot be reproduced with just one, global parameter.

Generally speaking, however, benchmarks encoding the degrees heterogeneity are to be preferred. Interestingly, (both) non-block models outperform block models, indicating that specifying additional information to the one encoded into local properties is indeed unnecessary. This is not surprising, however, when considering that the nodes belonging to the IN-component have zero in-degrees. The latter, in fact, are exactly reproduced by both the DCM and the RCM: the “peripheral” part of the network under analysis is, thus, automatically explained by a simpler kind of statistics with no need to invoke any *a priori* partition.

Let us compare our degree-informed models over the \mathbf{A}^i and \mathbf{A}^s sub-graphs. In the first case, the information carried by reciprocity is encoded into the degree sequence. The result $\mathcal{L}(\mathbf{A}^i)_{BCM} = \mathcal{L}(\mathbf{A}^i)_{BRCM}$ is, in fact, rooted into the observation that the links from the IN-component to the

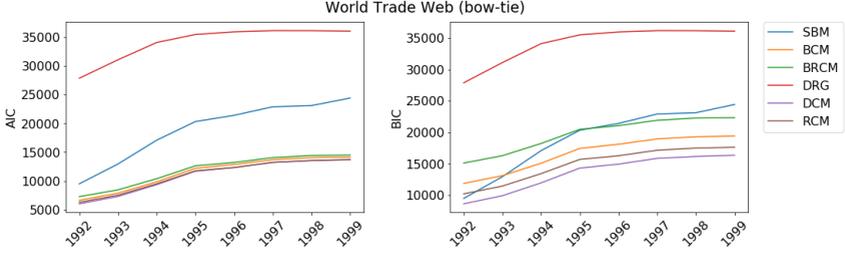


Figure 7: Evolution of the AIC and BIC values for the WTW across the years 1992-2002: while the SBM (blue trend) must be preferred to the traditional DRG (being the network composed by parts with different link densities), heterogeneous benchmarks are, generally speaking, to be preferred. Although the DCM and the RCM are characterized by very similar AIC values, AIC and BIC weights let always the DCM win.

SCC are not reciprocated. The same consideration together with the observation that the large $r(\mathbf{A}^S)$ value is due to reciprocal connections established between nodes *within* the SCC, leads to the result $\mathcal{L}(\mathbf{A}^S)_{BCM} = \mathcal{L}(\mathbf{A}^S)_{BRCM} \simeq \mathcal{L}(\mathbf{A}^S)_{RCM}$; similarly, $\mathcal{L}(\mathbf{A}^S)_{BRCM} \simeq \mathcal{L}(\mathbf{A}^S)_{RCM}$. As a consequence, being the two likelihood values (overall) very similar, the model with a larger number of parameters is more “penalized” (i.e. $AIC_{BRCM} > AIC_{RCM}$). On the other hand, comparing the BCM and the DCM on the SCC leads to the conclusion that, as the latter enlarges, $\mathcal{L}(\mathbf{A}^S)_{BCM} \simeq \mathcal{L}(\mathbf{A}^S)_{DCM}$, since the largest contribution to the nodes degrees comes from the connections established with other nodes within the SCC itself.

Apparently, thus, two non-block models compete, i.e. the DCM and the RCM (see Fig. 6). However, by computing AIC and BIC weights for each model m in our basket

$$w_m = \frac{e^{-\Delta_m/2}}{\sum_m e^{-\Delta_m/2}} \quad (2.7)$$

(with $\Delta_m = AIC_m - \min\{AIC_m\}_m$ and $\Delta_m = BIC_m - \min\{BIC_m\}_m$, respectively) one finds that the DCM always wins. The explanation of this result lies in the fact the WTW reciprocity is compatible with the

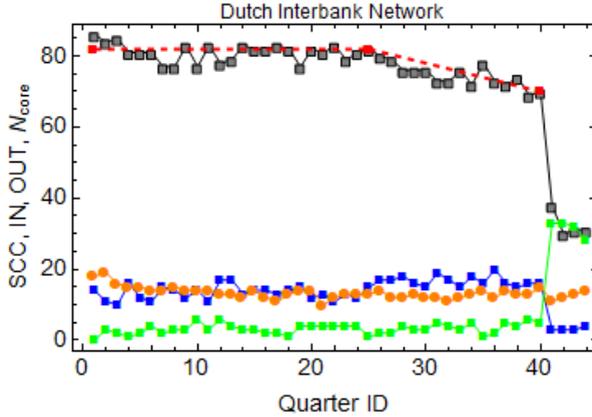


Figure 8: Evolution of the DIN bow-tie structure (the SCC is shown in gray, the IN-component is shown in blue and the OUT-component is shown in green). The crisis period (last four points) is signalled by a sharp decrease of the SCC and IN- components size (and a corresponding increase of the OUT-component size). The size of the SCC, however, starts shrinking in 2004Q1 (deviating from the approximately constant trend observed since 1998Q1), seemingly constituting an additional, early-warning signal of the upcoming crisis. On the other hand, the DIN core (shown in orange) doesn't undergo any significant variation throughout the whole temporal interval.

DCM prediction, as the computation of the index $\rho = \frac{r - \langle r \rangle}{1 - r}$ reveals (it amounts at $\simeq 0.05$ throughout our time interval) [59]. In other words, the seemingly peculiar mesoscale structure of the WTW is, to a good extent, reproduced by just specifying local constraints (in this case, the degree sequence).

The Dutch Interbank Network. According to the axiomatic model in [40], the DIN has been described as characterized by a well-defined core-periphery structure [70]. However, as it has been pointed out elsewhere [110], such a mesoscale organization is compatible with the predictions provided either by the DCM or the RCM, depending on the topological quantity inspected.

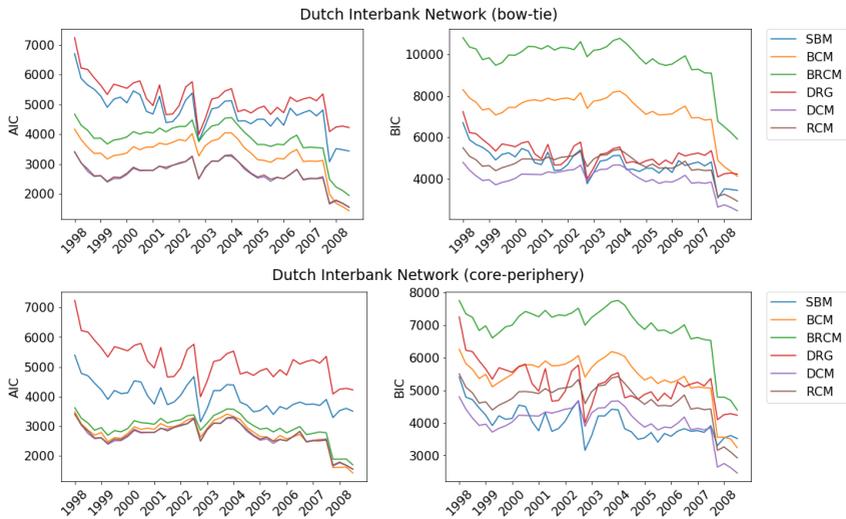


Figure 9: Evolution of the AIC and BIC values for the DIN across the quarters 1998Q1-2008Q4: while the SBM (blue trend) must be preferred to the traditional DRG (being the network composed by parts with different link densities), heterogeneous benchmarks are, generally speaking, to be preferred. Although the DCM wins in the vast majority of cases (both for the bow-tie and the core-periphery organizations), quarters exist where the DCM and the RCM compete. BIC, on the other hand, lets the SBM win when analysing the DIN core-periphery structure.

Notably, the DIN is also characterized by a certain degree of bow-tiness, given the presence of an SCC, an IN-component and, differently from the WTW, also a non-vanishing OUT-component: both the A^I and the A^O blocks, however, are empty, and nodes belonging to the IN- and OUT- components are not directly linked with each other (but only via the SCC nodes). From a purely empirical point of view, the evolution of the DIN bow-tie structure is much more informative than the evolution of its core-periphery structure: as Fig. 8 shows, while the size of the DIN SCC, in 2008, reduces to more than half its pre-crisis value - thus providing an additional, structural indicator of it - the number of nodes

belonging to the core shows no significant variations across the same period. However, the SCC starts shrinking well before 2008, a dynamics seemingly constituting an additional early-warning signal of the upcoming, topological change affecting the DIN. The IN-component, in turn, shrinks as well, while the OUT-component enlarges.

In order to individuate the best model to explain the DIN bow-tie structure, let us notice that its SCC can be imagined as a *weakly-connected, weakly-reciprocated* subgraph ($c(\mathbf{A}^s) \lesssim 0.1$ and $r(\mathbf{A}^s) \simeq 0.3$, except in 2008 where the SCC reciprocity drops to $\simeq 0.15$). More precisely, $c(\mathbf{A}^s) \gtrsim c(\mathbf{A}) \ll c(\mathbf{A}^\bullet)$, i.e. while the SCC connectance basically coincides with the one of the whole network, the core is much denser, an empirical observation that explains why the SBM provides a better explanation of the core-periphery structure (in fact, the AIC and BIC values for the SBM and the DRG are closer when considering the bow-tie structure - see Fig. 9).

Generally speaking, however, models accounting for the degree heterogeneity are to be preferred. As for the WTW, zero in-degree and zero out-degrees are exactly reproduced by non-block models like the DCM and the RCM. On top of this, the low reciprocity value of the DIN (amounting at $\lesssim 0.3$) allows us to imagine it playing a minor role in determining the nodes' degrees. As a consequence, the DCM and the RCM can be interpreted as different ways to rewrite the same (configuration) model. More quantitatively, $\mathcal{L}(\mathbf{A})_{RCM} \gtrsim \mathcal{L}(\mathbf{A})_{DCM}$.

Deviations from this idealized picture, however, exist. This is particularly evident when analysing the \mathbf{A}^s block, to fully understand which reciprocity indeed plays a role (in fact, $\mathcal{L}(\mathbf{A}^s)_{BRCM} > \mathcal{L}(\mathbf{A}^s)_{BCM}$); when considering the "peripheral" blocks, instead, one concludes that $\mathcal{L}(\mathbf{A}^\dagger)_{BRCM} \simeq \mathcal{L}(\mathbf{A}^\dagger)_{BCM}$, $\mathcal{L}(\mathbf{A}^\dagger)_{RCM} \gtrsim \mathcal{L}(\mathbf{A}^\dagger)_{DCM}$ and $\mathcal{L}(\mathbf{A}^{\dagger\dagger})_{BRCM} \simeq \mathcal{L}(\mathbf{A}^{\dagger\dagger})_{BCM}$, $\mathcal{L}(\mathbf{A}^{\dagger\dagger})_{RCM} \gtrsim \mathcal{L}(\mathbf{A}^{\dagger\dagger})_{DCM}$ (since the links from the IN-component to the SCC and from the SCC to the OUT-component are not reciprocated).

Consistently, AIC and BIC weights let the DCM win in the vast majority of cases, although in some periods the DCM and the RCM compete. Overall, this is valid when considering the DIN core-periphery structure

too.

2.5 Discussion

The WTW and the DIN represent two real-world systems characterized by (apparently) non-trivial mesoscale structures: while the first one is characterized by a (partial) bow-tie organization, in the second one the bow-tie partition co-exists with a core-periphery partition. Both kinds of mesoscale structures are characterized by interacting blocks whose internal topology is commonly believed to be determined by a non-trivial interplay between nodes connectivity and the reciprocity of connections. It is, thus, interesting to ask ourselves the extent to which such structures are, instead, accounted for by purely local information.

Remarkably, what our analysis points out is that specifying the degree sequences is often enough to reproduce these mesoscale structures, thus suggesting that the observed modules emerge as a consequence of local connectivity patterns between nodes: for example, the absence of incoming/outgoing connections for a set of nodes naturally leads them to be identified as an IN-/OUT-component.

Differences between systems, naturally, exist. Let us notice that, contrarily to what observed in the WTW case, AIC and BIC provide different answers to the question concerning the performance of block models in explaining the DIN core-periphery structure: while the Akaike criterion ranks the BCM first, the Bayesian criterion assigns the highest score to the SBM in the vast majority of temporal snapshots. If, on the one hand, this saves the role potentially played by blocks, on the other it points out that the large difference between the core and periphery connectivity values [110] provides - by itself - an effective explanation of this mesoscale organization.

A second comment about the DIN concerns the observation that, when considering the core-periphery structure, the AIC values of block models overlap with the AIC values of the simpler models to a larger extent (see Fig. 9): this may be a consequence of the fact that the core-periphery partition is, in some sense, less “neat” than the bow-tie one (the requirement

that nodes within the IN- and OUT- components have zero in- or out-degree represents a quite strong constraint); only apparently, however, the core-periphery organization seems to require additional information to be explained, as the explicit calculations of the Akaike weights confirms.

A third comment concerns reciprocity: although it plays a role in the definition of the “core” parts (i.e. the SCC and the properly-defined core), its explanatory power is much more limited than expected: as a result, the degree sequence seems to encode all relevant information to reproduce the mesoscale structures considered in the present paper, thus questioning the role supposedly played by some kind of higher-level information - e.g. a partition into blocks - to explain them.

Chapter 3

Detecting Mesoscale Network Structures

The content of this chapter is under review as a paper submitted to EuroPhysics Letters, and made public as a pre-print article on arXiv, see [122].

In Chapter 2 we look at null models for mesoscale network structures. We find that core-periphery and bow-tie structures in real-world economic and financial systems can be modelled with the configuration model, without specifically embedding the modular structure in this model, i.e. the observed modular structure is often not significant under the configuration model. In this chapter we propose a new method to detect these bimodular structures, using a more simplistic null model under which we can identify statistically-significant bimodular structures. It is based on a modification of the *surprise*, recently proposed for detecting communities. Our variant allows for bimodular node partitions to be revealed, by letting links to be placed either 1) within the core part and between the core and the periphery parts or 2) just between the (empty) layers of a bipartite network. From a technical point of view, this is achieved by employing a multinomial hypergeometric distribution instead of the traditional (binomial) hypergeometric one; as in the latter case, this allows a p-value to be assigned to any given (bi)partition of

the nodes. Our method has a clear advantage over some classic methods that search for the model best fitting an observed structure: we employ a benchmark that makes the statistical significance of a given structure explicit. In addition, besides having a method that works for undirected and directed graphs, we have a natural extension to analyse weighted graphs. We apply our method to various economic and financial systems to reveal significant core-periphery and bipartite structures.

3.1 Detecting core-periphery and bipartite structures

The intuitive notion of core-periphery network, as a structure consisting of a densely-connected bunch of nodes (i.e. the core) and low-degree nodes preferentially connected to the core (i.e. the periphery ones) has been firstly formalized by Borgatti & Everett: in [24] a score function indicating the extent to which a given graph partition deviates from an ideal core-periphery structure (whose core is *fully* connected and whose periphery nodes are *only* linked to the core ones) was defined. Several later works adopted the same approach [70, 54, 25], accompanying the error score with a significance level, computed on a properly-generated ensemble of networks (see [41] for a review on the topic). Detection of bipartitiveness has been approached similarly, by quantifying the deviation of an observed graph partition from the ideal bipartite graph with edges existing only between layers and not within them [69, 48].

Conversely, in recent years the detection of mesoscale structures has been faced by adopting a bottom-up approach, i.e. by defining a benchmark model against which to compare the actual network structure: in [133] the authors aim at identifying the most likely generative model that may have produced a given partition, in [13, 14] the authors compare the likelihood values of a Stochastic Block Model tuned to reproduce either a core-periphery or a bipartite structure; similarly, in [77] the authors adopt a Random Graph Model to find multiple core-periphery pairs in networks and in [76] the same authors employ the Configuration Model as a benchmark, showing that a single core-periphery structure can never

be significant under it, seemingly confirming the findings we presented in 2.

In this chapter we illustrate a novel method to detect statistically-significant bimodular structures (i.e. either bipartite or core-periphery ones - as shown in Fig. 3). To this aim, we build upon the results of the papers [3, 4, 90] and on the very last comment that can be found in [118], by adopting a surprise-like score function. Our choice is dictated by the versatility of this kind of quantity, allowing us to consider undirected as well as directed (binary) networks. This is a desirable feature many of the aforementioned algorithms do not have.

3.2 Methods

For traditional community detection, the surprise measure was introduced to represent a simple function whose maximization reveals the community structure of a network [3, 4]. This distribution function evaluates the link-density within clusters and in-between clusters. Where traditional surprise has a good performance in detecting traditional community structures, let us here first discuss the limitations of traditional surprise whenever employed to detect bimodular structures. In what follows we will implement the following definition of surprise [3, 4, 90]

$$S \equiv \sum_{i \geq l^*} \frac{\binom{V_{int}}{i} \binom{V - V_{int}}{L - i}}{\binom{V}{L}} \quad (3.1)$$

(the sum runs up to the value $i = \min\{L, V_{int}\}$) where $V = N(N - 1)$ is the volume of the network, coinciding with the total number of nodes pairs, V_{int} is the total number of intracluster pairs (i.e. the number of nodes pairs *within* the individuated communities), L is the total number of links and l^* is the observed number of intracluster links (i.e. *within* the individuated communities). The hypergeometric distribution shown in Eq. 3.1 describes the probability of observing i successes in L draws (without replacement) from a finite population of size V that contains exactly V_{int} objects with the desired feature (in our case, being an intracluster pair), each draw being either a success or a failure. Surprise is

the p-value of such an hypergeometric distribution, testing the statistical significance of the observed partition against the null hypothesis that the intracluster link density $p_{int} = \frac{l^*}{V_{int}}$ is compatible with the density $p = \frac{L}{V}$ predicted by the Directed Random Graph Model.

3.2.1 The limitations of surprise

While traditional surprise S is suited for community detection, it suffers from several limitations whenever employed to detect bimodular mesoscale structures.

Bipartite networks. Let us first consider a purely bipartite network, as the one shown in Fig. 3, whose first and second layer consist of N_1 and N_2 nodes respectively. Since we would like S to reveal two (empty) communities, we would be tempted to instantiate Eq. 3.1 with the values $V = (N_1 + N_2)(N_1 + N_2 - 1)$, $V_{int} = N_1(N_1 - 1) + N_2(N_2 - 1)$ and $l^* = 0$; upon considering, however, that $L \leq V_{int}$, the explicit computation of S reveals that $S = 1$ (as follows from the Vandermonde identity). Since S is nothing else than a p-value, a significant partition is expected to satisfy $S \leq S_{th}$, with S_{th} usually chosen to attain the value 0.01 or 0.05. In our case, however, the opposite result is obtained: the considered (bi)partition *cannot* be significant, independently from the actual number of connections characterizing the considered configuration. This example highlights one of the limitations of the definition provided in Eq. 3.1.

Star-like networks. Let us now consider proper core-periphery networks: according to the intuitive definition provided in [24], such configurations are characterized by a densely-connected portion, i.e. the core (in the ideal case $c_c \simeq 1$) and a sparsely-connected portion, i.e. the periphery (in the ideal case $c_p \simeq 0$). The density of the intermediate portion is variable, although the chain of inequalities $c_p \leq c_{cp} \leq c_c$ is always assumed to hold. Let us consider a peculiar example of this kind of networks, i.e. a configuration with a fully connected core plus a periphery of nodes, each of which is connected to just one core node (for the mo-

ment, let us suppose that the number of core nodes coincides with the number of periphery nodes - see Fig. 10). Let us now instantiate S on a partition that identifies each periphery node as a community on its own while considering the core as a traditional community. If we consider a core portion of N_1 nodes and $N_2 = N_1$ peripheral nodes, we have $V = (N_1 + N_2)(N_1 + N_2 - 1)$, $V_{int} = N_1(N_1 - 1)$, $L = N_1(N_1 - 1) + 2N_1$ and $l^* = N_1(N_1 - 1)$. In this case, only the addendum corresponding to $V - V_{int} = 3N_1^2 - N_1 = 2N_1 + N_1(N_1 - 1) + 2N_1(N_1 - 1)$ survives, leading to

$$S = \frac{\binom{N_1(3N_1-1)}{2N_1}}{\binom{2N_1(2N_1-1)}{N_1(N_1+1)}} \quad (3.2)$$

which is of the order of 10^{-3} for $N_1 = 3$ and rapidly decreases as N_1 grows (see Fig. 10). Since $S < S_{th} = 0.01$, such a partition is recovered as significant. As confirmed by running the PACO algorithm [90], such a configuration - constituted by an unreasonably large number of single-nodes communities - is indeed recognized as the optimal one. For the sake of comparison, let us calculate S for the “reasonable” partition identifying the core and the periphery as two separate communities: in this case, $V = (N_1 + N_2)(N_1 + N_2 - 1)$, $V_{int} = 2N_1(N_1 - 1)$, $L = N_1(N_1 - 1) + 2N_1$ and $l^* = N_1(N_1 - 1)$. As our explicit calculation reveals, such a partition can indeed be significant but it is not the optimal one (see also Fig. 10).

k-star networks. Let us now generalize the star-like network model, by considering a graph with k peripheral nodes linked to each core node. Instantiating S by considering each group of k leaves as a community on its own leads to

$$S = \sum_{i=l^*}^L \frac{\binom{V_{int}}{i} \binom{V - V_{int}}{L - i}}{\binom{V}{L}} \quad (3.3)$$

with $V = (N_1 + kN_1)(N_1 + kN_1 - 1)$, $V_{int} = N_1(N_1 - 1) + N_1k(k - 1)$, $L = N_1(N_1 - 1) + 2kN_1$ and $l^* = N_1(N_1 - 1)$ (as long as $k \geq 3$, in fact, $L \leq$

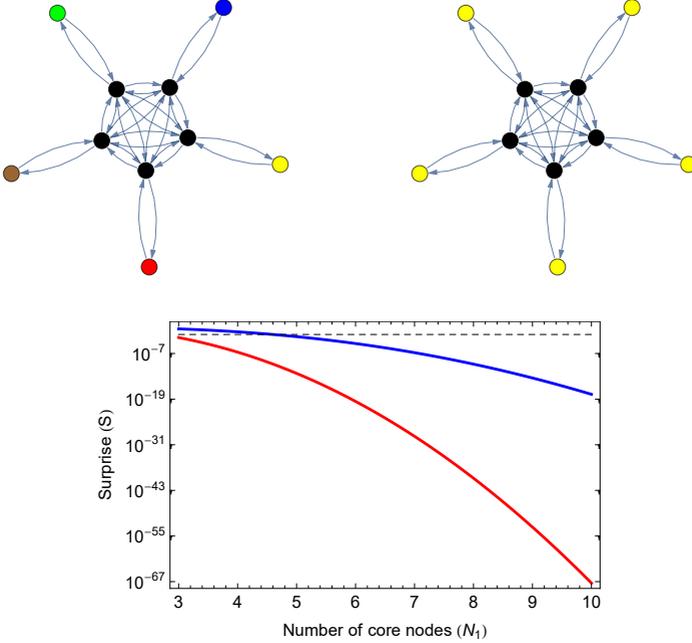


Figure 10: Lower panel: traditional surprise computed for the two partitions shown in the top panels. The red line refers to the partition constituted by 6 communities (top left panel), the blue line refers to the partition constituted by 2 communities (top right panel) and the black, dashed line corresponds to the value $S_{th} = 0.01$. As the number of core nodes is risen, both partitions become increasingly significant; the former, however, is always more significant than the latter. The network configuration shown in the top left panel is, in fact, recognized as the optimal one, as further confirmed by running the PACO algorithm [90].

V_{int}). The expression defined by Eq. 3.3 is significant only under certain conditions: in particular, *a*) for a given N_1 value, as k grows surprise becomes increasingly non-significant; *b*) for a given k value, as N_1 grows surprise becomes increasingly significant. Since the k nodes linked to each core node should be always considered as *non* constituting separate communities, irrespectively from the value of k , the findings above point out another detectability limit of surprise that, for certain values of the

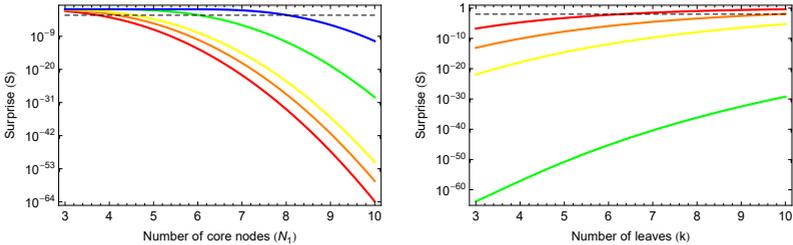
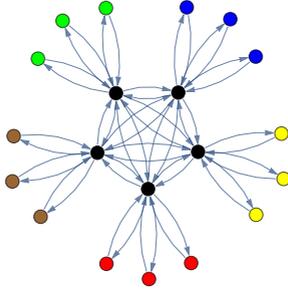


Figure 11: Traditional surprise optimization on a k-star network would lead to identify each group of peripheral nodes as a community on its own, although the intracluster density is zero. More precisely, for a given number of leaves ($k = 3, 4, 5, 10, 20$ as indicated by the red, orange, yellow, green, blue line respectively - lower left panel), as the number of core nodes rises, surprise is found to be increasingly significant. Consistently, for a given number of core nodes ($N_1 = 5, 6, 7, 10$ as indicated by the red, orange, yellow, green line respectively - lower right panel), as the number of leaves rises, surprise is increasingly non-significant (the black, dashed line corresponds to the value $S_{th} = 0.01$). These findings point out the existence of a region of the parameter space where surprise misinterprets the planet partition.

parameters, misinterprets the (planted) partition under analysis - it is, in fact, true that, for small networks, surprise optimization retrieves the aforementioned partition as non-significant; rising the number of core nodes, however, leads it to be detected as significant again (see also Fig. 11).

3.2.2 A bimodular surprise

The previous examples have shown that traditional surprise is still affected by a sort of resolution limit, whenever employed to detect bimodular structures. In order to overcome such a limitation, we introduce a variant of traditional surprise, specifically designed to detect bimodular mesoscale structures.

Whenever community detection is carried out by maximizing the surprise, links are understood as belonging to *two* different categories, i.e. the *internal* ones (the ones *within* clusters) and the *external* ones (the ones *between* clusters). On the other hand, whenever one is interested in detecting bimodular structures (be they bipartite or core-periphery), *three* different “species” of links are needed (e.g. core, core-periphery and periphery links). This is the reason why we need to consider the multinomial version of the surprise, whose definition reads

$$S_{\parallel} \equiv \sum_{i \geq l_c^*} \sum_{j \geq l_{cp}^*} \frac{\binom{V_c}{i} \binom{V_{cp}}{j} \binom{V - (V_c + V_{cp})}{L - (i + j)}}{\binom{V}{L}} \quad (3.4)$$

and that we will refer to as to the *bimodular surprise*. The index c labels the core part and the index cp labels the core-periphery part; whenever considering bipartite networks, the core-periphery portion will be assumed to indicate the inter-layer portion.

The presence of three different binomial coefficients allows three different kinds of links to be accounted for. From a technical point of view, S_{\parallel} is a p-value computed on a multivariate hypergeometric distribution describing the probability of $i + j$ successes in L draws (without replacement), from a finite population of size V that contains exactly V_c objects with a (first) specific feature *and* V_{cp} objects with a (second) specific feature, wherein each draw is either a success or a failure. Analogously to the univariate case, $i + j \in [l_c^* + l_{cp}^*, \min\{L, V_c + V_{cp}\}]$.

Bipartite networks. Let us now calculate S_{\parallel} for the bipartite case considered above, defined by the parameters values $V_c = N_1(N_1 - 1)$ (here, the label c indicates the internal volume of one of the two layers), $V_{cp} =$

$2N_1N_2$, $l_c^* = 0$ and $L = l_{cp}^*$. Since only one addendum survives, our bimodular surprise reads

$$S_{\parallel} = \frac{\binom{2N_1N_2}{l_{cp}^*}}{\binom{(N_1+N_2)(N_1+N_2-1)}{l_{cp}^*}} \quad (3.5)$$

which *can* be significant, as it should be: in fact, a number of inter-layer links exists above which the observed bipartite structure is significantly denser than its random counterpart (see also Fig. 12). Notice that Eq. 3.5 can be directly employed to test the significance of any bipartite configuration where no intra-layer links are observed: in other words, Eq. 3.5 tests the significance of the observed link density in this network portion against the null hypothesis that it is compatible with the one predicted by the Directed Random Graph Model.

Star-like networks. In the case of star-like networks, our parameters read $V_c = N_1(N_1 - 1)$ and $V_{cp} = 2N_1^2$: since, however, $l_c^* = N_1(N_1 - 1)$, the (only) sum indexed by j reduces to the single addendum

$$S_{\parallel} = \frac{\binom{2N_1^2}{2N_1}}{\binom{2N_1(2N_1-1)}{N_1(N_1+1)}}, \quad (3.6)$$

which is $\simeq 10^{-4}$ for $N_1 = 3$ and decreases (the corresponding partition, thus, becomes more and more significant) as N_1 increases. Notice that the traditional surprise would identify a community structure - with each peripheral node counted as a community on its own - with a comparable significance (see also Fig. 11): S_{\parallel} , however, is able to recover the ground-truth structure of the observed network.

k-star networks. Analogously, in the k-star case our parameters read $V = (N_1 + kN_1)(N_1 + kN_1 - 1)$, $V_c = N_1(N_1 - 1)$, $V_{cp} = 2kN_1^2$, $l_c^* = N_1(N_1 - 1)$ and $l_{cp}^* = 2kN_1$. Again, thus, the (only) sum indexed by j reduces to just one addendum

$$S_{\parallel} = \frac{\binom{2kN_1^2}{2kN_1}}{\binom{(N_1+kN_1)(N_1+kN_1-1)}{N_1(N_1-1)+2kN_1}}, \quad (3.7)$$

whose behaviour is shown in Fig. 12: briefly speaking, both in case the number N_1 of core nodes rises, while keeping the number of leaves fixed, and the number k of leaves rises, while keeping the number of core nodes fixed, the bimodular surprise becomes increasingly significant, always recovering the ground-truth partition.

3.2.3 Asymptotic results

The presence of binomial coefficients in the definition of S_{\parallel} may cause its explicit computation to be demanding from a purely numerical point of view. In order to speed up the computation of S_{\parallel} , this subsection is devoted to derive some asymptotic results. Similar calculations for what concerns the traditional surprise have been carried out in [118].

Let us start by considering Eq. 3.5. By Stirling expanding the binomial coefficients appearing in it, one obtains the expression

$$S_{\parallel} = \frac{\binom{V_{cp}}{l_{cp}^*}}{\binom{V}{l_{cp}^*}} \simeq \frac{p^{l_{cp}^*} (1-p)^{V-l_{cp}^*}}{p_{cp}^{l_{cp}^*} (1-p_{cp})^{V_{cp}-l_{cp}^*}} \quad (3.8)$$

having defined $p \equiv \frac{l_{cp}^*}{V}$ and $p_{cp} \equiv \frac{l_{cp}^*}{V_{cp}}$ (see the Appendix for the details of the calculations). The expression above makes it explicit that a given (bi)partition is statistically significant if the link density of the inter-layer network portion is large enough to let it be distinguishable from a typical configuration of the Directed Random Graph Model. In the sparse case, Eq. 3.8 reduces to $S_{\parallel} \simeq \left(\frac{p}{p_{cp}}\right)^{l_{cp}^*}$.

Let us now move to the core-periphery case and consider partitions satisfying the condition $l_c^* + l_{cp}^* = L < V_c + V_{cp}$: in this case, one can derive the result

$$S_{\parallel} \simeq \frac{p^L (1-p)^{V-L}}{p_c^{l_c^*} (1-p_c)^{V_c-l_c^*} \cdot p_{cp}^{l_{cp}^*} (1-p_{cp})^{V_{cp}-l_{cp}^*}} \quad (3.9)$$

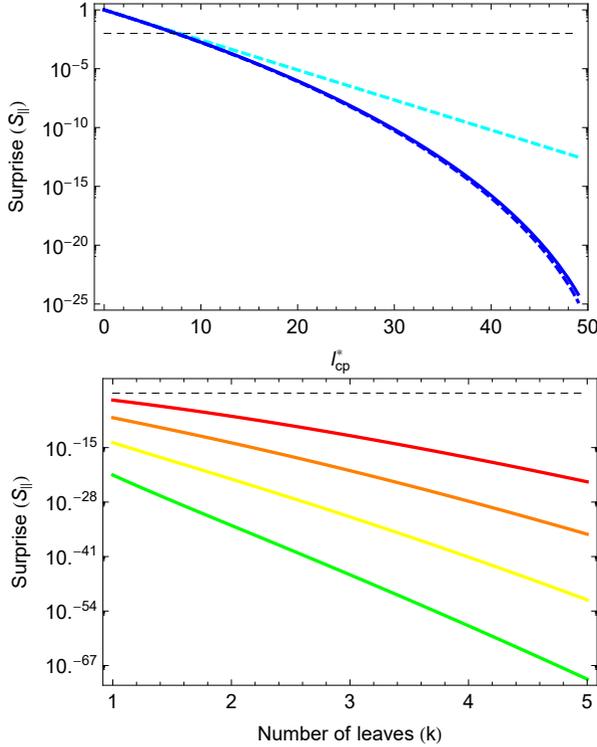


Figure 12: Upper panel: behaviour of $S_{||}$ as a function of l_{cp}^* , for a bipartite network with $N_1 = N_2 = 5$. The blue, solid line corresponds to the (full) expression shown in Eq. 3.5; the blue, dashed line corresponds to the asymptotic expression shown in Eq. 3.8 and the cyan, dashed line corresponds to its sparse-case approximation; for values $l_{cp}^* \gtrsim 10$ (corresponding to link densities $c \gtrsim 0.1$) the sparse-case approximation becomes increasingly less accurate. Lower panel: behaviour of $S_{||}$ for k -star network configurations (star-like networks are recovered as a particular case, when $k = 1$). For a given number of core nodes ($N_1 = 3, 4, 5, 6$ as indicated by the red, orange, yellow, green line respectively), as the number of leaves rises, surprise becomes increasingly significant. The black, dashed line corresponds to the value $S_{th} = 0.01$ in both cases.

having defined $p \equiv \frac{L}{V} = \frac{l_c^* + l_{cp}^*}{V}$, $p_c \equiv \frac{l_c^*}{V_c}$ and $p_{cp} \equiv \frac{l_{cp}^*}{V_{cp}}$. Even if interpreting Eq. 3.9 is less straightforward, it is, however, clear that the

significance of the observed partition is a consequence of the interplay between the link density value of the core and core-periphery regions (the link density of the periphery has been supposed to be zero - see also the Appendix for the details of the calculations). Notice also that as the core density is supposed to be zero as well, $p_c^{l_c^*} = 0$ (a result that is rigorously valid only in the limit $l_c^* \rightarrow 0$), $L = l_{cp}^*$ and Eq. 3.8 is recovered.

3.3 Detecting mesoscale structures in socio-economic and financial networks

Let us now move to analyse real-world systems with this bimodular surprise. We will start with various social and technical network examples, which are widely used in network science as graphs with community structure. Moving beyond these consistency tests, we analyse the bimodular structure of various economic and financial networks. We will employ our novel definition of surprise to understand if the considered networks have a significant bimodular structure (i.e. either bipartite or core-periphery). To this aim, we will search for the partition which minimizes $S_{||}$ (i.e. the optimal one) by employing a modified version of the PACO algorithm [90] whose pseudocode is shown in Appendix. Although, for the sake of generality, our discussion has focused on directed networks, in what follows we will consider directed as well as undirected networks.

Social networks. Let us start our analysis by considering a number of social networks (see Fig. 13). As a first example, let us consider the Zachary Karate Club. Although the latter is commonly employed as a benchmark for community detection, it is also characterized by a clear bimodular structure whose core nodes are represented by the masters, their close disciples and a fifth node “bridging” the two masters. Upon looking at the subgraphs constituted by the masters’ ego-networks, almost ideal (i.e. *a la Borgatti*) core-periphery networks are observable.

A similar comment can be done when considering the network of relationships among “Les Miserables” characters: the main characters

(e.g. Valjean, Javert, Cosette, Marius) belong to the core, while the large number of secondary characters linked to them constitute the periphery of such a network (see, for example, the nodes linked to Valjean); intuitively, again, core nodes are very densely inter-connected while the periphery internal link density is very low. As for the Zachary Karate Club network, there seem to be (core) nodes bridging two dense core subsets.

Let us now consider the (connected component of the) NetSci co-authorship network [127]. A core-periphery structure is, again, recovered (although the core is not very dense) where core nodes represent senior scientists (e.g. Stanely, Barabasi, Watts, Kertesz) and periphery nodes represent younger colleagues, students, etc. It is interesting to observe that the senior scientists share relatively few direct connections, while being connected to a plethora of younger collaborators; even more so, the structure of the co-authorship network seems to reflect the structure of the underlying collaboration network, with each research group seemingly being quite separated from the others.

A fourth social network is the one showing the relationships between US political blogs [2]. Any two blogs are linked if one of the two references the other. As shown in Fig. 14, a core of the most influential blogs (be they republican or democratic), surrounded by a periphery of loosely connected, less important blogs is clearly visible. Differently from the community structure that illustrates republican blogs and democratic blogs as belonging to different groups [72], our core-periphery structure highlights a different organizing principle, based on the blogs overall importance, irrespectively from their political orientation. Interestingly enough, the value of bimodular surprise indicates that a core-periphery structure is more significant than the traditional republicans versus democrats community structure.

Technical networks. A different example of networks is the US airports one (see Fig. 14). Core airports are the ones of New York, Indianapolis, Salt Lake City, Seattle, etc. The periphery airports are preferentially attached to the core ones. It shares interesting similarities with the NetSci co-authorship network: each core airport, in fact, seems to be surrounded

by a quite large number of periphery airports, sharing relatively few internal connections.

3.3.1 Economic networks

Let us now consider an economic network, i.e. the World Trade Web (WTW) in the years 1960, 1980 and 2000: as usual, nodes are world countries and links are trade relationships (i.e. exports, imports) between them. Upon running our bimodular surprise optimization we find a clear core-periphery structure with the core including the richest countries and several developing nations and the periphery including some of the poorest nations (e.g. several African nations throughout our dataset - see also Fig. 15). We also observe an interesting dynamics, causing the core size to rise (it represents the $\simeq 30\%$ of nodes in 1992 and the $\simeq 60\%$ of nodes in 2002) and progressively include countries previously belonging to the periphery. Such a dynamics - that can be interpreted as a signal of ongoing integration - confirms the results found in [123], where it was shown that the size of the WTW strongly connected component (SCC) increases with time as well. Although the SCC and the core portion of the World Trade Web do not perfectly overlap, many similarities between the two structures are indeed observable.

For a second economic network, we analyse the the production network of the United States. This network shows the input-output relations between sectors in the US [34]. Input-output networks show production networks on a sectoral level. The nodes in the network represent sectors of the US economy and edges indicate an input-supply between the sectors. Typical sectors here are: wholesale trade, advertisement and road transportation. Data comes from the US Bureau of Economic Analysis and describes the input-output data for the year 2002 [126]. In the network observe a particular core-periphery structure, where the core nodes have a relative small in-degree, but a relative higher out-degree, see Fig. 16. The core of the network thus consists of sectors that are mainly at the origins of the production network, although vastly connected to all other sectors.

3.3.2 Financial networks

Let us now consider a financial network, i.e. e-MID, the electronic Italian Interbank Market. Here, we compare two different datasets: the first one collects the 2005-2010 interbank transactions during the so-called maintenance periods [65]; the second one collects interbank transactions on a daily basis from 1999 to 2012 [13, 14]. The main difference between the two datasets lies in their level of aggregation: notice, in fact, the first one basically collects data on a monthly basis.

Let us start by analysing the first dataset. As Fig. 17 shows, its structure undergoes an interesting evolution: after an initial period of two years, where a large periphery of loosely connected nodes ($\simeq 70\%$) exists, a transient period of one year (i.e. 2007) during which the percentage of nodes belonging to the core rises, is visible. Afterwards, an equilibrium situation seems to be re-established with the percentage of core and periphery nodes basically coinciding. Even if the total number of banks registered in the dataset steadily decreases after 2007, this doesn't seem to affect the type of banks belonging to the core and to the periphery, i.e. Italian and foreign banks, respectively.

Let us now move to the analysis of the second dataset. As Fig. 18 shows, the link density analysis of the portions in which our bimodular surprise partitions the network reveals that, overall, a core-periphery structure seems to better characterize the daily data than a bipartite structure (two snapshots of the network are explicitly shown, illustrating the values of link density characterizing the three network portions). It should however be noticed that this picture seems to be less correct from 2008 on. In the last portion of the first panel of Fig. 18, in fact, it can be observed that a bipartite structures occur more often.

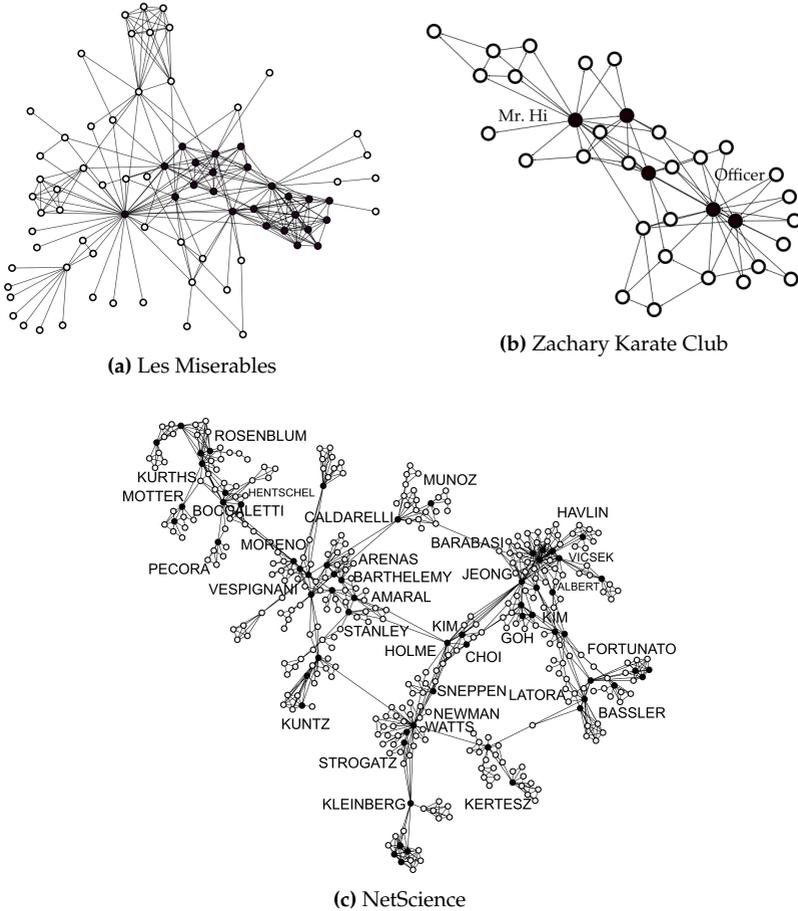


Figure 13: Bimodal structure of three real-world social networks (core nodes are drawn in black and periphery nodes are drawn in white). The core-periphery structure of the network of relationships among “Les Miserables” characters; the main characters (e.g. Valjean, Javert, Cosette, Marius) belong to the core. The Zachary Karate Club shows that while the two masters (plus some close disciples) belong to the core, the remaining disciples create a periphery around them, shaping a structure that is reminiscent of the Borgatti & Everett ideal structure [24]. In the NetScience co-authorship network we observe that while the senior scientists belong to the core - although sharing few direct connections - younger colleagues/students belong to node-specific peripheries connected to the former ones.

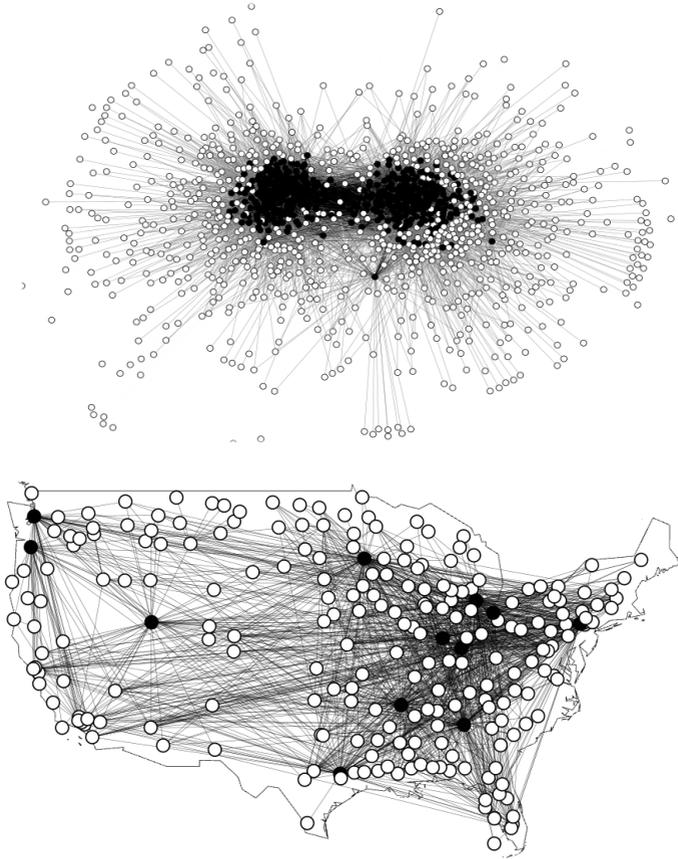


Figure 14: Upper panel: core-periphery structure of the US political blogs [2]: a core of the most influential blogs (be they republican or democratic), surrounded by a periphery of loosely connected, less important blogs is clearly visible. Notice that blogs are grouped independently from their political orientation. Lower panel: core-periphery structure of US airports. As for the NetSci co-authorship network, each core airport seems to be surrounded by a quite large number of periphery airports, sharing relatively few connections between themselves.

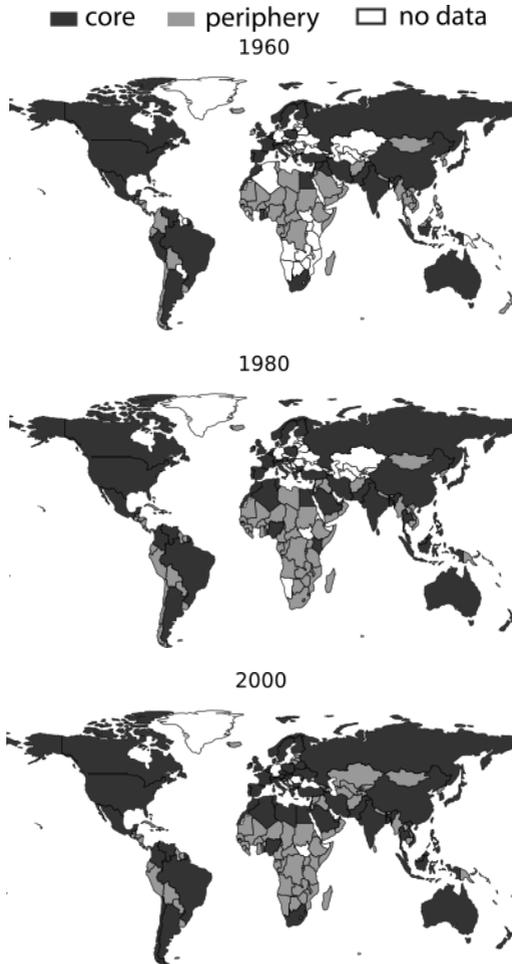
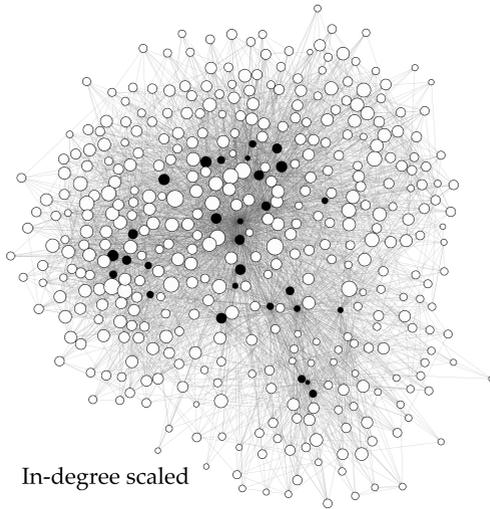
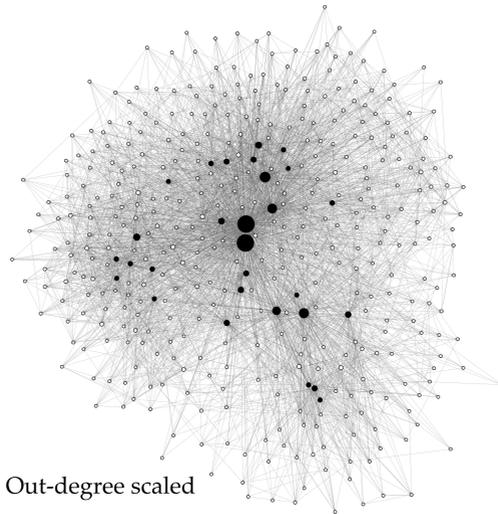


Figure 15: Core-periphery structure of the World Trade Web (black: core nodes; gray: periphery nodes). Loosely speaking, while the richest and several developing countries are found to belong to the core, the poorest nations belong to the periphery (e.g. several African nations, throughout our dataset). Notice that core size increases with time: apparently, thus, the system becomes increasingly integrated, confirming a result found in [123], where it was shown that the size of the WTW strongly connected component increases with time as well.



In-degree scaled



Out-degree scaled

Figure 16: Core-periphery structure of the US production network. Both graphs show the same network layout and core-periphery structure, but the node size is scaled by the in/out degree. We observe a particular core-periphery structure, where the core nodes have a relative small in-degree, but a relative higher out-degree. The core of the network consists of sectors that are mainly at the origins of the production network, although vastly connected to all other sectors.

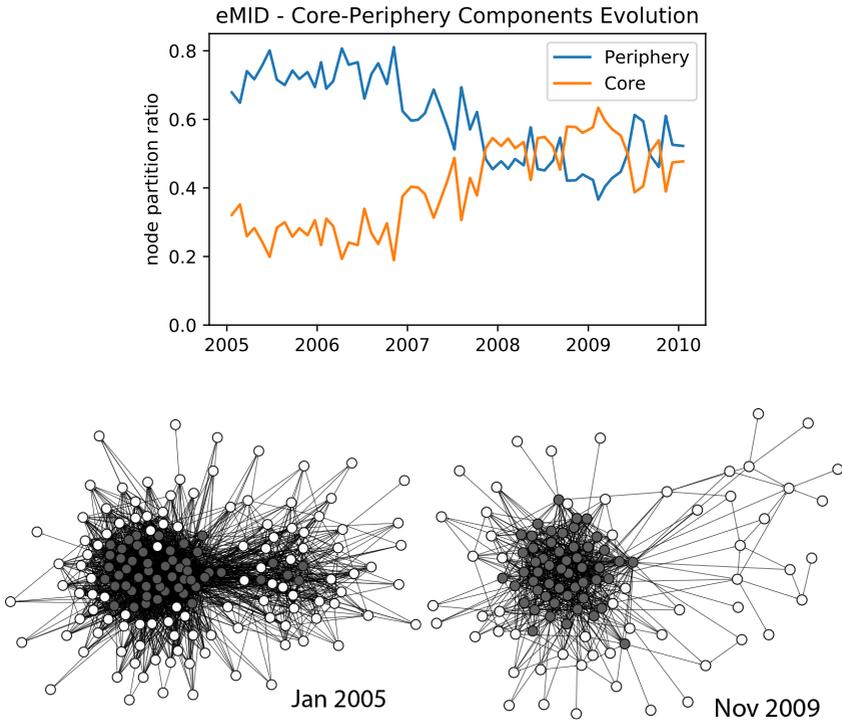


Figure 17: Core-periphery structure of e-MID maintenance periods (gray: core nodes; white: periphery nodes). After an initial period of two years characterized by an approximately constant value of the core and periphery size, a structural change takes place in 2007 and the percentage of nodes belonging to the core steadily rises until 2008. Afterwards, an equilibrium seems to be re-established. This may be due to a decrease in the total number of nodes which, however, does not affect the type of banks belonging to the core (Italian banks) and to the periphery (foreign banks). Networks are directed but we have omitted the link directionality for the sake of readability.

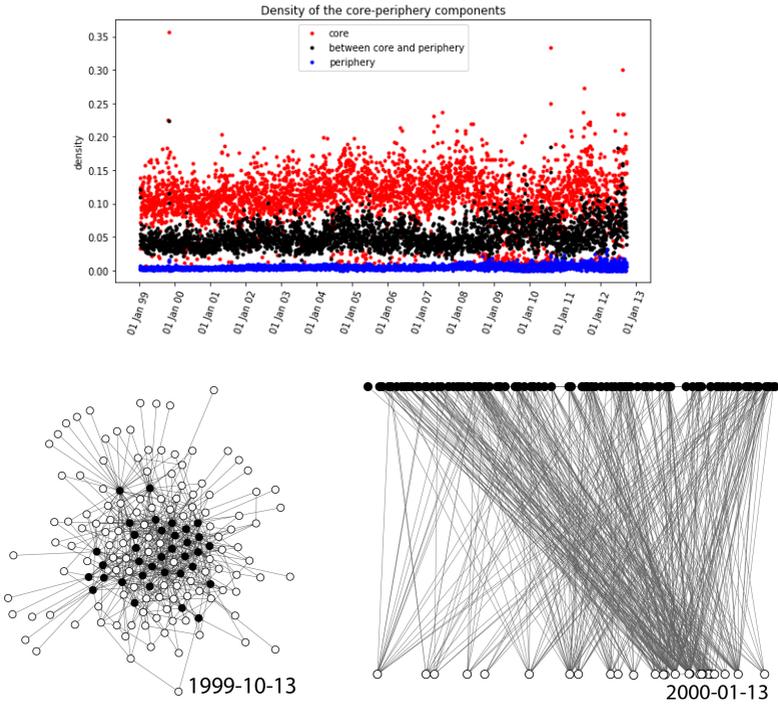


Figure 18: Mesoscale structure of e-MID daily data (gray: core nodes; white: periphery nodes). During the first snapshot, the chain of inequalities $c_p < c_{cp} < c_c$ holds. During the second snapshot, instead, a configuration for which $c_p \simeq c_c < c_{cp}$ is observed, indicating the presence of a bipartite structure (when referring to bipartite structures, the label cp will be assumed to indicate the inter-layer portion). Although for the vast majority of snapshots a core-periphery structure seems to be better represent the e-MID structure, the number of times a bipartite structure is observed increases after 2008. Networks are directed but we have omitted the link directionality for the sake of readability.

3.4 Discussion

It is hard to underestimate the importance of the presence of core-periphery structures in real-world networks: while the authors in [95] show that the most robust topology against random failures is the core-periphery one, understanding the relationship between a given node systemicness and its coreness is of paramount importance in finance [83]. In the same field, a core-periphery structure is believed to reflect the “essential” function of banks: the core ones tie the periphery ones into a single market through their intermediation activity [40].

Here we have proposed a novel measure for bimodular mesoscale structures detection. To this aim we have adopted a surprise-like score function, by considering the multivariate version of the quantity proposed in [90]. Employing this kind of quantities implements a bottom-up approach, where the modular structure is extrapolated from the data and not imposed *a priori* as in previous approaches [24, 14].

Most importantly, such a comparison is based on a properly-defined null model, allowing the significance of a partition to be quantified via, e.g. a p-value. As for the traditional surprise, the reference model is the Directed Random Graph Model that constrains the total number of observed connections, while randomizing everything else. The choice of employing such a benchmark is dictated by a number of recent results, pointing out that several mesoscale structures of interest (e.g. the core-periphery one, the bow-tie one, etc.) are actually compatible with - and hence undetectable under - a null model constraining the entire degree sequence [70, 76].

While solving the problem of consistently comparing an observed structure with a “random” model of it, our approach also solves a second drawback affecting the methods in [40, 24] and pointed out in [76]: ideal structures as the ones searched by algorithms *a la Borgatti* are very reliant on the nodes degree, with the core often composed of just the nodes with the largest number of neighbors. This is not necessarily true when a benchmark is adopted for comparison [133]: as previously discussed, the significance of a given partition detected by surprise results from the

interplay between the link density values of the different network areas.

This also sheds light on the relationship between apparently conflicting structures co-existing within the same network configuration: generally speaking, traditional and bimodular surprise optimization should be considered complementary (rather than mutually exclusive) steps of a more general analysis. As the example of the US political blogs confirms, it is indeed possible that a community structure co-exists with a core-periphery structure; a second, less trivial, example is provided by the World Trade Web, whose community structure has been studied in [12] but whose significance has, then, been questioned [96].

The two approaches to mesoscale structures detection that, so far, have been proposed in literature, i.e. comparing an observed structure with a benchmark [77, 76] and searching for the model best fitting a given partition [133, 13, 14, 72] can be supposed to be complementary, since a non-significant structure under a given benchmark is surely more compatible with it. Employing a benchmark, however, actually provides an advantage, i.e. making the statistical significance of a given structure explicit - something that remains “implicit” when employing the fitting procedure. In other words, searching for the best fit pushes one to enrich the model with an increasing amount information whose relevance cannot be easily clarified. Such a problem seems to affect all likelihood-based algorithms unless a more refined criterion to judge the goodness of a fit is employed: solutions like the one of adopting criteria like the Akaike Information Criterion *et similia* have been proposed [26].

The present work calls for a generalization to *weighted* mesoscale structures detection, a field where relatively little has been done so far [90, 49]. We present the first preliminary results of our work in this direction in appendix B.

Chapter 4

Unraveling the Multilayer Structure of Corporate Networks

The content of this chapter was published as an invited paper for the Focus on Multilayer Networks of the New Journal of Physics as [121].

Various company interactions can be described by networks, for instance the ownership networks and the board membership networks. The meso-scale structure of some of these corporate networks has been described as core-periphery and bow-tie. Using the bimodular surprise introduced in the previous chapter, we find that the aggregate of these interaction indeed shows a significant core-periphery structure. In this chapter we unravel this corporate network by analysing the different layers making up the aggregate corporate network. To this end we construct a new multiplex network of interactions between companies in Germany and in the United Kingdom, combining ownership links, social ties through joint board directors, R&D collaborations and stock correlations in one linked multiplex dataset. We describe the features of this network and show that the (mesoscale) structure of the different types of connection complement each other and together make an even smaller corporate

world than previously reported.

4.1 Corporate networks

Under corporate networks we consider all networks that describe interactions between companies. We have discussed how important the topology of economic and financial networks is for our understanding of these systems and phenomena like systemic stability. Corporate networks have also been analysed for such phenomena. In [16, 43] the authors study spreading processes of influence on the network of interlocking board members. Through board members that work for multiple companies, decision-making spreads and this network has been shown to exhibit herd behaviour. These board interlocks can limit competition as they obstruct independent decisions by boards [132].

Ownership networks constructed from (partial) ownership stakes of corporate entities are another type of corporate networks. These networks have revealed a small corporate world; a large concentration of ultimate ownership in a small group of core companies in this network [125]. In [101], ownership networks are studied to find the role of subsidiaries in control of large parent companies.

Also innovation dynamics have been studied by networks of R&D partnerships, showing for example the effective outsourcing of research by big corporations to start-ups [116]. On the financial side, corporate networks can be constructed from stock price correlations. With techniques like network backbone extraction one can identify influential companies from these networks [119, 23]. These studies have shown how networks can be used to study the dynamics of competition and innovation, or identify the influence of the corporate topology on some notion of control and influence between corporations.

Most research until now studied these different types of corporate networks in isolation. However, corporate networks are strongly interconnected; e.g. a cascading effect between board members will influence stock market fluctuations and vice versa. We therefore argue that these

systems should be studied in parallel, not in isolation. In this chapter we will unravel the aggregate corporate network and characterise the topology of the different layers of this system. To this end we construct a new multilayer network of interactions between companies. We identify four main connections: *ownership links, social ties through board members, research collaborations, and stock correlations.*

4.2 Data

We construct a unique multiplex dataset where we go beyond just listed companies and include all registered companies in Germany and in the United Kingdom and Ireland. The company information comes from the Amadeus database from Bureau van Dijk, this includes stock prices, ownership details, names of directors in the boards, patent data, and audit firm details. The data reflects the state of all registered companies in those countries as in February 2018. Daily stock prices are collected in the 2-year window before this date.

We select companies registered in Germany with at least 10 employees (if employee data is not available this is estimated by taking the industry average based on revenue). For the United Kingdom and Ireland we study all listed companies from data that comprises 1312 companies that are listed and registered within these countries.

Networks with links of different types are described by multilayer networks, where every layer corresponds to the interaction graph of a single type of interaction. The subset of these networks where there are no edges between nodes in different layers, is also called multiplex networks. In recent years both theory and applications of multilayer networks has shown that this richer description of a network can give a better representation than aggregate of single layer networks [45, 75, 22]. An example is the separation of long-term and short-term interbank exposures that can lead to a different estimation of systemic risk in the banking system [9]. This is driven by the fact that there are often multiple drivers of an effect, and as these are often connected but have different structures, the single or aggregate layers might under- or overestimate

the network effects.

4.2.1 Networks

- *Ownership.* Ownership ties are constructed from shareholder data [61]. We are interested in cross-shareholding for companies in our dataset, i.e. instances where the shareholder of a company is of itself a company in our data. We thus disregard all shareholders that are foreign or individuals.

Ownership can be defined in various ways. Strict ownership is usually set as stake of $> 50\%$. For control and influence a smaller stake (5, 10, 20%) is often considered sufficient [91, 97]. Peer effects of ownership can be robust under these different definitions [125]. We construct an edge between companies i and j when company i has a stake $\geq 10\%$ in company j following [125].

- *Board of Directors.* The board membership network describes companies that are linked through directors. The board makes strategic decisions for the company. Decision making can spread through directors that sit in multiple boards [16]. In this network an edge exists between company i and j when the boards of directors of company i and company j have a least one director in common. This can also be seen as the company-side projection of the bipartite network of directors and the companies they serve. We rely on unique person identifiers in the database, which prevents any problem with name disambiguation.
- *Research.* From patent application data we obtain two kinds of links; (i) companies that are joint assignees (owners) of a patent and (ii) inventors that worked for multiple companies at the same time. The first case is a clear sign of joint research which led to the joint patent application. Following other studies on R&D networks [63, 131], we also exploit inventors that appear on multiple patents with different assignee companies, i.e. inventors working with multiple companies. This can be used as a proxy of joint research [33]. Be-

cause of the long term characteristics of joint research projects, we use the patent application data between 2016-2018 (such a 2-year backward window is also used in [63, 33]).

- *Stock Correlations.* From stock price time-series we can identify significant relations between stocks. Here we follow earlier studies [23, 84], constructing a minimum spanning tree to identify a backbone network of important links. The method extracts a backbone network from pairwise correlations between the time-series. The pairwise correlations are computed on the logarithmic daily returns of a stock price. For stock s with price p , the daily log-return is : $r_t^s = \log p_t^s - \log p_{t-1}^s$. We use a 2-year window of the stock time-series in order to have $n_s < n_t$, for a well behaved multiple pairwise correlation matrix (in the Appendix we show that our results are robust also under a shorter time window). The 2-year window in combination with daily returns, rather than daily closing prices make the time series more comparable and diminishes effects of long term trends. In a further regression exercise we also use the full set of pairwise correlations without extracting the backbone network.

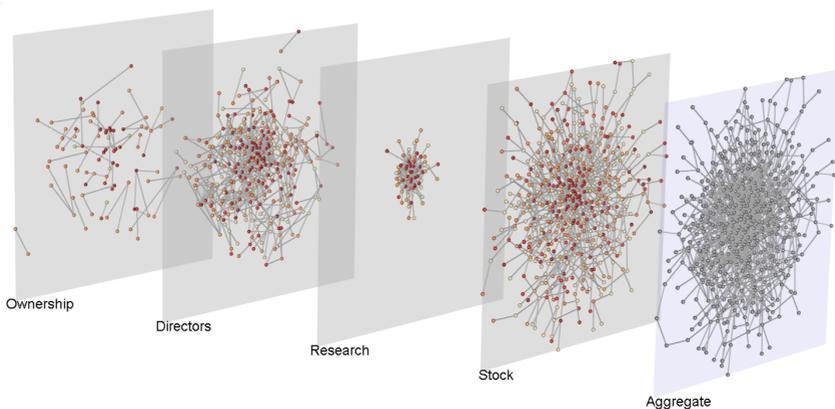


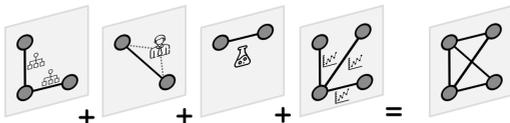
Figure 19: Multilayer network of listed German companies showing: cross-stockholding ownership network, joint board directors network, research interactions, minimum spanning tree of stock correlations and the aggregate. We show a multiplex layout; *the position of nodes is the same throughout the layers*, but isolated nodes in a layer are not shown. Nodes are coloured by their MultiRank centrality in the multiplex. The visualization illustrates clearly that these interactions have different structures. We observe for example that the research collaborations are really concentrated in the core of this multiplex. All multiplex figures are created with the MuxViz software [44]. This figure has been published under CC licence in [121].

4.3 Network topology of the corporate multilayer network

In this section we first describe the network topology of the various company-to-company interactions. We then discuss the significance of the multiplex structure and show that all layers convey different structural information.

4.3.1 Layer descriptions

The network of German companies results in a multiplex of over a hundred thousand companies, where the layers describing different types of company interactions are all described by different networks statistics.



		Ownership	Board	Research	Stock	Aggregate
German companies	Nodes	49385	90690	3744		105005
	Edges	38594	308123	8420		335213
	Average Degree	1.56	6.80	4.50		6.38
	Assortativity	-0.03	1.00	-0.18		0.98
Germany listed companies only	Nodes	86	330	64	535	535
	Edges	54	504	212	534	1214
	Average Degree	1.26	3.05	6.62	2.00	4.54
	Assortativity	-0.18	0.48	-0.42	-0.03	0.29
UK listed companies	Nodes	301	1043	25	1312	1312
	Edges	293	2778	25	1311	4354
	Average Degree	1.95	5.33	2.00	2.00	6.64
	Assortativity	-0.51	0.69	-0.19	-0.21	-0.11

Table 2: Multiplex descriptive statistics

For all German companies that have at least one connection to another company, we obtain an aggregate network of 105005 companies. 89% of the companies are connected through a joint board member, and slightly less than half of the companies are connected through an ownership link, see Table 2. The network of research interactions is much smaller compared to the other layers, but with an average degree of 4.5, this is more densely connected than the ownership topology which has only 1.26. The largest contribution of links in the aggregate network comes from board interactions. From the assortativity of the layers, the tendency of nodes to connect with nodes of the same size (degree), we see that while through board members companies connect very much to similar size companies, this effect is non existing in the ownership network. In the research network assortativity is negative, -0.51, driven by larger companies that collaborate with smaller companies (start-ups).

Listed companies

We will focus on listed companies, as more layers of the multiplex structure are specified. The subset of German companies that are listed has similar descriptive statistics to those of all German companies. When we compare the characteristics of the subset of listed companies to those of the whole system, many of the dynamics between the layers are the same, see Table 2. I.e. the relative differences in number of nodes, edges, average degree and assortativity between the layers is broadly conserved. This is in principle not surprising, however, in this case the subset comprises of a particular set of the most 'important' and mainly biggest companies in terms of turnover etc. Nearly all companies are connected by board members, showing the potentially large influence of this more informal network.

United Kingdom

For listed firms in the UK we observe a reasonable similarity to the German network, in terms of the assortativity of the different interactions and the relative sizes of the different layers. Main differences are a more dense board membership network, with an average degree of 5, and less research interactions, as the UK economy in general is more geared towards less research intensive services than the German economy, see Table 2.

Core-periphery structure of the aggregate corporate network

With the the bimodular surprise method introduced in the previous chapter we analyse whether this corporate network has a core-periphery structure. Fig. 20 shows the core-periphery structure as revealed by modular surprise on the aggregate of the multiplex of German listed companies.

4.3.2 Multiplex structure

In the previous section we described the topology of the isolated layers, but we are interested in these interactions in parallel. We want to un-

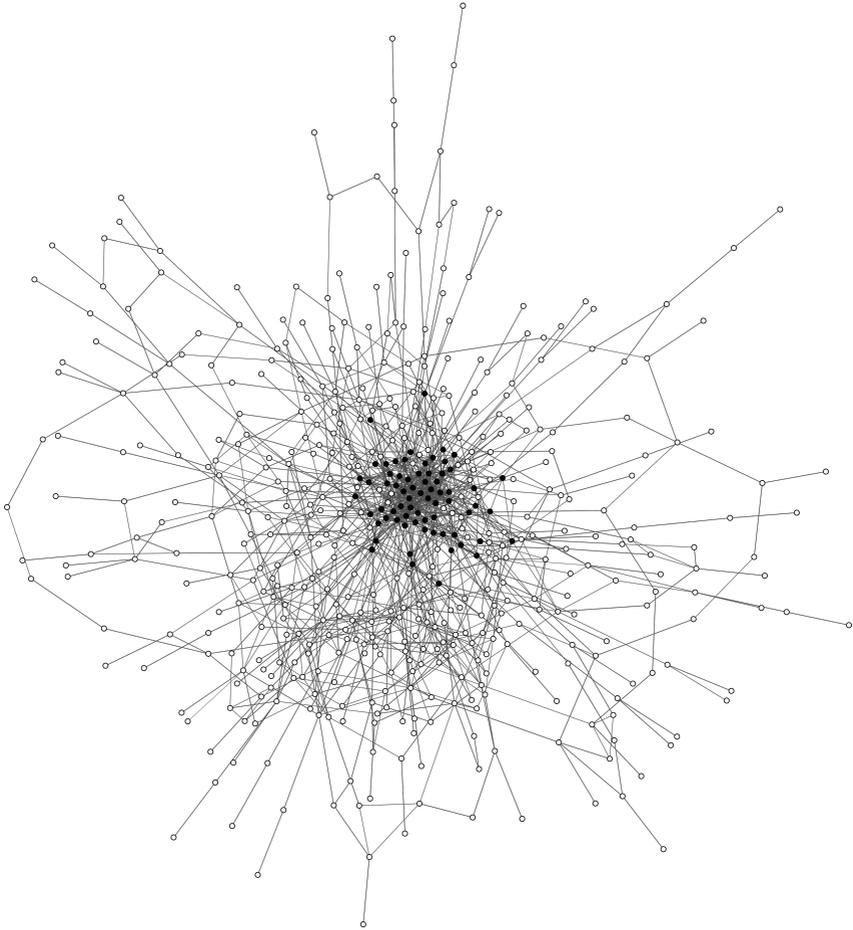


Figure 20: Core-periphery structure of the aggregate of the German multiplex of listed companies as revealed by bimodular surprise.

derstand whether we need to distinguish these different interactions, or whether the aggregate would suffice. As the multiplex literature is a recent one, there is not one established method to evaluate this question of the relevance of the multilayer structure. Therefore we answer this with three different available methods for the analysis of the multiplex structure:

1. the node and edge overlap of the layers;
2. the structural reducibility;
3. a network regression.

All these three methods on their own will show that all layers are significant and that the multiplex representation has an added value.

Node and edge overlap

Node and edge overlap shows similarities between the layers and the level of connectedness. The node and edge overlap in two networks is measured by the fraction of nodes (edges) that occur in both networks over the aggregate number of nodes (edges) in those two networks. Results in Table 3 indicate that all layers are connected since a significant fraction of the nodes overlaps. The overlapping edges are however fewer. This holds both for the German companies and the UK companies. The low edge overlap indicates that the different layers do not replicate many connections specified in other layers and thus that *these corporate networks do not simply overlap*.

Structural reducibility

Structural reducibility is a recently introduced measure [45] for multi-layer networks that indicates whether pairs of layers can be aggregated based on redundant information. The information encoded in a network can be quantified by the entropy. The structural reducibility quantity calculates the relative entropy between a network of multiple layers and its aggregate. By analysing whether some layers can be aggregated, without losing distinguishability from the aggregate, one can find the configuration of the multilayer that maximises the information in the system. The information is quantified by the entropy (Von Neumann entropy) of the network. Formally we maximise the value

$$q(G) = 1 - \frac{\text{combined entropy of the multiplex}}{\text{entropy of the aggregated network}} \quad (4.1)$$

(a) Node overlap of German multiplex

	Ownership	Board	Research	Stock
Ownership	1			
Board	0.767	1		
Research	0.190	0.174	1	
Stock	0.742	0.950	0.170	1

(b) Edge overlap of German multiplex

	Ownership	Board	Research	Stock
Ownership	1			
Board	0.067	1		
Research	0.015	0.044	1	
Stock	0.007	0.016	0.023	1

(c) Node overlap of UK multiplex

	Ownership	Board	Research	Stock
Ownership	1			
Board	0.235	1		
Research	0.019	0.023	1	
Stock	0.236	0.827	0.020	1

(d) Edge overlap of UK multiplex

	Ownership	Board	Research	Stock
Ownership	1			
Board	0.006	1		
Research	0	0.001	1	
Stock	0	0.009	0	1

Table 3: Layer overlap. Overlap of nodes and edges between the layers, as measured by the fraction of nodes/edges which appear in both layers over the aggregate number of nodes/edges of the two layers. Edge overlap is smaller than the node overlap, indicating layers are complementary. Nodes can be the same in the layers but the connections between them are different, this creates larger connected structures.

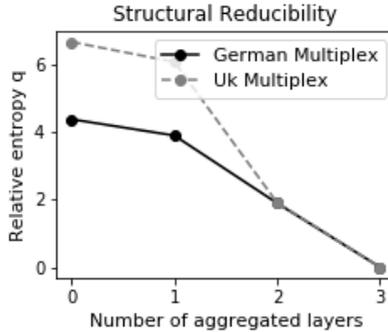


Figure 21: Structural reducibility of the multiplex based on entropy of the layers. The relative entropy q measures the structural information of the multiplex compared to its aggregate. The larger q , the more distinguishable the multiplex is from the aggregate network. When none of the layers are aggregated the entropy is maximal. I.e. none of the layers are structurally reducible based on redundant structural information measured from the entropy. For a complete description of the multiplex we do require all information encoded in all layers/types of interactions. This figure has been published under CC licence in [121].

See the Appendix for details and the formal introduction of this method. Calculating the structural reducibility quality between all the layers, we find the the optimal structure is the multiplex with all original layers present, see Fig. 21. This indicates that *all layers convey different structural information*.

Multiplex network regression

A third way we can probe the significance of the different layers is by means of a regression. The multiplex network describes different types of interaction between the actors. With a regression we can test the explanatory power of these interactions for an observed interaction. For sparse networks a classic regression, where observations are formed by all possible edges (all adjacency matrix entries), one would regress with many zero entries. We use a newly proposed method more suited for network regressions that uses a graph null model to look at significant

links. It is well known that interactions described by a network structure should be tested against a null model, to see which part of the observed network can be explained by randomness. With a null model we answer the question: out of the many combinations the network structure could possibly be configured, how (im)probable is the empirically observed structure? This way we can filter out observed interactions that arise from randomness from combinatorial factors.

To this end we use a recently introduced network regression model that uses a generalised hypergeometric graph ensemble as the graph null model. This regression estimates the influence of a layer of the multiplex and, using the null model, tests the statistical significance of the layer structure on observed interactions. A detailed explanation can be found in [35]. We setup a regression to see if the stock correlations can (in part) be explained by the network structures:

$$\text{stock correlation} \sim \text{ownership ties} + \text{board ties} + \text{research ties} \quad (4.2)$$

As dependent variable we take the pairwise correlation values between all stock time-series and scale them to integer values in the interval $[0, 100]$, as the multiplex regression model allows for the dependent variable to be weighted (positive, integer) - See [35] for details on this regression method. The dependent variables are the unweighed network structures; the Ownership network, the Board network and the Research network. *The regression identifies all dependent variables to be significant in explaining the stock correlations*, see Table 4. In C.2 we show that this result is robust with respect to the 2-year window of the stock data as discussed in Section 4.2.

On the basis of the above we conclude that the different layers have distinct information to convey. The three different methods have shown the significance of all layers in the multiplex. These results highlight that the different channels of influence between companies have very different structures. The complementary characteristics of the different layers show that the corporate world is an even smaller one than previously reported based on the studies of ownership [15] or the board membership

Table 4: Multiplex network regression on pairwise stock correlations. All layers are found significant independent variables for explaining the stock correlations.

	German stock correlations		UK stock correlations	
	Coef.	SE	Coef.	SE
Board	0.44	***	0.49	***
Ownership	0.89	***	0.72	***
Research	0.40	***	1.39	***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

network [43].

4.3.3 Network structure and company characteristics

We now take the first steps to explore the interaction of the network structure and company characteristics of this multilayer structure. The small world network that we described before is characterised by the presence of hub nodes. Such core nodes act as connectors of different parts of the network. We calculate the centrality of the nodes within the separate layers, and with the MultiRank centrality [98] quantify the node prominence in the multiplex. We regress node centralities on company characteristics like revenue, stock return, revenue growth, and the Sharpe ratio of the stock returns, in an ordinary least squares regression. We find that centrality in the multiplex is significantly correlated to revenue of the companies - see the appendix for more details. Larger companies, as measured by revenue, are more central in the network. The result is not necessarily surprising [125], as larger companies have more resources to form connections, like subsidiaries and R&D collaborations.

Now, we look at the position of the companies in the network, related to the stock performances. We measure stock performance with the Sharpe ratio instead of the direct stock returns. This ratio evaluates mean returns compensated for (high) volatility of the stock. A high Sharpe ratio corresponds to a high return and low volatility: a consistent steady

high return. The Sharpe ratio is calculated from the returns of a portfolio of stocks r as:

$$S = \frac{\langle d \rangle}{\sigma_d}, \quad (4.3)$$

where $d = r_{\text{portfolio}} - r_{\text{risk free}}$. We take the risk free return on capital as zero (an assumption which is valid as the return on German governmental bonds is currently zero, or even negative). We calculate the centrality also in the multiplex excluding the stock interactions layer.

Because not all layers are well connected with the presence of a giant component, we qualify the core-periphery structure by centrality rather than with the bimodular surprise. Calculating the Sharpe ratio of the portfolio of the highest ranked quantile, using the MultiRank centrality ranking on the multiplex network, we find the core companies perform significantly better compared to the rest of the network. In fact, we find a Sharpe ratio for the portfolio of companies in the core of 0.37, while only 0.18 for the rest of the network for the German multiplex (0.21 versus 0.07 for the UK multiplex). However, this effect seems mostly driven by the companies with largest revenue, rather than uniformly by all companies in the core. These preliminary results on the dynamics behind the multiplex structure indicate a relation between company performance and network formation, and invite more research on this topic.

4.4 Discussion

In this chapter we have described a uniquely compiled dataset which combines various known company-to-company interaction networks into one single multiplex structure. The layers of this system describe different types of interactions between the same set of companies. We have included ownership ties, social ties through joint board members, R&D collaborations, and stock correlations. With three separate methods we show the significance of the multiplex structure. Node and edge overlap highlights that different types of ties connect different sets of players; i.e. the structures are not overlapping and the layers complement each other. The structural reducibility quality was used to show that all layers

are structurally different and irreducible from an information theory perspective. In the third method we used a regression model to estimate the explanatory power of the multiplex structure on all pairwise stock correlations. The independent variables of network structures of ownership network, board membership network and R&D network were all found to be significant estimators of the structure in the stock correlations.

These three methods confirmed that the multiplex representation is different to the single layers or the aggregate, and that these interactions have different structures. For company interactions this indicates studies of peer effects of control should take these multiple connections into account. We evaluated the characteristics of companies related to the multiplex structure. These initial results indicate a relation between company performance and multiplex centrality.

Our results show that the corporate world is an even smaller world than the small world already described by various previous studies on corporate control and studies of the 'old boys network' of board rooms. The significance of the different layers of the corporate multiplex invite more research on the interconnectedness of diverse economic and financial networks.

Chapter 5

Maximum Entropy Approach to Link Prediction in Bipartite Networks

The content of this chapter is published as a pre-print article on ArXiv, see [8].

The same maximum entropy framework we used in Chapter 2 for reconstruction of network structures from limited information, can be exploited for link prediction. In fact, as stated before, this framework has been successfully applied in detection of significant structures, reconstruction, pattern recognition in networks, and recently to link prediction in monopartite networks. We will now extend this work to bipartite graphs. We test our method on two real world networks with different topological characteristics. Our performances are compared to state-of-the-art methods, and the results show that our entropy-based approach has a good overall performance.

5.1 Introduction

Until now we discussed monopartite networks; the connections between one class of nodes. Within these systems we briefly discussed bipartite networks; those describing the connections between two groups of nodes in the mesoscale structure e.g. connections between core and periphery. In general, bipartite networks describe the connections between two different classes of nodes, as customer-products [79] or countries and their exports [113, 103]. In fact, often monopartite networks are actually a projection of a bipartite network; a co-author network is the author-side projection of the bipartite network of authors and publications.

As argued for mesoscale structures, the network topology is relevant for many networks processes, such as diffusion phenomena and network resilience. Incomplete or incorrect knowledge over the network topology can cause biases in such analysis. Unfortunately, in real-world networks, the relationships among nodes are not always fully observable, and are subject to frequent changes over time. To overcome these issues, the objective of link prediction is to uncover unobserved or missing connections or forecast the emergence of future relationships from the current topological structure of the network [81, 32, 92].

The link prediction problem is an active research field and many methods have been proposed in the literature. Some methods make use of *local* information, i.e. at node level, while others are based on *global* approaches. In the following we will concentrate on the first class of methods. Also, we can distinguish methods based on similarity measures or likelihood functions. However, only few of the methods proposed in the literature have been applied in the case of bipartite networks [130, 42, 58]. Among the algorithms which admit bipartite configurations, there are several classes of techniques, such as global and kernel-based methods [78], extensions of results in monopartite networks to bipartite [42] and projections on the monopartite [58, 130].

In a recent work, the entropy-based approach as introduced in Chapter 1 was used for link predictions in (monopartite) trade networks, show-

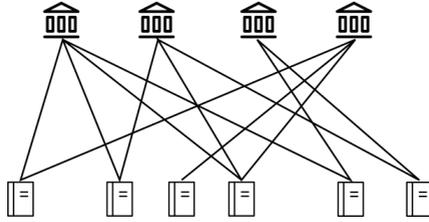


Figure 22: Visualization of a bipartite graph. A subgraph of the *Venezuelan Banks and Assets* bipartite network is shown, with banks on one layer linked to the assets they hold on the other layer.

ing good performances [93]. Here we extend this approach to the bipartite case on social and financial networks. As we will see in the following, the entropy-based approach has a good performance with respect to other available methods, as in the monoptartite case [93].

5.2 Methods

Let us indicate the two layers of the bipartite network as \top and \perp ; nodes on the layer \top are identified by Latin indices and nodes on the layer \perp with Greek ones. The number of nodes of the two layers is respectively N_{\top} and N_{\perp} . A bipartite network is described by a biadjacency matrix, i.e. the rectangular matrix $\mathbf{M}_{N_{\top} \times N_{\perp}}$ whose entries $m_{i\alpha}$ are 1 if there is an edge connecting i and α and 0 otherwise.

Let us indicate with the symbol $L = \sum_{i,\alpha} m_{i\alpha}$ the corresponding set of observed links and with the symbol $U = N_{\top} \times N_{\perp}$ the set of all nodes pair: as a consequence, $U \setminus L$ is the set of non-existent links in the network. In order to study the performance of a link prediction algorithm, the list of edges is usually divided into two separate sets: the training set L^T , used in the “calibration” phase of a given prediction algorithm, and a probe set $L^P = L \setminus L^T$ which is the set of removed links for testing the algorithm, thus constituting the actual “prediction target”. From those definitions, we can indicate with \mathbf{M}^T the portion of the adjacency matrix corresponding to the training set. Finally, the union of the missing-links set and the *non-existent links* set $L^N \equiv U \setminus L^T$ will be referred to as to the set

of *non-observed links*.

The following procedure is followed to test our link prediction method and to compare it with alternative algorithms:

1. the 10% of links are randomly removed. This operation is repeated 10 times;
2. on each of the reduced matrices we apply the link prediction algorithms;
3. the performance of each algorithm is evaluated by means of different evaluation measures which are then averaged across the 10 iterations.

5.2.1 Link prediction methods

Link-prediction algorithms output a list of scores to be assigned to non-observed links. The classification algorithms can be divided in two main classes:

- **Similarity-based algorithms** which employ local, quasi-local or global information, such as, respectively, the nodes degree, the degree of common neighbours and the length of paths connecting any two nodes;
- **Likelihood-based algorithms** defined by a likelihood function whose maximization provides the probability that any two nodes are connected.

The local similarity algorithm are based on the fact that the likelihood of an interaction between two non-adjacent nodes is strongly related with mechanisms of organization involving their first and/or second neighbour nodes. Upon indicating with $N(i)$ and $N(\alpha)$ respectively the set of neighbours of i and α and with $N(N(i))$ and $N(N(\alpha))$ respectively the set of the second-order neighbours of the nodes i and α , the main similarity indexes are the following:

- **Common neighbours (CN):**

$$s_{i\alpha}^{CN} = |(N(i) \cap N(N(\alpha))) \cup (N(\alpha) \cap N(N(i)))| \quad (5.1)$$

is an index counting the neighbors touched by the quadrangles that pass through the nodes i and α . In a bipartite network, of course, nodes in opposite layers cannot have common neighbors. Therefore instead of triangular closing in a monopartite graph, here we have this quadrangular closing: common neighbors defined as a closed quadrangular path on the two layers, see also [42];

- **Resource Allocation (RA):**

$$s_{i\alpha}^{RA} = \sum_{z \in ((N(i) \cap N(N(\alpha))) \cup (N(\alpha) \cap N(N(i))))} \frac{1}{|N(z)|}$$

assigns a different weight to the common neighbours of nodes i and α based on its degree;

- **Preferential Attachment (PA):**

$$s_{i\alpha}^{PA} = k_i \cdot k_\alpha$$

is simply the degree product of nodes i and α , can be used in bipartite networks.

- **Cosine Similarity (CS):**

$$s_{i\alpha}^{CS} = \frac{s_{i\alpha}^{CN}}{\sqrt{|k_i \cdot k_\alpha|}}$$

is based on the Cosine distance between two vectors of same length [82].

In contrast to the existing node-neighborhood-based approaches, the link prediction strategy of other similarity-based models focuses no longer only on groups of common nodes and their node neighbours, but also on the organization of the links between them. In those models, the information content related with the CN nodes is complemented with the

topological information emerging from the interactions between them. In order to demonstrate the validity of this theory on several classes of networks, different classical node-based link prediction techniques like CN, JC, RA and PA were reinterpreted. This mathematical reformulation represents the Cannistraci variations[42] of CN, RA and PA respectively renamed Cannistraci-Alanis-Ravasi (CAR), Cannistraci Resource Allocation (CRA) and Cannistraci Preferential Attachment (CPA) and defined in the following way:

- **CAR index:**

$$s_{i\alpha}^{CAR} = s_{i\alpha}^{CN} \cdot s_{i\alpha}^{LCL}$$

- **CRA index:**

$$s_{i\alpha}^{RA} = \sum_{z \in ((N(i) \cap N(N(\alpha))) \cup (N(\alpha) \cap N(N(i))))} \frac{|\gamma(z)|}{|N(z)|}$$

- **CPA index:**

$$s_{i\alpha}^{CPA} = e_i \cdot e_\alpha + e_i \cdot s_{i\alpha}^{CAR} + e_\alpha \cdot s_{i\alpha}^{CAR} + (s_{i\alpha}^{CAR})^2$$

where $s_{i\alpha}^{LCL}$ counts the links between the common neighbours of nodes i and α , $|\gamma(z)|$ is the number of links of z with the other neighbours of i and α , while $e(i)$ and $e(\alpha)$ are the number of external links respectively of nodes i and α .

5.2.2 The Bipartite Configuration Model approach

In our method, the probabilities of the Bipartite Configuration Model [104] (*BiCM*) are used as score function for predicting links [93].

Following the same maximum entropy framework laid out in the introduction, we derive the Bipartite Configuration Model. For bipartite graphs we denote the biadjacency matrix as \mathbf{M} , to differentiate from the adjacency matrix for monopartite graphs \mathbf{A} . The degree sequences of the two layers are indicates as: \vec{k}^\top for the degree sequence of the nodes in

Table 5: Data description of the MovieLens graph, and the Venezuelan Banks and Assets graph

Graph	Users (Banks)	Items	Nodes	Edges	Avg. De- gree	Avg. De- gree (Users)	Avg. De- gree (Items)
ML	943	1682	2625	100000	76.19	106.04	59.45
VBA ¹	45	20	65	912	28.06	20.27	45.60

the top layer, and \vec{k}^\perp for the bottom layer of the bipartite graph. The ensemble of graphs are in this case all bipartite graphs possible under the constraints. The constraints for the BiCM are the degree sequences of both layers; $\vec{k}^{\top*} = \langle \vec{k}^\top \rangle$ for the top layer, and $\vec{k}^{\perp*} = \langle \vec{k}^\perp \rangle$ for the bottom layer. The Hamiltonian of Eq. 1.3 with these constraints is

$$H(\mathbf{M}, \vec{\alpha}, \vec{\beta}) = \sum_i \alpha_i k_i^\top + \sum_j \beta_j k_j^\perp. \quad (5.2)$$

As such the probability per graph is

$$P(\mathbf{M}, \vec{\alpha}, \vec{\beta}) = \frac{e^{-H(\mathbf{M}, \vec{\alpha}, \vec{\beta})}}{Z(\vec{\alpha}, \vec{\beta})}. \quad (5.3)$$

Upon defining $x_i \equiv e^{-\alpha_i}$ and $y_j \equiv e^{-\beta_j}$, we can write the probability per graph as

$$P(\mathbf{M}, \vec{\alpha}, \vec{\beta}) = \prod_{i,j} \frac{(x_i y_j)^{m_{ij}}}{1 + x_i y_j}. \quad (5.4)$$

We can interpret the term $\frac{x_i y_j}{1 + x_i y_j}$ as the probabilities per link. The system of equation we solve to find the value of the Lagrange multipliers α and β is

$$\begin{cases} k_i^\top = \sum_j \frac{e^{-\alpha_i + \beta_j}}{1 + e^{-\alpha_i + \beta_j}}, \forall i \\ k_j^\perp = \sum_i \frac{e^{-\alpha_i + \beta_j}}{1 + e^{-\alpha_i + \beta_j}}, \forall j. \end{cases} \quad (5.5)$$

5.2.3 Evaluation measures

After the link-prediction algorithm has been performed, a number of statistical indices can be used to test its effectiveness. The first index we

have considered is the *True Positive Rate* (TPR) (also known with the name of *precision*) which is the percentage of missing-links that are correctly recovered, namely the number L_m of correctly identified missing-links, within the list of the first $|E^P|$ links with the largest score. The TPR is defined as:

$$\text{TPR} = \frac{L_m}{|L^P|} \quad (5.6)$$

Another evaluation index is the *area under the ROC curve*, or (AUC). This measures evaluates how many times a method (correctly) assigns a higher score to a missing link with respect to a non-existent one. It is formally defined as:

$$\text{AUC} = \frac{n' + 0.5n''}{n} \quad (5.7)$$

Specifically, for each combination of a missing and non-existent link, if the former scores higher than the latter, the index n' is raised by one unit. If the two links have the same score, n'' is raised. The denominator is given by the product of the number of missing links times the number of non-existent ones. If all scores were i.i.d. the AUC value should be distributed around an expected value of $1/2$: therefore, the extent to which the AUC value exceeds 0.5 provides an indication of how much better the algorithm performs than pure chance. Finally, the last index, called *accuracy* (ACC), quantifies the percentage of correctly classified links, namely both the missing ones and the non-existent ones L_{ne} , with respect to the total number of non-observed links $|L^N|$:

$$\text{ACC} = \frac{L_m + L_{ne}}{|L^N|} \quad (5.8)$$

5.3 Data

The following datasets have been employed to test the link-prediction method:

- **MovieLens (ML):** MovieLens [64] datasets were collected by the GroupLens Research Project at the University of Minnesota. This data set consists of 100000 ratings (1-5) from 943 users on 1682

movies. Each user has rated at least 20 movies and is characterized by some demographic information, such as age, job, sex, state and zipcode. The data was collected through the MovieLens web site (`movielens.umn.edu`) during the seven-month period from September 19th, 1997 through April 22nd, 1998. For the set of movies, there is information on the release year, title and genre. Each user can review a movie with a score that ranges from 1 to 5, according to his level of appreciation. We binarize the network by drawing an edge for a user-movie pair if the user has reviewed the movie;

- **Venezuelan Banks and Assets (VBA):** Bipartite networks of positions that 69 Venezuelan banks hold in 20 asset classes in the period between December 2013 and June 2015. The dataset was firstly presented and analyzed in [80]. The binarized network has an edge between a bank-asset pair, if the position the bank held in the asset class has a value greater than zero at a given timestamp.

The generic statistics of the graphs produced from the datasets that were used in this analysis are provided in Table 5.

5.4 Results

The results of the algorithm performances are presented in Table 6 and Fig. 23 for the MovieLens data set, and the metrics comparison for the Venezuelan Banks and Assets are in the Figures 24a, 24b and 24c. The results of the link prediction in bipartite networks are averaged over 10 iterations for each method, with exception for the BiCM method for the MovieLens data set the average is taken from 7 iterations. Table 6 shows the averaged measure and the standard deviation (SD).

The results for our entropy based algorithm (BiCM) are comparable and show strong performance as opposed to the benchmark algorithms. For the MovieLens network BiCM comes in third place for the accuracy (ACC) measure and is the fourth best algorithm for the precision (TPR) and AUC (see Table 6) and is closely trailing the best performers. For the

Table 6: Comparison of link prediction methods: MovieLens performance comparison, in *italic* the performance of our BiCM method.

Method	ACC (SD)	TPR (SD)	AUC (SD)
BiCM	<i>0.98873</i> (0.00003)	<i>0.15658</i> (0.002)	<i>0.8946</i> (0.002)
cosine	0.98836 (0.00005)	0.12905 (0.004)	0.8903 (0.0007)
car	0.98917 (0.00003)	0.18975 (0.003)	0.9028 (0.001)
CN	0.98860 (0.00003)	0.14713 (0.002)	0.8868 (0.0009)
cpa	0.98917 (0.00003)	0.18975 (0.003)	0.9028 (0.001)
cra	0.98929 (0.00003)	0.19856 (0.002)	0.9163 (0.001)
PA	0.98873 (0.00003)	0.15672 (0.002)	0.8932 (0.001)
RA	0.98793 (0.00004)	0.09712 (0.003)	0.8863 (0.0007)

Venezuelan Banks and Assets networks, BiCM link prediction algorithm is among the best five performers in precision and accuracy measures (see Fig. 24a for multiple time periods) while the three other methods are further behind. Furthermore, inspecting the Fig. 24c it can be observed that BiCM method dominates the others in AUC measure. For all algorithms in the Venezuelan Banks and Assets network analysis, precision ranges between 0.3875 and 0.7818, accuracy is between 0.8877 and 0.9577, and AUC values change from 0.8958 to 0.9730. All algorithms perform better on the Venezuelan Banks and Assets than on the MovieLens data set with respect to precision measure, i.e. the rate of true positive values.

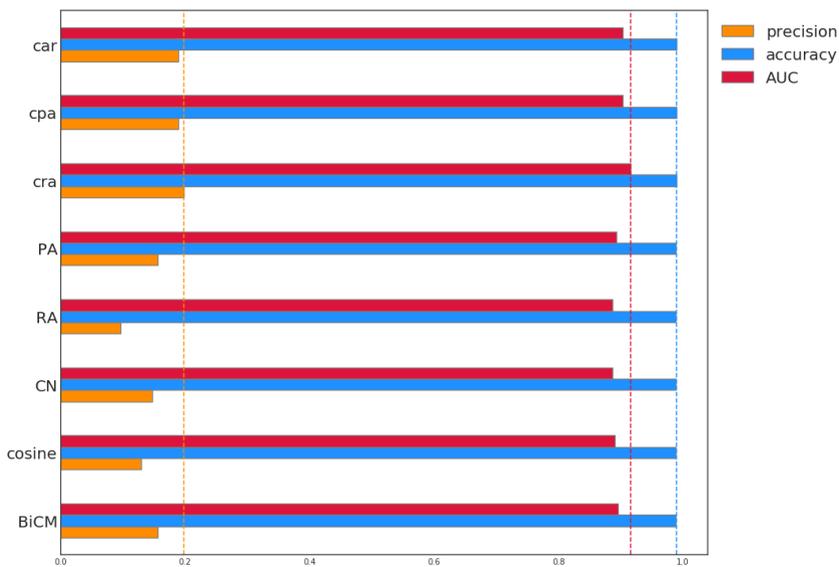


Figure 23: Link prediction performance on the MovieLens dataset, a bi-partite graph of users linked to movies they reviewed. The BiCM method performs on par with the alternative methods on all evaluation measures.

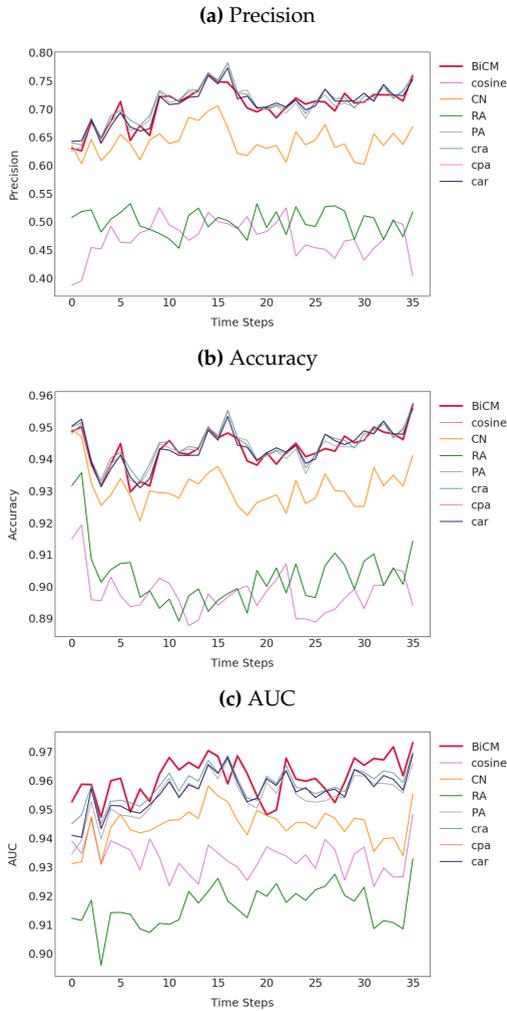


Figure 24: Link prediction performance on the Venezuelan Banks and Assets dataset, a bipartite graph which describes the types of assets in which banks held a position at various time-snapshots between December 2013 and June 2015. Performance is measured by Accuracy, Precision and AUC. The BiCM method is among the best performing methods in all time snapshots, and has best performance in some of these.

5.5 Discussion

Link-prediction is a method that can be leveraged for a wide array of tasks, as compensating for missing information [81]. Recently, it was proposed to employ entropy based null-model probabilities [94, 59, 109] as score function for predicting links [93]: missing links with high probability are likely to be present. In this chapter we extend this approach to bipartite networks, thus using as score function the probability of the bipartite configuration model [104].

In order to test our predictions, we first randomly remove a fraction of the links present in the real network and then use our procedure to predict the same amount of links. On the real world bipartite networks of user-movie ratings and bank-asset positions, we compared the performances of our proposed method to seven alternative local information based methods. On all datasets and all evaluation measures our method is able to consistently predict missing links.

It is not surprising that our approach has better performances on financial data, than on the social network of Movielens [64]: indeed in the latter case it is known that a collaborative filtering recommendation system [100] was employed [64]. Nevertheless, our results have similar performances with other known methods. Moreover, it is remarkable that our approach, that is based on local constraints has performances of the same order of quasi-local methods as the bipartite extension of Cannistraci corrected scores[32, 42]. For financial networks, for which the bipartite configuration model is known for having good performances, results are more promising: the BiCM induced link prediction is among the top link prediction methods on the tested datasets.

Our method can be naturally extended to (bipartite) review networks, as the ones in which a users can give a rating to a certain item. While the prediction of both the existence of links and their strength is not trivial, the recent extension to bipartite score network of the configuration model[21] makes the task more promising, thus overcoming the limitations of competing algorithms [32, 42].

Chapter 6

Conclusion

Networks arise naturally in economical and financial systems as these evolve around exchanges and transactions. In the past years, new data availability has driven new ways of analysing these systems. With data, for example on bank-to-bank loans, the analysis of these systems has moved away from assumption-rich macroscopic descriptions. In these interacting systems there has been a shift to model interactions between the actors in the system with the tools of *complex networks* theory. Complex networks represent heterogeneous systems characterized by emerging properties driven by interactions rather than individual characteristics: network science provides methods and tools to extract significant information from these systems.

In this thesis, we have presented novel methods for the analysis of mesoscale network structures and applied such methods to study real-world economic and financial networks. While a lot of attention has been focused on the community structure, many real world networks are characterized by mesoscale structures like core-periphery and bipartitions, especially so for economic and financial networks: whereas communities are characterized by a higher internal connectivity, the former mesoscale structures actually do have a significant connectivity *between* clusters. Detecting the presence of mesoscale structures in complex networks is of primary importance [53, 73]. This is especially true for fi-

nancial networks, whose structural organization deeply affects their resilience to shock propagation, node failures, etc. [40, 70, 54, 83].

When considering mesoscale structures, it is generally believed that null models encoding the modular organisation of nodes must be employed. We have explored the effectiveness of null models that constrain only local information in explaining mesoscale structures as the bow-tie and core-periphery structures. This was done by comparing competing null models and analysing their relative performance using model selection criteria. Specifically, we have considered the directed random graph, constraining the network edge density, the directed configuration model, constraining the directed degree sequence, and the reciprocated configuration model, which puts constraints on the reciprocated network degree sequence, beside their block-wise version, i.e. models encoding node-membership information. The World Trade Web, the network of worldwide bilateral import and exports, and eMID, the Italian interbank exposure network, provide the real world systems with a meso-scale structure to check model performances. Using AIC and BIC model selection criteria we found that it is often enough to constrain the network degree sequence, for reproducing the mesoscale structures.

The null models encoding the modular structure partition the network into asymmetric blocks, characterised by binary directed connections. In the process of defining these null models, we therefore also enriched the toolbox for analysis of bipartite networks by extending a recently proposed method to randomise undirected, bipartite, networks to the directed case; defining the directed bipartite configuration model [103, 113].

Our findings on the reconstruction of mesoscale structures indicate that these structures are often explained by the information encoded into the degree sequence, i.e. are compatible with the configuration model prediction. Another recent work also points out that a core-periphery structure cannot be significant under the configuration model [76]. To

still be able to detect such a core-periphery structure as a significant partition, we adopt a modification of the surprise, recently proposed for detecting communities [3, 4, 90, 118], thus turning to a “tuned down” null model, i.e. the (directed) random graph. Our variant allows for bimodular nodes partitions to be revealed, by letting links to be placed either 1) within the core part and between the core and the periphery parts or 2) just between the (empty) layers of a bipartite network. We evaluate such a partition of nodes with a multinomial hypergeometric distribution, an adaptation from the binomial hypergeometric one used in surprise, to allow for three types of links. This distribution allows a p-value to be assigned to any given (bi)partition of the nodes: we find core-periphery and bipartite structure by identifying the partition of the nodes that minimises this p-value, i.e. finding the most significant bimodular structure. To illustrate the performance of our method, we reported the results of its application to several real-world networks, including social, economic and financial ones.

The search for the partition which minimises bimodular surprise is increasingly computationally heavy for larger networks. We therefore introduced a search heuristic. It is based on a modification of the PACO heuristic introduced for surprise [90], and is driven by a greedy process which cycles through all sorted edges.

Other methods exist to detect core-periphery structures, mainly the ones derived from the Borgatti model [24, 70, 54, 25], which defines a distance function of an observed network to an idealised core-periphery structure. These methods are in general very reliant on node degree [76], with the core often composed of just the nodes with the largest number of neighbours. Bimodular surprise is driven by the interplay of link density values of the different network areas. By employing a benchmark, our method also has the advantage of making the statistical significance of a given structure explicit. Fitting methods ‘a la Borgatti’ leave this implicit, which pushes one to enrich the model with an increasing amount of information whose relevance cannot be easily clarified.

We have, then, moved to study corporate networks. The latter can be considered to be a mixture of economic and financial systems. Stock-

correlation networks describe a more financial system, where ownership between firms describe a more economical system. Several types of corporate networks have been studied before, such as ownership networks, research collaboration networks, and social networks of the boardroom elite [101, 132, 16, 43, 116]. These analyses have shaped our understanding of many important economic systems, like the concentration of ownership into a small group of corporation, and the spread of decision making in businesses through the social network of company directors [125, 16]. These studies have characterized many corporate networks as core periphery and bow-tie mesoscale structures. However, these networks have been studied almost exclusively in isolation. They are, in fact, all part of the complex system of firm interactions. We therefore studied these interactions in parallel and evaluated the structure of the disaggregated network. To this end, we constructed a new multilayer network dataset on interactions between companies in Germany and in the United Kingdom, combining ownership links, social ties through joint board directors, R&D collaborations and stock correlations in one linked multiplex dataset. We have described the features of this network with several network measures. With three separate methods we have shown the significance of layers within the aggregate multiplex structure; the node- and edge-overlap, the structural reducibility and a multiplex network regression. All these three methods confirm that the multiplex representation has an added value over the single layers. These results show that corporate control, boardroom influence and other connections have different but connected structures. This multilayer network shows that the corporate world is an even smaller world than the small world already described by various previous studies on corporate control and the 'old boys network' of board rooms.

As a side application of null models, we have approached the problem of link prediction in bipartite networks. We again employed the maximum entropy framework, building on a recent paper where this was used for monopartite networks [93]. The bipartite configuration model, in fact, defines all possible edge probabilities of the "null graph". The unobserved edges with highest probability are used for the link im-

putation problem. The MovieLens dataset, describing movie preferences of individuals, and the bipartite network of Venezuelan banks linked to the assets they hold have been used to test our method. We remove a random given fraction of the links in these datasets and use our link prediction method to predict these removed links. When compared to alternative state of the art link prediction methods, our bipartite configuration model method has almost always on par or better results than existing methods. An advantage of our method is that it only uses local information, e.g. degree sequences, where alternative measures use quasi-local information requiring the complete network configuration.

Throughout this thesis we have leveraged existing and newly proposed methods for the analysis of the (mesoscale) structure of economic, financial and corporate networks. We find that many of these systems show a modular organisation characterised by *core-periphery*, *bow-tie*, and *bipartite* structures. Our methods revealed the evolution of specific modules of the mesoscale structure in networks of global trade and exposures between banks. The overall mesoscale structure of these systems is however characterised as a stable core-periphery structure. We showed that the way in which network data is aggregated matters for the classification of the network structure. The mesoscale structure of Italian interbank loans is sometimes revealed as a bipartite structure when describing short time periods, and described as a core-periphery structure for longer time aggregation. For corporate networks, as we unravel the aggregate network and evaluate the different layers of this multilayer network, we found that the structure of the layers and the aggregate have vastly different characteristics.

Appendix A

Null Models

The null models used for reconstruction and link prediction are specific versions of the Exponential Random Graph Model as introduced in Chapter 1. In this appendix we derive the models used in Chapters 2 and 5 from this general model. In Chapter 2 we use models containing both node-specific local information and group membership of the nodes. We will start by deriving the model for the general case constraining the degree sequence and specifying the block-structure, and then continue to the other models by setting specific constraints.

We denote the adjacency matrix of a graph with \mathbf{A} , whose the entry a_{ij} signals the presence of a link between node i and node j . The degree of a node is k_i .

Degree informed

This first class of null models are the so-called *degree-informed null models*. Here specifically we will include information on the block structure in the models. All null models in this class are defined by constraints encoding node-specific local information (i.e. the directed degree sequences), and the membership of nodes to specified groups (labelled by the symbols $\{g_i\}$). Building on the Exponential Random Graph formalism introduced in Section 1.3, we combine these two kinds of information, obtaining block-specific directed degree sequences, definable

$$k_i^{r \rightarrow s} = \delta_{g_i r} \sum_{j(\neq i)} \delta_{g_j s} a_{ij}, \quad \forall i, r, s \quad (\text{A.1})$$

$$h_i^{s \rightarrow r} = \delta_{g_i r} \sum_{j(\neq i)} \delta_{g_j s} a_{ji}, \quad \forall i, r, s \quad (\text{A.2})$$

with $k_i^{r \rightarrow s}$ indicating the contribution to the out-degree of node i (belonging to block r) coming from block s (and analogously for $h_i^{s \rightarrow r}$). Remarkably, all null models in this class induce a probability for the generic network configuration \mathbf{A} reading

$$P(\mathbf{A}) = \prod_{i \neq j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1 - a_{ij}} \quad (\text{A.3})$$

with different null models inducing different functional forms for the edge-probability coefficients $\{p_{ij}\}$.

Reciprocity informed

When analysing directed networks, however, a non-trivial piece of information to be taken into account is represented by reciprocity [60]. For this reason, a second class of null models, the one including the so-called *reciprocity-informed null models*, is considered as well. Null models in this class are defined by constraints encoding the (non) reciprocal degree sequences, beside the usual nodes membership. In the most general case, the constraints defining such models can be written as

$$k_i^{\overrightarrow{rs}} = \delta_{g_i r} \sum_{j(\neq i)} \delta_{g_j s} a_{ij}^{\overrightarrow{rs}}, \quad \forall i, r, s \quad (\text{A.4})$$

$$k_i^{\overleftarrow{rs}} = \delta_{g_i r} \sum_{j(\neq i)} \delta_{g_j s} a_{ij}^{\overleftarrow{rs}}, \quad \forall i, r, s \quad (\text{A.5})$$

$$k_i^{\overleftrightarrow{rs}} = \delta_{g_i r} \sum_{j(\neq i)} \delta_{g_j s} a_{ij}^{\overleftrightarrow{rs}}, \quad \forall i, r, s. \quad (\text{A.6})$$

with $a_{ij}^{\rightarrow} = a_{ij}(1 - a_{ji})$, $a_{ij}^{\leftarrow} = a_{ji}(1 - a_{ij})$ and $a_{ij}^{\leftrightarrow} = a_{ij}a_{ji}$ [60] and $k_i^{\leftarrow r s}$ indicating the contribution to the reciprocal degree of node i (belonging to block r) coming from block s . All models in this second class induce a probability for the network \mathbf{A} reading

$$P(\mathbf{A}) = \prod_{i < j} (p_{ij}^{\rightarrow})^{a_{ij}^{\rightarrow}} (p_{ij}^{\leftarrow})^{a_{ij}^{\leftarrow}} (p_{ij}^{\leftrightarrow})^{a_{ij}^{\leftrightarrow}} (p_{ij}^{\leftarrow r})^{a_{ij}^{\leftarrow r}} (p_{ij}^{\leftarrow s})^{a_{ij}^{\leftarrow s}}; \quad (\text{A.7})$$

as before, different null models induce different functional forms for the probability coefficients $\{p_{ij}^{\rightarrow}\}$, $\{p_{ij}^{\leftarrow}\}$, $\{p_{ij}^{\leftrightarrow}\}$, $\{p_{ij}^{\leftarrow r}\}$.

For both classes of models, the likelihood function associated with $P(\mathbf{A})$ reads $\mathcal{L}(\mathbf{A}) = \ln P(\mathbf{A}) = \ln [\prod_b P(\mathbf{A}^{(b)})] = \sum_b \ln P(\mathbf{A}^{(b)})$. The second passage follows from the observation that each null model we consider in this chapter treats different nodes pairs as independent, thus inducing a factorized form of the probability coefficient $P(\mathbf{A})$ over the aforementioned blocks.

Block Configuration Model (BCM)

All null models in the first class are particular cases of the following hamiltonian

$$\begin{aligned} H &= \sum_{i \neq j} (\alpha_i^{g_i \rightarrow g_j} + \beta_j^{g_i \rightarrow g_j}) a_{ij} = \\ &= \sum_{i \neq j} \sum_r \sum_s \delta_{g_i r} \delta_{g_j s} (\alpha_i^{r \rightarrow s} + \beta_j^{r \rightarrow s}) a_{ij} \end{aligned} \quad (\text{A.8})$$

an expression inducing the following probability coefficients

$$p_{ij} = \frac{x_i^{g_i \rightarrow g_j} y_j^{g_i \rightarrow g_j}}{1 + x_i^{g_i \rightarrow g_j} y_j^{g_i \rightarrow g_j}} \quad (\text{A.9})$$

(where $x_i = e^{-\alpha_i^{g_i \rightarrow g_j}}$ and $y_i = e^{-\beta_i^{g_i \rightarrow g_j}}$) to be numerically determined by solving the likelihood equations

$$\begin{cases} k_i^{r \rightarrow s} = \langle k_i^{r \rightarrow s} \rangle, \forall i, r, s \\ h_i^{s \rightarrow r} = \langle h_i^{s \rightarrow r} \rangle, \forall i, r, s \end{cases} \quad (\text{A.10})$$

with $\langle k_i^{r \rightarrow s} \rangle = \delta_{g_i r} \sum_{j(\neq i)} \delta_{g_j s} p_{ij}$ and $\langle h_i^{s \rightarrow r} \rangle = \delta_{g_i r} \sum_{j(\neq i)} \delta_{g_j s} p_{ji}$.

Notice that the directed version of the Stochastic Block Model (SBM), the block model of the Directed Random Graph, can be recovered as a special case of the BCM, by posing $\alpha_i^{g_i \rightarrow g_j} = \alpha^{g_i \rightarrow g_j}$ and $\beta_j^{g_i \rightarrow g_j} = \beta^{g_i \rightarrow g_j}$ in Eq. A.8 and solving the equations $L_{rs} = \langle L_{rs} \rangle, \forall r, s$ with $L_{rs} = \sum_{i \neq j} \delta_{g_i r} \delta_{g_j s} a_{ij}$ and $\langle L_{rs} \rangle = \sum_{i \neq j} \delta_{g_i r} \delta_{g_j s} p_{ij}$.

The BCM extends the results in [72, 56] to the directed case. Interestingly, upon identifying $\alpha_i^{g_i \rightarrow g_j} \equiv \alpha_i + \frac{w_{g_i g_j}}{2}$ and $\beta_j^{g_i \rightarrow g_j} \equiv \beta_j + \frac{w_{g_i g_j}}{2}$ the *directed degree-corrected SBM* (ddc-SBM) of [72] is recovered. Upon retaining all multipliers in Eq. A.8 and defining $x_i \equiv e^{-\alpha_i}, y_i \equiv e^{-\beta_i}$ and $\chi_{g_i g_j} \equiv e^{-w_{g_i g_j}}$, one finds that

$$p_{ij} = \frac{x_i y_j \chi_{g_i g_j}}{1 + x_i y_j \chi_{g_i g_j}}; \quad (\text{A.11})$$

although formally equivalent, the expressions A.11 and A.9 are not when coming to estimate the unknown parameters: Eq. A.11 is, in fact, determined by solving the equations

$$\begin{cases} k_i = \langle k_i \rangle, \forall i \\ h_i = \langle h_i \rangle, \forall i \\ L_{rs} = \langle L_{rs} \rangle, \forall r, s \end{cases} \quad (\text{A.12})$$

The BCM in fact extends the results in [72, 134] to the non-sparse case.

Directed Configuration Model (DCM)

The local information model not specifying the block structure can be derived from the same A.8 equations by not encoding any block membership information. The DCM is obtained by posing $\alpha_i^{g_i \rightarrow g_j} = \alpha_i$ and $\beta_j^{g_i \rightarrow g_j} = \beta_j$ in Eq. A.8. Upon defining $x_i \equiv e^{-\alpha_i}$ and $y_i \equiv e^{-\beta_i}$, the surviving multipliers induce probability coefficients reading

$$p_{ij} = \frac{x_i y_j}{1 + x_i y_j} \quad (\text{A.13})$$

to be numerically determined by solving the likelihood equations

$$\begin{cases} k_i = \langle k_i \rangle, \forall i \\ h_i = \langle h_i \rangle, \forall i \end{cases} \quad (\text{A.14})$$

with the out- and in-degrees reading $k_i = \sum_{j(\neq i)} a_{ij}$ and $h_i = \sum_{j(\neq i)} a_{ji}$ respectively and $\langle k_i \rangle = \sum_{j(\neq i)} p_{ij}$, $\langle h_i \rangle = \sum_{j(\neq i)} p_{ji}$.

The Directed Random Graph Model (DRG) can be recovered as a particular case of the DCM, obtained by posing $\alpha_i \equiv \alpha$ and $\beta_j \equiv \beta$ in Eq. A.8. The only coefficient $p_{ij} \equiv p$ is determined by solving the equation $L = \langle L \rangle$ with $L = \sum_{i \neq j} a_{ij}$ and $\langle L \rangle = \sum_{i \neq j} p$.

Reciprocal Configuration Model (RCM)

We will now take into account the reciprocity of the network connections, by taking information from the reciprocated (or non-reciprocated) degrees. The RCM is defined by the following probability coefficients

$$p_{ij}^{\rightarrow} = \frac{x_i y_j}{1 + x_i y_j + y_j x_i + z_i z_j}, \quad (\text{A.15})$$

$$p_{ij}^{\leftarrow} = \frac{x_j y_i}{1 + x_i y_j + y_i x_j + z_i z_j}, \quad (\text{A.16})$$

$$p_{ij}^{\leftrightarrow} = \frac{z_i z_j}{1 + x_i y_j + y_i x_j + z_i z_j} \quad (\text{A.17})$$

to be numerically determined by solving the likelihood equations

$$\begin{cases} k_i^{\rightarrow} = \langle k_i^{\rightarrow} \rangle, \forall i \\ k_i^{\leftarrow} = \langle k_i^{\leftarrow} \rangle, \forall i \\ k_i^{\leftrightarrow} = \langle k_i^{\leftrightarrow} \rangle, \forall i \end{cases} \quad (\text{A.18})$$

with $\langle k_i^{\rightarrow} \rangle = \sum_{j(\neq i)} p_{ij}^{\rightarrow}$, $\langle k_i^{\leftarrow} \rangle = \sum_{j(\neq i)} p_{ij}^{\leftarrow}$, $\langle k_i^{\leftrightarrow} \rangle = \sum_{j(\neq i)} p_{ij}^{\leftrightarrow}$.

Block Reciprocal Configuration Model (BRCM)

The RCM can be re-defined in same block-wise fashion as the DCM, by specifying the probability coefficients defined by Eqs. A.15, A.16, A.17

for each block. A Block Reciprocal Configuration Model (BRCM) remains naturally defined, being determined by the system of equations

$$k_i^{\overrightarrow{rs}} = \langle k_i^{\overrightarrow{rs}} \rangle, \forall i, r, s \quad (\text{A.19})$$

$$k_i^{\overleftarrow{rs}} = \langle k_i^{\overleftarrow{rs}} \rangle, \forall i, r, s \quad (\text{A.20})$$

$$k_i^{\overleftrightarrow{rs}} = \langle k_i^{\overleftrightarrow{rs}} \rangle, \forall i, r, s \quad (\text{A.21})$$

with obvious meaning of the symbols.

Off-diagonal matrices

As the group membership might induce a partition of the adjacency matrix with asymmetric off-diagonal matrices, let us explicitly solve the BCM in the two, off-diagonal matrices \mathbf{A}^\top and \mathbf{A}^\perp (as in Eq. 2.1). In order to fix the formalism, let us suppose the two off-diagonal blocks \mathbf{A}^\top and \mathbf{A}^\perp to have dimensions $C \times P$ and $P \times C$, respectively. Analogously to the undirected case [104], solving the DCM within the off-diagonal blocks of the matrix \mathbf{A} induces the following probability coefficients

$$P(\mathbf{A}^\top) = \prod_c \prod_p p_{cp}^{a_{cp}^\top} (1 - p_{cp})^{1 - a_{cp}^\top} \quad (\text{A.22})$$

and

$$P(\mathbf{A}^\perp) = \prod_p \prod_c q_{pc}^{a_{pc}^\perp} (1 - q_{pc})^{1 - a_{pc}^\perp}; \quad (\text{A.23})$$

the probability that a link from a core node c to a periphery node p exists is $p_{cp} \equiv \frac{x_c^\top y_p^\top}{1 + x_c^\top y_p^\top}$ and the probability that a link from a periphery node p to a core node c exists is $q_{pc} \equiv \frac{x_p^\perp y_c^\perp}{1 + x_p^\perp y_c^\perp}$. Consistently, the vector $\vec{x} = \{\vec{x}_c^\top, \vec{x}_p^\perp\}$ is coupled to the outgoing degrees, while the vector $\vec{y} = \{\vec{y}_c^\perp, \vec{y}_p^\top\}$ is coupled to the incoming degrees.

The aforementioned probability coefficients are determined via the likelihood condition in A.10. Let us notice that the out-degree of core nodes and the in-degree of periphery nodes are measured on the matrix

\mathbf{A}^\top ; the converse is true for the matrix \mathbf{A}^\perp . More quantitatively, upon indicating with $\{\vec{k}, \vec{h}\}$ the core and periphery nodes degrees, one has

$$k_c^{out} = \sum_p a_{cp}^\top, k_c^{in} = \sum_p a_{pc}^\perp, \forall c \quad (\text{A.24})$$

and

$$h_p^{out} = \sum_c a_{pc}^\perp, h_p^{in} = \sum_c a_{cp}^\top, \forall p. \quad (\text{A.25})$$

The estimation step, thus, reads

$$\begin{cases} k_c^{out} &= \sum_p p_{cp}, \forall c, \\ h_p^{out} &= \sum_c q_{pc}, \forall p, \\ k_c^{in} &= \sum_p q_{pc}, \forall c, \\ h_p^{in} &= \sum_c p_{cp}, \forall p. \end{cases} \quad (\text{A.26})$$

The SBM can be recovered by posing $p_{cp} \equiv p$ and $q_{cp} \equiv q$, to be estimated by solving

$$p = \frac{L^\top}{C \cdot P} = \frac{\sum_{c,p} a_{cp}^\top}{C \cdot P} \text{ and } q = \frac{L^\perp}{C \cdot P} = \frac{\sum_{c,p} a_{cp}^\perp}{C \cdot P} \quad (\text{A.27})$$

with obvious meaning of the symbols.

Inserting the information about reciprocity into a bipartite null model leads to the following probability coefficient

$$P(\mathbf{B}) = \prod_c \prod_p (p_{cp}^{\rightarrow})^{a_{cp}^{\rightarrow}} (p_{cp}^{\leftarrow})^{a_{cp}^{\leftarrow}} (p_{cp}^{\leftrightarrow})^{a_{cp}^{\leftrightarrow}} (p_{cp}^{\leftrightarrow})^{a_{cp}^{\leftrightarrow}} \quad (\text{A.28})$$

that ‘‘mixes’’ the information coming from the two biadjacency matrices \mathbf{A}^\top and \mathbf{A}^\perp (whence the choice of a different symbol, \mathbf{B} , to indicate the bipartite network as a whole). The new variables read $a_{cp}^{\rightarrow} = a_{cp}^\top (1 - a_{pc}^\perp)$, $a_{cp}^{\leftarrow} = a_{pc}^\perp (1 - a_{cp}^\top)$, $a_{cp}^{\leftrightarrow} = a_{cp}^\top a_{pc}^\perp$ and $a_{cp}^{\leftrightarrow} = (1 - a_{cp}^\top)(1 - a_{pc}^\perp)$: while a_{cp}^{\rightarrow} indicates that a non-reciprocated link is present from the core node c to the periphery node p , a_{cp}^{\leftarrow} indicates that a non-reciprocated link is present from the periphery node p to the core node c ; naturally, a_{cp}^{\leftrightarrow} indicates that

both links are present between nodes c and p and a_{cp}^{\leftrightarrow} indicates that no link is present between the same nodes.

The probability coefficients defining our bipartite, reciprocal model read

$$p_{cp}^{\rightarrow} = \frac{x_c r_p}{1 + x_c r_p + y_c s_p + z_c t_p}, \quad (\text{A.29})$$

$$p_{cp}^{\leftarrow} = \frac{y_c s_p}{1 + x_c r_p + y_c s_p + z_c t_p}, \quad (\text{A.30})$$

$$p_{cp}^{\leftrightarrow} = \frac{z_c t_p}{1 + x_c r_p + y_c s_p + z_c t_p}, \quad (\text{A.31})$$

$$p_{cp}^{\leftrightarrow\leftrightarrow} = \frac{1}{1 + x_c r_p + y_c s_p + z_c t_p}, \quad (\text{A.32})$$

whose numerical value is determined by the following sufficient statistics, i.e. the reciprocal and non-reciprocal degrees of both core nodes

$$k_c^{\rightarrow} = \sum_p a_{cp}^{\rightarrow}, \quad k_c^{\leftarrow} = \sum_p a_{cp}^{\leftarrow}, \quad k_c^{\leftrightarrow} = \sum_p a_{cp}^{\leftrightarrow} \quad (\text{A.33})$$

(with $c = 1 \dots C$) and periphery nodes

$$h_p^{\rightarrow} = \sum_c a_{cp}^{\leftarrow}, \quad h_p^{\leftarrow} = \sum_c a_{cp}^{\rightarrow}, \quad h_p^{\leftrightarrow} = \sum_c a_{cp}^{\leftrightarrow} \quad (\text{A.34})$$

(with $p = 1 \dots P$). Notice that the binary variables defining h_p^{\leftarrow} (h_p^{\rightarrow}) are the ones defining also k_c^{\rightarrow} (k_c^{\leftarrow}): in fact, the non-reciprocated links outgoing from the core (periphery) are the same links incoming into the periphery (core). Finally, the estimation step for such a model reads

$$\left\{ \begin{array}{l} k_c^{\rightarrow} = \sum_p p_{cp}^{\rightarrow}, \quad \forall c, \\ h_p^{\rightarrow} = \sum_c p_{cp}^{\leftarrow}, \quad \forall p, \\ k_c^{\leftarrow} = \sum_p p_{cp}^{\leftarrow}, \quad \forall c, \\ h_p^{\leftarrow} = \sum_c p_{cp}^{\rightarrow}, \quad \forall p, \\ k_c^{\leftrightarrow} = \sum_p p_{cp}^{\leftrightarrow}, \quad \forall c, \\ h_p^{\leftrightarrow} = \sum_c p_{cp}^{\leftrightarrow}, \quad \forall p. \end{array} \right. \quad (\text{A.35})$$

Appendix B

Bimodular Surprise

B.1 Numerical approximations

The computation of binomial coefficients for large graphs can quickly become numerically demanding. In order to simplify the calculations of our bimodular surprise, let us proceed by steps. First, let us Stirling approximating the binomial coefficients:

$$\binom{V_c}{i} \simeq \left[(p_c)^i (1 - p_c)^{V_c - i} \right]^{-1}, \quad (\text{B.1})$$

$$\binom{V_{cp}}{j} \simeq \left[(p_{cp})^j (1 - p_{cp})^{V_{cp} - j} \right]^{-1}, \quad (\text{B.2})$$

$$\binom{V_p}{L - (i + j)} \simeq \left[(p_p)^{L - (i + j)} (1 - p_p)^{V_p - (L - (i + j))} \right]^{-1}, \quad (\text{B.3})$$

$$\binom{V}{L} \simeq \left[p^L (1 - p)^{V - L} \right]^{-1} \quad (\text{B.4})$$

having defined $V - (V_c + V_{cp}) \equiv V_p$, $p \equiv \frac{L}{V}$, $p_c \equiv \frac{i}{V_c}$, $p_{cp} \equiv \frac{j}{V_{cp}}$, $p_p \equiv \frac{L - (i + j)}{V_p}$. As a second step, let us substitute the expressions above into Eq. 3.4:

$$S_{\parallel} \simeq \sum_{i \geq l_c^*} \sum_{j \geq l_{cp}^*} \left(\frac{p}{p_p}\right)^L \left(\frac{1-p}{1-p_p}\right)^{V-L} \left(\frac{p_p}{p_c}\right)^i \left(\frac{1-p_p}{1-p_c}\right)^{V_c-i} \cdot \left(\frac{p_p}{p_{cp}}\right)^j \left(\frac{1-p_p}{1-p_{cp}}\right)^{V_{cp}-j}; \quad (\text{B.5})$$

in order to obtain a more explicit expression, let us limit ourselves to consider the leading term of the summation in Eq. B.5 that is readily obtained upon substituting i with l_c^* and j with l_{cp}^* :

$$S_{\parallel} \simeq \left(\frac{p}{p_p}\right)^L \left(\frac{1-p}{1-p_p}\right)^{V-L} \left(\frac{p_p}{p_c}\right)^{l_c^*} \left(\frac{1-p_p}{1-p_c}\right)^{V_c-l_c^*} \cdot \left(\frac{p_p}{p_{cp}}\right)^{l_{cp}^*} \left(\frac{1-p_p}{1-p_{cp}}\right)^{V_{cp}-l_{cp}^*} \quad (\text{B.6})$$

where, now, $p_c \equiv \frac{l_c^*}{V_c}$, $p_{cp} \equiv \frac{l_{cp}^*}{V_{cp}}$, $p_p \equiv \frac{L-(l_c^*+l_{cp}^*)}{V_p}$. Eq. B.6 already makes intuitively clear that our bimodular surprise is likely to be significant either when $p_c \simeq p_p$ but $p_{cp} \gg p_c \simeq p_p$ (i.e. in the case of bipartite networks - notice that Eq. 3.8 is recovered when $L = l_{cp}^*$) or when $p_c \gg p_p$ and $p_{cp} \gg p_p$ (i.e. in the core-periphery case - notice that Eq. 3.9 is recovered when $L = l_c^* + l_{cp}^*$).

A numerical check of the validity of the proposed approximation is shown in Fig. 25, where it is computed for k-star networks and compared to the full expression in Eq. 3.4.

Numerical surprise optimization: a modified PACO algorithm

In what follows we show a modified version of the PACO (PARTitioning Cost Optimization) algorithm pseudocode [90] by running which the partition that minimizes surprise can be found. The PACO algorithm implements an approach that is heuristic in nature since an exhaustive search of all possible partitions is not feasible when dealing with large graphs.

The idea is that of assigning every node to either one of two subsets - interpretable as the core and the periphery or the layers of a bipartite

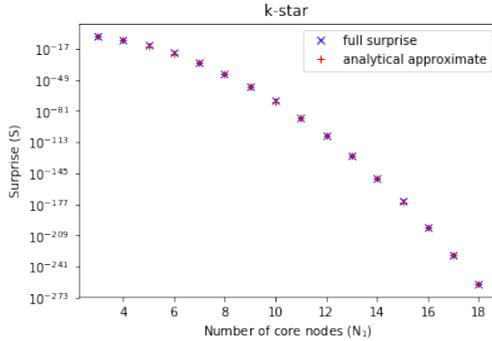


Figure 25: Numerical check of the validity of the approximation shown in Eq. B.6, computed for k-star networks and compared to the full expression in Eq. 3.4.

graph - by running a greedy process that takes as input pairs of nodes connected by an edge and evaluates whether those two nodes should belong to the same subset or not: the choice that minimizes the surprise is the one that is actually implemented. In order to speed up the calculations, the original PACO algorithm takes edges that are previously sorted according to their decreasing value of Jaccard index. Since the latter quantifies the fraction of common neighbours of the two connected nodes, nodes pairs with larger Jaccard index are also the ones most likely to be assigned to the same subset (e.g. a community).

Since for pure bipartite graphs the Jaccard index - as defined above - is zero for all edges (nodes connected by an edge always lie on different layers) we need to modify the score according to which we sort the edges. In our modified version of the PACO algorithm we sort links according to the number of z-motifs they belong to, the latter being defined as $z_{i\alpha} = \sum_{\beta,j} a_{i\beta} a_{i\alpha} a_{\alpha j}$: in other words, we evaluate the number of times a generic link is the “middle” one of a path whose length is 3. As with the original PACO algorithm, we progressively consider all edges, sorted as described above, evaluating whether the linked pairs should belong to the same subset or not.

As a final step, we consider a number of random reassignments of

nodes with the aim of preventing the possibility of getting stuck in a local minimum (a random move consists of selecting 3 random nodes belonging to the same group and evaluating if assigning them to different subsets would further minimize surprise).

The algorithm described above performs quite well in finding the global minimum of surprise on a range of different configurations we have tested. When considering low-density bipartite graphs, however, the algorithm does not always succeed in reaching the global minimum.

```

1: function CALCULATEANDUPDATESURPRISE( $C, C'$ )
2:    $S \leftarrow \text{calculateSurprise}(C)$ 
3:    $S' \leftarrow \text{calculateSurprise}(C')$ 
4:   if  $S' < S$  then
5:      $C \leftarrow C'$ 
6:      $S \leftarrow S'$ 
7:   end if
8:   return  $C$ 
9: end function
10:
11:  $C \leftarrow$  array of length  $N$  randomly initialized with binary entries (0 or
12: 1);
13:  $E \leftarrow$  sorted edges in decreasing order;
14: for edge  $(u, v) \in E$  do
15:    $C' \leftarrow C$ 
16:   if  $C'[u] \neq C'[v]$  then
17:      $C'[u] \leftarrow C'[v]$ 
18:      $C \leftarrow$  CALCULATEANDUPDATESURPRISE( $C, C'$ )
19:   else
20:      $C'[u] \leftarrow 1 - C'[v]$ 
21:      $C \leftarrow$  CALCULATEANDUPDATESURPRISE( $C, C'$ )
22:   end if
23:    $\Rightarrow$  randomly switch node membership for  $n = 3$  nodes in the
24:   same partition and accept move if  $S_{\parallel}$  decreases;
25: end for
26:  $\Rightarrow$  repeat several times the for-loop to improve the chance of finding
27: the optimal partition.

```

Table 7: Pseudocode for the PACO optimization heuristic for the minimalisation of bimodular surprise.

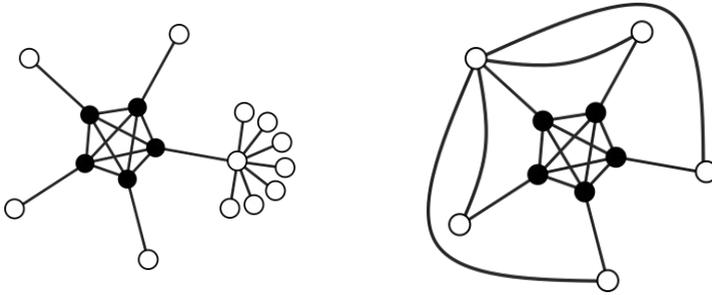


Figure 26: Two example networks with a connected clique, and a node with a higher degree than the nodes in the clique. Despite the higher degree these nodes are not classified as core-nodes by bimodular surprise. This illustrates that bimodular surprise does not solely rely on the degree of nodes, but also about the connections with other nodes in the same partition.

B.2 Core classification - node degree

We have pointed out that several alternative methods for core-periphery detection are highly correlated with node degree; the core nodes are basically the nodes with highest degree. With two examples we illustrate that nodes with high degree are not necessarily classified as core nodes by bimodular surprise. We create a toy graph with a central clique of nodes which are fully connected between them. to each of these clique nodes we connect a node. This creates a clear core-periphery structure. In the first example we connect several additional nodes to one of the peripheral nodes, see Fig. 26. This peripheral node now has a higher degree than the nodes in the clique. Despite this higher degree bimodular surprise classifies only the clique as the core partition. In the second example we connect one of the peripheral nodes to all other peripheral nodes. Also in this case the peripheral nodes now has a higher degree than the nodes in the clique. Bimodular surprise classified also in this case only the clique as core nodes. These examples illustrate that a core-periphery structure as revealed by bimodular surprise is not purely related to the node degree.

B.3 Weighted bimodular surprise

We now introduce the extension to the bimodular surprise to carry out the detection of mesoscale structures accounting for weights.

B.3.1 Method

We introduce a score function for carrying out the detection of weighted mesoscale structures adapted from the unweighed case of presented in Chapter 3. Let us start with the detection of communities, that is an adaptation of the traditional surprise, Eq. 3.1, to the weighted case:

$$S^W \equiv \sum_{w=W_{int}^*}^W \frac{\binom{V_{int}+w-1}{w} \binom{V-V_{int}+W-w}{W-w}}{\binom{V+W}{W}}, \quad (\text{B.7})$$

where $V = \frac{N(N-1)}{2}$ and W the total weight of the edges. The comparison is carried out with the Weighted Random Graph.

For what concerns the detection of core-periphery structures we introduce a score function, adapted from the bimodular surprise we proposed in Chapter 3 in Eq. 3.4 we can consider the multinomial negative hypergeometric distribution:

$$S_{||}^W \equiv \sum_i \sum_j \frac{\binom{V_c+i-1}{i} \binom{V_{cp}+j-1}{j} \binom{V_p+W-(i+j)}{W-(i+j)}}{\binom{\binom{N}{2}+W}{W}}. \quad (\text{B.8})$$

B.3.2 Results

Results on toy-models show that a high weight of the edge of a node can compensate for the lack of many connection to the node. In the case of a k-star core-periphery graph, peripheral nodes will be classified as belonging to the core if the weight on the edge to the core is increased sufficiently, see Fig. 27.

Similarly when we introduce significant weights on internal links of a revealed bipartition, the newly revealed partition can switch to include

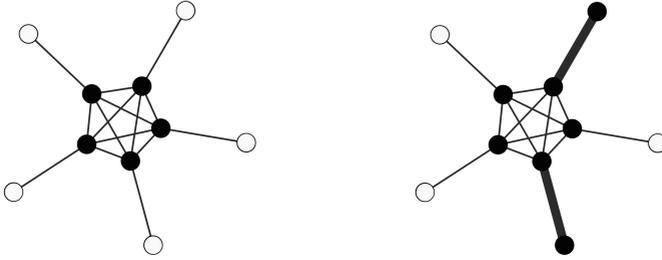


Figure 27: Core-periphery graph, the colour of the nodes indicate the two revealed partitions. When the weight on link to a peripheral nodes is increased sufficiently, such a node will be revealed as part of the core.

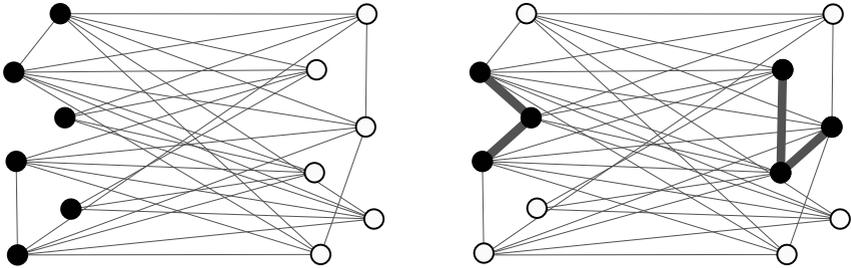


Figure 28: Bipartite graph, the color of the nodes indicate the two revealed partitions. When the weight on a few intra-cluster links is increased sufficiently the revealed partition by the weighted bimodular surprise turns from bipartite to core-periphery like.

the high weight edges in the 'core', see Fig. 28.

Appendix C

Multilayer Structure of Corporate Networks

C.1 Structural reducibility framework

The structural reducibility measure evaluates whether some layers of a multilayer network can be aggregated without loss of distinguishability from the aggregate [45]. Let our multiplex of four layers be, $\mathcal{A} = \{A^1, A^2, A^3, A^4\}$, where A^α denotes the adjacency matrix of layer α . The Von Neumann entropy of this multiplex is $H(\mathcal{A}) = \sum_1^4 h_{A^\alpha}$, where $h_{A^\alpha} = -\sum_{i=1}^N \lambda_i \log_2(\lambda_i)$, the Von Neumann entropy of $N \times N$ layer α with eigenvalues λ . We now sum two or more of the layers to find multiplex $\tilde{\mathcal{A}}$. The measure we are maximizing, the relative entropy is $q(\tilde{\mathcal{A}}) = 1 - \frac{\mathcal{H}(\tilde{\mathcal{A}})}{h_A}$, where h_A is the entropy of the aggregate of all layers.

C.2 Robustness of the multiplex regression

We test the multiplex network regression for robustness towards the 2-year time window we have considered for the stock-correlations. This window was chosen to have a well-behaved multiple pairwise correla-

Table 8: Robustness of the multiplex network regression on pairwise stock correlations. All layers are found significant independent variables for explaining the stock correlations, both for the 2-year and 1-year time window of stock correlations.

	German stock correlations				UK stock correlations			
	2-year		1-year		2-year		1-year	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Board	0.44	***	0.47	***	0.49	***	0.13	***
Ownership	0.89	***	1.06	***	0.72	***	0.18	***
Research	0.40	***	0.40	***	1.39	***	1.34	***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

tions matrix. The regression results are robust under a different time window. In Table 8 we show the regression both for the 2-year time window, and a 1-year time window (the entire year of 2017). Most importantly, in both cases all variables are significant, the magnitude of the coefficients varies but the relations between them are conserved.

C.3 Node properties regression

To establish properties of nodes that can drive the network dynamics we perform a simple regression of company properties to see which of these variables might drive the node ranking in the network. We include company characteristics revenue, revenue growth, average log return, Sharpe ratio in an ordinary least squares regression:

$$node\ ranking \sim \langle revenue \rangle + \langle revenue\ growth \rangle + \langle return \rangle + \langle Sharpe\ ratio \rangle. \quad (C.1)$$

We find that only revenue is a significant independent value of the node ranking. This results is robust for the multiplex ranking (with and without the stock interactions layer) and with the node centrality (PageRank) from the single layers.

C.4 Applications

We illustrate the use of this multiplex approach with two further examples of an interactions network of two companies before a partial takeover, and a network structure that shows clustering around audit firm choice.

C.4.1 Interwoven relations before a partial takeover: EON - RWE

The interactions between two German energy companies reveal different interactions throughout the network. Company mergers and takeovers are sometimes preceded by collaborations between the companies. From our network point of view these relations can be visualised, to help understand how interwoven companies are, but also to highlight possible conflicts of interest. As an example of the this we look at the recent deal between German energy conglomerates EON and RWE. In July 2018 the two companies reached a deal where EON will acquire the Innogy subsidiary of RWE, and RWE will end up with a significant stake in EON [47]. Our multiplex dataset shows a snapshot of the network from early 2018, well before the announcement of this deal. However we can observe a number of directors which are working both for EON and the RWE subsidiary, as well as a number of ownership ties between subsidiaries of the two companies, see Fig. 29.

C.4.2 Clustered auditor choice

Peer effects, whether through the old boys boardroom network or through herd behaviour, can steer decision making. All exchange traded companies must have their finances checked every year by an auditor. Big companies usually choose for one of the 'Big Four' audit firms: KPMG, Ernst&Young, PriceWaterhouseCoopers, and Deloitte. While there is a choice, a recent Financial Times article explains that there are limitations to this free choice [6]. Audit firms also do consultancy jobs for firms, which presents a conflict for also auditing the books. There are also ex-

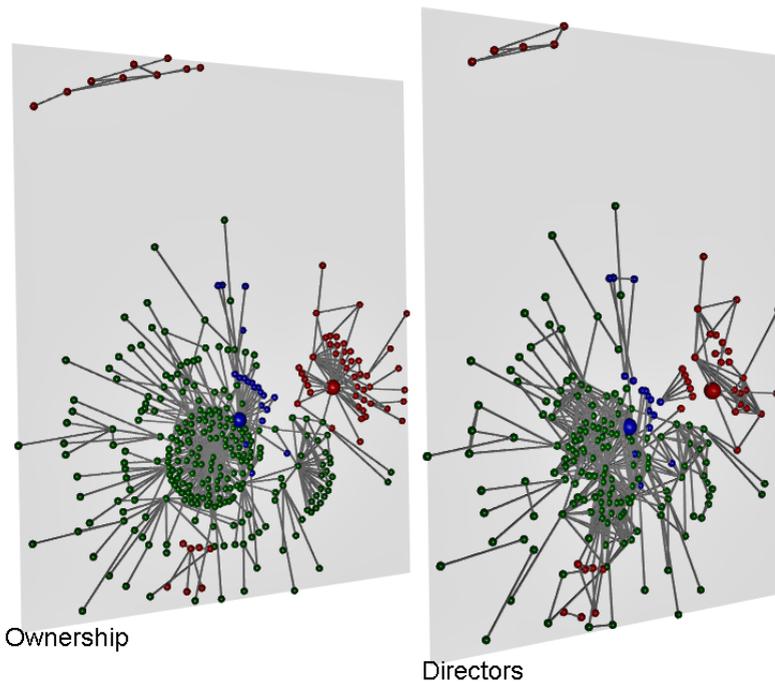


Figure 29: The multiplex network of Ownership and Board interactions between German energy conglomerates EON and RWE shows the firms are interconnected well before a partial takeover took place. This network shows relations months before the announcement of a deal in which EON (red) agreed to acquire the Innogy subsidiary (green) of RWE (blue). The two larger nodes are the stock market traded parent entities. Differences in connections in the two visualised layers show certain subsidiaries of EON that are connected to EON by ownership, but that are solely connected through common directors with RWE. Differences in these layers show the insight to gain from multiple types of corporate ties. This figure has been published under CC licence in [121].

amples where directors of a company are former partners at auditing firms, and therefore create a conflict of interest with that auditing firm [6]. The same goes for the relations between companies that we have described in our multiplex network. We can ask whether there is any re-

lation between existing connections between companies and their choice of an auditing firm. With our multiplex network, a simple test is to check whether connected nodes (companies) are more likely to use the same audit firm. For a given audit firm we calculate for all nodes, n , the average number of neighbours that use that audit firm. We then compare this average fraction of neighbours which uses an audit firm for all nodes, with the fraction for just the nodes that use that audit firm:

$$f_n = \sum_n \frac{\text{neighbors with audit firm } x}{\text{total number of neighbors}} \quad (\text{C.2})$$

$$f_{n_{\text{auditor}}} = \sum_{n_{\text{with audit firm } x}} \frac{\text{neighbors with audit firm } x}{\text{total number of neighbors}} \quad (\text{C.3})$$

We calculate these fractions for each of the big four audit firms individually. We find that companies which use the same audit firm are more connected among themselves than to other firms, especially for connections from the Board membership network. This suggests that there is a certain clustering of audit firm choice for connected firms. This effect might be due to other factors, as the co-evolution of the node attribute values (audit firm) and the network structure. The point is that results can be different on different layers of the multiplex and to show how a multiplex network can help in evaluating such questions.

Preferential choice of audit firm

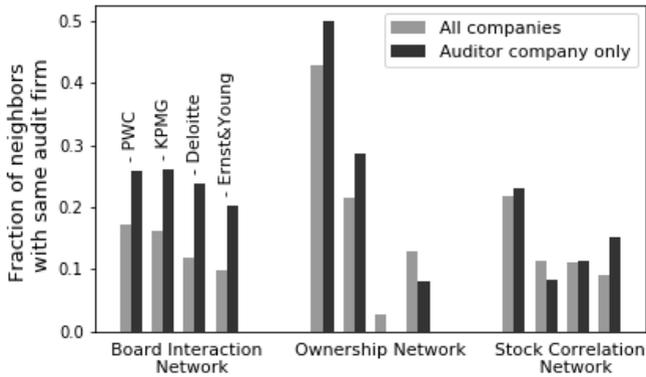


Figure 30: Firms that are connected are more likely to use the same audit firm, especially in the Board layer. Shown side by side is the average fraction of neighbours of a node that use the same indicated audit firm; for all nodes and for just the nodes of that that use that audit firm. The higher fraction for nodes of just the auditor indicates that connected firms are more likely to use the same audit firm compared to unconnected firms. There seems to be a relation between o.a. common board members and the choice for a specific audit firm. Such relation is not observed in the the stock interaction network. This figure has been published under CC licence in [121].

Appendix D

Alternative Link Prediction Methods

D.1 Other link prediction methods in terms of quantities per node

In the present section, the score functions introduced in Section 5.2.1 are rewritten in terms of the biadjacency matrix. The rationale is to provide a consistent formal framework in which all quantities can be expressed. Let us start with the CN: the bipartite extension of the Common Neighbours[42] is the number of nodes in the subgraph defined by the first neighbours of nodes (i, α) . In other words, the Common Neighbors counts the number of nodes involved in at least one “quadrangular”[42] (or *X-motif*[104]) if i and α were present. This can be expressed as:

$$\begin{aligned} s_{i\alpha}^{CN} &= \sum_j m_{j\alpha} \Theta_{\text{Heaviside}} \left(\sum_{\beta} m_{i\beta} m_{j\beta} \right) \\ &+ \sum_{\beta} m_{i\beta} \Theta_{\text{Heaviside}} \left(\sum_j m_{j\alpha} m_{j\beta} \right), \end{aligned} \tag{D.1}$$

where $\Theta_{\text{Heaviside}}$ is the Heaviside function which has a value equal to 1 if its argument is positive and to 0 otherwise. The first term in Eq. (D.1) considers the number of nodes of the layer \top that are involved in, at least,

one quadrangular insisting on (i, α) , the second is the analogous for the layer \perp . It is important to notice which it is not necessary to set $j \neq i$ or $\beta \neq \alpha$ in the summations since $m_{i\alpha} = 0$.

In the Resource Allocation, each of the terms contributing to $s_{i\alpha}^{CN}$ is weighted by the inverse of its degree, thus:

$$s_{i\alpha}^{RA} = \sum_j \frac{m_{j\alpha}}{k_j} \Theta_{\text{Heaviside}}\left(\sum_{\beta} m_{i\beta} m_{j\beta}\right) + \sum_{\beta} \frac{m_{i\beta}}{k_{\beta}} \Theta_{\text{Heaviside}}\left(\sum_j m_{j\alpha} m_{j\beta}\right), \quad (\text{D.2})$$

where we implicitly consider $m_{j\alpha}/k_j = 0$ if both the numerator and the denominator are 0.

In this formal framework, the expression of Local Community Links (LCL) is much simpler, since it is defined as the number of links in the subgraph defined by the neighbors of (i, α) . In fact, it is the number of quadrangular that would be closed by the presence of the link (i, α) , i.e.

$$s_{i\alpha}^{LCL} = \sum_{j, \beta} m_{j\alpha} m_{i\beta} m_{j\beta}. \quad (\text{D.3})$$

By construction, the value of the (bipartite) $s_{i\alpha}^{LCL}$ is limited from below by $s_{i\alpha}^{CN}$. The quantity γ [42] defined in Section 5.2 represents the degree in the subgraph of the neighbors of (i, α) . For a given node $j \in \top$, $\gamma(j)$, it can be expressed as:

$$\gamma(j) = m_{j\alpha} \sum_{\beta} m_{i\beta} m_{j\beta}$$

while the expression for a generic node $\beta \in \perp$ is analogous.

Bibliography

- [1] Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. “Systemic risk and stability in financial networks”. In: *American Economic Review* (2015). DOI: 10.1257/aer.20130456 (cit. on p. 2).
- [2] Lada a Adamic and Natalie Glance. “The Political Blogosphere and the 2004 U . S . Election : Divided They Blog”. In: *Methodology* (2005). DOI: 10.1145/1134271.1134277 (cit. on pp. 50, 54).
- [3] Rodrigo Aldecoa and Ignacio Marín. “Exploring the limits of community detection strategies in complex networks”. In: *Scientific Reports* (2013). DOI: 10.1038/srep02216 (cit. on pp. 40, 92).
- [4] Rodrigo Aldecoa and Ignacio Marín. “Surprise maximization reveals the community structure of complex networks”. In: *Scientific Reports* (2013). DOI: 10.1038/srep01060 (cit. on pp. 6, 40, 92).
- [5] Franklin Allen and Douglas Gale. “Financial Contagion”. In: *Journal of Political Economy* 108.1 (2000), pp. 1–33. DOI: 10.1086/262109 (cit. on p. 2).
- [6] “An illusion of choice: the conflicts that mire the audit world”. In: *Financial Times* (Aug. 2018). Accessed: 2018-10-01. URL: <https://www.ft.com/content/2085cab2-9af3-11e8-9702-5946bae86e6d> (cit. on pp. 113, 114).
- [7] Kartik Anand et al. “The missing links: A global study on uncovering financial network structures from partial data”. In: *Journal of Financial Stability* (2018). DOI: 10.1016/j.jfs.2017.05.012 (cit. on p. 14).

- [8] M Baltakiene et al. “Maximum entropy approach to link prediction in bipartite networks”. In: *arXiv preprint* (2018). arXiv: 1805.04307 (cit. on pp. x, 77).
- [9] L. Bargigli et al. “The multiplex structure of interbank networks”. In: *Quantitative Finance* 15.4 (2015), pp. 673–691. DOI: 10.1080/14697688.2014.968356 (cit. on p. 63).
- [10] Matteo Barigozzi, Giorgio Fagiolo, and Diego Garlaschelli. “Multinetwork of international trade: A commodity-specific analysis”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* (2010). DOI: 10.1103/PhysRevE.81.046104 (cit. on pp. 29, 31).
- [11] Matteo Barigozzi, Giorgio Fagiolo, and Giuseppe Mangioni. “Community structure in the multi-network of international trade”. In: *Communications in Computer and Information Science*. 2011. DOI: 10.1007/978-3-642-25501-4_17 (cit. on p. 17).
- [12] Matteo Barigozzi, Giorgio Fagiolo, and Giuseppe Mangioni. “Identifying the community structure of the international-trade multi-network”. In: *Physica A: Statistical Mechanics and its Applications* (2011). DOI: 10.1016/j.physa.2011.02.004 (cit. on pp. 6, 60).
- [13] Paolo Barucca and Fabrizio Lillo. “Disentangling bipartite and core-periphery structure in financial networks”. In: *Chaos, Solitons and Fractals* (2016). DOI: 10.1016/j.chaos.2016.02.004 (cit. on pp. 7, 17, 39, 52, 60).
- [14] Paolo Barucca and Fabrizio Lillo. “The organization of the interbank network and how ECB unconventional measures affected the e-MID overnight market”. In: *Computational Management Science* (2018). DOI: 10.1007/s10287-017-0293-6 (cit. on pp. 39, 52, 59, 60).
- [15] S. Battiston and M. Catanzaro. “Statistical properties of corporate board and director networks”. In: *The European Physical Journal B* 38.2 (Mar. 2004), pp. 345–352. DOI: 10.1140/epjpb/e2004-00127-8 (cit. on p. 73).
- [16] Stefano Battiston, Gerard Weisbuch, and Eric Bonabeau. “Decision spread in the corporate board network”. In: *Advances in Complex Systems* 06.04 (2003), pp. 631–644. DOI: 10.1142/S0219525903001109 (cit. on pp. 62, 64, 93).

- [17] Stefano Battiston et al. “Complexity theory and financial regulation”. In: *Science* 351.6275 (2016), pp. 818–819. DOI: 10 . 1126 / science . aad0299 (cit. on p. 1).
- [18] Stefano Battiston et al. “DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk”. In: *Scientific Reports* 2 (2012), pp. 1–6. DOI: 10 . 1038 / srep00541 (cit. on pp. 3, 16).
- [19] Stefano Battiston et al. “Liaisons dangereuses : Increasing connectivity , risk sharing , and systemic risk”. In: *Journal of Economic Dynamics and Control* 36.8 (2012), pp. 1121–1141. DOI: 10 . 1016 / j . jedc . 2012 . 04 . 001 (cit. on p. 2).
- [20] Stefano Battiston et al. “The structure of financial networks”. In: *Network Science*. Springer, 2010, pp. 131–163. DOI: 10 . 1007 / 978-1-84996-396-1_7 (cit. on p. 3).
- [21] C Becatti, G Caldarelli, and F Saracco. “Entropy-based randomisation of rating networks”. In: *ArXiv preprint* (May 2018). arXiv: 1805.00717 [physics.soc-ph] (cit. on p. 89).
- [22] S. Boccaletti et al. “The structure and dynamics of multilayer networks”. In: *Physics Reports* 544.1 (2014), pp. 1–122. DOI: <https://doi.org/10.1016/j.physrep.2014.07.001> (cit. on p. 63).
- [23] G. Bonanno et al. “Networks of equities in financial markets”. In: *The European Physical Journal B* 38.2 (Mar. 2004), pp. 363–371. DOI: 10 . 1140 / epjb / e2004-00129-6 (cit. on pp. 16, 62, 65).
- [24] Stephen P. Borgatti and Martin G. Everett. “Models of core/periphery structures”. In: *Social Networks* (2000). DOI: 10 . 1016 / S0378-8733 (99) 00019-2 (cit. on pp. 17, 39, 41, 53, 59, 92).
- [25] John P. Boyd, William J. Fitzgerald, and Robert J. Beck. “Computing core/periphery structures and permutation tests for social relations data”. In: *Social Networks* (2006). DOI: 10 . 1016 / j . socnet . 2005 . 06 . 003 (cit. on pp. 39, 92).
- [26] K.P. Burnham and D.R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed). 2002. DOI: 10 . 1016 / j . ecolmodel . 2003 . 11 . 004 (cit. on p. 60).
- [27] Kenneth P. Burnham, David R. Anderson, and Kathryn P. Huyvaert. *AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons*. 2011. DOI: 10 . 1007 / s00265-010-1029-6 (cit. on p. 26).

- [28] Fabio Caccioli, Paolo Barucca, and Teruyoshi Kobayashi. “Network models of financial systemic risk: a review”. In: *Journal of Computational Social Science* 1.1 (Jan. 2018), pp. 81–114. DOI: 10.1007/s42001-017-0008-3 (cit. on pp. 2, 16).
- [29] Fabio Caccioli et al. “Overlapping portfolios, contagion, and financial stability”. In: *Journal of Economic Dynamics and Control* (2015). DOI: 10.1016/j.jedc.2014.09.041 (cit. on p. 16).
- [30] Guido Caldarelli. *Scale-free networks: complex webs in nature and technology*. Oxford University Press, 2007. ISBN: 9780199211517 (cit. on p. 3).
- [31] Guido Caldarelli and Michele Catanzaro. *Networks: A very short introduction*. Vol. 335. Oxford University Press, 2012. ISBN: 9780199588077 (cit. on p. 4).
- [32] Carlo Vittorio Cannistraci, Gregorio Alanis-Lobato, and Timothy Ravasi. “From link-prediction in brain connectomes and protein interactomes to the local-community-paradigm in complex networks”. In: *Sci. Rep.* 3 (2013). DOI: 10.1038/srep01613 (cit. on pp. 78, 89).
- [33] Uwe Cantner and Holger Graf. “The network of innovators in Jena: An application of social network analysis”. In: *Research Policy* 35.4 (2006), pp. 463–480. DOI: <https://doi.org/10.1016/j.respol.2006.01.002> (cit. on pp. 64, 65).
- [34] Vasco M. Carvalho. “From Micro to Macro via Production Networks”. In: *Journal of Economic Perspectives* (2014). DOI: 10.1257/jep.28.4.23 (cit. on pp. 15, 51).
- [35] Giona Casiraghi. “Multiplex Network Regression: How do relations drive interactions?” In: *arXiv preprint* (2017). arXiv: 1702.02048 (cit. on p. 73).
- [36] Giulio Cimini, Andrea Gabrielli, and Francesco Sylos Labini. “The scientific competitiveness of nations”. In: *PLoS ONE* (2014). DOI: 10.1371/journal.pone.0113470 (cit. on p. 21).
- [37] Giulio Cimini et al. “Systemic Risk Analysis on Reconstructed Economic and Financial Networks”. In: *Scientific Reports* (2015). DOI: 10.1038/srep15758 (cit. on p. 11).
- [38] Giulio Cimini et al. “The statistical physics of real-world networks”. In: *Nature Reviews Physics* 1.1 (2019), p. 58. DOI: 10.1038/s42254-018-0002-6 (cit. on p. 11).

- [39] Aaron Clauset, Cosma Rohilla Shalizi, and M E J Newman. “Power-law distributions in empirical data.” In: *SIAM Review* (2007). DOI: 10.1109/ICPC.2008.18 (cit. on p. 17).
- [40] Ben Craig and Goetz von Peter. “Interbank Tiering and Money-Center Banks”. In: *Journal of Financial Intermediation* (2014). DOI: 10.1016/j.jfi.2014.02.003 (cit. on pp. 3, 33, 59, 91).
- [41] P. Csermely et al. “Structure and dynamics of core/periphery networks”. In: *Journal of Complex Networks* (2013). DOI: 10.1093/comnet/cnt016 (cit. on pp. 20, 39).
- [42] Simone Daminelli et al. “Common neighbours and the local-community-paradigm for topological link prediction in bipartite networks”. In: *New J. Phys.* 17.11 (2015). DOI: 10.1088/1367-2630/17/11/113037 (cit. on pp. 78, 81, 82, 89, 117, 118).
- [43] Gerald F. Davis, Mina Yoo, and Wayne E. Baker. “The Small World of the American Corporate Elite, 1982-2001”. In: *Strategic Organization* 1.3 (2003), pp. 301–326. DOI: 10.1177/14761270030013002 (cit. on pp. 62, 74, 93).
- [44] Manlio De Domenico, Mason A. Porter, and Alex Arenas. “MuxViz: a tool for multilayer analysis and visualization of networks”. In: *Journal of Complex Networks* 3.2 (2015), pp. 159–176. DOI: 10.1093/comnet/cnu038 (cit. on p. 66).
- [45] Manlio De Domenico et al. “Structural reducibility of multilayer networks”. In: *Nature communications* 6 (2015), p. 6864. DOI: <https://doi.org/10.1038/ncomms7864> (cit. on pp. 63, 70, 111).
- [46] Carsten F. Dormann and Rouven Strauss. “A method for detecting modules in quantitative bipartite networks”. In: *Methods in Ecology and Evolution* (2014). DOI: 10.1111/2041-210X.12139 (cit. on p. 21).
- [47] EON. “Joint Press Release: Innogy agrees with eon and rwe on fair integration processes and supports the planned transaction”. In: (). Accessed: 2018-06-01. URL: <https://www.eon.com/en/about-us/media/press-release/2018/innogy-agrees-with-eon-and-rwe-on-fair-integration-processes-and-supports-the-planned-transaction.html> (cit. on p. 113).

- [48] Ernesto Estrada and Juan A. Rodríguez-Velázquez. “Spectral measures of bipartivity in complex networks”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* (2005). DOI: 10.1103/PhysRevE.72.046105 (cit. on p. 39).
- [49] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. “On the topological properties of the world trade web: A weighted network analysis”. In: *Physica A: Statistical Mechanics and its Applications* (2008). DOI: 10.1016/j.physa.2008.01.050 (cit. on p. 60).
- [50] Giorgio Fagiolo, Tiziano Squartini, and Diego Garlaschelli. “Null models of economic networks: The case of the world trade web”. In: *Journal of Economic Interaction and Coordination* (2013). DOI: 10.1007/s11403-012-0104-7 (cit. on p. 29).
- [51] Niall Ferguson. *The Square and the Tower: Networks, Hierarchies and the Struggle for Global Power*. Penguin UK, 2017. ISBN: 9780141984810 (cit. on p. 1).
- [52] Santo Fortunato. *Community detection in graphs*. 2010. DOI: 10.1016/j.physrep.2009.11.002 (cit. on pp. 3, 20).
- [53] Santo Fortunato and Darko Hric. *Community detection in networks: A user guide*. 2016. DOI: 10.1016/j.physrep.2016.09.002 (cit. on pp. 3, 20, 90).
- [54] Daniel Fricke and Thomas Lux. “Core–Periphery Structure in the Overnight Money Market: Evidence from the e-MID Trading Platform”. In: *Computational Economics* (2014). DOI: 10.1007/s10614-014-9427-x (cit. on pp. 39, 91, 92).
- [55] Agata Fronczak and Piotr Fronczak. “Statistical mechanics of the international trade network”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* (2012). DOI: 10.1103/PhysRevE.85.056113 (cit. on p. 29).
- [56] Piotr Fronczak, Agata Fronczak, and Maksymilian Bujok. “Exponential Random Graph Models for Networks with Community Structure”. In: *Physical Review E* (2013). DOI: 10.1103/PhysRevE.88.032810 (cit. on p. 98).
- [57] Prasanna Gai and Sujit Kapadia. “Contagion in financial networks”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* (2010). DOI: 10.1098/rspa.2009.0410 (cit. on p. 16).

- [58] Man Gao et al. “Projection-based link prediction in a bipartite network”. In: *Information Sciences* 376 (2017), pp. 158–171. DOI: <https://doi.org/10.1016/j.ins.2016.10.015> (cit. on p. 78).
- [59] Diego Garlaschelli and Maria I Loffredo. “Fitness-Dependent Topological Properties of the World Trade Web”. In: *Phys. Rev. Lett.* 93.18 (Oct. 2004), p. 188701. DOI: 10.1103/PhysRevLett.93.188701 (cit. on pp. 33, 89).
- [60] Diego Garlaschelli and Maria I. Loffredo. “Multispecies grand-canonical models for networks with reciprocity”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* (2006). DOI: 10.1103/PhysRevE.73.015101 (cit. on pp. 25, 96, 97).
- [61] Diego Garlaschelli et al. “The scale-free topology of market investments”. In: *Physica A: Statistical Mechanics and its Applications* 350.2 (2005), pp. 491–499. DOI: <https://doi.org/10.1016/j.physa.2004.11.040> (cit. on p. 64).
- [62] Andrew G. Haldane and Robert M. May. “Systemic risk in banking ecosystems”. In: *Nature* (2011). DOI: 10.1038/nature09659 (cit. on p. 16).
- [63] Nobuyuki Hanaki, Ryo Nakajima, and Yoshiaki Ogura. “The dynamics of R&D network in the IT industry”. In: *Research Policy* 39.3 (2010), pp. 386–399. DOI: <https://doi.org/10.1016/j.respol.2010.01.001> (cit. on pp. 64, 65).
- [64] F. Maxwell Harper and Joseph A. Konstan. “The MovieLens Datasets”. In: *ACM Transactions on Interactive Intelligent Systems* (2015). DOI: 10.1145/2827872 (cit. on pp. 84, 89).
- [65] Vasilis Hatzopoulos et al. “Quantifying preferential trading in the e-MID interbank market”. In: *Quantitative Finance* (2015). DOI: 10.1080/14697688.2014.969889 (cit. on p. 52).
- [66] C. A. Hidalgo et al. “The product space conditions the development of nations”. In: *Science* (2007). DOI: 10.1126/science.1144581 (cit. on p. 15).
- [67] Cesar Hidalgo. *Why information grows: The evolution of order, from atoms to economies*. Basic Books, 2015. DOI: 10.1111/tesg.12240 (cit. on p. 9).

- [68] César A Hidalgo and Ricardo Hausmann. “The building blocks of economic complexity.” In: *Proc. Natl. Acad. Sci. U. S. A.* 106.26 (June 2009), pp. 10570–10575. DOI: 10 . 1073 / pnas . 0900943106 (cit. on p. 16).
- [69] Petter Holme et al. “Network bipartivity”. In: *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* (2003). DOI: 10 . 1103 / PhysRevE . 68 . 056107 (cit. on p. 39).
- [70] Daan in ’t Veld and Iman van Lelyveld. “Finding the core: Network structure in interbank markets”. In: *Journal of Banking and Finance* (2014). DOI: 10 . 1016 / j . jbankfin . 2014 . 08 . 006 (cit. on pp. 7, 17, 22, 33, 39, 59, 91, 92).
- [71] E. T. Jaynes. “Information Theory and Statistical Mechanics”. In: *Phys. Rev.* 106 (4 May 1957), pp. 620–630. DOI: 10 . 1103 / PhysRev . 106 . 620 (cit. on p. 8).
- [72] Brian Karrer and M. E.J. Newman. “Stochastic blockmodels and community structure in networks”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* (2011). DOI: 10 . 1103 / PhysRevE . 83 . 016107 (cit. on pp. 50, 60, 98).
- [73] Bisma S. Khan and Muaz A. Niazi. “Network Community Detection: A Review and Visual Survey”. In: *arXiv preprint* (2017). DOI: 10 . 1108 / IMDS - 09 - 2016 - 0408 (cit. on pp. 3, 20, 90).
- [74] Maksim Kitsak and Dmitri Krioukov. “Hidden variables in bipartite networks”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* (2011). DOI: 10 . 1103 / PhysRevE . 84 . 026114 (cit. on p. 21).
- [75] Mikko Kivelä et al. “Multilayer networks”. In: *Journal of Complex Networks* 2.3 (2014), pp. 203–271. DOI: 10 . 1093 / comnet / cnu016 (cit. on p. 63).
- [76] Sadamori Kojaku and Naoki Masuda. “Core-periphery structure requires something else in the network”. In: *New Journal of Physics* (2018). DOI: 10 . 1088 / 1367 - 2630 / aab547 (cit. on pp. 39, 59, 60, 91, 92).
- [77] Sadamori Kojaku and Naoki Masuda. “Finding multiple core-periphery pairs in networks”. In: *Physical Review E* (2017). DOI: 10 . 1103 / PhysRevE . 96 . 052313 (cit. on pp. 39, 60).

- [78] Jérôme Kunegis and Andreas Lommatzsch. “Learning spectral graph transformations for link prediction”. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 561–568. DOI: 10 . 1145 / 1553374 . 1553447 (cit. on p. 78).
- [79] Renaud Lambiotte and Marcel Ausloos. “Collaborative tagging as a tripartite network”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2006. DOI: 10 . 1007 / 11758532 _152 (cit. on p. 78).
- [80] Sary Levy-Carciente et al. “Dynamical macroprudential stress testing using network theory”. In: *Journal of Banking and Finance* (2015). DOI: 10 . 1016 / j . jbankfin . 2015 . 05 . 008 (cit. on p. 85).
- [81] David Liben-Nowell and Jon Kleinberg. “The Link Prediction Problem for Social Networks”. In: *Proc. Twelfth Annu. ACM Int. Conf. Inf. Knowl. Manag.* November 2003 (2003), pp. 556–559. DOI: 10 . 1002 / asi . v58 : 7 (cit. on pp. 78, 89).
- [82] L. Lü and T. Zhou. “Link prediction in complex networks: A survey”. In: *Physica A: Statistical Mechanics and its Applications* 390.6 (2011), pp. 1150–1170. DOI: 10 . 1016 / j . physa . 2010 . 11 . 027 (cit. on p. 81).
- [83] Duc Thi Luu et al. “Collateral Unchained: Rehypothecation networks, concentration and systemic effects”. In: *Sciences Po OFCE Working Paper n 07* (2018). DOI: 10 . 2139 / ssrn . 3123226 (cit. on pp. 1, 2, 16, 59, 91).
- [84] R. N. Mantegna. “Hierarchical structure in financial markets”. In: *The European Physical Journal B - Condensed Matter and Complex Systems* 11.1 (Sept. 1999), pp. 193–197. DOI: 10 . 1007 / s100510050929 (cit. on pp. 3, 16, 65).
- [85] Rossana Mastrandrea et al. “Enhanced reconstruction of weighted networks from strengths and degrees”. In: *New Journal of Physics* (2014). DOI: 10 . 1088 / 1367 - 2630 / 16 / 4 / 043022 (cit. on p. 20).
- [86] Rossana Mastrandrea et al. “Reconstructing the world trade multiplex: The role of intensive and extensive biases”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* (2014). DOI: 10 . 1103 / PhysRevE . 90 . 062804 (cit. on p. 29).

- [87] Robert M May, Simon A Levin, and George Sugihara. “Complex systems: Ecology for bankers”. In: *Nature* 451.7181 (2008), p. 893. DOI: 10.1038/451893a (cit. on pp. 1, 16).
- [88] M Newman. “Modularity and community structure in networks”. In: *Proceedings of the National Academy of Sciences* (2006). DOI: 10.1073/pnas.0601602103 (cit. on p. 6).
- [89] Mark Newman. *Networks: an introduction*. Oxford university press, 2010. DOI: 10.1093/acprof:oso/9780199206650.001.0001 (cit. on pp. 2–4, 9).
- [90] Carlo Nicolini and Angelo Bifone. “Modular structure of brain functional networks: Breaking the resolution limit by Surprise”. In: *Scientific Reports* (2016). DOI: 10.1038/srep19250 (cit. on pp. 40, 42, 43, 49, 59, 60, 92, 104).
- [91] OECD. “Direct investment – 10 percent threshold of voting power/equity ownership, employment”. In: *DITEG Issue papers* (2004). Accessed: 2018-06-01. URL: <https://www.imf.org/external/pubs/ft/bop/2004/04-31.pdf> (cit. on p. 64).
- [92] Liming Pan et al. “Predicting missing links and identifying spurious links via likelihood analysis”. In: *Sci. Rep.* 6 (2016). DOI: 10.1038/srep22955 (cit. on p. 78).
- [93] F Parisi, G Caldarelli, and T Squartini. “Entropy-based approach to missing-links prediction”. In: *ArXiv preprint* (Feb. 2018). arXiv: 1802.02064 [physics.soc-ph] (cit. on pp. 79, 82, 89, 93).
- [94] Juyong Park and Mark EJ Newman. “Statistical mechanics of networks”. In: *Phys. Rev. E* 70.6 (Dec. 2004), p. 66117. DOI: 10.1103/PhysRevE.70.066117 (cit. on p. 89).
- [95] Tiago P. Peixoto and Stefan Bornholdt. “Evolution of robust network topologies: Emergence of central backbones”. In: *Physical Review Letters* (2012). DOI: 10.1103/PhysRevLett.109.118703 (cit. on p. 59).
- [96] Carlo Piccardi and Lucia Tajoli. “Existence and significance of communities in the World Trade Web”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* (2012). DOI: 10.1103/PhysRevE.85.066119 (cit. on p. 60).

- [97] Rafael Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. “Corporate ownership around the world”. In: *The journal of finance* 54.2 (1999), pp. 471–517. DOI: 10.1111/0022-1082.00115 (cit. on p. 64).
- [98] Christoph Rahmede et al. “Centralities of nodes and influences of layers in large multiplex networks”. In: *Journal of Complex Networks* 6.5 (2018), pp. 733–752. DOI: 10.1093/comnet/cnx050 (cit. on p. 74).
- [99] Amanah Ramadiah, Fabio Caccioli, and Daniel Fricke. “Reconstructing and Stress Testing Credit Networks”. In: (2017). DOI: 10.2139/ssrn.3084543 (cit. on p. 14).
- [100] Paul Resnick et al. “GroupLens : An Open Architecture for Collaborative Filtering of Netnews”. In: *Proceedings of the 1994 ACM conference on Computer supported cooperative work* (1994). DOI: 10.1145/192844.192905 (cit. on p. 89).
- [101] Armando Rungi, Gregory Morrison, and Fabio Pammolli. “Global ownership and corporate control networks”. In: *IMT Working Paper Series* (2017). DOI: 10.2139/ssrn.3031955 (cit. on pp. 62, 93).
- [102] Fabio Saracco et al. “Detecting early signs of the 2007-2008 crisis in the world trade”. In: *Scientific Reports* (2016). DOI: 10.1038/srep30286 (cit. on pp. 11, 14).
- [103] Fabio Saracco et al. “Inferring monopartite projections of bipartite networks: An entropy-based approach”. In: *New Journal of Physics* (2017). DOI: 10.1088/1367-2630/aa6b38 (cit. on pp. 7, 21, 78, 91).
- [104] Fabio Saracco et al. “Randomizing bipartite networks: The case of the World Trade Web”. In: *Scientific Reports* (2015). DOI: 10.1038/srep10595 (cit. on pp. 21, 82, 89, 100, 117).
- [105] Frank Schweitzer et al. “Economic Networks: The New Challenges”. In: *Science* 325.5939 (2009), pp. 422–425. DOI: 10.1126/science.1173644 (cit. on pp. 1, 16).
- [106] Ma Ángeles Serrano and Marián Boguñá. “Topology of the world trade web”. In: *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* (2003). DOI: 10.1103/PhysRevE.68.015101 (cit. on p. 29).

- [107] Squartini Tiziano and Garlaschelli Diego. *Maximum-entropy networks. Pattern detection, network reconstruction and graph combinatorics*. Springer, 2017. ISBN: 9783319694368 (cit. on p. 14).
- [108] Tiziano Squartini, Giorgio Fagiolo, and Diego Garlaschelli. “Randomizing world trade. I. A binary network analysis”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* (2011). DOI: 10.1103/PhysRevE.84.046117 (cit. on pp. 14, 17).
- [109] Tiziano Squartini and Diego Garlaschelli. “Analytical maximum-likelihood method to detect patterns in real networks”. In: *New Journal of Physics* (2011). DOI: 10.1088/1367-2630/13/8/083001 (cit. on pp. 10, 12, 20, 89).
- [110] Tiziano Squartini, Iman Van Lelyveld, and Diego Garlaschelli. “Early-warning signals of topological collapse in interbank networks”. In: *Scientific Reports* (2013). DOI: 10.1038/srep03357 (cit. on pp. 11, 14, 17, 33, 36).
- [111] Tiziano Squartini et al. “Enhanced capital-asset pricing model for the reconstruction of bipartite financial networks”. In: *Physical Review E* (2017). DOI: 10.1103/PhysRevE.96.032315 (cit. on p. 21).
- [112] Mika J. Straka, Guido Caldarelli, and Fabio Saracco. “Grand canonical validation of the bipartite international trade network”. In: *Physical Review E* (2017). DOI: 10.1103/PhysRevE.96.022306 (cit. on pp. 15, 17).
- [113] Mika J Straka et al. “From Ecology to Finance (and Back?): Recent Advancements in the Analysis of Bipartite Networks”. In: *arXiv preprint* (2017), pp. 1–26. arXiv: 1710.10143 (cit. on pp. 7, 78, 91).
- [114] Giovanni Strona et al. “A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals”. In: *Nature Communications* (2014). DOI: 10.1038/ncomms5114 (cit. on p. 21).
- [115] Andrea Tacchella et al. “A new metrics for countries’ fitness and products’ complexity”. In: *Scientific Reports* (2012). DOI: 10.1038/srep00723 (cit. on p. 21).
- [116] Mario V. Tomasello et al. “The rise and fall of R and D networks”. In: *Industrial and Corporate Change* 26.4 (2017), pp. 617–646. DOI: 10.1093/icc/dtw041 (cit. on pp. 62, 93).

- [117] S. Torreggiani et al. “Identifying the community structure of the food-trade international multi-network”. In: *Environmental Research Letters* (2018). DOI: 10 . 1088 / 1748 - 9326 / aabf23 (cit. on p. 29).
- [118] V. A. Traag, R. Aldecoa, and J. C. Delvenne. “Detecting communities using asymptotical surprise”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* (2015). DOI: 10 . 1103 / PhysRevE . 92 . 022816 (cit. on pp. 40, 47, 92).
- [119] M. Tumminello et al. “A tool for filtering information in complex systems”. In: *Proceedings of the National Academy of Sciences* 102.30 (2005), pp. 10421–10426. DOI: 10 . 1073 / pnas . 0500298102 (cit. on pp. 16, 62).
- [120] Michele Tumminello et al. “Statistically validated networks in bipartite complex systems”. In: *PLoS ONE* (2011). DOI: 10 . 1371 / journal . pone . 0017994 (cit. on p. 21).
- [121] Jeroen van Lidth de Jeude, Tomaso Aste, and Guido Caldarelli. “The multilayer structure of corporate networks”. In: *New Journal of Physics* (2019). DOI: 10 . 1088 / 1367 - 2630 / ab022d (cit. on pp. x, 61, 66, 72, 114, 116).
- [122] Jeroen van Lidth de Jeude, Guido Caldarelli, and Tiziano Squartini. “Detecting Core-Periphery Structures by Surprise”. In: *arXiv preprint* (2018). arXiv: 1810 . 04717 (cit. on pp. x, 38).
- [123] Jeroen van Lidth de Jeude et al. “Reconstructing mesoscale network structures”. In: *Complexity* 2019 (2019), p. 13. DOI: 10 . 1155 / 2019 / 5120581 (cit. on pp. x, 19, 51, 55).
- [124] Yogesh Virkar and Aaron Clauset. “Power-law distributions in binned empirical data”. In: *Annals of Applied Statistics* (2014). DOI: 10 . 1214 / 13 - AOAS710 (cit. on p. 17).
- [125] Stefania Vitali, James B. Glattfelder, and Stefano Battiston. “The Network of Global Corporate Control”. In: *PLOS ONE* 6.10 (Oct. 2011), pp. 1–6. DOI: 10 . 1371 / journal . pone . 0025995 (cit. on pp. 6, 7, 15, 17, 18, 62, 64, 74, 93).
- [126] Website. *Benchmark Input-Output Data*. URL: <https://www.bea.gov/industry/benchmark-input-output-data> (visited on 10/31/2018) (cit. on p. 51).

- [127] Website. *NetScience Collaboration network in science of networks*. URL: <http://vlado.fmf.uni-lj.si/pub/networks/data/collab/netscience.htm> (visited on 05/01/2018) (cit. on p. 50).
- [128] Website. *UN Comtrade Database*. URL: <http://comtrade.un.org/> (cit. on p. 22).
- [129] Rong Yang, Leyla Zhuhadar, and Olfa Nasraoui. “Bow-tie decomposition in directed graphs”. In: *14th International Conference on Information Fusion* (2011) (cit. on p. 23).
- [130] Muhammed A Yildirim and Michele Coscia. “Using random walks to generate associations between objects”. In: *PloS one* 9.8 (2014). DOI: 10.1371/journal.pone.0104813 (cit. on p. 78).
- [131] Paolo Zacchia. “Knowledge Spillovers through Networks of Scientists”. In: (2017). URL: <http://eprints.imtlucca.it/id/eprint/3056> (cit. on p. 64).
- [132] Edward J. Zajac and James D. Westphal. “Director reputation, CEO/board power, and the dynamics of board interlocks”. In: vol. 1996. 1. 1996, pp. 254–258. DOI: 10.5465/ambpp.1996.4980568 (cit. on pp. 18, 62, 93).
- [133] Xiao Zhang, Travis Martin, and M. E.J. Newman. “Identification of core-periphery structure in networks”. In: *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* (2015). DOI: 10.1103/PhysRevE.91.032803 (cit. on pp. 39, 59, 60).
- [134] Yaojia Zhu, Xiaoran Yan, and Cristopher Moore. “Oriented and degree-generated block models: Generating and inferring communities with inhomogeneous degree distributions”. In: *Journal of Complex Networks* (2014). DOI: 10.1093/comnet/cnt011 (cit. on p. 98).