IMT School for Advanced Studies, Lucca
Lucca, Italy

Gravity Models of Networks

PhD Program in Systems Science
Track in Economics, Networks and Business Analytics
XXXIV Cycle

By

Marzio Di Vece

2024

The dissertation of Marzio Di Vece is approved.

PhD Program Coordinator: Prof. Alberto Bemporad, IMT School for Advanced Studies Lucca

Advisor: Prof. Diego Garlaschelli, IMT School for Advanced Studies Lucca

Co-Advisor: Prof. Tiziano Squartini, IMT School for Advanced Studies Lucca

The dissertation of Marzio Di Vece has been reviewed by:

Prof. Giorgio Fagiolo, Sant'Anna School of Advanced Studies

Prof. Maria Ángeles Serrano, University of Barcelona

IMT School for Advanced Studies Lucca
2024

To Orietta, Oscar and Valerio

Contents

List of Figures xi

List of Tables xiii

Acknowledgements xv

Vita and Publications xvii

Abstract xxi

1 Introduction 1

2 Discrete entropy-based econometric models 9
- 2.1 Modelling positive weights . . . 9
- 2.2 Statistical network models . . . 11
  - 2.2.1 Econometric models . . . 12
  - 2.2.2 Maximum-entropy models . . . 16
- 2.3 Results . . . 22
- 2.4 Discussion . . . 39

3 Continuous entropy-based econometric models 44
- 3.1 Introduction . . . 44
- 3.2 Statistical network models . . . 46
  - 3.2.1 Conditional models . . . 46
  - 3.2.2 Integrated models . . . 55
- 3.3 Results . . . 57
  - 3.3.1 Model selection via performance indicators . . . 58

3.3.2 Model selection via statistical tests 63
3.3.3 Model selection via information criteria 64
3.3.4 The Shannon-Fisher plane 66
3.4 Discussion 70
3.5 The DyGyS Python package 73

4 Beyond the deterministic estimation of parameters in conditional models 75
4.1 Introduction 76
4.2 Minimization of the KL divergence 76
4.3 Estimation of the parameters 78
4.3.1 'Deterministic' parameter estimation 78
4.3.2 'Annealed' parameter estimation 79
4.3.3 'Quenched' parameter estimation 80
4.4 Results 81
4.4.1 'Scalar' variant of the CEM 82
4.4.2 'Vector' variant of the CEM 84
4.4.3 'Econometric' variant of the CEM 87
4.5 Discussion 89

5 Triadic structure of the Dutch Production Multiplex 92
5.1 Introduction 92
5.2 The Dutch Production Multiplex 96
5.3 Maximum-entropy randomization methods 98
5.3.1 Binary benchmarks 99
5.3.2 Conditional weighted benchmarks 101
5.4 Results 106
5.4.1 Binary Motif Analysis 109
5.4.2 Weighted Motif Analysis 114
5.4.3 The NuMeTriS Python package 119
5.5 Discussion 119

6 Conclusions 123

A Discrete-valued models **128**
A.1 Estimating the GM parameters 128
A.2 Econometric models 130
A.2.1 Poisson model 130
A.2.2 Negative binomial model 131
A.2.3 Zero-inflated Poisson model 132
A.2.4 Zero-inflated negative binomial model 133
A.3 Maximum-entropy models 134

B Continuous-valued models **139**
B.1 Conditional models 139
B.1.1 Conditional exponential model 140
B.1.2 Conditional gamma model 141
B.1.3 Conditional Pareto model 142
B.1.4 Conditional log-normal model 142
B.2 Integrated models 143
B.3 Turning structural models into econometric ones 145
B.3.1 Conditional exponential model 146
B.3.2 Conditional gamma model 146
B.3.3 Conditional Pareto model 147
B.3.4 Conditional log-normal model 147
B.4 The Shannon-Fisher plane 148
B.4.1 Conditional exponential model 148
B.4.2 Conditional gamma model 149
B.4.3 Conditional Pareto model 149
B.4.4 Conditional log-normal model 150

C Three, different parameter estimation procedures **151**
C.1 The functional form of conditional models 151
C.2 Estimating the parameters 152
C.2.1 'Scalar' or homogeneous variant of the CEM 152
C.2.2 'Vector' or weakly heterogeneous variant of the CEM 155
C.2.3 'Tensor' variant of the CEM 155
C.2.4 'Econometric' variant of the CEM 156

D Maximum-entropy models for motifs detection **158**
D.1 Binary null models 158
D.1.1 The Directed Binary Configuration Model (DBCM) 158
D.1.2 The Reciprocal Binary Configuration Model (RBCM) 159
D.2 Conditional weighted null models 161
D.2.1 Conditional Reconstruction Model A 161
D.2.2 Conditionally Reciprocal Weighted Configuration Model 162

List of Figures

1 Scatter plot of the entire set of positive WTW weights w_{ij} versus the values predicted by different specifications of the GM. 10

2 Performance of econometric models versus the performance of ME models in reproducing binary network statistics. . . 24

3 Performance of econometric models versus the performance of ME models in reproducing weighted network statistics. 26

4 Percentage of times the empirical value of a given network statistics is compatible with its ensemble distribution, according to a KS test at the significance level of 5%. 29

5 Percentage of times the empirical value of a given network statistics is compatible with its ensemble distribution, according to a KS test at the significance level of 5%. 31

6 Reconstruction Accuracy for eigenvector, Katz, closeness and betweenness centrality measures. 32

7 Average Spearman Rank Correlations for eigenvector and Katz centrality measures. 34

8 Performance of the negative binomial model versus the performance of the ME model described by the Hamiltonian $H_{(2)}$, in reproducing binary and weighted network statistics. 41

9 Reconstruction accuracy for the network statistics of interest. 61

10 Reconstruction accuracy RA_m^w for the Gleditsch and the BACI datasets. 62
11 AIC values for the binary and the full log-likelihood of our ME models for the Gleditsch and the BACI datasets. 67
12 Shannon-Fisher plane for each, conditional ME model with continuous-valued weights. 69

13 Estimations of the parameter β, entering the definition of the homogeneous version of the CEM. 83
14 Estimations of the parameter β_{166} entering the definition of the weakly heterogeneous version of the CEM. 85
15 Estimations of the parameters β_0, ρ, α and γ, entering the definition of the econometric version of the CEM. 86

16 Graphical representation of a multi-layer production network and of the existing 13 triadic motifs. 93
17 Normalized Triadic Occurrences and Fluxes for the aggregated network and a sample of commodity layers. 108
18 Triadic, binary motifs analysis: directed vs reciprocal model. 110
19 Comparison of DBCM and RBCM on motif statistics across commodities 111
20 Triadic, weighted motifs analysis: directed vs reciprocal model. 115
21 Comparison of DBCM + $CReM_A$ and RBCM + CRWCM on motif statistics across commodities 116

22 Estimations of the parameters β_{168}, β_{170} and β_{171}, entering the definition of the weakly heterogeneous version of the CEM. 154
23 Empirical CDFs for the parameters entering the definition of the econometric version of the CEM. 156

List of Tables

1 Accuracy of ME and econometric models in reconstructing both the Gleditsch and the BACI datasets. 27

2 Ranking of ME and econometric models, according to the AIC and BIC values. 36

3 One-lagged accuracy, quantifying the performance of ME and econometric models in providing one-lagged predictions on both the Gleditsch and the BACI datasets. 38

4 Performance of ME and econometric models in providing accurate predictions of weights, on both the Gleditsch and the BACI datasets. 39

5 Summary of the Model specifications used in chapter 2. . . 42

6 Summary of the models and specific constraints used in chapter 2. 43

7 Compatibility between the distributions of the expected values of the statistics output by the ME models and their empirical counterparts. 60

8 Increments of the four indicators composing the confusion matrix when passing from the UBCM to the integrated exponential model. 65

9 Summary of the Model specifications used in chapter 3. . . 73

10 Summary of the models and specific constraints used in chapter 3. 74

11 Description of global network statistics across commodity layers. 105
12 Summary of the Model specifications used in chapter 5. . . 122
13 Summary of the models and specific constraints used in chapter 5. 122

Acknowledgements

I am grateful to my advisor, Diego Garlaschelli, for his guidance throughout my 'doctoral journey' and his contribution to the projects that led to the papers presented in chapter 2, chapter 3, chapter 4 and chapter 5.

I am also deeply indebted with my co-advisor, Tiziano Squartini, for his guidance, his feedback and his contribution to the papers upon which chapter 2, chapter 3 and chapter 4 are based.

I would also like to thank Frank P. Pijpers for the opportunity he gave me to conduct my research at CBS-Statistics Netherlands, in The Hague, and his contribution to the paper presented in chapter 5.

My PhD journey has been a long and winding road, filled with both triumphs and setbacks, breakthroughs and frustrations. I am immensely grateful for the many individuals who have enriched my experience during these years, starting with Ivan and Arturo, my very first friends at IMT.

My heartfelt appreciation extends to the entire Networks community, both seniors and juniors, whose insights, often shared during our lively Nedos discussions or aperitivo gatherings, have been invaluable in expanding my knowledge and broadening my perspectives. It would be impossible to single out each member of the IMT community, but I am deeply thankful for their camaraderie and the countless laughter shared over coffee breaks.

Special thanks also to my friends back in Salerno, especially Matteo, Rosario, Francesco & Anna Rita, Antonio, Vito, Benedetta and The Twins, who have always welcomed me with open

arms and made returning home a joyous occasion. Their unwavering support and willingness to engage in stimulating conversations have been a source of immense comfort and inspiration.

My deepest gratitude also goes to my family, who have been my unwavering pillars of support throughout my academic journey. I am forever indebted to my mother and father for their unwavering belief in me and their tireless efforts to ensure I received the education I desired. My brother Valerio deserves special recognition for instilling in me a strong sense of discipline and a problem-solving mindset.

Finally, I owe an immeasurable debt of gratitude to Veronica and her family. Her unwavering support, unwavering belief in me and genuine affection have been the driving forces behind my success. I cherish every moment we shared.

Vita

April 6, 1991	Born in Salerno (IT)
2014	Bachelor's Degree in Physics University of Salerno, Fisciano (IT)
2017	Master's Degree in Physics (110/110 cum laude) University of Salerno, Fisciano (IT)
2017	Visiting Erasmus Trainee IFISC Institute for Cross-Disciplinary Physics and Complex Systems, University of the Balearic Islands, Palma de Mallorca (E)
2018-2023	PhD Student in Systems Science (track in ENBA) IMT School for Advanced Studies, Lucca (IT)
2022	Statistical Researcher Intern Statistics Netherlands, The Hague (NL)
2022	Guest Researcher Lorentz Institute for Theoretical Physics, Leiden (NL)
2022-2023	Research Collaborator IMT School for Advanced Studies, Lucca (IT)
2023-2024	Post-Doctoral Researcher Scuola Normale Superiore, Pisa (IT)

Publications

1. M. Di Vece, F. P. Pijpers and D. Garlaschelli, *Commodity-specific triads in the Dutch inter-industry production network*, arXiv:2305.12179 (2023) (currently under review at Scientific Reports).
2. M. Di Vece, D. Garlaschelli and T. Squartini, *Deterministic, quenched, and annealed parameter estimation for heterogeneous network models*, Physical Review E **108**, 054301 (2023).
3. M. Di Vece, D. Garlaschelli and T. Squartini, *Reconciling econometrics with continuous maximum-entropy network models*, Chaos Solitons and Fractals **166**, 112958 (2023).
4. M. Di Vece, D. Garlaschelli and T. Squartini, *Gravity models of networks: integrating maximum-entropy and econometric approaches*, Physical Review Research **4**, 033105 (2022).

Presentations

1. *Commodity-specific triads in the Dutch inter-industry production network.*
   NetSci-X 2024 (Venezia, 22-25 January 2024).
2. *Reconciling econometrics with maximum-entropy network models*.
   NetSci 2023 (Vienna, 10-14 July 2023).
3. *Deterministic, quenched or annealed? Differences in the parameter estimation of heterogeneous network models*.
   FinEcoNets, NetSci 2023 (Vienna, 10 July 2023).
4. *Commodity-specific triads in the Dutch inter-industry production network.*
   'Complex networks: theory, methods, and applications' Lake Como School (Como, 22-26 May 2023).
5. *Maximum-Entropy models for Network Regression Analysis on Trade Data*.
   NetSci 2022 (online, 25-29 July 2022).
6. *Gravity models of networks: integrating maximum entropy and econometric approaches*.
   Networks 2021 (online, 5-10 July 2021).
7. *Gravity models of networks: integrating maximum entropy and econometric approaches*.
   Conference of the Italian Society of Statistical Physics (Parma, 23-25 June 2021).

Posters

1. *Network econometrics: unbiased estimation of extensive and intensive margins*.
   Entropy 2021 (5-7 May 2021).
2. *The topology of the International Trade: comparing network-based and econometric approaches*.
   NetSci 2020 (online, 17-25 September 2020).

Codes

1. **DyGyS**: *DYadic GravitY regression models with Soft constraints*
   Available at https://pypi.org/project/DyGyS/
2. **NuMeTriS**: *Null Models for Triadic Structures*
   Available at https://pypi.org/project/NuMeTriS/

Abstract

Trade networks are mathematical representations of the exchanges established by countries, industries, firms or individuals. The present thesis collects works aimed at overcoming the limitations characterizing the econometric recipes traditionally employed to study the aforementioned systems, by introducing a novel framework, based upon the maximum-entropy formalism. In chapter 2 we develop a novel class of models to study networks with discrete weights, capable of accommodating both structural and econometric parameters, finding that they outperform standard, econometric models [1]. In chapter 3 we extend the aforementioned set of models to study networks with continuous weights [2]. In chapter 4 we go beyond the 'deterministic' optimization procedure prescribed by econometrics to specify conditional models, considering two, alternative estimation recipes characterized by different ways of averaging over the topological randomness: what we find is that the 'annealed' recipe, prescribing to maximize a generalized likelihood function, is to be preferred, regardless of the heterogeneity of weights [3]. Finally, in Chapter 5, we delve into the extent to which the triadic structures embedded within the Dutch multi-commodity production network align with maximum-entropy conditional models [4]. Our findings reveal that for the vast majority of commodities, these models effectively replicate the observed triadic structures, exhibiting minimal deviations.

Chapter 1

Introduction

The ever-increasing availability of data is shedding light on the structure of human interactions, revealing that the latter can be well-represented by mathematical objects called *networks*, i.e. collections of nodes connected by edges - an abstraction which is so general to allow for the study of a wide range of interactions.

The 2008 financial crisis clearly pointed out that 'network effects matter' [5] with trade linkages being the preferential route of interaction between world countries: as the disappearance of a link, or a change in its 'weight', can lead to the transmission of shocks [6, 7, 8, 9, 10, 11, 12, 13], understanding the structural properties of economic networks has become necessary to increase the resilience of the entire economic-financial system.

An important class of economic networks is the one defined by trade exchanges, where the nodes (i.e. the 'actors') can be countries, industries or firms: two examples are provided by production networks, whose nodes are firms and whose edges are input/output relationships, and the World Trade Web (WTW), whose nodes are countries and whose edges are import/export relationships.

The surge in global trade data has spurred researchers from diverse fields like economics, network science, and social sciences to investigate

its network structure through empirical analysis, complementing the established theoretical framework in traditional trade economics.

These efforts have resulted in a substantial body of research exploring the structural characteristics of the World Trade Web (WTW), highlighting its unique features and patterns compared to other real-world networks [14, 16, 17, 18, 19, 20, 21, 22, 23, 5, 15, 9].

One strand of this research focuses on the binary properties of the WTW, examining the network's connectivity based solely on the presence or absence of trade relationships between countries. Studies in this area have revealed a high density of connections, a skewed and heavy-tailed distribution of the number of trade partners per country, disassortative mixing by degree (indicating a tendency for countries with many trade partners to connect with those with fewer), a hierarchical organization, the presence of a core of highly interconnected countries, and a bow-tie structure [19, 20, 22, 15, 24].

Another stream of research delves into the weighted properties of the WTW, considering the trade volume associated with each link and the overall trade volume of each country. These studies have uncovered right-skewed and heavy-tailed distributions of trade volume and trade partner strength (indicating a concentration of high-volume trade relationships), disassortative mixing by strength (showing a link between trade volume and connection strength), and a high weighted clustering coefficient (demonstrating that countries with many trade partners tend to have more connected partners) [6].

Theoretical models of international trade can be broadly divided into two categories: econometric models and network models, with the latter primarily rooted in statistical physics. The earliest international trade model, proposed in 1962 by Dutch economist Jan Tinbergen, employed a relationship analogous to the law of gravity to simulate trade flows between countries [25]. This model, known as the GM, has been shown to accurately predict positive trade volumes [26, 27, 28, 29]. While the GM has been successfully applied to various networks, including migration flows [30, 31], mobility and traffic patterns [32, 33, 34], communication streams [35], and spreading phenomena [36, 37], it has been criticized for

predicting that all countries will trade with each other. This prediction contradicts empirical data, which indicates that a significant proportion (up to half) of possible trade relationships are absent [38]. Since overestimating connections can lead to inaccurate network effects [5], the pure GM should not be considered a reliable model of the World Trade Web (WTW) [29].

Since the 1960s, economists have been working to address the limitations of the gravity model. Early approaches, such as those proposed by Eaton and Tamura [39] and Martin and Pham [40], involved rounding low-volume trade flows to zero, but the arbitrary threshold used in this method raised conceptual concerns. In response, Helpman, Melitz, and Rubinstein [41] introduced a two-step estimation procedure: first, a probit model is used to predict the likelihood of a trade relationship, and then, an Ordinary Least Squares (OLS) regression is employed to estimate the corresponding trade volume. Another alternative, proposed by Silva and Tenreyro [42], is to estimate the gravity equation in a multiplicative rather than additive manner using a Poisson Pseudo Maximum Likelihood (PPML) method. This approach provides robust estimates of trade flows, but it is sensitive to the presence of many zero trade flows. In response, zero-inflated (ZI) methods have been developed, such as those proposed by Burger and Winkelmann [43, 44]. These models involve a logit estimation to determine the presence of a trade relationship followed by either a Poisson (ZIP) or a negative binomial (ZINB) regression to estimate the corresponding trade volume. In 2013, Dueñas and Fagiolo demonstrated that the weighted structure of the World Trade Web (WTW) can be accurately predicted by gravity models only when the network's topological structure is also specified [45]. This finding shifts the focus from solely predicting trade flows to also reconstructing the network's topology.

Physics-inspired approaches, like the maximum-entropy framework, enable us to investigate structural network properties and determine the most unbiased probability distribution that aligns with these features. Studies have demonstrated that solely fixing the degree sequence of world countries can accurately replicate higher-order binary charac-

teristics, such as the average nearest-neighbor degree and the clustering coefficient [19, 46, 47, 48, 49, 50, 29]. Nevertheless, when the analysis is confined to a purely weighted framework, this outcome is not observed. In this instance, the strength sequence alone is not a reliable constraint for replicating the average nearest-neighbor strength and the weighted clustering coefficient [51, 52]. In fact, accurate reconstruction of weighted characteristics can only be achieved by simultaneously constraining both degrees and strengths [53, 54, 55].

So far, econometric and physics-inspired approaches have mostly proceeded on separate paths, the former focusing on the effects of economic covariates, such as gross domestic products (GDPs) and geographical distances, on the intensity of trade (usually defined as the yearly dollar value of the traded products) and the latter focusing on the accurate estimation of purely structural properties while ignoring the genuinely economic characteristics of the agents. More specifically, most econometric models estimate the parameters of a regression: for instance, the expected volume of trade between the countries i and j is estimated via the Gravity Model (GM) specification reading

$$\langle w_{ij} \rangle_{\mathrm{GM}} = \rho \, \mathrm{GDP}_i^{\alpha} \mathrm{GDP}_j^{\beta} d_{ij}^{\gamma} \tag{1.1}$$

where ρ is a dimensional parameter 'adjusting' the unit of measure, GDP_i is the gross domestic product of country i, GDP_j is the gross domestic product of country j and d_{ij} is their geographic distance[1]; instead, the parameters α, β, γ account for the impact of the economic factors onto the trade volumes. Interestingly, recipes like the one above are not capable of accurately reproducing the value of any network statistics without taking the whole structure as input [45]. By contrast, most of the physics-inspired models that have been proposed so far are based upon entropy-maximization [46, 51, 50, 56, 57, 58, 29], a general prescription leading to the definition of statistical ensembles induced by constraining a subset of network properties [59, 60, 61, 62, 63, 58, 51].

[1]The GM specification can be modified by adding other covariates such as common language, common religion and regional trade agreements [41].

A first attempt of combining the econometric and the physics-inspired approaches is represented by the reformulation of maximum-entropy models as *hidden variable*, or *fitness*, models [64, 19, 20, 65, 66, 29]. In this approach, the Lagrange multipliers, employed to enforce the chosen constraints, are identified with empirical, macroeconomic factors (e.g. the GDP of countries); a second attempt is represented by the approach proposed in [29] where a logit model describes the presence of each, single link and a GM specification describes the expected, conditional weight - although the total amount of empirical trade is not compatible with the one predicted by the model, which underestimates the former by a factor of 10^4 [1]. Another type of maximum-entropy model, inspired by the literature on random spatial networks, is discussed by Boguna et al. [67]. This model was found to exhibit small-worldness and non-zero clustering in the thermodynamic limit. This approach was extended to weighted networks using a two-step process [68] that controls for the dependence of link weight on topological measures such as degree and geometrical measures such as geographical distance. With a more abstract space, inspired by the coupling between popularity, resulting in higher degree centrality, and similarity, resulting in a function of trade friction in the trading of countries, Garcia-Perez et al. produced an unweighted backbone of the WTW [69] capable of exploring significant trade channels. Still, identifying the best way of integrating entropy-based models with econometric specifications (e.g. one guaranteeing that both network statistics *and* their variability is accurately reproduced in order to achieve a sufficiently precise description of economic, marginal effects) remains a debated issue.

In this thesis, we try to make progress by introducing a class of entropy-based models capable of intaking both structural constraints (e.g. the number of links, the degree sequence, the total volume of trade) and a flexible GM specification. As we will show in chapter 2 and chapter 3, these models admit two different formulations, i.e. the *integrated* and the *conditional* one, depending on whether the probability distributions describing the binary adjacency matrix and the weights are either dependent or independent; besides, we consider the two, aforementioned

formulations for models admitting both *discrete* and *continuous* weights. Remarkably, any of the aforementioned models can be derived by invoking the *Minimum Discrimination Information Principle*, prescribing to optimize either the discrete or the continuous version of the Kullback-Leibler divergence in a constrained fashion.

For what concerns discrete models, we find the physics-inspired ones to outperform traditional, econometric models, such as the Poisson and the negative binomial ones as well as their zero-inflated counterparts [43], in reproducing both network statistics and their variability, as confirmed by model selection criteria. For what concerns continuous models, we find each of them to have pros and cons: for instance, 1) models driven by the exponential and the gamma distributions lead to better reproduce higher-order network statistics such as the average nearest neighbors strength and the clustering coefficient; 2) models driven by the gamma and the lognormal distributions perform best in reproducing data variability; 3) the model driven by the exponential distribution leads to the most resilient configurations against shocks concerning trade volumes.

Conditional models are characterized by a log-likelihood that can be separated into a purely structural part and a conditional, weighted one. As a consequence, the optimization problem can be split into two subproblems, each one dealing with a smaller number of parameters: thus, from a purely computational perspective, solving conditional models is more 'convenient' than solving the integrated ones; still, individuating the best optimization procedure remains a debated issue.

In econometrics, conditional, or *hurdle*, models are optimized by tuning the parameters defining the distribution of weights on the unique, empirical realization of the binary adjacency matrix [70] although the weights are assigned to *any* realization output by the binary model of choice. To avoid inconsistencies, a recipe has been introduced in [63], where a generalized log-likelihood, obtained upon averaging over the topological randomness, is maximized. Chapter 4 is devoted to compare the two, aforementioned recipes with a third one, prescribing to estimate each parameter on the entire ensemble of configurations output by the

binary model of choice and, then, average the values constituting this set. The second and the third procedures differ in the way the average over the topological randomnes is carried out: according to the jargon employed in the physics of disordered systems, we name them *annealed* and *quenched*, respectively.

While the deterministic, annealed and quenched estimates are equivalent for models with homogeneous weights, annealed and quenched estimates are significantly different from the deterministic one for models constraining the strength sequences as well as the models introduced in [2]. A test on a snapshot of a sparse network, i.e. the Bitcoin Lightning Network, also reveals that the quenched estimate can be biased when the parameters are node-specific: this is the case of an empirically connected node that is found to be disconnected in certain model-induced instances, hence the corresponding parameter cannot be evaluated. This leads us to conclude that overcoming the limitations of the deterministic estimate is indeed possible although the quenched approach should be applied with caution (in fact, the sparser the network, the higher the potential bias).

In chapters 2, 3 and 4 we focus on the aggregated representation of the WTW: as this configuration is highly reciprocal, the system can be regarded as an undirected network. When, instead, the corresponding multiplex representation is considered (with different products being traded on different layers), the aforementioned, highly reciprocal structure is typically lost [14] and link directionality starts playing a major role. Chapter 5 is devoted to study the reciprocity structure, as well as its impact on the emergence of triadic motifs, of the 2018 domestic Dutch Production Multiplex, constructed from the data collected by CBS-Statistics Netherlands [71]. Employing conditional models encoding information 1) on link and weight directionality and 2) on link and weight reciprocity and using the annealed approach to estimate parameters allows us to conclude that 1) the aggregated version of the Dutch Production Network is not informative of the structure of the Dutch Production Multiplex (more specifically, of the abundance of motifs that are present

in the single layers); 2) for more than half of the layers, the abundance of binary and weighted motifs is well-described by a conditional model encoding information on link and weight reciprocity.

The aforementioned work represents a quite unique example in the literature, the only other study addressing the emergence of motifs - although for just a single, representative commodity - being [72], dealing with Japanese data and concluding that triads occuring more frequently than expected (i.e. the 'open' ones) are a byproduct of the intrinsic complementarity of the input/output relationships established by firms [73]; our study, on the contrary, proposes a simplified explanation for motif emergence, primarily involving constraints on lower-order quantities. However, situations may arise where these quantities alone are insufficient to fully comprehend the functional structures of individual commodities. In such instances, a meticulous case-by-case approach emerges as indispensable.

Chapter 2

Discrete entropy-based econometric models

This chapter illustrates the results of the analysis published in [1]. Upon implementing the Minimum Discrimination Information Principle in its simplest form, i.e. maximizing Shannon entropy in a constrained fashion, we derive physics-inspired models capable of accommodating both structural and econometric parameters. Then, we compare their performance with that of state-of-the-art econometric recipes in reproducing the structural properties of both the BACI dataset [74, 75] and the Gleditsch dataset [38], finding that physics-inspired models outperform econometric models.

2.1 Modelling positive weights

From a merely econometric point of view, the simplest exercise is that of reproducing the realized (positive) trade volumes (link weights) of the WTW. To this aim, we have considered two different datasets. The first one is curated by Gleditsch [38] and includes yearly trade volumes, yearly GDP values (both reported in millions of US dollars) and the (time-independent) matrix of geographic distances between capital cities of all countries in the data. The second one is the BACI dataset, a detailed description of which can be found in [74, 75]. For both datasets

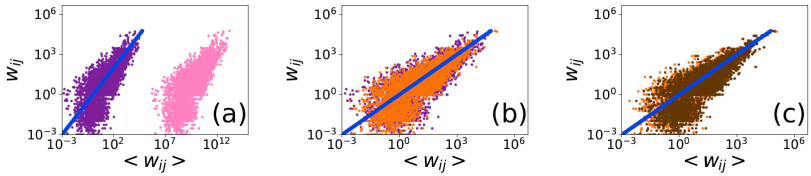

Figure 1: WTW weights w_{ij} versus the values predicted by different specifications of the GM: (1) GM with $\rho = \beta = 1$ and $\gamma = 0$, depicted in •; (2) GM with $\beta = 1$, $\gamma = 0$ and ρ tuned by requiring that the total weight is reproduced, i.e. $W = \sum_{i<j} w_{ij} = \sum_{i<j} \langle w_{ij} \rangle_{\rm GM} = \langle W \rangle_{\rm GM}$, depicted in •; (3) GM with $\beta = 1$, $\gamma = -1$ and ρ tuned as above, depicted in •; (4) full GM where all parameters are tuned as described in Appendix A.1, depicted in •. While moving from a parameter-free model to a single-parameter model largely improves the goodness-of-fit, see (a), the net effect of rising the number of explanatory variables is that of decreasing the dispersion around the identity, see (b). Further rising the number of parameters, instead, does not add much to the picture provided by the third specification of the GM, see (c). Results refer to the year 2000 of the dataset curated by Gleditsch [38].

we have selected and analyzed eleven years: 1990-2000 for the Gleditsch dataset and 2007-2017 for the BACI one. The year 2000 of the Gleditsch dataset - i.e. the one with the largest number of countries (176) - is the snapshot we have selected to graphically illustrate the results of our analyses. We have always considered the undirected (symmetrized) version of the weighted trade matrix, whose generic entry reads $w_{ij} \equiv (\exp_{ij} + \exp_{ji})/2$, i.e. w_{ij} is the bilateral trade volume defined as the arithmetic mean of the export volume from country i to country j and of the export volume from country j to the country i.

Our econometric exercise is carried out by comparing the empirical, positive weights of the WTW, in the year 2000, with four, different specifications of the following econometric function

$$\langle w_{ij} \rangle_{\rm GM} = f(\omega_i, \omega_j, d_{ij} | \underline{\theta}) = \rho (\omega_i \omega_j)^{\beta} d_{ij}^{\gamma} \tag{2.1}$$

where $\omega_i = {\rm GDP}_i / \overline{\rm GDP}$ is the GDP of country i divided by the arithmetic mean of GDPs, d_{ij} is the geographic distance between the capitals of countries i and j and $\underline{\theta} \equiv (\rho, \beta, \gamma)$. The first specification of the model

above is characterized by the assignment $\rho = \beta = 1$ and $\gamma = 0$ and its performance in reproducing the positive weights of the WTW is shown in Fig. 1(a). The overall poor performance of this version of the GM signals the presence of both a *scaling problem* - estimates are positively correlated with observations, only shifted to the right - and of a *dimensional problem* - in fact, pure numbers appear on the right-hand side while the WTW weights are measured in (multiples of) dollars.

A second specification of the GM solves both problems at once. This specification is characterized by the assignment $\beta = 1$ and $\gamma = 0$ while ρ is treated as a free parameter, tuned by requiring that the total weight is reproduced, i.e. that $W = \sum_{i<j} w_{ij} = \sum_{i<j} \langle w_{ij} \rangle_{\mathrm{GM}} = \langle W \rangle_{\mathrm{GM}}$: the fit is, now, much more accurate, as shown in Fig. 1(b). The model can be further enriched by adding dyadic factors such as the geographic distances between capitals: some more accuracy in the description of the empirical data is indeed gained, as the reduced dispersion of the cloud of points around the identity witnesses. Quite remarkably, the picture does not change much if, now, we let the entire set of parameters to be tuned as described in Appendix A.1 (see Fig. 1(c)).

Although the GM leads to a good prediction of the positive weights, its intrinsic limitation is that of not allowing the topological structure of the WTW to be correctly recovered. In fact, it outputs only positive weights, hence inducing a trivial, fully-connected structure: upon defining[1] $a_{ij} = \Theta[\langle w_{ij} \rangle_{\mathrm{GM}}]$, $\forall\, i < j$, it is evident that $a_{ij} = 1$, $\forall\, i < j$. In order to overcome such a limitation, one can 'refine' the plain gravity model by 'dressing' it with a probability distribution capable of accounting for the null entries as well.

2.2 Statistical network models

In very general terms, we need to define a *statistical network model*, i.e. a set of mathematical relationships between the random variables that are of interest for our network description. Two broad classes of such mod-

[1] The position $a_{ij} = \Theta[w_{ij}]$ means that $a_{ij} = 1$ whenever $w_{ij} > 0$ and $a_{ij} = 0$ if and only if $w_{ij} = 0$.

els can be identified, i.e. the econometric ones and the ones rooted into statistical physics. In what follows, we will deal with (either econometric or physics-rooted) discrete statistical models.

2.2.1 Econometric models

Let us start with the description of some of the most representative members of the econometric class of models.

Poisson model

The simplest model in this class prescribes to consider $\langle w_{ij} \rangle_{\text{GM}} = \rho(\omega_i \omega_j)^{\beta} d_{ij}^{\gamma}$ as the expected value of the Poisson probability mass function

$$q_{ij}^{\text{Pois}}(w_{ij}) = \frac{z_{ij}^{w_{ij}} e^{-z_{ij}}}{w_{ij}!}; \tag{2.2}$$

since $\langle w_{ij} \rangle_{\text{Pois}} = z_{ij}$, the explanatory power of the GM is retained upon requiring that $\langle w_{ij} \rangle_{\text{Pois}} = \langle w_{ij} \rangle_{\text{GM}}$, i.e. by posing

$$z_{ij} \equiv \rho(\omega_i \omega_j)^{\beta} d_{ij}^{\gamma}. \tag{2.3}$$

The expected topology of the network is determined by the entries of the adjacency matrix implied by the model, is captured by the expression $\langle a_{ij} \rangle_{\text{Pois}} = p_{ij}^{\text{Pois}} = 1 - q_{ij}^{\text{Pois}}(0) = 1 - e^{-z_{ij}}$ (see Appendix A.2 for a detailed description of the procedure to estimate the parameters of the Poisson model).

Negative binomial model

The main drawback of the Poisson model is that of predicting a variance of the weights that is necessarily equal to their average value. In general, this may be different from what empirical analyses suggest. In order to overcome this problem, econometricians have considered a different probability mass function, namely the negative binomial one with the introduction of the overdispersion[2] parameter $\alpha = m^{-1}$ [76]:

[2]In the Poisson case, one has that $\sigma^2_{\text{Pois}}[w_{ij}] = z_{ij} = \langle w_{ij} \rangle_{\text{Pois}}$; hence, variance cannot be adjusted independently from the mean. In the negative binomial case, instead, $\sigma^2_{\text{NB}}[w_{ij}] =$

$$q_{ij}^{\text{NB}}(w_{ij}) = \binom{m + w_{ij} - 1}{w_{ij}} \left(\frac{1}{1+\alpha z_{ij}}\right)^m \left(\frac{\alpha z_{ij}}{1+\alpha z_{ij}}\right)^{w_{ij}}. \tag{2.4}$$

One finds that $\langle w_{ij}\rangle_{\text{NB}} = m\alpha z_{ij} = z_{ij}$. The requirement that the expected value of the negative binomial distribution coincides with the prediction coming from the GM, i.e. $\langle w_{ij}\rangle_{\text{NB}} = \langle w_{ij}\rangle_{\text{GM}}$, can be, again, realized by posing $z_{ij} \equiv \rho(\omega_i\omega_j)^\beta d_{ij}^\gamma$. Predictions about the topology are, now, carried out via the expression $\langle a_{ij}\rangle_{\text{NB}} = p_{ij}^{\text{NB}} = 1 - q_{ij}^{\text{NB}}(0) = 1 - \left(\frac{1}{1+\alpha z_{ij}}\right)^m$ (see Appendix A.2 for a detailed description of the procedure to estimate the parameters of the negative binomial model).

Zero-inflated Poisson model (ZIP model)

The main drawback of the econometric models above is that of failing in reproducing the link density of the WTW. For instance, the latter equals $c = \frac{2L}{N(N-1)} \simeq 0.63$ in year 2000: it turns out that, while the Poisson model overestimates this quantity, the negative binomial one underestimates it, i.e.

$$\langle c\rangle_{\text{NB}} < c < \langle c\rangle_{\text{Pois}} \tag{2.5}$$

where $\langle c\rangle_{\text{Pois}} = \sum_{i<j} p_{ij}^{\text{Pois}} \simeq 0.68$ and $\langle c\rangle_{\text{NB}} \sum_{i<j} p_{ij}^{\text{NB}} \simeq 0.60$. For this reason, econometricians have defined ZI models, i.e. two-step recipes whose general form reads

$$\begin{aligned} Q(\mathbf{W}) = \prod_{i<j} q_{ij}(w_{ij}) &= \prod_{i<j} p_{ij}^{a_{ij}}(1-p_{ij})^{1-a_{ij}} \cdot q_{ij}(w_{ij}|a_{ij}) \\ &= \prod_{i<j} p_{ij}^{a_{ij}}(1-p_{ij})^{1-a_{ij}} \prod_{i<j} q_{ij}(w_{ij}|a_{ij}) = P(\mathbf{A})Q(\mathbf{W}|\mathbf{A}) \end{aligned} \tag{2.6}$$

a relationship indicating that the probability of the (network represented by the) weighted adjacency matrix $\mathbf{W}$ can be obtained as the product of the probability $P(\mathbf{A})$ of observing the purely binary adjacency matrix $\mathbf{A}$ and the conditional probability $Q(\mathbf{W}|\mathbf{A})$ - where, for consistency,

$z_{ij}(1+\alpha z_{ij}) = z_{ij}\left(1+\frac{z_{ij}}{m}\right) = \langle w_{ij}\rangle_{\text{NB}}(1+\alpha\langle w_{ij}\rangle_{\text{NB}}) = \langle w_{ij}\rangle_{\text{NB}}\left(1+\frac{\langle w_{ij}\rangle_{\text{NB}}}{m}\right)$ and the variance can be increased to overcome the problem of overestimating the link density.

$\mathbf{A} = \Theta[\mathbf{W}]$, i.e. $a_{ij} = \Theta[w_{ij}], \forall\, i < j$.

The simplest ZI model is the Poisson one, defined by the positions

$$p_{ij}^{\text{ZIP}} = \frac{G_{ij}}{1+G_{ij}}(1-e^{-z_{ij}}), \tag{2.7}$$

$$q_{ij}^{\text{ZIP}}(w_{ij}|a_{ij}=1) = \begin{cases} \frac{z_{ij}^{w_{ij}} e^{-z_{ij}}}{(1-e^{-z_{ij}})w_{ij}!}, & w_{ij} > 0 \\ 0, & w_{ij} \leq 0 \end{cases}. \tag{2.8}$$

Notice that $1 - p_{ij}^{\text{ZIP}} = \frac{1}{1+G_{ij}} + \frac{G_{ij}}{1+G_{ij}} e^{-z_{ij}}$, i.e. the connection between nodes i and j can be missing either because a link is not there (with probability $\frac{1}{1+G_{ij}}$) or because a link is there but has zero weight (with probability $\frac{G_{ij}}{1+G_{ij}} e^{-z_{ij}}$). Consistently, i and j are connected because the weight is not zero (with probability $1 - e^{-z_{ij}}$).

In order to 'dress' the GM, we need to identify some of the parameters of the Poisson model with the usual econometric function. Since

$$\langle w_{ij} \rangle_{\text{ZIP}} = p_{ij}^{\text{ZIP}} \langle w_{ij}|a_{ij}\rangle_{\text{ZIP}} = p_{ij}^{\text{ZIP}} \frac{z_{ij}}{1-e^{-z_{ij}}} = \frac{G_{ij}}{1+G_{ij}} z_{ij}, \tag{2.9}$$

we can make the identification $z_{ij} \equiv \rho(\omega_i \omega_j)^\beta d_{ij}^\gamma$. Upon doing so, we are treating z_{ij} as an 'effective' conditional weight: in fact, Eq. (2.9) can be understood as describing an aleatory experiment that combines a logit with a full Poisson step. According to this interpretation, z_{ij} would represent a Poisson-like expected weight, conditional to the success of the logit step, i.e. $z_{ij} = \frac{\langle w_{ij} \rangle_{\text{ZIP}}}{p_{ij}^{\text{logit}}}$, with $p_{ij}^{\text{logit}} = \frac{G_{ij}}{1+G_{ij}}$.

A second econometric identification is, however, needed: we will proceed by imposing

$$G_{ij} = \delta \omega_i \omega_j \tag{2.10}$$

(see Appendix A.2 for a detailed description of the procedure to estimate the parameters of the ZIP model).

Zero-inflated negative binomial model (ZINB model)

The ZI version of the negative binomial model, instead, is defined by

$$p_{ij}^{\text{ZINB}} = \frac{G_{ij}}{1+G_{ij}}(1-\tau_{ij}), \tag{2.11}$$

$$q_{ij}^{\text{ZINB}}(w_{ij}|a_{ij}=1) = \binom{m+w_{ij}-1}{w_{ij}} \left(\frac{1}{1-\tau_{ij}}\right) \left(\frac{1}{1+\alpha z_{ij}}\right)^{m} \left(\frac{\alpha z_{ij}}{1+\alpha z_{ij}}\right)^{w_{ij}} \tag{2.12}$$

where the conditional distribution $q_{ij}^{\text{ZINB}}(w_{ij}|a_{ij}=1)$ is so defined for $w_{ij} \geq 0$ and is 0 otherwise; as the for the plain negative binomial model, $\alpha = m^{-1}$ and $\tau_{ij} = \left(\frac{1}{1+\alpha z_{ij}}\right)^{m}$. Moreover, as for the ZIP model, $1 - p_{ij}^{\text{ZINB}} = \frac{1}{1+G_{ij}} + \frac{G_{ij}}{1+G_{ij}}\tau_{ij}$, i.e. the connection between nodes i and j can be missing either because a link is not there (with probability $\frac{1}{1+G_{ij}}$) or because a link is there but has zero weight (with probability $\frac{G_{ij}}{1+G_{ij}}\tau_{ij}$); consistently, i and j are connected because the weight is not zero (with probability $1-\tau_{ij}$).

In order to 'dress' the GM, we need to identify some of the parameters of the negative binomial model with the usual econometric function. Upon considering that

$$\langle w_{ij}\rangle_{\text{ZINB}} = p_{ij}^{\text{ZINB}}\langle w_{ij}|a_{ij}\rangle_{\text{ZINB}} = p_{ij}^{\text{ZINB}}\frac{z_{ij}}{1-\tau_{ij}} = \frac{G_{ij}}{1+G_{ij}}z_{ij}, \tag{2.13}$$

we can make the identification $z_{ij} \equiv \rho(\omega_i\omega_j)^{\beta}d_{ij}^{\gamma}$ and $G_{ij} \equiv \delta\omega_i\omega_j$. As for the ZIP case, we are treating z_{ij} as a negative binomial-like expected weight, conditional to the success of a logit step, i.e. $z_{ij} = \frac{\langle w_{ij}\rangle_{\text{ZINB}}}{p_{ij}^{\text{logit}}}$, with $p_{ij}^{\text{logit}} = \frac{G_{ij}}{1+G_{ij}}$ (see Appendix A.2 for a detailed description of the procedure to estimate the parameters of the ZINB model).

Let us notice that, while the ZIP model provides a better estimation of the link density than the Poisson model, the ZINB and the negative binomial ones basically perform in the same way. In fact,

$$\langle c\rangle_{\text{ZINB}} < c \simeq \langle c\rangle_{\text{ZIP}} \tag{2.14}$$

since $c = \frac{2L}{N(N-1)} \simeq 0.63$, $\langle c\rangle_{\text{ZIP}} = \sum_{i<j} p_{ij}^{\text{ZIP}} \simeq 0.63$ and $\langle c\rangle_{\text{ZINB}} = \sum_{i<j} p_{ij}^{\text{ZINB}} \simeq 0.60$, a result suggesting that both variants of the negative binomial model will perform poorly in reproducing the binary properties of the WTW.

2.2.2 Maximum-entropy models

The members of the second class of network models are the ones defined within the framework of traditional statistical mechanics. All of them can be derived by performing a constrained maximization of Shannon entropy [48] where the constraints represent the available information about the system at hand.

The simplest, yet non trivial, maximum-entropy (ME) model that can be considered comes from the maximization of the binary Shannon functional

$$S = -\sum_{\mathbf{A}} P(\mathbf{A}) \ln P(\mathbf{A}) \tag{2.15}$$

constrained to reproduce the entire degree sequence, $\{k_i(\mathbf{A})\}_{i=1}^{N}$, of the network. This model is known under the name of Undirected Binary Configuration Model (UBCM) and has been shown to accurately reproduce many (binary) properties of a wide spectrum of real-world systems [46]. The UBCM is described by the probability mass function

$$P(\mathbf{A}) = \prod_{i<j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1-a_{ij}} \tag{2.16}$$

which is factorized into the product of Bernoulli probability mass functions (one for each pair of nodes) with

$$p_{ij}^{\text{UBCM}} = \frac{x_i x_j}{1 + x_i x_j} \tag{2.17}$$

(where x_i is the Lagrange multiplier controlling for the degree of node i). Importantly, the logit model admitting the presence of a single, global constant can be derived from entropy maximization upon re-parametrizing the Lagrange multipliers of the UBCM [77]: the identification $x_i \equiv \sqrt{\delta}\omega_i$, in fact, leads to

$$p_{ij}^{\text{logit}} = \frac{G_{ij}}{1 + G_{ij}} = \frac{\delta\omega_i\omega_j}{1 + \delta\omega_i\omega_j}; \tag{2.18}$$

although the functional form above is not the most general one (for instance, dyadic factors such as geographic distances could be added as

well), it is the form we will adopt in what follows. In the network literature, the above formulation of the logit model has been named density-corrected Gravity Model (dcGM) [78] and popularized [19] as a particular case of the so-called fitness model [64]; interestingly, it has been proven to perform remarkably well for the task of reconstructing the topology of networks from partial information [77].

Conditional models

Since we are interested in reproducing the structural properties of weighted networks, we need to complement the purely binary step above with a recipe for reconstructing weights. The entropy-based framework handles such a requirement via the maximization of conditional Shannon functionals allowing the specification of $P(\mathbf{A})$ to be disentangled from that of $Q(\mathbf{W}|\mathbf{A})$ [63].

When discrete weighted models are considered, a useful quantity is the conditional Shannon entropy

$$S(\mathbf{W}|\mathbf{A}) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \sum_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) \log Q(\mathbf{W}|\mathbf{A}) \tag{2.19}$$

where the first sum runs over all binary configurations within the ensemble $\mathbb{A}$ and the second sum runs over all weighted configurations that are compatible with a specific binary structure, represented by the adjacency matrix $\mathbf{A}$, i.e. such that $\mathbb{W}_{\mathbf{A}} = \{\mathbf{W} : \Theta[\mathbf{W}] = \mathbf{A}\}$.

Conditional maximization proceeds by specifying a set of weighted constraints that, in the discrete case, read

$$1 = \sum_{\mathbb{W}_{\mathbf{A}}} P(\mathbf{W}|\mathbf{A}),\ \forall\, \mathbf{A} \in \mathbb{A} \tag{2.20}$$

$$\langle C_\alpha \rangle = \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \sum_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) C_\alpha(\mathbf{W}),\ \forall\, \alpha \tag{2.21}$$

the first condition ensuring the normalization of the conditional probability mass function and the vector $\{C_\alpha(\mathbf{W})\}$ representing the 'proper' set of weighted constraints. Differentiating the corresponding Lagrangean

functional with respect to $Q(\mathbf{W}|\mathbf{A})$ and equating the result to zero leads to

$$Q(\mathbf{W}|\mathbf{A}) = \begin{cases} \frac{e^{-H(\mathbf{W})}}{Z_{\mathbf{A}}}, & \mathbf{W} \in \mathbb{W}_{\mathbf{A}} \\ 0, & \mathbf{W} \notin \mathbb{W}_{\mathbf{A}} \end{cases} \tag{2.22}$$

where $H(\mathbf{W}) = \sum_{\alpha} \psi_{\alpha} C_{\alpha}$ is the so-called Hamiltonian, listing the constrained, weighted quantities and $Z_{\mathbf{A}} = \sum_{\mathbb{W}_{\mathbf{A}}} e^{-H(\mathbf{W})}$ is the partition function for fixed $\mathbf{A}$.

The explicit functional form of $Q(\mathbf{W}|\mathbf{A})$ can be obtained only once the functional form of the constraints has been specified as well. To this aim, let us consider the Hamiltonian

$$H(\mathbf{W}) = \sum_{i<j} \phi_{ij} w_{ij}, \tag{2.23}$$

where weights are modelled as non-negative integer variables, i.e. $w_{ij} \in \mathbb{N}, \forall\, i < j$. This choice induces a conditional probability mass function reading

$$Q(\mathbf{W}|\mathbf{A}) = \prod_{i<j} q_{ij}(w_{ij}|a_{ij}) = \prod_{i<j} y_{ij}^{w_{ij}-a_{ij}} (1-y_{ij})^{a_{ij}} \tag{2.24}$$

(with $e^{-\psi_{ij}} = e^{-\phi_{ij}} = y_{ij}$). Let us, now, turn the model above into a proper econometric one. To this aim, let us proceed by analogy. All zero-inflated recipes identify z_{ij} with a conditional expected weight, a prescription that, in our case, would translate into its identification with $\langle w_{ij}|a_{ij}\rangle = \frac{1}{1-y_{ij}}$. This choice, however, would lead to an inconsistency, since $z_{ij} > 0$ is a positive, real number while $\langle w_{ij}|a_{ij}\rangle$ must necessarily exceed 1 as it represents the expected weight conditional to the existence of a connection. An alternative, consistent identification reads

$$\frac{1}{1-y_{ij}} = 1 + z_{ij} \tag{2.25}$$

which, in turn, induces the conditional probability mass function

$$Q(\mathbf{W}|\mathbf{A}) = \prod_{i<j} q_{ij}(w_{ij}|a_{ij}) = \prod_{i<j} \left(\frac{z_{ij}}{1+z_{ij}}\right)^{w_{ij}-a_{ij}} \left(\frac{1}{1+z_{ij}}\right)^{a_{ij}} \tag{2.26}$$

Models of the kind are also known as hurdle models: quite remarkably, entropy maximization allows us to recover them in a fully principled way, i.e. by eliminating the (otherwise unavoidable) ambiguity that accompanies the choice of the distribution (supposedly) describing the positive values of an economic system.

The hurdle-geometric (HG) model derived above, however, suffers from a number of limitations, the most relevant of which is that of failing in reproducing basic network quantities such as the WTW total weight. As an illustrative example, while $W = \sum_{i<j} w_{ij} \simeq 10^9$, in the year 2000, we find that $\langle W \rangle_{\text{HG}} \simeq 10^5$. In order to overcome such a limitation, we have considered the conditional probability mass function induced by the Hamiltonian

$$H(\mathbf{W}) = \sum_{i<j} (\phi_0 + \phi_{ij}) w_{ij} = \phi_0 W + \sum_{i<j} \phi_{ij} w_{ij}. \tag{2.27}$$

Upon identifying $e^{-\phi_0} = y_0$ and $e^{-\phi_{ij}} = y_{ij} = \frac{z_{ij}}{1+z_{ij}}$ (see Eq. (2.25)), we obtain the modified, econometric model

$$q_{ij}(w_{ij}|a_{ij}) = \left(\frac{y_0 z_{ij}}{1 + z_{ij}} \right)^{w_{ij} - a_{ij}} \left(\frac{1 + z_{ij} - y_0 z_{ij}}{1 + z_{ij}} \right)^{a_{ij}}. \tag{2.28}$$

We are now ready to fully specify the suite of discrete entropy-models that we will compare with the aforementioned, purely econometric ones: to this aim, we need to fully specify the functional form

$$\begin{aligned} Q(\mathbf{W}) = \prod_{i<j} q_{ij}(w_{ij}) &= \prod_{i<j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1 - a_{ij}} \cdot q_{ij}(w_{ij}|a_{ij}) \\ &= \prod_{i<j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1 - a_{ij}} \prod_{i<j} q_{ij}(w_{ij}|a_{ij}) = P(\mathbf{A}) Q(\mathbf{W}|\mathbf{A}); \end{aligned} \tag{2.29}$$

the two, most obvious choices are represented by the models

$$Q_{\text{TSF}}(\mathbf{W}) = P_{\text{logit}}(\mathbf{A}) Q(\mathbf{W}|\mathbf{A}) = \prod_{i<j} (p_{ij}^{\text{logit}})^{a_{ij}} (1 - p_{ij}^{\text{logit}})^{1 - a_{ij}} \cdot q_{ij}(w_{ij}|a_{ij}) \tag{2.30}$$

and

$$Q_{\text{TS}}(\mathbf{W}) = P_{\text{UBCM}}(\mathbf{A})Q(\mathbf{W}|\mathbf{A}) = \prod_{i<j}(p_{ij}^{\text{UBCM}})^{a_{ij}}(1-p_{ij}^{\text{UBCM}})^{1-a_{ij}} \cdot q_{ij}(w_{ij}|a_{ij}) \tag{2.31}$$

combining the weighted, conditional step induced by the Hamiltonian defined in Eq. (2.27), respectively with the purely binary logit model and the Undirected Binary Configuration Model - the acronyms stand for 'two-step fitness' model and 'two-step' model and recall the names originally used to define them [20, 66].

Integrated models

Less trivial choices are represented by models whose both binary and weighted portions are jointly determined by the constraints. They can all be recovered as specifications of the generic Hamiltonian

$$H(\mathbf{W}) = \sum_{i<j}[\theta_{ij}a_{ij} + (\phi_0 + \phi_{ij})w_{ij}] = \sum_{i<j}\theta_{ij}a_{ij} + \phi_0 W + \sum_{i<j}\phi_{ij}w_{ij}; \tag{2.32}$$

in what follows, we will consider two different instances of such a function, defined by the choices $\theta_{ij} = \theta_0$ and $\theta_{ij} = \theta_i + \theta_j$. In other words, while we let the weighted parts of these models coincide and read as in Eq. (2.28), we allow for the binary part to vary, either constraining the total number of links, L, or the entire degree sequence, $\{k_i(\mathbf{A})\}_{i=1}^N$. In the first case, our Hamiltonian reads

$$H_{(1)}(\mathbf{W}) = \theta_0 L + \phi_0 W + \sum_{i<j}\phi_{ij}w_{ij} \tag{2.33}$$

and instances the model in Eq. (2.29) with

$$p_{ij}^{(1)} = \frac{xy_0z_{ij}}{1 + z_{ij} - y_0z_{ij} + xy_0z_{ij}} \tag{2.34}$$

$$q_{ij}^{(1)}(w_{ij}|a_{ij}) = \left(\frac{y_0z_{ij}}{1+z_{ij}}\right)^{w_{ij}-a_{ij}}\left(\frac{1+z_{ij}-y_0z_{ij}}{1+z_{ij}}\right)^{a_{ij}} \tag{2.35}$$

(having posed $e^{-\theta_0} = x$, $e^{-\phi_0} = y_0$ and $e^{-\phi_{ij}} = y_{ij} = \frac{z_{ij}}{1+z_{ij}}$); in the second case, it reads

$$H_{(2)}(\mathbf{W}) = \sum_i \theta_i k_i + \psi_0 W + \sum_{i<j} \psi_{ij} w_{ij} \tag{2.36}$$

and instances the model in Eq. (2.29) with

$$p_{ij}^{(2)} = \frac{x_i x_j y_0 z_{ij}}{1 + z_{ij} - y_0 z_{ij} + x_i x_j y_0 z_{ij}} \tag{2.37}$$

$$q_{ij}^{(2)}(w_{ij}|a_{ij}) = \left(\frac{y_0 z_{ij}}{1+z_{ij}}\right)^{w_{ij}-a_{ij}} \left(\frac{1+z_{ij}-y_0 z_{ij}}{1+z_{ij}}\right)^{a_{ij}} \tag{2.38}$$

(having posed $e^{-\theta_i} = x_i$, $e^{-\phi_0} = y_0$ and $e^{-\phi_{ij}} = y_{ij} = \frac{z_{ij}}{1+z_{ij}}$). Appendix A.3 provides a detailed description of the procedure we have adopted to estimate the parameters entering into the definition of our basket of discrete ME models.

So far, we have turned entropy-based models into econometric ones via a suitable, econometric transformation of the Lagrange multipliers defining the proper 'physical' models. The entropy-based formalism, however, also offers the opportunity to define statistical models in a fully data-driven fashion. To this aim, let us consider the Hamiltonian

$$H_{\text{UECM}}(\mathbf{W}) = \sum_i [\theta_i k_i + \phi_i s_i] = \sum_{i<j} [(\theta_i + \theta_j) a_{ij} + (\phi_i + \phi_j) w_{ij}] \tag{2.39}$$

that constrains both degrees and strengths. The model induced by the latter ones is called Undirected Enhanced Configuration Model (UECM) and represents the best-performing one for the task of network reconstruction in presence of full information about the constraints [56, 54].

Remarkably, all models considered in this section can be compactly derived by combining a logit-like probability mass function describing the binary network structure with the conditional expression defined in Eq. (2.28). To prove this, it is enough to notice that all the Bernoulli-like

probability mass functions characterizing our models can be compactly re-written as

$$p_{ij} = \frac{x'_i x'_j}{1 + x'_i x'_j} \tag{2.40}$$

where

$$(x'_i x'_j)^{\text{logit}} = \delta \omega_i \omega_j, \tag{2.41}$$

$$(x'_i x'_j)^{\text{UBCM}} = x_i x_j, \tag{2.42}$$

$$(x'_i x'_j)^{(1)} = \frac{x y_0 z_{ij}}{1 + z_{ij} - y_0 z_{ij}}, \tag{2.43}$$

$$(x'_i x'_j)^{(2)} = \frac{x_i x_j y_0 z_{ij}}{1 + z_{ij} - y_0 z_{ij}}, \tag{2.44}$$

$$(x'_i x'_j)^{\text{UECM}} = \frac{x_i x_j y_i y_j}{1 - y_i y_j}. \tag{2.45}$$

2.3 Results

Let us, now, test and compare the performance of our two, broad classes of models in reproducing the topological properties of the WTW. To this aim, let us consider both the local properties, such as the degrees and the strengths, and the higher-order ones such as the average nearest neighbors degree (ANND)

$$k_i^{nn} = \frac{\sum_{j(\neq i)=1}^{N} a_{ij} k_j}{k_i}, \quad \forall\, i \tag{2.46}$$

(which gives information about the degree correlations), the clustering coefficient (BCC)

$$c_i = \frac{\sum_{j(\neq i)=1}^{N} \sum_{k(\neq i,j)=1}^{N} a_{ij} a_{jk} a_{ki}}{k_i (k_i - 1)}, \quad \forall\, i \tag{2.47}$$

(which counts the percentage of node i's partners that are also partners themselves). We will also consider their weighted counterparts, i.e. the average nearest neighbors strength (ANNS)

$$s_i^{nn} = \frac{\sum_{j(\neq i)=1}^{N} a_{ij} s_j}{k_i}, \quad \forall\, i \tag{2.48}$$

(which gives information about the strength correlations) and the weighted clustering coefficient (WCC)

$$c_i^w = \frac{\sum_{j(\neq i)=1}^{N} \sum_{k(\neq i,j)=1}^{N} w_{ij} w_{jk} w_{ki}}{k_i(k_i - 1)}, \quad \forall\, i \tag{2.49}$$

(that weighs the closed triangular patterns that node i establishes with other trade partners). We will also test the accuracy of the reconstruction provided by the methods in our basket by considering properties like the true positive rate (TPR)

$$\langle \text{TPR} \rangle = \frac{\langle \text{TP} \rangle}{L} = \frac{\sum_{i<j} a_{ij} p_{ij}}{L} \tag{2.50}$$

(i.e. the percentage of links correctly recovered by a given reconstruction method), the specificity (SPC)

$$\langle \text{SPC} \rangle = \frac{\langle \text{TN} \rangle}{\binom{N}{2} - L} = \frac{\sum_{i<j} (1 - a_{ij})(1 - p_{ij})}{\binom{N}{2} - L} \tag{2.51}$$

(i.e. the percentage of zeros correctly recovered by a given reconstruction method), the positive predictive value (PPV)

$$\langle \text{PPV} \rangle = \frac{\langle \text{TP} \rangle}{\langle L \rangle} = \frac{\sum_{i<j} a_{ij} p_{ij}}{\langle L \rangle} \tag{2.52}$$

(i.e. the percentage of links correctly recovered by a given reconstruction method with respect to the total number of predicted links) and the accuracy (ACC)

$$\langle \text{ACC} \rangle = \frac{\langle \text{TP} \rangle + \langle \text{TN} \rangle}{\binom{N}{2}} \tag{2.53}$$

(measuring the overall performance of a given reconstruction method in correctly placing both links and zeros).

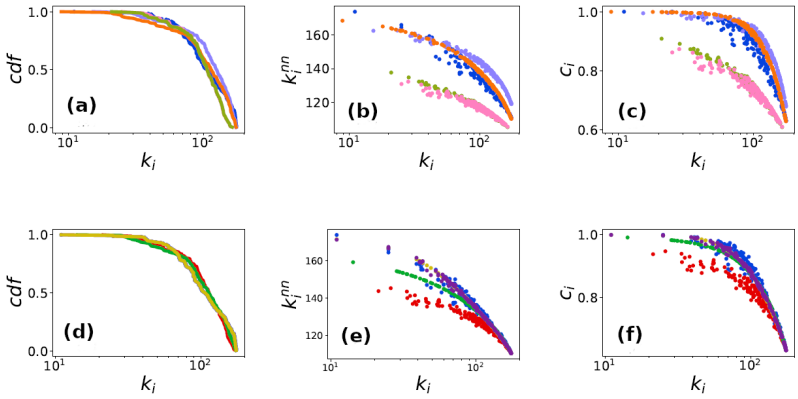

Figure 2: Performance of econometric models (top row) versus the performance of ME models (bottom row) in reproducing: (a), (d) the degrees; (b), (e) the ANND; (c), (f) the BCC. Empirical points are indicated as ●; econometric and ME models are indicated as follows: ● - Poisson; ● - Negative Binomial; ● - ZIP; ● - ZINB; ● - $H_{(1)}$; ● - $H_{(2)}$; ● - TS; ● - TSF. Results refer to the year 2000 of the dataset curated by Gleditsch [38].

Fig. 2 sums up the comparisons carried out between the econometric models and the ME ones. The comparison between the empirical cumulative density function (CDF) of the degrees and the ones output by the econometric models reveals the latter ones to be able to predict an overall similar functional form (see Fig. 2(a) and Fig. 2(d)); still, the prediction obtained by any of the ME models is much closer to the empirical trend. More quantitatively, one can implement the Kolmogorov-Smirnov (KS) test to check the goodness of any of the models considered in the present work to replicate the empirical degrees: while any of the ME models provides estimates of the degrees that are compatible with the empirical CDF (at the significance level of 5%), only the ZIP model predicts degrees that are compatible with the empirical ones: in fact, the p-values of the ME models read $p_{(1)} \simeq 0.06$, $p_{(2)} \simeq 0.99$, $p_{\rm TS} \simeq 0.99$, $p_{\rm TSF} \simeq 0.32$ while the p-values of the econometric models read $p_{\rm Pois} \simeq 0.001$, $p_{\rm NB} \simeq 0.0008$, $p_{\rm ZIP} \simeq 0.63$, $p_{\rm ZINB} \simeq 0.001$.

Coming to the higher-order properties, it is evident that the major-

ity of the econometric models fails to overlap with the empirical cloud of points (see Fig. 2(b)): the one providing the best prediction is the ZIP model, whose performance represents quite an improvement with respect to the one provided by the 'plain' Poisson model. While this is quite evident for what concerns the prediction of the ANND values, the performances of the ZIP and of the 'plain' Poisson model become less different when tested on the BCC values. On the contrary, the performance of the ZINB model closely resembles that of the negative binomial one when tested both on the ANND and on the BCC. As for the local properties, the KS test reveals that the only model outputting predictions compatible with the empirical values (at the significance level of 5%) is the ZIP one: in fact, $p^{\text{ANND}}_{\text{ZIP}} \simeq 0.38$, $p^{\text{BCC}}_{\text{ZIP}} \simeq 0.08$).

For what concerns ME models, the ones performing best are those constraining the degrees, i.e. the model induced by $H_{(2)}$ and its two-step counterpart, whose topological estimation step is carried out by employing p^{UBCM}_{ij}. The evidence that their performances in reproducing the purely binary structure of a network are very similar lets us suspect that $p^{(2)}_{ij} \simeq p^{\text{UBCM}}_{ij}$ and conclude that the purely econometric information encoded into $p^{(2)}_{ij}$ does not add much to what is already conveyed by the purely topological one. On the other hand, ME models not constraining the degrees provide predictions differing from the empirical trends to quite a large extent; still, as the KS test reveals, the only ME model outputting predictions that are not compatible with the empirical values (at the significance level of 5%) is the one induced by $H_{(1)}$.

The overall accuracy of our models in reproducing a network topology can be proxied by the index $\Delta_L = |\langle L \rangle - L|/L$ amounting at $\Delta^{\text{Pois}}_L \gtrsim \Delta^{\text{NB}}_L = \Delta^{\text{ZINB}}_L \simeq 6\%$ while $\Delta^{\text{ZIP}}_L \simeq 0.5\%$ and $\Delta^{\text{ME}}_L = 0$ for each ME model. This is confirmed by our analysis of single link statistics: in fact, $\langle \text{ACC} \rangle_{(2)} \simeq 0.83$ attains the largest value, followed by $\langle \text{ACC} \rangle_{\text{TS}} \simeq 0.81$ and $\langle \text{ACC} \rangle_{\text{ZIP}} \simeq 0.77$. Remarkably, $\langle \text{PPV} \rangle_{(2)} \simeq 0.86$ attains the largest value, indicating that the ME model induced by $H_{(2)}$ is the one placing links best among all the models in our basket.

Let us, now, consider the weighted properties (see Fig. 3). Overall, the

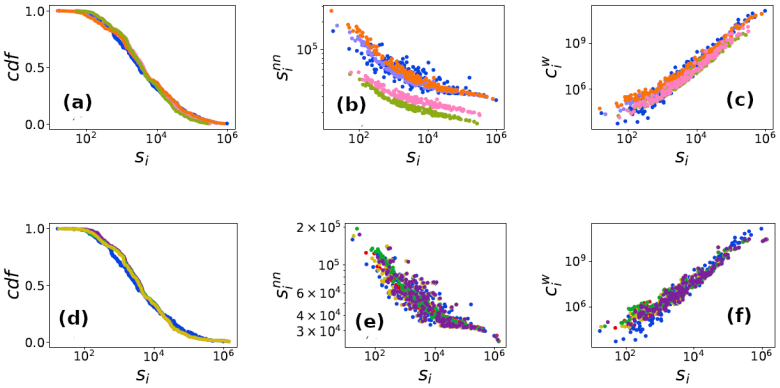

Figure 3: Performance of econometric models (top row) versus the performance of ME models (bottom row) in reproducing: (a), (d) the strengths; (b), (e) the ANNS; (c), (f) the WCC. Empirical points are indicated as ●; econometric and ME models are indicated as follows: ● - Poisson; ● - Negative Binomial; ● - ZIP; ● - ZINB; ● - $H_{(1)}$; ● - $H_{(2)}$; ● - TS; ● - TSF. Results refer to the year 2000 of the dataset curated by Gleditsch [38].

distribution of the strengths is reproduced quite well by all models considered here, although no one explicitly constrains them. This seems to indicate that the purely econometric information 'feeded' into our models indeed plays a role - which, however, is limited to ensure that the intensive margins (and the related properties, as we will see) are accurately predicted. The larger explanatory power of econometric models becomes now evident: all of them output predictions that are compatible with the empirical values. Although the same result holds true for ME models, the latter ones are outperformed by purely econometric models - the best-performing ones in predicting the strengths being the Poisson-like ones.

Coming to the higher-order properties, let us notice that the best-performing econometric models in reproducing the ANNS values are the ZIP and the 'plain' Poisson ones whose performances differ less than in the ANND case - although the KS test lets the ZIP model win. On the other hand, the ZINB and the negative binomial models (whose perfor-

Dataset	$H_{(1)}$	$H_{(2)}$	TSF	TS	Poisson	ZIP	Negative Binomial	ZINB
Gleditsch_{90}	0.72	**0.81**	0.73	0.79	0.75	0.75	0.68	0.70
Gleditsch_{91}	0.69	**0.78**	0.70	0.77	0.73	0.73	0.65	0.68
Gleditsch_{92}	0.70	**0.79**	0.71	0.77	0.73	0.73	0.65	0.68
Gleditsch_{93}	0.70	**0.79**	0.71	0.78	0.74	0.74	0.66	0.69
Gleditsch_{94}	0.70	**0.79**	0.72	0.78	0.75	0.75	0.66	0.69
Gleditsch_{95}	0.70	**0.80**	0.72	**0.80**	0.75	0.76	0.68	0.69
Gleditsch_{96}	0.72	**0.81**	0.73	**0.81**	0.76	0.77	0.68	0.69
Gleditsch_{97}	0.73	**0.82**	0.74	0.81	0.77	0.77	0.69	0.71
Gleditsch_{98}	0.73	**0.82**	0.74	0.81	0.77	0.77	0.69	0.71
Gleditsch_{99}	0.73	**0.82**	0.74	0.81	0.77	0.77	0.69	0.72
Gleditsch_{00}	0.74	**0.82**	0.74	0.81	0.78	0.77	0.70	0.72
BACI_{07}	0.83	**0.88**	0.83	0.87	0.84	0.84	0.76	0.77
BACI_{08}	0.83	**0.88**	0.83	0.87	0.85	0.84	0.76	0.77
BACI_{09}	0.84	**0.88**	0.84	0.87	0.84	0.84	0.76	0.77
BACI_{10}	0.85	**0.89**	0.85	0.88	0.85	0.85	0.77	0.78
BACI_{11}	0.85	**0.89**	0.85	0.88	0.86	0.85	0.77	0.78
BACI_{12}	0.85	**0.89**	0.85	0.88	0.86	0.85	0.77	0.78
BACI_{13}	0.85	**0.90**	0.85	0.89	0.86	0.86	0.77	0.78
BACI_{14}	0.85	**0.90**	0.85	0.89	0.86	0.85	0.77	0.78
BACI_{15}	0.85	**0.90**	0.85	0.89	0.85	0.84	0.77	0.78
BACI_{16}	0.84	**0.90**	0.84	0.89	0.85	0.84	0.76	0.78
BACI_{17}	0.85	**0.90**	0.85	0.89	0.85	0.85	0.77	0.78

Table 1: Accuracy of ME and econometric models in reconstructing both the Gleditsch and the BACI datasets. The best-performing models are $H_{(2)}$ and TS.

mances are, again, very similar) completely fail in capturing the empirical values. All predictions from ME models overlap with the empirical ANNS values: as the KS test reveals, the only ME model outputting predictions that are not compatible with the empirical values (at the significance level of 1%) is the one induced by $H_{(1)}$. For what concerns the values of the WCC, both the econometric and the ME models perform quite satisfactorily in capturing its rising trend; however, the KS test reveals that only the econometric models and the TS model output predictions compatible with the WCC empirical values (at the significance level of 1%).

To proxy the accuracy of our models in reproducing the weighted network structure we have considered the index $\Delta_W = |\langle W \rangle - W|/W$ amounting at $\Delta_W^{\text{ZINB}} \simeq 95\%$, $\Delta_W^{\text{NB}} = 60\%$, $\Delta_W^{\text{ZIP}} \simeq 0.3\%$ and $\Delta_W^{\text{Pois}} \simeq 0$ while $\Delta_W^{\text{ME}} = 0$ for the ME models constraining the total weight and $\Delta_W^{\text{TS}} \gtrsim \Delta_W^{\text{TSF}} \simeq 0.2\%$ for the ME two-step ones.

In order to understand if the conclusions above can be generalized,

let us calculate the accuracy of all models in our basket, for all years constituting our two datasets: the results, summed up in Tab. 1, confirm that model $H_{(2)}$ systematically outperforms all competing models. As an additional test, we computed a Kolmogorov-Smirnov (KS) compatibility frequency f_m^s. The KS compatibility frequency quantifies the proportion of times the distribution of a given expected statistic s under model m coincides with the empirical distribution, as determined by the two-sample KS test. A value of $f_m^s = 0.8$ indicates that the model's expected statistic distribution aligns with the empirical distribution for 80% of the data points. The results, shown in Fig. 4, point out that ME models are the ones for which compatibility is largest.

In addition to the measures above, we also investigate centrality metrics, which quantify the structural importance of individual nodes.

The eigenvector centrality of node i is defined as the i-th entry of the eigenvector corresponding to the largest eigenvalue of the (weighted) adjacency matrix. Since it is real and symmetric, its eigenvalues are guaranteed to exist and be real-valued.

The Katz centrality of node i is defined as the i-th entry

$$K_i = \left[(\mathbf{I} - \xi \mathbf{A})^{-1} \cdot \mathbf{1} \right]_i \tag{2.54}$$

where $\mathbf{I}$ is the identity matrix whose dimensions coincide with those of the adjacency matrix A, $\mathbf{1}$ is the N-dimensional vector of ones and ξ is a parameter strictly less than the reciprocal of the maximum eigenvalue. Katz centrality generalizes the degree centrality, accounting for the total number of possible walks originating from a source node. The weighted version of Katz Centrality can be calculated upon substituting $\mathbf{A}$ with $\mathbf{W}$.

The closeness centrality of node i is defined as

$$c_i = \frac{N-1}{\sum_j d_{i,j}} \tag{2.55}$$

where $d_{i,j}$ is the distance of the shortest path connecting node i and node i. It is the inverse of the farness, i.e. the average distance of a node from all the other ones.

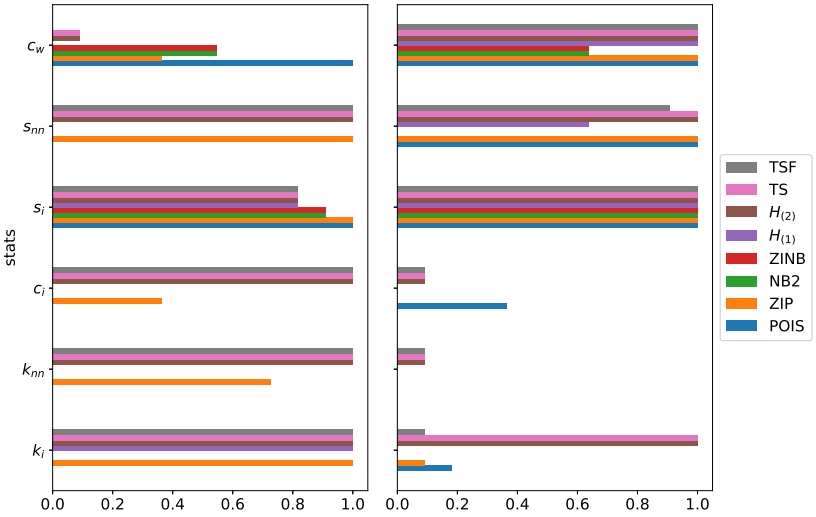

Figure 4: Percentage of times the empirical value of a given network statistics is compatible with its ensemble distribution, according to a KS test at the significance level of 5% (left, for the Gleditsch dataset; right, for the BACI dataset): a value of 1 means that the given network statistic is compatible with the model-induced ensemble distribution across all years of the considered dataset. Generally speaking, ME models are the ones for which compatibility is largest - although econometric models are characterized by a compatibility which is large as well whenever weighted measures are considered.

The betweenness centrality of node i is defined as

$$b_i = \sum_{j \neq i \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \tag{2.56}$$

i.e. as the sum, over all origin-destination pairs - with the exclusion of node i, of the fraction of shortest paths crossing node i. It quantify to which extent a given node acts as a bridge, connecting different pairs of nodes. Unfortunately, closeness and betweenness centrality lack a straightforward extension to weighted networks: therefore, we will focus on their binary variants only.

In Fig. 5, we inspect the compatibility between the empirical and the theoretical - according to the discrete models considered in the present chapter - distributions of our centrality measures. Analogously to the previous exercise, a compatibility $f = 1$ indicates the existence of a p-value larger than 0.05 (further implying that the null hypothesis that the empirical and the theoretical sets of measures come from the same distribution cannot be rejected), for each year. The results are strongly dataset dependent: for example, for what concerns the Gledistch dataset, ME methods generally outperform, or (at least) perform competitively when compared to, econometric approaches - the only exception being constituted by the weighted Katz centrality; on the other hand, for what concerns the BACI dataset, ME methods typically outperform econometric approaches only for closeness and betweenness centrality.

A different perspective is provided by the Reconstruction Accuracy (RA_m^s) of model m for statistic s depicted in Fig. 6. The reconstruction accuracy RA_m^s, expressed as a percentage, measures the fraction of node-specific values of statistic s falling within the confidence intervals (CIs) generated by sampling from the theoretical distribution of model m at a significance level of 5%. Figure 6 shows the average RA across the years, accompanied by confidence intervals illustrating the variability of this average. On both datasets, the ME models named TS and $H_{(2)}$ steadily outperform the others, including the econometric ones.

The ranking induced by the eigenvector centrality plays a critical role in International Trade Economics, as it reveals the network-related com-

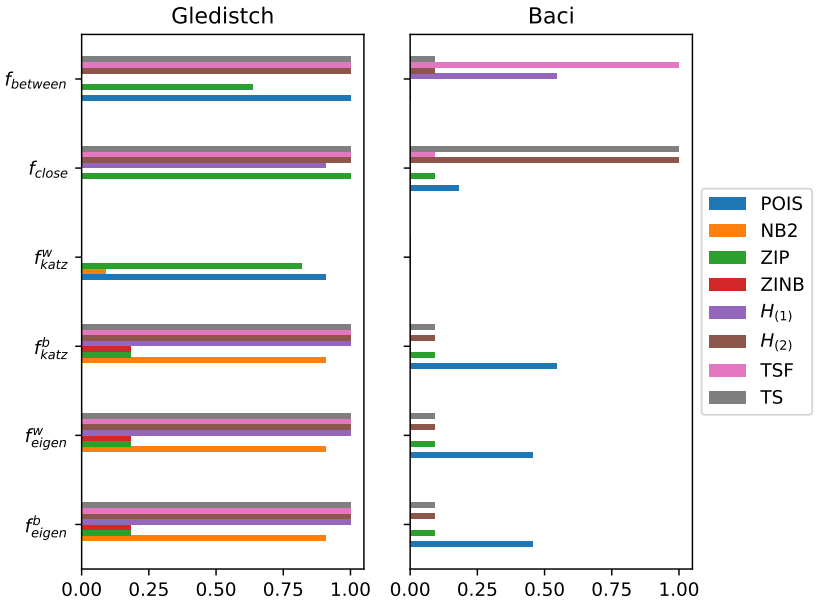

Figure 5: Percentage of times the empirical value of a given network statistics is compatible with its ensemble distribution, according to a KS test at the significance level of 5% (left, for the Gleditsch dataset; right, for the BACI dataset): a value of 1 means that the given network statistic is compatible with the model-induced ensemble distribution across all years of the considered dataset. Compatibility on these measures strictly depends on the dataset at hand. For Gledistch, high compatibility of ME models is observed, with the exception of the weighted Katz Centrality; for the Baci dataset, a higher compatibility (even if low on absolute terms) of eigenvector centralities is observed for the Poisson Model, while ME are performing better for closeness and betweenness centrality measures.

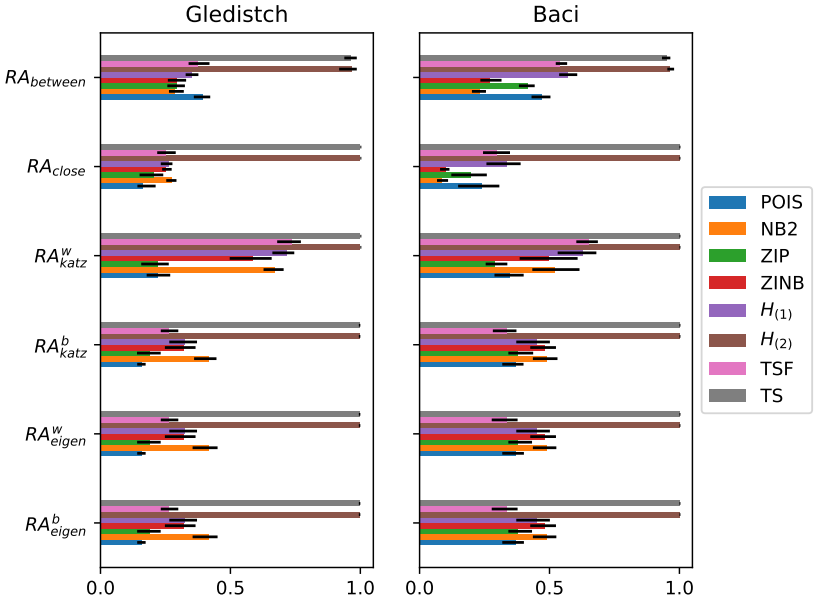

Figure 6: Reconstruction Accuracy for eigenvector, Katz, closeness and betweenness centrality measures, calculated as the percentage of node-specific values of the statistics s falling within the CI, induced by model m, at the significance level of 5%; yearly percentages are, then, averaged. The whiskers represent the 2.5 and 97.5 percentiles of each RA^{s}_{m} distribution, across different years. For both datasets (Gledistch on the left and Baci on the right), the methods $H_{(2)}$ and TS significantly outperform the others.

parative advantage that a country possesses in terms of influence. The accuracy in reproducing the empirical ranking can be assessed by computing the Spearman Rank Correlation R, a metric that lies within the interval $(-1, 1)$: a Spearman correlation of -1 indicates the presence of a perfectly negative correlation, a Spearman correlation of 0 indicates independence, a Spearman correlation of 1 the presence of a perfectly positive correlation. Therefore, a larger value hints at a larger similarity between the empirical and theoretical rankings. Figure 7 shows the average Spearman Rank Correlation, accompanied by confidence intervals illustrating the variability of this average, for both the Gledistch (left panel) and the BACI (right panel) datasets, and each model under investigation. As evident from the figure, the Poisson model and its Zero-Inflated counterpart consistently outperform other models in replicating the ranking of weighted centrality measures; conversely, the ME models named $H_{(2)}$ and TS outperform the others in replicating the ranking of binary centrality measures.

Let us, now, ask ourselves if a criterion exist to carry out a principled comparison of the performance of the models considered in the present work. The answer is positive and lays in the adoption of the popular Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), respectively defined as

$$\text{AIC} = 2K - 2\mathcal{L} \tag{2.57}$$

and

$$\text{BIC} = K \ln n - 2\mathcal{L} \tag{2.58}$$

where $\mathcal{L}$ is the log-likelihood of the tested model evaluated at its maximum, K is the number of parameters characterizing the model itself and n is the cardinality of the set of observations - estimated as $N(N-1)/2$ for undirected network data. Model selection based on these criteria prescribes to rank models according to either their AIC or BIC value and choose the one minimizing it. Table 2 shows both the AIC and the BIC values for all the models considered here.

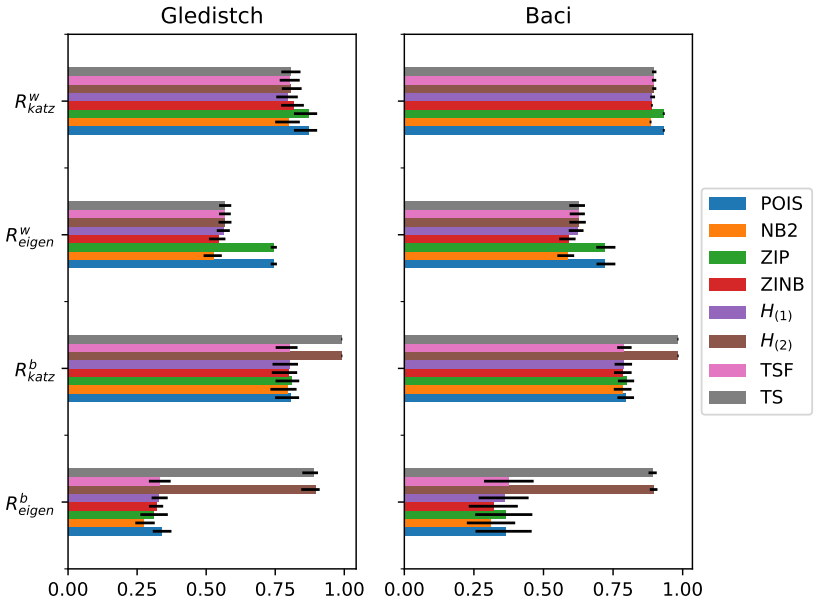

Figure 7: Average Spearman Rank Correlations for eigenvector and Katz centrality measures. An higher value corresponds to a higher correlation between the empirical and the model-induced ranking. The whiskers represent the 2.5 and 97.5 percentiles of the Spearman cross correlation measures, across different years. For binary measures, the ME methods significantly display a higher correlation in ranking; conversely, Poisson and its Zero-Inflated analog display higher ranking correlations for the weighted measures.

Quite surprisingly, the negative binomial model is the favored one among the econometric models, followed by its zero-inflated version; however, its bad performance in reproducing the empirical trends makes the choice of including it among the most suitable recipes for modelling trade data highly questionable. On the other hand, the ZIP model performs much better in reconstructing the trends of both local and higher-order properties although being much less parsimonious than both versions of the negative binomial model. Apparently, then, the question about which model to prefer - i.e. the favored one by information criteria or the best-performing one in reproducing trends? - cannot be properly answered by just considering purely econometric models. On the other hand, such a question can be unambiguously answered as soon as one switches to the class of maximum-entropy models: now, both the AIC- and the BIC-based rankings favor the model described by the Hamiltonian $H_{(2)}$ - the one encoding the information about the degree sequence and the total weight, plus admitting a tunable function of the weights - i.e. precisely the most accurate in replicating many (if not all of the) empirical trends.

For the sake of comparison, we have included into the basket of maximum-entropy models the Undirected Enhanced Configuration Model (UECM), i.e. the model performing best in presence of complete information about the constraints - degrees and strengths, in the specific case - of a given networked system: as evident from the table, it is disfavored with respect to the model described by the Hamiltonian $H_{(2)}$, an evidence signalling that while the information encoded into the degrees is essential (i.e. the latter ones must be explicitly constrained), the one carried by the strengths appears to be 'less fundamental' since providing a good approximation of them is enough to obtain an overall good reconstruction.

In order to understand if the conclusions above can be generalized, let us calculate the Akaike weights for the models in our basket $\mathcal{M}$. The Akaike weight for the i-th model is defined as

$$w_i = \frac{e^{-\Delta_i/2}}{\sum_m e^{-\Delta_m/2}}, \tag{2.59}$$

Model	Δ_L	Δ_W	AIC	BIC
Poisson	$\simeq 0.07$	0^*	$\simeq 5088080.1$	$\simeq 5088105.1$
ZI Poisson	$\simeq 4.5 \cdot 10^{-3}$	$\simeq 3.3 \cdot 10^{-3}$	$\simeq 5052767.3$	$\simeq 5052800.6$
Negative Binomial	$\simeq 0.06$	$\simeq 0.59$	$\simeq \mathbf{175693.2}$	$\simeq \mathbf{175726.6}$
ZI Negative Binomial	$\simeq 0.06$	$\simeq 0.95$	$\simeq 176071.4$	$\simeq 176113.1$
ME model $H_{(1)}$	0^*	0^*	$\simeq 182909.5$	$\simeq 182951.2$
ME model $H_{(2)}$	0^*	0^*	$\simeq \mathbf{174050.0}$	$\simeq \mathbf{175550.4}$
TS	0^*	$\simeq 6.8 \cdot 10^{-3}$	$\simeq 176129.1$	$\simeq 177629.5$
TSF	0^*	$\simeq 2.2 \cdot 10^{-3}$	$\simeq 185940.2$	$\simeq 185981.8$
UECM	0^*	0^*	$\simeq 186602.8$	$\simeq 189536.8$

Table 2: Ranking of ME and econometric models, according to the AIC and BIC values. While the negative binomial model is the favored one among the econometric models, it performs badly in reproducing both local and higher-order topological properties; on the other hand, the ZIP model reproduces the higher-order statistics quite accurately but is disfavored by both AIC and BIC. The solution to this dilemma comes from considering a different class of reconstruction models, i.e. the maximum-entropy ones: beside being the favored one by information criteria, the $H_{(2)}$ model achieves a very good reconstruction of both binary and weighted network properties (since $\Delta L = |\langle L \rangle - L|/L$ and $\Delta W = |\langle W \rangle - W|/W$, the symbol 0^* indicates that the model exactly reproduces the corresponding constraint). Results refer to the year 2000 of the dataset curated by Gleditsch [38].

with $\Delta_i = \text{AIC}_i - \min\{\text{AIC}_m\}_{m \in \mathcal{M}}$. Results on the dataset curated by Gleditsch show that the negative binomial and the $H_{(2)}$ models 'compete', in the sense that $H_{(2)}$ performs best (i.e. $w_{H_{(2)}} \simeq 1$) in the (bunches of) years 1990-1993 and 1997-2000 while the negative binomial model performs best (i.e. $w_{NB} \simeq 1$) in the (bunch of) years 1994-1996. For what concerns the BACI dataset, instead, the competing models are three: in fact, while $H_{(2)}$ performs best in the (bunches of) years 2007, 2009 and 2015-2017, the negative binomial outperforms the others in the (bunches of) years 2008 and 2010-2014; however, the ZINB model has a positive, non-negligible Akaike weight in the years 2008, 2012 and 2014, hence performing as well as the negative binomial one.

Let us now consider a couple of additional exercises, carried out on both datasets considered in the present work.

The first one concerns link prediction and was carried out by follow-

ing the reference [79]. Specifically, we have approached link prediction from a temporal perspective, inspecting the accuracy achieved by our reconstruction models at time $t+1$ given the knowledge about the network topology (for the maximum-entropy models) and other exogenous variables (for the purely econometric models) at time t. In other words, we opted for a one-lagged link prediction, calculating the log-likelihood

$$\mathcal{L}_{1l} = \ln \left[\prod_{i<j} \left(p_{ij}^{(t)} \right)^{a_{ij}^{(t+1)}} \left(1 - p_{ij}^{(t)} \right)^{1-a_{ij}^{(t+1)}} \right] \tag{2.60}$$

for each statistical model in our basket, the coefficients $\left\{ p_{ij}^{(t)} \right\}_{i,j=1}^{N}$ being the probabilities output by any model at time t and the coefficients $\left\{ a_{ij}^{(t+1)} \right\}_{i,j=1}^{N}$ being the entries of the adjacency matrix at time $t+1$. When carried out on the pairs of years 1993-1994, 1994-1995, 1995-1996, 1996-1997, 1997-1998, 1998-1999 and 1999-2000 of the dataset curated by Gleditsch and on the pairs of years 2008-2009, 2009-2010, 2010-2011, 2013-2014, 2014-2015,2015-2016 and 2016-2017 of the BACI dataset, the exercise above shows $H_{(2)}$ to outperform not only the entire class of econometric models but also the purely binary, maximum-entropy ones - as confirmed by the Akaike weights induced by the log-likelihood above.

To provide a more refined picture of the performance of our models in providing one-lagged predictions, we have also calculated their one-lagged accuracy, defined as

$$\langle \text{ACC} \rangle_{1l} = \frac{\langle \text{TP} \rangle_{1l} + \langle \text{TN} \rangle_{1l}}{\binom{N}{2}} \tag{2.61}$$

with $\langle \text{TP} \rangle_{1l} = \sum_{i<j} a_{ij}^{(t+1)} p_{ij}^{(t)}$ and $\langle \text{TN} \rangle_{1l} = \sum_{i<j} \left(1 - a_{ij}^{(t+1)} \right) \left(1 - p_{ij}^{(t)} \right)$. The results are reported in Tab. 3 and confirm what has been previously said: $H_{(2)}$ is the one performing best.

As a second exercise, we have tested the accuracy of our models in estimating link-specific weights - hence, carrying out what we have called a 'weight prediction' exercise. To this aim, we have considered each expected weight and calculated the confidence interval enclosing the 95%

Dataset	$H_{(1)}$	$H_{(2)}$	TSF	TS	Poisson	ZIP	Negative Binomial	ZINB
Gleditsch_{94}	0.70	**0.79**	0.72	0.77	0.75	0.74	0.66	0.69
Gleditsch_{95}	0.71	**0.80**	0.72	0.78	0.75	0.75	0.66	0.69
Gleditsch_{96}	0.72	**0.80**	0.73	**0.80**	0.76	0.76	0.67	0.70
Gleditsch_{97}	0.73	**0.82**	0.74	0.80	0.77	0.77	0.67	0.70
Gleditsch_{98}	0.73	**0.82**	0.74	0.81	0.77	0.77	0.69	0.71
Gleditsch_{99}	0.73	**0.82**	0.74	0.81	0.77	0.77	0.69	0.71
Gleditsch_{00}	0.73	**0.82**	0.74	0.81	0.77	0.77	0.69	0.71
BACI_{09}	0.84	**0.88**	0.84	0.87	0.85	0.84	0.76	0.77
BACI_{10}	0.84	**0.88**	0.84	0.87	0.85	0.84	0.76	0.77
BACI_{11}	0.85	**0.89**	0.85	0.88	0.86	0.85	0.77	0.78
BACI_{14}	0.85	**0.89**	0.85	0.88	0.86	0.85	0.77	0.78
BACI_{15}	0.85	**0.89**	0.85	0.88	0.86	0.85	0.77	0.78
BACI_{16}	0.84	**0.88**	0.84	0.88	0.85	0.84	0.77	0.78
BACI_{17}	0.85	**0.89**	0.85	0.88	0.85	0.84	0.77	0.78

Table 3: One-lagged accuracy, quantifying the performance of ME and econometric models in providing one-lagged predictions on both the Gleditsch and the BACI datasets. $H_{(2)}$ and TS are systematically the best-performing models.

of total probability around it. On the practical side, we have sampled 1000 configurations from the ensemble induced by each model in our basket and calculated the (ensemble-induced) 2.5 and 97.5 percentiles for each specific weight; then, we have calculated the percentage of empirical weights 'falling' within the corresponding CIs - now treated as error bars accompanying the point-estimate of each weight. The results of this exercise are reported in in Tab. 4: as it can be appreciated, maximum-entropy models compete with both the negative binomial and the ZINB ones - although the latter (slightly) outperform the former.

Finally, to assess the presence of heteroskedasticity, we employ the White and the Breusch-Pagan tests. It consistently return p-values lower than 0.05, indicating that the residuals exhibit heteroskedasticity, for all methods and across all years. As highlighted in the introduction, heteroskedasticity poses a significant challenge in parameter estimation, arising from various sources. In our case, employing the conventional gravity model for weight estimation raises concerns about the so-called 'omitted variable bias': this issue can be addressed by incorporating additional factors into the model specification, such as common language, colonial history, etc. Alternatively, we can either introduce fixed-effects

Dataset	$H_{(1)}$	$H_{(2)}$	TSF	TS	Poisson	ZIP	Negative Binomial	ZINB
Gleditsch_{90}	0.96	0.96	0.94	0.96	0.62	0.67	**0.98**	0.97
Gleditsch_{91}	0.96	0.96	0.94	0.96	0.63	0.71	**0.98**	0.97
Gleditsch_{92}	0.96	0.96	0.94	0.96	0.63	0.70	**0.98**	0.97
Gleditsch_{93}	0.96	0.96	0.94	0.96	0.64	0.69	**0.98**	0.97
Gleditsch_{94}	0.96	0.96	0.94	0.96	0.64	0.67	**0.98**	0.97
Gleditsch_{95}	0.96	0.96	0.94	0.96	0.61	0.68	**0.98**	0.97
Gleditsch_{96}	0.96	0.96	0.94	0.96	0.60	0.65	**0.98**	0.97
Gleditsch_{97}	0.96	0.96	0.94	0.96	0.60	0.63	**0.98**	0.97
Gleditsch_{98}	0.96	0.96	0.94	0.96	0.61	0.63	**0.98**	0.97
Gleditsch_{99}	0.96	0.96	0.94	0.96	0.61	0.62	**0.98**	0.97
Gleditsch_{00}	0.96	0.96	0.95	0.96	0.60	0.63	**0.97**	**0.97**
BACI_{07}	0.93	0.95	0.92	0.95	0.43	0.43	**0.97**	**0.97**
BACI_{08}	0.92	0.94	0.91	0.94	0.40	0.40	**0.97**	**0.97**
BACI_{09}	0.93	0.95	0.92	0.95	0.43	0.43	**0.97**	**0.97**
BACI_{10}	0.92	0.94	0.91	0.95	0.41	0.41	**0.97**	**0.97**
BACI_{11}	0.92	0.94	0.90	0.94	0.38	0.39	**0.97**	**0.97**
BACI_{12}	0.91	0.94	0.90	0.94	0.38	0.38	**0.97**	**0.97**
BACI_{13}	0.91	0.93	0.90	0.93	0.37	0.38	**0.97**	**0.97**
BACI_{14}	0.91	0.93	0.90	0.94	0.38	0.39	**0.97**	**0.97**
BACI_{15}	0.91	0.94	0.91	0.94	0.41	0.41	**0.97**	**0.97**
BACI_{16}	0.92	0.94	0.91	0.94	0.40	0.41	**0.97**	**0.97**
BACI_{17}	0.91	0.94	0.91	0.94	0.39	0.39	**0.97**	**0.97**

Table 4: Performance of ME and econometric models in providing accurate predictions of weights, on both the Gleditsch and the BACI datasets, quantified by calculating the percentage of empirical weights falling within the confidence interval enclosing the 95% probability around the corresponding expected value. Our maximum-entropy models compete with both the negative binomial and the ZINB ones - although the latter (slightly) outperform the former.

parameters or adopt a White Variance-Covariance matrix for the weights, in order to mitigate the impact of heteroskedasticity.

2.4 Discussion

The 2008 global financial crisis has dramatically clarified that bilateral trade relationships can explain only a small fraction of the impact that an economic shock, originating in a given country, can have on another country which is not a direct trade partner, urging researchers in economics to adopt a network-aware perspective; this, in turn, has motivated us to carry out a methodological comparison on real-world cases, with the aim of clarifying pros and cons of both approaches.

Researchers in economics have dealt with the issue of reconstructing

a network topology by approaching the simpler problem of reproducing the number of missing connections - or, equivalently, the link density. For instance, as we see from the chosen snapshot of the dataset curated by Gleditsch [38], although the error of the Poisson model in reproducing L is overall small (amounting at $\Delta_L^{\text{Pois}} \simeq 7\%$), it can be further reduced by adopting the zero-inflated version of it; on the contrary, inflating zeros does not improve the performance of the negative binomial model in reproducing the link density as it already underestimates L. The ability of a model in reproducing a global quantity such as the link density proxies its ability in providing a good estimation of local as well as higher-order topological properties (i.e. the degrees, the ANND and the BCC): from this point of view, the zero-inflated Poisson model is the one performing best among the econometric models. However, it is largely disfavored by information criteria such as AIC and BIC, a result suggesting that it may be not parsimonious enough.

Some of the problems of purely econometric models are solved by looking at a different class of statistical models, i.e. the physics-inspired ones. In particular, the model described by the Hamiltonian $H_{(2)} = \sum_i \theta_i k_i + \psi_0 W + \sum_{i<j} \psi_{ij} w_{ij}$ provides a very accurate reconstruction while being favored by information criteria. Remarkably, although it is defined by $N + 1$ purely topological constraints, both AIC and BIC reveal the latter to be 'irreducible', i.e. necessary to provide a satisfactory explanation of the network generating process. For the sake of comparison, Fig. 8 explicitly shows the performance of the models favored by the adopted information criteria (i.e. the negative binomial model and its zero-inflated version) with that of the ME model described by $H_{(2)}$: it is evident that the ME model outperforms the purely econometric ones, still achieving a good ranking.

Looking at the class of ME models in more detail, our analysis indicates that the information carried by the strengths is not as 'fundamental' as the one carried by the degrees: this is evident upon considering that 1) the UECM is always disfavored with respect to the models just constraining the degrees; 2) the second best-performing ME model is (always) the two-step one, defined by a first purely topological step, accounting for

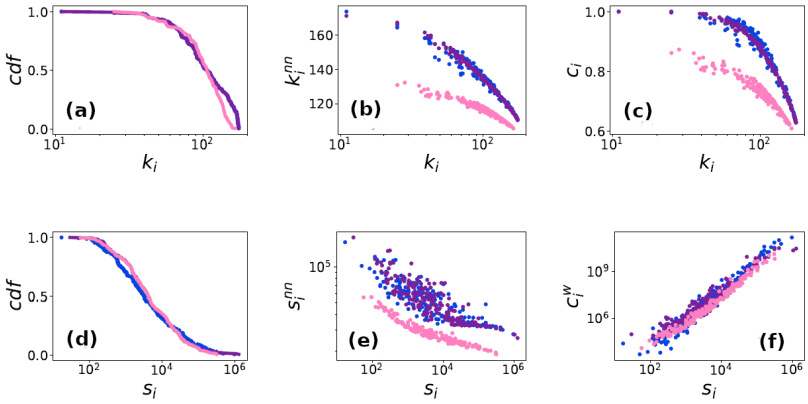

Figure 8: Performance of the negative binomial model versus the performance of the ME model described by the Hamiltonian $H_{(2)}$, in reproducing: (a) the degree distribution; (b) the ANND; (c) the BCC; (d) the strength distribution; (e) the ANNS; (f) the WCC. While a look at Tab. 1 suggests the negative binomial to be the econometric model performing best, this is definitely not the case as explicitly plotting its predictions against the empirical trends reveals. Notice that in (d) the distributions of strength induced by $H_{(2)}$ and by the negative binomial model overlap for a large range of values whereas, for other statistics, $H_{(2)}$ produces the best fit - confirming the importance of a good topological estimation. Empirical points are indicated •; econometric and ME models are indicated as follows: • - Negative Binomial; • - $H_{(2)}$. Results refer to the year 2000 of the dataset curated by Gleditsch [38].

the degrees, followed by an econometric-wise estimation of the weights. On top of that, we explicitly notice that topological information (e.g. the one provided by the link density or the degree sequence) usually correlates with node-specific, economic covariates; hence, excluding such information from the model may lead to the so-called 'omitted variable' bias. As proved by AIC and BIC, ME models represent the best compromise between goodness-of-fit and parsimony: in fact, they allow for structural information to be included, keeping the aforementioned type of bias low, while leading to a better description of economic systems than that provided by traditional, econometric models.

Model	p_{ij}	$\langle w_{ij}\vert a_{ij}=1\rangle$
POIS	$1-e^{-z_{ij}}$	$z_{ij}\left(1-e^{-z_{ij}}\right)^{-1}$
NB2	$1-(1+\alpha z_{ij})^{-m}$	$\frac{z_{ij}}{1-(1+\alpha z_{ij})^{-m}}$
ZIP	$\frac{\hat{x}_{ij}}{1+\hat{x}_{ij}}(1-e^{-z_{ij}})$	$\frac{z_{ij}}{1-e^{-z_{ij}}}$
ZINB	$\frac{\hat{x}_{ij}}{1+\hat{x}_{ij}}\left[1-(1+\alpha z_{ij})^{-m}\right]$	$\frac{z_{ij}}{1-(1+\alpha z_{ij})^{-m}}$
TSF	$\frac{\hat{x}_{ij}}{1+\hat{x}_{ij}}$	$\frac{1+z_{ij}}{1+z_{ij}-y_0 z_{ij}}$
TS	$\frac{x_i x_j}{1+x_i x_j}$	$\frac{1+z_{ij}}{1+z_{ij}-y_0 z_{ij}}$
$H_{(1)}$	$\frac{xy_0 z_{ij}}{1+z_{ij}-y_0 z_{ij}+xy_0 z_{ij}}$	$\frac{1+z_{ij}}{1+z_{ij}-y_0 z_{ij}}$
$H_{(2)}$	$\frac{x_i x_j y_0 z_{ij}}{1+z_{ij}-y_0 z_{ij}+x_i x_j y_0 z_{ij}}$	$\frac{1+z_{ij}}{1+z_{ij}-y_0 z_{ij}}$

Table 5: Models used in this chapter with columns indicating the model-induced probability connection p_{ij} and the weight conditional on the link presence $\langle w_{ij}|a_{ij}=1\rangle$. The parameter $z_{ij}=\exp\{\rho+\beta\ln(\omega_i\omega_j)+\gamma\ln(D_{ij})\}$ stands for the log-link gravity specification, while $\hat{x}_{ij}=\exp\{\theta+\ln(\omega_i\omega_j)-\ln(D_{ij})\}$ stands for the gravity specification used in the logistic binary model.

Our findings may indicate a route towards reconciling econometric and maximum-entropy approaches, suggesting how to build a model that combines the pros of both: the importance of purely structural information (highlighted by physics-inspired models) can be accounted for by a model whose first step that is purely topological in nature (notice, in fact, that the TSF model is disfavored with respect to the TS one) and a second step that takes care of estimating the weighted structure: such an estimation can rest upon econometric considerations, driving the reparametrization of otherwise purely structural models.

In Table 5 we sum up the weighted models explored in this chapter with the specifications related to the model-induced probability connection p_{ij} and to the conditional weight $\langle w_{ij}|a_{ij}=1\rangle$ (conditional on the established link), while in Table 6 the relative structural constraints are portrayed for each model. In this last table, objects the "hat" apex is used to identify the constrained quantities that are approximated with a log-link function in terms of constant and economic covariates.

Model	Type	Binary Constraints	Weighted Constraints
TSF	C-Int	$\{\hat{k}_i\}$	$W_{tot}, \hat{w}_{ij}$
TS	C-Int	$\{k_i\}$	$W_{tot}, \hat{w}_{ij}$
$H_{(1)}$	I-Int	$\sum_i k_i = L$	$W_{tot}, \hat{w}_{ij}$
$H_{(2)}$	I-Int	$\{k_i\}$	$W_{tot}, \hat{w}_{ij}$

Table 6: Entropy-based models used in this chapter with columns indicating the type of induced weights, relative binary constraints and weighted constraints. The type is annotated with C or I for conditional and integrated model respectively, and with Int and Cont for models inducing integer and continuous link weights. Constraints can be topological quantities or their approximation through economic factors. When an approximation in terms of a gravity log-link function is performed, it is indicated with the hat symbol.

Chapter 3

Continuous entropy-based econometric models

This chapter illustrates the results of the analysis published in [2]. Here, we extend the set of models introduced and discussed in the previous chapter to accommodate continuous-valued weights. Then, we test the performance of these novel models in reproducing the structural properties of both the BACI dataset [74, 75] and the Gleditsch dataset [38], describing the pros and cons of each of them; additionally, we study the resilience, against shocks on the weights, of the corresponding statistical ensemble by considering the related Shannon-Fisher plane. Finally, we develop the DyGyS Python package[1]*, equipped with the solvers for both discrete-valued and continuous-valued, entropy-based econometric models and routines to 1) explicitly sample the related statistical ensembles and 2) compute a bunch of network statistics.*

3.1 Introduction

As we have seen in the previous chapter, the traditional GM, although accurate in reproducing the positive trade volumes, cannot replicate structural network properties, unless the topology is completely known [6]. In order to overcome such a limitation, it needs to be 'dressed' with a

[1] Available at `https://github.com/MarsMDK/DyGyS`.

probability distribution $Q(\mathbf{W})$ that produces $\langle \mathbf{W} \rangle_{\text{GM}}$ as an expected value while accounting for null outcomes as well (i.e. those entries reading $w_{ij} = 0$ and representing missing links in the network) [41].

Although maximum-entropy models have been also studied from an econometric perspective (see [57] for a discussion of the economic relevance of the constraints defining the Poisson and the geometric network models), it is only recently that progresses have been made to reconcile the two approaches above, either allowing for economic factors parametrizing the maximum-entropy probability distribution producing links and weights [19, 20, 8, 65, 66, 29, 1] or by introducing network-related statistics into otherwise purely econometric models [80]. On the one hand, the novel framework enriches the methods developed by network scientists with an econometric interpretation; on the other, it enlarges the list of candidate distributions usable for econometric purposes.

Here, we refine the theoretical picture provided in the previous chapter, by introducing models to infer the topology and the weights of undirected networks defined by continuous-valued data. In order to do so, we present a theoretical, physics-inspired framework capable of accommodating both integrated and conditional, continuous models, our goal being threefold: 1) testing the performance of both classes of models on the WTW in order to understand which one is best suited for the task; 2) offering a principled derivation of the conditional, econometric models that are currently available; 3) enlarging the list of continuous-valued distributions to be used for econometric purposes. From an econometric point of view, our work moves along the methodological guidelines defining the class of Generalized Linear Models (GLMs) [81] while enriching it with distributions defined by both econometric and structural parameters; from the point of view of statistical physics, our work expands the class of maximum-entropy network models [58], endowing them with macroeconomic factors.

3.2 Statistical network models

3.2.1 Conditional models

Discrete maximum-entropy models can be derived by performing a constrained maximization of Shannon entropy [59, 60, 61]. In case of continuous probability distributions, mathematical problems are known to affect the definition of Shannon entropy and the resulting inference procedure. One must, thus, introduce the Kullback-Leibler (KL) divergence $D_{\mathrm{KL}}(Q||R)$ of a distribution Q from a prior distribution R and re-interpret the maximization of the entropy of Q as the minimization of $D_{\mathrm{KL}}(Q||R)$ from a given prior distribution R. In formulas, the KL divergence[2] is defined as

$$D_{\mathrm{KL}}(Q||R) = \int_{\mathbb{W}} Q(\mathbf{W}) \ln \frac{Q(\mathbf{W})}{R(\mathbf{W})} d\mathbf{W} \tag{3.1}$$

where $\mathbf{W}$ is one of the possible values of a continuous random variable (in our setting, an entire network with continuous-valued link weights), $\mathbb{W}$ is the set of possible values that $\mathbf{W}$ can take, $Q(\mathbf{W})$ is the (multivariate) probability density function to be estimated and $R(\mathbf{W})$ plays the role of prior distribution - the divergence of $Q(\mathbf{W})$ from which must be minimized. Such an optimization scheme embodies the so-called Minimum Discrimination Information Principle (MDIP), originally proposed by Kullback and Leibler [84] and implementing the idea that, given a prior distribution $R(\mathbf{W})$ and new information that becomes available, an updated distribution $Q(\mathbf{W})$ should be chosen in order to make its discrimination from $R(\mathbf{W})$ as hard as possible; equivalently, the MDIP demands that new data produce an information gain that is as small as possible.

In order to introduce the class of conditional models, we write the posterior distribution $Q(\mathbf{W})$ as

[2]The use of the KL divergence is, now, widespread in the fields of information theory [82] and machine learning [83], e.g. as a loss function within the Generative Adversarial Network (GAN) scheme - the aim of the 'generating' neural network being that of producing samples that cannot be distinguished from those constituting the training set by the 'discriminating' neural network).

$$Q(\mathbf{W}) = P(\mathbf{A})Q(\mathbf{W}|\mathbf{A}) \tag{3.2}$$

where $\mathbf{A}$ denotes the adjacency matrix for the binary projection of the weighted network $\mathbf{W}$. The above equation allows us to split the KL divergence into the following sum of three terms

$$D_{\mathrm{KL}}(Q||R) = S(Q,R) - S(P) - S(\overline{Q}|P) \tag{3.3}$$

where

$$S(P) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \ln P(\mathbf{A}) \tag{3.4}$$

is the *Shannon entropy* of the probability distribution describing the binary projection of the network structure,

$$S(\overline{Q}|P) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) \ln Q(\mathbf{W}|\mathbf{A}) d\mathbf{W} \tag{3.5}$$

is the *conditional Shannon entropy* of the probability distribution of the weighted network structure given the binary projection and

$$S(Q,R) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) \ln R(\mathbf{W}) d\mathbf{W} \tag{3.6}$$

is the *cross entropy* quantifying the amount of information required to identify a weighted network sampled from the distribution $Q(\mathbf{W})$ by employing the distribution $R(\mathbf{W})$. When continuous models are considered, $S(\overline{Q}|P)$ is defined by a first sum running over all the binary configurations within the ensemble $\mathbb{A}$ and an integral over all the weighted configurations that are compatible with a specific, binary structure - embodied by the adjacency matrix $\mathbf{A}$, i.e. such that $\mathbb{W}_{\mathbf{A}} = \{\mathbf{W} : \Theta[\mathbf{W}] = \mathbf{A}\}$.

The expression for $S(Q,R)$ can be further manipulated as follows. Upon separating the prior distribution itself into a purely binary part and a conditional, weighted one, one can write

$$R(\mathbf{W}) = T(\mathbf{A})R(\mathbf{W}|\mathbf{A}) \tag{3.7}$$

an expression that allows us to write $S(Q,R)$ as

$$S(Q,R) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \ln T(\mathbf{A}) - \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) \ln R(\mathbf{W}|\mathbf{A}) d\mathbf{W} \tag{3.8}$$

which, in turn, allows the KL divergence to be re-written as

$$D_{\mathrm{KL}}(Q||R) = D_{\mathrm{KL}}(P||T) + D_{\mathrm{KL}}(\overline{Q}||\overline{R}) \tag{3.9}$$

i.e. as a sum of the addenda

$$D_{\mathrm{KL}}(P||T) = \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \ln \frac{P(\mathbf{A})}{T(\mathbf{A})}, \tag{3.10}$$

$$D_{\mathrm{KL}}(\overline{Q}||\overline{R}) = \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) \ln \frac{Q(\mathbf{W}|\mathbf{A})}{R(\mathbf{W}|\mathbf{A})} d\mathbf{W} \tag{3.11}$$

with $T(\mathbf{A})$ representing the binary prior and $R(\mathbf{W}|\mathbf{A})$ representing the conditional, weighted one. In what follows, we will deal with completely uninformative priors: this amounts at considering the somehow 'simplified' expression

$$-S(Q) = -S(P) - S(\overline{Q}|P) \tag{3.12}$$

with

$$S(P) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \ln P(\mathbf{A}), \tag{3.13}$$

$$S(\overline{Q}|P) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) \ln Q(\mathbf{W}|\mathbf{A}) d\mathbf{W}. \tag{3.14}$$

The (independent) constrained optimization of $S(P)$ and $S(\overline{Q}|P)$ represents the starting point for deriving the members of the class of conditional models.

Choosing the binary constraints

As discussed in section 2.2.2, the functional form controlling for the binary part of conditional models can be derived by carrying out a constrained maximization of the binary Shannon entropy

$$S(P) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \ln P(\mathbf{A}). \tag{3.15}$$

Again, we will again consider the UBCM model, characterized by a probability that any two nodes establish a connection reading

$$p_{ij}^{\text{UBCM}} = \frac{x_i x_j}{1 + x_i x_j}, \tag{3.16}$$

in turn ensuring the entire degree sequence of the network at hand to be reproduced, and the fitness model (FM), characterized by the pair-specific probability coefficient

$$p_{ij}^{\text{FM}} = \frac{\delta\omega_i\omega_j}{1 + \delta\omega_i\omega_j}, \tag{3.17}$$

requiring the estimation of a global parameter only, i.e. δ. Remarkably, the FM has been proven to reproduce the (binary) properties of a wide spectrum of real-world systems [46, 20] as accurately as the UBCM, although requiring much less information.

Choosing the weighted constraints

The constrained maximization of $S(\overline{Q}|P)$ proceeds by specifying the following set of weighted constraints

$$1 = \int_{\mathbb{W}_{\mathbf{A}}} P(\mathbf{W}|\mathbf{A}) d\mathbf{W}, \; \forall\, \mathbf{A} \in \mathbb{A} \tag{3.18}$$

$$\langle C_\alpha \rangle = \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) C_\alpha(\mathbf{W}) d\mathbf{W}, \; \forall\, \alpha \tag{3.19}$$

the first condition ensuring the normalization of the probability distribution and the vector $\{C_\alpha(\mathbf{W})\}$ representing the 'proper' set of weighted constraints (weights are, now, treated as continuous random variables, i.e. $w_{ij} \in \mathbb{R}_0^+, \forall\, i < j$). They induce the distribution reading

$$Q(\mathbf{W}|\mathbf{A}) = \begin{cases} \frac{e^{-H(\mathbf{W})}}{Z_{\mathbf{A}}}, & \mathbf{W} \in \mathbb{W}_{\mathbf{A}} \\ 0, & \mathbf{W} \notin \mathbb{W}_{\mathbf{A}} \end{cases} \tag{3.20}$$

where $H(\mathbf{W}) = \sum_\alpha \psi_\alpha C_\alpha$ is the so-called Hamiltonian, listing the constrained quantities, and $Z_{\mathbf{A}} = \int_{\mathbb{W}_{\mathbf{A}}} e^{-H(\mathbf{W})} d\mathbf{W}$ is the partition function, conditional on the 'fixed topology' $\mathbf{A}$.

The explicit functional form of $Q(\mathbf{W}|\mathbf{A})$ can be obtained only once the functional form of the constraints has been specified. In what follows, we will deal with the Hamiltonian reading

$$H(\mathbf{W}) = \sum_{i<j} f(w_{ij}|\beta_{ij}, \xi_{ij}, \gamma_{ij}) \tag{3.21}$$

with the Lagrange multipliers $(\beta_{ij}, \xi_{ij}, \gamma_{ij})$ satisfying the following requirements:

- $\beta_{ij} \equiv \beta_0 + \beta_{ij}$, where β_0 is the Lagrange multiplier associated with the total weight $\sum_{i<j} w_{ij} \equiv W_1$ and β_{ij} encodes the dependence on purely econometric quantities;
- ξ_{ij} will be kept either in its dyadic form, to constrain the logarithm of each weight, or in its global form, $\xi_{ij} \equiv \xi_0$, to constrain the sum of the logarithms of the weights, i.e. $\sum_{i<j} \ln(w_{ij}) \equiv W_2$;
- $\gamma_{ij} \equiv \gamma_0$ plays the role of the Lagrange multiplier associated with (a function of) the total variance of the logarithms of the weights, i.e. $\sum_{i<j} \ln^2(w_{ij}) \equiv W_3$.

Conditional exponential model. Let us start by considering the simplest, conditional model, defined by the positions $\gamma_{ij} = \xi_{ij} = 0$ and inducing the Hamiltonian

$$H(\mathbf{W}) = \sum_{i<j} f(w_{ij}|\beta_0 + \beta_{ij}) = \sum_{i<j} (\beta_0 + \beta_{ij}) w_{ij}; \tag{3.22}$$

inserting the expression above into Eq. 3.20 leads to the distribution

$$Q(\mathbf{W}|\mathbf{A}) = \prod_{i<j} q_{ij}(w_{ij}|a_{ij}) = \prod_{i<j} \frac{e^{-(\beta_0+\beta_{ij})w_{ij}}}{\zeta_{ij}} = \prod_{i<j} (\beta_0 + \beta_{ij}) e^{-(\beta_0+\beta_{ij})w_{ij}} \tag{3.23}$$

and each node pair-specific distribution induces a (conditional) expected weight reading

$$\langle w_{ij}|a_{ij}=1\rangle=\frac{1}{\beta_0+\beta_{ij}}. \tag{3.24}$$

From a purely topological point of view, constraining each weight *and* their total sum is redundant. However, this is no longer true when turning the conditional, exponential model into a proper econometric one. Its econometric re-parametrization should be consistent with the literature on trade, stating that the weights are monotonically increasing functions of the gravity specification, i.e. $\langle w_{ij}\rangle_{\text{GM}}=e^{\rho+\beta\cdot\ln(\omega_i\omega_j)+\gamma\cdot\ln(d_{ij})}\equiv z_{ij}$, $\forall\, i<j$ and with $e^{\rho}\equiv\tau$; for this reason, the link function usually associated with the exponential distribution prescribes to identify the linear predictor with the inverse of the purely econometric parameter of the model, i.e.

$$\beta_{ij}\equiv z_{ij}^{-1}, \tag{3.25}$$

a position that turns Eq. 3.24 into

$$\langle w_{ij}|a_{ij}=1\rangle=\frac{1}{\beta_0+z_{ij}^{-1}}=\frac{z_{ij}}{1+\beta_0 z_{ij}}; \tag{3.26}$$

notice that the only structural constraint is, now, represented by the total weight (see also Appendix B.1).

Conditional gamma model. Let us, now, consider a different Hamiltonian, constraining each weight, their total sum and the sum of their logarithms, i.e.

$$\begin{aligned}H(\mathbf{W})&=\sum_{i<j}f(w_{ij}|\beta_0+\beta_{ij},\xi_0)\\&=\sum_{i<j}[(\beta_0+\beta_{ij})w_{ij}+\xi_0\ln(w_{ij})]=\sum_{i<j}\beta_{ij}w_{ij}+\beta_0 W_1+\xi_0 W_2;\end{aligned} \tag{3.27}$$

it induces the distribution reading

$$Q(\mathbf{W}|\mathbf{A}) = \prod_{i<j} q_{ij}(w_{ij}|a_{ij})$$
$$= \prod_{i<j} \frac{e^{-(\beta_0+\beta_{ij})w_{ij}-\xi_0 \ln(w_{ij})}}{\zeta_{ij}} = \prod_{i<j} \frac{(\beta_0+\beta_{ij})^{1-\xi_0}}{\Gamma(1-\xi_0)} e^{-(\beta_0+\beta_{ij})w_{ij}} w_{ij}^{-\xi_0}; \tag{3.28}$$

each node pair-specific distribution is characterized by a (conditional) expected weight reading

$$\langle w_{ij}|a_{ij}=1\rangle = \frac{1-\xi_0}{\beta_0+\beta_{ij}} \tag{3.29}$$

and by a (conditional) expected logarithmic weight reading

$$\langle \ln(w_{ij})|a_{ij}=1\rangle = \psi(1-\xi_0) - \ln(\beta_0+\beta_{ij}) \tag{3.30}$$

where the function $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the so-called digamma function.

Such a model can be turned into a proper, econometric one by considering the inference scheme of the gamma model with inverse response, which allows us to identify the linear predictor with the inverse of the purely econometric parameter of the model, i.e.

$$\beta_{ij} \equiv z_{ij}^{-1} \tag{3.31}$$

a position that, in turn, leads to the expressions

$$\langle w_{ij}|a_{ij}=1\rangle = \frac{1-\xi_0}{\beta_0+z_{ij}^{-1}} = \frac{(1-\xi_0)z_{ij}}{1+\beta_0 z_{ij}} \tag{3.32}$$

(allowing the conditional, exponential model to be recovered in case $\xi_0 = 0$, i.e. when the constraint on the sum of the logarithms of the weights is switched-off) and

$$\langle \ln(w_{ij})|a_{ij}=1\rangle = \psi(1-\xi_0) - \ln\left(\beta_0+z_{ij}^{-1}\right) \tag{3.33}$$

(see also Appendix B.1).

Conditional Pareto model. Constraining a slightly more complex function of the weights, i.e. their logarithm, leads to the Hamiltonian

$$H(\mathbf{W}) = \sum_{i<j} f(w_{ij}|\xi_{ij}) = \sum_{i<j} \xi_{ij} \ln(w_{ij}) \tag{3.34}$$

which, in turn, induces the distribution

$$\begin{aligned} Q(\mathbf{W}|\mathbf{A}) &= \prod_{i<j} q_{ij}(w_{ij}|a_{ij}) \\ &= \prod_{i<j} \frac{e^{-\xi_{ij} \ln(w_{ij})}}{\zeta_{ij}} = \prod_{i<j} \frac{w_{ij}^{-\xi_{ij}}}{\zeta_{ij}} = \prod_{i<j} \frac{(\xi_{ij}-1)}{m_{ij}^{1-\xi_{ij}}} w_{ij}^{-\xi_{ij}} \end{aligned} \tag{3.35}$$

where m_{ij} is the minimum, node pair-specific weight allowed by the model. Each node pair-specific distribution is characterized by a (conditional) expected weight reading

$$\langle w_{ij}|a_{ij}=1\rangle = \left(\frac{\xi_{ij}-1}{\xi_{ij}-2}\right) m_{ij}. \tag{3.36}$$

Such a model can be turned into a proper, econometric one by considering the positions

$$\xi_{ij} - 2 \equiv z_{ij}^{-1},\ m_{ij} \equiv w_{min} \tag{3.37}$$

ensuring that the expected weights are monotonically increasing functions of the gravity specification and leading to the expression

$$\langle w_{ij}|a_{ij}=1\rangle = (1+z_{ij}) w_{min} \tag{3.38}$$

where w_{min} is the empirical, minimum weight (see also Appendix B.1).

Let us explicitly notice that the derivation of the gamma and Pareto distributions within the maximum-entropy framework has been already studied in [85]; here, however, we aim at making a step further by individuating a suitable re-definition of these parameters capable of turning them into proper, econometric ones.

Conditional log-normal model. Adding a global constraint on (a function of) the total variance of the logarithms of the weights to the Hamiltonian defining the Pareto model leads to the expression

$$\begin{aligned} H(\mathbf{W}) &= \sum_{i<j} f(w_{ij}|\gamma_0, \xi_{ij}) \\ &= \sum_{i<j} [\xi_{ij} \ln(w_{ij}) + \gamma_0 \ln^2(w_{ij})] = \sum_{i<j} \xi_{ij} \ln(w_{ij}) + \gamma_0 W_3; \end{aligned} \quad (3.39)$$

the Hamiltonian above induces a distribution reading

$$\begin{aligned} Q(\mathbf{W}|\mathbf{A}) &= \prod_{i<j} q_{ij}(w_{ij}|a_{ij}) \\ &= \prod_{i<j} \frac{e^{-\xi_{ij} \ln(w_{ij}) - \gamma_0 \ln^2(w_{ij})}}{\zeta_{ij}} = \prod_{i<j} \frac{e^{-\xi_{ij} \ln(w_{ij}) - \gamma_0 \ln^2(w_{ij})}}{\sqrt{\frac{\pi}{\gamma_0}} e^{\frac{(\xi_{ij}-1)^2}{4\gamma_0}}}; \end{aligned} \quad (3.40)$$

each node pair-specific distribution is characterized by a (conditional) expected weight reading

$$\langle w_{ij}|a_{ij} = 1\rangle = e^{\frac{3-2\xi_{ij}}{4\gamma_0}}, \quad (3.41)$$

by a (conditional) expected, logarithmic weight reading

$$\langle \ln(w_{ij})|a_{ij} = 1\rangle = \frac{1-\xi_{ij}}{2\gamma_0} \quad (3.42)$$

and by a (conditional) logarithmic weight whose squared expectation reads

$$\langle \ln^2(w_{ij})|a_{ij} = 1\rangle = \frac{2\gamma_0 + (1-\xi_{ij})^2}{4\gamma_0^2}. \quad (3.43)$$

Such a model can be turned into a proper, econometric one by considering the position

$$1 - \xi_{ij} \equiv \ln(z_{ij}) \quad (3.44)$$

ensuring that the expected weights are monotonically increasing functions of the gravity specification and leading to the expressions

$$\langle w_{ij}|a_{ij}=1\rangle = e^{\frac{1+2\ln(z_{ij})}{4\gamma_0}}, \tag{3.45}$$

$$\langle \ln(w_{ij})|a_{ij}=1\rangle = \frac{\ln(z_{ij})}{2\gamma_0}, \tag{3.46}$$

$$\langle \ln^2(w_{ij})|a_{ij}=1\rangle = \frac{2\gamma_0+\ln^2(z_{ij})}{4\gamma_0^2} \tag{3.47}$$

(see also Appendix B.1).

3.2.2 Integrated models

The MDIP can be also implemented straightforwardly, i.e. by carrying out a constrained optimization of $D_{\mathrm{KL}}(Q||R)$. In this second case, the following set of constraints

$$1 = \int_{\mathbb{W}} Q(\mathbf{W})d\mathbf{W}, \tag{3.48}$$

$$\langle C_\alpha\rangle = \int_{\mathbb{W}} Q(\mathbf{W})C_\alpha(\mathbf{W})d\mathbf{W},\ \forall\,\alpha \tag{3.49}$$

can be specified, with obvious meaning of the symbols. Differentiating the corresponding Lagrangean functional with respect to $Q(\mathbf{W})$ and equating the result to zero leads to

$$Q(\mathbf{W}) = \frac{R(\mathbf{W})e^{-H(\mathbf{W})}}{\int_{\mathbb{W}} R(\mathbf{W})e^{-H(\mathbf{W})}d\mathbf{W}} \tag{3.50}$$

where $H(\mathbf{W}) = \sum_\alpha \psi_\alpha C_\alpha$ is, again, the Hamiltonian and $Z = \int_{\mathbb{W}} e^{-H(\mathbf{W})}d\mathbf{W}$ is the 'integrated' partition function.

The explicit functional form of $Q(\mathbf{W})$ can be obtained only once the functional form of both the prior distribution and the constraints has been specified as well. In what follows, we will deal with completely uninformative priors, a choice that amounts at considering the simplified expression

$$Q(\mathbf{W}) = \frac{e^{-H(\mathbf{W})}}{\int_{\mathbb{W}} e^{-H(\mathbf{W})}d\mathbf{W}}; \tag{3.51}$$

notice that the result above could have been also derived by carrying out a constrained minimization of

$$D_{\mathrm{KL}}(Q||R) = \int_{\mathbb{W}} Q(\mathbf{W}) \ln Q(\mathbf{W}) d\mathbf{W} \equiv -S(Q) \tag{3.52}$$

i.e. of (minus) the functional named *differential entropy* into which the KL divergence 'degenerates' in case completely uninformative priors are considered.

Choosing the constraints

Let us, now, specify the functional form of the constraints. In what follows, we will deal with a specific instance of the generic Hamiltonian

$$H(\mathbf{W}) = \sum_{i<j} f(w_{ij}|\alpha_{ij}, \beta_{ij}); \tag{3.53}$$

in particular, one could pose $\alpha_{ij} \equiv \alpha_0$, a choice that would lead to constrain the total number of links, or $\alpha_{ij} \equiv \alpha_i + \alpha_j$, a choice that would lead to constrain the whole degree sequence. In what follows we will employ the second functional form and pose $\beta_{ij} \equiv \beta_0 + \beta_{ij}$, where β_0 is the Lagrange multiplier associated with the total weight and β_{ij} encodes the dependence on purely econometric quantities. Our choices induce the Hamiltonian of the so-called *integrated exponential model*, i.e.

$$\begin{aligned} H(\mathbf{W}) &= \sum_{i<j} f(w_{ij}|\alpha_i + \alpha_j, \beta_0 + \beta_{ij}) \\ &= \sum_{i<j} [(\alpha_i + \alpha_j)a_{ij} + (\beta_0 + \beta_{ij})w_{ij}] = \sum_i \alpha_i k_i + \sum_{i<j} \beta_{ij} w_{ij} + \beta_0 W_1 \end{aligned}$$

that leads to the distribution

$$\begin{aligned} Q(\mathbf{W}) &= \prod_{i<j} q_{ij}(w_{ij}) \\ &= \prod_{i<j} \frac{(x_i x_j)^{a_{ij}} e^{-(\beta_0+\beta_{ij})w_{ij}}}{Z_{ij}} = \prod_{i<j} \frac{(x_i x_j)^{a_{ij}} e^{-(\beta_0+\beta_{ij})w_{ij}}}{1 + x_i x_j (\beta_0 + \beta_{ij})^{-1}} \end{aligned} \tag{3.54}$$

where $x_i \equiv e^{-\alpha_i}$. The generic node pair-specific distribution induces a probability for nodes i and j to be connected reading

$$p_{ij} = 1 - q_{ij}(0) = \frac{x_i x_j (\beta_0 + \beta_{ij})^{-1}}{1 + x_i x_j (\beta_0 + \beta_{ij})^{-1}} = \frac{x_i x_j \zeta_{ij}}{1 + x_i x_j \zeta_{ij}}; \tag{3.55}$$

besides, the corresponding expected weight reads

$$\langle w_{ij} \rangle = \frac{p_{ij}}{\beta_0 + \beta_{ij}}. \tag{3.56}$$

Equation 3.55 clarifies why the models considered in the present section are classified as 'integrated': each node pair-specific probability of connection is a function of the parameters controlling for both topological *and* weighted properties. Models of the kind are, thus, capable of 'integrating' information concerning a network structure with information concerning its weights, hence employing them in a joint fashion to define both inference steps.

The recipe for the econometric reparametrization of the integrated exponential model can read as the one of its conditional counterpart, i.e.

$$\beta_{ij} \equiv z_{ij}^{-1} \tag{3.57}$$

a position that turns Eq. 3.56 into

$$\langle w_{ij} \rangle = \frac{p_{ij}}{\beta_0 + z_{ij}^{-1}} \tag{3.58}$$

where

$$p_{ij} = \frac{x_i x_j}{x_i x_j + \beta_0 + z_{ij}^{-1}} \tag{3.59}$$

(see also Appendix B.2).

3.3 Results

The performance of the two classes of models in reproducing the topological properties of the WTW has been tested on the two, usual datasets, i.e. the Gleditsch one (covering 11 years, from 1990 to 2000 [38]) and the BACI one (covering 11 years, from 2007 to 2017 [75]).

To carry out our analyses, we have sampled the ensemble induced by each model as follows. First, the presence of a link connecting any two nodes i and j is established with probability p_{ij}; numerically, this is

realized by drawing a real number u_{ij} from $U(0,1)$, i.e. the uniform distribution with unit support, and comparing it with p_{ij}: if $u \leq p_{ij}$, then i and j are linked, otherwise they are not. Once the presence of a link is established, it is loaded with a weight by employing the inverse transform sampling technique: a second random variable η, uniformly distributed between 0 and 1, is set equal to the value of the complementary cumulative distribution function

$$F(v_{ij}) = \int_0^{v_{ij}} q(w_{ij}|a_{ij} = 1)dw_{ij}; \tag{3.60}$$

then, inverting the equation $F(v_{ij}) = \eta$, one obtains the value of the random variable v_{ij} to be assigned as a link weight to the pair i, j (with $i < j$). Each ensemble is repeatedly sampled in order to obtain 10^4 configurations. The error accompanying the estimate of any quantity of interest is quantified via the confidence intervals (CI) induced by the ensemble distribution of the quantity itself.

3.3.1 Model selection via performance indicators

Let us consider two measures of goodness-of-fit, i.e. the Kolmogorov-Smirnov (KS) compatibility frequency f_m^s and the reconstruction accuracy RA_m^s, already defined in the previous chapter.

The network statistics for which the values RA_m^s and f_m^s have been computed are the degree sequence

$$k_i = \sum_{j(\neq i)=1}^{N} a_{ij}, \quad \forall\, i \tag{3.61}$$

(which gives information about the tendency of node i to connect to other trade partners), the average nearest neighbors degree

$$k_i^{nn} = \frac{\sum_{j(\neq i)=1}^{N} a_{ij} k_j}{k_i}, \quad \forall\, i \tag{3.62}$$

the clustering coefficient

$$c_i = \frac{\sum_{j(\neq i)=1}^{N} \sum_{k(\neq i,j)=1}^{N} a_{ij} a_{jk} a_{ki}}{k_i(k_i - 1)}, \quad \forall\, i; \tag{3.63}$$

for what concerns the weighted statistics, we have considered the strength sequence

$$s_i = \sum_{j(\neq i)=1}^{N} w_{ij}, \quad \forall\, i \tag{3.64}$$

(which gives information about the trade flow of a country), the average nearest neighbors strength

$$s_i^{nn} = \frac{\sum_{j(\neq i)=1}^{N} a_{ij} s_j}{k_i}, \quad \forall\, i \tag{3.65}$$

the weighted clustering coefficient

$$c_i^w = \frac{\sum_{j(\neq i)=1}^{N} \sum_{k(\neq i,j)=1}^{N} w_{ij} w_{jk} w_{ki}}{k_i(k_i - 1)}, \quad \forall\, i. \tag{3.66}$$

KS compatibility frequency

Table 7 lists the values of f_m^s for both binary and weighted network statistics. For what concerns the binary statistics, we report the performance of three, different models, i.e. the UBCM, the FM and the integrated exponential model (denoted as I-Exp).

For what concerns the Gleditsch dataset, compatibility is observed for every year; for what concerns the BACI dataset, instead, this is no longer true: in fact, the FM outputs predictions that are not compatible with the empirical values for a large number of years and irrespectively from the considered quantity; the UBCM and the I-Exp (i.e. the models constraining the degrees), instead, output predictions whose compatibility depends on the considered quantity: higher-order statistics are the ones for which the two aforementioned models 'fail' to the larger extent. Overall, these results lead us to prefer the UBCM as the 'first step-algorithm' of our conditional models.

Let us, now, comment on the performance of our models in reproducing weighted statistics. As it can be appreciated upon looking at Tab. 7, the only models outputting predictions whose distributions are

Dataset	Model	f^{k_i}	$f^{k_{nn}}$	f^{c_i}
Gleditsch	I-Exp	1	1	1
Gleditsch	UBCM	1	1	1
Gleditsch	FM	1	1	1
BACI	I-Exp	1	0.09	0.09
BACI	UBCM	1	0.09	0.09
BACI	FM	0.09	0.09	0.09

Dataset	Model	f^{s_i}	$f^{s_{nn}}$	f^{c_w}
Gleditsch	I-Exp	1	1	0.18
Gleditsch	C-Exp	1	1	0.18
Gleditsch	C-Pareto	0	0	0
Gleditsch	C-Gamma	1	1	0.27
Gleditsch	C-Lognormal	1	0	0
BACI	I-Exp	1	1	1
BACI	C-Exp	1	1	1
BACI	C-Pareto	0	0	0
BACI	C-Gamma	1	1	1
BACI	C-Lognormal	0	0	0

Table 7: Compatibility between the distributions of the expected values of the statistics output by the models considered in the present work and their empirical counterparts for the Gleditsch (left) and the BACI (right) dataset. A large value of f_m^s indicates a large percentage of years for which the distribution of the network statistic predicted by model m is compatible with the empirical one. While all models seem to perform quite well in reproducing the binary statistics on the Gleditsch dataset, this is no longer true when considering the BACI dataset, on which the UBCM outperform the FM - a result that leads us to prefer the former as the 'first step-algorithm' of our conditional models. For what concerns the set of weighted statistics, the models constraining W_1 (i.e. the integrated exponential model, the conditional exponential model and the conditional gamma model) clearly outperform the others.

compatible with the empirical analogues are the integrated exponential one, the conditional exponential one and the conditional gamma one. On the other hand, employing only logarithmic constraints (as for the conditional Pareto model and the conditional log-normal model) does not help improving the accuracy of the description of the system at hand.

Reconstruction accuracy

So far, we have inspected the compatibility of the distributions of the empirical values of each network statistics with the ones of their expected values under each of our models. Let us, now, quantify the extent to which each model is able to recover node-wise information by computing the RA_m^s values.

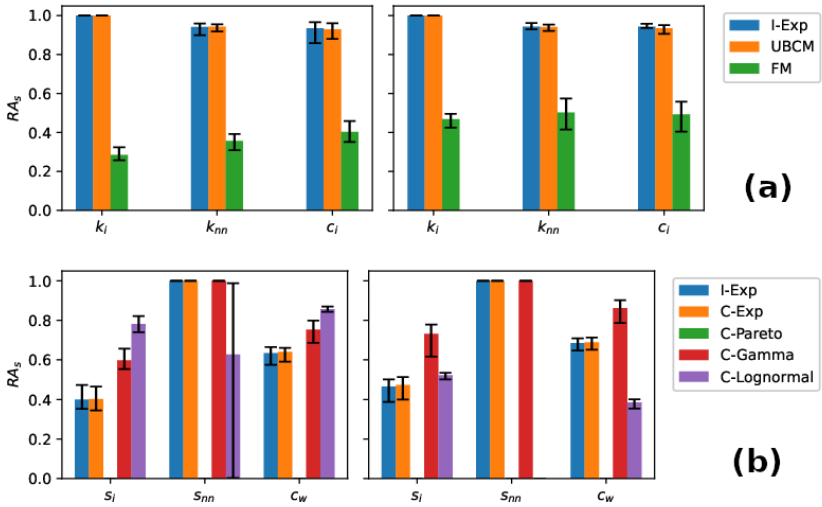

Figure 9: Reconstruction accuracy for the (a) binary and (b) weighted network statistics, calculated as the percentage of node-specific values of the statistics s falling within the CI, induced by model m, at the significance level of 5%; yearly percentages are, then, averaged. The whiskers represent the 2.5 and 97.5 percentiles of each RA_m^s distribution, across different years. Overall, all models perform quite well in reproducing the degrees, with the only exception of the FM - whence our choice of employing the UBCM as the 'first step-algorithm' of our conditional models. For what concerns higher-order, binary statistics, both the UBCM and the I-Exp perform quite well while the performance of the FM is much poorer - a result that holds true for both the Gleditsch (left panels) and the BACI (right panels) dataset. For what concerns the weighted statistics, only the average nearest neighbors strength is satisfactorily recovered - however, only by the (integrated and conditional) exponential models and the conditional gamma one.

Figure 9 shows the temporal average of the latter ones (i.e. across the years covered by our datasets), with the whiskers representing their variation, i.e. an indication of the stability of each model performance. For what concerns the binary statistics (see Fig. 9(a)), both the UBCM and the I-Exp perform quite well in reproducing them; on the other hand, the performance of the FM is much poorer. For what concerns the weighted statistics (see Fig. 9(b)), only the average nearest neighbors strength is satisfactorily recovered by the (integrated and conditional) exponential models and the conditional gamma one. Still, they are found to perform poorly on the other statistics, i.e. the strength, that is only recovered in distribution on both datasets, and the weighted clustering coefficient, that is only recovered in distribution on the BACI dataset. For what concerns the lognormal model, it performs better than competitors in reproducing the strength sequence and the weighted clustering coefficient on the Gleditsch dataset but worse than them on the BACI dataset, causing its behavior to be dataset-dependent.

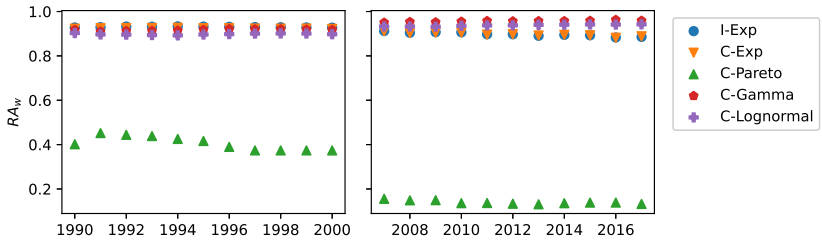

Figure 10: Reconstruction accuracy RA_m^w for the Gleditsch dataset (left panel) and the BACI dataset (right panel). All models perform quite well in reproducing the weights (across all years, on both datasets) with the only exception of the conditional Pareto model. Overall, the best-performing model on the Gleditsch dataset is the integrated exponential while the best-performing model on the BACI dataset is the conditional gamma model.

Finally, let us inspect the reconstruction accuracy of our models when tested on link weights. Specifically, let us consider RA_m^w, i.e. the percentage of empirical weights falling within the CIs, induced by model m, at the significance level of 5% [63]: as Fig. 10 shows, all models perform

quite well in reproducing the weights (across all years, on both datasets) with the only exception of the conditional Pareto model. Overall, the best-performing model on the Gleditsch dataset is the integrated exponential one while the best-performing model on the BACI dataset is the conditional gamma model.

Confusion matrix

The UBCM and the integrated exponential model perform similarly in reproducing the binary statistics, on both datasets. Let us, now, compare them in reproducing the four indicators composing the so-called confusion matrix, i.e. the true positive rate $\langle \text{TPR} \rangle = \langle \text{TP} \rangle / L = \sum_{i<j} a_{ij} p_{ij} / L$ (measuring the percentage of links correctly recovered by a given reconstruction method), the specificity $\langle \text{SPC} \rangle = \langle \text{TN} \rangle / (N(N-1)/2 - L) = \sum_{i<j} (1 - a_{ij})(1 - p_{ij}) / (N(N-1)/2 - L)$ (measuring the percentage of zeros correctly recovered by a given reconstruction method), the positive predictive value $\langle \text{PPV} \rangle = \langle \text{TP} \rangle / \langle L \rangle = \sum_{i<j} a_{ij} p_{ij} / \langle L \rangle$ (measuring the percentage of links correctly recovered by a given reconstruction method with respect to the total number of links predicted by it) and the accuracy $\langle \text{ACC} \rangle = (\langle \text{TP} \rangle + \langle \text{TN} \rangle) / N(N-1)/2$ (measuring the overall performance of a given reconstruction method in correctly placing both links and zeros).

The results are reported in Tab. 8, that shows the increments of the four indicators, defined as $\Delta_X = \langle X \rangle_{\text{I-Exp}} - \langle X \rangle_{\text{UBCM}}$ with $X = \text{TPR}$, SPC, PPV, ACC. Notice that each entry of the table is positive, a result signalling that the integrated exponential model steadily performs better than the UBCM. This is further confirmed by the (non-parametric) Wilcoxon rank-sum test on the ensemble distributions of the statistics to compare: all increments are significant, at the 1% level.

3.3.2 Model selection via statistical tests

Let us now rigorously test if constraining the entire degree sequence leads to a significantly better description of our data than that obtainable by just constraining the total number of links.

Upon solving the model constraining the entire degree sequence and the one constraining the total number of links, we are able to construct a vector reading (X_i^m, Y_i^m) where X_i^m is either RA_m^s or f_m^s for the i-th statistics under the 'L-constrained version' of model m; on the other hand, Y_i^m is either RA_m^s or f_m^s for the i-th statistics under the 'k-constrained version' of model m - naturally, both values have been considered for the same year, keeping the same set of weighted constraints. Pairing statistics as described above allows us to employ the (non-parametric) Wilcoxon signed-rank test for testing the hypotheses $\mathrm{RA}_k^s \leq \mathrm{RA}_L^s$ and $f_k^s \leq f_L^s$, i.e. that the models just constraining L perform better, in reproducing the statistics s, than those constraining the entire degree sequence.

Our results let us conclude that, for both datasets, constraining the degree sequence leads to a significant improvement, at the level of 5%, of the reconstruction accuracy of the average nearest neighbors degree, the clustering coefficient, the strength and the average nearest neighbors strength; on the other hand, constraining the degree sequence does not lead to any significant improvement of the reconstruction accuracy of the weighted clustering coefficient. For what concerns the KS compatibility frequency, a significant improvement, at the level of 5%, is observed in the description accuracy of the average nearest neighbors degree, the clustering coefficient and the average nearest neighbors strength.

3.3.3 Model selection via information criteria

Let us, now, compare the performance of our models in a more general fashion. To this aim, let us consider AIC [86], reading $\mathrm{AIC}_m = 2k - 2\mathcal{L}_m$ where k is the number of free parameters of model m and $\mathcal{L}_m$ is its log-likelihood, evaluated at its maximum.

The purely binary log-likelihood induced by model m is readily obtained from Eq. (2.16) and reads

$$\mathcal{L}_m^{(b)} = \ln P(\mathbf{A}) = \sum_{i<j} \left[a_{ij} \ln p_{ij} + (1 - a_{ij}) \ln(1 - p_{ij})\right] \tag{3.67}$$

where a_{ij} is the generic entry of the empirical adjacency matrix and p_{ij} is

Dataset	Δ_{TPR}	Δ_{SPC}	Δ_{PPV}	Δ_{ACC}
Gleditsch 90	0.017	0.026	0.017	0.020
Gleditsch 91	0.016	0.020	0.016	0.018
Gleditsch 92	0.015	0.019	0.015	0.017
Gleditsch 93	0.015	0.019	0.015	0.017
Gleditsch 94	0.013	0.018	0.013	0.015
Gleditsch 95	0.013	0.019	0.013	0.016
Gleditsch 96	0.012	0.019	0.012	0.015
Gleditsch 97	0.013	0.022	0.013	0.016
Gleditsch 98	0.013	0.022	0.013	0.016
Gleditsch 99	0.013	0.023	0.014	0.017
Gleditsch 00	0.014	0.023	0.014	0.017

Dataset	Δ_{TPR}	Δ_{SPC}	Δ_{PPV}	Δ_{ACC}
BACI 07	0.007	0.037	0.007	0.012
BACI 08	0.006	0.035	0.006	0.010
BACI 09	0.005	0.031	0.005	0.009
BACI 10	0.005	0.032	0.005	0.009
BACI 11	0.005	0.029	0.006	0.008
BACI 12	0.005	0.033	0.005	0.009
BACI 13	0.005	0.034	0.005	0.009
BACI 14	0.005	0.031	0.005	0.008
BACI 15	0.005	0.028	0.005	0.008
BACI 16	0.003	0.021	0.004	0.006
BACI 17	0.004	0.028	0.004	0.007

Table 8: Increments of the four indicators composing the confusion matrix, i.e. the true positive rate (TPR), the specificity (SPC), the positive predicted value (PPV) and the accuracy (ACC) when passing from the UBCM to the integrated exponential model for the Gleditsch (left table) and the BACI (right table) datasets. All increments are significant at the 1% level, according to the (non-parametric) Wilcoxon rank-sum test on the ensemble distributions of the statistics to compare.

the model-dependent probability that node i and node j establish a connection. The 'binary' AIC values (normalized by the yearly maximum, across models, for better visualization) are reported in Fig. 11(a): the integrated exponential model outperforms the others, across all years, for both datasets. This result suggests that the information gained by in-

cluding economic factors into the connection probabilities predicted by it does not affect the parsimony of its description, allowing it to perform better than the UBCM.

When, instead, the 'full' log-likelihood is considered, reading

$$\mathcal{L}^{(f)}_{\text{I-}m} = \ln Q(\mathbf{W}) = \sum_{i<j} \ln q_{ij}(w_{ij}) \tag{3.68}$$

for integrated models and

$$\begin{aligned}\mathcal{L}^{(f)}_{\text{C-}m} &= \ln P(\mathbf{A}) + \ln Q(\mathbf{W}|\mathbf{A}) \\ &= \sum_{i<j} [a_{ij} \ln p_{ij} + (1 - a_{ij}) \ln(1 - p_{ij}) + \ln q_{ij}(w_{ij}|a_{ij})]\end{aligned} \tag{3.69}$$

for conditional models (see Fig. 11(b)), the conditional log-normal and gamma models compete, outperforming the other ones - although the performance of the first one in predicting the network statistics of interest, on the BACI dataset, is less remarkable than that of the competing models (see Fig. 2b).

3.3.4 The Shannon-Fisher plane

We, now, complement the analysis of our models performance, in terms of realized likelihood, with an investigation of our models 'sensitivity', in terms of the variability of the likelihood across network configurations sampled from the model. To this end, for each conditional model we build the so-called Shannon-Fisher plane [87], a technique that has acquired some popularity in the study of time-series - for instance it has been employed to understand ordinal patterns [88], quantify the degree of stochasticity [89], classify financial stock markets [90] and build indicators of economic efficiency [91].

Within our context, we can use the Shannon-Fisher technique to project a given model onto a plane by assigning two coordinates to each connected dyad, i.e. to each pair of nodes such that $a_{ij} = 1$, where a_{ij} is taken from the empirical adjacency matrix of the network. The y coordinate in the plane is Shannon entropy

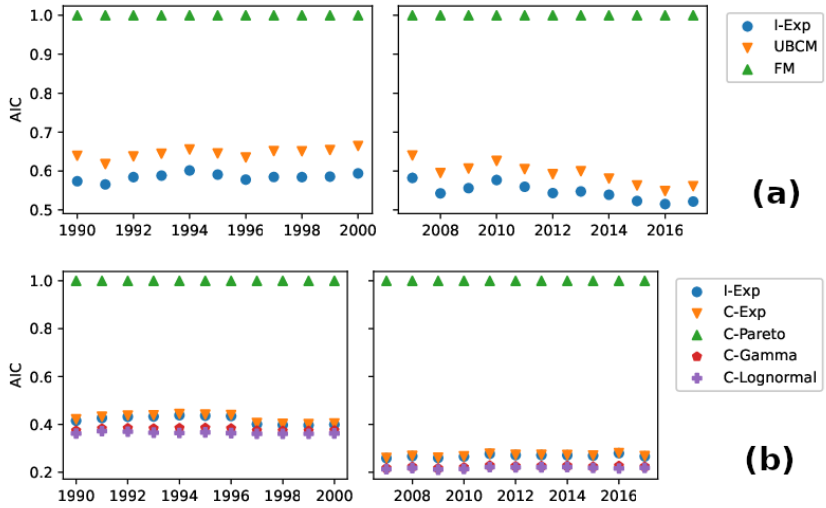

Figure 11: AIC values for the (a) binary and the (b) 'full' log-likelihood, normalized by the yearly maximum, across models, for better visualization, for the Gleditsch (left panels) and the BACI (right panel) datasets: a lower AIC value is associated to a better performance. Our plots clearly show that constraining the degree sequence increases a model performance in reproducing a network topology: moreover, the integrated exponential model steadily outperforms the UBCM, signalling that the information gain due to inclusion of economic variables does not affect the parsimony of its description. For what concerns the capability of our models in reproducing weighted properties, the conditional log-normal and gamma models compete, outperforming the other ones.

$$S_{ij} = -\int q_{ij}(w|a_{ij=1})\,\ln q_{ij}(w|a_{ij}=1)dw$$
$$= \int q_{ij}(w|a_{ij}=1)[H_{ij}(w)+\ln\zeta_{ij}]dw = \langle H_{ij}\rangle + \ln\zeta_{ij}, \qquad (3.70)$$

which quantifies the degree of uncertainty encoded into the link weight. Note that, since entropy is constructed from a continuous pdf, it can attain negative values: this is a well-known problem that can be regularized by introducing the KL divergence with respect to a continuous, uniform pdf; however the result will only consist in an overall shift and

rescaling of the y coordinate that are inessential for our discussion.

The x coordinate in the plane is the Fisher Information Measure (FIM), defined as

$$F_{ij} = \int q_{ij}(w|a_{ij}=1)\left(\frac{\partial \ln\, q_{ij}(w|a_{ij}=1)}{\partial w}\right)^2 dw \tag{3.71}$$

$$= \int q_{ij}(w|a_{ij}=1)\left(\frac{\partial[-H_{ij}(w)-\ln\zeta_{ij}]}{\partial w}\right)^2 dw \tag{3.72}$$

$$= \int q_{ij}(w|a_{ij}=1)\left(\frac{\partial H_{ij}(w)}{\partial w}\right)^2 dw \tag{3.73}$$

$$= \int q_{ij}(w|a_{ij}=1)\left(H'_{ij}(w)\right)^2 dw = \langle (H'_{ij})^2\rangle \tag{3.74}$$

and quantifying the (average) change in probability induced by small changes in the value of the link weight. In other words, the expression $F_{ij} = \langle (H'_{ij})^2\rangle$ captures the 'sensitivity' of the dyadic probability distribution with respect to small changes in the corresponding random variable[3]. Notice that this sensitivity is not captured by Shannon entropy which is indifferent to any reordering of the values of the random variable, provided each value retains its probability.

As different, connected dyads are described by probability distributions with different parameters, scattering all connected dyads in the plane provides an overall representation of the model identified by $q_{ij}(w|a_{ij}=1)$: different models are described by different probability distributions, hence have different projections in the Shannon-Fisher plane (see Appendix D for the explicit values of S_{ij} and F_{ij} for all the conditional models considered here).

As Fig. 12 shows, both the conditional exponential model and the conditional log-normal model follow a decreasing pattern; on average, however, the conditional exponential model is characterized by a smaller FIM, i.e. a smaller 'sensitivity' to variations of the related random variable. On the other hand, the conditional Pareto model collapses onto a single point while the conditional gamma model is characterized by a

[3]Notice that the presence of the derivative requires $q_{ij}(w|a_{ij}=1)$ to be continuous throughout the domain of integration: this is why we consider only conditional models with $a_{ij}=1$ so that there is no 'jump' in the unconditional $q_{ij}(w)$ from $w=0$ to $w>0$.

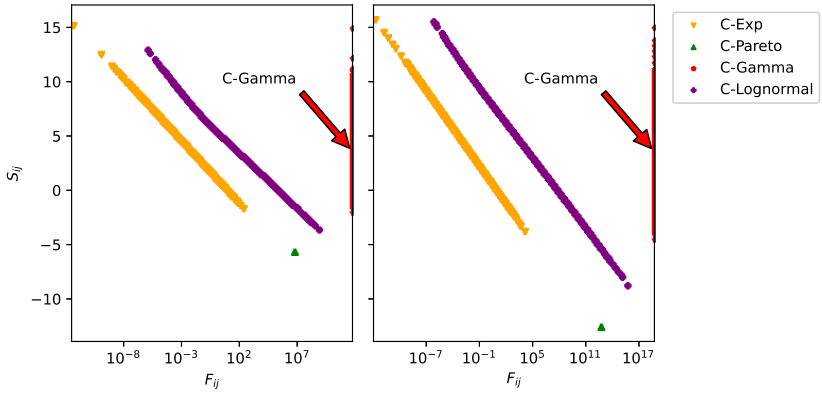

Figure 12: Shannon-Fisher plane for each conditional model considered in the present chapter. For each pair of nodes (i, j) that are connected in the real network ($a_{ij} = 1$), we consider the conditional weight distribution $q_{ij}(w|a_{ij} = 1)$ and plot the corresponding differential Shannon entropy (y-axis) versus the Fisher Information Measure (x-axis). The results shown correspond to the year 2000 of the Gleditsch dataset (left panel) and to the year 2017 of the BACI dataset (right panel). The dyads of both the conditional exponential and the conditional log-normal model follow a decreasing trend, those of the conditional Pareto model collapse onto a single point, while those of the gamma model are characterized by a diverging Fisher Information Measure independently of their entropy (symbolically depicted as a vertical line at the right edge of the plot). The results show that, while different models (except the Pareto) produce similar values of the entropy, their Fisher measure can be very different (note the logarithmic scale of the x-axis). The conditional exponential is the model for which, for a given value of the entropy, the Fisher measure is the minimum one, corresponding to the minimum variability in likelihood across different sampled configurations.

diverging FIM (because of the divergence of the first two negative moments, see Appendix D).

It is interesting to notice that, if we consider the sum of the y values of all the connected dyads for a given model (a sort of 'area under the curve'), we obtain the Shannon entropy for the entire network, conditional on the empirical, binary adjacency matrix $\mathbf{A}$:

$$S(\overline{Q}) = -\int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) \ln Q(\mathbf{W}|\mathbf{A}) d\mathbf{W} = \sum_{i<j|a_{ij}=1} S_{ij} \tag{3.75}$$

(notice that the dyadic entropy of $q(w_{ij}|a_{ij} = 0)$ is zero because the value $a_{ij} = 0$ leads to the value $w_{ij} = 0$ deterministically).

The above expression also coincides with 'minus' the average likelihood of the configurations sampled from the model, hence providing an (inverse) average 'goodness of fit' of the weighted model. Similarly, summing the x values of all the connected dyads gives an overall value of the FIM, hence the average change of the likelihood of the configurations sampled from the model. Therefore, the results shown in Fig. 12 indicate that while different models (except the Pareto) are characterized by similar values of the overall 'goodness of fit', the conditional exponential model attains the (overall) minimum FIM, thereby producing the most stable outcome, in terms of the likelihood of the realized configurations, when used to sample weighted networks.

3.4 Discussion

The analysis carried out in the previous chapter led the ZIP model to be identified as the one performing best among the econometric ones; still, it was also found to be largely disfavored by information criteria such as AIC and BIC. This dilemma was solved upon looking at a different class of statistical models, i.e. the physics-inspired ones: the latter have been found to outperform the purely econometric models for reconstruction purposes, the reason lying in the higher accuracy achieved by them in estimating the topological structure of networks.

Here, we extend the work carried out there by, first, introducing models to infer the topology and the weights of (undirected, weighted) networks defined by continuous-valued data and, then, turning them into proper, econometric ones: in order to do so, we present a theoretical, physics-inspired framework based upon the constrained minimization of the KL divergence - hence, implementing the Minimum Discrimination Information Principle - and capable of accommodating both inte-

grated and conditional (continuous) models.

The main difference between the models belonging to these classes lies in the way the estimation of the topology is carried out; while conditional models disentangle the purely binary step from the (conditional) weighted one, integrated models do not, letting both topological and weighted constraints determine all relevant, structural features of a network. An example of integrated models is provided by the UECM, defined by constraints such as the degree and the strength sequences and described by a mixed Bernoulli-geometric [56, 53] (also called Bose-Fermi [54]) distribution; examples of conditional models are provided by the CReM_A and the CReM_B [63]. From a more econometric perspective, hurdle models are conditional in nature while zero-inflated models can be thought as integrated, the estimation steps being carried out by selecting a distribution out of a basket of available ones.

Our analysis leads to several conclusions: 1) constraining the entire degree sequence leads to a statistically significant improvement in the reconstruction accuracy of the WTW. In particular, the integrated exponential model, described by the Hamiltonian $H(\mathbf{W}) = \sum_i \alpha_i k_i + \sum_{i<j} \beta_{ij} w_{ij} + \beta_0 W_1$, provides a very accurate, structural reconstruction while being favored by information criteria: although it is defined by $N + 1$ purely topological constraints, AIC reveals them as 'irreducible', i.e. necessary to provide a satisfactory explanation of the network generating process; 2) when considering weighted quantities, the conditional gamma model is the one performing best (although it competes with the integrated exponential one in reproducing properties such as the weights, on some of the temporal snapshots covered by our datasets), according to information criteria. To be noticed, however, that if strengths are not explicitly constrained - jointly with the degrees - maximum-entropy models recover them only 'in distribution' while failing to reproduce their exact values. The same consideration holds true for the weighted clustering coefficient.

Coming to comparing the models belonging to the classes considered in the present work, the two, best-performing ones are the integrated exponential model and the conditional gamma model, i.e. the ones con-

straining the total weight (although the conditional exponential model constrains the total weight as well, it is outperformed by the conditional gamma one within the class of conditional model): hence, W_1 seems to constitute a somehow fundamental quantity to be necessarily accounted for in order to achieve a good reconstruction accuracy. From an economic point of view, the parameter β_0 constraining the total weight can be interpreted as a sort of 'shadow price' to be paid by everyone to exchange goods.

Additional information is provided by our analysis of the Shannon-Fisher plane, which combines Shannon entropy, i.e. the (inverse) likelihood of a model, with the Fisher Information Measure, i.e. the average variability of the likelihood itself across different, sampled configurations: the conditional exponential model turns out the be the least variable, hence the most stable. It is worth noticing that our maximum-entropy approach is formulated for canonical ensembles, i.e. for 'soft constraints', which implies that different realizations of the network have fluctuating values of the weighted sufficient statistics: these fluctuations 'originate' the FIM; by contrast, if we were to formulate microcanonical models with 'hard constraints', the sufficient statistics would not fluctuate and the overall FIM would be zero. Therefore, the Shannon-Fisher plane shows that, among the canonical models considered here, the conditional exponential one is the closest to the 'least soft' extreme while the conditional gamma lies at the opposite 'softest' extreme, i.e. where the FIM diverges. As a question left for future research, it would be interesting to relate the behavior of the FIM to the phenomenon of ensemble non-equivalence [92].

Overall, we believe the framework proposed in this contribution to have the potential of reconciling the approach adopted by network scientists for reconstructing economic networks, and focusing on the purely structural aspects of a network formation, with the approach characterizing econometrics, tailored to inform these same models with macroeconomic quantities. From an operative point of view, our (classes of) models combine the pros of both approaches: the importance of purely structural information (highlighted by physics-inspired models) can be

Model	p_{ij}	$\langle w_{ij} \vert a_{ij}=1 \rangle$
I-Exp	$\frac{x_i x_j}{x_i x_j + \beta_0 + z_{ij}^{-1}}$	$\frac{z_{ij}}{1+\beta_0 z_{ij}}$
C-Exp	$\frac{x_i x_j}{1+x_i x_j}$	$\frac{z_{ij}}{1+\beta_0 z_{ij}}$
C-Gamma	$\frac{x_i x_j}{1+x_i x_j}$	$\frac{(1-\xi_0) z_{ij}}{1+\beta_0}$
C-Lognormal	$\frac{x_i x_j}{1+x_i x_j}$	$\exp\left\{\frac{3-2\xi_{ij}}{4\gamma_0}\right\}$
C-Pareto	$\frac{x_i x_j}{1+x_i x_j}$	$(1+z_{ij})w_{min}$

Table 9: Models used in this chapter with columns indicating the model-induced probability connection p_{ij} and the weight conditional on the link presence $\langle w_{ij}|a_{ij} = 1\rangle$. The parameter $z_{ij} = \exp\{\rho + \beta \ln(\omega_i \omega_j) + \gamma \ln(D_{ij})\}$ stands for the log-link gravity specification, while $\hat{x}_{ij} = \exp\{\theta + \ln(\omega_i \omega_j) - \ln(D_{ij})\}$ stands for the gravity specification used in the logistic binary model.

accounted for by constraining the entire degree sequence; on top of that, a second step is needed to estimate a network weighted structure: although the information provided by the total weight cannot be discarded without affecting the overall performance of a model, such an estimation can rests upon purely econometric considerations.

In Table 9 we sum up the weighted models explored in this chapter with the specifications related to the model-induced probability connection p_{ij} and to the conditional weight $\langle w_{ij}|a_{ij} = 1\rangle$ (conditional on the established link), while in Table 10 the relative structural constraints are portrayed for each model. In this last table, objects the "hat" apex is used to identify the constrained quantities that are approximated with a log-link function in terms of constant and economic covariates.

3.5 The DyGyS Python package

As an additional result, we release a Python package named DyGyS (an acronym stating for 'DYadic GravitY regression models with Soft constraints'), available at `https://github.com/MarsMDK/DyGyS`: it contains the routines to implement all the models discussed in the present chapter as well as those considered in the previous one.

Model	Type	Binary Constraints	Weighted Constraints
I-Exp	I-Cont	$\{k_i\}$	$W_{tot}, \hat{w}_{ij}$
C-Exp	C-Cont	$\{k_i\}$	$W_{tot}, \hat{w}_{ij}$
C-Gamma	C-Cont	$\{k_i\}$	$W_{tot}, \sum_{j<i} \ln(w_{ij}), \hat{w}_{ij}$
C-Lognormal	C-Cont	$\{k_i\}$	$\sum_{j<i} \ln^2(w_{ij}), \ln(\hat{w})_{ij}$
C-Pareto	C-Cont	$\{k_i\}$	$\ln(\hat{w})_{ij}$

Table 10: Entropy-based models used in this chapter with columns indicating the type of induced weights, relative binary constraints and weighted constraints. The type is annotated with C or I for conditional and integrated model respectively, and with Int and Cont for models inducing integer and continuous link weights. Constraints can be topological quantities or their approximation through economic factors. When an approximation in terms of a gravity log-link function is performed, it is indicated with the hat symbol.

Chapter 4

Beyond the deterministic estimation of parameters in conditional models

This chapter illustrates the results of the analysis documented in [3]. Here, we deal with the problem of estimating parameters of conditional models. Econometric recipes prescribe to carry out an optimization procedure treating the topology as deterministic, even when it is the outcome of a probabilistic model. In order to overcome the limitations of this approach, we devise two, different procedures (hereby named 'annealed' and 'quenched') that correspond to alternative ways of averaging over the topological randomness. For models whose weighted structure is homogeneous, the three procedures are equivalent, independently from the binary recipe; for models whose weighted structure is heterogeneous, the 'annealed' and 'quenched' approaches lead to the same estimate, which is often incompatible with the 'deterministic' one. When the considered configurations are sparse, however, the 'quenched' approach may lead to a biased estimation of node-specific parameters, a circumstance leading us to prefer the 'annealed' one (which is also the most convenient from a numerical perspective).

4.1 Introduction

Simultaneously modelling the establishment of a connection and the corresponding weight poses a serious challenge. Econometrics prescribes to estimate binary and weighted parameters either separately, within the context of hurdle models [70], or jointly, within the context of zero-inflated models [43]; in both cases, the GM specification [25] $\langle w_{ij} \rangle_{\mathrm{GM}} = f(\omega_i, \omega_j, d_{ij}|\underline{\phi}) = \rho(\omega_i \omega_j)^{\alpha} d_{ij}^{\gamma}$, where $\omega_i \equiv \mathrm{GDP}_i/\overline{\mathrm{GDP}}$ is the GDP of country i divided by the arithmetic mean of the GDPs of all countries, d_{ij} is the geographic distance between the capitals of countries i and j and $\underline{\phi} \equiv (\rho, \alpha, \gamma)$ is a vector of parameters, is interpreted as the expected value of a probability distribution whose functional form is arbitrary. On the other hand, the approach rooted in statistical physics constructs maximum-entropy distributions, constrained to satisfy certain network properties [82, 59, 93, 58, 62].

In chapter 2 and chapter 3 these approaches have been integrated within the framework induced by the constrained optimization of the KL divergence [84]: in particular, two, broad classes of models have been constructed, i.e. the integrated and conditional ones, defined by different, probabilistic rules to place links and load them with weights. For what concerns integrated models, they follow from a single, constrained optimization of the KL divergence [54]; for what concerns conditional models, they are disentangled and the functional form of the weight distribution follows from a conditional, optimization procedure [63]. Still, the prescriptions adopted by the two approaches to carry out the estimation of the parameters entering into the definition of each model differ.

4.2 Minimization of the KL divergence

As already discussed in section 3.2.1, the functional form of continuous, conditional network models can be identified through the constrained minimization of the KL divergence of a distribution Q from a prior distribution R, i.e.

$$D_{\mathrm{KL}}(Q||R) = \int_{\mathbb{W}} Q(\mathbf{W}) \ln \frac{Q(\mathbf{W})}{R(\mathbf{W})} d\mathbf{W} \tag{4.1}$$

where $\mathbf{W}$ is one of the possible values of a continuous random variable, $\mathbb{W}$ is the set of possible values that $\mathbf{W}$ can take, $Q(\mathbf{W})$ is the (multivariate) probability density function to be estimated and $R(\mathbf{W})$ plays the role of prior distribution, whose divergence from $Q(\mathbf{W})$ must be minimized: in our setting, $\mathbf{W}$ represents an entire network whose weights, now, obey the property $w_{ij} \in \mathbb{R}_0^+, \forall\, i < j$. In case of uninformative priors, this framework leads to considering the (somehow, simplified) expression

$$-S(Q) = -S(P) - S(\overline{Q}|P) \tag{4.2}$$

i.e. 'minus' the joint entropy, where

$$S(P) = -\sum_{\mathbf{A} \in \mathbb{A}} P(\mathbf{A}) \ln P(\mathbf{A}) \tag{4.3}$$

is the Shannon entropy of the probability distribution describing the binary projection of the network structure [58, 62] and

$$S(\overline{Q}|P) = -\sum_{\mathbf{A} \in \mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) \ln Q(\mathbf{W}|\mathbf{A}) d\mathbf{W} \tag{4.4}$$

is the conditional Shannon entropy of the probability distribution describing the weighted network structure [1, 2, 63]. The functional form of $P(\mathbf{A})$ can be determined by carrying out the usual, constrained maximisation of Shannon entropy [58, 62]; remarkably, any set of (binary) constraints considered in what follows will lead to the same expression for $P(\mathbf{A})$, i.e. $P(\mathbf{A}) = \prod_{i<j} p_{ij}^{a_{ij}} (1-p_{ij})^{1-a_{ij}}$ with $p_{ij} = x_{ij}/(1+x_{ij})$: specifically, the position $x_{ij} \equiv x$ individuates the Undirected Binary Random Graph Model (UBRGM), the position $x_{ij} \equiv x_i x_j$ individuates the Undirected Binary Configuration Model (UBCM) and the position $x_{ij} \equiv \delta\omega_i\omega_j$ individuates the Logit Model (LM) [64].

On the other hand, the functional form of $Q(\mathbf{W}|\mathbf{A})$ can be determined by carrying out the constrained maximization of $S(\overline{Q}|P)$, the set of constraints being, now,

$$1 = \int_{\mathbb{W}_{\mathbf{A}}} P(\mathbf{W}|\mathbf{A}) d\mathbf{W}, \; \forall \, \mathbf{A} \in \mathbb{A}, \tag{4.5}$$

$$\langle C_\alpha \rangle = \sum_{\mathbf{A} \in \mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) C_\alpha(\mathbf{W}) d\mathbf{W}, \; \forall \, \alpha; \tag{4.6}$$

while the first condition ensures the normalization of the probability distribution, the vector $\{C_\alpha(\mathbf{W})\}$ represents the proper set of weighted constraints. The distribution induced by such an optimisation problem reads

$$Q(\mathbf{W}|\mathbf{A}) = \frac{e^{-H(\mathbf{W})}}{Z_{\mathbf{A}}} = \frac{e^{-H(\mathbf{W})}}{\int_{\mathbb{W}_{\mathbf{A}}} e^{-H(\mathbf{W})} d\mathbf{W}} \tag{4.7}$$

if $\mathbf{W} \in \mathbb{W}_{\mathbf{A}}$ and 0 otherwise. While the Hamiltonian $H(\mathbf{W}) = \sum_\alpha \psi_\alpha C_\alpha(\mathbf{W})$ lists the constraints, the quantity at the denominator is the partition function, conditional on the fixed topology $\mathbf{A}$ [63].

For mathematical convenience, we will consider separable Hamiltonians, i.e. functions that can be written as sums of node pair-specific Hamiltonians, i.e. $H(\mathbf{W}) = \sum_{i<j} H_{ij}(w_{ij})$; this choice leads to the result

$$\begin{aligned} Q(\mathbf{W}|\mathbf{A}) &= \frac{e^{-\sum_{i<j} H_{ij}(w_{ij})}}{\int_{\mathbb{W}_{\mathbf{A}}} e^{-\sum_{i<j} H_{ij}(w_{ij})} d\mathbf{W}} \\ &= \prod_{i<j} \frac{e^{-H_{ij}(w_{ij})}}{\left[\int_{m_{ij}}^{+\infty} e^{-H_{ij}(w_{ij})} dw_{ij}\right]^{a_{ij}}} = \prod_{i<j} \frac{e^{-H_{ij}(w_{ij})}}{\zeta_{ij}^{a_{ij}}} \end{aligned} \tag{4.8}$$

(with m_{ij} being the pair-specific, minimum weight allowed by a given model and ζ_{ij} being the corresponding partition function), irrespectively from the specific, functional form of $H_{ij}(w_{ij})$ [2] (see also Appendix C.1).

4.3 Estimation of the parameters

Several, alternative recipes are viable to estimate the parameters entering into the definition of continuous, conditional network models.

4.3.1 'Deterministic' parameter estimation

The simplest one prescribes to consider the traditional likelihood function

$$\ln Q(\mathbf{W}^*) = \ln[P(\mathbf{A}^*)Q(\mathbf{W}^*|\mathbf{A}^*)] = \ln P(\mathbf{A}^*) + \ln Q(\mathbf{W}^*|\mathbf{A}^*) \quad (4.9)$$

with $\mathbf{W}^*$ ($\mathbf{A}^*$) being the empirical, weighted (binary) adjacency matrix; its maximization allows the parameters entering into the definition of the purely topological distribution and those entering into the definition of the conditional, weighted one to be estimated in a totally disentangled fashion [2]. In fact, maximizing

$$\begin{aligned}\mathcal{L}_{\underline{\psi}} &= \ln Q(\mathbf{W}^*|\mathbf{A}^*) = \\ &= -H(\mathbf{W}^*) - \ln Z_{\mathbf{A}^*} = -H(\mathbf{W}^*) - \ln\left[\int_{\mathbb{W}_{\mathbf{A}^*}} e^{-H(\mathbf{W})} d\mathbf{W}\right] \end{aligned} \quad (4.10)$$

with respect to the unknown parameters leads us to find the vector of values $\underline{\psi}^*$ satisfying the vector of relationships

$$\langle \mathbf{C} \rangle_{\mathbf{A}^*}(\underline{\psi}^*) \equiv \mathbf{C}^* \quad (4.11)$$

which stands for the set of relationships $\langle C_\alpha \rangle_{\mathbf{A}^*}(\underline{\psi}^*) \equiv C_\alpha^*, \forall\, \alpha$, each one equating the model-induced, average value of the corresponding constraint to its empirical value, marked with an asterisk.

This first approach to parameter estimation can be named as 'deterministic', to stress that $\mathbf{A}^*$ is considered as not being subject to variation; otherwise stated, this recipe - which is the most common in econometrics - prescribes to estimate the parameters entering into the definition of the conditional, weighted probability distribution by assuming the network topology to be fixed.

4.3.2 'Annealed' parameter estimation

Topology, however, is a random variable itself, obeying the probability distribution $P(\mathbf{A})$. As a consequence, the 'deterministic' recipe for parameter estimation could lead to inconsistencies, should the description of $\mathbf{A}^*$ provided by $P(\mathbf{A})$ be not accurate enough. The variability induced by $P(\mathbf{A})$ can be properly accounted for by considering the generalised likelihood [63]

$$\mathcal{G}_{\underline{\psi}} = \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \ln Q(\mathbf{W}^*|\mathbf{A}) =$$
$$= -H(\mathbf{W}^*) - \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \ln \left[\int_{\mathbb{W}_{\mathbf{A}^*}} e^{-H(\mathbf{W})} d\mathbf{W} \right] = \langle \mathcal{L}_{\underline{\psi}} \rangle \qquad (4.12)$$

whose maximization leads us to find the vector of values $\underline{\psi}^*$ satisfying the vector of relationships

$$\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \langle \mathbf{C} \rangle_{\mathbf{A}} (\underline{\psi}^*) = \langle \mathbf{C} \rangle (\underline{\psi}^*) = \mathbf{C}^* \qquad (4.13)$$

which stands for the set of relationships $\langle C_\alpha \rangle (\underline{\psi}^*) \equiv C^*_\alpha, \forall\ \alpha$. Taking this average is conceptually similar to taking the 'annealed' average in physics: parameter estimation is carried out while random variables - again, the entries of the adjacency matrix - are left to vary.

Interestingly, the 'deterministic' recipe is a special case of the 'annealed' recipe since the former can be recovered by posing $P(\mathbf{A}) \equiv \delta_{\mathbf{A},\mathbf{A}^*}$: in this case, in fact,

$$\mathcal{G}_{\underline{\psi}} = -H(\mathbf{W}^*) - \sum_{\mathbf{A}\in\mathbb{A}} \delta_{\mathbf{A},\mathbf{A}^*} \ln Z_{\mathbf{A}} = -H(\mathbf{W}^*) - \ln Z_{\mathbf{A}^*} = \mathcal{L}_{\underline{\psi}}; \quad (4.14)$$

similarly, $\sum_{\mathbf{A}\in\mathbb{A}} \delta_{\mathbf{A},\mathbf{A}^*} \langle \mathbf{C} \rangle_{\mathbf{A}} (\underline{\psi}^*) = \langle \mathbf{C} \rangle_{\mathbf{A}^*} (\underline{\psi}^*) = \mathbf{C}^*$.

4.3.3 'Quenched' parameter estimation

A viable alternative to properly account for the variability induced by $P(\mathbf{A})$ is that of reversing the two operations of 'likelihood maximization' and 'ensemble averaging': in other words, one can 1) numerically sample the ensemble of configurations induced by $P(\mathbf{A})$, 2) maximize the likelihood $\ln Q(\mathbf{W}^*|\mathbf{A})$ for each, generated network, 3) take the average of the resulting set of parameters, according to the formula

$$\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \underline{\psi}^*(\mathbf{A}) = \langle \underline{\psi}^* \rangle \qquad (4.15)$$

the estimation of the α-th parameter being assumed to coincide with the average $\langle \psi^*_\alpha \rangle$.

Taking this average is conceptually similar to taking the 'quenched' average in physics: random variables - in the specific case, the entries of the adjacency matrix - are frozen, parameter estimation is carried out and, only at the end, the values of the parameters are averaged over the ensemble of configurations induced by $P(\mathbf{A})$.

As our models inherit their functional form from the constrained minimization of the KL divergence, each parameter controls for a specific constraint: when employing the 'deterministic' recipe, such a circumstance makes each parameter configuration-dependent; when employing either the 'annealed' or the 'quenched' recipe, instead, accounting for the variability of a network structure induces a sort of 'loss of memory' about its empirical, purely topological details.

4.4 Results

In order to test if the 'deterministic', 'annealed' and 'quenched' prescriptions lead to the same estimation, let us focus on a number of variants of the Conditional exponential model (CEM), induced by the positions $H_{ij}^{\text{CEM}} = \beta_{ij} w_{ij}$ and $\zeta_{ij}^{\text{CEM}} = \beta_{ij}^{-1}$:

$$Q(\mathbf{W}) = P(\mathbf{A})Q(\mathbf{W}|\mathbf{A}) = \prod_{i<j} p_{ij}^{a_{ij}} (1-p_{ij})^{1-a_{ij}} \prod_{i<j} \beta_{ij}^{a_{ij}} e^{-\beta_{ij} w_{ij}}; \quad (4.16)$$

naturally, $q_{ij}(w_{ij} = 0|a_{ij} = 0) = 1$ (i.e. if nodes i and j are not connected, the weight of the corresponding link is zero with probability equal to one) and $q_{ij}(w_{ij} > 0|a_{ij} = 1) = \beta_{ij} e^{-\beta_{ij} w_{ij}}$.

In what follows, we will consider three, different instances of $p_{ij} = x_{ij}/(1 + x_{ij})$, corresponding to

- the UBRGM, defined by posing $x_{ij} \equiv x$ and induced by the maximization of $S(P)$ while constraining the total number of links, $L(\mathbf{A}^*) \equiv L^* = \sum_{i<j} a_{ij}^*$, i.e.

$$p_{ij}^{\text{UBRGM}} \equiv \frac{x}{1+x}; \quad (4.17)$$

- the UBCM, defined by posing $x_{ij} \equiv x_i x_j$ and induced by the maximization of $S(P)$ while constraining the whole degree sequence, $\{k_i(\mathbf{A}^*)\}_{i=1}^N \equiv \{k_i^*\}_{i=1}^N$ with $k_i^* = \sum_{j(\neq i)} a_{ij}^*$, i.e.

$$p_{ij}^{\text{UBCM}} \equiv \frac{x_i x_j}{1 + x_i x_j}; \tag{4.18}$$

- two, different instances of the LM, both representing a fitness-driven version of the UBCM, (again) induced by constraining the total number of links, $L(\mathbf{A}^*) \equiv L^* = \sum_{i<j} a_{ij}^*$. The first one is defined by posing $x_{ij} \equiv \delta\omega_i\omega_j$, i.e.

$$p_{ij}^{\text{LM}} \equiv \frac{\delta\omega_i\omega_j}{1 + \delta\omega_i\omega_j} \tag{4.19}$$

and has been employed to study the year 2017 of the BACI version of the WTW [75], that is a network of $N = 171$ nodes and a link density of $d = 0.87$. The second one is defined by posing $x_{ij} \equiv \delta s_i s_j$, i.e.

$$p_{ij}^{\text{LM}} = \frac{\delta s_i s_j}{1 + \delta s_i s_j} \tag{4.20}$$

and has been employed to study the 01/03/2019 snapshot of the Bitcoin Lightning Network (BLN) [94], that is a network of $N = 5012$ nodes and a link density of $d = 0.003$.

4.4.1 'Scalar' variant of the CEM

Let us start by considering the 'scalar' or homogeneous variant of the CEM, defined by the position $\beta_{ij} \equiv \beta$, $\forall\, i < j$. In this case, the 'deterministic' recipe for parameter estimation prescribes to maximize the likelihood

$$\mathcal{L}_{\underline{\psi}} = \sum_{i<j} [-\beta w_{ij}^* + a_{ij}^* \ln \beta] = -\beta W^* + L^* \ln \beta \tag{4.21}$$

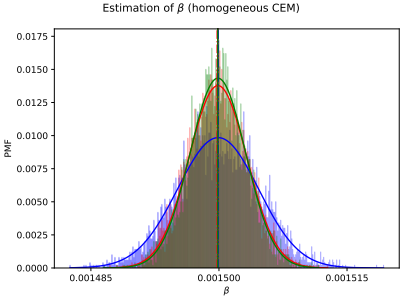

Figure 13: Estimations of the parameter β, entering the definition of the homogeneous version of the CEM, where the binary topology is either 'deterministic' (black) or generated via the UBRGM (blue), the UBCM (green) and the LM (red). The deterministic approach leads to a single estimate, while the other approaches lead to either a single, 'annealed' estimate (vertical, solid lines) or to a whole distribution of 'quenched' estimates (empirical distribution constructed over an ensemble of 5.000 binary configurations with theoretical curves, Binomial or Poisson-Binomial, dependent on the binary model; the corresponding average value is indicated by a vertical, dash-dotted line). The 'annealed' parameter estimates, the average values of the 'quenched' parameter distributions and the 'deterministic' parameter estimate coincide. Data refers to the year 2017 of the BACI version of the WTW [75].

where $W(\mathbf{W}^*) \equiv W^* = \sum_{i<j} w^*_{ij}$ and whose optimization leads to the expression $\beta = L^*/W^*$. The 'annealed' recipe prescribes to maximize the likelihood

$$\mathcal{G}_{\underline{\psi}} = \sum_{i<j} [-\beta w^*_{ij} + p_{ij} \ln \beta] = -\beta W^* + \langle L \rangle \ln \beta \tag{4.22}$$

whose optimization leads to the expression $\beta = \langle L \rangle / W^*$. The 'quenched' recipe, on the other hand, prescribes to calculate the average

$$\langle \beta \rangle = \sum_{\mathbf{A} \in \mathbb{A}} P(\mathbf{A}) \beta(\mathbf{A}) = \sum_{\mathbf{A} \in \mathbb{A}} P(\mathbf{A}) \frac{L(\mathbf{A})}{W^*} = \frac{\langle L \rangle}{W^*} \tag{4.23}$$

since, now, $\beta(\mathbf{A}) = L(\mathbf{A})/W^*$.

In the case of the 'scalar' variant of the CEM, the estimations coincide for any null model preserving the total number of links, i.e. ensur-

ing that $\langle L \rangle = L^*$, regardless of the network density. Such a result is confirmed by Fig. 13 where each recipe has been implemented on the WTW, by adopting the distributions induced by the UBRGM (blue), the UBCM (green) and the LM (red). Specifically, the 'deterministic' estimation (black, solid line) and the 'annealed' estimations (blue, green and red, solid lines) overlap; moreover, each 'annealed' estimation overlaps with the the corresponding, 'quenched' estimation, i.e. the average value of the related, 'quenched' distribution (blue, green and red, dash-dotted lines).

In the case of the UBRGM-induced, homogeneous version of the CEM, the 'quenched' distribution of the parameter $\beta(\mathbf{A}) = L(\mathbf{A})/W^*$ 'inherits' the distribution of the total number of links, i.e. $L \sim \text{Bin}(N(N-1)/2, p)$, with $p = 2L^*/N(N-1)$: more precisely, $W\beta \sim \text{Bin}(N(N-1)/2, p)$; analogously for the UBCM- and the LM-induced, homogeneous versions of the CEM - the only difference being that, now, L obeys two, different, Poisson-Binomial distributions.

4.4.2 'Vector' variant of the CEM

Let us, now, consider the 'vector' or weakly heterogeneous variant of the CEM, defined by the position $\beta_{ij} \equiv \beta_i + \beta_j, \forall\, i < j$. In this case, the 'deterministic' recipe for parameter estimation prescribes to maximize the likelihood

$$\mathcal{L}_{\underline{\psi}} = \sum_{i<j} [-(\beta_i + \beta_j) w^*_{ij} + a^*_{ij} \ln(\beta_i + \beta_j)] = -\sum_i \beta_i s^*_i + \sum_{i<j} a^*_{ij} \ln(\beta_i + \beta_j) \tag{4.24}$$

where $s_i(\mathbf{W}^*) \equiv s^*_i = \sum_{j(\neq i)} w^*_{ij}$ and whose optimization requires to solve the system of equations

$$s^*_i = \sum_{j(\neq i)} \frac{a^*_{ij}}{\beta_i + \beta_j}, \ \forall\, i. \tag{4.25}$$

The 'annealed' recipe, instead, prescribes to maximize the likelihood

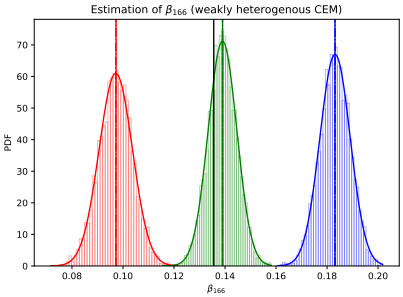

Figure 14: Estimations of the parameter β_{166} entering the definition of the weakly heterogeneous version of the CEM, where the binary topology is either 'deterministic' (black) or generated via the UBRGM (blue), the UBCM (green) and the LM (red). The deterministic approach leads to a single estimate, while the other approaches lead to either a single, 'annealed' estimate (vertical, solid lines) or to a whole distribution of 'quenched' estimates (histograms with normal density curves having the same average and standard deviation, constructed over an ensemble of 5.000 binary configurations; the average value is indicated by a vertical, dash-dotted line). Each 'annealed' parameter estimate coincides with the average value of the corresponding 'quenched' distribution although the distributions induced by the three, binary recipes are well separated. In addition, the 'deterministic' parameter estimate is very close to the UBCM-induced, 'annealed' one. Data refers to the year 2017 of the BACI version of the WTW [75].

$$\mathcal{G}_{\underline{\psi}} = \sum_{i<j}[-(\beta_i+\beta_j)w_{ij}^* + p_{ij}\ln(\beta_i+\beta_j)] = -\sum_i \beta_i s_i^* + \sum_{i<j} p_{ij}\ln(\beta_i+\beta_j) \tag{4.26}$$

whose optimization requires to solve the system of equations

$$s_i^* = \sum_{j(\neq i)} \frac{p_{ij}}{\beta_i+\beta_j},\ \forall\, i \tag{4.27}$$

(notice that both the 'deterministic' and the 'annealed' version of the 'vector' variant of the CEM are alternative instances of the so-called CReM_A, introduced in [63]). The 'quenched' recipe, on the other hand, requires to solve the system of equations $\langle\beta_i\rangle = \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A})\beta_i(\mathbf{A}),\ \forall\, i$ which no longer have an explicit expression. Devising some sort of approximation is, however, possible. Let us start by re-writing Eq. 4.27 as

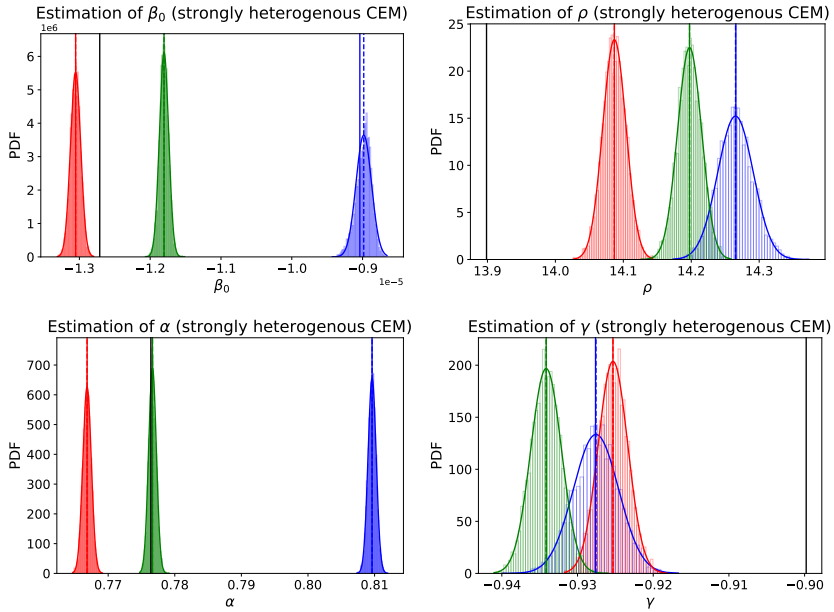

Figure 15: Estimations of the parameters (from top left to bottom right) β_0, ρ, α and γ, entering the definition of the econometric version of the CEM, where the binary topology is either 'deterministic' (black) or generated via the UBRGM (blue), the UBCM (green) and the LM (red). The deterministic approach leads to a single estimate, while the other approaches lead to either a single, 'annealed' estimate (vertical, solid lines) or to a whole distribution of 'quenched' estimates (histograms with kernel density curves, constructed over an ensemble of 5.000 binary configurations; the corresponding average value is indicated by a vertical, dash-dotted line). Each 'annealed' parameter estimate coincides with the average value of the corresponding 'quenched' distribution although the distributions induced by the three, binary recipes may overlap or not; the 'deterministic' estimate, instead, overlaps with the other, two ones only for the parameter α, under the UBCM-induced, binary recipe. Data refers to the year 2017 of the BACI version of the WTW [75].

$$\beta_i = \frac{1}{s_i^*} \sum_{j(\neq i)} \frac{p_{ij}}{1 + \beta_j/\beta_i}, \; \forall \, i \tag{4.28}$$

and consider the node whose coefficient is the largest one. This allows

us to write $\beta_i \simeq \sum_{j(\neq i)} p_{ij}/s_i^* = \langle k_i \rangle / s_i^*$: in case we implemented the UBRGM, we would obtain $\beta_i(\mathbf{A}) \simeq 2L(\mathbf{A})/Ns_i^*$, hence expecting the 'quenched' distribution of $Ns_i^*\beta_i/2$ to coincide with $\text{Bin}(N(N-1)/2, p)$; if, on the other hand, we implemented the UBCM, we would obtain $\beta_i(\mathbf{A}) \propto k_i(\mathbf{A})/s_i^*$, hence expecting the 'quenched' distribution of $s_i^*\beta_i$ to obey a Poisson-Binomial. Again, the estimations coincide for any null model preserving the structural properties characterizing the binary recipe implemented.

More generally, the mutual relationships between the estimations provided by the three recipes are node-dependent (see Fig. 14, illustrating the case-study of node 166 of the WTW and Fig. 22 in Appendix C.2): in general, however, each 'annealed' estimation overlaps with the average value of the related 'quenched' distribution. Moreover, the 'deterministic' estimation is very close to the UBCM-induced, 'annealed' one; such a result is a consequence of the accurate description of the empirical network topology provided by the UBCM - in fact, much more accurate than the ones provided by the UBRGM and the LM: indeed, the better the approximation $p_{ij} \simeq a_{ij}, \forall\, i < j$, the closer the 'annealed' estimation to the 'deterministic' one.

This is even more evident when considering the 'tensor' variant of the CEM, in which case the three optimization procedures lead to the expressions $\beta_{\text{det}} = a_{ij}^*/\hat{w}_{ij}, \forall\, i < j$ and $\beta_{\text{ann}} = \langle \beta \rangle_{\text{que}} = p_{ij}/\hat{w}_{ij}, \forall\, i < j$ - with $\hat{w}_{ij}$ representing an estimate of the empirical weight w_{ij}^*; if, however, $\hat{w}_{ij} \equiv w_{ij}^*, \forall\, i < j$ then, for consistency, $p_{ij} \equiv a_{ij}^*$ and the three recipes coincide.

4.4.3 'Econometric' variant of the CEM

As a third case-study, let us focus on the 'econometric' variant of the CEM, defined by posing $\beta_{ij} \equiv \beta_0 + z_{ij}^{-1}, \forall\, i < j$, where $z_{ij} \equiv e^{\rho}(\omega_i\omega_j)^{\alpha}d_{ij}^{\gamma}$ represents the GM specification traditionally employed to analyze undirected, weighted, trade networks and β_0 is a structural parameter to be tuned in order to ensure that $\langle W \rangle = W^*$. In this case, the 'deterministic' recipe for parameter estimation prescribes to maximize the likelihood

$$\mathcal{L}_{\underline{\psi}} = \sum_{i<j} [-(\beta_0 + z_{ij}^{-1}) w_{ij}^* + a_{ij}^* \ln\big(\beta_0 + z_{ij}^{-1}\big)] \tag{4.29}$$

whose optimization requires to solve the system of equations

$$W^* = \sum_{i<j} \frac{a_{ij}^*}{\beta_0 + z_{ij}^{-1}}, \tag{4.30}$$

$$\sum_{i<j} w_{ij}^* \cdot \frac{\partial z_{ij}^{-1}}{\partial \underline{\phi}} = \sum_{i<j} \frac{a_{ij}^*}{\beta_0 + z_{ij}^{-1}} \cdot \frac{\partial z_{ij}^{-1}}{\partial \underline{\phi}}. \tag{4.31}$$

The 'annealed' recipe, instead, prescribes to maximize the likelihood

$$\mathcal{G}_{\underline{\psi}} = \sum_{i<j} [-(\beta_0 + z_{ij}^{-1}) w_{ij}^* + p_{ij} \ln\big(\beta_0 + z_{ij}^{-1}\big)] \tag{4.32}$$

whose optimization requires to solve the system of equations

$$W^* = \sum_{i<j} \frac{p_{ij}}{\beta_0 + z_{ij}^{-1}}, \tag{4.33}$$

$$\sum_{i<j} w_{ij}^* \cdot \frac{\partial z_{ij}^{-1}}{\partial \underline{\phi}} = \sum_{i<j} \frac{p_{ij}}{\beta_0 + z_{ij}^{-1}} \cdot \frac{\partial z_{ij}^{-1}}{\partial \underline{\phi}}. \tag{4.34}$$

The 'quenched' recipe, on the other hand, requires to solve the system of equations $\langle \beta_0 \rangle = \sum_{\mathbf{A} \in \mathbb{A}} P(\mathbf{A}) \beta_0(\mathbf{A})$ and $\langle \underline{\phi} \rangle = \sum_{\mathbf{A} \in \mathbb{A}} P(\mathbf{A}) \underline{\phi}(\mathbf{A})$ which no longer have an explicit expression.

Figures 15 and 23 in Appendix C.2 illustrate the case-study of the WTW: although the 'quenched' distributions induced by the three, binary recipes are characterized by different shapes that may overlap (as in the case of the parameters ρ - under the UBRGM-induced and UBCM-induced binary recipes - and γ - under all, binary recipes) or not (as in the case of the parameters β_0 and α), 'annealed' and 'quenched' estimations always coincide (the only, small discrepancy being observable for the parameter β_0, under the UBRGM-induced, binary recipe). The 'deterministic' estimation, instead, is compatible with the other, two ones only for the parameter α, under the UBCM-induced, binary recipe.

Sparse networks deserve a separate discussion. The results concerning the homogeneous and econometric variant of the BLN, defined by posing $\beta_{ij} \equiv \beta_0 + z_{ij}^{-1}, \forall\, i < j$, with $z_{ij} \equiv e^{\rho}(s_i s_j)^{\alpha}$, are analogous to the ones shown for the WTW - in the latter case, the 'annealed' estimates of β_0, ρ and α are very close to their 'quenched' counterparts, the relative error $\mathrm{RE} = |(\phi_i^{\mathrm{ann}} - \phi_i^{\mathrm{que}})/\phi_i^{\mathrm{ann}}|$ amounting at $\simeq 10^{-3}$ for β_0 and $\simeq 10^{-4}$ for ρ, α. On the contrary, these conclusions no longer hold true when the weakly heterogeneous variant of the CEM is considered: in this case, in fact, carrying out the 'quenched' approach can lead to binary configurations with disconnected nodes, a circumstance that impairs the correct estimation of the corresponding parameters; carrying out the 'annealed' estimation, instead, remains a feasible task.

4.5 Discussion

The present contribution focuses on three recipes for estimating the parameters entering into the definition of statistical network models, i.e. the 'deterministic', 'annealed' and 'quenched' ones. In order to implement them, we have considered several variants of the CEM, i.e. the homogeneous one (defined by one, global parameter), the weakly heterogeneous one (defined by N, local parameters) and the econometric one (defined by four, global parameters), each one combined with three, different recipes for estimating the network topology (i.e. the UBRGM, the UBCM and the LM).

The 'deterministic' recipe, routinely employed in econometrics to determine the so-called hurdle models [70], prescribes to estimate the parameters associated with the weighted constraints on the empirical realization of the network topology. Since it considers $\mathbf{A}^*$ as not being subject to variation, its use is recommended whenever $\mathrm{Var}[a_{ij}] = p_{ij}(1 - p_{ij}) \simeq 0$ or, equivalently, $p_{ij} \simeq a_{ij}, \forall\, i < j$, i.e. whenever the binary random variables can be safely considered as deterministic or, more in general, whenever their (scale of) variation is negligible with respect to the (scale of) variation of the weighted random variables.

Accounting for such a variability in a fully consistent manner can

be achieved upon adopting either the 'annealed' recipe (according to which parameters are estimated on the average network topology) or the 'quenched' recipe (according to which parameters are, first, estimated on a large number of binary configurations and, then, averaged); the main difference between these procedures lies in the order in which the two operations of 'averaging' (of the entries of the binary adjacency matrix) and 'maximization' (of the related likelihood function) are taken. Interestingly, no variant of the CEM is sensitive to this choice (neither the purely structural ones nor the 'econometric' one); while, however, the coincidence of the 'annealed' and 'quenched' estimates for purely structural models can be explicitly verified, this is no longer true when the 'econometric' variant is considered: in this case, in fact, one can proceed only numerically.

This evidence reveals the main limitation of the 'quenched' approach, i.e. the need of resorting upon an explicit sampling of the chosen, binary ensemble. As any 'good' sampling algorithm must lead to a faithful representation of the parent distribution, we are left with the following question: is this always guaranteed, in all cases of interest to us?

This seems to be the case for dense networks. As shown in [48], a study of the coefficient of variation of the constraints defining the 'vector' variant of the CEM (i.e. the ratio between standard deviation and expected value of each degree) reveals it to vanish in the asymptotic limit: in other words, the fluctuations affecting each degree vanish, a result guaranteeing that the degree sequence of any configuration in the ensemble remains 'close enough' to the empirical one.

When sparse networks are, instead, considered, the coefficient of variation of the constraints defining the 'vector' variant of the CEM remains finite in the asymptotic limit: in other words, the fluctuations affecting each degree do not vanish, a result implying that the degree sequence of any configuration in the ensemble may largely differ from the empirical one; to provide a concrete example, nodes whose empirical degree is 'small' may disconnect, hence inducing the resolution of a system of equations which is not even compatible with the set of constraints defining the original problem. Overcoming such a limitation implies quantify-

ing the bias affecting the estimates in cases like these: although possible, calculations of this kind are far beyond the scope of the present paper.

Overall, then, two alternatives exist to overcome the main limitation of the 'deterministic' estimation recipe, i.e. that of ignoring the variety of structures that are compatible with a given probability distribution $P(\mathbf{A})$, namely the 'annealed' and 'quenched' ones. As the 'quenched' recipe requires an explicit sampling the ensemble - potentially leading to inconsistent estimates for sparse configurations - we believe the 'annealed' one to represent the better alternative, 1) being *unbiased* by definition, 2) being *convenient* from a numerical point of view, 3) reducing to the 'deterministic' recipe in case the empirical configuration is not subject to variation.

Chapter 5

Triadic structure of the Dutch Production Multiplex

This chapter illustrates the results of the analysis documented in [4]. Here, we test the maximally-random character of triadic structures in the Dutch Production Multiplex, constructed from the data collected by CBS-Statistics Netherlands. To this aim, we employ a novel (weighted, conditional) null model capable of controlling for both the directionality and the reciprocity of links - hence, of detecting deviations of the empirical number of triads from the expected one due to the abundance of patterns, the concentration of money on them or both. We find that 1) analyzing the aggregated structure, as routinely done in the literature on production networks, is not sufficient to characterize their layer-specific, triadic structure; 2) most layers are characterized by a maximally-random triadic structure while the remaining ones display (overall) small deviations from the expected behavior.

5.1 Introduction

The increasing availability of data at the industry and firm level has led to the appearance of a vast number of studies analyzing the system of

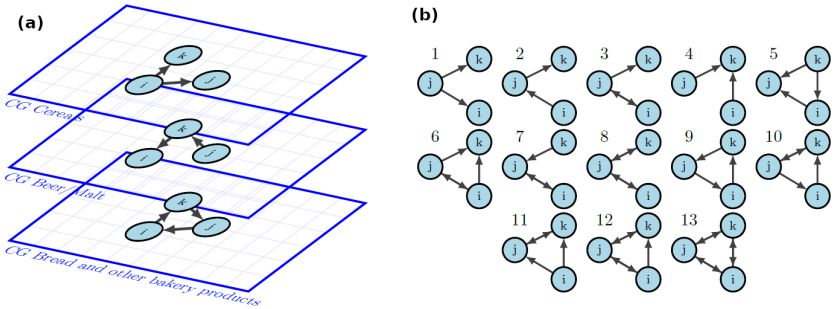

Figure 16: (a) Graphical representation of the Dutch multi-layer production network and (b) the existing 13 triadic motifs. For illustrative purposes, we represent three industries/firms i, j and k on three commodity group layers, namely from the top to the bottom (1) Cereals, (2) Beer/Malt, (3) Bread and other bakery products. Industries/Firms can trade different products by connecting themselves according to the 13 possible connected subgraphs.

customer-supplier trade relationships, i.e. the *production network*, established by industries [95, 96, 97, 98, 99, 100] or firms [101, 110, 116, 117, 118, 119, 120, 121, 122, 102, 71, 103, 104, 105, 106, 107, 108, 109, 73, 111, 112, 72, 113, 114, 115] and their impact on country-level macroeconomic statistics.

Even in the time of globalization - characterized by highly-interconnected, global supply chains - domestic production networks are still relevant[1]. The heterogeneity of production interlinkages plays an essential role in amplifying economic growth [111] and in propagating shocks [95, 118] related to both endogenous events (such as the 2008 financial crisis [123, 124]) and exogenous events (such as Hurricane Sandy [106], the Great East Asian Earthquake [117, 104], the Covid-19 pandemic [122, 105, 100]).

While aggregated information about single firms is contained in most National Statistical Institutes' repositories, reliable data on input/output relationships is available only for a small number of countries. For instance, 1) the Compustat dataset contains the major customers of the

[1]It has been shown that for a small country as Belgium, while almost all firms directly or indirectly import and export to foreign firms, these exchanges represent the minority of domestic firms' total revenues [121].

publicly listed firms in the USA [101]; 2) FactSet Revere dataset contains major customers of publicly listed firms at a global level, with a focus on the USA, Europe and Asia [107]; 3) the dataset collected by Tokyo Shoko Research Ltd. (TSR) [117] and the one collected by Teikoku DataBank Inc. (TDB) [112] (commercially available in Japan) are characterized by a large coverage of Japanese firms but with a limited amount of commercial partners; 4) country-specific datasets[2] contain the yearly value of transaction data among VAT-liable firms [120, 122]; 5) the Costa Rican and Dominican Republic datasets contains the yearly value of transaction data among all domestic firms [132, 133].

In production networks, user firms connect to supplier firms to buy goods for their own production: customer-supplier relationships are, thus, characterized by an intrinsic granularity that is usually neglected. The economic theory solves the problem of product granularity by assuming that industries and firms supply a specific product [95, 98], an assumption often not holding in reality, as single firms can possess more than a production pipeline, hence being capable of supplying multiple products - e.g. Samsung, a telecommunication company, sells household appliances too and multinational companies such as Amazon and Google supply a large number of different products.

CBS-Statistics Netherlands has recently produced two datasets concerning the multi-layer production network for domestic, intermediate trade of Dutch firms for 2012 [71] and 2018 [116], each layer corresponding to a different product exchanged by a firm for its own production process, as illustratively depicted in Fig. 16(a): the presence of product granularity makes it a valuable source for the analysis of commodity-specific structural patterns. These datasets have been recently used to prove the complementarity structure of production networks [73] by comparing the number of 3- and 4-cycles with the outcome of a null model taking into consideration the in-degree and out-degree distributions: firms are matched according to a deterministic procedure that is shown to decrease the dataset quality by originating a bias in the network density

[2]Among which the ones for Brazil [125], Belgium [120], Ecuador [126], Hungary [122], Kenya [127], Spain [128], Turkey [129], Rwanda and Uganda [130] and West Bengal [131].

and the degree distribution, as shown in [113] for a subset of known links of the production network collected by Dun & Bradstreet (whose dataset contains the customers of the largest 500 suppliers in the Dutch Economy).

Here, we consider the 2018 version of the CBS-Statistics Netherlands datasets to construct an inter-industry network over which inspecting the abundance of triadic motifs and anti-motifs, i.e. over- and under-occurrences of the patterns represented in Fig. 16(b). Triadic and tetradic subgraphs are understood to play the role of 'building blocks' of complex networks [134] (e.g. as functional modules in biological networks [135, 136], homophily-driven structures in social networks [137], complementarity-driven structures in production networks [72, 73]) and their temporal variability has been proven capable of providing early-warning signals of topological collapse in interbank networks [138, 10, 8] and stock market networks [124]. Besides, the triadic structure of the majority of real-world networks has been proven to be maximally-random [139] and, once constrained, capable of determining their global structure [140]. By contrast, research on weighted, triadic motifs and anti-motifs is still underdeveloped: to the best of our knowledge, only one study focuses on their abundance in trade networks, by implementing a probabilistic model based on random walks [141].

The task of detecting motifs calls for a definition of the randomization method employed for computing random expectations: here, we focus on the class of maximum-entropy ones. The latter [59, 60, 61] construct probability distributions that are maximally-random by construction. Available, global or local quantities are enforced as a constraints and their Lagrange multipliers are numerically determined by invoking the maximum-of-the-likelihood principle [142]. Such a framework has been proven to successfully reconstruct economic and financial systems [143, 62, 49, 58] such as the WTW (whose topology and weights can be accurately predicted [19, 46, 51] in both an integrated [56] and a conditional fashion [63], by either imposing purely structural constraints or defining hybrid models accommodating economic factors as well [29,

1, 2]), interbank networks [78] and interfirm networks (e.g. the ones determined by the payment flows established by Dutch firms that are clients of ABN-Amro or ING [103]). Two studies using the maximum-entropy formalism are worth to be mentioned: a theoretical one whose authors compute the z-score associated to each, triadic occurrence analytically [50] and an applied one where triadic motifs are recognized as early-warning indicators of the 2008 financial crisis [10]. Our contribution goes in this direction, using maximum-entropy methods that constrain the degree and the strength sequences to study the abundance of triadic connections and the amount of money associated with them characterizing the different layers of the Dutch Production Multiplex.

5.2 The Dutch Production Multiplex

The Dutch Production Multiplex has been constructed by CBS-Statistics Netherlands as follows. Firm-level data is obtained from the General Business Register (ABR), containing data for over $1.700.000$ firms: after removing the micro-firms with an annual net turnover below 10.000 €, $\simeq 900.000$ firms remain, accounting for $\simeq 99.5\%$ of the Dutch economy output in 2018.

The breakdown in commodities[3] is 1) derived from the Structural Business Statistics survey for commercial industries, 2) derived from the Prodcom survey for manufacturing industries and 3) estimated by National Accounts for non-commercial industries; a breakdown in intermediate supply/use per firm has been, then, carried out by employing intermediate purchases as the distributional key; a third breakdown in commodity groups has been realized by integrating data from the SBS and Prodcom surveys.

The resulting dataset has been, then, compared with the industry-level supply/use tables at the SBI4 Standard Industry Classification and appropriate rescaling of supply/use per firm has been performed by using the Iterative Proportion Fitting algorithm. Once supply/use per firm

[3]In most cases, data about the breakdown in commodities is available for industries as a whole and not for individual organizations within industries.

and per commodity is obtained, their in-degree distribution is estimated via the stylized facts concerning Japanese firms, connecting supply and use to out- and in-degree sequences, respectively [115].

Suppliers and users are, then, matched according to a deterministic procedure that takes into account (1) their trade capacity (encoded by their net turnover), (2) their mutual distance, (3) the presence of a link between their industries, (4) the observed relationship in the Dun & Bradstreet dataset. Finally, the resulting interfirm network at the 650 commodity level is compared with the (known) inter-industry network at the 250 commodity level and consequent adjustments are done to weights and links.

Even if the 2018 version of the CBS production network [116] represents an improvement of the 2012 version, having more auxiliary microdata and industry-level data, the intensive imputation procedure and the presence of biases contained in the previous version [113] let us prefer not to use the interfirm multilayer production network as it is. Instead, we took advantage of the tested coherence between the interfirm network at the 192 commodity level and the inter-industry network as the key point of our pre-processing: specifically, we aggregated the 650 commodity groups into 192, coherently with industry-level known tables and, then, we aggregated firms according to their SBI5 Standard Industry Classification. Passing from SBI4 to SBI5 leads to better resolve industries, thus increasing their number from 132 to 870. Finally, we cleaned for intra-industry trade[4] and obtain a multi-layer inter-industry production network containing linkages and weights for 862 industries (nodes) and 187 commodity groups (layers).

The resulting dataset, while being the most reliable and detailed multiproduct, inter-industry domestic production network for intermediate use, has, at least, two, clear limitations, i.e. 1) the lack of (known) firm granularity, which is a necessary ingredient to unveil more detailed net-

[4]Intra-industry trade is relevant for both intensive (weights) and extensive margins (links) but the lack of self-loops in triadic structures makes them irrelevant for the present analysis.

work anomalies - especially when intra-industry, supply-and-use relationships are considered; (2) the presence of potential imputation biases affecting the passage from SBI4 to SBI5. Nonetheless, we believe that our findings shed light on the role of product granularity in modeling production networks and characterizing their triadic structure.

5.3 Maximum-entropy randomization methods

The main goal of randomization methods is that of generating statistical ensembles of networks which are maximally-random given the available data[5]: in our case, we randomize each layer of our multiplex separately. Statistical measures of interest are, then, extracted as ensemble averages.

The available data - constraining the entropy maximization - consists of the supplier's (user's) tendency to supply (use) a specific commodity and its output(input): the obtained statistical ensemble of networks, thus, represents the possible realizations of the system once suppliers' and users' tendencies and taken into account. Hence, if the empirical statistics showed significant deviations from the model-induced ensemble averages, they would signal the presence of higher-order correlations not explained by the structural constraints themselves.

Reliable data is available for 187 commodity groups and not for each, specific commodity, an evidence implying that a group can contain similar, although distinct, products. A supplier of one of the products belonging to a group can simultaneously be a user of a different product within the same group, leading to possibly reciprocated links among industries: this leads us to consider models that encode the tendency to establish as well as reciprocate supply/use relationships.

[5]The maximum-entropy framework guarantee unbiasedness with respect to the missing data, as also proven by the independent tests [144, 145, 146, 147].

5.3.1 Binary benchmarks

For binary, directed graphs, the maximum-entropy formalism prescribes to maximize the functional reading

$$S(P) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \ln P(\mathbf{A}) \tag{5.1}$$

subject to the normalization of the probability distribution

$$\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) = 1 \tag{5.2}$$

and to the vector of constraints $\{C^*_\alpha\}$, i.e.

$$\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) C_\alpha(\mathbf{A}) = C^*_\alpha, \quad \forall\, \alpha. \tag{5.3}$$

Solving the optimization problem above leads to the canonical probability distribution reading

$$P(\mathbf{A}) = \frac{e^{-\sum_\alpha \theta_\alpha C_\alpha(\mathbf{A})}}{\sum_{\mathbf{A}\in\mathbb{A}} e^{-\sum_\alpha \theta_\alpha C_\alpha(\mathbf{A})}} \equiv \frac{e^{-H(\mathbf{A})}}{\sum_{\mathbf{A}\in\mathbb{A}} e^{-H(\mathbf{A})}} \tag{5.4}$$

where the *graph Hamiltonian* is defined as $H(\mathbf{A}) \equiv \sum_\alpha \theta_\alpha C_\alpha(\mathbf{A})$. Let us, now, focus on the binary benchmarks constraining local properties.

The Directed Binary Configuration Model (DBCM)

The DBCM is defined by the out-degree and in-degree sequences, representing the number of industries node i sells to and the number of industries node i buys from, respectively. Upon writing

$$H(\mathbf{A}) = \sum_i [\alpha_i^{out} k_i^{out}(\mathbf{A}) + \alpha_i^{in} k_i^{in}(\mathbf{A})] \tag{5.5}$$

the probability distribution can be re-written as the product of Bernoulli-like probabilities, i.e.

$$P(\mathbf{A}) = \prod_{i\neq j} p_{ij}^{a_{ij}} (1-p_{ij})^{1-a_{ij}} \tag{5.6}$$

where p_{ij} denotes the probability that supplier i and user j are connected and equals the expression

$$p_{ij} = \frac{x_i^{out} x_j^{in}}{1 + x_i^{out} x_j^{in}} \tag{5.7}$$

with $x_i^{out} \equiv e^{-\alpha_i^{out}}$ and $x_i^{in} \equiv e^{-\alpha_i^{in}}$.

The maximum-of-the-likelihood principle leads to numerically determine the Lagrange parameters $\{\alpha_i^{out}\}$ and $\{\alpha_i^{in}\}$ upon solving the system of $2N$ coupled equations

$$k_i^{out}(\mathbf{A}^*) = \sum_{j(\neq i)} a_{ij} = \sum_{j(\neq i)} p_{ij} = \langle k_i^{out} \rangle, \; \forall \, i \tag{5.8}$$

$$k_i^{in*}(\mathbf{A}^*) = \sum_{j(\neq i)} a_{ji} = \sum_{j(\neq i)} p_{ji} = \langle k_i^{in} \rangle, \; \forall \, i \tag{5.9}$$

where N is the number of industries in the network and $\langle k_i^{out} \rangle$ and $\langle k_i^{in} \rangle$ denote the ensemble average of the i-th out-degree and the i-th in-degree, respectively.

The Reciprocal Binary Configuration Model (RBCM)

The RBCM is defined by the sequences of non-reciprocated out-degrees $\{k_i^{\rightarrow}\}$, non-reciprocated in-degrees $\{k_i^{\leftarrow}\}$ and reciprocated degrees $\{k_i^{\leftrightarrow}\}$. Upon writing

$$H(\mathbf{A}) = \sum_i [\alpha_i^{\rightarrow} k_i^{\rightarrow}(\mathbf{A}) + \alpha_i^{\leftarrow} k_i^{\leftarrow}(\mathbf{A}) + \alpha_i^{\leftrightarrow} k_i^{\leftrightarrow}(\mathbf{A})] \tag{5.10}$$

where $k_i^{\rightarrow}(\mathbf{A}) = \sum_{j(\neq i)} a_{ij}(1 - a_{ji}) \equiv \sum_{j(\neq i)} a_{ij}^{\rightarrow}$, $k_i^{\leftarrow}(\mathbf{A}) = \sum_{j(\neq i)} a_{ji}(1 - a_{ij}) \equiv \sum_{j(\neq i)} a_{ij}^{\leftarrow}$ and $k_i^{\leftrightarrow}(\mathbf{A}) = \sum_{j(\neq i)} a_{ij} a_{ji} \equiv \sum_{j(\neq i)} a_{ij}^{\leftrightarrow}$, the probability distribution can be written as the product of Bernoulli-like probabilities, i.e.

$$P(\mathbf{A}) = \prod_{i<j} \left(p_{ij}^{\rightarrow}\right)^{a_{ij}^{\rightarrow}} \left(p_{ij}^{\leftarrow}\right)^{a_{ij}^{\leftarrow}} \left(p_{ij}^{\leftrightarrow}\right)^{a_{ij}^{\leftrightarrow}} \left(p_{ij}^{\nleftrightarrow}\right)^{a_{ij}^{\nleftrightarrow}} \tag{5.11}$$

where

$$p_{ij}^{\rightarrow} = \frac{x_i^{\rightarrow} x_j^{\leftarrow}}{1 + x_i^{\rightarrow} x_j^{\leftarrow} + x_i^{\leftarrow} x_j^{\rightarrow} + x_i^{\leftrightarrow} x_j^{\leftrightarrow}}, \tag{5.12}$$

$$p_{ij}^{\leftarrow} = \frac{x_i^{\leftarrow} x_j^{\rightarrow}}{1 + x_i^{\rightarrow} x_j^{\leftarrow} + x_i^{\leftarrow} x_j^{\rightarrow} + x_i^{\leftrightarrow} x_j^{\leftrightarrow}}, \tag{5.13}$$

$$p_{ij}^{\leftrightarrow} = \frac{x_i^{\leftrightarrow} x_j^{\leftrightarrow}}{1 + x_i^{\rightarrow} x_j^{\leftarrow} + x_i^{\leftarrow} x_j^{\rightarrow} + x_i^{\leftrightarrow} x_j^{\leftrightarrow}}, \tag{5.14}$$

$$p_{ij}^{\nleftrightarrow} = \frac{1}{1 + x_i^{\rightarrow} x_j^{\leftarrow} + x_i^{\leftarrow} x_j^{\rightarrow} + x_i^{\leftrightarrow} x_j^{\leftrightarrow}} \tag{5.15}$$

and $x_i^{\rightarrow} \equiv e^{-\alpha_i^{\rightarrow}}$, $x_i^{\leftarrow} \equiv e^{-\alpha_i^{\leftarrow}}$, $x_i^{\leftrightarrow} \equiv e^{-\alpha_i^{\leftrightarrow}}$ are the Lagrange multipliers controlling for the non-reciprocated out-degree, the non-reciprocated in-degree and the reciprocated degree of node i, respectively.

The maximum-of-the-likelihood principle leads to numerically determine the Lagrange multipliers above upon solving the system of $3N$ coupled equations

$$k_i^{\rightarrow}(\mathbf{A}^*) = \sum_{j(\neq i)} a_{ij}^{\rightarrow} = \sum_{j \neq i} p_{ij}^{\rightarrow} = \langle k_i^{\rightarrow} \rangle, \; \forall \, i \tag{5.16}$$

$$k_i^{\leftarrow}(\mathbf{A}^*) = \sum_{j(\neq i)} a_{ij}^{\leftarrow} = \sum_{j(\neq i)} p_{ij}^{\leftarrow} = \langle k_i^{\leftarrow} \rangle, \; \forall \, i \tag{5.17}$$

$$k_i^{\leftrightarrow}(\mathbf{A}^*) = \sum_{j(\neq i)} a_{ij}^{\leftrightarrow} = \sum_{j(\neq i)} p_{ij}^{\leftrightarrow} = \langle k_i^{\leftrightarrow} \rangle, \; \forall \, i \tag{5.18}$$

where N is the number of industries in the network.

5.3.2 Conditional weighted benchmarks

When inspecting network weights, the numeric nature of the involved trade volumes drive the choice towards the basket of benchmarks to be used: if the weights are discrete-valued, the constrained entropy maximization leads to a family of geometric distributions [1, 51, 148]; if, in contrast, the weights are continuous-valued, the constrained entropy maximization leads to a family of exponential distributions [63, 2]. Here, we stick to conditional models for continuous-valued weights, which are well-defined only after the functional form of $P(\mathbf{A})$ has been determined.

Maximizing the conditional Shannon entropy

$$S(\overline{Q}|P) = -\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) \ln Q(\mathbf{W}|\mathbf{A}) d\mathbf{W} \tag{5.19}$$

given the normalization of the related probability

$$\int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) = 1 \tag{5.20}$$

and the set of constraints $\{C_\alpha^*\}$

$$\sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) C_\alpha(\mathbf{W}) d\mathbf{W} = C_\alpha^*, \; \forall\, \alpha; \tag{5.21}$$

leads us to obtain

$$Q(\mathbf{W}|\mathbf{A}) = \begin{cases} \frac{e^{-H(\mathbf{W})}}{\int_{\mathbb{W}_{\mathbf{A}}} e^{-H(\mathbf{W})} d\mathbf{W}}, & \mathbf{W} \in \mathbb{W}_{\mathbf{A}} \\ 0, & \mathbf{W} \notin \mathbb{W}_{\mathbf{A}} \end{cases} \tag{5.22}$$

where $\mathbb{W}_{\mathbf{A}}$ stands for the ensemble of weighted configurations compatible with $\mathbf{A}$ and $H(\mathbf{W}) \equiv \sum_\alpha \beta_\alpha C_\alpha(\mathbf{W})$.

Parameters are estimated using the maximum-of-the-likelihood estimation, reading

$$\mathcal{L}_{\mathbf{W}|\mathbf{A}} = -H_{\vec{\beta}}(\mathbf{W}) - \ln Z_{\vec{\beta},\mathbf{A}} \tag{5.23}$$

where $Z_{\vec{\beta},\mathbf{A}}$ is the *conditional partition function* an explicit computation of which is feasible only once $\mathbf{A}$ is known. As we have seen in the previous chapter, however, estimating parameters on the empirical topology neglects its intrinsic variability: this problem can be solved by consideringthe generalized log-likelihood

$$\mathcal{G}_{\vec{\beta}} = -H_{\vec{\beta}}(\langle \mathbf{W} \rangle) - \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A}) \ln Z_{\vec{\beta},\mathbf{A}} \tag{5.24}$$

where $P(\mathbf{A})$ is the probability induced by the binary model. In what follows, we will employ the estimation based on $\mathcal{G}_{\vec{\beta}}$ [63].

The Conditional Reconstruction Model A (CReM$_\mathbf{A}$)

When randomizing a weighted adjacency matrix W, trade marginals such as the out-strength $s_i^{out} = \sum_{j(\neq i)} w_{ij}$ and the in-strength $s_i^{in} = \sum_{j(\neq i)} w_{ji}$ - indicating the total output and total input of industry i, respectively - are usually constrained [51, 56]. Solving such a problem leads to

$$H(\mathbf{W}) = \sum_i [\beta_i^{out} s_i^{out} + \beta_i^{in} s_i^{in}] \tag{5.25}$$

in turn, inducing

$$Q(\mathbf{W}|\mathbf{A}) = \prod_{\substack{i\neq j \\ a_{ij}=1}} q_{ij}(w_{ij}|a_{ij}=1) = \prod_{\substack{i\neq j \\ a_{ij}=1}} \left[\left(\beta_i^{out} + \beta_j^{in}\right) e^{-(\beta_i^{out}+\beta_j^{in})w_{ij}} \right]^{a_{ij}} \tag{5.26}$$

i.e. the product of dyadic exponential distributions, conditional on the establishment of the link and regulated by the related, node-specific Lagrange parameters. By maximizing the generalized log-likelihood[6] we obtain the system of $2N$ coupled equations reading

$$s_i^{out} = \sum_{j(\neq i)} \frac{f_{ij}}{\beta_i^{out} + \beta_j^{in}} = \langle s_i^{out} \rangle, \; \forall \, i \tag{5.27}$$

$$s_i^{in} = \sum_{j(\neq i)} \frac{f_{ji}}{\beta_i^{in} + \beta_j^{out}} = \langle s_i^{in} \rangle, \; \forall \, i \tag{5.28}$$

whose resolution leads us to find β_i^{out} and β_i^{in} for each industry.

The Conditionally Reciprocal Weighted Configuration Model (CRWCM)

In order to take into account reciprocity, we develop a novel model that accounts for the different nature of links whose weights are sampled, namely reciprocated and non-reciprocated ones. This choice leads to the definition of four trade marginals for each supplier/user, namely

[6] Such a procedure is equivalent at maximizing the sum of the logarithms of the dyadic conditional probabilities $q_{ij}(w_{ij}|a_{ij}=1) = \left[\left(\beta_i^{out} + \beta_j^{in}\right) e^{-(\beta_i^{out}+\beta_j^{in})} \right]^{f_{ij}}$.

- the non-reciprocated out-strength $s_i^{\rightarrow}$, which measures the output of supplier i to users from which it does not buy, i.e

$$s_i^{\rightarrow} = \sum_{j(\neq i)} a_{ij}^{\rightarrow} w_{ij} = \sum_{j(\neq i)} w_{ij}^{\rightarrow} \tag{5.29}$$

- the non-reciprocated in-strength $s_i^{\leftarrow}$, which measures the input of industry i from suppliers to which it does not supply, i.e

$$s_i^{\leftarrow} = \sum_{j(\neq i)} a_{ij}^{\leftarrow} w_{ji} = \sum_{j(\neq i)} w_{ij}^{\leftarrow} \tag{5.30}$$

- the reciprocated out-strength $s_i^{\leftrightarrow,out}$, measuring the output of supplier i to users from which it also buys from, i.e

$$s_i^{\leftrightarrow,out} = \sum_{j(\neq i)} a_{ij}^{\leftrightarrow} w_{ij} = \sum_{j(\neq i)} w_{ij}^{\leftrightarrow,out} \tag{5.31}$$

- and the reciprocated in-strength $s_i^{\leftrightarrow,in}$, measuring the input of user i from suppliers to which it also supplies to, i.e.

$$s_i^{\leftrightarrow,in} = \sum_{j(\neq i)} a_{ij}^{\leftrightarrow} w_{ji} = \sum_{j(\neq i)} w_{ij}^{\leftrightarrow,in} \tag{5.32}$$

Solving the constrained maximum-entropy problem implies writing the graph Hamiltonian

$$H(\mathbf{W}) = \sum_i [\beta_i^{\rightarrow} s_i^{\rightarrow} + \beta_i^{\leftarrow} s_i^{\leftarrow} + \beta_i^{\leftrightarrow,out} s_i^{\leftrightarrow,out} + \beta_i^{\leftrightarrow,in} s_i^{\leftrightarrow,in}] \tag{5.33}$$

leading to

$$Q(\mathbf{W}|\mathbf{A}) = \prod_{\substack{i \neq j \\ a_{ij}=1}} q_{ij}(w_{ij}|a_{ij}=1) \tag{5.34}$$

where $q_{ij}(w_{ij}|a_{ij})$ depends on the dyadic state, namely

$$\begin{cases} (\beta_i^{\rightarrow} + \beta_j^{\leftarrow}) e^{-(\beta_i^{\rightarrow} + \beta_j^{\leftarrow}) w_{ij}^{\rightarrow}}, & \text{for } w_{ij}^{\rightarrow} > 0 \\ (\beta_i^{\leftrightarrow,out} + \beta_j^{\leftrightarrow,in}) e^{-(\beta_i^{\leftrightarrow,out} + \beta_j^{\leftrightarrow,in}) w_{ij}^{\leftrightarrow,out}}, & \text{for } w_{ij}^{\leftrightarrow,out} > 0 \\ 0, & \text{for } w_{ij} = 0 \end{cases} \tag{5.35}$$

Layer-Statistics	Min	Lower Quartile	Median	Upper Quartile	Max
N	4	62	149	544	822
L	3	203	678	2076	15198
W_{tot}	0.95	239	768	2027	23767
r_t	0	0.01	0.05	0.14	0.78
r_w	0	0.01	0.08	0.28	0.78

Table 11: Description of the distribution of statistics such as the number of active industries N, the number of links L, the total weight W_{tot}, the topological reciprocity r_t and the weighted reciprocity r_w across commodity layers of the inter-industry network.

The resulting generalized log-likelihood is separable into a reciprocal and a non-reciprocal component, i.e. $\mathcal{G}_{\vec{\beta}} = \mathcal{G}^{\rightarrow}_{\vec{\beta}} + \mathcal{G}^{\leftrightarrow}_{\vec{\beta}}$ (see also Appendix B): the Lagrange parameters $\vec{\beta}$ are retrieved by maximizing $\mathcal{G}_{\vec{\beta}}$, which amounts at solving two uncoupled, systems of $2N$ coupled equations reading

$$s_i^{\rightarrow} = \sum_{j(\neq i)} \frac{f_{ij}^{\rightarrow}}{\beta_i^{\rightarrow} + \beta_j^{\leftarrow}} = \langle s_i^{\rightarrow} \rangle, \; \forall \, i \tag{5.36}$$

$$s_i^{\leftarrow} = \sum_{j(\neq i)} \frac{f_{ij}^{\leftarrow}}{\beta_i^{\leftarrow} + \beta_j^{\rightarrow}} = \langle s_i^{\leftarrow} \rangle, \; \forall \, i \tag{5.37}$$

for the non-reciprocal sub-problem and

$$s_i^{\leftrightarrow,out} = \sum_{j(\neq i)} \frac{f_{ij}^{\leftrightarrow}}{\beta_i^{\leftrightarrow,out} + \beta_j^{\leftrightarrow,in}} = \langle s_i^{\leftrightarrow,out} \rangle, \; \forall \, i \tag{5.38}$$

$$s_i^{\leftrightarrow,in} = \sum_{j(\neq i)} \frac{f_{ij}^{\leftrightarrow}}{\beta_i^{\leftrightarrow,in} + \beta_j^{\leftrightarrow,out}} = \langle s_i^{\leftrightarrow,in} \rangle, \; \forall \, i \tag{5.39}$$

for the reciprocal sub-problem.

5.4 Results

Measuring empirical reciprocity statistics

The presence of data on product granularity gives us the opportunity to study heterogeneity across commodity layers. Let us consider in Table 11 the number of layer-active industries N, the number of links L, the total weight W_{tot}, and reciprocity measures such as the *topological reciprocity* r_t, defined as the ratio of reciprocated links to L, i.e.

$$r_t = \frac{L^{\leftrightarrow}}{L} = \frac{\sum_{i,j\neq i} a_{ij}^{\leftrightarrow}}{\sum_{i,j\neq i} a_{ij}}. \tag{5.40}$$

and its *weighted* counterpart r_w, defined as the ratio of total weight on reciprocated links to W, i.e.

$$r_w = \frac{W_{tot}^{\leftrightarrow}}{W_{tot}} = \frac{\sum_{i,j\neq i} w_{ij}^{\leftrightarrow,out}}{\sum_{i,j\neq i} w_{ij}}. \tag{5.41}$$

The median for N is 149, meaning that for around 50% of commodity layers there are less than 149 active industries (as suppliers or users). At the same time, 25% of commodity layers have less than 62 industries, and another 25% have more than 544 industries. Consequently, industries are specialized among a small number of business activities for half of the commodity groups but, a small, and not negligible, number of layers is characterized by a high number of active industries and hence of industry heterogeneity. Some examples are suppliers of plastic goods that are sold to users with heterogeneous specializations, for instance, Bread, Beer, Cereals, Fish, etc. Also the distributions regarding the number of commodity-specific links L and the related total weight W_{tot} have wide distributions, with a minimum with few digits, respectively 3 and 0.95 (in millions of euro), and a maximum in 5 digits, respectively 15198 and 23767, implying a high degree of heterogeneity in network structure across commodity layers.

Passing from the commodity global statistics to r_t and r_w, we see a high degree of heterogeneity also in this case, namely a minimum value

of 0 stands for layers where no link is reciprocated, i.e. users and suppliers represent two distinct sets of nodes (bipartite graph). Instead, in the majority of the commodities (above 75%) there is a not-null reciprocity. In fact, the median is respectively 0.05 and 0.08. There is also the presence of a small number of commodities (below 10%) which are characterized by a large reciprocity, with a maximum of 0.78 for both r_t and r_w.

Reciprocity can arise for different reasons: (1) the aggregation from firms to industries or (2) the aggregation of products. To mention the first case, consider two firms A and B in the industry i and other two firms C and D in industry j. Suppose firm A supplies to firm D, while firm C supplies to firm B, in the same commodity layer. Once the firms are aggregated in the related industries, a reciprocated link emerges between them, even if reciprocity is not present at the firm level.
The second case follows from the fact that if each commodity layer represents a unique product, that could be represented by the finest CPA product classification (with around 5000 products), and we take into account only intermediate supply and use, it is not reasonable to think that firms are at the same time suppliers and users (of that specific product). Instead, in case of product aggregation, firms may be suppliers of a product inside that commodity group and also users of another product inside that same commodity group.

Let us now move to the analysis of triads. We define *triadic occurrences* N_m, the number of times a specific m-subgraph appears and *triadic fluxes* F_m, the total amount of money circulating on each m-subgraph. In Fig. 17, we depict their values normalizing by their sum across the m-types. The normalized N_m and F_m can be considered as the relative importance of a specific type of triadic subgraph in the network. The aggregated network (depicted in blue), where the weights of all commodity groups are summed, and three commodity layers, namely 'Cereals' (in green), 'Gas/Hot Water/City Heating (in orange) and 'Agricultural Services' (in pink) are displayed.

In the aggregated network, the structures that occur relatively more are $m = 1$, represented by a supplier connected to two users and $m = 13$, the totally reciprocated cyclical triad. While $m = 13$ is probably due to

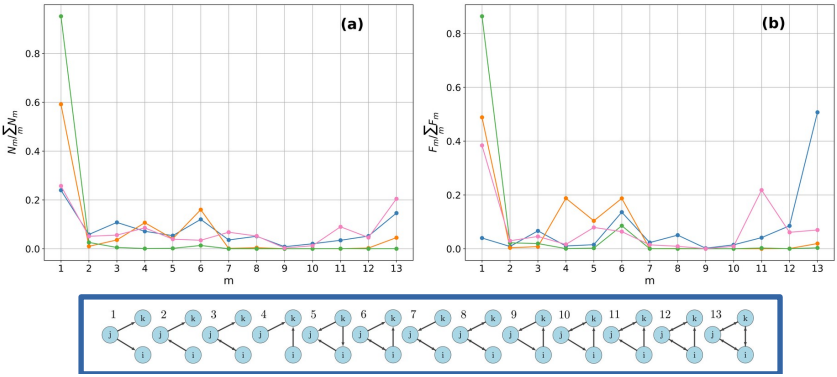

Figure 17: Normalized Triadic Occurrences (a) and Fluxes (b): the Aggregated Network (– • –) presents a high occurrence of subgraphs $m = 1$ and $m = 13$, representing open-Vs and completely reciprocated triads, respectively. The latter covers most of the total amount of money traded. The Cereals commodity layer (– • –), with a high occurrence of subgraph $m = 1$. A relatively high amount of money is distributed across $m = 1$, $m = 4$ and $m = 6$. Gas/Hot Water/City Heating layer (– • –) with a predominant occurrence and flux in subgraph $m = 1$. Agricultural Services layer (– • –), with a highly heterogenous spectrum of occurrences and fluxes. Completely cyclical triads have a high occurrence in the aggregated network, but break apart when passing to single commodity layers as G.H.C and Cereals, if not for rare cases such as Agricultural Services. In single commodity layers $m = 1$ receives the highest concentration of money, signalling a large amount of money flows over structures that greatly depend on a limited number of suppliers, which control the market.

product aggregation, the predominance of $m = 1$ is a signal of structural dependency on a limited number of suppliers. However, when normalized F_m are investigated, $m = 13$ still contain the majority of the volumes. A similar profile, in the binary case, is given by the Agricultural Services, with the predominance of $m = 1$ and $m = 13$. At the same time a relatively smaller amount of money is concentrated on $m = 13$ with respect to the aggregated case, while $m = 1$ and $m = 11$ carry a greater amount of money. During the product disaggregation weights on $m = 13$ in the aggregated network are redistributed on other subgraphs, especially $m = 1$. In 'Cereals' and 'Gas/Hot Water/City Heating' these

differences are even larger, with a relevant increase of triadic occurrences and fluxes on $m = 1$, further increasing the dependency of the network on a limited amount of suppliers. Note that when counting the different triads in Fig. 17 they are not nested, i.e. a subgraph of type $m = 8$ requires two reciprocated links and hence does not contain two subgraphs of type $m = 1$, which contain only non-reciprocated links. Consequently, the number and fluxes over all triadic subgraphs are structurally independent across different types.

5.4.1 Binary Motif Analysis

We analyze the number of occurrences N_m of all the possible triadic connected subgraphs, depicted in Fig. 1(b). To quantify their deviations to randomized expectations, we define the *binary z-score* of subgraph m

$$z\left[N_m\right] = \frac{N_m(A^*) - \langle N_m \rangle}{\sigma\left[N_m\right]} \tag{5.42}$$

where $N_m(A^*)$ is the number of occurrences of the m-type subgraph in the empirical adjacency matrix, $\langle N_m \rangle$ is its model-induced expected number of occurrences, and $\sigma\left[N_m\right]$ is the model-induced standard deviation.

An analytical procedure [50] has been developed to compute the binary z-scores for the binary case. However, the assumption on the confidence intervals - represented as the interval $(-3, 3)$ - holds true only if the ensemble distribution of N_m is Normal for each m. For all the commodities, m-types, and binary null models, we test the assumption using a Shapiro Test [149]. According to the test, N_m ensemble distributions are in a large proportion not normal at the 5% confidence level. Consequently, we must use a numeric approach. Networks are sampled according to the DBCM recipe by (1) computing the induced connection probability $p_{ij;DBCM}$ and (2) establishing a link between industry i and j if and only if a uniformly distributed random number $u_{ij} \in U(0, 1)$ is below $p_{ij;DBCM}$. The analogous recipe for RBCM requires (1) computing the set of connection probabilities for non-reciprocated connection between i and j, namely $p_{ij}^{\rightarrow}$, $p_{ij}^{\leftarrow}$ and $p_{ij}^{\nleftrightarrow}$, and reciprocated connection

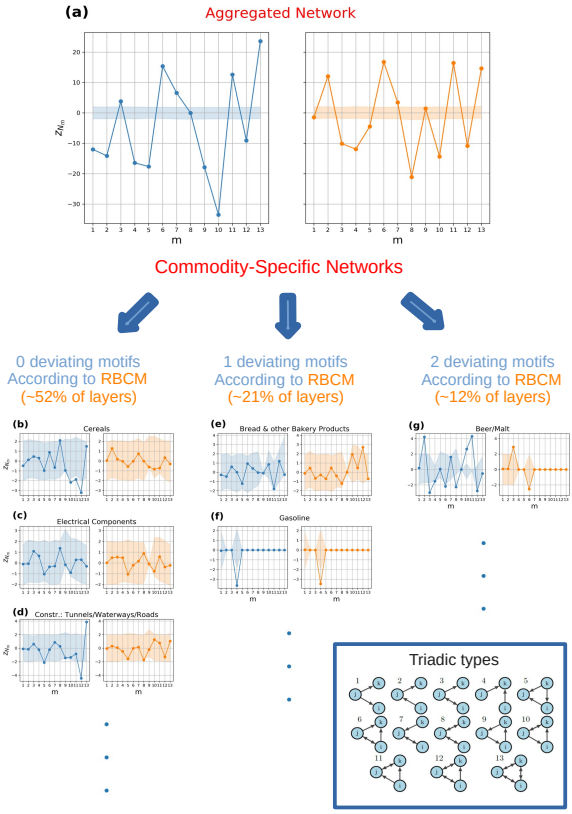

Figure 18: Triadic binary motif analysis: DBCM (●) vs RBCM (●). (a) Analysis of the aggregated network with a single representative commodity. Numerous motifs and anti-motifs are present using DBCM and RBCM as null models. (b-d) Commodity groups where RBCM reproduces all the triadic structures, and they are, respectively, Cereals, Electrical Components, and the Construction of Tunnels, Waterways, and Roads. (e-f) Commodity groups with one network motif, namely Bread and Gasoline. (g) Commodity group with two network motifs, namely Beer/Malt. The CIs are computed by extracting the 2.5-th and 97.5-th percentile from an ensemble distribution of 500 graphs. The numerous motifs and anti-motifs in the aggregated network can be seen as the aggregation of commodity groups presenting very few characteristic patterns.

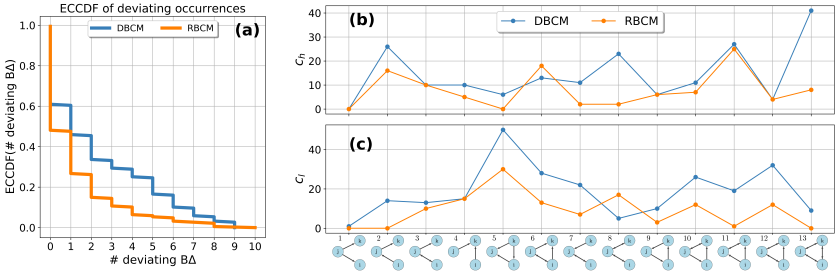

Figure 19: Comparison DBCM (●) vs. RBCM (●): (a) Empirical Counter Cumulative Distribution Function $ECCDF$ of the number of deviating binary triadic motifs and anti-motifs across commodity layers. (b) Number of commodities $c_h(m)$ having a m-type motif (overoccurrence). (c) Number of commodities $c_l(m)$ having a m-type anti-motif (underoccurrence). RBCM explains more triadic structures than DBCM, as shown in the difference of their $ECCDF$. Passing from DBCM to RBCM reduces the number of m motifs across commodities, with the exception of $m = 6$, and anti-motifs, with the exception of $m = 8$. The deviation of those triads is, hence, due to three-node correlations that go beyond directional and reciprocal tendencies of supply/use among industries. RBCM, hence, signals an increased vulnerability to demand shocks originating from the bankrupcty of industries of type k in sub-types $m = 6$ and an increased resiliency to supply/demand shocks of industries of type j in triadic formations $m = 8$.

$p_{ij}^{\leftrightarrow}$, generate a uniform random variable $u_{ij} \in (0, 1)$ and (2) establishing the appropriate links in the dyad in the following way:

- a non-reciprocated link from i to j if $u_{ij} \leq p_{ij}^{\rightarrow}$;
- a non-reciprocated link from j to i if $u_{ij} \in (p_{ij}^{\rightarrow}, p_{ij}^{\rightarrow} + p_{ij}^{\leftarrow}]$;
- a reciprocated link from i to j (and from j to i) if $u_{ij} \in (p_{ij}^{\rightarrow} + p_{ij}^{\leftarrow}, p_{ij}^{\rightarrow} + p_{ij}^{\leftarrow} + p_{ij}^{\leftrightarrow}]$;
- no links from i to j and from j to i otherwise.

In both cases, we generate a realization of A and extract the N_m statistic. $\langle N_m \rangle$ and $\sigma[N_m]$, are the average and standard deviation of N_m extracted from the ensemble distribution of 500 realizations of A. After having computed $z[N_m]$, we also extract the 2.5-th and 97.5-th per-

centiles from the ensemble distribution of N_m over all models and we standardize them using Eq. (5.42) by replacing the empirical N_m with the percentile. Such measures will serve as the 95% CI for the z-score.

The results for the aggregated inter-industry network are in Fig. 18(a). The z-scores computed with respect to the DBCM are depicted in blue on the left panel, while the z-scores computed with respect to the RBCM are depicted in orange on the right panel. The corresponding confidence intervals at the 5% percent are depicted with the same color (blue or orange) but in slight transparency. The majority of N_m are not reproduced by the randomized methods, i.e. the z-scores are outside the confidence intervals. Specifically, only N_8 is reproduced by the DBCM, while both N_1 and N_9 are reproduced by the RBCM. Discounting reciprocal information does not only increase the number of triads that are statistically well described, but potentially changes their type, implying a *qualitatively* different z-score profile. At the same time, in the aggregated picture, $m = 1$ and $m = 9$ are seen as described by a null model implementing reciprocity, i.e. neither high dependency on suppliers ($m = 1$), nor unstable feedback loops ($m = 9$), where industries supply to each other in a cyclical fashion, are revealed. The aggregated network, is hence, characterized by a multitude of structures that are not well described by the null model and are due to additional three-node correlations but is relatively resilient to supply shocks and cyclical input/output. By disaggregating from the aggregated monolayer to the multi-commodity network, the majority of commodity-layers have triadic structures which are statistically reproduced by the reciprocal null model. Only 1 or 2 motifs or anti-motifs are present for the majority of the remaining commodities, a result indicating that beneath the aggregated picture, commodity groups are characterized by a small number of *commodity-specific motifs* and *anti-motifs*.

In Fig. 18(b-d) z-score profiles for three commodity layers are displayed, namely Cereals, Electrical Components, and the Construction of Tunnels, Waterways, and Roads. RBCM well describes all subgraph occurrences (z_{N_m} is within CI), while the DBCM signals the presence of anti-motifs for $m = 10$, $m = 11$ and $m = 12$ for Cereals, and anti-motif

$m = 12$ and motif $m = 13$ for the Construction layer. In Fig. 18(e-f) two z-score profiles are displayed - namely for Bread & other Bakery Products and Gasoline - for which RBCM signals the presence of at least a motif or anti-motif. A motif $m = 12$ is present for the former layer while an anti-motif for $m = 4$ is present for the latter. Notice that for Bread the DBCM does not signal any motif or anti-motif, implying that deviations can emerge by introducing information on the reciprocal structure. Moreover, subgraph $m = 9$ in Bread and the majority of subgraphs in the Gasoline commodity layer are characterized by a degenerate Confidence Interval: in all of the generated synthetic networks $N_{m=9}$ correspond to the empirical N_9^* with null variance, i.e. the constraints imposed on the ensemble totally describe the specific m-type motif, a matter which can arise regardless of the lack of statistics in the related N_m. Finally, in Fig. 18(g) the z-profile for the commodity layer Beer/Malt is considered. The DBCM signals a large number of motifs, specifically for $m = 2$, $m = 10$, and $m = 11$, and anti-motifs for $m = 3$ and $m = 8$. In contrast, the RBCM signals a lone motif $m = 3$ and an anti-motif $m = 6$.
In Fig. 19(a), the empirical counter cumulative distribution for the number of deviating binary triads is shown. Introducing reciprocal structure information reduces the number of motifs and anti-motifs present across commodities. For instance, the percentage of commodities with at least a motif or anti-motif is 61% when compared to the DBCM, and 48% when compared to the RBCM, while the percentage of commodities having at least two motifs or anti-motifs is 46% when compared to the DBCM and 27% when compared to the RBCM.

Lastly, we identify the occurrence of m-type of motifs and anti-motifs across commodities by introducing two quantities, $c_h(m)$ and $c_l(m)$. $c_h(m)$ represents the number of commodities having a motif of type m while $c_l(m)$ represents the same measure for anti-motifs. The addition of the reciprocal structure reduces the number of commodity-specific motifs for each subgraph type, with the exception of motif $m = 6$ as depicted in Fig. 19(b), and the number of anti-motifs for each type, with the exception of anti-motif $m = 8$ as depicted in Fig. 19(c). The reciprocal null model, hence, reveals a higher number of commodities that are relatively

more vulnerable to demand shock due to bankruptcy of industries of type k in triadic formations $m = 6$, while it reveals an increased resilience to supply/demand shocks originating from bankruptcy of industries of type j in formations $m = 8$.

5.4.2 Weighted Motif Analysis

While the bankruptcy of an entire industry is unrealistic, a shock due to a decrease in the flow of goods among industries can propagate along the supply chain, with side effects on the real economy. This implies that not only binary information is important for shock propagation but also weighted information, namely the amount of money circulating on connected structures.

Consider the *triadic flux* F_m on motif m, defined as the total money circulating on triadic subgraphs of type m. We characterize the deviation of empirical F_m to null models by defining the *weighted z-scores* as

$$z\left[F_m\right] = \frac{F_m(W^*) - \langle F_m \rangle}{\sigma\left[F_m\right]} \tag{5.43}$$

where $\langle F_m \rangle$ is the model-induced average amount of money circulating on motif m and $\sigma\left[F_m\right]$ represents the model-induced standard deviation over the ensemble of network realizations.

The theoretical benchmark (or null model) is built by using a combination of binary and conditional weighted models, depending on the wanted constraints. If we deem reciprocal information of negligible importance we should use the combination of models given by DBCM, for the sampling of the binary adjacency matrix, and the CReM$_A$, constraining the out-strength and in-strength sequences. If we deem reciprocal information necessary, a combination of the RBCM and CRWCM should be used. We compare here the two to establish the importance of the addition of reciprocity information for the detection of weighted motifs.

In operative terms, using a two-step model such as the DBCM+CReM$_A$ reduces to (1) establishing a link between industries i and j when a uniform random number $u_{ij} \in U(0,1)$ is such that $u_{ij} \leq p_{ij;DBCM}$, (2) if i

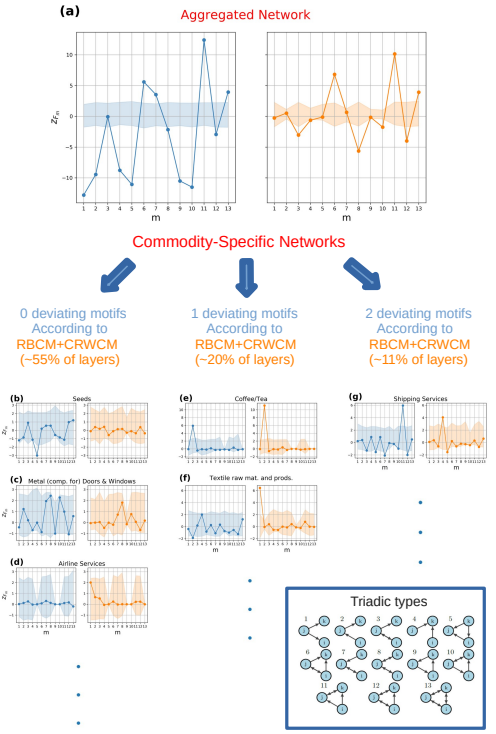

Figure 20: Triadic weighted motif analysis: DBCM+CReM$_A$ (●) vs RBCM+CRWCM (●). (a) Analysis of the aggregated network with a single representative commodity. A large number of motifs and anti-motifs are present when using DBCM+CReM$_A$, while three motifs are present when using the RBCM+CRWCM. (b-d) Commodity groups where RBCM+CRWCM reproduces all the triadic structures, and they are, respectively, Seeds, Metal Components for Doors & Windows, and Airline Services. (e-f) Commodity groups with one network motif, namely Coffee/Tea and Textile raw materials and products. (g) Commodity group with two network motifs, namely Shipping Services. The CIs are computed by extracting the 2.5-th and 97.5-th percentile from an ensemble distribution of 500 graphs. Passing from the aggregated network to the disaggregated product layers unveils the presence of a few commodity-specific motifs and anti-motifs.

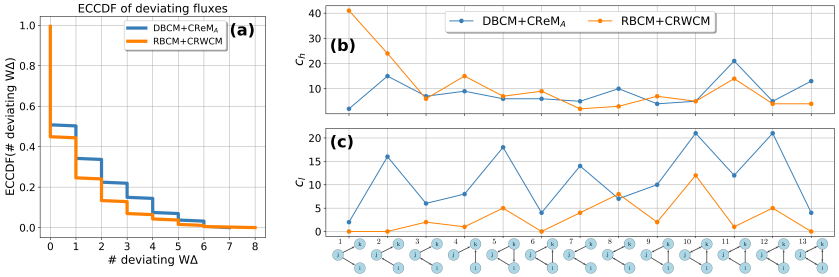

Figure 21: Comparison DBCM+CReM$_A$ (●) vs. RBCM+CRWCM (●): (a) Empirical Counter Cumulative Distribution Function $ECCDF$ of the number of deviating binary triadic motifs and anti-motifs across commodity layers. (b) The number of commodities $c_h(m)$ having a m-type motif. (c) The number of commodities $c_l(m)$ having a m-type anti-motif. RBCM+CRWCM explains slightly more triadic fluxes than DBCM+CReM$_A$, as shown in the difference of their $ECCDF$. Passing from the directed to the reciprocal model reduces the number of anti-motifs, with the exception of $m = 8$. In contrast, it changes qualitatively the motif profile, with a slight dominance of $m = 11$-type motifs when the directed model is used and a clear dominance of $m = 1$-type motifs when the reciprocal model is used. The reciprocated model unveils a vulnerability to supply shocks originating from a decrease in supply volumes of industries of type j in formations $m = 1$.

and j are connected, sampling w_{ij} by using the inverse transform sampling method technique, i.e., we generate a uniformly distributed random variable $\eta_{ij} \in U(0, 1)$ such that

$$F(v_{ij}) = \int_0^{v_{ij}} q_{CReM_A}(w_{ij}|a_{ij} = 1)dw_{ij} = \eta_{ij}, \tag{5.44}$$

then we invert the relationship finding the weight v_{ij} to load on the link (i, j).

The network sampling for the RBCM+CRWCM follows the same concepts with two major differences: (1) a link is established using the RBCM recipe and (2) the dyadic conditional weight probability $q_{CReM_A}(w_{ij}|a_{ij} = 1)$ is substituted with $q_{CRWCM}(w_{ij}|a_{ij} = 1)$ in the inverse transform sampling.

In Fig. 20(a) the z-score profile for the aggregated network with a single representative commodity is depicted using the directed (in blue

on the left panel) or the reciprocal models (in orange on the right panel). There is a large number of motifs and anti-motifs when the benchmark model is directed, only F_3 does not deviate significantly.

When reciprocity information is considered, the picture changes: only three motifs, namely $m = 6$, $m = 11$, and $m = 13$, are identified, and four anti-motifs, namely $m = 3$, $m = 8$, $m = 10$, and $m = 12$, are found when the reciprocal null model is employed. This model's enhanced accuracy unveils a higher-than-expected volume of financial activity on sub-types characterized by a single exclusive user and two suppliers utilizing each other's products ($m = 6$), two users supplying to each other while employing a product from the same supplier ($m = 11$), and entirely cyclical triads ($m = 13$). In contrast, a lower-than-expected level of financial activity transpires in open triads with two reciprocated ties ($m = 8$), one reciprocated link and one exclusive user ($m = 3$), or in closed triads of type $m = 10$ and $m = 12$. While it might be contended that the heightened concentration of funds on $m = 13$ is attributable to aggregation bias, it is crucial to recognize that aggregation solely accounts for the increased monetary worth of the particular sub-type in absolute terms, not for the weighted motif obtained after adjusting for the statistical null model. Please, bear in mind that the emergence of these specific motifs cannot be easily explained without delving into greater detail, given the representative commodity scheme, while the picture cannot be merely reduced to a higher activity on open triads and a lower activity on closed triads.

Similarly to the binary case, passing from the aggregated network to the disaggregated product-level layers, it is possible to identify a small number of *commodity-specific* weighted motifs and anti-motifs.

In Fig. 20(b-d) three commodity layers are depicted for which no motifs and anti-motifs are present when z-scores are computed using the reciprocal model. They are 'Seeds', 'Metal components for Doors & Windows' and 'Airline Services'. In the 'Seeds' layer, the directed model signals the presence of an anti-motif for $m = 5$. In the second layer, no deviations are registered by both null models but CIs are of different nature, in fact, the reciprocal model allows a more restricted range

of z-scores with respect to the directed model for $m = 9$. In the 'Airline Services' layer, for both models, no deviations are present and three CIs are degenerate for $m = 5$, $m = 9$, and $m = 10$. In Fig. 20(e-f) the z-scores relative to the commodity groups 'Coffee/Tea' and 'Textile raw materials and products' are depicted, for which 1 motif is present by using the reciprocal model. For both the directed and reciprocal models there is a weighted motif $m = 2$ in the 'Coffee/Tea' layer. In contrast, in the Textile products layer the directed model signals an anti-motif for $m = 2$, while the reciprocal model signals a motif for $m = 1$. If Fig. 20(g) the z-score profile for the commodity layer 'Shipping Services' is shown: the directed model signals a large number of anti-motifs, specifically for $m = 5$, $m = 7$ and $m = 12$, while it registers a motif for $m = 11$. The reciprocal model, instead, registers a motif for $m = 4$ and anti-motifs for $m = 5$ and $m = 12$. Different commodity layers call for different motifs and anti-motifs which are due to their specific characteristics. In this paper, we refrain from characterizing every single commodity layer, but a specific and thorough analysis is possible by visualizing the number of triadic sub-types, the z-score profile for N_m and their weighted analogs.

The empirical counter cumulative distribution ECCDF(# deviating WΔ) for the number of deviating weighted triads is depicted in Fig. 21(a). The number of deviating triadic fluxes is steadily lower using the reciprocal model. F_m are maximally random for 49% when the directed model benchmark is used and for 55% according to the reciprocal model. The reduction of the number of motifs is however not as significant as in the binary case.

In Fig. 21(b-c) we plot the weighted analogous of $c_h(m)$ and $c_l(m)$. Reciprocal information decreases the occurrence of all types of anti-motifs across commodities, with the exception of $m = 8$. Instead, the profile induced by $c_h(m)$ is *significantly* different using the two null models. For instance, according to the directed model, F_1 is almost always well predicted, instead, it is the most occurring motif according to the reciprocal model. At the same time, reciprocity unveils the dependency of more than 40 commodity layers on the supply of a limited amount of suppliers, which in this case control the market. In fact, the high presence of

$m = 1$ weighted motif signals the vulnerability of the industry-industry network to supply shocks provoked by a reduction of supply volumes.

5.4.3 The NuMeTriS Python package

As an additional result, we release a Python package named NuMeTriS (an acronym stating for 'Null Models for Triadic Structures'), available at `https://github.com/MarsMDK/NuMeTriS`: it contains the routines to implement all the models discussed in the present chapter.

5.5 Discussion

The study of triadic motifs on production networks is still in its infancy due to a scarcity of reliable data. In the existing literature, only binary triadic motifs on one production network, the Japanese one, have been characterized for a single representative commodity [72], while the Hungarian dataset has been analyzed only for triadic occurrences without recurring to a null model [150]. The Japanese study revealed a simple but significant pattern: open triadic subgraphs are over-represented while closed triadic subgraphs are under-represented. This phenomenon was attributed to complementarity, where economic actors connect in tetradic structures - better explained by open triads - due to complementary needs [73].

Our findings corroborate the notion that an analysis based on a single representative commodity is insufficient to fully characterize a production network. Product-level data is *essential* for disaggregating the network into layers that are characterized by commodity-specific binary motifs and anti-motifs. Moreover, we found that the majority of layers exhibit maximally random triadic structures when the reciprocal structure is considered.

At the level of binary motifs, we detected that cyclical reciprocated triadic subgraphs, which are dominant in the aggregated network, break up in the disaggregated product layers, where open triangles become dominant, especially $m = 1$. However, using the RBCM as a bench-

mark, we proved that $m = 1$ is always well described. Conversely, the completely cyclical triads, even if partially broken in the disaggregated layers, are often over-represented compared to the benchmark estimate. In general, constraining the reciprocation capacity of industries - by constraining the reciprocated degrees - is of the foremost importance when characterizing triadic motifs, as explained by the better accuracy and the decrease in binary triadic motifs and anti-motifs when using RBCM as a benchmark compared to DBCM.

We also characterized weighted motifs and anti-motifs, defined as the amount of money circulating on triadic subgraphs, with a novel model which constrains strengths, decomposing them according to the character of the corresponding links. This type of analysis is totally novel in the context of production networks, and rarely seen with benchmark models [141]. We find a non-trivial result already when analyzing the aggregated network, subgraphs that are well explained in binary terms - their occurrence is well described by the statistical ensemble induced by the DBCM or RBCM - can be not well described in weighted terms, meaning that even if a binary triadic subgraph has the expected occurrence it can accommodate an unexpected concentration of money. Furthermore, we identified a high presence of $m = 1$ weighted motifs across commodity layers, a signal of commodity-specific dependency on a limited number of suppliers, which control the market. This implies that a large number of layers are vulnerable to supply shocks, which can arise due to a decrease in supplied volumes (and not only to the supplier's bankruptcy as in the binary case).

Changing the benchmark from a directed to a reciprocal model significantly changes the identity of motifs and anti-motifs across commodities. Hence, it is essential to take into account the type of the corresponding link in which weights are sampled by constraining reciprocated and non-reciprocated strengths.

Overall, our results indicate that product-level information is *strictly* necessary to identify triadic structures and fluxes in production networks. We hope that our study can encourage Statistics Bureaus around the world to implement policies and techniques to reveal or reconstruct a re-

liable product heterogeneity for firm-level transaction data. Our analysis also shows that most commodity-specific layers can be reconstructed via null models that incorporate reciprocity while maintaining dyads independent. For these layers, network reconstruction methods of the type introduced in [103], if extended to incorporate reciprocity, are likely to perform well in replicating the properties of the entire layers starting from partial, node-specific information. Most other layers show at most one or a couple of deviating triadic motifs that are unexplained by the null model. For these layers, additional information is needed to achieve a good reconstruction. Once a rigorous product analysis has been performed, experts in the single commodity can interpret why such triadic formations over-occur or under-occur, accommodating an excessive or insufficient amount of trade volume, unveiling the detailed structure of the commodity-specific production networks.

In order to suggest improvements for further research, we conclude by noticing that our study is subject to two main limitations. First, an industry-level analysis inherently yields results that differ from those obtained from firm-level studies and underestimates the risk associated with exogenous and endogenous shocks [151]. Second, the dataset analyzed pertains to industries at the SBI5 level, a classification intermediate between firms and SBI4-level industries. Analyzing the dataset at the firm level was not feasible due to potential biases arising from the deterministic imputation and degree distribution assumptions, which are exacerbated when dealing with highly granular data. Conversely, analyzing industries at the SBI4 level, which encompasses a maximum of 132 industries, would imply that for a substantial number of commodities, very few industries are active. Consequently, the null model would trivially replicate, in a statistical sense, the triadic structures for the majority of commodity layers due to a lack of relevant observations. However, the same biases anticipated at the firm level can arise, even if mitigated, by selecting SBI5-level industries. This could potentially lead to biases in our analysis, especially in the type of motifs and anti-motifs found for each commodity. However, in order to validate all of our 'fingerprints' we would need fully empirical data for industries at the SBI5 level for

Model	p_{ij}	$\langle w_{ij}\|a_{ij}=1\rangle$
DBCM+CReM$_A$	$\frac{x_i^{out}x_j^{in}}{1+x_i^{out}x_j^{in}}$	$(\beta_i^{out}+\beta_j^{in})^{-1}$
RBCM+CRWCM	$\frac{x_i^{\rightarrow}x_j^{\leftarrow}+x_i^{\leftrightarrow}x_j^{\leftrightarrow}}{x_i^{\rightarrow}x_j^{\leftarrow}+x_i^{\leftarrow}x_j^{\rightarrow}+x_i^{\leftrightarrow}x_j^{\leftrightarrow}+1}$	$w_{ij}^{out,\rightarrow}+w_{ij}^{out,\leftrightarrow}$

Table 12: Models used in this chapter with columns indicating the model-induced probability connection p_{ij} and the weight conditional on the link presence $\langle w_{ij}|a_{ij}=1\rangle$. The parameter $z_{ij}=\exp\{\rho+\beta\ln(\omega_i\omega_j)+\gamma\ln(D_{ij})\}$ stands for the log-link gravity specification, while $\hat{x}_{ij}=\exp\{\theta+\ln(\omega_i\omega_j)-\ln(D_{ij})\}$ stands for the gravity specification used in the logistic binary model.

Model	Type	Binary Constraints	Weighted Constraints
DBCM+CReM$_A$	C-Cont	$\{k_i^{out},k_i^{in}\}$	$\{s_i^{out},s_i^{in}\}$
RBCM+CRWCM	C-Cont	$\{k_i^{\rightarrow},k_i^{\leftarrow},k_i^{\leftrightarrow}\}$	$\{s_i^{\rightarrow,out},s_i^{\rightarrow,in}\}\cap$ $\{s_i^{\leftrightarrow,out},s_i^{\leftrightarrow,in}\}$

Table 13: Entropy-based models used in this chapter with columns indicating the type of induced weights, relative binary constraints and weighted constraints. The type is annotated with C or I for conditional and integrated model respectively, and with Int and Cont for models inducing integer and continuous link weights. Constraints can be topological quantities or their approximation through economic factors. When an approximation in terms of a gravity log-link function is performed, it is indicated with the hat symbol.

each of the 187 commodities, an information that is not available in any country until now, to the best of our knowledge.

In Table 12 we sum up the weighted models explored in this chapter with the specifications related to the model-induced probability connection p_{ij} and to the conditional weight $\langle w_{ij}|a_{ij}=1\rangle$ (conditional on the established link), while in Table 13 the relative structural constraints are portrayed for each model.

Chapter 6

Conclusions

This thesis sums up our efforts to reconcile econometric and physics-inspired approaches for the description of economic systems.

Econometricians are typically interested in defining methods to estimate the unbiased effect of the independent variables, or regressors, onto the dependent one. However, when those methods are used to reconstruct the characteristics of the WTW, they fall short in describing the network statistics of interest [45]: more precisely, they either provide accurate point-wise estimations (as in the case of the ZIP model) or well-reproduce data variability (as in the case of the negative binomial model). On the other hand, we show in chapter 2 and chapter 3 that maximum-entropy models induced by both structural and econometric constraints are capable of achieving both goals, for both discrete-valued and continuous-valued systems. Besides, in the case of discrete systems, maximum-entropy models compete, if not outperform, the aforementioned, econometric recipes in all such aspects.

These methods can be defined either in a conditional or in an integrated fashion, depending on the recipe adopted to sample weights: if the binary probability distribution can be treated as independent from the weighted one, we are considering a conditional model; if the binary probability distribution cannot be treated as independent from the

weighted one, we are considering an integrated model. Interestingly, the best models to accurately predict over 20 years of WTW network statistics are (both the conditional and the integrated version of) the exponential model and the gamma model; according to information criteria, instead, the log-normal model should be preferred - a sort of trade-off that lead us to opt for exponential and gamma models, given their greater stability and an overall similar score to the one achieved by the log-normal model.

A Shannon-Fisher analysis of our models also reveals that the statistical ensemble induced by the exponential model is less susceptible to changes in the value of the weights while the gamma model is characterized by a diverging Fisher Information Measure, an evidence indicating that it is maximally susceptible to changes in the value of the weights: overall, then, in case of noisy data, the exponential model should be preferred.

When considering the problem of estimating parameters of conditional, weighted models, several solutions are viable. The estimation procedure traditionally pursued in econometrics considers a network topology as deterministic, hence ignoring the structural variability induced by the binary probability distribution.

In order to overcome this problem, we devise two alternative procedures in chapter 4, namely the 'annealed' and the 'quenched' one: the first approach estimates the parameters by defining a generalized log-likelihood where the topological randomness is averaged over the allowed binary configurations; the second approach estimates the parameters on each, sampled (binary) configuration, averaging them *a posteriori*. The two procedures are equivalent for models with a global constraint on weights and coincide with the 'deterministic' estimate. When considering models either constraining local properties or defined by econometric quantities, the two procedures lead to estimates that strongly depend on the adopted, binary model, with the 'annealed' and 'quenched' estimates that coincide while being significantly different from the 'deterministic' one.

Moreover, in case sparse graphs are considered, the 'quenched' approach can lead to biased estimates for nodes with a low degree - in fact, the ones that are poorly-connected in the empirical network can be disconnected in single realizations: whenever this happens, it is not possible to estimate the corresponding parameter. This problem leads us to prefer the 'annealed' approach which does not require any explicit sampling.

In chapter 5, we explore the triadic structure of the multi-commodity, inter-industry production network constructed by CBS-Statistics Netherlands. As information about domestic production networks is rarely available, the dataset our analysis is based upon represent a quite unique example in the current literature. The analysis of the binary version of this network reveals most layers to be characterized by an abundance of the 13 triadic motifs that is maximally-random when reciprocal and non-reciprocal degrees are constrained - an evidence showing that information on reciprocity must be accounted for by any model designed to reproduce third-order quantities. On the other hand, the analysis of the weighted version of this network reveals that models constraining reciprocity identify statistically significant structures that are quite different from those identified by models not constraining this information.

Overall, our results indicate that product-level information is strictly necessary to identify triadic structures and fluxes in production networks and that they can be effectively described by models that incorporate both individual and reciprocal tendencies for link establishment and supply/use for most commodities. We hope that our study can encourage Statistics Bureaus around the world to implement policies and techniques to reveal, or reconstruct, a reliable product heterogeneity for firm-level transaction data. Once a rigorous product analysis has been performed, experts in the single commodity can interpret why such triadic formations over-occur or under-occur, accommodating an excessive or insufficient amount of money, unveiling the detailed structure of the commodity-specific production networks.

While the present thesis offers methodologies to reconcile Network Science and Econometrics, the methods can be further enhanced. For

instance, as observed in chapter 2, Maximum Entropy (ME) methods with Gravity-like specifications exhibit heteroscedasticity. A simple solution would be incorporating a White Variance-Covariance matrix for the weights. However, we propose a more comprehensive approach by expanding the models to panel data, specifically adopting the Kullback-Leibler (KL) formalism for Temporal Networks with and without memory. This would provide econometricians with an information-theoretic alternative to the widely used Poisson Pseudo-Maximum Likelihood with dyadic fixed-effects for panel data.

Another potential area for improvement lies in the methods introduced in chapter 5, including the reciprocal model composed of the Reciprocal Binary Configuration Model (RBCM) for the binary part and the Conditionally Reciprocal Weighted Model (CRWCM) for the weighted part. We would like to investigate whether the CRWCM could be further refined by incorporating constraints on macro production functions, following the approach introduced in [103]. This modification would enhance the method's accuracy and enable the identification of triadic structures that are not adequately explained by individual or reciprocal tendencies but rather by production capabilities. Conversely, the reciprocal model could be simplified by employing a Gravity log-link specification instead of node-specific parameters.

Additionally, this thesis elucidates the process of transitioning from a Maximum-Entropy structural model to a more econometric-oriented model by approximating Lagrange parameters with constant and economic factors. This results in a more parsimonious model, as evidenced by reduced AIC and BIC values, without significantly lose overall performance.

Finally, chapter 4 examines the bias introduced in parameter estimation when the conventional deterministic approach is implemented in conditional models. This deterministic approach, widely used in hurdle models, is theoretically sound under the assumption of complete independence between the binary and weight distributions. However, it also entails the presumption that the empirical realization is the sole determinant, overlooking all other realizations generated by the chosen

binary model even though the weights are derived after sampling connections from the binary distribution. Our findings indicate that this approach results in parameter estimates that diverge from those obtained employing the annealed approach, which considers the entire probability matrix induced by the binary model, and are inconsistent with the ensemble distribution of parameter estimates produced by the quenched approach. Consequently, we advocate for the adoption of the annealed approach over the traditional deterministic one to reestablish sampling consistency.

Appendix A

Discrete-valued models

Appendix A, based on [1], further expands on the models presented in chapter 2. The first section discusses the estimation of the parameters defining the classical gravity model via the LS and PPML techniques. The second section discusses the estimation of the parameters defining the negative binomial model, the Poisson model and their zero-inflated counterparts via the likelihood maximization technique. The third section focuses on the discrete-valued distributions arising from the maximum-entropy principle and illustrates the set of first-order conditions to be satisfied for maximizing the corresponding log-likelihood function.

A.1 Estimating the GM parameters

Here, we consider two different specifications of the GM, i.e.

$$\langle w_{ij} \rangle_{\text{GM}}^{(1)} = \rho(\omega_i \omega_j) d_{ij}^{-1} \tag{A1}$$

and

$$\langle w_{ij} \rangle_{\text{GM}}^{(2)} = \rho(\omega_i \omega_j)^{\beta} d_{ij}^{\gamma}; \tag{A2}$$

the parameters appearing in both specifications of the GM can be estimated by implementing a Non linear-Least-Squares (NLS) regression, i.e. by solving the optimization problem

$$\arg\min_{\underline{\theta}} \left\{ \sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\mathrm{GM}} \right]^2 \right\} \tag{A3}$$

which, in turn, translates into solving the equation

$$\sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\mathrm{GM}} \right] \frac{\partial \langle w_{ij} \rangle_{\mathrm{GM}}}{\partial \theta_i} = 0, \quad \forall\, i; \tag{A4}$$

however, such a procedure is known to produce biased estimations. In fact, it leads to the set of conditions

$$\sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\mathrm{GM}}^{(1)} \right] \langle w_{ij} \rangle_{\mathrm{GM}} = 0 \tag{A5}$$

for the first GM specification and

$$\begin{cases} \sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\mathrm{GM}}^{(2)} \right] \langle w_{ij} \rangle_{\mathrm{GM}} & = 0 \\ \sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\mathrm{GM}}^{(2)} \right] \langle w_{ij} \rangle_{\mathrm{GM}} \ln(\omega_i \omega_j) & = 0 \\ \sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\mathrm{GM}}^{(2)} \right] \langle w_{ij} \rangle_{\mathrm{GM}} \ln d_{ij} & = 0 \end{cases} \tag{A6}$$

for the second GM specification. Since the conditions above are known to 'weigh' more larger weights, Silva and Tenreyro [42] propose to employ a Poisson Pseudo-Maximum Likelihood (PPML) estimator, leading to the set of conditions

$$\sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\mathrm{GM}}^{(1)} \right] = 0 \tag{A7}$$

for the first GM specification and

$$\begin{cases} \sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\mathrm{GM}}^{(2)} \right] & = 0 \\ \sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\mathrm{GM}}^{(2)} \right] \ln(\omega_i \omega_j) & = 0 \\ \sum_{i<j} \left[w_{ij} - \langle w_{ij} \rangle_{\mathrm{GM}}^{(2)} \right] \ln d_{ij} & = 0 \end{cases} \tag{A8}$$

for the second GM specification. Remarkably, the PPML estimator lets the purely topological condition $W = \sum_{i<j} w_{ij} = \sum_{i<j} \langle w_{ij} \rangle_{\mathrm{GM}} = \langle W \rangle_{\mathrm{GM}}$ (i.e. the preservation of the total weight) to be recovered for both GM specifications.

A.2 Econometric models

This section is devoted to the detailed description of the econometric models considered in the present work. When coming to estimate the parameters entering into the definition of any of the econometric models considered above, we invoke the maximum-of-the-likelihood principle, prescribing to maximize the function

$$\mathcal{L} = \ln Q(\mathbf{W}) \tag{A9}$$

where the optimization is carried out with respect to the set of parameters characterizing each specific model.

In the present contribution we have considered undirected networks, hence the parameters associated to node-specific regressors of the same economic variable (e.g. the GDPs of the country of origin and destination) are equal. In such a framework we have employed the specifications $z_{ij} = \rho(\omega_i\omega_j)^\beta d_{ij}^\gamma$ and $G_{ij} = \delta(\omega_i\omega_j)$ where $\omega_i = \frac{\text{GDP}_i}{\overline{\text{GDP}}}$, i.e. the GDP of each country is divided by the mean value of all GDPs. We would also like to stress that the parameters entering into the definition of z_{ij} are equal to those employed for the standard GM specification, i.e. $z_{ij} = e^{\underline{X}\cdot\underline{\theta}}$, as evident upon taking as regressors the natural logarithm of the GDPs and that of the geographic distance, i.e.

$$z_{ij} = e^{\underline{X}\cdot\underline{\theta}} = e^{\beta \ln \omega_i + \beta \ln \omega_j + \gamma \ln d_{ij} + c} = \rho(\omega_i\omega_j)^\beta d_{ij}^\gamma \tag{A10}$$

having posed $c = \ln \rho$.

A.2.1 Poisson model

The probability mass function of the Poisson model reads

$$q_{ij}^{\text{Pois}}(w_{ij}) = \frac{z_{ij}^{w_{ij}} e^{-z_{ij}}}{w_{ij}!}; \tag{A11}$$

where $z_{ij} = \rho(\omega_i\omega_j)^\beta d_{ij}^\gamma$. In what follows, we will substitute $w_{ij}!$ with $\Gamma[w_{ij}+1]$, as routinely done in the packages for solving econometric models. Its log-likelihood reads

$$\mathcal{L}_{\text{Pois}} = \sum_{i<j} [w_{ij} \ln z_{ij} - z_{ij} - \ln \Gamma[w_{ij} + 1]] \tag{A12}$$

whose optimization leads to the set of equations

$$\sum_{i<j} \left[\frac{w_{ij}}{z_{ij}} - 1 \right] \frac{\partial z_{ij}}{\partial \theta_i} = 0, \quad \forall\, i \tag{A13}$$

reading, more explicitly,

$$\begin{cases} \sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\text{Pois}}] & = 0 \\ \sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\text{Pois}}] \ln(\omega_i \omega_j) & = 0 \\ \sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{\text{Pois}}] \ln d_{ij} & = 0; \end{cases} \tag{A14}$$

notice that the set of equations above coincides with eqs. (A8). The Poisson model remains the most used one because it ensures that a network total weight is reproduced - a desirable feature to correctly estimate the weights of a network, as also observed for the 'plain' GM.

A.2.2 Negative binomial model

The probability mass function of the Negative Binomial model reads

$$q_{ij}^{\text{NB}}(w_{ij}) = \binom{m + w_{ij} - 1}{w_{ij}} \left(\frac{1}{1 + \alpha z_{ij}} \right)^m \left(\frac{\alpha z_{ij}}{1 + \alpha z_{ij}} \right)^{w_{ij}} \tag{A15}$$

where $\alpha = m^{-1}$ to handle overdispersion and $z_{ij} = \rho(\omega_i \omega_j)^\beta d_{ij}^\gamma$. By replacing each binomial coefficient with the corresponding Gamma function, one recovers the expression

$$\begin{aligned} \mathcal{L}_{\text{NB}} = \sum_{i<j} [& w_{ij} \ln(\alpha z_{ij}) - (m + w_{ij}) \ln(1 + \alpha z_{ij}) \\ & + \ln \Gamma[m + w_{ij}] - \ln \Gamma[w_{ij} + 1] - \ln \Gamma[m]]; \end{aligned} \tag{A16}$$

its optimization leads to the set of equations

$$\sum_{i<j} \left[\frac{w_{ij} - z_{ij}}{z_{ij}(1 + \alpha z_{ij})} \right] \frac{\partial z_{ij}}{\partial \theta_i} = 0, \quad \forall\, i \tag{A17}$$

and

$$\sum_{i<j}\left[\frac{w_{ij}-z_{ij}}{\alpha(1+\alpha z_{ij})}-m^2\left(\frac{\Gamma^{'}[m+w_{ij}]}{\Gamma[m+w_{ij}]}-\frac{\Gamma^{'}[m]}{\Gamma[m]}\right)\right]=0 \tag{A18}$$

reading, more explicitly,

$$\begin{cases}\sum_{i<j}\left[\frac{w_{ij}-\langle w_{ij}\rangle_{\mathrm{NB}}}{\mathrm{Var}_{\mathrm{NB}}[w_{ij}]}\right] & =0\\ \sum_{i<j}\left[\frac{w_{ij}-\langle w_{ij}\rangle_{\mathrm{NB}}}{\mathrm{Var}_{\mathrm{NB}}[w_{ij}]}\right]\ln(\omega_i\omega_j) & =0\\ \sum_{i<j}\left[\frac{w_{ij}-\langle w_{ij}\rangle_{\mathrm{NB}}}{\mathrm{Var}_{\mathrm{NB}}[w_{ij}]}\right]\ln d_{ij} & =0\\ \sum_{i<j}\left[\frac{w_{ij}-\langle w_{ij}\rangle_{\mathrm{NB}}}{\alpha(1+\alpha z_{ij})}-m^2\left(\frac{\Gamma^{'}[m+w_{ij}]}{\Gamma[m+w_{ij}]}-\frac{\Gamma^{'}[m]}{\Gamma[m]}\right)\right] & =0.\end{cases} \tag{A19}$$

Notice that the negative binomial model does not constrain the total weight of a network, whence its bad performance in reproducing the other weighted structural properties.

A.2.3 Zero-inflated Poisson model

The zero-inflated version of the Poisson model is defined by a functional form reading

$$\begin{aligned}Q(\mathbf{W}) &= \prod_{i<j} q_{ij}(w_{ij})\\ &= \prod_{i<j} p_{ij}^{a_{ij}}(1-p_{ij})^{1-a_{ij}}\cdot q_{ij}(w_{ij}|a_{ij}) = P(\mathbf{A})Q(\mathbf{W}|\mathbf{A})\end{aligned} \tag{A20}$$

and inducing the following log-likelihood

$$\begin{aligned}\mathcal{L} &= \ln Q(\mathbf{W})\\ &= \ln P(\mathbf{A}) + \ln Q(\mathbf{W}|\mathbf{A})\\ &= \sum_{i<j}[a_{ij}\ln p_{ij} + (1-a_{ij})\ln(1-p_{ij}) + \ln q_{ij}(w_{ij}|a_{ij})]\end{aligned} \tag{A21}$$

where

$$p_{ij}^{\mathrm{ZIP}} = \frac{G_{ij}}{1+G_{ij}}(1-e^{-z_{ij}}), \tag{A22}$$

$$q_{ij}^{\mathrm{ZIP}}(w_{ij}|a_{ij}) = \left[\frac{z_{ij}^{w_{ij}}e^{-z_{ij}}}{(1-e^{-z_{ij}})w_{ij}!}\right]^{a_{ij}} \tag{A23}$$

with $z_{ij} = \rho(\omega_i\omega_j)^\beta d_{ij}^\gamma$ and $G_{ij} = \delta\omega_i\omega_j$; hence,

$$\begin{aligned}\mathcal{L}_{\text{ZIP}} = \sum_{i<j}[&a_{ij}\ln G_{ij} - a_{ij}\ln\big(1+G_{ij}e^{-z_{ij}}\big) + \ln\big(1+G_{ij}e^{-z_{ij}}\big) - \ln(1+G_{ij}) \\ &+w_{ij}\ln z_{ij} - a_{ij}z_{ij} - a_{ij}\ln\Gamma[w_{ij}+1]].\end{aligned} \tag{A24}$$

Its optimization leads to the set of equations

$$\begin{cases} \sum_{i<j}\left[\frac{a_{ij}-p_{ij}^{\text{ZIP}}}{1+G_{ij}e^{-z_{ij}}}\right] & = 0 \\ \sum_{i<j}\left[w_{ij} - \left(\frac{a_{ij}+G_{ij}e^{-z_{ij}}}{1+G_{ij}e^{-z_{ij}}}\right)z_{ij}\right] & = 0 \\ \sum_{i<j}\left[w_{ij} - \left(\frac{a_{ij}+G_{ij}e^{-z_{ij}}}{1+G_{ij}e^{-z_{ij}}}\right)z_{ij}\right]\ln(\omega_i\omega_j) & = 0 \\ \sum_{i<j}\left[w_{ij} - \left(\frac{a_{ij}+G_{ij}e^{-z_{ij}}}{1+G_{ij}e^{-z_{ij}}}\right)z_{ij}\right]\ln d_{ij} & = 0 \end{cases} \tag{A25}$$

with a clear meaning of the symbols.

A.2.4 Zero-inflated negative binomial model

As for the ZIP model, the zero-inflated version of the negative binomial model induces a log-likelihood reading

$$\begin{aligned}\mathcal{L} &= \ln Q(\mathbf{W}) \\ &= \ln P(\mathbf{A}) + \ln Q(\mathbf{W}|\mathbf{A}) \\ &= \sum_{i<j}[a_{ij}\ln p_{ij} + (1-a_{ij})\ln(1-p_{ij}) + \ln q_{ij}(w_{ij}|a_{ij})]\end{aligned} \tag{A26}$$

where

$$p_{ij}^{\text{ZINB}} = \frac{G_{ij}}{1+G_{ij}}(1-\tau_{ij}), \tag{A27}$$

$$q_{ij}^{\text{ZINB}}(w_{ij}|a_{ij}) = \binom{m+w_{ij}-1}{w_{ij}}^{a_{ij}}\left(\frac{\tau_{ij}}{1-\tau_{ij}}\right)^{a_{ij}}\left(\frac{\alpha z_{ij}}{1+\alpha z_{ij}}\right)^{w_{ij}} \tag{A28}$$

with $\alpha = m^{-1}$, $\tau_{ij} = \left(\frac{1}{1+\alpha z_{ij}}\right)^m$, $z_{ij} = \rho(\omega_i\omega_j)^\beta d_{ij}^\gamma$ and $G_{ij} = \delta\omega_i\omega_j$. Hence,

$$\begin{aligned}\mathcal{L}_{\text{ZINB}} = \sum_{i<j}[&a_{ij}\ln G_{ij} - a_{ij}\ln(1+G_{ij}\tau_{ij}) + \ln(1+G_{ij}\tau_{ij}) - \ln(1+G_{ij}) \\ &+w_{ij}\ln(\alpha z_{ij}) - w_{ij}\ln(1+\alpha z_{ij}) - m a_{ij}\ln(1+\alpha z_{ij}) + a_{ij}\ln\Gamma[m+w_{ij}] \\ &-a_{ij}\ln\Gamma[w_{ij}+1] - a_{ij}\ln\Gamma[m]]\end{aligned} \tag{A29}$$

whose optimization leads to the set of equations

$$\begin{cases} \sum_{i<j}\left[\frac{a_{ij}-p_{ij}^{\text{ZINB}}}{1+G_{ij}\tau_{ij}}\right] & = 0 \\ \sum_{i<j}\left[w_{ij} - \left(\frac{a_{ij}+G_{ij}\tau_{ij}}{1+G_{ij}\tau_{ij}}\right)\right]\left(\frac{z_{ij}}{1+\alpha z_{ij}}\right) & = 0 \\ \sum_{i<j}\left[w_{ij} - \left(\frac{a_{ij}+G_{ij}\tau_{ij}}{1+G_{ij}\tau_{ij}}\right)\right]\left(\frac{z_{ij}}{1+\alpha z_{ij}}\right)\ln(\omega_i\omega_j) & = 0 \\ \sum_{i<j}\left[w_{ij} - \left(\frac{a_{ij}+G_{ij}\tau_{ij}}{1+G_{ij}\tau_{ij}}\right)\right]\left(\frac{z_{ij}}{1+\alpha z_{ij}}\right)\ln d_{ij} & = 0 \\ \sum_{i<j}\left(\frac{a_{ij}+G_{ij}\tau_{ij}}{1+G_{ij}\tau_{ij}}\right)\left(m^2\ln(1+\alpha z_{ij}) - \frac{m z_{ij}}{1+\alpha z_{ij}}\right) + & \\ +\sum_{i<j}\left[\frac{w_{ij}}{\alpha(1+\alpha z_{ij})} - m^2 a_{ij}\left(\frac{\Gamma^{'}[m+w_{ij}]}{\Gamma[m+w_{ij}]} - \frac{\Gamma^{'}[m]}{\Gamma[m]}\right)\right] & = 0. \end{cases} \tag{A30}$$

A.3 Maximum-entropy models

This section is devoted to the detailed description of the maximum-entropy models considered in the present work. As for the econometric ones, the estimation of the set of parameters defining each model is carried out by maximizing the corresponding log-likelihood function, $\mathcal{L} = \ln Q(\mathbf{W})$.

ME models are derived by maximizing Shannon entropy under a suitable set of constraints. The generic probability mass function reads

$$\begin{aligned} Q(\mathbf{W}) &= \prod_{i<j} q_{ij}(w_{ij}) \\ &= \prod_{i<j} p_{ij}^{a_{ij}}(1-p_{ij})^{1-a_{ij}} \cdot q_{ij}(w_{ij}|a_{ij}) \\ &= \prod_{i<j} p_{ij}^{a_{ij}}(1-p_{ij})^{1-a_{ij}} \cdot y_{ij}^{w_{ij}-a_{ij}}(1-y_{ij})^{a_{ij}} = P(\mathbf{A})Q(\mathbf{W}|\mathbf{A}) \end{aligned} \tag{A31}$$

and induces the log-likelihood

$$\begin{aligned}\mathcal{L} &= \ln Q(\mathbf{W}) \\ &= \ln P(\mathbf{A}) + \ln Q(\mathbf{W}|\mathbf{A}) \\ &= \sum_{i<j}[a_{ij}\ln p_{ij} + (1-a_{ij})\ln(1-p_{ij}) + \ln q_{ij}(w_{ij}|a_{ij})]. \end{aligned} \tag{A32}$$

In order to turn maximum-entropy models into proper econometric ones, we have posed $\frac{y_{ij}}{1-y_{ij}} = z_{ij} = \rho(\omega_i\omega_j)^\beta d_{ij}^\gamma$, with $\omega_i = \frac{\text{GDP}_i}{\overline{\text{GDP}}}$.

Let us start instantiating the ME formalism by considering the Hamiltonian

$$H_{(1)}(\mathbf{W}) = \theta_0 L + \psi_0 W + \sum_{i<j}\psi_{ij}w_{ij} \tag{A33}$$

that leads to

$$\begin{aligned}\mathcal{L}_{(1)} = & L\ln x + W\ln y_0 \\ & + \sum_{i<j}[w_{ij}\ln z_{ij} - w_{ij}\ln(1+z_{ij}) + \ln(1+z_{ij}-y_0 z_{ij}) \\ & - \ln(1+z_{ij}-y_0 z_{ij}+xy_0 z_{ij})]\end{aligned} \tag{A34}$$

whose optimization leads to solve the following set of equations

$$\begin{cases} \sum_{i<j}[a_{ij} - p_{ij}^{(1)}] & = 0 \\ \sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{(1)}] & = 0 \\ \sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{(1)}]\left(\frac{1}{1+z_{ij}}\right) & = 0 \\ \sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{(1)}]\left(\frac{\ln(\omega_i\omega_j)}{1+z_{ij}}\right) & = 0 \\ \sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{(1)}]\left(\frac{\ln d_{ij}}{1+z_{ij}}\right) & = 0 \end{cases} \tag{A35}$$

the first equation guaranteeing that the total number of links is preserved, i.e.

$$L = \langle L\rangle = \sum_{i<j}\frac{xy_0 z_{ij}}{1+z_{ij}-y_0 z_{ij}+xy_0 z_{ij}} = \sum_{i<j}p_{ij}^{(1)} \tag{A36}$$

and the second equation guaranteeing that the total weight is preserved, i.e.

$$W = \langle W \rangle = \sum_{i<j} \frac{p_{ij}^{(1)}(1+z_{ij})}{1+z_{ij}-y_0 z_{ij}} = \sum_{i<j} \langle w_{ij} \rangle_{(1)}. \tag{A37}$$

On the other hand, the Hamiltonian

$$H_{(2)}(\mathbf{W}) = \sum_i \theta_i k_i + \psi_0 W + \sum_{i<j} \psi_{ij} w_{ij} \tag{A38}$$

leads to the expression

$$\begin{aligned} \mathcal{L}_{(2)} = & \sum_i k_i \ln x_i + W \ln y_0 \\ & + \sum_{i<j} [w_{ij} \ln z_{ij} - w_{ij} \ln(1+z_{ij}) + \ln(1+z_{ij}-y_0 z_{ij}) \\ & - \ln(1+z_{ij}-y_0 z_{ij}+x_i x_j y_0 z_{ij})] \end{aligned} \tag{A39}$$

whose optimization requires to solve the set of equations below

$$\begin{cases} \sum_{j(\neq i)} [a_{ij} - p_{ij}^{(2)}] & = 0, \quad \forall\, i \\ \sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{(2)}] & = 0 \\ \sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{(2)}] \left(\frac{1}{1+z_{ij}} \right) & = 0 \\ \sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{(2)}] \left(\frac{\ln(\omega_i \omega_j)}{1+z_{ij}} \right) & = 0 \\ \sum_{i<j} [w_{ij} - \langle w_{ij} \rangle_{(2)}] \left(\frac{\ln d_{ij}}{1+z_{ij}} \right) & = 0 \end{cases} \tag{A40}$$

the first set of N equations guaranteeing that each degree is preserved, i.e.

$$k_i = \langle k_i \rangle = \sum_{j(\neq i)} \frac{x_i x_j y_0 z_{ij}}{1+z_{ij}-y_0 z_{ij}+x_i x_j y_0 z_{ij}} = \sum_{j(\neq i)} p_{ij}^{(2)} \tag{A41}$$

and the $(N+1)$-th equation guaranteeing that the total weight is preserved, i.e.

$$W = \langle W \rangle = \sum_{i<j} \frac{p_{ij}^{(2)}(1+z_{ij})}{1+z_{ij}-y_0 z_{ij}} = \sum_{i<j} \langle w_{ij} \rangle_{(2)}. \tag{A42}$$

The factorized probability mass function $Q(\mathbf{W}) = P(\mathbf{A})Q(\mathbf{W}|\mathbf{A})$ allows us to design two-step models, i.e. network models whose binary

estimation step can be defined independently from the weighted one. To this aim, let us combine the probability distribution of the Undirected Binary Configuration Model, inducing a single-link probability coefficient reading $p_{ij}^{\text{UBCM}} = \frac{x_i x_j}{1+x_i x_j}$, with the usual weighted, conditional one, i.e. $q_{ij}(w_{ij}|a_{ij}) = \left(\frac{y_0 z_{ij}}{1+z_{ij}}\right)^{w_{ij}-a_{ij}} \cdot \left(\frac{1+z_{ij}-y_0 z_{ij}}{1+z_{ij}}\right)^{a_{ij}}$. As a result, one obtains

$$\begin{aligned}\mathcal{L}_{\text{TS}} = &\sum_i k_i \ln x_i + (W-L)\ln y_0 \\ &+ \sum_{i<j}[-\ln(1+x_i x_j) + (w_{ij}-a_{ij})\ln z_{ij} - w_{ij}\ln(1+z_{ij}) \\ &+ a_{ij}\ln(1+z_{ij}-y_0 z_{ij})]\end{aligned} \tag{A43}$$

whose optimization requires to solve the set of equations below

$$\begin{cases} \sum_{j(\neq i)}[a_{ij} - p_{ij}^{\text{TS}}] & = 0, \quad \forall\, i \\ \sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{\text{TS}}] & = 0 \\ \sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{\text{TS}}]\left(\frac{1}{1+z_{ij}}\right) & = 0 \\ \sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{\text{TS}}]\left(\frac{\ln(\omega_i\omega_j)}{1+z_{ij}}\right) & = 0 \\ \sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{\text{TS}}]\left(\frac{\ln d_{ij}}{1+z_{ij}}\right) & = 0 \end{cases} \tag{A44}$$

the first set of N equations guaranteeing that each degree is preserved, i.e.

$$k_i = \langle k_i\rangle = \sum_{j(\neq i)} \frac{x_i x_j}{1+x_i x_j} = \sum_{j(\neq i)} p_{ij}^{\text{TS}} = \sum_{j(\neq i)} p_{ij}^{\text{UBCM}} \tag{A45}$$

and the $(N+1)$-th equation guaranteeing that the total conditional weight is preserved, i.e.

$$W = \langle W\rangle = \sum_{i<j} \frac{a_{ij}(1+z_{ij})}{1+z_{ij}-y_0 z_{ij}} = \sum_{i<j}\langle w_{ij}\rangle_{\text{TS}}. \tag{A46}$$

The amount of structural information to constrain can be further reduced by imagining a two-step model whose binary estimation step encodes a larger amount of 'economic' information. To this aim, let us combine the probability distribution of the density-corrected Gravity Model,

inducing a single-link probability coefficient reading $p_{ij}^{\text{TSF}} = \frac{G_{ij}}{1+G_{ij}}$, with the usual weighted, conditional one, i.e. $q_{ij}(w_{ij}|a_{ij}=1) = \left(\frac{y_0 z_{ij}}{1+z_{ij}}\right)^{w_{ij}-a_{ij}} \cdot \left(\frac{1+z_{ij}-y_0 z_{ij}}{1+z_{ij}}\right)^{a_{ij}}$. As a result, one obtains

$$\begin{aligned}\mathcal{L}_{\text{TSF}} = \sum_{i<j} [&a_{ij} \ln G_{ij} - \ln(1+G_{ij}) + (W-L)\ln y_0 \\ &+(w_{ij}-a_{ij})\ln z_{ij} - w_{ij}\ln(1+z_{ij}) + a_{ij}\ln(1+z_{ij}-y_0 z_{ij})]\end{aligned} \tag{A47}$$

where $G_{ij} = \delta\omega_i\omega_j$. Its optimization leads to the resolution of the following set of equations

$$\begin{cases} \sum_{i<j}[a_{ij} - p_{ij}^{\text{TSF}}] & = 0 \\ \sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{\text{TSF}}] & = 0 \\ \sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{\text{TSF}}]\left(\frac{1}{1+z_{ij}}\right) & = 0 \\ \sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{\text{TSF}}]\left(\frac{\ln(\omega_i\omega_j)}{1+z_{ij}}\right) & = 0 \\ \sum_{i<j}[w_{ij} - \langle w_{ij}\rangle_{\text{TSF}}]\left(\frac{\ln d_{ij}}{1+z_{ij}}\right) & = 0 \end{cases} \tag{A48}$$

the first equation guaranteeing that the total number of links is preserved, i.e.

$$L = \langle L\rangle = \sum_{j(\neq i)} \frac{G_{ij}}{1+G_{ij}} = \sum_{j(\neq i)} p_{ij}^{\text{TSF}} = \sum_{j(\neq i)} p_{ij}^{\text{dcGM}} \tag{A49}$$

and the second equation guaranteeing that the total conditional weight is preserved, i.e.

$$W = \langle W\rangle = \sum_{i<j} \frac{a_{ij}(1+z_{ij})}{1+z_{ij}-y_0 z_{ij}} = \sum_{i<j} \langle w_{ij}\rangle_{\text{TSF}}. \tag{A50}$$

Appendix B

Continuous-valued models

Appendix B, based on [2], further expands on the models presented in chapter 3. The first section discusses the so-called conditional models, characterized by a separate estimation of purely structural and weighted parameters: the corresponding log-likelihood as well as the set of first-order conditions required to maximize it are illustrated therein. The second section discusses the so-called integrated models, characterized by a joint estimation of purely structural and weighted parameters. The third section focuses on the model-dependent rules to enrich (otherwise) purely structural models with economic factors. The fourth section derives the expressions of the dyadic Shannon entropy and Fisher Information Measure for each, conditional model which are, then, used to construct the related Shannon-Fisher plane.

B.1 Conditional models

Any member of the class of conditional models is described by the expression

$$Q(\mathbf{W}|\mathbf{A}) = \frac{e^{-H(\mathbf{W})}}{\int_{\mathbb{W}_\mathbb{A}} e^{-H(\mathbf{W})} d\mathbf{W}} \tag{B1}$$

the single instances being characterized by different expressions of the Hamiltonian. Since, however, each Hamiltonian considered here is a sum

over the node pairs, the result

$$Q(\mathbf{W}|\mathbf{A}) = \frac{e^{-\sum_{i<j} H_{ij}(w_{ij})}}{\int_{\mathbb{W}_{\mathbb{A}}} e^{-\sum_{i<j} H_{ij}(w_{ij})} d\mathbf{W}} = \prod_{i<j} \frac{e^{-H_{ij}(w_{ij})}}{\left[\int_{m_{ij}}^{+\infty} e^{-H_{ij}(w_{ij})} dw_{ij}\right]^{a_{ij}}} = \prod_{i<j} \frac{e^{-H_{ij}(w_{ij})}}{\zeta_{ij}^{a_{ij}}} \tag{B2}$$

holds irrespectively from the specific functional form of $H_{ij}(w_{ij})$. Notice that m_{ij} is the minimum, pair-specific weight allowed by the model. The identifications

$$H_{ij}(w_{ij}) \equiv f(w_{ij}|\underline{\theta}, z_{ij}(\underline{\psi})) \tag{B3}$$

and

$$\mathcal{L} \equiv \ln Q(\mathbf{W}|\mathbf{A}) \tag{B4}$$

lead to estimate parameters by solving the (coupled) systems of purely structural equations $\frac{\partial \mathcal{L}}{\partial \underline{\theta}} = \underline{0}$ and econometric-like equations $\frac{\partial \mathcal{L}}{\partial \underline{\tau}} = \underline{0}$, i.e.

$$\begin{cases} \sum_{i<j|a_{ij}=1} \left[\frac{\partial f(w_{ij}|\underline{\theta}, z_{ij}(\underline{\psi}))}{\partial \underline{\theta}} + \frac{1}{\zeta_{ij}} \left(\frac{\partial \zeta_{ij}}{\partial \underline{\theta}} \right) \right] & = \underline{0} \\ \sum_{i<j|a_{ij}=1} \left[\frac{\partial f(w_{ij}|\underline{\theta}, z_{ij}(\underline{\psi}))}{\partial z_{ij}} + \frac{1}{\zeta_{ij}} \left(\frac{\partial \zeta_{ij}}{\partial z_{ij}} \right) \right] \frac{\partial z_{ij}}{\partial \underline{\psi}} & = \underline{0}. \end{cases} \tag{B5}$$

Notice that the estimation of the parameters carried out by maximizing the conditional likelihood, and letting only the positive weights to be accounted for, is perfectly consistent with the theory of hurdle models [152, 70]: although alternative estimation procedures can be devised (see, for example, [63]), in the present paper, we will stick to the proper, econometric one - which has been already employed in our companion paper [1], to estimate the parameters of conditional, discrete-valued models.

B.1.1 Conditional exponential model

The conditional exponential model is defined by the expression

$$H_{ij}(w_{ij}) = (\beta_0 + \beta_{ij}) w_{ij} \tag{B6}$$

that induces the following node pair-specific partition function

$$\zeta_{ij} = \int_0^{+\infty} e^{-(\beta_0+\beta_{ij})w_{ij}} dw_{ij} = \frac{1}{\beta_0+\beta_{ij}}. \tag{B7}$$

After the econometric reparametrization, according to which $\beta_{ij} \equiv z_{ij}^{-1}$, the log-likelihood function of the conditional exponential model reads

$$\mathcal{L} = \sum_{\substack{i<j \\ (a_{ij}=1)}} [-(\beta_0 + z_{ij}^{-1})w_{ij} - \ln(\zeta_{ij})]; \tag{B8}$$

hence, its maximization leads to the system of equations

$$\begin{cases} \sum_{i<j|a_{ij}=1}[\langle w_{ij}|a_{ij}=1\rangle - w_{ij}] & = 0 \\ \sum_{i<j|a_{ij}=1}[\langle w_{ij}|a_{ij}=1\rangle - w_{ij}]\partial_{\underline{\alpha}}(z_{ij}^{-1}) & = 0 \end{cases} \tag{B9}$$

where $\langle w_{ij}|a_{ij}=1\rangle = \frac{z_{ij}}{1+\beta_0 z_{ij}}$. Notice that we have a condition on the parameters, reading $\beta_0 + \beta_{ij} > 0$.

B.1.2 Conditional gamma model

The conditional gamma model is defined by the expression

$$H_{ij}(w_{ij}) = (\beta_0+\beta_{ij})w_{ij} + \xi_0 \ln(w_{ij}) \tag{B10}$$

and induces the following node pair-specific partition function

$$\zeta_{ij} = \int_0^{\infty} e^{-(\beta_0+\beta_{ij})w_{ij}} w_{ij}^{-\xi_0} dw_{ij} = \frac{\Gamma(1-\xi_0)}{(\beta_0+\beta_{ij})^{1-\xi_0}}. \tag{B11}$$

After the econometric reparametrization, according to which $\beta_{ij} \equiv z_{ij}^{-1}$, the log-likelihood function of the conditional gamma model reads

$$\mathcal{L} = \sum_{\substack{i<j \\ (a_{ij}=1)}} [-(\beta_0 + z_{ij}^{-1})w_{ij} - \xi_0 \ln(w_{ij}) - \ln(\zeta_{ij})]; \tag{B12}$$

hence, its maximization leads to the system of equations

$$\begin{cases} \sum_{i<j|a_{ij}=1}[\langle w_{ij}|a_{ij}=1\rangle - w_{ij}] & = 0 \\ \sum_{i<j|a_{ij}=1}[\langle \ln(w_{ij})|a_{ij}=1\rangle - \ln(w_{ij})] & = 0 \\ \sum_{i<j|a_{ij}=1}[\langle w_{ij}|a_{ij}=1\rangle - w_{ij}]\partial_{\underline{\alpha}}(z_{ij}^{-1}) & = 0 \end{cases} \tag{B13}$$

where $\langle w_{ij}|a_{ij}=1\rangle = \frac{z_{ij}(1-\xi_0)}{1+\beta_0 z_{ij}}$ and $\langle \ln(w_{ij})|a_{ij}=1\rangle = \psi(1-\xi_0) - \ln\big(\beta_0 + z_{ij}^{-1}\big)$. Notice that we have conditions on the parameters, reading $\beta_0 + \beta_{ij} > 0$ and $\xi_0 < 1$.

B.1.3 Conditional Pareto model

The conditional Pareto model is defined by the expression

$$H_{ij}(w_{ij}) = \xi_{ij} \ln(w_{ij}) \tag{B14}$$

that induces the following node pair-specific partition function

$$\zeta_{ij} = \int_{m_{ij}}^{+\infty} e^{-\xi_{ij}\ln(w_{ij})} dw_{ij} = \int_{m_{ij}}^{+\infty} w_{ij}^{-\xi_{ij}} dw_{ij} = \frac{m_{ij}^{1-\xi_{ij}}}{\xi_{ij}-1}. \tag{B15}$$

After the econometric reparametrization, according to which $\xi_{ij} - 2 \equiv z_{ij}^{-1}$ and $m_{ij} \equiv w_{min}$, the log-likelihood function of the conditional Pareto model reads

$$\mathcal{L} = \sum_{\substack{i<j \\ (a_{ij}=1)}} [-(2 + z_{ij}^{-1}) \ln(w_{ij}) - \ln(\zeta_{ij})]; \tag{B16}$$

hence, its maximization leads to the system of equations

$$\sum_{i<j|a_{ij}=1} [\langle \ln(w_{ij})|a_{ij}=1\rangle - \ln(w_{ij})]\partial_{\underline{\alpha}}(z_{ij}^{-1}) = 0 \tag{B17}$$

where $\langle \ln(w_{ij})|a_{ij}=1\rangle = \ln(w_{min}) + \frac{z_{ij}}{1+z_{ij}}$. Notice that we have conditions on the parameters, reading $w_{min} > 0$ and $\xi_{ij} > 2$.

B.1.4 Conditional log-normal model

The conditional log-normal model is defined by the expression

$$H_{ij}(w_{ij}) = \xi_{ij}\ln(w_{ij}) + \gamma_0 \ln^2(w_{ij}) \tag{B18}$$

that induces the following node pair-specific partition function

$$\begin{aligned}\zeta_{ij} &= \int_0^{+\infty} e^{-\xi_{ij}\ln(w_{ij})-\gamma_0\ln^2(w_{ij})}dw_{ij} \\ &= \int_{-\infty}^{+\infty} e^{(1-\xi_{ij})t_{ij}}e^{-\gamma_0 t_{ij}^2}dt_{ij} = \sqrt{\frac{\pi}{\gamma_0}}e^{\frac{(\xi_{ij}-1)^2}{4\gamma_0}}\end{aligned} \tag{B19}$$

a result that is readily obtained by putting $t_{ij} = \ln(w_{ij})$ and exploiting the relationship $\int_{-\infty}^{+\infty} e^{-ax^2+bx+c}dx = \sqrt{\frac{\pi}{a}}e^{\frac{b^2}{4a}+c}$.

After the econometric reparametrization, according to which $1-\xi_{ij} \equiv \ln(z_{ij})$, the log-likelihood function of the conditional log-normal model reads

$$\mathcal{L} = \sum_{\substack{i<j \\ (a_{ij}=1)}} [(\ln(z_{ij})+1)\ln(w_{ij}) - \gamma_0 \ln^2(w_{ij}) - \ln(\zeta_{ij})]; \tag{B20}$$

hence, its maximization leads to the system of equations

$$\begin{cases} \sum_{i<j|a_{ij}=1}[\langle \ln^2(w_{ij})|a_{ij}=1\rangle - \ln^2(w_{ij})] & = 0 \\ \sum_{i<j|a_{ij}=1}[\langle \ln(w_{ij})|a_{ij}=1\rangle - \ln(w_{ij})]\partial_{\underline{\alpha}}\ln(z_{ij}) & = 0 \end{cases} \tag{B21}$$

where $\langle \ln^2(w_{ij})|a_{ij}=1\rangle = \frac{2\gamma_0+\ln^2(z_{ij})}{4\gamma_0^2}$ and $\langle \ln(w_{ij})|a_{ij}=1\rangle = \frac{\ln(z_{ij})}{2\gamma_0}$. Notice that we have a condition on the parameters, reading $\gamma_0 > 0$.

B.2 Integrated models

Any member of the class of integrated models is described by the expression

$$Q(\mathbf{W}) = \frac{e^{-H(\mathbf{W})}}{\int_{\mathbb{W}} e^{-H(\mathbf{W})}d\mathbf{W}} \tag{B22}$$

the single instances being characterized by different expressions of the Hamiltonian. Since, however, each Hamiltonian considered here is a sum over the node pairs, the result

$$Q(\mathbf{W}) = \frac{e^{-\sum_{i<j} H_{ij}(w_{ij})}}{\int_{\mathbb{W}} e^{-\sum_{i,j} H_{ij}(w_{ij})} d\mathbf{W}}$$
$$= \prod_{i<j} \frac{e^{-H_{ij}(w_{ij})}}{\sum_{a_{ij}=0,1} \int_{\Theta[w_{ij}]=a_{ij}} e^{-H_{ij}(w_{ij})} dw_{ij}} = \prod_{i<j} \frac{e^{-H_{ij}(w_{ij})}}{Z_{ij}} \quad \text{(B23)}$$

holds irrespectively from the specific functional form of $H_{ij}(w_{ij})$. The identifications

$$H_{ij}(w_{ij}) \equiv f(w_{ij}|\underline{\theta}, z_{ij}(\underline{\psi})) \quad \text{(B24)}$$

and

$$\mathcal{L} \equiv \ln Q(\mathbf{W}) \quad \text{(B25)}$$

lead to estimate parameters by solving the (coupled) systems of purely structural equations $\frac{\partial \mathcal{L}}{\partial \underline{\theta}} = \underline{0}$ and econometric-like equations $\frac{\partial \mathcal{L}}{\partial \underline{\psi}} = \underline{0}$, i.e.

$$\begin{cases} \sum_{i<j} \left[\frac{\partial f(w_{ij}|\underline{\theta}, z_{ij}(\underline{\psi}))}{\partial \underline{\theta}} + \frac{1}{Z_{ij}} \left(\frac{\partial Z_{ij}}{\partial \underline{\theta}} \right) \right] = \underline{0} \\ \sum_{i<j} \left[\frac{\partial f(w_{ij}|\underline{\theta}, z_{ij}(\underline{\psi}))}{\partial z_{ij}} + \frac{1}{Z_{ij}} \left(\frac{\partial Z_{ij}}{\partial z_{ij}} \right) \right] \frac{\partial z_{ij}}{\partial \underline{\psi}} = \underline{0}. \end{cases} \quad \text{(B26)}$$

The integrated exponential model we have considered in the present paper is defined by the expression

$$H_{ij}(w_{ij}) = (\alpha_i + \alpha_j) a_{ij} + (\beta_0 + \beta_{ij}) w_{ij} \quad \text{(B27)}$$

that induces the following node pair-specific partition function

$$Z_{ij} = \sum_{a_{ij}=0}^{1} \int_{\Theta[w_{ij}]=a_{ij}} e^{-(\alpha_i+\alpha_j)a_{ij} - (\beta_0+\beta_{ij})w_{ij}} dw_{ij}$$
$$= 1 + e^{-(\alpha_i+\alpha_j)} \int_0^{+\infty} e^{-(\beta_0+\beta_{ij})w_{ij}} dw_{ij} = 1 + \frac{e^{-(\alpha_i+\alpha_j)}}{\beta_0 + \beta_{ij}}. \quad \text{(B28)}$$

After the econometric re-parametrization, according to which $\beta_{ij} \equiv z_{ij}^{-1}$, the log-likelihood function of the exponential model reads

$$\mathcal{L} = \sum_{i<j} [-(\alpha_i + \alpha_j) a_{ij} - (\beta_0 + z_{ij}^{-1}) w_{ij} - \ln(Z_{ij})]; \quad \text{(B29)}$$

hence, its maximization leads to the system of equations

$$\begin{cases} \langle k_i \rangle - k_i & = 0, \, \forall \, i \\ \langle W \rangle - W & = 0 \\ \sum_{i<j} [\langle w_{ij} \rangle - w_{ij}] \partial_{\underline{\alpha}} (z_{ij}^{-1}) & = 0 \end{cases} \tag{B30}$$

where $k_i = \sum_{j(\neq i)} a_{ij}$ is the empirical degree of node i, w_{ij} is the empirical, pair-specific weight and $W = \sum_{i<j} w_{ij}$ is the empirical, total weight; $\langle k_i \rangle = \sum_{j(\neq i)} p_{ij}$, $\langle w_{ij} \rangle$ and $\langle W \rangle = \sum_{i<j} \langle w_{ij} \rangle$ are their expected counterparts.

Notice that we have a condition on the parameters, reading $\beta_0 + \beta_{ij} > 0$.

B.3 Turning structural models into econometric ones

So far, we have derived two classes of models, by explicitly solving the constrained maximization of a number of functionals derived from the KL divergence. As the functional form of the probability distributions belonging to the two classes (solely) depends on the enforced constraints, such models 'are born' as purely structural ones.

In order to turn them into candidate models to be employed for econometric purposes, we need to properly transform (some of) the Lagrange multipliers into functions of the econometric quantities of relevance for the problem at hand. In this respect, the theory of GLMs provides helpful suggestions about how to proceed; besides, one can figure out some (sets of) basic requirements such a transformation should satisfy:

- the transformation should turn the expected values $\langle w_{ij} \rangle$ and $\langle w_{ij} | a_{ij} = 1 \rangle$ into positive, monotonically increasing functions of z_{ij};
- the transformation should not violate the mathematical requirements to have well-defined (first and second) distribution moments.

In what follows, we will focus on the conditional models.

B.3.1 Conditional exponential model

It is characterized by the expression

$$\langle w_{ij}|a_{ij}=1\rangle = \frac{1}{\beta_0+\beta_{ij}} \tag{B31}$$

that can be turned into an econometric one by posing

$$\langle w_{ij}|a_{ij}=1\rangle = \frac{1}{\beta_0+\beta_{ij}} \equiv g(z_{ij}) \tag{B32}$$

according to the prescription informing the so-called generalized linear models (GLMs). Before specifying the functional form of g, let us consider that the dyadic parameter β_{ij} must be decreasing in z_{ij} - a requirement that can be justified upon identifying β_{ij} as the 'shadow price' that countries i and j have to pay to trade a unity of goods [153]; analogously, β_0 can be interpreted as modelling a global tax that everyone has to pay to exchange goods - independently of its trade 'capacity'. These considerations lead us to impose $\beta_{ij} \equiv z_{ij}^{-1}$, a choice inducing the expression

$$\langle w_{ij}|a_{ij}=1\rangle = \frac{1}{\beta_0+z_{ij}^{-1}} = \frac{z_{ij}}{1+\beta_0 z_{ij}} \tag{B33}$$

which violates none of the requirements listed at the beginning of the section.

B.3.2 Conditional gamma model

It is characterized by the expressions

$$\langle w_{ij}|a_{ij}=1\rangle = \frac{1-\xi_0}{\beta_0+\beta_{ij}}, \tag{B34}$$

$$\langle \ln(w_{ij})|a_{ij}=1\rangle = \psi(1-\xi_0) - \ln(\beta_0+\beta_{ij}) \tag{B35}$$

(where $\psi(x) = \Gamma'(x)/\Gamma(x)$ is the digamma function) that can be turned into econometric ones by posing $\beta_{ij} \equiv z_{ij}^{-1}$, according to considerations which are analogous to those driving the econometric reparametrization of the conditional, exponential model. This choice induces the expressions

$$\langle w_{ij}|a_{ij}=1\rangle = \frac{1-\xi_0}{\beta_0 + z_{ij}^{-1}} = \frac{(1-\xi_0)z_{ij}}{1+\beta_0 z_{ij}}, \tag{B36}$$

$$\langle \ln(w_{ij})|a_{ij}=1\rangle = \psi(1-\xi_0) - \ln\big(\beta_0 + z_{ij}^{-1}\big); \tag{B37}$$

notice that the conditional, exponential model is recovered in case $\xi_0 = 0$ (i.e. when the constraint on the sum of the logarithms of weights is switched-off).

B.3.3 Conditional Pareto model

It is characterized by the expression

$$\langle w_{ij}|a_{ij}=1\rangle = \left(\frac{\xi_{ij}-1}{\xi_{ij}-2}\right) m_{ij} \tag{B38}$$

that can be turned into an econometric one by posing

$$\langle w_{ij}|a_{ij}=1\rangle = \left(\frac{\xi_{ij}-1}{\xi_{ij}-2}\right) m_{ij} \equiv g(z_{ij}) \tag{B39}$$

according to the prescription informing the GLMs. Upon considering that 1) the (conditional) expected value is well defined only if $\xi_{ij} - 2 > 0$ and that 2) a linear relationship between the former and z_{ij} would be desirable, a suitable reparametrization may read $\xi_{ij} - 2 \equiv z_{ij}^{-1}$ and $m_{ij} \equiv w_{min}$, in turn leading to

$$\langle w_{ij}|a_{ij}=1\rangle = (1+z_{ij})w_{min} \tag{B40}$$

which violates none of the requirements listed at the beginning of the section.

B.3.4 Conditional log-normal model

It is characterized by the expressions

$$\langle \ln(w_{ij})|a_{ij}=1\rangle = \frac{1-\xi_{ij}}{2\gamma_0}, \tag{B41}$$

$$\langle \ln^2(w_{ij})|a_{ij}=1\rangle = \frac{2\gamma_0 + (1-\xi_{ij})^2}{4\gamma_0^2}. \tag{B42}$$

Upon considering that the logarithm of weights can admit negative values, i.e. when $w_{ij} \in (0,1)$, and that there are no theoretical restrictions on the sign of $\langle \ln(w_{ij}) | a_{ij} = 1 \rangle$, a suitable reparametrization may read $1 - \xi_{ij} \equiv \ln(z_{ij})$, in turn leading to

$$\langle \ln(w_{ij}) | a_{ij} = 1 \rangle = \frac{\ln(z_{ij})}{2\gamma_0}, \tag{B43}$$

$$\langle \ln^2(w_{ij}) | a_{ij} = 1 \rangle = \frac{2\gamma_0 + \ln^2(z_{ij})}{4\gamma_0^2} \tag{B44}$$

which violate none of the requirements listed at the beginning of the section.

B.4 The Shannon-Fisher plane

Here, starting from the conditional probability density function $q_{ij}(w|a_{ij} = 1)$, for each connected dyad we compute explicitly the continuous Shannon entropy S_{ij} and the Fisher Information Measure (FIM) F_{ij} needed to construct the Shannon-Fisher plane introduced in section 3.3.4. We do so for each model separately.

B.4.1 Conditional exponential model

The conditional exponential model is defined by the probability distribution

$$q_{ij}(w_{ij}|a_{ij} = 1) = (\beta_0 + \beta_{ij}) e^{-(\beta_0 + \beta_{ij}) w_{ij}} \tag{B45}$$

inducing a Shannon entropy reading

$$S_{ij} = \langle H_{ij} \rangle + \ln \zeta_{ij} = 1 - \ln[\beta_0 + \beta_{ij}] \tag{B46}$$

and a FIM reading

$$F_{ij} = \langle (H'_{ij})^2 \rangle = (\beta_0 + \beta_{ij})^2; \tag{B47}$$

as the value of the parameter $\beta_0 + \beta_{ij}$ increases, Shannon entropy decreases while Fisher Information Measure increases as well.

B.4.2 Conditional gamma model

The conditional gamma model is defined by the probability distribution

$$q_{ij}(w_{ij}|a_{ij}=1)=\frac{(\beta_0+\beta_{ij})^{1-\xi_0}}{\Gamma(1-\xi_0)}w_{ij}^{-\xi_0}e^{-(\beta_0+\beta_{ij})w_{ij}} \tag{B48}$$

inducing a Shannon entropy reading

$$S_{ij}=\langle H_{ij}\rangle+\ln\zeta_{ij}=-\ln[\beta_0+\beta_{ij}]+\xi_0\psi(1-\xi_0)+\ln\Gamma(1-\xi_0)+(1-\xi_0) \tag{B49}$$

and a FIM reading

$$\begin{aligned}F_{ij}&=\langle(H'_{ij})^2\rangle=(\beta_0+\beta_{ij})^2+2\xi_0(\beta_0+\beta_{ij})\langle w_{ij}^{-1}|a_{ij}=1\rangle+\xi_0^2\langle w_{ij}^{-2}|a_{ij}=1\rangle\\&=(\beta_0+\beta_{ij})^2\left[1+2\xi_0\frac{\Gamma(-\xi_0)}{\Gamma(1-\xi_0)}+\xi_0^2\frac{\Gamma(-1-\xi_0)}{\Gamma(1-\xi_0)}\right];\end{aligned} \tag{B50}$$

the expression above does not diverge for the values of the parameter ξ_0 ensuring that the (first) two, negative moments, $\langle w_{ij}^{-1}|a_{ij}=1\rangle$ and $\langle w_{ij}^{-2}|a_{ij}=1\rangle$, of the (conditional) gamma distribution do not diverge as well, i.e. $\xi_0<-1$.

B.4.3 Conditional Pareto model

The conditional Pareto model is defined by the probability distribution

$$q_{ij}(w_{ij}|a_{ij}=1)=\frac{\xi_{ij}-1}{m_{ij}^{1-\xi_{ij}}}w_{ij}^{-\xi_{ij}} \tag{B51}$$

inducing a Shannon entropy reading

$$S_{ij}=\langle H_{ij}\rangle+\ln\zeta_{ij}=\left(\frac{\xi_{ij}}{\xi_{ij}-1}\right)-\ln[\xi_{ij}-1]+\ln m_{ij} \tag{B52}$$

and a FIM reading

$$F_{ij}=\langle(H'_{ij})^2\rangle=\frac{\xi_{ij}^2}{m^2}\left(\frac{\xi_{ij}-1}{\xi_{ij}+1}\right); \tag{B53}$$

the expression above holds true for the values of the parameter ξ_{ij} ensuring that the Pareto distribution exists, i.e. $\xi_{ij}>1$. Besides, the convergence of the second, negative moment of the (conditional) Pareto distribution ensures that its FIM does not diverge as well.

B.4.4 Conditional log-normal model

The conditional log-normal model is defined by the probability distribution

$$q_{ij}(w_{ij}|a_{ij}=1)=\frac{e^{-\xi_{ij}\ln(w_{ij})-\gamma_0\ln^2(w_{ij})}}{\sqrt{\frac{\pi}{\gamma_0}}e^{\frac{(\xi_{ij}-1)^2}{4\gamma_0}}} \tag{B54}$$

inducing a Shannon entropy reading

$$S_{ij}=\langle H_{ij}\rangle+\ln\zeta_{ij}=\frac{1-\xi_{ij}}{2\gamma_0}+\frac{1}{2}\left[1+\frac{1}{2}\ln\left(\frac{\pi}{\gamma_0}\right)\right] \tag{B55}$$

and a FIM reading

$$F_{ij}=\langle(H'_{ij})^2\rangle=e^{\xi_{ij}/\gamma_0}(1+2\gamma_0+\xi_{ij}+\xi_{ij}^2); \tag{B56}$$

the expression above holds true for the values of the parameter γ_0 ensuring that the log-normal distribution exists, i.e. $\gamma_0>0$.

Appendix C

Three, different parameter estimation procedures

Appendix C, based on [3], expands on the different procedures to estimate parameters presented in chapter 4. The first section provides a brief overview of the topic for conditional models. The second section focuses on the implementation of the three parameter estimation methods named 'deterministic', 'annealed' and 'quenched' for three types of models, i.e. scalar (or homogeneous), vector (or weakly heterogeneous), tensor (or strongly heterogeneous) ones.

C.1 The functional form of conditional models

The constrained maximization of $S(\overline{Q}|P)$ proceeds by specifying the set of weighted constraints reading

$$1 = \int_{\mathbb{W}_{\mathbf{A}}} P(\mathbf{W}|\mathbf{A}) d\mathbf{W}, \; \forall \, \mathbf{A} \in \mathbb{A}, \tag{C1}$$

$$\langle C_\alpha \rangle = \sum_{\mathbf{A} \in \mathbb{A}} P(\mathbf{A}) \int_{\mathbb{W}_{\mathbf{A}}} Q(\mathbf{W}|\mathbf{A}) C_\alpha(\mathbf{W}) d\mathbf{W}, \; \forall \, \alpha \tag{C2}$$

the first condition ensuring the normalisation of the probability distribution and the vector $\{C_\alpha(\mathbf{W})\}$ representing the proper set of weighted constraints. The distribution induced by them reads

$$
\begin{aligned}
Q(\mathbf{W}|\mathbf{A}) &= \frac{e^{-H(\mathbf{W})}}{Z_{\mathbf{A}}} \\
&= \frac{e^{-H(\mathbf{W})}}{\int_{\mathbb{W}_{\mathbf{A}}} e^{-H(\mathbf{W})} d\mathbf{W}} = \frac{e^{-\sum_{i<j} H_{ij}(w_{ij})}}{\int_{\mathbb{W}_{\mathbf{A}}} e^{-\sum_{i<j} H_{ij}(w_{ij})} d\mathbf{W}} \\
&= \prod_{i<j} \frac{e^{-H_{ij}(w_{ij})}}{\left[\int_{m_{ij}}^{+\infty} e^{-H_{ij}(w_{ij})} dw_{ij}\right]^{a_{ij}}} = \prod_{i<j} \frac{e^{-H_{ij}(w_{ij})}}{\zeta_{ij}^{a_{ij}}}
\end{aligned} \tag{C3}
$$

if $\mathbf{W} \in \mathbb{W}_{\mathbf{A}}$ and 0 otherwise - since each Hamiltonian considered in the present paper is separable, i.e. a sum of node pairs-specific Hamiltonians: in formulas, $H(\mathbf{W}) = \sum_{i<j} H_{ij}(w_{ij})$.

C.2 Estimating the parameters

Let us, now, provide general expressions for the 'deterministic' and the 'annealed' recipe for parameter estimation. The first one follows from writing

$$
\begin{aligned}
\mathcal{L}_{\underline{\psi}} &= \ln Q(\mathbf{W}^*|\mathbf{A}^*) \\
&= -H(\mathbf{W}^*) - \ln\left[\int_{\mathbb{W}_{\mathbf{A}^*}} e^{-H(\mathbf{W})} d\mathbf{W}\right] \\
&= \sum_{i<j} H_{ij}(w_{ij}^*) - \ln \prod_{i<j} \zeta_{ij}^{a_{ij}} = \sum_{i<j} [H_{ij}(w_{ij}^*) - a_{ij}^* \ln \zeta_{ij}]
\end{aligned} \tag{C4}
$$

while the second one follows from writing

$$
\mathcal{G}_{\underline{\psi}} = \sum_{\mathbf{A} \in \mathbb{A}} P(\mathbf{A}) \ln Q(\mathbf{W}^*|\mathbf{A}) = \langle \mathcal{L}_{\underline{\psi}} \rangle = \sum_{i<j} [H_{ij}(w_{ij}^*) - p_{ij} \ln \zeta_{ij}]. \tag{C5}
$$

C.2.1 'Scalar' or homogeneous variant of the CEM

In the particular case of the UBRGM-induced, homogeneous variant of the CEM, one can derive the 'quenched' distribution of the parameter β upon considering that it is a function of the discrete, random variable L. Since $L \sim \text{Bin}(N(N-1)/2, p)$, with $p = 2L^*/N(N-1)$, one finds that

$$
\beta \sim \binom{\frac{N(N-1)}{2}}{W^*\beta} p^{W^*\beta} (1-p)^{\frac{N(N-1)}{2} - W^*\beta} \tag{C6}
$$

an expression allowing us to derive the expected value of β, i.e.

$$\begin{aligned}\langle\beta\rangle &= \sum_{\beta=0}^{\frac{N(N-1)}{2W^*}} \beta \binom{\frac{N(N-1)}{2}}{W^*\beta} p^{W^*\beta}(1-p)^{\frac{N(N-1)}{2}-W^*\beta} \\ &= \frac{N(N-1)}{2W^*}p = \frac{\langle L\rangle}{W^*} = \frac{L^*}{W^*}\end{aligned} \tag{C7}$$

as well as its variance. Since

$$\begin{aligned}\langle\beta^2\rangle &= \sum_{\beta=0}^{\frac{N(N-1)}{2W^*}} \beta^2 \binom{\frac{N(N-1)}{2}}{W^*\beta} p^{W^*\beta}(1-p)^{\frac{N(N-1)}{2}-W^*\beta} \\ &= \frac{N(N-1)}{2(W^*)^2}p + \frac{N(N-1)}{2(W^*)^2}\left[\frac{N(N-1)}{2(W^*)^2}-1\right]p^2\end{aligned} \tag{C8}$$

we have that

$$\text{Var}[\beta] = \langle\beta^2\rangle - \langle\beta\rangle^2 = \frac{N(N-1)}{2(W^*)^2}p(1-p) = \frac{\text{Var}[L]}{(W^*)^2} = \frac{L^*}{(W^*)^2}\left[\frac{N(N-1)-2L^*}{N(N-1)}\right] \tag{C9}$$

with $\text{Var}[L] = N(N-1)/2 \cdot p(1-p)$. Since the distribution obeyed by L converges to the normal distribution $\mathcal{N}(L^*, \text{Var}[L])$, the distribution obeyed by β converges to the distribution

$$\begin{aligned}g(\beta) &= \frac{W^*}{\sqrt{2\pi\text{Var}[L]}} e^{-\frac{(W^*\beta-L^*)^2}{2\text{Var}[L]}} \\ &= \frac{1}{\sqrt{2\pi\text{Var}[L]/(W^*)^2}} e^{-\frac{(\beta-L^*/W^*)^2}{2\text{Var}[L]/(W^*)^2}} \frac{1}{\sqrt{2\pi\text{Var}[\beta]}} e^{-\frac{(\beta-\beta^*)^2}{2\text{Var}[\beta]}} = \mathcal{N}(\beta^*, \text{Var}[\beta])\end{aligned} \tag{C10}$$

with $\beta^* = L^*/W^*$ and $\text{Var}[\beta] = \text{Var}[L]/(W^*)^2$.

In the case of the UBCM-induced, homogeneous version of the CEM, L obeys the Poisson-Binomial (PB) distribution reading $\text{PB}(N(N-1)/2, \{p^{\text{UBCM}}\}_{i,j=1}^{N})$ whose normal approximation reads $\mathcal{N}(L^*, \text{Var}[L])$, with $\text{Var}[L] = \sum_{i<j} p_{ij}^{\text{UBCM}}(1-p_{ij}^{\text{UBCM}})$; as a consequence, the distribution obeyed by β converges to $\mathcal{N}(\beta^*, \text{Var}[\beta])$, with $\beta^* = L^*/W^*$ and $\text{Var}[\beta] = \text{Var}[L]/(W^*)^2$.

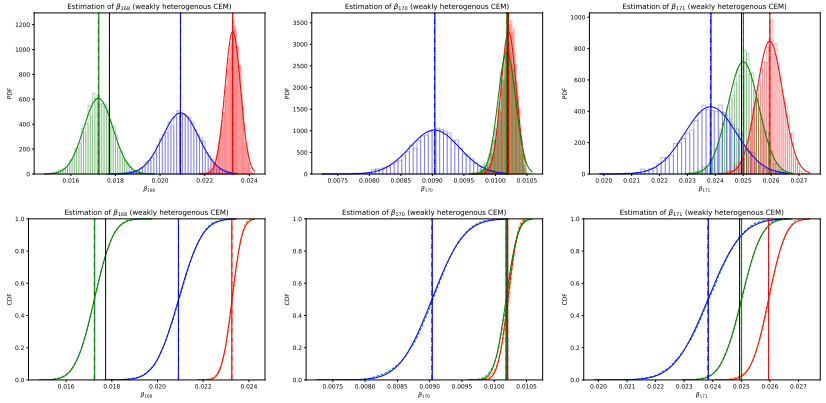

Figure 22: Estimations of the parameters β_{168}, β_{170} and β_{171} (top panels, from left to right), entering the definition of the weakly heterogeneous version of the CEM, where the binary topology is either 'deterministic' (black) or generated via the UBRGM (blue), the UBCM (green) and the LM (red). The deterministic approach leads to a single estimate, while the other approaches lead to either a single, 'annealed' estimate (vertical, solid lines) or to a whole distribution of 'quenched' estimates (histograms with normal density curves having the same average and standard deviation, constructed over an ensemble of 5.000 binary configurations; the average value is indicated by a vertical, dash-dotted line). Each 'annealed' estimate overlaps with the average value of the related 'quenched' distribution, although 1) the latter ones are well separated in the case of node 168, 2) only partly overlapped in the case of node 171, 3) the UBCM-induced and the LM-induced ones overlap while the UBRGM-induced one remains well separated in the case of node 170. Moreover, the 'deterministic' estimates are always very close to (if not overlapping with) the UBCM-induced, 'annealed' ones. Although the empirical and theoretical CDFs (respectively depicted as solid lines and dotted lines in the bottom panels) seem to be in a very good agreement, the Anderson-Darling test never rejects the normality hypothesis only for node 166 and does not reject the normality hypothesis in the case of the UBCM-induced distribution of estimates for node 168.

In the case of the LM-induced, homogeneous version of the CEM, L obeys the Poisson-Binomial distribution reading $\mathrm{PB}(N(N-1)/2, \{p^{\mathrm{LM}}\}_{i,j=1}^{N})$ whose normal approximation reads $\mathcal{N}(L^*, \mathrm{Var}[L])$, with $\mathrm{Var}[L] = \sum_{i<j} p_{ij}^{\mathrm{LM}}(1-p_{ij}^{\mathrm{LM}})$; as a consequence, the distribution obeyed by

β converges to $\mathcal{N}(\beta^*, \text{Var}[\beta])$, with $\beta^* = L^*/W^*$ and $\text{Var}[\beta] = \text{Var}[L]/(W^*)^2$.

C.2.2 'Vector' or weakly heterogeneous variant of the CEM

As pointed out in the main text, each 'annealed' estimation overlaps with the average value of the related 'quenched' distribution although 1) the latter ones are well separated, in the case of node 168, 2) only partly overlapped, in the case of node 171, 3) the UBCM-induced and the LM-induced ones overlap while the UBRGM-induced one remains well separated, in the case of node 170 (see Fig. 22). Moreover, the 'deterministic' estimation is always very close to the UBCM-induced, 'annealed' one - a result that may be a consequence of the accurate description of the empirical network topology provided by the UBCM - evidently, much more accurate than those provided by the UBRGM and the LM.

Each solid line in Fig. 22 represents a normal distribution whose average value and variance coincide with the ones of the corresponding sample distribution: although the empirical and theoretical CDFs seem to be in (a very good) agreement, the Anderson-Darling test never rejects the normality hypothesis only for node 166 and does not reject the normality hypothesis in the case of the UBCM-induced distribution of values for node 168.

C.2.3 'Tensor' variant of the CEM

Let us, now, leave β_{ij} in its tensor form and constrain the set of weight-specific estimates $\hat{w}_{ij}, \forall\, i < j$. In this case, the three recipes lead to the following estimates

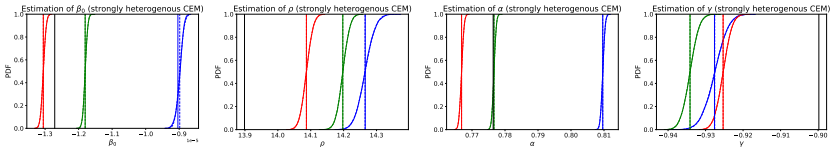

Figure 23: Empirical CDFs for the parameters (from left to right) β_0, ρ, α and γ entering the definition of the econometric version of the CEM, where the binary topology is either 'deterministic' (black) or generated via the UBRGM (blue), the UBCM (green) and the LM (red). The deterministic approach leads to a single estimate, while the other approaches lead to either a single, 'annealed' estimate (vertical, solid lines) or to a whole distribution of 'quenched' estimates (constructed over an ensemble of 5.000 binary configurations; the corresponding average value is indicated by a vertical, dash-dotted line). The shapes of the 'quenched', cumulative distributions induced by the three, binary recipes are very similar.

$$\mathcal{L}_{\underline{\psi}} = \sum_{i<j}[-\beta_{ij}\hat{w}_{ij} + a^*_{ij}\ln\beta_{ij}] \quad \Longrightarrow \quad \beta_{ij} = \frac{a^*_{ij}}{\hat{w}_{ij}} \tag{C11}$$

$$\mathcal{G}_{\underline{\psi}} = \sum_{i<j}[-\beta_{ij}\hat{w}_{ij} + p_{ij}\ln\beta_{ij}] \quad \Longrightarrow \quad \beta_{ij} = \frac{p_{ij}}{\hat{w}_{ij}} \tag{C12}$$

$$\langle\beta_{ij}\rangle = \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A})\beta_{ij}(\mathbf{A}) = \sum_{\mathbf{A}\in\mathbb{A}} P(\mathbf{A})\frac{a_{ij}}{\hat{w}_{ij}} \quad \Longrightarrow \quad \langle\beta_{ij}\rangle = \frac{p_{ij}}{\hat{w}_{ij}} \tag{C13}$$

a result signalling large differences between the 'deterministic' recipe, on the one hand, and the 'quenched' and 'annealed' recipes, on the other - that, instead, coincide. If, however, $\hat{w}_{ij} \equiv w^*_{ij}$, $\forall\, i<j$ then, for consistency, $p_{ij} \equiv a^*_{ij}$ and the three recipes coincide.

C.2.4 'Econometric' variant of the CEM

As Figs. 15 and 23 show, the 'deterministic' estimation is always quite different from the other, two ones - the only exception being represented by the parameter α, under the UBCM-induced, binary recipe. Such a result should warn from employing the 'deterministic' estimation recipe *tout court* as ignoring the variety of structures that are compatible with a given probability distribution $P(\mathbf{A})$ will, in general, affect the estimation

of the parameters of interest.

Appendix D

Maximum-entropy models for motifs detection

Appendix D, based on [4], expands on the structural models presented in chapter 5. The first section focuses on binary models, i.e. the Directed Binary Configuration Model (DBCM), constraining the out-degree and in-degree sequences, and the Reciprocal Binary Configuration Model (RBCM), constraining the non-reciprocated and reciprocated degrees. The second section focuses on weighted, conditional models, i.e. the CReM$_A$, constraining the out-strength and in-strength sequences, and the CRWCM, constraining the out-strength and in-strength sequences according to the type of link (i.e. reciprocated or not) supporting the corresponding weight. Both types of models are numerically determined via the 'annealed' approach discussed in chapter 4).

D.1 Binary null models

D.1.1 The Directed Binary Configuration Model (DBCM)

The DBCM is the maximum-entropy model constraining the out-degree and in-degree sequences. The corresponding Hamiltonian is

$$H(\mathbf{A}) = \sum_i [\alpha_i^{out} k_i^{out} + \alpha_i^{in} k_i^{in}] = \sum_{i \neq j} (\alpha_i^{out} + \alpha_j^{in}) a_{ij} \qquad \text{(D1)}$$

and the partition function reads

$$\begin{aligned} Z &= \sum_{\mathbf{A}} e^{-H(\mathbf{A})} \\ &= \sum_{\mathbf{A}} e^{-\sum_{i\neq i}(\alpha_i^{out}+\alpha_j^{in})a_{ij}} = \prod_{i\neq j}\sum_{a_{ij}=0,1}(x_i^{out}x_j^{in})^{a_{ij}} = \prod_{i\neq i}(1+x_i^{out}x_j^{in}) \end{aligned} \tag{D2}$$

where $x_i^{(\cdot)} = e^{-\alpha_i^{(\cdot)}}$ with $(\cdot) = \{out, in\}$. After computing the partition function, one gets

$$P(\mathbf{A}) = \frac{e^{-H(\mathbf{A})}}{Z} = \prod_{i\neq j}\frac{\left(x_i^{out}x_j^{in}\right)^{a_{ij}}}{1+x_i^{out}x_j^{in}} \tag{D3}$$

allowing us to define a the log-likelihood function as

$$\begin{aligned} \mathcal{L} &= \ln P(\mathbf{A}) \\ &= -H(\mathbf{A}) - \ln Z = \sum_i [k_i^{out}\ln x_i^{out} + k_i^{in}\ln x_i^{in}] - \sum_{i\neq j}\ln\left(1+x_i^{out}x_j^{in}\right); \end{aligned} \tag{D4}$$

parameters are, then, estimated by solving the following set of equations:

$$\frac{\partial\mathcal{L}}{\partial\alpha_i^{out}} = -k_i^{out} + \sum_{j(\neq i)}\frac{x_i^{out}x_j^{in}}{1+x_i^{out}x_j^{in}} = 0,\ \forall\, i \tag{D5}$$

$$\frac{\partial\mathcal{L}}{\partial\alpha_i^{in}} = -k_i^{in} + \sum_{j(\neq i)}\frac{x_i^{in}x_j^{out}}{1+x_i^{in}x_j^{out}} = 0,\ \forall\, i \tag{D6}$$

D.1.2 The Reciprocal Binary Configuration Model (RBCM)

The RBCM is the maximum-entropy model constraining the reciprocated and non-reciprocated degree sequences. The corresponding Hamiltonian is

$$\begin{aligned} H(\mathbf{A}) &= \sum_i(\alpha_i^{\rightarrow}k_i^{\rightarrow} + \alpha_i^{\leftarrow}k_i^{\leftarrow} + \alpha_i^{\leftrightarrow}k_i^{\leftrightarrow}) = \\ &= \sum_{i<j}(\alpha_i^{\rightarrow}+\alpha_j^{\leftarrow})a_{ij}^{\rightarrow} + (\alpha_i^{\leftarrow}+\alpha_j^{\rightarrow})a_{ij}^{\leftarrow} + (\alpha_i^{\leftrightarrow}+\alpha_j^{\leftrightarrow})a_{ij}^{\leftrightarrow} \end{aligned} \tag{D7}$$

and the partition function reads

$$\begin{aligned} Z &= \sum_{\mathbf{A}} e^{-H(\mathbf{A})} \\ &= \sum_{\mathbf{A}} \prod_{i<j} (x_i^{\rightarrow} x_j^{\leftarrow})^{a_{ij}^{\rightarrow}} (x_i^{\leftarrow} x_j^{\rightarrow})^{a_{ij}^{\leftarrow}} (x_i^{\leftrightarrow} x_j^{\leftrightarrow})^{a_{ij}^{\leftrightarrow}} \\ &= \prod_{i<j} \sum_{\{a_{ij}\}} (x_i^{\rightarrow} x_j^{\leftarrow})^{a_{ij}^{\rightarrow}} (x_i^{\leftarrow} x_j^{\rightarrow})^{a_{ij}^{\leftarrow}} (x_i^{\leftrightarrow} x_j^{\leftrightarrow})^{a_{ij}^{\leftrightarrow}} \\ &= \prod_{i<j} (1 + x_i^{\rightarrow} x_j^{\leftarrow} + x_i^{\leftarrow} x_j^{\rightarrow} + x_i^{\leftrightarrow} x_j^{\leftrightarrow}) \end{aligned} \tag{D8}$$

where $x_i^y = e^{-\alpha_i^y}$ with $\cdot = \{\rightarrow, \leftarrow, \leftrightarrow\}$. After computing the partition function, one gets

$$P(\mathbf{A}) = \frac{e^{-H(\mathbf{A})}}{Z} = \prod_{i<j} \frac{(x_i^{\rightarrow} x_j^{\leftarrow})^{a_{ij}^{\rightarrow}} (x_i^{\leftarrow} x_j^{\rightarrow})^{a_{ij}^{\leftarrow}} (x_i^{\leftrightarrow} x_j^{\leftrightarrow})^{a_{ij}^{\leftrightarrow}}}{1 + x_i^{\rightarrow} x_j^{\leftarrow} + x_i^{\leftarrow} x_j^{\rightarrow} + x_i^{\leftrightarrow} x_j^{\leftrightarrow}} \tag{D9}$$

allowing us to define the log-likelihood function as

$$\begin{aligned} &\mathcal{L} = \ln P(\mathbf{A}) = \\ &= -H(\mathbf{A}) - \ln Z = \\ &= \sum_i [k_i^{\rightarrow} \ln x_i^{\rightarrow} + k_i^{\leftarrow} \ln x_i^{\leftarrow} + k_i^{\leftrightarrow} \ln x^{\leftrightarrow}] - \sum_{i<j} \ln\big(1 + x_i^{\rightarrow} x_j^{\leftarrow} + x_i^{\leftarrow} x_j^{\rightarrow} + x_i^{\leftrightarrow} x_j^{\leftrightarrow}\big); \end{aligned} \tag{D10}$$

parameters are, then, estimated by solving the following set of equations:

$$\frac{\partial \mathcal{L}}{\partial \alpha_i^{\rightarrow}} = -k_i^{\rightarrow} + \sum_{j(\neq i)} \frac{x_i^{\rightarrow} x_j^{\leftarrow}}{1 + x_i^{\rightarrow} x_j^{\leftarrow} + x_i^{\leftarrow} x_j^{\rightarrow} + x_i^{\leftrightarrow} x_j^{\leftrightarrow}}, \ \forall\, i \tag{D11}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_i^{\leftarrow}} = -k_i^{\leftarrow} + \sum_{j(\neq i)} \frac{x_i^{\leftarrow} x_j^{\rightarrow}}{1 + x_i^{\rightarrow} x_j^{\leftarrow} + x_i^{\leftarrow} x_j^{\rightarrow} + x_i^{\leftrightarrow} x_j^{\leftrightarrow}}, \ \forall\, i \tag{D12}$$

$$\frac{\partial \mathcal{L}}{\partial \alpha_i^{\leftrightarrow}} = -k_i^{\leftrightarrow} + \sum_{j(\neq i)} \frac{x_i^{\leftrightarrow} x_j^{\leftrightarrow}}{1 + x_i^{\rightarrow} x_j^{\leftarrow} + x_i^{\leftarrow} x_j^{\rightarrow} + x_i^{\leftrightarrow} x_j^{\leftrightarrow}}, \ \forall\, i \tag{D13}$$

D.2 Conditional weighted null models

D.2.1 Conditional Reconstruction Model A

The CReM_A is the conditional maximum-entropy model constraining the out- and in-strength sequences. The corresponding Hamiltonian is

$$H(\mathbf{W}) = \sum_i (\beta_i^{out} s_i^{out} + \beta_i^{in} s_i^{in}) = \sum_{i \neq j} (\beta_i^{out} + \beta_j^{in}) w_{ij} \tag{D14}$$

and the partition function reads

$$Z_{\mathbf{A}} = \int_{\mathbb{W}_{\mathbf{A}}} \prod_{i \neq j} e^{-\left(\beta_i^{out} + \beta_j^{in}\right) w_{ij}} dw_{ij} = \left(\frac{1}{\beta_i^{out} + \beta_j^{in}} \right)^{a_{ij}}. \tag{D15}$$

After computing the partition function, one gets

$$Q(\mathbf{W}|\mathbf{A}) = \prod_{i \neq j} \left[(\beta_i^{out} + \beta_j^{in}) e^{-(\beta_i^{out} + \beta_j^{in}) w_{ij}} \right]^{a_{ij}} \tag{D16}$$

allowing us to define the log-likelihood function as

$$\mathcal{L} = -\sum_i (\beta_i^{out} s_i^{out} + \beta_i^{in} s_i^{in}) + \sum_{i \neq j} a_{ij} \ln\left(\beta_i^{out} + \beta_j^{in}\right); \tag{D17}$$

in order to account for the variability of the binary adjacency matrix $\mathbf{A}$, we average the log-likelihood function over the ensemble and obtain its generalized version

$$\mathcal{G} = -\sum_i (\beta_i^{out} s_i^{out} + \beta_i^{in} s_i^{in}) + \sum_{i \neq j} p_{ij} \ln\left(\beta_i^{out} + \beta_j^{in}\right) \tag{D18}$$

whose maximization leads to

$$\frac{\partial \mathcal{G}}{\partial \beta_i^{out}} = -s_i^{out} + \sum_{j(\neq i)} \frac{p_{ij}}{\beta_i^{out} + \beta_j^{in}}, \; \forall \, i \tag{D19}$$

$$\frac{\partial \mathcal{G}}{\partial \beta_i^{in}} = -s_i^{in} + \sum_{j(\neq i)} \frac{p_{ji}}{\beta_i^{in} + \beta_j^{out}}, \; \forall \, i \tag{D20}$$

D.2.2 Conditionally Reciprocal Weighted Configuration Model

The CRWCM is a conditional maximum-entropy model constraining out- and in-strength sequences, by dividing them according to the reciprocal charcater of the underlying links. The Hamiltonian is

$$\begin{aligned} H(\mathbf{W}) &= \sum_i (\beta_i^{\rightarrow} s_i^{\rightarrow} + \beta_i^{\leftarrow} s_i^{\leftarrow}) + \left(\beta_i^{\leftrightarrow,out} s_i^{\leftrightarrow,out} + \beta_i^{\leftrightarrow,in} s_i^{\leftrightarrow,in}\right) \\ &\equiv \sum_{i \neq j} H_{ij}(w_{ij}) \end{aligned} \tag{D21}$$

where

$$H_{ij}(w_{ij}) = (\beta_i^{\rightarrow} + \beta_j^{\leftarrow}) a_{ij}^{\rightarrow} w_{ij} + (\beta_i^{\leftrightarrow,out} + \beta_j^{\leftrightarrow,in}) a_{ij}^{\leftrightarrow} w_{ij}. \tag{D22}$$

It induces the partition function

$$Z_{\mathbf{A}} = \int_{\mathbb{W}_{\mathbf{A}}} e^{-\sum_{i \neq j} H_{ij}(w_{ij})} dw_{ij} = \prod_{i \neq j} \int_0^{\infty} e^{-H_{ij}(w_{ij})} dw_{ij} = \prod_{i \neq j} \zeta_{ij} \tag{D23}$$

where the dyadic-specific conditional partition function ζ_{ij} reads

$$\zeta_{ij} = \left(\frac{1}{\beta_i^{\rightarrow} + \beta_j^{\leftarrow}}\right)^{a_{ij}^{\rightarrow}} \left(\frac{1}{\beta_i^{\leftrightarrow,out} + \beta_j^{\leftrightarrow,in}}\right)^{a_{ij}^{\leftrightarrow}} \tag{D24}$$

i.e. $\zeta_{ij} = (\beta_i^{\rightarrow} + \beta_j^{\leftarrow})^{-1}$ if $a_{ij}^{\rightarrow} = 1$ and $\zeta_{ij} = (\beta_i^{\leftrightarrow,out} + \beta_j^{\leftrightarrow,in})^{-1}$ if $a_{ij}^{\leftrightarrow} = 1$.

After computing the partition function, one gets

$$Q(\mathbf{W}|\mathbf{A}) = \prod_{i \neq j} \frac{e^{-H_{ij}(w_{ij})}}{\zeta_{ij}}. \tag{D25}$$

allowing us to define the log-likelihood function as

$$\mathcal{L} = \sum_{i \neq j} [-H_{ij}(w_{ij}) + \ln \zeta_{ij}]; \tag{D26}$$

in order to account for the variability of the binary adjacency matrix $\mathbf{A}$, we average the log-likelihood function over the ensemble and obtain its

generalized version $\mathcal{G} = \mathcal{G}^{\rightarrow} + \mathcal{G}^{\leftrightarrow}$ that can be decoupled into a *non-reciprocated* component $\mathcal{G}^{\rightarrow}$ and a *reciprocated* component $\mathcal{G}^{\leftrightarrow}$, reading

$$\mathcal{G}^{\rightarrow} = \sum_{i \neq j} [-(\beta_i^{\rightarrow} + \beta_j^{\leftarrow}) w_{ij} + p_{ij}^{\rightarrow} \ln\big(\beta_i^{\rightarrow} + \beta_j^{\leftarrow}\big)] \tag{D27}$$

and

$$\mathcal{G}^{\leftrightarrow} = \sum_{i \neq j} \Big[-(\beta_i^{\leftrightarrow,out} + \beta_j^{\leftrightarrow,in}) w_{ij} + p_{ij}^{\leftrightarrow} \ln\Big(\beta_i^{\leftrightarrow,out} + \beta_j^{\leftrightarrow,in}\Big)\Big] \tag{D28}$$

respectively. Hence, maximizing it amounts at solving the two sub-problems

$$\frac{\partial \mathcal{G}}{\partial \beta_i^{\rightarrow}} = -s_i^{\rightarrow} + \sum_{j(\neq i)} \frac{p_{ij}^{\rightarrow}}{\beta_i^{\rightarrow} + \beta_j^{\leftarrow}}, \; \forall \, i \tag{D29}$$

$$\frac{\partial \mathcal{G}}{\partial \beta_i^{\leftarrow}} = -s_i^{\leftarrow} + \sum_{j(\neq i)} \frac{p_{ij}^{\leftarrow}}{\beta_i^{\leftarrow} + \beta_j^{\rightarrow}}, \; \forall \, i \tag{D30}$$

and

$$\frac{\partial \mathcal{G}}{\partial \beta_i^{\leftrightarrow,out}} = -s_i^{\leftrightarrow,out} + \sum_{j(\neq i)} \frac{p_{ij}^{\leftrightarrow}}{\beta_i^{\leftrightarrow,out} + \beta_j^{\leftrightarrow,in}}, \; \forall \, i \tag{D31}$$

$$\frac{\partial \mathcal{G}}{\partial \beta_i^{\leftrightarrow,in}} = -s_i^{\leftrightarrow,in} + \sum_{j(\neq i)} \frac{p_{ij}^{\leftrightarrow}}{\beta_i^{\leftrightarrow,in} + \beta_j^{\leftrightarrow,out}}, \; \forall \, i \tag{D32}$$

Bibliography

[1] Marzio Di Vece, Diego Garlaschelli, and Tiziano Squartini. "Gravity models of networks: Integrating maximum-entropy and econometric approaches". en. In: *Physical Review Research* 4.3 (Aug. 2022), p. 033105. ISSN: 2643-1564. DOI: 10.1103/PhysRevResearch.4.033105.

[2] Marzio Di Vece, Diego Garlaschelli, and Tiziano Squartini. "Reconciling econometrics with continuous maximum-entropy network models". en. In: *Chaos, Solitons & Fractals* 166 (Jan. 2023), p. 112958. ISSN: 09600779. DOI: 10.1016/j.chaos.2022.112958.

[3] Marzio Di Vece, Diego Garlaschelli, and Tiziano Squartini. "Deterministic, quenched, and annealed parameter estimation for heterogeneous network models". In: *Phys. Rev. E* 108 (5 2023), p. 054301. DOI: 10.1103/PhysRevE.108.054301.

[4] Marzio Di Vece, Frank P. Pijpers, and Diego Garlaschelli. *Commodity-specific triads in the Dutch inter-industry production network*. 2023. arXiv: 2305.12179 [physics.soc-ph]. (Currently under review at Scientific Reports).

[5] Frank Schweitzer et al. "Economic Networks: The New Challenges". en. In: *Science* 325.5939 (July 2009), pp. 422–425. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1173644.

[6] Giorgio Fagiolo. "The international-trade network: gravity equations and topological properties". en. In: *Journal of Economic Interaction and Coordination* 5.1 (June 2010), pp. 1–25. ISSN: 1860-711X, 1860-7128. DOI: 10.1007/s11403-010-0061-y.

[7] Fabio Saracco et al. "Detecting early signs of the 2007–2008 crisis in the world trade". en. In: *Scientific Reports* 6.1 (July 2016), p. 30286. ISSN: 2045-2322. DOI: 10.1038/srep30286.

[8] T. Squartini and D. Garlaschelli. “Stationarity, non-stationarity and early warning signals in economic networks”. en. In: *Journal of Complex Networks* 3.1 (Mar. 2015), pp. 1–21. ISSN: 2051-1310, 2051-1329. DOI: 10.1093/comnet/cnu012.

[9] Stefano Schiavo, Javier Reyes, and Giorgio Fagiolo. “International trade and financial integration: a weighted network analysis”. en. In: *Quantitative Finance* 10.4 (Apr. 2010), pp. 389–399. ISSN: 1469-7688, 1469-7696. DOI: 10.1080/14697680902882420.

[10] Tiziano Squartini, Iman van Lelyveld, and Diego Garlaschelli. “Early-warning signals of topological collapse in interbank networks”. en. In: *Sci Rep* 3.1 (Nov. 2013), p. 3357. ISSN: 2045-2322. DOI: 10.1038/srep03357.

[11] Raja Kali and Javier Reyes. “Financial Contagion on the International Trade Network”. en. In: *Economic Inquiry* 48.4 (Oct. 2010), pp. 1072–1101. ISSN: 00952583. DOI: 10.1111/j.1465-7295.2009.00249.x.

[12] Raja Kali and Javier Reyes. “The architecture of globalization: a network approach to international economic integration”. en. In: *Journal of International Business Studies* 38.4 (July 2007), pp. 595–620. ISSN: 0047-2506, 1478-6990. DOI: 10.1057/palgrave.jibs.8400286.

[13] Michele Starnini, Marián Boguñá, and M. Ángeles Serrano. “The interconnected wealth of nations: Shock propagation on global trade-investment multiplex networks”. en. In: *Scientific Reports* 9.1 (Sept. 2019), p. 13079. ISSN: 2045-2322. DOI: 10.1038/s41598-019-49173-2.

[14] Matteo Barigozzi, Giorgio Fagiolo, and Diego Garlaschelli. “Multi-network of international trade: A commodity-specific analysis”. en. In: *Physical Review E* 81.4 (Apr. 2010), p. 046104. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.81.046104.

[15] Stefania Vitali, James B. Glattfelder, and Stefano Battiston. “The Network of Global Corporate Control”. en. In: *PLoS ONE* 6.10 (Oct. 2011). Ed. by Alejandro Raul Hernandez Montoya, e25995. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0025995.

[16] Agata Fronczak. “Structural Hamiltonian of the international trade network”. In: *Acta Physica Polonica B* 5 (2012), pp. 31–46. DOI: 10.48550/ARXIV.1205.4589.

[17] Agata Fronczak and Piotr Fronczak. “Statistical mechanics of the international trade network”. en. In: *Physical Review E* 85.5 (May 2012), p. 056113. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.85.056113.

[18] Ma Ángeles Serrano and Marián Boguñá. “Topology of the world trade web”. en. In: *Physical Review E* 68.1 (July 2003), p. 015101. ISSN: 1063-651X, 1095-3787. DOI: 10.1103/PhysRevE.68.015101.

[19] Diego Garlaschelli and Maria I. Loffredo. “Fitness-Dependent Topological Properties of the World Trade Web”. en. In: *Physical Review Letters* 93.18 (Oct. 2004), p. 188701. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.93.188701.

[20] Diego Garlaschelli and Maria I. Loffredo. “Structure and evolution of the world trade network”. en. In: *Physica A: Statistical Mechanics and its Applications* 355.1 (Sept. 2005), pp. 138–144. ISSN: 03784371. DOI: 10.1016/j.physa.2005.02.075.

[21] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. “The evolution of the world trade web: a weighted-network analysis”. en. In: *Journal of Evolutionary Economics* 20.4 (Aug. 2010), pp. 479–514. ISSN: 0936-9937, 1432-1386. DOI: 10.1007/s00191-009-0160-x.

[22] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. “On the topological properties of the world trade web: A weighted network analysis”. en. In: *Physica A: Statistical Mechanics and its Applications* 387.15 (June 2008), pp. 3868–3873. ISSN: 03784371. DOI: 10.1016/j.physa.2008.01.050.

[23] Giorgio Fagiolo, Javier Reyes, and Stefano Schiavo. “World-trade web: Topological properties, dynamics, and evolution”. en. In: *Physical Review E* 79.3 (Mar. 2009), p. 036115. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.79.036115.

[24] Jeroen van Lidth de Jeude et al. “Reconstructing Mesoscale Network Structures”. en. In: *Complexity* 2019 (Jan. 2019), pp. 1–13. ISSN: 1076-2787, 1099-0526. DOI: 10.1155/2019/5120581.

[25] Jan Tinbergen. *Shaping the World Economy; Suggestions for an International Economic Policy*. en. Twentieth Century Fund, New York, Jan. 1962.

[26] James E. Anderson. “The Gravity Model”. en. In: *Annual Review of Economics* 3.1 (Sept. 2011), pp. 133–160. ISSN: 1941-1383, 1941-1391. DOI: 10.1146/annurev-economics-111809-125114.

[27] Peter A. G. van Bergeijk and Steven Brakman, eds. *The Gravity Model in International Trade: Advances and Applications*. 1st ed. Cambridge University Press, June 2010. ISBN: 9780511762109. DOI: 10.1017/CBO9780511762109.

[28] Luca De Benedictis and Daria Taglioni. “The Trade Impact of European Union Preferential Policies”. en. In: *The Trade Impact of European Union Preferential Policies*. Ed. by Luca De Benedictis and Luca Salvatici. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 55–89. ISBN: 978-3-642-16563-4. DOI: 10.1007/978-3-642-16564-1_4.

[29] Assaf Almog, Rhys Bird, and Diego Garlaschelli. “Enhanced Gravity Model of Trade: Reconciling Macroeconomic and Network Models”. In: *Frontiers in Physics* 7 (Apr. 2019), p. 55. ISSN: 2296-424X. DOI: 10.3389/fphy.2019.00055.

[30] E. G. Ravenstein. “The Laws of Migration”. In: *Journal of the Royal Statistical Society* 52.2 (June 1889), p. 241. ISSN: 09528385. DOI: 10.2307/2979333.

[31] Giorgio Fagiolo and Marina Mastrorillo. “International migration network: Topology and modeling”. en. In: *Physical Review E* 88.1 (July 2013), p. 012812. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.88.012812.

[32] George Kingsley Zipf. “The P 1 P 2 D Hypothesis: On the Intercity Movement of Persons”. In: *American Sociological Review* 11.6 (Dec. 1946), p. 677. ISSN: 00031224. DOI: 10.2307/2087063.

[33] Duygu Balcan et al. “Multiscale mobility networks and the spatial spreading of infectious diseases”. en. In: *Proceedings of the National Academy of Sciences* 106.51 (Dec. 2009), pp. 21484–21489. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0906910106.

[34] Woo-Sung Jung, Fengzhong Wang, and H. Eugene Stanley. “Gravity model in the Korean highway”. In: *EPL (Europhysics Letters)* 81.4 (Feb. 2008), p. 48005. ISSN: 0295-5075, 1286-4854. DOI: 10.1209/0295-5075/81/48005.

[35] Gautier Krings et al. “Urban gravity: a model for inter-city telecommunication flows”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2009.07 (July 2009), p. L07003. ISSN: 1742-5468. DOI: 10.1088/1742-5468/2009/07/L07003.

[36] Ling-ling Ma et al. “Identifying influential spreaders in complex networks based on gravity formula”. en. In: *Physica A: Statistical Mechanics and its Applications* 451 (June 2016), pp. 205–212. ISSN: 03784371. DOI: 10.1016/j.physa.2015.12.162.

[37] Zhe Li et al. “Identifying influential spreaders by gravity model”. en. In: *Scientific Reports* 9.1 (June 2019), p. 8387. ISSN: 2045-2322. DOI: 10.1038/s41598-019-44930-9.

[38] Kristian Skrede Gleditsch. “Expanded Trade and GDP Data”. In: *The Journal of Conflict Resolution* 46.5 (2002). Publisher: Sage Publications, Inc., pp. 712–724. ISSN: 0022-0027.

[39] Jonathan Eaton and Akiko Tamura. “Bilateralism and Regionalism in Japanese and U.S. Trade and Direct Foreign Investment Patterns”. en. In: *Journal of the Japanese and International Economies* 8.4 (Dec. 1994), pp. 478–510. ISSN: 08891583. DOI: 10.1006/jjie.1994.1025.

[40] William Martin and Cong S. Pham. “Estimating the gravity model when zero trade flows are frequent and economically determined”. en. In: *Applied Economics* 52.26 (June 2020), pp. 2766–2779. ISSN: 0003-6846, 1466-4283. DOI: 10.1080/00036846.2019.1687838.

[41] Elhanan Helpman, Marc Melitz, and Yona Rubinstein. “Estimating Trade Flows: Trading Partners and Trading Volumes*”. en. In: *Quarterly Journal of Economics* 123.2 (May 2008), pp. 441–487. ISSN: 0033-5533, 1531-4650. DOI: 10.1162/qjec.2008.123.2.441.

[42] J. M. C. Santos Silva and Silvana Tenreyro. “The Log of Gravity”. en. In: *The Review of Economics and Statistics* 88.4 (Nov. 2006), pp. 641–658. ISSN: 0034-6535, 1530-9142. DOI: 10.1162/rest.88.4.641.

[43] Martijn Burger, Frank van Oort, and Gert-Jan Linders. “On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation”. en. In: *Spatial Economic Analysis* 4.2 (June 2009), pp. 167–190. ISSN: 1742-1772, 1742-1780. DOI: 10.1080/17421770902834327.

[44] Rainer Winkelmann. *Econometric analysis of count data*. eng. 5th ed. Berlin: Springer, 2008. ISBN: 978-3-540-78389-3.

[45] Marco Dueñas and Giorgio Fagiolo. "Modeling the International-Trade Network: a gravity approach". en. In: *Journal of Economic Interaction and Coordination* 8.1 (Apr. 2013), pp. 155–178. ISSN: 1860-711X, 1860-7128. DOI: 10.1007/s11403-013-0108-y.

[46] Tiziano Squartini, Giorgio Fagiolo, and Diego Garlaschelli. "Randomizing world trade. I. A binary network analysis". en. In: *Physical Review E* 84.4 (Oct. 2011), p. 046117. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.84.046117.

[47] Tiziano Squartini et al. "Enhanced capital-asset pricing model for the reconstruction of bipartite financial networks". en. In: *Physical Review E* 96.3 (Sept. 2017), p. 032315. ISSN: 2470-0045, 2470-0053. DOI: 10.1103/PhysRevE.96.032315.

[48] Tiziano Squartini, Rossana Mastrandrea, and Diego Garlaschelli. "Unbiased sampling of network ensembles". In: *New Journal of Physics* 17.2 (Feb. 2015), p. 023052. ISSN: 1367-2630. DOI: 10.1088/1367-2630/17/2/023052.

[49] Giulio Cimini, Rossana Mastrandrea, and Tiziano Squartini. *Reconstructing Networks*. Elements in the Structure and Dynamics of Complex Networks. Cambridge University Press, 2021. DOI: 10.1017/9781108771030.

[50] Tiziano Squartini and Diego Garlaschelli. "Analytical maximum-likelihood method to detect patterns in real networks". en. In: *New J. Phys.* 13.8 (Aug. 2011), p. 083001. ISSN: 1367-2630. DOI: 10.1088/1367-2630/13/8/083001.

[51] Tiziano Squartini, Giorgio Fagiolo, and Diego Garlaschelli. "Randomizing world trade. II. A weighted network analysis". en. In: *Physical Review E* 84.4 (Oct. 2011), p. 046118. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.84.046118.

[52] Giorgio Fagiolo, Tiziano Squartini, and Diego Garlaschelli. "Null models of economic networks: the case of the world trade web". en. In: *Journal of Economic Interaction and Coordination* 8.1 (Apr. 2013), pp. 75–107. ISSN: 1860-711X, 1860-7128. DOI: 10.1007/s11403-012-0104-7.

[53] Rossana Mastrandrea et al. “Reconstructing the world trade multiplex: The role of intensive and extensive biases”. en. In: *Physical Review E* 90.6 (Dec. 2014), p. 062804. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.90.062804.

[54] Diego Garlaschelli and Maria I. Loffredo. “Generalized Bose-Fermi Statistics and Structural Correlations in Weighted Networks”. en. In: *Physical Review Letters* 102.3 (Jan. 2009), p. 038701. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.102.038701.

[55] Andrea Gabrielli et al. “Grand canonical ensemble of weighted networks”. en. In: *Physical Review E* 99.3 (Mar. 2019), p. 030301. ISSN: 2470-0045, 2470-0053. DOI: 10.1103/PhysRevE.99.030301.

[56] Rossana Mastrandrea et al. “Enhanced reconstruction of weighted networks from strengths and degrees”. In: *New Journal of Physics* 16.4 (Apr. 2014), p. 043022. ISSN: 1367-2630. DOI: 10.1088/1367-2630/16/4/043022.

[57] Leonardo Bargigli. “Statistical Ensembles for Economic Networks”. en. In: *Journal of Statistical Physics* 155.4 (May 2014), pp. 810–825. ISSN: 0022-4715, 1572-9613. DOI: 10.1007/s10955-014-0968-0.

[58] Tiziano Squartini and Diego Garlaschelli. *Maximum-Entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics*. SpringerBriefs in Complexity. Springer Cham, 2017. DOI: 10.1007/978-3-319-69438-2.

[59] E. T. Jaynes. “Information Theory and Statistical Mechanics”. en. In: *Physical Review* 106.4 (May 1957), pp. 620–630. ISSN: 0031-899X. DOI: 10.1103/PhysRev.106.620.

[60] E. T. Jaynes. “Information Theory and Statistical Mechanics. II”. en. In: *Physical Review* 108.2 (Oct. 1957), pp. 171–190. ISSN: 0031-899X. DOI: 10.1103/PhysRev.108.171.

[61] E.T. Jaynes. “On the rationale of maximum-entropy methods”. In: *Proceedings of the IEEE* 70.9 (1982), pp. 939–952. ISSN: 0018-9219. DOI: 10.1109/PROC.1982.12425.

[62] Giulio Cimini et al. “The statistical physics of real-world networks”. en. In: *Nature Reviews Physics* 1.1 (Jan. 2019), pp. 58–71. ISSN: 2522-5820. DOI: 10.1038/s42254-018-0002-6.

[63] Federica Parisi, Tiziano Squartini, and Diego Garlaschelli. "A faster horse on a safer trail: generalized inference for the efficient reconstruction of weighted networks". en. In: *New J. Phys.* 22.5 (May 2020), p. 053053. ISSN: 1367-2630. DOI: 10.1088/1367-2630/ab74a7.

[64] G. Caldarelli et al. "Scale-Free Networks from Varying Vertex Intrinsic Fitness". en. In: *Physical Review Letters* 89.25 (Dec. 2002), p. 258702. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.89.258702.

[65] Assaf Almog, Tiziano Squartini, and Diego Garlaschelli. "A GDP-driven model for the binary and weighted structure of the International Trade Network". In: *New Journal of Physics* 17.1 (Jan. 2015), p. 013009. ISSN: 1367-2630. DOI: 10.1088/1367-2630/17/1/013009.

[66] Assaf Almog, Tiziano Squartini, and Diego Garlaschelli. "The double role of GDP in shaping the structure of the International Trade Network". en. In: *International Journal of Computational Economics and Econometrics* 7.4 (2017), p. 381. ISSN: 1757-1170, 1757-1189. DOI: 10.1504/IJCEE.2017.086875.

[67] Marián Boguñá et al. "Small worlds and clustering in spatial networks". In: *Phys. Rev. Res.* 2 (2 2020), p. 023040. DOI: 10.1103/PhysRevResearch.2.023040. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.2.023040.

[68] Antoine Allard et al. "The geometric nature of weights in real complex networks". In: *Nature Communications* 8.1 (Jan. 2017), p. 14103. ISSN: 2041-1723. DOI: 10.1038/ncomms14103. URL: https://doi.org/10.1038/ncomms14103.

[69] Guillermo García-Pérez et al. "The hidden hyperbolic geometry of international trade: World Trade Atlas 1870–2013". In: *Scientific Reports* 6.1 (Sept. 2016), p. 33441. ISSN: 2045-2322. DOI: 10.1038/srep33441. URL: https://doi.org/10.1038/srep33441.

[70] John Mullahy. "Specification and testing of some modified count data models". en. In: *Journal of Econometrics* 33.3 (Dec. 1986), pp. 341–365. ISSN: 03044076. DOI: 10.1016/0304-4076(86)90002-3.

[71] Sjoerd Hooijmaaijers and Gert Buiten. "A methodology for estimating the Dutch interfirm trade network, including a breakdown by commodity". In: *OECD Conference: New Approaches to Economic Challenges* (2019), p. 18.

[72] Takaaki Ohnishi, Hideki Takayasu, and Misako Takayasu. "Network motifs in an inter-firm network". en. In: *J Econ Interact Coord* 5.2 (Dec. 2010), pp. 171–180. ISSN: 1860-711X, 1860-7128. DOI: 10.1007/s11403-010-0066-6.

[73] Carolina E. S. Mattsson et al. "Functional Structure in Production Networks". In: *Frontiers in Big Data* 4 (2021). ISSN: 2624-909X.

[74] Keith Head, Thierry Mayer, and John Ries. "The erosion of colonial trade linkages after independence". en. In: *Journal of International Economics* 81.1 (May 2010), pp. 1–14. ISSN: 00221996. DOI: 10.1016/j.jinteco.2010.01.002.

[75] Keith Head and Thierry Mayer. "Gravity Equations: Workhorse, Toolkit, and Cookbook". en. In: *Handbook of International Economics*. Vol. 4. Elsevier, 2014, pp. 131–195. ISBN: 978-0-444-54314-1. DOI: 10.1016/B978-0-444-54314-1.00003-3.

[76] Joseph M. Hilbe. *Negative Binomial Regression*. 2nd ed. Cambridge University Press, Mar. 2011. ISBN: 9780511973420. DOI: 10.1017/CBO9780511973420.

[77] Tiziano Squartini et al. "Reconstruction methods for networks: The case of economic and financial systems". en. In: *Physics Reports* 757 (Oct. 2018), pp. 1–47. ISSN: 03701573. DOI: 10.1016/j.physrep.2018.06.008.

[78] Giulio Cimini et al. "Systemic Risk Analysis on Reconstructed Economic and Financial Networks". en. In: *Sci Rep* 5.1 (Oct. 2015), p. 15758. ISSN: 2045-2322. DOI: 10.1038/srep15758.

[79] Akanda Wahid-Ul-Ashraf, Marcin Budka, and Katarzyna Musial. "How to predict social relationships — Physics-inspired approach to link prediction". en. In: *Physica A: Statistical Mechanics and its Applications* 523 (June 2019), pp. 1110–1129. ISSN: 03784371. DOI: 10.1016/j.physa.2019.04.246.

[80] Peter R. Herman. "Modeling complex network patterns in international trade". en. In: *Review of World Economics* 158.1 (Feb. 2022), pp. 127–179. ISSN: 1610-2878, 1610-2886. DOI: 10.1007/s10290-021-00429-y.

[81] J. A. Nelder and R. W. M. Wedderburn. "Generalized Linear Models". In: *Journal of the Royal Statistical Society. Series A (General)* 135.3 (1972), p. 370. ISSN: 00359238. DOI: 10.2307/2344614.

[82] C. E. Shannon. “A Mathematical Theory of Communication”. en. In: *Bell System Technical Journal* 27.3 (July 1948), pp. 379–423. ISSN: 00058580. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[83] Ian Goodfellow et al. “Generative Adversarial Nets”. In: *Advances in Neural Information Processing Systems*. Ed. by Z. Ghahramani et al. Vol. 27. Curran Associates, Inc., 2014.

[84] S. Kullback and R. A. Leibler. “On Information and Sufficiency”. en. In: *The Annals of Mathematical Statistics* 22.1 (Mar. 1951), pp. 79–86. ISSN: 0003-4851. DOI: 10.1214/aoms/1177729694.

[85] Matt Visser. “Zipf’s law, power laws and maximum entropy”. In: *New Journal of Physics* 15.4 (Apr. 2013), p. 043021. ISSN: 1367-2630. DOI: 10.1088/1367-2630/15/4/043021.

[86] Hirotogu Akaike. “Information Theory and an Extension of the Maximum Likelihood Principle”. In: *Selected Papers of Hirotugu Akaike*. Ed. by Emanuel Parzen, Kunio Tanabe, and Genshiro Kitagawa. Series Title: Springer Series in Statistics. New York, NY: Springer New York, 1998, pp. 199–213. ISBN: 978-1-4612-1694-0. DOI: 10.1007/978-1-4612-1694-0_15.

[87] C. Vignat and J.-F. Bercher. “Analysis of signals in the Fisher–Shannon information plane”. en. In: *Physics Letters A* 312.1-2 (June 2003), pp. 27–33. ISSN: 03759601. DOI: 10.1016/S0375-9601(03)00570-X.

[88] Osvaldo A. Rosso et al. “Causality and the entropy–complexity plane: Robustness and missing ordinal patterns”. en. In: *Physica A: Statistical Mechanics and its Applications* 391.1-2 (Jan. 2012), pp. 42–55. ISSN: 03784371. DOI: 10.1016/j.physa.2011.07.030.

[89] Martín Gómez Ravetti et al. “Distinguishing Noise from Chaos: Objective versus Subjective Criteria Using Horizontal Visibility Graph”. en. In: *PLoS ONE* 9.9 (Sept. 2014). Ed. by Satoru Hayasaka, e108004. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0108004.

[90] Yuanyuan Wang and Pengjian Shang. “Analysis of Shannon-Fisher information plane in time series based on information entropy”. en. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28.10 (Oct. 2018), p. 103107. ISSN: 1054-1500, 1089-7682. DOI: 10.1063/1.5023031.

[91] Leonardo H.S. Fernandes et al. “Macroeconophysics indicator of economic efficiency”. en. In: *Physica A: Statistical Mechanics and its Applications* 573 (July 2021), p. 125946. ISSN: 03784371. DOI: 10.1016/j.physa.2021.125946.

[92] Tiziano Squartini et al. “Breaking of Ensemble Equivalence in Networks”. en. In: *Physical Review Letters* 115.26 (Dec. 2015), p. 268701. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.115.268701.

[93] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. en. 1st ed. Wiley, Sept. 2005. ISBN: 978-0-471-24195-9. DOI: 10.1002/047174882X.

[94] Jian-Hong Lin et al. “The weighted Bitcoin Lightning Network”. In: *Chaos, Solitons & Fractals* 164 (2022), p. 112620. ISSN: 0960-0779. DOI: https://doi.org/10.1016/j.chaos.2022.112620.

[95] Daron Acemoglu et al. “The Network Origins of Aggregate Fluctuations”. en. In: *Econometrica* 80.5 (2012), pp. 1977–2016. ISSN: 1468-0262. DOI: 10.3982/ECTA9623.

[96] Daniel Aobdia, Judson Caskey, and N. Bugra Ozel. “Inter-industry network structure and the cross-predictability of earnings and stock returns”. en. In: *Rev Account Stud* 19.3 (Sept. 2014), pp. 1191–1224. ISSN: 1380-6653, 1573-7136. DOI: 10.1007/s11142-014-9286-7.

[97] Enghin Atalay. “How Important Are Sectoral Shocks?” en. In: *American Economic Journal: Macroeconomics* 9.4 (Oct. 2017), pp. 254–280. ISSN: 1945-7707. DOI: 10.1257/mac.20160353.

[98] Hafedh Bouakez, Emanuela Cardia, and Francisco J. Ruge-Murcia. “The Transmission of Monetary Policy in a Multi-Sector Economy”. en. In: *International Economic Review* 50.4 (Nov. 2009), pp. 1243–1266. ISSN: 00206598, 14682354. DOI: 10.1111/j.1468-2354.2009.00567.x.

[99] A. Brintrup et al. “Predicting Hidden Links in Supply Networks”. en. In: *Complexity* 2018 (Jan. 2018), e9104387. ISSN: 1076-2787. DOI: 10.1155/2018/9104387.

[100] Anton Pichler and J. Doyne Farmer. "Simultaneous supply and demand constraints in input–output networks: the case of Covid-19 in Germany, Italy, and Spain". In: *Economic Systems Research* 34.3 (July 2022), pp. 273–293. ISSN: 0953-5314. DOI: 10.1080/09535314.2021.1926934.

[101] Enghin Atalay et al. "Network structure of production". In: *Proceedings of the National Academy of Sciences* 108.13 (Mar. 2011), pp. 5199–5202. DOI: 10.1073/pnas.1015564108.

[102] Hayato Goto, Hideki Takayasu, and Misako Takayasu. "Estimating risk propagation between interacting firms on inter-firm complex network". eng. In: *PLoS One* 12.10 (2017), e0185712. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0185712.

[103] Leonardo Niccolò Ialongo et al. "Reconstructing firm-level interactions in the Dutch input–output network from production constraints". en. In: *Sci Rep* 12.1 (July 2022), p. 11847. ISSN: 2045-2322. DOI: 10.1038/s41598-022-13996-3.

[104] Hiroyasu Inoue and Yasuyuki Todo. "Firm-level propagation of shocks through supply-chain networks". en. In: *Nat Sustain* 2.9 (Sept. 2019), pp. 841–847. ISSN: 2398-9629. DOI: 10.1038/s41893-019-0351-x.

[105] Hiroyasu Inoue and Yasuyuki Todo. "The propagation of economic impacts through supply chains: The case of a mega-city lockdown to prevent the spread of COVID-19". en. In: *PLoS ONE* 15.9 (Sept. 2020). Ed. by Lizhi Xing, e0239251. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0239251.

[106] Yuzuka Kashiwagi, Yasuyuki Todo, and Petr Matous. "Propagation of economic shocks through global supply chains—Evidence from Hurricane Sandy". en. In: *Review of International Economics* 29.5 (2021), pp. 1186–1220. ISSN: 1467-9396. DOI: 10.1111/roie.12541.

[107] Michael D. König et al. "Aggregate fluctuations in adaptive production networks". In: *Proceedings of the National Academy of Sciences* 119.38 (Sept. 2022), e2203730119. DOI: 10.1073/pnas.2203730119.

[108] Edward Elson Kosasih and Alexandra Brintrup. "A machine learning approach for predicting hidden links in supply chain with graph neural networks". In: *International Journal of Production Research* 60.17 (Sept. 2022), pp. 5380–5393. ISSN: 0020-7543. DOI: 10.1080/00207543.2021.1956697.

[109] Julian Maluck et al. "Motif formation and industry specific topologies in the Japanese business firm network". In: *J. Stat. Mech.* 2017.5 (May 2017), p. 053404. ISSN: 1742-5468. DOI: 10.1088/1742-5468/aa6ddb.

[110] Andrew B. Bernard, Andreas Moxnes, and Yukiko U. Saito. "Production Networks, Geography, and Firm Performance". In: *Journal of Political Economy* 127.2 (Apr. 2019), pp. 639–688. ISSN: 0022-3808. DOI: 10.1086/700764.

[111] James McNerney et al. "How production networks amplify economic growth". In: *Proceedings of the National Academy of Sciences* 119.1 (2022), e2106031118. DOI: 10.1073/pnas.2106031118.

[112] Takayuki Mizuno, Wataru Souma, and Tsutomu Watanabe. "The Structure and Evolution of Buyer-Supplier Networks". en. In: *PLOS ONE* 9.7 (July 2014), e100712. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0100712.

[113] A Rachkov, Frank Pijpers, and D Garlaschelli. *Potential Biases in Network Reconstruction Methods Not Maximizing Entropy*. Mar. 2021. DOI: 10.13140/RG.2.2.31861.29925.

[114] Mathieu Taschereau-Dumouchel. "Cascades and Fluctuations in an Economy with an Endogenous Production Network". en. In: *2017 Meeting Papers* (2017).

[115] Hayafumi Watanabe, Hideki Takayasu, and Misako Takayasu. "Relations between allometric scalings and fluctuations in complex systems: The case of Japanese firms". In: *Physica A: Statistical Mechanics and its Applications* 392.4 (2013), pp. 741–756. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2012.10.020.

[116] Gert Buiten et al. *Reconstruction method for the Dutch interfirm network including a breakdown by commodity for 2018 and 2019 (v1.0)*. Dec. 2021. DOI: 10.13140/RG.2.2.16685.77286.

[117] Vasco M Carvalho et al. “Supply Chain Disruptions: Evidence from the Great East Japan Earthquake*”. In: *The Quarterly Journal of Economics* 136.2 (Dec. 2020), pp. 1255–1321. ISSN: 0033-5533. DOI: 10.1093/qje/qjaa044.

[118] Vasco M. Carvalho and Alireza Tahbaz-Salehi. “Production Networks: A Primer”. In: *Annual Review of Economics* 11.1 (2019), pp. 635–663. DOI: 10.1146/annurev-economics-080218-030212.

[119] Lauren Cohen and Andrea Frazzini. “Economic Links and Predictable Returns”. In: *The Journal of Finance* 63.4 (2008), pp. 1977–2011. ISSN: 0022-1082.

[120] Emmanuel Dhyne, Glenn Magerman, and Stela Rubínova. *The Belgian production network 2002-2012*. eng. Working Paper 288. NBB Working Paper, 2015.

[121] Emmanuel Dhyne et al. “Trade and Domestic Production Networks”. In: *The Review of Economic Studies* 88.2 (Mar. 2021), pp. 643–668. ISSN: 0034-6527. DOI: 10.1093/restud/rdaa062.

[122] Christian Diem et al. “Quantifying firm-level economic systemic risk from nation-wide supply networks”. eng. In: *Sci Rep* 12.1 (May 2022), p. 7719. ISSN: 2045-2322. DOI: 10.1038/s41598-022-11522-z.

[123] Julian Maluck and Reik V. Donner. “A Network of Networks Perspective on Global Trade”. en. In: *PLOS ONE* 10.7 (July 2015), e0133310. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0133310.

[124] Ze Wang et al. “Motif Transition Intensity: A Novel Network-Based Early Warning Indicator for Financial Crises”. In: *Frontiers in Physics* 9 (2022). ISSN: 2296-424X.

[125] Luca Mungo et al. “Reconstructing production networks using machine learning”. In: *Journal of Economic Dynamics and Control* 148 (2023), p. 104607. ISSN: 0165-1889. DOI: https://doi.org/10.1016/j.jedc.2023.104607.

[126] François Lafond et al. “Firm-level production networks: what do we (really) know?” In: *INET Oxford Working Papers* (2023).

[127] Peter Wankuru Chacha, Benard Kirui, and Verena Wiedemann. “Mapping Kenya’s Production Network Transaction by Transaction”. In: *SSRN* (2022). DOI: http://dx.doi.org/10.2139/ssrn.4315810.

[128] Jose-Luis Peydro et al. “Production and Financial Networks in Interplay: Crisis Evidence from Supplier-Customer and Credit Registers”. In: *CEPR Discussion Paper* No. DP15277 ().

[129] Banu Demir et al. “Financial Constraints and Propagation of Shocks in Production Networks”. In: *The Review of Economics and Statistics* (Jan. 2022), pp. 1–46. ISSN: 0034-6535. DOI: 10.1162/rest_a_01162.

[130] Richard Newfarmer, John Page, and Finn Tarp. *Industries Without Smokestacks: Industrialization in Africa Reconsidered*. Oxford University Press, 2018.

[131] Ashish Kumar et al. “Distress propagation on production networks: Coarse-graining and modularity of linkages”. In: *Physica A: Statistical Mechanics and its Applications* 568 (2021), p. 125714. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2020.125714.

[132] Alonso Alfaro-Ureña et al. “The costa rican production network: Stylized facts”. In: *Research Paper Series* (2018).

[133] Marvin Cardoza et al. “Worker Mobility and Domestic Production Networks”. In: *IMF Working Papers* 2020/205 (Sept. 2020).

[134] R. Milo et al. “Network Motifs: Simple Building Blocks of Complex Networks”. In: *Science* 298.5594 (Oct. 2002), pp. 824–827. DOI: 10.1126/science.298.5594.824.

[135] Shai S. Shen-Orr et al. “Network motifs in the transcriptional regulation network of Escherichia coli”. en. In: *Nature Genetics* 31.1 (May 2002), pp. 64–68. ISSN: 1546-1718. DOI: 10.1038/ng881.

[136] Alex Stivala and Alessandro Lomi. “Testing biological network motif significance with exponential random graph models”. en. In: *Appl Netw Sci* 6.1 (Dec. 2021), pp. 1–27. ISSN: 2364-8228. DOI: 10.1007/s41109-021-00434-y.

[137] Aili Asikainen et al. “Cumulative effects of triadic closure and homophily in social networks”. In: *Science Advances* 6.19 (May 2020), eaax7310. DOI: 10.1126/sciadv.aax7310.

[138] Tiziano Squartini and Diego Garlaschelli. “Triadic Motifs and Dyadic Self-Organization in the World Trade Network”. en. In: *Self-Organizing Systems*. Ed. by David Hutchison et al. Vol. 7166. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 24–35. ISBN: 978-3-642-28583-7. DOI: 10.1007/978-3-642-28583-7_3.

[139] Pol Colomer-de-Simón et al. “Deciphering the global organization of clustering in real complex networks”. en. In: *Sci Rep* 3.1 (Aug. 2013), p. 2517. ISSN: 2045-2322. DOI: 10.1038/srep02517.

[140] Almerima Jamakovic et al. “How small are building blocks of complex networks”. en. In: (Aug. 2009). DOI: 10.48550/arXiv.0908.1143.

[141] Francesco Picciolo et al. “Weighted network motifs as random walk patterns”. en. In: *New J. Phys.* 24.5 (June 2022), p. 053056. ISSN: 1367-2630. DOI: 10.1088/1367-2630/ac6f75.

[142] Diego Garlaschelli and Maria I. Loffredo. “Maximum likelihood: Extracting unbiased information from complex networks”. In: *Phys. Rev. E* 78.1 (July 2008), p. 015101. DOI: 10.1103/PhysRevE.78.015101.

[143] Marco Bardoscia et al. “The physics of financial networks”. en. In: *Nature Reviews Physics* 3.7 (June 2021), pp. 490–507. ISSN: 2522-5820. DOI: 10.1038/s42254-021-00322-5.

[144] Kartik Anand et al. “The missing links: A global study on uncovering financial network structures from partial data”. en. In: *Journal of Financial Stability*. Network models, stress testing and other tools for financial stability monitoring and macroprudential policy design and implementation 35 (Apr. 2018), pp. 107–119. ISSN: 1572-3089. DOI: 10.1016/j.jfs.2017.05.012.

[145] Michael Lebacher et al. *In Search of Lost Edges: A Case Study on Reconstructing Financial Networks*. 2019. DOI: 10.48550/ARXIV.1909.01274.

[146] Amanah Ramadiah, Fabio Caccioli, and Daniel Fricke. “Reconstructing and stress testing credit networks”. In: *Journal of Economic Dynamics and Control* 111 (2020), p. 103817. ISSN: 0165-1889. DOI: https://doi.org/10.1016/j.jedc.2019.103817.

[147] Piero Mazzarisi and Fabrizio Lillo. “Methods for Reconstructing Interbank Networks from Limited Information: A Comparison”. en. In: *Econophysics and Sociophysics: Recent Progress and Future Directions*. Ed. by Frédéric Abergel et al. New Economic Windows. Cham: Springer International Publishing, 2017, pp. 201–215. ISBN: 978-3-319-47705-3. DOI: 10.1007/978-3-319-47705-3_15.

[148] Tiziano Squartini, Francesco Picciolo, and Franco Ruzzenenti. "Reciprocity of weighted networks". In: *Scientific Reports* 3 (2013), p. 2729. ISSN: 0165-1889. DOI: 10.1038/srep02729.

[149] S. S. Shapiro and M. B. Wilk. "An analysis of variance test for normality (complete samples)†". In: *Biometrika* 52.3-4 (Dec. 1965), pp. 591–611. ISSN: 0006-3444. DOI: 10.1093/biomet/52.3-4.591.

[150] Andras Borsos and Martin Stancsics. *Unfolding the hidden structure of the Hungarian multi-layer firm network*. eng. MNB Occasional Papers 139. Budapest, 2020. URL: http://hdl.handle.net/10419/241068.

[151] Christian Diem et al. *Estimating the loss of economic predictability from aggregating firm-level production networks*. 2023. arXiv: 2302.11451 [econ.GN].

[152] John G. Cragg. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods". In: *Econometrica* 39.5 (Sept. 1971), p. 829. ISSN: 00129682. DOI: 10.2307/1909582.

[153] Leonardo Bargigli, Stefano Viaggiu, and Andrea Lionetto. "A Statistical Test of Walrasian Equilibrium by Means of Complex Networks Theory". en. In: *Journal of Statistical Physics* 165.2 (Oct. 2016), pp. 351–370. ISSN: 0022-4715, 1572-9613. DOI: 10.1007/s10955-016-1599-4.