

Empirical Applications of Network and Random Matrix Theories to Economic and Financial Complex Systems

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Wirtschafts- und Sozialwissenschaften
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
M.A., Duc Thi Luu
aus Thanh Hoa, Viet Nam

Kiel, February 6, 2017

Gedruckt mit Genehmigung der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

Dekan/in: *Prof. Dr. Till Requate*

Erstberichterstattender: *Prof. Dr. Thomas Lux*

Zweitberichterstattender: *Prof. Dr. Martin F. Quaas*

Tag der Abgabe der Arbeit: Octobre 11, 2016

Tag der mündlichen Prüfung: Januar 25, 2017

Dedicated to

my parents, all of my professors, and my love.

Acknowledgements

First of all, I would like to thank my supervisor Prof. Dr. Thomas Lux for his guidance and tremendous support during my doctoral studies. I am also grateful to Prof. Dr. Martin F. Quaas for his acceptance to be my second advisor. In addition, I wish to thank the doctoral Programme “*Quantitative Economics*” and the Federal State Funding at Kiel University for providing me financial support during the second and third years of my doctoral studies. My thanks go to Boyan Yanovski, who is one of my great collaborators as well as my closest friends in Kiel. I also would like to express my sincerely thanks to my parents, my brothers, and my girlfriend. This thesis could not have been written without their love, encouragement, and financial support.

Contents

Acknowledgements	i
List of Figures	vi
Notation and Abbreviations	xv
Preface	xvii
1 Introduction	1
2 Does the Heterogeneity in the Local Constraints Predominantly Determine Structural Correlations in the Italian Overnight Money Market?	8
2.1 Introduction	8
2.2 Structural correlations in complex networks	12
2.2.1 For undirected networks	12
2.2.2 For directed networks	14
2.2.3 Configuration models	20
2.3 Findings for the binary network	24
2.3.1 Structural correlations in the undirected binary e-MID network	24
2.3.2 Structural correlations in the directed binary e-MID network	28
2.3.3 Comparisons to the configuration models	37
2.4 Findings for the weighted network	50
2.4.1 Structural correlations in the undirected weighted e-MID network	50
2.4.2 Structural correlations in the directed weighted e-MID network	52
2.4.3 Comparisons to the weighted configuration models	60
2.4.4 z-scores analysis revealing structural changes in the weighted system	86
2.5 Conclusions	91
2.6 References	94

2.7	Appendix	96
2.7.1	Assortativity Coefficients	96
2.7.2	z-scores analysis of the indicators of structural correlations	100
3	An Approach to Identify Patterns in Structural Similarities in Financial Networks	102
3.1	Introduction	102
3.2	Similarity and dissimilarity in networks	104
3.2.1	General definitions and configuration models	104
3.2.2	Structural similarity	107
3.3	Findings	113
3.3.1	Binary analysis	113
3.3.2	Weighted analysis	121
3.3.3	Hidden variables, structural similarities, and non-random patterns	129
3.4	Conclusions	133
3.5	References	134
3.6	Appendix	137
3.6.1	Relevant proofs for rescaled similarities and z-scores evaluated under the various null models	137
3.6.2	Additional results	146
4	Identifying Patterns in the Bank-Sector Credit Network of Spain	163
4.1	Introduction	163
4.2	Data and Methods	166
4.2.1	Data	166
4.2.2	General definitions	166
4.2.3	Methods used for the analysis of binary bipartite networks	167
4.2.4	Methods used for the analysis of weighted bipartite networks	172
4.2.5	Bipartite Configuration Models	173
4.3	Findings	180
4.3.1	Binary bank-sector network	180
4.3.2	Weighted bank-sector network	193
4.4	Conclusions	199
4.5	References	201
4.6	Appendix	204

4.6.1	Motifs	204
4.6.2	Ensemble distributions of various network properties	207
4.6.3	z-scores of ANNSs and C_4^w evaluated under the BiWCM and the BiECM.	211
5	Multilayer Overlaps and Correlations in the Bank-Firm Credit Network of Spain	213
5.1	Introduction	213
5.2	Overlaps and Multiplexity in multilayer networks	216
5.2.1	General definitions, measures for overlaps, and null models	216
5.2.2	Multiplexity	224
5.3	Findings	231
5.3.1	Phenomenology to the bank-firm credit network of Spain	231
5.3.2	Binary analysis	233
5.3.3	Weighted analysis	236
5.3.4	Hubs distribution and hidden variables	240
5.4	Conclusions	243
5.5	References	244
5.6	Appendix	246
6	The Structure of the Cross-Correlation Matrices of Banks' Loan Portfolios in the Bank-Firm Credit Market of Japan	250
6.1	Introduction	250
6.2	Data and Methods	252
6.2.1	Data	252
6.2.2	Methods	253
6.3	Findings	258
6.4	Conclusions	273
6.5	References	274
6.6	Appendix	279
7	An Analysis of Systemic risk in Worldwide Economic Sentiment Indices	284
7.1	Introduction	284
7.2	Data and Methods	286
7.2.1	Data	286
7.2.2	Methods	286

7.3 Findings	291
7.4 Conclusions	302
7.5 References	302
Summary and Outlook	306
Affirmation	310

List of Figures

2.1	Degree-degree dependencies in the directed version	17
2.2	Directed triangles and the corresponding clusterings	17
2.3	Average degrees of nearest neighbors, local assortativity, and local clustering coefficients in the undirected binary e-MID network, in Q1 and Q48	25
2.4	Evolution of the overall assortativity in the undirected binary e-MID network	26
2.5	Evolution of the average of local clustering coefficients in the undirected binary e-MID network	27
2.6	Average degrees of nearest neighbors in the directed binary e-MID network, in Q1	29
2.7	Average degrees of nearest neighbors in the directed binary e-MID network, in Q48	30
2.8	Evolution of the overall assortativity indicators in the directed binary e-MID network	31
2.9	Local assortativity in the directed binary e-MID network, in Q1	32
2.10	Local assortativity in the directed binary e-MID network, in Q48	33
2.11	Local clustering coefficients in the directed binary e-MID network, in Q1 . .	35
2.12	Local clustering coefficients in the directed binary e-MID network, in Q48 . .	36
2.13	Evolution of the averages of local binary clustering coefficients in the directed binary e-MID network	37
2.14	Average degrees of nearest neighbors, local assortativity, and local clustering coefficients in the observed e-MID network and in the UBCM, in Q1 and Q48	38
2.15	Evolution of the average of average degrees of nearest neighbors, the overall assortativity, and the average of local clustering coefficients in the observed e-MID network and in the UBCM	39
2.16	Average degrees of nearest neighbors in the observed e-MID network and in the DBCM, in Q1	41

2.17	Average degrees of nearest neighbors in the observed e-MID network and in the DBCM, in Q48	42
2.18	Local assortativity in the observed e-MID network and in the DBCM, in Q1	43
2.19	Local assortativity in the observed e-MID network and in the DBCM, in Q48	44
2.20	Local clustering coefficients in the observed e-MID network and in the DBCM, in Q1	45
2.21	Local clustering coefficients in the observed e-MID network and in the DBCM, in Q48	46
2.22	Evolution of the averages of average degrees of nearest neighbors in the observed e-MID network and in the DBCM	47
2.23	Evolution of the global assortativity indicators in the observed e-MID network and in the DBCM	48
2.24	Evolution of the averages of clustering coefficients in the observed e-MID network and in the DBCM	49
2.25	Average strengths of nearest neighbors in the undirected weighted e-MID network, in Q1 and Q48	51
2.26	Evolution of global weighted assortativity in the undirected weighted e-MID network	51
2.27	Local weighted clustering coefficients in the undirected weighted e-MID network, in Q1 and Q48	52
2.28	Evolution of the average of local weighted clustering coefficients in the undirected weighted e-MID network	52
2.29	Average strengths of nearest neighbors in the directed weighted e-MID network, in Q1	54
2.30	Average strengths of nearest neighbors in the directed weighted e-MID network, in Q48	55
2.31	Evolution of the directed weighted assortativity indicators in the directed weighted e-MID network	56
2.32	Local weighted clustering coefficients in the directed weighted e-MID network, in Q1	57
2.33	Local weighted clustering coefficients in the directed weighted e-MID network, in Q48	58
2.34	Evolution of the averages of directed local weighted clustering coefficients in the directed weighted e-MID network	59

2.35 Average strengths of nearest neighbors in the observed network and in UWCM, in Q1 and Q48 61

2.36 Local weighted clustering coefficients in the observed network and in the UWCM, in Q1 and Q48 61

2.37 Average strengths of nearest neighbors in the observed network and in the UECM, in Q1 and Q48 62

2.38 Local weighted clustering coefficients in the observed network and in the UECM, in Q1 and Q48 62

2.39 z-scores of average strengths of nearest neighbors in the UWCM and the UECM, in Q1 and Q48 63

2.40 z-scores of local weighted clustering coefficients in the UWCM and the UECM, in Q1 and Q48 63

2.41 Evolution of the average of average strengths of nearest neighbors, the overall weighted assortativity, and the average of local weighted clustering coefficients in the observed e-MID network and in the UWCM 64

2.42 Evolution of the average of average strengths of nearest neighbors, the overall weighted assortativity, and the average of local weighted clustering coefficients in the observed e-MID network and in the UECM 65

2.43 Average strengths of nearest neighbors in the observed network and in the DWCM, in Q1 67

2.44 Average strengths of nearest neighbors in the observed network and in the DWCM, in Q48 68

2.45 Average strengths of nearest neighbors in the observed network and in the DECM, in Q1 69

2.46 Average strengths of nearest neighbors in the observed network and in the DECM, in Q48 70

2.47 z-scores of average strengths of nearest neighbors in the DWCM and the DECM, in Q1 71

2.48 z-scores of average strengths of nearest neighbors in the DWCM and the DECM, in Q48 72

2.49 Local weighted clustering coefficients in the observed e-MID network and in the DWCM, in Q1 73

2.50 Local weighted clustering coefficients in the observed e-MID network and in the DWCM, in Q48 74

2.51	Local weighted clustering coefficients in the observed e-MID network and in the DECM, in Q1	75
2.52	Local weighted clustering coefficients in the observed e-MID network and in the DECM, in Q48	76
2.53	z-scores of local weighted clustering coefficients in the DWCM and the DECM, in Q1	77
2.54	z-scores of local weighted clustering coefficients in the DWCM and the DECM, in Q48	78
2.55	Evolution of the averages of ANNs in the observed e-MID network and in the DWCM	80
2.56	Evolution of the global weighted assortativity indicators in the observed e-MID network and in the DWCM	81
2.57	Evolution of the averages of local weighted clustering coefficients in the observed e-MID network and in the DWCM	82
2.58	Evolution of the averages of average strengths of nearest neighbors in the observed e-MID network and in the DECM	83
2.59	Evolution of the global weighted assortativity indicators in the observed e-MID network and in the DECM	84
2.60	Evolution of the averages of local weighted clustering coefficients in the observed e-MID network and in the DECM	85
2.61	Evolution of z-scores for the average of average strengths of nearest neighbors, the overall weighted assortativity, and the average of local weighted clustering coefficients evaluated under the DWCM and the DECM	87
2.62	Evolution of z-scores for the global weighted assortativity indicators evaluated under the DWCM and the DECM	88
2.63	Evolution of z-scores for the averages of average strengths of nearest neighbors evaluated under the DWCM and the DECM	89
2.64	Evolution of z-scores for the averages of local weighted clustering coefficients evaluated under the DWCM and the DECM	90
2.65	In-coming, out-going degrees to two vertices of an edge in directed networks.	97
3.1	Binary structural similarities in the bank-firm credit network of Spain in 1997	115
3.2	Binary structural similarities in the bank-firm credit network of Spain in 2007	116
3.3	Evolution of the average of structural similarities in the bank-firm credit network of Spain	116

- 3.4 z-score analysis in the bank-firm credit network of Spain under the BCM . . . 117
- 3.5 Binary structural similarities in the e-MID network in the first quarter of 1999 119
- 3.6 Binary structural similarities in the e-MID network in the last quarter of 2010 120
- 3.7 Evolution of the average structural similarities in the e-MID network 120
- 3.8 z-score analysis in the e-MID network under the BCM 121
- 3.9 Weighted structural similarities in the bank-firm credit network of Spain in
1997 122
- 3.10 Weighted structural similarities in the bank-firm credit network of Spain in
2007 123
- 3.11 Evolution of the average of weighted structural similarities in the bank-firm
credit network of Spain 123
- 3.12 z-score analysis in the bank-firm credit network of Spain under the WCM and
the ECM 124
- 3.13 Weighted structural similarities in the e-MID network in the first quarter of
1999 126
- 3.14 Weighted structural similarities in the e-MID network in the last quarter of
2010 127
- 3.15 Evolution of the average of weighted structural similarities for the e-MID
network, compared to expected ones obtained from the WCM and the ECM 127
- 3.16 z-score analysis in the e-MID network under the WCM and ECM 128
- 3.17 Similarities in descending order of hidden variables in the bank-firm credit
network of Spain 130
- 3.18 Similarities in descending order of hidden variables in the e-MID network . . 130
- 3.19 Largest and statistically significant similarities between different groups of
banks in the bank-firm credit market of Spain 132
- 3.20 Binary structural similarities in the bank-firm credit network of Spain, in
1997, compared to the BCM 146
- 3.21 Binary structural similarities in the bank-firm credit network of Spain, in
2007, compared to the BCM 147
- 3.22 Binary structural similarities in the e-MID network, in the first quarter of
1999, compared to the BCM 148
- 3.23 Binary structural similarities in the e-MID network, in the last quarter of
2010, compared to the BCM 149

- 3.24 Weighted structural similarities in the bank-firm credit network of Spain, in 1997, compared to the WCM and the ECM 150
- 3.25 Weighted structural similarities in the bank-firm credit network of Spain, in 2007, compared to the WCM and the ECM 151
- 3.26 Weighted structural similarities in the e-MID network, in the first quarter of 1999, compared to the WCM and the ECM 152
- 3.27 Weighted structural similarities in the e-MID network, in the last quarter of 2010, compared to the WCM and the ECM 153
- 3.28 Comparison between z-scores evaluated under the WCM and ECM in the bank-firm credit network of Spain 154
- 3.29 Comparison between z-scores evaluated under the WCM and ECM in the e-MID network 154
- 3.30 Jaccard-based similarities in the bank-firm credit network of Spain in 1997, in the binary version 155
- 3.31 Jaccard-based similarities in the bank-firm credit network of Spain in 2007, in the binary version 156
- 3.32 Jaccard-based similarities in the e-MID network in the first quarter of 1999, in the binary version 157
- 3.33 Jaccard-based similarities in the e-MID network in the last quarter of 2010, in the binary version 158
- 3.34 Jaccard-based similarities in the bank-firm credit network of Spain in 1997, in the weighted version 159
- 3.35 Jaccard-based similarities in the bank-firm credit network of Spain in 2007, in the weighted version 160
- 3.36 Jaccard-based similarities in the e-MID network in the first quarter of 1999, in the weighted version 161
- 3.37 Jaccard-based similarities in the e-MID network in the last quarter of 2010, in the weighted version 162

- 4.1 Portfolio overlaps in the observed network and under the BiBCM 182
- 4.2 z-scores for the off-diagonal elements of the projection matrices evaluated under the BiBCM 183
- 4.3 The average degrees of nearest neighbors in the observed network and under the BiBCM 185

4.4	Evolution of the Pearson correlation coefficient of degrees in the observed network and under the BiBCM	186
4.5	Evolution of the averages of the average degrees of nearest neighbors in the observed network and under the BiBCM	186
4.6	Evolution of the numbers of motifs in the observed network and under the BiBCM	188
4.7	Local clustering coefficients in the observed network and under the BiBCM .	189
4.8	The adjacency matrix and nestedness structure in the bank-sector network .	190
4.9	Local nestedness in the observed network and under the BiBCM	191
4.10	Evolution of the overall nestedness indicators in the observed network and under the BiBCM	192
4.11	The average strengths of nearest neighbors in the observed network and under the BiWCM	195
4.12	Local weighted clustering coefficients in the observed network and under the BiWCM	196
4.13	The average strengths of nearest neighbors in the observed network and under the BiECM	197
4.14	Local weighted clustering coefficients in the observed network and under the BiECM	198
4.15	Different types of motifs in a bipartite network	204
4.16	Distribution of Pearson correlation coefficient of degrees in the ensemble generated by the BiBCM	208
4.17	Ensemble distributions of the numbers of Λ and V motifs under the BiBCM	209
4.18	Ensemble distributions of the numbers of X , M , and W motifs under the BiBCM	210
4.19	z -scores for the average strengths of nearest neighbors evaluated under the BiWCM and under the BiECM	211
4.20	z -scores for local weighted clustering coefficients evaluated under the BiWCM and under the BiECM	212
5.1	Distributions of banks' degree, firms' degree, and lending relations to sectors	232
5.2	Bank's diversification and sector's diversification	233
5.3	Binary multiplexities between layers of the bank-firm credit network of Spain	235
5.4	z -scores for binary multiplexities under the BRG model and the BCM	236
5.5	Comparison between binary and weighted multiplexities	236

- 5.6 Weighted multiplexities between layers of the bank-firm credit network of Spain, (*cont.*) 238
- 5.6 Weighted multiplexities between layers of the bank-firm credit network of Spain 239
- 5.7 z-scores for weighted multiplexities under the WRG model, the WCM, and the ECM 240
- 5.8 Hubs and hidden variables in the two example layers 242
- 5.9 Common hubs in the two example layers 242
- 6.1 Binary overlaps and portfolio similarities between banks in the three layers, in 2012 259
- 6.2 Temporal evolution of the distributions and fundamental statistics of the elements of the correlation matrices in the three layers 260
- 6.3 Distributions of the eigenvalues of the correlation matrices, compared with RMT, in the three layers, in 1980 and 2012 262
- 6.4 Evolution of the top 5 largest eigenvalues and absorption ratios for correlation matrices in the three layers 263
- 6.5 Inverse Participation Ratios of the eigenvectors of the correlation matrices in the three layers, in 1980 and 2012 265
- 6.6 The eigenvectors in descending order of the first largest and the fourth largest eigenvalues, in the three layers, in 2012 267
- 6.7 Correlations and filtered correlations between banks in lending to firms in the three layers, in 2012 268
- 6.8 Distribution of eigenvalues of the correlation matrices, compared with RMT, in the three layers, in 1980 and 2012, in the binary data 270
- 6.9 Evolution of the top 5 largest eigenvalues and the associated absorption ratios for correlation matrices in the three layers, in binary data 271
- 6.10 Inverse Participation Ratios of the eigenvectors of correlation matrices in the three layers, in 1980 and 2012, in the binary data 272
- 6.11 Distributions of the eigenvector components of four selected eigenvalues of the correlation matrices, in the three layers, in 2012, in the weighted data 279
- 6.12 Evolution of the eigenvector components of the four largest eigenvalues, in the layer 1 281
- 6.13 Evolution of the eigenvector components of the four largest eigenvalues, in the layer 2 282

6.14 Evolution of the eigenvector components of the four largest eigenvalues, in the layer 3 283

7.1 Evolution of the distribution of correlations for BCI data, in the OECD⁺ group 292

7.2 Evolution of the distribution of correlations for ESI data, in the EU⁺ group . 292

7.3 Identifying states of correlation matrix using similarity-based analysis 293

7.4 Evolution of eigenvalues and absorption ratios for BCI data, in the OECD⁺ group 294

7.5 Evolution of eigenvalues and absorption ratios for ESI data, in the EU⁺ group 295

7.6 Distribution of eigenvalues of the correlation matrix and the comparison with RMT 295

7.7 Evolution of the eigenvector components of the largest eigenvalue for BCI data, in the OECD⁺ group 296

7.8 Evolution of the eigenvector components of the largest eigenvalue, for ESI data in the EU⁺ group 297

7.9 Eigenvector components of the largest eigenvalue in different years 298

7.10 Inverse Participation Ratios 299

7.11 Raw and filtered correlation matrices for BCI data, in the OECD⁺ group . . 300

7.12 Raw and filtered correlation matrices for ESI data, in the EU⁺ group 301

Notations and Abbreviations

General Symbols

\approx	is approximated as
\sim	is similar to
$\overset{iid}{\sim}$	is identically and independently distributed according to
\equiv	is identical to
\gg	is much greater/larger than
\Rightarrow	that leads to
$\langle X \rangle$	the expectation of X
$\langle X \rangle_{\text{null model}}$	the expectation under the referenced null model
$\sigma[X]$	the standard deviation of X
$\sigma[X]_{\text{null model}}$	the standard deviation of X under the referenced null model
$\mu[X]_{\text{null model}}$	the rescaled quantity of X under the referenced null model

Abbreviations

ANND	Average degree of the nearest neighbors
ANNS	Average strength of the nearest neighbors
BiBCM	Bipartite Binary Configuration Model
BiCM	Bipartite Configuration Models
BiECM	Bipartite Enhanced Configuration Model
BiWCM	Bipartite Weighted Configuration Model
(U)BCM	(Undirected) Binary Configuration Model
(U)BRG	(Undirected) Binary Random Graph Model
CM	Configuration Models
DBCM	Directed Binary Configuration Model
DECM	Directed Enhanced Configuration Model
DWCM	Directed Weighted Configuration Model
ERG	Exponential Random Graph

(U)ECM	(Undirected) Enhanced Configuration Model
e-MID	Italian Market for Interbank Deposits
RGM	Random Graph Model
RMT	Random Matrix Theory
Sec.	Section
(U)WCM	(Undirected) Weighted Configuration Model
(U)WRG	(Undirected) Weighted Random Graph Model

Preface

This thesis is composed by six essays on the empirical applications of network theory and Random Matrix Theory on economic and financial systems. It is organized in a format with following chapters:

- Chapter 1: Introduction
- Chapter 2: Does the Heterogeneity in the Local Constraints Predominantly Determine Structural Correlations in the Italian Overnight Money Market? (*with Boyan Yanovski and Thomas Lux*).
- Chapter 3: An Approach to Identify Patterns in Structural Similarities in Financial Networks.
- Chapter 4: Identifying Patterns in the Bank-Sector Credit Network of Spain (*with Thomas Lux*).
- Chapter 5: Multilayer Overlaps and Correlations in the Bank-Firm Credit Market of Spain (*with Thomas Lux*).
- Chapter 6: The Structure of the Cross-Correlation Matrices of Banks' Loan Portfolios in the Bank-Firm Credit Market of Japan.
- Chapter 7: An Analysis of Systemic Risk in Worldwide Economic Sentiment Indices (*with Boyan Yanovski and Thomas Lux*).
- Chapter 8: Summary and Outlook.

Chapter 1

Introduction

This thesis contributes to the existing literature on the empirical applications of network theory and Random Matrix Theory to economic and financial complex systems.

It is based on six essays which can be structured in two parts. The first part consists of four essays (from chapter 2 to chapter 5), which are devoted to various applications of network theory to large data sets of credit relationships in the interbank market and between banks and other sectors of the economy. They study various properties in different structures of different financial networks. More specifically, the second and the third chapters respectively analyze structural correlations and structural similarities in one-mode networks. The fourth chapter deals with topological and structural properties in bipartite networks, and the fifth chapter is devoted to overlaps and correlations between layers in the multilayer structure of networks. Two essays in the second part (respectively in chapter 6 and chapter 7), based on the methods of Random Matrix Theory, analyze the structure of the cross-correlation matrices of banks' loan portfolios and the structure of the cross-correlation matrix of worldwide economic sentiment indices. The six essays from chapter 2 to chapter 7 in my thesis can be summarized as follows.

First essay: Does the Heterogeneity in the Local Constraints Predominantly Determine Structural Correlations in the Italian Overnight Money Market?

In this essay, we comprehensively study the second and the third order structural correlations in all network versions of the Italian Overnight Money Market (e-MID) over the period 1999-2010. First, analyzing the observed network, we show that the structural correlations vary across different versions of the network, i.e. undirected vs directed and unweighted vs weighted versions. Among other results, we find that, in the undirected case, both the binary and the weighted versions exhibit disassortative mixing. When the directions of the

edges are taken into account, two types of nodes (giving and receiving) give rise to four categories of mixing, i.e. in-in, in-out, out-in, and out-out. The overall assortativity in the out-in category is closest to the mixing observed in the undirected version of the network, meaning that nodes having a lot of outgoing edges tend to be connected to nodes with few incoming edges. Further analysis reveals that, overall, when the disassortativity in the network is strong enough, the global degree dependencies are mainly contributed by the hubs. We observe that the third order correlation structure among banks changes abruptly at the time of the outburst of the recent financial crisis and, on average, the weighted clustering coefficients are abnormally high during the period from the adoption of the Euro up until the time right before the crisis. We also detect indications of elevated systemic risk in the network, evidenced by the prevalence of the “middleman” and “inward” type of clustering in the network.

As a second step, we also examine the hypothesis that the structural correlations of the observed network stem from latent information embedded in the local constraints like the observed degree sequence(s) in the binary case and/or the observed strength sequence(s) in the weighted case. To do that, we employ a comprehensive family of configuration models that capture the intrinsic heterogeneity in the observed degree and/or strength distribution, and then compare the structural correlations of the observed network with those predicted by these models.

In the binary case, we find that, in both the undirected and directed versions of the observed network, the so-called Binary Configuration Models based on the observed degree sequence(s) can generally replicate the main features of the observed higher order structural correlations. However, not all second and third order correlations can be predicted by these models, showing that the structural correlations of the observed network in its binary version do exhibit some non-random patterns within the context of this randomization procedure.

In the weighted case, our findings suggest that, generally, the so-called Weighted Configuration Models (utilizing only information about the strength sequence(s)) cannot explain the weighted structural correlations. When each Weighted Configuration Model is allowed to employ also information about the degree(s) of the nodes (which is then referred to as a so-called Enhanced Configuration Model), its predictive power greatly improves. As in the binary version of the network, however, the information embedded in the local constraints is not enough to allow the model to account for all the patterns in the observed higher order structural correlations. In particular, one of the main features not explained by the sequences of observed degrees and strengths is the high level of clustering in the years pre-

ceding the crisis, i.e. the huge increase in various indirect exposures generated via more intensive interbank credit links.

Interestingly, we observe that, in certain structural properties the distance between the predictions of the Weighted Configuration Models and the Enhanced Configuration Models continuously increases from the adoption of the Euro up until the financial crisis in 2007 and then sharply decreases after that. This result can be interpreted as an indication of structural changes in the network associated with these two critical events. It also suggests that the importance of particular basic features of a network (like its degree sequence or its strength sequence) for the emergence of higher order correlation structures can vary over time.

This essay contributes to the existing literature on structural correlations in financial networks by comprehensively assessing the role of various local constraints in all versions of the Italian e-MID network. The essay has been written by myself together with Boyan Yanovski and Prof. Dr. Thomas Lux, and the three authors contributed equally to the essay.

Second essay: An Approach to Identify Patterns in Structural Similarities in Financial Networks

In the second essay, I investigate structural similarities between nodes based on the information of nodes' neighborhoods in one-mode networks. I introduce a method to filter the effects of different global or local constraints on such similarities. I then apply the method to analyze similarities between bank's loan portfolios in two real world financial networks, i.e. the Italian e-MID network and the bank-bank projection network obtained from the bank-firm credit network of Spain, for both the binary as well as weighted versions.

Among other results, my study finds that, first, in general, some banks' portfolios are very similar while some others exhibit a low level of similarity or no similarity at all. Second, the findings suggest that the random graph models specifying only global constraints (i.e. the average of degrees in the binary version or the average of strengths in the weighted version) cannot explain the main features of the observed similarities. Third, comparing the observed networks with different configuration models maintaining local constraints (i.e. the observed degree sequence and/or strength sequence), in both networks, almost all similarities between banks' portfolios in the binary version are removed when the effects of the observed degrees are filtered out. In the weighted version, the two networks show different behaviors. More specifically, in the e-MID network, the knowledge of the observed degree sequence in addition to the observed strength sequence results in a significant improvement in filtering the structural similarities. In contrast, in the bank-bank projection network, for many pairs of

banks, there is no significant difference between filtering the effects of the observed strength sequence alone and filtering the effects of the observed degree as well as strength sequences. Fourth, in both networks, I do observe the presence of patterns in the structural similarities between some banks that can not simply be explained by the global or local constraints. In the case of the bank-firm credit network of Spain, selecting the subset of largest and statistically significant components in the similarity matrices as examples, I find that this subset mainly consists of saving banks as well as commercial banks.

In a nut shell, the second essay presents another important application of the configuration models to the identification of non-random patterns in structural overlaps and similarities in undirected networks. It also introduces a new method used for filtering out the effects of fundamental constraints on portfolio overlaps and similarities in financial networks. This method can straightforwardly be extended to directed networks and bipartite networks (e.g. investor-asset networks, bank-firm credit networks, and so forth). This essay has been written by myself.

Third essay: Identifying Patterns in the Bank-Sector Credit Network of Spain

We now move on to an essay in the bipartite structure of networks. In this essay, we study the topological and structural properties of the bank-sector credit network of Spain over the period from 1997 to 2007. In particular, we start by analyzing assortativity, different types of motifs, and nestedness phenomenon both in the bipartite structure of the binary version of the network and in its weighted version. In order to assess the statistical significance of the network properties, we employ the family of the so-called Bipartite Configuration Models (BiCM) imposing initial constraints like the degree sequence and/or strength sequence of the observed bipartite network. Among other results, in the binary version, we find that the so called Bipartite Binary Configuration Model maintaining the observed degrees can replicate the main features of many properties of the observed network. In addition, the one-mode projection matrices indicating lending portfolio overlaps between banks and borrowing portfolio overlaps between sectors can mostly be predicted by information embedded in the degree sequence of the original bipartite structure. In the weighted version, we observe that the so-called Bipartite Enhanced Configuration Model, where both the degree as well as the strength sequences are preserved on average, outperforms the so-called Bipartite Weighted Configuration Model maintaining only the strength sequence in replicating the topological features of the network. Moreover, comparing the observed network to all the referenced null models, we still find a number of features of higher-order topological properties that can not simply be explained by information embedded in the observed degree and/or strength

sequence, in particular, an “excessive” weighted clustering of the banks and sectors with the highest degrees.

This essay not only contributes to the literature on the identification of higher-order patterns in real world bipartite networks in general but also provide some important implications for the reconstruction of the real credit networks from limited information. The essay has been written by myself together with Prof. Dr. Thomas Lux, and the two authors contributed equally to it.

Fourth essay: Multilayer Overlaps and Correlations in the Bank-Firm Credit Market of Spain

In this essay, under a multilayer network perspective, we investigate the structural dependencies in the bank-firm credit market of Spain. In particular, the original bipartite network is decomposed into different layers representing different industrial sectors. We then study the correlations between layers based on normalized measures of overlaps of links and weights of banks between layers. To assess the statistical significance of such correlations, we compare the observed values with the expected ones obtained from random graph models specifying only global constraints, i.e. the total degree or the total strength in single layers, and from configuration models capturing the intrinsic heterogeneity in the local constraints like the observed degree sequence and/or strength sequence in single layers.

We find that, first, the raw dependencies between layers of the observed network are highly heterogeneous. Second, when evaluated against the null models, on the one hand, the rescaled correlations after filtering out the effects of the global constraints typically display no significant difference to the observed correlations. In addition, in the binary version, almost all correlations are still present after subtracting the effects of the observed degree sequences in all layers. On the other hand, the observed correlations are partially explained by the local constraints maintained in the weighted configuration models. All in all, comparing the observed network with all referenced null models, we find that the multilayer credit network under scrutiny has a significant, non-random structure of correlations that cannot be explained by more primitive network properties alone. In the binary case, such a non-random structure is, for instance, typically observed in the pairs of layers that have high levels of overlaps and correlations. In contrast, in the weighted case, patterns are found in different pairs of layers that have various levels of overlaps and correlations.

This essay contributes to the existing literature on overlaps and correlations between layers in multiplex networks by assessing the role of various constraints in single layers. The essay has been written by myself together with Prof. Dr. Thomas Lux, and the two

authors contributed equally to it.

Fifth essay: The Structure of the Cross-Correlation Matrices of Banks' Loan Portfolios in the Bank-Firm Credit Market of Japan

In this essay, I explore the structure of the cross-correlation matrices of banks' loan portfolios in different lending layers (based on the maturity of loans) in the bank-firm credit market of Japan during the period from 1980 to 2012.

Employing the methods of Random Matrix Theory (RMT), my study aims to detect non-random patterns in the empirical cross-correlations as well as to identify different states of such correlations over time. Among other results, I find that on the one hand a majority of correlations between banks in lending relations to non-financial firms are contributed by noise. On the other hand, a group of largest eigenvalues of the correlation matrices always deviate from the spectrum explained by RMT, indicating the presence of patterns in such correlations. In addition, during the bubble period in Japan, banks' loan portfolios tend to be more correlated, implying a higher systemic risk in the Japanese bank-firm credit market. Furthermore, the eigenvectors corresponding to the top largest eigenvalues are not very localized, indicating that latent factors extracted from these largest eigenvalues have a wide effect on banks, although it should be emphasized that banks do not contribute homogeneously to these eigenvectors. Moreover, in general, the binary and weighted analyses provide a couple of common and independent results, showing that the binary version of bank-firm credit relationships also contain genuine information about the structure of the cross-correlation matrices.

This essay contributes to a deeper understanding of the structure of the cross-correlation matrices of banks' loan portfolios. The essay has been written by myself.

Sixth essay: An Analysis of Systemic Risk in Worldwide Economic Sentiment Indices

In this essay, we investigate the temporal dynamics of correlations between sentiment indices worldwide. Employing the tools of Random Matrix Theory and Principal Component Analysis (PCA), our study aims to extract latent information embedded in the interactions between economic sentiment indices around the world. We find that: (i) The dynamics of the sentiment indices across countries can be well explained by the evolution of a single factor (the "market mode"); (ii) during most periods, some groups of countries exhibit sentiment dynamics less associated with (or divergent from) the market mode, while (iii) during the financial crisis, no country or group of countries has been able to escape the

market mode, which accounts for almost all movements in the indices. We argue that strong “global” information signals, like the collapse of the US housing market in 2007, can lead to a homogenization of the expectation structure around the world, as such information can provide a coordination signal for a global phase of low confidence.

This essay shows that RMT and principal component analysis of the ensemble of worldwide or regional sentiment data can reveal important information on the correlations between economic sentiment indices in different countries. The tools and results presented in this paper should provide relevant input for business cycle forecasts and the analysis of international comovements of macroeconomic activity. The essay has been written by myself together with Boyan Yanovski and Prof. Dr. Thomas Lux, and the three authors contributed equally to the essay.

Chapter 2

Does the Heterogeneity in the Local Constraints Predominantly Determine Structural Correlations in the Italian Overnight Money Market?

Coauthored by: Boyan Yanovski and Thomas Lux.

Keywords: Interbank Network; Structural Correlations; Mixing Natures; Clustering Coefficients; Configuration Models.

2.1 Introduction

Understanding the topological structure of complex systems is crucial in many areas, e.g. in ecology, physics, neuroscience, epidemiology, economics, and finance. Statistics pertaining to properties related to single nodes, linked node pairs and linked node triplets are often referred to as structural correlations of the first, second and third order respectively. The study of these structural correlations is one of the most common approaches for examining the properties of a network. The degree and strength sequences are examples of first order structural correlations. Statistics pertaining to properties related to linked node pairs reveal information about the type of mixing (assortative vs disassortative) that takes place in the network, while those related to linked node triplets are indicative of the clustering behavior.

In terms of second order correlations, a network would exhibit assortative mixing if its nodes are predominantly connected to other nodes having similar degrees or strengths. In

contrast, disassortative mixing occurs when the connected nodes are dissimilar (see, for example, Newman, 2002; Newman, 2003a). This concept can be extended to directed networks yielding four mixing categories, i.e. in-in, in-out, out-in, and out-out mixing as illustrated in Figure (2.1) (see, for example, Foster et al., 2010; Piraveenan et al., 2012; van der Hoorn and Litvak, 2015). It should be emphasized that, in many real world networks, the mixing behavior of the directed version can differ a lot from the one observed in the undirected version. Furthermore, the same directed network can have assortative and disassortative aspects related to the mixing categories mentioned above (see, for example, Foster et al., 2010).

At the level of a single node, in the binary case, second order structural correlations can be expressed in terms of a relationship describing the average degree of the nearest neighbors (ANND) of a node as a function of that node’s own degree. If the ANND is an increasing function of degree, this can be considered evidence in favor of assortative mixing. In contrast, a decreasing function would signal disassortativity.

For the whole network, the Pearson correlation coefficient between the degrees of pairs of linked nodes is often used to assess whether a network displays disassortative or assortative mixing (Newman, 2002; Newman, 2003a). This indicator is nothing else but a function of node degree and can also be expressed in terms of the measures ANND collected for the whole network (see the Appendix for further details).

In addition, we can decompose the overall assortativity coefficient into the contributions of each node, i.e. we can measure the local assortativity associated with each node. Such a decomposition can reveal which nodes contribute to the overall observed mixing nature of the network and which are associated with the opposite type of mixing (see, for example, Piraveenan et al., 2012). For instance, a globally assortative network may be locally disassortative and vice versa. It is worth noting that two networks with the same degree distribution and the same global level of assortativity may display different patterns of local assortativity.

The analysis of the second order structural correlations in the binary case can be straightforwardly extended to weighted networks by employing a measure that takes the average strength of the nearest neighbors (ANNS) of a node or by computing the Pearson correlation coefficient between the strengths of pairs of linked nodes.

As is common in the literature, we use clustering coefficients as measures of the third order structural correlations in the network. A clustering coefficient measures the tendency of two neighbors of a particular node to also be connected to each other (e.g. Newman,

2003b). If we define a node triplet as three nodes connected by at least 2 edges, then, considering a network as a whole, the transitivity ratio (T) is equal to the number of triplets in which all three nodes are directly connected (forming a triangle) as a fraction of all node triplets (e.g. Newman, 2003b). An alternative measure is proposed by Watts and Strogatz (1998), which can capture the observed local clustering. The average of these local clustering coefficients can be used as an alternative measure of clustering for the whole network. The difference between the transitivity ratio and the average clustering coefficient is that, while in the former we calculate the ratio of the means, in the latter we take the mean of the ratios. In addition, for the directed version of a network, it is useful to differentiate between different relationship types depending on the direction of the edges in a triangle, i.e. inward, outward, cyclic, and middleman relationships, since as shown in Figure (2.2), the different relationships have different implications in terms of the risk exposure the individual banks are facing and in terms of systemic risk (see, for example, Fazio, 2007; Tabak et al., 2014). In weighted networks, the weighted clustering coefficients can be formulated in several ways, depending on how we take into account the roles of the strengths and weights of the nodes in each triangle (see, for example, Barrat et al., 2004; Onnela et al., 2005; Zhang and Horvath, 2005; Holme et al., 2007¹).

To assess whether the observed higher order structural correlations in a network are typical of a network with the observed lower order structural correlations, we can employ a randomization procedure based on the observed lower order patterns in the attempt to arrive at a suitable null model to test against for non-random patterns. Such null models create a whole ensemble of networks out of a subset of the information necessary to completely define the observed network. This is why this technique can also be used for filling in unavailable information. The most basic null models are the random graph models (RGM), which specify only global constraints such as the node degree average in the binary case or the node strength average in the weighted case. Since in these models, all nodes are treated homogeneously, there is no difference between the expected topological properties across nodes, which does not happen often in real world networks. In order to capture the intrinsic heterogeneity in the capacity of the individual nodes, a popular approach is to generate the microcanonical ensemble of networks having exactly the same degree sequence (or the same strength sequence in weighted networks) as the one in the observed network (see, for example, Maslov and Sneppen, 2002; Maslov et al., 2004; Zlatic et al., 2011). However, this “hard”

¹We refer the readers to Saramäki et al. (2007) for a comparison between different methods for calculating the local weighted clustering coefficients.

approach suffers from various limitations ². Based on the maximum-entropy and maximum-likelihood methods, recent advances in the specification of configuration models propose a “soft” approach that enforces the constraints on average over an ensemble of randomized networks (e.g. Garlaschelli and Loffredo, 2008; Squartini and Garlaschelli, 2011; Squartini et al., 2011a; Squartini et al., 2011b; Mastrandrea et al., 2014; Squartini et al., 2015). This approach allows us to sample network ensembles more efficiently and in an unbiased manner (Squartini et al., 2015).

In this paper we analyze the structural correlations in a particular financial system, i.e. the Italian electronic market for interbank deposits (e-MID). While some of the network properties of the e-MID market have been previously studied (see, for example, De Masi et al., 2006; Fricke, 2012; Fricke et al., 2013; Finger et al., 2013; Fricke and Lux, 2015a; Fricke and Lux, 2015b; Squartini et al., 2015; Cimini et al., 2015a), what is novel in our paper is that: (i) we provide a more comprehensive analysis of the structural correlations in all versions of the network, and employ both local as well as global measures for analyzing such patterns; (ii) we employ configuration models to investigate whether the intrinsic node heterogeneity represented by the degree sequence (in the binary network) and/or strength sequence (in the weighted network) can explain higher order structural correlations observed in the system; (iii) we utilize the so called Directed Enhanced Configuration Model as a null model for the directed weighted version of the network, which makes use of the available information about the direction of the edges in the network.

We use quarterly data for the e-MID network over the period 1999-2010 and restrict our analysis to the Italian banks participating in this market, because foreign banks are not frequently active in the market. Particularly, from the onset of the financial crisis in 2008 onward, non-Italian banks have basically withdrawn from this electronic market (e.g. see Fricke et al., 2013) ³.

The remainder of this paper is structured as follows. In Sec. 2.2 we provide a general framework for analyzing the structural correlations in different versions of the observed network as well as the algorithm for generating an ensemble of randomized networks from given constraints. In Sec. 2.3, we analyze the structural correlations in the undirected and directed binary versions of the e-MID network, and then compare the results to those

²See Squartini and Garlaschelli (2011) or Squartini et al. (2015) for a discussion of this “hard” approach and its limitations.

³The transactions between banks are aggregated into quarterly data, since at the higher frequencies the matrix of the trades between banks is really sparse. For a more detailed description of the e-MID dataset, we refer the readers to the studies of Fricke et al. (2013) and Finger et al. (2013), or to the e-MID website <http://www.e-mid.it/>.

obtained from the associated null models. In Sec. 2.4, we provide a similar analysis of the undirected as well as directed weighted versions of the network. Sec. 2.5 contains a discussion of the results as well as directions for future research. At the end of this paper, the Appendix provides additional details concerning the measures of structural correlations.

2.2 Structural correlations in complex networks

2.2.1 For undirected networks

2.2.1.1 General notation

In the undirected version, suppose we have a network (of size n) characterized by a symmetric adjacency matrix $A = \{a_{ij}\}_{n \times n}$ and a symmetric weighted matrix $W = \{w_{ij}\}_{n \times n}$ ($a_{ii} = w_{ii} = 0$). The degree and strength sequences for each node i are respectively defined as

$$k_i^{un} = \sum_{j=1}^n a_{ij}, \quad (2.1)$$

and

$$s_i^{un} = \sum_{j=1}^n w_{ij}. \quad (2.2)$$

The total degree and total strength over all nodes in the network are given by

$$m = \frac{1}{2} \sum_{i=1}^n k_i^{un}, \quad (2.3)$$

and

$$w_{tol} = \frac{1}{2} \sum_{i=1}^n s_i^{un}. \quad (2.4)$$

2.2.1.2 Structural correlations in undirected networks

Assortativity Analysis

Regarding assortativity, we use two measures, i.e. the average degree (strength in the weighted case) of the nearest neighbors as well as the Pearson correlation coefficient (hereafter: Pearson coefficient) between degrees (strengths in the weighted case).

The average degree and strength of the nearest neighbors

The average degree of the nearest neighbors (ANND) of node i in the binary version of

a network is given by

$$k_{nn,i}^{un} = \frac{\sum_{j=1}^n a_{ij} k_j^{un}}{k_i^{un}}. \quad (2.5)$$

For the weighted version, the average strength of the nearest neighbors (ANNS) of node i is defined as

$$s_{nn,i}^{un} = \frac{\sum_{j=1}^n a_{ij} s_j^{un}}{k_i^{un}}. \quad (2.6)$$

Treating k_{nn}^{un} as a function of k^{un} , an overall positive (negative) correlation between k_{nn}^{un} and k^{un} suggests assortative (disassortative) mixing in the binary version of the network. In the weighted case, a positive (negative) correlation between s_{nn}^{un} and s^{un} evidences assortative (disassortative) mixing.

We can also compute the averages of ANND and ANNS over the whole network respectively as

$$\bar{k}_{nn}^{un} = \frac{1}{n} \sum_{i=1}^n k_{nn,i}^{un} \quad (2.7)$$

and

$$\bar{s}_{nn}^{un} = \frac{1}{n} \sum_{i=1}^n s_{nn,i}^{un}. \quad (2.8)$$

Pearson correlation coefficient of the node degrees and of the node strengths

The second measure of mixing computes the Pearson's correlation between two degree sequences (see the Appendix for further details). Practically, the main idea to measure such a correlation is that, from the adjacency matrix, first, we obtain a list of m edges, that is the list of pairs of nodes (i_e, j_e) where $a_{i_e j_e} = 1$, (for $e = 1, 2, \dots, m$, $1 \leq i_e, j_e \leq n$). Next, for each e , we get two degrees $k_{i_e}^{un}$, $k_{j_e}^{un}$, and two strengths $s_{i_e}^{un}$, $s_{j_e}^{un}$ associated with the pair of nodes (i_e, j_e) . The correlation coefficient of the degrees (r_{bin}^{un}) is equal to the Pearson correlation coefficient between the degrees at either ends of an edge (e.g. Newman, 2002; Newman, 2003a). Similarly, we can define the correlation coefficient of the strengths (r_w^{un}) as the Pearson correlation coefficient between the strengths at either ends of an edge. In the binary case, if r_{bin}^{un} is negative, it signals the presence of disassortativity, while a positive value implies the opposite. The same interpretation holds for r_w^{un} in the weighted case, but r_{bin}^{un} and r_w^{un} are not necessary equal.

Clustering coefficients

According to Watts and Strogatz (1998), the undirected binary clustering associated with node i is defined as

$$C_{bin,i}^{un} = \frac{\sum_{j \neq i} \sum_{k \neq i,j} a_{ij} a_{jk} a_{jk}}{\sum_{j \neq i} \sum_{k \neq i,j} a_{ij} a_{ik}}. \quad (2.9)$$

Following Onnela et al. (2005), we obtain the local weighted clustering associated with node i in undirected version of the network as

$$C_{w,i}^{un} = \frac{\sum_{j \neq i} \sum_{k \neq i,j} w_{ij}^{\frac{1}{3}} w_{jk}^{\frac{1}{3}} w_{jk}^{\frac{1}{3}}}{\sum_{j \neq i} \sum_{k \neq i,j} a_{ij} a_{ik}}. \quad (2.10)$$

Note that $C_{w,i}^{un}$ in Eq. (2.10) is invariant to weight permutation for each triangle and it takes into account the weights of all associated edges. In addition, it is easy to show that if $A = W$, we will have $C_{w,i}^{un} = c_{bin,i}^{un}$.

To analyze the evolution of the third order correlations over time, we define the average of $\{C_{bin,i}^{un}\}_{i=1}^n$ as

$$\bar{C}_{bin}^{un} = \frac{1}{n} \sum_{i=1}^n C_{bin,i}^{un}, \quad (2.11)$$

and the average of $\{C_{w,i}^{un}\}_{i=1}^n$ as

$$\bar{C}_w^{un} = \frac{1}{n} \sum_{i=1}^n C_{w,i}^{un}. \quad (2.12)$$

2.2.2 For directed networks

2.2.2.1 General definitions

In a directed network, the two matrices A and W are not symmetric (i.e. $A \neq A^T$ and $W \neq W^T$). We distinguish between in-degree and out-degree for every node i as

$$k_i^{in} = \sum_{j=1}^n a_{ji}, \quad (2.13)$$

and

$$k_i^{out} = \sum_{j=1}^n a_{ij}. \quad (2.14)$$

Similarly, in-strength and out-strength for every node i are given by

$$s_i^{in} = \sum_{j=1}^n w_{ji}, \quad (2.15)$$

and

$$s_i^{out} = \sum_{j=1}^n w_{ij}. \quad (2.16)$$

The total in-degree and total out-degree over all nodes in the network are equal to

$$M = \sum_{i \neq j} a_{ij} = \sum_{j \neq i} a_{ji}. \quad (2.17)$$

The total in-strength and total out-strength over all nodes in the network are defined as

$$w_{tol} = \sum_{i \neq j} w_{ij} = \sum_{j \neq i} w_{ji}. \quad (2.18)$$

2.2.2.2 Structural correlations in directed networks

Assortativity Analysis

The average degree and strength of the nearest neighbors in directed networks

Taking the directions of edges into account (as in Figure (2.1)), two types of nodes (giving and receiving) give rise to four types of relationships and four versions of ANND for each node i :

$$k_{nn,i}^{in-in} = \frac{\sum_{j=1}^n a_{ji} k_j^{in}}{k_i^{in}}, \quad (2.19)$$

$$k_{nn,i}^{in-out} = \frac{\sum_{j=1}^n a_{ji} k_j^{out}}{k_i^{in}}, \quad (2.20)$$

$$k_{nn,i}^{out-in} = \frac{\sum_{j=1}^n a_{ij} k_j^{in}}{k_i^{out}}, \quad (2.21)$$

$$k_{nn,i}^{out-out} = \frac{\sum_{j=1}^n a_{ij} k_j^{out}}{k_i^{out}}. \quad (2.22)$$

Similarly we define different versions of ANNS for each node i :

$$s_{nn,i}^{in-in} = \frac{\sum_{j=1}^n a_{ji} s_j^{in}}{k_i^{in}}, \quad (2.23)$$

$$s_{nn,i}^{in-out} = \frac{\sum_{j=1}^n a_{ji} s_j^{out}}{k_i^{in}}, \quad (2.24)$$

$$s_{nn,i}^{out-in} = \frac{\sum_{j=1}^n a_{ij} s_j^{in}}{k_i^{out}}, \quad (2.25)$$

$$s_{nn,i}^{out-out} = \frac{\sum_{j=1}^n a_{ij} s_j^{out}}{k_i^{out}}. \quad (2.26)$$

In each version, the interpretation of the relationship between the ANND and node degree and between the ANNS and node strength is similar to the one for the measures discussed

in the undirected case. That is, a negative (positive) relationship signals disassortativity (assortativity) in the respective class of relationships.

The averages of the different versions of ANND are given by

$$\bar{k}_{nn}^{in-in} = \frac{1}{n} \sum_{i=1}^n k_{nn,i}^{in-in}, \quad (2.27)$$

$$\bar{k}_{nn}^{in-out} = \frac{1}{n} \sum_{i=1}^n k_{nn,i}^{in-out}, \quad (2.28)$$

$$\bar{k}_{nn}^{out-in} = \frac{1}{n} \sum_{i=1}^n k_{nn,i}^{out-in}, \quad (2.29)$$

$$\bar{k}_{nn}^{out-out} = \frac{1}{n} \sum_{i=1}^n k_{nn,i}^{out-out}. \quad (2.30)$$

For a directed and weighted network, the averages of the different versions of ANNS are defined as

$$\bar{s}_{nn}^{in-in} = \frac{1}{n} \sum_{i=1}^n s_{nn,i}^{in-in}, \quad (2.31)$$

$$\bar{s}_{nn}^{in-out} = \frac{1}{n} \sum_{i=1}^n s_{nn,i}^{in-out}, \quad (2.32)$$

$$\bar{s}_{nn}^{out-in} = \frac{1}{n} \sum_{i=1}^n s_{nn,i}^{out-in}, \quad (2.33)$$

$$\bar{s}_{nn}^{out-out} = \frac{1}{n} \sum_{i=1}^n s_{nn,i}^{out-out}. \quad (2.34)$$

Directed Pearson correlation coefficient of the node degrees and of the node strengths

Similarly, the four possible combinations between giving and receiving nodes are associated with four global assortativity coefficients, i.e. r_{bin}^{in-in} , r_{bin}^{in-out} , r_{bin}^{out-in} , and $r_{bin}^{out-out}$ (see the Appendix for further details). Their weighted counterparts are r_w^{in-in} , r_w^{in-out} , r_w^{out-in} , and $r_w^{out-out}$. The algorithm for calculating these binary (weighted) coefficients is still similar to the one used for r_{bin}^{un} (or r_w^{un} for the weighted case), except for the requirement that the directions of edges (see Figure (2.1)) must be taken into account.

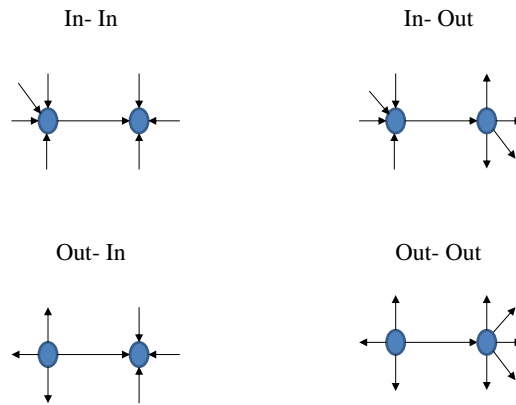


Figure 2.1: Degree-degree dependencies in the directed version.

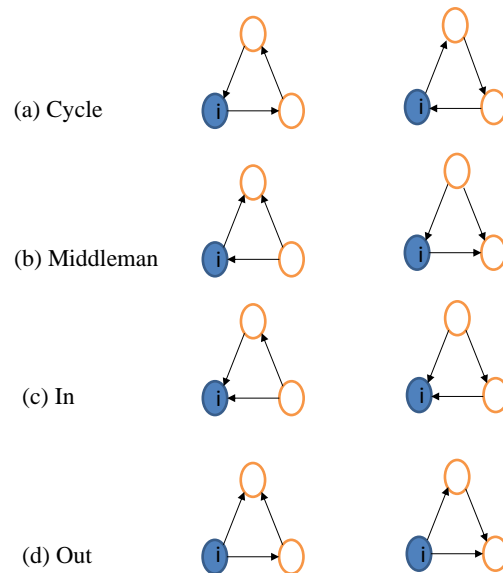


Figure 2.2: Directed triangles and the corresponding (binary) clusterings associated with a node i . (a) Cycle clustering, (b) Middleman clustering, (c) In clustering, (d) Out clustering.

Clustering coefficients

Directed binary clustering coefficients

As shown in in Figure (2.2), for each node i , we define four local binary clustering coefficients in a directed network associated with the respective type of relationship:

$$C_{bin,i}^{in} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} a_{jk} a_{ji} a_{ki}}{(\sum_{j \neq i} a_{ji})^2 - (\sum_{j \neq i} a_{ji}^2)}, \quad (2.35)$$

$$C_{bin,i}^{out} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} a_{ik} a_{ij} a_{jk}}{(\sum_{j \neq i} a_{ij})^2 - (\sum_{j \neq i} a_{ij}^2)}, \quad (2.36)$$

$$C_{bin,i}^{cyc} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} a_{ij} a_{jk} a_{ki}}{(\sum_{j \neq i} a_{ij} \sum_{j \neq i} a_{ji}) - (\sum_{j \neq i} a_{ij} a_{ji})}, \quad (2.37)$$

$$C_{bin,i}^{mid} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} a_{ik} a_{jk} a_{ji}}{(\sum_{j \neq i} a_{ij} \sum_{j \neq i} a_{ji}) - (\sum_{j \neq i} a_{ij} a_{ji})}. \quad (2.38)$$

Note that, in the binary case we have $a_{ij}^2 = a_{ij} (\forall i, j)$, therefore from Eqs. (2.35), (2.36), (2.37), (2.38) we have

$$C_{bin,i}^{in} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} a_{jk} a_{ji} a_{ki}}{k_i^{in} (k_i^{in} - 1)}, \quad (2.39)$$

$$C_{bin,i}^{out} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} a_{ik} a_{ij} a_{jk}}{k_i^{out} (k_i^{out} - 1)}, \quad (2.40)$$

$$C_{bin,i}^{cyc} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} a_{ij} a_{jk} a_{ki}}{k_i^{in} k_i^{out} - k_i^{\leftrightarrow}}, \quad (2.41)$$

$$C_{bin,i}^{mid} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} a_{ik} a_{jk} a_{ji}}{k_i^{in} k_i^{out} - k_i^{\leftrightarrow}}, \quad (2.42)$$

where k_i^{\leftrightarrow} is the number of nodes j in the neighborhood of the node i such that $a_{ij} = a_{ji} = 1$.

The averages over the local binary clustering coefficients are defined as

$$\bar{C}_{bin}^{in} = \frac{1}{n} \sum_{i=1}^n C_{bin,i}^{in}, \quad (2.43)$$

$$\bar{C}_{bin}^{out} = \frac{1}{n} \sum_{i=1}^n C_{bin,i}^{out}. \quad (2.44)$$

$$\bar{C}_{bin}^{cyc} = \frac{1}{n} \sum_{i=1}^n C_{bin,i}^{cyc}, \quad (2.45)$$

$$\bar{C}_{bin}^{mid} = \frac{1}{n} \sum_{i=1}^n C_{bin,i}^{mid}, \quad (2.46)$$

Directed weighted clustering coefficients

In a weighted directed network, among various measures, following Onnela et al. (2005), we define the local weighted clustering coefficients for each node i as

$$C_{w,i}^{in} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} w_{jk}^{\frac{1}{3}} w_{ji}^{\frac{1}{3}} w_{ki}^{\frac{1}{3}}}{(\sum_{j \neq i} a_{ji})^2 - (\sum_{j \neq i} a_{ji}^2)}, \quad (2.47)$$

$$C_{w,i}^{out} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} w_{ik}^{\frac{1}{3}} w_{ij}^{\frac{1}{3}} w_{jk}^{\frac{1}{3}}}{(\sum_{j \neq i} a_{ij})^2 - (\sum_{j \neq i} a_{ij}^2)}, \quad (2.48)$$

$$C_{w,i}^{cyc} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} w_{ij}^{\frac{1}{3}} w_{jk}^{\frac{1}{3}} w_{ki}^{\frac{1}{3}}}{(\sum_{j \neq i} a_{ij} \sum_{j \neq i} a_{ji}) - (\sum_{j \neq i} a_{ij} a_{ji})}, \quad (2.49)$$

$$C_{w,i}^{mid} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} w_{ik}^{\frac{1}{3}} w_{jk}^{\frac{1}{3}} w_{ji}^{\frac{1}{3}}}{(\sum_{j \neq i} a_{ij} \sum_{j \neq i} a_{ji}) - (\sum_{j \neq i} a_{ij} a_{ji})}. \quad (2.50)$$

We can see that in all local weighted clustering coefficients, the denominators are identical to those in the binary counterparts. Obviously, it is easy to show that for $A = W$, the binary clustering coefficients from Eqs. (2.35), (2.36), (2.37), (2.38) can be recovered respectively from Eqs. (2.47), (2.48), (2.49), (2.50).

Note that, similar to the binary case, one can also rewrite Eqs. (2.47), (2.48), (2.49), (2.50) as

$$C_{w,i}^{in} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} w_{jk}^{\frac{1}{3}} w_{ji}^{\frac{1}{3}} w_{ki}^{\frac{1}{3}}}{k_i^{in} (k_i^{in} - 1)}, \quad (2.51)$$

$$C_{w,i}^{out} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} w_{ik}^{\frac{1}{3}} w_{ij}^{\frac{1}{3}} w_{jk}^{\frac{1}{3}}}{k_i^{out} (k_i^{out} - 1)}, \quad (2.52)$$

$$C_{w,i}^{cyc} = \frac{\sum_{j \neq i} \sum_{k \neq i, j} w_{ij}^{\frac{1}{3}} w_{jk}^{\frac{1}{3}} w_{ki}^{\frac{1}{3}}}{k_i^{in} k_i^{out} - k_i^{\leftrightarrow}}, \quad (2.53)$$

$$C_{w,i}^{mid} = \frac{\sum_{j \neq i} \sum_{k \neq i,j} w_{ik}^{\frac{1}{3}} w_{jk}^{\frac{1}{3}} w_{ji}^{\frac{1}{3}}}{k_i^{in} k_i^{out} - k_i^{\leftrightarrow}}. \quad (2.54)$$

For the analysis of the evolution of the prevalence of a particular type of relationship observed in node triangles over time, we compute the averages of the local weighted clustering coefficients across all nodes as

$$\bar{C}_w^{in} = \frac{1}{n} \sum_{i=1}^n C_{w,i}^{in}, \quad (2.55)$$

$$\bar{C}_w^{out} = \frac{1}{n} \sum_{i=1}^n C_{w,i}^{out}, \quad (2.56)$$

$$\bar{C}_w^{cyc} = \frac{1}{n} \sum_{i=1}^n C_{w,i}^{cyc}, \quad (2.57)$$

$$\bar{C}_w^{mid} = \frac{1}{n} \sum_{i=1}^n C_{w,i}^{mid}. \quad (2.58)$$

2.2.3 Configuration models

In this subsection we will summarize the main ideas behind the algorithm involved in the extraction of hidden (latent) variables from an observed network and their role in the network randomization process (see, for example, Squartini and Garlaschelli, 2011; Squartini et al., 2011a; Squartini et al., 2011b; Mastrandrea et al., 2014; Squartini et al., 2015). For a more detailed explanation of the derivation of the family of Exponential Random Graph Model based on the maximum-entropy method, as well as on how to use the maximum-likelihood method to solve for the hidden variables under given constraints, we refer readers to the studies by Park and Newman (2004), Squartini and Garlaschelli (2011), and Squartini et al. (2015).

Undirected Binary Configuration model (UBCM)

In the UBCM, briefly, the entropy of a randomized ensemble of networks is maximized under the constraint that the node degrees in the observed network $\{k_i^{un}\}_{i=1}^n$ should match the averages of node degrees in the randomized ensemble. Mathematically, we need to solve the following system of n equations to obtain the non-negative hidden variables $\{x_i^*\}_{i=1}^n$ that carry the information from the constraints and allow us to perform an efficient unbiased

sampling of the ensemble

$$\sum_{j \neq i} \frac{x_i^* x_j^*}{1 + x_i^* x_j^*} = k_i^{un}, \forall i = 1, 2, \dots, n. \quad (2.59)$$

Once obtained, the hidden variable can be used to compute the probability p_{ij} of a link between any two nodes i and j , which in turn allows us to easily sample the ensemble associated with the above constraints

$$p_{ij} = \langle a_{ij} \rangle = \frac{x_i^* x_j^*}{1 + x_i^* x_j^*}, \quad (2.60)$$

where $\langle a_{ij} \rangle$ is the notation for the expectation of a_{ij} over the ensemble.

Directed Binary Configuration model (DBCM)

In the DBCM, the constraints are the observed out-degree and in-degree sequences $\{k_i^{out}\}_{i=1}^n$ and $\{k_i^{in}\}_{i=1}^n$. We need to solve the following system of $2n$ equations to obtain the associated non-negative hidden variables $\{x_i^*\}_{i=1}^n$ and $\{y_i^*\}_{i=1}^n$

$$\begin{cases} \sum_{j \neq i} \frac{x_i^* y_j^*}{1 + x_i^* y_j^*} = k_i^{out}, \forall i = 1, 2, \dots, n, \\ \sum_{j \neq i} \frac{x_j^* y_i^*}{1 + x_j^* y_i^*} = k_i^{in}, \forall i = 1, 2, \dots, n. \end{cases} \quad (2.61)$$

The probability of a link from node i to j is given by

$$p_{ij} = \langle a_{ij} \rangle = \frac{x_i^* y_j^*}{1 + x_i^* y_j^*}, \quad (2.62)$$

and the probability of a link from node j to i is given by

$$p_{ji} = \langle a_{ji} \rangle = \frac{x_j^* y_i^*}{1 + x_j^* y_i^*}. \quad (2.63)$$

Undirected Weighted Configuration model (UWCM)

Similarly, suppose that in an undirected weighted network we want to extract n hidden variables $\{x_i^*\}_{i=1}^n$ associated with the observed strength sequence $\{s_i^{un}\}_{i=1}^n$, ($\{x_i^*\}_{i=1}^n \in [0, 1)$). The maximum likelihood method involves solving the following system of n equations for the hidden variables

$$\sum_{j \neq i} \frac{x_i^* x_j^*}{1 - x_i^* x_j^*} = s_i^{un}, \forall i = 1, 2, \dots, n. \quad (2.64)$$

The expected link weight between node i and node j is given by

$$\langle w_{ij} \rangle = \frac{x_i^* x_j^*}{1 - x_i^* x_j^*}. \quad (2.65)$$

The probability of a link weight w_{ij} between node i and node j in the UWCM is

$$q(w_{ij}) = (p_{ij})^{w_{ij}} (1 - p_{ij}), \quad (2.66)$$

for $w_{ij} > 0$, where $p_{ij} = \langle a_{ij} \rangle$ is the probability of a link between two nodes (i, j) , which is given by

$$p_{ij} = \langle a_{ij} \rangle = x_i^* x_j^*. \quad (2.67)$$

Directed Weighted Configuration model (DWCM)

In the DWCM, the constraints are the observed out-strength and in-strength sequences (i.e. $\{s_i^{out}\}_{i=1}^n$ and $\{s_i^{in}\}_{i=1}^n$). Mathematically, we need to solve the following system of $2n$ equations to obtain the hidden variables $\{x_i^*\}_{i=1}^n$ and $\{y_i^*\}_{i=1}^n \in [0, 1)$, which are respectively associated with $\{s_i^{out}\}_{i=1}^n$ and $\{s_i^{in}\}_{i=1}^n$

$$\begin{cases} \sum_{j \neq i} \frac{x_i^* y_j^*}{1 - x_i^* y_j^*} = s_i^{out}, \forall i = 1, 2, \dots, n, \\ \sum_{j \neq i} \frac{x_j^* y_i^*}{1 - x_j^* y_i^*} = s_i^{in}, \forall i = 1, 2, \dots, n. \end{cases} \quad (2.68)$$

The expected link weights between node i and node j are given by

$$\langle w_{ij} \rangle = \frac{x_i^* y_j^*}{1 - x_i^* y_j^*}, \quad (2.69)$$

and

$$\langle w_{ji} \rangle = \frac{x_j^* y_i^*}{1 - x_j^* y_i^*}. \quad (2.70)$$

The probability of a link weight w_{ij} from node i to node j in the DWCM is

$$q(w_{ij}) = (p_{ij})^{w_{ij}} (1 - p_{ij}), \quad (2.71)$$

for $w_{ij} > 0$, where $p_{ij} = \langle a_{ij} \rangle$ is the probability of a link between two nodes (i, j) , given by

$$p_{ij} = \langle a_{ij} \rangle = x_i^* y_j^*. \quad (2.72)$$

Undirected Enhanced Configuration model (UECM)

In the UECM, we use both the degree sequence $\{k_i^{un}\}_{i=1}^n$ as well as the strength sequence $\{s_i^{un}\}_{i=1}^n$ as constraints. The associated non-negative hidden variables $\{x_i^*\}_{i=1}^n$ and $\{y_i^*\}_{i=1}^n$ ($\{y_i^*\}_{i=1}^n \in [0, 1)$) are then the solution to the following system of $2n$ equations

$$\begin{cases} \sum_{j \neq i} \frac{x_i^* x_j^* y_i^* y_j^*}{1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*} = k_i^{un}, \forall i = 1, 2, \dots, n, \\ \sum_{j \neq i} \frac{x_i^* x_j^* y_i^* y_j^*}{(1 - y_i^* y_j^*)(1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*)} = s_i^{un}, \forall i = 1, 2, \dots, n. \end{cases} \quad (2.73)$$

It should be noted that, in the UECM, the probability of a link (i.e. $\langle a_{ij} \rangle$) and the expected weight (i.e. $\langle w_{ij} \rangle$) between node i and node j depend on the information encoded in the strengths as well as in the degrees. More specifically, they are given by

$$p_{ij} = \langle a_{ij} \rangle = \frac{x_i^* x_j^* y_i^* y_j^*}{1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*}, \quad (2.74)$$

and

$$\langle w_{ij} \rangle = \frac{x_i^* x_j^* y_i^* y_j^*}{(1 - y_i^* y_j^*)(1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*)}. \quad (2.75)$$

In this model the probability of a link weight w_{ij} between two nodes (i, j) is given by

$$q(w_{ij}) = \begin{cases} 1 - p_{ij}, & \text{if } w_{ij} = 0 \\ p_{ij}(r_{ij})^{w_{ij}-1}(1 - r_{ij}), & \text{if } w_{ij} > 0, \end{cases} \quad (2.76)$$

where $r_{ij} = y_i^* y_j^*$, and p_{ij} is defined by Eq. (2.74).

Directed Enhanced Configuration model (DECM)

In the DECM, the non-negative hidden variables $\{x_i^*\}_{i=1}^n$, $\{y_i^*\}_{i=1}^n$, $\{z_i^*\}_{i=1}^n$, $\{t_i^*\}_{i=1}^n$ ($\{z_i^*\}_{i=1}^n$, $\{t_i^*\}_{i=1}^n \in [0, 1)$) extracted from the four sequences of constraints $\{k_i^{out}\}_{i=1}^n$, $\{k_i^{in}\}_{i=1}^n$, $\{s_i^{out}\}_{i=1}^n$, and $\{s_i^{in}\}_{i=1}^n$ are the solution to the following system of $4n$ equations

$$\begin{cases} \sum_{j \neq i} \frac{x_i^* y_j^* z_i^* t_j^*}{1 - z_i^* t_j^* + x_i^* y_j^* z_i^* t_j^*} = k_i^{out}, \forall i = 1, 2, \dots, n, \\ \sum_{j \neq i} \frac{x_j^* y_i^* z_j^* t_i^*}{1 - z_j^* t_i^* + x_j^* y_i^* z_j^* t_i^*} = k_i^{in}, \forall i = 1, 2, \dots, n, \\ \sum_{j \neq i} \frac{x_i^* y_j^* z_i^* t_j^*}{(1 - z_i^* t_j^*)(1 - z_i^* t_j^* + x_i^* y_j^* z_i^* t_j^*)} = s_i^{out}, \forall i = 1, 2, \dots, n, \\ \sum_{j \neq i} \frac{x_j^* y_i^* z_j^* t_i^*}{(1 - z_j^* t_i^*)(1 - z_j^* t_i^* + x_j^* y_i^* z_j^* t_i^*)} = s_i^{in}, \forall i = 1, 2, \dots, n. \end{cases} \quad (2.77)$$

Similar to the UECM, in the DECM, the probability of a link (i.e. $\langle a_{ij} \rangle$) and the expected weight (i.e. $\langle w_{ij} \rangle$) from node i to node j depend on information encoded in the two sequences

of observed degrees as well as in the two sequences of observed strengths. More specifically, we have

$$p_{ij} = \langle a_{ij} \rangle = \frac{x_i^* y_j^* z_i^* t_j^*}{1 - z_i^* t_j^* + x_i^* y_j^* z_i^* t_j^*}, \quad (2.78)$$

$$\langle w_{ij} \rangle = \frac{x_i^* y_j^* z_i^* t_j^*}{(1 - z_i^* t_j^*)(1 - z_i^* t_j^* + x_i^* y_j^* z_i^* t_j^*)}, \quad (2.79)$$

Similarly,

$$p_{ji} = \langle a_{ji} \rangle = \frac{x_j^* y_i^* z_j^* t_i^*}{1 - z_j^* t_i^* + x_j^* y_i^* z_j^* t_i^*}, \quad (2.80)$$

$$\langle w_{ji} \rangle = \frac{x_j^* y_i^* z_j^* t_i^*}{(1 - z_j^* t_i^*)(1 - z_j^* t_i^* + x_j^* y_i^* z_j^* t_i^*)}. \quad (2.81)$$

The probability of a link weight w_{ij} from node i to node j is now given by

$$q(w_{ij}) = \begin{cases} 1 - p_{ij}, & \text{if } w_{ij} = 0, \\ p_{ij}(r_{ij})^{w_{ij}-1}(1 - r_{ij}), & \text{if } w_{ij} > 0, \end{cases} \quad (2.82)$$

where $r_{ij} = z_i^* t_j^*$, and p_{ij} is defined by Eq. (2.78).

Note that, the expected values of the second and third structural correlations in the randomized networks can be analytically computed via the hidden variables extracted from each configuration model or numerically computed by taking the average over a simulated ensemble. In our study, for each considered null model, we generate an ensemble of 1000 randomized networks, and then take the averages of the measures in question over the ensemble.

2.3 Findings for the binary network

2.3.1 Structural correlations in the undirected binary e-MID network

We first investigate the degree dependencies in the undirected binary e-MID network by examining the relationship between the node degree (k^{un}) and the average degree of its neighbors (k_{nn}^{un}). The overall disassortativity in this version of the network is evidenced by the negative relationship between these two quantities as shown in panels (a) and (b) of Figure (2.3), in which the measures for the networks from Q1 (the first quarter in our data set) and from Q48 (the last quarter in our data set) are plotted as an example. Note that

the overall negative correlation between k_{nn}^{un} and k^{un} is also observed in all 48 quarters from 1999 to 2010. In addition, generally, we find that the absolute value of this correlation is declining over time. This also seems to be true for other observed dependencies.

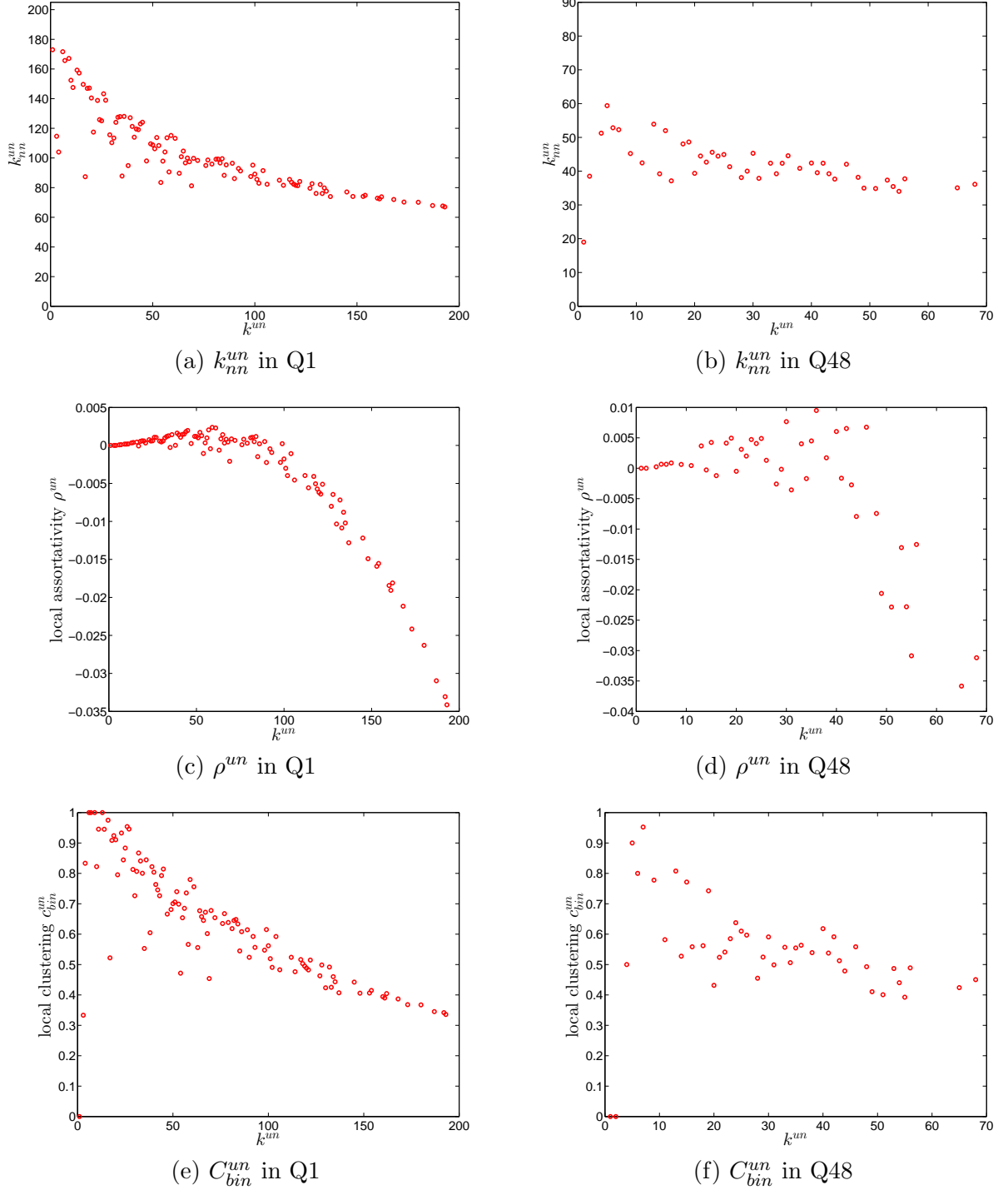


Figure 2.3: ANND (panels a, b), local assortativity ρ^{un} (panels c, d), and local clustering coefficients C_{bin}^{un} (panels e, f) in the undirected binary e-MID network, in Q1 and Q48.

Next, we now turn to the Pearson correlation coefficient of degrees r_{bin}^{un} as an overall indicator of degree dependencies in the network. As shown in Figure (2.4), over time, overall, the network exhibits disassortativity as signaled by the negative coefficient. Consistent with what we discovered in our analysis of the measure ANND, the absolute value of r_{bin}^{un} is also declining from 1999 to 2010.

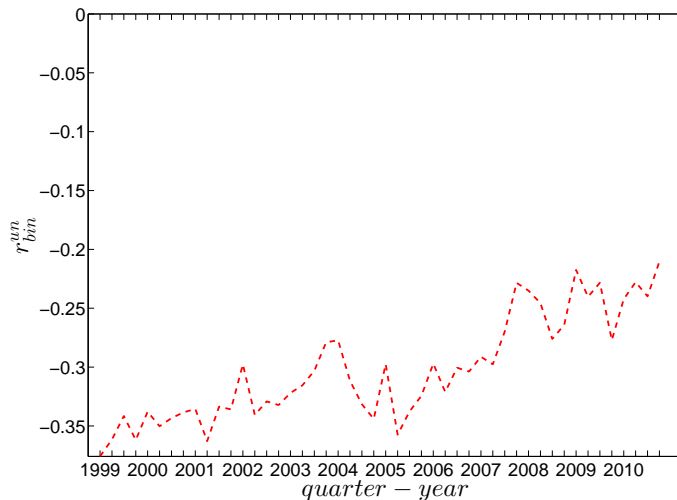


Figure 2.4: Evolution of the overall assortativity indicator r_{bin}^{un} in the undirected binary e-MID network.

For a more comprehensive assessment of the degree dependencies in the network, we employ the local assortativity coefficients ρ^{un} that expose the contribution of each node to the global level of assortativity r_{bin}^{un} (see the Appendix for further details). The basic idea is that the numerator in the Pearson correlation coefficient proposed by Newman (2002, 2003a) can be reformulated based on the contribution of the individual nodes instead of in term of the edges (see, for example, Piraveenan et al., 2010). It should be emphasized that we always have

$$r_{bin}^{un} = \sum_{i=1}^n \rho_i^{un}. \quad (2.83)$$

In panels (c) and (d) of Figure (2.3) we plot ρ^{un} against k^{un} to investigate which nodes (in terms of their degrees) contribute most to r_{bin}^{un} . It is clear that the hubs are the primarily contributors to the overall disassortativity of the network, while smaller degree nodes sometimes exhibit assortativity. This also reveals that adding or removing a hub from a network may have a large impact on its overall mixing nature.

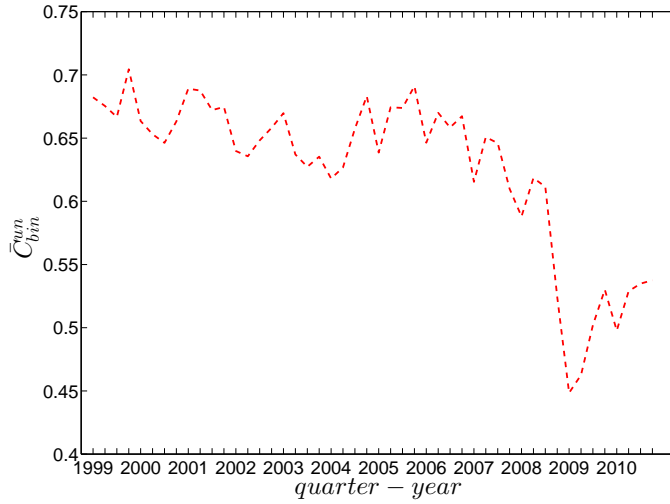


Figure 2.5: Evolution of the average of local clustering coefficients (i.e. \bar{C}_{bin}^{un}) in the undirected binary e-MID network.

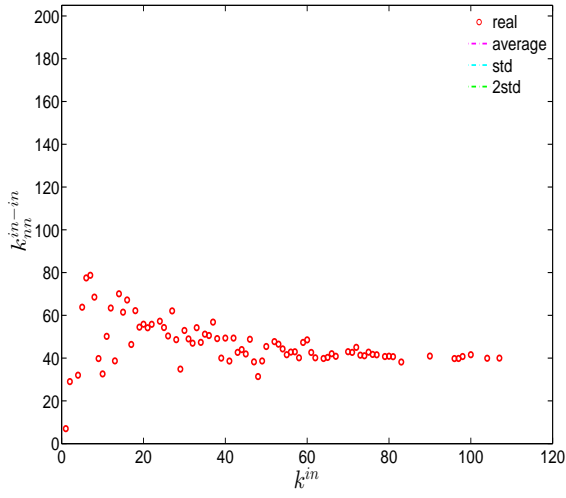
For the third order correlations, we employ the local clustering coefficient proposed by Watts and Strogatz (1998). In this simple version of the network (undirected binary case) clustering refers to the extent to which two connected nodes in the network have common neighbors. We observe that, overall, the undirected local clustering is a decreasing function of degree (panels (e) and (f) of Figure (2.3)), meaning that the neighbors of highly (poorly) connected banks are poorly (highly) interconnected. In fact, this relationship is typically found in many real world networks exhibiting a high heterogeneity in the degrees and a disassortative mixing nature (e.g. see Newman, 2003b). In our network, the bank degrees are highly heterogeneous, and the small (large) degree banks seem to have larger (smaller) local clustering coefficients because they are mostly connected to large (small) degree banks.

The evolution of the average of the undirected local binary clustering coefficients over all nodes is shown in Figure (2.5), where we can see a significant reduction in \bar{C}_{bin}^{un} around the time the financial crisis spreads to Europe.

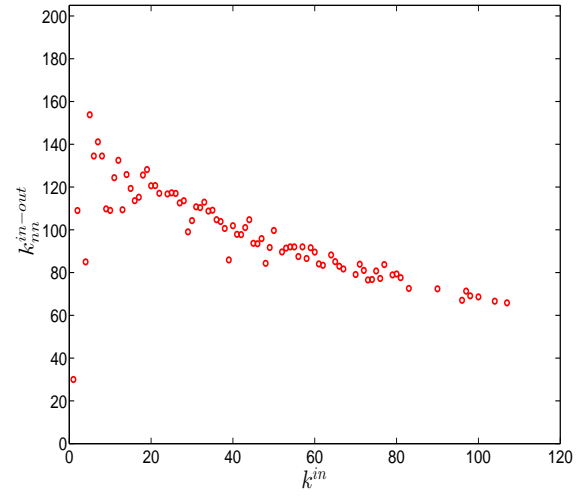
2.3.2 Structural correlations in the directed binary e-MID network

We now extend our analysis to the directed version of the binary network. Figures (2.6) and (2.7) show the relationship between ANND and node degree for the four types of mixing, i.e. in-in, in-out, out-in, and out-out. In the same network some types of mixing can be assortative, while others disassortative. For instance, while in Q1, overall, ANND is a decreasing function of the associated degree in all four cases, this relationship breaks down for the in-in and out-out mixing in Q48. In contrast, the overall negative correlation between ANND and the associated degree for the in-out and out-in mixing is observed in almost all quarters.

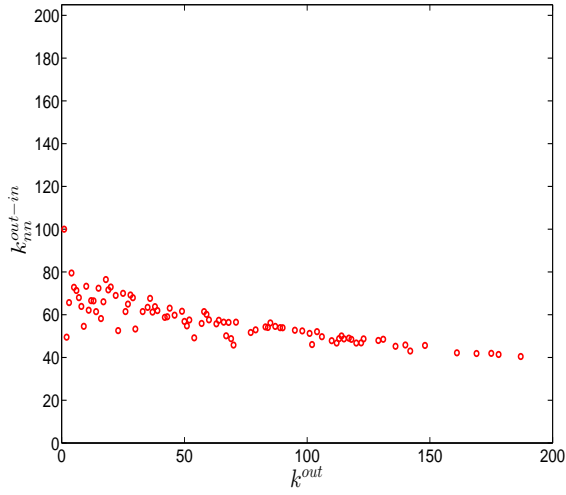
For a more general assessment of the overall mixing nature in the directed binary network, we calculate the Pearson correlation coefficient in each category of mixing and show its evolution over time (see Figure (2.8)). In comparison to the undirected version, the directed binary network displays more complicated degree dependencies. We can see that r_{bin}^{out-in} and r_{bin}^{in-out} display a different behavior than r_{bin}^{in-in} and $r_{bin}^{out-out}$. More specifically, while in the out-in and out-in categories we persistently observe disassortativity in all quarters, the other two categories switch between displaying assortativity and disassortativity over time. The interpretation of the mixing observed in the various categories is similar to the interpretation of mixing in an undirected binary network. For instance, a negative value of r_{bin}^{out-in} , signaling disassortativity in the out-in category, indicates that a high out-degree bank tends to have out-going links to low in-degree banks, and/or that a low out-degree bank tends to have out-going links to high in-degree banks. The mixing we observe in the out-in category (r_{bin}^{out-in}) comes closest to the one observed in the undirected network captured by r_{bin}^{un} . The similarity between these two quantities was mathematically proven by van der Hoorn and Litvak (2015). In addition, although the in-out mixing category exhibits disassortativity, in many quarters the coefficient r_{bin}^{in-out} is very close to zero.



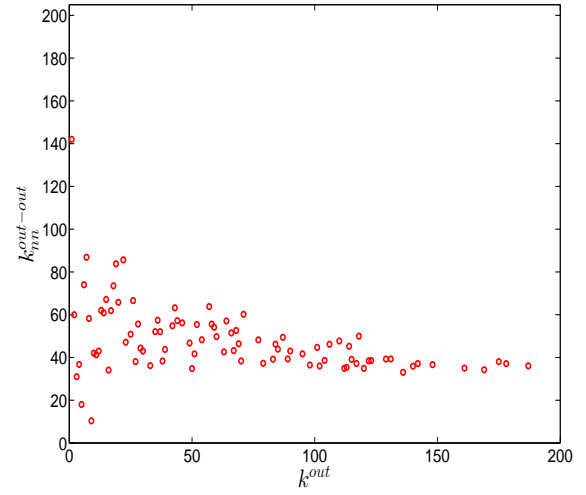
(a) k_{nn}^{in-in}



(b) k_{nn}^{in-out}



(c) k_{nn}^{out-in}



(d) $k_{nn}^{out-out}$

Figure 2.6: ANND in the directed binary e-MID network, in Q1. k_{nn}^{in-in} (panel a), k_{nn}^{in-out} (panel b), k_{nn}^{out-in} (panel c), $k_{nn}^{out-out}$ (panel d).

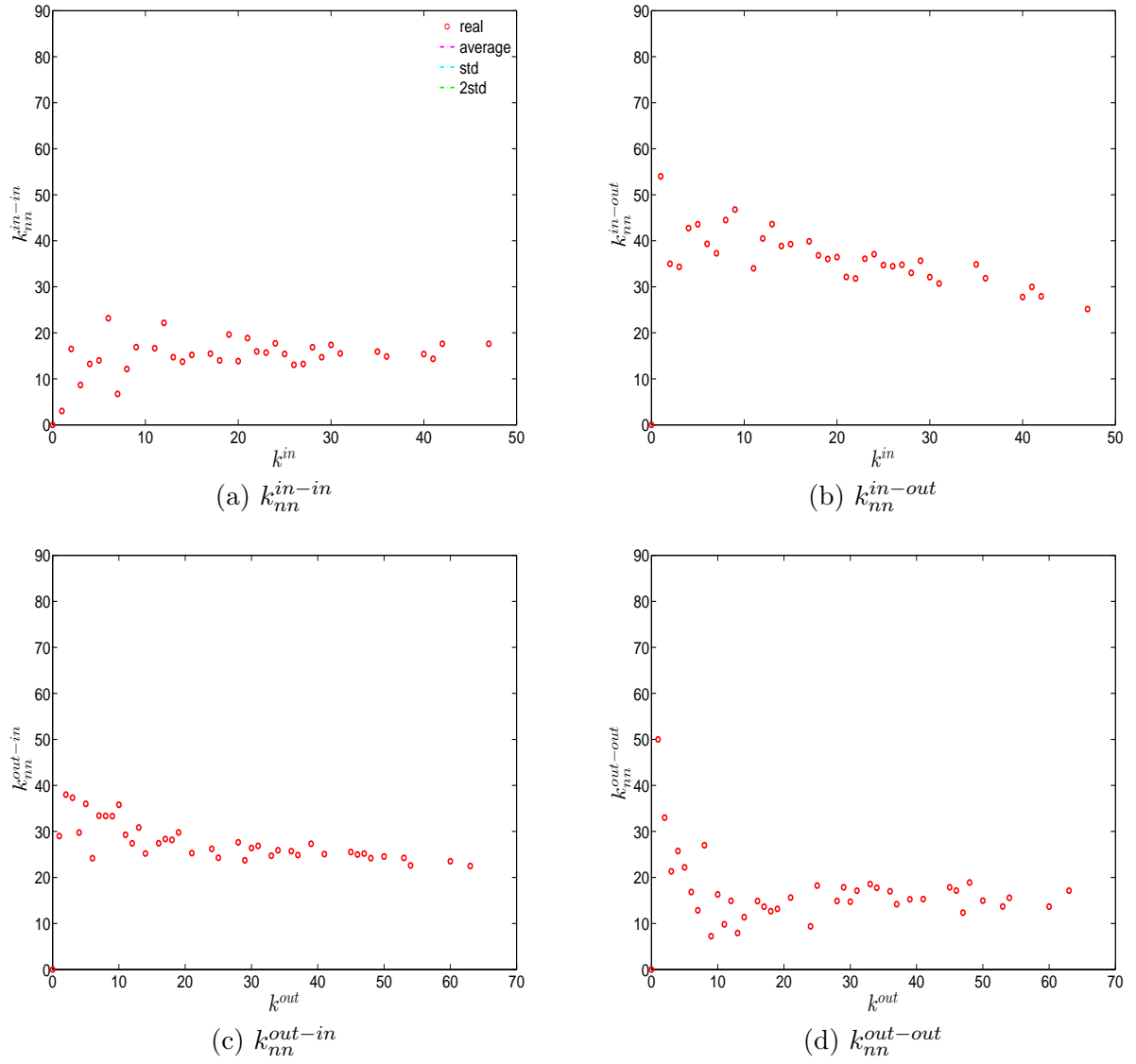


Figure 2.7: ANND in the directed binary e-MID network, in Q48. k_{nn}^{in-in} (panel a), k_{nn}^{in-out} (panel b), k_{nn}^{out-in} (panel c), $k_{nn}^{out-out}$ (panel d).

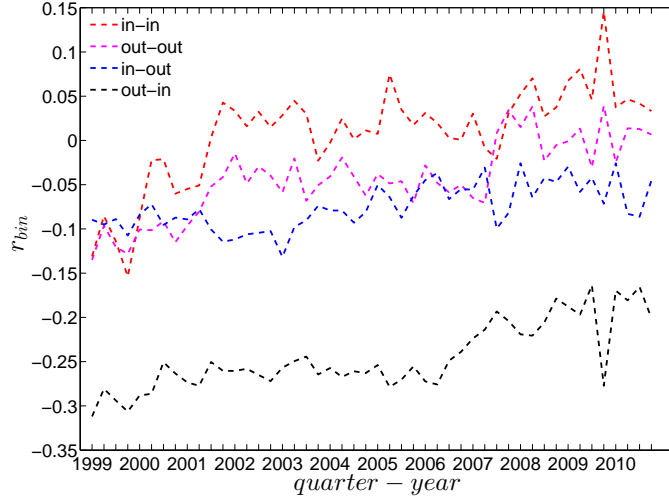


Figure 2.8: Evolution of the overall assortativity indicators in the directed binary e-MID network.

Similarly to the undirected case, we define the local assortativity measures for a given node i as ρ_i^{in-in} , ρ_i^{in-out} , ρ_i^{out-in} , and $\rho_i^{out-out}$ corresponding to the four mixing categories in the directed version of the network. Note that the following equalities must hold:

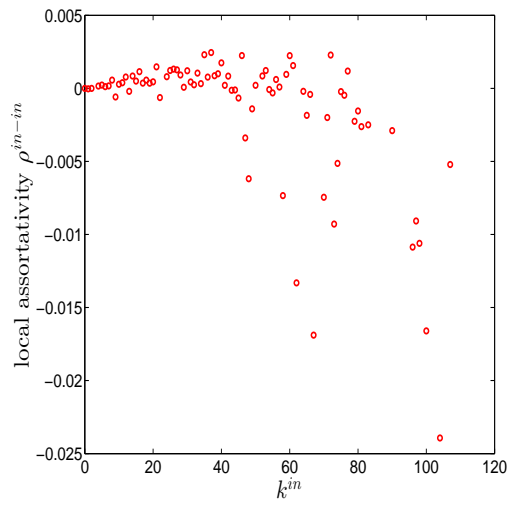
$$r_{bin}^{in-in} = \sum_{i=1}^n \rho_i^{in-in}, \quad (2.84)$$

$$r_{bin}^{in-out} = \sum_{i=1}^n \rho_i^{in-out}, \quad (2.85)$$

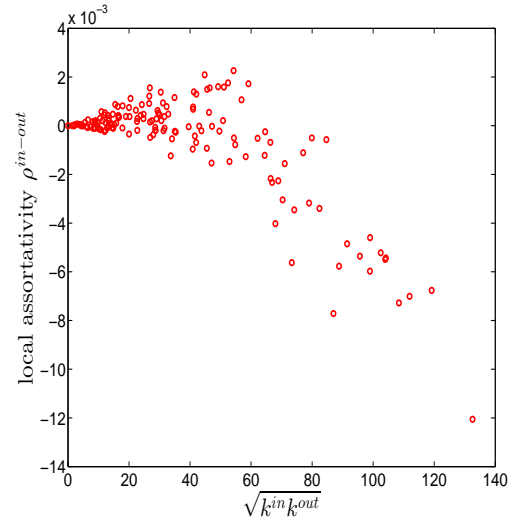
$$r_{bin}^{out-in} = \sum_{i=1}^n \rho_i^{out-in}, \quad (2.86)$$

$$r_{bin}^{out-out} = \sum_{i=1}^n \rho_i^{out-out}. \quad (2.87)$$

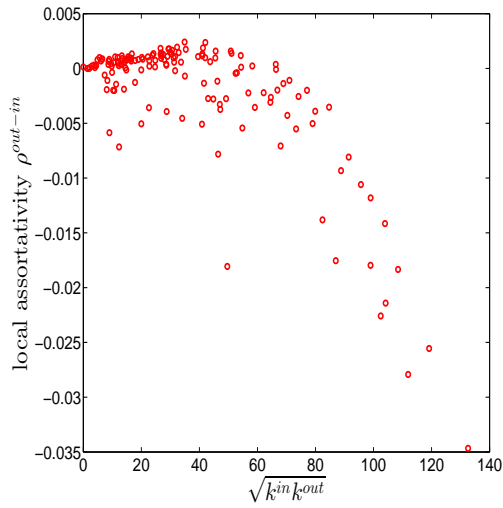
The measures ρ_i^{in-in} , ρ_i^{in-out} , ρ_i^{out-in} , and $\rho_i^{out-out}$ give us useful information about the contribution of each node to the respective overall assortativity indicators.



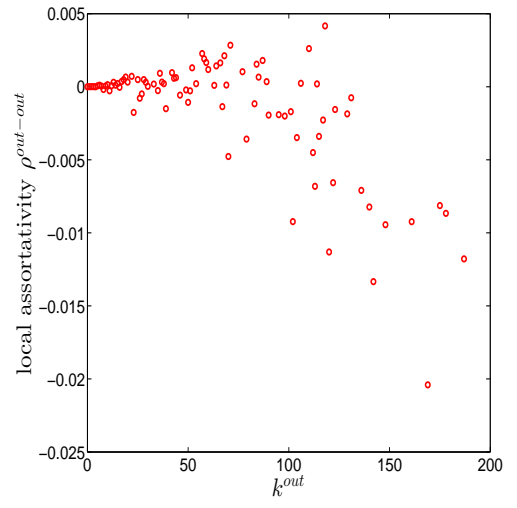
(a) ρ^{in-in}



(b) ρ^{in-out}



(c) ρ^{out-in}



(d) $\rho^{out-out}$

Figure 2.9: Local assortativity in the directed binary e-MID network, in Q1. ρ^{in-in} (panel a), ρ^{in-out} (panel b), ρ^{out-in} (panel c), $\rho^{out-out}$ (panel d).

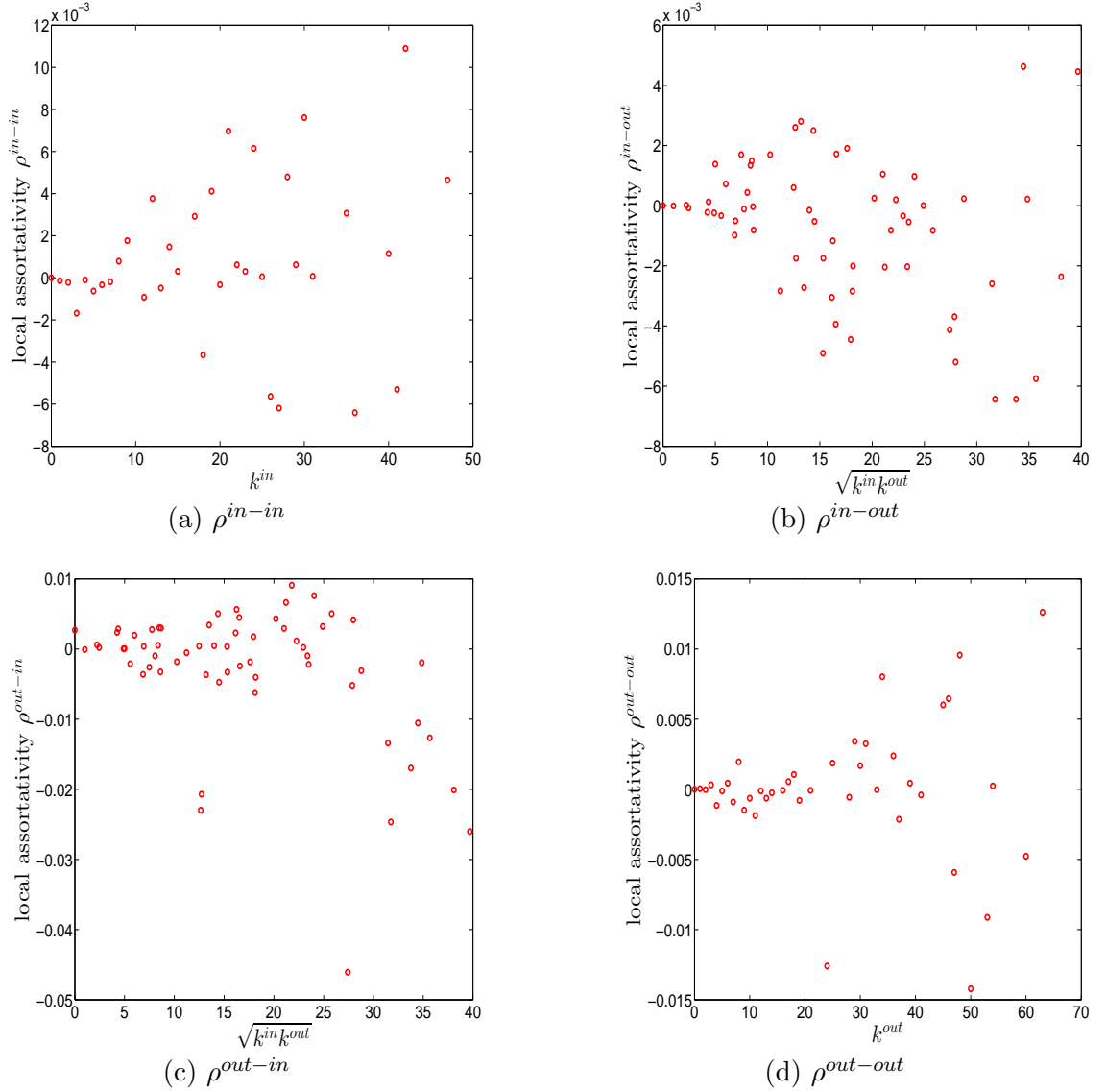


Figure 2.10: Local assortativity in the directed binary e-MID network, in Q48. ρ^{in-in} (panel a), ρ^{in-out} (panel b), ρ^{out-in} (panel c), $\rho^{out-out}$ (panel d).

The local assortativity indicators in the two quarters Q1 and Q48 are respectively shown in Figures (2.9) and (2.10). Note that for each local assortativity indicator, we consider it as the function of the corresponding degree ⁴. The results indicate that, first, given an overall level of assortativity in a particular category, the contribution of nodes of different degrees varies across the four mixing categories. In the out-in mixing category, we observe that, on the one hand, the hubs contribute most to the overall level of assortativity; on the other hand, small degree nodes can be associated with slight assortativity or disassortativity. In

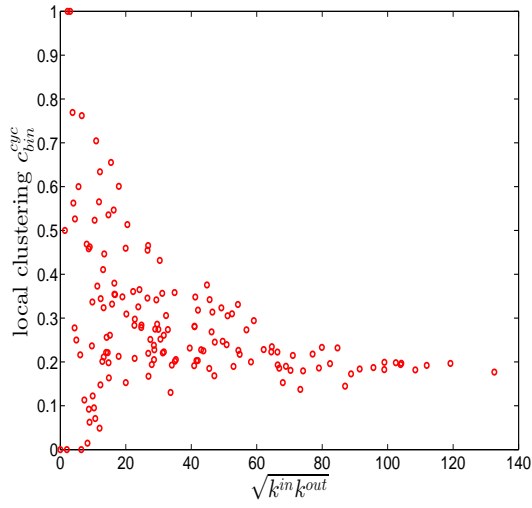
⁴In the cases of ρ^{in-out} and ρ^{out-in} , we plot them against $k^{in-out} = \sqrt{k^{in}k^{out}}$, since each of them depends on both k^{in} and k^{out} (see the Appendix for more detailed derivations).

addition, the contributions of medium degree nodes are more volatile than those of the small degree nodes. This is very similar to what we found in the undirected version of the network. However, the behavior of the local assortativity indicators becomes more complicated for the other mixing categories. For example, the contributions of hubs and medium degree nodes can fluctuate a lot, so that it becomes difficult to classify which type of nodes plays an important role for the overall level of assortativity.

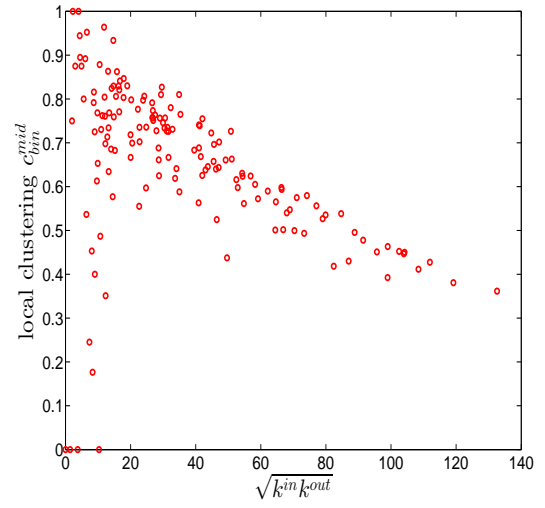
Next, we turn to the third order correlations between banks in the directed binary network. We focus on investigating local clustering as a function of degree for the four cases shown in Figure (2.2) (see, for example, Fagiolo, 2007). In the following discussion we will be referring to nodes i, j, k as an example of three vertices in a network building a triangle. It is clear that the directions of the edges now matter for the clustering analysis. The measures C^{mid} , C^{cyc} , C^{out} and C^{in} summarize the prevalence of a particular type of relationship that a node has with its neighbors. For instance, larger values of C^{mid} (see panel (b) of Figure (2.2)) may represent a higher systemic risk associated with that node, since bank i can be a source of risk as well as be exposed to risk from other banks. Clustering relationships of the type shown in panel (c) of Figure (2.2) are also conducive to systemic risk since a default of bank i would affect both its partners. Larger values of C^{in} therefore indicate a higher systemic risk related to overlapping portfolios in the banking system. This is, however, not the case for cyclical clustering relationships (captured by C^{cyc}) since in this type of relationships exposures can cancel each other out (see panel (a) of Figure (2.2)). Finally, large values of C^{out} associated with bank i tell a story about risk exposure affecting bank i itself, since both banks j and k can affect bank i in case either of them would default (see panel (d) of Figure (2.2)).

For each type of clustering relationship, we first consider the local clustering coefficient as the function of the corresponding degree ⁵. Typically, in each case, a general negative relationship is observed in the first quarters, but for later quarters this relationship becomes flatter (see, for example, Figures (2.11) and (2.12)).

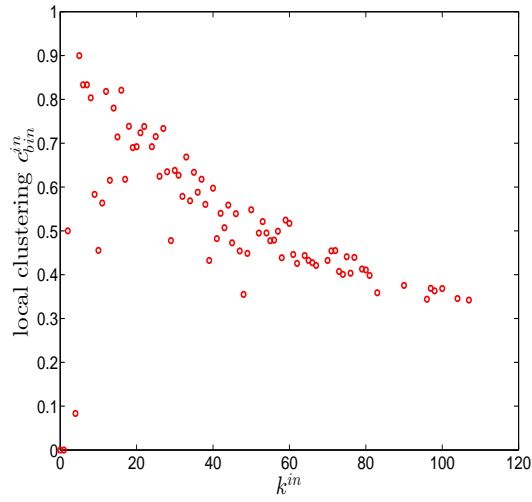
⁵In the cases of C_{bin}^{cyc} and C_{bin}^{mid} , we plot them against $k^{in-out} = \sqrt{k^{in}k^{out}}$.



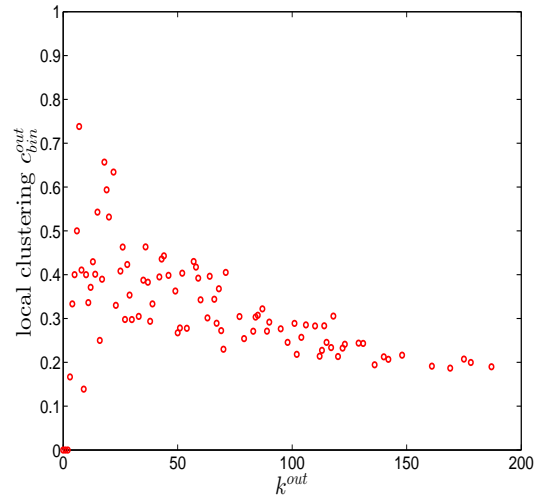
(a) C_{bin}^{cyc} in Q1



(b) C_{bin}^{mid} in Q1



(c) C_{bin}^{in} in Q1



(d) C_{bin}^{out} in Q1

Figure 2.11: Local clustering coefficients C_{bin}^{cyc} (panel a), C_{bin}^{mid} (panel b), C_{bin}^{in} (panel c), C_{bin}^{out} (panel d) in the directed binary e-MID network, in Q1.

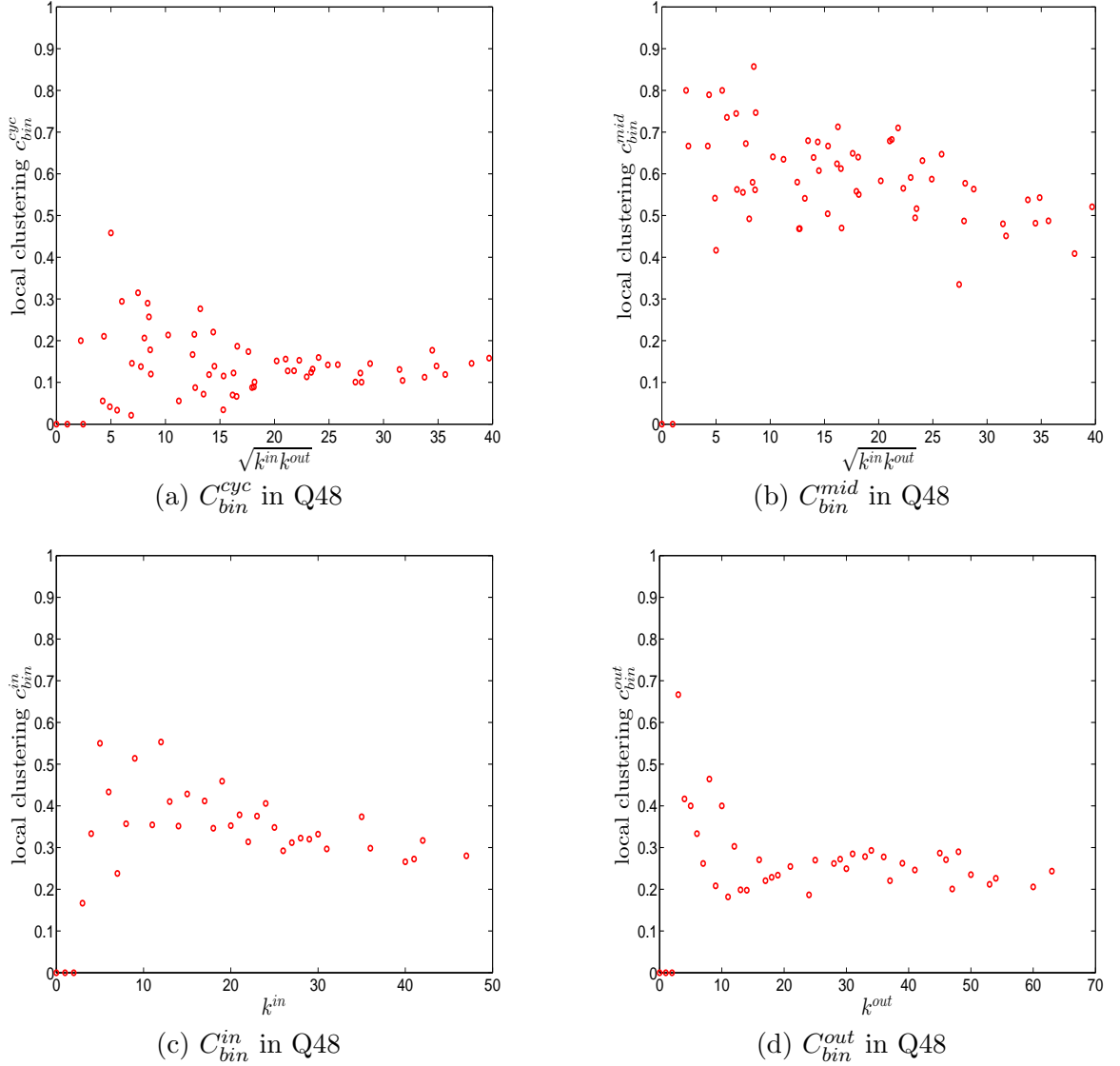


Figure 2.12: Local clustering coefficients C_{bin}^{cyc} (panel a), C_{bin}^{mid} (panel b), C_{bin}^{in} (panel c), C_{bin}^{out} (panel d) in the directed binary e-MID network, in Q48.

We now take the averages of the local clustering coefficients across all nodes and then investigate their evolution over time. We observe that, first, for the most part, the averages \bar{C}^{mid} , \bar{C}^{in} , \bar{C}^{out} , and \bar{C}^{cyc} are in descending order, with clustering relationships of the cyclical and out-type being much less common than the other two. We consider this prevalence of the middleman and in-type clustering relationships as evidence of the presence of systemic risk in the network. Second, similarly to what we observed in the undirected network for \bar{C}_{bin}^{un} , the averages of the local clustering coefficients for all four clustering types dramatically decrease around the time of the financial crisis, evidencing structural change in the third

order correlations between banks.

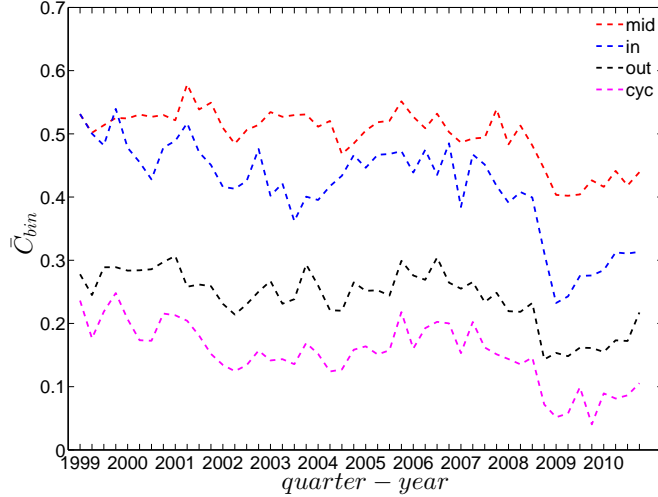
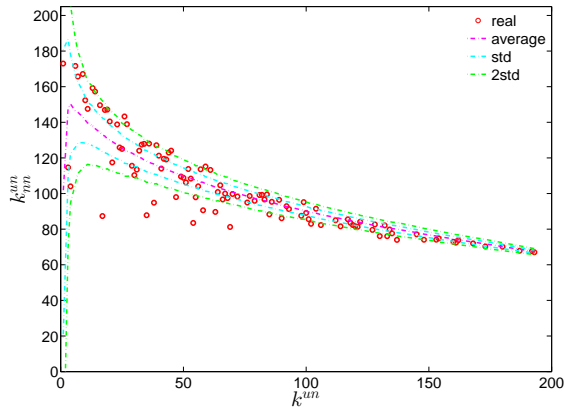


Figure 2.13: Evolution of the averages of local clustering coefficients (i.e. \bar{C}^{mid} , \bar{C}^{in} , \bar{C}^{out} , and \bar{C}^{cyc}) in the directed binary e-MID network.

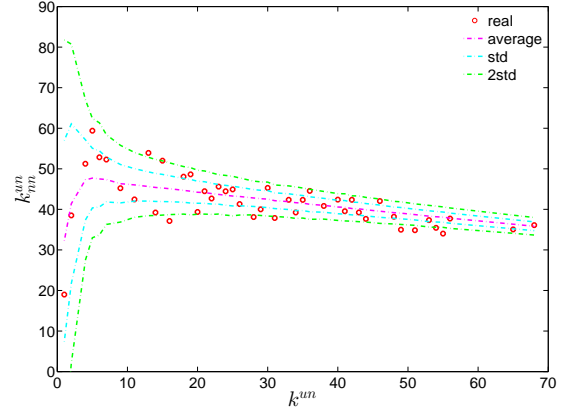
2.3.3 Comparisons to the configuration models

Undirected Binary Network

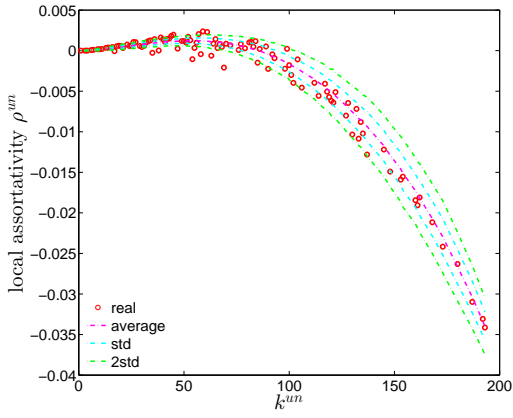
We first employ the undirected binary configuration model (UBCM), which maintains the intrinsic heterogeneity in the degree sequence of the undirected binary version of the observed e-MID network. Figure (2.14) shows a comparison between various higher order structural correlations observed in the e-MID network and the same structural correlations observed in the randomized ensemble for the first and last quarters. Note that, in each panel of Figure (2.14), besides the observed and the expected values (over the randomized ensemble), we also report the regions of ± 1 standard deviation (std.) and ± 2 std. away from the expectations. In most cases, as shown in panels (a) to (f), the local behavior of the structural correlations is well replicated by the UBCM. As shown in panel (a) of Figure (2.15), the average of the ANNDs over all nodes (\bar{k}_{nn}^{un}) is also located inside the ± 2 std. band when plotted over time. In contrast, in terms of our measure of global assortativity (r_{bin}^{un}), in almost all of the quarters, the observed values lie outside the ± 2 std. band (see panel (b) of Figure (2.15)). A similar result is obtained for the evolution of the average of the local clustering coefficients (\bar{C}_{bin}^{un}) with many significant deviations, but the main trends of the observed and the expected values are similar (see panel (c) of Figure (2.15)).



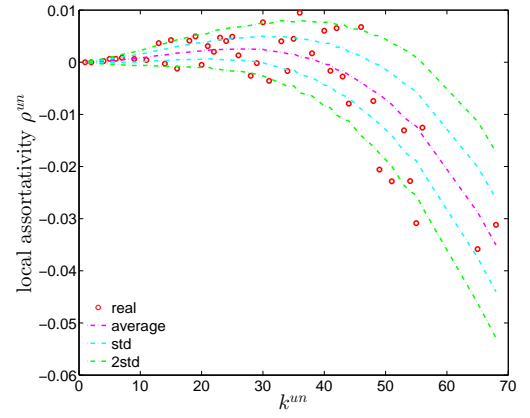
(a) k_{nn}^{un} in Q1



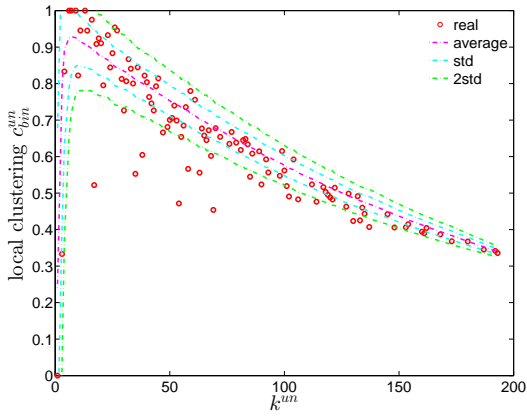
(b) k_{nn}^{un} in Q48



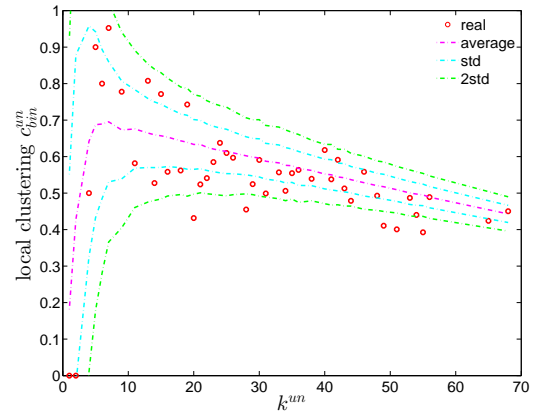
(c) ρ^{un} in Q1



(d) ρ^{un} in Q48



(e) C_{bin}^{un} in Q1



(f) C_{bin}^{un} in Q48

Figure 2.14: ANND (panels a, b), local assortativity ρ^{un} (panels c, d), local clustering coefficients C_{bin}^{un} (panels e, f) in the observed e-MID network and in the UBCM, in Q1 and Q48.

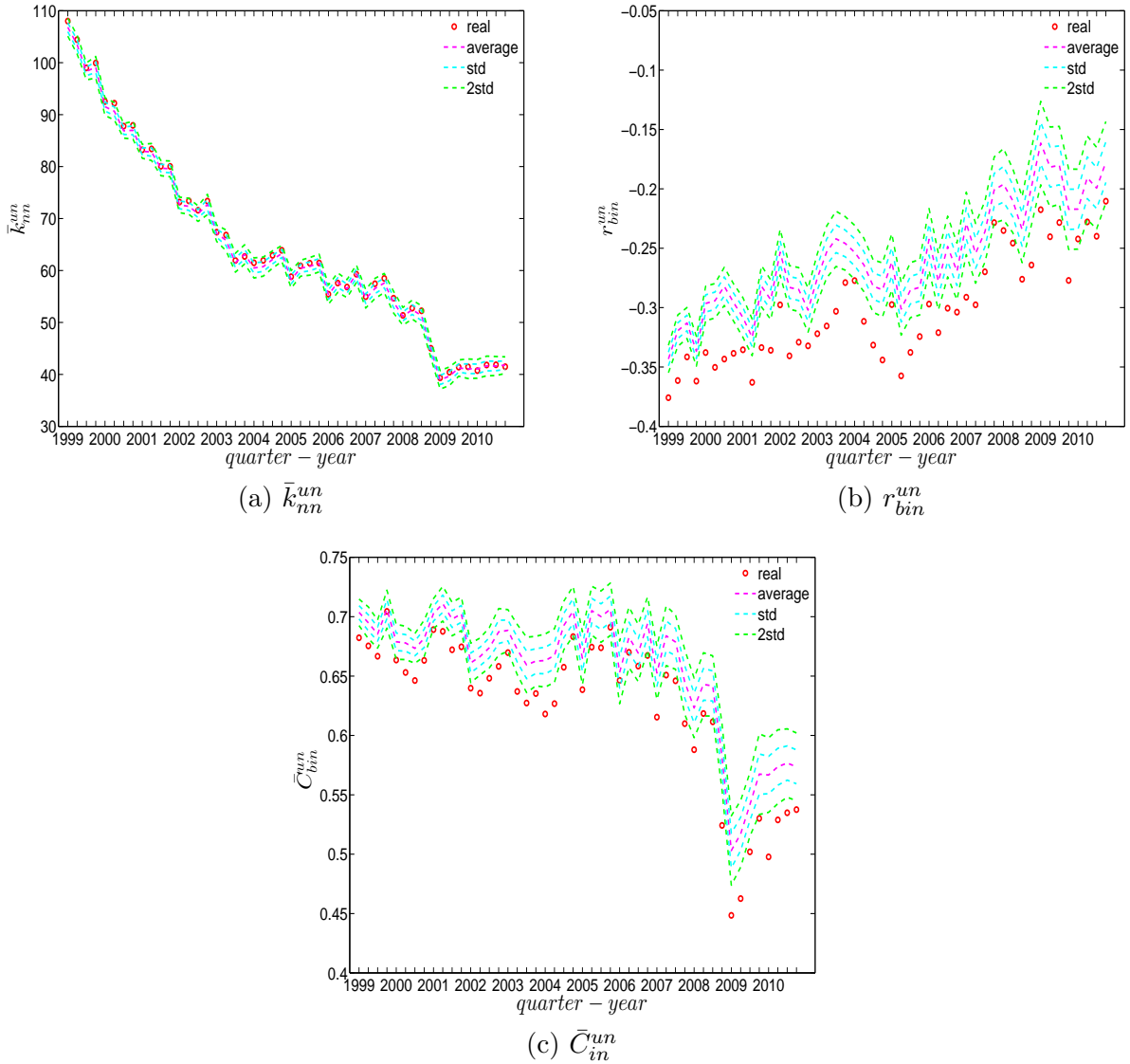


Figure 2.15: Evolution of \bar{k}_{nn}^{un} (panel a), r_{bin}^{un} (panel b), and \bar{C}_{in}^{un} (panel c) in the observed e-MID network and in the UBCM.

Directed Binary Network

Recalling that under the directed binary configuration model (DBCM), both out-going and in-coming degrees are enforced on average over the ensemble, we show the comparisons between the structural correlations of observed network and those obtained from that model in Figures (2.16) and (2.17) (for ANND), Figures (2.18) and (2.19) (for the local assortativity indicators), and Figures (2.20) and (2.21) (for the local clustering coefficients).

In addition, as for the undirected version, we also compare the evolution of the global indicators with the evolution of their expected values obtained from the DBCM. We show the results for the averages of the ANNDs (i.e. \bar{k}_{nn}^{in-in} , \bar{k}_{nn}^{in-out} , \bar{k}_{nn}^{out-in} , $\bar{k}_{nn}^{out-out}$) in Figure

(2.22), for the global assortativity indicators (r_{bin}^{in-in} , r_{bin}^{in-out} , r_{bin}^{out-in} , $r_{bin}^{out-out}$) in Figure (2.23), and for the averages of the local clustering coefficients (\bar{C}_{bin}^{in-in} , \bar{C}_{bin}^{in-out} , \bar{C}_{bin}^{out-in} , $\bar{C}_{bin}^{out-out}$) in Figure (2.24).

First, regarding the local indicators (see from Figure (2.16) to Figure (2.21)), in most cases, the observed ANNDs, local assortativity indicators, and local clustering coefficients are in agreement with those evaluated under the DBCM. Since the few observed points significantly deviating from the expected ones might not reveal any patterns (under the DBCM), they might be seen as the expected rejections one obtains for a large sequence of simultaneous tests.

Second, regarding the evolution of the averages of the ANNDs, Figure (2.22) shows that, \bar{k}_{nn}^{in-in} , \bar{k}_{nn}^{in-out} , and \bar{k}_{nn}^{out-in} always lie within the ± 2 std. band, while $\bar{k}_{nn}^{out-out}$ is underestimated for most of the time.

Third, in terms of the global assortativity indicators, for the most part, r_{bin}^{in-in} and r_{bin}^{out-in} are located inside the ± 2 std. band, while r_{bin}^{in-out} and $r_{bin}^{out-out}$ are mostly being overestimated (see Figure (2.23)).

Finally, over time, the averages of the local clustering coefficients \bar{C}_{bin}^{in-in} , \bar{C}_{bin}^{in-out} , \bar{C}_{bin}^{out-in} , and $\bar{C}_{bin}^{out-out}$ are generally in agreement with their expected values from the DBCM, as shown in Figure (2.24).

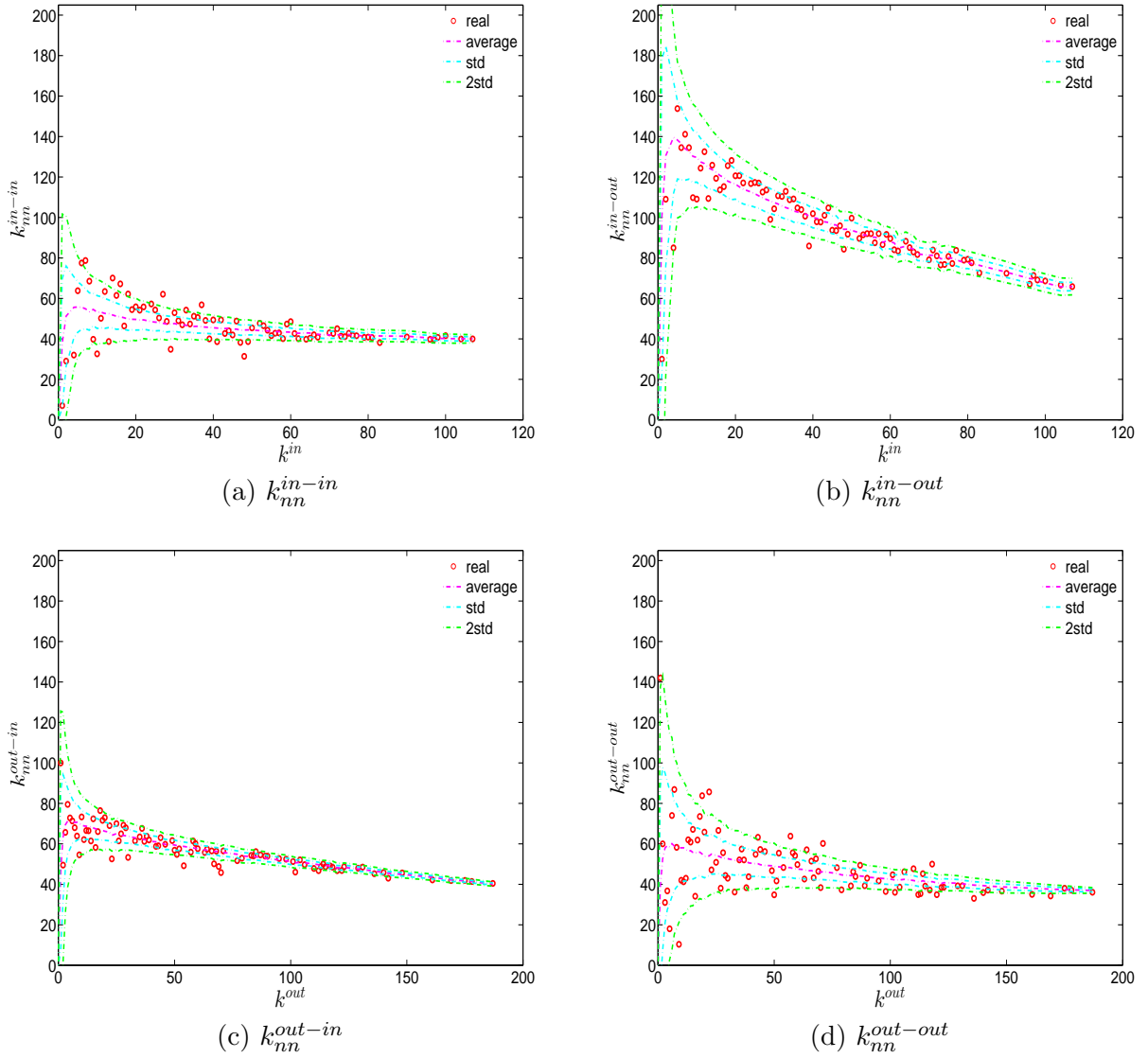


Figure 2.16: ANND in the observed e-MID network and in the DBCM, in Q1. k_{nn}^{in-in} (panel a), k_{nn}^{in-out} (panel b), k_{nn}^{out-in} (panel c), $k_{nn}^{out-out}$ (panel d).

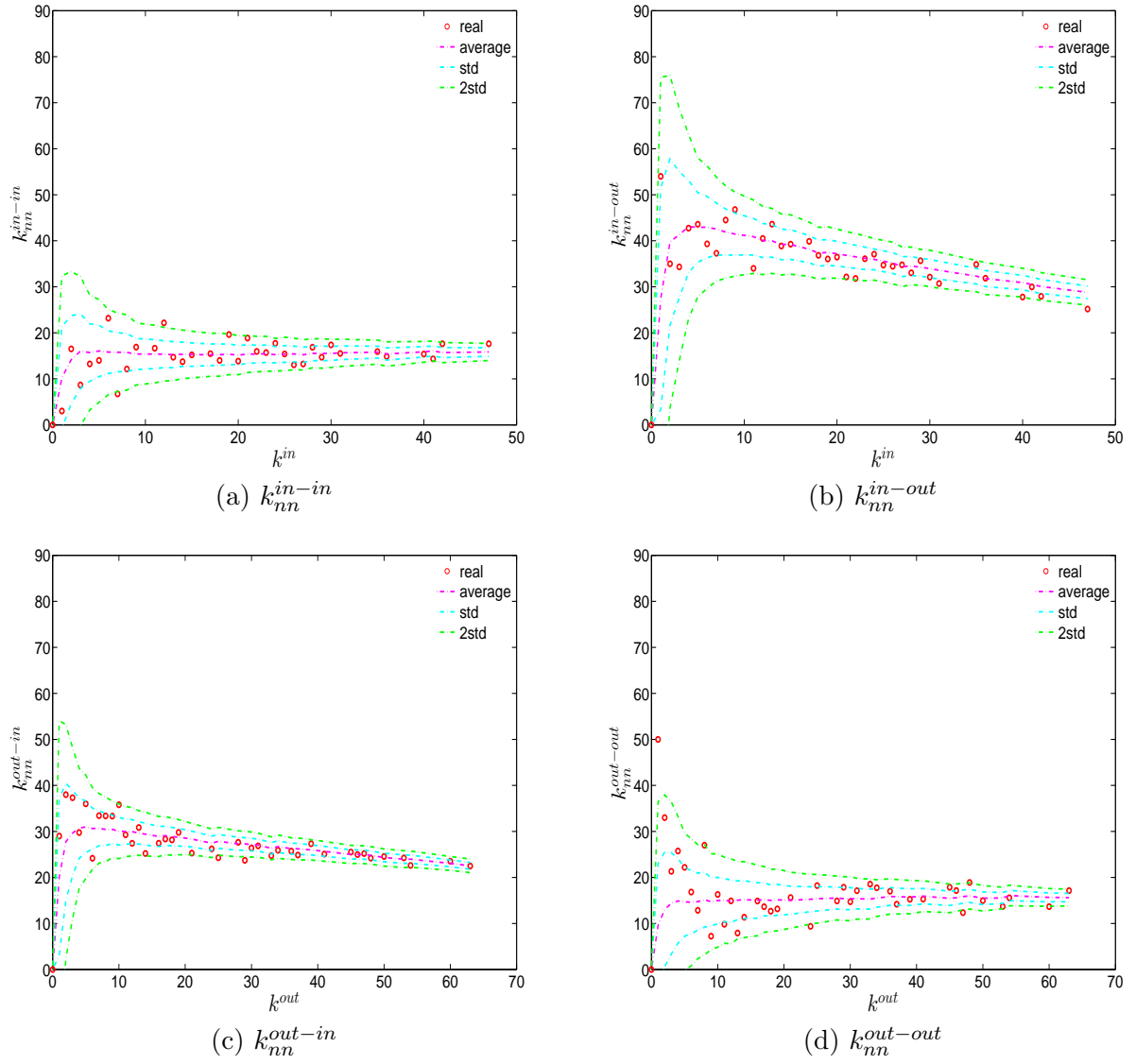


Figure 2.17: ANND in the observed e-MID network and in the DBCM, in Q48. k_{nn}^{in-in} (panel a), k_{nn}^{in-out} (panel b), k_{nn}^{out-in} (panel c), $k_{nn}^{out-out}$ (panel d).

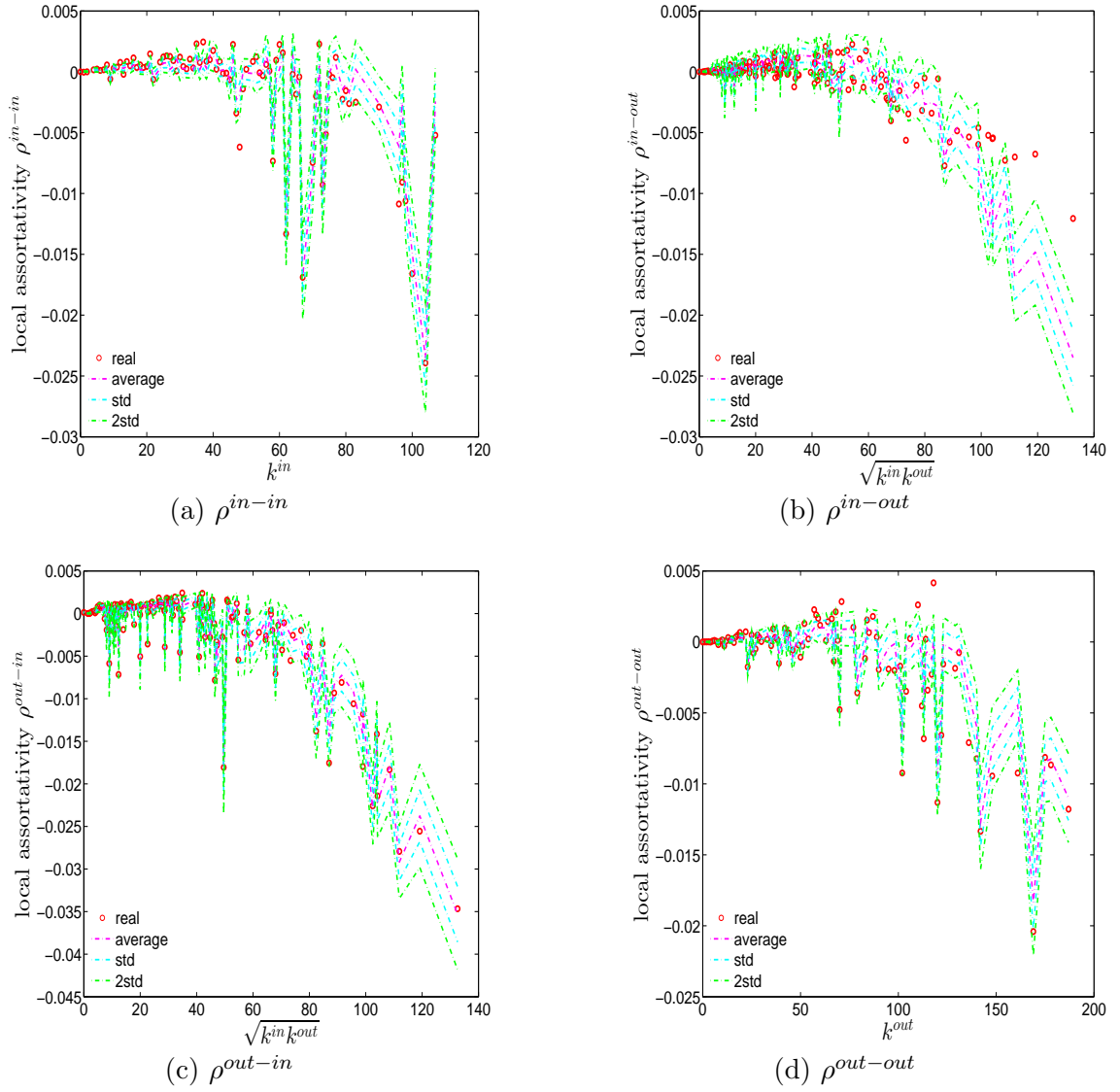


Figure 2.18: Local assortativity in the observed e-MID network and in the DBCM, in Q1. ρ^{in-in} (panel a), ρ^{in-out} (panel b), ρ^{out-in} (panel c), $\rho^{out-out}$ (panel d).

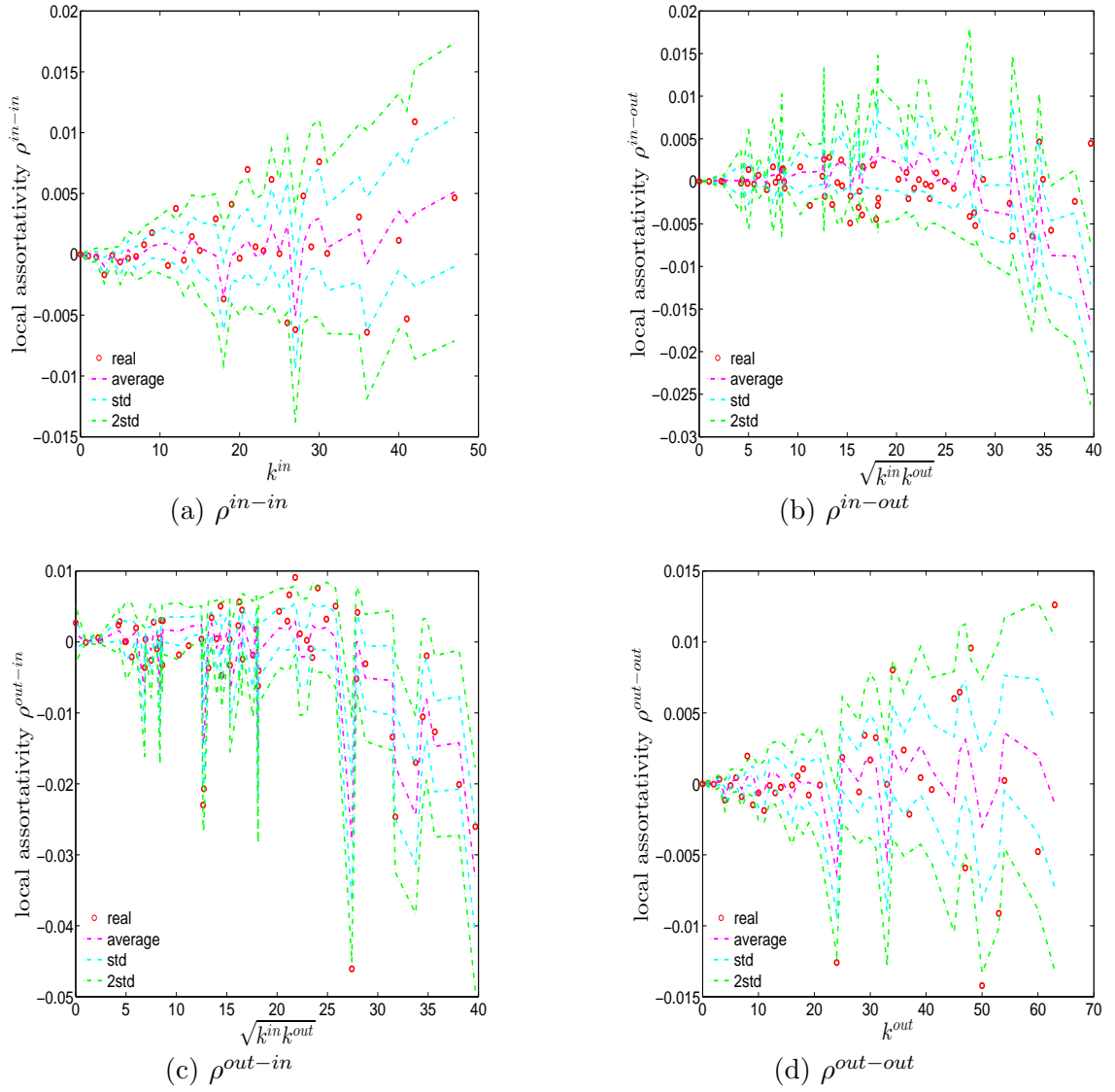
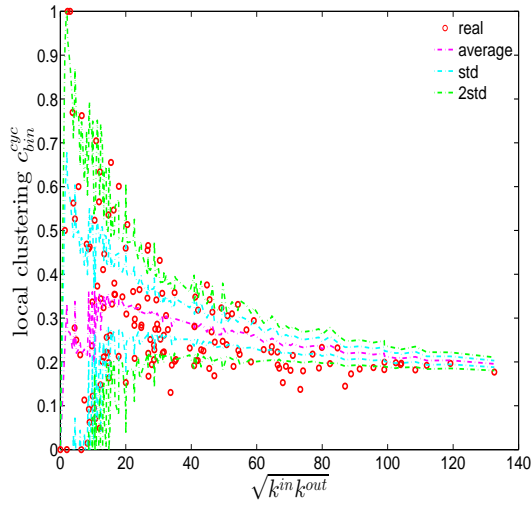
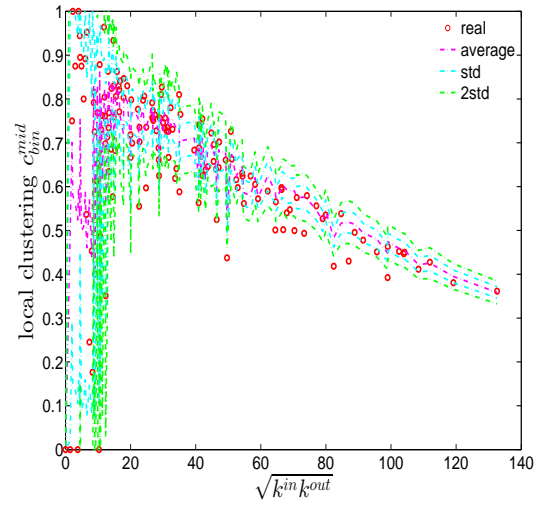


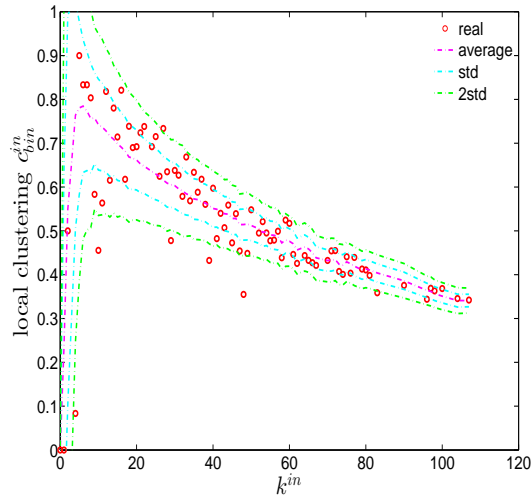
Figure 2.19: Local assortativity in the observed e-MID network and in the DBCM, in Q48. ρ^{in-in} (panel a), ρ^{in-out} (panel b), ρ^{out-in} (panel c), $\rho^{out-out}$ (panel d).



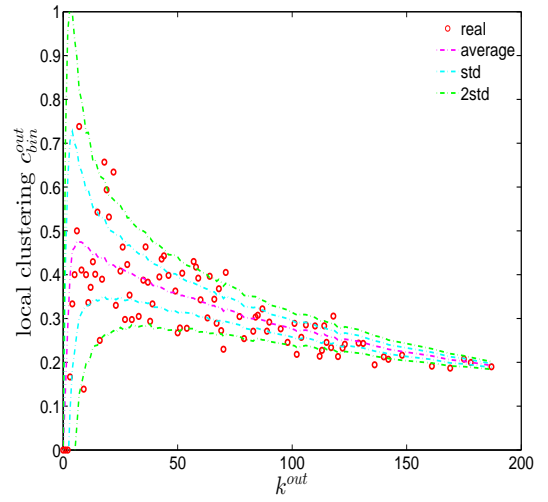
(a) C_{bin}^{cyc} in Q1



(b) C_{bin}^{mid} in Q1



(c) C_{bin}^{in} in Q1



(d) C_{bin}^{out} in Q1

Figure 2.20: Local clustering coefficients C_{bin}^{cyc} (panel a), C_{bin}^{mid} (panel b), C_{bin}^{in} (panel c), C_{bin}^{out} (panel d) in the observed e-MID network and in DBCM, in Q1.

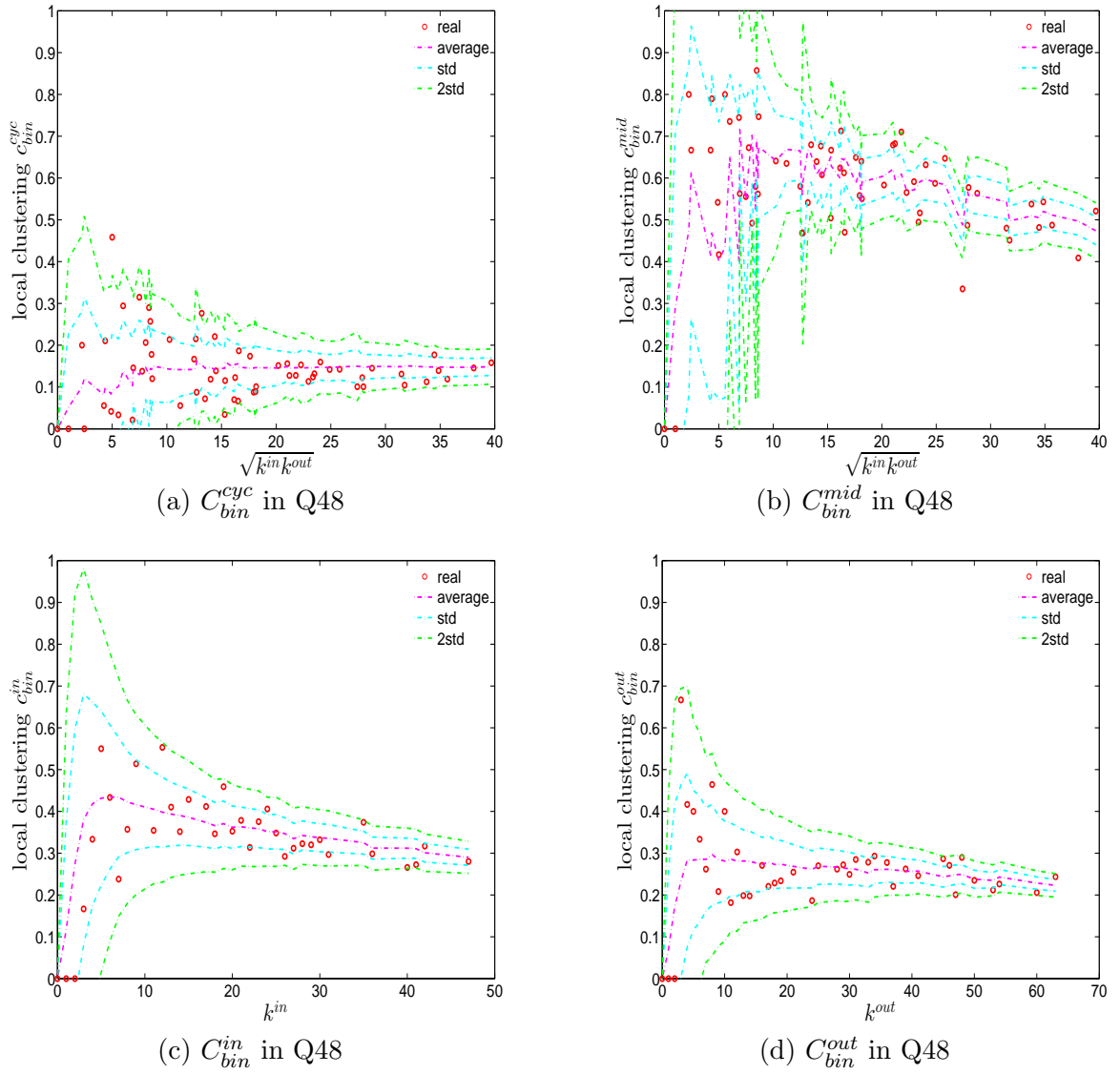


Figure 2.21: Local clustering coefficients C_{bin}^{cyc} (panel a), C_{bin}^{mid} (panel b), C_{bin}^{in} (panel c), C_{bin}^{out} (panel d) in the observed e-MID network and in DBCM, in Q48.

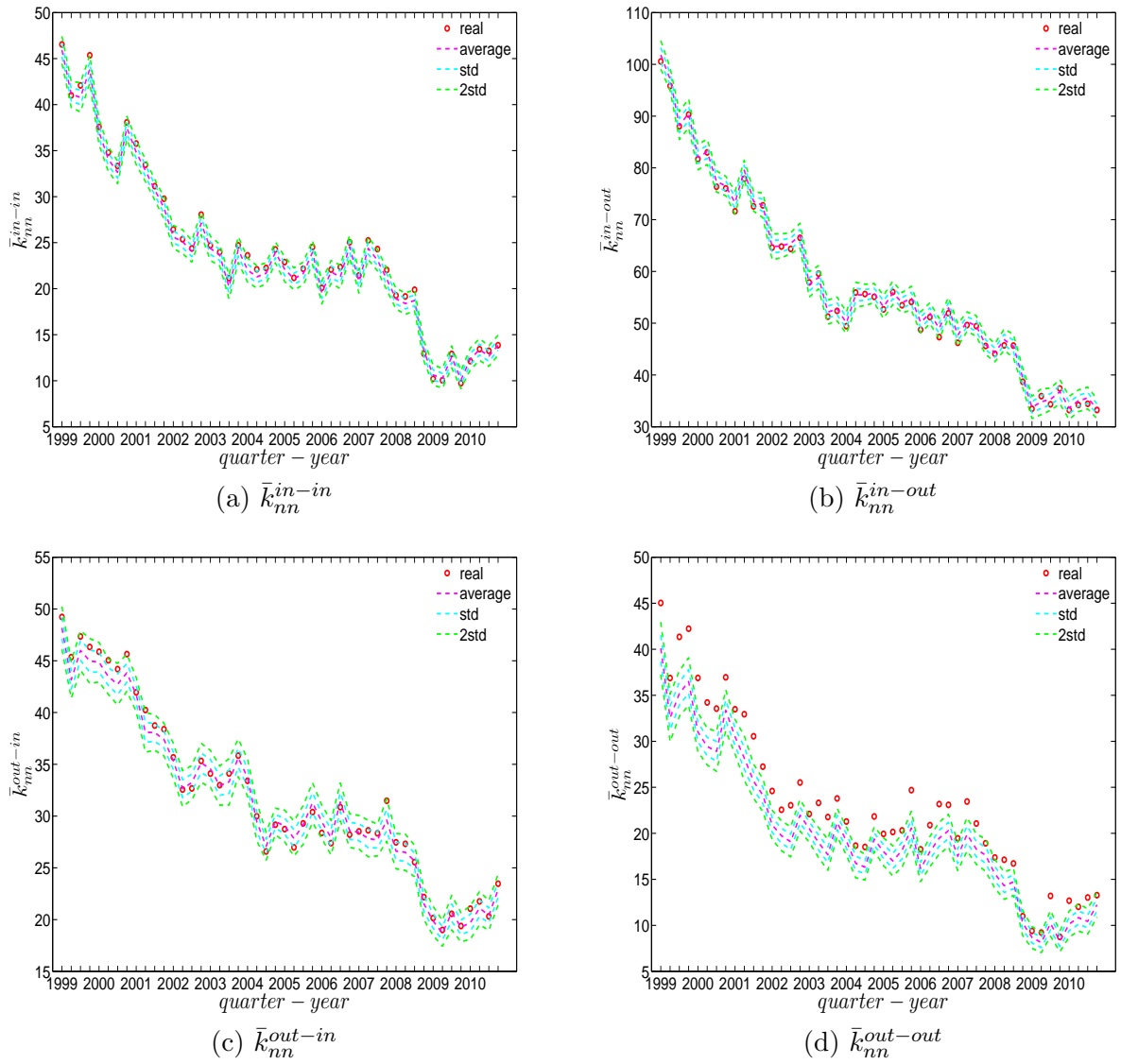


Figure 2.22: Evolution of the averages of ANNDs in the observed e-MID network and in the DBCM. \bar{k}_{nn}^{in-in} (panel a), \bar{k}_{nn}^{in-out} (panel b), \bar{k}_{nn}^{out-in} (panel c), $\bar{k}_{nn}^{out-out}$ (panel d).

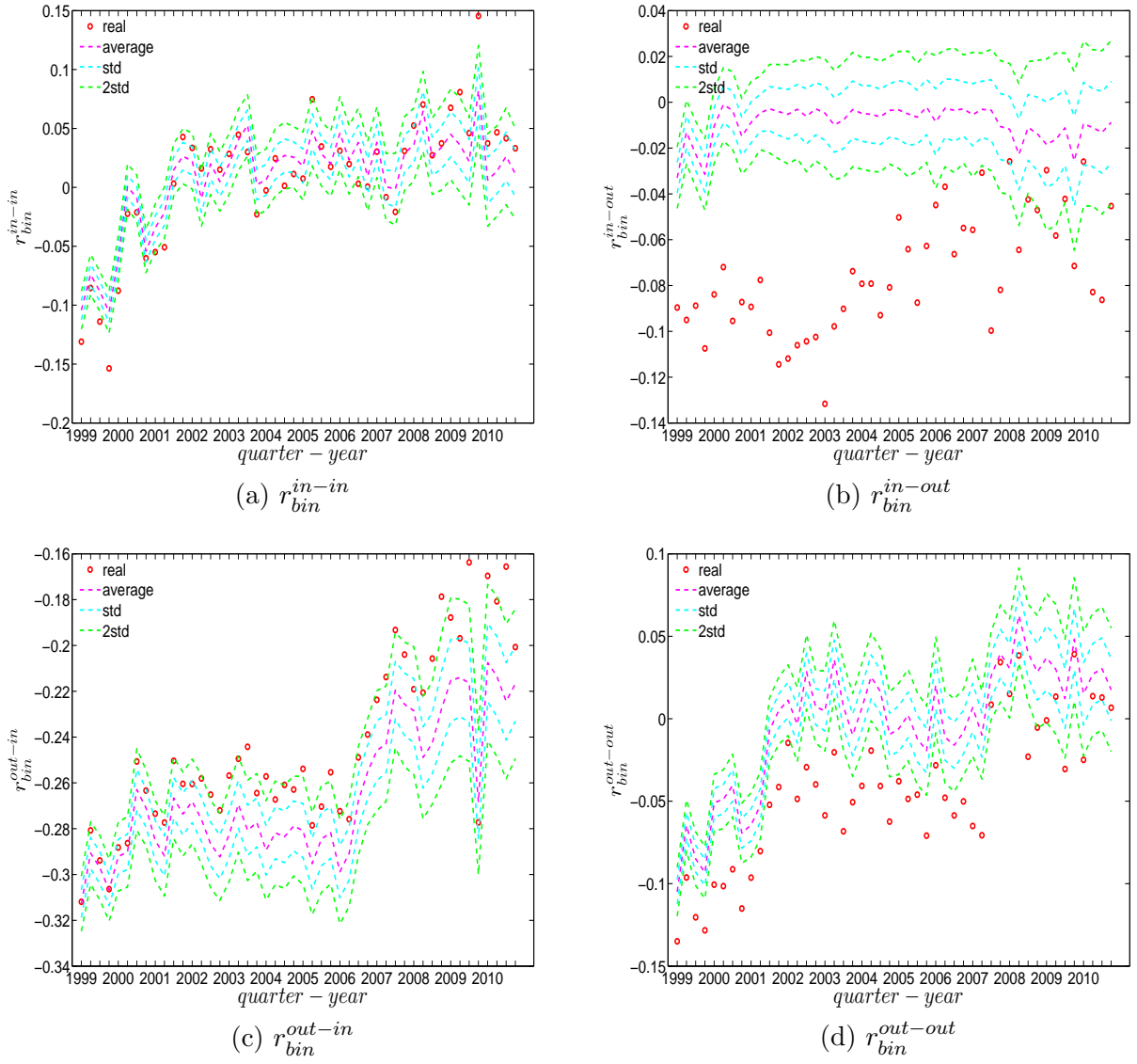


Figure 2.23: Evolution of the global assortativity indicators in the observed e-MID network and in the DBCM. r_{bin}^{in-in} (panel a), r_{bin}^{in-out} (panel b), r_{bin}^{out-in} (panel c), $r_{bin}^{out-out}$ (panel d).

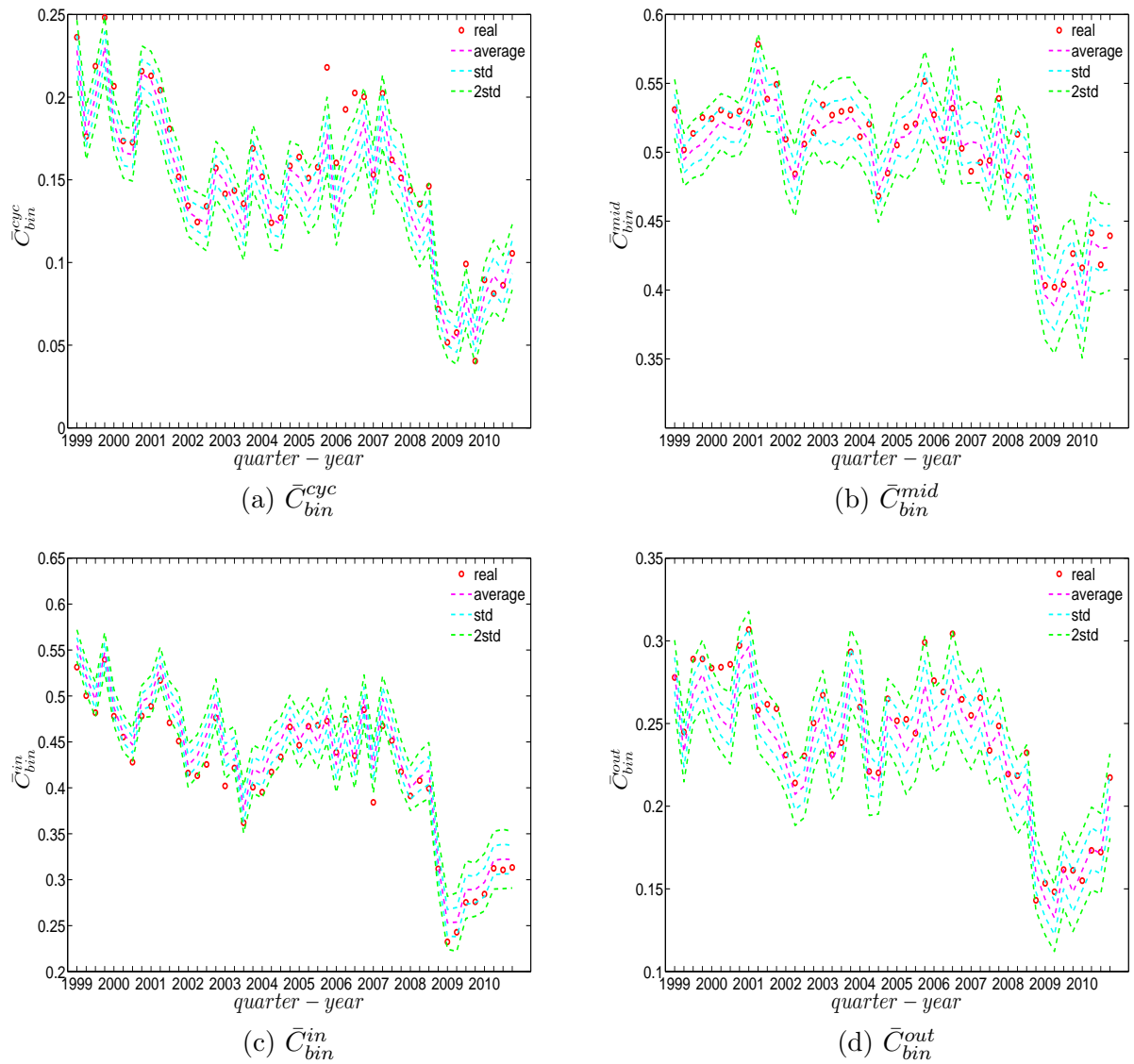


Figure 2.24: Evolution of the averages of clustering coefficients in the observed e-MID network and in the DBCM. \bar{C}_{bin}^{cyc} (panel a), \bar{C}_{bin}^{mid} (panel b), \bar{C}_{bin}^{in} (panel c), \bar{C}_{bin}^{out} (panel d).

2.4 Findings for the weighted network

In the binary version of the observed network, we treat all edges as if they were homogeneous. However, in reality, the capacity and intensity of the relations between banks can be very heterogeneous, consequently the weighted version can have different properties compared to its binary counterpart. In this section, we investigate the structural correlations in the weighted e-MID network. For the sake of simplicity, we do not consider the local weighted assortativity in this section, since breaking down the overall weighted assortativity measure into the contributions of the individual nodes is much more complicated than in the binary case.

Regarding the null models, instead of preserving the observed degree sequence(s) as in the Binary Configure Models (i.e. UBCM, DBCM), first, we employ the weighted configuration model preserving the observed strength sequence(s) (i.e. the UWCM model in the undirected case and the DWCM model in the directed case) and examine whether the chosen null models can replicate the structural correlations in the observed weighted network. As a second step, we consider the enhanced configuration models which maintain both the observed degree as well as strength sequences (i.e. the UECM and the DECM respectively in the undirected and directed cases) and repeat the same exercise.

2.4.1 Structural correlations in the undirected weighted e-MID network

We report the strength dependencies in Figure (2.25) by considering the relationship between s_{nn}^{un} (ANNS) and s^{un} in the first and last quarters. We observe that s_{nn}^{un} is generally a declining function of s^{un} . This relationship is confirmed by the negative value of the global weighted assortativity measure r_w^{un} (see Figure (2.26)). This signals the prevalence of disassortative mixing in the undirected weighted e-MID network, meaning that, in general, high strength banks tend to have relations with low strength banks. Furthermore, it should be emphasized that, in comparison to the undirected binary version of the network, the undirected weighted network exhibits less disassortativity overall, since r_w^{un} is smaller than r_{bin}^{un} in absolute value.

In our analysis of the third order correlations, in contrast to what we discovered in the binary version, we find that, on average the higher strength banks also have higher local clustering coefficients (see Figure (2.27)). This is mainly because the heterogeneity in the transaction volumes across banks in every triangle is now taken into account and

the average transactions of high strength banks are much larger than those of low strength banks. Furthermore, we observe three phases in the evolution of the average of the local weighted clustering coefficients, i.e. before 2002, from 2002 to 2006, and from 2007 onward, which might reflect effects arising from the adoption of the euro as well as from the recent financial crisis (see Figure (2.28)). In particular, we find that \bar{C}_w^{un} is strongly elevated from 2002 to 2006. The same results still hold if we normalize all weights by the total weight average.

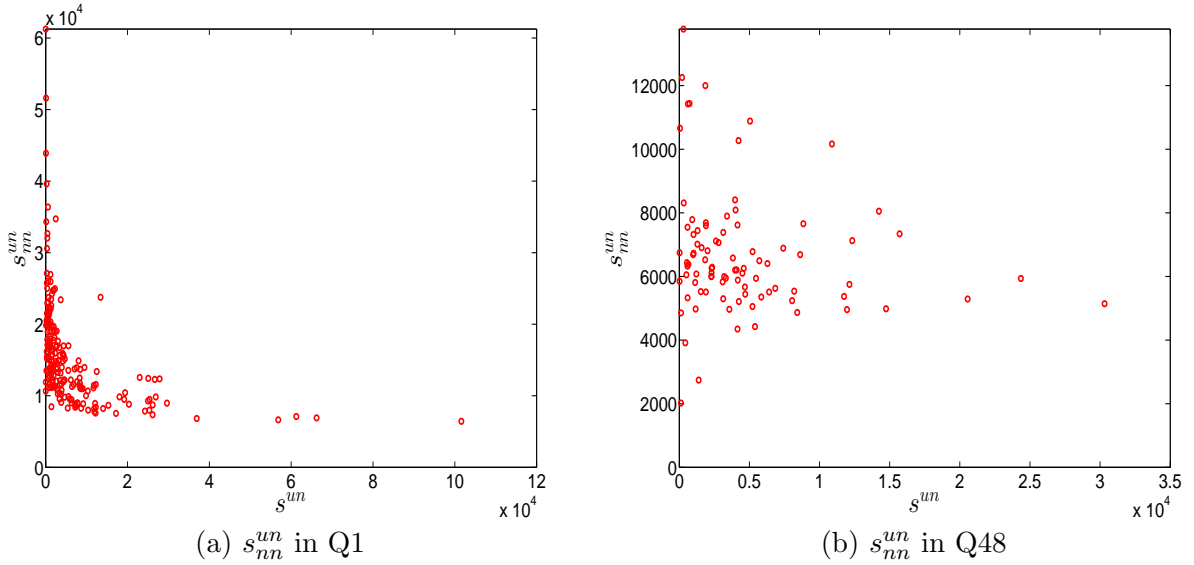


Figure 2.25: ANNS in the undirected weighted e-MID network, in Q1 and Q48.

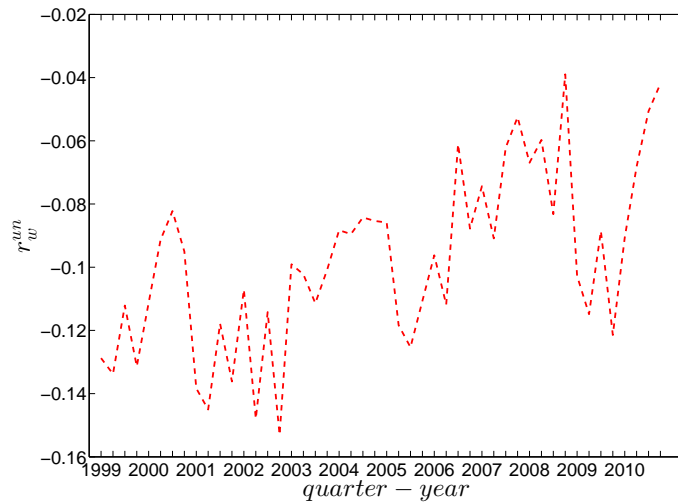


Figure 2.26: Evolution of global weighted assortativity r_w^{un} in the undirected weighted e-MID network.

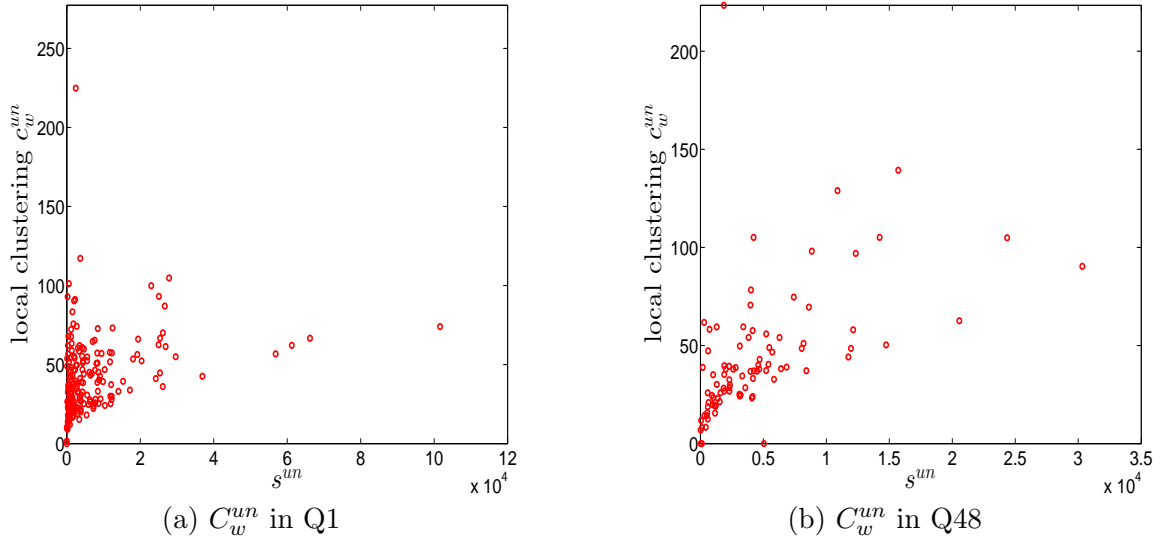


Figure 2.27: Local clustering coefficients C_w^{un} in the undirected weighted e-MID network, in Q1 and Q48.

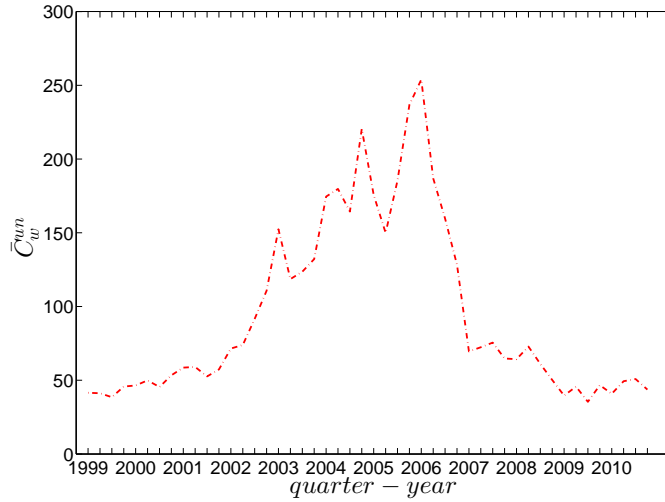


Figure 2.28: Evolution of the average of local weighted clustering coefficients (i.e. \bar{C}_w^{un}) in the undirected weighted e-MID network.

2.4.2 Structural correlations in the directed weighted e-MID network

In the directed weighted version of the e-MID network, to analyzing the structural correlations, we employ average nearest neighbor strength measures for the various mixing

categories $(s_{nn,i}^{in-in}, s_{nn,i}^{in-out}, s_{nn,i}^{out-in}, s_{nn,i}^{out-out})$, global weighted assortativity indicators $(r_w^{in-in}, r_w^{in-out}, r_w^{out-in}, r_w^{out-out})$, and weighted clustering coefficients $(C_w^{cyc}, C_w^{mid}, C_w^{in}, C_w^{out})$.

First, Figures (2.29) and (2.30) show the relationship between the ANNSs and the associated strengths for all four mixing categories in Q1 and Q48. Over time, while in the first quarters, the ANNSs are a declining function of the associated strengths (after certain truncated values of the associated strengths), in many later quarters these relationships break down, especially for the mixing categories in-in, in-out, and out-out. To obtain the overall level of strength dependency of bank interactions for each mixing category, we calculate the global assortativity indicators $r_w^{in-in}, r_w^{in-out}, r_w^{out-in}, r_w^{out-out}$, and show their evolution over time in Figure (2.31). The results indicate that, while the out-in mixing is disassortative for the most part, the other three categories do not seem to exhibit a distinct mixing nature. In comparison to the directed binary version, the absolute values of $r_w^{in-in}, r_w^{in-out}, r_w^{out-in}, r_w^{out-out}$ are smaller than those of $r_{bin}^{in-in}, r_{bin}^{in-out}, r_{bin}^{out-in}, r_{bin}^{out-out}$. An interesting observation is that, among the four mixing categories, the weighted assortativity in the out-in category is closest to the undirected weighted assortativity, i.e. $r_w^{un} \sim r_w^{out-in}$. For the binary versions of the network, when comparing the mixing patterns in the directed and undirected case, we made the same observation.

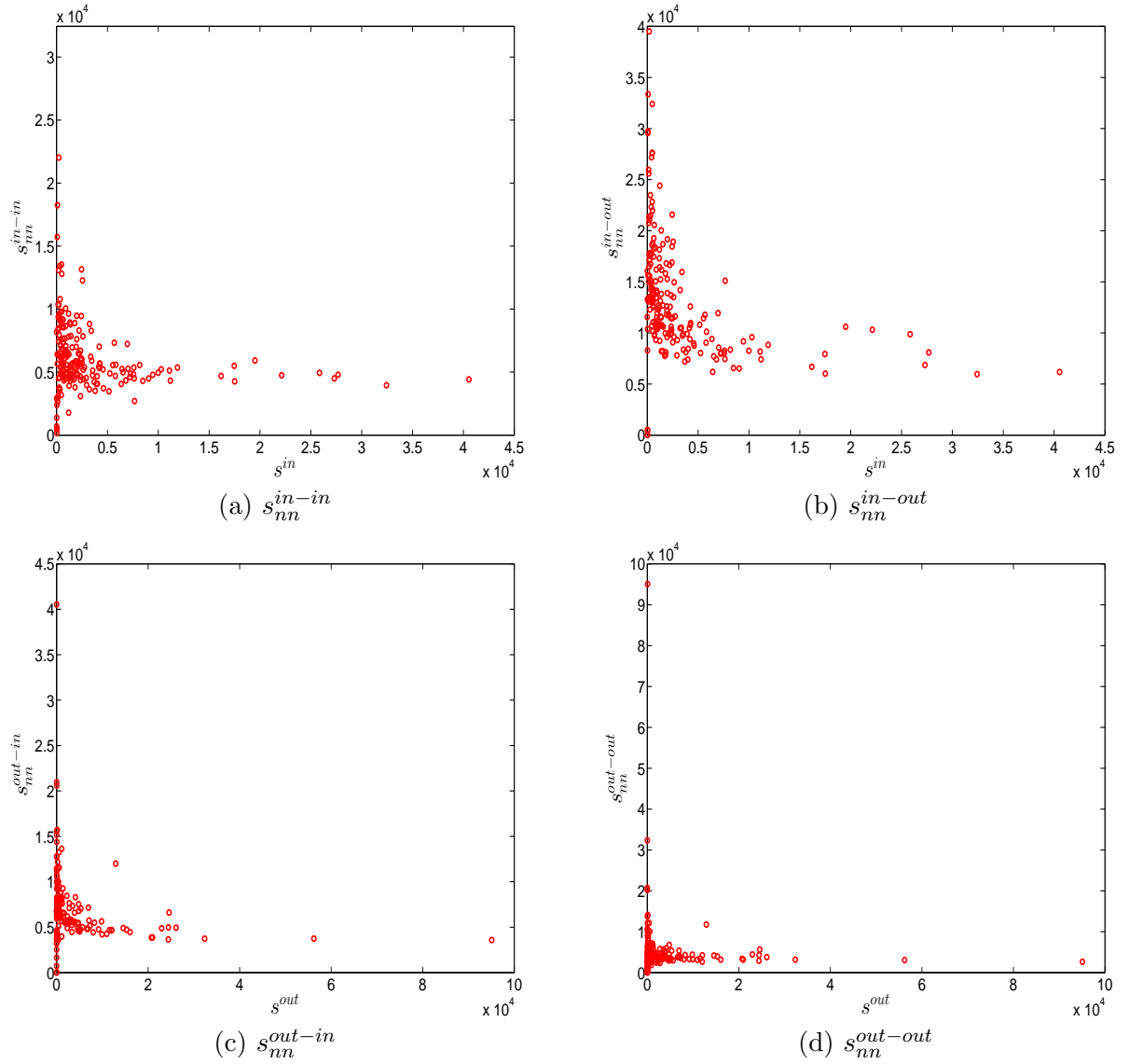


Figure 2.29: ANNs in the directed weighted e-MID network, in Q1. s_{nn}^{in-in} (panel a), s_{nn}^{in-out} (panel b), s_{nn}^{out-in} (panel c), $s_{nn}^{out-out}$ (panel d).

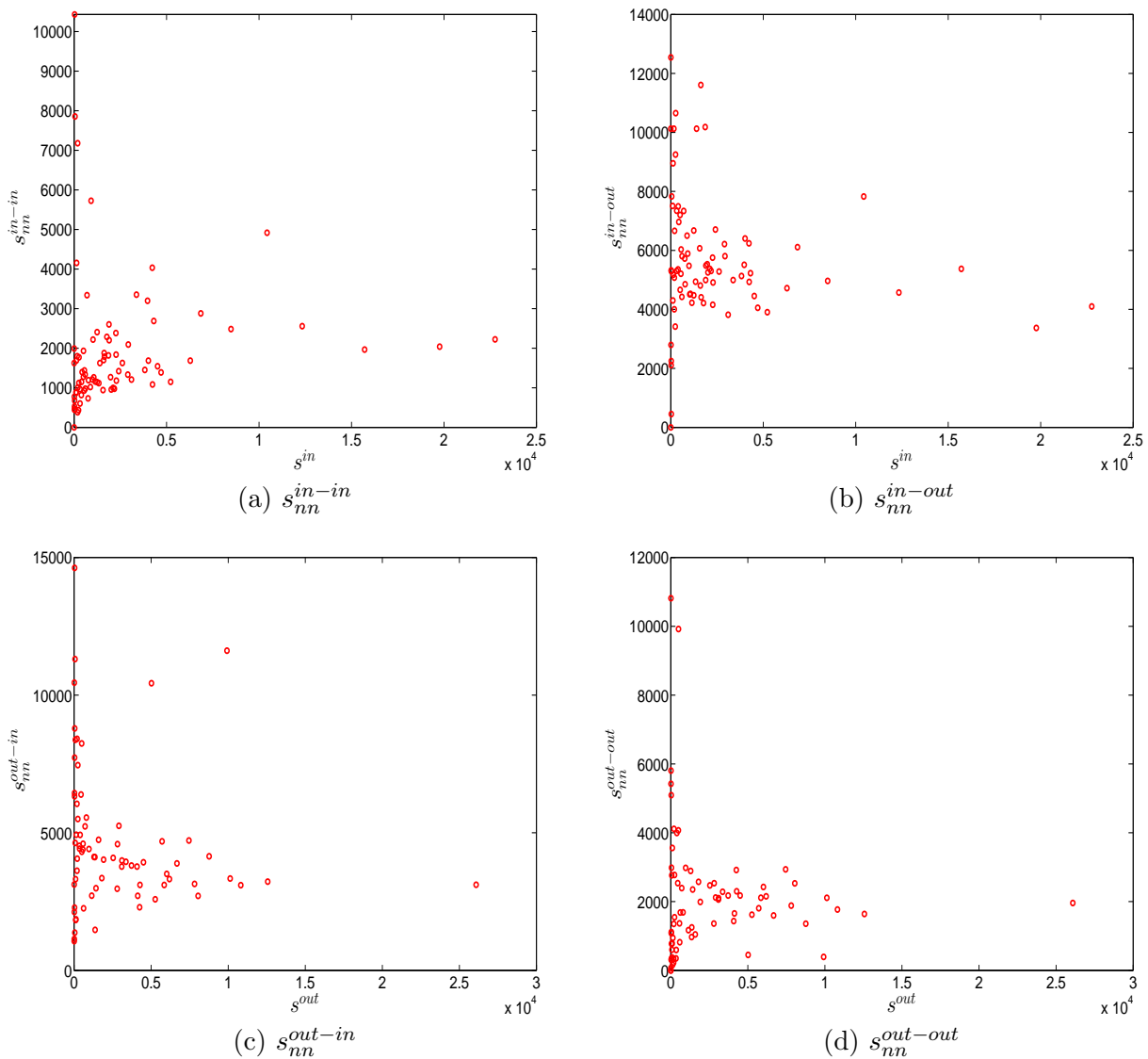


Figure 2.30: ANNSs in the directed weighted e-MID network, in Q48. s_{nn}^{in-in} (panel a), s_{nn}^{in-out} (panel b), s_{nn}^{out-in} (panel c), $s_{nn}^{out-out}$ (panel d).

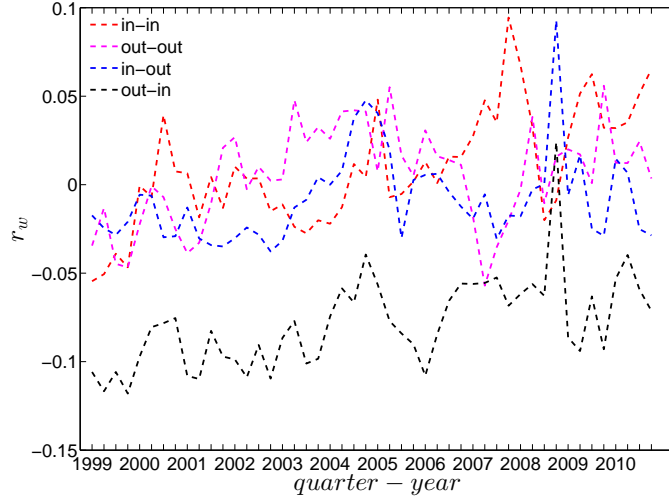


Figure 2.31: Evolution of the directed weighted assortativity indicators, i.e. r_w^{in-in} , r_w^{in-out} , r_w^{out-in} , and $r_w^{out-out}$ in the directed weighted e-MID network.

Second, the local weighted clustering coefficients for the four clustering types C_w^{cyc} , C_w^{mid} , C_w^{in} , C_w^{out} are plotted against the associated strengths in Figures (2.32) and (2.33) ⁶. We observe that, generally, higher (lower) strengths correspond to higher (lower) local weighted clustering coefficients.

The evolution of the averages of the local weighted clustering coefficients also exhibits three different phases, i.e. before 2002, from 2002 to 2006, and from 2007 onward. For all types of clustering, the averages in the period from 2002 to 2006 are higher than those in the other two periods. Recall that, on average, larger values of \bar{C}_w^{mid} , \bar{C}_w^{in} imply higher systemic risk, while larger values of \bar{C}_w^{out} reveal the high exposure of the associated bank to risk. The order and magnitude of different combinations of \bar{C}_w shown in Figure (2.34) thus reveal the importance of both types of risk in the period from 2002 to 2006 in the weighted version of the network. It should be emphasized that, even when all weights are normalized by the average weight over the whole network, we still observe a similar trend, signaling that the evolution of the averages of the directed local weighted clustering coefficients is not only driven by changes in the overall transaction volume (overall strength of the interactions) but also by changes in the frequency of aforementioned tripartite relations among banks.

⁶In the cases of C_w^{cyc} and C_w^{mid} , we plot them against $s^{in-out} = \sqrt{s^{in}s^{out}}$.

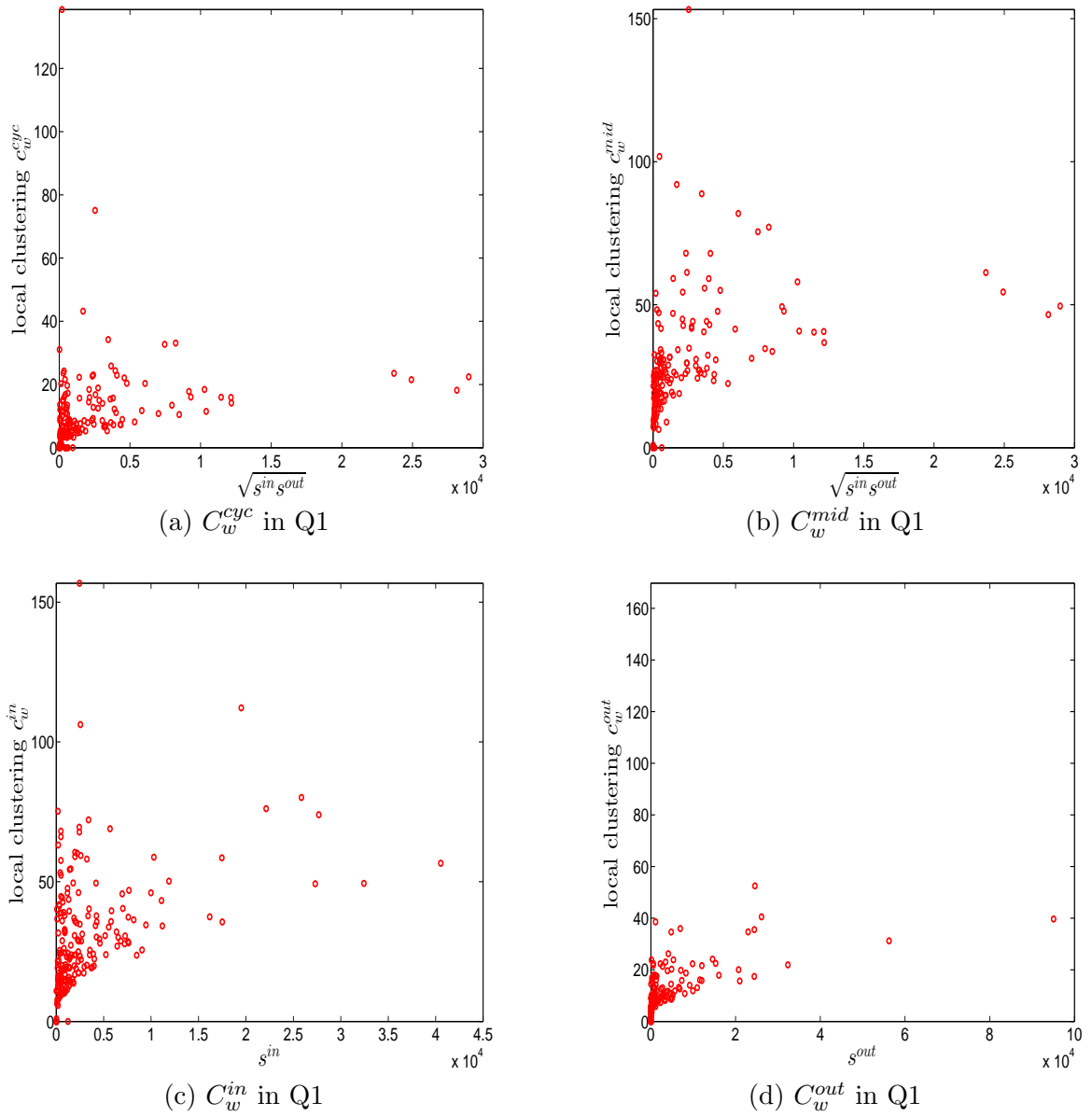
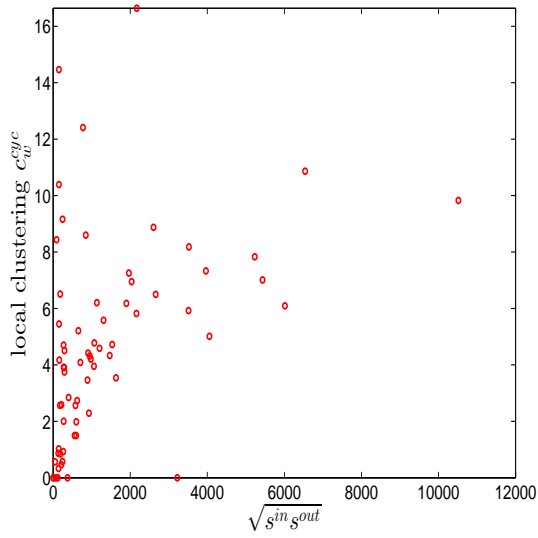
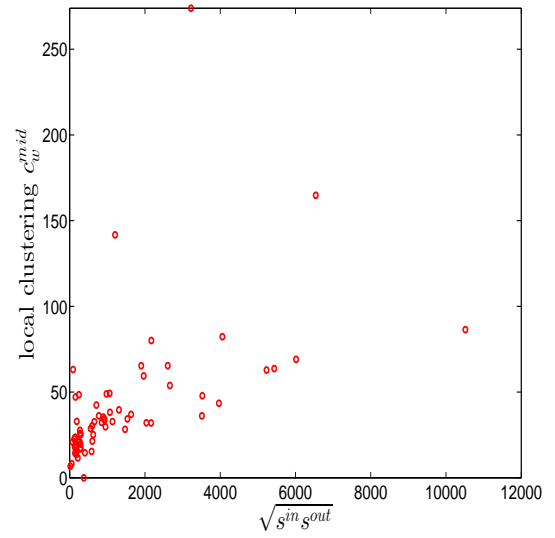


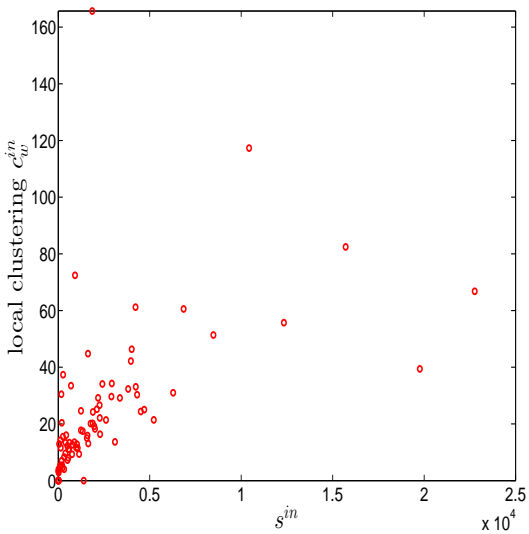
Figure 2.32: Local weighted clustering coefficients in the directed weighted e-MID network, in Q1. C_w^{cyc} (panel a), C_w^{mid} (panel b), C_w^{in} (panel c), C_w^{out} (panel d).



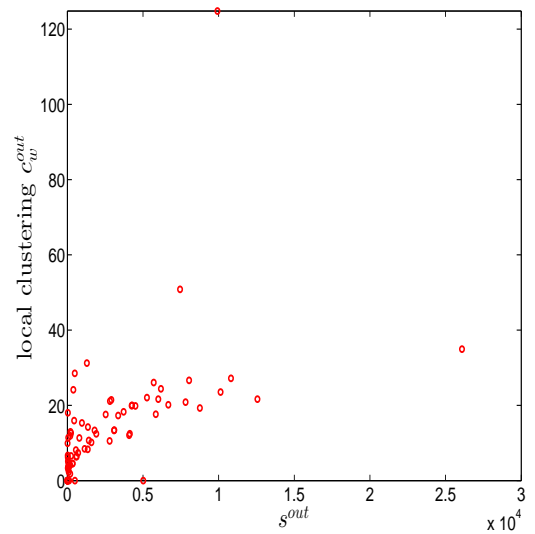
(a) C_w^{cyc} in Q48



(b) C_w^{mid} in Q48



(c) C_w^{in} in Q48



(d) C_w^{out} in Q48

Figure 2.33: Local weighted clustering coefficients in the directed weighted e-MID network, in Q48. C_w^{cyc} (panel a), C_w^{mid} (panel b), C_w^{in} (panel c), C_w^{out} (panel d).

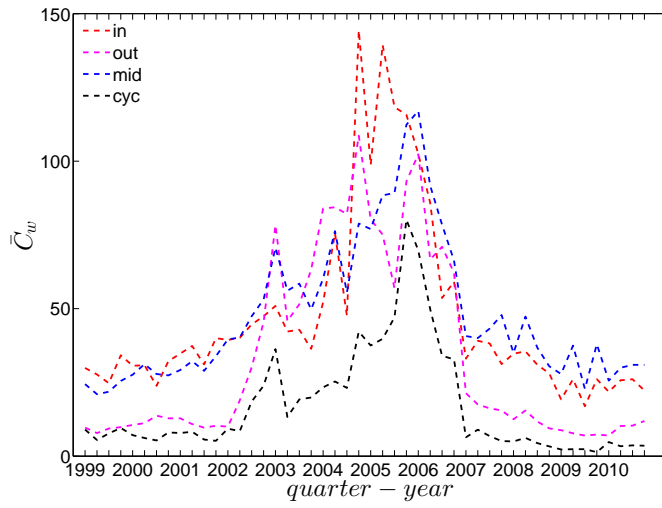


Figure 2.34: Evolution of the averages of local weighted clustering coefficients, i.e. \bar{C}_w^{cyc} , \bar{C}_w^{mid} , \bar{C}_w^{in} , and \bar{C}_w^{out} in the directed weighted e-MID network.

2.4.3 Comparisons to the weighted configuration models

Undirected Weighted Network

To examine the role of the heterogeneity in the local constraints for the emergence of higher order structural correlations in the weighted version of the observed network, we employ the UWCM, which preserves the observed strength sequence, and the UECM, which enforces both the observed degree as well as strength sequences on to the randomized ensemble.

First, the observed values of the measure ANNS as well as of the local weighted clustering coefficients (as can be seen in Figures (2.35) and (2.36)) strongly deviate from their respective expectations under the UWCM. In contrast, we find that the UECM model is able to reproduce the main features of such measures (see Figures (2.37) and (2.38)).

For a more detailed comparison between the two models, we compare the z-scores of the measure ANNS as well as of the local weighted clustering coefficients evaluated under the UWCM with those for the same measures evaluated under the UECM (see subsection B of the Appendix for a more detailed explanation). More specifically, for every bank i , we define the z-scores

$$z_{\text{ANNS}}^{\text{UWCM}}(i) = \frac{\text{ANNS}(i) - \langle \text{ANNS}(i) \rangle_{\text{UWCM}}}{\sigma[\text{ANNS}(i)]_{\text{UWCM}}}, \quad (2.88)$$

and

$$z_{\text{ANNS}}^{\text{UECM}}(i) = \frac{\text{ANNS}(i) - \langle \text{ANNS}(i) \rangle_{\text{UECM}}}{\sigma[\text{ANNS}(i)]_{\text{UECM}}}, \quad (2.89)$$

where $\langle \text{ANNS}(i) \rangle_{\text{UWCM}}$ and $\langle \text{ANNS}(i) \rangle_{\text{UECM}}$ are respectively the expected values of the measure ANNS for bank i evaluated under the UWCM and the UECM; and $\sigma(\text{ANNS}(i))_{\text{UWCM}}$ and $\sigma(\text{ANNS}(i))_{\text{UECM}}$ are respectively the standard deviations of ANNS(i) evaluated under the UWCM and the UECM ⁷.

Similarly, the z-scores for the local weighted clustering coefficients for bank i evaluated under the UWCM and the UECM are defined as

$$z_{C_w}^{\text{UWCM}}(i) = \frac{C_w^{\text{un}}(i) - \langle C_w^{\text{un}}(i) \rangle_{\text{UWCM}}}{\sigma[C_w^{\text{un}}(i)]_{\text{UWCM}}}, \quad (2.90)$$

and

$$z_{C_w}^{\text{UECM}}(i) = \frac{C_w^{\text{un}}(i) - \langle C_w^{\text{un}}(i) \rangle_{\text{UECM}}}{\sigma[C_w^{\text{un}}(i)]_{\text{UECM}}}. \quad (2.91)$$

We show the comparisons between the z-scores under the two considered configuration

⁷Throughout this paper, the notation $\langle X \rangle_{\text{null model}}$ and $\sigma[X]_{\text{null model}}$ are respectively the expected value and standard deviation of X evaluated under the referenced null model.

models in Figure (2.39) and Figure (2.40). For almost all banks, we find that $|z_{\text{ANNS}}^{\text{UECM}}| < |z_{\text{ANNS}}^{\text{UWCM}}|$ and $|z_{C_w^{un}}^{\text{UECM}}| < |z_{C_w^{un}}^{\text{UWCM}}|$.

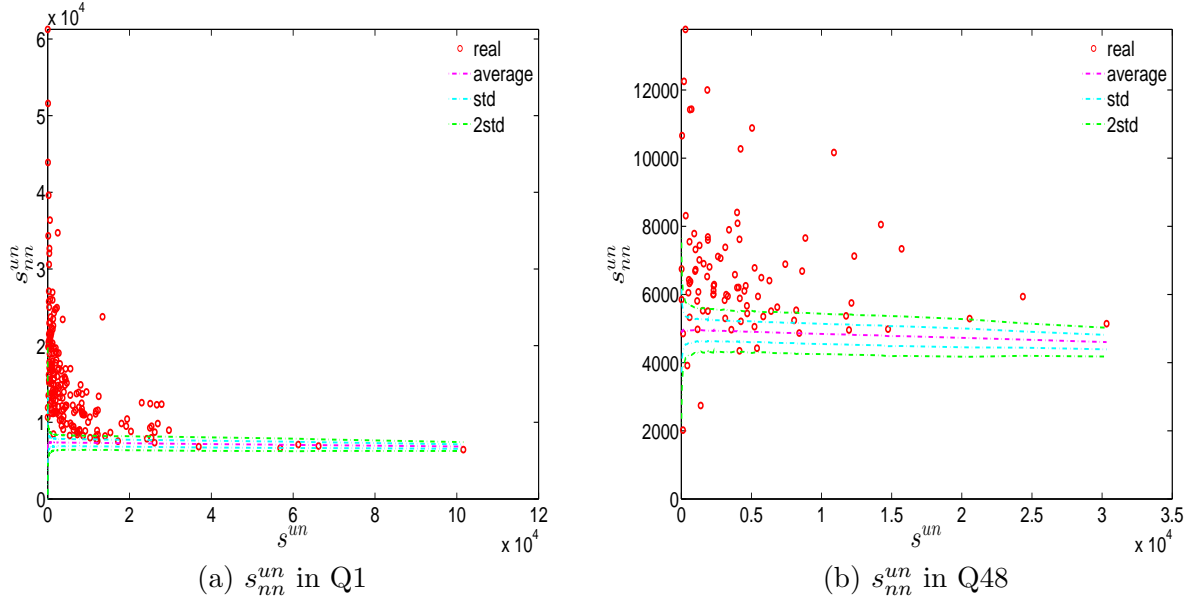


Figure 2.35: ANNS in the observed e-MID network and in the UWCM, in Q1 and Q48.

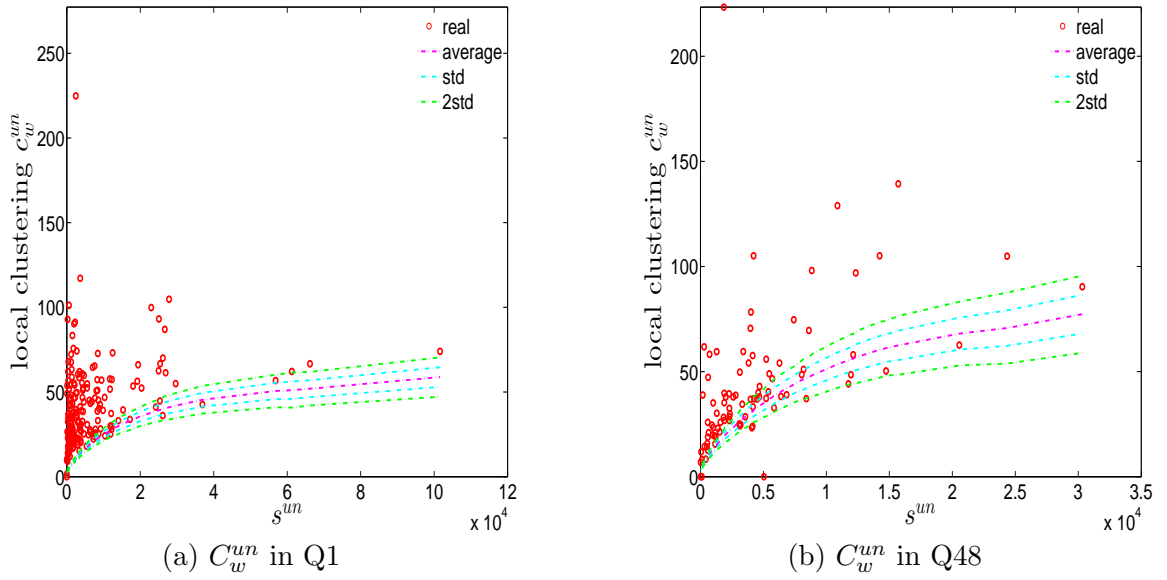


Figure 2.36: Local weighted clustering coefficients C_w^{un} in the observed e-MID network and in the UWCM, in Q1 and Q48.

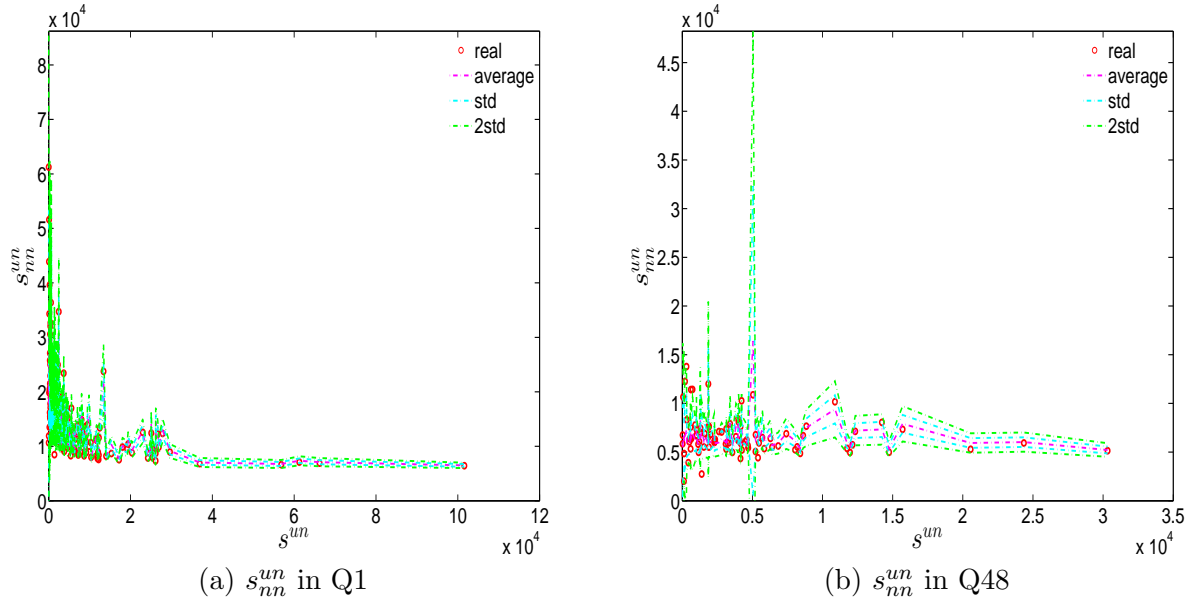


Figure 2.37: ANNS in the observed e-MID network and in the UECM, in Q1 and Q48.

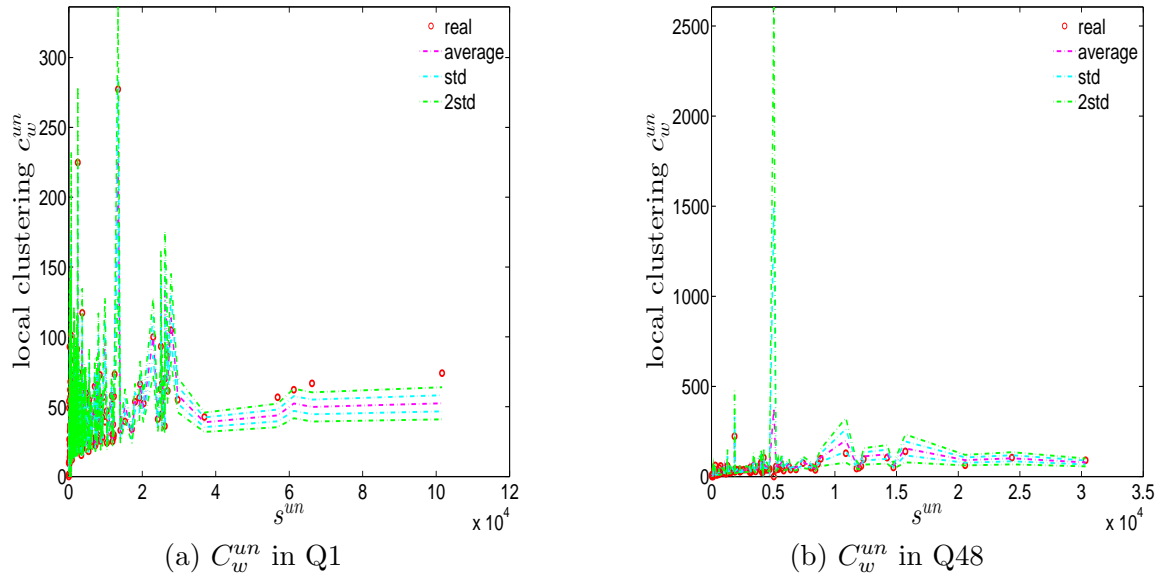


Figure 2.38: Local weighted clustering coefficients C_w^{un} in the observed e-MID network and in the UECM, in Q1 and Q48.

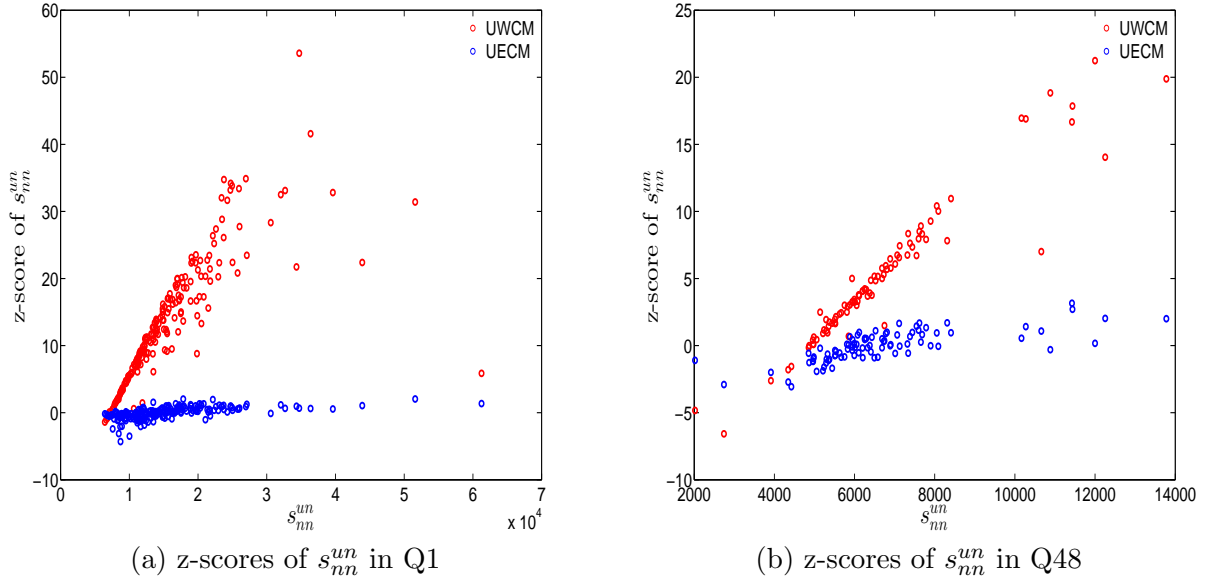


Figure 2.39: z-scores of s_{nn}^{un} vs. s_{nn}^{un} in the UWCM and the UECM, in Q1 and in Q48. Panel (a) for z-scores of s_{nn}^{un} in Q1, panel (b) for z-scores of s_{nn}^{un} in Q48.

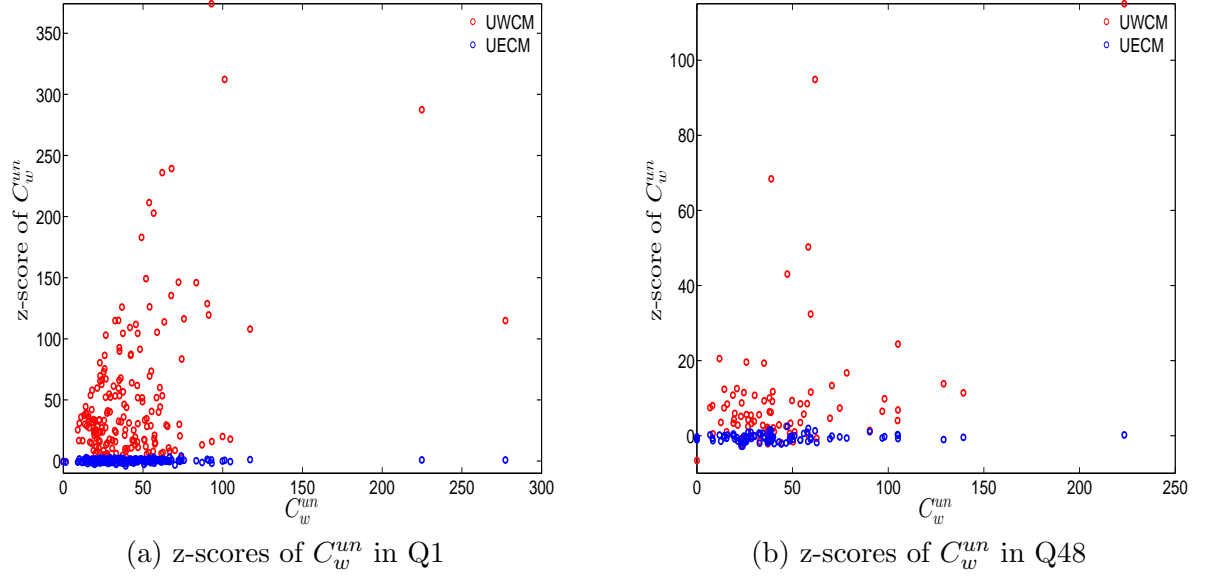


Figure 2.40: z-scores of C_w^{run} vs. C_w^{un} in the UWCM and the UECM. Panel (a) for z-scores of C_w^{run} in Q1, panel (b) for z-scores of C_w^{run} in Q48.

We now compare the evolution of \bar{s}_{nn}^{un} , r_w^{un} , and \bar{C}_w^{run} for the observed network with the one obtained for these measures under the UWCM and the UECM. In Figure (2.41), we see that for most of the time, the observed values of \bar{s}_{nn}^{un} , r_w^{un} , and \bar{C}_w^{run} lie outside the ± 2 bands associated with the UWCM. In contrast, in Figure (2.42), we see that the evolution

of these measures is well captured by the ECM. The observed values of \bar{s}_{nn}^{un} and \bar{C}_w^{un} and the expected ones obtained from the ECM are in very close agreement. Even in the case of r_w^{un} , for which several significant deviations are found, the main features of its evolution are well reproduced by the ECM.

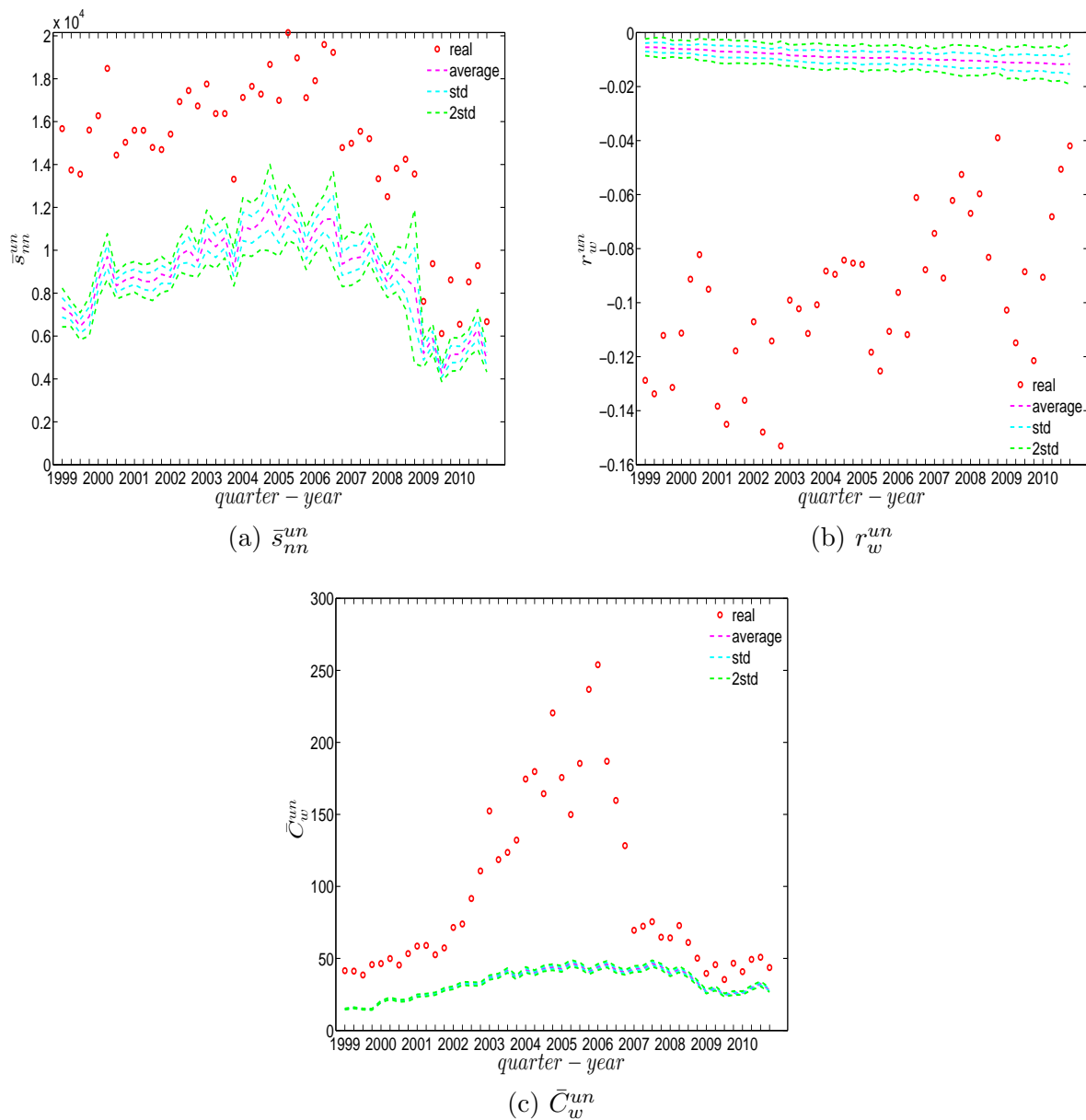


Figure 2.41: Evolution of \bar{s}_{nn}^{un} (panel a), r_w^{un} (panel b), and \bar{C}_w^{un} (panel c) in the observed e-MID network and in the UWCM.

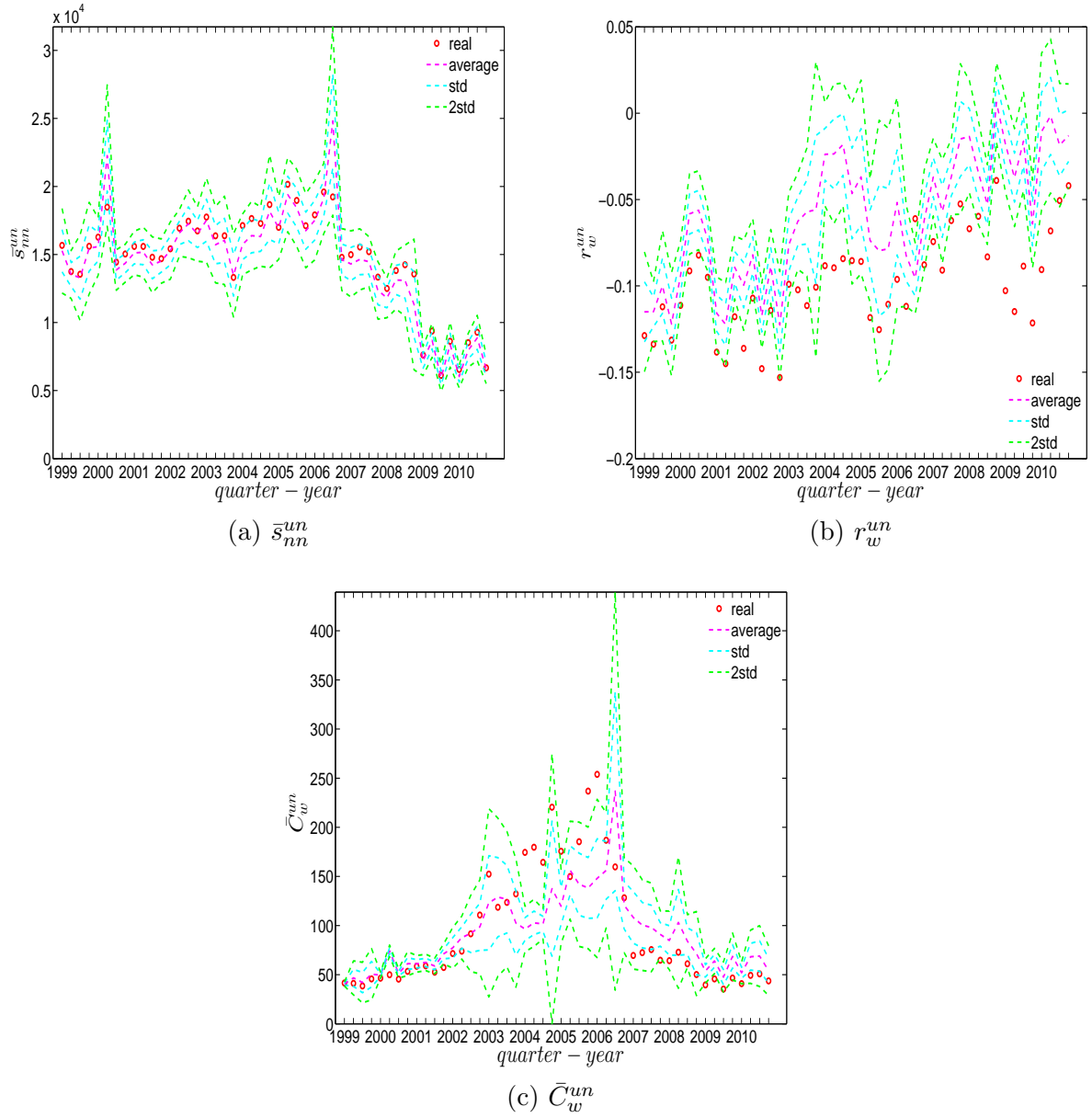


Figure 2.42: Evolution of \bar{s}_{nn}^{un} (panel a), r_w^{un} (panel b), and \bar{C}_w^{un} (panel c) in the observed e-MID network and in the UECM.

Directed Weighted Network

We now extend our comparison between the observed network and the configuration models to the directed weighted version. For this purpose, two null models are employed, i.e. the DWCM and the DECM.

First, regarding the directed versions of the measure ANNS, we compare s_{nn}^{in-in} , s_{nn}^{in-out} , s_{nn}^{out-in} , and $s_{nn}^{out-out}$ of the observed network in the two chosen quarters with those obtained from the DWCM in Figures (2.43) and (2.44), and with those obtained from the DECM

in Figures (2.45) and (2.46). Similar to the undirected weighted case, the z-scores of the directed weighted versions of the measure ANNS evaluated under these two models are also reported in Figures (2.47) and (2.48). Overall, we see that the main features of the measure ANNS are replicated much better by the DECM than by the DWCM. Furthermore, typically for almost all banks, we find that $|z_{\text{ANNS}}^{\text{DECM}}| < |z_{\text{ANNS}}^{\text{DWCM}}|$.

In terms of the third order structural correlations, the DECM again outperforms the DWCM in terms of reproducing the main features of local weighted clustering coefficients. This is visualized in Figures (2.49), (2.50), (2.51), and (2.52). In addition, for each type of local weighted clustering coefficients we also calculate the z-scores evaluated under the DWCM and the DECM. As shown in Figures (2.53) and (2.54), we typically observe that $|z_{C_w^{\text{cyc}}}^{\text{DECM}}| < |z_{C_w^{\text{cyc}}}^{\text{DWCM}}|$, $|z_{C_w^{\text{mid}}}^{\text{DECM}}| < |z_{C_w^{\text{mid}}}^{\text{DWCM}}|$, $|z_{C_w^{\text{in}}}^{\text{DECM}}| < |z_{C_w^{\text{in}}}^{\text{DWCM}}|$, and $|z_{C_w^{\text{out}}}^{\text{DECM}}| < |z_{C_w^{\text{out}}}^{\text{DWCM}}|$.

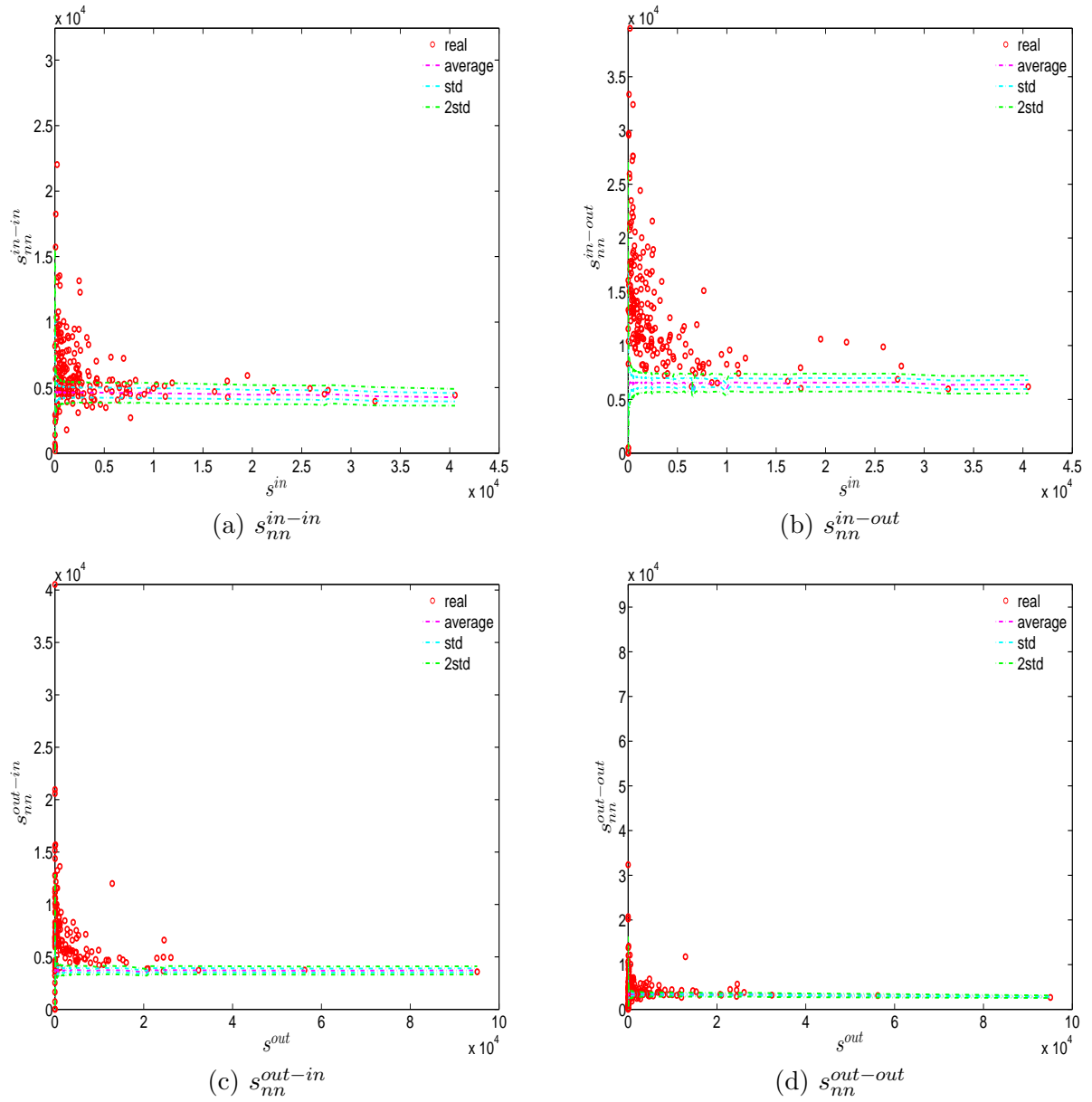


Figure 2.43: ANNs in the observed e-MID network and in the DWCM, in Q1. s_{nn}^{in-in} (panel a), s_{nn}^{in-out} (panel b), s_{nn}^{out-in} (panel c), $s_{nn}^{out-out}$ (panel d).

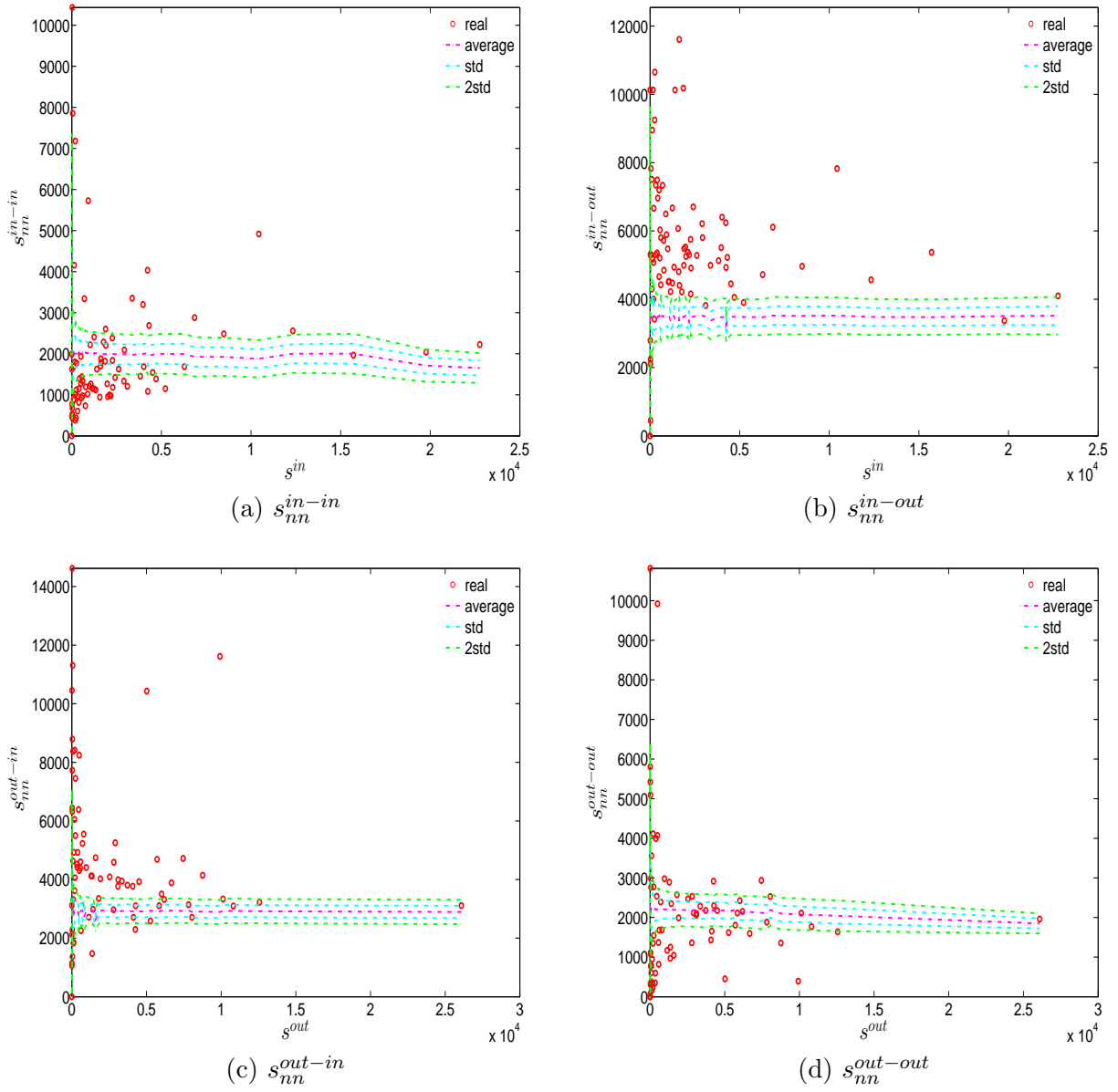


Figure 2.44: ANNs in the observed e-MID network and in the DWCM, in Q48. s_{nn}^{in-in} (panel a), s_{nn}^{in-out} (panel b), s_{nn}^{out-in} (panel c), $s_{nn}^{out-out}$ (panel d).

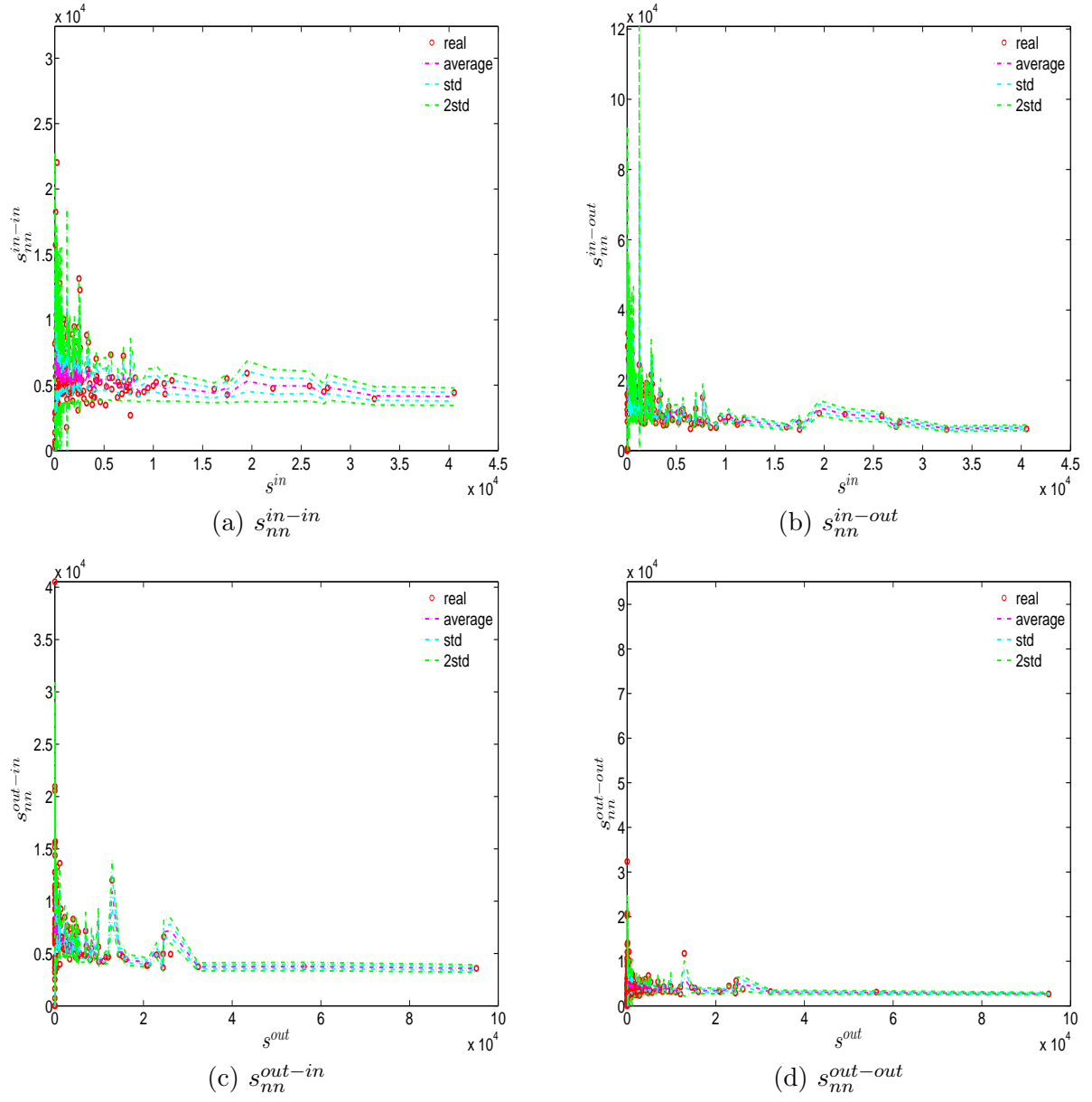


Figure 2.45: ANNSs in the observed e-MID network and in the DECM, in Q1. s_{nn}^{in-in} (panel a), s_{nn}^{in-out} (panel b), s_{nn}^{out-in} (panel c), $s_{nn}^{out-out}$ (panel d).

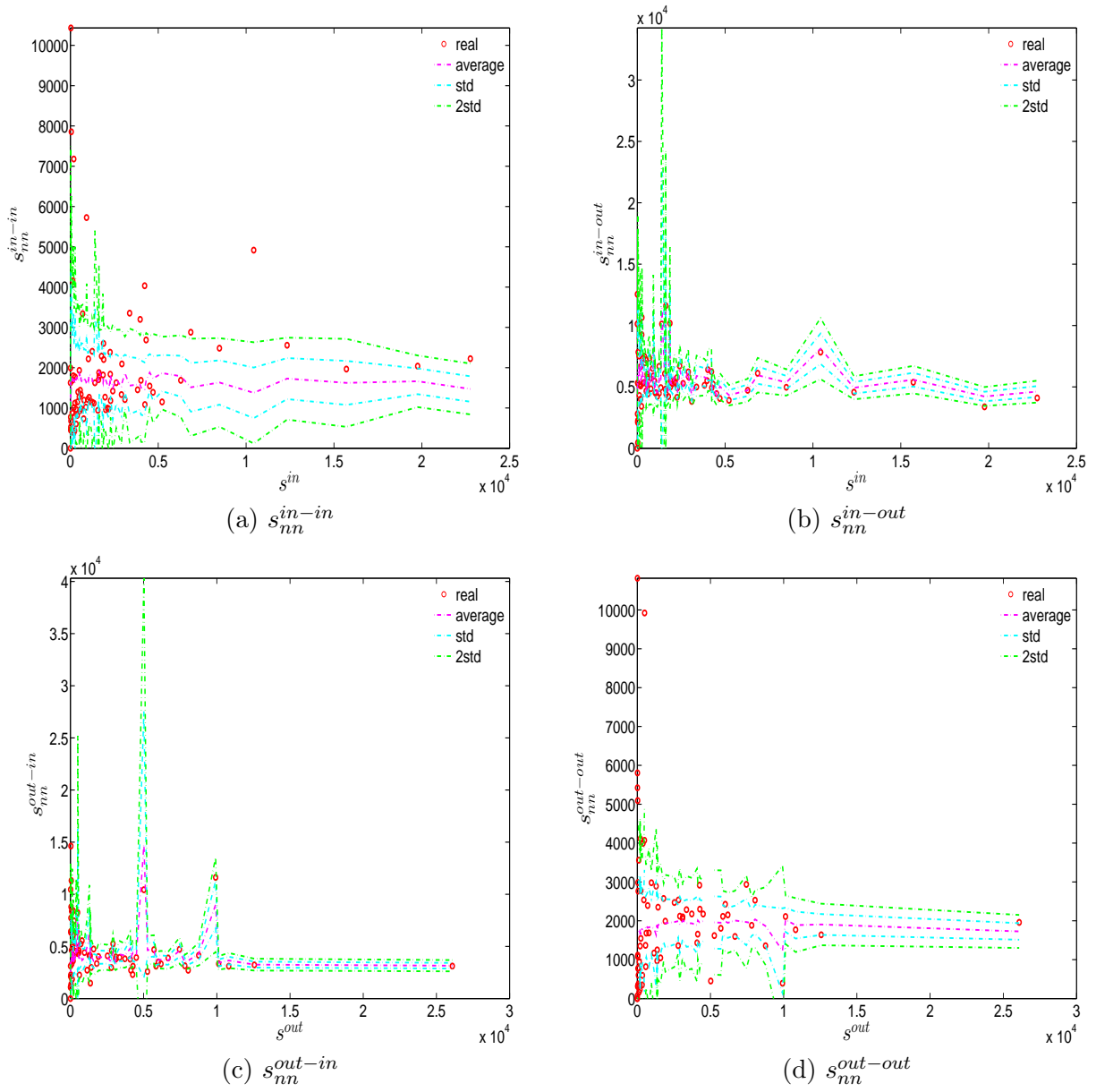


Figure 2.46: ANNs in the observed e-MID network and in the DECM, in Q48. s_{nn}^{in-in} (panel a), s_{nn}^{in-out} (panel b), s_{nn}^{out-in} (panel c), $s_{nn}^{out-out}$ (panel d).

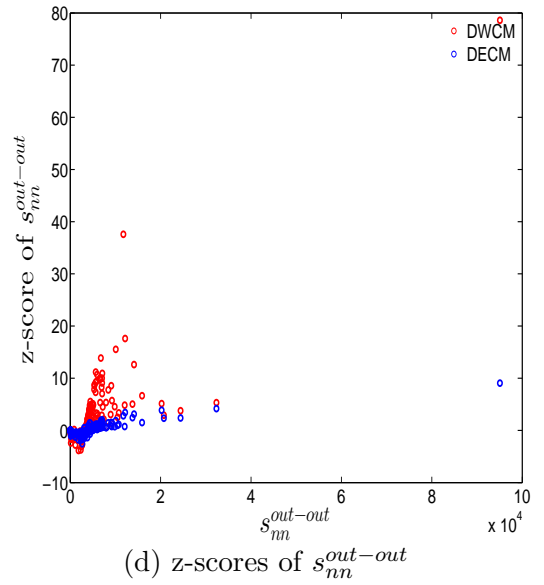
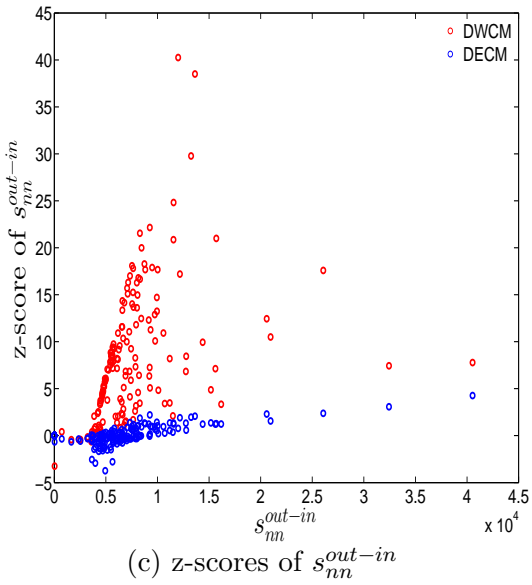
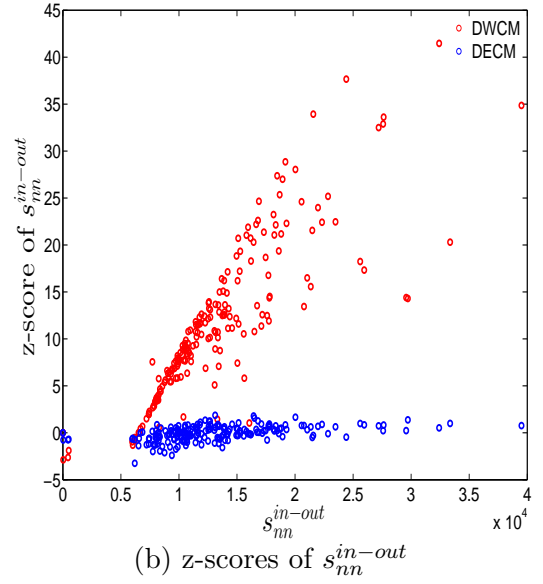
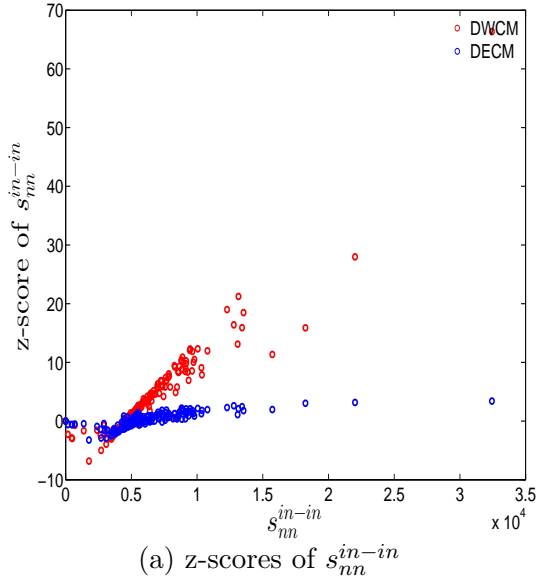
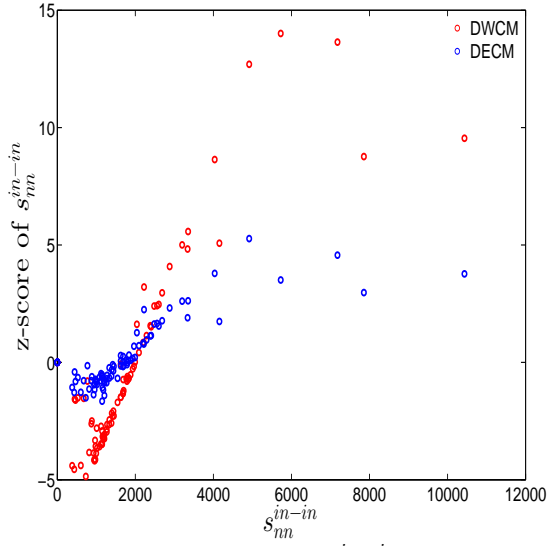
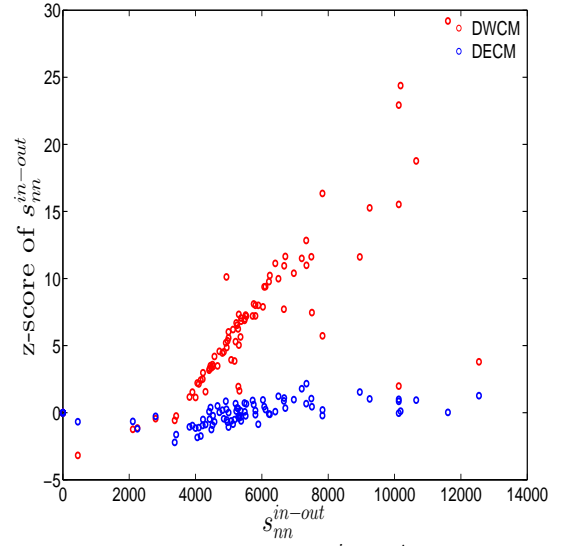


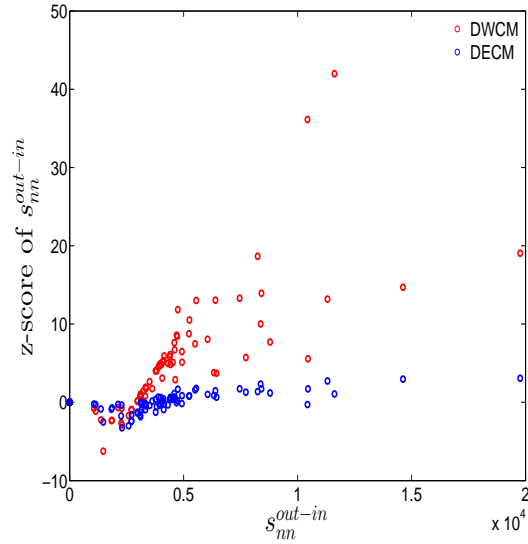
Figure 2.47: z-scores of ANNs vs. ANNs, in the DWCM and DECM models, in Q1. Panels (a) for s_{nn}^{in-in} , (b) for s_{nn}^{in-out} , (c) for s_{nn}^{out-in} , (d) for $s_{nn}^{out-out}$.



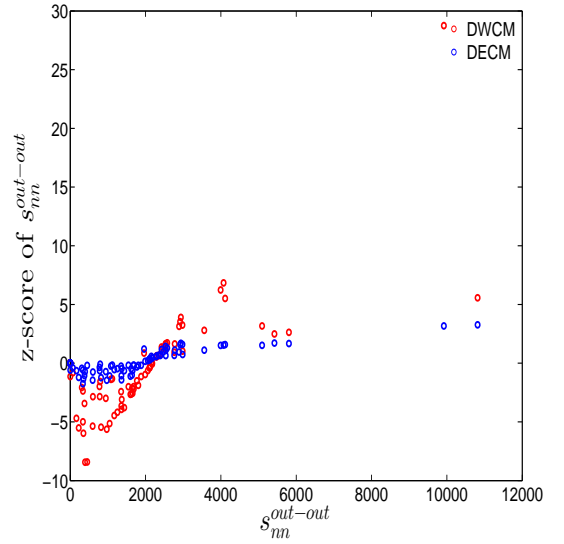
(a) z-scores of s_{nn}^{in-in}



(b) z-scores of s_{nn}^{in-out}



(c) z-scores of s_{nn}^{out-in}



(d) z-scores of $s_{nn}^{out-out}$

Figure 2.48: z-scores of ANNs vs. ANNs, in the DWCM and DECM models, in Q48. Panels (a) for s_{nn}^{in-in} , (b) for s_{nn}^{in-out} , (c) for s_{nn}^{out-in} , (d) for $s_{nn}^{out-out}$.

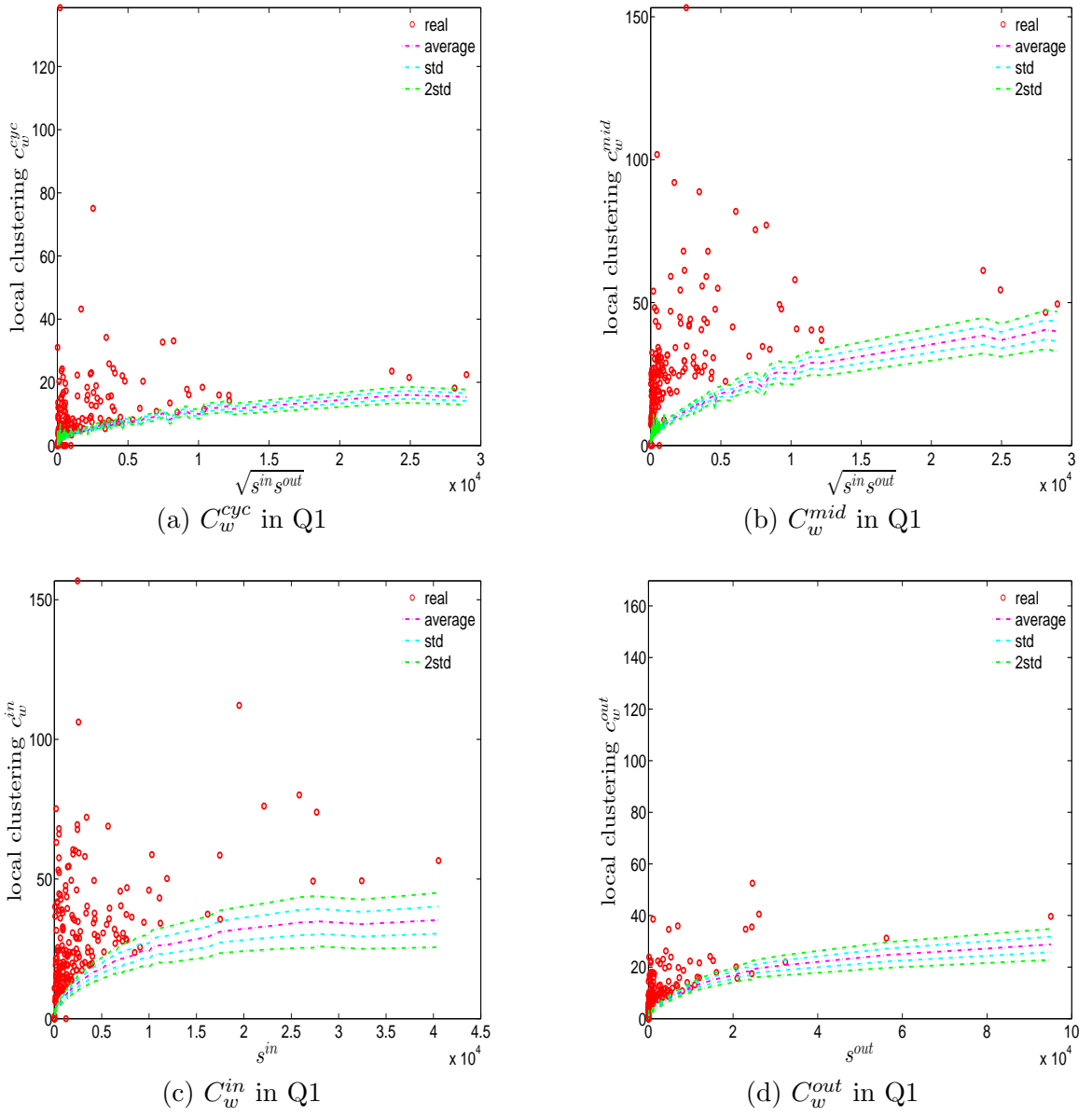
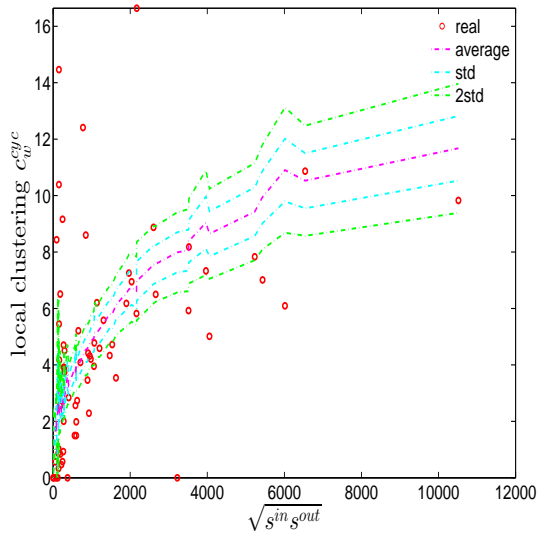
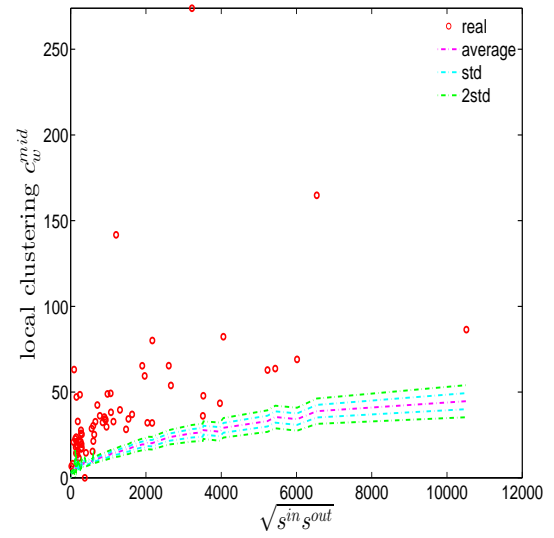


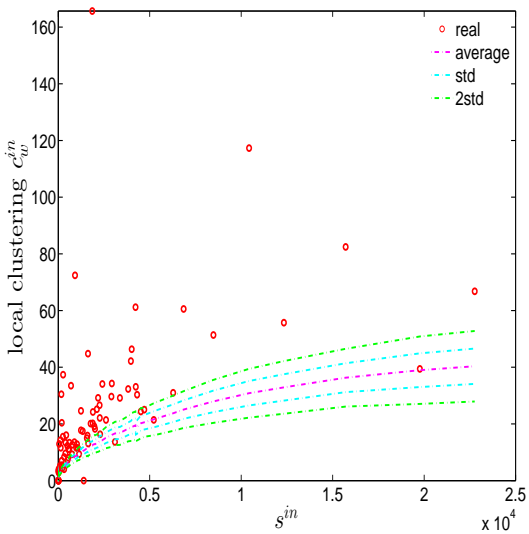
Figure 2.49: Local weighted clustering coefficients in the observed e-MID network and in the DWCM, in Q1. C_w^{cyc} (panel a), C_w^{mid} (panel b), C_w^{in} (panel c), C_w^{out} (panel d).



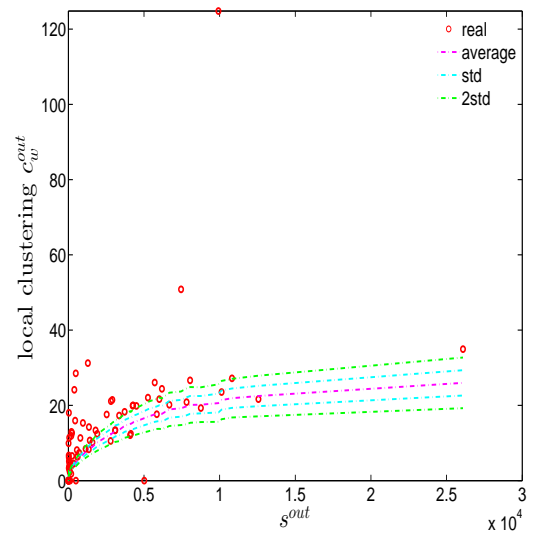
(a) C_w^{cyc} in Q48



(b) C_w^{mid} in Q48



(c) C_w^{in} in Q48



(d) C_w^{out} in Q48

Figure 2.50: Local weighted clustering coefficients in the observed e-MID network and in the DWCM, in Q48. C_w^{cyc} (panel a), C_w^{mid} (panel b), C_w^{in} (panel c), C_w^{out} (panel d).

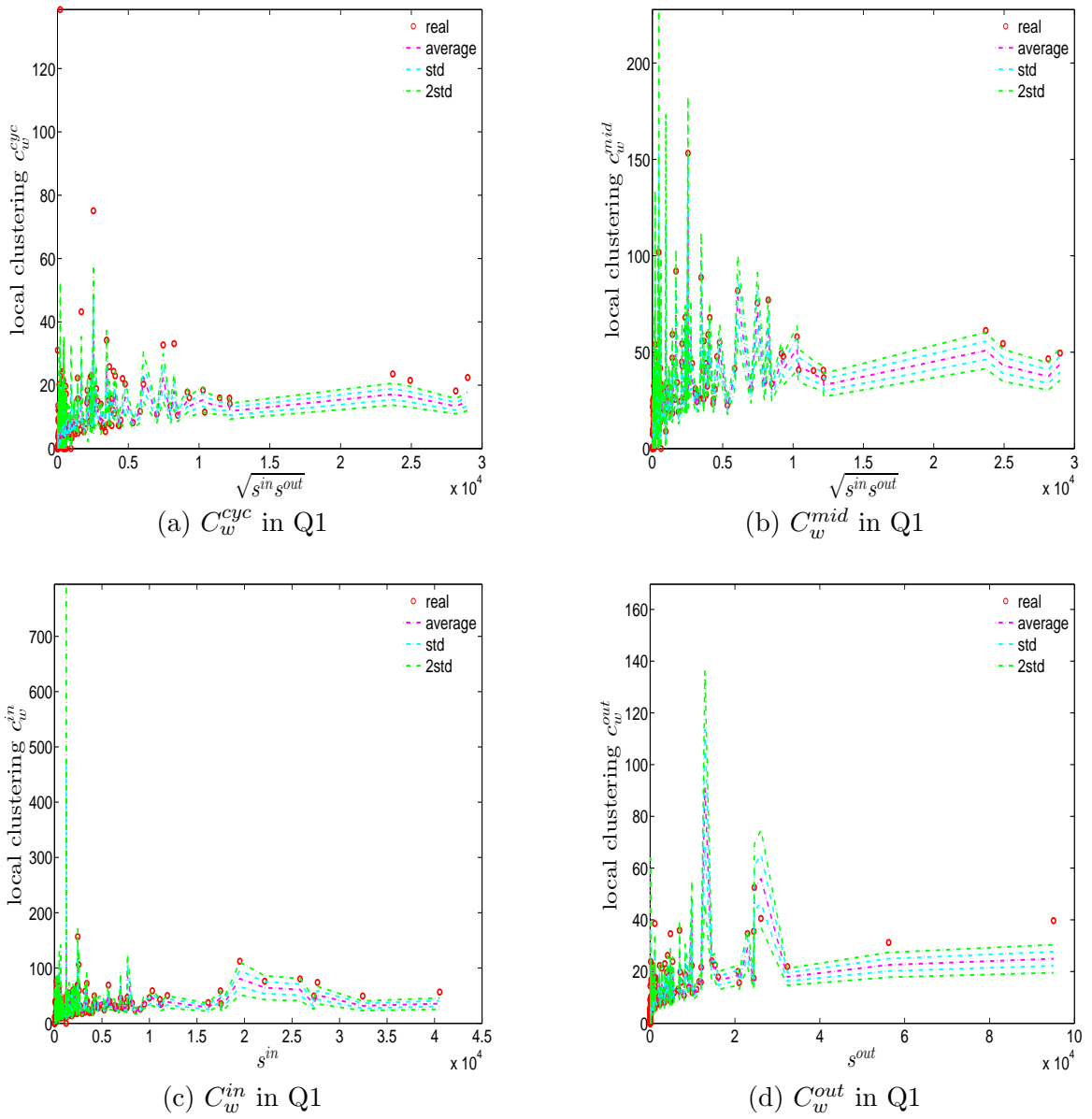
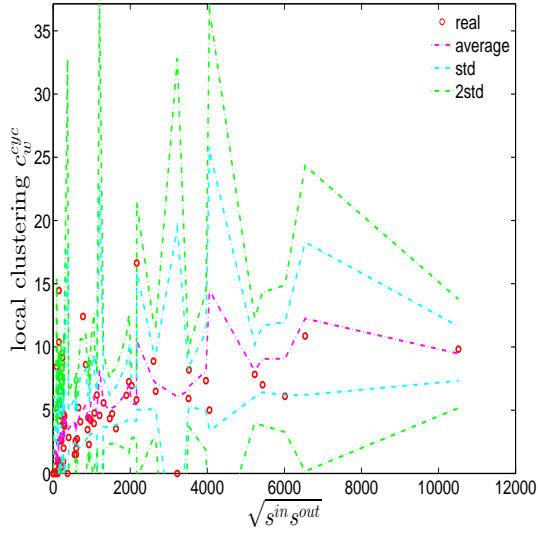
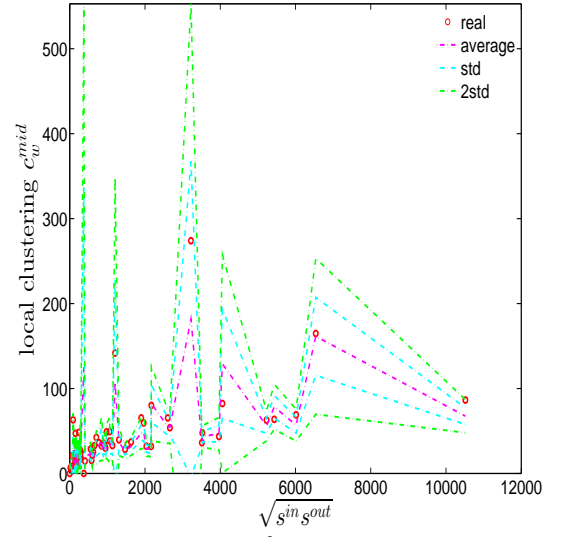


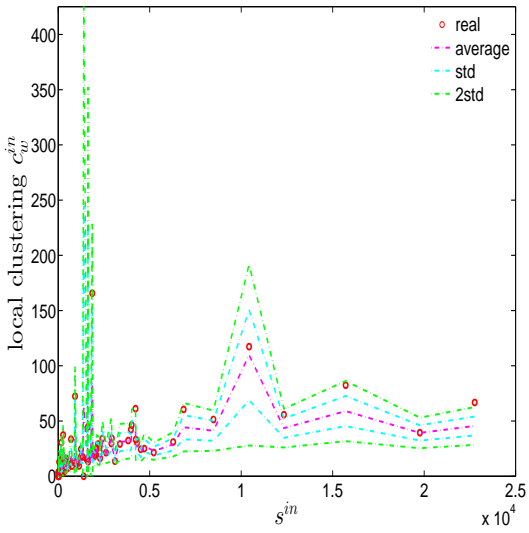
Figure 2.51: Local weighted clustering coefficients in the observed e-MID network and in the DECM, in Q1. C_w^{cyc} (panel a), C_w^{mid} (panel b), C_w^{in} (panel c), C_w^{out} (panel d).



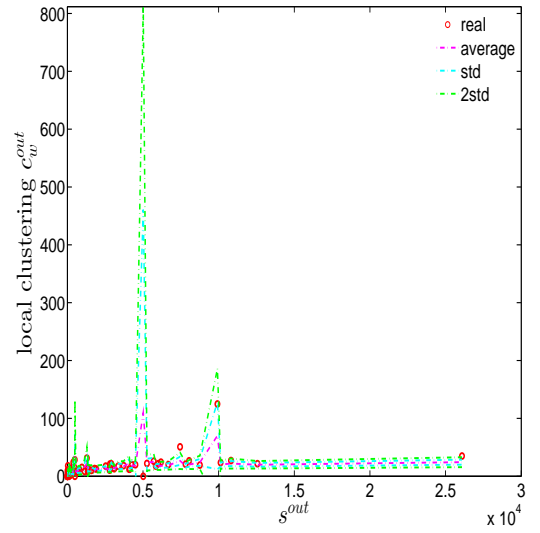
(a) C_w^{cyc} in Q48



(b) C_w^{mid} in Q48

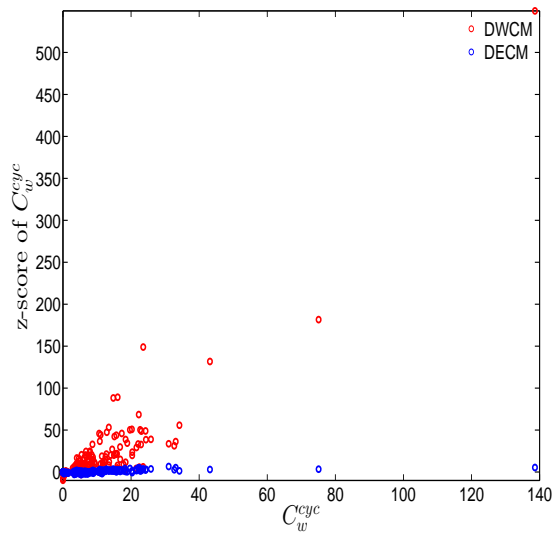


(c) C_w^{in} in Q48

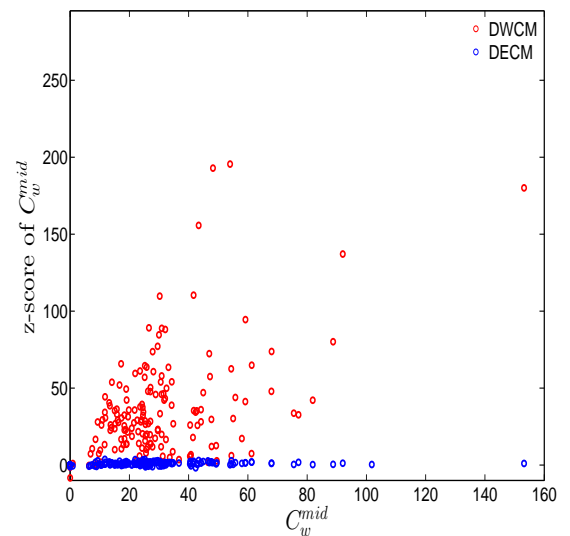


(d) C_w^{out} in Q48

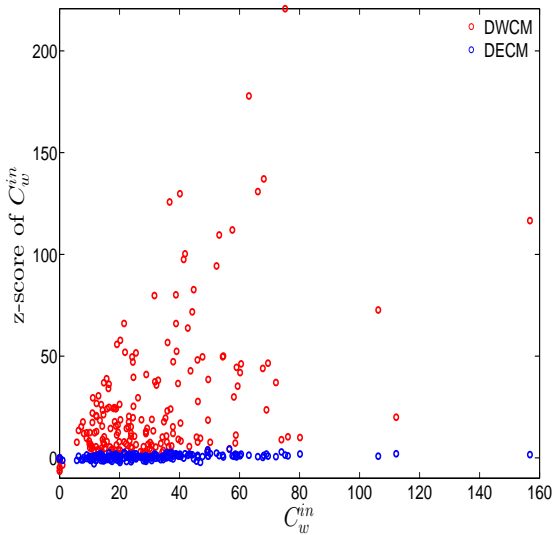
Figure 2.52: Local weighted clustering coefficients in the observed e-MID network and in the DECM, in Q48. C_w^{cyc} (panel a), C_w^{mid} (panel b), C_w^{in} (panel c), C_w^{out} (panel d).



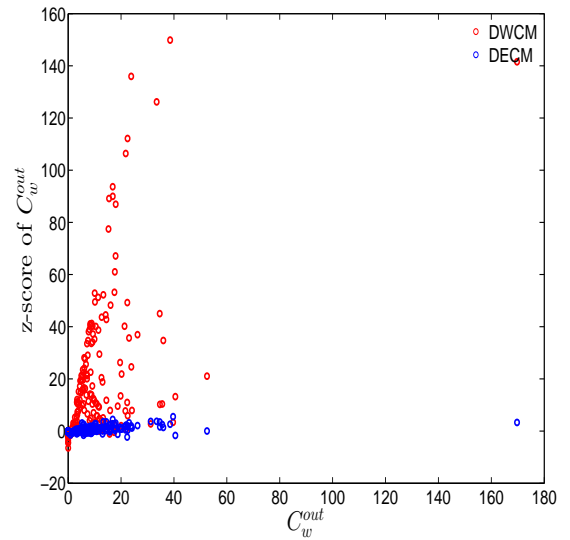
(a) z-scores of C_w^{cyc}



(b) z-scores of C_w^{mid}



(c) z-scores of C_w^{in}



(d) z-scores of C_w^{out}

Figure 2.53: z-scores of C_w vs. C_w , evaluated under the DWCM and DECM models, in Q1. Panel (a) for C_w^{cyc} , panel (b) for C_w^{mid} , panel (c) for C_w^{in} , panel (d) for C_w^{out} .

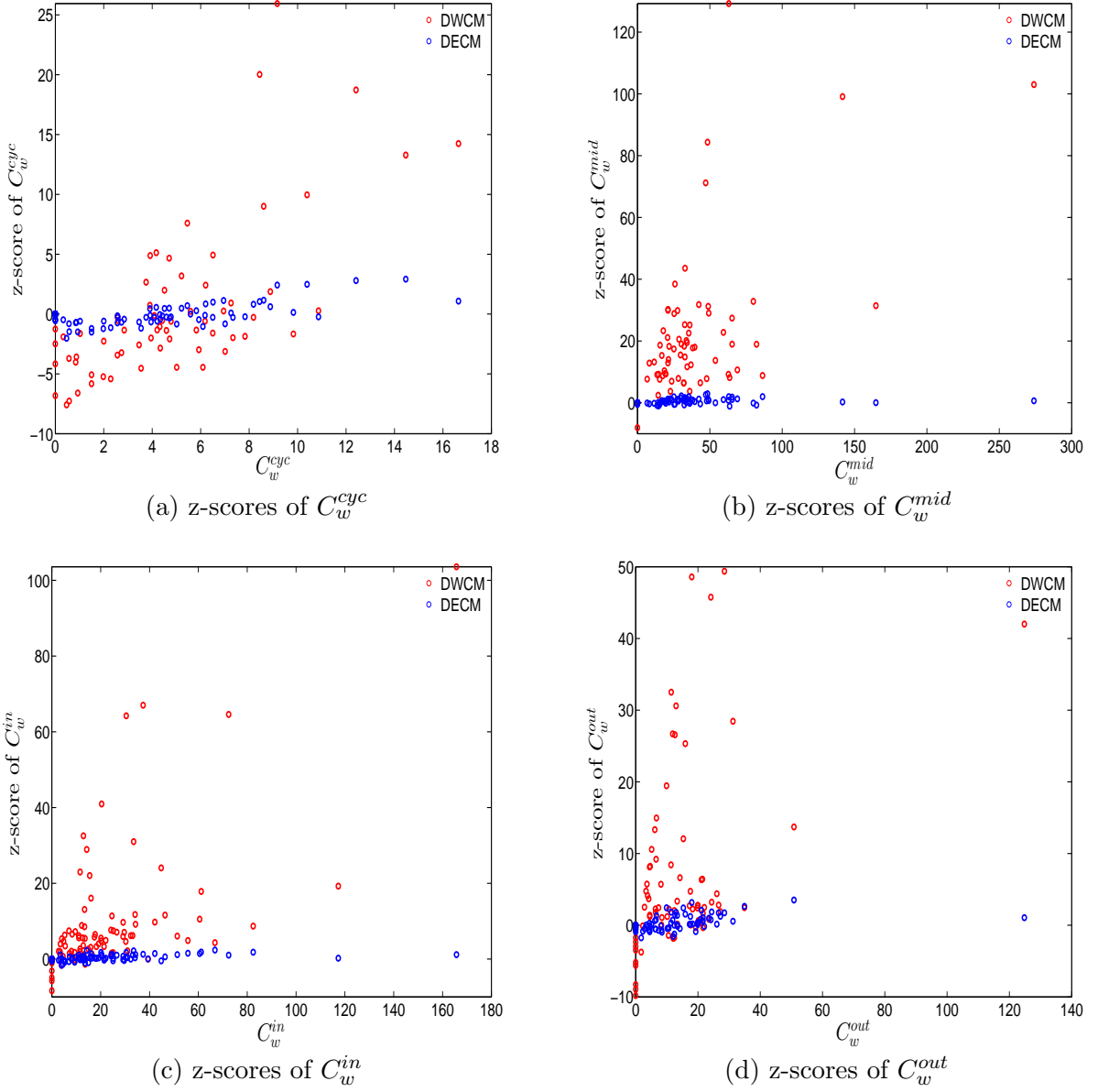


Figure 2.54: z-scores of C_w vs. C_w , evaluated under the DWCM and DECM models, in Q48. Panel (a) for C_w^{cyc} , panel (b) for C_w^{mid} , panel (c) for C_w^{in} , panel (d) for C_w^{out} .

Finally, we now analyze the predictive power of the two considered null models in terms of the evolution of the averages of the various versions of the measure ANNSs (i.e. \bar{s}_{nn}^{in-in} , \bar{s}_{nn}^{in-out} , \bar{s}_{nn}^{out-in} and $\bar{s}_{nn}^{out-out}$), the global weighted assortativity indicators (i.e. r_w^{in-in} , r_w^{in-out} , r_w^{out-in} and $r_w^{out-out}$), and the averages of the local weighted clustering coefficients (i.e. \bar{C}_w^{cyc} , \bar{C}_w^{mid} , \bar{C}_w^{in} and \bar{C}_w^{out}) (see also the next subsection section for a further comparison).

Figures (2.55), (2.56), and (2.57) show significant deviations of the observed network from the DWCM over time. A comparison between the DECM and the observed network in terms

of the aforementioned measures is shown in Figures (2.58), (2.59), and (2.60). Overall, we observe that, on the one hand, the DWCM is clearly dominated by the DECM, on the other hand, significant deviations from the DECM are still present in several quarters, regarding such as the average of the measure ANNS in the mixing category out-out ($\bar{s}_{nn}^{out-out}$), the global weighted assortativity indicators in the in-in and in-out categories, the average of the local weighted clustering coefficients \bar{C}_w^{cyc} and the average of the local weighted clustering coefficients \bar{C}_w^{out} .

We emphasize that one of the main features not explained by the sequences of degrees and strengths of the network nodes themselves is the high level of clustering in the years preceding the crisis, i.e. the huge increase in various indirect exposures generated via more intensive interbank credit links.

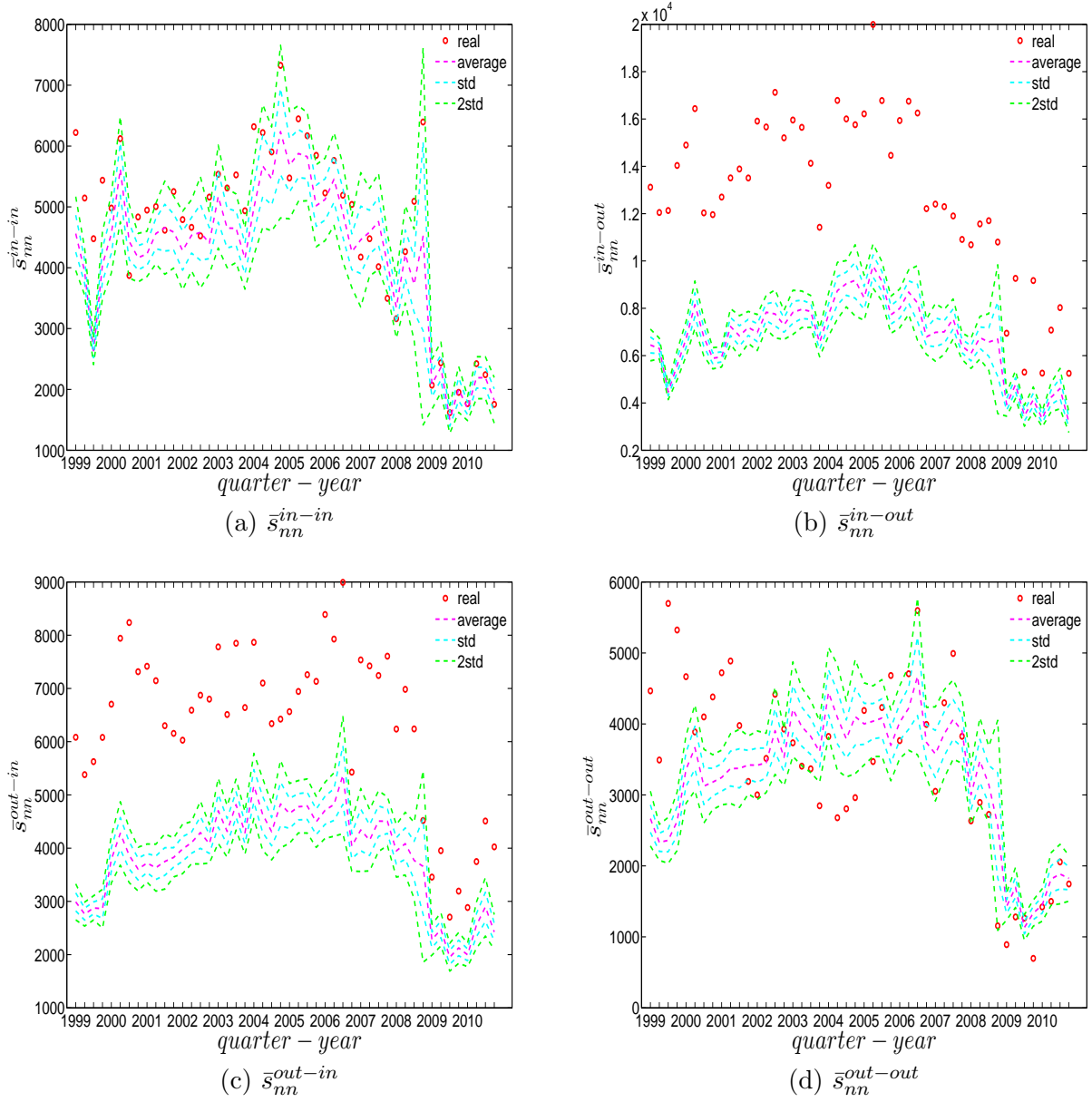


Figure 2.55: Evolution of the averages of ANNs in the observed e-MID network and in the DWCM. \bar{s}_{nn}^{in-in} (panel a), \bar{s}_{nn}^{in-out} (panel b), \bar{s}_{nn}^{out-in} (panel c), $\bar{s}_{nn}^{out-out}$ (panel d).

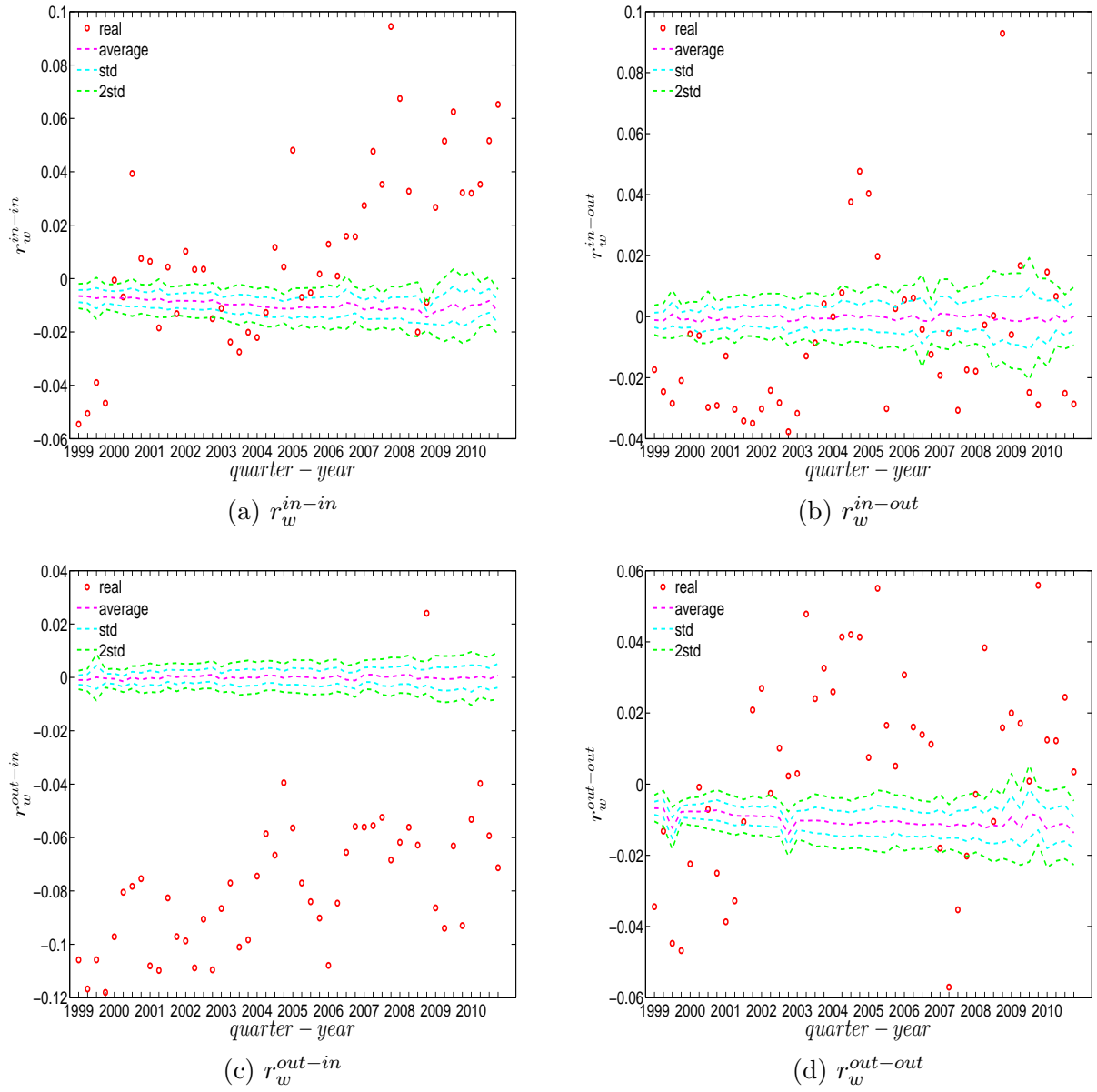


Figure 2.56: Evolution of the global weighted assortativity indicators in the observed e-MID network and in the DWCM. r_w^{in-in} (panel a), r_w^{in-out} (panel b), r_w^{out-in} (panel c), $r_w^{out-out}$ (panel d).

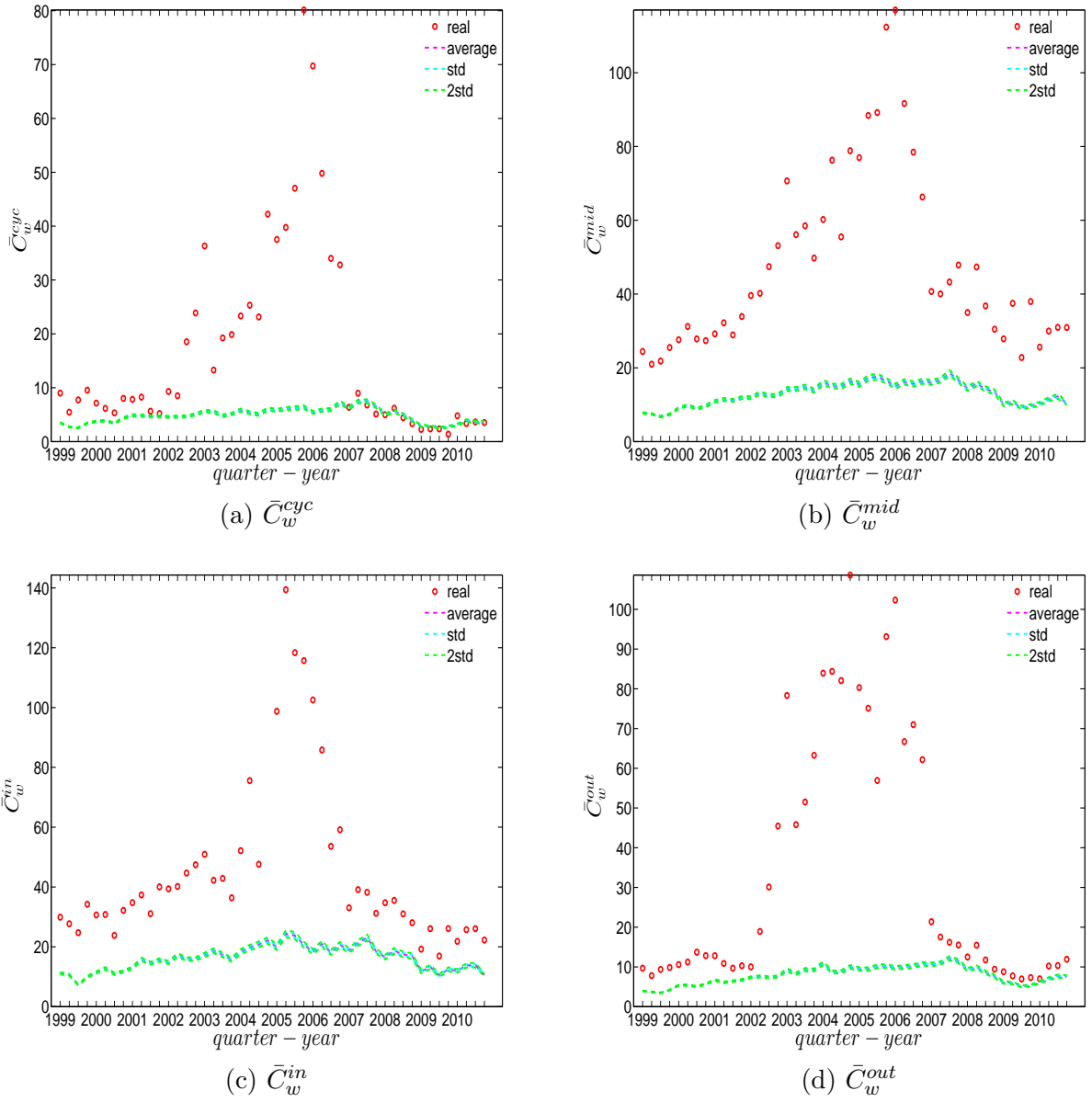


Figure 2.57: Evolution of the averages of local weighted clustering coefficients in the observed e-MID network and in the DWCM. \bar{C}_w^{cyc} (panel a), \bar{C}_w^{mid} (panel b), \bar{C}_w^{in} (panel c), \bar{C}_w^{out} (panel d).

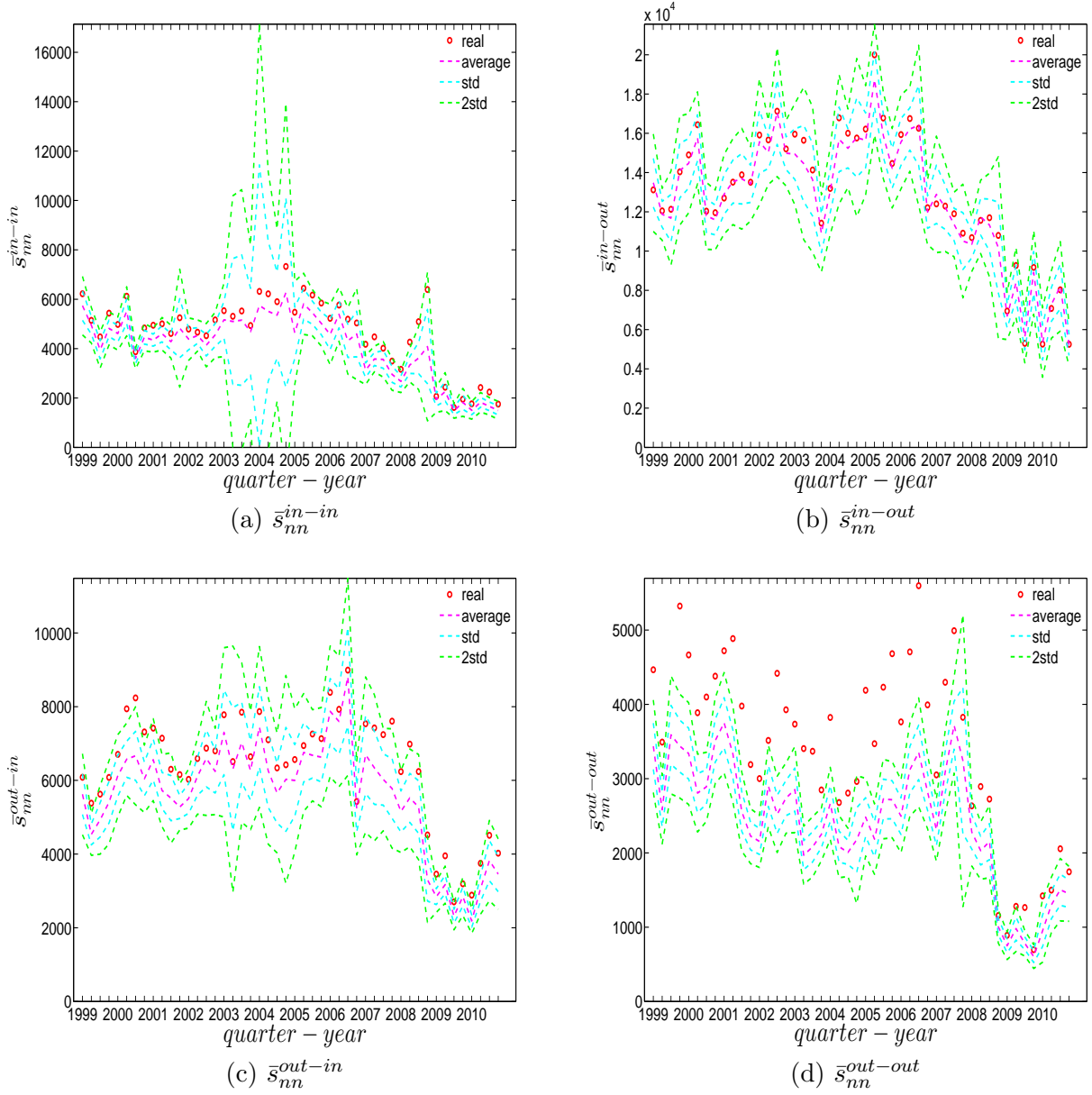


Figure 2.58: Evolution of the averages of ANNs in the observed e-MID network and in the DECM. \bar{s}_{nn}^{in-in} (panel a), \bar{s}_{nn}^{in-out} (panel b), \bar{s}_{nn}^{out-in} (panel c), $\bar{s}_{nn}^{out-out}$ (panel d).

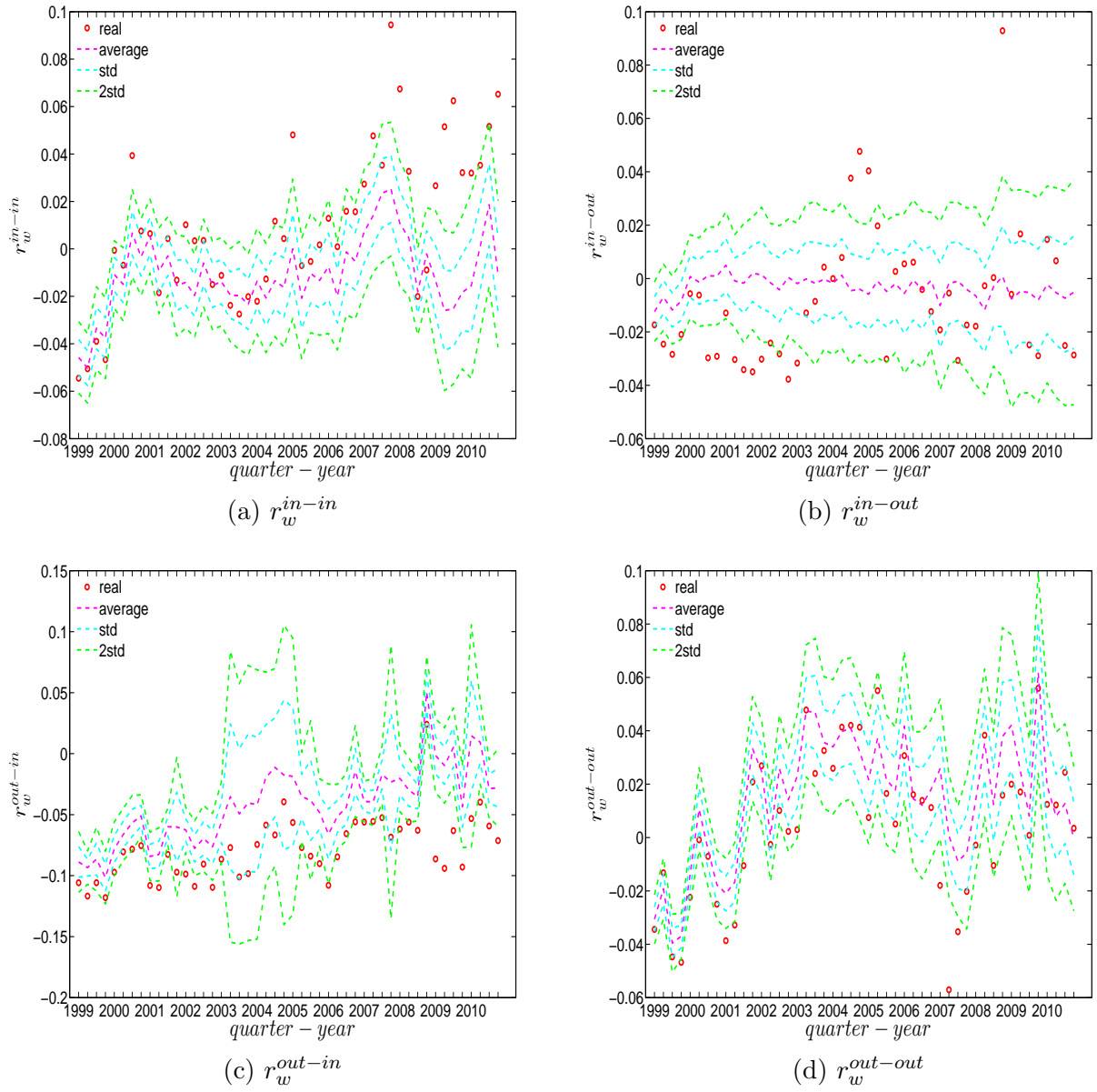


Figure 2.59: Evolution of the global weighted assortativity indicators in the observed e-MID network and in the DECM. r_w^{in-in} (panel a), r_w^{in-out} (panel b), r_w^{out-in} (panel c), $r_w^{out-out}$ (panel d).

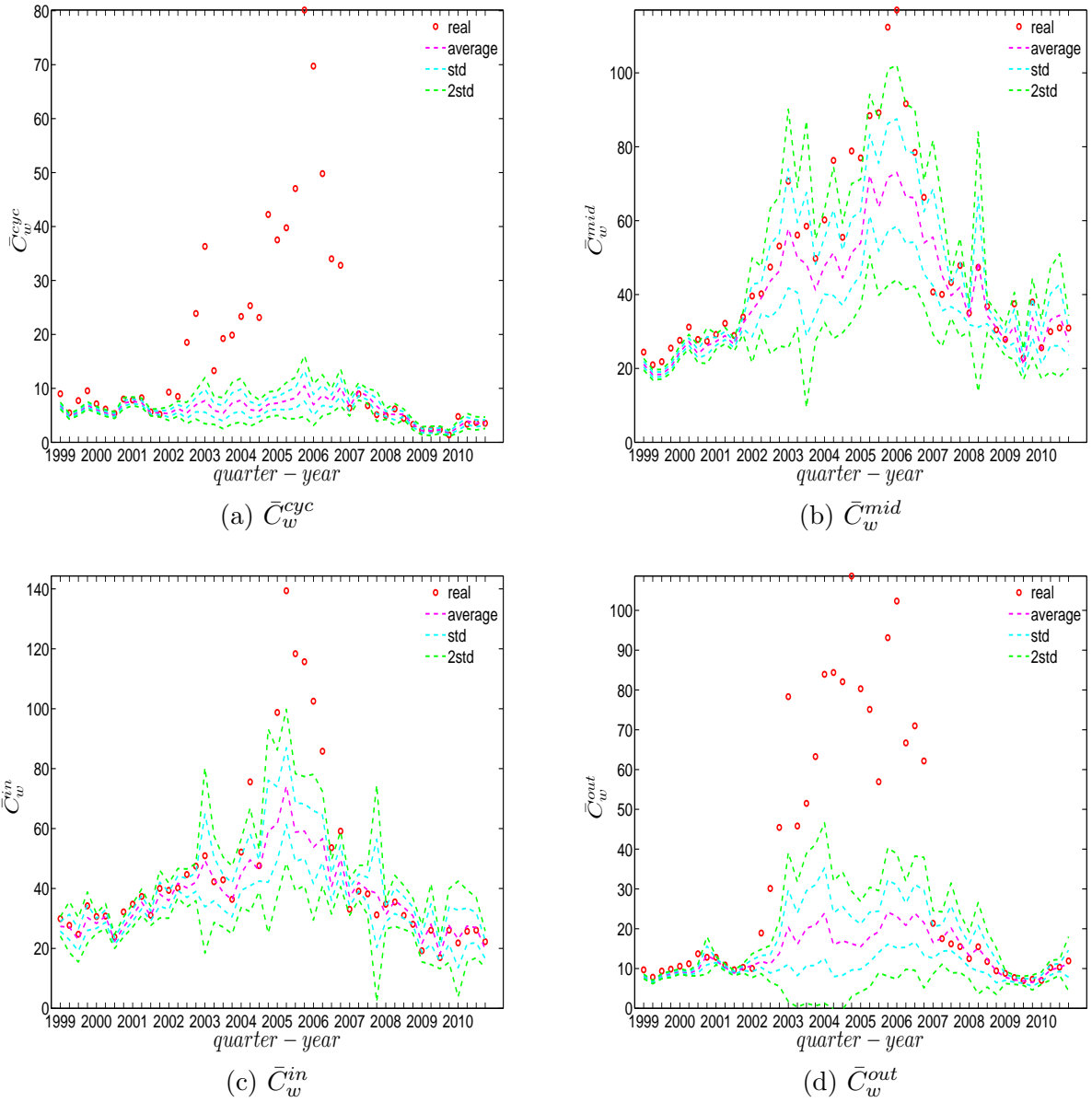


Figure 2.60: Evolution of the averages of local weighted clustering coefficients in the observed e-MID network and in the DECM. \bar{C}_w^{cyc} (panel a), \bar{C}_w^{mid} (panel b), \bar{C}_w^{in} (panel c), \bar{C}_w^{out} (panel d).

Note that, although, in general, we find that the family of Enhanced Configuration Models outperforms the family of Weighted Configuration Models in terms of replicating the main features of the structural correlations in the weighted version of the observed network, solving the system (2.77) to extract the hidden variables in the DECM (or system (2.73) in the UECM for the undirected version of the network) is much more computationally demanding than solving the system (2.68) for the DWCM (or sys. (2.64) for the UWCM for

the undirected version of the network)^{8,9}.

2.4.4 z-scores analysis revealing structural changes in the weighted system

To analyze the evolution of the discrepancies between the referenced models and the observed network, we define z-scores for the global indicators, i.e. for $\bar{s}_{nn}^{un}, r_w^{un}, \bar{C}_{un}^w$ in the undirected weighted network (evaluated under the UWCM and the UECM) and for $\bar{s}_{nn}^{in-in}, \bar{s}_{nn}^{in-out}, \bar{s}_{nn}^{out-in}, \bar{s}_{nn}^{out-out}, r_w^{in-in}, r_w^{in-out}, r_w^{out-in}, r_w^{out-out}, \bar{C}_w^{cyc}, \bar{C}_w^{mid}, \bar{C}_w^{in},$ and \bar{C}_w^{out} in the directed weighted network (evaluated under the DWCM and the DECM) (see subsection 2.7.2 in the Appendix for further details).

Before going into details, it should be noted that from Figure (2.61) to Figure (2.64), when comparing the UECM with the UWCM in the undirected version or the DECM with the DWCM in the directed version, some of the high z-scores under the UECM (or under the DECM) are blurred because of the presence of much larger z-scores under the UWCM (or under the DWCM).

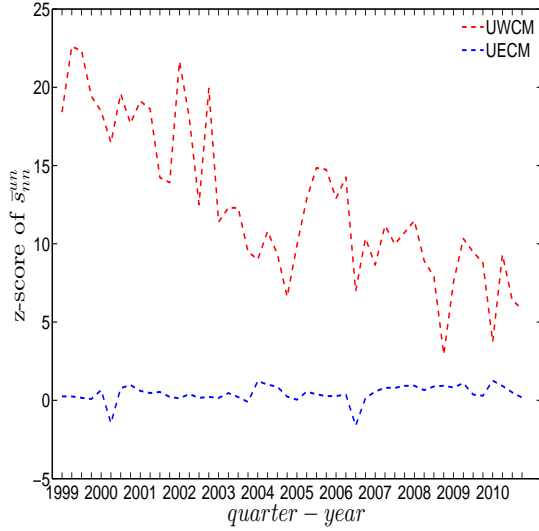
In the undirected weighted case, two important findings are obtained. First, overall, the z-scores are mostly smaller in absolute value under the UECM than under the UWCM (see Figure (2.61)). This is consistent with what we found for the local indicators and re-emphasizes the finding that the UECM out-performs the UWCM. Second, interestingly, in panel (c) of Figure (2.61) we see that, the distance between the z-scores for \bar{C}_w^{un} evaluated under the UWCM and the UECM increases over the period from 2002 to 2006, and then decreases sharply after the financial crisis. This suggests that the importance of particular basic features of a network (like its degree sequence or its strength sequence) for the emergence of higher order correlations structures can vary over time.

In the directed weighted case, similarly, we find that the z-scores under the DECM are much smaller in absolute value than those evaluated under the DWCM (see Figures (2.62), (2.63), and (2.64)). The third order correlations among banks in the directed case still seem to be much more informative than the second order ones if one would like to detect the effects

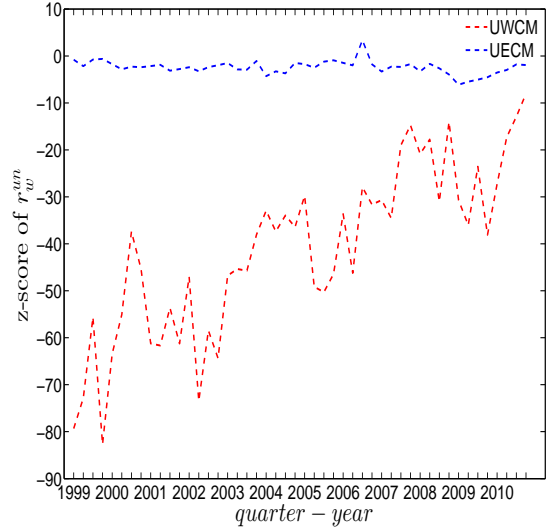
⁸According to Squartini et al. (2015), solving system (2.77) for the DECM and solving the system (2.73) for the UECM may be very time consuming if the strength distribution contains big outliers and the degree distribution is narrow. This also happens in our study, and in fact our data set shows that the strength distribution is much wider than the degree distribution.

⁹Following Mastrandrea et al. (2014) and Squartini et al. (2015), in order to speed up the process of solving system (2.73) for the UECM and system (2.77) for the DECM, we have to employ the iteration method, which uses the output of the previous iteration as the initial value for the current one. However, it remains a very time consuming process to obtain an acceptable solution for the hidden variables.

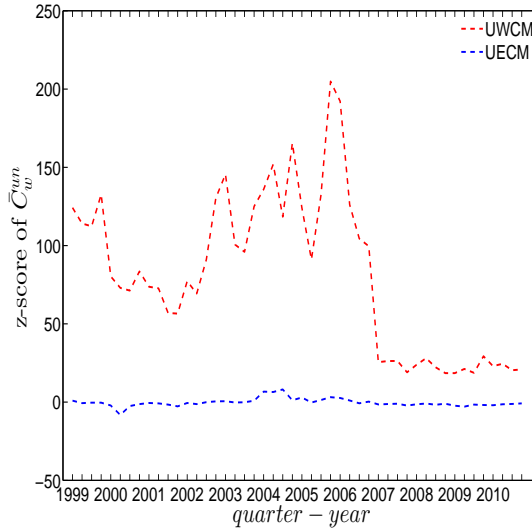
of critical events on the topology of the network. As we can see in Figure (2.64), similar to the undirected version, the distance between the z-scores for each of the \bar{C}_w evaluated under the DWCM and the DECM continuously increases during the period 2002 to 2006, and then decreases dramatically after the financial crisis.



(a) z-score of \bar{s}_{nn}^{un}

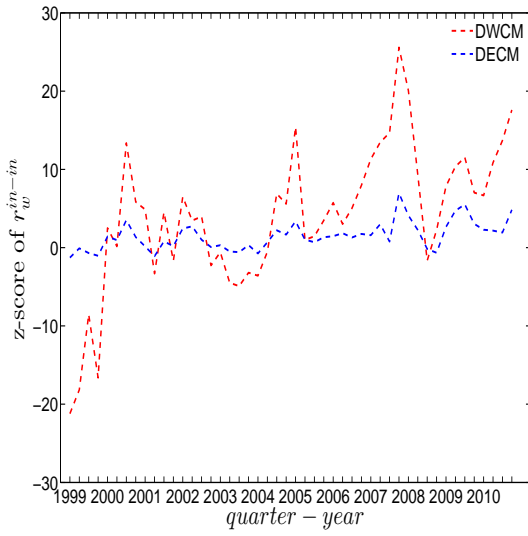


(b) z-score of r_w^{un}

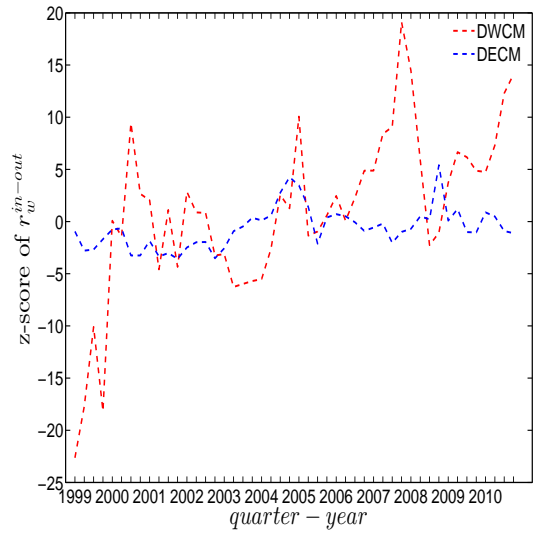


(c) z-score of \bar{C}_w^{un}

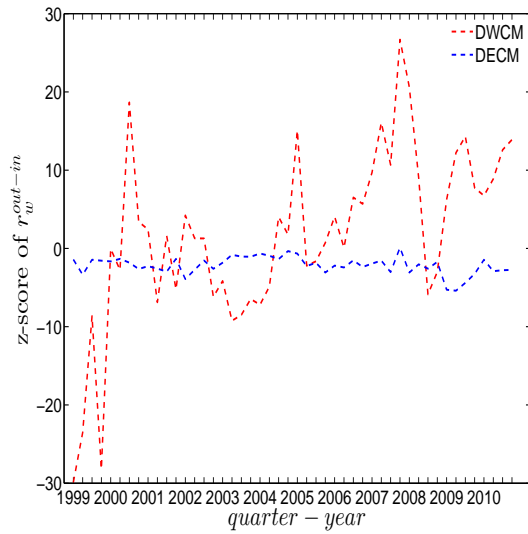
Figure 2.61: Evolution of z-scores for \bar{s}_{nn}^{un} (panel a), r_w^{un} (panel b), and \bar{C}_w^{un} (panel c) evaluated under the UWCM (red dashed lines) and the UECM (blue dashed lines).



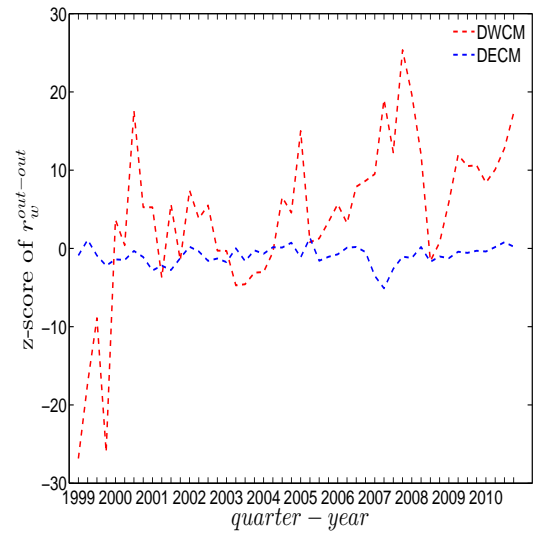
(a) z-score of r_w^{in-in}



(b) z-score of r_w^{in-out}

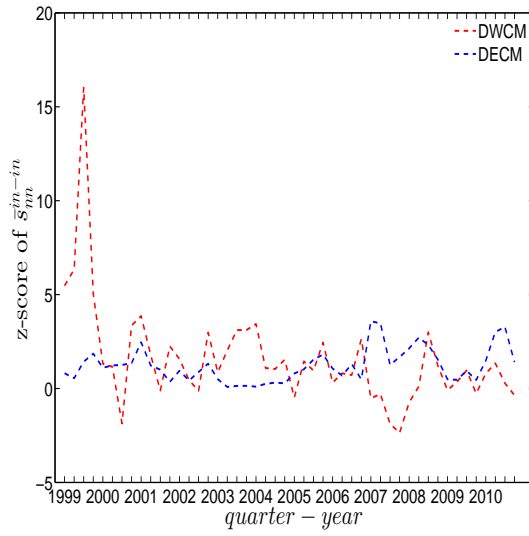


(c) z-score of r_w^{out-in}

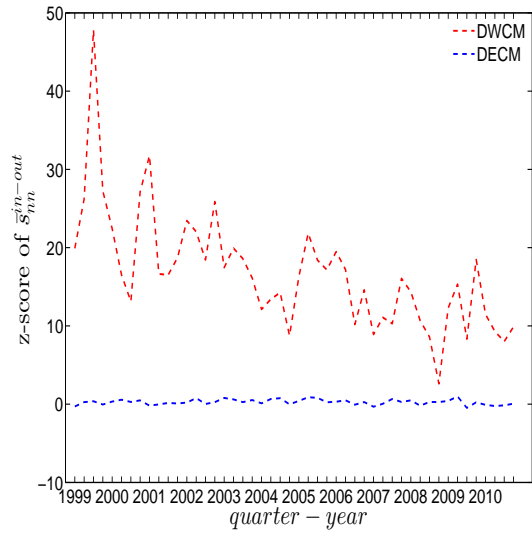


(d) z-score of $r_w^{out-out}$

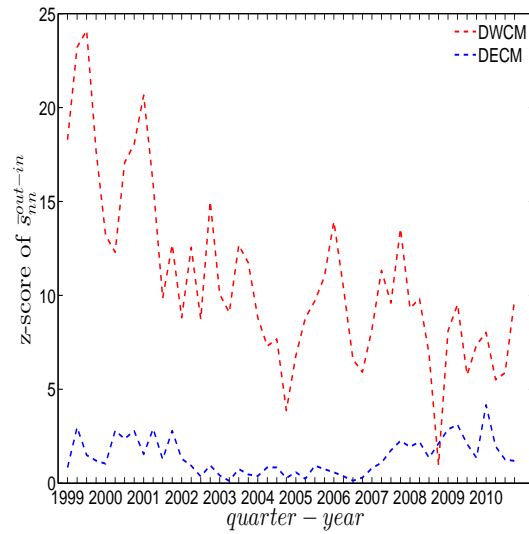
Figure 2.62: Evolution of z-scores for r_w^{in-in} (panel a), r_w^{in-out} (panel b), r_w^{out-in} (panel c), and $r_w^{out-out}$ evaluated under the DWCM (red dashed lines) and the DECM (blue dashed lines).



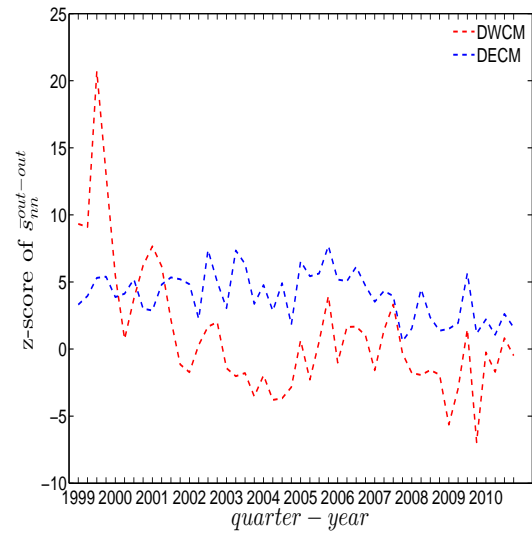
(a) z-score of \bar{s}_{nn}^{in-in}



(b) z-score of \bar{s}_{nn}^{in-out}

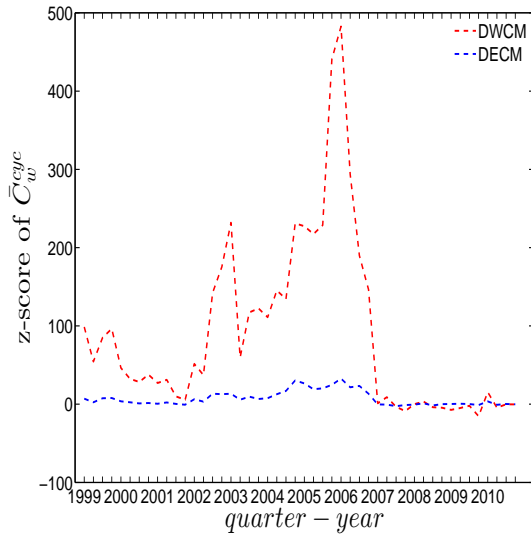


(c) z-score of \bar{s}_{nn}^{out-in}

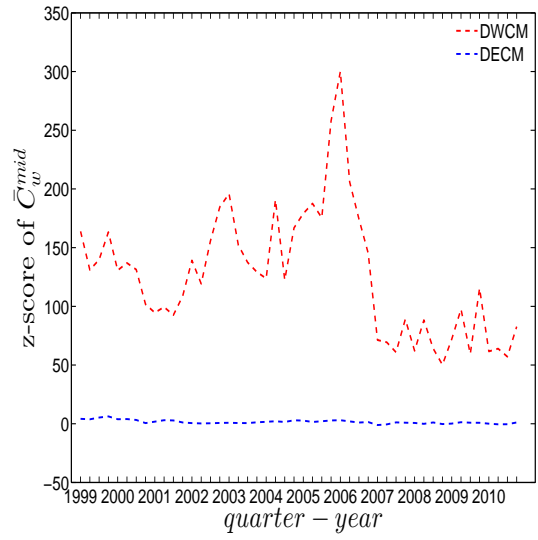


(d) z-score of $\bar{s}_{nn}^{out-out}$

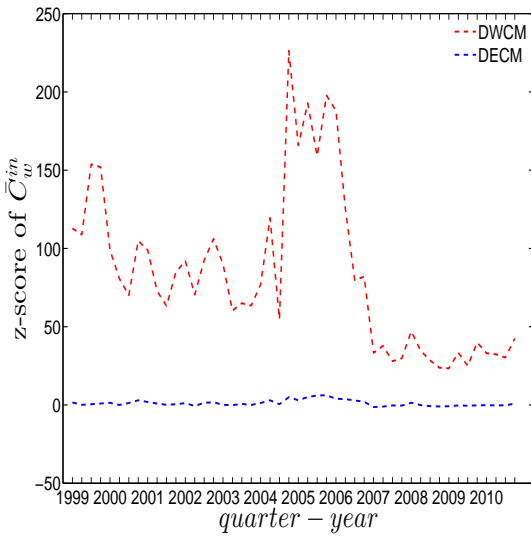
Figure 2.63: Evolution of z-scores for \bar{s}_{nn}^{in-in} (panel a), \bar{s}_{nn}^{in-out} (panel b), \bar{s}_{nn}^{out-in} (panel c), and $\bar{s}_{nn}^{out-out}$ evaluated under the DWCM (red dashed lines) and the DECM (blue dashed lines).



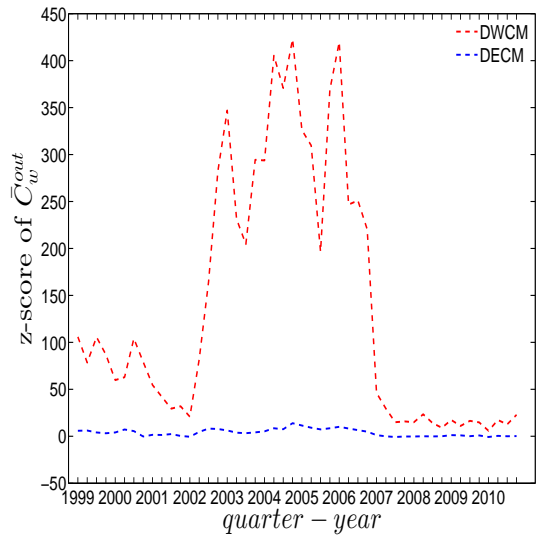
(a) z-score of \bar{C}_w^{cyc}



(b) z-score of \bar{C}_w^{mid}



(c) z-score of \bar{C}_w^{in}



(d) z-score of \bar{C}_w^{out}

Figure 2.64: Evolution of z-scores for \bar{C}_w^{cyc} (panel a), \bar{C}_w^{mid} (panel b), \bar{C}_w^{in} (panel c), and \bar{C}_w^{out} evaluated under the DWCM (red dashed lines) and the DECM (blue dashed lines).

2.5 Conclusions

In this study, we investigated the structural correlations in the e-MID network. We find that the observed structural correlations can vary across different versions of the network (binary vs weighted and undirected vs directed). In the undirected version of the network, the mixing is disassortative in both the binary and the weighted case. In addition, when the directions of the edges are taken into account, we find that among the four mixing categories (i.e. in-in, in-out, out-in, and out-out), the global assortativity in the out-in category comes closest to the mixing observed in the undirected network. The similarity between these two quantities is suggested in the study by van der Hoorn and Litvak (2015). Due to the fact that only in the out-in mixing category the considered edge (see the out-in category in Figure (2.1)) contributes to the node degrees on both of its sides, this mixing category can be considered a generalization of the mixing in undirected networks. During our analysis of the evolution of the third order correlations among banks over time, we detected dramatic changes in the network structure surrounding the recent financial crisis in 2007. More specifically, in the weighted network, the averages of the local weighted clustering coefficients appear elevated from the adoption of the Euro up until 2006, and then decrease dramatically around the time of the financial crisis. We also report strong indications of elevated systemic risk in the network, evidenced by the prevalence of the “middleman” and “inward” types of clustering in the network.

Moreover, by employing the various configuration models, we examined whether the information encoded in the local constraints (like the observed degree sequence and/or the strength sequence of a network) can explain higher order structural correlations. We find that, in the binary case, the degree sequence is informative in terms of explaining the main features of the structural correlations in the e-MID network. However, under scrutiny, the binary e-MID network does display some non-random patterns that cannot simply be explained by the degree sequence in conjunction with the configuration model.

In the weighted version of the network, for the most part, the structural correlations in the observed e-MID network are deviating strongly from their respective expectations evaluated under the Weighted Configuration Models, which capture only the heterogeneity in the strength sequence(s) (i.e. the UWCM in the undirected version and the DWCM in the directed version). One possible explanation is that while all measures of structural correlations used in the weighted network depend on the elements of both the adjacency as well as the weighting matrices, neither the UWCM or DWCM utilize information about the node degrees (degree sequence), which is, in fact, found to be more important than the

strength sequence in reproducing the topological properties of real world networks (see, for example, Squartini et al., 2011a, Squartini et al., 2011b; Squartini et al., 2015).

Due to the failures of the UWCM and the DWCM, we consider the family of Enhanced Configuration Models, which constrains the degree as well as the strength sequences of the randomized ensemble to match those of the observed network on average (i.e. UECM in the undirected case and the DECM in the directed case). Our findings indicate that the randomized ensembles produced by the Enhanced Configuration Model have a much greater predictive power. This is in line with what was found in previous studies such as Mastrandrea et al. (2014) and Squartini et al. (2015), and is not very surprising since the Enhanced Configuration Models utilize more information when replicating the structural correlations of the observed network. The results obtained from the analysis of the DECM confirms the role that the distribution of the in-coming and out-going degrees in directed weighted networks plays for the emergence of higher order structural correlations.

Still, a detailed comparison between the observed network and the Enhanced Configuration Models reveals that even this family of Configuration Models is not able to produce accurate estimates for all the measures of structural correlations we used, meaning that some of the patterns can be considered non-random or unexplained by the models. For instance, in the undirected network, we find that even when using the UECM, the weighted assortativity deviates significantly from the respective expected value in a couple of times. In the directed weighted network, the global weighted assortativity in the in-in as well as in the in-out mixing categories and the average of the local weighted clustering coefficients of “inward”, “outward”, and “cyclical” clustering also display non-random patterns in several quarters, mainly from 2002 to 2006. The high degree of clustering in this episode is the one characteristic that can not be explained satisfactorily via the influence of lower-order characteristics like the degree and strength sequences. Hence, this finding points to a behavioral change in the formation of the credit network: A deliberate increase of indirect exposure through multiple credit relations. Interestingly, with the crisis year 2007, we find an abrupt reduction of all clustering coefficients to their “normal” levels implied by the degree and strength sequences.

The Enhanced Configuration Models also fail to reproduce the local behavior of certain banks captured by the local indicators of structural correlations. Unfortunately, because of the lack of more detailed information about the banks in the system, we can not identify the factors for the formations of such non-random patterns.

Interestingly, similar to the study of Squartini et al. (2013)¹⁰, we also observe the evidence for structural changes when comparing the weighted version of the e-MID network with the weighted configuration models. More specifically, the distance between the predictions of the Weighted Configuration Models and of the Enhanced Configuration Models for the averages of local weighted clustering coefficients continuously increases from the adoption of the Euro up until the financial crisis in 2007 and then sharply decreases after that. This result can be interpreted as an indication of structural changes in the network associated with these two critical events. It also suggests that the importance of particular basic features of a network (like its degree sequence or its strength sequence) for the emergence of higher order correlation structures can vary over time.

Due to issues of confidentiality, in many cases, the biggest challenge in the analysis of complex real financial systems lies in the utilization of the limited available information. Our results can be understood as an evaluation of the potential of configuration models to reconstruct higher order topological properties of a network from limited information (e.g. see Mastrandrea et al., 2014; Cimini et al., 2015a). Successful information intensive network analysis, like, for example, systemic risk evaluation, can be conducted on reconstructed networks only to the extent to which the reconstruction is reliable (see, for example, Cimini et al., 2015b).

In addition, the configuration models translate the local constraints in the observed network into hidden variables associated with the individual banks. It would be interesting to investigate whether some individual node characteristics (i.e. non-topological properties) correlate with the extracted hidden variables (see, for example, Garlaschelli and Loffredo, 2004; Garlaschelli et al., 2007; Garlaschelli and Loffredo, 2008; Almog et al., 2015), however, such additional information is unfortunately not available in our data set. This can be a fruitful direction for future research into financial networks.

Moreover, since the Exponential Random Graph Model is generic and flexible enough, one may want to investigate the extent to which it can be useful to use other statistics of the observed network as ensemble constraints. For instance, the average degree of the nearest neighbors or the local clustering coefficients might also prove informative in explaining particular topological properties of the observed network (see, for example, Park and Newman (2004) and Bianconi (2009) for employing different constraints). In addition, since the second and third order structural correlations are the main focus of this study, we suggest that the role of various constraints for the emergence of higher order correlations (or motifs) and

¹⁰Squartini et al. (2013) focus on the analysis of the binary version of the network of interbank exposures among Dutch banks over the period 1998-2008.

for the meso-scale network structures such as the core-periphery and community structures should be studied further.

2.6 References

Almog A., Squartini T., Garlaschelli D. 2015. A GDP-driven model for the binary and weighted structure of the international trade network. *New Journal of Physics* 17.

Barrat A., Barthélemy M., Pastor-Satorras R., Vespignani A. 2004. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci.* 101 (11), pp. 3747-3752.

Bianconi G. 2009. Entropy of network ensembles. *Physical Review E* 79 (3).

Cimini G., Squartini T., Garlaschelli D., Gabrielli A. 2015a. Estimating topological properties of weighted networks from limited information. *Physical Review E* 92 (4).

Cimini G., Squartini T., Garlaschelli D., Gabrielli A. 2015b. Systemic risk analysis on reconstructed economic and financial networks. *Scientific Reports* 5.

De Masi G., Iori G., Caldarelli G. 2006. Fitness model for the Italian interbank money market. *Physical Review E* 74 (6).

Fagiolo G. 2007. Clustering in complex directed networks. *Physical Review E* 76 (2).

Finger K., Fricke D., Lux T. 2013. Network analysis of the e-MID overnight money market: the informational value of different aggregation levels for intrinsic dynamic processes. *Computational Management Science* 10 (2), pp. 187-211.

Foster J. G., Foster D. V., Grassberger P., Paczuski M. 2010. Edge direction and the structure of networks. *Proc. Natl. Acad. Sci.* 107 (24), pp. 10815-10820.

Fricke D. 2012. Trading strategies in the overnight money market: Correlations and clustering on the e-MID trading platform. *Physica A: Statistical Mechanics and its Applications*, 391 (24), pp. 6528-6542.

Fricke D., Finger K., Lux T. 2013. On assortative and disassortative mixing scale-free networks: The case of interbank credit networks. *Kiel Working Papers 1830*, Kiel Institute for the World Economy. Available at:

https://www.ifw-members.ifw-kiel.de/publications/on-assortative-and-disassortative-mixing-scale-free-networks-the-case-of-interbank-credit-networks/1830_KWP.pdf.

Fricke D., Lux T. 2015a. On the distribution of links in the interbank network: evidence from the e-MID overnight money market. *Empirical Economics* 49 (4), pp. 1463-1495.

Fricke D., Lux T. 2015b. Core-periphery structure in the overnight money market: Evi-

dence from the e-MID trading platform. *Computational Economics*. 45 (3), pp. 359-395.

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

Garlaschelli D., Di Matteo T., Aste T., Caldarelli G., Loffredo M. 2007. Interplay between topology and dynamics in the world trade web. *The European Physical Journal B* 57 (2), pp. 159-164.

Garlaschelli D., Loffredo M. I. 2008. Maximum likelihood: Extracting unbiased information from complex networks. *Physical Review E* 78 (1).

Holme P., Park S. M., Kim B. J., Edling C. R. 2007. Korean university life in a network perspective: Dynamics of a large affiliation network. *Physica A: Statistical Mechanics and its Applications* 373, pp. 821-830.

Maslov S., Sneppen K. 2002. Specificity and stability in topology of protein networks. *Science* 296 (5569), pp. 910-913.

Maslov S., Sneppen K., Zaliznyak A. 2004. Detection of topological patterns in complex networks: Correlation profile of the internet. *Physica A: Statistical Mechanics and its Applications* 333, pp. 529-540.

Mastrandrea R., Squartini T., Fagiolo G., Garlaschelli D. 2014. Enhanced reconstruction of weighted networks from strengths and degrees. *New Journal of Physics* 16.

Newman M. E. J. 2002. Assortative mixing in networks. *Physical Review Letters* 89 (20).

Newman M. E. J. 2003a. Mixing patterns in networks. *Physical Review E* 67 (2).

Newman M. E. J. 2003b. The structure and function of complex networks. *Society for Industrial and Applied Mathematics Review* 45 (2), pp. 167-256.

Onnela J. -P., Saramäki J., Kertész J., Kaski K. 2005. Intensity and coherence of motifs in weighted complex networks. *Physical Review E* 71 (6).

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

Park J., Newman M. E. J. 2004. Statistical mechanics of networks. *Physical Review E* 70 (6).

Piraveenan M., Prokopenko M., Zomaya A. 2010. Classifying complex networks using unbiased local assortativity. In Fellermann H., Dörr M., Hanczyc M. M., Ladegaard Laursen L., Maurer S., Merkle D., Monnard P. -A., Stoy K., Rasmussen S. (Eds.), *Artificial Life XII, Proc. 12th Int'l Conf. Synthesis and Simulation of Living Systems* (pp. 329-336).

Piraveenan M., Prokopenko M., Zomaya A. 2012. Assortative mixing in directed bio-

logical networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 9 (1), pp. 66-78.

Saramäki J., Kivela M., Onnela J., Kaski K., Kertész J. 2007. Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E* 75 (2).

Squartini T., Fagiolo G., Garlaschelli D. 2011a. Randomizing world trade. I. A binary network analysis. *Physical Review E* 84 (4).

Squartini T., Fagiolo G., Garlaschelli D. 2011b. Randomizing world trade. II. A weighted network analysis. *Physical Review E* 84 (4).

Squartini T., Garlaschelli D. 2011. Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics* 13.

Squartini T., van Lelyveld I., Garlaschelli D. 2013. Early-warning signals of topological collapse in interbank networks. *Scientific Reports* 3.

Squartini T., Mastrandrea R., Garlaschelli D. 2015. Unbiased sampling of network ensembles. *New Journal of Physics* 17.

Tabak B. M., Takami M., Rocha J. M. C., Cajueiro D. O., Souza S. R. S. 2014. Directed clustering coefficient as a measure of systemic risk in complex banking networks. *Physica A: Statistical Mechanics and its Applications* 394, pp. 211-216.

van der Hoorn P., Litvak N. 2015. Degree-degree dependencies in directed networks with heavy-tailed degrees. *Internet Mathematics* 11 (2), pp. 155-179.

Watts D. J., Strogatz S. H. 1998. Collective dynamics of “small-world” networks. *Nature* 393, pp. 440-442.

Zlatic V., Bianconi G., Díaz-Guilera A., Garlaschelli D., Rao F., Caldarelli G. 2009. On the rich-club effect in dense and weighted networks. *The European Physical Journal B* 67 (3), pp. 271-275.

Zhang B., Horvath S. 2005. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology* 4 (1).

2.7 Appendix

2.7.1 Assortativity Coefficients

Overall Assortativity

In an undirected network, define the list of m edges $\{A_e B_e\}_{e=1}^m$, where for each index e , the two nodes A_e, B_e stand for the ends of an edge. Note that, the overall assortativity

indicator r_{bin}^{un} can be calculated via $\{k_i^{un}\}_{i=1}^n$ and $\{k_{nn,i}^{un}\}_{i=1}^n$ as

$$r_{bin}^{un} = \frac{\sum_{i=1}^n (k_i^{un})^2 k_{nn,i}^{un} - \frac{1}{2m} [\sum_{i=1}^n (k_i^{un})^2]^2}{\sum_{i=1}^n (k_i^{un})^3 - \frac{1}{2m} [\sum_{i=1}^n (k_i^{un})^2]^2}, \quad (2.92)$$

where $m = \frac{1}{2} \sum_{i=1}^n k_i^{un}$ (e.g. Park and Newman, 2003).

In a directed network, suppose that we have a list of M edges $\{A_e B_e\}_{e=1}^M$, where for each index e , the two nodes A_e, B_e respectively stand for the source and target nodes (note that $M = \sum_{i=1}^n k_i^{in} = \sum_{i=1}^n k_i^{out}$).

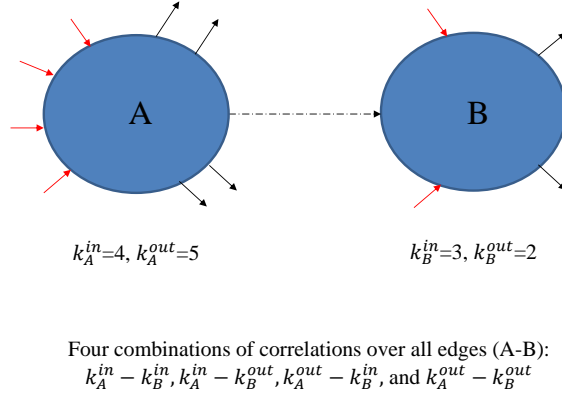


Figure 2.65: In-coming, out-going degrees to two vertices of an edge in directed networks.

Each node A_e or B_e has an in-coming degree and an out-going degree (see Figure (2.65)). Consequently, we have four combinations of degrees associated with each edge as mentioned in Figure (2.1). Therefore, regarding the degree dependencies, four separate indicators can be obtained, i.e. $r_{bin}^{in-in}, r_{bin}^{out-in}, r_{bin}^{in-out}, r_{bin}^{out-out}$. Similar to the undirected case, mathematically, these measures of overall assortativity actually depend on the degree sequences $\{k_i^{in}\}_{i=1}^n, \{k_i^{out}\}_{i=1}^n$ as well as the sequences of the average nearest neighbor degrees $k_{nn,i}^{in-in}, k_{nn,i}^{in-out}, k_{nn,i}^{out-in}, k_{nn,i}^{out-out}$ (e.g. Piraveenan et al., 2012; van der Hoorn and Litvak, 2015). More specifically, accordingly, they are given by

$$r_{bin}^{in-in} = \frac{\frac{1}{2} [\sum_{i=1}^n (k_i^{in})^2 k_{nn,i}^{in-in} + k_i^{in} k_i^{out} k_{nn,i}^{out-in}] - \frac{1}{M} [\sum_{i=1}^n (k_i^{in})^2 \sum_{i=1}^n (k_i^{in} k_i^{out})]}{\sqrt{\{\sum_{i=1}^n (k_i^{in})^3 - \frac{1}{M} [\sum_{i=1}^n (k_i^{in})^2]^2\} \{\sum_{i=1}^n (k_i^{in})^2 k_i^{out} - \frac{1}{M} [\sum_{i=1}^n (k_i^{in} k_i^{out})]^2\}}}, \quad (2.93)$$

$$r_{bin}^{in-out} = \frac{\frac{1}{2} \sum_{i=1}^n k_i^{in} k_i^{out} (k_{nn,i}^{in-in} + k_{nn,i}^{out-out}) - \frac{1}{M} [\sum_{i=1}^n (k_i^{in} k_i^{out})]^2}{\sqrt{\{\sum_{i=1}^n (k_i^{in})^2 k_i^{out} - \frac{1}{M} [\sum_{i=1}^n (k_i^{in} k_i^{out})]^2\} \{\sum_{i=1}^n (k_i^{out})^2 k_i^{in} - \frac{1}{M} [\sum_{i=1}^n (k_i^{out} k_i^{in})]^2\}}}, \quad (2.94)$$

$$r_{bin}^{out-in} = \frac{\frac{1}{2} \sum_{i=1}^n [(k_i^{out})^2 k_{nn,i}^{out-in} + (k_i^{in})^2 k_{nn,i}^{in-out}] - \frac{1}{M} [\sum_{i=1}^n (k_i^{in})^2 \sum_{i=1}^n (k_i^{out})^2]}{\sqrt{\{\sum_{i=1}^n (k_i^{in})^3 - \frac{1}{M} [\sum_{i=1}^n (k_i^{in})^2]^2\} \{\sum_{i=1}^n (k_i^{out})^3 - \frac{1}{M} [\sum_{i=1}^n (k_i^{out})^2]^2\}}}, \quad (2.95)$$

and

$$r_{bin}^{out-out} = \frac{\frac{1}{2} \sum_{i=1}^n [(k_i^{out})^2 k_{nn,i}^{out-out} + k_i^{out} k_i^{in} k_{nn,i}^{in-out}] - \frac{1}{M} [\sum_{i=1}^n (k_i^{out})^2 \sum_{i=1}^n (k_i^{in} k_i^{out})]}{\sqrt{\{\sum_{i=1}^n (k_i^{out})^3 - \frac{1}{M} [\sum_{i=1}^n (k_i^{out})^2]^2\} \{\sum_{i=1}^n (k_i^{out})^2 k_i^{in} - \frac{1}{M} [\sum_{i=1}^n (k_i^{in} k_i^{out})]^2\}}}. \quad (2.96)$$

Local Assortativity

The concept of local assortativity stems from the demand to calculate the (unbiased) contribution of individual nodes to the overall (global) assortativity. The basic idea is that the numerator in the Pearson correlation coefficient proposed by Newman (2003) can be reformulated based on the contribution of individual nodes instead of edges (e.g. Piraveenan et al., 2010; Piraveenan et al., 2012).

It should be emphasized that, for the directed version of the measure of local assortativity primarily introduced in Piraveenan et al. (2012), the two in-out and out-in degree dependencies are not differentiated, when in fact they exhibit totally different behaviors (as found in Foster et al. (2010) and in Sec. 2.3 of our study). In our study, the contributions to the in-out and out-in degree dependencies are distinguishable.

We denote the local assortativity measures for a given node i as ρ_i^{in-in} , ρ_i^{in-out} , ρ_i^{out-in} , and $\rho_i^{out-out}$ corresponding to the four mixing categories in the directed version and ρ_i^{un} is used for the undirected version. Note that the following equalities must hold:

$$r_{bin}^{un} = \sum_{i=1}^n \rho_i^{un}, \quad (2.97)$$

$$r_{bin}^{in-in} = \sum_{i=1}^n \rho_i^{in-in}, \quad (2.98)$$

$$r_{bin}^{in-out} = \sum_{i=1}^n \rho_i^{in-out}, \quad (2.99)$$

$$r_{bin}^{out-in} = \sum_{i=1}^n \rho_i^{out-in}, \quad (2.100)$$

$$r_{bin}^{out-out} = \sum_{i=1}^n \rho_i^{out-out}. \quad (2.101)$$

First, we define

$$\mu_{un} = \frac{1}{2m} \sum_{i=1}^n (k_i^{un})^2, \quad (2.102)$$

$$\mu_{in-in} = \frac{1}{M} \sum_{i=1}^n (k_i^{in})^2, \quad (2.103)$$

$$\mu_{out-out} = \frac{1}{M} \sum_{i=1}^n (k_i^{out})^2, \quad (2.104)$$

and

$$\mu_{in-out} = \mu_{out-in} = \frac{1}{M} \sum_{i=1}^n (k_i^{in} k_i^{out}). \quad (2.105)$$

Note that, in the undirected case, it can be shown that μ_{un} is equal to the average of the degrees of the target and source nodes in the edge list $\{A_e B_e\}_{e=1}^m$, i.e. $\mu_{un} = \frac{1}{2m} (\sum_{e=1}^m k_{A_e}^{un} + \sum_{e=1}^m k_{B_e}^{un})$. Similarly, in the directed case, given the edge list the edge list $\{A_e B_e\}_{e=1}^M$, it can be shown that μ_{in-in} and $\mu_{out-out}$ are respectively equal to the averages of the incoming and out-going degrees from target and source nodes in the edge list. Mathematically, $\mu_{in} = \frac{1}{M} \sum_{e=1}^M k_{B_e}^{in}$ and $\mu_{out} = \frac{1}{M} \sum_{e=1}^M k_{A_e}^{out}$. In contrast, μ_{in-out} (μ_{out-in}) tells us the average of out-going (in-coming) degrees of the target (source) nodes in the edge list. We have that $\mu_{in-out} = \frac{1}{M} \sum_{e=1}^M k_{A_e}^{in}$ and $\mu_{out-in} = \frac{1}{M} \sum_{e=1}^M k_{B_e}^{out}$.

Second, we define

$$\sigma_{un}^2 = \sum_{i=1}^n (k_i^{un})^3 - \frac{1}{2m} [\sum_i (k_i^{un})^2]^2. \quad (2.106)$$

$$\sigma_{in}^2 = \sum_{i=1}^n (k_i^{in})^3 - \frac{1}{M} [\sum_i (k_i^{in})^2]^2 \quad (2.107)$$

$$\sigma_{out}^2 = \sum_{i=1}^n (k_i^{out})^3 - \frac{1}{M} [\sum_i (k_i^{out})^2]^2, \quad (2.108)$$

$$\sigma_{in'}^2 = \sum_{i=1}^n (k_i^{in})^2 k_i^{out} - \frac{1}{M} \left[\sum_{i=1}^n (k_i^{in} k_i^{out})^2 \right], \quad (2.109)$$

and

$$\sigma_{out'}^2 = \sum_{i=1}^n (k_i^{out})^2 k_i^{in} - \frac{1}{M} \left[\sum_{i=1}^n (k_i^{in} k_i^{out})^2 \right]. \quad (2.110)$$

The denominators in Eqs. (2.92), (2.93), (2.94), (2.95), (2.96) are respectively equal to σ_{un}^2 , $\sigma_{in}\sigma_{in'}$, $\sigma_{in'}\sigma_{out'}$, $\sigma_{out}\sigma_{in}$, and $\sigma_{out}\sigma_{out'}$.

By decomposing the overall assortativity coefficient r_{bin}^{un} in Eq. (2.92), we obtain the local assortativity indicators. More specifically, the contribution of node i to r is

$$\rho_i^{un} = \frac{(k_i^{un})^2 k_{nn,i}^{un} - (k_i^{un})^2 \mu_{un}}{\sigma_{un}^2}. \quad (2.111)$$

Similarly, in the directed case, for each node i , we have four local assortativity indicators:

$$\rho_i^{in-in} = \frac{k_i^{in} [k_i^{in} * (k_{nn,i}^{in-in} - \mu_{in-out}) + k_i^{out} (k_{nn,i}^{out-in} - \mu_{in-in})]}{2\sigma_{in}\sigma_{in'}}, \quad (2.112)$$

$$\rho_i^{in-out} = \frac{[k_i^{in} k_i^{out} (k_{nn,i}^{out-out} + k_{nn,i}^{in-in}) - 2k_i^{in} k_i^{out} * \mu_{in-out}]}{2\sigma_{out'}\sigma_{in'}}, \quad (2.113)$$

$$\rho_i^{out-in} = \frac{[(k_i^{out})^2 * (k_{nn,i}^{out-in} - \mu_{in-in}) + (k_i^{in})^2 (k_{nn,i}^{in-out} - \mu_{out-out})]}{2\sigma_{out}\sigma_{in}}, \quad (2.114)$$

$$\rho_i^{out-out} = \frac{k_i^{out} [k_i^{out} * (k_{nn,i}^{out-out} - \mu_{out-in}) + k_i^{in} (k_{nn,i}^{in-out} - \mu_{out-out})]}{2\sigma_{out}\sigma_{out'}}. \quad (2.115)$$

2.7.2 z-scores analysis of the indicators of structural correlations

In the main text, in the undirected weighted network, the UWCM is compared with the UECM. Similarly, in the directed weighted network, the DWCM is compared with the DECM. For that purpose, we employ z-scores evaluated under each referenced null model, generally defined as

$$z_X^{\text{null model}} = \frac{X - \langle X \rangle_{\text{null model}}}{\sigma[X]_{\text{null model}}} \quad (2.116)$$

where X is a measured quantity of the observed network, $\langle X \rangle_{\text{null model}}$ and $\sigma[X]_{\text{null model}}$ are respectively the expected value and the standard deviation of X evaluated under the referenced null model. Obviously, the interpretation of the statistical significance of the

discrepancy between quantity X and its expected value is valid if and only if X follows a Gaussian; however, the value of $z_X^{\text{null model}}$ can still tell us by how many standard deviations the value of X in the observed network differs from the expected one (see, for example, Squartini et al., 2013).

As shown in the main text, one can define z-scores for the local indicators such as for the ANNSs as well as the local weighted clustering coefficients, and can then compare different models for every bank (e.g. Eqs. (2.88), (2.89), (2.90), (2.91) in the main text). We can also define such scores for global indicators such as $\bar{s}_{nn}^{un}, r_w^{un}, \bar{C}_{un}^w$ in the undirected weighted network (under the UWCM and the UECM), and for $\bar{s}_{nn}^{in-in}, \bar{s}_{nn}^{in-out}, \bar{s}_{nn}^{out-in}, \bar{s}_{nn}^{out-out}, r_w^{in-in}, r_w^{in-out}, r_w^{out-in}, r_w^{out-out}, \bar{C}_w^{cyc}, \bar{C}_w^{mid}, \bar{C}_w^{in},$ and \bar{C}_w^{out} in the directed weighted network (under the DWCM and the DECM).

Chapter 3

An Approach to Identify Patterns in Structural Similarities in Financial Networks

Keywords: Interbank Network; Bank-Firm Credit Network; Structural Similarity; Portfolio Similarity; Null Models.

3.1 Introduction

Structural similarity (homophily) and dissimilarity (heterophily) between nodes are the fundamental issues in network science. Given two vertices, a simple way to investigate the similarity between them is to compare their neighborhoods, i.e. two vertices are structurally similar if they are linked to many common partners. A number of applications based on the structural similarity have been proposed such as for community detection (e.g. Fortunato, 2010; Li et al., 2011), ranking and centrality of vertices (e.g. Zhu et al., 2012; Alvarez-Socorro et al., 2015), growing networks (e.g. Papadopoulos et al., 2012), reconstruction of propagation networks (Liao and Zeng, 2015), and so forth.

Overlap and similarity in financial portfolios have received remarkable attention. It is widely suggested that if portfolios are strongly overlapping, adjustments in the portfolio of an agent may have a large impact on the financial wealth of other agents as well (e.g. Huang et al., 2013; Caccioli et al., 2014; Greenwood et al., 2015). In addition, it is also suggested that, more similar portfolios may make the financial system less stable (e.g. Wagner, 2008; Acharya, 2009; Wagner 2010; Haldane and May, 2011).

A number of measures for the structural similarity between nodes have been proposed (e.g. see Leicht et al., 2006; Lu and Zhou, 2011; Alvarez-Socorro et al., 2015 for a more detailed explanation of various (dis-)similarity measures). Such measures are also widely used in the analysis of financial portfolios, such as in the studies of Brechler et al. (2014), Pool et al. (2015), Sias et al. (2015), Cai et al. (2016), Fricke (2016), and Blocher (2016). However, surprisingly, there has been very little research in answering the question of whether the structural similarity is just a purely random consequence of some fundamental network properties (e.g. whether it is a consequence of information embedded in the local constraints like the degree and/or strength sequences, or the global constraints like the average of degrees or the average of weights). In addition, in many cases, due to privacy issues, we may not be able to obtain a full knowledge of real networks. That leads to another important question of whether the main properties of the structural similarity of real networks can be predicted or reconstructed from limited information (e.g. Cimini et al., 2015a; Cimini et al., 2015b).

In this study, we propose a network-based method to filter the effects of fundamental constraints on the structural similarity in real networks. Practically, we compare the observed structural similarity with the expected one obtained from various null models and then transform the observed one by filtering out the effects of the preserved constraints on it. In particular, we consider the random graph models specifying only the average of links or the average of weights, and the configuration models based on the maximum entropy method capturing the intrinsic heterogeneity in the observed degree and/or strength sequences (e.g. Squartini et al., 2011a; Squartini et al., 2011b; Squartini and Garlaschelli, 2011; Mastrandrea et al., 2014; Squartini et al., 2015). We apply our method to identify patterns in similarities between banks' loan portfolios in two different real world financial networks, i.e. the Italian electronic market for interbank deposits (e-MID) from 1999 to 2010 (quarterly data) and the bank-firm credit network of Spain from 1997 to 2007 (yearly data).

Perhaps, among other studies on the applications of configuration models in identifying patterns in structural properties of real networks, the studies of Squartini et al. (2013), Gemmetto and Garlaschelli (2015), Gemmetto et al. (2015), and Luu and Lux (2016a) are closest to this study. However, we should emphasize that unlike our study where we focus on the similarity in node neighborhoods, Squartini et al. (2013) investigate the reciprocity between nodes in directed networks ¹, while the studies of Gemmetto and Garlaschelli (2015),

¹Generally speaking, reciprocity indicates the structure of pairs of links (in the binary case) or pairs of weights (in the weighted case) between nodes in both directions. More specifically, given an adjacency matrix and a weighted matrix defined in Sec. 3.2, the reciprocity structure between every two nodes i and j shows the relation between a_{ij} and a_{ji} in the binary case or between w_{ij} and w_{ji} in the weighted case.

Gemmetto et al. (2015), and Luu and Lux (2016a) focus on the overlaps and similarities between layers in multilayered networks ².

The remainder of this paper is structured as follows. In Sec. 3.2 we briefly describe the method used for the analysis of the structural similarity in undirected one-mode networks. Sec. 3.3 reports our main findings for the binary as well as weighted versions of the bank-firm credit market of Spain and the e-MID network. Discussions and concluding remarks are in Sec. 3.4. At the end of this paper, the Appendix provides the relevant methodological details for Sec. 3.2 as well as some additional results for Sec. 3.3.

3.2 Similarity and dissimilarity in networks

3.2.1 General definitions and configuration models

3.2.1.1 General definitions

To begin, let us briefly introduce some general notations and definitions for an undirected one-mode network. Suppose that we have a network of n nodes characterized by an adjacency matrix $A = \{a_{ij}\}_{n \times n}$ and a weighted matrix $W = \{w_{ij}\}_{n \times n}$ ($a_{ii} = w_{ii} = 0, \forall i$ from 1 to n). The degree and strength sequences are respectively defined as

$$k_i = \sum_{j=1}^n a_{ij}, \quad (3.1)$$

and

$$s_i = \sum_{j=1}^n w_{ij}. \quad (3.2)$$

The total degree and total strength over all nodes in the network are respectively given by

$$L_{\text{tol}} = \sum_{1 \leq i < j \leq n} a_{ij}, \quad (3.3)$$

$$W_{\text{tol}} = \sum_{1 \leq i < j \leq n} w_{ij}. \quad (3.4)$$

²Mathematically, these studies compare the adjacency matrices in the binary case (or the weighted matrices in the weighted case) in different layers.

3.2.1.2 Configuration models and latent variables

For random graph models (RGM), only the global constraints of the observed network are specified. Particularly, the average of links is controlled in the so-called Binary Random Graph (BRG) model, while in the so-called Weighted Random Graph model (WRG) the average of weights is preserved. In contrast, in the family of configuration models (CM), the observed node degrees and/or the observed node strengths are maintained (e.g. Squartini et al., 2011a; Squartini et al., 2011b; Squartini et al., 2015)³. Therefore, the intrinsic heterogeneity in the degree sequence and/or strength sequence is controlled in the configuration models. In the following, let us introduce the algorithm used for solving hidden variables in each configuration model.

In Undirected Binary Configuration Model ((U)BCM)

Under the BCM, the observed degree sequence $\{k_i\}_{i=1}^n$ is given. One needs to solve the following system of n equations to obtain n non-negative hidden variables $\{x_i^*\}_{i=1}^n$

$$\sum_{j \neq i} \frac{x_i^* x_j^*}{1 + x_i^* x_j^*} = k_i, \forall i = 1, 2, \dots, n. \quad (3.5)$$

The expected link between node i and node j is given by

$$\langle a_{ij} \rangle = p_{ij} = \frac{x_i^* x_j^*}{1 + x_i^* x_j^*}, \forall 1 \leq i \neq j \leq n. \quad (3.6)$$

Each element a_{ij} will follow a Bernoulli distribution with parameter p_{ij}

$$q(a_{ij}) = p_{ij}^{a_{ij}} (1 - p_{ij})^{1 - a_{ij}}. \quad (3.7)$$

In Undirected Weighted Configuration Model ((U)WCM)

Under the WCM, given the strength sequence $\{s_i\}_{i=1}^n$ of the observed network, we obtain n hidden variables $\{x_i^*\}_{i=1}^n$ ($\{x_i^*\}_{i=1}^n \in [0, 1)$) by solving

$$\sum_{j \neq i} \frac{x_i^* x_j^*}{1 - x_i^* x_j^*} = s_i, \forall i = 1, 2, \dots, n. \quad (3.8)$$

³Note that the BRG and WRG models are respectively the special cases of the BCM and WCM models (e.g. see Squartini et al., 2015).

The expected weight between node i and node j is given by

$$\langle w_{ij} \rangle = \frac{x_i^* x_j^*}{1 - x_i^* x_j^*}, \forall 1 \leq i \neq j \leq n. \quad (3.9)$$

The probability of a link of weight w_{ij} between node i and node j in this model is

$$q(w_{ij}) = \begin{cases} 1 - p_{ij}, & \text{if } w_{ij} = 0, \\ (p_{ij})^{w_{ij}} (1 - p_{ij}), & \text{if } w_{ij} > 0, \end{cases} \quad (3.10)$$

where p_{ij} depends on hidden variables x_i^* and x_j^*

$$p_{ij} = \langle a_{ij} \rangle = x_i^* x_j^*. \quad (3.11)$$

In Undirected Enhanced Configuration Model ((U)ECM)

Under the ECM, both the degree sequence $\{k_i\}_{i=1}^n$ and the strength sequence $\{s_i\}_{i=1}^n$ are preserved. The non-negative hidden variables $\{x_i^*\}_{i=1}^n$ and $\{y_i^*\}_{i=1}^n$ ($\{y_i^*\}_{i=1}^n \in [0, 1)$) are the solution to the system of $2n$ equations

$$\begin{aligned} \sum_{j \neq i} \frac{x_i^* x_j^* y_i^* y_j^*}{1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*} &= k_i, \forall i = 1, 2, \dots, n, \\ \sum_{j \neq i} \frac{x_i^* x_j^* y_i^* y_j^*}{(1 - y_i^* y_j^*)(1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*)} &= s_i, \forall i = 1, 2, \dots, n. \end{aligned} \quad (3.12)$$

The probability of a link between node i and node j is

$$p_{ij} = \langle a_{ij} \rangle = \frac{x_i^* x_j^* y_i^* y_j^*}{1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*}, \forall 1 \leq i \neq j \leq n. \quad (3.13)$$

The expected weight between node i and node j is

$$\langle w_{ij} \rangle = \frac{x_i^* x_j^* y_i^* y_j^*}{(1 - y_i^* y_j^*)(1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*)}, \forall 1 \leq i \neq j \leq n. \quad (3.14)$$

The probability of a link of weight w_{ij} is now given by

$$q(w_{ij}) = \begin{cases} 1 - p_{ij}, & \text{if } w_{ij} = 0, \\ p_{ij} r_{ij}^{w_{ij}-1} (1 - r_{ij}), & \text{if } w_{ij} > 0. \end{cases} \quad (3.15)$$

(where $r_{ij} = y_i^* y_j^*$).

Generally speaking, when hidden variables are solved under the referenced configuration model, we can use them to generate an ensemble \mathcal{G} of randomized networks that enforce the initial constraints of the observed network G^* on average. Note that, the average of any network property can be calculated over the ensemble (e.g. see Squartini et al., 2015). In addition, later we will see that, in fact, in a certain number of cases, the expected values of network properties can analytically be approximated/calculated via hidden variables.

3.2.2 Structural similarity

In this part, we will introduce a method for measuring the structural similarity in an undirected one-mode network. First, we will define the raw similarity for every pair of nodes in the observed network. Generally speaking, it indicates the overall level of similarity in the neighborhoods of two nodes. After that, we rescale this measure by filtering out the effects of the heterogeneity in the observed degrees and/or strengths. Next, in each referenced null model, we will introduce the z-score showing the dispersion between the raw similarity and the expected one. In the following, we will show that, in fact the rescaled measures as well as the z-scores evaluated under the various configuration models can analytically be calculated via hidden variables extracted from the associated constraints.

Note that, in the following we will focus on the minimum-based similarity, but further extensions to the maximum-based dissimilarity and to the generalized Jaccard index are straightforward. Further details about these extensions as well as the relevant proofs are mentioned in the Appendix.

3.2.2.1 Structural similarity in a binary network

Before going into details, it should be emphasized that, in our study we set the diagonal elements of the adjacency matrix, of the similarity matrices, of the rescaled similarity matrices, and of the z-score matrices to zero. In addition, for the special case that two nodes i and j are the only neighbors of each other, i.e. when $k_i = k_j = a_{ij} = 1$ ⁴, the structural similarity between them measured by Eq. (3.16) is equal to zero⁵.

Raw Similarity

⁴However, in our specific data sets, this is not often the case.

⁵Or in this case, one can define the similarity indicator similar to the modified Sørensen-Dice index proposed in Alvarez-Socorro et al. (2015), where both maximum and minimum based similarity between these two nodes is equal to 1.

For the binary version, the raw similarity between two different nodes i and j is defined as

$$S_{i,j}^{\text{bin},\text{min}} = \frac{2 \sum_k \min(a_{ik}, a_{jk})}{k_i + k_j}. \quad (3.16)$$

Note that the quantity in Eq. (3.16) is similar to the Sørensen-Dice similarity index (Dice, 1945; Sørensen, 1948). We can see that the the numerator on the right hand side of Eq. (3.16) captures the overall overlap between node neighborhoods, while the denominator is used to normalize that overlap such that the minimum-based similarity indicator $S_{i,j}^{\text{bin},\text{min}} \in [0, 1]$. For each pair of two nodes (i, j), the larger $S_{i,j}^{\text{bin},\text{min}}$, the higher the similarity between i and j . In addition, $S_{i,j}^{\text{bin},\text{min}}$ is maximally equal to 1 if i and j have the same neighborhood, and minimally equal to 0 when there is no overlap between their neighborhoods at all.

We can also define the maximum-based indicator as

$$S_{i,j}^{\text{bin},\text{max}} = \frac{2 \sum_k \max(a_{ik}, a_{jk})}{k_i + k_j}. \quad (3.17)$$

In contrast to the indicator $S_{i,j}^{\text{bin},\text{min}}$, the indicator $S_{i,j}^{\text{bin},\text{max}}$ now ranges in $[1, 2]$. Since the larger $S_{i,j}^{\text{bin},\text{max}}$, the lower similarity between two nodes i and j , we consider this indicator as a maximum-based dissimilarity measure.

We can also define the generalized Jaccard-based similarity between two nodes as

$$J_{i,j}^{\text{bin}} = \frac{\sum_k \min(a_{ik}, a_{jk})}{\sum_k \max(a_{ik}, a_{jk})}. \quad (3.18)$$

It is easy to show that

$$J_{i,j}^{\text{bin}} = \frac{S_{i,j}^{\text{bin},\text{min}}}{S_{i,j}^{\text{bin},\text{max}}}. \quad (3.19)$$

Rescaled similarity

In the binary version, for each referenced null model, the general form of the rescaled similarity $\mu[S_{i,j}^{\text{bin},\text{min}}]_{\text{null-model}}$ is defined as

$$\mu[S_{i,j}^{\text{bin},\text{min}}]_{\text{null-model}} = \frac{S_{i,j}^{\text{bin},\text{min}} - \langle S_{i,j}^{\text{bin},\text{min}} \rangle_{\text{null-model}}}{1 - \langle S_{i,j}^{\text{bin},\text{min}} \rangle_{\text{null-model}}} \quad (3.20)$$

where the notation $\langle X \rangle_{\text{null-model}}$ indicates the expected value of X evaluated under the null model ⁶.

⁶Throughout this paper, the notations $\langle X \rangle_{\text{null-model}}$ and $\mu[Y]_{\text{null-model}}$ respectively indicate the average of X and the rescaled quantity of Y evaluated under the referenced null model.

In the rescaled similarity, the effects of latent information encoded in the constraints (maintained in the null model) on the raw similarity are filtered out. It is easy to show that $\mu[S_{i,j}^{\text{bin,min}}]_{\text{null-model}} \leq 1$.

In particular, under the BCM the expectation of $S_{i,j}^{\text{bin,min}}$ is

$$\langle S_{i,j}^{\text{bin,min}} \rangle_{BCM} = \frac{2 \sum_k \langle \min(a_{ik}, a_{jk}) \rangle_{BCM}}{k_i + k_j}. \quad (3.21)$$

We can prove that, in the binary case, $\langle \min(a_{ik}, a_{jk}) \rangle_{BCM} = p_{ik}p_{jk}$ (see the Appendix), where p_{ik} is the probability of a link between two different nodes i and k , and $p_{ik} = \frac{x_i^* x_k^*}{1+x_i^* x_k^*}$ as in Eq. (3.6) ($1 \leq i \neq k \leq n$), and $\{x_i^*\}_{i=1}^n$ are hidden variables extracted from the degree sequence as in Eqs. (3.5). Therefore,

$$\langle S_{i,j}^{\text{bin,min}} \rangle_{BCM} = \frac{2 \sum_k p_{ik}p_{jk}}{k_i + k_j}. \quad (3.22)$$

It follows from Eqs. (3.16), (3.20), and (3.22) that the rescaled similarity in the BCM is analytically given by

$$\mu[S_{i,j}^{\text{bin,min}}]_{BCM} = \frac{2 \sum_k \min\{a_{ik}, a_{jk}\} - 2 \sum_k p_{ik}p_{jk}}{k_i + k_j - 2 \sum_k p_{ik}p_{jk}}. \quad (3.23)$$

Binary z scores

For every pair of nodes i and j , under each referenced null model, we measure the z score associated to $S_{i,j}^{\text{bin,min}}$ as

$$z[S_{i,j}^{\text{bin,min}}]_{\text{null-model}} = \frac{S_{i,j}^{\text{bin,min}} - \langle S_{i,j}^{\text{bin,min}} \rangle_{\text{null-model}}}{\sigma[S_{i,j}^{\text{bin,min}}]_{\text{null-model}}}, \quad (3.24)$$

with

$$\sigma[S_{i,j}^{\text{bin,min}}]_{\text{null-model}} = \frac{2\sigma[\sum_k \min\{a_{ik}, a_{jk}\}]_{\text{null-model}}}{k_i + k_j}, \quad (3.25)$$

and

$$\sigma[\min\{a_{ik}, a_{jk}\}]_{\text{null-model}} = \sqrt{\langle \min\{a_{ik}, a_{jk}\}^2 \rangle_{\text{null-model}} - \langle \min\{a_{ik}, a_{jk}\} \rangle_{\text{null-model}}^2}, \quad (3.26)$$

where the notation $\sigma[X]_{\text{null-model}}$ indicates the standard deviation of X evaluated under the referenced null model ⁷.

⁷Throughout this paper, the notation $\sigma[X]_{\text{null-model}}$ stands for the standard deviation of X , and

For every pair of nodes i and j , $z[S_{i,j}^{\text{bin},\text{min}}]_{\text{null-model}}$ shows how many standard deviations the raw value $S_{i,j}^{\text{bin},\text{min}}$ deviates from the expected value $\langle S_{i,j}^{\text{bin},\text{min}} \rangle$. Therefore, comparing to $\mu[S_{i,j}^{\text{bin},\text{min}}]_{\text{null-model}}$, $z[S_{i,j}^{\text{bin},\text{min}}]_{\text{null-model}}$ is more informative about the discrepancy between the observed similarity and the expected one obtained from the referenced null model. In addition, we can use $z[S_{i,j}^{\text{bin},\text{min}}]_{\text{null-model}}$ to assess, for example, whether small values of the elements of $\mu[S_{i,j}^{\text{bin},\text{min}}]_{\text{null-model}}$ (in the absolute terms) are statistically significant (e.g. Gemmetto and Garlaschelli, 2015; Gemmetto et al., 2015).

Particularly, under the BCM model, it can be shown that

$$\langle \min\{a_{ik}, a_{jk}\}^2 \rangle_{BCM} = p_{ik}p_{jk}, \quad (3.27)$$

and

$$\sigma[\min\{a_{ik}, a_{jk}\}]_{BCM} = \sqrt{p_{ik}p_{jk} - (p_{ik}p_{jk})^2}, \quad (3.28)$$

where $p_{ik} = \frac{x_i^* x_k^*}{1+x_i^* x_k^*}$ as in Eq. (3.6) (see the Appendix). Therefore, from Eqs. (3.16), (3.22), (3.24), and (3.28) we have

$$z[S_{i,j}^{\text{bin},\text{min}}]_{BCM} = \frac{\sum_k \min\{a_{ik}, a_{jk}\} - \sum_k p_{ik}p_{jk}}{\sqrt{\sum_k [p_{ik}p_{jk} - (p_{ik}p_{jk})^2]}}. \quad (3.29)$$

3.2.2.2 Structural similarity in a weighted network

Before going into details, in the weighted case, the diagonal elements of the weighted matrix, of the similarity matrices, of the rescaled similarity matrices, and of the z-score matrices are also equal to zero. For the special case that two nodes i and j are the only neighbors of each other (i.e. when $s_i = s_j = w_{ij}$ ⁸), the structural similarity between them measured by Eq. (3.30) is equal to zero.

Raw similarity

For an undirected weighted network, we define the minimum-based similarity for every pair of nodes i and j as

$$S_{i,j}^{\text{w},\text{min}} = \frac{2 \sum_k \min(w_{ik}, w_{jk})}{s_i + s_j}, \quad (3.30)$$

and the maximum-based dissimilarity as

$$S_{i,j}^{\text{w},\text{max}} = \frac{2 \sum_k \max(w_{ik}, w_{jk})}{s_i + s_j}. \quad (3.31)$$

$z[X]_{\text{null-model}}$ stands for the z score of X evaluated under the referenced null model.

⁸However, similar to the binary version, this is not often the case in our data sets.

The interpretations for $S_{i,j}^{\text{w},\text{min}}$ and $S_{i,j}^{\text{w},\text{max}}$ are respectively similar to the interpretations for $S_{i,j}^{\text{b},\text{min}}$ and $S_{i,j}^{\text{b},\text{max}}$ in the binary case. Clearly, when $w_{ij} = a_{ij}$ ($\forall i, j$), we immediately get the Eq. (3.16) and (3.17) from Eq. (3.30) and (3.31), respectively.

Note that we can obtain the generalized Jaccard index between two nodes in the weighted version (e.g. see Chierichetti et al., 2010) as

$$J_{i,j}^{\text{w}} = \frac{\sum_k \min(w_{ik}, w_{jk})}{\sum_k \max(w_{ik}, w_{jk})} = \frac{S_{i,j}^{\text{w},\text{min}}}{S_{i,j}^{\text{w},\text{max}}}. \quad (3.32)$$

Rescaled similarity

Generally, the rescaled similarity between two nodes i and j in the weighted version is defined as

$$\mu[S_{i,j}^{\text{w},\text{min}}]_{\text{null-model}} = \frac{S_{i,j}^{\text{w},\text{min}} - \langle S_{i,j}^{\text{w},\text{min}} \rangle_{\text{null-model}}}{1 - \langle S_{i,j}^{\text{w},\text{min}} \rangle_{\text{null-model}}}, \quad (3.33)$$

where $\langle S_{i,j}^{\text{w}} \rangle_{\text{null-model}}$ is the expected value of the weighted similarity under the chosen null model, which is given by

$$\langle S_{i,j}^{\text{w},\text{min}} \rangle_{\text{null-model}} = \frac{2 \sum_k \langle \min(w_{ik}, w_{jk}) \rangle_{\text{null-model}}}{s_i + s_j}. \quad (3.34)$$

Weighted z score

The z-score of the weighted similarity between two nodes i, j tells us by how many standard deviations the raw value $S_{i,j}^{\text{w},\text{min}}$ deviates from the expected value $\langle S_{i,j}^{\text{w},\text{min}} \rangle$. It is defined as

$$z[S_{i,j}^{\text{w},\text{min}}]_{\text{null-model}} = \frac{S_{i,j}^{\text{w},\text{min}} - \langle S_{i,j}^{\text{w},\text{min}} \rangle_{\text{null-model}}}{\sigma[S_{i,j}^{\text{w},\text{min}}]_{\text{null-model}}}, \quad (3.35)$$

where

$$\sigma[S^{\text{w},\text{min}}(i, j)]_{\text{null-model}} = \frac{2\sigma[\sum_k \min\{w_{ik}, w_{jk}\}]_{\text{null-model}}}{s_i + s_j}, \quad (3.36)$$

and

$$\sigma[\min\{w_{ik}, w_{jk}\}]_{\text{null-model}} = \sqrt{\langle \min\{w_{ik}, w_{jk}\}^2 \rangle_{\text{null-model}} - \langle \min\{w_{ik}, w_{jk}\} \rangle_{\text{null-model}}^2}. \quad (3.37)$$

We now consider the rescaled similarity and z score under the two null models, i.e. the WCM and the ECM. Similar to the binary case, we focus on the minimum-based similarity. Further extensions to the maximum-based dissimilarity and the generalized Jaccard index are explained in the Appendix.

Under the WCM

Under the WCM, as shown the Appendix, we have that

$$\langle \min(w_{ik}, w_{jk}) \rangle_{WCM} = \frac{p_{ik}p_{jk}}{1 - p_{ik}p_{jk}}, \quad (3.38)$$

with $p_{ik} = x_i^* x_k^*$ ($1 \leq i \neq k \leq n$) depending on hidden variables $\{x_i^*\}_{i=1}^n$ extracted from the strength sequence as in Eqs. (3.8). Therefore, from Eqs. (3.34) and (3.38) we have

$$\langle S_{ij}^{w,\min} \rangle_{WCM} = \frac{2 \sum_k \frac{p_{ik}p_{jk}}{(1-p_{ik}p_{jk})}}{s_i + s_j}. \quad (3.39)$$

It follows from Eqs. (3.30), (3.33), and (3.39) that the rescaled similarity evaluated under the WCM is

$$\mu[S_{i,j}^{w,\min}]_{WCM} = \frac{2 \sum_k \min\{w_{ik}, w_{jk}\} - 2 \sum_k \frac{p_{ik}p_{jk}}{(1-p_{ik}p_{jk})}}{s_i + s_j - 2 \sum_k \frac{p_{ik}p_{jk}}{(1-p_{ik}p_{jk})}}. \quad (3.40)$$

Moreover, we show in the Appendix that

$$\langle \min(w_{ik}, w_{jk})^2 \rangle_{WCM} = \frac{p_{ik}p_{jk}(1 + p_{ik}p_{jk})}{(1 - p_{ik}p_{jk})^2}. \quad (3.41)$$

From Eqs. (3.38), (3.41) we have

$$\sigma[\min(w_{ik}, w_{jk})]_{WCM}^2 = \frac{p_{ik}p_{jk}(1 + p_{ik}p_{jk})}{(1 - p_{ik}p_{jk})^2} - \left[\frac{p_{ik}p_{jk}}{(1 - p_{ik}p_{jk})} \right]^2 = \frac{p_{ik}p_{jk}}{(1 - p_{ik}p_{jk})^2}, \quad (3.42)$$

From Eqs. (3.30), (3.35), (3.39), and (3.42) we have

$$z[S_{i,j}^{w,\min}]_{WCM} = \frac{\sum_k \min\{w_{ik}, w_{jk}\} - \sum_k \frac{p_{ik}p_{jk}}{(1-p_{ik}p_{jk})}}{\sqrt{\sum_k \frac{p_{ik}p_{jk}}{(1-p_{ik}p_{jk})^2}}}. \quad (3.43)$$

Under the ECM

Under the ECM, as shown the Appendix, we have that

$$\langle \min(w_{ik}, w_{jk}) \rangle_{ECM} = \frac{p_{ik}p_{jk}}{1 - r_{ik}r_{jk}}, \quad (3.44)$$

where $p_{ij} = \frac{x_i^* x_j^* y_i^* y_j^*}{1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*}$ and $r_{ij} = y_i^* y_j^*$ ($1 \leq i \neq j \leq n$) depend on hidden variables extracted from the degree and strength sequences as in Eqs. (3.12).

Therefore, from Eqs. (3.34) and (3.44) we obtain the expected value of the weighted

similarity between every pair of nodes

$$\langle S_{ij}^{w,\min} \rangle_{ECM} = \frac{2 \sum_k \frac{p_{ik}p_{jk}}{(1-r_{ik}r_{jk})}}{s_i + s_j}. \quad (3.45)$$

It follows from Eqs. (3.30), (3.33), and (3.45) that the rescaled similarity under the ECM is

$$\mu[S_{i,j}^{w,\min}]_{ECM} = \frac{2 \sum_k \min \{w_{ik}, w_{jk}\} - 2 \sum_k \frac{p_{ik}p_{jk}}{1-r_{ik}r_{jk}}}{s_i + s_j - 2 \sum_k \frac{p_{ik}p_{jk}}{(1-r_{ik}r_{jk})}}. \quad (3.46)$$

In addition, as shown in the Appendix, we have

$$\langle \min(w_{ik}, w_{jk})^2 \rangle_{ECM} = \frac{p_{ik}p_{jk}(r_{ik}r_{jk} + 1)}{(1 - r_{ik}r_{jk})^2}. \quad (3.47)$$

From Eqs. (3.44), (3.47) we obtain

$$\sigma(\min(w_{ik}, w_{jk}))_{ECM}^2 = \frac{p_{ik}p_{jk}(r_{ik}r_{jk} + 1)}{(1 - r_{ik}r_{jk})^2} - \left(\frac{p_{ik}p_{jk}}{1 - r_{ik}r_{jk}}\right)^2 = \frac{p_{ik}p_{jk}(1 - p_{ik}p_{jk} + r_{ik}r_{jk})}{(1 - r_{ik}r_{jk})^2}. \quad (3.48)$$

From Eqs. (3.30), (3.35), (3.45), and (3.48), we get the z-score under the ECM as

$$z[S_{i,j}^{w,\min}]_{ECM} = \frac{\sum_k \min \{w_{ik}, w_{jk}\} - \sum_k \frac{p_{ik}p_{jk}}{(1-r_{ik}r_{jk})}}{\sqrt{\sum_k \left[\frac{p_{ik}p_{jk}(1-p_{ik}p_{jk}+r_{ik}r_{jk})}{(1-r_{ik}r_{jk})^2} \right]}}. \quad (3.49)$$

3.3 Findings

3.3.1 Binary analysis

3.3.1.1 Bank-firm credit network of Spain

First, we will investigate the bank-firm credit network of Spain over the period from 1997 to 2007. We project this binary bipartite network onto an un-weighted and undirected network of overlapping portfolio between banks, i.e. two banks are (indirectly) connected if they co-finance a firm.

In Figures (3.1) and (3.2), we respectively show the structural similarities in the two years 1997 and 2007 as examples. Let S^b be $S^{\text{bin},\min}$, in each year, the raw similarity matrix S^b is shown in panel (a). The color-coded matrix shows that some pairs of banks display a high level of similarities while some others show only a low level or no similarity at all.

We now compare the raw similarities with the rescaled ones obtained from the BRG

model (panel (b)) and from the BCM (panel (c)) (see Figures (3.1), (3.2)). Overall, there is no significant difference between S^b and $\mu[S^b]_{BRG}$. This indicates that specifying only the average of links in the BRG model can not replicate the observed structural similarities. In contrast, the rescaled similarity matrix $\mu[S^b]_{BCM}$ shows that most of the similarities are subtracted when the effects of the heterogeneity in the degree distribution are filtered out. Note that, a more detailed comparison between the raw similarities and the rescaled ones under the BCM is shown in Figures (3.20) and (3.21) in the Appendix, where we can see a clear difference between the distribution of the raw similarities and the distribution of the rescaled similarities under the BCM.

Figure (3.3) shows the temporal evolution of the average of similarities (over the whole network) over time. We observe that the average of similarities starts to increase from 2000 to 2007. That may be because the expansion in the credit market leads to more overlaps in banks' loan portfolios. Moreover, the main trend of the observed average is also reproduced by the BCM, although there are some slight deviations from ± 2 std. from the expected value in several years. Unfortunately, since we do not have information for years after 2007 for this data set, it is impossible to assess whether there is any structural change in portfolio overlaps and similarities due to the financial crisis.

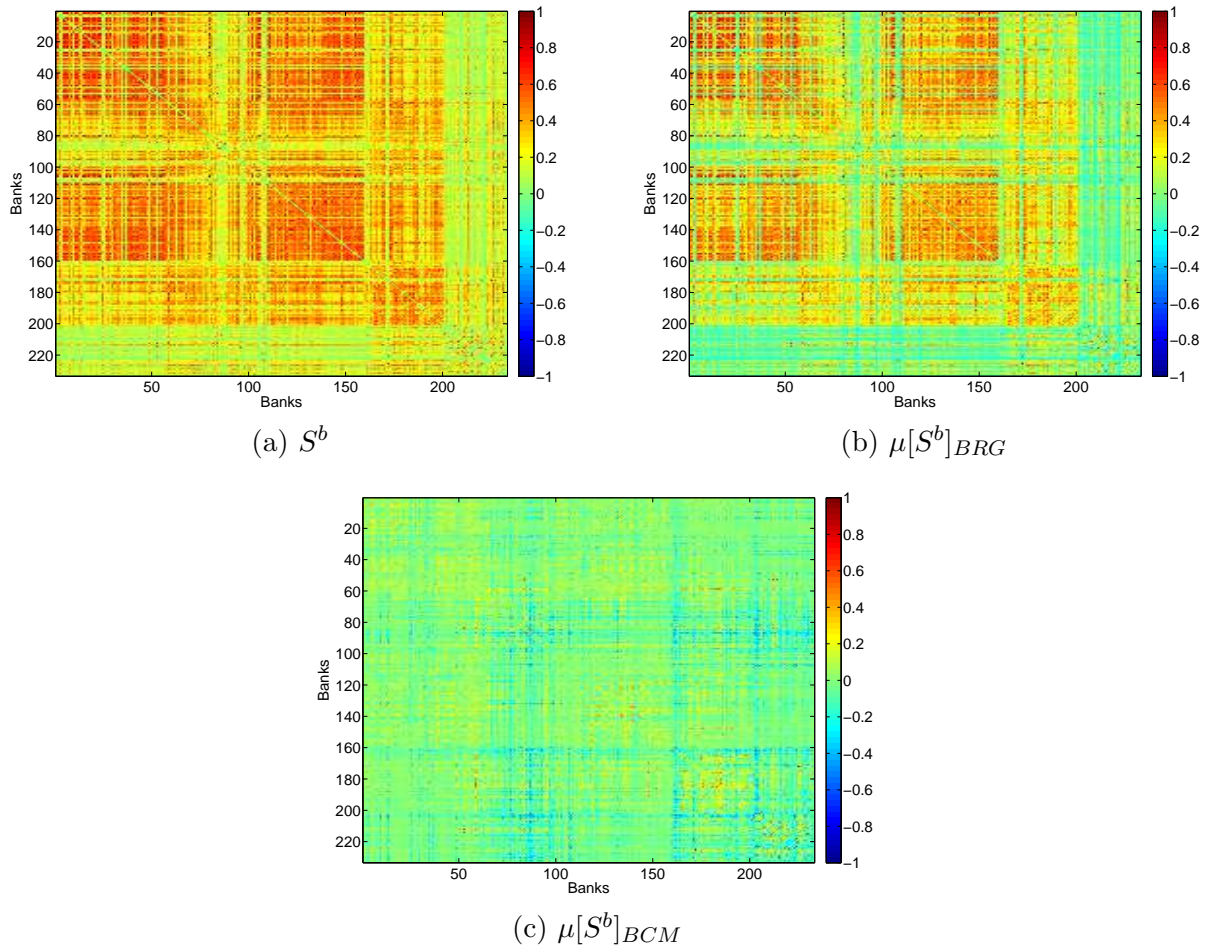


Figure 3.1: Binary structural similarities in the bank-firm credit network of Spain in 1997. Panel (a): raw similarity matrix. Panel (b): rescaled similarity matrix under the BRG. Panel (c): rescaled similarity matrix under the BCM.

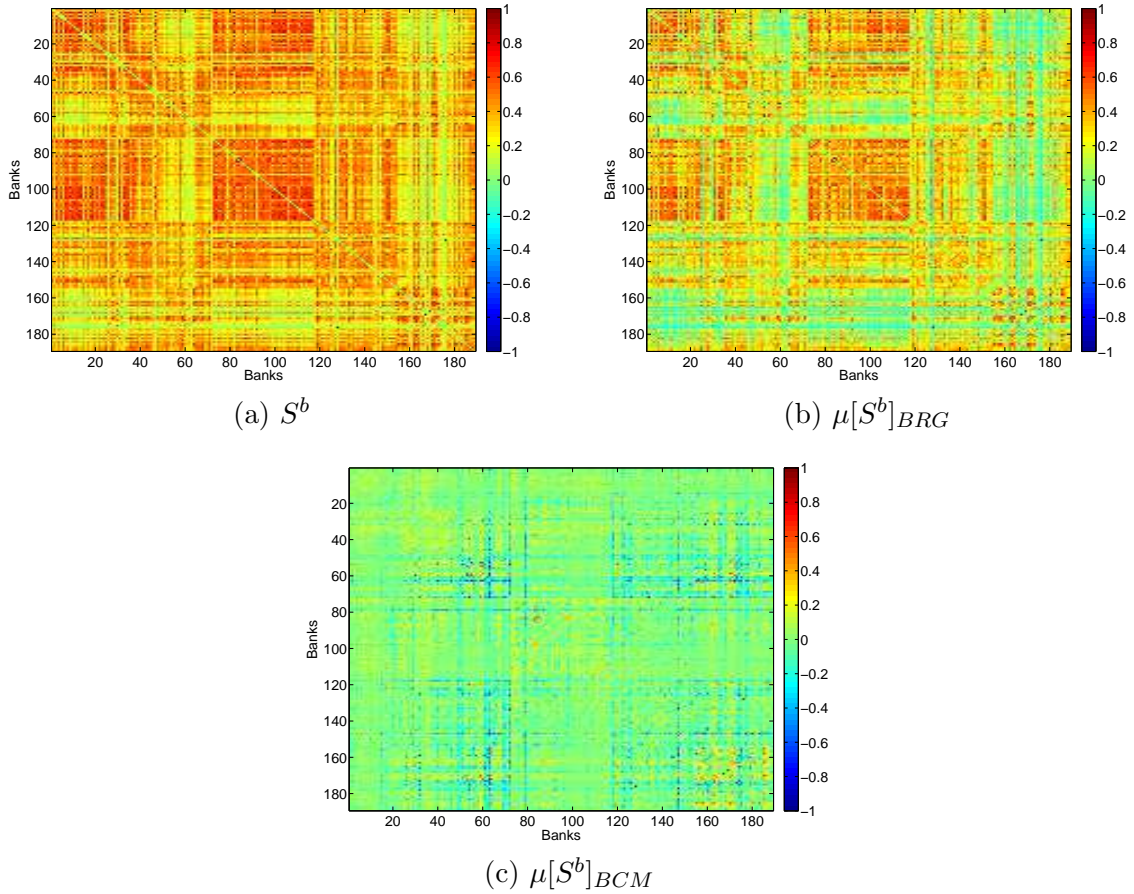


Figure 3.2: Binary structural similarities in the bank-firm credit network of Spain in 2007. Panel (a): raw similarity matrix. Panel (b): rescaled similarity matrix under the BRG. Panel (c): rescaled similarity matrix under the BCM.

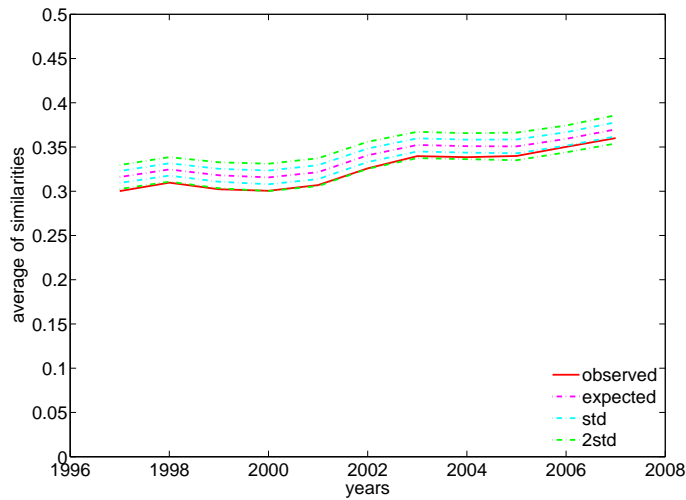


Figure 3.3: Evolution of the average of structural similarities in the bank-firm credit network of Spain from 1997 to 2007, in the binary case.

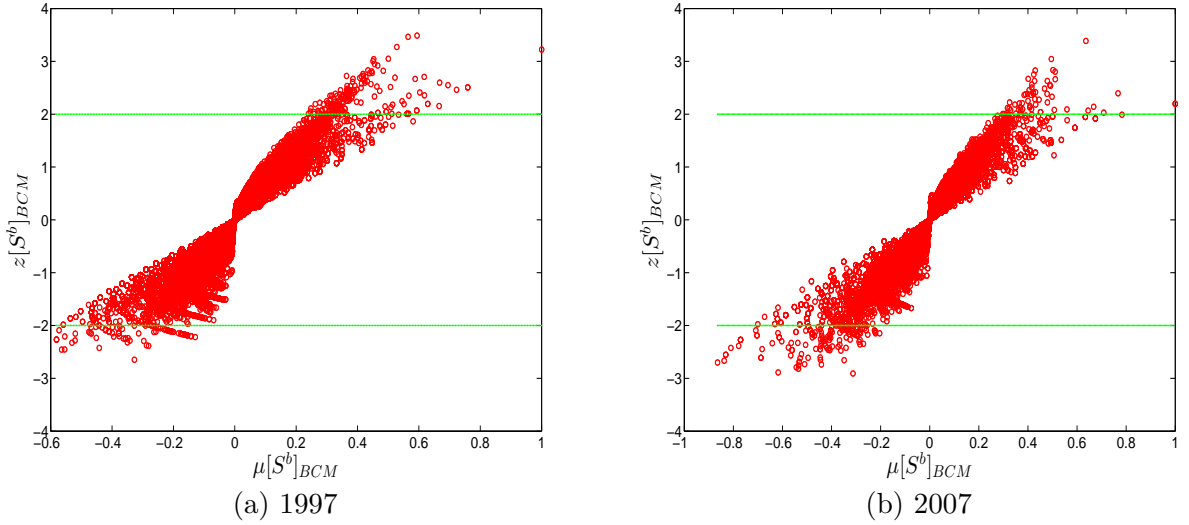


Figure 3.4: z-score analysis in the bank-firm credit network of Spain in 1997 (panel (a)) and 2007 (panel (b)) under the BCM. In each panel, we plot $z[S^b]_{BCM}$ vs. $\mu[S^b]_{BCM}$, and the green dashed lines represent the critical values $z^* = \pm 2$.

To assess the statistical significance of the elements of the structural similarity matrix S^b , we now employ the z-score analysis. We will focus on the BCM, since the BRG model can not capture the main features of S^b ⁹. In figure (3.4), we plot the z score $z[S^b]_{BCM}$ against the rescaled similarity $\mu[S^b]_{BCM}$ in the two years 1997 and 2007. We find that, on the one hand, almost all z-scores are located within the range $[-2, 2]$, implying that almost all the pairwise similarities are actually driven by latent information embedded in the observed degree sequence, no more no less. On the other hand, we do observe the presence of non-random patterns in the similarities between several banks that can not be completely explained by such latent information.

3.3.1.2 Italian e-MID network

We now analyze the structural similarity in the binary version of the e-MID network. In Figures (3.5) and (3.6), we show the raw similarities as well as the rescaled similarities in two example quarters, i.e. the 1st quarter in 1999 and the 4th quarter in 2010. As shown in the panel (a) of each Figure, many pairs of banks have a high level of similarity in their portfolios. In addition, similar to the previous network, in the panel (b) of Figures (3.5) and (3.6) we find that most of these similarities still prevail even when the effects of the average of links are discarded. Furthermore, a significant difference between S^b and $\mu[S^b]_{BCM}$ (panel

⁹In comparison, in absolute terms, typically we observe that z-scores evaluated under the BRG model are also much higher than those evaluated under the BCM.

(c) of Figures (3.5) and (3.6)) again indicates a crucial role of the intrinsic heterogeneity in the observed degree sequence for the emergence of the structural similarities between banks (see Figures (3.22), (3.23) in the Appendix for a more detailed comparison between the distribution of raw similarities and the distribution of the rescaled ones under the BCM model in the e-MID network).

Figure (3.7) shows the temporal dynamics of the average of similarities from the 1st quarter in 1999 to the last quarter in 2010. In the period around the financial crisis, we find that the average of similarities is strongly decreasing from the third quarter of 2008, which can be considered as an effect of the financial crisis. Intuitively, at that time, banks may have cut down on their lending/borrowing relations so that the portfolio similarities are also affected on average. Moreover, in comparison to the BCM, typically, for almost all quarters, the observed value of the average of similarities is completely located inside the region of ± 2 std. from the expected one.

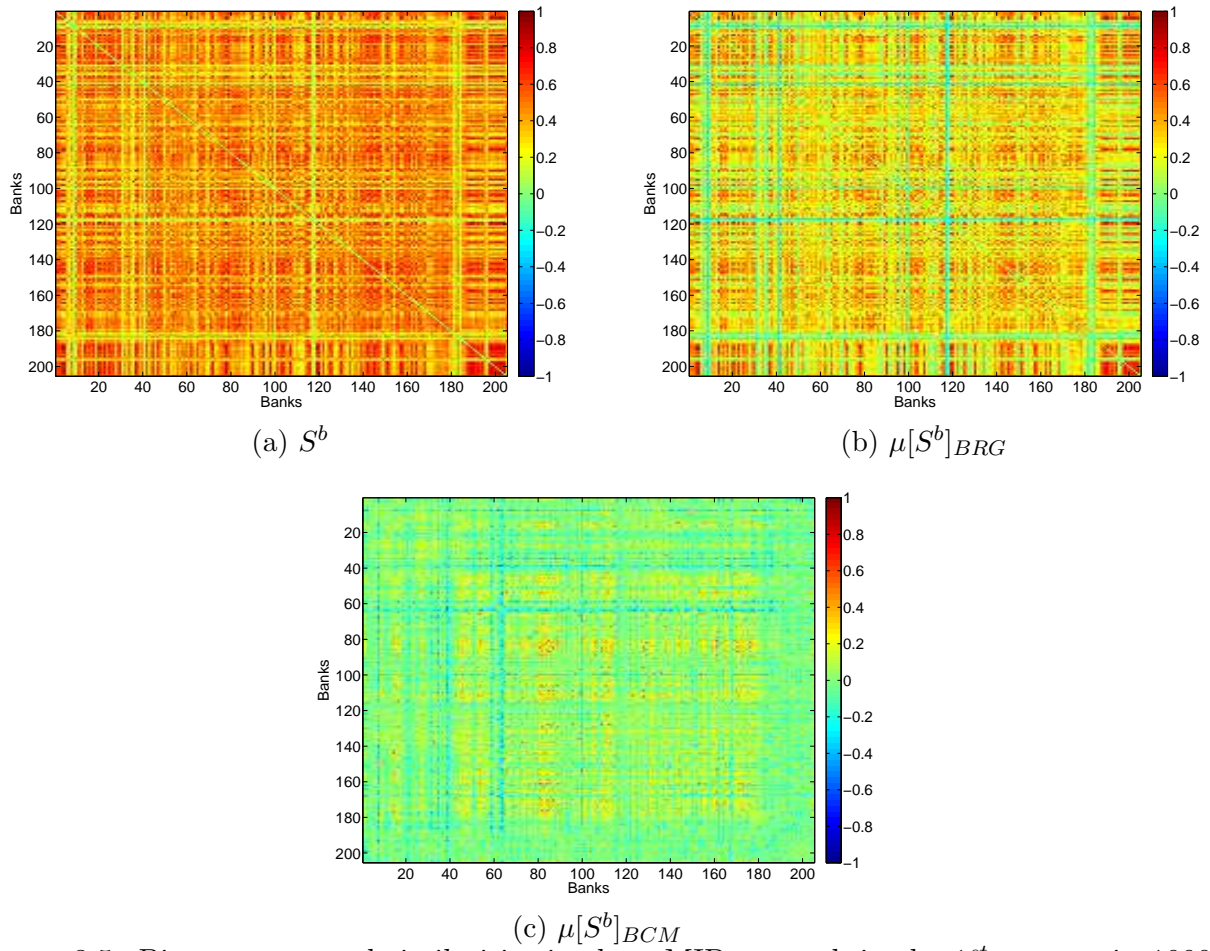


Figure 3.5: Binary structural similarities in the e-MID network in the 1st quarter in 1999. Panel (a): raw similarity matrix. Panel (b): rescaled similarity matrix under the BRG. Panel (c): rescaled similarity matrix under the BCM.

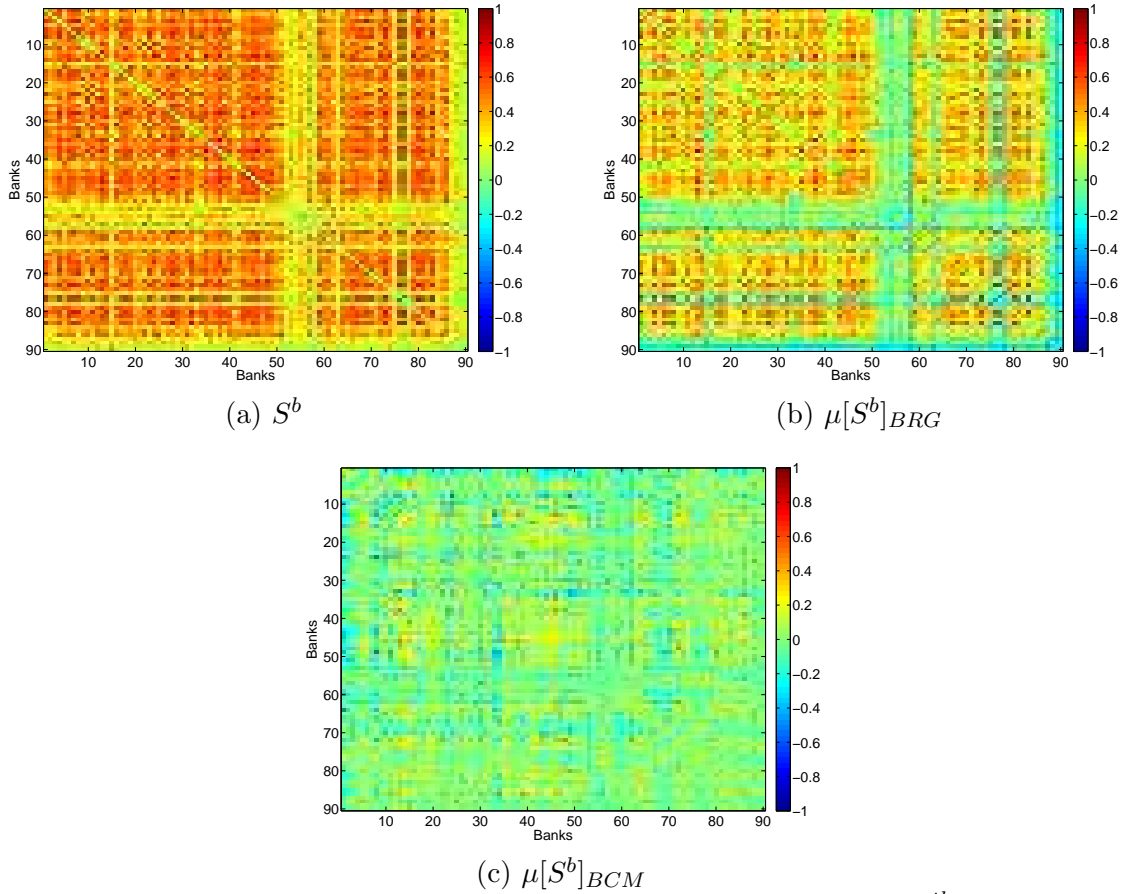


Figure 3.6: Binary structural similarities in the e-MID network in the 4th quarter in 2010. Panel (a): raw similarity matrix. Panel (b): rescaled similarity matrix under the BRG. Panel (c): rescaled similarity matrix under the BCM.

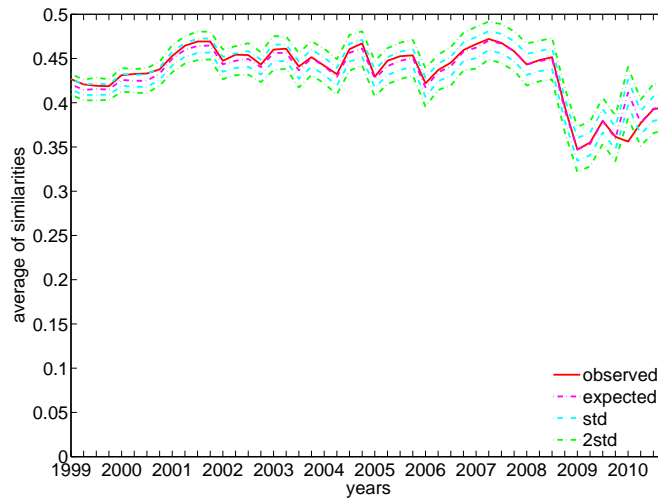


Figure 3.7: Evolution of the average structural similarities in the e-MID network from 1999 to 2010, in the binary case.

Next, we employ the z-score analysis to assess the statistical significance of the elements of the structural similarity matrix S^b in the e-MID data. In figure (3.8), we plot $z[S^b]_{BCM}$ against $\mu[S^b]_{BCM}$ in the two chosen quarters. Similar to the analysis of the previous network, we compare $z[S^b]_{BCM}$ with the critical values $z^* = \pm 2$. Overall, we find that, almost all pairwise similarities are completely explained by information encoded in the degree sequence, while some pairs of banks (especially for those corresponding to the smallest and largest values of $\mu[S^b]_{BCM}$) exhibit significant, non-random structural similarities.

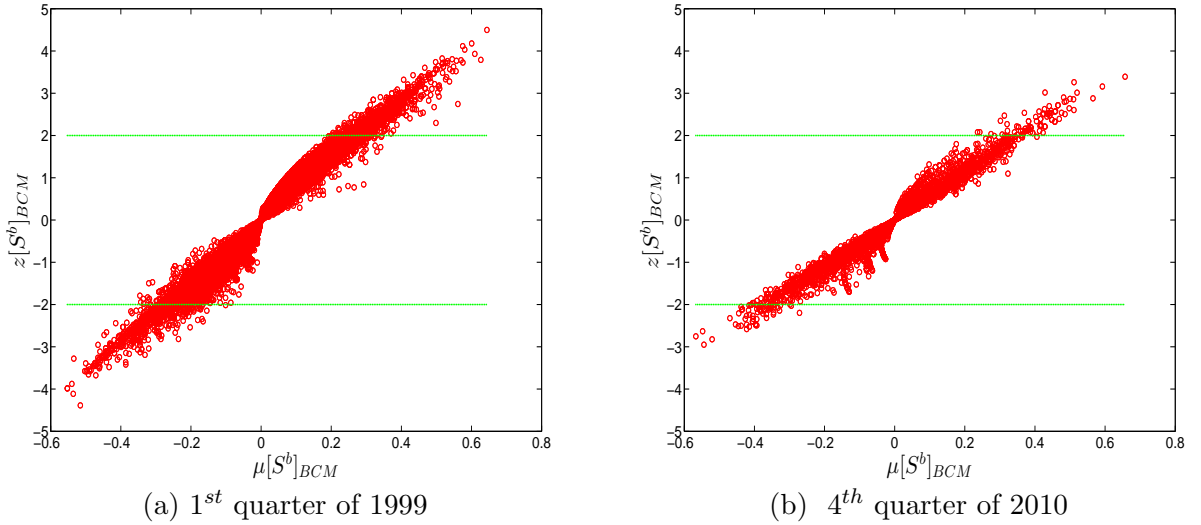


Figure 3.8: z-score analysis in the e-MID network in the 1st quarter of 1999 (panel (a)) and in the 4th quarter of 2010 (panel (b)) under the BCM. In each panel, we plot $z[S^b]_{BCM}$ vs. $\mu[S^b]_{BCM}$, and the green dashed lines represent for the critical values $z^* = \pm 2$.

3.3.2 Weighted analysis

3.3.2.1 Bank-firm credit network of Spain

Similar to the binary case, we project the binary bipartite bank-firm credit network onto a weighted network of portfolio overlaps between banks. Each element of the matrix of weights is the total number of firms that a pair of banks co-finance. Such a one-mode projected network indicates joint exposures to counterparty risk via loans to non-financial firms (e.g. Lux, 2016).

We observe that the color-coded matrix of structural similarities in the weighted version is less bold than that in the binary version, as shown in the panel (a) of Figures (3.9) and (3.10). More specifically, for almost all pairs of banks we find that $S^b > S^w$. The average of similarities in the weighted version is also smaller than in the binary version. Additionally,

the elements of the weighted similarity matrix are highly heterogeneous. Clearly, as shown in the panel (b) of these Figures, such a heterogeneous array can not be captured by the WRG model.

In the panels (c) and (d) of Figures (3.9) and (3.10), we find that for the two weighted configuration models WCM and ECM, on the one hand, the structural similarities between banks are partially filtered under these models. On the other hand, we observe that some pairwise similarities are still significantly present even when the effects of the strength sequence (i.e. in the WCM) or the effects of both the strength sequence as well as the degree sequence (in the ECM) are taken into account (see Figures (3.24), (3.25) in the Appendix for more details).

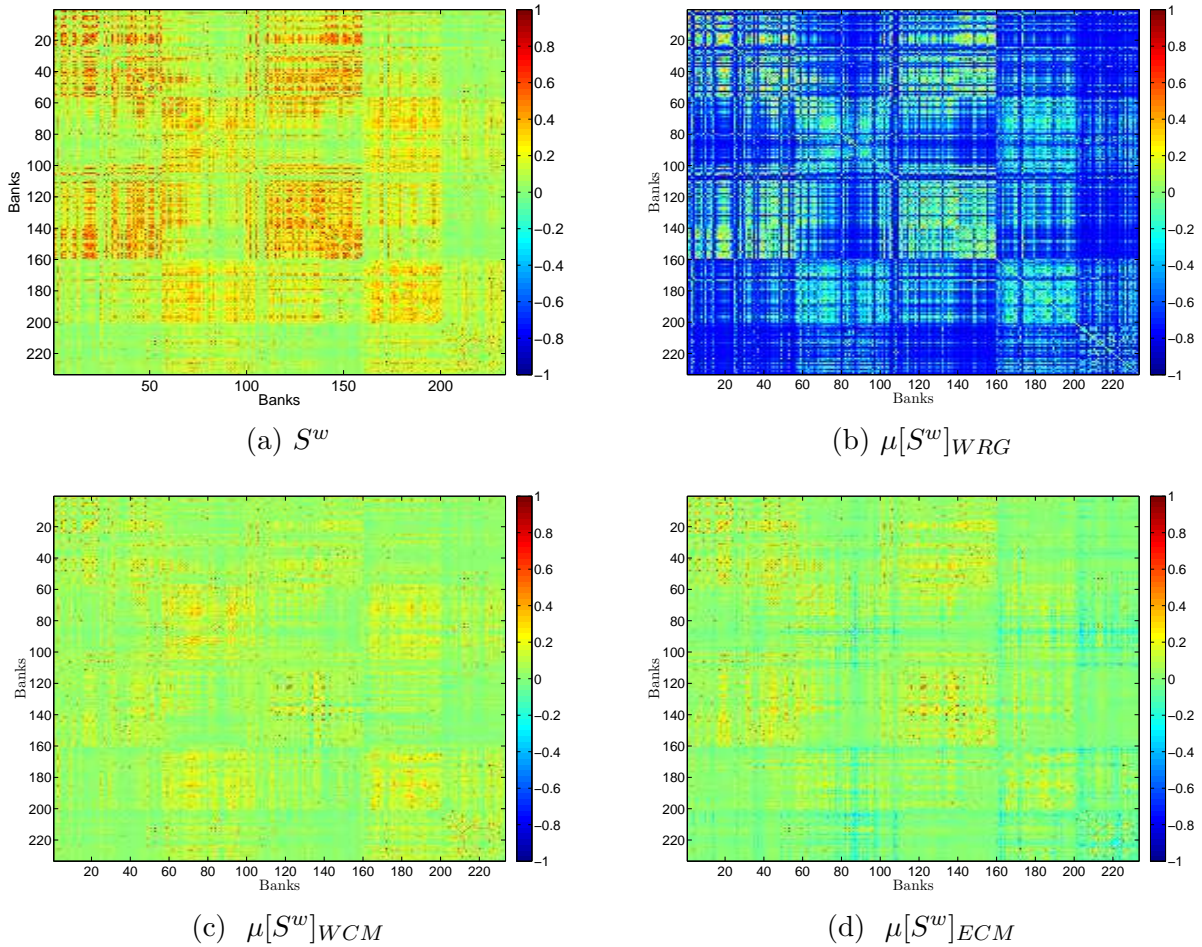


Figure 3.9: Weighted structural similarities in the bank-firm credit network of Spain in 1997. Panel (a): raw similarity matrix. Panel (b): rescaled similarity matrix under the WRG. Panel (c): rescaled similarity matrix under the WCM. Panel (d): rescaled similarity matrix under the ECM.

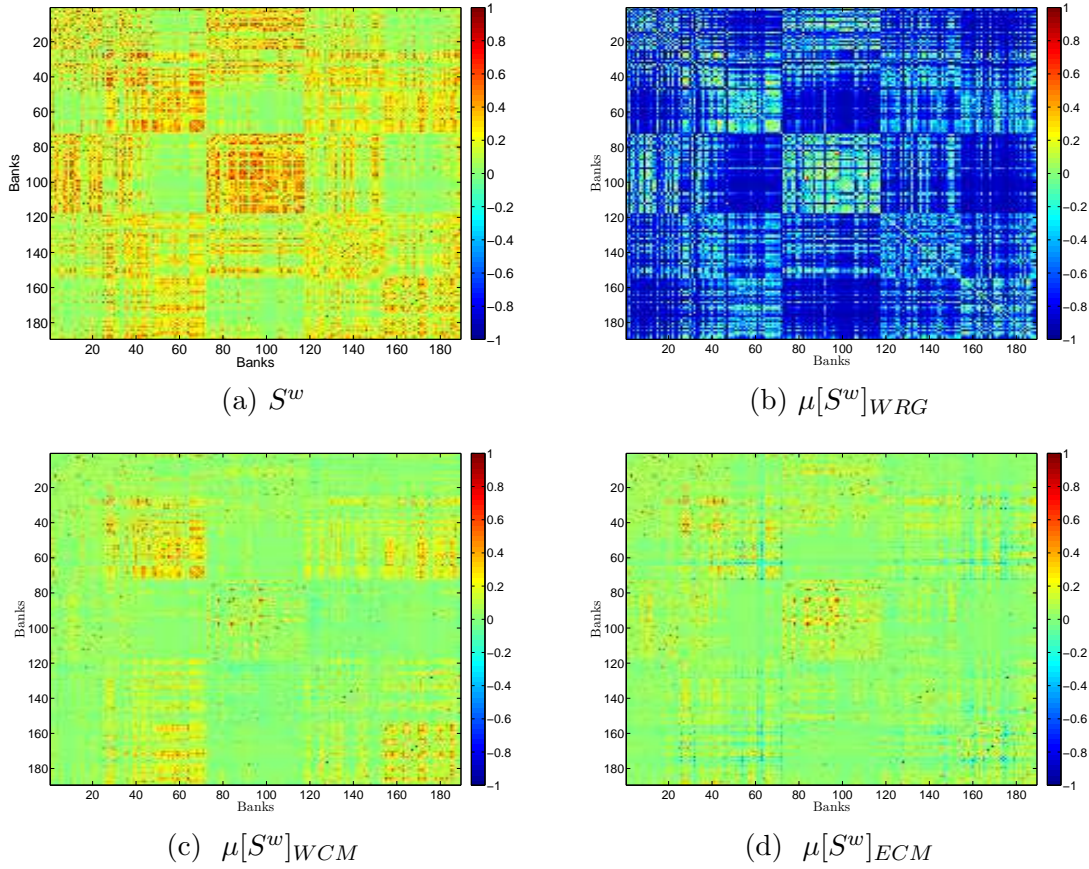


Figure 3.10: Weighted structural similarities in the bank-firm credit network of Spain in 2007. Panel (a): raw similarity matrix. Panel (b): rescaled similarity matrix under the WRG. Panel (c): rescaled similarity matrix under the WCM. Panel (d): rescaled similarity matrix under the ECM.

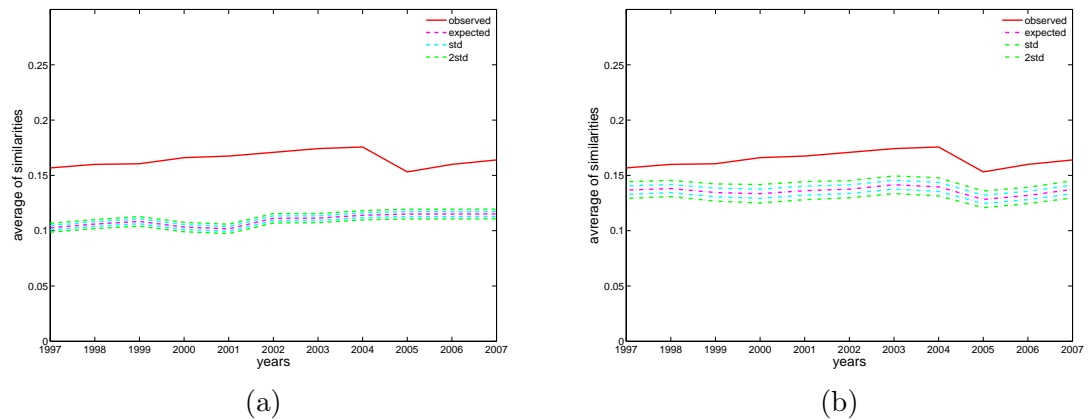


Figure 3.11: Evolution of the average of weighted structural similarities in the bank-firm credit network of Spain from 1997 to 2007, compared to expected ones obtained from the WCM (panel (a)) and the ECM (panel (b)).

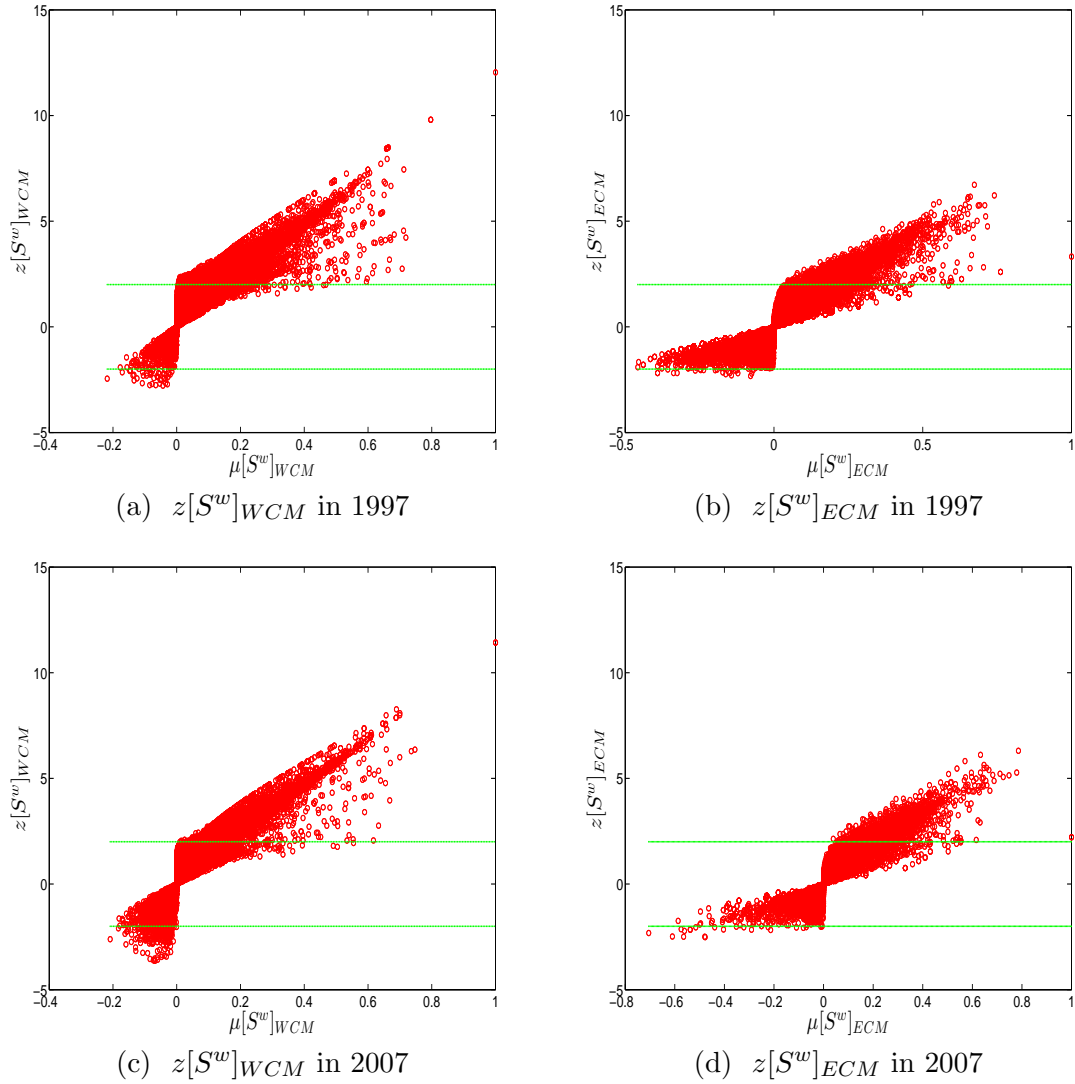


Figure 3.12: z-score analysis in the bank-firm credit network of Spain in 1997 and 2007. Panel (a): $z[S^w]_{WCM}$ vs. $\mu[S^w]_{WCM}$ in 1997. Panel (b): $z[S^w]_{ECM}$ vs. $\mu[S^w]_{ECM}$ in 1997. Panel (c): $z[S^w]_{WCM}$ vs. $\mu[S^w]_{WCM}$ in 2007. Panel (d): $z[S^w]_{ECM}$ vs. $\mu[S^w]_{ECM}$ in 2007. The green dashed lines indicate critical values $z^* = \pm 2$.

The evolution of the average of weighted similarities in the bank-firm credit network of Spain is shown in Figure (3.11). Over years, the observed value of the average significantly deviates from the expected ones obtained from the WCM (panel (a)) and from the ECM (panel (b)).

In Figure (3.12) we show the z-scores evaluated under the WCM and under the ECM. In line with the results obtained from the analysis of the color-coded matrices shown in Figures (3.9) and (3.10), we find that for many banks (especially for those have a high value of $\mu[S^w]_{WCM}$ or $\mu[S^w]_{ECM}$), the structural similarity between them can not simply be

explained by hidden variables extracted from the observed strengths or from the observed strengths as well as the observed degrees. Note that we also observe that even some small values of similarities are statistically significant.

A more detailed comparison between the z -scores evaluated under these two weighted configuration models is shown in Figure (3.28) in the Appendix. For pairs of banks i and j satisfying that $|z[S_{ij}^w]_{WCM}| > 2$ as well as $|z[S_{ij}^w]_{ECM}| > 2$, we also typically observe that $|z[S_{ij}^w]_{WCM}| > |z[S_{ij}^w]_{ECM}|$. This suggests that, for these pairs of banks, the discrepancies between the observed similarities and the expected ones in the ECM model are smaller than those in the WCM. However, we should emphasize that, it is difficult to conclude which weighted configuration model generally outperforms the other in terms of replicating the structural similarity matrix of the observed network.

3.3.2.2 Italian e-MID network

Figures (3.13) and (3.14) show the raw similarities as well as the rescaled ones in the weighted version of the e-MID network. Similar to the bank-firm credit network of Spain, typically we also find that $S^b > S^w$. Again, the WRG model shows that it is not able to filter the hierarchical structure of the raw weighted similarity matrix of the e-MID network. Between the two weighted configuration models, we find that $\mu[S^w]_{WCM}$ is negative for many pairs of banks, implying that the expected similarities between these banks under the WCM are higher than the observed ones. In contrast, in the rescaled similarity matrix $\mu[S^w]_{ECM}$, almost all similarities are removed, although several strictly positive similarities are still present. A more detailed comparison between the WCM and the ECM reveals that typically $|z[S^w]_{WCM}| > |z[S^w]_{ECM}|$ (see Figure (3.16) below and Figure (3.29) in the Appendix), implying that the observed similarities deviate less from the expected values obtained from the ECM than from the expected values obtained from the WCM.

In Figure (3.15) we show the evolution of the average of weighted similarities from 1999 to 2010. We find that the observed average of weighted similarities lies completely outside of ± 2 std. from the expected value obtained from the WCM. In contrast, for almost all quarters, we find a very close agreement between the observed average and the expected one obtained from the ECM.

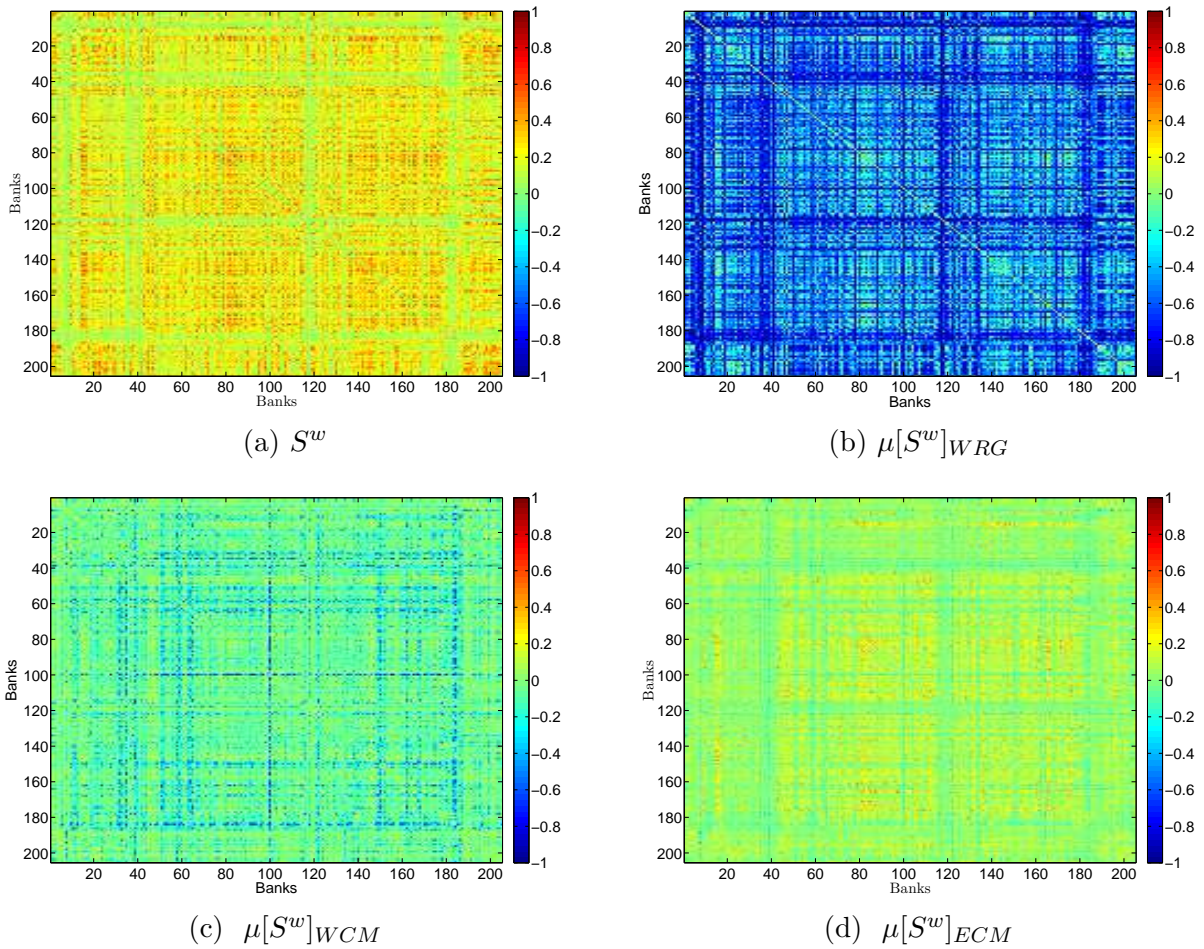


Figure 3.13: Weighted structural similarities in the e-MID network in the first quarter of 1999. Panel (a): raw similarity matrix. Panel (b): rescaled similarity matrix under the WRG. Panel (c): rescaled similarity matrix under the WCM. Panel (d): rescaled similarity matrix under the ECM.

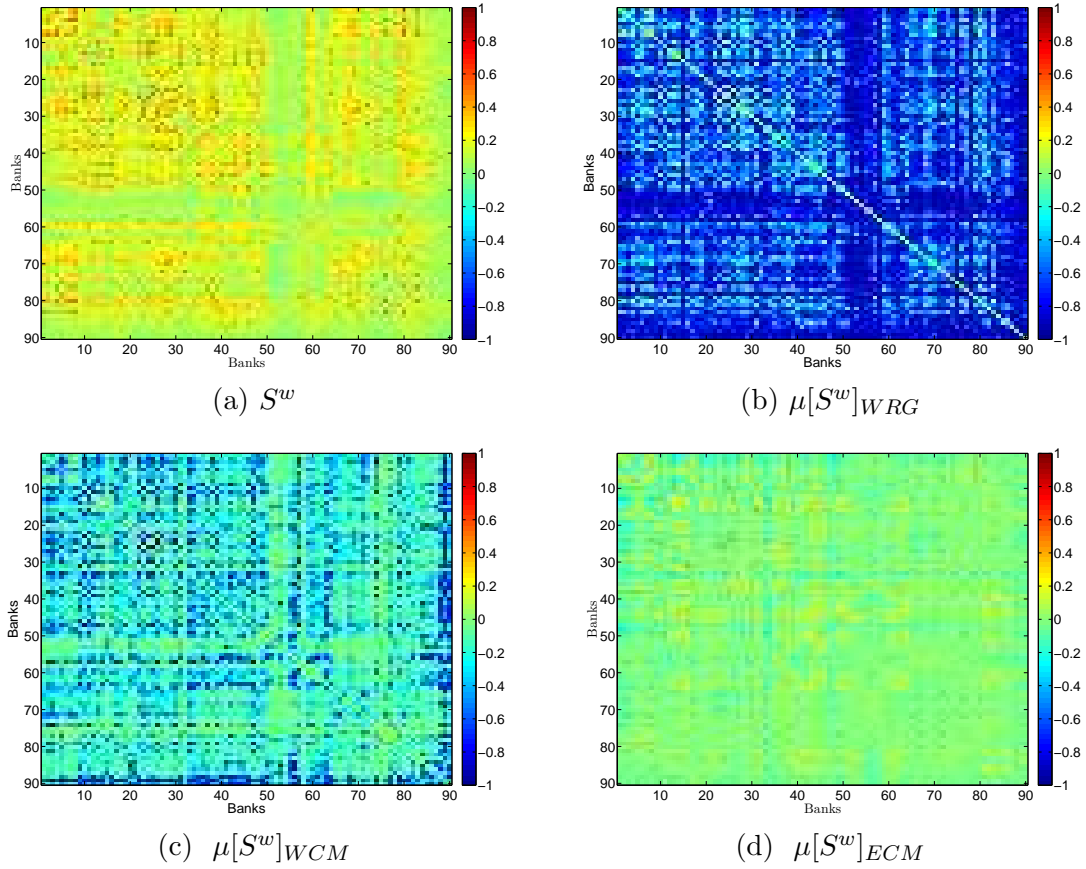


Figure 3.14: Weighted structural similarities in the e-MID network in the last quarter of 2010. Panel (a): raw similarity matrix. Panel (b): rescaled similarity matrix under the WRG. Panel (c): rescaled similarity matrix under the WCM. Panel (d): rescaled similarity matrix under the ECM.

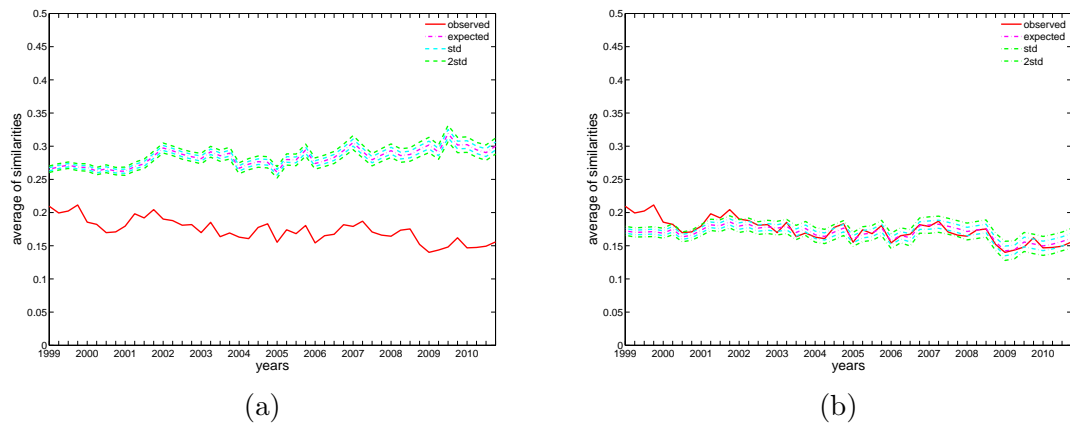


Figure 3.15: Evolution of the average of weighted structural similarities for the e-MID network from 1999 to 2010, compared to expected ones obtained from the WCM (panel (a)) and the ECM (panel (b)).

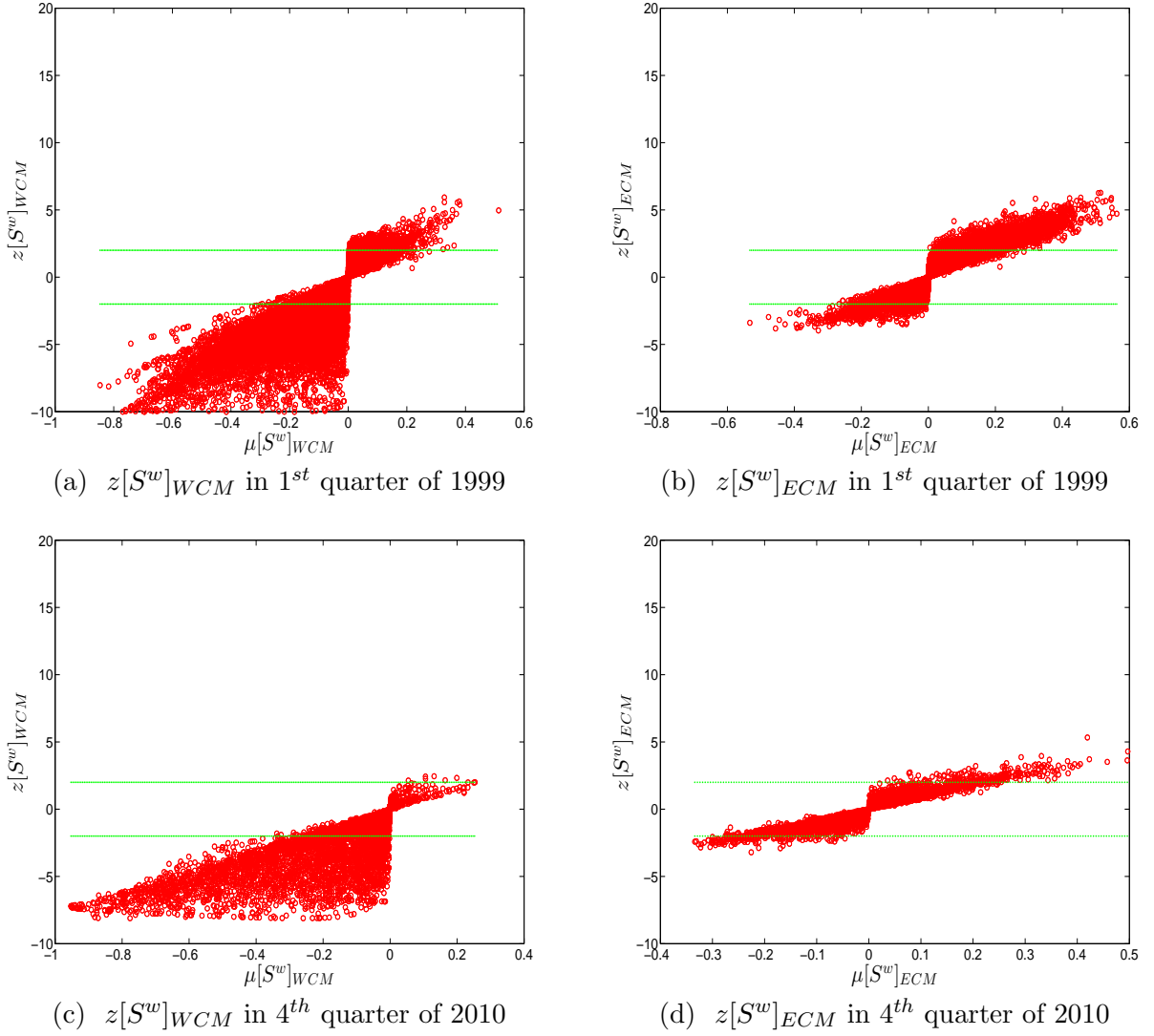


Figure 3.16: z-score analysis in the e-MID network. Panel (a): $z[S^w]_{WCM}$ vs. $\mu[S^w]_{WCM}$ in the 1st quarter of 1999. Panel (b): $z[S^w]_{ECM}$ vs. $\mu[S^w]_{ECM}$ in the 1st quarter of 1999. Panel (c): $z[S^w]_{WCM}$ vs. $\mu[S^w]_{WCM}$ in the 4th quarter of 2010. Panel (d): $z[S^w]_{ECM}$ vs. $\mu[S^w]_{ECM}$ in the 4th quarter of 2010. The green dashed lines indicate the critical values $z^* = \pm 2$.

We now assess the statistical significance of the pairwise similarities in the weighted version of the e-MID network. In the panels (a) and (c) of Figure (3.16) we can see that, when evaluated against the WCM, almost all observed pairwise similarities strongly deviate from the expected values obtained from that model. In contrast, as shown in panels (b) and (d) of Figure (3.16), overall we find that the discrepancies between observed pairwise similarities and the expected ones obtained from the ECM are significantly reduced. This reveals that, unlike in the previous network, adding the knowledge of the observed degrees in addition to the observed strength yields a remarkable improvement in replicating the similarities

between banks in the e-MID network. However, we should emphasize that the weighted version of the e-MID network under scrutiny does exhibit some non-random patterns of weighted similarities between several banks.

3.3.3 Hidden variables, structural similarities, and non-random patterns

In this part of the study we will provide a further explanation for the role of hidden variables extracted from the local constraints for the emergence of structural similarities.

We assume that each bank is assigned with a fitness representing its capacity. In the binary case, suppose that the fitness is positively correlated with the hidden variable extracted from the degree sequence under the BCM (e.g. Cimini et al., 2015a; Cimini et al., 2015b). If two different nodes i and j have similar hidden variables extracted from the degree sequence, i.e. $x_i^* \sim x_j^*$, then $\forall k \neq i, j$ we also have that

$$p_{ik} = \frac{x_i^* x_k^*}{x_i^* x_k^* + 1} \sim p_{jk} = \frac{x_j^* x_k^*}{x_j^* x_k^* + 1} \text{ (see Eq. (3.6)),}$$

meaning that the probability of a link between node i and node k is also similar to the probability of a link between node j and node k . As a consequence, node i and node j will also have a similar neighborhood. As two examples, in Figures (3.17) and (3.18), we examine the role of hidden variables in the structural similarities of the binary version of the observed network. Banks on the vertical as well as horizontal axes are in descending order of hidden variables. We can see that banks associated with similar hidden variables tend to have a higher level of structural similarity. In addition, the top-left corner of these Figures shows that the portfolios of a group of banks associated with the highest values of their hidden variables also have the highest degree of overlaps.

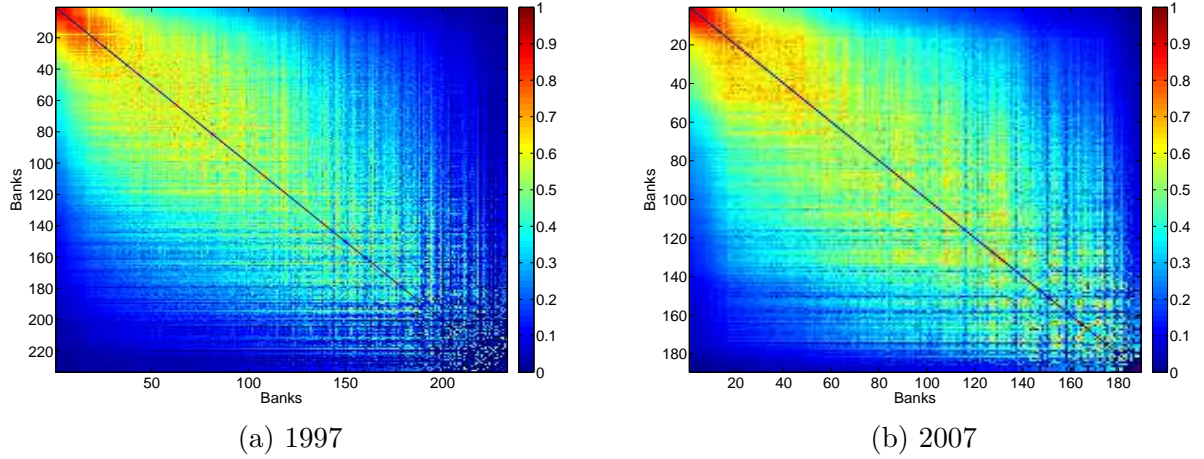


Figure 3.17: Similarities in descending order of hidden variables in the bank-firm credit network of Spain in 1997 (panel (a)), and 2007 (panel (b)).

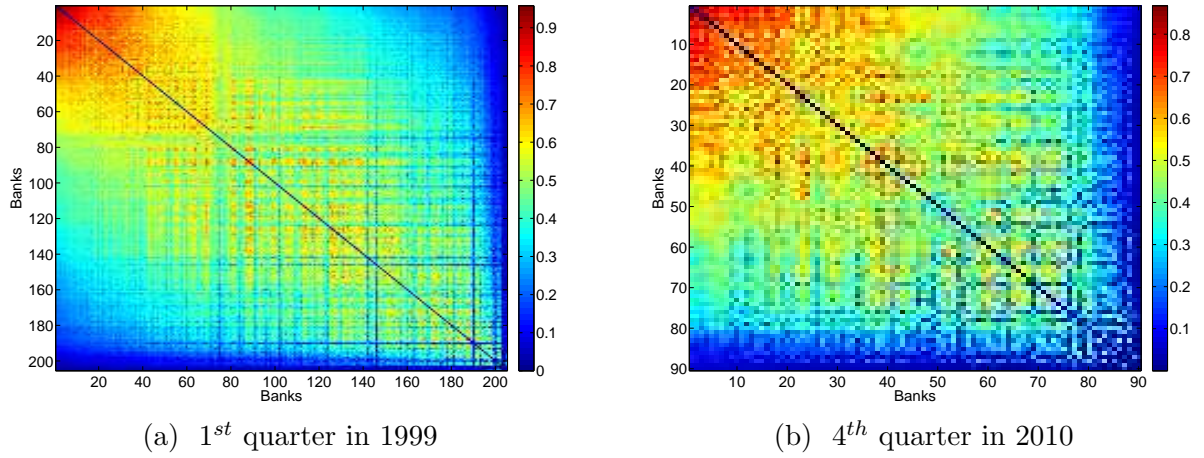


Figure 3.18: Similarities in descending order of hidden variables in e-MID network in the 1st quarter in 1999 (panel (a)) and in the 4th quarter in 2010 (panel (b)).

Note that, mathematically, a similar relation between the hidden variables extracted from the strength sequence and structural similarities in the WCM in the weighted version can easily be obtained. More specifically, under this model, if two different nodes i and j have similar hidden variables, i.e. $x_i^* \sim x_j^*$, then $\forall k \neq i, j$ we also have that

$$\langle w_{ik} \rangle = \frac{x_i^* x_k^*}{1 - x_i^* x_k^*} \sim \langle w_{jk} \rangle = \frac{x_j^* x_k^*}{1 - x_j^* x_k^*} \text{ (see Eq. (3.9))},$$

indicating that the expected weight between the two nodes i and k is similar to the expected weight between the two nodes j and k . However, since we always find that both observed net-

works significantly deviate from the WCM, the hidden variables extracted from the strength sequence can not explain the observed structural similarities in the weighted case. Under the ECM, which is found to have better performance than the WCM in the e-MID data, the expected weight between any two different nodes i and k is now determined by both the hidden variables extracted from the observed degrees as well as those extracted from the observed strengths as

$$\langle w_{ik} \rangle = \frac{x_i^* x_k^* y_i^* y_k^*}{(1 - y_i^* y_k^*)(1 - y_i^* y_k^* + x_i^* x_k^* y_i^* y_k^*)} \quad (\text{see Eq. (3.14)}).$$

Clearly, if for two different nodes i and j the hidden variables x_i^* and y_i^* are respectively similar to the hidden variables x_j^* and y_j^* , for any other node k we will also have that $\langle w_{ik} \rangle \sim \langle w_{jk} \rangle$. However, we should emphasize that unlike in the two previous configuration models, in this model the reverse relation does not necessarily hold.

Obviously, the presence of statistically significant similarities between some banks in both networks shows that they can not be completely explained by hidden variables extracted from the local constraints. For a further investigation of the largest and statistically significant similarities, we select the subset of pairs of banks that have $|z[S^b]_{BCM}| > 2$ as well as $S^b > 0.5$ in the binary case and the subset of pairs of banks that have $|z[S^w]_{ECM}| > 2$ as well as $S^w > 0.5$ in the weighted case.

Figure (3.19) shows the largest and significant components in the case of the bank-firm credit network of Spain. We find that, first, the emergence of these components in the weighted version is much denser than in the binary version. Second, among the three categories of banks, i.e. commercial banks, saving banks, and credit cooperatives, we find that the subset of banks having the largest and significant similarities mainly consists of banks of the first and the second categories. There is a tendency that banks in these two categories have a higher level of structural similarities with banks in the same category as well as with banks in the other category. It should be emphasized that in the presence of such patterns in the banking system, financial risks could be propagated from one category to the other.

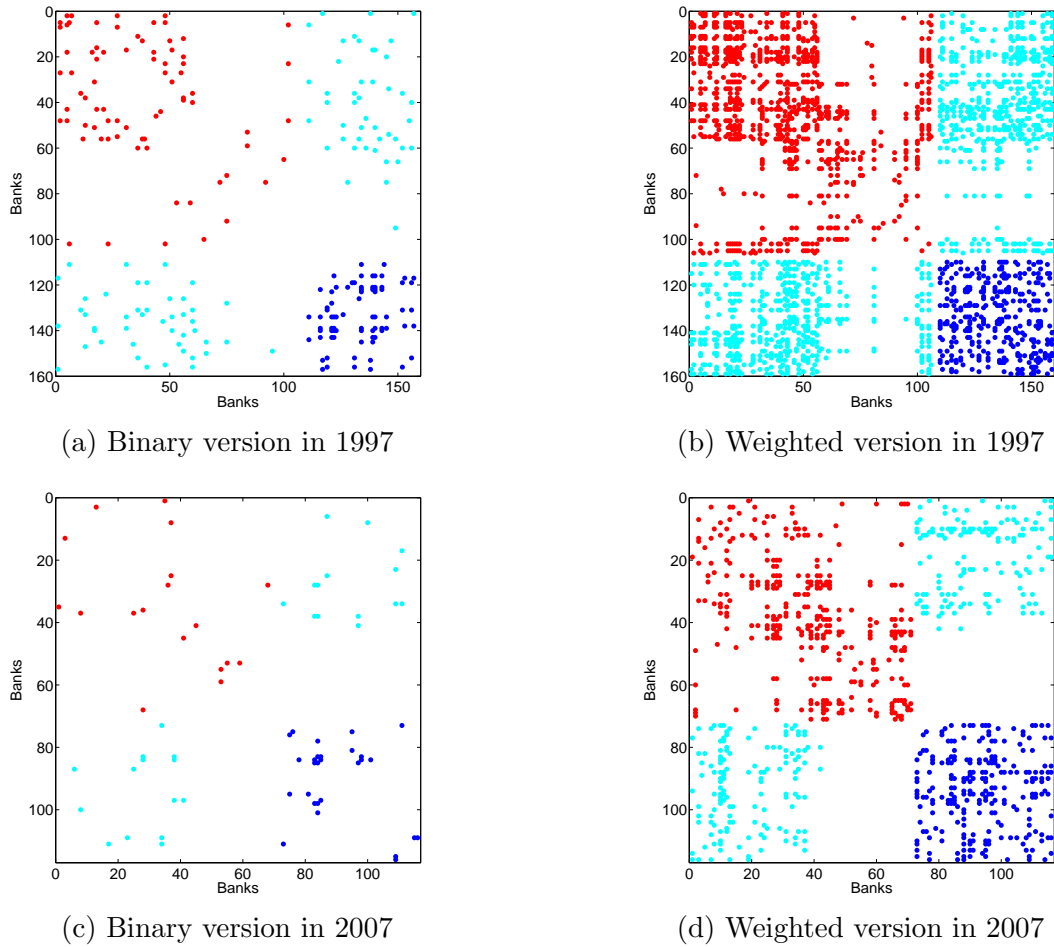


Figure 3.19: Largest and statistically significant similarities between different groups of banks in the bank-bank projection matrix of Spain. We select the subset of pairs of banks that have $|z[S^b]_{BCM}| > 2$ as well as $S^b > 0.5$ in the binary case and the subset of pairs of banks that have $|z[S^w]_{ECM}| > 2$ as well as $S^w > 0.5$ in the weighted case. Panels (a) (b) show the results in the binary and weighted versions in 1997. Panels (c) (d) show the results in binary and weighted versions in 2007. Red points stand for pairs of commercial banks. Blue points stand for pairs of saving banks. Cyan points represent pairs of commercial and saving banks.

Unfortunately additional information about non-topological properties such as the classification of banks is not available in the data set of the Italian e-MID network. Consequently, it is not possible to classify which categories of banks exhibit more significant structural similarities.

3.4 Conclusions

In this study, we apply a new method to filter the effects of the fundamental constraints on the structural similarities between nodes in one-mode networks, and then we apply this method to identify patterns in the portfolio overlaps and similarities in two real world financial networks, i.e. the bank-firm credit network of Spain and the Italian e-MID network.

Among the various null models, we find that the random graph models specifying only global constraints, i.e. the average of links in the binary version and the average of weights in the weighted version, cannot capture the main features of the similarity matrix of the observed network. This implies that they are not suitable models used for filtering the observed similarity matrices.

In contrast, almost all structural similarities between banks in the binary case are significantly removed when the effects of the observed degree sequence are filtered out. Furthermore, interestingly, we find that banks associated with similar hidden variables extracted from the observed degrees tend to have a similar portfolio. Intuitively, the portfolios of banks with similar characteristics (i.e. a similar level of hidden variables) may tend to be more similar.

The results for the weighted version of the two networks are mixed. More specifically, on the one hand, in the case of the e-MID network, additional knowledge of the degree sequence to the strength sequence results in a remarkable improvement of the reconstruction of the weighted similarity matrix. In other words, between the WCM and the ECM, the latter is a more proper model for filtering the weighted similarity matrix. On the other hand, in the case of the bank-firm credit network of Spain, we find that, for many pairs of banks, there is no significant difference between the filtered similarity under the WCM and the filtered similarity under the ECM.

It should be emphasized that, comparing the observed networks with the various null models, we do observe the presence of non-random patterns in the structural similarities between some banks in the two networks that can neither be explained by the global nor local constraints. Particularly, in the case of the bank-firm credit network of Spain, selecting the subset of largest and statistically significant components in the similarity matrices as examples, we find that this subset mainly consists of saving banks as well as commercial banks. We suggest that when more detailed information about the other characteristics of banks is available in both data sets, it would be interesting to investigate the role of these characteristics for the emergence of the statistically significant components of the structural similarity matrices.

This study contributes to the understanding of structural similarity in network science in general and in portfolio overlap and similarity between agents in the financial system in particular. The method employed in this study can straightforwardly be extended to directed networks and bipartite networks. More specifically, in the case of directed networks, a similar method can be proposed based on the configuration models defined for the directed version (e.g. see Squartini et al., 2011a; Squartini et al., 2011b; Squartini et al., 2013; Squartini et al., 2015; Luu et al., 2016). In the case of bipartite networks, one can, for instance, employ the recently proposed bipartite configuration models (e.g. see Saracco et al., 2015; Di Gangi et al., 2015; Luu and Lux, 2016b) as the null models to detect patterns in the structural similarity in different bipartite financial networks such as investor-asset networks, bank-firm credit networks, and so forth. Moreover, we believe that further extensions to other similarity measures (e.g. as summarized in Alvarez-Socorro et al., 2015) may provide a more comprehensive assessment of the structural similarity in complex networks.

3.5 References

Acharya V. V. 2009. A theory of systemic risk and design of prudential bank regulation. *Journal of Financial Stability* 5 (3), pp. 224-255.

Alvarez-Socorro A. J., Herrera-Almarza G. C., and González-Díaz L. A. 2015. Eigencentrality based on dissimilarity measures reveals central nodes in complex networks. *Scientific Reports* 5.

Blocher J. 2016. Network externalities in mutual funds. *Journal of Financial Markets* 30, pp. 1-26.

Brechler J., Hausenblas V., Komárková Z., Plašil M. 2014. Similarity and clustering of banks: The application on the credit exposures of the Czech banking sector. The Czech National Bank, Research and Policy Notes 4. Available at: http://www.cnb.cz/miranda2/export/sites/www.cnb.cz/en/research/research_publications/irpn/download/rpn_4_2014.pdf.

Caccioli F., Shrestha M., Moore C., Farmer J. D. 2014. Stability analysis of financial contagion due to overlapping portfolios. *Journal of Banking & Finance* 46, pp. 233-245.

Cai J., Saunders A., Steffen S. 2016. Syndication, interconnectedness, and systemic risk. Working Paper. Available at: <http://ssrn.com/abstract=1508642orhttp://dx.doi.org/10.2139/ssrn.1508642>.

Chierichetti F., Kumar R., Pandey S., Vassilvitskii S. 2010. Finding the Jaccard median.

- Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms.
- Cimini G., Squartini T., Garlaschelli D., Gabrielli A. 2015a. Estimating topological properties of weighted networks from limited information. *Physical Review E* 92 (4).
- Cimini G., Squartini T., Garlaschelli D., Gabrielli A. 2015b. Systemic risk analysis on reconstructed economic and financial networks. *Scientific Reports* 5.
- Dice L. R. 1945. Measures of the amount of ecologic association between species. *Ecology* 26 (3), pp. 297-302.
- Di Gangi D., Lillo F., Pirino D. 2015. Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. Working Paper. Available at: [arXiv:1509.00607](https://arxiv.org/abs/1509.00607).
- Fortunato S. 2010. Community detection in graphs. *Physics Reports* 486 (3-5), pp. 75-174.
- Fricke D. 2016. Has the banking system become more homogeneous? Evidence from banks' loan portfolios. *Economics Letters* 142, pp. 45-48.
- Haldane A., May R. M. 2011. Systemic risk in banking ecosystems. *Nature* 469, pp. 351-355.
- Huang X., Vodenska I., Havlin S., Stanley H. E. 2013. Cascading failures in bi-partite graphs: Model for systemic risk propagation. *Scientific Reports* 3.
- Gemmetto V., Squartini T., Picciolo F., Ruzzenenti F., Garlaschelli D. 2015. Multiplexity and multireciprocity in directed multiplexes. Available at: [arXiv:1411.1282](https://arxiv.org/abs/1411.1282).
- Gemmetto V., Garlaschelli D. 2015. Multiplexity versus correlation: the role of local constraints in real multiplexes. *Scientific Reports* 5.
- Greenwood R., Landier A., Thesmar D. 2015. Vulnerable banks. *Journal of Financial Economics* 115 (3), pp. 471-485.
- Leicht E. A., Holme P., Newman M. E. J. 2006. Vertex similarity in networks. *Physical Review E* 73 (2).
- Li Y., Ye Y., Du X. 2011. A new vertex similarity metric for community discovery: A local flow model. *Journal of Software* 6 (8), pp. 1545-1553.
- Liao H., Zeng A. 2015. Reconstructing propagation networks with temporal similarity. *Scientific Reports* 5.
- Lu L., Zhou T. 2011. Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications* 390 (6), pp. 1150-1170.
- Luu D. T., Lux T. 2016a. Multilayer overlaps and correlations in the bank-firm credit network of Spain. Working Paper.

- Luu D. T., Lux T. 2016b. Identifying patterns in the bank-sector credit network of Spain. Working Paper.
- Luu D. T., Lux T., Yanovski B. 2016. Does the heterogeneity in the local constraints predominantly determine structural correlations in the Italian overnight money market?. Working Paper.
- Lux T. 2016. A model of the topology of the bank-firm credit network and its role as channel of contagion. *Journal of Economic Dynamics and Control* 66, pp. 36-53.
- Mastrandrea R., Squartini T., Fagiolo G., Garlaschelli D. 2014. Enhanced reconstruction of weighted networks from strengths and degrees. *New Journal of Physics* 16.
- Papadopoulos F., Kitsak M., Serrano M. A., Boguna M., Krioukov D. 2012. Popularity versus similarity in growing networks. *Nature* 489, pp. 537-540.
- Park J., Newman M. E. J. 2004. Statistical mechanics of networks. *Physical Review E* 70 (6).
- Pool V. K., Stoffman N., Yonker S. E. 2015. The people in your neighborhood: Social interactions and mutual fund portfolios. *The Journal of Finance* 70 (6), pp. 2679-2732.
- Sias R., Turtle W., Zykaj B. 2015. Hedge fund crowds and mispricing. *Management Science* 62 (3), pp. 764 -784.
- Saracco F., Clemente R. D., Gabrielli A., and Squartini T. 2015. Randomizing bipartite networks: The case of the World Trade Web. *Scientific Reports* 5.
- Sørensen T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Kongelige Danske Videnskabernes Selskab* 5, pp. 1-34.
- Squartini T., Fagiolo G., Garlaschelli D. 2011a. Randomizing world trade.I. A binary network analysis. *Physical Review E* 84 (4).
- Squartini T., Fagiolo G., Garlaschelli D. 2011b. Randomizing world trade.II. A weighted network analysis. *Physical Review E* 84 (4).
- Squartini T., Garlaschelli D. 2011. Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics* 13.
- Squartini T., Picciolo F., Ruzzenenti F., Garlaschelli D. 2013. Reciprocity of weighted networks. *Scientific Reports* 3.
- Squartini T., Mastrandrea R., Garlaschelli D. 2015. Unbiased sampling of network ensembles. *New Journal of Physics* 17.
- Wagner W. 2008. The homogenization of the financial system and liquidity crises. *Journal of Financial Intermediation* 17, pp. 330-356.

Wagner W. 2010. Diversification at financial institutions and systemic crises. *Journal of Financial Intermediation* 19, pp. 373-386.

Zhu C., Kushwaha A., Berman K., Jegga A. G. 2012. A vertex similarity-based framework to discover and rank orphan disease-related genes. *BMC System Biology* 6.

3.6 Appendix

3.6.1 Relevant proofs for rescaled similarities and z-scores evaluated under the various null models

In this part, with a similar approach to Squartini et al. (2013), Gemmetto and Garlaschelli (2015), Gemmetto et al. (2015), and Luu and Lux (2016a), we will show that the expectations of $\min(a_{ik}, a_{jk})$, $\max(a_{ik}, a_{jk})$, $\min(a_{ik}, a_{jk})^2$, $\max(a_{ik}, a_{jk})^2$, $\min(w_{ik}, w_{jk})$, $\max(w_{ik}, w_{jk})$, $\min(w_{ik}, w_{jk})^2$, and $\max(w_{ik}, w_{jk})^2$ can analytically be calculated via hidden variables extracted from the constraints maintained in the various null models. In addition, we also obtain the analytical expressions for $\sigma^2(\min(a_{ik}, a_{jk}))$, $\sigma^2(\max(a_{ik}, a_{jk}))$, $\sigma^2(\min(w_{ik}, w_{jk}))$, $\sigma^2(\max(w_{ik}, w_{jk}))$.

Before going into details, let us start with some useful lemmas. For any real number y such that $0 < y < 1$, we have

Lemma 1

$$\sum_{n=0}^{\infty} y^n = \frac{1}{(1-y)}. \quad (3.50)$$

Eq. (3.50) from **Lemma 1** leads to **Lemma 2** and **Lemma 3** below

Lemma 2

$$\sum_{n=0}^{\infty} n y^n = \frac{y}{(1-y)^2}. \quad (3.51)$$

Lemma 3

$$\sum_{n=0}^{\infty} n^2 y^n = \frac{y^2 + y}{(1-y)^3}. \quad (3.52)$$

Lemma 4

–*Lemma 4.1*

$$\sum_{n=0}^{\infty} n(y^{n-1} - y^n) = \frac{1}{(1-y)}. \quad (3.53)$$

–Lemma 4.2

$$\sum_{n=0}^{\infty} n(y^n - y^{n+1}) = \frac{y}{(1-y)}. \quad (3.54)$$

Lemma 5

–Lemma 5.1

$$\sum_{n=0}^{\infty} n^2 y^{n-1} - n^2 y^n = \frac{y+1}{(1-y)^2}. \quad (3.55)$$

–Lemma 5.2

$$\sum_{n=0}^{\infty} n^2 y^n - n^2 y^{n+1} = \frac{y^2 + y}{(1-y)^2}. \quad (3.56)$$

Proof for the BCM

Recall that, in the BCM, the probability of a link between node i and node j is given by $p_{ij} = \frac{x_i^* x_j^*}{1+x_i^* x_j^*}$ ($\forall i \neq j$) as in Eq. (3.6), where $\{x_i^*\}_{i=1}^N$ are hidden variables extracted from the observed degree sequence. Since in the binary case, $\min(a_{ik}, a_{jk})$ is equal to 0 or 1, the expected value of $\min(a_{ik}, a_{jk})$ is

$$\begin{aligned} \langle \min(a_{ik}, a_{jk}) \rangle_{BCM} &= \min(a_{ik}, a_{jk}) \text{ prob.}(\min(a_{ik}, a_{jk}) = 1) \\ &= \text{prob.}(\min(a_{ik}, a_{jk}) = 1) = p_{ik} p_{jk}, \end{aligned} \quad (3.57)$$

where $\text{prob.}(\min(a_{ik}, a_{jk}) = 1)$ is the probability that $\min(a_{ik}, a_{jk}) = 1$.

Note that that in the binary case we always have $\min(a_{ik}, a_{jk}) \equiv \min(a_{ik}, a_{jk})^2$ and $\max(a_{ik}, a_{jk}) \equiv \max(a_{ik}, a_{jk})^2$. The expected value of $\min(a_{ik}, a_{jk})^2$ is then specified as

$$\begin{aligned} \langle \min(a_{ik}, a_{jk})^2 \rangle_{BCM} &= \min(a_{ik}, a_{jk})^2 \text{ prob.}(\min(a_{ik}, a_{jk}) = 1) \\ &= \text{prob.}(\min(a_{ik}, a_{jk}) = 1) = p_{ik} p_{jk}. \end{aligned} \quad (3.58)$$

From Eqs. (3.57) and (3.58), the variance of $\min(a_{ik}, a_{jk})$ is

$$\sigma^2(\min(a_{ik}, a_{jk}))_{BCM} = \langle \min(a_{ik}, a_{jk})^2 \rangle_{BCM} - \langle \min(a_{ik}, a_{jk}) \rangle_{BCM}^2 = (p_{ik} p_{jk}) - (p_{ik} p_{jk})^2. \quad (3.59)$$

Note that we can also obtain similar expressions for $\langle \max(a_{ik}, a_{jk}) \rangle_{BCM}$ as

$$\langle \max(a_{ik}, a_{jk}) \rangle_{BCM} = \langle a_{ik} \rangle + \langle a_{jk} \rangle - \langle \min(a_{ik}, a_{jk}) \rangle = p_{ik} + p_{jk} - p_{ik} p_{jk}, \quad (3.60)$$

and for $\langle \max(a_{ik}, a_{jk})^2 \rangle_{BCM}$ as

$$\begin{aligned} \langle \max(a_{ik}, a_{jk})^2 \rangle_{BCM} &= \max(a_{ik}, a_{jk})^2 \text{prob.}(\max(a_{ik}, a_{jk}) = 1) \\ &= \max(a_{ik}, a_{jk}) \text{prob.}(\max(a_{ik}, a_{jk}) = 1) = p_{ik} + p_{jk} - p_{ik}p_{jk}. \end{aligned} \quad (3.61)$$

Therefore,

$$\sigma^2(\max(a_{ik}, a_{jk}))_{BCM} = (p_{ik} + p_{jk} - p_{ik}p_{jk}) - (p_{ik} + p_{jk} - p_{ik}p_{jk})^2. \quad (3.62)$$

Proof for the BRG model

Mathematically, we can consider the BRG model as a special case of the BCM, when all nodes are homogeneous, i.e. $p_{ij} = p = \frac{2L_{tol}}{n(n-1)}$. From Eqs. (3.57) and (3.58), we get

$$\langle \min(a_{ik}, a_{jk}) \rangle_{BRG} = p^2, \quad (3.63)$$

and

$$\langle \min(a_{ik}, a_{jk})^2 \rangle_{BRG} = p^2. \quad (3.64)$$

Therefore, the variance of $\min(a_{ik}, a_{jk})$ under the BRG model is

$$\sigma^2(\min(a_{ik}, a_{jk}))_{BRG} = p^2 - p^4. \quad (3.65)$$

Proof for WCM model

In the WCM, for every pairs of nodes i and j , the probability of a link of weight w_{ij} is

$$q(w_{ij}) = \begin{cases} 1 - p_{ij}, & \text{if } w_{ij} = 0, \\ (p_{ij})^{w_{ij}}(1 - p_{ij}), & \text{if } w_{ij} > 0, \end{cases} \quad (3.66)$$

where $p_{ij} = x_i^* x_j^*$ as given in Eq. (3.11), and $\{x_i^*\}_{i=1}^n$ are the solution to Sys. (3.8).

Assuming that weights are integers, we obtain the CCDF of w_{ij} as

$$q(w_{ij} \geq w) = 1 - \sum_{t=0}^{w-1} q_{ij}(t). \quad (3.67)$$

It follows from Eq. (3.66) and Eq. (3.67) that the CCDF can be rewritten as

$$q(w_{ij} \geq w) = 1 - \left[\sum_{t=0}^{w-1} (p_{ij}^t - p_{ij}^{t+1}) \right] = p_{ij}^w. \quad (3.68)$$

Hence the CDF function is

$$q(w_{ij} \leq w) = \sum_{t=0}^w q_{ij}(t) = 1 - p_{ij}^{w+1}. \quad (3.69)$$

Eqs. (3.68) and (3.69) lead to

$$q(\min(w_{ik}, w_{jk}) \geq w) = q(w_{ik} \geq w)q(w_{jk} \geq w) = p_{ik}^w p_{jk}^w. \quad (3.70)$$

and

$$q(\max(w_{ik}, w_{jk}) \leq w) = q(w_{ik} \leq w)q(w_{jk} \leq w) = (1 - p_{ik}^{w+1})(1 - p_{jk}^{w+1}). \quad (3.71)$$

In general, the expected value of $\min(w_{ik}, w_{jk})$ can be expressed as

$$\langle \min(w_{ik}, w_{jk}) \rangle = \sum_{w'=0}^{\infty} w' q(\min(w_{ik}, w_{jk}) \geq w') - w' q(\min(w_{ik}, w_{jk}) \geq (w' + 1)). \quad (3.72)$$

From *Lemma 4.2* (Eq. (3.54)), Eq. (3.70), and Eq. (3.72) we obtain

$$\langle \min(w_{ik}, w_{jk}) \rangle_{WCM} = \sum_{w'=0}^{\infty} w' [(p_{ik} p_{jk})^{w'} - (p_{ik} p_{jk})^{w'+1}] = \frac{p_{ik} p_{jk}}{1 - p_{ik} p_{jk}}. \quad (3.73)$$

Eq. (3.73) leads to

$$\langle \min(w_{ik}, w_{jk}) \rangle_{WCM}^2 = \left(\frac{p_{ik} p_{jk}}{1 - p_{ik} p_{jk}} \right)^2. \quad (3.74)$$

Note that, in a similar way we can show that

$$\langle \min(w_{ik}, w_{jk})^2 \rangle = \sum_{w'=0}^{\infty} w'^2 [q(\min(w_{ik}, w_{jk}) \geq w') - q(\min(w_{ik}, w_{jk}) \geq w' + 1)]. \quad (3.75)$$

From *Lemma 5.2* (Eq. (3.56)), Eqs. (3.70) and (3.75) we obtain

$$\langle \min(w_{ik}, w_{jk})^2 \rangle_{WCM} = \frac{p_{ik} p_{jk} (p_{ik} p_{jk} + 1)}{(1 - p_{ik} p_{jk})^2}. \quad (3.76)$$

Therefore, the variance of $\min(w_{ik}, w_{jk})$ is obtained from Eqs. (3.74), (3.76) as

$$\sigma^2(\min(w_{ik}, w_{jk}))_{WCM} = \frac{p_{ik} p_{jk} (p_{ik} p_{jk} + 1)}{(1 - p_{ik} p_{jk})^2} - \left(\frac{p_{ik} p_{jk}}{1 - p_{ik} p_{jk}} \right)^2 = \frac{p_{ik} p_{jk}}{(1 - p_{ik} p_{jk})^2}. \quad (3.77)$$

Note that we can also obtain similar expressions for $\langle \max(w_{ik}, w_{jk}) \rangle_{WCM}$ and $\sigma^2(\max(w_{ik}, w_{jk}))_{WCM}$. With the CDF function of $q(w_{ij})$ is given in Eq. (3.69), the expected value of $\max(w_{ik}, w_{jk})$ is given by

$$\begin{aligned}
\langle \max(w_{ik}, w_{jk}) \rangle_{WCM} &= \sum_{w'=0}^{\infty} w' q(\max(w_{ik}, w_{jk}) \leq w') - w' q(\max(w_{ik}, w_{jk}) \leq (w' - 1)) \\
&= \sum_{w'=0}^{\infty} w' [(1 - p_{ik}^{w'+1})(1 - p_{jk}^{w'+1}) - (1 - p_{ik}^{w'}) (1 - p_{jk}^{w'})] \\
&= \sum_{w'=0}^{\infty} w' [(p_{ik}^{w'} - p_{ik}^{w'+1}) + (p_{jk}^{w'} - p_{jk}^{w'+1}) - (p_{ik}^{w'} p_{jk}^{w'} - p_{ik}^{w'+1} p_{jk}^{w'+1})]
\end{aligned} \tag{3.78}$$

Applying *Lemma 4.2* (Eq.(3.54)), we obtain

$$\begin{aligned}
\langle \max(w_{ik}, w_{jk}) \rangle_{WCM} &= \frac{p_{ik}}{1 - p_{ik}} + \frac{p_{jk}}{1 - p_{jk}} - \left(\frac{p_{ik} p_{jk}}{1 - p_{ik} p_{jk}} \right) \\
&= \frac{p_{ik}}{1 - p_{ik}} + \frac{p_{jk}}{1 - p_{jk}} - \langle \min(w_{ik}, w_{jk}) \rangle_{WCM}.
\end{aligned} \tag{3.79}$$

Remember that, under the WCM we have $\langle w_{ik} \rangle = \frac{p_{ik}}{1 - p_{ik}}$ and $\langle w_{jk} \rangle = \frac{p_{jk}}{1 - p_{jk}}$. Therefore, Eq. (3.79) satisfies the equality that $\max(x, y) = x + y - \min(x, y)$, $\forall x, y$ are real numbers.

Similarly, we have

$$\begin{aligned}
\langle \max(w_{ik}, w_{jk})^2 \rangle_{WCM} &= \sum_{w'=0}^{\infty} w'^2 q(\max(w_{ik}, w_{jk}) \leq w') - w'^2 q(\max(w_{ik}, w_{jk}) \leq (w' - 1)) \\
&= \sum_{w'=0}^{\infty} w'^2 [(1 - p_{ik}^{w'+1})(1 - p_{jk}^{w'+1}) - (1 - p_{ik}^{w'}) (1 - p_{jk}^{w'})] \\
&= \sum_{w'=0}^{\infty} w'^2 [(p_{ik}^{w'} - p_{ik}^{w'+1}) + (p_{jk}^{w'} - p_{jk}^{w'+1}) - (p_{ik}^{w'} p_{jk}^{w'} - p_{ik}^{w'+1} p_{jk}^{w'+1})].
\end{aligned} \tag{3.80}$$

Applying *Lemma 5.2* (Eq. (3.56)), we have

$$\begin{aligned}
&\langle \max(w_{ik}, w_{jk})^2 \rangle_{WCM} = \\
&\frac{(1 + p_{ik})p_{ik}}{(1 - p_{ik})^2} + \frac{(1 + p_{jk})p_{jk}}{(1 - p_{jk})^2} - \frac{(1 + p_{ik}p_{jk})p_{ik}p_{jk}}{(1 - p_{ik}p_{jk})^2}.
\end{aligned} \tag{3.81}$$

Therefore, from Eqs. (3.79), (3.81), the variance of $\max(w_{ik}, w_{jk})$ is

$$\begin{aligned} \sigma^2(\max(w_{ik}, w_{jk}))_{WCM} &= \frac{(1+p_{ik})p_{ik}}{(1-p_{ik})^2} + \frac{(1+p_{jk})p_{jk}}{(1-p_{jk})^2} - \frac{(1+p_{ik}p_{jk})p_{ik}p_{jk}}{(1-p_{ik}p_{jk})^2} \\ &\quad - \left[\frac{p_{ik}}{1-p_{ik}} + \frac{p_{jk}}{1-p_{jk}} - \left(\frac{p_{ik}p_{jk}}{1-p_{ik}p_{jk}} \right) \right]^2. \end{aligned} \quad (3.82)$$

Proof for WRG model

In the WRG model, the probability of having a link of weight w_{ij} between two nodes i and j is given by

$$q(w_{ij}) = \begin{cases} 1 - p_w, & \text{if } w_{ij} = 0 \\ (p_w)^{w_{ij}}(1 - p_w), & \text{if } w_{ij} > 0 \end{cases} \quad (3.83)$$

where $p_w = \frac{2W_{\text{tot}}}{n(n-1)+2W_{\text{tot}}}$.

This model can be considered as a special case of the WCM when all strengths are equal. With a similar approach, it is easy to show that

$$\langle \min(w_{ik}, w_{jk}) \rangle_{WRG} = \frac{p_w^2}{1 - p_w^2}. \quad (3.84)$$

and

$$\langle \min(w_{ik}, w_{jk})^2 \rangle_{WRG} = \frac{p_w^4 + p_w^2}{(1 - p_w^2)^2}. \quad (3.85)$$

Therefore

$$\sigma^2(\min(w_{ik}, w_{jk}))_{WRG} = \frac{p_w^2}{(1 - p_w^2)^2}. \quad (3.86)$$

Proof for the ECM

In the ECM, the probability of a link between node i and node j is $p_{ij} = \frac{x_i^* x_j^* y_i^* y_j^*}{1 - y_i^* y_j^* + x_i^* x_j^* y_i^* y_j^*}$ as given in Eq. (3.13), and probability of a link of weight w_{ij} is

$$q(w_{ij}) = \begin{cases} 1 - p_{ij}, & \text{if } w_{ij} = 0, \\ p_{ij} r_{ij}^{w_{ij}-1} (1 - r_{ij}), & \text{if } w_{ij} > 0, \end{cases} \quad (3.87)$$

where $r_{ij} = y_i^* y_j^*$, and $\{x_i^*\}_{i=1}^n$ and $\{y_i^*\}_{i=1}^n$ are the solution to Sys. (3.12).

Similar to the WCM, we obtain the CCDF of w_{ij} as

$$q(w_{ij} \geq w) = 1 - \sum_{t=0}^{w-1} q_{ij}(t). \quad (3.88)$$

It follows from Eq. (3.87), and Eq. (3.88) that the CCDF can be rewritten as

$$q(w_{ij} \geq w) = 1 - [1 - p_{ij} + \sum_{t=1}^{t=w-1} p_{ij} r_{ij}^{t-1} (1 - r_{ij})] = p_{ij} r_{ij}^{w-1}. \quad (3.89)$$

Hence the CDF of w_{ij} is

$$q(w_{ij} \leq w) = \sum_{t=0}^w q_{ij}(t) = 1 - p_{ij} r_{ij}^w. \quad (3.90)$$

Another important result of Eq. (3.89) is that

$$q(\min(w_{ik}, w_{jk}) \geq w) = q(w_{ik} \geq w)q(w_{jk} \geq w) = (p_{ik} r_{ik}^{w-1})(p_{jk} r_{jk}^{w-1}). \quad (3.91)$$

Recalling that the expected value of $\min(w_{ik}, w_{jk})$ can generally be expressed as

$$\langle \min(w_{ik}, w_{jk}) \rangle = \sum_{w'=0}^{\infty} w' q(\min(w_{ik}, w_{jk}) \geq w') - w' q(\min(w_{ik}, w_{jk}) \geq (w' + 1)). \quad (3.92)$$

From *Lemma 4.1* (Eq. (3.53)), Eq. (3.91), and Eq. (3.92) we obtain

$$\langle \min(w_{ik}, w_{jk}) \rangle_{ECM} = p_{ik} p_{jk} \sum_{w'=0}^{\infty} w' [(r_{ik} r_{jk})^{w'-1} - (r_{ik} r_{jk})^{w'}] = \frac{p_{ik} p_{jk}}{1 - r_{ik} r_{jk}}. \quad (3.93)$$

That leads to

$$\langle \min(w_{ik}, w_{jk}) \rangle_{ECM}^2 = \left(\frac{p_{ik} p_{jk}}{1 - r_{ik} r_{jk}} \right)^2. \quad (3.94)$$

Note that, in a similar way, we can obtain

$$\begin{aligned} \langle \min(w_{ik}, w_{jk})^2 \rangle_{ECM} &= \sum_{w'=0}^{\infty} w'^2 [q(\min(w_{ik}, w_{jk}) \geq w') - q(\min(w_{ik}, w_{jk}) \geq w' + 1)] \\ &= p_{ik} p_{jk} \sum_{w'=0}^{\infty} w'^2 [(r_{ik} r_{jk})^{w'-1} - (r_{ik} r_{jk})^{w'}]. \end{aligned} \quad (3.95)$$

Lemma 5.1 (Eq. (3.55)), Eq. (3.91), and Eq. (3.95) lead to

$$\langle \min(w_{ik}, w_{jk})^2 \rangle_{ECM} = \frac{p_{ik} p_{jk} (r_{ik} r_{jk} + 1)}{(1 - r_{ik} r_{jk})^2}. \quad (3.96)$$

Therefore, from Eqs. (3.94) and (3.96) the variance of $\min(w_{ik}, w_{jk})$ is

$$\sigma^2(\min(w_{ik}, w_{jk}))_{ECM} = \frac{p_{ik}p_{jk}(r_{ik}r_{jk} + 1)}{(1 - r_{ik}r_{jk})^2} - \left(\frac{p_{ik}p_{jk}}{1 - r_{ik}r_{jk}}\right)^2 = \frac{p_{ik}p_{jk}(1 - p_{ik}p_{jk} + r_{ik}r_{jk})}{(1 - r_{ik}r_{jk})^2}. \quad (3.97)$$

We can also obtain similar expressions for $\max(w_{ik}, w_{jk})$ under the ECM. Note that the expected value of $\max(w_{ik}, w_{jk})$ is

$$\langle \max(w_{ik}, w_{jk}) \rangle_{ECM} = \sum_{w'=0}^{\infty} w' q(\max(w_{ik}, w_{jk}) \leq w') - w' q(\max(w_{ik}, w_{jk}) \leq (w' - 1)). \quad (3.98)$$

with the CDF function of w_{ij} is given in Eq. (3.90). Equivalently

$$\begin{aligned} \langle \max(w_{ik}, w_{jk}) \rangle_{ECM} &= \sum_{w'=0}^{\infty} w' [(1 - p_{ik}r_{ik}^w)(1 - p_{jk}r_{jk}^w) - (1 - p_{ik}r_{ik}^{w-1})(1 - p_{jk}r_{jk}^{w-1})] \\ &= \sum_{w'=0}^{\infty} w' [(p_{ik}r_{ik}^{w-1} - p_{ik}r_{ik}^w) + (p_{jk}r_{jk}^{w-1} - p_{jk}r_{jk}^w) - (p_{ik}p_{jk}r_{ik}^{w-1}r_{jk}^{w-1} - p_{ik}p_{jk}r_{ik}^wr_{jk}^w)]. \end{aligned} \quad (3.99)$$

Applying *Lemma 4.1* (Eq. (3.53)), we obtain

$$\begin{aligned} \langle \max(w_{ik}, w_{jk}) \rangle_{ECM} &= \frac{p_{ik}}{1 - r_{ik}} + \frac{p_{jk}}{1 - r_{jk}} - \left(\frac{p_{ik}p_{jk}}{1 - r_{ik}r_{jk}}\right) \\ &= \frac{p_{ik}}{1 - r_{ik}} + \frac{p_{jk}}{1 - r_{jk}} - \langle \min(w_{ik}, w_{jk}) \rangle_{ECM}. \end{aligned} \quad (3.100)$$

Note that $\langle w_{ik} \rangle_{ECM} = \frac{p_{ik}}{1 - r_{ik}}$ and $\langle w_{jk} \rangle_{ECM} = \frac{p_{jk}}{1 - r_{jk}}$. Therefore, Eq. (3.100) justifies that $\max(x, y) = x + y - \min(x, y)$, $\forall x, y$ are real numbers.

In addition, by a similar approach we have that

$$\langle \max(w_{ik}, w_{jk})^2 \rangle_{ECM} = \sum_{w'=0}^{\infty} w'^2 [q(\max(w_{ik}, w_{jk}) \leq w') - q(\max(w_{ik}, w_{jk}) \leq w' - 1)]. \quad (3.101)$$

Therefore,

$$\begin{aligned} \langle \max(w_{ik}, w_{jk})^2 \rangle_{ECM} &= \\ \sum_{w'=0}^{\infty} w'^2 [(p_{ik}r_{ik}^{w-1} - p_{ik}r_{ik}^w) + (p_{jk}r_{jk}^{w-1} - p_{jk}r_{jk}^w) - (p_{ik}p_{jk}r_{ik}^{w-1}r_{jk}^{w-1} - p_{ik}p_{jk}r_{ik}^wr_{jk}^w)]. \end{aligned} \quad (3.102)$$

Applying *Lemma 5.1* (Eq. (3.55)), we get

$$\langle \max(w_{ik}, w_{jk})^2 \rangle_{ECM} = \left[\frac{p_{ik}(1+r_{ik})}{(1-r_{ik})^2} + \frac{p_{jk}(1+r_{jk})}{(1-r_{jk})^2} - \frac{p_{ik}p_{jk}(r_{ik}r_{jk}+1)}{(1-r_{ik}r_{jk})^2} \right]. \quad (3.103)$$

The variance of $\max(w_{ik}, w_{jk})$ is obtained from Eqs. (3.100) and (3.103) as

$$\begin{aligned} \sigma^2(\max(w_{ik}, w_{jk}))_{ECM} = & \left[\frac{p_{ik}(1+r_{ik})}{(1-r_{ik})^2} + \frac{p_{jk}(1+r_{jk})}{(1-r_{jk})^2} - \frac{p_{ik}p_{jk}(r_{ik}r_{jk}+1)}{(1-r_{ik}r_{jk})^2} \right] \\ & - \left[\frac{p_{ik}}{1-r_{ik}} + \frac{p_{jk}}{1-r_{jk}} - \left(\frac{p_{ik}p_{jk}}{1-r_{ik}r_{jk}} \right) \right]^2. \end{aligned} \quad (3.104)$$

3.6.2 Additional results

Distributions of the elements of the similarity matrices

In this part we will provide a more detailed comparison between the observed similarities and the rescaled ones obtained from the various configuration models. In particular, in the binary analysis, from Figures (3.20) to (3.23), we compare S^b with $\mu[S_{i,j}^{\text{bin},\text{min}}]_{BCM}$. From Figures (3.24) to (3.27), we compare S^w with $\mu[S_{i,j}^w]_{WCM}$ and $\mu[S_{i,j}^w]_{ECM}$. In each Figure, the color-coded matrices of the raw and the rescaled similarities are shown in the left panels, while the distributions of the elements of these matrices are shown in the right panels.

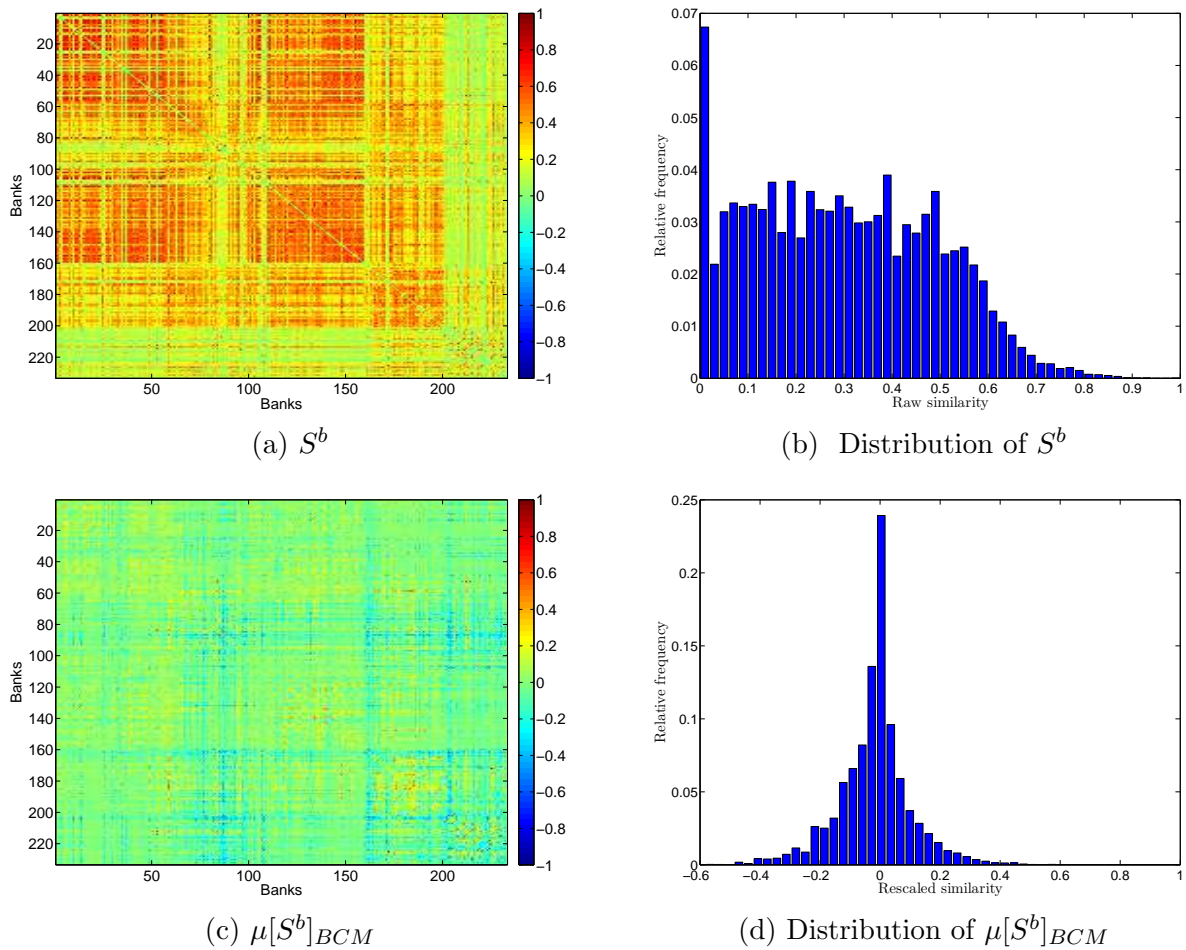
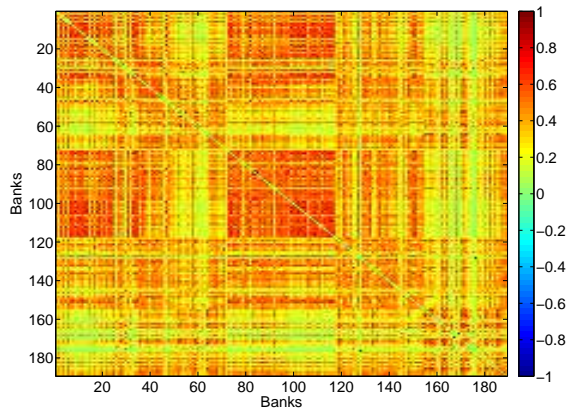
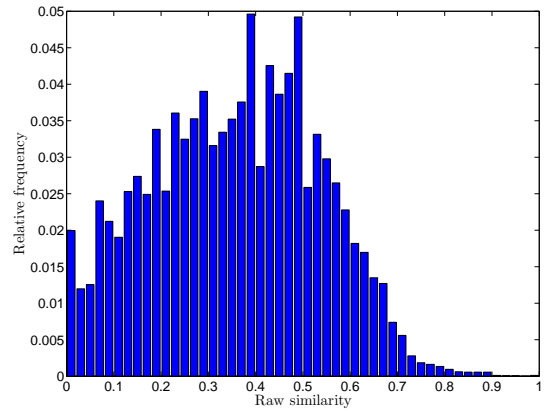


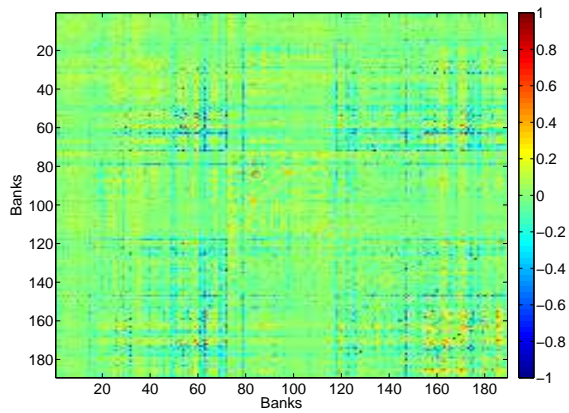
Figure 3.20: Binary structural similarities in the bank-firm credit network of Spain, in 1997. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c) (d): rescaled similarities under the BCM and the corresponding distribution.



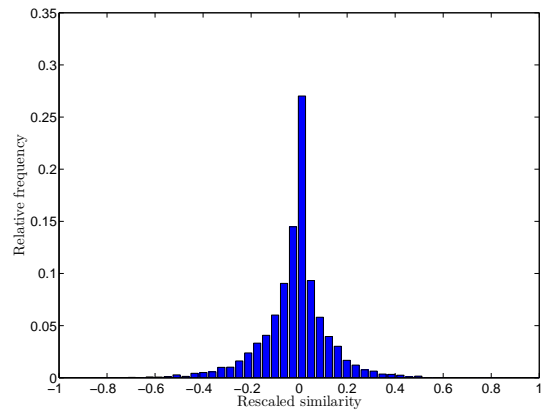
(a) S^b



(b) Distribution of S^b

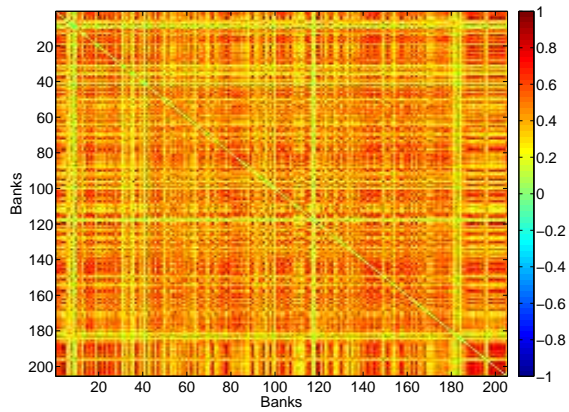


(c) $\mu[S^b]_{BCM}$

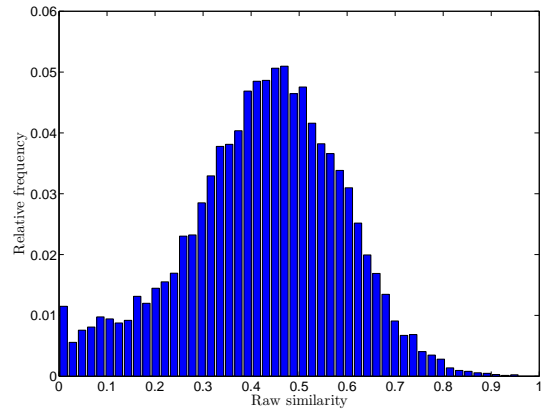


(d) Distribution of $\mu[S^b]_{BCM}$

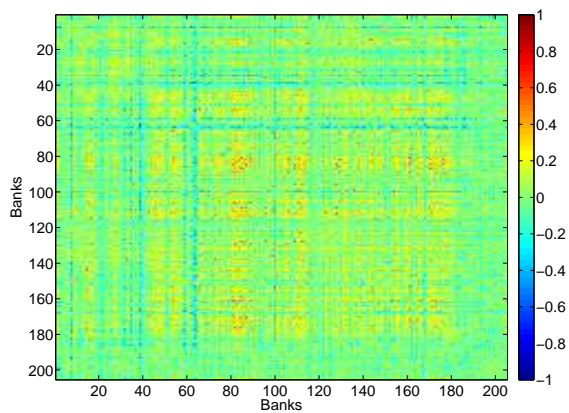
Figure 3.21: Binary structural similarities in the bank-firm credit network of Spain, in 2007. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c) (d): rescaled similarities under the BCM and the corresponding distribution.



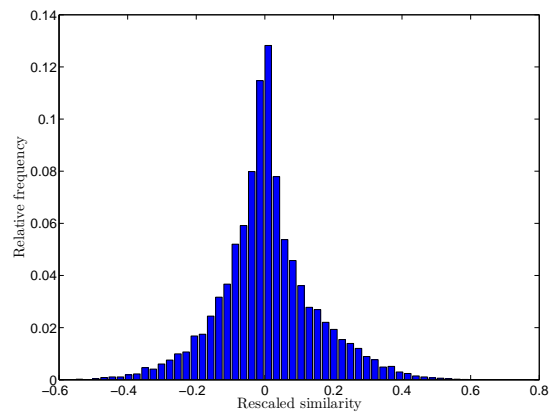
(a) S^b



(b) Distribution of S^b

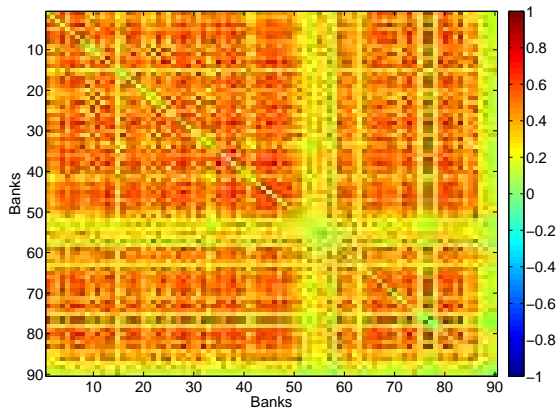


(c) $\mu[S^b]_{BCM}$

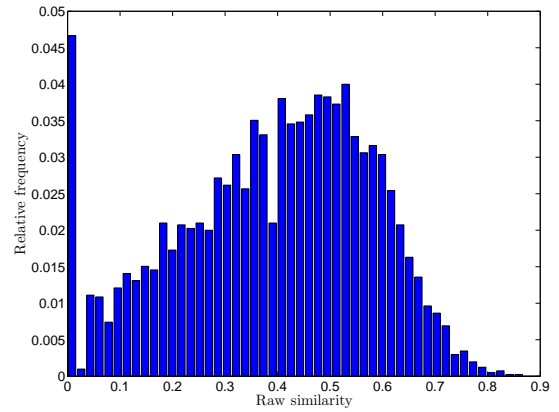


(d) Distribution of $\mu[S^b]_{BCM}$

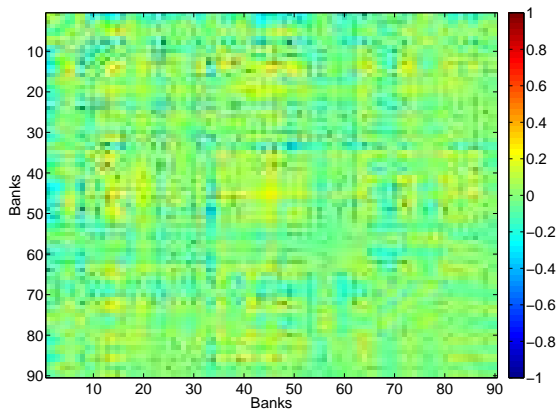
Figure 3.22: Binary structural similarities in the e-MID network, in the first quarter of 1999. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the BCM and the corresponding distribution.



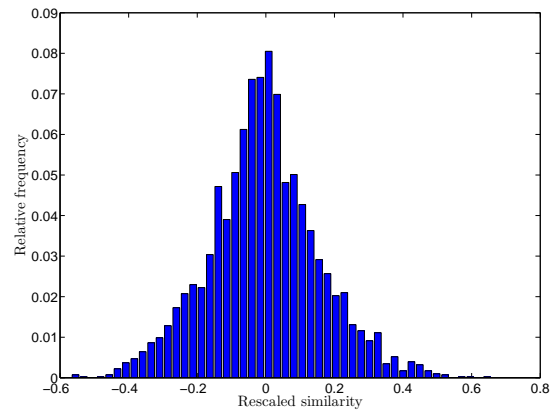
(a) S^b



(b) Distribution of S^b



(c) $\mu[S^b]_{BCM}$



(d) Distribution of $\mu[S^b]_{BCM}$

Figure 3.23: Binary structural similarities in the e-MID network, in the last quarter of 2010. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the BCM and the corresponding distribution.

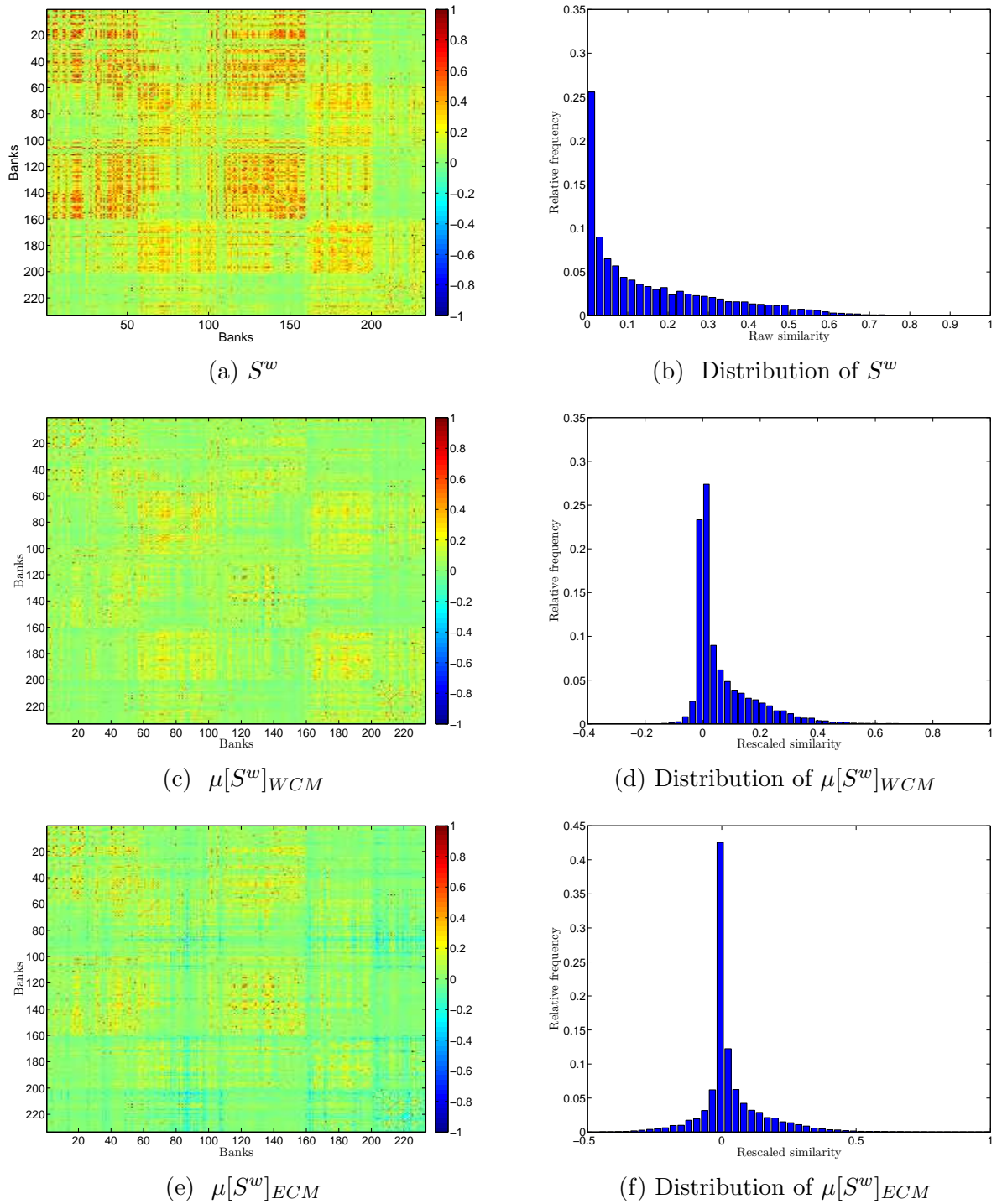


Figure 3.24: Weighted structural similarities in the bank-firm credit network of Spain, in 1997. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the WCM and the corresponding distribution. Panels (e),(f): rescaled similarities under the ECM and the corresponding distribution.

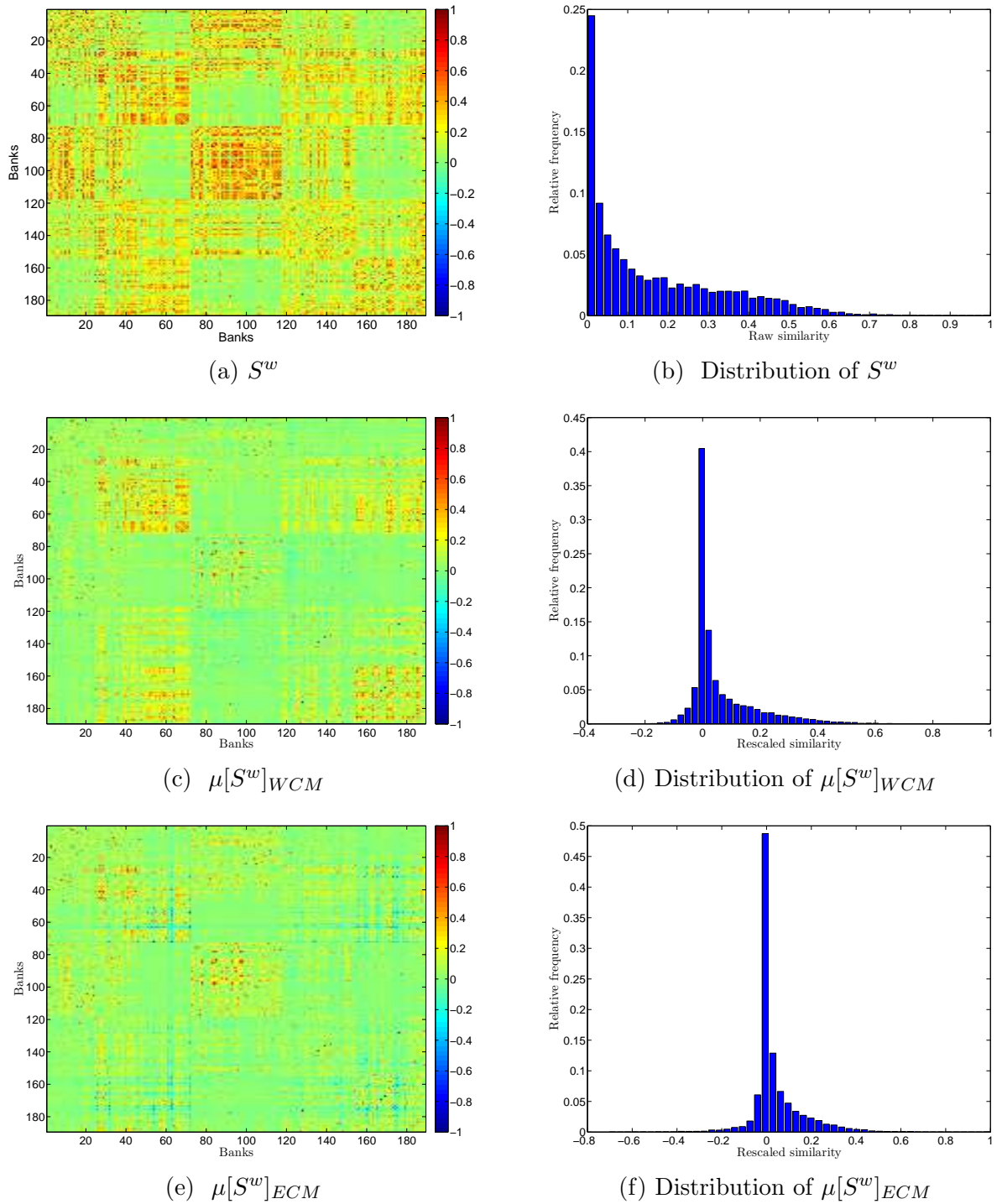
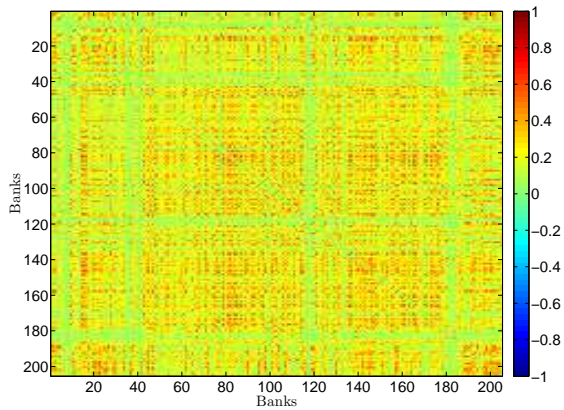
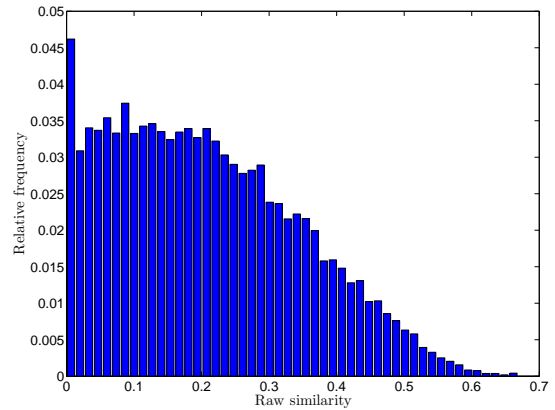


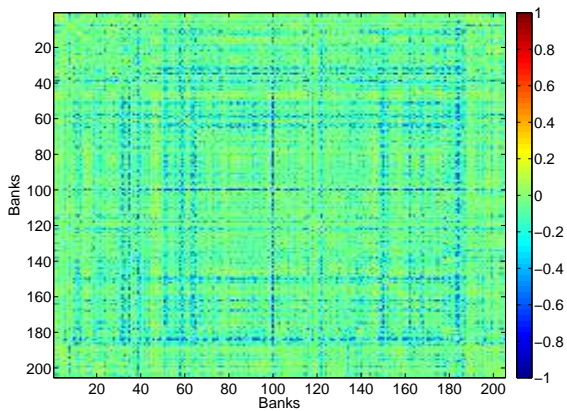
Figure 3.25: Weighted structural similarities in the bank-firm credit network of Spain, in 2007. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the WCM and the corresponding distribution. Panels (e), (f): rescaled similarities under the ECM and the corresponding distribution.



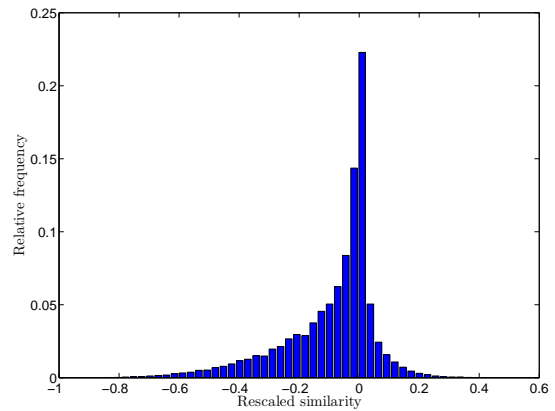
(a) S^w



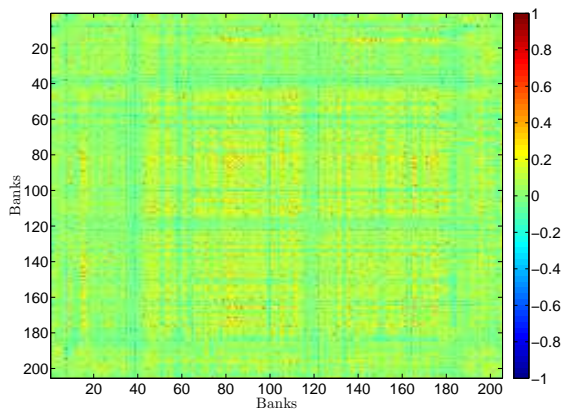
(b) Distribution of S^w



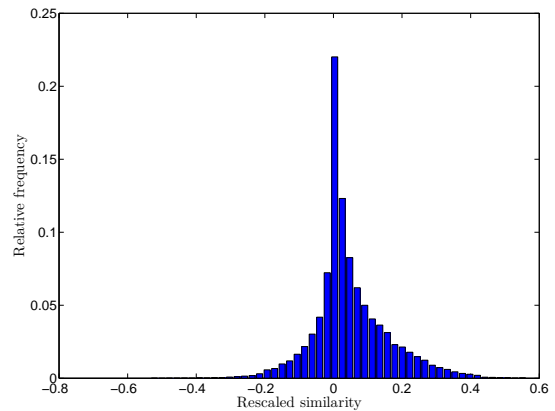
(c) $\mu[S^w]_{WCM}$



(d) Distribution of $\mu[S^w]_{WCM}$

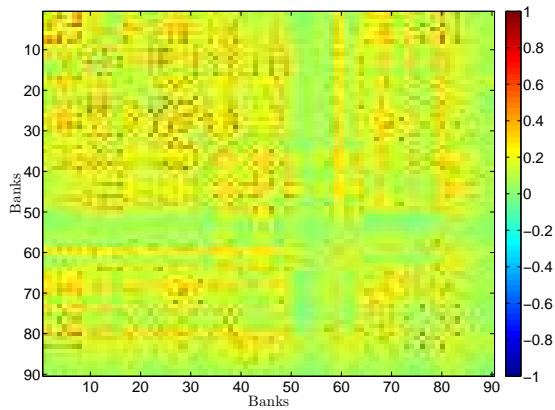


(e) $\mu[S^w]_{ECM}$

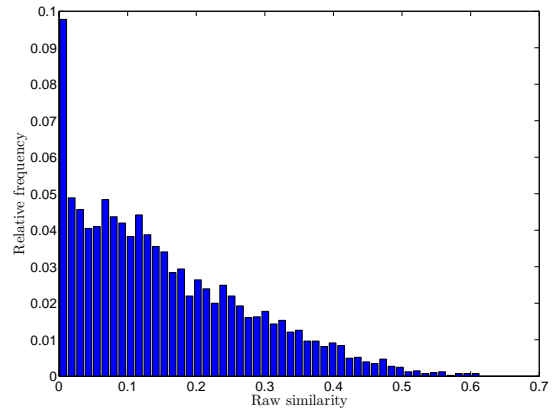


(f) Distribution of $\mu[S^w]_{ECM}$

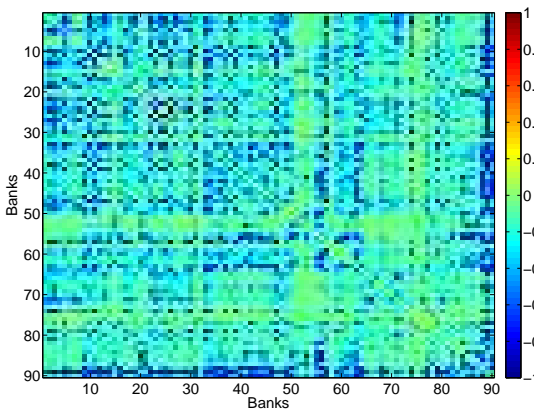
Figure 3.26: Weighted structural similarities in the e-MID network, in the first quarter of 1999. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the WCM and the corresponding distribution. Panels (e), (f): rescaled similarities under the ECM and the corresponding distribution.



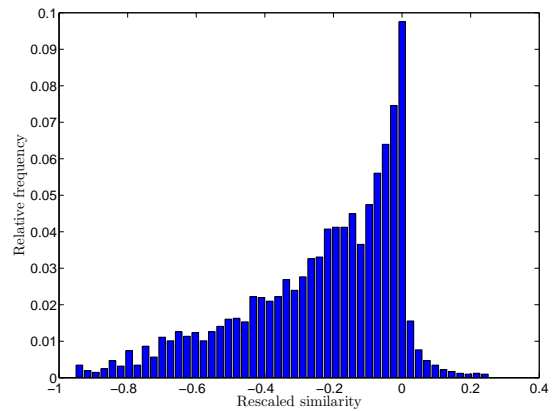
(a) S^w



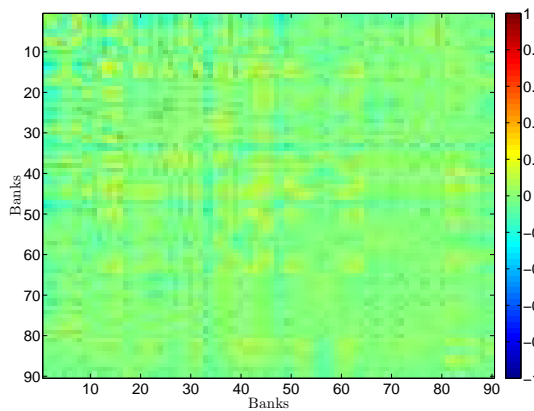
(b) Distribution of S^w



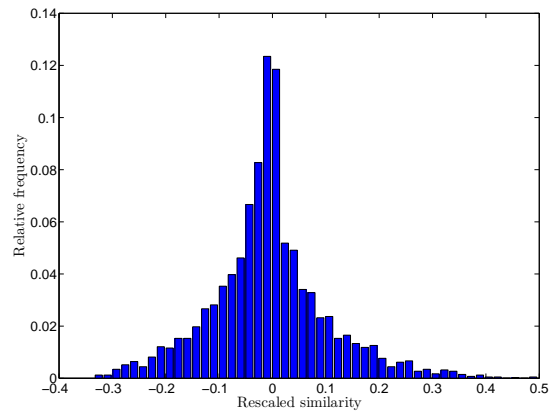
(c) $\mu[S^w]_{WCM}$



(d) Distribution of $\mu[S^w]_{WCM}$



(e) $\mu[S^w]_{ECM}$



(f) Distribution of $\mu[S^w]_{ECM}$

Figure 3.27: Weighted structural similarities in the e-MID network, in the last quarter of 2010. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the WCM and the corresponding distribution. Panels (e), (f): rescaled similarities under the ECM and the corresponding distribution.

Comparisons between z-scores evaluated under the WCM and the ECM

In Figures (3.28) and (3.29) we show the comparisons between the z-scores evaluated under the WCM and ECM in the two networks. In the case of the bank-firm credit network of Spain (Figure (3.28)), we can see that, for pairs of banks i and j satisfying that $|z[S_{ij}^w]_{WCM}| > 2$ as well as $|z[S_{ij}^w]_{ECM}| > 2$, typically we also observe that $|z[S_{ij}^w]_{WCM}| > |z[S_{ij}^w]_{ECM}|$. In the case of the e-MID network, for almost all pairs of banks we have that $|z[S_{ij}^w]_{WCM}| > |z[S_{ij}^w]_{ECM}|$.

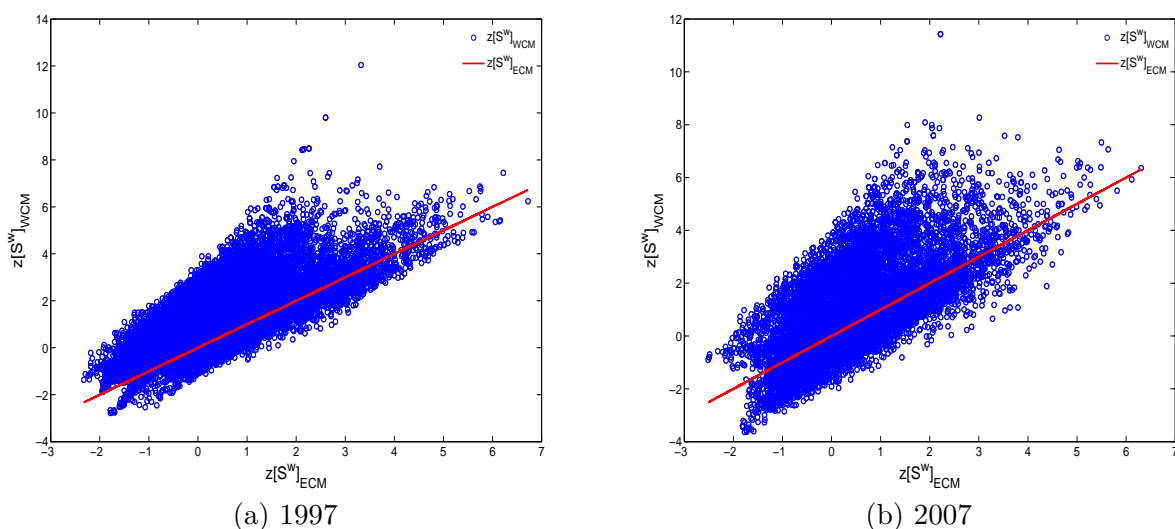


Figure 3.28: Comparison between z-scores evaluated under the WCM and ECM in the bank-firm credit network of Spain, in 1997 (panel a) and 2007 (panel b).

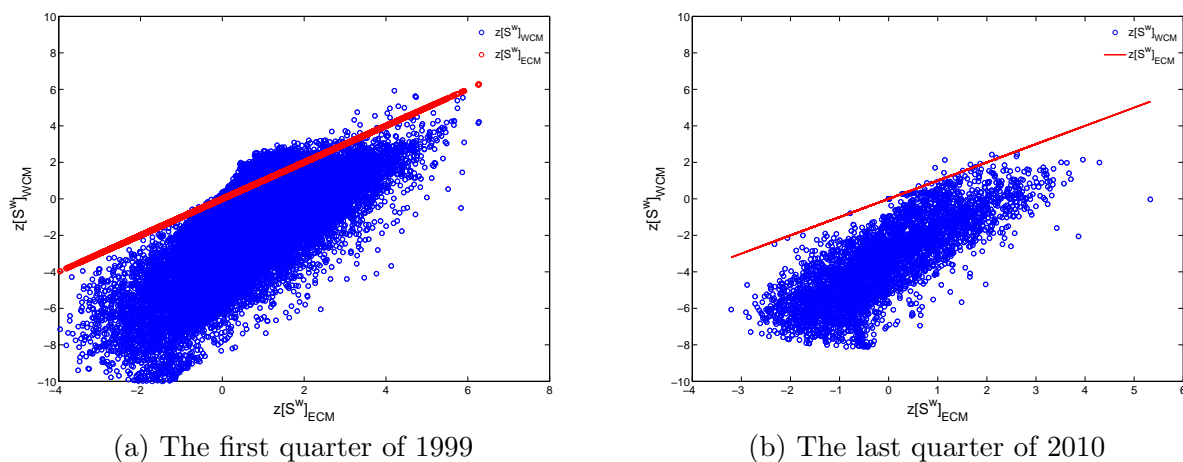


Figure 3.29: Comparison between z-scores evaluated under the WCM and ECM in the e-MID network, in the first quarter of 1999 (panel a) and in the last quarter of 2010 (panel b).

Jaccard-based similarity

As we have mentioned in the main text, we can easily extend our analysis to other measures for the structural similarity such as the Jaccard-based similarity and its rescaled quantity. For the sake of conciseness, in the following, we will compare each observed network with the BCM (in the binary case) and with the ECM (in the weighted case).

The binary version of the Jaccard-based similarities between nodes are shown in Figures (3.30), (3.31) for the bank-firm credit network of Spain and in Figures (3.32), (3.33) for the e-MID network. The weighted version of the Jaccard-based similarities between nodes are shown in Figures (3.34), (3.35) for the bank-firm credit network of Spain and in Figures (3.36), (3.37) for the e-MID network. Overall, we observe similar results between the minimum-based similarity and the Jaccard-based similarity in both versions of the two networks.

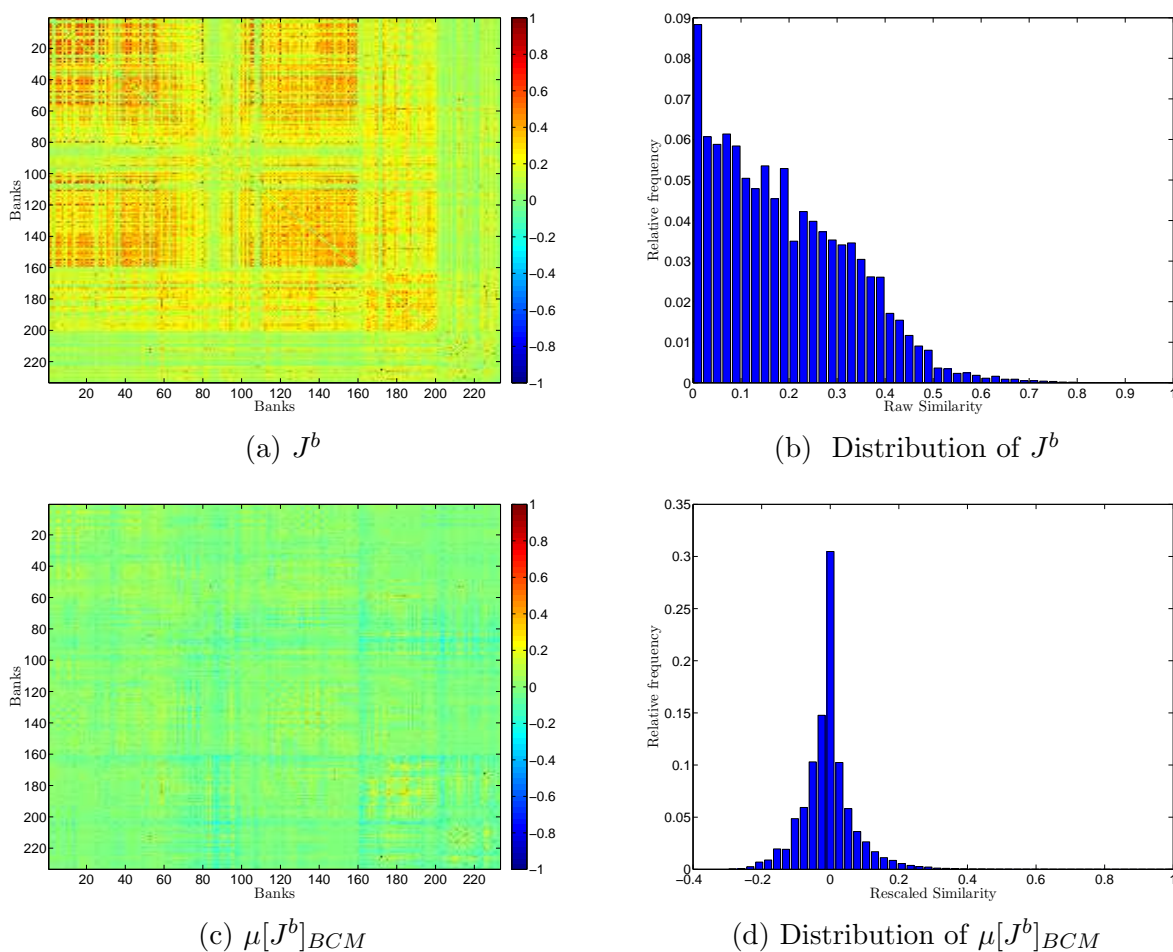


Figure 3.30: Jaccard-based similarities in the bank-firm credit network of Spain in 1997, in the binary version. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the BCM and the corresponding distribution.

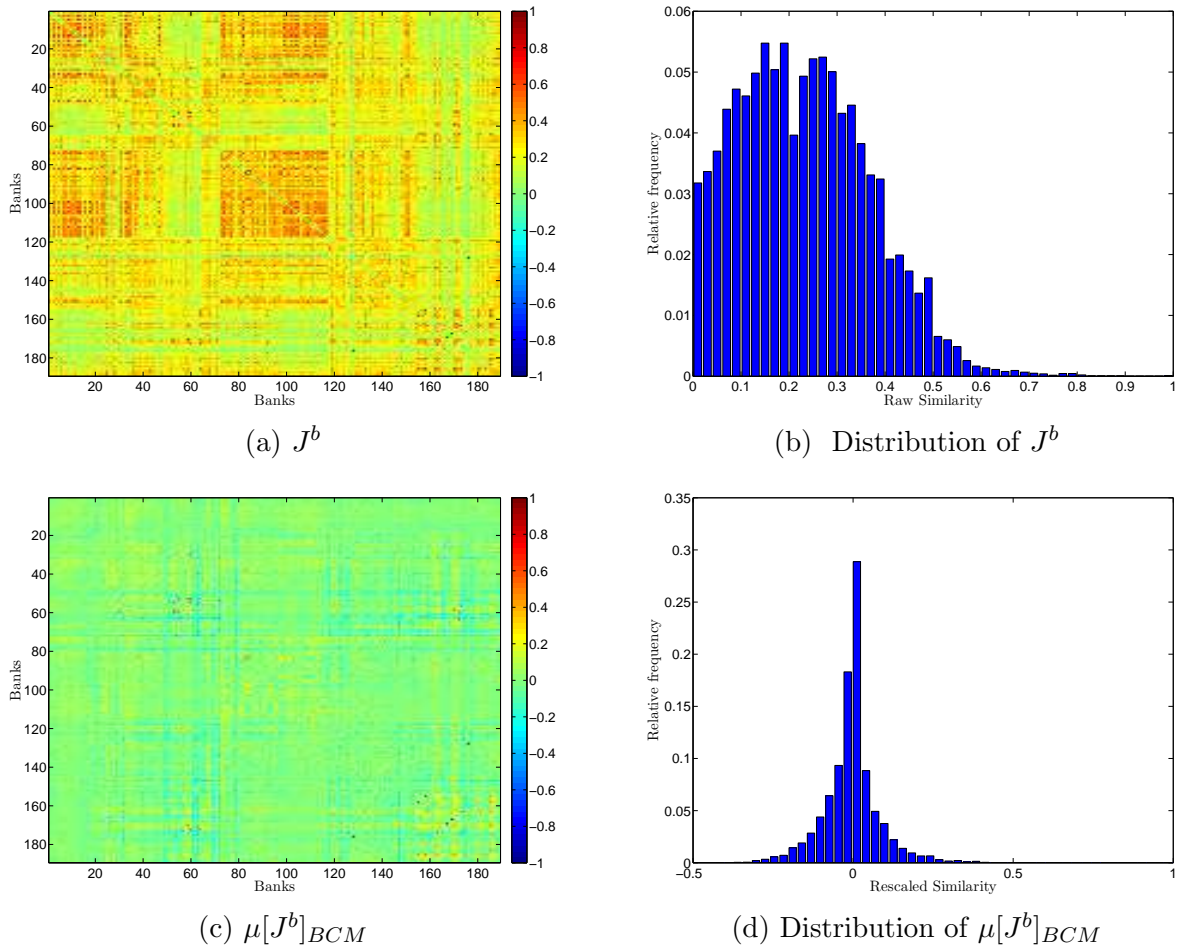


Figure 3.31: Jaccard-based similarities in the bank-firm credit network of Spain in 2007, in the binary version. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the BCM and the corresponding distribution.

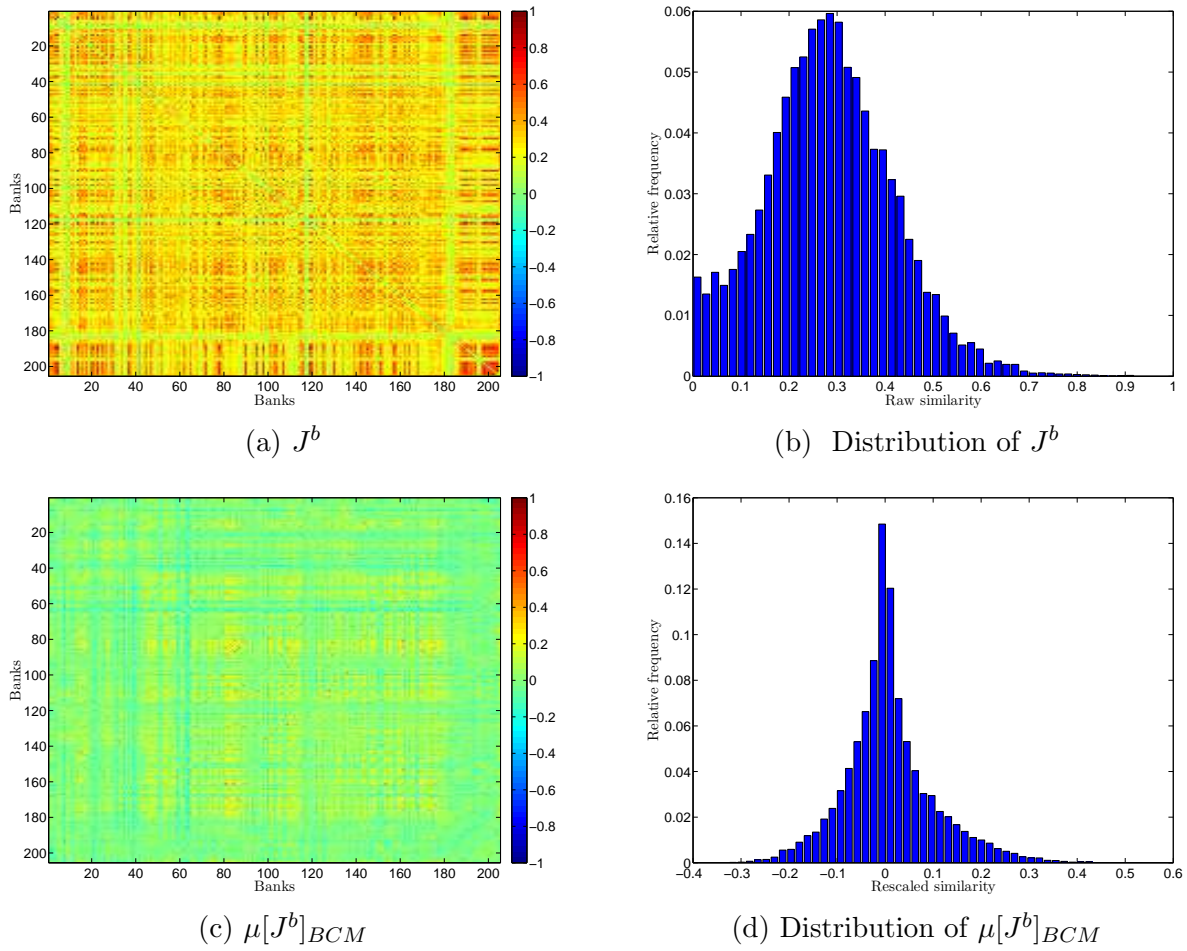
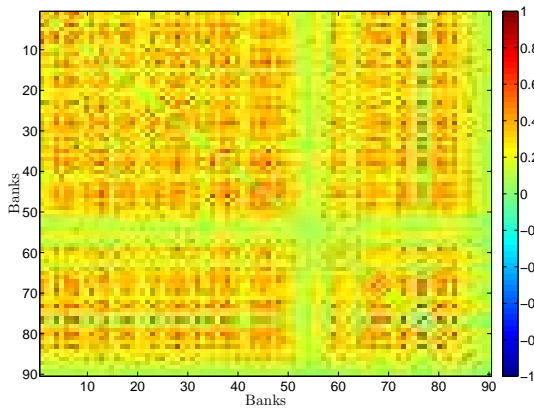
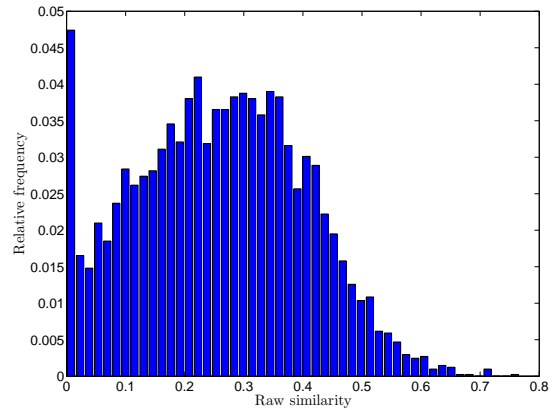


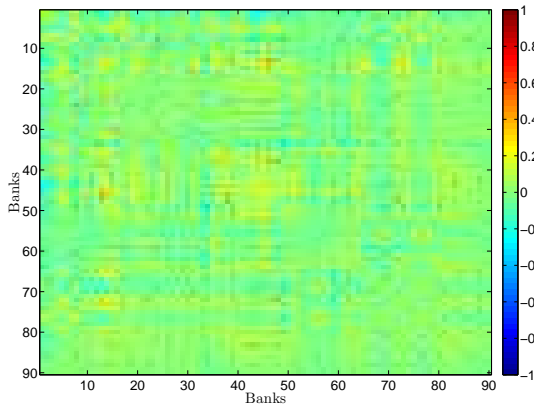
Figure 3.32: Jaccard-based similarities in the e-MID network in the first quarter of 1999, in the binary version. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the BCM and the corresponding distribution.



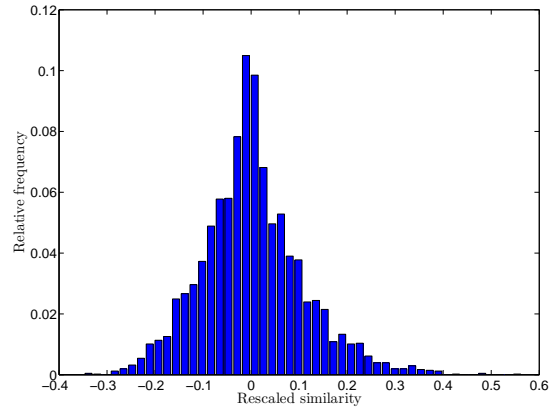
(a) J^b



(b) Distribution of J^b



(c) $\mu[J^b]_{BCM}$



(d) Distribution of $\mu[J^b]_{BCM}$

Figure 3.33: Jaccard-based similarities in the e-MID network in the last quarter of 2010, in the binary version. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the BCM and the corresponding distribution.

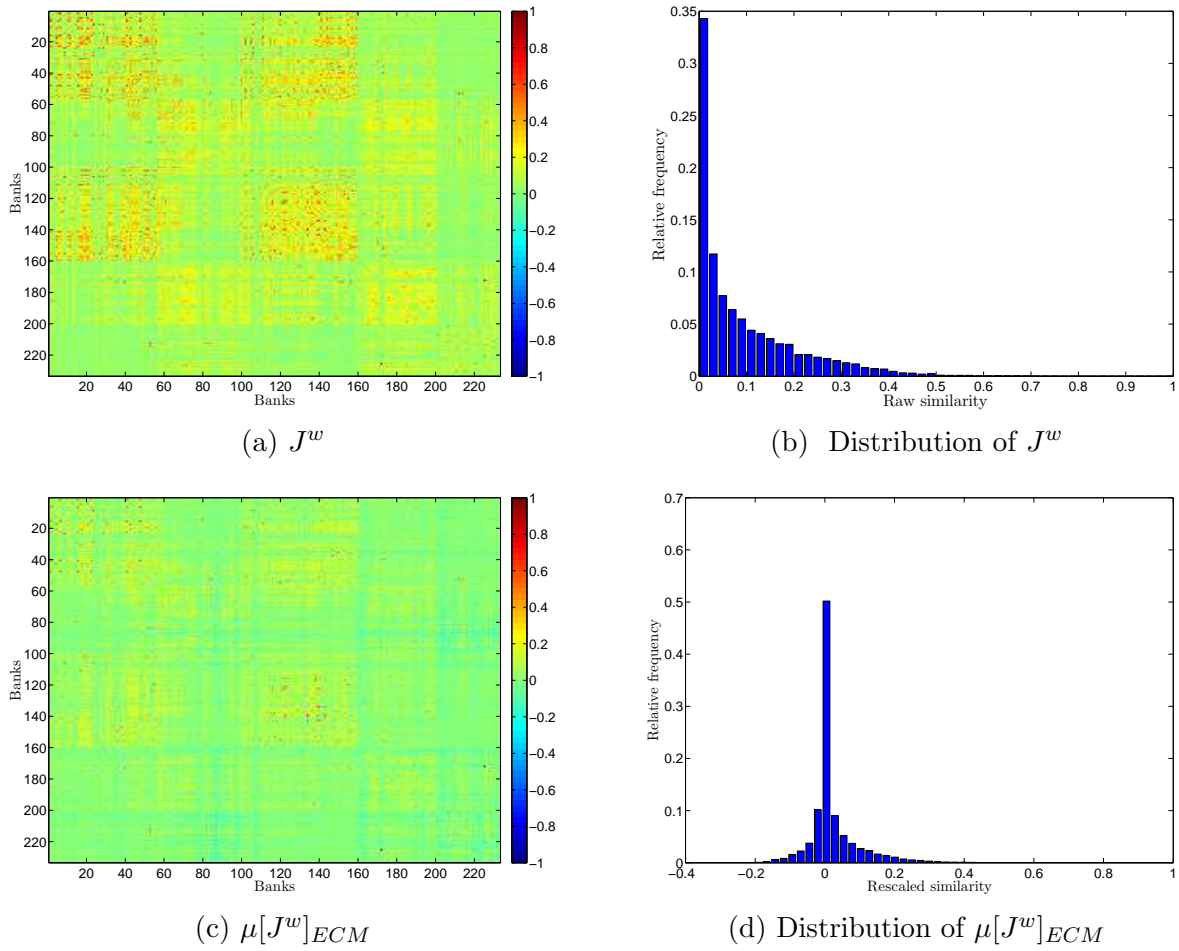


Figure 3.34: Jaccard-based similarities in the bank-firm credit network of Spain in 1997, in the weighted version. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the ECM and the corresponding distribution.

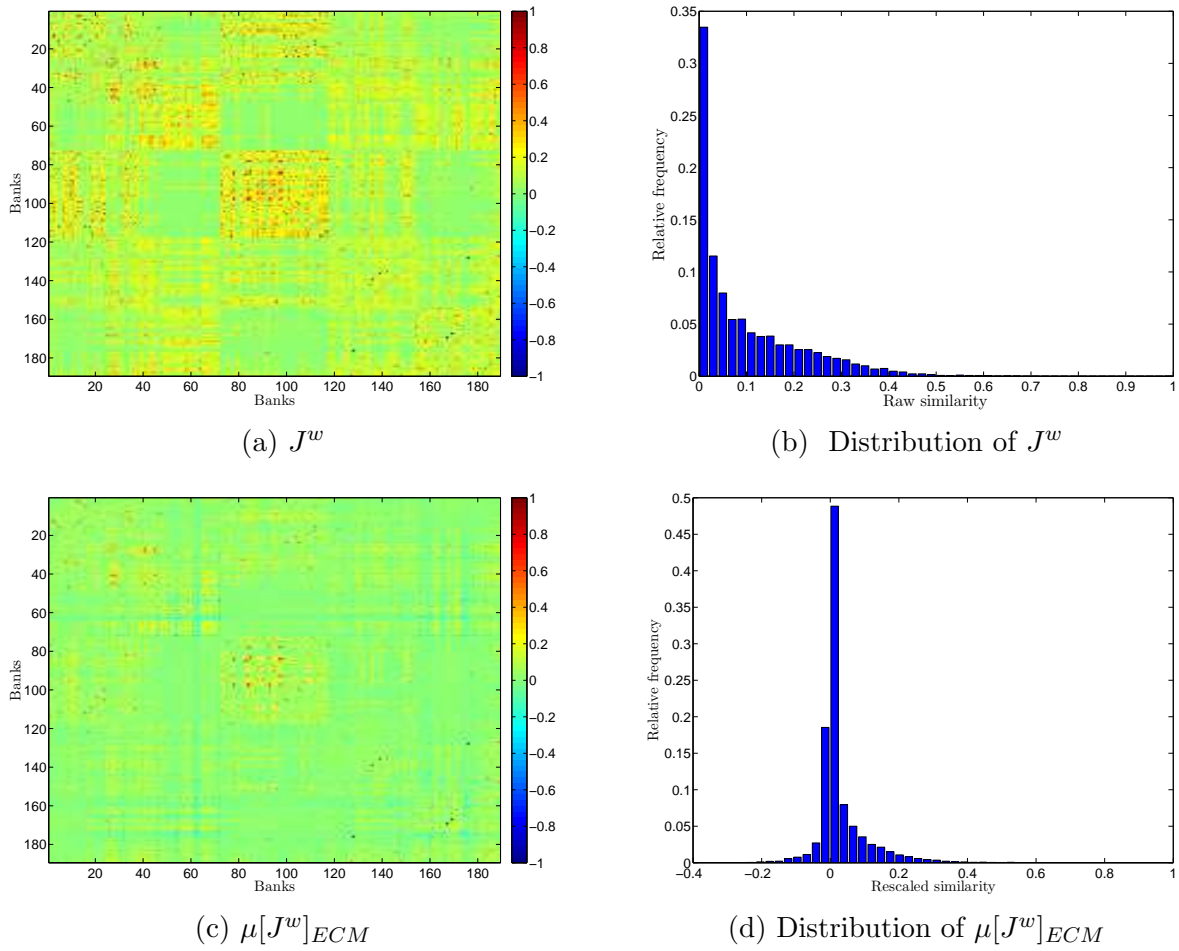


Figure 3.35: Jaccard-based similarities in the bank-firm credit network of Spain in 2007, in the weighted version. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the ECM and the corresponding distribution.

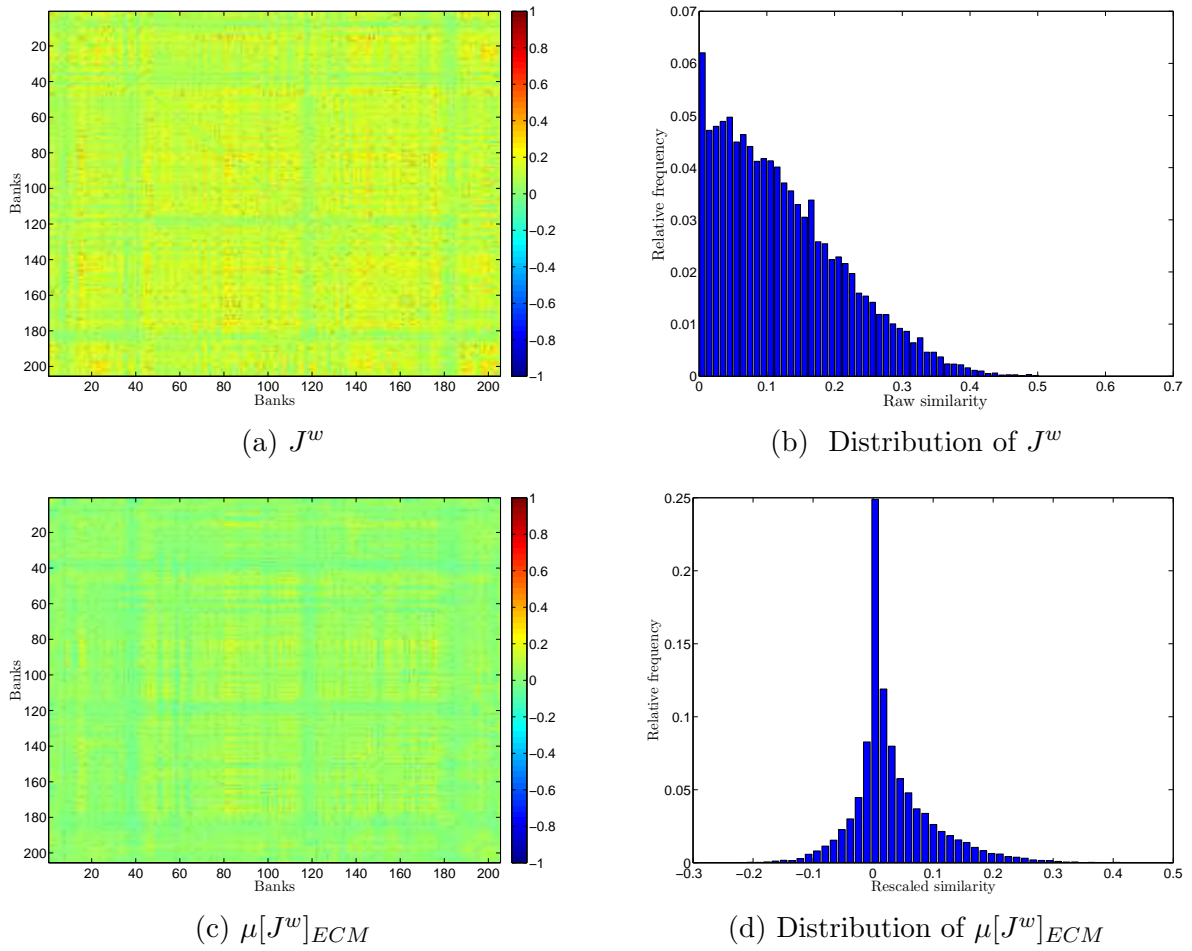
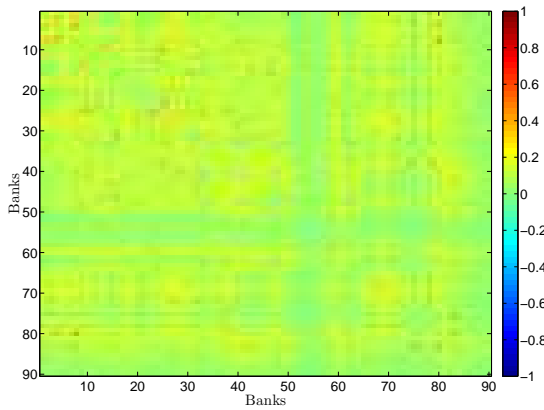
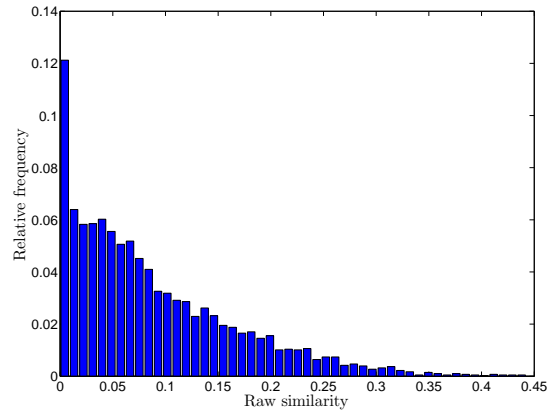


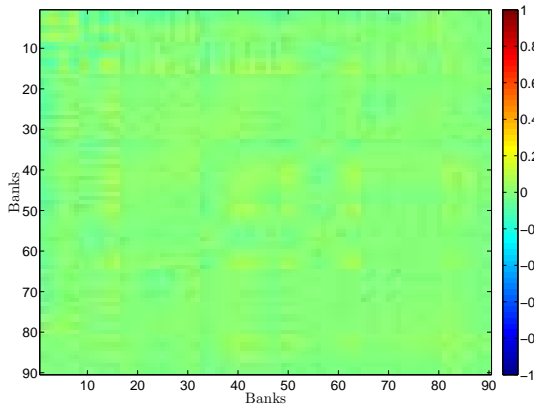
Figure 3.36: Jaccard-based similarities in the e-MID network in the first quarter of 1999, in the weighted version. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the ECM and the corresponding distribution.



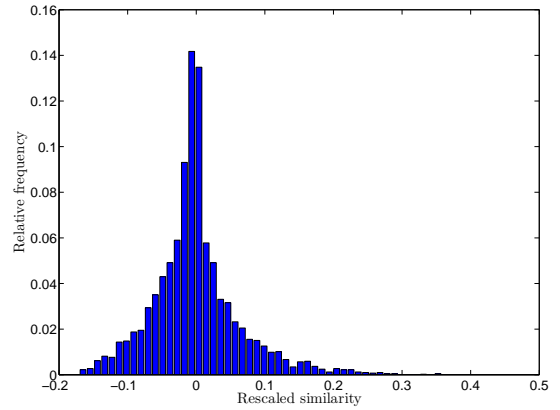
(a) J^w



(b) Distribution of J^w



(c) $\mu[J^w]_{ECM}$



(d) Distribution of $\mu[J^w]_{ECM}$

Figure 3.37: Jaccard-based similarities in the e-MID network in the last quarter of 2010, in the weighted version. Panels (a), (b): raw similarities and the corresponding distribution. Panels (c), (d): rescaled similarities under the ECM and the corresponding distribution.

Chapter 4

Identifying Patterns in the Bank-Sector Credit Network of Spain

Coauthored by: Thomas Lux.

Keywords: Bipartite Networks; Bank-Sector Credit Network; Bipartite Binary Configuration Model; Bipartite Weighted Configuration Model; Bipartite Enhanced Configuration Model.

4.1 Introduction

Many complex networks have a bipartite structure, in which interactions are established between nodes in two different modes. For instance, plant-animal networks in ecology (e.g. Bascompte et al., 2003; Bascompte and Jordano, 2007), scientific collaborations in social networks (e.g. Newman, 2001), document-word networks in information system (e.g. Hofmann, 1999), country-product networks in international trade (e.g. Hidalgo and Hausmann, 2009; Tacchella et al. 2012; Saracco et al., 2015; Saracco et al., 2016), bank-firm credit networks (e.g. De Masi et al., 2011; De Masi and Gallegati, 2012), and bank/investor-asset networks (e.g. Huang et al., 2013; Levy-Carciente et al., 2015; Di Gangi et al., 2015; Squartini et al., 2016) in financial system, city-firm networks in geographical economics (e.g. Garas et al., 2015), and so forth are naturally bipartite. Bipartite networks are mostly analyzed via the projection method by which the original networks are projected into two one-mode networks. This method is convenient as there exists a large number of analytical tools that can be used for analyzing one-mode networks. However, a limitation of the projection method is that the one-mode projection networks are always less informative than the original ones.

For instance, according to Lehmann et al. (2008), (i) all (often sparse) information about the original bipartite relations is projected into denser one-mode networks, and (ii) weights associated with bipartite links are often discarded.

In addition, in order to assess the statistical significance of the properties of the observed networks, one needs to compare such properties with the expected ones obtained from proper null models. However, in contrast to a remarkable number of studies on null models defined for one-mode networks (e.g. see Squartini et al., 2011a; Squartini et al., 2011b; Squartini and Garlaschelli, 2011; Squartini et al., 2013; Mastrandrea et al., 2014; Squartini et al., 2015), less attention has been devoted to defining appropriate null models for bipartite networks. The often used approach is to apply the local rewiring algorithm to generate an ensemble of networks preserving exactly the same degrees as in the observed one. Unfortunately, this method suffers from several limitations such as high computational demand, and lack of analytical penetration of the appropriate null network model ¹. In the recent works of Sarraco et al. (2015) and Sarraco et al. (2016), the so-called Bipartite Binary Configuration Model (BiBCM) maintaining the observed degree sequence is proposed. Moreover, configuration models defined for the weighted version of bipartite networks, i.e. the so called Bipartite Weighted Configuration Model (BiWCM) enforcing the observed strength sequence and the so-called Bipartite Enhanced Configuration Model maintaining both the degree as well as the strength sequences of the observed network, have also recently been introduced in the study of Di Gangi et al. (2015). Unlike in the family of homogeneous benchmark models (i.e. the family of random graph models), in this family of configuration models, the intrinsic heterogeneity in the degrees and/or strengths is preserved. In fact, these bipartite configuration models are the straightforward extensions of the configuration models defined for one-mode networks (e.g. see Squartini et al., 2011a; Squartini et al., 2011b; Squartini and Garlaschelli, 2011; Mastrandrea et al., 2014; Squartini et al., 2015).

In economic and financial networks, the credit linkages between financial institutions and non-financial institutions have been suggested as an important channel of risk contagion in the economy (e.g. Aoyama et al., 2013; Aoyama 2014; Lux, 2016). There exists already some empirical work focusing on the topological features of the empirical bank- firm credit networks (e.g. De Masi et al., 2011; De Masi and Gallegati, 2012; Fujiwara et al., 2009). However, the predominant objective of these early studies has been on basic statistics of network properties such as the distributions of weights, degrees, and strengths. The analysis

¹See Squartini et al. (2015) for a detailed discussion of the limitation of the local rewiring algorithm or Sarraco et al. (2015) for further discussion about the limitations of other null models used for bipartite networks.

of higher order correlations between nodes were often implemented for one-mode projection networks only.

The purposes of our study are twofold: (i) to fill the gap in the analysis of the topological properties of bipartite credit networks; and (ii) to investigate the role of the intrinsic heterogeneity in the observed degrees and/or strengths on the higher-order properties of networks.

This study uses a data set of the undirected binary bank-firm credit network of Spain over the period from 1997 to 2007. We then aggregate the data set into bank-sector relationships based on the industrial codes of firms. We obtain two versions of the bank-sector credit network: (i) the undirected binary bank-sector network, and (ii) the undirected weighted bank-sector network². The advantages of this aggregation procedure are twofold. First, we can reduce the size (typically very huge in our data set) of the original bipartite bank-firm network; second, we obtain also a weighted version of the aggregated network so that we are able to employ the weighted configuration models to assess the statistical significance of the properties of the weighted connections between banks and different sectors of the real economy.

In our analysis, first, focusing on the assortativity, different orders of motifs, clustering coefficients, and the nestedness phenomenon, we will directly analyze the original bipartite structure of the bank-sector network instead of using the one-mode projection method. Next, employing the various bipartite configuration models, we also aim to explore whether the observed values of the network properties can be explained by information encoded in the fundamental constraints of the network like the degree and/or the strength sequences, or whether significant, non-random patterns can be identified that can not straightforwardly be explained by lower-order characteristics of the data. This is very important in not only defining proper null models but also in reconstructing the real bipartite credit networks from limited information.

The remainder of this paper is structured as follows. In Sec. 4.2, after briefly describing the data, we introduce general definitions, the methods used for measuring topological and structural properties of a bipartite network, and the family of the bipartite configuration models used for randomizing networks from given constraints. Sec. 4.3 summarizes the main findings for the binary and weighted versions of the bank-sector credit network of Spain. Discussions and concluding remarks are in Sec. 4.4. At the end of this paper, the Appendix provides further explanations of the methods and additional results.

²Each weight indicates the total number of firms in each sector that a bank lend to.

4.2 Data and Methods

4.2.1 Data

In this study, we analyze a yearly data set on bank-firm credit relationships in Spain over the period from 1997 to 2007. The information of firms is based on the so-called SABI database (Sistema de Análisis de Balances Ibéricos) by Bureau van Dijk³. The database provides accounting and financial data as well as additional information for the characteristics of firms (e.g. their headquarter locations and sectors). It also provides the ID numbers of banks that firms have financial relationships with. We assume that if a firm reports the ID number of a bank, it indicates that the firm borrows from that bank as well. For a more detailed explanation of the evolution of the bank-firm credit market of Spain over the sample period, we refer the readers to, for example, Illueca et al. (2013).

Note that, the sector codes are based on “The Statistical Classification of Economic Activities in the European Community” (NACE), level 2, which is identified by two-digit numerical codes (01 to 99). Overall, firms are located in around 60 industrial sectors over the period from 1997 to 2007. In the next part, we will explain the procedure used for obtaining the binary as well as weighted versions of bank-sector credit relationships from the original bank-firm credit market.

4.2.2 General definitions

Suppose that a bipartite bank-sector credit network consists of two modes $B = \{b_i\}_{i=1}^n$ standing for n banks and $S = \{s_j\}_{j=1}^m$ standing for m sectors. It is characterized by an adjacency matrix $A = A^{B-S} = \{A_{ij}\}_{n \times m}$ and a weighted matrix $W = W^{B-S} = \{W_{ij}\}_{n \times m}$.

From the economic perspective, in the binary version, for every pair of i and j ($1 \leq i \leq n$, $1 \leq j \leq m$), $A_{ij} = 1$ if bank i lends to at least one firm in sector j , and $A_{ij} = 0$ otherwise. In the weighted version, since we do not have detailed information about the amount of loans from banks to firms in the original bank-firm credit network, here we consider the total number of firms in each sector that borrow from each bank as the proxy for the weight between that bank and that sector. Therefore, in our study, mathematically, W_{ij} is nothing else but the total number of firms in sector j borrowing from bank i .

³For the introduction of the database, see, for example, Bureau van Dijk’s website: <http://www.bvdinfo.com/en-gb/our-products/company-information/national-products/sabi>.

We define the total degree as

$$L(A) = \sum_{i=1}^n \sum_{j=1}^m A_{ij}. \quad (4.1)$$

The degree of each node i in the mode B and the degree of each node j in the mode S are respectively given by

$$\text{kbs}_i = \sum_{j=1}^{j=m} A_{ij}, \quad (4.2)$$

and

$$\text{ksb}_j = \sum_{i=1}^{i=n} A_{ij}. \quad (4.3)$$

Similarly, in the weighted version of the network, the total strength is defined as

$$S(W) = \sum_{i=1}^n \sum_{j=1}^m W_{ij}. \quad (4.4)$$

The strength of each node i in the mode B and the strength of each node j in the mode S are respectively given by

$$\text{sbs}_i = \sum_{j=1}^{j=m} W_{ij}, \quad (4.5)$$

and

$$\text{ssb}_j = \sum_{i=1}^{i=n} W_{ij}. \quad (4.6)$$

4.2.3 Methods used for the analysis of binary bipartite networks

Projection matrices and overlaps between nodes

Taking the one-mode projection of an original bipartite network, we will obtain the two matrices, i.e. $A^{B-B} = \{A_{ij}^{B-B}\}_{n \times n}$ indicating the overlaps between nodes in the mode B and $A^{S-S} = \{A_{ij}^{S-S}\}_{m \times m}$ indicating the overlaps between nodes in the mode S. Mathematically,

$$A^{B-B} = A^{B-S} A^{S-B}, \quad (4.7)$$

and

$$A^{S-S} = A^{S-B} A^{B-S}, \quad (4.8)$$

where A^{S-B} is the transpose of A^{B-S} .

Note that in the case of the bank-sector credit market, each off-diagonal element of the symmetric matrix A^{B-B} is the number of sectors in which a pair of banks provides loans to at least one common company. Similarly, each off-diagonal element of the symmetric matrix A^{S-S} is the number of common banks that a pair of sectors obtain loans from. The diagonal elements of A^{B-B} (A^{S-S}) are nothing else but the node degrees in the mode B (S) as given in Eq. (4.2) (Eq. (4.3)).

So far, the analysis of bipartite networks has often been implemented with one-mode projection networks obtained from Eq. (4.7) and Eq. (4.8). The corresponding null models used to assess the statistical significance of the topological properties of these networks are therefore also null models defined for one-mode networks. As a consequence, it is difficult to assess in how far the properties of the original bipartite network, for instance the degree and/or strength distributions, can explain the particular empirical structure of the original bipartite network as well as that of the one-mode projected networks. We are going to answer that question in the next section.

Assortativity

Regarding the assortativity level of a network, we select two measures, i.e. the average degree of the nearest neighbors (ANND) as well as the Pearson correlation coefficient of degrees. Respectively, the average degree of the nearest neighbor, $kbs_{nn}(i)$, of each node i in the mode B and the average degree of the nearest neighbors, $ksb_{nn}(j)$, of each node j in the mode S are given by

$$kbs_{nn}(i) = \frac{\sum_{j=1}^n A_{ij}^{B-S} ksb_j}{kbs_i}, \quad (4.9)$$

and

$$ksb_{nn}(j) = \frac{\sum_{i=1}^n A_{ij}^{B-S} kbs_i}{ksb_j}. \quad (4.10)$$

Generally speaking, an overall positive (negative) correlation between ANNDs and node degrees indicates an assortative (disassortative) mixing nature in the network. Practically, we will plot kbs_{nn} and ksb_{nn} respectively against kbs and ksb to obtain the overall trend in the mixing nature.

The second measure is based on the Pearson correlation coefficient of degrees between pairs of linked nodes (e.g. Newman, 2002; Newman, 2003). The main idea is that, first, from the adjacency matrix A^{B-S} , we obtain a list of L edges, i.e. the list of pairs of two nodes (i_e, j_e) where $A_{i_e j_e}^{B-S} = 1$ for every e from 1 to L ($1 \leq i_e \leq n$, $1 \leq j_e \leq m$). From that, we get two degrees kbs_{i_e} and ksb_{j_e} associated with each pair of two nodes (i_e, j_e) . The correlation coefficient of degrees (r) is then obtained by measuring the the Pearson correlation coefficient

between two sequences $\{\text{kbs}_{i_e}\}_{e=1}^{e=L}$ and $\{\text{ksb}_{j_e}\}_{e=1}^{e=L}$. If r is negative, it signals the presence of disassortativity, while positive value of r implies the opposite mixing nature.

Bipartite motifs

Recurrent patterns of connections within subsets of a network are called "motifs" in pertinent literature. Unlike in one-mode networks, motifs in bipartite networks must consist of nodes in two different modes. In our study, following Saracco et al. (2015) and Saracco et al. (2016), we analyze different families of motifs that capture different types of structural correlations between nodes in bipartite networks.

In the following, first we will define the so-called $V(k)$ and $\Lambda(k)$ motifs, and then we move on to the definitions of other types of motifs (see Figure (4.15) in the Appendix). Latter, we will see that, in fact, the higher-order correlations between nodes can be inductively formulated as $V(k)$ and $\Lambda(k)$ motifs.

*V(k) and $\Lambda(k)$ motifs*⁴

Briefly, for every $k \geq 2$, the number of $V(k)$'s (i.e. when k nodes in the mode B all link to a node in the mode S) and the number of $\Lambda(k)$'s (i.e. when k nodes in the mode S all link to a node in the mode B) are given by

$$N_{V(k)}(A) = \sum_{j=1}^{j=m} \binom{\text{ksb}_j}{k}, \quad (4.11)$$

and

$$N_{\Lambda(k)}(A) = \sum_{i=1}^{i=n} \binom{\text{kbs}_i}{k} \quad (4.12)$$

(see Figure (4.15)). Note that, if $k=2$ we will obtain the number of motif types $V=V(2)$ and $\Lambda=\Lambda(2)$ indicating the overall overlaps between nodes in a bipartite network (for instance, in the bank-sector network, $V(2)$ and $\Lambda(2)$ are respectively the total number of pairs of banks and the total number of pairs of sectors that have a common neighbor).

X motifs

The X motifs represent the 4-cycles in a bipartite network, in which two nodes in the mode B are all connected to two nodes in the mode S (see Figure (4.15)). The number of these motifs is given by

$$N_X(A) = \sum_{1 \leq i < i' \leq n} \binom{A_{ii'}^{B-B}}{2} = \sum_{1 \leq j < j' \leq m} \binom{A_{jj'}^{S-S}}{2}. \quad (4.13)$$

⁴Throughout this paper, the notation N_Y stands for the number of motifs Y.

Decomposing $N_X(A)$ in Eq. (4.13) into the contribution of each node, we will obtain the local number of X motifs. More specifically, the local number of X motifs of a node i_0 in the mode B is $N_X(i_0) = \sum_{i' \neq i_0} \binom{A_{i_0 i'}^{B-B}}{2}$, and the local number of X motifs of a node j_0 in the mode S is $N_X(j_0) = \sum_{j' \neq j_0} \binom{A_{j_0 j'}^{S-S}}{2}$.

Based on the local number of of X motifs, the local clustering coefficient of a node h (h is either in the mode B or in the mode S) is then given by

$$C_4^{bin}(h) = \frac{N_X(h)}{K_{nn}(h) \frac{k(h)(k(h)-1)}{2}}, \quad (4.14)$$

with $k(h)$ and $K_{nn}(h)$ are respectively the degree and the number of the second nearest neighbors⁵ of the node h (e.g. see Lind et al., 2005; Fernandez et al., 2012; Gilarranz et al., 2012). We can see that $C_4(h)$ is nothing else but the probability that two neighbors of the node h have another common neighbor in addition to the node h . In fact, this quantity is often used as the local clustering coefficient for each node in a binary bipartite network.

M and W motifs

Graphically, each M motif is composed by two nodes in the mode B and three nodes in the mode S, while each W motif is composed by three nodes in the mode B and two nodes in the mode S (see Figure (4.15)). Mathematically, we have

$$N_M(A) = \sum_{1 \leq i < i' \leq n} \binom{A_{ii'}^{B-B}}{3}, \quad (4.15)$$

and

$$N_W(A) = \sum_{1 \leq j < j' \leq m} \binom{A_{jj'}^{S-S}}{3}. \quad (4.16)$$

Obviously, the generalization of motifs can capture the higher-order structural correlations between nodes in a bipartite network. For example, we can define a family of motifs (or cliques) $K(h,k)$ as a complete sub-bigraph of A , with h nodes in the mode B are all linked to k nodes in the mode S. In addition, from that, a $K(h,k)$ community can be defined as a group of all possible $K(h,k)$ that each of them can be reached from each other through series of adjacent $K(h,k)$ motifs (e.g. Lehmann et al., 2008). For the detailed derivations and extensions of motifs, we provide more information in the Appendix. Within the scope of our study, we focus on the V, A, X, M, and W families of motifs.

⁵For example, given a node i in the mode B, all nodes in the mode S that are connected to the node i are defined as the first nearest neighbors of the node i . The second nearest neighbors of the node i are then nodes in the mode B that are connected to its first nearest neighbors.

It is worthwhile to emphasize that a particular application of a motifs-based analysis in bipartite credit networks is that the motifs reveal different degrees of overlaps between lending portfolios and between borrowing portfolios.

Nestedness

The phenomenon of “nestedness” is found in many mutualistic networks in ecology (e.g. Bascompte et al., 2003; Almeida-Neto et al., 2008; Staniczenko et al., 2013). For example, considering interactions between species in an ecological network, nestedness occurs when the more specialist species interact only with proper subgroups of those species interacting with the more generalist ones (Bascompte et al., 2003). In the binary case, this phenomenon is often measured through the so-called presence-absence matrix. Graphically, after reordering vertices in each mode according to a certain criterion (e.g. based on node degrees), the presence of upper triangular structure of the adjacency matrix reveals the nestedness phenomenon. In this paper, in line with the previous studies of Saracco et al. (2015) and Saracco et al. (2016), in order to measure the nestedness, we employ the method of overlap and decreasing completeness proposed by Almeida-Neto et al. (2008). Given the adjacency matrix A , suppose that two matrices S and T are defined as

$$S_{ii'}^{B-B} = \begin{cases} \frac{\sum_j A_{ij}A_{i'j}}{\min(kbs_i, kbs_{i'})}, & \text{if } kbs_i \neq kbs_{i'} \\ 0, & \text{otherwise,} \end{cases} \quad (4.17)$$

and

$$T_{jj'}^{S-S} = \begin{cases} \frac{\sum_i A_{ij}A_{ij'}}{\min(ksb_j, ksb_{j'})}, & \text{if } ksb_j \neq ksb_{j'} \\ 0, & \text{otherwise.} \end{cases} \quad (4.18)$$

The normalized nestedness is given by

$$NODF = 2 \frac{\sum_{1 \leq i < i' \leq n} S_{ii'} + \sum_{1 \leq j < j' \leq m} T_{jj'}}{n(n-1) + m(m-1)}. \quad (4.19)$$

Here we consider $NODF$ as the overall indicator for the nestedness of the whole network in the binary version. We can define the overall nestedness indicator $NODF_B$ for all nodes in the mode B as

$$NODF_B = 2 \frac{\sum_{1 \leq i < i' \leq n} S_{ii'}}{n(n-1)}, \quad (4.20)$$

and similarly the overall nestedness indicator $NODF_S$ for all nodes in the mode S as

$$NODF_S = 2 \frac{\sum_{1 \leq j < j' \leq m} T_{jj'}}{m(m-1)}. \quad (4.21)$$

Given the network's size, a higher value of NODF indicates a stronger level of nested structure in the network. If $NODF_B$ is larger (smaller) than $NODF_S$, nodes in the mode B contribute more (less) to the overall nestedness indicator NODF than nodes in mode S do. In addition, the local nestedness indicators can be defined as the contribution of single nodes to the overall nestedness indicators in each mode.

4.2.4 Methods used for the analysis of weighted bipartite networks

In this part, we will briefly introduce the methods used for the analysis of weighted bipartite networks. For the sake of simplicity, we focus on measures for weighted assortativity, weighted X motifs, and local clustering coefficients.

Weighted assortativity

Similar to the definition of ANND in the binary version, the average strengths of the nearest neighbors (ANNS) in the weighted version are defined as

$$sbs_{nn}(i) = \frac{\sum_{j=1}^m A_{ij}^{B-S} sbs_j}{kbs_i} \quad (4.22)$$

and

$$ssb_{nn}(j) = \frac{\sum_{i=1}^n A_{ij}^{B-S} sbs_i}{ksb_j} \quad (4.23)$$

A positive (negative) correlation between ANNSs and strengths will indicate an assortative (disassortative) mixing nature in the weighted network.

Note that, we can also measure the Pearson correlation coefficient of strengths. Similar to the binary case, given the list of edges L , it is equal to the correlation coefficient r_w between two strength sequences $\{s_{i_e}^b\}_{e=1}^{e=L}$ and $\{s_{j_e}^s\}_{e=1}^{e=L}$. If r_w is positive (negative), it implies the presence of disassortativity (assortativity) in the weighted network.

Weighted motifs and clusterings

Recalling Eq. (4.14), in the weighted version, for each link between two nodes, we should take into account the associated weight. Following the idea of Onnela et al. (2005), for each node $i_0 \in [1, n]$ in the mode B and each node $j_0 \in [1, m]$ in the mode S, the local weighted

clustering coefficients based on X motifs are defined as

$$C_4^w(i_0) = \frac{\sum_{i' \neq i_0} \sum_{1 \leq j < j' \leq m} w_{i_0 j}^{\frac{1}{4}} w_{i' j}^{\frac{1}{4}} w_{i_0 j'}^{\frac{1}{4}} w_{i' j'}^{\frac{1}{4}}}{K_{nm}(i_0) \text{kbs}_{i_0} (\text{kbs}_{i_0} - 1)}, \quad (4.24)$$

and

$$C_4^w(j_0) = \frac{\sum_{j' \neq j_0} \sum_{1 \leq i < i' \leq n} w_{i j_0}^{\frac{1}{4}} w_{i' j_0}^{\frac{1}{4}} w_{i j'}^{\frac{1}{4}} w_{i' j'}^{\frac{1}{4}}}{K_{nm}(j_0) \text{ksb}_{j_0} (\text{ksb}_{j_0} - 1)}. \quad (4.25)$$

Similar to the binary case, the denominators in Eqs. (4.24) and (4.25) are the total number of possible 4-paths around the considered nodes (e.g. see Fernandez et al. 2012; Gilarranz et al., 2012). It is easy to show that, in the special case when $W \equiv A$, we immediately get $C_4^w \equiv C_4^{bin}$.

4.2.5 Bipartite Configuration Models

We now introduce the Exponential Random Graph Model (e.g. Park and Newman, 2004) for complex networks and the derivation for the family of bipartite configuration models maintaining the intrinsic heterogeneity in the observed degrees and/or strengths. More specifically, in the Bipartite Binary Configuration Model (BiBCM), the observed degree sequence is enforced on average over the ensemble of randomized networks. In contrast, in the Bipartite Weighted Configuration Model (BiWCM), the observed strength sequence is preserved. In the Bipartite Enhanced Configuration Model (BiECM), both the observed degree sequence as well as the observed strength sequence are maintained.

Exponential Random Graph Model

Based on the Exponential Random Graph Model (e.g. Park and Newman, 2004), we define a group \mathcal{G} of N graphs $\{G_j\}_{j=1}^{j=N}$ ($N \in [1, +\infty)$) such that

$$\langle c_i(G) \rangle = \sum_G P(G) c_i(G) = c_i^*, \forall i = 1, 2, \dots, k, \quad (4.26)$$

where $P(G)$ is the probability distribution of G in \mathcal{G} , and $C(G) = \{c_i^*\}_{i=1}^{i=k}$ are the given constraints, such as the topological properties of the observed network that we want to preserve ($k \ll N$).

One can define the Shannon-Gibbs entropy as a function of $P(G)$

$$S = - \sum_G P(G) \ln P(G). \quad (4.27)$$

The maximum entropy distribution $P(G)$ of a family of graphs G satisfying k constraints in (4.26) is obtained by finding P such that

$$\begin{cases} \max_{P(G)} S(P(G)) \\ s.t. : \\ \sum_{G \in \mathcal{G}} P(G) = 1 \\ \sum_{G \in \mathcal{G}} P(G) c_i(G) = c_i^*, \forall i = 1, 2, \dots, k. \end{cases} \quad (4.28)$$

Since there are now $(k+1)$ constraints, to solve (4.28) we need to define $(k+1)$ Lagrange multipliers, including $\{\lambda_i\}_{i=1}^{i=k}$ associated with (4.26) and α for the normalization condition of $P(G)$ in \mathcal{G} . That leads to

$$\frac{\partial}{\partial P(G)} [S + \alpha(1 - \sum_G P(G)) + \sum_i \lambda_i (c_i^* - \sum_G P(G) c_i(G))] = 0, \forall G \in \mathcal{G}. \quad (4.29)$$

From (4.29) we obtain

$$\ln P(G) + 1 + \alpha + \sum_{i=1}^k \lambda_i c_i(G) = 0, \quad (4.30)$$

Expressed differently, we have

$$P(G) = \frac{e^{-\sum_{i=1}^k \lambda_i c_i(G)}}{e^{\alpha+1}}, \forall G \in \mathcal{G}. \quad (4.31)$$

We define the graph Hamiltonian function as

$$H(G) = \sum_{i=1}^k \lambda_i c_i(G), \quad (4.32)$$

and the partition function as

$$Z = \sum_{G \in \mathcal{G}} e^{-H(G)}. \quad (4.33)$$

The normalization condition of P requires $Z = e^{\alpha+1}$. It is clear that H and Z depend on the Lagrange multipliers. From Eq. (4.31), (4.32), and (4.33) we get the general form for $P(G)$ as

$$P(G) = \frac{e^{-H(G|\vec{\theta})}}{Z(\vec{\theta})}, \forall G \in \mathcal{G}, \quad (4.34)$$

where here $\vec{\theta}$ is the vector of Lagrange multipliers associated with the constraints ⁶. The model explained by Eq. (4.32), (4.33), and (4.34) is the so-called Exponential Random Graph (ERG) model, i.e., the distribution of graph probability over a specified ensemble that maximizes the Shannon-Gibbs entropy subject to given constraints (Park and Newman, 2004).

Note that, in all above derivations we have not specified the type of graphs G or the constraints we want to preserve yet. If we consider only one-mode networks, we can define the family of configuration models preserving the given constraints in one-mode networks, for instance as in the study of Squartini et al. (2015). In this paper, we will focus on describing the configuration models particularly used for bipartite networks. We consider three types of constraints: the degree sequence, the strength sequence, and both the degree as well as the strength sequences.

Bipartite Configuration Models

Bipartite Binary Configuration Model (BiBCM)

We need to define a set \mathcal{A} of the bipartite binary networks A (A consists of nodes in two modes $B = \{B_i\}_1^n, S = \{S_j\}_1^m$) that enforces the degree sequence $C(A) = \{\{kbs_i\}_1^n, \{ksb_j\}_1^m\}$ of A . Under the BiBCM, let $P_{\text{BiBCM}}(A)$ be the probability of each A in the ensemble \mathcal{A} . The maximum entropy problem is

$$\begin{cases} \max_{P_{\text{BiBCM}}(A)} S(P_{\text{BiBCM}}(A)) \\ s.t. : \\ \sum_{A \in \mathcal{A}} P_{\text{BiBCM}}(A) = 1 \\ \langle kbs_i(A) \rangle = kbs_i, \forall i = 1, 2, \dots, n, \\ \langle ksb_j(A) \rangle = ksb_j, \forall j = 1, 2, \dots, m. \end{cases} \quad (4.35)$$

Following what we have generally explained for the ERG model, we obtain the solution to $P_{\text{BiBCM}}(A)$ in (4.35) as

$$P_{\text{BiBCM}}(A|\vec{\theta}) = \frac{e^{-H(A|\vec{\theta})}}{Z(\vec{\theta})}, \forall A \in \mathcal{A}, \quad (4.36)$$

where

$$H(A|\vec{\theta}) = \sum_{i=1}^n \lambda_i kbs_i(A) + \sum_{j=1}^m \eta_j ksb_j(A), \quad (4.37)$$

⁶Now, in a more generic form, we allow for the variation in the Lagrange multipliers.

and the partition function

$$Z(\vec{\theta}) = \sum_A e^{-H(A|\vec{\theta})}, \quad (4.38)$$

where $\vec{\theta}$ is the vector of the Lagrange multipliers associated with the degrees in problem (4.35).

Note that

$$H(A|\vec{\theta}) = \sum_{i=1}^n \lambda_i \text{kbs}_i(A) + \sum_{j=1}^m \eta_j \text{ksb}_j(A) = \sum_{i=1}^n \sum_{j=1}^m A_{ij} (\lambda_i + \eta_j), \quad (4.39)$$

and

$$Z(\vec{\theta}) = \sum_A e^{-\sum_{i=1}^n \lambda_i \text{kbs}_i(A) - \sum_{j=1}^m \eta_j \text{ksb}_j(A)} = \sum_A \prod_{i=1}^n \prod_{j=1}^m e^{-A_{ij} (\lambda_i + \eta_j)}. \quad (4.40)$$

In addition, according to Park and Newman (2004), as A_{ij} takes all of its possible realizations, we have

$$Z(\vec{\theta}) = \prod_{i=1}^n \prod_{j=1}^m [1 + e^{-(\lambda_i + \eta_j)}]. \quad (4.41)$$

Let $\{\vec{x}, \vec{y}\}$ be hidden variables (where $x_i = e^{-\lambda_i}$ for $\forall i$ from 1 to n , $y_j = e^{-\eta_j}$ for $\forall j$ from 1 to m , $\vec{x} = \{x_i\}_1^n$, and $\vec{y} = \{y_j\}_1^m$), and $P_{\text{BiBCM}}(A|\vec{\theta})$ be the probability mass function of a binary bipartite network A given the Lagrange multipliers $\vec{\theta}$. From Eqs. (4.36), (4.39), and (4.41), the function $P_{\text{BiBCM}}(A|\vec{\theta})$ can be expressed as

$$P_{\text{BiBCM}}(A|\vec{\theta}) = \prod_{i=1}^n \prod_{j=1}^m [(x_i y_j)^{A_{ij}} (1 + x_i y_j)^{-1}]. \quad (4.42)$$

Equivalently,

$$P_{\text{BiBCM}}(A|\{\vec{x}, \vec{y}\}) = \prod_{i=1}^n x_i^{\text{kbs}_i(A)} \prod_{j=1}^m y_j^{\text{ksb}_j(A)} \prod_{i,j} (1 + x_i y_j)^{-1}. \quad (4.43)$$

Clearly, $P_{\text{BiBCM}}(A|\{\vec{x}, \vec{y}\})$ depends on $(m+n)$ unknown parameters $\{x_i\}_1^n, \{y_j\}_1^m$. Now we define the log-likelihood function $L_{\text{BiBCM}}(\vec{x}, \vec{y})$ as

$$L_{\text{BiBCM}}(\vec{x}, \vec{y}) = \ln(P_{\text{BiBCM}}(A|\vec{\theta})) = \sum_{i=1}^n \text{kbs}_i(A) \ln(x_i) + \sum_{j=1}^m \text{ksb}_j(A) \ln(y_j) - \sum_{i=1}^n \sum_{j=1}^m \ln(1 + x_i y_j). \quad (4.44)$$

To maximize $L_{\text{BiBCM}}(\vec{x}, \vec{y})$ under the conditions that $\text{kbs}_i = \text{kbs}_i^*, \text{ksb}_j = \text{ksb}_j^*$, we need

to find $\{x_i^*\}_1^n, \{y_j^*\}_1^m \in [0, \infty)$ such that

$$\begin{cases} \frac{\partial L}{\partial x_i} = 0, \forall i = 1, 2, \dots, n, \\ \frac{\partial L}{\partial y_j} = 0, \forall j = 1, 2, \dots, m. \end{cases} \quad (4.45)$$

Equivalently, one has to solve the system of $(m+n)$ non-linear equations

$$\begin{cases} \sum_{j=1}^m \frac{x_i^* y_j^*}{1+x_i^* y_j^*} = kbs_i^*, \forall i = 1, 2, \dots, n, \\ \sum_{i=1}^n \frac{x_i^* y_j^*}{1+x_i^* y_j^*} = ksb_j^*, \forall j = 1, 2, \dots, m. \end{cases} \quad (4.46)$$

to obtain $(m+n)$ latent variables $\{x_i^*\}_1^n, \{y_j^*\}_1^m$.

Bipartite Weighted Configuration Model (BiWCM)

In this model, the strength sequence of the observed network is preserved on the average over the ensemble of randomized networks. The procedure used to define the probability mass function of a weighted bipartite network under the BiWCM is similar to the BiBCM in the binary version. For the sake of conciseness, in the following we only summarize the main relevant results of this model. We refer the readers to Di Gangi et al. (2015) for more detailed derivations of this model.

Let $P_{\text{BiWCM}}(W|\vec{\theta})$ be the probability mass function of a weighted bipartite network W , given the Lagrange multipliers $\vec{\theta}$ associated with the strength sequence. According to Di Gangi et al. (2015), $P_{\text{BiWCM}}(W|\vec{\theta})$ is given by

$$P_{\text{BiWCM}}(W|\vec{\theta}) = \prod_{i=1}^n \prod_{j=1}^m [(x_i y_j)^{w_{ij}} (1 - x_i y_j)]. \quad (4.47)$$

Equivalently,

$$P_{\text{BiWCM}}(W|\{\vec{x}, \vec{y}\}) = \prod_{i=1}^n x_i^{\text{sbs}_i(W)} \prod_{j=1}^m y_j^{\text{ssb}_j(W)} \prod_{i,j} (1 - x_i y_j). \quad (4.48)$$

The log-likelihood function $L_{\text{BiWCM}}(\vec{x}, \vec{y})$ is defined as

$$L_{\text{BiWCM}}(\vec{x}, \vec{y}) = \ln(P_{\text{BiWCM}}(W|\vec{\theta})) = \sum_{i=1}^n \text{sbs}_i(W) \ln(x_i) + \sum_{j=1}^m \text{ssb}_j(W) \ln(y_j) + \sum_{i=1}^n \sum_{j=1}^m \ln(1 - x_i y_j). \quad (4.49)$$

To maximize the log-likelihood function $L_{\text{BiWCM}}(\vec{x}, \vec{y})$ under the conditions that $\{\text{sbs}_i\}_1^n =$

$\{\text{sbs}_i^*\}_1^n, \{\text{ssb}_j\}_1^m = \{\text{ssb}_j^*\}_1^m$, we need to find $\{x_i^*\}_1^n, \{y_j^*\}_1^m \in [0, 1)$ such that

$$\begin{cases} \sum_{j=1}^m \frac{x_i^* y_j^*}{1 - x_i^* y_j^*} = \text{sbs}_i^*, \forall i = 1, 2, \dots, n, \\ \sum_{i=1}^n \frac{x_i^* y_j^*}{1 - x_i^* y_j^*} = \text{ssb}_j^*, \forall j = 1, 2, \dots, m. \end{cases} \quad (4.50)$$

Bipartite Enhanced Configuration Model (BiECM)

In contrast to the two previous configuration models, in the Bipartite Enhanced Configuration Model (BiECM), both the observed degree sequence as well as the observed strength sequence are maintained on the average over the ensemble of randomized networks. Let us summarize the main results of the BiECM. For further details, we refer the readers to the study of Di Gangi et al. (2015).

Suppose that four sequences $\{x_i\}_{i=1}^{i=n}$, $\{y_j\}_{j=1}^{j=m}$, $\{z_i\}_{i=1}^{i=n}$, and $\{t_j\}_{j=1}^{j=m}$ respectively represent hidden variables associated with the degree and the strength sequences $\{\text{kbs}_i\}_1^n$, $\{\text{kbs}_j\}_1^m$, $\{\text{sbs}_i\}_1^n$, $\{\text{ssb}_j\}_1^m$. Given the Lagrange multipliers $\vec{\theta}$ associated with these constraints, $P_{\text{BiECM}}(W|\vec{\theta})$ is defined as the probability mass function of a weighted bipartite network W . According to Di Gangi et al. (2015), the function $P_{\text{BiECM}}(W|\vec{\theta})$ is given by

$$P_{\text{BiECM}}(W|\vec{\theta}) = \prod_{i=1}^n \prod_{j=1}^m \frac{(1 - z_i t_j)(z_i t_j)^{w_{ij}} (x_i y_j)^{h(w_{ij})}}{1 - z_i t_j (1 - x_i y_j)}, \quad (4.51)$$

where $h(x)$ is a function defined as $h(x) = 1$ if $x > 0$ and $h(x) = 0$ otherwise.

Equivalently,

$$P_{\text{BiECM}}(W|\vec{x}, \vec{y}, \vec{z}, \vec{t}) = \prod_{i=1}^n x_i^{\text{kbs}_i(A)} \prod_{j=1}^m y_j^{\text{kbs}_j(A)} \prod_{i=1}^n z_i^{\text{sbs}_i(W)} \prod_{j=1}^m t_j^{\text{ssb}_j(W)} \prod_{i=1}^n \prod_{j=1}^m \frac{(1 - z_i t_j)}{1 - z_i t_j (1 - x_i y_j)}, \quad (4.52)$$

where in this case A is the adjacency matrix obtained from the weighted matrix W .

The log-likelihood function $L_{\text{BiECM}}(\vec{x}, \vec{y}, \vec{z}, \vec{t})$ is then defined as

$$\begin{aligned} L_{\text{BiECM}}(\vec{x}, \vec{y}, \vec{z}, \vec{t}) = \ln(P_{\text{BiECM}}(W|\vec{\theta})) &= \sum_{i=1}^n \text{kbs}_i(A) \ln(x_i) + \sum_{j=1}^m \text{kbs}_j(A) \ln(y_j) + \sum_{i=1}^n \text{sbs}_i(W) \ln(z_i) \\ &+ \sum_{j=1}^m \text{ssb}_j(W) \ln(t_j) + \sum_{i=1}^n \sum_{j=1}^m [\ln(1 - z_i t_j) - \ln(1 - z_i t_j (1 - x_i y_j))]. \end{aligned} \quad (4.53)$$

To maximize the log-likelihood function $L_{\text{BiECM}}(\vec{x}, \vec{y}, \vec{z}, \vec{t})$ under the conditions that $\text{kbs}_i =$

$kbs_i^*, ksb_j = ksb_j^*, sbs_i = sbs_i^*$, and $ssb_j = ssb_j^*$, we need to find $\{x_i^*\}_1^n, \{y_j^*\}_1^m \in [0, +\infty)$ and $\{z_i^*\}_1^n, \{t_j^*\}_1^m \in [0, 1)$ ⁷ such that

$$\begin{cases} \sum_{j=1}^m \frac{x_i^* y_j^* z_i^* t_j^*}{(1-z_i^* t_j^*)[1-z_i^* t_j^*(1-x_i^* y_j^*)]} = sbs_i^*, \forall i = 1, 2, \dots, n, \\ \sum_{i=1}^n \frac{x_i^* y_j^* z_i^* t_j^*}{(1-z_i^* t_j^*)[1-z_i^* t_j^*(1-x_i^* y_j^*)]} = ssb_j^*, \forall j = 1, 2, \dots, m, \\ \sum_{j=1}^m \frac{x_i^* y_j^* z_i^* t_j^*}{1-z_i^* t_j^*(1-x_i^* y_j^*)} = kbs_i^*, \forall i = 1, 2, \dots, n, \\ \sum_{i=1}^n \frac{x_i^* y_j^* z_i^* t_j^*}{1-z_i^* t_j^*(1-x_i^* y_j^*)} = ksb_j^*, \forall j = 1, 2, \dots, m. \end{cases} \quad (4.54)$$

Hidden variables and the projection matrices of a bipartite network

In general, the expectation of one-mode projection matrices can be calculated via hidden variables extracted from the referenced configuration model. More specifically,

$$\langle A_{i,k}^{B-B} \rangle = \sum_{j=1}^m f(i, j) f(k, j), (1 \leq i, k \leq n, i \neq k), \quad (4.55)$$

and

$$\langle A_{j,k}^{S-S} \rangle = \sum_{i=1}^n f(i, j) f(i, k), (1 \leq j, k \leq m, j \neq k), \quad (4.56)$$

where $\langle A_{i,k}^{B-B} \rangle$ ($\langle A_{j,k}^{S-S} \rangle$) is the notation for the expectation of $A_{i,k}^{B-B}$ ($A_{j,k}^{S-S}$)⁸, and for every pair of nodes i and j , $f(i, j)$ is the probability function that the node i in the mode B links to the node j in the mode S. Obviously, the function f depends on hidden variables extracted from the given constrains preserved in each configuration model.

In particular, under the BiBCM we have

$$f(i, j) = \frac{x_i^* y_j^*}{1 + x_i^* y_j^*}, \forall i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, m, \quad (4.57)$$

where $\{x_i^*\}_1^n, \{y_j^*\}_1^m$ are the solutions of Eqs. (4.46).

Under the BiWCM we have

$$f(i, j) = x_i^* y_j^*, \forall i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, m, \quad (4.58)$$

where $\{x_i^*\}_1^n, \{y_j^*\}_1^m$ are the solutions of Eqs. (4.50).

⁷We can relax this condition as $\{z_i^*\}_1^n, \{t_j^*\}_1^m \in [0, +\infty)$, and $z_i^* t_j^* < 1 \forall i = 1, 2, \dots, n$ and $\forall j = 1, 2, \dots, m$.

⁸Throughout this paper, $\langle X \rangle$ is the notation for expectation of X under the referenced null model.

In the case of BiECM, $f(i, j)$ is given by

$$f(i, j) = \frac{x_i^* y_j^* z_i^* t_j^*}{1 - z_i^* t_j^* (1 - x_i^* y_j^*)}, \forall i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, m, \quad (4.59)$$

where $\{x_i^*\}_{i=1}^{i=n}$, $\{y_j^*\}_{j=1}^{j=m}$, $\{z_i^*\}_{i=1}^{i=n}$, and $\{t_j^*\}_{j=1}^{j=m}$ are the solutions of Eqs. (4.54).

Hidden variables and expected network properties

Note that, the expected value of each network property can numerically be calculated over the ensemble of randomized networks as

$$\langle Y \rangle = \sum_{G \in \mathcal{G}} P(G) Y(G), \quad (4.60)$$

where $P(G)$ is probability of G in \mathcal{G} defined by the referenced configuration model and $Y(G)$ is the value of the considered network property, namely Y , of G .

In certain cases, $\langle Y \rangle$ can analytically be calculated/approximated under a simple form. For instance, according to Saracco et al. (2015) and Saracco et al. (2016), under the BiBCM, the expected numbers of V , Λ , M , and W motifs can be evaluated by the expected adjacency matrix $A^{\text{BiBCM}} = \{A_{ij}^{\text{BiBCM}}\}_{n \times m} = \left\{ \frac{x_i^* y_j^*}{x_i^* y_j^* + 1} \right\}_{n \times m}$, where $\{x_i^*\}_1^n, \{y_j^*\}_1^m$ are the solution of Eqs. (4.46) ⁹.

4.3 Findings

We now investigate the topological and structural properties of the bank-sector credit network, compared with the expected values of such properties evaluated under the various configuration models. Three bipartite configuration models are employed here, i.e. the BiBCM (in the binary version), and the BiWCM as well as BiECM (in the weighted version). In each model, we will generate an ensemble consisting of 1000 randomized networks ¹⁰.

4.3.1 Binary bank-sector network

Projection matrices and overlaps

Under the BiBCM, if $\{x_i^*\}_1^n$ and $\{y_j^*\}_1^m$ are hidden variables extracted from the observed degree sequence, the expectation of the elements of the projection matrix between nodes in

⁹We refer the readers to Saracco et al. (2015), Saracco et al. (2016) and for further details.

¹⁰In fact, there are little differences when we increase the size of the ensemble from 500 to 1000 or higher.

mode B is

$$\langle A_{ik}^{B-B} \rangle = \sum_{j=1}^m \frac{x_i^* y_j^*}{1 + x_i^* y_j^*} \frac{x_k^* y_j^*}{1 + x_k^* y_j^*}, \forall i, k = 1, 2, \dots, n, \quad (4.61)$$

and the expectation of the elements of the projection matrix between nodes in mode S is

$$\langle A_{jk}^{S-S} \rangle = \sum_{i=1}^n \frac{x_i^* y_j^*}{1 + x_i^* y_j^*} \frac{x_i^* y_k^*}{1 + x_i^* y_k^*}, \forall j, k = 1, 2, \dots, m. \quad (4.62)$$

Figure (4.1) shows the comparison between the two projected matrices of the observed network and the expected ones obtained from the BiBCM (element by element) for the first and the last year of the sample. Results for other years are qualitatively very similar. For every pair of nodes (i, k) in the bank-bank projection matrix, we compare $\langle A_{ik}^{B-B} \rangle$ with A_{ik}^{B-B} , and for every pair of nodes (j, k) in the sector-sector projection matrix, we compare $\langle A_{jk}^{S-S} \rangle$ with A_{jk}^{S-S} . As we can see, the overlaps in lending and borrowing portfolios between banks and between sectors, especially for the pairs of highly overlapping portfolios, can be replicated very well by hidden variables via equations (4.61) and (4.62). To assess the statistical significance of the portfolio overlaps between banks and between sectors, we employ the z-score analysis for the elements of the projection matrices. More specifically, we define $z_{ik}^{B-B} = \frac{A_{ik}^{B-B} - \langle A_{ik}^{B-B} \rangle}{\sigma[A_{ik}^{B-B}]}$ and $z_{jk}^{S-S} = \frac{A_{jk}^{S-S} - \langle A_{jk}^{S-S} \rangle}{\sigma[A_{jk}^{S-S}]}$ ¹¹. In Figure (4.2), we plot z_{ik}^{B-B} against A_{ik}^{B-B} (for every pair of nodes (i, k) in the bank-bank projection matrix) and z_{jk}^{S-S} against A_{jk}^{S-S} (for every pair of nodes (j, k) in the sector-sector projection matrix). It shows that, almost all the z-scores are less than or equal to $z^* = 2$, in terms of absolute values. Especially, for the large values of A_{ik}^{B-B} and A_{jk}^{S-S} , the values of the corresponding z-scores completely lie within the range $[-2, 2]$, indicating that these overlaps are simply explained by information embedded in the degree sequence of the original bipartite network. Obviously, the few z-scores lying outside that range might imply the presence of non-random patterns in the overlaps between several pairs of banks and between several pairs of sectors.

¹¹Throughout this paper, $\sigma[X]$ is the notation for standard deviation of X under the referenced null model.

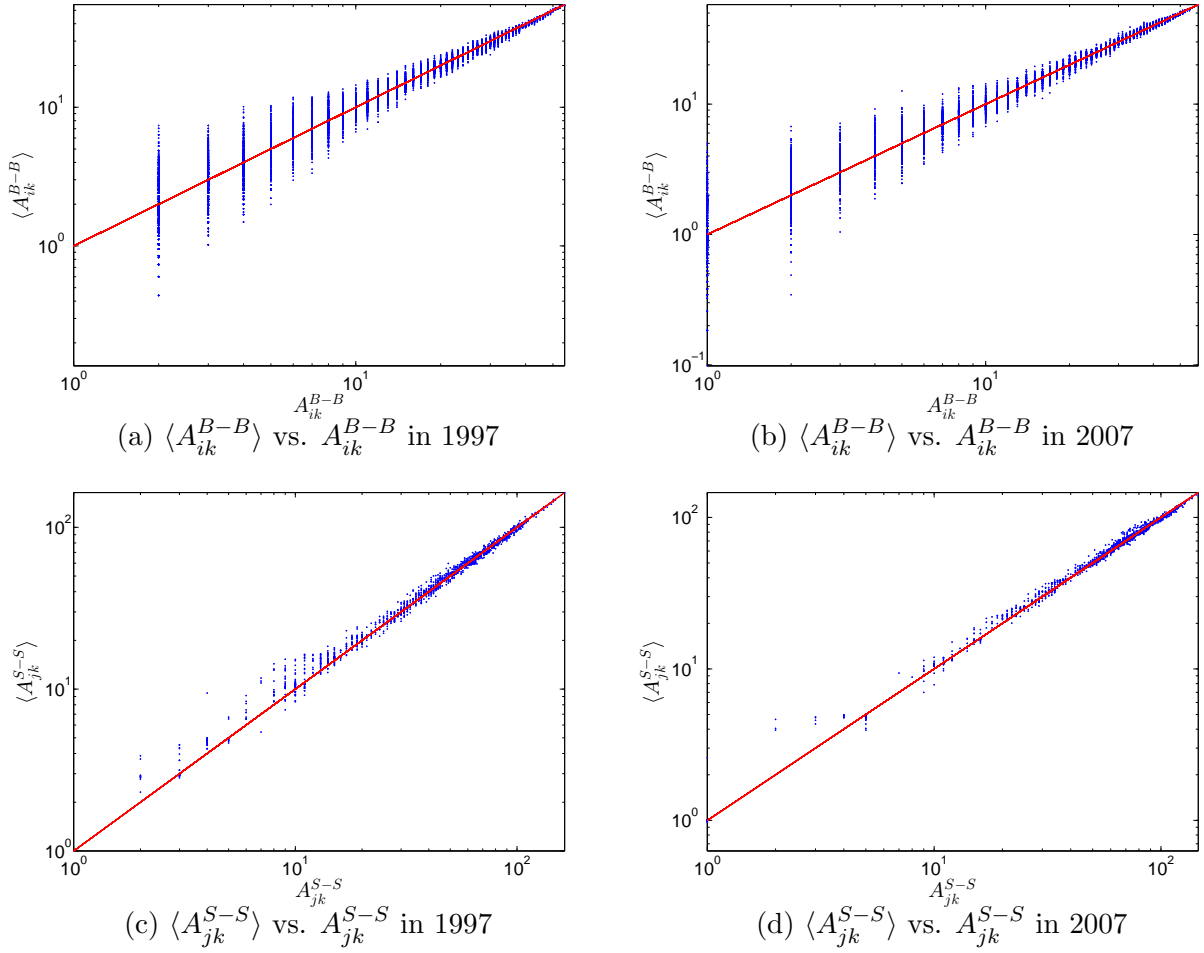


Figure 4.1: Scatter plots of portfolio overlaps in the observed network and under the BiBCM. Panels (a), (b) show the off-diagonal elements of the projection matrices between banks in the observed network and in the BiBCM, in 1997 and 2007. Panels (c), (d) show the off-diagonal elements of the projection matrices between sectors in the observed network and in the BiBCM, in 1997 and 2007. In each panel, we compare each element of the observed matrix with its expectation under the BiBCM. The identity line is shown in red, expected points are in blue.

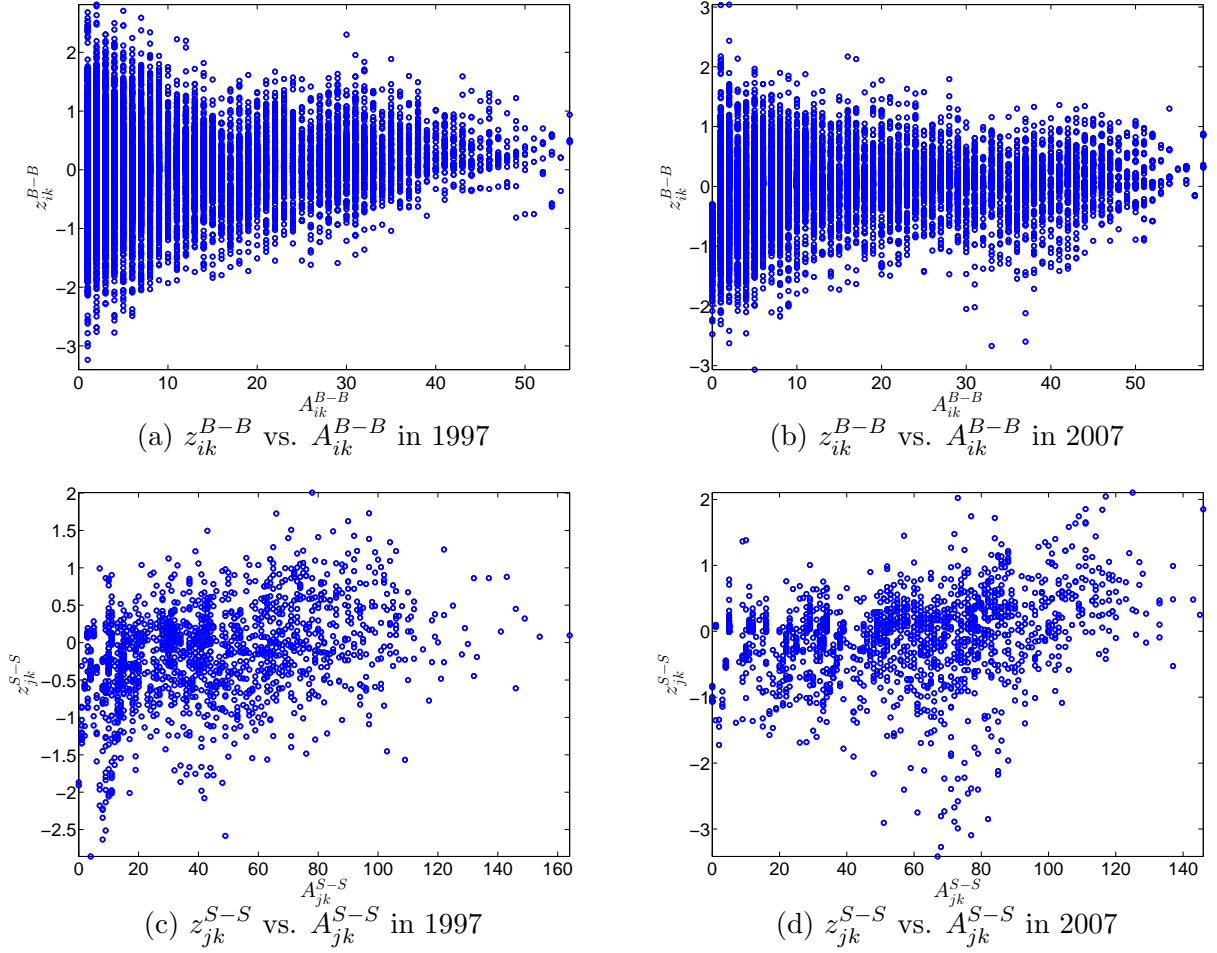


Figure 4.2: Scatter plots of z-scores of the off-diagonal elements of the projection matrices A^{B-B} and A^{S-S} , evaluated under the BiBCM, in 1997 and 2007. Panel (a) z_{ik}^{B-B} vs. A_{ik}^{B-B} in 1997, panel (b) z_{ik}^{B-B} vs. A_{ik}^{B-B} in 2007. Panel (c) z_{jk}^{S-S} vs. A_{jk}^{S-S} in 1997, panel (d) z_{jk}^{S-S} vs. A_{jk}^{S-S} in 2007.

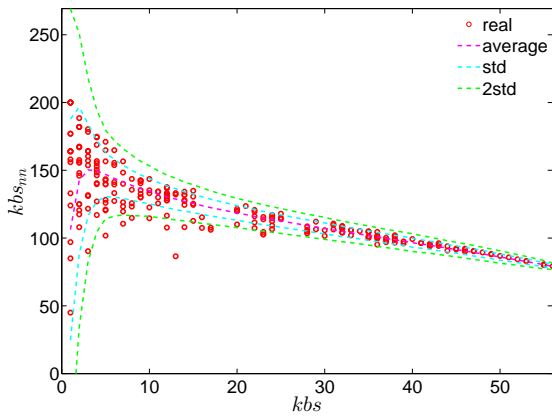
Assortativity

We now turn to the observed and expected values of the average degree of nearest neighbors (ANND) and the overall Pearson correlation coefficient r in Figure (4.3) and Figure (4.4). First, we find that the observed network is disassortative. More specifically, the ANND is a decreasing function of the degree, and the overall Pearson correlation coefficient r is always less than zero. This indicates that, overall, the high degree banks tend to lend to the low degree sectors, while the low degree banks tend to lend to the high degree sectors. Furthermore, such a disassortative mixing nature between nodes is also mostly persistent over years. Latter, in the analysis of the nestedness, we will see how this structure can be used together with the disassortative nature to enrich the qualitative characterization of

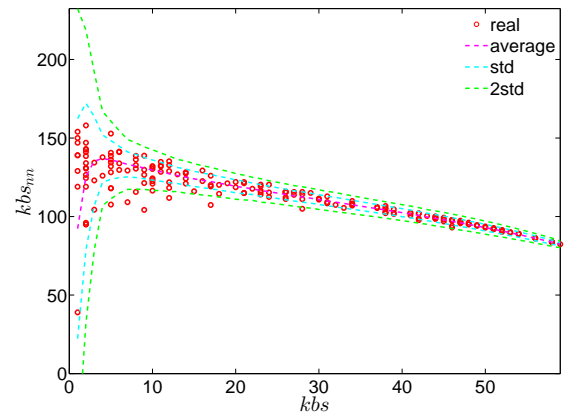
connections among nodes in the network.

Second, in comparison to the BiBCM benchmark, we observe mixed results. More specifically, as shown in Figure (4.3), almost all observed values of ANND are located inside ± 2 standard deviations (std.) from the expected ones. This suggests that, for most of the banks and sectors, the intrinsic heterogeneity in the observed degree sequence is sufficient to explain the average degree of nearest neighbors. Additionally, Figure (4.5) shows that, on the one hand, imposing the degree constraints can reproduce the overall trend in the evolution of the average of ANNDs over the sample period. On the other hand, there are actually some slight deviations of the observed average of ANNDs from ± 2 std. band in several years (see panel (b) of Figure (4.5)). In addition, we find that, over the years, the observed value of r persistently deviates by more than ± 2 std. from the expected one, although we should emphasize that the main trends of the observed and the expected values of r are still in close agreement (see Figures (4.4) and (4.16)).

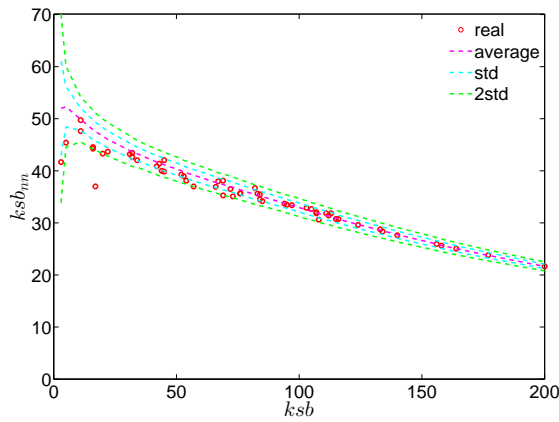
In summary, the above results indicate that, although the observed degree sequence plays a crucial role in replicating the disassortative nature of the bank-sector credit network of Spain, we find that it is not completely explained by the BiBCM, and in fact, the Pearson correlation coefficient is consistently slightly (but significantly so) higher (less negative) than expected on the base of the null model.



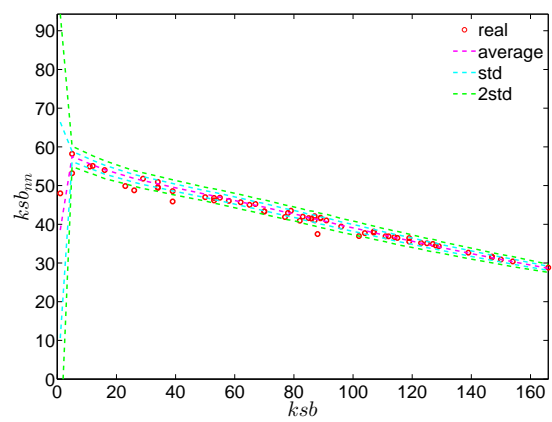
(a) kbs_{nn} in 1997



(b) kbs_{nn} in 2007



(c) ksb_{nn} in 1997



(d) ksb_{nn} in 2007

Figure 4.3: ANNDs in the observed network and under the BiBCM. Panels (a), (b) report kbs_{nn} as a function of kbs , respectively in 1997 and 2007. Panels (c), (d) show ksb_{nn} as a function of ksb , respectively in 1997 and 2007. Observed points are red circles. Ensemble averages are represented by dashed curves in magenta. The cyan and green dashed curves respectively indicate the ± 1 and ± 2 std. from the averages.

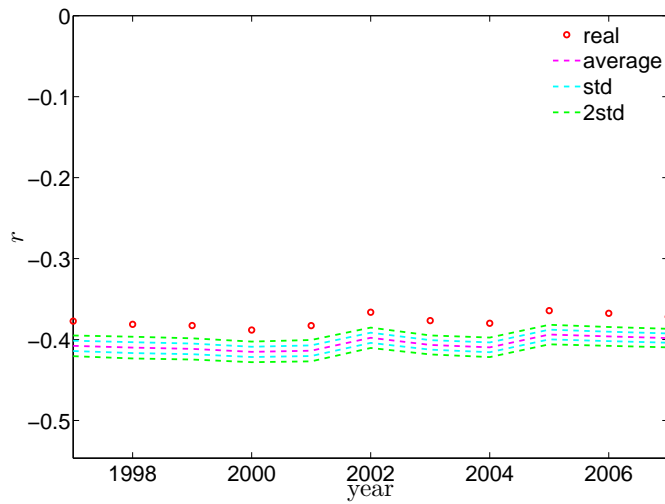
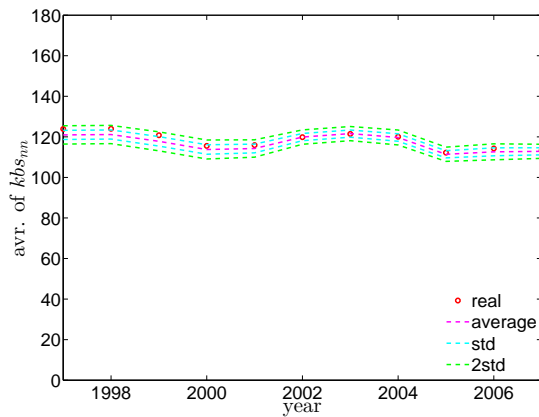
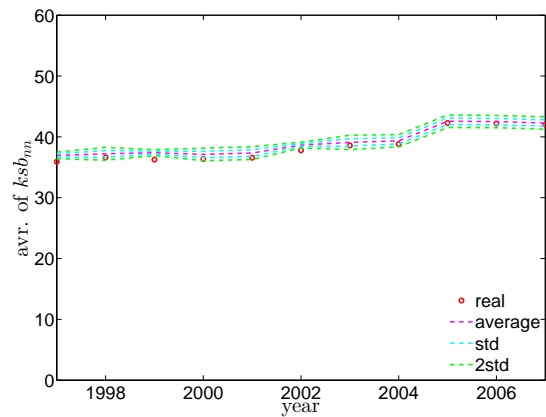


Figure 4.4: Evolution of the Pearson correlation coefficient of degrees in the observed network and under the BiBCM, from 1997 to 2007. Observed points are red circles. Ensemble averages are represented by dashed curves in magenta. The cyan and green dashed curves respectively indicate the ± 1 and ± 2 std. from the averages.



(a) Evolution of avr. of kbs_{nn}



(b) Evolution of avr. of ksb_{nn}

Figure 4.5: The evolution of the average of ANNDs in the observed network and under the BiBCM. Panel (a) shows the average of kbs_{nn} , panel (b) shows the average of ksb_{nn} . Observed points are red circles. Ensemble averages are represented by dashed curves in magenta. The cyan and green dashed curves respectively indicate the ± 1 and ± 2 std. from the averages.

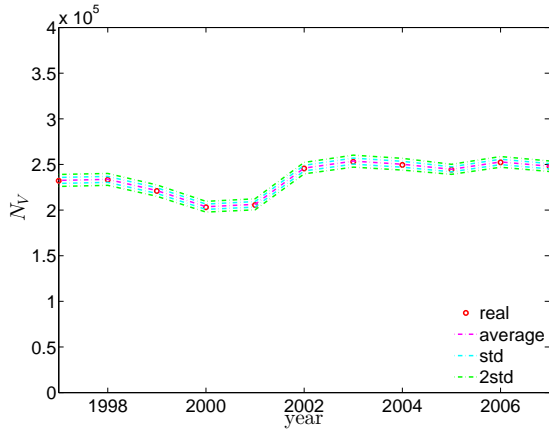
Motifs

We now analyze different types of motifs capturing various orders of structural correlations and overlaps between banks and the other sectors of the economy. Each panel of Figure (4.6) shows the evolution of the number of motifs in the observed network, the evolution of the average of the number of motifs evaluated under the BiBCM, and ± 1 and ± 2 standard deviations from that average.

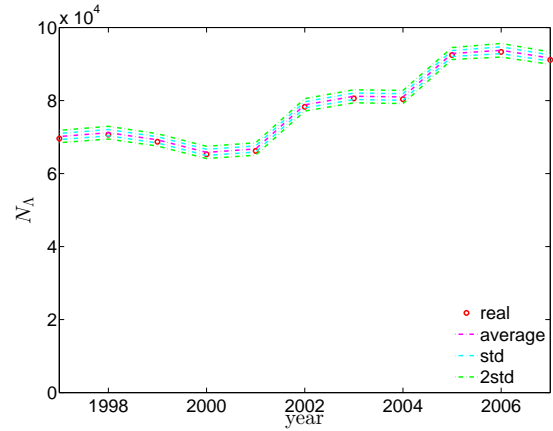
We find that, first, except for the case of V motifs, all other types of motifs including the Λ , X, M, and W motifs show a monotonic increase in their total numbers in the period from 2001 to 2006. This suggests a tendency of an increase in structural correlations and overlaps between banks and non-financial sectors over that period. Interestingly, we also observe that the numbers of the considered motifs all decline somewhat in 2007, which may signal a structural change in the network. Unfortunately, since we do not have information for the bank-sector credit relations after 2007, it is impossible to assess the effects of the global financial crisis on the network's structure.

Second, the evolution of N_V , N_Λ , N_X , N_M , and N_W can be explained very well by the BiBCM. Over the years, in each type of motifs, the observed number of motifs is completely located inside ± 2 std. from the expected value. Note that, we also show the ensemble distributions of N_V , N_Λ , N_X , N_M , and N_W in two years 1997 and 2007 in Figures (4.17), (4.18) in the Appendix.

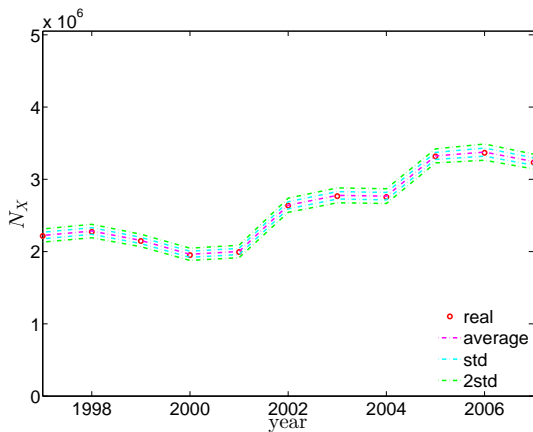
In order to have a more comprehensive analysis of the local correlations between nodes, one should consider the local motifs as well. Here we select the family of X motifs as an example to illustrate this analysis. Based on the number of X motifs surrounding each node in the network, we measure the local clustering coefficients. In Figure (4.7), we examine the local clustering coefficient as a function of node degree. Clearly, we can see that, in all modes of the observed network, the local binary clustering coefficient is a decreasing function of degree except for a few outliers on the left-hand side of the graphs. In addition, in comparison to the BiBCM, for most part, there is no significant difference between the observed values and the expected ones. This indicates that the local binary clustering can also mostly be explained by information encoded in the observed degree sequence. Few significant deviations obviously indicate the local non-random patterns in some structural correlations of the observed network.



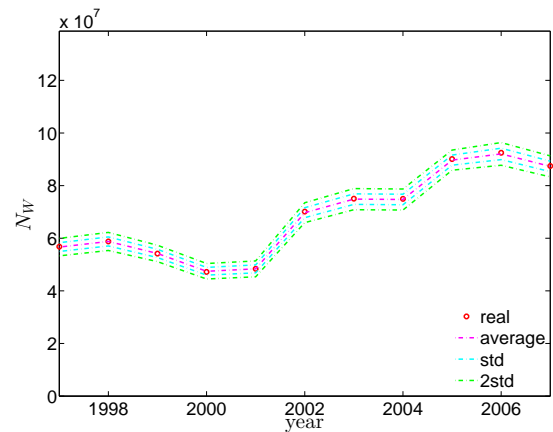
(a) Evolution of N_V



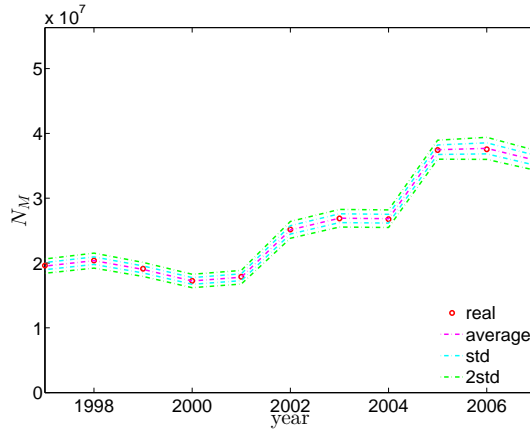
(b) Evolution of N_Λ



(c) Evolution of N_X



(d) Evolution of N_W



(e) Evolution of N_M

Figure 4.6: Evolution of the numbers of motifs over years in the observed network and under the BiBCM. Panel (a) represents N_V , panel (b) represents N_Λ , panel (c) represents N_X , panel (d) represents N_W , and panel (e) represents N_M . In each panel, observed points are red circles, ensemble averages are represented by dash-dotted curves in magenta, the cyan and green dash-dotted curves respectively indicate the ± 1 and ± 2 std. from the averages.

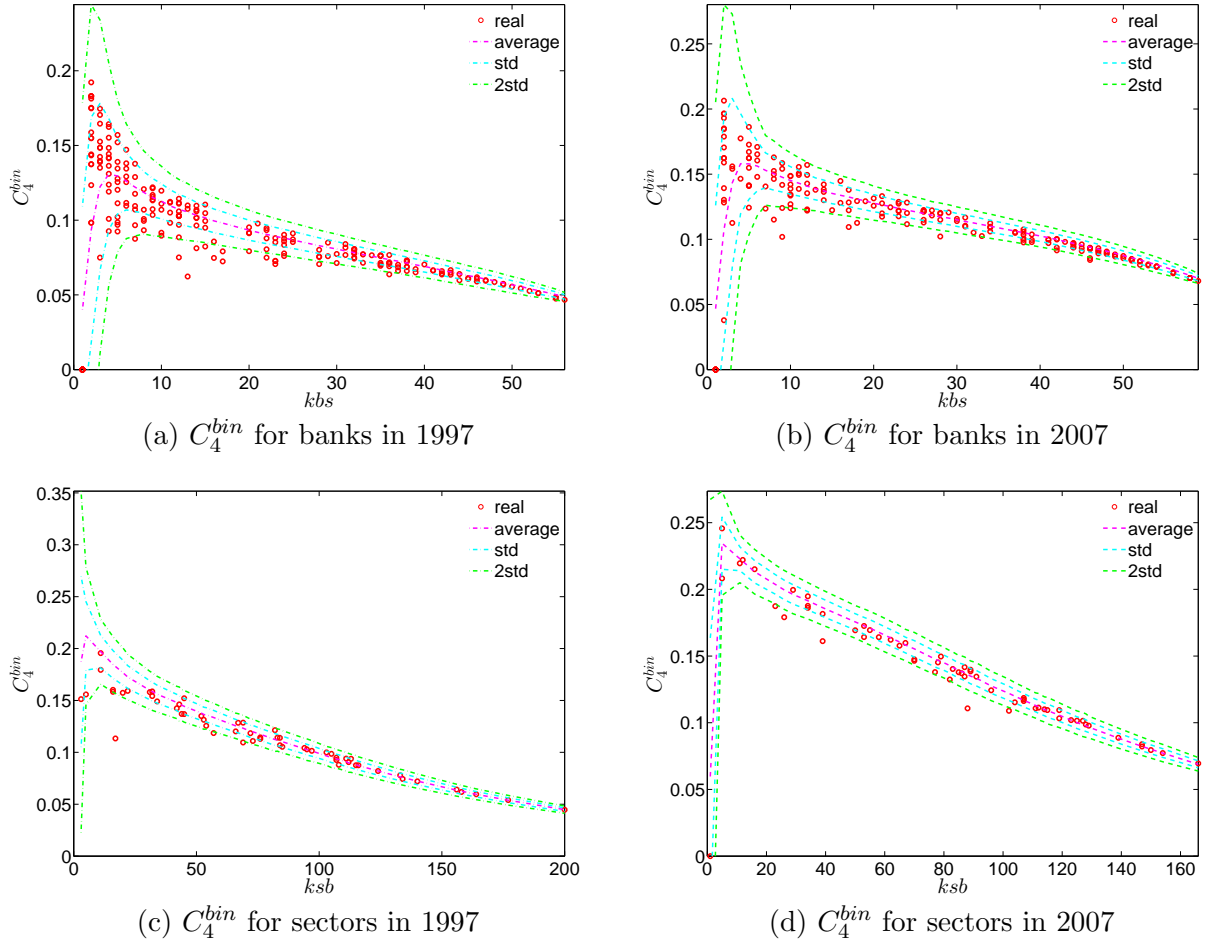


Figure 4.7: Local clustering coefficients (based on X motifs) in the observed network and under the BiBCM. Panels (a), (b) show the binary local clustering coefficient C_4^{bin} for banks as a function of kbs, respectively in 1997 and 2007. Panels (c), (d) show the binary local clustering coefficient C_4^{bin} for sectors as a function of ksb, respectively in 1997 and 2007. Observed points are red circles. Ensemble averages are represented by magenta dashed curves. The cyan and green dashed curves respectively indicate the ± 1 and ± 2 std. from the averages.

Nestedness

The “nestedness” structure can be graphically illustrated by the nested adjacency matrix, where the interactions among “generalists” (i.e. the high degree nodes or highly diversified nodes) constitute a dense “core” within the adjacency matrix. In addition, “specialists” (i.e. the low degree nodes or less diversified nodes) also tend to interact with “generalists”, rather than among themselves (e.g. Bascompte et al., 2003). That structure in the bipartite bank-sector credit network of Spain is illustrated by the nested matrix in Figure (4.8).

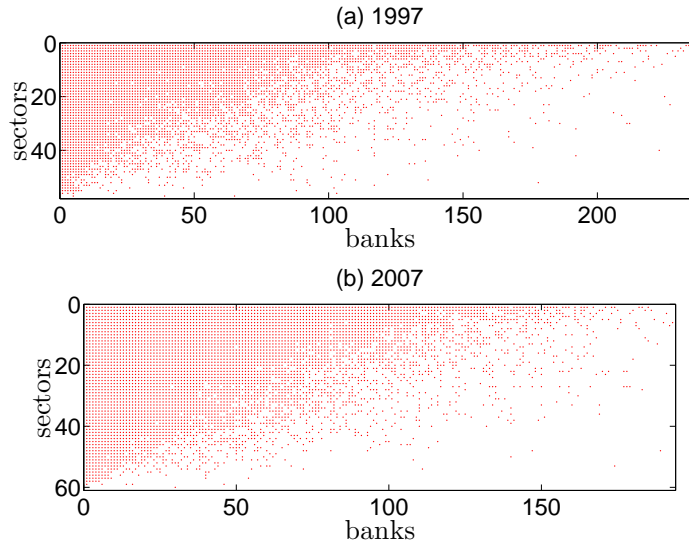
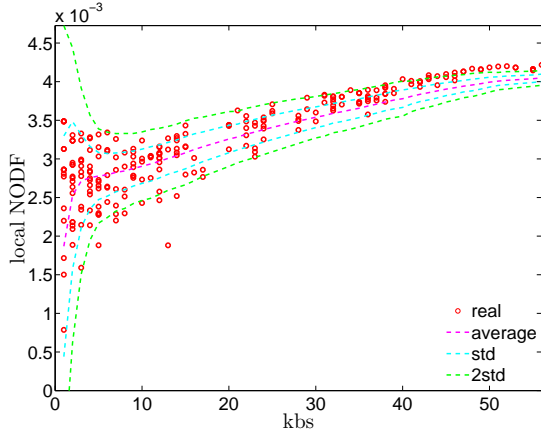


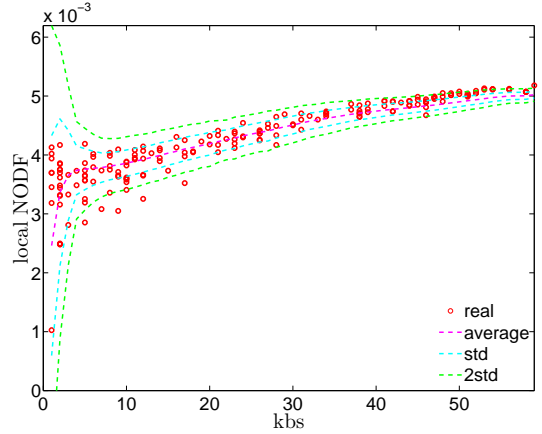
Figure 4.8: The adjacency matrix and nestedness structure of the observed bank-sector network in 1997 and 2007. Rows and columns are in descending order of the node degrees. Each red point indicates a credit link between a bank and a sector.

Clearly, the top left corner of each panel of Figure (4.8) reveals the presence of an approximately triangular shape of the distribution of 1's in the reordered adjacency matrix. Several 1's deviating from the benchmark of perfect nestedness signal that the nestedness structure is not perfect ¹². It should be emphasized that the analysis of nestedness can be used complementarily to the analysis of assortativity to shed light on systematic tendencies for nodes to form connections. In particular, such a structure indicates that the generalist sectors tend to borrow from almost all types of banks, and similarly the generalist banks also tend to lend to almost all sectors; in contrast, specialist banks concentrate their lending on the subset of generalist sectors, and similarly specialist sectors mainly borrow from generalist banks. This is somehow similar to the core-periphery structure found in many one-mode banking networks (e.g. van Lelyveld and in't Veld, 2014; Fricke and Lux, 2015). Generally speaking, a network with a core-periphery structure will consist of a group of densely connected nodes surrounded by a periphery of comparatively sparse connected nodes.

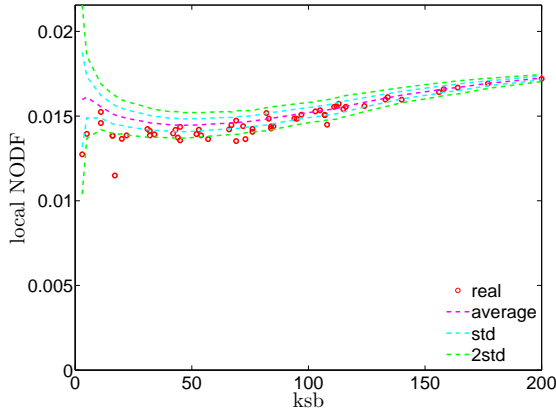
¹²Here 1 stands for elements $A_{ij} = 1$ of the adjacency matrix A .



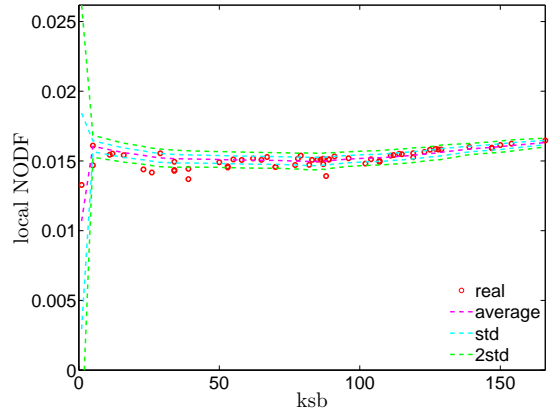
(a) local NODF for banks in 1997



(b) local NODF for banks in 2007



(c) local NODF for sectors in 1997



(d) local NODF for sectors in 2007

Figure 4.9: Local nestedness in the observed network and under the BiBCM. Panels (a), (b) report the local NODF as a function of the bank degree k_{bs} , respectively in 1997 and 2007. Panels (c), (d) show the local NODF as a function of the sector degree k_{sb} , respectively in 1997 and 2007. Observed points are represented by red circles. Ensemble averages are represented by dashed curves in magenta. The cyan and green dashed curves respectively indicate the ± 1 and ± 2 standard deviations from the averages.

We now measure the local and overall nestedness indicators (NODF) in the network. In general, we find a positive correlation between the local nestedness indicator and the node degree in both modes. More specifically, as shown in Figure (4.9), in both modes, overall, except for some outliers with a small overall degree, nodes with higher degrees also correspond to a higher local nestedness indicator. This implies that hubs in the network contribute more to the overall nestedness. In addition, most of the observed points are located inside ± 2 std. from their expected values, indicating the ability of the BiBCM of replicating the behavior of the local NODF of the observed network. This result is in line with what has been found in recent studies suggesting the important role of the heterogeneity in the distribution of the

degrees for the emergence of a nested structure in bipartite networks (e.g. Johnson et al. 2013; Feng and Takemoto, 2014). Again, several points deviating from the ± 2 std. band are observed, revealing the likely presence of non-random patterns in some parts of the observed network.

The temporal dynamics of the overall NODF for the whole network, the overall NODF for banks, and the overall NODF for sectors from 1997 to 2007 are shown in Figure (4.10). Over time, the overall nestedness for the whole network, for banks, and for sectors seem to undergo hardly any changes. Again, in all panels of Figure (4.10), the main trend is mostly captured by the BiBCM, but we do observe slight but statistically significant deviations from the expected values in a couple of years.

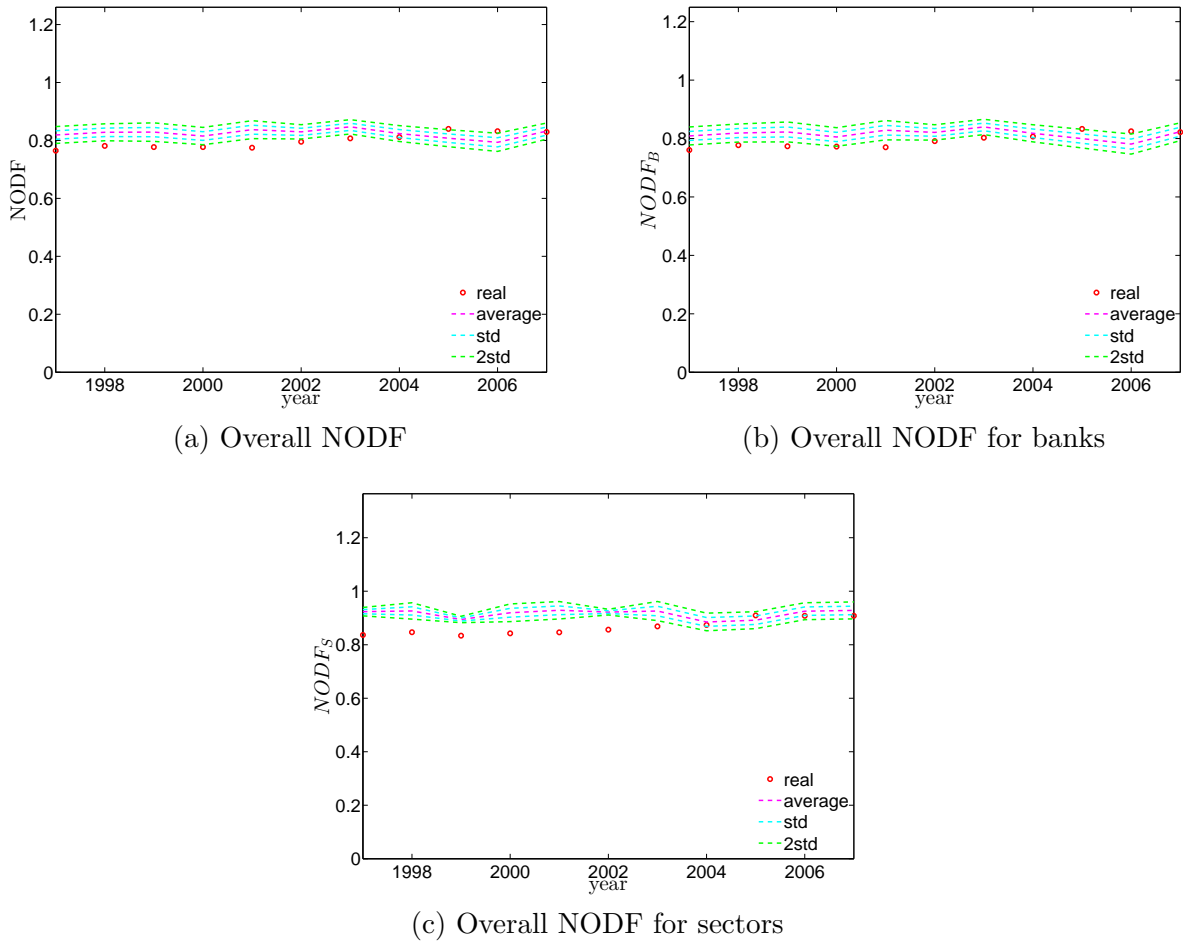


Figure 4.10: Evolution of the overall nestedness indicators over years in the observed network and under the BiBCM. Observed points are represented by red circles. Ensemble averages are represented by dashed curves in magenta. The cyan and green dashed curves respectively indicate the ± 1 and ± 2 standard deviations from the averages.

4.3.2 Weighted bank-sector network

In the weighted version of the bank-sector network, we will focus on the analysis of the weighted assortativity and local weighted clustering coefficients. For the sake of consistency, we again select the two years 1997 and 2007 as the examples for the presentation of the details of the analysis. In addition, regarding the weighted clustering coefficients, we only consider the local clustering coefficients based on X motifs consisting of two nodes in each mode. Two weighted configuration models are employed, i.e. the BiWCM maintaining the observed strength sequence, and the BiECM preserving both the observed degree sequence as well as the observed strength sequence.

Before moving on to the comparison between the observed properties with the expected ones obtained from the two configuration models, let us first briefly summarize the main results of the analysis of the weighted version of the observed network. The disassortative mixing nature is still found in the weighted version, as revealed by an overall negative correlation between the ANNS and the strength (e.g. see red points in Figure (4.11)). In line with such a negative correlation, the overall Pearson correlation coefficient of strengths is also less than zero ($r_w = -0.123$ in 1997, and $r_w = -0.095$ in 2007). However, since typically $|r_w| < |r|$, the weighted version is less disassortative than the binary one. In terms of the local clustering coefficients, unlike in the binary version, we observe that, in general, nodes with a higher degree are now characterized by a larger value of the local weighted clustering coefficient (see Figure (4.12)). This indicates the role of link weights in the structural correlations between banks and sectors.

In the following, we will compare the observed values of the considered properties with those obtained from the two configuration models.

Overall, we find that the knowledge of the strength sequence is not enough to reproduce the properties of the weighted version of the observed network. More specifically, Figures (4.11) and (4.12) show that, the main features of the average strength of the nearest neighbors and the local weighted clustering coefficients can not be replicated by latent information extracted from the observed strength sequence only. For most part, the observed values of these properties are significantly different from the expected ones obtained from the BiWCM. This observation is in line with previous studies of the weighted configuration model defined for one-mode networks. It has repeatedly been found that the knowledge of the observed strengths is not enough to explain the structural correlations in real world networks (e.g. Squartini et al., 2011b; Mastrandrea et al., 2014; Squartini et al., 2015; Luu et al., 2016).

Moving on to the BiECM, as can be seen from Figures (4.13) and (4.14) that, compared

to the BiWCM, we find that the main trend of the average strengths of nearest neighbors and the local weighted clustering coefficients now can be replicated better by hidden variables extracted from the BiECM, although there are still several significant deviations from the expected values, especially in the case of the local weighted clustering coefficients C_4^w . For the latter, the most connected banks and sectors both show levels of clustering beyond what is already implied by their degree and strength sequences.

For a more detailed comparison between the two models, we define the following z-scores of ANNSs and C_4^w evaluated under the BiWCM and the BiECM. Generally, for each value of a property X ($sbs_{nn}(i)$, $sb_{nn}(i)$, or $C_4^w(i)$ of any node i) of the network, we define

$$z[X]_{\text{null-model}} = \frac{X - \langle X \rangle_{\text{null-model}}}{\sigma[X]_{\text{null-model}}}, \quad (4.63)$$

where the null model here can be BiWCM or the BiECM. In summary, we typically find that in terms of absolute values, the z-scores evaluated under the BiECM are smaller than those evaluated under the BiWCM (see Figures (4.19) and (4.20) in the Appendix).

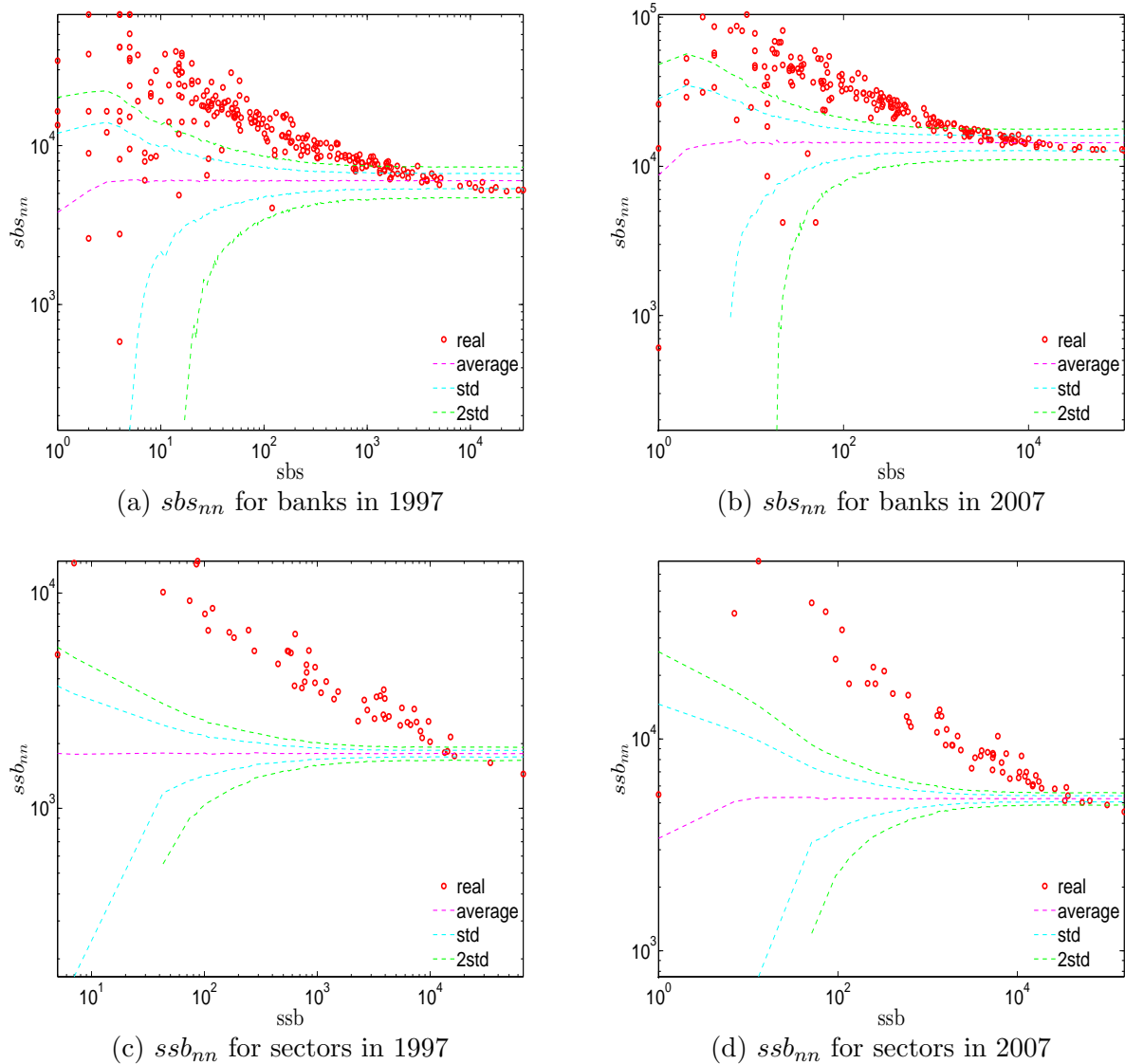
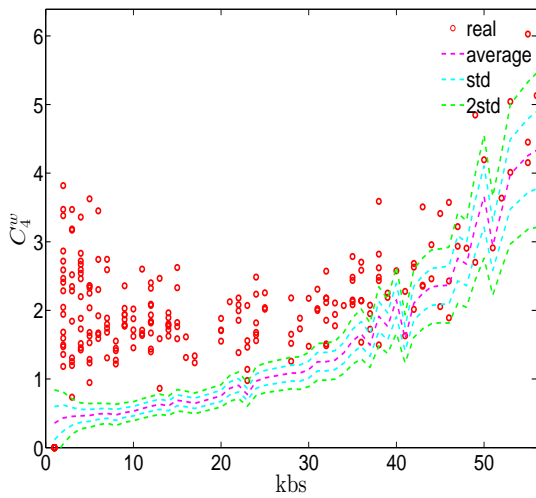
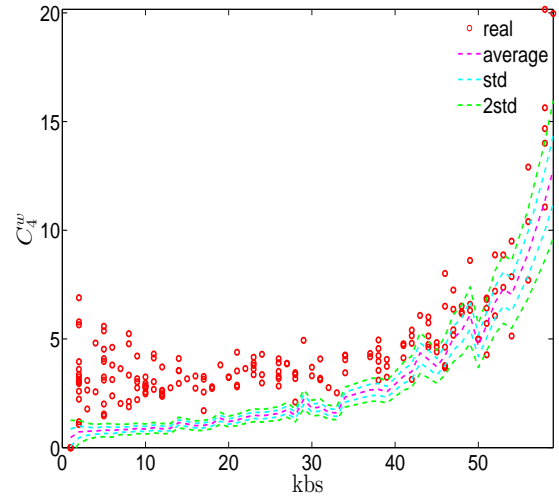


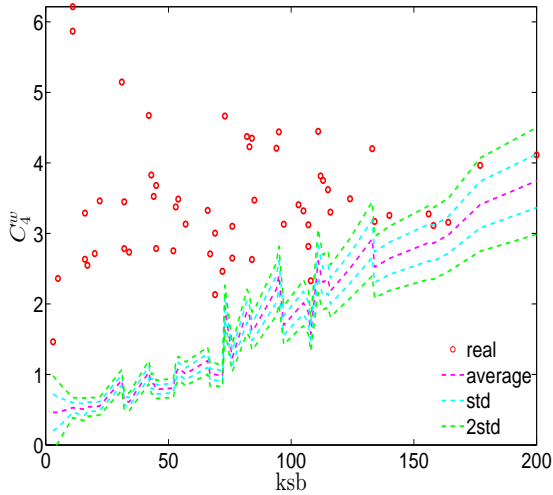
Figure 4.11: ANNSs in the observed network and under the BiWCM. Panels (a), (b) show sbs_{nn} as a function of sbs , in 1997 and 2007, respectively. Panels (c), (d) show sbs_{nn} as a function of ssb , in 1997 and 2007, respectively. Observed points are represented by red circles. Ensemble averages are represented by dashed curves in magenta. The cyan and green dashed curves respectively indicate the ± 1 and ± 2 standard deviations from the averages.



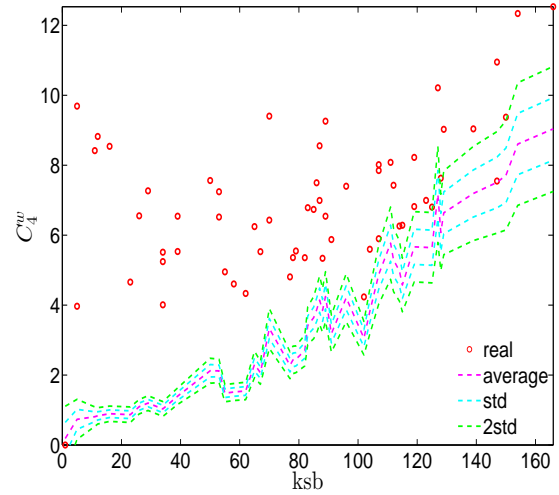
(a) C_4^w for banks in 1997



(b) C_4^w for banks in 2007

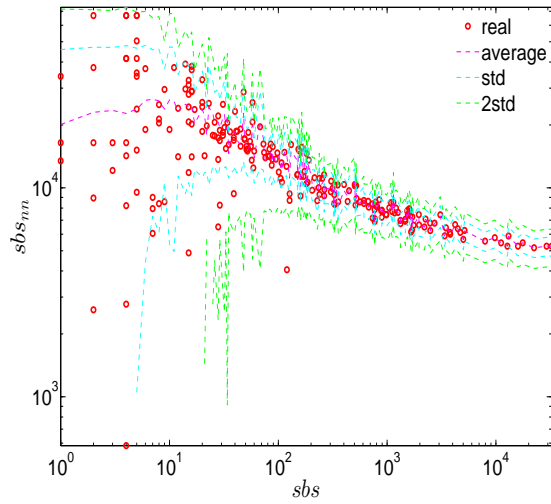


(c) C_4^w for sectors in 1997

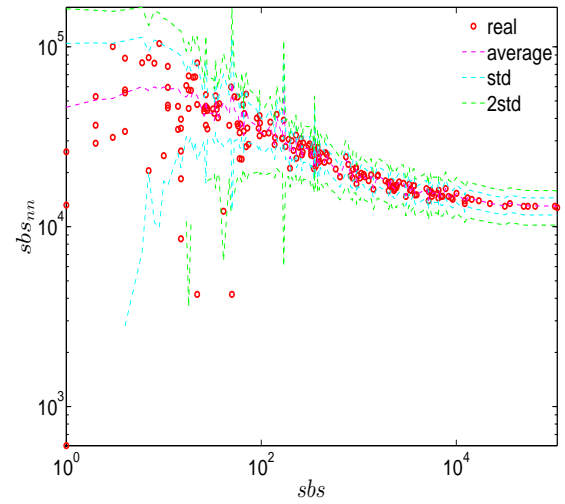


(d) C_4^w for sectors in 2007

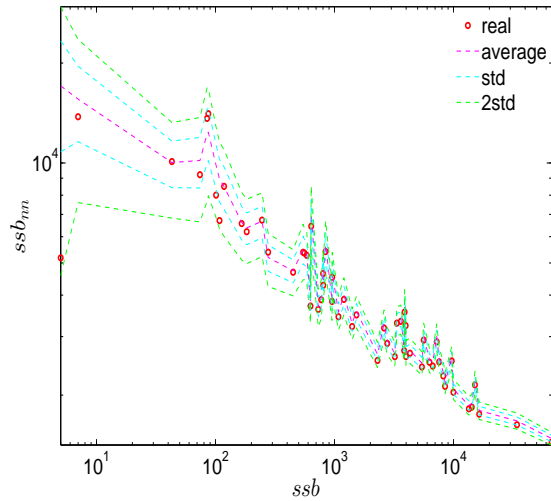
Figure 4.12: Local weighted clustering coefficients in the observed network and under the BiWCM. Panels (a), (b) show the weighted local clustering C_4^w for banks as a function of kbs, respectively in 1997 and 2007. Panels (c), (d) show the weighted local clustering C_4^w for sectors as a function of ksb, respectively in 1997 and 2007. Observed points are represented by red circles. Ensemble averages are represented by dashed curves in magenta. The cyan and green dashed curves respectively indicate the ± 1 and ± 2 std. from the averages.



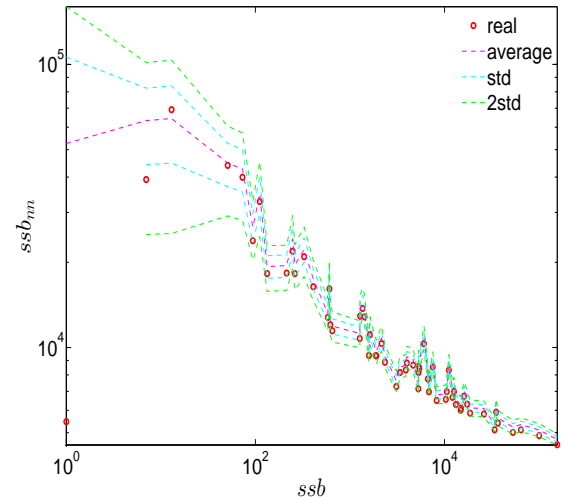
(a) sbs_{nn} for banks in 1997



(b) sbs_{nn} for banks in 2007

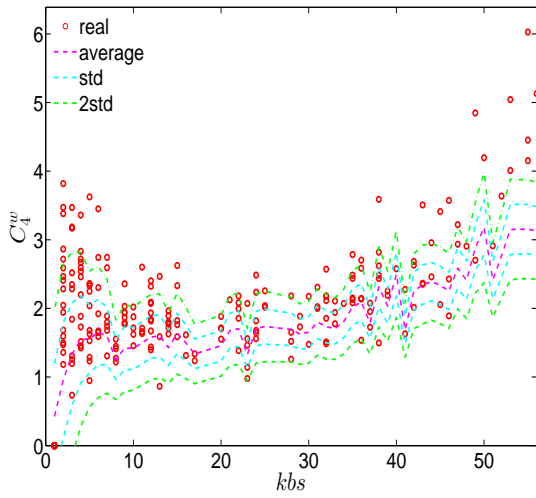


(c) sbs_{nn} for sectors in 1997

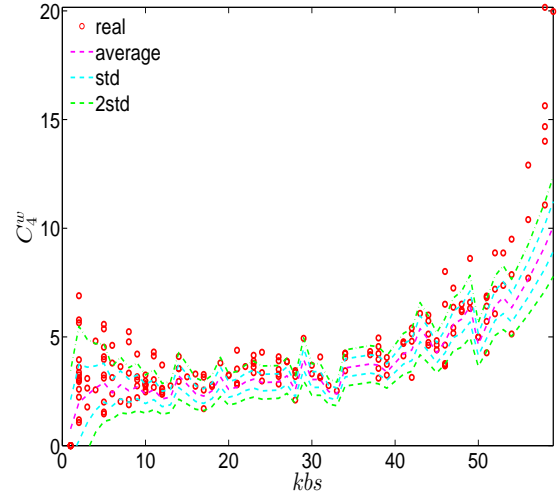


(d) sbs_{nn} for sectors in 2007

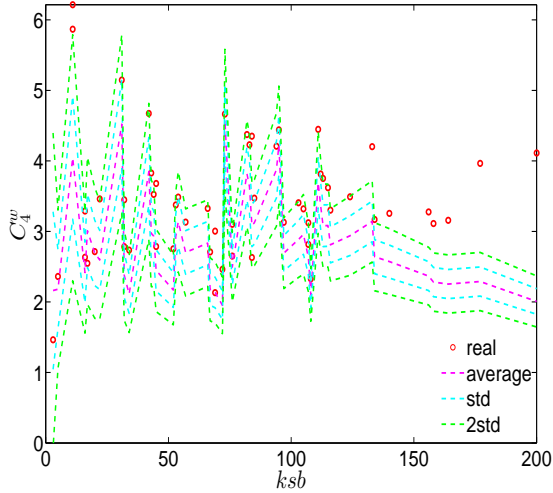
Figure 4.13: ANNSs in the observed network and under the BiECM. Panels (a), (b) show sbs_{nn} as a function of sbs , in 1997 and 2007 respectively. Panels (c), (d) show sbs_{nn} as a function of sbs , in 1997 and 2007 respectively. Observed points are represented by red circles. Ensemble averages are represented by dashed curves in magenta. The cyan and green dashed curves respectively indicate the ± 1 and ± 2 standard deviations from the averages.



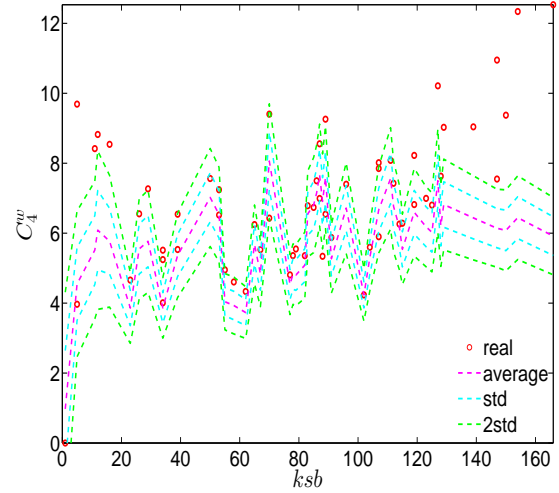
(a) C_4^w for banks in 1997



(b) C_4^w for banks in 2007



(c) C_4^w for sectors in 1997



(d) C_4^w for sectors in 2007

Figure 4.14: Local weighted clustering coefficients in the observed network and under the BiECM. Panels (a), (b) show the weighted local clustering C_4^w for banks as a function of kbs , respectively in 1997 and 2007. Panels (c), (d) show the weighted local clustering C_4^w for sectors as a function of ksb , respectively in 1997 and 2007. Observed points are represented by red circles. Ensemble averages are represented by dashed curves in magenta. The cyan and green dashed curves respectively indicate the ± 1 and ± 2 std. from the averages.

4.4 Conclusions

We have examined the topological and structural properties of the bank-sector credit market of Spain over the period from 1997 to 2007. We have also assessed whether such properties can be reproduced by information embedded in the observed degree sequence and/or the observed strength sequence by employing the various configuration models used for bipartite networks.

In the analysis of the binary version, we observe disassortative mixing as well as the presence of a nested structure in the network. This indicates that, in general, high degree banks tend to diversify their lending to almost all sectors, whereas low degree banks preferentially specialize their lending on high degree sectors. Similarly, high degree sectors tend to borrow from almost all banks, while low degree sectors prefer to borrow from high degree banks. Such interactions are also investigated in other bipartite bank-sector credit networks (e.g. see Fricke and Roukny, 2016), but here in line with recent studies (e.g. Johnson et al. 2013; Feng and Takemoto, 2014) we show that the intrinsic heterogeneity in the distribution of the degrees plays a crucial role for the emergence of this nested structure in the observed network. Moreover, in the analysis of the weighted version, we still observe the presence of disassortative mixing in the observed network, but the weighted version is less disassortative than the binary version.

Comparing the observed properties with the expected ones obtained from the different configuration models, we find that, for the binary case, on the one hand, over the years, the main trend of almost all properties can be simply reproduced by the BiBCM. On the other hand, for some higher-order properties, there are still some significant deviations from the expected ones obtained from the BiBCM, suggesting that these parts of the network are not completely explained by information embedded in the observed degrees. In addition, the elements of the one-mode projection matrices indicating the portfolio overlaps between banks and between sectors are mostly predicted by latent variables extracted from the degree sequence of the original bipartite network. This shows the relevance of the configuration models defined for the original bipartite structure instead of employing the null models defined for the one-mode projection networks.

In the weighted version, we consider two null models, i.e. the BiWCM preserving the observed strength sequence and the BiECM preserving both the observed degree as well as the strength sequences. Overall, we find that while maintaining only the strength sequence can not explain the main features of the network properties, the reproduction of such features is considerably improved by the additional knowledge of the degree sequence. Nevertheless,

we still observe that certain local properties can not be explained by these models, suggesting the presence of higher-order patterns in the weighted version that can not simply be explained by the lower-order properties of the observed network like the degree sequence and/or the strength sequence. One of the strongest deviations from the null models is observed in an unexpected high (under the null) clustering of the banks and the sectors with the highest degrees. Note that this statement indicates “excessive” clustering beyond the relatively high level that would be expected anyway because of the central position of these models.

This paper not only contributes to the literature on the identification of higher-order patterns in real world bipartite networks in general but also provides some important implications for the reconstruction of the real credit networks from limited information. Our findings suggest that most of the higher-order properties of the present network can be preserved when randomizing and reconstructing bipartite networks under local constraints like the degree sequence in the binary case, and both the degree as well as strength sequences in the weighted case. This is in line with what is found in the analysis of one-mode networks (e.g. Squartini et al., 2011a; Squartini et al., 2011b; Squartini and Garlaschelli, 2011; Masrandrea et al. 2014; Squartini et al., 2015). Obviously, the presence of significant deviations from the various configuration models indicates that further studies are needed in order to capture such subtle differences. In addition, we suggest that further extensions to the fitness-induced configuration models could provide a useful way to an assessment of the role of intrinsic non-topological properties associated with each node on the network topology (e.g. Garlaschelli and Loffredo, 2008; Cimini et al., 2015; Squartini et al., 2016).

It should be stressed that the methods as well as the bipartite configuration models employed in this study are generic enough to apply the same analysis to many other bipartite networks. Unfortunately, our original data set of the bank-firm credit network of Spain does not have detailed information about the amount of loans from banks to firms, consequently, using the BiWCM or BiECM is not suitable for the bank-firm credit network. In addition, the number of firms is very large in comparison to the number of banks so that applying the configuration models to randomize this network would impose a very high computational burden. We suggest that the evaluation of the topological and structural properties of the bank-firm credit network against the expected properties obtained from the various bipartite configuration models should be studied further in order to obtain a more comprehensive assessment of real world credit networks from different perspectives.

4.5 References

- Almeida-Neto M. A., Guimaraes P., Guimarães P. R., Loyola R. D., Ulrich W. 2008. A consistent metric for nestedness analysis in ecological systems: Reconciling concept and measurement. *Oikos* 117 (8), pp. 1227-1239.
- Aoyama H. 2014. Systemic risk in Japanese credit network. In Abergel F., Aoyama H., Chakrabarti B. K., Chakraborti A., Ghosh A. (Eds.), *Econophysics of Agent-Based Models* (pp. 219-228). Springer International Publishing.
- Aoyama H., Battiston S., Fujiwara Y. 2013. DebtRank analysis of the Japanese credit network. RIETI Discussion Paper Series 13-E-087. Available at: <http://www.rieti.go.jp/jp/publications/dp/13e087.pdf>.
- Bascompte J., Jordano P., Melián C. J., Olesen J.M. 2003. The nested assembly of plant-animal mutualistic networks. *Proc. Natl. Acad. Sci.* 100 (16), pp. 9383- 9387.
- Bascompte J., Jordano P. 2007. Plant-animal mutualistic networks: The architecture of biodiversity. *Annual Review of Ecology, Evolution, and Systematics* 38, pp. 567-593.
- Cimini G., Squartini T., Musmeci N., Puliga M., Gabrielli A., Garlaschelli D., Battiston S., Caldarelli G. 2015. Reconstructing topological properties of complex networks using the fitness model. In Aiello L. M., McFarland D. (Eds.), *Social Informatics* (pp. 323-333), Springer International Publishing.
- De Masi G., Fujiwara Y., Gallegati M., Greenwald B., Stiglitz J. E. 2011. An analysis of the Japanese credit network. *Evolutionary and Institutional Economic Review* 7 (2), pp. 209-232.
- De Masi G., Gallegati M. 2012. Bank-firms topology in Italy. *Empirical Economics* 43 (2), pp. 851-866.
- Di Gangi D., Lillo F., Pirino D. 2015. Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. Working Paper. Available at: [arXiv: 1509.00607](https://arxiv.org/abs/1509.00607).
- Feng W., Takemoto K. 2014. Heterogeneity in ecological mutualistic networks dominantly determines community stability. *Scientific Reports* 4.
- Fernandez M., Javier G., Hidalgo C. 2012. Bipartite networks provide new insights on international trade markets. *Networks and Heterogeneous Media* 7 (3), pp. 399-413.
- Fricke D., Lux T. 2015. Core-periphery structure in the overnight money market: Evidence from the e-MID trading platform. *Computational Economics* 45 (3), pp. 359-395.
- Fricke D., Roukny T. 2016. Generalists and specialists in the credit market. Working Paper. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2752413.

Fujiwara Y., Aoyama H., Ikeda Y., Iyetomi H., Souma W. 2009. Structure and temporal change of the credit network between banks and large firms in Japan. *Economics: The Open-Access, Open-Assessment E-Journal* 3.

Jonhson S., Domínguez-García V. and Muñoz M. A. 2013. Factors determining nestedness in complex networks. *PLoS ONE* 8 (9).

Garlaschelli D., Loffredo M. 2008. Maximum likelihood: Extracting unbiased information from complex networks. *Physical Review E* 78 (1).

Garas A., Rozenblat C., Schweitzer F. 2015. The network structure of city-firm relations. Working Paper. Available at: [arXiv:1512.02859](https://arxiv.org/abs/1512.02859).

Gilarranz L. J., Pastor J. M. Javier G. 2012. The architecture of mutualistic weighted networks. *Oikos* 121, pp. 1154-1162.

Hidalgo C. A., Hausmann R. 2009. The building blocks of economic complexity. *Proc. Natl. Acad. Sci.* 106 (26), pp. 10570-10575.

Hofmann T. 1999. Probabilistic latent semantic analysis. In *Proc. of Uncertainty in Artificial Intelligence, UAI 99*, Stockholm.

Huang X., Vodenska I., Havlin S., Stanley H. E. 2013. Cascading failures in bi-partite graphs: Model for systemic risk propagation. *Scientific Reports* 3.

Illueca M., Norden L., Udell G. F. 2013. Liberalization and risk-taking: Evidence from government-controlled banks. *Review of Finance* 18 (4), pp. 1217-1257.

Lehmann S., Schwartz M., Hansen L. K. 2008. Biclique communities. *Physical Review E* 78 (1).

Levy-Carciente S., Kenett D. Y., Avakian A., Stanley H. E., Havlin S. 2015. Dynamical macroprudential stress testing using network theory. *Journal of Banking & Finance* 59, pp. 164-181.

Lind P. G., González M. C., Herrmann H. J. 2005. Cycles and clustering in bipartite networks. *Physical Review E* 72 (5).

Luu D.T., Lux T., Yanovski B. 2016. Does the heterogeneity in the local constraints predominantly determine structural correlations in the Italian overnight money market?. Working Paper.

Lux T. 2016. A model of the topology of the bank-firm credit network and its role as channel of contagion. *Journal of Economic Dynamics and Control* 66, pp. 36-53.

Mastrandrea R., Squartini T., Fagiolo G., Garlaschelli D., 2014. Enhanced reconstruction of weighted networks from strengths and degrees. *New Journal Physics* 16.

Newman M. E. J. 2001. The structure of scientific collaboration networks. *Proc. Natl.*

Acad. Sci. 98 (2), pp. 404-409.

Newman M. E. J. 2002. Assortative mixing in networks. *Physical Review Letters* 89 (20).

Newman M. E. J. 2003. Mixing patterns in networks. *Physical Review E* 67 (2).

Onnela J. P., Saramäki J., Kertész J., and Kaski K. 2005. Intensity and coherence of motifs in weighted complex networks. *Physical Review E* 71 (6).

Park J., Newman M. E. J. 2004. The statistical mechanics of networks. *Physical Review E* 70 (6).

Saracco F., Clemente R. D., Gabrielli A., Squartini T. 2015. Randomizing bipartite networks: the case of the world trade web. *Scientific Reports* 5.

Saracco F., Clemente R. D., Gabrielli A., Squartini T. 2016. Detecting early signs of the 2007–2008 crisis in the world trade. *Scientific Reports* 6.

Squartini T., Fagiolo G., Garlashedelli D. 2011a. Randomizing world trade.I. A binary network analysis. *Physical Review E* 84 (4).

Squartini T., Fagiolo G., Garlashedelli D. 2011b. Randomizing world trade.II. A weighted network analysis. *Physical Review E* 84 (4).

Squartini T., Garlashedelli D. 2011. Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics* 13.

Squartini T., van Lelyveld I., Garlaschelli D. 2013. Early-warning signals of topological collapse in interbank networks. *Scientific Reports* 3.

Squartini T., Mastrandrea R., Garlaschelli D. 2015. Unbiased sampling of network ensembles. *New Journal of Physics* 17.

Squartini T., Caldarelli G., Cimini G. 2016. Stock markets reconstruction via entropy maximization driven by fitness and density. Working Paper. Available at: [arXiv:1606.07684](https://arxiv.org/abs/1606.07684).

Staniczenko P. P. A., Kopp J. C., Allesina S. 2013. The ghost of nestedness in ecological networks. *Nature Communication* 4.

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

van Lelyveld I., in't Veld D. 2014. Finding the core: Network structure in interbank markets. *Journal of Banking & Finance* 49, pp. 27–40.

4.6 Appendix

4.6.1 Motifs

Since there are no odd cycles in a bipartite structure, the often used clustering coefficients based on triangles in the one-mode networks must be modified. In this study, following the recent studies of Saracco et al. (2015) and Saracco et al. (2016), we employ a new class of motifs, which is basically composed by $V(k)$ motifs (i.e. all k nodes in the mode B links to a common node in the mode S) and $\Lambda(k)$ (i.e. all k nodes in S links to a common node in the mode B) motifs. For the sake of convenience, we suppose that $i = 1, 2, \dots, n$ represent n nodes in the mode B (i.e. banks in our study) and $j = 1, 2, \dots, m$ represent m nodes in the mode S (i.e. sectors in our study), and the notation $N_Y(A)$ stands for the number of motifs type Y in a bipartite network characterized by an adjacency matrix $A = \{A_{ij}\}_{n \times m}$. Figure (4.15) shows how different types of interactions between nodes in two modes can be represented by combinations of various $V(k)$ and $\Lambda(k)$ motifs.

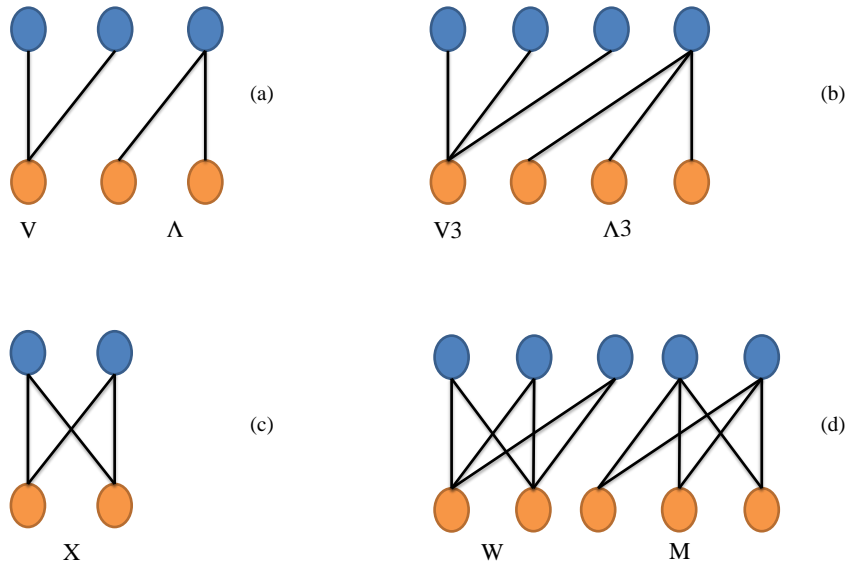


Figure 4.15: Different types of motifs in a bipartite network. Panel (a) shows V and Λ motifs, panel (b) shows $V(3)$ and $\Lambda(3)$ motifs, panel (c) shows a X motif, and panel (d) show W and M motifs. Different colors represent for different modes. The representation of the various motifs in this figure is based on Saracco et al. (2015), Saracco et al. (2016).

In the following, we will show how to measure the numbers of different types of motifs in a bipartite network.

V and Λ motifs

When $k=1$, $V(k), \Lambda(k)$ are nothing else but the degree sequence $\{kbs_i\}_1^n, \{ksb_j\}_1^m$. If $k=2$, $V(2)$ and $\Lambda(2)$ respectively are V and Λ motifs shown in the panel (a) of Figure (4.15). The number of motifs type V is given by

$$N_V(A) = \sum_{1 \leq i < i' \leq n} \sum_{j=1}^m A_{ij} A_{i'j} \quad (4.64)$$

$N_V(A)$ can be rewritten as a function of degree sequence $\{ksb_j\}_1^m$ in the mode S as

$$N_V(A) = \sum_{j=1}^{j=m} \binom{ksb_j}{2} \quad (4.65)$$

Similarly, the number of motifs type Λ is defined as

$$N_\Lambda(A) = \sum_{1 \leq j < j' \leq m} \sum_{i=1}^n A_{ij} A_{ij'}, \quad (4.66)$$

and $N_\Lambda(A)$ can be rewritten as a function of degree sequence $\{kbs_i\}_1^n$ in the mode B as

$$N_\Lambda(A) = \sum_{i=1}^{i=n} \binom{kbs_i}{2}. \quad (4.67)$$

$V(k), \Lambda(k)$ motifs ($k \geq 3$)

Extending $V(2)$ and $\Lambda(2)$, we can obtain $V(k), \Lambda(k)$ motifs for $k \geq 3$ (see the panel (b) of Figure (4.15) for an example of $k=3$). The numbers of $V(k)$ and $\Lambda(k)$ motifs are respectively given by

$$N_{V(k)}(A) = \sum_{1 \leq i^1 < i^2 < \dots < i^k \leq n} \sum_{j=1}^m A_{i^1 j} A_{i^2 j} \dots A_{i^k j}, \quad (4.68)$$

and

$$N_{\Lambda(k)}(A) = \sum_{1 \leq j^1 < j^2 < \dots < j^k \leq m} \sum_{i=1}^n A_{i j^1} A_{i j^2} \dots A_{i j^k}, \quad (4.69)$$

Equivalently, we have

$$N_{V(k)}(A) = \sum_{j=1}^{j=m} \binom{ksb_j}{k}, \quad (4.70)$$

and

$$N_{\Lambda(k)}(A) = \sum_{i=1}^{i=n} \binom{kbs_i}{k}. \quad (4.71)$$

As can be seen in Eqs. (4.68) and (4.69), in each case, all k nodes in one mode are linked to a node in another mode. These families of motifs therefore capture different orders of indirect interlocks between banks and between sectors via the bank-sector credit relations.

X motifs

Since each X motif consists of two nodes in the mode B and two nodes in the mode S (see Figure (4.15)), mathematically the number of X motifs is given by

$$N_X(A) = \sum_{1 \leq i < i' \leq n} \sum_{1 \leq j < j' \leq m} A_{ij} A_{ij'} A_{i'j} A_{i'j'}. \quad (4.72)$$

It can be shown that Eq. (4.72) has an alternative expression as a function of $\{A_{ii'}^{B-B}\}$ or $\{A_{jj'}^{S-S}\}$. More specifically,

$$N_X(A) = \sum_{1 \leq i < i' \leq n} \binom{A_{ii'}^{B-B}}{2} = \sum_{1 \leq j < j' \leq m} \binom{A_{jj'}^{S-S}}{2}. \quad (4.73)$$

Now we consider the local number of X motifs surrounding a node i_0 (i.e. $N_X(i_0)$) in the mode B or a node j_0 (i.e. $N_X(j_0)$) in the mode S. From Eq. (4.72) and Eq. (4.73), we obtain

$$N_X(i_0) = \sum_{i' \neq i_0} \sum_{1 \leq j < j' \leq m} A_{i_0j} A_{i_0j'} A_{i'j} A_{i'j'} = \sum_{i' \neq i_0} \binom{A_{i_0i'}^{B-B}}{2}, \quad (4.74)$$

and

$$N_X(j_0) = \sum_{j' \neq j_0} \sum_{1 \leq i < i' \leq n} A_{ij_0} A_{ij'} A_{i'j_0} A_{i'j'} = \sum_{j' \neq j_0} \binom{A_{j_0j'}^{S-S}}{2}. \quad (4.75)$$

M and W motifs

Graphically, each M (W) motif is a complete sub-network of two nodes in the mode B (mode S) and three nodes in the mode S (mode B) (see Figure (4.15)). The number of M motifs is given by

$$N_M(A) = \sum_{1 \leq i < i' \leq n} \sum_{1 \leq j < j' < j'' \leq m} A_{ij} A_{ij'} A_{ij''} A_{i'j} A_{i'j'} A_{i'j''}. \quad (4.76)$$

Equivalently,

$$N_W(A) = \sum_{1 \leq i < i' \leq n} \binom{A_{ii'}^{B-B}}{3}. \quad (4.77)$$

In a similar way, we obtain the number of W motif as

$$N_W(A) = \sum_{1 \leq j < j' \leq m} \sum_{1 \leq i < i' < i'' \leq n} A_{ij} A_{ij'} A_{i'j} A_{i'j'} A_{i''j} A_{i''j'} = \sum_{1 \leq j < j' \leq m} \binom{A_{jj'}^{S-S}}{3}. \quad (4.78)$$

One can also define the generalization of M and W motifs as $K_{k,l}$ motifs consisting of k nodes in the mode B and l nodes in the mode S (in this case, all k nodes in the mode B are connected to all l nodes in the mode S). The number of $K_{k,l}$ motifs is given by

$$N_{K_{k,l}}(A) = \sum_{1 \leq i^1 < i^2 < \dots < i^k \leq n} \sum_{1 \leq j^1 < j^2 < \dots < j^l \leq m} \prod_{u=1}^k \prod_{v=1}^l A_{i^u j^v}, \quad (4.79)$$

where $1 < k \leq n$, and $1 < l \leq m$. This is similar to the definition of K-bicliques proposed by Lehmann et al. (2008).

Note that in a special case $k=2$ or $l=2$ we obtain

$$N_{K_{2,l}}(A) = \sum_{1 \leq i < i' \leq n} \binom{A_{ii'}^{B-B}}{l}, \quad (4.80)$$

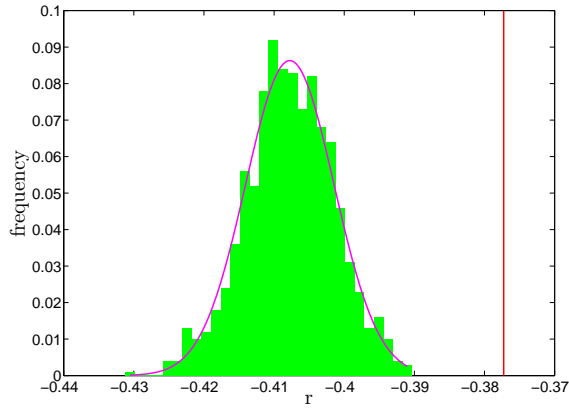
and

$$N_{K_{k,2}}(A) = \sum_{1 \leq j < j' \leq m} \binom{A_{jj'}^{S-S}}{k}. \quad (4.81)$$

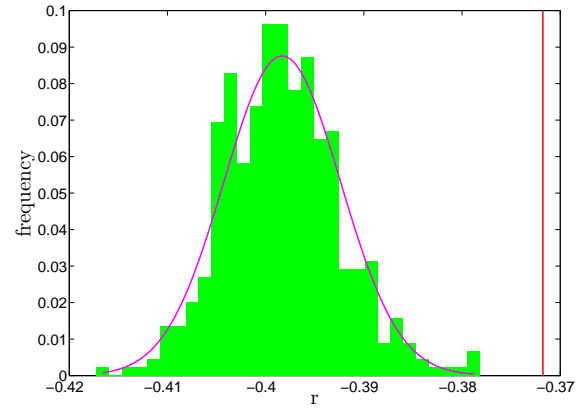
It is easy to show that Eqs. (4.73), (4.77), and (4.78) are just the special cases of Eq. (4.80) and Eq. (4.81).

4.6.2 Ensemble distributions of various network properties

In Figures (4.16), (4.17), and (4.18), we show the ensemble distributions of the overall assortativity and the numbers of the various motifs. Note that, fits in these Figures (the magenta solid lines) are obtained by superimposing normal distribution with the sample average and variance.

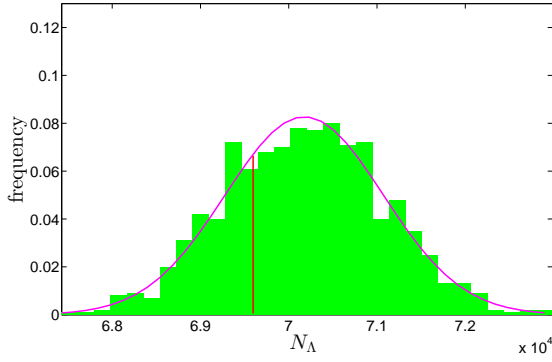


(a) Dist. of r in 1997

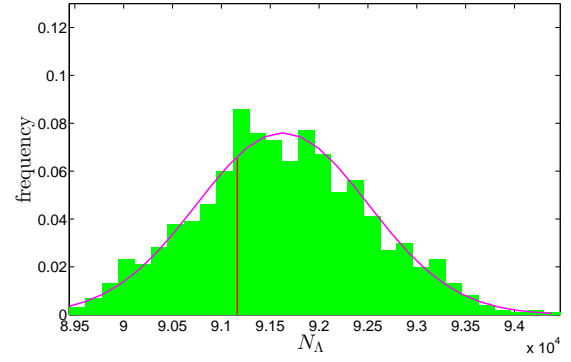


(b) Dist. of r in 2007

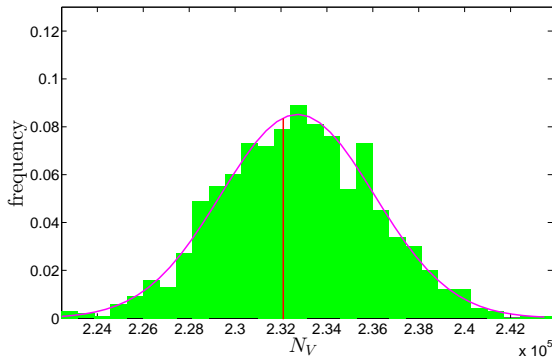
Figure 4.16: Distribution of Pearson correlation coefficient of degrees in the ensemble generated by the BiBCM. Panels (a), (b) report the ensemble distribution of r , respectively in 1997 and 2007 as examples. The red vertical line stands for the observed r .



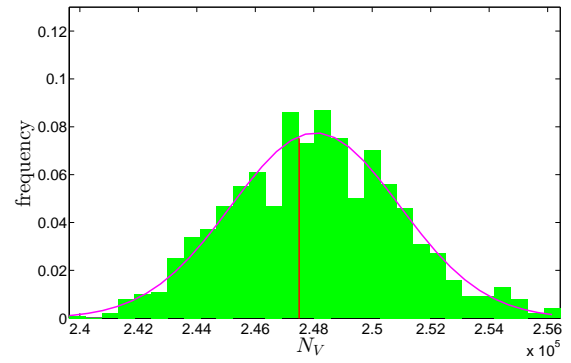
(a) Dist. of N_Λ in 1997



(b) Dist. of N_Λ in 2007

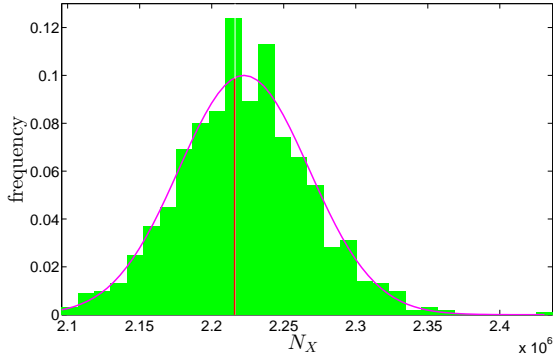


(c) Dist. of N_V in 1997

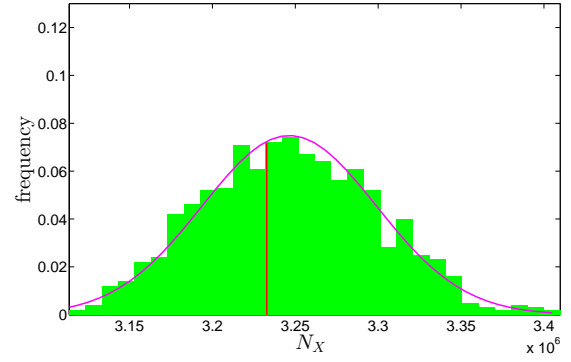


(d) Dist. of N_V in 2007

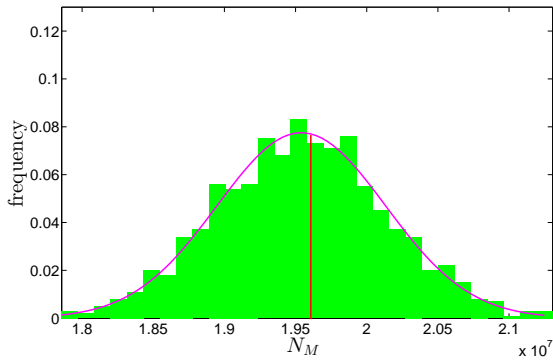
Figure 4.17: Ensemble distributions of N_Λ and N_V under the BiBCM. Panels (a), (b) are the ensemble distribution of N_Λ in 1997 and 2007, respectively. Panels (c), (d) are ensemble distribution of N_V in 1997 and 2007, respectively. In each panel, the red vertical line stands for the observed value.



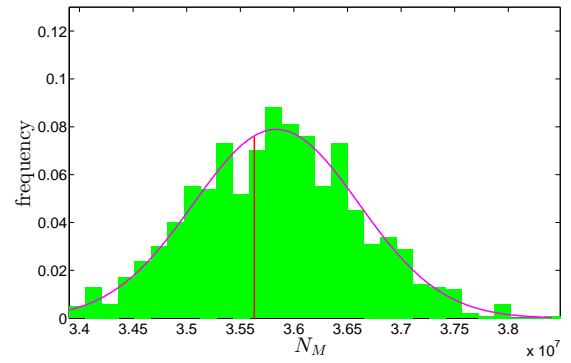
(a) Dist. of N_X in 1997



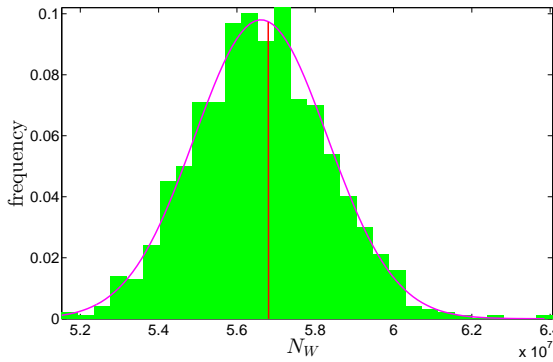
(b) Dist. of N_X in 2007



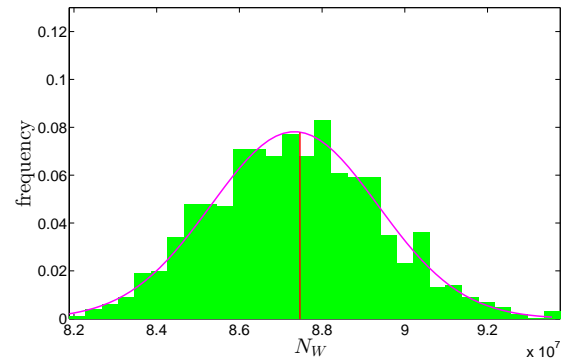
(c) Dist. of N_M in 1997



(d) Dist. of N_M in 2007



(e) Dist. of N_W in 1997



(f) Dist. of N_W in 2007

Figure 4.18: Ensemble distributions of N_X , N_M , N_W under the BiBCM. Panels (a), (b) are the ensemble distribution of N_X in 1997 and 2007, respectively. Panels (c), (d) are the ensemble distribution of N_M in 1997 and 2007, respectively. Panels (e), (f) are the ensemble distribution of N_W in 1997 and 2007, respectively. In each panel, the red vertical line stands for the observed value.

4.6.3 z-scores of ANNs and C_4^w evaluated under the BiWCM and the BiECM.

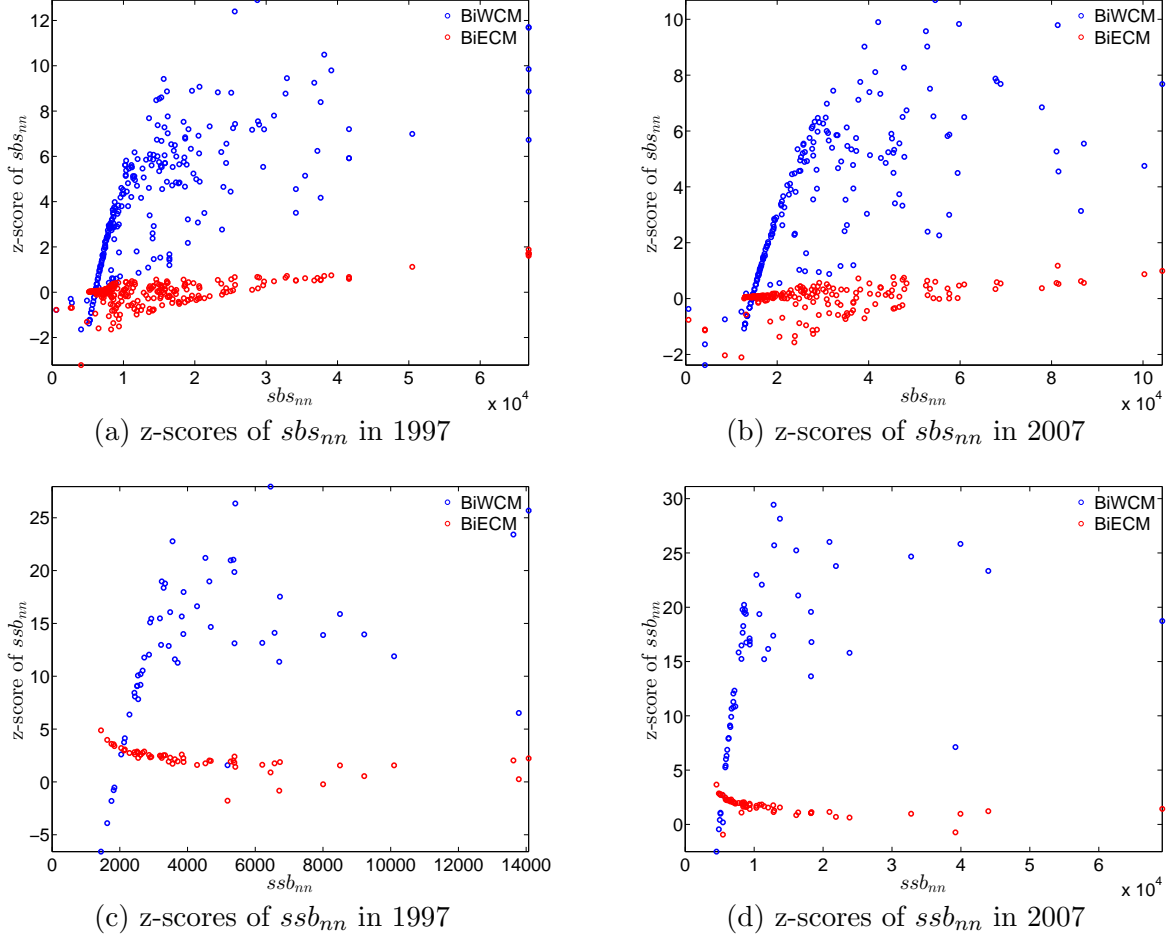
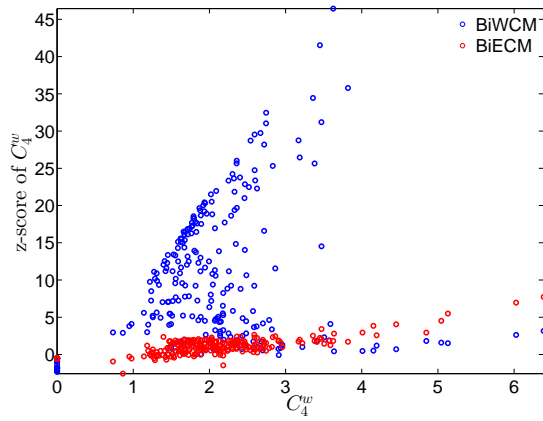
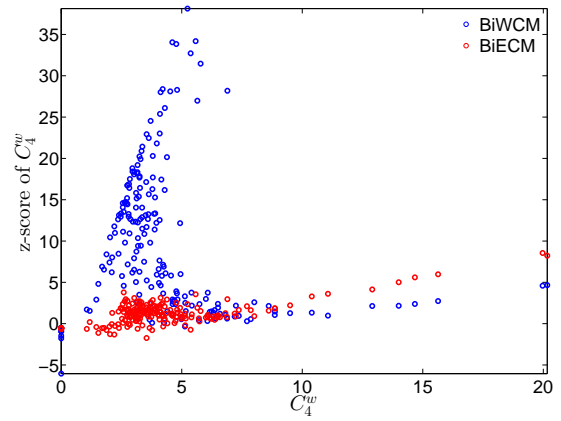


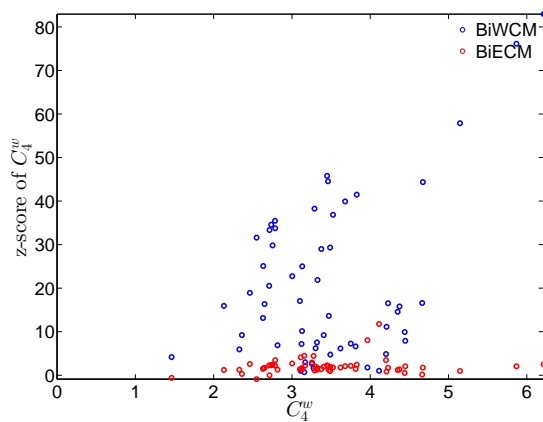
Figure 4.19: z-scores of ANNs evaluated under the BiWCM (blue circles) and under the BiECM (red circles), in 1997 and 2007. Panels (a), (b): z-score of sbs_{nn} vs. sbs_{nn} , respectively in 1997 and 2007. Panels (c), (d): z-score of ssb_{nn} vs. ssb_{nn} , respectively in 1997 and 2007.



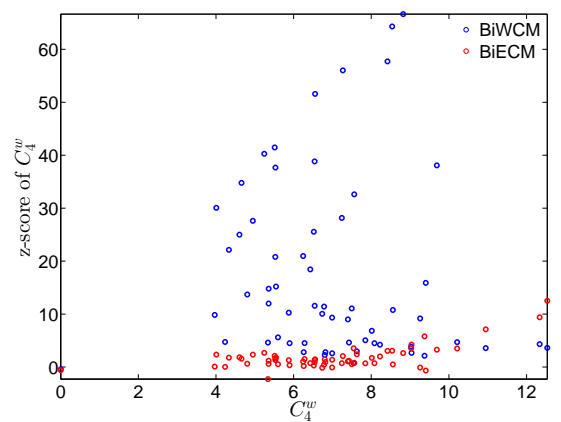
(a) z-scores of C_4^w for banks in 1997



(b) z-scores of C_4^w for banks in 2007



(c) z-scores of C_4^w for sectors in 1997



(d) z-scores of C_4^w for sectors in 2007

Figure 4.20: z-scores of C_4^w evaluated under the BiWCM (blue circles) and under the BiECM (red circles) vs. C_4^w . Panel (a) for banks in 1997, panel (b) for banks in 2007. Panel (c) for sectors in 1997, panel (d) for sectors in 2007.

Chapter 5

Multilayer Overlaps and Correlations in the Bank-Firm Credit Network of Spain

Coauthored by: Thomas Lux.

Keywords: Bank-Firm Credit Network; Multilayer Network; Multiplexity; Overlaps; Correlations.

5.1 Introduction

The recent financial crisis shows the danger of financial and economic links in propagating distress and creating systemic risk. Much of the analysis of risk contagion in the economic and financial system has so far focused on directed exposures in single mode networks (e.g. in the interbank lending network, in production networks). Less attention has, however, been devoted to the propagation of distress between the financial and the real sector of the economy.

In the last few years, there has been a growing interest in studying different channels of financial contagion by adding additional layers of connectivity within the network structure of the banking system. In a multilayer architecture, a failure of nodes in one layer may lead to other failures of connected nodes in other layers, and then the accumulated repercussion between all layers may eventually lead to a cascade of failures.

For example, in a model for systemic risk propagation in a bipartite network of banks' portfolio overlaps, the study of Huang et al. (2013) suggests that fluctuations in bank asset

values may affect the financial robustness of a group of banks, and in turn it may force these banks to sell their assets. If this selling causes the market price of the assets to decrease, the balance sheets of all other banks who hold these assets will be affected due to mark-to-market evaluation. If this loss is large enough, further selling pressure may cause further devaluation of assets in the market. The failures may continue to spread to other banks and other assets and finally lead to a large decline of the banks' balance sheets as well as the capitalization in the assets market. In a similar study, Caccioli et al. (2014) develop a bank-asset bipartite network approach to amplification of financial contagion due to the combination of common asset holdings and leverage. Their findings suggest a "robust yet fragile" nature of the financial system. A higher degree of diversification in asset holdings of banks can create dangerous systemic effects, and too high leverage ratios can cause the unstable region of the financial network to grow.

Considering relationships between the banking system to non-financial firms, in a recent study, Lux (2016) suggests that to add bank-firm credit relations as a second important layer to a standard model of the interbank credit market. Although the bipartite structure in the bank-firm credit network is much sparser than in the interbank network, this does not imply that lending to non-financial firms should be of less concern from the viewpoint of contagious spread of stress. Through computational experiments, his study suggests that in a certain number of cases, for instance due to a high local interlock between banks, a default of single business firms in the real sector may trigger a large scale collapse of the network.

There exist a small number of empirical studies on the bank-firm credit markets focusing on the topological properties of the aggregate bipartite structure or one-mode projection networks (e.g. De Masi et al., 2011; De Masi and Gallegati, 2012). However, such an aggregate perspective neglects the differences between various sectors of the real sphere of the economy and effectively treats credit to different sectors as perfect substitutes in banks' balance sheets. For instance, the issues of sectoral diversifications, sectoral overlaps, and sectoral correlations in banks' lending can not be addressed in such a setting. In fact, if we look at the portfolios of banks we can see that banks' lending distribution is diversified and that the lending strategies may also be heterogeneous. Some banks may prefer to diversify their portfolios, while others concentrate on a certain number of sectors and firms since they might have acquired expertise in lending, for example, due to lower costs of monitoring. To gain a deeper understanding of such differences, it is necessary to consider the multilayer structure of bank-firm credit relationships.

In the past few years, in network science, major advances in the understanding of the

statistical properties and mechanics of multilayer networks have been made. Among others, the issue of overlaps and dependencies between layers has received significant attention since it is important to understand the interactions between layers (e.g. Bianconi, 2014; Boccaletti et al., 2014; Menichetti, Remondini, and Bianconi, 2014; Menichetti et al., 2014; Kivela et al., 2014). It is commonly observed that the robustness of a multiplex network is significantly affected by the interactions between its layers (e.g. Bianconi, 2014; Chen et al., 2015; Menichetti et al., 2016).

Among a limited number of empirical studies on economic and financial multiplex networks, probably the International Trade Network (ITN) has been studied most intensively due to advantages in data availability ¹. For instance, Barigozzi et al. (2010) decompose the ITN into almost 100 layers based on commodity codes. Their findings show that the distributions of link weights in single layers are highly heterogeneous, and the network properties at commodity levels are very different from those of the aggregated one. In addition, the analysis of cross-layer correlations suggests that the ITN is not simply an aggregate network of many independent layers.

In order to assess the statistical significance of network properties, one needs to compare the observed properties with those of an appropriate null model. To construct such null models for multilayer networks, Mastrandrea et al. (2014b) adopt the benchmark of the so-called configuration models (e.g. Squartini et al., 2011a; Squartini et al., 2011b; Mastrandrea et al., 2014a; Squartini et al., 2015) to investigate the role of local constraints like the degree distribution and the strength distribution in replicating some structural correlations of the World Trade multiplex. Decomposing the aggregate trade network into nearly 100 commodity specific layers, their findings indicate that the main features of the structural correlations like average nearest neighbor degrees as well as clustering coefficients are well replicated at both the layer and aggregate levels by the local constraints. Two recent studies of Gemmetto and Garlaschilli (2015), Gemmetto et al. (2015) have applied multiplexity measures to the International Trade Network as well as the European Airport Network, and then compare these networks with random graph models and configuration models including the so called binary configuration model (Squartini et al., 2011a) and the so called weighted configuration model (Squartini et al., 2011b). Gemmetto and Garlaschilli (2015)

¹In fact, there are also several attempts on analyzing multilayer financial networks, such as Aldasoro and Alves (2015), Bargigli et al. (2015), and Poledna et al. (2015). They also suggest that the multilayer nature of financial networks should be taken into account instead of analyzing aggregate or single-layer networks. In their studies, the classification of layers is often based on direct relations/transactions between banks. In contrast, in this study we focus on the multilayer structure of joint exposures of banks when lending to non-financial firms in different industrial sectors.

and Gemmetto et al. (2015) show that the local constraints in single layers such as the degree sequence and the strength sequence play an important role in explaining the dependencies between layers, highlighting the important role played by the hubs across layers. Their findings also suggest that the use of homogeneous benchmarks (i.e. the family of random graph models), which treat all agents as homogeneous in each layer, can lead to misleading results.

Following recent advances in the study of multiplex networks (e.g. Bianconi, 2013; Gemmetto and Garlaschilli, 2015; Gemmetto et al., 2015), in our paper we will analyze the bank-firm credit network of Spain with a focus on the overlaps and correlations between layers representing industrial sectors of firms ². In particular, we investigate whether the observed correlations between layers differ significantly from the expected values obtained from random graph models specifying only global constraints and from configuration models preserving the intrinsic heterogeneity in the degree sequence and/or strength sequence of the observed network. Our aims are twofold: (i) to investigate the roles of different constraints in explaining the dependencies between layers in the observed multiplex network; and (ii) to identify the presence of patterns in such dependencies. Regarding the null models, we add an important novelty to the theoretical framework for measuring multiplexity in multilayer networks, namely the rescaled multiplexity evaluated under the so-called enhanced configuration model (ECM) where both the observed degree as well as strength sequences are enforced in each layer.

The remainder of this paper is structured as follows. In Sec. 5.2 we briefly describe the methods used for measuring the overlaps and multiplexity in a multilayer network. Sec. 5.3 reports our main findings for the bank-firm credit market of Spain. Discussions and concluding remarks are in Sec. 5.4. At the end of this paper, the Appendix provides an outline of the assessment of significance of multiplex features under the ECM.

5.2 Overlaps and Multiplexity in multilayer networks

5.2.1 General definitions, measures for overlaps, and null models

In this part, first, we briefly define some basic notations and definitions of a multiplex network. After that we will explain the method used to quantify the overlaps between layers. Basically, we follow the approach proposed in Bianconi (2013) and in Menichetti et

²For more detailed analyses of the evolution of the bank-firm credit market and the bank-sector credit network of Spain, we refer the readers to, for example, Illueca et al. (2013) and Luu and Lux (2016).

al. (2014), respectively, for binary and weighted multiplex networks ³. Next, we introduce different null models employed in this study including the family of random graph models and the family of configuration models.

5.2.1.1 General definitions

Generally, suppose we have N nodes and NS layers such that nodes may be connected with others in any layer. A generic multiplex can be defined by $\vec{G} = (G^1, G^2, \dots, G^\alpha, \dots, G^{NS})$, where the network of each layer α , namely G^α , is characterized by a binary adjacency matrix $A_\alpha = \{a_{ij}^\alpha\}$ and/or a weighted matrix $W_\alpha = \{w_{ij}^\alpha\}$, where i, j are indexed from 1 to N representing the N nodes in the network. It should be noted that, in an undirected network, the adjacency and weighted matrices are symmetric.

Similar to elementary one-mode networks, we define the degree and strength of node i in each layer $\alpha = 1, 2, \dots, NS$ as the (unweighted and weighted) sums of their links:

$$k_i^\alpha = \sum_{j \neq i} a_{ij}^\alpha, \quad (5.1)$$

$$s_i^\alpha = \sum_{j \neq i} w_{ij}^\alpha. \quad (5.2)$$

The total number of links L^α and total weight W^α in layer α in an undirected network are given by

$$L^\alpha = \sum_i \sum_{i < j} a_{ij}^\alpha, \quad (5.3)$$

$$W^\alpha = \sum_i \sum_{i < j} w_{ij}^\alpha. \quad (5.4)$$

Now, we will relate the above definitions to our particular data, i.e. the bank-firm credit network. Assume that we have a binary, undirected, bipartite credit network without sectoral differentiation of N banks lending to F firms in NS sectors. We define the adjacency matrix of that network as

$$A^{B-F} = \{a_{ij}^{b-f}\}_{N \times F} \quad (5.5)$$

where $a_{ij}^{b-f} = 1$ if bank i lends to firm j , and zero otherwise. The degrees of bank i and firm

³We refer the readers to the papers of Bianconi (2013) and Menichetti et al. (2014) for a more detailed description of the multiplexity and overlaps in a multilayer network. A more comprehensive review on multilayer networks can be found in the study of Kivela et al. (2014), for instance.

j in the bank-firm credit network without sectoral differentiation are respectively given by

$$k_{bf}(i) = \sum_{k=1}^F a_{ik}^{b-f} \quad (5.6)$$

and

$$k_{fb}(j) = \sum_{k=1}^B a_{kj}^{b-f}. \quad (5.7)$$

Considering bank-sector relations, we define the adjacency matrix of the bipartite bank-sector credit network consisting of N banks and NS sectors as

$$M^{B-S} = \{m_{ij}^{b-s}\}_{N \times NS} \quad (5.8)$$

where $m_{ij}^{b-s} = 1$ if bank i lends to at least one firm in sector j , and zero otherwise.

Denote

$$k_{bs}(i) = \sum_{j=1}^{NS} m_{ij}^{b-s}, \quad (5.9)$$

$$k_{sb}(j) = \sum_{i=1}^N m_{ij}^{b-s}, \quad (5.10)$$

respectively the degrees of bank i and sector j in the bank-sector credit network.

The degrees k_{bf} and k_{bs} indicate, respectively, the diversification in the loan portfolio of banks at firm and sector levels, while the degrees k_{sb} and k_{fb} show the diversification in borrowing of firms and sectors respectively.

In addition, we can decompose the original bank-firm credit network A^{B-F} into bank-firm lending relations in a particular sector. $M^{B-F,\alpha} = \{m_{ij}^{b-f,\alpha}\}_{N \times F_\alpha}$ denotes the adjacency matrix showing the credit relations between banks and firms in each sector α , where $m_{ij}^{b-f,\alpha} = 1$ if bank i lends to firm j in sector α and zero otherwise; and F_α is the number of firms in sector α .

From the adjacency matrix of each sector α , we obtain a network structure of joint exposures of banks by projecting the bipartite bank-firm network $M^{B-F,\alpha}$ onto an undirected bank-bank network with an adjacency matrix $A_\alpha = \{a_{ij}^\alpha\}_{N \times N}$ and a weighted matrix $W_\alpha = \{w_{ij}^\alpha\}_{N \times N}$ as mentioned at the beginning of this subsection.

Note that, from the economic point of view, $a_{ij}^\alpha = 1$ if there is a firm in sector α that two banks i and j co-finance. In addition, w_{ij}^α is nothing else but the total number of common borrowers in sector α of two banks i and j . In addition, k_i^α (Eq. (5.1)) and s_i^α (Eq. (5.2)),

respectively, indicate the degree and the strength of bank i in the layer α . Consequently, the aggregate bank-firm credit network is now reduced to a multiplex network of N banks in NS layers, and each of NS layers shows the overlaps between the N banks in their lending to the industrial sector of firms associated to that layer.

5.2.1.2 Measures for overlaps

For two different layers α and β , the overall binary overlap is given by

$$O_{bin}^{\alpha,\beta} = \sum_i \sum_{i<j} a_{ij}^\alpha a_{ij}^\beta. \quad (5.11)$$

The local binary overlap for node i in two layers α, β is defined as

$$o_{i,bin}^{\alpha,\beta} = \sum_{j \neq i} a_{ij}^\alpha a_{ij}^\beta. \quad (5.12)$$

Another way to define the overlaps between layers is based on the minimum values of the elements in two adjacency matrices. These measures are proposed by Gemmetto and Garlaschelli (2015). More specifically, the overall binary overlap between two layers α and β is given by

$$O_{bin}^{\alpha,\beta} = \sum_i \sum_{i<j} \min(a_{ij}^\alpha, a_{ij}^\beta). \quad (5.13)$$

It is easy to show that $\min(a_{ij}^\alpha, a_{ij}^\beta) = a_{ij}^\alpha a_{ij}^\beta$, consequently Eq. (5.11) and Eq. (5.13) are equivalent. Similarly, ones can also define the local binary overlap for node i as

$$o_{i,bin}^{\alpha,\beta} = \sum_{j \neq i} \min(a_{ij}^\alpha, a_{ij}^\beta), \quad (5.14)$$

which is again identical to Eq. (5.12).

For weighted networks, both definitions do lead to different measurements, and it is mainly the second version that is intuitively appealing. The overall weighted overlap between two layers α and β is then defined as

$$O_w^{\alpha,\beta} = \sum_i \sum_{i<j} \min(w_{ij}^\alpha, w_{ij}^\beta). \quad (5.15)$$

The local weighted overlap for node i in two layers α, β is defined as

$$o_{i,w}^{\alpha,\beta} = \sum_{j \neq i} \min(w_{ij}^\alpha, w_{ij}^\beta). \quad (5.16)$$

The overlap coefficients indicate the similarities between layers. In the binary case, while $O_{bin}^{\alpha,\beta}$ is the total number of pairs of nodes linked at the same time by an edge in layer α and an edge in layer β , the local measure $o_{i,bin}^{\alpha,\beta}$ is the total number of other nodes connected to node i at the same time by an edge in α and an edge in layer β . The overall weighted overlap $O_w^{\alpha,\beta}$ and the local weighted overlap $o_{i,w}^{\alpha,\beta}$ are interpreted in a similar way but they take into account the weights of the links in both layers α and β . Based on these definitions, in the next subsection we will define the raw binary multiplexity and the raw weighted multiplexity showing the correlations between layers in a multilayer network.

5.2.1.3 Null models

The seminal work of Bianconi (2013) provides a framework for the formal characterization of correlated multiplex ensembles satisfying certain constraints. Accordingly, in a multiplex ensemble, where the probability for each possible outcome is specified by $P(\vec{G})$, the (overall) entropy of the multiplex is defined as

$$S = - \sum_G P(\vec{G}) \ln P(\vec{G}). \quad (5.17)$$

In the case of an uncorrelated multiplex, i.e., if the correlation between any two separate layers is equal to zero, we can decompose the overall entropy in Eq. (5.17) into the sum of the single layers' entropy

$$S = \sum_{\alpha=1}^{NS} S^\alpha = - \sum_{\alpha=1}^{NS} P_\alpha(G^\alpha) \ln P_\alpha(G^\alpha) \quad (5.18)$$

since we have

$$P(\vec{G}) = \prod_{\alpha=1}^{NS} P_\alpha(G^\alpha) \quad (5.19)$$

where $P_\alpha(G^\alpha)$ is the probability of G^α in the layer α . In contrast, if layers are correlated, we can not factorize the overall entropy as in Eq. (5.18).

We follow the ideas of Gemmetto and Garlaschelli (2015) and Gemmetto et al. (2015) to consider the role of constrains in single layers on correlations between layers of a multiplex

network. For instance, for each layer α of a multiplex network that we have defined above, we may want to preserve the total degree L^α , or the degree sequence $\{k_i^\alpha\}_{i=1}^{i=N}$, or the total weight W^α , or the strength sequence $\{s_i^\alpha\}_{i=1}^{i=N}$, or both $\{k_i^\alpha\}_{i=1}^{i=N}$ as well as $\{s_i^\alpha\}_{i=1}^{i=N}$.

Recent studies (e.g. Squartini et al., 2011a; Squartini et al., 2011b; Mastrandrea et al., 2014a; Squartini et al., 2015) provide a general method for unbiased randomizing ensembles for different versions of networks with various constraints. Based on maximum entropy and maximum likelihood methods, hidden variables are analytically extracted from the initially given constraints. In the following, we briefly introduce the family of random graph models and the family of configuration models. The main difference between both families is that in the former, we specify only global constraints and therefore treat all nodes in the network homogeneously, while in the latter the intrinsic heterogeneity in the degree and/or strength distributions is maintained.

Random Graph models

In the undirected Binary Random Graph model (BRG), with the total degree L^α of N nodes in each layer α , the probability of a link between any two different nodes in that layer is given by

$$p^\alpha = \frac{2L^\alpha}{N(N-1)} = \langle a_{ij}^\alpha \rangle_{BRG}, \quad (5.20)$$

where $\langle a_{ij}^\alpha \rangle_{BRG}$ is the notation for the expectation of a_{ij}^α under the BRG model ⁴.

In the undirected Weighted Random Graph model (WRG), the total strength W^α of N nodes in each layer α , the probability of a link of weight w_{ij}^α between any two different nodes (i, j) in the layer α is equal to

$$q(w_{ij}^\alpha) = \begin{cases} 1 - p_w^\alpha, & \text{if } w_{ij}^\alpha = 0, \\ (p_w^\alpha)^{w_{ij}^\alpha} (1 - p_w^\alpha), & \text{if } w_{ij}^\alpha > 0, \end{cases} \quad (5.21)$$

where the parameter p_w^α is given by

$$p_w^\alpha = \frac{2W^\alpha}{N(N-1) + 2W^\alpha}. \quad (5.22)$$

The expected link weight between node i and node j is

$$\langle w_{ij}^\alpha \rangle_{WRG} = \frac{2W^\alpha}{N(N-1)}. \quad (5.23)$$

⁴Throughout this paper, $\langle X \rangle_Y$ is the notation for expectation of X under the referenced null model Y .

The probability of a link between node i and node j is

$$\langle a_{ij}^\alpha \rangle_{WRG} = p_w^\alpha = \frac{2W^\alpha}{N(N-1) + 2W^\alpha}. \quad (5.24)$$

Configuration models

We now move on to configuration models that preserve the degree and/or strength sequence of a given network.

In the undirected Binary Configuration model (BCM), in each layer α , the degree sequence $\{k_i^\alpha\}_{i=1}^N$ is given. Mathematically, we then need to solve a system of N equations in each layer α to obtain N non-negative hidden variables $\{x_i^\alpha\}_{i=1}^N$ that determine the link formation probability

$$\sum_{j \neq i} \frac{x_i^\alpha x_j^\alpha}{1 + x_i^\alpha x_j^\alpha} = k_i^\alpha, \forall i = 1, 2, \dots, N. \quad (5.25)$$

The probability of a link between node i and node j in layer α is given by

$$\langle a_{ij}^\alpha \rangle_{BCM} = p_{ij}^\alpha = \frac{x_i^\alpha x_j^\alpha}{1 + x_i^\alpha x_j^\alpha}. \quad (5.26)$$

In the undirected Weighted Configuration model (WCM), in each layer α , the strength sequence $\{s_i^\alpha\}_{i=1}^N$ is given. Hidden variables $\{x_i^\alpha\}_{i=1}^N \in [0, 1)$ associated with N nodes are again obtained by solving

$$\sum_{j \neq i} \frac{x_i^\alpha x_j^\alpha}{1 - x_i^\alpha x_j^\alpha} = s_i^\alpha, \forall i = 1, 2, \dots, N. \quad (5.27)$$

The probability of a link of weight w_{ij}^α between node i and node j in the WCM is

$$q(w_{ij}^\alpha) = \begin{cases} 1 - p_{ij}^\alpha, & \text{if } w_{ij}^\alpha = 0, \\ (p_{ij}^\alpha)^{w_{ij}^\alpha} (1 - p_{ij}^\alpha), & \text{if } w_{ij}^\alpha > 0, \end{cases} \quad (5.28)$$

where

$$p_{ij}^\alpha = x_i^\alpha x_j^\alpha. \quad (5.29)$$

The expected link weight between node i and node j is given by

$$\langle w_{ij}^\alpha \rangle_{WCM} = \frac{x_i^\alpha x_j^\alpha}{1 - x_i^\alpha x_j^\alpha}. \quad (5.30)$$

The probability of a link between node i and node j in layer α is specified as

$$\langle a_{ij}^\alpha \rangle_{WCM} = p_{ij}^\alpha = x_i^\alpha x_j^\alpha. \quad (5.31)$$

In the undirected Enhanced Configuration model (ECM), in each layer α , the degree sequence $\{k_i^\alpha\}_{i=1}^N$ as well as the strength sequence $\{s_i^\alpha\}_{i=1}^N$ are given. The non-negative variables $\{x_i^\alpha\}_{i=1}^N$ and $\{y_i^\alpha\}_{i=1}^N$ ($\{y_i^\alpha\}_{i=1}^N \in [0, 1)$) are the solution to the system of $2N$ equations

$$\begin{aligned} \sum_{j \neq i} \frac{x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha}{1 - y_i^\alpha y_j^\alpha + x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha} &= k_i^\alpha, \forall i = 1, 2, \dots, N, \\ \sum_{j \neq i} \frac{x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha}{(1 - y_i^\alpha y_j^\alpha)(1 - y_i^\alpha y_j^\alpha + x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha)} &= s_i^\alpha, \forall i = 1, 2, \dots, N. \end{aligned} \quad (5.32)$$

The probability a link of weight w_{ij}^α is now given by

$$q(w_{ij}^\alpha) = \begin{cases} 1 - p_{ij}^\alpha, & \text{if } w_{ij}^\alpha = 0, \\ p_{ij}^\alpha r_{ij,\alpha}^{w_{ij}^\alpha - 1} (1 - r_{ij,\alpha}), & \text{if } w_{ij}^\alpha > 0, \end{cases} \quad (5.33)$$

where

$$p_{ij}^\alpha = \frac{x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha}{1 - y_i^\alpha y_j^\alpha + x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha}, \quad (5.34)$$

and

$$r_{ij,\alpha} = y_i^\alpha y_j^\alpha. \quad (5.35)$$

The expected link weight between node i and node j is given by

$$\langle w_{ij}^\alpha \rangle_{ECM} = \frac{x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha}{(1 - y_i^\alpha y_j^\alpha)(1 - y_i^\alpha y_j^\alpha + x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha)}, \quad (5.36)$$

and the probability of a link between node i and node j is

$$\langle a_{ij}^\alpha \rangle_{ECM} = p_{ij}^\alpha = \frac{x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha}{1 - y_i^\alpha y_j^\alpha + x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha}. \quad (5.37)$$

One important result of the configuration models is that, mathematically, the expected values of the properties of networks over the ensemble can be calculated via hidden variables (see, for instance, Squartini and Garlaschelli, 2011; Squartini et al., 2015). In addition, in a multilayer network, the expected values of the overlaps and correlations between layers still depend on the properties of the expected adjacency matrices (in the binary case) as well as

the expected weighted matrices (in the weighted case), and they can analytically be computed based on hidden variables extracted in the single layers (Gemmetto and Garlaschelli, 2015; Gemmetto et al., 2015). In the following subsection, we will show how to measure the expected and rescaled multiplicities under the various null models.

5.2.2 Multiplexity

We will in the following, first define the multiplicities of an observed multilayer network, and then introduce the expectation for these measures under the various null models. In addition, to assess the statistical significance of the observed multiplicities, we employ the analysis of z-scores indicating the difference between the observed and the expected multiplicities in units of the standard deviation of the pertinent measure under any null model.

For a more detailed description and proof for the expected multiplicities and z-scores under the BRG model, the BCM, the WRG model, and the WCM, we refer the readers to the study of Gemmetto and Garleaschelli (2015). We add the analytical details for the ECM in the Appendix.

5.2.2.1 Undirected binary multiplexity

Raw binary multiplexity

For binary network, the raw (observed) multiplexity between two layers α and β is defined as

$$m_{bin}^{\alpha,\beta} = \frac{2 \sum_i \sum_{i < j} \min(a_{ij}^\alpha, a_{ij}^\beta)}{L^\alpha + L^\beta}. \quad (5.38)$$

The multiplexity indicator between any two layers α and β ranges in $[0,1]$, and it is maximally equal to 1 if α and β are identical, and minimally equal to 0 when there is no overlap between the two layers. Clearly, the raw binary multiplexity is actually based on the overall overlap between two layers as achieved in Eq. (5.13). The only difference here is the term $\frac{L^\alpha + L^\beta}{2}$, which is used to normalize $O_{bin}^{\alpha,\beta}$.

We rescale this measured multiplexity by

$$\mu_{bin}^{\alpha,\beta} = \frac{m_{bin}^{\alpha,\beta} - \langle m_{bin}^{\alpha,\beta} \rangle}{1 - \langle m_{bin}^{\alpha,\beta} \rangle}, \quad (5.39)$$

where $\langle m_{bin}^{\alpha,\beta} \rangle$ is the expected value of the binary multiplexity under the chosen null model. This rescaled multiplexity measures the relative difference between the observed multiplexity

and the expected one with respect to the referenced null model. It is easy to show that $\mu_{bin}^{\alpha,\beta} < 1$.

Mathematically, the general formulation of $\langle m_{bin}^{\alpha,\beta} \rangle$ is given by

$$\langle m_{bin}^{\alpha,\beta} \rangle = \frac{2 \sum_i \sum_{i < j} \langle \min(a_{ij}^\alpha, a_{ij}^\beta) \rangle}{L^\alpha + L^\beta}. \quad (5.40)$$

In the binary case, it can be shown that

$$\langle \min(a_{ij}^\alpha, a_{ij}^\beta) \rangle = p_{ij}^\alpha p_{ij}^\beta, \quad (5.41)$$

where p_{ij}^α is the probability of a link between nodes i and j in layer α . It should be noted that under different null models (i.e. BRG, BCM), p_{ij}^α will have different forms.

Binary z score

While the rescaled quantities can explain the similarity between layers, we still need another measure that shows how much the the raw value $m_{bin}^{\alpha,\beta}$ deviates from the expected value $\langle m_{bin}^{\alpha,\beta} \rangle$. This leads to the definition of the z score of $m_{bin}^{\alpha,\beta}$ (binary z score in short) under the chosen null model, which is given by

$$z[m_{bin}^{\alpha,\beta}] = \frac{m_{bin}^{\alpha,\beta} - \langle m_{bin}^{\alpha,\beta} \rangle}{\sigma[m_{bin}^{\alpha,\beta}]}, \quad (5.42)$$

where $\sigma[m_{bin}^{\alpha,\beta}]$ is the standard deviation of $m_{bin}^{\alpha,\beta}$ ⁵.

Note that

$$\sigma[m_{bin}^{\alpha,\beta}] = \frac{2\sigma[\sum_i \sum_{i < j} \min\{a_{ij}^\alpha, a_{ij}^\beta\}]}{L^\alpha + L^\beta}, \quad (5.43)$$

and

$$\sigma^2[\min\{a_{ij}^\alpha, a_{ij}^\beta\}] = \langle \min\{a_{ij}^\alpha, a_{ij}^\beta\}^2 \rangle - \langle \min\{a_{ij}^\alpha, a_{ij}^\beta\} \rangle^2. \quad (5.44)$$

In the following, we will consider the rescaled quantity and the z-score in two particular cases, i.e. under the BRG model and under the BCM.

Under BRG model

Under the BRG model, it can be shown that

$$\langle \min\{a_{ij}^\alpha, a_{ij}^\beta\} \rangle_{BRG} = p^\alpha p^\beta, \quad (5.45)$$

⁵Throughout this paper, $z[X]_Y$ and $\sigma[X]_Y$ are respectively the notations for standard deviation and z score of X under the referenced null model Y .

where for each layer α , $p^\alpha = \frac{2L^\alpha}{N(N-1)}$ as in Eq. (5.20).

The expectation of $m_{bin}^{\alpha,\beta}$ under the BRG model is then

$$\langle m_{bin}^{\alpha,\beta} \rangle_{BRG} = \frac{2 \sum_i \sum_{i<j} p^\alpha p^\beta}{L^\alpha + L^\beta}. \quad (5.46)$$

Defining $\mu_{BRG}^{\alpha,\beta}$ as the rescaled multiplexity under the BRG model, we have

$$\mu_{BRG}^{\alpha,\beta} = \frac{2 \sum_i \sum_{i<j} \min(a_{ij}^\alpha, a_{ij}^\beta) - 2 \sum_i \sum_{i<j} p^\alpha p^\beta}{L^\alpha + L^\beta - 2 \sum_i \sum_{i<j} p^\alpha p^\beta}. \quad (5.47)$$

Note that, according to Gemmetto and Garleaschelli (2015), the rescaled multiplexity under the BRG model in fact can be reduced to the usual correlation coefficient between the elements of the adjacency matrix for any pair of layers α and β .

In addition, under the BRG model we have

$$\sigma[\min\{a_{ij}^\alpha, a_{ij}^\beta\}]_{BRG} = \sqrt{p^\alpha p^\beta - (p^\alpha p^\beta)^2}. \quad (5.48)$$

Therefore, we obtain the z-score of $m_{bin}^{\alpha,\beta}$ under the BRG model as

$$z_{BRG}^{\alpha,\beta} = \frac{\sum_i \sum_{i<j} \min\{a_{ij}^\alpha, a_{ij}^\beta\} - \sum_i \sum_{i<j} p^\alpha p^\beta}{\sqrt{\sum_i \sum_{i<j} [p^\alpha p^\beta - (p^\alpha p^\beta)^2]}}. \quad (5.49)$$

Under the BCM

Under the BCM, it can be shown that

$$\langle \min\{a_{ij}^\alpha, a_{ij}^\beta\} \rangle_{BCM} = p_{ij}^\alpha p_{ij}^\beta, \quad (5.50)$$

where as mentioned in Eq. (5.26), $p_{ij}^\alpha = \frac{x_i^\alpha x_j^\alpha}{1+x_i^\alpha x_j^\alpha}$ for each layer α .

That leads to the following expression for the expectation of $m_{bin}^{\alpha,\beta}$ under the BCM

$$\langle m_{bin}^{\alpha,\beta} \rangle_{BCM} = \frac{2 \sum_i \sum_{i<j} p_{ij}^\alpha p_{ij}^\beta}{L^\alpha + L^\beta}. \quad (5.51)$$

Therefore, we have

$$\mu_{BCM}^{\alpha,\beta} = \frac{2 \sum_i \sum_{i<j} \min(a_{ij}^\alpha, a_{ij}^\beta) - 2 \sum_i \sum_{i<j} p_{ij}^\alpha p_{ij}^\beta}{L^\alpha + L^\beta - 2 \sum_i \sum_{i<j} p_{ij}^\alpha p_{ij}^\beta}, \quad (5.52)$$

where $\mu_{BCM}^{\alpha,\beta}$ as the rescaled multiplexity under the BCM model.

Additionally, the standard deviation of $\min\{a_{ij}^\alpha, a_{ij}^\beta\}$ is given by

$$\sigma[\min\{a_{ij}^\alpha, a_{ij}^\beta\}]_{BCM} = \sqrt{p_{ij}^\alpha p_{ij}^\beta - (p_{ij}^\alpha p_{ij}^\beta)^2}. \quad (5.53)$$

From that, we obtain the z-score of $m_{bin}^{\alpha,\beta}$ under the BCM model as

$$z_{BCM}^{\alpha,\beta} = \frac{\sum_i \sum_{i<j} \min\{a_{ij}^\alpha, a_{ij}^\beta\} - \sum_i \sum_{i<j} p_{ij}^\alpha p_{ij}^\beta}{\sqrt{\sum_i \sum_{i<j} [p_{ij}^\alpha p_{ij}^\beta - (p_{ij}^\alpha p_{ij}^\beta)^2]}}. \quad (5.54)$$

5.2.2.2 Undirected weighted multiplexity

Raw weighted multiplexity

For undirected weighted multiplexities, similar to the binary case, we start by defining the generic weighted multiplexity as

$$m_w^{\alpha,\beta} = \frac{2 \sum_i \sum_{i<j} \min(w_{ij}^\alpha, w_{ij}^\beta)}{W^\alpha + W^\beta}. \quad (5.55)$$

The interpretation for $m_w^{\alpha,\beta}$ is similar to that for $m_{bin}^{\alpha,\beta}$, but here we take into account the role of link weights between nodes in every layer.

Rescaled weighted multiplexity

The rescaled quantity for $m_w^{\alpha,\beta}$ is defined as

$$\mu_w^{\alpha,\beta} = \frac{m_w^{\alpha,\beta} - \langle m_w^{\alpha,\beta} \rangle}{1 - \langle m_w^{\alpha,\beta} \rangle}, \quad (5.56)$$

where $\langle m_w^{\alpha,\beta} \rangle$ is the expected value under the referenced null model.

Mathematically, $\langle m_w^{\alpha,\beta} \rangle$ can generally be expressed as

$$\langle m_w^{\alpha,\beta} \rangle = \frac{2 \sum_i \sum_{i<j} \langle \min(w_{ij}^\alpha, w_{ij}^\beta) \rangle}{W^\alpha + W^\beta}. \quad (5.57)$$

Weighted z scores

Similar to the binary case, weighted z scores can tell us how many standard deviations the raw weighted multiplexity differs from the expected value under the considered null model. The general form for the z-score of weighted multiplexity (weighted z-score in short)

is defined as

$$z[m_w^{\alpha,\beta}] = \frac{m_w^{\alpha,\beta} - \langle m_w^{\alpha,\beta} \rangle}{\sigma[m_w^{\alpha,\beta}]}.$$
 (5.58)

In addition, in general, we have that

$$\sigma[m_w^{\alpha,\beta}] = \frac{\sigma[\sum_i \sum_{i<j} \min\{w_{ij}^\alpha, w_{ij}^\beta\}]}{W^\alpha + W^\beta},$$
 (5.59)

and

$$\sigma^2[\min\{w_{ij}^\alpha, w_{ij}^\beta\}] = \langle \min\{w_{ij}^\alpha, w_{ij}^\beta\}^2 \rangle - \langle \min\{w_{ij}^\alpha, w_{ij}^\beta\} \rangle^2.$$
 (5.60)

Under the WRG model

Under the WRG model, it can be shown that

$$\langle \min(w_{ij}^\alpha, w_{ij}^\beta) \rangle_{WRG} = \frac{p_w^\alpha p_w^\beta}{1 - p_w^\alpha p_w^\beta},$$
 (5.61)

where $p_w^\alpha = \frac{2W^\alpha}{N(N-1)+2W^\alpha}$ for each layer α as in Eq. (5.24).

We then obtain

$$\langle m_w^{\alpha,\beta} \rangle_{WRG} = \frac{2 \sum_i \sum_{i<j} \frac{p_w^\alpha p_w^\beta}{(1-p_w^\alpha p_w^\beta)}}{W^\alpha + W^\beta}.$$
 (5.62)

The rescaled quantity for $m_w^{\alpha,\beta}$ under the WRG models is then specified by

$$\mu_{WRG}^{\alpha,\beta} = \frac{2 \sum_i \sum_{i<j} \min\{w_{ij}^\alpha, w_{ij}^\beta\} - 2 \sum_i \sum_{i<j} \frac{p_w^\alpha p_w^\beta}{(1-p_w^\alpha p_w^\beta)}}{W^\alpha + W^\beta - 2 \sum_i \sum_{i<j} \frac{p_w^\alpha p_w^\beta}{(1-p_w^\alpha p_w^\beta)}}.$$
 (5.63)

In addition, we have

$$\sigma[\min\{w_{ij}^\alpha, w_{ij}^\beta\}]_{WRG} = \sqrt{\frac{p_w^\alpha p_w^\beta}{(1 - p_w^\alpha p_w^\beta)^2}}.$$
 (5.64)

Therefore, the z-score under the WRG model is expressed as

$$z_{WRG}^{\alpha,\beta} = \frac{\sum_i \sum_{i<j} \min\{w_{ij}^\alpha, w_{ij}^\beta\} - \sum_i \sum_{i<j} \frac{p_w^\alpha p_w^\beta}{(1-p_w^\alpha p_w^\beta)}}{\sqrt{\sum_i \sum_{i<j} [\frac{p_w^\alpha p_w^\beta}{(1-p_w^\alpha p_w^\beta)^2]}}}$$
 (5.65)

Under the WCM

Under the WCM, we have that

$$\langle \min(w_{ij}^\alpha, w_{ij}^\beta) \rangle_{WCM} = \frac{p_{ij}^\alpha p_{ij}^\beta}{1 - p_{ij}^\alpha p_{ij}^\beta}, \quad (5.66)$$

where $p_{ij}^\alpha = x_i^\alpha x_j^\alpha$ as in Eq. (5.31).

The expected multiplexity under the WCM is then given by

$$\langle m_w^{\alpha,\beta} \rangle_{WCM} = \frac{2 \sum_i \sum_{i<j} \frac{p_{ij}^\alpha p_{ij}^\beta}{(1-p_{ij}^\alpha p_{ij}^\beta)}}{W^\alpha + W^\beta}. \quad (5.67)$$

Therefore, the rescaled multiplexity in the WCM is specified as

$$\mu_{WCM}^{\alpha,\beta} = \frac{2 \sum_i \sum_{i<j} \min\{w_{ij}^\alpha, w_{ij}^\beta\} - 2 \sum_i \sum_{i<j} \frac{p_{ij}^\alpha p_{ij}^\beta}{(1-p_{ij}^\alpha p_{ij}^\beta)}}{W^\alpha + W^\beta - 2 \sum_i \sum_{i<j} \frac{p_{ij}^\alpha p_{ij}^\beta}{(1-p_{ij}^\alpha p_{ij}^\beta)}}. \quad (5.68)$$

Moreover, it can be shown that

$$\sigma[\min\{w_{ij}^\alpha, w_{ij}^\beta\}]_{WCM} = \sqrt{\frac{p_{ij}^\alpha p_{ij}^\beta}{(1 - p_{ij}^\alpha p_{ij}^\beta)^2}}. \quad (5.69)$$

We obtain the following expression for the z-score under the WCM

$$z_{WCM}^{\alpha,\beta} = \frac{\sum_i \sum_{i<j} \min\{w_{ij}^\alpha, w_{ij}^\beta\} - \sum_i \sum_{i<j} \frac{p_{ij}^\alpha p_{ij}^\beta}{(1-p_{ij}^\alpha p_{ij}^\beta)}}{\sqrt{\sum_i \sum_{i<j} \left[\frac{p_{ij}^\alpha p_{ij}^\beta}{(1-p_{ij}^\alpha p_{ij}^\beta)^2} \right]}}. \quad (5.70)$$

Under the ECM

Under the ECM, we can show that

$$\langle \min(w_{ij}^\alpha, w_{ij}^\beta) \rangle_{ECM} = \frac{p_{ij}^\alpha p_{ij}^\beta}{1 - r_{ij}^\alpha r_{ij}^\beta}, \quad (5.71)$$

where $p_{ij}^\alpha = \frac{x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha}{1 - y_i^\alpha y_j^\alpha + x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha}$ as in Eq. (5.34), and $r_{ij,\alpha} = y_i^\alpha y_j^\alpha$ as in Eq. (5.35).

The expectation of $m_w^{\alpha,\beta}$ under the ECM is then given by

$$\langle m_w^{\alpha,\beta} \rangle_{ECM} = \frac{2 \sum_i \sum_{i < j} \frac{p_{ij}^\alpha p_{ij}^\beta}{1 - r_{ij,\alpha} r_{ij,\beta}}}{W^\alpha + W^\beta}. \quad (5.72)$$

The rescaled multiplexity under the ECM is therefore equal to

$$\mu_{ECM}^{\alpha,\beta} = \frac{2 \sum_i \sum_{i < j} \min \{w_{ij}^\alpha, w_{ij}^\beta\} - 2 \sum_i \sum_{i < j} \frac{p_{ij}^\alpha p_{ij}^\beta}{1 - r_{ij,\alpha} r_{ij,\beta}}}{W^\alpha + W^\beta - 2 \sum_i \sum_{i < j} \frac{p_{ij}^\alpha p_{ij}^\beta}{(1 - r_{ij,\alpha} r_{ij,\beta})}}. \quad (5.73)$$

In addition, we can show that

$$\sigma[\min \{w_{ij}^\alpha, w_{ij}^\beta\}]_{ECM} = \sqrt{\frac{p_{ij}^\alpha p_{ij}^\beta (1 - p_{ij}^\alpha p_{ij}^\beta + r_{ij,\alpha} r_{ij,\beta})}{(1 - r_{ij,\alpha} r_{ij,\beta})^2}}. \quad (5.74)$$

We obtain the z-score under the ECM as

$$z_{ECM}^{\alpha,\beta} = \frac{\sum_i \sum_{i < j} \min \{w_{ij}^\alpha, w_{ij}^\beta\} - \sum_i \sum_{i < j} \frac{p_{ij}^\alpha p_{ij}^\beta}{(1 - r_{ij,\alpha} r_{ij,\beta})}}{\sqrt{\sum_i \sum_{i < j} \left[\frac{p_{ij}^\alpha p_{ij}^\beta (1 - p_{ij}^\alpha p_{ij}^\beta + r_{ij,\alpha} r_{ij,\beta})}{(1 - r_{ij,\alpha} r_{ij,\beta})^2} \right]}}. \quad (5.75)$$

5.3 Findings

In our study, bank-firm credit relationships are obtained from the so-called SABI database (Sistema de Análisis de Balances Ibéricos ⁶) by Bureau van Dijk. The data set consists of 193 banks and 202,691 firms in 2007 (firms located in 60 industry sectors ⁷). Our procedure for the analysis of multilayer bank-firm credit network is organized as follows. First, we will provide an overview of the original bipartite bank-firm credit network as well the bank-sector credit network. In the next step, we break down firms into sectors based on the associated industrial sector codes, and then use the projection method to derive the matrices of overlaps between banks.

Note that we treat each of 60 industrial sectors as a separate layer. Considering a system of networks rather than the aggregate of all credit relations reflects the fact that firms in different sectors may have different characteristics like sector-specific risk, firm sizes, etc., and that banks may consider the sector that a borrower is located in as one of their criteria for screening. In each layer, we obtain two versions of projection networks, the binary version indicating whether two banks lend to at least one common firm, and the weighted version showing how many firms each pair of two banks co-finances. We then investigate the overlaps and dependencies between layers based on the measures for multiplexity introduced in the previous section, for both the binary as well as weighted cases.

5.3.1 Phenomenology to the bank-firm credit network of Spain

We report some basic statistics for the diversification of banks, firms, and sectors based on their degrees in Table (5.1). The distributions of banks' degree k_{bf} , firms' degree k_{fb} , and lending relations to sectors k_{bs} are shown in Figure 5.1. We observe that the distribution of banks' degree is much wider than that of firms' degree. While many banks diversify their lending serving often more than 1000 firms, almost all firms concentrate their borrowing on 1 to 3 banks (see panels (a), (b) of Figure 5.1). In addition, lending is heterogeneously distributed across sectors, as shown in the panel (c) of Figure 5.1.

⁶See, for instance, the introduction to SABI by Bureau van Dijk, available at: <http://www.bvdinfo.com/en-gb/our-products/company-information/national-products/sabi>.

⁷Sector codes are based on "The Statistical Classification of Economic Activities in the European Community" (NACE), level 2, which is identified by two-digit numerical codes (01 to 99).

Degrees	min	max	avg.	std.
k_{bf}	1	47075	1771	5877
k_{fb}	1	8	1.68	0.93
k_{bs}	1	59	22.71	18.11
k_{sb}	1	162	73.06	40.98

Table 5.1: Basic statistics for the diversification of banks, firms, and sectors in 2007.

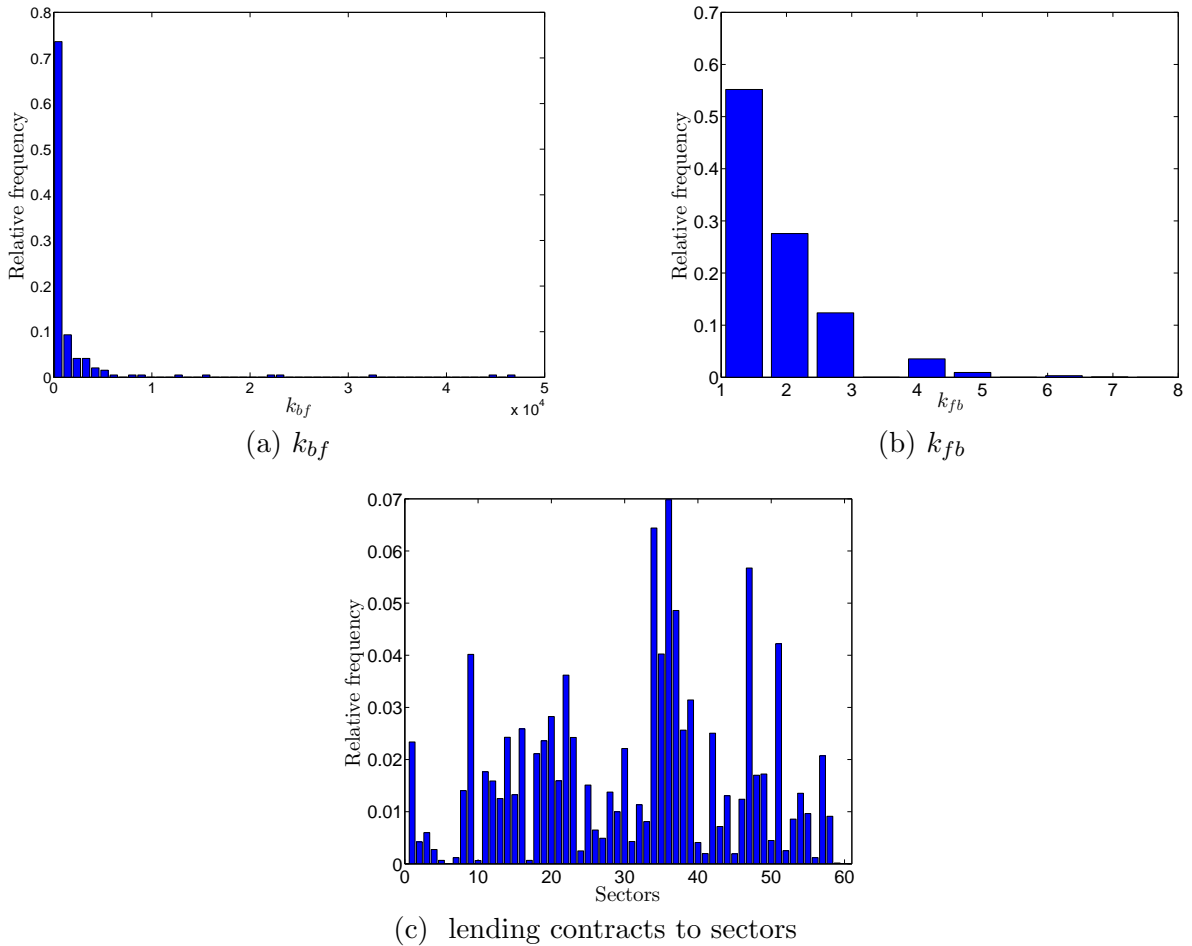


Figure 5.1: Distributions of banks' degree, firms' degree, and lending relations to sectors in 2007. Panel (a) banks' degree k_{bf} . Panel (b) firms' degree k_{fb} . Panel (c) lendings to sectors.

In Figure 5.2 (a) we exhibit the relation between $\log(k_{bf})$ and k_{bs} . It shows that a high number of firms that a bank lends to will also correspond to a high level of sector diversification of that bank ⁸. In contrast, banks with smaller degrees concentrate their

⁸Note that the estimated parameters in this Figure are just used for the purpose of illustration of the correlations between k_{bs} and k_{bf} in panel (a) as well as between k_{sb} and k_{sf} in panel (b).

lending on a limited number of sectors.

For each sector, let k_{sf} denote the number of firms in that sector that borrow from all banks ⁹. In Figure 5.2 (b), a positive correlation between $\log(k_{sf})$ and $\log(k_{sb})$ is found, suggesting that in terms of log-log scale, a large (small) number of firms borrowing from banks reveals a high (low) level of sectors' borrowing diversification.

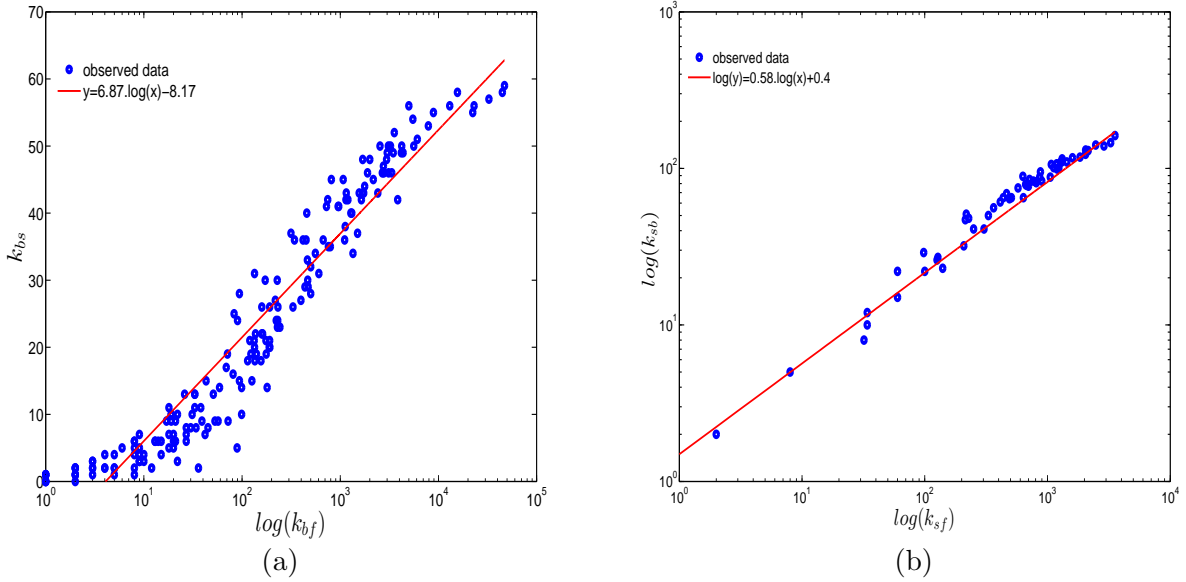


Figure 5.2: Bank's diversification and sector's diversification in 2007. Panel (a) the lending to sectors (k_{bs}) vs. the lending to firms (k_{bf}) in semilogx plot. Panel (b) the number of banks lending to each sector (k_{sb}) vs. the number of firms in that sector borrowing from banks (k_{sf}) in log-log scale.

The results show the relevance of the presence of sector specific layers in the bank-firm credit network of Spain. In the following subsections, we will focus on the overlaps and the correlations between layers in different versions (i.e. the binary and weighted versions) of the network.

5.3.2 Binary analysis

For the binary version, we show the raw binary multiplexity and its corresponding distribution respectively in Figure 5.3 (a) and Figure 5.3 (b). We find that, some of the layers

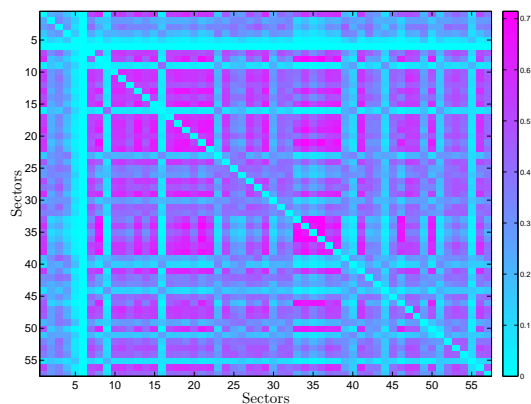
⁹Here we assume that k_{sf} , i.e. the number of firms in each sector financed by all banks, is a proxy for sectors' borrowing demand.

exhibit a high correlation in their borrowing structure while other pairs show little overlap in their lender banks.

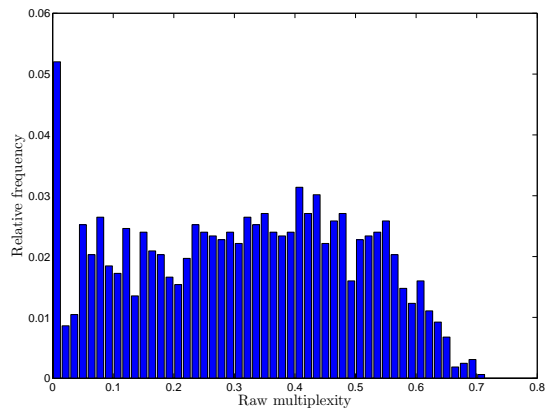
The rescaled multiplexity under the BRG model and its distribution are displayed in Figure 5.3 (c) and Figure 5.3 (d). In addition, at the bottom of Figure 5.3, in panels (e) and (f), we, respectively, display the rescaled multiplexity and the corresponding distribution under the BCM. Clearly, it is difficult to find significant differences between the rescaled quantities and the raw ones by mere inspection of these plots.

Next, since the analysis based on the color-coded multiplexity matrices does not allow any conclusion on the discrepancy of the observed values to the corresponding expected ones under the null models (e.g. Gemmetto and Garlaschelli, 2015; Gemmetto et al., 2015), we also exhibit z -scores evaluated under the BRG model and the BCM. More specifically, we plot z -scores against the rescaled multiplexity in the BRG model (Figure 5.4 (a)) and in the BCM (Figure 5.4 (b)). We observe that, in both null models, almost all z -scores are much higher than the critical value $z^*=2$. This indicates that, compared to both null models, our multilayer network in its binary version has a significant, non-random structure of correlations between layers.

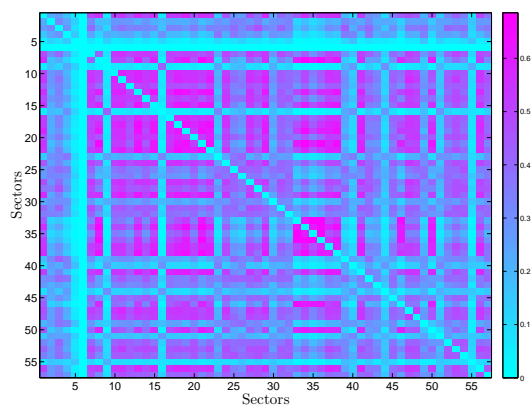
Moreover, a general positive correlation between z_{BRG} and μ_{BRG} is clearly shown in Figure 5.4 (a), meaning that high (low) μ_{BRG} corresponds to high (low) z_{BRG} . In Figure 5.4 (b), we also observe a similar correlation between z -scores and rescaled multiplexities under the BCM model. Furthermore, a majority of z_{BCM} are located in the interval $[0, 40]$, while almost all of the z_{RGM} are much larger than 40. This shows that for small values of multiplexity, the observed multiplexity deviates somewhat less on average from the expected multiplexity in the BCM than in the BRG model.



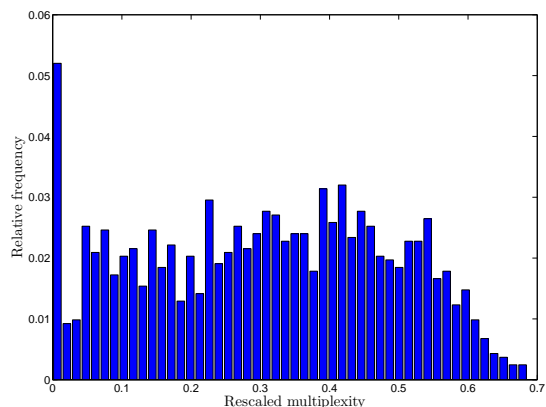
(a) m_{bin}



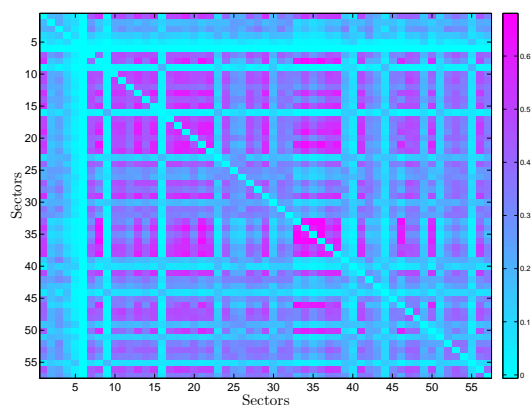
(b) Distribution of m_{bin}



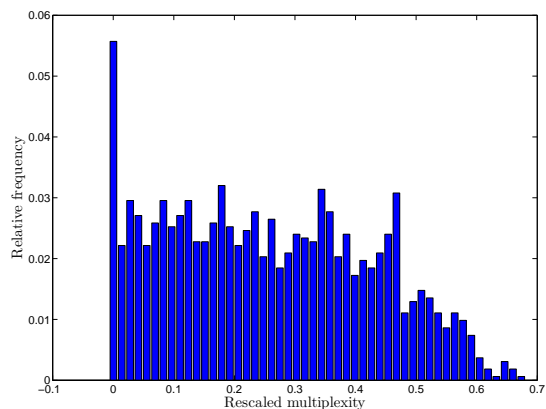
(c) μ_{BRG}



(d) Distribution of μ_{BRG}



(e) μ_{BCM}



(f) Distribution of μ_{BCM}

Figure 5.3: Binary multiplexities between layers of the bank-firm credit network in 2007. Panels (a), (b): Raw multiplexity m_{bin} and the corresponding distribution. Panels (c), (d): Rescaled multiplexity under the BRG (μ_{BRG}) and the corresponding distribution. Panels (e), (f): Rescaled multiplexity under the BCM (μ_{BCM}) and the corresponding distribution.

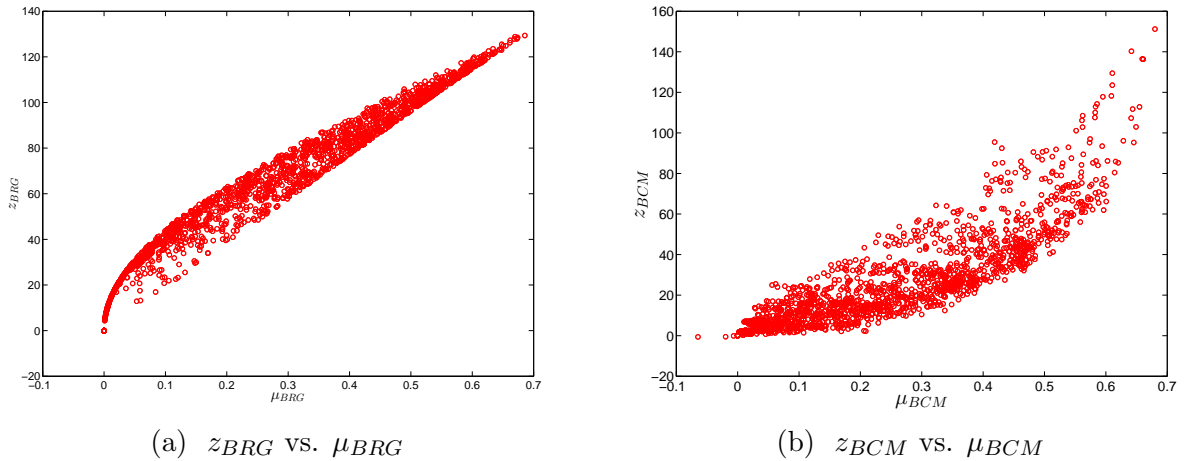


Figure 5.4: Binary z-scores in 2007. Panel (a) z_{BRG} vs. μ_{BRG} . Panel (b) z_{BCM} vs. μ_{BCM} .

5.3.3 Weighted analysis

We now move on to the analysis of the weighted version of the multilayer bank-firm credit network. To begin, in Figure 5.5, we compare the raw weighted multiplexity matrix with the raw binary multiplexity matrix, element by element. We find a tendency that levels of correlations between sectors in the weighted version are often smaller than in the binary version, indicating the role of the link intensity in weighted multiplexities.

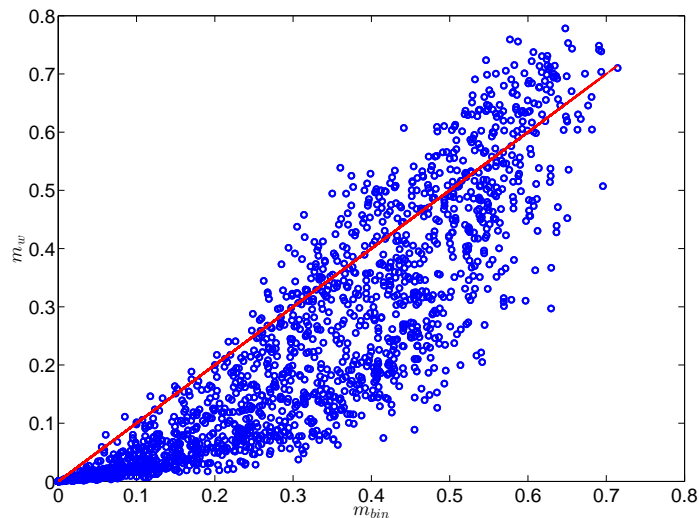


Figure 5.5: Comparison between binary and weighted multiplexities, i.e. m_w vs. m_{bin} . The red line stands for the identity line. For almost all pairs of sectors we find that $m_w < m_{bin}$.

Next, we show the color-coded matrix of the raw weighted multiplexity and the corre-

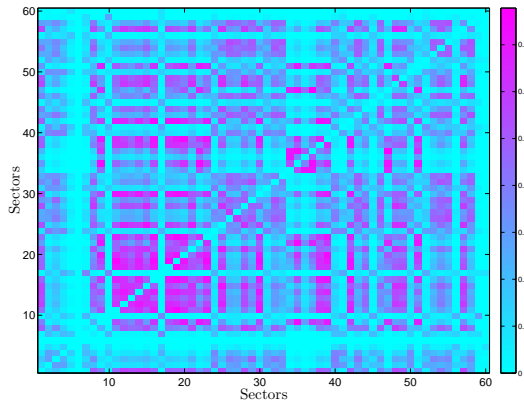
sponding distribution in the panels (a) and (b) of Figure 5.6. Similar to the binary version, we can see that the color-coded matrix as well as the distribution of the raw weighted multiplexity reveal a very hierarchical structure of correlations between layers. More specifically, on the one hand, a number of clusters of layers are highly correlated to each other, while on the other hand, many other pairs are characterized by very small (non-negative) values of multiplexity.

The rescaled multiplexities for the weighted version, i.e. μ_{WRG} (under the WRG model), μ_{WCM} (under the WCM), and μ_{ECM} (under the ECM) and their corresponding distributions are shown from panels (c) to (h) of Figure 5.6. Regarding the WRG model, we find that there is no significant difference between m_w and μ_{WRG} (see the panels (c) and (d) of Figure 5.6)), pointing out that specifying only the information about the total weight in each layer is not enough to replicate the observed correlations between layers. This result is similar to what we have obtained from the comparison between the raw multiplexity and the rescaled one under the BRG model in the binary version.

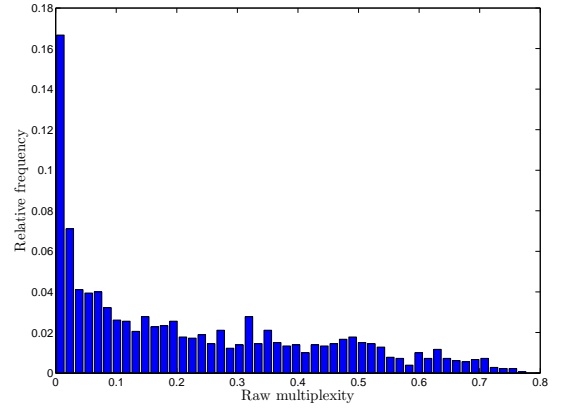
In contrast, when the strength sequence is maintained in the WCM, in comparison between Figure 5.6 (a) and Figure 5.6 (e) as well as between Figure 5.6 (b) and Figure 5.6 (f), we find that the correlations between many sectors are partially reduced in μ_{WCM} . In addition, compared to the WCM, although in the ECM the knowledge of the observed degree sequence in addition to the observed strength sequence in each layer of the observed multiplex network yields a slight improvement in replicating the elements of the weighted multiplexity matrix, significant correlations between some layers still emerge after filtering the effects of the local constraints (see Figure 5.6 (g) and Figure 5.6 (h)).

In Figure 5.7 (a), (b), and (c) we report the z-scores evaluated under the three null models in the weighted version. Similar to the relation between z_{BRG} and μ_{BRG} in the BRG model, we also observe an overall positive correlation between z_{WRG} and μ_{WRG} . This implies that high values of μ_{WRG} correspond to large discrepancies between the observed multiplexity m_w and the expected one obtained from the WRG model, and vice-versa. In contrast, under the WCM and the ECM, it becomes more difficult to detect any overall significant correlation between the elements of the rescaled multiplexity matrix and the corresponding z-scores, since now both low as well as high levels of rescaled multiplexities can have high z-scores. However, it is worthwhile to emphasize that for $\mu_{WCM} > 0.2$ or $\mu_{ECM} > 0.2$, the values of the corresponding z-scores are also typically higher than the critical value $z^*=2$.

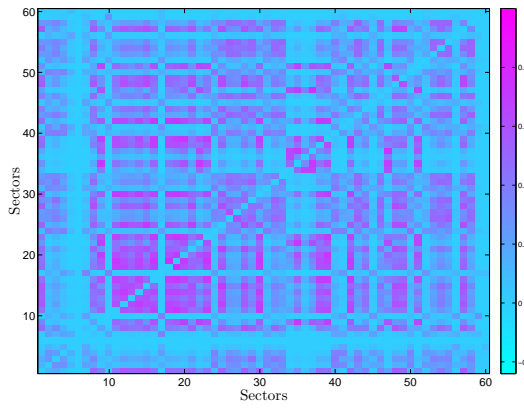
In addition, among the three null models, the absolute value of z-scores under the WRG is usually much larger than the corresponding ones evaluated under the WCM and the ECM.



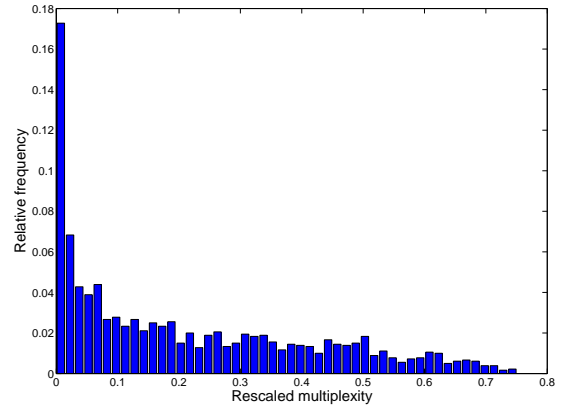
(a) m_w



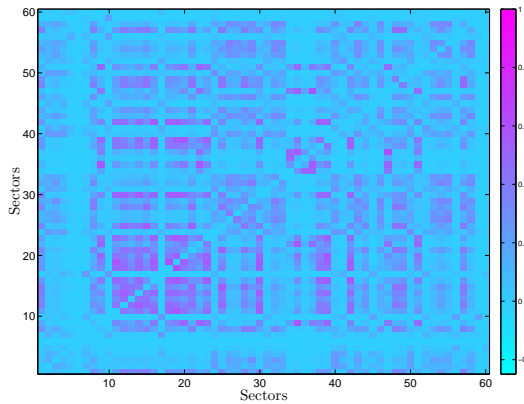
(b) Distribution of m_w



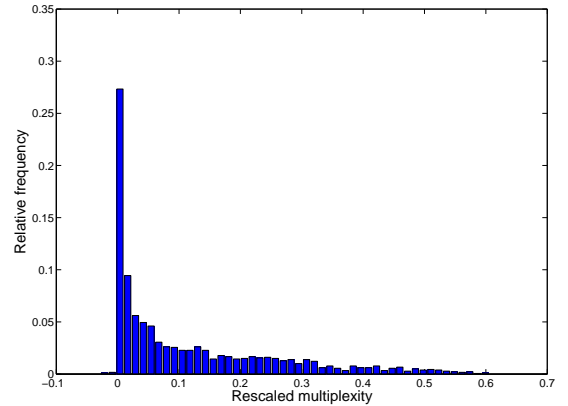
(c) μ_{WRG}



(d) Distribution of μ_{WRG}

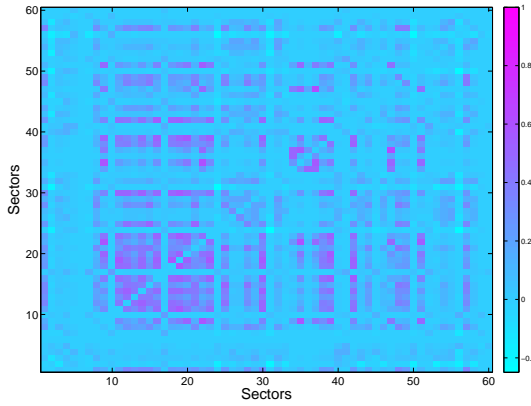


(e) μ_{WCM}

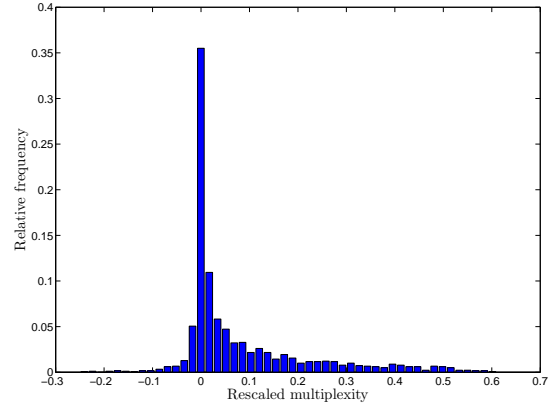


(f) Distribution of μ_{WCM}

Figure 5.6: Weighted multiplexities between layers of the bank-firm credit network of Spain in 2007. Panels (a), (b): m_w and the corresponding distribution. Panels (c), (d): μ_{WRG} and the corresponding distribution. Panels (e), (f): μ_{WCM} and the corresponding distribution. Panels (g), (h): μ_{ECM} and the corresponding distribution. (*cont.*)



(g) μ_{ECM}



(h) Distribution of μ_{ECM}

Figure 5.6: Weighted multiplexities between layers of the bank-firm credit network of Spain in 2007. Panels (a), (b): m_w and the corresponding distribution. Panels (c), (d): μ_{WRG} and the corresponding distribution. Panels (e), (f): μ_{WCM} and the corresponding distribution. Panels (g), (h): μ_{ECM} and the corresponding distribution.

In particular, we find that, on the one hand, $-1 < z_{WCM} < 10$ and $-6 < z_{ECM} < 12$; on the other hand, almost all of z_{WRG} are greater than 50. That means, overall, the observed multiplexity deviates less from the expected multiplexity under the WCM and under the ECM than under the WRG model.

We should emphasize that, in all three null models employed for the weighted version, many z-scores (especially for the ones associated with high values of the rescaled multiplexity) are larger than the critical value $z^*=2$, showing that the weighted version of the network does display a significant, non-random structure of correlations between layers.

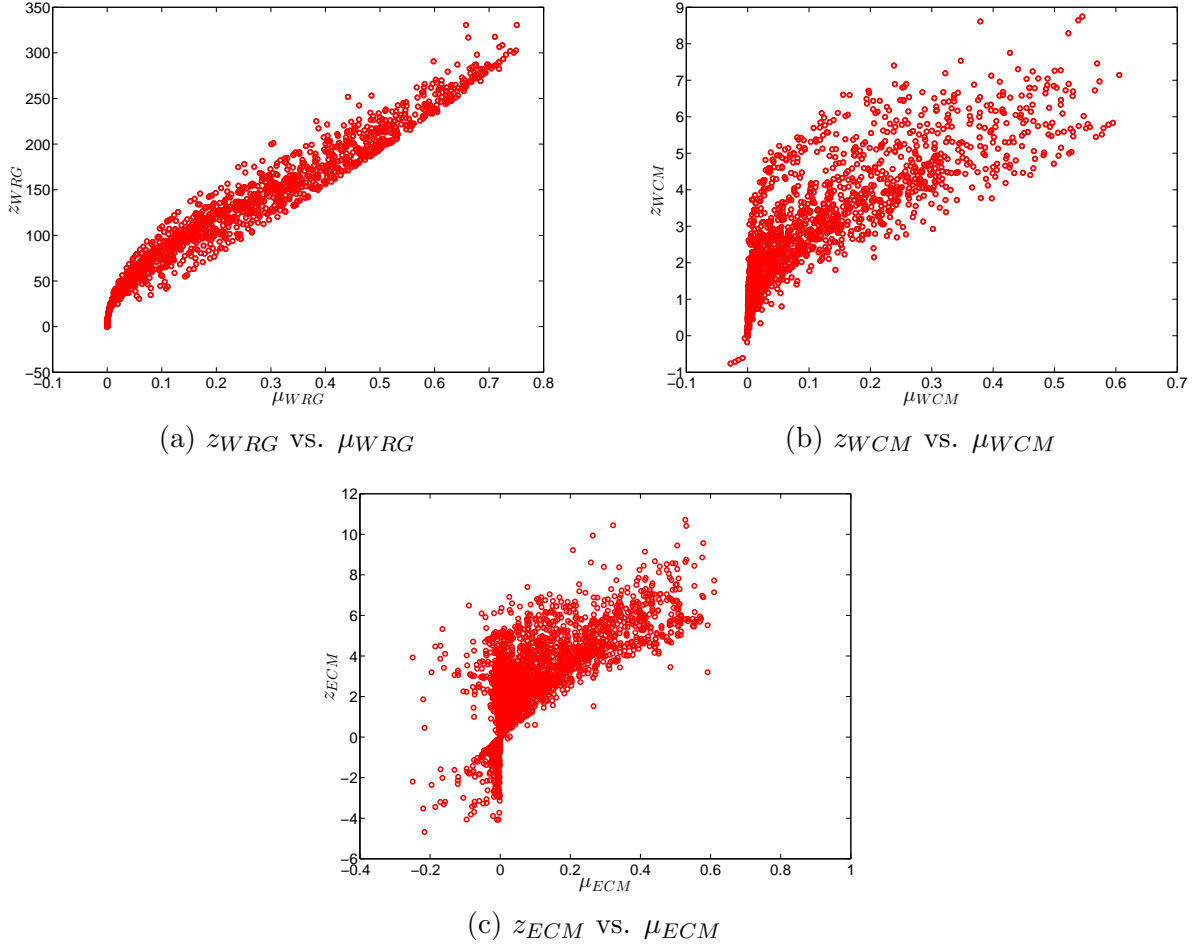


Figure 5.7: Weighted z-scores in 2007. Panel (a) z_{WRG} vs. μ_{WRG} , panel (b) z_{WCM} vs. μ_{WCM} , panel (c) z_{ECM} vs. μ_{ECM} .

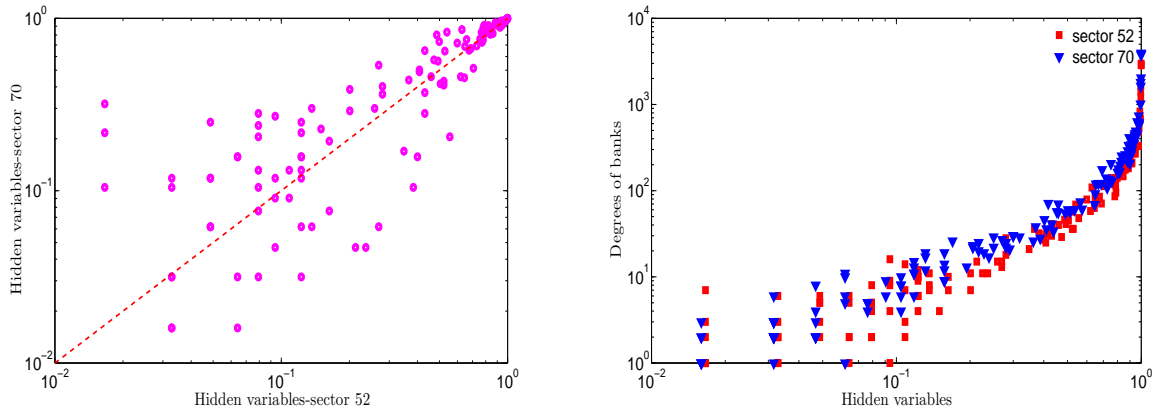
5.3.4 Hubs distribution and hidden variables

Recalling Eq. (5.26), given two nodes in any layer associated with two hidden variables (x, y) extracted under the BCM, the probability of a link between them is equal to $\frac{xy}{xy+1}$. Similarly, from Eq. (5.30), given two nodes in any layer associated with two hidden variables (z, t) extracted from the WCM, the expected weight between them is $\frac{zt}{1-zt}$. It is easy to show that $\frac{xy}{xy+1}$ is an increasing function of x and y in $(0, \infty)$, and $\frac{zt}{1-zt}$ is also an increasing function of z and t in $[0, 1)$. Therefore, in each layer, nodes associated with a higher level of hidden variables will attract more links (in the BCM) or more weights (in the WCM) from other nodes. When looking across layers, obviously if the order of hidden variables in a layer α (i.e. $\{x_i^\alpha\}_{i=1}^N$) is similar to the order of hidden variables in another layer β (i.e. $\{x_i^\beta\}_{i=1}^N$), this will lead to a higher overlap and correlation between these two layers.

In the bank-firm credit network of Spain, similar to what is observed in the case of the International Trade Network (see Gemmetto and Garlaschelli, 2015), we find that a number of pairs of sectors share the same hubs. This may partially explain why these sectors exhibit more overlap and correlation. In Figure 5.8 we show an example of the distribution of hubs in the pair of layers that have the highest value of the weighted multiplexity, i.e. the layer associated with the industrial sector coded 52 (namely layer 52) and the layer associated with the industrial sector coded 70 (namely layer 70). According to the classification based on NACE (Rev. 2), the former corresponds to the industry of warehousing and support activities for transportation, and the latter corresponds to the industry of activities of head offices and management consultancy activities.

Let $\{x_i^{52}\}_{i=1}^N$ and $\{x_i^{70}\}_{i=1}^N$ be hidden variables obtained from the WCM model in these two layers. As shown in Figure 5.8 (a), we observe that banks with higher values of x_i^{52} also tend to have higher values of x_i^{70} . We then investigate the relation between hidden variables and the degree of banks in the original bipartite structure, i.e. total number of borrowers of banks in each sector. Figure 5.8 (b) shows that in each layer, a larger value of hidden variables reveals a higher degree of banks, and vice-versa. Figure (5.9) visualizes the bank-bank projection networks in the layer 52 and the layer 70. Size and color of each node are proportional to its degree in each layer. A close inspection of the center of each panel shows that both layers share a certain number of common hubs. It should be emphasized that we also typically observe this behavior in other pairs of highly correlated layers.

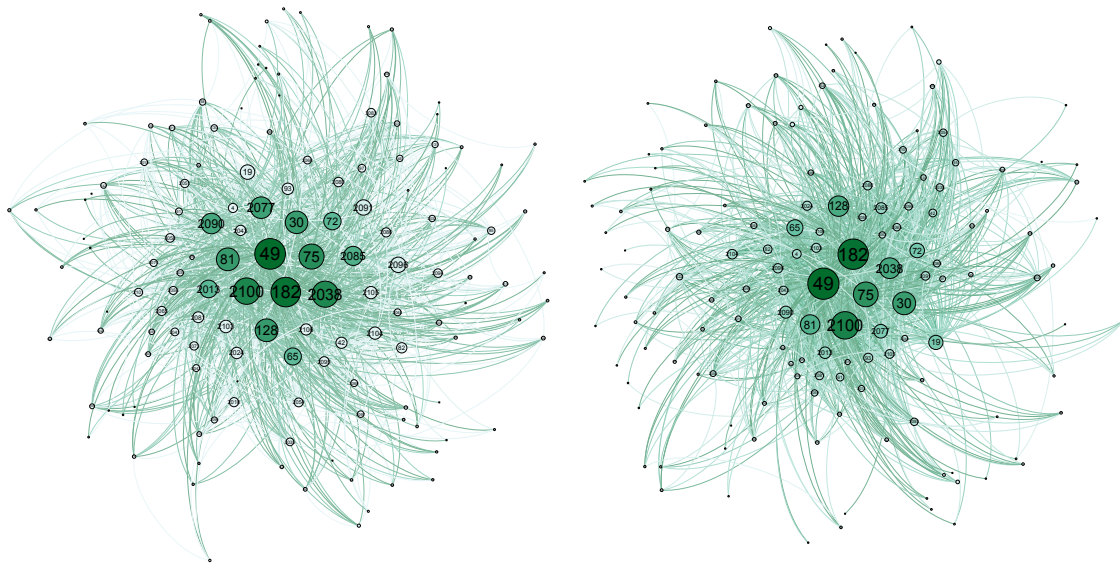
Obviously, since many significant correlations are still present after filtering the effects of the various constraints in all layers, other factors than the distribution of hubs across layers should be studied further. Such factors may be associated with higher-order topological properties of the network and/or non-topological properties of banks, sectors, and firms.



(a) Hidden variables in layers 52 and 70

(b) Hidden variables and bank degrees

Figure 5.8: Hubs and hidden variables in the two example layers (sectors) 52 and 70. Panel (a): x_i^{70} vs. x_i^{52} , where the red-dashed line stands for the identity line. Panel (b): bank degrees vs. hidden variables in the two layers 52 and 70.



(a) Layer 52

(b) Layer 70

Figure 5.9: Common hubs in the two example layers (sectors) 52 (panel a) and 70 (panel b). Nodes represent banks. Size and color of a node are proportional to its degree in that layer; the number associated with each node is the bank code in the data set.

5.4 Conclusions

We have applied recent advances in measuring overlaps and correlations between layers in multilayer networks to the bank-firm credit network of Spain, for both the binary as well as weighted versions. Each layer represents an industrial sector where firms (i.e. borrowers) are located in. We find that, first, the dependencies between layers of the observed network are highly heterogeneous. In such a hierarchical structure, layers in some clusters are highly overlapped and correlated to each other, whereas many other pairs of layers exhibit only a small level of overlaps and correlations.

Second, comparing the observed multiplexities to the expected ones obtained from the various null models maintaining different fundamental constraints in each layer, we find that there is no significant difference between the observed multiplexities and the rescaled ones filtering out the effects of the global constraints, indicating that these constraints play no significant role in the cross-layer dependencies of the observed network. When the role of local constraints is taken into account, on the one hand, in the binary version, almost all correlations are still present after subtracting the effects of the observed degree sequences in all layers. On the other hand, when filtering the effects of the heterogeneity in the observed strength and/or degree sequences in all layers, the weighted multiplexities are partially reduced. However, many significant correlations still emerge after filtering, implying that the observed network does display a non-random structure of dependencies between layers that can not simply be explained by more primitive properties of the network alone. In particular, in both versions of the network, a significant deviation from all null models is always observed in the highest level of dependencies between layers. Furthermore, interestingly, in the weighted version, many smaller values of multiplexities also strongly deviate from the expected ones evaluated under the various weighted null models.

Our study contributes to the existing literature on the role of the different fundamental constraints in the overlaps and correlations between layers of multilayer networks. Since the method used in this study is generic, we suggest that future studies should also consider to investigate the structural overlaps and dependencies in other multilayer financial networks, for instance in networks where different types of assets or financial transactions represent different layers.

Moreover, in our study, each layer is analyzed in forms of the one-mode projection network of the banking sector that it generates. Similarly, in the previous studies of Barigozzi et al. (2010), Gemmetto and Garlaschelli (2015), and Gemmetto et al. (2015), each layer is also studied in forms of the one-mode networks. In contrast, up till now, the role of constraints

in the original bipartite structure of single layers has received little attention. Since the one-mode projection network is always less informative than the original bipartite network (e.g. Lehmann et al., 2008; Luu and Lux, 2016) we believe that further research considering the bipartite structure in each layer may provide a more comprehensive analysis of the role of the different fundamental constraints in the multilayer structure of real bank-firm credit networks. In that research direction, the null models defined for bipartite networks should be employed (e.g. see Saracco et al., 2015; Di Gangi et al., 2015; Luu and Lux, 2016).

5.5 References

- Aldasoro I., Alves I. 2015. Multiplex interbank networks and systemic importance: An application to European data. SAFE Working Paper No. 102. Available at: <http://ssrn.com/abstract=2603732>.
- Bargigli L., di Iasio G., Infante L., Lillo F., Pierobon F. 2015. The multiplex structure of interbank networks. *Quantitative Finance* 15 (4), pp. 673-691.
- Barigozzi M., Fagiolo G., Garlaschelli D. 2010. Multinetwork of international trade: A commodity-specific analysis. *Physical Review E* 81 (4).
- Bianconi G. 2013. Statistical mechanics of multiplex ensembles: Entropy and overlap. *Physical Review E* 87 (6).
- Bianconi G. 2014. Multilayer networks: Dangerous liaisons?. *Nature Physics* 10, pp. 712-714.
- Boccaletti S., Bianconi G., Criado R., del Genio C. I., Gomez-Gardenes J., Romance M., Sendina-Nadal I., Wang Z., Zanin M. 2014. Structure and dynamics of multilayer networks. *Physics Reports* 544 (1), pp. 1-122.
- Caccioli F., Shrestha M., Moore C., Farmer J. D. 2014. Stability analysis of financial contagion due to overlapping portfolios. *Journal of Banking & Finance* 46, pp. 233-245.
- Chen Y. -Z., Huang Z. -G., Zhang H. -F., Eisenberg D., Seager T. P., Lai Y. -C. 2015. Extreme events in multilayer, interdependent complex networks and control. *Scientific Reports* 5.
- De Masi G., Fujiwara Y., Gallegati M., Greenwald B., Stiglitz J. E. 2011. An analysis of the Japanese credit network. *Evolutionary and Institutional Economics Review* 7 (2), pp. 209-232.
- De Masi G., Gallegati M. 2012. Bank-firms topology in Italy. *Empirical Economics* 43 (2), pp. 851-866.

Di Gangi D., Lillo F., Pirino D. 2015. Assessing systemic risk due to fire sales spillover through maximum entropy network reconstruction. Working Paper. Available at: [arXiv:1509.00607](#).

Gemmetto V., Garlaschelli D. 2015. Multiplexity versus correlation: the role of local constraints in real multiplexes. *Scientific Reports* 5.

Gemmetto V., Squartini T., Picciolo F., Ruzzenenti F., Garlaschelli D. 2015. Multiplexity and multireciprocity in directed multiplexes. Working Paper. Available at: [arXiv:1411.1282](#).

Huang X., Vodenska I., Havlin S., Stanley H. E. 2013. Cascading failures in bi-partite graphs: Model for systemic risk propagation. *Scientific Reports* 3.

Illueca M., Norden L., Udell G. F. 2013. Liberalization and risk-taking: Evidence from government-controlled banks. *Review of Finance* 18 (4), pp. 1217-1257.

Kivela M., Arenas A., Barthelemy M., Gleeson J. P., Moreno Y., Porter J. P. 2014. Multilayer networks. *J. Complex Networks* 2 (3), pp. 203-271.

Lehmann S., Schwartz M., Hansen L. K. 2008. Biclique communities. *Physical Review E* 78 (1).

Luu D. T., Lux T. 2016. Identifying patterns in the bank-sector credit network of Spain. Working Paper.

Lux T. 2016. A model of the topology of the bank-firm credit network and its role as channel of contagion. *Journal of Economic Dynamics and Control* 66, pp. 36-53.

Mastrandrea R., Squartini T., Fagiolo G., Garlaschelli D. 2014a. Enhanced reconstruction of weighted networks from strengths and degrees. *New Journal of Physics* 16.

Mastrandrea R., Squartini T., Fagiolo G., Garlaschelli D. 2014b. Reconstructing the world trade multiplex: The role of intensive and extensive biases. *Physical Review E* 90 (6).

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

Menichetti G., Remondini D., Bianconi G. 2014. Correlation between weights and overlap in ensembles of weighted multiplex networks. *Physical Review E* 90 (6).

Menichetti G., Dall'Asta L., Bianconi G. 2016. Control of multilayer networks. *Scientific Reports* 6.

Poledna S., Molina-Borboa J. L., Martínez-Jaramillo S., van der Leij M., Thurner S. 2015. The multi-layer network nature of systemic risk and its implications for the costs of financial crises. *Journal of Financial Stability* 20, pp. 70-81.

Saracco F., Clemente R. D., Gabrielli A., Squartini T. 2015. Randomizing bipartite

networks: the case of the World Trade Web. *Scientific Reports* 5.

Squartini T., Fagiolo G., Garlaschelli D. 2011a. Randomizing world trade: I. A binary network analysis. *Physical Review E* 84 (4).

Squartini T., Fagiolo G., Garlaschelli D. 2011b. Randomizing world trade: II. A weighted network analysis. *Physical Review E* 84 (4).

Squartini T., Garlaschelli D. 2011. Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics* 13.

Squartini T., Mastrandrea R., Garlaschelli D. 2015. Unbiased sampling of network ensembles. *New Journal of Physics* 17.

5.6 Appendix

In this section we will show that under the ECM, for every pair of two layers α, β , the expectations $\langle \min(w_{ij}^\alpha, w_{ij}^\beta) \rangle_{ECM}$, $\langle \min(w_{ij}^\alpha, w_{ij}^\beta)^2 \rangle_{ECM}$, and consequently the standard deviation $\sigma[\min(w_{ij}^\alpha, w_{ij}^\beta)]_{ECM}$ can analytically be calculated via hidden variables extracted from the observed degree as well as strength sequences in the two layers. That will lead to the expressions for $\langle m_w^{\alpha, \beta} \rangle_{ECM}$, $\mu_{ECM}^{\alpha, \beta}$, and $z_{ECM}^{\alpha, \beta}$ as respectively shown in Eqs. (5.72), (5.73), and (5.75) in the main text. For the proofs under other null models including the BRG model, the BCM, the WRG model, and the WCM, we refer the readers to the study of Gemmetto and Garlaschelli (2015).

First, we will start with some relevant lemmas used in our proof.

For $0 < y < 1$ we have:

Lemma 1

$$\sum_{n=0}^{\infty} y^n = \frac{1}{(1-y)}. \quad (5.76)$$

Eq. (5.76) leads to

Lemma 2

$$\sum_{n=0}^{\infty} n y^n = \frac{y}{(1-y)^2}. \quad (5.77)$$

From Eq. (5.77) we have

Lemma 3

$$\sum_{n=0}^{\infty} n^2 y^n = \frac{y^2 + y}{(1-y)^3}. \quad (5.78)$$

From Eqs. (5.76), (5.77), and (5.78) we have:

Lemma 4

$$\sum_{n=0}^{\infty} n(y^{n-1} - y^n) = \frac{1}{(1-y)}, \quad (5.79)$$

and

Lemma 5

$$\sum_{n=0}^{\infty} n^2 y^{n-1} - n^2 y^n = \frac{y+1}{(1-y)^2}. \quad (5.80)$$

As shown in the main text, in the ECM, with hidden variables $\{x_i^\alpha\}_{i=1}^N$ and $\{y_i^\alpha\}_{i=1}^N$ extracted from the observed degree and strength sequences in each layer α (see Sys. (5.32)), the probability of a link between node i and node j is

$$p_{ij}^\alpha = \frac{x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha}{1 - y_i^\alpha y_j^\alpha + x_i^\alpha x_j^\alpha y_i^\alpha y_j^\alpha} \quad (5.81)$$

and probability of a link of weight w_{ij}^α between node i and node j is

$$q(w_{ij}^\alpha) = \begin{cases} 1 - p_{ij}^\alpha, & \text{if } w_{ij}^\alpha = 0 \\ p_{ij}^\alpha r_{ij,\alpha}^{w_{ij}^\alpha - 1} (1 - r_{ij,\alpha}), & \text{if } w_{ij}^\alpha > 0 \end{cases} \quad (5.82)$$

where $r_{ij,\alpha} = y_i^\alpha y_j^\alpha$.

Considering that weights are integers, we obtain the CCDF of a link of weight w_{ij}^α as

$$q(w_{ij}^\alpha \geq w) = 1 - \sum_{t=0}^{w-1} q_{ij}(t). \quad (5.83)$$

It follows from Eq. (5.82) and Eq. (5.83) that the CCDF of a link of weight w_{ij}^α can be rewritten as

$$q(w_{ij}^\alpha \geq w) = 1 - [1 - p_{ij}^\alpha + \sum_{t=1}^{w-1} p_{ij}^\alpha (r_{ij,\alpha}^{t-1} - r_{ij,\alpha}^t)] = p_{ij}^\alpha r_{ij,\alpha}^{w-1}. \quad (5.84)$$

From (5.84), we have

$$q(\min(w_{ij}^\alpha, w_{ij}^\beta) \geq w) = q(w_{ij}^\alpha \geq w)q(w_{ij}^\beta \geq w) = (p_{ij}^\alpha r_{ij,\alpha}^{w-1})(p_{ij}^\beta r_{ij,\beta}^{w-1}). \quad (5.85)$$

Note that, the expected value of $\min(w_{ij}^\alpha, w_{ij}^\beta)$ under ECM model can be formulated as

$$\langle \min(w_{ij}^\alpha, w_{ij}^\beta) \rangle_{ECM} = \sum_{w'=0}^{\infty} w' q(\min(w_{ij}^\alpha, w_{ij}^\beta) \geq w') - w' q(\min(w_{ij}^\alpha, w_{ij}^\beta) \geq (w' + 1)). \quad (5.86)$$

From Lemma 4 (Eq. (5.79)), Eq. (5.85) and Eq. (5.86), we get

$$\langle \min(w_{ij}^\alpha, w_{ij}^\beta) \rangle_{ECM} = p_{ij}^\alpha p_{ij}^\beta \sum_{w'=0}^{\infty} w' [(r_{ij,\alpha} r_{ij,\beta})^{w'-1} - (r_{ij,\alpha} r_{ij,\beta})^{w'}] = \frac{p_{ij}^\alpha p_{ij}^\beta}{1 - r_{ij,\alpha} r_{ij,\beta}}. \quad (5.87)$$

Note that, with a similar approach we have

$$\langle \min(w_{ij}^\alpha, w_{ij}^\beta)^2 \rangle_{ECM} = \sum_{w'=0}^{\infty} w'^2 [q(\min(w_{ij}^\alpha, w_{ij}^\beta) \geq w') - q(\min(w_{ij}^\alpha, w_{ij}^\beta) \geq w' + 1)]. \quad (5.88)$$

or equivalently

$$\langle \min(w_{ij}^\alpha, w_{ij}^\beta)^2 \rangle_{ECM} = p_{ij}^\alpha p_{ij}^\beta \sum_{w'=0}^{\infty} w'^2 [(r_{ij,\alpha} r_{ij,\beta})^{w'-1} - (r_{ij,\alpha} r_{ij,\beta})^{w'}]. \quad (5.89)$$

From Lemma 5 (Eq. (5.80)) and Eq. (5.89), we obtain

$$\langle \min(w_{ij}^\alpha, w_{ij}^\beta)^2 \rangle_{ECM} = \frac{p_{ij}^\alpha p_{ij}^\beta (r_{ij,\alpha}^\alpha r_{ij,\beta}^\beta + 1)}{(1 - r_{ij,\alpha} r_{ij,\beta})^2}. \quad (5.90)$$

In addition, from Eq. (5.87) we have

$$\langle \min(w_{ij}^\alpha, w_{ij}^\beta) \rangle^2 = \left(\frac{p_{ij}^\alpha p_{ij}^\beta}{1 - r_{ij,\alpha} r_{ij,\beta}} \right)^2. \quad (5.91)$$

Therefore, from Eqs. (5.89), (5.91), the variance of $\min(w_{ij}^\alpha, w_{ij}^\beta)$ is specified as

$$\sigma(\min(w_{ij}^\alpha, w_{ij}^\beta))_{ECM}^2 = \frac{p_{ij}^\alpha p_{ij}^\beta (r_{ij,\alpha}^\alpha r_{ij,\beta}^\beta + 1)}{(1 - r_{ij,\alpha} r_{ij,\beta})^2} - \left(\frac{p_{ij}^\alpha p_{ij}^\beta}{1 - r_{ij,\alpha} r_{ij,\beta}} \right)^2 = \frac{p_{ij}^\alpha p_{ij}^\beta (1 - p_{ij}^\alpha p_{ij}^\beta + r_{ij,\alpha} r_{ij,\beta})}{(1 - r_{ij,\alpha} r_{ij,\beta})^2}. \quad (5.92)$$

Eqs. (5.87), (5.90), (5.92) imply that under the ECM, for every pair of two layers α, β ,

the expectations $\langle \min(w_{ij}^\alpha, w_{ij}^\beta) \rangle_{ECM}$, $\langle \min(w_{ij}^\alpha, w_{ij}^\beta)^2 \rangle_{ECM}$, and consequently the standard deviation $\sigma[\min(w_{ij}^\alpha, w_{ij}^\beta)]_{ECM}$ can analytically be calculated via hidden variables extracted from the observed degree as well as strength sequences in the two layers. From these results we can easily obtain the expressions for $\langle m_w^{\alpha,\beta} \rangle_{ECM}$, $\mu_{ECM}^{\alpha,\beta}$, and $z_{ECM}^{\alpha,\beta}$ as respectively shown in Eqs. (5.72), (5.73), and (5.75) in the main text.

Chapter 6

The Structure of the Cross-Correlation Matrices of Banks’ Loan Portfolios in the Bank-Firm Credit Market of Japan

Keywords: Credit Market; Portfolio Correlations; Random Matrix Theory.

6.1 Introduction

Over the last few years, catastrophic cascade of failures in interdependent networks has received a remarkable attention in network science (e.g. Smart et al., 2008; Buldyrev et al., 2010; Huang et al., 2011; Brummitt et al., 2012; Reis et al., 2014; Liu et al., 2016). One of the most important results of this research direction is that the robustness of a complex network depends on its internal structure as well as its pattern of connections to other interdependent networks.

In the banking system, the recent financial turmoil also shows that not only direct relations but also indirect relations between banks are suggested as the relevant channels of financial contagion. For instance, overlapping portfolios due to common asset holdings are widely considered as a major source of risk contagion (e.g. May and Arinaminpathy, 2010; Beale et al., 2011; Corsi et al., 2013; Huang et al., 2013; Caccioli et al., 2014; Lillo and Pirino, 2015; Greenwood et al., 2015; Levy-Carciente et al., 2015).

Another similar channel via the bank-firm credit relationships is also suggested as an

important channel of the propagation of financial distress between the financial and the real sectors of the economy (e.g. Aoyama et al., 2013; de Castro Miranda and Tabak, 2013; Aoyama, 2014; Lux, 2016). Accordingly, since banks can be indirectly linked through a set of joint exposures to firms, the distress originating from a group of banks or firms can be propagated through the whole system. In addition, from the banks' side, one of the most crucial issues in the study of this channel is to analyze the dynamics of correlations (or co-movements) between banks in lending to firms. That is because, from the macroeconomic perspective, changes in the credit supply of banks to firms have been widely suggested to have important impacts on macroeconomic fluctuations (e.g. Gertler and Kiyotaki, 2010; Jordà et al., 2013; Bassett et al., 2014; Gambetti and Musso, 2016). Furthermore, with regard to bank managers, the understanding of correlations between banks is also important to the portfolio analysis and the risk management. However, so far, the key drivers of such correlations still remain a puzzling mystery.

This study contributes to a deeper understanding of the structure of the cross-correlation matrix of banks' loan portfolios. In particular, we aim to address the following main research questions that have not been studied intensively yet: (i) how to detect non-random patterns and latent factors (if they exist) in the empirical cross-correlation matrix of banks' loan portfolios; and (ii) how to identify different states of temporal dynamics of the correlation matrix. Using a detailed data set of the bank-firm credit market of Japan, we consider three separate maturity-based layers, i.e. the short-term lending, the long-term lending, and the total lending to firms. Furthermore, we also compare the analysis of the weighted data with the analysis of the binary data to examine whether the binary and weighted relations between banks and firms share similar information about the structure and the dynamics of cross-correlations between banks.

In terms of methodology, while a number of empirical studies on the bank-firm credit market of Japan are based on network theory (e.g. Fujiwara et al., 2009; De Masi et al., 2011; Marotta et al., 2015; Fricke, 2016), in our work we will employ the methods of Random Matrix Theory (RMT), which have been widely used in the analysis of the empirical cross-correlation matrix of stock price changes (e.g. Laloux et al., 1999; Plerou et al., 1999; Laloux et al., 2000; Plerou et al., 2000; Plerou et al., 2002; Kim and Jeong, 2005; Bouchaud and Potters, 2009; Meng et al., 2014; Jiang et al., 2014; Uechi et al., 2015; MacMahon and Garlaschelli, 2015) and recently in the study of the empirical correlations between economic variables (e.g. Ormerod, 2008; Iyetomi et al., 2011; Kichikawa et al., 2015; Yoshikawa et al., 2015; Luu et al., 2016). Practically, we will compare the eigenvalue spectrum of the

empirical cross-correlation matrix with the spectrum predicted by RMT. Focusing on a group of largest eigenvalues deviating from random bulk predicted by RMT, we then extract latent information from the corresponding eigenvectors. In addition, analyzing the evolution of these eigenvalues and eigenvectors allow us to identify different states in the market. Perhaps, among several studies on the bank-firm credit relations, the work of Fujiwara et al. (2009) is closest to our study. Based on eigenvalue analysis, Fujiwara et al. (2009) measure the fragility scores in terms of dynamical propagation of influences. However, it should be emphasized that the main difference to the present paper is, instead of analyzing the eigenspectrum of the cross-correlation matrix, Fujiwara et al. (2009) define a new weight matrix P capturing the accumulated reflections and influences between two modes. In such a setting, the largest eigenvalue is always equal to 1, and few top largest eigenvalues, i.e. the second and the next largest eigenvalues of P are important in the propagation of influences from banks to firms and vice versa over a finite time-scale ¹.

The remainder of this paper is structured as follows. In Sec. 6.2 we briefly describe the data and methods employed in our study. Sec. 6.3 reports our main findings. Discussions and concluding remarks are in Sec. 6.4. In the Appendix, we provide additional results for the distribution and the evolution of the eigenvector components.

6.2 Data and Methods

6.2.1 Data

We use a large data set containing detailed credit relations between banks and large firms in Japan over the period from 1980 to 2012 ². Based on the maturity of loans, three layers of lending relationships from banks to firms are considered, i.e. total lending relationships (namely layer 1), short-term lending relationships (namely layer 2), and long-term lending relationships (namely layer 3). Note that, in our study, long-term lending relationships consist of all contracts that exceed 1 year.

¹In their study, briefly, from the matrix of weights W^{bf} representing loans from banks to firms defined in the next part of the present study, they obtain a matrix A that stands for relative amounts of lending by banks to firms and a matrix B that stands for the relative amount of borrowing by firms from banks. Then the matrix $P = AB$ will capture the accumulated reflections and influences between banks and firms. Assume that $\lambda_1^{(P)} \geq \lambda_2^{(P)} \geq \dots \geq \lambda_N^{(P)}$ are the eigenvalues of P , it can be shown that $\lambda_1^{(P)} = 1$. Therefore, over a finite time-scale, only few next largest eigenvalues $\lambda_2^{(P)} \dots$ are important in the dynamics of the propagation of influences from banks to firms and vice versa. For more details, we refer readers to the study of Fujiwara et al. (2009).

²For a more detailed explanation about the data set, we refer readers to, for instance Fujiwara et al. (2009), Marotta et al. (2015).

For the sake of conciseness, we exclude other financial institutions including insurance companies and a small number of aggregated banks^{3 4}. In addition, we should emphasize that, in each year, for each maturity-based layer, we focus on the aggregated loans from each bank to each firm, meaning that we ignore contract-specific frequencies (i.e. how many times a bank lend to a firm in each year).

6.2.2 Methods

Network and Similarity Matrices

To begin, we will briefly define some basic notations and definitions of a bank-firm credit network. Assume that, in each year we have a weighted bipartite network consisting of N banks lending to NF firms. Let $W^{bf} = \{w_{ij}^{bf}\}_{N \times NF}$ be the matrix of weights (where i runs from 1 to N standing for N banks, and j runs from 1 to NF standing for NF firms), the value of w_{ij}^{bf} is the amount of loan that bank i lends to firm j (can be the long term loan, or the short term loan, or the total loan). From W^{bf} , we obtain the adjacency matrix $A^{bf} = \{a_{ij}^{bf}\}_{N \times NF}$, where $a_{ij}^{bf} = 1$ if $w_{ij}^{bf} > 0$, and zero otherwise.

From the perspective of network theory, one can measure the structural similarity between banks' loan portfolios based on various measures of the similarity between two vectors (e.g. Pool et al., 2015; Alvarez-Socorro et al. 2015; Cai et al., 2016; Luu and Lux, 2016a; Luu, 2016; Frick, 2016). The main idea of this network-based approach is that, for each pair of two nodes in the same mode in the bipartite network, the similarity between them is based on the similarity of their neighborhoods. For instance, in the weighted version, we can, respectively, define the similarity matrices of banks and of firms based on the generalized Jaccard index as

$$S_{ij}^{B-B} = \frac{\sum_k \min(w_{ik}^{bf}, w_{jk}^{bf})}{\sum_k \max(w_{ik}^{bf}, w_{jk}^{bf})}, (\forall 1 \leq i \neq j \leq N), \quad (6.1)$$

and

$$S_{ij}^{F-F} = \frac{\sum_k \min(w_{ki}^{bf}, w_{kj}^{bf})}{\sum_k \max(w_{ki}^{bf}, w_{kj}^{bf})}, (\forall 1 \leq i \neq j \leq NF). \quad (6.2)$$

It can easily be shown that S_{ij}^{B-B} and S_{ij}^{F-F} range in $[0, 1]$, and $S_{ij}^{B-B} = 1$ (or $S_{ij}^{F-F} = 1$) if and only if $w_{ik}^{bf} = w_{jk}^{bf}$ for all firms k (or $w_{ki}^{bf} = w_{kj}^{bf}$ for all banks k). Furthermore, $S_{ij}^{B-B} = 0$ (or $S_{ij}^{F-F} = 0$) if and only if the portfolios of the two banks i and j (or of the two firms i and j) have no overlap at all. Therefore, each element of the matrix S^{B-B} (or S^{F-F}) will

³Each aggregated bank consists of a group of small/ local banks.

⁴However, the main conclusions do not change very much when we include insurance companies and aggregated banks in the data, since the market shares of these companies and banks are small.

show the degree of similarity between a pairs of banks' loan portfolios (or a pairs of firms' borrowing portfolios).

In addition, the one-mode projection matrices of a bank-firm credit network also indicate the overlapping portfolios between banks or between firms (e.g. see Luu and Lux, 2016a; Luu and Lux, 2016b). More specifically, defining

$$A^{B-B} = A^{bf} A^{fb}, \quad (6.3)$$

we will obtain a matrix in which each off-diagonal elements is nothing else but the number of firms that a pair of banks co-finances ⁵. In a similar way, each off-diagonal element of the A^{F-F} matrix, with

$$A^{F-F} = A^{fb} A^{bf}, \quad (6.4)$$

indicates the number of common banks that a pair of two firms borrows from. A limitation of the network projection approach is that the information about the value of loans from banks to firms is partially or completely discarded.

Correlation matrix

We now introduce the method to measure the empirical cross-correlation matrix of banks' loan portfolios. Similar to the procedure often implemented in the analysis of stock prices, the normalized amount of loan from bank i to firm j is defined as

$$X_{i,j} = \frac{w_{i,j}^{bf} - \langle w_{i,j}^{bf} \rangle}{\sigma_{i,j}}, \quad (6.5)$$

where for each bank i , $\langle w_{i,j}^{bf} \rangle$ and $\sigma_{i,j}$ are respectively the average and the standard deviation of $w_{i,j}^{bf}$ over all NF firms.

Denote $X = \{X_{i,j}\}_{N \times NF}$, the cross-correlation matrix of N banks is defined as

$$C = \{C_{ij}\}_{N \times N} = \frac{1}{NF} X X^T, \quad (6.6)$$

where X^T is the transposition of X .

From an economic point of view, the value of C_{ij} denotes the correlation between bank i and bank j in lending to all firms. Note that, $-1 < C_{ij} < 1$, for $1 \leq i < j \leq N$, and for all i we always have $C_{ii} = 1$. In addition, $C_{ij} > 0$ (< 0) indicates that two banks i and j are positively (negatively) correlated, while $C_{ij} = 0$ shows there is no correlation between the

⁵The matrix A^{fb} is defined as the transposition of A^{bf} .

two banks at all.

Random Matrix Theory

The methods of RMT have been widely used to extract latent information embedded in the empirical financial correlations (e.g. Laloux et al., 1999; Laloux et al., 2000; Plerou et al., 2002; Kim and Jeong, 2005; Bouchaud and Potters, 2009; Wang et al., 2011; Meng et al., 2014; Jiang et al., 2014; Uechi et al., 2015; MacMahon and Garlaschelli, 2015), and recently in empirical economic correlations (e.g. Ormerod, 2008; Iyetomi et al., 2011; Kichikawa et al., 2015; Yoshikawa et al., 2015; Luu et al., 2016). In fact, a number of studies have also been devoted to the applications of RMT in the analysis of various types of complex networks such as one-mode networks, bipartite networks (e.g. Bandyopadhyay and Jalan, 2007; de Carvalho et al., 2009; Tran et al., 2013; Dumitriu and Johnson, 2016).

In the following, we will introduce the relevant results of RMT which will be used latter to analyze the cross-correlations matrix defined in Eq. (6.6). Suppose that $\{\lambda_i\}_{i=1}^N$ are the eigenvalues of a correlation matrix C, and $\rho(\lambda)$ is then the density of the eigenvalues of C, given by

$$\rho(\lambda) = \frac{dn(\lambda)}{d(\lambda)}, \quad (6.7)$$

where $n(\lambda)$ is the number of eigenvalues of C less than λ .

As suggested by RMT, if $X \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$, in the limit $N, NF \rightarrow \infty$ with $Q = \frac{NF}{N} \rightarrow a = \text{constant} > 1$, the eigenvalue distribution $\rho(\lambda)$ (i.e. the Marchenko-Pastur distribution) is analytically given by

$$\rho(\lambda) = \begin{cases} \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max}^{RMT} - \lambda)(\lambda - \lambda_{min}^{RMT})}}{\lambda} & \text{if } \lambda_{min}^{RMT} \leq \lambda \leq \lambda_{max}^{RMT}, \\ 0 & \text{elsewhere,} \end{cases} \quad (6.8)$$

where σ^2 is the variance of the elements of X (the variance is equal to 1 in our case), and λ_{max}^{RMT} and λ_{min}^{RMT} are respectively the upper and lower bounds of eigenvalues predicted by RMT, which are given by

$$\lambda_{max}^{RMT} = (1 + \sqrt{1/Q})^2, \text{ and } \lambda_{min}^{RMT} = (1 - \sqrt{1/Q})^2. \quad (6.9)$$

Generally speaking, if some of the eigenvalues of the empirical cross-correlation matrix are beyond the interval $[\lambda_{min}^{RMT}, \lambda_{max}^{RMT}]$, they carry genuine information about the correlations between banks (e.g. Plerou et al., 2000; Plerou et al., 2002). We now introduce one of the most important implications of RMT, i.e. how to extract latent information from largest

eigenvalues deviating from the random bulk.

Without loss of generality, we assume that $\lambda_1 > \dots > \lambda_k > \lambda_{max}^{RMT} > \dots > \lambda_N$ are the eigenvalues of the empirical correlation matrix C , and u_1, u_2, \dots, u_N are their corresponding eigenvectors. Mathematically, the correlation matrix C can be diagonalized as

$$C = U\Lambda U^T, \quad (6.10)$$

where $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_N\}$, and $\{U\}_{N \times N}$ is an orthonormal matrix, whose i^{th} column is the normalized eigenvector u_i associated with λ_i . From Eq. (6.10), we obtain

$$\lambda_i = u_i^T C u_i = u_i^T \text{Cov}(X) u_i = \text{Var}(u_i^T X). \quad (6.11)$$

Let $y_i = u_i^T X$, the total variance of $\{X_i\}_{i=1}^N$ is then

$$\sum_{i=1}^N \text{Var}(X_i) = N = \sum_{i=1}^N \lambda_i = \sum_{i=1}^N \text{Var}(y_i). \quad (6.12)$$

Eqs. (6.11) and (6.12) show that $\frac{\lambda_i}{N}$ indicates the percentage of the total variance $\sum_{i=1}^N \text{Var}(X_i)$ explained by the principal component $y_{i,t} = u_i^T X$ (e.g. see Jolliffe, 1986).

In addition, Eq. (6.10) can be rewritten as

$$C = \sum_{i=1}^{i=N} \lambda_i u_i u_i^T. \quad (6.13)$$

The so-called ‘market’ mode, i.e. the common factor influencing all banks, can be extracted from the largest eigenvalue and its eigenvector (e.g. Plerou et al, 2001; Plerou et al., 2002). Its effect on C is represented by

$$C^m = \lambda_1 u_1 u_1^T, \quad (6.14)$$

and the rest of C is

$$FC^m = C - C^m, \quad (6.15)$$

which is the filtered matrix after the market-wide effect is subtracted.

Obviously, genuine information can still be extracted from the next largest eigenvalues if they are still comparably larger than the theoretical upper bound λ_{max}^{RMT} predicted by RMT. Such information may be associated with the level of sub-groups (namely the ‘group’ mode) (e.g. Kim and Jeong, 2005; Shen and Zheng, 2009; Jiang and Zheng, 2012). In that case,

following Kim and Jeong (2005), we can decompose C into three parts C^m , C^g , and C^r as

$$C = C^m + C^g + C^r, \quad (6.16)$$

where C^g and C^r respectively indicate the effects of the ‘group’ mode and the rest of C (e.g. random component). Mathematically,

$$C^g = \sum_{i=2}^{i=N_g} \lambda_i u_i u_i^T, \quad (6.17)$$

and the rest

$$C^r = \sum_{i=N_g+1}^{i=N} \lambda_i u_i u_i^T, \quad (6.18)$$

where N_g is determined based on the number of the next largest eigenvalues larger than λ_{max}^{RMT} (e.g. see Kim and Jeong (2005), MacMahon and Garlaschelli (2015) for further discussions about the selection of N_g). Let $FC^{m,g} = C^r$ be the the filtered matrix after the effects of the ‘market’ and ‘group’ modes are subtracted.

Absorption ratio

Recalling from Eq. (6.12) that $\sum_{i=1}^N \text{Var}(X_i) = N = \sum_{i=1}^N \lambda_i = \sum_{i=1}^N \text{Var}(y_i)$, we define the absorption ratios as

$$E_i = \frac{\sum_{j=1}^i \lambda_j}{N} \quad (\text{for } i = 1, 2, \dots, N), \quad (6.19)$$

which is equal to the fraction of the total variance of $\{X_i\}_{i=1}^N$ explained by the first i principal components.

Since $E_N = 1$ stands for 100% of the total variance as shown in Eq. (6.12), comparing E_1, \dots, E_k (where k is the largest integer that $\lambda_k > \lambda_{max}^{RMT}$) to 1 allows us to assess the contribution of the top largest eigenvalues to the total variance. As another application of the absorption ratios, for instance, we can use the absorption ratios to analyze the systemic risk in the market, for example, if E_1 accounts for most of E_N when λ_1 increases, it indicates a higher level of systemic risk in the market (e.g. Pukthuanthong and Roll, 2009; Billio et al., 2012; Zheng et al., 2012; Meng et al., 2014).

Inverse Participation Ratio

The Inverse Participation Ratio (IPR), an indicator of the contribution to the eigenvectors

of the correlation matrix (e.g. Plerou et al., 1999; Plerou et al., 2002), is defined as

$$\text{IPR}_i = \sum_{j=1}^N u_i(j)^4. \quad (6.20)$$

The inverse of each IPR indicates the number of eigenvector components that contribute significantly to the associated eigenvector. More specifically, a low IPR implies that banks contribute more equally. In contrast, a large IPR points out that a smaller number of banks dominate in the eigenvector. Since in the normalized version, for every i , $\sum_{j=1}^N u_i(j)^2 = 1$ ($\forall i = 1, 2, \dots, N$). Consequently, IPR_i has two limiting cases, i.e.

$$1 \geq \text{IPR}_i \geq \frac{1}{N}, \quad \forall i = 1, 2, \dots, N, \quad (6.21)$$

where $\text{IPR}_i = \frac{1}{N}$ if and only if $u_i(j) = \sqrt{\frac{1}{N}}$ for all $j = 1, 2, \dots, N$; while $\text{IPR}_i = 1$ if and only if only one element contributes the eigenvector u_i and all other elements must equal to zero.

6.3 Findings

To begin, we first briefly summarize the results of the network-based analysis of loan portfolio overlaps and portfolio similarities between banks. Selecting 2012 as an example, in Figure (6.1) (a), (b), and (c), we observe that many banks have high overlaps in their loan portfolios when lending to non-financial firms, especially in the layer 1 and layer 3. Furthermore, the analysis of the weighted portfolio similarities based on the generalized Jaccard index shows that some pairs of banks exhibit a high level of similarity in their loan structure, while other pairs display only a low level of similarity or no similarity at all (see Figure (6.1) (d), (e), and (f))⁶. This indicates the effect of the heterogeneity in the link intensity between banks and firms in the weighted similarity matrix.

Next, we focus on the analysis of the cross-correlation matrix of banks' loan portfolios. We start by analyzing the temporal dynamics of the fundamental statistics of the elements of the correlation matrices (C_{ij}) in three layers over the period from 1980 to 2012. We summarize the main results in Figure (6.2). The time-varying distributions of C_{ij} in different layers are

⁶In the the analysis of weighted portfolio similarities, for the sake of convenience, we use the same order of banks as in the analysis of binary overlaps in Figure (6.1) (a), (b), and (c), i.e. banks are sorted in the descending order of the eigenvector centrality of the adjacency matrix of the bank-bank projection network obtained from A^{B-B} . We do not conclude any nature of the relationship between that order and the elements of the weighted portfolio matrix S^{B-B} .

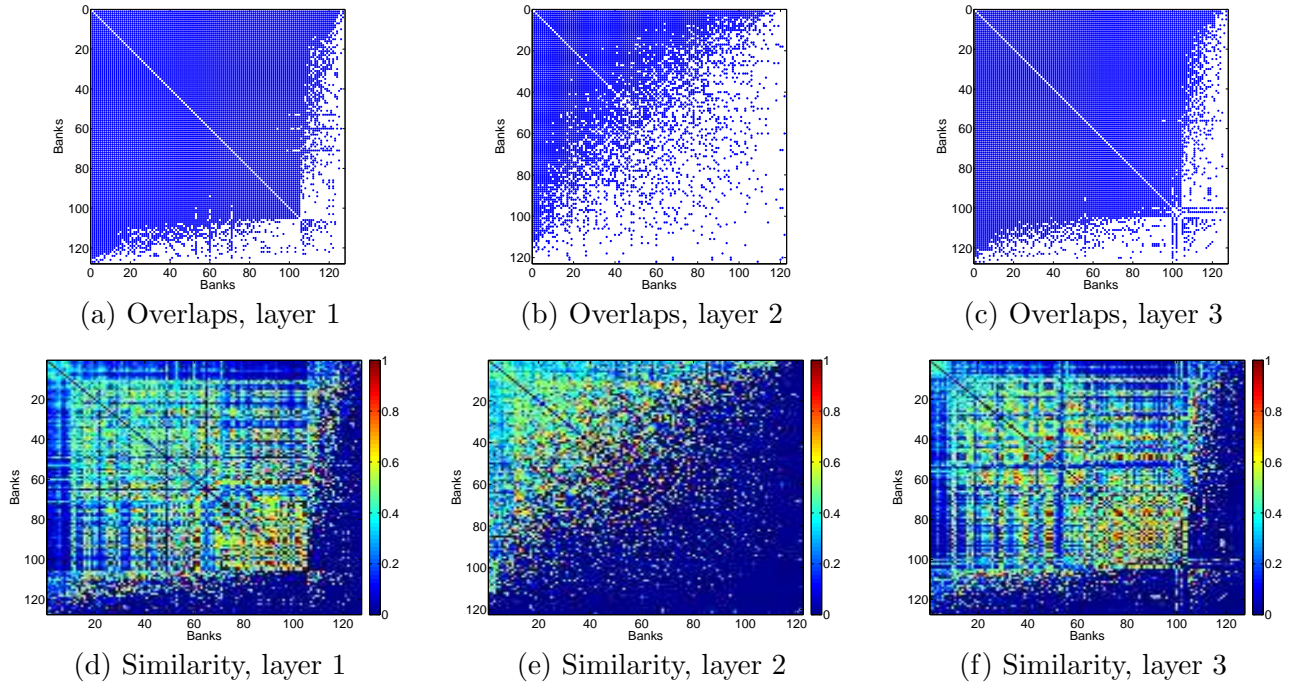


Figure 6.1: Binary overlaps and portfolio similarities between banks in the three layers, in 2012. Banks are sorted in the descending order of the eigenvector centrality of the adjacency matrix of the bank-bank projection network obtained from A^{B-B} . Panels (a) to (c): The binary bank-bank projection matrices in three layers. Panels (d) to (f): The similarity matrices in three layers.

shown in panels (a) to (c), while the fundamental statistics are shown in the remaining panels. As shown in the panels (d) to (f) of Figure (6.2), overall, we observe that the averages of correlations in all layers are always small and positive. In addition, these distributions exhibit positive skewness (panels (g) to (i)) and high kurtosis (panels (j) to (l)). Furthermore, when aggregating short-term loans (i.e. layer 2) with long-term loans (i.e. layer 3), we observe that the statistical properties of the elements of the cross-correlation matrix of total loans (i.e. layer 1) are mainly driven by those of long-term loans.

To investigate whether patterns can be detected in the correlations between banks' loan portfolios, in each layer, we decompose the empirical correlation matrix and then calculate its eigenvalues and eigenvectors. After that, we will compare the distribution of eigenvalues of the empirical correlation matrix C with the one predicted by RMT. In Figure (6.3) we can see that a small group of eigenvalues (typically from 10 % to 20 % of eigenvalues) is always larger than the upper bound λ_{max}^{RMT} . In addition, in the rest, although only less than half of the eigenvalues lie between $[\lambda_{min}^{RMT}, \lambda_{max}^{RMT}]$, the contribution of the smallest eigenvalues (which are less than the lower bound λ_{min}^{RMT}) to the total variance is relatively small.

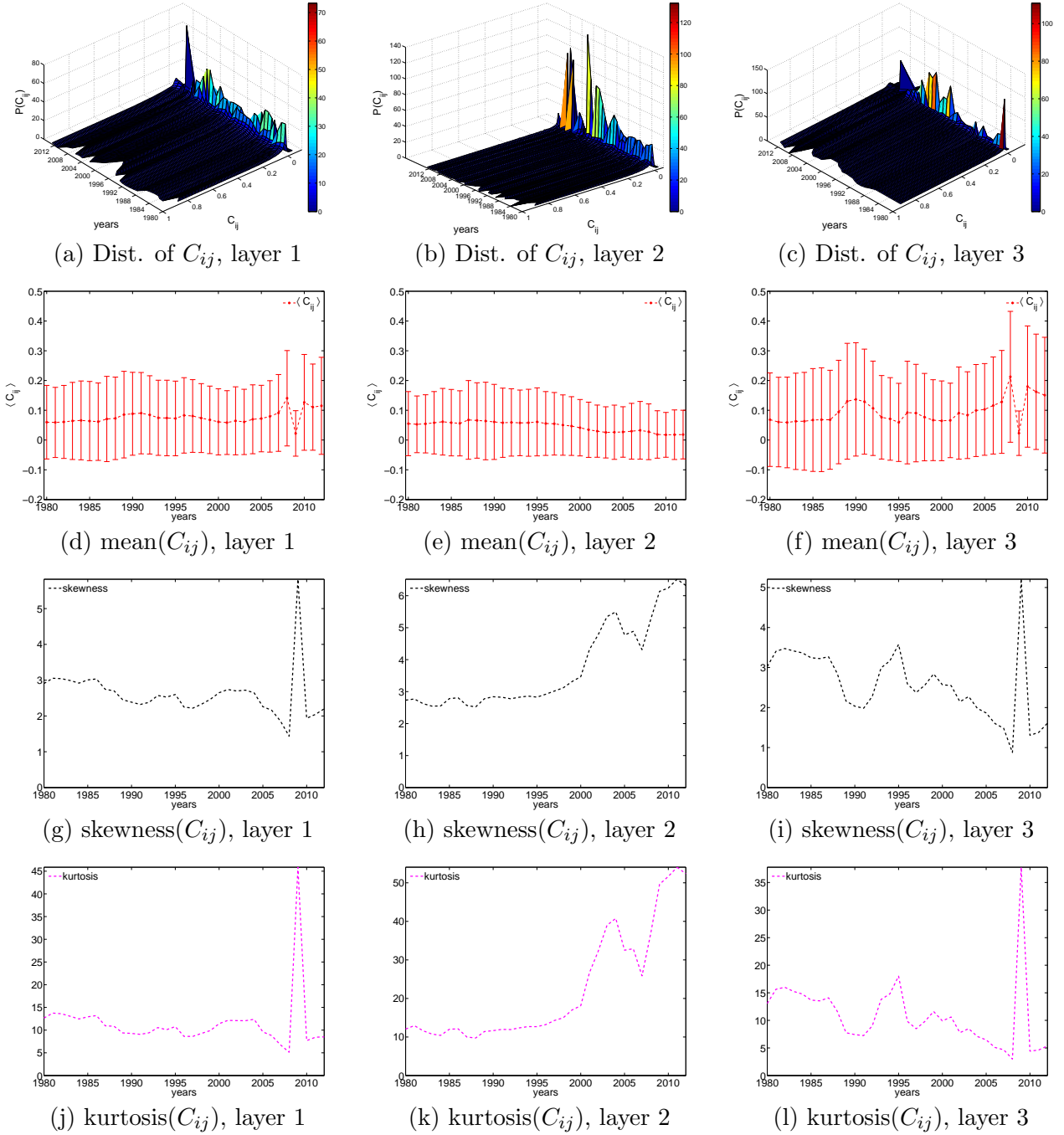


Figure 6.2: Temporal evolution of the distributions and fundamental statistics of C_{ij} in the three layers. Panels (a) to (c): The distributions of C_{ij} in the three layers. Panels (d) to (f): The averages (together with their 95 % confidence intervals) of C_{ij} in the three layers. Panels (g) to (i): The values of the skewness of C_{ij} in the three layers. Panels (j) to (l): the values of the kurtosis of C_{ij} in the three layers.

For the sake of conciseness, we select the top 5 largest eigenvalues of the empirical correlation matrix in each layer, which are always beyond the maximum RMT value λ_{max}^{RMT} ,

and then investigate their temporal evolution over time. We also measure their contribution to the total variance based on the associated absorption ratios. As shown in Figure (6.4), a couple of states in the market can be identified. For example, the evolution shows that almost all these eigenvalues and the absorption ratios increase during the period of the Japanese asset price bubble (1986-1991), indicating an increase in the level of systemic risk during that period as well (e.g. Kritzman et al., 2011; Billio et al., 2012; Meng et al., 2014). The largest eigenvalue and the associated absorption ratio (in the layer 1 and layer 3) also increase during the period from 2002 to 2008 when the Bank of Japan (BOJ) implements monetary stimulus policy (e.g. Bowman et al., 2015). In addition, a significant decrease in these eigenvalues and absorption ratios in 2009 (especially in the layers of long-term and total loans) can be interpreted as a consequence of the economic contraction in Japan hit by the global crisis (e.g. Kawai and Takagi, 2009).

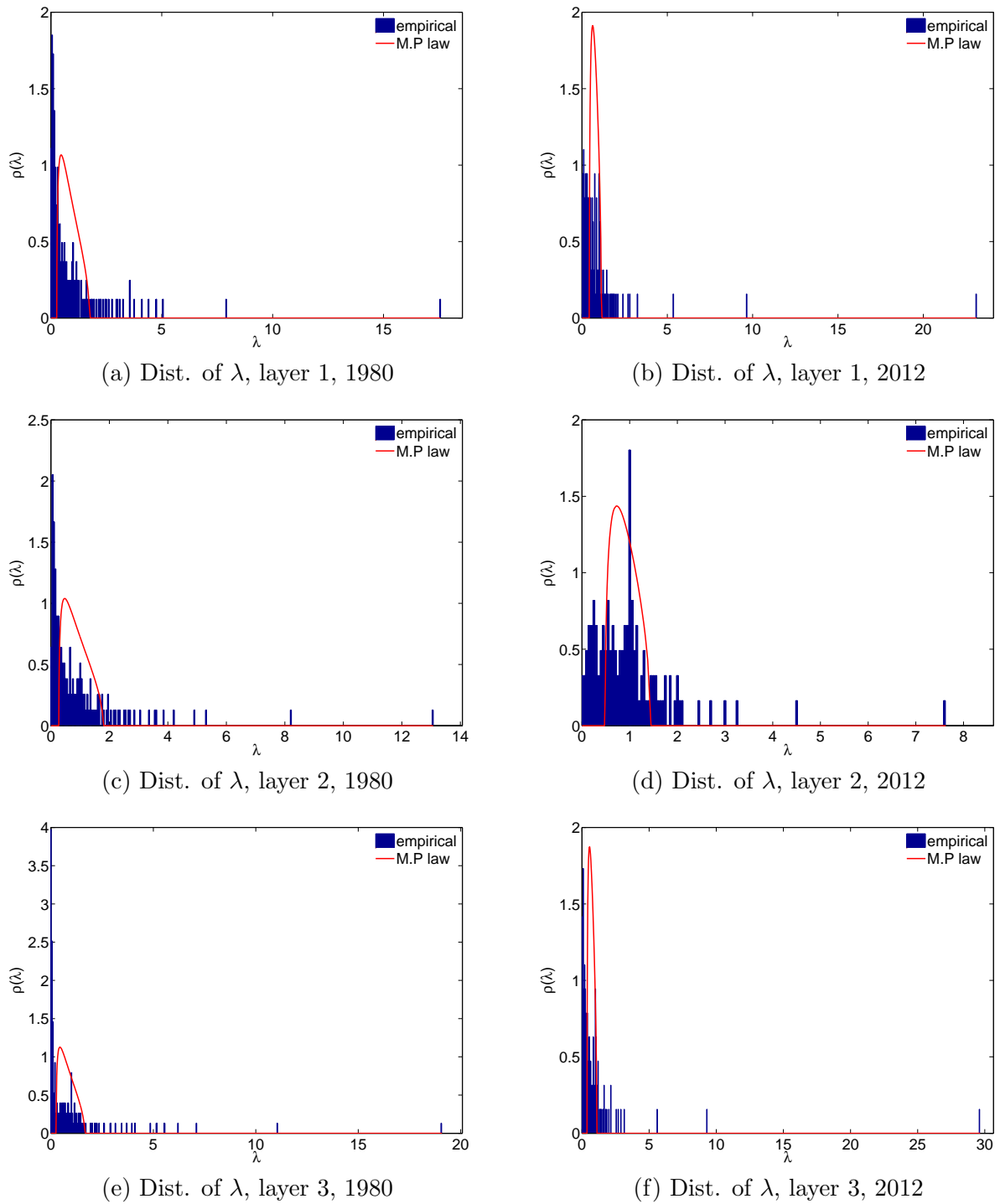
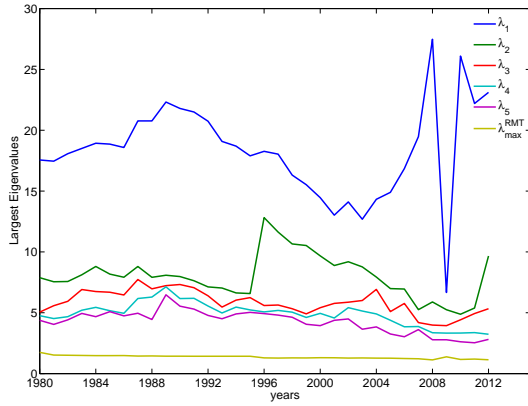
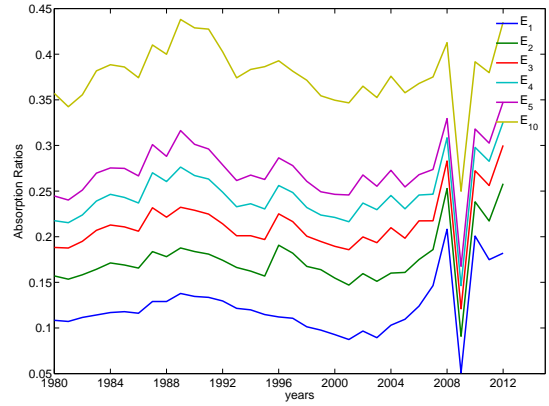


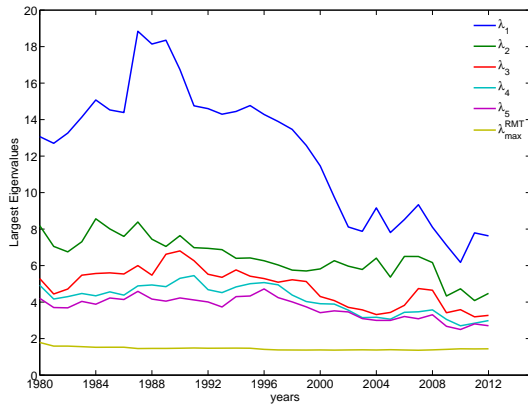
Figure 6.3: Distributions of the eigenvalues of the correlation matrices, compared with RMT, in the three layers, in 1980 and 2012. Panels (a), (c), (e) show the distributions of λ in 1980, and panels (b), (d), (f) show the distributions of λ in 2012.



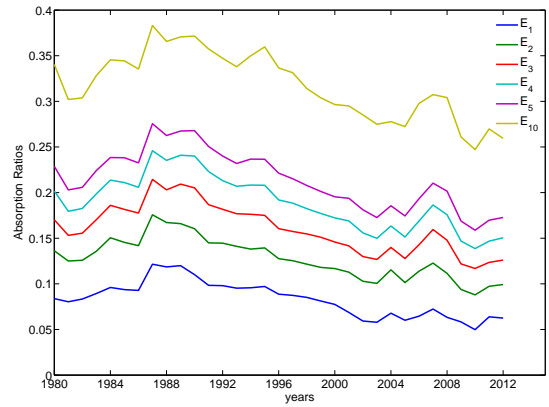
(a) Largest eigenvalues, layer 1



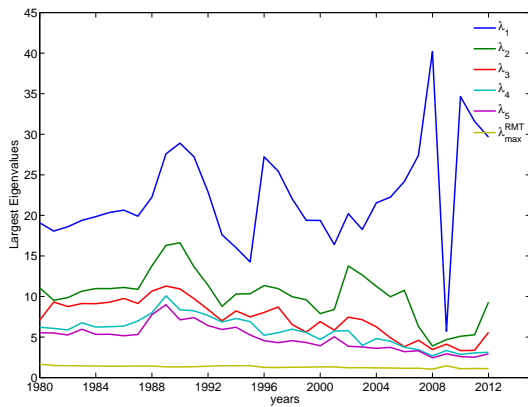
(b) Absorption ratios, layer 1



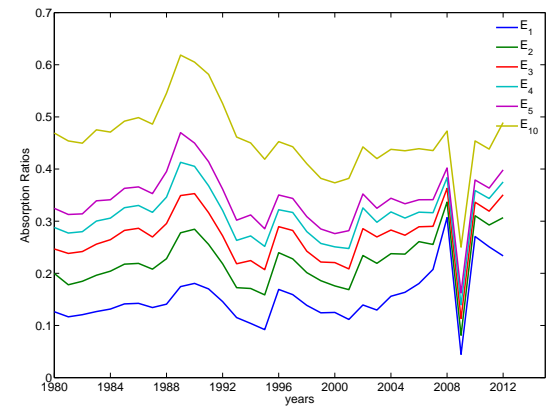
(c) Largest eigenvalues, layer 2



(d) Absorption ratios, layer 2



(e) Largest eigenvalues, layer 3



(f) Absorption ratios, layer 3

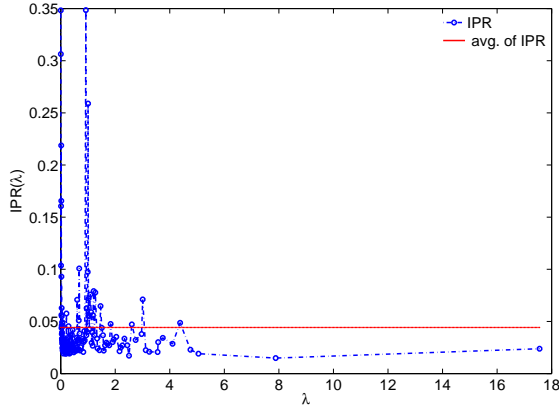
Figure 6.4: Evolution of the top 5 largest eigenvalues and absorption ratios for the correlation matrices in the three layers. The panels on the left show the evolution of the top 5 largest eigenvalues (these eigenvalues are larger than λ_{max}^{RMT}). The right panels show the 5 associated absorption ratios and the absorption ratio of the top 10 largest eigenvalues.

We now move on to the analysis of the contribution of banks to each eigenvector of C . Recalling that the i^{th} element of an eigenvector u_k indicates the contribution of the bank i to u_k , we calculate the Inverse Participation Ratios IPR_k and then examine the relationship between them and the corresponding eigenvalues. Large values of IPR_k reveal few banks contribute to u_k , suggesting the presence of a localized behavior of banks. In contrast, a small value of IPR_k indicates that many banks contribute to u_k , showing that the factor (if it exists) associated with that eigenvector has a wide effect on banks.

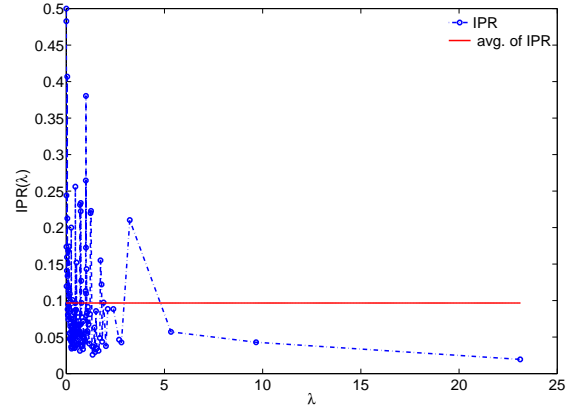
In all panels of Figure (6.5), we can see that there is a common observation among the three layers, i.e. the smallest eigenvalues often have highest IPR and their IPR ratios deviate from the average of $\{IPR_k\}_{k=1}^N$. This indicates that their associated eigenvectors are localized, i.e. a small group of banks contribute to them. Nevertheless, since these smallest eigenvalues are much smaller than 1, the contribution of the components associated with these eigenvectors to the total variance is actually negligible (see Eqs. (6.11) and (6.12)).

Additionally, we also observe a large group of small values of IPR_k associated with eigenvalues located in the center of the distribution of eigenvalues. These two results are very similar to what one often observes in the analysis of the stock price changes (e.g. Plerou et al., 1999). Interestingly, in the case of the eigenvector associated with the first largest eigenvalue, there are many banks contributing to this eigenvector (typically IPR_1 is less than the average of $\{IPR_k\}_{k=1}^N$). This indicates a wide effect of the ‘market’ mode on many banks. Such behavior still emerges in a couple of the next largest eigenvalues deviating from the random bulk, suggesting that the elements of these eigenvectors may represent large groups of banks that are more correlated. It should be emphasized that, under a detailed inspection, in each year, we do observe the localization behavior in at least one of the largest eigenvalues, but the behavior of that eigenvector is not persistent over time. For instance, in 2012, we can see that in the first and the third layers, IPR_4 is larger than the average of $\{IPR_k\}_{k=1}^N$ (see panels (b), (f) of Figure (6.5)). In the second layer, such behavior can be found in the case of IPR_8 .

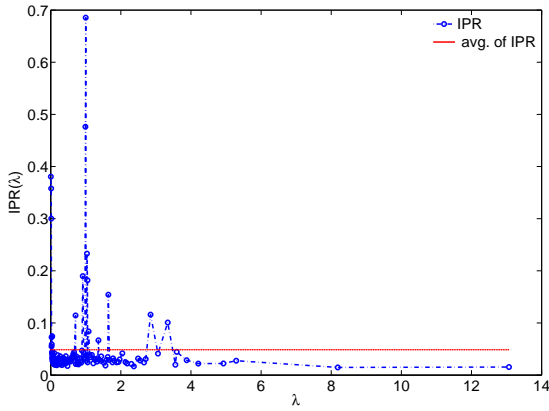
Note that we can compare the distributions of the empirical eigenvector components with the theoretical distribution predicted by RMT. According to the Gaussian prediction of RMT, the components of each eigenvector will follow $P^{RMT}(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$ (e.g. Plerou et al., 2002). We typically find that the distribution of the elements of eigenvectors corresponding to the largest eigenvalues significantly deviate from the Gaussian prediction of RMT (e.g. see Figure (6.11) in the Appendix).



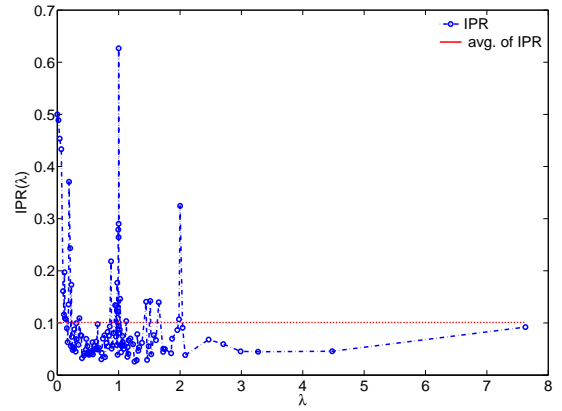
(a) IPR, layer 1, 1980



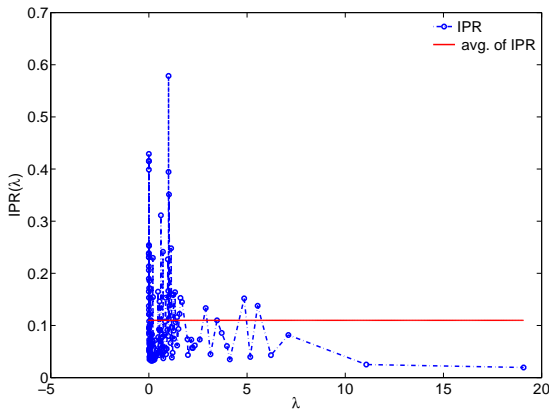
(b) IPR, layer 1, 2012



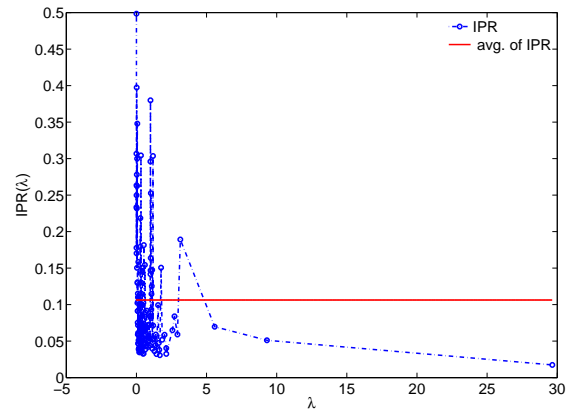
(c) IPR, layer 2, 1980



(d) IPR, layer 2, 2012



(e) IPR, layer 3, 1980

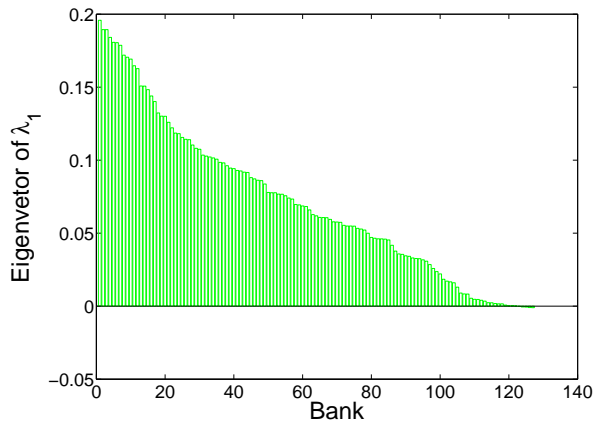


(f) IPR, layer 3, 2012

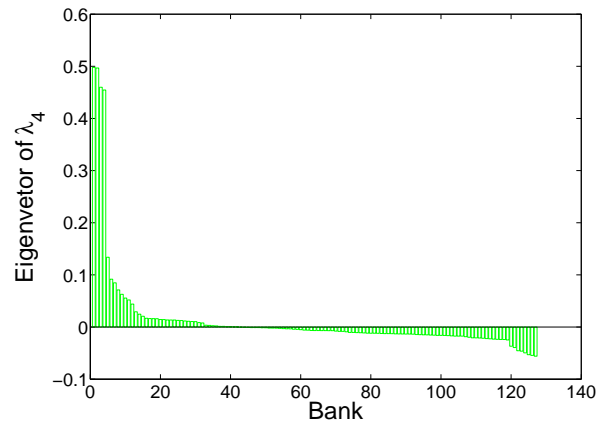
Figure 6.5: Inverse Participation Ratios of the eigenvectors of the correlation matrices in the three layers. Panels (a), (b) show IPR vs. λ in the layer 1, in 1980 and 2012. Panels (c), (d) show IPR vs. λ in the layer 2, in 1980 and 2012. Panels (e), (f) show IPR vs. λ in the layer 3, in 1980 and 2012. The red lines stand for the averages of $\{IPR_k\}_{k=1}^N$.

For a more detailed assessment of bank contributions to eigenvectors, we show the elements of two eigenvectors u_1 and u_4 in 2012 as an example (see Figure (6.6)). In this case, one can see that, in the first and third layers, the localization in u_4 is stronger than in u_1 , i.e. there are more banks that have significant components in the eigenvector u_1 while few banks dominate in the components of the eigenvector u_4 . Furthermore, it should be emphasized that even in the case of u_1 , where IPR_1 is smaller than the average of $\{\text{IPR}_k\}_{k=1}^N$, the elements of u_1 are not homogeneous. In addition, the second layer shows a different behavior. More specifically, although both IPR_1 and IPR_4 are smaller than the average of $\{\text{IPR}_k\}_{k=1}^N$, IPR_1 is larger than IPR_4 so that in general there are fewer banks contributing to u_1 , compared to those contributing to u_4 .

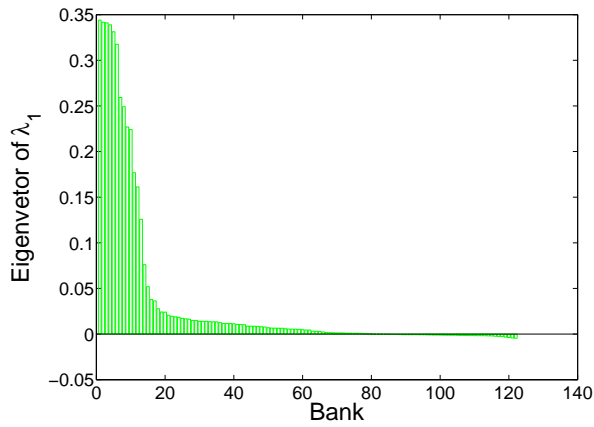
Moreover, if the largest eigenvalue and the next largest one respectively stand for the ‘market’ mode and ‘group’ mode, it is plausible to compare the raw correlation matrix with the filtered one when the effects of ‘market’ mode and/or ‘group’ mode are subtracted. As can be seen in the panels (a) to (f) of Figure (6.7), after excluding the effects of market mode on cross-correlation matrix C , we still observe significant correlations in the ‘market’ mode filtered matrix FC^m , which may be driven by ‘group’ mode associated with the next largest eigenvalues. Further decomposition can be implemented to separate the ‘market’ mode filtered matrix into the correlations driven by the ‘group’ mode and by noise. The ‘market’ as well as ‘group’ modes-filtered matrix $FC^{m,g}$ in three layers is shown in the panels (g) to (i) of Figure (6.7), which indicates that almost all correlations are significantly subtracted after excluding the effects of the two modes.



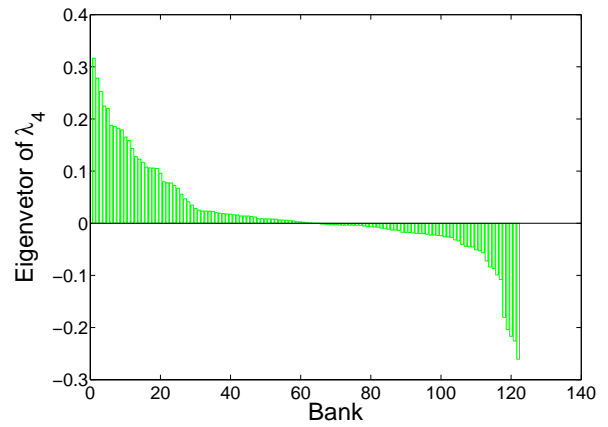
(a) Evec1 in 2012, layer 1



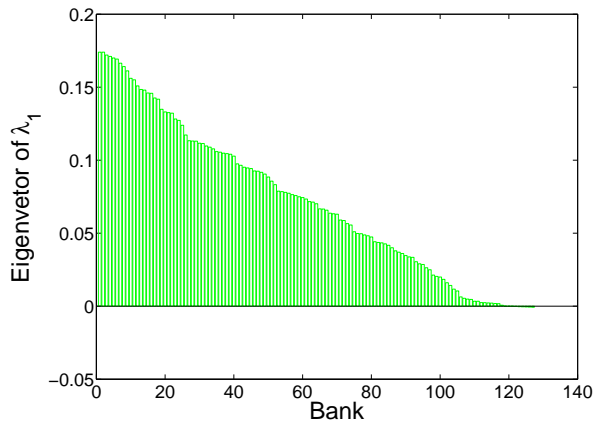
(b) Evec4 in 2012, layer 1



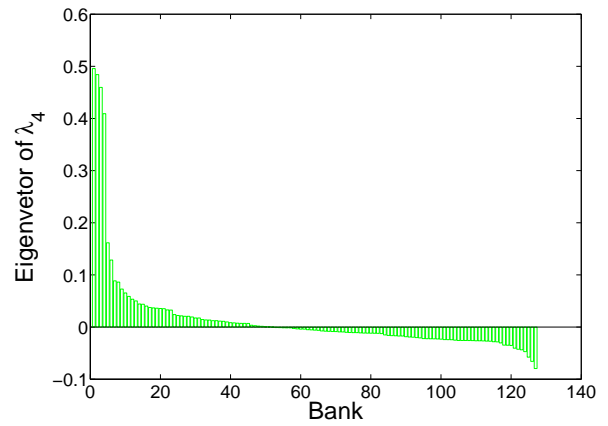
(c) Evec1 in 2012, layer 2



(d) Evec4 in 2012, layer 2



(e) Evec1 in 2012, layer 3



(f) Evec4 in 2012, layer 3

Figure 6.6: The eigenvectors (sorted in descending order) of λ_1 and λ_4 , in the three layers, in 2012. Without loss of generality, in each eigenvector we assume that the sign of the element that has the largest absolute value is non-negative. Panels (a), (b) respectively show the eigenvectors of λ_1 and λ_4 in the layer 1. Panels (c), (d) respectively show the eigenvectors of λ_1 and λ_4 in the layer 2. Panels (e), (f) respectively show the eigenvectors of λ_1 and λ_4 in the layer 3.

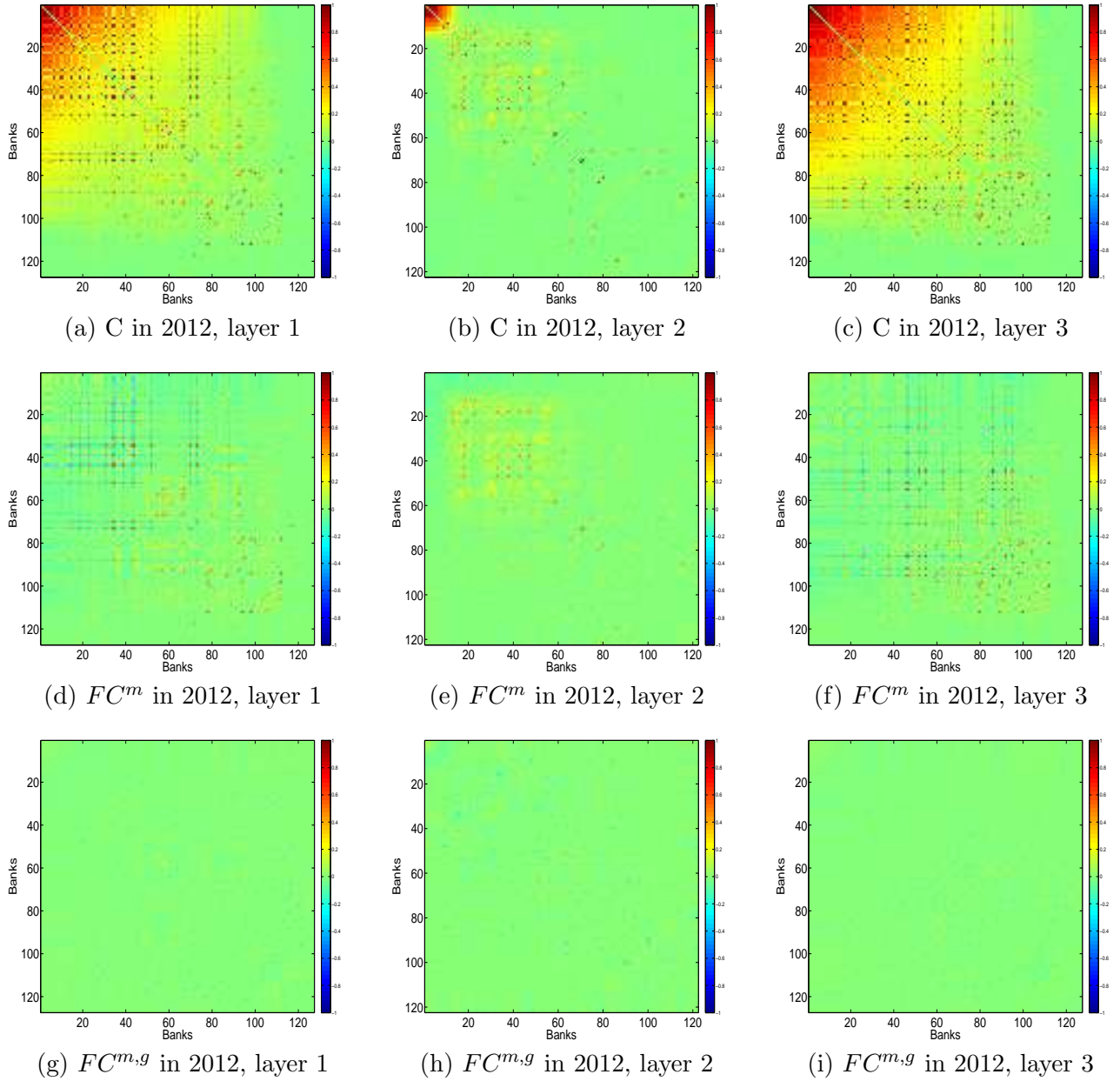


Figure 6.7: Correlations and filtered correlations between banks in lending to firms in the three layers, in 2012. The three upper panels from (a) to (c) are the raw correlation matrices. The three panels from (d) to (f) are the filtered correlation matrices by RMT method, subtracting the ‘market’ mode. The three panels from (g) to (i) are the filtered correlation matrices by RMT method, subtracting ‘market’ mode as well as ‘group’ mode.

Interestingly, in general, the analysis of the empirical cross-correlation matrix C^{bin} of the binary credit relationships between banks and firms also provide several similar as well as independent results. Recalling the method used to calculate that matrix in the previous section, now we consider the case of the categorical data where the binary value of $a_{i,j}$

indicates whether bank i lend to firm j . We define the matrix $X_{bin} = \{X_{i,j,bin}\}_{N \times NF}$, which each element is defined as

$$X_{i,j,bin} = \frac{a_{i,j} - \langle a_{i,j} \rangle}{\sigma(a_{i,j})}, \quad (6.22)$$

where for each bank i , $\langle a_{i,j} \rangle$ and $\sigma(a_{i,j})$ are respectively the average and the standard deviation of $a_{i,j}$ over all NF firms. The correlation matrix C^{bin} of the binary data is defined as

$$C^{bin} = \{C_{ij}^{bin}\}_{N \times N} = \frac{1}{NF} X_{bin} X_{bin}^T, \quad (6.23)$$

with X_{bin}^T is the transposition matrix of X_{bin} .

For the sake of conciseness, in the following we will only summarize the main results for the analysis of the binary data. First, comparing to RMT, we also find that a group of largest eigenvalues (typically from 10 % to 15 %) always deviate from the bulk (e.g. see Figure (6.8)). In addition, we often observe that more than half of the eigenvalues lies between the RMT bounds.

Compared to the weighted data, for the layers of total loans and long-term loans, we also find similar results for the evolution of the largest eigenvalues (larger than λ_{max}^{RMT}) and the absorption ratios, although in general these largest eigenvalues relatively contribute less to the total variance in the binary data than in the weighted data (e.g. see Figure (6.9), then compare to Figure (6.4)). In contrast, the evolution in the layer of short-term loans exhibits different behaviors. It shows that the top 5 largest eigenvalues and their contribution to the total variance decrease from 1995. This may be due to the fact that, the binary bipartite structure of the bank-firm credit network in the short-term lending relations is much more sparse than those in other layers. However, it is worthwhile to emphasize that, when aggregating short-term and long-term loans together, we persistently observe that many features of the correlation matrix in the layer of the total loans are similar to those of the correlation matrix in the layer of the long-term loans.

Moreover, on the one hand, the localization behavior still emerges in the eigenvectors corresponding to the smallest eigenvalues, as we can see that their IPR ratios are much higher than the average of IPR. On the other hand, the ‘market’ mode as well as the ‘group’ mode still have a wide effect on many banks, which is revealed by the fact that many banks contribute to the eigenvectors of the top largest eigenvalues (e.g. see Figure (6.10)).

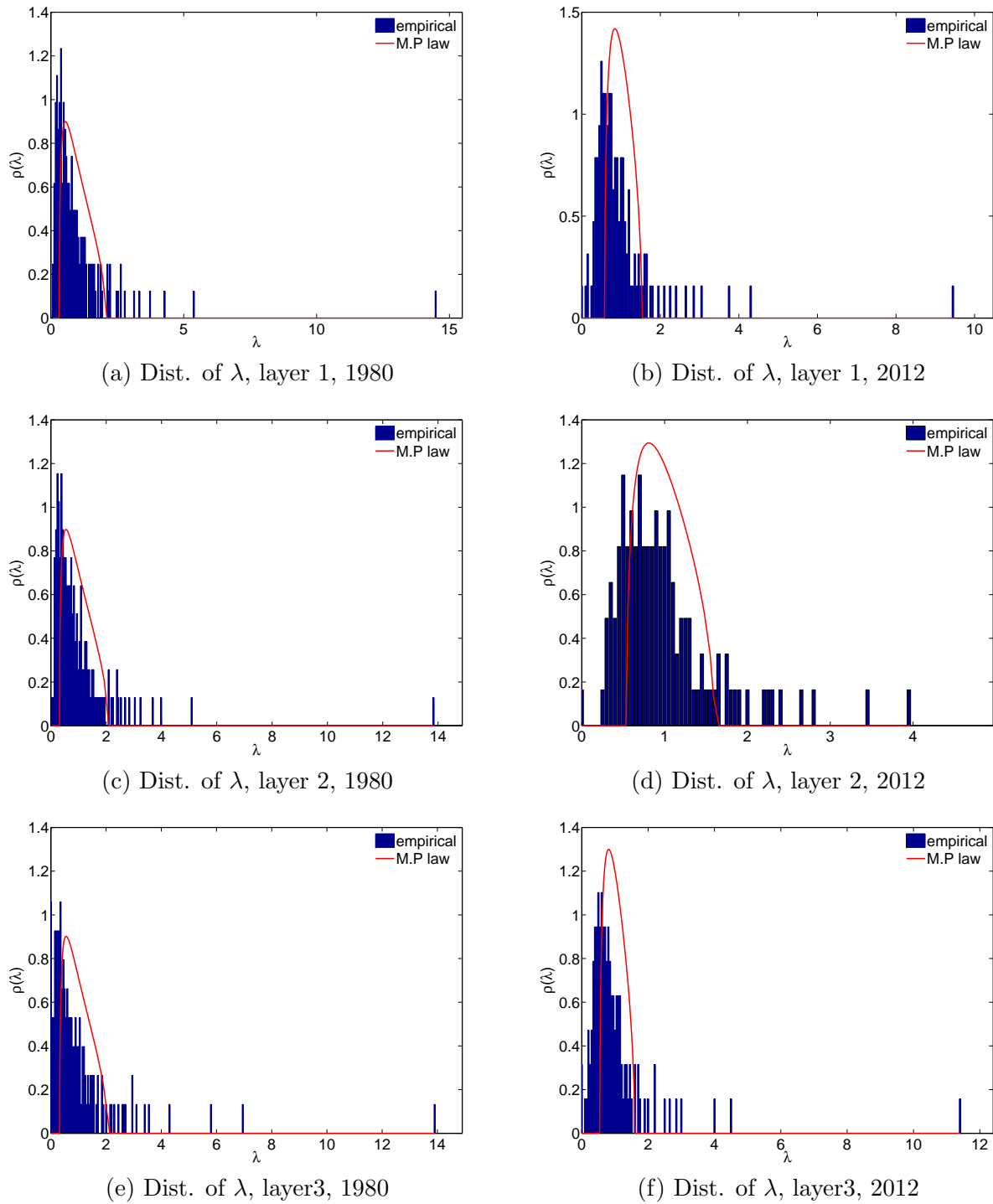
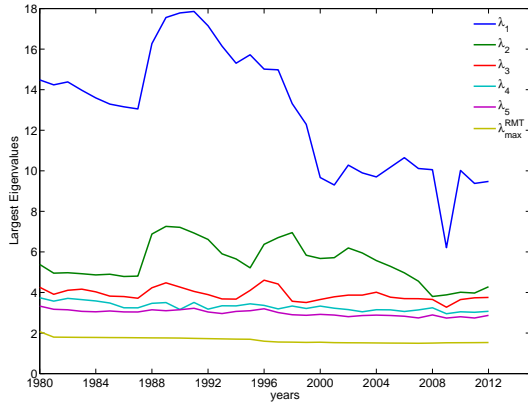
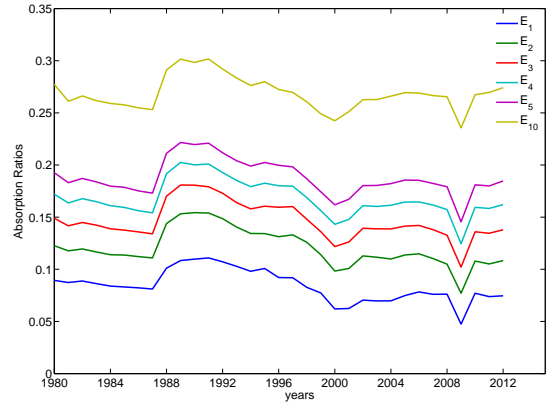


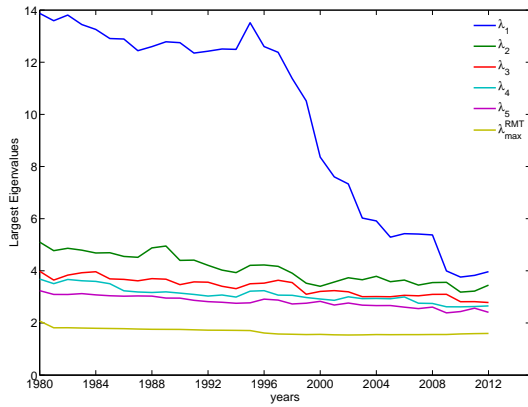
Figure 6.8: Binary data: Distribution of eigenvalues of C^{bin} , compared with RMT, in the three layers, in 1980 and 2012. Panels (a), (c), (e) show the distributions of λ in 1980, and panels (b), (d), (f) show the distributions of λ in 2012.



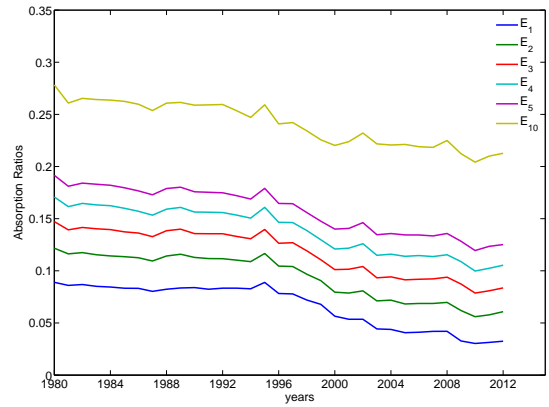
(a) Top eigenvalues, layer 1



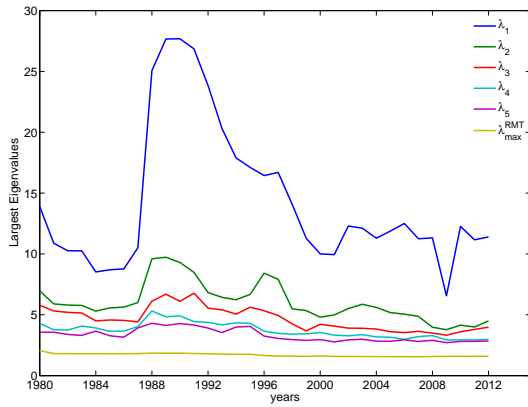
(b) Absorption ratios, layer 1



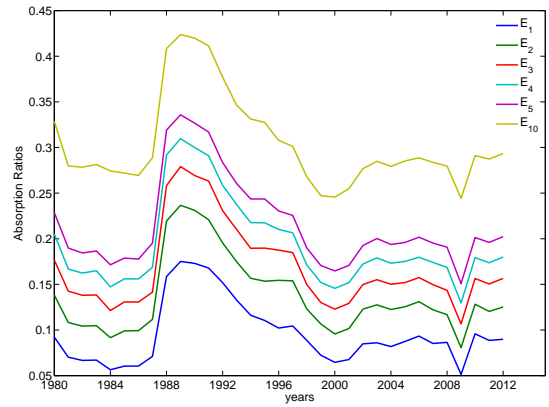
(c) Top eigenvalues, layer 2



(d) Absorption ratios, layer 2

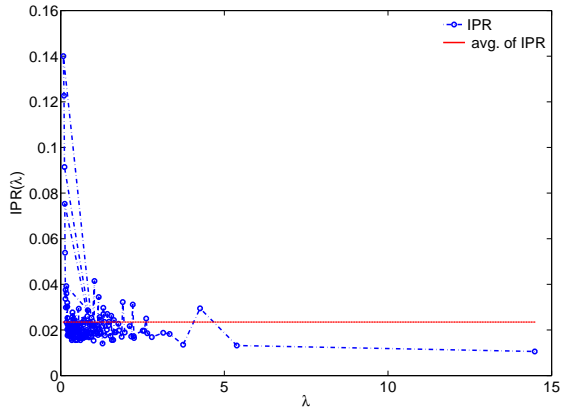


(e) Top eigenvalues, layer 3

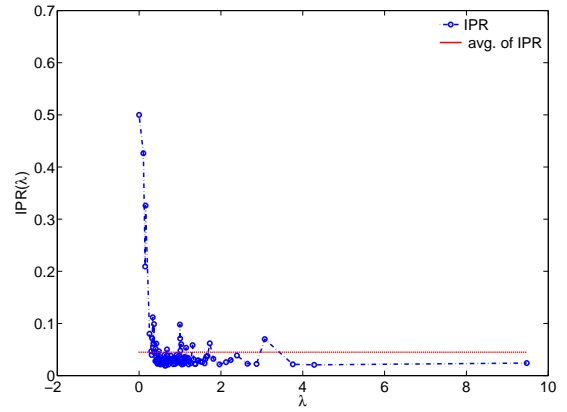


(f) Absorption ratios, layer 3

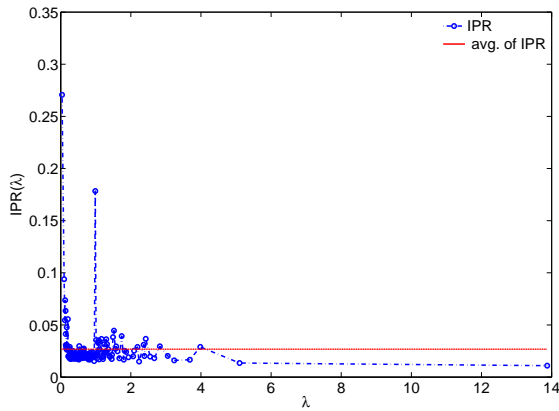
Figure 6.9: Binary data: Evolution of the top 5 largest eigenvalues and the associated absorption ratios for C^{bin} in the three layers. The panels on the left show the evolution of the top 5 largest eigenvalues (these eigenvalues are larger than λ_{max}^{RMT}). The panels on the right show the associated absorption ratios and the absorption ratio of the top 10 largest eigenvalues.



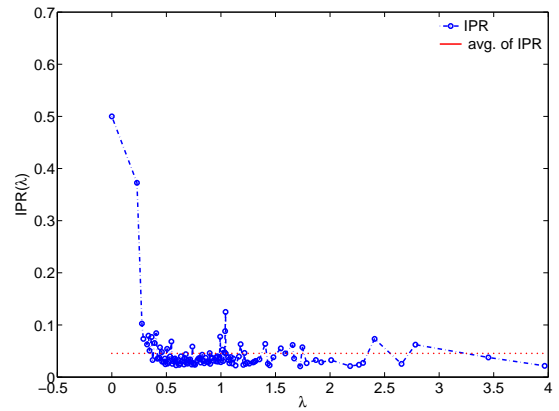
(a) IPR, layer 1, 1980



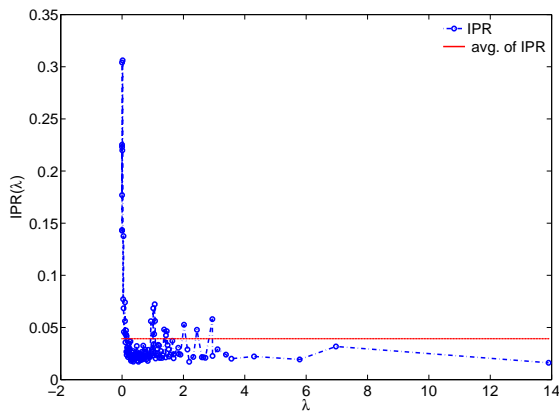
(b) IPR, layer1, 2012



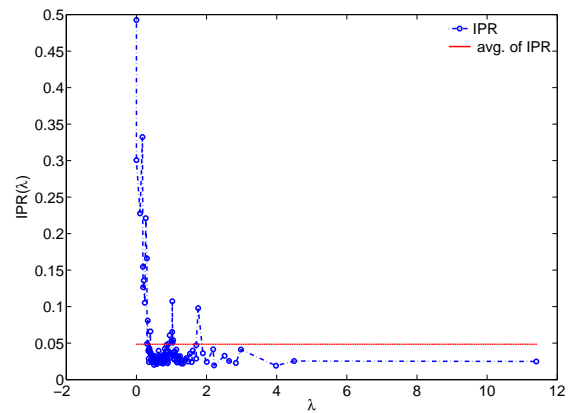
(c) IPR, layer 2, 1980



(d) IPR, layer 2, 2012



(e) IPR, layer 3, 1980



(f) IPR, layer 3, 2012

Figure 6.10: Binary data: Inverse Participation Ratios of the eigenvectors of C^{bin} in the three layers, in 1980 and 2012. Panels (a), (b) show IPR vs. λ in the layer 1, in 1980 and 2012. Panels (c), (d) show IPR vs. λ in the layer 2, in 1980 and 2012. Panels (e), (f) show IPR vs. λ in the layer 3, in 1980 and 2012. The red lines stand for the averages of $\{IPR_k\}_{k=1}^N$.

6.4 Conclusions

In this paper, we have examined the structure of the correlation matrices for the banks' loan portfolios in the various maturity-based layers in the bank-firm credit market of Japan. Although from network theory's perspective, banks show high overlaps in their lending to firms, having applied Random Matrix Theory to detect patterns in the empirical correlation matrices, we observe that in general, actually less than 20% of the eigenvalues have genuine information while the rest actually corresponds to noise or just have a small effect on the correlations between banks. This implies that the methods of RMT can be used for filtering noise from the empirical correlation matrices of banks' loan portfolios.

The dynamics of the largest eigenvalues (which deviate from the random bulk explained by RMT) and the associated absorption ratios suggest that there was an increase in the level of systemic risk during the period of the Japanese asset price bubble (1986-1991). A significant decrease in these eigenvalues and absorption ratios in 2009 can be interpreted as a consequence of the economic contraction in Japan hit by the global crisis (e.g. Kawai and Takagi, 2009).

Based on the eigenvector components of the largest eigenvalues, we have further investigated the clusters of banks. We find that, on the one hand, in the case of the eigenvectors corresponding to a group of smallest eigenvalues, there are very few banks contributing to them. Such a localized behavior is similar to what is often found in analyses of stock price changes (e.g. Plerou et al., 1999). However, since these eigenvalues are much smaller than 1, the contribution of the corresponding eigenvector components to the total variance is also negligible. On the other hand, we observe that the eigenvectors corresponding to the largest eigenvalues are not very localized. This indicates that latent factors extracted from these largest eigenvalues have a wide effect on banks, although it should be emphasized that, banks do not contribute homogeneously to these eigenvectors.

The comparison between the analysis of the binary data and the analysis of the weighted data provides an interesting observation, i.e. both cases share several common as well as independent results, indicating that the binary version of bank-firm credit relationships also contains significant information about the cross-correlations between banks' loan portfolios. Such a relationship between binary and non-binary data is also found in some empirical studies of stock price changes (e.g. Almog and Garalascelli, 2014; Almog et al., 2015).

Obviously, the effects of the bank characteristics and bank categories on the formation of clusters of banks in the correlation matrices in the various layers should be studied further. In addition, as we can see in the analysis of eigenvector components, it is often observed

in a couple of the eigenvectors of the top largest eigenvalues that a large group of banks have a higher degree of correlations. Therefore, it is also interesting to analyze the mesoscopic structure in the correlation matrices (e.g. see Almog et al., 2015; MacMahon and Garlaschelli, 2015).

6.5 References

Almog A., Garlaschelli D. 2014. Binary versus non-binary information in real time series: empirical results and maximum-entropy matrix models. *New Journal of Physics* 16.

Almog A., Besamusca F., MacMahon M., Garlaschelli D. 2015. Mesoscopic community structure of financial markets revealed by price and sign fluctuations. *PLoS ONE* 10 (7).

Alvarez-Socorro A. J., Herrera-Almarza G. C., and González-Díaz L. A. 2015. Eigencentality based on dissimilarity measures reveals central nodes in complex networks. *Scientific Reports* 5.

Aoyama H. 2014. Systemic risk in Japanese credit network. In Abergel F., Aoyama H., Chakrabarti B. K., Chakraborti A., Ghosh A. (Eds.), *Econophysics of Agent-Based Models* (pp. 219-228), Springer International Publishing.

Aoyama H., Battiston S., Fujiwara Y. 2013. DebtRank analysis of the Japanese credit network. RIETI Discussion Paper Series 13-E-087. Available at: <http://www.rieti.go.jp/jp/publications/dp/13e087.pdf>.

Bandyopadhyay J. N., Jalan S. 2007. Universality in complex networks: random matrix analysis. *Physical Review E* 76 (2).

Bassett W. F., Chosak M. B., Driscoll J. C., Zakrajsek E. 2014. Changes in bank lending standards and the macroeconomy. *Journal of Monetary Economics* 62, pp. 23-40.

Beale N., Rand D. G., Battey H., Croxson K., May R. M., Nowak M. A. 2011. Individual versus systemic risk and the regulator's dilemma. *Proc. Natl. Acad. Sci.* 108 (31), pp. 12647-12652.

Billio M., Getmansky M., Lo A. W., Pelizzon L. 2012. Econometric measures of systemic risk in the finance and insurance sectors. *Journal of Financial Economics* 104 (3), pp. 535-559.

Bouchaud J. -P., Potters M. 2009. Financial applications of random matrix theory: a short review. Working Paper. Available at: [arXiv:0910.1205](https://arxiv.org/abs/0910.1205).

Bowman D., Cai F., Davies S., Kamin S. 2015. Quantitative easing and bank lending: Evidence from Japan. *Journal of International Money and Finance* 57, pp. 15-30.

- Brummitt C. D., D'Souza R. M., Leicht E. A. 2012. Suppressing cascades of load in interdependent networks. *Proc. Natl. Acad. Sci.* 109 (12), pp. 680-689.
- Buldyrev S. V., Parshani R., Paul G., Stanley H. E., Havlin S. 2010. Catastrophic cascade of failures in interdependent networks. *Nature* 464, pp. 1025-1028.
- Cai J., Saunders A., Steffen S. 2016. Syndication, interconnectedness, and systemic risk. Working Paper. Available at: <http://ssrn.com/abstract=1508642orhttp://dx.doi.org/10.2139/ssrn.1508642>.
- Caccioli F., Shrestha M., Moore C., Farmer J. D. 2014. Stability analysis of financial contagion due to overlapping portfolios. *Journal of Banking & Finance* , 46, pp. 233-245.
- Corsi F., Marmi S., Lillo F. 2013. When micro prudence increases macro risk: The destabilizing effects of financial innovation, leverage, and diversification. Working Paper. Available at: <http://ssrn.com/abstract=2278298>.
- de Carvalho J. X., Jalan S., Hussein M. S. 2009. Deformed Gaussian-orthogonal-ensemble description of small-world networks. *Physical Review E* 79 (5).
- de Castro Miranda R. C., Tabak B. M. 2013. Contagion risk within firm-bank bivariate networks. Working Paper 322, Central Bank of Brazil. Available at: <https://www.bcb.gov.br/pec/wps/ingl/wps322.pdf>.
- De Masi G., Fujiwara Y., Gallegati M., Greenwald B., Stiglitz J. E. 2011. An analysis of the Japanese credit network. *Evolutionary and Institutional Economic Review* 7 (2), pp. 209-232.
- Dumitriu I., Johnson T. 2016. The Marčenko-Pastur law for sparse random bipartite biregular graphs. *Random Structures and Algorithms* 48 (2), pp. 313-340.
- Fricke D. 2016. Has the banking system become more homogeneous? Evidence from banks' loan portfolios. *Economics Letters* 142, pp. 45-48.
- Fujiwara Y., Aoyama H., Ikeda Y., Iyetomi H., Souma W. 2009. Structure and temporal change of the credit network between banks and large firms in Japan. *Economics: The Open-Access, Open-Assessment E-Journal* 3.
- Gambetti L., Musso A. 2016. Loan supply shocks and the business cycle. *Journal of Applied Econometrics*.
- Gertler M., Kiyotaki N. 2010. Financial intermediation and credit policy in business cycle analysis. In Friedman B. M., Woodford M. (Eds.), *Handbook of Monetary Economics* 3 (pp. 547-599), North Holland.
- Greenwood R. M., Landier A., Thesmar D. 2015. Vulnerable banks. *Journal of Financial Economics* 115 (3), pp. 471-485.

- Huang X., Gao J., Buldyrev S. V., Havlin S., Stanley H. E. 2011. Robustness of interdependent networks under targeted attack. *Physical Review E* 83 (6).
- Huang X., Vodenska I., Havlin S., Stanley and H. E. 2013. Cascading failures in bi-partite graphs: Model for systemic risk propagation. *Scientific Reports* 3.
- Iyetomi H., Nakayama Y., Yoshikawa H., Aoyama H., Fujiwara Y., Ikeda Y., Souma W. 2011. What causes business cycles? Analysis of the Japanese industrial production data. *Journal of the Japanese and International Economies* 25 (3), pp. 246-272.
- Jiang X. F., Zheng B. 2012. Anti-correlation and subsector structure in financial systems. *Europhysics Letters* 97 (4).
- Jiang X. F., Chen T. T., Zheng B. 2014. Structure of local interactions in complex financial dynamics. *Scientific Reports* 4.
- Jolliffe I. T. 1986. *Principal component analysis*. Springer-Verlag, New York.
- Jordà O., Schularick M., Taylor A. M. 2013. When credit bites back. *Journal of Money, Credit and Banking* 45 (2), pp. 3-28.
- Kawai M., Takagi S. 2009. Why was Japan hit so hard by the global financial crisis?. ADBI Working Paper 153. Available at: <http://www.adb.org/sites/default/files/publication/156008/adbi-wp153.pdf>.
- Kichikawa Y., Kita Y., Iyetomi H. 2015. Complex principle component analysis on dynamic correlation structure in price index data. *Procedia Computer Science* 60, pp. 1836-1845.
- Kim D. -H., Jeong H. 2005. Systematic analysis of group identification in stock markets. *Physical Review E* 72 (4).
- Kritzman M., Li Y., Page S., Rigobon R. 2011. Principal components as a measure of systemic risk. *The Journal of Portfolio Management* 37 (4), 112-126.
- Laloux L., Cizeau P., Bouchaud J. -P., Potters M. 1999. Noise dressing of financial correlation matrices. *Physical Review Letters* 83 (7), pp. 1467-1470.
- Laloux L., Cizeau P., Potters M., Bouchaud J. -P. 2000. Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance* 3 (3), pp. 391-397.
- Levy-Carciente S., Kenetta D. Y., Avakiana A., Stanley H. E., Havlin S. 2015. Dynamical macro-prudential stress testing using network theory. *Journal of Banking & Finance* 59, pp. 164-181.
- Lillo F., Pirino D. 2015. The impact of systemic and illiquidity risk on financing with risky collateral. *Journal of Economic Dynamics and Control* 50, pp. 180-202.

- Liu R. - R., Li M., Jia C. -X., Wang B. -H. 2016. Cascading failures in coupled networks with both inner-dependency and inter-dependency links. *Scientific Reports* 6.
- Luu D. T., Lux T. 2016a. Multilayer overlaps and correlations in the bank-firm credit network of Spain. Working Paper.
- Luu D. T., Lux T. 2016b. Identifying patterns in the bank-sector credit network of Spain. Working Paper.
- Luu D. T., 2016. An approach to identify patterns in structural similarities in financial networks. Working Paper.
- Luu D. T., Lux T., Yanovski B. 2016. An analysis of systemic risk in worldwide economic sentiment indices. Working Paper.
- Lux T. 2016. A model of the topology of the bank-firm credit network and its role as channel of contagion. *Journal of Economic Dynamics and Control* 66, pp. 36-53.
- MacMahon M., Garlaschelli D. 2015. Community detection for correlation matrices. *Physical Review X* 5 (2).
- Marotta L., Miccichè S., Fujiwara Y., Iyetomi H., Aoyama H., Gallegati M., Mantegna R. N. 2015. Backbone of credit relationships in the Japanese credit market. Working Paper. Available at: <http://ssrn.com/abstract=2694005>.
- May R. M., Arinaminpathy N. 2010. Systemic risk: the dynamics of model banking systems. *Journal of the Royal Society Interface* 7 (46), pp. 823-838.
- Meng H., Xie W. -J., Jiang Z. -Q., Podobnik B., Zhou W. -X., Stanley H. E. 2014. Systemic risk and spatiotemporal dynamics of the US housing market. *Scientific Reports* 4.
- Ormerod P. 2008. Random matrix theory and macro-economic time-series: An illustration using the evolution of business cycle synchronisation, 1886–2006. *Economics: The Open-Access, Open-Assessment E-Journal* 2.
- Plerou V., Gopikrishnan P., Rosenow B., Luís A., Amaral N., Stanley H. E. 1999. Universal and nonuniversal properties of cross correlations in financial time series. *Physical Review Letters* 83 (7), pp. 1471-1474.
- Plerou V., Gopikrishnan P., Rosenow B., Amaral L. A. N., Stanley H. E. 2000. A random matrix theory approach to financial cross-correlations. *Physical A: Statistical Mechanics and its Applications* 287, pp. 374-382.
- Plerou V., Gopikrishnan P., Rosenow B., Amaral L. A. N., Guhr T., Stanley H. E. 2002. Random matrix approach to cross correlations in financial data. *Physical Review E* 65 (6).
- Pool V. K., Stoffman N., Yonker S. E. 2015. The people in your neighborhood: social interactions and mutual fund portfolios. *The Journal of Finance* 70 (6), pp. 2679-2732.

- Pukthuanthong K., Roll R. 2009. Global market integration: An alternative measure and its application. *Journal of Financial Economics* 94 (2), pp. 214-232.
- Reis S. D. S., Hu Y., Babino A., Andrade J. S. Jr., Canals S., Sigman M., Makse H. A. 2014. Avoiding catastrophic failure in correlated networks of networks. *Nature Physics* 10, pp. 762-767.
- Shen J., Zheng B. 2009. Cross-correlation in financial dynamics. *Europhysics Letters* 86 (4).
- Smart A. G., Amaral L. A. N., Ottino J. M. 2008. Cascading failure and robustness in metabolic networks. *Proc. Natl. Acad. Sci.* 105 (36), pp. 13223-13228.
- Tran L. V., Vu V. H., Wang K. 2013. Sparse random graphs: Eigenvalues and eigenvectors. *Random Structures Algorithms* 42 (1), pp. 110-134.
- Uechi L., Akutsua T., Stanley H. E., Marcus A. J., Kenett D. Y. 2015. Sector dominance ratio analysis of financial markets. *Physica A: Statistical Mechanics and its Applications* 421, pp. 488-509.
- Wang D., Podobnik B., Horvatic D., Stanley H. E. 2011. Quantifying and modeling long-range cross correlations in multiple time series with applications to world stock indices. *Physical Review E* 83 (4).
- Yoshikawa H., Aoyama H., Iyetomi H., Fujiawa Y. 2015. Deflation/inflation dynamics: Analysis based on micro prices. RIETI Discussion Paper Series 15-E-026. Available at: <http://www.rieti.go.jp/jp/publications/dp/15e026.pdf>.
- Zheng Z., Podobnik B., Feng L., Li B. 2012. Changes in cross-correlations as an indicator for systemic risk. *Scientific Reports* 2.

6.6 Appendix

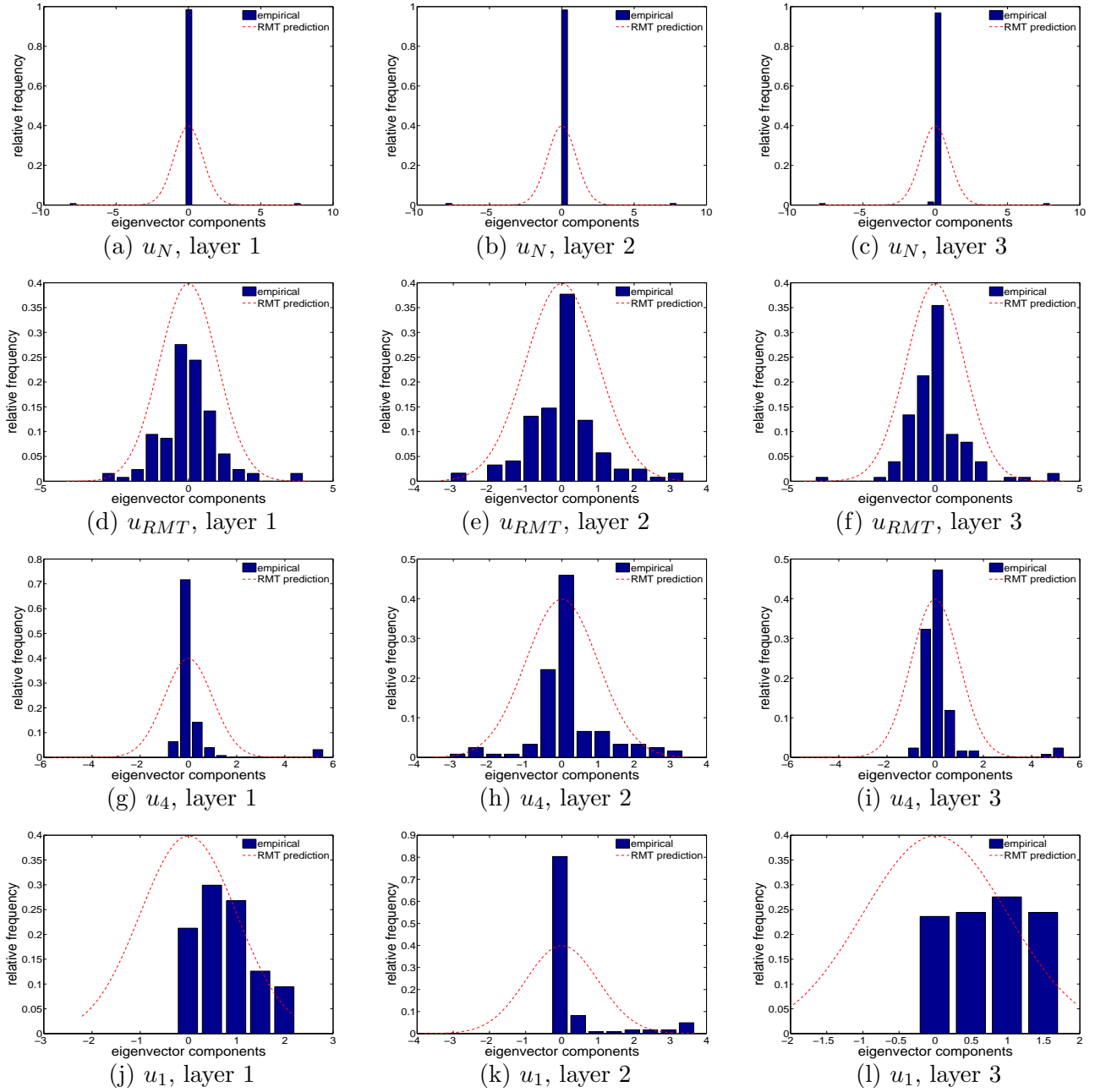
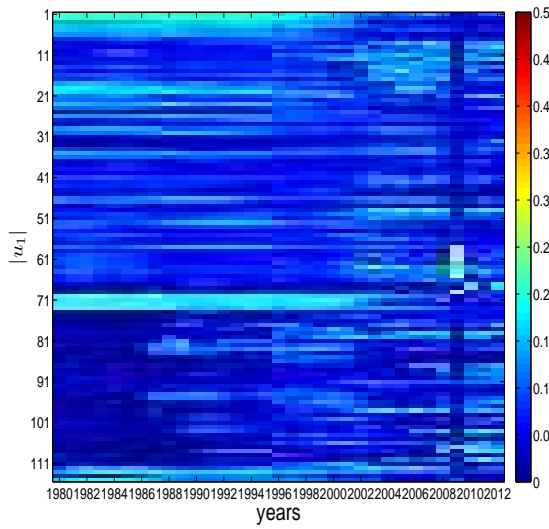


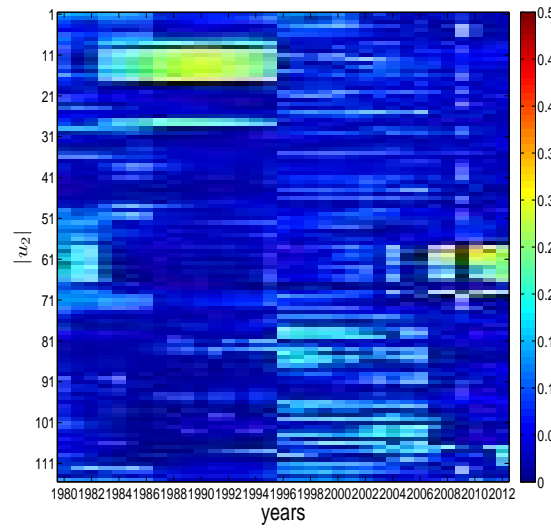
Figure 6.11: Distributions of the eigenvector components of four selected eigenvalues of C , in 2012, in the three layers, in the weighted data. Panels (a) to (c) show the distribution of the eigenvector components of u_N . Panels (d) to (f) show the distribution of the eigenvector components of an eigenvector in the RMT bulk. Panels (g) to (i) show the distribution of the eigenvector components of u_4 . Panels (j) to (l) show the distribution of the eigenvector components of u_1 .

In Figure (6.11), we compare the distributions of four typical eigenvector components of C , i.e. the eigenvector components of u_N (of the smallest eigvalue), the eigenvector components of an eigenvalue located in the RMT bulk (namely u_{RMT}), the eigenvector components of u_4 , and the eigenvector components of u_1 . Note that for the sake of convenience, in this Figure we renormalize the eigenvector components such that $\sum_{j=1}^N u_i(j)^2 = N$. We can see that, first, u_N has only few significant components, indicating that this eigenvector is very localized. Second, the distributions of the components of u_1 and u_4 are significantly different from the distribution of the components of u_{RMT} .

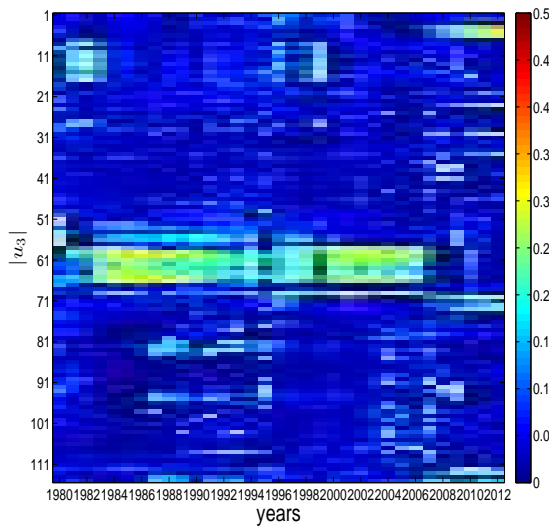
In Figures (6.12), (6.13), and (6.14), we show the evolution of the eigenvectors (in absolute terms) of the top four largest eigenvalues in three layers. Here we focus on banks that activate in the bank-firm credit market of Japan in all years from 1980 to 2012. In these Figures, we can find that some banks correspond to higher eigenvector components in certain subperiods, but overall, in almost all years, the distributions of these eigenvector components are not very wide. This result is consistent with the results obtained from the analysis of IPR.



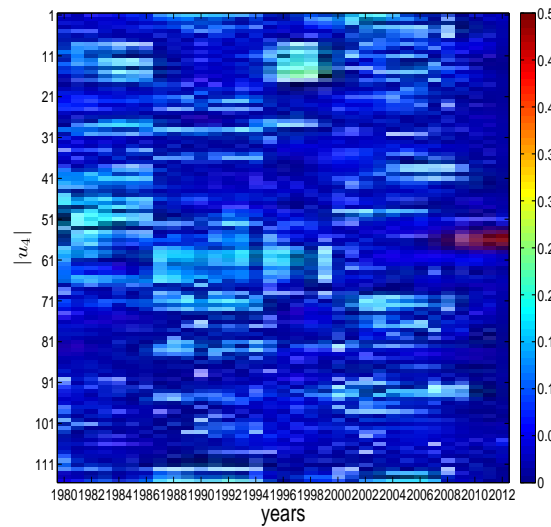
(a) Evec. components of λ_1



(b) Evec. components of λ_2

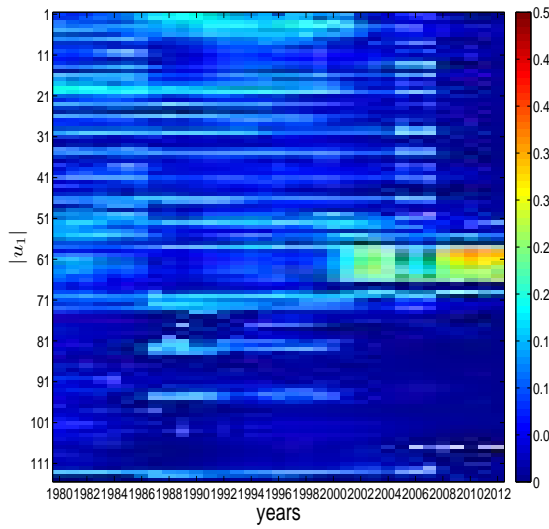


(c) Evec. components of λ_3

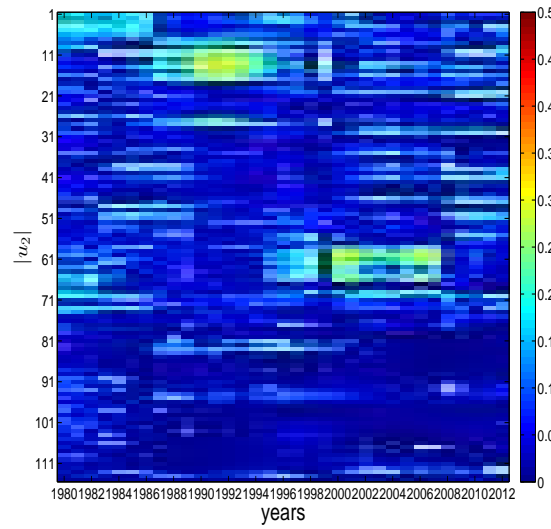


(d) Evec. components of λ_4

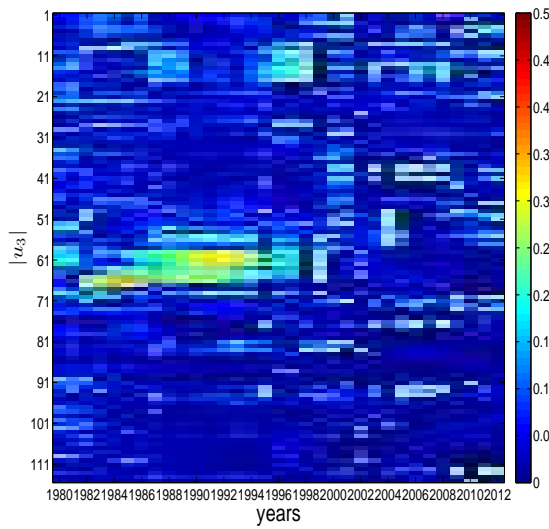
Figure 6.12: Evolution of the eigenvector components of λ_1 (panel (a)), λ_2 (panel (b)), λ_3 (panel (c)), and λ_4 (panel (d)), for 115 banks fixed over years, in weighted data, in the layer 1.



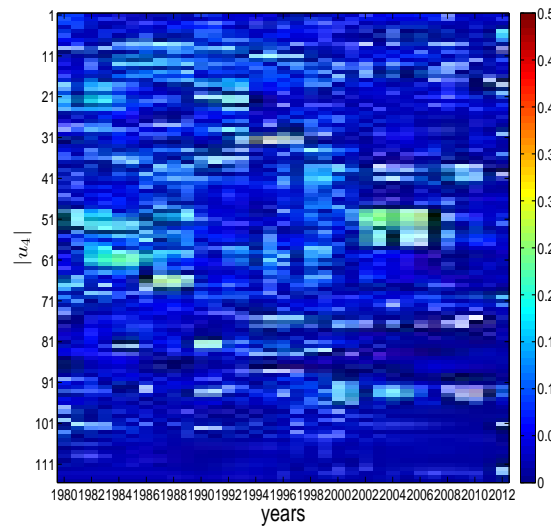
(a) Evec. components of λ_1



(b) Evec. components of λ_2

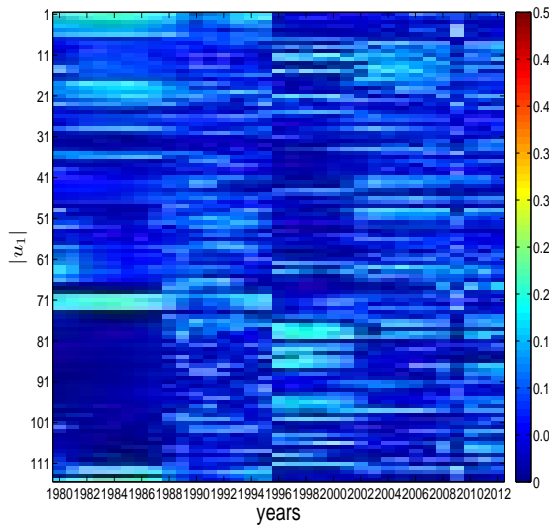


(c) Evec. components of λ_3

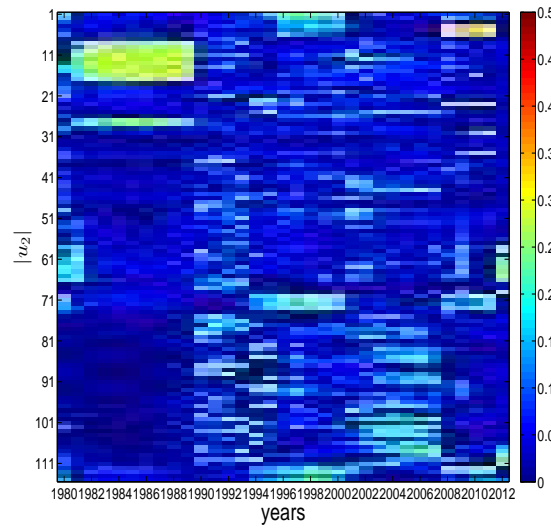


(d) Evec. components of λ_4

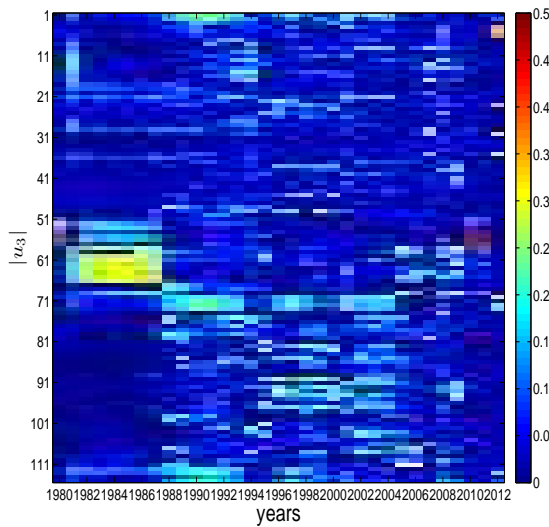
Figure 6.13: Evolution of the eigenvector components of λ_1 (panel (a)), λ_2 (panel (b)), λ_3 (panel (c)), and λ_4 (panel (d)), for 115 banks fixed over years, in weighted data, in the layer 2.



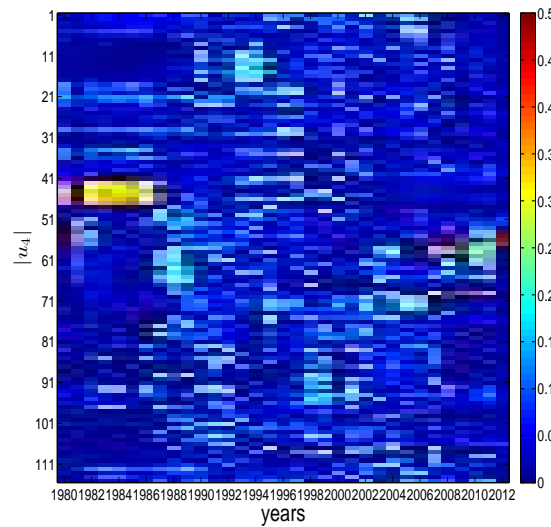
(a) Evec. components of λ_1



(b) Evec. components of λ_2



(c) Evec. components of λ_3



(d) Evec. components of λ_4

Figure 6.14: Evolution of the eigenvector components of λ_1 (panel (a)), λ_2 (panel (b)), λ_3 (panel (c)), and λ_4 (panel (d)), for 115 banks fixed over years, in weighted data, in the layer 3.

Chapter 7

An Analysis of Systemic risk in Worldwide Economic Sentiment Indices

Coauthored by: Boyan Yanovski and Thomas Lux.

Keywords: Sentiment Index; Correlations; Random Matrix Theory; Principal Components.

7.1 Introduction

The analysis of business cycles in different countries is one of the fundamental issues in international economics. So far, much of the analysis is often directed at investigating the synchronization and convergence of “tangible” macroeconomic variables like GDP growth rates, unemployment rates, and so forth (e.g. Bordo and Hebling, 2003; Bordo and Hebling, 2011; Artis et al., 2011; Kose et al., 2012; Ferroni and Klaus, 2015). However, up until now, issues related to the correlations between the expectation structures across different countries have been receiving less attention.

Expectations are a key driver of fluctuations in economic activity since most economically relevant decisions have a strong inter-temporal component (e.g. investment, consumption or saving decisions). This was emphasized, in particular, by Keynes (1936), and later by Minsky (1977) and Akerlof and Shiller (2009). Empirically, such claims are supported by studies by authors like Santero and Westerlund (1996), Howrey (2001), Taylor and McNabb (2007), Carriero and Marcellino (2011), Milani (2011), van Aarle and Kappler (2012), or Mi-

lani (2014) in which the structure of the expectations is measured by sentiment or confidence indices. The expectations themselves are formed on the basis of past experience or on currently incoming information signals from the economy. We argue that “global” information signals, like the collapse of the US housing market in 2007, can lead to a homogenization of the expectation structure around the world, as such information can provide a coordination signal for a global phase of pessimistic expectations. Here we confine ourselves to the phenomenological analysis of coordination of expectations. Whether this synchronization is justified in fundamental terms by the spillovers between countries in real economic activity, or whether it constitutes another, psychological factor of contagion, should be investigated in subsequent research.

This study contributes to the understanding of cross- correlations between economic and business sentiment indices worldwide. We aim to answer three main research questions: (i) how many statistically significant common factors can we extract from the joint dynamics of the sentiment indices worldwide; (ii) how well do these common factors account for the dynamics of the individual indices; and (iii) how does the weight of these factors change over time?

We analyze two data sets, i.e. the Business Confidence Index (BCI) and the Economic Sentiment Indices (ESI) ¹. In terms of methods, instead of using traditional approaches based on econometric models, we employ Random Matrix Theory (see, for example, Laloux et al., 1999; Bouchaud and Potters, 2009) and Principal Component Analysis (see, for example, Jolliffe, 1986; Billio et al., 2012; Wang et al., 2011) to investigate the dynamics of the correlation matrix of country-specific sentiment/confidence indices. We extract the hidden factors encoded in the empirical correlations across countries by analyzing the group of eigenvalues (and their corresponding eigenvectors) deviating from the random bulk. In this way, we can capture the evolution of the statistically significant factors underlying the dynamics of the correlation matrix. The extent to which different countries are affected by these factors can be thought of as the risk of sentiment contagion that the individual countries are facing during a particular period.

This paper is structured as follows. In Sec. 7.2 we briefly describe the data and methods employed in our study. Sec. 7.3 reports our main findings. Discussions and concluding remarks are found in Sec. 7.4.

¹See the next section for a more detailed description of the two data sets.

7.2 Data and Methods

7.2.1 Data

We consider two data sets containing country-specific sentiment indices. The first data set is collected by the OECD, which consists of all the OECD members and several other countries including China, Russia, India, Turkey, and Brazil. We name this data set OECD⁺. The data set captures the Business Confidence Index (BCI) measured monthly for each country. The index is based on the entrepreneurs' assessments of their current production, orders and stocks, as well as on their expectations for the immediate future (e.g. OECD, 2016, ²). To avoid the problem of missing data in some reported countries, we confine our analysis to the period from January 2002 until the end of 2015. This gives us data on the monthly business confidence indices in 33 countries.

The second data set reports the Economic Sentiment Indices (ESI) of Eurozone members and other European countries. The ESI summarizes consumer confidence, as well as the developments and expectations in the other surveyed sectors, i.e. industry (manufacturing), services, retail trade and construction sectors (e.g. EC, 2016, ³). In our analysis, we name this group of countries EU⁺. We use this data set for the period from January 1997 to December 2015, which gives us 24 monthly economic sentiment indices associated with 24 European countries.

7.2.2 Methods

Correlation matrix

Given the reported N indices for every month $\{SI_{i,t}\}_{i=1:N}$ from time $t=1$ to $t=T$, we apply a standard normalization procedure to the data ⁴. First, we consider the difference in logs across periods

$$I_{i,t} = \ln(SI_{i,t+1}) - \ln(SI_{i,t}). \quad (7.1)$$

As a second step, we define the normalized log-sentiment index for the time horizon T as

$$X_{i,t} = \frac{I_{i,t} - \langle I_{i,t} \rangle}{\sigma_{i,t}}, \quad (7.2)$$

²See OECD. 2016. Business Confidence Index (indicator). [doi:10.1787/3092dc4f-en](https://doi.org/10.1787/3092dc4f-en) (Accessed on 29 January 2016).

³See EC. 2016. Economic Sentiment Index (ESI). http://ec.europa.eu/economy_finance/db_indicators/surveys/index_en.htm (Accessed on 29 January 2016).

⁴In our study, we choose $T=36$ (months), which satisfies the condition that $T > N$.

where $\langle I_i(t) \rangle$ and $\sigma_i(t)$ are respectively the time average and the standard deviation of $I_i(t)$ over the time horizon T . Now we have $\langle X_i \rangle = 0$ and $\text{Var}(X_i) = 1$. Next, we consider the rectangular matrix $X = \{X_{i,t}\}_{N \times T}$ and the associated correlation matrix of the N normalized log-sentiment indices

$$C = \{C_{ij}\}_{N \times N} = \frac{1}{T} X X^T, \quad (7.3)$$

where the notation X^T stands for the matrix transposition of X . The value of C_{ij} denotes the correlation between country index i and j , where $-1 < C_{ij} < 1$, for $1 \leq i < j \leq N$. Note that, for any i we always have $C_{ii} = 1$. In case $C_{ij} > 0$ (< 0) the two countries i and j are positively (negatively) empirically correlated, while $C_{ij} = 0$ indicates a lack of any correlation.

Similarity matrix

One of the methods we use, to study the central question of the evolution of the sentiment correlation matrix over time, is the method proposed in Münnix et al. (2012), which is often applied when identifying states of stock markets⁵. The main idea is to come up with a measure of the similarity between correlation matrices from different periods. Suppose we observe two correlation matrices $C(t_1)$ and $C(t_2)$ associated with the two distinct periods t_1 and t_2 from the sample $\{1, 2, \dots, \mathcal{T}\}$, then the similarity S between those two matrices is defined as

$$S_{t_1, t_2} = \langle |C(t_1) - C(t_2)| \rangle, \quad (7.4)$$

where $|\dots|$ is the notation for the absolute value. Note that a higher value of S_{t_1, t_2} indicates that the “distance” between two correlation matrices is higher; in contrast, a smaller value of S_{t_1, t_2} reveals a higher level of similarity between the two matrices.

Random Matrix Theory

RMT, which was originally developed in nuclear physics by Wigner and Dirac to explain complex quantum systems, has emerged as one of the most important techniques for extracting latent information embedded in empirical correlations from the financial sector (e.g. Laloux et al., 1999; Laloux et al., 2000; Plerou et al., 2002; Kim and Jeong, 2005; Meng et al., 2014; Jiang et al., 2014; Uechi et al., 2015; MacMahon and Garlaschelli, 2015)⁶. Surprisingly, the applications of RMT in macroeconomic time series are very limited. Only a few studies, such as the studies by Ormerod and Mounfield (2002) and Ormerod (2008), have employed that technique to investigate the phenomenon of business cycle synchronization

⁵One can also use other similarity measures such as the one proposed in Münnix et al. (2010).

⁶We suggest the readers to, for instance, Bouchaud and Potters (2009) for a more detailed review of the financial applications of RMT.

over time.

Define $\{\lambda_i\}_{i=1}^{i=N}$ to be the eigenvalues of the correlation matrix C and consider the probability density function of these eigenvalues

$$\rho_C(\lambda) = \frac{dn(\lambda)}{d(\lambda)}, \quad (7.5)$$

where $n(\lambda)$ is the number of eigenvalues of C less than λ .

According to RMT, if all $X_{it} \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2)$, for $N, T \rightarrow \infty$ and $Q = \frac{T}{N} \rightarrow a = \text{constant} > 1$, the probability density function $\rho_C(\lambda)$ of eigenvalue λ will follow the Marchenko-Pastur (M-P) law

$$\rho_C(\lambda) = \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max}^{RMT} - \lambda)(\lambda - \lambda_{min}^{RMT})}}{\lambda}, \text{ for } \lambda_{min}^{RMT} \leq \lambda \leq \lambda_{max}^{RMT}, \quad (7.6)$$

and $\rho_C(\lambda) = 0$ elsewhere, with λ_{max}^{RMT} and λ_{min}^{RMT} are respectively the upper and lower bounds of eigenvalues associated with a random correlation matrix with the same variance and the same Q . According to RMT these bounds are given by

$$\lambda_{max}^{RMT} = \sigma^2(1 + \sqrt{1/Q})^2, \text{ and } \lambda_{min}^{RMT} = \sigma^2(1 - \sqrt{1/Q})^2. \quad (7.7)$$

We are interested in the latent information encoded in the eigenvectors corresponding to the largest eigenvalues deviating from the bulk of eigenvalues associated with a random correlation matrix with the same variance and the same Q . Suppose $\lambda_1 > \dots > \lambda_k > \lambda_{max}^{RMT} > \dots > \lambda_N$ are the eigenvalues of the empirical correlation matrix C in descending order and their corresponding eigenvectors are u_1, u_2, \dots, u_N . The elements of the eigenvector u_1 can be interpreted as the effect of the strongest factor (extracted from the correlation matrix) on all country-specific indices (see, for example, Plerou et al., 2002). In the following, we will be referring to this factor as the “market mode” or the “market factor”. In our study, we will investigate the temporal dynamics of the largest eigenvalues (larger than λ_{max}^{RMT}) and their corresponding eigenvectors, in order to identify periods with distinct cross-country sentiment correlation structures, as well as, to quantify the systemic risk associated with these periods (see, for example, Billio et al., 2012; Zheng et al., 2012; Meng et al., 2014; Nobi and Lee, 2016).

Decomposition and noise filtering

Note that we can diagonalize the correlation matrix C as

$$C = U\Lambda U^T, \quad (7.8)$$

where $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_N\}$ and the matrix $\{\mathbf{U}\}_{N \times N}$ is orthonormal, whose i^{th} column is the normalized eigenvector u_i associated with λ_i . From Eq. (7.8) we have

$$\lambda_i = \mathbf{u}_i^T \mathbf{C} \mathbf{u}_i = \mathbf{u}_i^T \text{Cov}(\mathbf{X}_t) \mathbf{u}_i = \text{Var}(\mathbf{u}_i^T \mathbf{X}_t). \quad (7.9)$$

The total variance of \mathbf{X}_t is then

$$\sum_{i=1}^N \text{Var}(X_{i,t}) = N = \sum_{i=1}^N \lambda_i = \sum_{i=1}^N \text{Var}(\mathbf{u}_i^T \mathbf{X}_t). \quad (7.10)$$

Now we can see that λ_i indicates the portion of total variance of \mathbf{X}_t contributed by the principal component $y_{i,t} = \mathbf{u}_i^T \mathbf{X}_t$ (e.g. Jolliffe, 1986).

We can rewrite Eq. (7.8) as

$$\mathbf{C} = \sum_{i=1}^{i=N} \lambda_i \mathbf{u}_i \mathbf{u}_i^T. \quad (7.11)$$

The expression $\mathbf{C}^m = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T$ represents the part of the sentiment correlation structure accounted for by the market mode (recall that the eigenvalues are indexed in descending order). We can filter the market mode away from \mathbf{C} . Following Kim and Jeong (2005), we define the filtered correlation matrix

$$\mathbf{M} = \mathbf{C} - \mathbf{C}^m. \quad (7.12)$$

From Eq. (7.11), we can express \mathbf{M} in the following way

$$\mathbf{M} = \mathbf{C} - \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T = \sum_{i=2}^{i=N} \lambda_i \mathbf{u}_i \mathbf{u}_i^T. \quad (7.13)$$

The latent information encoded in the eigenvectors of the second largest eigenvalue can also be useful if it is still large enough not to fall within the random bulk (i.e. if $\lambda_2 > \lambda_{max}^{RMT}$). In general, information embedded in other eigenvalues larger than λ_{max}^{RMT} is associated with important factors other than the market mode. In that case, the correlation matrix can be decomposed as

$$\mathbf{C} = \mathbf{C}^m + \mathbf{C}^g + [\mathbf{C} - \mathbf{C}^m + \mathbf{C}^g], \quad (7.14)$$

where \mathbf{C}^g accounts for correlations captured by the second most important factor. For instance, in the analysis of stock markets, it is repeatedly suggested that the sectoral component can be captured by the eigenvectors associated with the second largest eigenvalues.

In our study, we can think of the cultural and economic peculiarities of particular countries or groups of countries (e.g. emerging markets) as being such a factor.

Absorption ratio

From Eq. (7.8) and Eq. (7.9), the absorption ratios are given by

$$E_i = \sum_{j=1}^i \lambda_j/N \quad (\text{for } i = 1, 2, \dots, N). \quad (7.15)$$

E_i represents the fraction of the total variance of X_t explained by the first i principal components (since $\sum_{j=1}^N \lambda_j/N = \frac{\text{trace}(C)}{N} = 1$, we always have $E_N = 1$). What we are interested in, are comparisons between E_1, \dots, E_k and $E_N = 1$ (i.e. E_N stands for 100% of the variance as shown in Eq. (7.10)), where k is the largest integer for which $\lambda_k > \lambda_{max}^{RMT}$ is true. Besides using the average of correlations, the absorption ratios can be used to infer the systemic risk in the market (see, for example, Pukthuanthong and Roll, 2009; Billio et al., 2012; Zheng et al., 2012; Meng et al., 2014). For instance, high values of E_1 associated with a high λ_1 signal a strong co-movement of the individual sentiment indices, which implies a high systemic risk.

Inverse Partition Ratio

The inverse of the Inverse Partition Ratio (IPR) measures the number of eigenvector components (i.e. countries) strongly associated with a particular factor (the market mode, for example). It is defined as

$$\text{IPR}(i) = \sum_{j=1}^N u_i(j)^4. \quad (7.16)$$

Recalling that the elements of each eigenvector are normalized, i.e. $\sum_{j=1}^N u_i(j)^2 = 1$ ($\forall i = 1, 2, \dots, N$), it is easy to show that for all $i = 1, 2, \dots, N$, we have

$$\frac{1}{N} \leq \text{IPR}(i) \leq 1, \quad (7.17)$$

where $\text{IPR}(i) = \frac{1}{N}$ if and only if $u_i(j)^2 = \frac{1}{N}$ for all $j = 1, 2, \dots, N$; while $\text{IPR}(i) = 1$ if and only if only one element of the eigenvector u_i is different than zero, which implies that only this element (country) contributes to this particular factor. Overall, the inverse of the IPR indicates the number of eigenvector components that contribute significantly to that eigenvector. More specifically, a low IPR indicates that countries contribute more equally. In contrast, a large IPR would imply that the factor is driven by the dynamics of a small number of countries.

7.3 Findings

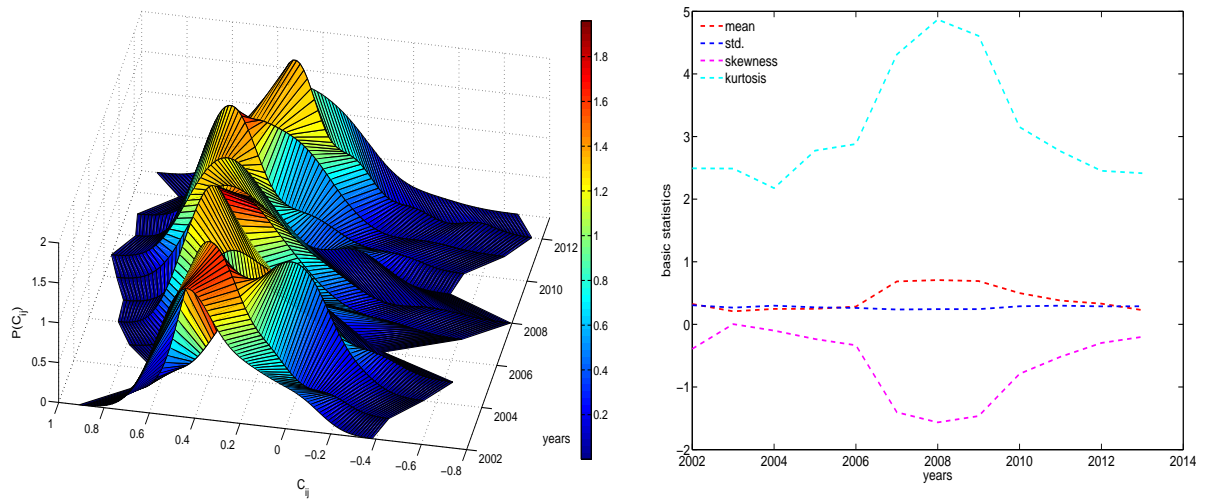
We report the temporal dynamics of the distribution of correlations between sentiment indices in Figure (7.1) for the BCI data (OECD⁺ group) and in Figure (7.2) for the ESI data (EU⁺ group)⁷. Our first observation is that the distribution of C_{ij} is generally asymmetric, and its shape is not stable over time. A noticeable change can be easily detected for the period of the financial crisis (2007 to 2009). More specifically, the average of correlations and the kurtosis increase during that time, while the skewness decreases significantly. In addition, we find that for all years, in the case of the BCI data, the average of correlations is always positive and the distribution always is left-skewed, signaling that the mass of the distribution of correlations is concentrated on the positive side. This implies that, overall, countries tend to be more positively than negatively correlated (see, for example, Plerou et al., 2002). A similar observation can be made for the ESI data, except for several years, during which a positive skewness is observed (in particular, around the period when the Eurozone was implemented). We provide the following potential explanation for the increased number of negative correlations during that period. Before the introduction of the Euro the interest rates in the “periphery” (Spain, Italy, Portugal, etc.) were much higher than those in the “core” (Germany, Netherlands, France, etc.) of the monetary union. Thus, during the implementation of the Eurozone, the sentiment in the “periphery” of the union might have been positively affected by the convergence of the interest rates across the Eurozone members, while the effect on the sentiment in the “core” might have been negative⁸. In other words, the convergence implies an increase of interest rates in the “core” and a fall in the “periphery”, which might result in opposing sentiment dynamics in the two areas of the union. In this rare case, the global information signal can have implications that differ across countries.

Since we observe that the sentiment correlation matrix is not stable over time, the question of how to identify the different states of C comes to the fore. In the previous section, we have introduced a method to quantify the similarity between correlation matrices (see Eq. (7.4)). This method allows us to identify particular states of the sentiment correlation matrix.

Figure (7.3) shows the similarity between the temporal correlation matrices for the BCI data and for the ESI data, respectively in panels (a) and (b). Three states can be identified

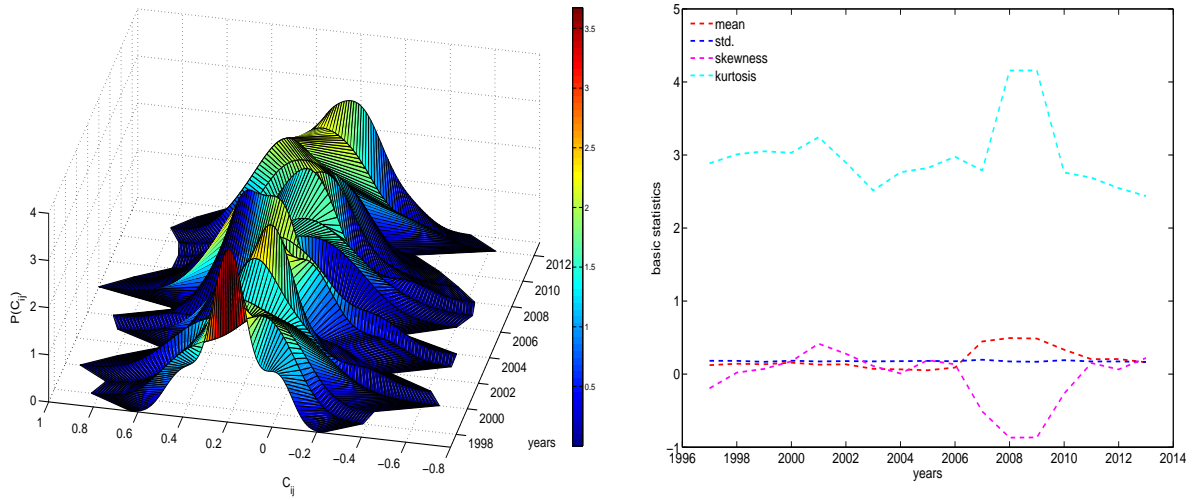
⁷The entries of the correlation matrix for each year have simply been pooled, after which a kernel density estimator has been used to arrive at a distribution in a particular year.

⁸For a discussion of the interest rate convergence in Eurozone see, for example, Arghyrou et al., 2009.



(a) Distribution of C , BCI data, OECD⁺ group (b) Statistics of C , BCI data, OECD⁺ group

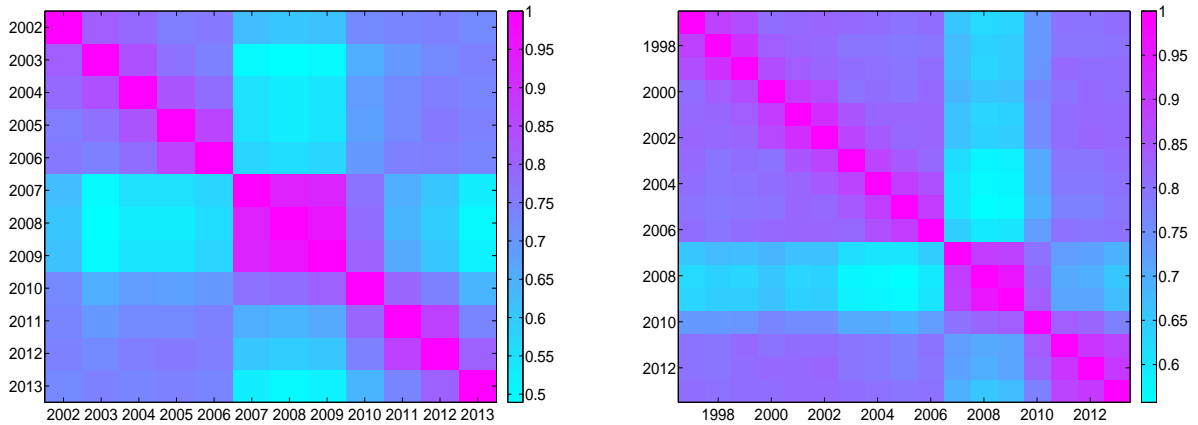
Figure 7.1: Evolution of the distribution of the elements of C , for BCI data in the OECD⁺ group. Panel (a) shows the distribution of the elements of C from 2002 to 2013. Panel (b) reports the basic statistical indicators of the elements of C including the mean, standard deviation (std.), skewness, and kurtosis.



(a) Distribution of C , ESI data, EU⁺ group (b) Statistics of C , ESI data, EU⁺ group

Figure 7.2: Evolution of the distribution of the elements of C , for ESI data in the EU⁺ group. Panel (a) shows the distribution of the elements of C from 1997 to 2013. Panel (b) reports the basic statistical indicators of the elements of C including the mean, standard deviation, skewness, and kurtosis.

from these panels, i.e. before 2007, from 2007 to 2009, and after 2009. We can see that the period from 2007 to 2009 is very homogeneous with respect to the correlation structures observed. The correlation matrices from the years 2007, 2008 and 2009 are very similar



(a) Similarity among C, BCI data, OECD⁺ group (b) Similarity among C, ESI data, EU⁺ group

Figure 7.3: Identifying states of correlation matrix using similarity-based analysis. Panel (a) shows the similarity among correlation matrices C from 2002 to 2013, for BCI data in the OECD⁺ group. Panel (b) shows the similarity among correlation matrices C from 1997 to 2013, for ESI data in the EU⁺ group.

to each other compared to matrices from other periods. This is consistent with what we have found during our analysis of the evolution of the basic statistics of the correlation matrices over time. In addition, in panel (b) we see that the correlation matrices for the EU⁺ group associated with the period of the European debt crisis (2011 to 2013) exhibit strong similarities relative to correlation matrices from other periods. We can thus conclude that for the EU⁺ we can detect an additional distinct state of the correlation matrix associated with the time of the debt crisis in Europe. In the following, we are going to look more closely at what these distinct states are characterized by.

We start by investigating the spectrum of the correlation matrix and its evolution over time. In Figure (7.4) and Figure (7.5), panel (a) we see that the largest eigenvalue λ_1 is typically more than three times larger than the upper bound λ_{max}^{RMT} for the OECD⁺ group, and more than 1.3 times larger for the EU⁺ group. In all years, λ_1 always deviates from the random bulk associated with the M-P law. Figure (7.6) shows the distribution of the eigenvalues of a random correlation matrix (with the same variance and the same Q) compared to the actual distribution of the eigenvalues in 2007 for both groups of countries. For the EU⁺ group, during the whole sample period, only the first eigenvalue λ_1 is larger than the upper bound λ_{max}^{RMT} , while for the OECD⁺ group, in some years, a second eigenvalue λ_2 crosses this upper bound ⁹. The second factor may be interpreted as a group factor. On some

⁹We also observe that λ_1 is always very similar to $N\langle C_{ij} \rangle$ (where $\langle C_{ij} \rangle$ stands for the pooled average of C), which supports the presence of one common factor affecting all indices.

rare occasions (e.g. like in the years 2003, 2004 and 2010), particular countries (including some “emerging markets”) can have sentiment dynamics opposing those of the rest of the world. We can detect this by looking at the eigenvector elements associated with countries like Mexico, Turkey, Slovakia, Russia, China for in some years and recognizing that these elements have the opposite sign of the elements associated with countries from the rest of the world. Some developed countries like New Zealand and Australia for instance, also show a similar behavior. All this suggests that the cross-country sentiment dynamics are driven primarily by a single factor (the market mode) and only on rare occasions does a second factor become marginally significant. In the following, we will thus be concentrating on the market mode and on the relationship that countries or groups of countries have with it.

The evolution of the importance of the market factor for the cross-country sentiment dynamics can also be observed in Figure (7.4) and Figure (7.5). We see that during the financial crisis the importance of the market factor becomes overwhelming since both the largest eigenvalue and the associated absorption ratio jump by approximately 100%. Since the largest eigenvalue and the associated absorption ratio increase together, we can say that the systemic worldwide component of sentiment was high during the period 2007 to 2009. We can also see that, for the EU^+ group, the absorption ratios after 2009 are still higher than during the period before the financial crisis. The perceived threat to the Eurozone’s stability stemming from the risk of sovereign default of some member states might have prevented individual sentiment indices in Europe from diverging from the market mode.

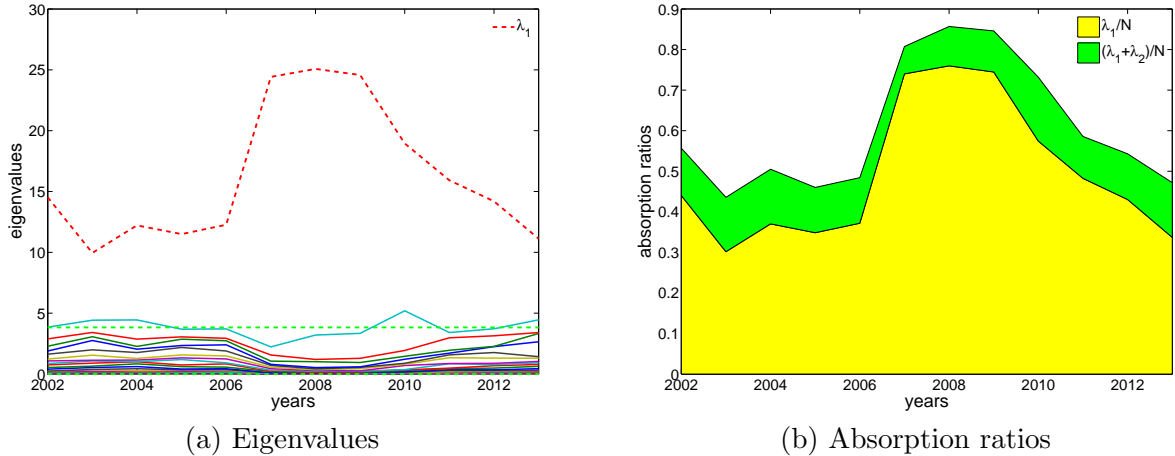


Figure 7.4: Evolution of eigenvalues and absorption ratios for BCI data in the $OECD^+$ group. In panel (a), the red dashed line stands for the largest eigenvalue, the green dashed lines stand for the interval $[\lambda_{min}^{RMT}, \lambda_{max}^{RMT}]$ explained by RMT. Panel (b) shows the absorption ratios associated with the first and the second largest eigenvalues.

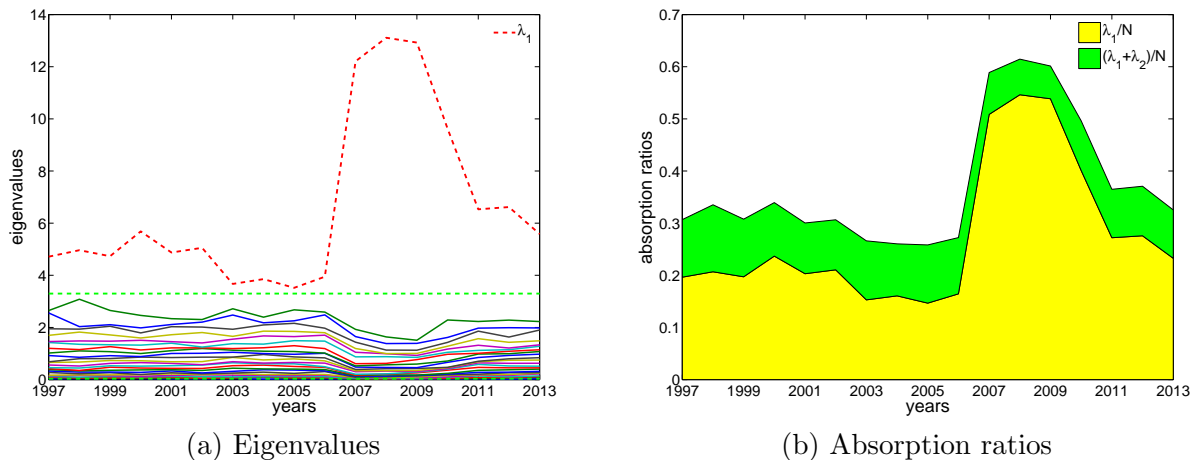


Figure 7.5: Evolution of eigenvalues and absorption ratios for ESI data in the EU^+ group. In panel (a), the red dashed line stands for the largest eigenvalue, the green dashed lines stand for interval $[\lambda_{min}^{RMT}, \lambda_{max}^{RMT}]$ explained by RMT. Panel (b) shows the absorption ratios associated with the first and the second largest eigenvalues.

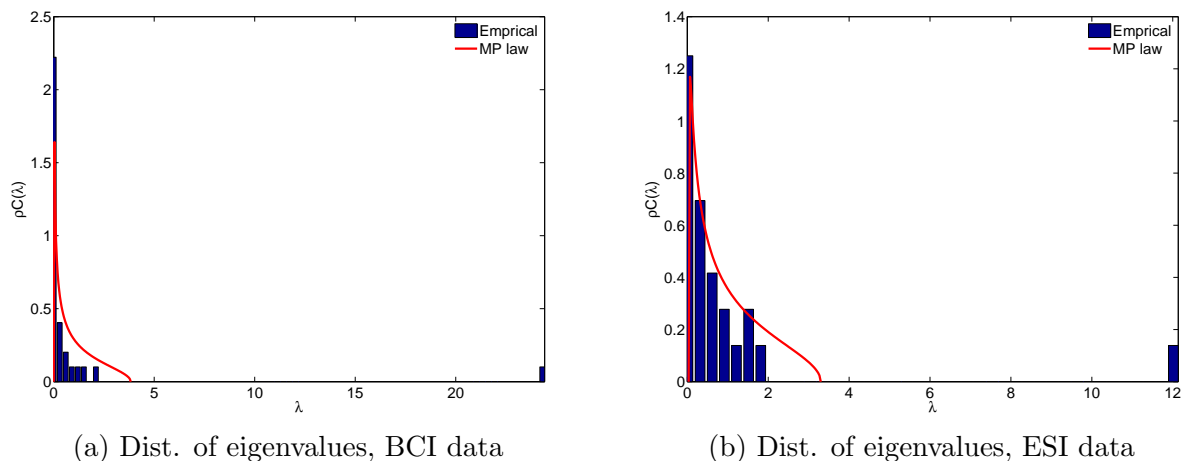


Figure 7.6: Distribution of eigenvalues of C in 2007, compared with RMT. Panel (a) for BCI data in the $OECD^+$ group. Panel (b) for ESI data in the EU^+ group.

Next, let us look in more detail at the components of the eigenvector u_1 associated with the market mode. We find that in the period of financial crisis (2007 to 2009), the components of u_1 become more homogeneous, evidencing the synchronization of the sentiment indices around the world (see Figures (7.7), (7.8), and (7.9)). This result holds true for both the EU^+ and $OECD^+$ groups. Still, a few countries like China, South Africa or New Zealand seem to be somewhat less influenced by the market mode during the crisis. During normal times, much more divergent behavior is observed. More specifically, in case of the $OECD^+$

group, countries like Italy, France, Belgium, Hungary, U.K., Austria, Slovenia, Denmark, Chile, Netherlands or Germany contribute the most to the market mode, while the sentiment dynamics in other countries like Finland, Korea, Mexico, New Zealand, Turkey, Slovakia, Australia, Russia, China or India can be divergent to a certain extent. For the EU⁺ group, countries like Belgium, Denmark, Germany, France, Italy, Netherlands or Sweden contribute the most to the market mode, while the sentiment dynamics in other countries like Bulgaria, Latvia, Poland, Romania, Slovakia or Finland can be divergent to a certain extent.

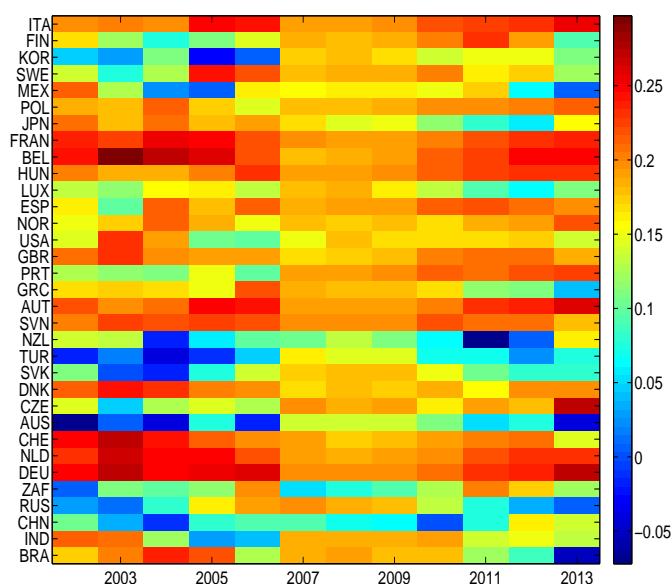


Figure 7.7: Evolution of the eigenvector components of the largest eigenvalue (λ_1), for BCI data in the OECD⁺ group. Without loss of generality, we assume that the sign of the eigenvector element that has the largest absolute value is non-negative. The financial crisis from 2007 to 2009 is captured by drastic changes in the the largest eigenvector components, i.e. they become more homogeneous during that period.

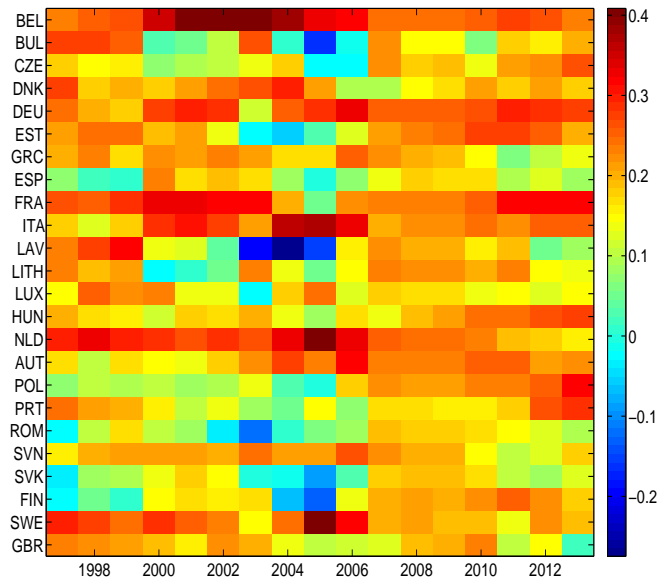
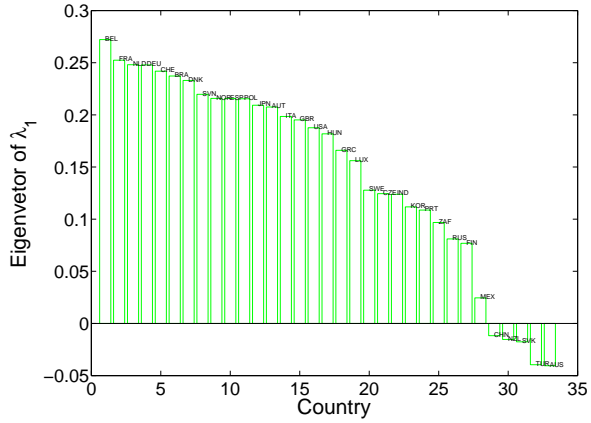
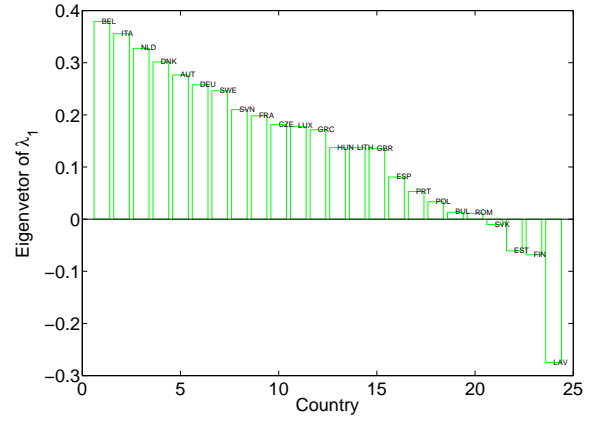


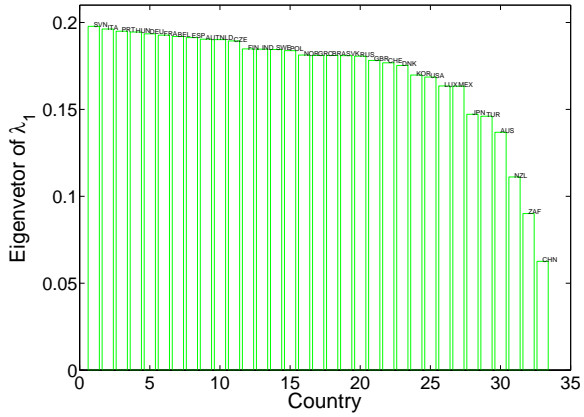
Figure 7.8: Evolution of the eigenvector components of the largest eigenvalue (λ_1), for ESI data in the EU⁺ group. Without loss of generality, we assume that the sign of the eigenvector element that has the largest absolute value is non-negative. The financial crisis from 2007 to 2009 is captured by drastic changes in the the largest eigenvector components, i.e. they become more homogeneous during that period.



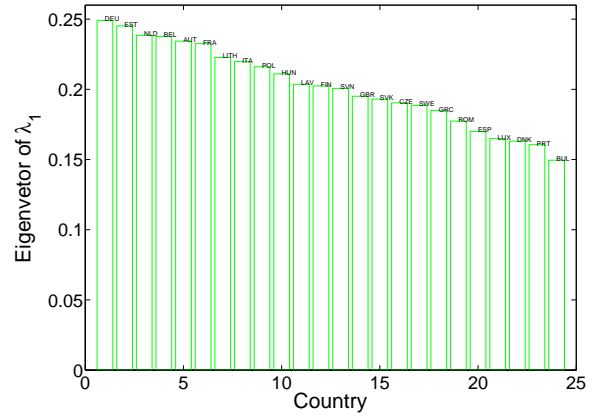
(a) Eivec1 in 2003, BCI



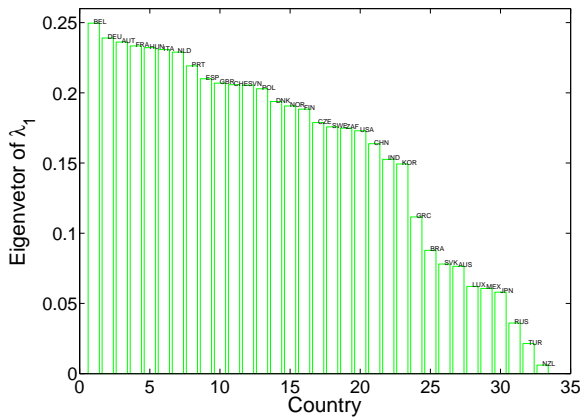
(b) Eivec1 in 2003, ESI



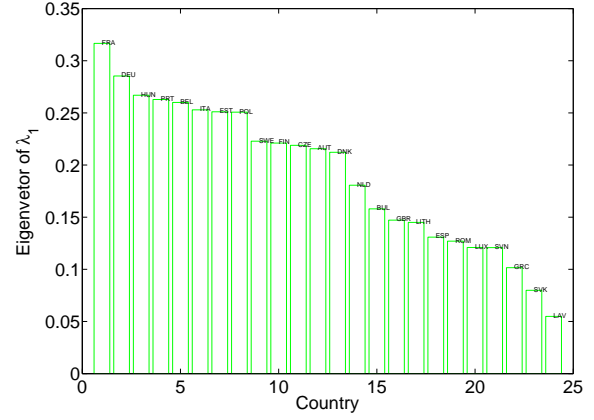
(c) Eivec1 in 2007, BCI



(d) Eivec1 in 2007, ESI



(e) Eivec1 in 2011, BCI



(f) Eivec1 in 2011, ESI

Figure 7.9: Eigenvector components of λ_1 (Eivec1). Without loss of generality, we assume that the sign of the eigenvector element that has the largest absolute value is non-negative. The three left panels (a), (c), and (e) are for BCI data in the OECD⁺ group, in 2003, 2007, and 2011. The three right panels (b), (d), and (f) are for ESI data in the EU⁺ group, in 2003, 2007, and 2011. The country code associated with each eigenvector component is also reported. We can see that in the three example years, the eigenvector components of λ_1 are more homogeneous in 2007.

We further analyze the eigenvectors by calculating the IPR. Figure (7.10) shows the IPR versus the corresponding eigenvalues for the years 2003, 2007, and 2011 as examples. Panels (a) and (b) do this for the OECD⁺ and for the EU⁺ groups, respectively. We find that for those years, in which if the largest eigenvalue strongly deviates from the random bulk, the associated eigenvector also exhibits the largest inverse of IPR, meaning that the sentiment dynamics in the majority of the countries is influenced by the market mode.

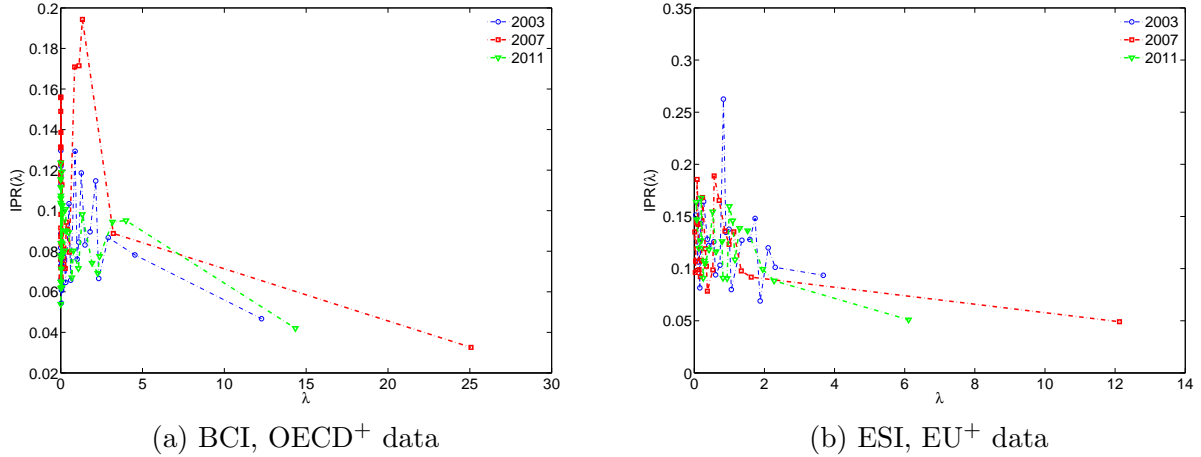


Figure 7.10: Comparison of Inverse Participation Ratios of the eigenvectors of the correlation matrix for periods before the financial crisis, during the financial crisis, and after the financial crisis. Panel (a) shows the IPR for BCI in the OECD⁺ group, panel (b) shows the IPR for ESI in the EU⁺ group. In both panels, three years, (i) 2003, (ii) 2007, and (iii) 2011, are chosen as the examples.

Above, we have detected a common factor underlying the dynamics of the sentiment indices. Now, we will compare the sentiment correlation matrix over time before and after filtering the effect of that factor. The results are shown in Figures (7.11) and (7.12), respectively for the OECD⁺ and EU⁺ groups. Overall, the raw correlations are significantly reduced after the information encoded in the largest eigenvalue and its corresponding eigenvector is subtracted. In addition, we can see that at the time of financial crisis (exemplified here by the graph for 2007, which is the same for 2008 and 2009), the raw correlations between countries increase but their filtered counterparts exhibit reduced correlations. In other words, the increase in the raw correlations were accompanied by an increase in the fraction of the correlations associated with the market mode. Note that for the OECD⁺ group, some significant correlations still appear after the filtering in some years of the sample period. They are actually mainly contributed by emerging markets like China or India, for which the informational signal associated with the collapse of the US housing market might have

been less relevant due to their limited exposure to toxic securities.

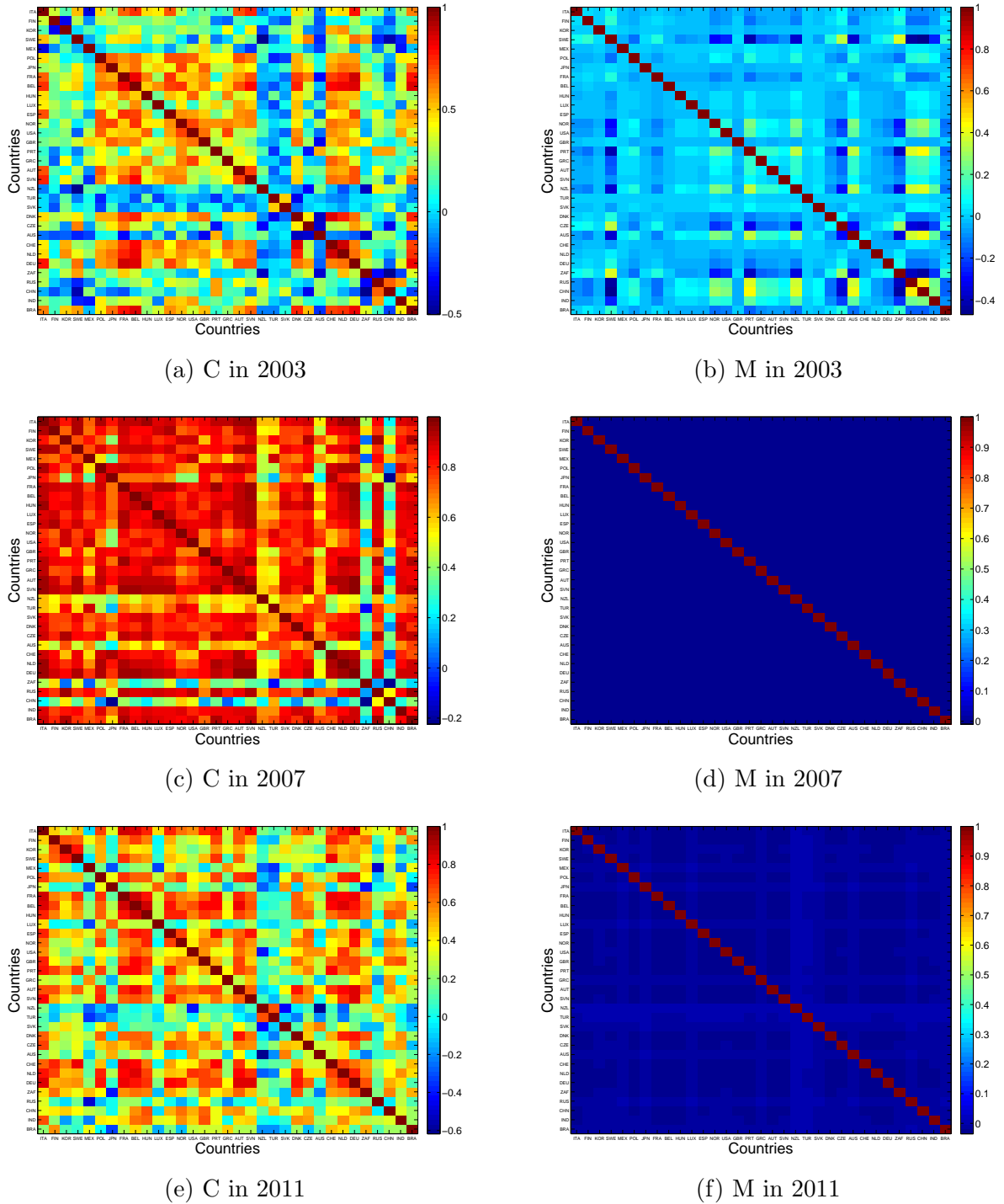
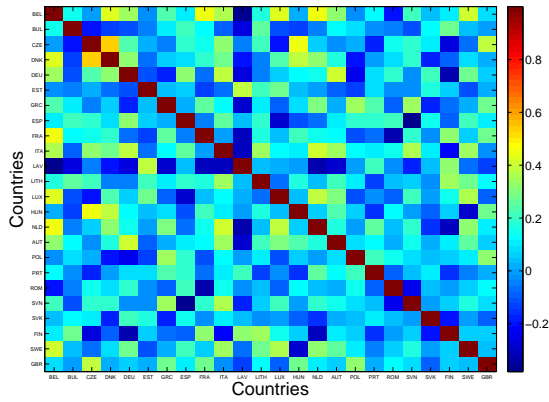
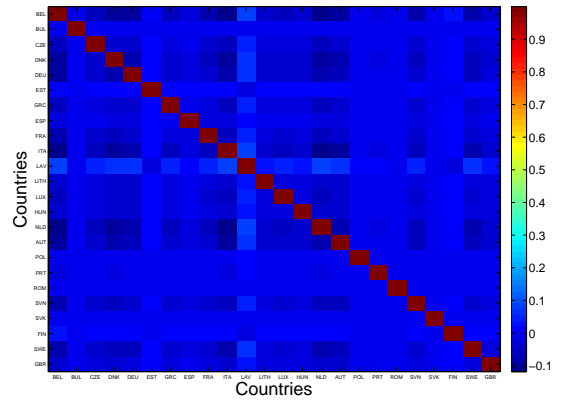


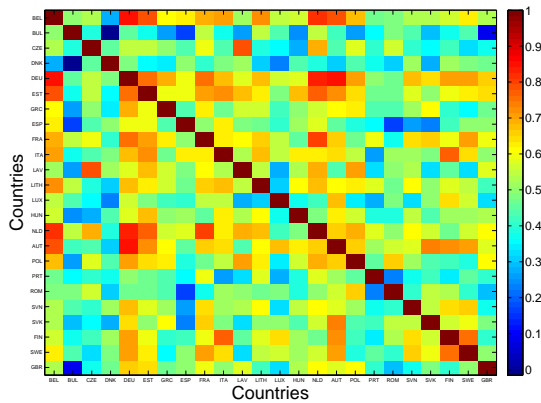
Figure 7.11: Correlations between countries, for BCI data in the OECD⁺ group. The three left panels (a), (c), and (e) are the raw correlation matrices in 2003, 2007, and 2011. The three right panels (b), (d), and (f) are the correlation matrices filtered by the RMT method in 2003, 2007, and 2011.



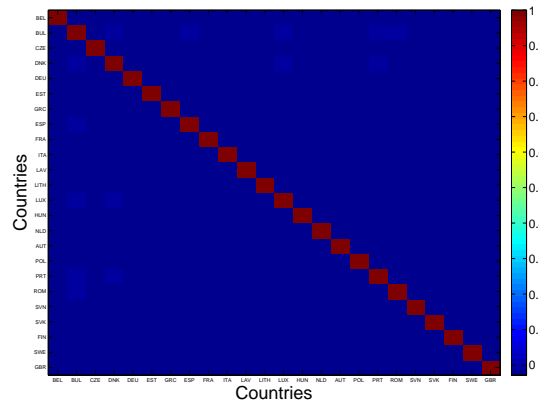
(a) C in 2003



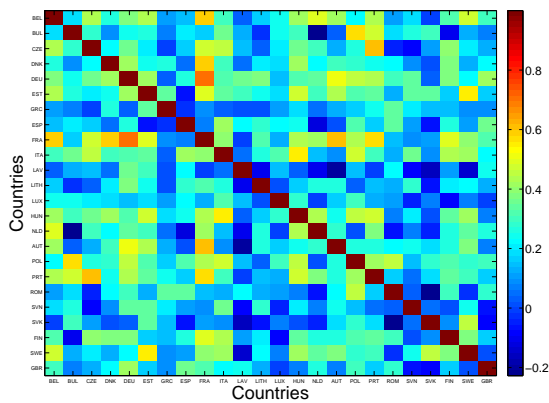
(b) M in 2003



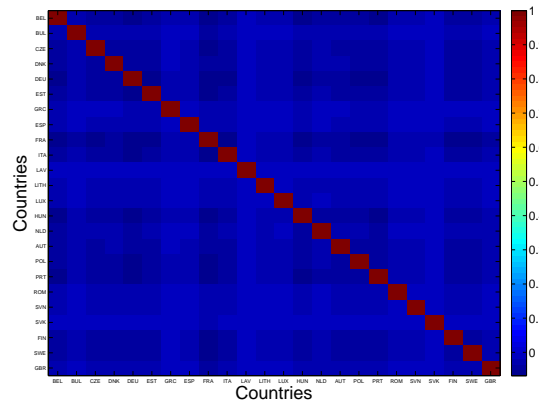
(c) C in 2007



(d) M in 2007



(e) C in 2011



(f) M in 2011

Figure 7.12: Correlations between countries, for ESI data in the EU⁺ group. The three left panels (a), (c), and (e) are the raw correlation matrices in 2003, 2007, and 2011. The three right panels (b), (d), and (f) are the correlation matrices filtered by the RMT method in 2003, 2007, and 2011.

7.4 Conclusions

In this paper, we have analyzed the evolution of the empirical correlations between the macroeconomic sentiment indices in different countries. Overall, we observe different states in the correlation structure, associated with a varying importance of the common factor (“market mode”). The correlations between indices are significantly reduced after the effects of that common factor are subtracted.

It would seem that many of the information signals worldwide have a common component, since generally sentiment indices tend to comove. During normal times, however, the sentiment in various countries or groups of countries can “resist” the common factor or can even, on rare occasions, “swim against the tide”. This is the case for some emerging markets like China, Turkey and other countries like Australia, New Zealand. However, in the presence of strong global information signals, we observe a strong synchronization of the sentiment dynamics all over the world. We consider the collapse of the US housing market (2007-2009) as an example of such global signals. In the case of the Eurozone debt crisis (2011-2013), the sentiment synchronization is high only within Europe, which can be interpreted as an indication that the Eurozone debt crisis is not perceived as a global information signal in countries outside of Europe. Information signals can also cause the sentiments to diverge, if the information has different implications for particular countries or groups of countries. We consider the interest rate convergence associated with the establishment of the Eurozone around the year 2000 to be an example of such an effect.

Overall, we believe that RMT and principal component analysis of the ensemble of worldwide or regional sentiment data can reveal important information on the correlations between business and consumer sentiment in different countries. The tools and results presented in this paper should provide relevant input for business cycle forecasts and the analysis of international co-movements of macroeconomic activity.

7.5 References

Akerlof G. A., Shiller R. J. 2009. *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. Princeton University Press.

Arghyrou M. G., Gregoriou A. and Kontonikas A. 2009. Do real interest rates converge? Evidence from the European union. *Journal of International Financial Markets, Institutions and Money* 19 (3), pp. 447- 460.

Artis M., Chouliarakis G., Harischandra P. K. G. 2011. Business cycle synchronization

since 1880. *The Manchester School* 79 (2), pp. 173-207.

Billio M., Getmansky M., Lo A. W., Pelizzon L. 2012. Econometric measures of systemic risk in the finance and insurance sectors. *Journal of Financial Economics* 104 (3), pp. 535-559.

Bouchaud J. -P., Potters M. 2009. Financial applications of random matrix theory: a short review. Working Paper. Available at: [arXiv:0910.1205](https://arxiv.org/abs/0910.1205).

Bordo M. D., Helbling T. 2003. Have national business cycles become more synchronized?. NBER Working Paper 10130. Available at: <http://www.nber.org/papers/w10130>.

Bordo M. D., Helbling T. 2011. International business cycle synchronization in historical perspective. *The Manchester School* 79 (2), pp. 208-238.

Carriero A., Marcellino M. 2011. Sectoral survey-based confidence indicators for Europe. *Oxford Bulletin of Economics and Statistics* 73 (2), pp. 175-206.

EC. 2016. Economic sentiment index (indicator). Available at: http://ec.europa.eu/economy_finance/db_indicators/surveys/index_en.htm (Accessed on 29 January 2016).

Ferroni F., Klaus B. 2015. Euro area business cycles in turbulent times: Convergence or decoupling. ECB Working Paper 1819. Available at: <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1819.en.pdf?8aacf9b049f3a7e6360b69d13e435844>.

Howrey E. P. 2001. The predictive power of the index of consumer sentiment. *Brookings Papers on Economic Activity* 2001 (1), pp. 175-207.

Jiang X. F., Chen T. T., Zheng B. 2014. Structure of local interactions in complex financial dynamics. *Scientific Reports* 4.

Jolliffe I.T. 1986. *Principal component analysis*. Springer-Verlag, New York.

Keynes J. M. 1936. *The general theory of employment, interest and money*. New York: Harcourt, Brace & World.

Kim D. -H., Jeong H. 2005. Systematic analysis of group identification in stock markets. *Physical Review E* 72 (4).

Kose M. A., Otrok C., Prasad E. 2012. Global business cycles: Convergence or decoupling?. *International Economic Review* 53 (2), pp. 511-538.

Laloux L., Cizeau P., Bouchaud J. -P., Potters M. 1999. Noise dressing of financial correlation matrices. *Physical Review Letters* 83 (7), pp.1467-1470.

Laloux L., Cizeau P., Potters M., Bouchaud J. -P. 2000. Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance* 3 (3), pp. 391-397.

MacMahon M., Garlaschelli D. 2015. Community detection for correlation matrices.

Physical Review X 5 (2).

Milani F. 2011. Expectation shocks and learning as drivers of the business cycle. *The Economic Journal* 121 (552), pp. 379-401.

Milani F. 2014. Sentiment and the U.S. business cycle. Working Paper. Available at: <http://www.economics.uci.edu/files/docs/workingpapers/2014-15/14-15-04-1.pdf>.

Minsky H. 1977. The financial instability hypothesis: An interpretation of Keynes and an alternative to 'standard' theory. *Nebraska Journal of Economics and Business* 16 (1), pp. 5-16.

Meng H., Xie W. -J., Jiang Z. -Q., Podobnik B., Zhou W. -X., Stanley H. E. 2014. Systemic risk and spatiotemporal dynamics of the US housing market. *Scientific Reports* 4.

Münnix M. C., Shimada T., Schäfer R., Leyvraz F., Seligman T. H., Guhr T., Stanley H. E. 2012. Identifying states of a financial market. *Scientific Reports* 2.

Münnix M. C., Schäfer R., Grothe O. 2010. Estimating correlation and covariance matrices by weighting of market similarity. Working Paper. Available at: [arXiv:1006.5847](https://arxiv.org/abs/1006.5847).

Nobi A., Lee J. W. 2016. State and group dynamics of world stock market by principal component analysis. *Physica A: Statistical Mechanics and its Applications* 450, pp. 85-94.

OECD. 2016. Business confidence index (indicator). Available at: <https://data.oecd.org/leadind/business-confidence-index-bci.htm#indicator-chart> (Accessed on 29 January 2016).

Ormerod P., Mounfield C. 2002. The convergence of European business cycles 1978-2000. *Physica A: Statistical Mechanics and its Applications* 307 (3-4), pp. 494-504.

Ormerod P. 2008. Random matrix theory and macro-economic time-series: An Illustration using the evolution of business cycle synchronisation, 1886-2006. *Economics: The Open-Access, Open-Assessment E-Journal* 2, pp. 1-10.

Plerou V., Gopikrishnan P., Rosenow B., Amaral L. A. N., Guhr T., Stanley H. E. 2002. Random matrix approach to cross correlations in financial data. *Physical Review E* 65 (6).

Pukthuanthong K., Roll R. 2009. Global market integration: An alternative measure and its application. *Journal of Financial Economics* 94 (2), pp. 214-232.

Santero T., Westerlund N. 1996. Confidence indicators and their relationship to changes in economic activity. OECD Economics Department Working Papers, No. 170, OECD Publishing. Available at: <http://dx.doi.org/10.1787/537052766455>.

Taylor K., McNabb R. 2007. Business cycles and the role of confidence: Evidence for Europe. *Oxford Bulletin of Economics and Statistics* 69 (2), pp. 185-208.

Uechi L., Akutsu T., Stanley H. E., Marcus A. J., Kenett D. Y. 2015. Sector dominance

ratio analysis of financial markets. *Physica A: Statistical Mechanics and its Applications* 421, pp. 488–509.

van Aarle B., Kappler M. 2012. Economic sentiment shocks and fluctuations in economic activity in the euro area and the USA. *Intereconomics* 47 (1), pp. 44-51.

Wang D., Podobnik B., Horvatic D., Stanley H. E. 2011. Quantifying and modeling long-range cross correlations in multiple time series with applications to world stock indices. *Physical Review E* 83 (4).

Zheng Z., Podobnik B., Feng L., Li B. 2012. Changes in cross-correlations as an indicator for systemic risk. *Scientific Reports* 2.

Summary and Outlook

This dissertation covers the empirical applications of network theory and Random Matrix Theory to large data sets of credit relationships in the interbank market as well as between banks and non-financial firms, and to economic and financial correlations.

In the first essay (chapter 2), we provide a comprehensive analysis of the second as well as the third order structural correlations in all versions (binary vs weighted and undirected vs directed) of the Italian e-MID network. Our findings suggest that the observed structural correlations can vary across different versions of the network. During our analysis of the evolution of the third order correlations among banks over time, we detected dramatic changes in the network structure surrounding the recent financial crisis in 2007. Moreover, by employing the various configuration models, we examined whether the information embedded in the observed degree sequence and/or the strength sequence can explain observed higher order structural correlations. The results show that, in the binary case, the degree sequence is informative in terms of explaining the main features of the structural correlations in the e-MID network. In the weighted case, the randomized ensembles produced by the Enhanced Configuration Models, which constrains both the degree as well as the strength sequences, have a much greater predictive power than the randomized ensembles produced by the Weighted Configuration Models, which constrains only the strength sequences. However, under scrutiny, both the binary as well as the weighted versions of the observed e-MID network do exhibit some non-random patterns that cannot completely be explained by the degree sequence(s) and/or strength sequence(s). One main feature not explained by the sequences of degrees and strengths of the network nodes themselves is the high level of clustering in the years preceding the crisis, i.e. the huge increase in various indirect exposures generated via more intensive interbank credit links. Interestingly, the distance between the predictions of the Weighted Configuration Models and of the Enhanced Configuration Models for the averages of the measures for the third order correlations continuously increases from the adoption of the Euro up until the financial crisis in 2007 and then sharply decreases after

that, revealing structural changes in the network associated with these two critical events. It also suggests that the importance of the degree sequence(s) and/or strength sequence(s) for the emergence of higher order correlation structures can vary over time. For future studies, it would be interesting to investigate whether some characteristics of banks correlate with the extracted hidden variables. In addition, we suggest that the role of various constraints for the emergence of higher order structural correlations (or motifs) and for the meso-scale network structures such as the core-periphery and community structures should be studied further.

In the second essay (chapter 3), I show another application of the configuration models in the identification of non-random patterns in portfolio overlaps and similarities between banks in two real world financial networks, i.e. the bank-bank projection network of the bank-firm credit market of Spain and the Italian e-MID network. In the binary case, almost all pairwise similarities are significantly removed when the effects of heterogeneity in the observed degree sequence are filtered out. In the weighted case, on the one hand, the Undirected Enhanced Configuration Model preserving both observed degrees as well as strengths seems to be a proper model used for filtering the structural similarities in the Italian e-MID network. On the other hand, in the case of the bank-firm credit network of Spain, many significant structural similarities still emerge even when the effects of both sequences are filtered out. In particular, a subset of banks having the largest and significant similarities mainly consists of the a group of commercial banks as well as a group of saving banks. These banks have a higher degree of similarities with those in the same category as well as with those in the other category. When more detailed information about the other characteristics of banks is available, further research should clarify other factors determining significant structural similarities between some banks. Future studies should also analyze structural similarities in other types of networks such as directed networks and bipartite networks. Besides, further extensions to other similarity measures may provide a more comprehensive assessment of the structural similarity in complex networks.

The third essay (chapter 4) deals with another structure of financial networks, in which interactions are established between different types of agents (i.e. the bipartite structure). We study the topological and structural properties of the bank-sector credit network of Spain. Among other results, we find that high degree banks tend to diversify their lending to almost all sectors, whereas low degree banks preferentially specialize their lending on high degree sectors. Similarly, high degree sectors tend to borrow from almost all banks, while low degree sectors prefer to borrow from high degree banks. We show that the intrinsic heterogeneity

in the distribution of the degrees of banks and sectors plays a crucial role for the emergence of this nested structure of interactions between banks and other sectors of the real economy. In addition, interestingly, the elements of the one-mode projection matrices indicating the portfolio overlaps between banks and between sectors are also mostly predicted by latent variables extracted from the degree sequence of the original bipartite network. This shows the relevance of the configuration models defined for the original bipartite structure instead of employing the null models defined for one-mode networks. In the weighted version, similar to the findings obtained from the first essay, we find that maintaining only the strength sequence can not replicate the main features of the network properties, while the reproduction of such features is considerably improved by adding the knowledge of the degree sequence. Nevertheless, we still observe that certain local properties can not simply be explained by the lower-order properties of the observed network like the degree sequence and/or the strength sequence. The presence of significant deviations from the various bipartite configuration models indicates that further studies are needed in order to capture such subtle differences. Furthermore, the evaluation of the topological and structural properties of other financial networks (e.g. bank-firm credit networks, investor-asset networks, and so forth) against the expected properties obtained from the various bipartite configuration models can be a fruitful direction for future research.

In the fourth essay (chapter 5), we move on to the multilayer architecture of joint exposures of banks when lending to non-financial firms. First, the bank-firm credit network of Spain is decomposed into different layers representing different industrial sectors. We then study the dependencies between layers based on normalized measures of overlaps of links and weights of banks between layers. Our findings show that the dependencies between layers of the observed network exhibit a very hierarchical structure. After that, we examine the role of various constraints in all layers for the emergence of such dependencies. To do that, we compare the observed values with the expected ones obtained from random graph models specifying only the total degree or the total strength in single layers, and from configuration models capturing the intrinsic heterogeneity in the observed degree sequence and/or strength sequence in single layers. We find that the bank-firm credit network of Spain has a significant, non-random structure of correlations between layers that cannot be explained by the primitive network properties alone. Particularly, in both versions of the network, a significant deviation from all null models is always observed in high levels of multiplexities. Additionally, in the weighted version, many smaller values of multiplexities also strongly deviate from the expected ones evaluated under the various weighted null models. Further

research considering the bipartite structure of the bank-firm credit relations in each industrial sector-based layer would provide a more comprehensive analysis of the role of the different constraints for the emergence of dependencies between layers. Furthermore, we suggest that future studies should also consider to investigate the structural dependencies in the multi-layer architecture of other financial networks, for instance in networks where different types of assets or financial transactions represent different layers.

The fifth essay (chapter 6) analyzes the structure of the cross-correlation matrices of banks' loan portfolios in the various lending layers (based on the maturity of loans) in the bank-firm credit market of Japan during the period from 1980 to 2012. Although from network theory's perspective, banks show high overlaps in their lending to firms, having applied the methods of RMT, I show that a majority of correlations between banks' loan portfolios are contributed by noise. In addition, a group of largest eigenvalues of the cross-correlation matrices always deviate from the spectrum predicted by RMT, indicating the presence of patterns in such correlations. Furthermore, during the bubble period in Japan, I find that banks' loan portfolios tend to be more correlated, implying a higher systemic risk in the market. Moreover, the binary version of bank-firm credit relationships also contain genuine information about the structure of the cross-correlation matrices of banks' loan portfolios. Obviously, the effects of the characteristics of banks on the formation of patterns in the correlation matrices in the various lending layers should be studied further. In addition, it is also interesting to analyze the mesoscopic structure in these matrices.

The sixth essay (chapter 7) analyzes the evolution of the empirical correlations between the macroeconomic sentiment indices in different countries. We observe that the dynamics of the sentiment indices across countries can be well explained by the evolution of a common component. However, during normal times, the sentiment in some countries can "resist" the common factor or can even, on rare occasions, "swim against the tide". Nevertheless, in the presence of strong global information signals like the collapse of the US housing market (2007-2009), we observe a strong synchronization of the sentiment dynamics all over the world. In the case of the Eurozone debt crisis (2011-2013), the sentiment synchronization is high only within Europe, which can be interpreted as an indication that the Eurozone debt crisis is not perceived as a global information signal in countries outside of Europe. We believe that a similar analysis of the empirical co-movements between other macroeconomic indicators of different countries can be a fruitful direction for future research. Additionally, the methods employed in this study can be used to decompose the correlations between macroeconomic variables into components.

Affirmation

I hereby affirm that I have completed my doctoral thesis entitled, “Empirical Applications of Network and Random Matrix Theories to Economic and Financial Complex Systems” entirely on my own and unassisted, and that I have specially marked all of the quotes I have used from other authors as well as those passages in my work that are extremely close to the thoughts presented by other authors, and listed the sources in accordance with the regulations I have been given.

February 6, 2017

Luu Duc Thi

CURRICULUM VITAE

PERSONAL DETAILS

- Full name: Duc Thi Luu
- Address (Current- For contact): Wilhelm-Seelig-Platz 1, Zimmer 504, Universität Kiel, Institut für Volkswirtschaftslehre, Olshausenstraße 40, D-24118 Kiel, GERMANY
- Email: luucthi@gmail.com or d.t.luu@economics.uni-kiel.de
- Date of Birth: 19/05/1984
- Country of Citizenship: VIETNAM

EDUCATION

- 10/2012-Now: Kiel University, Germany. PhD candidate in Quantitative Economics Program.
- 2010-2012: International Christian University, Tokyo, Japan. M.A. in Public Economics.
- 2009-2010: Leeds Metropolitan University (Leeds Beckett University), England and Hanoi Finance Academy, Vietnam. M.A. in International Trade and Finance.
- 2002-2006: Foreign Trade University, Vietnam. Degree: B.A. in International Economics.
- 1999-2002: Lam Son High School for the Gifted, ThanhHoa, Vietnam. Specialized in Mathematics.

SHORT TRAINING COURSES

- 08/2013-05/2014: Kiel Institute for World Economy. Participated in Advanced Studies Program in International Economics.
- 06/2012-09/2012: International Christian University, Tokyo, Japan. Global Leadership Studies Program.
- 04/2011-05/2011: India Institute for International Trade, New Delhi, India. Research Methodology in Trade.

RESEARCH INTERESTS

- Quantitative Economics and Finance, Financial Macroeconomics.
- Network Science, Complex System Analysis in Economics and Finance, Social Networks.
- Social Dynamic Interactions, Agent Based Models.
- Econophysics.

- Big Data, Data Mining.

CONFERENCES AND WORKSHOPS

Selected:

- Complex Networks: from theory to interdisciplinary applications, July 11-13, 2016, in Marseilles, France – POSTER.

- Statistical Physics of Financial and Economic Networks, 15-16 July 2016, University Paris 1, France – CONTRIBUTED TALK.

ACADEMIC REFEREES

1. Prof. Dr. Thomas Lux (First advisor)

University of Kiel, Germany

Email: lux@economics.uni-kiel.de